Proteomic characterisation of diabetic pig models - diabetic treatment effect on metabolic tissues

von Erik Gunnar Emanuel Ländström

Inaugural-Dissertation zur Erlangung der Doktorwürde
(Dr.rer.biol.vet)
der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Proteomic characterisation of diabetic pig models - diabetic treatment effect on metabolic tissues

von

Erik Gunnar Emanuel Ländström

aus Göteborg, Schweden

München 2020

Aus dem
Veterinärwissenschaftlichen Department der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München
Lehrstuhl für Molekulare Tierzucht und Biotechnologie

Arbeit angefertigt unter der Leitung von:
Univ.-Prof. Dr. Eckhard Wolf

im
Labor für Funktionale Genomanalyse (LAFUGA)
Genzentrum der Ludwig-Maximilians-Universität München
Proteomics

Mentor: Dr. Thomas Frölich

Gedruckt mit Genehmigung der Tierärztlichen Fakultät der
Ludwig-Maximilians-Universität München

Dekan:	Univ.-Prof. Dr. Reinhard K. Straubinger, Ph.D
Berichterstatter:	Univ.-Prof. Dr. Eckhard Wolf
Korreferenten:	Univ.-Prof. Dr. Cornelia A. Deeg Univ.-Prof. Dr. Johannes Hirschberger Priv.-Doz. Dr. Florian M. Trefz

Tag der Promotion: 25. Juli 2020

Table of Contents

List of Figures

List of Tables

Index of Abbreviations

ABC Ammonium Bicarbonate

AC Alternating Current

ACOT4 Acyl–Coenzyme A Thioesterase 1–Like

ADA American Diabetes Association

AGT Angiotensinogen

AK2 Adenylate Kinase 2

ALDH1A1 Aldehyde Dehydrogenase 1 family, member A1

ALDH1A2 Aldehyde dehydrogenase 1 family, member A2

ALOX15 Arachnidonate 15–Lipoxygenase

AMT Aminomethyltransferase, mito-

chondrial

ANOVA Analysis of Variance

ARFGAP3 ADP ribosylation factor GTPase activating protein 3

AT Adipose Tissue

ATOX1 Copper Transport Protein ATOX1

ATP Adenosine Triphosphate

atRA all–trans–Retinoic Acid

BAT Brown Adipose Tissue

BCAA Branched-chain Amino Acids

BLAST Basic Local Alignment Search Tool

BMP7 Bone Morphogenetic Protein 7

CBR2 Carbonylreductase [NADHPH] 2

CES1 Carboxylesterase 1

CFD Complement Factor D

CIAPIN1 Anamorsin

CID Collision-Induced Disassociation

CLIC5 Chloride Intracellular Channel Protein 5

COL3A1 Collagen Alpha–1(III) chain

COX7A1 Cytochrome C Oxidase Subunit 7A1, Mitochondrial Precursor

CRISPR Clustered Regularly Interspaced Palindromic Repeats

CYP3A39 Cytochrome

P450 3A39

CYP3A4 Cytochrome P450 3A4

DC Direct Current

DDA Data-Dependent Acquisition

DHRS4 Dehydrogenase/Reductase SDR Family Member 4

DLST Component of 2-Oxoglutarate Dehydrogenasecomplex, Mitochondrial Precursor

DM Diabetes Mellitus

DR Direct Repeat

DTT Dithiothreitol

EEF1D Elongation Factor 1-delta

EIF3C Eukaryotic Translation Iniation Factor 3 Subunit C

EIF4E2 eukaryotic translation iniation factor 4E type 2

ER Endoplasmic Reticulum

ESI Electrospray Ionization

FBP1 Fructose-1,6 Bisphosphatase 1

FDR False Discovery Rate

FFA Free Fatty Acid

FOXO1 Forkhead Box Protein O1

FPR False Positive Rate

FWER Family–Wise Error Rate

G6P Glucose-6-Phospate

GALM Aldose 1-Epimerase

GCK Glucokinase
GIP Glucose-dependent Insulinotropic Polypeptide
GIPR Glucose-dependent Insulinotropic Polypeptide Receptor
GLP1 Glucagon-Like Peptide 1
GLUT Glucose Transporter
GMDS GDP-Mannose-4,6-Dehydratase
GO Gene Ontology
GOBP Gene Ontology Biological Process
GOCC Gene Ontology Cellular Component
GOMF Gene Ontology Molecular Function
GSEA Gene Set Enrichment Analysis

HADHA Trifunctional Enzyme Subunit Alpha, Mitochondrial
HADHB Trifunctional Enzyme Subunit Beta, Mitochondrial
HIF1 Hypoxia-Induced Factor 1
HILIC Hydrophilic Interaction Liquid Chromatography
HK1 Hexokinase 1
HMGCS2 Hydroxymethylglutaryl-CoA Synthase, Mitochondrial
HNF1A Hepatocyte Nuclear Factor 1A
HNF4A Hepatocyte Nuclear Factor 4A
HPLC High Perfor-

mance Liquid Chromatography

HSD Honest Significant Differences

HSP90AA1 Heat Shock Protein 90–Alpha

IAH1 Isoamyl Acetate-Hydrolyzing Esterase 1 Homolog

ICAT Isotope-Coded Affinity Tag

INS Insulin

ITRAQ Isobaric Tag for Relative and Absolute Quantitation

KEGG Kyoto Encyclopedia of Genes and Genomes

KO Knockout

KRT19 Keratine Type I Cytoskeletal 19

KRT8 Keratine Type II Cytoskeletal 8

L2FC Log2 Fold Change

LAMC1 laminin subunit gamma-1

LFQ Label Free Quantification

LIT Linear Ion Trap

LMAN1 Lectin Mannose Binding 1

LysC Endoproteinase LysC

m/z Mass-to-charge Ratio

MALDI Matrix-Assisted Laser Desorption/Ionization

MAOA Amine Oxidase [Flavine Containing] A

MAR Missing At Random

MAT Mesenteric Adipose Tissue
MCAR Missing Completely At Random
MFAP2 Microfibrillar-Associated Protein 2
MIDY Mutant INS Gene–induced Diabetes of Youth
MNAR Missing Not At Random
MODY Maturity Onset Diabetes of the Young
mRNA Messenger Ribonucleic Acid
MS Mass Spectrometry
MTP Mitochondrial Trifunctional Protein

NES Normalized Enrichment Score
NXN Nucleoredoxin

OAT Ornithine Aminotransferase
OGN Mimecan
OPLAH 5-Oxoprolinase

PC Principal Component
PCA Principal Component Analysis
PCNA Proliferating Cell Nuclear Antigen
PCOLCE Procollagen C-Endopeptidase Enhancer 1
PCYOX Prenylcysteine Oxidase 1
PDX1 Duodenal Homeobox 1
PEF1 Peflin
PHGDH D-3-Phosphoglycerate Dehydrogenase
PND Permanent Neonatal Diabetes

PRDM16 Positive Regulatory Domain Containing 16
PRKAR2A cAMP-Dependent Protein Kinase Type II-α Regulatory Subunit
PSM Peptide Spectrum Match
PSMD5 26S Proteasome Non–ATPase Regulatory Subunit 5

QMF Quadropole Mass Filter
QTL Quantitative Trait Locus

RA Retinoic Acid
RAR Retinoic Acid Receptor
RARE Retinoic Acid Response Element
RBX1 E3 Ubiquitin-Protein Ligase RBX1
RDH16 Retinol Dehydrogenase 16
RNA Ribonucleic Acid
ROS Reactive Oxygen Species
RPLC Reversed Phase Liquid Chromatography
RXR Retinoic X Receptor

S100A1 Protein S100A1
SAX Strong-Anion Exchange
SCAT Subcutaneous Adipose Tissue
SCP2 Non–Specific Lipid–Transfer Protein 2
SCX Strong-Cation Exchange

SDC Sodium Deoxycholate
SDS Sodium Dodecyl Sulfate
SILAC Stable Isotope Labeling by/with Amino acids in Cell culture
STEAP4 Metalloreductase STEAP4

T1DM Type 1 Diabetes Mellitus
T2DM Type 2 Diabetes Mellitus
TCA Trichloroacetic Acid
TCEP Tris(2-Carboxyethyl)Phosphine
TF Transcription factor
TFA Trifluoroacetic Acid
TKT Transketolase
TMT Tandem Mass Tag
tRNA Transfer RNA

UCHL3 Ubiquitin Carboxyl-Terminal Hydrolase 3
UCP1 Uncoupling Protein 1
UPR Unfolded Protein Response

VAT Visceral Adipose Tissue

WAT White Adipose Tissue
WT Wildtype

Chapter 1

Introduction

Diabetes mellitus (DM) is a metabolic syndrome with impaired glucose homeostastis caused by defective insulin action or pancreatic β-cell insulin secretion. DM is one of the biggest global health challenges, with worldwide prevalence numbers increasing. The number of people diagnosed with DM is expected to reach 629 million by 2045, similarly the cost for treating these patients is also predicted to increase (**?**). Therefore, there is a need to study the different forms of DM and to develop and test new forms of efficient treatments for the different subclasses.

The two major forms of DM are type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM). T1DM is caused by autoimmune destruction of the pancreatic β-cells, disrupting insulin secretion, and is the most common DM form in children and adolescents. Most T1DM patients must be treated with lifelong insulin therapy (**?**). T2DM is the most prevalent type of DM and it is characterized by insulin resistance, which increases the demand for insulin in target tissues. The increased demand for insulin can lead to progressive β-cell destruction. T2DM is often linked to obesity and metabolic syndrome (**?**).

Animal models allow researchers to control genetic and environmental factors *in vivo* that influences the progression of DM, giving researchers essential new insights to complications and treatment of the disease in human patients. As such, animal models have provided researchers with invaluable contributions to the understanding of the pathogenesis and treatment of DM (**?**). The most commonly used animal models for studying DM are mice and rat models, but other animals with human-like characteristics have also been used (**???**). There are significant differences between rodents and humans (e.g. size and metabolism) and

data derived from rodent models can be insufficient in translating scientific findings from the animal model to human pathologies (**?**).

Therefore, focus has been shifting to larger animal models including *Sus scrofa* in order to bridge the translational gap between rodent models and humans (**?**). This shift has been enabled by the sequencing of the pig genome and advances in new gene editing tools such as zinc finger nucleases and CRISPR/CAS method (**?**). Pigs are relevant models for translational research in different medical fields, as pigs share many similarites with human anatomy, physiology and metabolism; e.g. similarites in pancreas morphology and metabolic regulation makes pigs suitable for diabetes research (**?**). Furthermore, pigs have a relatively short generation span, early sexual maturity, and on average 10 offspring per litter (**?**), which is beneficial for effective biomedical research.

Proteomics is the study of all the proteins that control and catalyse all cellular processes at a given time (**?**). Nowadays mass spectrometry is the most commonly used method for studying the proteome of cells and tissues. As mRNA and protein expression levels do not always show perfect correlation (**?**), it is neces-

sary to study the proteome in order to get a complete picture of the functional state of the tissue. As such, proteomics as emerged as a powerful tool that has provided insights into complex biological processes and phenotypes (**?**).

In this doctoral thesis, proteomic differences between treated diabetic pigs and control groups in selected metabolic tissues have been studied using mass spectrometry in order to better understand the effects of diabetes treatments and to establish pigs as a suitable animal model for translational diabetes research.

In the first study, proteomic differences in mesenteric and subcutaneous adipose tissue between suboptimally insulin treated two year old female MIDY (a large animal model of neonatal diabetes) and WT pigs have been compared. An increase in expression of retinol dehydrogenase 16 (RDH16) in both types of adipose tissue was observed in MIDY pigs, suggesting that retinol metabolism and retinoic acid induced transcriptional regulation is an important adaptation to low insulin levels.

In the second study, transgenic pigs, a large animal model for T2DM, carrying a dominant-negative mutation in the glucose-dependent insulinotropic polypep-

tide receptor ($GIPR^{dn}$ pigs) have been used to study the effects of liraglutide (a glucagon-like peptide 1 (GLP1) receptor agonist). Liraglutide stimulates üancreatic β-cell insulin secretion (**?**) and the $GIPR^{dn}$ pigs were treated with either liraglutide or placebo, and the proteome differences between the two treatment groups were examined. Mitochondrial hydroxymethylglutaryl-CoA synthase (HMGCS2) was detected as less abundant in liver from liraglutide treated pigs. HMGCS2 is the rate-limiting enzyme for ketogenesis (**?**), suggesting that liraglutide induced insulin secretion inhibits ketogenesis in diabetic pigs.

Chapter 2

Literature Review

2.1 Diabetes

In 2017, 425 million people were estimated to be affected by diabetes and the prevalence of diabetes is expected to increase by 48% until 2045 (**?**). Similarly the cost of treating diabetes is also predicted to increase in the near future (**?**) there is thus a need to develop more effective and efficient treatments.

Hyperglycemia, high concentrations of blood sugar, is the fundamental hallmark of diabetes, a highly het-

erogeneous disease, that is caused by relative or absolute loss of insulin. The American Diabetes Association (ADA) divides diabetes into four subgroups based on the etiological characterization of the disease (**?**):

- Class I is type 1 diabetes mellitus (T1DM), which is characterized by an absolute loss of insulin production due to autoimmune-mediated destruction of the pancreatic β -cells. Type 1 diabetes accounts for 5-10% of diabetic patients.

- Class 2 is type 2 diabetes mellitus (T2DM), the most common form of the disease. Type 2 diabetes is not an autoimmune disorder, instead it is usually caused by insulin resistance coupled with a relative insulin deficiency due to a progressive loss of pancreatic β-cells. This form of diabetes is often linked with obesity.

- Class 3 is gestational diabetes, which is diabetes diagnosed during the second or third trimester of diabetes.

- Class 4 is other specific types of diabetes. This class covers a wide range of different causes in-

cluding genetic defects causing lower insulin secretion or insulin action, diseases in the exocrine pancreas, drug or chemical induced diabetes, and β-cell destruction caused by virus infections. Monogenic diabetes belongs to this class, which accounts for 1-5% of all diabetes cases (**?**).

Other ways of classifying the different diabetic diseases exist as well, where T1DM and T2DM can be further classified into subgroups based on severity of the disease, the time of disease onset, etc (**?**). Here, permanent neonatal diabetes (PND) and T2DM will be described in more detail as they are of primary concern for this thesis.

2.1.1 Obesity, metabolic syndrome, and type 2 diabetes mellitus

Obesity has consistently increased over the last decades, and in 2015, 604 million adults were considered to suffer from obesity (**?**). Obesity may now be the leading cause of premature mortality in the world (**?**). Visceral obesity is linked to metabolic disorders such as

insulin resistance and dyslipidemia, and obese individuals are more prone to other adverse health effects such as T2DM and cancer (**?**). The metabolic syndrome, which is defined as having 3 or more of the risk factors obesity, high triglycerides, high cholesterol, hyperglycemia, and hypertension, is correlated with increased risk of cardiovascular disease (**??**).

Subcutaneous adipose tissue (SCAT) depots are generally protective against the metabolic syndrome, whereas visceral adipose tissue (VAT) depots are associated with the metabolic syndrome (**?**).

2.1.1.1 Treatments of diabetes

T1DM needs to be treated with lifelong insulin injections together with blood glucose monitoring (**??**). Unlike T1DM, there are many different ways to treat T2DM beyond insulin injections (**?**). Treatment options vary by severity of the diseases and range from lifestyle changes (such as changes to diet and exercise) and treatment of obesity to treatments with for example metformin (decreases fasting blood sugar levels), sulfonylureas (stimulates pancreatic β-cell insulin secretion) and insulin injections (**??**).

Glucagon-like peptide 1 (GLP1) receptor agonists have been identified as an effective treatment of T2DM (**??**). GLP1 is a peptide hormone that is secreted from the intestine in response to food intake (**?**) and stimulates pancreatic insulin secretion while inhibiting pancreatic glucagon release, which reduces blood sugar levels (**?**). The use of GLP1 is limited by its short half-life (1-2 min), therefore many GLP1 receptor agonists have been developed that mimic the effects of the hormone while having a longer action duration (**?**). There are short term and long term GLP1 receptor agonists (**??**). One long lasting GLP1 receptor agonist is liraglutide, which has a 97% homology to GLP1, which should be administered daily by subcutaneous injections (**?**). Compared to other treatments, treatment with GLP1 receptor agonists reduces blood glucose levels while having a reduced risk of hypoglycemia (**?**).

2.1.2 Monogenic causes of diabetes

As mentioned above, diabetes of class 3 includes a wide variety of mechanisms that cause diabetes. Maturity onset diabetes of the young (MODY) is a form of class

3 diabetes that is caused by monogenic defects in pancreatic β-cell function. Mutations in many different genes are known to cause MODY of varying degrees of severity and it is inherited in an autosomal dominant heterozygous fashion (**??**). They are characterized by dysfunctional insulin secretion often with no effect on insulin action (**??**). Patients with MODY are defined by early onset of the disease and normally lack features of the metabolic syndrome or autoimmunity (**??**). Common genes with mutations known to cause MODY include: glucokinase (*GCK*) genes, transcription factors (TFs), hepatocyte nuclear factor 1A and 4A (*HNF1A* and *HNF4A*) and pancreatic and duodenal homeobox 1 (*PDX1*) (**??**).

Another type of monogenic diabetes is permanent neonatal diabetes (PND). Permanent neonatal diabetes is defined by an extremely early onset of diabetes, often within the first months of life (**?**). Mutations in either the sulfonylurea receptor/Kir6.2 complex or in the insulin gene are the primary causes of permanent neonatal diabetes(**?**). The latter form is called Mutant *INS*-gene induced Diabetes of youth (MIDY) and is caused by mutations in the proinsulin gene resulting in an accumulation of misfolded insulin and endoplas-

mic reticulum(ER) stress, which ultimately leads to the destruction of the pancreatic β-cells (**??**).

Because many genes causing monogenic diabetes have also found to be implicated in T1DM or T2DM (or both), the study of monogenic diabetic pathologies can also give valuable insight about the more common and complex variants of the diabetic condition (**?**).

2.1.2.1 Insulin folding and ER stress

Protein misfolding and ER stress were also found to be a factor for developing T1DM and T2DM (**?**). Up to 50% of total protein synthesis in pancreatic β-cells is made up of insulin production and a higher requirement for synthesis and folding of proinsulin makes the β-cell sensitive to ER stress (**?**). Proper folding and post-translational modification of the insulin protein is thus essential for proper cell functionality (**??**).

The insulin gene encodes a single transcript called preproinsulin which consists of four different domains. Upon translocation to the ER membrane, proinsulin is created by cleavage of the signal peptide of the preproinsulin mRNA. Next, proinsulin undergoes conformational change in the ER lumen, where three disul-

phide bridges are formed between specific cysteine amino acids in the protein (**??**). The formation of these three disulphide bridges are essential for the protein's bioactivity (**?**). Upon correct folding, proinsulin is transported to the Golgi apparatus where the C-peptide is removed by prohormone convertases and insulin gets its mature state and are stored in granules until they are ready to be secreted (**?**).

Several studies have shown that proinsulin misfolding and ER stress caused by genetic mutations in the insulin (*INS*) gene or adverse ER folding conditions can lead to β-cell failure and onset of diabetes (reviewed in (**?**)). Disturbance of the post-translational ER machinery leads to the initiation of the unfolded protein response (UPR) in order to limit the accumulation of misfolded proteins inside the ER. UPR activation leads to a slower translation rate, increased ER membrane surface area and an increase in expression of chaperonins. If this doesn't help to alleviate the ER stress, prolonged UPR activation leads to apoptosis (**?**).

2.1.2.2 Mutations leading to MIDY

Mutations in the *INS* gene can cause many different forms of the diabetic condition (**?**). At least 30 different autosomal dominant mutations have been found that are known to cause the MIDY syndrome by affecting proinsulin folding in the ER and 15 of these mutations have been experimentally determined to do so (**?**). More than half of the mutations generates free, unpaired cysteine residues, either by mutating an existing native cysteine or by changing a non-cysteine residue with a cysteine (**?**).

Similarly, Stoy and coworkers reported 10 autosomal-dominant mutations in the insulin gene causing MIDY in 16 different probands (**?**). Of these, six mutations either remove a cysteine or add an addional cysteine residue, these mutations are predicted to disrupt normal folding of proinsulin and lead to activation of the unfolded protein response (UPR) in the ER, ultimately leading to destruction of β-cells in a similar fastion as mutations in *Ins2* in the Akita and Munich mice (**?**).

Liu, *et al* (2010) found that exposed cysteine residues in proinsulin facilitates the formation of abnormal disulfide-linked protein complexes. This leads to non-canonical

pairing of cysteine residues and formation of protein aggregates, and this perturbation initiates the pathogenesis of MIDY (**?**).

2.2 Animal models

Animal models have been used for a long time in biomedical research to gain insights about human biology as well as to test novel treatments for human diseases such as cancer and diabetes (**?**). Different animal models have their own benefits and disadvantages, therefore the animal model should be chosen with care in order to use the best model for a given research topic.

Mus musculus is for many reasons the most popular model organism for studying human diseases, including low maintenance costs, short generation time, high ethical acceptance and well established genome modification methods (**?**). However, therapeutic findings in mouse models are not always predictive of therapeutic results in humans because of differences in size, anatomy and metabolism. Hence, animal models more similar to human patients are need to better test the efficacy and safety of novel treatments in translational

research.

Recently, *Sus scrofa* has gained popularity as a large animal model in translational research that is used to bridge the gap between rodents and humans (**???**). Pigs share many physiological, anatomical, and metabolic similarities with humans (**??**) and genetic analyses indicates that pigs are more closely related to humans than rodents
(**?**). While rodent models in general are cheaper to develop and maintain, pigs are generally a good model candidate where translational research in mouse models are limited by differences in size or life span, or where mouse models don't accurately mimic the characteristics of the human disease in question(**?**). Compared to non-human primates, pigs are less expensive and more ethically acceptable (**?**). Many genetic methods have been established for *Sus scrofa* (**?**), for example for diseases such as cystic fibrosis (**?**), Duchenne muscular dystrophy (**?**), and Huntington's disease
(**?**).

Many different animal models have been developed to study different types of diabetes, both rodent and non-rodent models (**?**). Pigs are generally considered to be a good animal model for the different types of di-

abetes, helping bridging the gap between mice and humans in translational and comparative research (**??**).

Here, two previously established swine models that were used for the work in this thesis will be described:

- Munich GIPR^{dn} pigs, a model for studying effects of incretin-based T2DM therapies.
- The Munich MIDY pig, a transgenic pig carrying a mutation in the insulin gene that makes it prone to misfolding, which is a large animal model for PND.

2.2.1 Transgenic dominant-negative glucose-dependent insulinotropic polypeptide receptor pigs

Based on the function of the incretin glucose-dependent insulinotropic polypeptide (GIP) is impaired in patients with T2DM (whereas the incretin hormone glucagon-like peptide 1 (GLP1) retains its effects in diabetic patients), a transgenic pig model was created, carrying a dominant-negative GIP receptor (GIPR^{dn}) to study

the role of the impaired GIP function in translational research (**??**). The mutant pig carries a GIPR copy that has an 8 amino acid deletion and an alanine to glutamine substitution at position 340, which is essential for signal transduction, controlled by the rat insulin 2 promoter (**?**).

These pigs have been shown to have reduced glucose tolerance and progressive β-cell destruction and impaired insulin secretion, hallmarks of T2DM. These results were similar to the results seen in a GIPR^{dn} mouse mutant that the transgenic pigs were based on (**?**). The GIPR^{dn} pig is an attractive model for developing new incretin-based diabetes therapies given the similarities between human and pig physiology and pathogenesis (**?**). In fact, treatment of GIPR^{dn} pigs with the GLP1 receptor antagonist liraglutide had similar effects to what has been seen in human patients, like reduced body mass and improved glycemic control (**?**).

2.2.2 Transgenic *INS*^{C94Y} pigs an animal model for human permanent neonatal diabetes

Renner and colleagues (**?**) have previously established the Munich MIDY (Mutant *INS* gene-induced Diabetes of Youth) pig as an animal model for studying long term effect of diabetes and hyperglycemia. The Munich MIDY-Pig biobank (**??**) consists of samples from different tissues and body fluids of 4 two-year-old female MIDY pigs and 5 wild-type littermate controls. Thus, the MIDY-Pig biobank is a unique resource that provides the scientific community with samples that can be used to study organ crosstalk within a multi-omics framework.

The Munich MIDY pig is based on the Akita mouse, a diabetic mouse model of insulin dependent diabetes mellitus (IDDM), that exhibits progressive hyperglycemia, β-cell dysfunction, and impaired insulin secretion (**??**). The Akita mice carries a missense mutation in the *Ins2* gene, C96Y, changing a cysteine to tyrosine affecting a disulfide bond, disrupting folding of the proinsulin gene (**??**). This corresponds to the C96Y mutation

found in the insulin gene in a human patient with MIDY syndrome (**?**). In the MIDY pigs the corresponding mutation in the insulin gene is C94Y.

The MIDY pigs develop a progressive diabetic phenotype that after a few months show significantly lower body mass, β-cell mass, and insulin levels when compared to non-diabetic litter mates (**?**). The MIDY pigs exhibit a stable diabetic phenotype that can be rescued by treatment with insulin, and are a suitable model for testing therapies such as β-cell transplantation in a human sized model organism (**?**).

2.3 Metabolism

Glucose homeostasis is primarily regulated by insulin and glucagon, and it involves several different organs (**?**). In response to elevated plasma glucose levels (such as after food intake), insulin is secreted to plasma from β-cells in the pancreas (**??**). Insulin's main action is to stimulate glucose removal, and it does so by having different effects on different target organs (**?**). In peripheral tissues, such as skeletal muscle and adipose tissue, insulin promotes glucose uptake, usage, and

storage (**?**). In liver, insulin inhibits gluconeogenesis and glycogenolysis, and promotes glycogen synthesis from glucose (**??**). Furthermore, insulin also inhibits secretion of glucagon from pancreatic α-cells (**?**).

In response to low blood glucose levels, glucagon is released from pancreatic α-cells. Glucagon promotes hepatic gluconeogenesis, production of glucose from hepatic glycogen, and glucose secretion. This raises the blood sugar to normal levels. In diabetic patients, the absence of insulin thus leads to a hyperglycemic state since there is no feedback mechanism to counteract the effects of glucagon (**??**).

There are other hormones beyond insulin and glucagon that also regulate glucose homeostasis, for example leptin and adiponectin that is released from fat cells (**??**).

2.3.1 Liver metabolism

The liver is an essential metabolic organ, and hepatic metabolic activity is regulated by insulin and other hormones. Glucose enters the hepatocytes via plasma membrane glucose transporters (GLUT)s, glucose is then phosphorylated by glucokinase into glucose-6-phosphate

(G6P). G6P is the precursor molecule for glycogen synthesis after feeding, it can also be converted into pyruvate and ATP through glycolysis and the citric acid cycle respectively. Pyruvate can also be used as a precursor for lipogenesis, the production of fatty acids (**?**).

During short-term fasting, glycogen is converted into glucose, which is then secreted from liver and transported to other tissues. Whereas during long-term fasting when the glycogen storages are depleted, the liver produces glucose through gluconeogenesis from lactate, amino acids and glycerol (waste products from skeletal muscle and adipose tissue respectively) (**?**).

2.3.2 Adipose tissue metabolism

White adipose tissue (WAT) is no longer considered to be a static organ that only stores fatty acids (**??**). In fact adipose tissue AT is an endocrine organ that is responsible for the production and secretion of a plurality of different hormones, impacting a wide variety of biological functions (**?**).

The accumulation of fat in WAT is governed by two processes: (i) the synthesis of fat (lipogenesis),

and (ii) fatty acid oxidation (lipolysis). Lipogenesis produces triglycerides for energy storage and occurs preferentially in adipose tissue, although it can also occur in the liver (**?**).

Lipogenesis is stimulated by a high intake of carbohydrates, whereas fasting and high levels of polyunsaturated fatty acids inhibits triglyceride generation by lipogenesis
(**?**). This regulation of lipogenesis is partly facilitated by hormones, e.g. leptin inhibits lipogenesis while insulin and angiotensin promotes it (**??**). Insulin promotes lipogenesis by: (i) increasing glucose uptake in adipose tissue, (ii) activating lipogenic and glycolytic enzymes, and (iii) inducing expression of lipogenic genes (**?**).

Breakdown of fat (lipolysis) occurs in adipose tissue during times of metabolic stress, such as fasting, when the fat storages are hydrolysed into glycerol and free fatty acids (FFAs) (**?**). The fatty acids are then secreted to the blood and transported to other tissues to be used as a source for energy generation (**?**). FFAs can be converted into ketone bodies in the mitochondria of the liver through fatty acid β-oxidation, ketone bodies are important energy substrates for extrahep-

atic tissue (**?**)

2.3.2.1 Differences between visceral and subcutaneous adipose tissue

It has been known for a long time that obesity has a higher correlation with body fat distribution than to total body fat mass (**?**). Visceral adipose tissue (VAT) is present mainly in the mesentery and omentum, and an excess abundance of VAT is linked to T2DM and cardiovascular diseases (**?**).

Adipocytes in VAT have a higher metabolic activity than adipocytes in subcutaneous adipose tissue (SCAT), which is reflected in a higher rate of lipolytic activity (**?**). Unlike SCAT, visceral fat is directly connected to the liver through the portal vein (**??**) giving secreted hormones and FFAs from VAT immediate access to the liver (**?**).

2.3.3 Retinoid metabolism

Retinoids (Vitamin A derivatives) have two physiological roles, (i) 11-cis-retinal is the chromophore of visual pigments in the eye, and (ii) all-trans- and 9-cis-

retinoic acid (RA) are responsible for regulating gene expression through the two nuclear receptors, retinoic acid receptors (RARs) and retinoid X receptors (RXRs) (**??**). Here, I will focus on the latter function of the hormonal role of retinoids.

During fasting the active retinoid precursor, all-trans-retinol (Vitamin A), is transported in plasma (**?**). After all-trans-retinol is imported by the target cells, it is converted in two steps into all-trans-retinoic acid (atRA, **Figure 2.1**). Retinol dehydrogenase 16 (RDH16) is one of many enzymes that catalyses the first oxidation step of this conversion, from all-trans-retinol into all-trans-retinal
(**?**). Retinol dehydrogenases catalyses the rate-limiting step of RA metabolism (**?**).

AtRA and 9-cis-RA regulate the expression of many different genes (**??**). Both atRA and 9-cis-RNA can bind to RARs, whereas 9-cis-RA is the only RA that binds to RXRs (**?**). To activate transciption, RARs and RXRs bind to retinoic acid response elements (RAREs) as heterodimers, RAREs are genomic elements that are composed of tandem repeats of 5'-(A/G)G(G/T)TCA-3' (**?**), most frequently as direct repeats (DRs) with two hexameric motifs interspersed by 1, 2, or 5 base-

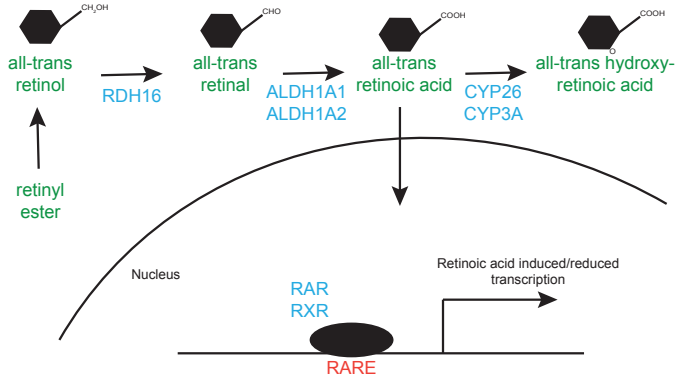


Figure 2.1. Schematic representation of retinol metabolism. RDH16 catalyzes the reaction from all-trans retinol to all-trans retinal, then the aldehyde dehydrogenase 1 protein family (ALDH1A) enzymes convert all-trans retinal to atRA. AtRA enters the nucleus and binds to the transcription factors RAR or RXR and activates them. RAR and RXR regulate transcription by binding promoter regions as either homodimers or as heterodimers with each other or other transcription factors. AtRA can also be converted into all-trans hydroxy-retinoic acid by cytochrome P450 enzymes. Metabolites are colored in green, proteins in blue, and DNA elements in red. All proteins named are examples of enzymes that catalyses the specific reaction.

pairs (DR1, DR2, and DR5 respectively) (**?**). Other non-canonical spacings can also occur.

Studies of the association between retinoid metabolism and diabetes in humans have shown that serum retinol levels are higher in type 2 diabetes and lower in type 1 diabetic patients respectively (**?**). RA has also been shown to affect the function of pancreatic β-cells by increasing insulin production and secretion (**?**).

AtRA inhibits preadipocyte differentiation into WAT (**?**), and atRA also stimulate lipolysis, decreases triacylglycerol concentration and increases fatty acid oxidation (**?**). Interestingly, Ziouzenkova and colleagues (**?**) have shown that retinal have distinct functions in adipose tissue of rodents, independent of its conversion to RA. Furthermore, they showed that retinal can interact with RARs and RXRs and that retinal and RA have different effects on adiposity in mice. All-trans-retinal was shown in the study to inhibit adipogenesis, and peroxisome proliferator-activated receptor and RXR signalling (**?**).

2.4 Proteomics

The proteome is the entire set of proteins that control and catalyse all processes in the cell in conjunction with other proteins or biomolecules at specific times and cellular locations, determining the functional state and phenotype of the cell (**?**). Because of this, and since mRNA and protein levels don't show perfect correlations
(**?**), the study of the proteome is important for giving us a complete picture to the inner workings of cells. Here, mass spectrometry (MS)-based proteomics will be described in greater detail. Briefly, purified proteins are digested into peptides by a peptidase followed by injection onto a liquid chromatography system where the peptides are separated by for example hydrophobicity. The peptides are then eluted and enter the mass spectrometer as ionized ions, where the precursor ions mass-to-charge (m/z) ratios are measured. Then selected ions are fragmented and the fragment spectra are collected. The peptides and proteins are identified and quantified using the precursor and fragment ion spectra together with a protein database, bioinformatic analysis is then performed in order to for ex-

ample detect differentially abundant proteins between two samples.

2.4.1 Sample preparation

Because MS analysis of intact proteins is inadequate for protein identification, the protein samples first have to be digested into smaller peptides using a peptidase (**??**). This is because sequence information is needed, and mass spectrometers are better equiped to obtain sequence information from peptides of sizes < 20 amino acids than from intact proteins (**?**). The most common peptidase used in proteomics is trypsin, which cleaves the polypeptide backbone after arginine and lysine residues (**???**). Other peptidases can also be used (**??**), e.g. digestion with the protease LysC, which cuts after lysine residues, can be performed in highly solubilizing conditions (e.g. 8 M urea) (**?**). Digestion with LysC followed by trypsin digestion has been reported to be more efficient than only using trypsin alone (**?**).

Since many proteins have a compact tertiary structure and hydrophobic properties (**?**), proteins must be denatured before being trypsinated. One common way to denature proteins is to use high concentrations of

urea, since it can bind to proteins and compete with native bonds between the amino acids (**?**). A benefit of using urea is that it is compatible with electrospray ionization (ESI)-MS since it doesn't interact with commonly used liquid chromatography columns and is thus eluted from the column before any of the peptides (**?**). Other denaturants can be used as well, such as sodium dodecyl sulfate SDS and sodium deoxycholate SDC. SDC can in contrast to urea be used at high concentrations without impacting trypsin's protease activity (**?**). But unlike urea these denaturants are not compatible with LC-ESI-MS, and have to be removed prior to chromatagraphic separation (**?**). This can easily be done for SDC using either a acid precipitation method or a phase transfer method (**??**). SDC has been found to increase the performance of detecting peptides derived from membrane proteins (**?**), zebrafish liver tumors (**?**), and proteins from AT (**?**).

To enhance the digestion of the proteins, the samples are also treated with dithiothreitol (DTT) and tris(2-carboxyethyl)phosphine (TCEP) to reduce disulphide bridges between cysteine residues. The free cysteine amino acids are then alkylated with iodoacetamide to prevent the cysteine residues to form new disulphide

bridges (**?**).

2.4.2 Liquid chromatography

Due to the different functional groups of peptides (hydrophilic, hydrophobic, basic, acidic), many options exist to choose from when choosing a separation technique
(**?**). The most popular separation methods use reversed phase liquid chromatography (RPLC), strong-cation exchange (SCX), strong-anion exchange (SAX), and hydrophilic interaction liquid chromatography (HILIC). Of these, RPLC is the most popular for HPLC proteomics, due to its robustness, separation efficiency, and selectivity(**?**). In RPLC, the columns are packed with a hydrophobic stationary phase, of which octadecyl carbon chain (C18)-bound silica is the most common (**??**), and for the mobile phase, water together with an organic solvent such as acetonitrile or methanol is used (**?**). Furthermore, the RPLC method's eluents are fully compatible with peptide ionization methods (**??**).

The peptides generated by protein digestion are injected onto a high performance liquid chromatog-

raphy (HPLC) micro capillary column (RPLC) that is directly coupled to the mass spectrometer. The peptides are eluted from the column using a solvent gradient of increasing organic content and enter the mass spectrometer according to their hydrophobicity, hydrophilic peptides elute first and hydrophobic peptides elute later (**?**). Since the sensitivity of the LC-MS analysis is directly proportional to the peptide concentrations eluting from the chromatography column, the chromatography is performed at very low flow-rates.This is achieved by using capillary columns with small inner diameters, usually on the scale of 50-150 μm, which allow for flow rates as low as 500 nl min^{-1} (**?**).

2.4.3 Mass spectrometry

A mass spectrometer consists of an ion source, a mass analyser that computes the mass-to-charge ratios (m/z) of the ions of interest, and an ion detector that for each m/z value counts the number of ions (**?**).

2.4.3.1 Electrospray ionization

There are two methods, which define the ion source, to ionize the peptides before entering the mass spectrometer, ESI and matrix-assisted laser desorption/ionization (MALDI) (**??**). ESI has now become the state of the art method for comprehensive study of the proteome, due to its compatibility with HPLC methods (**??**). Recent research has shown that overlap of identified peptides between the two methods are quite low and that they are complementary to each other (**?**).

In ESI, charged (mostly protonated) peptides are eluted in droplets from the spray needle, which is held at a high electrical potential with respect to the inlet of the mass spectrometer. As the droplets travel towards the mass spectrometry inlet, the solvent evaporates, which increases the charge concentration on the droplets' surface area. Single, naked charged peptides are generated either (i) by the fission of the droplets or (ii) single ions can be emitted from the droplet when the electrical field is sufficiently high on the surface of the droplets (**??**). The charged, single peptides then enters the vacuum of the mass spectrometer ready to be measured.

2.4.3.2 Mass spectrometry and mass analysers

There are different ways a mass spectrometer can operate. In the data-dependent acquisition (DDA, which is the most common mode of operation for proteomics), the mass spectrometer records the intensity signals for all ions' m/z ratio, thus generating a mass spectrum (at the MS^1 level) (**??**). Because each ion signal consists of a cluster of isotope peaks, the charge state of the peptides can easily be determined by the shape of the ion signal. The mass of a peptide can thus be derived from the detected m/z ratio and the peptide charge (**?**).

Since primary structure (sequence) information is needed to unambiguously identify a protein, peptides are fragmented by collision with an inert gas (such as nitrogen, helium or argon) (**?**). The most common type of product ions generated from the precursor ion by collision-induced disassociation (CID) are derived from breakage of the peptide bond, b and y ions, depending on which side retains the proton (**??**). The tandem-MS (MS/MS or MS^2) spectrum is then obtained. Throughout a LC-MS run, the mass spectrometer cycles between obtaining MS^1 and MS^2 spectras.

Once the gradient finishes, peptides are identified and quantified by a computational analysis.

Several different types of mass analysers exist, and each come with their own strengths and weaknesses that make them suitable for different applications. Examples of different mass spectrometers include: linear ion traps (LIT), quadropole mass filters (QMF), time-of-flight instruments, and high-resolution Orbitraps.

LIT mass analysers use four quadropolar rods that are segmented into three parts each. AC is applied to pairs of electrodes at opposite sign to trap ions using electromagnetic forces in the radial dimension, whereas the ions are trapped axially by a potential well generated by applying different DC currents to the three different segments of the rods. Ion motion within the trap is proportional to the amplitude of the applied AC current and the size of the ions, with higher amplitude and larger size leading to larger motions. Ions are then scanned by a process called resonance ejection. Resonance ejection works by that an additional, increasing AC current is applied to the exit rods, and when ions are in resonance with the main AC current they exit through slits in the rods. The increasing AC current ensures that ions with lower m/z values are

ejected first (**?**).

Similarly to LITs, ions in QMFs are radially confined by AC. What differentiates the QMF to the LIT is that a DC current is applied to each rod pair with opposite sign, this produces a complex electric field that allows for ion motion both radially and axially. In a scan, both electrical currents are increasing (while the proportion of the currents are kept the same) so that ions of increasing mass pass through the QMF (**?**).

Orbitraps measure mass-to-charge ratios from the ion's oscillatory motion frequency using Fourier transformation. In an Orbitrap mass analyser, ions are oscillating around a spindle electrode, and axial oscillating movements are detected by electrodes as transients. Transients are the frequencies of oscillations, and are unique for each m/z value. The transients are then transformed into m/z values using Fourier transformation. Unlike quadropole based mass analysers, all ions are detected simultaneously in an Orbitrap and no electric ramp is required
(**???**).

These different types of mass analysers can be used alone, or together in hybrid instruments that takes advantage of the strengths of each type of mass analyser.

For example, the high speed and sensitivity of LITs can be combined with quadropoles enhanced capabilities for precursor isolation to create a hybrid instrument with increased dynamic range (**?**). Another example of a hybrid instrument is the Q Exactive (Thermo Fisher Scientific) family series of instruments that compines a QMF with an Orbitrap (**???**). The quadropole-Orbitrap hybrids have been proven to be very powerful for mass spectrometry based proteomics, significantly increasing the number of identified proteins per hour in DDA
(**?**).

DDA-MS has seen striking developments in both computational and sample preparation methods, as well as in instrument performance during the last couple of years, with the complete yeast proteome being obtained (**??**) and over 10,000 proteins have been identified in human cell line samples (**??**). Tissue proteomics presents its own challenges due to the heterogeneous cellular composition and dynamic protein concentration range of most tissues. Each tissue has its own particular challenge for proteomics, for example the high dynamic range of skeletal muscle (**?**), the high levels of proteases in the pancreas (**?**), and the

high concentrations of lipids and hydrophobic proteins in adipose tissue (**?**).

2.4.4 Computational and bioinformatic analysis

Computational proteomics concerns the mathematical and statistical algorithms used to identify peptide and protein identification as well as their quantification, whereas the bioinformatic analysis concerns the downstream functional analysis that leads to biologically relevant and interpretable results (**?**).

2.4.4.1 Identifying peptides: from spectra to proteins

Peptide sequence can be deduced from the distances between neighbouring peaks in the MS^2 spectra, since each fragment ion differs from its neighbour by one amino acid. But the advent of whole-genome sequencing made it possible to match peptide-fragmentation spectras to theoretical spectras derived from sequencing databases, turning the issue of peptide identification into a database matching problem (**??**). Compu-

tationally this is more efficient since only a fraction of all possible amino acid combinations have to be considered (**?**).

The goal of a tandem-MS database search is to detect the best matching sequence to each fragment spectrum. Four general methods for performing and scoring peptide database searches: descriptive, interpretative, stochastic, and statistical and probability based modelling (**?**). Here, I will focus on probability based methods and describe this method in more detail.

Probability based modelling was popularized by the commercial Mascot software
(**??**). A probability based approach relates features sequences to the spectra, for example the frequency of matches of b and y ions are calculated and used to calculate a identification probability (**?**).

The free software MaxQuant is one of the most popular data analysis programs for handling MS-based proteomics data (**??**). MaxQuant uses it's own probability based peptide search enginge called Andromeda (**?**). Andromeda's peptide search algorithm first calculates the n number of theoretical fragment ions and the k number of matches between the spectrum and

the theoretical fragment ions, the higher k is relative to n the higher the probability is that these matches were not due to random chance. Andromeda uses this information to calculate an approximate p-value for the null hypothesis that there is no similarity between the theoretical and observed fragment ion masses. A successful match is called a peptide spectrum match (PSM) (**?**).

Due to the vast amount of hypotheses being evaluated and that many of the acquired spectra might be due to chemical and electrical noise, many of the PSMs might be due to random chance and thus be false positives (**???**). To alleviate this, the detected PSMs should be corrected with a false discovery method (FDR) that limits the number of false positives while containing as many of the true positives as possible. In proteomics, FDR values are usually calculated from a decoy database (**?**). There are many ways to construct a decoy database, of which a reversed proteins database is the simplest and most commonly used (**?**).

A reversed decoy database is created by reversing the protein FASTA sequences used for obtaining the theoretical peptides. Using the reversed method yields a decoy database with similar features to the target

database (in respect to number of proteins and peptides, protein length, and sequence redundancy) (**?**). The number of positive hits in the decoy database is used to estimate the false positives of the target database, assuming that the probability of finding false PSMs is equal for the two databases (**?**).

Identified peptides are then matched to the proteins in the target database. The proteome contain a high degree of sequence redundancy, which is due to different evolutionary processes and that a single protein can have multiple isoforms, this leads to difficulties in assigning a peptide to one single origin protein. To solve this, Maxquant merges all proteins that can't be distinguished by the identified peptides into protein groups (**??**). An FDR filter then applied in order to filter out false positives. An FDR threshold of 1% is usually applied in proteomics, both for peptide and protein identifications (**???**).

2.4.4.2 Quantitative proteomics

There are two general approaches to quantitate proteins in DDA-MS, labelled or label-free quantification. Both of these techniques are relative quantification

methods, meaning that two or more sample groups are compared (**??**).

Label-based quantification, uses stable isotope labels such as C13 and N15. These stable isotopes are chemically more or less identical to the most abundant isotopes. Thus, labelling with stable isotopes ensures that the mass spectrometer can distinguish between proteins that are eluted in parallel but derived from different samples. Label-based approaches includes: (i) metabolic (such as SILAC), (ii) isotopic (ICAT), and (iii) isobaric (ITRAQ and TMT) labeling. The labeling step can be performed at the peptide or protein level depending on the method used (**?**). The advantage of using a label-based approach is that they can be more accurate than label-free methods, whereas disadvantages can include (depending on the method) that the number of samples that can be labeled is limited, variability in labeling efficiency, and that additional sample preparation steps are included where loss of proteins/peptides can occur (**??**).

Label-free quantification is a cheap and easy method for relative protein quantification and two major strategies exist: spectral counting or peptide peak intensity measurements (**??**). They also have the advantage of

having a higher dynamic range than label-based approaches (**?**). Spectral counting methods infer protein abundance from the number of PSMs detected for each protein, whereas peak intensity methods use the peptide's intensity, which is derived from integration of chromatographic peaks, for quantification (**?**). Generally, ion intensity methods perform better than spectral counting methods as it more robustly quantitates low intensity peptides (**?**) and offer in general better performance than spectral counting (**?**).

MaxQuant uses its own quantification algorithm based on peptide intensity quantification called MaxLFQ (**?**). MaxLFQ first calculates protein intensity ratios from the median of the peptide intensity ratios between all samples that share a predetermined number of peptides (default is 2 shared peptides). The protein intensities for each sample are then solved by a least-squares analysis that will satisfy all viable protein ratios between the samples to calculate the optimal abundance profile for the protein (**?**).

There are many R (**?**) packages developed by the proteomics research community, both for bioinformatic analyses and data visualization (**??**), of which for example DEP is a package developed specifically for com-

parative proteomics (**?**).

2.4.4.3 Missing value imputation

Typically, global proteomic experiments generally suffer from a large degree of missing values(**???**). There are many reasons why a peptide could be missing in the dataset, both experimental and bioinformatic mechanisms can lead to missing peptides and the underlying cause are often unknown and complex (**??**). There are three categories of missing values: missing completely at random (MCAR), missing at random (MAR), and missing not at random MNAR. MCAR values are usually derived from small measurement errors or stochastic fluctuations whereas MNAR values are derived from targeted effects (such as the instrument detection limit or that the protein doesn't exist in the sample) (**?**). Since ignoring missing values would dramatically reduce the size of the data and that many statistical tools need complete datasets (**?**), missing value imputation and/or combined with missing value filtering is usually applied to the datasets.

Choosing which imputation method to use is not trivial, as it can affect how the data is interpreted.

There are many available imputation methods, but no single approach is consistently better than other approaches (**??**). This is probably due to that different methods are developed for different scenarios, for example a probabilistic minimum imputation methods, that imputes missing values from the lower part of normal distribution, generally performs well for MNAR values but not for MCAR values (**??**). Ideally, the type of missing value should guide the use of the imputation method to use, but in practice it is difficult to determine whether a missing value observation is stochastic or deterministic in nature (**?**).

2.5 Statistics

Comparative studies where the researchers compare the outcomes of two different scenarios, e.g.: mutant versus wildtype, or placebo versus treated, are among the most common types of experiments in the life sciences.

2.5.1 Statistical hypothesis testing

Statistical hypothesis testing is thus performed in order to detect differences between the study groups (**?**). When performing a statistical test we assume that there is no difference between the groups (the null hypothesis), and after the test the hypothesis is either accepted or rejected based on the test statistic.

A two-sample t-test compares the means of the 2 groups relative to the standard deviation of the samples and returns a p-value. If the p-value is larger than a predetermined threshold, α (usually set to 0.05), then we accept the null hypothesis that there is no difference between the 2 groups. Contrary, if the p-value is smaller than α then the null hypothesis is rejected and the other hypothesis is accepted that there is a difference between the two groups (**?**).

A two-sample t-test has three assumptions (**?**):

1. The sample populations should be normally distributed.
2. The sample populations should have the same variance.
3. The sample populations should be uncorrelated.

2.5.1.1 Different outcomes of statistical testing

When performing a statistical test, there are four different potential outcomes
(**?**). These are illustrated in **Table 2.1**, two of these outcomes are correct whereas the other two are incorrect (errors). There are two different types of errors, type I errors and type II errors. When a type I error is made, a difference is deemed significant even though there is no true difference between the groups (and vice versa for a type II error). In the scenario when one statistical test is performed with a significance level of α-level = 0.05, there's a 5% chance that a significant outcome will be wrong 5% of the time. That is 1 out of 20 times, we will accept that we get a type I error (or in other words yield a false positive result) (**??**).

An important factor to take into consideration is the statistical power, which is the probability of detecting a true difference between the two groups (**??**). With increasing statistical power, the likelihood of doing a type II error decreases. Thus, statistical power determines the ability of the test to find true significant differences between the groups. Sample size is a

		TRUTH	
		Difference	No
RESULT OF STATISTICAL TEST	Significant	True positive	Fal
	Not significant	False positive	Tru

Table 2.1. The four outcomes of a statistical test. Correctly identified outcomes are colored in blue, whereas incorrectly identified outcomes are colored in red.

major factor that influences the statistical power of a study, the statistical power increases with increasing sample size (**??**).

2.5.2 Multiple hypotheses testing

When performing multiple statistical tests at the same time (as in most omics experiments) p-values are misleading since they only inform us about a singular outcome. Consider an experiment with 10000 statistical tests and an α-level of 0.05, then 500 of these tests will be deemed significant by random chance even if there's no true difference between the samples (**?**). Therefore, p-values will have to be adjusted and reinterpreted,

which is done by using different multiple hypotheses correction methods.

The simplest multiple hypotheses correction method, is the Bonferroni's. The Bonferroni method controls the family-wise error rate (FWER, that is the probability of detecting one false positive) by multiplying the p-values with the number of tests. This greatly reduces the false positive rate (FPR, the probability of inferring an effect even though no effect is present) and the statistical power, and increases the yield of false negatives (**?**).

Instead, a more appropriate method for the analysis of proteomics datasets is the Benjamini-Hochberg method that controls the false discovery rate (FDR, the proportion of false positives to all significant tests) (**?**). The benefit of methods that controls FDR instead of the FWER, is that they maintain statistical power and limits the number of false positives (**?**). An FDR of 0.05 means that of all significant tests 5% will be false positives.

A useful method for controlling the FDR has been developed by Storey, 2002. This method controls the FDR by using the fact that non-significant p-values are uniformly distributed whereas lower, significant p-

values are enriched for values close to 0 . Storey's method incorporates an estimation of the fraction of tests for which the null hypothesis is true to adjust the p-values and yield q-values (**?**).

2.5.3 Analysis of variance, linear models and *post hoc* tests

For experimental studies that includes more than two sample groups or more than one factor, the t-test is obsolete. In fact, the t-test is a distinct case of an analysis of variance (ANOVA) test, and ANOVA has the same requirements as the t-test. The null hypothesis for an ANOVA is that all samples are derived from the same distribution and have the same means, and when the null hypothesis is rejected it is concluded that all means are not equal and further tests are needed to find out which means are not equal (if the number of groups > 2) (**??**).

An ANOVA is based on linear models. A linear model tries to fit a known function in order to find the "best" line through the data by minimizing the sum of squares of the residuals. In an ANOVA the

predicted values are the factor level means and the sum of squares represent the variation that is not accounted for by the factor (**?**).

If there are more than two groups (factor levels), one most use a *post hoc* test in order to investigate which group means that are significantly different from each other. One of the most common methods for follow up tests on ANOVA is Tukey's Honest Significant Differences (HSD) (**?**). Tukey's HSD method compares all pairwise comparisons and calculates p-values for each comparison that are adjusted for multiple hypotheses testing using FWER to limit the type 1 error rate. Important to note is that when we have many features (as in most omics experiments), Tukey's method applies correction for multiple hypotheses on each feature separately.

2.5.4 Principal component analysis

Principal component analysis (PCA) is an unsupervised statistical method that reduces dimensions of complex, high-dimensional data (such as omics data with multiple genes/proteins) while preserving trends and patterns of the data. PCA does so by project-

ing the high-dimensional data onto a series one dimensional principal components (PC). The first PC is chosen to minimise the sum of the squared distances between the data points and PC, this have the effect that the variance of the projected points is maximized. The second PC (and all following PCs) is selected in a similar way with the additional constraint that it must be orthogonal to first PC, PCA is most often used in order to easier visualize the high-dimensional data (**?**).

2.5.5 Gene set enrichment analysis

The Gene Ontology GO database was created to construct a consistent and curated language for functional annotation of proteins (**???**). Three different types of GO categories exists: molecular funcion (GOMF), biological process (GOBP), and cellular compartment (GOCC). GO categories are strictly hierarchical, most classes thus have both parent and child terms that describe protein functionality at different levels of detail (**?**).

Gene set enrichment analysis (GSEA) is a method for detecting functional enrichments in biological datasets (**?**). A GSEA algorithm performs a statistical test to

see whether more genes/proteins are associated with a gene set (e.g. a GO class) than by random chance (**?**). This can be done in different ways. First, a list of significantly more abundant genes (as determined by a statistical test) can be tested for enrichments of gene sets. This is the method of bioinformatic tools such as STRING (**?**) and DAVID (**?**). The second method doesn't need any prefiltering of significant features, instead it takes into account all genes/proteins in a dataset together with a quantitative measurement (e.g. expression levels, L2FC, and p-values). This is reflecting that a smaller change in expression among a subset of genes sharing a similar function or belonging to the same molecular pathway might be more biologically important than a large change in one gene (**?**). This is the method of the software GSEA. Other types of databases for GSEA can also be used such as pathway information (Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein domains (Interpro) (**??**).

Due to the hierarchical structure of the GO database, GSEA will often find enriched categories at different depths from the same hierarchical paths (**?**). Methods have been developed to remove redundant GO terms

from a GO list to make it easier to interpret, e.g. GO trimming that removes redundant term based on the number of shared proteins between the two classes (**?**).

Chapter 3

Materials and methods

3.1 Adipose tissue from the Munich MIDY pig biobank

5 WT and 5 MIDY female littermates were maintained for two years. MIDY pigs started receiving treatment with both long-lasting (Lantus®; Sanofi) and short-acting insulin (NovoRapid®; NovoNordisk) at the age of two months, in order to mimic suboptimal diabetes

treatment (moderately hyperglycemic). To remove effects from estrous cycle, the 2 year old pigs were estrus synchronized and inseminated 12 days before necropsy. A clinical examination was performed the day before necropsy to record data about the general well-being of the pigs. Pigs were fasted overnight before necropsy, and tissue samples were collected according to standardized sampling procedures (described in Surma, *et al* (2015) (**?**). Tissue samples for further molecular studies were shock frozen within 20 min of acquirement and stored in -80°C.

3.1.1 Proteomic sample preparation

Two tissue samples from different locations were obtained from each pig and AT type (4 MIDY pigs and 5 WT pigs, SCAT and MAT) from the Munich MIDY pig biobank (**Table 3.1**, the harvesting of the samples have been described in (**??**)). The two tissue samples from each pig and tissue group were pooled (see **Table 3.2** for the outline of experimental design, and approximately 100 mg tissue were collected from each pooled sample. Each pooled sample was suspended in a solution containing 1% SDC and 50 mM

Table 3.1. Characteristics of the pigs and adipose tissue samples from the MIDY-pig biobank.

Animal ID	Date of birth	Genotype	Section number	H
1861	25/10/2012	WT	S 736/14	
1857	25/10/2012	MIDY	S 737/14	
1885	14/11/2012	WT	S 738/14	
1856	25/10/2012	MIDY	S 739/14	
1859	25/10/2012	MIDY	S 740/14	
1877	07/11/2012	WT	S 741/14	
1875	07/11/2012	WT	S 743/14	
1886	07/11/2012	MIDY	S 744/14	
1878	07/11/2012	WT	S 745/14	

ABC[1]. The samples were homogenized using a homogenizer (ART-MicraD8, ART Prozess- & Labortechnik, Müllheim, Germany) at 23,500 rpm for 2×1 min. Samples were kept on ice for 30 min and were then centrifuged for 5 min at 16,000×g, 4°C. The aqueous layer beneath the lipid layer was harvested and transferred to new Eppendorf tubes. Protein concentration was measured with a NanoDrop ND-1000 spectrophotometer (Marshall Scientific), using the A280 method for proteins/peptides.

		Genotype	
		MIDY	**WT**
Adipose tissue	**Mesenteric**	MIDY, MAT	WT, MAT
	Subcutaneous	MIDY, SCAT	WT, SCAT

Table 3.2. The multifactorial experimental design for MIDY AT study showing the two independent variables (genotype and AT source) and the four subsequent 2-factor groups.

50 μg of protein in 50 μl 1%, 50 mM ABC were diluted to 100 μl 0.5% SDC, 50 mM ABC. The cys-

[1] all water used were HPLC grade

teine residues were reduced with DTT and TCEP at a final concentration of 4 and 2 mM respectively for 30 min at 56 °C and cysteine residues were blocked using iodoacetamide for 30 min at a final concentration of 8 mM in the dark. DTT was added to a final concentration of 10 mM and the samples were incubated for 15 min before protein digestion. Proteins were first digested with 1 μg LysC (Wako) for 4 hours followed by digestion with 1 μg trypsin (Promega) for 16 hours at 37°C.

SDC was removed using the acid precipitation method (**???**). The digests were acidified to a final concentration of 1% TFA and incubated on ice for 5 min, before the samples were centrifuged at 16,000×g for 15 min. The supernatant were transferred to a new Eppendorf tube and evaporated in a vacuum centrifuge (Bachhofer). The samples were stored at -20°C until sequencing by mass spectrometry. The samples were resuspended in 0.1 % FA, and 1.5 μg peptides in 15 μl were injected into the LC-MS/MS system.

3.1.2 Mass spectrometry and proteomic analysis

The nano-LC-MS/MS analysis was performed using a Q-exactive HF-X mass spectrometer coupled to an Ultimate 3000 nano-LC system (Thermo Scientific) as described in (**?**). Briefly, 1.5 μg peptides were separated at 200 nl/min using sequential linear gradients from 1% to 5% solvent B (0.1% FA in acetonitrile, whereas solvent A was 0.1% FA, 1% acetonitrile) for 10 min, followed by 5% to 25% B for 115 min, and lastly 25% to 50% B in 20 min. Spectra were acquired using a precursor ion scan at a resolution of 120,000 from 380 to 2000 m/z, followed by MS/MS scans of the 24 peaks with highest intensity at a resolution of 15,000.

Peptides and proteins were identified (FDR < 0.01) using MaxQuant 1.6.3.4 (**??**) with the NCBI RefSeq *Sus scrofa* proteome database (version 20180313) and quantified using MaxLFQ with the match between run feature enabled (**?**). For details about parameters used see **Supplementary Table C.1**.

3.1.3 Bioinformatics

The bioinformatic analysis was done in R 3.6.0 (**?**). Gene names were identified based on the NCBI RefSeq identifiers using bioDBnet (**?**) and BioMart (**?**) using the biomaRt package in R (**?**). The proteins LFQ intensity values were $\log_2$ transformed. Proteins were filtered using 75% valid values in at least one group (i.e. only one missing value allowed in at least one group). Additional normalization (variance stabilizing transformation) and imputation were performed using the DEP package (**?**). The missing proteins were imputed using random draws from a left shifted Gaussian distribution with parameters: scale = 0.3 and shift = 1.8 (scale defines the width of the Gaussian distribution relative to the standard deviation of the data, and shift defines how much the distribution that the random values are drawn from are shifted downwards). For statistical analysis R's base functions were used together with the tidyverse suite of packages (**??**). Multiple hypotheses correction was performed using the false discovery rate method provided by the R qvalue package (**?**).

Significant proteins (q-value < 0.05) were uploaded

to the STRING database (version 11.0) (**?**) separately for each group. Default parameters were used, except for minimum interaction score (0.15 for significant proteins in genotype categories and 0.4 for significant proteins in tissue categories). Selected enriched gene sets were selected and highlighted on the protein networks.

Indenpendently for both the genotype and tissue factors, the p-values for each factor from the 2-way ANOVA were transformed to $-log_{10}$ format. The $-log_{10}(p-values)$ were then assigned to be negative if the L2FC were negative and vice versa, and the proteins were than ordered according to these transformed values. The preranked protein sets were analysed using GSEA 3.0 (**?**) using the GSEApreranked method with KEGG pathways and all three GO categories. The preranked GSEA was run with default parameters except for the min size parameter which was set to 10.

GOtrim 2.0 (**?**) was used to trim GO enrichments (with default parameters) independently for each group. REVIGO (**?**) and the Cluego 3.7.1 app (**?**) Enrichment Map 3.2.1 (**?**) were used to visualize GSEA enrichment results.

3.2 GIPR^{dn} liraglutide treatment study

GIPR^{dn} treatment and sample acquisition, has been described previously by Streckel, *et al*, (2015) (**?**). Briefly, 18 heterozygous GIPR^{dn} pigs were randomly assigned to be treated with liraglutide or placebo. Pigs were subcutaneously injected with pre-filled pens once daily for 90 days with either liraglutide (Victoza®, 6 mg/ml, Novo Nordisk A/S) or placebo (0.9% NaCl, B. Braun). Liraglutide doses were based on human dosages (0.6-1.2 mg per day) and corrected for pig body mass. Pigs received treatment between the age of 2-5 months. At the age of 2 months, GIPR^{dn} pigs had impaired oral glucose tests and delayed insulin secretion. Pigs were fed a standard diet and other living conditions were the same between the two treatment groups. Following the end of the treatment period, pigs were euthanized and selected organs were weighed. Tissue samples were shock frozen and stored at -80°C.

One piece of flash frozen liver sample was obtained from each pig in the GIPR^{dn} liraglutide treatment study (**Table 3.3**, see **Figure 3.1** for a visual summarization

of the methodology), resulting in sample groups consisting of 8 LT and 9 PT GIPR^{dn} pigs respectively. Approximately 30-50 mg of tissue were taken from each sample tube, and 15 μl 8 M urea, 400 mM ABC per mg liver tissue. The tissues were lysed using a homogenizer (ART-MicraD8, ART Prozess- & Labortechnik, Müllheim, Germany) at 23,500 rpm for 30 seconds. The lysates were then centrifuged through a QIA-gen shredder device (Qiagen, Hilden, Germany) for 30 second at max speed. The samples were then stored at -20°C until the next step.

The protein concentration was determined using a Pierce®660 nm Protein Assay (Thermo Scientific) and a DU®640 spectrophotometer (Beckman Coulter). The bovine serum albumin (BSA, Thermo Scientific) standard dilution series was as following: 0, 50, 100, 250, 500, 750, 1000, 1500, and 2000 mg/ml. The liver samples mean protein concentrations were measured using a 1/10 and 1/20 dilution. After the protein concentrations were determined, they were adjusted to 2 mg/ml by adding 8 M urea, 400 mM ABC independently for each sample.

100 μg of protein (50 μl lysate) were reduced using a final concentration of 4 mM DTT and 2 mM TCEP

Table 3.3. Characteristics of pigs and liver samples from the GIPR*dn* liraglutide treatment study.

ID	Genotype	Treatment	Gender	Birthdate	Sampli
1311	GIPRdn	Placebo	male	07/07/2011	19/1
1312	GIPRdn	Liraglutide	female	07/07/2011	19/1
1313	GIPRdn	Liraglutide	female	07/07/2011	19/1
1315	GIPRdn	Placebo	female	07/07/2011	20/1
1316	GIPRdn	Placebo	female	07/07/2011	20/1
1317	GIPRdn	Placebo	female	07/07/2011	21/1
1323	GIPRdn	Placebo	male	08/07/2011	21/1
1326	GIPRdn	Liraglutide	male	08/07/2011	20/1
1331	GIPRdn	Liraglutide	female	08/07/2011	21/1
1488	GIPRdn	Liraglutide	male	17/11/2011	02/0
1492	GIPRdn	Liraglutide	male	17/11/2011	02/0
1497	GIPRdn	Placebo	female	17/11/2011	02/0
1501	GIPRdn	Placebo	male	17/11/2011	03/0
1502	GIPRdn	Placebo	male	17/11/2011	03/0
1503	GIPRdn	Liraglutide	male	17/11/2011	03/0
1505	GIPRdn	Placebo	female	17/11/2011	02/0
1506	GIPRdn	Liraglutide	female	17/11/2011	03/0

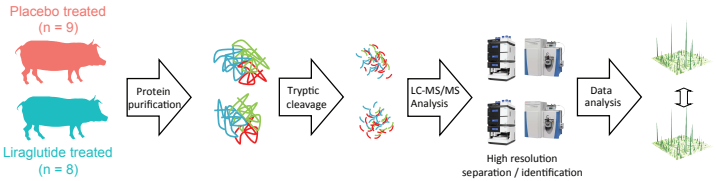


Figure 3.1. Experimental outline. Liver samples from liraglutide (n = 8) and placebo (n = 9) treated GIPR^{dn} pigs were independently analysed using label free quantification (LFQ). Proteins were digested into peptides using LysC followed by trypsin. Peptides were detected using a Q-Exactive HF-X instrument coupled to a 50 cm nano-LC column without prefractionation. Peptides and proteins were identified and quantified using MaxQuant and label free quantification (LFQ). The bioinformatic analysis was performed to find significant, relative proteome differences between the two treatment group

for 30 min at 56°C. Following the reduction, cysteine residues were blocked using a final concentration of 8 mM IAA for 30 min at room temperature and in the dark. DTT was added to a final concentration of 10 mM and the samples were incubated for 15 min at room temperature. The proteins were first digested with LysC at an enzyme to protein ratio of 1/50 for 4 hours at 37°. 419 μl H20 were added to each sample to reduce the urea concentration to below 1 M. Trypsin was added at an enzyme to protein ratio of 1/50 for 16 hours at 37°C. 5 μl FA was added to each sample to reduce the pH to below 3 in order to stop the trypsin protein digestion.

3.2.1 Mass spectrometry and proteomic analysis

The nano-LC-MS/MS analysis were performed using a Q Exactive HF-X mass spectrometer coupled to an Ultimate 3000 nano-LC system (Thermo Scientific) as described in (**?**). Briefly, 2 μg peptides were separated at 200 nl/min using sequential linear gradients from 1% to 5% solvent B (0.1% FA in acetonitrile, whereas

solvent A was 0.1% FA, 1% acetonitrile) for 10 min, followed by 5% to 25% B for 115 min, and lastly 25% to 50% B in 20 min. Spectra were acquired using a precursor ion scan at a resolution of 120,000 from 380 to 2000 m/z, followed by MS/MS scans of the 24 peaks with highest intensity at a resolution of 15,000.

Peptides and proteins were identified (FDR < 0.01) using MaxQuant 1.6.1.0 (**??**) with the NCBI RefSeq *Sus scrofa* proteome database (version 20180313) and quantified using MaxLFQ and with the match between run feature enabled (**?**). Apart from MaxQuant version number used, the parameters used were the same as for the study of AT in MIDY pigs. For details about parameters used see C.1.

3.2.2 Bioinformatics

The following bioinformatic analysis was done in R 3.6.0 (**?**). Gene names were identified based on the NCBI RefSeq identifiers using a custom R script together with the BioMart database (**?**) using the biomaRt package in R (**?**). All proteins LFQ intensity values were $\log_2$ transformed. Proteins were filtered using 55% valid values in at least one group (i.e. allowing

for 3 or 4 missing values at least one group, depending on the group (since n is different for the two treatment groups). Additional normalization (variance stabilizing transformation) and imputation were performed using the DEP package (**?**). The missing proteins were imputed using a random draws from a left shifted Gaussian distribution with parameters: scale = 0.3 and shift = 1.8 (scale defines the width of the Gaussian distribution relative to the standard deviation of the data, and shift defines how much the distribution that the random values are drawn from are shifted downwards. A linear model was used with LFQ intensity as the dependent variable and gender as a independent variable 3.3 in order to filter out the effects of gender on the protein LFQ intensities. For all downstream analysis the resulting residuals from the linear model was used. For further statistical analysis R's base functions were used together with the tidyverse suite of packages (**??**). Multiple hypotheses correction was performed using the false discovery rate method provided by the R qvalue package (**?**).

Significant proteins (q-value < 0.05) were uploaded to the STRING database (version 11.0) (**?**) separately for each group with default parameters. Selected en-

riched gene sets were selected among the significantly enriched gene sets and highlighted on the protein network maps.

The residuals from the linear model together with the gene names for all proteins (whether significant or non-significant) were uploaded to GSEA 4.0.1 (**?**). GSEA analysis was performed using GOBP, GOCC, GOMF, and KEGG gene sets. Gene sets larger than 500 proteins or smaller than 15 were excluded from the analysis. The following parameters were used: permutation type was set to gene_set, metric for ranking gene sets were set to Ttest, and collapse dataset to gene symbols was set to false. Default settings were used if not specified elsewhere.

Chapter 4

Results

4.1 Differential proteome analysis of adipose tissue from the Munich MIDY pig biobank

In order to investigate how hypoinsulenemia in the MIDY pigs effects the proteome in adipose tissue, two different types of adipose tissue were selected from the Munich MIDY pig biobank (**?**), for both wildtype (WT, n = 5) and MIDY pigs (n = 4): (i) abdominal

subcutaneous adipose tissue and (ii) mesenteric adipose tissue (MAT) (a type of visceral adipose tissue).

During previous studies of MIDY liver tissue, it was detected that one of the MIDY samples didn't cluster very well with the other transgenic pigs (data not shown). It was later discovered that this pig had an insulinoma in the pancreas that was able to produce insulin. For this reason, this pig was not used in the study of the liver (**?**), and it was not used in this study of MIDY AT for this reason either.

The purified peptides from respective sample were analysed using a Q-Exactive HF-X mass spectrometer. Protein intensity quantification was made using MaxQuant's label free quantification (LFQ) method (**?**). Across all samples, a total of 2779 proteins were identified using this approach. Identifications were filtered for proteins with only one missing value in one sample group (a sample group is defined as one adipose tissue type plus the genotype of the samples, in total there are four groups). This filter only keeps the proteins that have been quantified with high confidence, limiting the number of uses of missing value imputation and increasing the robustness of the downstream statistical tests.

As can be seen in the principal component analysis (PCA) of the two first principal components (PCs) in **Figure 4.1**, the samples can be separated into two clusters containing the samples that originated from mesenteric and subcutaneous adipose tissue (SCAT) respectively. Moreover, clusters between MIDY and WT within each AT cluster can be detected, but they are not as well defined as between the two tissue types.

To alleviate the small sample size of the Munich MIDY pig biobank, linear modelling using a 2-way ANOVA was performed in order to detect global proteomic changes in adipose tissue derived from diabetic and non-diabetic pigs. This approach will also be used to detect protein abundance changes between the second factor group, adipose tissue type, as well as the interaction effect between the genotype and adipose tissue factors.

As can be seen in the p-value histograms (using the p-values derived from the ANOVA) **Figure 4.2A**, there is an enrichment of low p-values for the adipose tissue factor and a smaller enrichment of low p-values for the genotype factor. Contrary, the p-value histogram for the interaction factor is almost uniformly distributed, meaning that we don't expect

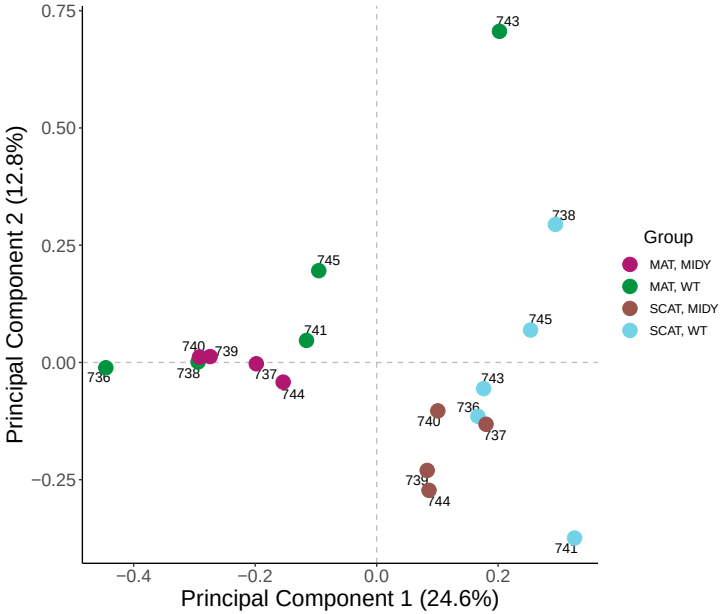


Figure 4.1. First two principal components of the adipose tissue MIDY pigs PCA. The percentages are the proportion of variance explained by each principal component.

to find many proteins to be significant for the interaction effect after multiple hypotheses testing. After pooling all p-values together, there's still a pronounced enrichment of low p-values (**Figure 4.2B)**, telling us that there are an enrichment of small p-values among all tested hypotheses and that FDR correction can be applied to the grouped set of p-values.

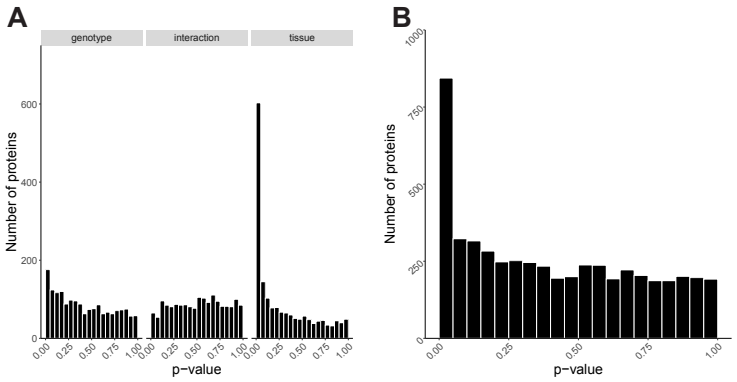


Figure 4.2. P-value histograms with p-values from the 2-way ANOVA. (**A**) Separate p-value histograms for each 2-way ANOVA factor (genotype, interaction, and tissue), (**B**) pooled p-value histogram using all p-values derived from the 2-way ANOVA.

4.1.1 Genotype effects

The 2-way ANOVA in conjunction with multiple hypotheses correction (using the qvalue (**?**) package in R) approach found 14 proteins to be significantly (FDR < 0.05) more abundant and 9 to be significantly less abundant in adipose tissue from MIDY pigs **(Figure 4.3A)**.

14 proteins were found to be upregulated in the adipose tissue of MIDY pigs, however one protein, tetratricopeptide repeat protein 38 (TTC38), also had a significant interaction effect (FDR = 0.01). Thus, TTC38 will be considered for its significant Genotype x Tissue interaction effect.

Note that the log2 fold changes used here are not derived from the 2-way ANOVA's linear model, they are calculated from the original imputed dataset and should serve as a tool to elucidate the direction of the effect for a significant difference detected by the 2-way ANOVA.

Among the significant proteins with the highest abundance increase in MIDY adipose tissue were retinol dehydrogenase 16 (RDH16, l2fc = 4.84, FDR = 0.000007), carboxyl esterase 1 (CES1, L2FC = 2.21, FDR = 0.03),

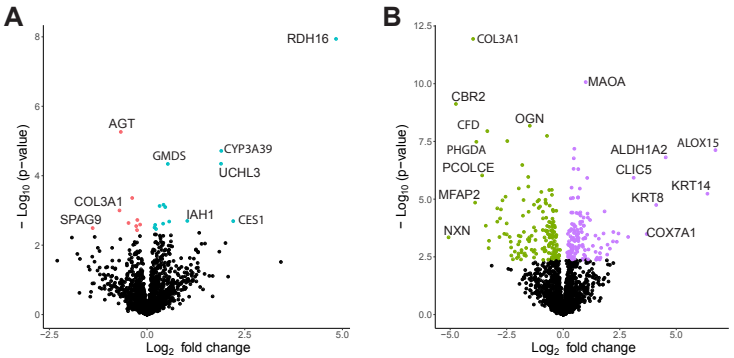


Figure 4.3. MIDY 2-way ANOVA volcano plot. (**A**) Volcano plot showing the results for the genotype factor, significant proteins (q-value < 0.05) enriched in AT of MIDY and WT pigs are colored in green and red respectively. (**B**) Volcano plot showing the results for the tissue factor, significant proteins (q-value < 0.05) are colored in purple for MAT and in yellow for SCAT.

cytochrome P450 3A39 (CYP3A39, l2fc = 1.90, FDR = 0.002), and ubiquitin carboxyl-terminal hydrolase 3(UCHL3, l2fc = 1.89, FDR = 0.003) (see **Table 4.1** for all proteins significantly more abundant in MIDY pigs).

RDH16 was identified in all MIDY samples (both from SCAT and VAT), but was only detected in one WT sample (derived from SCAT) at a lower intensity level. This explains the high log2 fold change and high significance (low FDR value) of RDH16, since LFQ values in WT adipose tissue were not quantified and had to be imputed at lower intensity levels (**Figure 4.4**). RDH16 has previously been found to be significantly more abundant in the liver of MIDY pigs (**?**).

The two subunits of the mitochondrial trifunctional protein (MTP), trifunctional enzyme subunit alpha, mitochondrial (HADHA, l2fc = 0.47, FDR = 0.02) and trifunctional enzyme subunit beta, mitochondrial (HADHB, l2fc = 0.42, FDR = 0.01) were also detected to be significantly more abundant in adipose tissue from MIDY pigs.

The proteins that had the significantly highest abundance decrease in MIDY adipose tissue were angiotensinogen (AGT, l2fc = -0.67, FDR = 0.0008), fructose-1,6-

Table 4.1. Significant proteins upregulated in adipose tisuee from MIDY pigs.

Gene name	Protein name	p-value	q
RDH16	retinol dehydrogenase 16	1.1E-08	7
CES1	liver carboxylesterase precursor	0.0020	
CYP3A39	cytochrome P450 3A39	1.9E-05	
UCHL3	ubiquitin carboxyl-terminal hydrolase	4.5E-05	
IAH1	isoamyl acetate-hydrolyzing esterase 1 homolog	0.0020	
AK2	adenylate kinase 2, mitochondrial	0.0021	
GMDS	GDP-mannose 4,6 dehydratase	4.6E-05	
HADHA	trifunctional enzyme subunit alpha, mitochondrial	8.0E-04	
HADHB	trifunctional enzyme subunit beta, mitochondrial	6.9E-04	
ACOT4	acyl-coenzyme A thioesterase 1-like	0.0024	
PSMD5	26S proteasome non-ATPase regulatory subunit 5	7.5E-04	
TTC38	tetratricopeptide repeat protein 38	0.0034	
PRKAR2A	cAMP-dependent protein kinase type II-alpha regulatory subunit	0.0026	
GALM	aldose 1-epimerase	0.0031	

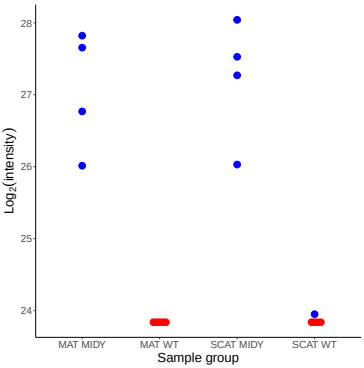


Figure 4.4. Distribution of observed values for RDH16. Intensity values are colored blue if RDH16 was detected in a sample and red if it was missing. Missing values were imputed by a constant value of 10% of the mean for all samples.

bisphosphatase 1 (FBP1, l2fc = -0.47, FDR = 0.03), hexokinase 1 (HK1, l2fc = -0.38, FDR = 0.01) and elongation factor 1-delta (EEF1D, l2fc = -0.28, FDR = 0.04) (**Table 4.2)**.

Table 4.2. Significant proteins downregulated in adipose tisuee from MIDY pigs.

Gene name	Protein name	p-value
SPAG9	C-Jun-amino-terminal kinase-interacting protein 4	0.0032
COL3A1	collagen alpha-1(III) chain precursor	0.0010
AGT	angiotensinogen	5.5E-06
FBP1	fructose-1,6-bisphosphatase 1	0.0023
HK1	hexokinase-1	0.0004
EEF1D	elongation factor 1-delta	0.0028
ATOX1	copper transport protein	0.0037
EIF3C	eukaryotic translation initiation factor 3 subunit C	0.0019
HSP90AA1	heat shock protein HSP 90-alpha	0.0025

4.1.1.1 Gene Ontology enrichment analysis using the STRING database

Gene ontology and KEGG pathway enrichment analysis was performed using STRING, (**?**) and functional annotations from *Homo sapiens* and the significantly differentially expressed proteins. The resulting Proteomaps are shown in **Figure 4.5**.

For the protein set that were upregulated in AT of MIDY pigs, CYP3A39 were replaced with its human paralog CYP3A4, since CYP3A39 doesn't exist in *H. sapiens*. Porcine CYP3A39 and human CYP3A4 share 77.4% identity (BLAST (**?**), a multiple sequences alignment is shown in **Supplementary Figure A.1**) and the two proteins share high homology in key active sites (**?**), it is thus probable that they have similar functions in the two different organisms. In the text I will still refer to the protein as CYP3A39.

Proteins that are significantly more abundant in MIDY pigs are involved in fatty acid (ACOT4, CES1, CYP3A4, HADHA, and HADHB) and lipid metabolic processes (ACOT4, CES1, CYP3A4, HADHA, HADHB, IAH1, and RDH16). RDH16 and CYP3A4 are both annotated as being part of the KEGG retinol metabolism

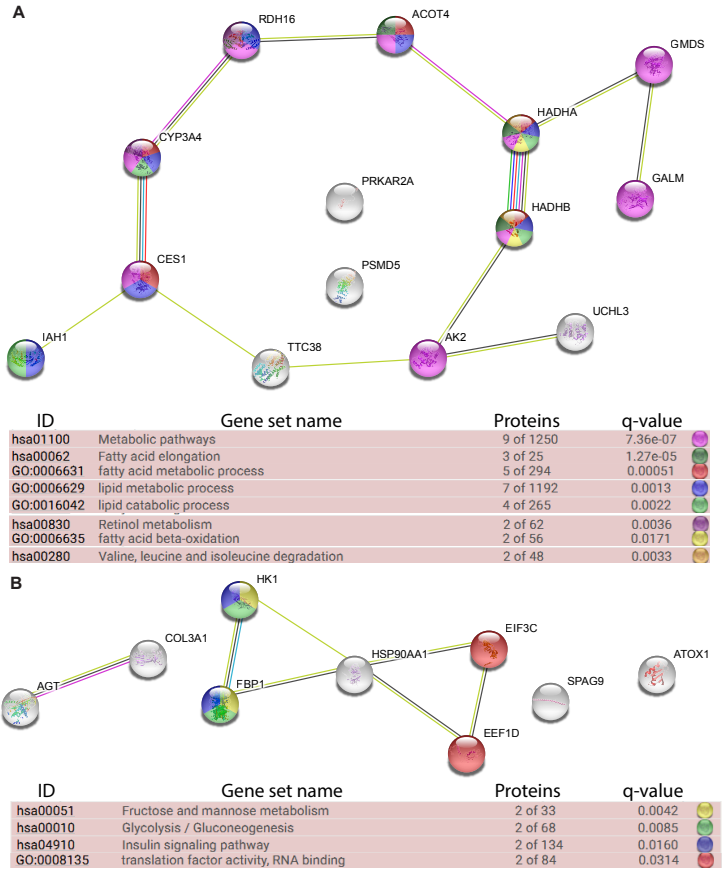


Figure 4.5. Gene set enrichment analysis using the STRING database. (**A**) GSEA using the significant proteins more abundant in AT from MIDY pigs and(**B**) WT pigs. Proteins belonging to gene set are colored according to the table, edges between proteins indicates that there is an interaction between the two proteins (an interaction score of 0.15 was used). The Proteins column indicates how many proteins in the

pathway, which is enriched in MIDY pigs according to the STRING analysis.

Interestingly, valine, leucine and isoleucine metabolism was also found to be enriched in MIDY pigs, as both MTP subunits were annotated as being involved in this KEGG pathway. The high confidence of a functional interaction between HADHA and HADHB is not surprising since they are both subunits of the same protein complex (**?**).

In the smaller protein set of proteins that were downregulated in MIDY pigs, both FBP1 and HK1 are annotated by KEGG to be part of fructose and mannose metabolism, glycolysis/gluconeogenesis and the insulin signalling pathway. Both EEF1D and EIF3C have translation factor activity and both proteins bind to mRNA.

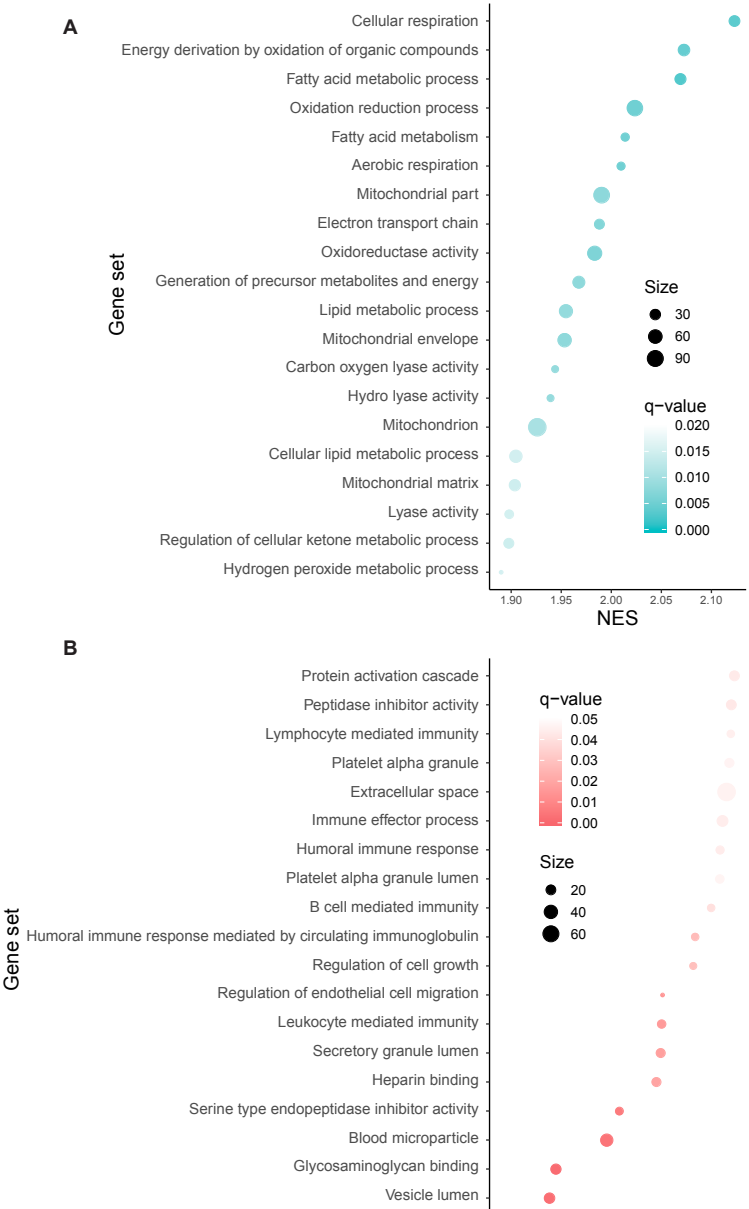
4.1.1.2 Gene set enrichment analysis

The proteins were ranked according to their $-log_{10}$ p-values, and the log-transformed p-values were assigned a minus sign if the L2FC was negative or a plus sign if it was positive. A preranked GSEA analysis (**?**) using 1734 gene sets (minimum gene set size = 10 and maxi-

mum size = 500) found that 54 gene sets were enriched for MIDY pigs (**Figure 4.6 A, Supplementary table A.3**) and 25 gene sets were enriched for WT at q-value < 0.05 (**Figure 4.6 B, Supplementary table A.4, Figure 4.7**).

Among the 54 gene sets enriched in MIDY pigs were 6 GOCC categories related to mitochondria (mitochondrial part, mitochondrial envelope, mitochondrion, mitochondrial matrix, inner mitochondrial membrane protein complex, and mitochondrial protein complex), 5 gene sets related to fatty acid metabolism (fatty acid metabolic process, fatty acid metabolism, fatty acid catabolic process, regulation of fatty acid metabolic process, and fatty acid β-oxidation), as well as 5 GOBP gene sets involved in lipid metabolic processes (lipid metabolic process, cellular lipid metabolic process, lipid oxidation, cellular lipid catabolic process, and lipid modification).

The enriched GO classes were then trimmed to remove mainly redundant classes using the software GO trimming (**?**). After trimming, 31 GO classes remained (21 GOBP, 8 GOCC, and 6 GOMF respectively) for enrichments in MIDY pigs. Among the enriched GOBP classes (**Figure 4.8**) there are a 16 dif-


A
Cellular respiration
Energy derivation by oxidation of organic compounds
Fatty acid metabolic process
Oxidation reduction process
Fatty acid metabolism
Aerobic respiration
Mitochondrial part
Electron transport chain
Oxidoreductase activity
Generation of precursor metabolites and energy
Lipid metabolic process
Mitochondrial envelope
Carbon oxygen lyase activity
Hydro lyase activity
Mitochondrion
Cellular lipid metabolic process
Mitochondrial matrix
Lyase activity
Regulation of cellular ketone metabolic process
Hydrogen peroxide metabolic process
Gene set
Size
30
60
90
q−value
0.020
0.015
0.010
0.005
0.000
1.90
1.95
2.00
2.05
2.10
NES
B
Protein activation cascade
Peptidase inhibitor activity
Lymphocyte mediated immunity
Platelet alpha granule
Extracellular space
Immune effector process
Humoral immune response
Platelet alpha granule lumen
B cell mediated immunity
Humoral immune response mediated by circulating immunoglobulin
Regulation of cell growth
Regulation of endothelial cell migration
Leukocyte mediated immunity
Secretory granule lumen
Heparin binding
Serine type endopeptidase inhibitor activity
Blood microparticle
Glycosaminoglycan binding
Vesicle lumen
Gene set
q−value
0.05
0.04
0.03
0.02
0.01
0.00
Size
20
40
60

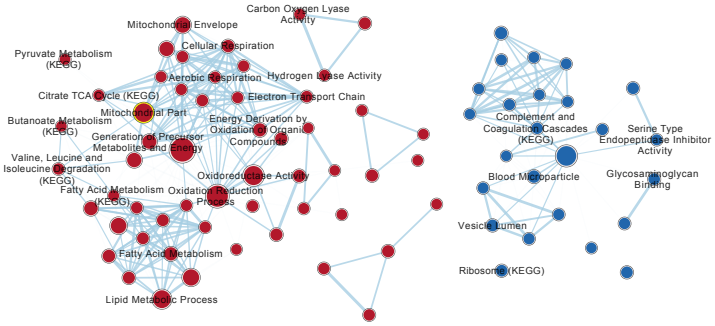


Figure 4.7. Network map showing significant (q-value < 0.05) GO ontology and KEGG categories enriched in MIDY (red) and WT (blue) pigs using GSEA. Annotated categories were either KEGG pathways or GO cateories significant at q-value < 0.01. The size of the nodes reflect the gene set size and the size of the edges reflects similarity between the two gene sets. Network map was made using the Cytoscape plugin Enrichment Map.

ferent clusters. There are two major clusters, of which the first includes lipid metabolic process, cellular respiration, and oxidation-reduction process and the second includes fatty acid metabolic process, organic acid metabolic process, primary alcohol metabolic process, and tricarboxylic acid metabolic process. Two smaller clusters include processes regulating fatty acid and ketone metabolism and anion transport. 5 gene sets doesn't belong to a single cluster, they include response to hydrogen peroxide, positive regulation of canonical Wnt signaling pathway, generation of precursor metabolites and energy, hydrogen peroxide catabolic process, and reactive oxygen species metabolic process.

Interestingly, 21 proteasomal subunits were found to be included in at least one enriched category (including PSMD5 that were found to be significant in the 2-way ANOVA). These proteins were mainly found in GO categories such as positive regulation of wnt signaling and in regulation of ketone metabolism. Cellular respiration was found to be enriched, as well as gene sets that involved response and breakdown of hydrogen peroxide and processes related to reactive oxygen species (ROS).

If one looks beyond categories that were enriched

at q-value < 0.05, additional categories of interest can also be found to be enriched in MIDY pigs such as response to insulin that was enriched at q-value $= 0.076$ and lipid catabolic process (q-value $= 0.050$).

There were also 5 KEGG pathways that were detected to be enriched in AT from MIDY pigs, they were: fatty acid metabolism, citrate cycle TCA cycle, valine, leucine and isoleucine degradation, butanoate metabolism, and propanoate metabolism .

After trimming the 23 GO gene sets enriched in AT from WT pigs, 8 GOBP, 4 GOCC, and 3 GOMF classes remained. For biological processes there was an enrichment of gene sets related to immunological responses, and regulation of cell migration, cell growth, wounding, and protein maturation. Enriched cellular compartments included vesicle lumen and extracellular space. There were two KEGG pathways among the enriched gene sets, ribosome and complement and coagulation cascade.

4.1.2 Tissue effects

The distribution of observed proteins in the four different sample groups is shown in **Figure 4.9**. As can

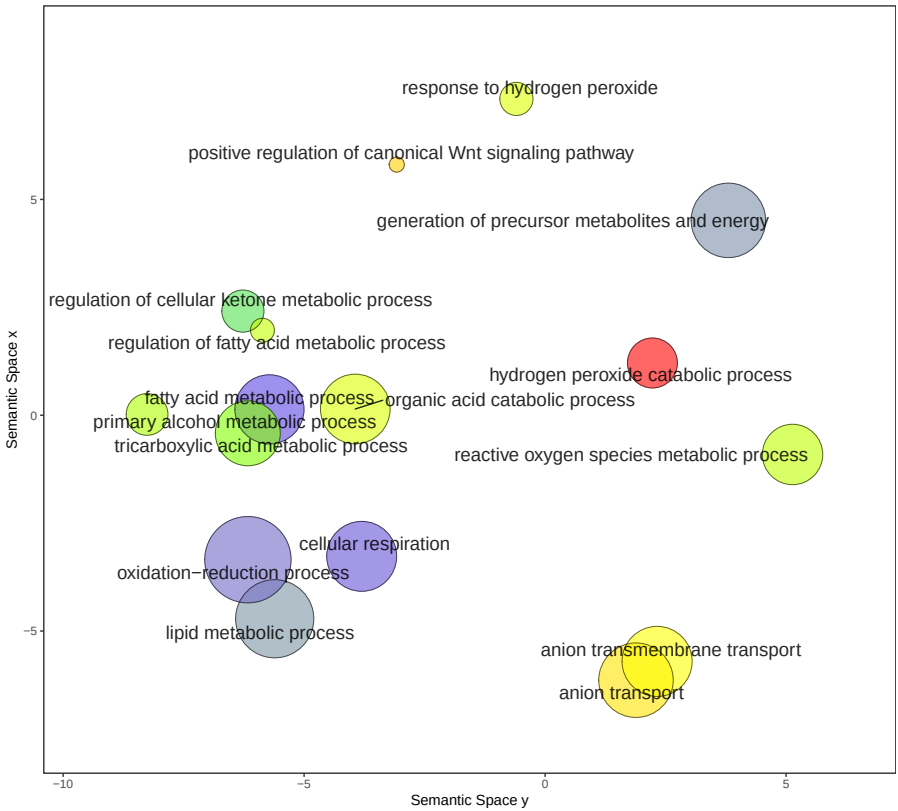


Figure 4.8. Significant GOBP categories in MIDY after trimming away redundant categories. The size of the circles are the log percentages of genes annotated with the GO term in the Uniprot database. The colors are q-values. The distance between two terms are a relative to the how semantically similar the two terms are, i.e. the closer two terms are the more similar they are.

be seen there aren't many proteins that are observed in one genotype group but not in the other, whereas there are considerably more that is only found in one of the tissue groups. Thus, we expect more statistical significant differences in the tissue factor in the 2-way ANOVA analysis.

The 2-way ANOVA found 302 proteins to be significantly differentially expressed after multiple hypotheses testing for the adipose tissue factor at FDR < 0.05. Of these, 141 proteins had a l2fc > 0, meaning that they were more abundant in mesenteric adipose tissue. Conversely, 161 of the significant proteins had a negative l2fc (**Figure 4.3B**, indicating that they had a higher expression in subcutaneous adipose tissue). Here, I will describe the most significantly altered proteins for each adipose tissue group (tables with all significantly altered proteins can be found in the **Supplementary Tables A.1** and **A.2**).

As can be seen in the volcano plot (**Figure 4.3B**, there are considerably more proteins that have larger l2fc (> 3) than in the genotype comparison. Among the proteins significantly more abundant in MAT with the highest l2fc inludes arachidonate 15-lipoxygenase (ALOX15, l2fc = 6.73, FDR = 0.000026), cytoskele-

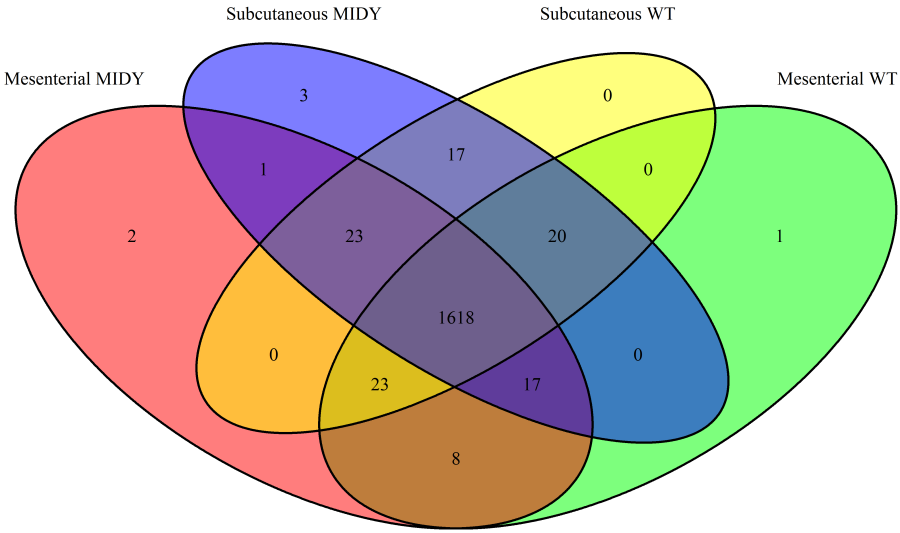


Figure 4.9. Venn diagram showing number of observed proteins in each group. The data were filtered allowing for 1 missing value in at least one group. A protein counts as observed if it has been detected in at least one sample in a group.

tal keratine type I 19 (KRT19, l2fc = 6.38, q-value = 0.00077), Aldehyde dehydrogenase 1 family, member A2 (ALDH1A2, l2fc = 4.54, FDR = 0.000049), keratin type II cytoskeletal 8 (KRT8, l2fc = 4.11 , q-value = 0.0015), cytochrome c oxidase subunit 7A1, mitochondrial precursor (COX7A1, l2fc = 3.69, q-value = 0.0099), and chloride intracellular channel protein 5 (CLIC5, l2fc = 3.12, q-value = 0.00023).

Table 4.3 shows the 15 proteins with most significant FDR values among the significant proteins that had a positive l2fc (more abundant in mesenteric adipose tissue). Among these proteins, the most significant proteins were amine oxidase [flavine containing] A (MAOA, l2fc = 1.00, FDR = 0.00000016), heat shock protein HSP 90-alpha (HSP90AA1, l2fc = 0.49, FDR = 0.000025), ALOX15 (l2fc = 6.73, q-value = 2.64E-05), ALDH1A2 (l2fc = 4.54, q-value = 4.90E-05), and component of 2-oxoglutarate dehydrogenase complex, mitochondrial precursor DLST, l2fc = 0.47, q-value = 4.99E-05).

Interestingly, five proteins that were significantly more abundant in MIDY pigs were also significantly more abundant in mesenteric adipose tissue (AK2, ACOT4, HADHA, HADHB, and PSMD5). Four of these pro-

Table 4.3. Top 15 most significant proteins more abundant in mesenteric adipose tissue.

Gene name	Protein name	p-
MAOA	amine oxidase [flavin-containing] A	8.5
HSP90AA1	heat shock protein HSP 90-alpha	6.5
ALOX15	arachidonate 15-lipoxygenase	7.5
ALDH1A2	aldehyde dehydrogenase 1 family, member A1	1.5
DLST	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial precursor	1.6
HSP90AB1	heat shock protein HSP 90-beta	5.0
SUCLG2	succinate–CoA ligase	4.8
CLIC5	chloride intracellular channel protein 5	1.1
SOD2	superoxide dismutase 2, mitochondrial	1.2
CCT2	T-complex protein 1 subunit beta	3.1
PRDX5	peroxiredoxin-5, mitochondrial	3.4
KRT19	keratin, type I cytoskeletal 19	5.8
ALDH9A1	4-trimethylaminobutyraldehyde dehydrogenase	8.2
IVD	isovaleryl-CoA dehydrogenase, mitochondrial	8.6
PSME1	proteasome activator complex subunit 1	8.8

teins (all except PSMD5) are involved in metabolic processes (**Figure 4.5**). HSP90AA1 was also found to be significant for the genotype factor, but were more abundant in WT pigs.

Nucleoredoxin (NXN, l2fc = -5.07, q-value = 0.011), carbonyl reductase [NADPH] 2 (CBR2, l2fc = -4.75, FDR = 0.0000010, collagen alpha-1(III) chain (COL3A1, l2fc = -3.99, FDR = 0.0000000044), microfibrillar-associated protein 2 (MFAP2, l2fc = -3.90, q-value = 0.0013), D-3-phosphoglycerate dehydrogenase (PHGDH, l2fc = -3.84, q-value = 1.40E-05), procollagen C-endopeptidase enhancer 1 (PCOLCE, l2fc = -3.59, q-value = 0.00021), and complement factor D (CFD, l2fc = -3.36, q-value = 7.33E-06) were the the most significantly abundant proteins in SCAT (**Figure 4.3**).

Table 4.4 shows the 15 proteins with the most significant FDR values among the significant proteins that had a negative l2fc (more abundant in subcutaneous adipose tissue). In this group, COL3A1, CBR2, mimecan (OGN, l2fc = -1.48, FDR = 0.0000064) and CFD were the proteins with the most significant q-values .

Three proteins (AGT, COL3A1, EIF3C) found to be significantly more abundant in subcutaneous tissue

Table 4.4. Top 15 most significant proteins more abundant in subcutaneous adipose tissue.

Gene name	Protein name	p-va
COL3A1	collagen alpha-1(III) chain precursor	1.15E
CBR2	carbonyl reductase [NADPH] 2	7.56E
OGN	mimecan isoform X1	6.64E
CFD	complement factor D	1.14E
FERMT2	fermitin family homolog 2	1.82E
MMP2	72 kDa type IV collagenase precursor	3.06E
PHGDH	D-3-phosphoglycerate dehydrogenase	3.29E
SDR16C5	epidermal retinol dehydrogenase 2	3.33E
PCOLCE	procollagen C-endopeptidase enhancer 1	9.33E
DCN	decorin precursor	1.10E
LUM	lumican precursor	2.84E
SGCA	alpha-sarcoglycan precursor	3.41E
TUBB	tubulin beta chain	3.98E
SNX3	sorting nexin-3	4.05E
SORBS1	sorbin and SH3 domain-containing protein 1	4.89E

were also found to be significantly enriched in WT pigs.

4.1.2.1 GSEA: gene sets enriched in different AT depots

A similar preranked GSEA analysis was performed using the p-values for the AT factor of the 2-way ANOVA. 154 gene sets were enriched in MAT and 174 gene sets were enriched in SCAT (q-value < 0.05) (**Supplementary tables** A.5 and A.6).

Among the 154 enriched gene sets in pig MAT included 102 GOBP, 18 GOCC, 22 GOMF, and 11 KEGG gene sets (**Table A.5**). **Figure 4.10A** shows the top 20 most enriched gene sets in MAT. It can be seen that a majority of the top 20 most enriched gene sets are involved in metabolism and energy derivation or located in mitochondria. The most significantly enriched gene sets were mitochondrial part, mitochondrial matrix, cellular respiration, generation of precursor metabolites and energy, and energy derivation by oxidation of organic compounds.

Among the enriched GOBP categories were process involved in respiration and energy generation, such as cellular and aerobic respiration and electron transport

chain. Metabolic gene sets enriched in MAT were categories such as several gene sets related to fatty acid metabolism and lipid metabolism. Also the gene sets glucose and branched-chain amino acid metabolism were enriched in MAT. Other GOBP categories included categories involved in processes such as protection against oxidative damage, cofactor metabolic process, oxidation-reduction process, tricarboxylic acid metabolic process, and protein polyuqibuitination. Among the 18 GOCC categories enriched in MAT, 8 gene sets were related to mitochondria. Other GOCC categories of note were two proteasome related categories and respiratory chain complex.

The 11 KEGG pathways that were also enriched in MAT were mainly involved in metabolism and energy generation. These inluded TCA cycle, fatty acid metabolism, oxidative phosphorylation, butanoate, propanoate, and pyruvate metabolism, and valine, isoleucine, and leucine amino acid degradation. The KEGG pathways proteasome and peroxisome were also significantly enriched in pig MAT.

The MIDY and MAT significantly enriched categories share a large overlap with each other, with 46 out of 54 categories enriched in MIDY pigs also be-

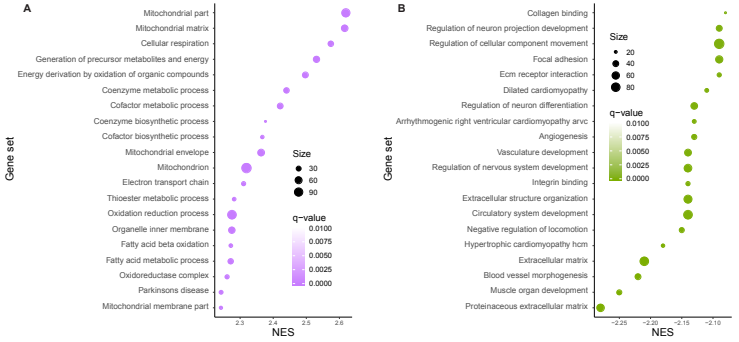


Figure 4.10. Top 20 most enriched gene sets from the tissue GSEA. (**A**) The top 20 most enriched gene sets in MAT and (**B**) the top 20 most enriched gene sets in SCAT. Gene sets are ordered according to their normalized enrichment score (NES). Size is the number of proteins detected in each gene set.

ing enriched in MAT at a FDR level of 0.05. The eight gene sets that were only enriched in MIDY but not MAT included: the KEGG pathway butanoate metabolism, and the GOBP categories primary alcohol metabolic process and regulation of fatty acid metabolic process.

The the top 20 most enriched gene sets in SCAT can be seen in (**Figure 4.10**). The most significantly enriched gene sets are collagen binding, regulation of neuron projection development, regulation of cellular component movement, focal adhesion, and ECM receptor interaction. Among the 174 gene sets enriched in SCAT, several gene sets related to the extracellular matrix was detected to be enriched here, as well as gene sets related to blood and angiogenesis. Gene sets related to the cytoskeleton were also enriched in SCAT, such as actin binding and actomyosin. Other features include gene sets involved in neuron and muscle development, integrin binding, cell adhesion, and skin development.

Among the 11 KEGG pathways significantly enriched in SCAT, were gene sets related to heart and muscle tissue, ECM receptor interaction, focal adhesion and gap junction, and endocytosis. There were

also two KEGG pathways related to cancer that were enriched in SCAT.

Only four out of the 25 gene sets enriched WT adipose tissue were detected as significantly enriched in SCAT. These four gene sets were glycosaminoglycan binding, secretory granule lumen, heparin binding, and extracellular space. This is considerably smaller than the overlap between the enriched gene sets detected in MIDY adipose tissue and MAT.

4.1.3 Interaction effects

As predicted from the p-value histograms (**Figure 4.2**) there were considerably fewer proteins significant for the Genotype x Tissue interaction factor than the non-interaction factors. Only five proteins were found to be significant for the interaction effect between genotype and adipose tissue after multiple hypotheses correction with FDR correction. The five proteins were were proliferating cell nuclear antigen (PCNA), RNA-binding protein with multiple splicing (RBPMS), non-specific lipid-transfer protein 2 (SCP2), TTC38, and vacuolar protein sorting-associated protein 29 (glsvps29).

Of these, PCNA and RBPMS were heavily imputed

(15 and 7 missing values respectively) and, when identified and quantified in a sample, were lowly expressed. RBPMS is involved in the processing of RNA and was significantly more abundant in MIDY MAT than in WT MAT. It was detected in all MIDY MAT samples, but only one WT MIDY sample (at a similar LFQ intensity as the MIDY samples). Since I can't find any information that links RBPMS to adipose tissue and diabetes (and that it was detected in all sample groups), it's possible that this significant difference is an artefact from quantification and missing value imputation. Contrary, SCP2, TTC38, and VPS29 were detected in all samples.

To elucidate which comparisons were significant for these 3 proteins, a Tukey's honest significant differences (HSD) test was performed. Tukey's HSD test compares the means of the different factor levels and adjusts p-values for multiple hypotheses by using family-wise error rate (FWER) within each group. We will focus on significant differences between comparable groups.

Table 4.5 shows the results of Tukey's HSD test, 7 comparisons were found to be significant at an adjusted p-value < 0.05 and meeting the above criteria. PCNA was significant in all comparisons includ-

ing MAT from MIDY samples. SCP2 was more significantly abundant in the SCAT from WT pigs when compared to SCAT from MIDY pigs (l2fc = 0.49, adjusted p-value = 0.0028) and it was also significantly less abundant SCAT when compared to MAT in MIDY pigs (l2fc = -0.40, adjusted p-value = 0.021). TTC38 was less significantly abundant in WT pigs when compared to MIDY pigs in SCAT (l2fc = -0.51, adjusted p-value = 0.00037) and VPS29 was less abundant in SCAT when compared to MAT in MIDY pigs (l2fc =-0.44, adjusted p-value = 0.00045).

4.2 GIPR pigs

In order to study the effects of liraglutide treatment on T2DM pigs, liver samples were investigated from 8 GIPR^dn pigs treated with liraglutide and 9 GIPR^dn pigs treated with placebo, in order to understand how liraglutide treatment affect the liver proteome of T2DM pigs. The samples were digested and the subsequent peptides were analysed using a Q Exactive HF-X mass spectrometer.

A total of 3404 protein groups were identified after

Table 4.5. Significant comparisons for the 5 proteins that were significant for the interaction factor in the 2-way ANOVA (q-value < 0.05). All significant comparisons with adjusted p-values from the Tukey HSD test are shown (adjusted p-value < 0.05).

Protein	Comparison	Adjusted p-value	Log 2 fold cha
PCNA	midy:sc-midy:m	0.000018	-2.41
PCNA	wt:m-midy:m	0.00011	-1.93
PCNA	wt:sc-midy:m	0.000282991	-1.77
TTC38	wt:sc-midy:sc	0.000372538	-0.51
VPS29	midy:sc-midy:m	0.000453679	-0.44
SCP2	wt:sc-midy:sc	0.002782554	0.49
SCP2	midy:sc-midy:m	0.021274227	-0.40
VPS29	midy:sc-wt:m	0.021551092	-0.26
TTC38	wt:sc-wt:m	0.023981781	-0.29
RBPMS	wt:m-midy:m	0.035444865	-1.45
VPS29	wt:sc-midy:m	0.035633954	-0.24
SCP2	midy:sc-wt:m	0.04432736	-0.33

filtering out potential contaminants and reverse hits without any prefractionation. Of these 2611 proteins could be quantified in at least 5 samples in at least one group, and missing values were imputed from the lower end of a normal distribution.

As the treatment groups included both male and female pigs (**Table 3.3**), an initial statistical test using a 3-way ANOVA was carried out to see how different factors affected differential protein expression. As can be seen in **Figure 4.11**, there were considerably more proteins found to be significant for the treatment factor after multiple hypotheses correction at a q-value $<$ 0.05 than for any other factor or interaction between factors. Based on this, it was decided to focus on the differences observed between the LT and PT treated GIPR^{d}n pigs. To increase the statistical power of the study, a linear model was applied to the protein LFQ intensities in order to filter out gender effects from the different proteins' expression values.

Comparing PCA made on the original LFQ intensities and the residuals from the linear model, it was seen that the linear model increased the proportion of variance explained by the principal components of the PCA (data not shown). As can be seen in the

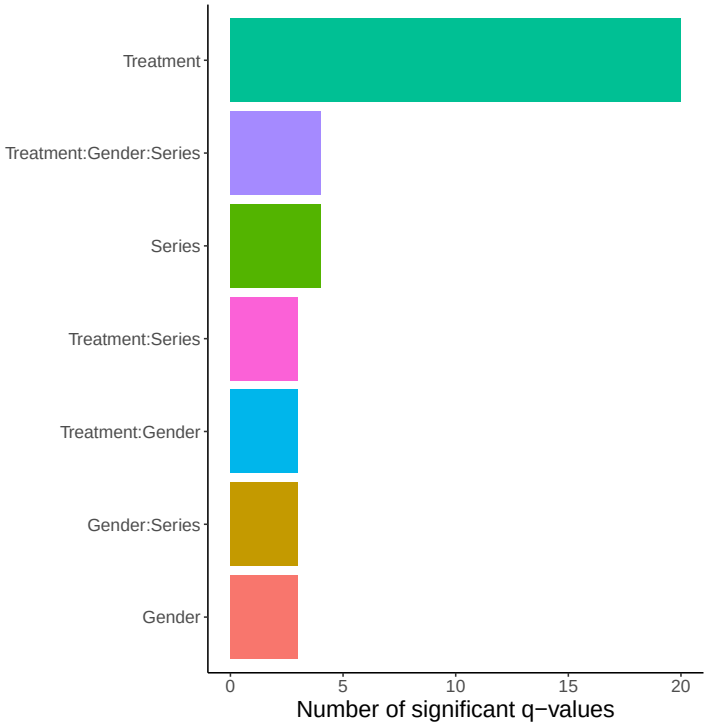


Figure 4.11. Number of proteins significantly altered in abundance after multiple hypotheses correction for the different factors and interactions. 3-way ANOVA was used to get p-values for each comparison. The series factor reflects the time when the experiments were performed.

PCA, the two first principal components separates the two different treatment groups into two separate clusters that defines the two different treatments (**Figure 4.12**).

After performing an equal variance Student's t-test on the resulting residuals from the linear model and adjusting the hypotheses using multiple hypotheses correction, a total of 127 proteins were identified as being differentially abundant at a significance level of q-value < 0.05 (**Figures 4.13, 4.14**).

4.2.1 Proteins and gene sets significantly enriched in liraglutide treated pigs

64 proteins were detected to be significantly more abundant in liver samples from liraglutide treated GIPR^{dn} pigs (**Supplementary Table B.1**). E3 ubiquitin-protein ligase RBX1 (RBX1) was the protein with highest fold change that was significant in liraglutide pigs (l2fc = 1.69, q-value = 0.0012). Other significant proteins with large log2 fold changes in liver from LT GIPR^{dn} pigs (**Table 4.6**) were the antiapoptotic protein anamorsin (CIAPIN1, l2fc = 1.28, q-value =

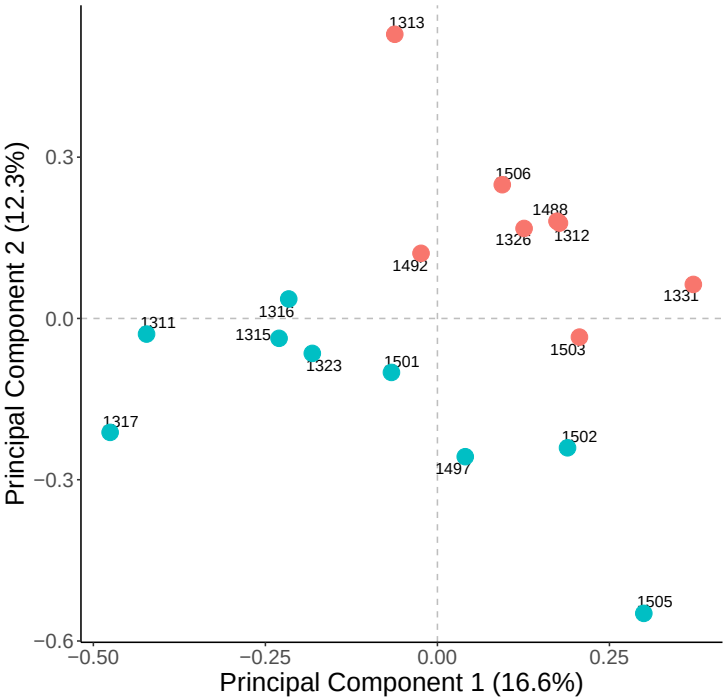


Figure 4.12. Dimension reduced visualization of the proteomic data from the liraglutide (red) and placebo (green) treated $GIPR^dn$ pigs using PCA. The plot shows the two first principal components.

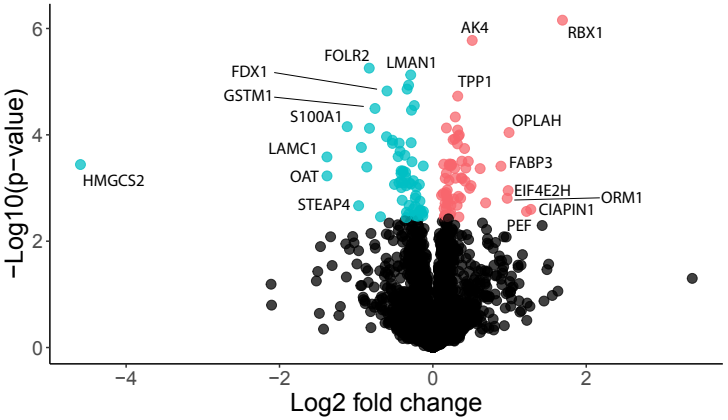


Figure 4.13. Volcano plot using the residual l2fc and p-values from the t-test using the residuals. Significant proteins (q-value < 0.05) are colored in red for proteins enriched in the liver of liraglutide treated pigs and green for proteins enriched in placebo treated pigs. Selected proteins are labelled.

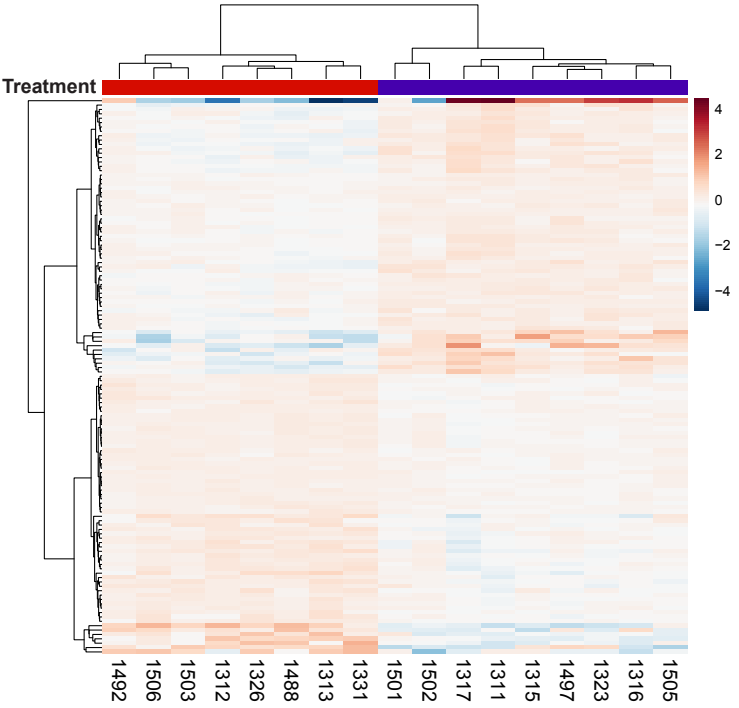


Figure 4.14. Heatmap showing LFQ values of proteins significantly altered in abundance between liver from LT and PL treated GIPR^{dn} pigs, residuals from the gender linear model and the ward.D2 clustering methods were used. Treatment annotation row: blue = PT and red = LT

0.042), the calcium binding protein peflin (PEF1, l2fc = 1.22, q-value = 0.044), 5-oxoprolinase (OPLAH, l2fc = 0.99, q-value = 0.0092), and eukaryotic translation iniation factor 4E type 2 (EIF4E2, l2fc = 0.98, q-value = 0.025).

Table 4.6. Top 10 significantly more abundant proteins in liver samples from liraglutide treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value
RBX1	E3 ubiquitin-protein ligase RBX1	1.69	6.98E-07
CIAPIN1	anamorsin	1.28	0.0025
PEF1	peflin	1.22	0.0028
OPLAH	5-oxoprolinase	0.99	9.05E-05
EIF4E2	eukaryotic translation initiation factor 4E type 2	0.98	0.0011
ORM1	alpha-1-acid glycoprotein	0.97	0.0016
FABP3	fatty acid-binding protein, heart	0.89	3.90E-04
GPAM	TRPM8 channel-associated factor 2-like	0.69	0.0019
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	0.62	4.30E-04

Table 4.6. Top 10 significantly more abundant proteins in liver samples from liraglutide treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value
AK4	adenylate kinase 4, mitochondrial	0.51	1.68E-06
GARS	glycine–tRNA ligase	0.5	9.00E-04

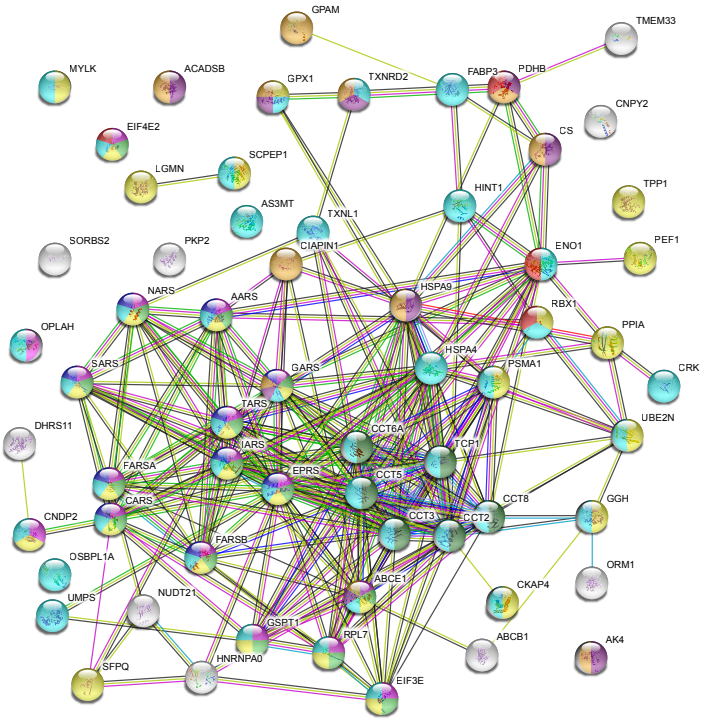
The 64 proteins enriched in liraglutide treated GIPR^{dn} were submitted to the STRING database to detect correlation between the proteins and GO and KEGG gene set enrichments (**Figure 4.15**). As can be seen in **Figure 4.15** the majority (41 proteins) of the enriched proteins were annotated as being cytosolic according to GOCC, there were also 11 proteins that were from the mitochondrial part. Interestingly, 6 enriched proteins were part of the chaperonin-containing ring-complex, which mediates protein folding.

Moreover, 15 proteins were annotated as being part of the translation GOBP gene set. Of these 15 proteins 10 were various aminoacyl-tRNA ligases. The majority of the proteins are involved in various metabolic

processes, such as the GOBP gene sets: cellular amide metabolic process (22 proteins) and protein metabolic process (29 proteins). 4 proteins (RBX1, EIF4E2, PDHB, and ENO1) were part of the hypoxia-induced factor 1 (HIF1) signalling pathway from the KEGG annotation database.

4.2.2 Proteins and gene sets significantly enriched in placebo treated pigs

In placebo treated GIPR^{dn} pigs, 63 proteins were found to be significantly more abundant in the liver (**Supplementary Table B.2**). The most significantly differentially expressed proteins was hydroxymethylglutaryl-CoA synthase, mitochondrial (HMGCS2, l2fc = -4.60, q-value = 0.015). HMGCS2 had the highest absolute log2 fold change of all proteins (**Figure 4.13**). Other proteins that were among the most significantly enriched proteins in placebo treated GIPR^{dn} were laminin subunit gamma-1 (LAMC1, l2fc = -1.38, q-value = 0.014), ornithine aminotransferase (OAT, l2fc = -1.38, q-value = 0.018), probable calcium ion signal transducer protein S100A1 (S100A1, l2fc = -1.12, q-value = 0.0087),



ID	Gene set name	Proteins	q-value
hsa00970	Aminoacyl-tRNA biosynthesis	10 of 44	8.33e-14
GO:0043043	peptide biosynthetic process	17 of 386	1.59e-12
GO:0006412	translation	15 of 362	1.07e-10
GO:0005832	chaperonin-containing T-complex	6 of 11	1.22e-09
GO:0005829	cytosol	41 of 4958	1.64e-09
GO:0019538	protein metabolic process	29 of 4194	0.00032
GO:0044429	mitochondrial part	11 of 1015	0.0041
hsa04066	HIF-1 signaling pathway	4 of 98	0.0128

Figure 4.15. Gene set enrichment analysis using the STRING database for proteins enriched in the liver of LT GIPR^{dn} pigs. Proteins belonging to a gene set are colored according to the table, edges between proteins indicates that there is an interaction between the two proteins (an interaction score of 0.4 was used). The Proteins column indicates the number of proteins in the given dataset detected for a given gene set.

and metalloreductase STEAP4 (STEAP4, l2fc = -0.97, q-value = 0.037).

Table 4.7. Top 10 significantly more abundant proteins in liver samples from placebo treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value
HMGCS2	hydroxymethylglutaryl-CoA synthase, mitochondrial precursor	-4.60	3.60E-04
LAMC1	laminin subunit gamma-1 precursor	-1.38	2.60E-04
OAT	ornithine aminotransferase, mitochondrial	-1.38	5.90E-04
S100A1	protein S100-A1	-1.12	7.02E-05
STEAP4	metalloreductase STEAP4	-0.97	0.0022
DHRS7	dehydrogenase/reductase SDR family member 7	-0.93	1.70E-04
GNMT	glycine N-methyltransferase	-0.86	4.00E-04
FOLR2	folate receptor 2	-0.83	5.57E-06
GATM	glycine amidinotransferase, mitochondrial	-0.83	7.53E-05
GSTM1	glutathione S-transferase Mu 1	-0.75	3.19E-05

Table 4.7. Top 10 significantly more abundant proteins in liver samples from placebo treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value

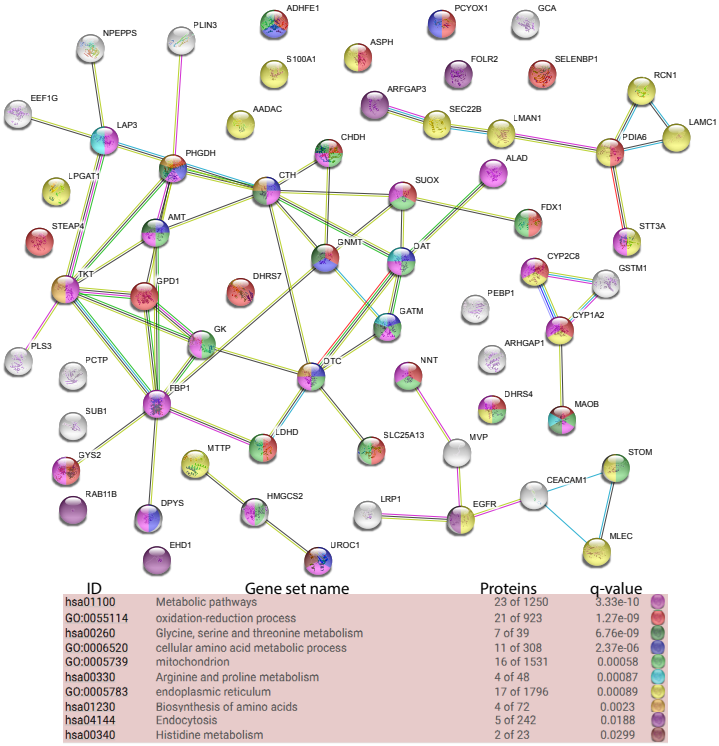
To find enrichments and correlations among the proteins enriched in the placebo treated group, the 63 proteins enriched were uploaded to the STRING database. However, no corresponding human gene called CYP2C33 was found. To include the finding when performing GSEA with STRING, the closest human homolog was included CYP2C8 (the two proteins share $> 60\%$ homology).

The majority of the proteins were, similarly to the proteins more abundant in the liraglutide group, part of metabolic processes (53 proteins). Enriched GOBP gene sets included oxididation-reduction process (21 proteins), cellular amino acid metabolic (11 proteins) and catabolic processes (6 proteins) (**Figure 4.16)**. 4 GOBP gene sets related to lipid and fatty acid metabolic processes were also significant, they were lipid metabolic process, lipid biosynthetic process, cellular lipid catabolic

process, and long-chain fatty acid biosynthetic process.

Also among this set of proteins cytoplasmic part was also a common GOCC (60 proteins). But 49 proteins also had a GOCC annotation named organelle part, including 16 proteins annotated as being located in the mitochondrion and 17 proteins as being located in the ER.

Among significantly enriched KEGG pathways were metabolic pathways (23 proteins) and endocytosis (5 proteins). There were four KEGG gene sets that were involved in amino acid metabolism that were enriched in the liver of placebo treated pigs. They were glycine, serine, and threonine metabolism (7 proteins), arginine and proline metabolism (4 proteins), biosynthesis of amino acids (4 proteins), and histidine metabolism (2 proteins). 5 proteins upregulated in placebo treated pigs were enriched in the KEGG pathway endocytosis. Other KEGG pathways of note were carbon metabolism (AMT, FBP1, TKT, and PHGDH), retinol metabolism (CYP1A2, CYP2C8, and DHRS4), terpenoid backbone biosynthesis (HMGCS2 and PCYOX), linoleic acid metabolism (CYP1A2 and CYP2C8), and pentose phophate pathway (FBP1 and TKT).



ID	Gene set name	Proteins	q-value
hsa01100	Metabolic pathways	23 of 1250	3.33e-10
GO:0055114	oxidation-reduction process	21 of 923	1.27e-09
hsa00260	Glycine, serine and threonine metabolism	7 of 39	6.76e-09
GO:0006520	cellular amino acid metabolic process	11 of 308	2.37e-06
GO:0005739	mitochondrion	16 of 1531	0.00058
hsa00330	Arginine and proline metabolism	4 of 48	0.00087
GO:0005783	endoplasmic reticulum	17 of 1796	0.00089
hsa01230	Biosynthesis of amino acids	4 of 72	0.0023
hsa04144	Endocytosis	5 of 242	0.0188
hsa00340	Histidine metabolism	2 of 23	0.0299

Figure 4.16. Gene set enrichment analysis using the STRING database for proteins significantly enriched in placebo treated pigs. Proteins belonging to a gene set are colored according to the table, edges between proteins indicates that there is an interaction between the two proteins (an interaction score of 0.4 was used). The Proteins column indicates the number of proteins detected in a given gene set. Gene sets are either from the GOBP, GOCC or KEGG databases.

4.2.3 GSEA analysis using all proteins quantified in the pig liver

A GSEA analysis was also performed using all proteins and their respective residual values using GSEA 4.0.1. The GSEA analysis found 189 gene sets to be enriched in the liver of liraglutide treated GIPR^{dn} pigs and 20 gene sets that were enriched in placebo treated GIPR^{dn} pigs (**Supplementary Figure B.1, Supplementary Tables B.3 and B.4**).

The top 21 most enriched gene sets in LT pigs can be seen in **Figure 4.17A**). Similarly to the STRING database analysis, the KEGG pathway aminoacyl tRNA biosynthesis was the most significantly enriched gene set in liver from liraglutide treated GIPR^{dn} pigs (NES = 2.60, q-value = 0, **Figure 4.17A, Supplementary Table B.3**). As can be seen in **Figure 4.18** the expression of the amino acid tRNA ligases are heavily skewed towards the liraglutide group. Other enriched KEGG pathways in liraglutide treated pigs included proteasome (NES = 2.27, q-value = 0), also other GO gene sets related to the proteasome and ubiquitination), lysosome (NES = 1.96, q-value = 0.002), spliceosome (NES = 1.86, q-value = 0.009), oocyte meiosis

(NES = 1.75, q-value = 0.02), and WNT signaling pathway (NES = 1.67, q-value = 0.04).

Since there were so many gene sets enriched in this group, I will briefly summarise the results here. There were many gene sets involved in regulation of cell cycle and/or cell cycle transition. Other "housekeeping" gene sets enriched in liraglutide treated pigs included RNA metabolic process (NES = 2.36, q-value = 0) and several other related gene sets (including both metabolism of non coding and mRNA), as well as gene sets involved in translation initiation and translation. Gene sets related to respiration and electron transport chain were also found to be enriched in liraglutide treated pigs.

There were a few different signaling pathways enriched in the liraglutide group, e.g. interleukin 1 mediated signaling pathway (NES = 2.36, q-value = 0), innate immune response activating cell surface receptor signaling pathway (NES = 2.24, q-value = 0), and different Wnt signaling pathways. The interleukin 1 mediated signaling pathway included the most differentially expressed protein in the liraglutide group, RBX1.

Among the 20 gene sets that were enriched in livers from placebo treated GIPR^{dn} pigs, the most enriched

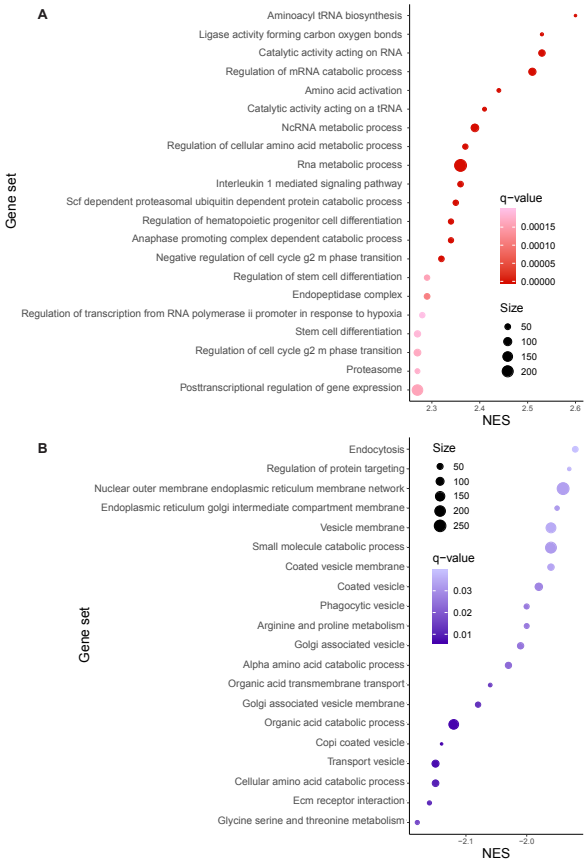


Figure 4.17. The most enriched gene sets in the GSEA analysis for each treatment group. (**A**) Top 21 most enriched gene sets in LT pig livers. (**B**) Top 20 most enriched gene sets in PT pig livers. Size = the number of proteins detected in the given gene set, color = q-value significance, and NES = normalized enrichment score.

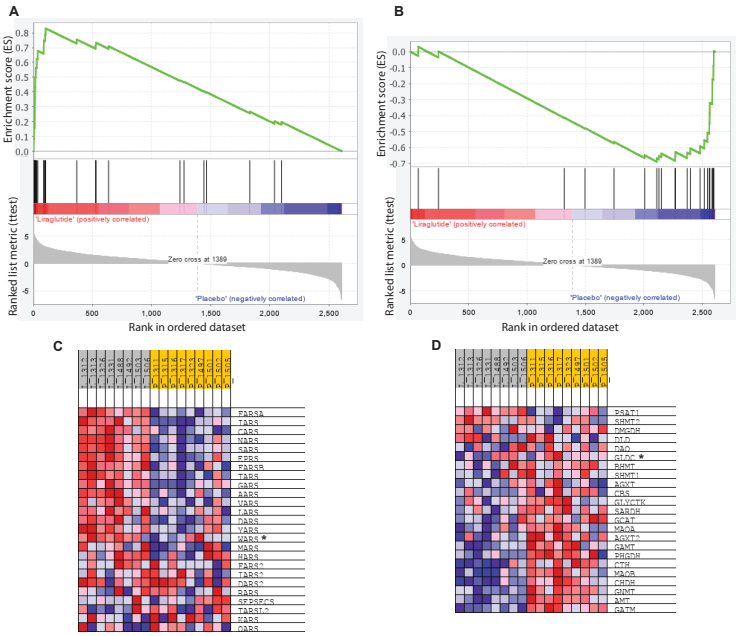


Figure 4.18. Examples of the most enriched gene sets in liraglutide and placebo treated GIPR^{dn} pigs. (**A**) Aminoacyl tRNA biosynthesis (KEGG) the most significantly enriched gene set in liver from liraglutide treated pigs and (**B**) glycine, serine, and threonine metabolism (KEGG), the most significantly enriched gene set in placebo treated pigs. Green line is the running ES score, black lines indicates a protein belonging to the gene set (a hit), and grey line is the ranking metric score. Heatmaps showing the quantified proteins belonging to the aminoacyl tRNA biosynthesis (**C**) and glycine, serine, and threonine metabolism (**D**). Heatmap colors indicates protein expression values, where red means high values and blue means low values. The asterisk indicates the first protein that is

gene set were the KEGG pathway glycine, serine, and threonine metabolism (NES = -2.18, q-value = 0.017) **Figure 4.17B, Supplementary Table B.4**). Proteins that are annotated as being part of the metabolism of the glycine, serine, and threonine amino acids are having L2FC values that show higher expression in the placebo group (**Figure 4.18B**). The KEGG pathway arginine and proline metabolism (NES = -2.00, q-value = 0.026) was also enriched in liver of placebo treated pigs. Other related metabolic gene sets that were enriched in placebo treated pigs were the GO categories: cellular amino acid catabolic process (NES = -2.15, q-value = 0.01), organic acid catabolic process (NES = -2.12, q-value = 0.007), alpha amino acid catabolic process (NES = -2.03, q-value = 0.023), and small molecule catabolic process (NES = -1.96, q-value = 0.032). Other amino acid metabolic pathways were enriched at lower levels of significance, such as tyrosine metabolism (NES = -1.85, q-value = 0.063). Lipid catabolic process were also enriched at a lower level (NES = -1.81, q-value = 0.069).

Two other KEGG pathways were enriched in placebo treated pigs, extracellular matrix (ECM) receptor interaction (NES = -2.16, q-value = 0.012) and endocy-

tosis (NES = -1.92, q-value = 0.039). There were also 9 gene sets related to vesicle transmembrane transport and vesicles enriched in placebo treated pigs at q-value < 0.05. In the vesicle related gene sets ADP ribosylation factor GTPase activating protein 3 (ARFGAP3) was the most enriched protein. Finally, 3 gene sets related to transmembrane protein transport and protein localization was also detected to be enriched in placebo treated pigs (ER Golgi intermediate compartment membran (NES = -1.95, q-value = 0.032), nuclear outer membrane ER membrane network (NES = -1.94, q-value = 0.033), and regulation of protein targeting (NES = -1.93, q-value = 0.037). In these 3 gene sets lectin mannose binding 1 (LMAN1) was the most enriched protein.

Chapter 5

Discussion

5.1 Differential proteome analysis of MIDY pig AT

Here, the proteome differences between littermate transgenic MIDY and WT pigs in two different AT depots, MAT and SCAT, were analyzed using mass spectrometry. In total 2779 protein groups were identified, a reasonable high number comparable to other proteomic studies using AT for mass spectrometry-based proteomics (**?**).

5.1.1 Proteomic differences between MAT and SCAT

There were considerably more significant differences found between MAT and SCAT than the number of significant differences between the two genotypes, reflecting the findings from both the p-value histograms (**Figure 4.2**) and that the two AT types were better separated in the PCA (**Figure 4.1**). This probably reflects differences in physiology and function between the two different AT depots.

5.1.1.1 Genes with large log2 fold changes

I will focus on significant proteins with large absolute l2fc values, as they are probably more linked to each AT types specialized characteristics. ALOX15 plays a role in the inflammatory response and has been found to have functions in adipogenesis and adipocyte differentiation with links to PPARγ, the adipocyte differentiation master regulator, activation (**?**). Dobrian and colleagues have previously described that ALOX15 was detected at high levels in human omental fat (VAT) and low levels in SCAT (**?**). Similarly, ALOX15 was

detected at high levels in all MAT samples, but only at a low level in one SCAT sample in the MIDY pigs. So ALOX15

Together KRT19 and KRT8 help link the contractile apparatus to dystrophin in muscles (**?**). KRT19 has been detected in VAT in rats (**?**), and at higher levels in omental AT than in SCAT in human men using DNA microarray (**?**) and in obese women using real-time PCR (**?**). KRT19 is a mesothelial marker protein (**?**). KRT8 is also a mesothelial biomarker and has been detected in VAT from mice using single cell RNA-sequencing (**?**).

ALDH1A2 converts retinal to atRA. Similar to the pigs in this study, ALDH1A2 mRNA expression is higher in VAT than in SCAT in humans
(**???**). The expression of ALDH1A2 is controlled by the mesothelial development and transcription factor Wilms tumor protein (**?**).

COX7A1 is a marker for brown/beige adipocytes (**??**). COX7A1 was only detected in MAT but only in seven out of nine MAT samples. Suggesting that pig MAT is more metabolically active than SCAT. RA increases COX7A1 expression in brown adipocytes (**?**). However, no significant differences were detected be-

tween the two genotypes in MAT, which could be due to low sample size.

NXN controls redox homeostasis in cells, and is important for adipocyte differentiation via the Wnt/β-catenin pathway (**?**). Overexpression of NXN leads to WAT lipid accumulation (**?**). Why NXN was detected in more SCAT than MAT samples remains unclear at this point.

CBR2 is an ortholog of mouse CBR2, a protein that doesn't seem to exist in humans (BLAST, data not shown). In mice, CBR2 is primarily expressed in lung and at lower levels in AT (**?**). CBR2 is tetrameric form unlike the monomeric carbonyl reductase forms, CBR1 and CBR3, that exist in humans as well (**?**). CBR2 is part of arachidonic acid metabolism, and is important for the synthesis of prostaglandins (**?**). CBR2 levels have previously been detected to be increased in peripheral blood in pregnant pigs (**?**). CBR2 was only detected in SCAT samples whereas CBR3 was detected in all samples, suggesting that CBR2 is important for SCAT metabolism. Interestingly, as mentioned above ALOX15 is also part of arachidonic acid metabolic pathway.

MFAP2 is an ECM protein that is involved in reg-

ulating energy expenditure in SCAT cells in both humans and mice, and MFAP2 transcript levels are higher in SCAT than in BAT (**?**). MFAP2 deficient mice have increased TGF-β activity, leading to reduced activity of BAT differentiation regulators PPARγ and PRDM16, and ultimately lower UCP1 expression in SCAT (**?**). However, no significant differences were detected between the genotypes which could be due to the study design since many samples had imputed intensity values.

5.1.1.2 Gene sets involved in metabolism and energy derivation were significantly upregulated in MAT

There were many significant gene sets related to metabolism and energy derivation in MAT when compared to SCAT. Gene sets related to lipid metabolism were enriched in MAT. It is known that MAT is more lipolytically active than SCAT, and MAT is aksi known to contribute more to the FFA plasma levels than SCAT (**?**). In obese individuals this difference is even more apparent, where increased lipolytic activity is linked to an upregulation of leptin and downregulation of adiponectin in

MAT (**?**). This could be a possible explanation for the upregulation seen in MIDY AT too, since MIDY also have an impaired glucose metabolism. But leptin was not detected in the proteomic study, so this can't be confirmed in this study.

Mitochondria are essential for fatty acid metabolism, with more than 98% of fatty acid oxidation occurring in mitochondria (**?**). It is known that VAT adipocytes have more mitochondria than SCAT adipocytes in both rats and humans (**??**). This explains the enrichment of gene sets related to fatty acid metabolism, mitochondria and cellular respiration seen in MAT.

That gene sets related to ROS and hydrogen peroxide metabolism were enriched in both MAT and MIDY can probably be explained by the higher metabolic activity in those groups than in their comparisons (SCAT and WT respectively). Mitochondrial hydrogen peroxide production is higher when fatty acids are used as substrates instead of carbohydrates (**?**). This might be an explanation for the increase in ROS related gene sets enriched in MIDY AT too.

Gene sets related to branched-chain (valine, leucine, and isoleucine) amino acids (BCAA) were enriched in both MAT and MIDY AT. VAT is known to catabo-

lize BCAA in mice (**?**), and BCAA ratios influence AT lipid metabolism in pigs
(**?**). BCAA has also been suggested to play a part in adipocyte differentiation (**?**). PPARγ is major regulator of plasma level BCAA and BCAA catabolism in both WAT and BAT in mice (**?**).

5.1.1.3 Gene sets upregulated in SCAT

Gene sets related to collagen and ECM were enriched in SCAT. Both of these categories have previously been found to be enriched in SCAT in a microarray and histological study comparing SCAT and WAT in rats, and SCAT might play a role in ECM development and maintenance via production of ECM proteins (**?**).

Many gene sets related to cytoskeleton, actin, myosins and cellular movement were also enriched in SCAT. Other gene sets included classes related to development/differentiation of different tissues, such as neurons and the nervous system, blood and the vasculature system, skin, and muscle. This might be explained by the location of SCAT near connective tissue and thus close to other tissues such as skin and muscle tissue.

5.1.2 Genotype effects

We decided to use a 2-way ANOVA as t-tests for each individual tissue group didn't yield any significantly, differentially expressed proteins after correcting for multiple hypotheses (at a level of q-value < 0.05). This is may be due to the limited number of available samples, preventing the detection of weaker abundance alterations in the analyzed proteomes. RDH16 was found to be significantly more abundant in MIDY pigs for both tissue groups when using the t-test method (at a lower significance threshold of q-value < 0.1), but it was the only protein that passed this more relaxed significance threshold. With a bigger sample size (e.g. 10 MIDY versus 10 WT pigs) a two-sample t-test would probably have yielded more significant results for each AT source at q-value < 0.05 as the statistical power would increase.

The 2-way ANOVA essentially doubles the statistical power of the study yielding groups with sample groups with 8-10 samples depending on the factor and group. In this case it seemed to have worked well, as we are detecting global proteomic changes in AT caused by the MIDY syndrome.

Another explanation for the relatively few observed differentially expressed proteins could be that the MIDY pigs were treated with insulin (**?**). The insulin treatment would potentially make the two pig cohorts more similar, which would result in fewer detected significant proteins. However, the diabetic pigs would probably not survive 2 years without insulin treatment. Even though this approach might have yielded fewer significant differentially expressed proteins, this approach gives insight into how insulin treatment affects diabetic patients.

Thirdly, it is probable that MIDY affects liver more than AT. Thus, it would be more likely to detect statistically significant differences between the sample groups in liver than in AT. This assumption seems to be correct given the data from Backman, *et al* (2019) (**?**). This method would probably find more significant differences in AT when comparing T2DM with WT pigs, since T2DM is more closely linked to obesity which affects AT more than T1DM and MIDY.

Assigning significance at a level of q-value < 0.1 could also be an alternative, with the drawback that 10% of the significant proteins would be false positives instead of 5%. For exploratory studies, this could be

an acceptable number of false positives in my opinion. However, the FDR level that is deemed significant should be decided during the experimental planning of the study and has to be decided on a case by case basis.

5.1.2.1 RDH16 is upregulated in AT in MIDY pigs and potential links to BAT

RDH16 was both the most significant differentially expressed protein upregulated in MIDY AT and had the largest L2FC, since it was detected in all MIDY samples but only in one WT sample (**Figure 4.4**). RDH16 catalyses the conversion of all-trans-retinol into all-trans-retinal (**?**). RDH16 has previously been found to be upregulated in liver tissue in MIDY pigs (**?**). Backman and colleagues also detected that the concentrations of retinal and retinoic acid was increased in the liver of MIDY pigs when compared to WT pigs (**?**). A transcriptomic analysis of four different ATs in MIDY pigs detected that RDH16 was the only protein consistently upregulated in MIDY AT (Backman, unpublished data). This suggests that RDH16 might be globally upregulated in MIDY pigs. However, more tissues from the Munich-MIDY pig biobank need to

be analysed to establish this, especially other organs important in the context of diabetes such as skeletal muscle and intestine. Similarly to the Backman, *et al.* study (**?**), it would also be of great interest to confirm if all-trans-retinal and atRA are also more abundant in MIDY AT tissue. A similar metabolic method as the one established in the liver study to measure all-trans-retinol derivatives could be used.

Obrochta, *et al*, found that insulin suppresses RDH16 expression in a hepatoma cell line through inhibition of FOXO1 (**?**). Thus, insulin seems to be an important negative regulator in both AT and liver. Both atRA and all-trans-retinal are known to interact with RXR and RAR transcription factors (**?**), and atRA is known to affect the gene expression of a multitude of different genes (**???**). There are many links between retinoid metabolism and diabetes (**?**). Furthermore, retinal present in rodent AT inhibits adipogenesis (**?**), which may partially explain why RDH16 is more abundant in AT from MIDY pigs. But further studies would be needed to elucidate this.

Recently, Krois and colleagues discovered that RDH16 suppresses obesity in mice by increasing the amount of BAT present in AT (**?**). Furthermore, loss of RDH16

affected the expression of over 400 different gene transcripts, leading to changes in whole-body metabolism. They also found that RDH16 was expressed in tissues with high metabolic activity, such as liver and brown adipose tissue (BAT), but not in epididymal WAT. The ablation of RDH16 leads to decreased lipid metabolic activity, abnormal mitochondria function, and lower respiration rates. An RDH16 KO pig model would be of interest to see if the effects of RDH16 deletion seen in mice would also be translated into pigs. Similarly, proteomic and transcriptomic analysis of important diabetic tissues from the Akita mice would also be of interest to establish whether the effects seen on RDH16 in MIDY pigs also occurs in the mouse model.

In mice, transplantation of BAT has been shown to improve glucose homeostasis and reverse the T1DM phenotype (**?**). The link between RDH16 and BAT is interesting (**?**), since the question whether pigs have BAT (unlike many other mammals) is a contested research topic. In the 80s, morphological studies showed evidence for BAT or browning of WAT in pigs (**?**). However, with the advent of genomic methods it was discovered that pigs lack a functional uncoupling pro-

tein 1 UCP1, which is a biomarker for BAT (**??**). UCP1 is known to be regulated by atRA (**?**). Despite the lack of UCP1, pigs has been proposed to be valuable animal model for human energy expenditure since adults and obese humans have low expression of UCP1 compared to mice (**?**). Even though pigs don't have a functional UCP1 gene, the GSEA analysis of the MIDY pigs indicates that similar metabolic changes occurred in the AT of MIDY pigs as were observed by Krois, *et al* in mice (**?**). This includes an increase of proteins in MIDY pigs involved in lipid and fatty acid metabolism, cellular respiration, and an increase of mitochondrial proteins (**Figure 4.8**). So atRA might regulate genes involved in these processes in the AT of pigs, increasing fatty acid metabolism and respiration, independently of UCP1.

RA has been found to induce mouse WAT browning (**?**). In mice, positive regulatory domain containing 16 (*PRDM16*) gene is the dominant transcriptional regulator of the browning of white adipocytes (**???**), and RA increases PRDM16 transcription in mice (**?**). Even though pig lacks the UCP1 gene, it has been shown that overexpression of PRDM16 in porcine preadipocytes inhibited differentation into WAT and induced tran-

scription of genes responsible for lipolysis and fatty acid oxidation (**?**). E.g., the UCP1 homolog UCP3 was upregulated in porcine preadipocytes with overexpressed PRDM16 (**?**). Cold resistant pig strains have been shown to have non-shivering thermogenesis that is dependent on increasing expression of UCP3 and browning of WAT (**?**). Chen and Yu (**?**) found that bone morphogenetic protein 7 (BMP7) can trigger differentiation into "brown-like" adipocytes in pigs, with an increase of PRDM16, UCP2, and UCP3 expression, as well as an increase of the number of mitochondria in differentiated "brown-like" adipocytes. WAT of UCP1 knock-in pigs have lower fat content linked to a higher rate of lipolysis (**?**). The UCP1-3 genes have all been reported to be upregulated by retinoic acids (**??**). Whether PRDM16, UCP2 or UCP3 are upregulated in AT in MIDY pigs would need to be further studied to clarify if this could be an explanation for the RDH16 upregulation detected in MIDY pigs and if "browning" occurs. BAT contains more mitochondria than WAT and instead of single, large lipid droplet, BAT have many smaller lipid droplets (**?**). Thus histological analysis to study if there is an increase in mitochondria or more smaller lipid droplets in AT cells from

MIDY pigs or fluorescence studies targeting the UCP proteins or PRDM16 would be interesting to carry out.

5.1.2.2 Proteomic changes in proteins involved in fatty acid and lipid metabolism

CES1, CYP3A39, HADHA, HADHB, and ACOT4 were annotated as being part of the GOBP fatty acid metabolism gene set. Of these, carboxylesterase 1 (CES1) was found to be significantly more abundant in the adipose tissue from MIDY pigs and was the second significantly more abundant protein in MIDY AT after RDH16. Carboxylesterases are abundantly expressed in liver and AT (**?**), and are located primarily in the ER of many tissues where they are responsible for hydrolysing both endogenous and xenobiotic compounds (**?**), e.g. caboxylesterases have been proposed to hydrolyse retinyl esters, the precursors of all-trans-retinol (**??**). The mouse paralog CES1d is known to have triglyceride lipase activity in WAT (**?**), and human CES1 is also thought to be a part of lipid catabolic processes. CES1 expression is higher in adipose tissue from obese individuals than adipose tissue from lean

individuals in human subjects (**????**), but there were no significant expression differences detected between different adipose tissue depots
(**?**). Yang and colleagues (**?**), have recently shown that insulin suppresses transcription of *CES1 in vitro*. It has previously been shown that CES1 is significantly differentially regulated in BAT from non-obese and obese male rats (**?**). *In vitro* glucose restriction leads to an upregulation of CES1 in adipocytes (**?**). The CES1 mouse homolog CES1d is highly expressed in both WAT and BAT (**?**).

CYP3A39 was the third most significantly upregulated protein in MIDY AT. The cytochrome P450 superfamily comprises of known catabolic enzymes that can oxidize atRA, primarily CYP26 isoforms (**??**). Other cytochromes also have atRA oxidative activity, such as the CYP3A4
(**??**). Since pig CYP3A39 is highly homologous to human CYP3A4, there's a reason to believe that CYP3A39 may also have a role in the catabolism of retinoic acid. CYP3A activity in AT has previously been described, and cytochrome P450 enzyme activity and expression in pigs is generally similar to human physiology (**?**). Retinoids, including atRA, have been found to increase

the expression of CYP3A4 *in vitro* (**??**). CYP3A4 expression was induced by treating hepatocytes with diabetic rat serum (**?**) and CYP3A has been implicated in obesity in mice (**?**).

Both subunits of the mitochondrial trifunctional protein (MTP), trifunctional enzyme subunit alpha and subunit beta (HADHA and HADHB respectively), were detected as being more abundant in adipose tissue from MIDY pigs. The MTP protein complex is composed of four alpha subunits and four beta subunits, and is responsible for catalysing three out of four steps of chain-shortening reactions in mitochondrial β-oxidation of fatty acids (**?**). Ageing mice carrying a heterozygous loss-of-function mutation for both subunits develop insulin resistance (**??**). HADHA transcription has previously been shown to be upregulated during fasting in pigs, mice and rats (**?**). A study in mice using *hadh* KOs showed that HADH is important for adaptive thermogenesis, body weight maintenance and pancreatic β-cell insulin secretion (**?**). HADHA has previously been described as being regulated by RA (**?**) and both proteins are considered to be markers for beige and brown adipocytes (**?**). HADHB correlates with an increase in BAT mass in ground squirrels

during hibernation (**?**).

ACOT4 is an acyl-CoA thioesterase that hydrolyses various acyl-CoAs into FFAs and coenzyme A, and its function may be to promote peroxisomal β-oxidation
(**?**). The annotation for this protein group identifies the peptides as coming from either ACOT1 or ACOT4. Low expression of ACOT1 and ACOT4 has been detected in WAT, with higher expression of ACOT4 seen in BAT (**?**). The increased expression of ACOT4 may be explained by that fatty acid oxidation is increasingly important during diabetes (**?**). ACOT4 transcripts has previously been found to be upregulated in WAT from obese humans (**?**).

Of these proteins together with RDH16, 3 proteins (RDH16, CES1, and CYP3A39) have a function related to retinol metabolism. And a majority of these genes seems to have a connection to browning/beigeing of adipocytes.

IAH1 together with RDH16 and the 5 proteins mentioned above were found to be part of the MIDY enriched gene set lipid metabolic process. IAH1 is a lipase that is responsible for hydrolysing esters in *Saccharomyces cerevisiae* (**??**), but not much is known

about IAH1 protein in mammalians. It was identified in a quantitative trait locus QTL for fatty liver disease in mice, and it was subsequently shown that *in vitro* overexpression of IAH1 suppressed genes involved in triglyceride synthesis and lipid metabolism (**?**).

In the genotype GSEA analysis, the many gene sets related to mitochondria, respiration, metabolism, and oxidation in MIDY AT might suggest that there was an increase of mitochondria in MIDY AT cells, as well as an increase in respiratory and oxidative activity, features that are reflective of BAT.

5.1.2.3 Other proteins significantly more abundant in MIDY AT

UCHL3 has been shown to promote insulin signaling and is important for AT differentiation in mice, and UCHL3 deficient mice had less VAT relative to WT mice (**?**). Suzuki *et al* (**?**), also proposed that UCHL3 not only regulates AT differentiation by regulating insulin signaling, but also by regulating PPARγ activity, the master regulator of adipocyte differentiation, by deubiquitination of PPARγ. PPARγ is known to be degraded by the proteasome (**??**). UCHL3 has been

found to be upregulated in BAT during winter hibernation in ground squirrels (**?**).

The proteasomal subunit PSMD5 was significantly upregualted in MIDY AT, and 21 proteasomal subunits were detected to be detected in least one leading edge in gene sets enriched in MIDY AT. Alterations in the ubiquitin-proteasome pathway are important for adipocyte differentiation (**???**) and are linked to increased oxidative stress in AT (**?**), proteins involved in insulin signaling are also affected by the ubiquitin-proteasome pathway (**?**). NRF1 induced proteasome activity is important for BAT function (**?**) and RA is known to regulate proteasomal degradation activity (**???**).

AK2 is a mitochondrial enzyme that produces 2 ADP molecules from one ATP and one AMP molecule. AK2 is markedly induced during adipocyte and β-cell differentiation, and inhibition of AK2 leads to lower adiponectin secretion and activation of UPR during differentiation (**?**). AK2 is important for mitochondrial function and oxidation (**?**). Taken together, these results could indicate that there is indeed adipocyte differentiaton occuring MIDY AT and strengthen the case that there might be "browning" of the AT in

MIDY pigs. However, further studies is needed to establish this.

GDP-mannose-4,6-dehydratase GMDS is responsible for converting GDP-D-mannose to GDP-4-dehydro-6-deoxy-D-mannose. Recently, it was shown that giving mice oral supplements of D-mannose can suppress T1DM (**??**). The supplementation of D-mannose induces T-cell differentiation, and it was also observed that D-mannose increased ROS and fatty acid oxidation (**?**), processes that were significantly enriched in AT in MIDY pigs in the GSEA (**Figure 4.8**). It might be possible that GMDS has a role to play in this process, but further studies would be needed to determine this.

cAMP-dependent protein kinase type II-α regulatory subunit (PRKAR2A) was identified to be associated with adiponectin, a hormone produced in AT that increases insulin sensitivity and fatty acid oxidation (**?**), levels in an exome quantitative trait locus study (**?**). AtRA has been demonstrated to stimulate the cAMP/PKA signaling pathway (**??**). *Prkar2a* KO mice resist diet-induced obesity and related metabolic alterations (**?**). If fatty acid oxidation levels are increased in MIDY AT remains to be seen, but it would

be consistent with other findings and might support the theory of a "browning" of MIDY AT.

Aldose 1-epimarase (GALM) catalyses the reaction α-D-glucose or α-D-galactose to β-D-glucose or β-D-galactose. AtRA is known induce the transcription of GALM and other galactose-related genes (**?**). GALM was detected to be important for insulin signaling in mouse (pre)adipocytes in a multiomics study (**?**). Galactose intake may protect against adiposity and improve metabolic and overall health in female mice (**?**). Increased expression of GALM could be a response to a greater need for improved insulin signaling in MIDY pigs.

5.1.2.4 Gene sets enriched in MIDY but not MAT

Of the eight gene sets enriched in MIDY but not in MAT, six gene sets were similar or related to one or more gene sets detected to be significantly enriched in both MIDY and MAT. The two gene sets with a unique process enriched in MIDY were primary alcohol metabolic process and butanoate metabolism.

Primary alcohol metabolic process included other

proteins together with RDH16 involved in retinol metabolism, e.g. aldehyde dehydrogenase 1 family, member A1 (ALDH1A1) that catabolizes the conversion of all-trans-retinal to atRA and dehydrogenase/reductase SDR family member 4 (DHRS4) that can catalyse both the conversion from retinol to retinal and retinal to RA. This further strengthens the hypothesis that retinol metabolism and RA regulated gene expression might be important for diabetic adaption in MIDY pigs.

Proteins involved in the KEGG pathway butanoate metabolism were enriched in MIDY pigs. Butanoate supplement has been shown to increase lipolysis, as well as mitochondrial protein expression in WAT in mice (**?**). Butanoate supplement also increases oxidation of fat and activates BAT (**??**) and butyrate treatment increases the thermogenic capacity in mice (**?**). Similar increases in fatty acid oxidation and lipolysis has been seen in pigs following butanoate supplement (**?**).

Treatment with sodium butanoate alleviates adipocyte inflammation in mice
(**?**). This could potentially explain why there are gene sets related to immune responses enriched in WT AT, but this would have to be verified in more detailed

studies.

5.1.2.5 Angiotensinogen is downregulated in MIDY AT

AGT was the most significantly enriched protein in AT from WT pigs. In both humans and animal models, AGT is highly expressed in AT as well as inherently secreted from adipocytes (**?**). AGT is also known to stimulate lipogenesis
(**?**). Lemieux and colleagues have previously shown that *Agt*-KO mice had reduced AT inflammation and increased metabolic activity (**?**). Furthermore, they also found that *Agt*-KO mice had increased gene expression of genes involved in fatty acid metabolism and glucose uptake, as well as increased mitochondrial respiration in *Agt*-KO mice. Overexpression of AGT in AT induces adipose inflammation, decreased glucose tolerance, and insulin resistance (**?**). This is in line with the results from the GSEA analysis, where many gene sets involved in fatty acid metabolism, respiration, and mitochondria were enriched in MIDY pigs that had lower AGT expression. Furthermore, gene sets involved in glucose uptake were enriched in AT of

MIDY pigs at q-value < 0.1. Also, the GOBP regulation of inflammatory response were enriched in AT from WT pigs, which had a higher expression of AGT.

The renin-angiotensin aldosterone system is affected by atRA, as evidenced by that angiotensin can be downregulated by atRA in kidneys from rats with nephritis
(**?**) or glomerulosclerosis (**?**). AT expresses all components of the renin-angiotensin system that are necessary to produce the active angiotensin II
(**??**). Angiotensin II upregulates lipogenesis and downregulates lipolysis in AT via the angiotensin receptors (**?**). It has previously been shown that insulin increases AGT expression in subcutaneous adipocytes (**?**). In mice, AT derived AGT is important for AT development and the endocrine system
(**?**), as 30% of plasma AGT levels originates from AT. It has also been reported that active forms of AGT (such as angiotensin II and angotensin 1-7) stimulate browning of WAT and brown adipocyte differentiation (**??**), and AGT plays a role in general adipocyte differentiation
(**?**). One could speculate whether the downregulation of AGT seen in MIDY AT is caused by higher cleav-

age of AGT into active forms or due to a lower translation of the AGT protein. However, none of AGT's derivatives were detected in the proteomic analysis so if there are changes in expression of e.g. angiotensin II is not known. AGT was also detected to be significantly more abundant in SCAT than in MAT, potentially highlighting that it is more important for SCAT function, and angiotensin-converting enzyme 2 (ACE2) was significantly more abundant in MAT.

5.1.2.6 Glycolytic enzymes were upregulated in WT adipose tissue

Both FBP1 and HK1 were significantly enriched in WT pigs. Both enzymes are linked to diabetes since they are involved in glycolysis and insulin signaling. FBP1 is responsible for the reversible reaction fructose-1,6-bisphosphate to fructose-6-phospate in glycolysis/gluconeogenesis FBP1 has a RARE in its promoter region and can thus be regulated by atRA (**?**).

HK1 phosporylates glucose to yield glucose-6-phosphate, the first step of glycolysis. Hexokinase was detected to be downregulated in AT from diabetic mice (**?**). If there is less available glucose for the AT cells due to

impaired insulin signaling in diabetic pigs, a consequence of this would probably be that glycolytic enzymes would be downregulated in AT as insulin is important for regulating these genes (**?**).

5.1.2.7 Immune response proteins were more abundant in WT

The genotype GSEA analysis revealed that there were a few GO gene sets related to immunological responses enriched in WT AT, this effect was also seen in the liver study of the MIDY pigs (**?**). Backman, *et al.* argued that a decreased AKT activation by insulin in MIDY pigs could confine Toll-like receptor mediated PI3K-AKT-mTOR signaling and affect immune homeostasis.

The collagen COL3A1 was detected to be more abundant in both WT (for the genotype factor) and SCAT (for the tissue factor). This is in agreement with a study that detected decreased COL3A1 levels in SCAT from T2DM subjects linking it to a lower rate of AT fibrosis (**?**). COL3A1 has immune response regulatory properties (**?**) and has previously been found to be regulated by atRA in skin (**?**). Hyperinsuline-

mia is known to induce transcriptional changes of the COL3A1 gene in different tissues from rats (**?**), which might be comparable to the findings present here since COL3A1 was more abundant in WT AT with normal insulin production.

5.1.2.8 Other proteins upregulated in WT adipose tissue

Copper transport protein ATOX1 (ATOX1) was significantly enriched in WT AT, copper is known to have an important role in cell proliferation. ATOX1 is involved in the insulin signaling in vascular tissue (**?**). ATOX1 also acts as an antioxidant against hydrogen peroxide, and the protein has been suggested to be a thereapetical target in diabetes (**?**).

Both EEF1D and EIF3C have functions involved in translation. The EEF1D protein has been found to be downregulated by RA in cell cultures (**?**) and elongation factor activity are known to be stimulated by insulin (**?**). Many translation initiation factors are stimulated by insulin (**?**), and EIF3C binds to EIF4G when mTOR is stimulated by insulin and increases translation by increasing translation iniation (**?**). Since

insulin is known to stimulate proteins synthesis (**?**), these results could mean that the translation rate in MIDY pig AT is lower due to the lack of endogenous insulin. The GSEA analysis also revealed that several gene sets, such as the KEGG pathway ribosome, related to translation and protein synthesis were enriched in WT AT.

HSP90AA1 is a heat shock protein and important for protein folding during stress, and it is regulated by atRA and retinoid receptors *in vitro* and in sheep (**??**). Inhibition of HSP90 improved insulin signaling in diabetic mice (**?**) and amplified AKT insulin induced phosphorylation in cell cultures
(**?**).

The significance of SPAG9, the significant protein in WT with the highest l2fc, might be an artifact of the missing value imputation since it was imputed in 11 samples and detected at the lower end of the intensity distribution. The relative low q-value (= 0.042) might support this hypothesis.

5.1.3 Overlap between enriched gene sets between genotype and tissue

The two GSEA analyses also revealed that there was a considerable overlap between the gene sets enriched between MIDY and MAT. This could imply that the diabetic phenotype of the MIDY pigs manifests itself more in MAT than in SCAT. The transcriptomic analysis also supports this since the most significant differences between MIDY and WT were detected in MAT when compared to other within tissue comparisons. However, to prove if this is the case for the proteome as well one would need to perform further studies with higher statistical power.

In contrast, the overlap between enriched gene sets in WT and SCAT were much lower. Only four out of 25 enriched gene sets in WT were found to be significantly enriched in SCAT as well. This indicates that even though many gene sets enriched in MIDY and MAT were the same, the underlying comparison that was tested was different.

5.1.4 Only a few proteins had significant Genotype x Tissue interaction effects

There were only five proteins detected that had a significant Genotype x Tissue interaction effect. Of these five proteins, three were detected in every sample (TTC38, SCP2, VPS29). The relative low number of significant hits might indicate that the MIDY syndrome doesn't affect the majority of proteins from the two AT source types differently.

Of these five proteins, PCNA is probably the most interesting protein. PCNA was detected in three out of four MIDY MAT samples and not in any other sample group, and was significant in all three comparisons with other sample groups that only had missing values (**Figure 4.5**). PCNA has roles in housekeeping functions such as DNA replication and interaction with DNA polymerase δ (**?**). Nuclear PCNA is a marker for cell proliferation, whereas cytoplasmic PCNA is impacting energy metabolism (**??**). PCNA is important for correct DNA replication in eukaryotes (**?**). PCNA has been found to correlate with cell

proliferation and expression of PPARγ expression levels during BAT differentiation in human fetuses (**?**). PCNA correlates with BAT hyperplasia induced by cold exposure or metformin treatment in rodents but were not detected in WAT (**???**). However, it is worth noting that PCNA has also been detected in proliferating white adipocytes in obese rats (**?**). Since MIDY pigs don't exhibit obesity and with the other proteomic data that might suggest browning of AT, the expression of PCNA might suggest the presence of proliferative brown/beige adipocytes. To confirm the role of PCNA in MIDY AT it would be interesting to stain for PCNA and to study the localization and expression levels of PCNA in the two AT types. PCNA regulates retinoid receptors by direct protein-protein inteaction (**?**).

TTC38 was significantly upregulated in MIDY SCAT versus WT SCAT and WT MAT versus WT SCAT. However, TTC38 is a protein with unknown function.

VPS29 was significantly differentially abundant in three comparisons, of which the upregulation in the MIDY SCAT versus MIDY MAT comparison was the most significant. VPS29 is part of the retromer protein compelex which is responsible for transport from

endosomes to the Golgi apparatus (**?**). There might be possible connection between the retromer and diabetes (**?**), and the retromer is important for GLUT4 translocation to the plasma membrane in adipocytes and AT cell differentiation (**?**). However, no significant intra-tissue differences were observed between the genotypes.

SCP2 was detected at signicantly lower levels in MIDY SCAT than in the other three sample groups. SCP2 has previously been found to be differentially regulated in diabetic rats (**?**), and is important for regulating the composition of lipid droplets (**?**).

5.1.5 Conclusions and further studies

The most significant result from the comparison between AT between MIDY and WT pigs, was the upregulation of RDH16 in both AT types examined in this study. This together with an enrichment of proteins involved in lipid and fatty acid metabolism in MIDY AT suggests that one adaptation to the MIDY syndrome could be a shift to a more metabolically active AT and perhaps a "browning" of AT in MIDY pigs, especially in MAT. However, for this to be established

this would have to be examined in further studies. For example, the expression of BAT protein markers or mitochondrial content using histological analysis could be studied in MIDY AT.

The benefit of working with pigs is that it is a more relevant translational animal model when compared to for example rodent models, while at the same time cost and time limits the number of samples that can be obtained. Doubling the number of pigs in each sample group in the MIDY study would most likely lead to a discovery of more significant proteome differences between MIDY and WT pigs since the statistical power would be greater. However, I think it is more beneficial to focus further studies based on the findings presented here, as RDH16 is a largely unstudied protein in both pigs and humans.

Now that both the liver (**?**) and AT from the Munich MIDY pig biobank have been profiled by proteomics and transcriptomics, I propose that other relevant tissues from the biobank should be studied. Particularly, to study if RDH16 is more abundant in other tissues in MIDY pigs as well. This could answer wheter this is a global effect or limited to these two metabolically active tissue. In my opinion, the most interesting

tissues would be skeletal muscle, pancreas, and intestine. These are all relevant tissues that are affected by diabetes.

As can be seen in this study, a 2-way ANOVA can be used to elucidate differences in protein expression in AT even when the statistical power is low.

Another important question is whether the upregulation of RDH16 in the diabetic pigs is limited to this porcine diabetic model or if humans and other model organisms also exhibit this adaptation. A good place to start would be to study if Akita mice (the mouse model that MIDY pigs are based on) have a similar upregulation of RDH16 in liver and AT. In humans, proteomic and transcriptomic profiling of liver and AT biopsies could be used to test if the upregulation of RDH16 observed here occurs in T1DM diabetic patients as well.

In order to get a better understanding of RDH16's protein function, a RDH16 deficient pig model (similar to the mouse model studied by (**?**), *et al*) would be interesting to study. An INS^{C94Y} and RDH16 KO double mutants might also be interesting to study.

Furthermore, there seems to be a shift in metabolic activity in the AT in the diabetic pigs. Fatty acid and

lipid metabolic proteins and pathways were upregulated in AT from MIDY pigs, while glycolytic enzymes were detected to have higher expression levels in WT AT. Further research is required in order to elucidate whether retinol signaling is part of regulating this shift in protein expression.

The differences between MAT and SCAT in pigs suggests that the visceral AT type is more metabolically active while structural proteins are more highly expressed in SCAT. This is likely an effect of their location relative to other tissues, e.g. MAT is close to the liver whereas SCAT is located near skin and ECM.

Chapter 6

Summary

In this thesis, the effects of diabetes treatments on the proteome of selected tissues have been examined using mass spectrometry and relative protein quantification.

In the first study, the proteome differences between MIDY pigs, a large animal model for mutant INS gene-induced diabetes of youth, and WT littermates in two types of adipose tissue (AT) were investigated. In total 2779 proteins were detected in the AT samples and using a 2-way ANOVA, taking into account the factors group (diabetic vs. non-diabetic) and type of AT (subcutaneous vs. mesenterial), significantly differentially

expressed proteins were detected. Among the significantly differentially expressed proteins, retinol dehydrogenase 16 (RDH16) was detected to be the significantly most abundant protein in AT and enriched in MIDY AT. As all-trans retinoic acid is important for regulating transcript levels, this suggests that retinol is involved in the adaptation to diabetes in AT of the MIDY pigs. The presence of RDH16 in the AT of MIDY pigs might also indicate that a form of "browning" of the fat cells has occurred, as RDH16 is only expressed in BAT in rodent models. However, this would require further studies to elucidate.

Other proteins that were significantly more abundant in MIDY AT were proteins involved in fatty acid and lipid metabolic processes such as carboxylesterase 1 (CES1) and mitochondrial trifunctional enzyme subunit alpha and beta (HADHA and HADHB), whereas the glycolytic enzymes fructose-1,6 bisphosphatase 1 (FBP1) and hexokinase 1 (HK1) were significantly more abundant in WT AT. This could indicate that there is a shift in metabolic activity in AT as a response to the insulin deficiency. Whether retinol signaling is involved in regulating this metabolic shift remains unclear and further studies are required to understand

this better.

The results from the 2-way ANOVA also showed that mesenterial AT (MAT) is more metabolically active than subcutaneous AT (SCAT), as MAT expressed more proteins involved in various metabolic pathways. On the other hand, SCAT expressed more structural proteins such as proteins related to collagen and extracellular matrix (ECM). This can probably be explained by the location of the two tissue types, as MAT is situated close to the liver and SCAT is located near skin and ECM.

The second study used pigs carrying a dominant negative mutation in the glucose-dependent insulinotropic polypeptide receptor ($GIPR^{dn}$), a type 2 diabetes mellitus (T2DM) large animal model, to study what effect treatment with the glucagon-like peptide 1 (GLP1) receptor agonist liraglutide has on the proteome of the liver of $GIPR^{dn}$ pigs and compare these results to placebo treated $GIPR^{dn}$ pigs. In total 3404 protein groups were detected in the livers of $GIPR^{dn}$ pigs, of these 127 proteins were found to be significantly differentially expressed after multiple hypotheses correction (q-value < 0.05). Of these, mitochondrial hydroxymethylglutaryl-CoA synthase (HMGCS2) was detected at lower ex-

pression levels in liraglutide treated pigs and with the largest absolute log2 fold change. This suggests that increased insulin secretion stimulated by the liraglutide treatment inhibits ketogenesis in diabetic pigs, as HMGCS2 is the rate-limiting enzyme for ketogenesis.

The liraglutide treatment also induced changes in amino acid metabolism, as proteins involved in glycine, serine, and threonine metabolism and arginine and proline metabolism were enriched in placebo treated pigs. Expression of proteins involved in translation and protein folding, as translation factors and tRNA ligases, were significantly higher in liraglutide treated pigs.

These results show that the pig is a suitable large animal model for translational and proteomic research for research into diabetes and diabetes treatments, that can bridge the gap between findings in mice and human patients.

Chapter 7

Zusammenfassung

Proteomische Charakterisierung von porzinen Diabetesmodellen – Behandlungseffekte auf metabolische Gewebe

Im Rahmen dieser Dissertation wurden an zwei verschiedenen Schweinemodellen für die translationale Diabetesforschung an ausgewählten Geweben mit Relevanz für den Stoffwechsel Proteomuntersuchungen mittels Massenspektrometrie und relativer Proteinquantifizierung durchgeführt. In einer ersten Studie wurden die Proteomprofile von subkutanem und mesenterialem Fettgewebe zwischen MIDY-Schweinen, einem

Großtiermodell für mutant INS gene-induced diabetes of youth, und gesunden Geschwistertieren verglichen. Insgesamt wurden in dieser Studie 2779 Proteine detektiert und hinsichtlich ihrer Abundanzen mittels ANOVA auf die Beeinflussung durch Gruppe (diabetisch vs. nicht-diabetisch) und Fettgewebstyp (subkutan vs. mesenterial) untersucht. Das in beiden Fettgewebstypen bei den diabetischen Schweinen am meisten abundanzvermehrte Protein war die Retinol-Dehydrogenase 16 (RDH16), die für die Synthese von all-trans Retinsäure (atRA) eine entscheidende Rolle spielt. Nachdem atRA die Expression von zahlreichen Genen reguliert, könnte die vermehrte Abundanz von RDH16 eine Anpassung des Fettgewebes an die diabetische Stoffwechsellage reflektieren. Die Anwesenheit von RDH16 im Fettgewebe von diabetischen Schweinen könnte auch als Hinweis auf browning gewertet werden, da RDH16 bei Mäusen nur im braunen Fettgewebe exprimiert wird. Dies bedarf weiterer Untersuchungen. Neben RDH16 waren im Fettgewebe der diabetischen Schweine Enzyme aus dem Fettsäure- und Fettstoffwechsel abundanzvermehrt. Beispiele sind die Carboxylesterase 1 (CES1) sowie die alpha- und beta-Untereinheiten des mitochondrialen trifunktionalen Enzyms (HADHA und HADHB),

dagegen waren die Konzentrationen der Enzyme Fructose-1,6-Bisphosphatase 1 (FBP1) und Hexokinase 1 (HK1) im Fettgewebe der Kontrollschweine höher als bei den diabetischen Tieren. Dies weist auf metabolische Veränderungen im Fettgewebe als Folge eines chronischen Insulinmangels hin. Ob Veränderungen im Retinolstoffwechsel dafür eine Rolle spielen, ist gegenwärtig unklar und Gegenstand weiterer Untersuchungen. Neben den Unterschieden zwischen diabetischen und nicht-diabetischen Tieren zeigte die ANOVA deutliche Unterschiede zwischen den zwei untersuchten Fettgewebstypen. So lagen im mesenterialen Fettgewebe vor allem Proteine aus verschiedenen metabolischen Pathways in höherer Abundanz vor als im subkutanen Fettgewebe, was auf eine höhere metabolische Aktivität des mesenterialen Fettgewebes hinweist. Umgekehrt waren im subkutanen Fettgewebe vor allem Kollagene und andere extrazelluläre Matrixproteine in höherer Konzentration vorhanden als im mesenterialen Fettgewebe. Diese Befunde spiegeln die klaren strukturellen und funktionellen Unterschiede der beiden Fettgewebstypen auf molekularer Ebene wider. Die zweite Studie beschäftigt sich mit einem Schweinemodell, das einen dominant-negativen Rezeptor für das Glukose-abhängige insulinotrope

Polypeptid (GIP) exprimiert und Charakteristika des Diabetes Typ 2 zeigt. Hier war die Fragestellung, welche Proteomveränderungen in der Leber die Behandlung mit dem GLP1-Rezeptor Agonisten Liraglutide im Vergleich zu einer Placebo-Behandlung induziert. In dieser Studie konnten insgesamt 3404 Proteine identifiziert werden, von denen nach Korrektur auf multiples Testen 127 Proteine in ihrer Abundanz signifikant (q-value ¡ 0.05) unterschiedlich zwischen den beiden Behandlungsgruppen waren. Davon war die mitochondriale Hydroxymethylglutaryl-CoA Synthase (HMGCS2) durch die Liraglutide-Behandlung am stärksten abundanzvermindert. Die HMGCS2 ist das Schlüsselenzym der Ketogenese. Die verminderte Expression ist durch die vermehrte Insulinfreisetzung nach Liraglutide-Behandlung und die hemmende Wirkung von Insulin auf die Expression der HMGCS2 erklärbar. Darüber hinaus induzierte die Liraglutide-Behandlung Abundanzveränderungen von Proteinen aus dem Funktionsbereich Aminosäure-Stoffwechsel. So waren Proteine aus dem Glycin-, Serin- und Threonin-Stoffwechsel sowie Proteine aus dem Arginin- und Prolin-Metabolismus in der Leber der Liraglutide-behandelten Tiere in ihrer Konzentration vermehrt. Gleiches gilt für Proteine, die bei der Translation bzw.

Proteinfaltung eine Rolle spielen, sowie für bestimmte tRNA-Ligasen Diese Ergebnisse zeigen, dass Proteomstudien an klinisch relevanten Großtiermodellen für die Diabetesforschung molekulare Einblicke in Krankheitsmechanismen und Therapiewirkungen geben können, die helfen, die translationale Lücke zwischen Untersuchungen an Nagermodellen und klinischen Studien zu schließen.

Appendix A

Supplementary material for MIDY adipose tissue analysis

Table A.1. All significant proteins more abun
in mesenteric adipose tissue in the 2-way ANOVA
value < 0.05).

Gene name	Protein name	p-v
MAOA	amine oxidase	8.58

Gene name	Protein name	p-v
HSP90AA1	heat shock protein HSP 90-alpha	6.5
ALO5	arachidonate 15-lipoxygenase	7.5
ALDH1A2	aldehyde dehydrogenase 1 family, member A2	1.5
DLST	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial precursor	1.6
HSP90AB1	heat shock protein HSP 90-beta	5.0
SUCLG2	succinate–CoA ligase	4.8
CLIC5	chloride intracellular channel protein 5	1.1
SOD2	superoxide dismutase	1.2
CCT2	T-complex protein 1 subunit beta	3.1
PRD	peroxiredoxin-5, mitochondrial	3.4
KRT14	keratin, type I cytoskeletal 19	5.8
ALDH9A1	4-trimethylaminobutyraldehyde dehydrogenase	8.2
IVD	isovaleryl-CoA dehydrogenase, mitochondrial	8.6
PSME1	proteasome activator complex subunit 1	8.8
GSTO1	glutathione S-transferase omega-1	9.1
KRT8	keratin, type II cytoskeletal 8	1.8
FASN	fatty acid synthase	3.3

Gene name	Protein name	p-v
SLC25A4	ADP/ATP translocase 1	3.73
PRD	peroxiredoxin-2	3.81
PGAM1	phosphoglycerate mutase 1	3.80
ACADVL	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	3.96
PRD	thioredoxin-dependent peroxide reductase, mitochondrial	4.19
HSD17B10	3-hydroxyacyl-CoA dehydrogenase type-2	4.24
HSPE1	10 kDa heat shock protein, mitochondrial	5.14
HSD11B1	corticosteroid 11-beta-dehydrogenase isozyme 1 2	6.28
ECH1	delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	6.27
ME1	NADP-dependent malic enzyme	8.01
PDHB	pyruvate dehydrogenase E1 component subunit beta, mitochondrial	9.37
HSPB1	heat shock protein beta-1	9.80
DLAT	dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial precursor	0.0
LMAN2	vesicular integral-membrane protein VIP36	0.0
NP_001172070	ornithine aminotransferase, mitochondrial	0.0

Gene name	Protein name	p-v
ACSS2	acetyl-coenzyme A synthetase, cytoplasmic	0.0
ITIH1	inter-alpha-trypsin inhibitor heavy chain H1 precursor	0.0
TXN	thioredoxin	0.0
MDH2	malate dehydrogenase, mitochondrial	0.0
CYP2B22	cytochrome P450 2B4-like	0.0
ST13	hsc70-interacting protein	0.0
NDUFS1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	0.0
AIFM1	apoptosis-inducing factor 1, mitochondrial	0.0
ETFB	electron transfer flavoprotein subunit beta	0.0
UGP2	UTP–glucose-1-phosphate uridylyltransferase	0.0
CCT8	T-complex protein 1 subunit theta	0.0
SLC25A22	mitochondrial glutamate carrier 1	0.0
A2M	alpha-2-macroglobulin	0.0
ACO2	aconitate hydratase, mitochondrial precursor	0.0
CACYBP	calcyclin-binding protein	0.0
ACADS	short-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	0.0
S100A12	protein S100-A12	0.0
EZR	ezrin	0.0

Gene name	Protein name	p-v
NPEPPS	puromycin-sensitive aminopeptidase	0.0
COX7A1	cytochrome c oxidase subunit 7A1, mitochondrial precursor	0.0
ATP5F1D	ATP synthase subunit delta, mitochondrial	0.0
HSPA9	stress-70 protein, mitochondrial	0.0
ACO	peroxisomal acyl-coenzyme A oxidase 1	0.0
PCBD2	pterin-4-alpha-carbinolamine dehydratase 2	0.0
SLC25A5	ADP/ATP translocase 2	0.0
HSPD1	60 kDa heat shock protein, mitochondrial	0.0
STIP1	stress-induced-phosphoprotein 1	0.0
PTGES3	prostaglandin E synthase 3	0.0
ITIH2	inter-alpha-trypsin inhibitor heavy chain H2 precursor	0.0
ATP5ME	ATP synthase subunit e, mitochondrial	0.0
COA	cytochrome c oxidase subunit 5A, mitochondrial	0.0
PSME2	proteasome activator complex subunit 2	0.0
SQOR	sulfide:quinone oxidoreductase, mitochondrial	0.0
PCNA	proliferating cell nuclear antigen	0.0
SLC2A4	solute carrier family 2, facilitated glucose transporter member 4	0.0

Gene name	Protein name	p-v
GOT2	aspartate aminotransferase, mitochondrial precursor	0.0
ACAA1	3-ketoacyl-CoA thiolase, peroxisomal	0.0
AK2	adenylate kinase 2, mitochondrial	0.0
GOT1	aspartate aminotransferase, cytoplasmic	0.0
NDUFS3	NADH dehydrogenase	0.0
PSMB3	proteasome subunit beta type-3	0.0
LNPEP	leucyl-cystinyl aminopeptidase	0.0
LYPLA1	acyl-protein thioesterase 1	0.0
ACSF2	acyl-CoA synthetase family member 2, mitochondrial	0.0
PTGR1	prostaglandin reductase 1	0.0
HES1	ES1 protein homolog, mitochondrial	0.0
LTF	lactotransferrin precursor	0.0
GNG7	guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-7	0.0
VPS29	vacuolar protein sorting-associated protein 29	0.0
CYP27A1	sterol 26-hydroxylase, mitochondrial	0.0
PZP	pregnancy zone protein	0.0
PSMC3	26S protease regulatory subunit 6A	0.0
PSMD5	26S proteasome non-ATPase regulatory subunit 5	0.0

Gene name	Protein name	p-v
TCP1	T-complex protein 1 subunit alpha	0.0
NDUFB10	NADH dehydrogenase	0.0
G6PD	glucose-6-phosphate 1-dehydrogenase	0.0
SDHB	succinate dehydrogenase	0.0
ACLY	ATP-citrate synthase	0.0
PSMA2	proteasome subunit alpha type-2	0.0
ACSL1	long-chain-fatty-acid–CoA ligase 1	0.0
SYPL1	synaptophysin-like protein 1	0.0
SLC25A3	phosphate carrier protein, mitochondrial	0.0
GYG2	glycogenin-2	0.0
ATP5B	ATP synthase subunit beta, mitochondrial	0.0
PXN	paxillin X7	0.0
ACE2	angiotensin-converting enzyme 2	0.0
UQCRC1	cytochrome b-c1 complex subunit 1, mitochondrial	0.0
NNT	NAD(P) transhydrogenase, mitochondrial	0.0
NDUFA8	NADH dehydrogenase	0.0
TGM2	protein-glutamine gamma-glutamyltransferase 2	0.0
HADHB	trifunctional enzyme subunit beta, mitochondrial	0.0
CS	citrate synthase, mitochondrial	0.0
BLVRB	flavin reductase (NADPH)	0.0

Gene name	Protein name	p-v
CCT7	T-complex protein 1 subunit eta	0.0
IMMT	MICOS complex subunit MIC60	0.0
MCCC1	methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	0.0
LRPPRC	leucine-rich PPR motif-containing protein, mitochondrial	0.0
ACOT4	acyl-coenzyme A thioesterase 1-like	0.0
HSPA4	heat shock 70 kDa protein 4	0.0
PDCD6IP	programmed cell death 6-interacting protein	0.0
TUFM	elongation factor Tu, mitochondrial	0.0
CHCHD3	MICOS complex subunit MIC19	0.0
PDHX	pyruvate dehydrogenase protein X component, mitochondrial	0.0
PHB2	prohibitin-2	0.0
SLC25A10	mitochondrial dicarboxylate carrier	0.0
ACSS3	acyl-CoA synthetase short-chain family member 3, mitochondrial	0.0
HPRT1	hypoxanthine-guanine phosphoribosyltransferase	0.0
SLC25A6	ADP/ATP translocase 3	0.0
SLC25A1	tricarboxylate transport protein, mitochondrial	0.0

Gene name	Protein name	p-v
SLC25A13	calcium-binding mitochondrial carrier protein Aralar2	0.0
RPE	ribulose-phosphate 3-epimerase	0.0
PSMD2	26S proteasome non-ATPase regulatory subunit 2	0.0
ATP5H	ATP synthase subunit d, mitochondrial	0.0
FKBP4	peptidyl-prolyl cis-trans isomerase FKBP4	0.0
NPG1	protegrin-1-like	0.0
ECHDC1	ethylmalonyl-CoA decarboxylase	0.0
ATP5F1A	ATP synthase subunit alpha, mitochondrial	0.0
HADHA	trifunctional enzyme subunit alpha, mitochondrial	0.0
HBB	hemoglobin subunit beta	0.0
UQCRFS1	cytochrome b-c1 complex subunit Rieske, mitochondrial	0.0
CHP1	calcineurin B homologous protein 1	0.0
ACACA	acetyl-CoA carboxylase 1	0.0
BCKDHA	2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial	0.0
NAMPT	nicotinamide phosphoribosyltransferase	0.0
SDHA	succinate dehydrogenase	0.0
GPM6A	neuronal membrane glycoprotein M6-a	0.0
C3	complement C3	0.0

Gene name	Protein name	p-v
NDUFB4	NADH dehydrogenase	0.0

Table A.2. All significant proteins more abun
in subcutaneous adipose tissue in the 2-way AN
(q-value < 0.05).

Gene name	Protein name	p-v
COL3A1	collagen alpha-1(III) chain precursor	1.15
CBR2	carbonyl reductase	7.56
OGN	mimecan	6.64
CFD	complement factor D	1.14
FERMT2	fermitin family homolog 2	1.82
MMP2	72 kDa type IV collagenase precursor	3.06
PHGDH	D-3-phosphoglycerate dehydrogenase	3.29
SDR16C5	epidermal retinol dehydrogenase 2	3.33
PCOLCE	procollagen C-endopeptidase enhancer 1	9.33
DCN	decorin precursor	1.10
LUM	lumican precursor	2.84
SGCA	alpha-sarcoglycan precursor	3.41
TUBB	tubulin beta chain	3.98
SNX3	sorting nexin-3	4.05

Gene name	Protein name	p-v
SORBS1	sorbin and SH3 domain-containing protein 13	4.89
PLCD1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	6.54
IKBIP	inhibitor of nuclear factor kappa-B kinase-interacting protein	7.83
XP_020930866	tensin-1	1.06
SERPINF1	pigment epithelium-derived factor	1.09
PRELP	prolargin	1.10
XP_020917965	olfactomedin-like protein 1	1.12
F13A1	coagulation factor XIII A chain	1.24
DPYSL2	dihydropyrimidinase-related protein 2	1.40
MFAP2	microfibrillar-associated protein 2	1.41
TMSB4X	thymosin beta-4	1.56
HSPA2	heat shock-related 70 kDa protein 2	1.73
AGT	angiotensinogen	1.89
UGDH	UDP-glucose 6-dehydrogenase	2.08
LEMD2	LEM domain-containing protein 2	2.43
MYOC	myocilin precursor	2.49
XP_020923800	C-type mannose receptor 2	2.77
TLN1	talin-1	2.90
YKT6	synaptobrevin homolog YKT6	3.18
ITGB1	integrin beta-1 precursor	3.33

Gene name	Protein name	p-v
RENBP	N-acylglucosamine 2-epimerase	3.48
DMD	dystrophin3	3.52
ASPN	asporin	5.20
CYB5R3	NADH-cytochrome b5 reductase 3-like	6.38
GNB4	guanine nucleotide-binding protein subunit beta-4	6.77
ECM1	extracellular matrix protein 1 isoform X3	6.79
PPIB	peptidyl-prolyl cis-trans isomerase B	7.01
ARHGDIA	rho GDP-dissociation inhibitor 1	9.48
XP_005655478	olfactomedin-like protein 3	9.43
CD276	CD276 antigen	9.72
ADD1	alpha-adducin	0.0
SCP2.1	podocan	0.0
TWF2	twinfilin-2	0.0
SPTBN1	spectrin beta chain, non-erythrocytic 1	0.0
AKAP12	A-kinase anchor protein 12	0.0
COL1A2	collagen alpha-2(I) chain precursor	0.0
XP_020938403	glutathione S-transferase P-like isoform X3	0.0
EPB41L2	band 4.1-like protein 2	0.0
ILK	integrin-linked protein kinase	0.0
SPTAN1	spectrin alpha chain, non-erythrocytic 1	0.0
PI16	peptidase inhibitor 16 precursor	0.0
LGALS1	galectin-1	0.0

Gene name	Protein name	p-v
KANK2	KN motif and ankyrin repeat domain-containing protein 2	0.0
SEPTIN9	septin-9	0.0
SNX2	sorting nexin-2	0.0
PEPD	xaa-Pro dipeptidase	0.0
SNCG	gamma-synuclein	0.0
KDELC2	KDEL motif-containing protein 2 isoform X9	0.0
DDAH2	N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	0.0
PARVA	alpha-parvin	0.0
IGFBP6	insulin-like growth factor-binding protein 6 precursor	0.0
VPS26C	Down syndrome critical region protein 3	0.0
HMGB1	high mobility group protein B1	0.0
IFI30	gamma-interferon-inducible-lysosomal thiol reductase precursor	0.0
ITGA7	integrin alpha-7	0.0
RSU1	ras suppressor protein 1 isoform X3	0.0
XP_020943699	eukaryotic translation initiation factor 3 subunit C	0.0

Gene name	Protein name	p-v
XP_020941305	LIM and senescent cell antigen-like-containing domain protein 1 isoform X8	0.0
RUFY1	RUN and FYVE domain-containing protein 1	0.0
LIMCH1	LIM and calponin homology domains-containing protein 1	0.0
TLN2	talin-2	0.0
DNM2	dynamin-29	0.0
DNAJC13	dnaJ homolog subfamily C member 13	0.0
AP2A2	AP-2 complex subunit alpha-2	0.0
GBE1	1,4-alpha-glucan-branching enzyme	0.0
TUBB4B	tubulin beta-4B chain	0.0
CRYM	ketimine reductase mu-crystallin	0.0
FKBP9	peptidyl-prolyl cis-trans isomerase FKBP9 precursor	0.0
ASPA	aspartoacylase	0.0
ITIH5	inter-alpha-trypsin inhibitor heavy chain H5	0.0
ITGA5	integrin alpha-5	0.0
NXN	nucleoredoxin	0.0
CORO1B	coronin-1B	0.0
COL1A1	collagen alpha-1(I) chain isoform X3	0.0
CIRBP	cold-inducible RNA-binding protein	0.0

Gene name	Protein name	p-v
INPP1	inositol polyphosphate 1-phosphatase	0.0
SULT1A3	sulfotransferase 1A3	0.0
ABI3BP	target of Nesh-SH39	0.0
EMD	emerin	0.0
LAMA5	laminin subunit alpha-5	0.0
HNMT	histamine N-methyltransferase	0.0
FBLN2	fibulin-2	0.0
SH3GLB2	endophilin-B21	0.0
SERPINH1	serpin H1 precursor	0.0
LAMB2	laminin subunit beta-2	0.0
CACNA2D1	voltage-dependent calcium channel subunit alpha-2/delta-19	0.0
FAP	prolyl endopeptidase FAP	0.0
RAP1A	ras-related protein Rap-1A	0.0
ITGAV	integrin alpha-V isoform X3	0.0
MAPK1	mitogen-activated protein kinase 1	0.0
FN1	fibronectin	0.0
KDSR	3-ketodihydrosphingosine reductase	0.0
COL6A3	collagen alpha-3(VI) chain	0.0
EHD4	EH domain-containing protein 4	0.0
FNDC1	fibronectin type III domain-containing protein 1	0.0
VCAN	versican core protein precursor	0.0

Gene name	Protein name	p-v
LRP1	prolow-density lipoprotein receptor-related protein 1	0.0
LGALS3	galectin-3	0.0
XP_003132938	phosphatidylethanolamine-binding protein 1	0.0
PDIA5	protein disulfide-isomerase A5	0.0
DYNC1H1	cytoplasmic dynein 1 heavy chain 1	0.0
PFKP	ATP-dependent 6-phosphofructokinase, platelet type	0.0
DBN1	drebrin isoform X3	0.0
ERLIN2	erlin-2	0.0
AOX1	aldehyde oxidase	0.0
XP_020938158	collagen alpha-2(VI) chain	0.0
RTN4.1	reticulon-4 isoform X3	0.0
MYH14	myosin-14	0.0
XP_020926755	collagen alpha-1(VI) chain	0.0
VWF	von Willebrand factor	0.0
CD163	scavenger receptor cysteine-rich type 1 protein M130	0.0
CREG1	protein CREG1 isoform X3	0.0
SH3GLB1	endophilin-B1	0.0
XP_005667226	mth938 domain-containing protein	0.0
SGCD	delta-sarcoglycan	0.0
HPX	hemopexin precursor	0.0

Gene name	Protein name	p-v
MYO1B	unconventional myosin-Ib	0.0
THBS4	thrombospondin-4	0.0
RHOG	rho-related GTP-binding protein RhoG	0.0
FSTL1	follistatin-related protein 1	0.0
XP_020922632	membrane primary amine oxidase	0.0
KPNB1	importin subunit beta-1	0.0
CYP2A19	cytochrome P450 2A19	0.0
COPA	coatomer subunit alpha	0.0
NAXD	ATP-dependent (S)-NAD(P)H-hydrate dehydratase isoform 2	0.0
APOE	apolipoprotein E precursor	0.0
MYOF	myoferlin	0.0
CTSB	cathepsin B	0.0
UAP1	UDP-N-acetylhexosamine pyrophosphorylase isoform X3	0.0
COPG1	coatomer subunit gamma-1	0.0
RRAS	ras-related protein R-Ras	0.0
NP_001230809	PRA1 family protein 2	0.0
XP_020921390	BTB/POZ domain-containing protein KCTD12	0.0
ENOSF1	mitochondrial enolase superfamily member 1	0.0
CTTN	src substrate cortactin	0.0

Gene name	Protein name	p-v
BGN	biglycan	0.0
ARCN1	coatomer subunit delta	0.0
PGM3	phosphoacetylglucosamine mutase	0.0
SNX6	sorting nexin-6	0.0
NPTN	neuroplastin	0.0
PDIA3	protein disulfide-isomerase A3 precursor	0.0
HSPB6	heat shock protein beta-6	0.0
FAM129A	protein Niban	0.0
MYDGF	myeloid-derived growth factor	0.0
UAP1L1	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1	0.0
DHDH	trans-1,2-dihydrobenzene-1,2-diol dehydrogenase	0.0
DPT	dermatopontin	0.0

Table A.3. GSEA: Significantly enriched gene sets i
adipose tissue from MIDY pigs (q-value < 0.05).

Category	Name	Size[1]	NES[2]	p-
GO	Fatty acid metabolic process	34	2.07	0
GO	Cellular respiration	33	2.12	0

Category	Name	Size[3]	NES[4]	p-
GO	Energy derivation by oxidation of organic compounds	39	2.07	0
GO	Aerobic respiration	15	2.01	0
GO	Oxidation reduction process	83	2.02	0
KEGG	Fatty acid metabolism	16	2.01	0
GO	Oxidoreductase activity	66	1.98	0
GO	Electron transport chain	26	1.99	0
GO	Mitochondrial part	86	1.99	0
GO	Generation of precursor metabolites and energy	43	1.97	0
GO	Mitochondrial envelope	58	1.95	0
GO	Lipid metabolic process	55	1.95	0
GO	Carbon oxygen lyase activity	10	1.94	0
GO	Hydro lyase activity	10	1.94	0
GO	Mitochondrion	112	1.93	0
GO	Regulation of cellular ketone metabolic process	26	1.90	0
GO	Mitochondrial matrix	36	1.90	0
GO	Cellular lipid metabolic process	47	1.90	0
GO	Lyase activity	18	1.90	0
GO	Hydrogen peroxide metabolic process	4	1.89	0
GO	Organelle inner membrane	47	1.88	0

Category	Name	Size[3]	NES[4]	p-
GO	Oxidative phosphorylation	20	1.89	0
GO	Lipid oxidation	18	1.87	0
KEGG	Citrate cycle tca cycle	14	1.86	0
GO	Tricarboxylic acid metabolic process	13	1.86	0
GO	Monocarboxylic acid metabolic process	47	1.84	0
GO	Electron carrier activity	25	1.84	0
GO	Oxidoreductase activity acting on the ch oh group of donors nad or nadp as acceptor	20	1.84	0
GO	Primary alcohol metabolic process	8	1.84	0
GO	Inner mitochondrial membrane protein complex	20	1.83	0
GO	Fatty acid catabolic process	19	1.83	0
GO	Reactive oxygen species metabolic process	10	1.83	0
GO	Regulation of fatty acid metabolic process	10	1.82	0
GO	Respiratory chain	20	1.82	0
GO	Fatty acid beta oxidation	17	1.81	0

Category	Name	Size[3]	NES[4]	p-
KEGG	Valine leucine and isoleucine degradation	14	1.81	0
GO	Oxidoreductase complex	24	1.80	0
GO	Organic acid catabolic process	29	1.80	0
GO	Positive regulation of wnt signaling pathway	27	1.81	0
GO	Response to hydrogen peroxide	3	1.81	0
GO	Cellular lipid catabolic process	23	1.81	0
GO	Lipid modification	19	1.81	0
GO	Oxidoreductase activity acting on ch oh group of donors	20	1.79	0
GO	Anion transmembrane transport	9	1.79	0
GO	Anion transport	18	1.78	0
GO	Positive regulation of canonical wnt signaling pathway	26	1.78	0
GO	Anion transmembrane transporter activity	8	1.77	0
KEGG	Pyruvate metabolism	10	1.77	0
KEGG	Butanoate metabolism	9	1.75	0
GO	Mitochondrial protein complex	22	1.75	0
GO	Cellular response to reactive oxygen species	6	1.74	0
GO	Microbody	17	1.74	0

Category	Name	Size[3]	NES[4]	p-
GO	Hydrogen peroxide catabolic process	4	1.74	(
GO	Monocarboxylic acid catabolic process	19	1.74	(

Table A.4. GSEA: Significantly enriched gene sets i
adipose tissue from WT pigs (q-value < 0.05).

Category	Name	Size[5]	NES[6]	p-
KEGG	Complement and coagulation cascades	24	-2.38	(
GO	Glycosaminoglycan binding	22	-2.27	(
GO	Vesicle lumen	23	-2.28	(
GO	Blood microparticle	33	-2.17	(
GO	Serine type endopeptidase inhibitor activity	13	-2.15	(
GO	Regulation of endothelial cell migration	7	-2.07	(
GO	Leukocyte mediated immunity	15	-2.07	(

[3]Number of proteins in leading edge.
[4]Normalized enrichment score.

Table A.4. GSEA: Significantly enriched gene sets i
adipose tissue from WT pigs (q-value < 0.05).

Category	Name	Size[5]	NES[6]	p-
GO	Secretory granule lumen	17	-2.07	0
GO	Heparin binding	17	-2.08	0
GO	Humoral immune response mediated by circulating immunoglobulin	12	-2.00	0
GO	Regulation of cell growth	11	-2.01	0
GO	B cell mediated immunity	12	-1.97	0
GO	Fibrinolysis	10	-1.92	0
GO	Regulation of protein maturation	15	-1.92	0
GO	Peptidase inhibitor activity	20	-1.94	0
GO	Negative regulation of response to wounding	18	-1.91	0
KEGG	Ribosome	21	-1.93	0
GO	Protein activation cascade	20	-1.93	0
GO	Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	12	-1.91	0
GO	Humoral immune response	14	-1.96	0

Table A.4. GSEA: Significantly enriched gene sets i
adipose tissue from WT pigs (q-value < 0.05).

Category	Name	Size[5]	NES[6]	p-
GO	Immune effector process	25	-1.95	0
GO	Lymphocyte mediated immunity	12	-1.94	0
GO	Extracellular space	74	-1.94	0
GO	Platelet alpha granule	17	-1.94	0
GO	Platelet alpha granule lumen	15	-1.96	0

Table A.5. GSEA: Significantly enriched gene sets i
MAT in pigs (q-value < 0.05).

Category	Name	Size[7]	NES[8]	p-
GO	Mitochondrial part	88	2.62	
GO	Mitochondrial matrix	54	2.62	
GO	Cellular respiration	37	2.57	
GO	Generation of precursor metabolites and energy	49	2.53	
GO	Energy derivation by oxidation of organic compounds	42	2.50	
GO	Coenzyme metabolic process	37	2.44	
GO	Cofactor metabolic process	41	2.42	

Category	Name	Size	NES	p-
GO	Coenzyme biosynthetic process	14	2.38	
GO	Cofactor biosynthetic process	18	2.37	
GO	Mitochondrial envelope	57	2.36	
GO	Mitochondrion	118	2.32	
GO	Electron transport chain	23	2.31	
GO	Thioester metabolic process	19	2.28	
GO	Oxidation reduction process	98	2.28	
GO	Organelle inner membrane	52	2.28	
GO	Fatty acid beta oxidation	19	2.27	
GO	Fatty acid metabolic process	36	2.27	
GO	Oxidoreductase complex	23	2.26	
KEGG	Parkinsons disease	21	2.24	
GO	Mitochondrial membrane part	17	2.24	
KEGG	Citrate cycle tca cycle	18	2.23	
GO	Mitochondrial protein complex	25	2.22	
GO	Aerobic respiration	19	2.21	
GO	Mitochondrion organization	29	2.21	
GO	Monocarboxylic acid metabolic process	53	2.18	
GO	Lipid oxidation	20	2.17	
KEGG	Oxidative phosphorylation	17	2.15	
GO	Fatty acid catabolic process	19	2.15	
GO	Thioester biosynthetic process	11	2.14	

Category	Name	Size	NES	p-
GO	Inner mitochondrial membrane protein complex	21	2.13	
GO	Tricarboxylic acid metabolic process	18	2.13	
GO	Oxidoreduction coenzyme metabolic process	19	2.12	
GO	Oxidoreductase activity	80	2.11	
GO	Oxidative phosphorylation	17	2.11	
GO	Respiratory chain	17	2.09	
GO	Nucleobase containing small molecule metabolic process	41	2.09	
GO	Tumor necrosis factor mediated signaling pathway	25	2.09	
GO	Monocarboxylic acid catabolic process	25	2.07	
GO	Cellular lipid catabolic process	28	2.07	
GO	Proteasome complex	25	2.07	
GO	Positive regulation of ligase activity	22	2.02	
GO	Oxidoreductase activity acting on the ch ch group of donors	13	2.02	
GO	Lipid modification	21	2.02	
GO	Envelope	66	2.02	

Category	Name	Size	NES	p-
KEGG	Fatty acid metabolism	16	2.02	0
GO	Regulation of ligase activity	22	2.01	
KEGG	Huntingtons disease	23	2.00	
GO	Protein polyubiquitination	25	1.99	
GO	Nucleoside monophosphate metabolic process	23	1.99	
GO	Positive regulation of protein modification by small protein conjugation or removal	23	1.98	
GO	Nik nf kappab signaling	23	1.98	
GO	Nucleoside triphosphate metabolic process	24	1.98	
GO	Fatty acyl coa binding	9	1.97	
GO	Monosaccharide biosynthetic process	14	1.96	
GO	Fatty acyl coa metabolic process	9	1.96	
GO	Anaphase promoting complex dependent catabolic process	21	1.95	
GO	Fatty acid biosynthetic process	8	1.95	
KEGG	Proteasome	19	1.95	
GO	Organophosphate metabolic process	47	1.94	

Category	Name	Size	NES	p-
GO	Mitochondrial respiratory chain complex assembly	13	1.94	
GO	Cofactor binding	40	1.94	
GO	Lyase activity	18	1.94	
KEGG	Peroxisome	12	1.94	
GO	Protein tetramerization	9	1.92	
GO	Oxidoreductase activity acting on nad p h quinone or similar compound as acceptor	13	1.92	
GO	Mitochondrial respiratory chain complex i biogenesis	13	1.92	0
GO	Regulation of protein ubiquitination involved in ubiquitin dependent protein catabolic process	23	1.92	
GO	Innate immune response activating cell surface receptor signaling pathway	22	1.91	
GO	Dicarboxylic acid metabolic process	19	1.91	
GO	Carbon oxygen lyase activity	12	1.90	0
GO	Ligase activity forming carbon sulfur bonds	8	1.90	

Category	Name	Size	NES	p-
KEGG	Propanoate metabolism	15	1.90	0
GO	Regulation of cellular ketone metabolic process	23	1.89	
GO	Fatty acid beta oxidation using acyl coa dehydrogenase	9	1.89	
GO	Protein ubiquitination	30	1.89	
GO	Hydro lyase activity	12	1.89	
GO	Myelin sheath	36	1.88	
GO	Flavin adenine dinucleotide binding	14	1.88	0
GO	Carbohydrate biosynthetic process	17	1.88	
GO	Nadh dehydrogenase complex	12	1.87	
GO	Heat shock protein binding	11	1.87	
GO	Acyl coa dehydrogenase activity	7	1.87	0
GO	Branched chain amino acid metabolic process	9	1.87	
GO	Organic acid metabolic process	78	1.86	
GO	Cellular response to reactive oxygen species	6	1.85	
GO	Cellular protein complex assembly	21	1.85	
GO	Anion transmembrane transport	7	1.85	

Category	Name	Size	NES	p-
KEGG	Valine leucine and isoleucine degradation	18	1.84	0
GO	Proteasome accessory complex	10	1.84	0
GO	Lipid catabolic process	30	1.84	
GO	Nucleoid	7	1.82	0
GO	Response to tumor necrosis factor	28	1.81	0
GO	Glycosyl compound metabolic process	25	1.81	
GO	Proteasomal protein catabolic process	23	1.81	
GO	Glyoxylate metabolic process	9	1.80	
GO	Glucose metabolic process	15	1.80	
GO	Protein homotetramerization	6	1.80	0
GO	Mitochondrial transport	16	1.80	
KEGG	Pyruvate metabolism	12	1.80	0
GO	Regulation of cellular amino acid metabolic process	21	1.80	
GO	Regulation of cellular amine metabolic process	21	1.79	0
GO	Organic acid catabolic process	33	1.79	
GO	Small molecule metabolic process	106	1.79	
GO	Coenzyme binding	29	1.79	0

Category	Name	Size	NES	p-
GO	Apoptotic mitochondrial changes	5	1.79	0
GO	Regulation of protein modification by small protein conjugation or removal	24	1.78	0
GO	Regulation of cofactor metabolic process	7	1.78	
GO	2 oxoglutarate metabolic process	6	1.78	0
GO	Microbody	17	1.78	0
GO	Nadp metabolic process	6	1.78	
GO	Positive regulation of protein localization to nucleus	6	1.78	
GO	Nad metabolic process	10	1.77	0
GO	Cellular lipid metabolic process	46	1.77	
GO	Cellular response to oxidative stress	7	1.76	
GO	Lipid metabolic process	57	1.76	
GO	Electron carrier activity	25	1.76	
GO	Response to toxic substance	14	1.75	0
GO	Hydrogen peroxide metabolic process	5	1.75	0
GO	Lipid biosynthetic process	18	1.75	
GO	Mitochondrial membrane organization	7	1.74	

Category	Name	Size	NES	p-
GO	Hydrogen transport	7	1.74	0
GO	Hydrogen ion transmembrane transport	6	1.73	0
GO	Hydrogen peroxide catabolic process	5	1.73	0
GO	T cell receptor signaling pathway	23	1.72	0
GO	Monocarboxylic acid biosynthetic process	13	1.72	0
GO	Protein localization to mitochondrion	10	1.72	0
GO	Oxidoreductase activity acting on peroxide as acceptor	8	1.71	0
GO	Fertilization	7	1.71	0
GO	Regulation of RNA stability	26	1.70	
GO	Lipid homeostasis	12	1.70	0
GO	Glutathione metabolic process	9	1.70	0
GO	Unfolded protein binding	14	1.70	0
GO	Transmembrane transport	18	1.70	
KEGG	Alzheimers disease	19	1.69	0
GO	Protein targeting to mitochondrion	9	1.69	0
GO	Purine containing compound metabolic process	25	1.68	

Category	Name	Size	NES	p-
GO	Microbody part	10	1.68	0
GO	Antigen receptor mediated signaling pathway	23	1.68	0
GO	Oxidoreductase activity acting on the aldehyde or oxo group of donors nad or nadp as acceptor	12	1.68	0
GO	Unsaturated fatty acid metabolic process	8	1.67	0
GO	Oxidoreductase activity acting on nad p h	16	1.67	0
GO	Small molecule biosynthetic process	31	1.67	
GO	Steroid hormone receptor binding	5	1.67	0
GO	Nucleoside phosphate biosynthetic process	5	1.66	0
GO	Protein stabilization	13	1.66	0
GO	Nadh metabolic process	9	1.66	0
GO	Protein modification by small protein conjugation or removal	36	1.65	0
GO	Ligase activity	18	1.65	0
GO	Carbon carbon lyase activity	6	1.65	0

Category	Name	Size	NES	p-
GO	Anion transmembrane transporter activity	7	1.65	0
GO	Reactive oxygen species metabolic process	11	1.64	0
GO	Hexose metabolic process	15	1.64	
GO	Antigen processing and presentation of exogenous peptide antigen via mhc class i	17	1.64	0
GO	Positive regulation of canonical wnt signaling pathway	22	1.63	0

Table A.6. GSEA: Significantly enriched gene sets
SCAT in pigs (q-value < 0.05).

Category	Name	Size[9]	NES[10]	p
GO	Proteinaceous extracellular matrix	62	-2.28	
GO	Muscle organ development	32	-2.25	
GO	Blood vessel morphogenesis	40	-2.22	
GO	Extracellular matrix	81	-2.21	
KEGG	Hypertrophic cardiomyopathy hcm	22	-2.18	

Category	Name	Size	NES	p
GO	Negative regulation of locomotion	32	-2.15	
GO	Extracellular structure organization	71	-2.14	
GO	Regulation of nervous system development	64	-2.14	
GO	Integrin binding	26	-2.14	
GO	Circulatory system development	82	-2.14	
GO	Vasculature development	52	-2.14	
KEGG	Arrhythmogenic right ventricular cardiomyopathy arvc	24	-2.13	
GO	Angiogenesis	33	-2.13	
GO	Regulation of neuron differentiation	50	-2.13	
KEGG	Dilated cardiomyopathy	24	-2.11	
KEGG	Ecm receptor interaction	26	-2.09	
GO	Regulation of neuron projection development	41	-2.09	
KEGG	Focal adhesion	58	-2.09	
GO	Regulation of cellular component movement	97	-2.09	
GO	Collagen binding	19	-2.08	

Category	Name	Size	NES	p
GO	Integrin mediated signaling pathway	21	-2.07	
GO	Regulation of cell differentiation	128	-2.04	
GO	Regulation of cellular response to growth factor stimulus	29	-2.04	
GO	Regulation of cell development	88	-2.04	
GO	Muscle structure development	55	-2.01	
GO	Skeletal system development	35	-2.00	
GO	Actin binding	93	-2.00	
GO	Response to mechanical stimulus	21	-2.00	
GO	Positive regulation of neuron differentiation	26	-1.98	
GO	Regulation of stat cascade	12	-1.97	
GO	Positive regulation of neuron projection development	24	-1.97	
GO	Regulation of peptidyl tyrosine phosphorylation	25	-1.97	0
GO	Cell adhesion molecule binding	40	-1.96	
GO	Positive regulation of cell differentiation	82	-1.96	
GO	Cell substrate junction	162	-1.95	
GO	Cell substrate adhesion	38	-1.95	

Category	Name	Size	NES	p
GO	Positive regulation of nervous system development	33	-1.95	
GO	Extracellular matrix component	38	-1.94	
GO	Anchoring junction	168	-1.94	
GO	Positive regulation of cell development	52	-1.93	
GO	Regulation of ras protein signal transduction	17	-1.92	
GO	Cell projection organization	90	-1.92	
GO	Movement of cell or subcellular component	153	-1.92	
GO	Cell matrix adhesion	30	-1.92	
GO	Cell cell junction	56	-1.91	
GO	Sulfur compound catabolic process	11	-1.91	
GO	Glycosaminoglycan binding	36	-1.91	
GO	Negative regulation of developmental process	70	-1.90	
GO	Carbohydrate derivative catabolic process	25	-1.90	
GO	Positive regulation of peptidyl tyrosine phosphorylation	17	-1.89	
GO	Cell part morphogenesis	64	-1.88	

Category	Name	Size	NES	p
GO	Cell junction	207	-1.88	
GO	Extracellular matrix structural constituent	19	-1.88	
GO	Biological adhesion	107	-1.88	
GO	Positive regulation of developmental process	108	-1.88	
GO	Skin development	18	-1.88	0
GO	Cell junction assembly	33	-1.87	
GO	Cell junction organization	37	-1.87	
GO	Regulation of cell morphogenesis	76	-1.87	
GO	Aminoglycan biosynthetic process	12	-1.87	0
GO	Regulation of cell projection organization	65	-1.86	
GO	Neurogenesis	129	-1.86	
GO	Neuron projection development	60	-1.86	
GO	Regulation of phosphatidylinositol 3 kinase signaling	12	-1.86	
GO	Actin filament bundle	24	-1.86	0
GO	Anatomical structure formation involved in morphogenesis	82	-1.85	

Category	Name	Size	NES	p
GO	Substrate adhesion dependent cell spreading	10	-1.85	
GO	Actomyosin	24	-1.85	0
GO	Collagen trimer	15	-1.85	
GO	Calmodulin binding	25	-1.84	0
GO	Cell morphogenesis involved in differentiation	56	-1.84	
GO	Extracellular matrix binding	20	-1.84	
GO	Lamellipodium	41	-1.84	
GO	Response to growth factor	54	-1.83	
GO	Golgi lumen	16	-1.82	0
GO	Protein complex binding	167	-1.82	
GO	Regulation of cell morphogenesis involved in differentiation	47	-1.82	
GO	Cytoskeleton organization	115	-1.82	
GO	Regulation of extent of cell growth	13	-1.82	
GO	Cellular component morphogenesis	91	-1.81	
GO	Regulation of axonogenesis	20	-1.81	
GO	Regulation of cell substrate adhesion	35	-1.81	

Category	Name	Size	NES	p
GO	Skeletal muscle organ development	14	-1.81	
GO	Aminoglycan catabolic process	12	-1.80	
GO	Cell development	129	-1.80	
GO	Synapse	71	-1.80	
GO	Negative regulation of cellular component organization	101	-1.80	
GO	Collagen fibril organization	11	-1.79	0
GO	Regulation of rho protein signal transduction	11	-1.79	
GO	Negative regulation of multicellular organismal process	103	-1.79	
GO	Heparin binding	30	-1.78	
GO	Plasma membrane receptor complex	15	-1.78	0
GO	Cellular response to mechanical stimulus	13	-1.78	0
GO	Muscle tissue development	31	-1.78	0
GO	Multicellular organismal macromolecule metabolic process	17	-1.77	
GO	Glial cell differentiation	21	-1.77	0
GO	Actin filament based process	86	-1.77	
GO	Gastrulation	19	-1.77	0

Category	Name	Size	NES	p
GO	Positive regulation of cell projection organization	40	-1.77	0
GO	Regulation of extrinsic apoptotic signaling pathway	21	-1.76	0
GO	Platelet degranulation	38	-1.76	
GO	Single organism cell adhesion	54	-1.76	
GO	Transforming growth factor beta receptor signaling pathway	15	-1.76	0
GO	Regulation of cell adhesion	77	-1.76	
GO	Positive regulation of locomotion	51	-1.75	0
KEGG	Regulation of actin cytoskeleton	51	-1.75	
KEGG	Viral myocarditis	15	-1.75	0
GO	Negative regulation of cell adhesion	25	-1.75	0
GO	Macromolecular complex binding	198	-1.75	
GO	Cytoskeletal protein binding	157	-1.74	
GO	Neuron projection morphogenesis	47	-1.74	
GO	Cell projection part	117	-1.74	
GO	Cardiocyte differentiation	11	-1.74	0
GO	Heart development	42	-1.74	0
GO	Neuron development	69	-1.74	
GO	Adherens junction organization	15	-1.73	0

Category	Name	Size	NES	p
GO	Regulation of synaptic plasticity	11	-1.73	0
GO	Positive regulation of cell substrate adhesion	27	-1.73	0
GO	Cellular response to amino acid stimulus	13	-1.73	0
GO	Gliogenesis	28	-1.73	
GO	Cell activation	58	-1.72	0
GO	Formation of primary germ layer	15	-1.72	0
GO	Peptidyl serine modification	13	-1.72	
GO	Secretion by cell	67	-1.71	
GO	Endodermal cell differentiation	10	-1.71	0
GO	Cell projection	216	-1.71	
GO	Cilium morphogenesis	13	-1.71	0
KEGG	Small cell lung cancer	11	-1.71	0
GO	Plasma membrane protein complex	68	-1.71	0
GO	Cell cell adhesion	52	-1.71	0
GO	Regulation of multicellular organismal development	176	-1.71	
GO	Cell growth	20	-1.70	0
KEGG	Endometrial cancer	10	-1.70	0
GO	Lymphocyte differentiation	10	-1.70	0
GO	Platelet activation	35	-1.70	0

Category	Name	Size	NES	p
GO	Vacuolar part	118	-1.70	
KEGG	Endocytosis	30	-1.70	0
GO	Endoderm development	11	-1.70	0
GO	Endoderm formation	10	-1.69	
GO	Positive regulation of cell adhesion	54	-1.69	
GO	Positive regulation of cell morphogenesis involved in differentiation	29	-1.69	0
GO	Secretory granule lumen	28	-1.69	0
GO	Glial cell development	13	-1.69	
GO	Secretion	82	-1.69	
GO	Negative regulation of nervous system development	29	-1.69	0
GO	Basement membrane	27	-1.69	0
GO	Locomotion	115	-1.69	0
GO	Tissue remodeling	14	-1.69	
GO	Cell leading edge	71	-1.68	
GO	Regulation of small gtpase mediated signal transduction	24	-1.68	0
GO	Cell cell adherens junction	16	-1.68	
GO	Extracellular space	206	-1.68	
GO	Head development	65	-1.68	0

Category	Name	Size	NES	p
GO	System process	126	-1.68	
GO	Single organism behavior	28	-1.68	
GO	Early endosome membrane	18	-1.67	
GO	Behavior	37	-1.67	
GO	Response to transforming growth factor beta	22	-1.67	
GO	Response to wounding	101	-1.67	
KEGG	Gap junction	20	-1.67	
GO	Telencephalon development	18	-1.66	
GO	Exocytosis	54	-1.66	
GO	Negative regulation of epithelial cell migration	10	-1.66	
GO	Calcium ion binding	82	-1.66	
GO	Regulated exocytosis	46	-1.66	
GO	Aminoglycan metabolic process	16	-1.66	
GO	Regulation of sodium ion transmembrane transport	10	-1.65	
GO	Sensory perception of light stimulus	14	-1.65	
GO	Negative regulation of cell differentiation	44	-1.65	
GO	Regulation of dendrite development	13	-1.65	

Category	Name	Size	NES	p
GO	Negative regulation of cellular response to growth factor stimulus	17	-1.65	0
GO	Plasma membrane region	92	-1.65	
GO	Regulation of protein kinase b signaling	10	-1.65	
GO	Cell substrate junction assembly	14	-1.65	

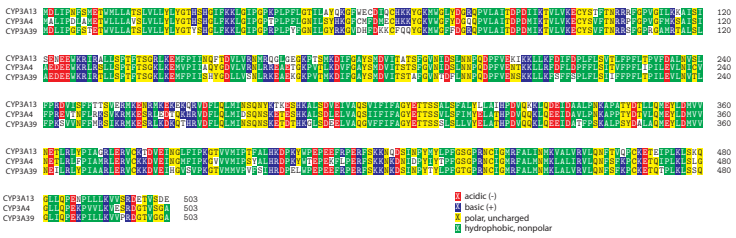


Figure A.1. Cytochrome p450 3A proteins have a high sequence homology between pig, human, and mouse proteins. CYP3A13 is a mouse protein, CYP3A4 is a human protein, and CYP3A39 is a pig protein. Color shows the polarity of each amino acid. The multiple sequence alignment was made with the R package msa **?** using the ClustalOmega method.

Appendix B

Supplementary materials for GIPR^{dn} pigs liraglutide treatment study

Table B.1. Significantly more abundant proteins in liver samples from liraglutide treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value
RBX1	E3 ubiquitin-protein ligase RBX1	1.69	6.98E-07
CIAPIN1	anamorsin	1.28	0.0025
PEF1	peflin	1.22	0.0028
OPLAH	5-oxoprolinase	0.99	9.05E-05
EIF4E2	eukaryotic translation initiation factor 4E type 2	0.98	0.0011
ORM1	alpha-1-acid glycoprotein	0.97	0.0016
FABP3	fatty acid-binding protein, heart	0.89	3.90E-04
GPAM	TRPM8 channel-associated factor 2-like	0.69	0.0019
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	0.62	4.30E-04
AK4	adenylate kinase 4, mitochondrial	0.51	1.68E-06
GARS	glycine–tRNA ligase	0.5	9.00E-04
CNDP2	cytosolic non-specific dipeptidase	0.48	0.001
SARS	serine–tRNA ligase, cytoplasmic	0.46	3.20E-04

Gene name	Protein name	l2fc	p-value
TARS	threonine–tRNA ligase, cytoplasmic	0.43	4.20E-04
GGH	gamma-glutamyl hydrolase	0.42	7.50E-04
CARS	cysteine–tRNA ligase, cytoplasmic	0.41	1.80E-04
GPAM	glycerol-3-phosphate acyltransferase 1, mitochondrial	0.38	6.70E-04
NARS	asparagine–tRNA ligase, cytoplasmic	0.38	3.10E-04
ACADSB	short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	0.37	6.40E-04
TMEM33	transmembrane protein 33	0.36	0.0016
DHRS11	dehydrogenase/reductase SDR family member 11	0.34	1.00E-04
CKAP4	cytoskeleton-associated protein 4	0.34	0.0035
AARS	alanine–tRNA ligase, cytoplasmic	0.33	0.0014
IARS	isoleucine–tRNA ligase, cytoplasmic	0.33	1.10E-04
TPP1	tripeptidyl-peptidase 1	0.32	1.87E-05
TXNL1	thioredoxin-like protein 1	0.32	8.14E-05

Gene name	Protein name	l2fc	p-value
HINT1	histidine triad nucleotide-binding protein 1	0.32	1.50E-04
SCPEP1	retinoid-inducible serine carboxypeptidase	0.32	0.0021
CS	citrate synthase, mitochondrial	0.3	6.50E-04
FARSA	phenylalanine–tRNA ligase alpha subunit	0.29	4.61E-05
NUDT21	tubulin alpha-1C chain	0.29	1.20E-04
LGMN	legumain	0.28	4.30E-04
HNRNPA0	heterogeneous nuclear ribonucleoprotein A0	0.27	0.0016
CRK	adapter molecule crk	0.26	1.20E-04
FARSB	phenylalanine–tRNA ligase beta subunit	0.25	3.60E-04
SORBS2.1	sorbin and SH3 domain-containing protein 2	0.24	3.60E-04
TXNRD2	thioredoxin reductase 2, mitochondrial	0.24	3.80E-04
CNPY2	protein canopy homolog 2	0.23	0.0021
PDHB	pyruvate dehydrogenase E1 component subunit beta, mitochondrial	0.23	0.0029
AS3MT	arsenite methyltransferase	0.23	0.0032

Gene name	Protein name	l2fc	p-value
CCT3	T-complex protein 1 subunit gamma	0.22	3.50E-04
EPRS	bifunctional glutamate/proline–tRNA ligase	0.21	3.80E-04
CCT6A	T-complex protein 1 subunit zeta	0.21	7.20E-04
CCT5	T-complex protein 1 subunit epsilon	0.2	0.0016
OSBPL1A	oxysterol-binding protein-related protein 1	0.19	0.0017
PPIA	peptidyl-prolyl cis-trans isomerase A	0.19	0.0019
TCP1	T-complex protein 1 subunit alpha	0.19	0.0034
CCT2	T-complex protein 1 subunit beta	0.19	0.0011
GPX1	glutathione peroxidase 1	0.18	0.0024
EIF3E	eukaryotic translation initiation factor 3 subunit E	0.18	7.44E-05
GSPT1	eukaryotic peptide chain release factor GTP-binding subunit ERF3A	0.18	5.30E-04
HSPA9	stress-70 protein, mitochondrial	0.18	0.0013

Gene name	Protein name	l2fc	p-value
ABCE1	ATP-binding cassette sub-family E member 1	0.17	0.0012
NUDT21	cleavage and polyadenylation specificity factor subunit 5	0.17	0.0035
CCT8	T-complex protein 1 subunit theta	0.17	0.0037
ENO1	alpha-enolase	0.16	6.60E-04
UMPS	uridine 5-monophosphate synthase	0.16	0.0019
UBE2N	ubiquitin-conjugating enzyme E2 N	0.16	0.0032
PSMA1	proteasome subunit alpha type-1	0.16	0.0033
MYLK	myosin light chain kinase, smooth muscle	0.15	0.0036
PKP2	plakophilin-2	0.15	0.0024
HSPA4	heat shock 70 kDa protein 4	0.14	3.60E-04
SFPQ	splicing factor, proline- and glutamine-rich	0.13	0.0013
RPL7	60S ribosomal protein L7	0.11	0.0014

Table B.2. Significantly more abundant proteins in liver samples from placebo treated GIPR^{dn} pigs (q-value < 0.05).

Gene name	Protein name	l2fc	p-value
HMGCS2	hydroxymethylglutaryl-CoA synthase, mitochondrial precursor	-4.60	3.60E-04
LAMC1	laminin subunit gamma-1 precursor	-1.38	2.60E-04
OAT	ornithine aminotransferase, mitochondrial	-1.38	5.90E-04
S100A1	protein S100-A1	-1.12	7.02E-05
STEAP4	metalloreductase STEAP4	-0.97	0.0022
DHRS7	dehydrogenase/reductase SDR family member 7	-0.93	1.70E-04
GNMT	glycine N-methyltransferase	-0.86	4.00E-04
FOLR2	folate receptor 2	-0.83	5.57E-06
GATM	glycine amidinotransferase, mitochondrial	-0.83	7.53E-05
GSTM1	glutathione S-transferase Mu 1	-0.75	3.19E-05
PCTP	phosphatidylcholine transfer protein	-0.68	0.0035
SUOX	sulfite oxidase, mitochondrial	-0.60	1.10E-04

Gene name	Protein name	l2fc	p-value
FDX1	adrenodoxin, mitochondrial precursor	-0.60	1.50E-05
AADAC	arylacetamide deacetylase-like	-0.53	1.30E-04
FBP1	fructose-1,6-bisphosphatase 1	-0.53	1.40E-04
LDHD	probable D-lactate dehydrogenase, mitochondrial	-0.50	8.60E-04
CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1	-0.45	2.60E-04
CYP2C33	cytochrome P450 2C33	-0.44	7.90E-04
AMT	aminomethyltransferase, mitochondrial	-0.44	1.40E-04
PLS3	plastin-3	-0.43	2.00E-04
MVP	major vault protein	-0.41	5.30E-04
CYP1A2	cytochrome P450 1A2	-0.41	7.40E-04
GK	glycerol kinase	-0.41	4.70E-04
OTC	ornithine carbamoyltransferase, mitochondrial	-0.41	9.80E-04
MAOB	amine oxidase [flavin-containing] B	-0.40	0.0017
CHDH	choline dehydrogenase, mitochondrial	-0.39	5.00E-04
EHD1	EH domain-containing protein 1	-0.39	4.80E-04

Gene name	Protein name	l2fc	p-value
DHRS4	dehydrogenase/reductase SDR family member 4	-0.39	8.00E-04
PLIN3	perilipin-3	-0.36	2.40E-04
GYS2	glycogen [starch] synthase, liver	-0.36	5.10E-04
LPGAT1	acyl-CoA:lysophosphatidylglycerol acyltransferase 1	-0.35	6.20E-04
SELENBP1	selenium-binding protein 1	-0.35	9.10E-04
RCN1	reticulocalbin-1	-0.35	0.0036
EGFR	epidermal growth factor receptor precursor	-0.34	0.0029
SLC25A13	calcium-binding mitochondrial carrier protein Aralar2	-0.34	1.38E-05
NNT	NAD(P) transhydrogenase, mitochondrial	-0.32	1.16E-05
DPYS	dihydropyrimidinase	-0.31	0.0031
MTTP	microsomal triglyceride transfer protein large subunit precursor	-0.31	0.0021
STOM	erythrocyte band 7 integral membrane protein	-0.31	8.00E-04
LMAN1	protein ERGIC-53	-0.29	7.46E-06
GCA	grancalcin	-0.29	8.40E-04

Gene name	Protein name	l2fc	p-value
ASPH	aspartyl/asparaginyl beta-hydroxylase9	-0.28	1.40E-04
ARFGAP3	ADP-ribosylation factor GTPase-activating protein 3	-0.28	3.44E-05
UROC1	urocanate hydratase	-0.27	3.20E-04
PHGDH	D-3-phosphoglycerate dehydrogenase	-0.25	0.0025
LAP3	cytosol aminopeptidase	-0.25	7.30E-04
ADHFE1	hydroxyacid-oxoacid transhydrogenase, mitochondrial	-0.25	0.0033
GPD1	glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	-0.24	2.81E-05
CTH	cystathionine gamma-lyase	-0.23	0.0022
ALAD	delta-aminolevulinic acid dehydratase	-0.23	0.0016
PEBP1	phosphatidylethanolamine-binding protein 1	-0.22	0.0033
MLEC	malectin	-0.22	0.0014
PCYOX1	prenylcysteine oxidase 1	-0.21	9.60E-04
SUB1	activated RNA polymerase II transcriptional coactivator p15	-0.20	0.0034

Gene name	Protein name	l2fc	p-value
STT3A	dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit STT3A	-0.18	8.60E-04
ARHGAP1	rho GTPase-activating protein 1	-0.17	0.0027
PDIA6	protein disulfide-isomerase A6 precursor	-0.17	0.0028
RAB11B	ras-related protein Rab-11B	-0.17	0.0018
EEF1G	elongation factor 1-gamma	-0.16	0.0026
LRP1	prolow-density lipoprotein receptor-related protein 1	-0.14	0.0030
SEC22B	vesicle-trafficking protein SEC22b	-0.13	0.0033
TKT	transketolase	-0.13	3.90E-04
NPEPPS	puromycin-sensitive aminopeptidase	-0.12	0.0028

Table B.3. GSEA: Significantly enriched gene set
in liraglutide treated in GIPR^{dn} pig livers (q-value <
0.05).

Category	Name	Size	NES	p-va
KEGG	Aminoacyl tRNA biosynthesis	25	2.6	0
GO	Ligase activity forming carbon oxygen bonds	26	2.53	0
GO	Catalytic activity acting on RNA	63	2.53	0
GO	Regulation of mRNA catabolic process	79	2.51	0
GO	Amino acid activation	29	2.44	0
GO	Catalytic activity acting on a tRNA	30	2.41	0
GO	NcRNA metabolic process	88	2.39	0
GO	Regulation of cellular amino acid metabolic process	44	2.37	0
GO	Interleukin 1 mediated signaling pathway	48	2.36	0
GO	Rna metabolic process	237	2.36	0
GO	Scf dependent proteasomal ubiquitin dependent protein catabolic process	45	2.35	0
GO	Regulation of hematopoietic progenitor cell differentiation	43	2.34	0
GO	Anaphase promoting complex dependent catabolic process	43	2.34	0

GO	Negative regulation of cell cycle g2 m phase transition	47	2.32	0
GO	Endopeptidase complex	50	2.29	0
GO	Regulation of stem cell differentiation	45	2.29	0
GO	Regulation of transcription from RNA polymerase ii promoter in response to hypoxia	45	2.28	0
GO	Stem cell differentiation	60	2.27	0
KEGG	Proteasome	38	2.27	0
GO	Regulation of cell cycle g2 m phase transition	68	2.27	0
GO	Posttranscriptional regulation of gene expression	182	2.27	0
GO	Protein modification by small protein removal	72	2.26	0
GO	Cell cycle g2 m phase transition	76	2.26	0
GO	Innate immune response activating cell surface receptor signaling pathway	45	2.24	0
GO	Hematopoietic stem cell differentiation	43	2.24	0
GO	Hematopoietic progenitor cell differentiation	50	2.22	0

GO	Protein polyubiquitination	63	2.21	0
GO	Negative regulation of wnt signaling pathway	57	2.21	0
GO	Negative regulation of cell cycle process	69	2.21	0
GO	Regulation of chromosome organization	47	2.2	0.0
GO	TRNA metabolic process	46	2.2	0
GO	Canonical wnt signaling pathway	72	2.2	0
GO	Ribonucleoprotein complex	261	2.2	0
GO	Rna localization	71	2.2	0
GO	Translation initiation factor activity	30	2.18	0
GO	Response to interleukin 1	66	2.18	0
GO	Ribonucleoprotein complex biogenesis	118	2.18	0
GO	Non canonical wnt signaling pathway	55	2.16	0
GO	Negative regulation of cell cycle phase transition	57	2.16	0
GO	Tumor necrosis factor mediated signaling pathway	52	2.16	0

GO	Regulation of DNA templated transcription in response to stress	52	2.16	0
GO	Chromosome organization	126	2.15	0
GO	Regulation of establishment of planar polarity	52	2.15	0
GO	Positive regulation of protein localization to nucleus	29	2.14	0
GO	Peptide biosynthetic process	230	2.14	0
GO	Regulation of telomere maintenance via telomere lengthening	18	2.14	0
GO	Fc epsilon receptor signaling pathway	47	2.13	0
GO	Regulation of wnt signaling pathway	78	2.13	0
GO	Rna catabolic process	176	2.13	0
GO	Regulation of morphogenesis of an epithelium	57	2.13	0
GO	Regulation of telomere maintenance	21	2.13	0
GO	Positive regulation of chromosome organization	32	2.13	0

GO	Rna dependent DNA biosynthetic process	21	2.12	0
GO	Catalytic complex	301	2.12	0
GO	Positive regulation of telomere maintenance via telomere lengthening	15	2.12	0
GO	Rna splicing via transesterification reactions	109	2.12	0
GO	Mitochondrial respiratory chain complex i	36	2.11	0
GO	Negative regulation of mitotic cell cycle	64	2.11	0
GO	Telomere maintenance via telomere lengthening	22	2.11	0
GO	Positive regulation of wnt signaling pathway	56	2.11	0
GO	Nadh dehydrogenase activity	33	2.09	0
GO	Threonine type peptidase activity	17	2.08	0
GO	Positive regulation of telomere maintenance	17	2.08	0
GO	Rna splicing	125	2.08	0
GO	Nik nf kappab signaling	59	2.08	0

GO	Regulation of cellular ketone metabolic process	55	2.07	0
GO	Positive regulation of canonical wnt signaling pathway	53	2.07	0
GO	MRNA metabolic process	130	2.07	0
GO	Regulation of protein localization to nucleus	34	2.06	0
GO	Peptidase complex	55	2.04	0
GO	Proteasomal ubiquitin independent protein catabolic process	17	2.04	0
GO	T cell receptor signaling pathway	60	2.04	0
GO	Ribonucleoprotein granule	58	2.04	0
GO	Protein modification by small protein conjugation or removal	156	2.02	0
GO	Cytoplasmic stress granule	23	2.02	0
GO	Telomere organization	28	2.02	0
GO	Amide biosynthetic process	267	2.02	0
GO	Dna metabolic process	103	2.02	0
GO	MRNA binding	82	2.02	0
GO	Antigen processing and presentation of exogenous peptide antigen via mhc class i	47	2.01	0
GO	Chaperone complex	16	2.01	0

GO	Nadh dehydrogenase complex assembly	36	2.01	0
GO	Antigen receptor mediated signaling pathway	60	2	0
GO	Protein modification by small protein conjugation	133	2	0
GO	Mitochondrial electron transport nadh to ubiquinone	37	2	0
GO	Cell cell signaling by wnt	105	2	0
GO	Translation factor activity RNA binding	45	1.99	0
GO	Translation regulator activity nucleic acid binding	53	1.98	0
GO	Morphogenesis of a polarized epithelium	58	1.97	0
GO	Proteasome accessory complex	22	1.97	0
GO	Regulation of cellular amine metabolic process	49	1.96	0
KEGG	Lysosome	43	1.96	0
GO	Regulation of cellular amide metabolic process	102	1.96	0
GO	Cell surface receptor signaling pathway involved in cell cell signaling	109	1.96	0

GO	Proteasome core complex	18	1.95	0.0
GO	TRNA binding	23	1.95	0.0
GO	Proteasomal protein catabolic process	111	1.95	
GO	Ribonucleoprotein complex subunit organization	86	1.94	
GO	Translation regulator activity	61	1.94	
GO	Dna biosynthetic process	34	1.93	
GO	Nucleobase containing small molecule biosynthetic process	31	1.93	0.0
GO	Negative regulation of cellular amide metabolic process	39	1.92	
GO	Ligase activity	73	1.92	
GO	Regulation of cell cycle phase transition	94	1.92	
GO	Ribonucleoprotein complex binding	47	1.92	
GO	Regulation of animal organ morphogenesis	63	1.92	
GO	Negative regulation of cell cycle	108	1.91	
GO	Single stranded RNA binding	32	1.9	
GO	Positive regulation of DNA biosynthetic process	22	1.89	0.0

GO	Modification dependent macromolecule catabolic process	127	1.88	0
GO	Regulation of translational initiation	24	1.88	0.0
GO	Cytoplasmic translational initiation	18	1.87	0
GO	Spliceosomal complex	68	1.87	0
GO	Regulation of cell cycle process	132	1.86	0
GO	Purine ntp dependent helicase activity	16	1.86	0.0
GO	Mitochondrial respiratory chain complex assembly	43	1.85	0
KEGG	Spliceosome	60	1.85	0.0
GO	NcRNA processing	58	1.85	0.0
GO	Catalytic step 2 spliceosome	45	1.84	0
GO	Dna geometric change	18	1.84	0.0
GO	Translational initiation	115	1.83	0
GO	Catalytic activity acting on DNA	16	1.83	0.0
GO	Cellular macromolecule catabolic process	314	1.83	0
GO	Negative regulation of RNA catabolic process	15	1.83	0.0
GO	Atp synthesis coupled electron transport	55	1.82	0

GO	Protein stabilization	56	1.82	0
GO	Regulation of steroid biosynthetic process	26	1.81	0.0
GO	MRNA 3 utr binding	23	1.81	0.0
GO	Regulation of DNA metabolic process	46	1.81	0
GO	Helicase activity	20	1.81	0
GO	Respiratory chain complex	49	1.8	0.0
GO	Regulation of DNA biosynthetic process	25	1.8	0.0
GO	Regulation of cell cycle	185	1.79	0
GO	Nucleoplasm part	136	1.79	0
GO	Protein localization to nucleus	67	1.79	0
GO	Cell cycle phase transition	116	1.79	0.0
GO	Regulation of nucleobase containing compound metabolic process	103	1.78	0
GO	Cell recognition	22	1.78	0
GO	Cellular amide metabolic process	334	1.78	0
GO	Activation of innate immune response	89	1.77	0
GO	Lysosomal lumen	43	1.77	0
GO	Single fertilization	18	1.77	0.0
GO	Macromolecule catabolic process	370	1.76	0

GO	Regulation of innate immune response	117	1.76	0
GO	Fertilization	21	1.76	0.0
GO	Regulation of alcohol biosynthetic process	23	1.76	0.0
GO	Regulation of cholesterol biosynthetic process	22	1.75	0.0
GO	Regulation of mitotic cell cycle	122	1.75	0
KEGG	Oocyte meiosis	25	1.75	0
GO	Ribosome biogenesis	59	1.75	0.0
GO	Negative regulation of nucleobase containing compound metabolic process	29	1.74	0.0
GO	Fc receptor signaling pathway	68	1.74	0
GO	Unfolded protein binding	56	1.74	0.0
GO	Cellular response to DNA damage stimulus	88	1.73	0
GO	Cellular response to oxygen levels	77	1.73	0
GO	Response to tumor necrosis factor	76	1.73	0.0
GO	Nuclear body	117	1.72	0.0
GO	Rna helicase activity	15	1.72	0.0

GO	Nucleoside phosphate catabolic process	22	1.71	0.0
GO	Regulatory region nucleic acid binding	52	1.71	0
GO	Negative regulation of RNA metabolic process	18	1.71	0.0
GO	Positive regulation of DNA metabolic process	34	1.71	0.0
GO	Regulation of cellular catabolic process	214	1.71	0
GO	Dna repair	59	1.71	0.0
GO	Atpase activity coupled	80	1.7	0
GO	Respirasome	53	1.69	0.0
GO	Nucleic acid phosphodiester bond hydrolysis	35	1.69	0.0
GO	Cytoplasmic translation	50	1.68	0.0
GO	AlteRNAtive mRNA splicing via spliceosome	25	1.68	0.0
GO	Immune response regulating cell surface receptor signaling pathway	87	1.68	0
GO	Transcription factor binding	80	1.68	0.0

GO	Positive regulation of nucleobase containing compound metabolic process	54	1.67	0.0
GO	Positive regulation of defense response	117	1.67	0
KEGG	Wnt signaling pathway	20	1.67	0.0
GO	Negative regulation of growth	23	1.67	0.0
GO	Cell cycle	267	1.67	0
GO	Covalent chromatin modification	37	1.66	0.0
GO	Protein kinase c binding	16	1.66	0.0
GO	Oxidative phosphorylation	74	1.66	0
GO	Dna conformation change	36	1.66	0.0
GO	Oxidoreductase activity acting on nad p h quinone or similar compound as acceptor	42	1.66	0.0
GO	Microtubule cytoskeleton	176	1.66	0
GO	Transcription coupled nucleotide excision repair	15	1.65	0.0
GO	Cell cycle process	204	1.65	0
GO	Organic cyclic compound catabolic process	253	1.64	0
GO	Mitotic cell cycle	170	1.64	0
GO	RRNA metabolic process	39	1.64	0.0
GO	Transferase complex	84	1.64	0.0

GO	Cellular ketone metabolic process	82	1.64	0.0

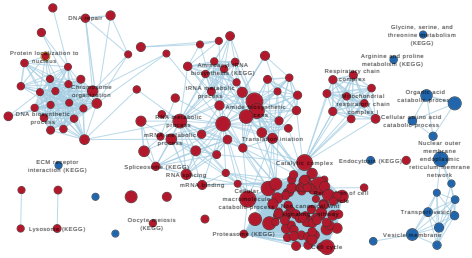


Figure B.1. Network map showing significant (q-value < 0.05) GO ontology and KEGG categories enriched in LT (red) and PT (blue) pigs using GSEA. Annotated categories were either KEGG pathways or GO cateories significant at q-value < 0.01. The size of the nodes reflect the gene set size and the size of

Table B.4. GSEA: Significantly enriched gene sets in placebo treated in GIPR^{dn} pigs (q-value < 0.05).

Category	Name	Size	NES	p-va
KEGG	Glycine serine and threonine metabolism	23	-2.18	0
KEGG	Ecm receptor interaction	23	-2.16	0
GO	Cellular amino acid catabolic process	63	-2.15	0
GO	Transport vesicle	74	-2.15	0
GO	Copi coated vesicle	15	-2.14	0
GO	Organic acid catabolic process	161	-2.12	0
GO	Golgi associated vesicle membrane	38	-2.08	0
GO	Organic acid transmembrane transport	19	-2.06	0
GO	Alpha amino acid catabolic process	55	-2.03	0
GO	Golgi associated vesicle	56	-2.01	0
GO	Phagocytic vesicle	34	-2	0
KEGG	Arginine and proline metabolism	32	-2	0
GO	Coated vesicle	84	-1.98	0
GO	Vesicle membrane	169	-1.96	0
GO	Coated vesicle membrane	59	-1.96	0.0
GO	Small molecule catabolic process	212	-1.96	0
GO	Endoplasmic reticulum golgi intermediate compartment membrane	26	-1.95	0
GO	Nuclear outer membrane endoplasmic reticulum membrane network	264	-1.94	0
GO	Regulation of protein targeting	18	-1.93	0.0

Appendix C

Additional materials and methods

Table C.1. All parameters used in the prote
analysis in Maxquant.

Parameter	
Version	
User name	
Machine name	E
Date of writing	02/0
Include contaminants	

Parameter	
PSM FDR	
PSM FDR Crosslink	
Protein FDR	
Site FDR	
Use Normalized Ratios For Occupancy	
Min. peptide Length	
Min. score for unmodified peptides	
Min. score for modified peptides	
Min. delta score for unmodified peptides	
Min. delta score for modified peptides	
Min. unique peptides	
Min. razor peptides	
Min. peptides	
Use only unmodified peptides and	
Modifications included in protein quantification	Oxidation (M);A
Peptides used for protein quantification	
Discard unmodified counterpart peptides	
Label min. ratio count	
Use delta score	
iBAQ	
iBAQ log fit	
Match between runs	
Matching time window [min]	

Parameter	
Alignment time window [min]	
Find dependent peptides	
Fasta file	GCF_000003025.
Decoy mode	
Include contaminants	
Advanced ratios	
Fixed andromeda index folder	
Temporary folder	
Combined folder location	
Second peptides	
Stabilize large LFQ ratios	
Separate LFQ in parameter groups	
Require MS/MS for LFQ comparisons	
Calculate peak properties	
Main search max. combinations	
Advanced site intensities	
Write msScans table	
Write msmsScans table	
Write ms3Scans table	
Write allPeptides table	
Write mzRange table	
Write pasefMsmsScans table	
Write accumulatedPasefMsmsScans table	

Parameter
Max. peptide mass [Da]
Min. peptide length for unspecific search
Max. peptide length for unspecific search
Razor protein FDR
Disable MD5
Max mods in site table
Match unidentified features
Epsilon score for mutations
Evaluate variant peptides separately
Variation mode
MS/MS tol. (FTMS)
Top MS/MS peaks per Da interval. (FTMS)
Da interval. (FTMS)
MS/MS deisotoping (FTMS)
MS/MS deisotoping tolerance (FTMS)
MS/MS deisotoping tolerance unit (FTMS)
MS/MS higher charges (FTMS)
MS/MS water loss (FTMS)
MS/MS ammonia loss (FTMS)
MS/MS dependent losses (FTMS)
MS/MS recalibration (FTMS)
MS/MS tol. (ITMS)
Top MS/MS peaks per Da interval. (ITMS)

Parameter
Da interval. (ITMS)
MS/MS deisotoping (ITMS)
MS/MS deisotoping tolerance (ITMS)
MS/MS deisotoping tolerance unit (ITMS)
MS/MS higher charges (ITMS)
MS/MS water loss (ITMS)
MS/MS ammonia loss (ITMS)
MS/MS dependent losses (ITMS)
MS/MS recalibration (ITMS)
MS/MS tol. (TOF)
Top MS/MS peaks per Da interval. (TOF)
Da interval. (TOF)
MS/MS deisotoping (TOF)
MS/MS deisotoping tolerance (TOF)
MS/MS deisotoping tolerance unit (TOF)
MS/MS higher charges (TOF)
MS/MS water loss (TOF)
MS/MS ammonia loss (TOF)
MS/MS dependent losses (TOF)
MS/MS recalibration (TOF)
MS/MS tol. (Unknown)
Top MS/MS peaks per Da interval. (Unknown)
Da interval. (Unknown)

Parameter	
MS/MS deisotoping (Unknown)	
MS/MS deisotoping tolerance (Unknown)	
MS/MS deisotoping tolerance unit (Unknown)	
MS/MS higher charges (Unknown)	
MS/MS water loss (Unknown)	
MS/MS ammonia loss (Unknown)	
MS/MS dependent losses (Unknown)	
MS/MS recalibration (Unknown)	
Site tables	Oxidati

Acknowledgements

First of all I would like to thank Prof. Dr. Eckhard Wolf for supervising my doctoral thesis, and for his advise and help during my time at the Gene Center.

Thank you to the people of the Arnold/Fröhlich lab for housing me. Special thanks to Dr. Thomas Fröhlich and Dr. George Arnold for all help I have received in matters of science and proteomics. Similarly, thank you to Dr. Florian Flenkenthaler for advice regarding all things proteomics.

Thank you to the graduate school Quantitative Biosciences Munich (QBM) for funding my PhD and giving me this opportunity. I would like to give a special thanks to Prof. Dr. Ulrike Gaul for creating this programme. Thank you all members of the QBM team

for all help I have received throughout my years in Munich, and for organizing courses and scientific gatherings. Thanks to all friends I have made in QBM and for making my time in Munich enjoyable!

Thank you to Mattias Backman for many interesting discussions regarding the science of our parallel projects and many other topics!

I would like to thank my family and friends for their support during these last four years.

Por último y no menos importante, gracias a mi novia Daniela por su apoyo.