With rapid and dramatic success, the Genome-Wide Association Study (GWAS) has proven to be an effective tool for the discovery of unsuspected genetic determinants of common disorders and has opened a new armamentarium for the pathophysiologic exploration of numerous diseases. The best illustration of the feasibility and strength of the GWAS approach was demonstrated in June of 2007 by a consortium of more than 50 British research groups participating in the Wellcome Trust Case-Control Consortium (WTCCC). Working collaboratively, the WTCCC investigators studied 14,000 cases of 7 common diseases and 3000 shared controls and identified 24 independent association signals, 1 in bipolar disease, 1 in coronary disease, 9 in Crohn disease, 3 in rheumatoid arthritis, 7 in type 1 diabetes, and 3 in type 2 diabetes, each with a statistical effect approaching or exceeding genome-wide levels of significance (1). Remarkably, the key finding in this study for coronary heart disease—a clear association between vascular risk and common variation in the region of chromosome 9p21.3—was rapidly validated in a series of similar GWAS studies, including the Cardiogenics Consortium; the Ottawa, Dallas, and Framingham Heart Studies; and the DeCode Genetics program in Iceland (2–5). The chromosome 9p21.3 region contains the coding sequences of genes for 2 cyclin-dependent kinase inhibitors known to play roles in tumor suppression, cell proliferation, and apoptosis. Thus, these validated GWAS findings for coronary disease not only raise the concept of a novel genetic determinant of disease, but also provide strong pathophysiologic support for prior work linking each of these processes directly to atherogenesis and plaque disruption.

What is less clear from emerging GWAS studies is whether or not the discovery of new gene-disease associations will ultimately help identify persons at high risk, particularly for complex disorders such as atherothrombosis, for which major environmental determinants exist. As described by the Framingham Heart Study investigators as far back as 1961, older age, smoking, hypertension, hyperlipidemia, and diabetes are common determinants of coronary heart disease (6), and these “traditional” risk factors have been codified into global risk scores for the prediction of cardiovascular risk (7, 8). It is also widely recognized, however, that 15% to 20% of cases of incident myocardial infarction or ischemic stroke occur among individuals without these risk factors, and that nearly 50% occur in the absence of hyperlipidemia. Thus, there is considerable interest in finding new markers able to predict the occurrence of atherothrombosis more accurately and with improved risk classification.

To date, work on novel atherothrombotic risk factors has largely focused on biomarkers of hemostasis and thrombosis (fibrinogen, von Willebrand factor, tissue plasminogen activator, plasminogen activator inhibitor, factor VII) or on biomarkers of inflammation [C-reactive protein (CRP), serum amyloid A, interleukin-6, sCD40L, intercellular adhesion molecule-1] (9). Unfortunately, with the exception of CRP, most of these novel biomarkers have failed to show additive value for the prediction of future vascular events. For example, in the development and validation of the Reynolds Risk Score, in which a panel of 35 factors was assessed among 24,558 initially healthy American women, the best simplified prediction model for future vascular events included the traditional Framingham risk determinants and only 2 other factors—high-sensitivity CRP (representing inflammatory risk) and parental history of myocardial infarction before age 60 (representing genetic risk) (10). These 2 new factors reclassified 40% to 50% of those at intermediate risk into higher or lower risk categories, an effect with important clinical implications for the targeting of preventive therapies (see www.reynoldsriskscore.org).

In this issue of Clinical Chemistry, Talmud et al. directly address the clinical question of whether or not detection of polymorphism in the 9p21.3 region impacts on global risk prediction (11). In a carefully performed and thoughtful analysis, the authors first evaluated the relationship of rs10757274 (a previously defined tag single-nucleotide polymorphism for the 9p21.3 region) as a determinant of vascular risk in the prospective Northwick Park Heart Study-II (NPHS-II) in which 2742 men were followed over a 15-year period for incident coronary heart disease. As in prior studies, polymorphism at this locus was common (frequency of the G allele 0.48, 95% CI 0.47–0.50), and compared to men homozygous for the common A allele, men homozygous for the G allele had a 1.6 times greater hazard ratio for future vascular events (95% CI 1.12–2.28). These effects were largely independent of traditional risk markers and consistent for the individual endpoints of myocardial infarction and coronary artery
bypass surgery. Somewhat surprisingly, the magnitude of effect on risk associated with polymorphism at rs10757274 was not materially altered in analyses after further adjustment for family history, despite the fact that family history is a major risk in the NPHS-II population (12).

What makes the data from Talmud et al. novel is that the authors then take a second important step and ask whether polymorphism at rs10757274 substantively adds to risk prediction based on traditional markers such as those used in the Framingham Risk Score. Here, readers of the genetic-epidemiology literature will find data to satisfy a full range of opinions. For those who are skeptical that genetic information can improve risk prediction, ample data are supplied, in that the area under the ROC curve (as defined by the c-statistic) increased from 0.62 to only 0.64 when data on rs10757274 were added to a panel of traditional risk factors, a small and nonsignificant effect (P = 0.14). Thus, the authors conclude that, on its own, genetic variation near chromosome 9p21.3 does not add to the overall risk prediction in the NPHS-II cohort.

On the other hand, statisticians have recently come to recognize that the c-statistic (an effective tool for evaluating diagnostic tests) may be ill suited for the evaluation of risk prediction models. For example, Cook has demonstrated that widely accepted risk factors such as LDL and HDL cholesterol, hypertension, and even smoking typically have multivariable relative risks <2.0 and thus on their own would also not substantively improve the c-statistic (13). Further, in risk prediction modeling, calibration and reclassification may play as great a role in defining clinical utility as does discrimination, the test characteristic summarized by the c statistic (14).

Recognizing this issue, Talmud et al. also provide promising data for those who believe polymorphism data will ultimately impact on our ability to predict clinical risk. Specifically, despite marginal improvement in the c statistic, the authors also show that knowledge of rs10757274 genotype does improve model fit as described by the Bayes Information Criteria and that model calibration as assessed by the Hosmer-Lemeshow statistic improves when genotype information is included. Although the total number of individuals reclassified as a result of rs10757274 data was small in the NPHS-II cohort, those reclassified were often at intermediate risk, the category for which improved risk prediction is likely to be most helpful. To further address reclassification, Talmud et al. additionally model the effect of 10 hypothetical randomly assigned gene variants with similar allele frequencies and risk characteristics as rs10757274. Although such variants are not known to exist, the authors find that this multimarker approach could indeed effectively re-classify large proportions of patients in much the same way the multimarker Framingham or Reynolds Risk Scores combine data from several pathways to improve risk detection.

Thus, the study from Talmud et al. nicely demonstrates the clinical complexity that will accompany attempts to treat specific single-nucleotide polymorphisms as risk factors for multifactorial disorders such as atherothrombosis. This complexity is not limited to genetic discovery, but will also impact on the development of other high-throughput technologies in molecular biology, including proteomics. In our opinion, whether or not fishing for new genes will ultimately catch heart disease far enough upstream for prevention to be successful remains an open question. In the meantime, considerable pathophysiologic insight will come from ongoing GWAS studies, and the potential to find new targets for therapy will remain a driving force for this arena of research.

Grant/funding Support: Supported by research grants to J.E. and P.M.R. from the Leducq Foundation, Paris FR.

Financial Disclosures: None declared.

-- Joseph Emmerich1,2*
-- Paul M Ridker3
1  INSERM
Paris, France
2  Université Paris Descartes
Faculté des Sciences Biologiques et Pharmaceutiques
Paris, France
3  Center for Cardiovascular Disease Prevention
Brigham and Women’s Hospital
Harvard Medical School
Boston, MA
* Address correspondence to this author at:
INSERM U765
Faculté des Sciences Biologiques et Pharmaceutiques
4 avenue de l’Observatoire
75006 Paris
Fax 33-1-44-07-17-72
E-mail joseph.emmerich@egp.aphp.fr

References
1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature (Lond) 2007;447:661–78.


DOI: 10.1373/clinchem.2007.100313