

## Genome Wide Association Studies – Do Not Judge a Book by Its Cover

Robert Roberts

Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

The Era of the Genome-Wide Association Studies (GWAS) has arrived and in a short span of 5 years its prolific record has been nothing short of outstanding. GWAS will likely be a landmark of the 21st century, analogous to the Human Genome Project of the 20th century. The GWAS design provides an unbiased approach to mapping the chromosomal location of common genetic risk variants.<sup>1,2</sup> A common risk variant is arbitrarily defined as one having a frequency in the general population of equal to or greater than 5%.<sup>3</sup> However, the lower limit of sensitivity for most GWAS published to date has been only at a 10% frequency.<sup>4</sup> Rare ( $\leq 5\%$  frequency) variants are not expected to be detected by GWAS, unless the sample size is exceptionally large. A systematic approach to detect rare variants will probably require direct DNA sequencing.

One major feature of the GWAS has been the large international consortiums formed around several diseases. The largest consortium, Coronary Artery Disease Genome-wide Replication and Meta Analysis (CARDIOGRAM),<sup>5</sup> dedicated to coronary artery disease (CAD), has already phenotyped and genotyped a discovery population of 82 000 and a replication population of more than 40 000. These consortiums bring together not just the large sample sizes necessary to map the risk variants but also integrate and enrich the expertise and resources necessary to interpret and analyze the data. As a result of these large consortiums, it is expected that most of the genetic variants, predisposing to common diseases such as CAD

and cancer will be mapped within the near future. Despite, the success of GWAS to detect over 400 loci associated with disease<sup>6</sup>, there is skepticism as to whether GWAS will advance the management of disease.<sup>7,8</sup> The expected mechanisms whereby knowing genetic risk variants would improve management of these diseases would include genetic screening for diagnosis and prevention, pharmacogenetics to provide the right drug for the right person at the right dose, and lastly targets for development of new and more appropriate drugs. We should also be reminded that hidden treasures may evolve from improved understanding of the molecular basis for disease, other than those expected. One example of the latter could be vaccine development, often facilitated by knowing the DNA sequence.

The concern over limitations of GWAS will continue to be a legitimate issue until we have indeed shown their application is beneficial. This concern is heightened by the observation that despite many common risk variants having been mapped for diseases such as CAD, their cumulative relative risk does not account for most of the expected genetic predisposition.<sup>7,9,10</sup> Based on current data, the absolute risk due to these variants probably accounts for only 20% to 30% of predicted risk. What is responsible for the remaining risks? The answer is certainly not obvious at this time but we should be reminded it is less than 5 years since the technology to map common loci became available and thus one would expect many more common variants are yet to be discovered. Another possibility is that rare variants<sup>11</sup> which we would expect to have much greater effect have yet to be identified because GWAS does not have the sensitivity to detect rare variants. Do risk variants act in a linear fashion, or are they interactive with each other, giving a greater effect than simply adding up the risk of each individual variant? In this issue of *Revista Española de Cardiología* a study by Lluís-Ganella et al<sup>12</sup> has attempted to address this issue.

In this study the investigators performed an in silico analysis of a sample size of 7368 from the

SEE ARTICLE ON PAGES 925-33

Correspondence: R. Roberts, MD, FRCPC, MACC, President & Chief Executive Officer, Director, Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, 40, Ruskin Street, Ottawa, ON K1Y 4W7 Canada  
E-mail: roberts@ottawaheart.ca

Wellcome Trust Case Control Consortium database, consisting of 1988 CAD cases and 5380 controls. From the literature, they selected 9 proven variants associated with increased risk for CAD. Since each individual has 2 copies of each variant, 1 could be heterozygous, homozygous or have neither of the risk alleles. Thus, the total number of risk alleles for any one individual would be 18. The 9 selected variants manifest their risk for CAD independent of classical risk factors such as cholesterol, diabetes, or hypertension. The investigators provided interesting and unique analysis showing that the number of risk variants per individual varied from 1 to 13 with the median number per individual being 7. The results of this analysis clearly indicate that the greater the number of risk alleles the greater the risk. The risk was additive in a linear relationship. The linear relationship excludes any gene to gene interaction, indicating the accumulative risk is simply the total of the risk exhibited by each individual variant. If these results are confirmed in other populations and for other genes, it would exclude gene to gene interaction, which would be quite surprising. Gene to gene interaction is expected, which in turn would provide greater risk along with a greater environmental response. However, it is premature to exclude gene to gene interaction on the basis of this study alone. This study did not attempt to assess whether application of the genetic score would add to risk stratification over that of conventional factors. The investigators strongly urge that a prospective population cohort is most appropriate to attempt to respond to this question.

While the concerns for application and benefits from genetic risk variants are legitimate and appropriate, it should accelerate rather than deter the search for more common and rare variants. The evidence for genetic predisposition to CAD<sup>13</sup> (as for other diseases) comes primarily from epidemiological studies of familial and population cohorts. We now have independent and additive proof of this genetic predisposition. Through the application of GWAS, hundreds of chromosomal loci have been shown to be associated with diseases. Secondly, as predicted, the genetic predisposition to these diseases is polygenic, with each allele contributing only minimal to moderate risk. We must be reminded not to judge a book solely by its cover. The loci that predispose to CAD and other diseases have been located on the basis of an association between single nucleotide polymorphisms (SNPs) markers within the loci and the disease phenotype. Most of the SNPs are simply markers in linkage disequilibrium with the actual causative SNP. The loci currently mapped by the case control association (GWAS) vary in length

from 50 000 bps to 100 000 bps. Yet we know that for most loci there is a single SNP that is causative and responsible for the genetic predisposition. Secondly, most of the loci are not in protein coding regions but in introns and promoter regions.<sup>8,9</sup> Furthermore, not infrequently the regions include non-coding RNAs. This is exemplified by the 9p21 risk variant<sup>14,15</sup>, a long non-coding RNA referred to as ANRIL.<sup>16</sup> To evaluate and apply this information scientifically or clinically, we should identify the causative sequence and determine its function. It is reasonable to assume that while genetic screening for prevention will be helpful, the greater benefit could possibly be using the variant sequences as targets for novel therapy.

## REFERENCES

1. Wang WY, Barratt B, Clayton DG, Todd JA. Genome-Wide Association Studies: Theoretical and Practical Concerns. *Nat Rev Genet.* 2005;6:109-18.
2. Hirshhorn JN, Daly MJ. Genome-wide Association Studies for Common Diseases and Complex Traits. *Nat Rev Genet.* 2005;6:95-108.
3. Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics.* 2003;19:376-82.
4. Dandona S, Stewart A.F.R., Roberts R. Genomics in coronary artery disease: Past, present and future. *Can J Cardiol.* 2010;26 Suppl A:56A-9A.
5. CARDIoGRAM consortium. Design of the Coronary Artery Disease Genome-wide Replication and Meta Analysis (CARDIoGRAM) Consortium - a prospective meta-analysis of 13 genome-wide association studies. *Circ Cardiovasc Genet.* 2009. In press.
6. Donnelly P. Progress and challenges in genome-wide association studies in humans. *Nature.* 2008;458:728-31.
7. Goldstein DB. Common genetic variation and human traits. *New Engl J Med.* 2009;360:1699-701.
8. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-53.
9. McClelland JH, King M-C. Genetic Heterogeneity in Human Disease. *Cell.* 2010;141:210-7.
10. Paynter NP, Chasman DI, Pare G, Buring J, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA.* 2010;303:631-7.
11. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol. *Science.* 2004;305:869-72.
12. Lluís-Ganella C, Lucas G, Subirana I, Sentí M, Jiménez-Conde J, Marrugat J, et al. Efecto aditivo de diferentes variantes genéticas en el riesgo de cardiopatía isquémica. *Rev Esp Cardiol.* 2010;63:925-33.
13. Roberts R. Personalized Medicine: A Reality within this Decade. *J Cardiovasc Transl Res.* 2008;1:11-6.
14. McKenna D, Summerour SR, Villareal FJ. Role of mechanical factors in modulating cardiac fibroblast function and

- extracellular matrix synthesis. *Cardiovasc Res.* 2000;46:257-63.
15. Slaughter LM. Your Genes and Privacy. *Science.* 2007;316:797.
16. Jarinova O, Stewart AFR, Roberts R, Wells G, Lau P, Naing T, et al. Functional Analysis of the Chromosome 9p21.3 Coronary Artery Disease Risk Locus. *Arterioscler Thromb Vasc Biol.* 2009;29:1671-7.