GWAS or Gee Whiz, PSAS or Pshaw: Elucidating the biologic and clinical significance of genetic variation in cardiovascular disease

Andrew P. Landstrom, BS and **Michael J. Ackerman, MD, PhD**

Departments of Medicine, Pediatrics, and Molecular Pharmacology & Experimental Therapeutics and the Divisions of Cardiovascular Diseases and Pediatric Cardiology, Mayo Clinic, Rochester, Minnesota

The initial decoding of a handful of human genomes in February 2001 ushered in a new age of genetic discovery that has revolutionized our understanding of health and disease.[1,2] Over the last two decades, the genetic underpinnings of heritable cardiovascular disease have been unveiled with the discovery of rare pathogenic mutations that cause cardiac channelopathies and cardiomyopathies. In addition, the influence of common genetic variation (polymorphisms) on both physiology and pathology is slowly being illuminated. As technology has advanced, so has our ability to further interrogate the genome. With the completion of the final draft of the Human Genome Project in 2004,[3] the International HapMap Project in 2005,[4] and the ENCODE Project in 2007,[5] a new wave of genetic association studies have started to detail the role of genetic variation in human disease. Despite significant advancements, the complexity of the genome and the diseases being studied have also brought new challenges.

Methodology for uncovering the pathogenetic basis of human disease

For uncommon heritable diseases, the vast majority of disease susceptibility genes have been discovered using either genome-wide linkage analysis or candidate gene analysis. A highly penetrant, multigenerational pedigree is required for the former, whereas a cohort of phenotypically rich, unrelated probands is preferred for the latter. No *a priori* hypothesis is needed for genome-wide linkage analysis except to hypothesize that the family's disease locus is, in fact, concealed somewhere in their human genome and that the panel of genome-wide markers will expose its location (locus). Conversely, for candidate gene analysis, the "biologic plausibility" of the gene's translated product is the cornerstone for this hypothesis-driven strategy. For example, when specific mouse models of cardiac hypertrophy, dilation, and failure undergo cardiac remodeling, reduced expression of *JPH2*-encoded junctophilin type 2 is observed.[6,7] Subsequently, genetic mutations in *JPH2* were implicated in human hypertrophic cardiomyopathy based on the candidate gene approach.[8] Although these approaches have netted nearly 50 cardiomyopathy/channelopathy susceptibility genes, they

Address reprint requests and correspondence: Dr. Michael J. Ackerman, Windland Smith Rice Sudden Death Genomics Laboratory, Guggenheim 501, Mayo Clinic, Rochester, Minnesota 55905. ackerman.michael@mayo.edu.

have fallen short in identifying the cause(s) of complex multifactorial diseases such as hypertension and coronary artery disease.

With the recent development of high-throughput genotyping arrays, genome-wide association studies (GWAS) have broadened genetic analysis to the entire genome. These high-density, high-resolution arrays allow for the genotyping of usually approximately 500,000 single nucleotide polymorphisms (SNPs) for an individual in an unbiased, hypothesis-free manner. When applied to large cohorts of patients with a given disease/trait and compared to appropriately matched healthy volunteers, there is significant power to detect which SNPs, and the genomic loci for which they serve as markers of linkage dysequilibrium, may play a role in disease pathogenesis. The last 3 years have seen an explosion of these studies with hundreds, perhaps thousands, of SNPs associated with nearly 100 physiologic traits or disease processes.[9] GWAS have identified novel genetic loci associated with the development of coronary artery disease,[10] hypertension,[11] stroke,[12] and quantitative physiologic traits such as the QT interval.[13,14] For some investigators, the need for large cohorts of patients and additional validation/replication studies using an independent disease cohort can make GWAS cost-prohibitive.

Consequently, a hybrid approach between a hypothesis-driven candidate gene study and a hypothesis-free GWAS is emerging, referred to here as pathway-specific association studies (PSAS). More commonly used to explore the genetic modifiers of drug efficacy and toxicity, [15] PSAS have rarely been used to explore the pathogenesis of cardiac diseases. In this issue of *Heart Rhythm*, Tseng et al[16] take this relatively novel approach to examine genetic susceptibility to sudden cardiac arrest (SCA) and use a PSAS to explore the potential sudden death association of SNPs in genes encoding components of the transforming growth factor-β (TGF-β) signaling pathway.

Reproducibility dictates relevance

Tseng et al interrogated 617 SNPs spanning 12 genes that encode components of the TGF-β signaling cascade, including TGF-β2 and TGF-β3, which are two ligands for the TGF-β receptors TGFBR1 and TGFBR2, and various SMAD proteins that serve as either stimulatory or inhibitory downstream effectors of receptor signaling. The investigators identified a SNP (rs9838682) that reportedly localized to the TGF-β receptor 2 locus (*TGFBR2*) and imparted an age- and sex-adjusted odds ratio of 1.66 for SCA. While appearing modest at first blush, this odds ratio is relatively high compared to most GWAS published to date. In a recent GWAS meta-analysis, Hindorff et al[9] identified a median odds ratio of 1.33 (interquartile range 1.20–1.61) among 531 SNP-trait/disease associations in 151 GWAS.

As the authors recognized, the location of the SNP outside of a coding exon makes it difficult to establish/predict its functional sequelae on *TGFBR2* signaling or any other biologic target within that locus. It is more likely that this apparent pro-SCA SNP lies in linkage dysequilibrium with other SNPs in the same genetic region, hinting at a DNA locus, rather than a single SNP, which may be associated with SCA. In fact, the authors report that

rs9838682 lies in a region of high linkage dysequilibrium encompassing 44 kb of DNA, which is actually what was associated with increased susceptibility for SCA in their cohort.

Ultimately, the true test of the clinical relevance of this, or any other SNP, will be time and the reproducibility it will need to bring. As the authors point out, an important prerequisite before determining whether this scientific observation may yield a clinically actionable biomarker with therapeutic/prognostic relevance is replication of the rs9838682-SCA association in an independent cohort. Should this finding be reproduced in other genetic association studies of SCA and the increased risk associated with minor allele status confirmed, then prospective analyses into the prognostic implications could be pursued. Among the plethora of GWAS published to date, replication of an identified association is seldom validated in an independent cohort within the same study. Indeed, those that have demonstrated reproducibility, such as the 9p21 loci association with coronary artery disease, [17–19] may hold future clinical utility. Unfortunately, when a replication study is conducted, the results frequently do not identify the same loci.[20]

Quite independent of the future clinical potential of the SNP as an informative biomarker is the possibility of uncovering novel biology. Although TGF-β signaling has been established as a critical modulator of cardiac fibrosis,[21] its role in fatal arrhythmogenesis has yet to be elucidated. However, because rs9838682 is a noncoding polymorphism, direct functional evaluation of the minor allele on the signal transduction capability of the receptor cannot be ascertained, and dissecting out its biologic impact will be daunting. This is reminiscent of one of the first, and most intensely studied, GWAS signals. The 9p21 locus associates with coronary artery disease and myocardial infarction, yet it lies outside of any known gene or microRNA sequence.[17–19] By one estimate, greater than 40% of SNP trait/disease associations identify SNPs that are intergenic and do not immediately offer a clear genetic/mechanistic explanation behind the observed association.[9]

Careful navigation of the genetic sea

As the "omics" age (genomics, transcriptomics, proteomics, metabolomics, and beyond) advances, we must carefully navigate an ever treacherous sea of information. Publically accessible databases, which house growing caches of information, must be used thoughtfully and confirmed independently. Although Tseng et al have truly demonstrated an intriguing association between the rs9838682 SNP and SCA, they may have inadvertently mislocalized the SNP. According to NCBI, Ensembl,[22] and the UCSC Genome Browser,[23] rs9838682 is located more than 230 kb upstream from the 5′ untranslated region, the first exon, of *TGFBR2*. Although these informatics or info-omics databases place *TGFBR2* well outside the locus in linkage dysequilibrium with rs9838682, a yet-to-be explained sudden death predisposing signal associates with this particular SNP. Furthermore, as this SNP does not localize to any known gene or regulatory element (indeed, the closest genetic feature is *TGFBR2*), the authors may have serendipitously identified a novel SCA-associated regulatory element. As with all association studies, further investigation is needed to confirm or disprove this possibility.

As we navigate the "-omics" sea to further explore the fundamental causes of genetic disease and uncover the influence of genetic variation in multifactorial disease, the ultimate key—for the benefit of our patients—will be the wise interpretation of the mistakes/variants/glitches divined from our "-omics" code. If clinically actionable biomarkers in terms of either therapeutics/risk stratification or novel biology fail to surface from the current litany of literally hundreds, perhaps thousands, of odds ratio <2 SNPs, then GWAS potentially could mutate, with a combination of insertions and deletions, to "Gee Whiz" and PSAS to "Pshaw."

References

1. International Human Genome Sequencing Consortium (IHGSC). Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

2. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]

3. International Human Genome Sequencing Consortium (IHGSC). Finishing the euchromatic sequence of the human genome. Nature. 2004; 431:931–945. [PubMed: 15496913]

4. International HapMap Consortium (IHMC). A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]

5. ENCODE Project Consortium (EPC). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]

6. Minamisawa S, Oshikawa J, Takeshima H, et al. Junctophilin type 2 is associated with caveolin-3 and is down-regulated in the hypertrophic and dilated cardiomyopathies. Biochem Biophys Res Commun. 2004; 325:852–856. [PubMed: 15541368]

7. Xu M, Zhou P, Xu S-M, et al. Intermolecular failure of L-type Ca^{2+} channel and ryanodine receptor signaling in hypertrophy. PLoS Biol. 2007; 5:e21. [PubMed: 17214508]

8. Landstrom AP, Weisleder N, Batalden KB, et al. Mutations in JPH2-encoded junctophilin-2 associated with hypertrophic cardiomyopathy in humans. J Mol Cell Cardiol. 2007; 42:1026–1035. [PubMed: 17509612]

9. 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; 106:9362–9367. [PubMed: 19474294]

10. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007; 357:443–553. [PubMed: 17634449]

11. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009; 41:677–687. [PubMed: 19430479]

12. Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. N Engl J Med. 2009; 360:1718–1728. [PubMed: 19369658]

13. Pfeufer A, Sanna S, Arking DE, et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nat Genet. 2009; 41:407–414. [PubMed: 19305409]

14. Newton-Cheh C, Eijgelsheim M, Rice KM, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat Genet. 2009; 41:399–406. [PubMed: 19305408]

15. Wang L, Weinshilboum RM. Pharmacogenomics: candidate gene identification, functional validation and mechanisms. Hum Mol Genet. 2008; 17:R174–R179. [PubMed: 18852207]

16. Tseng Z, Vittinghoff E, Musone S, et al. Association of TGFBR2 polymorphism with risk of sudden cardiac arrest in patients with coronary artery disease. Heart Rhythm. 2009; 66:1745–1750.

17. Wellcome Trust Case Control Consortium (WTCCC). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

18. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007; 316:1488–1491. [PubMed: 17478681]

19. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–1493. [PubMed: 17478679]

20. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008; 9:356–369. [PubMed: 18398418]

21. Leask A. TGFbeta, cardiac fibroblasts, and the fibrotic response. Cardiovasc Res. 2007; 74:207–212. [PubMed: 16919613]

22. Hubbard T, Andrews D, Caccamo M, et al. Ensembl 2005. Nucleic Acids Res. 2005; 33:D447–D453. [PubMed: 15608235]

23. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]