Metabolic profiling in diabetes

Karsten Suhre¹,²

¹Department of Physiology and Biophysics, Qatar Foundation – Education City, Weill Cornell Medical College – Qatar, PO Box 24144, Doha, Qatar
²Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

Abstract

Metabolic profiling, or metabolomics, has developed into a mature science in recent years. It has major applications in the study of metabolic disorders. This review addresses issues relevant to the choice of the metabolomics platform, study design and data analysis in diabetes research, and presents recent advances using metabolomics in the identification of markers for altered metabolic pathways, biomarker discovery, challenge studies, metabolic markers of drug efficacy and off-target effects. The role of genetic variance and intermediate metabolic phenotypes and its relevance to diabetes research is also addressed.

Key Words

- diabetes
- metabolism
- lipid
- insulin resistance
- glucose metabolism

Introduction

Metabolic profiling, or metabolomics, aims at determining all relevant small molecules (metabolites) in a biological sample (Pauling et al. 1971). Recent technology advances have allowed the characterization of hundreds of metabolites from a small amount of blood or urine in a single experiment (Nicholson et al. 1999). The metabolic profile of an individual patient, a lab animal, or even a cell culture provides a functional readout of its current metabolic state (Griffin et al. 2001, Fiehn 2002, Nicholson et al. 2002, Bain et al. 2009). Metabolomics can thus be viewed as the equivalent to an ‘imaging approach’ for the biochemical processes that occur within an organism. The biochemical knowledge that has been accumulated in over a century of individual experiments on cells and tissue cultures provides a thorough basis for the interpretation of perturbations in the metabolic state of a patient and of the underlying metabolic connections inside the human body, as they are revealed by studies of metabolic profiling in diabetes (German et al. 2005, Wenk 2005). Application of metabolomics approaches to metabolic disorders, especially diabetes, is particularly promising since deregulations of metabolic processes are expected to be directly related to relevant disease end-points. Friedrich (2012) recently reviewed current findings of metabolic research regarding diabetes in animal models and humans in this journal. I therefore first give an introduction to metabolomics with an update on current diabetes studies using metabolomics (Table 1), followed by a focus on study design considerations, current challenges and future directions of metabolic profiling, and how metabolomics can be applied most successfully in diabetes research (Table 2).

Today, two fundamentally different techniques are mainly used to obtain metabolomics data on larger sets of samples: mass spectrometry (MS) and proton (¹H) nuclear magnetic resonance (NMR) spectroscopy (Nicholson & Lindon 2008). Both techniques have their strengths and weaknesses. MS-based methods are in general more sensitive, but they are also more complex in their implementation with a higher risk of generating technical artifacts and measurement errors. NMR-based methods are more robust and straightforward to implement, cheaper, faster, and especially the reproducibility of NMR spectra is excellent. While MS calls for complex sample extraction methods, NMR requires higher sample volumes.
MS is most often coupled to additional gas- or liquid-phase chromatography separation steps (GC–MS and LC–MS) that allow for the stratification of complex mixtures. As a consequence, MS-based methods provide an overall larger area under a curve using arbitrary units. Targeted methods with absolute quantification have the advantage that measurements from different studies can be easily combined and compared. While non-targeted metabolomics approaches allow for the discovery of new molecules that associate with the phenotype under investigation. Targeted approaches typically attempt to quantify the measured metabolites on an absolute scale, using external or internal standards as a reference, while non-targeted approaches provide only semi-quantitative measures, such as ion counts or areas under a curve using arbitrary units. Targeted methods with absolute quantification have the advantage that measurements from different studies can be easily combined and compared. They typically have lower technical variance since more time can be spent on measurements from different studies can be easily combined and compared. While non-targeted metabolomics approaches allow for the discovery of new molecules that associate with the phenotype under investigation. Targeted approaches typically attempt to quantify the measured metabolites on an absolute scale, using external or internal standards as a reference, while non-targeted approaches provide only semi-quantitative measures, such as ion counts or areas under a curve using arbitrary units. Targeted methods with absolute quantification have the advantage that measurements from different studies can be easily combined and compared. They typically have lower technical variance since more time can be spent on measuring a smaller set of metabolic features. However, discovery of new processes and pathways is more likely to be done using non-targeted methods, aiming at measuring the widest possible range of metabolites.

Table 1  Selected human metabolomics studies with diabetes-related phenotypes

<table>
<thead>
<tr>
<th>Metabolomics method</th>
<th>Experimental model</th>
<th>Key results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targeted metabolomics</td>
<td>Incident type 2 diabetes (T2D)</td>
<td>Plasma phospholipid markers of T2D</td>
<td>Wang et al. (2005)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Lean/obese with insulin resistance</td>
<td>Branched-chain amino acids (BCAAs) differ-entiate obese and lean individuals and contribute to insulin resistance</td>
<td>Newgard et al. (2009)</td>
</tr>
<tr>
<td>Multi-platform metabolomics</td>
<td>Euglycemic clamp</td>
<td>Metabolomic profiles reflective of glucose homeostasis in nondiabetic and T2D obese African–American women</td>
<td>Gall et al. (2010)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Incident T2D</td>
<td>Metabolic signatures of insulin resistance in young adults</td>
<td>Fiehn et al. (2010)</td>
</tr>
<tr>
<td>1H NMR metabolomics</td>
<td>HOMA-IR (patients with T2D were excluded)</td>
<td>Metabolic signatures of insulin resistance in young adults</td>
<td>Würzt et al. (2012)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Incident T2D and impaired fasting glucose</td>
<td>Confirms the role of catabolism of BCAAs in T2D and impaired fasting glucose</td>
<td>Menni et al. (2013)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Incident T2D, multi-matrix (saliva, plasma, urine)</td>
<td>1,5-anhydroglucitol (1,5-AG) in saliva is a noninvasive marker of short-term glycemic control</td>
<td>Mook-Kanamori et al. (2014)</td>
</tr>
<tr>
<td>Prospective studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-targeted metabolomics (targeted for replication)</td>
<td>Individuals who developed T2D over 12 years</td>
<td>BCAAs are predictive of diabetes</td>
<td>Wang et al. (2011)</td>
</tr>
<tr>
<td>Targeted metabolomics</td>
<td>Individuals who developed insulin resistance or T2D over 7 years</td>
<td>Glycine and L-GPC are predictors of impaired glucose tolerance and of T2D</td>
<td>Wang-Sattler et al. (2012)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Two cohorts (3 and 9.5 years follow-up) with diverse diabetes relevant traits</td>
<td>AHB and L-GPC are independent predictors of impaired glucose tolerance</td>
<td>Ferrannini et al. (2013)</td>
</tr>
<tr>
<td>Targeted metabolomics</td>
<td>Two cohorts with a 7-year follow-up and diabetes relevant traits</td>
<td>Sugar metabolites, amino acids, and choline-containing phospholipids are independently associated with risk of T2D 2-aminoacidic acid is a biomarker for diabetes risk</td>
<td>Floegel et al. (2013)</td>
</tr>
<tr>
<td>Non-targeted metabolomics</td>
<td>Individuals who developed T2D over 12 years</td>
<td></td>
<td>Wang et al. (2013)</td>
</tr>
</tbody>
</table>

Depending on the objective, targeted or non-targeted metabolomics approaches may be favored. Targeted approaches focus on a specific subset of the metabolome and provide data only on a predefined set of molecules, while non-targeted metabolomics approaches allow for the discovery of new molecules that associate with the phenotype under investigation. Targeted approaches typically attempt to quantify the measured metabolites on an absolute scale, using external or internal standards as a reference, while non-targeted approaches provide only semi-quantitative measures, such as ion counts or areas under a curve using arbitrary units. Targeted methods with absolute quantification have the advantage that measurements from different studies can be easily combined and compared. They typically have lower technical variance since more time can be spent on measuring a smaller set of metabolic features. However, discovery of new processes and pathways is more likely to be done using non-targeted methods, aiming at measuring the widest possible range of metabolites.
The Human Metabolome Database (Wishart et al. 2013), which is probably the largest and most comprehensive metabolomics database available to date (http://www.hmdb.ca/), contains over 40,000 entries. Note, however, that this number includes an important number of ‘expected’ metabolites. These are small molecules that have been predicted to exist based on the metabolic capabilities of known metabolic pathways, but that these ‘expected’ metabolites have not yet been detected in a biological sample. These ‘expected’ metabolites comprise, in particular, lipid species that are made up of several fatty acid side chains with different chain lengths and degrees of desaturation that may occur in a combinatorial fashion (Wenk 2005). The number of metabolites becomes much smaller when the scope is limited to confirmed metabolites in a single biological fluid: in an attempt to detect as many known metabolites as possible in a single sample, a multi-platform metabolomic approach using NMR and a set of different types of MS identified 445 and quantified 378 unique metabolites or metabolite species in human urine (Bouatra et al. 2013). Overall, the human urine metabolome database describes 2651 confirmed human urine metabolite species (http://www.urinemetabolome.ca). To date, no single existing technique alone can be expected to provide the full picture in any diabetes study with metabolomics. In expectation of future technical improvements, it is thus advisable to store valuable specimens in adequate bio-banks for future re-analysis, ideally aliquoted into small volumes so that no additional thawing cycles are required (Fliniaux et al. 2011, Yin et al. 2013).

**Cross-sectional studies for the identification of altered metabolic pathways**

Altered metabolite profiles can be the indicators of changes in disease-relevant metabolic processes. Metabolites that associate with relevant disease phenotypes can be developed into biomarkers that are indicative of a metabolic deregulation. However, the question of whether a perturbation in a metabolite is causal to a disease phenotype, consequential, or just driven by a confounding factor needs to be addressed eventually. Inter-correlations between different metabolites can help with answering this question. The ‘simplest’ design of a diabetes study with metabolomics is a case–control design, in which blood is collected at a single time point from a group of patients with diabetes and from a ‘healthy’ group of controls. Statistically significant differences in metabolite concentrations between the groups are then analyzed and interpreted. Cases and controls can be matched to reduce the influence of confounding factors, but modeling of covariates into statistical models appears to be the preferable alternative (Faresjo & Faresjo 2010, de Graaf et al. 2011). Often, cross-sectional studies draw their samples from existing larger population studies that have an epidemiological background. A pilot study (Suhre et al. 2010), conducted in the general German Cooperative Health Research in the Region of Augsburg (KORA) population (Wichmann et al. 2005) using 40 cases of type 2 diabetes (T2D) and 60 controls, showed that diabetes-related complications can already be detected under subclinical conditions by screening over 420 unique small molecules on three different metabolomics platforms (Fig. 1). This study identified diabetes-associated perturbations of metabolic pathways linked to kidney dysfunction (3-indoxyl sulfate), indicators of interaction with the gut microflora (bile acids), and lipid metabolism (glycerophospholipids, free fatty acids). The latter observation replicated early work on plasma phospholipid metabolic profiling by Wang et al. (2005), who reported associations of different types of phospholipids with T2D. In a cohort of 399 nondiabetic subjects with a broad spectrum of insulin sensitivity and glucose tolerance and using a nontargeted metabolomics approach, Gall et al. (2010) found that α-hydroxybutyrate (AHB) is an early marker for both insulin resistance and

**Table 2** Examples of application of metabolomics in diabetes research

<table>
<thead>
<tr>
<th>Application</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic biomarker discovery</td>
<td>1,5-AG is a marker of short-term glycemic control (Yamanouchi, et al. 1996)</td>
</tr>
<tr>
<td>Prognostic biomarker discovery</td>
<td>Branched chain amino acids are predictive of insulin resistance (Wang et al. 2011)</td>
</tr>
<tr>
<td>Elucidation of disease pathways</td>
<td>Identification of the role of solute carriers in the physiological response to glucose challenge (Deo et al. 2010)</td>
</tr>
<tr>
<td>Identification of drug side effects</td>
<td>Methylglutaryl carnitine is oppositely affected by rosiglitazone treatment in diabetic and WT mice (Altmairer et al. 2008)</td>
</tr>
<tr>
<td>Functional biomarker for drug action</td>
<td>Ratios between specific phospholipid species are markers for FABP4 inhibition (Suhre et al. 2011a)</td>
</tr>
</tbody>
</table>
impaired glucose regulation. The underlying biochemical mechanisms may involve increased lipid oxidation and oxidative stress. Menni et al. (2013) conducted a much larger cross-sectional metabolomics study on diabetes, with a broad nontargeted panel of 447 metabolites in blood plasma from participants of the TwinsUK study (Moayyeri et al. 2013). The authors found molecular markers that arise before and after hyperglycemia in a large population-based study of 2204 female participants, of whom 115 had T2D and 192 had impaired fasting glucose. In their study, 42 metabolites from three major energy sources, carbohydrates, lipids, and proteins, associate significantly with T2D after adjusting for multiple testing. Twenty-two of these metabolites had already previously been reported in association with T2D or insulin resistance, 20 were newly discovered.

Mook-Kanamori et al. (2014) conducted a case–control study to screen for diabetes-relevant metabolic readouts in saliva, using blood and urine as a reference. Using a nontargeted MS platform, they tested 2178 metabolites in saliva, blood plasma, and urine samples from 369 individuals for association with T2D. They observed a strong association of T2D with 1,5-anhydroglucitol (1,5-AG) in saliva that was robust across three different non-Caucasian ethnicities, irrespective of BMI, age, and gender. 1,5-AG in blood has been established in the past as a reliable marker of short-term glycemic control (Kim et al. 2012), indicating that 1,5-AG in saliva can be used as a noninvasive marker for screening undiagnosed diabetes.

Prospective studies and biomarker discovery

While cross-sectional case–control studies can provide valuable insights into the mechanisms and processes related to the actual disease state and its co-morbidities, and are particularly interesting for biomarker discovery for incident diabetes, observations taken before disease onset are required in order to identify metabolic biomarkers that are predictive of disease development and its related co-morbidities. Such longitudinal studies require samples to be collected for metabolic analysis even before concrete plans for these measurements had been conceived, which makes such samples particularly valuable. A hallmark study of this design is that of Wang et al. (2011). Using a panel of amino acids, amines, and other polar metabolites in a longitudinal setting with 2422 individuals followed for 12 years (201 of whom developed diabetes), the authors found that all three branched chain amino acids (BCAAs) – isoleucine, leucine, valine – and two aromatic amino acids – tyrosine and phenylalanine – are associated with future diabetes. Furthermore, a subset of three of them, with the ‘top combination’ of isoleucine, phenylalanine, and tyrosine, could be used to predict future diabetes. It is of note here that BCAAs have previously been associated with incident T2D in cross-sectional studies, indicating the importance of perturbed amino acid metabolism in patients with T2D (Newgard et al. 2009, Fieln et al. 2010, Gall et al. 2010, Suhre et al. 2010). Newgard et al. (2009) actually tested the effect of a high-fat diet supplemented with BCAA on insulin resistance in rats. Their findings indicated that in the context of a dietary pattern that includes high fat consumption, BCAAs contribute to the development of obesity-associated insulin resistance.

Würtz et al. (2012) studied the metabolic signatures of insulin resistance in 7098 young adults. They again confirmed the association of branched-chain and aromatic amino acids, gluconeogenesis intermediates, ketone bodies, and fatty acids with insulin resistance. Moreover, due to the large number of samples, they were able to identify statistically significant sex- and obesity-dependent interactions, showing that leucine, isoleucine, valine, and tyrosine were significant in women only if they were abdominally obese.

Using a lipidomics-oriented panel with 140 metabolites in 4297 fasting serum samples from a longitudinal population-based study, covering a time span of 7 years, Wang-Sattler et al. (2012) identified candidate biomarkers for pre-diabetes. Their study indicated that the levels of the three metabolites glycine,
linoleoylglycerophosphocholine (lyso-phosphatidylcholine 18:2, L-GPC), and acetyl carnitine were altered in individuals with impaired glucose tolerance (IGT). Lower levels of glycine and L-GPC were predictors of IGT and T2D. To test the predictivity of AHB and L-GPC for incident dysglycemia, Ferrannini et al. (2013) enrolled 1261 nondiabetic participants from the Relationship between Insulin Sensitivity and Cardiovascular Disease (RISC) study and 2580 from the Botnia Prospective Study, with 3- and 9.5-year follow-up data respectively. They found that AHB and L-GPC are independent predictors of degrading glucose tolerance and are physiologically consistent with a joint signature of insulin resistance and β-cell dysfunction. Floegel et al. (2013) used a targeted metabolomics approach to identify serum metabolites that are associated with the risk of T2D in a study comprising 800 incident cases of T2D and 2282 controls, randomly drawn from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam. The mean time of follow-up for this study was 7 years. In addition to associations found between several phospholipids, hexoses, and phenylalanine, lower levels of glycine and L-GPC were again found to be associated with an increased risk of developing T2D. In a nested case–control design with 188 cases and 188 matched controls of the Framingham Heart Study, Wang et al. (2013) identified 2-amino adipic acid (2-AAA) as being most strongly associated with the risk of developing diabetes within a 12 years follow-up time. Study participants in the top quartile of 2-AAA concentrations had a four times higher risk of developing T2D.

After identification of an association of a metabolic trait with a disease-related trait, the next challenge is how to develop this knowledge into a clinically applicable biomarker (Batch et al. 2014). Milburn & Lawton (2013) present an example of how a set of metabolites that is associated with a T2D-related phenotype can be translated into a fasting blood biomarker for the diagnosis and monitoring of insulin resistance. Cobb et al. (2013) conducted a clinical study of the resulting clinical test, called Quantose.

From these cross-sectional and prospective studies, it emerges that diabetes-related perturbations of metabolic pathways involve many different classes of metabolites, including carbohydrates, amino acids, in particular BCAAs and glycine, and certain lipids, such as phospholipids and L-GPC. Many associations were coherently associated with T2D-related phenotypes in several studies. In many cases the specific nature of the metabolic phenotypes provides deeper functional insights into the metabolic perturbations that may be at the basis of these associations (Newgard 2012). However, most of these studies focused on broad outcomes, such as incident diabetes or insulin resistance. It can be expected that future metabolomics studies on diabetes with more detailed characterization of the patients’ diabetes state and a specific focus on diabetes-related co-morbidities, such as skin disorders, micro-vascular, or macro-vascular incidents shall reveal even deeper insights into the pathophysiology of this complex disease.

**Challenge studies to investigate physiological perturbations**

Metabolomics measurement taken during physiological challenges may allow the identification of metabolic signatures that are not detectable in the fasting state (Shaham et al. 2008, Pellis et al. 2012, Ho et al. 2013). Krug et al. (2012) revealed the dynamic range of the human metabolome by submitting 15 young healthy male volunteers to a highly controlled challenge protocol over a 4-day period. The challenges included a 36 h fasting period, oral glucose and lipid tests, controlled liquid test meals, physical exercise, and cold stress challenges. The study resulted in a rich data set of 275 metabolic traits from blood, urine, exhaled air, and breath exhale condensate, measured at up to 56 time points and provides a unique reference for future metabolomics studies. The data collected for this study is freely available at http://metabolomics.helmholtz-muenchen.de/humet/. The authors show that physiological challenges increase the variation that is observed between individuals, revealing metabolic phenotypes that are not observable under baseline (typically overnight-fasting) conditions. Protocols developed within this study were applied by Wahl et al. (2014) to elucidate the role of genetic variance in the FTO gene, one of the major diabetes and obesity risk genes (Frayling et al. 2007). The metabolic response to five defined challenge tests, consisting of an oral glucose tolerance test, a standardized high-fat high-carbohydrate meal, and a lipid tolerance test, as well as an intravenous glucose tolerance test and a euglycemic hyperinsulinemic clamp, was investigated in the plasma samples of 25 homozygous carriers of the FTO risk allele (rs9939609 AA genotype) and 31 carriers of the protective TT genotype. Although this study found only minor effects of FTO genotype on the metabolite fluxes after standardized metabolic challenges, it highlights how gene and environment interactions can be investigated in future diabetes studies with metabolomics, using well-designed
physiological challenge protocols. Similarly, metabolomics-based challenge studies can be conducted using animal models of diabetes. For instance, transgenic pigs that express a dominant-negative glucose-dependent insulinotropic polypeptide receptor (GIPRdn) develop impaired glucose control and display loss of β-cell mass. These animals provide a unique opportunity for studying metabolic changes that occur before the onset of overt diabetes. Renner et al. (2012) report results from a metabolomics study from intravenous glucose tolerance tests of 2.5- and 5-month-old GIPRdn transgenic and control animals. They found that seven amino acids (Phe, Orn, Val, XLeu, His, Arg, and Tyr) were increased in 2.5-month-old pigs, but decreased in 5-month-old GIPRdn transgenic pigs compared with controls, and that certain sphingomyelins and phospholipids were decreased in the plasma of 5-month-old GIPRdn transgenic pigs. The observed changes in metabolite concentrations were associated with gene expression levels of relevant pathways in the liver and provide new insights into diabetes-related pathways that cannot be studied easily in humans.

Identification of metabolic markers of drug efficacy and off-target effects

Currently, the most frequently applied drugs for the therapy of T2D include suppressors of hepatic gluconeogenesis (metformin) (Hundal et al. 2000), insulin-sensitizing PPAR agonists (pioglitazone) (Hirose et al. 2002), and insulin secretagogues (sulfonylureas) (Groop et al. 1992). However, therapeutics that target the multifactorial nature of T2D to increase drug efficiency are still in an early phase of the development. In this study, metabolomics finds its application in the development and screening of new drugs (pharmacometabolomics) (Kaddurah-Daouk et al. 2008) by generating a largely unbiased metabolome-wide view of drug-induced perturbations that can be applied at each level of the therapeutics discovery pipeline from cellular and animal disease models to the preclinical and clinical translation (Robertson & Prevent 2013). Pharmacometabolomics can capture both a drug’s pathway of action and unexpected off-target effects of a new drug. In an early metabolomics study in mice, Watkins et al. (2002) found that the PPARγ antagonist rosiglitazone induced hypolipidemia and elicited an unusual accumulation of polyunsaturated fatty acids within adipose tissue in mice. Using a similar study design, Altmaier et al. (2008) found that methylglutaryl carnitine is oppositely affected by rosiglitazone to phosphatidylcholine ratios.

Starting from the observation that metabolomics may indicate functionally relevant perturbations in metabolism, the technique can also be used to identify mechanistic markers for drug action, which can then be used for screening and efficacy tests. For instance, the fatty acid-binding protein 4 (FABP4) is part of a family of lipid chaperones that control intracellular fatty acid fluxes. Mice lacking FABP4 are protected against genetic or diet-induced insulin resistance, making FABP4 inhibition a target for anti-diabetes drug development. Using a targeted metabolomics platform with a focus on lipid-related species, Suhre et al. (2011a) showed that FABP4 inhibition leads to a clear metabolic signature in the form of a shift between lipid species that contain different amounts of mono- and unsaturated C16 and C18 fatty acids (C16:0, C16:1, C18:0, and C18:1). Note that these fatty acids are substrates of FABP4. The biomarker identified in that study was a ratio between two phospholipids and can be implemented in high-throughput screenings of potential FABP4-inhibitory molecules. Metabolite profiling should thus have a firm place in the development of anti-diabetes drugs, be it for the investigation of functional mechanisms, the identification of drug side effects, the development of mechanistic markers of drug efficacy, or the evaluation of changes in metabolic profiles related to treatment response.

The role of genetic variance and the intermediate metabolic phenotypes

Diabetes is a metabolic disorder with a considerable heritable contribution. Studies indicate up to 69% heritability for T2D in patients, with age at onset of 35–60 years (Almgren et al. 2011). When investigating metabolic perturbations in the context of diabetes, it is therefore important to consider the role of genetic variance in modification of a patient’s body’s metabolic capabilities. Genome-wide association studies (GWAS) with metabolic phenotypes have identified many genetic loci that modulate the metabolic phenotype as a function of genotype (reviewed in Suhre & Gieger (2012)). These so-called genetically influenced metabolotypes (GIMs) are generally constituted of relatively frequent genetic variants, with minor allele frequencies of 20% and above, show large effect sizes that explain 10–20% of the observed variance in the metabolic trait, and are most often located in or near metabolically active genes, such as enzymes, transporters, and regulators thereof, and in
mostly all cases the associated metabolic phenotypes then match the function of the genes (Suhre et al. 2011b).

Despite its high heritability, GWAS with T2D have failed to identify genetic variants that explain the larger part of the genetic risk of developing diabetes for an individual. This lack of explained genetic risk is referred to as the ‘missing heritability’ (Maher 2008) and indicates that more complex interactions between many genetic variants need to be considered. To approach this problem, Kronenberg (2012) suggested investigating intermediate metabolic phenotypes as a means to connect genetic variants with disease end-points. For instance, the first large-scale association study for T2D (WTCCC 2007) found a marginal association of the LIPC locus with T2D (SNP rs4775041; $P = 0.061$). At that time, this locus also associated only weakly ($P = 0.025$) with blood triglyceride levels (Kathiresan et al. 2008). However, the first GWAS with metabolic traits (Gieger et al. 2008) identified a nearly genome-wide significant association of several phosphatidylethanolamine species (PE aa C38:6, $P = 9.6 \times 10^{-8}$) at this locus. Subsequent GWAS confirmed the association of the LIPC locus with triglycerides (Teslovich et al. 2010) and with phosphatidylethanolamines (Shin et al., Nature Genetics (In Press)), and revealed an association with metabolic syndrome (Kraja et al. 2011). Despite its low statistical power, the Gieger et al. (2008) GWAS with metabolomics could hence identify an intermediate metabolic trait that arguably lies on the pathway between genetic variants of the hepatic lipase LIPC and diabetes-related outcomes, such as triglyceride levels and metabolic syndrome. It is noteworthy herein that phosphatidylethanolamines were also found to be associated with diabetes in a subsequent study (Suhre et al. 2010).

A second example where genetics is linked to metabolomics in the study of diabetes is the glucokinase (hexokinase 4) regulator (GCKR) locus. This is a major pleiotropic risk locus associated with diabetes- and cardiometabolic-related traits, such as fasting glucose (Dupuis et al. 2010) and insulin levels (Aulchenko et al. 2009), triglyceride levels, and chronic kidney disease (Kottgen et al. 2010). It is also associated with many metabolic traits (Illig et al. 2010, Würzt et al. 2012). A recent GWAS for metabolic traits has also found a strong association of the GCKR locus with mannose-to-glucose ratios (Suhre et al. 2011b). The fasting level of mannose is lower in carriers of the risk allele, as opposed to that of glucose, which is higher. Little is known about the physiological role of mannose, other than its use in protein glycosylation. Mannose enters the cell via a specific transporter that is insensitive to glucose (Panneerselvam & Freeze 1996), and hepatic glycogen breakdown is implicated in the maintenance of plasma mannose concentrations (Taguchi et al. 2005). These observations and the association with GCKR detected in the GWAS study, which is even stronger than that of glucose with GCKR, call for further investigation of the role of mannose as a differential biomarker, or even as a potential point of intervention in diabetes care. A third example of a diabetes-relevant GIM is SNP rs10830963 in the gene encoding the melatonin receptor (MTNR1B) which is associated with fasting glucose (Prokopenko et al. 2009). The same SNP in associated with tryptophan-to-phenylalanine ratios in the study reported by Illig et al. (2010). Phenylalanine is a precursor of melatonin, therefore indicating a possible functional relationship between this pathway and the regulation of glucose homeostasis (Illig et al. 2010).

Beyond inherited genetic variance alone, epigenetic modifications of DNA are suspected to be associated with diabetes phenotypes (Ng et al. 2010, Drong et al. 2012). A recent epigenome-wide association study (EWAS) of DNA methylation with metabolomics has shown that DNA methylation also plays an important role in human metabolism (Petersen et al. 2014). In this study, we briefly highlight one diabetes-relevant example from that study: the thioredoxin-interacting protein (TXNIP) plays a functional role in glucose regulation. DNA methylation levels of CpG locus cg19693031 near the TXNIP gene were associated with levels of chylomicrons in blood plasma, and also with known metabolic markers of diabetes, including different phospholipids, hexose, and AHB. In a previous study, TXNIP expression was elevated in the muscle of pre-diabetic and diabetic study participants (Parikh et al. 2007). However, there was no evidence for association between common genetic variations in TXNIP and T2D. Hence, DNA methylation may play a regulatory role of TXNIP, resulting in the observed metabolic phenotype. These examples show that by combining information from different types of genetic and disease association studies, it is possible to extract new insights into metabolic pathways relevant to diabetes.

**Future directions: study design and data analysis**

A number of lessons can be learned from the studies cited in this review. Clearly, metabolomics is one of the methods of choice that should be applied whenever the resources and sample availability permit it.
Existing samples from previous and ongoing epidemiological cohort studies provide easy access to extensive phenotype datasets in cross-sectional and prospective designs and have been the prime source for large-scale metabolomics studies on diabetes so far. Drawbacks of these studies are in particular a relatively coarse characterization of the T2D phenotype, many studies lacking for instance data on glucose tolerance tests, not to mention euglycemic clamp tests or more elaborate metabolic challenges. Moreover, in most cross-sectional studies, participants with T2D constitute a very heterogeneous group. For instance, individuals with well-adjusted diabetes are possibly more similar to members of the control group than some potentially pre-diabetic ‘healthy’ study participants. Medical treatment and development of comorbidities vary for every patient. Detailed phenotyping of study participants is hence of utmost importance, where at the same time large participant numbers are required in order to obtain statistically significant association results when testing multiple metabolic traits. In a world of limited resources, this necessarily leads to a trade-off between depth of phenotyping and study size. We shall probably find studies in the future that cover the whole spectrum between large number of participants and exhaustive phenotypic and metabolic characterization, ideally also some combined with physiological challenges. All those studies have their merit and shall complement each other. In particular, we expect to see more extensive metabolic coverage of less commonly studied biological samples, including saliva, stools, tears, sweat, hair, fingernails, and, wherever accessible, tissue from biopsies. This may in particular include combined studies of the gut microbiome (Martin et al. 2007) and studies of patients undergoing bariatric surgery (Friedrich et al. 2012). New prospective studies should in particular attempt to enroll patients that are newly diagnosed with diabetes and yet untreated, in order to be able to investigate individual responses to treatment from the start. A lesson learned from past epidemiological studies is that one should collect and store as many samples of different kinds of biomaterial as possible for future use, and for as many time points as possible, even if at the time of study design neither plans for analysis nor funding for their analysis are available. Wisely built-up biobanks should be able to attract such funding easily at a later stage when samples are readily available in a freezer (Watts 2012).

This review has focused mostly on human studies, but the value of metabolomics studies in cell culture and animal models should not be underestimated. Here again, a combination of deep metabolic phenotyping with large sample numbers, including biological replicates, multiple conditions, and time points, are key to success. In contrast to studies on human samples, animal studies allow for access to a combination of tissues from several organs, while work on cell culture allows for the variation of a single parameter at a time and potentially even investigation of the metabolome of a single cell (Ibanez et al. 2013), possibly limited to specific subcellular organelles. Moreover, although only briefly mentioned in this review, the value of parallel access to complementary large-scale data, such as genome-wide genetic variation, DNA methylation, gene expression, noncoding RNA levels, and proteomics cannot be overestimated. Metabolomics is only one facet of the greater picture that describes a complex disease (albeit one of the most informative ones when it comes to metabolic disorders). One major challenge for the future lies in the integrative interpretation of these datasets, going beyond simple association tests on single phenotypes. New data analysis approaches naturally rely on systems biology methods, including for instance the unbiased analysis of all possible pairs of ratios between metabolite concentrations (Petersen et al. 2012), their connection in Gaussian graphical models (Krumsieck et al. 2011, 2012), and the interpretation of metabolomic profiles using unbiased pathway models (Deo et al. 2010).

Declaration of interest
The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the review.

Funding
K S is supported by ‘Biomedical Research Program’ funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. The statements made herein are solely the responsibility of the author.

Acknowledgements
The author thanks Anna Halama for critical reading of the present manuscript.

References


Group LC 1992 Sulfonylureas in NIDDM. Diabetes Care 15 737–754. (doi:10.2337/diacare.15.6.737)


Kim WJ, Park CY, Lee KB, Park SE, Rhee EJ, Lee WY, Oh KW & Park SW 2012 Serum 1,5-anhydroglucitol concentrations are a reliable index of glycemic control in type 2 diabetes with mild or moderate renal dysfunction. Diabetes Care 35 281–286. (doi:10.2337/dc11-1462)


Watts G 2012 UK Biobank opens it data vaults to researchers. *BMJ* **344** e2459. (doi:10.1136/bmj.e2459)


WTCCC 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447** 661–678. (doi:10.1038/nature05911)


Received in final form 4 March 2014
Accepted 11 March 2014