Metabolomics: Approaches and Applications to Diabetes Research

Anju Preet1, Tejaswita M. Karve1, Nasser Rizk1 and Amrita K. Cheema1*

1Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, USA
2Department of Biochemistry and Cellular and Molecular Biology, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, USA
3Department of Health Sciences, Qatar University, Doha, Qatar

Abstract

Diabetes mellitus is a multi-factorial disease associated with a dysregulated metabolism. A holistic approach emphasizing each of the individual factors with regards to complete pathophysiology of the disease is critical to our understanding of this heterogeneous disease. Several technical advances in the field of functional genomics such as metabolomics and proteomics aid in comprehending the state of the overall biological system and thus can be utilized to decipher the complex interactions among components of the metabolic system in human diabetes. Above-mentioned techniques combined with a set of bioinformatics tools and available databases aim to profile wide array of proteins and metabolites repertoire in humans. The accurate and comprehensive measurements of these molecules is employed to investigate complex interactions of metabolites and proteins not only among themselves but also with genes, transcripts and other small molecules to decipher cellular microenvironment and the effects due to drug intervention. This review provides an overview of the applications of metabolomics in human diabetes research. We also discuss the potential of combining different bioinformatics tools with the omics approach to advance the scientific knowledge towards discovery of biomarkers to monitor and regulate general human health with respect to the deregulated metabolic state, a characteristic of diabetes mellitus.

Keywords: Diabetes; Metabolomics; Gas chromatography; Liquid chromatography; Non-magnetic resonance

Introduction

Diabetes mellitus is a chronic metabolic disease, characterized by elevated blood glucose level–hyperglycemia, due to a deficient insulin secretion by pancreas or an inefficient insulin action in body tissues [1,2]. World Health Organization (WHO) estimated 3.4 million diabetes related deaths in 2004 and projected the incidence of diabetes to double between 2005 and 2030 [3,4]. Four types of diabetes mellitus have been identified: Type1, Type 2, “other specific types” and gestational diabetes [5]. Type 2 Diabetes (T2D) is the most prevalent form of diabetes characterized by decreased insulin sensitivity as opposed to deficient insulin secretion due to autoimmune destruction of pancreatic beta-cells, the accepted causal factor for T1D [2,6]. Persistent hyperglycemia in diabetes leads to metabolic dysfunction and manifests in form of series of complications like retinopathy, neuropathy, nephropathy and cardiovascular disease [7]. Extensive research efforts have been invested in order to understand the metabolic signature of T2D that would augment early detection of the disease and the development of effective therapeutics [8]. Metabolite analysis in body fluids such as blood and urine is routinely practiced to assess diabetes risk [9,10]. Metabolomics is a powerful tool to study the complexities of T2D development and progression. Metabolic phenotype is a reflection of genetic makeup that reflects the changes induced by cellular and external environmental conditions that manifest in the form of diseases of altered metabolism such as T2D. Better understanding of metabolic status in T2D would aid clinical interventions to direct the metabolism in more favorable direction. This review summarizes the applications of metabolomics in diabetes research.

Metabolomics as a Tool for Diabetes Research

Metabolomics is defined as a comprehensive characterization of endogenous metabolites representing the “metabolome”. It provides global analysis of small molecules, which are either substrates or are products of metabolism.

Metabolomics investigates the unique metabolic phenotype or fingerprint that provides a snapshot of all metabolic pathways in an organism at any given time. It is emerging as an important tool for the study of diseases of dysregulated systemic metabolism such as obesity, cardiovascular disease, diabetes mellitus and associated complications, as it focuses on identifying biochemical pathways and their interactive roles within systemic metabolism [11,12]. Metabolome, the quantitative complement of metabolites in a biological system, is traditionally placed at the lowest tier of the biochemical information flow, originating from genome transcriptome that translates to proteome to metabolome (Figure 1).

Metabolome is considered a sensitive indicator of both genetic and environmental perturbations. Metabolic reconstructions suggest that the changes in the metabolome are usually greater than those observed at protein or gene level [13,14]. Each level of a biological system interacts with each other and elicits a characteristic response to intrinsic as well as intrinsic environmental challenges (diet, lifestyle, drug and disease) to determine the resultant phenotype. Metabolomics can be a top down study of a biological system, using a holistic approach involving the study of components and interactions of the complete system. Bottoms-up strategy on the other hand, refers to the study of specific components and interactions within the system. The study of metabolites provides insights into biological processes, facilitating the understanding and manipulation of complex biological systems for diagnostic, prognostic and therapeutic purposes.
Two major approaches are currently used in metabolomics: targeted and untargeted. Targeted studies focus on quantitative measurements of specific metabolites with high precision. These studies involve the use of biochemical and analytical tools for the quantification of known metabolites of biological interest. Targeted studies assist the investigator with hypothesis testing and require the addition/spiking of several stable isotope-labeled standards as appropriate internal standards to ensure accuracy and specificity of quantitation. Targeted, quantitative metabolomics have facilitated the characterization of known as well as novel metabolic changes in experimental diabetic mice as well as diabetes patients [15-17]. Recent studies using a targeted metabolomics approach revealed altered and modified metabolic phenotype in diabetes and drug treatment respectively [16,18,19]. Analysis of dysregulated metabolites associated with diabetes, provide a functional readout of the metabolic state of subject under study and helps identification of candidate markers of metabolic pathways affected by the disease and/or treatment. Pathway specific perturbations of metabolic homeostasis in individuals can help identify patients at high risk and can predict diagnosis and prognosis of the disease state [18,20].

Non-targeted metabolomic profiling on the other hand, does not require prior knowledge and can thus be used to identify novel metabolic biomarkers of disease and drug efficacy besides analyzing the global metabolic profile of the whole system. Comprehensive biochemical profiling using metabolomics has provided insights into the pathophysiological progression of diabetes not only in clinical but also in per-clinical conditions [15]. The metabolome, can be analyzed from different matrices such as serum, urine, cerebrospinal fluid, breath and tissues. Metabolomics studies have revealed significant elevation of certain amino acids and their derivatives in serum that strongly correlate with fasting hemoglobin A1c (HbA1c). HbA1C is a clinically accepted form of hemoglobin that is measured primarily to identify the average plasma glucose concentration over prolonged periods of time [21]. Increased HbA1C values are strong indicators of developing diabetic complications [7]. Urine metabolome analysis across different species in diabetes have identified significant changes in nucleotide metabolism, including that of N-methylnicotinamide and N-methyl-2-pyridone-5-carboxamide, which may provide unique biomarkers for following T2D mellitus progression [17,22,23].

Thus, metabolomics provides a comprehensive snapshot of the disease process and helps investigator to assess the disease status of the subject under study and inform therapeutic decisions. Metabolomics applied to diabetes research, helps to obtain an overview of the disease onset and progression. Several therapeutic targets have been discovered through metabolomics [12,24]. Cross-species mapping of the lipid profile using metabolomics have helped investigators develop models to investigate early disease pathophysiology of diabetes [25,26]. The understanding of lipid profile at a tissue-specific level further facilitates understanding of the etiology of different complications associated with diabetes [25,27].

Comprehensive analytical studies in multi-factorial diseases as T2D have an edge over isolated knowledge of individual components. This approach offers an accurate mechanistic understanding of the complex disease phenotype since integrated behavior of a system is likely to be different from that of a single component [11,12]. Global biochemical studies began in the mid-twentieth century with the use of gas chromatography-mass spectrometry (GC-MS) and were accentuated by the availability of non-magnetic resonance (NMR) spectrometers [28,29]. Eventually Mass spectrometry MS or NMR evolved as analytical instruments of choice to detect metabolic changes for diagnostic purposes. Advancements in computational technologies helped integrate and annotate the endogenous metabolite data quantitated through one of the existing analytical platforms such as MS, NMR and chromatographic systems helping the researchers investigate the effect of integrated metabolism on human health [30].

The intrinsic diversity in chemical structure, size, abundance and reactivity of the pool of metabolites in any biological samples makes it challenging to identify and quantify all metabolites simultaneously in a single, high throughput platform. Majority of the reported studies apply MS or NMR spectroscopy as the analytical instrument of choice [31-34]. However, many other techniques including Fourier transform infrared and Raman spectroscopy have also been used [35-37]. Electrochemical detection for identification and quantification of electrochemically active metabolites from redox pathways has also been reported [38,39]. The use of different analytical platforms provides complimentary information that can be integrated for deeper metabolome coverage [40].

Overview of Metabolomics Work Flow

Typically, metabolomics studies are investigative in nature involving delineation of biomarkers of a disease [6,19,41]. These can further be subdivided as metabolic profiling, using an untargeted approach or metabolite identification and quantitation using a targeted approach. It is critical that the samples from each group under investigation are collected, stored and processed in standardized manner. A combination of analytical methods, 1H NMR spectroscopy and LC-MS have been used to provide information on metabolic pathways known to be altered by insulin deficiency in diabetes [11,15,26,32]. In absence of a single common platform to identify and quantitate all metabolites in the same sample simultaneously, the comprehensive metabolic changes are assembled by consolidating data from different platforms.
Raw data collected from different analytical platforms are preprocessed to remove inconsistencies resulting from instrument performance or sampling. QTOF Mass spectrometry is capable of resolving thousands of molecules in a single experiment and making accurate mass assignments. Greater sensitivity allows for the detection of low abundance metabolites missed by NMR [42,43]. The MS spectral data is further filtered and is de-convoluted. Standard publicly available tools like XCMS, MetaAlign, MZmine and MathDAMP help data pre-processing [44,45]. The conversion of continuous NMR data into a segmented version by bucketing or binning corrects data for peak shifts due to pH or ionic strength variations across different samples as well as reduces the data significantly simplifying the subsequent data analysis [46,47]. Another approach involves de-convolution of NMR spectra into individual components allowing identification and quantitation of individual components from a complex NMR spectrum [48].

The pre-processed metabolomics data can be further analyzed using supervised and unsupervised algorithms. Supervised methods include partial least squares discriminate analysis (PLS-DA), support vector machines (SVM) and discriminant function analysis (DFA) and univariate ANOVA and median fold change (MFC). Unsupervised methods include principal component analysis (PCA), supervised and self-organized maps. Schematic of a typical metabolomics study workflow is illustrated in Figure 2. These tools help the researcher delineate candidate markers that are significantly dysregulated in the experimental data set. A select panel of candidate markers is then further validated using independent cohorts or by repeating the study.

Metabolomics for Diabetes Research

Mass Spectrometry (MS) has been successfully employed to investigate different processes central to diabetes, such as non-enzymatic protein glycation where different hexose sugars would modify the proteins leading to their altered or diminished functions [49] or to obtain metabolic profiles in T2D patients [32,50-53] and for diabetes risk assessment [18]. Due to sensitivity and diverse chemical identification capabilities, MS is the tool of choice for obtaining broad metabolic profiles in conjunction with gas or liquid chromatography [54,55].

Gas chromatography-mass spectroscopy (GC-MS)

GC-MS is the oldest and a robust tool for qualitative metabolic profiling. GC-MS provides high chromatographic resolution and allows for non-targeted profiling for the discovery of novel metabolites and metabolic pathways [56,57]. GC-MS involves electron impact ionization wherein the GC column eluants are introduced into the source, ionized and fragmented to generate a characteristic fragmentation pattern and mass spectrum that is typically used for chemical identification. GC-MS has been extensively used as a discovery tool in steroid characterizations for clinical purposes [56,58,59].

GC-MS has been used for the study of pathways of oxidative stress activated in diabetic macrovascular disease, both in primate and rodent models [60-62]. These studies emphasize the role of oxidized amino acids as potential markers for the assessment of oxidative damage. The chromatographic resolution capacity of the conventional GC has been further enhanced by a more recent technique known as Comprehensive GC X GC-MS that has been applied successfully in metabolomics [63,64]. This technique uses an additional column for two dimensional separation that significantly increases the analytical performance by improving the chromatography thereby expanding metabolome coverage [65,66].

Liquid chromatography-mass spectrometry (LC-MS)

LC-MS involves interfacing of liquid chromatography platforms with mass spectrometers. LC provides metabolite separation by equilibration between a mobile liquid phase and a stationary solid (or liquid) phase. The coupling of liquid systems to mass spectrometry is facilitated by the use of electrospray as the commonly applied ionization technique. Application of LC-MS as a reliable technology has increased during the previous decade [18,67-69]. LC-MS metabolite profiling of 20 non-obese and obese individuals showed a strong correlation between fasting concentrations of branched-chain and aromatic amino acids and serum insulin [70]. A strong correlation has also been reported between branched-chain amino acid (BCAA) catabolism and insulin resistance [71]. LC-MS was used to generate metabolic profiles from 2,422 normo-glycemic individuals followed over a period of twelve years of which 201 eventually developed diabetes [18]. This study reported a panel of more than 60 metabolites including branched-chain and aromatic amino acids as predictors of development of diabetes over the standard risk factors such as fasting glucose, body mass index (BMI) etc.

![Figure 2: Schematic representation of a metabolomics experimental workflow.](image-url)
Capillary electrophoresis- mass spectrometry (CE-MS)

Capillary Electrophoresis- Mass Spectrometry (CE-MS) is yet another analytical tool utilized for metabolite separation and detection. Metabolites are first separated by CE based on charge and size, and then selectively detected using MS by monitoring a large range of m/z values. CE is particularly suited for the separation of polar and charged compounds and can provide complementary information to LC-MS on the biological composition of sample [72]. It has been successfully used in different studies to detect and quantitate cationic and anionic metabolites not only across different species but also across different sample types as bio-fluids, cells and tissues [73-78]. Cross-platform analysis utilizing CE-MS fingerprinting augmented the identification of metabolites galactosylhydroxylysine, l-carnitine, among others which markedly increased in urine from diabetic rats as compared to control animals [79].

Nuclear magnetic resonance spectroscopy (NMR)

NMR spectroscopy is a quantitative, highly reproducible and non-selective analytical technique for metabolic profiling [80]. It is independent of the hydrophobicity or pKa of the compounds being analyzed. It has been extensively used for metabolic profiling for more than 20 years. This technique interrogates all the molecules present in the sample simultaneously by using the active NMR of hydrogen (1H) or carbon (13C)-the so called common magnetic nuclei [81]. The qualitative limitation of NMR lies in its inherent insensitivity and is hence suitable only for detection and quantification of metabolites present in relatively high concentration [12]. Alternative strategies are being currently developed to increase the sensitivity of NMR including the use of cryoprobes in improving signal to noise for 13C NMR based metabolomics [82-84]. Another area of improvement is the use of hyperpolarized substrates to selectively enhance the resonance of key metabolites.

Low circulating levels of plasma phosphatidylcholine and high levels of methylamines were detected in plasma and urine samples from 129S6 mice, a mouse strain known to be susceptible to hepatic steatosis and insulin resistance in comparison to BALBc (relative resistance) using 1H NMR-based metabolic profiling [85]. In a recent study, assessment of biochemical process of diabetes has been done utilizing quantitative 1H NMR-based metabonomics to analyze urine, serum, and liver extracts from streptozotocin-induced diabetic rats [86]. This study identified a number of metabolic alterations in liver samples from diabetic rats including metabolites participating in nitrogen and carbon metabolism. Salek et al used 1H NMR to compare metabolic alterations not only in diabetic animal models but also in humans [22]. The study involved NMR based urine analysis from 12 healthy and 30 T2D patients. A clear group separation based on a large number of metabolites that included amino acids such as alanine, ornithine, etc. was observed. Although a robust technique, the detection limits of 1H NMR is compromised by the large number of co-resonances that may be somewhat improved by use of 2-dimensional NMR or with use of nuclei, such as 13C that have more dispersion.

Animal Models of Diabetes Mellitus

Animal models, including but not limited to rabbits, dogs, monkeys and various murine species (e.g. rat and mice), serve a critical function in understanding of the pathophysiology, early embryonic clues of disease predisposition(s) and genetic basis of the disease as well as developing therapeutic (such as drug efficacy and toxicity) and/ or preventive strategies and tracking disease prognosis e.g. rhesus monkey models have been used to understand T1D as well as developing insulin administration strategies as a medical intervention to alleviate T1D [87]. Given the multi-faceted nature of diabetes mellitus, there are numerous murine models representing individual factors shown to be responsible for developing T1D or T2D. These murine models can be grouped as spontaneous or genetic models, diet or nutritional induced models, environmental or chemically induced models, surgically induced models and transgenic or knockout models [88,89]. Even though, none of the rodent models may accurately correlate with the human disease pathology for diabetes singularly, since most of these animals display an array of symptoms that resemble the human disease, they can be manipulated for studying the impact of a particular component such as genetics or environmental effect(s) [89,90].

Spontaneous murine models, known since early 1980’s, are available for both T1D and T2D, and are helpful in understanding the genetics of the disease, including consequences of inbreeding, insulin resistance mainly due to glucose toxicity, ketosis, obesity and hyperinsulinemia [89,91]. Some of the most commonly used murine models for genetic studies of diabetes are NOD (non-obese diabetic) mice, KK (Kuondo) mice and BB (bio breeding) rats [88,91]. Mouse models such as NOD-Pdcd1-/- (programmed cell death 1 [PD-1, Pdcd1], an immune-inhibitory receptor from the CD28/cytotoxic T lymphocyte-associated antigen-4 family), have been further developed to specifically study T1D [92]. A recent study illustrating urine profile in a spontaneous non-human primate T2D model, Rhesus monkey (Rhesus macaques), was able to detect a defective Na (+)-dependent transporter, SLC6A20, in proximal tubules of kidneys [93]. Further, this study confirmed similar observation in db/db mouse model and thus reflects the basic functional changes at the cellular level in a disease state of T2DM.

Diet or nutritionally-induced models play an important role in gaining insights specifically concerning T2D, which is mainly linked with the obesity and development of insulin resistance leading to glucose toxicity. As per the information by the American Diabetic Association, T2D is much more common in ethnic minorities as well as other non-white communities such as Asian Americans in the United States and is linked mainly to the diet. A diet high in saturated fats and poor in nutritional value is a major contributing element to obesity and in turn developing T2D. Number of studies using animal models have highlighted the role of factors apart from high-fat diet that may influence the predisposition of these populations for developing T2D; source of protein in diet [94]. In these studies, rats that were on diets high in fats and had most of their protein share from casein-based food or soy products showed insulin resistance while the group of animals with high fat diet and most of its proteins coming from cod (fish-based) showed no insulin resistance [94]. Thus, these studies suggest that the overall food pyramid distribution and sources of each particular food groups influence the outcome, at least, in predicting the predisposition of a population for developing T2D.

Another subtype of murine model, environmental or chemical-induced rodent model, has been useful in various studies focused on disorders related to diabetes. One of the most commonly used chemical-induced diabetic murine models is streptozotocin (STZ) induced diabetic rat [88,95]. Diabetes, a heterogeneous disease, is known to affect a number of normal functions in various human organs. One of the common disorders observed in diabetic men is diabetes-related erectile dysfunction (ED), which has been widely studied in chemical-induced murine models [95,96]. Maggi et al reported that STZ-diabetic rats suffered from the hypogonadism, a condition frequently observed in diabetics, along with low testosterone production and atrophy in a
number of androgen-dependent accessory glands as well as induction of RhoA/Rho-kinase (ROCK) signaling pathway [95,96]. Further, the authors found that normalizing the testosterone production along with the introduction of ROCK inhibitors showed significant improvement in alleviating the ED in these animals and may thus have a promising potential for clinical implications for the patients suffering from diabetes-related ED.

A different type of animal model, a surgically-induced model in mice, pigs, dogs and rats, has been helpful in studying diabetic-related retinopathy and the role of pancreas in T1D and T2D [97]. The surgical models are generally developed by complete or partial removal of pancreas known as pancreatectomy, He, et al. have shown that complete pancreatectomy is one of the two factors in diabetic monkeys that can lead to severe hypoglycemic conditions [98]. Further, combining the pancreatectomy along with the other diabetic murine models such as spontaneous rat model has led to some interesting findings. Plachot, et al. found that even partial pancreatectomy in GK (Goto-Kakizaki) rats, a commonly used spontaneous rat model to study T2D further accelerates the disease initiation by reducing beta-cell proliferation and insulin secretion, a critical element in diabetes [99]. Hyperglycemia, responsible for the loss or reduction of beta-cell proliferation has also been connected to the expression of c-myc, a known oncogene, and other transcription factors in these murine models and thus, may indirectly control the insulin production in these animals [100].

Transgenic or knockout murine models are other common tools that are used to study the role of specific gene(s) associated with a disease such as neurodegenerative disorder, cancer and many other genetic diseases. Most of the transgenic rodent models for diabetes, T1D and T2D, are associated with single or double knockout of various genes in the insulin production pathway [101]. However, due to number of components in a insulin synthesis pathway and resultant insulin resistance in diabetes, transgenic models have been disappointing to this end [101]. Nevertheless, transgenic or knockout models are useful when used in conjunction with targeted molecular and biochemical studies. Transgenic mouse models are also helpful in studying diabetes in conjunction with other obesity-related metabolic disorders/ syndromes such as atherosclerosis, dyslipidemia and insulin resistance and thus are frequently used in metabolomics studies, specifically LC-MS-based lipidomics and liver and blood profiling in these animal models [90,102,103]. At least one MS-based lipidomics study found that the stimulation of PPARγ (peroxisome proliferation activated receptors) by an agonist rosiglitazone as detected in blood plasma of obese T2D mouse model and thus has a promising potential as a target for drug therapy in T2D [102,104].

Future Directions/ Challenges

Given the recently released predictions about escalating prevalence of diabetes worldwide by the WHO, it is extremely critical to integrate biochemical, molecular and clinical knowledge to design diabetes treatment and preventive strategies that are affordable as well as effective. The modern improvements in technology such as advances in the multi-platform and highthroughput omics; transcriptomics, proteomics and metabolomics, and vast bioinformatics toolset, as well as advanced knowledge acquired by molecular and biochemical studies are certainly helpful. However, the immense data obtained from the “omics” need to be interpreted with a great caution; giving due consideration to the limitations of animal models and the statistical tools used. Results should thus be interpreted in conjunction with experimental evidence from biochemical and molecular studies. Currently metabolomics approaches lack standardized procedures for sample preparation, data analysis and interpretation. Thus, there is a critical need to develop common benchmarks describing the experimental set-up and ontology for metabolomics similar to the universal standards available for other “omics” such as transcriptomics e.g. MIAME (minimum information associated with a microarray experiment) [105] and proteomics e.g. MIAPE (minimum information about a proteomics experiment) [106]. Additionally, there is a lack of comprehensively annotated metabolite databases for unambiguous metabolite identification [107].

Similar caution is required in interpreting data obtained from the animal models since most animal models do not entirely complement various disease states as found in T1D and T2D in humans. The use of non-human primates closest to humans such as monkeys can be successfully substituted for a better understanding of T1D and T2D. However, the usage of primates requires complex animal care protocols and is not cost-effective in a large experimental set-up. Additionally, since they have a longer life span the interpretation of results can be delayed and complex.

Nonetheless, the data obtained from integrating various “omics” studies, may be useful in discovering molecular markers and/ or drug targets that can be of clinical significance for developing therapeutic interventions as well as for gaining mechanistic insights into disease onset and progression. In addition, these and other studies may also have implications in unraveling genetic markers predicting individual’s predisposition for developing diabetes and thus can be valuable in formulating diagnostic tests for early detection of diabetes.

Acknowledgements

Dr. Cheema and Rizk are also supported in part, by the Qatar National Research Foundation grant (NPRP 06-740-3-148).

References

inflammation and arterial stiffness in men with newly diagnosed type 2 Diabetes. Clin Endocrinol: 1365-2285.


