Human Metabolic Individuality in Biomedical and Pharmaceutical Research

Siddharth Prakash, MD, PhD


Study Hypothesis

Although genome-wide association studies have discovered candidate loci for hundreds of human traits, from height to lipid concentrations, the relationships between genetic variants and the phenotypes that they influence remain unclear. In many cases, the functions of candidate genes are not previously known or discoverable by use of standard techniques. Suhre et al hypothesized that biochemical analysis of serum metabolites can provide an important new resource to determine how some of these risk alleles may promote disease. They leveraged an extensive library of metabolites obtained by profiling human serum from 2 large European cohorts to discover previously unsuspected networks of regulatory genes, metabolites, and diseases. In combination with known human disease associations and biochemical pathways, they extracted novel gene-disease associations that merit testing in follow-up studies.

How Was the Hypothesis Tested?

Metabolites were measured in fasting serum on the Metabolon platform with 2 separate ultra-high-performance liquid chromatography/tandem mass spectrometry injections and 1 gas chromatography/mass spectrometry injection per sample. In the first phase of the study, the concentrations of 258 individual metabolites and all possible ratios of metabolites (32,499) were screened for associations with a panel of 534,635 single-nucleotide polymorphisms (SNPs) by use of a log-linear model in 2 different cohorts: the German KORA F4 study (n=1768, genotyped with the Affymetrix GeneChip array 6.0) and the British TwinsUK study (n=1052, genotyped with a combination of Illumina HumanHap300, HumanHap610Q, 1M-Duo, and 1.2M-Duo 1M). Data for 37 additional metabolites were of insufficient quality and were discarded. KORA F4 was a population-based study of unrelated adults from southern Germany. The TwinsUK cohort is a nationally recruited registry of predominately adult female (97%) twins. Associations that were significant in both cohorts were selected for a secondary meta-analysis with imputed HapMap2 SNPs to derive the final probability values that were reported in the study. In addition to conservative Bonferroni corrections for individual metabolites, P-gain statistics were used to select ratios of metabolites with significant interactions for further analysis.

Principal Findings

In a genome-wide association study of more than 250 sentinel metabolites from 2820 individuals, Suhre et al identified 37 loci with significant effects on blood metabolite concentrations, which they named “genetically determined metabolotypes” (GDMs). They created regional association plots of their imputed and meta-analyzed data to evaluate all candidate genes in linkage disequilibrium with the most strongly associated SNPs. Almost all of these SNPs were common, with minor allele frequencies exceeding 10%. Thirty of the 37 loci encoded proteins that were previously known to regulate the synthesis, degradation, or transport of the associated metabolites. Fourteen of these loci overlapped genome-wide association study hits in the National Human Genome Research Institute catalog, and 10 candidate genes in GDM loci were previously associated with common human disorders, including ABO and FUT2, which were previously implicated as risk factors for coronary artery disease and venous thromboembolism. Proteins encoded by these genes interact in a biochemical pathway that regulates serum fibrinogen, a clotting factor that is an independent risk factor for myocardial infarction. Using a series of randomly selected SNPs with similar properties, they demonstrated that the GDM loci are highly enriched for disease-associated variants. They also identified a previously unrecognized function of the SLC16A9 protein as a carnitine transporter, which supports
their observation that a SNP in the SLC16A9 gene is significantly associated with serum carnitine concentrations.

**Implications**

Over the past decade, metabolic profiling with $^1$H nuclear magnetic resonance or mass spectroscopy has been applied to a wide variety of human disorders to discover new biomarkers of disease progression or severity and to generate testable hypotheses about disease etiology. Metabolome-wide association studies may provide a more accurate and comprehensive assessment of complex disease susceptibility than typical genome-wide association studies, because metabolic profiles reflect the global physiological impact of genetic, epigenetic, transcriptomic, and proteomic processes with environmental and lifestyle factors. This plausibly explains the large effect sizes (10% to 60%) of GDMs in this study. Suhre et al pioneered the integration of metabolic and genome-wide data to discover new functional relationships between disease-associated variants. Notably, they also presented the largest reported sample of metabolic data and greatly expanded the number of verified GDM loci with exceptionally stringent criteria, requiring genome-wide significance in 2 independent cohorts. In a comparable study of 526 urine and plasma metabolites in 211 individuals, Nicholson et al identified 4 GDMs with genome-wide significance. Suhre et al also found that their GDM loci were enriched for disease-associated candidate genes and tended to interact in common biochemical pathways. One important caveat of this study is that chromatographic separation of metabolites before mass spectroscopy is associated with potentially significant experimental variability compared with nuclear magnetic resonance, which has been used in most mammalian metabolomics publications. Nevertheless, the groundbreaking results of Suhre et al demonstrate how metabolomics can generate new insights into the pathogenesis of complex disorders, which opens exciting avenues for biomarker discovery and targeted therapies.

**Acknowledgments**

The author is a member of the Early Career Committee of the American Heart Association Functional Genomics and Translational Biology Council.

**Disclosures**

None.