Delineating a Metabolomic Signature for the Transition from Gestational Diabetes to Type 2 Diabetes

By

Amina Allalou

A thesis submitted in conformity with the requirements for the degree of Masters of Science in Medical Sciences

Institute of Medical Science
University of Toronto

© Copyright by Allalou Amina 2015
Delineating a Metabolomic Signature for the Transition from Gestational Diabetes to Type 2 Diabetes

Amina Allalou
Masters of Science
Institute of Medical Science
University of Toronto
2015

Abstract
Although gestational diabetes (GDM) is of a transient nature, up to 50% of women with GDM develop type 2 diabetes (T2D) within 5 years. Despite this extremely high risk, post-partum screening remains low. Here, we delineated a predictive metabolomic signature of disease transition. The study patients were part of the SWIFT Cohort, which follows 1010 GDM women, 109 of which developed T2D two years post-partum. Future cases were matched to future controls based on ethnicity, age and pre-pregnancy BMI. Hexoses, specific amino acids, specific sphingomyelin and specific phosphatidylcholine species associated with T2D. Through a J48 decision tree predictive modelling in the training set, future T2D incidence was predicted in an independent testing set with discriminative power 0.769 (P<0.001) and sensitivity 74%, which was more predictive than fasting plasma glucose alone. This study represents the first metabolomics study of the transition from GDM to T2D.
Acknowledgments

I am grateful to all members of the Wheeler Lab, to our study collaborators, to the staff at the AFBM, as well as all committee members for taking the time to provide feedback. Sincere commendations are also due to the building custodians, and to my family, neighbours and friends for their continuous support.
List of Contributions

The author gratefully acknowledges the contributions of the Analytical Facility of Small Active Biomolecules (The Hospital for Sick Children, Toronto, ON, Canada) for the mass spectrometry work in the analysis of metabolite concentrations. The author also acknowledges Katherine Leavey for the construction of the PVCA plots.
# Table of Contents

Acknowledgments...

List of Contributions

Table of Contents

List of Abbreviations

List of Tables

List of Figures

List of Appendices

1 Literature Review

1.1 Introduction: Diabetes Epidemiology

1.2 Type 2 Diabetes

1.2.1 Definition

1.2.2 Causes and Correlates of T2D: -Omic Studies

1.2.2.1 Carbohydrates

1.2.2.2 Free Fatty Acids

1.2.2.3 Amino Acids

1.2.2.4 Sphingomyelin

1.2.2.5 Phosphatidylcholine Species

1.2.3 Risk Factors for T2D

1.2.3.1 Physical activity

1.2.3.2 Diet

1.2.3.3 Obesity

1.2.3.4 Age

1.2.3.5 Ethnicity and Family History

1.2.3.6 Gestational Diabetes (GDM)
5.3 Design of Experiment .......................................................... 86
  5.3.1 Anthropomorphic measures ........................................... 86
  5.3.2 Correlations .............................................................. 86
  5.3.3 Metabolite Associated with T2D Incidence ....................... 87
     5.3.3.1 Hexoses are elevated in future T2D ......................... 87
     5.3.3.2 Free Fatty Acids are not associated with future T2D .... 88
     5.3.3.3 Sphingomyelins are decreased in Incident T2D .......... 89
     5.3.3.4 Amino Acids are altered in Incident T2D ................. 90
     5.3.3.5 Ether-linked Phosphatidylcholine Species are decreased in T2D .... 92
  5.3.4 Metabolites Predictive of T2D ...................................... 93
  5.4 Data Strengths ................................................................ 96
  5.5 Limitations of Study ....................................................... 97
  5.6 Division of Duties .......................................................... 99
6 Conclusion ............................................................................ 100
7 Future Directions ................................................................ 101
  7.1 Expanding Signature Generalizability .............................. 101
  7.2 Exploring Mechanistic Implications of Metabolite Signature ... 103
  7.3 Facilitating Metabolomic Investigations ......................... 104
References ............................................................................. 106
Appendices ........................................................................... 122
Copyright Acknowledgements (if any) ................................. 135
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2hPG</td>
<td>2-hour post glucose load</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acids</td>
</tr>
<tr>
<td>AC</td>
<td>Acylcarnitine</td>
</tr>
<tr>
<td>ADA</td>
<td>American Diabetes Association</td>
</tr>
<tr>
<td>BCAA</td>
<td>Branched-chain amino acids</td>
</tr>
<tr>
<td>CDA</td>
<td>Canadian Diabetes Association</td>
</tr>
<tr>
<td>FFA</td>
<td>Free fatty acid</td>
</tr>
<tr>
<td>FPG</td>
<td>Fasting plasma glucose</td>
</tr>
<tr>
<td>GC/MS</td>
<td>Gas chromatography/Mass Spectrometry</td>
</tr>
<tr>
<td>GDM</td>
<td>Gestational Diabetes Mellitus</td>
</tr>
<tr>
<td>GSIS</td>
<td>Glucose-stimulated insulin secretion</td>
</tr>
<tr>
<td>IADPSG</td>
<td>International Association of Diabetes and Pregnancy Study Groups</td>
</tr>
<tr>
<td>IDF</td>
<td>International Diabetes Federation</td>
</tr>
<tr>
<td>IFG</td>
<td>Impaired fasting glucose</td>
</tr>
<tr>
<td>IGT</td>
<td>Impaired glucose tolerance</td>
</tr>
<tr>
<td>IRS-1</td>
<td>Insulin Receptor Substrate 1</td>
</tr>
<tr>
<td>LC/MS</td>
<td>Liquid chromatography/Mass Spectrometry</td>
</tr>
<tr>
<td>LPC</td>
<td>Lysophosphatidyl choline</td>
</tr>
<tr>
<td>OGGT</td>
<td>Oral glucose tolerance test</td>
</tr>
<tr>
<td>PP</td>
<td>Post-partum</td>
</tr>
<tr>
<td>RandomPG</td>
<td>Random plasma glucose</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Curve</td>
</tr>
<tr>
<td>SM</td>
<td>Sphingomyelins</td>
</tr>
<tr>
<td>T1D</td>
<td>Type 1 Diabetes Mellitus</td>
</tr>
<tr>
<td>T2D</td>
<td>Type 2 Diabetes Mellitus</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
List of Tables

**Table 1** Approximation equation for measures of insulin resistance and β-cell function, first derived by Matthews et al (1985). Glucose in mM and insulin in mU/L.

**Table 2** Classification of Type 2 Diabetes (T2D), Impaired Glucose Tolerance (IGT) and Impaired Fasting Glucose (IFG) according to various health associations. All glucose values in mM.

**Table 3** Categories and definitions of race and ethnicity according to the FDA 2014 guidelines.

**Table 4** Latest GDM Diagnostic criteria, per organization. Presently, the two-step system has been replaced in favour with a direct OGTT. Note that the WHO criteria is the most inclusive, and may thus inflate future increased GDM prevalence. All glucose values given are in mM.

**Table 5** Terminology used in predictive modelling

**Table 6** Mass transition monitored for amino acid quantification

**Table 7** Summary of metabolite subpanels

**Table 8** Internal controls used to validate method and to correct for batch effect

**Table 9** Baseline (6-9 weeks post-partum) and post-partum characteristics of SWIFT patients in the training and testing set (n=126 pairs).

**Table 10** Individual metabolites and metabolite classes that significantly differ in future T2D in the training set (n=160). Fold change refers to T2D/nonT2D.

**Table 11** Paired t-test values of metabolites associated with future T2D incidence, as well as significance following removal of all outliers with the sensitivity analysis P value. Fold change=T2D/nonT2D

**Table 12** Descriptive of validation set of T2D cases beyond 2 years.

**Table 13** Baseline (6-8 weeks) characteristics of normal and glucose impaired women in the SWIFT cohort (n=156)

**Table 14** Association with impaired status of top 24 metabolites t (n=158). Fold change refers to Imp/NGT.
List of Figures

**Fig 1** Natural History of T2D, depicting mean OGTT values at various stages of hyperglycemia
A) Progressive decline in glucose uptake as insulin resistance increases; B) Increasing insulin secretion to compensate for IR (Starling’s curve of the pancreas); C) Increasing plasma glucose levels as disease progresses; OB: Obese; T2D Early: low IR; T2D late: high IR (Adapted from DeFronzo, 2009).

**Fig 2** Sphingomyelin synthesis and breakdown. Initial condensation reaction in the endoplasmic reticulum by serine palmitoyl transferase (SPT) results in 3-keto sphinganine, which is rapidly reduced by KSA-reductase and acylated by ceramide synthase (CerS) to produce dihydroceramide. In the next major step which occurs in the trans Golgi or at the plasma membrane, a phosphocholine head group is transferred to ceramide, yielding a sphingomyelin and diacylglycerol as a by-product. The breakdown of SM by SMase either in the plasma or within the lysosome results in the release of ceramide. Adapted from Slotte et al 2013.

**Fig 3** Work Flow for Big Data predictive modeling. Data must first be collected, curated and then pre-processed, the strategy chosen dependent on the deployed algorithm.

**Fig 4** Understanding outcomes of predictive modeling, definitions described in the text. Sensitivity, also known as the True Positive Rate and Recall; Specificity, or True Negative Rate; Precision, or Positive Predictive Value; n is the sample size, or total population

**Fig 5** The ROC space. A model along the diagonal indicates no predictive ability better than chance. Perfect classification is presented by the red dot.

**Fig 6** Study design of the SWIFT cohort. A total of 1035 women were enrolled, and followed annually for 2 years, with an annual retention rate of 96% and 84%, respectively. By 2 year post-partum, 109 women (10.8%) had developed T2D. An additional 21 women developed T2D beyond 2 years post-partum. Both GDM and T2D were diagnosed as per ADA criteria.

**Fig 7** Workflow for SWIFT plasma samples through the various metabolomics subpanels. Upon arrival, SWIFT samples were aliquoted and processed through each subpanel for a concentration read out, quality control. A total of 185 metabolites were assayed. B-HB: Beta-hydroxybutyrate

**Fig 8** A) Chromatograph of FFA in the standards, showing excellent resolution; B) Zoomed in chromatograph of oleate and vaccinate, the double peak providing able to resolve the two FFA species (21.86 min and 21.91 min); C) Zoomed in chromatograph of a random sample, demonstrating overlap of the two peaks and thus inability to independently characterize concentration of both species (time 21.88 min)

**Fig 9** Bland-Altman plot, comparing the summation of Leu and Ile (xLeu) from the Biocrates and Special AA subpanels. The majority of metabolites fell within 2 STDEV, as presented by the red lines. There was a significant decrease of 16 uM in the Special AA subpanel compared to the Biocrates technology. This decrease was not correlated to increasing total concentration, as demonstrated by the regression line (in blue, $r^2 = 0.007$) and its 95% confidence interval.
Fig 10 Scatter plot of hexose and glucose values from the two different methods, showing the expected positive correlation.

Fig 11 Box plot demonstrating median and IQR of metabolite concentration across various batches for glucose and isoleucine. Circle (•) indicates an outlier, while a star (*) indicates an extreme outlier.

Fig 12 Two-tailed Pearson correlation r heat-map of all 111 metabolites to each other (n=262).

Fig 13 PVCA identified Ethnicity, pre-pregnancy BMI and group membership to be responsible for high portions of variance of metabolites.

Fig 14 Histogram with normal curve fit of the top 4 metabolites. No one metabolite demonstrated a complete separation in the histogram of future T2D compared to those who remain nonT2D. (n=156)

Fig 15 ROC of the J48 algorithm on the training set (A) and testing set (B), performing with discriminative power 0.835 (P< 10^-6) and 0.755 (P<0.001), respectively. This is much greater than FPG alone (0.584 and 0.618, respectively, NS).

Fig 16 Decision tree by J48 demonstrating the use of PCaeC40:5, hexose, branched chain amino acids (BCAA) and SM(OH)C14:1. The grey boxes indicate the metabolite chosen for the node, while the clear numbered boxes indicate the concentration threshold used for branching (in uM for all metabolites except for hexose, which is in mM). A blue or green box indicates the group (T2D or nonT2D) associated with the concentration range. The percentage below each group indicates percent of instances correctly classified.

Fig 17 Baseline amino acid concentrations correlations with baseline HOMA-IR, a measure of insulin resistance. HOMA IR was moderately correlated with Ile (0.40), 2-AAA (0.36) and Gly (-0.33), and weakly correlated with Tyr (0.27).

Fig 18 Baseline SM and PC concentrations correlations with baseline HDL cholesterol levels. HDL was moderately correlated with the sum of all SM (0.47) but only weakly with SMC20:2 (0.25). Furthermore, PCaeC40:5 and PCaeC44:5 were both weakly correlated with HDL (0.27 and 0.29, respectively).

Fig 19 Amino acids glycine, isoleucine, 2-AAA and tyrosine across lactation intensity group. Exclusive formula-feeding was associated with decreased glycine, increased isoleucine and increased 2-AAA (P<0.05).

Fig 20 Proposed mechanism combining major metabolomics players in progression to T2D, as expanded upon in the text.
List of Appendices

Appendix 1 Test Run Quality Control Parameters

Appendix 2 SWIFT Sample Run Quality Control Parameters

Appendix 3 Metabolite Concentration Distribution and Outliers
1 Literature Review

1.1 Introduction: Diabetes Epidemiology

The first recorded mention of diabetes dates back to over 3000 years ago by the ancient Egyptian Chief of Dentists and Physicians Hesy-Ra, which he termed “the passing of too much urine”. In 250 BCE, Greek physician Apollonius, born in Egypt, would use the Greek word for ‘passing through’ or diabetes, to name the disease. This diagnosis of diabetes from urine would be espoused by physicians from all around the world, including the Chinese, Indian, and Persian civilizations. Treatment however following such a diagnosis was scant. It was only until the 1900, with the discovery of insulin as an effective treatment by Canadian scientists Banting and Best at the University of Toronto, that the role of the pancreas in diabetes mellitus was established. What was once death sentence had now potential treatment.

Since the industrial revolution, diabetes prevalence has increased at epidemic proportions. With a worldwide prevalence of 317 million (WHO), 3 million of which reside in Canada (CDA), the personal, social and economic impact of diabetes render it a disease to be reckoned. This increase is commonly ascribed to multiple risk factor, including changes in lifestyle and food consumption. Exploring the *tolle causem* (etiology, pathogenesis and pathophysiology) of diabetes are thus of great relevance.

Diabetes is typically classified into 3 major types: type 1 diabetes mellitus (T1D) refers to an autoimmune dysfunction of the beta cells, resulting in deregulated insulin secretion and glucose homeostasis. Type two diabetes mellitus (T2D), which accounts for over 90% of diabetes cases, is a failure of pancreatic beta cells fail to compensate for increasing insulin demand. The third type of diabetes, gestational diabetes mellitus (GDM), is a transient condition in which glucose intolerance develops during the late 2\textsuperscript{nd} trimester, and is resolved upon parturition. GDM is frequently thought of as a transient form of T2D due to similar underlying causes, such as increased insulin resistance. This classification system is however a gross oversimplification, and hence it may be more useful to think of T1D and T2D as a spectrum, with latent autoimmune diabetes in adults (LADA), maturity-onset diabetes of the young (MODY) and other subtypes as intermediaries between T1D and T2D (Prasad and Groop 2015).
1.2 Type 2 Diabetes

1.2.1 Definition

T2D is first characterized by peripheral insulin resistance and compensatory beta cell action, resulting in both plasma hyperglycemia, hyperinsulinemia, as well as beta cell hyperplasia and hypertrophy. This compensated state of high insulin resistance may persist for years and in fact may not progress to T2D. The gold standard approach for measuring insulin resistance (IR) is through a hyperinsulinemic euglycemic clamp. However, due to the complicated, costly, labour intensive and risk of hypoglycemic events in using this approach, IR is commonly calculated in large epidemiological studies using the Homeostatic Model Assessment (HOMA). The HOMA-IR requires fasting plasma glucose and insulin values only (Buchanan et al 2010), and physiologically crudely refers to the feedback loop between hepatic glucose efflux and beta cell insulin secretion in 1) glucose- and insulin-dependent uptake by skeletal muscle and adipose tissue and 2) glucose-dependent brain uptake and kidney loss (Wallace et al 2004). This model, initially written by Matthews et al (1985) and updated by Levy et al (1998), correlates well with the clamp method (coefficient of variation of 0.88) and was originally calibrated to give a normal IR of 1. Currently, there are no standardized ranges that delineate normal HOMA-IR values. This is because, although a popular threshold of 2.5 exists, with higher values indicating IR (Salgado et al 2010), this threshold differs depending on gender and ethnicity (Qu et al 2011). Therefore, the HOMA-IR is a useful measure to compare between and within groups, but not necessarily to assess IR in one individual patient (Wallace et al 2004). The HOMA also includes a calculation of beta cell activity, or HOMA-%B, calibrated to give normal function as 100%. The approximation equations for both HOMA-IR and HOMA-%B are summarized in Table 1.

Table 1 Approximation equation for measures of insulin resistance and β-cell function, first derived by Matthews et al (1985). Glucose in mM and insulin in mU/L.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOMA-IR</td>
<td>$\frac{\text{fasting glucose} \times \text{fasting insulin}}{22.5}$</td>
</tr>
<tr>
<td>HOMA-%B</td>
<td>$\frac{20 \times \text{fasting insulin}}{\text{glucose} - 3.5} %$</td>
</tr>
</tbody>
</table>
The pathogenesis of diabetes is traditionally represented by Starling’s curve (Fig 1). As mean glucose uptake decreases in insulin responsive skeletal and adipose tissue over increasing BMI (Fig 1A), plasma insulin proportionally increase (Fig 1B). This compensation is imperfect, as noted by the gradual increase in plasma glucose (Fig 1C). Ultimately, by a mechanism still unclear, in a proportion of individuals the beta cell fails to keep up with ever increasing insulin secretion in response to increasing insulin resistance, as represented by the peak in Fig1B. This decompensation leads to an initial increase in post-prandial glucose (such as 2hPG), followed by an increase in fasting plasma glucose (FPG, Fig 1C), and ultimately T2D. The levels of FPG and 2hPG that define T2D have been the subject of much controversy in the past. Table 2 illustrates the thresholds used by various organizations. These thresholds were initially chosen as they marked increased prevalence and incidence of diabetic retinopathy, a complication of T2D (National Diabetes Data Group 1979). In 2006, the WHO and IDF collaboratively released new guidelines for the definition and classification of diabetes mellitus (WHO 2006). In 2011, the WHO also approved the use of levels of glycated hemoglobin (HbA1c) of greater than 6.5% as indicative of T2D. A value equal or below 6.5% does not however exclude T2D cases, and plasma glucose values must be measured for validation (WHO 2011). Although it is costly, the HbA1c allows for measurement of long term glucose concentrations, and is thus a better predictor of microvascular and macrovascular complications.

The natural history of T2D also involves multiple organ systems, including the pancreatic beta cell and alpha cell, liver, muscle, adipose tissue, gastrointestinal tract, the kidney and the brain, parsimoniously referred to as the ominous octet by DeFronzo (2009). In other words, T2D effects are beyond the pancreas, and involve deregulated metabolic homeostasis at the level of the organism. System’s biology-based approaches to the study of T2D are thus of paramount importance. Importantly, a variety of complications arise from untreated T2D, including diabetic retinopathy (the most common complication), neuropathy, nephropathy, peripheral vascular diseases such gangrene leading to amputation, and cardiovascular disease. In fact, extensive tissue damage to these organs can occur before noticeable hyperglycemia and decompensation (Nerpin et al 2008). Presently, 50% of diabetic patients die of cardiovascular disease, primarily due to heart attack and stroke (Moorish et al 2001)
Fig 1 Natural History of T2D, depicting mean OGTT values at various stages of hyperglycemia; A) Progressive decline in glucose uptake as insulin resistance increases; B) Increasing insulin secretion to compensate for IR (Starling’s curve of the pancreas); C) Increasing plasma glucose levels as disease progresses; OB: Obese; T2D Early: low IR; T2D late: high IR (Adapted from DeFronzo, 2009).
Table 2 Classification of Type 2 Diabetes (T2D), Impaired Glucose Tolerance (IGT), Impaired Fasting Glucose (IFG) and Prediabetes according to various health associations. Glucose in mM.

<table>
<thead>
<tr>
<th></th>
<th>World Health Organization - 2011</th>
<th>Canadian Diabetes Association</th>
<th>American Diabetes Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2D</td>
<td>FPG ≥ 7.0 or 2hPG ≥ 11.1 or HbA1c ≥ 6.5%</td>
<td>FPG ≥ 7.0 or 2hPG ≥ 11.1 or HbA1c ≥ 6.5% or RandomPG ≥ 11.1 + symptoms</td>
<td>FPG ≥ 7.0 or 2hPG ≥ 11.1 or HbA1c ≥ 6.5% or RandomPG ≥ 11.1 + symptoms</td>
</tr>
<tr>
<td>IGT</td>
<td>2hPG 7.8– 11.1 and FPG &lt; 7.0</td>
<td>2hPG 7.8– 11.0</td>
<td>2hPG 7.8– 11.0</td>
</tr>
<tr>
<td>IFG</td>
<td>FPG 6.1– 6.9 and 2hPG &lt; 7.8</td>
<td>FPG 6.1– 6.9</td>
<td>FPG 5.6– 6.9</td>
</tr>
<tr>
<td>Prediabetes or Impaired</td>
<td>2hPG 7.8– 11.1 or FPG 6.1– 6.9</td>
<td>HbA1C 6.0– 6.4%</td>
<td>HbA1C 5.7– 6.4%</td>
</tr>
</tbody>
</table>

1.2.2 Causes and Correlates of T2D: -Omic Studies

As a multi-organ disease, diabetes is the result of multiple converging factors. The current rise in T2D incidence has been primarily attributed to the parallel increase in obesity and sedentary life style, both of which promote the development of insulin resistance (Koivisto et al 1986; Assali et al 2001). Various -omic studies have shed much insight into the etiology of the disease in Homo sapiens.

In genomic studies, to date, approximately 70 SNPs have been associated with T2D incidence (Reviewed by Prasad and Groop 2015). Nevertheless, while genomics can explain 80% of T1D heritability, this figure dramatically falls to 10% for T2D (Sun et al 2014). The heritability for T2D percentage may increase in certain ethnicities known to have greater T2D risk – however this remains to be studied.

Proteomics studies have also advanced our understanding of T2D. Of the most consistent trends, apolipoproteins A1 and E, which remove cholesterol from the blood, were extremely reduced in T2D patients and hence worsen the dyslipidemia characteristic of the disease. On the other hand, apolipoproteins CII and CIII, which promote atherosclerotic plaques and coronary
heart disease, were elevated in T2D patients (Li et al, 2008; Liu et al, 2014). Nevertheless, it must be noted that only a few plasma proteomic studies of T2D have been carried out, and the latter suffer from small cohort size and do not consistently specify the fasting/non-fasting status of the subjects. Furthermore, the systematic proteomic approach is still faced with significant methodical challenges with regards to the great chemical heterogeneity of proteins (hydrophobic vs hydrophilic) as well as the large dynamic range in concentration of proteins.

As a metabolic syndrome, metabolomics may provide a greater understanding of cause of T2D. In contrast to genomic and proteomic approaches, metabolomics focuses on the end product, the metabolites, giving a more accurate picture of the beta cell extracellular environment. By far, the most prevalent mechanistic explanation of T2D is through the glucolipotoxicity model, a model that posits various metabolites as key players in beta cell dysfunction (Poitout et al 2010). This model asserts that the synergetic interaction between hyperglycemia (metabolites such as glucose and other carbohydrates) and hyperlipidemia (metabolites such as free fatty acids and other lipids) induces toxic pathways, resulting in beta cells apoptosis and/or reduced functionality.

Over a dozen untargeted and targeted metabolomics studies on T2D have been published over the last decade using this novel technological advancement (Yi et al, 2008; Tan et al, 2009; Zhang et al, 2009; Adams et al, 2009; Fiehn et al, 2010; Mihalik et al, 2010; Suhre et al 2010; Wang et al, 2011; Rui Wang et al, 2012; Ha et al, 2012; Mihalik et al, 2012; Floegel et al, 2012; Kaur, et al 2012; Wang et al 2013; Xu et al, 2013; Menni et al, 2013; Ferrannini et al, 2013; Prentice et al, 2014). These studies varied with regards to sample size, population characteristics (ethnicity, age, and others) as well as study design.

design from several large and well established cohorts (the Framingham Heart Study, KORA, EPIC-Postdam and Malmö Diet and Cancer Study), allowing for characterization of metabolomic changes years before T2D incidence. Lastly, while two studies were found to have focused on Black incident T2D cases (Adams et al, 2009; Fiehn et al 2010), no cohort with a predominantly Hispanic population has been investigated. In addition, several of the aforementioned studies compared other clinical categories of relevance in data interpretation, such as comparing the impaired status metabolome to T2D incidence (Rui Wang et al 2012), the T2D metabolome with T2D complicated by cardiovascular disease (Yi et al 2008), obese T2D patients compared to non-obese T2D patients (Fiehn et al 2010), or focused on adolescents afflicted by T2D (Mihalik et al 2012).

The results from all of the above 18 metabolomic studies were aggregated to delineate any consistent pattern. Notably, over 560 unique metabolites were analyzed in all of the combined studies, with 304 metabolites found significant in at least 1 study, and 87 significant in greater than 2 studies. Importantly, retrospective studies revealed changes in particular metabolite concentrations years before T2D incidence. These metabolites belonged to several macromolecule classes, including amino acids, free fatty acids, sphingomyelin species, phosphatidylcholine species and carbohydrates. The effects and proposed pathways by which the metabolites may induce insulin resistance and/or beta cell dysfunction are explored below.

1.2.2.1 Carbohydrates

Hyperglycemia being the *sine qua non* of T2D, it would come to no surprise that levels of glucose are elevated in T2D compared to controls. Whether glucose itself causes T2D (metabolic dysregulation) or is instead a marker of T2D is not simply a matter of semantics. Janket et al (2003) did not find an association between total carbohydrate intake and future T2D incidence in a large cohort of 39 000 women. In contrast, Sluijs et al (2010) investigating in a cohort of a similar size, found a significant increased diabetes risk in individuals with diets high in glucose. In general, glucose levels gradually rise before T2D onset, in a state termed ‘pre-diabetes’, which is further explored in section 1.2.4. The current understanding is that glucose is both a cause as well as a marker of metabolic dysregulation, feeding a progressive vicious cycle of deterioration.

The first phase in the pathogenesis of T2D, characterized by increasing insulin resistance, is not commonly ascribed to chronically elevated levels of plasma glucose. Instead, elevated
levels of glucose constantly activate beta cell insulin secretion (and hence hyperinsulinemia), and this produces a toxic effect via reduced insulin gene expression (beta cell desensitization), increased endoplasmic reticulum stress due to the constant demand on insulin synthesis (beta cell exhaustion), and increased oxidative stress ultimately leading to apoptosis (beta cell death) (Reviewed by Cernea and Dobrenau, 2013). The damaged beta cell exacerbates the existing hyperglycemia, further accelerating beta cell insufficiency (Maedler et al, 2003).

Furthermore, elevated glucose levels are a more explanatory variable for the risk of diabetic complications. In the presence of repeated spikes of high plasma glucose, the intima-media of the vasculature thickens, increasing the risk for cardiovascular disease, including myocardial infraction and stroke (Temelkova-Kurktschiev et al, 2000). In a murine model of age-related macular degeneration, a diet that induced hyperglycemia resulted in the accumulation of advanced glycation end products (AGE) in multiple organs. Furthermore, there is increased glycation of the ubiquitin-proteasome system under hyperglycemia, thereby further inhibiting the degradation of AGE and inducing oxidative stress. This resulted in increased lesions in the retina of the mice models (Uchiki et al, 2012).

Importantly and of potential surprise, other carbohydrates besides glucose are also reported to be altered in T2D. For example, fructose and mannose were observed elevated in Black, White and Asian cohorts (Fiehn et al 2010; Menni et al, 2013; Xu et al, 2013). Interestingly, a high fructose diet was positively associated with T2D in a Finnish nutrition study, although the investigators could not separate the effect of a high glucose diet (Montonen et al 2007). Furthermore, Xiao et al (2013) have demonstrated that fructose injections increases intestinal lipoprotein particle release and dietary fat absorption, ensuing augmented hypertriglyceridemia. Vitally, in contrast to glucose, fructose is known to stimulate hepatic lipogenesis (Reviewed by Samuel, 2011), which induces further insulin resistance, a key step in T2D etiology.

The sum total of all 6 carbon sugars, or hexoses, was significantly elevated at least 5 years before a T2D diagnosis in 2 nested case-control studies (Floegel et al, 2012; Rui Wang et al, 2012). In addition to glucose, fructose and mannose, hexose also includes sugars such as allose, gulose, idose, galactose and talose. In other words, global dysregulation of carbohydrate metabolism may be a marker of exacerbated future glucose dysregulation. Alternatively, it may
be that a greater proportion of future cases was already in the impaired state at baseline compared to control in the aforementioned studies. In fact, the mean concentration of hexose in future cases in Floegel et al (2012) was 5.7 mM at baseline, compared to 4.7 mM in future controls.

1.2.2.2 Free Fatty Acids

In the various metabolomic T2D studies, there is a consistent significant trend of decrease concentration of short-chain free fatty acids, but an increase concentration in medium and long chain free fatty acids in T2D cases. In particular, the saturated free fatty acid myristate (C14:0), palmitate (C16:0) and stearate (C18:0) as well as the poly-unsaturated fatty acids palmitoleate (C16:1) and oleate (C18:1) were observed elevated across cohorts of different ethnicities (Yi et al, 2008; Tan et al, 2009; Suhre et al, 2010; Ha et al, 2012; Xu et al, 2013; Menni et al, 2013), although BMI was not well matched in all of the aforementioned studies. However, all of these studies were cross-sectional. Dismally, no metabolomics nested-case control study assayed free fatty acids to determine if there is a temporal correlation between increased free fatty acids and future T2D incidence.

Notwithstanding, it is well established that obesity induces insulin resistance, and thereby increasing the risk of T2D incidence. Insulin resistance is characterized by a combination of impaired insulin-stimulated glucose uptake (particularly in adipose and skeletal tissue) and defects in glycogen synthesis (particularly in the liver). In fact, as skeletal muscle accounts for approximately 40% of body weight and 80% of insulin-mediated glucose disposal, it is believed to contribute the most to peripheral insulin resistance (Bruni and Donati, 2009). Free fatty acids inhibit glucose uptake in skeletal muscle (even in the presence of insulin) by inducing the insulin receptor inhibition via protein kinase C, as well as inducing oxidative stress in the mitochondria by activating pro-inflammatory transcription factor NF-kB (Reviewed in Krebs and Roden, 2005). Hepatic insulin resistance is also induced by the increased portal flux of FFA, in combination with increased hormone leptin and decreased adiponectin, causing even greater hepatocellular lipid concentrations. These FFA in turn inhibit the insulin-dependent inhibition of endogenous glucose production, as well as inhibit the suppression of glycogenolysis (Boden et al 2002). The net result is thus an increase in glucose production, further exacerbating hyperglycemia.
Moreover, under the glucolipotoxicity model, it is proposed that the increase in free fatty acids (FFA) such as palmitate and stearate induces apoptosis of the beta cell (Welters et al, 2004). A variety of mechanisms have been proposed, including the formation of toxic ceramides (Maedler et al, 2000), induction of ER stress (Cui et al, 2013) and mitochondrial overload due to incomplete free fatty acid oxidation (Koves et al, 2008). Interestingly, saturated FFA have been shown to induce inflammation in adipocytes and macrophages through activation of NFkB and TLRs (Suganami et al, 2007). In addition, other FFA such as the unsaturated fatty acid arachidonic acid and alpha-linolenic acid, which also play a key role in inflammation, are also altered concentrations in diabetic serum (Tan et al, 2009). Whether inflammation due to FFA also plays a role in beta cell death is currently not known. Ultimately, the increased insulin resistance along with beta cell toxicity results in T2D incidence.

1.2.2.3 Amino Acids

Traditionally, T2D incidence was not known to associate with levels of amino acids. This changed when in Wang et al (2011) noted an increase in the branched chain amino acids (BCAA) valine, leucine and isoleucine, up to 12 years before T2D incidence. Multiple studies, both cross-sectional and longitudinal, have since replicated this trend (Floegel et al 2012; Fiehn et al 2010; Xu et al 2013; Menni et al 2013; Ferrannini et al 2013). Of note, Mihalik et al (2012) and Zhang et al (2009) reported significant yet decreased BCAA in T2D. This discrepancy may be due to the study populations. Mihalik et al (2012) investigated youth with T2D, a subset of T2D with plausibly different pathological mechanisms. In contrast, Zhang et al (2009) T2D was composed of both newly diagnosed or T2D patients treated with short acting oral hypoglycemic agents – the effect of these drugs on the metabolomics profile remains to be investigated.

While T2D was not known to associate with BCAA, the observation that obese insulin-resistant individuals have elevated BCAA dates back to the 1960s (Felig et al, 1969). This connection of BCAA and insulin resistance was further corroborated in a study by Newgard et al (2009) and Fiehn et al (2010), where a BCAA metabolic signature was correlated to insulin resistance in obese individuals. Lynch et al (2014) also found a positive association between plasma BCAA and models of obesity with insulin resistance (compared to lean) in several mammalian species.
Numerous hypotheses have been proposed positing elevated BCAA in the etiology of diabetes. As essential amino acids, the BCAA are acquired through primarily diet, and hence the correlation with obesity may be more direct. Nevertheless, BCAA plasma concentration are due to at least four factors: diet, genetics, gut microbiota and BCAA metabolism (Shah et al, 2011).

The connection between how these amino acids specifically contribute to insulin resistance has been studied intensively. In a study using two obese rat models, obesity was associated with higher BCAA levels, with reduced activity of key BCAA catabolic enzymes in the liver and adipose tissue (She et al, 2007). BCAA may also induce muscular insulin resistance through interaction of the IRS1 with S6K1 (Tremblay et al, 2007). A study by Krebs et al (2002) further implicates mTOR and JUN signaling pathway, while Huffman et al (2009) noted activation of the JNK phosphorylation activity in a rat model on a high fat diet supplemented with BCAA.

Alternatively, BCAA have been reported to stimulate beta cell insulin secretion (Floyd et al, 1966). Nilsson et al (2007) designed a clinical study where participants drank different glucose-equivalent drinks, and noted the most significant insulin secretagogues to be the BCAA along with lysine and threonine. Xu et al (2001) investigated the mechanism for leucine-mediated insulin secretion. Leucine was observed to allosterically activate glutamate dehydrogenase, further providing additional intermediates for the Krebs cycle. The researchers also noted stimulation by leucine of p70S6K phosphorylation via the mTOR pathway, leading to increased insulin synthesis. In the context of T2D induction, it may be that as hyperaminoacidemia promotes hyperinsulinemia, this may potentially result in beta cell exhaustion, although this has not yet been shown. Lastly, as the context of diabetes frequently involves lipotoxicity, an interplay may exist between BCAA and lipids. Through an animal model, Newgard et al (2012) proposed that BCAA result in excess flux of catabolic by products such as acylcarnitine C3 and C5, further worsening insulin resistance.

Conversely, BCAA may simply be associated with T2D incidence, as opposed to being causal. It has been reported that lactation is protective of future T2D incidence (Gunderson et al, 2012). Lei et al (2012a, 2012b) explored mammary epithelial cell role in BCAA catabolism, and found that BCAT activity to be highly induced during lactation. Given that ~20% of dietary
proteins are composed of BCAA and that milk is a high-protein substance, it may thus be the case that women who lactate have reduced plasma BCAA and reduced T2D incidence rate.

1.2.2.4 Sphingomyelin

In addition to the observed changes in amino acids, metabolomics studies of diabetes have revealed a distinct class of lipids, the sphingomyelins, to be decreased in T2D. Mielke et al (2014) noted a decrease in diabetic and pre-diabetic men and women in an older cohort. Similarly, Floegel et al (2012) also noted a decreased level of SMC16:1 in T2D, which was inversely proportional to insulin secretion and positively correlated to plasma HDL cholesterol. Rui Wang et al (2012) investigated a T2D cohort, and confirmed a decrease in SMC20:2, SMC16:0, SMC16:1 and other species. Both Floegel et al (2012) and Rui Wang et al (2012) conducted a nested-case control study, and hence this decrease was noted up to 7 years before T2D incidence.

Sphingomyelin (SM) plasma concentrations are regulated by the action of two opposing enzymes – SM synthase or sphingomyelinase. Initially, SM are synthesized in two major steps. In the first step, L-serine and palmitoyl-CoA undergo a condensation reaction in the endoplasmic reticulum by serine palmitoyl transferase (SPT). The product is then rapidly reduced and acylated to produce dihydroceramide by the enzyme ceramide synthase. In the next major step, which occurs in the trans Golgi (or at the plasma membrane), a phosphocholine head group is transferred to ceramide, yielding a SM and diacylglycerol as a by-product (Slotte et al, 2013). Conversely, the breakdown of SM by SMase either in the plasma or within the lysosome results in the release of ceramide. The synthesis and breakdown of SM is illustrated in Fig 2.

The cause of decreased level of SMs in T2D currently remains unknown. Theoretically, this may be due to a genetic susceptibility, an inhibition of SM synthase activity, or activation of SMase. With regards to genetic susceptibility, in a metabolomics analysis, carriers of T2D risk allele TCF7L2 had significantly elevated levels of total SM, yet upon a glucose challenge, had a significantly higher clearance rate (Then et al, 2013). The TCF7L2 mutation impairs beta cell insulin secretion, disrupts the incretin effect and enhances hepatic glucose production (Pilgaard et al, 2009). However, as previously discussed, T2D heritability is low and hence this pathway may be of limited importance in T2D pathogenesis. The inhibition of SM synthase activity or
activation of SMase is discussed below in the context of insulin resistance and beta cell dysfunction, the two hallmarks of T2D.

The mechanism by which this decrease in SM may increase insulin resistance remains the subject of much research. Translocation to the plasma membrane and activation of SMase may occur via TNFα, an adipokine involved in systemic inflammation (Clarke et al, 2008). As many T2D are obese, and obesity is regarded a chronic low-grade inflammatory state, the elevated levels of TNFα (Tzanavari et al 2010) in obesity may thus contribute to increased SM breakdown. Furthermore, Samad et al (2006) compared genetically obese mice model to lean controls. The investigators noted decreased SM and increased expression of SMases, SPT and ceramide in the adipose tissue. However, the investigators also reported an increase in both SM and ceramide levels in the plasma of these same mice, and hence the mechanism may not be as direct. In a similar study using diet induced obesity, Shah et al (2008) noted an increase in plasma ceramide, as well as SMases and SPT. Although the authors did not assay SM levels, it may be assumed that increased activity of SMases would result in decreased levels of SM – however, such a decrease may not be observed if SM levels are constantly replenished through diet. In contrast to studies on SM and adipose insulin resistance, few have investigated the effect on skeletal insulin resistance. It is well established that the accumulation of ceramide impairs insulin sensitivity by perturbing the Akt/PKB pathway, thereby inhibiting translocation of GLUT4 into the membrane (Holland et al 2007). The manipulation of SM and the effect on GLUT4 translocation would thus be of great interest in future research.

Furthermore, whether decreased sphingomyelin (or increased SM breakdown) impair beta cell secretion also remains to be explored. A study using SM synthase-null mice associated reduced SM synthesis with increased reactive oxygen species formation, as well as impaired insulin secretion (Yano et al, 2011). Moreover, while it is true that the breakdown of SM results in ceramide production (which induce apoptosis in beta cell), it must be placed within the lipotoxic context of the pancreatic beta cell. Importantly, the accumulation of free fatty acid (lipotoxicity) also promotes ceramide synthesis and apoptosis (Maedler et al, 2001). The SM breakdown may thus work synergistically with the accumulation of free fatty acids for beta cell deterioration (Russo et al, 2013).
Overall, the cause and role of SM in T2D pathology represents a territory that remains to be explored. While the above cited studies revealed potential mechanistic pathway, the investigation of specific SM metabolites remains undetermined.

**Fig 2** Sphingomyelin synthesis and breakdown. Initial condensation reaction in the endoplasmic reticulum by serine palmitoyl transferase (SPT) results in 3-keto sphinganine, which is rapidly reduced by KSA-reductase and acylated by ceramide synthase (CerS) to produce dihydroceramide. In the next major step, which occurs in the trans Golgi or at the plasma membrane, a phosphocholine head group is transferred to a ceramide species, thus producing a sphingomyelin and a diacylglycerol molecule. The breakdown of SM by SMase either in the plasma or within the lysosome results in the release of ceramide. *Adapted from Slotte et al 2013.*
1.2.2.5 Phosphatidylcholine Species

Phosphatidylcholine species (PCs) have been assayed, to our knowledge, in only 3 T2D metabolomic studies (Suhre et al, 2010; Rui Wang et al, 2012; Floegel et al, 2012), with encouraging results. Suhre et al (2010) observed a significant decrease in PC species PCaaC34:3 only. Of 65 PC species assayed, Rui Wang et al (2012) reported a decrease in PCaaC32:1, PCaeC34:2, PCaeC34:3, PCaeC36:2, PCaeC36:3, PCaeC38:2, PCaeC38:3, PCaeC40:3 and PCaeC40:5. No PC species were reported significantly increased. Similarly, of 70 PC species assayed, Floegel et al (2012) observed a decrease in 17 PC species, and an increase in 5 PC species. Specifically, the species PCaaC32:1, PCaeC34:2, PCaeC34:3 PCaeC36:2, PCaeC36:3 and PCaeC40:5 overlapped in the two latter studies, with change in future T2D decreased at baseline.

Chemically, PC species are comprised of a nonpolar phospholipid bound to a polar choline head group. Importantly, although the molecule is overall neutral, there is a dipole. By containing both a polar head and a hydrophobic tail, PC species are the fundamental building blocks of mammalian cell membranes. The phosphatidyl chain is made of a glycerophosphoric acid (glycerol bound to a phosphate group) attached to two free fatty acids (or acyl chains). The linkage between the acyl chain and glycerol backbone may either be an ester, forming a diacylglycerolphosphatidyl choline, or PCaa (Paoletti et al 2011). Alternatively, the linkage between the acyl chain and glycerol backbone may be an ether bond, or PCae. Investigators Guler et al (2009) have reported that ether linkage results in a lower membrane dipole potential. Lastly, the number after the colon denotes the degree of unsaturation found within the two acyl chains. However, the location of these double bonds is not specified within the name.

The majority of PC species identified in T2D metabolomics studies represent ether-linked acyl chains. In physiology, plasmalogen represents the major source of ether lipids (Da Sila et al 2012). Although found in all tissues, plasmalogen PCs are highly enriched in nerve myelin tissue (52%, O’Brien and Samson, 1965) and heart (45%, Warner and Lands, 1961). A deficiency in this molecule is associated with peroxisomal disorders, the site of ether phospholipid metabolism (Wanders et al 2010).
The association of reduced ether phospholipids and type 2 diabetes remains to be investigated. In a recent combined genomic/metabolomic study in healthy human serum, levels of plasmalogen PCae such as PCaeC36:4, PC ae C38:4, and PC ae C40:5, were associated with the gene fatty acid delta-5 desaturase (FADS1), whose product is an enzyme that metabolizes long-chain polyunsaturated fatty acids (Gieger et al, 2008). Specifically, individuals with the SNP allele rs174548 had the lowest levels of these PC species. Of vital interest, the allele rs174548 has also been associated with fasting plasma glucose level in T2D GWAS studies (Dupuis et al, 2010).

1.2.3 Risk Factors for T2D

While metabolomics has provided new leads into risk factors for T2D, traditionally the emerging T2D epidemic has been attributed to an increase sedentary lifestyle and a wanting diet, advanced age, membership of certain ethnicities and family history of T2D. Lastly, the presence of other pathologies, such as hypertension and gestational diabetes, is also associated with greater T2D risk.

1.2.3.1 Physical activity

Physical activity refers to the duration and frequency of activities (typically listed in a questionnaire), weighted by the metabolic equivalent estimate of that activity, or intensity (Diabetes Prevention Program Research Group, 2002). The actual weights and activities listed depend upon the questionnaire being used. Questionnaires have also been developed for specialized subgroups such as pregnant women (Chasan-Taber et al, 2004) and post-partum women (Gunderson et al, 2011).

Several studies have provided high-quality evidence solidifying the increased physical activity as protective of T2D incidence, as well as the corollary, that decreased physical activity is a risk factor for T2D. Tuomilehto et al (2001) found that lifestyle intervention, including ~150 min/week of exercise and diet-induced weight loss decreased T2D incidence in impaired individuals by 58% compared to control (cumulative incidence in the intervention group 11% compared to control group cumulative incidence of 23%). This study however was not design to gauge the effect of exercise alone. Pan et al (1997) stratified impaired individuals into 4 groups –
diet only, exercise only, diet and exercise, and no intervention (control). Exercise alone was associated with a 46% decrease in T2D incidence over 6 year follow up.

1.2.3.2 Diet

In addition to a lack of physical activity, a diet in excess of nutrients or rich in specific compounds has been correlated with increased T2D incidence. Nutrient excess is habitually measured by caloric intake. Usually a self-reported measure, multiple questionnaires exist to assess dietary intake including the Semi-Quantitative Food Frequency Questionnaire by Willet et al (1992) and PrimeScreen by Rifas-Shiman et al (2001).

Reduction in caloric intake is known to decrease the risk of T2D. In the same aforementioned study by Pan et al (1997), the diet-only group were prescribed a diet designed to gradually induce weight loss at 0.5-1.0 kg a month. Six years later, the diet-only intervention resulted in a 31% significantly reduced risk of T2D incidence in impaired patients compared to control. Interestingly, there was even greater risk reduction of 42% in diet and exercise intervention group, although this did not differ from the exercise-only group. In a very recent study by Weiss et al (2015), it appears that for the same loss of weight relative to body weight, exercise combined with caloric restriction induces greater improvement in insulin sensitivity, compared to exercise alone or caloric restriction alone.

Several investigators have also explored which types of diets most increase T2D risk. A meta-analysis by McEvoy et al (2014) highlighted that a diet rich in refined carbohydrates, processed meats, and low in complex carbohydrates, fruits and vegetables, had a 41% increased risk of T2D. Admittedly, the best diet as well as markers of diet quality (such as the glycemic index, dietary fiber and whole grain markers) continue to be the subject of intense research. Presently, the Mediterranean diet has been demonstrated to result in great decrease in risk of T2D in meta-analysis (see Koloverou et al 2014), a diet that was applied to treat T2D as early as the turn of the millennia by Avicenna in his encyclopedia titled The Canon of Medicine (1025).

1.2.3.3 Obesity

The result of low exercise and high caloric intake is inevitably obesity. Several measures for obesity are currently in use, including the body mass index (BMI) and the waist-hip ratio. By far, the most common measure is the BMI, calculated by dividing the mass of the individual over
height, using SI units. However, while BMI highly correlates with percentage body fat (Meuwsen et al, 2010), it is limited as it does not take into account the localization of adipose tissue. This is pertinent as visceral (or abdominal fat) has been independently associated with metabolic syndrome (Carr et al, 2004). In addition, markers of obesity such as the levels of triglycerides (TG) and the ratio of high-density lipoprotein (HDL) to low-density-lipoprotein have also been associated with metabolic risk (Boorsma et al, 2008). Nevertheless, BMI remains a valuable, easily and widely collected measure, and may be calibrated for age, ethnicity or pregnancy status for a defined normal range. The World Health Organization has defined a BMI of greater than 25 kg/m² as overweight, and a BMI of greater than 30 kg/m² as obese.

Using the Third National Health and Nutrition Examination Survey (NHANES III) comprising over 16 000 adults, Must et al (1999) uncovered a much greater prevalence of type 2 diabetes among overweight (BMI > 25 kg/m²) or obese individuals (BMI > 30 kg/m²), across gender and ethnic groups. This association between obesity and T2D was replicated in multiple other cohorts, and progressively worsens with increasing obesity (Nguyen et al, 2008). In Canada, the prevalence of obesity in adults has dramatically increased from 10% in the 1970s to 26% in 2010, with 61 to 74% of T2D cases estimated to be directly associated with this risk factor (Janssen et al, 2013). However, it must be noted that some investigators hold that a subset of obese individuals, also known as “Metabolically Normal but Obese Individuals” (MNOB), are not at higher risk of T2D. Nonetheless, MNOBs represent a low percentage of the obese population, and other investigators do not recognize any protective effect (Kuk and Ardern, 2009).

1.2.3.4 Age

Diabetes risk increases with age. In practice, administrative guidelines generally classify the elderly as an individual over 65 years old, a threshold characterized by a continuous gradual decline in function. Importantly, the symptoms of T2D in the elderly do not present themselves as in the young due to the normal process of aging. For example, the renal threshold for glucose increases as one ages, and thus glycosuria is not seen at similar levels, along with a loss in first-phase insulin release (Chau and Edelman, 2001, reported originally in Meneilly 2000).

The incidence of T2D diabetes increases as one ages, as demonstrated by a national epidemiological study by Tan and MacLean (1995). In fact, over 50% of Europeans with
diabetes are over the age of 60 (Shaw et al, 2010). The CDA recommends screening for T2D in all individuals over 40 years old, as Leiter et al (2001) noted a great prevalence of T2D and impaired status using that threshold. Feig et al (2008) also noted increased T2D in women with advanced maternal age.

### 1.2.3.5 Ethnicity and Family History

Ethnicity, or race, is a self-reported socio-political construct as opposed to being defined biologically. All clinical research funded by the National Institute of Health (NIH) or the FDA must collect ethnicity using standardized categories defined by the respective organizations, and allow participants to select more than one category (U.S. Department of Health and Human Services Food and Drug Administration 2005), with definitions for each race summarized in Table 3.

**Table 3** Categories and definitions of race and ethnicity according to the FDA 2014 guidelines.

<table>
<thead>
<tr>
<th>Race or Ethnicity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>Of Spanish origin or culture, such as Cuba, Mexico, Puerto Rico, South or Central America, regardless of race.</td>
</tr>
<tr>
<td>White</td>
<td>Of European, North African or Middle Eastern origin.</td>
</tr>
<tr>
<td>Asian</td>
<td>Of origin from the Far East, Southeast Asia, or the Indian subcontinent. This includes Pakistan, India, China, Japan and the Philippine Islands.</td>
</tr>
<tr>
<td>Black</td>
<td>Of African origin, whether African American or Haitian.</td>
</tr>
<tr>
<td>Other</td>
<td>All other races, including Native Americans, Native Alaskan and Native Hawaiian.</td>
</tr>
</tbody>
</table>

In general, members of the Hispanic, Asian and Black ethnic/race group have increased risk of T2D prevalence. Carter et al (1996) demonstrated 2 to 6 times greater prevalence of T2D and T2D complications in these groups compared to Whites in the USA. Similarly, Mukerji et al (2012) calculated an age-adjusted prevalence of diabetes increase of 41% and 145% in Canadian Chinese and South-Asians women compared with White women, respectively. Fittingly, both the ADA and CDA list membership to an ethnicity other than White as a risk factor for T2D. Similarly, there is augmented risk of T2D with a family history of T2D.
Evidently, specific ethnicity membership and patients related to individuals with diabetes share both genetic and environmental factors that may explain increased T2D risk (factors such as diet, exercise, body adipose distribution, socioeconomic status, language, culture and access to healthcare). As previously explored (section 1.2.2), GWAS have attributed the heritability of T2D to only 10%. In addition, due to differences in general body frame and fat localization (Araneta et al 2005), lower different BMI thresholds for obesity for certain ethnic groups such as Asians have been implemented recently (ADA Standards of Medical Care in Diabetes 2015).

1.2.3.6 Gestational Diabetes (GDM)

Gestational diabetes, a distinct form of diabetes that is temporarily present during pregnancy, is also a risk factor for T2D. The etiology for this condition is further explored in section 1.3, and the presence of GDM as a risk factor for T2D is investigated in section 1.4.

1.2.4 Risk Categories for T2D

The state of hyperglycemia referred to in the natural history of T2D is a continuum. Although thresholds have been specified to define T2D (see Table 2), elevated levels of both FPG and 2hPG (below the diabetic threshold) may progressively increase until T2D incidence. This state is typically referred to as a pre-diabetic or an impaired state, and was specified as a separate category so as to better identify and follow up on patients at risk of T2D. Large meta-analyses have indicated an annual conversion rate from pre-diabetes to T2D between 5-12% (Gerstain et al, 2007). Prediabetes may refer to either impaired glucose tolerance (IGT), whence the FPG is normal while the post-prandial 2hPG is elevated, or to impaired fasting glucose (IFG), whence the FPG is elevated while the 2hPG levels remain within the normaglycemic range. A prediabetic may also have both IGT and IFG if hyperglycemia is noted in both the fasting and post-prandial state.

The cut-off values for IFG and IGT are summarized in Table 2 by organization. Importantly, the values are similar except that the IFG threshold are lower in the ADA criteria, thereby classifying a greater proportion of individuals as impaired. This was not selected based on the sharp rise in risk of any particular metabolic outcome (Shaw et al, 2000), but was based on observations that the higher FPG threshold predicted a lower proportion of T2D compared to the IGT threshold. Hence, the FPG threshold was lowered so as to represent similar proportion of
the population, and this in turn increased the sensitivity and specificity of IFG (Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 2003). In addition, the gold standard for diagnosis of IGT is through the OGTT, a time-consuming and unpleasant procedure requiring ingestion of an extremely sweet fluid of 75 g sugar, followed by frequent blood sampling. Alternative methods for a proxy to the OGTT from a fasting blood sample have been proposed using metabolomics technology (Cobb et al 2014), with sensitivity and specificity 78% and 72% respectively.

1.2.5 Risk Scores for T2D

In order to better quantify the afore-mentioned risk factors, several risk scores for T2D have been developed. Griffin et al (2000) developed a risk score for patients at risk of undetected diabetes. Using routinely collected clinical variables, various logistical regression models were built optimizing sensitivity and specificity. Variables such as age, BMI, gender, family history of T2D, smoking history and hypertensive medication enabled earlier T2D detection with sensitivity 77% and specificity 72%. Similarly, Rosella et al (2011) built a risk score for a Canadian population to predict 9-year T2D risk with 0.77-0.8 discriminative power. Variables such as BMI, age, ethnicity, hypertension, immigration status, smoking, level of education and cardiovascular disease were important predictors in a sex-specific Weibell survival function. In an independent validation test, Abbasi et al (2012) applied the 25 existing prediction models of T2D to the EPIC-NL cohort. In the basic models containing only routine clinical parameters, the discriminative power varied from 0.73-0.84 for a 7.5 year risk. In contrast, the discriminative power for models including biomarkers ranged from 0.81-0.93, a considerable improvement. Lastly, Savvidou et al (2010) applied a comparable method to develop a risk score for GDM from first trimester variables. Logistic regression models resulted in a discriminative power of 0.861, and included variables such as age, BMI, ethnicity, history of GDM, family history of diabetes, systolic blood pressure, as well as concentrations of t-PA and HDL cholesterol.
1.3 Gestational Diabetes

1.3.1 Definition

In 1998, the Fourth International Workshop-Conference on Gestational Diabetes Mellitus organizing committee defined GDM as “any degree of glucose intolerance with onset or first recognition during pregnancy” (Metzger et Coustan, 1998). Due to such an ambiguous definition, there is currently no international consensus for the diagnostic criteria and diagnostic method of GDM. The International Association of Diabetes and Pregnancy Study Groups (IADPSG), in an attempt to reach universal criteria, published in 2010 a new set of criteria, which was subsequently adopted by the WHO. These new thresholds reflect a greater than 1.75 odds for a pregnancy with birth weight, cord C peptide and percent body fat (markers of a higher risk pregnancy) to be over the 90th percentile (IADPSG 2010). In addition, the two-step method of an initial glucose challenge followed by an OGTT was replaced with an OGTT directly, given the weak performance of the FPG alone (Cypryk et al 2004). Table 4 summarizes the various diagnostic criteria for GDM used presently.

<table>
<thead>
<tr>
<th>Organization</th>
<th>FPG</th>
<th>OGTT</th>
<th>1hPG</th>
<th>2hPG</th>
<th>3hPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO (IADSP)</td>
<td>≥ 5.1</td>
<td>75 g</td>
<td>≥ 10</td>
<td>≥ 8.5</td>
<td></td>
</tr>
<tr>
<td><em>One value sufficient</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDA</td>
<td>≥ 5.3</td>
<td>75 g</td>
<td>≥ 10.6</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td><em>2 values exceeded required</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADA, ACOG</td>
<td>≥ 5.3</td>
<td>100 g</td>
<td>≥ 10</td>
<td>≥ 8.6</td>
<td>≥ 7.8</td>
</tr>
<tr>
<td><em>2 values exceeded required</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDDG</td>
<td>≥ 5.8</td>
<td>100 g</td>
<td>≥ 10.6</td>
<td>9.2</td>
<td>≥ 8.1</td>
</tr>
<tr>
<td><em>2 values exceeded required</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Latest GDM Diagnostic criteria, per organization. Presently, the two-step system has been replaced in favour with a direct OGTT. Note that the WHO criteria is the most inclusive, and may thus inflate future increased GDM prevalence. All glucose values given are in mM.
Furthermore, there remains controversy as to what extent hyperglycemia during pregnancy results in adverse perinatal outcome – some advocating adverse effects only for overt diabetes, whereas others claim continuous increased risk of adverse events with increasing hyperglycemia (Ferrara 2007). GDM is often asymptomatic, yet poses great risk to the fetus and mother. These risks include being large for gestational age, neonatal hypoglycemia, increased cord C peptide and intensive neonatal care, as well as to the mother, including pre-eclampsia, pre-term delivery, shoulder dystocia/birth injury, increased risk of caesarean section and hyperbilirubinemia (IADPSG 2010). In addition, women with pre-GDM or GDM remain at increased risk of congenital anomalies and perinatal mortality compared to nondiabetic women in an Ontarian cohort (Feig et al 2014).

1.3.2 Causes and Correlates

The causes or physiological mechanisms of GDM are not fully defined. During a normal pregnancy, lactogenic hormones (prolactin and human placental lactogen hormone) stimulate beta cell proliferation, whereas counter-regulatory hormones (human placental growth hormone, glucocorticoid cortisol, progesterone and proinflammatory cytokine TNFα) induce IR (Barbour et al, 2007). This IR is multifactorial, and occurs via several mechanisms, such as by 1) inhibiting insulin phosphorylation of the insulin receptor and reducing IRS-1 expression (Vejrazkova et al, 2014); 2) diminishing IRS-1 tyrosine kinase activity (Altinova et al, 2007); and 3) suppressing adiponectin transcription, an endogenous insulin sensitizing hormone (Heitritter et al, 2005). Furthermore, whereas early pregnancy is marked by an increase in fat store, an increase in lipolysis (via PPARγ) is observed in late gestation. This increase in FFA further induces IR in the adipose tissue, liver and skeletal muscle (Barbour et al 2007). In other words, in a physiologic pregnancy, there is a decreased insulin sensitivity compensated by increased beta cell mass and insulin secretion. This adaptation enables more carbohydrates to be allocated for the growing fetus.

In contrast, dysregulation at the level of insulin-responsive cells and insulin secreting beta cells results in a GDM pregnancy. First, in GDM cohorts, it was observed that IR was present even before the index pregnancy (Catalano et al, 2009). In other words, chronic IR before pregnancy, in addition to the physiologic IR during pregnancy, places additional stress in metabolic regulation, resulting in a greater risk of GDM (Metzger et al, 2007). Mechanistically,
lower levels of IRS-1 expression (Barbour et al 2007), IRS1-tyrosine kinase activity (Shao et al 2000), reduced insulin-stimulated GLUT4 translocation (Friedman et al 1999) and reduced plasma adiponectin (Williams et al, 2004) have been recorded for women with GDM, in peripheral tissue compared to normal BMI-matched pregnant controls. The added effect of all these mechanisms (and others) may induce greater IR characteristic of GDM.

Secondly, an inherent defect may also be present in the pancreatic beta cell in women genetically susceptible to GDM, rendering them unable to compensate for the increase IR. Importantly, for the same level of IR, GDM beta cells have lower insulin secretion compared to NGT beta cells (Metzger et al 2007). Genetically, Zhang et al (2013) identified 12 SNPs on 9 genes specific to gestational diabetes. Of these, 6 regulate insulin secretion – transcription factor 7-like 2 (TCF7L2), glucokinase (GCK), potassium inwardly rectifying channel, subfamily J, member 11 (KCNJ11), CDK5 regulatory subunit associated protein 1-like 1 (CDKAL1), insulin-like growth factor 2 mRNA-binding protein 2 (IGF2BP2), and melatonin receptor 1B (MTNR1B). However, the odds ratio for each gene ranged from 1.15 to 1.46, indicated a small effect size. Interestingly, some of these genes have also been identified in T2D GWAS. This is not surprising given the 7 fold risk of developing T2D in women with prior GDM (Bellamy 2009).

Interestingly, many of the T2D genetic variants also associated with GDM in a Danish and Korean cohort (Lauenborg et al 2009, Cho et al 2009). These include CDKALI and TCF7L2 variants, known to impair compensatory insulin secretion in response to increasing insulin resistance

In contrast to T2D, few plasma GDM metabolomics studies have been published, and these report inconsistent findings. For example, while some studies found higher levels of BCAA (Metzger et al 1980; Butte et al 1999) and aromatic acid tyrosine in GDM Hispanic women (Butte et al 1999), more recent and larger studies have reported the contrary (Cetin et al 2005; Pappa et al 2007). Interestingly, Pappa et al (2007) reported an increase in beta-hydroxybutyrate and a decrease in glycine in GDM plasma, although this has not been replicated. Elevation of saturated FFA in GDM women was also reported by Idzior-Walus et al (2008), Chen et al (2010), and Pappa et al (2007), although these observations were not all adjusted for
BMI or caloric intake. Even fewer prospective GDM metabolomic studies are available, and the majority suffer from small sample sizes and inconsistent results (reviewed in Huynh et al 2014).

Interestingly, studies have shown that the sex of the fetus has important implications in the risk of GDM in the mother. By reviewing Canadian data, researchers uncovered an association between a male fetus and increased risk of gestational diabetes in both the first and second pregnancy (Retnakaran and Shah, 2015). Mechanistically, this may be due to the release of debris (such as hormones and metabolites) by the fetus into the mother’s circulation via the placenta. Although the effect is small based on a systematic review by Jaskolka et al (2015), with relative risk of 1.04 or 1.03 in a sensitivity analysis, the effect is significant. The released debris is thus sex-dependent – how this may affect glucose metabolism and subsequent diabetes risk especially in genetically-susceptible individuals remains the subject of future research. Along this vein, Hocher et al (2010) investigated a maternal PPARγ2 polymorphism in combination with fetal sex. Counter to the observed risk with male fetus, the investigators observed greater levels of glycated hemoglobin in mothers carrying a female fetus compared to mothers carrying a male fetus, or mothers without genetic susceptibility. More recently, Xia et al (2014) recorded greater insulin resistance in mothers with a female compared to a male fetus, despite similar levels of glucose. Therefore, the mechanism for the increased diabetes risk with a male fetus remains to be investigated.

1.4 The transition from GDM to T2D

Notably, any degree of glucose intolerance during pregnancy predicts intolerance post-partum (Retnakaran et al, 2008). In fact, the original thresholds for GDM diagnosis developed by O’Sullivan and Mahan (1964) were selected based on which one resulted in the highest risk of future T2D incidence. Importantly, although 90% of GDM cases resolve post-partum (Feig et al 2008), a meta-analysis reported that between 5-20% of these will develop T2D within 2 years post-partum (and up to 70% within 10 years) (Kim et al 2002). More recent studies report a similar rate of conversion (Metzger 2007)

Several organizations have put forward recommendations to address this increased diabetes risk. In 2009, the American College of Obstetricians and Gynecologists' (ACOG) recommended that GDM be screened 6-12 weeks post-partum, with preference for an OGTT as it identifies both IFG and IGT women. The Canadian Diabetes Association lists a past GDM
pregnancy as a risk factor for T2D (CDA, 2013). Presently, the ADA recommends T2D screening 6 weeks post-partum, and annually if this test identifies any degree of impairment (IGT or IFG) or every 3 years if this test is negative (NGT) (ADA 2004).

Despite the evidence that a GDM pregnancy is a robust marker for increased diabetes risk, follow-up OGTT with women in this high risk group remains extremely low, be it a large multi-centre Canadian cohort (18%, Shah et al, 2011), an Australian cohort (27%, Morisson et al, 2009), or across American cohorts (21%, Lawrence et al, 2010. For a review see England et al, 2009). Of concern, in a large cohort of over 700 women with GDM, while 57% of women returned for post-partum screening, those tended to have had less severe GDM (Hunt and Conway, 2008). There is thus a global lack of awareness of the high risk GDM poses, possibly due to the fact that many women with prior GDM cases do not present any additional risk factors. Reasons for low rates also include logistical difficulties and unpleasant side-effects of an OGTT in new mothers (Fachnie et al, 1988), the change in care providers and lack of awareness of primary care physician that the woman had GDM (Rodgers et al, 2014), unclear responsibility between primary physician and obstetrician for ordering screening test (Bentley-Lewis et al, 2008) and lack of knowledge of testing recommendations (Shah et al, 2011). In addition, there is low risk perception of the new mother, fear of receiving bad news (Bennett et al, 2011, Kim et al 2007) and the stress and time-commitment of caring for a newborn (Schaefer-Graf et al, 2009, Razee et al, 2010) compounded with potential return to work (Conway et al, 1999). Walker et al (2007) developed the Risk Perception Survey-Diabetes Mellitus (RPS-DM) as a tool for diabetes educators. This was used by Zera et al (2013) in a small population of women post a GDM pregnancy. The survey results revealed that while 95% of the participants correctly identified GDM as a T2D risk factor, only 61% perceived themselves at high risk. This may thus be a reflection of the optimistic bias, which upholds that ‘bad things only happen to other people’. Improving risk perception is imperative as it translates to more protective behaviour such as increased diabetes screening per year (Lavielle et Wacher, 2014). Personal conversations with endocrinologists and obstetricians who are members of the Diabetes in Pregnancy Study Group, based in Toronto, Ontario, confirmed the need for a more quantitative measure of risk as a more effective means of improving risk perception and addressing low post-partum follow up. It is thus imperative that better be done to identify women most at risk in this cohort in a more quantitative approach.
1.4.1 Risk Factors for the transition from GDM to T2D

Many studies have examined risk factors for this transition to T2D, many of which overlap with the traditional risk factors explored in section 1.2.3. While Kim et al (2002) found a similar rate of conversion regardless of ethnicity (Kim et al, 2002), Mukerji et al (2012) found a difference in the incremental impact of a GDM pregnancy on T2D incidence between ethnicities. Using the Ontario Diabetes database comprising over a million women, incidence of T2D was compared between White, Chinese and South Asian women with and without a GDM pregnancy. A GDM pregnancy conferred a 13 fold T2D risk in White women, but only a 9 fold risk in South Asian and Asian women. Similarly, Xiang et al (2011) noted increased T2D transition in Black women who had GDM compared to White women enrolled in the Kaiser Permanente Southern California health plan, even after adjustment for confounding variables.Nevertheless, as observed by Kim et al, both Mukerji et al and Xiang et al showed a much greater cumulative incidence of T2D in women with an index GDM pregnancy compared to women with a normal pregnancy.

In a Canadian study encompassing over 600 000 women with no pre-existing diabetes, Feig et al (2008) reported a T2D incidence rate of 3.7% for women 9 months post-partum a GDM pregnancy, and 18.9% for women 9 years post-partum, with a higher incidence in women with advanced maternal age, or later delivery. Similarly, Aberg et al (2002) reported maternal age as a risk factor for T2D conversion within 1 year post-partum. However, this study compared women with GDM to women with a healthy pregnancy, and thus does not inform which women with GDM are at most risk. In addition, while some studies report increased risk of progression with additional pregnancies (Peters et al, 1996), this has not been replicated by others (Kim et al 2007). Additional pregnancies may thus be confounded with concurrently increasing age.

The degree of severity of hyperglycemia during a GDM pregnancy has also been reported as a risk factor for T2D progression (Scahefer-Graf et al, 2009; Aberg et al, 2002; Buchanan et al, 1999; Kjos et al, 1995; Metzger et al, 1993; reviewed in Baptiste-Roberts et al 2009). Since the severity of GDM is directly correlated to the method of treatment, gestational age of GDM diagnosis and insulin use during pregnancy is also used as a proxy measure (such as in Baptiste-Roberts et al 2009). Interestingly, Metzger et al (1993) showed that this measure is more predictive of early T2D conversion (within 6 months post-partum) as opposed to 5 years. This is
in agreement with Schaefer et al (2009), who determined T2D incidence 3 months post-partum, and Aberg et al (2002), which only assessed women 1 year post-partum.

Consistently, pre-pregnancy, antenatal and post-partum BMI, associate with T2D progression (reviewed in Baptiste-Roberts et al 2009). Schaefer et al (2009) noted an antenatal BMI > 30 kg/m$^2$ as a marker of increased T2D incidence. Similarly, Metzger et al (1993) found an association with pre-pregnancy BMI and T2D incidence 5 years post-partum. Dalfra et al (2001) also noted BMI of post-GDM NGT women at 1 year post-partum (in a cohort of 70) to be associated with T2D incidence 5 years post-partum. However, this study is limited as a great proportion of women with prior GDM remained impaired post-partum.

Contraception type and use is also highly relevant in a post-GDM population, given that some contraceptives may increase T2D progression. In 1998, Kjos et al reported that while nonhormonal contraception and estrogen-progestin combination oral contraceptive did not increase risk of T2D incidence, the use of progestin-only oral contraceptives increased relative risk by almost 3 folds. Importantly, progestin-only oral contraceptives are usually prescribed to breast-feeding mothers. Yet, breast-feeding has been shown to be protective against future T2D (Gunderson et al, 2012). More recently, Kramer et al (2014) also noted higher odds of GDM incidence in women who used hormonal contraceptives compared to no contraceptives.

In intervention studies, the progression from GDM to T2D was greatly abated by either a lifestyle modification (including weight loss and increased physical activity) or metformin treatment (Diabetes Prevention Research Group 2002). Recently a 10 year follow up was completed in this cohort (Diabetes Prevention Research Group 2015), and confirmed the beneficial effect of lifestyle modification or metformin treatment, with T2D incidence of 10-15% 10 year T2D incidence rate, compared to 50% in the non-intervention group. The greatest decrease was in the lifestyle group, and this is in agreement with the former studies citing BMI as a risk factor.

None of the T2D risk scores previously mentioned (section 1.2.5) considers a GDM pregnancy nor quantifies a specific risk score for the transition from GDM to T2D. While some of the aforementioned T2D metabolomics studies (section 1.2.2) identified biomarker metabolites predictive of T2D, only one very recent study investigated a cohort of women following a GDM pregnancy. Lappas et al (2015) used lipidomics on plasma collected 12-weeks
post the index GDM pregnancy to predict T2D incidence 10 years after delivery in 104 NGT women. Using a model of age, BMI, pregnancy FPG, postnatal FPG, triacylglycerol and total cholesterol in addition to 3 metabolites enabled prediction with discriminative power 0.865, specificity 89.9% and sensitivity 59%. The biomarkers alone resulted in discriminative power 0.845, specificity 92.4% and sensitivity 39.8%. However, this predictive model was not validated in an independent testing set and had very low sensitivity. SM species as a class were also significantly different in future T2D.

Lastly, a small study by Bentley-Lewis et al (2014) investigated the levels of 23 amino acids in a post-partum GDM population with plasma samples taken within 3 years post-partum (sample taken at mean time 16-19 months post-partum). While some amino acid levels decreased following an OGTT in both NGT and impaired women, this response was blunted in impaired women. In contrast, longer duration of breast-feeding was associated with the greatest number of amino acid changes.

1.5 Predictive Modeling

Predictive modeling refers to the use of various statistical methods to calculate the probability of an outcome given a set number of events and their associated attributes. Traditional statistical methods include regression models, whereas more advanced methods are explored under the machine learning section 1.5.1. Regardless of the chosen method, any design of experiment must include considerations for validation and optimization (section 1.5.2). Depending on the field or statistical program, different terminology may be used interchangeably, and this is summarized in Table 5.
An important distinction must be made between prediction research and explanatory research. Whereas explanatory research begins with an a priori hypothesis that ideally test for causality, prediction research does not begin with preconceived specified hypotheses, and takes a discovery-based approach. The attributes that a predictive model highlights may provide insight into the pathophysiology, or may simply be associated with future disease incidence (Waljee et al., 2014). In other words, an association may be predictive but not causal.

### 1.5.1 Machine Learning

With the advances in technology and computing, the modern world is effectively logging new data constantly, at a scale beyond human comprehension. The field of medicine not exempt, the development of genomic, proteomic, and metabolomics studies have resulted in enormous data with incredibly complex patterns. As indicated in section 1.5, traditional regression methods do not work well with high-dimensional data, where the number of attributes is greater than the number of classes to be predicted. Furthermore, the human brain has difficulty untangling complex patterns. New problems require new solutions. Combining epidemiology, biostatistics, artificial intelligence, database technology and computer science, machine learning was thus born.

In machine learning, various algorithms attempt to identify general structures within the data using dimension-reduction techniques, in order to make prediction on new observations.
This is not an exact science, or in other words, prediction is never 100% accurate. The field is typically divided into two major categories – supervised learning and unsupervised learning. In supervised learning, the data set is split into a training set and a testing set. The selected algorithm uses the training set to develop a model for predicting the class of interest. The model then applies the developed model upon the testing set, and outputs the accuracy of the model. In contrast, in unsupervised learning, the designated algorithm looks for patterns and structures within the data of interest. It may note that the data set clearly divides into the two classes, which may roughly map onto the classes of interest. Many algorithms of supervised and unsupervised learning are available, and the unsupervised algorithms may also be combined as a pre-processing step. These are further explored under section 1.5.2.

In the case of supervised learning, the appropriate algorithm is contingent upon the outcome of interest is categorical/qualitative (such as future T2D vs no T2D) or a quantitative value (such as the 2hPG value). Several algorithms in which the outcome of interest is categorical are explored below. Importantly, some of these methods have been a part of the statistical literature before the advancement in wide-scale data mining, and have thus been adapted for machine learning.

1.5.1.1 K Nearest Neighbors

Initially developed in the 1950s, in k-nearest neighbors (k-NN), each data instance is mapped onto a space that is an aggregate of all attributes around it (Cover and Hart 1967). The algorithm decides class for a new unknown instance based the class of k instances nearest to it. In other words, the new instance finds itself in a neighborhood of instances with known classes, and the algorithm gives this new instance the same class as the neighbors. This is also known as instance-based learning as learning is represented by the very instances (Witten et al 2011). Generally, distance between instances is calculated using standard Euclidean distances.

This algorithm can be fine-tuned in several ways to produce better results. For example, because of the reliance on distances between instances, it is frequently recommended that the attributes be normalized so that the data is not skewed by concentrations on different scales. In addition, in the case of noisy data, the number of neighbors that ultimately determine class can be specified. A k value of 1 indicates that the unknown instance class will be based on the class of the closest neighbor. In contrast, a k of 5 indicates that the algorithm will take the majority
class of the 5 closest neighbors. It also follows that as $k$ reaches sample size, the accuracy converges to the baseline probability (Witten et al 2011).

This method is limited in several ways. Importantly, k-NN assumes that all attributes are equally important. If this is not true, weighting of attributes should be considered. More importantly, k-NN may be forbiddingly time-consuming should it store all training instances, as all training instances are used for every test data point each time.

1.5.1.2 Logistic Regression

In a regression analysis, model variables (in this case a metabolite) are used to predict the outcome variable (in this case T2D incidence). In the case where the outcome variable is binary as opposed to being continuous, a logistic regression is used. This is because a dichotomous outcome does not have a normal distribution but a Bernoulli distribution. Since the mathematics of linear equations are much easier to work with, statisticians have developed the logit function, or the natural logarithm of the odds ratio. The logit therefore links any linear combination of the model variables onto the Bernoulli distribution of the dependent variable. Importantly, the odds ratio is linked to probability, a measure more intuitively interpreted. Probability is defined as the likelihood of an outcome of interest (or event) divided by all possible outcomes. In turn, odds are the probability of an event divided by the probability of a non-event. Odds ratio are simply the odds of an event divided by the odds of a second event. The odds ratio in a logistic regression can be interpreted as the change in odds for a 1 unit change in the outcome variable. In other words, the odds ratio does not change with different starting points, but is fundamentally the same for the same interval of change. Evidently, an odds ratio of 1 indicates no significant change.

The output of the logistic regression analysis is an estimated regression equation, whose B coefficients can be thought of as ‘weights’ for each model variable similar to the case in a linear equation. Ultimately, from the probabilities of a logistic regression analysis, we can develop a model of an event occurring. In addition, we can use this model to predict and classify the outcome of a new observation. The advantage of logistic regression is that the results are relatively easily explainable. Importantly, attributes are given different weights in the predictive model, represented as coefficient B in the output.
1.5.1.3 Support Vector Machine

Support vector machines (SVM) can be thought of as the merging of linear modeling and instance-based learning. Similar to k-NN, support vector machine (SVM) maps individual instances onto a space that is an aggregate of all attributes (Platt 1998). In contrast to k-NN however, SVM computes a boundary (hyperplane) that separates the two different classes using the instances on the edge of the class group in the space (also called the support vectors). This boundary, or hyperplane, may be linear, in which case it is also called the maximum margin hyperplane (Witten et al 2011). Some classes are however best separated by curved lines, and this is possible in SVM through the use of kernels. A new instance is then given a class based on which side of the boundary it falls upon in the space.

Similar to k-NN, the mapping of instances onto a space by SVM requires dimension-reduction techniques, along with considerations for distance calculation and normalization. Since SVM prediction of new instances does not rely on individual data instances, SVM tends to be more resilient to over-fitting. Despite the advantages of SVM, this algorithm does not cope well with data that divides into greater than 2 classes. In addition, the output of SVM is not easily explainable in terms of the translation into a clinical setting. The boundary equation may clearly delineate events from non-events, but this represents attributes that are on a hyperplane.

1.5.1.4 Naïve Bayes

Similar to k-NN and SVM, Naïve Bayes uses all attributes to develop a predictive model. However, Naïve Bayes takes a probabilistic approach, by taking the product of the probability of each event given an attribute, followed by normalization (John and Langley 1995). This approach assumes that all attributes are equally important and independent of one another, although the method works even when the assumption is violated (Witten et al 2011). This is because all attributes are selected, and hence is overall a stable learning scheme.

1.5.1.5 Decision Trees

In decision tree methods such as J48, the algorithm uses a top down approach to build a tree where each root (node) represents a variable with a threshold, and instances are split into the appropriate branch (Quinlan 1993). This is done recursively for each branch until all instances in the node have the same class, if defined as such.
The algorithm J48 attempts to build the smallest tree with the purest split at each branch, pure referring to a greatest majority class represented in the instances. Which attribute to select for splitting is determined using information theory (Shannon 1948), which computes the difference between the entropy of distribution before the split with the entropy of distribution after the split.

Ideally, the algorithm will find attributes for which it can recursively splits instances until only one class is represented in a node. Therefore, practically, J48 will continue to split until one class is represented in a node, or until the node only holds less than 3 instances. This of course results in an over fitted model with little generalizability. Statisticians thus developed ‘pruning’ methods, or methods that result in a simplified tree, sometimes with even better accuracy (Witten et al 2011). In one method, the minimum number of instances in a node can be specified – the higher the number, the more representative that attribute is in a generalizable model. Alternatively, one can specify the confidence factor for each node, or selectively prune an interior node only.

The advantage of decision trees is that they are reliable, and easily understandable and implemented. In contrast to k-NN and SVM, only some attributes are selected in decision trees. These represent the defining differences in the classes of interest. Furthermore, the data need not be transformed nor normalized since decision trees convert numerical attributes into ordinal attributes with a defined threshold.

1.5.1.6 Rules

Similar to decision trees, rules algorithms such as PRISM, PART and JRIP model classes based on converting each attribute into an ordinal variable, and using a bottom-up approach, form decision lists (Frank and Witten, 1998). Rules and decision trees often have the same expressive power and produce the same predictions, except that rules are more perspicuous. It is critical that rules are executed in the order given – ie, only if the first rule fails, should the second one be considered and so on. New rules are added until all instances are covered.

1.5.1.7 Ensemble Learning

A recent advancement in machine learning, ensemble learning is a meta technique, in which multiple models are combined into an ensemble, and allowed to ‘vote’ on a decision.
ultimate result is often improved accuracy, although the output is generally complex to interpret. Importantly, the ensemble aspect is not itself the learning scheme algorithm, but rather a way in which the learning scheme algorithm is applied. The ‘vote’ may be done in a variety of ways, including bagging, randomization, boosting and stacking (Witten et al 2011).

With bagging, the training set is formulated multiple times, and a model is built for each one, and ultimately combined (Breiman 1996). For example, if the decision tree J48 is bagged, then this means multiple decision trees are built for each training set, and then the averaged decision tree is tested in the testing set. Bagging therefore tends to produce much more stable outputs as outliers would not be included in all of the training sets.

Similarly, with randomization, instead of selecting the best model given by an algorithm, one of the best models is randomly selected. Also called random forests, several ‘one of the best’ tree models are built to produce a forest, and an averaged decision tree is then used to provide accuracy (Breiman 2001). The number of trees, number of attributes, and depth of trees are all variables that can be fine-tuned. Randomization usually improves accuracy, similar to bagging.

In boosting ensemble learning, new algorithm outputs are built for cases not yet covered by the current outputs (Freund and Schapire 1996). For analogy, the algorithm outputs can be thought of as committee members that complement each other’s expertise. Using trees, this refers to building a forest iteratively, with new trees more accurately predicting class of previous error instances.

Lastly, stacking uses completely different learning schemes, or base learners, to build a predictive model (Wolpert 1992). While the output is much more difficult to interpret, accuracy is generally improved.

### 1.5.2 Best Practices for Predictive Modeling

In predictive modeling, many factors must be considered before, during, and after deploying an algorithm. Predictive modeling strategies must first consider the design of the experiment, whether it is prospective, retrospective or cross-sectional, as well as the method of validation (section 1.5.2.1). Next, the work-flow for predictive modeling involves many elements, as presented in section 1.5.2.2. Lastly, the selection of the ultimate model will depend on several measures of outcomes, explored in section 1.5.2.3.
1.5.2.1 Design of Experiment

Predictive modeling is a technique that is highly adaptive to the design of the experiment. Nevertheless, there are some key concerns that must be considered in the design, so as to best optimize the generalizability of the model.

Clinical trials are generally grouped into three types of design. In case-control or cross-sectional studies, a group free of pathology (controls) and a group with pathology (cases) are selected and compared. In cohort or prospective studies, a large number of individuals free of pathology are selected. These are followed over time for incidence of disease. While much more financially demanding and resource intensive requiring large sample size, the advantage of cohort studies is the establishment of a time-sequence before disease, and the accuracy of exposure data. In the third type of design, the nested case-control study, somewhat represents the best of case-control and cohort studies. In a pre-existing cohort, future cases in this cohort are time-matched to controls, and only these are analyzed for differences. Nested case-control studies can be retrospective (where cases have already been identified) or prospective (where cases are added as they occur) (Ernester 1994). The advantage of this method is that it greatly reduces cost, includes a time-sequence and thereby reinforcing potential causality.

More often than not, predictive modeling is applied on an already existing data set initially collected for a separate study, or as a nested case-control study. The best use of this data set in answering the new research question will highly depend upon careful deliberation. First, as opposed to utilizing all cases in the model, a subset of the data may be selected. The total sample size, when given the option, should be one that is statistically powered to detect differences at a power of 0.8 and significance of <0.05. Secondly, the ratio of cases and controls will vary depending on the design of study. While a ratio of cases: controls of 1:1 is standard, this is generally thought to be insufficient for retrospective studies. This is primarily a concern if the cases and controls do not originate from the same underlying population (Ernester 1994). A second concern in favour of increased number of controls to cases is to reflect the heterogeneity in human populations – a disease with prevalence of 10% in a specified population would ideally be modeled in a cohort with case: control ratio of 1:9. However, when the data has already been matched, there appears to be no added benefit of a ratio beyond 1:1 (Miettinen 1985). This is because matching serves to protect against differential distribution of confounders between cases.
and controls. In fact, it is generally better to add more case: control pairs in the study for increased precision as opposed to including additional controls (Wacholder 1992). Financial constraints may also limit the use of more than one control per case, and in fact, a control may be used as a control for more than one case in different stratifications.

The design of experiment must also include methods of validation of results for reliable evaluation. A model can be internally validated using cross-validation. In this method, the data set is split into a certain number of equal parts (the number of folds). In general, the training and testing set are split in a ratio of training: testing 70:30 or 80:20. The algorithm utilizes \( n-1 \) folds to develop a model, and tests that model against the remaining unused fold. The algorithm does this iteratively until every fold has been excluded once, and one last time against the entire data set. The result is a model with the highest accuracy amongst the different fold iterations.

Alternatively, in external validation, the original data set is split into a training set and a testing set, which are completely independent of one another. The training set is used for predictive modeling, then applied to the test set. This ensures that the predictive modeling is generalizable as opposed to utilizing unique structures of the training set. However, some statisticians recommend partitioning the data into 3 subsets: a training, testing and a validation set. This is because, as the training is optimized to improve the testing set performance, the testing set becomes no longer fully independent. The 3rd set, the validation set, ensures that the model is truly generalizable.

1.5.2.2 Work Flow of Modeling

Predictive modeling is comprised of 6 general steps, as summarized in Fig 3. There is no one universal best algorithm, and so much trial and error is required in selecting and after optimizing a method.

First, the data must be integrated into one major data base. Next, the data must be curated or cleaned. A careful process must be applied so as to avoid common errors, such as misalignment of patient IDs and data. In addition, different programs may assign different symbols for similar meaning. For example, a missing value may be represented as blank, 0, NULL, etc., and thus a uniform symbol should be selected. If the data assays represent different batches, a proper correction variable must be applied to account for this. Curating must also
ensure that the final data set only includes variables with a high degree of confidence. All quality control outcomes must be satisfactory.

In pre-processing the data, several strategies may be deliberated, which will highly depend upon the algorithm deployed in the next step. In attribute selection, only a subset of attributes are provided to the model for predictive modelling. This reduction in attributes for which the algorithm should consider may improve results. Several algorithms exist upon which to select attributes for modeling, including CfsSubsetEval, which selects attributes that are highly correlated to the class but not to each other (Hall 1998). If using the standard t-test, attribute selection means that only metabolites significant in the t-test are available to the algorithm. Lastly, in principle component analysis (PCA), the algorithm sums attributes that relate in an orthogonal space at different weights to produce factors. These factors can be used as new attributes for predictive modeling.

In discretization, numerical attributes are transformed into nominal ones. This technique is implicitly used by tree classifiers in choosing an attribute for a node as well as the threshold against which to split on. The difference is that discretization during pre-processing means that this technique is applied earlier.

Normalization, standardization or both changes the scale and range of numerical values of an attribute, so as to be more reflective of each other. This is critical in algorithms that use distances such as SVM and K nearest neighbors. In normalization, numeric variables are scaled from 0 to 1. The chosen equation varies, each with advantages and drawbacks especially in the effects of outliers. In standardization (or z-transformation), the data is transformed to have a mean and unit variance of 0. Standardization does not bound the data set, and hence would not be as vulnerable to the effect of outliers.
Fig 3 Work Flow for Big Data predictive modeling. Data must first be collected, curated and then pre-processed, the strategy chosen dependent on the deployed algorithm.
Clustering algorithms explore the data set for inherent partitioning in the structure without knowledge of class. This unsupervised learning technique may precede a supervised learning algorithm, if the clusters are selected as new ‘transformed’ attributes. In Kmeans clustering, the number of clusters desired is selected (k), and the algorithm maps new clusters minimizing the total squared distance from local centroids.

Following pre-processing, an algorithm is deployed. Should results be non-satisfactory, alternative pre-processing strategies or different algorithms may be attempted. Ultimately, the reported model performs best in the outcomes of interest.

1.5.2.3 Choosing the Best Model

The best model ultimately to be selected is dependent upon which parameter one wishes to optimize. The first step is to consider a classification plot, which considers actual observations and model predictions, for cases and controls. Fig 4 summarizes some of the measures that may be considered, using the example of T2D incidence. Precision refers to the proportion of correctly predicted T2D compared to the sum of all predicted to be T2D. Although similar, sensitivity is defined as the proportion of correctly identified T2D compared to total observed T2D cases. In contrast, specificity refers to the proportion of nonT2D correctly identified as nonT2D in the model. Accuracy and Youden’s J statistic can be thought of as a measure that combines sensitivity and specificity, giving equal weights to each. While accuracy is the proportion of cases or controls correctly identified by the model, Youden’s J statistic is the summation of sensitivity and specificity subtracted by 1. The F score is sometimes also used in modeling, especially when there is emphasis in developing a model that catches all cases, even with low specificity. In this example, this would mean developing a model where all those who develop T2D are identified, even if many nonT2D are also identified as T2D. In all three measures (accuracy, J statistic and the F score), the closer the value of 1, the better the model performs. A value of 1 indicates neither any false positive nor any false negatives.
With regards to graphical representation, the receiver operating curves (ROC) are the standard for binary classifications. In a ROC, sensitivity is plotted against (1-specificity) for each individual. As illustrated in Fig 5, the area under the curve represents the discriminatory power of the model. The closer the AUC is to 1, the greater discrimination, or c-statistic, of the model. A perfect classification of AUC=1 is represented by a red dot in Fig 5. In contrast, a model that does no better than chance would fall along the dotted line, with AUC=0.5, or a non-significant p-value.

Predictive models may also include additional measures to assess model performances, including overall performance through Brier scores, calibration through the Hosmer-Lemeshow test, and net reclassification improvement. The application of these additional measures depends upon the selected algorithm, as well as the sample size.
**Fig 5** The ROC space. A model along the diagonal indicates no predictive ability better than chance. Perfect classification is presented by the red dot.
2 Research Aims and Hypothesis

2.1 Rationale

According to multiple organizations including the World Health Organization, it is the non-communicable chronic diseases that will pose the greatest public health challenge at the turn of the 21st century. In 2013, the Public Health Agency of Canada (2013) published “Preventing Chronic Disease Strategic Plan 2013-2016”, with a focus on preventing these debilitating conditions. Diabetes, amongst a plethora of other chronic diseases, is increasing at epidemic proportions with both personal, social and economic costs. Therefore, it is imperative that better be done to identify women most at risk in this cohort.

Given that gestational diabetes (GDM) prevalence of 3-14% of pregnant women (Ferrara 2007), and that 20-50% of women develop T2D within 5 years of the index pregnancy (Kim et al 2002), the monitoring of this high risk cohort would improve health outcomes. Currently, the American Diabetes Association recommends T2D screening every 1 to 3 years post a GDM pregnancy, via an OGTT. The Canadian Diabetes Association does not have specific recommendations for this population beyond an OGTT 6-12 weeks post-partum. Nevertheless, post-partum screening of this high-risk group remain sub-optimal, with only 16% (Shah et al 2011) in a Canadian cohort. Reasons for low rates include logistical and unpleasant side effects of an OGTT in new mothers (Fachnie et al 1988), the change in care providers after delivery (Rodgers et al 2014), and faulty low risk perception of the new mother (Bennett et al 2011).

Alternative methods such as a risk score for prediction of T2D transition, are presently not available. Although several risk scores have been developed for T2D (Griffin et al 2000; Rosella et al 2011; Savvidou et al 2010), none consider a GDM diagnosis. There is thus an unmet need to be able to predict T2D in women with a previous GDM pregnancy.

With the growing diabetes epidemic, the ideal approach would be practical at a large-scale, as well as inform pathophysiology risk. Ultimately it is a failure of the β-cell to compensate for increasing insulin resistance that leads to diabetes. Diabetes is a metabolic disorder, and hence a metabolomics approach may yield valuable insight of changes that occur before changes in glucose levels. Metabolites represent the final product interaction between genes and the environment, rendering it more relevant outcome than other -omics. Already,
several metabolomic investigations of T2D have revealed significant differences between T2D and NGT controls (Tan et al 2009; Suhre et al 2010). However, no metabolomics study has been conducted on the transition from GDM to T2D. Whether a similar and predictive metabolomics signature specific to a post-GDM pregnancy is possible remains to be explored.

There is thus an unmet need to be able to predict T2D in women with a previous GDM pregnancy early post-partum, in order to optimize preventative interventions. While some of the aforementioned T2D metabolomics studies (section 1.2.2) identified biomarker metabolites of T2D, none investigated a cohort of women following a GDM pregnancy. Interestingly, as the pathology of GDM and T2D overlaps to a certain extent (section 1.3.2) and that GDM may transition into T2D frequently (section 1.4), it may also be the case that metabolites found significant in T2D may be significant in GDM women, or at least those GDM patients who transition into T2D. As diabetes is fundamentally a progressive metabolic disease, we hypothesized that certain metabolites differentially present at baseline could predict future T2D incidence compared to those who remain nonT2D.

2.2 Aims

Our aims were two-fold. First, we intended to undertake a metabolomics study of women recovering from a GDM pregnancy at baseline using fasting plasma samples. We then further correlated the metabolomics data with future T2D incidence and build predictive models of T2D. Secondly, we correlated baseline metabolomics data with baseline glucose status, and noted whether any predictive metabolite biomarker from the fasting plasma samples associated with impaired status.

2.3 Research Questions

1. What metabolites are significantly different in future T2D compared to those who remain nonT2D?
2. Would combining significantly changed metabolites at baseline better predict T2D development compared to currently used indicators, namely FPG and 2hPG?
3. Can any of the significantly altered metabolite at baseline serve as a proxy for the impaired status?
2.4 Hypotheses

As a study falling under systems epidemiology, this research is more discovery-based as opposed to following from specified hypotheses.

1. I hypothesize that metabolites found significant in previous T2D metabolomics studies, specifically belonging to the carbohydrate, free fatty acid, phosphatidylcholines and sphingomyelin class, will significantly differ in future T2D compared to control for those with a previous GDM pregnancy.

2. I hypothesize that a combination of these metabolites, as opposed to any one individually, will be more predictive of T2D than FPG alone.

3. I hypothesize that some of the metabolites that are predictive of T2D are also predictive of baseline impaired status.
3 Methodology

3.1 The SWIFT Cohort

3.1.1 Study Design

The Study of Women, Infant Feeding, and Type 2 diabetes mellitus after GDM pregnancy (SWIFT) is an NIH prospective cohort within the Kaiser Permanent Northern California integrated health care system (R01 HD050625, Gunderson PI). A total of 1035 women were enrolled as they fulfilled the following inclusion criteria: age 20-45 at delivery, singleton pregnancy delivered >35 weeks, GDM during gestation as diagnosed by ADA criteria, no T2D at 6-9 weeks post-partum as diagnosed by ADA criteria, free of other serious medical condition at baseline, no use of steroid or medication affecting glucose tolerance at baseline, and willing to abide by the lactation regiment of their choice. Patients were excluded if they failed to meet any of the inclusion criteria or failed to provide consent for future contact. For the feasibility of the study, patients also had to be English or Spanish-speaking, not planning to move out of the area, and not planning an additional pregnancy within the next 2 years. Patients further had to consent to fill a monthly questionnaire, a follow up interview at 6 months, and return for 3 visits: at 6-9 weeks post-partum (baseline; V1), 1 year post-partum (V2) and 2 years post-partum (V3). While the funding of this study only extended until 2 years post-partum, many women continued to be tracked within the integrated health system beyond this time frame.

At each visit, trained staff collected various parameters, including FPG and 2hPG through a standard 75 g OGTT for T2D diagnosis, fasting insulin, 2h-post prandial insulin, body size and central adiposity, using calibrated research equipment and validated questionnaires. All extra biospecimen, including plasma samples, were stored at -80C. In additional, anthropometric measures such as dietary intake, caffeine intake, physical activity, sleep questionnaire, depression and clinical risk factors were recorded. This study was approved by the KPNC Institutional Review Board. Please refer to Gunderson et al (2011) for additional detail of methodology and design.
3.1.2 The SWIFT Cohort Design of Experiment

To answer the research questions (section 2.3), or compare the baseline metabolomics parameters of future T2D cases compared to those who remained nonT2D, a prospective nested case-control study was designed. As presented in Fig 6, by 2 years post-partum, 109 women had developed T2D, or 10.8% of the cohort. The majority of T2D cases are therefore identified from the annual screening through an OGTT in person. A small number of cases were identified based on alternative electronic medical records used for clinical diagnoses. No T2D case was identified through self-reporting. Due to the prospective nature of the study, an additional 21 women were identified as T2D at 3 or more years post-partum by June 2015. Therefore, a total of 130 T2D cases were identified in the SWIFT cohort.

![Fig 6 Study design of the SWIFT cohort. A total of 1035 women were enrolled, and followed annually for 2 years, with an annual retention rate of 96% and 84%, respectively. By 2 year post-partum, 109 women (10.8%) had developed T2D. An additional 21 women developed T2D beyond 2 years post-partum. Both GDM and T2D were diagnosed as per ADA criteria.](image)

Patients with T2D were matched to nonT2D based on age, pre-pregnancy BMI and ethnicity, the latter three factors selected as each is associated with increased diabetes risk, as previously discussed. Cases were matched to control as follows for the first 86 pairs: race exact match (white, black, Hispanic or Asian), age ± 2 years, BMI ± 1 kg/m². No suitable control could be selected for the remaining cases, and hence these were matched as follows: race exact match (Hispanic vs nonHispanic, age ± 7 years, BMI ± 8 kg/m²). Of the 130 T2D cases, 8 cases could not be matched due to high pre-pregnancy BMI, and hence a total of 122 T2D were assayed for this study. A ratio of 1:1 between cases: controls was chosen (see section 1.5.2).
Cases within 2 years post-partum were subsequently split into a training and a testing set in a 75:25 ratio. Importantly, training set controls were also time time-matched to mitigate any effect of a control converting to a case. In other words, all controls were nonT2D at the V3 (2 year post-partum visit), even if these were paired with a case of T2D at V2 (1 year post-partum). The remaining 21 cases beyond year 2 were thereafter added to the testing set – these were not time-matched due to the limited number of controls available at this time point. In other words, 80 pairs comprised the trainings set, while the remaining 42 pairs formed the testing set.

3.1.3 Relevant Paperwork

A collaboration was established with Dr. Erica Gunderson at KPNC for this study. For this, several important documents had to be completely, negotiated and successfully approved. All of the relevant documents are listed below:

- Research Ethics Board application and Approval letter
- Material Transfer Data Agreement
- Import Permit
- Biosafety Certification
- TCPS2 Certification

3.2 Metabolomics Assay

3.2.1 Development of a Metabolomic Panel

As developing an individual extraction protocol for each metabolite would require a large plasma volume and be financially prohibitive, we used a combination of methods to assay metabolite concentrations. By optimizing plasma volume with cost of running the sample per assay, we succeeded in assaying a panel of 185 metabolites, divided over 7 subpanels, and requiring a total of only 250 ul of plasma. Broadly, the metabolomics assay was developed using a combination of the p150 AbsoluteIDQ™ plate technology (Biocrates, Austria), which quantifies 163 metabolites, in addition to 18 metabolites previously associated with T2D following a literature review of all published metabolomics studies that could be located (Yi et al, 2008; Tan et al, 2009; Zhang et al, 2009; Zhang et al, 2009; Adams et al, 2009; Fiehn et al, 2010; Mihalik et al, 2010; Suhre et al 2010; Wang et al, 2011; Rui Wang et al, 2012; Ha el al, 2012; Mihalik et al, 2012; Floegel et al, 2012; Kaur, et al 2012; Wang et al 2013; Xu et al, 2013; Menni et al, 2013; Ferrannini et al, 2013; Prentice et al, 2014). These 18 metabolites had been
significant in 2 or more studies, and showed a consistent direction of change across studies (>75%).

The first subpanel consisting of the p150 AbsoluteIDQ™ plate technology (Biocrates, Austria), which quantifies over 150 metabolites consisting of acylcarnitine (AC), phosphatidylcholine (PC), lysophosphatidylcholine (LPC), sphingomyelins (SMs), amino acids (AA) and hexose. This plate technology provided semi-quantitative values for the SM, LPC and PC species, yet absolute concentrations for the hexose, AC and AA species.

The second subpanel entailed 4 special amino acids not covered by the Biocrates plate but which appeared highly promising from the literature review. The first member, 2-aminoadipate, was identified by Wang et al (2011) in a nested case-control study as well as in studies on murine models of impacting beta cell function. Similarly, phenylacetyl glutamine was identified by Suhre et al (2010) as well as by previous laboratory pilot studies in metabolomics as promising. Lastly, as the plate technology could not differentiate between leucine and isoleucine, we developed a protocol for differentiating between the two in this panel.

The third subpanel absolutely quantified 12 free fatty acids identified in the literature as significant in T2D. These included both the saturated, unsaturated, and polyunsaturated medium and long chain free fatty acids: C14:0, C16:0, C16:1n-7, C16:1n-9, C18:0, C18:1n-9, C18:1n-7, C20:1n-9, C20:4n-6, C20:5n-3 (EPA), C22:5n-3 (DPA) and C22:6n-3 (DHA).

CMPF, previously shown to be significantly elevated in both GDM and T2D (Prentice et al 2014), was detected using two methods in two different subpanels. In the fourth subpanel, free CMPF was determined via GC/MS. In contrast, in the 5th subpanel, total CMPF was assayed with a competitive ELISA kit (NovaTein Biologicals, USA).

The penultimate subpanel (6th) assayed beta-hydroxybutyrate, a ketone body found significantly elevated in 3 studies (Xu et al 2013; Fiehn et al 2010; Suhre et al 2010). Lastly, glucose was assayed in the 7th panel. This value would provide the FPG, and so provide a reference value for comparison to the metabolomics signature.

Therefore, the combination of all of these subpanels resulted in a total of 185 metabolites to be assayed on 244 plasma samples, each requiring a minimum of 250 ul.
3.2.2 Protocol for Metabolomic Methods

The first four sub-panels were assayed using GC/MS and LC/MS by the Analytical Facility of Small Bioactive Molecules (Hospital for Sick Children, Canada), whereas the latter three were assayed by the author and/or fellow lab members. The method for each sub-panel is described below.

Metabolites of the first sub-panel were assayed as per the manufacturer instructions for each kit. Using 10 µl of plasma, the p150 assays 15 sphingomyelin species, 15 lysophosphatidylcholine species, 77 phosphatidylcholine species, 41 acylcarnitines, 14 amino acids and the sum of all 6 carbon sugars (hexose). These plates come in formats of 24, 56 and 96 wells and are one time use only. In total, one 24 well plate was purchased for a test run, and four 96 well plates for the sample run. Manually handled samples were run through a 5500 QTRAP mass spectrometer (AB Sciex, Concord, ON) coupled to an Agilent 1290 Infinity Series HPLC (Agilent Technologies Canada Inc., Mississauga, ON), including a binary pump and autosampler and controlled by the Analyst 1.6 software. Metabolite data quantification and concentrations as well as quality assessment was performed with the MetIQ™ software package, an integral part of the Absolute IDQ™ p150 kit. This novel plate technology is further described in Römisch-Margl et al (2011), and has further been shown to be in conformance with the US Department of Health and Human Services FDA guidelines (2001).

The second sub-panel quantitatively measured 4 amino acids by first spiking amino Acid standards (0.05-50 µg/mL Leu and Ile, 0.005-5 µg/mL AAA and PAG) and 10 µl plasma samples with an internal standard mixture of labeled amino acids (5 µg/mL Leu-d10 and Glu-d3, 0.5 µg/mL PAG-d5 in H2O + 0.1% FA). Thereafter, both standards and plasma samples were deproteinized using 600 µL methanol, vortexed, and centrifuged for 10 min at 20 000 × g for and 4°C. Under N2, samples and standards were evaporated until dry, and then derivatized by adding 100 µL 3N HCL in n-butanol at 65°C for 20 min. Samples and standards were once again evaporated until dry under N2 and then reconstituted in 500 µL of 5 mM ammonium formate (pH 3.2) in a ratio of 10/90 water/acetonitrile. Standards and samples were analyzed via LC/MS/MS using an Agilent 1290 HPLC with a Q-Trap 5500 mass spectrometer (AB Sciex) and injected into a Kinetex HILIC column (2.6 µm 100 Å, 50 x 4.6 mm) (Phenomenex), running at a flow rate of 500 µL/min. Using 5 mM ammonium formate (pH 3.2) in 10/90 water/acetonitrile,
internal standards and samples were separated isocratically, and quantification in MRM mode was used to monitor the desired mass transitions, as detailed in Table 6.

<table>
<thead>
<tr>
<th>Metabolite/Internal Standard</th>
<th>Parent Ion (m/z)</th>
<th>Product Ion (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile</td>
<td>188.1</td>
<td>43.0</td>
</tr>
<tr>
<td>Leu</td>
<td>188.1</td>
<td>69.0</td>
</tr>
<tr>
<td>AAA</td>
<td>274.4</td>
<td>172.25</td>
</tr>
<tr>
<td>PAG</td>
<td>321.4</td>
<td>84.1</td>
</tr>
<tr>
<td>Leu-d10</td>
<td>198.2</td>
<td>48.1</td>
</tr>
<tr>
<td>Glu-d3</td>
<td>263.3</td>
<td>189.1</td>
</tr>
<tr>
<td>PAG-d5</td>
<td>326.4</td>
<td>84.1</td>
</tr>
</tbody>
</table>

In the third subpanel, fatty acid standards and 20 ul of plasma samples were spiked with the same amounts of each internal standard. Thereafter, ultra-pure water was added to each sample, before being acidified with 1 M HCl. After an addition of 1 mL of hexane, samples and standards were vortexed, chilled and centrifuged for 10 mi at 2,000 x g. The supernatant was collected into separate test tubes. The lower aqueous phase was re-extracted with 1 mL of hexane, vortexed, chilled and centrifuged. Both hexane phases were then combined and dried under nitrogen gas. An equal amount of 1% pentafluorobenzyl bromide and 1% diisopropylamine (in acetonitrile) was added to reconstitute the fatty acids and convert them to their pentafluorobenzyl esters via incubation for 20 min at room temperature. Standard and samples were dried under N₂ once again, and reconstituted in 200 ul of hexane, then analyzed by injection into the GC-MS system Agilent 7890A/5975C GC-MS system. Excellent separation on the chromatograph for each standard was observed for each free fatty acid. Similarly, chromatograph separation was complete in the plasma samples, except for the free fatty acids oleate and cis-vaccenate, which differ in only the location of the double bond. These two were thus combined to give a total concentration for C18:1.

CMPF was characterized in the fourth subpanel. CMPF-d₅ internal standard was used to spike plasma samples and CMPF standards in a 4% BSA in PBS matrix. 480 µL of ultrapure water was added to dilute the 20 ul of plasma or standard samples. After addition of 80% phosphoric acid, samples were vortexed one-by-one, immediately diluted into 1.5 mL of ethyl
acetate (EtOAc) and then briefly vortexed once more. After the last sample, samples were vortexed more thoroughly on a multi-tube vortexer and thereafter chilled on ice. The aqueous and organic phase was then separated via centrifugation, and upper organic EtOAc layer was dispensed into new test tubes. Another 1.5 ml of EtOAc was added to the aqueous phase for re-extraction, and once again, vortexed, chilled and then centrifuged. The two EtOAc layers were combined and then dried under nitrogen gas. 200 ul of acetonitrile was added to reconstitute the pellet. The sample was analyzed by the Agilent 1200 Series HPLC was coupled to an AB SCIEX API 4000 triple-quadrupole mass spectrometer.

CMPF was additionally characterized using a competitive ELISA kit which detects total CMPF (NovaTein Biosciences, MA, USA), as described in Prentice et al (2014). Initially, plasma sample, with its own endogenous CMPF, is added into a well with pre-coated anti-Human CMPF antibody. Immediately after, CMPF-HRP conjugate is added, and all are incubated for 1 hour. The CMPF-HRP competes with endogenous CMPF for binding to the anti-Human CMPF antibody. The higher the endogenous concentration of CMPF, the lower the amount of CMPF-HRP binding. Afterwards, the plate is washed five times, removing any unbound endogenous CMPF or CMPF-HRP conjugate. The plate is then incubated with a substrate for the HRP enzyme, which cleaves HRP resulting in a blue color proportional to the number of CMPF-HRP conjugate and anti-Human antibody pairs. In other words, a lower amount of CMPF-HRP binding results in a fainter blue colour. The addition of the stop solution converts the blue colour to a stable yellow colour. The plate is read at 450 nm on a PHERAstar FS (BMG LabTech, Ortenberg, Germany), and fitted with a 5 parametric fit curve using the MARS Data Analysis Software (BMG LabTech, Ortenberg, Germany).

Ketone body beta-hydroxybutyrate (β-HB) was assayed using a colorimetric kit, following manufacturer instructions (700190; Cayman Chemicals, USA). First, β-HB is specifically oxidized with NAD+ by the enzyme 3-hydroxybutyrate dehydrogenase to D-3-Hydroxybutyrate, concomitantly forming the NADH cofactor. NADH reacts with the colorimetric detector WST-1 in the presence of diaphorase, resulting in WST-1 formazan dye, which produces a colour best read at wavelength 450 nm. Theoretically, the greater the concentration of endogenous β-HB, the greater the concentration of NADH produced, and subsequently the more intense the colour of the formazan dye. To avoid any matrix effect, samples were not diluted, and absorbance was read at 450 nm on a PHERAstar FS (BMG
LabTech, Ortenberg, Germany), before the 1 hour time limit of the stability of the colorimetric detector. Concentration was calculated by fitting a linear regression of the blank corrected absorbance of the standard duplicates using the MARS Data Analysis Software (BMG LabTech, Ortenberg, Germany).

Similarly, fasting glucose was assayed following manufacturer instructions (10009582; Cayman Chemicals, USA). Samples were homogenously diluted in a 1:5 ratio to ensure values within the dynamic range of the assay, as recommended by the kit. Similarly to the β-HB kit, glucose is first oxidized by glucose-oxidase FAD to produce δ-D-gluconolactone and glucose oxidase-FADH$_2$. Glucose oxidase-FADH$_2$ reacts with O$_2$ to produce H$_2$O$_2$. The latter is catalyzed by horseradish peroxidase to react with 3, 5-dichloro-2-hydroxybenzenesulfonic acid and 4-aminoantipyrine to produce a pink dye, with a more intense colour the greater the levels of endogenous glucose. Absorbance was read at 514 nm within the 1 hr time limit of stability of the enzyme mix on a PHERAstar FS (BMG LabTech, Ortenberg, Germany). A linear regression of the blank-corrected duplicates was used to compute plasma concentrations using the MARS Data Analysis Software (BMG LabTech, Ortenberg, Germany). All subpanels are summarized in Table 7.

<table>
<thead>
<tr>
<th>Subpanel #</th>
<th>Name</th>
<th># of Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Biocrates (Biocrates)</td>
<td>163</td>
</tr>
<tr>
<td>2</td>
<td>Special Amino Acids (AA)</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Free Fatty Acids (FFA)</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>MS CMPF (Free CMPF)</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>ELISA CMPF (Bound CMPF)</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Beta-hydroxybutyrate (BHB)</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Glucose (FPG)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7 Summary of metabolite subpanels
3.2.3 Validation of Metabolomic Methods

Each subpanel was validated by a test run of pooled female plasma internal control (FPIC) via three tests: a stability test, a precision test and an accuracy test. Each test is further elaborated below.

A stability test was designed and performed to test the effect of freeze/thaw cycles. The SWIFT plasma samples used in the analysis had already been freeze/thawed once. Upon receipt, each plasma sample was to be aliquoted for each subpanel, and then frozen once more. In other words, plasma samples would have undergone a minimum of 2 freeze/thaw cycles before analysis, and potentially higher if a sample had failed in an analysis or required further dilution. Duplicates of FPIC freeze/thawed two (FPT2) or three times (FPT3) were run for each panel. The average of FPT2 and FPT3 were compared and an independent t-test was performed to determine for significance. For the FFA panel, due to difficulty in developing a method that assays each of the FFA with high quality, an additional internal control of female plasma (FPAA) from another source was used. In this case, FPIC and FPA were freeze/thawed 1, 3 and 5 times. Of important note, the SWIFT samples run underwent a maximum of 2 cycles for all of the subpanels, except for the FFA panel, which underwent 4 cycles.

The FPIC was also spiked as part of the accuracy test to determine any matrix effect and computer percent recovery for each subpanel. A standard from the lower and/or higher concentration range was spiked to FPIC in a 1:1 volume ratio in duplicates. The detected concentration was divided by the theoretical concentration and multiplied by 100 to give percent recovery. The percent recovery was deemed valid for the multiplex assays if it ranged between 70-130%, and 80-120% for uniplex assays. These ranges are standard in quality control CITE.

Precision was measured by calculating the coefficient of variation (%CV), a measure of the extent of variability from well to well, for each duplicate of the FPIC in the assay. This was critical as it informs whether we could run plasma samples in singlicate or whether we required plasma assayed in duplicates or triplicates for an accurate reading. A cut-off of 15% was selected for uniplex metabolites, and 20% for metabolites assayed in a multiplex assay.
Other quality parameters were also considered. The standard curve $r^2$ was to be $>0.98$ for each subpanel. In addition, accepted values had to be greater than the limit of detection (LOD) and within the dynamic range of the method. Furthermore, a high resolution for each metabolite on the chromatograph must be present so as to accurately calculate concentrations.

![Diagram](image)

**Fig 7** Workflow for SWIFT plasma samples through the various metabolomics subpanels. Upon arrival, SWIFT samples were aliquoted and processed through each subpanel for a concentration read out, quality control. A total of 185 metabolites were assayed.

### 3.2.4 Workflow of SWIFT Plasma Metabolomics

Upon arrival, the SWIFT samples were immediately stored at -80°C. Although the container arrived broken, there remained dry ice in it and the samples had not thawed. Thereafter, samples were thawed in batches of 86, the maximum number of samples allowed per plate, and aliquoted into various tubes for each subpanel, before being subsequently frozen once more. Several aliquots were transported on dry ice to the AFBM (The Hospital for Sick Children, Canada) while others remained in the laboratory for analysis. Each plasma sample was assayed for each subpanel, given sufficient volume. If there was not sufficient volume, the sample was
run through the p150 AbsoluteIDQ™ plate technology only. Refer to Fig 7 for a visual representation of the work flow.

Due to technical limitations, plasma samples were assayed over several batches completed over a period of 6 months. Results for each batch would be analyzed before submission of the next batch of samples, to ensure that the quality is maintained. For each batch in each subpanel, duplicates of the FPIC were included in addition to duplicates of the standard, blanks and any additional technical controls. From the FPIC or FPAA, the %CV and, where possible, %Recovery, were computed to validate batch results. Any metabolite with a value greater than the specified threshold formerly specified was rejected (intra %CV<20 in multiplex, <15 otherwise). The inter-assay CV was also computed and flagged when it exceeded 20%.

The fold change from each average of the FPIC or FPAA as it related to the first batch of SWIFT plasma analysis was also computed for each batch. Any fold change outside of the 0.7-1.3 range was flagged. This fold change also served as the correction factor for all subpanels, except for the CMPF Elisa and fasting plasma glucose kit which used an aggregate standard curve as explained above. In the case of CMPF, the raw absorbance values of the standards from all 4 batches were combined to construct a new combined standard curve using MasterPlex Reader Fit (Hitachi Solutions America, Ltd., San Bruno, CA USA) to fit a 4 parameter fit, with CV<20%. Similarly for the fasting plasma glucose assay, the standard curve from each batch was combined to create an aggregate standard curve, with CV<10%. Afterwards, batch-corrected data was aggregated into one main database for statistical analysis.

### 3.3 Statistical Analysis

Several software were utilized to apply the workflow upon the SWIFT data set. Data was first aggregated on Excel 2013 (Microsoft, Washington, USA). As certain tests assumed a normal distribution, normality was assessed through 3 tests: a z-score within ±2.58 for skewness and kurtosis, a visual judgement of the histogram and a visual inspection of the Normal Q-Q plot for linearity. Data was deemed normal if it passed 2 out of these 3 tests.

To assess the sources of variability of the metabolite concentrations, a principal variance component analysis was executed by loading training set data into R 3.1.3 and thereafter converting data into an ExpressionSet object, using the pvca package from Bioconductor.
Missing values were given the column mean. All variables were binned as follows: maternal ages (20-24, 25-29, 30-34, 35-39, and 40-44 years), pre-pregnancy BMIs (Underweight: < 18.5; Normal: 18.5-25; Overweight: 25-30; Obese class I: 30-35; Obese class II: 35-40; and Obese class III: >40 kg/m²). Ethnicity/race was assessed as both a five level factor (White, Asian, Black, Hispanic, and Other). Ethnicity was also assessed as a two level factor (Hispanic and Non-Hispanic). The correlation heat map was produced using Metaboanalyst 3.0 (The Metabolomics Innovation Center: Canada).

In association analysis, both an independent and a paired t-test were computed using SPSS Statistics version 20 (SPSS Inc IBM: USA). Some tests were also sensitive to outliers, and so a sensitivity analysis when applicable was executed by applying Tukey’s correction (outside of 1.5 times the interquartile range).

Predictive modeling using machine learning algorithms were explored using WEKA (University of Waikato, New Zealand). The logistic regression modeling were done using SPSS Statistics version 20 (SPSS Inc IBM: USA. As alluded to in section 1.5.2.3, various measures of diagnostic test specificity, sensitivity, discriminative power and efficiency were calculated. Significance was set at a threshold of P<0.05. Figures were also built from the SPSS platform.
4 Results

4.1 Quality Control Measures

4.1.1 Quality Control: Test Run

To minimize loss of valuable plasma samples, a test run of each subpanel was carried out using two sources of plasma (Table 8), pooled female plasma from a population that is similar to the subjects in our study. Female plasma from a single volunteer was used as an additional means to validate the free fatty acid method. As previously explained, the free fatty acid subpanel samples underwent additional freeze/thaw cycles, and these are accordingly validated in the testing set. The results of each metabolite is summarized in Appendix 1. Overall, the majority of metabolites passed all quality control thresholds, with CV<15%, percent recovery between 80-120%, and an $r^2$ value for the standard curve >0.95 for the majority of the metabolites. The freeze/thaw test indicated high stability for most metabolites, except for 2-AAA, C20:4n-6 and C20:5n-3, which increased slightly over additional cycles. However, plasma samples within one panel always had the same number of freeze/thaw cycles. Of important note, many of the acylcarnitines had concentrations well below the limit of detection. Any metabolite that failed quality control was not used for modeling. Overall, these results indicated that the method was robust for patient sample analysis.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Pooled Female Plasma Internal control (FPIC)</th>
<th>Female Plasma (FPAA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic</td>
<td>White</td>
</tr>
<tr>
<td>Age years (STDEV)</td>
<td>27.3 (1.5)</td>
<td>25</td>
</tr>
<tr>
<td>BMI (STDEV)</td>
<td>34.3 (1.2)</td>
<td>24.3</td>
</tr>
</tbody>
</table>

Table 8 Internal controls used to validate method and to correct for batch effect
4.1.2 Quality Control: SWIFT Sample Runs

Similarly to the testing run, appropriate quality control samples (duplicates of FPIC or FPAA, with spiking as required) were included for each batch to ensure reliability of concentration measurement. The precision, accuracy, batch effect, and dynamic range for each metabolite for each batch, is provided in Appendix 2. Overall, of the initial 185 metabolites, a total of 111 metabolites were of sufficient quality for inclusion in modeling.

With regards to the first subpanel, the samples were run in 4 batches with 86, 86, 86 and 4 samples respectively. A total of 92 metabolites (out of 163) were of consistent quality for each of the first 3 runs. Of the remaining 71 metabolites, while some may have been within the specified threshold in one batch, they would not be in one of the 3 batches. The decision was thus made to exclude these metabolites, as opposed to input the concentration of these metabolites as missing for that specific batch, since a third of missing values is quite substantial. With regards to the 4th batch, only 1 metabolite was of insufficient quality in this run that was not previously excluded in the first 3 runs. Since this 4th batch only assayed 4 plasma samples, the concentration values for AC10 were simply inputted as missing. This is presented in Appendix 2, Table 1.

With regards to the other subpanels, the %CV and range of fold change between batches is presented in Appendix 2, Table 2. The BHB, special amino acids and CMPF MS presented no issue for any batch for any metabolite. However, the free fatty acid presented two specific issues. First, with regards to the resolution, all metabolites could be appropriately resolved within samples except for the FFAs C18:1n7 (cis-vaccenate) and C18:1n9 (oleate) (See Fig 8 for representative chromatographs). The latter’s peaks overlapped, and hence the data is presented as C18:1, or the summation of C18:1n7 and C18:1n9. Secondly, the %CV for stearate in the FPAA internal control was unacceptably high, despite low CV in other quality control samples (FPIC, as well as spiked FP). A comparison of the FPAA over all batches revealed that the 3rd FPAA in the triplicate sample was definitively an outlier. This was thus discarded, and a %CV and fold change was thus recalculated. Lastly, arachidate appeared below the limit of detection for the majority of samples as well as the control. The decision was thus made to exclude C20:0 from further analysis.
Fig 8 A) Chromatograph of FFA in the standards, showing excellent resolution; B) Zoomed in chromatograph of oleate and vaccinate, the double peak providing able to resolve the two FFA species (21.86 min and 21.91 min); C) Zoomed in chromatograph of a random sample, demonstrating overlap of the two peaks and thus inability to independently characterize concentration of both species (time 21.88 min)
In the case of the CMPF assayed with an ELISA, the %CV of the aggregated standard curve was below 20%. Similarly, in the case of the fasting plasma glucose assay, batch 3a and 3b had much lower concentrations compared to the previous batches in the FPIC aliquot, despite an excellent standard curve within similar absorbance range. As samples could not be re-assayed, the standard curve from each batch was instead combined to create an aggregate standard curve, with CV<10%.

Since we characterized Leu and Ile using two independent methods (through the Biocrates plate technology which provides the summation of Leu and Ile, along with a separate MS method that separately characterizes Leu and Ile), we compared the read out of both methods. The majority of the values fell within 20% of each other, which is within the expected range given that the %CV is less than 10 for each assay. However, that left about a quarter of samples outside of this range. The values were correlated to one another, although the $r^2$ value

**Fig 9** Bland-Altman plot, comparing the summation of Leu and Ile (xLeu) from the Biocrates and Special AA subpanels. The majority of metabolites fell within 2 STDEV, as presented by the red lines. There was a significant decrease of 16 uM in the Special AA subpanel compared to the Biocrates technology. This decrease was not correlated to increasing total concentration, as demonstrated by the regression line (in blue, $r^2 = 0.007$) and its 95% confidence interval.
was only 0.388. A Bland-Altman plot was thus constructed (Fig 9), which is a method used to see if two methods are in fact measuring the same compound. In a scatter plot, mean difference was plotted against average concentration. This plot revealed that there was a highly significant mean difference of 16 uM between the two methods. However, because 95% of the samples fell within 2 SDEV, we could be sure that the methods are in fact measuring the same analyte. Overall, since the individual mass spec method had a higher precision, the individual metabolites Leu and Ile were used for modeling.

Similarly, the Biocrates subpanel provided a concentration of the total of all 6 carbon sugars, or hexoses. Using a colorimetric kit, we had assayed plasma glucose only. Since, theoretically, the level glucose should not be higher than the level of hexoses, we compared the two methods. From a scatterplot, there was a positive correlation between hexose and FPG, as expected, with \( r = 0.366, r^2 = 0.134 \) (Fig 10). Encouragingly, no hexose concentration was greater than 7 mM, the threshold for T2D. However, one plasma sample (220452) recorded an FPG of 8.62 mM, a concentration that exceeds T2D thresholds. In contrast, the hexose concentration was 4.94 mM. This sample was thus considered an outlier and deleted. As mass spectrometry is a technique with a much higher precision than an enzymatic assay, any FPG values minus 2STDEV that remains greater than 2 STDEV from the hexose concentration were discarded as outliers. This resulted in the removal of 1 sample (hexose= 2.67 mM, FPG= 6.58 mM), which physiologically indicated the patient was severely hypoglycemic. Furthermore, a total of 5 other samples (ID 122377, 111897, 220964, 203554 and 120860) had hypoglycemic FPG values (FPG< 2.8 mM), despite normal hexose concentrations, but these were kept as it is still physiologically possible.

**Fig 10** Scatter plot of hexose and glucose values from the two different methods, showing the expected positive correlation.
Lastly, we ensured that concentration readouts were similar between batches, to ensure the removal of batch effect and that the initial batch randomization was truly random. This we did by taking a random subset of metabolites representative of each major class in the Biocrates plate technology, along with each metabolite for all other panels, and comparing box plots for any consistent trends. We did not find any significant differences between batches. An example is presented in Fig 11 for FPG and Ile. Overall, of 185 metabolites, 112 passed all quality control parameters for each batch, or a percentage of 61.

![Fig 11 Box plot demonstrating median and IQR of metabolite concentration across various batches for glucose and isoleucine. Circle (*) indicates an outlier, while a star (⋆) indicates an extreme outlier.](image-url)
4.2 Metabolites predictive of future T2D

4.2.1 Subject Anthropomorphic Measures

Baseline sociodemographic and clinical characteristics of the training and testing set are summarized in Table 9. The patients’ age ranged 21 to 44 years, with an average of 34 years. Although the mean age in the testing set was significantly higher than the training set (training 33.2 years, testing 35.1 years), this is not clinically significant. Furthermore, the proportion of Hispanics, Asians, Whites and Blacks were very similar in both sets, with the majority of pairs of Hispanic background. Cases and controls in both sets did not clinically differ with regards to parity, hypertension history, pre-pregnancy BMI and post-partum BMI. A greater proportion of cases in the testing set had a family history of diabetes however whereas this was not the case in the testing set. Furthermore, lifestyle choices such as smoking history, physical activity, lactation intensity group and total energy intake were similar between cases and controls, with the exception of a decrease in caloric intake in testing set cases. Importantly, lactation intensity group was trending significance (P=0.08). In both sets, cases had significantly elevated levels of parameters typically associated with T2D risk, including elevated FPG, 2hPG, fasting insulin and insulin resistance. HOMA-B was normal between cases and controls in both sets. At two year follow up, controls had a significantly much greater rate of subsequent birth in the testing set. Follow up was similar in the training and testing set, with a greater follow up amongst testing set cases given the inclusion of 21 cases beyond 2 years post-partum.
Table 9  Baseline (6-9 weeks post-partum) and post-partum characteristics of SWIFT patients in the training and testing set (n=122 pairs)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases (n= 80)</td>
<td>Controls (n=80)</td>
</tr>
<tr>
<td><strong>Sociodemographic/Clinical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years),</td>
<td>33.3 (5.2)</td>
<td>33.1 (4.5)</td>
</tr>
<tr>
<td>Race/Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>12 (15)</td>
<td>13 (16)</td>
</tr>
<tr>
<td>Asian, (East, South, Southeast)</td>
<td>26 (33)</td>
<td>26 (33)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>10 (12)</td>
<td>10 (12)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>31 (39)</td>
<td>31 (39)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Parity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primiparous (1 birth)</td>
<td>26 (33)</td>
<td>31 (39)</td>
</tr>
<tr>
<td>Biparous (2 births)</td>
<td>29 (36)</td>
<td>27 (34)</td>
</tr>
<tr>
<td>Multiparous (&gt;2 births)</td>
<td>25 (31)</td>
<td>22 (27)</td>
</tr>
<tr>
<td>GDM prenatal treatment, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet Only</td>
<td>33 (41)</td>
<td>50 (63)</td>
</tr>
<tr>
<td>Oral Medications</td>
<td>38 (48)</td>
<td>28 (35)</td>
</tr>
<tr>
<td>Insulin</td>
<td>9 (11)</td>
<td>2 (2)*</td>
</tr>
<tr>
<td>Pre-pregnancy BMI (kg/m²)</td>
<td>33.5 (8.4)</td>
<td>33.3 (8.3)</td>
</tr>
<tr>
<td>Postpartum 6-9 weeks BMI (kg/m²)</td>
<td>33.5 (7.7)</td>
<td>33.2 (7.8)</td>
</tr>
<tr>
<td>Hypertension history , (%)</td>
<td>19 (24)</td>
<td>16(20)</td>
</tr>
<tr>
<td>Family history of diabetes (%)</td>
<td>45 (56)</td>
<td>42 (53)</td>
</tr>
<tr>
<td><strong>6-9 weeks Postpartum, Lifestyle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoker, n (%)</td>
<td>4 (5)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Physical activity (met-hrs/week)</td>
<td>54.2 (25.1)</td>
<td>47.4 (21.0)</td>
</tr>
<tr>
<td>Total Energy intake (Kcal/day)</td>
<td>805 (338)</td>
<td>811 (319)</td>
</tr>
<tr>
<td>Lactation Intensity Groups, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exclusive lactation</td>
<td>10 (12)</td>
<td>20 (25)</td>
</tr>
<tr>
<td>Mostly lactation</td>
<td>28 (35)</td>
<td>30 (38)</td>
</tr>
<tr>
<td>Mostly formula/Mixed</td>
<td>19 (24)</td>
<td>18 (22)</td>
</tr>
<tr>
<td>Exclusive formula</td>
<td>23 (29)</td>
<td>12 (15)</td>
</tr>
<tr>
<td><strong>6-9 weeks Postpartum, Plasma</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting glucose (mg/dl)</td>
<td>103 (10.5)</td>
<td>95 (8.4)*</td>
</tr>
<tr>
<td>2-hr Post 75 g OGTT glucose, (mg/dl)</td>
<td>132 (29.5)</td>
<td>109 (25.9)*</td>
</tr>
<tr>
<td>Fasting insulin (µU/ml)</td>
<td>33 (17.7)</td>
<td>26 (14.8)*</td>
</tr>
<tr>
<td>Fasting triglycerides (mg/dl)</td>
<td>150 (105.2)</td>
<td>128 (90.7)</td>
</tr>
<tr>
<td>Fasting HDL-C, (mg/dl)</td>
<td>49 (13.0)</td>
<td>49 (13.2)</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>8.6 (5.0)</td>
<td>6.1 (3.7)*</td>
</tr>
<tr>
<td>HOMA-B</td>
<td>305 (156)</td>
<td>299 (183)</td>
</tr>
<tr>
<td><strong>Post-baseline, 2-Year Follow Up</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsequent Birth, n (%)</td>
<td>5 (6)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>Follow up (months), median (IQR)</td>
<td>16.4 (11.6)</td>
<td>22.4 (1.9)*</td>
</tr>
</tbody>
</table>

Values are Mean (SD) unless otherwise noted or n (%). *P<0.05 between cases and controls; ^P<0.05 between training and testing set
4.2.2 Correlations and Covariations

A Pearson r correlation heat map was constructed for all 111 metabolites (Fig 12). The heat map indicates that, generally, members of the same class are highly correlated to each other. Furthermore, the hydrophobic metabolite classes SMs and PCs, tend to positively correlate with each other. While LPCs weakly correlated to various PC species, the former correlate even more minimally to SMs. In addition, classes such as the AC and amino acids tended to only correlate with each other. Within the free fatty acids, the saturated and mono-unsaturated tended to correlate to one another, while the polyunsaturated FFA correlated with each other. CMPF also highly correlated with the polyunsaturated FFAs such as EPA, DHA and DPA, along with specific PC species.

Metabolites were also correlated with pre-pregnancy BMI and age. No metabolite strongly or moderately correlated with either parameter (|r|>0.4). Threonine, isoleucine and free CMPF significantly yet weakly correlated with pre-pregnancy BMI (r = 0.359, 0.306 and -0.302, respectively). Several PC species also correlated with age, yet the r value was <0.3 in all cases.

Fig 12 Two-tailed Pearson correlation r heat-map of all 111 metabolites to each other (n=262).
A PVCA was also constructed inputting all batches, as well as available clinical parameters (Fig 13). The variable ethnicity (binned as Hispanic, Asian, White, Black and other) was identified as responsible for 8.3% of the metabolite concentration variance. This percentage remained high even after ethnicity was binned in only 2 variables (Hispanic vs non-Hispanic, 5.6%). Group membership (T2D or nonT2D), and pre-pregnancy BMI also accounted for 2.5% of proportion variance. Overall, the large residual of 77.9% indicates that other clinical variables are important. Alternatively, they may suggest the high heterogeneity in human plasma samples in this population. Lastly, as inherent in any method in the real world, a part of the 77.9% may also be noise.

Fig 13 PVCA identified Ethnicity, pre-pregnancy BMI and group membership to be responsible for high portions of variance of metabolites.
4.2.3 Distribution of Metabolites

Since some of the subsequent statistical analysis assumes a normal distribution or may be sensitive to outliers, the distribution for each metabolite as well as the presence of outliers was noted for cases and controls separately in the training set. These results are summarized in Appendix 3. Due to the high number of outliers as well as a non-normal distribution, all FFA along with BHB were log transformed.

4.2.4 Association Analyses in Training set

To gauge for any association, an initial independent two-tailed T-test was executed on the training set data, with results in Table 10. A total of 18 metabolites were found to significantly differ between cases and controls. The most significant species were increased hexoses in future T2D (P<0.0001), followed by a decrease in sphingomyelin species SMC20:2 (P<0.001). The branched chain amino acids leucine, valine and isoleucine were the next most significant metabolites (P<0.01). The metabolite 2-AAA along with other amino acids were also significantly elevated, except in the case of glycine, which was decreased in incident T2D. Notably, only the free fatty acid palmitoleate (C16:1n9) was found significantly altered in future T2D. Several long chain phosphatidylcholine species were similarly decreased in future T2D. Interestingly, no LPC species was associated with future T2D incidence in this multi-ethnic cohort.
As we used a matched design, the paired t-test results are also presented in Table 11. All previously identified metabolites were significant, as well as additional SM and PC species. Fasting plasma glucose was also slightly but significantly elevated in future T2D.

Table 10 Individual metabolites and metabolite classes that significantly differ in future T2D in the training set (n=160). Fold change refers to T2D/nonT2D.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Class</th>
<th>Fold Change</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Hexose</td>
<td>Carbs</td>
<td>1.097</td>
</tr>
<tr>
<td>2</td>
<td>SMC20:2</td>
<td>SMs</td>
<td>0.819</td>
</tr>
<tr>
<td>3</td>
<td>Tyr</td>
<td>AAs</td>
<td>1.121</td>
</tr>
<tr>
<td>4</td>
<td>Val</td>
<td>AAs</td>
<td>1.094</td>
</tr>
<tr>
<td>5</td>
<td>SMC18:1</td>
<td>SMs</td>
<td>0.891</td>
</tr>
<tr>
<td>6</td>
<td>Leu</td>
<td>AAs</td>
<td>1.098</td>
</tr>
<tr>
<td>7</td>
<td>2-AAA</td>
<td>AAs</td>
<td>1.20</td>
</tr>
<tr>
<td>8</td>
<td>Ile</td>
<td>AAs</td>
<td>1.095</td>
</tr>
<tr>
<td>9</td>
<td>SMC24:1</td>
<td>SMs</td>
<td>0.913</td>
</tr>
<tr>
<td>10</td>
<td>Trp</td>
<td>AAs</td>
<td>1.057</td>
</tr>
<tr>
<td>11</td>
<td>Thr</td>
<td>AAs</td>
<td>1.097</td>
</tr>
<tr>
<td>12</td>
<td>PCaeC42:5</td>
<td>PCs</td>
<td>0.916</td>
</tr>
<tr>
<td>13</td>
<td>SMC18:0</td>
<td>SMs</td>
<td>0.919</td>
</tr>
<tr>
<td>14</td>
<td>Gly</td>
<td>AAs</td>
<td>0.897</td>
</tr>
<tr>
<td>15</td>
<td>C16:1n9</td>
<td>FFAs</td>
<td>0.890</td>
</tr>
<tr>
<td>16</td>
<td>SM(OH)C16:1</td>
<td>SMs</td>
<td>0.914</td>
</tr>
<tr>
<td>17</td>
<td>SM(OH)C22:2</td>
<td>SMs</td>
<td>0.977</td>
</tr>
<tr>
<td>18</td>
<td>PCaeC40:5</td>
<td>PCs</td>
<td>0.906</td>
</tr>
<tr>
<td>19</td>
<td>PCaeC44:5</td>
<td>PCs</td>
<td>0.922</td>
</tr>
<tr>
<td>20</td>
<td>AC3</td>
<td>ACs</td>
<td>1.104</td>
</tr>
<tr>
<td>21</td>
<td>AC10</td>
<td>ACs</td>
<td>0.907</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metabolite Classes</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SMs</td>
<td>0.939</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>2</td>
<td>BCAA</td>
<td>1.095</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
The assumptions of the paired T-test include a normal distribution and no outlier, which was not met for all metabolites. A paired test is considered robust to deviations from normality especially when the sample size is not small (>50), and hence metabolites were not transformed. In contrast, the paired test is known to be sensitive to outliers, and thus a sensitivity analysis was effected by removing all concentration values outside of 1.5 times the interquartile range by inspection of the boxplot, also known as the Tukey correction. This was done to see whether any metabolites deemed significant were due to outliers. As demonstrated in the last column, the majority of metabolites remained significant or became more significant. The data with the outliers was thus used for all subsequent analyses.

<table>
<thead>
<tr>
<th>#</th>
<th>Metabolite</th>
<th>Fold Change</th>
<th>P value</th>
<th>SA Fold Change</th>
<th>SA P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hexose</td>
<td>1.098</td>
<td>&lt;0.000001</td>
<td>1.102</td>
<td>&lt;0.0000001</td>
</tr>
<tr>
<td>2</td>
<td>SMC20:2</td>
<td>0.819</td>
<td>&lt;0.0001</td>
<td>0.781</td>
<td>&lt;0.000001</td>
</tr>
<tr>
<td>3</td>
<td>Val</td>
<td>1.094</td>
<td>&lt;0.001</td>
<td>1.095</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>Tyr</td>
<td>1.121</td>
<td>&lt;0.001</td>
<td>1.111</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>Leu</td>
<td>1.098</td>
<td>&lt;0.01</td>
<td>1.104</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>SMC18:1</td>
<td>0.891</td>
<td>&lt;0.01</td>
<td>0.875</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>7</td>
<td>Ile</td>
<td>1.095</td>
<td>&lt;0.01</td>
<td>1.081</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>8</td>
<td>Trp</td>
<td>1.056</td>
<td>&lt;0.01</td>
<td>1.044</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>9</td>
<td>SMC24:1</td>
<td>0.913</td>
<td>&lt;0.01</td>
<td>0.913</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>10</td>
<td>2-AAA</td>
<td>1.197</td>
<td>&lt;0.01</td>
<td>1.153</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>11</td>
<td>PCaeC42:5</td>
<td>0.916</td>
<td>&lt;0.05</td>
<td>0.916</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>12</td>
<td>Thr</td>
<td>1.097</td>
<td>&lt;0.05</td>
<td>1.087</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>13</td>
<td>SMC18:0</td>
<td>0.919</td>
<td>&lt;0.05</td>
<td>0.904</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>14</td>
<td>PCaeC40:5</td>
<td>0.906</td>
<td>&lt;0.05</td>
<td>0.899</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>15</td>
<td>Gly</td>
<td>0.897</td>
<td>&lt;0.05</td>
<td>0.923</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>16</td>
<td>AC10</td>
<td>0.907</td>
<td>&lt;0.05</td>
<td>0.880</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>17</td>
<td>C16:1n9</td>
<td>0.888</td>
<td>&lt;0.05</td>
<td>0.880</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>18</td>
<td>FPG</td>
<td>1.047</td>
<td>&lt;0.05</td>
<td>1.044</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>19</td>
<td>SM(OH)C22:2</td>
<td>0.924</td>
<td>&lt;0.05</td>
<td>0.911</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>20</td>
<td>SM(OH)C16:1</td>
<td>0.914</td>
<td>&lt;0.05</td>
<td>0.893</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>21</td>
<td>AC3</td>
<td>1.104</td>
<td>&lt;0.05</td>
<td>1.091</td>
<td>NS (p=0.07)</td>
</tr>
<tr>
<td>22</td>
<td>PCaeC34:1</td>
<td>0.933</td>
<td>&lt;0.05</td>
<td>0.940</td>
<td>NS (p=0.06)</td>
</tr>
<tr>
<td>23</td>
<td>PCaeC44:5</td>
<td>0.922</td>
<td>&lt;0.05</td>
<td>0.908</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>24</td>
<td>PCaeC40:2</td>
<td>0.927</td>
<td>&lt;0.05</td>
<td>0.919</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>25</td>
<td>SMC16:0</td>
<td>0.946</td>
<td>&lt;0.05</td>
<td>0.939</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

Table 11: Paired t-test values of metabolites associated with future T2D incidence, as well as significance following removal of all outliers with the sensitivity analysis P value. Fold change=T2D/nonT2D.
In addition, similar to the independent t-test results, the BCAA as a class were significantly elevated in T2D incident cases (Fold change 1.095, P<0.001). In addition, the SM were decreased in future T2D, with the majority of the absolute mean difference due to a decrease in non-hydroxylated SM.

The data was subsequently stratified by ethnicity as Hispanic vs non-Hispanic, as suggested by the PVCA plot (sample size n=31 and 49, respectively). The top 2 metabolites, hexoses and SMC20:2, were shared in both groups. Furthermore, the BCAA were trending significant and significant in the Hispanic and non-Hispanic group, respectively (P=0.059, P<0.01). Overall, a total of 36 metabolites significantly differed between Hispanic cases and controls, the majority of which were PC species. In contrast, only 11 metabolites reached significance between non-Hispanic cases and controls.

Since the top 4 metabolites were shared between both t-test, a violin plot was constructed to gauge the degree of overlap (Fig 14). While the peak of each distribution was clearly shifted for incident T2D, the distribution of each metabolite indicated much overlap. The endogenous concentration of SMC20:2 in nonT2D (0.416 uM, STDEV 0.120) was comparable to the mean FPIC concentration (0.393 uM), both higher than the T2D value (0.341 uM, STDEV 0.126). In contrast, hexose, valine and tyrosine were each elevated in incident T2D (concentrations 5.2 ± 0.6 mM vs 4.7 ± 0.5 mM, 252 ± 47 uM vs 231 ± 30 uM, 106 ± 26 uM vs 95 ± 18 uM, respectively).

Lastly, to account for multiple testing, the Bonferroni p value was also computed for each metabolite, where 0.05 is divided by the number of metabolites in a given panel. This resulted in a threshold for subpanels Biocrates, Special AA, and FFA of p<0.0003, p<0.01 and p<0.004 respectively. Of the metabolites associated with T2D, hexoses, SMC20:2, valine, leucine, isoleucine and 2-AAA associated at the more stringent Bonferroni threshold.
**Fig 14** Histogram with normal curve fit of the top 4 metabolites. No one metabolite demonstrated a complete separation in the histogram of future T2D compared to those who remain nonT2D. (n=160)
4.2.5 Predictive Modeling

Various statistical algorithms were attempted, as described in section 3.2.4. In this section, the best model will first be presented, followed by important insights derived from other algorithms. Modeling is not dependent upon using the paired design. Instead, as described in section 1.5.2 and 3.1.2, the use of the data varies from algorithm to algorithm.

4.2.5.1 Decision Tree

Following optimization by setting a minimum of 14 instances per leaf and a minimum confidence factor of 0.5, the decision tree algorithm provided the best predictability of T2D incidence in the training and testing set (Fig 15) using only 2 metabolites (PCaeC40:5 and SM(OH)C14:1) and 2 metabolite groups (hexoses and BCAA), as presented in Fig 16. The model first splits cases based on concentration of PCaeC40:5, with those lower than 3.394 uM designated as future T2D, with 31 out of 35 instances correctly classified. Next, the tree splits cases with PCaeC40:5 greater than 3.394 uM based on hexose concentration of 4.934 mM. Cases with hexose<4.934 mM are split based on the sum of BCAA, with those lower than 426.567 uM as future nonT2D (correct classification 52 out of 59 cases, or 88%). Conversely, Instances with BCAA>427.567 were classified as T2D (correct classification 21 out of 31 cases, or 68%). On the other side of the tree, instances with Hexose > 4.934 mM were split based on SM (OH) C14:1, with instances at concentration less than 4.413 uM as future nonT2D (correctly classified 14 out of 18 instances, or 78%) and those with concentration greater than 4.413 uM as future T2D (correctly classified 42 out of 53, or 79% correct).

In the training set, the algorithm sensitivity, specificity, accuracy, precision and F-score reached 86.3%, 68.8%, 77.5%, 73.4% and 79.3% respectively. This resulted in a discriminative power of 83%, and a best model score of 1.623. This was replicated in the independent testing set, with sensitivity 73.8%, specificity 69%, accuracy 72%, precision 71%, F-score 72%, discriminative power 76.9%, and a best model score of 1.49.
Fig 15 ROC of the J48 algorithm on the training set (A) and testing set (B), performing with discriminative power 0.835 (P< 10^{-6}) and 0.755 (P<0.001), respectively. This is much greater than FPG alone (0.584 and 0.618, respectively, NS).
Fig 16 Decision tree by J48 demonstrating the use of PCaeC40:5, hexose, branched chain amino acids (BCAA) and SM(OH)C14:1 in predicting future T2D status. The grey boxes indicate the metabolite chosen for the node, while the clear numbered boxes indicate the concentration threshold used for branching (in uM for all metabolites except for hexose, which is in mM). A blue or green box indicates the group (T2D or nonT2D) associated with the concentration range. The percentage below each group indicates percent of instances correctly classified.
As illustrated in Fig 16, many of the metabolites selected by this algorithm overlap with those identified by the independent t-test in Table 10. A notable exception is the use of SM(OH)C14:1 which was identified to have predictive capability of identifying future T2D cases. The exclusion of this metabolite still results in an excellent decision tree, with sensitivity 72.5%, specificity 61.9%, accuracy 70.2%, precision 67%, F-score 73%, discriminative power 76%, and a best model score of 1.48, which is significantly lower than the aforementioned model. In addition, the summation of the branched chain amino acids (valine, leucine and isoleucine) resulted in a more parsimonious model, and hence the selection in this model.

Training and testing set probabilities of this decision tree were also stratified based on ethnicity (Hispanic vs nonHispanic). The metabolomic signature performed well in both groups with AUC 0.840 and 0.834 in Hispanic and non-Hispanic cases in the training set respectively, and AUC 0.747 and 0.775 in Hispanic and non-Hispanic cases in the testing set, respectively.

4.2.5.2 Other Predictive Models

While the other algorithms did not perform as well as the J48 decision tree, important insights could still be derived from the results of some algorithms. For example, k-NN was applied against the data. Given that the k-NN is highly affected by the scale, data was normalized on a scale from 0 to 1, without regards to class. While the algorithm could correctly classify instances on training set data (AUC 0.994, F score 0.994), this would not translate into the training set (AUC 0.548, F 0.513). Furthermore, a 10 fold cross validation on trainings set data resulted in a much lower AUC of 0.556 and F score 0.503. The greatest source of error was misclassification of T2D cases, an essential outcome for our signature. Various techniques were applied in improving the model, such as by increasing the number of neighbours, with minimal improvement. Similarly, SVM performed only slightly better under a 10 fold cross validation (AUC 0.600 and F score 0.595), but this did not translate into the testing set (AUC 0.536 and F score 0.451), despite normalization of concentrations. Due to the nature of the hyperplane, individual metabolite contributions could not be assessed. In contrast, Naïve Bayes algorithm could be optimized using supervised discretization on the training set up to a discriminative power of 0.725 and F-score of 0.543, for a best model score of 1.268. Similarly, logistic regression models following attribute selection by Cfs subset evaluator, which only selects
independent metabolites, resulted in a high discriminative power of 0.748 but a low F score of 0.529. Some metabolites previously used appeared (such as SMC16:1, SMC20:2, SMC24:1, PCaeC40:5, hexose and BCAA. Interestingly, no FFA were selected for modelling.

4.2.6 Metabolites Associated with Basal Impairment

Given that we found some metabolites predictive of T2D incidence, and that impaired status (IGT or IFG) at baseline is also predictive of future T2D incidence, we next explored whether any of the previously identified metabolites could be associated with impaired status at baseline. Patient characteristics of the training set are summarized in Table 13. There was no difference in NGT and impaired with regards to age or pregnancy BMI. A greater proportion of Asians were impaired at baseline, while a greater proportion of Whites were NGT at baseline. In addition, as expected, baseline FPG was significantly increased in impaired women, and an impaired status was much more associated with T2D incidence at 2 years post-partum.

We next performed an independent t-test (Table 14). A total of 11 of the 24 metabolites were associated with impaired status, with the same direction of change. This included the top metabolites hexose, the BCAA, 2-AAA and the top sphingomyelin SM C20:2. Of the 10 non-significant metabolites, 4 trended towards significance (P<0.1), while the remaining 6 (Trp, FFA C16:1n9, PCaeC40:5, SMC18:1, Thr and Tyr) did not associate with impaired status.

<table>
<thead>
<tr>
<th>Table 13 Baseline (6-8 weeks) characteristics of normal and glucose impaired women in the SWIFT cohort (n=158)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical parameters</strong></td>
</tr>
<tr>
<td>Age, years</td>
</tr>
<tr>
<td>Ethnicity</td>
</tr>
<tr>
<td>White (%)</td>
</tr>
<tr>
<td>Asian (%)</td>
</tr>
<tr>
<td>Black (%)</td>
</tr>
<tr>
<td>Hispanic (%)</td>
</tr>
<tr>
<td>Other (%)</td>
</tr>
<tr>
<td>Pre-pregnancy BMI, kg/m²</td>
</tr>
<tr>
<td>Baseline FPG, mM</td>
</tr>
<tr>
<td>2Y T2D incidence (%)</td>
</tr>
</tbody>
</table>

Values are AVE (STDEV) or n (%). *P<0.05
As suggested by the literature, amino acid metabolites were also correlated with HOMA-IR, a measure of insulin resistance. The branched chain amino acids Ile, Leu and Val had a positive Pearson correlation coefficient of 0.40, 0.32 and 0.29 respectively (p<0.0001 for all). Tyrosine was similarly positively correlated with HOMA-IR (r=0.27, p<0.0001). Furthermore, glycine had a negative r value of -0.33 p<0.0001), with the negative direction in agreement with the decreased levels of glycine seen in future T2D cases. Lastly, 2-AAA was also moderately correlated with insulin resistance (Fig 17).

**Table 14** Association with impaired status of top 24 metabolites t (n=158). Fold change refers to Imp/NGT.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Class</th>
<th>Fold Change</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCaeC44:5</td>
<td>PCs</td>
<td>0.849</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Leu</td>
<td>AAs</td>
<td>1.134</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Ile</td>
<td>AAs</td>
<td>1.127</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Hexose</td>
<td>Carb</td>
<td>1.150</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Glucose</td>
<td>Carb</td>
<td>1.111</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>SM(OH)C14:1</td>
<td>SMs</td>
<td>0.867</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PCaeC44:6</td>
<td>PCs</td>
<td>0.867</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SMC16:1</td>
<td>SMs</td>
<td>0.911</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SMC20:2</td>
<td>SMs</td>
<td>0.850</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>2-AAA</td>
<td>AAs</td>
<td>1.232</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SMC16:0</td>
<td>SMs</td>
<td>0.915</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PCaeC44:4</td>
<td>PCs</td>
<td>0.893</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>AC3</td>
<td>ACs</td>
<td>1.159</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SMC24:1</td>
<td>SMs</td>
<td>0.909</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>SM(OH)C22:2</td>
<td>SMs</td>
<td>0.910</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>PCaaC28:1</td>
<td>PCs</td>
<td>0.905</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>SM(OH)C16:1</td>
<td>SMs</td>
<td>0.903</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>PCaeC40:4</td>
<td>PCs</td>
<td>0.904</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Val</td>
<td>AAs</td>
<td>1.067</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>SM(OH)C24:1</td>
<td>SMs</td>
<td>0.906</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>SMC24:0</td>
<td>SMs</td>
<td>0.923</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>PCaaC42:0</td>
<td>PCs</td>
<td>0.921</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>AC5</td>
<td>ACs</td>
<td>1.099</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>BHB</td>
<td>Ketone</td>
<td>1.298</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>
Fig 17 Baseline amino acid concentrations correlations with baseline HOMA-IR, a measure of insulin resistance. HOMA IR was moderately correlated with Ile (0.40), 2-AAA (0.36) and Gly (-0.33), and weakly correlated with Tyr (0.27). (p<0.0001 for all correlations)
Similarly, as suggested in previous findings, levels of sphingomyelins and PC species were correlated with baseline HDL cholesterol. While total SM levels moderately correlated with baseline HDL levels \( (r=0.47) \), this was not the case for the top metabolite SMC20:2 \( (r=0.25) \). In addition, the top PC species PCaeC40:5 and PCaeC42:5 only weakly yet significantly correlated with HDL levels (Fig 18).

**Fig 18** Baseline SM and PC concentrations correlations with baseline HDL cholesterol levels. HDL was moderately correlated with the sum of all SM \((0.47)\) but only weakly with SMC20:2 \((0.25)\). Furthermore, PCaeC40:5 and PCaeC44:5 were both weakly correlated with HDL \((0.27\) and \(0.29\), respectively). \( p<0.0001\) for all.
Fig 19 Amino acids glycine, isoleucine, 2-AAA and tyrosine across lactation intensity group. Exclusive formula-feeding was associated with decreased glycine, increased isoleucine and increased 2-AAA (P<0.05).
Lastly, previous SWIFT publications have reported the benefits of lactation as a protective effect against future T2D incidence. Given that milk is a high protein product, levels of amino acids associated with T2D were stratified based on the 4 lactation intensity groups: breast feeding only (BF only), mostly breastfeeding (mostly BF), mostly formula feeding (High FF), and exclusively formula feeding (FFonly). Lower glycine levels were significantly associated with the formula feeding only group (Fig 19) compared the exclusive breast feeding group. In contrast, higher levels of BCAA isoleucine as well as the amino acid 2-AAA were significantly associated with formula feeding only. The amino acid tyrosine, as well as valine and leucine, only trended higher across the four groups.
5 Discussion

GDM represents one of the greatest risk factor for the development of T2D and 20-50% of women with GDM develop T2D within 5 years as seen across many studies (Kim et al 2002; Metzger 2007). Risk factors such as ethnicity (Xiang 2011; Mukerji 2012), advanced maternal age (Aberg et al 2002; Feig et al 2008), increased antenatal 2hPG and insulin therapy (Metzger et al 1993; Kjos et al 1995; Buchanan et al 1999) and pre-pregnancy BMI (Dalfra et al 2001; Schaefer et al 2009) are known to associate with future T2D risk. However, no risk score has been developed that can delineate which of the GDM women is most likely to progress to T2D. As 90% of the women with prior GDM return to nonT2D glycemic levels 6-12 weeks post-partum (Feig et al 2008), the ADA recommends OGTT screening tests of everyone in this high risk population every 1 to 3 years.

This recommendation however is rarely applied with low rates of screening across various cohorts (reviewed by England et al 2009). Reasons for such low compliance range from personal obstacles such as care for newborn (Razee et al 2010), returning to work (Conway et al 1999), health system gaps such as unclear responsibility of who should order the screening test (Bentley-Lewis et al 2008) and lack of knowledge of testing recommendations (Shah et al 2011). Furthermore, many of the women with prior GDM along with primary care physicians hold a faulty risk perception of the risk of T2D progression, many times unaware of the 50% conversion rate within only 5 years (Bennett et al 2011; Rodger et al 2014).

As GDM and T2D are ultimately metabolic disorders, a metabolomics approach may provide important insight into the etiology and pathogenesis of this debilitating condition with no present cure. Notably, preventative measures before T2D development are also possible through metabolomics analysis of samples taken before incidence. We thus adapted and validated a metabolomics assay. Next, using the SWIFT cohort, we were able to develop a signature with high specificity, precision and sensitivity, as further explored below.

5.1 Method Validation

In this master’s project, a metabolomics assay of 185 metabolites was successfully systematized and validated using female plasma or pooled female plasma. Our results indicate the high stability of the majority of metabolites over multiple freeze-thaw cycles, if stored at -
84C. Specifically, the majority of free fatty acids were stable in up to 5 freeze-thaw cycles, with the exception of C20:4n-6 and C20:5n-3.

Stringent quality control measures were also applied to each batch of the SWIFT samples. Of the four Biocrates plates over which the SWIFT samples were assayed, the majority of plates had ~35 metabolites out of 163 that presented with high CV or were below LOD, plate 2 had a total of 62 metabolites that failed quality control, or almost double. To avoid a data set with too many missing values, all 62 metabolites, the majority acylcarnitines and phosphatidyl species, were not included in modelling. With regards to the special amino acid, free CMPF and β-HB subpanels, all batches were well within the specified %CV and %Recovery ranges. The free fatty acid panel initially proved more difficult, due to the high likelihood of environmental contamination, however, this too was resolved in time although samples used for concentrations had been freeze/thawed 4 times total. Our validation with the FPIC and FPAA however showed that these metabolites are stable over this number of cycles. Lastly, the bound CMPF ELISA kit and FPG kit were both analyzed through an aggregated standard curve, with % CV <20% and <10% respectively. We could thus be confident in the read out for each of the 111 metabolites used for predictive modeling, with minimal missing values.

5.2 SWIFT Cohort

As one of the largest and most diverse post-GDM cohort followed prospectively from delivery, the SWIFT cohort, led by Dr. Gunderson, is ideal for the development of a risk signature predictive of T2D incidence. Currently, the two year T2D incidence of the entire cohort has been tabulated, with approximately 10% of women developing T2D within this time frame, in agreement with previously published meta-analysis (Kim et al 2002). Continued follow up beyond 2 years post-partum will thus no doubt be able to tabulate additional cases, with so far 21 having been recorded. This may also suggest that present controls used in the study may become T2D in the future. To tackle this issue, we applied three strategies. First, we specified in the research question that we aim to develop a metabolic signature that is predictive of T2D by two years post-partum. In other words, the signature developed from the training set represents answers the research question is "what biomarkers at baseline are predictive of diabetes 2 years post-partum?" This also meant that for each control selected must have been a control for the entire time frame of interest (2 years), even if it was matched to a year 1 case. Secondly, we
decided to add the 21 cases beyond 2 years into the testing set. While we do not know the status of every control beyond the 2 year time frame due to the prospective nature of the study, applying the signature to cases beyond year 2 would give us insight into the applicability of the signature beyond 2 years. Thirdly, when possible we chose controls for which the status beyond 2 years was known. In total, 58% and 54% of controls in the training and testing set, respectively, had a known status nonT2D beyond 2 years, with the majority of these NGT (63% and 64% in training and testing set respectively).

In addition, while constructing the matches for each case-control pair, we had a choice between using time-dependent statistical techniques (such as the cox proportional hazard ratios) or time-independent techniques such as logistic regression, with ‘time’ referring to the time of incidence of T2D. Given the context of the research question and that the SWIFT cohort has only been completely followed up for 2 years, it became clear that of primary concern to a clinician at baseline is not whether the women with GDM will develop T2D at 1 year or 2 year post-partum, but whether she is at higher risk of developing T2D within the time frame of the SWIFT cohort. In addition, since the present ADA recommendations for an OGTT are every 1 to 3 year, the inclusion of cases beyond 2 years ensures the generalizability of the signature within this time frame. Nevertheless, should the SWIFT cohort participants be re-contacted 5 years post-partum, time-dependent analysis would certainly be much more appropriate.

Lastly, an important consideration in predictive modelling is the statistical power of the study. The sample size in this prospective cohort was limited by the number of conversions, and hence a post-hoc analysis need be done. In addition, due to the novel aspect of –omic strategies, many methods have been proposed with no one set as the standard, due to the diversity in classification algorithm and dimension-reducing techniques employed in the field (mostly applied in a genomic setting). Research is still ongoing in this regard, especially due to the highly correlated nature of many metabolic pathways (Hendricks et al 2011). Nevertheless, as the cases were split into a training and testing set, this ensured that the model was at least generalizable to an independent testing set.
5.3 Design of Experiment

5.3.1 Anthropomorphic measures

As illustrated in Table 9, the training and testing set were overall very similar with regards to sociodemographic parameters. Testing set patients were on average 2 years older, a statistically significant different but not clinically significant, given that CDA recommendations for T2D screening is for individuals over 40 years old. The proportions of each ethnicity was similarly distributed, with the greatest representation in Hispanic pairs, followed by Asians. Both pre-pregnancy and post-partum BMI were greater than 30 kg/m², or in the obese range.

Comparing cases and controls, there was no difference in hypertension, smoking history, physical activity and caloric intake between cases and controls, as well as between the training and testing set. Multiparous women were not at greater risk of T2D progression in either the training or testing set, although a significant greater number of controls in the testing set had 3 or more children. Women with prenatal insulin treatment were more likely to be cases in the training set, but only trending in the testing set (P<0.05 and P=0.1 respectively). Nevertheless, in terms of absolute numbers, an approximately equal proportion of women who later progress to T2D were either diet- or orally- treated for GDM, and this did not differ between cases and controls in both sets. As previously discussed in section 1.4, while the original GDM thresholds were chosen to indicate highest risk of T2D progression, not all GDM women develop T2D. Therefore, relying on the women with the most severe prenatal hyperglycemic condition would still miss the majority of T2D cases (over 87% of cases). While cases and controls in the training set did not differ in percentage of family history of diabetes, a significantly much greater proportion of cases in the testing set did. In contrast, cases in the training set were more likely to be impaired, have higher FPG, 2hPG, fasting insulin and higher insulin resistance (HOMA-IR). This may indicate that cases in the training set T2D incidence was more associated with genetic or latent factors (given the high family history).

5.3.2 Correlations

The correlation heat map (Fig 12) visually demonstrated the high degree of correlation between metabolites of the same class. This is expected given the high degree of biochemical pathway interconnectivity or cellular localization between such metabolites.
5.3.3 Metabolite Associated with T2D Incidence

Both the independent and paired t-test resulted in over a dozen metabolites associated with future T2D status. Due to the nature of the t-test, this observation does not mean that these metabolites cause T2D. However, since there is a temporal lag between the altered concentration of these metabolites and future T2D, causality may be the case. Importantly, many of these metabolites, including hexoses, sphingomyelins, amino acids and acylcarnitine C3 have been previously reported to differ in T2D cross-sectional studies. Potential explanations and pathways are further explored below.

5.3.3.1 Hexoses are elevated in future T2D

In our study, hexoses are the most significantly associated with T2D incidence, with an average of 4.7 mM in controls and 5.2 mM in cases. Importantly, hexoses were also reported elevated in 2 other nested case-control studies (Floege et al, 2012; Rui Wang et al, 2012), with up to 5 years before T2D incidence. As introduced in section 1.2.2.1, T2D is defined as the elevation of plasma glucose beyond a set threshold. This increase in hexoses may either be due to 1) a greater proportion of prediabetic women in cases or 2) elevation of other 6 carbon sugars.

The increase in hexoses before T2D incidence may simply reflect an increase in women with impaired status. In our cohort, impaired women have a relative risk of 2.4 to be future T2D cases compared to controls, In other words, 71% of future T2D cases were impaired, compared to 29% of impaired women who remained norm glycemic 2 years post-partum. If we only consider the proportion of women with IFG only, given that the plasma sample taken was analyzed, the relative risk remains at 2.4. Therefore, this increase in hexoses may indeed reflect a greater proportion of IFG women who become T2D.

Alternatively, given the average glucose value of 4.6 in controls and 4.8 mM in cases, with an average glucose difference of only 0.2 mM (compared to an average hexose difference of 0.5 mM), and that the average glucose and hexose value in cases is below the ADA IFG threshold of 5.6 mM (Table 2), it remains possible that other carbohydrates are significantly altered between cases and controls. For example, fructose and mannose were individually assayed, and reported elevated in T2D in 3 multi-ethnic studies (Fiehn et al 2010; Menni et al,
In our study, the different 6 carbon sugars were not individually assayed, and so further study would be required for confirmation.

When stratified by ethnicity/race, this elevation in hexoses was observed across the 3 main groups, an encouraging result since the signature would ideally be applicable to all groups. Overall, our study reports is that women most likely to progress to T2D have an overall elevation of 6 carbon-sugars at least 2 years post-partum.

### 5.3.3.2 Free Fatty Acids are not associated with future T2D

In our nested-case control study, only 1 free fatty acid (FFA C16:1n9) was associated with T2D incidence. Previous metabolomics studies demonstrating elevation in FFA in diabetes were all cross-sectional (Yi et al, 2008; Tan et al 2009; Suhre et al 2010; Ha et al 2012; Xu et al 2013; Menni et al 2013). This overall negative result may be explained by the high obese BMI in the SWIFT cohort cases and controls. In fact, the average pre-pregnancy BMI was 33.6 and 33.3 in cases and controls, and the average post-partum BMI was 33.7 and 33.0 in cases and controls, respectively. Nevertheless, discretization of pre-pregnancy BMI selected at the obesity threshold of 30 kg/m² did not result in any significant association between future T2D status and FFA concentration in either non-obese or obese women, although EPA was trending higher in future cases (P=0.051). Therefore, in a post-partum cohort, FFA levels may not provide much insight into future T2D incidence, despite the known contribution of FFA in inducing insulin resistance and beta cell dysfunction.

With regards to FFA C16:1n9 (palmitoleate), we observed a decrease in future T2D compared to control. Unsaturated free fatty acids such as palmitoleate have been reported to be protective of beta cell function in both in vivo and in vitro experiments (Maedler et al 2001, Welters et al 2004). This observed decrease may thus suggest a beta cell extracellular environment with reduced factors that safeguard function. The cause of this decrease is unknown. In a very recent and large epidemiological study, palmitoleate and oleate consumption was computed to be lower in individuals who later develop T2D (Wang et al 2015). In other words, should this result be due to an unbalanced diet, then intervention is feasibly with a great return of investment.
Interestingly, we observed a decrease in acylcarnitine AC10 (decanoylcarnitine), from 0.223 uM in cases to 0.247 uM in controls. Medium-chain acylcarnitine metabolites are intrinsically linked to mitochondrial fatty acid oxidation, and are observed elevated in fetuses affected by medium-chain acyl-CoA dehydrogenase (MCAD) deficiency (Nada et al 1996). To our knowledge, no other study has reported a decrease in AC10. Interestingly, Mai et al (2013) noted a gradual increase (as opposed to decrease) in AC10 in NGT, Impaired and T2D patients. Similarly, Adams et al (2009) noted a non-significant increase in AC10 in T2D women compared to control. Lastly, AC10 levels were decreased in patients with celiac disease (Bene et al 2005), although with a much greater reduction. Further study would be required to elucidate the link between a decrease in AC10 and future T2D status.

5.3.3.3 Sphingomyelins are decreased in Incident T2D

In our study, the second most significant metabolite associated with future T2D was SMC20:2. Several other SM species were also decreased in future diabetics, including SMC18:1, SMC24:1, SM(OH)C16:1, SMC18:0 and SM(OH)C22:2, with fold change ranging from 0.815 to 0.924. Furthermore, SMs as a class were significantly of lower abundance in T2D cases compared to controls.

This reduction in SM is consistent with T2D metabolomics studies (Mielke et al 2014; Rui Wang et al 2012; Floegel et al 2012), with the latter noting decreased level of SM C16:1 as inversely proportional to insulin secretion and positively correlated to plasma HDL cholesterol. Furthermore, Rui Wang (2012) also reported a decrease in SMC20:2. While further research is required in understanding the mechanism behind this early decrease in SM, it is known that the breakdown of SM results in ceramides, a product known to induce beta cell apoptosis (Maedler et al 2001). Other potential mechanisms were explored in section 1.2.2.4, including potential SM role in exacerbating insulin resistance or beta cell dysfunction. Importantly, research into the role of specific SM species is currently extremely limited due to the limited ability to synthesize with great purity a specific SM.

We correlated SM levels to HDL cholesterol, as such an association has been observed previously (Floegel et al 2012). While total SM were significantly, positively and moderately correlated with HDL (Fig 18, r=0.47), this value dramatically fell to 0.25 for SMC20:2 specifically. This suggests alternate pathways or variables may better explain the decrease in
Physiologically, sphingomyelin breakdown produces ceramides, metabolites known for their role in autophagy, thus suggesting an avenue worth exploring.

5.3.3.4 Amino Acids are altered in Incident T2D

In the SWIFT cohort, the branched chain amino acids valine, leucine and isoleucine, as well as the amino acids tyrosine, tryptophan, glycine and threonine were altered in T2D cases. Furthermore, the amino acid derivative 2-amino adipic acid (2-AAA) was also significantly different in cases. All of the aforementioned amino acids were elevated in future T2D, except for glycine, which was decreased. In addition, acylcarnitine C3, a metabolite linked to BCAA catabolism, was increased, while acylcarnitine C5 (also linked to BCAA catabolism) was only trending (P=0.15).

Branched chain amino acids have consistently been found to be significantly altered in a variety of studies. In T2D metabolomics studies, BCAA were elevated in at least 5 studies comprising both cross-sectional and longitudinal design with up to 12 years before T2D incidence (Floegel et al 2012; Fiehn et al 2010; Xu et al 2013; Menni et al 2013; Ferrannini et al 2013). Although 2 studies reported the opposite trend (Mihalik et al 2012; Zhang et al 2009), this may be explained due to the specific and smaller study populations. In the few GDM metabolomics studies, the results have been conflicting, with a reported increase in 2 studies (Metzger et al 1980; Butte et al 1999) and a decrease in 2 more recent studies (Cetin et al 2005; Pappa et al 2007). Butte et al (1999) specifically assayed insulin treated Hispanic women, and found amino acid levels elevated in GDM compared to healthy pregnant women both ante-partum and post-partum. The investigators also reported overall much higher amino acid levels post-partum compared to ante-partum. In contrast, Pappa et al (2007) only characterized BCAA in diet controlled White GDM women compared to control healthy pregnant women ante-partum only. All 4 studies had a relatively small sample size of ~30 women per group. In the only post-partum study recently published, Anderson et al (2014) reported an increase in the BCAA leucine only in women with a previous GDM pregnancy compared to healthy controls. In our study, we observed an increased concentration of BCAA in those who are at most risk of progressing to T2D. This may thus potentially reconcile between the 4 conflicting GDM studies as follows – women with extreme hyperglycemia ante-partum have both elevated BCAA post-partum and are at most risk of T2D progression, or that this trend is specific to Hispanic women,
coincidentally the group most represented in our study. Given that stratification by ethnicity revealed that various BCAA are associated with future T2D in White, Asian and Hispanic GDM women, the former explanation may provide a better lead for future investigations.

Ultimately, BCAA levels are the summation of the rate of appearance and the rate of disappearance. Up-regulation of processes such as food intake, protein degradation and downregulation of BCAA catabolism would increase the rate of appearance of BCAA levels, whereas up-regulation of processes such as excretion, protein synthesis, and BCAA catabolism would increase the rate of disappearance of BCAA levels. Exactly which processes are most relevant to the reported increase in BCAA remains the subject of intense research. Encouragingly, the significant increase in Acylcarnitine C3, and the trending increase of acylcarnitine C5, may hint at increased in BCAA catabolism, an observation replicated in an animal model by Newgard et al (2012). In a review on the association of BCAA and T2D, Lynch and Adams (2014) reported a correlation between BCAA and insulin resistance. In our study, the BCAA and 2-AAA were correlated with insulin resistance, in agreement with the aforementioned studies.

Conversely, BCAA may simply be associated with T2D incidence, as opposed to being causal. It has been shown that in this same SWIFT cohort, that lactation is protective of future T2D incidence (Gunderson et al, 2012). Lei et al (2012) explored mammary epithelial cell role in BCAA catabolism and noted that, during lactation, BCAT activity is highly induced. This is agreement with other observations, including that 1) ~20% of dietary proteins are composed of BCAA (these AA are the most hydrophobic and confer special functions), 2) milk is a high-protein substance and 3) BCAA catabolism forms glutamate, an AA with many important function. It may thus be the case that women who lactate have reduced plasma BCAA and reduced T2D incidence rate. In our study, lower BCAA and 2-AAA levels were observed in women who breast-fed exclusively. The analyzed plasma samples were taken an average of 6-9 weeks post-partum, and the lactation baseline intensity group strongly predicts duration of lactation. Therefore, it may be the case that in a post-GDM cohort, the association of BCAA and future T2D status may be a result of breastfeeding status. Interestingly, glycine, a metabolite observed lower in formula-feeding patients, is also observed to be negatively associated with future T2D. Lastly, we report for the first time that 2-AAA is associated with a different lactation regiment in a post-GDM cohort, with higher levels associated with future T2D. All of the above
may therefore provide mechanistic clues into the protective effect of lactation in future T2D incidence. In addition, they provide a feasible avenue of intervention.

The observed decrease in glycine is consistent with other studies (Fiehn et al 2010; Floegel et al 2012; Mihalik et al 2012; Rui Wang et al 2012), where glycine was associated with progression to both IGT and T2D. Plasma glycine levels have been reported to be positively associated with insulin sensitivity (Floegel et al 2012). Various pathways have been proposed to account for this observation. At the genetic level, insulin is known to repress expression of ALAS-H, an enzyme that catalyzes the condensation between glycine and succinyl-coA (Phillips and Kushner 2005). Therefore, during hyperinsulinemic stage of T2D progression, increased expression of ALAS-H may further reduce glycine levels. Alternatively, some propose the decrease in glycine to be due to increased glutathione synthesis driven by increased ROS characteristic of T2D natural history (Sekhar et al 2011).

In contrast to the BCAAs and glycine, the metabolite 2-AAA was only recently identified in a European cohort as increased up to 12 years before T2D incidence (Wang et al 2013). In in vivo experiments, this metabolite further decreased FPG in murine models and enhanced glucose-stimulated insulin secretion in beta cell models. To our knowledge, this metabolite has only been characterized once before in a T2D cohort by Fiehn et al (2010). In this cross-sectional study of Black women, there was no significant difference between T2D women compared to controls. In our study, we report that 2-AAA is only elevated in GDM women who later progress to T2D. Therefore, consideration of GDM status of the women by Fiehn et al (2010) may result in significance.

5.3.3.5 Ether-linked Phosphatidylcholine Species are decreased in T2D

In our study, we reported a significant decrease in five PCae species: PCaeC42:5, PCaeC40:5, PCaeC34:1, PCaeC44:5 and PCaeC40:2. Overall, a total of 46 PC species were analyzed, 25 of which were ester-linked and 21 of which were ether-linked. Importantly, all of the PC species associated with future T2D were ether-linked.

The ether-linked PC decrease is consistent with the 3 previous T2D metabolomics studies that assayed PC levels (Suhre et al, 2010; Floegel et al, 2012; Rui Wang et al, 2012). Specifically, PC specie PCaeC40:5 overlapped in these two studies and ours. Critically,
PCaeC40:5 was selected by our modelling algorithm as the best metabolite for the first split in the decision tree.

As alluded to previously, the observation of decreased PC species in general, and PCaeC40:5, up to 12 years before T2D incidence, remains unknown. The association of lower PC species with SNPs of the gene fatty acid delta-5 desaturase (FADS1), may suggest dysregulated fatty acid metabolism at the genetic level (Gieger et al 2008). Notably, we did not characterize the genome of the SWIFT women, and hence such a conclusion may be preliminary. In addition, given that cardiovascular complications are of the most common (and fatal) diabetes complications, and that ether-linked PC species are particularly enriched in heart tissue (Warner and Lands 1961), may suggest a mechanistic link between T2D incidence and heart disease.

Lastly, building on previous association of PC levels and HDL cholesterol (Gieger et al 2008), we correlated PCaeC40:5 and PCaeC44:5 level to HDL. These PC species were significantly positively yet weakly correlated to HDL cholesterol. This suggests alternate pathways may be explanatory in the observed change in PC levels in incident T2D cases (Fig 18).

5.3.4 Metabolites Predictive of T2D

Using a training and testing set design of experiment, we have developed a metabolic signature capable of predicting T2D incidence with discriminative power of over 77%. This signature represents a novel way of stratifying GDM women most at risk of T2D progression. Furthermore, this signature is optimized to detect future T2D cases, as per the use of the F score in choosing the best model. The signature performed similarly to other T2D risk scores, as reviewed by Abbasi et al 2012. However, whereas many of these risk scores were applied for general populations, this signature is specific for women with GDM.

Interestingly, while many metabolites chosen were associated with basal impairment status (IGT or IFG), in both our cohort, the top metabolite PCaeC40:5 was not. Nevertheless, recent developed risk scores have been developed that attempt to bypass the OGTT. For example, Cobbs et al (2014) developed a metabolomics signature as a proxy for OGTT, with sensitivity, specificity, and AUC of 78%, 72% and 0.82. In their study, of 22 metabolites
associated with IGT, the metabolites beta-hydroxybutyrate acid, 2-aminoacidipic acid, glycine, isoleucine, leucine, tyrosine and valine overlapped with our study.

In the J48 decision tree algorithms, metabolites PCaeC40:5, SM(OH)C14:1, leucine, isoleucine, valine and hexoses provided the best prediction for future T2D or nonT2D status. Potential causal pathophysiological mechanisms were explored in section 5.3.3, while Fig 20 illustrates a summary of the observed changes, along with a potential explanatory variable. In the SWIFT cohort, 98% of women with GDM returned to nonT2D glycemic levels at baseline, which is consistent with other studies (Feig et al 2008). Of the 124 women who subsequently develop T2D, a combination of genetic and environmental factors synergistically interact to ultimately induce T2D. In the figure, it is proposed that the decrease in PC levels such as PCaeC40:5 is primarily due to the genetic background of the patients. This is indirectly supported by the greater family history of T2D in the testing set cases, although, importantly, family members also share a similar environment.

With regards to environmental factors, the association of lactation intensity with BCAA levels suggests a behavioural role. However, BCAA levels may also increase with diet – while total caloric intake did not differ in the two groups, specific protein or BCAA calculations was not available for the SWIFT cohort. Nevertheless, it may be assumed that specific food groups such as hexoses may be more highly consumed in future cases.

As the average BMI in this cohort encompasses the obese range, this implies a state of chronic low-grade inflammation (Tzanavari et al 2009), and therefore elevated TNFα levels. TNFα has been mechanistically reported to increase SM breakdown, and may thus be an explanatory variable for the decrease in SM such as SMC20:2 in the plasma of T2D incidence.

Ultimately, all the identified factors induce greater insulin resistance and/or beta cell dysfunction, two critical phases of T2D incidence.
Fig 20 Proposed mechanism combining major metabolomics players in progression to T2D, as expanded upon in the text.
5.4 Data Strengths

The strengths of the afore-presented data may be summarized into 3 general aspects: concentration quality, mechanistic insight, and the application of metabolomics in predicting incidence in a new type of T2D cohort.

First, rigorous quality control parameters ensure that any concentration read out from the metabolomics subpanels was reliable for modeling. As expanded upon in sections 4.1 and appendices 1 and 2, only 111 out of the initial 185 metabolites were selected for modeling. In other words, approximately 40% of metabolites were discarded from the study, the majority of which were acylcarnitines and phosphatidylcholine species. Each metabolite had undergone a stability, precision and accuracy test, with results for each summarized in the Appendices. In addition, the resolution of each metabolite in the chromatograph was diligently scrutinized.

Secondly, the use of the independent and paired t-test revealed many associations previously unknown in this post-partum GDM population. To our knowledge, we are the first study to comprehensively characterize the metabolome of a post-partum GDM multi-ethnic cohort. Importantly, we have reported the elevation of BCAA in women more at risk of T2D progression from a GDM pregnancy, and thereby delineating ‘low risk’ compared to ‘high risk’ GDM women. In addition, we report for the first time an elevation of 2-amino adipic acid in women who progress to T2D, a metabolite that remained obscure until the 2013 publication by Wang et al. Furthermore, we are of the very few studies to assay FFA levels before T2D incidence, and report no association with the outcome of interest in the SWIFT cohort. Lastly, the significant decrease of the SMC20:2 species in women who progress to T2D is novel. These observations may thus serve as insight into the systematic biology in a multi-organ disease such as T2D. As presented in Fig 20, potential pathways of relevance in the progression of T2D are highlighted – these include causes for which intervention is possible (such as a change in diet) or alternatively, for which risk can be better characterized (such as the genetic background of the individual.

Thirdly, we have applied a novel technology, metabolomics, to develop a predictive model of T2D incidence in women with a previous GDM pregnancy. This metabolite signature thus answers a question, and holds the potential to boost preventative measures. Furthermore, by providing a numerical number for risk, this signature potentially provides women afflicted with a
GDM pregnancy a more accurate risk perception. This is critical given the observation that many women with GDM hold a low risk perception (Bennett et al 2011, Kim et al 2007). In addition, this is a group that is easily identifiable, and for which intervention is possible. Critically, intervention in this group benefits not only the mother, but also the newborn. Importantly, current ADA recommendations for this group include an OGTT every 1 to 3 years – this suggests that while the OGTT has good diagnostic value, it is a poor prognostic predictor due to the need for continuous replication. In contrast, our signature holds great prognostic value if able to be replicated in an external cohort.

5.5 Limitations of Study

While the study metabolite signature accurately predicts future T2D within 2 years post-partum, this study is limited in 3 major ways: the number of metabolites assayed, the follow up time/sample size and external validation.

As previously mentioned in section 4.1, about 70 metabolites were discarded due to low quality control in one of the batches. It is entirely possible that some of these metabolites would be significantly associated with future T2D status. In addition, the entire plasma metabolome is thought to comprise about 4000 metabolites (Psychogios et al 2014), a number much greater than the 111 metabolites assayed in this study. For example, individual 6 carbon sugars were not assayed in this study, limited our understanding of the role of carbohydrates other than glucose in metabolic dysregulation. It may be the case that fructose and mannose are specifically elevated years before T2D incidence.

Secondly, the SWIFT cohort, as of the beginning of this master’s project, had only been completed followed up for 2 years post-partum. While our signature accurately predicts which women will transition to T2D within this time frame, it is entirely possible that women selected as controls will progress to T2D at a later date. This is especially relevant given that approximately 50% of GDM women progress to T2D at 5 years (Metzger 2007). Therefore, our signature is potentially more adapted to early T2D converters, as opposed to later. Furthermore, while stratification by ethnicity revealed important insight into shared metabolites, the study was not sufficiently powered for the development of ethnic-specific signatures, if indeed such an approach would result in higher accuracy. However, when we stratified our model based on ethnicity, the signature was stable in both the Hispanic and the non-Hispanic pairs.
Lastly, this signature remains to be validated in a completely external cohort, to test the generalizability of the metabolite signature. Evidently, the formation of such cohorts requires a great undertaking by the primary investigator, and hence, it may be best to approach already formulated post-partum cohorts with a blood bank. Only following this external validation can we conclude with greater firmness that this signature indeed accurately predicts T2D conversion. This is because the use of the training and testing set, while presenting many advantages, can be biased as the training set model is continuously optimized to ultimately result in the best discriminative power and F-score in the testing set. Nevertheless, given that many of our observations have been consistent with other T2D metabolomics studies (albeit with some important novel additions), this signature holds great potential to be generalizable.
5.6 Division of Duties

The author of this thesis’s role was comprehensive, from coordinating the metabolomics assay to establishing a collaboration with the SWIFT Cohort PI Dr. Erica Gunderson. From the metabolomics aspect, the author was responsible for carrying out a literature review, selecting metabolites to be assayed, contacting the AFBM at the Hospital for Sick Children for panel setup, and analyzing all quality control data in assay development. Of note, before the start of this thesis, about two dozen metabolites were identified by other lab members as important in a future metabolomics study, and this the author ensured were included. With regards to the plasma samples, the author dutifully followed up on any legal and ethical paperwork required for the study’s execution. Once samples were received, the author aliquoted all plasma samples and provided them to the AFBM (for the Biocrates, special amino acid, and free fatty acid panel) or performed the assay herself (BHB and glucose). Once results were in, the author performed all of the statistical analysis, with the guidance of committee members Drs. Lisa Strug and Brian Cox. Colleague Katherine Leavey performed the PVCA analysis, for which the author is grateful. The PI Dr. Erica Gunderson provided the correlation of HDL and HOMA-IR with the various metabolites, as well as the average concentration of select metabolite in the different lactation intensity groups.
6 Conclusion

This metabolomics signature is highly predictive of T2D transition in women with previous GDM pregnancy, some of which have not been identified in previous studies. Although developed for T2D transition within 2 years, this signature was able to discriminate with 76% accuracy for cases 2 or more year post-partum. Furthermore, this signature was predictive regardless of ethnicity, with various ethnicities sharing similar metabolites significantly altered. Interestingly, many of metabolites found significant were correlated with impaired status at baseline. This represents the first metabolomics study for the transition from GDM to T2D.

Our initial hypotheses consisted of three parts. In the first part, we hypothesized that the metabolome of GDM women who progress to T2D is different from those who do not. This we were able to prove in section 4.2.4, where about a score of metabolites associated with T2D. These metabolites included carbohydrates, PCs, SMs and one free fatty acid. In the second part, we hypothesized that a combination of metabolites as opposed to only 1 metabolite is more predictive of future T2D. This hypothesis was supported in section 4.2.5, where predictive modeling included more than 1 metabolite to improve prediction. Lastly, as we hypothesized in the 3rd part, some metabolites such as hexoses and BCAA were similarly associated with baseline impaired status.

The implications of this study are that women who progress to T2D years after a GDM pregnancy may be metabolically different as early as 6-9 weeks post-partum. This signature may thus serve as a proxy, and bypass the ADA recommendation of an OGTT every 1 to 3 years, following external validation. Furthermore, this signature provides a quantitative measure of risk, a much needed tool in addressing the low screening rates post-partum, as well as boosting preventative measures in the women most at risk of T2D progression.
7 Future Directions

This study represents the first comprehensive metabolomics study of a post-GDM cohort with a signature that predicts T2D 2 or more years post-partum. The signature was also able to delineate 19 T2D cases beyond this time frame. However, this small sample size limits the generalizability of the model. The future directions suggested will thus broadly cover 3 aspects of the signature, all of which ultimately address the limitations of the present study (as explored in section 5.6). The first direction suggested focuses on expanding the generalizability of this signature, either over time or in alternate cohorts, to determine the clinical utility of the signature. The second direction involves in depth exploration of the mechanisms that the identified metabolites may participate in leading to T2D onset. The third direction will focus on developing more cost-efficient way of characterizing of these metabolites.

7.1 Expanding Signature Generalizability

The above signature was efficient in discriminating future T2D cases at almost 80% in the training set and the independent testing set of cases within 2 years post-partum. An important consideration as alluded to is that progression to T2D in women with GDM will continue beyond this time frame, with 20-50% developing T2D within 5 years (Kim et al 2002), with the fastest rate of increase within 1 year post-partum (Metzger 2007), and possibly a plateau beyond 10 years 10 years of 50 to 70% T2D incidence (Diabetes Prevention Research Group 2015; Kim et al 2002). Therefore, the SWIFT cohort should be followed up annually for the next 8 years, at which each visit an OGTT and other relevant criteria previously mentioned should be collected.

To ensure a great rate of return and minimize loss of follow up, incentives for study participation should be provided. Furthermore, each of the remaining 800 samples should undergo the metabolomics assay of either the top 21 metabolites or all 185 metabolites. At the 5 and 10 year time point, the entire metabolomics data should be revisited, and relevant changes to the status of controls over time should be noted. Based on the current prevalence statistic, of the 1010 plasma samples, between 200-500 will become incident T2D cases by year 5, and 500-700 by year 10 post-partum. As matching cases and controls based on age, ethnicity and pre-pregnancy BMI may not be possible given the almost 1:1 proportion of future cases and controls, all samples in one group should be analyzed, with post-hoc adjustment of all potential confounders. Multiple analyses should subsequently be carried out, and outcomes such as yearly rate of incidence of
T2D, and relative risk of T2D given a combination of the top 21 metabolites identified in this study. The larger sample size would also provide more power in the conclusions from results stratified by ethnicity, year of T2D incidence, age, pre-pregnancy BMI and others.

Furthermore, the top 21 metabolites or all 185 should be characterized in completely independent post-GDM cohorts with a plasma sample collection at baseline to ascertain generalizability beyond patients of Northern California. Several such prospective cohorts presently exist, including the Nurse’s Health Study which currently cites 867 GDM cases, although it is not clear whether plasma is available (Bao et al 2014). Alternatively, the Danish National Birth Cohort followed 60,000 women between 1991 and 2007 during and after pregnancy (Olsen et al 2001). Closer to home, the ADAPT-M (NCT01918345) study led by Dr. Lipscombe et al (2014) is currently recruiting post-GDM women and includes baseline plasma collection. Lastly, setting up new cohorts is also a considerations, to expand the international representativeness of this signature. We are currently forming collaborations in China and India, some of the countries most affected by the diabetes epidemic.

Evidently, access to cohorts involves the completion of the paperwork aforementioned, but more importantly, the ability to connect with other scientists as collaborators. This no doubt requires much effort, perseverance, patience and flexibility. Following access to these cohorts, appropriate consideration for the time of follow up should guide any statistical analysis path chosen. A similar strategy as presented here may yield great results. However, other metabolites may prove more predictive. Determining which metabolites are shared amongst international cohorts and which are specific would certainly be an exciting endeavor, and further inform us on the etiology of disease. Such activities would fall under the umbrella of knowledge translation and exchange activities.

Lastly, while we did not have access to the clinical variables of the SWIFT cohort so as to form a combined clinical and metabolomics signature, such an avenue is definitely worth exploring. The combination of easily available information with metabolites may therefore be a better prediction than either metabolites or clinical variables alone.
7.2 Exploring Mechanistic Implications of Metabolite Signature

Metabolites identified by metabolomics are either associated or predictive of diabetes. It also follows that some of these metabolites may be causative of beta cell failure and hence diabetes. Given the prospective nature of this study, it is reasonable to assume that some of the top circulating metabolites identified may in fact play a causal role in diabetes development. Theoretically, these compounds could either cause impaired insulin secretion in β-cells or induce insulin resistance in peripheral or liver tissue. Hence, as a future direction, appropriate in vitro and in vivo studies could be designed and executed on candidate metabolites.

The primary read out to determine whether a metabolite causes impaired or altered insulin secretion in β-cells should be through a direct glucose-stimulated insulin secretion test on β-cell lines such as MIN6 cells INS1 cells, followed by freshly isolated mice or human islets. In the case of cell lines, cells may first be seeded on a 96 well plate. Once 90% confluence is reached, cells should subsequently treated with the metabolite (and in the case of a hydrophobic metabolite, a literature review should be conducted to see if this compound is BSA-bound) dissolved in the appropriate media in a range of concentration spanning the physiological and pathological range. Incubation time of metabolite should also be varied between 4hr and 24hr to reflect acute and chronic exposure. Afterwards, cells will be washed three times in KRB solution to remove any media, incubated in low glucose KRB for 1hr to allow cells to reach basal levels of insulin secretion. This would be followed by stimulation of low glucose or high glucose KRB solution (actual concentrations previously determined on untreated cells dose-response curve). Solution collected from each of the two conditions can be assayed for insulin concentration, and results may be normalized to DNA or protein. All individual conditions should be performed in triplicates, and all conditions replicated at least 3 times, and a p value of less than 0.05 will be deemed significant.

Following confirmation of an effect at a specified dose on β-cells, detailed mechanistic studies should follow. A Caspase assay could be used to gauge the rate of apoptosis, and a PCR could be carried out for ER stress markers such as XBP1. Additional assays for IL-1, TNFα and other inflammatory markers may also be performed.

With regards to effect on insulin resistance, a similar strategy could be applied on in vitro on liver or muscle cell lines, or ex vivo on primary hepatocytes and muscle extracts. After
treatment of compound at the relevant normoglycemic and hyperglycemic concentration observed from various studies, level of GLUT4 phosphorylation or translocation and other relevant markers following insulin treatment should be characterized. As for studies on β-cells, the timing is also critical and effect of acute and chronic exposure should be reported separately.

Ultimately, in vivo experiments in mice or rat models may also provide insight into the effect seen ex vivo. The design of the experiment should reflect the human conditions as closely as possible. For example, if a metabolite is known to be elevated in a high BMI background, obesity can be induced in the chosen mice strain before treatment of the compound. Euglycemic-hyperglycemic clamps should be incorporated into the design as they provide the best measure of insulin resistance and β-cell function. For an example metabolite of the above strategy, refer to Prentice et al (2014).

Recently, great interest has fostered in the influence of the brain on metabolism, such as for example appetite. For metabolites that may cross the blood-brain barrier, appropriate experiments should be designed that explore any effect on hypothalamic insulin signaling. This may be achieved through hyperinsulinemic euglycemic clamp with an injection directly into the CNS, or through knockout of the brain insulin receptor. Key exchange activities will involve collaborating with experts in the field.

Lastly, the –omic strategies discussed above in the analysis of human plasma could similarly be applied on animal tissue. For example, a proteomics or metabolomics of β-cell in response to incubation of a given metabolite could provide a system’s view of the changes that occur. Similarly, a proteomics analysis of primary hepatocytes following chronic exposure to a metabolite may provide important insights into the state of various proteins, and which of these becomes phosphorylated.

### 7.3 Facilitating Metabolomic Investigations

Currently, a great hindrance in metabolomics studies is the high cost associated with the technology. Characterizing the levels of each metabolite would require a prohibitive amount of blood as well as prove financially unfeasible. While the invention of technologies such as the Biocrates plate technology allow for more cost-efficient analysis, these are limited by the number of metabolites chosen by the manufacturer of highest interest.
Several potential solutions exist for this solution that should be explored in future research. First, it is possible that as more researchers explore metabolomics as an option, the cost per sample would reduce. Secondly, given that we have identified a metabolomics signature of only 6 metabolites, developing and executing a protocol that characterizes these metabolites only may be more economic than characterizing all 185 metabolites. This would involve research in sample handling, extraction and isolation, as well as an understanding of which metabolites may be grouped together in an assay. Thirdly, for some metabolites, it may be possible to develop ELISAs or enzymatic-based colorimetric kits. To do so, research into the biochemistry of these metabolites as well as whether a highly specific and sensitive enzyme to the metabolite of interest can be isolated should be funded. A more economic metabolomics assays will ultimately result in the ability of more researchers to characterize plasma samples, and potentially result in more accurate predictions at the level population health, and greater understanding of disease etiology at the level of the cell. Ultimately, such is the goal of this new emerging field of system epidemiology. Crossing the barriers between the bench and the bedside, system epidemiology is integral in promoting knowledge translation.
References


carriers of the transcription factor 7-like 2 polymorphism rs7903146. PloS One, 8(10), e78430.


Appendix 1: Test Run Quality Control Parameters

Table 1 Test run quality control parameters for each subpanel, with regards to stability (from the freeze-thaw P value), precision (through the %CV), accuracy (%Recovery) as well as the female plasma dynamic range (FP DR). The dynamic range may be quantitative (QNT), semi-quantitative (SQ), below the lower limit of quantification (LLOQ) or below the limit of detection (LOD).

<table>
<thead>
<tr>
<th>Order</th>
<th>Metabolite</th>
<th>Subpanel</th>
<th>#Cycles</th>
<th>F/T Test (P value)</th>
<th>Precision (%CV)</th>
<th>Accuracy (%Recovery)</th>
<th>FP DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SM(OH)C14:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>2</td>
<td>SM(OH)C16:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>3</td>
<td>SM(OH)C22:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>4</td>
<td>SM(OH)C22:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>5</td>
<td>SM(OH)C22:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>6</td>
<td>SMC16:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>7</td>
<td>SMC16:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>8</td>
<td>SMC18:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>9</td>
<td>SMC18:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>10</td>
<td>SMC20:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>11</td>
<td>SMC22:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>12</td>
<td>SMC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.1</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>13</td>
<td>SMC24:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>14</td>
<td>SMC26:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>15</td>
<td>SMC26:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>16</td>
<td>LPCaC14:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>LOD</td>
</tr>
<tr>
<td>17</td>
<td>LPCaC16:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>18</td>
<td>LPCaC16:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>19</td>
<td>LPCaC17:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>20</td>
<td>LPCaC18:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>21</td>
<td>LPCaC18:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>22</td>
<td>LPCaC18:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>23</td>
<td>LPCaC20:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>24</td>
<td>LPCaC20:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>25</td>
<td>LPCaC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>26</td>
<td>LPCaC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>27</td>
<td>LPCaC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>28</td>
<td>LPCaC28:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>29</td>
<td>LPCaC28:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>70-130</td>
<td>SQ</td>
</tr>
<tr>
<td>30</td>
<td>LPCaC6:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.1</td>
<td>&lt;5%</td>
<td>70-130</td>
<td>SQ</td>
</tr>
<tr>
<td>31</td>
<td>PCaaC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>70-130</td>
<td>SQ</td>
</tr>
<tr>
<td>32</td>
<td>PCaaC24:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>33</td>
<td>PCaaC28:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>PCaaC30:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>35</td>
<td>PCaaC30:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>LOD</td>
</tr>
<tr>
<td>36</td>
<td>PCaaC32:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>37</td>
<td>PCaaC32:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>38</td>
<td>PCaaC32:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>39</td>
<td>PCaaC32:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>40</td>
<td>PCaaC34:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>41</td>
<td>PCaaC34:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>42</td>
<td>PCaaC34:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>43</td>
<td>PCaaC34:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>44</td>
<td>PCaaC36:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>45</td>
<td>PCaaC36:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>46</td>
<td>PCaaC36:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>47</td>
<td>PCaaC36:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>48</td>
<td>PCaaC36:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>49</td>
<td>PCaaC36:5</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>50</td>
<td>PCaaC36:6</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>51</td>
<td>PCaaC38:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>52</td>
<td>PCaaC38:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>53</td>
<td>PCaaC38:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>54</td>
<td>PCaaC38:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>55</td>
<td>PCaaC38:5</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>56</td>
<td>PCaaC38:6</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>57</td>
<td>PCaaC40:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>LOD</td>
</tr>
<tr>
<td>58</td>
<td>PCaaC40:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>59</td>
<td>PCaaC40:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>60</td>
<td>PCaaC40:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>61</td>
<td>PCaaC40:5</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>62</td>
<td>PCaaC40:6</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>63</td>
<td>PCaaC42:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>64</td>
<td>PCaaC42:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>65</td>
<td>PCaaC42:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>66</td>
<td>PCaaC42:4</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>67</td>
<td>PCaaC42:5</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>68</td>
<td>PCaaC42:6</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>LOD</td>
</tr>
<tr>
<td>69</td>
<td>PCaeC30:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>70</td>
<td>PCaeC30:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td>SQ</td>
</tr>
<tr>
<td>71</td>
<td>PCaeC30:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>72</td>
<td>PCaeC32:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>73</td>
<td>PCaeC32:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>74</td>
<td>PCaeC34:0</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>75</td>
<td>PCaeC34:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>76</td>
<td>PCaeC34:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td>77</td>
<td>PCaeC34:3</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>SQ</td>
</tr>
<tr>
<td></td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td></td>
<td>80-120</td>
<td>SQ</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---------------</td>
<td>------</td>
<td>----</td>
<td>----</td>
<td>--------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>PCaeC36:0</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>PCaeC36:1</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>PCaeC36:2</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>PCaeC36:3</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>PCaeC36:4</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>PCaeC36:5</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>PCaeC38:0</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>PCaeC38:1</td>
<td></td>
<td></td>
<td>&gt;20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>PCaeC38:2</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>PCaeC38:3</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>PCaeC38:4</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>PCaeC38:5</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>PCaeC38:6</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>PCaeC40:0</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>PCaeC40:1</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>PCaeC40:2</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>PCaeC40:3</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>PCaeC40:4</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>PCaeC40:5</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>PCaeC40:6</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>PCaeC42:0</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>PCaeC42:1</td>
<td></td>
<td></td>
<td>&gt;20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>PCaeC42:2</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>PCaeC42:3</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>PCaeC42:4</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>PCaeC42:5</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>PCaeC44:3</td>
<td></td>
<td></td>
<td>&lt;20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>PCaeC44:4</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>PCaeC44:5</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>PCaeC44:6</td>
<td></td>
<td></td>
<td>&lt;15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>AC0</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>AC2</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>AC3</td>
<td></td>
<td></td>
<td>&lt;5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC3-DAC (AC4-OH)</td>
<td>2, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC3-OH</td>
<td>2, 3</td>
<td>NS</td>
<td>20%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC3:1</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC4</td>
<td>2, 3</td>
<td>NS</td>
<td>20%</td>
<td>70-130</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC4:1</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC5</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC5-DAC (AC6-OH)</td>
<td>2, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC5-M-DAC</td>
<td>2, 3</td>
<td>&lt;0.05</td>
<td>&lt;5%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AC5-OH (AC3-DAC-M)</td>
<td>2, 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LLOQ</td>
<td>2, 3</td>
<td></td>
<td>&lt;5%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOD</td>
<td>2, 3</td>
<td></td>
<td>&lt;15%</td>
<td>80-120</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analyte</td>
<td>Source</td>
<td>S/N</td>
<td>CV</td>
<td>LOD</td>
<td>LOD</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>AC5:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>AC5:1-DAC</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>AC6 (AC4:1-DAC)</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>AC6:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>Out of range</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>AC7-DAC</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.05</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>AC8</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>AC8:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>AC9</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.1</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>AC10</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>AC10:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>AC10:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>AC12</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>AC12-DAC</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>AC12:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>AC14</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>AC14:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>136</td>
<td>AC14:1-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>AC14:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.1</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>AC14:2-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>AC16</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>AC16-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>AC16:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>142</td>
<td>AC16:1-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>AC16:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>70-130</td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>AC16:2-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>AC18</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>AC18:1</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>AC18:1-OH</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>&lt;0.05</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>148</td>
<td>AC18:2</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>70-130</td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>Arg</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>Gln</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>Gly</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>His</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>Met</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&gt;20%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>Orn</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>Phe</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>156</td>
<td>Pro</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>Ser</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>158</td>
<td>Thr</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>159</td>
<td>Trp</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>Tyr</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>Val</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td>162</td>
<td>xLeu</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td>80-120</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>Leu</td>
<td>Special AA</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>164</td>
<td>Ile</td>
<td>Special AA</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>2-AAA</td>
<td>Special AA</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>PAG</td>
<td>Special AA</td>
<td>2, 3</td>
<td>&lt;0.05</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>Hexose</td>
<td>Biocrates</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>Glucose</td>
<td>Glucose</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>BHB</td>
<td>BHB</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>CMPF MS</td>
<td>Free CMPF Bound</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>CMPF ELISA</td>
<td>CMPF</td>
<td>2, 3</td>
<td>NS</td>
<td>&lt;20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>FFAC14:0</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>173</td>
<td>FFAC16:0</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>FFAC16:1 n-7</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>FFAC16:1 n-9</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>176</td>
<td>FFAC18:0</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>FFAC18:1</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>FFAC18:2</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>FFAC18:3</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>FFAC20:0</td>
<td>FFA</td>
<td>1 to 5</td>
<td>N/A</td>
<td>&gt;20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>FFAC20:1</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>FFAC20:4</td>
<td>FFA</td>
<td>1 to 5</td>
<td>&lt;0.05</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>FFAC20:5</td>
<td>FFA</td>
<td>1 to 5</td>
<td>&lt;0.05</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>184</td>
<td>FFAC22:5</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>FFAC22:6</td>
<td>FFA</td>
<td>1 to 5</td>
<td>NS</td>
<td>&lt;15%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: SWIFT Sample Run Quality Control Parameters

Table 1 Quality Control Parameters of metabolites from the Biocrates plate technology, for each batch.
A total of 86, 86, 86 and 4 samples were assayed in batches 1, 2, 3 and 4 respectively. The precision of the FPIC for each metabolite was noted (FP %CV), along with the swift plasma samples dynamic range (PL DR). The dynamic range may be quantitative (QNT), semi-quantitative (SQ), below the lower limit of quantification (LLOQ) or below the limit of detection (LOD).

<table>
<thead>
<tr>
<th>Order</th>
<th>Metabolite</th>
<th>FP CV</th>
<th>PL DR</th>
<th>FP CV</th>
<th>PL DR</th>
<th>FP CV</th>
<th>PL DR</th>
<th>FP CV</th>
<th>PL DR</th>
<th>Include (Y) or Reject (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SM(OH)C14:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>SM(OH)C16:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>SM(OH)C22:1</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>SM(OH)C22:2</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>SM(OH)C24:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>SMC16:0</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>SMC16:1</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>SMC18:0</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>SMC18:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>SMC20:2</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>SMC22:3</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>SMC24:0</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>13</td>
<td>SMC24:1</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>SMC26:0</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>15</td>
<td>SMC26:1</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>16</td>
<td>LPCaC14:0</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>LPCaC16:0</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>18</td>
<td>LPCaC16:1</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>19</td>
<td>LPCaC17:0</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>20</td>
<td>LPCaC18:0</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>21</td>
<td>LPCaC18:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>22</td>
<td>LPCaC18:2</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>23</td>
<td>LPCaC20:3</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>24</td>
<td>LPCaC20:4</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>25</td>
<td>LPCaC24:0</td>
<td>&lt;15%</td>
<td>SQLOD</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>26</td>
<td>LPCaC26:0</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>27</td>
<td>LPCaC26:1</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>28</td>
<td>LPCaC28:0</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>29</td>
<td>LPCaC28:1</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>30</td>
<td>LPCaC6:0</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>31</td>
<td>PCaaC24:0</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>32</td>
<td>PCaaC26:0</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>PCaaC28:1</td>
<td>&lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>PCaaC30:0</td>
<td>&lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>PCaaC30:2</td>
<td>&gt;20% SQLOD &gt;20% SQLOD &gt;20% SQ &gt;20% SQ N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>PCaaC32:0</td>
<td>&lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>PCaaC32:1</td>
<td>&lt;15% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>PCaaC32:2</td>
<td>&lt;15% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>PCaaC32:3</td>
<td>&lt;5% SQ &gt;20% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>PCaaC34:1</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>PCaaC34:2</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;20% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>PCaaC34:3</td>
<td>&lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>PCaaC34:4</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>PCaaC34:5</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>PCaaC34:6</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>PCaaC36:0</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>PCaaC36:1</td>
<td>&lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>PCaaC36:2</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>PCaaC36:3</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>PCaaC36:4</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>PCaaC36:5</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>PCaaC36:6</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>PCaaC38:0</td>
<td>&lt;15% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>PCaaC38:1</td>
<td>&lt;15% SQ &gt;20% SQ &lt;5% SQ &lt;15% SQ N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>PCaaC38:2</td>
<td>&lt;5% SQ &lt;15% SQ &lt;15% SQ &lt;5% SQ &lt;5% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>PCaaC38:3</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>PCaaC38:4</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>PCaaC38:5</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>PCaaC38:6</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>PCaaC38:7</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>PCaaC38:8</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>PCaaC38:9</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>PCaaC38:10</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>PCaaC38:11</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>PCaaC38:12</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>PCaaC38:13</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>PCaaC38:14</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>PCaaC38:15</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>PCaaC38:16</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>PCaaC38:17</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>PCaaC38:18</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>PCaaC38:19</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>PCaaC38:20</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>PCaaC38:21</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>PCaaC38:22</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>PCaaC38:23</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>PCaaC38:24</td>
<td>&lt;5% SQ &lt;15% SQ &lt;5% SQ &lt;15% SQ Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCaeC34:3</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>------</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>77</td>
<td>PCaeC36:0</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>78</td>
<td>PCaeC36:1</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>79</td>
<td>PCaeC36:2</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>80</td>
<td>PCaeC36:3</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>81</td>
<td>PCaeC36:4</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>82</td>
<td>PCaeC36:5</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>83</td>
<td>PCaeC38:0</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>84</td>
<td>PCaeC38:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>85</td>
<td>PCaeC38:2</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>86</td>
<td>PCaeC38:3</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>87</td>
<td>PCaeC38:4</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>88</td>
<td>PCaeC38:5</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>89</td>
<td>PCaeC38:6</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>90</td>
<td>PCaeC40:0</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>91</td>
<td>PCaeC40:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>92</td>
<td>PCaeC40:2</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>93</td>
<td>PCaeC40:3</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>94</td>
<td>PCaeC40:4</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>95</td>
<td>PCaeC40:5</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>96</td>
<td>PCaeC40:6</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>97</td>
<td>PCaeC42:0</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQLQD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>98</td>
<td>PCaeC42:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>99</td>
<td>PCaeC42:2</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>100</td>
<td>PCaeC42:3</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>101</td>
<td>PCaeC42:4</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&gt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>102</td>
<td>PCaeC42:5</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>103</td>
<td>PCaeC44:3</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>104</td>
<td>PCaeC44:4</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>105</td>
<td>PCaeC44:5</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>106</td>
<td>PCaeC44:6</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>107</td>
<td>AC0</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>108</td>
<td>AC10:2</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>109</td>
<td>AC12</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>110</td>
<td>AC12-DC</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>111</td>
<td>AC12:1</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>112</td>
<td>AC14</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>113</td>
<td>AC14:1</td>
<td>&lt;15%</td>
<td>SQLOD</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>N</td>
</tr>
<tr>
<td>114</td>
<td>AC14:1-OH</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>115</td>
<td>AC14:2</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQLOD</td>
<td>&lt;15%</td>
<td>SQLOD</td>
<td>&lt;15%</td>
<td>SQLOD</td>
<td>N</td>
</tr>
<tr>
<td>116</td>
<td>AC14:2-OH</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>117</td>
<td>AC16</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>N</td>
</tr>
<tr>
<td>118</td>
<td>AC16-OH</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>119</td>
<td>AC16:1</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>AC16:1-OH</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>122</td>
<td>AC16:2</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>123</td>
<td>AC16:2-OH</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>124</td>
<td>AC18</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>125</td>
<td>AC18:1</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>SQLQ</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>126</td>
<td>AC18:1-OH</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>127</td>
<td>AC18:2</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>128</td>
<td>AC2</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>129</td>
<td>AC3</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>Y</td>
</tr>
<tr>
<td>130</td>
<td>AC3-DC (AC4-OH)</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>131</td>
<td>AC3-OH</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>132</td>
<td>AC3:1</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>133</td>
<td>AC4</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>Y</td>
</tr>
<tr>
<td>134</td>
<td>AC4:1</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>135</td>
<td>AC5</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>Y</td>
</tr>
<tr>
<td>136</td>
<td>AC5-DC (AC6-OH)</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>137</td>
<td>AC5-M-DC</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>138</td>
<td>AC5-OH</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>SQLOD</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>139</td>
<td>AC5:1</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>140</td>
<td>AC5:1-DC</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>141</td>
<td>AC6 (AC4:1-DC)</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOQ</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>142</td>
<td>AC6:1</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>143</td>
<td>AC7-DC</td>
<td>&gt;20%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>144</td>
<td>AC8</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>145</td>
<td>AC8:1</td>
<td>&lt;15%</td>
<td>SQ</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>&lt;5%</td>
<td>SQ</td>
<td>Y</td>
</tr>
<tr>
<td>146</td>
<td>AC9</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>N</td>
</tr>
<tr>
<td>147</td>
<td>AC10</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>LLOQ</td>
<td>&lt;15%</td>
<td>LLOQ</td>
<td>&lt;20%</td>
<td>LOD</td>
<td>Y</td>
</tr>
<tr>
<td>148</td>
<td>AC10:1</td>
<td>&lt;15%</td>
<td>LOD</td>
<td>&lt;20%</td>
<td>SQLOD</td>
<td>&lt;5%</td>
<td>LOD</td>
<td>&gt;20%</td>
<td>SQLOD</td>
<td>N</td>
</tr>
<tr>
<td>149</td>
<td>Arg</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>150</td>
<td>Gln</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>151</td>
<td>Gly</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>152</td>
<td>His</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>153</td>
<td>Met</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>154</td>
<td>Orn</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>155</td>
<td>Phe</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>156</td>
<td>Pro</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>157</td>
<td>Ser</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>158</td>
<td>Thr</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>159</td>
<td>Trp</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>160</td>
<td>Tyr</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>161</td>
<td>Val</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>162</td>
<td>xLeu</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
<tr>
<td>167</td>
<td>Hexose</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;5%</td>
<td>QNT</td>
<td>&lt;15%</td>
<td>QNT</td>
<td>Y</td>
</tr>
</tbody>
</table>
**Table 2** Quality Control Parameters of metabolites from the special amino acid, free CMPF and free fatty acid subpanels, for each batch. The precision of the FPIC for each metabolite was noted (FP %CV). All metabolites were within the dynamic range and quantitative, except for FFAC20:0 which was not used in modeling.

<table>
<thead>
<tr>
<th>Order</th>
<th>Metabolites</th>
<th>%CV of FPIC or FPAA</th>
<th>Fold Change Range (Correction Factor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>163</td>
<td>Leu</td>
<td>&lt;5% &lt;15%</td>
<td>1 0.8-1.2 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>164</td>
<td>Ile</td>
<td>&lt;15% &lt;5%</td>
<td>1 0.7-1.3 0.7-1.4 0.7-1.3 0.7-1.3 0.8-1.2</td>
</tr>
<tr>
<td>165</td>
<td>2-AAA</td>
<td>&lt;5% &lt;15%</td>
<td>1 0.8-1.2 0.7-1.3 0.7-1.3 0.8-1.2 1.4</td>
</tr>
<tr>
<td>166</td>
<td>PAG</td>
<td>&lt;5% &lt;5% N/A</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>169</td>
<td>BHB</td>
<td>&lt;15% &lt;15% &lt;5% &lt;5%</td>
<td>1 0.7-1.3 0.7-1.3 0.8-1.2 0.7-1.3 0.8-1.2</td>
</tr>
<tr>
<td>170</td>
<td>Free CMPF</td>
<td>&lt;15% &lt;5% &lt;5% &lt;5% &lt;5%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>172</td>
<td>FFAC14:0</td>
<td>&lt;5% &lt;15% &lt;5% &lt;5% &lt;5%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>173</td>
<td>FFAC16:0</td>
<td>&lt;15% &lt;15% &lt;5% &lt;5% &lt;5%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>174</td>
<td>FFAC16:1n-7</td>
<td>&lt;5% &lt;15% &lt;5% &lt;5% &lt;5%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>175</td>
<td>FFAC16:1n-9</td>
<td>&lt;15% &lt;5% &lt;15% &lt;5% &lt;5% &lt;15% &lt;5% &lt;5% &lt;5%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>176</td>
<td>FFAC18:0</td>
<td>&lt;15% &lt;15% &lt;15% &lt;5% &lt;5% &lt;15% &lt;15%</td>
<td>1 0.7-1.3 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>177</td>
<td>FFAC18:1</td>
<td>&lt;5% &lt;15% &lt;5% &lt;15% &lt;5% &lt;15%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>178</td>
<td>FFAC18:2</td>
<td>&lt;5% &lt;15% &lt;5% &lt;5% &lt;15% &lt;15%</td>
<td>1 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>179</td>
<td>FFAC18:3</td>
<td>&lt;5% &lt;5% &lt;15% &lt;5% &lt;15% &lt;15%</td>
<td>1 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>180</td>
<td>FFAC20:0</td>
<td>LOD &lt;15% &lt;15% &lt;5%</td>
<td>0.8-1.2 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>181</td>
<td>FFAC20:1</td>
<td>&lt;5% &lt;15% &lt;15% &lt;5% &lt;5%</td>
<td>0.8-1.2 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>182</td>
<td>FFAC20:4</td>
<td>&lt;15% &lt;15% &lt;20% &lt;15% &lt;15%</td>
<td>0.8-1.2 0.7-1.3 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>183</td>
<td>FFAC20:5</td>
<td>&lt;20% &lt;15% &lt;5% &lt;15% &lt;15%</td>
<td>0.8-1.2 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
<tr>
<td>184</td>
<td>FFAC22:5</td>
<td>&lt;5% &lt;15% &lt;20% &lt;15% &lt;5%</td>
<td>0.8-1.2 0.8-1.2 0.6 0.8-1.2 0.7-1.3</td>
</tr>
<tr>
<td>185</td>
<td>FFAC22:6</td>
<td>&lt;15% &lt;5% &lt;15% &lt;15% &lt;15%</td>
<td>0.8-1.2 0.8-1.2 0.7-1.3 0.8-1.2 0.8-1.2</td>
</tr>
</tbody>
</table>
Appendix 3: Metabolite Concentration Distribution and Outliers

**Table 1** Distribution as well as presence of outliers in metabolites, in training set samples.

<table>
<thead>
<tr>
<th>Order</th>
<th>Metabolite</th>
<th>Normality</th>
<th>Outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SM(OH)C14:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>2</td>
<td>SM(OH)C16:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>3</td>
<td>SM(OH)C22:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>4</td>
<td>SM(OH)C22:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>5</td>
<td>SM(OH)C24:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>6</td>
<td>SMC16:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>7</td>
<td>SMC16:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>8</td>
<td>SMC18:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>9</td>
<td>SMC18:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>10</td>
<td>SMC20:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>12</td>
<td>SMC24:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>13</td>
<td>SMC24:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>17</td>
<td>LPCaC16:0</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>18</td>
<td>LPCaC16:1</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>19</td>
<td>LPCaC17:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>20</td>
<td>LPCaC18:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>21</td>
<td>LPCaC18:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>22</td>
<td>LPCaC18:2</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>23</td>
<td>LPCaC20:3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>24</td>
<td>LPCaC20:4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>33</td>
<td>PCaaC28:1</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>34</td>
<td>PCaaC30:0</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>36</td>
<td>PCaaC32:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>37</td>
<td>PCaaC32:1</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>38</td>
<td>PCaaC32:2</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>40</td>
<td>PCaaC34:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>41</td>
<td>PCaaC34:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>42</td>
<td>PCaaC34:3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>43</td>
<td>PCaaC34:4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>45</td>
<td>PCaaC36:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>46</td>
<td>PCaaC36:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>47</td>
<td>PCaaC36:3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>48</td>
<td>PCaaC36:4</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>49</td>
<td>PCaaC36:5</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>50</td>
<td>PCaaC36:6</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td></td>
<td>PCaaC38:0</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>51</td>
<td>PCaaC38:3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>53</td>
<td>PCaaC38:4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>54</td>
<td>PCaaC38:5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>55</td>
<td>PCaaC38:6</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>56</td>
<td>PCaaC40:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>58</td>
<td>PCaaC40:4</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>60</td>
<td>PCaaC40:5</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>61</td>
<td>PCaaC40:6</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>63</td>
<td>PCaaC42:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>66</td>
<td>PCaaC42:6</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>72</td>
<td>PCaeC32:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>74</td>
<td>PCaeC34:0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>75</td>
<td>PCaeC34:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>76</td>
<td>PCaeC34:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>77</td>
<td>PCaeC34:3</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>79</td>
<td>PCaeC36:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>80</td>
<td>PCaeC36:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>81</td>
<td>PCaeC36:3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>82</td>
<td>PCaeC36:4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>83</td>
<td>PCaeC36:5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>88</td>
<td>PCaeC38:4</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>89</td>
<td>PCaeC38:5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>90</td>
<td>PCaeC38:6</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>93</td>
<td>PCaeC40:2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>95</td>
<td>PCaeC40:4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>96</td>
<td>PCaeC40:5</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>97</td>
<td>PCaeC40:6</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>103</td>
<td>PCaeC42:5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>104</td>
<td>PCaeC44:3</td>
<td>Not Normal</td>
<td>Few</td>
</tr>
<tr>
<td>105</td>
<td>PCaeC44:4</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>106</td>
<td>PCaeC44:5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>107</td>
<td>PCaeC44:6</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>108</td>
<td>AC0</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>109</td>
<td>AC2</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>110</td>
<td>AC3</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>114</td>
<td>AC4</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>116</td>
<td>AC5</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>126</td>
<td>AC8:1</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>128</td>
<td>AC10</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>146</td>
<td>AC18:1</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>148</td>
<td>AC18:2</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>149</td>
<td>Arg</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>150</td>
<td>Gln</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>151</td>
<td>Gly</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>152</td>
<td>His</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>153</td>
<td>Met</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>154</td>
<td>Orn</td>
<td>Not Normal</td>
<td>Few</td>
</tr>
<tr>
<td>155</td>
<td>Phe</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>156</td>
<td>Pro</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>157</td>
<td>Ser</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>158</td>
<td>Thr</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>159</td>
<td>Trp</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>160</td>
<td>Tyr</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>161</td>
<td>Val</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>162</td>
<td>xLeu</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>163</td>
<td>Leu</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>164</td>
<td>Ile</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>165</td>
<td>2-AAA</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>166</td>
<td>PAG</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>167</td>
<td>Hexose</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>168</td>
<td>Glucose</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>169</td>
<td>BHB</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>170</td>
<td>CMPF MS</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>171</td>
<td>CMPF ELISA</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>172</td>
<td>FFAC14:0</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>173</td>
<td>FFAC16:0</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>174</td>
<td>FFAC16:1 n-7</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>175</td>
<td>FFAC16:1 n-9</td>
<td>Normal</td>
<td>Few</td>
</tr>
<tr>
<td>176</td>
<td>FFAC18:0</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>177</td>
<td>FFAC18:1</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
<tr>
<td>178</td>
<td>FFAC18:2</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>179</td>
<td>FFAC18:3</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>181</td>
<td>FFAC20:1</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>182</td>
<td>FFAC20:4</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>183</td>
<td>FFAC20:5</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>184</td>
<td>FFAC22:5</td>
<td>Normal</td>
<td>Many</td>
</tr>
<tr>
<td>185</td>
<td>FFAC22:6</td>
<td>Not Normal</td>
<td>Many</td>
</tr>
</tbody>
</table>
Copyright Acknowledgements (if any)