DIETARY PATTERNS AND METABOLOMICS

IN THE NUTRITION ETIOLOGY OF TYPE 1 DIABETES

by

RANDI KAY JOHNSON

B.S., Lipscomb University, 2006

M.P.H., Colorado School of Public Health, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Epidemiology Program

2019

This thesis for the Doctor of Philosophy degree by

Randi Kay Johnson

has been approved for the

Epidemiology Program

by

Tessa Crume, Chair

Jill Norris, Advisor

Katerina Kechris

Ulla Uusitalo

Oliver Fiehn

Date: May 18, 2019

Johnson, Randi Kay (PhD, Epidemiology Program)

Dietary Patterns and Metabolomics in the Nutrition Etiology of Type 1 Diabetes

Thesis directed by Professor and Chair Jill M. Norris

ABSTRACT

The incidence of type 1 diabetes (T1D) is increasing worldwide, likely due to the increase (or decrease, if protective) of key non-genetic factors. Though diet has been implicated in T1D development, its role is poorly understood, perhaps due to oversimplification of dietary exposures or failure to account for differences in biological processing of nutrients and foods. Metabolites mark response to diet, and may help elucidate this role when used in combination with reported intake. We examined combinations of dietary factors in the development of preclinical autoimmunity and T1D in two prospective studies: The Environmental Determinants of Diabetes in the Young (TEDDY) and the Diabetes Autoimmunity Study in the Young (DAISY). First, we identified individual metabolites and chemically similar metabolite groups associated with the development of multiple autoantibodies (mAb+) and T1D in metabolome-wide association studies. Candidate metabolites were used to capture disease-related dietary and nutrient patterns, which were subsequently tested for association with disease endpoints.

In TEDDY, unsaturated phosphatidylcholines, sphingomyelins, phosphatidylethanolamines, glucosylceramides, and phospholipid ethers in infancy were inversely associated with mAb+ risk, while dicarboxylic acids were associated with an increased risk. A dietary pattern explaining these metabolites in infancy was associated with decreased risk of mAb+ but was not generalizable when applied to similarly at-risk infants. In DAISY, a nutrient pattern explaining unknown metabolites at

seroconversion was associated with increased risk of progression to T1D. Nutrients highly contributing to this pattern included both established (linoleic acid, total sugars, niacin) and novel (vitamin C, riboflavin, vitamin K) dietary risk factors for T1D.

We made significant and novel contributions to the understanding of metabolomics and dietary patterns in the natural history of T1D. Our findings suggest that the contents of the diet may work in combination to affect the development of T1D, and should be considered jointly in future work. As demonstrated by this dissertation, expanding our methodological repertoire beyond individual risk factor epidemiology will improve our understanding of the disease process, and ultimately lead to more effective interventions to prevent, reverse, or delay development of T1D.

The form and content of this abstract are approved. I recommend its publication.

Approved: Jill M. Norris

Everyone needs food.

Not all bodies self-attack.

Should you change habits?

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ASK   Autoimmunity Screening in Kids

DAISY   Diabetes Autoimmunity Study in the Young

DCCT   Diabetes Complications and Control Trial

EtherPL   Phospholipid ethers

GADA   GAD as the first and only persistent confirmed autoantibody at IA

GlcCer   Glucosylceramides

HLA-DR3/4 Human leukocyte antigen genotype DR3-DQA1*05:01-DQB1*02:01/DR4-DQA1*03:01-DQB1*03:02

IA   Islet autoimmunity

IAA   Insulin as the first and only persistent confirmed autoantibody at IA

LOESS   Locally weighted scatterplot smoothing

mAb+   Multiple autoantibody positive, or stage 1 type 1 diabetes

PC   Unsaturated phosphatidylcholines

PE   Phosphatidylethanolamines

SM   Sphingomyelins

T1D   Type 1 Diabetes

TEDDY   The Environmental Determinants of Diabetes in the Young

TCA   Tricarboxylic acid

CHAPTER I

INTRODUCTION

Overview

Comprehensive dietary intake measurements and analyses are paramount to understanding the causes of nutrition-related chronic diseases. Type 1 diabetes (T1D) is a chronic autoimmune disease characterized by destruction of the insulin-producing beta-cells of the pancreas. Incidence of T1D is increasing worldwide, suggesting that an essential environmental (non-genetic) factor is increasing (or decreasing, if protective) with time. Diet may be an important environmental factor in the development of T1D and its preclinical phase, islet autoimmunity (IA), which is defined by circulating autoantibodies. Importantly, diet is a modifiable risk factor for IA and T1D.

Decades of research in the appearance of IA and progression to T1D have identified several putative nutrients and food groups that are associated with risk. For example, omega-3 fatty acids have been shown to decrease risk, and cow's milk protein intake increases risk. However, inconsistencies across studies suggest that focusing on single nutrients or foods may oversimplify the exposure. Nutrients enter the body as part of foods that contain other nutrients that may have synergistic or antagonistic effects. These nutrient intakes are so highly correlated that their effects are difficult to separate using traditional analytical approaches. To better understand the effects of diet on IA, we examined a more comprehensive picture of intake than previous studies using a summary measure of foods eaten in combination, hereafter referred to as a dietary pattern.

In addition to dietary intake, the biological response to what is eaten may also be implicated in the disease process. Once in the body, the pathways by which diet affect IA are not known. Ingested foods are digested, absorbed, and metabolized by the body, resulting in small molecules (i.e., metabolites) circulating in the bloodstream. These metabolites are important intermediaries because they reflect both what is ingested, and how it is processed by the body. Metabolites are often used as independent markers of dietary intake, as a solution to the inaccuracies associated with self-reported measures. However, metabolites can also be used to identify important biological pathways in disease etiology when used in conjunction with comprehensive dietary intake information. Previous studies have shown that changes in diet are reflected by changes in metabolites. Despite the important link between dietary intake and metabolites, no IA or T1D studies have examined them together. In order to understand the true effects of diet on IA, and ultimately T1D, we need new approaches to characterizing dietary exposure that account for complex dietary intake and metabolism.

The primary purpose of this study was to identify metabolite-related dietary patterns associated with IA or T1D risk. By creating dietary patterns reflective of important metabolites, we gained new insights into the mechanism by which diet may exert influence in the development of IA/T1D. While dietary pattern analyses have been increasingly utilized in nutritional epidemiologic studies, this was the first application of these methods to IA or T1D etiologic research. The study was conducted in two prominent ongoing T1D cohorts that both recruited children at high genetic risk of T1D at birth and still follow them prospectively for the development of the disease: The

Environmental Determinants of Diabetes in the Young (TEDDY), and the Diabetes

Autoimmunity Study in the Young (DAISY). We accomplished the objective in three

primary aims:

1. Identify metabolites associated with IA/T1D.

   Hypothesis 1.1: Nutrition-related metabolites are associated with development of

   IA/T1D.

2. Create dietary patterns reflective of important metabolites.

3. Test the association of dietary patterns with development of IA/T1D.

   Hypothesis 3.1: At least one dietary pattern is associated with the development of

   IA/T1D.

 Metabolites are products of interactions among the genome, microbiome, and

dietary intake. Using the metabolome to identify potential mechanisms by which diet

may lead to IA can focus future etiologic research to those upstream factors in identified

pathways. Overall, this study's examination of the effect of combinations of metabolites

and food was a critical next-step for the long-term goal of developing dietary

interventions to prevent IA/T1D.

Innovation

 The study utilized innovative approaches to estimate dietary exposure and risk

of IA/T1D. To our knowledge, this was the first study to look at metabolite-associated

dietary patterns in risk of IA/T1D. It was also the first in the nutrition epidemiology

literature to examine nutrients in the identification of dietary patterns—previous

studies have exclusively used food group intake. Combining multiple food, nutrient, and

biomarker indicators allowed for a more complete picture of dietary exposure than

previous studies of IA and T1D, which have exclusively looked at individual rather than combinations of dietary risk factors. The use of metabolites to focus dietary patterns was a novel approach to "big data" studies.

The TEDDY study is an international, multi-site consortium that provides a rich opportunity to examine diet and metabolomics in a pediatric population that has been followed closely since birth. TEDDY also offers enough power and sample size to answer research questions utilizing novel empirically driven statistical methods, such as those proposed in this project. While the DAISY study is smaller, the ability to implement prospective dietary pattern and metabolomics analyses in two comparable yet independent populations uniquely positioned us to conduct the proposed innovative research. **Figure I-1** summarizes the prospective follow-up of children in the TEDDY and DAISY studies, and an overview of the identification of dietary patterns and metabolomics in the development of IA/T1D.
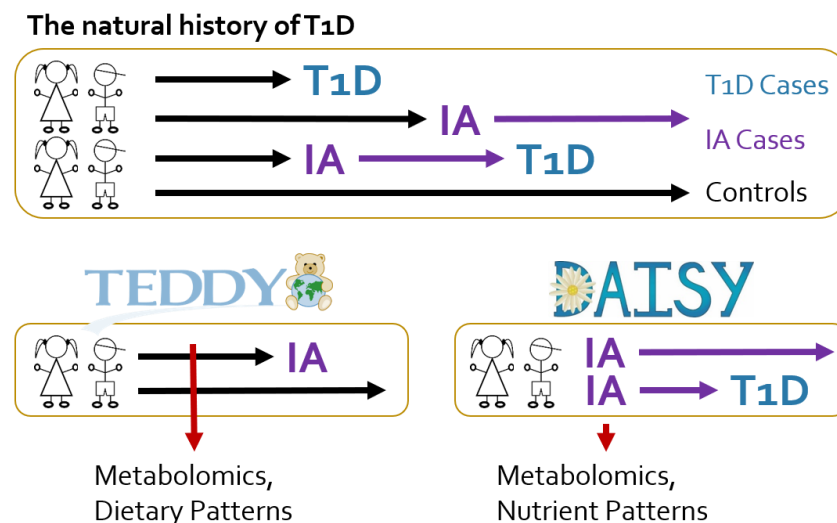
Figure I-1: Dissertation introduction. The prospective TEDDY and DAISY cohorts follow children for the development of islet autoimmunity (IA) and type 1 diabetes (T1D). We identified dietary patterns and metabolomics that were associated with the development of IA in the TEDDY study, and nutrient patterns and metabolomics that were associated with progression from IA to T1D in DAISY.

CHAPTER II

LITERATURE REVIEW

Natural History of Type 1 Diabetes

T1D is a chronic, immune-mediated disease characterized by the destruction of the insulin-producing beta-cells of the pancreas by autoreactive T-cells.[1] Development of these islet reactive autoantibodies is thought to be induced after a genetically-susceptible individual is exposed to some putative environmental risk factor that triggers a loss of immune regulation.[2] Destruction of the beta-cells leads to a decrease in insulin secretion, the development of hyperglycemia, and ultimately T1D diagnosis.

Preclinical Markers

T1D is preceded by IA, a preclinical, asymptomatic phase in which autoantibodies to beta-cell antigens are detectable in serum.[3] Four autoantibodies are considered biomarkers for IA, including insulin,[4] GAD,[5] IA-2,[6] and zinc transporter.[7] Seroconversion to IA-positivity occurs when one of the four autoantibodies are persistently detectable by immunoassay, typically no earlier than 3 months of age and by 2-4 years of age for the majority of people who progress on to T1D.[8–10]

The autoimmune process preceding T1D is heterogeneous—the type and number of first-appearing autoantibodies distinguish subgroups with shared characteristics in the disease process.[11] For example, while incidence of any IA peaks at 2 years of age, the incidence of insulin as the first-appearing autoantibody (IAA) peaks at age 1-year and rapidly declines while incidence of GAD as the first-appearing autoantibody (GADA) appears later with steadier incidence.[12,13] Other genetic and environmental factors, such as human leukocyte antigen (*HLA*) genotype, SNPs in

PTPN22, *ERBB3*, *INS*, and *BACH2* genes, age-at first-appearing autoantibody, male sex, and father as the diabetic proband may distinguish autoimmune phenotypes and the risk for developing a second autoantibody and/or T1D.[12] A study pooling data from three prominent T1D cohorts showed that 70% of children with multiple autoantibody positivity (mAb+, or stage 1 T1D) progressed on to T1D within 10 years of seroconversion, and the progression was faster for 1) children who seroconverted at less than 3 years of age compared to those 3 years and older, 2) children with the high-risk HLA DR3/DR4 genotype compared to other HLA genotypes, and 3) girls compared to boys.[14]

Genotype is another marker used to identify at-risk populations, and is used to recruit patients into most T1D etiologic studies. The Type 1 Diabetes Genetic Consortium was an international, multi-site collaborative that made major contributions to the understanding of genetic risk predictors for T1D,[15] including the identification of over 40 genetic loci associated with T1D risk.[16] The HLA region on chromosome 6 accounts for over 50% of the genetic risk for the disease, with some HLA class II haplotypes increasing disease risk (i.e., *DRB1*0301-DQA1*0501-DQB1*0201, DRB1*0405-DQA1*0301-DQB1*0302*), and some haplotypes decreasing risk (i.e., *DRB1*1501-DQA1*0102-DQB1*0602, DRB1*1401-DQA1*0101-DQB1*0503*).[17] However, the frequency of high-risk HLA genotypes among T1D cases has decreased over time, indicating that genetic predisposition is less predictive than in the past.[18,19]

<u>Diagnosis and Treatment</u>

T1D diagnoses occurs in stages, following guidelines set forth by the American Diabetes Association.[20] Given the high probability of progression to T1D once multiple

autoantibodies are detectable, Stage 1 of T1D is characterized by multiple autoantibody positivity, normoglycemia, and is generally presymptomatic. Patients with multiple autoantibody positivity and no clinical symptoms progress to Stage 2 T1D when dysglycemia begins, as measured by impaired fasting glucose or impaired glucose tolerance. Further onset of hyperglycemia and clinical symptoms indicates Stage 3 T1D, diagnosed by standard criteria where fasting blood glucose is higher than 7 mmol/L (126 mg/dL), 2-hour blood glucose is 11.1 mmol/L (200 mg/dL) or higher, symptoms of hyperglycemia accompanied by any blood glucose test is 11.1 mmol/L (200 mg/dL) or higher, or glycated hemoglobin (HbA$_{1c}$) test is 48 mmol/mol (6.5%) or higher.

After disease onset, T1D cases are treated with exogenous insulin for the remainder of their lives. Modern technological advances have improved treatment of T1D and development of complications through the use of devices that better monitor and regulate insulin and blood glucose levels (i.e., insulin pump, continuous glucose monitor, artificial pancreas device system), and by improved insulin preparations (i.e., long-acting versus rapid-acting insulin).[21] The Diabetes Complications and Control Trial (DCCT), which began in 1983, showed that intensive treatment aimed at keeping blood glucose close to normal levels was more effective at reducing diabetic-induced eye, kidney, and nerve damage than the standard treatment aimed at avoiding symptomatic hyper- or hypoglycemia.[22] These results changed the standard of clinical care for T1D patients. However, much of this progress is limited to the developed-world, as vast global disparities in insulin access, disease management, and technology exist.

While T1D has been associated historically with increased risk of early

mortality,[23,24] survival has improved over the last 80 years, especially in high-income

countries. Successes in insulin availability and treatment have led to increased life-

expectancy and decreased mortality among those diagnosed in childhood. The

Pittsburgh Epidemiology of Diabetes Complications study found that children

diagnosed with T1D from 1965 to 1980 were expected to live 68.8 years at birth, almost

15 years longer than those diagnosed between 1950 and 1964, and only 3.6 years less

than the general population.[25] From 2008-2009 to 1968-1969, the CDC diabetes death

rate decreased 68% among US children less than 20 years of age (to 1.05 per million

persons), and 78% among US children less than 10 years old.[26] In contrast, life-

expectancy is seen as low as 0.6 years among children under 15 years of age diagnosed

with T1D in low- and middle-income countries.[27] Limited access to trained healthcare

professionals, diagnostic tools, and insulin likely contribute to these disparities.[28,29]

T1D often leads to kidney and cardiovascular complications, which are the

primary causes of early mortality.[30] The observational follow-up study of 90% of DCCT

participants, called the Epidemiology of Diabetes Interventions and Complications,

showed that the intensive treatment received during the trial had long-term benefits,

reducing the development of renal disease,[22,31] cardiovascular disease,[32] and all-cause

mortality[33] among T1D patients. In the absence of renal disease, the 20-year mortality

risk for T1D patients diagnosed in childhood is now comparable to the nondiabetic

population.[34] While these improvements in T1D management, treatment and care have

led to successes in life-expectancy and mortality in those living with T1D, effective primary and secondary prevention strategies remain unclear.

<u>Primary and Secondary Prevention</u>

Given the likely role of environmental factors in disease pathogenesis, the early age at IA onset, and the high likelihood that children with IA will progress on to T1D, the few primary prevention trials have focused on early life dietary modifications to prevent IA. The Trial to Reduce IDDM in the Genetically at Risk showed that hydrolyzed infant formula was not associated with IA or T1D compared to cow's milk formula, among genetically at risk children who did not have breast milk available.[35,36] While dietary interventions for primary prevention have not yet been successful, participation and compliance in the trial indicates that a large-scale dietary intervention in young children is feasible.

Secondary prevention trials to prevent or delay the progression to T1D among those with IA have been largely unsuccessful. Neither oral insulin,[37] nicotinamide,[38] nor intranasal insulin[39] have been very effective at preventing or delaying diabetes onset. Reversal or halting of beta-cell destruction after T1D onset is not yet possible, though alternative treatments to exogenous insulin therapy, such as islet-cell transplantation, show promise.[40] Development of successful prevention strategies have been hindered by the inability to identify environmental risk factors unequivocally associated with disease, but are important to develop given the increasing presence of T1D worldwide.

Population-based screening for the high-risk multiple autoantibody positive children, whose lifetime risk of symptomatic (Stage 3) T1D approaches 100%,[14] has been initiated by the Fr1da Study in Germany[41] and the Autoimmunity Screening in

Kids (ASK) in the United States.[42] Earlier detection of Stage 1 T1D may reduce costly morbidities at symptom onset and improve longer-term diabetic control.[43–45] The identification of these high-risk populations also enhances the ability to conduct trials, and will allow for faster implementation of effective interventions to prevent or delay the progression from Stage 1 to Stage 3 T1D.

Epidemiology of Type 1 Diabetes

T1D affects approximately 40 million people worldwide,[28] including 1.4 million Americans.[46] Over 542,000 children are affected, making it the most common chronic disease in children.[28] The WHO DIAMOND project estimated T1D incidence in the 1990s to range from over 40 per 100,000 persons per year (in Finland) to less than 1 per 100,000 persons per year (in Venezuela) using in-country diabetes registries. Since then, global incidence of T1D has been increasing, especially in central and eastern European countries and among children less than 5 years of age.[47] The rate of increase in incidence is estimated between 3-4% annually in European children less than 15 years of age,[48] and 2.7% annually in the US among children less than 20 years of age.[49] Recent reports from Finland,[50] which has the highest incidence rate of T1D in the world at more than 60 per 100,000 persons per year,[51] and Sweden[52] indicate that the rate of increase among children is beginning to plateau, though some question the validity of those conclusions due to short periods of categorization used in time trend analyses.[53] Furthermore, T1D is often misclassified as type 2 diabetes, particularly in adults, likely leading to a gross underestimate of the true incidence of T1D.[54]

Age at T1D onset and sex differences in incidence vary geographically. Incidence rates tend to be higher in higher income countries and lower in African and other

developing nations, though incidence data are sparse in many low income countries.[47]

While T1D can develop at any age, the incidence in European countries is higher in children aged 0-14 years than in young adults aged 15 to 29 years.[55] T1D diagnoses tend to peak between 5-7 years of age and again near puberty in these populations.[56] In contrast, peak incidence in the few African nations with available data occurs in young adults aged 15 to 29 years.[57–59] Differences in T1D incidence by sex are slight, unlike most other autoimmune diseases. Males tends to have a slight excess in high-incidence countries, whereas females tend to have a slight excess in low-incidence countries.[60] These geographic differences in the epidemiology of T1D may be explained by differences in detection or heterogeneity in exposure to environmental disease triggers.

T1D incidence also displays seasonal patterns. More cases of T1D are diagnosed in the autumn and winter months, and fewer in the summer, regardless of age, sex, hours of sunshine, and average temperature.[61,62] Being born in the spring is also associated with an increased risk of T1D among U.S. youth, which may relate to sunlight exposure in either the prenatal or postnatal environment.[63] Heterogeneity in T1D incidence over time, geography, age, and season all support a theoretical role for environmental factors in driving its development. Furthermore, by age 60 years, only 65% of monozygotic twins are concordant for T1D status,[64–66] implicating environmental factors in disease etiology.

Though genetics explain the majority of risk for T1D, genetic changes cannot account for the rapid increases in T1D incidence.[67] Geographic, seasonal, and age differences in the incidence of T1D, combined with the lessening contribution of genetic factors, suggest environmental (non-genetic) influences are important.[68] Given the

peaks in IA incidence during infancy and adolescence, diet has been implicated as an environmental risk factor that triggers development of T1D on top of a background of higher genetic risk.

While in utero exposures[69–71] and the timing of food introduction in the infant diet[72–75] may make early contributions to the development of disease, childhood dietary intake is an important ongoing exposure. New results in children at high-genetic risk for T1D suggest that later introduction of gluten-containing cereals may increase the risk for IA, but the timing of introduction of other solid foods do not show a clear association.[76] Decades of research have identified several dietary factors that are associated with IA or T1D (**Table II-1**). Cow's milk protein[77,78], omega-3 fatty acids[79–81], glycemic index[82], and sugar[83] have been shown to be associated with either IA or T1D, while soluble fiber[84] and omega-6 fatty acids[79] have not been. Other dietary factors, such as vitamin D intake, may influence risk of IA or T1D, but only among subgroups with specific genotypes.[85,86] However, results are inconsistent across studies,[87] and no particular dietary agent has yet been confirmed in T1D pathogenesis.[68] Difficulties with dietary analytical approaches and exposure assessment may explain some of the inconsistency in T1D nutritional epidemiology research, and can be addressed by using metabolomics and dietary patterns.

Metabolomics and Dietary Intake in the Development of IA and T1D

Small molecules present in biological fluids (blood, saliva, urine), or metabolites, are increasingly studied as alternative markers of nutrition or disease status. Metabolites are products of interactions among the genome, microbiome, and dietary intake,[89] and are reflective of many factors related to human health.[90] Biomarker

discovery often focuses on identifying a single marker that represents a particular food group; however, a combination of metabolites may be a more accurate measure of dietary exposure or disease status.[91] Metabolites have been used in conjunction with dietary patterns to identify changes reflective of reported nutrition,[92] new biomarkers of foods or nutrients,[93] and biological pathways important for disease progression.[94] The few small IA and/or T1D studies that have explored metabolomics have identified serum methionine,[95] triglycerides,[95,96] antioxidant phospholipids,[96,97] and amino acids[95–97] as important predictors (**Table II-2**). Interpreted independently, these findings are inconclusive as these metabolites could reflect immune, genetic, dietary, or other biological changes.[98] Metabolites discovered in the context of dietary intake, on the other hand, are more interpretable, and their combined use has been shown to reduce bias and increase statistical power to detect diet-disease relationships.[99,100]

Table II-1: Previously studied dietary risk factors for IA/T1D.

Food or nutrient	Associated with IA or T1D	Direction of increased intake on risk	Reference
Cow's milk protein	IA*; T1D	↑; ↑	77,78
Omega-3 fatty acids	IA	↓	80
Fruit and berry juices	IA	↑	78
Sugar	T1D	↑	83
Soluble fiber	Neither		84
Omega-6 fatty acids	Neither		79,80
Vitamin D	Neither		85
Gluten	Neither		88

only for subjects with low risk HLA genotypes

Table II-2: Previously studied metabolomics risk factors for IA/T1D.

Metabolite	Associated with IA or T1D	Direction of increased levels on risk	Reference
Methionine	IA	↓*	95
Amino acids	T1D	↑	96,97
Triacylglycerols (TG)	IA; T1D	↑*; ↓	95,96
Phosphatidylcholines (PC)	T1D; young T1D; IA; T1D	↓; ↓; ↓; ↑	95–97,101–103
Ether phospholipids (ether PL)	T1D	↓	96
Sphingomyelins (SM)	T1D; not IA	↓	101,103
Phosphatidylethanolamines (PE)	Young T1D	↓	102
Polar metabolites (full panel)	IA	None	104

early seroconverters (<2 years) compared to late seroconverters (>=8 years)

Traditional investigation of diet and disease examines the effects of individual nutrients or foods, and does not account for the complexity of the diet—the effects of single nutrients and foods are often too small to identify, or too highly correlated to separate.[100] Newer methods that overcome the challenges of examining diet comprehensively are increasingly used to identify dietary patterns that integrate both food intake and biomarker status.[105–107] Dietary patterns represent a more complete picture of diet and are more strongly associated with disease than single nutrients or food.[108] For example, dietary patterns have been particularly useful in research on cardiovascular disease (CVD), where a Mediterranean diet pattern was consistently identified as protective in observational studies and intervention trials.[109] Dietary patterns offer a solution for examining whole-diet effects in the development of disease.

Comprehensive dietary intake measurement and analysis are paramount to understanding the causes of nutrition-related chronic diseases, such as T1D. Incidence of T1D is increasing worldwide, suggesting that an essential environmental (non-genetic) factor is increasing (or decreasing, if protective) with time. Diet may be an important, modifiable risk factor in the development of IA or T1D. Decades of research in the role of diet in the appearance of IA and progression to T1D have identified several putative nutrients and food groups that are associated with risk. For example, omega-3 fatty acids have been shown to decrease risk, and cow's milk protein intake increases risk. However, inconsistencies across studies suggest that focusing on single nutrients or foods may oversimplify the exposure (**Figure II-1**). Nutrients enter the body as part of foods that contain other nutrients that may have synergistic or antagonistic effects. These nutrient intakes are so highly correlated that their effects are difficult to separate using traditional analytical approaches. To better understand the effects of diet on IA, we will examine a more comprehensive picture of intake than previous studies using a summary measure of foods or nutrients eaten in combination, or a dietary pattern.

In addition to dietary intake, the biological response to what is eaten may also be implicated in the disease process. Once in the body, the pathways by which diet affect IA are not known. Ingested foods are digested, absorbed, and metabolized by the body, resulting in small molecules (i.e., metabolites) circulating in the bloodstream. These metabolites are important intermediaries because they reflect both what is ingested, and how it is processed by the body. Metabolites are often used as independent markers of dietary intake, as a solution to the inaccuracies associated with self-reported

measures. However, metabolites can also be used to identify important biological pathways in disease etiology when used in conjunction with comprehensive dietary intake information. Previous studies have shown that changes in diet are reflected by changes in metabolites. Despite the important link between dietary intake and metabolites, no IA or T1D studies have examined them together. In order to understand the true effects of diet on IA, and ultimately T1D, we need new approaches to characterizing dietary exposure that account for complex dietary intake and metabolism.

Figure II-1: Complexities in correlation of dietary intake exposures. a) Previous studies of dietary intake in IA/T1D oversimplify dietary exposures by considering only individual foods or nutrients, when in reality dietary intake is highly correlated. b) Nutrient and food intake effects are also highly correlated and their effects difficult to separate with traditional analytical approaches.

The overall objective of this dissertation is to identify metabolite-related dietary patterns associated with IA risk. We will accomplish this objective in three distinct scientific aims: 1) identify metabolites associated with IA/T1D, 2) create dietary patterns reflective of important metabolites, and 3) test the association of dietary patterns with development of IA/T1D (**Figure II-2**). By creating dietary patterns

reflective of important metabolites, we will gain new insights into the mechanism by which diet may exert influence in the development of IA or T1D. While dietary pattern analyses have been increasingly utilized in nutritional epidemiologic studies, this will be the first application of these methods to IA or T1D etiologic research. The study will be conducted in two prominent ongoing T1D cohorts that both recruited children at high genetic risk of T1D at birth, and follow them prospectively for development of the disease: The Environmental Determinants of Diabetes in the Young (TEDDY) study, and the Diabetes Autoimmunity Study in the Young (DAISY). TEDDY is a multi-center consortium that began following 8,676 newborn infants at high-genetic risk for T1D in the US, Germany, Sweden, and Finland in 2004.[110] DAISY has followed 2,547 children at high-genetic risk for T1D in Colorado for development of IA and further progression to T1D since 1993.[111]
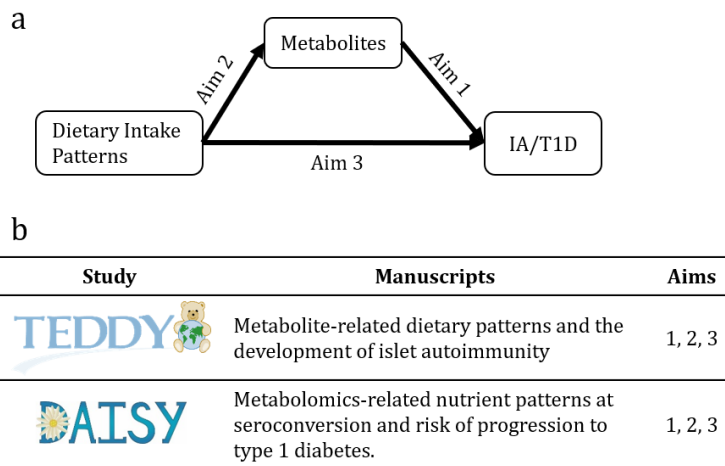
Figure II-2: Dissertation organization. a) This dissertation accomplished three aims in two prospective studies of T1D: The Environmental Determinants of Diabetes in the Young (TEDDY) and the Diabetes Autoimmunity Study in the Young (DAISY). b) Manuscripts were organized by study, including all three aims in each publishable unit.

CHAPTER III

APPROACH

Overview

To achieve our goal of elucidating more of the nutrition etiology of IA/T1D by identifying combinations of metabolites and foods that affect IA or T1D risk, we first test metabolites for risk of disease (Aim 1). Then we will identify dietary intake patterns that best explain the variation in those important metabolite candidates (Aim 2). Finally, we will apply those dietary patterns longitudinally to test their association with the development of IA or T1D (Aim 3). All aims were conducted independently in the TEDDY and DAISY studies, with alterations to the methods as appropriate for each study's design (**Table III-1**). This shift to combinations of dietary factors, rather than a reductionist one-at-a-time approach, may help clarify the inconsistencies in the role of diet in development of IA and T1D. Results from a dietary intake-level exploration that is focused by biologically meaningful metabolites are more interpretable than previous approaches and will highlight important biological mechanisms on the pathway from dietary intake to IA and T1D. Comparison of the measures available in both studies are described below (**Table III-2**).

Table III-1: Summary of approach by Aim and study.

Aim	Study	Study Design	Exposure(s)	Primary Outcome(s)	Statistical Method
Aim 1	TEDDY	Nested case-control	Metabolite or lipid	IA, mAb+	Conditional logistic regression
	DAISY	Cohort	Metabolite or lipid	T1D	Cox regression
Aim 2	TEDDY	Cross-sectional	Foods	Metabolites	Reduced rank regression
	DAISY	Cross-sectional	Nutrients	Metabolites	Reduced rank regression
Aim 3	TEDDY	Cohort	Dietary pattern	mAb+	Cox regression
	DAISY	Cohort	Nutrient pattern	T1D	Cox regression

Table III-2: Availability of relevant measures in TEDDY and DAISY.

	Follow-up in Months							Every 3 months (24 - 48)	Every 6 months (24 - 48)	Every 6 months (> 48)	Annually (> 24)
	Initial Screen	3	6	9	12	15	24				
Blood Draw	D T*	T	T	D T	T	D T	D T	T		T	D
Food Frequency Questionnaire							D				D
24-hr recall		T									
3-Day Diet Record			T	T	T		T		T	T	
Height and Weight		D T	D T	D T	D T	D T	D T	T		T	D

Study Designs, Populations, and Measures

<u>TEDDY Study</u>

From 2004-2010, TEDDY screened and enrolled 8,676 newborn infants with a high- or moderate-risk HLA genotype at six locations around the world: Georgia/Florida, Colorado, and Washington (in the USA), and Finland, Sweden, and Germany.[110] Dietary assessment is carried out by 24-hour recall at the first clinic visit at 3 months of age, then by 3-day food record every 3 months until 12 months of age, and then every 6 months thereafter until follow up stops at age 15 years.[112] The foods consumed were quantified into food groups (e.g., cereals, fruits and berries, etc.) and subgroups (e.g., wheat, rice, oats, citrus fruits, apple, berries, etc.). From quantities of foods consumed, the amount of nutrients contained therein were then quantified using national food composition databases unique to each country. Results of a detailed nutrient harmonization study documented 21 nutrients that were comparable across study centers, despite potential differences in food composition databases.[112] Results

from food group harmonization indicated 15 main food groups, and 89 subgroups were comparable among TEDDY study centers.[113]

Blood samples were collected on TEDDY participants every 3 months until 4 years of age, and biannually thereafter until age 15 (**Table III-2**).[114] A nested case-control study was conducted in TEDDY to reduce the amount of resources needed to assay samples for metabolomics and other measures. A case of IA was defined as the presence of an autoantibody (GAD65A, IA-2A or IAA) at two or more consecutive visits. Risk set sampling[115] was used to match 418 cases of confirmed persistent IA to 1253 controls (matched 1:3) who were autoantibody free at the time of the case's IA seroconversion visit, based on clinical center, sex, and family history of T1D.[116] At the time of selection in May 2012, the mean age of the cases was 24 months, 23% had a first degree relative with T1D, and 44% were female.[116] Untargeted primary metabolism and complex lipid panels were run at The NIH West Coast Metabolomics Center, University of California Davis, CA, USA on 9,394 plasma samples from every available time point prior to the time of IA seroconversion in cases and controls.[116]

DAISY Study

From 1993-2004 DAISY screened and enrolled 2,547 children in Colorado with high-risk HLA genotype or a first-degree relative with T1D, 1,886 of whom entered the study prior to age 15 months.[73] Prospective dietary assessment was collected via a food frequency questionnaire (FFQ) completed by the parent beginning when the child was 2 years old concerning the diet in the previous year, and continuing annually until age 10, when the subjects were asked to self-report diet using a youth adolescent questionnaire (YAQ). Data collected by FFQ and YAQ dietary assessment tools have been shown

comparable if an indicator is used designating the assessment tool.[117] The FFQ has been validated in the DAISY population using multiple 24-hour recalls and biomarkers.[118,119]

Blood samples were collected on DAISY participants at age 9, 15, and 24 months, and annually thereafter. A nested case-control study was conducted in DAISY using a similar case definition as in TEDDY, where a case of confirmed persistent IA was defined as the presence of one confirmed autoantibody (GAD65A, IA-2A IAA, or ZnT8) on two or more consecutive visits. Cases were frequency matched to controls by age, ethnicity, and sample availability for important time points of interest—e.g., birth, early visit (9 or 15 months of age), the visit prior to conversion to IA, the IA visit, and the visit just prior to T1D diagnosis (**Figure III-1**). Upon selection in 2015, the mean age of the cases was 12.9 years, 48% were female, and 79% were non-Hispanic white. Like in TEDDY, untargeted primary metabolism and complex lipid panels were run at The NIH West Coast Metabolomics Center, University of California Davis, CA, USA.
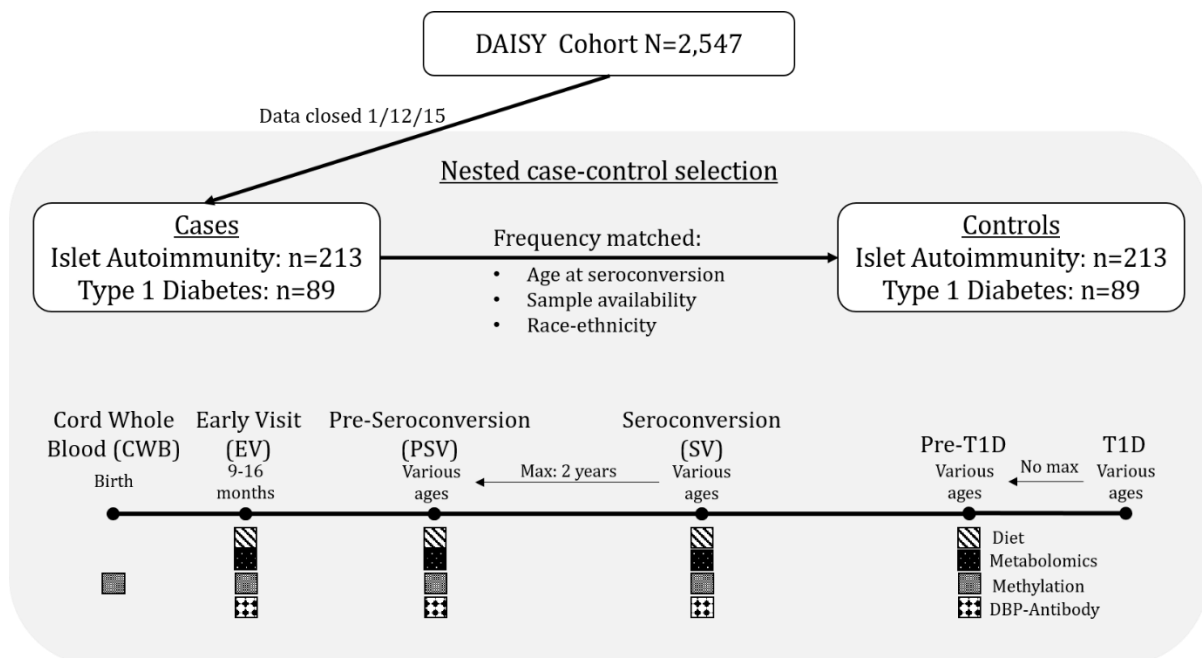
Figure III-1: DAISY nested case-control study design.

Specific Aim 1: Identify Metabolites Associated With IA/T1D

Metabolite Quantification and Pre-Processing

Led by dissertation committee member Dr. Oliver Fiehn at The NIH West Coast Metabolomics Center, University of California Davis, CA, USA, TEDDY measured metabolites on the 9,394 plasma samples in the nested case-control study. Two platforms were utilized: 1) CSH-QTOF MS platform which quantifies more than 350 lipids covering 12 lipid classes, and 2) GC-TOF MS platform which quantifies up to 170 primary metabolites such as amino acids, sugars, and hydroxyl acids. Initial data processing was completed at UC Davis and included locally estimated scatterplot smoothing (LOESS)[120] normalization to internal standards and adjustment for batch or series effects. After laboratory pre-processing, TEDDY received measurements on 364 metabolites from the GC-TOF platform and 816 lipids from the CSH-QTOF platform. These data became available on the TEDDY online data sharing platform in December 2016.

DAISY metabolomics were similarly acquired from the NIH West coast Metabolomics Center, using the GC-TOF platform for primary metabolites and the CSH-QTOF for lipids. An additional panel of data was acquired in DAISY, covering biogenic amines using HILIC-QTOF-MS. There were 734 DAISY visits selected for the nested case-control study with sufficient sample available for metabolomics, resulting in 2,720 metabolite features received from all 3 platforms. **Table III-3** summarizes the number of untargeted metabolites received by panel and study.

Table III-3: Summary of the number of untargeted metabolites received from The NIH West Coast Metabolomics Center by panel and study.

	TEDDY	DAISY
GC-TOF (primary metabolites)	364	363
CSH-QTOF (lipids)	816	1,181
HILIC QTOF-MS (biogenic amines)	NA	1,176
Total Number of Metabolites	1,180	2,720

Statistical Analysis

We conducted a metabolome-wide association study (MWAS) on the subset of children in the TEDDY cohort selected for the biomarker nested case-control study. Primary metabolomics analyses were conducted at the first of two consecutive clinic visit positive for autoantibodies (seroconversion, SV), just prior to development of IA (pre-seroconversion, PSV), and prior to the detection of autoantibodies at 9-months (infancy). Conditional logistic regression was used to examine the association between each metabolite (including unknowns) and being a case, controlling the false discovery rate (FDR) using the Benjamini-Hochberg multiple comparisons correction.[121] Models were adjusted for matching factors (clinical center, sex, and family history of T1D), HLA genotypes, and age. We considered several outcomes, including: persistent confirmed positivity to any autoantibody (IA), or to insulin only (IAA), GAD only (GADA), or IA2 only (IA-2A) at IA event time. IA-2A was ultimately excluded as an outcome since very few cases developed IA2 as their first and only persistent confirmed autoantibody at IA case-time. Multiple autoantibody positivity (mAb+) was defined as any subject positive

for more than one autoantibody at IA event time, or who developed more than one autoantibody during follow-up.

To better understand the relationship between nominally significant annotated metabolites, and to account for the correlation between metabolites, results from the TEDDY MWAS were used to conduct enrichment analysis using the ChemRICH tool.[122] Key metabolites identified from significant (p<0.05) chemically similar groups were used for dietary pattern analyses in Aim 2.

We also conducted a MWAS in DAISY to identify seroconversion metabolomics associated with progression from IA to T1D. The 213 IA cases selected for the DAISY – omics nested case-control study were eligible to have metabolomics, which represents all IA and T1D cases identified in DAISY as of January 2015. All metabolites were included in analyses, including those quantified from the panels used in TEDDY (primary metabolites, complex lipids), and the new panel of biogenic amines. Primary analyses were conducted using Cox regression to model the hazard of metabolite level at seroconversion and time to T1D diagnosis event, adjusting for age at seroconversion and HLA genotype. Both annotated and unknown significant metabolites were retained for the creation of dietary patterns in DAISY.

Expected Outcomes and Power

We hypothesized that both annotated, nutrition-related metabolites and unknown metabolites would be associated with IA and T1D outcomes after FDR correction. Though not named, unknown metabolites were uniquely indicated in the Fiehn laboratory database, and could therefore be interrogated for function, identity, and occurrences in other studies. Significant unknown metabolites were evaluated

using the laboratory's in-house BinVestigate tool. BinVestigate looks for the unknown compound in the Fiehn laboratory database containing 1,850 metabolomics studies representing 90,000 processed samples, and reports similar characteristics of the unknown across studies, including tissue type analyzed (stool, plasma, etc.) and outcome evaluated against.

The TEDDY nested case–control study selected 418 cases and 1253 controls, but the number of pairs for metabolomics analyses was reduced to 1002 due to sample availability. The study had 80% or greater power at a significance level of 5% to detect ≥2.01 relative risk (RR) with 1002 pairs if the proportion of exposure was 5%.[116] Bonferroni adjustment for ~1,000 tests (alpha = 0.00005) will allow us to detect RR ≥ 2.92. DAISY had 80% or greater power at a significance level of 5% to detect ≥ 2.61 RR with 213 pairs if the proportion of exposure was 5%. Correcting for Bonferroni adjustment for 1,000 tests, DAISY had power to detect ≥ 4.12 RR. All power analyses were conducted using SAS® software version 9.4.

Specific Aim 2: Create Dietary Patterns Reflective of Important Metabolites

<u>Statistical Analysis</u>

Dietary patterns were identified in both studies using reduced rank regression (RRR). RRR is a statistical method that uses linear combinations of predictors to explain the maximum covariance of a set of response variables.[123] For creating dietary patterns that are relevant for IA outcomes in TEDDY, food group intake was used as predictors to explain metabolites (response variables) that were identified from Aim 1 as significantly associated with development of IA outcomes in TEDDY. All food subgroups that were shown to be comparable across countries based on the TEDDY food

harmonization study, and survived the pre-processing pipeline were used in the analysis. Food groups were standardized to the full 9-month TEDDY population for RRR analyses. Metabolites were standardized by all metabolomics in the TEDDY nested case-control study.

We conducted a similar RRR analysis in DAISY to create nutrient patterns of intake that explained the most variation in the significant metabolites identified in Aim 1 in DAISY. Metabolites were used as RRR response variables. Due to smaller sample size in DAISY, nutrients used as predictor variables in RRR were selected after a multi-step filtering process. First, only nutrients consistently obtained from questionnaires over the entire course of the DAISY study were considered for inclusion in nutrient pattern analysis (n=38). Then, using stepwise selection, only nutrients that were significant predictors (p<0.1) for at least 2 of the metabolite response variables were included in RRR. Nutrients and metabolites were standardized to the 132 IA cases with both measures at seroconversion.

All RRR analyses were conducted using SAS® software version 9.4. In RRR analysis, the maximum number of dietary patterns identified, or the rank of the RRR model, is equal to the number of response variables included. The number of dietary or nutrient patterns needed to best explain the variation in metabolites was chosen by the van der Voet T^2 statistic, which indicates the least number of extracted factors whose residuals are not significantly greater than those of the model with minimum error.[124] Alternately, the number of dietary/nutrient patterns can be chosen by minimizing the predicted residual sum of squares (PRESS). The factors, or dietary/nutrient patterns, created as output of the RRR procedure consist of relative weights, called factor

loadings, given to food groups or nutrients. The factor loadings of food groups/nutrient

on each dietary/nutrient pattern and the partial correlation with metabolite response

variables were used to interpret each pattern. The goodness-of-fit of the

dietary/nutrient patterns was evaluated by examining the percentage of metabolite

variation explained overall and by each pattern.

<u>Expected Outcomes and Power</u>

This aim produced the factor structure and description of dietary patterns that

explain variation in IA-associated metabolites (TEDDY), and nutrient patterns that

explained variation in T1D-associated metabolites (DAISY). We hypothesized that RRR

analyses would identify at least one metabolite-related dietary pattern in TEDDY and at

least one metabolite-related nutrient pattern in DAISY. We expected the total amount of

metabolite variation explained by the patterns to range from 3-8%, typical of other RRR

analyses using biomarkers or metabolites as response variables and dietary intake as

predictor variables.[107,125,126] Factors loadings for food groups or nutrients on the

pattern were also expected to fall within previous published ranges, generally less than

0.4.[127,128]

Power calculations for RRR depend on the degrees of freedom in the model.

Unlike other covariance modeling strategies, such as principal components regression

or structural equation modeling, RRR degrees of freedom depend on the response

variable matrix, and methods to calculate them are still in development.[129] However,

compared to other studies successfully using RRR, such as European Prospective

Investigation into Cancer and Nutrition (EPIC) (n=402 to 2,267),[123,125] Western

Australian Pregnancy (Raine) Cohort Study (n=783),[130] Framingham Offspring Study

(n=2879),[131] and UK Avon Longitudinal Study of Parents and Children (ALSPAC) (n=6106 to 7285),[132] we have comparable sample size of 1,002 pairs in the TEDDY study. Though we are much less powered in the DAISY study, the filtering of predictor nutrients for inclusion in RRR improved our ability to identify metabolite-related nutrient patterns.[133]

Specific Aim 3: Test the Association of Dietary Patterns with Development of IA/T1D

Statistical Analysis

Reported food group intake for each TEDDY participant was scored by the RRR procedure based on each pattern identified in Aim 2. The score for each pattern represented how similar their dietary intake was to the dietary pattern. Multiple dietary patterns were produced as a result of the RRR procedure, and could be subsequently modeled together because they were statistically independent.

Dietary pattern scores were first tested in the nested case-control study using conditional logistic regression to account for matching strata, and adjusted for high-risk HLA genotype and age at blood draw. Then we used Cox regression to test the association of the dietary pattern score in infancy for the full TEDDY cohort with risk of mAb+ adjusting for clinical center, high-risk HLA genotype, family history of type 1 diabetes, total energy intake, and sex. Time-to-event analyses were performed to evaluate whether metabolite-related dietary pattern scores were associated with mAb+. Cases included those selected for the nested case-control study plus any additional cases that developed in TEDDY by January 2018.

In DAISY, subjects received a score reflecting how similar their nutrient intake at seroconversion was to the nutrient pattern, using the RRR procedure. The association of the nutrient pattern score and progression to T1D was tested using Cox regression models adjusted for age at seroconversion and high-risk HLA genotype.

Expected Outcomes and Power

Given that the dietary and nutrient patterns explain variation in outcome-related metabolites, we hypothesized that any patterns identified in Aim 2 would be associated with outcomes in Aim 3. The direction of association is expected to be consistent with the hypothesized direction of dietary factors loading highly onto the pattern. For example, a dietary pattern characterized by intake high in foods containing omega-3 fatty acids (which is thought to be protective for IA or T1D risk)[80] would be inversely associated with IA or T1D risk.

Using Cox proportional hazards, a two-sided log rank test in a sample size of 8,676 TEDDY subjects with an overall IA rate of 9% in 5 years achieves 80% power at a 0.05 significance level to detect a hazard ratio (HR) > 1.23 assuming the proportion of children lost to follow-up is 20% (**Table III-4**). In a similar Cox proportional hazards framework, a two-sided log rank test in a sample size of 213 DAISY IA cases with an overall T1D rate of 40% in 5 years achieves 80% power at a 0.05 significance level to detect a hazard ratio (HR) > 1.75 assuming the proportion of children lost to follow-up is 20%. HRs are based on a one standard deviation change in dietary or nutrient pattern score. We have good power to detect dietary pattern associations with IA in TEDDY, and less but still adequate power in DAISY.

Table III-4: Detectable HR with various power for Aim 3.

	TEDDY* (n=8,676)	DAISY** (n=213)
Power	HR	HR
70%	1.2	1.65
80%	1.23	1.75
90%	1.26	1.90

*Assuming 9% IA rate and 20% loss to follow-up in 5 years
**Assuming 40% T1D rate and 20% loss to follow-up in 5 years

Potential Pitfalls and Alternative Approaches

Given the exploratory nature of our Aims, and the unique strength of having access to two similarly conducted prospective studies of T1D, there were many alternative approaches considered for identifying combination of dietary factors associated with IA and T1D outcomes. Ultimately, we selected the most statistically sound and interpretable approaches to account for the complex nature (high-dimensional, correlated) of metabolomics and dietary intake, and the two different nested study designs.

Alternative statistical frameworks commonly used for metabolomics analyses were considered, but ultimately deemed inappropriate due to the longitudinal, matched study design with multiple adjustment factors. Unsupervised models used for high-dimensional data, such as principal components analysis (PCA)[134], K-means clustering[135], and self-organizing maps (SOM)[136], are useful for exploration and visualization of metabolomics data that might inform further statistical analyses. However, they cannot account for the correlation of repeated measures on study subjects, do not perform hypothesis testing, and cannot incorporate an adjustment for confounding factors. While components identified in PCA can be applied longitudinally

to repeated measures as a reduced, composite measure of metabolite exposure, the components summarizing the most variability in the data (the first ones, which are typically used in this manner) may not reflect disease-causing metabolites. Supervised learning methods, such as partial least squares (PLS, PLS-DA)[137], identify metabolites that discriminate between outcomes (case versus control), however it is difficult to incorporate confounding variables and to account for the matching strata. Selecting cross-sections of interest in the natural history of T1D allowed us to account for matching strata from the matched case-control design in TEDDY and to adjust for confounding factors in both studies.

Principal components analysis (PCA), partial least squares, cluster analysis, and RRR are all common dimension reduction methods for combining multiple food or nutrient exposures in a way that accounts for their complex correlation. Multiple studies assessing differences in the methods as applied to nutrition studies conclude that RRR is most appropriate for investigating disease etiology or pathophysiology, as it incorporates both *a posteriori* (food group intake) and *a priori* (selection of metabolites) information to create dietary patterns.[105,123,127,138,139]

However, the RRR procedure cannot account for repeated measures or for confounding variables—which could cause the dietary patterns created in Aim 2 to be confounded by the structural differences in dietary intake by age. For example, food groups eaten at PSV for a child aged 1 year may be very different than for a child aged 5 years, and this variability may have undue influence in the creation of the dietary pattern. The inability to look at dietary pattern structural stability longitudinally is a weakness shared by many data-driven dietary pattern methods, including PCA, factor

analysis, and RRR. Attempts to examine whether dietary pattern structures (loadings on the food groups) are stable with age have been limited to comparing loadings at two cross-sectional time points.[140–143]

In a pilot study, we applied a new method called three-way PCA which created dietary patterns on intake collected at multiple time points in healthy DAISY subjects, and described how childhood dietary pattern structures may change over time. Results of the analysis indicate that three childhood dietary patterns (indicative of "food pyramid", "prudent/western", and "allergy avoidance" diets) are relatively stable from ages 1 to 8 years, and a fourth dietary pattern characterized by high intake of fruit, fruit juices, and dairy is evident in the first two years of life (unpublished data). Given that cross-sections selected based on the natural history of disease (i.e., pre-seroconversion, seroconversion) can include many ages, we might expect dietary pattern structures to be confounded by age. Findings from the pilot study motivated our decision to restrict at least one cross-section to a single age of subjects (TEDDY, infancy cross-section).

Availability of and access to both the TEDDY and DAISY studies was a unique opportunity for replication. As with the era of large-scale candidate gene studies, metabolomics studies are producing many candidate biomarkers that may or may not be able to be replicated in independent populations, though replication is universally recognized as important.[144,145] The decision to organize manuscripts by study necessitated the investigation of different disease endpoints to provide novel contributions to the literature, precluding the ability to replicate as part of the dissertation.   Differences between the studies were also great enough regarding the storage and quantification of metabolomics and the assessment tools for dietary intake,

that the ability to replicate true associations (if they existed) was questionable. For example, assessing dietary intake with food records (TEDDY) over three 24-hour periods tends to capture only commonly consumed foods compared to the FFQ (DAISY), which may capture rarely consumed foods but tends to overestimate dietary intake. A comparison of food groupings available in DAISY and TEDDY highlights differences in the food groups available (**Appendix A**). Additionally, different blood collection anticoagulants were used in the two studies that would have contributed to replication failure. The TEDDY study used citrate, which has been shown to interfere with primary metabolite and lipid quantification,[146,147] while DAISY used EDTA anticoagulant. Ultimately, replication did not seem feasible for discovering combinations of dietary intake associated with IA and T1D. However, a targeted replication could be considered as a next step for the dietary and nutrient patterns identified herein.

CHAPTER IV

METABOLITE-RELATED DIETARY PATTERNS AND THE DEVELOPMENT OF ISLET AUTOIMMUNITY

Abstract

The role of diet in type 1 diabetes development is poorly understood. Metabolites, which reflect dietary response, may help elucidate this role. We explored metabolomics and lipidomics differences between 352 cases of islet autoimmunity (IA) and controls in the TEDDY (The Environmental Determinants of Diabetes in the Young) study. We created dietary patterns reflecting pre-IA metabolite differences between groups and examined their association with IA. Secondary outcomes included IAA or GADA as first-appearing autoantibody, or multiple autoantibodies (mAb+). The association of 853 plasma metabolites with outcomes was tested at seroconversion to IA, just prior to seroconversion, and during infancy. Key compounds in enriched metabolite sets were used to create dietary patterns reflecting metabolite composition, which were then tested for association with outcomes in the nested case-control subset and the full TEDDY cohort. Unsaturated phosphatidylcholines, sphingomyelins, phosphatidylethanolamines, glucosylceramides, and phospholipid ethers in infancy were inversely associated with mAb+ risk, while dicarboxylic acids were associated with an increased risk. An infancy dietary pattern representing higher levels of unsaturated phosphatidylcholines and phospholipid ethers, and lower sphingomyelins was protective for mAb+ in the nested case-control study only. The risk for mAb+ was associated with metabolomics factors related to differences in lipid-related food intake.

Introduction

Type 1 diabetes affects over 500,000 children globally, making it one of the most common metabolic illnesses in children.[28] Autoimmune destruction of the insulin-producing beta-cells in the pancreas results in hyperglycemia and lifelong insulin dependency. Genetic risk factors are well described and likely interact with non-genetic risk factors to influence disease progression, though exact pathogenesis remains unclear.[148] The appearance of autoantibodies can be detected as early as 3 months of age and defines the beginning of islet autoimmunity (IA), the preclinical stage of the disease.[10] Efforts to better characterize metabolic dysregulation around the time of seroconversion and prior to the detection of autoantibodies may allow earlier identification of at-risk children and better understanding of the processes involved.

Metabolites reflect the interaction of numerous biological factors, including many that may influence the development of autoimmune diabetes, such as genetics, microbiome, and dietary intake. Metabolomics differences between IA cases and controls mostly have been found at the time of seroconversion, but are inconsistent across studies conducted in country-specific populations.[95,101,102,104] Previous studies indicated that metabolic disturbances exist prior to autoimmunity, and as early as birth, in Finnish children who later progressed to type 1 diabetes.[96,101,103] The identification of generalizable metabolic profiles related to the development of early stages of the disease may inform dietary interventions to prevent type 1 diabetes, which have so far proven unsuccessful.[36]

Traditional investigations of diet in the development of type 1 diabetes has examined effects of individual foods or food groups and nutrients such as cow's

milk,[36,77,78] fatty acids,[79–81,149,150] or vitamin D.[151,85,152,86] However, these approaches do not account for the complexity of the diet—the effects of single nutrients and foods are often too small to identify, or too highly correlated to be separated from each other.[100] Examining combinations of foods and metabolites may better elucidate the role of diet in IA, as it can account for synergistic or antagonistic effects of foods or nutrients contained in the diet, and differences in how they are processed in the body.

We aimed to identify metabolite-related dietary patterns associated with IA in the multinational The Environmental Determinants of Diabetes in the Young (TEDDY) study (**Figure IV-1**). We conducted a metabolome- and lipidome-wide association study to better characterize plasma metabolites and lipids distinguishing cases and controls both at the time of the first autoantibody detection, and prior to its development. We created dietary patterns summarizing candidate metabolites identified pre-IA, and tested the longitudinal association of those metabolite-related dietary patterns with the development of IA.
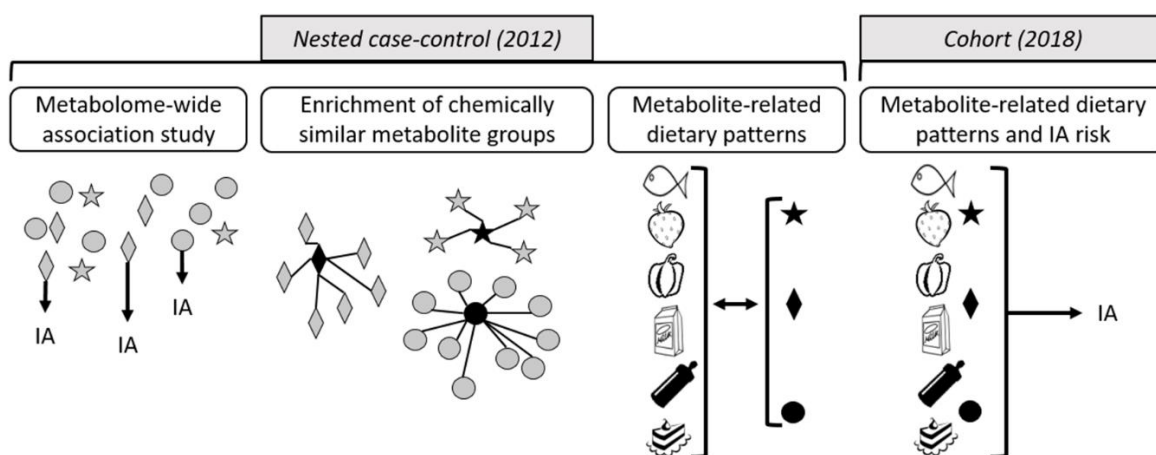
Figure IV-1: Study design for identifying metabolite-related dietary patterns associated with the development of IA in TEDDY.

Research Design and Methods

TEDDY Study Design

TEDDY is an international consortium that enrolled 8,676 newborn infants with a high- or moderate-risk class II HLA genotype between 2004 and 2010.[110] Participants are closely followed for the development of IA or type 1 diabetes, with study visits every three months from birth to age 48 months, and every three or six months thereafter depending on autoantibody status until age of 15 years. Participating study centers include: Georgia/Florida, Colorado, and Washington in the U.S., and Finland, Sweden, and Germany in Europe. IA cases are defined by confirmed autoantibody positivity to either insulin (IAA), GAD (GADA), or IA-2 (IA-2A) on two consecutive study samples, the first of which defines the case's event age. Institutional review board approval for the study and written informed consents were obtained for all participating children.

A nested case-control biomarker study was designed using risk set sampling to select three controls per IA case (n=418) that had developed in TEDDY as of May 2012. Eligible controls were autoantibody-negative at the case's event age, and further matched on clinical center, sex, and family history of type 1 diabetes as previously described.[116] Secondary outcomes included persistent confirmed positivity to insulin only (IAA), GAD only (GADA), or IA2 only (IA-2A) at IA event time. IA-2A was excluded as an outcome since very few cases developed IA2 as their first and only persistent confirmed autoantibody at IA case-time. Multiple autoantibody positivity (mAb+) was defined as any subject positive for more than one autoantibody at IA event time, or who developed more than one autoantibody during follow-up.

Metabolomics Data Pre-Processing

Metabolomics abundance measures (metabolites and lipids) were obtained for all cases and controls for each available study visit from birth until the case event time. Primary metabolites and complex lipids were quantified from citrate plasma using GC-TOF MS and CSH-QTOF MS data acquisition, respectively, at the NIH West Coast Metabolomics Center at the University of California, Davis. GC-TOF MS data were acquired as previously described,[153] with data processing and compound identification using the BinBase algorithm.[154] GC-TOF data were sum normalized followed by LOESS (locally weighted scatterplot smoothing) normalization. For complex lipids, samples were extracted by methyl-tert-butyl ether/methanol/water,[155] followed by chromatogram peak detection and alignments using Mass Profiler Professional (Agilent, Santa Clara, CA). Peaks detected in a minimum of 30% of samples were identified and quantification back-filled using the Fiehn laboratory's LipidBlast spectral library, as previously described.[120] LOESS followed by batch ratio (QC samples were used to adjust sample batch median to global study median) normalization was performed across all the samples to estimate and remove analytical variance.

Prior to transformation, data quality checks included evaluation at the metabolite- and sample-level. Metabolites that were not detected in more than 10% of samples (6 metabolites), or with a coefficient of variation greater than or equal to 100% (286 metabolites) were excluded from further analyses. Samples with missing or zero values in greater than 10% of metabolomics features (n=5) or with values more extreme than 4 standard deviations above or below the mean in greater than 30% of metabolomics features (n=6) were removed from analyses. A total of 853 metabolites

and lipids and 11,556 samples passed the quality checks. All metabolites were transformed using Box-Cox transformation analysis, and scaled.[156]

Dietary Intake and Food Groupings

Dietary assessment was carried out by 24-hour recall at the first clinic visit at 3-4.5 months of age, then by 3-day food record every 3 months until 12 months of age, and then every 6 months thereafter. TEDDY research staff provided detailed instruction and examples to families regarding completion of food records, as previously described.[84] From quantities of foods and dishes consumed, the amounts of energy and single foods contained therein were computed using in-house food record processing programs and food composition databases unique to each country.[113] The foods and dishes (e.g., wheat bread, apple-oat meal) consumed were quantified into main food groups (e.g., cereals, fruits and berries, etc.) and subgroups (e.g., wheat, rice, oats, citrus fruits, apple, berries, etc.) in grams per day (g/day) of intake. After quantification, the three food records were averaged to calculate the mean energy and food intake for each study subject on each study visit. Results of detailed harmonization studies of these country-specific food composition databases documented that the energy values, 14 main food groups, and 85 food subgroups used in this study were comparable across the TEDDY countries.[113]

For any food record where a subject was indicated as breast fed, we estimated the amount of breastmilk consumption using an algorithm developed by the Institute of Medicine.[157] First, we calculated the estimated energy requirement based on age and weight. The difference in the estimated energy requirement and the mean energy reported on the food record from food and formula was attributed to breastmilk. We

calculated the amount (grams) of breastmilk consumed to achieve that energy intake using a conversion factor of energy density per 100g, as follows: 65.3 kcal/100g in Finland, 69 kcal/100g in Germany, 68 kcal/100g in Sweden, and 70 kcal/100g in the U.S.

Statistical Analyses

Figure IV-2 summarizes the population and data flow for all aims and analyses.

Metabolites Associated with IA

Conditional logistic regression was used to calculate odds ratios (ORs) for the association of each transformed metabolite with the development of IA, adjusting for high-risk HLA genotype (DR3-DQA1*05:01-DQB1*02:01/DR4-DQA1*03:01-DQB1*03:02 versus all other), and age at blood draw. Some TEDDY subjects follow a long-distance protocol, in which blood is drawn and shipped to clinical centers before being processed for biomarker identification. Since plasma primary GC-TOF MS metabolic profiles are less stable with centrifugation delay,[158] we required an additional match for the long distance protocol between case and control samples.

Metabolomics analyses were run in three cross-sections. First, we examined metabolite differences between IA cases and controls at the first detection of autoantibody positivity (seroconversion), defined as the first of the two consecutive autoantibody positive visits for cases. Then, to identify metabolites and lipids that may differentiate IA cases and controls prior to the detection of autoantibodies (pre-seroconversion) we selected the most recent IA-free visit for cases. Finally, since autoimmunity can begin very early in life and both metabolomics and dietary factors are strongly related to age, we identified metabolites distinguishing IA cases and
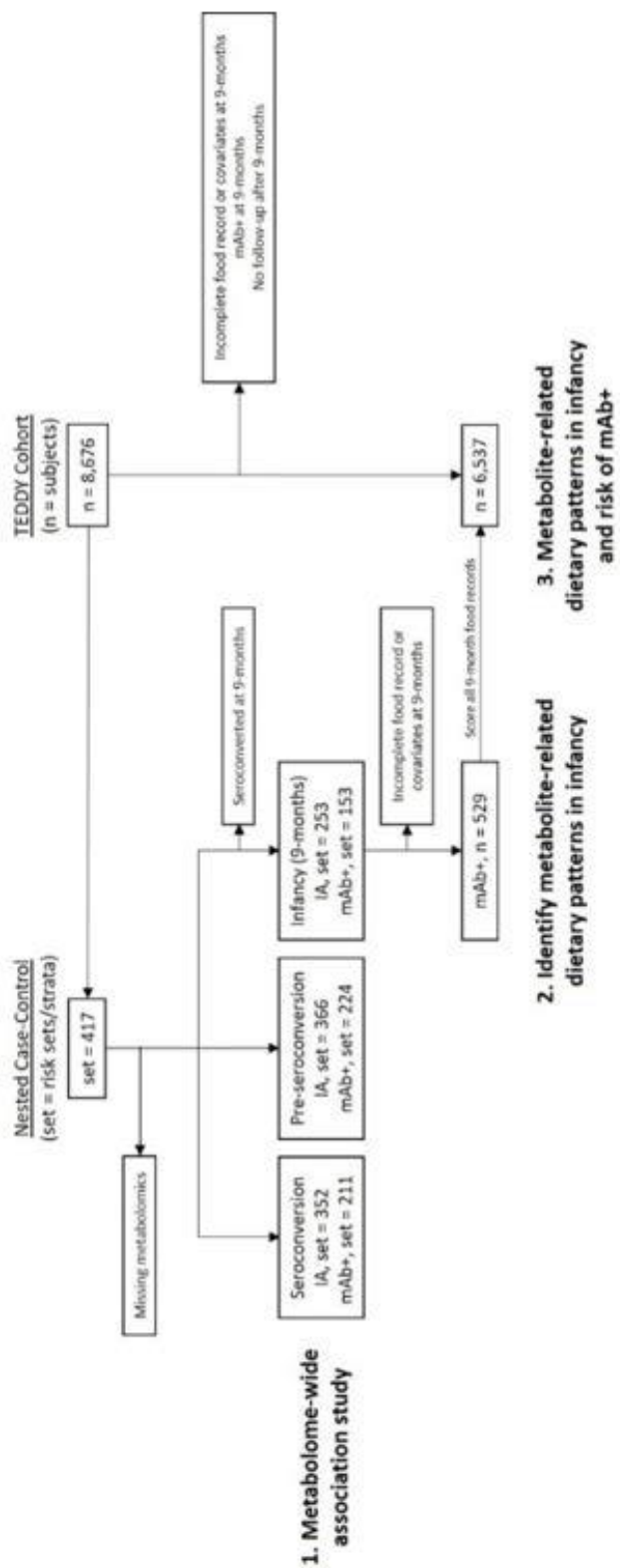
Figure IV-2: Data flow diagram summarizing selection of analysis population for all aims (numbered). For the nested case-control study, the number of informative risk sets (1 cases and at least 1 control) are shown. The number of subjects are identified for non-matched analyses. IA=islet autoimmunity. mAb+=multiple autoantibody positive.

controls prior to the appearance of autoantibodies in infancy. The "infancy" cross-section was defined as an IA-free visit at 9 months of age for cases. For controls, the visit corresponding to the case visit was selected for all cross-sections. Children positive for autoantibodies at 9-months (n=48 IA cases) and their matched controls were excluded from the infancy cross-section (**Figure IV-2**).

We tested the association of each metabolite with the secondary outcomes described above: IAA, GADA, and mAb+. We considered p-value<0.05 significant since traditional approaches for multiple comparison correction may be too strict for the unusually highly correlated metabolomics data or inappropriate given the exploratory nature of our study aims.[159] SAS version 9.4 was used for these analyses.

We also focused on pathway enrichment, given that metabolites may capture perturbations in many upstream biological systems thereby complicating interpretation of individual associations. Using the ChemRICH metabolomics set enrichment tool,[122] we tested whether groups of structurally similar metabolites were associated with any autoantibody outcome. ChemRICH forms non-overlapping groups of metabolites based on chemical similarity and ontology mapping.[122] It calculated a single p-value for each group, and identified the most significant metabolite in each group as the "key compound" (http://chemrich.fiehnlab.ucdavis.edu/). Inputs for the ChemRICH analyses included the nominal p-value and odds ratio from the individual conditional logistic regression models, and chemical structure information from well-characterized annotated metabolites and lipids (p=315, **Appendix B**). Metabolite groups with p-value<0.05 were considered significant.

Metabolite-Related Dietary Patterns Preceding the Appearance of Autoantibodies

Reduced rank regression (RRR) was used to identify dietary patterns reflecting metabolites associated with IA. RRR creates linear combinations of foods (dietary patterns) that explain the maximum covariation in a second set of intermediate response variables (metabolites),[123] thereby capturing disease-related variation in the diet rather than general eating behaviors identified from other dietary pattern methods.[139] We focused dietary pattern analysis on the infancy cross-section, since it is prior to the beginning of the autoimmune process and all children were the same age.

Food groups from the diet record were combined into 43 subgroups based on nutrient content and culinary usage.[113] Many foods had a large proportion of subjects with no reported intake, given that food records only capture dietary intake on three days within each study visit window and that many foods may not be introduced yet for visits occurring during infancy. Therefore we applied a filtering criteria for inclusion of a food subgroup into the dietary pattern analysis, as is common in studies using RRR.[133] Food subgroups that were shown to be comparable across TEDDY countries, and were eaten by at least 40% of subjects in infancy were included in the creation of dietary patterns, since highly zero-inflated foods might perform sub-optimally. Food subgroups were standardized to the age-specific mean and standard deviation of all TEDDY food records for dietary pattern analyses. The key compound in each significantly enriched metabolite group (identified by ChemRICH) was used as RRR response variables.

The number of dietary patterns needed to best explain the variation in metabolites was selected using the van der Voet T2 statistic.[124] The loadings (or relative

weights) of food groups on each dietary pattern and the partial correlation with

metabolite response variables was used to interpret each dietary pattern.

Metabolite-Related Dietary Patterns and Risk of IA

Infancy metabolite-related dietary patterns were first tested in the nested case-

control study using conditional logistic regression as described above. Then we applied

them to the full TEDDY cohort at 9 months of age, using the food group loadings to

generate one score per pattern for each subject with complete food records (n=6,845).

The score indicates how similar the reported dietary intake is to the dietary pattern.

Multiple dietary patterns were produced as a result of the RRR procedure and could be

subsequently modeled together because they are statistically independent.

Cox proportional-hazards models were used to test the association of

metabolite-related dietary pattern scores at 9-months on risk of mAb+, adjusting for

clinical center, high-risk HLA genotype, family history of type 1 diabetes, total energy

intake, and sex. Time-to-event analyses were performed to evaluate whether

metabolite-related dietary pattern scores were associated with mAb+ by the age of 6

years. Cases included those selected for the nested case-control study plus any

additional cases that developed by January 2018. Given that risk factors for IA may

differ by age, we restricted follow up to 6 years in the TEDDY cohort to ensure the

cohort analysis represented a similarly-aged case-population as the nested case-control

study. For consistency with the nested case-control study, the time-to-event was

defined as the time from birth to the appearance of the first persistent confirmed

autoantibody among IA cases who developed a second persistent confirmed

autoantibody at any point. Subjects without complete covariate information, those

developing mAb+ at 9-months (n=53), or who had no follow-up after 9-months (n=243) were excluded from survival analysis (**Figure IV-2**).

Results

<u>Metabolic Dysregulation Apparent at Seroconversion and Infancy</u>

From the nested case-control study, there were 352 matched sets with metabolomics measures for 1 case and at least 1 control at seroconversion (mean (SD) case-age=722 (446) days), 366 sets at pre-seroconversion (mean (SD)=625 (412) days), and 253 sets at the 9-month infancy visit (mean (SD)=283 (14) days) (**Table IV-1**). For secondary outcomes, 49% of IA cases were positive for IAA, 32% for GADA, and 60% for mAb+. The distribution of secondary outcomes was consistent across the seroconversion, pre-seroconversion, and 9-month infancy cross-sections. The majority of IA cases were from Sweden (32%) and Finland (28%). IA cases were 55% male, 22% had a first degree relative, and 12% had their seroconversion blood-draw following TEDDY's long distance protocol (**Table IV-2**). The distribution of matched sets by matching factors (clinical center, sex, first-degree relative status, and long distance protocol) were similar in the secondary outcomes compared to primary IA. Approximately 9% of the pre-seroconversion case samples (n=34) occurred during the 9-month infancy visit.

Table IV-1: Description of matched sets (1 case and 1, 2 or 3 controls) for metabolomics analyses by outcome and cross-section.

	IA		IAA*			GADA*			mAb+*		
Cross-section	**N**	**Case-age†, mean (SD)**	**n**	**% of IA cases**	**Case-age, mean (SD)**	**n**	**% of IA cases**	**Case-age, mean (SD)**	**n**	**% of IA cases**	**Case-age, mean (SD)**
Seroconversion	352	722 (446)	171	48.6	586 (370)	113	32.1	888 (509)	211	59.9	655 (365)
Pre-Seroconversion	366	625 (412)	180	49.2	505 (366)	116	31.7	786 (445)	224	61.2	541 (346)
Infancy 9-months	253	283 (14)	114	45.1	283 (14)	83	32.8	284 (16)	153	60.5	282 (14)

Secondary outcomes defined as IAA or GADA as first-appearing and only autoantibody at IA case-time (mutually exclusive), while mAb+ indicates IA case developed more than 1 persistent confirmed Ab at any point during follow-up.

†*Age at the time of metabolomics blood draw, in days*

Table IV-2: Characteristics and outcomes of TEDDY nested case-control 1:3 matched study at seroconversion.

	Outcomes			
	IA	**IAA**	**GADA**	**mAb+**
Matching Characteristic	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Clinical Center				
Colorado	51 (14.5)	22 (12.9)	17 (15.0)	30 (14.2)
Georgia	28 (8.0)	8 (4.7)	12 (10.6)	18 (8.5)
Washington	32 (9.1)	16 (9.4)	14 (12.4)	15 (7.1)
Finland	100 (28.4)	59 (34.5)	23 (20.4)	72 (34.1)
Germany	28 (8.0)	11 (6.4)	9 (8.0)	16 (7.6)
Sweden	113 (32.1)	55 (32.2)	38 (33.6)	60 (28.4)
Sex				
Female	158 (44.9)	71 (41.5)	56 (49.6)	91 (43.1)
Male	194 (55.1)	100 (58.5)	57 (50.4)	120 (56.9)
FDR/GP Status				
First Degree Relative	79 (22.4)	37 (21.6)	27 (23.9)	53 (25.1)
General Population	273 (77.6)	134 (78.4)	86 (76.1)	158 (74.9)
Long Distance Protocol				
Yes	43 (12.2)	17 (9.9)	16 (14.2)	24 (11.4)
No	309 (87.8)	154 (90.1)	97 (85.8)	187 (88.6)

Conditional logistic regression results from the metabolome-wide association study indicated metabolic dysregulation in cases compared to matched controls at seroconversion and during infancy (**Table IV-3, Figure IV-3**). Over 6% (p=54) of metabolites and lipids were significantly different at p<0.05 among IA cases and matched controls at seroconversion, and 7.5% (p=64) in infancy. More metabolites and lipids were different by case status when restricting to the mAb+ outcome—10% (p=91) were different between mAb+ and controls at seroconversion, and 15% (p=130) in infancy. There were few metabolites associated with the secondary outcomes IAA first and GADA first in any cross-section (**Figure IV-4**).

Table IV-3: Summary of individual metabolites significantly associated (p<0.05) with diabetes autoimmunity outcomes in three cross-sections*.

		Outcomes					
		IA			mAb+		
		All		OR>1	OR<1	All		OR>1	OR<1
Cross-section	p	p	%	p	p	p	%	p	p
Seroconversion*	853	54	6.3	21	33	91	10.7	17	74
Pre-seroconversion	853	27	3.2	8	19	19	2.2	11	8
Infancy 9-months	853	64	7.5	21	43	130	15.2	20	110

Conditional logistic regression models adjusted for HLA (DR3/4) and age at blood draw date
**Visit at which an autoantibody (or autoantibodies) was first detected*
p=number of metabolites

We focused the metabolomics set enrichment analyses on the mAb+ outcome, which represented 60% of IA cases and had the largest signal of metabolomics differences between cases and controls. ChemRICH identified seven groups of chemically similar metabolites that were significantly different among mAb+ cases and controls at seroconversion, one group at pre-seroconversion, and six groups in infancy (**Table IV-4**).

Of the metabolite groups identified as different among mAb+ cases, only unsaturated phosphatidylcholines (PC) were consistently dysregulated in all three analyses (p-value for group in seroconversion=2.1 x 10^{-5}, pre-seroconversion=0.013, infancy=2.2 x 10^{-20}), with the majority of the individual metabolites being lower in mAb+ cases compared with controls (OR<1) (**Figure IV-5**). Similarly, phosphatidylethanolamines (PE) were lower in mAb+ cases (OR<1) at both seroconversion (p-value=0.0047) and in infancy (p-value=2.9 x 10^{-6}).

Other than PCs and PEs, distinct metabolite groups distinguished mAb+ cases from their controls at the time of seroconversion to primary IA compared to during infancy prior to the appearance of any autoantibodies (**Table IV-4**, **Figure IV-5**). At seroconversion, mAb+ cases had lower levels of unsaturated triglycerides (p-value=3.1 x 10^{-15}), amino acids (p-value=0.0074), diglycerides (p-value=0.014), and aromatic amino acids (p-value=0.03), and higher levels of saturated fatty acids (p-value=0.0074). In infancy, other phospholipids were significantly protective for mAb+ (majority of OR<1), including sphingomyelins (SM, p-value=1.1 x 10^{-8}) and phospholipid ethers (EtherPL, p-value=0.0032), along with the glucosylceramides (GlcCer, p-value=4.4 x 10^{-5}). Three dicarboxylic acids were significantly higher in mAb+ cases compared to controls in infancy (OR>1, p-group=0.0019).
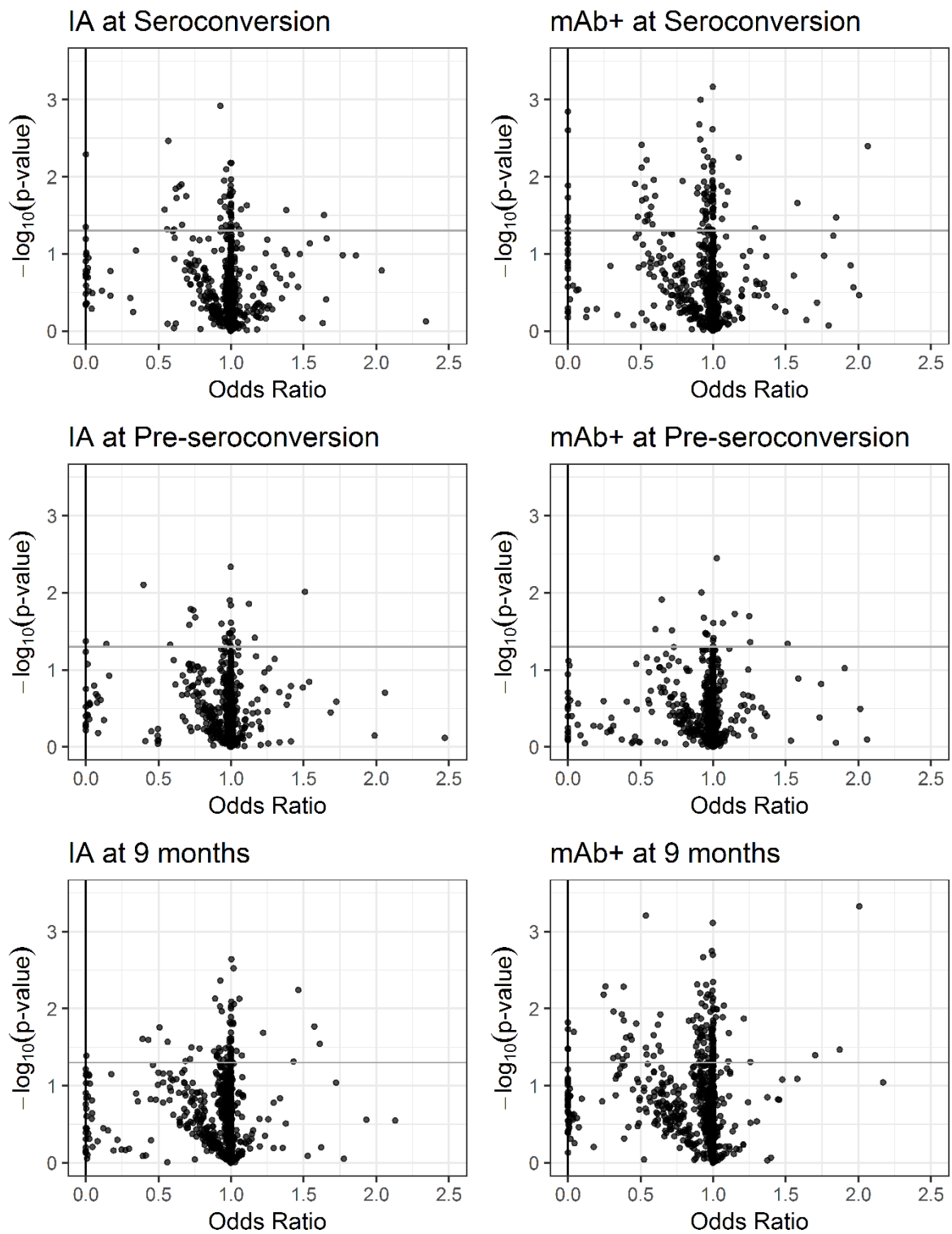
Figure IV-3: Volcano plot of the association between 853 metabolites and islet autoimmunity (IA), multiple autoantibody positivity (mAb+) in the TEDDY study.
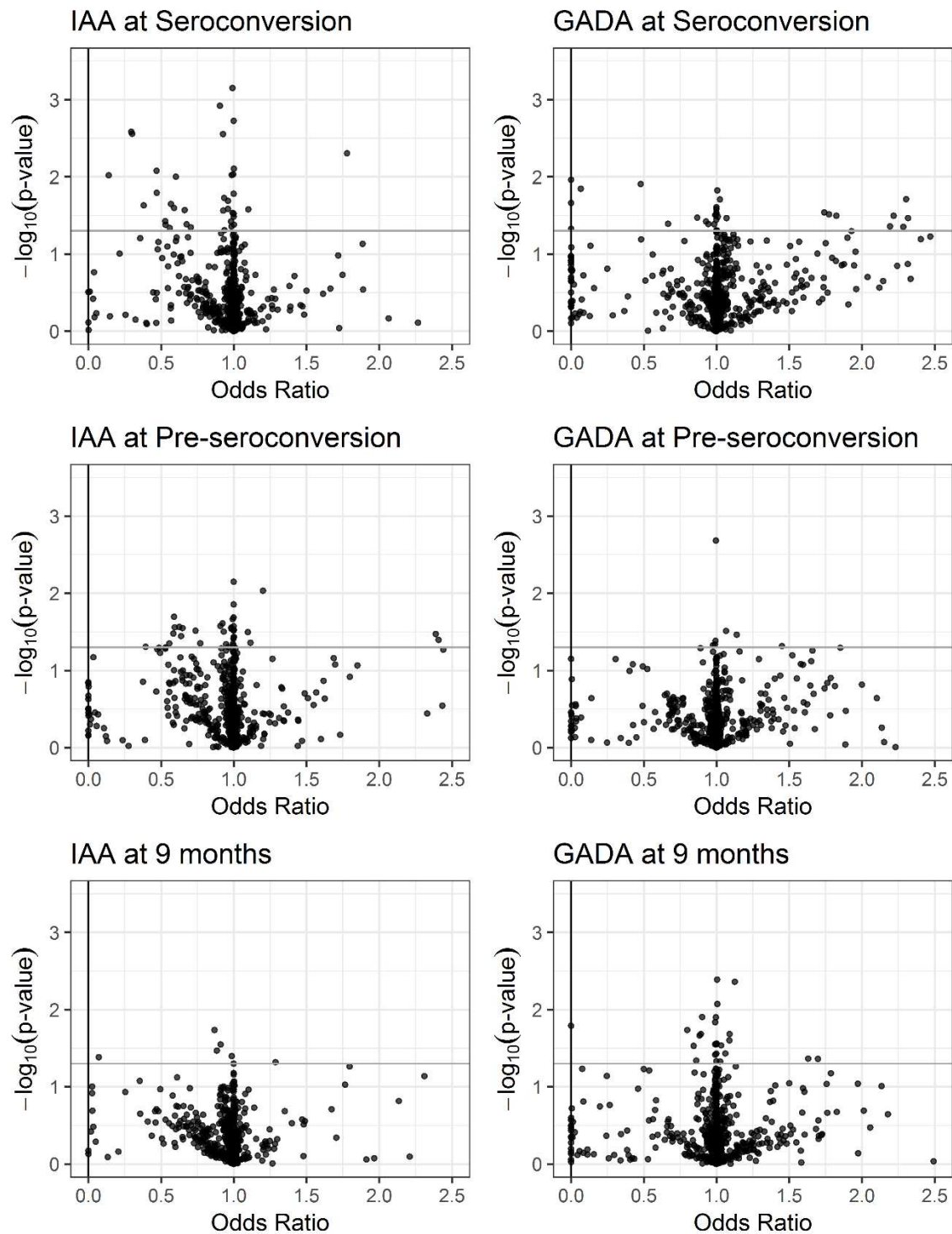
Figure IV-4: Volcano plot of the association between 853 metabolites and secondary outcomes insulin-first at IA (IAA), and GAD-first at IA (GADA) in the TEDDY study.

Table IV-4: ChemRICH summary of chemically similar metabolite groups significantly associated with mAb+ (nominal p-value for group<0.05).

		Seroconversion				Pre-seroconversion				Infancy			
Metabolite Groups		Metabolite		Cluster		Metabolite		Cluster		Metabolite		Cluster	
Name	Total n	n, p<0.05	n, OR<1	p-value	q-value	n, p<0.05	n, OR<1	p-value	q-value	n, p<0.05	n, OR<1	p-value	q-value
Unsaturated Phosphatidyl-cholines	76	5	3	0.000021	0.00029	2	1	0.013	0.35	29	28	2.2E-20	6.6E-19
Sphingomyelins	16									10	10	0.000000011	0.00000016
Phosphatidyl-ethanolamines	10	3	3	0.0047	0.042					3	3	0.0000029	0.000029
Galactosyl-ceramides	4									2	2	0.000044	0.00033
Dicarboxylic Acids	5									3	0	0.0019	0.012
Phospholipid Ethers	6									2	2	0.0032	0.016
Unsaturated Triglycerides	43	17	15	3.1E-15	8.7E-14								
Amino Acids	6	2	2	0.0074	0.042								
Saturated FA	7	3	0	0.0074	0.042								
Diglycerides	11	3	3	0.014	0.064								
Amino Acids, Aromatic	4	2	2	0.03	0.12								

ChemRICH inputs: p-value and OR from conditional logistic regression models adjusted for HLA (DR34) and age at draw date
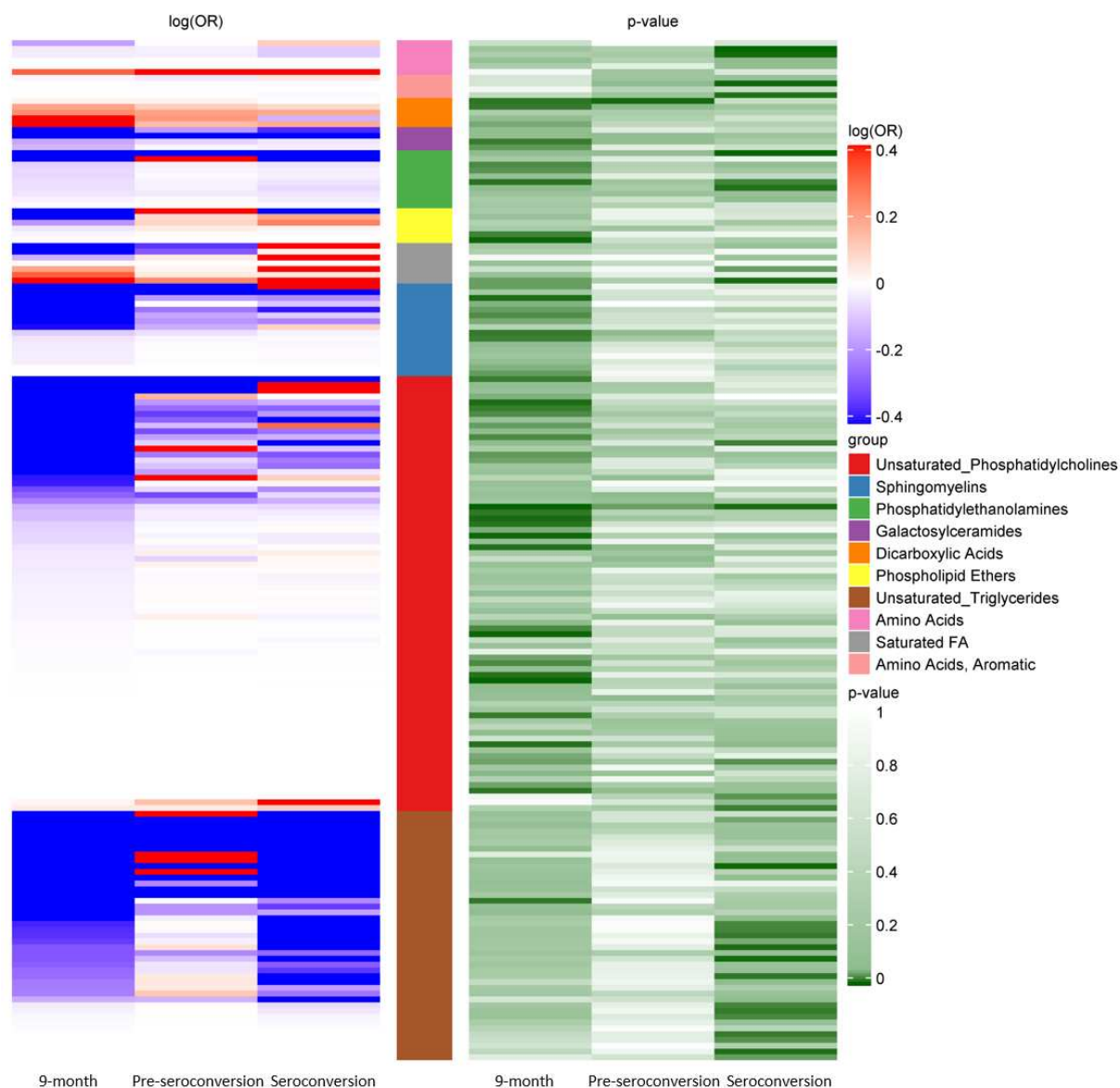
Figure IV-5: Chemically similar metabolite sets identified as significantly associated with mAb+ by ChemRICH. Each row is an individual metabolite, grouped by ChemRICH set and sorted by log(OR) within each set. Log(OR) > 0 (red) indicates a positive association between metabolite and mAb+, whereas log(OR) < 0 (blue) indicates an inverse association between metabolite and mAb+. Phosphatidylcholines were significantly lower in cases compared to controls in infancy (9-month), just prior to seroconversion, and at seroconversion to primary IA. Other phospholipids were significantly lower in cases only in infancy, while other metabolite groups, such as unsaturated triglycerides and amino acids, distinguished cases and controls at seroconversion.

<u>Infant Metabolite-Related Dietary Patterns and Risk of mAb+</u>

From each of the six groups identified by ChemRICH in infancy, we used the key metabolite (most significant one) as a response variable in dietary pattern analyses, including: PC (34:3), SM (d41:2) A, PE (34:2), GlcCer (d41:1), adipic acid, and PC (p-32:0) or PC (o-32:1) (EtherPL).

Reduced rank regression identified three dietary patterns that explained 8% of the variation in metabolites and 29.3% of the food variation. Food groups factor loadings and metabolite variable weights for each dietary pattern are shown in **Figure IV-6**. Infants scoring high on Dietary Pattern 1 ate more non-gluten containing cereals, onions, vegetable oils, and fat-free milk, and less breast milk. This diet corresponded to higher levels of PE (34:2). Whereas infants scoring high on Dietary Pattern 2 ate diets with higher saturated fats, fat-free milk, poultry, and infant formula, and lower in potatoes and vegetable oils. This diet corresponded to higher levels of SM (d41:2) A, GlcCer (d41:1), and PC (p-32:0) or PC (o-32:1). Finally, 9-month infants scoring high on Dietary Pattern 3 ate diets higher in breast milk, red meat, potatoes, and cereals, and lower in processed fruits, legumes, and infant formula. High scores on Dietary Pattern 3 corresponded to higher levels of PC (34:3) and PC (p-32:0) or PC (o-32:1), and lower levels of SM (d41:2) A.
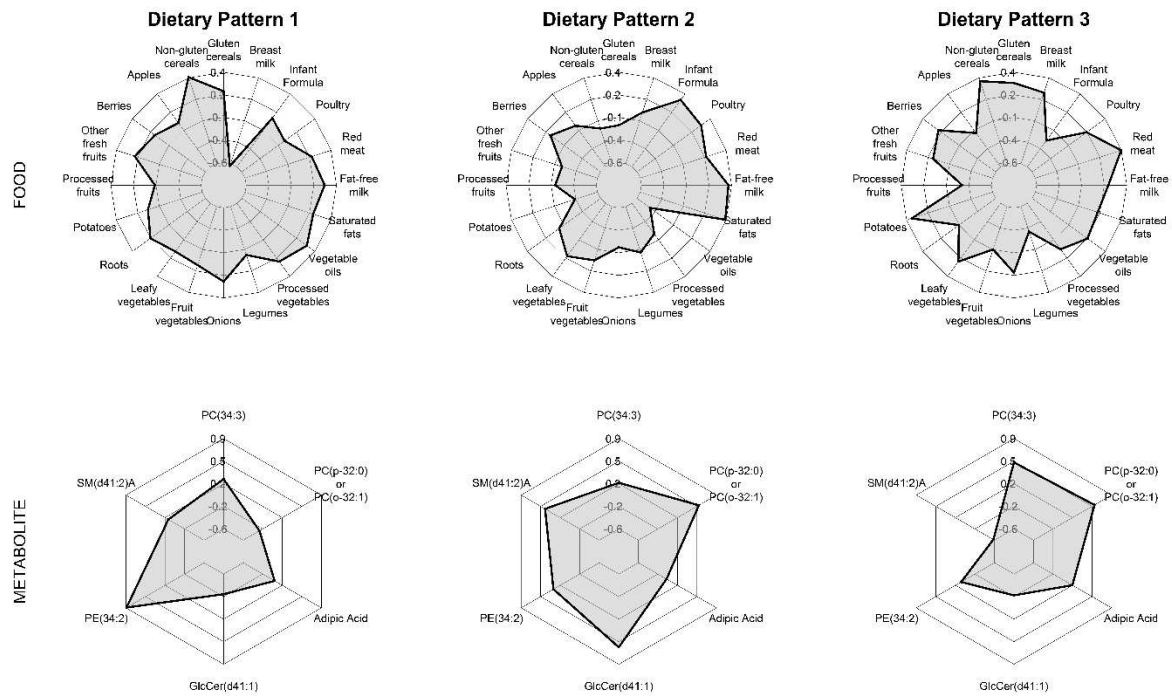
Figure IV-6: Food group loadings and metabolite weights for metabolite-related dietary patterns. In total, the three dietary patterns explained 8% of metabolite variation and 29.3% of food variation. For food, the radial axis indicates the loading on each dietary pattern (Range: -0.6 to 0.4), and is used to interpret which combinations of foods are influential in the dietary pattern. Similarly, the metabolite radial axis indicates the weight of each metabolite on each dietary pattern (Range: -0.6 to 0.9), indicating which combination of metabolites are explained by each dietary pattern. For example, subjects scoring high on dietary pattern 1 had diets higher in non-gluten containing cereals, onions, vegetable oils, and fat-free milk, and lower in breast milk. This diet corresponded to higher levels of PE (34:2).

Dietary patterns generated from metabolites and food intake in the nested case-control study were applied to the full cohort to generate one metabolite-related dietary pattern score for each dietary pattern on all 9-month diet records (**Table IV-5**). Subjects developing mAb+ by age 6 years were more likely to have a first-degree relative (FDR) with type 1 diabetes and to have high-risk HLA-DR3/4 genotypes. No dietary patterns were univariately associated with becoming a case of mAb+. While dietary pattern 3 was significantly protectively associated with mAb+ in the nested

54

case-control study (OR=0.67, 95%CI=0.48-0.94, **Table IV-6**), there was no association

seen in time-to-event analyses applied to the whole cohort and adjusted for clinical

center, sex, HLA-DR3/4, and FDR (HR=0.98, 95%CI=0.83-1.16, **Table IV-6).** No other 9-

month metabolite-related dietary patterns were associated with development of mAb+

in the TEDDY cohort. Results did not change in a sensitivity analysis in which we

adjusted for an additional 17 covariates that TEDDY has identified as associated with

development of IA (race-ethnicity, maternal education, maternal age, introduction of

probiotics before 28 days, introduction of probiotics at or after 28 days, weight for age

z-score at 12 months, and number of minor alleles for rs2476601, rs2816316,

rs11711054, rs10517086, rs4948088, rs1004446, rs7111341, rs2292239, rs3184504,

rs3825932, rs12708716) (data not shown).

Table IV-5: Cohort characteristics of TEDDY children with food records at age 9-months.

	mAb+ by age 6 yr		mAb- at age 6 yr		
	n	%	N	%	p-value
Total Subjects	300		6237		
Clinical Center					0.196
Colorado	52	17.33	987	15.82	
Georgia	25	8.33	638	10.23	
Washington	29	9.67	837	13.42	
Finland	77	25.67	1406	22.54	
Germany	23	7.67	365	5.85	
Sweden	94	31.33	2004	32.13	
Female	139	46.33	3075	49.3	0.315
First degree relative	71	23.67	613	9.83	<0.001
HLA-DR3/4	169	56.33	2354	37.74	<0.001
	Mean	**SD**	**Mean**	**SD**	**p-value**
Age at first Ab (days)	990.1	517.3	--		
Dietary Pattern Scores					
1	-0.055	1.07	0.004	0.98	0.349
2	-0.076	0.88	0.005	0.93	0.120
3	0.082	1.06	0.008	1.04	0.238

Table IV-6: Dietary patterns at 9-months of age associated with risk of mAb+ in TEDDY.

Metabolite-related dietary patterns	Nested Case-Control[*] n=147 mAb+ cases			Cohort[†] n=300 mAb+ cases by 6 years		
	OR	95%CI		HR	95%CI	
1	0.85	0.68	1.05	0.95	0.83	1.08
2	0.81	0.61	1.08	0.89	0.78	1.02
3	0.67	0.48	0.96	0.98	0.83	1.16

Conditional logistic regression models adjusted for age at metabolomics blood draw and total energy
†Survival models adjusted for clinical center, sex, FDR, total energy, and HLA DR3/4
OR=Odds Ratio, CI=confidence interval, HR=Hazard Ratio

Discussion

We identified dysregulated metabolism at the onset of and preceding stage 1 diabetes (mAb+) in a multi-national, prospective type 1 diabetes study. Different groups of metabolites distinguished cases and controls depending on whether it was directly after or prior to the detection of autoantibodies. Only PC and PE metabolites were consistently decreased in mAb+ cases compared to controls. At seroconversion, unsaturated triglycerides and amino acid groups were lower among mAb+ cases; whereas SM, GlcCer, and EtherPL lipids were lower among mAb+ cases in infancy. Infancy dietary patterns explaining choline- and sphingosine- containing lipids were not associated with the development of mAb+ in the full TEDDY cohort.

Dicarboxylic acids were the only metabolite group we found associated with increased risk of mAb+ in infancy. Adipic acid, the key compound of the dicarboxylic acids group, is not a substantial component of food other than as a biomarker of jello intake,[160] and as such was poorly explained by dietary pattern analysis. Other dicarboxylic acids associated with increased mAb+ risk included the tricarboxylic acid (TCA) cycle intermediaries succinic acid and malic acid. Through their regulation of

demethylase activity, succinic acid and other TCA cycle intermediaries may be important regulators of DNA and histone methylation,[161] which have been implicated in pathogenesis of type 1 diabetes.[162] While previous metabolomics studies have similarly identified early life TCA cycle metabolite levels (glutamic acid, succinic acid) as important for future development of type 1 diabetes among Finnish children,[96] the effects are not always consistent with the present study. Measurement of TCA cycle metabolites, particularly citric acid, was limited by the use of citrate tubes for plasma collection and storage in TEDDY. Future work should better characterize the role of TCA cycle metabolites in type 1 diabetes using plasma stored in a different medium.

Sphingolipid metabolism plays a role in diabetic pathologies, including regulating beta-cell apoptosis, proinsulin and insulin folding in the endoplasmic reticulum, and cytokine secretion.[163] The evidence supporting this connection has been recently extended from animal models into human islet cells.[164] We identified two sphingolipid groups as significantly lower in infancy for mAb+ cases versus controls, including the SM group, which were previously identified in type 1 diabetes metabolomics studies,[96,103] and the GlcCer group. As a whole, sphingolipids have been characterized as both pro- and anti-inflammatory. Endogenous sphingolipids are metabolically involved in T-cell regulation, autoimmunity, and inflammation,[165] yet consumption of dietary sphingolipids have been linked to anti-inflammatory responses.[166] The protective effects may operate via changes in gut microbiota or by activating other cofactors such as peroxisome proliferator-activated receptor γ expression,[167] both of which have been implicated in type 1 diabetes,[168] While the sphingolipid metabolites were explained by Dietary Pattern 2 and trended toward

57

protective, it was only marginally significant in the full cohort analysis. Targeted

characterization of the relationship between sphingolipid dietary intake and metabolite

levels might help to disentangle reported contrasting effects, which likely depend on

other factors such as specific sphingolipid structure and existing metabolic state.

We identified choline-containing lipid groups (PC, SM, EtherPL) as protective for

development of mAb+ at 9 months of age, consistent with previous studies conducted at

birth and 3 months of age.[96,102,103] Choline is important for rapid growth and

development in infancy, as a constituent of phospholipid cellular membranes.

Additionally, it may play a role in insulin resistance or energy metabolism, perhaps

through its role as a methyl donor for epigenetic changes.[169] Breastmilk, infant formula,

and milk are important sources of choline in the infant diet, though the amount and

type of choline-containing compounds vary widely across dietary sources.[170,171]

Correlation of choline-containing compounds with each other, and with breastmilk and

infant formula were identified in Dietary Pattern 3, which was protective for mAb+ in

the nested case-control study only.

However, no metabolite-related dietary pattern at 9-months was associated with

the development of mAb+ by the age of six years in the full TEDDY cohort. The percent

of metabolomics variation explained by the dietary patterns is modest, yet comparable

to other dietary pattern studies utilizing targeted metabolomics measures.[126] The

remainder of the variability is likely related to other environmental, lifestyle, and

genomic factors that influence metabolism.

There are several factors that could contribute to the lack of dietary pattern

association found, despite overlap between the mAb+ cases on whom the patterns were

created and all mAb+ cases included in cohort analyses. First, untargeted metabolomics and 3-day food records may not be measured precisely enough to successfully identify disease-related dietary patterns at such a young age where variability in both is large. Second, using the most significant metabolite in each group may not be the best choice of response variable, which is the variable that determines the ability of reduced rank regression to capture disease-related variation in the diet.[172] Metabolomics measures were further limited because they were quantified from non-fasting samples, which has been shown to differentially impact serum metabolic profiles related to dietary factors.[173]

Despite the large variation in untargeted metabolomics panels and diet, both of which vary by geography in our multi-national setting, we were able to identify metabolic dysregulation prior to the detection of autoantibodies that distinguished children whose lifetime risk for symptomatic (Stage 3) type 1 diabetes approaches 100%.[174] Our findings confirmed smaller, country-specific studies and expanded the knowledge base of metabolite groups in the more heterogeneous TEDDY study population. Though results generalize to a broader population than previous studies, the very young cases (<= 9-months of age) were excluded from our study. The discovery of only one group of metabolites (PCs) that was identified across the pre-autoimmune period is consistent with other type 1 diabetes studies where the effect of non-genetic factors could be age-related or have an effect only during certain susceptible windows. While novel application of dietary patterns summarizing candidate metabolites did not successfully extend outside of the nested case-control

study, the approach may show promise for future work with targeted measurement of disease-related metabolites.

In summary, we identified metabolomics differences between cases and controls both at the time of autoantibody detection, and prior to its development. Metabolomics differences were more apparent when comparing the high-risk mAb+ group to controls than comparing all cases of IA to controls. Few differences were identified by the type of first-appearing autoantibody. Dietary patterns explaining candidate metabolites were not significantly associated with development of mAb+ in the full TEDDY cohort. The risk for mAb+ was associated with decreased levels of choline- and sphingolipid-related metabolomics factors in infancy, possibly due to changes in lipid-related dietary intake.

Acknowledgments

CHAPTER V

METABOLOMICS-RELATED NUTRIENT PATTERNS AT SEROCONVERSION AND RISK OF PROGRESSION TO TYPE 1 DIABETES

Abstract

Objective

To elucidate the role of diet in type 1 diabetes by examining combinations of nutrient intake in the progression from islet autoimmunity (IA) to type 1 diabetes.

Research Design and Methods

We measured 2,457 metabolites and dietary intake at the time of seroconversion in 132 IA-positive children in the prospective DAISY study. IA was defined as the first of two consecutive visits positive for at least one autoantibody (insulin, GAD, IA-2, or ZnT8). By December 2018, 40 children progressed to type 1 diabetes. Intakes of 38 nutrients were estimated from semi-quantitative food frequency questionnaires. We tested the association of each metabolite with progression to type 1 diabetes using multivariable Cox regression. Nutrient patterns that best explained variation in these significant metabolites were identified using reduced rank regression (RRR), and their association with progression to type 1 diabetes was tested using Cox regression adjusting for age at seroconversion and high-risk HLA genotype.

Results

In stepwise selection, 22 nutrients significantly predicted at least two of the 13 most significant metabolites associated with progression to type 1 diabetes, and were included in RRR. A nutrient pattern corresponding to intake lower in linoleic acid, niacin, and riboflavin, and higher in total sugars, explained 18% of metabolite

variability. Children scoring higher on this metabolite-related nutrient pattern at seroconversion had increased risk for progressing to type 1 diabetes (HR=3.31, 95%CI=1.46-7.47).

<u>Conclusions</u>

Combinations of nutrient intake reflecting candidate metabolites are associated with increased risk of type 1 diabetes, and may help focus dietary prevention efforts.

Introduction

Type 1 diabetes is characterized by destruction of the insulin-producing beta-cells in the pancreas, and affects over 100,000 children and adolescents in the U.S.[175] Clinical type 1 diabetes is preceded by a period of detectable and persistent autoimmunity to islet antigens, called islet autoimmunity (IA). While the incidence of childhood type 1 diabetes has been increasing worldwide for decades,[175] the incidence of IA appears stable,[176] suggesting that an increase in progression from IA to clinical diabetes may help explain these incidence trends. Therefore, it is important to identify factors influencing progression from IA to type 1 diabetes.

Non-genetic factors are involved in the development of type 1 diabetes, which also has a well-defined genetic risk component.[16] Dietary intake in early life and throughout childhood has been implicated in type 1 diabetes risk, though without conclusive findings of a single responsible risk factor.[177] Higher intake of some foods or nutrients at seroconversion has been associated with increased risk of progression to type 1 diabetes, including: glycemic load,[82] total sugars and sugar-sweetened beverages,[83] and cow's milk protein.[77] Other dietary factors, including vitamin D intake and status,[85,178] and omega-3 fatty acid intake and status,[79] were not associated with

progression. These traditional investigations of individual dietary risk factors may oversimplify true dietary exposure since multiple nutrients are contained in foods, and nutrients may work synergistically or antagonistically in the body. We sought to improve upon previous investigations into diet and progression from islet autoimmunity to type 1 diabetes by considering combinations of nutrient intake, or nutrient patterns.

Our aim was to identify nutrient patterns of intake at seroconversion associated with risk of progression to type 1 diabetes in the Diabetes Autoimmunity Study in the Young (DAISY), as shown in **Figure V-1**. To capture disease-related variation in nutrient intake, first we conducted a metabolome-wide association study to identify metabolites and lipids associated with progression from IA to type 1 diabetes. As markers of biological processing and response to nutrition, metabolites have been used together with reported dietary intake to elucidate the role of diet in the development or prevention of other outcomes, such as cardiovascular disease [179]. Next, we used reduced rank regression to create nutrient patterns best explaining those candidate metabolites. Finally, we tested the nutrient pattern with the risk of progression from IA to type 1 diabetes.
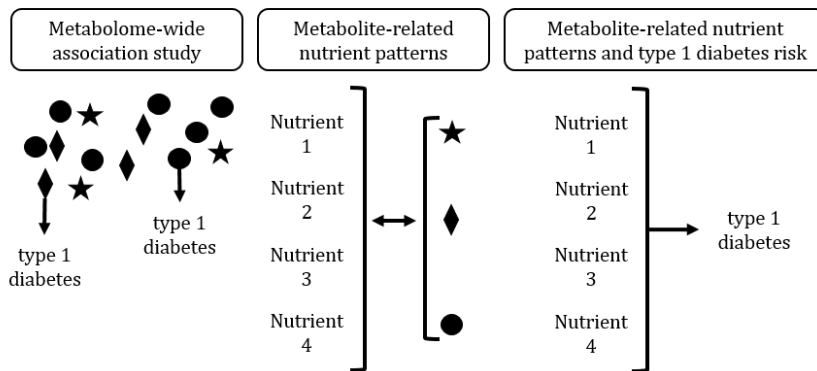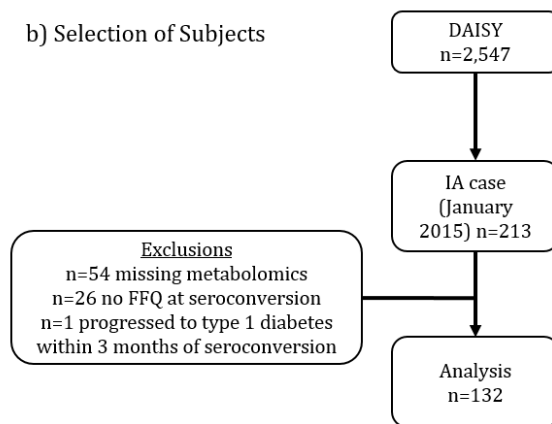
Figure V-1: Study design for identifying metabolite-related nutrient patterns associated with progression from islet autoimmunity (IA) to type 1 diabetes. a) We conducted a metabolome-wide association study to identify seroconversion lipidomics and metabolomics associated with progression to type 1 diabetes by December 2018. Significant metabolites and nutrient patterns that best summarized them were identified using reduced rank regression. Metabolite-related nutrient pattern scores were tested for the risk of type 1 diabetes. b) Of the 213 IA cases identified in DAISY by January 2015, 132 had untargeted metabolomics measures and dietary intake available at the time of seroconversion.

Research Design and Methods

Study Design

      From 1993-2004, DAISY recruited 2,547 children in Colorado who were at high

risk for developing type 1 diabetes.[111] Prospective follow-up for the development of

islet autoantibodies and type 1 diabetes is ongoing, and has included clinic visits at 9, 15, and 24 months, and annually thereafter.[180] An IA case is defined by the presence of one or more confirmed autoantibody to insulin, GAD65, IA-2, or ZnT8 on two or more consecutive clinic visits. Seroconversion is defined as the first visit at which IA was detected. IA cases follow an accelerated protocol with clinic visits every 3-6 months until type 1 diabetes is diagnosed by a physician following standardized criteria, including typical symptoms of polyuria and/or polydipsia and a random glucose > 11.1 mmol/l or an OGTT with a fasting plasma glucose ≥ 7.0 mmol/l or 2-h glucose > 11.1 mmol/l.[20]

To identify seroconversion metabolomics and nutrient patterns associated with progression from IA to type 1 diabetes, we examined IA cases who seroconverted prior to January 2015 and their prospective follow-up for type 1 diabetes through December 2018.

Metabolomics Data Acquisition, Normalization, and Pre-Processing

Untargeted metabolomics measures were acquired from serum samples using GC-TOF MS (primary metabolism), CSH-QTOF MS (complex lipids), and HILIC-QTOF MS (biogenic amines), at the NIH West Coast Metabolomics Center at the University of California, Davis. Samples were kept at -80°C prior to analysis. Samples were allowed to thaw on wet-ice and kept cold during extraction, once thawed samples were inverted multiple times to ensure serum homogeneity. Samples were extracted by taking 30 μL of serum and performing a liquid-liquid extraction first previously described[181] with modification, including addition of labeled internal standards for quality control and retention time correction.[155] The aqueous phase, containing polar metabolites, was split

into two equal volumes and dried then re-suspended in 1:1 acetonitrile:water to precipitate any remaining lipids. One polar aliquot was then prepared for and analyzed by HILIC-QTOF MS/MS.[182] The second polar aliquot was analyzed by GC-TOF-MS.[183] The non-polar phase was analyzed by CSH-QTOF MS/MS for identification and relative quantification of complex lipids and free fatty acids.[155]

Peak picking, integration, alignment and annotation of GC-TOF-MS data was performed using BinBase.[184] Liquid chromatography (LC) data (CSH-QTOF-MS and HILIC-QTOF-MS) were processed with MS-Dial,[185] and complex lipids were annotated with LipidBlast[186] and Massbank of North America (http://mona.fiehnlab.ucdavis.edu/). MS-FLO was used to remove erroneous peaks and reduce the false discovery rate in LC datasets.[187]

After data were collected, annotated and post-processed they were normalized to account for any instrument drift, variation, or batch effects which may have occurred. Data were normalized by our in-house normalization algorithm; systematic error removal using random forest (SERRF).[188] SERRF is a QC-based normalization method using the random forrest algorithm.[189] Briefly, SERRF was designed to be used in place of compound independent normalization methods such as locally weighted smoothing (LOESS). While LOESS normalizes compounds independently, SERRF takes the metabolites' correlation into consideration and automatically assigns higher weight to important compounds. SERRF is nonparametric, nonlinear, less prone to overfitting, robust to outliers and noise, and fast to train.

After normalization, we performed quality control checks at the sample and metabolite level. Samples with a high proportion of features estimated to be zero were

considered low abundance (n=2) and excluded from analyses. Each metabolite was transformed using box-cox transformation analysis, to improve heteroscedasticity and normality.[156] Metabolites with a coefficient of variation more than 2 absolute deviations from the median were excluded from analyses (p=344), as they may be unreliable or produce unstable estimates. Coefficient of variation median and median absolute deviations were calculated separately by panel. After quality checks, there were 2,457 untargeted metabolites for statistical analyses, including 1,905 (77%) unknown metabolites.

Dietary Intake Assessment

Dietary intake was measured annually using the Willett semi-quantitative food frequency questionnaire (FFQ) starting at age two years. Parents completed an FFQ representing their child's average intake over the previous year until age 10 years, after which study participants self-reported using the Youth Adolescent Questionnaire (YAQ). The FFQ has been previously validated,[118] and the FFQ and YAQ shown to be comparable in the DAISY study population.[117]

The average amounts of daily nutrient intakes were estimated from the FFQ and YAQ at the Channing Laboratory, Harvard, MA. Nutrients that were measured throughout the DAISY study were considered for inclusion in nutrient pattern analyses (p=40). Where available, nutrient variables included the intake from both foods and supplements. We used the residual method to calculate the energy-adjusted nutrient intakes.[100] There were 38 energy-adjusted nutrients available for nutrient pattern analysis (**Appendix C**).

Statistical Methods

IA cases missing metabolomics at seroconversion (n=54), those without a FFQ at seroconversion (n=25), those reporting unreasonable total calories (>5,000 kcal, n=1), or whose seroconversion visit occurred less than 3 months prior to type 1 diabetes diagnosis (n=1) were excluded from all analyses (**Figure V-1**). All metabolites were standardized prior to analyses to allow for direct comparison of the magnitude of association with type 1 diabetes. We used Cox regression to identify seroconversion metabolomics associated with progression from IA to type 1 diabetes, adjusting for age at seroconversion and high-risk HLA (DR3/4) genotype. The time-to-event was defined as the time from seroconversion (the first of two consecutive positive visits) to diagnosis with type 1 diabetes. We adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate.[190]

Using the most significant metabolites identified in discovery (minimum FDR *q-value*), we used reduced rank regression (RRR) to identify nutrient patterns that best explained metabolite variation. RRR helps summarize high-dimensional data by creating linear combinations of nutrients (nutrient patterns) that maximize the variance explained in a set of response variables (metabolites).[123] We reduced the number of nutrients included in RRR by first applying a stepwise regression selection procedure for each metabolite.[133] Nutrients that significantly predicted at least 2 metabolites (*p-value*<0.1, **Appendix C**) were included in nutrient pattern analysis. Energy-adjusted nutrient intakes were standardized prior to the RRR. Results of RRR included metabolite weights, nutrient loadings, and a nutrient pattern score for each

participant, indicating how similar their intake was to the nutrient pattern. Weights and loadings more extreme than ±0.2 were used to interpret the nutrient pattern.

We tested the association of metabolite-related nutrient pattern scores with progression from IA to type 1 diabetes using Cox regression. For comparison with the nutrient pattern, we included an additional analysis testing individual nutrient associations with progression from IA to type 1 diabetes for nutrients that loaded strongly (more extreme than ±0.2) on the nutrient pattern. All multivariable Cox regression models were adjusted for age at seroconversion and high-risk HLA (DR3/4) genotype. SAS 9.4 (SAS Institute Inc., Cary, NC, USA) was used for all statistical analyses.

Results

Of 132 IA cases with metabolomics and dietary intake measured at the time of seroconversion, 40 progressed to type 1 diabetes by December 2018. Characteristics of the study population are shown in **Table V-1**. Type 1 diabetes progressors were 48% male, mostly non-hispanic white, and 63% had a first-degree relative with type 1 diabetes. These characteristics were similar in the non-progressor group. However, compared to non-progressors, those who progressed to type 1 diabetes had higher-risk HLA genotypes (43% versus 28%) and seroconverted to IA positivity at younger ages (5 years versus 7.8 years).

From individual multivariable Cox regression models, 201 metabolites (8%) had nominal associations (p-value<0.05) with risk of progression to type 1 diabetes (**Figure V-2**). None of these associations reached significance after adjustment for multiple comparisons (q-value<0.05). The 13 metabolites that were tied at the minimum q-value of 0.3451, including threonine, histidine, choline, and 10 unknown compounds from the

biogenic amines and complex lipids panels, were used in the nutrient pattern analysis.

Mass spectra for the compounds are included in **Figure V-3**. The majority of nominally

significant metabolites were associated with increased risk of progression to T1D

(**Figure V-2**).

Table V-1: Characteristics of 132 IA cases at seroconversion

	Progressed to type 1 diabetes (n=40)		Non-progressors (n=92)		
	n	%	n	%	*p-value**
Male	19	47.50	45	48.91	0.883
Non-Hispanic white	36	90.00	68	73.91	0.079
High-risk HLA (DR3/4)	17	42.50	26	28.26	0.046
First degree relative with type 1 diabetes	25	62.50	52	56.52	0.595
	Mean	SD	Mean	SD	
Age (years)	5.00	3.82	7.83	4.28	0.025
Total calories	2146.04	492.47	2077.66	738.63	0.811

*From Cox regression

To identify nutrient patterns using RRR, we first selected nutrients that were

significant predictors of the 13 metabolites identified in discovery. Stepwise selection

results indicated that 22 of 39 nutrients predicted at least 2 metabolites (**Appendix C**).

From RRR, one nutrient pattern was identified that explained 18% of metabolite

variation and 5% of nutrient variation. The metabolite weights and nutrient loadings

that define the nutrient pattern are shown in **Figure V-4**. A higher score on the nutrient

pattern corresponded to a diet with higher intake of total sugars, vitamin C, and

monounsaturated fat, and lower in linoleic acid, niacin, riboflavin, vitamin K, vitamin

B12, and caffeine (**Figure V-4a).** Higher scores on the nutrient pattern also

corresponded to higher amounts of unknown metabolites hilic_291, hilic_294, hilic_179,

hilic_353, hilic_1015, hilic_383, and lower amounts of hilic_243 (**Figure V-4b**).

Annotated metabolites threonine, histidine, and choline were poorly explained by the

nutrient pattern, as indicated by weights closer to 0.

In multivariable Cox regression, nutrient pattern scores at seroconversion were

associated with increased risk of progression from IA to type 1 diabetes (**Figure V-4c,**

HR=3.31, 95%CI=1.46-7.47). In analyses of the nine nutrients that loaded high on the

nutrient pattern, only vitamin B12, riboflavin, niacin, and total sugars had significant

individual associations with progression from IA to type 1 diabetes.
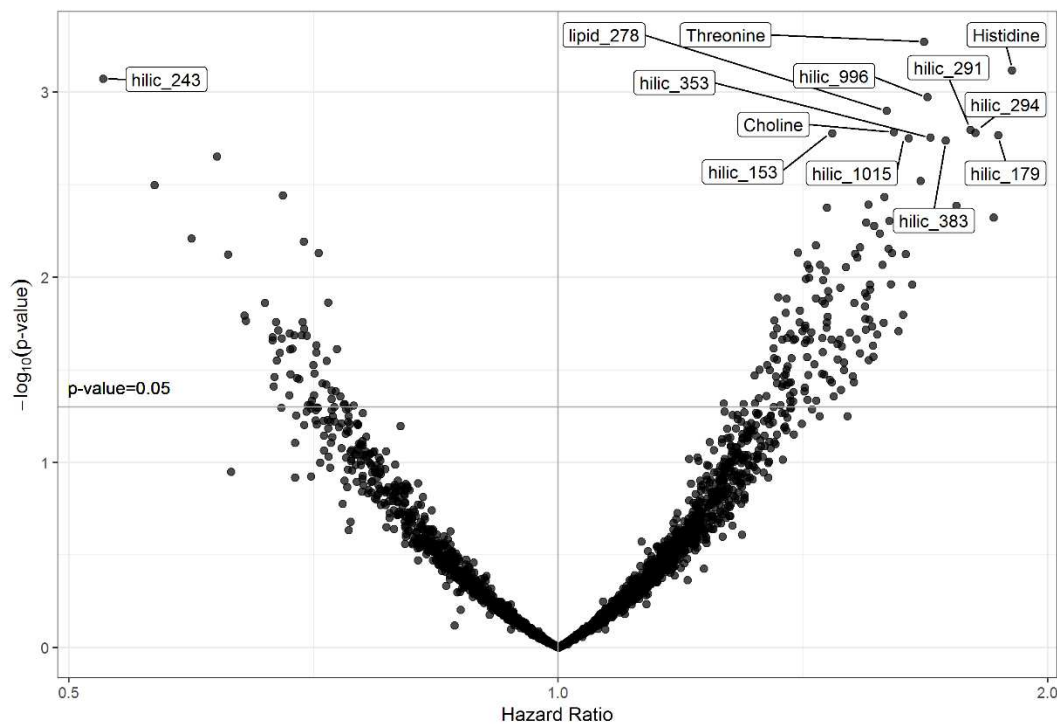
Figure V-2: Cox regression results showing the association of 2,457 metabolites with progression from IA to type 1 diabetes. 201 metabolites (8%) had nominal *p-value*<0.05 (above the guideline), and none were significant at *q-value*<0.05. The 13 labeled metabolites had the minimum *q-value* of 0.3451 and were selected for nutrient pattern analyses. These included threonine, histidine, choline, and 10 unknown compounds. Unknown compounds were labeled by the untargeted platform on which they were measured (e.g., HILIC_243 was measured using HILIC-QTOF-MS)
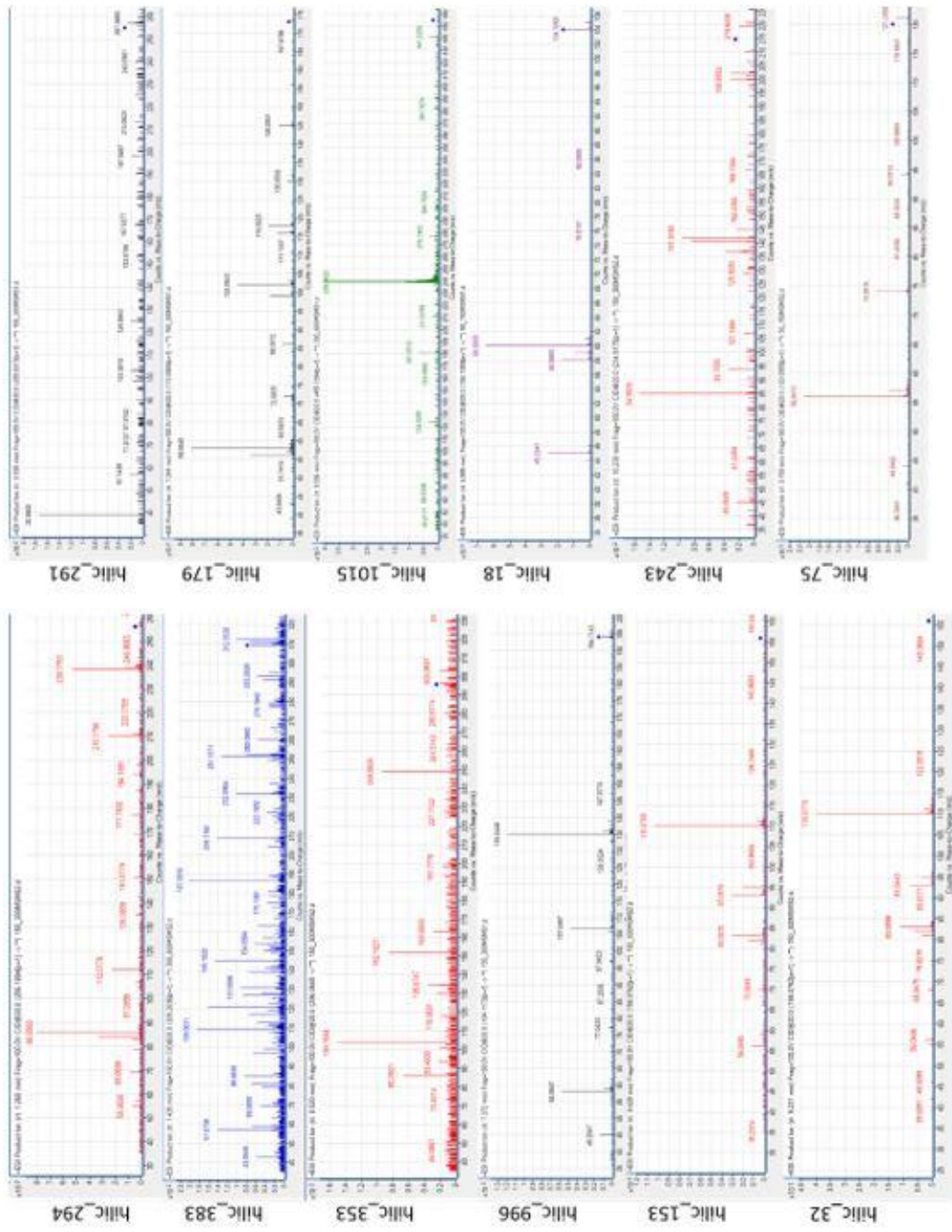
Figure V-3: Mass spectra for 12 of the 13 most significant metabolites associated with progression from IA to T1D. The unknown lipid_278 did not have MS/MS available.

73
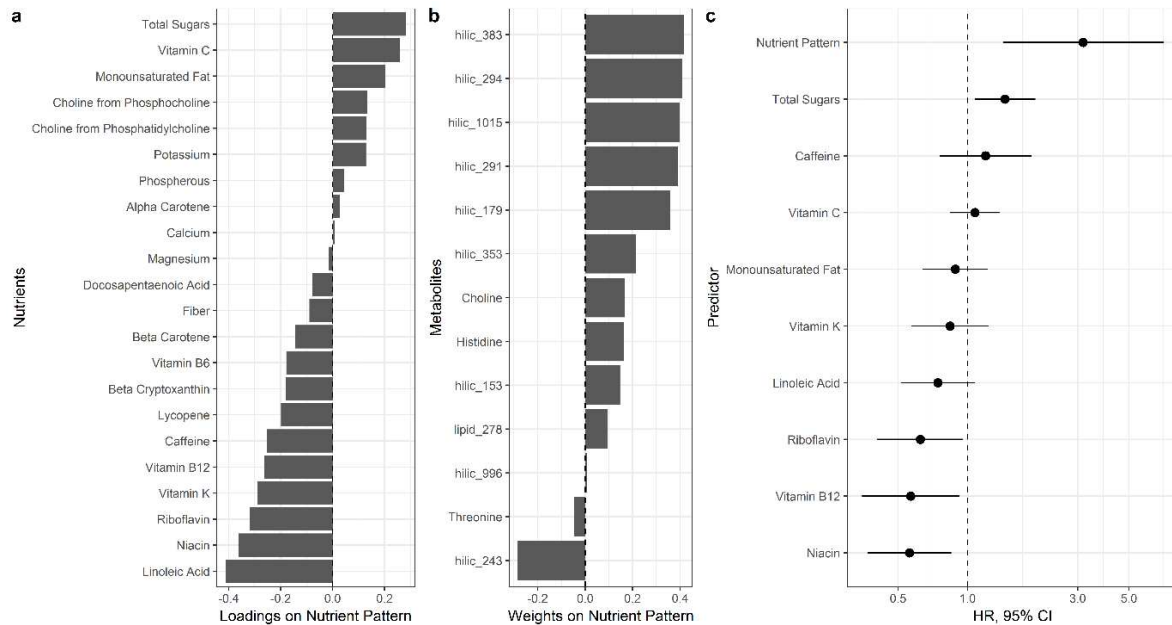
Figure V-4: A seroconversion metabolite-related dietary pattern and its association with progression from IA to type 1 diabetes. One dietary pattern explained 5% of nutrient variation and 18% of metabolite variation. a) The loadings indicate which combinations of nutrients defined the nutrient pattern scores. b) The weights indicate which metabolites were best explained by the nutrient pattern. c) Multivariable Cox regression results showing the associations of the nutrient pattern score and individual nutrients in progression from IA to type 1 diabetes.

Conclusions

We identified a nutrient pattern of intake at seroconversion that was significantly associated with increased risk of progression to type 1 diabetes. Higher nutrient pattern scores (and increased risk of progression) corresponded to higher intake of total sugars, vitamin C, and monounsaturated fat, and lower in linoleic acid, niacin, riboflavin, vitamin K, vitamin B12, and caffeine. Accounting for combinations of nutrient intake in the diet using a nutrient pattern score resulted in a stronger association with progression to type 1 diabetes (HR=3.31) than the individual nutrients

74

highly contributing to the pattern (HR range=0.56-1.45), and the metabolites used to inform the pattern (HR range=0.52-1.9).

Several of the nutrients that highly contributed to the nutrient pattern have been investigated in the development of IA or type 1 diabetes for years, often resulting in small or inconsistent associations with disease endpoints. For example, higher intake of total sugars was previously associated with increased risk of progression in DAISY [83]. Dietary intake of the omega-6 fatty acid, linoleic acid, was not associated with progression from IA to type 1 diabetes.[79] However, linoleic acid status has been inconsistently identified as protective for development of IA.[150] Similarly, the niacin derivative nicotinamide was protective for type 1 diabetes in mice, but shown to be ineffective in a prevention trial.[38] The failure to account for the inter-relationship between nutrients may explain the inconsistencies surrounding the role of dietary intake in type 1 diabetes development.

Other nutrients contributing to the pattern have not been directly studied, but have plausible connection to type 1 diabetes. For example, riboflavin[191] and vitamin K (K_1, phylloquinone)[192] have antioxidant properties that may protect islet cells from destruction by free radicals and oxidative stress. Through its role as a cofactor in one-carbon metabolism, vitamin B12 affects DNA methylation, which is also implicated in the disease process (cite methylation). Notably, vitamin C contributes in an unexpected direction to the nutrient pattern. Due to its antioxidant and anti-inflammatory properties, vitamin C would be expected to have an inverse relationship with risk of progression to type 1 diabetes. We speculate that the correlation between vitamin C and sugars intake in foods may explain this finding.

The metabolite discovery was primarily used as a means to capture meaningful, disease related variation in nutrient pattern intake. However, a broad conclusion of our metabolomics findings indicate that metabolic differences at seroconversion distinguish which IA cases will go on to type 1 diabetes, similar to previous studies comparing IA cases and type 1 diabetes cases to autoantibody negative controls.[95,96] Though conducted in different disease stages, each study has identified various amino acids as dysregulated prior to the development of different disease endpoints. In our study, however, threonine and histidine were poorly correlated with diet and therefore less influential in the nutrient pattern. The other study comparing IA cases to type 1 diabetes cases investigated the lipidome in cord blood.[193] The majority of our candidate metabolites were unknown compounds measured on the HILIC panel, making comparison infeasible.

The thirteen candidate metabolites were mostly associated with increased risk of progression to type 1 diabetes. While these associations were only nominal, our nutrient pattern explained over 18% of metabolite variability—over three times the amount of variability explained in previous dietary pattern studies,[128] and studies of individual dietary factors and targeted metabolites.[126] This suggests that the metabolites used as response variables for identifying the nutrient pattern are related to nutrition. Future work should identify and annotate these unknown compounds, and consider the strong correlation between metabolites which makes the use of typical multiple comparison procedures too strict. The novel use of nutrients rather than food groups in pattern analysis likely contributes to the increased variability explained by the nutrient pattern—nutrient intakes have more continuous distributions and may

perform better in correlation-based analyses such as RRR. While nutrient patterns may be harder to interpret than dietary patterns of food, nutrients are more comparable across different populations and therefore may be more amenable to the development and cross-cultural implementation of dietary interventions to prevent type 1 diabetes.

As hypothesized, examining nutrient intakes in combinations, to account for their correlation in intake and in biological processing, resulted in a stronger magnitude of association than any of the individual constituent nutrients. The relationship of highly contributing nutrients to the nutrient pattern is consistent with existing literature and hypotheses, suggesting that the contents of the diet may work in combination to affect the development of type 1 diabetes, and should be considered jointly in future work.

Acknowledgments

Thank you to the participants and families of the DAISY study, whose continued commitment make such research possible.

CHAPTER VI

DISCUSSION

Using newer nutritional epidemiologic methods, this dissertation explored the role of metabolomics and dietary patterns in the nutrition etiology of type 1 diabetes. First, we identified individual metabolites and chemically similar metabolite groups associated with the development of mAb+ and T1D in metabolome-wide association studies of two prospective cohorts (TEDDY and DAISY). Candidate metabolites were used to capture disease-related dietary and nutrient patterns, which were subsequently tested for association with disease endpoints. In TEDDY, a dietary pattern explaining metabolites in infancy was associated with decreased risk of mAb+ but was not generalizable when applied to similarly at-risk infants. In DAISY, a nutrient pattern explaining unknown metabolites at seroconversion was associated with increased risk of progression to T1D.

We made significant and novel contributions to the understanding of metabolomics in the natural history of T1D, though the metabolomics discovery was primarily used as a means for identifying disease-related variation in dietary intake. Our DAISY study was the first investigation of metabolomics in T1D development in a U.S. population, and by far had the best coverage of the metabolome—we measured almost 2,500 metabolites, in contrast to previous studies which detected anywhere from 106[102] to 540[95] metabolites. With participants from the U.S., Finland, Germany, and Sweden, the TEDDY study had a larger sample size representing more geographic areas than the smaller, country-specific studies of metabolomics in T1D risk previously conducted in Germany,[95] Sweden,[102] Norway,[104] and Finland.[96,101,103,193]

Increased sample size, a more generalizable population (multi-national), and greatly expanded coverage of the metabolome allowed us to comprehensively characterize sphingomyelin, choline, and TCA cycle dysregulation in infancy that may lead to the development of mAb+ later in life. Previous studies have identified a few compounds in each of these classes; however, direct comparisons with them are complicated by the varying number of metabolites tested, the stage of disease studied, and the time at which metabolites were measured. For example, only two[101,104] of the seven previous metabolomics studies conducted in prospective T1D cohorts also examined cases positive for multiple autoantibodies compared to controls. Of those, one measured only lipidomics (p=159) in cord blood, and the other measured only metabolomics (p=279) repeatedly over time beginning at 3 months.

Comparison of our key metabolomics findings from the DAISY study with previous literature is similarly complicated. Only three of the 13 metabolites important for progression from IA to T1D were annotated—many metabolomics studies either exclude unknown compounds from analyses, or fail to provide distinguishing characteristics (e.g., mass spectra) for them. Of the seven previous prospective metabolomics T1D studies, only the most recent investigations in Finland compared T1D cases to IA cases who did not progress to T1D; however, this study only included annotated compounds in analyses.[103,193] These difficulties highlight current challenges for the use of high-throughput metabolomics in epidemiologic research.

Metabolomics is one of the newer "—omics" fields to gain popularity in human research. The increasing technological capability and decreasing cost of comprehensively interrogating small molecules in a biospecimen contributes to its

growing use, which in turn is driving a period of rapid growth and progression in the field. For example, in only five years between the acquisition of TEDDY and DAISY metabolomics data, significant improvements were made to compound deconvolution, identification and normalization methods. These, and other, changes resulted in a different number of metabolites and proportion of annotated metabolites available for statistical analyses in our studies, despite using the same two metabolomics platforms.

Differently shaped volcano plots from the MWAS discoveries in the two studies highlight the potential impacts of changes to platforms or analytical pipelines over time. In the TEDDY study, the lipids were mostly protective for the development of outcomes and represented the highest proportion of metabolites tested, giving the volcano plots a heavier left cloud (OR<1). The DAISY volcano plot has the opposite skew, with the majority of candidate probes in the right cloud (HR>1). Interrogation of different disease endpoints may explain this difference—perhaps early life metabolomics signatures are protective for mAb+, while seroconversion metabolomics differences reflect only the increased risk for progression to T1D. Similarly likely, however, the right-skewed volcano plot in DAISY reflects the addition of the HILIC panel, which doubled the total number of metabolites tested, many of which were associated with increased risk of outcome. Similarly, the TEDDY volcano plots showed a pattern of metabolites with extreme magnitudes of association that do not achieve significance, near the 0 limit of the OR. Further investigation revealed that those metabolites had little variability, which led to unstable estimates of association. Changes to the pre-processing pipeline to filter metabolites with both extremely high and extremely low coefficient of variation removed that artefact, giving the DAISY volcano plot a different

shape. Finally, standardization (mean, standard deviation) of metabolites in the DAISY study also changed the shape of the volcano plot—giving it a butterfly, rather than cloud, shape. Changes to pre-processing and analytical pipelines complicates comparisons between metabolomics studies, both internal to our own two studies, and externally to other metabolomics studies of type 1 diabetes.

Limitations of available statistical methods may also influence the ability to discover metabolites strongly, significantly, and specifically associated with disease endpoints using high-throughput metabolomics. The need to account for an unknown background of metabolites that should be detectable in biospecimens and for the unusually highly correlated features (compared to other –omics) are two specific challenges that should be addressed. Both of these attributes may have influenced our finding of less extreme p-values and adjusted p-values in the much larger and better-powered TEDDY study than in the smaller DAISY study. Unlike genomics, which has matured over decades of research in humans, metabolomics lacks normative procedures for almost all parts of the pipeline from bench to publication.

Standardizing protocols, analytical pipelines, and developing appropriate statistical methods will greatly improve the utility, interpretation, and reproducibility of metabolomics findings in epidemiologic research. Specific to our findings of metabolomics in T1D, immediate next steps include: 1) better characterization of the relationship between annotated and unknown metabolites to appropriately be able to adjust for multiple comparisons, 2) examination of longitudinal changes in metabolomics prior to disease development utilizing the rich TEDDY resource with measurements every 3 months since birth, and 3) identification of unknown

metabolites highly explained by the T1D-related nutrient pattern identified in DAISY. Despite the challenges of conducting research in a new and rapidly changing field, our metabolomics discovery was useful for capturing dietary and nutrient pattern intake related to T1D risk.

With our investigation of dietary patterns, we made significant and novel contributions to both nutrition epidemiology and to the nutrition etiology of T1D. Ours were the first two studies in the nutrition literature of T1D risk to consider dietary patterns—previous investigations focused solely on evaluating nutrient or food group risk factors individually.[76,78,83] In the DAISY study, the combination of nutrient intake best explaining unknown metabolites was more strongly associated with T1D risk than any of the constituent nutrients or metabolites that defined the pattern. Higher nutrient pattern scores (and increased risk of progression) corresponded to higher intake of total sugars, vitamin C, and monounsaturated fat, and lower intake of linoleic acid, niacin, riboflavin, vitamin K, vitamin B12, and caffeine. Many of the constituent nutrients to the pattern have been inconsistently associated with T1D for decades (**Table VI-1**),[87,177] providing some confidence in the validity of the pattern identified. Other nutrients, such as vitamin C, riboflavin, and vitamin K were novel to this dissertation, and provided evidence that consideration of the whole diet is important in T1D research, to account for possible synergism or antagonism between dietary components.

Table VI-1: Summary of previous individual findings for constituent nutrients of the nutrient pattern associated with increased risk of progression from IA to T1D.

Nutrient in Pattern	Direction of Pattern Intake Associated with Increased T1D Risk	Previous Individual Findings	Hypothesis
Total Sugars	↑	↑ Total sugars at IA = ↑ T1D	
Linoleic Acid (LA)	↓	LA at IA ≠ T1D ↑ LA = ↓ IA and ≠ ↓ IA	
Niacin	↓	↑ Nicotinamide = ↓ T1D (MICE) Nicotinamide ≠ ↓ T1D (HUMANS)	
Riboflavin	↓		Antioxidant, anti-inflammatory
Vitamin K	↓		
Vitamin C	↑		

The infancy dietary patterns identified in TEDDY were less successful—they explained less metabolite variation than the nutrient patterns (8% versus 18%) and were less strongly associated with disease (OR=0.67 versus HR=3.31). The investigation of two different stages in the natural history of T1D is perhaps the largest distinction between our two studies, and may explain why patterns explaining mAb+ were less successful than patterns explaining progression to T1D. Risk factors affecting development of IA are often different than those affecting progression to T1D.[11] Therefore, it is possible that diet is more influential after the appearance of autoimmunity than in its appearance.

The dietary pattern associated with the development of mAb+ in the nested case-control study association did not reproduce when applied to the full TEDDY cohort. However, this may not indicate less success than the nutrient patterns in DAISY, as there were insufficient additional IA cases in DAISY to investigate replication of the nutrient pattern. Since replication of RRR dietary patterns in independent population is generally recommended,[172] future investigation of both nutrient and dietary patterns in an independent population will lend more confidence to the generalizability of our findings.

Other reasons for the differences in our findings between studies may stem from differences between TEDDY and DAISY. In TEDDY, dietary intake is collected by food records covering three 24-hour periods, whereas DAISY collects FFQs covering average diet in the previous year. While both methods have limitations (the food record may not capture infrequently eaten foods, the FFQ overestimates intake), dietary patterns identified using RRR have been shown to be comparable for both assessments among adolescents.[130] TEDDY dietary intake is also complicated by the need to harmonize national food composition databases between four countries.[112] Dietary pattern procedures founded on the correlation structure between food intake may suffer from residual heterogeneity across countries, despite extensive harmonization efforts. The slightly different distribution of mAb+ cases by country in the nested case-control (**Table IV-2**) versus cohort (**Table IV-5**) analyses may partially explain why the protective association of the TEDDY dietary pattern with mAb+ did not reproduce.

Alternately, less successful identification of disease-related dietary patterns in TEDDY compared to DAISY may relate to differences in our patterns approach between

the studies. While RRR was used for both studies, the selection of response variables (and their correlation structure) heavily influences the resulting pattern.[172] Indeed the pattern cannot be interpreted independently of the response variables. Our selection of response variables differed by study, as they were determined by the metabolomics discovery Aim conducted independently in TEDDY and DAISY. While we only considered annotated metabolite groups as identified by ChemRICH enrichment analysis in TEDDY, we allowed for the inclusion of unknown metabolites in DAISY where selection was based solely on significance. The difference in metabolite variability explained by the patterns may indicate that the unknown metabolites were more strongly related to diet, or it may reflect that the more continuously distributed nutrients perform better in pattern analysis than food groups. The identification of patterns using nutrients rather than food groups was a major difference between the approaches, and a novel contribution of this dissertation to dietary patterns literature.

Much like metabolomics, the data-driven dietary patterns literature has few normative procedures, leaving much room for improvement in the search for combinations of dietary factors that may affect T1D risk. For example, by adding a filtering step to include only the nutrients that correlated well with metabolite response variables, we likely improved our ability to detect patterns associated with T1D. Similar filtering of the metabolites (or other response variables) to include only diet-related variables in RRR would likely improve the derivation of patterns associated with disease endpoints, if they truly exist. Prior to RRR, we adjusted our nutrient variables for total energy using the residual method.[100] Extending this method to additional potential confounders of dietary intake and response variables would allow for a less

biased interpretation of how dietary risk factors explain response variables. Studies with long-term longitudinal follow-up throughout childhood and adolescence (such as TEDDY and DAISY), would benefit from statistical methodological improvements to allow for repeated measures in the creation of patterns. To enhance the ability to compare patterns across populations, nutrient patterns should be further explored as they may be easier to harmonize across dietary assessment methods and cross-nationally.

While many general improvements could be made in the search for combinations of dietary risk factors that may influence T1D development, specific next steps generated from this dissertation include: 1) refinement and simplification of the nutrient pattern to enhance the ability to reproduce in other populations,[194] 2) replication of T1D-related nutrient patterns in the TEDDY study, and 3) investigation of additional systems-level biological changes (transcriptome, proteome, etc.) that might mediate the effects of the nutrient pattern on progression to T1D.

Dietary patterns, foods, and nutrients are inextricably linked,[195] yet their interconnection is insufficiently studied in etiologic research of T1D. As with many other nutrition-related complex diseases, multiple nutrients and foods have been weakly or inconsistently implicated in the disease process. Feasibility of a dietary intervention in young children has been established,[35,36] and at-risk populations for intervention are being collected via population-based screening studies.[41,42] Identification of an effective intervention is the crucial next step for nutrition research in T1D, which could grow from the foundation built in this dissertation project.

We established that investigating complex dietary intake and related molecular-level changes is an important part of the ongoing search for risk factors amenable to intervention in children. In addition to the future work specifically outlined above, advancement of the field studying the nutrition etiology of T1D might benefit from a framework shift from a one-at-a-time approach, to the "top down approach" that has successfully been used to develop dietary guidelines to prevent CVD.[196] After establishing evidence of dietary patterns associated with disease (the Mediterranean Diet), the role of foods (nuts, olive oil) and action of nutrients within those foods (fatty acids, polyphenols) were examined to provide context, explanation, and possible mechanisms.[195] This project provided evidence of system-level dysregulation of lipids in infancy that occur prior to the appearance of autoimmunity, and that may relate to dietary intake. As breastmilk was the most highly contributing food to the protective dietary pattern (**Figure IV-6**), understanding its nutrient content and potential mechanisms in T1D development might be a logical next step. We also identified patterns of nutrient intake including known and novel individual factors that influence progression from seroconversion to T1D. Characterization of the foods providing those nutrients, and investigation of the novel nutrients contributing to the pattern are important extensions from the pattern to the food and nutrient level. Replication and fine-tuning of the patterns identified here is a necessary next step in this process.

Complex, chronic diseases such as T1D are increasing worldwide. Despite general consensus that diet may act as an environmental (non-genetic) risk factor responsible for these increases and decades of investigation into putative "triggers" of disease processes, identification of a single causative dietary factor remains elusive in

T1D research. As demonstrated by this dissertation, expanding our methodological

repertoire beyond individual risk factor epidemiology will improve our understanding

of the disease process, and ultimately lead to more effective interventions to prevent,

reverse, or delay development of T1D.

REFERENCES

1. Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*. 2010;464(7293):1293-1300.

2. Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med*. 1986;314(21):1360-1368. doi:10.1056/NEJM198605223142106

3. Verge CF, Gianani R, Kawasaki E, et al. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes*. 1996;45(7):926–933.

4. Palmer JP, Asplin CM, Clemons P, et al. Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science*. 1983;222(4630):1337-1339. doi:10.1126/science.6362005

5. Baekkeskov S, Aanstoot HJ, Christgau S, et al. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature*. 1990;347(6289):151-156. doi:10.1038/347151a0

6. Rabin DU, Pleasic SM, Shapiro JA, et al. Islet cell antigen 512 is a diabetes-specific islet autoantigen related to protein tyrosine phosphatases. *J Immunol Baltim Md 1950*. 1994;152(6):3183-3188.

7. Wenzlau JM, Juhl K, Yu L, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A*. 2007;104(43):17040-17045. doi:10.1073/pnas.0705894104

8. Ziegler A-G, Bonifacio E, BABYDIAB-BABYDIET Study Group. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia*. 2012;55(7):1937-1943. doi:10.1007/s00125-012-2472-x

9. Parikka V, Näntö-Salonen K, Saarinen M, et al. Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia*. 2012;55(7):1926-1936. doi:10.1007/s00125-012-2523-3

10. Krischer JP, Lynch KF, Schatz DA, et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia*. 2015;58(5):980-987. doi:10.1007/s00125-015-3514-y

11. Rewers M, Hyöty H, Lernmark Å, et al. The Environmental Determinants of Diabetes in the Young (TEDDY) Study: 2018 Update. *Curr Diab Rep*. 2018;18(12):136. doi:10.1007/s11892-018-1113-2

12. Krischer JP, Lynch KF, Lernmark Å, et al. Genetic and Environmental Interactions Modify the Risk of Diabetes-Related Autoimmunity by 6 Years of Age: The TEDDY Study. *Diabetes Care*. 2017;40(9):1194-1202. doi:10.2337/dc17-0238

13. Ilonen J, Hammais A, Laine A-P, et al. Patterns of β-cell autoantibody appearance and genetic associations during the first years of life. *Diabetes*. 2013;62(10):3636-3640. doi:10.2337/db13-0300

14. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to Multiple Islet Autoantibodies and Risk of Progression to Diabetes in Children. *JAMA*. 2013;309(23):2473-2479. doi:10.1001/jama.2013.6285

15. Rich SS, Akolkar B, Concannon P, et al. Overview of the Type I Diabetes Genetics Consortium. *Genes Immun*. 2009;10(S1):S1-S4. doi:10.1038/gene.2009.84

16. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703-707. doi:10.1038/ng.381

17. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes*. 2008;57(4):1084-1092. doi:10.2337/db07-1331

18. Gillespie KM, Bain SC, Barnett AH, et al. The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet Lond Engl*. 2004;364(9446):1699-1700. doi:10.1016/S0140-6736(04)17357-1

19. Steck AK, Armstrong TK, Babu SR, Eisenbarth GS, Type 1 Diabetes Genetics Consortium. Stepwise or linear decrease in penetrance of type 1 diabetes with lower-risk HLA genotypes over the past 40 years. *Diabetes*. 2011;60(3):1045-1049. doi:10.2337/db10-1419

20. Association AD. 2. Classification and Diagnosis of Diabetes. *Diabetes Care*. 2017;40(Supplement 1):S11-S24. doi:10.2337/dc17-S005

21. Hirsch IB. Clinical review: Realistic expectations and practical use of continuous glucose monitoring for the endocrinologist. *J Clin Endocrinol Metab*. 2009;94(7):2232-2238. doi:10.1210/jc.2008-2625

22. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med*. 1993;329(14):977-986. doi:10.1056/NEJM199309303291401

23. Borch-Johnsen K, Kreiner S, Deckert T. Mortality of type 1 (insulin-dependent) diabetes mellitus in Denmark: a study of relative mortality in 2930 Danish type 1 diabetic patients diagnosed from 1933 to 1972. *Diabetologia*. 1986;29(11):767-772.

24. Harjutsalo V, Forsblom C, Groop P-H. Time trends in mortality in patients with type 1 diabetes: nationwide population based cohort study. *BMJ*. 2011;343:d5364. doi:10.1136/bmj.d5364

25. Miller RG, Secrest AM, Sharma RK, Songer TJ, Orchard TJ. Improvements in the life expectancy of type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complications study cohort. *Diabetes*. 2012;61(11):2987-2992. doi:10.2337/db11-1625

26. Centers for Disease Control and Prevention. Diabetes death rates among youths aged ≤ 19 years--United States, 1968-2009. *MMWR Morb Mortal Wkly Rep*. 2012;61(43):869-872.

27. Beran D, Yudkin JS, de Courten M. Access to care for patients with insulin-requiring diabetes in developing countries: case studies of Mozambique and Zambia. *Diabetes Care*. 2005;28(9):2136-2140.

28. International Diabetes Federation. *IDF Diabetes Atlas*. Brussels: International Diabetes Federation; 2015.

29. Gale EA. Dying of diabetes. *The Lancet*. 2006;368(9548):1626-1628. doi:10.1016/S0140-6736(06)69672-4

30. Dorman JS, Laporte RE, Kuller LH, et al. The Pittsburgh insulin-dependent diabetes mellitus (IDDM) morbidity and mortality study. Mortality results. *Diabetes*. 1984;33(3):271-276.

31. DCCT/EDIC Research Group, de Boer IH, Sun W, et al. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *N Engl J Med*. 2011;365(25):2366-2376. doi:10.1056/NEJMoa1111732

32. Nathan DM, Cleary PA, Backlund J-YC, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353(25):2643-2653. doi:10.1056/NEJMoa052187

33. Orchard TJ, Nathan DM, Zinman B, et al. Association Between 7 Years of Intensive Treatment of Type 1 Diabetes and Long-term Mortality. *JAMA*. 2015;313(1):45-53. doi:10.1001/jama.2014.16107

34. Orchard TJ, Secrest AM, Miller RG, Costacou T. In the absence of renal disease, 20 year mortality risk in type 1 diabetes is comparable to that of the general population: a report from the Pittsburgh Epidemiology of Diabetes Complications Study. *Diabetologia*. 2010;53(11):2312-2319. doi:10.1007/s00125-010-1860-3

35. Knip M, Åkerblom HK, Becker D, et al. Hydrolyzed infant formula and early β-cell autoimmunity: a randomized clinical trial. *JAMA*. 2014;311(22):2279-2287. doi:10.1001/jama.2014.5610

36. Knip M, Åkerblom HK, Taji EA, et al. Effect of Hydrolyzed Infant Formula vs Conventional Formula on Risk of Type 1 Diabetes: The TRIGR Randomized Clinical Trial. *JAMA*. 2018;319(1):38-48. doi:10.1001/jama.2017.19826

37. Skyler JS, Type 1 Diabetes TrialNet Study Group. Update on worldwide efforts to prevent type 1 diabetes. *Ann N Y Acad Sci*. 2008;1150:190-196. doi:10.1196/annals.1447.055

38. Gale E a. M, Bingley PJ, Emmett CL, Collier T, European Nicotinamide Diabetes Intervention Trial (ENDIT) Group. European Nicotinamide Diabetes Intervention Trial (ENDIT): a randomised controlled trial of intervention before the onset of type 1 diabetes. *Lancet Lond Engl*. 2004;363(9413):925-931. doi:10.1016/S0140-6736(04)15786-3

39. Näntö-Salonen K, Kupila A, Simell S, et al. Nasal insulin to prevent type 1 diabetes in children with HLA genotypes and autoantibodies conferring increased risk of disease: a double-blind, randomised controlled trial. *The Lancet*. 2008;372(9651):1746-1755. doi:10.1016/S0140-6736(08)61309-4

40. Shapiro AM, Lakey JR, Ryan EA, et al. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med*. 2000;343(4):230-238. doi:10.1056/NEJM200007273430401

41. Raab J, Haupt F, Scholz M, et al. Capillary blood islet autoantibody screening for identifying pre-type 1 diabetes in the general population: design and initial results of the Fr1da study. *BMJ Open*. 2016;6(5):e011144. doi:10.1136/bmjopen-2016-011144

42. Rasmussen CRG, Rewers M, Baxter J, et al. Population Screening for T1D and Celiac Disease—Autoimmunity Screening for Kids (ASK). *Diabetes*. 2018;67(Supplement 1):182-OR. doi:10.2337/db18-182-OR

43. Elding Larsson H, Vehik K, Bell R, et al. Reduced prevalence of diabetic ketoacidosis at diagnosis of type 1 diabetes in young children participating in longitudinal follow-up. *Diabetes Care*. 2011;34(11):2347-2352. doi:10.2337/dc11-1026

44. Winkler C, Schober E, Ziegler A-G, Holl RW. Markedly reduced rate of diabetic ketoacidosis at onset of type 1 diabetes in relatives screened for islet autoantibodies. *Pediatr Diabetes*. 2012;13(4):308-313. doi:10.1111/j.1399-5448.2011.00829.x

45. Duca LM, Wang B, Rewers M, Rewers A. Diabetic Ketoacidosis at Diagnosis of Type 1 Diabetes Predicts Poor Long-term Glycemic Control. *Diabetes Care*. 2017;40(9):1249-1255. doi:10.2337/dc17-0558

46. Libman I, Songer T, LaPorte R. How Many People in the U.S. Have IDDM? *Diabetes Care*. 1993;16(5):841-842. doi:10.2337/diacare.16.5.841

47. The DIAMOND Project Group. Incidence and trends of childhood Type 1 diabetes worldwide 1990–1999. *Diabet Med*. 2006;23(8):857-866. doi:10.1111/j.1464-5491.2006.01925.x

48. Patterson CC, Gyürüs E, Rosenbauer J, et al. Trends in childhood type 1 diabetes incidence in Europe during 1989-2008: evidence of non-uniformity over time in rates of increase. *Diabetologia*. 2012;55(8):2142-2147. doi:10.1007/s00125-012-2571-8

49. Lawrence JM, Imperatore G, Dabelea D, et al. Trends in Incidence of Type 1 Diabetes Among Non-Hispanic White Youth in the U.S., 2002–2009. *Diabetes*. 2014;63(11):3938-3945. doi:10.2337/db13-1891

50. Harjutsalo V, Sund R, Knip M, Groop P-H. Incidence of Type 1 Diabetes in Finland. *JAMA*. 2013;310(4):427-428. doi:10.1001/jama.2013.8399

51. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G, EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet Lond Engl*. 2009;373(9680):2027-2033. doi:10.1016/S0140-6736(09)60568-7

52. Berhan Y, Waernbaum I, Lind T, Möllsten A, Dahlquist G, Swedish Childhood Diabetes Study Group. Thirty years of prospective nationwide incidence of childhood type 1 diabetes: the accelerating increase by time tends to level off in Sweden. *Diabetes*. 2011;60(2):577-581. doi:10.2337/db10-0813

53. Tuomilehto J. The Emerging Global Epidemic of Type 1 Diabetes. *Curr Diab Rep*. 2013;13(6):795-804. doi:10.1007/s11892-013-0433-5

54. Tuomi T. Type 1 and type 2 diabetes: what do they have in common? *Diabetes*. 2005;54 Suppl 2:S40-45.

55. Kyvik KO, Nystrom L, Gorus F, et al. The epidemiology of Type 1 diabetes mellitus is not the same in young adults as in children. *Diabetologia*. 2004;47(3):377-384. doi:10.1007/s00125-004-1331-9

56. Harjutsalo V, Sjöberg L, Tuomilehto J. Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet Lond Engl*. 2008;371(9626):1777-1782. doi:10.1016/S0140-6736(08)60765-5

57. Lester FT. The clinical pattern of diabetes mellitus in Ethiopians. *Diabetes Care*. 1984;7(1):6-11.

58. Kalk WJ, Huddle KR, Raal FJ. The age of onset and sex distribution of insulin-dependent diabetes mellitus in Africans in South Africa. *Postgrad Med J*. 1993;69(813):552-556.

59. Marshall SL, Edidin D, Sharma V, Ogle G, Arena VC, Orchard T. Current clinical status, glucose control, and complication rates of children and youth with type 1 diabetes in Rwanda. *Pediatr Diabetes*. 2013;14(3):217-226. doi:10.1111/pedi.12007

60. Karvonen M, Pitkäniemi M, Pitkäniemi J, Kohtamäki K, Tajima N, Tuomilehto J. Sex difference in the incidence of insulin-dependent diabetes mellitus: an analysis of the recent epidemiological data. World Health Organization DIAMOND Project Group. *Diabetes Metab Rev*. 1997;13(4):275-291.

61. Moltchanova EV, Schreier N, Lammi N, Karvonen M. Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. *Diabet Med J Br Diabet Assoc*. 2009;26(7):673-678. doi:10.1111/j.1464-5491.2009.02743.x

62. Patterson CC, Gyürüs E, Rosenbauer J, et al. Seasonal variation in month of diagnosis in children with type 1 diabetes registered in 23 European centers during 1989-2008: little short-term influence of sunshine hours or average temperature. *Pediatr Diabetes*. 2015;16(8):573-580. doi:10.1111/pedi.12227

63. Kahn HS, Morgan TM, Case LD, et al. Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. *Diabetes Care*. 2009;32(11):2010-2015. doi:10.2337/dc09-0891

64. Barnett AH, Eff C, Leslie RD, Pyke DA. Diabetes in identical twins. A study of 200 pairs. *Diabetologia*. 1981;20(2):87-93.

65. Kaprio J, Tuomilehto J, Koskenvuo M, et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia*. 1992;35(11):1060-1067.

66. Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. Concordance for islet autoimmunity among monozygotic twins. *N Engl J Med*. 2008;359(26):2849-2850. doi:10.1056/NEJMc0805398

67. Soltesz G, Patterson C, Dahlquist G, EURODIAB Study Group. Worldwide childhood type 1 diabetes incidence – what can we learn from epidemiology? *Pediatr Diabetes*. 2007;8:6-14. doi:10.1111/j.1399-5448.2007.00280.x

68. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *The Lancet*. 2014;383(9911):69-82. doi:10.1016/S0140-6736(13)60591-7

69. Lamb MM, Myers MA, Barriga K, Zimmet PZ, Rewers M, Norris JM. Maternal diet during pregnancy and islet autoimmunity in offspring. *Pediatr Diabetes*. 2008;9(2):135-141. doi:10.1111/j.1399-5448.2007.00311.x

70. Virtanen S, Uusitalo L, Kenward M, et al. Maternal food consumption during pregnancy and risk of advanced β-cell autoimmunity in the offspring. *Pediatr Diabetes*. 2011;12(2):95-99. doi:10.1111/j.1399-5448.2010.00668.x

71. Niinistö S, Takkinen H-M, Uusitalo L, et al. Maternal dietary fatty acid intake during pregnancy and the risk of preclinical and clinical type 1 diabetes in the offspring. *Br J Nutr*. 2014;111(05):895–903. doi:10.1017/S0007114513003073

72. Kostraba JN, Cruickshanks KJ, Lawler-Heavner J, et al. Early Exposure to Cow's Milk and Solid Foods in Infancy, Genetic Predisposition, and Risk of IDDM. *Diabetes*. 1993;42(2):288-295. doi:10.2337/diab.42.2.288

73. Frederiksen B, Kroehl M, Lamb MM, et al. Infant Exposures and Development of Type 1 Diabetes Mellitus. *JAMA Pediatr*. 2013;167(9):808-815. doi:10.1001/jamapediatrics.2013.317

74. Hyppönen E, Kenward MG, Virtanen SM, et al. Infant feeding, early weight gain, and risk of type 1 diabetes. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes Care*. 1999;22(12):1961-1965. doi:10.2337/diacare.22.12.1961

75. Virtanen SM, Räsänen L, Ylönen K, et al. Early Introduction of Dairy Products Associated with Increased Risk of IDDM in Finnish Children. *Diabetes*. 1993;42(12):1786-1790. doi:10.2337/diab.42.12.1786

76. Uusitalo U, Lee H-S, Andrén Aronsson C, et al. Early Infant Diet and Islet Autoimmunity in the TEDDY Study. *Diabetes Care*. 2018;41(3):522-530. doi:10.2337/dc17-1983

77. Lamb MM, Miller M, Seifert JA, et al. The effect of childhood cow's milk intake and HLA-DR genotype on risk of islet autoimmunity and type 1 diabetes: The Diabetes Autoimmunity Study in the Young. *Pediatr Diabetes*. 2015;16(1):31-38. doi:10.1111/pedi.12115

78. Virtanen SM, Nevalainen J, Kronberg-Kippilä C, et al. Food consumption and advanced β cell autoimmunity in young children with HLA-conferred susceptibility to type 1 diabetes: a nested case-control design. *Am J Clin Nutr*. 2012;95(2):471-478. doi:10.3945/ajcn.111.018879

79. Miller MR, Yin X, Seifert J, et al. Erythrocyte membrane omega-3 fatty acid levels and omega-3 fatty acid intake are not associated with conversion to type 1 diabetes in children with islet autoimmunity: The Diabetes Autoimmunity Study in the Young (DAISY). *Pediatr Diabetes*. 2011;12(8):669-675. doi:10.1111/j.1399-5448.2011.00760.x

80. Norris JM, Yin X, Lamb MM, et al. Omega-3 polyunsaturated fatty acid intake and islet autoimmunity in children at increased risk for type 1 diabetes. *JAMA*. 2007;298(12):1420-1428. doi:10.1001/jama.298.12.1420

81. Norris JM, Kroehl M, Fingerlin TE, et al. Erythrocyte membrane docosapentaenoic acid levels are associated with islet autoimmunity: the Diabetes Autoimmunity Study in the Young. *Diabetologia*. 2014;57(2):295-304. doi:10.1007/s00125-013-3106-7

82. Lamb MM, Yin X, Barriga K, et al. Dietary Glycemic Index, Development of Islet Autoimmunity, and Subsequent Progression to Type 1 Diabetes in Young Children. *J Clin Endocrinol Metab*. 2008;93(10):3936-3942. doi:10.1210/jc.2008-0886

83. Lamb MM, Frederiksen B, Seifert JA, Kroehl M, Rewers M, Norris JM. Sugar intake is associated with progression from islet autoimmunity to type 1 diabetes: the Diabetes Autoimmunity Study in the Young. *Diabetologia*. 2015;58(9):2027-2034. doi:10.1007/s00125-015-3657-x

84. Beyerlein A, Liu X, Uusitalo UM, et al. Dietary intake of soluble fiber and risk of islet autoimmunity by 5 y of age: results from the TEDDY study. *Am J Clin Nutr*. 2015;102(2):345-352. doi:10.3945/ajcn.115.108159

85. Simpson M, Brady H, Yin X, et al. No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia*. 2011;54(11):2779-2788. doi:10.1007/s00125-011-2278-2

86. Norris JM, Lee H-S, Frederiksen B, et al. Plasma 25-Hydroxyvitamin D Concentration and Risk of Islet Autoimmunity. *Diabetes*. 2018;67(1):146-154. doi:10.2337/db17-0802

87. Norris JM. Infant and Childhood Diet and Type 1 Diabetes Risk: Recent Advances and Prospects. *Curr Diab Rep*. 2010;10(5):345-349. doi:10.1007/s11892-010-0131-5

88. Lund-Blix NA, Dong F, Mårild K, et al. Gluten Intake and Risk of Islet Autoimmunity and Progression to Type 1 Diabetes in Children at Increased Risk of the Disease: The Diabetes Autoimmunity Study in the Young (DAISY). *Diabetes Care*. February 2019:dc182315. doi:10.2337/dc18-2315

89. Zheng Y, Yu B, Alexander D, Steffen LM, Boerwinkle E. Human Metabolome Associates With Dietary Intake Habits Among African Americans in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2014;179(12):1424-1433. doi:10.1093/aje/kwu073

90. Watkins SM, German JB. Toward the implementation of metabolomic assessments of human health and nutrition. *Curr Opin Biotechnol*. 2002;13(5):512-516. doi:10.1016/S0958-1669(02)00363-4

91. Gibbons H, O'Gorman A, Brennan L. Metabolomics as a tool in nutritional research: *Curr Opin Lipidol*. 2015;26(1):30-34. doi:10.1097/MOL.0000000000000140

92. Altmaier E, Kastenmüller G, Römisch-Margl W, et al. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol*. 2011;26(2):145-156. doi:10.1007/s10654-010-9524-7

93. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr*. 2011;93(2):314-321. doi:10.3945/ajcn.110.000950

94. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448-453. doi:10.1038/nm.2307

95. Pflueger M, Seppänen-Laakso T, Suortti T, et al. Age- and Islet Autoimmunity–Associated Differences in Amino Acid and Lipid Metabolites in Children at Risk for Type 1 Diabetes. *Diabetes*. 2011;60(11):2740-2747. doi:10.2337/db10-1652

96. Orešič M, Simell S, Sysi-Aho M, et al. Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J Exp Med*. 2008;205(13):2975-2984. doi:10.1084/jem.20081800

97. Balderas C, Rupérez FJ, Ibañez E, et al. Plasma and urine metabolic fingerprinting of type 1 diabetic children: General. *ELECTROPHORESIS*. September 2013:n/a-n/a. doi:10.1002/elps.201300062

98. Frohnert BI, Rewers MJ. Metabolomics in childhood diabetes. *Pediatr Diabetes*. 2016;17(1):3-14. doi:10.1111/pedi.12323

99. Freedman LS, Kipnis V, Schatzkin A, Tasevska N, Potischman N. Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? *Epidemiol Perspect Innov EPI*. 2010;7:2. doi:10.1186/1742-5573-7-2

100. Willett W. *Nutritional Epidemiology*. OUP USA; 2012.

101. Orešič M, Gopalacharyulu P, Mykkänen J, et al. Cord Serum Lipidome in Prediction of Islet Autoimmunity and Type 1 Diabetes. *Diabetes*. 2013;62(9):3268-3274. doi:10.2337/db13-0159

102. Torre DL, Seppänen-Laakso T, Larsson HE, et al. Decreased Cord-Blood Phospholipids in Young Age–at–Onset Type 1 Diabetes. *Diabetes*. 2013;62(11):3951-3956. doi:10.2337/db13-0215

103. Lamichhane S, Ahonen L, Dyrlund TS, et al. Dynamics of Plasma Lipidome in Progression to Islet Autoimmunity and Type 1 Diabetes – Type 1 Diabetes Prediction and Prevention Study (DIPP). *Sci Rep*. 2018;8(1):10635. doi:10.1038/s41598-018-28907-8

104. Jørgenrud B, Stene LC, Tapia G, et al. Longitudinal plasma metabolic profiles, infant feeding, and islet autoimmunity in the MIDIA study. *Pediatr Diabetes*. 2017;18(2):111-119. doi:10.1111/pedi.12360

105. Biesbroek S, A DL van der, Brosens MC, et al. Identifying cardiovascular risk factor–related dietary patterns with reduced rank regression and random forest in the EPIC-NL cohort. *Am J Clin Nutr*. May 2015:ajcn092288. doi:10.3945/ajcn.114.092288

106. Lucas M, Chocano-Bedoya P, Shulze MB, et al. Inflammatory dietary pattern and risk of depression among women. *Brain Behav Immun*. 2014;36:46-53. doi:10.1016/j.bbi.2013.09.014

107. Lamichhane AP, Liese AD, Urbina EM, et al. Associations of dietary intake patterns identified using reduced rank regression with markers of arterial stiffness among youth with type 1 diabetes. *Eur J Clin Nutr*. 2014;68(12):1327–1333.

108. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13(1):3-9.

109. Martinez-Gonzalez MA, Bes-Rastrollo M. Dietary patterns, Mediterranean diet, and cardiovascular disease: *Curr Opin Lipidol*. 2014;25(1):20-26. doi:10.1097/MOL.0000000000000044

110. The TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr Diabetes*. 2007;8(5):286-298. doi:10.1111/j.1399-5448.2007.00269.x

111. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia*. 1996;39(7):807-812. doi:10.1007/s001250050514

112. Uusitalo U, Kronberg-Kippila C, Aronsson CA, et al. Food composition database harmonization for between-country comparisons of nutrient data in the TEDDY Study. *J Food Compos Anal Off Publ U N Univ Int Netw Food Data Syst*. 2011;24(4-5):494-505. doi:10.1016/j.jfca.2011.01.012

113. Joslowski G, Yang J, Andrén Aronsson C, et al. Development of a harmonized food grouping system for between-country comparisons in the TEDDY Study. *J Food Compos Anal*. 2017;63:79-88. doi:10.1016/j.jfca.2017.07.037

114. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci*. 2008;1150(1):1-13. doi:10.1196/annals.1447.062

115. Rothman KJ, Professor TLLA, Greenland S. *Modern Epidemiology*. Third, Mid-cycle revision edition. Philadelphia: LWW; 2012.

116. Lee H-S, Burkhardt BR, McLeod W, et al. Biomarker discovery study design for type 1 diabetes in The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Diabetes Metab Res Rev*. 2014;30(5):424-434. doi:10.1002/dmrr.2510

117. Lamb MM, Ross CA, Brady HL, Norris JM. Comparison of children's diets as reported by the child via the Youth/Adolescent Questionnaire and the parent via the Willett food-frequency questionnaire. *Public Health Nutr*. 2007;10(07):663–670. doi:10.1017/S1368980007226059

118. Parrish LA, Marshall JA, Krebs NF, Rewers M, Norris JM. Validation of a Food Frequency Questionnaire in Preschool Children: *Epidemiology*. 2003;14(2):213-217. doi:10.1097/01.EDE.0000041256.12192.23

119. Brady H, Lamb MM, Sokol RJ, et al. Plasma micronutrients are associated with dietary intake and environmental tobacco smoke exposure in a paediatric population. *Public Health Nutr*. 2007;10(07):712–718. doi:10.1017/S1368980007662296

120. Dunn WB, Broadhurst D, Begley P, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. 2011;6(7):1060-1083. doi:10.1038/nprot.2011.335

121. Benjamini YH, Hochberg Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *ResearchGate*. 1995;57:289-300. doi:10.2307/2346101

122. Barupal DK, Fiehn O. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Sci Rep*. 2017;7(1):14567. doi:10.1038/s41598-017-15231-w

123. Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U, Boeing H. Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *Am J Epidemiol*. 2004;159(10):935-944. doi:10.1093/aje/kwh134

124. van der Voet H. Comparing the predictive accuracy of models using a simple randomization test. *Chemom Intell Lab Syst*. 1994;25(2):313-323. doi:10.1016/0169-7439(94)85050-X

125. Kröger J, Ferrari P, Jenab M, et al. Specific food group combinations explaining the variation in intakes of nutrients and other important food components in the European Prospective Investigation into Cancer and Nutrition: an application of the reduced rank regression method. *Eur J Clin Nutr*. 2009;63(S4):S263-S274. doi:10.1038/ejcn.2009.85

126. Floegel A, von Ruesten A, Drogan D, et al. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr*. 2013;67(10):1100–1108.

127. Emmett PM, Jones LR, Northstone K. Dietary patterns in the Avon Longitudinal Study of Parents and Children. *Nutr Rev*. 2015;73(suppl 3):207-230. doi:10.1093/nutrit/nuv055

128. Meyer J, Döring A, Herder C, Roden M, Koenig W, Thorand B. Dietary patterns, subclinical inflammation, incident coronary heart disease and mortality in middle-aged men from the MONICA/KORA Augsburg cohort study. *Eur J Clin Nutr*. 2011;65(7):800-807. doi:10.1038/ejcn.2011.37

129. Mukherjee A, Chen K, Wang N, Zhu J. On the degrees of freedom of reduced-rank estimators in multivariate regression. *Biometrika*. 2015;102(2):457-477. doi:10.1093/biomet/asu067

130. Appannah G, Pot GK, O'Sullivan TA, Oddy WH, Jebb SA, Ambrosini GL. The reliability of an adolescent dietary pattern identified using reduced-rank regression: comparison of a FFQ and 3 d food record. *Br J Nutr*. 2014;112(04):609–615. doi:10.1017/S0007114514001111

131. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. *Am J Clin Nutr*. 2009;90(4):1075-1083. doi:10.3945/ajcn.2009.28009

132. Ambrosini GL, Emmett PM, Northstone K, Howe LD, Tilling K, Jebb SA. Identification of a dietary pattern prospectively associated with increased adiposity during childhood and adolescence. *Int J Obes*. 2012;36(10):1299–1305.

133. Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *Am J Clin Nutr*. 2005;82(3):675-684.

134. Jolliffe I. Principal Component Analysis. In: *Wiley StatsRef: Statistics Reference Online.* John Wiley & Sons, Ltd; 2014. http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06472/abstract. Accessed December 23, 2016.

135. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *J R Stat Soc Ser C Appl Stat*. 1979;28(1):100-108. doi:10.2307/2346830

136. Kohonen T. The self-organizing map. *Proc IEEE*. 1990;78(9):1464-1480. doi:10.1109/5.58325

137. Wold H. Estimation of Principal Components and Related Models by Iterative Least squares. In: *Multivariate Analysis.* Academic Press; 1966:391-420.

138. Nettleton JA, Steffen LM, Schulze MB, et al. Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr*. 2007;85(6):1615-1625.

139. DiBello JR, Kraft P, McGarvey ST, Goldberg R, Campos H, Baylin A. Comparison of 3 Methods for Identifying Dietary Patterns Associated With Risk of Disease. *Am J Epidemiol*. 2008;168(12):1433-1443. doi:10.1093/aje/kwn274

140. Northstone K, Emmett PM. Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. *Br J Nutr*. 2008;100(05):1069. doi:10.1017/S0007114508968264

141. Northstone K, Emmett PM. A comparison of methods to assess changes in dietary patterns from pregnancy to 4 years post-partum obtained using principal components analysis. *Br J Nutr*. 2008;99(5):1099-1106. doi:10.1017/S0007114507842802

142. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth MEJ. Longitudinal changes in dietary patterns during adult life. *Br J Nutr*. 2006;96(4):735-744.

143. Harrington JM, Dahly DL, Fitzgerald AP, Gilthorpe MS, Perry IJ. Capturing changes in dietary patterns among older adults: a latent class analysis of an ageing Irish cohort. *Public Health Nutr*. 2014;17(12):2674-2686. doi:10.1017/S1368980014000111

144. Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA. Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on -Omic Technologies. *Am J Epidemiol*. 2014;180(2):129-139. doi:10.1093/aje/kwu143

145. Mäkinen V-P, Ala-Korpela M. Metabolomics of aging requires large-scale longitudinal studies with replication. *Proc Natl Acad Sci*. 2016;113(25):E3470-E3470. doi:10.1073/pnas.1607062113

146. Drake SK, Bowen RAR, Remaley AT, Hortin GL. Potential Interferences from Blood Collection Tubes in Mass Spectrometric Analyses of Serum Polypeptides. *Clin Chem*. 2004;50(12):2398-2401. doi:10.1373/clinchem.2004.040303

147. Gonzalez-Covarrubias V, Dane A, Hankemeier T, Vreeken RJ. The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC–MS lipidomic profiling. *Metabolomics*. 2013;9(2):337-348. doi:10.1007/s11306-012-0450-4

148. Katsarou A, Gudbjörnsdottir S, Rawshani A, et al. Type 1 diabetes mellitus. *Nat Rev Dis Primer*. 2017;3:17016. doi:10.1038/nrdp.2017.16

149. Virtanen SM, Niinistö S, Nevalainen J, et al. Serum fatty acids and risk of advanced β-cell autoimmunity: a nested case–control study among children with HLA-conferred susceptibility to type I diabetes. *Eur J Clin Nutr*. 2010;64(8):792–799.

150. Niinistö S, Takkinen H-M, Erlund I, et al. Fatty acid status in infancy is associated with the risk of type 1 diabetes-associated autoimmunity. *Diabetologia*. 2017;60(7):1223-1233. doi:10.1007/s00125-017-4280-9

151. Hyppönen E, Läärä E, Reunanen A, Järvelin MR, Virtanen SM. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet Lond Engl*. 2001;358(9292):1500-1503. doi:10.1016/S0140-6736(01)06580-1

152. Mäkinen M, Mykkänen J, Koskinen M, et al. Serum 25-Hydroxyvitamin D Concentrations in Children Progressing to Autoimmunity and Clinical Type 1 Diabetes. *J Clin Endocrinol Metab*. 2016;101(2):723-729. doi:10.1210/jc.2015-3504

153. Fiehn O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol*. 2016;114:30.4.1-30.4.32. doi:10.1002/0471142727.mb3004s114

154. Kind T, Wohlgemuth G, Lee DY, et al. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem*. 2009;81(24):10038-10048. doi:10.1021/ac9019522

155. Cajka T, Smilowitz JT, Fiehn O. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. *Anal Chem*. 2017;89(22):12360-12368. doi:10.1021/acs.analchem.7b03404

156. Box GE, Cox DR. An analysis of transformations. *J R Stat Soc Ser B Methodol*. 1964:211–252.

157. Trumbo P, Schlicker S, Yates AA, Poos M, Food and Nutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. *J Am Diet Assoc*. 2002;102(11):1621-1630.

158. Malm L, Tybring G, Moritz T, Landin B, Galli J. Metabolomic Quality Assessment of EDTA Plasma and Serum Samples. *Biopreservation Biobanking*. 2016;14(5):416-423. doi:10.1089/bio.2015.0092

159. Bender R, Lange S. Adjusting for multiple testing—when and how? *J Clin Epidemiol*. 2001;54(4):343-349. doi:10.1016/S0895-4356(00)00314-0

160. Glick N, Fischer M. Adipic aciduria, a dietary artefact. *J Inherit Metab Dis*. 1991;14(5):849-850.

161. Salminen A, Kaarniranta K, Hiltunen M, Kauppinen A. Krebs cycle dysfunction shapes epigenetic landscape of chromatin: novel insights into mitochondrial regulation of aging process. *Cell Signal*. 2014;26(7):1598-1603. doi:10.1016/j.cellsig.2014.03.030

162. Paul DS, Teschendorff AE, Dang MAN, et al. Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat Commun*. 2016;7:ncomms13555. doi:10.1038/ncomms13555

163. Boslem E, Meikle PJ, Biden TJ. Roles of ceramide and sphingolipids in pancreatic β-cell function and dysfunction. *Islets*. 2012;4(3):177-187. doi:10.4161/isl.20102

164. Holm LJ, Krogvold L, Hasselby JP, et al. Abnormal islet sphingolipid metabolism in type 1 diabetes. *Diabetologia*. 2018;61(7):1650-1661. doi:10.1007/s00125-018-4614-2

165. Maceyka M, Spiegel S. Sphingolipid metabolites in inflammatory disease. *Nature*. 2014;510(7503):58-67. doi:10.1038/nature13475

166. Norris GH, Blesso CN. Dietary and Endogenous Sphingolipid Metabolism in Chronic Inflammation. *Nutrients*. 2017;9(11). doi:10.3390/nu9111180

167. Mazzei JC, Zhou H, Brayfield BP, Hontecillas R, Bassaganya-Riera J, Schmelz EM. Suppression of intestinal inflammation and inflammation-driven colon cancer in mice by dietary sphingomyelin: importance of peroxisome proliferator-activated receptor γ expression. *J Nutr Biochem*. 2011;22(12):1160-1171. doi:10.1016/j.jnutbio.2010.09.017

168. Pörksen S, Nielsen LB, Mortensen HB, et al. Variation within the PPARG gene is associated with residual beta-cell function and glycemic control in children and adolescents during the first year of clinical type 1 diabetes. *Pediatr Diabetes*. 2008;9(4 Pt 1):297-302. doi:10.1111/j.1399-5448.2008.00398.x

169. Zeisel SH. Metabolic crosstalk between choline/1-carbon metabolism and energy homeostasis. *Clin Chem Lab Med CCLM FESCC*. 2013;51(3):467-475. doi:10.1515/cclm-2012-0518

170. Davenport C, Yan J, Taesuwan S, et al. Choline intakes exceeding recommendations during human lactation improve breast milk choline content by increasing PEMT pathway metabolites. *J Nutr Biochem*. 2015;26(9):903-911. doi:10.1016/j.jnutbio.2015.03.004

171. Holmes-McNary MQ, Cheng WL, Mar MH, Fussell S, Zeisel SH. Choline and choline esters in human and rat milk and in infant formulas. *Am J Clin Nutr*. 1996;64(4):572-576. doi:10.1093/ajcn/64.4.572

172. Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. *Curr Opin Clin Nutr Metab Care*. 2016;19(5):341-346. doi:10.1097/MCO.0000000000000308

173. Sedlmeier A, Kluttig A, Giegling I, et al. The human metabolic profile reflects macro- and micronutrient intake distinctly according to fasting time. *Sci Rep*. 2018;8(1):12262. doi:10.1038/s41598-018-30764-4

174. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*. 2013;309(23):2473-2479. doi:10.1001/jama.2013.6285

175. Mayer-Davis EJ, Lawrence JM, Dabelea D, et al. Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002–2012. *N Engl J Med*. 2017;376(15):1419-1429. doi:10.1056/NEJMoa1610187

176. Ziegler A-G, Pflueger M, Winkler C, et al. Accelerated progression from islet autoimmunity to diabetes is causing the escalating incidence of type 1 diabetes in young children. *J Autoimmun*. 2011;37(1):3-7. doi:10.1016/j.jaut.2011.02.004

177. Virtanen SM. Dietary factors in the development of type 1 diabetes. *Pediatr Diabetes*. 2016;17(S22):49-55. doi:10.1111/pedi.12341

178. Raab J, Giannopoulou EZ, Schneider S, et al. Prevalence of vitamin D deficiency in pre-type 1 diabetes and its association with disease progression. *Diabetologia*. 2014;57(5):902-908. doi:10.1007/s00125-014-3181-4

179. Martinez-Gonzalez MA, Bes-Rastrollo M. Dietary patterns, Mediterranean diet, and cardiovascular disease: *Curr Opin Lipidol*. 2014;25(1):20-26. doi:10.1097/MOL.0000000000000044

180. Norris JM, Barriga K, Klingensmith G, et al. Timing of initial cereal exposure in infancy and risk of islet autoimmunity. *JAMA*. 2003;290(13):1713-1720. doi:10.1001/jama.290.13.1713

181. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49(5):1137-1146. doi:10.1194/jlr.D700041-JLR200

182. Showalter MR, Nonnecke EB, Linderholm AL, et al. Obesogenic diets alter metabolism in mice. Fornace AJ, ed. *PLOS ONE*. 2018;13(1):e0190632. doi:10.1371/journal.pone.0190632

183. Fiehn O, Wohlgemuth G, Scholz M, et al. Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J Cell Mol Biol*. 2008;53(4):691-704. doi:10.1111/j.1365-313X.2007.03387.x

184. Skogerson K, Wohlgemuth G, Barupal DK, Fiehn O. The volatile compound BinBase mass spectral database. *BMC Bioinformatics*. 2011;12:321. doi:10.1186/1471-2105-12-321

185. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;12(6):523-526. doi:10.1038/nmeth.3393

186. Kind T, Liu K-H, Yup Lee D, DeFelice B, Meissen JK, Fiehn O. LipidBlast - in-silico tandem mass spectrometry database for lipid identification. *Nat Methods*. 2013;10(8):755-758. doi:10.1038/nmeth.2551

187. DeFelice BC, Mehta SS, Samra S, et al. Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing. *Anal Chem*. 2017;89(6):3250-3255. doi:10.1021/acs.analchem.6b04372

188. Fan S, Kind T, Cajka T, et al. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal Chem*. February 2019. doi:10.1021/acs.analchem.8b05592

189. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

190. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300.

191. Cobianchi L, Fornoni A, Pileggi A, et al. Riboflavin inhibits IL-6 expression and p38 activation in islet cells. *Cell Transplant*. 2008;17(5):559-566.

192. Varsha MKNS, Thiagarajan R, Manikandan R, Dhanasekaran G. Vitamin K1 alleviates streptozotocin-induced type 1 diabetes by mitigating free radical stress, as well as inhibiting NF-κB activation and iNOS expression in rat pancreas. *Nutr Burbank Los Angel Cty Calif*. 2015;31(1):214-222. doi:10.1016/j.nut.2014.05.012

193. Lamichhane S, Ahonen L, Sparholt Dyrlund T, et al. Cord-Blood Lipidome in Progression to Islet Autoimmunity and Type 1 Diabetes. *Biomolecules*. 2019;9(1):33. doi:10.3390/biom9010033

194. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *Br J Nutr*. 2003;89(03):409–418. doi:10.1079/BJN2002778

195. Tapsell LC, Neale EP, Satija A, Hu FB. Foods, Nutrients, and Dietary Patterns: Interconnections and Implications for Dietary Guidelines. *Adv Nutr*. 2016;7(3):445-454. doi:10.3945/an.115.011718

196. Jacobs DR, Tapsell LC. Food synergy: the key to a healthy diet. *Proc Nutr Soc*. 2013;72(2):200-206. doi:10.1017/S0029665112003011

TEDDY AND DAISY FOOD GROUPINGS

TEDDY Food Groups and Subgroups	DAISY Food Groups and Serving Size
CEREALS	
Rice	White rice (1 cup)
	Brown rice (1 cup)
Wheat	Dark bread (1 slice)
	White bread (1 slice)
	Wheat germ (1 tbs)
Rye	
Oats	Cooked breakfast cereal e.g. oatmeal (1 cup)
Barley	
Corn	
Other gluten free flours and starches	
FRUIT AND BERRIES	
Citrus fruit	Grapefruit (1/2)
	Oranges (1)
Apple	Fresh apples or pears (1)
Berries	Blueberries, fresh, frozen or canned (1/2 cup)
	Strawberries, fresh, frozen or canned (1/2 cup)
Other fruits	Bananas (1)
	Cantaloupe (1/4 melon)
	Watermelon (1 slice)
	Peaches, Apricots or Plums (1 fresh, or 1/2 cup canned)
Canned fruits	
Dried fruits and berries	Prunes (1/2 cup)
	Raisins (1 oz. or small pack) or grapes
Citrus juice	Grapefruit juice (small glass)
	Orange juice (small glass)
Apple juice	Apple juice or cider (small glass)
Berry juices	
Other fruit juices	Other fruit juices (small glass)
Mixed Juices	
Juices, mixture of vegetable and fruit	
VEGETABLES	
Potatoes	French fried potatoes (4 oz.)
	Potato chips (small bag or 1 oz.)
	Potatoes, baked, boiled (1) or mashed (1 cup)

Roots, sweet potatoes	Carrots (1 whole or 1/2 cup cooked)
	Yams or sweet potatoes (1/2 cup)
Leafy vegetables	Iceberg or head lettuce (1 serving)
	Kale, mustard or chard greens (1/2 cup)
	Romaine or leaf lettuce (1 serving)
	Spinach, cooked (1/2 cup)
	Spinach, raw as in salad
Cabbages	Broccoli (1/2 cup)
	Brussels sprouts (1/2 cup)
	Cabbage or coleslaw (1/2 cup)
	Cauliflower (1/2 cup)
Fruit vegetables	Corn (1 ear or 1/2 cup frozen or canned)
	Tomatoes (1)
	Yellow (winter) squash (1/2 cup)
	Eggplant, Zucchini, or other summer squash (1/2 cup)
Onions	
Mushrooms	
Legumes, beans, peas	Beans or lentils, baked or dried (1/2 cup)
	Peas or lima beans (1/2 cup fresh, frozen, canned)
	String beans (1/2 cup)
Vegetables, dried	
Vegetables, canned	
Vegetable juices	Tomato juice (small glass)
SOY	
Soy beans	
Soy products	Tofu or soybeans (3-4 oz.)
Soy milk	
Soy dessert	
Soy sausage	
Soy sauce	
NUTS, SEEDS	
Nuts, seeds	Nuts (small packet or 1 oz.)
	Peanut butter (1 Tbs)
FATS AND OILS	
Vegetable oils	
Solid vegetable fat	
Fish oil	
Animal fats	
Butter	Butter (pat) added to food or bread, exclude use in cooking

Margarines	Margarine (pat) added to food or bread, exclude use in cooking
Butter-margarine mixtures	
MILK AND MILK PRODUCTS	
Breast milk	
Fat-free milk	
Low-fat milk	Skim or low fat milk (8oz. Glass)
High-fat milk	Whole milk (8 oz. glass)
Other animal milk	
Creams	Cream e.g. coffee, whipped (TBS)
Ice cream	Milk shake or frappe (1)
	Ice cream (1/2 cup)
	Ice cream sundae
Sour milk and sour milk products	Yogurt (1cup)
Cheese	Other cheese, e.g., American, etc., plain or as part of a dish (1 slice or 1oz. Serving)
	Cream cheese (1 oz)
	Cottage or Ricotta chz (1/2 cup)
Whey	
NON-DAIRY PRODUCTS	
Non-dairy products	Sherbet or ice milk (1/2 cup
	Non-Dairy Coffee Whitener (tsp)
MEAT AND MEAT PRODUCTS	
Pork	
Beef	Hamburger (1 patty)
	Beef, pork, or lamb as a main dish (4-6oz.)
Poultry	Chicken or turkey, with skin (4-6 oz.)
	Chicken or turkey, without skin (4-6 oz.)
Lamb, goat, horse	
Game	
Processed meats and sausages	Processed meats, e.g., sausage, Salami, bologna, etc. (piece or slice)
	Bacon (2 slices)
Organ meats/offals	Liver (3-4 oz.)
FISH AND FISH PRODUCTS	
Fresh and frozen fish	Other fish (3-5 oz.)
	Dark meat fish (3-5 oz.) salmon, sardines
Processed fish	
Canned fish	Canned tuna fish (3-4 oz.)
Shellfish, other seafood	Shrimp, lobster, scallops as a main dish
EGGS	
Eggs	Eggs (1)

BEVERAGES	
Coffee	
Tea	
Light Beverages including cola	Other low calorie carbonated beverage, e.g. Fresca, Diet 7-up, Diet Ginger Ale
	Low-calorie caffeine-free cola, e.g. pepsi free
	Low-calorie cola e.g. tab with caffeine
Sugar sweetened beverages including cola	Hawaiian punch, lemonade, or other non-carbonated fruit drinks (1 glass, bottle, can)
	Other carbonated beverage with sugar e.g. 7-up, ginger-ale
	Coke, pepsi, or other cola with sugar
	Caffeine free coke, pepsi, or other cola with sugar
Alcohol	
CONFECTIONARY	
Sweets	Candy without chocolate (1 oz.)
Chocolate	chocolate pieces e.g. Hershey's, M&M's
TEDDY ONLY: INFANT FORMULAS	
Regular cow's milk based infant formulas	
Partially hydrolyzed cow's milk based infant formulas	
Fully hydrolyzed cow's milk based infant formulas and amino acid (elemental) formulas	
Soy base infant formulas	
Other non-dairy infant formulas	
Other animal (dairy) infant formulas	
TEDDY ONLY: MISC.	
Ketchup	
	DAISY ONLY: COMPLEX FOODS
	Brownies (1)
	Cake ready made (slice)
	Cake homebaked (slice)
	other Candy bars, e.g. snickers, Milky Way, mars
	Cookies ready made (1)
	Chowder or cream soup (1 cup)
	Cold breakfast cereal (1 cup)
	Cookies home baked (1)
	Corn chips, e.g. fritos, doritos (small bag or 1 oz.)

	Crackers, triskets, wheat thins,etc. (1)
	Doughnuts (1)
	English muffins, Bagels, or rolls (1)
	Hot dogs (1)
	Jams, Jellies, preserves, syrup, or honey (1 Tbs)
	Mayo or other creamy salad dressing (1 TBS)
	Mixed vegetables (1/2 cup)
	Muffins or biscuits (1)
	Oil and vinegar dressing, e.g. italian (1 TBS)
	Onion rings, french fried
	Pancakes or waffles (serving)
	Pasta, e.g. spaghetti, noodles, etc. (1 cup)
	Pie, ready made (slice)
	Pie, homemade (slice)
	Pizza (2 slices)
	Popcorn (1 cups)
	Red chili sauce (1 TBS)
	Beef, pork, or lamb as a sandwich or mixed dish, stew casserole, lasagna
	Sweet roll, coffee cake or other pastry ready made (serving)
	Sweet roll, coffee cake or other pastry homebaked (serving)
	Tomato sauce (1/2 cup) e.g., spaghetti sauce

APPENDIX B

CHEMICALLY ANNOTATED METABOLITES USED IN CHEMRICH METABOLOMICS

SET ENRICHMENT ANALYSIS (TEDDY)

Variable	Compound Name	InChiKeys	Pubchem ID	SMILES
t74	3-aminoisobutyric acid	QCHPKSFMDHPSNR-UHFFFAOYSA-N	25201103	CC(C[NH3+])C(=O)[O-]
t147	Acylcarnitine (C10:1)	GOOOCIIXFLVRAG-UHFFFAOYSA-N	53481651	C[N+](C)(C)CC(CC(=O)[O-])OC(=O)CCCCCCCC=C
t148	Acylcarnitine (C14:2)	HXOGMKPCIDSSKJ-NKBLVAAJSA-N	53481681	CCCCCCCC/C=C/C=C/CC(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
t149	Acylcarnitine (C18:3)	DFVGGGHKDAHYIU-UHMZJXMFSA-N	53477821	CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
t150	Acylcarnitine (C8:0)	CXTATJFJDMJMIY-CYBMUJFWSA-N	11953814	CCCCCCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
t151	Acylcarnitine (C8:1)	YMIVWYONPRZBEJ-LXKVQUBZSA-N	53481667	CCCCC/C=C/C(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
t141	Acylcarnitine C10:0	LZOSYCMHQXPBFU-UHFFFAOYSA-N	10245190	CCCCCCCCCC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
t142	Acylcarnitine C12:0	FUJLYHJROOYKRA-QGZVFWFLSA-N	168381	CCCCCCCCCCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
t144	Acylcarnitine C18:0	FNPHNLNTJNMAEE-UHFFFAOYSA-N	6426855	CCCCCCCCCCCCCCCCCC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
t145	Acylcarnitine C18:1	HITOYGLMAFIRNI-YSESTWPTSA-N	53477830	CCCCCC/C=C/CCCCCCCCCC(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
t146	Acylcarnitine C18:2	MJLXQSQYKZWZCB-DQFWFXSYSA-N	6450015	CCCCC/C=C\C/C=C\CCCCCCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
t73	adipic acid	WNLRTRBMVRJNCN-UHFFFAOYSA-N	196	C(CCC(=O)O)CC(=O)O
t72	alanine	QNAYBMKLOCPYGJ-REOHCLBHSA-N	5950	C[C@@H](C(=O)O)N
t70	asparagine	DCXYFEDJOCDNAF-REOHCLBHSA-N	6267	C([C@@H](C(=O)O)N)C(=O)N
t69	aspartic acid	CKLJMWTZIZZHCS-REOHCLBHSA-N	44367445	C([C@@H](C(=O)O)[NH3+])C(=O)[O-]
t68	benzoic acid	WPYMKLBDIGXBTP-UHFFFAOYSA-N	20144841	[H+].C1=CC=C(C=C1)C(=O)[O-]
t67	capric acid	GHVNFZFCNZKVNT-UHFFFAOYSA-N	2969	CCCCCCCCCC(=O)O
t66	caprylic acid	WWZKQHOCKIZLMA-UHFFFAOYSA-N	379	CCCCCCCC(=O)O
t433	CE (16:1)	HODJWNWCVNUPAQ-FSAOOAOSSA-N	22833543	CCCCC/C=C\CCCCCCCC(=O)OC1CC[C@@]2(C3CC[C@@]4([C@H](CCC4C3CC=C2C1)[C@H](C)CCCC(C)C)C)C
t434	CE (18:1)	RJECHNNFRHZQKU-RMUVNZEASA-N	5283632	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H]1CC[C@@]2([C@H]3CC[C@]4([C@H]([C@@H]3CC=C2C1)CC[C@@H]4[C@H](C)CCCC(C)C)C)C
t435	CE (18:2)	NAACPBBQTFFYQB-LJAITQKLSA-N	5287939	CCCCC/C=C\C/C=C\CCCCCCCC(=O)O[C@H]1CC[C@@]2([C@H]3CC[C@]4([C@H]([C@@H]3CC=C2C1)CC[C@@H]4[C@H](C)CCCC(C)C)C)C
t436	CE (18:3)	FYMCIBHUFSIWCE-WVXFKAQASA-N	6436907	CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)O[C@H]1CC[C@@]2([C@H]3CC[C@]4([C@H]([C@@H]3CC=C2C1)CC[C@@H]4[C@H](C)CCCC(C)C)C)C
t152	CE (20:3)	MLPRJPSMAFZPLA-PJSAOELNSA-N	53477892	CCCCC/C=C\C/C=C\C/C=C\CCCCCCC(=O)O[C@H]1CC[C@@]2(C3CC[C@]4(C(C3CC=C2C1)CCC4[C@H](C)CCCC(C)C)C)C
t437	CE (20:4)	IMXSFYNMSOULQS-BEDFLICRSA-N	6479222	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)O[C@H]1CC[C@@]2([C@H]3CC[C@]4([C@H]([C@@H]3CC=C2C1)CC[C@@H]4[C@H](C)CCCC(C)C)C)C

t438	CE (20:5)	XZFUGMCJZFRBKF-JIKDAPOUSA-N	53477889	CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)O[C@H]1CC[C@@]2(C3CC[C@]4(C(C3CC=C2C1)CCC4[C@H](C)CCCC(C)C)C)C
t439	CE (22:6)	VOEVEGPMRIYYKC-HNJOWPRISA-N	14274978	CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCC(=O)O[C@H]1CC[C@@]2([C@H]3CC[C@]4([C@H]([C@@H]3CC=C2C1)CC[C@@H]4[C@H](C)CCCC(C)C)C)C
t848	Ceramide (d33:1)	QBFXCLDNTKBAPQ-STSAHMJASA-N	52931112	CCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCC)O
t849	Ceramide (d34:0)	GCGTXOVNNFGTPQ-JHOUSYSJSA-N	5283572	CCCCCCCCCCCCCCC[C@H]([C@H](CO)NC(=O)CCCCCCCCCCCCCC)O
t850	Ceramide (d34:1)	YDNKGFDKKRUKPY-TURZORIXSA-N	5283564	CCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCC)O
t851	Ceramide (d34:2)	XXWRZIYYFPIQHE-XPAOSYCESA-N	52931118	CCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCC/C=C\CCC)O
t852	Ceramide (d36:1)	MJQIARGPQMNBGT-WWUCIAQXSA-N	6442676	CCCCCCCCCCCCCCC[C@H]([C@H](CO)NC(=O)CCCCCCCC/C=C\CCCCCCCC)O
t853	Ceramide (d38:1)	XWBWIAOWSABHFI-NUKVNZTCSA-N	5283566	CCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCCCC)O
t854	Ceramide (d39:1)	WYSRACVJQVNCRW-PQPBPFPMSA-N	11273482	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCC)O
t855	Ceramide (d40:0)	SXPRAKSDHOEHIG-ZESVVUHVSA-N	5283575	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](CCCCCCCCCCCCCCC)O
t856	Ceramide (d40:1)	KEPQASGDXIEOIL-GLQCRSEXSA-N	5283567	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCCCC)O
t857	Ceramide (d40:2)	HILTUFAERVOALR-MQXYEJFFSA-N	52931123	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCC/C=C\CCC)O
t859	Ceramide (d42:0)	BPLYVSYSBPLDOA-WVILEFPPSA-N	5283577	CCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](CCCCCCCCCCCCCCC)O
t860	Ceramide (d42:1)	ZJVVOYPTFQEGPH-AUTSUKAISA-N	5283571	CCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCCCC)O
t861	Ceramide (d42:2)	VJSBNBBOSZJDKB-KPEYJIHVSA-N	5283568	CCCCCCCCCCCCC/C=C/[C@H]([C@H](CO)NC(=O)CCCCCCCCCCCC/C=C\CCCCCCCC)O
t863	Ceramide (d43:1)	QHPYSHVSWAOLHS-PVNBSDFKSA-N	9547202	CCCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCCCC)O
t864	Ceramide (d44:1)	CJROVRTUSFQVMR-GVOPMEMSSA-N	5283570	CCCCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO)[C@@H](/C=C/CCCCCCCCCCCCC)O
t64	citramalic acid	XFTRTWQBIOMVPK-UHFFFAOYSA-N	1081	CC(CC(=O)O)(C(=O)O)O
t63	creatinine	DDRJAANPRJIHGJ-UHFFFAOYSA-N	588	CN1CC(=O)N=C1N
t62	cystine	LEVWYRKDKASIDU-UHFFFAOYSA-N	24798687	C(C(C(=O)[O-])[NH3+])SSCC(C(=O)[O-])[NH3+]
t61	deoxypentitol	FJGNTEKSQVNVTJ-UHFFFAOYSA-N	270738	CC(C(C(CO)O)O)O
t444	DG (32:1)	XEQQGHISHUGMIP-ASUORMEESA-N	14275341	CCCCCCCCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\CCCCCCCC
t445	DG (34:1)	YEJYLHKQOBOSCP-OZKTZCCCSA-N	5282283	CCCCCCCCCCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\CCCCCCCC
t156	DG (36:1)	SAEPUUXWQQNLGN-LVVMQYBKSA-N	6443547	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\CCCCCCCC
t447	DG (36:2)	AFSHUZFNMVJNKX-LLWMBOQKSA-N	9543716	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\CCCCCCCC
t448	DG (36:3)	BLZVZPYMHLXLHG-JOBMVARSSA-N	9543722	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t157	DG (36:4)	MQGBAQLIFKSMEM-ZHARMHCNSA-N	9543729	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t159	DG (36:5)	PGXBELQFNRPKBC-WBVIKXMWSA-N	9543737	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCC/C=C\C/C=C\C/C=C\CC
t160	DG (38:0)	IQNYOCFHHRCMKY-KDXMTYKHSA-N	53478362	CCCCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](CO)OC(=O)CCCCCCCCCCCCCCC

t161	DG (38:3)	ADXAWIIUCSQOAS-SMJOZILHSA-N	9543766	CCCCCCCC/C=C\CCCCCCCCC(=O)OC[C@@H](CO)OC(=O)CCCCCCCCC/C=C\C/C=C\CCCCC
t449	DG (38:5)	GRGDLDNREYVILP-CNWVQWJYSA-N	9543784	CCCCC/C=C\C/C=C\CCCCCCCCCC(=O)O[C@@H](CO)COC(=O)CCCCCCC/C=C\C/C=C\C/C=C\CC
t450	DG (38:6)	YDVDXUYJFQLPEG-CTHJWPIASA-N	9543795	CCCCC/C=C\C/C=C\C/C=C\CCCCCCC(=O)O[C@@H](CO)COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CC
t60	erythronic acid lactone	SGMJBNSHAZVGMC-PWNYCUMCSA-N	5325915	C1[C@H]([C@H](C(=O)O1)O)O
t514	FA (16:0)	IPCSVZSSVZVIGE-UHFFFAOYSA-N	985	CCCCCCCCCCCCCCCC(=O)O
t515	FA (16:1)	SECPZKHBENQXJG-FPLPWBNLSA-N	445638	CCCCCC/C=C\CCCCCCCC(=O)O
t516	FA (18:0)	QIQXTHQIDYTFRH-UHFFFAOYSA-N	5281	CCCCCCCCCCCCCCCCCC(=O)O
t517	FA (18:1)	ZQPPMHVWECSIRJ-KTKRTIGZSA-N	445639	CCCCCCCC/C=C\CCCCCCCC(=O)O
t518	FA (18:2)	JBYXPOFIGCOSSB-XBLVEGMJSA-N	5282796	CCCCCC/C=C/C=C/CCCCCCCC(=O)O
t519	FA (20:4)	YZXBAPSDXZZRGB-DOFZRALJSA-N	444899	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)O
t520	FA (22:4)	TWSWSIQAPQLDBP-DOFZRALJSA-N	5497181	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCCC(=O)O
t59	fucose	SHZGCJCMOBCMKK-FPRJBGLDSA-N	439650	C[C@@H]1[C@@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O
t58	galactonic acid	RGHNJXZEOKUKBD-MGCNEYSASA-N	128869	C([C@H]([C@@H]([C@@H]([C@H](C(=O)O)O)O)O)O)O
t451	Gal-Gal-Cer(d18:1/16:0) or Lactosylceramide(d18:1/16:0)	HLIJNIKSBCIDGO-QKLMXXKVSA-N	53477895	CCCCCCCCCCCCCCCC(=O)N[C@@H](CO[C@H]1[C@@H]([C@H]([C@@H]([C@H](O1)CO)O[C@H]2[C@@H]([C@H]([C@H]([C@H](O2)CO)O)O)O)O)O)[C@@H](/C=C/CCCCCCCCCCCCC)O
t865	GlcCer (d38:1)	DFELABABMXOKTD-IYFIADHGSA-N	20057356	CCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO[C@H]1[C@@H]([C@H]([C@@H]([C@H](O1)CO)O)O)O)[C@@H](/C=C/CCCCCCCCCCCCC)O
t452	GlcCer (d40:1)	QMYGQSYNQBLKGY-YAEABVQUSA-N	70699232	CCCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO[C@H]1C(C([C@@H]([C@H](O1)CO)O)O)O)[C@@H](/C=C/CCCCCCCCCCCCC)O
t867	GlcCer (d41:1)	SJGWLQDELUWRDA-MUYAOIFFSA-N	52931253	CCCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](CO[C@H]1C(C([C@@H]([C@H](O1)CO)O)O)O)[C@@H](/C=C/CCCCCCCCCCCCC)O
t453	GlcCer (d42:1)	SNVYJLGKALMFOP-MOLROICHSA-N	6321367	CCCCCCCCCCCCCC[C@H]([C@H](CO[C@H]1C(C([C@@H]([C@H](O1)CO)O)O)O)NC(=O)CCCCCCCCCCCCC/C=C\CCCCCCCC)O
t454	GlcCer (d42:2)	WBOZIXHPUPAOIA-RBELZSLISA-N	6321360	CCCCCCCCCCCCC/C=C/[C@H]([C@H](CO[C@H]1C(C([C@@H]([C@H](O1)CO)O)O)O)NC(=O)CCCCCCCCCCCC/C=C\CCCCCCCC)O
t870	GlcCer(d14:1(4E)/20:0(20H))	YYILQTLJZBSOCA-IMJQGFQJSA-N	70699246	CCCCCCCCCCCCCCCCCCCC(C(=O)N[C@@H](CO[C@H]1C(C([C@@H]([C@H](O1)CO)O)O)O)[C@@H](/C=C/CCCCCCCCC)O)O
t57	gluconic acid	RGHNJXZEOKUKBD-QTBDOELSSA-N	6857417	C([C@@H]([C@H]([C@@H]([C@@H](C(=O)O)O)O)O)O)O
t56	glucose	WQZGKKKJIJFFOK-VFUOTHLCSA-N	64689	C([C@@H]1[C@H]([C@@H]([C@H]([C@@H](O1)O)O)O)O)O
t55	glutamine	ZDXPYRJPNDTMRX-VKHMYHEASA-N	5961	C(CC(=O)N)[C@@H](C(=O)O)N
t54	glutaric acid	JFCQEDHGNNZCLN-UHFFFAOYSA-N	23322899	[H+].[H+].C(CC(=O)[O-])CC(=O)[O-]
t53	glycerol	PEDCQBHIVMGVHV-UHFFFAOYSA-N	753	C(C(CO)O)O
t52	glycine	DHMQDGOQFOQNFH-UHFFFAOYSA-N	5257127	C(C(=O)[O-])[NH3+]
t51	heptadecanoic acid	KEMQGTRYUADPNZ-UHFFFAOYSA-N	10465	CCCCCCCCCCCCCCCCC(=O)O

t50	hexitol	FBPFZTCFMRRES A-UHFFFAOYSA-N	453	C(C(C(C(C(CO)O)O)O)O)O
t49	histidine	HNDVDQJCIGZPNO -YFKPBYRVSA-N	6274	C1=C(NC=N1)C[C@@H](C(=O)O)N
t47	hydroxylamine	AVXURJPOCDRRF D-UHFFFAOYSA-N	787	NO
t46	indole-3-acetate	SEOVTRFCIGRIMH -UHFFFAOYSA-N	802	C1=CC=C2C(=C1)C(=CN2)CC(=O)O
t45	indoxyl sulfate	BXFFHSIDQOFMLE -UHFFFAOYSA-N	10258	C1=CC=C2C(=C1)C(=CN2)OS(=O)(=O)O
t44	isoleucine	AGPKZVBTJJNPAG -WHFBIAKZSA-N	6306	CC[C@H](C)[C@@H](C(=O)O)N
t43	isothreonic acid	JPIJQSOTBSSVTP- GBXIJSLDSA-N	151152	C([C@H]([C@@H](C(=O)O)O)O)O
t42	lactic acid	JVTAAEKCZFNVCJ- UHFFFAOYSA-N	19789253	[H+].CC(C(=O)[O-])O
t455	Lactosylceramide (d18:1/24:1(15Z))	MKOKWBRPIBQYJ J-LWQSSKHKSA-N	20057309	CCCCCCCCCCCCCC/C=C/[C@H]([C@H](CO[C@H]1C([C@H]([C@@H]([C@H](O1)CO)O[C@H]2C[C@@H]([C@H]([C@H]([C@H](O2)CO)O)O)O)O)O)O)NC(=O)CCCCCCCCCCCCCCCCCCCCC/C=C\CCCCCCCC)O
t41	lauric acid	POULHZVOKOAJM A-UHFFFAOYSA-N	3893	CCCCCCCCCCCC(=O)O
t40	leucine	ROHFNLRQFUQHC H-YFKPBYRVSA-N	6106	CC(C)C[C@@H](C(=O)O)N
t39	linoleic acid	OYHQOLUKZRVUR Q-HZJYTTRNSA-N	5280450	CCCCC/C=C\C/C=C\CCCCCCCC(=O)O
t521	LPC (14:0)	VXUOFDJKYGDUJI- OAQYLSRUSA-N	460604	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t163	LPC (15:0)	RJZVWDTYEWCUA R-JOCHJYFZSA-N	24779458	CCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t871	LPC (16:0)	ASWBNKHCZGQVJ V-HSZRJFAPSA-N	460602	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t522	LPC (16:1)	LFUDDCMNKWEO RN-ZXEGGCGDSA- N	24779461	CCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t166	LPC (17:1)	LPMGFNAQZPADD Z-FJIRUFBNSA-N	24779451	CCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t523	LPC (18:0)	IHNKQIMGVNPMT C-RUZDIDTESA-N	497299	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t525	LPC (18:1)	YAMUFBLWGFFIC M-PTGWMXDISA- N	16081932	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t872	LPC (18:2)	SPJFYYJXNPEZDW -FTJOPAKQSA-N	11005824	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t170	LPC (18:3)	WKQNRCYKYCKES D-YVHLTTHBSA-N	24779469	CC/C=C\C/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t171	LPC (20:0)	UATOAILWGVYRQ S-HHHXNRCGSA-N	24779473	CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t526	LPC (20:1)	GJTDRNFWIDPAR Y-GTPZACKGSA-N	24779475	CCCCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t527	LPC (20:2)	YYQVCMMXPIJVH Y-ZOIJLGJPSA-N	52924053	CCCCC/C=C\C/C=C\CCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t528	LPC (20:3)	BBNHCUBQEQJHI G-FZZJNMCHSA-N	52924055	CCCCC/C=C\C/C=C\C/C=C\CCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t175	LPC (20:4)	GOMVPVRDBLLHQ C-VEJNOCSESA-N	53480469	CC/C=C\C/C=C\C/C=C\C/C=C\CCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t176	LPC (20:5)	PDIGSOAOQOXRD U-WJPZTBRDSA-N	11757087	CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t177	LPC (22:4)	ZOJBSSVHFSBHMP -JJJSWPRASA-N	52924039	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t529	LPC (22:5)	YBUXFQUGNPBZP S-YNBHEIDWSA-N	53480473	CCCCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t179	LPC (22:6)	LSOWKZULVQWM LY-APPDJCNMSA- N	10415542	CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t180	LPC (o-16:0)	VLBPIWYTPAXCFJ -DEOSSOPVSA-N	10480367	CCCCCCCCCCCCCCCCOC[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)O

t181	LPC (p-16:0) or LPC (o-16:1)	HTZINLFNXLXRBC-CQLBIITFSA-N	10917802	CCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)O
t530	LPE (16:0)	CKPBBEOJHAPPBT-HXUWFJFHSA-N	53480922	CCCCCCCCCCCCCCCC(=O)O[C@H](CO)COP(=O)(O)OCCN
t531	LPE (18:2)	DBHKHNGBVGWQJE-USWSLJGRSA-N	52925130	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)O
t532	LPE (20:4)	JPNPIRVRGLGTRE-YSKCIPFOSA-N	53480952	CC/C=C\C/C=C\C/C=C\C/C=C\CCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)O
t533	LPE (22:6)	XEVRBOQZSXWGQO-PAUXXPOVSA-N	52925132	CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCC(=O)OC[C@H](COP(=O)(O)OCCN)O
t38	lysine	KDXKERNSBIXSRK-YFKPBYRVSA-N	5962	C(CCN)C[C@@H](C(=O)O)N
t37	lyxitol	HEBKCHPVOIAQTA-IMJSIDKUSA-N	439255	C([C@@H](C([C@H](CO)O)O)O)O
t36	malic acid	BJEPYKJPYRNKOW-UHFFFAOYSA-N	20130941	[H+].[H+].C(C(C(=O)[O-])O)C(=O)[O-]
t35	methanolphosphate	CAAULPUQFIIOTL-UHFFFAOYSA-N	13130	COP(=O)(O)O
t34	methionine	FFEARJCKVFRZRR-BYPYZUCNSA-N	6137	CSCC[C@@H](C(=O)O)N
t32	myo-inositol	CDAISMWEOUEBRE-UHFFFAOYSA-N	892	C1(C(C(C(C(C1O)O)O)O)O)O
t31	N-methylalanine	GDFAOVXKHJXLEI-VKHMYHEASA-N	5288725	C[C@@H](C(=O)O)NC
t30	nornicotine	MYKUKUCHPMASKF-UHFFFAOYSA-N	412	C1CC(NC1)C2=CN=CC=C2
t28	ornithine	AHLPHDHHMVZTML-BYPYZUCNSA-N	6262	C(C[C@@H](C(=O)O)N)CN
t27	oxoproline	ODHCTXKNWHHXJC-VKHMYHEASA-N	7405	C1CC(=O)N[C@@H]1C(=O)O
t184	PC (16:0/9:0(CHO))	PPTNNIINSOQWCE-WJOKGBTCSA-N	46907874	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCC=O
t185	PC (28:0)	CITHEXJVPOWHKC-UUWRZZSWSA-N	5459377	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCC
t186	PC (30:0)	RFVFQQWKPSOBED-PSXMRANNSA-N	129657	CCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t187	PC (30:1)	ANKCYRKQDLQXGL-MRDDHZETSA-N	52922250	CCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCC)COP(=O)([O-])OCC[N+](C)(C)C
t188	PC (31:0)	NPGWXTIWUUFYAB-DIPNUNPCSA-N	24778680	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCCC
t189	PC (31:1)	QFVHCMLUKNHDSH-WTWBAFHPSA-N	24778657	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCC
t873	PC (32:0)	KILNVBDSWZSGLL-KXQOOQHDSA-N	452110	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCCCC
t534	PC (32:1)	QIBZFHLFHCIUOT-NPBIGWJUSA-N	6443788	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCC
t192	PC (32:2)	GPWHCUUIQMGELX-VHQDNGOZSA-N	24778764	CCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCC
t193	PC (32:3)	UXEFXNOSLOCOLX-ZCHSEWAGSA-N	52922763	CCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCC/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t194	PC (33:0)	FHENRYRLCPXONH-LDLOPFEMSA-N	52922645	CCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCCCCCCC
t196	PC (33:2)	SBNDHGBVMZMSNL-UESLNCBNSA-N	52922715	CCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCC)COP(=O)([O-])OCC[N+](C)(C)C
t197	PC (34:0)	PZNPLUBHRSSFHT-RRHRGVEJSA-N	24778686	CCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCCCCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t874	PC (34:1)	WTJKGGKOPKCXLL-VYOBOKEXSA-N	5497103	CCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCCCCC

t875	PC (34:2)	JLPULHDHAOZNQI-ZTIMHPMXSA-N	5287971	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t539	PC (34:3)	CNNSEHUKQJCGTE-UPPWDXJYSA-N	24778699	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCC/C=C\C/C=C\C/C=C\CCCCC
t200	PC (34:4)	YWDDIWXKFJEMKF-JTZVLWBESA-N	52922891	CCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCC/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t201	PC (35:1)	MFHIZGSSDZJFKD-IYEJTHTFSA-N	52922679	CCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t541	PC (35:2)	ZSKWZJYUVZYDQU-WESJWMGVSA-N	52922491	CCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t203	PC (35:2) B	LNGBVAOHJZCRIL-GPDPEMMZSA-N	52923157	CCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCCC/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t204	PC (35:3)	AYXGHIQPMDYMJC-AHMBLZLYSA-N	52924614	CCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\C/C=C\C/C=C\CCCCC)COP(=O)(O)OCCN
t205	PC (35:4)	OROZWUJCDDCYAU-IPUAOQJZSA-N	52922204	CCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t543	PC (36:1)	ATHVAWFAEPLPPQ-VRDBWYNSSA-N	24778825	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCCCCC
t876	PC (36:2)	SNKAWJBJQDLSFF-NVKMUCNASA-N	10350317	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCCCCC
t545	PC (36:3)	YPAZQMWFRMHBBM-CLKMJQEKSA-N	53478785	CCCCCC/C=C\CCCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t208	PC (36:3) A	BXRLDROZWDUSGM-ZRYFCQOPSA-N	24778937	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CCCC/C=C\C/C=C\CCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t211	PC (36:4) A	NKQPOVROGSWLTO-NVPMBMBWSA-N	52922783	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CCCC/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t212	PC (36:4) C	IIZPXYDJLKNOIY-JXPKJXOSSA-N	10747814	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t548	PC (36:5)	DYDDZDMJSQYFGN-OIVUZXIWSA-N	24778771	CCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t214	PC (36:5) B	SUZYROYNFNQALJ-MHEIZRSESA-N	52923341	CCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t215	PC (36:6)	SPWBDEZMKCRQSX-NGPPOSSDSA-N	52922847	CCCCC/C=C\C/C=C\C/C=C\CCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t550	PC (37:2)	MCZUABDVGPPWPM-HJTCUGKVSA-N	52922735	CCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCC/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t217	PC (37:3)	OOYQEEUUQRMQKL-JUUDQZDJSA-N	52922851	CCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t218	PC (37:4)	QRPUCJXFPYFTMB-FBFLODOBSA-N	52922853	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t219	PC (37:5)	URYYGMVXBWUJFP-LQYSTYLLSA-N	53478655	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC
t220	PC (37:6)	GEINPYKZLFHHIL-HEXXMCQTSA-N	52922342	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC
t552	PC (38:2)	KXXLFCAPKGRXBT-FMJYHZMHSA-N	24779263	CCCCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-

				])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t877	PC (38:3)	OJHJKEBRZSDTTL-VHWCKNCUSA-N	52922741	CCCCCCCC/C=C\CCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t223	PC (38:4) A	PSVRFUPOQYJOOZ-QNPWAGBNSA-N	16219824	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\CCCCC
t224	PC (38:4) B	DNYKSJQVBCVGOF-LCKGXUDJSA-N	52923291	CCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t553	PC (38:5)	YLWBKBDNHWQEFU-YJXJLLHLSA-N	53479033	CCCCCC/C=C\CCCCCCCCCC(=O)O[C@H](COC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t225	PC (38:5) A	SUACBSWYGWBPFC-GPUJSUHJSA-N	52923235	CCCCC/C=C\C/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t555	PC (38:6)	PLZBTDKJYHXIEW-DZUXOTHRSA-N	52923295	CCCCC/C=C\C/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t229	PC (38:7)	BNNUJTATKJXKJP-XCDHYEIISA-N	53479075	CCCCC/C=C\C/C=C\C/C=C\CCCCC(=O)O[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t230	PC (39:6)	QMCWOGICYCFNBF-BWHZRABLSA-N	52922637	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCC)COP(=O)([O-])OCC[N+](C)(C)C
t556	PC (40:4)	QQIYXJBHMDYXHH-NMUBDWGHSA-N	52923573	CCCCC/C=C\C/C=C\CCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t558	PC (40:5)	LJFKFKIYUJNFPZ-ZLFSCUDPSA-N	52923133	CCCCCCCC/C=C\CCCCCCCCCCC(=O)O[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\CCCCC
t232	PC (40:5) A	IJTJDJOOHZVSAC-NDRUHXFFSA-N	53479083	CCCCCCCC/C=C\CCCCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t233	PC (40:5) B	SFESOYFQZQJCOY-FXYWPAEZSA-N	52923365	CCCCCCCCCCCCCCCCCCCC(=O)O[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t879	PC (40:6)	FYVNIFOYDIIODX-KNKJIUSSSA-N	24778900	CCCCCC/C=C\CCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC
t234	PC (40:6) B	TYRTWVKQVGNGSZ-RGBTVBCDSA-N	52923195	CCCCC/C=C\C/C=C\CCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t560	PC (40:7)	BPUROMFCPFGBOT-ZEGPSQTJSA-N	24778982	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t561	PC (40:8)	BFCSBEFTXQRIOJ-IMYLGOOQSA-N	53479093	CC/C=C\C/C=C\C/C=C\C/C=C\CCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCC/C=C\C/C=C\C/C=C\C/C=C\CC
t237	PC (42:10)	GILJCAGAMFVHNE-QEOOLSSISA-N	53479133	CCCCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCC(=O)O[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t238	PC (42:5)	APYSSUSAYQRESE-VZWUYPTESA-N	52923591	CCCCC/C=C\C/C=C\CCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCC/C=C\C/C=C\C/C=C\CCCCC
t239	PC (42:6)	DSVRMAGYENFTLY-GDDYDVMSSA-N	52923651	CCCCC/C=C\C/C=C\CCCCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)COP(=O)([O-])OCC[N+](C)(C)C

t562	PC (o-32:0)	SVWBXNAUENUO NE-LDLOPFEMSA-N	173570	CCCCCCCCCCCCCCCCCO[C@H](COC(=O)CCCCCCCCCCC CCCC)COP(=O)([O-])OCC[N+](C)(C)C
t241	PC (o-34:0)	ZKTXOJMFIAILJG-VQJSHJPSSA-N	24779361	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OCCCCCCCCCCCCCCCCCC
t563	PC (p-32:0) or PC (o-32:1)	KEVGQWGZKKFG DC-JCUPVDEDSA-N	53478671	CCCCCCCCCCCCCCCC(=O)OCC(COP(=O)([O-])OCC[N+](C)(C)C)O/C=C\CCCCCCCCCCCCCCC
t243	PC (p-32:1) or PC (o-32:2)	FZMYLOBGNYZPQ O-QLSONYGBSA-N	52923882	CCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCC
t564	PC (p-34:0) or PC (o-34:1)	QCGUXAIDEOWPB V-SNKLRXETSA-N	53481719	CCCCCCCCCCCCCCCC(=O)O[C@H](COCCCCCCCCCCC/C=C\CCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t566	PC (p-34:1) or PC (o-34:2)	MBRHHFWRXQYY AN-RTVLTNFHSA-N	53480735	CCCCCCCCCCCCCCCC(=O)OC(CO/C=C\CCCCCCCC/C=C\CCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t245	PC (p-34:1) or PC (o-34:2) A	MBRHHFWRXQYY AN-JEPFLRBFSA-N	70698781	CCCCCCCCCCCCCCCC(=O)O[C@H](CO/C=C\CCCCCCCC/C=C\CCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t246	PC (p-34:1) or PC (o-34:2) B	KMNVIRCHUMQG HD-RCINKDPXSA-N	52923934	CCCCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\CCCCC
t567	PC (p-34:2) or PC (o-34:3)	QLEHHUPUHJPURI -PWYDUFMYSA-N	24779386	CCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t568	PC (p-36:1) or PC (o-36:2)	ZOTYCHIFTCFAHC -KOUVQCMKSA-N	53480797	CCCCCCCCCCCCCCCCCC(=O)O[C@H](CO/C=C\CCCCC/C=C\CCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t249	PC (p-36:1) or PC (o-36:2) B	ZYLPVUZBZNMVM R-ZBBHDILGSA-N	52923754	CCCCCCCCCCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCC/C=C\C/C=C\CCCC
t250	PC (p-36:2) or PC (o-36:3)	DIHWZUCEXWUH OD-IIVNATNGSA-N	53480801	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](CO/C=C\CCCCC/C=C\CCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t570	PC (p-36:3) or PC (o-36:4)	SOUZQPFUXRVDG K-KCTKZSJBSA-N	53481701	CCCCCCCCCCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CC
t880	PC (p-36:4) or PC (o-36:5)	IOYKZPNDXIIXLN-LOQSCQKMSA-N	24779388	CCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t571	PC (p-38:3) or PC (o-38:4)	GWBOVQHRCURS PU-QMFAPAEZSA-N	53480815	CCCCCCCC/C=C\CCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCCC/C=C\C/C=C\CCCC
t572	PC (p-38:4) or PC (o-38:5)	DBQMOXDLWKVK KG-REWQMPQJSA-N	53480761	CCCCCC/C=C\CCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCC/C=C\C/C=C\C/C=C\CCCCC
t255	PC (p-38:4) or PC (o-38:5) A	YPAPIRJFGNBODV -AMFPDOHCSA-N	53480715	CCCCCCCCCCCCCCCC/C=C\OCC(COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CC
t256	PC (p-38:4) or PC (o-38:5) B	DYFGXBAIDOATA E-DSLLYOFSSA-N	53480759	CCCCCCCC/C=C\C/C=C\C/C=C\CCCC(=O)OC(CO/C=C\CCCCCCCC/C=C\CCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t573	PC (p-38:5) or PC (o-38:6)	ATTCDOPAYPGSL E-LQULQHAGSA-N	53479121	CCCCCCCC/C=C\CCCCCC/C=C\OC(COC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t257	PC (p-38:5) or PC (o-38:6) A	FAKYQMLQEAQOL K-LHZZQLRFSA-N	53480695	CCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t574	PC (p-40:1) or PC (o-40:2)	KYEGAYPFBCAHD L-SPACVREBSA-N	53480827	CCCCCCCCCCCCCCCCCCCCCC(=O)O[C@H](CO/C=C\CCCCC/C=C\CCCCCCCC)COP(=O)([O-])OCC[N+](C)(C)C
t261	PC (p-40:3) or PC (o-40:4)	RDNHPNJCALITSY -MBZSPAKGSA-N	52923852	CCCCCCCCCCCCCCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t262	PC (p-40:4) or PC (o-40:5)	UWNFEVACEPZILS -RNNLSGHUSA-N	52924022	CCCCCCCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC

t263	PC (p-40:5) or PC (o-40:6)	KUHMJRMPHBRA MY-DDURNVNNSA-N	53480775	CCCCCC/C=C\CCCCCCCCC/C=C\OCC(COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t264	PC (p-40:6) or PC (o-40:7) A	XUVCLJCZWPTAIO-OAKHYACESA-N	53479405	CCCCCCCC/C=C\CCCCCC/C=C\OC(COC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t265	PC (p-40:6) or PC (o-40:7) B	FMBYBTSZVHUJM V-DIVFMYBRSA-N	53479425	CCCCCCCCCCCCCCCC/C=C\OC(COC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)([O-])OCC[N+](C)(C)C
t267	PC (p-42:2) or PC (o-42:3)	ZLQCRABYIDJNLT-AZPQEANBSA-N	53479549	CCCCCCCC/C=C\CCCCCCCCCCCCCC(=O)OCC(COP(=O)([O-])OCC[N+](C)(C)C)O/C=C\CCCCC/C=C\CCCCCCCC
t269	PC (p-42:4) or PC (o-42:5)	NLEDXBSUDVLSE N-UFFJXODHSA-N	52924034	CCCCCCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t270	PC (p-42:5) or PC (o-42:6) A	QZMFOSCFDWPIS C-KQHSIIGPSA-N	53481769	CCCCC/C=C\C/C=C\C/C=C\CCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCC/C=C\C/C=C\C/C=C\CCCCC
t271	PC (p-42:5) or PC (o-42:6) B	CMUBJJYJAQLOOD-XSOFIKLRSA-N	52923864	CCCCCCCCCCCCCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CC
t272	PC (p-44:4) or PC (o-44:5)	CEZAZXUWDFPTT E-FHIJIHMSSA-N	53481767	CCCCC/C=C\C/C=C\CCCCCCCCCCCOC[C@H](COP(=O)([O-])OCC[N+](C)(C)C)OC(=O)CCCCCCCC/C=C\C/C=C\C/C=C\CCCCC
t274	PE (34:2)	HBZNVZIRJWODIB-NHCUFCNUSA-N	46891780	CCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[NH3+])OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t275	PE (36:1)	JQKOHRZNEOQNJ E-ZZEZOPTASA-N	25244969	CCCCCCCCCCCCCCCCCC(=O)OCC(COP(=O)([O-])OCC[NH3+])OC(=O)CCCCCCC/C=C\CCCCCCCC
t276	PE (36:4)	KZLUVTCXBFEIFJ-XGLJQOENSA-N	52924904	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)OC(=O)CCCC/C=C\C/C=C\C/C=C\C/C=C\CC
t277	PE (38:4)	ANRKEHNWXKCX DB-BHFWLYLHSA-N	46891781	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COP(=O)([O-])OCC[NH3+])OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t278	PE (38:6)	LFGBKOUQHCWB QI-BZGLIJSBSA-N	52924893	CCCCC/C=C\C/C=C\CCCCCCCCC(=O)OC[C@H](COP(=O)(O)OCCN)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t279	PE (p-34:1) or PE (o-34:2)	SMPXBIVJXNXOAL-PRWZWGSOSA-N	53479657	CCCCCCCCCCCCCC/C=C\O[C@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)COP(=O)(O)OCCN
t281	PE (p-36:2) or PE (o-36:3)	CFANDHZPOSNKN O-UDHSZFGOSA-N	53480897	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](CO/C=C\CCCCCC/C=C\CCCCCCCC)COP(=O)(O)OCCN
t282	PE (p-36:4) or PE (o-36:5)	ADWDFBQPQIEGR Z-XBICFDGKSA-N	53480870	CCCCCC/C=C\CCCCCCCC/C=C\OC[C@H](COP(=O)(O)OCCN)OC(=O)CCCC/C=C\C/C=C\C/C=C\CCCCC
t283	PE (p-38:4) or PE (o-38:5)	ZTZQZGHJLWFLFQ-VZBWJDOASA-N	53480855	CCCCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)(O)OCCN)OC(=O)CCCCCC/C=C\C/C=C\C/C=C\C/C=C\CC
t284	PE (p-38:5) or PE (o-38:6)	IQSPCSIULMCRPM-ZAJUHDLGSA-N	53479831	CCCCCCCC/C=C\CCCCCC/C=C\O[C@H](COC(=O)CCCCC/C=C\C/C=C\C/C=C\C/C=C\CC)COP(=O)(O)OCCN
t286	PE (p-40:5) or PE (o-40:6)	HHQFKPJXVYWLLJ-ABYSKWQHSA-N	53480857	CCCCCCCCCCCCCCCC/C=C\OC[C@H](COP(=O)(O)OCCN)OC(=O)CC/C=C\C/C=C\C/C=C\C/C=C\C/C=C\CCCC
t25	pelargonic acid	FBUKVWPVBMHYJ Y-UHFFFAOYSA-N	8158	CCCCCCCCC(=O)O
t24	phenylalanine	COLNVLDHVKWL RT-QMMMGPOBSA-N	6140	C1=CC=C(C=C1)C[C@@H](C(=O)O)N
t23	phosphate	NBIIXXVUZAFLBC-UHFFFAOYSA-N	1004	OP(=O)(O)O
t22	proline	ONIBWKKTOPOVI A-BYPYZUCNSA-N	145742	C1C[C@H](NC1)C(=O)O
t21	pseudo-uridine	PTJWIQPHWPFNB W-GBNDHIKLSA-N	15047	C1=C(C(=O)NC(=O)N1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O
t19	salicylaldehyde	SMQUZDBALVYZA C-UHFFFAOYSA-N	6998	C1=CC=C(C(=C1)C=O)O
t18	serine	MTCFGRXMJLQNB G-REOHCLBHSA-N	5951	C([C@@H](C(=O)O)N)O

t287	SM (d30:1)	HZCLJRFPXMKWH R-FEBLJDHQSA-N	44260123	CCCCCCCCCCCCC/C=C/[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCC)O
t288	SM (d32:0)	MJAFYELZQYPMQ G-MPQUPPDSSA-N	44260138	CCCCCCCCCCCCCC[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCC)O
t289	SM (d32:1)	KYICBZWZQPCUM O-PSALXKTOSA-N	11433862	CCCCCCCCCCCCC/C=C/[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCCC)O
t291	SM (d33:1)	LQINJRUGTUOHGS -YPDYIYJKSA-N	52931139	CCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t292	SM (d34:0)	QHZIGNLCLJPLCU-QPPIDDCLSA-N	9939965	CCCCCCCCCCCCCCCC[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCCCCCC)O
t293	SM (d34:1)	RWKUXQNLWDTS LO-GWQJGLRPSA-N	9939941	CCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t294	SM (d34:2)	YLWSJLLZUHSIEA-CKSUKHGVSA-N	52931235	CCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCC/C=C\CC)O
t295	SM (d36:0)	JCELSEVNSMXGKA -IOLBBIBUSA-N	44260130	CCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](CCCCCCCCCCCCCCC)O
t296	SM (d36:1)	LKQLRGMMMAHR EN-YJFXYUILSA-N	6453725	CCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t297	SM (d36:2)	NBEADXWAAWCC DG-QDDWGVBQSA-N	6443882	CCCCCCCCCCCCC/C=C/[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCC/C=C\CCCCCCCC)O
t298	SM (d36:3)	YMTVMVYOUDDT QJ-UOMMIRHQSA-N	52931155	CCCCCCCC/C=C\CCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCC/C=C\CC)O
t301	SM (d38:1)	AADLTHQNYQJHQ V-SVLGDMRNSA-N	44260124	CCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t302	SM (d38:2)	MDRFMTLYKHBJT F-NQYLGBTJSA-N	52931179	CCCCCCCCCCCCC/C=C/[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCC/C=C\CCCCCC)O
t305	SM (d40:0)	FONAXCRWZQFJH Y-JCGOJSMZSA-N	44260132	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](CCCCCCCCCCCCCCC)O
t307	SM (d40:2)	FOULCGVQZYQEQ M-DNXGLLHMSA-N	52931201	CCCCCCCCCCCC/C=C/[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCCC/C=C\CCCCCC)O
t309	SM (d41:1)	SXZWBNWTCVLZJ N-NMIJJABPSA-N	46891684	CCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t310	SM (d41:2) A	JBDGKEXQKCCQFK -JWQIMADESA-N	52931209	CCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCC/C=C\CC)O
t312	SM (d42:1)	QEDPUVGSSDPBM D-XTAIVQBESA-N	44260127	CCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCCCCCC)O
t314	SM (d42:2)	DACOGJMBYLZYD H-GXJPFUDISA-N	52931217	CCCCCCCCCCCCCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCC/C=C\CC)O
t315	SM (d42:3)	TXFLWJQVQCDUD Z-BRUGZULGSA-N	52931215	CCCCCCCC/C=C\CCCCCCCCCCCCC(=O)N[C@@H](COP(=O)([O-])OCC[N+](C)(C)C)[C@@H](/C=C/CCCCCCCC/C=C\CC)O
t316	SM (d43:1)	LXMARZYBSFYVSY -KUQVZNNZSA-N	52931225	CCCCCCCCCCCCCCCC[C@H]([C@H](COP(=O)([O-])OCC[N+](C)(C)C)NC(=O)CCCCCCCCCCCC/C=C\CCCCCC)O
t16	succinic acid	KDYFGRWQOYBRF D-UHFFFAOYSA-N	1110	C(CC(=O)O)C(=O)O

t15	taurine	XOAAWQZATWQO TB-UHFFFAOYSA-N	1123	C(CS(=O)(=O)O)N
t457	TG (48:0)	PVNIQBQSYATKKL-UHFFFAOYSA-N	11147	CCCCCCCCCCCCCCCC(=O)OCC(COC(=O)CCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCC
t458	TG (48:1)	FEKLSEFRUGWUOS-DLOIZKPKSA-N	9543986	CCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCCCCCCCCCC
t459	TG (48:2)	RUOVJPPUXXFZPC-YZEIBMOJSA-N	9543987	CCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\CCCCCC
t320	TG (49:0)	TTWJTJMWHOYBPQ-ANFMRNGASA-N	9543988	CCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCC
t460	TG (49:1)	VYYGQDOPVVYUKW-UKFBYESTSA-N	9543991	CCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCCC)OC(=O)CCCCCCCCCCCCCCC
t322	TG (49:2)	QZYSUBAQYSVFNX-PSMULLBHSA-N	9543993	CCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\CCCCCC
t461	TG (49:3)	DIGMYZZFQSIQBD-PNLKURBTSA-N	56938088	CCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\CCCCCC
t462	TG (50:0)	MARPCPMDFOPPJX-UHFFFAOYSA-N	545588	CCCCCCCCCCCCCCCCCC(=O)OCC(COC(=O)CCCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCC
t463	TG (50:1)	YHMDGPZOSGBQRH-YYSBDVFPSA-N	25240460	CCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC
t464	TG (50:2)	QEZWFCZNHWUARW-XQCAQTCHSA-N	9544010	CCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\CCCCCC
t465	TG (50:3)	UFHNZOACKFBCOM-YXKNDSBASA-N	25240357	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\CCCCC
t466	TG (50:4)	PVMBAGXWHHZKFP-JMPJWMFJSA-N	25240359	CCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\CCCCC
t467	TG (50:5)	AFTBPUXZTDLRSP-UDQIKIEDSA-N	9544045	CCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CC)OC(=O)CCCCCCC/C=C\CCCCC
t468	TG (51:1)	OZAXLAGNPZMZAD-BOEMPQCLSA-N	9544006	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC
t469	TG (51:2)	NSNSZGBCOIKUBU-SZOKBDNISA-N	9544013	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\CCCCCC
t470	TG (51:3)	ISSGPXMQOMAFMJ-DMGKHJLRSA-N	9544023	CCCCCCC/C=C\CCCCCCCC(=O)OCC(OC(=O)CCCCCCC/C=C\CCCCCCC)COC(=O)CCCCCCC/C=C\CCCCCCC
t471	TG (51:4)	IIRQXNVLAXQEKB-KBEZCZBDSA-N	9544052	CCCCCCCC/C=C\CCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCC
t472	TG (52:0)	SDNYRTVJOFMYIW-OIVUAWODSA-N	545690	CCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC
t473	TG (52:1)	NPCZZYKITFKRQZ-RFBIWTDZSA-N	5365005	CCCCCCCCCCCCCCCCCC(=O)OC(COC(=O)CCCCCCCCCCCCCCC)COC(=O)CCCCCCC/C=C\CCCCCCC
t474	TG (52:2)	TXMWKTABZBAJCW-QLHBUVOUSA-N	56938176	CCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCC
t475	TG (52:3)	DQXQIWIQYYEGLG-MMWLGPPDSA-N	56938177	CCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCC
t476	TG (52:4)	WHSWXEYWNPTUPW-HNJDVRDNSA-N	25240364	CCCCCCCC/C=C\CCCCCCCC(=O)O[C@H](COC(=O)CC CCC/C=C\CCCCCC)COC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t477	TG (52:5)	CQZAAIKPSLHIBC-KDJOUNIJSA-N	25240366	CCCCCC/C=C\CCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t478	TG (52:6)	SSOSFUDNINFYLJ-KIYGNKBKSA-N	56938180	CCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t479	TG (53:2)	RSINITWKVQRWSZ-RFVLVDBCSA-N	9544102	CCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCC/C=C\C/C=C\CCCC)OC(=O)CCCCCCCCCCCCCCCCC

122

t480	TG (53:3)	ZNQBEJJYVJSZLM-LEDQTTRKSA-N	9544126	CCCCCCCC/C=C\CCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCCC)OC(=O)CCCCCCC/C=C\CCCCCCC
t324	TG (53:4)	BMSDHYZLQWTKSQ-LSJAAEOESA-N	9544152	CCCCCCCC/C=C\CCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCC)OC(=O)CCCCCCC/C=C\CCCCCCC
t481	TG (53:5)	QHYAATSKYBYSLG-BXDFBOBBSA-N	9544183	CCCCCCC/C=C\CCCCCCCCC(=O)OC[C@H](COC(=O)CCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t483	TG (54:1)	YFFIQXNTTVSKJC-NZEOUKRFSA-N	16058371	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\CCCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC
t484	TG (54:2)	WUUWGORPQFKFQN-XDVOZUNOSA-N	56938183	CCCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCC
t326	TG (54:3)	PHYFQTYBJUILEZ-IUPFWZBJSA-N	5497163	CCCCCCCC/C=C\CCCCCCCCC(=O)OCC(OC(=O)CCCCC/C=C\CCCCCCCC)COC(=O)CCCCCCC/C=C\CCCCCCCC
t329	TG (54:4)	BRLGHZXETDWABO-NOFIOOQLSA-N	9544255	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCCC/C=C\C/C=C\C/C=C\CC)OC(=O)CCCCCCC/C=C\CCCCCCC
t486	TG (54:5)	OEJXMJPFOHYSIU-GRLFFVHSSA-N	9544294	CCCCCCCC/C=C\CCCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CC)OC(=O)CCCCCCC/C=C\CCCCCCCC
t488	TG (54:6)	CDNDFDKFZBPPFW-AXJGXPKFSA-N	9544363	CCCCCCCC/C=C\CCCCCCCCC(=O)O[C@H](COC(=O)CCCCC/C=C\CCCCCC)COC(=O)CCC/C=C\C/C=C\C/C=C\CCCCC
t489	TG (54:8)	BMPVTDWOWBNPJU-NYRSPQLFSA-N	9544413	CCCCC/C=C\C/C=C\C/C=C\C/C=C\CCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCC)OC(=O)CCCCCC/C=C\C/C=C\CCCC
t490	TG (56:2)	PDEQUPGHMOMBFC-FYEHETCMSA-N	9544390	CCCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCCCCCCCCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t491	TG (56:3)	QXMHHXQBBKDSSL-BAQZNRHJSA-N	9544447	CCCCCCCC/C=C\CCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)OC(=O)CCCCCCC/C=C\CCCCCCCC
t330	TG (56:4)	YONCDTJKIZDSKQ-IYASBODOSA-N	25240379	CCCCCCCC/C=C\CCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t493	TG (56:5) A	UHEJWASONFIROS-YPSHDQQVSA-N	25240380	CCCCCCCC/C=C\CCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t494	TG (56:6)	ZTNDRFCABXFVMY-WJTCTALZSA-N	9544625	CCCCC/C=C\C/C=C\CCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t332	TG (56:7) B	DODZUDCYRVWEOJ-GKZBLMSTSA-N	9544695	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COC(=O)CCCCCC/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCCCCCC/C=C\C/C=C\CCCCC
t496	TG (56:8)	UBGUHMDKBGQUND-VPFWBQFRSA-N	9544762	CCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t498	TG (58:1)	OWZMHFAFGQCCNI-FBXRAONGSA-N	25240381	CCCCCCCCCCCCCCCCCCCCCC(=O)OC[C@@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)OC(=O)CCCCCCCCCCCCCCCCC
t497	TG (58:10)	GXWBCAVCOMAOHT-VMCJOIRWSA-N	9545277	CCCCC/C=C\C/C=C\CCCCCCCC(=O)OC[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t499	TG (58:6)	GSNFRUMSEHHPSY-LCXCSEBNSA-N	9544977	CCCCCCCC/C=C\CCCCCCCCCC(=O)O[C@H](COC(=O)CCCCCCC/C=C\CCCCCCCC)COC(=O)CCC/C=C\C/C=C\C/C=C\CCCCC
t336	TG (58:8)	KWIGMCRWEINBIR-HUPVKWKYSA-N	9545124	CCCCCCCCCCCCCCCCCC(=O)OC[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC
t500	TG (58:9)	RVXFSLZMZOFGEQ-SWIIBWKZSA-N	9545200	CCCCCCCC/C=C\CCCCCCCCC(=O)OC[C@H](COC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC)OC(=O)CCC/C=C\C/C=C\C/C=C\C/C=C\CCCCC

t14	threitol	UNXHWFMMPAW VPI-QWWZWVQMSA-N	169019	C([C@H]([C@@H](CO)O)O)O
t13	threonic acid	JPIJQSOTBSSVTP-STHAYSLISA-N	5460407	C([C@@H]([C@H](C(=O)O)O)O)O
t12	threonine	AYFVYJQAPQTCCC-GBXIJSLDSA-N	6288	C[C@H]([C@@H](C(=O)O)N)O
t11	tocopherol alpha-	NCYCYZXNIZJOKI-OVSJKPMPSA-N	638015	CC1=C(C(CCC1)(C)C)/C=C/C(=C/C=C/C(=C/C=O)/C)/C
t10	tocopherol gamma-	QUEDXNHFTDJVIY-DQCZWYHMSA-N	92729	CC1=C(C=C2CC[C@@](OC2=C1C)(C)CCC[C@H](C)CCC[C@H](C)CCCC(C)C)O
t9	trans-4-hydroxyproline	PMMYEEVYMWAS QN-DMTCNVIQSA-N	5810	C1[C@H](CN[C@@H]1C(=O)O)O
t8	tryptophan	QIVBCDIJIAJPQS-VIFPVBQESA-N	6305	C1=CC=C2C(=C1)C(=CN2)C[C@@H](C(=O)O)N
t7	tyrosine	OUYCCCASQSFEM E-QMMMGPOBSA-N	6057	C1=CC(=CC=C1C[C@@H](C(=O)O)N)O
t6	urea	XSQUKJJJFZCRTK-UHFFFAOYSA-N	1176	C(=O)(N)N
t5	uric acid	LEHOTFFKMJEON L-UHFFFAOYSA-N	1175	C12=C(NC(=O)N1)NC(=O)NC2=O
t4	uridine	DRTQHJPVMGBUC F-XVFCMESISA-N	6029	C1=CN(C(=O)NC1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O
t3	valine	KZSNJWFQEVHDM F-BYPYZUCNSA-N	6287	CC(C)[C@@H](C(=O)O)N
t2	xanthine	LRFVTYWOQMYA LW-UHFFFAOYSA-N	1188	C1=NC2=C(N1)C(=O)NC(=O)N2
t1	xylulose	LQXVFWRQNMED EE-PYHARJCCSA-N	439205	C1[C@@H]([C@H](C(O1)(CO)O)O)O

DAISY NUTRIENTS EVALUATED FOR SIGNIFICANTLY PREDICTING CANDIDATE

METABOLITES USING STEPWISE SELECTION

Nutrient	hilic_1015	Threonine	Histidine	hilic_243	hilic_996	lipid_278	hilic_291	Choline	hilic_294	hilic_153	hilic_179	hilic_353	hilic_383
Alpha Carotene mcg		1	1				1						
AOAC fiber gm		1		1									
Vitamin B12 mcg	1			1				1					1
Vitamin B1 mg	1												
Vitamin B2 mg	1						1		1				
Vitamin B6 mg						1			1				
Beta Carotene mcg				1				1					
Beta Cryptoxanthin mcg	1		1			1	1		1		1		1
Caffeine mg		1		1					1				
Calcium mg	1							1	1				
Copper mg													
Linoleic gm			1		1				1	1	1		1
Linolenic fatty acid gm					1								
Arachadonic fatty acid gm													
Eicosapentaenoic fatty acid (EPA) gm									1				
Docosapentaenoic fatty acid (DPA) gm	1				1		1	1			1		1
Docosahexaenoic fatty acid (DHA) gm					1								
Free Choline, choline-contributing metabolite mg	1												
Choline from Glycerophos								1					

phocholine, mg													
Iron mg											1		
Potassium mg	1						1	1					
Lutein and Zeaxanthin mcg													
Lycopene mcg	1		1				1	1			1		
Magnesium mg											1		1
Manganese mg						1							
Monounsaturated fat gm				1				1				1	
Vitamin B3 mg									1				1
Choline from Phosphocholine, mg	1							1					
Phospherous mg								1					1
Choline from Phosphatidylcholine, mg	1			1	1				1				1
Saturated fat gm													
Choline from Sphingomyelin, mg					1								
Total Sugars gm		1		1	1	1		1				1	
Protein gm	1												
Vitamin E										1			
Vitamin C mg	1						1			1	1		1
Vitamin D IU				1									
Phylloquinone Vitamin K1, mcg	1	1					1		1				1
Zinc mg						1							

1 = nutrient was selected as a predictor of the metabolite in stepwise regression