THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GC-MS based metabolomics –

Development of a next generation GC-MS method and its application to nutrition and biomarker research

OTTO SAVOLAINEN

Food and Nutrition Science
Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017

GC-MS based metabolomics - Development of a next generation GC-MS method and its application to nutrition and biomarker research

Otto Savolainen

ISBN: 978-91-7597-531-3

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie (ISSN 0346-718X)

Chalmers University of Technology
Department of Biology and Biological Engineering
Food and Nutrition Science
SE-412 96 GÖTEBORG

Printed by Chalmers Reproservice
Gothenburg, Sweden 2017

Front cover by Dimitra Lappa

GC-MS based metabolomics-
Development of a next generation GC-MS method and its application to nutrition and biomarker research

OTTO SAVOLAINEN
Department of Biology and Biological Engineering
Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

Metabolomics, the measurement of a broad range of small molecules in a sample, is maturing as a field in analytical chemistry and becoming a standard research tool in biological sciences for helping to generate a wider understanding of the complex biological mechanisms behind, for example, development of disease and effects of diet on health. Metabolomics analysis is generally divided into either targeted or untargeted metabolomics. The former 'targets' a specific set of molecules, often quantitatively, while the latter aims to detect as many metabolites as possible in the sample. Both of these approaches have their advantages and disadvantages, and combining both into one single analytical method could allow the specificity, sensitivity, and quantitation of targeted metabolomics to be combined with the broad coverage and potential to find unknown compounds that are the key advantages of untargeted metabolomics.

In this thesis, a novel gas chromatography triple quadrupole metabolomics (GC-MS) method that exploits fast data acquisition to simultaneously acquire both targeted quantitative data using multiple reaction and selected ion monitoring, and untargeted qualitative data using full spectrum scanning. The method was developed for human blood plasma and applied to two studies. The first was a crossover intervention study comparing metabolic effects of consumption of herring with chicken and pork in 15 adults. The second was a prospective cohort of 600 older Swedish women for discovering predictive biomarkers of development of type 2 diabetes (T2D) and for exploration of the associations between potential dietary and nutrient biomarkers and glucose tolerance status and development of T2D in the same cohort.

The evaluated method parameters (linearity, limit of detection and quantification, precision, accuracy, number of spectral features) supports the approach of combining targeted quantitative and untargeted qualitative data acquisition as a way to improve GC-MS metabolomics. The method was successfully applied to the analysis of 1200 samples from 600 subjects, measured

over 24 separate analytical batches with an average within batch metabolite variation of 10 %, based on control sample metabolites detected using multiple reaction monitoring.

In the intervention study, the new metabolomics method detected 190 identified metabolites, of which 18 were altered when subjects ate herring instead of chicken and pork. These changes were mainly around the tricarboxylic acid and urea cycles, with an apparent differential effect related to arginine metabolism. This finding was supported by finding that circulating nitric oxide was higher in male subjects after the herring diet compared to the chicken and pork diets. In the cohort study, we established that the best predictive markers of T2D detected using metabolomics improved on or had similar prediction to established predictors of T2D. Using a combination of both the targeted and untargeted data, we were also able to detect 10 dietary and nutrient biomarkers, many of which were strongly associated with glucose tolerance status and development of T2D in this cohort. This is one of the first studies using multiple dietary and nutrient biomarkers that suggests a clear role of diet in prevention of T2D. This supports the current dietary guidelines for eating whole grains and fish for preventing T2D.

In conclusion, the metabolomics method developed as part of this thesis detects a wide array of known biomarkers in blood plasma while still providing global untargeted information and having the possibility of expansion by addition of targeted molecules of interest. The method proved to be robust during the analysis of a moderately sized sample set, supporting its further use in observational cohorts. Results from the application of this metabolomics method further support the potential of metabolomics to add value to biological research by highlighting diverse outcome-metabolite relationships that may otherwise be overlooked.

LIST OF PUBLICATIONS

This doctoral thesis is based on the work contained in 5 papers:

I. Otto Savolainen, Ann-Sofie Sandberg and Alastair Ross. **A simultaneous metabolic profiling and quantitative multimetabolite metabolomic method for human plasma using gas-chromatography tandem mass spectrometry.** Journal of Proteome Research, 2016, 15(1), 259-265.

II. Andrew Vincent, Otto Savolainen, Partho Sen, Nils-Gunnar Carlsson, Annette Almgren, Helen Lindqvist, Mads Vendelbo Lind, Ingrid Undeland, Ann-Sofie Sandberg and Alastair Ross. **Herring and chicken/pork meals lead to differences in plasma levels of TCA intermediates and arginine metabolites in overweight and obese men and women.** Molecular Nutrition & Food Research. 00, 1-9, 2016

III. Mads Vendelbo Lind, Otto Savolainen and Alastair Ross. **The use of mass spectrometry for analyzing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large number of biological samples.** European Journal of Epidemiology, 2016, 31, 717.

IV. Otto Savolainen, Björn Fagerberg, Mads Vendelbo Lind, Ann-Sofie Sandberg, Alastair Ross and Göran Bergström. **Biomarkers for predicting type 2 diabetes development – can metabolomics improve on existing biomarkers?** *Submitted*

V. Otto Savolainen, Mads Vendelbo Lind, Göran Bergström, Björn Fagerberg, Ann-Sofie Sandberg and Alastair Ross. **Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women.** *Submitted*

Associated papers not included in the doctoral thesis:

I. Otto Savolainen, Rossana Coda, Katja Suomi, Kati Katina, Riikka Juvonen, Kati Hanhineva and Kaisa Poutanen. **The role of oxygen in the liquid fermentation of wheat bran.** Food Chemistry, 153:424-431, 2014.

II. Jenna Pekkinen, Natalia N Rosa, Otto Savolainen, Pekka Keski-Rahkonen, Hannu Mykkänen, Kaisa Poutanen, Valérie Micard and Kati Hanhineva. **Disintegration of wheat aleurone structure has an impact on the bioavailability of phenolic compounds and other phytochemicals as evidenced by altered urinary metabolite profile of diet-induced obese mice.** Nutrition & Metabolism, 11(1):1, 2014.

III. Kati Hanhineva, Pekka Keski-Rahkonen, Jenni Lappi, Kati Katina, Jenna Pekkinen, Otto Savolainen, Oskari Timonen, Jussi Paananen, Hannu Mykkänen and Kaisa Poutanen. **The Postprandial Plasma Rye Fingerprint Includes Benzoxazinoid-Derived Phenylacetamide Sulfates.** Journal of Nutrition, 144(7), 2014.

IV. Isabel Bondia-Pons, Otto Savolainen, Riitta Törrönen, Martinez Alfredo, Kaisa Poutanen and Kati Hanhineva. **Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry.** Food Research International, 63, 2014.

V. Otto Savolainen, Jenna Pekkinen, Kati Katina, Kaisa Poutanen and Kati Hanhineva. **Glycosylated benzoxazinoids are degraded during fermentation of wheat bran.** Journal of Agricultural and Food Chemistry, 63(25), 2015.

VI. Nikul Soni, Alastair Ross, Nathalie Scheers, Otto Savolainen, Intawat Nookaew, Britt Gabrielsson and Ann-Sofie Sandberg. **Eicosapentaenoic acid and docosahexaenoic acid-enriched high fat diet delays skeletal muscle degradation in mice.** Nutrients, 8(9):543, 2016.

VII. Alastair Ross, Cecilia Svelander, Otto Savolainen, Mads Vendelbo Lind, John Kirwan, Isabelle Breton, Jean-Philippe Godin and Ann-Sofie Sandberg. **A high throughput method for LC-MS/MS determination of plasma alkylresorcinols, biomarkers of whole grain wheat and rye intake.** Analytical Biochemistry, 499, 1-7, 2016

VIII. Nikul Soni, Alastair Ross, Nathalie Scheers, Otto Savolainen, Intawat Nookaew, Britt Gabrielsson and Ann-Sofie Sandberg. **Splenic Immune Response Is Down-Regulated in C57BL/6J Mice Fed Eicosapentaenoic Acid and Docosahexaenoic Acid Enriched High Fat Diet.** Nutrients, 9(1):50, 2017.

IX. Alastair Ross, Cecillia Svelander, Göran Karlsson and Otto Savolainen. **Identification and quantification of even and odd chained 5-n alkylresorcinols, branched chain-alkylresorcinols and methylalkylresorcinols in Quinoa (*Chenopodium quinoa*).** Food Chemistry. 220, 344-351, 2017.

X. Katharina Dihm, Mads Vendelbo Lind, Henrik Súnden, Alastair Ross and Otto Savolainen. **Quantification of benzoxazinoids and their metabolites in Nordic breads**. Food Chemistry, *Under revision*

CONTRIBUTION REPORT

Paper I: Otto Savolainen (OS), was involved in the design of the study, performed the method development and evaluation and was responsible for writing the manuscript.

Paper II: OS planned and performed the metabolomics analysis, supervised metabolomics data processing and part of the statistical analysis and was involved in the writing of the manuscript.

Paper III: OS was involved in the planning and writing of the manuscript

Paper IV: OS was involved in the design of the metabolomics study, planned and performed the metabolomics analyses, performed the statistical analyses, led the interpretation of the data and was responsible for writing the manuscript.

Paper V: OS was involved in the design of the biomarker association study, planned and performed the metabolomics analyses, performed the statistical analyses, led the interpretation of the data and was responsible for writing the manuscript.

ABBREVIATIONS

1D	One Dimensional
2D	Two Dimensional
ANCOVA	Analysis of Covariance
APCI	Atmospheric Pressure Chemical Ionization
AR	Alkylresorcinol
AROC	Area Under the Receiver Operator Characteristics Curve
BCAA	Branched Chain Amino Acid
BMI	Body Mass Index
CI	Chemical Ionization
CID	Collision Induced Dissociation
CMPF	3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid
CVD	Cardiovascular Disease
DHA	Docosahexaenoic acid
EI	Electron Impact Ionization
EPA	Eicosapentaenoic acid
EPIC	European Prospective Investigation into Cancer and Nutrition
ESI	Electrospray Ionization
FDA	United States Food and Drug Administration
FDR	False Discovery Rate
FFQ	Food Frequency Questionnaire
FHD	Family History of Diabetes
FINRISC	Cardiovascular Risk in Young Finns study
GC	Gas Chromatography
GTS	Glucose Tolerance Status
HDL	High Density Lipoprotein
HOMA-IR	Homeostatic Model Assessment-Insulin Resistance
HbA_{1c}	Glycated Haemoglobin
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IR	Insulin Resistance
KORA	Cooperative Health Research in the Region Augsburg
LC	Liquid Chromatography

LDL	Low Density Lipoprotein
LLE	Liquid Liquid Extration
LOD	Limit Of Detection
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
NGT	Normal Glucose Tolerance
NHANES	National Health and Nutrition Examination Survey
NMR	Nuclear Magnetic Resonance Spectroscopy
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
PCA	Principal Component Analysis
PLS-DA	Partial Least Square Discriminant Analysis
PPT	Protein Precipitation
RI	Retention Index
RISC	Relationship Between Insulin Sensitivity and Cardiovascular Disease
SIM	Selected Ion Monitoring
SPE	Solid Phase Extraction
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TOF	Time Of Flight
WG	Whole Grain

TABLE OF CONTENTS

INTRODUCTION

Biological systems such as cells, tissues and organisms are composed of an interactive network of genes, proteins and metabolites. Understanding how different parts of the networks interact and function is a fundamental aspect of biological sciences and an increasingly important component of understanding disease risk and the effects of lifestyle factors such as diet on human health. This 'systems' approach to biology requires techniques and methods that allow the measurement of many parameters simultaneously, including genomics, transcriptomics, proteomics and metabolomics, which together can provide broad, yet detailed information on biological mechanisms and function.

Metabolomics, the measurement of a wide range of small molecules in a sample is rapidly maturing to the point where it can be readily applied in many studies in a routine manner. The hundreds or thousands of molecules measured by metabolomics offer a possibility to get an overview of the metabolic composition and status of a sample making it a valuable tool for a wide range of fields in the biological sciences. Metabolomics has commonly been used to discover mechanisms behind biological phenomena, and is now frequently being applied to other areas including discovery of biomarkers.

Biomarkers, indicators of biological phenomena such as disease, infection or environmental exposure, are closely linked to metabolomics due to the possibility to discover potential biomarkers and measure many known biomarkers. For nutrition research metabolomics is a new tool for discovering dietary biomarkers that can potentially reflect dietary intake and for finding associations between diet and diet-related diseases such as type 2 diabetes (T2D). To date methods for measuring dietary biomarkers have generally focused on the analysis of a single or a class of dietary biomarkers. The use of metabolomics, however, offers the possibility to measure multiple compound classes simultaneously enabling multiple potential dietary biomarkers to be analysed at the same time. Although the number of multi-biomarker methods based on metabolomics approaches is increasing, this concept is yet to be applied to dietary biomarkers. Another application of metabolomics in nutrition research is to understand how food, a highly complex mixture of thousands of different compounds, interacts with our body and its own 'metabolome', and how different food sources can influence our overall metabolism. Over the past decade metabolomics applied to nutrition has been used to probe the effects of dietary patterns on metabolic health, yet there are still many basic nutritional questions that are yet to be investigated using metabolomics.

In the field of T2D research, there is a need for tools that can accurately predict people at early-stage risk for the disease. Metabolomics has been applied to many studies related to T2D, leading to the discovery of an association between development of T2D and branched chain amino acids. This has led to potential new risk markers for T2D, as well as better mechanistic understanding of the disease. Further research on T2D using metabolomics is still needed as reliable markers for detecting all people that will develop T2D do not yet exist. Investigating a wide variety of subject

groups using different kinds of metabolomics methods is one way towards reliable risk predictors of T2D.

As with all fields of science, there is always a cycle of continuous improvement as new technologies become available. Metabolomics is no exception and rapid advancements in the sensitivity and robustness of mass spectrometers enables new approaches to metabolomics that can improve on the number and type of metabolites that can be reliably detected. Recent advances in quadrupole mass spectrometers have opened new possibilities to metabolomics data acquisition and research outlined in this thesis takes advantage of a latest generation instrumentation and discusses the development of a novel metabolomics method for human blood plasma and its application to finding metabolic effects of replacing dietary chicken and pork with herring, to prediction of T2D development and lastly to exploring the associations between potential dietary and nutrient biomarkers and T2D.

OBJECTIVES

The overall aim of this thesis was to develop a new method utilizing fast scanning gas chromatography coupled to tandem mass spectrometry (GC-MS/MS) instrumentation for simultaneous targeted quantitative and untargeted qualitative metabolomics. The aim was further to apply this method to finding metabolic effects of fish and meat in an intervention study, and to the prediction of future development of type 2 diabetes based on both a broad range of metabolites and potential dietary and nutrient biomarkers as well as to explore the associations between glucose tolerance status and the potential dietary and nutrient biomarkers.

The specific objectives of the research of this thesis are:

- To develop a robust high-throughput simultaneous targeted quantitative and untargeted qualitative GC-MS/MS metabolomics method for human blood plasma

- To apply the GC-MS/MS metabolomics method to finding the metabolic effects of replacing chicken and pork with herring in obese and overweight men and women in a crossover intervention trial.

- To apply the GC-MS/MS metabolomics method for finding biomarkers that can predict future development of type 2 diabetes in a cohort of older Swedish women and to evaluate if such biomarkers can improve on prediction based on already established predictors of T2D

- To search the generated GC-MS/MS metabolomics data for potential dietary and nutrient biomarkers and to test if these were associated with and could predict glucose tolerance status and development of T2D in a cohort of older Swedish women.

LITERATURE OVERVIEW

Metabolomics

Metabolomics is a field of applied analytical chemistry that aims to measure a wide range of metabolites using sensitive instrumentation combined with chemometric and statistical tools. The use of metabolomics has been growing rapidly during the 21st century with almost no publications on metabolomics in 2000, to over 3300 in 2016 (**Figure 1**). Metabolomics provides complementary information to the other ´omics´ technologies (e.g. genomics, transcriptomics and proteomics) and together can contribute to the understanding of biology of an organism. As the name suggests, metabolomics focuses on analysing the metabolome of an organism or sample. The term metabolome was first introduced in 1998 (1,2) and refers to the set of small molecules, metabolites, generally defined as <1500 Da found in a sample, where the sample can be anything from a cell organelle to a complete organism. Currently, there is no single analytical method that can analyse the whole metabolome of an organism or even a subset such as the serum metabolome, estimated to contain approximately 4200 metabolites (3). Nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) instruments are the two major instrumentation types used in metabolomics (2) and allow coverage of a wide proportion of metabolome.

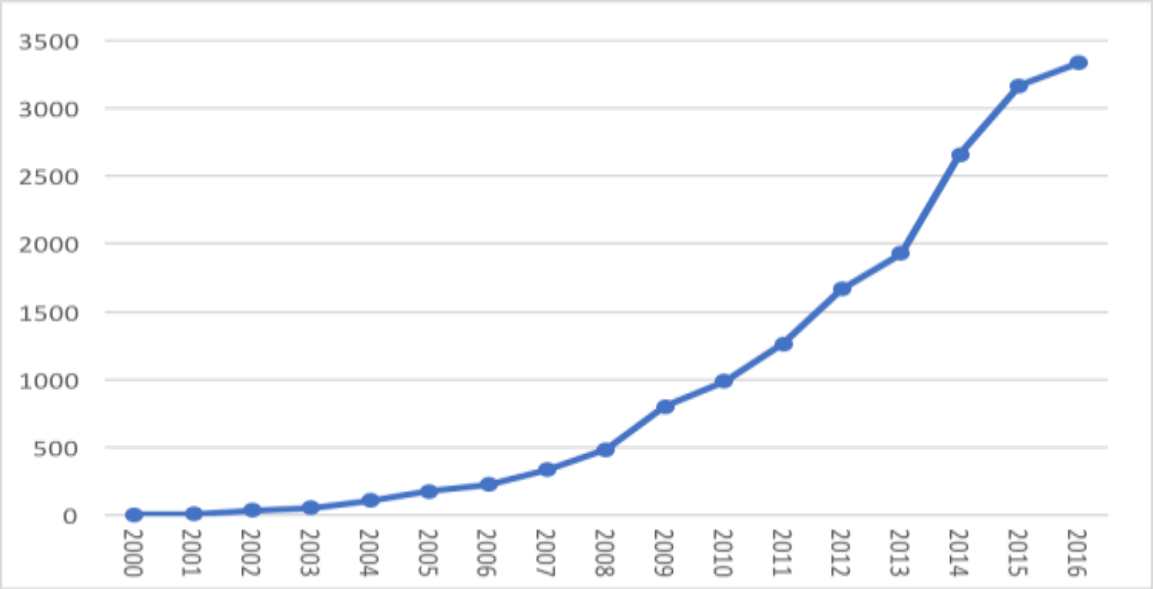

Figure 1. Number of publications listed in the Pubmed database found by using search term "metabolomics" or "metabonomics"

There are many aspects to metabolomics, from study design through to data analysis. The goal of this literature overview is to not give a comprehensive review of all aspects of metabolomics, and instruments which are not covered by the methods used in this thesis (e.g. NMR and LC-MS), as well as standardised protocols that have not been developed as part of this work (e.g. plasma

extraction) are not discussed in detail. In addition we have focused on applied examples for diet and type 2 diabetes as these are related to the work carried out in this thesis.

Mass spectrometry

Mass spectrometry (MS) based methods for measuring the metabolome have gained popularity due to their selectivity and sensitivity (2). The development of MS instruments during the 21st century has improved their usability and robustness for both assays requiring extreme sensitivity and for assays demanding detection of a wide range of different biomolecules at various concentrations in different biological matrices (1,4). Also, the relative cost of MS measurements has decreased making them more accessible to more laboratories. A review on principles of mass spectrometry is given in Paper III and this literature overview focuses on gas chromatography tandem mass spectrometry (GC-MS/MS) and its application to metabolomics.

MS is both sensitive and partially selective (the ability to detect a specific molecule), and new developments in the field often focus on improving selectivity and sensitivity. In metabolomics this has practical implications on several aspects including how much sample is needed and what compounds can be reliably detected. Further selectivity can be achieved with the utility of multiple reaction monitoring (MRM) using collision induced dissociation (CID) where a precursor ion is first selected in one quadrupole, this ion is then fragmented in a collision cell followed by a selection of a specific fragment ion with a second quadrupole (**Figure 2**). This can also improve sensitivity by reducing noise generated by other molecules entering the MS simultaneously. Fragmentation also aids in molecule identification, as molecular fragmentation patterns are consistent from instrument to instrument under set conditions. For GC-MS, this allows the use of 'mass spectral libraries' which contain the fragmentation pattern of thousands of molecules, and can be used to help identify unknown molecules.

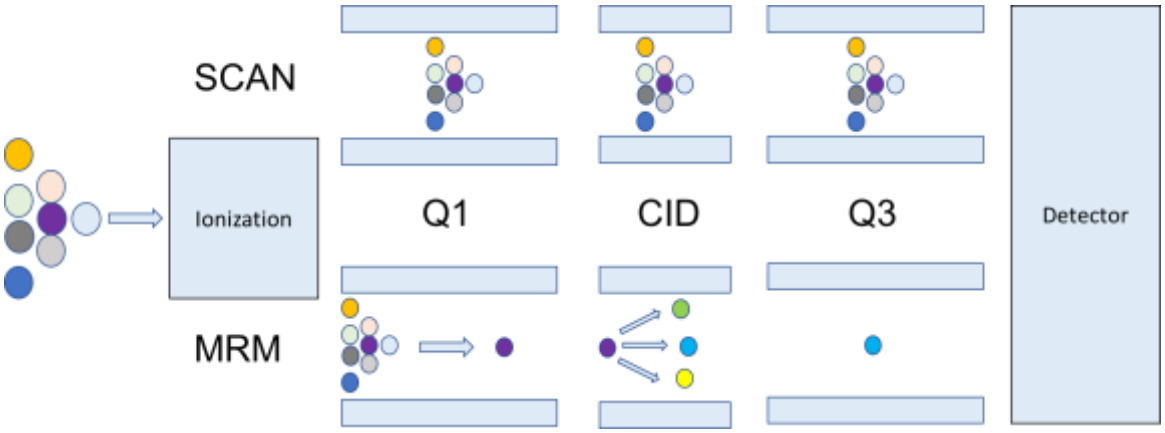

Figure 2. *A schematic illustration of scan and MRM data acquisition using a triple quadrupole mass spectrometer. In scan mode, ionised molecules within a pre-set scan range (e.g. m/z 50-700) pass through quadrupole 1 (Q1), collision cell (CID) and quadrupole 3 (Q3) and all ions within this range are detect. In MRM data acquisition, Q1 is adjusted so that only one molecule or mass (the 'precursor ion') passes through to the CID. The precursor ion is fragmented in the CID, and only one of the mass fragment ions is selected to pass through Q3 to the detector.*

In metabolomics MS is used for both targeted quantitative measurements as well as for untargeted profiling experiments (2) both utilizing different types of mass analysers. Some of the relative merits of each type of mass analyser are shown in **Figure 3**.

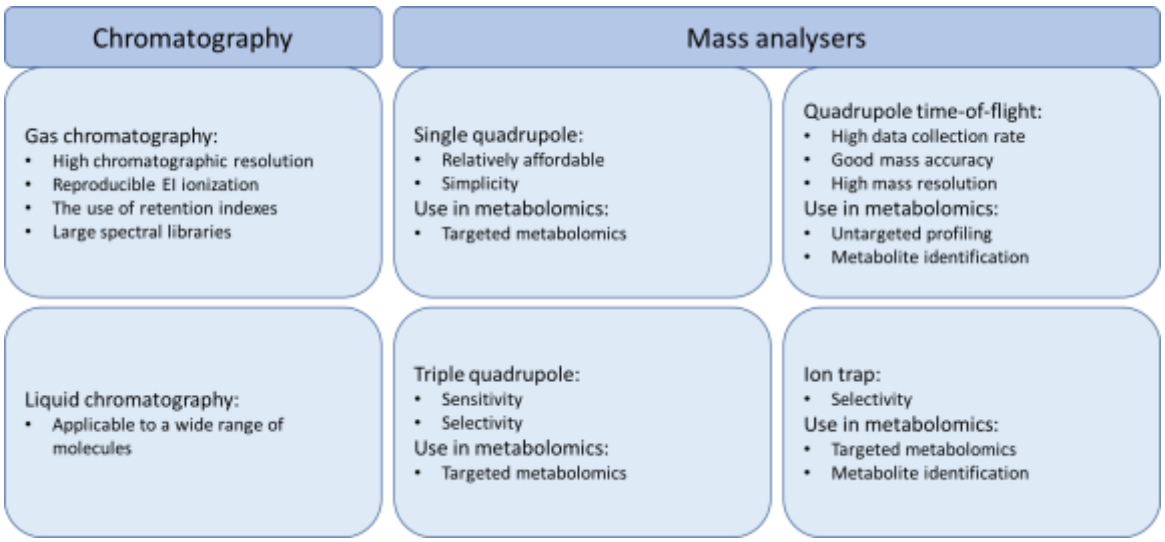

Figure 3. *Relative advantages of gas- and liquid chromatography as well as different types of mass analysers in metabolomics (Paper II)*

Gas chromatography-mass spectrometry

Due to the complex nature of biological samples, some separation of the molecules in a sample is before they enter the MS is preferable. In general, there are two major separation techniques used in metabolomics: GC (5) and LC (6), although other techniques, such as capillary electrophoresis (7) are also used. The basic purpose of both liquid- and gas chromatography in metabolomics analyses is to separate molecules based on their physicochemical properties making detection of the molecules with MS easier by reducing the number of molecules entering the MS at the same time (8). Some of the relative merits of both gas- and liquid chromatography in metabolomics are outlined in **Figure 3**, though a comprehensive comparison of the different MS-based metabolomics methods is outside the scope of this literature overview.

GC in connection with MS is a robust separation technique that has long been applied in metabolomics analyses. GC enables separation of molecules that are or can be made volatile by chemical derivatization and has been applied for measurement of different targeted sets of biomolecules, such as amino- and organic acids (9) and steroids (10) and also for measurement of untargeted metabolomics (11). Prior analysis by GC sample analytes need to be volatile or made volatile by adding a derivatization step into the sample preparation process (12). The need for derivatisation also stops larger non-volatile molecules from being analysed by GC. MS instruments used in GC based metabolomics most often include either single quadrupole (13), triple quadrupole (14), or unit mass time-of-flight instruments (15-18). More recently accurate mass instruments have also been used for GC-MS metabolomics (19,20).

The use of two-dimensional GC-MS (2D-GC-MS) in metabolomics increases the chromatographic resolution over the more traditional one dimensional (1D) approach and provides additional selectivity over the 1D approaches. In a 2D-GC-MS the samples are separated by a combination of two chromatographic columns connected to each other with a modulator enabling separation of the molecules, that could not be separated in the first dimension (first column), in the second dimension (second column), thus, allowing chromatographic separation of otherwise unresolved molecules (5,21). TOF or fast quadrupole MS detectors are preferred in connection with 2D-GC-MS for detection of the typically narrow chromatographic peaks that are further focused in the modulator (22).

Compared to LC-MS there have been limited method developments or instrumental advances in GC-MS metabolomics during recent years. Accurate mass GC-MS has only recently been used for metabolomics (23), while equivalent instruments for LC-MS have been used for metabolomics for over a decade. 2D-GC-MS has been the basis of excellent results (24-27), but its use has not become widespread. Also, the use atmospheric pressure chemical ionization (APCI) in GC-MS metabolomics has been studied (19,20) but its general use is still limited. The added selectivity and sensitivity of GC-MS/MS instruments have been used in targeted metabolomics (14,28,29), though <100 metabolites are detected with such methods. GC-MS and GC-TOFMS systems remain the mainstay of GC-MS metabolomics. However, the recent advances enabling a GC-MS instrument to scan at 20 000 Da/s enables not only the use of this type of quadrupole detector for 2D-GC-MS (30), but in the context of GC-MS/MS, could allow simultaneous collection of targeted MRM and untargeted scanning data. This would allow collection of both targeted and untargeted metabolomics data in the same run, which could allow the best of both worlds of metabolomics.

Targeted metabolomics

Targeted metabolomics is an extension of standard way of measuring small molecules in a sample. Standard targeted methods are primarily used for the testing of a predefined hypothesis and, thus the basic requirement for the method in this case is that it needs to be able to measure a compound with adequate accuracy and precision. Targeted metabolomics can make use of the same analytical methods, but the analysis is done without any predefined hypothesis. Advances in analytical methods and instruments, and mass spectrometry in particular, has allowed the quantification of many different metabolites in the same run, blurring the line between targeted methods and metabolomics. Importantly targeted measures can be quantitative which makes it easier to compare results between studies, something which is currently not reliably done in untargeted metabolomics. Over the past decade commercially available targeted metabolomics method packages have entered the market enabling easier access to targeted metabolomics workflows and the solutions have been used successfully in biological research, including work aiming to discover markers of type 2 diabetes (31,32).

Sample preparation may differ for targeted metabolomics, depending on the compound classes of interest, although methods that have wide coverage often use similar methods to those used in untargeted metabolomics, based on protein precipitation. MS instruments used in targeted metabolomics assays are most often triple quadrupole instruments with either gas- or liquid

chromatography hyphenation (2). The earlier mentioned advantages of triple quadrupole mass spectrometers in improvement of sensitivity and specificity of target compound detection comes at the expense of not detecting other molecules which may also be relevant for the research question.

Untargeted metabolomics

Untargeted metabolomics in comparison to targeted metabolomics aims to detect as many metabolites in the metabolome as possible (2). This makes the untargeted workflow especially useful for explorative studies aiming at finding potential biological processes, such as pathways, involved in the study question, but also requires considerably more time as potential compounds of interest need to be identified. MS instruments used for an untargeted metabolomics workflow most often include accurate mass instruments for hyphenation with LC (33-38) and quadrupole (13) or time-of-flight (15-18) for hyphenation with GC. A further compromise with MS-based untargeted metabolomics is the need to have general MS settings which should allow ionisation and detection of most molecules, but may be far from optimal for some. Targeted metabolomics methods allow for optimisation of MS settings to get the best possible sensitivity, which in turn can impact method variability parameters.

The untargeted metabolomics workflow is ideal when used for generating biological hypothesis without any predefined knowledge and has for example been extensively used in studies aiming at finding associations between whole grain intake and health and metabolism (38-45) as well as in nutritional metabolomics in general (46). However, it is useful to remember that there is no such thing as a universal analytical method and each of the parameters chosen for measuring the metabolome introduces some level of bias into the results

Data processing – making sense of metabolomics data

Both univariate and multivariate statistics are commonly used for analysing metabolomics data. Univariate statistical tools have been designed for testing one or a few different variables, however metabolomics data characteristically has hundreds to thousands of variables, meaning that with a standard significance threshold of 5 %, many ´significant´ results may be obtained purely by chance. This gives rise to the problem of multiple testing, for which correction methods such as Bonferroni (47) and false discovery rate (FDR) (48) are often applied. However, these methods can also exclude relevant results which might not be desirable in an explorative study (49).

A limitation of univariate statistics for metabolomics is that it does not account for how all the data from a sample behaves together and if there are any overall patterns in the data. Multivariate data analysis tools including tools such as principal component analysis (PCA) (50) and partial least square discriminant analysis (PLS-DA) (51) are ideal for metabolomics as they facilitate the analysis of large numbers of samples with large numbers of variables measured. PCA is commonly used to detect outlier samples and for quality control in metabolomics, while PLS-DA and orthogonal-PLS-DA are used to detect differences due to group or outcome. In both PCA and PLS-DA models,

overfitting is a common problem and as with univariate statistics, it is important that the multivariate models are thoroughly evaluated (52,53). A combination of multivariate and univariate statistical methods can help overcome the limitations of each approach.

In studies aiming at predicting or classifying an outcome such as type 2 diabetes, area under the receiver operator characteristics curve (AROC) is a commonly used method (54). AROC is a method for measuring discrimination of a binary variable and can be created by plotting the rate of true positives against the rate of false positives (**Figure 4**). The method with the highest AROC yields the best prediction and overall the closer to the upper left corner the curve reaches the better the prediction whereas a straight line diagonally across the plot gives a random prediction.

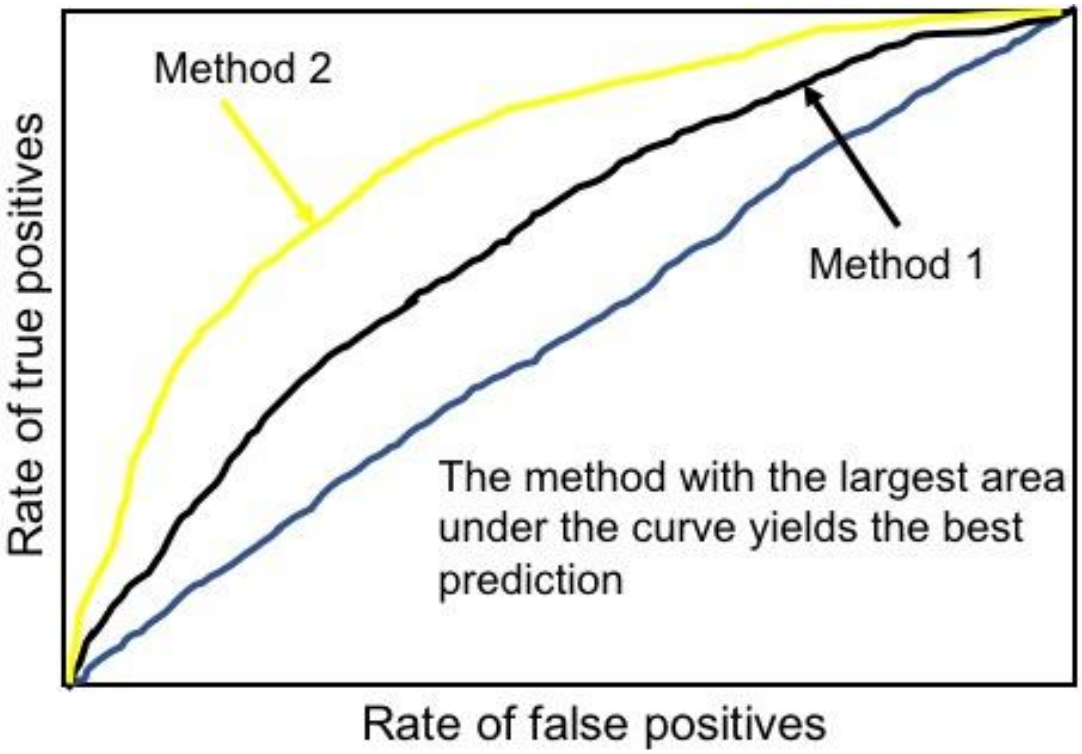

***Figure 4**. An example of an area under the receiver operator characteristics curve (AROC) plot. The yellow line (method 2) represents better prediction than the black line (method 1). If a line follows the blue line, then the model has no predictive ability.*

Applying MS-based metabolomics to large studies

The combination of large numbers of samples from studies with the large number of metabolites that can be detected with metabolomics is potentially a powerful combination for finding new associations and developing new hypotheses on lifestyle and health. MS-based metabolomics has been used in epidemiological settings for exploring relationships between metabolome and coronary heart disease (55), excess alcohol consumption (56) and dietary patterns (57,58). Although widespread application of MS-based metabolomics to epidemiology could lead to new breakthroughs, a number of factors need to be considered.

As with all analytical methods, precision, selectivity, sensitivity, linearity and stability among others are factors affecting the validity of the data and awareness of the effect of these in the data is needed as the analytical variation can make it hard to distinguish biological variation (59-61). Thus, proper evaluation of the analytical method is needed, for which guidelines are available (62,63). For MS based methods one particular source of analytical error is the dirtiness of the ion source, ion optics and collision cell (64) leading to changes in ionization efficiency that can have an impact on within and between batch variation as well as on MS signal drift. Thus, normalization of this and other analytical variation is needed. Various methodologies for data normalization exist based on internal standards and quality control samples (60,65-67). In addition to analytical variation, pre-analytical variation contributes to the total variation and sample collection, storage and preparation are important factors affecting the end result and need to be taken into consideration in the study design (68). Implementation of standardized standard operating procedures for sampling and storing samples are needed for limiting the introduction of unnecessary bias into the results (60,68,69) and for enabling comparison between studies.

Validation of analytical methods is a key aspect of ensuring robustness of a method and for comparing results from batch to batch and lab to lab. A review of the main method validation parameters recommended by the United States Food and Drug Administration is given in Paper III. Although these parameters are well defined for targeted methods, they are less easy to apply to a metabolomics study. There are several papers outlining standardisation approaches for metabolomics (16,70,71) which are needed for standardizing the analytical workflow for guaranteeing valid data. The first steps towards harmonization of reporting standards were made in 2007 (72), suggesting minimum reporting requirements for metabolomics. Complying with and improving on these guidelines should do much to increase the quality of metabolomics data.

Metabolomics in food and nutrition research

In nutrition research, the development of metabolomics has provided new tools for assessing relationships between diets and/or food molecules with diseases such as obesity (73) and type-2-diabetes (74,75) beyond single nutrients or constituents. In addition, given the complex nature of the interaction between food and humans, the wide range of metabolites measured in metabolomics can better describe complex biological responses to diet than measurement of a few biomarkers.

The food metabolome has been defined as “the sum of all metabolites directly derived from the digestion of foods, their absorption in the gut, and biotransformation by the host tissues and the microbiota” (76). The enormous complexity of biology dealt with in exposure studies such as dietary interventions is highlighted by the estimate that foods eaten by humans contain >25000 different types of compounds (77) that interact with endogenous metabolites. Furthermore, food constituents not directly used in metabolism, such as polyphenols, undergo xenobiotic metabolism during which they are biotransformed by enzymes for easier excretion via urine, as well as microbial metabolism which both further add to the complexity of the food metabolome (78).

The utilisation of metabolomics both in research related to food science and nutritional epidemiology has been increasing rapidly as highlighted by a few examples that follow. Starting from raw food material, metabolomics has been applied for discovering new constituents of raw materials, and led to the discovery of novel glycoside forms of benzoxazinoids, a class of natural pesticides, in the edible parts of wheat and rye (79). Since then, benzoxazinoid glycosides have been found to undergo microbial degradation during fermentation and bread baking highlighting the importance of food processing for preservation or degradation of these molecules in foods (37,80). Later, their absorption and metabolism in humans was confirmed, largely based on metabolomics studies (81-85).

Global effects of replacing meat with fish on metabolism

Although many dietary guidelines recommend reducing meat intake and increasing fish intake for better health, few studies have examined the effects of replacing meat with fish on the metabolome in humans. Some population-based studies have examined the global metabolic fingerprint of people who eat meat versus vegetarian or vegan diets (57,86-88), though these are confounded by a large number of lifestyle factors that limit direct conclusions about the role of meat on metabolic health. Many studies have used metabolomics to look for biomarkers of meat or fish intake (89,90). Only two human intervention trials comparing meat with fish and using metabolomics were found in the literature. One study has compared the intake of lean seafood (white fish and scallops) with a mixed meat diet (chicken, lean beef, turkey and pork) in a four week intervention with a crossover design on the urinary metabolome measured using NMR and LC-MS (91). The study found that energy metabolism was likely impacted based on reduced urinary excretion of carnitine, 2,6-dimethylheptanoylcarnitine, and N-methyl-2-pyridone-5-carboxamide, which the authors related to mitochondrial lipid and energy metabolism. The authors also found an increase in trimethylamine-N-oxide after the lean seafood diet, and 3-methylhistidine and guanidineoacetate after the meat diet, linked to the intake of the diets themselves rather than any endogenous metabolic effect. The post-prandial effect of a single meal of baked herring compared to baked beef has also been studied using metabolomics, providing a dynamic view of metabolism (92). Among other findings, metabolomics analysis detected higher circulating levels of 2-aminoadipic acid and a branched chain amino acid (BCAA) leucine to be associated with beef as a protein source in comparison to herring (92). The results are interesting as both 2-aminoadipic acid and BCAA have been associated with elevated risk of type 2 diabetes (T2D) (93,94) supporting the earlier associations between red meat intake and incidence of T2D (95-97). These few studies that have been carried out show that there is much promise in using metabolomics to understand the global effects of eating fish instead of meat, though more studies are required to understand the differences in effects of different types of fish and meat on human metabolism.

Type 2 diabetes

T2D is one of the major global health challenges and its global incidence is expected to increase dramatically during the coming years (98). Development of T2D is generally slow, and results from

an inability to efficiently use insulin to control blood glucose concentration (referred to as insulin resistance) leading to prolonged periods with elevated blood glucose concentrations. Progression of the disease has severe consequences for the patient including neuropathy and microvascular complications (99). Insulin is a hormone secreted from the β-cells in the pancreas in response to elevated glucose concentrations, binding to insulin receptors on cell membranes that in turn send a signal to increase synthesis of glucose transporters that migrate to the cell membrane and allow uptake of glucose into the cell. T2D results from an inability of cells to effectively take up glucose from the bloodstream after a meal as the insulin resistance (IR) in the cells has reached a point where pancreas can no longer produce enough insulin to stimulate glucose uptake. Beyond new developments with bariatric surgery (100), there are no means to cure T2D, so preventing the onset of the disease is the most effective means to reduce rates of T2D.

One method of T2D diagnosis is by measuring venous blood plasma glucose concentrations both at fasting state and two hours after an oral glucose tolerance test during which subjects are given a 75 g glucose load (101). According to World Health Organisation criteria subjects with fasting plasma glucose levels of $\geq$ 7.0 mmol/l or two hour plasma glucose of $\geq$ 11.1 mmol/l are diagnosed as T2D. Diagnosis of T2D is often binary, though the disease itself represents a continuum from normal (healthy) glucose tolerance through to T2D. A third diagnosis can be made from glucose tolerance tests, 'impaired glucose tolerance' (IGT), or 'pre-diabetic'. People with a fasting glucose level of $<$ 7.0 mmol/l and two hour glucose between $\geq$ 7.8 and $<$ 11.1 mmol/l are classified as IGT, and are at high risk for developing T2D (102). Other tests used to diagnose and characterise T2D are glycated haemoglobin (HbA_{1c}) which reflects longer term (2-3 months) blood glucose levels (103) and homeostatic model assessment (HOMA) that uses fasting glucose and insulin (HOMA-IR) in a computational model for estimating insulin resistance (104).

Development of T2D is a multifactorial process affected by both genetic and environmental factors among which obesity is one of the main predictors of development of T2D, which in turn is tightly connected to excess energy intake (105). In this context, the global increase in prevalence of obesity makes it likely that the future global burden of T2D will rise (106). This makes developing therapies and prevention strategies for T2D essential. At present, pharmaceutical treatments for T2D beyond insulin injections have limited success in long-term treatment of T2D, and cannot reverse the disease. As obesity is strongly implicated in the development of T2D, prevention through improvement in lifestyle habits including diet and physical activity should be an effective strategy (107-109). One way to encourage people to take steps to improve their lifestyle to prevent T2D is to identify those at high risk for developing T2D well before they actually develop the disease. Basic clinical parameters such as fasting glucose and oral glucose tolerance test (OGTT) together with factors such as body mass index (BMI), smoking and heredity are all associated with development of T2D (110) but their use, however, has so far not been successful in prevention of the disease as illustrated by the increasing incidence of T2D (111), though lack of widespread screening programmes may also play a role.

Metabolomics and type 2 diabetes

Metabolomics could serve as an alternative means for population level screening to detect T2D risk. Several studies have used metabolomics for exploring the associations of T2D and the metabolome and the findings suggest several metabolite classes are associated with elevated risk of T2D. Among those, BCAA including leucine, isoleucine and valine, have been repeatedly associated with elevated risk of IGT (112-117) or T2D (31,113,116,118-120) suggesting a relationship between BCAA and IGT and T2D. Similarly to BCAA, aromatic amino acids have been repeatedly associated with both IGT (112,114,116,117) and T2D (116,118,120). In addition, various lipid classes (31,32,113,118,121-123), sugars (31,113,116,118,121) and ketones (32,113,116-118,124,125) have been associated with IGT, T2D and development of T2D. The wealth of information created in metabolomics studies opens up new, previously unhypothesised possibilities for mechanistic evaluation of the disease as highlighted recently by the discovery of a furan fatty acid, 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF), to be elevated in subjects with T2D (126). CMPF was also found to induce β-cell dysfunction (126) and its rapid elevation may accelerate development of the disease (127). Interestingly, CMPF has also been associated with increased intake of fatty fish (42) and in the same population no associations between CMPF and glucose tolerance were found (128). This suggests a far more complex relationship between diet and disease, and that one metabolite may predict different things in different populations or circumstances. This illustrates the complexity of understanding the role of different metabolites detected using metabolomics, the difficulty in disentangling the endogenous and exogenous (e.g. food) metabolomes, and the need for scientists to work across disciplines to get the best possible understanding of what metabolomics results may mean.

Prediction of the onset of T2D using metabolomics could offer the possibility for improved detection of the subjects at high risk and screening of the at-risk subjects using a single fasting plasma sample instead of the more common OGTT that requires a longer visit at the clinic and more time with health professionals. However, established predictors such as HbA_{1c}, a marker for longer term elevated blood glucose levels, already can provide reasonable detection of the subjects at risk (129) although in general better prediction has been achieved when using a combination of classical predictors and metabolomics (**Table 1**). Prospective cohorts, where subjects are followed for several years to capture new T2D cases during the follow-up period, are especially useful for evaluating if metabolomics can be used for prediction of future T2D or improve on the existing markers. Large prospective cohorts including but not limited to the Framingham Heart Study (130), Cooperative Health Research in the Region Augsburg (KORA) (131), the Cardiovascular Risk in Young Finns study (FINRISC) (132), the Relationship between insulin sensitivity and Cardiovascular disease study (RISC) (133), the Botnia study (134) and EPIC (135) have been used so far for testing prediction based on metabolomics (**Table 1**). In the Framingham cohort prediction of future T2D was improved from AROC=0.52 to AROC=0.65 when using a combination of five amino acids and age, BMI and glucose in comparison to using age, BMI and glucose (136). In the KORA study prediction based on age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol yielded an AROC of 0,742 whereas adding HbA1c, fasting glucose and insulin into the model improved the prediction to AROC=0.818 (31). Adding

metabolomics results to the first model gave a modest improvement to the prediction from AROC=0,742 to AROC=0.754 and adding metabolomics results to the latter model improved prediction from AROC=0,818 to 0,828 (31). In the EPIC cohort prediction based on diabetes risk score (DRS) gave an AROC of 0,847 whereas prediction based on metabolites had similar prediction power with an AROC of 0,849 (121). Combining DRS with metabolites improved prediction (AROC=0,890) and further with a combination of DRS, glucose and HbA_{1c} (AROC=0,901), whereas best predictive power was achieved by using DRS, glucose, HbA_{1c} and metabolites (AROC=0,912)(121). In the Botnia study family history of diabetes (FHD), sex, age and BMI predicted T2D with an AROC of 0,794, addition of fasting glucose into the model increased prediction to 0,766, addition of fasting and 2-h glucose to 0,788 whereas the best prediction was achieved with further addition of two metabolites (AROC=0,796) (32). These results suggest that metabolomics can identify metabolites that can lead to modest improvements to existing measures used to predict diabetes. More work, including a cost-benefit analysis would be required to see if these improvements would make any practical improvement in reducing diabetes incidence and associated costs. These studies have focused on predicting diabetes, and there may be another potential application of these results in studying how they could improve monitoring of diabetes.

The application of metabolomics in research related to T2D suggest it to be an effective tool for exploring associations between different outcomes related to glucose tolerance and metabolome as highlighted by the replication of BCAA associations in many studies (137). Also, as the studies discussed earlier suggest prediction of future T2D based on metabolomics is possible with reasonable sensitivity and selectivity providing an improvement or an additional means for screening of at risk population. However, it can be argued that already the existing tools i.e. fasting glucose, HbA_{1c} and risk factors already provide sufficient sensitivity and selectivity for detection of at risk subjects and, thus, the focus should be more on better utilization of the existing tools.

Table 1. *Studies using prospective data and metabolomics for prediction of development of dysglycaemia or type 2 diabetes*

Author	Year	Reference	Cohort	Sample	Outcome	Metabolomics method	Predictive metabolites	Prediction model	AROC
Wang et al.	2011	(136)	Framingham	Plasma	incident T2D	Targeted, LC-MS/MS	Isoleucine, leucine, valine, tyrosine and phenylalanine	age, BMI, glucose	0.52
								age, BMI, glucose, isoleucine, phenylalanine, tyrosine	0.65
								age, BMI, glucose, isoleucine, phenylalanine, tyrosine, leucine, valine	0.66
Wang-Sattler	2012	(31)	KORA	serum	Incident T2D	Targeted, FIA-MS/MS and LC-MS/MS	Glycine, LPC C18:2, C2	age, sex, BMI, PA, AI, smoking, SBT, HDL	0.742
								age, sex, BMI, PA, AI, smoking, SBT, HDL, metabolites	0.754
								age, sex, BMI, PA, AI, smoking, SBT, HDL, HbA1c, FG, FI	0.818
								age, sex, BMI, PA, AI, smoking, SBT, HDL, HbA1c, FG, FI, metabolites	0.828
Wang-Sattler	2012	(31)	KORA	serum	incident IGT	Targeted, FIA-MS/MS and LC-MS/MS	Glycine, LPC C18:2, C2	age, sex, BMI, PA, AI, smoking, SBT, HDL	0.638
								age, sex, BMI, PA, AI, smoking, SBT, HDL, metabolites	0.671
								age, sex, BMI, PA, AI, smoking, SBT, HDL, HbA1c, FG, FI	0.656
								age, sex, BMI, PA, AI, smoking, SBT, HDL, HbA1c, FG, FI, metabolites	0.683
Floegel et al.	2013	(121)	EPIC	serum	incident T2D	Targeted, FIA-MS/MS	hexose, phenylalanine, glycine, SM(C16:1), AAPC((C32:1), (C36:1), (C38:3), (C40:5)), A-APC((C34:3), (C40:6), (C42:5), (C44:4), (C44:5)), LPC(18:2)	DRS	0.847
								Metabolites	0.849
								DRS+metabolites	0.890
								DRS+Glc+HbA1c	0.901

								DRS+GlCc+HbA1c+metabolites	0.912
Ferrannini et al.	2013	(32)	RISC	plasma	incident dysglycaemia	Targeted, LC-MS/MS	LGPC, AHB	FHD, sex, age, BMI	0.718
								FHD, sex, age, BMI, FG	0.762
								FHD, sex, age, BMI, FG, 2HG	0.786
								FHD, sex, age, BMI, FG, metabolites	0.790
								FHD, sex, age, BMI, FG, 2HG, metabolites	0.804
Ferrannini et al.	2013	(32)	Botnia	plasma	incident T2D	Targeted, LC-MS/MS	LGPC, AHB	FHD, sex, age, BMI	0.749
								FHD, sex, age, BMI, FG	0.766
								FHD, sex, age, BMI, FG, 2HG	0.788
								FHD, sex, age, BMI, FG, metabolites	0.783
								FHD, sex, age, BMI, FG, 2HG, metabolites	0.796
Wurtz et al.	2013	(138)	FINRISC	IR[1]	serum	Targeted, NMR	isoleucine, leucine, valine, phenylalanine, tyrosine	HOMA-IR, BMI, apoB, PA	0.826
								HOMA-IR, BMI, apoB, PA, metabolites[2]	0.876
								HOMA-IR, BMI, apoB, PA	0.858
								HOMA-IR, BMI, apoB, PA, metabolites[2]	0.856
Yengo et al.	2016	(139)	D.E.S.I.R	serum	incident T2D	Untargeted, GC-MS and LC-MS/MS	isoleucine, isovalerylcarnitine, phenylalanine, pro-hydroxy-pro, serine, tyrosine, fructose, mannose, glucose, 1,5-anhydroglucitol, L-GPC, 1-palmitoylglycerol, γ-glutamylphenylalanine, γ-glutamyltyrosine, cotinine, piperine	metabolites	0.86

[1]Defined as HOMA-IR > 90th percentile at 6-y follow-ip
[2]sum of concentrations of the predictive metabolites
Abbreviations used in the table: PA=physical activity; AI=alcohol intake; LPC(18:2)=lysophosphatidylcholine C18:2; C2=acetylcarnitine; FG=fasting glucose; FI=fasting insulin; DRS= (7.4 × waist circumference [cm]) − (2.4 × height [cm]) + (4.3 × age [years]) + (46 × hypertension [self-report]) + (49 × red meat [each 150 g/day]) – (9 × whole-grain bread [each 50 g/day]) – (4 × coffee [each 150 g/day]) – (20 × moderate alcohol [between 10 and 40 g/day]) – (2 × physical activity [h/week]) + (24 × former smoker) + (64 × current heavy smoker [≥ 20 cigarettes/day]); SM=sphingomyelin; AAPC=diacylphosphatidylcholine; A-APC=acyl-alkyl-phosphatidylcholine; LGPC=linoleoylglycerophosphocholine; AHB=α-hydroxybutyrate; FHD=family history of diabetes; 2HG=plasma glucose concentration 2h after an oral glucose load; apoB=apolipoprotein B; IR=insulin resistance

Biomarkers

Biological markers, more commonly referred to as biomarkers, are measures that can be used to objectively indicate a status of a subject (140). By definition a biomarker is:

- 'A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention' (141)
- 'Any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease' (142)
- 'Almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological. The measurement response may be functional and physiological, biochemical at the cellular level, or a molecular interaction' (143)

Based on these broad definitions a biomarker can be almost any measurable biological parameter that has a relevant association with a condition, disease, intervention or a response. At its simplest, biomarkers can be physiological measurements such as measurement of blood pressure reflecting cardiovascular risk (144) or weight or waist and hip circumference reflecting obesity. Well known biomarkers of T2D measured in blood include fasting glucose and HbA_{1c} (110,145,146).

Based on their use biomarkers can be classified into three classes as follows: Biomarkers of Exposure, Biomarkers of Effect, and Biomarkers of Susceptibility. Biomarkers of exposure are measurable substances, either parent molecules or metabolites that can reflect exposure to an external factor such as environmental pollutants or diet. Thus, the exposure biomarkers can be used to finding associations between the external factor and health without additional data on the magnitude of exposure (143). Dietary biomarkers fall into the category of biomarkers of exposure and can provide a link between diet and an outcome when dietary intake data may not be accurate or is not available (147). Biomarkers of effect can be endogenous substances which when they change concentration reflects or is associated with an outcome (143), often disease risk, such as T2D. Lastly, biomarkers of susceptibility are indicators of an inherent or acquired ability that makes the host more susceptible to an outcome such as disease (148). Such a marker can be for example a genetic factor leading to elevated disease risk (149).

Dietary Biomarkers

Estimating food intake is central in nutrition research. Estimations rely primarily on methods such as food frequency questionnaires (FFQs), food diaries and 24 h recalls. However these tools are subjective and have several well-known limitations, including subjects not always recalling all foods or food components they have consumed, difficulties in estimation of portion sizes, and incorrect reporting of what has been eaten (150-152). Measuring biomarkers of exposure to diet, or dietary biomarkers, can provide an independent measure of food intake that is not influenced

by biases related to self-reporting. Dietary biomarkers can be used in conjunction with other measures of intake to identify major discrepancies between diet estimation methods, and also in situations where no dietary intake data is available. This latter case can be especially important in cohort-based studies or when performing case-control studies based on biobanked samples, which may not have dietary intake data associated with them. Numerous molecules have been suggested as possible biomarkers for intake of different food items or diets (77) although to date few extensively evaluated dietary biomarkers exist (153-155).

In general, dietary biomarkers can be categorized into three classes, recovery, concentration and predictive biomarkers. Recovery biomarkers, such as urinary nitrogen for protein intake (156), are markers that quantitatively reflect dietary intake and thus need to be properly validated in calibration studies for establishing their use in diets and situations (157). Recovery biomarkers can also be used as a tool for validating other dietary intake measurement methods such as via the method of triads (158). Concentration biomarkers, such as alkylresorcinols for whole grain wheat and rye intake (159), can be used as reference measurements for dietary intake and their concentration correlates with dietary intake but as such the information cannot be translated into actual intake values (157). Lastly predictive biomarkers, such as urinary sucrose and fructose for sugar consumption (160), that can be used for predicting dietary intake rather than measuring it due to overall low recovery of the biomarker (157,160).

An optimal dietary biomarker would accurately and specifically reflect dietary intake and it should in addition be applicable to many populations (157). Typical biomarkers are, however, not ideal and there are several factors affecting how good/valid a dietary biomarker is stemming from food (compositional variation, food processing and preparation), human physiology (including absorption, metabolism and excretion) and from the measurement including sample collection, sample preparation and both biological and analytical reproducibility (153,161). Dietary biomarkers should be specific meaning that the measurement should reflect intake of the food item or nutrition under examination (161). A number of factors such as genetic variation, lifestyle, dietary factors and biomarker measurement, can confound the relationship between the biomarker and exposure (157) and working towards taking these into account is important (162), or at least knowing the extent these confounders can effect a result. The individual differences in metabolism possess a challenge for the use of dietary biomarkers as the variation between individuals is not well understood and absorption rates, affected by a number of factors, can vary not only between individuals but also within individuals (153). Metabolic rates may also make time of sampling critical, with differences in pharmacokinetic half-life potentially leading to great differences is a fasting blood sample is collected at for example 8 or 12 hour after the last meal. A biomarker should also be reproducible both in physiological response and in analytical measurement i.e. if food intake is stable, the biomarker measurement should be similar across several days (161).

Among the few well studied biomarkers urinary sucrose and fructose have been established as a biomarker for total sugar intake in both normal and obese subjects and further detect changes in consumption (160). Furthermore, using these biomarkers subjects can be divided into high and low consumers of sugar (163). Alkylresorcinols (AR) reflect whole grain (WG) rye and wheat intake

(159) and several studies have found correlations between dietary intake of whole grain rye and wheat and plasma concentrations of AR suggesting suitability of AR as biomarkers for whole grain rye and wheat intake (164-167) and AR have been used for confirming compliance in several studies (168-170). AR have been further exploited to understand more about cereal intake, including using the ratio between the AR homologues C17:0 and C21:0 to determine subjects who eat more wheat or rye (166,167). Also, the potential of AR as biomarkers of gluten free diet important in treatment of celiac disease patients has been suggested recently (171). In addition, it was recently discovered that quinoa, a South American pseudocereal, contains including an array of AR that opens the possibility to separate WG rye and wheat users from quinoa users (172). The varying examples of AR as biomarkers demonstrates the potential of dietary biomarkers not only as markers of dietary intake, but also as an additional tool in treatment of disease.

A few studies has demonstrated the potential of using dietary biomarkers in intervention studies (165,173-175) which could be used as an ´unbiased´ measure of either compliance or diet or to capture dietary patterns reflecting diet quality as demonstrated recently (176,177). Overall the biggest limitation of the application of biomarker approach for assessment of nutrition is the lack of valid biomarkers both in restricted and larger populations. Although many of the nutritional metabolomics studies undertaken focusing on food intake have found potential dietary biomarkers (77), the necessary validation work for the suggested dietary biomarkers remains to be done. Large cohorts and a diverse range of intervention studies with dietary intake data are needed for assessing the accurateness and robustness of these biomarkers and in addition the approaches of measuring them which is imperative for increasing valid use of biomarkers in dietary assessment (178).

Discovery of dietary biomarkers using metabolomics

Metabolomics based approaches for biomarker discovery have aided the identification of many potential new dietary biomarkers (77). Certainly, the monitoring of hundreds of small molecules simultaneously offers new possibilities for biomarker discovery in addition to knowledge based measurement of food constituents (58). However, a bottleneck in biomarker discovery work is the validation work that needs to be done after discovery of a potential biomarker. Scalbert et al (77) suggested division of dietary biomarker discovery approaches into hypothesis-driven and data-driven, where metabolomics can be applied in both approaches. These approaches are in principle analogous to the division of metabolomics analyses into targeted and untargeted; in hypothesis-driven approaches, prior knowledge based on, for example, available biomarker databases are used to limit the candidate biomarkers that will be measured whereas in data-driven approach all measured metabolites are considered as candidate biomarkers (77). Also, the application of targeted metabolomics methods will direct the biomarker discovery towards the set of measured molecules (177) in contrast to using untargeted metabolomics workflows allowing a wider screen of potential biomarkers. Examples among the potential biomarkers are the urinary concentrations of 1- and 3-methyl-histidine (179-181), and plasma concentrations of β-alanine and 4-hydroxyproline (92) which have been found to be associated with meat intake in metabolomics studies. There is still a need to validate these potential biomarkers, and in particular establish how

well they can be distinguished from breakdown of muscle, which also adds similar compounds to the overall metabolome.

In general, both intervention and cohort studies can be used for dietary biomarker discovery. Short and medium intervention studies where subjects consume a known amount of a given food either once, or over a period of time can be used for discovering specific biomarkers (77,182). In a single-meal study baseline and postprandial samples are collected and, in general, in a longer intervention study samples are collected at baseline and at the end of the intervention period. However, intervention studies lack the variability of a habitual diet and thus the discovered putative dietary biomarkers may not be sufficiently specific at population level. Cohort studies with food intake data allow the identification of high and low consumers of certain foods making it possible to cross-sectionally examine the relationships of possible biomarkers with the food intake data in the cohort population. However, as with intervention studies some specificity problems arise also from cohort studies as in many cases specificity of a biomarker to a certain food cannot be confirmed, and requires further work in other cohorts and intervention studies to confirm the relationship.

METHODS

Sample preparation and GC-MS/MS metabolomics method

A novel GC-MS/MS method aiming to simultaneously collect MRM and full scan data was developed using common protocols for preparation of blood plasma for analysis. In brief, blood plasma collected using sodium citrate as an anticoagulant was extracted and derivatised using the method of A et al (15). Ten stable isotopes were included as internal standards for correcting within and between batch variation, and dried plasma extracts and standards were derivatised using methoxymation followed by silylation. Analyses were carried out on a Shimadzu GCMS TQ-8030 GC-MS/MS system. Optimized conditions for MS data collection were scan time (50-700 m/z) 75 ms, MRM transitions 45 ms, for a total loop time of 120 ms (Paper I).

A MATLAB (Mathworks, Natick, MA, USA) script and database developed at the Swedish Metabolomics Centre (Umeå, Sweden) was used for targeted analysis of the full scan data (183). Internal standard normalisation of raw data was performed using the method of Jonsson et al (183).

GC-MS/MS method evaluation

The performance parameters used for evaluation of the GC-MS/MS metabolomics method were full scan and loop time, linearity, limit of detection (LOD), plasma volume, precision of the targeted and untargeted part of the method, and accuracy (**Table 2**).

Table 2. *Parameters and criteria used for evaluation of the GC-MS/MS metabolomics method.*

Parameter	Assessment	Criteria
MRM optimization	9-45 eV	Sensitivity and selectivity
Full scan time	50-75 ms	Number of deconvoluted spectral features
Loop time	100-150 ms	Number of deconvoluted spectral features
Linearity	Standard and plasma	R^2
Limit of detection	Calculation	Concentration
Plasma volume	40-160 ul	Reproducibility of MRM analytes
Precision (MRM)	Intra- and interbatch precision	RSD%
Precision (scan)	Intra- and interbatch precision	Number of features passing three RSD% thresholds
Accuracy	Comparison against reference	Percentage difference

Application to an intervention study: Comparing herring with chicken and pork intake

Study population and design

This study was conducted as a data analyser blinded randomised crossover intervention trial approved by the Ethical Committee of the Western Götaland region (359-03) and carried out in 2003. Details of the study design have been published previously (184). Briefly, fifteen overweight and obese men and women (mean age 50.5 y, mean BMI 32.6 kg/m^2, mean weight 97.5 kg) with no known chronic or serious health issues were recruited for this study. Participants were provided with frozen meals including 150 g of baked herring or 120-150 g chicken and lean pork per meal and were instructed to eat five of these meals per week for lunch or dinner during each four week intervention period. The two intervention periods were separated by a two-week washout period. Fasting blood samples used in this study were collected at baseline, week two and week four in each study period. Thirteen subjects completed the study, with two dropouts due to use of analgesic medication and inability to attend all parts of the study and there was insufficient plasma for metabolomics to be measured for two subjects. Samples were frozen and stored in -80°C.

Study foods

The study foods have been previously described (184). In brief, fresh herring, lean pork and chicken fillets were baked in the oven at 150°C until protein denaturation temperature and prepared as part of frozen ready meals within three days of catching or slaughter. Herring, lean pork or chicken fillets were made into complete meals that contained mashed potatoes, baked potato wedges, pasta or rice, with different vegetables and sauces to increase meal variety. The dishes were identical except for herring vs. chicken/lean pork between the two interventions. The average nutrient content of the meals was similar between the two interventions, aside from fat content (184). The herring meals contributed 3.4 g/d of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), while the chicken and pork meals contained negligible amounts of these LC n-3 PUFA. Meals were stored frozen at -40°C until delivery to the subjects, who were instructed to keep the meals in their freezers until reheating in a microwave using standardised cooking times.

Chemical analyses

GC-MS metabolomics was performed using the method as outlined above and with the same instrument with the exception that only scan data was acquired and used. Nitric oxide was determined using a colorimetric kit (Cell Biolabs, San Diego, CA, USA). Amino acids were measured in the same sample extract that was used for metabolomics analysis after storage at -20°C for two years. Amino acid analysis of amino acids of interest based on the GC-MS results was done using LC-MS (185). Ornithine (m/z 133), Arginine (m/z 175), Asparagine (m/z 133) and Glutamine (m/z

147) were detected and quantified against an external standard curve. Citrulline and aspartate could not be detected in the plasma samples.

Data analysis

Global differences in plasma response between the meals were investigated with PCA and orthogonal partial least squares-discriminant analysis (OPLS-DA) using the SIMCA software (Umetrics AB, Umeå, Sweden)(186). After normalization based on internal standard, data were corrected for baseline by subtracting the baseline values from the week 2 and week 4 time points, and data were subjected to unit variance (UVN) scaling and mean centring. The plasma responses to the baked herring meal were compared to the pork/chicken-based meal in an OPLS-DA analysis, where the difference between the meals was explained by the first predictive component. The OPLS model was made from all time points in the dataset, i.e. 0-4 weeks, with a 10-fold validation. Discriminating variables detected using OPLS-DA were further analysed for differences between diets using mixed model analysis and post hoc t-tests to determine differences between diets. Mixed models were used to determine diet x time differences for individual metabolites. Meal type and time were included as fixed factors, adjusted for age and gender with subject as a random variable. Data was corrected for baseline by subtracting the baseline values from the corresponding week 2 and week 4 values. XLSTAT for Excel 2013 (Addinsoft, Paris, France) was used for all mixed model and univariate data analyses. Uncorrected P values < 0.05 are reported as of interest in line with the exploratory nature of this work. We have reported trend results at $P < 0.1$ if they were potentially related to significant metabolites.

To determine the overall metabolic pathways affected by the difference in diets, the significantly changed metabolites and their intensities were mapped to the human metabolic network and pathway over-representation analysis (187) was performed. A hypergeometric test estimated the relative significance of the over-represented pathways against the background KEGG pathways for *Homo sapiens* (188). The Relative Importance of the Metabolites (RIM) involved in pathway(s) were estimated by betweenness centrality (189).

Application to a population based cohort: The DIWA study

Study population

A population-based cohort established by inviting all 64-year-old (n=4856) women identified through the County Register in Gothenburg region to participate in a screening examination that took place in 2001-2004 **(Figure 5)** (190). The definition of IGT and diabetes were based on WHO criteria (191). The means of the two oral glucose tolerance tests were used in the classification of IGT at baseline. The screening examination included fasting capillary whole blood glucose measurements in women with overt diabetes and repeated OGTT (110,190) in women without overt diabetes. Among the 2595 women screened, prevalence of diabetes was 9.5% and prevalence of IGT was 14.4%. Women with diabetes, similarly sized groups of women with IGT and normal glucose tolerance (NGT) were randomly selected and invited to participate in a nested

case–control study, which included a baseline examination. In total, 629 women participated in the first examination and samples from 607 subjects were analysed by GC-MS/MS metabolomics (T1D n = 14, T2D n = 202; IGT, n = 203; NGT, n = 188) (110). T1D was defined as serum glutamic acid decarboxylase antibody level ≥4.6. A nested case-control re-examination was performed after 5.5 years in 500 women with diabetes (n = 159), IGT (n = 174) or NGT (n = 167) according to the classification at baseline. The remaining 129 women from the original study did not participate in the re-examination because of death (n = 23), severe disease (n = 3), because they were no longer living in the area (n = 12) or were unwilling to participate (n = 91) (110). For a comprehensive description of the cohort it is recommended to refer to the original description of the cohort (190).

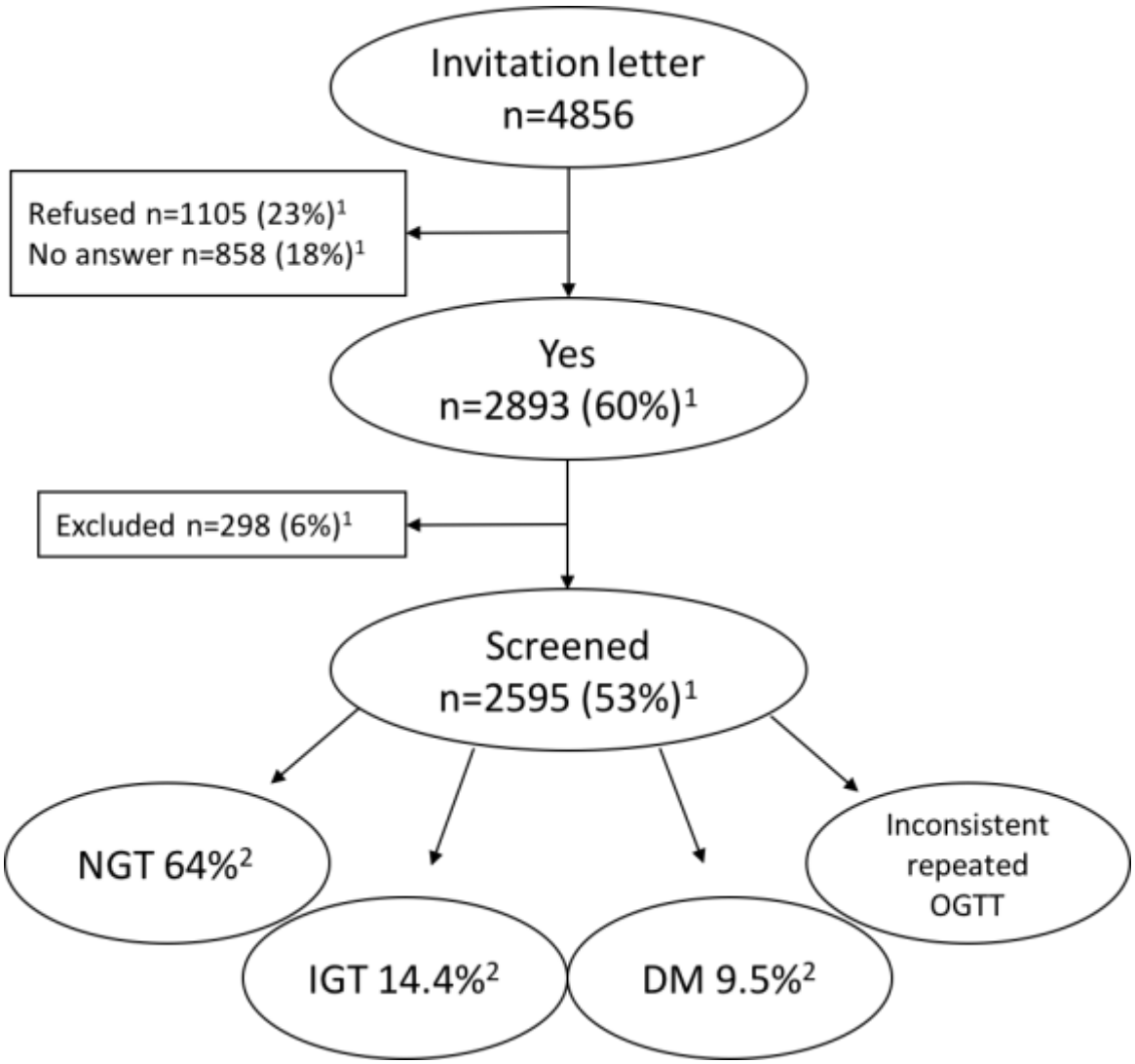

Figure 5. *The screening procedure. [1]Percentage of invited and [2]Percentage of screened.*

Clinical biochemistry

Blood glucose, insulin and adiponectin as well as total cholesterol, high density lipoprotein (HDL), low density lipoprotein (LDL), free fatty acid and triglyceride measurements were performed using standard clinical chemistry techniques as previously described at the Clinical Chemistry Laboratory of the Sahlgrenska University Hospital, Gothenburg (110).

Data analysis strategy and statistical analyses

Associations between potential dietary and nutrient biomarkers and glucose tolerance status

Due to the possibility to simultaneously detect multiple molecules the GC-MS/MS metabolomics data was searched for potential food derived molecules that potentially reflect either dietary intake or nutrient status. The following potential biomarkers were found: α- and γ-tocopherol [vitamin E], alkylresorcinols C17 and C19 [wholegrain wheat and rye], β-alanine [meat], CMPF [fish], lauric acid [saturated fat], oleic acid [found in animal fat and olive oil], eicosapentaenoic acid (EPA) [fish] and linoleic acid [often associated with margarine and vegetable oil] (**Table 5**). These compounds were used in the data analysis plan outlined below.

Samples taken at baseline were used for exploring the associations between the potential dietary and nutrient biomarkers with glucose tolerance groups (GTS) defined as NGT, IGT and T2D, and new T2D cases at follow-up, correcting for common risk factors of T2D and cardiovascular disease (CVD). Analysis of covariance (ANCOVA) and Tukey´s post hoc t-test were used for comparing levels of the potential biomarkers between the three GTS groups and logistic regression for testing if baseline values of potential biomarkers were associated with new T2D at follow-up. All data were checked for normal distribution and skewed variables were log transformed. For logistic regression analysis each biomarker was scaled to have a mean of zero and a standard deviation of one to have comparable units for all biomarkers. As the focus of this work was to find relationships between potential dietary and nutrient biomarkers and incident T2D, cases with previously diagnosed T2D or T1D were excluded from the analyses. Waist circumference, total cholesterol, HDL, LDL, free fatty acids, systolic blood pressure and triglycerides were considered as potential confounders and were first individually tested for each biomarker and outcome combination (Stage 1). After this, a step-wise model reduction was performed starting from the model including all potential confounders ($P<0.05$ at Stage 1 testing) for each biomarker individually followed by removal of the least significant variables one by one and ending with a model including only significant ($P<0.05$) confounders (Stage 2) (**Table 5**). Results for the logistic regression are reported as odds ratios (OR) and 95% CI for 1 standard deviation change in concentration of biomarker for significant biomarkers. Results from ANCOVA models comparing GTS were reported as fold differences compared to NGT and to IGT in the case of T2D. As this study is explorative, testing for multiple comparisons was not done and results were considered significant at $P<0.05$.

Finding predictive markers of Type 2 diabetes

Baseline data was used to explore potential metabolomics markers of T2D after exclusion of medicated T2D cases and the 69 participants who developed T2D during follow-up. The candidate markers for detection of incident T2D were tested on those subjects who developed T2D during 5.5 years of follow-up. Sixty of the 69 women who developed T2D during the follow up period had IGT at baseline. The subjects with T1D at baseline were not included in the analysis.

For finding predictive biomarkers the NGT and IGT groups, excluding the 69 women who developed T2D during follow up, were compared with the T2D group in two pairwise comparisons (NGT vs. T2D and IGT vs. T2D) using multivariate logistic regression models adjusted for analytical batch. Metabolites found to be significant in both of the comparisons were considered as potential predictive metabolites. To avoid statistical overfitting of the prediction models the number of metabolites used for prediction of incident T2D was restricted to the nine that separated T2D at baseline with the highest statistical difference. Before statistical analyses data was checked for normal distribution and skewed variables were log transformed. To put equal statistical weight for all metabolites, each metabolite was scaled to have a mean of zero and a standard deviation of one. Correction for multiple testing was done using the Benjamini-Hochberg method (48). Metabolomics data results are reported as odds ratio for 1 standard deviation change in concentration of metabolite for significant metabolites.

Prediction of development of Type 2 diabetes

In total ten different prediction models (**Table 7**) were tested for detection of the subjects that developed T2D (n=69). Models anthropometry model(AM), anthropometry + metabolomics model (AM+M), adiponectin model (AdM) and adiponectin + metabolomics model (AdM+M) were based on known markers of T2D risk with and without the strongest metabolite predictors. Metabolomics model (MM) was based on the 9 strongest candidate predictors and optimized metabolomics model (OM) was developed to determine if the 9 metabolites used in the MM could be reduced with minimal loss of prediction to make an optimised metabolomics model. The OM model was further combined with the glucose measures IGT and impaired fasting glucose (IFG) to test if these will further improve the prediction. Furthermore, the potential dietary and nutrient biomarkers were used for prediction of the new T2D cases alone (model NB) and in combination with AdM model. Area under the receiver operating characteristics curve (AROC) was used to assess the discriminating power of the prediction models. AROC was calculated with the trapezoidal rule using the R-package pROC (version 1.8). Using the same R-package the different AROC curves were tested for significant differences using a non-parametric comparison as described earlier (192). Model reduction for the 9 significant metabolites was done by finding the optimal model using R package glmulti (193). Akaikes information criterion was used to evaluate goodness of fit of the models (194).

RESULTS

A new GC-MS/MS metabolomics method

Optimization and testing of instrumental parameters and spectral features in full scan data

MRM parameters were optimized for 37 molecules and 10 labelled internal standards (**Table 3**). In the case of the two AR molecules, selected ion monitoring (SIM) of the ion 268 m/z was used instead of MRM as this gave better sensitivity than MRM. Depending on the number of MRMs measured simultaneously, dwell time varied between 1.5 and 20.5 ms.

The time spent measuring the untargeted scanning data had an effect on the number of detected spectral features from the data. The amount ranged between 265 and 411, with the highest number of peaks detected with a 70 ms scan time, and the lowest with a 65 ms scan time. Both 70 and 75 ms scan times resulted in a similar number of features and 75 ms was selected for the method as it is not close to the scan-rate threshold.

The tested loop times did not have a clear pattern in the number of detected features as a function of loop time. Varying loop time between 100 and 150 ms yielded 413 to 456 detected features. 120 ms was selected for the method for providing sufficient time for allocation of multiple MRMs while still maintaining sufficient spectral acquisition rate (8.22 Hz) for the scanning data.

The average number of raw detected spectral features in the scanning data was 2979 which was further reduced to 1054 by application of a minimum normalised peak area filter of 6000. The number of spectral features passing the reproducibility threshold levels, 30, 25 and 20 RSD%, was 601, 501 and 373 respectively for intra-batch variation and 486, 297 and 143 for inter-batch variation. Deconvolution of a matrix free blank sample yielded 20, 4 and 1 features with the same threshold limits.

Table 3. *Optimized MRM transitions with collision energy (CE) for quantifier and qualifier ions as well as accuracy as percentage of the measured value over the reference value.*

	Accuracy (µmol/l)			Quantifier		Reference		
	Reference	Measured	Accuracy (%)[1]	Transition	CE	Transition	CE	Retention index
cis-Aconitic acid	-	-	-	229.00>147.10	12	375.00>147.10	18	1745
Alanine	8,9	11,8	132,8	116.00>73.10	12	147.00>73.10	18	1101
β-Alanine	-	-	-	174.00>73.00	18	248.00>147.00	18	1423
Alkylresorcinol C17:0[2]	-	-	-	268.00[2]	-	-	-	2890
Alkylresorcinol C19:0[2]	-	-	-	268.00[2]	-	-	-	3085
Arginine/Ornithine	-	-	-	142.00>73.00	15	174.00>73.00	15	1813
Asparagine	-	-	-	231.00>73.10	27	188.00>73.10	9	1665
Cholesterol	-	-	-	329.00>95.00	9	129.00>73.00	30	3149
Citric acid	-	-	-	273.00>73.10	24	273.00>183.20	12	1809
CMPF[3]	-	-	-	327.00>147.20	18	266.00>147.10	24	1968
Dimethylarginine	-	-	-	256.00>157.20	9	256.00>75.10	24	1828
Fumaric acid	-	-	-	245.00>73.10	24	143.00>75.10	12	1351
Glutamic acid	3,3	3,5	106,9	246.00>73.00	27	128.00>73.00	12	1617
Glycine	6,1	5,7	92,8	174.00>73.10	15	248.00>147.10	15	1302
Histidine	3,8	7,4	198	154.00>73.10	18	254.00>73.10	24	1912
2-Hydroxybutyric acid	-	-	-	131.00>73.10	12	205.00>147.10	12	1131
3-Hydroxybutyric acid	-	-	-	191.00>147.10	12	233.00>147.20	12	1163
myo-Inositol	-	-	-	217.00>73.10	18	305.00>73.10	27	2081
Isoleucine	2,4	2,6	107,5	158.00>73.00	18	218.00>73.00	18	1291
Ketoleucine	-	-	-	200.00>82.1	15	216.99>89.1	12	1215
α-Ketoglutaric acid	-	-	-	198.00>73.10	12	288.00>73.10	24	1574
Lactate	-	-	-	191.00>73.10	9	191.00>147.10	9	1130
Leucine	4,4	4,7	106,6	158.00>73.00	15	102.00>73.00	9	1271
Lysine	6,8	7,9	115,6	156.00>73.00	18	174.00>73.00	18	1917
Methionine	1,1	1,1	94,9	176.00>128.00	9	128.00>73.00	9	1515
3-Methyl-2-oxopentanoic acid	-	-	-	189.00>79.1	15	189.00>99.1	9	1195
Phenylalanine	2,8	2,7	96,2	218.00>73.00	15	192.00>73.00	15	1622
Proline	6,8	6,6	96,5	142.00>73.00	18	216.00>147.00	15	1292
trans-4-hydroxy-L-proline	-	-	-	230.00>73.00	27	230.00>140.00	9	1521
Serine	3,4	1,4	40,5	204.00>73.00	15	218.00>73.00	15	1359
Succinic acid	-	-	-	247.00>147.20	9	247.00>73.10	27	1328
Threonine	4,7	4,7	98,4	218.00>73.00	18	291.00>101.00	12	1384
α-Tocopherol	2,7	3,2	117,4	237.00>73.00	15	502.00>73.00	36	3133
γ-Tocopherol	0,6	0,9	153,5	223.00>73.00	36	488.00>73.00	12	2989
Tryptophan	-	-	-	202.00>73.00	18	291.00>101.00	18	2207
Urea	78,1	76,1	97,4	189.00>147.20	12	189.00>73.10	24	1249
Valine	7,1	8,5	119,6	144.00>73.10	15	218.00>73.10	21	1214
Internal standards								
Cholesterol-D_7				336.00>121.20	18	129.00>73.1	9	3118
Glucose-$^{13}C_6$				323.00>73.10	30	147.00>73.10	18	1899
Glutamic acid-$^{13}C_5$,^{15}N				251.00>73.10	30	251.00>132.10	15	1624
α-Ketoglutarate-$^{13}C_4$				202.00>73.10	18	147.00>73.10	30	1580
Methyl stearate				143.00>55.10	18	298.00>101.10	18	2124
Proline-$^{13}C_5$				146.00>73.10	15	220.00>147.10	15	1294
Putrescine-D_4				174.00>73.10	18	174.00>86.10	12	1728
Salicylic acid-D_4				271.00>73.10	30	149.00>133.10	9	1501
Succinic acid-D_4				147.00>73.10	18	309.00>293.80	18	1317
Sucrose-$^{13}C_2$				367.00>73.10	30	367.00>174.10	9	2641

[1](measured value)/reference value * 100, [2]Measured in single ion monitoring mode, [3]3-carboxy-4-methyl-5-propyl-2-furanopropanoic acid

Linearity, Plasma Volume and Limits of detection

Using a plasma sample volume of 100 and 120 µL in the extraction protocol were best with average RSD% for the compounds measured by MRM between 3.5 and 4.1. The use of lower plasma volumes led to increased variation.

The majority of the compounds measured by MRM had a good linearity measured both from standard solution and plasma matrix with $R^2 > 0.99$ (**Table 4**). Furthermore, compounds with structurally similar internal standard had higher R^2 values indicating that use of more internal standards may lead to better linearity for more compounds. In the plasma matrix, α- and γ-tocopherol had inverse U shaped calibration curves, with reduced response when larger volumes of plasma were used.

Limits of detections were assessed for both the scanning data and MRM compounds (**Table 4**). Overall the estimated LODs varied between 0.05 and 102.8 µmol/L for the MRM data and between 0.23 and 2118.4 µmol/L for the scanning data.

Precision and Accuracy

Intra-batch variation ranged between 1.0 and 16.7 % with 31 compounds <10 % and 6 between 10 and 19.1 % whereas inter batch variation was under 10 for 22 compounds, between 10 and 20 % for 10 compounds and between 20 and 30 % for three compounds (γ-tocopherol, dimethylarginine and tryptophan) (**Table 4**).

Overall accuracy based on analysis of a recognised standard plasma sample ranged between 93 and 133 %, except for 3 compounds (γ-tocopherol 154 %, histidine 198% and serine 41%) (**Table 3**).

Table 4. *Overview of the established method parameters.*

	Calibration range in plasma (µM)		Coefficient of determination (R^2)			Limits of Detection (µM)		Precision of MRMs (RSD%)		Precision of full scan peaks (RSD%)	
	Low	High	Standard	Plasma MRM	Plasma full scan	Standard MRM	Standard full scan	Intra-batch[1]	Inter-batch[2]	Intra-batch[1]	Inter-batch[2]
cis-Aconitic acid	2.15	17.23	0.999	-	-	0.17	0.41	3.5	7.8	-	-
Alanine	24.00	481.52	0.999	0.989	0.965	14.47	129.14	5.1	9.5	6.2	10.2
β-Alanine	0.24	3.37	0.999	0.930	-	0.07	0.32	3.7	16.0	-	-
Alkylresorcinol C17:0	0.06	0.86	0.998	-	-	0.05	-	-	-	-	-
Alkylresorcinol C19:0	0.06	1.59	0.997	-	-	0.05	-	-	-	-	-
Arginine/Ornithine	4.80	95.34	0.999	0.992	-	-	-	16.5	17.1	-	-
Asparagine	2.40	47.68	0.998	0.971	0.987	6.81	0.49	12.7	14.3	10.6	15.6
Cholesterol	120.00	2399.82	0.999	0.993	0.130	-	-	7.5	10.2	-	-
Citric acid	4.80	96.81	0.949	0.985	0.859	-	-	14.9	18.4	15.3	30.0
CMPF[3]	3.00	59.93	0.997	0.988	-	3.83	-	7.3	12.2	-	-
Dimethylarginine	0.06	0.49	0.939	-	-	-	-	3.0	29.6	-	-
Fumaric acid	0.43	2.58	0.999	-	-	0.06	0.23	2.8	4.8	-	-
Glutamic acid	2.40	48.94	0.998	1.000	0.989	0.43	2.39	2.4	3.1	1.8	7.1
Glycine	12.00	239.79	0.999	0.949	0.951	3.13	23.51	3.0	4.7	3.1	3.2
Histidine	6.00	119.88	0.999	0.975	0.988	-	-	4.7	10.0	9.3	56.5
2-Hydroxybutyric acid	1.20	23.05	0.993	0.998	-	-	-	1.8	3.2	-	-
3-Hydroxybutyric acid	4.80	95.10	0.970	0.998	0.965	0.43	1.75	1.9	8.6	6.0	12.9
myo-Inositol	0.60	11.66	0.998	0.994	-	0.36	4.81	3.1	5.5	-	-
Isoleucine	6.00	121.21	0.999	0.998	0.970	1.05	22.54	5.5	5.9	7.0	13.8
Ketoleucine	6.44	51.48	0.999	-	-	3.02	-	0.9	5.3	-	-
α-Ketoglutaric acid	2.74	22.59	0.997	-	-	0.64	6.62	0.9	10.1	-	-
Lactate	343.04	2754.28	0.999	-	-	58.38	818.13	1.6	7.6	-	-
Leucine	6.00	121.21	0.999	0.998	0.971	1.64	44.38	3.0	4.4	14.7	25.4

Lysine	12.00	240.10	0.999	0.992	0.880	29.06	6.72	1.8	5.1	5.7	14.6
Methionine	1.20	24.13	0.996	0.997	0.992	0.24	-	2.6	3.4	-	-
3-Methyl-2-oxopentanoic acid	4.42	34.58	0.999	-	-	2.44	-	2.1	9.2	-	-
Phenylalanine	6.00	119.86	0.998	1.000	0.993	1.06	14.88	2.7	2.6	1.6	8.1
Proline	12.00	239.73	0.999	1.000	0.966	1.86	36.92	2.2	8.6	4.1	8.8
trans-4-hydroxy-L-proline	1.20	25.17	0.996	0.993	-	0.45	-	2.9	3.1	-	-
Serine	9.60	191.26	0.992	0.961	-	3.59	10.93	8.8	18.0	-	-
Succinic acid	3.29	25.40	0.999	-	-	2.38	3.93	5.1	5.9	-	-
Threonine	9.60	191.40	0.999	0.936	-	7.39	-	16.6	16.1	-	-
α-Tocopherol	2.40	48.06	0.995	0.812	-	2.92	5.97	1.8	4.9	-	-
γ-Tocopherol	0.24	5.04	0.998	0.200	-	0.20	0.59	16.7	25.1	-	-
Tryptophan	4.80	95.48	0.999	0.967	0.959	2.28	8.24	1.0	25.1	4.6	21.4
Urea	240.00	4800.20	0.997	0.999	-	102.84	2184.08	7.2	9.0	-	-
Valine	18.00	361.09	0.999	0.993	0.964	5.33	111.33	3.6	8.4	-	-

[1]Average variation of the three batches, [2]Calculated from all analysed samples over the three days, [3]3-carboxy-4-methyl-5-propyl-2-furanopropanoic acid

Impact of replacing dietary chicken and pork with herring on human metabolism

Out of the 199 detected metabolites 46 were observed to best differentiate between the two diets in the OPLS-DA model (R^2X=0.528, Q^2=0.616). No outlier samples were found, and all time points for all subjects were kept in the analysis. From these 46 metabolites, eighteen metabolites were found to have significant diet x time interactions based on the follow-up mixed-model analysis. In order to determine if there were any pathway specific effects of replacing chicken/pork with herring, we performed a pathway over-representation analysis based on mapping the significantly changed metabolites onto the human metabolic pathway map which found that the TCA cycle, glyoxylate-dicarboxylate and arginine metabolism were highly enriched (-log10(p-values)>5) after the herring diets.

The herring diet led to seven metabolites increasing in comparison to the chicken and lean pork intake, while eleven metabolites increased on the meat diet relative to the herring diet. The herring diet led to increases (p value for diet x time interactions) of plasma asparagine (P=0.019), erythritol (P=0.020), ornithine (P=0.005), glutamine (P=0.012), glucosamine (P=0.005), shikimic acid (P=0.021) and p-coumaric acid (P=0.037). Of these only glucosamine was significant at both two weeks (P=0.016) and four weeks (P=0.039) whereas glutamine, ornithine and shikimic acid were found to be significant only after four weeks. Metabolites increased by the chicken/pork diet (p-value for diet x time interaction) were agmatine (P<0.001), cellobiose (P=0.045), citric acid (P=0.027), eicosanoic acid (P=0.009), fumaric acid (P=0.028), gluconic 1,4-lactone (P=0.045), glycolic acid (P=0.005), isocitric acid (P=0.002), oxalic acid (P=0.001), methylhistidine (P=0.001) and sorbose (P=0.041). Only isocitric acid and oxalic acid were significant at both two weeks (P=0.023, P=0.025) and four weeks (P = 0.029, P = 0.006) while citric acid and agmatine were only increased on the chicken/pork diet after four weeks. To aid in the interpretation of these changes, differences to the central carbon metabolites that were detected were mapped for the TCA and urea cycles (**Figure 6**).

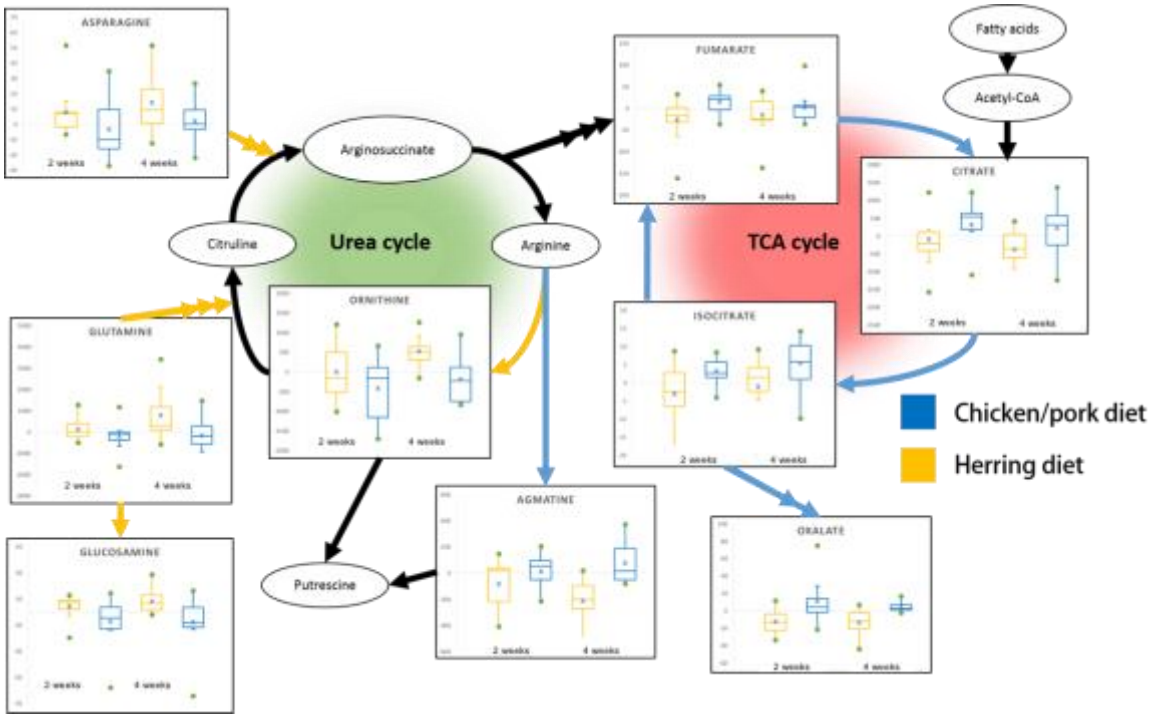

Figure 6. *Metabolite map of a selection of significant metabolites (box plots) differing between the herring and chicken/pork diet. The yellow arrows show the pathway of the metabolites increased by the herring diet and the blue lines for the metabolites increased by the chicken/pork diet. The black arrows indicate connections between metabolites where no intermediates were detected.*

Because metabolites in the TCA tended to be higher after the chicken/pork diet, and higher in the urea cycle after the herring diet, aside from agmatine, we hypothesised that there was an effect on arginine metabolism, with a possible downstream effect on nitric oxide, a cellular messenger synthesised from arginine by nitric oxide synthase. Nitric oxide concentrations tended to increase in plasma on the herring diet (P=0.09), with an interaction for gender. Analysis of each gender found that there was no effect of diet on nitric oxide concentrations in women but there was a significant difference in nitric oxide between diets in men after two weeks (P=0.03), and a tendency for a difference after four weeks (P=0.11). Conversely there were no significant differences for the amino acids analysed for men, while asparagine was higher after the herring diet in women (P<0.001), and ornithine tended to be higher after the herring diet (P=0.05). No differences for arginine or glutamine were observed.

Potential dietary and nutrient biomarkers and Type 2 diabetes

Associations between the potential diet and nutrient biomarkers and glucose tolerance status

β-alanine (P<0.001), alkylresorcinol homologues C17:0 (P=0.003) and C19:0 (P=0.011), CMPF (P=0.002), EPA (P=0.041), linoleic acid (P<0.001), oleic acid (P=0.003) and α-tocopherol (P<0.001) were different when comparing the three GTS groups with each other (**Table 5**). All of the

previously named biomarkers, except for CMPF, were significant when comparing subjects with NGT to subjects with IGT, linoleic acid and CMPF when comparing NGT to T2D and CMPF, linoleic acid, β-alanine and α-tocopherol when comparing IGT to T2D (**Table 5**). As illustrated by fold chance (FC) in Table 5, subjects with NGT were had lower circulating levels of β-alanine and CMPF compared to subjects with IGT and T2D. The two alkylresorcinols were higher in subjects with NGT compared to subjects with IGT, α-tocopherol higher in NGT subjects when compared to IGT and T2D subjects. Oleic acid and EPA were higher in subjects with NGT when compared to IGT subjects, linoleic acid higher in NGT when compared to IGT subjects and higher in T2D when compared to NGT and IGT subjects.

Baseline values of the potential biomarkers and future development of Type 2 diabetes

α-Tocopherol and CMPF were inversely associated with odds of developing T2D (P=0.01, OR: 0.72 (95% CI 0.56-0.93 AND P=0.04, OR: 0.71(95% CI 0.51-0.98, respectively).

Table 5. *P-values and odds ratios with 95% confidence intervals for the associations between the potential dietary and nutrient biomarkers and development of type 2 diabetes as well as ANCOVA P-values and post-hoc P-values with fold changes for the associations between the potential dietary and nutrient biomarkers with glucose tolerance status. Main dietary sources in Sweden.*

		Difference in glucose tolerance group							Confounders		Development of T2D	
			P-value			Fold Change						
Biomarker	Main dietary sources of the biomarker in Sweden	ANCOVA P	IGT-NGT	T2D-NGT	T2D-IGT	IGT/NGT	T2D/NGT	T2D/IGT	Stage 1	Stage 2	OR (95% CI)	P-value
β-Alanine	Circulating levels linked to consumption of meat	0.000	0.003	0	0.092	1.14	1.22	1.07	HDL	HDL	0.82(0.61-1.10)	0.185
Alkylresorcinol C17	Whole grain wheat and rye	0.003	0.001	0.073	0.288	0.87	0.93	1.06	FFA	FFA	0.86(0.63-1.14)	0.303
Alkylresorcinol C19	Whole grain wheat and rye	0.011	0.003	0.178	0.237	0.9	0.94	1.05	-	-	0.81(0.59-1.10)	0.19
CMPF	Fish	0.002	0.232	0	0.008	1.05	1.2	1.14	LDL	LDL	0.71(0.51-0.98)	0.041
EPA	Fish	0.041	0.013	0.145	0.92	0.93	0.93	1	CHOL, HDL, LDL	CHOL, LDL	1.06(0.80-1.40)	0.673
Lauric acid	Meat, dairy, coconut, palm kernel oil	0.05	0.033	0.861	0.093	1.05	1	0.95	FFA, HDL	FFA, HDL	1.35(1.02-1.81)	0.038
Linoleic acid	Seeds, nuts and vegetable oils	0	0	0.004	0.021	0.9	1.02	1.14	FFA, HDL	FFA, HDL	0.79(0.58-1.08)	0.135
Oleic acid	Olive and rapeseed oil, almond, hazelnuts, avocado	0.003	0.003	0.776	0.121	0.96	0.99	1.04	FFA, CHOL, HDL	FFA, CHOL, HDL	0.83(0.62-1.09)	0.186
α-Tocopherol	Pastry, margarine, eggs, bread, vegetables and fruits	0	0	0	0.983	0.84	0.84	1	CHOL, HDL, LDL, TG	CHOL, LDL	0.72(0.56-0.93)	0.013
γ-Tocopherol	Corn and soy bean, margarine	0.17	0.089	0.137	0.852	1.07	1.08	1.01	FFA, CHOL, HDL, LDL, TG, W	CHOL, HDL, LDL	1.00(0.77-1.29)	0.992

Prediction of Type 2 diabetes using metabolomics and potential dietary and nutrient biomarkers

Metabolomics

The comparisons of the glucose tolerance groups between each other found 41 significant (P_{corr}<0.05) metabolites between NGT and T2D subjects and 12 when comparing IGT and T2D subjects (**Table 6**). The two lists were combined and nine metabolites were selected for the prediction study based on lowest corrected P-value. These metabolites were sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid and 2-hydroxybutyric acid.

Table 6. *Significant metabolites with chemical class, P-value, corrected P-value (Pcorr) and odds ratios with 95% confidence intervals (OR (95% CI)) from pairwise comparisons of NGT and T2D and IGT and T2D using logistic regression models adjusted for analytical batch.*

		NGT vs. T2D			IGT vs. T2D		
Metabolites	Class	Pvalue	Pcorr	OR (95% CI)	Pvalue	Pcorr	OR (95% CI)
Alanine	Amino acid	-	-	-	0,005	0,039	1.57(1.16-2.19)
Asparagine	Amino acid	-	-	-	0,002	0,019	1.68(1.23-2.36)
Cystine	Amino acid	0,009	0,027	1.46(1.10-1.95)	-	-	-
Glutamic acid	Amino acid	0,001	0,004	1.67(1.24-2.28)	-	-	-
Glutamine	Amino acid	0,001	0,004	1.68(1.25-2.33)	-	-	-
Lysine	Amino acid	0,000	0,000	1.85(1.37-2.52)	-	-	-
Phenylalanine	Amino acid	0,012	0,032	1.47(1.09-1.94)	-	-	-
Proline	Amino acid	0,004	0,012	1.55(1.16-2.11)	-	-	-
Sarcosine	Amino acid	0,014	0,036	1.42(1.08-1.88)	-	-	-
β-Alanine	Amino acid	0,002	0,006	1.75(1.24-2.53)	-	-	-
Isoleucine	BCAA	0,000	0,000	1.92(1.42-2.65)	-	-	-
Leucine	BCAA	0,000	0,000	1.89(1.41-2.58)	-	-	-
Valine	BCAA	0,001	0,004	1.64(1.23-2.23)	-	-	-
2-Methylmalic acid	Dicarboxylic acid	0,000	0,000	0.53(0.38-0.70)	-	-	-
Fumaric acid	Dicarboxylic acid	0,001	0,002	1.66(1.25-2.24)	-	-	-
Malonic acid	Dicarboxylic acid	0,010	0,028	1.50(1.11-2.05)	-	-	-
Oxalic acid*	Dicarboxylic acid	0,000	0,000	2.99(2.08-4.48)	0,003	0,027	1.64(1.20-2.30)
Eicosapentaenoic acid	Fatty acid	0,003	0,011	0.64(0.47-0.86)	-	-	-
Octadecanoic acid	Fatty acid	0,008	0,025	1.45(1.10-1.93)	-	-	-
Lactate	Hydroxy acid	0,002	0,007	1.59(1.20-2.14)	-	-	-
2-Hydroxybutyric acid*	Hydroxybutyric acid	0,000	0,000	3.72(2.19-6.89)	-	-	-
3-Hydroxybutyric acid	Hydroxybutyric acid	0,000	0,000	2.17(1.53-3.22)	-	-	-
α-ketoglutaric acid	Keto acid	0,000	0,000	1.90(1.41-2.63)	-	-	-
3-Methyl-2-oxopentanoic acid*	Keto acid (metabolite of isoleucine)	0,000	0,000	2.20(1.61-3.08)	-	-	-
Ketoleucine	Keto acid (metabolite of leucine)	0,001	0,004	1.66(1.24-2.26)	-	-	-
1,5-Anhydro glucitol	Monosaccharide	0,015	0,038	0.70(0.53-0.93)	-	-	-
Arabinose	Monosaccharide	-	-	-	0,008	0,049	1.57(1.14-2.22)
Fructose	Monosaccharide	0,000	0,000	2.11(1.53-2.98)	-	-	-
Galactose*	Monosaccharide	0,000	0,000	2.54(1.74-3.87)	0,000	0,006	1.91(1.36-2.76)
Glucaric acid-1,4-lactone*	Monosaccharide	0,000	0,000	5.40(3.33-9.25)	0,007	0,049	1.80(1.192-2.82)
Mannose*	Monosaccharide	0,000	0,000	4.40(2.81-7.30)	0,000	0,003	2.07(1.46-3.03)
Inosine	Nucleoside	0,008	0,025	1.69(1.18-2.59)	0,004	0,035	2.56(1.46-5.30)
cis-Aconitic acid	Organic acid	0,014	0,036	1.48(1.10-2.05)	-	-	-
Uric acid*	Purine	0,000	0,000	3.44(2.34-5.29)	0,002	0,019	1.68(1.23-2.35)
Cholesterol	Sterol	-	-	-	0,001	0,013	0.58(0.42-0.80)
Galacticol*	Sugar alcohol	0,000	0,000	8.72(5.00-16.60)	0,000	0,003	2.10(1.49-3.07)
Sorbitol*	Sugar alcohol	0,000	0,000	13.58(7.13-28.70)	0,000	0,000	2.98(1.98-4.72)
Glucose-6-phosphate	Sugar phosphate	0,018	0,043	0.70(0.51-0.93)	-	-	-
γ-Tocopherol	Tocopherol	0,020	0,047	1.44(1.07-1.99)	-	-	-
α-Tocopherol (MRM)	Tocopherol	0,000	0,000	0.34(0.23-0.49)	-	-	-
Tyrosine	Tyrosine	0,000	0,000	1.87(1.39-2.56)	-	-	-

*Metabolites used for prediction of development of type 2 diabetes

Prediction of Type 2 diabetes using metabolomics

From the tested models the NB+AdM model, combining adiponectin (AdM) and potential dietary and nutrient biomarkers (NB) models had the highest selectivity and sensitivity (AROC=0.85[0.81-0.90]) followed by model AdM+M combining adiponectin (AdM) and metabolomics model (MM) (AROC=0.81[0.75-0.87]) and adiponectin model (AdM) (AROC=0.79[0.74-0.85]) (**Table 7**)(**Figure 7**). The potential dietary and nutrition biomarkers model (NB) alone provided a sensitivity and selectivity (AROC=0.72[0.66-0.78]) that was higher than that of metabolomics model (MM) (AROC=0.66[0.57-0.74]). Anthropometry model (AM) (AROC=0.64[0.56-0.71]) and optimized metabolomics model (AROC=0.66[0.58-0.73]) performed similarly to the MM model. Furthermore, the addition of baseline values of IFG and IGT to the optimized metabolomics model further increased sensitivity and selectivity over the optimized metabolomics model only (AROC=0.78[0.71-0.84]).

***Figure 7**. AROC (AROC(95% CI)) for predicting incident T2D during the 5.5 year follow up including a) smoking, alcohol consumption, waist circumference, systolic blood pressure and heredity, b) potential dietary and nutrient biomarkers, c) serum adiponectin concentration, HOMA-IR, smoking and IGT and IFG at baseline and d) a combination of b and c, as variables in the model.*

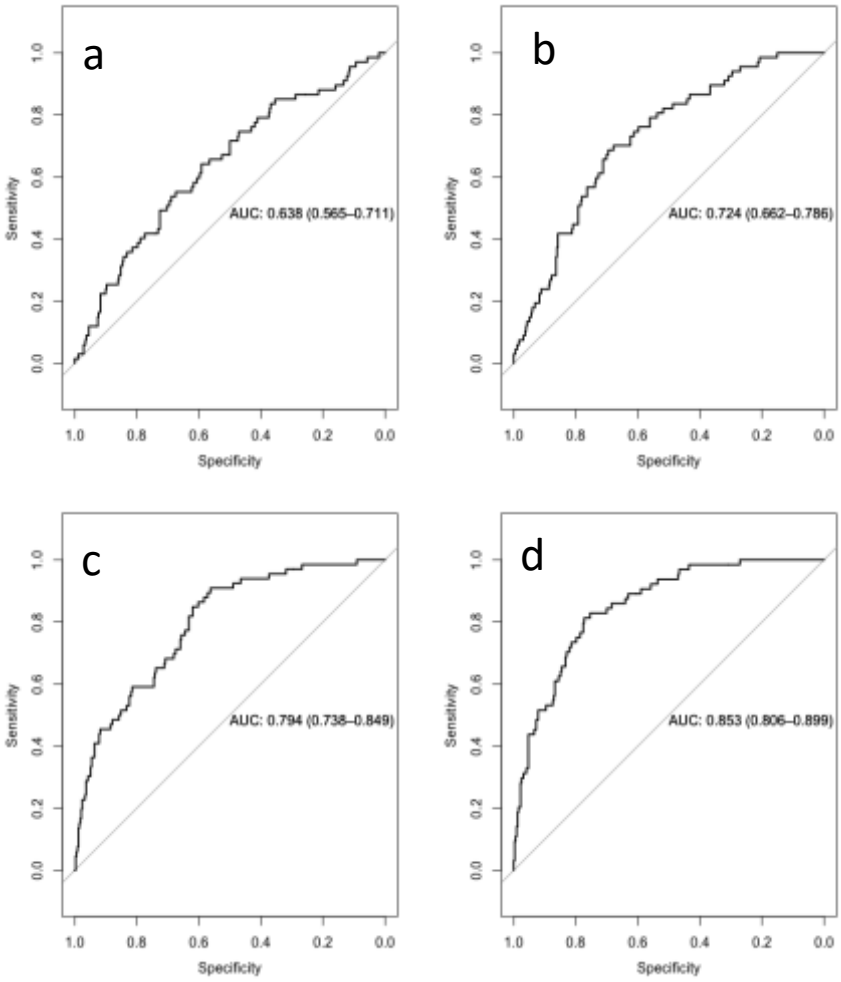

Table 7. *Different prediction models used for testing of prediction of development of T2D with abbreviation, AROC with 95% confidence interval, percentage increase in AROC in comparison to anthropometry model and variables used for the prediction.*

Model	Abbreviation	AROC (95% CI)	% increase in AROC	Variables
Dietary and nutrient biomarkers with adiponectin model	NB+AdM	0.853(0.806-0.899)	33.7	NB + AdM
Adiponectin + metabolomics	AdM+M	0.808(0.750-0.867)	26.6	AdM+MM
Adiponectin	AdM	0.794(0.738-0.850)	24.5	serum adiponectin, HOMA IR, smoking, IGT, IFG
Optimized metabolomics with glucose measures	OM2	0.776(0.713-0.839)	21.6	OM + IFG, IGT
Dietary and nutrient biomarkers	NB	0.724(0.662-0.776)	13.5	α- and γ-tocopherol, alkylresorcinols C17 and C19, β-alanine, 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF), lauric acid, oleic acid, eicosapentaenoic acid and linoleic acid
Anthropometry + metabolomics	AM+M	0.700(0.626-0.773)	9.7	AM + sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid, 2-hydroxybutyric acid
Metabolomics	MM	0.657(0.577-0.736)	3.0	sorbitol, galacticol, mannose, galactose, uric acid, oxalic acid, glucaric acid-1,4-lactone, 3-methyl-2-oxopentanoic acid, 2-hydroxybutyric acid
Optimized metabolomics	OM	0.656(0.579-0.734)	2.8	galacticol, mannose, galactose and 2-hydroxybutyric acid
Adiponectin w/o glucose measures	AdM2	0.655(0.580-0.729)	2.7	serum adiponectin, HOMA IR, smoking
Anthropometry	AM	0.638(0.565-0.711)	-	waist circumference, alcohol consumption, smoking, systolic blood pressure, family history of T2D

[1]Increase in AROC as percentage in comparison to anthropometry model

DISCUSSION

Simultaneous acquisition of targeted and untargeted metabolomics data

Spectral Features in Untargeted Data

Overall the method development was successful and provided necessary sensitivity to cover biological concentration ranges of the targeted MRM analytes, comparable number of detected features to that of earlier published methods and sufficient reproducibility of both targeted and untargeted data. The number of spectral features detected with an untargeted metabolomics method has in many times been used as an indicator for the goodness of the method (195) and our findings on the number of detected spectral features using different reproducibility thresholds are comparable to that of earlier reported for GC-MS metabolomics using TOF instruments (15). For providing a more realistic and comparable picture on the number of spectral features it is necessary to take data processing method into account as the results are affected strongly by these parameters. For example, in our development the number of features was reduced from 2979 to 1054 with an application of a minimum intensity filter. Furthermore, the application of a reproducibility threshold filtering for controlling both intra- and inter batch variation, based on guidelines such as FDA guideline for biological method validation (62) is necessary for only including the features that can provide meaningful results without excess analytical variation, which can be of special importance in untargeted metabolomics that is often followed by resource demanding identification of unknowns (196).

Data Collection Rate

In metabolomics studies data collection rate affects reproducibility via inconsistent peak shape if insufficient data collection rates are used. As a rule of thumb data collection rate should be high enough for providing at least 10 data points per peak. In this method development, the data collection rate was determined by the loop time consisting of both the time used in acquisition of the untargeted scanning data and the time measuring the multiple MRMs (**Figure 8**). After optimisation of the time spent measuring scan and MRM data a loop time of 120 ms was used in the final method providing a data collection rate of 8.22 Hz which equals to 10 data points covering 1.22 s, which provides sufficient data points for most GC peaks in this method. The data collection rate is lower than 20-30 Hz what has been used in reported methods based on TOF instruments (15,16,183) although some TOF methods have also reported lower acquisition rates of 1-3 Hz (19,20,23). However, the use of data filtering based on reproducibility threshold removes peaks with greater analytical variation and, thus, the remaining features present good, usable data for further data analyses.

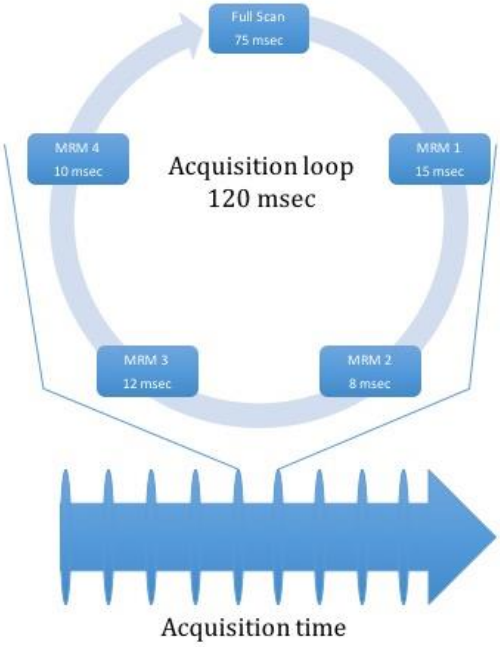

Figure 8. *The concept of simultaneous acquisition of non-targeted full scan data and targeted quantitative data using multiple reaction monitoring (Paper I).*

Linearity and Limits of Detection

Assessment of linearity was done both by using standard solutions and by using different volumes of plasma in the sample preparation for mimicking concentration changes. Furthermore, for linearity in plasma matrix, we used both MRM and scanning data for linearity assessment. As expected the overall best linearity was obtained from the MRM data measured from standards followed by MRM measurement of plasma and full scan measurement of plasma. Further testing of LOD by using MRM and scanning data supported these findings as the LOD values for MRM data acquisition were in general lower than for scan data acquisition. In the plasma matrix both α- and γ-tocopherol had inverse U shape calibration curves with reduced responses when higher plasma volumes were used. A likely reason for this observation is an interfering compound in the plasma matrix as this phenomenon was only observed in connection with the two tocopherols and there is a 10-fold difference in biological concentration of these (197-199). Another possible explanation is a limited dynamic range of fast scanning quadrupole GC-MS (30), though the concentration range tested for the two tocopherols was not especially wide. The added mass separation dimension when utilising MRM data acquisition provides a more selective detection of the target molecules by removing interfering molecules during the precursor selection – collision induced dissociation – product ion selection cycle in comparison to full scan data acquisition where all the ions within the acquisition mass range are being transferred to the detector. This was seen as an overall improved performance of the targeted MRM part of the method over the untargeted scanning part of the method. However, data processing methods can significantly affect this comparison and an optimal comparison taking all variables into consideration is hard to establish.

Precision and Accuracy

Precision of the method was assessed both within and between batches for providing a realistic picture on how reliable data the method produces in real life analysis conditions. From the two types of variations within batch precision gives a figure on how well the analytical method is performing for analysing one batch of samples and between batch precision describes how well the measured analytes can be detected over a longer time period. Both within and between batch variation were acceptable for all metabolites (<20 %), except for tryptophan, dimethylarginine and γ-tocopherol having higher between batch variation. Dimethylarginine and γ-tocopherol are present at blood plasma at lower concentrations than most of the other metabolites (low µmol-nmol/L range) (197-199) which likely contributes to the higher between batch variation. Tryptophan, on the other hand, is relatively abundant in plasma (197-199) but its stability is a known issue in GC-MS analysis (200). Accuracy based on measurement of National Institute of Standards and Technology (NIST) reference plasma and comparison to reported reference values was in most cases good except for serine, histidine and γ-tocopherol. However, as precision for these molecules was good, accurate estimation, although not often needed in metabolomics studies, could be achieved with application of a correction factor.

What does the new GC-MS/MS metabolomics method add to the field?

The new GC-MS/MS method developed as part of this thesis demonstrates the utility of combined collection of targeted quantitative and untargeted qualitative data increasing the limits of what can be achieved with one analytical injection. The method can also be adapted and expanded by adding more targeted analytes into the method to fit the sample set at hand. As for all methods, there are also several limitations with this way of acquiring metabolomics data. While it was clear that MRM acquisition improves sensitivity of the detection, the utility of MRM is limited by EI fragmentation that in most cases does not produce high intensity molecular ions necessary for very sensitive MRM. Previous metabolomics methods utilising GC-MS/MS (14,201,202) have only used a targeted metabolomics approach and use chemical ionization (CI) that, as a softer ionization technique, provides stronger molecular ions that can improve sensitivity of the method. Virtually all the large spectral libraries for GC-MS have been constructer using EI limiting usability of CI in untargeted metabolomics. In addition to CI, APCI in combination with a HRMS has been used for metabolite profiling (19,20) enabling metabolite identification with the combination of accurate mass information of the molecular ions and fragmentation information when instruments with collision cell are used. However, the utilisation of multiple MRM acquisition with modern fast GC-MS/MS instrument makes it possible, in theory, to measure thousands of molecules simultaneously in one chromatographic run. The application of such a multiMRM method would further break the barriers of untargeted and targeted metabolomics as in principle all the known detectable molecules from a sample type previously detected using untargeted metabolomics could be converted into MRMs and used in a 'targeted' metabolomics method that would be much larger in scale than current methods. However this would still exclude the possibility of detection of novel molecules in studies where this is likely, returning to the need to choose the most appropriate method(s) for each study.

While the data collection rate was sufficient for peaks slightly wider than 1s, narrower peaks of for example 0.5 s are not reliably detected with the current method due to insufficient data collection rate. Faster scan speeds exceeding 10 000 Da/s led to reduced number of spectral features in the scanning data and, thus, the current loop time is a compromise between providing sufficient time for MRM and for scan data collection. A two-dimensional GC-MS metabolomics method utilizing similar fast scanning quadrupole instrument than in the current development has demonstrated the use of 20 000 Da/s scan speeds in metabolomics analysis (30) encouraging further exploration of applicability of even faster scan rates in other applications. Further, the added resolution of the second chromatographic dimension in 2D-GC-MS aids in resolving co-eluting molecules (203) though the usability 2D-GC-MS/MS approach in metabolomics use has not been yet evaluated. Additionally our results clearly demonstrate that more isotopically labelled internal standards could substantially improve precision. Here, the possibility of using labelled derivatization reagents for generating labelled internal standard counterparts for virtually to all molecules present in the sample is tempting as demonstrated before with a limited set of molecules (14,202,204).

The division of metabolomics methods into 'targeted' and 'untargeted' can be a limiting factor for researchers using metabolomics, as both cost and sample availability often limits what can be analysed. If a researcher chooses to analyse known biomarkers relevant to the research question, they remove the possibility of new findings from their samples, but will have data that can be easily compared with previous literature. If they choose targeted metabolomics analysis, they will limit their findings to those compound classes covered by the targeted metabolomics method, but will also have quantitative or semi-quantitative data that can be compared with previous literature. If they choose an untargeted metabolomics method, they will have the potential to see the widest range of new findings from their data, but will lose sensitivity compared to targeted methods, and may have difficulty in making direct comparisons with past literature due to untargeted metabolomics not being quantitative. Because of these differences in what can be obtained from different types of metabolomics, drawing conclusions based on several published studies is not straightforward due to difficulties in comparing data between studies. Therefore, verification of metabolomics results using appropriate quantitative measures is of importance for making across studies comparisons possible. Our novel GC-MS/MS simultaneous targeted and untargeted method is a step towards getting the best of both worlds, which can add value for researchers looking to get the most from one sample injection, though the general problems of comparability with different methods still remain.

Different effects of meat and fish diet on metabolism as determined by GC-MS metabolomics

Of the eighteen metabolites that differed between the herring and chicken/pork diets, many were involved in either the tricarboxylic acid cycle or the urea cycle, the first being a major pathway for energy metabolism and for providing building blocks for biosynthesis of both amino acids and lipids, while the latter is primarily involved in converting ammonia from amino acid metabolism to urea. TCA cycle intermediates that were detected by the GC-MS method were higher after the chicken and pork intervention, while urea cycle metabolites were higher after the herring diet. Agmatine and ornithine, both metabolites of the amino acid arginine, responded differently to the two diets – the former being relatively higher after the chicken and pork diet, and the latter higher after the herring diet, suggesting that the two diets differ in their effect on arginine metabolism. As arginine is also a precursor to nitric oxide, an important cellular regulator (205), we hypothesised that nitric oxide would be different between the two diets. Analysis of nitric oxide found that the herring diet increased concentrations in men, but not women, in part supporting the metabolomics results. One uncertainty in this analysis is that the GC-MS method used leads to partial conversion of arginine to ornithine, potentially confounding interpretation of this result. Amino acids were analysed using LC-MS, with results that were in part inconsistent with the GC-MS results. This inconsistency may stem from a difference in sample preparation and that the samples for GC-MS and LC-MS were prepared and analysed nearly two years apart, which may have given time for partial degradation of the sample. The overall consistency of the GC-MS metabolomics results between the two pathways suggests that these findings are not random, though further work is needed to confirm the findings and to understand what component or factor in the two different diets led to these different results. The only other study to investigate the longer-term effects of eating fish and meat on metabolism found that urinary excretion of several metabolites related to mitochondrial energy production was lower after eating lean fish instead of meat (91), supporting an effect of replacing meat with fish on overall energy metabolism.

Earlier work on samples from this study found that concentrations of the n-3 fatty acids EPA and DHA were increased after the herring diet, but not after the chicken/pork diet (184), confirming the often repeated observation that fatty fish increase n-3 fatty acids in plasma. In Paper II we found that methylhistidine, either a metabolite of anserine (1-methylhistidine) or a product of protein breakdown (3-methylhistidine) was elevated after the chicken and pork diet compared to the herring diet. Previously urinary methylhistidine was proposed as a marker for meat intake (206), and our finding is the first that we are aware of to demonstrate that plasma concentrations of methylhistidine may also be useful as a marker of chicken and pork intake. Recently plasma 3-methylhistidine was found to be increased after a chicken meal, but relatively unchanged after red meat or processed meat (89), suggesting that the rise in methylhistidine is more likely due to the chicken in the chicken/pork diet. These results from a small dietary intervention study support the ability of the GC-MS/MS method, in this case only utilising the full scan data, to be applied to intervention studies to identify metabolic mechanisms.

Associations of potential dietary and nutrient biomarkers with glucose tolerance status and development of type 2 diabetes

β-alanine and 3-carboxy-4-methyl-5-propyl-2-furanopropanoic acid (CMPF)

The metabolites β-alanine was positively associated with diabetes classification suggesting that a greater circulating β-alanine is related to T2D. The concentration of β-alanine is higher in meat as it is in other common protein sources including fish and chicken and recently it was found that circulating levels β-alanine markedly increased after a beef-based meal, but not after a herring-based meal (92). β-alanine is also a breakdown product of carnosine which is a vital molecule for normal muscle function and, thus, β-alanine can also be used as a supplement to improve physical capacity by increasing its carnosine content (207,208).

The furan fatty acid metabolite CMPF, previously associated with fish intake (42) was positively associated with GTS and inversely with odds of T2D. The positive association with GTS suggest a loss of glucose control in the subjects having higher circulating concentrations of CMPF. This is in line with recent results suggesting increased circulating levels of CMPF to lead to the loss of glucose control via causing β-cell dysfunction (126). Also, as in our study, elevated baseline concentrations of CMPF have been recently found in subjects developing diabetes in a prospective study (127). On the other hand, the major dietary source of CMPF is fatty fish and circulating levels of CMPF can be used to distinguish high and low fish consumers (42). In the same population including men and women with characteristics of metabolic syndrome and IGT, associations between IGT and CMPF were not found (128). Furthermore, our finding of inverse associations between CMPF and glucose tolerance group suggest that increased levels of CMPF are not necessarily indicative of T2D risk. Although these results seem to point in different directions, it could also be an indication that CMPF as a marker could indicate two separate situations depending on the population; as a possible biomarker of fish intake, suggesting healthier dietary choices being protective against T2D, and in populations with low fish intake, as an indicator of β-cell dysfunction leading to greater risk of T2D.

Fatty acids

Of the 3 fatty acids associated with GTS, EPA and linoleic acid are both food derived polyunsaturated fatty acids (PUFAs), major dietary sources for these fatty acids being fatty fish for EPA and vegetable oils, seeds and nuts for linoleic acid, all of which are reflecting a balanced and healthy diet. EPA is an n-3 PUFA found in fatty fish and recommended for improved cardiometabolic health (209) while linoleic acid has been suggested to have adverse health effects (210-212). In the current study, we found both EPA and linoleic acid to be negatively associated with glucose tolerance status pointing towards increased proportion of these fatty acids in the diet being protective against impaired glucose tolerance and T2D. Oleic acid is a monounsaturated fatty acid present in both animal and plant derived foods, olive and rapeseed oil being especially

rich in it (213). In type 2 diabetic men, diet rich in oleic acid has been shown to yield lower fasting glucose and insulin levels as well as lower plasma cholesterol and LDL in comparison to diet rich in linoleic acid (214). As with EPA and linoleic acid, we found negative association between oleic acid and glucose tolerance suggesting, similarly to EPA and linoleic acid, a positive effect of higher proportion of oleic acid containing foods in diet.

Alkylresorcinols

Alkylresorcinols are strongly associated with whole grain wheat and rye intake in several studies (159,165,173) and two of the homologues, C17:0 and C19:0, were inversely associated with GTS supporting the status of whole grains in healthy nutrition. Previously, associations between total circulating concentrations of alkylresorcinols and T2D were not found, and rather the ratio of C17:0 and C21:0 homologues, depicting a diet rich in rye, was inversely associated with future T2D (215), fasting insulin concentration (169) and positively associated with insulin sensitivity (169). Also, alkylresorcinols have been found to improve insulin sensitivity in rats (216) though no direct studies for effects in humans have been undertaken and the results presented here are the first describing associations between circulating alkylresorcinols and GTS. In this metabolomics method, AR homologues C21:0 and above are difficult to detect in plasma, even if they are often present in higher concentrations than C17:0 and C19:0 (217), likely due to the generic extraction method used, which is not optimal for highly lipophilic compounds.

α-tocopherol

Circulating tocopherols have been associated with either intake of foods rich in tocopherols or with supplementation with tocopherols, though the association are stronger when supplementation is taken into account (218) and recently correlations between α-tocopherol intake and circulating concentrations were found using metabolomics (219). The positive association of α-tocopherol both with GTS and development of T2D suggests a positive effect of vitamin E in relation to glucose metabolism and a protective effect against development T2D. A major dietary source of α-tocopherol in this population in sunflower and olive oils, which are also part of the local dietary guidelines (220). In the cohort used in this study consisting of older women supplementation has to also be taken into account as a possible confounder. The body of literature suggesting possible protective effect of vitamin E intake and supplementation against T2D complications is not conclusive (221-225) and it has been suggested that a modest supplementation does not protect against the disease (221,226). As the finding that circulating α-tocopherol was negatively associated with GTS and T2D, further study should be undertaken to investigate whether supplementation, diet, or other factors explain this finding in this population.

What could dietary biomarkers explain in relation to type 2 diabetes?

Eight of the ten measured biomarkers were associated with GTS and 2 with odds of developing T2D. The associations were generally robust and not attenuated by common risk factors of CVD and T2D including obesity suggesting that these food components may have a direct association

with the disease not mediated by obesity or other related variables. Overall the results suggest that a diet rich in whole grains, fatty fish and vegetable oils is associated with a healthier GTS whereas higher consumption of meat is associated with worse GTS – matching healthy diet guidelines of American Diabetes Association (227) and Nordic Nutrition Recommendations (220).

A limitation of the current study is the lack of dietary intake data in the cohort which makes cross-comparison of the biomarker data with the intake data impossible and limits the discovery of associations between GTS and diet to the selection of the potential biomarkers measured by the metabolomics method. Notwithstanding the biomarker panel measured in Paper V captures some important aspects of healthy nutrition such as whole grain and fish intake and vitamin E status. However, the use of biomarkers is, however, not without controversy as the concentration of several of the biomarkers can be impacted by endogenous metabolism, further complicating the association between the biomarkers and GTS. An example of this is the endogenous production of EPA and docosahexaenoic acid (DHA) from α-linolenic acid with a conversion rate in males 0.5 % and 5 % respectively (228) although a mechanism for increased production of EPA and DHA α-linolenic acid seems to exist as when the need for these is high, such as in pregnancy, the production capacity can increase (229). Thus, the underlying metabolism of the biomarkers can affect the robustness of the biomarker likely not affecting population level generalisation, but limiting their use at individual level if the effect of metabolism and potential endogenous production rate is not known. These limitations are well known, although more research is required to accurately describe them so that they can be accounted for in future dietary biomarker research, as no dietary biomarker will be entirely free of this type of confounding. Not all potential dietary biomarkers measurable with the GC-MS/MS method were associated with any effects on T2D diabetes. Methylhistidine, associated with greater intake of chicken and pork (Paper II), and likely associated with chicken intake (89) was not found to differ between any of the glucose tolerance groups. In this context it would suggest that chicken has no effect on GTS in this population. This study is the first to use potential dietary biomarkers to estimate the influence of diet on disease without using dietary intake data, which is on one hand a major limitation, though also represents a tentative step forward where the relationship between diet and disease is not necessarily dependent on subjective measures of dietary intake.

Although to date only a few potential dietary and nutrient biomarkers have been extensively validated, there are many candidates that can be detected in plasma. In the development of our GC-MS/MS method, we selected several of these that we knew to be detectable using GC-MS for specific detection using MRM and SIM, including alkylresorcinols C17:0 and C19:0, CMPF, β-alanine and the two E vitamers α- and γ-tocopherol. In addition the data processing method used in the work presented in this thesis also detected several other compounds that have been associated with foods or dietary intake, including methylhistidine, EPA and oleic, linoleic and lauric acid. These were identified based on a combination of mass spectral matching and retention index. Not all potential diet and nutrient biomarker compounds of potential interest were included, in part because several such as proline betaine and carotenoids were not compatible with the used analysis method, and in part because the goal was to develop a metabolomics method rather than a method that would specifically target diet and nutrient biomarkers. To our

surprise many of the compounds detected by the method were strongly associated with glucose tolerance in the DIWA cohort underlining the perhaps underrated potential of using dietary and nutrient biomarkers in this type of observational trial and the use of metabolomics to detect potential food and diet related biomarkers in studies where dietary intake data is missing.

Prediction of future Type 2 diabetes using metabolomics

Sensitivity and selectivity

Based on the results presented in this thesis prediction of the development of T2D within the 5.5 year follow-up period with reasonable sensitivity and specificity is possible using different prediction models. In general, the addition of metabolomics over established predictors of T2D improved the prediction. Setting the anthropometric model as a reference, the highest increase in AROC (+33.7 %) was achieved with the model combining the potential dietary and nutrient biomarkers with the adiponectin model followed by the model combining adiponectin and metabolomics (+26.6 % in AROC) closely followed by the adiponectin model (+ 24.5 % in AROC). The metabolomics model gave similar prediction than the reference model with 3.0 % increase in AROC and the addition of the preconditions of the disease, IFG and IGT, into the optimized metabolomics model improved the prediction yielding a 21.6 % increase in AROC in comparison to the reference model and similar AROC to the adiponectin model. The biomarkers alone increased the AROC by 13.5 % in comparison to the reference model. The AROC values found in this research compare well to previous studies utilizing prospective data and metabolomics where AROC values have ranged between 0.52 and 0.91 depending on the cohort and the variables used for prediction (**Table 1**).

Can metabolomics improve on existing measures of type 2 diabetes risk?

In this cohort of older Swedish women, the prediabetic states, IFG and IGT, were among the strongest predictors of incident T2D, in line with the literature that links pre-diabetes with a strong risk of developing T2D (102). The addition IFG and IGT into the optimised metabolomics models increased the strength of prediction, while their removal from the previously established adiponectin-based predictive model decreased AROC-based prediction. This is not surprising as IFG and IGT are directly related to circulating glucose concentrations that are the main diagnostic tools of T2D making them by their nature predictive for T2D. Diagnosis of T2D is complicated by the application of a binary diagnosis to a disease that exists along a continuum of increasing glucose tolerance. Thus prediction of glucose tolerance based on three discrete groups may not be ideal for developing a predictive marker. Similarly in this cohort, there is likely to be some confounding in the T2D group based on those who made lifestyle changes because of the T2D diagnosis at baseline. These factors could make finding the perfect universal biomarker for T2D difficult, and potentially impossible. Recent work on the concept of personalised nutrition found that individuals had widely varying predictors of blood glucose response, based on differences in gut microbiota, anthropometry and disease risk biomarkers in blood (230), supporting the idea that a perfect universal predictor of T2D may be unobtainable. Lower circulating levels of

adiponectin have been frequently associated with future T2D (231) and its use in prediction of the disease in the cohort used in this thesis has also been exploited before (110). Unlike IFG and IGT, that are based on cut-off values of fasting and 2h-glucose concentrations, adiponectin is a hormone that improves insulin sensitivity in muscle and liver (232). It is synthetized in adipocytes and its synthesis is decreased in obesity (231) and so has a logical connection with T2D development. The effectiveness of food intake and nutrient status biomarkers in prediction of T2D is interesting as they present a group of molecules not directly related to the molecular development of the disease and in this cohort were not confounded by other risk factors of T2D, instead reflecting a part of the subject's lifestyle. Diet is one of the important modifiable lifestyle factors that can be used for prevention of T2D (233). Consumption of whole grains, fruits, vegetables, legumes, and nuts, moderate in alcohol consumption and lower in refined grains, red or processed meats, and sugar-sweetened beverages have been linked to reduced risk of diabetes (234) and are part of recommendations of American Diabetes association (227) and Nordic Nutrition Recommendations (220). The results presented in this thesis support these guidelines.

The cohort used for this study was relatively homogenous as all subjects were women of the same age and from the same city, although there were socioeconomic differences. Thus, the prediction models are expected to work well in this specific population and generalization of these results into the general population is limited. Also, the cohort overlooks any possible new T2D cases after the single 5.5 year follow-up examination that a second follow-up examination would capture. As for the cohort, the method used for metabolomics analyses needs to be considered in relation to the discovered biomarkers as different approaches will lead to a different discovered set of candidate biomarkers and to different prediction. Of the previous work using metabolomics for T2D (**Table 1**), all except one have used targeted metabolomics methods and targeting similar metabolite pools, limiting their ability to find new biomarker associations. The simultaneous targeted and untargeted GC-MS/MS method used in this thesis made it possible to detect a wider range of compound classes, though targeted methods can provide excellent coverage of certain molecule classes. Needless to say that the use of untargeted LC-MS metabolomics would lead to higher number of likely less precise candidate biomarkers whereas the use of 2D-GC-MS would likely offer the best selectivity for a number of small molecules of the available methods. Overall, the use of several metabolomics methods in the same cohort would allow for wider screen of candidate biomarkers than any single method and should thus be encouraged.

One goal of discovering new predictive markers of T2D is to screen and discover people at risk of T2D early enough for them to make lifestyle changes to prevent progression to T2D. The traditional diagnostic tools of T2D, insulin and OGTT, are not surprisingly among the strongest predictors of future T2D but it can be argued that other tools, such as tests derived from metabolomics, might be better suited for population-wide screening as prediction based on a single fasting blood sample is less of a burden for the patient and saves valuable time at healthcare clinics. An additional advantage of metabolomics if used in population screening, is that it could detect risk of other diseases risks as well as T2D with the use of the same sample and method. On the other hand, as the existing predictors offer sufficient detection of the subjects at risk, finding new tools for prediction of T2D is less of an issue than their effective establishment in the health

care systems globally. Therefore, the question to ask is what is sufficient level of certainty in the detection of subjects at risk of developing T2D and perhaps more crucially, and well beyond the remit of this thesis, how do we motivate these subjects to make the lifestyle changes needed to reverse the development of T2D? If metabolomics is to be increasingly used in T2D prevention priority needs to be given to testing of the current metabolomics findings in the literature in the available prospective cohorts for achieving a consensus on the effectiveness of these in the general population.

CONCLUSIONS

- A rapid metabolomics method for simultaneous acquisition of targeted quantitative and untargeted qualitative data utilizing a fast scanning GC-MS/MS instrument was developed and evaluated, and found to have good performance characteristics.

- In samples from a human dietary intervention study, the method was able to determine novel differences in the metabolic effects of herring diet compared to chicken and pork diet, suggesting that the interface between the TCA and urea cycles is metabolically sensitive to changes in dietary protein source, and may have downstream effects on nitric oxide production.

- Diet, as seen by a panel of potential dietary and nutrient biomarkers, was associated with both glucose tolerance status and development of type 2 diabetes, independent of common risk factors for cardiovascular diseases and type 2 diabetes, including obesity. The result supports current dietary guidelines for prevention of type 2 diabetes.

- The use of multiple potential dietary biomarkers as an objective tool for estimating dietary intake seems promising and likely can capture elements of nutrition important for human health.

- Prediction of development of type 2 diabetes using GC-MS/MS metabolomics was possible with similar or improved prediction in comparison to established predictors of type 2 diabetes. Prediction based on metabolomics in combination with other predictors of type 2 diabetes yielded comparable sensitivity and selectivity to previously published prediction models utilising metabolomics.

FUTURE RESEARCH

- Further expansion of the method presented in this thesis with the addition of targeted MRM analytes should be done. Evaluation of normalisation methods for improving precision by using more labelled internal standards or labelled biomass should be tested.

- Increasing use of metabolomics in large cohort studies can expand the potential to link lifestyle, disease mechanisms and health outcomes. For this to become widely feasible, robust, most preferably quantitative metabolomics methods suitable for large scale studies are needed. Further application of methods allowing review of metabolites in a single-cell will consequently lead to improved understanding of cellular mechanisms.

- Large cohorts and intervention studies with food intake data in different populations are needed for both discovery and validation of potential dietary biomarkers. A natural continuum of the discovery work to validation work should be established for evaluating the selectivity and sensitivity of the biomarkers, possible quantitative relationship between intake of the food item and the biomarker as well as usability of the biomarker in epidemiological use.

- The use of dietary biomarkers together with food intake data in nutritional epidemiology should be encouraged for overcoming some of the limitations of more traditional methods of dietary intake assessment.

- The integration of dietary biomarkers into metabolomics is a promising approach to understand the relationship between diet and molecular mechanisms, especially in cases where no dietary data is available. Further work should test this approach in cohorts where dietary data is available to confirm a relationship between diet and the biomarkers measured using this GC-MS/MS method, and look to develop a separate method focused on multiple dietary biomarkers.

- The predictive ability of the array of predictive metabolites of type 2 diabetes reported in the literature including the metabolites listed in this thesis should be evaluated using available prospective cohorts.

ACKNOWLEDGMENTS

There are many people that have helped me along the way to completion of this PhD thesis and I would like to express my sincere gratitude to all of you!

First I want to thank my supervisors Alastair Ross and Ann-Sofie Sandberg for recruiting me to do this work and for all the support and encouragement on the way to this point. Thank you for the freedom you have given me and for allowing me to make all the mistakes, that could have most probably been avoided, and providing me with all the time and resources that I needed for completing this thesis. This has meant a lot to me.

I want to thank all the people at Food and Nutrition science for the welcoming and supportive work environment. A special thanks goes to Nisse and Annette for all the help you have provided along the way. This has been much appreciated! Mi-Lo and Helena, thank you for always arranging everything. Mads, thank you for the time together in the office and lab, and for all the help in statistics. I really enjoyed to work with you and really appreciated all the inspiring discussions! Also, I extend my sincere gratitude to all the past and present fellow ´doktorander´ at the institute.

Göran and Björn, thank you for the collaboration with the DIWA samples and for always proving challenging feedback.

Area Of Advance Life Science Engineering is acknowledged for funding of this thesis.

Finally, I want to thank my family for all the support and encouragement on the way. Most importantly, I want to thank Piia for the endless patience and encouragement that has carried me through this thesis and my children for all the joy You bring to my life! This thesis is to you!

REFERENCES

1. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends in Biotechnology. 1998;16:373–8.
2. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13:263–9.
3. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, et al. The human serum metabolome. Flower D, editor. PloS one. Public Library of Science; 2011;6:e16957.
4. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry. 2005;24:285–94.
5. Pasikanti KK, Ho PC, Chan ECY. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2008;871:202–11.
6. Forcisi S, Moritz F, Kanawati B, Tziotis D, Lehmann R, Schmitt-Kopplin P. Liquid chromatography-mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. Journal of chromatography A. 2013;1292:51–65.
7. Ramautar R, Somsen GW, de Jong GJ. CE-MS for metabolomics: developments and applications in the period 2010-2012. 2012 ed. 2013;34:86–98.
8. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. Journal of pharmaceutical and biomedical analysis. 2014;87:12–25.
9. Villas-Boas SG, Smart KF, Sivakumaran S, Lane GA. Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? Metabolites. Molecular Diversity Preservation International; 2011;1:3–20.
10. Krone N, Hughes BA, Lavery GG, Stewart PM, Arlt W, Shackleton CHL. Gas chromatography/mass spectrometry (GC/MS) remains a pre-eminent discovery tool in clinical steroid investigations even in the era of fast liquid chromatography tandem mass spectrometry (LC/MS/MS). J Steroid Biochem Mol Biol. 2010;121:496–504.
11. Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2013;22:2182–201.
12. Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot. Oxford University Press; 2005;56:219–43.
13. Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, Hanson RW, Kalhan SC, Ryals JA, Milburn MV. Analysis of the adult human plasma metabolome. Pharmacogenomics. Future Medicine Ltd London, UK; 2008;9:383–97.
14. Kvitvang HF, Andreassen T, Adam T, Villas-Boas SG, Bruheim P. Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. Analytical chemistry. 2011 ed. 2011;83:2705–11.
15. A J, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, Marklund SL, Moritz T. Extraction and GC/MS analysis of the human blood plasma metabolome. Analytical chemistry. 2005 ed. 2005;77:8086–94.
16. Begley P, Francis-McIntyre S, Dunn WB, Broadhurst DI, Halsall A, Tseng A, Knowles J, Goodacre R, Kell DB, Consortium H. Development and Performance of a Gas Chromatography-Time-of-Flight Mass Spectrometry Analysis for Large-Scale Nontargeted Metabolomic Studies of Human Serum. Analytical chemistry. 2009;81:7038–46.
17. Lu Y, A J, Wang G, Hao H, Huang Q, Yan B, Zha W, Gu S, Ren H, Zhang Y, et al. Gas chromatography/time-of-flight mass spectrometry based metabonomic approach to differentiating hypertension- and age-related metabolic variation in spontaneously hypertensive rats. Rapid communications in mass spectrometry : RCM. 2008 ed. 2008;22:2882–8.
18. Major HJ, Williams R, Wilson AJ, Wilson ID. A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition. Rapid communications in mass spectrometry : RCM. John Wiley & Sons, Ltd; 2006;20:3295–302.
19. Wachsmuth CJ, Almstetter MF, Waldhier MC, Gruber MA, Nürnberger N, Oefner PJ, Dettmer K. Performance Evaluation of Gas Chromatography–Atmospheric Pressure Chemical Ionization–Time-of-Flight Mass Spectrometry for Metabolic Fingerprinting and Profiling. Analytical chemistry. American Chemical Society; 2011;83:7514–22.
20. Wachsmuth CJ, Hahn TA, Oefner PJ, Dettmer K. Enhanced metabolite profiling using a redesigned atmospheric pressure chemical ionization source for gas chromatography coupled to high-resolution time-of-flight mass spectrometry. Anal Bioanal Chem. Springer Berlin Heidelberg; 2015;407:6669–80.

21. Koek MM, van der Kloet FM, Kleemann R, Kooistra T, Verheij ER, Hankemeier T. Semi-automated non-target processing in GC × GC-MS metabolomics analysis: applicability for biomedical studies. Metabolomics : Official journal of the Metabolomic Society. 2011 ed. 2011;7:1–14.
22. Adahchour M, Brandt M, Baier H-U, Vreuls RJJ, Batenburg AM, Brinkman UAT. Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: principles and applications. Journal of chromatography A. 2005;1067:245–54.
23. Carrasco-Pancorbo A, Nevedomskaya E, Arthen-Engeland T, Zey T, Zurek G, Baessmann C, Deelder AM, Mayboroda OA. Gas chromatography/atmospheric pressure chemical ionization-time of flight mass spectrometry: analytical validation and applicability to metabolic profiling. Analytical chemistry. American Chemical Society; 2009;81:10071–9.
24. Suvitaival T, Mantere O, Kieseppä T, Mattila I, Pöhö P, Hyotylainen T, Suvisaari J, Oresic M. Serum metabolite profile associates with the development of metabolic co-morbidities in first-episode psychosis. Transl Psychiatry. Nature Publishing Group; 2016;6:e951.
25. Hyötyläinen T, Mattila I, Wiedmer SK, Koivuniemi A, Taskinen M-R, Yki-Järvinen H, Orešič M. Metabolomic analysis of polar metabolites in lipoprotein fractions identifies lipoprotein-specific metabolic profiles and their association with insulin resistance. Mol Biosyst. The Royal Society of Chemistry; 2012;8:2559–65.
26. Bondia-Pons I, Maukonen J, Mattila I, Rissanen A, Saarela M, Kaprio J, Hakkarainen A, Lundbom J, Lundbom N, Hyötyläinen T, et al. Metabolome and fecal microbiota in monozygotic twin pairs discordant for weight: a Big Mac challenge. FASEB J. Federation of American Societies for Experimental Biology; 2014;28:4169–79.
27. Hyötyläinen T, Jerby L, Petäjä EM, Mattila I, Jäntti S, Auvinen P, Gastaldelli A, Yki-Järvinen H, Ruppin E, Orešič M. Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. Nat Commun. 2016;7:8994.
28. Bruheim P, Kvitvang HFN, Villas-Boas SG. Stable isotope coded derivatizing reagents as internal standards in metabolite profiling. Journal of chromatography A. 2013;1296:196–203.
29. Luan H, Yang L, Ji F, Cai Z. PCI-GC-MS-MS approach for identification of non-amino organic acid and amino acid profiles. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2016.
30. Weinert CH, Egert B, Kulling SE. On the applicability of comprehensive two-dimensional gas chromatography combined with a fast-scanning quadrupole mass spectrometer for untargeted large-scale metabolomics. Journal of chromatography A. 2015;1405:156–67.
31. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol. 2012;8:615.
32. Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam K-P, Milburn MV, Kastenmüller G, Adamski J, Tuomi T, et al. Early Metabolic Markers of the Development of Dysglycemia and Type 2 Diabetes and Their Physiological Significance. Diabetes. American Diabetes Association; 2013;62:1730–7.
33. Gürdeniz G, Jensen MG, Meier S, Bech L, Lund E, Dragsted LO. Detecting Beer Intake by Unique Metabolite Patterns. J Proteome Res. American Chemical Society; 2016;15:4544–56.
34. Andersen M-BS, Rinnan Å, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, Astrup A, Dragsted LO. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res. American Chemical Society; 2014;13:1405–18.
35. Zhao Y-Y, Chen H, Tian T, Chen D-Q, Bai X, Wei F. A pharmaco-metabonomic study on chronic kidney disease and therapeutic effect of ergone by UPLC-QTOF/HDMS. Dzeja P, editor. PloS one. Public Library of Science; 2014;9:e115467.
36. Pekkinen J, Olli K, Huotari A, Tiihonen K, Keski-Rahkonen P, Lehtonen M, Auriola S, Kolehmainen M, Mykkänen H, Poutanen K, et al. Betaine supplementation causes increase in carnitine metabolites in the muscle and liver of mice fed a high-fat diet as studied by nontargeted LC-MS metabolomics approach. Molecular nutrition & food research. 2013 ed. 2013;57:1959–68.
37. Savolainen O, Pekkinen J, Katina K, Poutanen K, Hanhineva K. Glycosylated Benzoxazinoids Are Degraded during Fermentation of Wheat Bran. Journal of agricultural and food chemistry. American Chemical Society; 2015;63:5943–9.
38. Bondia-Pons I, Barri T, Hanhineva K, Juntunen K, Dragsted LO, Mykkanen H, Poutanen K. UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. Molecular nutrition & food research. Germany: WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim; 2013;57:412–22.
39. Moazzami AA, Bondia-Pons I, Hanhineva K, Juntunen K, Antl N, Poutanen K, Mykkänen H. Metabolomics reveals the metabolic shifts following an intervention with rye bread in postmenopausal women--a randomized control trial. Nutr J. BioMed Central Ltd; 2012;11:88.

40. Hanhineva K, Aura AM, Rogachev I, Matero S, Skov T, Aharoni A, Poutanen K, Mykkanen H. In vitro microbiotic fermentation causes an extensive metabolite turnover of rye bran phytochemicals. PloS one. United States; 2012;7:e39322.
41. Pekkinen J, Rosa NN, Savolainen O-I, Keski-Rahkonen P, Mykkänen H, Poutanen K, Micard V, Hanhineva K. Disintegration of wheat aleurone structure has an impact on the bioavailability of phenolic compounds and other phytochemicals as evidenced by altered urinary metabolite profile of diet-induced obese mice. Nutrition & Metabolism. BioMed Central Ltd; 2014;11.
42. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, de Mello V, Sola R, Lehtonen M, Poutanen K, et al. Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, fatty fish, and bilberries in a randomized controlled trial. The Journal of nutrition. American Society for Nutrition; 2015;145:7–17.
43. Moazzami AA, Zhang J-X, Kamal-Eldin A, Åman P, Hallmans G, Johansson J-E, Andersson S-O. Nuclear magnetic resonance-based metabolomics enable detection of the effects of a whole grain rye and rye bran diet on the metabolic profile of plasma in prostate cancer patients. The Journal of nutrition. American Society for Nutrition; 2011;141:2126–32.
44. Zhu Y, Wang P, Sha W, Sang S. Urinary Biomarkers of Whole Grain Wheat Intake Identified by Non-targeted and Targeted Metabolomics Approaches. Scientific Reports. Nature Publishing Group; 2016;6:36278.
45. Pekkinen J, Rosa Sibakov N, Micard V, Keski-Rahkonen P, Lehtonen M, Poutanen K, Mykkänen H, Hanhineva K. Amino acid-derived betaines dominate as urinary markers for rye bran intake in mice fed high-fat diet—A nontargeted metabolomics study. Molecular nutrition & food research. 2015;59:1550–62.
46. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. The American journal of clinical nutrition. 2005;82:497–503.
47. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. British Medical Journal Publishing Group; 1995;310:170–0.
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995.
49. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol. Frontiers; 2015;3:23.
50. Bro R, Smilde AK. Principal component analysis. Analytical Methods. Royal Society of Chemistry; 2014;6:2812–31.
51. Fonville JM, Richards SE, Barton RH, Boulange CL, Ebbels TMD, Nicholson JK, Holmes E, Dumas M-E. The evolution of partial least squares models and related chemometric approaches in metabonomics and metabolic phenotyping. Journal of Chemometrics. John Wiley & Sons, Ltd; 2010;24:636–49.
52. Kjeldahl K, Bro R. Some common misunderstandings in chemometrics. Journal of Chemometrics. John Wiley & Sons, Ltd; 2010;24:558–64.
53. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. CBIO. Bentham Science Publishers; 2012;7:96–108.
54. Herder C, Kowall B, Tabak AG, Rathmann W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. Diabetologia. 2014;57:16–29.
55. Ganna A, Salihovic S, Sundström J, Broeckling CD, Hedman AK, Magnusson PKE, Pedersen NL, Larsson A, Siegbahn A, Zilmer M, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. Gibson G, editor. PLoS Genet. Public Library of Science; 2014;10:e1004801.
56. Jaremek M, Yu Z, Mangino M, Mittelstrass K, Prehn C, Singmann P, Xu T, Dahmen N, Weinberger KM, Suhre K, et al. Alcohol-induced metabolomic differences in humans. Transl Psychiatry. Nature Publishing Group; 2013;3:e276.
57. Floegel A, Ruesten von A, Drogan D, Schulze MB, Prehn C, Adamski J, Pischon T, Boeing H. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. European journal of clinical nutrition. 2013;67:1100–8.
58. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. The American journal of clinical nutrition. American Society for Nutrition; 2014;100:208–17.
59. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc. 2011;6:1060–83.
60. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC

samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis. Future Science Ltd London, UK; 2012;4:2249–64.
61. Cuadros-Rodríguez L, Bagur-González MG, Sánchez-Viñas M, González-Casado A, Gómez-Sáez AM. Principles of analytical calibration/quantification for the separation sciences. Journal of chromatography A. 2007;1158:33–46.
62. FDA Guidance for Industry: Bioanalytical Method Validation.
63. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics: Official journal of the Metabolomic Society. 2007;3:211–21.
64. Pitt JJ. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. The Clinical Biochemist Reviews. The Australian Association of Clinical Biochemists; 2009;30:19–34.
65. De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Tull D, Roessner U, McConville M, Speed TP. Normalizing and Integrating Metabolomics Data. Analytical chemistry. American Chemical Society; 2012;84:10768–76.
66. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. J Proteome Res. American Chemical Society; 2009;8:5132–41.
67. Brunius C, Shi L, Landberg R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. Metabolomics: Official journal of the Metabolomic Society. 2016;12:173.
68. Kellogg MD, Ellervik C, Morrow D, Hsing A, Stein E, Sethi AA. Preanalytical considerations in the design of clinical trials and epidemiological studies. Clinical chemistry. Clinical Chemistry; 2015;61:797–803.
69. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal Bioanal Chem. Springer-Verlag; 2012;403:1523–48.
70. Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, O'Hagan S, Knowles JD, Halsall A, Wilson ID, et al. Development of a Robust and Repeatable UPLC-MS Method for the Long-Term Metabolomic Study of Human Serum. Analytical chemistry. 2009;81:1357–64.
71. Dunn WB, Broadhurst D, Ellis DI, Brown M, Halsall A, O'Hagan S, Spasic I, Tseng A, Kell DB. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. Int J Epidemiol. Oxford University Press; 2008;37 Suppl 1:i23–30.
72. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner LW, Goodacre R, Hardy NW, Taylor C, et al. The metabolomics standards initiative (MSI). Metabolomics: Official journal of the Metabolomic Society. 2007;3:175–8.
73. Schwab U, Seppänen-Laakso T, Yetukuri L, Ågren J, Kolehmainen M, Laaksonen DE, Ruskeepää A-L, Gylling H, Uusitupa M, Orešič M, et al. Triacylglycerol Fatty Acid Composition in Diet-Induced Weight Loss in Subjects with Abnormal Glucose Metabolism – the GENOBIN Study. Risérus U, editor. PloS one. Public Library of Science; 2008;3:e2630.
74. Hu FB, Narasimhan K. Nutritional genomics and metabolomics in obesity and type 2 diabetes. BMC Genomics. BioMed Central; 2014;15:O10.
75. Urpi-Sarda M, Almanza-Aguilera E, Tulipani S, Tinahones FJ, Salas-Salvadó J, Andres-Lacueva C. Metabolomics for Biomarkers of Type 2 Diabetes Mellitus: Advances and Nutritional Intervention Trends. Curr Cardiovasc Risk Rep. Springer US; 2015;9:12.
76. Fardet A, Llorach R, Orsoni A, Martin J-F, Pujos-Guillot E, Lapierre C, Scalbert A. Metabolomics provide new insight on the metabolism of dietary phytochemicals in rats. The Journal of nutrition. 2008;138:1282–7.
77. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hooft JJJ, Wishart DS. The food metabolome: a window over dietary exposure. The American journal of clinical nutrition. American Society for Nutrition; 2014;99:1286–308.
78. van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. Microbial metabolomics: toward a platform with full metabolome coverage. Analytical Biochemistry. 2007;370:17–25.
79. Hanhineva K, Rogachev I, Aura AM, Aharoni A, Poutanen K, Mykkanen H. Qualitative characterization of benzoxazinoid derivatives in whole grain rye and wheat by LC-MS metabolite profiling. Journal of agricultural and food chemistry. United States; 2011;59:921–7.
80. Pedersen HA, Laursen B, Mortensen A, Fomsgaard IS. Bread from common cereal cultivars contains an important array of neglected bioactive benzoxazinoids. Food Chem. 2011;127:1814–20.
81. Adhikari KB, Laursen BB, Gregersen PL, Schnoor HJ, Witten M, Poulsen LK, Jensen BM, Fomsgaard IS. Absorption and metabolic fate of bioactive dietary benzoxazinoids in humans. Molecular nutrition & food research. 2013 ed. 2013;57:1847–58.
82. Adhikari KB, Laerke HN, Mortensen AG, Fomsgaard IS. Plasma and Urine Concentrations of Bioactive

Dietary Benzoxazinoids and Their Glucuronidated Conjugates in Rats Fed a Rye Bread-Based Diet. J Agric Food Chem. American Chemical Society; 2012;60:11518–24.
83. Hanhineva K, Keski-Rahkonen P, Lappi J, Katina K, Pekkinen J, Savolainen O, Timonen O, Paananen J, Mykkänen H, Poutanen K. The postprandial plasma rye fingerprint includes benzoxazinoid-derived phenylacetamide sulfates. The Journal of nutrition. 2014 ed. American Society for Nutrition; 2014;144:1016–22.
84. Jensen BM, Adhikari KB, Schnoor HJ, Juel-Berg N, Fomsgaard IS, Poulsen LK. Quantitative analysis of absorption, metabolism, and excretion of benzoxazinoids in humans after the consumption of high- and low-benzoxazinoid diets with similar contents of cereal dietary fibres: a crossover study. Eur J Nutr. Springer Berlin Heidelberg; 2015;1–11.
85. Beckmann M, Lloyd AJ, Haldar S, Seal C, Brandt K, Draper J. Hydroxylated phenylacetamides derived from bioactive benzoxazinoids are bioavailable in humans after habitual consumption of whole grain sourdough rye bread. Molecular nutrition & food research. 2013.
86. Wittenbecher C, Mühlenbruch K, Kröger J, Jacobs S, Kuxhaus O, Floegel A, Fritsche A, Pischon T, Prehn C, Adamski J, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. The American journal of clinical nutrition. American Society for Nutrition; 2015;101:1241–50.
87. Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, Cross AJ, Gunter MJ, Fensom GK, Appleby PN, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. The American journal of clinical nutrition. American Society for Nutrition; 2015;102:1518–26.
88. Carrizo D, Chevallier OP, Woodside JV, Brennan SF, Cantwell MM, Cuskelly G, Elliott CT. Untargeted metabolomic analysis of human serum samples associated with different levels of red meat consumption: A possible indicator of type 2 diabetes? Food Chem. 2017;221:214–21.
89. Cheung W, Keski-Rahkonen P, Assi N, Pietro Ferrari, Freisling H, Rinaldi S, Slimani N, Zamora-Ros R, Rundle M, Frost G, et al. A metabolomic study of biomarkers of meat and fish intake. American Journal of Clinical Nutrition. American Society for Nutrition; 2017;ajcn146639.
90. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. American Journal of Clinical Nutrition. American Society for Nutrition; 2014;100:ajcn.078758–217.
91. Schmedes M, Aadland EK, Sundekilde UK, Jacques H, Lavigne C, Graff IE, Eng Ø, Holthe A, Mellgren G, Young JF, et al. Lean-seafood intake decreases urinary markers of mitochondrial lipid and energy metabolism in healthy subjects: Metabolomics results from a randomized crossover intervention study. Molecular nutrition & food research. 2016;60:1661–72.
92. Ross AB, Svelander C, Undeland I, Pinto R, Sandberg A-S. Herring and Beef Meals Lead to Differences in Plasma 2-Aminoadipic Acid, β-Alanine, 4-Hydroxyproline, Cetoleic Acid, and Docosahexaenoic Acid Concentrations in Overweight Men. The Journal of nutrition. American Society for Nutrition; 2015;145:2456–63.
93. Wang TJ, Ngo D, Psychogios N, Dejam A, Larson MG, Vasan RS, Ghorbani A, O'Sullivan J, Cheng S, Rhee EP, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. The Journal of clinical investigation. 2013 ed. 2013;123:4309–17.
94. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. Nature Publishing Group; 2011;17:448–53.
95. InterAct Consortium, Bendinelli B, Palli D, Masala G, Sharp SJ, Schulze MB, Guevara M, van der ADL, Sera F, Amiano P, et al. Association between dietary meat consumption and incident type 2 diabetes: the EPIC-InterAct study. Diabetologia. Springer-Verlag; 2013;56:47–59.
96. Männistö S, Kontto J, Kataja-Tuomola M, Albanes D, Virtamo J. High processed meat consumption is a risk factor of type 2 diabetes in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention study. Br J Nutr. Cambridge University Press; 2010;103:1817–22.
97. Fung TT, Schulze M, Manson JE. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. ACC Curren; 2005.
98. World Health Organization. Global Report on Diabetes. WHO. 2016.
99. Bailes BK. Diabetes Mellitus and its Chronic Complications. AORN Journal. 2002;76:265–82.
100. Schauer PR, Kashyap SR, Wolski K, Brethauer SA, Kirwan JP, Pothier CE, Thomas S, Abood B, Nissen SE, Bhatt DL. Bariatric Surgery versus Intensive Medical Therapy in Obese Patients with Diabetes. http://dxdoiorg/101056/NEJMoa1200225. Massachusetts Medical Society; 2012;366:1567–76.
101. World Health Organization, Federation ID. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. WHO; 2006. Available from:

http://apps.who.int/iris/bitstream/10665/43588/1/9241594934_eng.pdf
102. Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet. 2012;379:2279–90.
103. Nathan DM, Turgeon H, Regan S. Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia. Springer-Verlag; 2007;50:2239–44.
104. Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. Diabetes care. 2004;27:1487–95.
105. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. American Diabetes Association; 2011. pp. 1249–57.
106. World Health Organization. Global status report on noncommunicable diseases.World HealthOrganization;2014.Availablefrom:http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1
107. Diabetes Prevention Program Research Group, KNOWLER WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, Brenneman AT, Brown-Friday JO, Goldberg R, Venditti E, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet. 2009;374:1677–86.
108. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Keinänen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med. Massachusetts Medical Society; 2001;344:1343–50.
109. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, Li H, Li H, Jiang Y, An Y, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet. 2008;371:1783–9.
110. Fagerberg B, Kellis D, Bergstrom G, Behre CJ. Adiponectin in relation to insulin sensitivity and insulin secretion in the development of type 2 diabetes: a prospective study in 64-year-old women. Journal of internal medicine. Blackwell Publishing Ltd; 2011;269:636–43.
111. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract. Elsevier; 2014;103:137–49.
112. Wurtz P, Mäkinen V-P, Soininen P, Kangas AJ, Tukiainen T, Kettunen J, Savolainen MJ, Tammelin T, Viikari JS, Rönnemaa T, et al. Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes. American Diabetes Association; 2012;61:1372–80.
113. Menni C, Fauman E, Erte I, Perry JRB, Kastenmüller G, Shin S-Y, Petersen A-K, Hyde C, Psatha M, Ward KJ, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. Diabetes. American Diabetes Association; 2013;62:4270–6.
114. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell metabolism. 2009;9:311–26.
115. Newbern D, Gumus Balikcioglu P, Balikcioglu M, Bain J, Muehlbauer M, Stevens R, Ilkayeva O, Dolinsky D, Armstrong S, Irizarry K, et al. Sex differences in biomarkers associated with insulin resistance in obese adolescents: metabolomic profiling and principal components analysis. The Journal of Clinical Endocrinology & Metabolism. 2014;99:4730–9.
116. Xu F, Tavintharan S, Sum CF, Woon K, Lim SC, Ong CN. Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. The Journal of Clinical Endocrinology & Metabolism. 2013;98:E1060–5.
117. Cobb J, Eckhart A, Perichon R, Wulff J, Mitchell M, Adam K-P, Wolfert R, Button E, Lawton K, Elverson R, et al. A novel test for IGT utilizing metabolite markers of glucose tolerance. J Diabetes Sci Technol. SAGE Publications; 2015;9:69–76.
118. Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. Breant B, editor. PloS one. 2010;5:e13953.
119. Kaur P, Rizk N, Ibrahim S, Luo Y, Younes N, Perry B, Dennis K, Zirie M, Luta G, Cheema AK. Quantitative metabolomic and lipidomic profiling reveals aberrant amino acid metabolism in type 2 diabetes. Mol Biosyst. The Royal Society of Chemistry; 2013;9:307–17.
120. Walford GA, Porneala BC, Dauriz M, Vassy JL, Cheng S, Rhee EP, Wang TJ, Meigs JB, Gerszten RE, Florez JC. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. Diabetes care. American Diabetes Association; 2014;37:2508–14.
121. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost H-G, Fritsche A, Häring H-U, Hrabé de Angelis M, Peters A, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. American Diabetes Association; 2013;62:639–48.
122. Palmer ND, Stevens RD, Antinozzi PA, Anderson A, Bergman RN, Wagenknecht LE, Newgard CB, Bowden

DW. Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of Clinical Endocrinology & Metabolism. Endocrine Society Chevy Chase, MD; 2014;100:E463–8.
123. Zhao J, Zhu Y, Hyun N, Zeng D, Uppal K, Tran VT, Yu T, Jones D, He J, Lee ET, et al. Novel metabolic markers for the risk of diabetes development in American Indians. Diabetes care. American Diabetes Association; 2015;38:220–7.
124. Padberg I, Peter E, González-Maldonado S, Witt H, Mueller M, Weis T, Bethan B, Liebenberg V, Wiemer J, Katus HA, et al. A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology. Hribal ML, editor. PloS one. Public Library of Science; 2014;9:e85082.
125. Gall WE, Beebe K, Lawton KA, Adam K-P, Mitchell MW, Nakhle PJ, Ryals JA, Milburn MV, Nannipieri M, Camastra S, et al. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. Federici M, editor. PloS one. 2010;5:e10883.
126. Prentice KJ, Luu L, Allister EM, Liu Y, Jun LS, Sloop KW, Hardy AB, Wei L, Jia W, Fantus IG, et al. The Furan Fatty Acid Metabolite CMPF Is Elevated in Diabetes and Induces beta Cell Dysfunction. Cell metabolism. 2014;19:653–66.
127. Liu Y, Prentice KJ, Eversley JA, Hu C, Batchuluun B, Leavey K, Hansen JB, Wei DW, Cox B, Dai FF, et al. Rapid Elevation in CMPF May Act As a Tipping Point in Diabetes Development. Cell Reports. 2016;14:2889–900.
128. Lankinen MA, Hanhineva K, Kolehmainen M, Lehtonen M, Auriola S, Mykkänen H, Poutanen K, Schwab U, Uusitupa M. CMPF Does Not Associate with Impaired Glucose Metabolism in Individuals with Features of Metabolic Syndrome. Passi AG, editor. PloS one. Public Library of Science; 2015;10:e0124379.
129. Edelman D, Olsen MK, Dudley TK, Harris AC, Oddone EZ. Utility of hemoglobin A1c in predicting diabetes risk. J Gen Intern Med. Springer-Verlag; 2004;19:1175–80.
130. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. Elsevier; 2014;383:999–1008.
131. Holle R, Happich M, Löwel H, Wichmann HE, MONICA/KORA Study Group. KORA--a research platform for population based health research. Gesundheitswesen. © Georg Thieme Verlag KG Stuttgart · New York; 2005;67 Suppl 1:S19–25.
132. Borodulin K, Vartiainen E, Peltonen M, Jousilahti P, Juolevi A, Laatikainen T, Männistö S, Salomaa V, Sundvall J, Puska P. Forty-year trends in cardiovascular risk factors in Finland. Eur J Public Health. Oxford University Press; 2015;25:539–46.
133. Hills SA, Balkau B, Coppack SW, Dekker JM, Mari A, Natali A, Walker M, Ferrannini E, EGIR-RISC Study Group. The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and objectives. Diabetologia. 2004;47:566–70.
134. Groop L, Forsblom C, Lehtovirta M, Tuomi T, Karanko S, Nissén M, Ehrnström BO, Forsén B, Isomaa B, Snickars B, et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. Diabetes. 1996;45:1585–93.
135. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondière UR, Hémon B, Casagrande C, Vignat J, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. Cambridge University Press; 2002;5:1113–24.
136. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. 2011 ed. 2011;17:448–53.
137. Giesbertz P, Daniel H. Branched-chain amino acids as biomarkers in diabetes. Curr Opin Clin Nutr Metab Care. 2016;19:48–54.
138. Wurtz P, Soininen P, Kangas AJ, Rönnemaa T, Lehtimaki T, Kahonen M, Viikari JS, Raitakari OT, Ala-Korpela M. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diabetes care. 2013;36:648–55.
139. Yengo L, Arredouani A, Marre M, Roussel R, Vaxillaire M, Falchi M, Haoudi A, Tichet J, D.E.S.I.R Study Group, Balkau B, et al. Impact of statistical models on the prediction of type 2 diabetes using non-targeted metabolomics profiling. Mol Metab. 2016;5:918–25.
140. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5:463–6.
141. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89–95.
142. World Health Organization, Safety IPOC. Biomarkers in risk assessment: validity and validation.Geneva:WorldHealthOrganization;2001.Availablefrom:http://www.who.int/iris/handle/10665/42363
143. World Health Organization, Safety IPOC. Biomarkers and risk assessment: concepts and principles.

Geneva : World Health Organization; 1993.
144. Kannel WB. Elevated systolic blood pressure as a cardiovascular risk factor. American Journal of Cardiology. 2000;85:251–5.
145. Selvin E, Steffes MW, Gregg E, Brancati FL, Coresh J. Performance of A1C for the classification and prediction of diabetes. Diabetes care. American Diabetes Association; 2011;34:84–9.
146. Kodama S, Horikawa C, Fujihara K, Hirasawa R, Yachi Y, Yoshizawa S, Tanaka S, Sone Y, Shimano H, Iida KT, et al. Use of high-normal levels of haemoglobin A(1C) and fasting plasma glucose for diabetes screening and for prediction: a meta-analysis. Diabetes Metab Res Rev. 2013;29:680–92.
147. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. American Journal of Clinical Nutrition. American Society for Nutrition; 1997;65:1232S–1239S.
148. Branca F, Hanley AB, Pool-Zobel B, Verhagen H. Biomarkers in disease and health. Br J Nutr. 2001;86 Suppl 1:S55–92.
149. Lyssenko V, Laakso M. Genetic screening for the risk of type 2 diabetes: worthless or valuable? Diabetes care. 2013;36 Suppl 2:S120–6.
150. Illner AK, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol. Oxford University Press; 2012;41:1187–203.
151. Byers T. Food frequency dietary assessment: how bad is good enough? Am J Epidemiol. Oxford University Press; 2001;154:1087–8.
152. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2005;14:2826–8.
153. Potischman N. Biologic and methodologic issues for nutritional biomarkers. The Journal of nutrition. 2003;133 Suppl 3:875S–880S.
154. McCabe-Sellers B. Advancing the Art and Science of Dietary Assessment through Technology. Journal of the American Dietetic Association. Journal of the American dietetic association; 2010;110:52–4.
155. Hardin DS. Validating Dietary Intake with Biochemical Markers. Journal of the American Dietetic Association. Journal of the American Dietetic Association; 2009;109:1698–9.
156. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. American Journal of Clinical Nutrition. American Society for Nutrition; 1985;42:1276–89.
157. Jenab M, Slimani N, Bictash M, Pietro Ferrari, Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum Genet. Springer-Verlag; 2009;125:507–25.
158. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. American Journal of Clinical Nutrition. 1997;65:1232S–1239S.
159. Ross AB. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. J Nutr Metab. Hindawi Publishing Corporation; 2012;2012:462967–12.
160. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary Sucrose and Fructose as Biomarkers for Sugar Consumption. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2005;14:1287–94.
161. Marshall JR. Methodologic and statistical considerations regarding use of biomarkers of nutritional exposure in epidemiology. The Journal of nutrition. 2003;133 Suppl 3:881S–887S.
162. Hedrick VE, Dietrich AM, Estabrooks PA, Savla J, Serrano E, Davy BM. Dietary biomarkers: advances, limitations and future directions. Nutr J. 2nd ed. BioMed Central; 2012;11:109.
163. Joosen AMCP, Kuhnle GGC, Runswick SA, Bingham SA. Urinary sucrose and fructose as biomarkers of sugar consumption: comparison of normal weight and obese volunteers. International Journal of Obesity. Nature Publishing Group; 2008;32:1736–40.
164. Ross AB, Bourgeois A, Macharia HN. Plasma alkylresorcinols as a biomarker of whole-grain food consumption in a large population: results from the WHOLEheart Intervention Study. The American journal …; 2012.
165. Andersson A, Marklund M, Diana M, Landberg R. Plasma alkylresorcinol concentrations correlate with whole grain wheat and rye intake and show moderate reproducibility over a 2- to 3-month period in free-living Swedish adults. The Journal of nutrition. American Society for Nutrition; 2011;141:1712–8.
166. Landberg R, Kamal-Eldin A, Andersson A, Vessby B, Åman P. Alkylresorcinols as biomarkers of whole-grain wheat and rye intake: plasma concentration and intake estimated from dietary records. The American journal of clinical nutrition. 2008;87:832–8.
167. Landberg R, Kamal-Eldin A, Andersson SO. Reproducibility of plasma alkylresorcinols during a 6-week rye intervention study in men with prostate cancer. The Journal of …; 2009.
168. Kristensen M, Toubro S, Jensen MG, Ross AB, Riboldi G, Petronio M, Bügel S, Tetens I, Astrup A. Whole

grain compared with refined wheat decreases the percentage of body fat following a 12-week, energy-restricted dietary intervention in postmenopausal women. The Journal of nutrition. American Society for Nutrition; 2012;142:710–6.

169. Magnusdottir OK, Landberg R, Gunnarsdottir I, Cloetens L, Åkesson B, Landin-Olsson M, Rosqvist F, Iggman D, Schwab U, Herzig K-H, et al. Plasma alkylresorcinols C17:0|[sol]|C21:0 ratio, a biomarker of relative whole-grain rye intake, is associated to insulin sensitivity: a randomized study. European journal of clinical nutrition. Nature Publishing Group; 2014;68:453–8.

170. Ross AB, Bruce SJ, Blondel-Lubrano A. A whole-grain cereal-rich diet increases plasma betaine, and tends to decrease total and LDL-cholesterol compared with a refined-grain diet in healthy subjects. British journal of …. 2011.

171. Choung RS, Murray JA, Marietta EV, Van Dyke CT, Ross AB. Serum alkylresorcinols as biomarkers of dietary gluten exposure in coeliac disease. Aliment Pharmacol Ther. 2017.

172. Ross AB, Svelander C, Karlsson G, Savolainen OI. Identification and quantification of even and odd chained 5-n alkylresorcinols, branched chain-alkylresorcinols and methylalkylresorcinols in Quinoa (Chenopodium quinoa). Food Chem. 2017;220:344–51.

173. McKeown NM, Marklund M, Ma J, Ross AB, Lichtenstein AH, Livingston KA, Jacques PF, Rasmussen HM, Blumberg JB, Chen C-YO. Comparison of plasma alkylresorcinols (AR) and urinary AR metabolites as biomarkers of compliance in a short-term, whole-grain intervention study. Eur J Nutr. Springer Berlin Heidelberg; 2016;55:1235–44.

174. Marklund M, Magnusdottir OK, Rosqvist F, Cloetens L, Landberg R, Kolehmainen M, Brader L, Hermansen K, Poutanen KS, Herzig K-H, et al. A dietary biomarker approach captures compliance and cardiometabolic effects of a healthy Nordic diet in individuals with metabolic syndrome. The Journal of nutrition. American Society for Nutrition; 2014;144:1642–9.

175. Marklund M, Landberg R, Andersson A, Åman P, Kamal-Eldin A. Alkylresorcinol metabolites in urine correlate with the intake of whole grains and cereal fibre in free-living Swedish adults. Br J Nutr. Cambridge University Press; 2013;109:129–36.

176. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, Hansen T, Beckmann M, Pedersen O, Elliott P, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. Lancet Diabetes Endocrinol. 2017.

177. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, et al. Identifying biomarkers of dietary patterns by using metabolomics. The American journal of clinical nutrition. American Society for Nutrition; 2016;ajcn144501.

178. Brennan L. The nutritional metabolomics crossroads: how to ensure success for dietary biomarkers. The American journal of clinical nutrition. 2017;ajcn150847.

179. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, van der Ouderaa F, Bingham S, Cross AJ, Nicholson JK. Susceptibility of Human Metabolic Phenotypes to Dietary Modulation. J Proteome Res. 2006;5:2780–8.

180. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. Anal Bioanal Chem. Springer-Verlag; 2010;396:1451–63.

181. Myint T, Fraser GE, Lindsted KD. Urinary 1-methylhistidine is a marker of meat consumption in Black and in White California Seventh-day Adventists. American journal of …. 2000.

182. Gibbons H, Brennan L. Metabolomics as a tool in the identification of dietary biomarkers. The Proceedings of the Nutrition Society. Cambridge University Press; 2016;:1–12.

183. Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, Grung B, Marklund S, Sjostrom M, Antti H, Moritz T. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Analytical chemistry. 2005 ed. 2005;77:5635–42.

184. Lindqvist H, Langkilde AM, Undeland I, Rådendal T, Sandberg A-S. Herring (Clupea harengus) supplemented diet influences risk factors for CVD in overweight subjects. European journal of clinical nutrition. Nature Publishing Group; 2007;61:1106–13.

185. Özcan S, Şenyuva HZ. Improved and simplified liquid chromatography/atmospheric pressure chemical ionization mass spectrometry method for the analysis of underivatized free amino acids in various foods. Journal of chromatography A. 2006;1135:179–85.

186. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics. John Wiley & Sons, Ltd; 2007;20:341–51.

187. Drăghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA. Global functional profiling of gene expression☆☆This work was funded in part by a Sun Microsystems grant awarded to S.D., NIH Grant HD36512 to S.A.K., a Wayne State University SOM Dean's Post-Doctoral Fellowship, and an NICHD Contraception and Infertility Loan to G.C.O. Support from the WSU MCBI mode is gratefully

appreciated. Genomics. 2003;81:98–104.
188. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. Oxford University Press; 2014;42:D199–D205.
189. Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. Bioinformatics (Oxford, England). Oxford University Press; 2010;26:2342–4.
190. Brohall G, Behre CJ, Hulthe J, Wikstrand J, Fagerberg B. Prevalence of diabetes and impaired glucose tolerance in 64-year-old Swedish women: experiences of using repeated oral glucose tolerance tests. Diabetes care. 2006 ed. 2006;29:363–7.
191. Alberti KGMM, Zimmet PZ, WHO Consultation. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. Diabet Med. John Wiley & Sons, Ltd; 1998;15:539–53.
192. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44:837.
193. Calcagno V, de Mazancourt C. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. Journal of Statistical Software. 2010;34:1–29.
194. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. Breakthroughs in Statistics. New York, NY: Springer New York; 1992. pp. 610–24.
195. Mahieu NG, Huang X, Chen Y-J, Patti GJ. Credentialing Features: A Platform to Benchmark and Optimize Untargeted Metabolomic Methods. Analytical chemistry. American Chemical Society; 2014;86:9583–9.
196. Benton HP, Ivanisevic J, Mahieu NG, Kurczy ME, Johnson CH, Franco L, Rinehart D, Valentine E, Gowda H, Ubhi BK, et al. Autonomous metabolomics for rapid metabolite identification in global profiling. Analytical chemistry. American Chemical Society; 2015;87:884–91.
197. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res. Oxford University Press; 2012;41:gks1065–D807.
198. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, et al. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. Oxford University Press; 2009;37:D603–10.
199. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res. Oxford University Press; 2007;35:D521–6.
200. McGaw EA, Phinney KW, Lowenthal MS. Comparison of orthogonal liquid and gas chromatography-mass spectrometry platforms for the determination of amino acid concentrations in human plasma. Journal of chromatography A. 2010;1217:5822–31.
201. Kvitvang HFN, Kristiansen KA, Bruheim P. Assessment of capillary anion exchange ion chromatography tandem mass spectrometry for the quantitative profiling of the phosphometabolome and organic acids in biological extracts. Journal of chromatography A. 2014;1370:70–9.
202. Lien SK, Kvitvang HF, Bruheim P. Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography-tandem mass spectrometry quantification of silylated metabolites. Journal of chromatography A. 2012 ed. 2012;1247:118–24.
203. Hartonen M, Mattila I, Ruskeepää A-L, Orešič M, Hyötyläinen T. Characterization of cerebrospinal fluid by comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. Journal of chromatography A. 2013;1293:142–9.
204. Lien SK, Kvitvang HFN, Bruheim P. Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography-tandem mass spectrometry quantification of silylated metabolites. Journal of chromatography A. 2012;1247:118–24.
205. Wink DA, Hines HB, Cheng RYS, Switzer CH, Flores-Santana W, Vitek MP, Ridnour LA, Colton CA. Nitric oxide and redox mechanisms in the immune response. J Leukoc Biol. Society for Leukocyte Biology; 2011;89:873–91.
206. Cross AJ, Major JM, Sinha R. Urinary Biomarkers of Meat Consumption. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2011;20:1107–11.
207. del Favero S, Roschel H, Solis MY, Hayashi AP, Artioli GG, Otaduy MC, Benatti FB, Harris RC, Wise JA, Leite CC, et al. Beta-alanine (Carnosyn™) supplementation in elderly subjects (60–80 years): effects on muscle carnosine content and physical capacity. Amino Acids. 2011;43:49–56.
208. Stout JR, Graves BS, Smith AE, Hartman MJ, Cramer JT, Beck TW, Harris RC. The effect of beta-alanine supplementation on neuromuscular fatigue in elderly (55-92 Years): a double-blind randomized study. J Int Soc Sports Nutr. BioMed Central; 2008;5:21.
209. Eyre H, Kahn R, Robertson RM, Clark NG, Doyle C, Gansler T, Glynn T, Hong Y, Smith RA, Taubert K, et al. Preventing Cancer, Cardiovascular Disease, and Diabetes: A Common Agenda for the American Cancer Society, the American Diabetes Association, and the American Heart Association*†. CA: A

Cancer Journal for Clinicians. John Wiley & Sons, Ltd; 2004;54:190–207.
210. Choque B, Catheline D, Rioux V, Legrand P. Linoleic acid: Between doubts and certainties. Biochimie. 2014;96:14–21.
211. Ramsden CE, Zamora D, Majchrzak-Hong S, Faurot KR, Broste SK, Frantz RP, Davis JM, Ringel A, Suchindran CM, Hibbeln JR. Re-evaluation of the traditional diet-heart hypothesis: analysis of recovered data from Minnesota Coronary Experiment (1968-73). BMJ. British Medical Journal Publishing Group; 2016;353:i1246.
212. Ramsden CE, Zamora D, Leelarthaepin B, Majchrzak-Hong SF, Faurot KR, Suchindran CM, Ringel A, Davis JM, Hibbeln JR. Use of dietary linoleic acid for secondary prevention of coronary heart disease and death: evaluation of recovered data from the Sydney Diet Heart Study and updated meta-analysis. BMJ. British Medical Journal Publishing Group; 2013;346:e8707–7.
213. Gunnerud U, Jansson N. Hälsomässiga effekter av palmolja - från vetenskap till folkhälsa [Internet]. Swedish nutrition foundation. Available from: http://snf.ideon.se/wp-content/uploads/2015/01/palmoljerapport_FINALversion.pdf
214. Madigan C, Ryan M, Owens D, Collins P, Tomkin GH. Dietary unsaturated fatty acids in type 2 diabetes Higher levels of postprandial lipoprotein on a linoleic acid-rich sunflower oil diet compared with an oleic acid-rich olive oil diet. Diabetes care. 2000;23:1472–7.
215. Biskup I, Kyrø C, Marklund M, Olsen A, van Dam RM, Tjønneland A, Overvad K, Lindahl B, Johansson I, Landberg R. Plasma alkylresorcinols, biomarkers of whole-grain wheat and rye intake, and risk of type 2 diabetes in Scandinavian men and women. The American journal of clinical nutrition. American Society for Nutrition; 2016;:ajcn133496.
216. Oishi K, Yamamoto S, Itoh N, Nakao R, Yasumoto Y, Tanaka K, Kikuchi Y, Fukudome S-I, Okita K, Takano-Ishikawa Y. Wheat alkylresorcinols suppress high-fat, high-sucrose diet-induced obesity and glucose intolerance by increasing insulin sensitivity and cholesterol excretion in male mice. The Journal of nutrition. American Society for Nutrition; 2015;145:199–206.
217. Ross AB, Svelander C, Savolainen OI, Lind MV, Kirwan JP, Breton I, Godin J-P, Sandberg A-S. A high-throughput method for liquid chromatography-tandem mass spectrometry determination of plasma alkylresorcinols, biomarkers of whole grain wheat and rye intake. Analytical Biochemistry. 2016;499:1–7.
218. Zhao Y, Monahan FJ, McNulty BA, Gibney MJ, Gibney ER. Effect of vitamin E intake from food and supplement sources on plasma α- and γ-tocopherol concentrations in a healthy Irish adult population. Br J Nutr. Cambridge University Press; 2014;112:1575–85.
219. Lampe JW, Huang Y, Neuhouser ML, Tinker LF, Song X, Schoeller DA, Kim S, Raftery D, Di C, Zheng C, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. The American journal of clinical nutrition. 2016;:ajcn144840.
220. Bere E, Brug J. Towards health-promoting and environmentally friendly regional diets – a Nordic example. Public Health Nutr. Cambridge University Press; 2009;12:91–6.
221. Xu R, Zhang S, Tao A, Chen G, Zhang M. Influence of Vitamin E Supplementation on Glycaemic Control: A Meta-Analysis of Randomised Controlled Trials. Brennan L, editor. PloS one. Public Library of Science; 2014;9:e95008.
222. Farvid MS, Jalali M, Siassi F, Hosseini M. Comparison of the Effects of Vitamins and/or Mineral Supplementation on Glomerular and Tubular Dysfunction in Type 2 Diabetes. Diabetes care. American Diabetes Association; 2005;28:2458–64.
223. Gómez-Pérez FJ, Valles-Sanchez VE, López-Alvarenga JC, Choza-Romero R, Ibarra Pascuali JJ, González Orellana R, Pérez Ortiz OB, Rodríguez Padilla EG, Aguilar Salinas CA, Rull JA. Vitamin E modifies neither fructosamine nor HbA1c levels in poorly controlled diabetes. Rev Invest Clin. 1996;48:421–4.
224. Montonen J, Knekt P, Järvinen R, Reunanen A. Dietary Antioxidant Intake and Risk of Type 2 Diabetes. Diabetes care. American Diabetes Association; 2004;27:362–6.
225. Mayer-Davis EJ, Costacou T, King I, Zaccaro DJ, Bell RA. Plasma and Dietary Vitamin E in Relation to Incidence of Type 2 Diabetes. Diabetes care. American Diabetes Association; 2002;25:2172–7.
226. Suksomboon N, Poolsup N, Sinprasert S. Effects of vitamin E supplementation on glycaemic control in type 2 diabetes: systematic review of randomized controlled trials. Journal of Clinical Pharmacy and Therapeutics. Blackwell Publishing Ltd; 2011;36:53–63.
227. Evert AB, Boucher JL, Cypress M, Dunbar SA, Franz MJ, Mayer-Davis EJ, Neumiller JJ, Nwankwo R, Verdi CL, Urbanski P, et al. Nutrition therapy recommendations for the management of adults with diabetes. Diabetes care. American Diabetes Association; 2013;36:3821–42.
228. Plourde M, Cunnane SCCC. Extremely limited synthesis of long chain polyunsaturates in adults: implications for their dietary essentiality and use as supplements. Applied Physiology, Nutrition, and Metabolism. 2007;32:619–34.
229. Burdge GC, Calder PC. Conversion of alpha-linolenic acid to longer-chain polyunsaturated fatty acids in

human adults. Reprod Nutr Dev. 2005;45:581–97.
230. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015;163:1079–94.
231. Li S, Shin HJ, Ding EL, van Dam RM. Adiponectin Levels and Risk of Type 2 Diabetes: A Systematic Review and Meta-analysis. JAMA. American Medical Association; 2009;302:179–88.
232. Kadowaki T, Yamauchi T, Kubota N, Hara K, Ueki K, Tobe K. Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. The Journal of clinical investigation. American Society for Clinical Investigation; 2006;116:1784–92.
233. Steyn NP, Mann J, Bennett PH, Temple N, Zimmet P, Tuomilehto J, Lindström J, Louheranta A. Diet, nutrition and the prevention of type 2 diabetes. Public Health Nutr. 2004;7:147–65.
234. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. The Lancet. 2014;383:1999–2007.