TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

COBLL1 genotype driven Metabotyping for early T2D Biomarker Discovery

Loubna Mohammad Ayman Alsadat

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung einer Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Martin Klingenspor
Prüfer der Dissertation:	1. Prof. Dr. Johann J. Hauner
	2. apl. Prof. Dr. Philippe Schmitt-Kopplin

Die Dissertation wurde am 04.04.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 08.09.2022 angenommen.

„By the name of GOD I start"

Acknowledgment

My long Journey started in 2014, and about to end soon, I collected a lot of experiences and knowledge. However, this Journey couldn't be completed and achieved without the support from important persons on my way. Therefore, my first page will be acknowledging to thank all the people who lend a helping hand to me along my way.

First and before all, I must express my gratitude to Almighty God for his blessings, guidance, and his power to me throughout this period. My faith and strong will helped me to step to what I'm in today.

I want to thank Prof. Hauner for allowing me to start this adventure in his Chair and for always providing me with great support for all of the years.

I would also like to thank Prof. Melina, our collaboration partner at the Broad Institute of MIT and Harvard, Boston, USA, for her significant input in the project.

A special thank you to my mentor and a good friend Sara Forcisi for all her support.

I gratefully thank my dearest friend, Katharina Kappo, who was always beside me, cheering me in good and bad times.

I thank all the colleagues for the nice time we had; I didn't only gain knowledge but friends and memories which I will always remember :)

Most importantly, I would like to say a heartfelt thank you to my lovely family, Dad, Mom, my sisters Farah and Leen, for their support, encouragement, and faith in making me believe that sky is my limit. I would've not made it this far without them; with all honor, I would like to dedicate the Ph.D. thesis to them and to the world, hoping that it's a step in making it healthier.

And finally, to my life partner Mohammad, who has been by my side throughout this Ph.D., believing in me and living every single minute of it, he made my Journey brighten with his love and encouragement.

Table of Content

List of Figures

List of Tables:

Abstract

Type 2 diabetes (T2D) is among the most widespread diseases in almost every country, and a particularly high prevalence is observed in Germany. The pathogenesis of T2D involves the interaction of both genetic and environmental factors, where diet and lifestyle are long acknowledged as the most critical environmental risk factors. Evidence from twin studies suggested a strong genetic component with an estimated heritability of 54-83%. Recent genome-wide association studies (GWAS) have provided evidence for a polygenic contribution to T2D. To date, GWAS has revealed more than 700 genetic risk loci for T2D. Genetic variants associated with metabolic changes are expected to display more significant effect sizes because of their direct involvement in metabolite conversion underlying molecular disease-causing mechanisms. Thus, to deeply understand the pathogenesis of T2D and translate GWAS findings into functions, we combined GWAS information with metabolomics data. Ultra-high resolution mass-spectrometry was used to unveil biochemical patterns and potential mechanisms. The overall goal of modern genetics is to translate genetic information into daily clinical practice and elucidate the molecular mechanism of a disease. In the present work, we studied the effects of the *COBLL1* locus, at the SNP rs6712203, on metabolic phenotypes using mass-spectrometry based on untargeted metabolomics. We investigated the metabotype profile of human blood plasma from 127 female subjects (91 *COBLL1* rs6712203 allele carriers/36 non-allele carriers), recruited from the pre-diabetes Lifestyle Intervention Study (PLIS), following a nutritional challenge (OGTT). We performed direct-infusion ion-cyclotron-resonance Fourier-transform mass-spectrometry (DI-ICR-FT MS), in virtue of its ultra-high resolution and mass-accuracy, to link the genome and metabolome levels. Untargeted metabolomics allows for comprehensive screening of numerous known and unknown plasma metabolites. Our results may help to understand which biochemical pathways are altered when a disruption of glucose uptake, adipogenesis, lipolysis, and fat storage occurs due to a disturbance of actin filament formation. We generated two different hypotheses, which involve alternative energy pathways. The study results provide a solid base for future investigations in understanding the biochemical pathways involved in the scenario of T2D.

Zusammenfassung

Der Typ 2 Diabetes (T2D) zählt zu den am weitesten verbreiteten chronischen Krankheiten weltweit. Deutschland ist eines der Länder mit der höchsten T2D-Prävalenz in Europa. Die Krankheitsentstehung von T2D beruht auf einer Wechselwirkung von genetischen und umweltbezogenen Faktoren, wobei die Ernährungsweise und der Lebensstil die kritischsten Risikofaktoren darstellen. Es gibt wachsende Evidenz, z.B. aus Zwillingsstudien, dass genetische Komponenten eine geschätzte Erblichkeit von 54-83 % aufweisen. Besonders aufschlußreich sind genomweite Assoziationsstudien (GWAS), um die Erblichkeit des T2D abzuschätzen. Bis heute haben GWAS mehr als 700 genetische Risiko-Loci für T2D identifiziert. Es ist zu erwarten, dass genetische Varianten, die mit Veränderungen des Stoffwechsels assoziiert sind, deutlich höhere Effektstärken aufweisen, da sie direkt an metabolischen Prozessen beteiligt sind, die den molekularen Mechanismen der Krankheitsentstehung zugrunde liegt. Um die Pathogenese der Krankheit mit Hilfe von GWAS besser zu verstehen, ist es eine neue Strategie, GWAS-Daten mit Metabolomics zu kombinieren. Massenspektrometrie mit ultra-hoher Auflösung wurde zur Erfassung biochemischer Muster und deren Mechanismen angewandt. Die globalen Ziele der heutigen genetischen Forschung sind die Übersetzung genetischer Informationen in die klinische Praxis sowie die Aufklärung der molekularen Krankheitsmechanismen. In der vorliegenden Arbeit haben wir die Effekte des *COBLL1*-Lokus am SNP rs6712203 auf metabolische Phänotypen mithilfe von Massenspektrometrie-basierter ungezielter Metabolomik *(untargeted Metabolomics)* untersucht. Dafür haben wir das Metaboliten-Profil im Blutplasma von 127 weiblichen Probanden (91 *COBLL1* rs6712203 Risiko Allel-Träger/ 36 Nichtrisiko-Allel Träger) analysiert, welche aus der Prädiabetes Lebensstil Interventionsstudie (PLIS) rekrutiert wurden und die sich einem oralen Glukosetoleranztest (OGTT) unterzogen hatten. Wir haben Direktinfusions-Fourier-Transform Ionenzyklotronresonanz Massenspektrometrie (DI-ICR-FT MS)-Messungen durchgeführt, die eine ultrahohe Auflösung und Massengenauigkeit ermöglicht. Ungezielte Metabolomik erlaubt ein breites Screening von vielen bekannten und unbekannten Metaboliten. Unsere Ergebnisse erlauben es, nachzuvollziehen, welche biochemischen Signalwege verändert sind, wenn eine Störung der Glukoseaufnahme, Adipogenese, Lipolyse oder Fettspeicherung vorliegt. Darauf basierend haben wir zwei unterschiedliche Hypothesen aufgestellt, die alternative Energiesignalwege betreffen. Die Studienergebnisse bieten eine solide Grundlage für zukünftige Untersuchungen zum Verständnis biochemischer Signalwege im Zusammenhang mit T2D.

1. Introduction

1.1. Diabetes - a challenge for health care and society

Metabolic disorders such as diabetes have reached epidemic levels, particularly type 2 diabetes (T2D), the most common and widespread type in almost every country (Shaw et al., 2010). The term "diabetes mellitus" is defined by the World Health Organization as "a chronic, metabolic disease characterized by elevated levels of blood glucose, which leads over time to severe damage to the heart, blood vessels, eyes, kidneys, and nerves" (Roglic, 2016). Diabetes mellitus is a disease that is difficult to treat and expensive to manage (Guariguata et al., 2014). Without effective prevention and management programs, this disease is set to increase globally and continuously. Furthermore, diabetes mellitus may lead to several long-term complications, thus playing a significant role in increasing morbidity and mortality in affected patients (Guariguata et al., 2014).

In the past three decades, the prevalence of T2D has risen dramatically in countries of all income levels (Roglic, 2016). Differences in T2D prevalence worldwide are based on many reasons such as cultural and social changes, aging, urbanization, increased prevalence of obesity, physical inactivity, and other unhealthy behaviors (Cecchini et al., 2010). In recent years, diabetes has become one of the leading causes of worldwide death. According to the World Health Organization, around 1.5 million people worldwide died due to diabetes in 2019 (WHO, 2021). In addition, it is estimated that 463 million people are living with diabetes all over the world. By 2045, projections show this number will rise to around 700 million people globally (WHO, 2021).

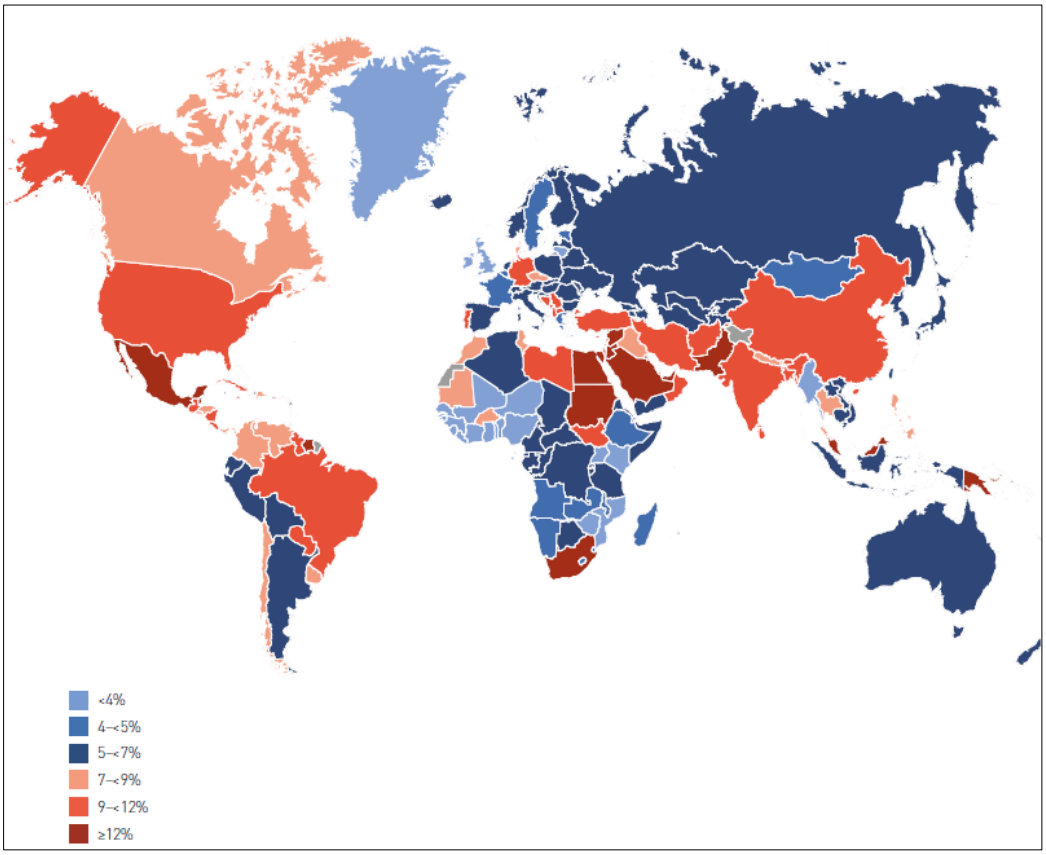


Figure 1: Estimated Age-**adjusted prevalence of Diabetes in adults (20-79 years), in 2019.** Reprinted with permission from the International Diabetes Federation. *IDF Diabetes Atlas, 9th Ed.* Brussels, Belgium: International Diabetes Federation, 2019.

One of the countries in Europe with the highest prevalence of T2D is Germany; according to International Diabetes Federation (IDF) estimates, about 7 million people suffer from diabetes without undiagnosed cases (IDF, 2017).

Diagnosis of diabetes is frequently delayed because symptoms at the early stages of diabetes (pre-diabetes) are usually missing or only mild. Diabetes can be diagnosed through simple blood tests, for example, via oral glucose tolerance test (OGTT) (IDF, 2017). The classification criteria for diabetes and the two high-risk states of abnormal glucose metabolism are IFG and IGT; IFG means impaired fasting glucose (fasting plasma glucose level 100-125 mg/dL), IGT is defined as impaired glucose tolerance (plasma glucose level 140-199mg/dL) (WHO, 2006; ADA, 2017). Diabetes can also be diagnosed by measuring glycated hemoglobin (HbA1c),

raised levels which reflect the concentration of the blood glucose average over the past few weeks to start at 5.8% (39 mmol/mol) (WHO, 2006; ADA, 2017).

It is better to be diagnosed as early as possible as long as the chances of preventing harmful and costly complications will be higher (Dall et al., 2014). Therefore, there is an urgent demand to screen, diagnose, and supply appropriate health care to people with pre-diabetes.

1.1.1. Risk factors of type 2 diabetes

Some risk factors for T2D, such as genetics, ethnicity, and age, are not modifiable. Others, such as being overweight or obese, insufficient physical activity, unhealthy diet, and smoking, are modifiable through behavioral and environmental changes (ADA, 2017).

T2D development results from the interaction between environmental factors and a strong heredity component (Ali et al., 2013). Evidence from twin studies suggests that the concordance rate is higher in monozygotic twins, about 70%. In contrast, in dizygotic twins the rate observed is 20%-30%—estimates of the heritability of T2D range from 20% to 80%. The incidence rate is 40% for individuals with one relative with T2D and 70% if both parents are affected. The disease-prone to run in families and occurs more in ethnic groups (Asian and Hispanic), and the disease increases risk in populations that rapidly adopted a western lifestyle (i.e., Pima Indians) (Ali et al., 2013). Environmental risk factors that are known to increase the development of T2D include obesity, especially abdominal obesity, sedentary lifestyle, pre-diabetes or IGT, smoking, diabetes during a previous pregnancy, other nutritional factors such as unhealthy dietary habits (IDF, 2017). The presence of overweight or obesity increases the risk progressively, proportional to Body Mass Index (BMI) and waist circumference (WC) (Poretsky, 2010). Therefore, people with increased risk factors of T2D should be screened for T2D regularly and supported by suitable strategies to prevent or delay the development of T2D.

1.1.2. Pathogenesis of T2D

The pathogenesis of T2D involves the interaction of both genetic and environmental factors (Gerich et al., 2007). These factors influence glucose homeostasis, which is tightly regulated with dynamic interactions between insulin secretion and tissue sensitivity to insulin. The development of diabetes includes several pathogenic processes (Gerich et al., 2007). This may encompass from autoimmune destruction of the pancreas ß-cells with subsequent insulin deficiency to abnormalities that end in resistance to insulin action (ADA, 2011). In T2D, these mechanisms are impaired, with the consequence that the two primary pathophysiological defects are reduced insulin secretion through a dysfunction of the pancreatic ß-cell, and impaired insulin action through insulin resistance (Leahy, 2005).

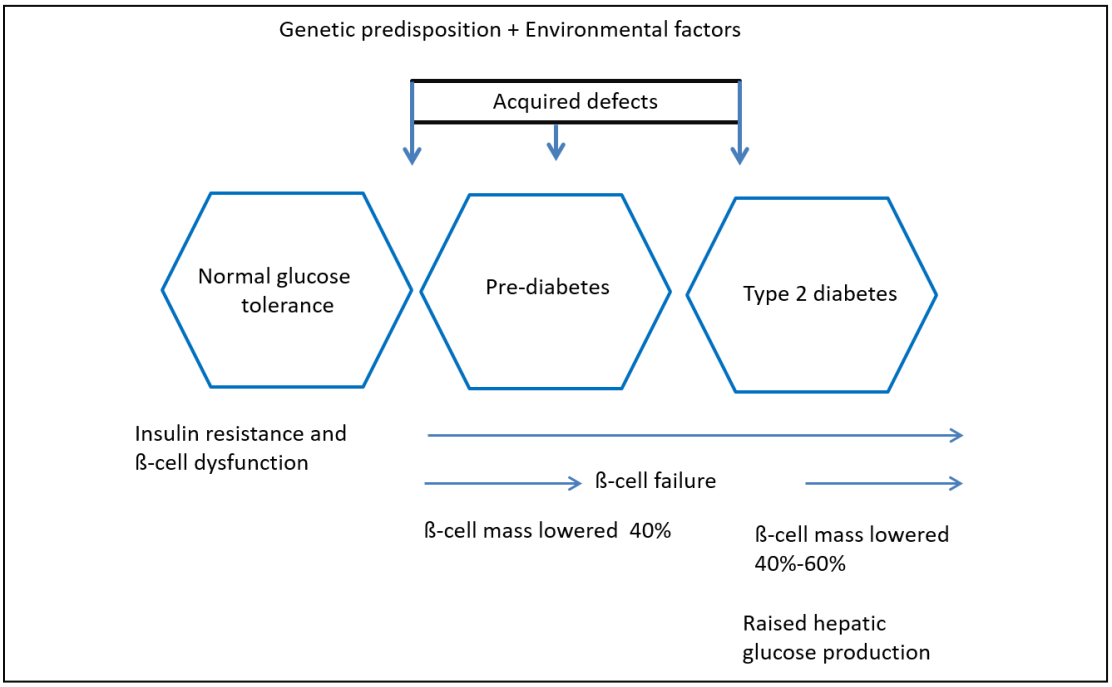


Figure 2: Pathogenesis of T2D. The main pathogenic factors for developing T2D, are a combination of genetic predisposition (insulin resistance and ß-cell dysfunction) and environmental factors leading to pre-diabetes and, eventually, T2D.

Prospective studies indicate that individuals prone to develop T2D pass through five stages (Gerich et al., 2007). The first stage starts at birth, when glucose homeostasis is normal. Individuals may be at risk of T2D for the reason of genetic variants that may predispose them to become obese and/or may limit the ability of their pancreatic ß-cells to compensate for insulin resistance. During stage 2, insulin sensitivity decreases due to this genetic predisposition and an unhealthy lifestyle, which is at first compensated by increased ß-cell insulin secretion so that glucose tolerance remains normal. During stage 3, both ß-cell function and insulin sensitivity are deteriorating. At this point, the ß-cell function is abnormal but sufficient to maintain average fasting plasma glucose concentration. In Stage 4, plasma glucose concentration increases because of a progressing deterioration of ß-cell function and worsening of insulin sensitivity. Finally, in stage 5, both fasting and postprandial glucose levels reach diabetic levels as a result of an advanced deterioration in ß-cell function (Gerich et al., 2007).

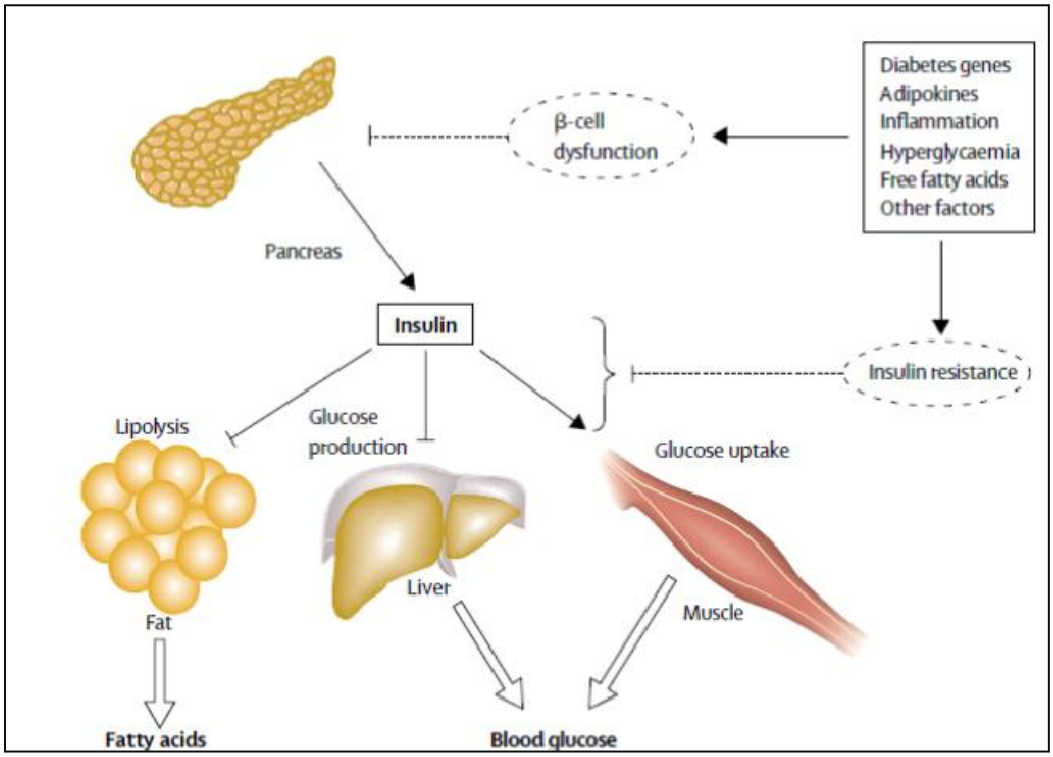


Figure 3: Insulin Secretion and insulin action in T2D pathogenesis. Reprinted from *The Lancet*, 365, Stumvoll M., Goldstein B. J., van Haeften T. W., Type 2 diabetes: principles of pathogenesis and therapy, 1333-1346, Copyright (2012), with permission from Elsevier.

1.1.3. Symptoms and complications of T2D

The symptoms of T2D due to high blood glucose include in particular: excessive thirst and dry mouth, frequent and abundant urination, lack of energy, extreme tiredness (fatigue), tingling or numbness in hands and feet, slow wound healing, and blurred vision (IDF, 2017).

When T2D is not well managed, it can lead to severe complications in many organs. Continuous high blood glucose levels can cause vascular damage affecting the heart, eyes, kidneys, and nerves. T2D is one of the leading causes of; cardiovascular disease (CVD), blindness, kidney failure, and lower-limb amputation as a consequence of the diabetic foot syndrome. Cardiovascular and renal complications are the leading cause of death in people with T2D around the world, and this can be avoided with appropriate treatment (Sargsyan et al., 2019). Besides, diabetes is associated with increased cancer rates, physical and cognitive disability, tuberculosis, and depression. Moreover, in pregnancy, poorly controlled T2D increases the occurrence of maternal and fetal complications (Glümer et al., 2004)

Patient self-management is critical to prevent or delay diabetes complications. Some people who have T2D can achieve near-normal blood glucose levels with diet and exercise alone; on the other hand, many need diabetes medication (metformin, etc.) or insulin therapy. The decision is made depending on the level of blood glucose and the risk of complications.

1.1.4. Prevention and Treatment of T2D

Modification of lifestyle provides an opportunity to reverse the diabetes trend. Saying differently, we cannot change our genetic makeup, but we can modify environmental factors and lifestyles (Darnton-Hill et al., 2004). Risk factors such as diet, adiposity, physical activity, and environmental expositions are modifiable by applying a combination of approaches at both the population and individual level (Darnton-Hill et al., 2004). Many of the prevention programs developed so far have focused on lifestyle modification addressing the previously mentioned modifiable risk factors; meanwhile, other strategies also included the use of pharmacological agents, which improve ß-cell function or insulin resistance (Hussain et al., 2007).

At the individual level, intensive interventions that improve nutrition and physical activity can prevent or delay the outset of T2D in people at high-risk. The importance and effectiveness of prevention of T2D were shown in many clinical trials, such as the Finnish Diabetes Prevention Study (Tuomilehto et al., 2001), the Diabetes Prevention Program Trail in the US (Knowler et al., 2002), and smaller trials such as the Tübingen Lifestyle Intervention Program (TULIP) and the Danish study (Schäfer et al., 2007; Glümer et al., 2004).

Study	Intervention	Relative risk reduction	The number needed to prevent	Time (years)	Reference
DPS	Lifestyle	58	7	3	Tuomilehto J et al., 2001
DPP	Lifestyle	58	7	3	Knowler WC et al., 2002
IDPP	Lifestyle	29	6	3	Ramachandran A et al., 2006
Da Qing	Lifestyle	42	4,5	6	Pan XR et al., 1997
TRIPOD	Troglitazone	49	6	2,5	Buchanan TA et al., 2002
DREAM	Rosiglitazone	60	7	3	Gerstein HC et al., 2006
STOP-NIDDM	Acarbose	25	11	3	Chiasson JL et al., 2002
DPP	Metformin	31	14	3	Knowler WC et al., 2002
XENDOS	Orlistat	37	10	4	Torgerson JS et al., 2004

Table 1: Prevention Studies of T2D. Including randomized controlled trials and pharmacological trials.

The main aim of most intervention programs to prevent T2D is achieving and maintaining a healthy body weight by combining a healthy diet and physical activity in individuals with impaired glucose tolerance as a high-risk group, particularly in those with a family history of diabetes (Hussain et al., 2007). Dietary recommendations in most studies are the following; reduction of fat intake, limiting saturated fatty acid intake to less than 10% of total energy intake, high intake of dietary fiber through consumption of wholegrain cereals, legumes, fruits, and vegetables, and moderate calorie restriction in overweight/obese subjects (Roglic, 2016). Regarding physical activity, it is recommended to spend 30-40 min of moderate physical activity mostly daily or some days a week.

Research groups in different parts of the world, including the USA, Finland, China, India, and other countries, have conducted proof-of-principle studies showing that lifestyle modification with physical activity and a healthy diet can delay or prevent the onset of type 2 diabetes (Roglic, 2016).

The first landmark study to examine the effects of dietary modification, weight loss, and increased physical activity was the Da Qing trial in China, which investigated the effect of a 6-year diet and exercise intervention in Chinese subjects with IGT (Pan et al., 1997). After 6 years of follow-up, the authors concluded that the incidence of conversion to diabetes was 68% in the control subjects and significantly lower in the intervention groups (Pan et al., 1997).

Another landmark study, The Finnish Diabetes Prevention Study (DPS), examined the effect of an intensive lifestyle program in middle-aged and overweight men and women with IGT (Tuomilehto et al., 2001). This combined lifestyle intervention produced long-term beneficial changes in diet, physical activity, consecutive improvement of clinical and biochemical parameters, and reduction of the incidence of T2D. Such a program should be implemented in the primary health care system (Tuomilehto et al., 2001).

The Diabetes Prevention Program (DPP) is the most extensive study to date to examine the efficacy of a lifestyle modification program and metformin in preventing the development of T2D, especially in people at high-risk (Knowler et al., 2002). This study was conducted in 27 centers in

the USA and randomized more than 3000 middle-aged, overweight men and women. After three years, the DPP showed a 58% (lifestyle) and 31% (metformin) reduction in the development of T2D compared with participants in the control group (Knowler et al., 2002).

Likewise, the Indian Diabetes Prevention Program (IDDP) was conducted to examine the combination of a lifestyle program and treatment with metformin on the development of T2D (Ramachandran et al., 2006). The reduction of T2D incidence was 28.5% with lifestyle modification, 26.4% with metformin, and 28.2% with a combination of both over 3 years (Ramachandran et al., 2006).

These studies uniformly demonstrated the importance of the two angles of lifestyle risk factors, dietary factors and physical activity, which are closely related to the risk of T2D. Therefore, both should be combined in order to prevent, manage and control T2D.

For example, several pharmacological intervention studies, TRIPOD, DREAM, EXENDOS, and STOP-NIDDM, have convincingly shown that T2D can be prevented or delayed with different drugs such as Rosiglitazone, Acarbose, Metformin, or Orlistat (Poretsky, 2010). Nevertheless, in the majority of studies, this intervention was not as effective as changes in diet and physical activity (Roglic, 2016).

1.2. Genetic architecture of type 2 diabetes

The development of T2D results from the interaction between environmental factors and the genetic background (Hara et al., 2014). Even with the same environmental exposures, some people are more prone to developing diabetes than others. This means the significant risk appears to be partly inherited. During the last decades, many studies have provided insight into the genetic architecture underlying complex diseases. T2D has been at the frontline of human diseases and phenotypic traits studied by new genetic methodologies (Hara et al., 2014).

The genetic architecture of T2D, defined by the number, frequencies, and effect sizes of causal alleles (Prasad et al., 2015), is still incompletely understood. To understand it is challenging

due to several confounding factors: one is the incoherent and complicated sum of heterogeneous phenotypes summarized by the term T2D. To sustainably improve our understanding of the genetic architecture of T2D, decoding and understanding the genome-wide association studies GWAS signals is essential.

Since (GWAS) was launched, they have changed the landscape of diabetes genetics. Before the availability of GWAS, the primary methods used to determine a link between genotype and phenotype were candidate gene approaches and linkage analyses (Poretsky, 2010).

In the last years, GWAS have increased sample sizes and associated more than 600 genetic variants to T2D (Billings et al., 2010). This number is even markedly higher when glycemic and lipid traits, like insulin secretion, insulin action, obesity, adverse lipid profiles, high blood pressure, and inflammation, are separately considered (Billings et al., 2010).

T2D genes identified by linkage analysis were Transcription factor 7-Like 2 *(TCF7L2)* and Calpain 10 *(CAPN 10)* (Wijmenga et al., 2018). Additionally, peroxisome proliferator-activated receptor gamma *(PPAR-γ)*, and insulin receptor substrate 1 *(IRS1*) are involved in tissue-specific body fat storage, and subsequent effects on cardio-metabolic diseases such as T2D. IRS-2, potassium inwardly-rectifying channel subfamily J member 11 *(KCNJ11)*, Wolfram syndrome 1 *(WFS1)*, HNF1 homeobox A *(HNF1A)*, HNF1 homeobox B *(HNF1B)*, and Hepatocyte Nuclear Factor 4 Alpha (*HNF4A)* were identified through the candidate gene method. The necessity to have other more powerful techniques was obvious to look for variants that were not easily identified by these methods (Wijmenga et al., 2010).

1.2.1. Genome-Wide Association Studies (GWAS)

GWAS is a genetics research tool to look for an association between many specific genetic variations (common SNPs) and particular diseases or traits (Billings et al., 2010). These studies have been powered to consider a massive number of polymorphisms in the human genome. However, the identification of disease-causing variants within association loci remains a significant challenge. Yet, GWAS has some limitations: first, associated genetic variants are not

necessarily causative due to the haplotype structure of the human genome. On the other hand, the requirement of a minimum allele frequency is often set at 5% in the study population. This means that the associated mutation should be ancient enough to be widespread in the study population, and rare variants cannot be easily linked to phenotypes if not sequenced (Billings et al., 2010).

Nevertheless, since the first identification of SNPs analyzed for associations with macular degeneration and myocardial infarction by GWAS, the GWAS catalog has now grown to contain tens of thousands of SNPs associated with hundreds of common diseases and phenotypes (Ali, 2013).

Additionally, over the last decade, GWAS has identified 686 genetic variants at 403 loci associated with T2D (Vujkovic et al., 2020; Cai et al., 2020). The first GWAS for T2D was performed in a French discovery cohort, where the authors identified novel, reproducible association signals at the Zinc transporter 8 (*SLC30A8),* and they were able to validate the well-known association at Transcription factor 7-Like 2 (*TCF7L2)* (Sladek et al., 2007). *SLC30A8* is a gene encoded for a protein involved in the storage and secretion of insulin granules, and is expressed at a high level, specifically in the pancreas (Lefebvre et al., 2012). On the other hand, *TCF7L2* is the most reliable and most replicated variant in the context of T2D studied to date (Damcott et al., 2006; Tong et al., 2009). Moreover, the glucokinase regulatory protein *(GCKR)* gene was identified in GWAS of T2D to be associated with fasting serum triglyceride levels (Saxena et al., 2007). The *GCKR* encodes a regulatory protein in the liver that inhibits glucokinase activity, the enzyme responsible for regulating metabolism, uptake, and storage of circulating glucose. Furthermore, Transducin-like enhancer protein 1 *(TLE1),* Melatonin receptor type 1B *(MTNR1B),* Growth factor receptor-bound protein 14- Cordon-bleu protein-like 1 (*GRB14-COBLL1)* are other relevant T2D loci recently identified by GWAS (Zeggini et al., 2008; Prokopenko et al., 2010; Manning et al., 2012).

A meta-analysis containing eight metabolic traits and six inflammatory markers using existing GWAS published genetic summary results, with about 2.5 million SNPs from the twelve largest GWAS consortia. The analyses yielded 130 unique genomic regions with pleiotropic

associations. One of the genomic regions involved the two genes *GRB14* and *COBLL1* (Kraja et al., 2014).

Manning and colleagues have developed a joint meta-analysis (JMA) approach to identify SNPs significantly associated with either fasting glucose and/or fasting insulin with both adjusting for BMI (Manning et al., 2012). Through this approach, six loci were identified (*IRS1, GRB14-COBLL1,* protein Phosphatase 1 Regulatory Subunit 3B *(PPP1R3B),* platelet Derived Growth Factor C *(PDGFC),* UHRF1 Binding Protein 1 *(UHRF1BP1), and* lysophospholipase-like 1 *(LYPLAL1)*). The *GRB14-COBLL1* locus was associated with increased triglycerides (TGs), decreased high-density lipoprotein (HDL), increased fasting insulin, T2D, and increased low-density lipoprotein (LDL), and is known to interact with receptor tyrosine kinases such as insulin and insulin-like growth factor receptor. It is also associated with triglyceride and insulin levels, consistent with the previous association of this locus with HDL cholesterol (Brown et al., 2016; Abou Ziki et al., 2016).

As long as overweight and obesity are major risk factors and contributors to the development of T2D, the GWAS investigated whether the relationship of the sentry SNPs with T2D might be mediated through adiposity (Wijmenga et al., 2018; Travers et al., 2011). Because of the strong link between obesity and T2D, genes that increase the risk of obesity also show up in GWAS for T2D, including frequently replicated genes Alpha-ketoglutarate dependent dioxygenase *(FTO),* melanocortin receptor 3 *(MC3R)*, and melanocortin receptor 4 (*MC4R)* (Wijmenga et al., 2018; Travers et al., 2011). One of the strong reproducible GWAS BMI-associated signals is the *FTO* gene (fat mass and obesity-associated gene) (McCarthy, 2010). The latest follow-up of mechanistic studies showed pervasive pleiotropy at the locus implicating repression of mitochondrial thermogenesis in adipocyte precursor cells and shift of their differentiation from beige (energy-dissipating) cells to white (energy-storing) adipocytes (Claussnitzer et al., 2015), as well as effects on hypothalamic neurons and sweet preference (Joslin et al., 2021).

Interestingly, Laber and Forcisi et al. have discovered an rs1421085 related decrease of steroids and their derivatives in rs1420185-DEL82 mice under high-fat-diet (HFD) conditions

compared to controls using untargeted, ultra-high-resolution metabolome analysis. Besides, two independent human cohorts following an oral glucose challenge have shown a significant increase of the steroids compound class in male risk carriers compared to non-risk carriers in the immediate response (0h-1h) and a subsequent decrease in the short-term (1h-2h). The mitochondrial characteristics (number of mitochondria, marker genes for thermogenesis and browning) in murine adipose tissue of rs1421085-DEL82 have been in line with findings by Claussnitzer et al., 2015 (Laber et al., 2021).

1.2.2. *COBLL1* as a gene of interest

Since the *COBLL1* gene is among the strongest signals from GWAS that looked at associations with T2D and related traits (McCarthy et al., 2009), our group has been working on elucidating the molecular mechanisms underlying such associations; this gene was also selected as a gene of interest in our ongoing metabolomics project.

The *COBLL1* (COBL-like 1) gene was cloned in 1999 and initially specified as KIAA0977 (Gordon et al., 2011). The deduced protein contained 1,166 amino acids and was found to be expressed at high levels in the lung, liver, kidney, pancreas, ovary, spinal cord, brain, fetal liver, and all specific adult brain regions. In 2003, Carroll et al. renamed the gene *COBLL1* based on the homology to the newly discovered *COBL* (i.e., cordon-bleu homolog, mouse) gene. The specific roles of the gene is not known, but each is assumed to play a role in embryogenesis based on temporal expression patterns during development (Gordon et al., 2011).

COBLL1 is a gene with close to 9500 SNPs and has two isoforms, *COBLL1*a and *COBLL1*b (Park et al., 2015). In Mancina et al., the association between the SNP *COBLL1* rs7607980 C allele and lower insulin levels as lower insulin resistance in overweight and obese children was described for the first time (Mancina et al., 2013). These findings were also confirmed in another study where Manning et al. proved the association of the rs7607980 C allele with lower insulin resistance in adult Europeans (Manning et al., 2012). Finally, in an interesting study by Desmarchelier et al. on a healthy male population, *COBLL1* SNPs were found to be significantly

associated with the postprandial chylomicron and triacylglycerol response, which was positively associated with atherosclerosis and cardiovascular disease risk (Desmarchelier et al., 2014)

In a large GWAS meta-analysis, the *GRB14/COBLL1* locus was significantly identified out of 13 novel loci for WHR, which was adjusted for BMI (Heid et al., 2010; Morris et al., 2012).

Furthermore, the *GRB14/COBLL1* locus was determined to be associated with body fat distribution assessed by WHR adj BMI. In another study, loci near the *GRB14/COBLL1* locus identified body fat percentage (BF %) rather than BMI. This is possible, as fat mass/adiposity is not fully captured by BMI (which represents both lean and fat mass) (Heid et al., 2010; Morris et al., 2012). The association signature of the *GRB14/COBLL1* locus is consistent with the observation that its BF% increasing allele is associated with lower WHR adj BMI, suggestive of a preferential gluteal rather than abdominal fat storage and nominal significance with subcutaneous fat (SAT), but not with the metabolically more harmful visceral fat (VAT) (Lu et al., 2016). GWAS reported a sexual dimorphism at the *GRB14/COBLL1* locus, with a stronger association with women's waist-to-hip ratio (WHR) (Heid et al., 2010; Morris et al., 2012).

Interestingly, a study focusing on body fat distribution found 49 loci whose putative regulatory elements were enriched in adipose tissue, linking adipogenesis and insulin resistance to regulating body fat distribution (Shungin et al., 2015). The association with body fat distribution at the *GRB14/COBLL1* locus may also be mediated by adipose insulin response (Kan et al., 2016). Claussnitzer et al. identified a sexual dimorphism in their experiments, indicating an effect of rs6712203 on stimulated lipolysis and potentially fat distribution in women only. It is interesting to consider that a decreased adipogenic capacity due to *COBLL1* locus perturbation may reduce the adipocyte insulin response in subcutaneous adipocytes, together with the ability to store an excess of energy in women (Glunk et al., under Revision).

The cross-phenotype associations and the phenotypic correlation between BF% and cardio-metabolic traits are inconsistent (Lu et al., 2016). The *GRB14/COBLL1* locus is an example: it was previously identified for associations with fasting insulin, TG, HDL-C, and T2D risk. However, it was shown for the first time that the BF% increasing allele is related to a protective

effect on cardiometabolic health, significantly lowering TG levels, increasing HDL-C levels, leading to a reduced T2D risk. This association signature of the *GRB14/COBLL1* locus is consistent with the observation that its BF% increasing allele was associated with lower WHR and BMI (Lu et al., 2016).

In contrast to the *GRB14* gene, the functional role of the *COBLL1* gene in metabolic diseases is poorly understood in humans and other mammals (Lumish et al., 2020).

The *Cobl*-like gene is linked to diabetes and obesity (Mancina et al., 2013; Sharma et al., 2017). Therefore, Claussnitzer et al. studied different SNPs that are around the *COBLL1* gene (Table 2) and used the bioinformatics phylogenetic module complexity analysis (PMCA) method to prioritize functional genetic variants at the *GRB14/COBLL1* locus. This method was developed to identify potentially relevant SNPs by searching for conserved co-occurring transcription factor binding sites (TFBS) patterns, organized within cis-regulatory modules (CRMs), to predict cis-regulatory variants (i.e., variants affecting gene expression) (Claussnitzer et al., 2014).

Based on this technology, the SNPs presented in (Table 2) were assumed to be functional for the T2D association signal at the *GRB14/COBLL1* locus and had the top scores of T2D and insulin sensitivity (Claussnitzer et al., 2014). Based on this analysis, we intended to investigate subjects genotyped for the SNP rs6712203, an intronic SNP located on the 2nd chromosome. We recruited allele carriers vs. non-risk allele carriers for this investigation who underwent a nutritional challenge (OGTT) in a female sub-population to unveil conditions that cannot be observed in a basal status. Yet, the cellular functions of *COBLL1* and the underlying molecular mechanisms are entirely unknown.

Tag SNP	Reported gene locus	proxy SNP	PMCA result	Chr	Position [GRCh37/hg19][bp]
rs3923113	*GRB14*	rs10184004	complex SNP region	chr2	165508389
rs3923113	*GRB14*	rs10179126	complex SNP region	chr2	165511794
rs3923113	*GRB14*	rs10195252	complex SNP region	chr2	165513091
rs3923113	*GRB14*	rs10187501	complex SNP region	chr2	165532454
rs3923113	*GRB14*	rs6753142	complex SNP region	chr2	165544071
rs3923113	*GRB14*	**rs6712203**	**complex SNP region**	**chr2**	**165557318**
rs3923113	*GRB14*	rs3923113	non-complex SNP region	chr2	165501849
rs3923113	*GRB14*	rs13389219	non-complex SNP region	chr2	165528876

Table 2: Overlap of complex regions and non-complex regions with evolutionary constraint elements and localization to next Transcriptional Start Sites (TSSs). The tag SNPs, the reported gene locus, and the proxy SNPs are listed. Proxy SNPs are derived from SNAP viewer data (Johnson, et al., 2008), and SNP regions of the 1,465 candidate SNPs located at the 47 GWAS T2D risk loci were analyzed by PMCA and sorted into SNPs surrounded by complex regions and non-complex regions. SNP positions of the 1,465 SNPs were used as anchors in a Genome Inspector (Genomatix) correlation with evolutionary constraint elements (Lindblad-Toh et al., 2011).

Our research group was interested in further investigating the *GRB14/COBLL1* risk locus and the *COBLL1* gene and the risk for T2D. Claussnitzer and colleagues performed a broad series of experiments to elucidate the function of *COBLL1* as a gene of interest. As a result, they demonstrated that *COBLL1* mRNA expression could be linked to actin filamentation, cell differentiation, insulin-stimulated glucose uptake, lipid accumulation, and regulation of lipolysis in subcutaneous human adipocytes (unpublished data). Additionally, since the group found that the *COBLL1* locus is involved in the regulation of F-actin, a dynamic regulation is essential to adapt to changes in the cell's microenvironment. They proposed that a disrupted regulation of the *COBLL1* gene in rs6712203 risk allele carriers may lead to decreased actin stress fiber production, which subsequently does not allow the production of cortical actin and may lead to consecutively disturbed glucose uptake, adipogenesis, lipolysis, and TG storage. This molecular mechanism could partially explain the observed decreased hip circumference and unfavorable body fat distribution in rs6712203-C risk allele carriers, particularly in obese women. A disturbance in the energy metabolism of adipocytes and resulting challenges in energy storage are known risk factors for T2D, and this could be the missing link in the GWAS risk signal for T2D at the *GRB14/COBLL1* locus (Figure 4).

The Claussnitzer team identified at least two tagging variants within the *GRB14/COBLL1* locus and associated with T2D, highlighting the complex haplotype structure and the great challenge to identify the disease-causing variants. This computational and experimental model elucidated a *POU2F2*-dependent up-regulation of *COBLL1* mRNA in rs6712203-T non-risk adipocytes, possibly contributing to the *GRB14/COBLL1* GWAS risk associated with insulin resistance and T2D (Morris et al., 2012; Konner et al., 2011; Mahajan et al., 2014). The next aim was to determine the likely regulatory variant(s) responsible for the T2D association. Using the cross-species conservation analysis PMCA, a computational approach calculating regulatory probability scores (range 1-9) using conserved transcription factor binding site (TFBS) patterns within cis-regulatory modules (CRMs) (Claussnitzer et al., 2014), rs6712203 was identified as the highest-scoring variant with a striking clustering of 312 TFBS ($p<0.0001$), 32402 TFBS modules ($p<0.0001$), 763 TFBS in these modules ($p<0.0001$), resulting in the combined overall score of 9

(Claussnitzer et al., 2014). Notably, the rs6712203 SNP is localized within an adipose-specific enhancer region (Glunk et al., under Revision).

This work provides evidence that *COBLL1* is involved in adipocyte actin remodeling, maturation, and metabolism, which has not been reported before. Although the *GRB14* locus encodes an adaptor protein for the insulin receptor (Scharf et al., 2004), which has been suggested as being potentially causal for the T2D risk association at the *GRB14/COBLL1* locus (Schleinitz et al., 2004), the regulatory circuitry underlying the genome-wide disease association has not previously been identified. To unravel the regulatory variant, its regulator, and the affected gene at the *GRB14/COBLL1* locus, they integrated publicly available data with specifically designed targeted perturbations in human adipocytes.

Glunk et al. identified an adipocyte-specific enhancer region surrounding the intronic *COBLL1* variant rs6712203, its upstream regulator *POU2F2*, and the up-regulation of the target gene *COBLL1* in rs6712203-T non-risk allele carriers. Until now, the function of the *COBLL1* protein in adipocytes has not been described. Following *COBLL1* locus perturbation, they found a lower ability to remodel F-actin stress fibers into cortical actin, decreased adipogenesis, decreased insulin-stimulated glucose uptake, triglyceride storage, and decreased stimulated lipolysis. The observed cellular phenotypes are relevant to T2D and consistent with the association of the *GRB14/COBLL1* locus with T2D across populations resulting from peripheral insulin resistance with a dysfunctional energy regulation in adipose tissue.

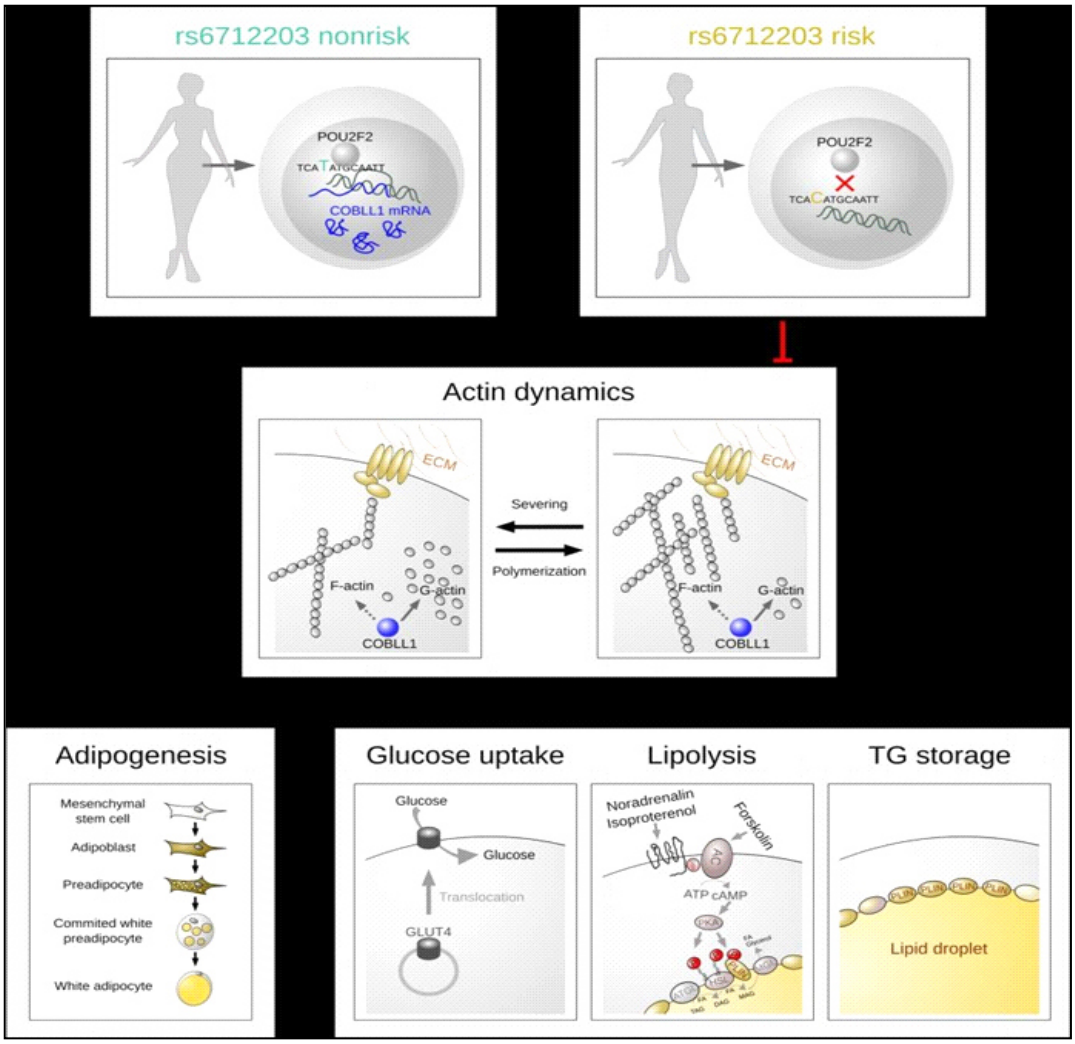


Figure 4: Mechanistic model detailing the POU2F2-dependent up-regulation of *COBLL1* expression in rs6712203-T non-risk allele carriers. In rs6712203-T non-risk human adipocytes, the expression of *COBLL1* is POU2F2 dependent upregulated. In rs6712203-C risk adipocytes, the POU2F2 motif is partially disrupted, which prevents the up-regulation of *COBLL1. COBLL1* perturbation in human adipocytes leads to a disturbed remodeling of F-actin fibers from stress fibers to cortical actin, which is essential for adipogenesis and the adaptation to changes in the cell's microenvironment. We propose that a disrupted regulation of *COBLL1* in rs6712203-risk adipocytes may lead to lower cortical actin structures, which can cause disturbances during adipogenesis and result in a decreased insulin-stimulated glucose uptake lipolysis and TG storage. Disturbed energy regulation in adipocytes and resulting challenges in energy storage in the human body are known risk factors for T2D. Therefore, they could contribute to the GWAS risk signal for T2D at the *GRB14/COBLL1* locus. Provided by M. Claussnitzer et al. (Glunk et al., under review)

The mentioned mechanisms and pathways may significantly control essential metabolic pathways in the human body, starting with adipogenesis, which is the cell differentiation process

by which pre-adipocytes become adipocytes (fat cells) (Ghaben et al., 2019). This process is highly regulated by counter-regulatory hormones, which these cells are susceptible to. For example, insulin contributes to triglyceride storage in fat cells, whereas the catabolic hormones epinephrine, glucagon, and ACTH promote lipid mobilization (Ghaben et al., 2019).

Mutations of genetic alleles that regulate the production of actin or its associated protein can cause an enormous number of diseases (Laing et al., 2009). The generation of actin is also vital for the infection process by some pathogenic microorganisms (Laing et al., 2009). Mammals have six actin genes; four are expressed in muscle cells, and two in non-muscle cells (Perrin et al., 2010). Any mutations in the different genes that regulate actin production in humans can cause a muscular disease (Laing et al., 2009). In addition, the composition of the cytoskeleton is related to the pathogenicity of intracellular bacteria and viruses, particularly in the processes associated with avoiding the actions of the immune system (Laing et al., 2009).

The cortical actin filament is found in mature adipocytes, while pre-adipocytes have stress fibers, which are remodeled into cortical actin during adipogenesis (Yang et al., 2014). A whole and dynamic actin remodeling is also crucial for insulin-stimulated glucose uptake into the cells, since cortical actin disruption and stabilization inhibits *GLUT4* translocation (Kanzaki et al., 2001). In adipocytes, insulin is essential for the breakdown of F-actin stress fibers (Martina et al., 1996). Insulin stimulation of adipocytes initially results in cortical actin remodeling, followed by increased polymerized actin to enable *GLUT4* translocation to the plasma membrane (Kanzaki et al., 2001).

In summary, the actin cytoskeleton is involved in many cellular processes and provides structural support and vesicle trafficking. Coordinated regulation of the cytoskeleton is essential during adipogenesis (Kawaguchi et al., 2003; Yang et al., 2014; Kanzaki et al., 2001), *GLUT4* translocation to the plasma membrane (Kanzaki et al., 2001), the maintenance of lipid droplets, and lipid droplet dynamics during lipolysis (Husson et al., 2011). Taking into account that the *COBLL1* homolog *COBL* is a known F-actin dynamizer (Orlicky et al., 2013), and GO-term pathways

list *COBLL1* as actin interacting protein (Ashburner et al., 2000), we investigated whether *COBLL1* may play a role in the regulation of the actin cytoskeleton in adipocytes, thereby affecting *GLUT4* vesicle trafficking, lipid metabolism, and adipocyte differentiation (Ashburner et al., 2000).

1.3. Metabolomics and system biology

Metabolomics is an emerging field in systems biology, providing a direct readout of physiological status, biochemical mechanisms, and enzymatic activities within an individual at a specific point in time (Ramautar et al., 2013). Additionally, it is an essential technique for systems biology and translational medicine, particularly in combination with other -omics technologies such as genomics, transcriptomics, or proteomics (Villas-Boas et al., 2007; Nicholson et al., 2006). It has the potential to deliver novel diagnostic biomarkers for the detection and prognosis of diseases (Villas-Boas et al., 2007; Nicholson, 2006). Metabolomics refers to the systemic identification and quantification of the small molecules and products (the metabolome) of a biological system (cell, tissue, organ, biological fluid, or organism) at a specific time point (Klassen et al., 2017).

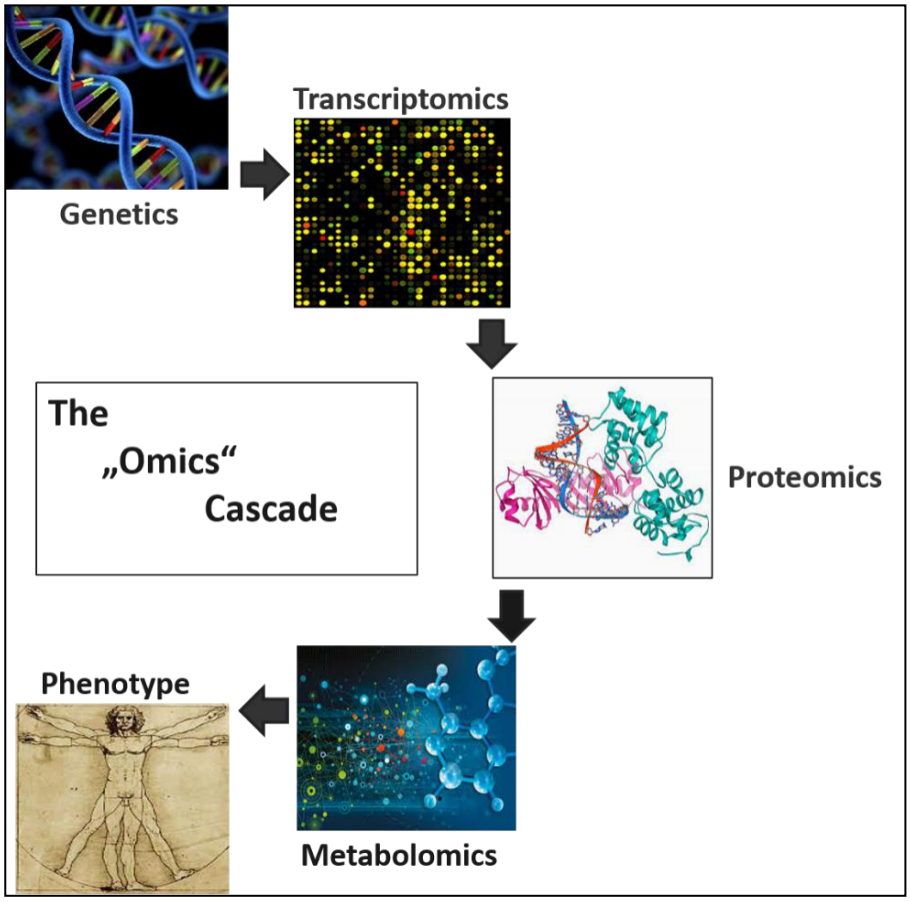


Figure 5: The omics cascade. A typically integrated omics cascade starts with datasets inputs, datasets outputs, and results. Using individual omics datasets that are closer to genotype (genomics and transcriptomics), and those closer to phenotype (proteomics and metabolomics), plus a host of other omics platforms, and datasets that are integrated using statistical or advanced machine learning approaches. Results may be simple pathways or complex networks and include both known and novel molecules. In addition, results may predict health or disease states, provide insights for effective therapeutic interventions, or reveal space-time regulation of systems such as cell, tissue, or organ type specificity.

Metabolites are small molecules and products of cellular regulatory processes; many factors influence their levels, including disease status, environment, medications, diet, and genetic factors (Nicholson, 2008). Thus, metabolites are helpful for diagnosis and prognosis as well as predicting and monitoring the efficacy of treatment (Likić et al., 2010).

According to the human metabolome database (http://www.hmdb.ca), human metabolites can range from 1000 until 108766 (Wishart et al., 2009).

Consequently, the strong links between an individual genetic profile and metabolomics allow the investigation of the pathways behind changes at the level of the metabolites (Suhre et al., 2012). Therefore, in the field of metabolomics, researchers need to take into account the genetic factors underlying the production of metabolites and their possible role in disease processes (Suhre et al., 2012; Johnson et al., 2016).

Metabolomics is defined as the (semi-)quantitative measurement of a complex system's multi-parametric time-related metabolic response to an intervention or genetic modification (Nicholson et al., 1999). The metabotype is defined as a metabolic profile, which relates genetic variations of an organism (Gavaghan et al., 2000) to environmental factors such as age, gender, lifestyle, diet, stress level, and gut microbiota (Nicholson et al., 2002; Daviglus et al., 2004; Li et al., 2008). Now, GWAS can be carried out with large panels of metabolite concentrations (Suhre et al., 2012). While using this largely hypothesis-free approach, common genetic variants in genes encoding enzymes transporter proteins have been identified with substantial influences on human metabolic traits. These so-called genetically influenced metabotypes (GIMs) are starting to be combined with increasing knowledge of disease-associated genetic loci to uncover new complex risk factors of common diseases and to provide functional insights into the pathophysiology of related disorders (Suhre et al., 2012).

Metabolomics strategies have been divided into two different approaches: untargeted and targeted metabolomics (Johnson et al., 2016). **The targeted metabolomics** approach refers to a method, which characterizes specific classes of known compounds, and biochemically annotated metabolites, focusing on one or more known related pathways of interest. This strategy provides measurements that are more precise and is easy to replicate but are limited to a subset of pre-selected compounds (Patti et al., 2012). This approach is widely applied in the pharmaceutical field and provides quantifications using isotope-labeled external standards.

The other kind of metabolic profiling is the **untargeted metabolomics** approach, by means of which a comprehensive and extensive range of compounds are measured, without bias and including unknown metabolites (Naz et al., 2014). This approach increases the need to follow multiple screenings during analysis (Roberts et al., 2014). However, it detects the maximum

possible amount of metabolites without prior knowledge of the extracted compounds, making this a valuable approach for identifying novel or unveiling new mechanisms (Roberts et al., 2014).

1.3.1. System biology in the "omics" era

Nowadays, expansive amounts of biological data can be produced by advanced high throughput technologies allowing not only the identification and the quantification of individual components of a system (e.g., genes, proteins, or metabolites), but the generation of extensive networks describing the potential interactions between their components (Wang et al., 2015). Due to such massive availability of "omics" datasets, the system's way of thinking has become an accomplishable goal (Likić et al., 2010). However, any single type of such high throughput data, representing only one dimension of complex biological systems, is unable to uncover new functions. Different "omics" levels provide alternately complementary information on corresponding mechanisms. The complete picture can only be constructed by studying the relationships between genes, transcripts, proteins, and metabolites, enhancing the importance of the integration of heterogeneous and large "omics" data. The "omics" era meets the challenge of mining biological knowledge and generates novel insights and reliable hypotheses from the excess of available data.

1.3.2. Analytical tools in metabolomics

Advances in analytical technologies certainly drive scientific knowledge (Zhang et al., 2012). The development in the detection and identification of small molecules, including amino acids, peptides, lipids, carbohydrates, etc., dramatically depends on the corresponding methods and tools. Finding the most comprehensive platform is not always feasible; since every analytical technique has its advantages and disadvantages. Thus, combining analytical approaches is an essential step towards studying the global metabolome (Zhang et al., 2012; Forcisi et al., 2015).

Metabolomics experiments can be performed on many biofluids and tissue types via the application of different analytical platforms: mass-spectrometry (MS) and Nuclear magnetic

resonance (NMR) spectroscopy, allowing metabolite identification and quantification (Zhang et al., 2012; Forcisi et al., 2015).

To examine such a broad chemical diversity, global untargeted metabolomics requires various and complementary analytical tools to achieve a coverage of all metabolites, endogenous and exogenous, present in a biological sample (Zhang et al., 2012; Forcisi et al., 2015).

Over the last decades, the application of metabolomics has gained increasing interest. Several analytical platforms are commonly used in untargeted metabolomics research applying NMR spectroscopy, Liquid chromatography (LC-MS), Gas chromatography (GC-MS), and Direct Infusion-Ion-Cyclotron-resonance Fourier-Transform Mass-Spectrometry (DI-ICR-FT MS)

(Villas-Bôas et al., 2005; Wilson et al., 2005; Lenz et al., 2007; Lu et al., 2008; Theodoridis et al., 2008; Wu et al., 2009; Wishart et al., 2011; Dunn et al., 2008; Dunn et al., 2005; Nicholson et al., 2007).

NMR spectroscopy is particularly appropriate for analysis of bulk metabolites, GC-MS to the analysis of volatile organic compounds and primary derivative metabolites, LC-MS is highly appropriate to analyze a wide range of semi-polar compounds, and (DI-ICR-FT MS) provides an extraordinarily high mass-resolution and mass-accuracy, together with a high-order-of-magnitude intensities range. Since LC-MS can avoid chemical derivatization, it is a vastly used instrument in combination with high-resolution mass-spectrometry (Villas-Bôas et al., 2005; Wilson et al., 2005; Lenz et al., 2007; Lu et al., 2008; Theodoridis et al., 2008; Wu et al., 2009; Wishart et al., 2011; Dunn et al., 2008; Dunn et al., 2005; Nicholson et al., 2007). High selectivity and sensitivity for the identification and quantification of metabolites are offered by MS-based metabolomics, combined with advanced high throughput separation techniques that can reduce the complexity of metabolite separation (Zhang et al., 2012; Forcisi et al., 2015; Villas-Bôas et al., 2005; Wilson et al., 2005; Lenz et al., 2007; Lu et al., 2008; Theodoridis et al., 2008; Wu et al., 2009; Wishart et al., 2011; Dunn et al., 2008; Dunn et al., 2005; Nicholson et al., 2007; Shulaev et al., 2006).

The following (Table 3) shows a comparison of the principal technologies applied in metabolomics research

Technique	NMR	GC-MS	UPLC-MS	DI-ICR-FT MS
Sample preparation	Dilution	1. Minimization of matrix suppression effect 2. minimizing the content of salt, proteins, and artifacts	1. Minimization of matrix suppression effect 2. minimizing the content of salt, proteins, and artifacts	1. Minimization of matrix suppression effect 2. minimizing the content of salt, proteins, and artifacts
Range of metabolites	All the range of metabolites in high concentration	Volatile compounds, non-polar volatile compounds with masses < 600 amu	Polar and polar non-volatile compounds	Polar and polar non-volatile compounds with m/z > 120 Da.
Advantages	1. Information on the molecular structure 3. Robust quantification without standards	1. High reproducibility 2. Separation and concentration of different classes of compounds according to Physio-chemical properties	1. Detection of isomers and isobars 2. Separation and concentration of different classes of compounds Physio-chemical properties	1. High resolution 2. High mass precision formula calculation and compounds identification 3. High throughput 4. Long term ion storage for MS/MS/MS experiments

Table 3: Comparison of the principal technologies applied in metabolomics research. Advantages are highlighted (Forcisi, S. et al., 2015; González-Dominguez et al., 2017).

2. Study aim

This work aims to connect variants at the *GRB14/COBLL1* metabolic risk locus and the effector gene *COBLL1* with metabolic changes in the plasma, applying metabolic profiling to understand better the underlying mechanisms of the risk locus on disease-relevant traits and to pave the way for better disease prediction through biomarkers. An individual predisposition towards diseases can be determined by studying the effects of genetic variants on metabolic phenotypes (metabotypes) (Gienger et al., 2008; Suhre et al., 2011; Long et al., 2017). In my thesis, I sought to investigate the metabolic profiles of the subjects genotyped for the *GRB14/COBLL1* risk haplotype by the rs6712203 variant, which the Claussnitzer team has previously shown to mediate part of the metabolic risk by affecting actin cytoskeleton remodeling in adipocytes. In my work, I included subjects from the PLIS cohort genotyped for the SNP rs6712203 (91 risk allele carriers vs. 36 non-risk allele carriers) who underwent nutritional challenges to unveil their metabolic profile under specific stimuli.

I studied the effect of SNP rs6712203 on metabolic phenotypes by utilization of an untargeted metabolome analysis that allows us to have a comprehensive picture by screening all metabolites as well as the ones with unknown biotransformation mechanisms within an unknown chemical space, especially after we observed no significant differences values in all glucose and NEFA parameters between the measured clinical parameters such as TG, HBA1C, CHO, HDL, LDL, CRP. Furthermore, we used the direct platform infusion, Fourier transforms ion cyclotron resonance mass-spectrometry (DI-ICR-FT MS) because of its ultra-high-resolution and mass-accuracy to link the rs6712203 genotype and metabolome.

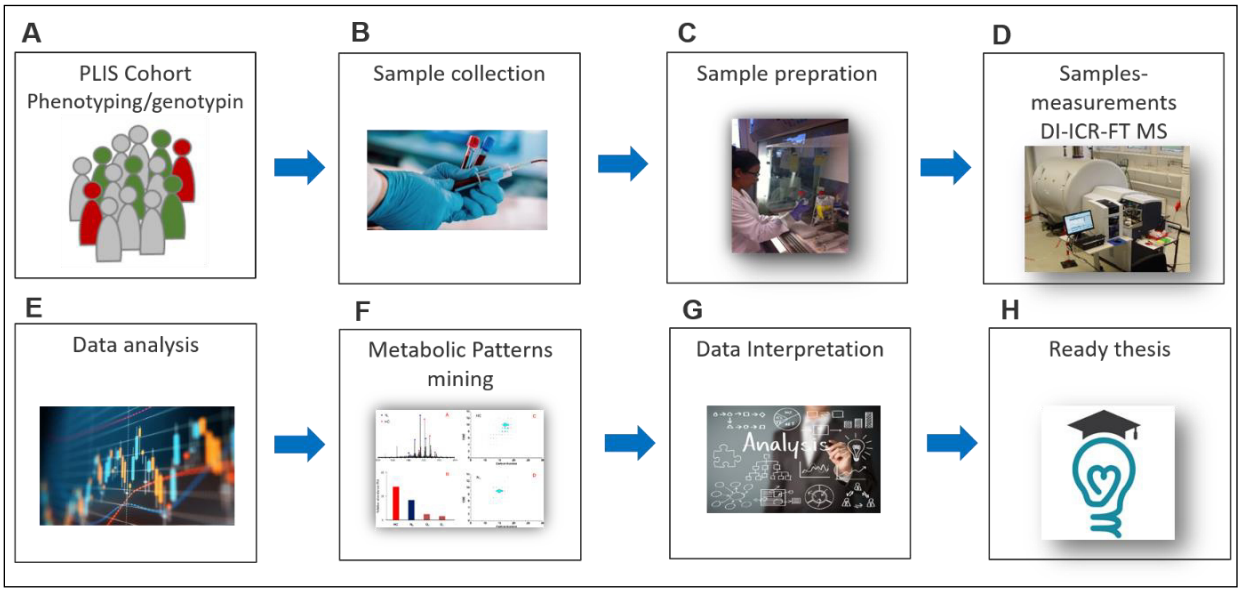


Figure 6: Workflow description of the project. Starting with the PLIS Cohort, we wanted to study the metabolites of subjects recruited using untargeted metabolomics and measure them with the high resolution and mass-accuracy platform DI-ICR-FT MS. Starting with phenotyping and genotyping of the PLIS cohort subjects **(A)**, subjects blood samples were collected **(B)** and prepared **(C).** Afterward, measurements of samples via DI-ICR-FT MS were performed **(D),** followed by data analysis **(E)** and mining of metabolic patterns (F). After data interpretation **(G),** results are ready to be presented **(H)**

My main study objective is to identify metabolite patterns in circulating plasma that reflects the metabolic dysregulation caused by genetic variation, specifically the *COBLL1* gene, leading to an increased risk of T2D. This took place in a cohort of subjects participating in the pre-diabetes intervention study (PLIS). In order to accomplish this goal, the untargeted human blood metabolites were investigated using Direct Infusion-Ion-Cyclotron-resonance Fourier-Transform Mass-Spectrometry DI-ICR-FT MS. This high-resolution and high accuracy approach focused on screening a maximum number of metabolites, enabling top confidence feature assignment to molecular formulas.

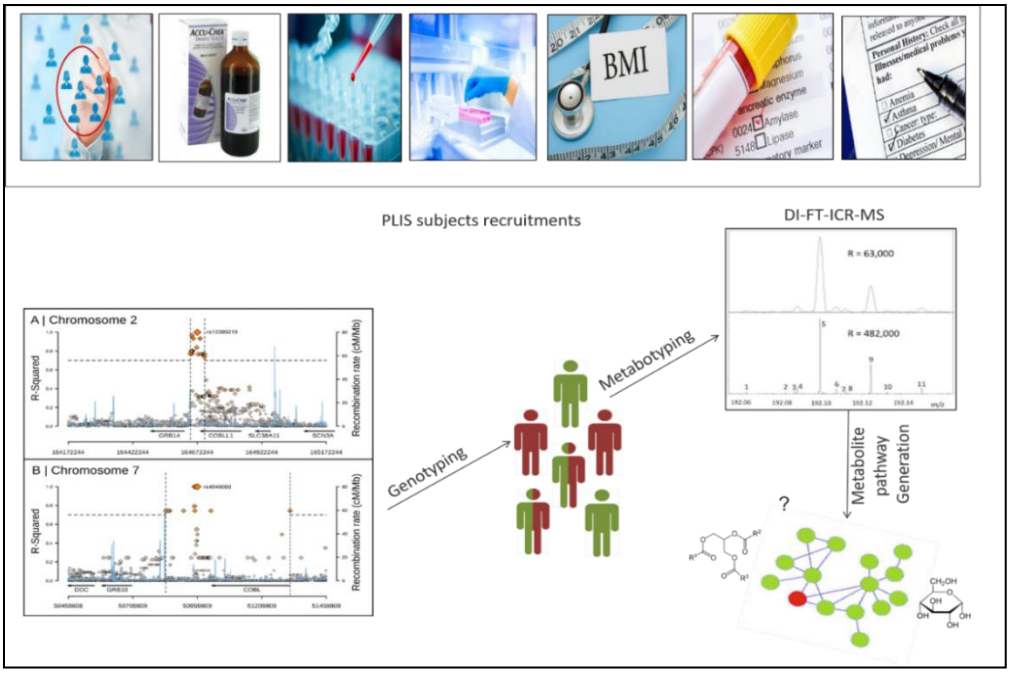


Figure 7: Study objective

In this experimental work, we applied the pipeline established by the Analytical BioGeoChemistry (BGC) research unit at Helmholtz Zentrum München, using the SOPs established by them in the framework of a collaborative project. I was doing the workflow encompassing sample collection, sample preparation (metabolites extraction), sample analysis via DI-ICR-FT MS, and the rest of the steps; data post-processing, statistical evaluation, features assignment to chemical molecular formulas, and annotation via DBs was done by Dr. Forcisi. The last step included the use of mass-difference network analysis (MDiN) for biochemical data interpretation. After the sample collection step, all the steps were conducted at the BGC research unit.

3. Methods

3.1. The Study Cohort (Pre-diabetes Lifestyle intervention Study- PLIS)

Over the past years, several studies have consistently shown that prevention of T2D is possible. These studies used a combination approach of increased physical activity (e.g., at least 30 minutes daily) and a healthy diet. Further goals are a body weight reduction of more than 5%. Details of such prevention programs were already described in the introduction (Tuomilehto et al., 2001; Knowler et al., 2002; Pan et al., 1997; Ramachandran et al., 2006).

As resources for healthcare are limited, the cost-effectiveness of lifestyle intervention must be increased. It has also been shown that some of the individuals at risk benefit from the intervention, while others respond less or not at all to a lifestyle program ("responders" and "non-responders"). The distinction between the two groups should be made in advance. The conventional lifestyle intervention may be sufficient for a subgroup. In contrast, others may need and get an intensified lifestyle intervention and, thus, may benefit to the same extent. For this purpose, factors must be identified that can predict the success of lifestyle interventions.

Pre-diabetic people with relatively high insulin sensitivity and increased insulin secretion at the beginning of the intervention respond better. In contrast, high circulating levels of inflammatory and immune parameters were predictive factors for a relatively weak response to the intervention. In addition, various genetic factors that primarily affect insulin action had a predictive effect on the success of the intervention.

The Tübingen Lifestyle Intervention Program (TULIP) was dealing with the identification of predictive factors for the success of lifestyle intervention with the goal to facilitate individualized prevention of T2D (Uusitupa et al., 2009). The TULIP study involved over 400 people at risk for T2D and cardiovascular disease (obese or obese and/or first-degree relatives of patients with T2D and/or previous gestational diabetes and/or known IGT). The objectives of the intervention were similar to those of the DPS study (Tuomilehto et al., 2001). In addition, subjects were also examined for the genetic background for the risk of diabetes (Schäfer et al., 2007).

At the first follow-up of the TULIP study, the subjects had lost an average of 2.7 kg (approx. 3%) in body weight, 8.7% in total body fat, 14.4% in visceral (abdominal) fat content, and even 30% in liver fat content, and their insulin sensitivity and fitness improved by approximately 15% and 5%, respectively. However, as with other extended lifestyle interventions, the results were slightly less pronounced after two years. At the same time, it was confirmed that subjects responded differently to the intervention and that subjects even deteriorated despite following the recommendations. On this basis, the TULIP study identified valuable, both genetic and environmental factors that predict the success of lifestyle intervention (Schäfer et al., 2007).

Among the factors listed in (Figure 8), insulin secretory dysfunction, decreased insulin sensitivity, and increased liver fat was identified to be general risk factors for the development of T2D. Therefore, these factors should define a high-risk group for diabetes not responding to conventional lifestyle interventions.

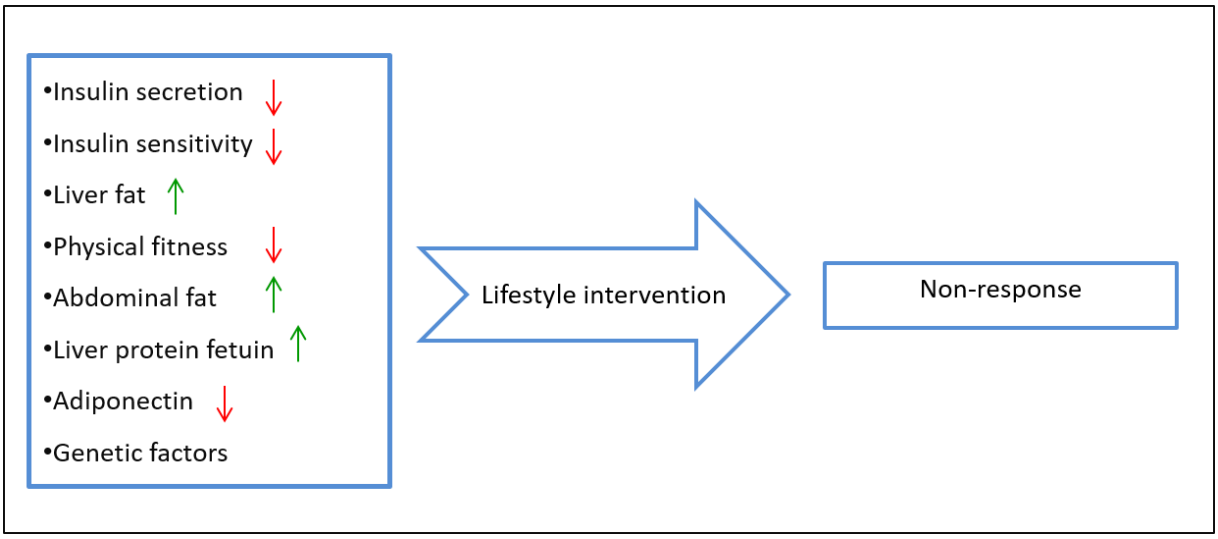


Figure 8: Selection of results from the TULIP studies: factors associated with a reduced response to lifestyle intervention (non-response).

3.1.1. Aims of the PLIS study

The following hypotheses were tested in the PLIS trial:

1. In people with pre-diabetes and at high-risk, intensive lifestyle intervention improves postprandial blood glucose more than conventional lifestyle intervention.

2. With conventional lifestyle intervention, high-risk people with conventional lifestyle interventions show worse postprandial blood glucose over time than those with pre-diabetes and at lower risk.

3. In low-risk people, conventional lifestyle intervention over control (without lifestyle intervention) improves postprandial blood glucose.

4. People with pre-diabetes and at high-risk who are characterized by an insulin secretion disorder show a higher postprandial blood sugar after lifestyle intervention than people with pre-diabetes and high-risk who are characterized by insulin resistance with elevated liver fat content. The difference is independent of intensive or traditional lifestyle intervention.

3.1.2. Study design

The PLIS study is a prospective, randomized, multicenter intervention trial designed to investigate whether intensive lifestyle intervention is superior to a conventional lifestyle intervention in high-risk groups for non-response to diabetes prevention. Besides, intensive phenotyping defines subgroups with an increased risk of T2D (Fritsche et al., 2021).

The study consists of intensive phenotyping at baseline, identifying people with pre-diabetes. These high-risk individuals are randomized into two arms: conventional lifestyle intervention and intensive lifestyle intervention, each with equal numbers of subjects (n=250), whose results are compared at the end of the study. The low-risk population was randomized into two arms: conventional lifestyle intervention and control without lifestyle intervention, each with the same number of subjects (n=250) (Figure 9).

After a screening examination, the intervention arms were followed by a 12-months intervention study. The different therapy groups described above were formed. The subjects in the intensive intervention group received 16 consultation sessions. The subjects in the traditional lifestyle intervention received 8 consultation sessions. The control group got a single education session about a healthy lifestyle. Throughout the study, continuous care by doctors and nutritionists occurred, and the participants had to additionally document nutrition and movement protocols and unbiased measurements.

At the beginning and the end of the study, the examined subjects were extensively metabolically characterized. This also applied to subjects with pre-diabetes who were not in the high-risk group and who did not receive any intervention. Follow-up examinations were carried out 1, 2, and 3 years after inclusion in the study and randomization process. To investigate the long-term effect of the lifestyle intervention and further examinations, follow-up visits after 6, 9, and 12 years will take place.

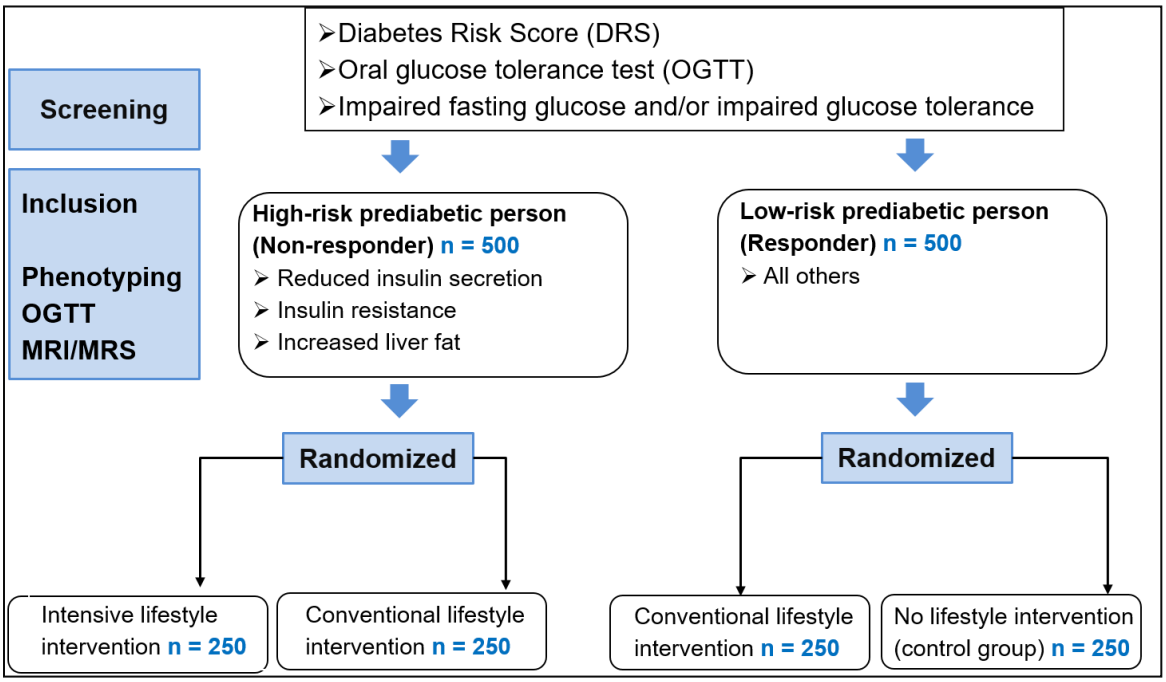


Figure 9: PLIS Study design

A steering group of the German Center of Diabetes Research initiated and designed this prospective, randomized, multicenter trial with 8 study centers (Tübingen, 2 x Munich, Heidelberg, Berlin/Potsdam, Dresden, Düsseldorf, and Leipzig).

3.1.3. Duration and procedures of the study

The study started in May 2012, and the recruitment duration was until June 2017. The design of the study is shown in (Figure 9). The information of participants and informed consent took place before any data were collected or measurements were done (Appendix 1). During the screening visit (V1), a diabetes risk screening test was carried out (DRS, Appendix 1). Recruitment into the study was only possible if the DRS score was 50 or more (Appendix 1). After consent and signature of the screened subject, a dietary protocol (documentation over three days) was requested.

After meeting the inclusion criteria, the baseline examination V0 included; history, complete physical examination, an OGTT, a physical activity test, and a whole-body MRI. The MRI scan was performed up to 3 weeks after the OGTT. Furthermore, blood was collected for DNA analysis. Based on the OGTT values, it was decided whether the subject was eligible for inclusion in the study. This was the case when pre-diabetes was present (impaired glucose tolerance (IGT), impaired fasting glucose (IFG), or both).

Allocation to the high-risk group depended on the insulin values for the calculation of insulin sensitivity and insulin secretion as well as on MRI results regarding liver fat content. The laboratory values were determined centrally in Tübingen and measurement of liver fat locally by the involved MR radiologists.

- **High-risk group**

If a participant was eligible for the high-risk group, randomization was made for one of the following intervention groups:

- The subgroup with intensive lifestyle intervention

- The subgroup with conventional lifestyle intervention
- Control group without lifestyle intervention

- Criteria for the high-risk group: A + B **or** A + C **or** A + B + C **or** B + C

A. Decreased Insulin Secretion: Disposition Index (Insulinogenic Index) * ISI <760

B. Insulin resistance: Insulin sensitivity index (ISI) according to Matsuda / deFronzo <9.2

C. Increased Liver Fat: Magnetic Resonance Spectroscopy: fat content > 5.56%

- **Low-risk group**

The following criteria defined the low-risk group: Participants meeting these criteria were randomized to one of the following intervention groups:

- Conventional lifestyle intervention
- Control group without lifestyle intervention

- Criteria for the low-risk group: **No** A, **or B and** C alone.

A. Decreased Insulin Secretion: Disposition Index (Insulinogenic Index) * ISI <760

B. Insulin resistance: Insulin sensitivity index (ISI) according to Matsuda / deFronzo <9.2

C. Increased Liver Fat: Magnetic Resonance Spectroscopy> 5.56%

After allocating participants to either the high-risk or low-risk group, they were randomized to the treatment arms. At V1, subjects in the intervention groups also received an accelerometer to document exercise and calories consumption.

3.1.4. Recruitment

Participation in the study was voluntary, and the volunteers were invited to the study center. Many efforts were made to address volunteers: advertising, such as local newspapers, flyers, or other channels.

3.1.5. Participant selection

Inclusion criteria were:

- Ability to consent and a signed consent form.
- Age between 18 - 75 years.
- A DRS value is greater than 50.

Inclusion Criteria	Exclusion Criteria
• Increased fasting glucose (IFG), fasting blood glucose between 99 and 126 mg/dl and / Or • Impaired glucose tolerance (IGT), 120 minutes of blood glucose in the oral glucose tolerance test with 75 g of glucose between 139 and 200 mg/dl. Or • Increased fasting glucose (IFG), fasting blood glucose between 99 and 126 mg/dl and / Or • Impaired glucose tolerance (IGT), 120 minutes of blood glucose in the oral glucose tolerance test with 75 g of glucose between 139 and 200 mg/dl	• Diabetes mellitus type 1 or 2 • Existing pregnancy or breastfeeding mother • BMI > 45 kg/m² • Severe symptomatic malignancy weight loss > 10% within the past 6 months • Severe liver or kidney disease (Transaminase elevation > 3 times the upper limit, GFR <50 ml / min/ 1.73 m2) • Systemic infection (CRP > 1 mg/dl) • Therapy with steroids • Drug abuse • Non-compliance during study • Severe mental illness • Serious illnesses such as cancer • symptomatic coronary heart disease • Potentially unreliable subjects • Contraindications for MRI such as any kind of metal in the body [pacemaker, artificial heart valves, metal prostheses, implanted magnetic metal parts[screws, plates of operations] Spiral, metal fragments/ shrapnel, fixed braces, acupuncture needle, insulin pump, tattoos, permanent lid shadow. • Persons with limited thermal sensors or increased sensitivity to heat • Persons who report hearing impairment

Table 4: Inclusion and Exclusion Intervention Study Criteria

Criteria for additional spiroergometry were:

- Acute coronary syndrome
- High-grade cardiac arrhythmias
- Heart failure
- Acute carditis
- Pulmonary embolism
- Acute phlebothrombosis of the lower extremities
- Hypokalemia

3.1.6. Number of study participants

1000 subjects were planned to be included for the PLIS trial in all study centers together (according to the power calculation). However, more than 3000 subjects at risk were screened in all centers. The duration of the study, from the time of written consent of the first subject at V0 to the final examination V17 of the last subject, was expected to be 45 months (May 2012-December 2017). The final total number of subjects screened in our study center was 338.

3.1.7. Randomization

If a screened subject met all inclusion criteria for the PLIS study and none of the exclusion criteria were met, they were invited to participate and sign the consent form.

Randomization was performed on a pre-determined randomization list after (V0) by assigning each included subject the following free number on the randomization list. The randomization was done independently at each study center using randomization lists, which were centrally produced at the University Hospital of Tübingen. Therefore, V0 and the start of intervention V1 should not be more than 2 weeks apart.

Randomization to conventional lifestyle intervention or intensive lifestyle intervention occurred when additional criteria for high-risk (pre-diabetes) were fulfilled:

A+ b **or** A+ C **or** A+ B + C **or** B + C

A) Decreased insulin secretion

B) Insulin resistance

C) Increased liver fat

Randomization to conventional lifestyle intervention and the control group occurred when criteria for high-risk (pre-diabetes) were fulfilled.

3.1.8. Intervention

Overview of the differences between the intervention arms, both concerning the intensive lifestyle intervention as well as the conventional lifestyle intervention. The following goals were defined:

- Reduction of body weight by ≥ 5% in overweight subjects (BMI ≥ 25kg/m^2)
- Less than 30% of calorie intake through total fat
- Less than 10% of calorie intake due to saturated fat
- More than 15g of fiber per 1000 kcal of energy intake

❖ **The intensive lifestyle intervention** consisted of 6 hours of physical activity per week. In addition, the intervention involved 16 sessions a year with a lifestyle counselor.

❖ **The conventional lifestyle intervention** consisted of 3 hours of physical activity per week. The intervention involved 8 sessions a year with a lifestyle counselor.

- **The control group** received one lifestyle counseling of 30 mins duration, where the OGTT results and the other laboratory values were presented and discussed. A target weight was recommended (-5%, if BMI ≥ 25kg/m^2).

A dietitian supervised the participants, and the visits took between 30 and 60 minutes. The sessions consisted of nutritional advice based on the 4 days of dietary protocols, and the guidance indicated how nutritional goals could be achieved. The dietary protocols were evaluated per study center with the "Optidiet' program (version 5).

Furthermore, physical activity was assessed using protocols and pedometers. The protocol included every kind of sports and physical activity besides walking, such as swimming, biking, etc. In addition, to encourage participants, they received a pedometer (Beurer AS 50, activity sensor) to monitor their walking activities. The device documents steps, active time, and distance. Afterward, data were evaluated (Beurer EsyFit version 2.2) on request by the examined person himself during the visit.

3.1.9. Clinical examinations

- **History and physical examination**

During the screening visit (V0), the subject underwent a comprehensive baseline examination, anamnesis, and detailed medical history, performance of medical examinations, anthropometric measurements, and clinical parameters.

Anthropometric measurements	– Height – Weight – BMI – Waist circumference – Waist-to-hip ratio
Clinical parameters	**Basal blood sampling** – Electrolytes – Serum creatinine – Urea – Uric acid – Glutamat-Oxalacetate-Transaminase **(**GOT) – Alanine-Aminotransferase (ALT) – Gamma-Glutamyl Transferase (GGT) – Bilirubin – C reactive protein (CRP) – Iron – Ferritin – TSH – Glucose – HbA1c – Insulin – C-peptide – Cholesterol – Triglycerides – Blood counts – Plasmatic coagulation – Free fatty acids – GAD-II antibodies

Table 5: Anthropometric and clinical parameters/measurements of PLIS participants

- **Oral glucose tolerance test (OGTT)**

The examination was started at 8:00 or 8:30 a.m. after an overnight fasting period of the subject after venous access (Abbocath 20G) has been placed in an elbow vein. After taking a basal blood sample, the subject drank 75g of a standardized glucose solution (OGTT Roche Acuu-Check Dextrose O.G.-T.). The sampling schedule was 0, 30, 60, 90, 120 min (collection volume: 5X20 ml

whole blood). Plasma glucose, insulin, proinsulin, and C peptide were determined. These values were used to calculate pre-specified parameters of insulin secretion and sensitivity.

- **Calculation of insulin sensitivity index (ISI) (Matsuda et al., 1999)**

$$ISI = \frac{10000}{\sqrt{\begin{array}{l} BZN0 \times INS0 \times \\ \left(\frac{(BSN0 + BS30 + BS60 + BS90 + BS120)}{5}\right) \times \\ \left(\frac{[INS0 + INS30 + INS60 + INS90 + INS120]}{5}\right) \end{array}}}$$

- BS = blood Sugar; 0, 30, 60, 90, 120 = Minutes in OGTT
- INS = Insulin

- **Calculation of insulinogenic index (Herzberg-Schäfer et al., 2010)**

$$IGI = \frac{INS30 - INS0}{BS30 - BS0}$$

- **DNA isolation and Genotyping**

After the study cohort was completed, we ran DNA isolation from EDTA plasma samples using the Qiagen DNeasy Blood& Tissue Kit (cat. nos. 69504 and 69506) in our laboratory in Freising (Appendix 2). Afterward, the samples were genotyped for SNP rs6712203 in the *COBLL1* gene using 15 SNPs (RQ-008136). Double strand DNA was isolated from EDTA full blood or buffy coat using the DNeasy® Blood& Tissue Kit with DNeasy Mini spin-columns (QIAGEN, Hilden, Germany) following manufacturers' protocol. The genotyping was partially performed by GENEWIZ® UK (Takeley, Essex, United Kingdom) and at the Broad Institute of MIT and Harvard

(Cambridge, MA, USA). Following are the primer sequences for genotyping rs6712203 (COBLL1). The annealing temperature is 58°C.

GAA CTC TCC ACT ACC ATT GC	for_seq_COBLL1
CAA AAT TCC TTC CTT GCC AG	rev_seq_COBLL1

3.1.10. Standardized health questionnaires

The subject was asked to complete standardized health questionnaires to the fullest on the following items:

- Diabetes Risk Test (DRT) (Appendix 1)

- Physical activity, calculation of work, sport and leisure-time physical activity index

- PHQ-D health questionnaire for patients (Appendix 1)

- SCL-90-R

- Visual analog scale (VAS)

- Medical history

3.1.11. Blood volume and storage of blood samples

The quantity of blood taken during the initial phase of the study (1 year) was approximately 310 ml. Then, in the second and third year, respectively, once again 100 ml, divided into 5 blood sampling times per OGTT. The blood and DNA samples were given an ID number and stored locally at - 80C. The procedures followed SOPs provided by the Tübingen group.

3.1.12. Magnetic resonance imaging and Spectroscopy

Accurate quantification of body fat compartments and determination of ectopic fat deposits, such as liver fat, was a part of the assessments within the study. The examination was done in the morning in a fasting state at the University Hospital (München Rechts der Isar) and took about 45-60 minutes. The person evaluating the results of the MRI scans was blinded, which means they did not know to which study arm the subject was assigned.

Before the spectroscopic measurements, images were taken utilizing MR imaging (MRI) to assess body fat. First, total body fat (TAT), visceral adipose tissue (VAT), and abdominal subcutaneous adipose tissue (SCAT) volumes were cautiously recorded using axial umbilical T1-weighted MRI scanning (Schulze et al., 2012; Schwenzer et al., 2010). Next, proton magnetic resonance spectroscopy was applied to determine the liver fat content using a 3.0 T whole-body imager (Magnetom Sonata, Siemens Medical Solutions, Erlangen, Germany) (Machann et al., 2006)

The volumes of the adipose tissue compartments were determined as described by Machann et al. (Machann et al., 2005). In addition, the MRI examination program of the local study center of the Technical University of Munich included an additional measurement of the fat content in the bone marrow and assessment of the saturated and unsaturated fatty acids in the subcutaneous, visceral fat.

In addition, the MRI examination was also optionally provided for participants who did not meet the inclusion criteria for the intervention, and the examination program was identical. All MRI examinations were carried out in cooperation with the Department of Diagnostic and Interventional Radiology at the Klinikum Rechts der Isar of TUM (Head: Prof. Dimitrios Karampinos)

3.1.13. Bioelectrical impedance measurement (BIA)

Bioelectrical Impedance Analysis (BIA) was used to determine body fat percentage and lean body mass. The measurement is based on a weak alternating current via two electrodes; via two further electrodes inside this field, the voltage drop and the phase shift of the signal voltage are measured (four-wire measurements). BIA measurement was usually taking place at every visit.

3.1.14. Spiroergometry for determination of physical fitness

Spiroergometry is a method for determining cardiovascular/pulmonary capacity. The examination takes around 20 min. It consists of measuring oxygen consumption and carbon dioxide production during exercise on a cycle ergometer, where the wattage is continuously increased (steep ramp). In addition, blood pressure and heart rate are automatically measured throughout the test. The following instrument was used: ergometrics 800 S (bicycle ergometer); Ergoline GmBH& Co. KG, with electromagnetic barking cycle. Analysis software: MetaSoft CPX software (lead display 3); CORTEX Biophysics GmBH, Bitz, Germany.

Study participants were instructed to hold a speed of 60 U/ min throughout the test. After a warm-up period of about 2 minutes at 0 W, the test started with a force of 20 W, and each minute was followed by a gradual increase in resistance by 40 W to exhaustion. Lean body mass was given as (ml/min) per kilogram of lean body mass.

Spiroergometry was performed in cooperation with Prof. Renate Oberhoffer-Fritz of the Chair of Preventive Pediatrics.

- **Spiroergometry Termination criteria**

Subject's exhaustion, RR over 200 mmHg systolic and 120 mmHg diastolic, heartbeat over 200/ minute were defined criteria for termination, at the discretion of the attending physician.

3.1.15. Privacy policy

The study was conducted following the Helsinki Declaration and GCP guidelines and the German Data Protection Act principles.

The data were initially recorded on paper and electronic data carriers. The transfer of the collected data to third parties (i.e., persons entrusted with the further processing of the data), the evaluation as well as the publication of the data took place exclusively in a pseudonymized form (a number replaced, i.e., the name of the test person and other identification features to the purpose to rule out the identification of the person). The name list corresponds to which number is kept locked in each study center and only possible for the principal investigator and the study's data monitoring and safety board. Medical confidentiality is required for all data collected. A central database was established at the University Hospital Tübingen for all participating centers.

In the case of incidental MRI findings, the morphological changes were communicated to the patient, and they were asked to have them clarified further.

If study participants withdrew, they had the right to decide whether their data may be deleted or the existing data or sample may continue to be used anonymously.

3.1.16. Database

The data were recorded in an access database. This database is located in a central server at the German Center for Diabetes Research (DZD). For this purpose, a "partner network" (Microsoft Share-Point system) was set up, which allows common information and documents, for example, standard operating procedures (SOPs), to be managed. Therefore, access to this common database was warranted.

The data were initially collected via hard copy Case Record Forms (CRFs) and entered into the database by hand. The construction of the Electronic Case Record Form (eCRF) for all studies and its implementation in the partner network was performed by the DZD study-coordinating

center. The “Koordinierungszentrum für Klinische Studien” (KKS) of the University Hospital Düsseldorf was responsible for data monitoring. The software was set up to directly transfer laboratory data from central laboratories to the main database in the partner network.

The collected samples from the study subjects were stored in different locations: Blood samples (the measurement of insulin and NEFA) were collected from each center and shipped for central storage in Tübingen. The rest of the samples (blood and urine) were stored at the Institute of Nutritional Medicine in Freising. The shipping and storage of samples followed the respective SOPs of the study.

3.1.17. Data storage and safekeeping

According to the valid data protection law, the PLIS trial data is stored for 20 years. The biosamples have an unlimited period of storage in Tübingen, where the central biobank is located. This is true for the subject who signed the separate biobank informed consent form.

3.1.18. Encryption

All personal data is stored in a pseudonymized form. An assignment of the data to the respective study subjects will only be possible for the participating investigators (principle investigator of the respective study center, one additional person to be named per study center). All those involved were subjects to medical confidentiality and were allowed to use this right only in case of emergency.

3.1.19. SOPs

In the PLIS study, all procedures followed commonly established SOPs by the study center under the lead of the University Hospital Tübingen and the DZD to achieve a high level of conformity among the national study centers.

- The day **before the examination,** we label the tubes: Patient ID, time point, and sample type, especially For E + HCL and E+ A.

- **<u>On the morning of the examination:</u>**
  - Take the glucose cuvettes out of the refrigerator (if the cuvettes are taken out of the freezer, they need about 30 minutes to come to room temperature)
  - Cool the centrifuge down to 4°C once while it is idling
  - When the participants are there, pipette 80 µl of aprotinin into the EDTA tubes E + A

- **<u>blood collection:</u>**
  - Fill out the lab sheet with the patient information: date of birth, ID, and Initials
  - One 2.7 ml EDTA tube per participant is frozen at -80°C without centrifugation (label: DNA patient pseudonym 3 digits)

- **<u>Aliquoting for study participants</u>**:
  - Cyrus are labeled with particular Etiquette (PLIS, ID, Date of the visit).
  - According to the blood sugar measurement V0, at V10 or V17, aliquots are made in a complete set if the inclusion criteria are met.
  - At V0, 2 centrifuged EDTA tubes (9 ml, use supernatant for aliquoting) and the RNA tube are stored in the refrigerator, while for V10 and V17, the centrifuged EDTA tubes are frozen at - 80°C.
  - A box is created for each participant and labeled with PLIS TUM V0, 10 or 17, and patient ID
  - The Cyrus for insulin and NEFAs are frozen in an extra box at -80°C until dispatch.

The following (Table 6) and (table 7) showed the blood withdrawal checklist and aliquoting for each PLIS participant.

Tubes	0	30	60	90	120
PAXGene (2,5 ml)	✓				
EDTA (4 x 9 ml)	✓				
EDTA + HCl (2,7 ml)	✓	✓	✓	✓	✓
EDTA + A (2,7 ml)	✓	✓			✓
Citrate (9 ml)	✓				
Li-Heparin (5,5 ml)	✓				
Serum (7,5 ml)	✓				
Serum (5,5 ml)	✓	✓	✓	✓	✓
NaF (2,7 ml)	✓				

Table 6: PLIS blood withdrawal checklist

Tubes	0	30	60	90	120
PAX Gene(2,5 ml)					

EDTA (4 x 9 ml)	24 x 300 µl				
EDTA + HCl (2,7 ml)	1 x 500 µl	1 x 500 µl	1 x 500 µl	1 x 500 µl	1 x 500 µl
EDTA + A (2,7 ml)	2 x 300 µl 1 x 500 µl	1 x 500 µl			1 x 500 µl
Citrat (9 ml)	10 x 300 µl				
Li-Heparin (5,5 ml)	5 x 300 µl				
Serum (7,5 ml)	20 x 300 µl				
Serum (5,5 ml)	1 x 1000 µl	1 x 1000 µl	1 x 1000µl	1 x 1000µl	1 x 1000µl
NaF (2,7 ml)	3 x 250 µl	3 x 250 µl			3 x 250 µl
Urin (10 ml)	3 x 1000 µl				

Table 7: PLIS Aliquoting for each participant

I was involved in the PLIS team from the beginning and had specific responsibilities for the study until the present time. First, I was involved with the recruitment procedure, starting with advertising, screening, and obtaining consent. Moreover, I was a member of the study team with many duties such as handling the whole sample set after blood withdrawal in the lab, including centrifuging, pipetting, storing, and sending samples packages to our head study center in Tübingen. In addition, I was assisting the Spiroergometry challenge and was running all the physical examinations for the patient, starting with the height, weight, and BIA measurements.

Moreover, I analyzed the dietary protocols of the subjects. Finally, I was responsible for completing and correcting CRFs and eCRFs entries.

3.2. Metabolomics: pre-analytical sample preparation

The increasing number of metabolomics applications in clinical research has led to a significant demand for the pre-analytical processes to be standardized during blood collection, blood preparation, storage, and transport (Leichtle et al., 2013). This necessity is specifically high in clinical studies, where poor sample quality may heavily bias the final results. Therefore, the pre-analytical phase needs to be tightly controlled to avoid adverse effects on the metabolite measurements. This process is usually rigorously regulated by standard operation procedures (SOP) (Yin et al., 2013). Such SOPs are commonly used during clinical trials, and they are of great importance for targeted and untargeted clinical metabolomics studies (Yin et al., 2015). The following section describes the pre-analytical steps referring to untargeted metabolomics studies, from the experimental design to the final interpretation of the data as practiced in this study.

3.2.1. SOPs establishment and experimental design

The blood metabolite pattern is a compactly controlled balanced system. Still, a diversity of exogenous factors and physiological conditions may lead to dynamic changes in blood metabolites with possible substantial effects on the pre-analytical phase (Yin et al., 2013). Furthermore, the blood metabolome composition is also affected by multiple intrinsic and extrinsic factors, including physiological rhythm, circadian, sex, age, diet, exercise, drugs, and nutritional supplements such as vitamins (Gibney et al., 2015; Minami et al., 2009; Narayanan et al., 2000). Therefore, planning of the conditions is needed before sample collection for any metabolomics studies (Yin et al., 2015).

3.2.2. Sample collection

Standardized collection methods are essential to compare different samples, preserving long-term quality, reproducibility, and stability. The principal variability points involve; the kinds of additives used for blood collection to avoid blood clotting, the types of collection tubes, the handling temperature, and the degree of hemolysis, sample storage, and freeze-thawing cycles. In addition, these checklists are critical in the context of untargeted metabolic technologies where the interest is to detect the maximum extent of metabolites in terms of mass range (50-1200 m/z) and chemical properties.

- **Tubes:** In the PLIS study, we used Plastic Tubes for sample collection with different anticoagulants (EDTA, NaF, Serum, Li-heparin, and citrate), and to avoid Plastic polymers effect on the metabolites, a tube pre-test was done. We preferred EDTA, NaF, and citrate blood-collection tubes for metabolomics investigations because of their stability.

3.2.3. Sample preparation

Sample preparation is a crucial step, especially in untargeted metabolomics studies. Plasma is a very complex matrix. Therefore, an incisive sample pre-treatment is fundamental.

We used solid-phase extraction (SPE) to
To avoid hemolysis, we placed whole blood immediately in ice water. Next, it was centrifuged at 4 °C, in order to the obtained plasma was then aliquoted in small volumes of 150 - 200 µl to avoid freeze-thawing.

In our study, the preparation of blood plasma for the DI-ICR-FT MS analysis was conducted according to the protocol of Forcisi et al. Before the procedure; the initially frozen plasma samples were thawed on ice and vortex mixed for 30 s. to avoid degradation and protein kinase activation that can occur at room temperature. Afterward, 50µL of plasma was transferred into a tube containing 50 µL of 2% phosphoric acid. The resulting 100 µL solution was vortex mixed

for 30 s. The extraction of metabolites was performed by solid-phase extraction (SPE) technology to extract the most substantial number of metabolites and remove compounds that could interfere with a mass-spectrometer (e.g., salts) or with the chromatographic system. We used Omix C18 100 µL tips (Varian, Palo Alto, California, USA) and followed the manufacturer's instructions.

The conditioning and equilibration steps (before loading the sample onto the SPE tip) included flushing the tips with methanol and 2% formic acid, respectively. After loading the sample, the tips were washed with 2% formic acid three times, followed by methanol's elution step. Finally, the eluate was stored in a new tube at - 80C until further processing.

After we prepared the samples and handled them following the aforementioned SOPs, they are now ready to be measured through the Platform Direct Infusion-Ion-Cyclotron-resonance Fourier-Transform Mass-Spectrometry (DI-ICR-FT MS).

3.2.4. Direct Infusion-Ion-Cyclotron-resonance Fourier-Transform Mass-Spectrometry (DI-ICR-FT MS)

Our study used a platform based on direct infusion-Fourier transform-mass-spectrometry (DI-ICR-FT MS). This technology provides an extraordinarily high mass-resolution, and mass-accuracy, together with a high-order-of-magnitude intensities range for metabolite detection. Asaph Aharoni first applied it to metabolomics in 2002 (Ghaste et al., 2016). Since then, several studies have been published in different fields, including plant sciences, wine analysis, and nutritional studies (Pinu et al., 2019). As a result, it is considered the mass-spectrometric technology with the highest mass-accuracy and resolution (Pinu et al., 2019).

Its principle of mass-detection is based on the circular oscillation that charged ions exhibit once they are introduced into the homogenous magnetic field (Forcisi et al., 2013). Thus, DI-ICR-FT MS is a type of mass-spectrometry, which determines the mass-to-charge ratio (m/z) of ions, in a fixed magnetic field, based on the cyclotron frequency of the ions (Forcisi et al., 2013).

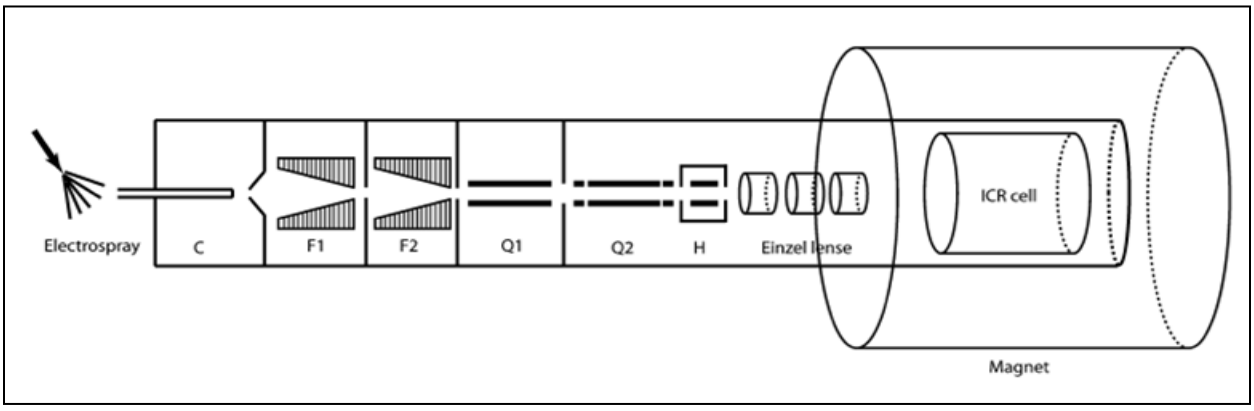


Figure 10: Schematic representation of DI-ICR-FT MS. Reprinted with permission from Elsevier (Forcisi et al., 2013)

The DI-ICR-FT MS technology consists of three essential modules: an *ion source,* which transforms the molecules in a sample into ionized fragments; a *mass-analyzer,* which sorts the ions by their masses by applying electric and magnetic fields; and a *detector,* which measures the value of some quantity indicator and thus provides data for calculating the abundances of each ion fragment present (Forcisi et al., 2013).

The following steps explain how the instrument works. After the injection, the molecules of interest are first introduced into the ionization source of the mass-spectrometer, where they are first ionized to acquire positive or negative charges. In our case, we set a positive ionization mode. The ions then travel through linear ion beam guides to reach the mass-analyzer and oscillate at different cyclotron frequencies, inversely proportional to their m/z ratios. A time-domain transient is generated during ion detection and recorded by a computer system. The computer subsequently performs a Fourier Transform (FT) to the time-domain transient and converts it to the frequency spectrum, which can be converted afterward to a mass-spectrum.

Finally, a mass-spectrum displays the signals graphically as a histogram showing the relative abundance of the signals (intensity) according to their m/z ratio (Forcisi et al., 2013).

3.2.5. Electrospray Ionization ESI

In most metabolomics MS applications, the method of choice is electrospray ionization (ESI) because of its relative soft ionization nature and no requirements on prior chemical modification (Gibson et al., 2010). Most scientists favor it since it has a well-balanced ionization efficacy in terms of chemical compound classes and is a soft ionization method since there is very little fragmentation (Gibson et al., 2010). In addition, this ionization method does not rely on rapid but on the gentle vaporization of the solvent since ions can be generated in solution (Cech et al., 2001). The ESI capillary, which introduces the sample spray into the system, is placed at a close distance to a counter electrode. The potential difference between the ESI capillary and the center electrode ranges between 3000V and 4500V. This electric potential difference depends on the distance between the ESI capillary and the counter electrode. An electric field of 1000V/cm is maintained (Cech et al., 2001).

The process involves transmitting sample ions from the solution to the gas phase by subjecting a strong electric field at atmospheric pressure. Vaporization of the solvents containing charged analytes leads to colombic explosions of the spray droplets, releasing ions having multiple charge states z to the gas phase (Forcisi et al., 2013; Banerjee et al., 2012). The generated ions are collected by a skimmer and guided through ion funnels, Quadrupole 1, 2, and hexapole to the analytical part (Forcisi et al., 2013; Cech et al., 2001). They arrive afterward at the ion cyclotron resonance (ICR) cell, where the ions are confined and detected. The cell is located in the center of a super conducting magnet that provides a *magnetic field* with the highest possible homogeneity (Forcisi et al., 2013; Cech et al., 2001).

Ions are cohesively excited to a larger radius orbit by using a pulse of the radio frequency electric field, and the detection of their image charge is on receiver plates as a time-domain signal. The time-domain signal Fourier transformation results in a frequency domain signal converted to a mass-spectrum, based on the inverse relationship between cyclotron frequency and m/z (Forcisi et al., 2013; Nikolaev et al., 2013). The whole process of electrospray ionization is summarized in (figure 11).

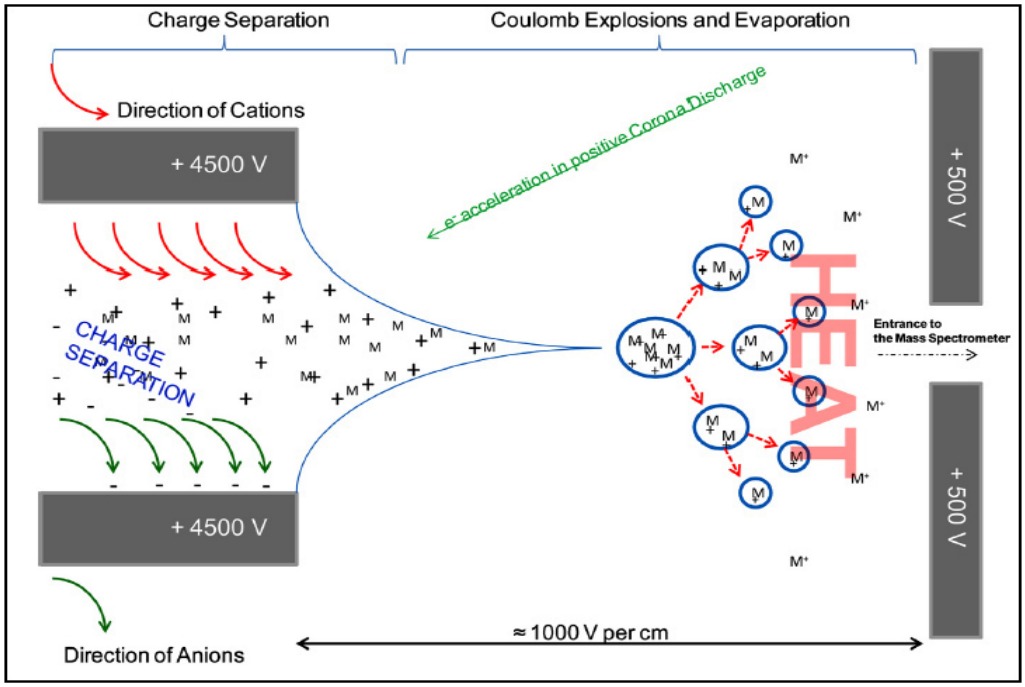


Figure 11: Schematic representation of the electrospray ionization (Gibson et al., 2010). Reprinted from *Journal of Chromatography A*, 1292, 2013 with permission from Elsevier.

3.2.6. The instrumental analysis

Mass-spectrometry (MS) and nuclear magnetic resonance (NMR) are the most commonly used detection methods (Villas-Bôas et al., 2005). Therefore, in our study, the first question necessary to be answered was the suitable kind of anticoagulant in DI-ICR-FT MS applications in terms of information, ion suppression due to the specific matrix effects, and artifacts possibly derived from the different collection tubes. Often there is a difference, especially in how the

samples need to be analyzed; therefore, we considered automation, batch analysis, and the use of quality control (QC) or comparison of the different batches ((Villas-Bôas et al., 2005).

A total of 381 plasma samples were analyzed via DI-ICR-FT MS. Prior analyses, the metabolites (from 50µl of blood plasma) were extracted by C18 solid-phase extraction (SPE) technology, using Omix C18 100µl tips (Varian) and following the protocol described in Forcisi et al. The extracts, diluted in methanol by a factor of 50, were analyzed in positive electrospray ionization mode (ESI) via DI-ICR-FT MS, using a Bruker SolariX instrument equipped with a 12 T magnet (Bruker Daltonic GmbH, Bremen, Germany). The instrument was externally calibrated by injecting a 10 µg/ml solution of arginine and observing corresponding peaks with m/z values equal to 175.11895 (M+H)+, 349.23062 (2M+H)+, 523.34230 (3M+H)+, 697.45397 (4M+H)+. In the experiment, the infusion flow rate was set to 120 µL/h. 400 scans, each corresponding to 4 mega words in the interval from 147.4 to 1000.0 m/z, were acquired and averaged. The time of accumulation ion t was set to 0.7 s, and the time of flight to the detector was set to 1 ms. The voltages of capillary and spray shields were set to 3800 V and -500 V, respectively. The flow rate of nebulizer gas was kept at 2.2 bar, and the drying gas flow rate was set to 4 L/min [at a temperature of 180^{o} C].

3.2.7. Data processing

The collected data from the instrument are usually multi-dimensional, including interference from chemical noise. The data post-processing commonly includes data reduction, de-noising, metabolite extraction, alignment, and matrix generation (Bijlsma et al., 2006). The detected metabolites are small molecules with molecular weights below 1 kDa. Features peaks with specific retention time and mass to charge ratio *m/z*); peak area, which is usually the preferred parameter to represent the relative abundance of each metabolite in different samples (Liu et al., 2017).

3.2.8. The result interpretation

In this step, the chemical structure of the potential biomarkers and biochemical mechanisms pathways should be examined (Eichner et al., 2014). The increased availability of high-resolution mass-spectrometry (HR-MS) in chemical analysis has dramatically improved the detection and identification of compounds (using MS/MS techniques) in different matrices. However, confidence in this HR-MS-based determination varies among studies and substances since it is not always possible or even meaningful to synthesize each element or confirm them via complementary methods.

3.3. Analysis of the metabolomics data

3.3.1. ORA: Over-Representation Analysis

The over-representation analysis represents a knowledge-driven method commonly used for estimating the significance of specific pathways involved in the investigated phenomenon (Eichner et al., 2014). In our study, the metabolic features are first scored (based on a statistical model) and projected, if possible, on the corresponding biochemical pathways. Second, a statistically interesting subset is taken from the entire set of features according to the calculated score. Then, the probability of finding labels of pathways or compound classes within the chosen subset is computed, indicating their over-representation in the subset (Eichner et al., 2014).

3.3.2. Over-representation analysis of compound classes (database driven)

Frequently, little or inscrutable information can be obtained based merely on individual metabolites shown to be significant in a study (Eichner et al., 2014; Schymanski et al., 2014; González-Domínguez et al., 2017). Thus, special attention is paid to investigating the metabolite patterns' mutual behavior or their cooperative dynamics. The study applied broad patterns untargeted metabolites methodology to reveal hidden 1000 compound classes via different ORA

tools. Metabolites are involved in different biochemical pathways affected by a particular biological phenomenon of interest (González-Domínguez et al., 2017). To assess these effects mathematically, several methods have been extensively utilized (González-Domínguez et al., 2017). First, we resorted to over-representation analysis (ORA), representing a knowledge-driven method that estimates the significance of specific pathways involved in the investigated phenomenon (Ghaste et al., 2016). Next, the metabolic features are scored based on a statistical model and projected, if possible, on the corresponding has known biochemical pathways by mapping them onto a database such as HMDB. Secondarily, a small subset is taken from an entire set of features according to their significance. Afterward, the probability is computed, indicating whether a specific pathway is significantly enriched by assessing whether corresponding metabolites are over-represented in the considered subset (Ghaste et al., 2016).

3.3.3. Mass-difference enrichment analysis (over-representation analysis of mass-differences)

A way to monitor (bio) chemical reactions and consequently the activity of genes with enzyme-encoding exons is to investigate the mass-differences (MDs) between detected m/z-signals (Ghaste et al., 2016). The MDs between each substrate and product of a biochemical reaction is specific for each reaction type, where two metabolites/substrates react to give a product. Untargeted metabolomics enables the investigation of wide chemical space, including known and unknown metabolites. We annotate thousands of possible metabolites belonging to different compound classes at the hand of databases. Using mass-difference analysis (MDA), we could assign molecular formulas with high accuracy and explore the nature even of signals unknown to databases (Ghaste et al., 2016). In this work, we used MDA, and a graph-based approach, to examine the mass-differences between each feature and map them against a theoretical library of mass-differences that describe possible occurring biotransformation.

3.3.4. Mining of the DI-ICR-FT MS data

The acquired spectra were exported using the vendor software (Data Analysis) provided by Bruker Daltonics M/z peaks were picked at S/N ≥ 4, and a minimum intensity of 1.5* 10^6 counts was set. Gibbs peaks were removed by Data Analysis automatically. All spectra were exported as tab-separated as-files and calibrated, according to Smirnov et al., 2019. Using an in-house peak alignment algorithm, these spectra were aligned with the mass tolerance window set to 0.5 ppm. M/z features that occurred in less than 10% of all samples were discarded in the generated matrix. Molecular formula assignment was performed following the mass-difference network approach, described in Tziotis et al., 2011 and Moritz et al., 2017. Finally, compound class enrichment analysis was applied for the whole population (female, heterozygous).

Since mass-difference enrichment analysis (MDEA) was implemented to infer molecular pathways based on statistics, one of the essential steps in MDEA is selecting the theoretical mass-differences that allow for a targeted investigation of the imbalance caused by *COBLL1* in circulating blood (Moritz et al., 2012). Therefore, we created a list of co-regulated genes of *COBLL1* in order to map the mass-differences related to the considered mechanism.

I used the list of the co-regulated genes from Claussnitzer et al. (figure 12). The Claussnitzer group created a list of *COBLL1* co-regulated genes, where they investigated co-expression analyses from human adipocyte microarray gene expression data and found enrichment for genes from the integrin and inflammation signaling pathways for positive correlation with *COBLL1* mRNA expression (Figure 12) and genes from the gonadotropin-releasing hormone receptor and p38 MAPK pathways enriched for the negative correlation. Interestingly, in data from the gene enrichment profiler (Benita et al., 2010), a database that uses publicly available microarray data from 142 (16 cancer, 126 average) different tissues and Cells for gene expression analysis – the group also found *ADAM12* co-regulated with *COBLL1*. *ADAM12* was previously reported to induce actin remodeling during adipogenesis by regulating β1 integrin function (Kawaguchi et al., 2003). Integrins enable outside-in signaling from the extracellular

matrix (ECM) into the cell (Morandi et al., 2016). As such, integrins are involved in cellular processes like adipocyte proliferation and differentiation. It is not yet completely understood how integrins transmit ECM signals into the cells, but F-actin has been shown to play a role in this process (Morandi et al., 2016). In prostate cancer cells, *COBLL1* has been demonstrated to enable actin filamentation, suggesting a comparable function in adipocytes (Mishra et al., 2019). Furthermore, Claussnitzer et al. found *actinin alpha 4 (ACTN4)* and Actin-related protein 2/3 complex subunit 2 (*ARPC2)* co-expressed with *COBLL1. ACTN4* links filamentous actin to membrane proteins and is involved in insulin-stimulated *GLUT4* trafficking in muscle cells (Talior-Volodarsky et al., 2008). *ARPC2* is one subunit of the *actin-related protein 2/3 complex subunit 2 (Arp2/3)*, which is responsible for the nucleation of branched actin filaments (Weaver et al., 2003). Large-scale Affinity Capture-MS experiments found an interaction of *COBLL1* with protein kinase C and casein kinase substrate in neurons protein 1 (*PACSIN1)*, *PACSIN2,* and *PACSIN3* (Stark et al., 2006). *PACSINs* are highly conserved Src-homology 3 (SH3)-domain-containing proteins involved in cytoskeletal mediated membrane

They are trafficking through interaction with WH2 containing proteins like the *Arp2/3* complex activator *N-WASP* (Kessels et al., 2004).

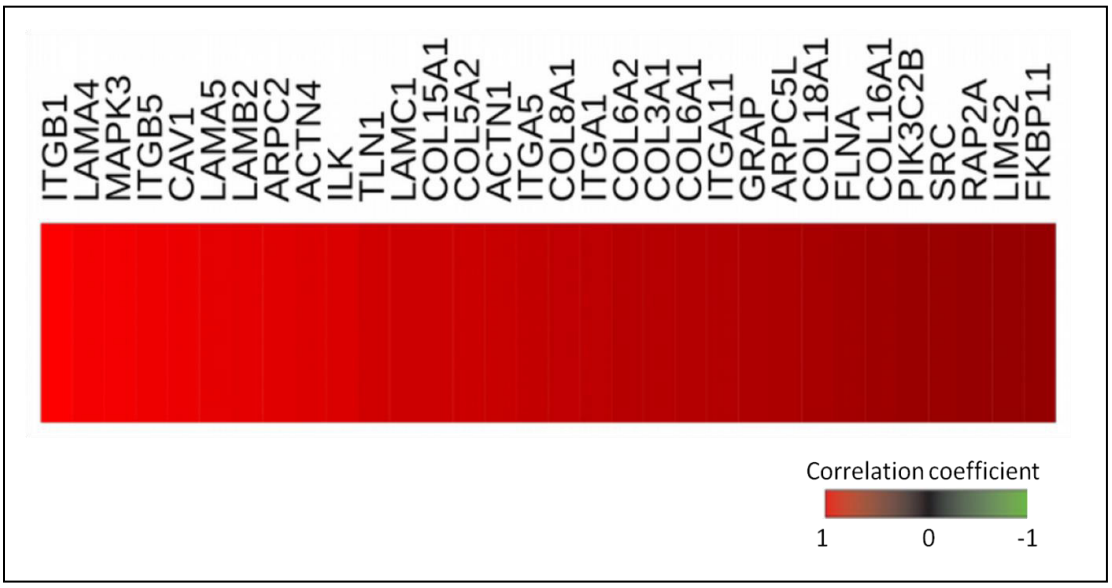


Figure 12: Positive correlation of *COBLL1* mRNA expression (*COBLL1* probe ILMN_1761260) with the expression of integrin pathway genes. Pearson's correlation coefficient between *COBLL1* and each gene was calculated using microarray data from primary hAC. In addition, Gene enrichment was tested using the PANTHER pathway analysis. The top-scoring pathway was the positively correlated integrin pathway (p-value = 6.32x10-8). All genes listed in this pathway and significantly correlated with the expression of *COBLL1* are shown in the heat map, where red indicates a positive correlation and black no correlation**.** (Data under revision).

In my thesis work, I extended the list by searching for more involved co-regulated genes of *COBLL1* with respect of mechanisms of interest, using several databases, including Uniprot (Consortium, 2019), Reactome (Jassal et al., 2020), and HMDB (Wishart et al., 2009). As a result, after modifications on the list, we identified in total 44 co-regulated genes of *COBLL1* (Appendix 3).

(Table 8) summarizes the final 11 genes extracted from the original modified list of co-regulated genes (Appendix 3) that we included in our dataset.

Gene	Protein
PIK3C2B	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta
ADAM12	Disintegrin and metalloproteinase domain-containing protein 12
ITGB1	Integrin beta-1
ITGA5	Integrin alpha-5
SRC	Proto-oncogene tyrosine-protein kinase Src
ITGB5	Integrin beta-5
MAPK3	Mitogen-activated protein kinase 3
PPAR-γ	Peroxisome proliferator-activated receptor gamma
PLIN5	Perilipin-5
ILK	Integrin-linked protein kinase
DGAT2	Diacylglycerol O-acyltransferase 2

Table 8: Selection of co-regulated genes of *COBLL1*. The table shows a selection of 11 genes involved in our analysis dataset. We have excluded the genes encoding proteins involved in the phosphorylation of proteins since we cannot detect and follow any of these MDs.

We created a list of candidate mass-differences based on the selected genes and merged them to a manually curated list of theoretical MDs. Finally, the comprehensive list of MDs was searched in our experimental mass-difference space.

4. Results

4.1. The PLIS Study Cohort

The duration of the study, from the time of written consent of the first subject at V0 to the final examination V17 of the last subject, was expected to be 45 months (May 2012-December 2017).

The final total number of subjects screened in our study center was 338 (table 9). Further description of the subject characteristics is presented in (Appendix 4).

Variable			
	Male **Mean±S.D.**	**Female** **Mean±S.D.**	**P-value**
Total	**102**	**236**	
Age V0	50±11.0	49±11.1	0.441
Height (cm)	180±6.7	166±6.9	1.128
Weight (kg)	104±21.0	87±19.9	1.103
BMI (kg/m²)	32±6.0	32±6.9	0.705
Waist (cm)	110±14.6	115±15.0	5.487
Hip (cm)	111±13.0	1.0±15.7	0.022

WHR (cm)	1.0±0.0	1.0±0.1	6.447
WHtR (cm)	1.0±0.0	41±0.1	0.241
Fat (%)	28±6.7	36±7.8	9.685
Fat mass (kg)	30±12.3	50±14.2	9.311

Table 9: Anthropometric and clinical parameters of the PLIS subjects at the TUM study center. P-value was calculated using a t-test.

The rest of the PLIS cohort subjects, not included in the study, were defined as screen failures. Those either met the exclusion criteria, were healthy, or couldn't complete their OGTT for any health reason. Eventually, their samples were frozen at - 80 C and considered for any future planned projects.

4.2. Experimental design and metabotyping via DI-ICR-FT MS

After accomplishing the PLIS cohort recruitment, which contains males and females with different heterozygous carriers and homozygous allele carriers. In my thesis, we investigated if the intronic *COBLL1* variant rs6712203 reveals a nutritionally induced impact on the metabolic phenotype (metabotype). We included from the whole PLIS cohort only female and homozygous as depicted in (figure 10), we performed the metabotype profiling of 127 female subjects (91 *COBLL1* rs 6712203 homozygous risk allele carriers/36 homozygous non-risk allele carriers) (Table 10) (Appendix 5 for the extended table 10), recruited from a pre-diabetes clinical cohort, employing ultra-high resolution mass-spectrometry. In addition, all the participants underwent an OGTT, and sodium fluoride blood plasma (NaF) was collected at different time points (0, 30, 60, 120 minutes).

Variable	Females		
	Non-risk Mean±	Risk Mean±	p-value
Total	36	91	
Genotype	rs6712203	rs6712203	
Age (years)	48±11.2	48 ±10.8	0,259
BMI (kg/m²)	32±7.8	32±7.2	0,828
Waist circumference (cm)	97±15.3	99±14.6	0,718
WHR (cm)	1.0±0.0	1.0±0.0	0,452
WHtR (cm)	1.0±0.0	1.0±0.0	0,911
Fat (%)	39±10.1	41±7.3	0,499
Fat mass (kg)	36±16.1	36±14.6	0,882

Table 10: Study cohort. Total of 127 females (91 *COBLL1* rs 6712203 risk allele carriers/36 non-risk allele carriers)

After the plasma samples were analyzed via DI-ICR-FT Ms, the data went through a pipeline of analysis, including mining of the DI-ICR-FT MS data, molecular formulas assignment by means of mass-difference-network (MDiN) analysis, and database annotation for the visualization of metabolic patterns. All these steps enabled the data interpretation, and the generation of hypotheses for future investigations.

4.3. Analysis of metadata and quantitative metabotypes

We investigated the differences of several clinical blood parameters between the two-genotype groups. (Figure 13 and 14) depict the levels of glucose and non-esterified fatty acids (NEFA), respectively, across the different time points of the OGTT. In both cases, we observed no significant differences. Further analyses of clinical blood parameters such as TG, HBA1C, CHO, HDL, LDL, CRP, and anthropometric parameters (i.e., BMI, waist, hip, WHR, fat %, fat mass, lean mass) did not reveal a significant difference.

Anthropometric parameters	Clinical Blood Parameters
BMI	TG
Waist circumference	HBA1C
HIP	CHO
WHR	HDL
Fat %	LDL
Fat mass	CRP
Lean mass	

Table 11: Investigated Anthropometric and Clinical parameters for the metabolomics metadata and quantitative metabotype

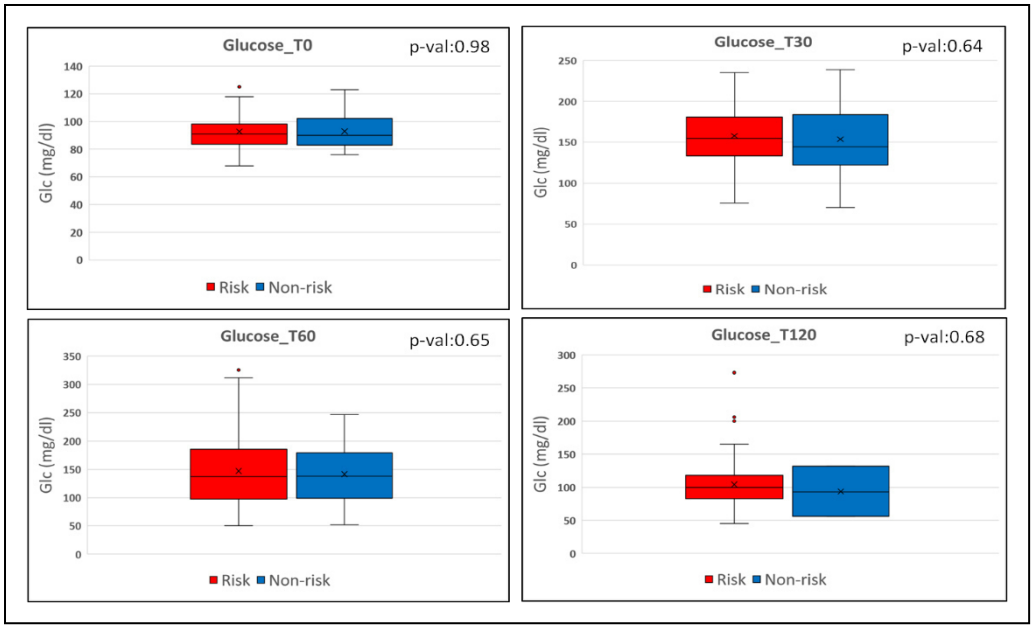


Figure 13: Quantitative data of glucose (mg/dl) across the different time points of the OGTT (0, 30, 60, 120mins). We analyzed glucose across different time **points of the OGTT between homozygous carriers and non-carriers of rs 6712203. 127 female subjects (91 *COBLL1* rs 6712203 homozygous risk allele carriers/36 homozygous non-risk allele carriers.** The measurements were performed fully automatically using a clinical-chemical enzyme immunoassay analyzer.

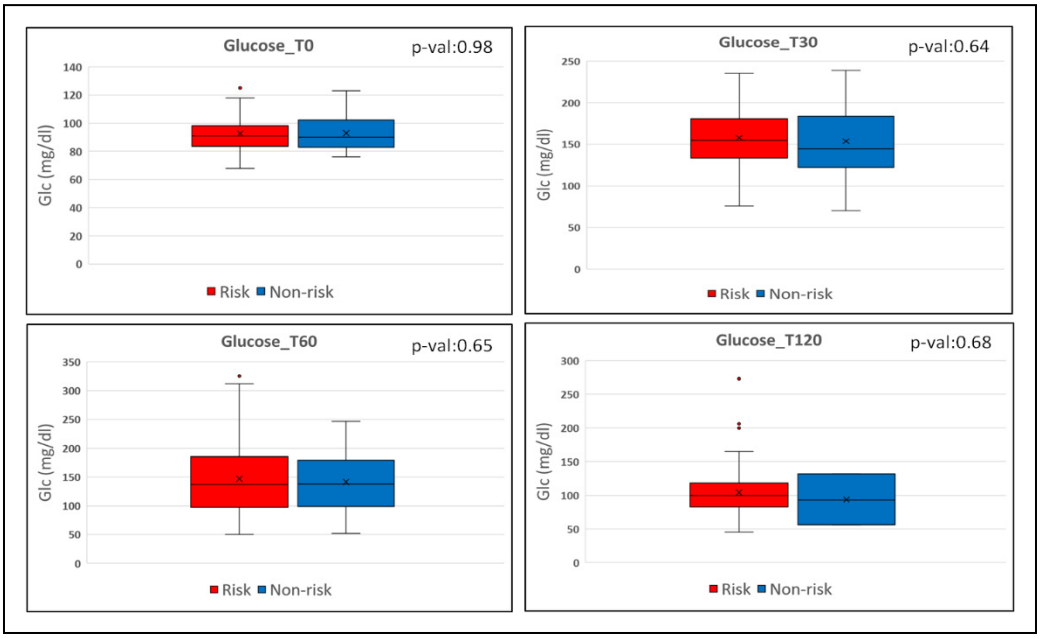


Figure 14: Quantitative data of NEFA (mg/dl) across different time points of OGTT (0, 30, 60, 120mins). We analyzed NEFA across different time **points of the OGTT between homozygous carriers and non-carriers of rs 6712203. 127 female subjects (91 *COBLL1* rs 6712203 homozygous risk allele carriers/36 homozygous non-risk allele carriers** NEFA determination has been standardized according to the WHO international reference material (WHO standard 66/304).

4.4. Over-representation analysis of compound classes (database driven)

The metabolic features are scored based on a statistical model and projected, if possible, on the corresponding which has known biochemical pathways by mapping them onto a database such as HMDB (Moritz et al., 2019). Secondarily, a small subset is taken from an entire set of features according to their significance. Afterward, the probability is computed, indicating whether a specific pathway is significantly enriched by assessing whether corresponding metabolites are over-represented in the considered subset (Moritz et al., 2019).

In our work, we describe Effect 1, which we consider as the effect between time point 0 (basal) and time point 60min, while Effect 2 is considered as the impact between time point 60min and time point 120min.

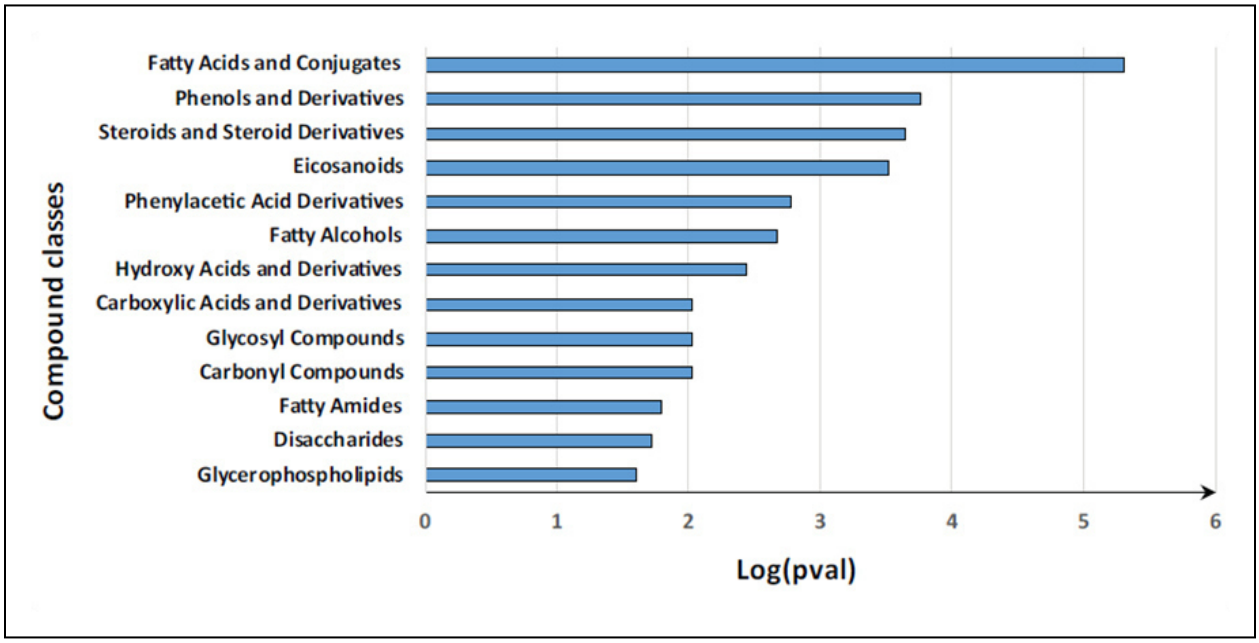


Figure 15: Results of over-representation analysis of HMDB compound classes annotated against HMBD. The figure depicts the chemical compound classes over-represented in a specific subset in **Effect 1** via DI-ICR-FT MS in risk allele female carriers (homozygous). **127 female subjects (91 *COBLL1* rs 6712203 homozygous risk allele carriers/36 homozygous non-risk allele carriers.** The experimental masses were assigned to the known metabolites from the HMDB database. The enrichment analysis was applied in order to see which classes of compounds are prevalent in the subset of features chosen from a PLS model. A high score means that the corresponding metabolic class is over-represented in a subset. No significant scores were observed for the up and down-regulation.

(Figure 15) depicts the compound classes over-represented in risk allele carriers at Effect 1, which is conceived as the difference between time 1h and baseline of the OGTT. **Fatty acids and conjugates** (log p-value 5.29), **phenols and derivatives** (log p-value 3.76), **steroids and steroid derivatives** (log p-value 3.65), and **eicosanoids** (log p-value 3.51) are the most over-represented classes, showing significant p-values. Three additional classes of interest are glycosyl compounds (log p-value 2.01), disaccharides (log p-value 1.71), and glycerophospholipids (log p-value 1.60), which could also be involved in the mechanisms of glucose uptake and GLUT4.

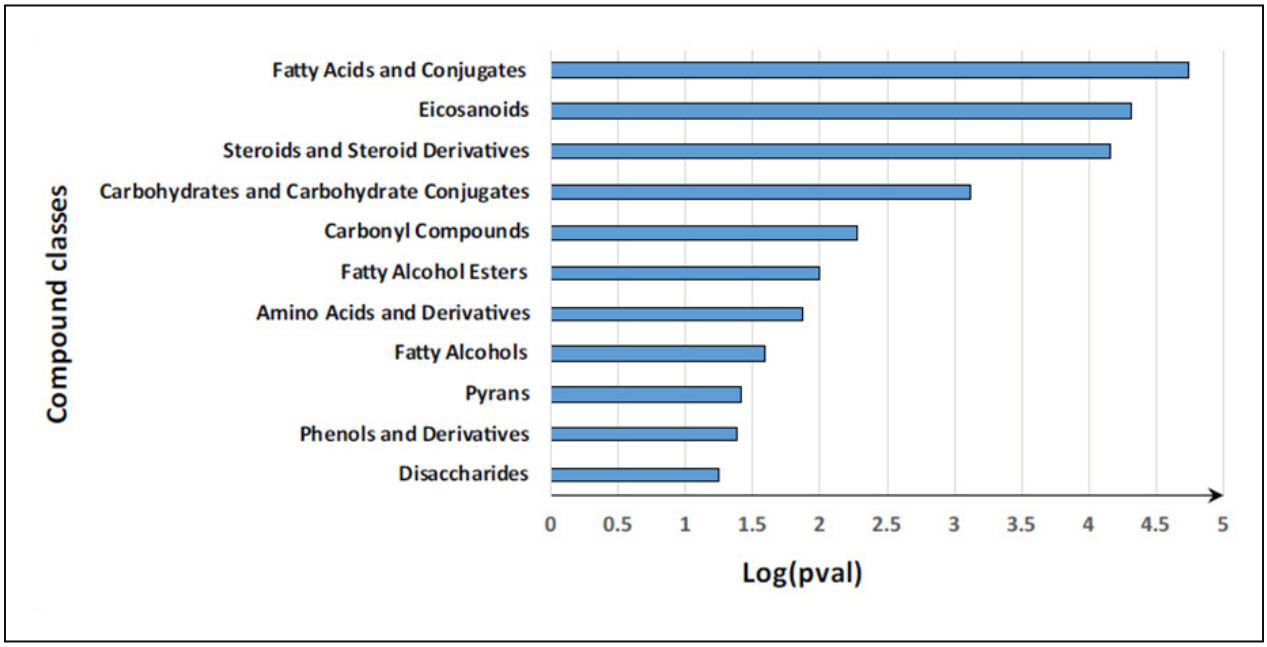


Figure 16: Results of over-representation analysis of HMDB Compound classes annotated against HMBD. The figure depicts the over-represented chemical compound classes in a specific subset in **Effect 2** via DI-ICR-FT MS in risk allele female carriers (heterozygous). **127 female subjects (91 *COBLL1* rs 6712203 homozygous risk allele carriers/36 homozygous non-risk allele carriers.** The experimental masses were assigned to the known metabolites from the HMDB database. The enrichment analysis was applied in order to see which classes of compounds are prevalent in the subset of features chosen from a PLS model. A high score means that the corresponding metabolic class is over-represented in a subset. No significant scores were observed for the up and down-regulation.

In (figure 16), there is a description of the compounds regulated risk allele carrier subjects in effect2: the difference between time 2h - 1h that means OGTT at time point 120 to time point 60. As shown, **fatty acids and conjugates, eicosanoids, steroids, and steroid derivatives occupy**

the first place. Again, similar to Effect 1, in Effect 2, the prominent class is **fatty acids and conjugates** with a significant value (log p-value 4.75). Besides that, there are some classes of interest, such as carbohydrates and carbohydrate conjugates (log p-value 3.11) and disaccharides (log p-value 1.25), which could be related to the mechanisms of our attention.

The following tables (table 12 and table 13) translate the figures 15 and 16 precisely to the compound classes with their p-value.

Effect1	
Compound classes	**Log(p-val)**
Glycerophospholipids	1,601855
Disaccharides	1,715716
Fatty Amides	1,79213
Carbonyl Compounds	2,015617
Glycosyl Compounds	2,015617
Carboxylic Acids and Derivatives	2,022025
Hydroxy Acids and Derivatives	2,446832
Fatty Alcohols	2,676833

Phenylacetic Acid Derivatives	2,778556
Eicosanoids	3,51683
Steroids and Steroid Derivatives	3,653774
Phenols and Derivatives	3,761765
Fatty Acids and Conjugates	5,299266

Table 12: Results of over-representation analysis of HMDB Compound classes annotated against HMBD. The figure depicts the chemical compound classes that are over-represented in a specific subset in **Effect 1** via DI-FT-ICR-MS in risk allele female carriers (Homozaygos)**.** The experimental masses were assigned to the known metabolites from the HMDB database. The enrichment analysis was applied in order to see which classes of compounds are prevalent in the subset of features chosen from a PLS model

Effect 2	
Compound classes	**Log(p-val)**
Fatty Acids and Conjugates	4,750694583
Eicosanoids	4,323942728
Steroids and Steroid Derivatives	4,160064341
Carbohydrates and Carbohydrate Conjugates	3,111890001

Carbonyl Compounds	2,270893919
Fatty Alcohol Esters	1,997124126
Amino Acids and Derivatives	1,876066987
Fatty Alcohols	1,586136293
Pyrans	1,412965168
Phenols and Derivatives	1,383591659
Disaccharides	1,252576136

Table 13: Result of over-representation analysis of HMDB Compound classes annotated against HMBD. The figure depicts the chemical compound classes that are over-represented in a specific subset in **Effect 2** via DI-FT-ICR-MS in risk allele female carrier (heterozygous). The experimental masses were assigned to the known metabolites from HMDB database. The experimental masses were assigned to the known metabolites from HMDB database. The enrichment analysis was applied in order to see which classes of compounds are prevalent in the subset of features chosen from a PLS model.

4.5. MDEA Result:

Mass-difference enrichment analysis (MDEA) was implemented to infer molecular pathways based on statistics (Moritz et al., 2017). One of the essential steps in MDEA is selecting the theoretical mass-differences that allow for a targeted investigation of the imbalance caused by *COBLL1* in circulating blood. Therefore, we created a list of co-regulated genes of *COBLL1* in order to map the mass-differences related to the considered mechanism (Moritz et al., 2017).

From the final co-regulated genes list of *COBLL1* (table 8), we could detect the MDs related to three genes: ***PPARG, PLIN5, and ILK*.** Following is a description of their function. Next, we created a list of candidate MDs based on the selected genes and merged them to a manually created list of theoretical MDs. Finally, the comprehensive list of MDs was searched in our experimental mass-difference space.

The ***PPARG*** gene (Peroxisome proliferator-activated receptor gamma) encodes for a nuclear receptor with molecular functions such as activating transcription factor binding, DNA, fatty acid, and lipid-binding (Tyagi et al., 2011). Moreover, it played an essential role in cell differentiation. It was characterized as the master regulator of adipogenesis, fatty acid metabolism, fatty acid oxidation, glucose homeostasis, lipid metabolism in a broader sense, and white fat cell differentiation (Tyagi et al., 2011).

The ***PLIN5*** gene (Perilipin-5) is a lipid droplet-associated protein, maintaining the balance between lipogenesis and lipolysis and regulating fatty acid oxidation in oxidative tissue (fatty acids released from mitochondrial fatty acids) (Kimmel et al., 2014). Moreover, this gene induces mitochondria translocation to the lipid droplets surface and is involved in lipid droplet homeostasis by fatty acid regulation in the form of triglycerides (Kimmel et al., 2014).

The ***ILK*** gene (Integrin-linked protein kinase) is an enzyme that regulates integrin-mediated signal transduction (Sakai et al., 2003). Thus, it plays a role in cell-matrix adhesion, cell population proliferation, integrin-mediated signaling pathway, and regulation of actin cytoskeleton organization (Sakai et al., 2003)

The main results of MDEA are represented in (figures 17-20). (Figures 17 and 18) describe the over-represented mass-differences (reaction products, up and down-regulated) in Effect 1.

In contrast, (figures 19 and 20) depict the over-represented mass-differences (reaction products, up and down-regulated) in Effect 2.

In Effect 1, the over-represented mass-differences associated with the up-regulation of reaction products in the risk allele carriers encompass tricarboxylic acid cycle (TCA)-metabolites and carboxylic acids (Figure 17). (Figure 18) describes the over-represented mass-differences associated with the down-regulation of reaction products in risk allele carriers, including carboxylic acid and sugars

In Effect 2, amino acids (AA) were among the over-represented mass-differences associated with the up-regulation of reaction products in the risk allele carriers (Figure 19). (Figure 20) describes the over-represented mass-differences related to the down-regulation of reaction products in risk allele carriers, including carboxylic acids and tryptophan.

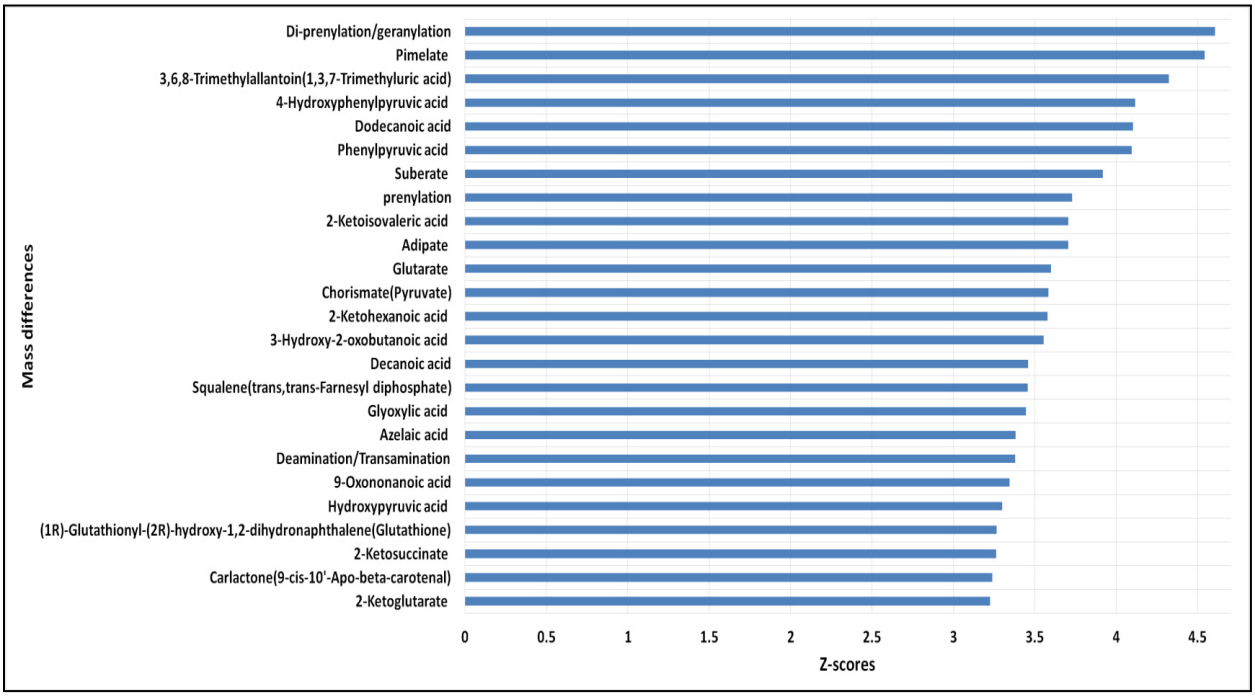


Figure 17: Description of the over-represented mass-differences associated with the <u>UP- regulation</u> of reaction products in the risk allele carriers during Effect 1 [T1h-T0]. TCA-metabolites and carboxylic acid are among the prominent mass-differences. The gene involved in the detected MDs is ILK, which encodes an integrin-linked protein kinase (EC.2.7.11.1) (59 kDa serine/threonine-protein kinase).

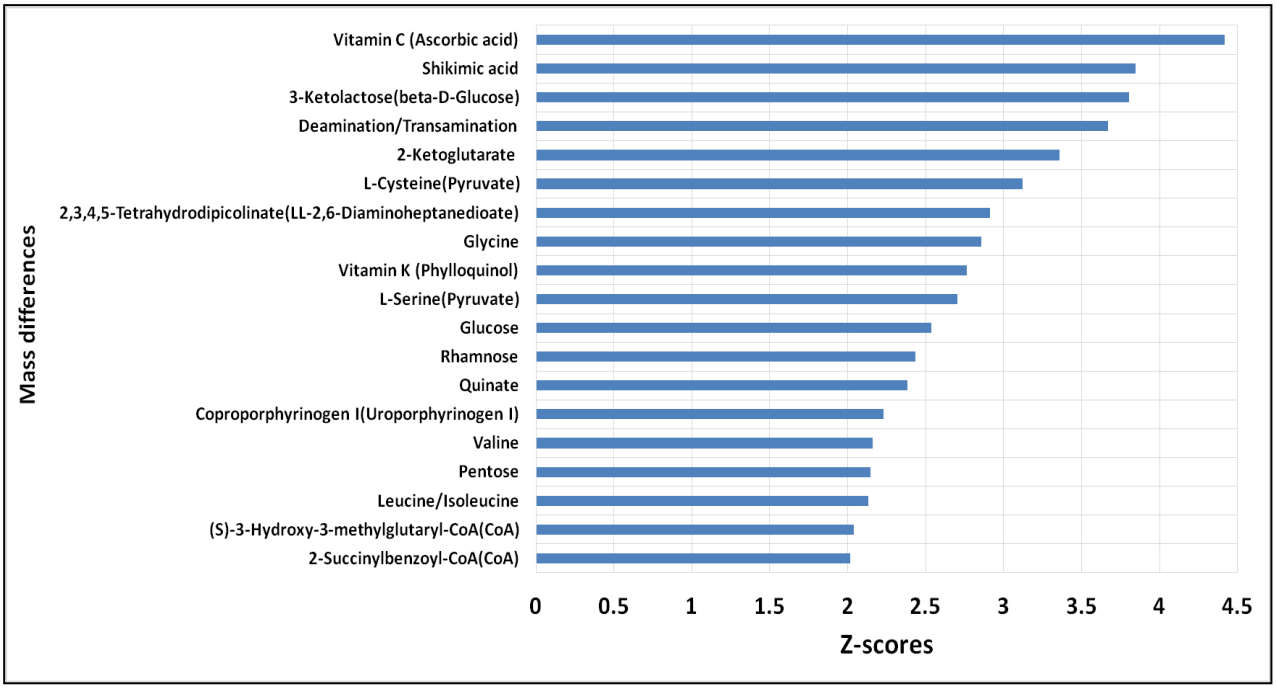


Figure 18: Description of the over-represented mass-differences associated with the <u>DOWN- regulation</u> of reaction products in the risk allele carriers during Effect 1 [T1h-T0]. Carboxylic acids and sugar are included in the prominent mass-differences. The genes involved in the detected MDs are *PPARY*, which encodes peroxisome proliferator-activated receptor gamma, and *PLIN5,* which encodes Perilipin-5 (Lipid storage droplet protein 5).

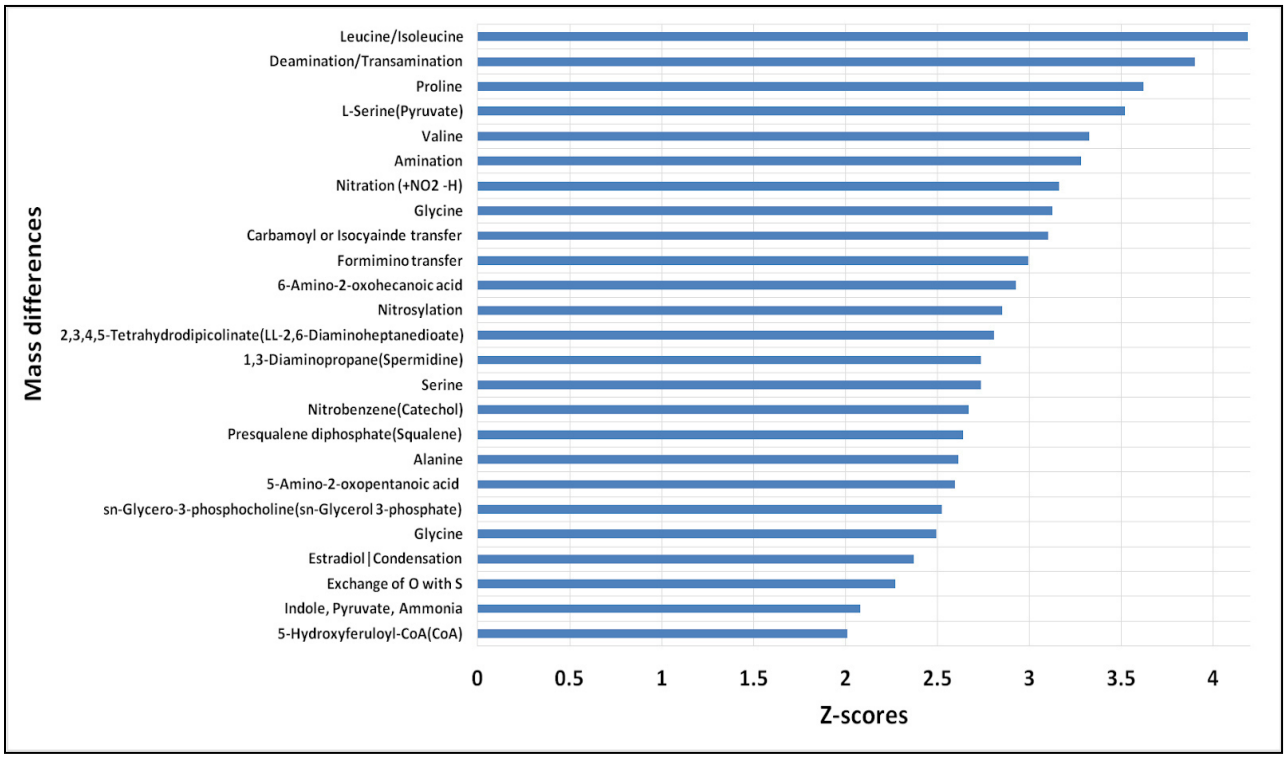


Figure 19: Description of the over-represented mass-differences associated with the <u>UP- regulation</u> of reaction products in the risk allele carriers during Effect 2 [T2h-T1h]. Amino acids (AA) are among the prominent mass-differences. The genes involved in the detected MDs are PPAR-y, which encodes peroxisome proliferator-activated receptor gamma, and PLIN5, which encodes Perilipin-5 (Lipid storage droplet protein 5).

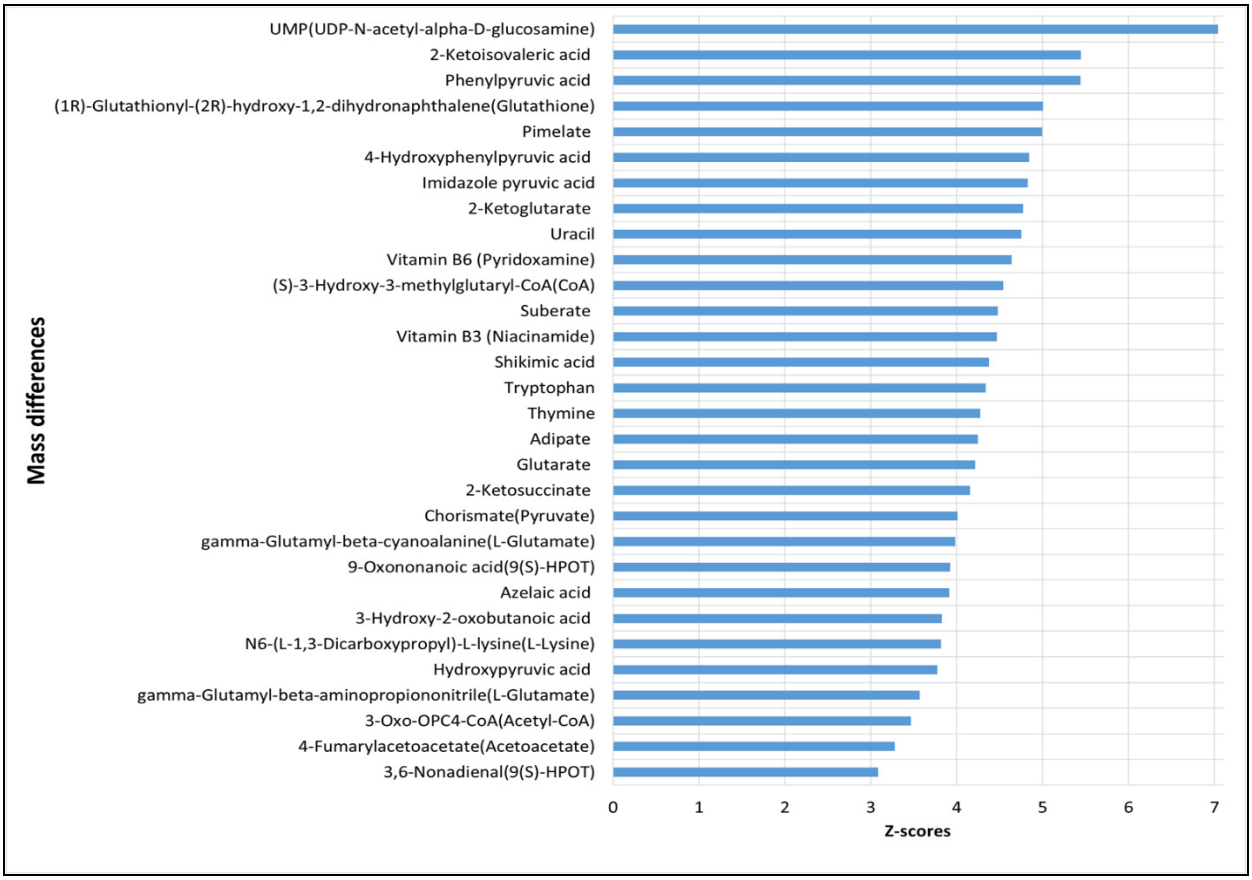


Figure 20: Description of the over-represented mass-differences associated with the DOWN- regulation of reaction products in the risk allele carriers during Effect 2 [T2h-T1h]. Carboxylic acids and tryptophan are included in the prominent mass-differences. The gene involved in the detected MDs is *PPARY*, which encodes peroxisome proliferator-activated receptor gamma

The following tables are the Description of the over-represented mass-differences associated with the UP and DOWN regulation of products in the risk allele carrier. Effect 1 (T1h-T0) and Effect 2 (T2h-T1h) shown in figure (17-20)

Effect 1_target uP				
Mass difference	**Z-scores**	**MD_Ids**	**Mass**	**Counts**
Di-prenylation/geranylation	4,606989	236	136,1252	2701

Pimelate	4,544503	246	142,063	3073
3,6,8-Trimethylallantoin(1,3,7-Trimethyluric acid)	4,322368	10	9,98435	5462
4-Hydroxyphenylpyruvic acid	4,11564	199	118,0419	2404
Dodecanoic acid	4,104119	241	138,1772	2311
Phenylpyruvic acid	4,096139	159	102,047	2480
Suberate	3,916314	267	156,0786	2909
prenylation	3,72961	84	68,0626	4972
2-Ketoisovaleric acid	3,706212	149	98,03678	4062
Adipate	3,704718	119	84,05752	4852
Glutarate	3,59929	184	114,0317	3411
Chorismate(Pyruvate)	3,583745	239	138,0317	2422
2-Ketohexanoic acid	3,578353	176	112,0524	3871
3-Hydroxy-2-oxobutanoic acid	3,554974	69	56,02622	5521
Decanoic acid	3,459387	174	110,1459	3096
Squalene(trans,trans-Farnesyl diphosphate)	3,456385	37	28,26022	48

Glyoxylic acid	3,444691	11	12	6396
Azelaic acid	3,382911	291	170,0943	2640
Deamination/Transamination	3,380043	1	0,984016	3091
9-Oxononanoic acid	3,345443	240	138,1045	3125
Hydroxypyruvic acid	3,30012	124	86,0004	3568
(1R)-Glutathionyl-(2R)-hydroxy-1,2-dihydronaphthalene(Glutathione)	3,265144	251	144,0575	1635
2-Ketosuccinate	3,262093	88	70,00548	4349
Carlactone(9-cis-10'-Apo-beta-carotenal)	3,237184	102	74,08842	2549
2-Ketoglutarate	3,225248	117	84,02113	4257

Table 14: Description of the over-represented mass-differences associated with the UP- regulation of products in the risk allele carrier. Effect 1 (T1h-T0).

Effect 1_target DOWN				
Mass differences	**Z-scores**	**MD_Ids**	**Mass**	**Counts**
Vitamin C (Ascorbic acid)	4,417027871	270	158,0215	1806
Shikimic acid	3,846658933	266	156,0423	2220

3-Ketolactose(beta-D-Glucose)	3,803452786	275	160,0372	2108
Deamination/Transamination	3,668391347	2	1,031634	2907
2-Ketoglutarate	3,35944526	212	128,011	2355
L-Cysteine(Pyruvate)	3,121839503	45	33,00371	261
2,3,4,5-Tetrahydrodipicolinate(LL-2,6-Diaminoheptanedioate)	2,911571781	27	19,0422	2488
Glycine	2,855875388	13	13,03163	3008
Vitamin K (Phylloquinol)	2,764929372	1625	434,3549	76
L-Serine(Pyruvate)	2,70488178	24	17,02655	3022
Glucose	2,537078622	280	162,0528	2397
Rhamnose	2,433097217	253	146,0579	3000
Quinate	2,383753188	296	174,0528	1981
Coproporphyrinogen I(Uroporphyrinogen I)	2,22785557	299	175,9593	522
Valine	2,158749118	67	55,07858	2564
Pentose	2,14464514	229	132,0423	3091
Leucine/Isoleucine	2,129811559	87	69,09423	2310

(S)-3-Hydroxy-3-methylglutaryl-CoA(CoA)	2,03795528	250	144,0423	2700
2-Succinylbenzoyl-CoA(CoA)	2,017015637	331	204,0423	661

Table (15): Represents the description of the over-represented mass-differences associated with the down-regulation of products in risk allele carriers. Effect 1 (T1h-T0).

Effect 1_target uP				
Mass differences	**Z-scores**	**MD_Ids**	**Mass**	**Counts**
Leucine/Isoleucine	4,187767713	87	69,094234	2310
Deamination/Transamination	3,898155298	2	1,031634	2907
Proline	3,621674329	62	53,062934	2664
L-Serine(Pyruvate)	3,5178133	24	17,026549	3022
Valine	3,326987255	67	55,078584	2564
Amination	3,279898167	16	15,010899	3105
Nitration (+NO2 -H)	3,161642912	57	44,985079	2123
Glycine	3,124466142	75	59,037114	2634
Carbamoyl or Isocyainde transfer	3,102728974	53	43,005814	2639

Formimino transfer	2,99515503	33	27,010899	2981
6-Amino-2-oxohecanoic acid	2,9267196	116	83,073499	2461
Nitrosylation	2,851746985	38	28,990164	2673
2,3,4,5-Tetrahydrodipicolinate(LL-2,6-Diaminoheptanedioate)	2,806380076	27	19,042199	2488
1,3-Diaminopropane(Spermidine)	2,73598746	93	71,073499	2602
Serine	2,735738315	54	43,042199	2947
Nitrobenzene(Catechol)	2,670152378	12	12,995249	3013
Presqualene diphosphate(Squalene)	2,641291518	298	175,927579	21
Alanine	2,613606345	34	27,047284	2903
5-Amino-2-oxopentanoic acid	2,596671984	130	87,068414	2339
sn-Glycero-3-phosphocholine(sn-Glycerol 3-phosphate)	2,522627312	123	85,089149	2380
Glycine	2,493234613	13	13,031634	3008
Estradiol|Condensation	2,371921125	409	254,167065	272
Exchange of O with S	2,269579702	18	15,977156	449
Indole, Pyruvate, Ammonia	2,080391984	98	73,016379	2079

5-Hydroxyferuloyl-CoA(CoA)	2,009296226	319	192,04226	986

Table 16: Represents the description of the over-represented mass-differences associated with the UP- regulation of products in the Risk allele carrier. Effect 2 (T2h-T1h).

Effect 2_target DOWN				
Mass differences	**Z-scores**	**MD_Ids**	**Mass**	**Counts**
UMP(UDP-N-acetyl-alpha-D-glucosamine)	7,040271	502	283,0457	33
2-Ketoisovaleric acid	5,44312	149	98,03678	4062
Phenylpyruvic acid	5,440976	205	120,0575	2869
(1R)-Glutathionyl-(2R)-hydroxy-1,2-dihydronaphthalene(Glutathione)	5,005099	251	144,0575	1635
Pimelate	4,996164	246	142,063	3073
4-Hydroxyphenylpyruvic acid	4,844738	199	118,0419	2404
Imidazole pyruvic acid	4,82868	139	92,03745	994
2-Ketoglutarate	4,772561	212	128,011	2355
Uracil	4,752991	142	94,01671	908
Vitamin B6 (Pyridoxamine)	4,638872	259	150,0793	607

(S)-3-Hydroxy-3-methylglutaryl-CoA(CoA)	4,544363	250	144,0423	2700
Suberate	4,479299	267	156,0786	2909
Vitamin B3 (Niacinamide)	4,468829	165	104,0374	882
Shikimic acid	4,375866	266	156,0423	2220
Tryptophan	4,338377	309	186,0793	303
Thymine	4,272168	171	108,0324	865
Adipate	4,248299	213	128,0473	3270
Glutarate	4,216233	184	114,0317	3411
2-Ketosuccinate	4,154295	88	70,00548	4349
Chorismate(Pyruvate)	4,009598	239	138,0317	2422
gamma-Glutamyl-beta-cyanoalanine(L-Glutamate)	3,984995	145	96,03236	956
9-Oxononanoic acid(9(S)-HPOT)	3,926221	240	138,1045	3125
Azelaic acid	3,911704	291	170,0943	2640
3-Hydroxy-2-oxobutanoic acid	3,828011	154	100,016	3554
N6-(L-1,3-Dicarboxypropyl)-L-lysine(L-Lysine)	3,81484	221	130,0266	2772

Hydroxypyruvic acid	3,773764	124	86,0004	3568
gamma-Glutamyl-beta-aminopropiononitrile(L-Glutamate)	3,569511	61	52,04253	1213
3-Oxo-OPC4-CoA(Acetyl-CoA)	3,466655	320	192,115	1676
4-Fumarylacetoacetate(Acetoacetate)	3,280408	148	98,0004	3054
3,6-Nonadienal(9(S)-HPOT)	3,083496	294	172,1099	2826

Table 17: Represents the description of the over-represented mass-differences associated with the down-regulation of products in risk allele carriers. Effect 2 (T2h-T1h).

5. Discussion:

Modifying lifestyle provides an opportunity to reverse the diabetes trend, and it's the concept procedure for type 2 diabetes. Risk factors such as diet, adiposity, physical activity, and environmental expositions are modifiable by applying a combination of approaches at the population and individual level (Darnton-Hill et al., 2004). Many prevention programs have focused on lifestyle modification, addressing modifiable risk factors. Those studies have shown that lifestyle intervention is effective in preventing Diabetes. Different prospective randomized studies (Tuomilehto J et al., 2001, Knowler WC et al., 2002, Ramachandran A et al., 2006, Li G et al., 2008) have indicated that diabetes risk can be reduced by modifying diet and physical activity.

Despite that, there is an imperative need for making the lifestyle intervention more effective for diabetes prevention; the reason is that a significant proportion of participants do not benefit from the intervention in lifestyle intervention trials. These participants are usually referred to as non-responders (Perreault L et al., 2009, Schmid V et al., 2017). For instance, in the Diabetes Prevention Program (DPP), every fifth patient of the Lifestyle intervention developed type 2 diabetes within 4 years (Knowler WC et al., 2002). On the other hand, individuals with pre-diabetes do not progress to diabetes during 11 years follow up even without intervention. Therefore there's an important question whether Lifestyle intervention is essential in all individuals with pre-diabetes.

These observations of non-response to lifestyle intervention and non-progression to diabetes focus attention on the need for risk classify intervention strategies in individuals with pre-diabetes. Next, an essential question is which phenotype determines the risk for diabetes, especially the response and non-response to lifestyle intervention. A recent analysis of the DPP showed that response varies based on diabetes risk, suggesting that the lifestyle intervention should be adapted based on individual risk.

The Tübingen Lifestyle Intervention Program (TULIP) (PLIS pilot study) identified a phenotype of high risk (HR) associated with a higher probability of short-term and long-term non-response to Lifestyle intervention. Furthermore, TULIP showed that risk stratification can identify severe disease courses and increased risk for diabetes-related complications in populations before diabetes onset and with diabetes. Accordingly, it's definitive to improve the efficiency and effectiveness of lifestyle intervention programs in high-risk subjects to conquer no response to preventive interventions.

Therefore, the PLIS study was designed to answer 2 important questions: can intensive lifestyle intervention overcome non-response in high risk individuals with pre-diabetes? And is lifestyle intervention effective in low-risk individuals with pre-diabetes? The details of the PLIS study participants, randomization, intervention, etc., is mentioned in the methods part of the thesis.

There was no significant difference in the low-risk group after lifestyle intervention after 2 years concerning the following parameters: glucose, insulin sensitivity, hepatic fat, cardiovascular risk. Meanwhile, there were significant differences between high-risk conventional and high-risk intervention for the same mentioned parameters for the high-risk group.

The result after 3 years of observation (1 year of lifestyle intervention and additional 2 years follow up), there were significant differences for the lower risk group (control to conventional) and high-risk group (Conventional to intensive).

The data from the PLIS study indicate that conventional lifestyle intervention as were applied DPP and DPS can be successfully intensified. However, the implementation of intensified lifestyle intervention was effective on BMI, insulin sensitivity, and liver fat content were more pronounced. In contrast, intensified lifestyle intervention did not improve insulin secretion capacity compared to conventional lifestyle intervention. For that reason, the effect of intensive lifestyle intervention on post-challenge glucose was most probably due to reduced fat liver

content and improved insulin sensitivity. The improvement of insulin sensitivity in successful lifestyle intervention is consistent with the DPP and DPS trials' findings.

Interestingly, the exercise volume and the number of counseling sessions differed between the intensified and conventional interventions in PLIS. However, the weight reduction goal's accomplishment was significantly associated with reducing 2 hours post glucose challenge during 1 year of intervention in all treatment groups. This suggests that the number of counseling sessions was either motivational or guidance from the lifestyle advisor is underlying the higher efficacy of the intensive intervention group. Therefore, qualified lifestyle counselors and sufficient counseling frequency should be the central factors in lifestyle intervention planning.

In the PLIS cohort, the beneficial effect of intensive lifestyle intervention goes beyond glucose control and affects liver fat content. After the intensive intervention, the liver fat content was close to the normal threshold of 5.6%, suggesting a clinically relevant effect as a target for future approaches in diabetes prevention.

Furthermore, the cardiovascular risk is decreased in the participants for the high-risk group with nearly doubling of the risk reduction for the high-risk intervention group compared with the conventional group.

The limitation of the study includes a short Lifestyle intervention duration (12 months). One more possible limitation is the heterogeneity of lifestyle counseling throughout different study centers. Moreover, the study did not include an intensified intervention in the Low-risk group, and there was no control group without intervention for high-risk. The last limitation is that the high-risk and low-risk groups were unbalanced.

Future studies are needed to investigate the question in Lifestyle individuals; nevertheless, screening and treatment approaches in preventing type 2 diabetes should include risk stratification and individualized interventions.

Our sub study identified metabolite patterns in plasma that reflect the metabolic dysregulation caused by genetic variation, specifically the *COBLL1* gene, leading to an increased risk of T2D. This work aimed to connect variants at the *GRB14/COBLL1* metabolic risk locus and the effectors' gene *COBLL1* with metabolic changes in plasma, applying metabolic profiling to further the mechanistic understanding of the risk locus on disease-relevant traits and to pave the way for better disease prediction through biomarkers. I sought to investigate the metabolic profiles of the subjects genotyped for the *GRB14/COBLL1* risk haplotype by the rs6712203 variant, which our group has previously shown to mediate part of the metabolic risk by affecting the actin cytoskeleton remodeling in (pre-)adipocytes. In my work, I included subjects from the PLIS cohort genotyped for the SNP rs6712203 (91 risk allele carriers vs. 36 non-risk allele carriers) who underwent nutritional challenges to unveil their metabolic profile under specific stimuli. In order to accomplish this goal, the untargeted human blood metabolites were investigated using direct infusion ultrahigh-resolution mass-spectrometry DI-ICR-FT-MS. This high-resolution and high accuracy approach focused on screening a maximum number of metabolites, enabling top confidence feature assignment to molecular formulas. This will provide promising targets for future disease prediction and prevention strategies.

For the over-representation analysis of compound classes (database driven), the compound classes over-represented in risk allele carriers at Effect 1, which is conceived as the difference between time 1h and baseline of the OGTT. **Fatty acids and conjugates** (log p-value 5.29), **phenols and derivatives** (log p-value 3.76), **steroids and steroid derivatives** (log p-value 3.65), and **eicosanoids** (log p-value 3.51) were the most over-represented classes, showing significant p-values. Three additional classes of interest were glycosyl compounds (log p-value 2.01), disaccharides (log p-value 1.71), and glycerophospholipids (log p-value 1.60), which could also be involved in the mechanisms of glucose uptake and GLUT (Figure 4).

In Effect 2, the difference between time 2h - 1h means OGTT at time point 120 to time point 60. **Fatty acids and conjugates, eicosanoids, steroids, and steroid derivatives occupied the first place**. Again, similar to Effect 1, in Effect 2, the prominent class is **fatty acids and conjugates** with a significant value (log p-value 4.75). Besides that, there were some classes of

interest, such as carbohydrates and carbohydrate conjugates (log p-value 3.11) and disaccharides (log p-value 1.25), which could be related to the mechanisms of interest (Figure 16).

One of the interesting compound classes is Steroids and steroids derivatives. The transportation of the steroids happens through the bloodstream to the cells of different organs. There they act as regulators of an extensive range of physiological functions. When steroid hormones pass across the target cell membrane, they may cause fundamental physiological changes. After they go through the membrane, they bind to particular receptors in the cytoplasm; afterward, this complex receptor demands the production of mRNA molecules, which code numerous proteins (Pandey et al., 2009).

As mentioned previously, actin is involved in the intracellular transport of organelles, vesicles, and muscular contractions (Bugyi et al., 2020). Thus, people, who carry a mutation in the *COBLL1* gene, show disturbances in actin morphology that lead to an essential change in embryogenesis, as reported in the work of Claussnitzer group. That could explain why steroid hormones compound classes are significantly shown in our results (figure 16 and 17). Moreover, the association between *COBLL1* with WHR adj BMI is five times clearer in women than in men, showing a significant sexual dimorphism effect (Heid et al., 2010). These differences become apparent during puberty and are generally attributed to the influence of sex hormones (Heid et al., 2010).

Glunk et al. identified a sexual dimorphism in their experiments, indicating an effect of rs6712203 on stimulated lipolysis and potentially fat distribution only in women. It is intriguing to speculate that a decreased adipogenic capacity due to *COBLL1* perturbation in subcutaneous adipocytes may decrease the adipocyte insulin response and safely store excess energy in women. Still, more work will be required to investigate the effects of *COBLL1* on female fat distribution. In addition, insight may be gained from comparing visceral and subcutaneous fat depots. Taken together, we here provide some evidence for a female-specific effect of *COBLL1* on fat distribution, which may contribute to the risk for insulin resistance and T2D.

Another compound class is interested, fatty acids and conjugates. It is known that the variants at the *GRB14/COBLL1* locus have also been associated with HDL cholesterol[42,175] and triglyceride levels[42,176]. The variant rs6712203 has been associated with GC/MS metabolite measurement using the KORA and Twins UK cohorts with glycerol and n-Butyl Oleate release[177]. The results by our team substantiate the association with HDL cholesterol, but it remains to be investigated how the role of *COBLL1* in adipocytes is involved in these associations.

The actin cytoskeleton is involved in many cellular processes, besides providing structural support and vesicle trafficking. Coordinated regulation of the cytoskeleton is essential during adipogenesis (Kawaguchi et al., 2003, Yang et al., 2014; Kanzaki et al., 2001). GLUT4 translocation to the plasma membrane (Kanzaki et al., 2001), the maintenance of lipid droplets, and lipid droplet dynamics during lipolysis (Orlicky et al., 2013). Considering that the *COBLL1* homolog *COBL* is a known F-actin dynamizer (Husson et al., 2011), and GO-term pathways list *COBLL1* as an actin interacting protein (Ashburner et al., 2000), we investigated whether *COBLL1* may play a role in the regulation of the actin cytoskeleton in adipocytes, thereby affecting GLUT4 vesicle trafficking, lipid metabolism, and adipocyte differentiation.

In the studies of Glunk et al., a disrupted regulation of *COBLL1* was found in risk allele carriers in rs6712203-C, leading to disturbed glucose uptake, lipolysis, adipogenesis, and increased TG storage. Moreover, they demonstrated that all cell functions and oxidation processes happening in the cell are entirely disrupted (Glunk et al., submitted).

As shown in (Figure 4), Glunk et al. reported that all the oxidation processes are disrupted by the *COBLL1* mutation, provoking an accumulation of citrate and pyruvate. Consequently, less glucose is transported and taken up, and several other pathways such as lipolysis, adipogenesis, and fat storage are also impaired.

5.1. *COBLL1* and Glucose uptake

Insulin-stimulated *GLUT4* translocation in adipocytes is dependent on cortical actin remodeling (Kanzaki et al., 2001). Insulin receptor activation triggers a massive increase in the rate of *GLUT4* vesicle exocytosis, with a slight decrease in the frequency of internalization by endocytosis. Several studies demonstrated that:

1. Cell cytoskeleton has a notable influence over vesicle trafficking events.

2. Actin cytoskeleton influence regulated exocytosis.

3. A thick sheet of F-actin beneath and juxtaposed to the plasma membrane in most secretory cells, so-called cortical actin.

4. Actin function as a physical barrier to vesicle docking is based on its transient polymerization during exocytosis, and the secretion happens at the site where the actin cortex is thin (Kanzaki et al., 2001).

Both stabilization and disruption of adipocyte cortical actin inhibit insulin-stimulated *GLUT4* translocation and cortical actin remodeling. Insulin's dynamic actin rearrangement process is induced, which is necessary for insulin-stimulated *GLUT4* translocation. Insulin-stimulated membrane ruffling is not displayed by cortical actin but appears to undergo dynamic remodeling (polymerization/depolymerization). These results are consistent with F-actin analysis in primary rat adipocytes, displaying a predominantly cortical actin network (Norris et al., 2018).

In summary: differentiated adipocytes firstly express cortical actin, which must go through active insulin-stimulated remodeling, essential for translocation of insulin-stimulated *GLUT4* (Norris et al., 2018; Lowe et al., 2011). The cortical actin network is regulated by TC10, a member of the Rho family of small GTP-binding protein (Lowe et al., 2011). Besides, cortical actin plays an essential positive role in trafficking *GLUT4* vesicles (Lowe et al., 2011). Glunk et al. identified a decreased sensitivity of *COBLL1* perturbed cells to insulin, which leads to reduced glucose uptake, potentially caused by the effect of *COBLL1* on actin fermentation (Glunk et al., submitted).

5.2. *COBLL1* and adipogenesis

Adipogenesis is a physiological process that promotes the tissue's ability to isolate lipids and prevent lipotoxicity in peripheral organs safely (Karpe et al., 2011). Furthermore, the hormone insulin supports expansion, whereas glucagon, epinephrine, and ACTH promote mobilization (Karpe et al., 2011). Therefore, individuals at high-risk of developing diabetes, characterized by inappropriately enlarged adipose cells relative to their BMI (hypertrophic obesity), is due to reduced adipogenesis, insulin resistance, and adipose tissue inflammation. Besides, adipose *GLUT4* levels are reduced in adipocytes from individuals with T2D (Ghaban et al., 2019).

The results from Glunk et al. link *COBLL1* expression and adipogenesis with an effect on lipid accumulation. In rs6712203-C, risk allele carriers could translate into lower adipogenesis due to a lower ability to up-regulate *COBLL1* mRNA expression as long as mature adipocytes have an increased insulin-stimulated glucose uptake when compared to undifferentiated cells (Salans et al., 1968). Since the maturation of adipocytes and their size is essential for the response of the cells to insulin, they evaluated the relationship between *COBLL1* mRNA expression and adipocyte size by using fractions of isolated mature primary human adipocytes (Newsholme et al., 2010). In the large portion, they found some evidence for a positive correlation between cell size and *COBLL1* mRNA expression (p-value = 0.01). This finding may be explained by the increased capability of the cells to store lipids when *COBLL1* is not perturbed (Newsholme et al., 2010).

5.3. *COBLL1* lipolysis and TGs storage

Triglycerides are formed primarily from dietary fats: They are hydrolyzed to monoglycerides and FAs by lipases (Newsholme et al., 2010). The liver and adipose tissue are the primary organs for endogenous TGs synthesis. Under normal conditions, hepatic TGs are secreted as very-low-density lipoprotein (VLDL). Under specific pathological states, TGs accumulate in hepatocytes, leading to hepatic steatosis (Newsholme et al., 2010). When the hormone-sensitive lipase (HSL)

is activated by glucagon and adrenaline, fatty acids are mobilized from adipose tissue TGs (Harris et al., 2014).

Glunk et al. found evidence for higher *COBLL1* levels in mature and large adipocytes, confirming the previous observation of a diminished maturation capacity and lipid storage when *COBLL1* is reduced. In addition, the finding that leptin mRNA expression levels, an adipokine produced in proportion to the size of fat depots (Smas et al., 1995), correlate with *COBLL1* mRNA expression further supports the role of *COBLL1* in fat storage (Smas et al., 1995).

In summary, the actin cytoskeleton and the ability of cells to remodel actin filaments have been implicated in the storage of lipids (Orlicky et al., 2013; Greenberg et al., 2011), which suggests the role of *COBLL1* in actin remodeling could also be necessary for lipid storage in differentiated adipocytes (Luo et al., 2011; Greenberg et al., 2011),

In mature unilocular white adipocytes, energy is mainly stored in one lipid droplet (Karpe et al., 2011). Which occupies most of the adipocyte, controls the volume (Prasad et al., 2015), and leads to a rounded cell shape (Orlicky et al., 2013; Greenberg et al., 2011). Actin dynamics control the number and size of adipocytes (Prasad et al., 2015) and accommodate the appearance and growth of lipid droplets (Prasad et al., 2015). Body mass index (BMI), a measure for obesity, is a significant predictor for T2D. Still, some genetically predisposed individuals develop T2D with a lower BMI than others, highlighting the importance of genetic factors (Perry et al., 2012), influencing the ability to safely store excess energy (Henninger et al., 2014). A combined defective and overloaded “healthy” subcutaneous adipose tissue (Asterholm et al., 2014) expansion may lead to unfavorable lipid storage in visceral adipose tissue or non-adipocyte cells and may - accompanied by fibrosis - lead to insulin resistance (Ahmadian et al., 2007). Glunk et al. provide evidence for an increased expression of *COBLL1* in “healthy” mature adipocytes.

Moreover, a disturbing *COBLL1* expression or a lower ability to regulate *COBLL1* expression was associated with disturbances in adipocyte hypertrophy and potentially hyperplasia (Smas et al., 1995). The correlations between leptin and *COBLL1* supported these observations. Leptin is produced in proportion to the size of fat depots in the human body. Therefore, when *COBLL1* is perturbed or decreased, we found a lower amount of *leptin*, substantiating the observation that

COBLL1 might be necessary during adipogenesis and lipid storage, enabling a “healthy” expansion of human adipose tissue. Furthermore, the results of Glunk et al. show that the up-regulation of *COBLL1* during adipogenesis could be essential to remodel actin from stress fibers to cortical actin, influencing the adipogenic capacity (Smas et al., 1995).

Lipolysis is defined as a catabolic process leading to the breakdown of TAGs, stored in cellular lipid droplets, into FFAs and glycerol (Tansey et al., 2001). After the FFAs are released into the blood, they are transported and taken by other tissues for ß-oxidation and ATP generation. Not all FFAs leave the cell; some are re-esterified into TAGs intracellularly. Lipolysis usually takes place in the cytoplasm. After lipolysis, the released glycerol from TGs directly enters the glycolysis pathways as DHAP (Tansey et al., 2001).

In cells with *COBLL1* perturbation, lipids' storage and release may be disturbed. Glunk et al. found a decreased lipolysis rate and reduced perilipin levels when *COBLL1* was reduced. Interestingly, similar phenotypes were observed (Grant et al., 2009) in *COBLL1* perturbed human adipocytes. In conclusion, they found a decreased β-adrenergic lipolytic response in *COBLL1* perturbed adipocytes. In combination with a decreased adipogenic capacity, this may lead to the inability to safely store excess energy (Almadian et al., 2007).

The actin cytoskeleton and the ability of cells to remodel actin filaments have been implicated in the storage of lipids (Orlicky et al., 2013; Greenberg et al., 2011). Therefore, the role of *COBLL1* in actin remodeling could also be necessary for lipid storage in differentiated adipocytes. In mature white adipocytes, energy is mainly stored in one lipid droplet (Karpe et al., 2011), which occupies most of the adipocyte, controls the volume (Prasad et al., 2015), and leads to a rounded cell shape (Orlicky et al., 2013; Greenberg et al., 2011).

Based on the work of Glunk et al. and understanding the actin dynamics and distribution of pathways in a *COBLL1* mutation discussed above, we could, in our study cohort, focus on the most significant compound classes circulating in blood plasma which reflect the pathways perturbation.

Table 18 depicts the most significant classes in my over-represented analysis (ORA) work. The table compares compound classes from ORA and MDs. In Effect 1 (T1h-T0), glucose and FAs are the most significant compound classes, while in Effect 2 (T2h-T1h), glucose and AAs are the most significant compounds. We compared the over-represented mass-differences associated with the up and down-regulation of products in risk allele carriers. We found up-regulated MDs in Effect 1 (T1h-T0) carboxylic acid and TCA products, mainly pyruvate and ketoglutarate. In contrast, in Effect 2(T2h-T1h), amino acids and pyruvate were prominently up-regulated MDs. This is because pyruvate and ketoglutarate in Effect 1 and pyruvate in Effect 2 are intermediate products in the process of the TCA cycle.

For the down-regulated MDs In Effect 1 (T1h-T0h), glucose and the TCA product 2-ketoglutarate were down-regulated, while in Effect 2 (T2h-T1h), tryptophan and TCA products, pyruvate, glutamate, 2-keto-succinate, hydroxyl pyruvate acid, were the most prominently down-regulated mass-differences. 2-ketoglutarate in Effect 1 and pyruvate, glutarate, 2-keto-succinate, and hydroxyl pyruvate acid in Effect 2 were the most prominent MDs considered intermediate products in the process of the TCA cycle.

	Compounds	Mass differences UP regulated	Mass differences Down regulated
Effect 1 (T1h-T0)	Glucose FA's	↑ TCA (pyruvate, Ketoglutarate)	↓ Glucose ↓ TCA(2-Ketoglutrate)
Effect 2 (T2h-T1h)	Glucose AA's	↑ AA ↑ TCA(Pyruvate)	↓ Tryptophan ↓ TCA(pyruvate, Glutarate, 2-keto-succinate, Hydroxyl pyruvate acid)

Table 18: Comparison of compounds classes enrichment analysis outcomes and Mass- differences in Effect 1 and Effect 2. In Effect 1, Glucose and FA's, while in Effect 2, Glucose and AA's are the most prominent compound classes, respectively. In Effect 1, Up-regulated mass- differences were TCA (pyruvate, Ketoglutarate) and Carboxylic acid, while glucose and TCA (2-Ketoglutarate) were down-regulated mass-differences. In Effect 2, AA and TCA (pyruvate) were Up-regulated mass-differences, while Tryptophan and TCA (pyruvate, glutarate, 2-keto-succinate, and hydroxyl pyruvate acid) were down-regulated mass-differences.

Eventually, I was able to conclude my hypothesis based on the results and interpretation of the data as follows:

5.4. Glucose uptake hypothesis

The following figure represents the Scenarios of glucose uptake.

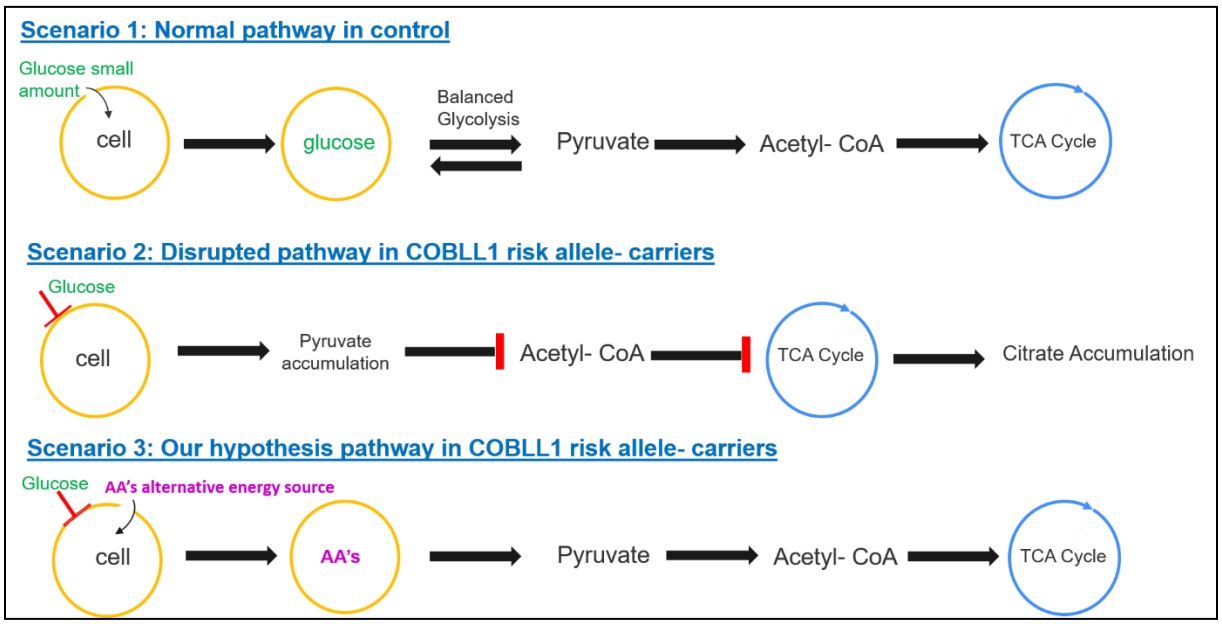


Figure 21: Hypothesis for glucose uptake and transport. Scenario 1: Normal pathway of glucose uptake. Scenario 2: Disrupted glucose uptake in *COBLL1* risk allele carriers. Scenario 3: Our hypothesis pathway in *COBLL1* risk allele- carriers

In (figure 21), we describe the glucose uptake Scenarios. In Scenario 1, where the normal pathway of glucose uptake is described, i.e., *COBLL1* non-risk allele carrier. Glucose uptake is fully functioning, and glucose after uptake is converted to pyruvate during glycolysis (Berg et al., 2002; Lehninger et al., 2008). Afterward, pyruvate is converted to acetyl-CoA and enters the TCA cycle to continue its normal pathway (Berg et al., 2002; Lehninger et al., 2008). Thus, scenario 1 explains the up-regulation of TCA products in Effect 1, where the products are consumed rapidly and immediately to complete the cycle. Additionally, the two processes, namely glycolysis, and glycogenesis, are balanced in the usual case.

Coming to Scenario 2 may explain the disruption of the glucose uptake pathway in *COBLL1* risk allele carriers. When the disturbance is happening, glucose uptake cannot be increased in the cells following insulin stimulation, as the dynamic regulations of the actin cytoskeleton is a prerequisite for the cellular translocation of *GLUT4* (Kanzaki et al., 2001). Therefore, a decreased regulation of *COBLL1* expression in rs6712203-C risk allele carriers may result in an impaired actin-mediated *GLUT4* vesicle trafficking to the plasma membrane. As a result, pyruvate is alternatively generated from lipolysis, which results in the accumulation of pyruvate, and acetyl CoA cannot be generated to enter the TCA cycle, which leads to citrate accumulation.

The body recognizes the lack of energy needed for metabolic processes and starts to generate the essential energy. Therefore, in Scenario 3, we hypothesized that the alternative energy source is AAs.

Amino acids were found to be up-regulated in Effect 2 (Table 18). Thus, AAs are used alternatively to glucose, which allows the generation of pyruvate, which may be converted to acetyl- CoA and able to enter the TCA cycle, which could complete its normal pathway to generate energy.

5.5. Lipolysis, fat storage, and adipogenesis hypothesis

The following figure represents the Scenarios of lipolysis, fat storage, and adipogenesis.

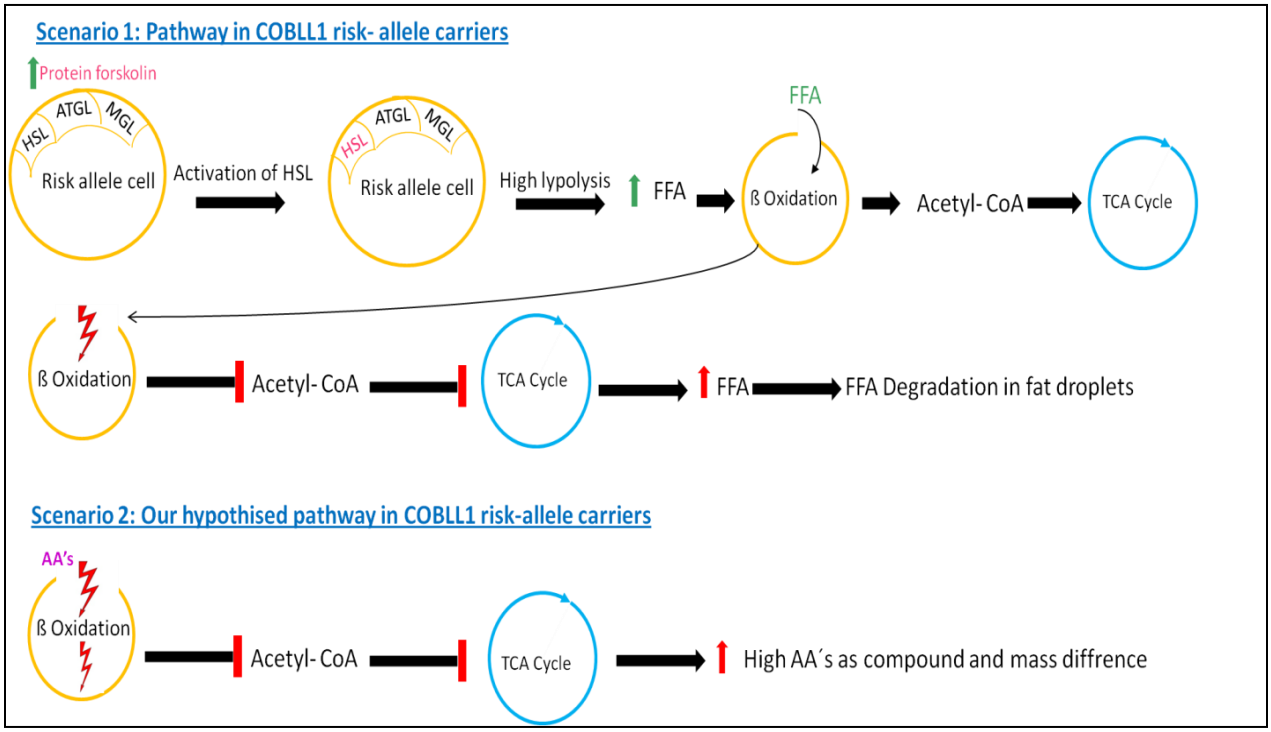


Figure 22: Hypothesis in lipolysis, fat storage, and adipogenesis. Scenario 1: pathway in COBLL1 risk allele carriers and disruption of ß-oxidation in *COBLL1* risk allele carriers. Scenario 3: Our hypothesized pathway in *COBLL1* risk allele carriers.

In female risk allele carriers, the protein forskolin is regulated—this protein regulates HSL activation (Grant et al., 2009). When HSL is activated, the lipolysis process increases, leading to increased FFAs, which explains why FAs are up-regulated in Effect 1, Scenario 1 (Figure22). Then the pathway continues normally, where ß-oxidation is converting FFAs to acetyl-CoA, and this is entering the TCA cycle, Scenario 1 (Figure 22). However, in *COBLL1* risk allele carriers, ß-oxidation is impaired, leading to a degeneration of acetyl-CoA followed by a decrease of the TCA cycle. In addition, the disruption in lipolysis is also leading to excessive degradation of FFAs storage in lipid droplets (Figure 22, Scenario 1). Interestingly, it was shown that lipolysis in risk allele carriers is increased, which means we have excess FFAs, which leads to impairment of ß-oxidation, and leads to a deranged cell function. Moreover, when ß-oxidation is impaired, changes in lipolysis

and lipogenesis are taking place; however, with impairment of lipolysis, leading to down-regulation of BCAA, which may yield less acetyl-CoA available for the TCA cycle (Heinonen et al., 2020).

In Scenario 2, we hypothesized that due to the disruption of ß-oxidation, AAs could not get oxidized. Therefore, they will not be available to be converted into acetyl-CoA and enter the TCA cycle. This hypothesis may explain the up-regulation of AAs in Effect 2 (Table 18).

The failure of fatty acid oxidation can lead to the accumulation of FAs derivatives in muscle or liver (e.g., fatty acyl-CoA, diacylglycerol) that can impair the ability of insulin to stimulate glucose uptake and disposal muscle as well as glucose conversion to glycogen in the liver (Sears et al., 2015). These changes could promote a chronic elevation of blood glucose levels. This, together with the reduced ability to dislocate glucose from the blood, resulting in glucose intolerance, can lead to insulin resistance and, ultimately, T2D. As well-known FAs occur within mitochondria, any disorder in the electron transfer chain will limit the oxidation of FAs, which can eventually result in several clinical issues. The low rate of fat oxidation leads to obesity since any un-oxidized fat will be stored as triacylglycerol in adipose tissue (Sears et al., 2015). Moreover, mitochondrial dysfunction in adipose tissue results in the down-regulation of mitochondrial adipogenesis (Heinonen et al., 2020). This leads to a reduction in the *GLUT4* translocation to the adipocyte cell membrane, which reduces glucose uptake into adipose tissue (Kanzaki et al., 2001), and thus the level of pyruvate available to the TCA cycle.

6. Summary and outlook

During the last 10 years, GWAS has paved the way towards an exciting journey to better understand the genetic architecture of T2D and its related traits (Visscher et al., 2017). The experimental design of GWAS has contributed to a remarkable range of discoveries in human genetics, detecting associations between standard DNA variants and several human diseases and disorders (Schwintzer et al., 2011). Moreover, GWAS led to new developments in disease epidemiology and the discovery of therapeutic drugs (Schwintzer et al., 2011). Those discoveries will detect or unravel new biochemical mechanisms, specifically for T2D prevention and early intervention (Claussnitzer et al., 2014). The overall goal of modern genetics is to translate genetic information into daily clinical practice and elucidate the molecular mechanism of disease (Claussnitzer et al., 2014). We investigated the *COBLL1* locus by studying risk allele carriers of the target SNP rs6712203. We applied untargeted metabolome analysis using ultra-high-resolution mass-spectrometry. Our results revealed possible biotransformations and biochemical pathways involved in glucose uptake, adipogenesis, lipolysis, and TGs storage. Our results agree with the hypotheses published from the *in vitro* findings from Glunk et al. (unpublished, manuscript submitted). Our findings allow us to understand the link between the effect of *COBLL1* on the metabolic phenotypes, linking the impact of *COBLL1* locus and T2D risk. We studied the effect of SNP rs6712203 on metabolic phenotypes by utilizing an untargeted metabolomics approach that allows us to have a comprehensive picture by screening all detectable metabolites and the ones with unknown biotransformation mechanisms from unknown chemical space.

We were able to support the hypothesis that a cell actin dynamic imbalance may lead to perturbations in metabolic pathways. We achieved that by using the platform DI-ICR-FT MS because of its ultra-high-resolution and mass-accuracy (Ramautar et al., 2013) to link genome and metabolome. Our results are promising, and further experiments are required to substantiate our proposed hypotheses and confirm the biochemical mechanisms using adipocyte cells. This cell type was chosen since Glunk et al. identified the function of *COBLL1* in adipose tissue, where they established *COBLL1* gene expression in the adipocyte. Adipose tissue is the primary storage organ of the human body (Newsholme et al., 2010). When surplus energy needs

to be stored, adipocytes can massively expand in number (hyperplasia) or size (hypertrophy) to prevent ectopic fat accumulation (Newsholme et al., 2010). Hypertrophy and hyperplasia are dependent on structural changes, which are mainly controlled by actin dynamics and accommodate the emergence and growth of lipid droplets: actin remodeling and the regulation of adipogenesis ((Newsholme et al., 2010; Langin et al., 2006). Additionally, an essential part of the human body's energy homeostasis is maintained by adipocytes and tightly regulated by chemical, enzymatic, and mechanical influences (Hauner et al., 1998). This explains our interest in investigating the effect of *COBLL1* in adipose tissue energy uptake.

We proposed some possible new mechanisms that would reveal new pathways in the energy metabolism of *COBLL1* carriers. Future studies are required to elucidate these mechanisms further and better understand the link to T2D as an essential step towards the more global aim of personalized prevention, treatment, and medicine.

References:

Abou Ziki, M.D. and A. Mani, *Metabolic syndrome: genetic insights into disease pathogenesis.* Current opinion in lipidology, 2016. **27**(2): p. 162

Ahmadian, M., et al., *Triacylglycerol metabolism in adipose tissue.* Future lipidology, 2007. **2**(2): p. 229-237.

Ali, O., *Genetics of type 2 diabetes.* World journal of diabetes, 2013. **4**(4): p. 114.

American Diabetes Association. 2. Classification and Diagnosis of Diabetes. Diabetes Care. 2017.**40**(Suppl 1):S11-S24

American Diabetes Association. Standards of medical care in diabetes--2011. *Diabetes Care*. 2011; **34**(Suppl 1):S11-S61.

Ashburner, M., et al., *Gene ontology: tool for the unification of biology.* Nature genetics, 2000. **25**(1): p. 25-29.

Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* International journal of analytical chemistry, 2012. 2012

Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*.The cetric acid cycle, 2002, New York: WH Freeman.

Bijlsma, S., et al., *Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation.* Analytical chemistry, 2006. **78**(2): p. 567-574.

Billings, L.K. and J.C. Florez, *The genetics of type 2 diabetes: what have we learned from GWAS?* Annals of the New York Academy of Sciences, 2010. **1212**: p. 59.

Brown, A.E. and M. Walker, *Genetics of insulin resistance and the metabolic syndrome.* Current cardiology reports, 2016. **18**(8): p. 1-8.

Cai, L., et al., *Genome-wide association analysis of type 2 diabetes in the EPIC-InterAct study.* Scientific Data, 2020. **7**(1): p. 1-6.

Cariou, B., et al., *Increased adipose tissue expression of Grb14 in several models of insulin resistance.* The FASEB journal, 2004. **18**(9): p. 965-967.

Cecchini, M., et al., *Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness.* The Lancet, 2010. **376**(9754): p. 1775-1784.

Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals.* Mass spectrometry reviews, 2001. **20**(6): p. 362-387.

Claussnitzer, M., C.-C. Hui, and M. Kellis, *FTO Obesity Variant and Adipocyte Browning in Humans.* The New England journal of medicine, 2015. **374**(2): p. 192-193.

Claussnitzer, M., et al., *Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms.* Cell, 2014. **156**(1-2): p. 343-358.

Consortium, U., *UniProt: a worldwide hub of protein knowledge.* Nucleic acids research, 2019. **47**(D1): p. D506-D515.

Dall, T.M., et al., *The economic burden of elevated blood glucose levels in 2012: diagnosed and undiagnosed diabetes, gestational diabetes mellitus, and prediabetes.* Diabetes care, 2014. **37**(12): p. 3172-3179.

Damcott, C.M., et al., *Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance.* Diabetes, 2006. **55**(9): p. 2654-2659.

Darnton-Hill, I., C. Nishida, and W. James, *A life course approach to diet, nutrition and the prevention of chronic diseases.* Public health nutrition, 2004. **7**(1a): p. 101-121.

Daviglus, M.L., et al., *Favorable cardiovascular risk profile in young women and long-term risk of cardiovascular and all-cause mortality.* Jama, 2004. **292**(13): p. 1588-1592.

Desmarchelier, C., et al., *The postprandial chylomicron triacylglycerol response to dietary fat in healthy male adults is significantly explained by a combination of single nucleotide polymorphisms in genes involved in triacylglycerol metabolism.* The Journal of Clinical Endocrinology & Metabolism, 2014. **99**(3): p. E484-E488.

Dumas, M.-E., et al., *Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study.* Analytical chemistry, 2006. **78**(7): p. 2199-2208.

Eichner, J., et al., *Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software.* Journal of Chromatography B, 2014. **966**: p. 77-82.

Forcisi, S., et al., *Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra-high pressure liquid chromatography coupling.* Journal of Chromatography A, 2013. **1292**: p. 51-65.

Forcisi, S., et al., *Solutions for low and high accuracy mass spectrometric data matching: a data-driven annotation strategy in nontargeted metabolomics.* Analytical chemistry, 2015. **87**(17): p. 8917-8924.

Fritsche, Andreas, et al. "Different Effects of Lifestyle Intervention in High-and Low-Risk Prediabetes: Results of the Randomized Controlled Prediabetes Lifestyle Intervention Study (PLIS)." *Diabetes*, 2021. P. 2785-2795.

Gavaghan, C.L., et al., *An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk: ApfCD mouse.* FEBS letters, 2000. **484**(3): p. 169-174.

Gerich JE., Szoke E. Pathogenesis of Type 2 Diabetes: Atlas of diabetes. *Current Medicin LLC,*Philadelphia, US,2007:111-117.

Ghaben, A.L. and P.E. Scherer, *Adipogenesis and metabolic health.* Nature reviews Molecular cell biology, 2019. **20**(4): p. 242-258.

Ghaste, M., R. Mistrik, and V. Shulaev, *Applications of Fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics.*

International journal of molecular sciences, 2016. **17**(6): p. 816.

Gibney, M.J., et al., *Metabolomics in human nutrition: opportunities and challenges–.* The American journal of clinical nutrition, 2005. **82**(3): p. 497-503.

Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-56.

Gieger, C., et al., *Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum.* PLoS genetics, 2008. **4**(11): p. e1000282.

Glümer, C., et al., *A Danish diabetes risk score for targeted screening: the Inter99 study.* Diabetes care, 2004. **27**(3): p. 727-733.

González-Domínguez, R., A. Sayago, and Á. Fernández-Recamales, *Direct infusion mass spectrometry for metabolomic phenotyping of diseases.* Bioanalysis, 2017. **9**(1): p. 131-148.

Gordon, G.J., R. Bueno, and D.J. Sugarbaker, *Genes associated with prognosis after surgery for malignant pleural mesothelioma promote tumor cell survival in vitro.* BMC cancer, 2011. **11**(1): p. 1-9.

Grant, R.W., A.F. Moore, and J.C. Florez, *Genetic architecture of type 2 diabetes: recent progress and clinical implications.* Diabetes care, 2009. **32**(6): p. 1107-1114.

Greenberg, A.S., et al., *The role of lipid droplets in metabolic disease in rodents and humans.* The Journal of clinical investigation, 2011. **121**(6): p. 2102-2110.

Guariguata, L., et al., *Global estimates of diabetes prevalence for 2013 and projections for 2035.* Diabetes research and clinical practice, 2014. **103**(2): p. 137-149.

Han, J.-L. and H.-L. Lin, *Intestinal microbiota and type 2 diabetes: from mechanism insights to therapeutic perspective.* World journal of gastroenterology: WJG, 2014. **20**(47): p. 17737.

Hara, K., et al., *Genetic architecture of type 2 diabetes.* Biochemical and biophysical research communications, 2014. **452**(2): p. 213-220.

Harris, R.B., *Direct and indirect effects of leptin on adipocyte metabolism.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2014. **1842**(3): p. 414-423.

Hauner, H., et al., *Development of insulin-responsive glucose uptake and GLUT4 expression in differentiating human adipocyte precursor cells.* International journal of obesity, 1998. **22**(5): p. 448-453.

Heid, I.M., et al., *Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution.* Nature genetics, 2010. **42**(11): p. 949-960.

Heinonen, S., et al., *White adipose tissue mitochondrial metabolism in health and in obesity.* Obesity Reviews, 2020. **21**(2).

Henninger, A.J., et al., *Adipocyte hypertrophy, inflammation and fibrosis characterize subcutaneous adipose tissue of healthy, non-obese subjects predisposed to type 2 diabetes.* PloS one, 2014. **9**(8): p. e105262.

Herzberg-Schäfer, S.A., et al., *Evaluation of fasting state-/oral glucose tolerance test-derived measures of insulin release for the detection of genetically impaired β-cell function.* PloS one, 2010. **5**(12): p. e14194.

Hussain, A., et al., *Prevention of type 2 diabetes: a review.* Diabetes research and clinical practice, 2007. **76**(3): p. 317-326.

Husson, C., et al., *Cordon-Bleu uses WH2 domains as multifunctional dynamizers of actin filament assembly.* Molecular cell, 2011. **43**(3): p. 464-477.

International Diabetes Federation, (2017).p. 164-173.

International Diabetes Federation, *IDF diabetes atlas 8th edition.* 2017: p. 905-911.

International Diabetes Federation. IDF Diabetes Atlas, 8th edn. Brussels, Belgium.

Jassal, B., et al., *The reactome pathway knowledgebase.* Nucleic acids research, 2020. **48**(D1): p. D498-D503.

Johnson, A.D., et al., *SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.* Bioinformatics, 2008. **24**(24): p. 2938-2939.

Johnson, C.H., J. Ivanisevic, and G. Siuzdak, *Metabolomics: beyond biomarkers and towards mechanisms.* Nature reviews Molecular cell biology, 2016. **17**(7): p. 451-459.

Joslin, A.C., et al., *A functional genomics pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci.* Nature communications, 2021. **12**(1): p. 1-15.

Kan, M., et al., *Rare variant associations with waist-to-hip ratio in European-American and African-American women from the NHLBI-Exome Sequencing Project.* European Journal of Human Genetics, 2016. **24**(8): p. 1181-1187.

Kanzaki, M. and J.E. Pessin, *Insulin-stimulated GLUT4 Translocation in Adipocytes Is Dependent upon Cortical Actin Remodeling* 210.* Journal of Biological Chemistry, 2001. **276**(45): p. 42436-42444.

Karpe, F., J.R. Dickmann, and K.N. Frayn, *Fatty acids, obesity, and insulin resistance: time for a reevaluation.* Diabetes, 2011. **60**(10): p. 2441-2449.

Kawaguchi, N., et al., *ADAM12 induces actin cytoskeleton and extracellular matrix reorganization during early adipocyte differentiation by regulating β1 integrin function.* Journal of cell science, 2003. **116**(19): p. 3893-3904.

Kessels, M.M. and B. Qualmann, *The syndapin protein family: linking membrane trafficking with the cytoskeleton.* Journal of cell science, 2004. **117**(15): p. 3077-3086.

Kimmel, A.R. and C. Sztalryd, *Perilipin 5, a lipid droplet protein adapted to mitochondrial energy utilization.* Current opinion in lipidology, 2014. **25**(2): p. 110.

Klassen, A., et al., *Metabolomics: definitions and significance in systems biology.* Metabolomics: From Fundamentals to Clinical Applications, 2017: p. 3-17.

Knowler, WC., et al., *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin,* 2002: 393.

Kraja, A.T., et al., *Pleiotropic genes for metabolic syndrome and inflammation.* Molecular genetics and metabolism, 2014. **112**(4): p. 317-338.

Laber, S., et al., *Linking the FTO obesity rs1421085 variant circuitry to cellular, metabolic, and organismal phenotypes in vivo*. Sci Adv. 2021. **7**(30).

Laing, N.G., et al., *Mutations and polymorphisms of the skeletal muscle α-actin gene (ACTA1).* Human mutation, 2009. **30**(9): p. 1267-1277.

Langin, D., *Adipose tissue lipolysis as a metabolic pathway to define pharmacological strategies against obesity and the metabolic syndrome.* Pharmacological Research, 2006. **53**(6): p. 482-491.

Leahy, J.L., *Pathogenesis of type 2 diabetes mellitus.* Archives of medical research, 2005. **36**(3): p. 197-209.

Lefebvre, B., et al., *Regulation and functional effects of ZNT8 in human pancreatic islets.* Journal of endocrinology, 2012. **214**(2): p. 225.

Lehninger, A.L., et al., *principles of biochemistry*.5th edition. Newyork, USA. 2008.

Leichtle, A.B., J.-F. Dufour, and M. Fiedler, *Potentials and pitfalls of clinical peptidomics and metabolomics.* Swiss medical weekly, 2013. **143**: p. w13801.

Likić, V.A., et al., *Systems biology: the next frontier for bioinformatics.* Advances in bioinformatics, 2010.

Lindblad-Toh, K., et al., *A high-resolution map of human evolutionary constraint using 29 mammals.* Nature, 2011. **478**(7370): p. 476-482.

Liu, X. and J.W. Locasale, *Metabolomics: a primer.* Trends in biochemical sciences, 2017. **42**(4): p. 274-284.

Long, T., et al., *Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites.* Nature genetics, 2017. **49**(4): p. 568-578.

Lowe, C.E., S. O'Rahilly, and J.J. Rochford, *Adipogenesis at a glance.* Journal of cell science, 2011. **124**(16): p. 2681-2686.

Lu, Y., et al., *New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk.* Nature communications, 2016. **7**(1): p. 1-15.

Lumish, H.S., M. O'Reilly, and M.P. Reilly, *Sex differences in genomic drivers of adipose distribution and related cardiometabolic disorders: opportunities for precision medicine.* Arteriosclerosis, thrombosis, and vascular biology, 2020. **40**(1): p. 45-60.

Luo, P. and M.-H. Wang, *Eicosanoids, β-cell function, and diabetes.* Prostaglandins & other lipid mediators, 2011. **95**(1-4): p. 1-10.

Machann, J., et al., *Hepatic lipid accumulation in healthy subjects: a comparative study using spectral fat-selective MRI and volume-localized 1H-MR spectroscopy.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2006. **55**(4): p. 913-917.

Machann, J., et al., *Standardized assessment of whole body adipose tissue topography by MRI.* Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2005. **21**(4): p. 455-462.

Mahajan, A., et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.* Nature genetics, 2014. **46**(3): p. 234-244.

Mancina, R.M., et al., *The COBLL1 C allele is associated with lower serum insulin levels and lower insulin resistance in overweight and obese children.* Diabetes/metabolism research and reviews, 2013. **29**(5): p. 413-416.

Manning, A.K., et al., *A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance.* Nature genetics, 2012. **44**(6): p. 659-669.

Matsuda, M. and R.A. DeFronzo, *Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp.* Diabetes care, 1999. **22**(9): p. 1462-1470.

McCarthy, M.I. and E. Zeggini, *Genome-wide association studies in type 2 diabetes.* Current diabetes reports, 2009. **9**(2): p. 164-171.

McCarthy, M.I., *Genomics, type 2 diabetes, and obesity.* New England Journal of Medicine, 2010. **363**(24): p. 2339-2350.

Mergenthaler, P., et al., *Sugar for the brain: the role of glucose in physiological and pathological brain function.* Trends in neurosciences, 2013. **36**(10): p. 587-597.

Minami, Y., et al., *Measurement of internal body time by blood metabolomics.* Proceedings of the National Academy of Sciences, 2009. **106**(24): p. 9890-9895.

Mishra, R., et al., *Epigenetic changes in fibroblasts drive cancer metabolism and differentiation.* Endocrine-related cancer, 2019. **26**(12): p. R673-R688.

Morandi, E., et al., *ITGAV and ITGA5 diversely regulate proliferation and adipogenic differentiation of human adipose derived stem cells.* Scientific reports, 2016. **6**(1): p. 1-14.

Moritz F.,Kaling M.,Schmitt-Kopplin P.,Schnitzler JP. *Characterization of poplar metabotypes via mass difference enrichment analysis*.Plant Cell Environ.2017.**40**:1057-1073.

Moritz, F., et al., *Mass differences in metabolome analyses of untargeted direct infusion ultra-high resolution MS data*, in *Fundamentals and Applications of Fourier Transform Mass Spectrometry*. 2019, Elsevier. p. 357-405.

Morris, A.P., et al., *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes.* Nature genetics, 2012. **44**(9): p. 981.

Müssig, K., et al., *Genetic variants affecting incretin sensitivity and incretin secretion.* Diabetologia, 2010. **53**(11): p. 2289-2297.

Müssig, K., et al., *The ENPP1 K121Q polymorphism determines individual susceptibility to the insulin-sensitising effect of lifestyle intervention.* Diabetologia, 2010. **53**(3): p. 504-509.

Narayanan, S., *The preanalytic phase: an important component of laboratory medicine.* American journal of clinical pathology, 2000. **113**(3): p. 429-452.

Naz, S., et al., *Method validation strategies involved in non-targeted metabolomics.* Journal of Chromatography A, 2014. **1353**: p. 99-105.

Newsholme, E. and A. Leech, *Functional biochemistry in health and disease*. Fat metabolism 2010: John Wiley & Sons.

Newsholme, E. and A. Leech, *Functional biochemistry in health and disease*. Lipid 2010: John Wiley & Sons.

Newsholme, E. and A. Leech, *Functional biochemistry in health and disease*. Lipid biosynthesis 2010: John Wiley & Sons.

Newsholme, P., L. Brennan, and K. Bender, *Amino acid metabolism, β-cell function, and diabetes.* Diabetes, 2006. **55**(Supplement 2): p. S39-S47.

Nicholson, J., *LindonJC.* Systems biology: Metabonomics, 2008. **455**(7216): p. 1054-1056.

Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function.* Nature reviews Drug discovery, 2002. **1**(2): p. 153-161.

Nicholson, J.K., *Global systems biology, personalized medicine and molecular epidemiology*. 2006, John Wiley & Sons, Ltd Chichester, UK.

Nicholson, J.K., J.C. Lindon, and E. Holmes, *'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.* Xenobiotica, 1999. **29**(11): p. 1181-1189.

Norris, D.M., et al., *Glucose transport: Methods for interrogating GLUT4 trafficking in adipocytes*, in *Glucose Transport*. 2018, Springer. p. 193-215.

Orlicky, D.J., et al., *Dynamics and molecular determinants of cytoplasmic lipid droplet clustering and dispersion.* PloS one, 2013. **8**(6): p. e66837.

Pan, X.-R., et al., *Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study.* Diabetes care, 1997. **20**(4): p. 537-544.

Pandey, K.B. and S.I. Rizvi, *Plant polyphenols as dietary antioxidants in human health and disease.* Oxidative medicine and cellular longevity, 2009. **2**(5): p. 270-278.

Park, J.-W., et al., *Molecular Analysis of Alternative Transcripts of the Equine Cordon-Bleu WH2 Repeat Protein-Like 1 (COBLL1) Gene.* Asian-Australasian journal of animal sciences, 2015. **28**(6): p. 870.

Patti, G.J., O. Yanes, and G. Siuzdak, *Metabolomics: the apogee of the omics trilogy.* Nature reviews Molecular cell biology, 2012. **13**(4): p. 263-269.

Perreault L, Kahn SE, Christophi CA, Knowler WC; Diabetes Prevention Program Research Group. Regression from pre-diabetes to normal glucose regulation in the diabetes prevention program. Diabetes Care 2009; 32:1583–1588 12.

Perrin, B.J. and J.M. Ervasti, *The actin gene family: function follows isoform.* Cytoskeleton, 2010. **67**(10): p. 630-634.

Perry, J.R., et al., *Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases.* PLoS genetics, 2012. **8**(5): p. e1002741.

Pinu, F.R., S.A. Goldansaz, and J. Jaine, *Translational metabolomics: current challenges and future opportunities.* Metabolites, 2019. **9**(6): p. 108.

Poretsky, L., *Principles of diabetes mellitus*. Vol. 21. 2010: Springer.

Prasad, R.B. and L. Groop, *Genetics of type 2 diabetes—pitfalls and possibilities.* Genes, 2015. **6**(1): p. 87-123.

Prokopenko, D.J.L.C., *I Saxena R Soranzo N Jackson AU Wheeler E Glazer NL Bouatia-Naji N Gloyn AL et al. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nature genetics. **42**: p. 105-116.

Ramachandran, A., et al., *The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1).* Diabetologia, 2006. **49**(2): p. 289-297.

Ramautar, R., et al., *Human metabolomics: strategies to understand biology.* Current opinion in chemical biology, 2013. **17**(5): p. 841-846.

Roberts, L.D., A. Koulman, and J.L. Griffin, *Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome.* The lancet Diabetes & endocrinology, 2014. **2**(1): p. 65-75.

Roglic, G., *Global report on diabetes*. 2016: World Health Organization.

Sakai, T., et al., *Integrin-linked kinase (ILK) is required for polarizing the epiblast, cell adhesion, and controlling actin accumulation.* Genes & development, 2003. **17**(7): p. 926-940.

Salans, L.B., J.L. Knittle, and J. Hirsch, *The role of adipose cell size and adipose tissue insulin sensitivity in the carbohydrate intolerance of human obesity.* The Journal of clinical investigation, 1968. **47**(1): p. 153-165.

Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-1336.

Schäfer, S., et al., *Lifestyle intervention in individuals with normal versus impaired glucose tolerance.* European journal of clinical investigation, 2007. **37**(7): p. 535-543.

Scharf, P.J., et al., *Solution structure of the human Grb14–SH2 domain and comparison with the structures of the human Grb7–SH2/erbB2 peptide complex and human Grb10–SH2 domain.* Protein science, 2004. **13**(9): p. 2541-2546.

Schleinitz, D., et al., *Fat depot-specific mRNA expression of novel loci associated with waist–hip ratio.* International journal of obesity, 2014. **38**(1): p. 120-125.

Schmid V, Wagner R, Sailer C, et al. Non-alcoholic fatty liver disease and impaired proinsulin conversion as newly identified predictors of the long-term non-response to a lifestyle intervention for diabetes prevention: results from the TULIP study. Diabetologia 2017; **60**:2341–2351

Schulze, M., et al., *Body adiposity index, body fat content and incidence of type 2 diabetes.* Diabetologia, 2012. **55**(6): p. 1660-1667.

Schwenzer, N.F., et al., *Quantitative analysis of adipose tissue in single transverse slices for estimation of volumes of relevant fat tissue compartments: a study in a large cohort of subjects at risk for type 2 diabetes by MRI with comparison to anthropometric data.* Investigative radiology, 2010. **45**(12): p. 788-794.

Schwintzer, L., et al., *The functions of the actin nucleator Cobl in cellular morphogenesis critically depend on syndapin I.* The EMBO journal, 2011. **30**(15): p. 3147-3159.

Sears, B. and M. Perry, *The role of fatty acids in insulin resistance.* Lipids in health and disease, 2015. **14**(1): p. 1-9.

Sharma, V., et al., *Replication of newly identified type 2 diabetes susceptible loci in Northwest Indian population.* Diabetes research and clinical practice, 2017. **126**: p. 160-163.

Shaw, J.E., R.A. Sicree, and P.Z. Zimmet, *Global estimates of the prevalence of diabetes for 2010 and 2030.* Diabetes research and clinical practice, 2010. **87**(1): p. 4-14.

Shungin, D., et al., *New genetic loci link adipose and insulin biology to body fat distribution.* Nature, 2015. **518**(7538): p. 187-196.

Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-885.

Smas, C.M. and H.S. Sul, *Control of adipocyte differentiation.* Biochemical Journal, 1995. **309**(3): p. 697-710.

Suhre, K. and C. Gieger, *Genetic variation in metabolic phenotypes: study designs and applications.* Nature reviews genetics, 2012. **13**(11): p. 759-769.

Suhre, K., et al., *Human metabolic individuality in biomedical and pharmaceutical research.* Nature, 2011. **477**(7362): p. 54-60.

Talior-Volodarsky, I., et al., *α-Actinin-4 is selectively required for insulin-induced GLUT4 translocation.* Journal of Biological Chemistry, 2008. **283**(37): p. 25115-25123.

Tansey, J., et al., *Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity.* Proceedings of the National Academy of Sciences, 2001. **98**(11): p. 6494-6499.

Tong, Y., et al., *Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis.* BMC medical genetics, 2009. **10**(1): p. 1-25.

Travers, M.E. and M.I. McCarthy, *Type 2 diabetes and obesity: genomics and the clinic.* Human genetics, 2011. **130**(1): p. 41-58.

Tuomilehto, J., et al., *Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.* New England Journal of Medicine, 2001. **344**(18): p. 1343-1350.

Tyagi, S., et al., *The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases.* Journal of advanced pharmaceutical technology & research, 2011. **2**(4): p. 236.

Tziotis D, Hertkorn N, Schmitt-Kopplin P. Kendrick-Analogous *Network Visualisation of Ion Cyclotron Resonance Fourier Transform Mass Spectra: Improved Options for the Assignment of Elemental Compositions and the Classification of Organic Molecular Complexity*. European Journal of Mass Spectrometry. 2011; **17**(4):415-421.

Uusitupa, M., et al., *Ten-year mortality and cardiovascular morbidity in the Finnish Diabetes Prevention Study—secondary analysis of the randomized trial.* PloS one, 2009. **4**(5): p. e5656.

Villas-Bôas SG, Mas S, Akesson M, Smedsgaard J, Nielsen J. Mass spectrometry in metabolome analysis. Mass Spectrom Rev. 2005. **24**(5):613-46

Villas-Boas, S.G., et al., *Metabolome analysis: an introduction*. Vol. 24. 2007: John Wiley & Sons.

Visscher, P.M., et al., *10 years of GWAS discovery: biology, function, and translation.* The American Journal of Human Genetics, 2017. **101**(1): p. 5-22.

Vujkovic, M., et al., *Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis.* Nature genetics, 2020. **52**(7): p. 680-691.

Wang, R.S., B.A. Maron, and J. Loscalzo, *Systems medicine: evolution of systems biology from bench to bedside.* Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2015. **7**(4): p. 141-161.

Weaver, A.M., et al., *Integration of signals to the Arp2/3 complex.* Current opinion in cell biology, 2003. **15**(1): p. 23-30.

Wijmenga, C. and A. Zhernakova, *The importance of cohort studies in the post-GWAS era.* Nature genetics, 2018. **50**(3): p. 322-328.

Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome.* Nucleic acids research, 2009. **37**(suppl_1): p. D603-D610.

World Health Organization, *Biomarkers and Risk Assessment: Concepts and Principles-Environmental Health Criteria 155*. 1993.

World Health Organization*, Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation*. 2006.

World Health Organization. Global report on diabetes. *World Health Organization*. (2021). Geneva

Yang, W., et al., *BSCL2/seipin regulates adipogenesis through actin cytoskeleton remodelling.* Human molecular genetics, 2014. **23**(2): p. 502-513.

Yin, P. and G. Xu, *Metabolomics for tumor marker discovery and identification based on chromatography–mass spectrometry.* Expert review of molecular diagnostics, 2013. **13**(4): p. 339-348.

Yin, P., et al., *Preanalytical aspects and sample quality assessment in metabolomics studies of human blood.* Clinical chemistry, 2013. **59**(5): p. 833-845.

Yin, P., R. Lehmann, and G. Xu, *Effects of pre-analytical processes on blood samples used in metabolomics studies.* Analytical and bioanalytical chemistry, 2015. **407**(17): p. 4879-4892.

Zeggini, E., et al., *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* Nature genetics, 2008. **40**(5): p. 638-645.

Zhang, A., et al., *Modern analytical techniques in metabolomics analysis.* Analyst, 2012. **137**(2): p. 293-300.

Appendix 1

Einwilligungserklärung zur Studie
„Individualisierte Lebensstilintervention bei Prädiabetes“
(PLIS - Prediabetes Lifestyle Intervention Study)

Name und Anschrift des lokalen Prüfzentrums, in dem die klinische Prüfung durchgeführt wird:

Institut für Ernährungsmedizin
Klinikum rechts der Isar der TU München
München Uptown Campus D
Georg-Brauchle-Ring 62
80992 München
Tel.: 089/289 249-25
Fax: 089/289 249-13

Name und Telefonnummer des/r aufklärenden Prüfarztes/ärztin:

……
……
……...

„Individualisierte Lebensstilintervention bei Prädiabetes“

Vorname, Name: .....

Geburtsdatum: .....

Ich wurde über das Ziel dieser klinischen Studie, die Dauer, den Ablauf, den Nutzen sowie sämtliche Risiken und Komplikationen der Studienteilnahme ausführlich informiert.
Ich hatte dabei ausgiebig Zeit und Gelegenheit, noch offene Fragen zum Ziel der klinischen Studie und zu den möglichen Risiken der Untersuchungen zu stellen. Im Einzelnen wurden folgende Fragen besprochen:

.....

.....

Zurzeit habe ich keine weiteren Fragen, bin aber darüber informiert, dass ich mich jederzeit während oder nach der klinischen Studie an meine/n Prüfarzt/ärztin wenden kann.

Ich wurde darüber informiert, dass für die Abschätzung des Risikos meiner Teilnahme an dieser klinischen Studie die wahrheitsgemäßen Angaben über frühere bzw. derzeit bestehende Erkrankungen oder die Einnahme von Arzneimitteln und Drogen von besonderer Bedeutung sind. Ich erkläre, dass ich frühere Erkrankungen und eingenommene Medikamente nicht absichtlich verschweige.
Ich bin darüber informiert, dass meine Studienteilnahme vollkommen freiwillig ist und dass ich die Einwilligung zur Teilnahme jederzeit ohne Angabe von Gründen widerrufen kann, ohne dass dies für mich nachteilige Folgen hat oder meine weitere Behandlung beeinträchtigt.
Die schriftliche Patienteninformation zu dieser klinischen Studie und die Einwilligungserklärung mit Datenschutzerklärung habe ich erhalten und genau gelesen. Ich hatte genügend Zeit, meine Teilnahme an der klinischen Prüfung zu überdenken.

Ich bin mit der Teilnahme an dieser klinischen Studie einverstanden.

☐ ja ☐ nein

Ich bin damit einverstanden, dass mir im Rahmen der klinischen Studie Blutproben entnommen und für 10 Jahre im Rahmen der DZD Biobank zugangskontrolliert im Labor des Prüfzentrums Institut/Lehrstuhl für Ernährungsmedizin der TU München, Standorte Uptown München Campus D und Freising/Weihenstephan () in pseudonymisierter (verschlüsselter) Form, d.h. ohne Nennung von Name oder Geburtsdatum aufbewahrt und zur Analyse studienspezifischer Fragestellungen verwendet werden.

☐ ja ☐ nein

München,

.....
Datum, Vor-/Zuname in Druckschrift **Unterschrift Patient/in**

München,

.....
Datum, Vor-/Zuname in Druckschrift **Unterschrift Prüfarzt/ärztin**

Else Kröner-Fresenius-Zentrum für Ernährungsmedizin
Klinikum rechts der Isar
Technische Universität München

Direktor: Univ.-Prof. Dr. med. Hans Hauner

Datenschutzerklärung zur Studie
„Individualisierte Lebensstilintervention bei Prädiabetes“

Mir ist bekannt, dass bei dieser klinischen Prüfung personenbezogene Daten, insbesondere medizinische Befunde, über mich erhoben, gespeichert und ausgewertet werden sollen. Die Verwendung der Angaben über meine Gesundheit erfolgt nach gesetzlichen Bestimmungen und setzt vor der Teilnahme an der klinischen Prüfung folgende freiwillig abgegebene Einwilligungserklärung voraus, d.h. ohne die nachfolgende Einwilligung kann ich nicht an der klinischen Prüfung teilnehmen.

Einwilligungserklärung zum Datenschutz

1) Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten, insbesondere Angaben über meine Gesundheit, in Papierform und auf elektronischen Datenträgern im Prüfzentrum *(Institut für Ernährungsmedizin, Klinikum rechts der Isar der TU München, München Uptown Campus D, Georg-Brauchle-Ring 62, 80992 München)* und im Koordinierungszentrum für klinische Studien Düsseldorf *(KKS, Medizinische Fakultät, Universitätsklinikum Heinrich-Heine-Universität, Moorenstr. 5, 40225 Düsseldorf)* aufgezeichnet werden. Soweit erforderlich, dürfen die erhobenen Daten pseudonymisiert (verschlüsselt) weitergegeben werden:

   a) an den Sponsor *(Universitätsklinikum Tübingen, Verantwortlicher für die Sponsorpflichten Prof. Dr. Häring)* oder eine von diesem beauftragte Stelle zum Zwecke der wissenschaftlichen Auswertung

   b) im Falle unerwünschter Ereignisse: an den Sponsor *(Universitätsklinikum Tübingen, Verantwortlicher für die Sponsorpflichten Prof. Dr. Häring),* an die zuständige Ethik-Kommission

2) Außerdem erkläre ich mich damit einverstanden, dass autorisierte und zur Verschwiegenheit verpflichtete Beauftragte des Sponsors (*Universitätsklinikum Tübingen, Verantwortlicher für die Sponsorpflichten Prof. Dr. Häring)* sowie die zuständigen Überwachungsbehörden und zuständige Ethik-Kommission in meine beim Prüfarzt vorhandenen personenbezogenen Daten, insbesondere meine Gesundheitsdaten, Einsicht nehmen, soweit dies für die Überprüfung der ordnungsgemäßen Durchführung der Studie notwendig ist. Für diese Maßnahme entbinde ich das Studienteam von der Verpflichtung auf das Daten- und Gesundheitsgeheimnis.

3) Die Einwilligung zur Erhebung und Verarbeitung meiner personenbezogenen Daten, insbesondere der Angaben über meine Gesundheit, ist unwiderruflich. Ich bin bereits darüber aufgeklärt worden, dass ich jederzeit die Teilnahme an der Studie beenden kann. Im Fall eines solchen Widerrufs meiner Einwilligung, an der Studie teilzunehmen, erkläre ich mich damit einverstanden, dass die bis zu diesem Zeitpunkt gespeicherten Daten ohne Namensnennung weiterhin verwendet werden dürfen, soweit dies erforderlich ist, um sicherzustellen, dass meine schutzwürdigen Interessen nicht beeinträchtigt werden.

4) Ich erkläre mich damit einverstanden, dass meine Daten nach Beendigung oder Abbruch der Prüfung mindestens 15 Jahre aufbewahrt werden. Danach werden meine personenbezogenen Daten gelöscht, soweit nicht gesetzliche, satzungsmäßige oder vertragliche Aufbewahrungsfristen entgegenstehen.

Eine Kopie des Informationsblattes und der Einverständniserklärung habe ich erhalten.

München,

.....

Datum, Vor-/Zuname in Druckschrift **Unterschrift Patient/in**

München,

.....

Datum, Vor-/Zuname in Druckschrift **Unterschrift Prüfarzt/ärztin**

Einwilligungserklärung zum DNA-Screening

Ich bin bereit, an dem Projekt **„Individualisierte Lebensstilintervention bei Prädiabetes – Diabetes-Screening“** teilzunehmen.

Ich bin mit der genetischen Analyse meiner Erbinformation zu wissenschaftlichen Zwecken einverstanden. Die Analyse der Erbinformation beschränkt sich auf die Gene, die möglicherweise für den Diabetes mellitus und mit ihm zusammenhängende Stoffwechselerkrankungen verantwortlich gemacht werden können. Die Ergebnisse der genetischen Analyse dienen nur zu Forschungszwecken. Deshalb werden Ergebnisse der Forschung insgesamt und individuelle Befunde weder Ihnen noch Ihrem Arzt zugänglich gemacht.

Das Arztgeheimnis wird gewahrt. Daten werden nicht an Dritte weitergegeben. Mit der Veröffentlichung der Ergebnisse der Studie in anonymisierter Form bin ich einverstanden.

☐ Ich bin mit der einmaligen Entnahme von 12,5 Millilitern Blut und der zugangskontrollierten Lagerung des genetischen Materials im Hauptprüfzentrum in Tübingen (Universitätsklinikum Tübingen, Medizinische Klinik IV) für 10 Jahre einverstanden.

☐ Ich möchte NICHT am Diabetesgen-Screening teilnehmen und möchte KEINE zusätzliche Blutentnahme.

Ich erkläre, dass meine Teilnahme an diesem Teil der Untersuchungen ebenfalls freiwillig ist. Ich weiß, dass ich keine Nachteile habe, wenn ich nicht am Diabetesgen-Screening teilnehme. Die Teilnahme an der aktuellen Untersuchung wird dadurch nicht berührt, insbesondere entstehen mir keine Nachteile für eine weitere Betreuung in der Klinik.

Ich kann meine Zustimmung zur Teilnahme am Diabetesgen-Screening jederzeit auch nachträglich ohne Angabe von Gründen wieder zurückziehen.

Eine Kopie des Informationsblattes und der Einverständniserklärung habe ich erhalten.

München,

.....
Datum, Vor-/Zuname in Druckschrift **Unterschrift Patient/in**

München,

.....
Datum, Vor-/Zuname in Druckschrift **Unterschrift Prüfarzt/ärztin**

DZD
Deutsches Zentrum
für Diabetesforschung

PLIS	
Datum	
Visite	
Screennr.	
Randomnr.	

Diabetes Prevention Programme (DPP) Resource Utilization and Costs of Intervention Questionnaire - Übersetzte Fragen

1. Wie viel Zeit in Stunden verbringen Sie in einer typischen Woche durchschnittlich damit, Lebensmittel einzukaufen und Essen für sich zuzubereiten?

☐☐ Stunden

2. Denken Sie an alle Übungen oder körperlichen Aktivitäten abgesehen von Ihrer Arbeit, die Sie aktuell für Ihre Gesundheit ausüben und bewerten Sie, wie viel Freude oder Zufriedenheit Sie insgesamt daraus schöpfen. **Kreuzen Sie bitte nur ein Kästchen an.**

1. ☐ Ich mag/genieße diese Aktivitäten, sie verschaffen mir Zufriedenheit.
2. ☐ Neutral.
3. ☐ Ich mag/genieße diese Aktivitäten nicht, sie verschaffen mir keine Zufriedenheit.

Die folgenden Fragen sind erst für die Nachbefragungen der Studien relevant. Bei Erstbefragung bitte überspringen.

3. Haben Sie in den letzten 6 Monaten dafür bezahlt, an kommerziellen Programmen zum Gewichtsverlust, etwa Weight Watchers, Optifast, etc. teilzunehmen?

☐ JA, (was?_____) ☐ NEIN

Wenn ja: Wie hoch sind die gesamten Kosten in den letzten 6 Monaten?

_____ EUR (wenn Sie den genauen Betrag nicht kennen, schätzen Sie ihn bitte)

DZD
Deutsches Zentrum
für Diabetesforschung

PLIS	
Datum	
Visite	
Screennr.	
Randomnr.	

4. Haben Sie in den letzten 6 Monaten einen oder mehrere der nachfolgenden Artikel gekauft, um Ihre Fitness, Gesundheit und Ihr Wohlbefinden zu fördern? Wenn ja, geben Sie bitte den entsprechenden Betrag in EUR an (wenn Sie den genauen Betrag nicht kennen, schätzen Sie ihn bitte).

	NEIN	JA	Betrag in EUR
1. Fahrrad	☐	☐	
2. Langlaufski	☐	☐	_____
3. Abfahrtski oder Snowboard	☐	☐	_____
4. Trainingsvideos	☐	☐	_____
5. Gewichte (Hanteln)	☐	☐	_____
6. Golfschläger	☐	☐	_____
7. Fitnessgerät für zu Hause	☐	☐	_____
8. Schlittschuhe	☐	☐	_____
9. Rollschuhe oder Inline Skates	☐	☐	_____
10. Rudergerät	☐	☐	_____
11. Crosstrainer/Skigerät	☐	☐	_____
12. Schneeschuhe	☐	☐	_____
13. Stepper	☐	☐	_____
14. Hometrainer (Fahrrad)	☐	☐	_____
15. Step (für Aerobic)	☐	☐	_____
16. Tennisschläger	☐	☐	_____
17. Laufband	☐	☐	_____
18. Sonstiges (was? _____)	☐	☐	_____

DZD
Deutsches Zentrum
für Diabetesforschung

PLIS	
Datum	
Visite	
Screennr.	
Randomnr.	

5. Haben Sie in den letzten 6 Monaten eine oder mehrere dieser kostenpflichtigen Dienstleistungen in Anspruch genommen, um Ihre Fitness, Gesundheit und Ihr Wohlbefinden zu fördern? Wenn ja, geben Sie bitte die monatlichen Kosten für die jeweilige Dienstleistung an (wenn Sie den Betrag nicht kennen, schätzen Sie ihn bitte).

	NEIN	JA	Betrag in EUR
1. Sport- oder Aerobic-Kurse	☐	☐	_____
2. Kochkurse	☐	☐	_____
3. Mitgliedschaft in einem Gesundheitsverein oder Fitnessstudio	☐	☐	_____
4. Wellness-Center oder Kur zum Gewichtsverlust	☐	☐	_____
5. Persönlicher Trainer	☐	☐	_____
6. Sonstiges (was? _____)	☐	☐	_____

Gesundheitsfragebogen

Deutsche Version für Deutschland

(German version for Germany)

Bitte kreuzen Sie unter jeder Überschrift DAS Kästchen an, das Ihre Gesundheit HEUTE am besten beschreibt.

BEWEGLICHKEIT / MOBILITÄT

Ich habe keine Probleme herumzugehen ❑
Ich habe leichte Probleme herumzugehen ❑
Ich habe mäßige Probleme herumzugehen ❑
Ich habe große Probleme herumzugehen ❑
Ich bin nicht in der Lage herumzugehen ❑

FÜR SICH SELBST SORGEN

Ich habe keine Probleme, mich selbst zu waschen oder anzuziehen ❑
Ich habe leichte Probleme, mich selbst zu waschen oder anzuziehen ❑
Ich habe mäßige Probleme, mich selbst zu waschen oder anzuziehen ❑
Ich habe große Probleme, mich selbst zu waschen oder anzuziehen ❑
Ich bin nicht in der Lage, mich selbst zu waschen oder anzuziehen ❑

ALLTÄGLICHE TÄTIGKEITEN *(z. B. Arbeit, Studium, Hausarbeit, Familien- oder Freizeitaktivitäten)*

Ich habe keine Probleme, meinen alltäglichen Tätigkeiten nachzugehen ❑
Ich habe leichte Probleme, meinen alltäglichen Tätigkeiten nachzugehen ❑
Ich habe mäßige Probleme, meinen alltäglichen Tätigkeiten nachzugehen ❑
Ich habe große Probleme, meinen alltäglichen Tätigkeiten nachzugehen ❑
Ich bin nicht in der Lage, meinen alltäglichen Tätigkeiten nachzugehen ❑

SCHMERZEN / KÖRPERLICHE BESCHWERDEN

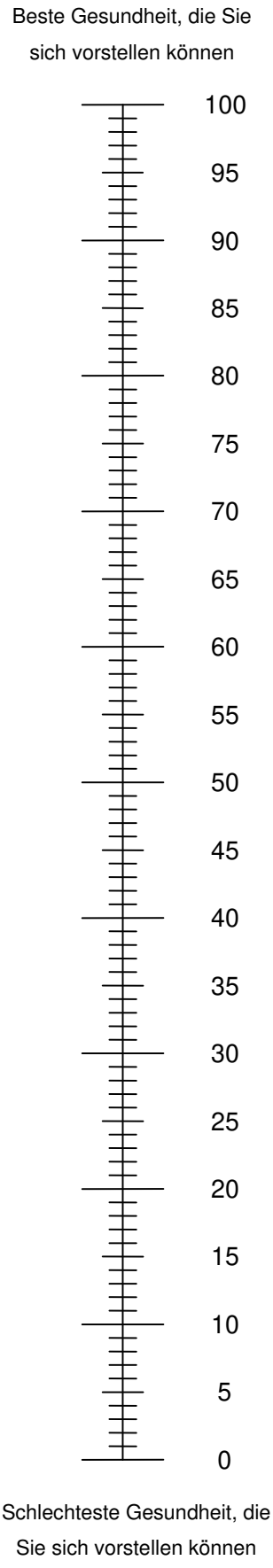
Ich habe keine Schmerzen oder Beschwerden ❑
Ich habe leichte Schmerzen oder Beschwerden ❑
Ich habe mäßige Schmerzen oder Beschwerden ❑
Ich habe starke Schmerzen oder Beschwerden ❑
Ich habe extreme Schmerzen oder Beschwerden ❑

ANGST / NIEDERGESCHLAGENHEIT

Ich bin nicht ängstlich oder deprimiert ❑
Ich bin ein wenig ängstlich oder deprimiert ❑
Ich bin mäßig ängstlich oder deprimiert ❑
Ich bin sehr ängstlich oder deprimiert ❑
Ich bin extrem ängstlich oder deprimiert ❑

- Wir wollen herausfinden, wie gut oder schlecht Ihre Gesundheit HEUTE ist.
- Diese Skala ist mit Zahlen von 0 bis 100 versehen.
- 100 ist die beste Gesundheit, die Sie sich vorstellen können.
  0 (Null) ist die schlechteste Gesundheit, die Sie sich vorstellen können.
- Bitte kreuzen Sie den Punkt auf der Skala an, der Ihre Gesundheit HEUTE am besten beschreibt.
- Jetzt tragen Sie bitte die Zahl, die Sie auf der Skala angekreuzt haben, in das Kästchen unten ein.

IHRE GESUNDHEIT HEUTE = ☐



Appendix 2

SOP No. 16

Title: DNA isolation from blood/ tissue/ cell lysates	
Version: 4	**Date:.16.04.18**
Author: CS	
Reviewer: JH,MH	

Principle

The DNeasy Blood&Tissue Kit by Qiagen provides a silica-based DNA purification**.** After isolation approximately 500ng DNA are expected. For genotyping, usually 1500 ng qPCR product in 15 µl are required.

Materials

	company	**order no.**
1.5ml DNA Low Bind tubes	Eppendorf	0030 108.051
2 ml DNA Low Bind tubes	Eppendorf	0030.108.078
Beads silica-zirconia	Roth	11079105
DNeasy Blood&Tissue Kit	Qiagen	69504
Ethanol 100% p.A.	VWR	20821.330
FastPrep 24 homogenizer	MP Biomedicals	
Microtubes	Sarstedt	72.693.465
PBS	Merck	L 182-50
PCR-Filter Tips	Sarstedt	
Proteinase K	Qiagen	19131
RNase A (17.500U)	Qiagen	19101
Trypan Blue solution 0.4%	Sigma RT	T8154
Trypsin/EDTA (TRY)	Sigma Freezer 2	T3924

Solutions

1. **PBS solution** — C_{final} 0.995 %
   47.75 g PBS
   dissolve in 5 l pure dest. H_2O, autoclave, MHD 2-3 month

2. **AW1 (component of the DNeasy kit)**
   Add EtOH p.A. according to the volume stated on the flask

3. **AW2 (component of the DNeasy kit)**
   Add EtOH p.A. according to the volume stated on the flask

4. **ATL & AL buffer ready to use** (precipitates? > 56°C until disintegration)

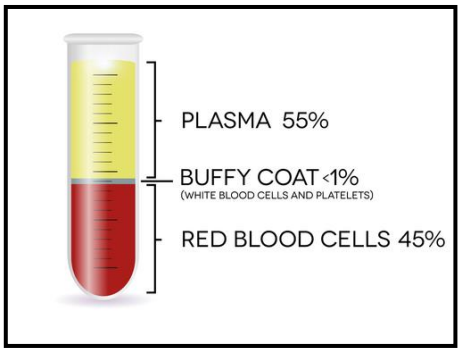
5. **Ethanol/AL mixture (only for tissue)** per sample mix 200µl EtOH abs.
   + 200µl AL buffer

Sampling:

Wear your S2 lab coat and gloves. Clean work surface with RNase Zap.
Preheat the thermo shaker with the according insert (1.5 ml or 2 ml) to 56°

Whole blood fresh:

1. Centrifuge 2.500g/rcf, 10min, RT
2. Remove plasma
3. Use 100µl of the buffy coat for isolation



Whole blood frozen after centrifugation:

1. Defrost on ice (Attention: 9ml tubes need 2h)
2. Don't shake or vortex
3. Remove plasma
4. Slowly collect buffy coat (grey slurry) in a swirling motion
5. Use 100µl of the buffy coat for isolation

Whole blood frozen without centrifugation:

1. Defrost on ice (Attention: 9ml tubes need 2h)
2. Do not centrifuge!
3. Use 200µl whole blood from the upper third for isolation

Fat tissue:

1. Weight approximately 200 mg **Beads** in a microtube
2. Fill tubes in a bag, autoclave & put bag under the UV cleaner for 5 h
3. Transfer fat tissue (max. 100 mg > 50ng/µl, more overload the column) in a microtube
4. Label tubes on top & side
5. Add **180µl ATL Puffer** & close tubes and seal with parafilm
6. Fast prep samples (MP2x24, Speed 6, 30s) > store samples 30 seconds on ice
7. Repeat fast prep for 2 more times (between samples on ice)
8. Centrifuge 12.000g/rcf, 10 min, 4°C
9. Transfer lower phase with an insulin syringe in a new 2ml DNA LoBind tube (don´t transfer fat!) alternative use long 200 µl Sorenson tips

Cells (PACs 1x T75er flask = 2x10^6 / maximum 5x10^6 cells):

1. Aspirate culture medium & wash cells once with PBS
2. Add 1 ml **TRY**/T75 flask directly on to the surface
3. Incubate for 5-7 min (max. 10 min) at 37°C
4. Further detach cells by agitating the culture flask
5. Check detachment of cells (by microscope)
6. Add 9 ml pre-warmed proliferation medium (PM see SOP 5)
7. Transfer everything to a 50 ml tube
8. Centrifuge 300 g/rcf / RT/ 10 min
9. Gently resuspend cells in 200µl PBS in a new 1.5 ml DNA LoBind tube
10. Freeze at -80°C until further procedure

Lysis:

1. Heat up Eppendorf thermos block 1.5 ml tubes, 56°C
2. Pipette components into a 1.5 ml DNA LoBind tube:

	Buffy coat	**Whole blood**	Cell lysate	Tissue
Sample	100 µl	200 µl	200µl in PBS	lower phases
Proteinase K	20 µl	40 µl	20 µl	20 µl
RNase A	4 µl	8 µl	4 µl	
PBS	100 µl	200 µl	/	
	Vortex & incubate 2 min RT			
AL-buffer	200 µl	400 µl	200 µl	
Vortex & incubate 10 min 56°C				
RNase A				4 µl
AL-buffer				200 µl vortex
EtOH	200 µl	400 µ	200 µl	200 µl
Vortex & transfer everything onto the spin columns				
Spin columns	one	**two**	one	one
Spin through ≥ 6.000g/rcf (8.000 rpm), 1min, RT				

After lysis, cool down the thermo shaker to 37°C and pre heat AE buffer

Washing

1. Discard flow through
2. Transfer spin column on a new collection tube (liquids transfer in waste bottle)
3. Add 500µL AW1 buffer / column
4. Centrifuge :≥ 6.000g/rcf (8.000rpm), 1min, RT
5. Transfer spin column on a new collection tube (liquids transfer in waste bottle)
6. Wash column with 500µL AW2 buffer
7. Centrifuge: **14.000g/rcf,** 3min, RT
8. Transfer spin column on a new DNA LoBind tube

Elution for 100µl buffy coat & tissue; one spin column

1. Add 60µL AE-buffer (**37°C for BC / 50°C for tissue**) directly (!) on the membrane
2. Centrifuge: 8.000g, 1min, RT (place the lid of the tube on the rotor)
3. Add 40µL AE-buffer directly (!) on the membrane
4. Centrifuge: 8.000g/rcf, 1min, RT (place the lid of the tube on the rotor)

Elution for 200µl whole blood; two spin columns

1. Transfer the first and second spin column onto two separate DNA LoBind tubes
2. Add 60µL AE-buffer (37 °C) directly (!) on the membrane of column 1
3. Centrifuge: 8.000g/rcf, 1min, RT (place the lid of the tube on the rotor)
4. Add 40µL AE-buffer (37 °C) directly (!) on the membrane of column 1
5. Centrifuge: 8.000g/rcf, 1min, RT (place the lid of the tube on the rotor)
6. Use the eluate from column 1 to eluate column 2 in the same manner, this approximately doubles your yield

Quality Control (Tecan or Nano Drop)

1. Use the Tecan photometer with the NanoQuant Plate (260/280 & (260/230))
2. In the software select the applications tab & choose dsDNA as sample type
3. Blank with AE-buffer
4. Use a sample volume of 1.5 µl and measure in duplicates
5. When highly accurate concentrations are required (e.g. array based analysis) a fluorescent dye should be used (e.g. PicoGreen or Qubit)
6. For the rs1421085, Simple Probe (SOP No.21) assay ~50 ng are required!
7. Samples can be stored at -4°C for 2 month / long time -20°C

References

[1] Qiagen DNeasy handbook
[2] Thermo fisher nucleic acid quantitation and quality control

Appendix 3

Entry	Co-regulation Genes	Proteins	Definition	Molecular function	Bilogical Process	Ligand	Links
1	**ITGB1**	Integrin beta-1	Receptor	*Actin binding *integrin binding	*Cell adhesion *Host-virus interaction	*Calcium, Magnesium and Metal-binding	https://www.uniprot.org/uniprot/P05556
2	**LAMA4**	Laminin subunit alpha-4		*extracellular matrix structural constituent *signaling receptor binding	*Cell adhesion *Extracellular matrix organization *Regulation of cell adhesion *Regulation of cell migration	_____	https://www.uniprot.org/uniprot/Q16363
3	**MAPK3**	Mitogen-activated protein kinase 3	Enzyme EC:2.7.11.24	*ATP binding *Identical protein binding *Kinase activity *MAP kinase activity *Phosphatase binding *Protein serine/threonine kinase activity *Scaffold protein binding	*Activation of MAPK activity *Arachidonic acid metabolic process *Caveolin-mediated endocytosis *Cell cycle *Cellular response to amino acid starvation *lipopolysaccharide-mediated signaling pathway	*ATP-binding *Nucleotide-binding	https://www.uniprot.org/uniprot/P27361 https://reactome.org/content/detail/R-HSA-73724
4	**ITGB5**	Integrin beta-5	Receptor	*Integrin binding *Signaling receptor activity	*Cell adhesion mediated by integrin *Cell-matrix adhesion *Cell migration *Endodermal cell differentiation *Extracellular matrix organization *Integrin-mediated signaling pathway *Muscle contraction *Stress fiber assembly *Transforming growth factor beta receptor signaling pathway	_____	https://www.uniprot.org/uniprot/P18084
5	**CAV1**	Caveolin-1		*Cholesterol binding *Enzyme binding *Identical protein binding *Protein binding, bridging *Protein kinase binding *Signaling receptor binding *Structural molecule activity	*Cell differentiation *Cholesterol homeostasis *Cholesterol transport *Lipid storage *Regulation of fatty acid metabolic process *Response to progesterone	_____	https://www.uniprot.org/uniprot/Q03135
6	**LAMA5**	Laminin subunit alpha-5		*Extracellular matrix structural constituent *Integrin binding	*Cell differentiation *Cell migration *Cell population proliferation *Cell recognition	_____	https://reactome.org/content/detail/R-HSA-215956 https://www.uniprot.org/uniprot/Q03135
7	**LAMB2**	Laminin subunit beta-2		*Extracellular matrix structural constituent *Integrin binding	*Cell adhesion *Cellular protein metabolic process	_____	https://www.uniprot.org/uniprot/P55268
8	**ARPC2**	Actin-related protein 2/3 complex subunit 2		*Actin filament binding *Structural constituent of cytoskeleton	*Actin filament polymerization *Actin polymerization-dependent cell motility *Arp2/3 complex-mediated actin nucleation *Membrane organization	_____	https://www.uniprot.org/uniprot/O15144
9	**ACTN4**	Alpha-actinin-4		*Actin binding *Actin filament binding *Integrin binding *Nuclear hormone receptor binding	*Actin filament bundle assembly *Protein transport	*Calcium and Metal-binding	https://www.uniprot.org/uniprot/O43707
10	**ILK**	Integrin-linked protein kinase	Enzyme EC:2.7.11.1P	*ATP binding *Integrin binding *Protein kinase binding	*Cell-matrix adhesion *Cell population proliferation *Integrin-mediated signaling pathway *Regulation of actin cytoskeleton organization	*ATP-binding *Nucleotide-binding	https://www.uniprot.org/uniprot/Q13418
11	**TLN1**	Talin-1		*Actin filament binding *Integrin binding *Structural constituent of cytoskeleton *Vinculin binding	*Cortical actin cytoskeleton organization *Cytoskeletal anchoring at plasma membrane *Integrin activation *Integrin-mediated signaling pathway	_____	https://reactome.org/content/detail/R-HSA-350713 https://www.uniprot.org/uniprot/Q9Y490
12	**LAMC1**	Laminin subunit gamma-1		*Extracellular matrix constituent conferring elasticity *Extracellular matrix structural constituent	*Cell adhesion *Cell migration *Cellular protein metabolic process *Extracellular matrix disassembly *Extracellular matrix organization *Positive regulation of epithelial cell proliferation	_____	https://www.uniprot.org/uniprot/P11047
13	**COL15A1**	Collagen alpha-1(XV) chain		*Developmental protein	*Cell adhesion *Cell differentiation *Extracellular matrix organization	_____	https://www.uniprot.org/uniprot/P39059
14	**COL5A2**	Collagen Type V Alpha 2 Chain		*Extracellular matrix structural constituent	*Cellular response to amino acid stimulus *Extracellular matrix organization	*Calcium and Metal-binding	https://www.uniprot.org/uniprot/P05997
15	**ACTN1**	Alpha-actinin-1		*Actin filament binding *Integrin binding *Vinculin binding	*Actin filament bundle assembly *Actin filament network formation *Actin filament organization	*Calcium and Metal-binding	https://www.uniprot.org/uniprot/P12814

16	**ITGA5**	Integrin alpha-5	Receptor	*Integrin binding *Epidermal growth factor receptor binding	*Cell adhesion *Extracellular matrix organization *Endodermal cell differentiation *Cell adhesion mediated by integrin	*Calcium and Metal-binding	https://www.uniprot.org/uniprot/P08648
17	**COL8A1**	Collagen alpha-1(VIII) chain		*Extracellular matrix structural constituent	*Extracellular matrix organization *Cell adhesion *Endodermal cell differentiation	____	https://www.uniprot.org/uniprot/P27658
18	**ITGA1**	Integrin alpha-1	Receptor	*Collagen binding *Protein phosphatase binding *Signaling receptor binding	*Cell-matrix adhesion *Extracellular matrix organization *Integrin-mediated signaling pathway	*Calcium, magnisum and Metal-binding	https://www.uniprot.org/uniprot/P56199
19	**COL6A2**	Collagen alpha-2(VI) chain		*Extracellular matrix structural constituent conferring tensile strength	*Response to glucose *Extracellular matrix organization *Cell adhesion	_____	https://www.uniprot.org/uniprot/P12110
20	**COL3A1**	Collagen alpha-2(VI) chain		*Extracellular matrix structural constituent conferring tensile strength	*Response to glucose *Extracellular matrix organization *Cell adhesion	_____	https://www.uniprot.org/uniprot/P02461
21	**COL6A1**	Collagen alpha-1(VI) chain		*Extracellular matrix structural constituent conferring tensile strength	*Cell adhesion *Extracellular matrix organization *Cellular response to amino acid stimulus *Endodermal cell differentiation	_____	https://www.uniprot.org/uniprot/P12109
22	**ITGA11**	Integrin alpha-11	Receptor	*Collagen binding *Collagen binding involved in cell-matrix adhesion *Collagen receptor activity	*Cell adhesion *Extracellular matrix organization *Cell-matrix adhesion *Cell adhesion mediated by integrin	*Calcium, magnisum and Metal-binding	https://www.uniprot.org/uniprot/Q9UKX5
23	**GRAP**	GRB2-related adapter protein		*SH3/SH2 adaptor activity	*Cell-cell signaling *Ras protein signal transduction	_____	https://www.uniprot.org/uniprot/Q13588
24	**ARPC5L**	Actin-related protein 2/3 complex subunit 5-like protein		*Actin filament binding	***Arp2/3*** **complex-mediated actin nucleation** **(which is involved in regulation of actin polymerization)** *Cell migration	_____	https://www.uniprot.org/uniprot/Q9BPX5
25	**COL18A1**	Collagen alpha-1(XVIII) chain		*Extracellular matrix structural constituent *Extracellular matrix structural constituent conferring tensile strength	*Cell adhesion *Extracellular matrix organization *Endothelial cell morphogenesis	*Metal-binding and Zinc	https://www.uniprot.org/uniprot/P39060
26	**FLNA**	Filamin-A		*Actin filament binding	*Actin crosslink formation *Actin cytoskeleton reorganization *Positive regulation of actin filament bundle assembly *Positive regulation of integrin-mediated signaling pathway *Protein localization to cell surface	_____	https://www.uniprot.org/uniprot/P21333 https://reactome.org/content/detail/R-HSA-210019
27	**COL16A1**	Collagen alpha-1(XVI) chain		*Extracellular matrix structural constituent *Extracellular matrix structural constituent conferring tensile strength *Integrin binding	*Cell adhesion *Extracellular matrix organization *Cell adhesion mediated by integrin *Cellular response to amino acid stimulus *Integrin-mediated signaling pathway *Integrin activation	_____	https://www.uniprot.org/uniprot/Q07092
28	**PIK3C2B**	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	Enzyme EC:2.7.1.154	*ATP binding *Lipid kinase activity *Phosphatidylinositol binding	*Cell migration *Protein kinase B signaling	*ATP-binding *Nucleotide-binding	https://www.uniprot.org/uniprot/O00750
29	**SRC**	Proto-oncogene tyrosine-protein kinase Src	Non- Receptor	*ATP binding *Enzyme binding *Insulin receptor binding *Integrin binding	*Cell differentiation *Cellular response to fatty acid *Cellular response to insulin stimulus *Integrin-mediated signaling pathway *Cellular response to lipopolysaccharide	*ATP-binding *Nucleotide-binding	https://www.uniprot.org/uniprot/P12931
30	**RAP2A**	Ras-related protein Rap-2a		*GDP binding *GTPase binding *GTP binding	*Actin cytoskeleton reorganization *Cellular protein localization *Positive regulation of protein phosphorylation	*ATP-binding *Nucleotide-binding	https://www.uniprot.org/uniprot/P10114
31	**LIMS2**	LIM and senescent cell antigen-like-containing domain protein 2		*Metal ion binding	*Positive regulation of integrin-mediated signaling pathway	*Metal-binding and Zinc	https://www.uniprot.org/uniprot/Q7Z4I7
32	**FKBP11**	Peptidyl-prolyl cis-trans isomerase FKBP11	Enzyme EC:5.2.1.8	*Isomerase, Rotamase	*PPIases accelerate the folding of proteins during protein synthesis.	_____	https://www.uniprot.org/uniprot/Q9NYL4
33	**SLC2A4**	**GLUT4** Solute carrier family 2, facilitated glucose transporter member 4	Transporter protein	*Glucose transmembrane transporter activity	*Brown fat cell differentiation *Carbohydrate metabolic process *Cellular response to insulin stimulus *Glucose homeostasis *Glucose import e *Glucose import in response to insulin stimulus *Glucose transmembrane transport *Sugar transport	_____	https://www.uniprot.org/uniprot/Q9NR83

34	**LEPR**	Leptin receptor	Receptor	*Cytokine binding *Cytokine receptor activity *Identical protein binding *Leptin receptor activity *Peptide hormone binding *Transmembrane signaling receptor activity	*Cell surface receptor signaling pathway *Cholesterol metabolic process *Energy homeostasis *Energy reserve metabolic process *Glucose homeostasis *Leptin-mediated signaling pathway *Response to leptin	_____	https://www.uniprot.org/uniprot/P48357
35	**PPAR-γ**	Peroxisome proliferator-activated receptor gamma	Nuclear receptor	*Activating transcription factor binding *Alpha-actinin binding *Arachidonic acid binding *DNA binding *Estrogen receptor binding *Fatty acid binding *Lipid biding *Peptide biding	*Cell differentiation *Cellular response to insulin stimulus *Cellular response to retinoic acid *Cellular response to vitamin E *Fatty acid metabolic process *Fatty acid oxidation *Glucose homeostasis *Lipid metabolic process *Lipoprotein transport *Long-chain fatty acid transport *White fat cell differentiation	*Metal-binding and Zinc	https://www.uniprot.org/uniprot/P37231
36	**C/EBPβ**	CCAAT/enhancer-binding protein beta	transcription factor	*DNA binding *Glucocorticoid receptor binding *Kinase binding	*Brown fat cell differentiation *Cellular response to amino acid stimulus *Transcription regulation	_____	https://www.uniprot.org/uniprot/P17676
37	**C/EBPα**	CCAAT/enhancer-binding protein alpha	transcription factor	*DNA binding *Development protein *Kinase binding	*Brown fat cell differentiation *Fat cell differentiation *Transcription regulation *Cholestrol metabolic process *Glucose homeostasis *White fat cell differentiation	_____	https://www.uniprot.org/uniprot/P49715
38	**APLN**	Apelin		*Hormone activity	*Apelin receptor signaling pathway	_____	https://www.uniprot.org/uniprot/?query=APLN&sort=score
39	**SLC7A1**	**GLUT1** Solute carrier family 2, facilitated glucose transporter member 1	Transporter protein	*Glucose transmembrane transporter activity *Identical protein binding *Kinase binding	*Glucose transmembrane transport *Lactose biosynthetic process *Regulation of glucose transmembrane transport *Regulation of insulin secretion *Response to insulin *Sugar transport	_____	https://www.uniprot.org/uniprot/P30825
40	**IRS1**	Insulin receptor substrate 1	Receptor	*Insulin receptor binding	*Respnse to insulin *Glucose homeostasis	_____	https://www.uniprot.org/uniprot/P35568
41	**ADAM12**	Disintegrin and metalloproteinase domain-containing protein 12		*Hydrolase, Metalloprotease, Protease	*Cell adhesion *Extracellular matrix organization	_____	https://www.uniprot.org/uniprot/O43184
42	**CAV1**	Caveolin-1		*Cholestrol binding *Enzyme binding	*Cell differentiation *Cholestrol homeostasis *Cholestrol transport *Lipid storage *Regulation of fatty acid metabolic process *Triglyceride metabolic process	_____	https://www.uniprot.org/uniprot/Q03135
43	**PLIN5**	Perilipin-5		*Lipase binding	*Lipid droplet organization *Lipid storage *Positive regulation of fatty acid beta-oxidation *Positive regulation of lipase *Positive regulation of lipid storage *Positive regulation of triglyceride biosynthetic process	_____	https://www.uniprot.org/uniprot/Q00G26
44	**DGAT2**	Diacylglycerol O-acyltransferase 2	Enyzme EC:2.3.1.20	*Acyltransferase, Transferase	*Cellular triglyceride homeostasis *Cholesterol homeostasis *Fatty acid homeostasis *Lipid storage *Positive regulation of triglyceride biosynthetic process *Regulation of cholesterol metabolic process *Regulation of lipoprotein metabolic process *Triglyceride biosynthetic process	_____	https://www.uniprot.org/uniprot/Q96PD7

Appendix 4

Variable			
	Male Mean±	**Female Mean±**	**P-value**
Total	**102**	**236**	
Age V0 (years)	50±11.00	49±11.10	0.441
Height (cm)	180±6.73	166±6.98	2.481
Weight (kg)	104±21.04	87±19.90	1.103
BMI (kg/m²)	32±6.02	32±6.95	0.705
Waist (cm)	110±14.69	115±15.03	5.487
Hip (cm)	111±13.08	1.0±15.70	0.022
WHR (cm)	1.0±0.06	1.0±0.06	6.447
WHtR *cm)	1.0±0.08	41±0.09	0.241
Fat (%)	28±6.75	36±7.81	9.685
Fat mass (kg)	30±12.33	50±14.21	9.311
Lean mass (kg)	73±10.44	127±6.72	9.408
BP systol R (mm Hg)	135±16.39	127±14.77	0.000
BP diastol R (mm Hg)	82±9.57	79±8.03	0.001
BP systol L (mm Hg)	132±15.86	135±103.62	0.688
BP diastol L (mm Hg)	83±10.10	78±7.42	0.001
SU_ph	6.0±0.79	6.0±0.95	0.026
ERY (Mio/µl)	5.0±0.35	5.0±0.30	1.407
HB (g/dL)	15±1.01	14±0.95	2.155
HBE (pg)	30±1.40	30±1.62	0.001
MCV (fl)	88±3.86	88±4.03	0.877

HKT (%)	45±2.63	41±2.45	7.739
MCHC (g/dl)	35±0.888	34±0.88	4.612
THROMBOTSD (Tausend/µl)	231±48.40	262±55.16	4.619
LEUKO (1/µl)	6±1.54	6±1.39	0.641
GOT_IFCC (U/l)	30±11.83	23±8.23	9.866
GPT_IFCC (U/l)	43±25.62	27±16.63	1.440
GGT_IFCC (U/l)	43±40.16	26±21.60	9.386
SODIUM (mmol/l)	140±2.65	139±2.74	0.286
POTASSIUM (mmol/l)	6.0±4.83	5.0±0.40	0.225
CAM (mmol/l)	2.0±0.10	2.0±0.11	0.585
PHOS (mg/dl)	3.0±0.46	3.0±0.50	2.744
Glc_0 (mg/dl)	100±23.14	93±12.26	0.013
Glc_30 (mg/dl)	170±39.74	156±35.99	0.004
Glc_60 (mg/dl)	165±66.09	146±51.17	0.020
Glc_90 (mg/dl)	134±63.05	124±42.53	0.251
Glc_120 (mg/dl)	109±55.748	106±32.186	0.794
Ins 0 (pmol/l)	97±91.551	72±59.907	0.010
Ins 30 (pmol/l)	587±483.50	503±731.38	0.133
Ins 60 (pmol/l)	727±620.73	662±562.21	0.236
Ins 90 (pmol/l)	642±662.01	568±540.45	0.328
Ins 120 (pmol/l)	403±475.70	416±436.85	0.815
ISI Matsuda (mg/dl,µU/ml)	17±15.37	20±17.93	0.063
Insulinogenic Index	152±277.25	109±322.30	0.231
Disposition Index	1753±2920.45	1173±9812.38	0.416

NEFA 0 (µmol/l)	381±165.11	511±228.58	4.112
NEFA 30 (µmol/l)	252±110.64	290±142.79	0.010
NEFA 60 (µmol/l)	140±73.35	111±68.44	0.000
NEFA 120 (µmol/l)	70±41.68	51±38.39	0.000
HBA1C (mmol/mol)	6±0.78	6±0.44	0.112
CHOMG (mg/dl)	207±37.99	210±36.33	0.506
TGMG (mg/dl)	166±129.85	125±58.77	0.002
HDLMG (mg/dl)	45±10.37	59±15.59	1.736
LDLMG (mg/dl)	136±32.98	136±34.95	0.885
Albumin_S (mg/dl)	4504±266.27	4391±223.87	0.000
KREATININ (mg/dl)	1.0±0.13	1.0±0.11	1.079
Urea_S (mmol/l)	35±8.20	30±6.98	1.129
Uric acid (mg/dl)	7.0±1.18	5.0±1.28	1.138
SU_Protein (g/l)	8.0±6.60	6.0±4.25	0.002
SU_Albumin (mg/l)	2.0±3.90	2.0±3.24	0.837
Iron (µg/dl)	109±31.45	101±35.47	0.035
Ferritin (µg/dl)	218±159.24	84±81.86	5.731
CRP (mg/dl)	0.0±0.44	0.0±0.45	0.037
sum_Protein_S (g/l)	74±4.28	73±4.09	0.275
SU_Kreatinin (mg/dl)	156±76.19	103±67.38	1.103
TSH (mU/l)	19±175.77	42±281.15	0.360

Table shows Anthropometric and clinical parameters of the PLIS subjects at the TUM study center. P-value was calculated using a t-test.

Appendix 5

Variable	Females		
	Non risk Mean±	Risk Mean±	P-value
Total	36	91	
Genotype	COBLL1 rs6712203	COBLL1 rs6712203	
Age (years)	48±11.22	48 ±10.85	0,259
BMI (kg/m²)	32±7.85	32±7.21	0,828
Waist circumference (cm)	97±15.36	99±14.63	0,718
WHR (cm)	1.0±0.06	1.0±0.06	0,452
WHtR (cm)	1.0±0.09	1.0±0.08	0,911
Fat percentage (%)	39±10.18	41±7.38	0,499
Fat mass (kg)	36±16.17	36±14.69	0,882
Lean mass (kg)	51±6.76	51±7.37	0.980
BP systol R (mm Hg)	127±16.11	126±14.40	0.716
BP diastol R (mm Hg)	78±8.18	78±8.27	0.876
BP systol L (mm Hg)	130±10.75	149±166.34	0.428
BP diastol L (mm Hg)	79±7.33	70±7.09	0.147
SU_ph	6.0±1.11	6.1±0.96	0.791
ERY (Mio/µl)	5.0±0.31	4.7±0.29	0.962
HB (g/dL)	14±0.90	13.7±0.90	0.559
HBE (pg)	30±1.50	29.5±1.78	0.492
MCV (fl)	89±3.50	87.3±4.31	0.543
HKT (%)	41±2.30	40.5±2.26	0.638
MCHC (g/dl)	34±0.90	33.8±0.85	0.621

THROMBOTSD (Tausend/µl)	256±70.4	264.4±53.01	0.529
LEUKO (1/µl)	6±1.30	6.1±1.60	0.455
GOT_IFCC (U/l)	24±6.70	22.6±7.64	0.506
GPT_IFCC (U/l)	28±16	26.8±15.99	0.582
GGT_IFCC (U/l)	28±20.50	25.8±21.75	0.538
SODIUM (mmol/l)	140±2.10	139.4±3.18	0.288
POTASSIUM (mmol/l)	5.0±0.41	5.0±0.43	0.128
CAM (mmol/l)	2.0±0.10	2.3±0.11	0.229
PHOS (mg/dl)	3.0±0.41	3.4±0.53	0.372
Glc_0 (mg/dl)	95±13.90	92.7±12.10	0.460
Glc_30 (mg/dl)	155±41.10	157.5±34.75	0.783
Glc_60 (mg/dl)	142±51.70	145.4±54.7	0.774
Glc_90 (mg/dl)	123±40	123.2±45.70	0.998
Glc_120 (mg/dl)	104±25	104.5±35.15	0.951
Ins 0 (pmol/l)	70.1±55.91	73.7±68.94	0.767
Ins 30 (pmol/l)	451.4±373.71	516.7±442.34	0.408
Ins 60 (pmol/l)	635.6±568.91	714.4±619.24	0.500
Ins 90 (pmol/l)	576.7±531.80	586.7±576.02	0.926
Ins120 (pmol/l)	436.9±392.50	423.0±463.63	0.866
ISI Matsuda (mg/dl,µU/ml)	23.6±23.71	20.2±17.15	0.440
Insulinogenic Index	116.9±155.10	126.5±137.53	0.750
Disposition Index	1181.6±2669.41	1818.7±3116.88	0.258
NEFA 0 (µmol/l)	492.0±154.80	509.3±189.79	0.429

NEFA 30 (µmol/l)	293.6±155.51	289.0±145.98	0.885
NEFA 60 (µmol/l)	118.5±67.00	115.5±81.70	0.837
NEFA 120 (µmol/l)	55.4±66.40	49.4±32.93	0.660
HBA1C (mmol/mol)	5.6±0.50	5.5±0.44	0.451
CHO (mg/dl)	206.5±32.20	209.4±36.13	0.655
TG (mg/dl)	121.4±71.40	127.0±65.93	0.191
HDL (mg/dl)	57.3±14.91	59.8±16.92	0.408
LDL (mg/dl)	132.1±37.50	133.2±34.22	0.878
Albumin S (mg/dl)	4372.6±236.60	4411.8±217.95	0.384
KRATININ (mg/dl)	0.7±0.11	0.7±0.11	0.169
Urea S (mmol/l)	29.3±7.00	30.3±6.73	0.441
Uric acid (mg/dl)	5.5±1.20	5.2±1.25	0.198
SU_Protein (g/l)	5.7±4.20	6.2±4.71	0.743
SU_Albumin (mg/l)	0.8±0.30	2.0±3.63	0.060
Iron (µg/dl)	109.4±37.81	98.3±33.98	0.124
Ferritin (µg/dl)	76.2±64.50	80.8±101.48	0.755
CRP (mg/dl)	0.4±0.50	0.4±0.39	0.910
Sum Protein S (g/l)	72.5±3.90	73.6±4.36	0.148
SU Kreatinin (mg/dl)	103.4±80.40	96.4±61.63	0.642
TSH (mU/l)	2.0±1.10	29.5±263.76	0.321

Table of Study cohort. Total of 127 females (91 *COBLL1* rs 6712203 risk allele carriers/36 non-risk allele carriers)