The search for new cardiovascular biomarkers

Robert E. Gerszten[1,2] & Thomas J. Wang[1]

Despite considerable advances in the treatment of cardiovascular disease, it remains the leading cause of death in developed countries. Assessment of classic cardiovascular risk factors — including high blood pressure, diabetes and smoking — has a central role in disease prevention. However, many individuals with coronary heart disease (a narrowing of the blood vessels that supply the heart) have only one, or none, of the classic risk factors. Thus, new biomarkers are needed to augment the information obtained from traditional indicators and to illuminate disease mechanisms.

From a clinical perspective, biomarkers have a variety of functions, which correspond to different stages in the development of a disease. Biomarkers can assist in the care of patients who have no apparent disease (screening biomarkers), those who are suspected to have disease (diagnostic biomarkers) and those with overt disease (prognostic biomarkers). At present, diagnostic and prognostic cardiovascular biomarkers are available, but there are no widely accepted biomarkers for screening. This has been an active area of investigation, because preventing events in those at risk of cardiovascular disease is likely to have a substantial impact on the overall public-health burden. In this progress article, we discuss the search for clinically useful biomarkers, focusing on how emerging technologies are being integrated into current efforts.

Current status of cardiovascular biomarkers

Circulating biomarkers that have been successfully incorporated into cardiology practice fall into the category of diagnostic biomarkers: troponin I and troponin T for myocardial infarction (heart attack) and brain natriuretic peptide (BNP) for heart failure. These biomarkers also seem to have prognostic value in patients presenting with acute myocardial infarction or heart failure, and studies are ongoing to assess whether the biomarkers could be used to guide specific treatment decisions. Prognostic biomarkers might also have valuable roles as surrogate end points for therapy or clinical trials.

Several potential screening biomarkers have attracted attention because of their ability to predict future cardiovascular events and their mechanistic involvement in atherosclerosis-associated pathways. These include biomarkers associated with inflammation (C-reactive protein, interleukin-6 and lipoprotein-associated phospholipase A_2), haemostasis/thrombosis (fibrinogen and plasminogen-activator inhibitor 1), neurohormone activation (renin and BNP), insulin resistance (insulin and haemoglobin A1C) and endothelial dysfunction (homocysteine and urinary albumin). The value and appropriate use of these biomarkers remain a source of debate, however[1].

A recent investigation from the Framingham Heart Study illustrates some of the uncertainties surrounding the use of the available screening biomarkers in ambulatory individuals. The study evaluated 10 cardiovascular biomarkers in more than 3,000 people who were followed for nearly a decade[2]. Several biomarkers were found to be significant predictors of death (C-reactive protein, BNP, urinary albumin, renin and homocysteine) or cardiovascular events (BNP and urinary albumin). When biomarkers were combined into a 'multimarker' score, individuals with high scores had a fourfold higher risk of death and a twofold higher risk of major cardiovascular events than people with low scores. However, the multimarker score was associated with only a moderate increase in the area under the receiver-operating-characteristic curve (AUC) compared with a risk score based on conventional risk factors alone (Fig. 1). The AUC incorporates two features of a screening test — sensitivity and specificity — and is an objective measure of the test's ability to distinguish between individuals with and without disease. Although complementary metrics for evaluating new risk markers exist, including model calibration and reclassification percentage, these findings suggest that current screening biomarkers add only moderately to the ability of classic risk factors to predict future events in an individual person.

Limitations of the available screening biomarkers warrant consideration. As is the case for any biological analyte, biomarker concentrations are broadly distributed. Thus, even if the underlying distribution differs according to disease status, concentrations in individuals with and without disease overlap substantially[3]. Moreover, most current biomarkers participate in pathways that are known to be associated with atherosclerotic cardiovascular disease, such as those involved in inflammation and cholesterol biosynthesis. Consequently, the available biomarkers provide information that is often correlated with what is already known or being measured. Although correlated biomarkers can underscore the importance of a biological pathway, they might not provide a substantial increase in predictive value. This point is illustrated by Margaret Pepe and Mary Lou Thompson[4], who carried out simulations using two hypothetical cancer biomarkers. Assuming an AUC of 0.80 with one biomarker, they showed that the inclusion of a second biomarker raised the AUC to 0.88 if the two biomarkers were weakly correlated but only to 0.83 if the two biomarkers were moderately correlated. This result translates into a sensitivity of 80% with two weakly correlated biomarkers but a sensitivity of 70% with two moderately correlated biomarkers (assuming a false positive rate of 20%). The implication is that an additional 10 individuals for every 100 people destined to develop disease would be identified with the use of less-correlated biomarkers, a clinically meaningful difference. The difference would be further magnified if multiple biomarkers were included, emphasizing that a large number of correlated biomarkers is substantially less informative than a small number of uncorrelated biomarkers.

[1]Massachusetts General Hospital, Cardiology Division, 55 Fruit Street, Boston, Massachusetts 02114, USA. [2]Massachusetts General Hospital, Center for Immunology and Inflammatory Diseases, 149 13th Street, Charlestown, Massachusetts 02129, USA.

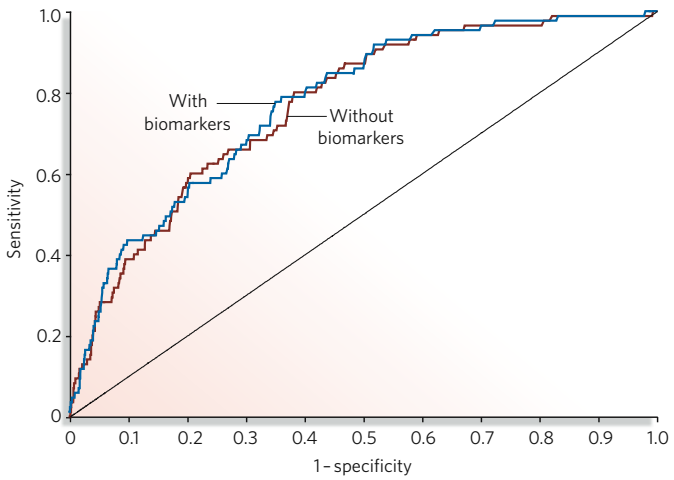

Figure 1 | Receiver-operating-characteristic curves for the prediction of cardiovascular events. The curves depict results using risk prediction models that include classic cardiovascular risk factors alone or with multiple biomarkers (that is, incorporating the multimarker score). The diagonal line denotes an uninformative test, with an AUC of 0.50. A test with perfect discrimination yields an area under the receiver-operating-characteristic curve of 1.0. Sensitivity refers to the proportion of diseased individuals with a positive test (the true positive rate). Specificity refers to the proportion of non-diseased individuals with a negative test (the true negative rate). (Reproduced, with permission, from ref. 3).

Despite the desirability of using multiple biomarkers, there are barriers to identifying new biomarkers, particularly for screening or prognostic uses. One difficulty is the requirement for large, adequately powered clinical studies. Large studies are necessary because the predictive effects of new biomarkers might be smaller than those observed with classic risk factors and because multiple biomarkers are often studied concurrently. Meta-analyses of individual participant data from multiple cohorts provide a potentially valuable tool for circumventing sample-size limitations in single cohorts.

Genetics and transcriptomics

The limitations of the available biomarkers for screening or prognostic uses underscore the importance of identifying 'orthogonal' (that is, uncorrelated) biomarkers associated with new disease pathways. Most available biomarkers have been developed as an extension of targeted physiological studies, investigating known pathways such as those involved in inflammation or haemostasis. By contrast, emerging technologies are beginning to allow the systematic, unbiased characterization of variation in genes, RNA, proteins and metabolites associated with disease conditions (Fig. 2).

Genetic studies will undoubtedly identify variants that could be biomarkers themselves or will point to circulating markers for further exploration. Current technology allows the examination of hundreds of thousands of single-nucleotide polymorphisms (SNPs) in affected and unaffected individuals to search for significant associations with disease. Although numerous studies have reported associations between variants in candidate genes and coronary heart disease, these findings have almost uniformly failed to be reproducible in subsequent analyses[5]. In contrast to candidate gene studies, genome-wide association studies provide an unbiased scan of genomic sequence variants, an approach likely to reveal new disease-associated pathways. Using this method, three groups recently identified a locus on chromosome 9p21 that is associated with early-onset myocardial infarction[6–8]. The chromosomal region identified did not contain genes recognizably associated with established risk factors for coronary heart disease such as plasma lipoproteins, hypertension or diabetes. Intriguingly, the genes encoding the cyclin-dependent kinase inhibitors INK4A and INK4B — which are known to affect cellular senescence, apoptosis and stem-cell function — are located near this chromosomal region. A recent genome-wide association study with more direct implications for the development of plasma biomarkers identified variants in the gene encoding the chemoattractant cytokine CXCL12 in individuals with premature atherosclerosis[6]. CXCL12 is involved in a variety of pathways, including cardiac development, platelet activation and stem-cell recruitment[9–11]. In the absence of expression and functional data about CXCL12 variants, however, speculation on the biological importance of this finding remains at its earliest stages.

Genetic data also provide an opportunity to assess the causality of biomarkers for disease. For a biomarker that has a causal role, the expected random distribution in a population of a polymorphism that determines high or low biomarker concentrations would be skewed in individuals, depending on their disease status. Data from 'mendelian randomization' studies are accumulating for several biomarkers such as C-reactive protein, fibrinogen and homocysteine[12,13].

The relative maturity of transcript-profiling techniques, which have been used successfully for cancer diagnostics, has led to their integration into the cardiovascular biomarker field. The application of transcriptional approaches towards the identification of new cardiac biomarkers in humans is, clearly, limited by the availability of the most relevant tissue, the heart. This difficulty has been circumvented in several recent studies[14,15]. For example, Richard Lee and colleagues[14] discovered that *ST2* messenger RNA is markedly upregulated in cultured cardiomyocytes after applying mechanical strain, an *in vitro* model that recapitulates some aspects of human pathophysiology. ST2 is a member of the interleukin-1-receptor family and exists in two forms: a membrane-bound receptor, and a truncated receptor that is soluble and can be detected in human serum. The *in vitro* data of Lee and colleagues[14] suggest that ST2 might be produced in conditions of myocardial overload such as congestive heart failure. Indeed, soluble ST2 serum concentrations predict outcomes in patients with heart failure, and an increase in soluble ST2 concentrations over time is associated with worsening prognosis[16]. Furthermore, concentrations of soluble ST2 also predict mortality and heart failure in patients after myocardial infarction[17].

In addition to the limitation of obtaining relevant tissue from myocardium or blood vessels, another obstacle to discovering biomarkers of acute heart disease is that the biomarkers identified might reflect pathological mechanisms that are associated not with events that trigger acute disease (for example, plaque rupture and thrombosis in the case of myocardial infarction) but, instead, with the downstream consequences of the resultant pathology. To address this potential pitfall, Daniel Simon and colleagues[15] profiled genes expressed by circulating platelets, which lack nuclear DNA but retain megakaryocyte-derived mRNAs and the translational machinery for protein biosynthesis. Transcriptional profiling of platelets can thus provide a window on gene expression that preceded the onset of acute events such as myocardial infarction.

Using this approach, one of the strongest discriminators between patients with acute myocardial infarction and those with stable coronary heart disease was the secreted protein myeloid-related protein 14 (MRP14). The diagnostic utility of MRP14 was validated in a prospective, nested, case–control study among apparently healthy women to assess the association of plasma MRP14 with the risk of future cardiovascular events, including myocardial infarction, stroke and cardiovascular-associated death[15]. In this study, the risk of a first cardiovascular event increased with each quartile of MRP14 concentration such that women with the highest concentrations had a fourfold higher risk of any cardiovascular event. The risk conferred by increasing MRP14 concentrations was independent of classic risk factors and C-reactive protein concentration. Thus, study of the platelet transcriptome led to the identification of a biomarker that can predict the risk of future cardiovascular events in healthy individuals.

Proteomics and metabolomics

Of the emerging platforms for biomarker discovery, perhaps none has garnered more recent attention than proteomics and metabolomics.

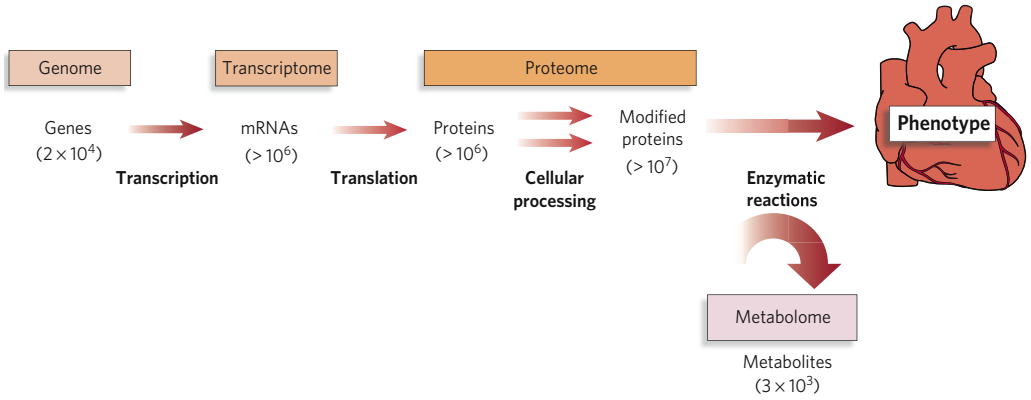

Figure 2 | The conceptual relationship of the genome, transcriptome, proteome and metabolome. Informational complexity increases from genome to transcriptome to proteome. The estimated number of entities of each type of molecule in a typical cell is indicated in parentheses.

Although still in their infancy compared with other approaches, these technologies offer complementary insight into the full complexity of the disease phenotype (Fig. 2). The set of proteins and metabolites in a cell can change rapidly in response to environmental cues, so the proteome and metabolome — the latter being defined as biochemicals, including lipids, sugars, nucleotides, amino acids and related amines, of less than 2 kDa — reflect the state of a cell or group of cells at a given time.

This complexity presents an analytical challenge, particularly as it applies to searching for biomarkers in the blood. Many cell types contribute to the plasma proteome and metabolome, which have so far been poorly characterized. In the case of the plasma proteome, the 22 most abundant proteins, including albumin and the immunoglobulins, constitute 99% of the total proteome mass[18]. Many of the biologically interesting molecules relevant to human disease are low-abundance proteins. For example, cardiac biomarkers such as troponin are found in the nanomolar range, insulin in the picomolar range, and tumour-necrosis factor in the femtomolar range. There are estimated to be tens of thousands of unique protein species in the plasma, with concentrations spanning a range of more than ten orders of magnitude. Indeed, it has been suggested that the entire set of polypeptides resulting from alternative splicing and post-translational modifications is represented in the plasma proteome in humans (estimated at more than 300,000 species)[18]. This is because the protein content of plasma includes proteins of all functional classes and from apparently all cellular localizations — most of the low-abundance proteins in plasma are intracellular or membrane-associated proteins that are present in plasma as a result of cellular turnover[19]. Recent estimates suggest that the human metabolome comprises about 3,000 small molecules and is thus more tractable than the human proteome, although in the absence of definitive data sets this remains speculative. Recently, collaborative efforts have been organized to catalogue the plasma metabolome[20].

Two core technologies have emerged as the workhorses of plasma metabolite profiling: nuclear magnetic resonance (NMR) spectroscopy and tandem mass spectrometry. NMR spectroscopy, which is used almost exclusively for analysing small biochemicals in the blood, requires relatively little sample preparation and is non-destructive, allowing further analyses. However, the method tends to have low sensitivity and can detect only highly abundant analytes. By contrast, tandem mass spectrometry, coupled with liquid chromatography, has a much higher sensitivity for both small molecules and peptides and is also applicable to a wide range of biological fluids (including serum, plasma and urine). Recent advances in tandem mass-spectrometry technology are now enabling researchers to determine analyte masses with such high precision and accuracy that peptides and metabolites can be identified unambiguously even in complex fluids.

These technologies can be used to characterize biological fluids in either a targeted manner or a pattern-discovery (fingerprint) manner. In the former, the investigator targets a predefined set of analytes to be quantified. In the latter, the investigator is faced with a complex pattern of peaks, and the molecular identities of the species giving rise to many of these peaks are generally not known. The targeted approach is more limiting than the pattern-discovery approach; however, the analysis of the data is more straightforward in this case, because the analytes giving rise to the signals are already known. By contrast, the pattern-discovery approach is inherently less biased; however, the unambiguous identification of the peaks can be laborious and difficult, and the observed associations might be spurious.

Future directions

The application of mass spectrometry and related techniques to biomarker discovery is based on decades of intensive efforts to understand and diagnose congenital errors of metabolism in infants. David Millington and colleagues[21] pioneered the use of methods based on tandem mass spectrometry for monitoring fatty-acid oxidation products, as well as organic acids and amino acids. Their work has culminated in universal neonatal screening for metabolic disorders in many geographic locations[22], allowing the identification of infants with fatty-acid oxidation disorders, organic acidaemias and aminoacidopathies. In many instances, rapid identification of these disorders triggers intervention in the form of dietary modulation, with beneficial therapeutic effects. A global metabolomic or proteomic analysis in more common diseases might similarly spotlight pathways that could be modulated by diet or drugs.

The application of proteomics or metabolomics to common cardiovascular diseases has potential obstacles, however. For acute events, such as myocardial infarction, the inherent unpredictability of when the event occurs often precludes prior blood sampling. Furthermore, the effects on biomarker concentrations are likely to be far more subtle than in the case of congenital errors of metabolism, and the extent of interindividual variability of the human proteome and metabolome remains unclear. It is clear, however, that this variability can be further compounded by environmental factors, including drug exposures. Indeed, one report applied pattern-discovery techniques to proton NMR spectra of human serum to aid in the non-invasive diagnosis of chronic coronary heart disease[23]. However, the pattern of metabolites that generated several of the spectroscopy peaks ultimately proved to be confounded by statin therapy[24,25]. Although studying samples from large patient cohorts, stratified by known risk factors or exposures, could minimize the impact of confounding clinical variables, the throughput of the detection technologies is not yet adequate for such analysis. An initial strategy to overcome such problems is to focus on well-characterized individuals who are given a physiological challenge — for example physical exercise[26] or glucose loading — and sampled over time. These individuals can thus function as their own biological control, and this type of dynamic analysis is likely to prove more reliable than analyses based on static phenotypes. Biomarkers derived from smaller, carefully phenotyped cohorts can subsequently be validated in large, more heterogeneous populations.

The identification of new biomarkers of cardiovascular disease will depend on the complementary power of genetics, transcriptional profiling, proteomics and metabolomics. The ultimate test for new

biomarkers will be to ask whether, when combined with existing clinical risk factors, they improve the prediction of risk in an individual and hence can contribute to personalized medicine. In addition to screening biomarkers, diagnostic biomarkers are needed to aid the difficult diagnoses of acute events such as reversible myocardial ischaemia, pulmonary embolism and aortic dissection. It is a long journey towards the identification of a clinical biomarker and an arduous transition from the research environment to routine clinical practice. However, there is a clear mandate for harnessing emerging technologies to systematically assess variation in genes, RNA, proteins and metabolites, as well as for identifying orthogonal biomarkers, which are unlikely to be found by focusing on well-studied pathways. ■

1. Wilson, P. W. CDC/AHA workshop on markers of inflammation and cardiovascular disease: Application to clinical and public health practice: Ability of inflammatory markers to predict disease in asymptomatic patients: a background paper. *Circulation* **110,** e568–e571 (2004).
2. Wang, T. J. *et al.* Multiple biomarkers for the prediction of first major cardiovascular events and death. *N. Engl. J. Med.* **355,** 2631–2639 (2006).
3. Ware, J. H. The limitations of risk factors as prognostic tools. *N. Engl. J. Med.* **355,** 2615–2617 (2006).
4. Pepe, M. S. & Thompson, M. L. Combining diagnostic test results to increase accuracy. *Biostatistics* **1,** 123–140 (2000).
5. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4,** 45–61 (2002).
6. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357,** 443–453 (2007).
7. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316,** 1488–1491 (2007).
8. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316,** 1491–1493 (2007).
9. Tachibana, K. *et al.* The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* **393,** 591–594 (1998).
10. Peled, A. *et al.* The chemokine SDF-1 stimulates integrin-mediated arrest of $CD34^+$ cells on vascular endothelium under shear flow. *J. Clin. Invest.* **104,** 1199–1211 (1999).
11. Abi-Younes, S. *et al.* The stromal cell-derived factor-1 chemokine is a potent platelet agonist highly expressed in atherosclerotic plaques. *Circ. Res.* **86,** 131–138 (2000).
12. Timpson, N. J. *et al.* C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet* **366,** 1954–1959 (2005).
13. Wald, D. S., Law, M. & Morris, J. K. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *B. Med. J.* **325,** 1202 (2002).
14. Weinberg, E. O. *et al.* Expression and regulation of ST2, an interleukin-1 receptor family member, in cardiomyocytes and myocardial infarction. *Circulation* **106,** 2961–2966 (2002).
15. Healy, A. M. *et al.* Platelet expression profiling and clinical validation of myeloid-related protein-14 as a novel determinant of cardiovascular events. *Circulation* **113,** 2278–2284 (2006).
16. Weinberg, E. O. *et al.* Identification of serum soluble ST2 receptor as a novel heart failure biomarker. *Circulation* **107,** 721–726 (2003).
17. Shimpo, M. *et al.* Serum levels of the interleukin-1 receptor family member ST2 predict mortality and clinical outcome in acute myocardial infarction. *Circulation* **109,** 2186–2190 (2004).
18. Anderson, N. L. *et al.* The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3,** 311–326 (2004).
19. Omenn, G. S. *et al.* Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5,** 3226–3245 (2005).
20. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35,** D521–D526 (2007).
21. Roe, C. R., Millington, D. S. & Maltby, D. A. Identification of 3-methylglutarylcarnitine. A new diagnostic metabolite of 3-hydroxy-3-methylglutaryl-coenzyme A lyase deficiency. *J. Clin. Invest.* **77,** 1391–1394 (1986).
22. Frazier, D. M. *et al.* The tandem mass spectrometry newborn screening experience in North Carolina: 1997–2005. *J. Inherit. Metab. Dis.* **29,** 76–85 (2006).
23. Brindle, J. T. *et al.* Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nature Med.* **8,** 1439–1444 (2002).
24. Nielsen, E.-M. D. *et al.* The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* **52,** 573–577 (2003).
25. Kirschenlohr, H. L. *et al.* Proton NMR analysis of plasma is a weak predictor of coronary artery disease. *Nature Med.* **12,** 705–710 (2006).
26. Sabatine, M. S. *et al.* Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* **112,** 3868–3875 (2005).

Acknowledgements The authors are grateful for support from the National Institutes of Health (to R.E.G. and T.J.W.), the Donald W. Reynolds Foundation (to R.E.G.), the Fondation Leducq (to R.E.G.) and the American Heart Association (to T.J.W.).

Author Information Reprints and permissions information is available at npg.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to R.E.G. (rgerszten@partners.org).