It is estimated that at least 29.1 million Americans, or 9.3% of the US population, currently have diabetes (1), a disease characterized by impaired insulin action and/or production. Although type 2 diabetes (T2D),4 which accounts for >90% of diagnosed diabetes, is largely predictable through anthropometric, lifestyle, and clinical factors, and is preventable through diet and lifestyle modifications, the metabolic pathways underlying its development and progression are incompletely understood. The rapidly developing area of metabolomics, which is designed to quantitatively profile a large number of small molecules in cells or biofluids, has emerged as a promising approach to elucidate altered metabolic pathways and discover novel biomarkers in T2D.

The past several years have seen the initial success of metabolomics in identifying novel biomarkers for insulin resistance and T2D. In 2009, Newgard et al. (2) compared 131 targeted metabolites between 74 obese and 67 lean subjects and found that plasma concentrations of branched-chained amino acids (BCAAs) were strongly correlated with obesity and insulin resistance. In 2011, Wang et al. (3) measured a panel of 61 metabolites and found that 5 BCAAs and aromatic amino acids (i.e., isoleucine, leucine, valine, tyrosine, and phenylalanine) were predictive of developing diabetes in the Framingham Offspring Study and the Malmö Diet and Cancer Study. In 2012, Wang-Sattler et al. (4) quantified 140 metabolites in 4297 fasting serum samples in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam cohort, which has collected blood samples, detailed measurements of anthropometric parameters and blood pressure, dietary and lifestyle questionnaires, and verified clinical outcomes during a follow-up period of approximately 20 years. The study included 300 incident T2D cases and 300 randomly selected controls matched on age, sex, fasting time, time of day of blood sampling, and season at blood sampling. The study population was then randomly split into 2 internal substudies, each containing 150 matched case-control pairs to verify the internal consistency of findings. The authors observed that diverse altered classes of metabolites, including 6 lipids and 7 nonlipids, preceded the onset of overt T2D by a median of 6 years. More specifically, higher serum concentrations of lipids in the phosphatidylcholine (PC) class, PC(22:4/18:0) and PC(O-18:0/22:5), and lower concentrations of 2-aminoadipic acid were associated with a significantly increased risk of T2D among a panel of 70 metabolites during 12 years of follow-up.

The above studies have revealed several promising novel biomarkers of T2D; however, they have been limited in the number of metabolites detected and analyzed (approximately 100–200 targeted metabolites). In an article in the current issue of Clinical Chemistry, Drogan et al. (7) applied an untargeted metabolomic approach with a coverage of >4500 metabolite features by ultraperformance LC-MS with a protocol specifically designed for large-scale metabolomic studies regarding robustness and repeatability. The study was nested in the well-characterized European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam cohort, which has collected blood samples, detailed measurements of anthropometric parameters and blood pressure, dietary and lifestyle questionnaires, and verified clinical outcomes during a follow-up period of approximately 20 years. The study included 300 incident T2D cases and 300 randomly selected controls matched on age, sex, fasting time, time of day of blood sampling, and season at blood sampling. The study population was then randomly split into 2 internal substudies, each containing 150 matched case-control pairs to verify the internal consistency of findings. The authors observed that diverse altered classes of metabolites, including 6 lipids and 7 nonlipids, preceded the onset of overt T2D by a median of 6 years. More specifically, higher serum concentrations of lipids in the phosphatidylcholine (PC) class, PC(22:4/dm18:0) and PC(O-18:0/22:5), and lower concentrations of PC(O-20:0/O-20:0) were related to a higher risk of T2D. Higher serum concentrations of 2-aminoadipic acid were associated with a significantly increased risk of T2D among a panel of 70 metabolites during 12 years of follow-up.

Although further validation in independent external cohorts is warranted, these findings underscore the po-

---

1 Department of Nutrition and 2 Department of Epidemiology, Harvard School of Public Health, Boston, MA; 3 Channing Laboratory, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA.

* Address correspondence to this author at: Department of Nutrition, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115. Fax 617-432-2435; e-mail frank.hu@channing.harvard.edu.

Received December 28, 2014; accepted December 29, 2014.

Previously published online at DOI: 10.1373/clinchem.2014.235986

© 2014 American Association for Clinical Chemistry

4 Nonstandard abbreviations: T2D, type 2 diabetes; BCAA, branched-chained amino acid; EPIC, European Prospective Investigation into Cancer and Nutrition; PC, phosphatidylcholine; MS, mass spectrometry; NMR, nuclear magnetic resonance; GWAS, genome-wide association study.
tential of using untargeted metabolomic approaches with a wide coverage for novel biomarker discovery.

The findings of the present study, together with the results of previous studies, have advanced our understanding of metabolic pathways underlying the development of T2D. Thus far, several metabolite classes have emerged as candidate biomarkers of T2D. The most consistent findings are centered on the metabolism of amino acids and lipids. With respect to individual metabolite classes, amino acids, in particular BCAAs (3), seem to hold the most promise as markers of insulin resistance and prediabetes from large population-based studies and animal models. Whether these potential biomarkers could be used in clinical settings remains unclear, however, because the cut points of BCAAs for increased risk of T2D have not been defined. In addition, whether these biomarkers add to the prediction of T2D beyond traditional risk factors is yet to be determined. Lipids, including lyso-phosphatidylcholines (4, 7), fatty acids (8), and acylcarnitines (9), are also highlighted as potential biomarkers of T2D. Although these biomarkers can help us better understand the role of nutrient and food metabolism in the development of T2D, their clinical utility is likely to be limited at this time.

Currently, 2 main platforms [mass spectrometry (MS) and nuclear resonance spectroscopy (NMR)] and 2 main approaches (targeted and untargeted) are routinely applied in the metabolomic field. MS-based methods characterize a metabolite by the retention time and mass-to-charge ratio (m/z) when combined with chromatography (10). In the NMR-based methods, a molecule is identified by a specific pattern, i.e., chemical shift in the resonance spectrum of its protons when excited by an oscillating magnetic field (11). Compared to the NMR-based methods, the MS-based protocols are more analytically sensitive and have lower costs for instrumentation. Therefore, MS-based protocols are more suitable and more widely used in large population studies. Targeted approaches are more appropriate for quantifying only a selected set of known metabolites, whereas untargeted approaches are more global in scope and typically involve numerous previously uncataloged metabolites. There is usually a trade-off when choosing between a wide and largely untargeted panel, which often comes at the cost of lower data quality, and a narrower targeted panel, which may miss potentially interesting metabolites. With rapid advances in profiling technologies, the untargeted metabolomics approach may have wider applications in the future.

Metabolomics is designed to profile the end products of the upstream genome, transcriptome, and proteome. Integration of metabolomic data with the upstream omics data could maximize the potential impact of metabolomics in T2D research. Also, in combination with metabolomics, these upstream omics studies can gain functional correlates at the level of a metabolic readout. Genome-wide association studies (GWAS) have identified >70 susceptibility loci associated with T2D (12). However, these loci typically have fairly modest effects and do not add to the clinical prediction of diabetes beyond that of traditional risk factors. GWAS performed in the context of a high-throughput assessment of intermediate molecular traits (i.e., metabolomics) can lead to a better understanding of how sequence variation in diabetes-associated regions mediates pathology. For example, Yu et al. (13) presented a well-conducted GWAS of 308 untargeted metabolites among 1977 African Americans from the Atherosclerosis Risk in Communities study. Nineteen significant common variant-metabolite associations and 4 potential disease-associated pathways were identified. As one example, variation of the gene for trehalose hydrolysis (TREH, trehalase [brush-border membrane glycoprotein]) was related to the concentrations of the metabolite trehalose, which can be cleaved to 2 molecules of glucose; the metabolite trehalose was strongly associated with fasting glucose concentrations and showed a positive relation with incident diabetes after a mean 7 years of follow-up, even with adjustment for fasting glucose values. This study highlights the value of using systematic measures of metabolites proximal to gene function to offer new insights into diabetes pathology. In addition, several recent studies have attempted to integrate multiple omics technologies. Ferrara et al. (14) demonstrated that genomic analysis could be integrated with mRNA expression and metabolite profiling data to construct potentially causal pathway networks of diabetes pathophysiology in animal models. Chen et al. (15) presented an integrative personal omics profile by combining genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from a single individual over a 14-month period. This profile revealed novel biological pathways underlying various health risks; e.g., a viral infection triggered onset of the increased glucose response.

In population-based studies, traditional epidemiology typically links environmental factors such as diet and lifestyle to diabetes and its various clinical parameters. This approach has successfully identified numerous diabetic risk factors. However, the underlying mechanisms regarding these risk factors and T2D are largely unknown. Recent advances in omics technologies have enabled the emergence of the “systems epidemiology” approach to improving our understanding of these biological mechanisms and to possibly unlocking the “black box” (16). Systems epidemiology integrates a wide range of high-throughput technologies such as genetic predisposition (genome), epigenetic changes (epigenome), the expression of genes (transcriptome), proteins (proteome), metabolites (metabolome), and gut microbiota (microbiome) into well-characterized large, prospective
cohort studies for which biological samples are available. This approach has the potential to advance our understanding of pathophysiology of T2D and to enable early detection and interventions by identifying high-risk individuals.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References