Metabolomics

Integrated Metabolomics and Genomics
Systems Approaches to Biomarkers and Mechanisms of Cardiovascular Disease

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

Abstract—The genetic architecture underlying the heritability of cardiovascular disease is incompletely understood. Metabolomics is emerging as an important technology platform for understanding of mechanisms underlying common chronic diseases, such as diabetes mellitus, obesity, and cardiovascular disease (CVD). Metabolomics measures chemistry, and chemistry represents an integrated readout of upstream genetic, transcriptomic, and proteomic variation. It is also becoming apparent that metabolomics can be integrated with these other omic technologies to identify novel biological pathways and disease mechanisms. At the simplest level, metabolite biomarkers can be combined with genetics, other biomarkers, and clinical variables to provide incremental gain in diagnostic or risk prediction models. At another level, metabolites can serve as intermediate phenotypes for genetic studies. Finally, at a more biologically and analytically complex level, metabolomics can be integrated with multiple omic platforms in a systems biology approach. A major goal of systems biology is to assemble a global map of the functional relationships and interactions between physical entities in the cell (genes, proteins, metabolites, etc), as well as a roadmap of the interaction of different tissues and organ systems for affecting physiological homeostasis.

The promise of metabolomics and integrated omics analyses for chronic disease research is just beginning to be realized. Initial successes have fueled a surge in interest, but investigators need to be aware of study design and analytic and interpretative challenges. Here, we review the status of this emergent area, as illustrated by provocative proof-of-principle examples.

Genetics and Metabolomics of CVD
It is well established that CVD is heritable. Before the Human Genome Project, hundreds of candidate gene studies were published, but few variants were consistently associated with disease and fewer still with detailed functional data. After publication of the initial draft of the human genome, many genome-wide association studies (GWAS) for CVD were conducted, consistently identifying the same intergenic locus on chromosome 9p21. However, the effect size of this single-nucleotide polymorphism (SNP) is modest and has unclear functional effects. Larger GWAS through collaborative consortia have uncovered additional variants with weaker effects, but their effect on functions related to CVD and their overall effect on development of disease remain to be determined.

Although static genetic biomarkers are key components of heritability, it makes biological sense that with a chronic systemic disease, such as CVD, with early manifestations that can begin to develop in childhood, molecular signals move proximal to the disease process might serve as stronger biomarkers. Certainly, proteins reflecting the diverse biological processes of CVD are commonly used in practice (ie, high-sensitivity C-reactive protein, pro–brain natriuretic peptide, and troponin). The strong association of CVD with broad metabolic perturbations, such as hyperglycemia and hyperlipidemia, suggests that a more extensive survey of metabolic variation through application of metabolomics could identify metabolites with diagnostic and mechanistic relevance.

Indeed, metabolomics has already been used with some success to identify cardiometabolic disease biomarkers. An early study uncovered differences in nuclear magnetic resonance–derived metabolite peaks in a small group of patients with coronary artery disease compared with controls. Enthusiasm for the approach was tempered when a subsequent study found that the original association was likely confounded by statin medication use and sex of subjects, emphasizing the
importance of evaluating medications, comorbidities, diet, and other confounders. Fortunately, several subsequent studies have emerged that include validation cohorts identifying biomarkers strongly associated with CVD and contributing conditions including insulin resistance, diabetes mellitus, and inflammation.

For example, our group has demonstrated the association of a cluster of branched chain amino acids (BCAA) and related metabolites with insulin resistance.\(^4\) Similar association has been observed in the Framingham Heart Study,\(^9\) and a subsequent study made the important observation that baseline BCAA levels predicted future development of type 2 diabetes mellitus.\(^10\) Baseline BCAA levels also predict improvement in insulin resistance with weight loss,\(^11\) and BCAA levels are lower in metabolically healthy versus metabolically unwell overweight/obese individuals.\(^12\) Interestingly, BCAA levels decrease more dramatically in response to bariatric surgery compared with a dietary intervention, even when matched for the amount of weight loss, consistent with the greater effect of the surgical intervention on glucose homeostasis.\(^13\) Finally, levels of BCAA and related metabolites are associated with coronary artery disease–independent of type 2 diabetes mellitus.\(^14,15\) The potential mechanistic implication of these findings is illustrated by feeding studies in rats, in which supplementation of high-fat diets with BCAA promoted insulin resistance despite lesser weight gain compared with rats fed unsupplemented high-fat diet.\(^7\)

More recently, metabonomic studies have identified additional biomarkers of obesity and diabetes mellitus. One study demonstrated a strong association of 2-aminoacidic acid levels and incident diabetes mellitus that remained significant when corrected for BCAA levels.\(^16\) 2-Aminoacidic acid is a lysine metabolite. The reasons for its association with new-onset diabetes mellitus are not known, but interestingly, this metabolite was shown to increase insulin secretion in rodent and human islet cells. The authors suggest that elevated levels of 2-aminoacidic acid in the prediabetic state could cause hyperinsulinism, leading to insulin resistance. Further studies are required to test this mechanism. Another recent study identified β-aminoisobutyric acid (BAIBA) as a metabolite secreted from myocytes with forced expression of the transcriptional coactivator PGC-1α.\(^17\) BAIBA levels were found to increase in response to exercise, and BAIBA infusion was found to induce the appearance of brown fat. These studies suggest that BAIBA can act as a small molecule myokine that increases energy expenditure and participates in exercise-induced protection from metabolic diseases.\(^17\) Interestingly, one potential source of BAIBA is valine metabolism, but it can also be generated by degradation of thymine. BAIBA levels in obese and insulin resistant states have not been investigated, and such studies will be required to fully understand its relationship with metabolic function.

In CVD itself, several biomarkers have emerged from metabonomics efforts, some with mechanistic implications. A previous study capitalized on a human model of planned myocardial infarction (MI): plasma was collected at several time points before and after alcohol septal ablation performed for management of hypertrophic cardiomyopathy, a surrogate model for spontaneous MI.\(^14\) Several metabolites changed in blood, including a signature consisting of aconitic acid, hypoxanthine, trimethylamine N-oxide (TMAO), and threonine, with metabolites increasing as early as 10 minutes after initiation of MI and earlier than traditional markers (ie, troponin). Although the pathophysiology of the planned MI model differs from that of spontaneous MI, levels of the same set of metabolites were also different in patients with spontaneous MI when compared with patients undergoing angiography who were not having MI. Thus, these metabolites may serve as earlier and more sensitive markers of MI.\(^18\)

Metabonomic profiling has also identified markers predicting incident CVD events, a phenotype where clinical prediction models are incomplete. Using targeted mass spectrometry–based metabolic profiling in baseline blood samples from 314 patients with coronary artery disease, we found that a cluster of short-chain dicarboxylycarnitines (SCDA) discriminated the 74 individuals who went on to suffer death or MI, with over a 2-fold increased risk of CVD events for every 1-U increase in the levels of the SCDA metabolic signature.\(^1\) These results were validated in a case–control cohort (n=129),\(^7\) in a sequential cohort of 203 individuals,\(^15\) and in patients undergoing coronary artery bypass grafting (Figure 1A and 1B).\(^19\) Although many studies show independent association of biomarkers (ie, after adjusting for relevant clinical variables), we demonstrated that SCDA metabolites added incremental predictive capabilities on top of clinical models. Specifically, 27% of individuals who would have classified as intermediate risk from a robust model, including 23 clinical variables, were correctly reclassified into low or high risk by inclusion of metabolites (Figure 1C).\(^20\) Such analyses of incremental discrimination/prediction are important in assessing the significance of omic biomarkers and will facilitate the translation of these biomarkers into clinical practice. Interestingly, little is known about the source of SCDA metabolites and how they link to CVD at a mechanistic level, but studies are underway in both areas.

Interest has emerged on the role of the gut microbiome in the development of chronic cardiometabolic diseases. For example, using non-targeted metabonomics, higher circulating levels of TMAO, choline, and betaine were observed in individuals with CVD events.\(^20\) The dietary lipid phosphatidylcholine is the primary dietary source of choline, and catabolism of betaine and choline by intestinal microbes leads to TMAO production. Dietary supplementation with these metabolites in mice promoted atherosclerosis, and ablation of the microbiota with antibiotics prevented the effect of dietary choline in enhancing atherosclerosis. Consistent with these results, antibiotic treatment caused TMAO levels in humans to decrease in response to an oral phosphatidylcholine challenge.\(^21\) Moreover, TMAO levels were higher in humans after l-carnitine ingestion (a trimethylamine abundant in red meat). Carnitine can also serve as a TMAO precursor via microbial metabolism.\(^22\) Interestingly, different proportions of bacterial taxa were present in the feces of vegetarians when compared with omnivores, and several of these taxa were associated with plasma TMAO concentrations. This suggests that diet could modulate intestinal microbiota composition and concomitantly, the ability to synthesize trimethylamine and TMAO from dietary l-carnitine, thus providing a mechanistic link between diet, gut microbiome, and atherosclerosis.\(^22\)
Several studies have also linked the microbiome to metabolic diseases. A recent study of twins discordant for obesity involving transplantation of fecal material into germ-free mice revealed that mice harboring the transplanted microbiota from the obese twins had higher circulating BCAA levels and higher muscle acylcarnitine levels, an indicator of muscle insulin resistance. Remarkably, transcriptomic analysis of the microbiota revealed the induction of the entire pathway of BCAA biosynthesis in the obese microbes (this pathway is present in bacteria but not in mammals), and metabolomic analysis of the obese microbiota demonstrated higher BCAA production. Although the relationship between microbial metabolite synthesis/production and changes in host metabolism requires further study, these new findings highlight the importance of the interactions between microbial and host genetics and metabolism.

A more integrated omics approach might also contribute to better understanding of the genetic architecture underlying CVD heritability. Although large CVD genetics meta-analyses are able to identify statistically significant genetic variants, they have primarily relied on lumping of CVD phenotypes, often resulting in a low effect size for the variants. Such studies are certainly vital for elucidating the polygenic nature of CVD, but there is also a need for parallel evaluations that split CVD into more discrete intermediate phenotypes. Metabolomic profiling holds great potential for enabling such phenotyping by reporting on subclinical cellular dysfunction that may not be embodied in more traditional measures. Metabolic profiling could also identify pathophysiologic changes at earlier time points in the CVD temporal continuum.

**Omics Technologies and the Concept of Systems Biology**

Advances in our understanding of disease pathophysiology, normal human physiology, and environmental influences on both have been driven recently by application of high-throughput molecular technologies (Figure 2). It is now possible to measure millions of genetic variants in thousands of samples in a short period of time (GWAS). Next generation sequencing allows analysis of the entire exome or genome, with many centers having the ability to perform studies on hundreds to thousands of samples, in contrast to the handful of individuals sequenced for the initial draft of the Human Genome Project. Sequencing technologies have also advanced the scientific community’s ability to measure a larger and more comprehensive number of RNA transcripts and microRNAs. Technologies...
are likewise evolving for high-throughput epigenetic profiling. These tools are only recently being applied to cardiometabolic diseases, but they show great promise for shedding light on gene–environment interactions. Finally, in parallel, steady advances in nuclear magnetic resonance and mass spectrometry methods have enabled more accurate and comprehensive profiling of metabolites and proteins in tissues and blood.

The concept of systems biology moves beyond studying individual molecules or single reactions, integrating orthogonal data from diverse biological datasets, including the genes, epigenetic modifications, RNAs, proteins, metabolites, environmental inputs, clinical variables, and other factors, providing an analytic snapshot into normal and dysregulated biological function. The molecular phenotypes can be tested for relationships with each other and with clinical traits/diseases. These concepts are particularly relevant to cardiometabolic diseases with their dynamic temporal nature and multiple, varying environmental inputs. Such a systems biology approach may help the field to overcome the somewhat disappointing lack of clarity about the genetic architecture of CVD derived from studies of static DNA variation in isolation. It seems reasonable to postulate that clear understanding of the genetic underpinnings of CVD pathophysiology will benefit from a more holistic analysis of gene expression, proteomic, and metabolic consequences of altered gene expression, and the role of incremental clinical/environmental inputs.

In addition, although individual genetic loci identified from GWAS of disease end points can be tested mechanistically using traditional molecular biology experiments, this approach is difficult given the modest effects of the identified variants and the uncertainty surrounding the specific gene or gene modifier that is driving the specific SNP association. Systems biology approaches can be useful in this regard, enabling the analysis of molecular interactions in the context of multiple genetic polymorphisms influencing traits and disease. In the hopes of translating biomarkers into clinical practice, it is important to assess not just independent but also incremental association. For example, SCDA metabolites have been shown to add incremental predictive capability on top of a robust clinical model inclusive of 23 clinical variables with a net classification index of 8.8%. Integration of genetics and metabolomics can also be useful in the context of a focused pathway of interest. The

Applications of Integrated Metabolomic Genetic Analysis

Metabolomics and genetics can be integrated for different purposes including (1) improving clinical risk models, (2) evaluating metabolites associated with a single genotype to generate hypotheses for mechanisms underlying a specific SNP/gene, (3) using metabolites as phenotypes for genetic studies (ie, GWAS), and (4) more unbiased, hypothesis-generating studies integrating metabolomics with genetics, transcriptomics, proteomics, and other omics to identify the pathways of disease. Although this latter systems biology approach with large orthogonal datasets may prove to be the most powerful for identification of novel mechanisms of cardiometabolic disease, integration of metabolomics with genomics at a simpler level has already led to useful advances as now reviewed.

One powerful feature of a multiomics approach is that each molecular platform provides measures of different biological inputs to disease, meaning that their contribution to disease models can be orthogonal, thereby providing incremental discriminative/predictive capability. An early example of this approach demonstrated that inclusion of several protein-based biomarkers incrementally improved the risk prediction of acute coronary syndrome. Another study showed the potential power of an integrated genetic risk score inclusive of SNPs for predicting CVD events, with the genotype score modestly improving risk reclassification incrementally to a clinical prediction model. In the newer metabolomics arena, investigators have sought to establish independent association (ie, adjusting statistically for potential confounders) but not always incremental association (ie, of metabolites with disease in models that account first for clinical factors). In the hopes of translating biomarkers into clinical practice, it is important to assess not just independent but also incremental association. For example, SCDA metabolites have been shown to add incremental predictive capability on top of a robust clinical model inclusive of 23 clinical variables with a net classification index of 8.8%.

Integration of genetics and metabolomics can also be useful in the context of a focused pathway of interest.
initial studies identifying TMAO linked host metabolism with microbiome metabolism. The same group has now demonstrated that 2 flavin monooxygenase family members oxidize trimethylamine to TMAO. They further show that the FMO3 gene contributes to variations in TMAO levels in mice, that mice that express variants in this gene have increased susceptibility to atherosclerosis. Metabolomics can also help inform functional annotation for single-candidate SNPs/genes. TCF7L2 SNPs have been shown to be associated with type 2 diabetes mellitus, apparently because of deficiencies in insulin secretion, although the molecular mechanism of β-cell dysfunction remains unknown. Metabolomic profiling has revealed alterations in phospholipid metabolism in response to glucose tolerance testing in individuals with the risk TCF7L2 genotype. The authors conclude that these results may reflect a genotype-mediated link to early metabolic abnormalities that occur before the development of impaired glucose tolerance.

A similar approach can be used in model organisms. Metabolomic profiling in mice with gene deletions resulting in inactivation of xanthine oxidoreductase identified, in addition to the expected derangements in purine metabolism, dysregulation of several other pathways, including pyrimidine, nicotinamide, tryptophan, and phospholipid metabolism, demonstrating the power of metabolomics for systematic assessment of direct and indirect consequences of gene mutations. Proteomics and metabolomics were combined in mice with transgenic manipulation of protein kinase C-ε, providing evidence for a role of protein kinase C-ε in modulating cardiac glucose metabolism. In another study, metabolomic profiling of hearts from vascular endothelial growth factor-B transgenic mice (which exhibit cardiac hypertrophy without cardiomyopathy) revealed apparent mitochondrial lipotoxicity, suggesting that vascular endothelial growth factor-B regulates lipid metabolism, a heretofore unrecognized function for this angiogenic growth factor.

Metabolites can also be used as phenotypes (metabolic quantitative trait loci) for genetic evaluations by serving as intermediate early reporters on the temporal continuum of CVD development. Furthermore, metabolites are more closely related to genes of interest, serving as intermediates between genes and clinical end points, and thus, mapping metabolites has potential for identification of genetic variants with stronger effect sizes than seen with mapping of CVD per se. Moreover, the pathway in which the metabolite plays a role may provide insight into the underlying biological mechanism responsible for the development of the associated disease. Concordantly, metabolite levels have been shown to be heritable.

A previous study integrated metabolomics with GWAS in a small cohort of 284 participants from the Cooperative Health Research in the Region Augsburg (KORA) study. Using individual metabolites and ratios of metabolite concentrations (as proxies for enzymatic activity), they found that SNPs explained ≤28% of the observed variance in metabolite levels. Four of the most significant SNPs were within genes encoding enzymes in the pathway of the corresponding metabolite. In a GWAS of 6 plasma polyunsaturated fatty acids in 1075 participants in the Invecchiare in Chianti (InCHIANTI) study of aging, genome-wide associations were found in a region of chromosome 11 encoding 3 fatty acid desaturases (FADS1, FADS2, and FADS3), providing further evidence that genetic variation contributes to variation in plasma fatty acid levels. Another study integrating GWAS with 163 metabolites in a larger sample of 1809 KORA subjects found 8 genetic loci meeting genome-wide significance, with most of the loci again located in or near enzyme or solute-carrier coding genes involved in processing of the associated metabolite (Figure 3).

Figure 3. Manhattan plot of genome-wide association studies (GWAS) of metabolites from the KORA study. Displayed is the strength of association with metabolite concentrations (top; \( P<10^{-1} \) in red) and concentration ratios (bottom; \( P<10^{-3} \) in red). Reprinted from Illig et al with permission of the publisher. Copyright © 2009, Nature Publishing Group.
error of metabolism (medium-chain acyl-CoA dehydrogenase deficiency). This suggests that common SNPs in genes that cause rare Mendelian diseases may lead to a less severe and potentially subclinical phenotype that could only be discovered by mapping the metabolite itself. Several subsequent studies combining GWAS with metabolomics have been published (Table). It is important to note that although some of the identified SNPs have been associated with disease phenotypes in different cohorts, these studies have not shown that metabolite-associated SNPs also associate with disease in the same cohort. In fact, many of the aforementioned studies were not performed in disease-bearing cohorts, reducing the power for triangulating metabolic, genetic, and disease associations.

**Integrated Metabolomics Genetics: Systems Biology Examples**

Integrated approaches have been applied in model organism studies, both for focused hypothesis testing and for hypothesis generation. As an example of the latter, a study integrating metabolomics, transcriptomics, and genetics in liver samples from an F2 intercross between diabetes mellitus–resistant and diabetes mellitus–susceptible mouse strain connected variations in metabolites and transcripts to regions of the genome and constructed associative networks controlling liver metabolic processes (Figure 4A).44 On the basis of advanced computational analysis of this multiomic data set, a causal network linking variations in glutamate to regulation of the key gluconeogenic enzyme phosphoenolpyruvate carboxykinase was identified, and importantly, experimentally validated by showing that glutamate-induced expression of phosphoenolpyruvate carboxykinase and other genes in the network (Figure 4B).45 Studies of this nature serve as a proof-of-principle for use of systems biology in identification of plausible and testable metabolic control networks.

In humans, using network analysis from multiple omics platforms in a large population–based cohort from Finland, the authors elucidated functional effects of a lipid signaling module composed of a set of highly correlated genes with a prominent role in regulating the levels of 80 metabolites and providing new links between inflammation, metabolism, and adiposity.39 Such systems biology approaches may be useful even in small sample sizes. As an extreme example, an integrative Personal Omics Profile was performed on a single individual that integrated omics over multiple time points for an 14-month period.40 Whole-genome sequencing identified a genetic variant predisposing the individual to increased risk of type 2 diabetes mellitus, and monitoring of glucose and HbA1c levels subsequently revealed the onset of the disease despite a normal body mass index. Integrated molecular profiles changed concordant with respiratory infections, showing dynamic molecular changes in response to disease. Such studies not only aid in identification of novel mechanisms of disease pathogenesis but also perhaps project future approaches to personalized medicine.

**Analytic and Bioinformatic Considerations**

Analysis of millions of data points per single-study subject poses unique challenges in omic sciences. In some instances, traditional techniques can be used. For example, analysis of individual analytes in GWAS or metabolomic studies is facilitated by adjustment for multiple comparisons. However, approaches that explicitly incorporate the collinearity and multidimensionality of the complex data structure, take biological pathway information into account, and analyze patterns or networks within the data may have greater statistical power and enable better mechanistic hypothesis generation. Such approaches can be unsupervised (ie, variation within the molecular data only) or supervised (ie, variations in the molecular data and the disease state). Analytic approaches used include factor analysis, hierarchical clustering algorithms, Gaussian graphical modeling, and pathway- and network-based analyses that can integrate data from disparate omic platforms and identify molecular interactions and pathways. Programs have been developed to aid in data visualization, statistical analysis, and bioinformatic annotation and interpretation. These programs often incorporate information from publically available databases and use pathway and network statistical analysis techniques.

Concurrently, the metabolomics community is making advances in metabolite annotation, nomenclature, and cataloging, including the Human Metabolome Project and the LIPID Metabolites and Pathways Strategy. One issue that remains to be overcome involves methods of quantification of metabolites. Some laboratories, including ours, emphasize targeted approaches that make use of extensive libraries of stable isotope-labeled standards.34 For example, when measuring amino acids, known quantities of multiple stable isotope-labeled amino acids are added to the sample, and the concentration of the native analyte is calculated by reference to its cognate internal standard. In contrast, nontargeted methods that report analytes as relative peak areas unreferenced to specific internal standards are often used, even in epidemiological studies. It is important to note that correlation or association does not define directionality of a molecular relationship. Statistical methods for assessing cause-and-effect, such as Mendelian randomization, are becoming increasingly used but does not replace the ultimate need for testing of predictions emanating from statistical analyses through biological experiments.

**Study Design, Challenges, and Future Directions**

Although the studies we have detailed highlight the great promise of combining high-throughput molecular data, many considerations need to be addressed to enable dissection of the signal-to-noise in such data, avoid type 1 and type 2 errors, and ensure accurate data interpretation. Investigators must carefully decide on a study design from among various options, including population-based cohorts, disease case–control studies, or evaluation of extremes of a clinical or molecular trait. Statistical power should be considered, particularly in the context of large numbers of molecular biomarkers measured in smaller sample sizes. The depth of molecular profiling needs to be weighed against the quantitative precision (or lack thereof) of the measurements. Analytic strategies for multidimensional data reduction, pathway/network analysis,
**Table. Genome-Wide Association Studies of Metabolites**

<table>
<thead>
<tr>
<th>Cohort population–based cohort&lt;sup&gt;46&lt;/sup&gt;</th>
<th>Platform</th>
<th>n</th>
<th>Results</th>
<th>Significance</th>
<th>Adjustment for (1) multiple comparisons and (2) clinical variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>KORA population–based cohort&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Targeted tandem MS, 163 metabolites</td>
<td>2231</td>
<td>9 replicating loci (FADS1, ELOVL2, ACADS, ACADL, ACADL, SPATL3, ETFDH, SLC16A9, and PLEKHH1)</td>
<td>$P=3.10^{-29}$ to $6.5.10^{-129}$</td>
<td>(1) Bonferroni for SNPs and metabolic combinations ($P&lt;2.44\times10^{-11}$) and (2) none</td>
</tr>
<tr>
<td>InCHIANTI study on aging&lt;sup&gt;47&lt;/sup&gt;</td>
<td>GC, 6 unsaturated fatty acids</td>
<td>1075</td>
<td>FADS1-3 and ELOVL2</td>
<td>$P=1.10^{-6}$ to $6.0.10^{-60}$</td>
<td>(1) Bonferroni at level of SNPs ($P&lt;1.0\times10^{-4}$) and (2) sex, age, and age2</td>
</tr>
<tr>
<td>Prostate cancer, Swedish men&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Nontargeted UPLC-MS, 6138 molecular features</td>
<td>893</td>
<td>7 replicating loci (PYROX2D, FADS1, PON1, CYP4F2, UGT1A8, ACADL, and LIPC); pathway analysis reflected enrichment of genes with acyl-CoA dehydrogenase activity</td>
<td>$P=8.8.10^{-13}$ to $3.4.10^{-38}$</td>
<td>(1) Bonferroni ($P=2.44\times10^{-11}$) and (2) none</td>
</tr>
<tr>
<td>Finnish individuals&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Nontargeted NMR, 117 metabolites</td>
<td>8330</td>
<td>31 loci (including SLC1A4, PPM1K, F12, SLC25A1, GCKR, GP62C, CPT1A, PCSK9, ANGPTL3, LPL, ABCA1, FADS1-3, LIPC, CETP, LPL, LDR4, APOE, and PLTP); metabolites showed heritability</td>
<td>$P=8.7.10^{-11}$ to $2.5.10^{-38}$</td>
<td>(1) Bonferroni ($P=2.31\times10^{-11}$) and (2) age, sex, and 10 genetic principal components</td>
</tr>
<tr>
<td>Two European cohorts&lt;sup&gt;43&lt;/sup&gt;</td>
<td>Nontargeted NMR in urine and plasma, 526 metabolite peaks</td>
<td>211</td>
<td>3 replicating loci (PYROX2D, NAT8, and AGX72)</td>
<td>$P=8.6.10^{-11}$ to $2.8.10^{-38}$</td>
<td>(1) Permutation-based procedure constraining genome-wide false discovery probability to be &lt;0.001 for each metabolite GWAS and (2) age and sex</td>
</tr>
<tr>
<td>KORA population–based cohort&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Targeted tandem MS, 363 metabolites</td>
<td>284</td>
<td>No individual metabolite genome-wide significant; 4 loci genome-wide significant for metabolite ratios (FADS1, LIPC, SCAD, and MCD2)</td>
<td>$P=2.0.10^{-3}$ for best single metabolite; $P=10^{-31}$ to $10^{-38}$ for metabolite ratios</td>
<td>(1) Bonferroni ($P=1.33\times10^{-5}$) and (2) none</td>
</tr>
<tr>
<td>European cohorts&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Targeted electrospray ionization tandem MS, 33 sphingolipids</td>
<td>4400</td>
<td>5 loci with the strongest associations in/near 7 genes involved in ceramide biosynthesis and trafficking (SPATL3, LASS4, SGP1, ATP100, and FADS1-3); SNPs in 3 loci, but not necessarily the same SNPs (ATP100, FADS3, and SPATL3) were also associated with myocardial infarction in different cohorts</td>
<td>$P=1.5.10^{-19}$ to $2.2.10^{-31}$</td>
<td>(1) Bonferroni ($P=1.6.10^{-11}$) and (2) age and sex</td>
</tr>
<tr>
<td>German cohort&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Nontargeted GC tandem MS, 517 metabolic traits</td>
<td>1768</td>
<td>34 loci with strongest results for 7 loci (SLC22A2, COMT, CYP3A5, CYP2C18, GB3A, UGT3A1, and rs12411395); also overlaid metabolic networks to generate hypotheses for unknown compounds and performed experimental validation of those compounds</td>
<td>$P=2.3.10^{-13}$ to $3.2.10^{-25}$</td>
<td>(1) Bonferroni ($P=4.5.10^{-11}$) and (2) age</td>
</tr>
<tr>
<td>SHIP and KORA population–based studies&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Nontargeted NMR in urine, 59 metabolites</td>
<td>2893</td>
<td>5 loci validated (SLC7A9, NAT2, SLC6A20, ACGT2, and WDR66)</td>
<td>$P=2.3.10^{-13}$ to $2.3.10^{-25}$</td>
<td>(1) Bonferroni ($P=4.5.10^{-11}$) and (2) age</td>
</tr>
<tr>
<td>KORA population–based and TwinsUK studies&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Nontargeted tandem MS, &gt;250 metabolites</td>
<td>2820</td>
<td>37 loci (including NAT8, GCKR, ABO, ACADS, ACADL, UGT1A, CPS1, and ELOVL2); 30 of these mapped to protein biochemically linked to associated metabolites; 15 associated with disease end points from previous studies in other cohorts</td>
<td>$P=1.4.10^{-12}$ to $4.4.10^{-33}$</td>
<td>(1) Bonferroni ($P=2.0.10^{-11}$) and (2) age, sex, and family structure</td>
</tr>
<tr>
<td>Meta-analysis of 5 European family–based studies&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Targeted tandem MS, 153 phospholipids and sphingolipids</td>
<td>4034</td>
<td>35 loci (including FADS1-2-3, PAQR9, AGPAT1, PKD2L1, PDXC1, PLD2, APOE, PNLP, P2R, APOH3, APOA1, ELOVL2, LIPC, and APOE); 5 associated with disease phenotypes in other cohorts (FADS1-2-3, AGPAT1, and APOA1)</td>
<td>$P=4.9.10^{-8}$ to $9.9.10^{-26}$</td>
<td>(1) Bonferroni ($P=2.2.10^{-5}$) and (2) familial relatedness</td>
</tr>
<tr>
<td>Cardiovascular risk in YFS and NFBC66&lt;sup&gt;44&lt;/sup&gt;</td>
<td>Nontargeted NMR in serum, 130 metabolite measures</td>
<td>6608</td>
<td>34 loci (including PCSK9, APOE, GCKR, ELOVL2, LPL, ABCA1, FADS1-2-3, and CETP); SERPINA1 and APOA1 showed eQTL, upregulated in atherosclerotic plaques</td>
<td>$P=3.9.10^{-9}$ to $3.9.10^{-34}$</td>
<td>(1) Metabolic networks used as traits; significance set at $P=4.5.10^{-6}$ and (2) none</td>
</tr>
<tr>
<td>KORA population–based and TwinsUK studies&lt;sup&gt;50&lt;/sup&gt;</td>
<td>LC- and GC–tandem MS in plasma/serum, 529 metabolites</td>
<td>7824</td>
<td>145 loci (including NAT8, LIPC, ANGPTL3, FM03, ELOVL2, FADS1, CETP, CPS1, APOE, PPM1K, and 84 new loci); integrated with gene expression; created web-based resources for data mining and results visualization</td>
<td>$P=5.2.10^{-3}$ to $6.2.10^{-40}$</td>
<td>(1) Bonferroni ($P=1.03\times10^{-11}$ for individual metabolites; $P=5.08\times10^{-37}$ for metabolite ratios) and (2) age and sex</td>
</tr>
</tbody>
</table>

eQTL indicates expression quantitative trait loci; GC, gas chromatography; LC, liquid chromatography; MS, mass spectrometry; NFBC66, Northern Finland Birth Cohort 1966; NMR, nuclear magnetic resonance; SHIP Study of Health In Pomerania; SNP, single-nucleotide polymorphism; UPLC, ultrahigh performance liquid chromatography; and YSF, Young Finns Study.
and adjustment for multiple comparisons should be delineated early in study implementation. Built into study design should be plans for laboratory validation of molecular targets, replication in other cohorts, and experimental validation of pathways identified from statistical analysis. For example, the association between BCAA, SCDA, betaine-derived metabolites, and cardiometabolic diseases has been validated in independent cohorts, as have several of the metabolic quantitative trait loci identified in metabolic GWAS, but other findings reviewed herein have not. Other challenges being tackled by the scientific community include a need for a common nomenclature for metabolomics, standardization across molecular platforms, and development of more robust analytic techniques for the large \( P \) and small \( n \) issue (ie, thousands to millions of molecular data points for each sample, but a relatively small number of samples). Systems biology is collaborative and multidisciplinary by nature, but it fails if study teams are lacking in individuals with the relevant expertise and ability to converse across fields. Finally, the availability of relevant biological samples with well-annotated clinical phenotypes is vital to maximize the signal/noise ratio.

In the future, we can anticipate the need to further refine analytic and bioinformatics approaches to accommodate even more high-dimensional datasets. Molecular omic datasets will be enhanced by integration with more detailed environomes, exposomes, and phenomes. Research on the role of the gut microbiome in regulation of host metabolism, hormonal milieu, and inflammatory tone is exploding, and a complete molecular profile will soon come to include information about gut microbiome composition, genetics, and metabolism. Technologies are also evolving to enable high-throughput molecular phenotyping on single cells, facilitating mechanistic studies in heterogeneous tissues. Perhaps most importantly, omic analyses of the future will integrate all of these tools to define cause-and-effect mechanisms and to guide experimental validation of identified pathways.

In conclusion, although investigators should be careful about analytic and bioinformatics challenges, integrated metabolomic genetic analyses and systems biology approaches hold great potential for furthering our understanding of biomarkers and mechanisms of health and disease and
moving the scientific community closer to an eventual goal of more personalized medicine.

Acknowledgments

We thank our colleagues from the Duke Molecular Physiology Institute for their critical contributions to the work cited from our group.

Sources of Funding

Work cited in this article was supported by National Institutes of Health grants HL095987 (Shah) and P01-DK58398 (Newgard) and by a sponsored research agreement from Pfizer.

Disclosures

The authors hold a patent on a related finding. Dr Newgard is a member of the Pfizer CVMED scientific advisory board.

References


Key Words: genomics ■ metabolism ■ metabolomics ■ risk factor
Integrated Metabolomics and Genomics: Systems Approaches to Biomarkers and Mechanisms of Cardiovascular Disease
Svati H. Shah and Christopher B. Newgard

Circ Cardiovasc Genet. 2015;8:410-419
doi: 10.1161/CIRCGENETICS.114.000223

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/8/2/410

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics 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: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org/subscriptions/