Metabolomic Profiling for the Identification of Novel Biomarkers and Mechanisms Related to Common Cardiovascular Diseases

Form and Function

Svati H. Shah, MD, MHS; William E. Kraus, MD; Christopher B. Newgard, PhD

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, but their molecular etiology remains poorly understood, in part because they develop slowly as a result of a mixture of genetic and environmental factors. Given the complex nature of CVD, molecular profiling of processes more "proximal" to the disease than genetic markers may have great promise in revealing both "form" (novel biomarkers with clinical potential) and "function" (mechanisms of disease development) of CVD. CVD is often associated with conditions of perturbed energy homeostasis and metabolism, including obesity, insulin resistance, and diabetes. The strong relationship between CVD and certain circulating lipids such as cholesterol and triglycerides has been recognized for decades. However, beyond these well-established associations, there is a strong possibility that other links between metabolic dysregulation and CVD remain to be discovered, and as a corollary, that new metabolite signatures can be identified that enhance risk prediction models. Comprehensive metabolic profiling, or "metabolomics" is increasingly being applied to CVD, leading to recent discoveries with both form and function implications. Here we review recent progress in this rapidly expanding area.

Evolving Metabolomics Technologies

Metabolomics is a term used to describe the measurement of multiple small-molecule metabolites in biological specimens, including bodily fluids (urine, blood, saliva), tissues, and breath exhalate. Metabolomics is perceived as a very recent addition to the omics platforms relative to genomics, transcriptomics, or proteomics, but in fact, its origins are in the venerable discipline of analytic biochemistry. Recent technological advances have enabled high-throughput profiling of large numbers of metabolites in biological samples, with increasing application to disease research, including CVD. One seemingly attractive feature of metabolic profiling is the relatively small number of human metabolites (≈7000) relative to the estimated numbers of genes (25,000), transcripts (100,000), and proteins (1,000,000). However, in reality, metabolites exist in a very broad range of concentrations and exhibit remarkable chemical diversity, so there is no one instrument that can reliably measure all of the metabolites in the human metabolome in a single analysis. Instead, practitioners of this technology generally use a suite of instruments, most often involving different combinations of liquid chromatography (LC) or gas chromatography coupled with mass spectrometry (MS), to obtain broad coverage of metabolic space. More than tractability, the allure of metabolomics lies with the concept that metabolites fall downstream of genetic, transcriptomic, proteomic, and environmental variation, thus providing the most integrated and dynamic measure of phenotype and medical condition.

Metabolomic profiling is most often performed with either nuclear magnetic resonance (NMR) or MS. Individual analytes within the sample are separated by their magnetic resonance shift or mass-to-charge ratio, respectively, resulting in a spectral profile of separation. When working with biological fluids, NMR often does not require chemical manipulation of a sample, whereas MS usually requires derivatization of metabolites to produce ionic species that are more readily separated by the mass-to-charge ratio. MS enjoys a strong advantage in sensitivity compared with NMR and, through application of different combinations of chromatographic methods and mass/charge separation technologies, provides the ability to measure a broader array of small-molecule metabolites.

Metabolic profiling is often referred to as targeted or nontargeted. In the targeted approach, specific metabolites of known identity are profiled. In the case of targeted MS, this often involves the addition of multiple stable isotope-labeled standards to the biological sample before the extraction and derivatization steps to control for differences in analyte loss during sample processing and to compensate for ionization-suppression effects. Targeted analysis is often performed in a modular format in which a sample is split into aliquots that...
are extracted with different solvents and subjected to different derivatization steps tailored for the chemical properties of a cluster of analytes such as amino acids, acylcarnitines, acyl CoAs, organic acids, and nucleotides, in each case with the addition of a cocktail of cognate stable isotope standards to the sample. When performed with appropriate internal standards, an advantage of targeted methods is their quantitative precision. A disadvantage is their limitation in breadth of analysis; generally, several hundred metabolites in 6 to 7 chemical classes are covered by these methods. Nevertheless, targeted methods provide an excellent survey of metabolic fuel selection and a profile of energy-yielding metabolic pathways, including elements of mitochondrial metabolism.

Nontargeted profiling involves the use of NMR or MS for simultaneous measurement of as many metabolites as possible in a biological specimen. These approaches are generally used to compare 2 biological or clinical states and to report on differences between the 2 states based on peak areas of raw spectral data. Often, the chemical identity of the NMR- or MS-resolved peaks is not known a priori, and significant chemical/spectral analysis must be performed to define the molecular species. NMR spectroscopy is theoretically an excellent tool for nontargeted metabolic profiling because the method is based on the detection of spectral features emanating from any molecules that contain carbon or hydrogen and can be conducted on unadulterated biological samples. However, the method suffers from poor sensitivity and difficulties in deconvolution and normalization of complex spectra in biological samples. Generally, this limits NMR-based analysis to 100 of the most abundant metabolites in a sample.

Modern MS platforms such as those that incorporate time of flight, Orbitrap, and Fourier-transform ion cyclotron resonance mass analyzers offer very high mass resolution and mass accuracy. Coupling such MS instrumentation with high-resolution chromatographic technologies (e.g., ultra-high-pressure chromatography) has made it possible to resolve literally thousands of individual small molecules. The high mass accuracy of these methods facilitates peak identification through databases such as METLIN, KEGG, and the Human Metabolome Database. Less sophisticated analytic platforms such as gas chromatography/MS can also be used for nontargeted analysis, with peak identification aided by the use of retention time-locked spectral libraries. However, in the absence of the addition of a panel of cognate stable isotope standards, these methods are semiquantitative, meaning that when applied to clinical epidemiology, care must be taken in the interpretation of differences between groups. A combination of targeted and nontargeted approaches has been applied to CVD research in the past several years, and selected pertinent examples are discussed here.

Form: Metabolomic Profiling Identifies Novel CVD Biomarkers

Metabolomics Profiles of Coronary Artery Disease

Early attempts to define metabolic profiles of CVD involved proton NMR-based profiling. In one of the first studies, this method was used to compare the serum from 36 individuals with severe coronary artery disease (CAD) and 30 individuals with angiographically normal coronary arteries. The spectral profiles differed significantly and provided a >90% predictive power for discrimination between the 2 groups. Several of the discriminatory peaks were found in the major lipid regions of the spectra, leading to the suggestion that choline-containing metabolites were particularly diagnostic. However, a subsequent independent study using similar NMR-based methods found that the associations of specific peaks with CVD were weak when corrected for factors such as sex and use of statins. This second group of authors demonstrated that the 1H NMR technique could identify male versus female subjects with 100% accuracy but was much less able to identify statin users or subjects with CVD, despite expectations of substantial changes in lipid profiles in such groups.

In more recent work, targeted MS/MS-based methods were used to profile 45 plasma acylcarnitines and 15 amino acids in a larger study of CAD. With the use of principal components analysis for data reduction, 2 principal components analysis–derived metabolic factors were found to be associated with CAD: 1 composed of branched-chain amino acids (BCAAs) and their associated metabolites and 1 composed of urea cycle metabolites, including arginine and citrulline (Table). These metabolite clusters discriminated individuals with CAD from those without CAD in both discovery and validation data sets. An independent study that applied MS/MS in 1010 subjects undergoing elective cardiac evaluation also demonstrated a strong association of arginine and its downstream metabolites ornithine and citrulline with CAD and with major adverse cardiovascular events, including death, myocardial infarction (MI), and stroke. The authors used these analytes to calculate a GABR ratio (defined as arginine/[ornithine+citrulline]) and determined that a lower GABR score was strongly correlated with CAD and major adverse cardiovascular event risk. In all of these studies, the implicated metabolites were independently associated with CAD even after adjustment for traditional CVD risk factors and were found to have incremental value for discrimination of individuals with CAD relative to the common factors. Further discussion of the potential mechanistic significance of dysregulated BCAA and arginine metabolism is provided below.

Myocardial Infarction

Prior work has identified certain proteins such as troponin I, troponin T, C-reactive protein, and B-type natriuretic peptide as diagnostic markers for CVD events and heart failure. More recently, several studies have emerged that used metabolomics for predicting MI and its consequences. In our own work, targeted, MS/MS-based metabolomic profiling has identified novel metabolic biosignatures that predict CVD events independently and incrementally to known clinical risk factors. In an initial discovery cohort of 314 individuals with CAD, we found that a principal components analysis–derived metabolite factor composed of small- to medium-chain dicarboxylated acylcarnitines predicted death and MI (hazard ratio, 2.17; 95% confidence interval, 1.23–3.84; P=0.007). These results were replicated in a validation
cohort of 129 CAD subjects (63 with events, 66 control subjects; odds ratio, 1.52; 95% confidence interval, 1.08–2.14; \( P = 0.01 \)). To obtain further validation, we studied 2023 consecutive individuals undergoing cardiac catheterization from the CATHGEN cohort, including 38% found to have no clinically significant CAD on angiography.\(^{11}\) At the median follow-up of 3.1 years, there were 232 deaths within this group. We again found that principal components analysis–derived metabolite signatures composed of dicarboxylated acylcarnitines (again including the short-chain species but now also a separate long-chain cluster) were independently associated with mortality (hazard ratio, 1.17; 95% confidence interval, 1.05–1.31; \( P = 0.005 \); and hazard ratio, 1.14; 95% confidence interval, 1.05–1.25; \( P = 0.002 \), respectively), as well as with death or MI (Figure 1). We further showed the incremental utility of these (and other) metabolic signatures in discriminating CVD events, including a demonstration of statistical improvement of receiver-operating characteristic curves when comparing a robust, multivariable clinical model with a model inclusive of clinical variables and metabolites. We also showed potential incremental clinical utility for mortality models using reclassification analyses, with 27% of intermediate-risk individuals correctly reclassified (net reclassification improvement, 8.8%).

More recently, a study of a subset of 478 CATHGEN subjects who underwent coronary artery bypass grafting revealed that the same short- to medium-chain dicarboxylated acylcarnitine cluster (principal component) identified in the incident event studies was associated with adverse outcome after surgery in univariate analysis (\( P = 0.002 \)) and remained independently predictive of adverse outcomes in multivariable time-to-event analysis (adjusted hazard ratio, 1.23; \( P < 0.001 \)).\(^{12}\) Thus, the small- to medium-chain dicarboxylated acylcarnitine cluster is associated with CVD events in multiple studies of different designs. The recurrent association of dicarboxylated acylcarnitines with CVD events is encouraging and speaks to the importance of performing index, validation, and large cohort confirmatory studies when metabolomics profiling is used. However, we acknowledge that the recurrent nature of these associations is also likely to be influenced by the selective nature of the profiling conducted to date. A broader survey of metabolites with nontargeted or more inclusive targeted methods could yield other metabolites or metabolite clusters that are equally or more significantly associated with CVD events compared with the dicarboxylated acylcarnitine cluster, and this is currently under investigation. Current studies are focused on understanding the mechanistic implications of this metabolic signature via approaches discussed in the Function section that follows.

Another recent study has identified a fascinating link between the diet, gut microflora, host metabolism, and metabolomic biomarkers of risk for incident CVD events.\(^{13}\) The authors used a nontargeted LC-MS–based metabolomics approach to profile stable patients who subsequently experienced MI, stroke, or death over the ensuing 3-year period compared with age- and sex-matched control subjects who did not experience events. Among 18 analytes found to be associated with events in both the index (50 cases and 50 controls) and validation (25 cases and 25 controls) cohorts, 3

---

**Table. Association of Metabolite Clusters With Coronary Artery Disease**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Name</th>
<th>Individual Components</th>
<th>Initial Cohort ( P )</th>
<th>Replication Cohort ( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Medium-chain acylcarnitines</td>
<td>C8, C10:1, C12, C10, C12:1, C10-OH:C8:DC, C6:DC, C8:1-DC, C14:1, C14:2, C8:1-OH:C6:1-DC, C2</td>
<td>0.001, 0.01</td>
<td>0.15, 0.03</td>
</tr>
<tr>
<td>2</td>
<td>Long-chain acylcarnitines</td>
<td>C18:1, C18:2, C18, C16, C16:1, C20:4, C14:1, C14:2, C16:2, C14:1-OH</td>
<td>0.28, 0.34</td>
<td>0.03, 0.01</td>
</tr>
<tr>
<td>3</td>
<td>Long-chain dicarboxy/hydroxyl acylcarnitines</td>
<td>C18-0H/C16-DC, C20-0H/C18-0DC, C20:1-0H/C18:1-DC, C16-0H/C14-DC, C18-1-0H/C16-1-DC, C14-0H/C12-DC, C12-0H:C10-DC, C14:1-0H, C20</td>
<td>0.10, 0.36</td>
<td>&lt;0.0001, 0.004</td>
</tr>
<tr>
<td>4</td>
<td>BCAA related</td>
<td>Phe, Tyr, Leu/Ile, Met, Val, C5, Ala</td>
<td>0.002, 0.02</td>
<td>0.01, 0.03</td>
</tr>
<tr>
<td>5</td>
<td>Ketone related</td>
<td>Ket, Hbut, Ala (–), C2, C4:OH, C14:1</td>
<td>0.18, 0.33</td>
<td>0.54, 0.41</td>
</tr>
<tr>
<td>6</td>
<td>Various</td>
<td>C8:1, C10:3</td>
<td>0.56, 0.75</td>
<td>0.79, 0.92</td>
</tr>
<tr>
<td>7</td>
<td>Amino acids</td>
<td>Ser, Gly, FFA (–)</td>
<td>0.19, 0.13</td>
<td>0.04, 0.18</td>
</tr>
<tr>
<td>8</td>
<td>Dicarboxyls</td>
<td>C5-DC, C8:1-0H/C6:1-DC, C6-DC, C8:1, C18:2, C18, C16, C16:1, C16:1-DC, C14:1, C14:2, C16:2, C14:1-OH</td>
<td>0.73, 0.34</td>
<td>0.05, 0.57</td>
</tr>
<tr>
<td>9</td>
<td>Urea cycle related</td>
<td>Arg, His, Cit, C14-DC:C4:DC (–)</td>
<td>0.0004, 0.004</td>
<td>0.01, 0.01</td>
</tr>
<tr>
<td>10</td>
<td>Short-chain acylcarnitines</td>
<td>C3, C4:0:C4, C5</td>
<td>0.02, 0.19</td>
<td>0.72, 0.92</td>
</tr>
<tr>
<td>11</td>
<td>Various</td>
<td>C5:1, C18:2-0H (–), C22 (–)</td>
<td>0.62, 0.13</td>
<td>0.03, 0.01</td>
</tr>
<tr>
<td>12</td>
<td>Various</td>
<td>Aox, C22</td>
<td>0.12, 0.83</td>
<td>&lt;0.0001, &lt;0.0001</td>
</tr>
</tbody>
</table>

BCAA indicates branched-chain amino acid; FFA, free fatty acid. Targeted metabolomics analysis was performed on blood samples from patients with and without coronary artery disease, involving an index (initial) cohort of 174 cases and 174 controls and a replication cohort of 140 cases and 140 controls (adapted from Shah et al).\(^{8}\)

*Adjusted for diabetes mellitus, hypertension, smoking, dyslipidemia, family history, and body mass index; replication group was additionally adjusted for age, race, and sex.
were highly correlated with each other and biochemically related as intermediates of choline metabolism. Rigorous, targeted MS methods were used to identify the 3 analytes as choline, betaine, and trimethylamine N-oxide (TMAO). Diet and gut microflora are implicated because lipid-rich foods contain phosphatidylcholine, which is converted to free choline and then metabolized to trimethylamine by gut bacteria. Betaine is also derived from choline via 2 enzymes present in mitochondria of human cells. Host metabolism is implicated because trimethylamine enters the circulation and is further metabolized to TMAO by hepatic flavin monooxygenases. Importantly, these authors also performed a follow-up study in which targeted LC-MS/MS analysis of choline, betaine, and TMAO was performed on samples from 1876 subjects undergoing elective cardiac evaluation. They found a strong association between these metabolites and a range of CVD phenotypes, including peripheral artery disease, CAD, and history of MI, even after adjustment for traditional risk factors. Overall, this work suggests that diet and gut microflora composition can make substantial contributions to the levels of metabolites that are diagnostic for CVD disease risk. Further discussion of the cause-and-effect relationships is provided in the Function section that follows.

With regard to profiling of the consequences of MI, a remarkable study was performed involving LC-MS–based profiling of serial blood samples from 36 individuals undergoing alcohol septal ablation for treatment of hypertrophic obstructive cardiomyopathy. This is considered a human model for planned MI (PMI). Changes occurring in response to PMI were normalized against samples from subjects undergoing elective diagnostic coronary angiography (negative control) or patients subjected to acute coronary angiography owing to spontaneous MI (positive control). From among ~200 metabolites profiled, 4 were found to be altered in the same direction in both PMI compared with baseline and spontaneous MI compared with baseline diagnostic angiography (TMAO and threonine decreased, aconitic acid and hypoxanthine increased). Positive features of this study include the detailed time course over which metabolites were surveyed after PMI (as early as 10 minutes after the intervention and stretching to 24 hours), the ability to use
subjects as their own controls by study of preintervention and postintervention metabolites, the comparison of PMI and spontaneous MI, and the demonstration that some of the altered metabolites are likely derived from the injured heart tissue based on sampling from the coronary sinus in a subset of subjects. Interestingly, studies involving exposure of cultured rat cardiomyocytes to hypoxia revealed that some of the metabolites regulated in PMI altered the apoptotic response to hypoxic challenge, including inosine (diminished apoptosis) and hypoxanthine (enhanced apoptosis). It should be acknowledged that biomarkers for diagnosing myocardial ischemia (ie, troponin) are highly sensitive, and circulating metabolites may represent diverse tissue origins, thus raising uncertainty about the future broad utility of metabolomics for detecting acute myocardial ischemia. Nevertheless, studies performed to date help shed light on the mechanisms of ischemia, and the identified metabolites could still serve as important biomarkers for the stratification of patients at higher risk of recurrent events. Future studies are needed to address these issues.

Heart Failure
Metabolomics has also been applied to studies of heart failure. Our group used targeted MS/MS to measure 63 metabolites in arterial and coronary sinus blood obtained during cardiac surgery, before and after ischemia/reperfusion. This work demonstrated that the preexisting ventricular state (left ventricular dysfunction, CAD, or neither condition) was associated with clear differences in myocardial fuel uptake, both at baseline and after ischemia/reperfusion. In particular, left ventricular dysfunction was associated with global suppression of metabolic fuel intake (glucose, fatty acid, and amino acids) and limited myocardial metabolic reserve and flexibility after ischemia/reperfusion. In another study, gas chromatography/MS was used for targeted analysis of serum samples of 52 patients with systolic heart failure (ejection fraction of <40% and symptoms of failure) and 57 control subjects, resulting in identification of pseudouridine and 2-oxoglutarate (α-ketoglutarate) as 2 potential biomarkers of the failing heart. One limitation of this study was the extensive differences between groups with regard to the use of medications that could have influenced the metabolic profiles. In addition, pseudouridine was highly correlated with serum creatinine in these studies, suggesting a contribution of failing renal function to the profile. Given the strong relationship between renal dysfunction and CVD risk and the strong impact of renal dysfunction on metabolomics profiles, the contribution of renal function to the metabolomics profiles of CVD risk should be monitored carefully. Studies of human heart failure will also benefit tremendously in the future from the profiling of heart tissue samples, both in fulminant heart failure and in response to device-assisted ventricular unloading, and such studies are in progress.

Studies of Exercise Intervention in CVD
Metabolomics can be an excellent tool for the assessment of whole-body and target tissue metabolic responses to an exercise stimulus, both short and long term, and for assessing the mechanistic relationship to the health benefits of exercise training. For the last decade, the Studies of a Targeted Risk Reduction Intervention Through Defined Exercise (STRRIDE) trial has examined the impact of low-, moderate-, and high-intensity exercise interventions on cardiac risk markers, including insulin resistance. This includes a study of the association of 67 peripheral blood metabolites in the fasted state with insulin sensitivity measured with a frequently sampled intravenous glucose tolerance test both before and after exercise training in 72 individuals assigned to 1 of 4 exercise groups: (1) inactive control, (2) moderate-intensity/low-amount aerobic exercise training, (3) vigorous-intensity/low-amount aerobic exercise training, and (4) vigorous-intensity/high-among aerobic exercise. Targeted metabolic profiling revealed an association of BCAAs and related metabolites with insulin resistance in STRRIDE subjects at baseline, whereas a factor containing free fatty acids and another containing glycerol and proline accounted for 59% of the variability in the change scores in insulin sensitivity with training.

Complementing the longitudinal training approach in the STRRIDE study, another group has applied metabolic profiling in patients undergoing functional cardiac stress testing and cardiopulmonary exercise testing. In an early study of this type, LC-MS was applied to plasma from 18 individuals with exercise-induced ischemia compared with 18 individuals without evidence of ischemia on stress testing. A number of metabolites were found to be discordant between the 2 groups, including lactate, byproducts of AMP metabolism, and metabolites of the citric acid cycle. A metabolic ischemia risk score was derived from the 6 most discordantly regulated metabolites, although its broader applicability remains uncertain, given the small size of the study and the differences in sex and race between the 2 study groups. More recently, the same group assayed 200 metabolites before and after exercise testing in 103 individuals undergoing acute cardiopulmonary exercise testing with incrementally ramped upright cycle ergometry, respiratory gas exchange, and continuous hemodynamic monitoring. They identified correlations between plasma markers of glycogenolysis (glucose-6-phosphate), tricarboxylic acid cycle activity (succinic acid, malic acid, and fumaric acid), metabolites that serve as modulators of insulin sensitivity (niacinamide), and fatty acid oxidation (pantothenic acid) with hemodynamic responses to acute exercise and showed that these metabolites recovered appropriately to normal after the exercise bout. Interestingly, exercise-induced increases in circulating glycerol levels were strongly related to fitness level and were attenuated in individuals with cardiac ischemia. They also studied 25 runners before and after the Boston Marathon and observed that the circulating metabolites changed in a fashion consistent with observations in the 103 acutely tested subjects but also reflected the significant substrate use shift that occurs during prolonged exercise (late-stage shifting to complete fatty acid oxidation and consumption of amino acids and lactate for gluconeogenesis). Finally, they reported that a subset of metabolites upregulated at peak short-term exercise (glycerol, niacinamide, glucose-6-phosphate, pantothenate, and succinate) induce expression of the orphan nuclear receptor Nr4a1, which is known...
to regulate glucose use, diet-induced obesity, and lipid metabolism in cultured cells and animal models, thereby identifying a possible molecular mediator of the fitness adaptation.

**Function: Metabolomic Profiling for Identification of Novel Disease Mechanism(s)**

**Retrotranslation of Clinical Findings to Animal and Cellular Studies**

Although there has been much recent progress in the use of metabolomics tools to identify CVD biomarkers, the emergence of each new marker/cluster of metabolites raises questions of mechanistic relevance. In other words, are the metabolites simply a biomarker, or are they related in a causal way to disease pathogenesis? In addition, in the event that they serve as a marker of disease risk, what cellular processes are they reporting on that may relate to CVD development? Here, we provide selected examples of the use of emergent metabolomics-derived biomarkers to gain insights into disease mechanisms, culled from both the diabetes mellitus and CVD research arenas.

Targeted metabolomics has been used to identify a signature of dysregulated metabolism of BCAAs in multiple cohorts of insulin-resistant humans, including obese, insulin-resistant compared with lean, insulin-sensitive subjects, in a cross-sectional study of subjects with metabolic syndrome of varying body mass index, and in Chinese and Asian-Indian subjects in whom body mass index was controlled (average body mass index, 24 kg/m²). In each case, principal components analysis identified a cluster of metabolites comprising BCAAs, aromatic amino acids (Aro AA-Phe, Tyr), Glu, Ala, and C3 and C5 acylcarnitines that were strongly associated with insulin resistance, more so than any lipid-related factor. The analytes in this diagnostic cluster are all connected in various ways to BCAA catabolism. These findings were extended via application of LC-MS–based metabolomics in a nested case-control study of Framingham Heart Study (FHS) participants. During 12 years of follow-up, the strongest associations with incident diabetes mellitus were observed for 5 BCAAs and Aro AA: Ile, Leu, Val, Phe, and Tyr (P<0.001 for all). These findings were replicated in the Malmo Diet and Cancer Cohort.

BCAAs and related metabolites are also related to therapeutic outcomes. Thus, obese subjects undergoing gastric bypass surgery have a much more dramatic decline in circulating BCAAs, C3 and C5 acylcarnitines, Phe, and Tyr than found in response to dietary intervention, despite equal weight loss. This is significant because gastric bypass causes a greater improvement in glucose homeostasis than dietary intervention. In addition, in studies of 500 overweight/obese subjects from the Weight Loss Maintenance (WLM) trial, the BCAA-related principal component factor score at baseline was a strong predictor of improvement in insulin sensitivity with intervention (P<0.001), whereas lipid-related factors and most clinical variables, including the amount of weight lost, had little or no predictive association. Taken together, a diverse array of recent clinical studies has revealed that BCAAs and related metabolites are associated with insulin resistance, are predictive of diabetes mellitus development and intervention outcomes, and are highly and uniquely responsive to therapeutic interventions.

The strong association of BCAAs and related metabolites with cardiometabolic diseases described above suggests but does not prove a cause-effect relationship between elevated BCAAs and disease development. To investigate this further, we fed normal rats a high-fat diet supplemented with BCAAs. These animals developed insulin resistance despite eating less food and gaining less weight than rats fed a high-fat diet alone. In contrast, rats pair-fed a high-fat diet to match the caloric intake of the high-fat/BCAA group did not become insulin resistant. Strikingly, targeted metabolomics analysis of muscle extracts from the different rat groups revealed that animals fed a BCAA-supplemented high-fat diet accumulate incompletely processed mitochondrial metabolites in muscle to the same extent observed in heavier high-fat–fed animals and in genetic models of obesity/diabetes mellitus such as Zucker fatty rats. Accumulation of acylcarnitines, especially medium-chain and long-chain species in skeletal muscle, has been described as a signature of incomplete fatty acid oxidation in animal models of obesity. The direct, causal relationship between these mitochondrial metabolites and insulin resistance has been demonstrated by studies in transgenic mice deficient for the expression of malonyl CoA carboxylase or carnitine acyltransferase. Knockout of malonyl CoA carboxylase limits the entry of fatty acid into the mitochondria through the carnitine palmitoyltransferase system and relieves diet-induced insulin resistance. Knockout of carnitine acyltransferase impairs the removal of excess acyl CoA from mitochondria via conversion to their cognate acylcarnitine species and accordingly exacerbates the insulin-resistant state.

The possible mechanistic link underlying the association between BCAAs and CAD is currently less clear. This association remains intact after correction for diabetes mellitus in the CATHGEN cohort, suggesting that it may not be driven entirely by insulin resistance. The mechanistic relationship between arginine metabolites and CVD is also incompletely resolved. Tang et al have reported a decline in GABR associated with CAD and incident major adverse cardiovascular events. They and others have also reported associations between variations in a host of methylated arginine species measured by targeted LC-MS/MS methods, including symmetrical dimethylarginine, monomethylarginine, and asymmetrical dimethylarginine, with either CAD or major adverse cardiovascular events. Arginine is a key physiological precursor for nitric oxide production, an important modulator of vascular tone, and asymmetrical dimethylarginine and monomethylarginine but not symmetrical dimethylarginine are potent inhibitors of endogenous nitric oxide synthase, suggesting that low GABR and increased methylated arginine species could work through a unified mechanism of inhibition of nitric oxide production. Interestingly, 2 separate studies on large and independent cohorts report that incident major adverse cardiovascular events are associated with elevations in both symmetrical dimethylarginine and asymmetrical dimethylarginine, whereas 1 study reports that the risk for significantly obstructive CAD is associated with high symmetrical dimethylarginine levels but not with elevations in monomethylarginine and asymmetrical dimethylarginine. In addition, the methyltransferases involved in arginine methyl-
ation and other enzymes involved in the metabolism of methylated arginine species are influenced by inflammation and oxidative stress, and the findings in aggregate suggest that some of the associations reported between arginine metabolites and CVD may be independent of nitric oxide production. Further work is required in this fascinating area.

More recently, a study of targeted metabolites in the FHS and the Malmö Diet and Cancer Study corroborated the role of BCAAs in insulin resistance but also identified glutamine, glutamate, and the glutamine-to-glutamate ratio as important biomarkers of multiple metabolic and insulin-resistance phenotypes. The authors further demonstrated that oral supplementation of glutamine to mice resulted in increased glucose tolerance and lower blood pressure compared with mice fed standard chow or standard chow supplemented with glutamate, suggesting that diminished glutamine may contribute to metabolic disease risk. These findings expand on an emerging appreciation of a role for altered amino acid metabolism in cardiometabolic disease risk.

Important mechanistic studies have also been performed in the context of the recently reported triad of choline metabolites that associate with incident CVD events, choline, betaine, and TMAO. An obligate role for gut microflora in the formation of TMAO was demonstrated by incorporation of labeled choline into TMAO in untreated mice but not in mice treated with broad-spectrum antibiotics. Direct cause-and-effect relationships were demonstrated by studies in which atherosclerosis-prone apoE−/− mice were placed on normal chow diets with or without choline or TMAO supplementation, revealing a clear increase in plaque size in the supplemented animals that was proportional to the increase in circulating TMAO levels. Finally, the ability of dietary choline to enhance lesion formation in apoE−/− mice is abrogated in mice treated with broad-spectrum antibiotics, providing evidence of a direct link between diet, the gut microbiome, and biological events associated with CVD.

Integration of Genetics and Metabolomics for the Identification of Novel Disease Pathways

Another potential avenue for translating metabolomics-derived biomarkers to disease mechanisms is the integration of metabolomics with other “omics” methods. Human genome-wide association studies have mapped loci associated with polygenic disorders like CVD and diabetes mellitus, but they account for only a small fraction of these diseases and have made limited contributions to knowledge-based therapeutic interventions. This results in part because both conditions are actually a family of diseases in which genetic variability, environmental factors, and resultant perturbations in metabolic control within multiple tissues and organs conspire to disrupt fuel homeostasis and tissue functions. Perhaps when viewed in this context, it is less surprising that individual loci mostly fail to explain these diseases because convergence of multiple intermediate phenotypes may be required to create them. Now that the metabolic signatures that predict incident cardiometabolic disease and intervention outcomes are beginning to emerge, there is fresh opportunity to define underlying genetic architecture. Through an understanding of genetic influences on intermediate traits, including variations in metabolite levels, a better understanding of disease pathogenesis may emerge. Here, we highlight selected examples of the integration of metabolomics with genomics and transcriptomics to gain insight into cardiometabolic phenotypes in mouse models and in human studies.

We have applied an integrated genetics/transcriptomics/metabolomics approach to the study of F2 mice generated from a cross of diabetes mellitus–susceptible BTBR-ob/ob mice with diabetes mellitus–resistant C57BL/6-ob/ob mice. Mice were subjected to whole-genome single-nucleotide polymorphism analysis, transcriptomic profiling, and targeted metabolomics analyses. Importantly, variations in many of the measured metabolites were associated with specific regions of the genome, suggesting that metabolites can be used to define quantitative trait loci (Figure 2). When these data are analyzed by advanced computational methods that integrate data for comapping transcripts, gene–transcript–metabolite and gene–metabolite–transcript networks can be identified that include arrows of causality. Application of these methods to the F2 mice from the BTBR-ob/ob×C57BL/6-ob/ob cross revealed a novel network by which glutamine controls the expression of the key gluconeogenic enzyme PEPCK. The existence of this network was confirmed in cellular studies, and its potential relevance to disease pathogenesis is illustrated by its perturbation in the diabetes mellitus–susceptible BRBR-ob/ob strain.

Further insights may be gained from new resources that are coming available for mouse genetics research. For example, the Collaborative Cross involves the interbreeding of 8 mouse strains that together represent ~90% of the genetic diversity of all inbred mouse strains. Interbreeding is followed by inbreeding to generate ~300 recombinant strains, achieving an unprecedented level of genetic diversity. The Collaborative Cross strains attempt to simulate some of the genetic diversity present in outbred humans and have advantages relative to F2 intercrosses that include expanded genetic space, improved mapping resolution, and the ability to perform studies with unlimited biological replication in each of the 300 recombinant inbred strains. Detailed metabolomics, transcriptomic, and proteomic studies are planned in these strains. When complete, this can allow investigators to search for extremes in metabolites implicated in metabolomics surveys of human disease states among the strains and to then use those strains to define genomic and transcriptomic variation that drives variation in the metabolites.

It has also become clear that metabolites are heritable traits in human subjects. Thus, MS/MS and gas chromatography/mass spectrometry–based profiling of fatty acids, acylcarnitines, and amino acids was performed on plasma samples from proband and offspring subjects in 8 multiplex families with familial early-onset CVD. Even after adjustment for variables such as diabetes mellitus, hypertension, dyslipidemia, body mass index, age, and sex, multiple individual metabolites and metabolite clusters identified by principal components analysis were found to be highly heritable within families, including groups of amino acids (arginine, glutamate, alanine, ornithine, valine, leucine/isoleucine), free fatty acids (arachidonic, linoleic), and acylcarnitines. Interestingly, families in this study showed 2 distinct metabolite...
profiles (including 1 that contained the arginine-ornithine group referred to previously) that tracked with their clinical characteristics, suggesting different genetic backgrounds and consequent variation in control of key metabolic pathways that converge on CVD.

The Challenge of Identifying Genetic Variants That Control Metabolism and Disease in Humans

The hierarchy of molecular control of biological systems passes from gene through gene modification (epigenetics) to gene expression (transcriptomics) to protein expression (proteomics) to metabolites/metabolic pathways (metabolomics) to clinical phenotype or outcome (Figure 3, top). In this scheme, metabolism sits closest in proximity to clinical phenotype and is therefore a strong candidate for identifying causative molecular mechanisms for disease. However, in identifying genes responsible for the metabolic intermediates that affect outcome, it is clear that the biological complexity of the system increases as one passes from gene to outcome, thus making the task of identifying the specific pathways involved in genetic cause or mediation of disease quite daunting—conceptually, experimentally, and statistically.

The experimental challenge is illustrated by the fact that the 9p21 genetic locus undoubtedly contains the strongest and most consistent genetic marker of CVD risk identified so far, yet 5 years after its discovery, the biological mechanism or pathway by which this genetic variant mediates disease is still a mystery.

The statistical challenge is illustrated by our own recent work and similar studies being performed by others (Figure 3, bottom). In our CATHGEN cohort, we have genome-wide association study data containing $10^5$ single-nucleotide polymorphisms on 3500 individuals, whole-genome transcript data with $2 \times 10^4$ gene tags on 1500 of the same individuals, data for $>70$ metabolites on 3500 individuals, and at least 10 outcomes or phenotypes on close to 10 000 individuals. Our work to date reveals strong associations of specific genetic loci with the cluster of small- to medium-chain dicarboxylated acylcarnitines found in the same population to be prognostic for incident cardiovascular events. The potential functional relationships between these genetic loci and variation in the dicarboxylated acylcarnitines are currently being investigated. The statistical complexity and the risk for false discovery (the multiple-comparisons problem) are daunting, exceeding $10^{13}$ comparisons in each individual in the study population. Thus, in moving forward, we propose to adopt both the classic and retrograde approaches summarized in Figure 3. In the classic genetic variant mapped to outcome...
approach, used in the majority of genome-wide association studies published to date, the risk is that although one might identify a genetic variant conferring some modest degree of risk, the underlying biological mechanisms and therefore biological targets for pharmacological intervention might remain elusive. In the retrograde approach, one runs the risk of missing genetic variants associated with the clinical phenotype of interest but retains a higher probability of identifying relevant biological pathways and pharmacological targets.

Other studies have begun to survey the relationships between genetic and metabolic variability in the general population. Targeted51 and nontargeted52 MS-based metabolic profiling has been applied to the Cooperative Health Research in the Region of Augsburg (KORA) and TwinsUK cohorts, in concert with whole-genome genotyping. With the use of targeted metabolite profiling methods, associations were defined between frequent single-nucleotide polymorphisms and variability in plasma metabolite levels in 1809 KORA subjects and 422 TwinsUK participants.53 Remarkably, in 8 of 9 cases, variation in specific metabolite levels mapped to loci in or near genes encoding metabolic enzymes or carriers (FADS1, LIPC, SCAD, MCAD) that participate in the metabolism of the particular analyte. Similar findings occurred in an expanded study using nontargeted metabolomics, which also provided clues about the function of 2 orphan genes.52 For example, changes in carnitine levels were associated with the single-nucleotide polymorphism rs7094971 in the SLC16A9 gene, also known as MCT9, which has homology to monocarboxylic acid transporters. Expression of this gene in Xenopus oocytes demonstrated its ability to function as a carnitine transporter.

Conclusions

One can be encouraged by a few examples of elegant studies that have successfully navigated from genomic variation to mechanism. In 1 such study, genome-wide association study defined a locus on chromosome 1p13 that is strongly linked to MI and low-density lipoprotein cholesterol levels.53 The mechanism. In 1 such study, genome-wide association study

defined a locus on chromosome 1p13 that is strongly linked to MI and low-density lipoprotein cholesterol levels.53 The definition of this gene in

Xenopus oocytes demonstrated its ability to function as a carnitine transporter.

Conclusions

One can be encouraged by a few examples of elegant studies that have successfully navigated from genomic variation to mechanism. In 1 such study, genome-wide association study defined a locus on chromosome 1p13 that is strongly linked to MI and low-density lipoprotein cholesterol levels. The definition of this gene in

Xenopus oocytes demonstrated its ability to function as a carnitine transporter.

Acknowledgments

We wish to thank our colleagues in the Sarah W. Stedman Nutrition and Metabolism Center and Center for Human Genetics at Duke University for their strong contributions to work cited in this article from our laboratories.

Sources of Funding

Work from our groups is supported by National Institutes of Health grants DK58398 (to Dr Newgard), HL095987 (Dr Shah), HL101621 (Dr Kraus), and AG02871 (Drs Newgard and Kraus).

Disclosures

Dr Newgard has a sponsored research agreement with Pfizer that supports a portion of the work in his laboratory related to the topic of this article, and he is also a consultant for the Pfizer CVMED group. The other authors report no conflicts.

References


**KEY WORDS:** biological markers. cardiovascular diseases. metabolism
Metabolomic Profiling for the Identification of Novel Biomarkers and Mechanisms Related to Common Cardiovascular Diseases: Form and Function
Svati H. Shah, William E. Kraus and Christopher B. Newgard

Circulation. 2012;126:1110-1120
doi: 10.1161/CIRCULATIONAHA.111.060368
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/126/9/1110

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/