Putting genome analysis to good use: Lessons from C-reactive protein and cardiovascular disease

ABSTRACT

New methods of studying the human genome offer novel ways to examine the relationship between biomarkers and common, chronic human diseases. As an example, we will review a large genomics study (Elliott et al, JAMA 2009; 302:37–48) that concluded that C-reactive protein (CRP) is likely not a cause of coronary heart disease, although it is a marker for it.

KEY POINTS

Genome-wide association studies can uncover associations between genetic markers and medical conditions, but they fall short of establishing causality or even clear biologic interactions between a genetic variant and a disease state.

Mendelian randomization is a method for addressing the relationship between genetic variants and disease, ie, whether a biomarker affected by the variant is a cause of the disease or merely a bystander.

CRP, an acute-phase reactant produced by the liver in response to inflammation, is one of many inflammatory markers whose levels correlate with coronary disease and which has been suggested to play a role in its pathogenesis.

The findings of Elliott et al suggest that therapies that specifically lower CRP levels are not likely to affect coronary artery disease.

doi:10.3949/ccjm.79a.09169

GENOMICS RESEARCH is paying off, not only by identifying people at risk of rare inherited diseases but also by clarifying the pathogenic mechanisms of important, common ones.

Thanks to advances in technology, we can now, at a reasonable cost, simultaneously screen for millions of genetic variants in thousands of people to find variants that are more common in people with a given disease than without the disease, a fruitful method called a genome-wide association study. Moreover, an epidemiologic method called mendelian randomization takes advantage of the natural re-shuffling of the genetic deck that occurs with each generation to give an estimate of whether certain gene products are mediators—or merely markers—of disease.

In a landmark study published in 2009, Elliott et al1 used mendelian randomization to evaluate the role of C-reactive protein (CRP) in coronary artery disease.

Here, we review the use of genetic tools in a clinical context, highlighting CRP to illustrate some of the potential uses and limitations of applied genomics in clinical investigation.

NATURE VS NURTURE: AN AGE-OLD DEBATE

The relative contributions of genetic and environmental factors to human health and disease—nature vs nurture—is an age-old debate in which interest has been renewed in this era of intensive research in molecular genetics.

In the 19th century, Charles Darwin proposed that evolution proceeds through natural selec-
tion of variations in inherited traits. His con-
temporary, Gregor Mendel, showed that traits
are inherited in discrete units, later named genes.
Just what genes were and how they worked had
to await the discovery of the structure of DNA in
1953, by Watson and Crick.2

Since then, progress has accelerated. Ad-
vances in recombinant DNA and DNA-se-
quencing technologies enabled sequencing of
the entire human genome only 50 years later.
More recently, we have seen automated rapid
sequencing, the HapMap project (more on
this below), and the advent of genome-wide
association studies that uncover genetic vari-
ants correlated with or predisposing to com-
mon, complex human diseases.

Until recent years, medical genetics was
mostly confined to the study of rare syn-
dromes, such as Huntington disease, that
are due either to a change in a single gene
or to abnormal quantities of large swaths of
chromosomes containing many genes. It had
little application to most of the common dis-
orders seen by primary care physicians. How-
ever, the genes and pathways implicated in
rare monogenic disorders have provided key
insights into common diseases. For example,
defining the genes and mutations underlying
familial hypercholesterolemia highlighted
the role of low-density lipoprotein choles-
terol (LDL-C) in the pathogenesis of athero-
sclerotic disease.

■ 3.4 BILLION BASE PAIRS, 23,000 GENES

The DNA molecule consists of two strings of
the nucleotides guanine (G), cytosine (C),
thymine (T), and adenine (A). The human
geneome contains about 3.4 billion of these
nucleotides, also called base pairs, as they bind
G to C and A to T across the length of the
double helix of the DNA molecule.

Only about 2% of these 3.4 billion base
pairs make up genes, ie, sequences that are
transcribed into RNA and then translated
into protein. Humans have only about 23,000
genomes, which is less than in some plant species.

What about the rest of the human genome,
ie, most of it? Previously dismissed as “junk,”
these regions likely possess more elusive regu-
laratory functions, controlling gene expression
(ultimately, the production of protein), which
varies considerably from tissue to tissue and
over a person’s lifetime.

It is the orchestration of gene expression
over time and cell type that gives the human
body its intricate complexity. The study of
how all our genes and gene products inter-
act is called genomics and is part of the larger
topic of the network of protein interactions
(proteomics) and of the integration of various
protein pathways (metabolomics).

We are all 99% identical—or 12 million nucleotides different
Human genome sequences are 99% identical
across populations. But the remaining 1% is
still a big number: there are more than 12 mil-
lion variants between any two individuals’ ge-
nomes. These variants include:

• Single-nucleotide polymorphisms (SNPs),
  ie, a single-nucleotide change that is pres-
  ent in at least 1% of the population
• Copy number variants (CNVs), ie, a
  stretch of DNA that is either missing or
duplicated
• Repeating patterns of DNA that vary in
  the number of repeated sequences.

■ THE EVOLUTION
OF GENOMICS RESEARCH

Much of the initial focus of research in the
genomics era consisted of identifying these
variants and discovering associations between
them and particular human diseases or clinical
outcomes. In this way, we uncovered a multi-
tude of potential new biomarkers and therapeu-
tic targets, requiring further investigation
into the connection between the DNA vari-
ant and the clinical state.

At the close of the 20th century, genetic
factors were correlated with human disease by
linkage analysis (a method of mapping pat-
tterns of markers that congregate in relatively
narrow regions of DNA in families with spe-
cific diseases), and candidate gene approaches,
whereby genes were investigated on the basis
of their postulated biology and of previous
studies. These techniques were relatively low-
yield and cumbersome; years of work uncov-
ered only a handful of genes proven to be as-
sociated with diseases.

Newer tools can look at scores of genes

Studies of familial hyper-
cholesterolemia

shed light on
the role
of LDL-C in
atherosclerosis
linked to common diseases. Researchers now rely on sophisticated DNA sequencing tools and interpretation software to sift masses of data to find meaningful markers (DNA variants or mutations).

Genomics research in the past few years has been mostly hypothesis-independent. Investigators are no longer limited to the small cache of genes whose corresponding proteins are well characterized, but can instead probe the entire genome for connections between our DNA and our physiology.

The rise of genome-wide association studies
Over the past decade, much clinically useful information has been gathered in genome-wide association studies.

The rise of this type of study rested on our emerging understanding of the architecture of our genome. When the genomes of multiple humans were fully sequenced, we discovered that specific variants do not occur randomly in relation to each other. Rather, they tend to be inherited in particular blocks called haplotypes, and some SNPs or combinations of SNPs are very rare or essentially never seen.

In its first phase, the HapMap project organized these useful blocks of variants, genotyping 1 million SNPs for each of 270 individuals from mother-father-offspring trios from distinct geographic regions of the world. The second phase of the HapMap project extended the analysis to more than 3 million SNPs and to other populations.

While the HapMap should be generally applicable to other populations not yet studied, limitations of the first two HapMap phases include rare SNPs or CNVs, or variants outside of haplotype regions.

The 1,000 Genomes Project, now underway, will develop an even more comprehensive catalog of human genetic variants in much broader populations.

The success of genome-wide association studies is also partly attributable to progress in DNA-sequencing technology. Using microarray chips, we can now look at millions of SNPs per patient or the entire coding sequence of the genome (termed the exome) in a single experiment that is both time-efficient and cost-effective.

What is a genome-wide association study?
A genome-wide association study generally compares genetic variants between patients with a particular clinical condition (cases) and people without the condition (controls), looking for statistically significant differences. As a tool for genetic discovery, these studies have revealed many avenues for further investigation in the pathogenesis of disease, as well as potential targets of therapy.

Using these studies, research groups around the world have found reproducible correlations between genetic variants and susceptibility to common adult-onset diseases.

Although many of the variants identified in these studies are associated with only a slightly higher risk of disease, the method is free of many of the inherent biases associated with clinical research. These studies permit a comprehensive, hypothesis-independent and unbiased scan of the genome to identify novel susceptibility factors, whereas earlier genetic epidemiology studies could take on only a handful of variables to evaluate at a time. Additionally, they are powered to detect very small increases (or decreases) in disease risk, previously outside the reach of linkage analysis. Polymorphisms (or, presumably, non-disease-causing DNA changes) discovered using these studies often correlate with clinical phenotypes or with levels of biomarkers, even if the genetic variants are not necessarily pathologic in themselves.

Thus, genome-wide association studies have led to important insights into the pathogenesis of multiple common diseases, such as inflammatory bowel disease and diabetes mellitus, and they are facilitating new treatment approaches. For instance, multiple studies have reproduced an association between Crohn disease and variation in the gene NOD2, whose protein product is implicated in bacterial product recognition, autophagy, and apoptosis. This discovery led to the investigation of new potential therapies for Crohn disease, ie, the tyrosine kinase inhibitors gefitinib (Iressa) and erlotinib (Tarceva), known to inhibit NOD2 activity, and to the prognostic use of the NOD2 genotype in Crohn disease (a field of study known as genotype-phenotype correlation).

Future advances will likely come from
looking at combinations of variants, which may carry a higher risk of disease than single variants.

### CORONARY HEART DISEASE: FRESH INSIGHT INTO AN OLD PROBLEM

Cardiovascular disease accounts for 30% of deaths worldwide. Of all the cardiovascular disorders, coronary heart disease is rising most rapidly in incidence, as the rest of the world adopts Western practices such as a high-calorie, high-fat, high-glycemic diet.

Hundreds of risk factors for coronary heart disease have been described. Three of them are clearly modifiable participants in the pathogenesis of atherosclerosis: hypertension, smoking, and elevated LDL-C. These and others form the basis for risk-assessment tools such as the Framingham risk score and the Prospective Cardiovascular Münster (PROCAM) study score. Other possible markers require further evaluation as to whether they are clinically useful and are direct mediators of coronary heart disease.

Because up to 40% of coronary deaths occur in people who lack conventional risk factors for it (eg, they do not smoke and they have normal levels of LDL-C and blood pressure), researchers are searching hard for new, potentially treatable risk factors. Of particular interest are components of inflammatory pathways linked with atherosclerosis and coronary heart disease. The identity of the key inflammatory factors that cause arterial plaque formation and rupture continues to be studied.

CRP, an acute-phase reactant produced by the liver in response to inflammation, has received much attention, as serum CRP levels correlate with a number of other, confounding risk factors. Therefore, much basic research and clinical observation had to be done before we could say that LDL-C plays a role in the pathogenesis of coronary heart disease.

Initially an association between LDL-C and heart disease was noted. Then, studies of familial hypercholesterolemia uncovered genetic abnormalities that increase LDL-C levels and, thereby, the risk of coronary heart disease—eg, mutations in the LDL receptor gene, the apolipoprotein B (APOB) gene at its LDL receptor-binding domain, LDL-RAP1 (a gene encoding an accessory adaptor protein that interacts with the LDL receptor), and PCSK9 (a gene that codes for proprotein convertase subtilisin-kexin type 9 protease). Conversely, specific loss-of-function truncating mutations of PCSK9 that reduce LDL-C levels are associated with strong protection against coronary heart disease. Other gene mutations that reduce LDL-C also lower the risk.

Further, a genome-wide association study identified multiple genetic variations associated with different forms of dyslipidemia, uncovering additional links between LDL-C and coronary heart disease.

Finally, randomized controlled trials of niacin, fibrates, and statins showed that these potent LDL-C-lowering agents reduce the rate of development or progression of coronary heart disease.

### C-reactive protein: Marker or mediator?

Unlike LDL-C, no familial syndromes of coronary heart disease have been recognized in patients who have isolated high serum levels of CRP. Since many substances in addition to CRP increase in concentration in both acute and chronic inflammatory states, agents that lower CRP in a targeted manner would be needed for large prospective, randomized trials to show whether CRP plays a direct role in coronary heart disease. A specific CRP inhibitor, 1,6-bis(phosphocholine)-hexane, may aid in these efforts, although it is not orally bioavailable and has a very short serum half-life.

The JUPITER trial. Statins lower levels of both LDL-C and CRP. The justification for the Use of Statins in Primary Prevention: an
Intervention Evaluating Rosuvastatin (JUPITER) trial was designed to find out whether statins alter coronary risk in patients with “normal” LDL-C levels (< 130 mg/dL) and elevated CRP levels (> 2 g/L).25

In this prospective, randomized trial, statin treatment resulted in a dramatic risk reduction of 40% to 50% in multiple coronary end points, as well as a reduction in CRP levels of 37% compared with placebo. However, LDL-C levels fell by 50%, confounding the effect on CRP, as the lower coronary event rate could alternatively be explained by the effect of lower-than-normal LDL-C levels. Thus, a causative link between CRP and coronary heart disease could not be proved.26

Though ongoing trials may further illuminate the role of inflammation in the development of coronary heart disease, and specific CRP inhibitors are in development, we have few tools to answer the fundamental question of whether CRP itself is an active participant in cardiovascular disease progression or if it is a bystander marker, helping to define risk for patients who develop coronary heart disease without other known risk factors.

Of note, adding CRP to the Framingham risk score does not improve its predictive power very much in any age group.27,28 Nevertheless, for certain end points, such as the long-term rate of death after percutaneous coronary intervention29 or of cardiovascular death immediately after coronary artery bypass grafting,30 CRP levels predict coronary events reliably.

Mendelian randomization offers a way to evaluate the correlation between coronary heart disease development and CRP levels independent of other risk factors.

How many heart attacks in people with or without polymorphisms?
Mendelian randomization takes advantage of a basic genetic principle, ie, the independent assortment of traits. According to Mendel’s second law, alleles for different traits are inherited independently of one another. Therefore, the gene that encodes CRP and other genes that influence its circulating level are presumably inherited independently from other genes that influence coronary risk.

In typical studies of CRP, participants are grouped according to whether they have high or low CRP levels. In these studies, confounding variables congregate in these two groups. For example, people with high CRP may be more likely to smoke and to have a higher body mass index and higher lipid levels—all of which influence cardiovascular outcomes. It is therefore difficult to tease out the effect of CRP levels from other background risk factors.

In contrast, in studies using mendelian randomization, patients are grouped according to whether they have a variant that affects the substance being studied (eg, CRP), and outcomes are compared between the two genetic groups.

Strengths and limitations of this method
By randomizing research subjects by gene variants affecting CRP levels, it is theoretically possible to achieve more equal stratification and minimize confounding between subgroups.33

Mendelian randomization should also address the possibility of “reverse causality,” when the intermediate trait with a potential role in disease development (eg, CRP) is actually regulated by the disease state itself (ie, “inflammation of atherosclerotic cardiovascular disease”).34

A limitation of mendelian randomization is that different genes influencing the biomarker under investigation must be proven to be truly randomly assorted among populations. It cannot be assumed that levels of a
biomarker are equally distributed across cases and controls when there may in fact be non-random genetic associations.

For instance, if SNPs in various genes that affect creatine kinase levels were being compared to cardiovascular outcome, it would be important to take into account that baseline creatine kinase levels are higher in African Americans as well as in men in interpreting the study data.\textsuperscript{35}

\subsection*{The Elliott Study (2009)}

In a study published in 2009, Elliott et al\textsuperscript{1} mined genome-wide data collected over the last decade to bring more clarity to the issue of causality between elevated CRP and heart disease.

To accomplish mendelian randomization, the authors assessed SNPs that affect circulating CRP levels in combined sets of 28,000 cases and 100,000 controls—robust population sizes. The SNP variants included were associated with approximately 20\% lower CRP levels. This degree of CRP reduction should correspond to a 6\% reduction in coronary risk as predicted by meta-analysis of observational studies.

\textbf{No association between low-CRP variants and heart disease}

The authors found significant associations between these SNPs and CRP levels and between CRP levels and coronary heart disease, but not between the SNPs and coronary disease when results for three SNPs were combined and standardized to a 20\% lower CRP level (odds ratio 1.00, 95\% confidence interval 0.97–1.02).\textsuperscript{1}

In view of the lack of association between coronary heart disease and SNPs that affect CRP levels, the authors suggested that the observational data linking CRP levels and coronary disease may have been confounded by other risk factors, or that the trend is due to reverse causation (the inflammatory response associated with atherosclerosis elevates CRP) rather than CRP’s directly causing heart disease.

These findings have important implications for management of cardiovascular disease, as therapeutic strategies to reduce plasma CRP levels are less likely to be beneficial.

The authors also described other genetic variants that may affect coronary heart disease. Carriers of minor alleles of SNPs in the gene for the leptin receptor \textit{LEPR} and the \textit{APOE-CI-CII} cluster showed a significantly higher risk of coronary heart disease.\textsuperscript{1} However, both variants were associated with lower levels of CRP (and, for the SNP in \textit{LEPR}, lower body weight and body mass index), suggesting that the links with coronary heart disease are not mediated by CRP. These findings illustrate the ability of genome-wide association studies to identify novel susceptibility loci for complex disease without limiting investigation to genes previously thought to take part in coronary heart disease.

In view of the evidence from this study, it seems that the benefits accruing to patients with high CRP from lipid-lowering therapy as demonstrated in the JUPITER trial are likely not the result of CRP-lowering per se, but rather are the result of action on the underlying pathology that leads to elevation of inflammatory markers, including CRP. As an editorial accompanying the study by Elliott et al pointed out, the work not only provides important information in the effort to identify genetic markers associated with complex disease, but it also helps discern the role of the genes and their products in the progress and treatment of common diseases.\textsuperscript{36}

Subsequent studies of CRP and the “directionality” of its role in coronary disease,\textsuperscript{37} as well as in other conditions such as obesity and cancer,\textsuperscript{38,39} have carried on the strategy of Elliott et al, providing further evidence for the function of CRP as a bystander in the inflammatory response and complex disease progression.

\subsection*{Implications of these Findings}

Tools now exist to leapfrog the randomized controlled trials that have been the primary way of examining the role of potential mediators of common diseases. Mendelian randomization aids in determining whether biomarkers are involved in disease pathogenesis, are simply bystanders, or are secondary markers caused by the disease itself. While randomized controlled trials will still be important,
C-REACTIVE PROTEIN

this new approach offers the power of evaluating much larger sample sizes and more equally stratifying confounding factors between study groups by relying on independent assortment of genetic traits.

In medical care today, the prevention of coronary heart disease entails aggressive treatment of hypertension and hyperlipidemia, along with lifestyle modifications such as balanced diet, routine exercise, and smoking cessation. Given the large numbers of patients at risk, even with low risk scores using currently identified risk factors, more specific and sensitive markers (or panels of such markers) of cardiovascular risk are needed.

In the personalized medicine of the future, we will rely on markers that not only identify people at higher risk, but also tell us who would benefit from certain therapies. From the JUPITER trial, we understand that patients with elevated CRP levels may be appropriate candidates for statin therapy even if they have normal levels of LDL-C.36 The study by Elliott et al steers us away from using CRP-affecting SNPs in predicting the course of disease and also from the belief that targeting CRP alone would be a worthwhile therapeutic strategy.

The inflammatory hypothesis of coronary heart disease remains a very important area of investigation, and CRP may turn out to be one of the best biomarkers we have to predict the progression of coronary diseases. But the study by Elliott et al demonstrates that CRP-lowering drugs are unlikely to be magic bullets.

Most importantly, geneticists will partner with clinical researchers to answer important questions about biomarkers and genes, capitalizing on large sets of population data.

REFERENCES


ADDRESS: Leslie Cole Manace, MD, MPhil, Department of Genetics, Kaiser Permanente Oakland Medical Center, 280 West MacArthur Boulevard, Oakland, CA 94611-5693; e-mail leslie.c.manace@kp.org.