

2006 Cardiovascular Biomarkers And Surrogate Endpoint Symposium

September 18-20, 2006

**Bethesda North Marriott Hotel & Conference Center
Bethesda, Maryland**

NIH Cardiovascular Biomarker Initiatives

Christopher O'Donnell, M.D.

Senior Advisor to the Director for Genome Research
and Associate Director of the Framingham Heart Study
NHLBI, National Institutes of Health
Bethesda, MD

Within the National Institutes of Health (NIH), and particularly the National Heart, Lung, and Blood Institute (NHLBI) many genetics and genomics projects are underway that are expected to have profound implications in the biomarker field. The application of genomics knowledge and technology is an entirely new branch of information that is going to flow very rapidly into the current river of information utilized to understand disease, biomarkers of disease, and genetic modification of the disease process.

There is an underlying assumption that common genetic variation has something to do with the common manifestations of cardiovascular disease (CVD). The type of polymorphism that is of most interest in this regard is the single nucleotide polymorphism (SNP), where a single base pair alteration along the genetic sequence potentially leads to some functional or structural change downstream. While examples of rare Mendelian conditions, such as familial hypercholesterolemia leading to myocardial infarction, are well-characterized, these conditions make up a very small fraction of the overall CVD in the population. Most disease involves more complex modes of transmission, with a genetic component which is not Mendelian.

There certainly is very good evidence that there is a heritable genetic component to CVD. The Framingham study which demonstrated the association between premature CVD in parents, offspring, and siblings, is part of the large body of evidence supporting the idea that CVD, and its manifestations such as myocardial infarction, congestive heart failure, and arrhythmias all have a heritable component. Therefore genetics likely underlies part of these disease processes. Framingham and several other studies have demonstrated that intermediate phenotypes, the phenotypes that lie between the genetic and environmental interaction and the clinical outcome, have a heritable component and that genetics appears to underlie the variability in these traits. For example, the heritability calculated by sibling correlations for a variety of traditional risk factors demonstrates that between 30-40% of the individual variability is explained by genetic factors. So similarly for biomarkers, there is a significant proportion of variability driven by heritable components. For sub-clinical disease traits, which are often referred to as surrogate measures of disease, there is also a fairly significant heritable component. Therefore it is reasonable to look for both the genetic and environmental underpinnings of these traits.

There have been two approaches to understanding how genetics underlies the disease traits in CVD: the Genome-Wide Association (GWA) approaches and Candidate Gene Association approaches. The general notion has been to discover strong genotype-phenotype associations that are replicated independently in many cohorts, then to do fine mapping, other genetic studies, and ultimately functional studies of these putative variants in order to work toward personalized medicine approaches for disease prediction and prevention. Much work needs to be done before findings can be applied toward personalized medicine, but much progress has recently been made in this direction. A review of individual studies seems to indicate that none of the studied variants are conferring a very strong relative risk for disease. However, when data is analyzed in a meta-analysis, statistically significant associations seem to emerge for these some variants. It is important to recognize that only single variants in these genes and not the entire totality of genetic variation across the gene region have been studied. The results have been less than satisfying because it appears that the magnitude of risk conferred by any one variant is not very great, and that it takes a lot of participants to find any true associations.

Framingham was able to study many of these same single variants in relationship not only to CVD outcomes, but to biomarkers that are so called intermediate phenotypes located downstream of the variants. A strong association between some single variants and blood levels of biomarkers (Apo E and lipoprotein sub-fractions, PAI-1 gene and PAI-1 blood levels, etc.) have been consistently found. However with other markers such as sub-clinical carotid endothelial thickness, the association with CVD outcome, which is many steps downstream, has often not been as consistent.

Much larger sample sizes may be needed to detect some of these associations. As a result of the complete sequencing of the human genome, first announced in the year 2000, and then published less than two years ago, the ability to find important associations has improved. The human genome project sequenced 2.85 billion nucleotides and a total of 20-25,000 protein-coding genes were estimated to exist in the human genome. One of the remarkable elements of this project, aside from the fact that it occurred, is that all of the data is available on the internet for anyone to use. This model of free access to data is being used with the NIH clinical genomic programs.

Additionally, a complete haplotype map (HapMap) of the human genome has been published. It is known from the human genome sequencing project that there is one single nucleotide polymorphism (SNP) occurring per approximately 1,000 bases. Therefore, there are approximately 10 million SNPs across the human genome. The hypothesis is while 99.9% of the human genome sequence is identical, these variations, about 1/1,000 bases, underlie much of the differences between us as humans. Also learned from the human genome sequencing and haplotype map projects is the fact that SNPs that reside close to each other along the human genome sequence actually run together in correlated groups. So if there are five of these SNPs running along the human genome, there are potentially 2^5 possible combinations or haplotypes. But, what has been observed is that far fewer common haplotypes exist because there is conservation of the structure of these SNPs across the small segments of the DNA due to the lack of

recombination during the reproductive process of meiosis. Therefore, one can take advantage of the limited number of haplotypes to construct different types of genetic research projects that can make use of far fewer SNPs than one might have to otherwise genotype.

For example, by genotyping SNPs across the entire plasminogen activator 1 (PAI-1) gene region it was found that there are correlated blocks of linkage disequilibrium or correlations across this gene region, and a limited number of haplotypes are formed by these SNPs. Further evaluations looking at the associations of these haplotypes and the individual SNPs that make up these haplotypes with PAI-1 levels in the blood found some very strong associations between not only the known 4G-5G polymorphism, but also several other SNPs in this region, and these SNPs also made up haplotypes that are strongly associated. It was determined that approximately 5% of the variability in PAI-1 levels are driven by variation at this gene region. This may not seem like a very large proportion, but in fact there are not many biomarkers where single polymorphisms in one gene region define more than 1% of the variability. This finding suggests that 5% of the variability is driven by the gene variability in the PAI-1 locus, but perhaps the rest of the variability influenced by genetics can be explained by variation in other regions of the genome yet to be discovered.

As a result of the haplotype map, the human genome sequence, and the availability of high throughput genotyping at a reasonable cost, it is possible to conduct Genome-Wide Association Studies. Several companies are now selling screening chips and the prices of these chips have been decreasing precipitously. The possibility of research being conducted with these chips has never been more imminent. In some sense technology has outpaced the conduct of research because research studies are now being initiated in areas where there is little prior experience.

Similarly, not only at the locus, but on the entire chromosome, one can define by a limited number of SNPs, so called blocks of linkage disequilibrium. It is estimated for example, for European-Americans that there are about 100,000 blocks of linkage disequilibrium. Therefore approximately 276,000 SNPs would be needed to define the entirety of variability across all of the 22 autosomes. So the notion upon which these chips are being developed is that one can define a set of SNPs that defines all variability across the genome, and use those for discovery purposes.

In the past year, several proof-of-principle studies have been published. The first of these studies was not in the area of cardiovascular disease, but in fact in macular degeneration, a degenerative condition in the eye, where a variant in the complement Factor H gene was associated strongly with macular degeneration. One of the SNPs emerged as strongly significant and was replicated in other studies, suggesting this is a real finding that derived from this screen of a 100,000 SNPs.

NHLBI recently convened a working group of experts in the field of genetics and genomics. This working group concluded that large-scale GWA studies should proceed, but that they should be conducted in existing cohorts with well-designed phenotypes and

those with public health importance, i.e. biomarkers. The group also concluded that immediate access to data should be made possible and a centralized data repository should hold these data.

In the area of cardiovascular disease, another proof-of-concept GWA study in the Framingham population involved genotyping only 200 subjects at the extremes of QT interval length. Interestingly, even with this limited population, a single variant in the CAPON gene nitric oxide synthase (one regulator gene implicated in QT interval length) was uncovered. This particular gene was never before known or even suspected to be involved in QT interval or repolarization. This finding demonstrates how a result, when strongly replicated, may lead to a new area of research and potentially a new pathway to evaluate. Similarly, another study of the Framingham population evaluating 100,000 SNPs had shown that a variant in the INSIG2 gene is associated with body mass index, a finding which was replicated in 4 of 5 other studies.

The SNP Health Association Resource (SHARe) program is another GWA study involving approximately 10,000 Framingham subjects from all three Framingham generations and genotyping of 550,000 SNPs. Included in this database are all available phenotypes that have been measured in Framingham and a comprehensive web-based database of the genotypes and phenotypes will be assembled. The planned initiation is June, 2007. The data from the SHARe project will be rapidly shared with the scientific community after an application has been submitted. There will also be some intellectual property (IP) guidelines put forward which will strongly discourage users of these data from placing IP claims on the data in order to prevent a block of the wide spread use of these data. Many pharmaceutical companies are in agreement with this type of policy.

Another program is the ENDGAME program which will focus on optimal designs and methods for analysis of GWA studies. STAMPEED (SNP Typing for Association with Multiple Phenotypes from Existing Epidemiologic Data) is another project of GWA studies, most of which focus on disease based studies, such as myocardial infarction. The CARE project is a project for genotyping and creating a genotype-phenotype database in 50,000 persons from 8 different NHLBI Cohorts, in which SNPs in 1,700 or more candidate genes will be genotype or crossed to all 50,000 subjects. Additionally, a small GWA study will be performed to determine whether or not associations are real across a diversity of populations. The Gene and Environment Initiative is a project designed to accelerate the understanding of genetic and environmental contributors in disease. One of the components is a genetic analysis of case control studies using GWAs technologies. The other is development of technologies for measuring environmental exposures.

The hope is that these GWA studies will generate a number of new and interesting associations that after independent replication will become targets and potentially enter into the river of information that can be used for biomarker research and discovery. These programs just outlined are expected to contribute actively to the sort of movement of information from GWAs studies down to ultimately the personalized medicine approach.

There is now a portfolio of genomic studies that has been developed in NHLBI and NIH with a strong emphasis on widespread accessibility of data to researchers which is revolutionary in its scale and extent. A growing number of whole genome association and related studies, many funded by NHLBI, will yield new information regarding cardiovascular disease genes and pathways. The challenges are many in terms of the interpretation of these data. The massive data and the volume of data are orders of magnitude greater than what anyone has ever dealt with before in the genetics field. Therefore, the analytic challenges are formidable. But hopefully with funding and the use of new informatics tools, useful knowledge will be gained. Many opportunities exist for researchers in the field to analyze and incorporate these data. Perhaps the potential short and intermediate term outcomes of these projects will be an impact on the biomarker pipeline and the development of personalized medicine. This information may be useful to dissect genetically the biomarker and disease traits, to understand the genetic underpinnings of biomarkers in surrogate measures, and potentially to discover new targets for drugs. Ultimately the hope is to incorporate this information into tests and strategies, using both genetic and known biomarker information to predict, prevent, and pre-empt disease.