Published in final edited form as:

Pharmacogenomics. 2010 March ; 11(3): 305–308. doi:10.2217/pgs.10.6.

Genome-wide association studies in pharmacogenetics research debate

Kent R Bailey^{1,†} and Cheng Cheng²

¹Division of Biomedical Statistics and Informatics, Harwick 7, Mayo Clinic, 200, First Street SW, Rochester, MN 55905, USA

²Department of Biostatistics, St Jude Children's Research Hospital, Memphis, TN, USA

Abstract

Will genome-wide association studies (GWAS) 'work' for pharmacogenetics research? This question was the topic of a staged debate, with pro and con sides, aimed to bring out the strengths and weaknesses of GWAS for pharmacogenetics studies. After a full day of seminars at the Fifth Statistical Analysis Workshop of the Pharmacogenetics Research Network, the lively debate was held – appropriately – at Goonies Comedy Club in Rochester (MN, USA). The pro side emphasized that the many GWAS successes for identifying genetic variants associated with disease risk show that it works; that the current genotyping platforms are efficient, with good imputation methods to fill in missing data; that its global assessment is always a success even if no significant associations are detected; and that genetic effects are likely to be large because humans have not evolved in a drug-therapy environment. By contrast, the con side emphasized that we have limited knowledge of the complexity of the genome; limited clinical phenotypes compromise studies; the likely multifactorial nature of drug response clouding the small genetic effects; and limitations of sample size and replication studies in pharmacogenetic studies. Lively and insightful discussions emphasized further research efforts that might benefit GWAS in pharmacogenetics.

Pharmacogenetics is based on the association of genetic variation with variation in drug effect, either beneficial response or adverse events. Genome-wide association studies (GWAS) have revolutionized approaches to untangling the genetic basis of common disease susceptibility, but will GWAS work for pharmacogenetics research?

To highlight strengths and limitations of GWAS, and expose potential hurdles when using GWAS in pharmacogenetics, a debate was held in conjunction with the Pharmacogenetics Research Network meeting in Rochester (MN, USA). After a day-long meeting of the Fifth Pharmacogenetics Research Network Statistical Analysis Workshop on 15 April 2009, an evening's 'entertainment' in the form of a lively debate was held, appropriately, at Goonies Comedy Club in Rochester. The main focus of the debate was on the question "Will GWAS

[†]Author for correspondence: Tel.: +1 507 284 5581 Fax: +1 507 284 9542, baileyk@mayo.edu.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

advance the search for the genetic basis of response to pharmacotherapy, or are they just another fad?”

Professor Michael Province from Washington University (MO, USA) took the pro position, while professor Daniel Schaid from the Mayo Clinic (MN, USA) took the con position. Professor Nancy Cox from the University of Chicago (IL, USA) officiated by flipping a coin to determine the order of presentations by the debaters. Curiously, both debaters felt a coin flip could have easily determined their pro/con position, because there are numerous strengths and limitations of GWAS for pharmacogenetics. Nonetheless, Province won the coin toss and chose Schaid to present first. The drama unfolded.

Con side of GWAS

In a professorial manner, Schaid laid out five points against GWAS as an approach to understand the genetic basis of response to drug therapy:

- Limited knowledge of the genome;
- Limitations of clinical phenotypes;
- The likely multifactorial nature of drug response;
- Limitations of sample size and the challenges to find adequate replication studies;
- Conceptual limitations of attempting to base analyses on past paradigms while using current technologies.

To emphasize our limited knowledge of the genome, Schaid raised the question of how many genes are likely to be involved in response to pharmacotherapy. He acknowledged the few cases where a single gene, or a few genes, determines a large proportion of the variation in response to a drug. For example, 40% of the variation in response to warfarin treatment has been attributed to two genes, *VKORC1* and *CYP2C9* [1]. However, Schaid posited that this example is likely to be a rare exception. He questioned whether we have found most of the ‘low-hanging fruit.’ While acknowledging the recent successes of GWAS in the context of discovering disease risk alleles (400 alleles reported for 75 diseases in 230 GWA studies in the past 3 years [2]), to date, these alleles have limited clinical utility because their associated odds ratios are small, and they account for a very small fraction of the purported heritability of disease susceptibility.

To emphasize the limited understanding of polygenic phenotypes, Schaid considered the ability to predict height. Although height is estimated to be 80–90% heritable, a large GWAS was only able to find approximately 40 SNPs that account for only 5% of its variance [3]. In a similar manner, the proportion of disease heritability that has been attributable to specific SNPs has been only on the order of 1%.

Schaid argued that our current SNP arrays only give us a ‘peek’ at the underlying genetic complexity, emphasizing that most tagging SNPs are within intergenic regions, so-called gene ‘deserts’. There are few coding SNPs on current arrays and few causal alleles have been found. An open question is whether resequencing will overcome these limitations.

Furthermore, Schaid emphasized that genes do not work in isolation, but rather work in conjunction with other genes, as well as with the environment. Current GWAS are targeted at independent effects of single genes, and may miss interaction effects. He pointed out that regulation of gene transcription is highly complex and interdependent, involving functions of noncoding RNAs, *cis* effects of SNPs and weaker *trans* effects of SNPs, which need a large sample to be detected. In addition, current GWAS cannot address how large-scale genomic structures, such as the highly repetitive ‘junk’ of heterochromatin – the ‘dark matter’, affect cellular processes and drug metabolism.

Turning to the limitations of phenotypes, Schaid pointed out that in classical genetic studies, phenotypes were regarded as the manifestation of single genes. This works well when the phenotype is close to the gene product. However, for current pharmacogenetic studies, phenotypes are typically based on clinically defined diseases – collections of crude measures of disease symptoms. It is unlikely that these ‘beanbags’ of signs and symptoms can be adequately explained by genetic variation. Furthermore, to achieve large sample sizes, some GWAS must pool samples across institutions, hence potentially diluting the phenotype because of varying institutional definitions of symptoms and diseases. Schaid emphasized that a single phenotype can have multiple causes, so it is important to control for nongenetic factors, leading naturally to his next point, multifactorial etiology.

Schaid emphasized that even if a phenotype was well-defined, its etiologic mechanism is likely to be multifactorial. We know that genetic variation, such as SNPs, copy-number variation and structural changes, influence phenotypes. However, so too does noninherited epigenetic variation, such as DNA methylation, histone modification and genomic imprinting. Then there are noncoding RNAs that regulate gene expression. Furthermore, nongenetic factors, such as gender, age, diet and other drugs, play major roles in the response to drug treatments. Beyond these complexities lie interactions among the genetic and nongenetic factors. To emphasize debates about the complexity of the genome, Schaid referred to a debate between Ronald A Fisher and Sewall Wright regarding the forces of selection and evolution, held in the 1920s. Fisher advocated the idea of many small additive genetic effects, while Wright advocated the idea of fewer genes that had large interactions [4]. Schaid acknowledged that the evidence from GWAS to date supports the Fisher model. Schaid also proposed an evolutionary argument favoring multiple rare recessive effects. If heterozygotes produce sufficient enzyme to be functionally normal, then most deleterious genotypes would be recessive. As recessive genotypes are rare, they are harder to purge from the population.

Quoting a study by Alan Wright *et al.* [5], Schaid further suggested that late-onset disease is more likely to be polygenic, with heritability decreasing with age. If true, as supported by current GWAS, then this requires individuals to be at the extreme of the distribution of risk alleles to have an unusually increased risk of disease.

Schaid then turned to the problems of sample size, power and replication. When screening for common risk alleles, to surpass the genome-wide significance level that controls the testing of approximately half a million SNPs, odds ratios of 1.3 are detectable with sample sizes in the order of 1000, while detection of odds ratios of 1.1 require sample sizes on the

order of 10,000. Most associations detected in GWAS will be false positives, so replication studies of at least equal sample size are needed. This is a major challenge for GWAS in pharmacogenetics. The first study is often based on samples from large collaborative clinical trials, which are unique *per se*. How can a replication study be found that is similar to the first study? Furthermore, when evaluating rare adverse drug reactions, it can be quite challenging, if not impossible, to obtain a sufficient sample size for the initial study. How, then, can a replication study be found? To offer hope, Schaid raised the question of whether cell lines, while not strictly replication, might be a viable alternative.

Schaid ended his con discussion by commenting that in the current GWAS era, we are “acting globally but thinking locally”. That is, the GWAS method is a global approach, yet we are still thinking one SNP at a time. In effect, we are straddling past and future, which might not be optimal. We are still hoping for a few ‘big hits’. He ended by emphasizing that the global approach requires better methods for detecting gene–gene interactions, as well as pathways or networks of combined gene effects, along with better-defined traits and study designs, larger samples and better functional assays.

Pro side of GWAS

In a short time, the audience saw that Province would not follow Schaid’s cautious, professorial manner. Province, in his inimitable style, gave glowing tribute to the potential and actual successes of GWAS. He shouted “GWAS works!” Province emphasized the many successes of GWAS to detect alleles for many common diseases, appearing monthly in high-profile journals. His humorous parody of the many popular spinoffs of the television series CSI emphasized the many spinoff GWAS.

Province reviewed reasons for these successes. GWAS is efficient. The cost of SNP genotyping has been going down, thanks to the power of competition between Affymetrix (CA, USA) and Illumina (CA, USA); he alluded in another parody to the famous competition between Coke versus Pepsi. Other reasons for success are high call rates, high quality and reproducibility of genotyping calls, and the ability to provide error estimates so that we even know when things are not working well. Genotype imputation and error estimates will get even better with better reference samples, such as the upcoming 1000 Genomes Project. Other strengths are that GWAS accommodate many study designs and that many software tools are available for analyses. Province turned the multiple comparisons issue from a weakness into a strength, noting that GWAS provides a framework in which multiple testing issues can be formally accommodated. By contrast, the ‘good old days’ of candidate genes had an undetermined number of statistical tests with rolling lists of candidate genes, making it impossible to correct for multiple testing at any given point in time.

Province emphasized the successes of GWAS in discovering new risk alleles and new biology, and asserted that it would be only a matter of time before GWAS discovered all of the common variants associated with disease risk. He recognized that GWAS will miss rare variants, but noted that common variants are the most likely targets for pharmacotherapy and have a larger public health impact. Turning to a mathematical demonstration, he showed that

rare and common variants cannot be in high linkage disequilibrium and, therefore, common variants cannot act as markers of rare variants, and *vice versa*. He emphasized that the next generation of GWAS will extend to rare variants, given the 1000 Genomes Project and the increasingly large GWAS conducted.

Next, Province achieved a coup by arguing that, even if a GWAS ‘fails’, that is, it does not demonstrate statistically significant SNP associations, it is nonetheless successful because it provides valuable information negating associations. Province used this argument to imply that a GWAS “cannot lose”. To emphasize this view, Province referred to the case of Christopher Columbus. Columbus failed several times in his ‘specific aim’ of discovering a new trade route to the riches of China and India. Despite mistaking his preliminary data (estimated distance), his main conclusion (confusing India with America), and failing to accomplish his specific aims, his complete failure to accomplish his preconceived aims was nonetheless a rousing success: he discovered a new world, America. Province’s analogy was brilliant. It showed the value of ‘failed’ research. It also implied that opposition to GWAS is anti-American!

When considering the role of evolution to eliminate common variants predisposing to disease, Province emphasized that this view fails when applied to pharmacogenomics. Humans did not evolve in an environment of drug therapies, and therefore, there is no evolutionary pressure on responses to recently developed pharmacologic agents. Province suggested that the search for common genetic variants affecting response to pharmacotherapy is likely to be successful because the genetic effect size is likely to be larger than that found for disease susceptibility. To emphasize this view, Province cited success stories (e.g., warfarin, mercaptopurine and irinotecan), pointing out that heritabilities are high for many pharmacodynamic phenotypes, and they might not be as polygenic as height.

In summary, Province described GWAS as the “most exciting success story in 21st Century medicine”; it is inexpensive, of high quality, efficient, flexible and effective at either finding common risk or response alleles or strongly falsifying common variant hypotheses. He emphasized that GWAS is so good that it is almost required for all research efforts. His final slide showing a not so futuristic hand-held ‘iGWAS’ hammered the finishing touch on his powerful case for the many pro sides of GWAS.

Discussant

Responding to Schaid’s emphasis that GWAS has not adequately accounted for gene–environment interactions, Cox emphasized that in pharmacogenomics we do know the environment, and in fact measure and use environmental factors. She raised concerns that Province might have been “blinded by rose-colored glasses” while looking at GWAS, and pointed out that effect sizes must be large enough to have clinical utility. Cox emphasized that a major challenge of GWAS in pharmacogenetics is replication: clinical trials based on a GWAS will be extremely difficult to replicate; adverse events are typically rare. Cox ended by emphasizing the need for creative use of intermediate phenotypes, biomarkers and nonpharmacophenotypes as warranted.

Bibliography

1. Limdi NA, Veenstra DL. Warfarin pharmacogenetics. *Pharmacotherapy*. 2008; 28(9):1084–1097. [PubMed: 18752379]
2. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*. 2008; 118(5):1590–1605. [PubMed: 18451988]
3. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008; 456(7218):18–21. [PubMed: 18987709]
4. Provine, WB. *The Origins of Theoretical Population Genetics*. University of Chicago Press; IL, USA: 1971.
5. Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H. A polygenic basis for late-onset disease. *Trends Genet*. 2003; 19(2):97–106. [PubMed: 12547519]