Bioinformatic analysis of multiomic data from the Munich MIDY Pig biobank

von Mattias Emanuel Backman

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

Bioinformatic analysis of multiomic data from the Munich MIDY Pig biobank

von

Mattias Emanuel Backman

aus Kungälv, 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

Angefertigt am Laboratorium für Funktionale Genomanalyse (LAFUGA)
Genzentrum der Ludwig-Maximilians-Universität
Genomics

Mentor: Dr. Helmut Blum

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
Korreferent/en:	Priv.-Doz. Dr. Sven Reese Univ.-Prof. Dr. Markus Meißner Priv.-Doz. Dr. Matthias Eddicks Univ.-Prof. Dr. Mathias Ritzmann

Tag der Promotion: 08.02.2020

Table of Contents

List of Figures

List of Tables

Index of Abbreviations

AAA	Aromatic amino acid
AASS	Aminoadipate-semialdehyde synthase
Ac-Orn	N-acetylornithine
AC	Acyl carnitine
ADAMTS17	A disintegrin and metalloproteinase with thrombospondin type 1 motif 17
ADMA	Asymmetric dimethylarginine
AFMID	Arylformamidase
AGEs	Advanced glycation end products
AGXT	Alanine-glyoxylate aminotransferase
AHCY	Adenosyl- homocysteinase
AKT	Protein kinase B
ALDH7A1	Aldehyde dehydrogenase 7 family member A1
AMPK	AMP-activated protein kinase
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
APOA	Apolipoprotein A
ARG1	Arginase 1
ASS1	Argininosuccinate Synthase 1
ATP	Adenosine triphosphate
atRA	All-trans retinoic acid
BAT	Brown adipose tissue
BCAA	Branched-chain amino acid
BCA	Bicinchoninic acid
BDH1	3-hydroxybutyrate dehydrogenase 1
BH	Benjamini Hochberg
BLAST	Basic local alignment search tool
bp	Base pairs
C(n)	Acyl carnitine chain length
CBS	Cystathionine beta-synthetase
CCA	Canonical correlation analysis
cDNA	Complementary DNA
CDO1	Cysteine dioxygenase 1
Cer	Ceramide
CE	Cholesteryl ester
Chol	Cholesterol
CL	Cardiolipin
CoA	Coenzyme A
COL	Collagen
CPS1	Carbamoyl-phosphate synthase 1
CPT1	Carnitine Palmitoyltransferase 1
CRP	C-reactive protein

CTH	Cystathionine gamma-lyase
DAG	Diacylglycerol
DAMP	Damage-associated molecular patterns
DC	Dicarboxylated
DM1	Diabetes mellitus type 1
DM2	Diabetes mellitus type 2
DMA	Dimethylarginine
DNA	Deoxyribonucleic acid
ER	Endoplasmatic reticulum
ES	Enrichment score
FASTA	FASTA sequence data format
FASTQ	FASTQ sequence data format
Fbpase	Fructose 1,6-bisphosphatase
FDR	False discovery rate
FFA	Free fatty acid
FOXA2	Forkhead Box A2
FOXO1	Forkhead box protein O1
FWER	Familywise error rate
G6Pase	Glucose 6 phosphatase
GLM	Generalized linear model
GLS2	Glutaminase
GLUT	Glucose transporter
GMPR	GMP reductase 1
GNMT	Glycine N-methyltransferase
GOT1	Glutamicoxaloacetic transaminase 1
GO	Gene ontology
GPT2	Glutamic-pyruvic transaminase 2
GPX	Glutathione peroxidases
GSEA	Gene set enrichment analysis
GSH	Glutathione
GSK3B	Glycogen synthase 3 beta
GSSG	Glutathione disulfide
GSTM2	Glutathione S-transferase mu 2
GYS2	Glycogen synthase 2
H1	Hexose
H2O2	Hydrogen peroxide
HAL	Histidine ammonia-lyase
HGD	Homogentisate 1,2-dioxygenase
HGNC	Human gene nomenclature committee
HLA	Human leukocyte antigen
HMGB1	High mobility group protein B1
HMGCL	3-hydroxy-3-methylglutaryl-CoA lyase
HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2
HPLC	High-performance liquid chromatography

HSC	Hepatic stellate cell
IBA1	Ionized calcium-binding adapter molecule 1
IDE	Intergeated development environment
IDH1	Isocitrate dehydrogenase 1
IGV	Integrated genomics viewer
Ins2	Insulin 2 gene of mouse
INSR	Insulin receptor
INS	Insulin
IRS	Insulin receptor substrate
KEGG	Kyoto encyclopedia of genes and genomes
KYAT1	Kynurenine aminotransferase 1
l2fc	Log2 fold change
LC	Liquid chromatography
LDH	L-lactate dehydrogenase
LDL	Low-density lipoprotein
LGMN	Legumain
LOC	Locus
LOD	Limit of Detection
log2	Binary logarithm
Lys-C	Endoproteinase Lys-C
MAPK3	Mitogen-activated protein kinase 3
MAT1A	Methionine adenosyltransferase 1A
Met-SO	Methionine sulfoxide
MHC	Major histocompatibility complex
MIDY	Mutant INS-gene-induced diabetes of youth
MODY	Maturity-onset diabetes of the youth
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MSigDB	Molecular signatures database
MS	Mass spectrometry
MTHF	Methyltetrahydrofolate
mTORC1	Mechanistic target of rapamycin complex 1
mTOR	Mechanistic target of rapamycine
MTR	Methionine synthase
MUFA	Mono-unsaturated fatty acid
NADPH	Nicotinamide adenine dinucleotide phosphate
NAFLD	Non-alcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
na	Not available
NA	Not available
NES	Normalized enrichment score
NMR	Nuclear magnetic resonance

OAS2	2-5-oligoadenylate synthetase 2
OH	Hydroxylated
ORF	Open reading frame
OTC	Ornithine carbamoyltransferase
PAMP	Pathogen-associated molecular pattern
PC(n)	Principal component number n
PCA	Principal component analysis
PCK1	phosphoenolpyruvate carboxykinase
PCR	Polymerase chain-reaction
PC	Glycerophosphocholines
PC	Pyruvate carboxylase
PDPK1	3-phosphoinositide-dependent protein kinase-1
PD	Privatdozent
PEA	Phenethylamine
PEPCK	Phosphoenolpyruvate carboxykinase
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
PI3K	Phosphoinositide 3-kinase
PKB	Protein Kinase B
ployT	Polytyrosine
polyA	Polyadenyl
PPAR	Peroxisome Proliferator Activated Receptor
PPP2R1A	Protein phosphatase 2 scaffold subunit A alpha
PP	Pancreatic polypeptide
PRR	Pattern recognition receptor
PSME2	Proteasome activator complex subunit 2
PUFA	Poly-unsaturated fatty acid
PVDF	Polyvinylidene fluoride
p	Probability value
qPCR	Quantitative polymerase chain reaction
RAC1	Rac family small GTPase 1
RAGE	Receptor for advanced glycation end products
RAM	Random access memory
RAR	Retinoic acid receptor
RDH	Retinol dehydrogenase
RFX5	Regulatory factor X5
RIN	RNA integrity number
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rRNA	ribosomal RNA
RXR	Retinoid X receptor
S100A1	S100 calcium-binding protein A1
SAH	S-adenosyl homocysteine

SAM	S-adenosylmethionine / Significance analysis of microarrays
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDS	Serine dehydratase
SEC61B	Protein transport protein Sec61 subunit beta
SEM	Standard error of the mean
SFA	Saturated fatty acid
SHROOM3	Shroom Family Member 3
SLA	Sus leukocyte antigen
SLC	Solute carrier
SM	Sphingomyelins
SREBF	Sterol regulatory element-binding factor
SREBP	Sterol regulatory element-binding protein
STZ	Streptozocin
SV(n)	Surrogate variable number n
SVA	Surrogate variable analysis
t4-OH-Pro	Hydroxyproline
TAG	Triacylglycerol
TAG	Triacylglycerol
TCA	Tricarboxylic acid
TG	Transgenic
THF	Tetrahydrofolate
TLR	Toll-like receptor
TPM	Transcripts per million
UBE2D2	Ubiquitin-conjugating enzyme E2 D2
UCP	Uncoupling protein
UMI	Unique molecular identifier
WT	Wild type

Chapter 1

Introduction

Diabetes mellitus (DM) is the fastest growing disease in the world with over 400 million sufferers [Atlas]. As the world grows wealthier and access to large quantities of food with high glucose content increases the incidence of metabolic diseases follows. This is especially true in developing countries where the local culture regarding food has to adapt to not only an increase in the quantity of food but also an influx of western-style foods, many of which are more energy dense than the native diet [Hu, 2011]. As an example, many pacific islands have an adult population that are over 70% overweight, with American Samoa reaching close to 90% and an associated diabetes rate of almost 50% [Hawley and McGarvey, 2015]. Furthermore, some populations, such as south Asian people, are thought to have a genetic predisposition to diabetes and associated diseases [Chowdhury et al., 2014].

Aside from being a common disease DM is also one of the world's oldest known. The name "diabetes" derives from "to pass through" in ancient Greek, which relates to the excessive urination of sufferers. DM shares this name with the unrelated disease diabetes insipidus. The word "mellitus" means sweet and refers to the excess of glucose in the urine making it smell sweeter than normal. Even with such ancient origins, it took until the turn of the century until it was understood as a disease of the pancreas, and about 20 years more until insulin could be isolated [Eknoyan and Nagy, 2005]. Before the discovery of insulin as the treatment of diabetes very few sufferers survived for long, especially in the case of type 1 diabetes as it is more severe and occurs early in life. Nowadays people with well-controlled diabetes live relatively normal lives although usually with eventual chronical problems at older ages and slightly shorter life spans [Miller et al., 2012].

With the frequency of diabetes increasing much research is focused on understanding the origin of the disease. An adjacent field, more commonly studied by medical professionals,

are the effects of diabetes on the sufferers. While complications such as diabetic foot ulcers and diabetic glaucoma are well known, not as much is known about the holistic chronic effects of hyperglycemia on the body as most research is performed on mice, an animal not so close to humans in metabolism. This work is thus concerned about the effect of chronic insulin deficiency and hyperglycemia in pigs, an animal remarkably similar to humans in metabolism and anatomy. The Munich MIDY Pig biobank, which was conceived shortly before the start of this project as a source for controlled and translational diabetes research, serves as the perfect research subject to answer these type of questions. This work describes the thorough multiomics analysis of the MIDY pig liver as well as the more recent analysis of adipose tissues in the biobank.

The work described in this dissertation is also part of two publications:

M. Backman, F. Flenkenthaler, Andreas B., M. Dahlhoff, E. Ländström, S. Renner, et al. Multi-omics insights into functional alterations of the liver in insulin-deficient diabetes mellitus. *Mol Metab*, 26:30–44, 2019.

A. Blutke, S. Renner, F. Flenkenthaler, M. Backman, S. Haesner, E. Kemter, et al. The munich midy pig biobank - a unique resource for studying organ crosstalk in diabetes. *Mol Metab*, 6(8):931–940, 2017.

Chapter 2

Literature review

2.1 Sub-types of Diabetes mellitus

Traditionally diabetes as a disease is divided into three major classes along with numerous smaller groups.

2.1.1 Diabetes mellitus type 1 (DM1)

DM1 is a chronic disease that typically manifests during childhood or teenage years. The disease currently has no cure, however, treatment is sufficiently advanced that sufferers can live a relatively normal life when handled well. The origin of DM1 is complex, environmental triggers lead to an autoimmune attack on the insulin-producing beta cells of the pancreas which causes a reduced insulin output and eventually a total loss of beta cells leading to chronic hyperglycemia. The exact cause of this autoimmune response has not yet been elucidated, it is thought that antibodies against beta cell antigens in combination with dysregulated immune cells are at least part of the pathogenesis [Kahaly and Hansen, 2016]. There is a genetic component as the disease tends to run in families with multiple genes known to increase susceptibility. Most notable are the genes of the MHC class II, aka the HLA genes, as they modulate histocompatibility which is important for autoimmune responses. The incidence of DM1 worldwide is 40 million [Tuomilehto, 2013]. The incidence rate varies greatly between parts of the world with richer countries, particularly those of northern Europe, generally having a higher rate of incidence, however, the number of sufferers have been steadily rising every year.

2.1.2 Diabetes mellitus type 2 (DM2)

Contrary to DM1, DM2 is a disease of metabolic origin. The root of the disease is insulin resistance, a condition where the cells of the body lose sensitivity to insulin signaling [Cefalu, 2001]. The primary tissues to be affected by insulin resistance are skeletal muscle and adipose tissue as these tissues feature the insulin-dependent transporter GLUT4 [Mueckler, 2001]. Insulin resistance is not enough for a DM2 diagnosis, DM2 is a combination of insulin resistance and beta-cell failure [Matveyenko and Butler, 2008]. DM2 generally occurs later in life and is closely associated with obesity and a sedentary lifestyle. As the symptoms of insulin resistance develop slowly, the period of transition to full DM2 is termed as prediabetes. Due to the lifestyle origin of DM2, it is usually seen as more preventable compared to DM1. However, it should be noted that DM2 has a very strong co-occurrence between identical twins even if the disease sets in at different times, suggesting a significant genetic background [Medici et al., 1999]. Worldwide the incidence of DM2 is around 350 million, a number that is quickly rising due to mentioned lifestyle changes.

2.1.3 Gestational diabetes

Gestational diabetes is the final major diabetes group. Just as in DM2 the cause of gestational diabetes is insulin resistance, however, with gestational diabetes this is temporary and associated with pregnancy. Although the pathogenesis of gestational diabetes is unknown, it is thought to originate from already fragile beta-cells in combination with the changes in hormone signaling during pregnancy [Kaaja and Ronnemaa, 2008]. The incidence of gestational diabetes is around 5% per pregnancy, however very few women retain the diabetic phenotype after the gestational period is over, although some develop DM2 afterwards [Kampmann et al., 2015]. Notably, children from mothers with gestational diabetes are more likely to be obese and to develop DM2 at an early age [Group, 2002].

2.1.4 Other diabetes types

The fourth category of diabetes is composed of several different origins grouped together. These are diabetes types that occur in low frequency and thus are not in the focus of most research. Maturity-onset diabetes of the youth (MODY) is a group of monogenic diabetic diseases that due to an autosomal dominant mutation hinder insulin production. Other causes can be related to another disease such as pancreatitis or cystic fibrosis, or when diabetes is induced chemically or by drugs, for example by glucocorticoids.

2.1.5 Diabetes as a spectrum

While the disease diabetes is normally divided into these groups, in reality the disease is more of a spectrum, where DM1 partients sometimes also have lower levels of insulin resistance and DM2 patients usually have some decrease in beta cell mass [Brooks-Worrell and Palmer, 2011]. Thus, while the diseases have differences in their origins and exact effect upon the sufferer, research findings from one subtype can still have an impact on the handling of the other. Even the origin of the diseases might be more closely linked than previously thought; a recent hypothesis is that beta cell fragility, i.e. the susceptibility of beta cells to apoptosis, might be a common factor that contributes to both diseases [Dooley et al., 2016].

2.2 The long term complications of diabetes

Prolonged hyperglycemia and reduced insulin signaling affect the body in many different ways. Two of the most common origins of complications are angiopathy and defective wound healing. The glucose intake of the endothelial cells that line vasculature are independent of insulin, which results in an increased sugar uptake proportional to the blood glucose content into the endothelia during hyperglycemic conditions [Vithian and Hurel, 2010]. The high level of intracellular glucose also leads to higher levels of reactive oxygen species (ROS). ROS originate from the energy metabolism of the mitochondria and are one of the main perpetrators of oxidative stress in the cell [Brownlee, 2005]. Over time this will lead to an increase in advanced glycation end products (AGEs) on the plasma membrane of these cells, leading to a thicker basal membrane and thus vascular hypertension [Charonis et al., 1990]. Additionally, AGEs can interact with the receptor for advanced glycation end products (RAGE), which can further stimulate an increased ROS production, along with an increase in anti-inflammatory response [Kay et al., 2016]. The production of ROS can be altered by controlling the mitochondrial activity, for example by mitochondrial uncoupling [Mailloux and Harper, 2011a]. Mitochondrial uncoupling is when electrons are not used for ATP generation because of some mechanism, such as an open proton transporter through the inner membrane.

The effects of capillary damage are different depending on the location [Fowler, 2008]. Diabetic neuropathy is a progressive damage to neurons due to worsening blood flow [Malik, 2014]. Diabetic retinopathy has a similar pathogenesis with the retina degenerating and malformed due to a lack of blood access [Ahsan, 2015]. Diabetic retinopathy is one

of the leading causes of blindness and affects a majority of diabetes sufferers. Diabetic nephropathy stems from the small blood vessels of the glomerulus, the blood filtration unit, which are damaged during diabetes [Cao and Cooper, 2011]. Diabetic foot is a consequence of neuropathy and decreased wound healing where a small wound can lead to the amputation of a whole leg due to bad circulation and healing [Dinh and Veves, 2005].

2.3 Insulin expression

The hormone insulin is produced from the *INS*-gene, which is expressed only in the beta cells of the pancreas. The beta cells are organized together with glucagon producing alpha cells to form the Islets of Langerhans, cell clusters that are spread throughout the pancreas in islands, which also is the origin of the name insulin. In addition, the Islets of Langerhans include several other minor endocrine cells such as the delta cells which produce the endocrine regulating hormone somatostatin, epsilon cells which produce the hunger hormone ghrelin, and PP cells which produce pancreatic polypeptides that regulate the other cells of the pancreas. In total, the beta cells make up only around 1% of the pancreatic cells [McEvoy, 1981]. As the beta cells' main function is the production of insulin, the biology of the cells is centered around this task as well. When glucose reaches the cells after food intake, it activates several transcription factors that activate the beta cells and induces insulin production [Andrali et al., 2008]. Since the functional insulin hormone is a processed polypeptide, several steps of posttranslational modifications are carried out. As the translation of the polypeptide is carried out, the signal peptide will anchor the ribosome to the ER and the growing peptide, the pre-proinsulin, will complete its formation inside the organelle [Kim et al., 2012]. Once translation has completed, proinsulin is formed by the removal of the signaling peptide. Proinsulin is composed of three peptide chains; the A-chain, the C-peptide, and the B-chain. In the final insulin product the C-peptide is removed and two disulfide bonds are formed between the A-chain and the B-chain cysteines, while a third disulfide bond is formed between cysteines in the A-chain. The processing steps to this point also involve the packaging of the proinsulin into secretory vesicles by the Golgi apparatus [Huang and Arvan, 1994, Weiss, 2009]. Conveniently the C-peptide is released together with the finished insulin product, allowing for the quantification of endogenous insulin when insulin is administered as a treatment [Van Cauter et al., 1992]; although it should be noted that C-peptide has some functions of its own [Marques et al., 2004]. It also happens that proinsulin is released before it is processed, and as proinsulin

is more stable than insulin a moderate amount can be found in the blood. In most cases of diabetes the expression of insulin is altered. DM2 usually has a minor to an intermediate reduction in beta-cell function, while in DM1 the beta cell mass is almost completely lost.

2.4 Mutant Insulin induced Diabetes of the Youth

Within the MODY denomination of diabetes an insulin specific mutant group named Mutant INS-gene-induced Diabetes of the Youth, aka MIDY exists [Liu et al., 2010]. The individuals who have this type of mutation are typically diabetic from birth and require insulin treatment their entire life. The molecular origin of the MIDY condition depends on which specific amino acid is mutated, however most mutations lead to errors in the folding of the proinsulin peptide. Most commonly a cysteine residue is mutated resulting in an altered disulfide bond configuration and thus a misfolded peptide. The misfolded proinsulin peptides form complexes which impair the normal proinsulin trafficking to the Golgi apparatus. Eventually, the proinsulin buildup in the ER will overwhelm the cellular machinery and induce apoptosis of the beta cells, something already described in other forms of diabetes [Eizirik et al., 2007]. As the beta cells die off, fewer insulin producers remain which increases the demand on the remaining cells and thus induce a positive reinforcement loop where all beta cells eventually die off. The MIDY subsection of diabetes thus mimics DM1 without including the autoimmune component. As insulin secretion occurs already in embryos [Ashworth et al., 1973], sufferers will exhibit a neonatal diabetes which requires treatment to survive.

Several animal models of MIDY exist, the most notable of them being the Akita mouse [Yoshioka et al., 1997]. The Akita mouse has a C96Y (cysteine mutated to tyrosine at position 96) mutation in the *Ins2* gene [Wang et al., 1999], one of the two insulin genes in mouse, which causes a phenotype of total beta-cell loss. This reflects a rare human mutation requiring life-long diabetes care [Ahamed et al., 2008]. The Akita mouse is used to study complications such as renal dysfunction [Kitada et al., 2016] and retinopathy [Han et al., 2013].

2.5 The liver

As the main organ regulating metabolism, the liver is closely connected to diabetes. The liver has muliple functions, such as handling toxic metabolites [Jaeschke et al., 2012],

producing bile acids [Russell, 2003], and maintaining homeostatic levels of metabolites [Klover and Mooney, 2004, Cynober, 2002, Raghow et al., 2008]. New nutrients enter the bloodstream by the intestines and are then directly transported to the liver by the hepatic portal vein to be processed. The liver is supplied with blood from two sources; the aforementioned hepatic portal vein which brings nutrient-filled but deoxygenated blood, and the hepatic artery which brings the oxygen required for the organ to function [Waseem and Chen, 2016]. Importantly, the portal vein also brings insulin directly from the pancreas to the liver, allowing for a rapid response to changes in blood glucose. When insulin stimulates the liver it is taken up and partially degraded by the hepatocytes, which means that the rest of the body experiences a lower insulin dose compared to the liver [Geho, 2014]. Around 12000 genes are transcribed in the liver, which is around 60% of all active genes in humans [Uhlén et al.].

2.5.1 Structure of the liver

The human liver is made up of four lobes, widely different in size but similar in function [Bismuth, 1982]. The porcine liver has three lobes structured differently, however these are quite similar to humans as the sub-segmentation matches between the species [Court et al., 2003]. The organ as a whole is relatively homogeneous without great differences in function for the different lobes. On a microscopic level however, the liver is highly structured and heterogeneous. Hepatic lobules are structural units of the liver that are hexagonal and works as the building blocks of the liver. They contain, in the inward direction, the portal triad, the sinusoids, and the central vein. The portal triad consists of the portal venule and hepatic arteriole bringing blood to the tissue together, as well as the bile duct leading bile away from the tissue. The blood is led from the portal triad towards the central vein passing through the hepatic sinusoids, with hepatocytes closer to the portal vein having functions and oxygen levels different to the cells close to the central vein [Matsumura et al., 1986].

2.5.2 The sinusoid

The sinusoid is the functional unit of the liver. The sinusoid is surrounded by epithelial cells which are highly fenestrated (that is to say contain many openings) and are in turn surrounded by hepatocytes [Braet and Wisse, 2002]. The space between the epithelial cells and the hepatocytes are called the perisinusoidal space, or the space of Disse, and contain

blood plasma filtered by the epithelial cells. Additionally, the perisinusoidal space contains the Hepatic Stellate Cells (HSC) which are the pericytes of the liver, comprising around 15% of the livers cells [Geerts, 2001]. As the name suggests these cells are star-shaped and surround the capillaries. During normal hepatic conditions, the HSCs are used as storage for retinol (vitamin A) containing 80% of the body's total retinoid content [Blomhoff and Blomhoff, 2006]. During times of stress however, the HSCs become activated and drastically change their function into fibroblasts, disposing collagen into the perisinusoidal space [Mormone et al., 2011]. Overproduction of this collagen is the pathogenesis of cirrhosis, the final stage of fatty-liver. The final important cell type of the sinusoid is the Kupffer cell, the liver's resident macrophage. The Kupffer cells are scavengers which by endocytosis take up and break down bacteria and red blood cells. As the liver is located directly after the intestines in the blood flow and thus receives the brunt of endotoxic and xenobiotic material [Fox et al., 1987], the immune system of the liver is more immunotolerant compared to other organs [Thomson and Knolle, 2010]. The Kupffer cells reflect this immunotolerance by being able to suppress t-cell function in several ways [Ju et al., 2003], one being the suppression of regulatory t-cell activity in the liver [Breous et al., 2009].

2.6 Metabolism of the liver

Metabolically the liver is responsible for maintaining the glucose homeostasis along with taking part of regulating the concentrations of other metabolites such as amino acids and lipids [Klover and Mooney, 2004, Cynober, 2002, Raghow et al., 2008]. Hepatocytes have two metabolic modes; when glucose is available in high levels glycolysis is active, and when glucose is low gluconeogenesis is active. Glycolysis is the conversion of glucose to ATP, a process that is active in most cells. This conversion happens through several metabolic steps that end with the citric acid cycle (TCA cycle). When glycolysis is active the liver focuses on maintaining the levels of other metabolites, mainly lipids in the form of triacylglycerol (TAG). The liver then exports TAG to the bloodstream where it can be taken up and stored by adipose tissue. During gluconeogenesis however, the liver reverses the glycolysis reaction and starts producing glucose from ATP and TCA cycle components in order to export glucose into the blood and maintain blood sugar levels. This is a reaction that mainly occurs in hepatocytes [Widmaier et al., 2008]. The gluconeogenesis pathway uses the same metabolic intermediates as glycolysis with three enzymes, glucokinase, phosphofructokinase, and pyruvate kinase being replaced by gluconeogenic enzymes glucose 6

phosphatase (G6Pase), fructose 1,6-bisphosphatase (Fbpase), pyruvate carboxylase (PC), and phosphoenolpyruvate carboxykinase (PEPCK) [Oh et al., 2013]. To feed gluconeogenesis, hepatocytes start to break down amino acids and fatty acids so that they can be converted into TCA components. During the breakdown of fatty acids, the amount of acetyl-CoA is disproportionately high compared to oxaloacetate, and can thus not enter the TCA cycle properly. When this happens acetyl-CoA can be funneled into producing ketone bodies through the process known as ketogenesis [Newman and Verdin, 2014]. Ketone bodies are an alternative energy source to glucose, and more specifically the ketone body beta-hydroxybutyrate is exported to the blood from the liver under low glucose conditions. Ketone bodies are however acidic meaning that there is a limit on how many of them can be released before blood acidity becomes damaging.

2.7 Diabetes and the liver

Diabetes is known to affect the liver in multiple ways. As the liver is the first organ to receive secreted insulin through the portal vein it also receives the highest amount of the hormone. Since the liver clears most of the insulin after endocytosis, it is exposed to a far higher level of insulin compared to other tissues [Wang et al., 2014, Bojsen-Møller et al., 2018, Geho, 2014]. Contrary to adipose and muscle tissues which use GLUT4 as a glucose transporter, the liver uses GLUT2 which is not affected by insulin [Dimitriadis et al., 2011]. Instead, insulin affects the rate of glycolysis which in turn lowers the intracellular glucose levels facilitating glucose intake. In a diabetic organism with high blood sugar and no insulin, constitutive gluconeogenesis and glucose export from the liver will lead to even higher blood glucose levels. Since ketogenesis is ongoing, ketone bodies will make the blood progressively more acidic, and in DM1, where insulin is totally absent, this can lead to the condition of diabetic ketoacidosis where the blood becomes so acidic that the sufferer is in mortal danger [Chua et al., 2011]. It has been shown that during treatment of diabetes most subcutaneously administered insulin is absorbed by periphery organs with barely any hormone reaching the liver [Geho et al., 2009].

Associated with DM2 is NAFLD, aka non-alcoholic fatty liver disease. Although diabetic NAFLD is not entirely understood, it is thought to be related to insulin resistance [Tilg et al., 2017]. Increased fat deposition in the liver eventually causes storage of fatty acids to reach damaging levels as the tissue is inflamed and scarring occurs. Even in DM1 this can be an issue, although to a lower degree. DM1 fatty liver is not well studied,

although there are hypotheses for its origin including dysregulated lipoprotein ratios and abnormal lipid biogenesis [Regnell and Lernmark, 2011]. There are some evidence that NAFLD and insulin resistance leads to a higher level of endotoxins in sufferers [Harte et al., 2010].

2.8 Insulin signaling in the liver

Since much of research on the diabetic liver is focused on NAFLD and insulin resistance, not as much is known about the effects of pure insulin deficiency on the organ. Nevertheless, research has been done on the insulin signaling within the organ. Insulin is bound to the insulin receptor (INSR) which starts a chain reaction of signaling protein activations [White, 1998]. Insulin receptor substrate 2 (IRS2) and to a lesser degree also insulin receptor substrate 1 (IRS1) are activated by the insulin receptor [Rother et al., 1998, Kido et al., 2000], which in turn activate PI3K, a regulatory enzyme involved in many processes. PI3K activates AKT, which in turn inhibits gluconeogenesis by inhibition of PEPCK and G6Pase, and enhances glycogen synthesis. Lipid catabolism is also stimulated by insulin through sterol regulatory element-binding proteins (SREBPs) transcribed by the SREBF1 and SREBF2 genes. SREBF1 is responsible for lipogenesis and low-density lipoprotein (LDL) production while SREBF2 activation leads to transcription of enzymes involved in sterol synthesis [Horton et al., 2002].

2.9 Adipose tissue and diabetes

Contrary to insulin signaling in the liver, insulin directly influences the uptake of glucose into adipose tissue by GLUT4 [Kandror and Pilch, 1996]. In the absence of insulin GLUT4 is stored in intracellular vesicles which are translocated to the cell membrane upon insulin stimulation and activation of PI3K and AKT. The main goal of insulin in adipose tissue is to suppress lipolysis, the enzymes responsible for freeing fatty acids from the adipocytes. With low levels of insulin more than normal lipase activity is happening, leading to an increase in FFA.

2.10 Gluconeogenesis and amino acids

Amino acids can be categorized in several different ways. Glycogenic amino acids are the amines which when metabolized for energy are converted into a Krebs cycle intermediate and can thus be utilized in the production of glucose through gluconeogenesis. Ketogenic amino acids are in contrast converted into acetyl-CoA and can thus not be utilized for glucose production under gluconeogenic conditions. Out of the proteogenic amino acids, only lysine and leucine are fully ketogenic, while tryptophan, tyrosine, threonine, isoleucine, and phenylalanine are both ketogenic and glycogenic.

2.10.1 Amino acids relevant to diabetes

In addition to the glucogenic/ketogenic categorization, amino acids have other metabolic functions relevant to the diabetic condition. When amino acids are broken down they leave behind nitrogen in the form of ammonia, which is toxic. To get rid of the ammonia the liver uses the urea cycle to convert it into urea which is soluble and can be excreted in the urine. Glutamate functions as the carrier of ammonium and enters the urea cycle together with aspartate, eventually generating urea along with metabolites alpha-ketoglutarate and oxaloacetate. Arginine is a part of the urea cycle together with non-proteinogenic amino acids ornithine and citrulline [Morris Jr, 2002]. Branched-chain amino acids (BCAAs) are mainly metabolized in the skeletal muscle, contrary to other amino acids which are metabolized in the liver [Vary and Lynch, 2007]. BCAAs have an anabolic effect and are linked to the development of DM2 by their activation of mTORC1 [Yoon, 2016].

The amino acids serine, glycine, and methionine are part of the folate metabolism, an anabolic network that produces both building blocks for cell growth in the form of one-carbon units, as well as methylations [Locasale, 2013]. The system relies on tetrahydrofolate (THF) to manage one-carbon units. Cleaving glycine through the glycine cleavage system produces methyltetrahydrofolate (MTHF) from THF. Serine in reaction with THF creates glycine as well as MTHF. Methionine is a source of methylations in the form of S-adenosylmethionine (SAM), and methionine can later be renewed from homocysteine by using MTHF.

Some amino acids also have anti-oxidative properties, such as cysteine and glycine. Cysteine has a thiol group which gives it strong antioxidant ability, however due to its instability, it is mainly used in the cell in the form of glutathione, a tripeptide consisting of cysteine, glycine and glutamate [Wu et al., 2004]. Cysteine can also be converted into

taurine, a compound which is part of bile and also has antioxidant properties. As the final stage of methionine metabolism is cysteine, methionine can also be considered a source of anti-oxidants.

2.11 Animal models of diabetes

Traditionally the animal used to study most human diseases has been the mouse. Mice are quick to breed and the costs of maintenance are relatively low compared to other mammals. Additionally, as the mouse is the most studied mammalian animal model, the accumulated knowledge about its biology and genetic techniques makes it a convenient choice. However, since diabetes is a disease of metabolic origin the organism in where it should be studied has to be relatively similar to humans as metabolism varies greatly between organisms of different sizes and with differing diets. Furthermore, to study the effect of diabetes on different organs the anatomy of the model organism also needs to be similar to humans for the study to have translational value. Thus while mice are suited to mechanistic studies, once the disease is studied as a whole, a more suitable animal model is needed. To induce diabetes in rodents, chemical, surgical, viral, dietary, and genetic methods have been used [King, 2012].

2.11.1 Diabetic pigs

An interesting animal of choice for diabetes studies is the pig (reviewed in [Wolf et al., 2014]). The metabolism and physiology of pigs are remarkably similar to that of humans due to both organisms being mammal omnivores of similar size. Furthermore, pigs have short generation time for a large animal and a large litter size, allowing for efficient breeding [Aigner et al., 2010]. There are several ways of inducing diabetes in pigs: by surgically removing parts of the pancreas [Morel et al., 1991] or by injecting the pig with Alloxan [Tyrberg et al., 2001] or STZ [Dufrane et al., 2006] a phenotype similar to that of DM1 can be established. However, these techniques are invasive which leads to questions about the inherent translational value of the disease model. Several different models of DM2 also exist [Bellinger et al., 2006].

2.12 Biobanks

An inherent issue in working with large animal models is the maintenance cost and time involved in breeding and maintaining the animals until adulthood. Conversely, an inherent advantage of such a model organism is the large amount of available tissue from a single animal. This makes large animal models perfect candidates for biobanking, an emerging practice of storing tissue and body fluids for use at a later point [Agca, 2012]. For large animal models of diseases such as the MIDY pig this allows for the storage of all tissues and subsequently a truly full-scale characterization of the diseased animal. Except for the scientific value biobanks of large animal models provide, they also have an ethical value as they reduce the number of needed test animals.

2.13 Munich MIDY Pig Biobank

With the goal of studying the chronic effects of diabetes and hyperglycemia on the body of a clinically relevant organism, the Munich MIDY pig biobank was established [Blutke et al., 2017]. Under controlled conditions, 10 sows, 5 wild-type and 5 diabetic, were housed for 2 years together. The diabetic animals were all from the MIDY subtype and in order to minimize the outside effects all pigs were littermates. During this time the animals were regularly measured for their insulin and glucose content. The diabetic animals were treated with both short and long-acting insulin in order to keep them alive as the diabetes would have caused an early death otherwise. The animals were only treated during the day and were affected by hyperglycemia during the night in order to mimic the conditions of a poorly compliant diabetes sufferer. At several points in the pigs' lifetime clinical blood measurements were taken. At the age of 2 years the pigs were euthanized. Following the moment of death, an organized dissection of the animal's tissues occurred whereupon the sampled tissues were stored in a biobank. The procedure for each organ is described in Albl et al. [2016a], and the procedure was highlighted in Abbott [2015]. In total the biobank has around 19000 samples from over 50 different tissues and body fluids, prepared and stored in a way suitable for a future analysis type such as omics analysis, histology, etc. (Figure 2.1).

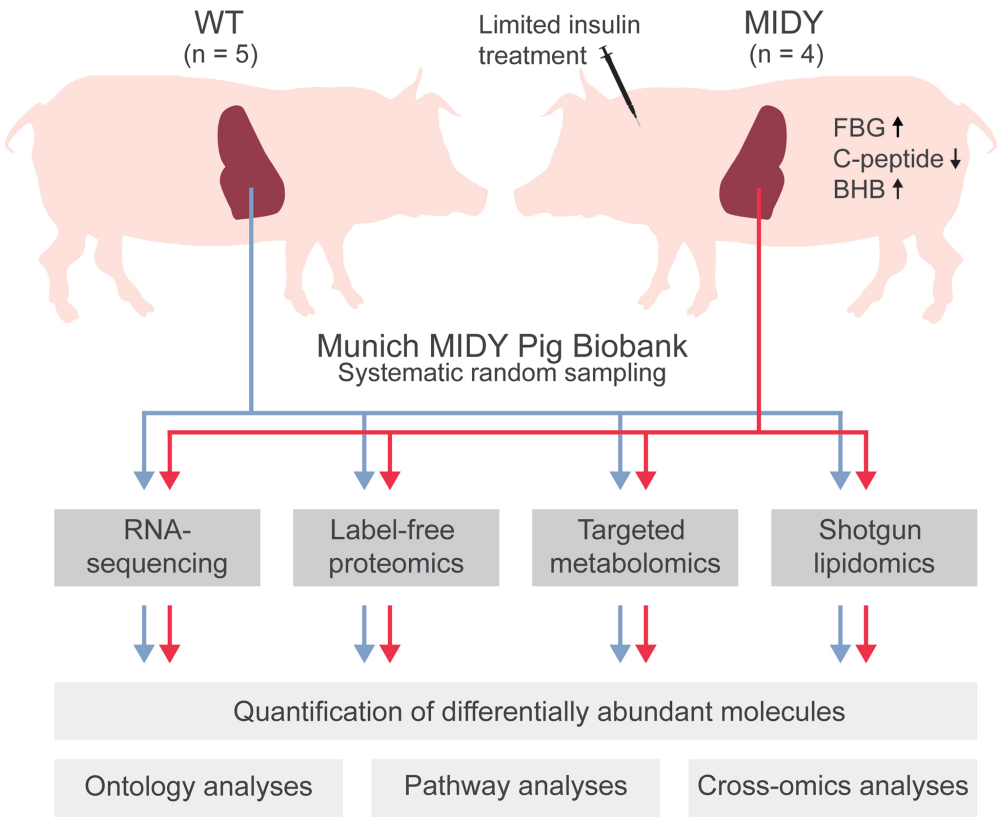

Figure 2.1: Figure summarizing the Munich MIDY biobank. MIDY diabetic pigs are grown along with WT litter mates under treatment until the age of two, when they are euthanized and dissected, and then stored in a biobank. Heterogeneous tissues such as the liver is randomly sampled. Tissues are prepared in such a way to make it convenient for several different analysis including omics analyses and subsequently gene ontology and pathway analyses. Adapted from Backman et al. [2019].

2.14 Molecular characterization of tissue

Omics is the word for high dimensional biological data obtained by experiments intended to capture every type of a biological molecule or biological information. As data storage space and processing speed have rapidly increased in recent years along with decreased costs of data acquisition, omics data is more and more becoming a central part of life sciences. The most common types of omics studies are genomics which is concerned about finding the exact genotype of an organism by DNA sequencing [Koboldt et al., 2013], transcriptomics which studies the changes in mRNA concentrations between organisms by sequencing of cDNA libraries [Conesa et al., 2016], proteomics which establishes the state of an organism's protein content by mass spectrometry [Bantscheff et al., 2012], and metabolomics which

measures the concentration of different metabolites again by mass spectrometry [Guasch-Ferré et al., 2016]. Each omics data can either be looked at separately or together with other omics types if available. The simultaneous analysis of several omics types is called multi-omics and is one of the emerging fields in computational life science.

2.15 Transcriptomics

Contrary to the expression of proteins, which requires both transcription and translation before they can respond to change, the expression of mRNA only requires transcription and is thus very dynamic and rapidly changes whenever the cell is developing or adapting to stimuli. mRNA concentration is determined by two things, transcription speed, which varies but is on average around 3.1-3.8 kb/min, and degradation speed which varies but is usually in the order of hours [Maiuri et al., 2011, Horvathova et al., 2017]. Due to this, the study of transcriptome profiles by RNA-Seq experiments is popular when comparing the biology of perturbed organisms to that of control animals. Especially in exploratory studies this is useful as by design the entirety of the animals mRNA is captured. In the past, microarrays were the primary tool of expression analysis, however with the advances in sequencing technologies and the associated reduction in cost, sequencing of mRNAs have become the preferred technology [Bottomly et al., 2011]. Another goal of transcriptomics is finding alternate splicing events and alternatively spliced transcripts [Lee and Rio, 2015]. In recent years, the prospect of sequencing single cells has also become a reality.

2.15.1 Library preparation

The procedure of preparing RNA for sequencing is called library construction and is a vital part of the RNA-Seq workflow. The first step is the conversion of non-ribosomal RNA to double stranded DNA. This is followed by ligation of sequencing adapters and a PCR reaction to amplify to levels required for sequencing. RNA is extracted with Trizol or a similar reagent and purified after thorough homogenization of the tissues. It is important to prevent RNAse activity as it will quickly degrade RNA. After this step library preparation varies a lot depending on the technique. As described in the next section, rRNA removal is an important issue, as is the way of obtaining shorter fragments that are needed for Illumina sequencing. Tailored fragmentation of cDNA is common here using sonication or tagmentation, however other techniques like Lexogene sense circumvent that by the use of start/stop primers. Another approach is to not sequence the whole transcript and instead

only sequence one read per transcript at it´s 3´end. While this is simpler and allows for differential expression analysis, it is not possible to investigate splicing differences using these techniques as all measured transcripts will be biased towards the 3' end.

2.15.2 Ribosomal RNA

A major issue of mRNA analysis is the fact that 85% of the RNA in any organism is composed of ribosomal RNA (rRNA) [Scott et al., 2010]. There are several approaches to enriching mRNA from this total RNA pool. If this step is not done the required sequencing depth will be much higher and an RNA-Seq experiment will be far more expensive. One way to get rid of rRNA is to enrich only mRNAs with a polyA tail using polyT coated beads. Another technique is to bias the random hexamer reverse transcription primers against rRNA sequences and thus minimize rRNA cDNAs available in the library [Adomas et al., 2010, Bhargava et al., 2013]. Both these techniques will have some bias against different mRNA species, however the bias should be the same for genes in the same species even if the experimental condition is different, unless alternate splice isoforms are used. A downside of using the polyA approach is that non-coding RNAs will be excluded.

2.15.3 Sequencing of cDNA libraries

Once cDNA libraries have been established some sequencing technique has to be used for the measurement of transcripts. The most common sequencing technique for short length reads is the Illumina array based sequencing by synthesis. Fragments are attached to the so called flow cell, the glass slide with reaction chambers coated with primers fitting to adaptors used for library construction. Once attached by hybridization, the DNA strand is converted to a double strand. After denaturing of the double stranded fragment, the free end of the fragment hybridizes to a complementary oligonucleotide attached at the surface of the flow cell. The elongation of this oligo creates a double-stranded DNA bridge that is used to generate clusters of DNA by solid phase amplification. After several rounds of amplification, sequencing is performed using dyed nucleotides blocked at the 3´-end. The complementing strand is used as a template for the stepwise incorporation of the dyed nucleotides, followed by capturing the color of clusters by excitation. Serially taken pictures are used to infer the nucleotide sequence. The alternate to short read sequencing is long read sequencing, typically performed by PacBio, or recently the nanopore.

2.15.4 Data processing

Once cDNA is sequenced several steps of data processing commence. Each read contains additional information like a unique identifier and quality score for each nucleotide. The sequence data is stored in a FASTQ file, a modification of the standard FASTA sequence data file with added quality scores [Cock et al., 2009]. Usually most reads will be in the length of 100bp, the standard for short read sequencing, but occacionally it can be even lower at 50bp or less. A de-multiplexing step is usually performed if multiple samples are involved as each sample would be assigned to its own tag. This is so that all samples will be treated equally when sequenced. Following this step, a round of quality control reveals if there is any general sequencing error or contamination in the data. Several library preparation methods will require that adapters are trimmed before the read is ready for further analysis.

2.16 Bioinformatics analysis of transcriptomic data

Once data is processed each read is aligned to the reference genome of the sequenced organism, a process commonly referred to as mapping. As transcriptomics data is gapped due to introns, a reference file containing known gene locations usually has to be supplied, although some mappers can find these splice sites by themselves. For transcriptomics data several well-known read mappers exist. In the past the Tophat aligner [Kim et al., 2013], a gapped version of the Bowtie DNA mapper, was dominant as the quickest mapper, now several algorithms outcompete it. Most used is STAR, an extremely fast mapper that is specifically tailored for reporting splicing events [Dobin et al., 2013].

Once mapping is done the analysis path is dependent upon the type of experiment conducted. For differential expression analysis, the number of reads mapping to each gene needs to be calculated. Several methods exist for this purpose, one example being HTSeq counts [Anders et al., 2015]. The resulting count data is what is then used as an estimation of gene expression for each gene in each sample.

When determining whether or not a scientific hypothesis is significant or not some type of statistical test is employed. For biological data which is usually low in sample size, Student's t-test tends to be the method of choice. If multiple factors are involved Analysis of Variance (ANOVA) is the preferred method for finding which factor or factor combination induces the most variance to the data. For counts data such as the results from an RNA-Seq experiment however, a more sophisticated method is required. Tools such as

DESeq2 [Love et al., 2014] and EdgeR [Robinson et al., 2010] are popular; both providing methods for sophisticated dispersion shrinkage and uses Generalized Linear Models (GLM) [Nelder and Wedderburn, 1972].

2.16.1 Gene set identification

While the identification of individual genes is informative on the biological perturbations in a tissue, pathways and sets of genes are typically more relevant for carrying out different functions. Several databases exist which assign classifiers to individual genes and sort them into groups; the GO consortium [Consortium, 2019] divides genes into different gene ontologies in three domains: cellular component, molecular function, and biological process. Another database, Kyoto encyclopedia of genes and genomes aka KEGG [Kanehisa and Goto, 2000], has a highly curated collection of biological pathways where genes are sorted depending on their function. Several other databases exist, such as the reactome [Joshi-Tope et al., 2005] and biocarta [Nishimura, 2001, Schaefer et al., 2008]. All the above pathway databases are compiled in the MSigDB, a large collection of gene sets [Liberzon et al., 2011].

When analyzing over-represented pathways within high dimensional data, one important question is how to select what genes to include in the enrichment analysis. Two main schools of thought exist here: to take only genes above a certain threshold e.g. only statistically significant genes or those with a large enough effect size (or both), or to take every gene and include a ranking criteria to indicate the importance of individual genes [Tarca et al., 2013]. In the first approach you risk losing information from small changes in gene expression, while in the second you increase the influence of noise on your analysis. Fischer's exact test is the typical statistical test for the first approach where one treats the significant genes as a set of their own.

Gene set enrichment analysis (GSEA) is the most well-known rank-based method for finding overrepresented pathways [Subramanian et al., 2005]. GSEA depends on a Kolmogorov-Smirnoff-like running score which reads the barcode of genes belonging to the investigated pathway in the ranked list. The list will be ranked depending on certain criteria such as the signal to noise ratio in the data. A high enrichment score will appear when most genes show up at either end of the list and thus create a large peak or trough. The enrichment score, ES, will be normalized into NES depending on the gene set size and then used to calculate significance. Significance is established by permuting the list and recalculating the score a large number of times. The permutation can be done either by

switching the labels of genes or the labels of phenotypes. Alternatively one can supply an already ranked list to GSEA which will start from the scoring step and exclusively uses gene label permutation.

2.17 Proteomics

The proteome consists of all proteins expressed in a measured biological source. The analysis of differential expression of proteins is similar to that of differential expression of mRNA in that in both cases what is measured is gene expression, though proteins are one regulatory level after mRNA. However, the analysis of proteins differs from the analysis of mRNA and DNA as there is no convenient way to do sequencing by synthesis for peptides. Instead, peptides are cut into smaller segments which are analyzed individually by a mass spectrometer and then computationally reassembled. By the nature of this approach, the analysis is less accurate than the sequencing-based approach, and many of the proteins from genes with a lower expression level are unquantifiable. Another issue that faces proteomics is alternate isoforms, leading to a proteome that contains up to 100,000 proteins [Horgan and Kenny, 2011].

2.18 Metabolomics

A metabolite is a small molecule that is an intermediate product in metabolism. Metabolomics, the study of chemical metabolites, is measured similarly as in proteomics using a mass spectrometer [Dettmer et al., 2007] or alternatively using nuclear magnetic resonance, NMR [Nicholson and Wilson, 2003]. In contrast to transcriptomics and proteomics, there is no sequence information in metabolomics data. Instead, the challenge of metabolomics is the large amount of different metabolic molecules and their derivations that are available to be measured in the cell. For example, in plants the number of metabolites is in the order of 200,000 [Weckwerth, 2003]. Therefore, some techniques, called targeted metabolomics, focus on only identifying the amounts of metabolites important to the scientific question at hand, such as amino acids, sugars, and lipids in the case of diabetes.

2.19 Lipidomics

Lipidomics is a subsection of metabolomics where instead of measuring the whole metabolome the focus is on lipids. The challenge is also similar to metabolomics since lipids are possibly the metabolites with most structural variation [Wang et al., 2016]. As many lipids contain one or more long carbon tails that can vary in bond configuration in combination with one of several lipid heads, the possible configurations are far more than can be measured. For questions such as the influence of metabolic disease, lipidomics can be more precise compared to general targeted metabolomics approaches [Lydic and Goo, 2018].

2.20 Multi-omics

Integrating different omics techniques is one of the current hot topics of the bioinformatics world [Sun and Hu, 2016]. Due to the properties of the different omics data, there are great difficulties in making conclusive and coherent analyses. There are many different approaches to integrating multi-omics data, however, so far none has shown itself to be superior [Huang et al., 2017]. Network-based approaches are greatly limited by the current biological knowledge and by missing data. Multivariate techniques are powerful but can only be used for correlations and associations. Furthermore, all techniques are limited by small sample sizes, a common occurrence in biological studies.

2.21 Statistical treatment of multivariate data

Ordinary linear regression is a technique for modeling data where a response variable, i.e. gene expression in terms of counts or mass-spectrometry intensities, is dependent on one or more independent variables such as genotype, sex, drug treatment, etc [Harrell Jr, 2015]. Ordinary linear models effectively model the means of the response variable with the deviating spread of data points being referred to as residuals. Residuals should follow a normal distribution for the linear model to have been used accurately. One of the strengths of linear model regression is the ability to model out the influence of one independent variable that could somehow influence the desired comparison. GLMs are a way of modeling data using linear regression but allowing for non-normally distributed data to be analyzed [Nelder and Wedderburn, 1972]. In the case of modeling transcriptomics counts data the negative binomial distribution is the most accurate [Robinson and Smyth, 2007].

Removing batch effects by linear modeling is possible when using the batch as an extra factor. If the batch effect is hidden, methods such as Surrogate Variable Analysis (SVA) [Leek and Storey, 2007] can help with identification and removal. Methods for estimating differential expression such as DESeq2 [Love et al., 2014] are estimating a null-distribution from the observed data which is then used to calculate significance using Wald's test. In this case outlier samples or hidden batch effects can cause an under or over-estimation of the null-distribution variance which will lead to an abnormal number of false positives. To help remove this incorrect estimation the null-distribution can be explicitly modeled using the test score, here z-score, in some programs such as fdrtool [Strimmer, 2008].

2.22 Multiple testing correction

In many biological and medical studies there is a small set of hypotheses being considered, such as the measurement of a set of genes by qPCR. Normally the end product is the p-value, the well-known, controversial, and often misinterpreted probability value which indicates the probability of obtaining a more extreme result than obtained in the experiment under the null-hypothesis [Greenland et al., 2016]. A cutoff, usually at <0.05, is used to reject the null-hypothesis and to consider the result significant. The cutoff however is not an accurate indicator when many hypotheses are under consideration. When testing many hypotheses simultaneously a number of them will be true due to chance as p-values are uniformly distributed under the null-hypothesis [Shaffer, 1995]. To counteract this a correction to the significance cutoff is required in order to minimize the false positives. A simple solution is to divide the cutoff with the number of tested hypotheses, the so-called Bonferroni correction. However, with large number of tests this correction tends to be conservative [Narum, 2006]. More exactly, the Bonferroni correction controls the familywise error rate (FWER), which is the probability of producing one false positive.

For a study with a large number of hypotheses such as an exploratory omics study, it is preferred to find the false discovery rate, aka the FDR. The FDR reports the proportion of hypotheses that will be false positives and can be computed by the Benjamini-Hochberg procedure [Benjamini and Hochberg, 1995]. Alternatively one can empirically establish the FDR by permutating labels such as sample or gene labels [Goeman and Solari, 2014]. Several other methods exist for FDR calculation, some measure the local FDR which is distinct from the normal FDR as it is a Bayesian posterior probability [Efron, 2005].

Chapter 3

Materials and Methods

3.1 Munich MIDY Pig Biobank

A cohort of 10 sows consisting of 5 MIDY and 5 WT pigs, were brought up together until the age of 2 years. During this time they were fed once daily and had free access to water. The sows were treated with both long-acting and short-acting insulin. The treatment was only during daytime in order to simulate poorly controlled diabetes. At the age of 2, the pigs were estrous synchronized [Kurome et al., 2015] and then inseminated in order to exclude the effects of estrous cycles on the molecular profiles of tissues and body fluids, 12 days after this they were euthanized. Systemic and controlled necropsies were performed, dissecting complex tissues and performing "systemic random sampling" on homogeneous tissues. Body fluids were sampled as well including both blood plasma and serum. Collected samples were shock-frozen on dry ice and stored at -80 degrees within 20 minutes after death. All samples were stored in Munich MIDY Biobank [Blutke et al., 2017] until further analysis was carried through. For all subsequent steps, all samples were processed in parallel.

3.2 RNA Sequencing

Preparation and sequencing of MIDY RNA-Seq libraries were carried out by Julia Philippou-Massier, Stefan Krebs, and Helmut Blum. All samples for RNA-Sequencing originated from the Munich MIDY Pig Biobank. Four different library preparation methods were used in this thesis. For the liver, Nugen Encore Complete, Lexogen Sense, Lexogen QuantSeq and a bulk version of the single-cell SCRB-Seq was used, while for adipose tissue only Lexogen

Sense was used for all four tissues types. Samples were homogenized in Trizol and total RNA was isolated with chloroform following the manufacture's protocol. Isolated total RNA was quantified using s Nanodrop and quality controlled by a Bioanalyzer 2100. Good quality RNA (RIN >7.0) was used to construct sequencing libraries.

The NuGEN library was performed using the rRNA avoiding "not-so-random" priming of the first strand synthesis, followed by second strand synthesis after RNA degradation. The second strand contains nucleotide analogs used for strand selection. cDNAs are sonicated using Covaris. Adaptors are added, the forward adaptor contains the same nucleotide analogs as used for second strand synthesis. Base excision and heat is then used to remove miss-aligned adaptors to maintain strand specificity for the sense direction. This is followed by a PCR reaction for amplification.

The Lexogen Sense library preparation started with denatured RNA mixed with oligodT beads, capturing only polyadenylated RNAs. Other RNAs are washed away, purifying mRNA. Starter/stopper heterodimers are added to the mix which randomly hybridizes to the RNA strand. Heterodimers are used to perform reverse transcription, creating short cDNA reads. RNA is then hydrolyzed when the second strand is synthesized, creating double stranded cDNA. After magnetic beads are washed away, a PCR reaction is performed and adaptor sequences are added.

The QuantSeq library preparation started with oligodT priming, synthesizing the first strand. RNA is then removed, allowing for random priming of second strand synthesis. A PCR reaction is then performed for amplification.

SCRB Seq library preparation starts with the priming of a cDNA strand containing oligodT to the polyA tail. The second strand is then synthesized. Tagmentation is used for inserting Nextera primer that are used for PCA amplification.

All libraries were sequenced on a HiSeq 1500 (Illumina) as 100 bp single reads with the exception of SCRB-Seq, which use 50 bp reads. Due to poor mRNA quality, the adipose tissue had to be sequenced in two batches.

3.3 Analysis environments

The R statistical computational environment [Rco] along with the Rstudio IDE [Team et al., 2015] was utilized for data processing and statistical analyses. The Galaxy platform [Giardine et al., 2005, Afgan et al., 2018] was used for data pre-processing, processing, storage and organization of sequencing data.

3.4 RNA-Seq data pre-processing

For all data, raw FASTQ files were demultiplexed using 6-nucleotide long barcodes with an in-house script allowing for one miss-match. For Lexogen Sense data, a head-crop was performed in order to remove the 12 first bases using the Trimmomatic tool [Bolger et al., 2014]. Encore Complete RNA-Seq library system originating data required no head crop. The FastQC tool [Andrews, 2010] was used along with the MultiQC tool [Ewels et al., 2016] in order to visualize and quality control sequencing data. If Illumina adaptor content was high enough for a data set to give a warning, Trimmomatic was once again used for adaptor clipping.

3.5 RNA-Seq data mapping

Mapping of short reads against the susScr11.1.91 ensembl reference genome (FASTA genome and gtf annotation files) was performed using the STAR [Dobin et al., 2013] ultra-fast short read mapper. Default parameters were used with an sjdbOverhang of 100. Mapping efficiency was analyzed by MultiQC on the STAR output logs. For control of mapped transcripts the Integrative Genomics Viewer, IGV [Robinson et al., 2011, Thorvaldsdóttir et al., 2013], was used.

Counting of reads mapped to specific genes was performed by HTSeq [Anders et al., 2015]. The susScr11.1.91 gff3 file was used as a reference, 'exon' was used as feature and 'gene' was used as feature ID. Minimum alignment quality was set to 10 and 'intersection-strict' was used as counting mode. For Lexogen Sense libraries, the reverse strand interpretation was used.

For the SCRB-Seq library preparation method a paired-end single-cell method used here on bulk tissue. The Je suite [Girardot et al., 2016] was used for both UMI clipping and PCR duplicate removal with 1 miss-match allowed.

3.6 Annotation

The R package biomaRt [Durinck et al., 2005, Smedley et al., 2009] was used for conversion of ENSEMBL identifiers into gene symbols. Matching was done with the "sscrofa_gene_ensembl" database against "hgnc_symbol", "external_gene_name", and "wikigene_name" in that order. For converting gene symbols to full gene names the HGNC

database tool multi-symbol checker [Yates et al., 2016] was used. When such matches could not be found, i.e. when “LOC”, “orf” or “SLA” genes were retrieved from biomaRt, a BLAST [Altschul et al., 1990] was attempted in order to find the closest matching ortholog gene.

3.7 RNA-Seq normalization

For comparison of gene expression levels between different genes, TPM (Transcripts Per Million) were used [Li et al., 2010]. TPM normalizes for sample depth and transcript length. TPM is calculated as follows:

1. Divide each transcript's read counts by its length in bases.
2. Divide the length normalized counts by the sum of all length normalized counts in the sample.
3. Multiply by 10^6.

$$TPM_i = \frac{X_i}{\tilde{l}_i} * \left(\frac{1}{\sum_j \frac{X_j}{\tilde{l}_j}} \right) * 10^6 \tag{3.1}$$

Where l is the gene length and X is raw-counts data.

For gene expression estimation within a condition the DESeq2 [Love et al., 2014] normalized counts data was used. The DESeq2 normalization, commonly referred to as "median of ratios", are done by dividing all the counts of a sample by a "size factor". The size factor is calculated by taking the geometric mean of all sample counts within a gene and then calculating the ratio of each count to that geometric mean. For each sample, the median of all ratios within the sample is the size factor. The calculation looks as follows:

$$s_j = median \frac{X_{ij}}{\left(\prod_{j=1}^{m} X_{ij} \right)^{1/m}} \tag{3.2}$$

Where s_j is the size factor for a particular sample, X_{ij} is raw-counts data and m is sample size.

3.8 Differential expression analysis

Individual HTSeq output columns were labeled according to their pig of origin and combined into an R data frame. Genes with an average count of less than 10 were excluded from further analysis in order to remove noise. DESeq2 [Love et al., 2014] was used for the establishment of differential expression. Outlier replacement and independent filtering were used for all experiments. A DESeq2 design formula of ~Genotype was used in the case of the liver NuGEN analysis ("MIDY" vs "WT"). For analysis of different adipose tissues SVA [Leek and Storey, 2007] was used with 1-3 surrogate variables estimated from the counts data and as such the design formula was ~SV1 + ... + Genotype. For PCA's of sequencing data, the "rlog" function in the DESeq2 package was used. The significance cutoff was set at an FDR of 0.05.

The VennCounts function from the limma R package [Ritchie et al., 2015] was utilized together with the VennDiagram function from the same package for creating Venn diagrams. Different data were merged using the gene or gene set name as a common identifier, with NA's and non-common features being filled as 0. When the significance of genes or gene sets was included, the cutoff was at 0.05 for the adjusted p-values.

3.9 Gene set enrichment analysis

Gene names along with logarithmized and signed p-values from DESeq2, alternatively corrected log2 fold changes from DESeq2, were imported to the GSEA [Subramanian et al., 2005] program. Gene Ontology (GO) [Consortium, 2019], Kyoto Encyclopedia of Genes and Genomes (KEGG) [Kanehisa and Goto, 2000] and the Reactome [Joshi-Tope et al., 2005] was the primary gene set databases used. These datasets are part of MSigDB [Liberzon et al., 2011], however a pig specific KEGG gene set collection was also created by pulling data from the KEGG repository and filtering out unrelated disease gene sets. In this way larger gene sets made up of several smaller ones such as "Amino Acid Metabolism" were also constructed. Additionally, gene sets were constructed manually in order to check certain hypotheses, using literature as a guide. The pre-ranked analysis was used with 1000 permutations as standard. The maximum gene set size was put at 1000 and minimum at 15. Both weighted and unweighted analyses were performed. The significance cutoff was set at 0.05 FDR.

3.10 Gene network visualization

The network visualization program Cytoscape [Shannon et al., 2003] was used along with the plug-ins ClueGO [Bindea et al., 2009] and Cluepedia [Bindea et al., 2013]. GO and KEGG gene set collections for pig was utilized. Genes significant at an FDR cutoff of 0.05 were considered as significant and genes were seperated into up and down-regulated groups. The parameter network specificity was set to default in the case of KEGG and increased for analyzing GOs due to a large number of connections.

3.11 Proteomics

The proteome analysis was performed on the same liver localizations as RNA-Seq measurements by Dr. Florian Flenkenthaler and Dr. Thomas Fröhlich. Sample were homogenized as described [Fröhlich et al., 2016]. Protein concentration was determined as described in [Antharavally et al., 2009]. For each sample 100 µg of protein was digested with Lys-C and trypsin as described in [Blutke et al., 2017]. A Q Exactive HF-X mass spectrometer was used for nano-LC-MS/MS analysis. 2.5 µg of peptides were separated at 200 nL/min. Spectra were acquired at 120,000 resolution between 380 and 2000 m/z. MS/MS scans of the 24 most intense peaks were performed at a resolution of 15,000. MaxQuant [Cox and Mann, 2008] and the NCBI RefSeq Sus scrofa database was used for protein identification and label-free quantification. For each group (MIDY, WT) 3 valid identifications were required to keep the protein for further analysis. Perseus was used for normal distribution based imputation [Tyanova et al.]. Significance was calculated using the SAM statistic [Chu et al., 2001].

3.12 Metabolomics

In order to analyze metabolites in plasma and tissues, the team of Dr. Jerzy Adamski performed liquid chromatography-electrospray ionization-tandem mass spectrometry and flow injection analysis-electrospray ionization tandem mass spectrometry in a targeted metabolomics approach with the AbsoluteIDQTM p180 Kit (Biocrates Life Sciences AG). Liver tissue first had to be homogenized using an ethanol/phosphate buffer. In total 188 metabolites were quantified. The Hamilton Microlab STAR robot and an Ultravap nitrogen evaporator was used for sample handling. API 4000 triple quadrupole system

equpped with an 1200 Series HPLC and a HTC PAL auto sampler was used for mass spectrometry analysis. This was controlled by the software Analyst 1.6.2. Metabolites were quantified using the MultiQuant 3.0.1 software and the MetIDQ software package. Concentrations of all metabolites were calculated using internal standards. For a more detailed description of methods, refer to [Zukunft et al., 2018, 2013].

3.13 Metabolomics data analysis

For all metabolomics datasets, raw metabolite concentrations were imported into R for processing. Limit of Detection (LOD) for each metabolite was calculated by multiplying the median value of three ethanol/phosphate zero-samples by three. Metabolites missing more than half of their data points were excluded. Remaining missing data points were imputed using half the value of the lowest metabolite in said batch (multiplied by a randomized value between 0.75 and 1.25). Normalization was done by plate for each metabolite using the following formula:

$$Norm(m_x) = \frac{P_a(m_x)}{P(m_x)} * C(m_x) \tag{3.3}$$

Where:

$$P(m_x) = \frac{\sum C(m_x)}{N^{sample}} \tag{3.4}$$

And:

$$P_a(m_x) = \frac{\sum P(m_x)}{N^{plate}} \tag{3.5}$$

m_x is one metabolite needing to be normalized, the calculation should be carried out for each metabolite individually (m_1,m_2...m_n where n is the number of metabolites). $C(m_x)$ is the concentration of a metabolite in the plate, N^{sample} is the number of samples for the plate and N^{plate} is the number of plates. P is the mean of plate, while P_a is the mean of means of plates. The ratio $P_a(m_x)/P(m_x)$ signifies the plate factor which is used to normalize the concentrations.

When modeling the metabolites for t-tests, each data point was logarithmized by a

generalized logarithm:

$$log_2 * \left(\frac{x + \sqrt{x^2 + a^2}}{2} \right) \tag{3.6}$$

Where x is a data point and a is a constant with a value of 1.

Metabolites were then Pareto scaled:

$$\tilde{x}_{ij} = \frac{x_{ij} - \tilde{x}_i}{\sqrt{s_i}} \tag{3.7}$$

Where x_{ij} is a data point and s_i is the standard deviation of each metabolite.

Student's t-tests were performed assuming equal variance and multiple testing were calculated using Benjamini Hochberg (BH) [Benjamini and Hochberg, 1995] or the qvalue [Dabney et al., 2010] R package. An FDR of below 0.05 was considered significant. PCA was performed using the mixOmics [Rohart et al., 2017] R package with centered and scaled values. Barplots were constructed using an R script. Heatmaps were made using the pheatmap package.

3.14 Lipidomics

Lipidomics experiments were performed by the team of Dr. Ünal Coskun as described in [Sampaio et al., 2011]. Liver samples were homogenized on ice in an ammonium-bicarbonate buffer (150 mM ammonium bicarbonate, pH 7) with an ultra-turrax homogenizer. Lipids were extracted using a chloroform/methanol procedure [Ejsing et al., 2009], followed by extraction of the organic phase which was transferred to an infusion plate and then dried in a speed vacuum concentrator. Protein content was assessed using BCA Protein Assay Kit (Thermo Fisher). Equivalents of 20 µg of protein were taken for mass spectrometry analysis. 1st step dry extract was re-suspended in 7.5 mM ammonium acetate (Sigma) in chloroform/methanol/propanol (1:2:4, v:v:v) and 2nd step dry extract in 33 ethanol solution of methylamine in chloroform/methanol (0.003:5:1; v:v:v). The Hamilton Robotics Starlet robotic platform was used for sample handling. At a resoultion of $R_{m/z=200}$ = 280,000 for MS and $R_{m/z=200}$ = 17,500 for MS/MS experiments, samples were analyzed in both positive and negative ion modes. An inclusion list of corresponding MS mass ranges scanned in 1 Da increments were used for triggering MS/MS.

3.15 Lipidomics data analysis

Data were analyzed with a lipid identification software based on LipidXplorer [Herzog et al., 2011]. Only lipid identifications with a signal-to-noise ratio >5, and a signal intensity 5-fold higher than in corresponding blank samples were considered for further data analysis. Raw lipid expression files were imported into R. Lipid species with more than five missing values were excluded from the analysis. PCA was performed using the mixOmics [Rohart et al., 2017] R package with centered and scaled values. Lipid species were grouped by lipid family and normalized as a percentage of total lipid content. Student's t-test was applied in the same way as the metabolomics data.

3.16 Multiomics integration

For the integration of transcriptomics and proteomics data the 2D-enrichment method [Cox and Mann, 2012] of the Perseus software [Tyanova et al.] was used. Identifier matching was done using an R script. 0.05 was the significance cutoff for pathway enrichment.

3.17 Retinoid Measurements

Retinoid measurements was performed by Dr Micheal Rothe. Folch's protocol was used for lipid extraction from homogenized liver tissue [Folch et al., 1957]. HPLC-MS/MS was used with an Agilent 1290 HPLC coupled with a triple quadrupole mass spectrometer Agilent 6470 equipped with an electrospray jetstream ion source. Positive/negative switching mode was used to capture retinoic acid as its signal is more selective in negative mode. Retinoid contents were normalized to tissue protein content and then tested for significance using student's t-test. Significance was considered at p below 0.05.

3.18 Glutathione Measurements

Glutathione measurements were performed by Dr. Simone Renner. Both glutathione (GSH) and the fraction of its oxidized form (GSSG) were measured using a Glutathione Colorimetric Detection Kit (EIAGSHC, Invitrogen) according to the instructions of the manufacturer. Free glutathione concentration was calculated by subtracting GSSG from

GSH. All measurements were normalized to the protein content of the samples and then tested for significance using student's t-test. Significance was considered at p below 0.05.

3.19 Quantification of Kupffer cells

Kupffer cells were detected using immunohistochemistry on randomly sampled liver tissue samples, performed by PD Dr. Andreas Parzefall. Goat polyclonal anti-IBA1 antibody and biotinylated rabbit anti-goat Ig secondary antibody were used. The principle of Delesse was used to calculate the volume density of Kupffer cells. Differential points counting was used for calculating area densities [Albl et al., 2016b].

3.20 Western blotting

Insulin receptor signaling molecules and their phosphorylation levels were measured by Western blot [Streckel et al., 2015, Hinrichs et al., 2018], peformed by PD Dr. Maik Dahlhoff. SDS-PAGE was used to separate protein from liver tissue homogenized in Laemmli buffer. Proteins were transferred to PVDF membranes (Millipore) by electro-blotting. After processed as described above, bound antibodies were detected using the ECL Advance Western Blotting Detection Kit (GE Healthcare). The Imagequant software (GE Healthcare) was used to quantify band intensities. A square root transformation of the densiometric data approximated a normal distribution wich was assesed for differences using Student's t-test.

Chapter 4

Results

4.1 RNA Sequencing of MIDY liver by NuGEN Encore complete

As the first transcriptome sequencing was performed with the NuGEN Encore Complete library preparation kit, the mRNA enrichment procedure was based on rRNA depletion. rRNA levels in NuGEN samples were still fairly high at around 38.7% (**Table 4.2**). This is likely due to the non-random priming process not being intended for pigs as it was originally derived from humans and mice. Due to this, extra depth for the NuGEN experiment was generated and combined with previous reads, allowing for an average read depth of 31.9 million reads per sample.

4.1.1 Removal of outlier sample 742

One transgenic sow, pig 742, was discovered to have an unusual insulin expression, reaching several magnitudes higher than other pigs of both conditions (**Figure 4.1**). The pig was an outlier in several datasets, including the metabolomics (**Figure 4.2**) and proteomics data, although it generally clustered with the other transgenic sows. The glucagon expression of 742 was almost identical to that of WT pigs, while beta-hydroxybutyrate levels were only slightly increased. As glucose levels were still moderately increased, this indicated that the animal had mainly an increase in levels of misfolded non-functional insulin, possibly functional enough to have a minor effect on insulin signaling. As the MIDY pig model organism is based on total insulin deficiency, the animal was removed from further analysis. Even with the decrease in power from a 4vs5 design compared to a balanced 5vs5 design,

Sample	Sex	Pregnancy status	Genotype	Weight (kg)
S736	female	+	WT	244
S737	female	-	TG	215
S738	female	-	WT	224.5
S739	female	+	TG	185
S740	female	+	TG	184
S741	female	+	WT	237.5
S742	female	+	TG	210
S743	female	+	WT	247
S744	female	+	TG	217
S745	female	-	WT	238

Table 4.1: Table describing the MIDY samples of the biobank. Sample S742 was removed from the biobank due to aberrant behaviour.

Sample	mapped (M)	rRNA (M)	% rRNA	mapped to genes (M)
S736	44.51	17.22	38.70	27.28
S737	52.53	21.33	40.61	31.19
S738	60.19	25.04	41.59	35.16
S739	50.37	22.20	44.08	28.16
S740	54.57	18.95	34.72	35.62
S741	56.97	25.35	44.50	31.62
S743	47.35	18.09	38.19	29.27
S744	50.98	15.86	31.10	35.13
S745	50.38	17.07	33.88	33.31
Mean	51.98	20.12	38.71	31.86

Table 4.2: Sequencing statistics for the MIDY biobank livers using NuGEN Encore Complete. Reads are counted in million reads (M).

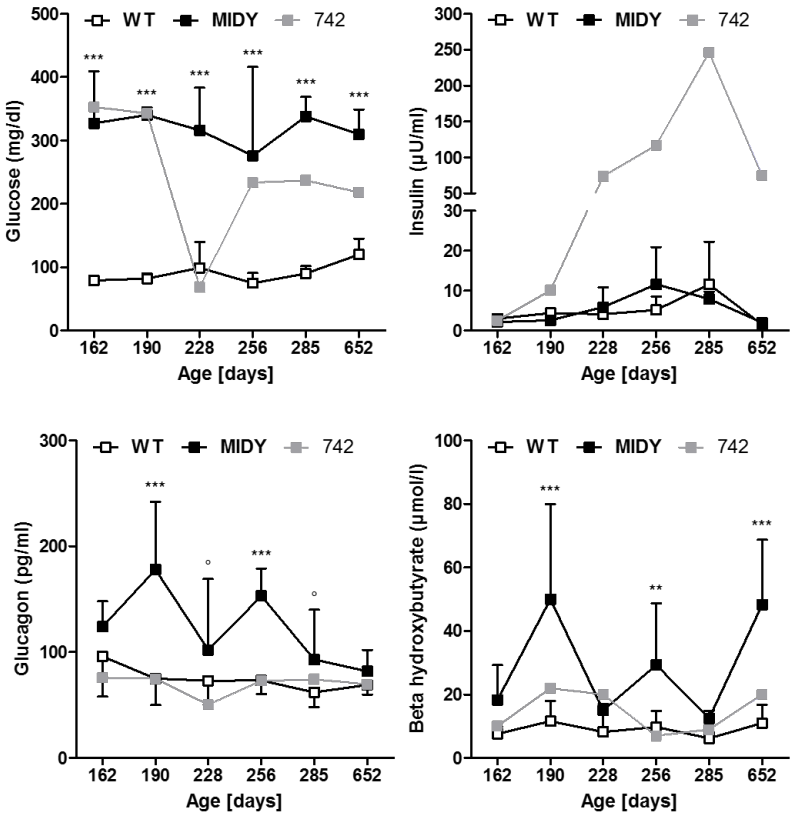

Figure 4.1: The concentrations of blood glucose, insulin, glucagon and beta-hydroxybutyrate during the lifetime of the pig. Means and standard deviations for each group is shown. MIDY animal 742 is displayed as grey and not grouped with other MIDY animals in order to show its aberrant phenotype. Significance is indicated between groups excluding 742 (*p <0.05; **p <0.01; ***p <0.001). Borderline significance (p <0.08) is indicated by °.

the number of significantly differentially expressed genes grew on the transcriptomics level, suggesting that pig 742 was more similar to the WT pigs. While the suggestion that the pig was suffering from insulinoma was made, no evidence of said growth could be detected. Another possibility could be alternate regulation of the mutant insulin cassette due to mutation or homologous recombination. The decision was made to not analyze sow 742 at all in other tissues.

4.1.2 Differential expression

In the MIDY vs WT analysis of hepatic tissue 533 significantly differentially expressed genes were detected, 320 upregulated and 213 downregulated (**Figure 4.3 A, B**), (**Table**

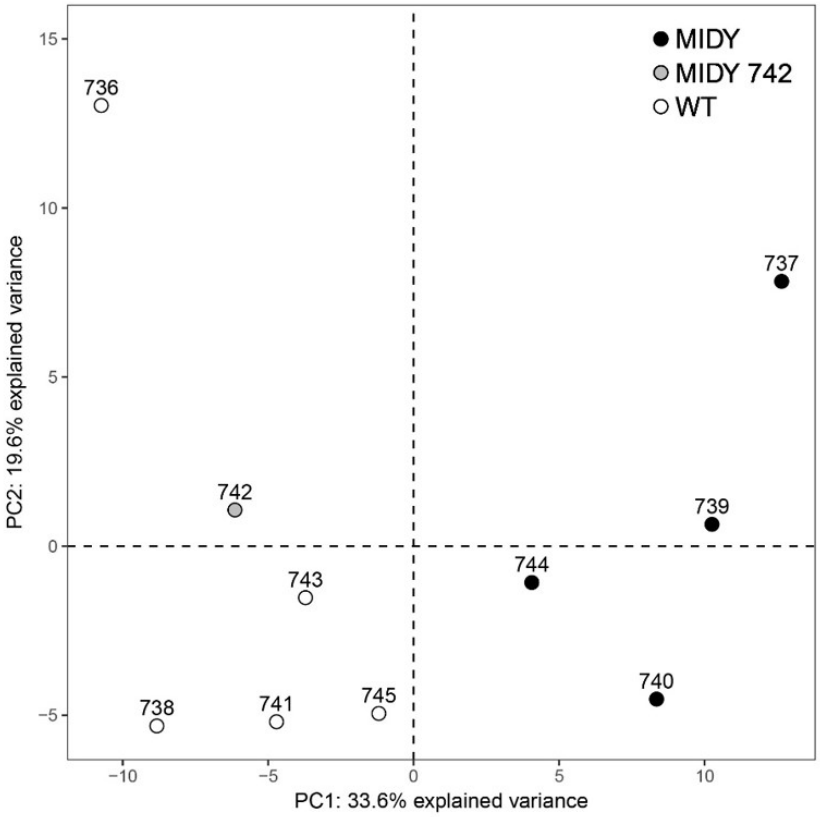

Figure 4.2: PCA of metabolomics data from the MIDY liver samples showing the seperation of MIDY and WT samples. Animal 742 is marked as grey, showing its clustering with WT samples.

4.3, Table 4.4). Most highly upregulated was ADAMTS17, which encodes ADAM with thrombospondin type 1 motif 17. ADAMTS17 has a relatively low expression in the WT pigs where the TPM is around 0.1, whereas the TPM in the MIDY pig is around 1.4. With DESeq2 corrected fold changes this gives a log2fold change of 2.7. The most highly downregulated gene was SHROOM3, with a log2fold change of -1.25. As reflected in the number of differentially expressed genes the log2fold changes of upregulated genes were generally higher than that of the downregulated ones (**Figure 4.4**). A list of all significant genes including their p-values and log2 fold changes discussed in this thesis can be found in (**Supplementary Table A.1**).

Gene symbol	Mean expression (counts)	l2fc (MIDY/WT)	p-value	FDR
ADAMTS17	39.89	2.69	6.37E-31	9.42E-27
SLC25A47	1539.32	1.83	5.88E-23	4.35E-19
BMP8B	50.19	2.02	1.42E-18	6.98E-15
SLX4IP	887.15	1.13	1.03E-17	3.82E-14
CBR4	456.42	1.23	1.46E-15	4.33E-12
GPT2	6504.60	1.47	2.96E-15	7.30E-12
KLHL3	1383.27	0.92	2.83E-14	5.99E-11
RDH11	3019.80	1.00	4.60E-14	8.51E-11
GOT1	1147.73	1.16	5.30E-14	8.72E-11
LOC102165634	68.65	1.89	7.85E-14	1.06E-10
LOC106505246	50.25	1.88	1.03E-13	1.27E-10
CTH	5813.01	1.27	1.82E-13	2.07E-10
HMGCS2	1613.50	1.79	1.37E-12	1.45E-09
GYS2	9890.09	0.93	1.51E-12	1.49E-09
ELOVL6	6439.98	0.94	7.09E-12	6.55E-09
ARFGAP3	1829.70	0.93	1.22E-11	1.06E-08
ASS1	9331.15	0.80	1.31E-11	1.08E-08
MIGA2	220.90	1.01	3.81E-11	2.96E-08
SEC16A	1099.80	0.78	5.62E-11	4.15E-08
LOC106506349	47.33	1.61	2.04E-10	1.44E-07

Table 4.3: The 20 most significantly upregulated genes in the MIDY pig liver. Mean expression and log2 fold change (l2fc) are DESeq2 normalized. LOCXXXXXXXXX genes are unannotated genes.

4.1.3 Amino acid metabolism

The most common theme in the upregulated transcripts is one of amino acid metabolism. Most amino acids had one or more catabolic enzymes upregulated. Specifically, the increased enzymes were: glutamic-pyruvic transaminase 2 (GPT2; alanine), glutamic-oxaloacetic transaminase 1 (GOT1; aspartate), glutaminase (GLS2; glutamine), arylformamidase (AFMID; tryptophan), homogentisate 1,2-dioxygenase (HGD; tyrosine, phenylalanine), serine dehydratase (SDS; serine), histidine ammonia-lyase (HAL; histidine), aminoadipate-semialdehyde synthase (AASS; lysine), aldehyde dehydrogenase 7 family member A1 (ALDH7A1; lysine), and kynurenine aminotransferase 1 (KYAT1; tryptophan, cysteine conjugates).

Gene symbol	Mean expression (counts)	l2fc (MIDY/WT)	p-value	FDR
SHROOM3	232.91	-1.25	6.16E-14	9.12E-11
IGFALS	641.04	-0.80	2.45E-10	1.65E-07
MEGF9	253.65	-1.03	5.14E-10	2.62E-07
CRP	3637.87	-1.18	6.15E-10	2.93E-07
PFKFB3	80.71	-1.15	7.63E-09	3.05E-06
THRB	837.77	-0.68	2.27E-08	7.63E-06
SFRP1	378.50	-0.99	1.05E-07	3.05E-05
TMSB4X	3052.66	-0.51	3.16E-07	7.53E-05
GIPC2	114.86	-0.97	3.39E-07	7.95E-05
PDE4B	201.77	-0.83	3.65E-07	8.31E-05
MOB3B	167.14	-0.94	3.87E-07	8.68E-05
AMPD3	532.75	-0.77	4.68E-07	9.75E-05
SGK1	2932.51	-0.88	6.24E-07	0.000123991
VIL1	92.67	-1.05	9.56E-07	0.000181266
SETBP1	424.61	-0.77	1.29E-06	0.000235104
LOC110258394	32.74	-1.19	1.41E-06	0.000250803
NREP	405.63	-1.06	1.81E-06	0.000310762
LOC110256685	151.76	-0.81	2.43E-06	0.000386522
LURAP1L	482.52	-0.72	2.79E-06	0.000430362
LOC110255180	191.79	-0.92	3.46E-06	0.000511224

Table 4.4: The 20 most significantly downregulated genes in the MIDY pig liver. Mean expression and log2 fold change (l2fc) are DESeq2 normalized. LOCXXXXXXXXX genes are unannotated genes.

4.1.4 Gene set enrichment analysis

Due to the count-based nature of RNA-Seq, the standard microarray-based GSEA was avoided and pre-ranked GSEA was utilized. Both weighted and unweighted ("classic") analyses were performed, however the unweighted approach was given more importance as the weighted analysis is not designed for the pre-ranked analysis type. For the ranked approach either signed log-transformed p-values or log2 fold change values for genes can be used as input. Both approaches were found to give similar results, and for compatibility with other approaches using p-values as cutoffs, the preferred input was -log10 transformed p-values signed by the direction of the change. In total, 40 gene sets were found significantly increased in the MIDY liver, while 90 in total were found decreased (**Table 4.5, Table 4.6**), showing a reversion of the gene-level trend where upregulated transcripts were more common (**Figure 4.4, Supplementary Figure B.1**). Increased gene sets in the MIDY

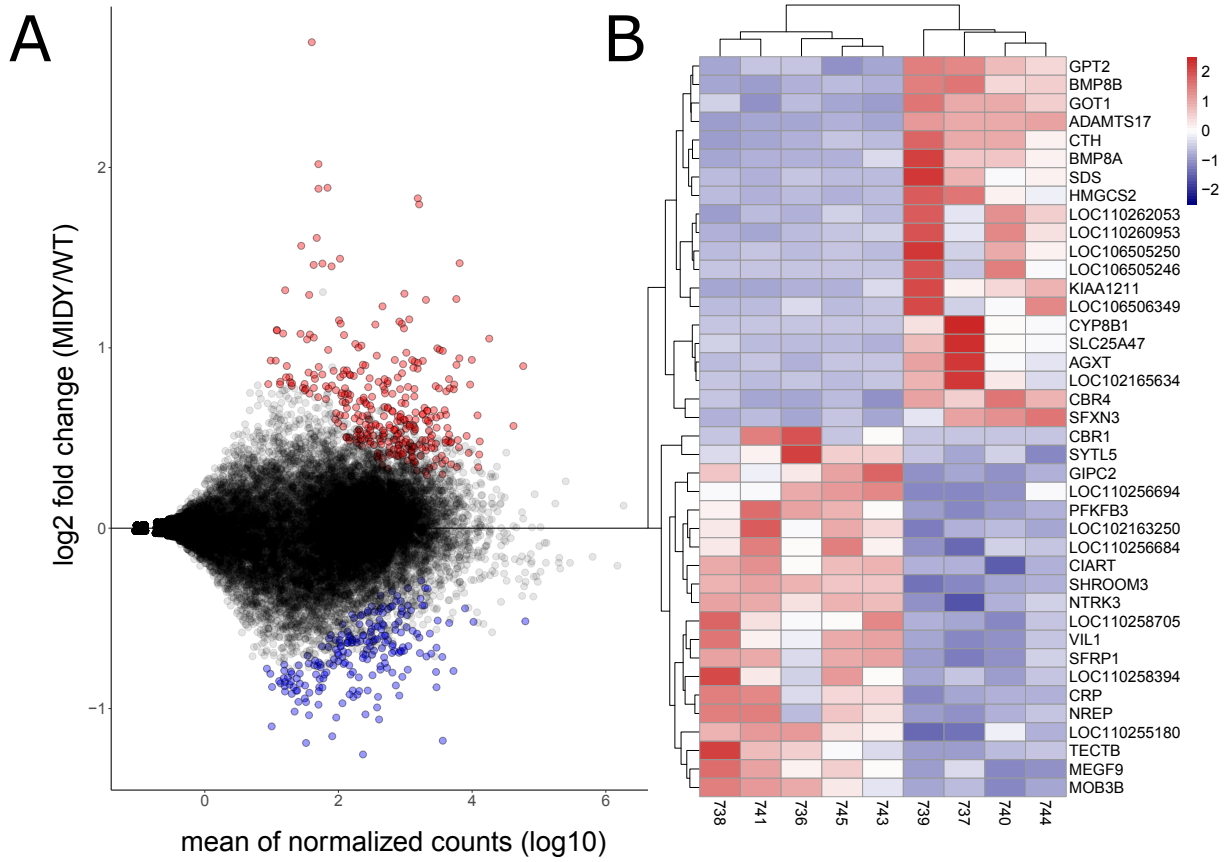

Figure 4.3: **A** MA-plot showing the distribution of expression and differential expression among MIDY transcripts. Log2 fold change here describes the DESeq2 corrected log2 fold change, while the normalized counts is the DESeq2 normalization. Red circles represent significantly upregulated transcripts, blue circles represent significantly downregulated transcripts ($p<0.05$). **B** Heatmap showing the expression of the top 20 most significantly changed genes in each direction. Each gene is normalized in expression. Adapted from [Backman et al., 2019].

pig were almost exclusively metabolic in nature. The most enriched gene set was of amino acid metabolism, with the gene sets of carbohydrate metabolism, propanoate metabolism and TCA-cycle following. Several individual pathways of amino acid metabolism were also increased. Non-metabolism connected pathways included folding, sorting and degradation, autophagy, and aging. Insulin signaling and insulin resistance pathways were also significant to a smaller degree. In the gene sets found decreased in the MIDY liver were an abundance of immune response-related gene sets. Infectious Diseases: Parasitic was the most enriched of all gene sets, followed by Hematopoietic Cell Lineage, Development, and Immune Diseases. A focus seemed also to be on Focal Adhesion and ECM interaction.

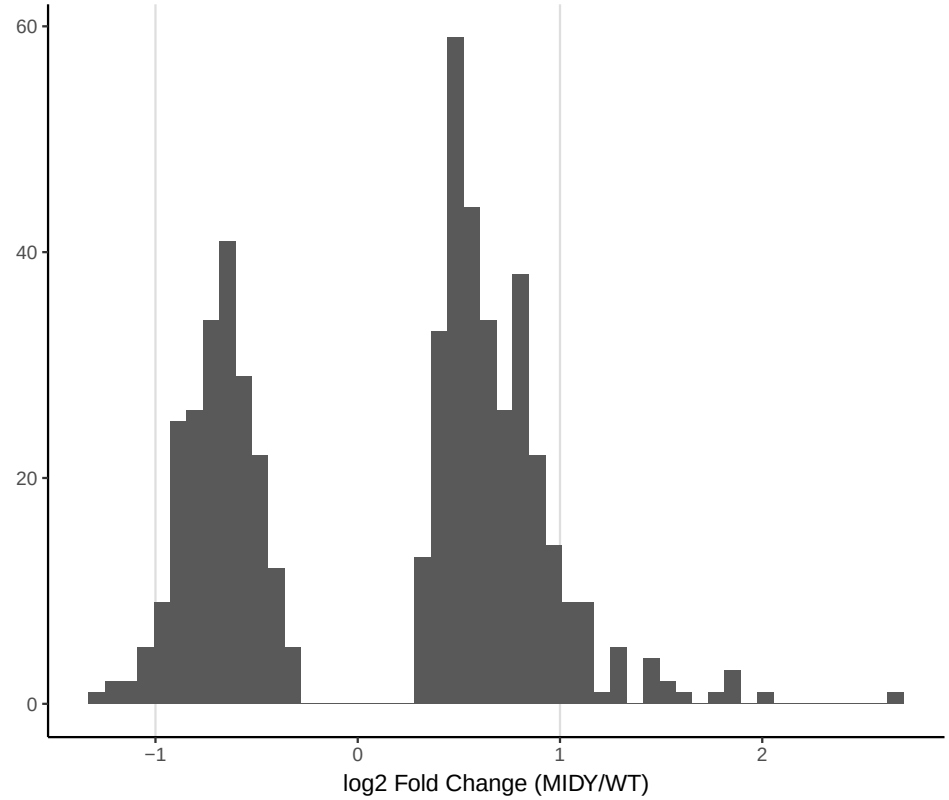

Figure 4.4: Histogram of the distribution of log2fold changes in mRNA between MIDY and WT in the liver showing more upregulated genes than downregulated ones.

Many gene sets here were irrelevant to the liver such as osteoclast differentiation, digestive system, and hypertrophic cardiomyopathy. The Diabetes Type I pathway was also found to be decreased.

4.1.5 Gene network

The alternate approach to the rank-based GSEA is to make an interaction network using only the significant genes. This was done using ClueGO, generating a network where similar trends to the results found in the GSEA approach could be identified (**Figure 4.5**). Metabolism pathways were dominant among genes increased in MIDY liver while Immune Response and ECM interaction were featured among genes decreased in the MIDY liver. Signaling pathways such as PPAR signaling and Sphingolipid signaling pathways served as the link between the two groups of genes.

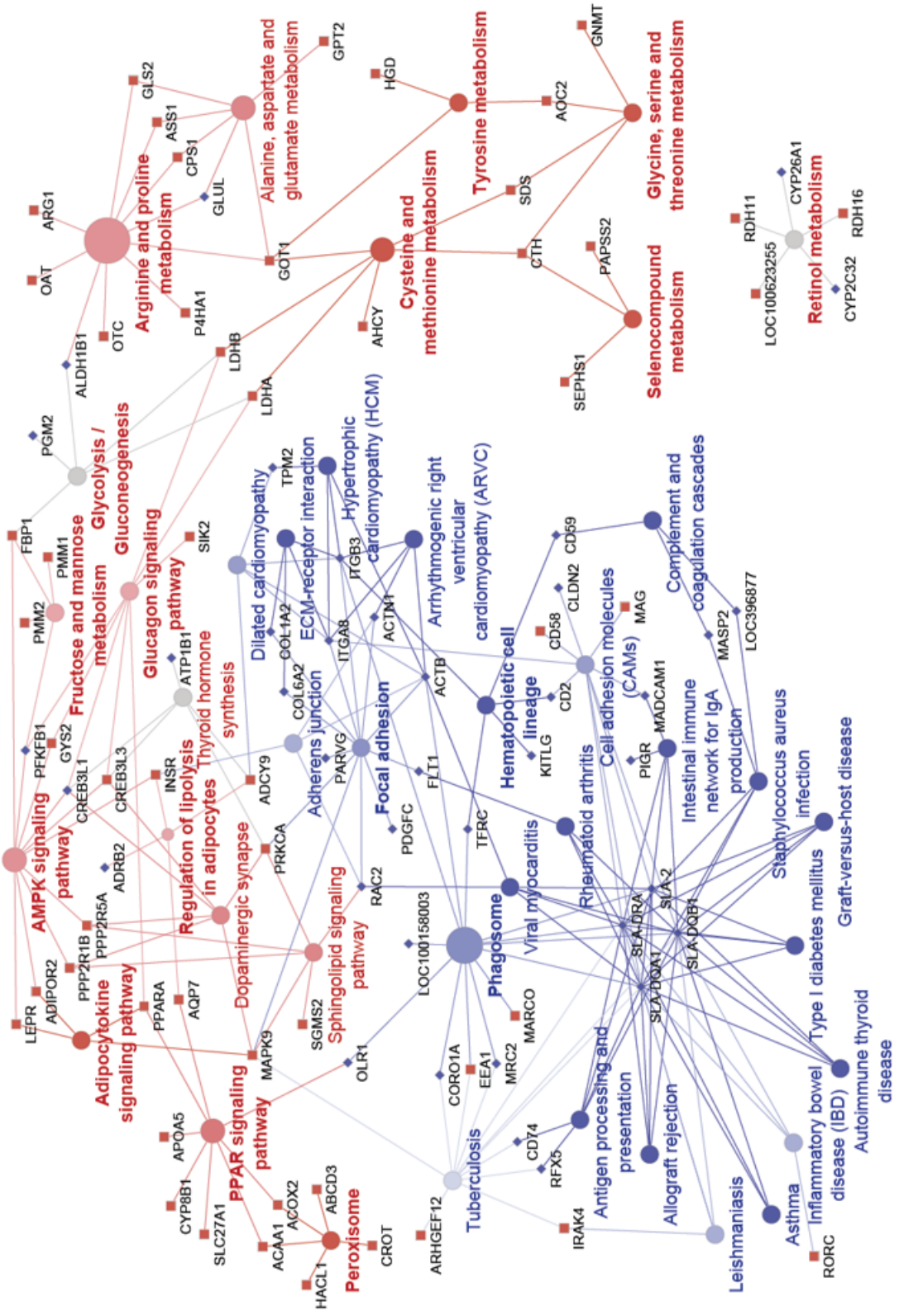

Figure 4.5: ClueGO gene network of the significant genes of the MIDY liver. Circles indicate gene sets while squares and diamonds represent up and downregulated genes respectively. Each line represents a membership of a gene in a set. Red color indicates an increase in MIDY while blue indicates a decrease. Adapted from Backman et al. [2019].

Gene set name	Size	NES	p-value	FDR
amino acid metabolism	259	4.43	0.00E+00	0.00E+00
carbohydrate metabolism	304	4.01	0.00E+00	0.00E+00
propanoate metabolism	30	3.55	0.00E+00	0.00E+00
citrate cycle (tca cycle)	29	3.45	0.00E+00	0.00E+00
valine, leucine and isoleucine degradation	48	3.28	0.00E+00	0.00E+00
metabolism of cofactors and vitamins	172	3.15	0.00E+00	0.00E+00
folding, sorting and degradation	441	3.09	0.00E+00	0.00E+00
arginine biosynthesis	19	2.78	0.00E+00	2.11E-04
pyruvate metabolism	36	2.50	0.00E+00	1.90E-03
mtor signaling pathway	143	2.48	2.02E-03	1.88E-03
glycine, serine and threonine metabolism	36	2.45	1.98E-03	2.03E-03
protein processing in endoplasmic reticulum	157	2.45	0.00E+00	2.15E-03
cysteine and methionine metabolism	42	2.45	0.00E+00	1.99E-03
metabolism of other amino acids	91	2.43	0.00E+00	2.20E-03
ampk signaling pathway	116	2.37	0.00E+00	3.12E-03
selenocompound metabolism	17	2.34	2.09E-03	3.50E-03
autophagy animal	122	2.34	0.00E+00	3.29E-03
butanoate metabolism	23	2.34	0.00E+00	3.16E-03
autophagy other	31	2.28	0.00E+00	5.25E-03
glycolysis / gluconeogenesis	57	2.27	0.00E+00	5.16E-03

Table 4.5: Upregulated gene sets from the GSEA of the liver tissue in the Munich MIDY Pig biobank.

4.2 Sequencing of MIDY liver by alternate sequencing libraries

Three alternate library sequencing techniques, Lexogen Sense, Lexogen QuantSeq, and SCRBSeq, were tested on the MIDY biobank liver samples in order to establish the preferred sequencing method for the remaining tissues. All three sequencing kits are cheaper than the NuGEN Encore Complete kit, which makes them preferred alternatives if effective.

4.2.1 Lexogen Sense library

The Sense library kit gave results similar to the ones from the sequencing of the NuGEN kit, however seemingly with lower power as only around two thirds as many significantly differentially expressed genes were found when sequencing depth was much larger (**Figure 4.7**). Of note is that the Sense kit cannot find non-polyA RNA as it enriches mRNA by

Gene set name	Size	NES	p-value	FDR
infectious diseases: parasitic	237	-5.03	0.00E+00	0.00E+00
hematopoietic cell lineage	79	-4.31	0.00E+00	0.00E+00
development	263	-4.23	0.00E+00	0.00E+00
immune diseases	187	-4.16	0.00E+00	0.00E+00
cytokine-cytokine receptor interaction	192	-4.14	0.00E+00	0.00E+00
focal adhesion	184	-4.06	0.00E+00	0.00E+00
cell adhesion molecules (cams)	124	-3.86	0.00E+00	0.00E+00
graft-versus-host disease	28	-3.75	0.00E+00	0.00E+00
complement and coagulation cascades	75	-3.69	0.00E+00	0.00E+00
infectious diseases: bacterial	309	-3.68	0.00E+00	0.00E+00
cardiovascular diseases	253	-3.64	0.00E+00	0.00E+00
intestinal immune network for iga production	36	-3.63	0.00E+00	0.00E+00
ecm-receptor interaction	75	-3.58	0.00E+00	0.00E+00
leukocyte transendothelial migration	96	-3.39	0.00E+00	0.00E+00
allograft rejection	30	-3.20	0.00E+00	8.13E-05
tight junction	149	-3.17	0.00E+00	7.62E-05
pi3k-akt signaling pathway	305	-3.07	0.00E+00	7.17E-05
chemokine signaling pathway	162	-3.07	0.00E+00	6.78E-05
axon guidance	171	-3.07	0.00E+00	6.42E-05
age-rage signaling pathway in diabetic complications	97	-3.03	0.00E+00	6.10E-05

Table 4.6: Downregulated gene sets from the GSEA of the liver tissue in the Munich MIDY Pig biobank.

binding to the tail, effectively lowering the potential amount of transcripts to be found. Still, the amount of non-coding transcripts is not enough to explain the discrepancy between the techniques. Samples still separated clearly on PC1 (**Figure 4.6 A**) and the p-value distribution was normal (**Figure 4.6 B**), even as the Lexogen samples were sequenced in two different runs, adding a batch effect. Gene set enrichment results were similar to that of NuGEN however with emphasis on different pathways (**Supplementary Figure A.4, A.5**).

4.2.2 Lexogen QuantSeq library

The QuantSeq kit from Lexogen yielded 74 significant genes (**Table 4.7**), representing a way to get a quick and cheap overview of transcription change. QuantSeq only sequences the 3' end of the transcript and as such cannot estimate alternative splicing events. Issues in the library preparation protocol were found as while PC2 of the PCA corresponded to

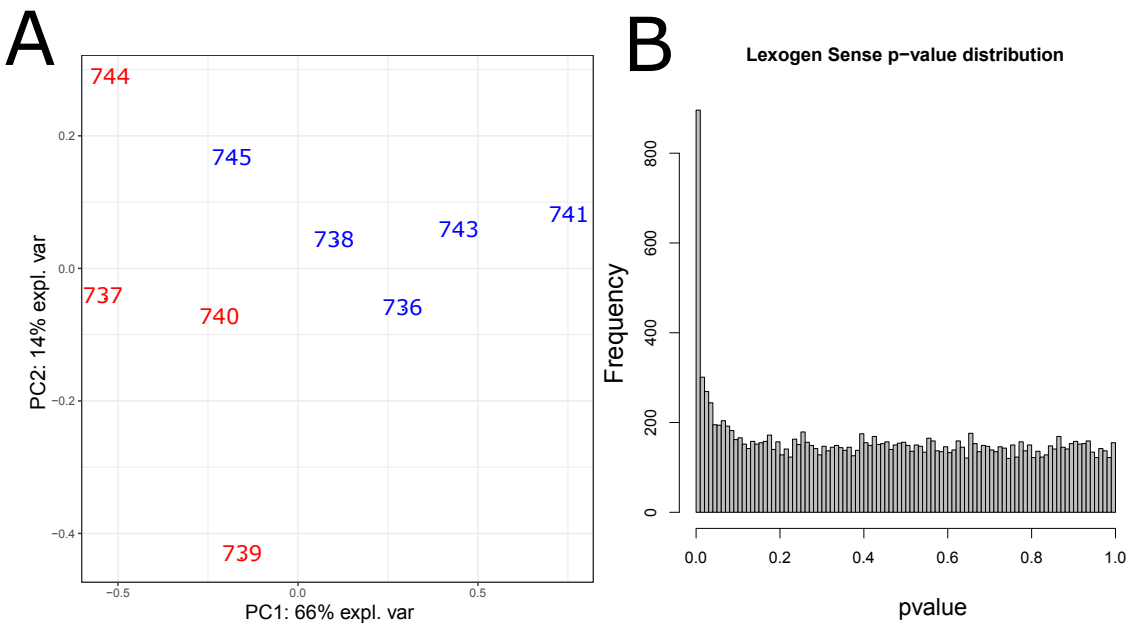

Figure 4.6: **A** PCA of the Lexogen Sense sequencing library. **B** A histogram showing the distribution of p-values from the differential expression analysis of the Lexogen Sense sequencing library.

the genotype separation, PC1 was found to correspond to the sample handling number (**Supplementary Figure B.2**), suggesting an important step of pipetting needed to be handled quicker, alternatively simultaneously. Due to this, a change was made in the protocol to use a multi-pipet, which solved the issue after the new procedure and sequencing were carried out. Separation of samples on PCA was not strong but visible, and the p-value histogram showed a balanced experiment (**Figure 4.7**).

4.2.3 SCRB-Seq library

SCRB-Seq (Single-Cell RNA Barcoding and Sequencing) yielded 19 significant genes (**Table 4.7**). Few PCR duplicates were detected, arguing against the technique as a correction against PCR duplicates. There are no clear separation on the PCA and the p-value histogram is not uniformly distributed (**Figure 4.8**), although this might be because of the low depth of the technique. Nevertheless, the 19 genes are very representative of the most interesting findings in the other library preparation methods, and as such, the method is still usable for a quick analysis.

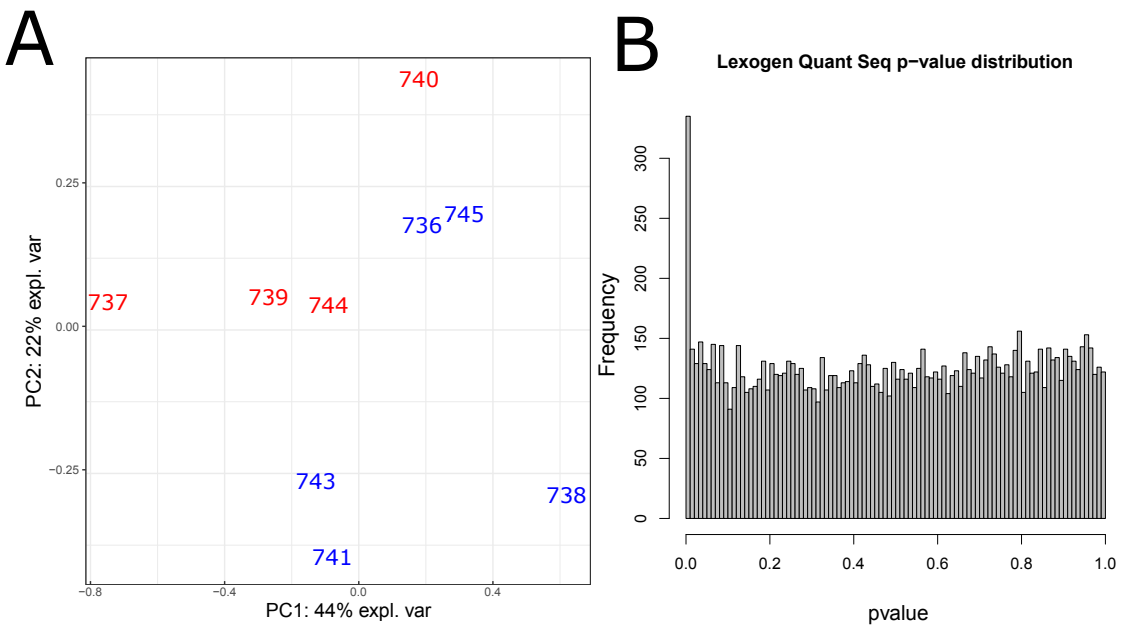

Figure 4.7: **A** PCA of the Lexogen QuantSeq sequencing library. **B** A histogram showing the distribution of p-values from the differential expression analysis of the Lexogen QuantSeq sequencing library.

4.3 Comparison of sequencing techniques

The different depth of the sequencing libraries prevents an easy comparison for the ideal sequencing library, and in addition the goal of the different techniques is also different. **Table 4.7** shows a comparison of the different techniques together with additional details. **Supplementary Figure B.3** shows the gene coverage saturation plots, clearly showing that both NuGEN and Lexogen Sense have reached saturation. While QuantSeq is almost at saturation. SCRB-Seq requires far more depth to reach saturation, however the two last techniques are intended to be used for quick overviews of the data rather than in-depth analyses and as such do not aim to reach the same saturation level. While QuantSeq finds more significant genes in general, comparison of the overlap between significant genes in different techniques (**Figure 4.9 A**) reveals that 81% of the significant genes in QuantSeq can be found in either of the two deeper techniques, for SCRB Seq this number is 95%. This suggests that while SCRB seq has a very shallow depth, it is very accurate with the genes it does find. When looking at gene sets found by GSEA however (**Figure 4.9 B**) there is not much difference as 92% of QuantSeq gene sets are found in NuGEN and Lexogen Sense, while for SCRB Seq the number is 89%. In the end, due to the low cost and high output, the Sense library preparation was chosen to be used for the rest of the

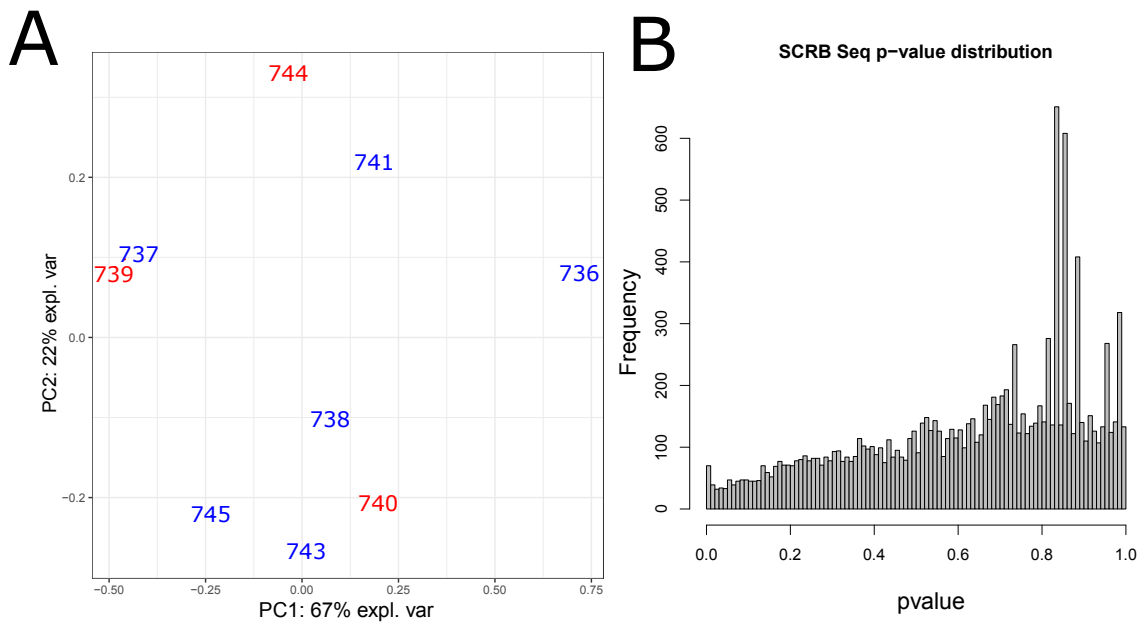

Figure 4.8: **A** PCA of the SCRB Seq sequencing library. **B** A histogram showing the distribution of p-values from the differential expression analysis of the SCRB Seq sequencing library.

biobank. For mass sequencing with a lower budget however, the QuantSeq technique might be more suitable.

4.4 Proteomics

2,535 proteins were identified with high confidence (FDR <0.01, [Elias et al., 2005]), and samples separated clearly into MIDY and WT on both hierarchical clustering and PCA (**Figure 4.10 A, B**). 60 of the identified proteins were considered significantly more abundant while 84 were significantly less abundant after multiple testing. RDH16, retinol dehydrogenase 1, had the highest abundance increase at a log2 fold change of 4.7, followed by HMGCS2, 3-hydroxy-3-methylglutaryl-CoA synthase 2 at a log2 fold change of 2.7. GO annotations and KEGG pathways were found using STRING analysis (**Supplementary Figure B.4**) which similar to the transcriptomics analysis found an increase in amino acid metabolism and central energy metabolism, while pathogen defense response, cellular stress response, and cell signaling were found lowered.

Library	NuGEN	Sense	QuantSeq	SCRB-Seq
M Reads	32	117	13.9	0.8
M Reads mapped to genes	16.6	89.7	6.6	0.3
Mapped to genes %	52%	77%	47%	33%
No. 10x covered genes	12311	13573	9492	2034
Genes found significant	533	367	74	19
Length (time)	08:30	06:00	05:00	08:30
Cost*	70€/sample	50€/sample	25€/sample	3€/sample
Notes	3x runs	2x runs		

Table 4.7: Overview of the different sequencing techniques used on the liver tissue in the Munich MIDY Pig biobank. Read depth is counted in million (M) reads. Cost only refers to the cost of library preparation, not sequencing. Both the NuGEN and Sense libraries were run multiple times as indicated.

4.5 Integration of transcriptomics and proteomics data

Transcript IDs matching protein IDs were used for data integration and concurrent analysis. 1572 matches were found and thus used for both integration and correlation. Individual gene correlation was created using log2 differential abundance values from both experiments, which in the case of transcriptomics were corrected for dispersion (**Figure 4.11 A**). Correlation between mRNA and protein was moderate (R = 0.34) which was as expected, since transcript amounts alone are not able to accurately predict protein amounts [Liu et al., 2016]. However, for the more strongly affected proteins such as HMGCS2, RDH16, SLC22A7 and COL1A1, transcript abundances correlated strongly, suggesting that transcripts can work as a replacement for proteins where proteomics fails, for example with lowly abundant proteins and low confidence measurements. The 2D pathway enrichment of transcripts and proteins showed few significant anti-correlating pathways, with most gene sets ending up on the correlating diagonal (**Figure 4.11 B**). Results are similar to the enrichments found in the individual omics enrichments, with metabolic pathways such as gluconeogenesis, urea cycle, glucagon signaling pathway, and biosynthesis of amino acids found increased while ECM and immune response were found decreased.

4.6 Metabolomics characterization of MIDY pig plasma

For the initial characterization of the Munich MIDY pig biobank, a metabolomics measurement of the plasma was made in addition to the clinical characterization. The MIDY and

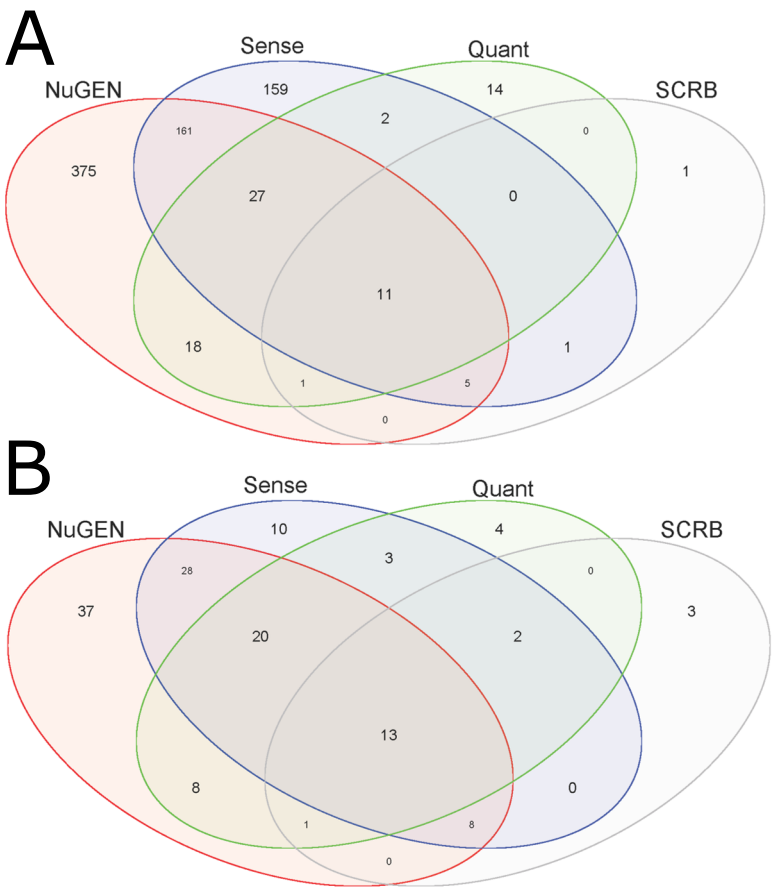

Figure 4.9: Venn diagrams showing the overlap between sequencing techniques, comparing **A** overlapping significant genes, and **B** overlapping significant GSEA KEGG pathways.

WT pigs separated clearly on the first PC (**Figure 4.12 A**) indicating substantial differences in the metabolic makeup. As expected, hexose levels were significantly increased in the MIDY pig, as well as the proportion of unsaturated fatty acids. The ratio of long-chain acylcarnitines to free carnitine, aka the CPT1 ratio, was significantly increased, showing increased beta-oxidation. Other changes included an increase in branched-chain amino acids and the amino acids lysine, phenylalanine and tryptophan. Metabolites decreased in the MIDY pigs were sphingomyelins and the ratio of total dimethylated arginine to total unmodified arginine (**Figure 4.12 B**) (**Supplementary Table A.2**). Free fatty acids were measured as they regulate hepatic triglyceride content. Concentration of FFA was increased in MIDY plasma, although not significantly (**Supplementary Figure B.5**).

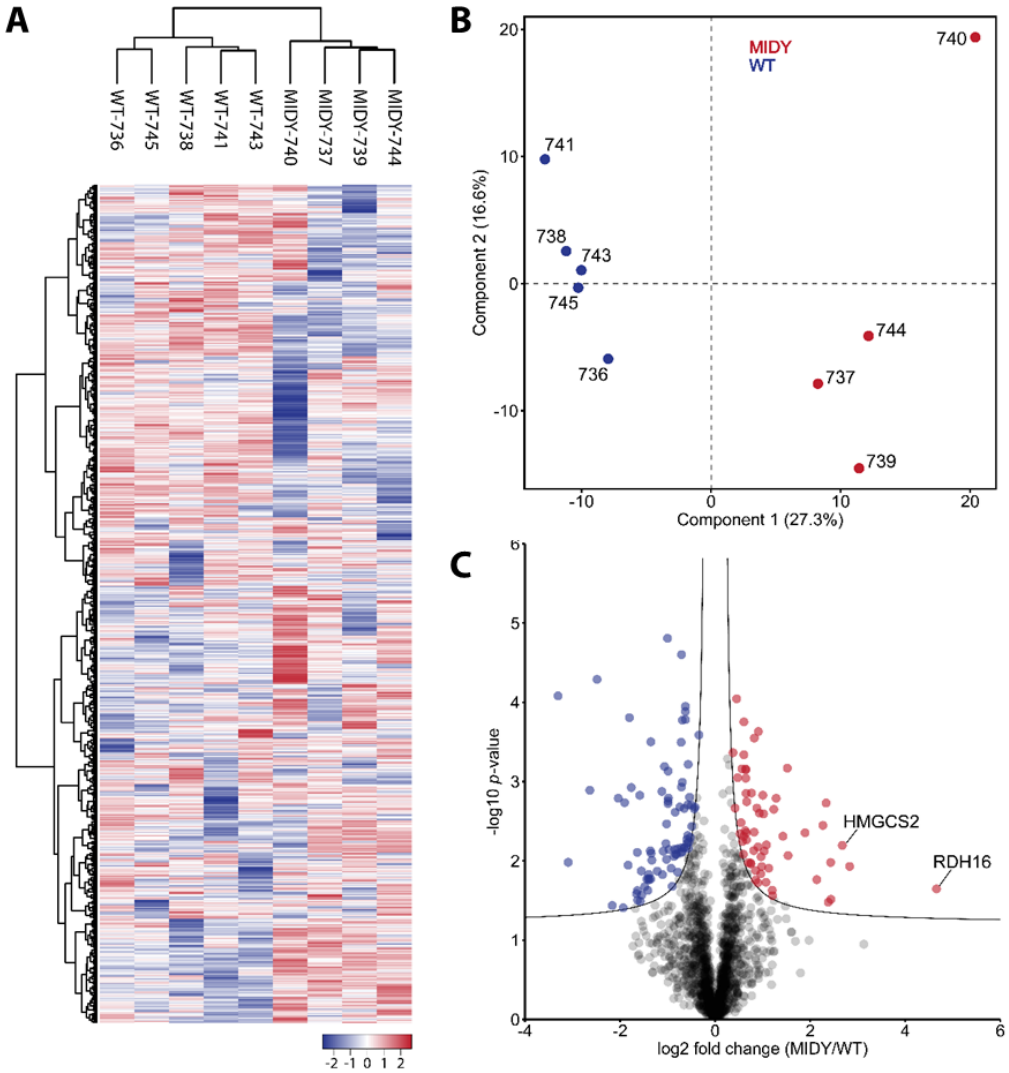

Figure 4.10: Proteome analysis of MIDY and WT pigs using mass spectrometry. **A** shows unsupervised hierarchical clustering of normalized LFQ intensity values. **B** shows a PCA that separates MIDY and WT pigs on the first component. **C** shows a volcano plot of log2 fold changes between MIDY and WT. Red and blue circles indicate significantly differentially abundant proteins. The black curves show the FDR cutoff for significance. Adapted from [Backman et al., 2019].

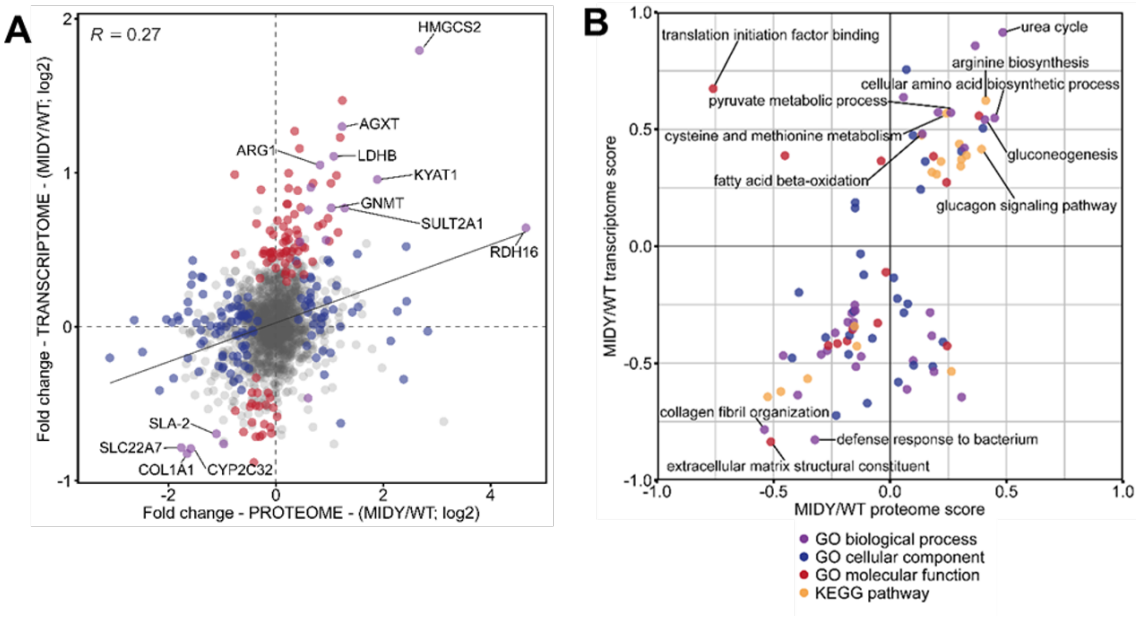

Figure 4.11: Correlation between transcriptomics and proteomics measurements on genes found in both measurements. **A** shows the correlation between transcriptomics and proteomics data. Red color indicates significance in transcriptomics but not in proteomics, blue color indicates significance in proteomics but not in transcriptomics, purple indicates significance in both. **C** shows the 2D annotation enrichment analysis between the two data sets. Transcriptomics and proteomics scores are displayed only for pathways with a p-value of $p < 0.01$. Pathways are colored according to the database they belong to. Adapted from Backman et al. [2019].

4.7 Metabolomics characterization of MIDY pig liver

As a trend, amino acids were increased in the MIDY pigs (**Figure 4.13 A**). Especially increased were the amino acids lysine and methionine, which were around 170% of their WT concentration. Branched amino acids were also increased to around 125% of their WT concentrations. Only three amino acids were lower in the MIDY pigs; arginine was not detectable in the MIDY condition and thus could not be analyzed for significance. Serine was only around 60% of its WT concentration while glycine had a very minor decrease in concentration. Several biogenic amines were also changed (**Figure 4.13 B**), such as an increase in spermine and histamine levels and a decrease in creatinine and serotonin levels in MIDY. Several other metabolites such as kynurenine and dimethylated arginine were irregularly increased.

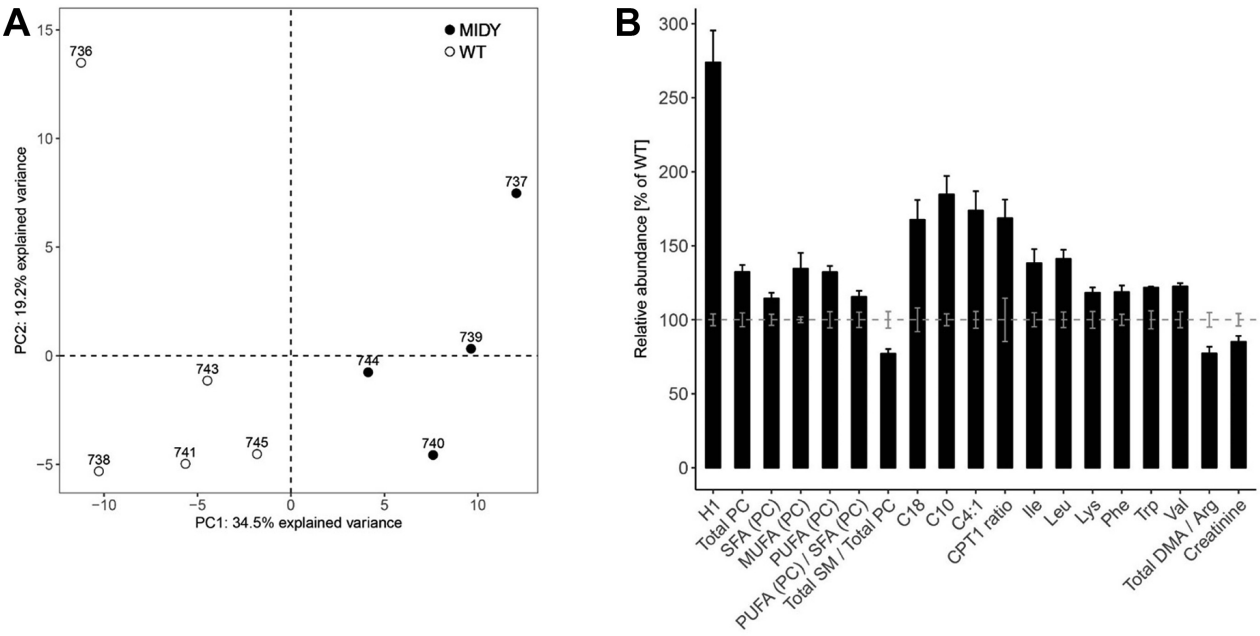

Figure 4.12: The results of the metabolomics characterization of the MIDY plasma. **A** shows the PCA of MIDY and WT control samples, **B** shows the relative change in expression of selected relevant metabolites using WT as a base line (gray striped line). SEM are indicated for both conditions. Adapted from Blutke et al. [2017]

4.7.1 Acylcarnitines

The indicator of beta-oxidation, the CPT1 ratio, was significantly increased in MIDY pigs (**Figure 4.13 C**). The CPT1 ratio is the ratio of long-chain acylcarnitines, C16+C18, to that of free carnitine, C0. C0 did not have any change, however both C16 and C18 were significantly increased in MIDY pig liver. Furthermore, short length acylcarnitines and the total amount of acylcarnitines were decreased. Modified acylcarnitines, i.e. dicarboxy-acylcarnitines and hydroxy-acylcarnitines, were increased in comparison with total acylcarnitine levels.

4.7.2 Lipids

Lipidomics and metabolomics experiments both found phosphocholines to be unchanged while sphingomyelins were reduced in the MIDY pigs. As a trend other lipids were reduced in MIDY samples (**Figure 4.13 D**), notable exceptions being diacylglycerol (DAG) and triacylglycerol (TAG). lyso-phosphatidylserine (LPS) and phosphatidic acid (PA) had a marked reduction, while cholesterol and phosphatidylserine (PS) had minor reductions.

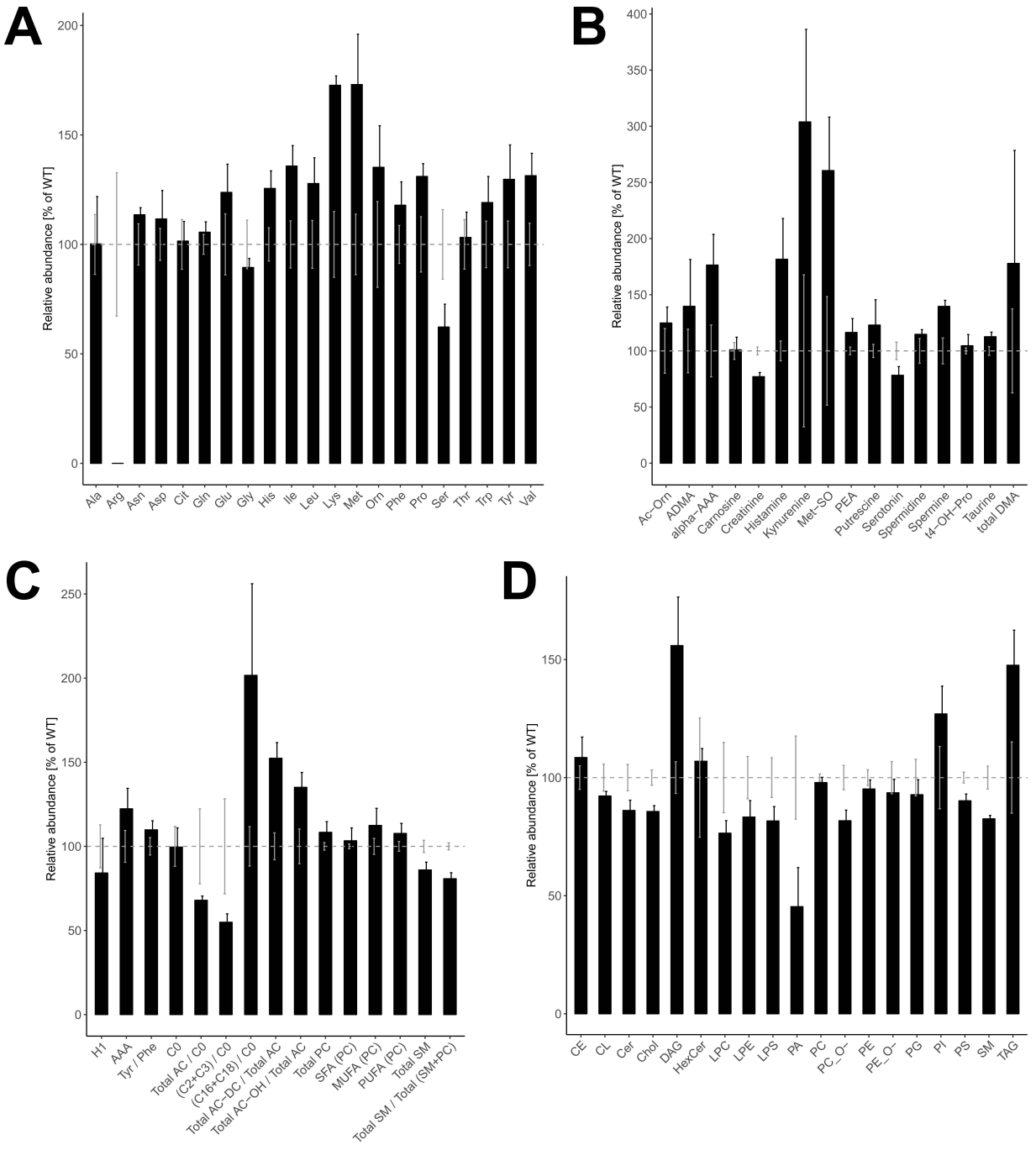

Figure 4.13: The results of the metabolomics and lipidomics characterization of the MIDY liver. The graphs show the relative abundances of amino acids **A**, biogenic amines **B**, and selected metabolic indicators **C**, determined by targeted metabolomics. Graph **D** shows the relative abundance of lipid groups determined by lipidomics. The abundance is shown as a percentage of WT expression, which is shown as 100% (gray striped line). SEM is indicated for both conditions. Adapted from Backman et al. [2019].

4.8 Insulin signalling

The phosphorylation state of several important proteins of the insulin signaling pathway was determined by western blot in order to verify the effect of lower plasma insulin (**Figure 4.14, Supplementary Figure B.6**). Insulin receptor (INSR) phosphorylation levels were significantly reduced in contrast to INSR transcript and protein amounts which were significantly increased. Phosphoinositide 3-kinase PI3K phosphorylation levels were not changed although the protein and transcript amounts were as a trend reduced. MIDY 3-phosphoinositide-dependent protein kinase-1 (PDPK1), protein kinase B (PKB, AKT) and glycogen synthase 3 beta (GSK3B) all had significantly reduced phosphorylation levels compared to WT. Both forkhead box protein O1 (FOXO1) protein and phosphorylation levels were lower in MIDY, although not significantly. Mechanistic target of rapamycin (mTOR), AMP-activated protein kinase (AMPK), and ribosomal protein S6 were not changed either in protein abundance or transcript amount.

4.9 Retinoid signaling

The substantial increase in abundance of RDH16 led to a quantification of retinoid levels in the MIDY liver (**Figure 4.15**). Retinol, retinal and retinoic acid (atRA) were quantified by mass spectrometry, revealing a significant increase in retinol and atRA levels, while retinol remained largely the same.

4.10 Glutathione

For measurement of oxidation levels, glutathione (GSH) and its reduced form GSSG were measured (**Supplementary Figure B.7**). While no significant differences were noticed, the amount of GSSG in MIDY pigs was close to being significantly increased ($p=0.076$). Another potential indicator of oxidation, taurine, was also almost significantly increased in the MIDY pig ($p=0.06$). Methionine sulfoxide is also a reliable indicator of oxidation, however the compound had missing values in the WT pigs, which along with the increased methionine content made the compound an unreliable indicator.

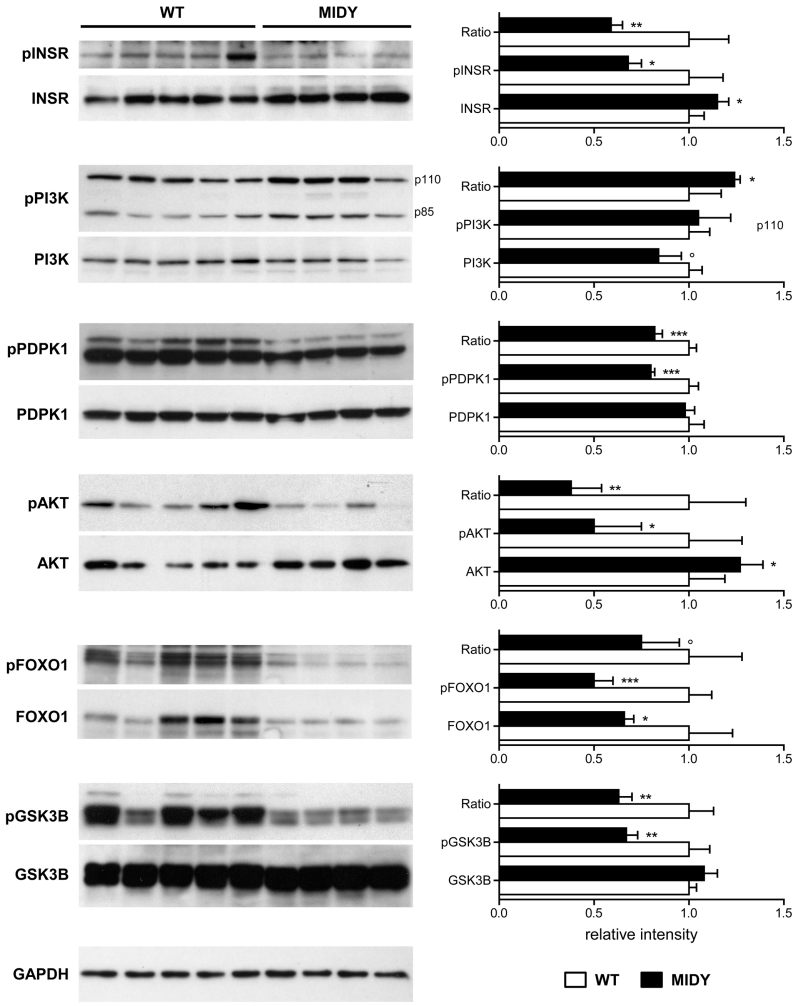

Figure 4.14: Western blot analysis of activation in insulin receptor and downstream molecules. The bar diagrams show means and standard deviations. Significant differences between MIDY and WT pigs are indicated by asterisks: *p <0.05; **p <0.01; ***p <0.001; °borderline significance (p <0.07). Adapted from Backman et al. [2019].

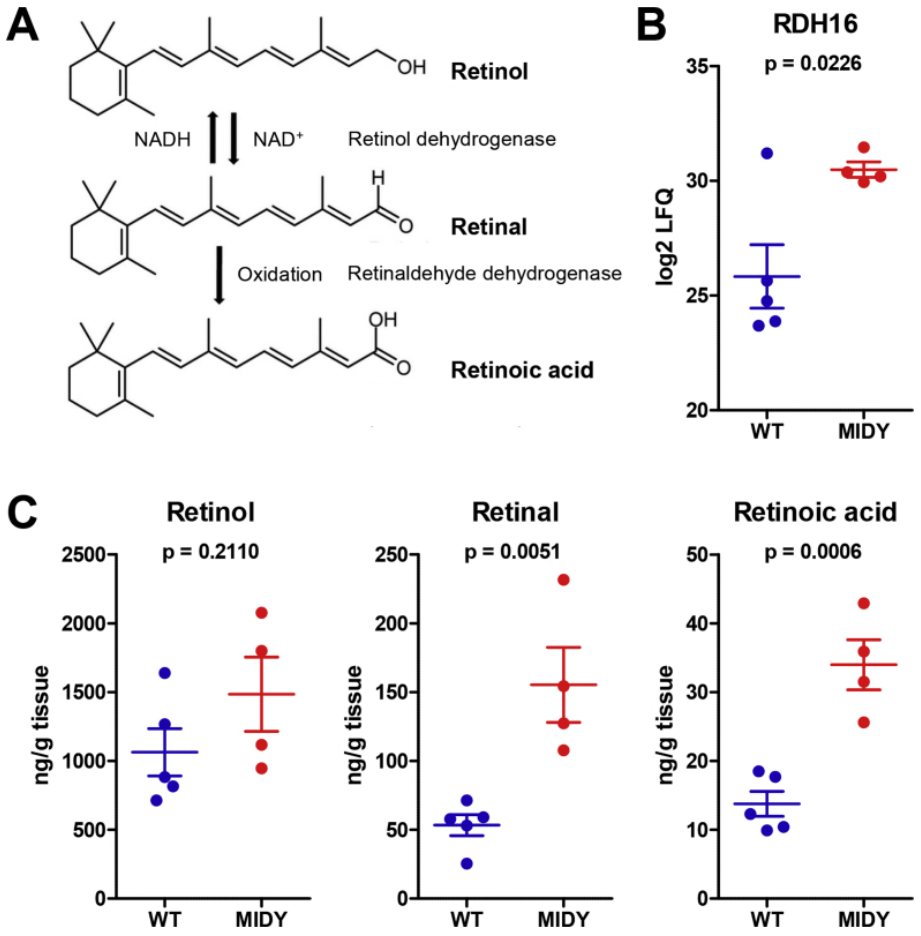

Figure 4.15: Summary of retinol metabolism and levels in MIDY liver. **A** shows the pathway from retinol to active atRA. **B** shows RDH16 levels in MIDY liver which is correlated to **C** which describes the levels of retinol, retinal and atRA in the MIDY liver. Concentrations are in ng per g liver tissue. Significance was determined using student's t-test. Adapted from Backman et al. [2019].

4.11 Kupffer cells

Kupffer cells were measured in order to investigate whether the immune system between the MIDY and WT livers was activated in different levels. No significant differences could be found between the two tissues, however a minor decrease and a more narrow variation in volume density could be found in the MIDY pigs (**Figure 4.16**).

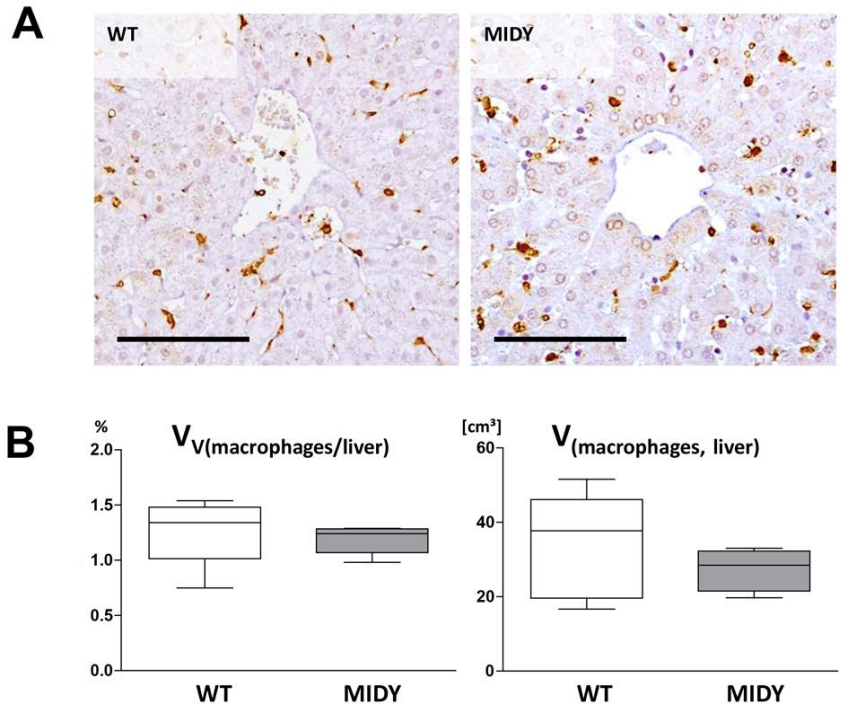

Figure 4.16: Quantification of IBA1-positive macrophages (Kupffer cells) in MIDY and WT liver. **A** shows macrophages detected by immunohistochemistry and **B** shows the volume density as a percentage of total liver cell volume. Adapted from Backman et al. [2019].

4.12 Adipose tissue sequencing

Four different fat tissues were available for sequencing in the MIDY biobank: Subcutaneous adipose tissue from the abdomen and the back, mesenteric adipose, and perirenal adipose tissue. Sequencing of subcutaneous back and mesenteric tissue was performed first, due to quality issues resequencing was required, adding batch effects. In the end, complete resequencing of the mesenteric tissue and sequencing of the other two adipose tissues was performed. While this new sequencing of the mesenteric tissue retained an outlier, it was good enough to be used for analysis. All tissues were distinguishable from each other when looking at a PCA on PC2 and PC3 (**Figure 4.17**).

4.12.1 Differential expression analysis of adipose tissue

None of the adipose tissues were able to be analyzed by simply applying DESeq2 to the samples, some type of uncorrected batch factor seemed to exist even when the different

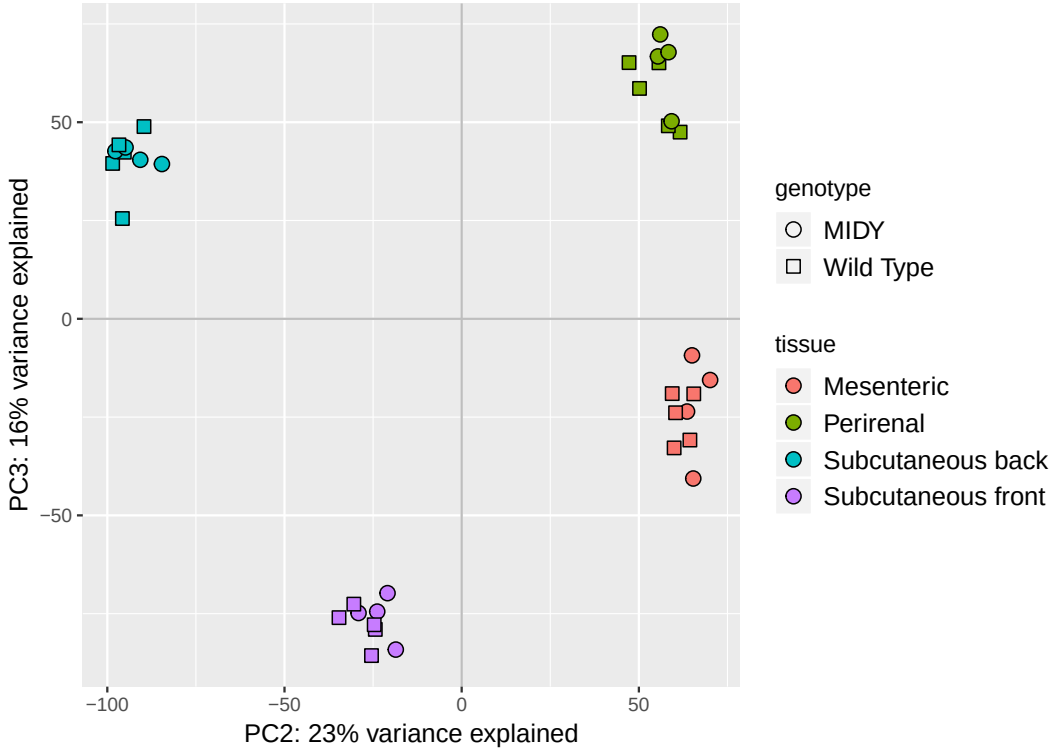

Figure 4.17: PCA showing how the adipose tissue samples from the MIDY biobank cluster together on the second and third principal component.

sequencing runs and even the pregnancy status of the animals were taken into account. This effect could be seen on the p-value histograms, which were either overly liberal as in the "subcutaneous abdomen" or overly conservative as seen in the rest (**Figure 4.18, Supplementary Figure B.8**). Due to this SVA was employed to correct the distribution. "Subcutaneous rear" was corrected using 3 SVs yielding 161 significant genes, "subcutaneous abdomen" using 2 SVs yielding 84 significant genes, "mesenteric" by 2 SV yielding 160 significant genes, and "perirenal" by 2 SVs yielding 91 significant genes (**Figure 4.18, Supplementary Figure B.8**). A Venn diagram (**Figure 4.19 A**) of the significantly differentially expressed genes showed a minor overlap between the different tissues, with the only gene to be altered in all tissues being retinol dehydrogenase 16, aka RDH16, a prominent gene in the hepatic tissue analysis as well. The minor overlap could signify a too heavy correction by SVA, or simply more difference between locations than expected. A Venn diagram over gene sets found significant by GSEA showed a similar overlap as on the gene level, with mostly the mesenteric tissues showing a strong phenotype (**Figure 4.19 B**). Only TCA-cycle was found upregulated in all tissues.

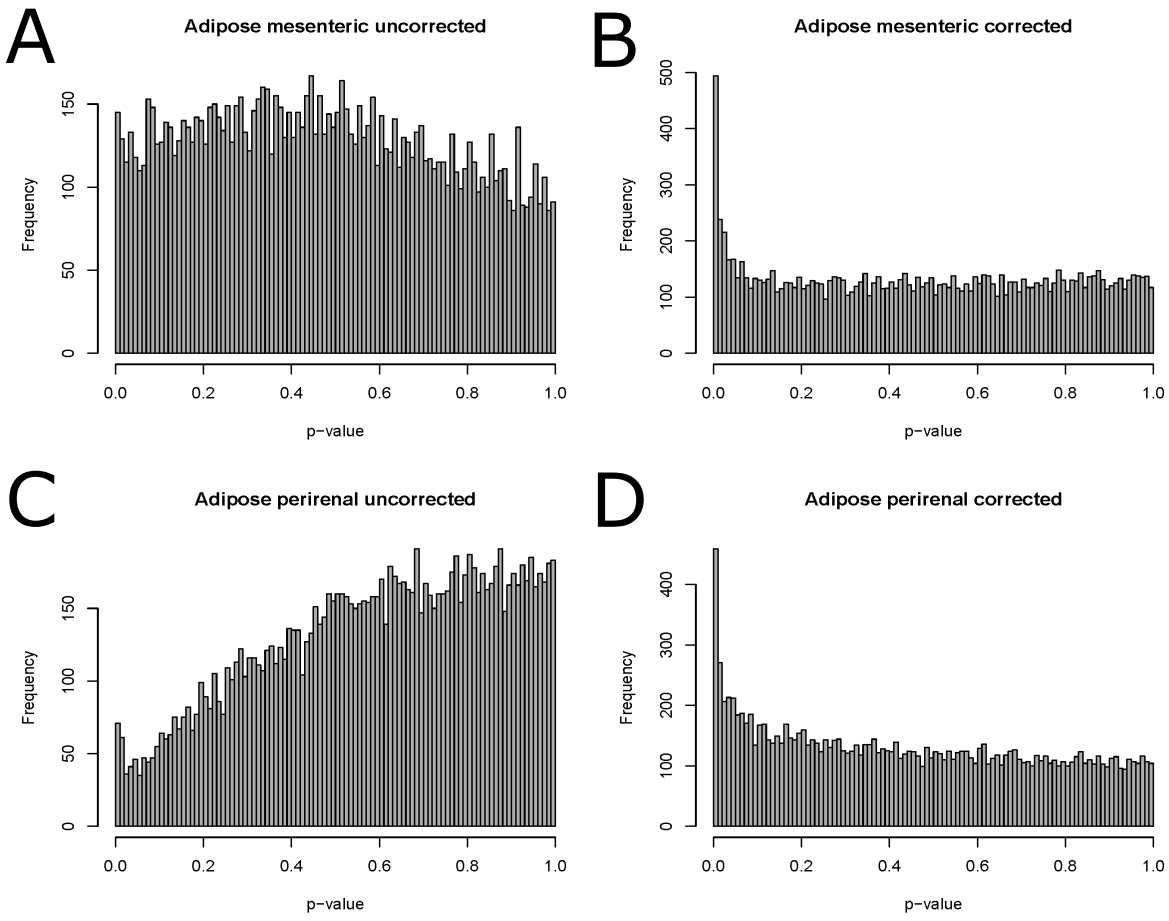

Figure 4.18: Plots showing the SVA correction of the two visceral adipose tissues with **A,C** showing the p-value histogram before correction while **B,D** show the p-value histogram after correction.

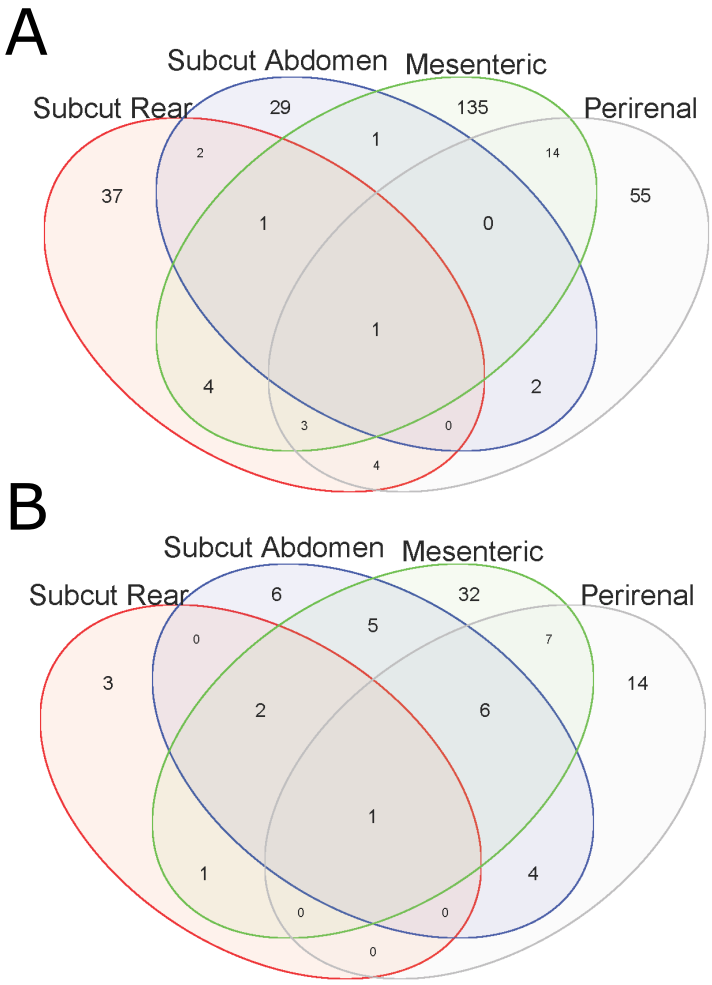

Figure 4.19: Venn diagram of the **A** genes and **B** gene sets significantly differentially expressed genes in MIDY adipose tissue

Chapter 5

Discussion

In this thesis, the livers of Munich MIDY Pig biobank sows were analyzed on different omics levels. These levels contribute to the holistic profile of the organ differently: RNA-Seq captures changes in the whole genome and is sensitive to genes which have lower expression, proteomics represent fewer but more functionally interpretable changes, and metabolic/lipidomic characterization show a limted but direct readout of cellular perturbations. The multiomics analysis, being exploratory in nature, reveals both expected and unexpected biological perturbations. While they are expected, the details of the metabolic changes still reveal new findings.

5.1 Evidence for stimulated gluconeogenesis

While not quite significant on the mRNA level, the parallel proteomics experiment found the abundance of phosphoenolpyruvate carboxykinase (PCK1), the rate-limiting enzyme of gluconeogenesis [Obrochta et al., 2015], to be significantly increased. The reason for the discrepancy between mRNA and protein could be time related as proteins represent long term changes while transcripts are short term changes. The mRNA of several other enzymes involved in gluconeogenesis, i.e., glutamic-pyruvic transaminase 2 (GPT2), L-lactate dehydrogenase A chain (LDHA), L-lactate dehydrogenase B chain (LDHB), and alanine-glyoxylate aminotransferase (AGXT) were significantly increased in the MIDY samples. In concordance, since gluconeogenesis and glycolysis cannot be active at the same time, the transcript level of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3), a key stimulator of glycolysis (reviewed in [De Bock et al., 2013]), was significantly decreased.

5.2 No increased glycogen synthesis in spite of upregulated GYS2

The rate-limiting step of glycogen synthesis in the liver is glycogen synthase 2 (GYS2) [Roach et al., 2012], which was significantly more abundant in the MIDY samples than in WT. Since the phosphorylation level of GSK3B (an inhibitor of GYS2) was reduced however (**Figure 4.14**), increased glycogen synthesis in the liver of MIDY pigs was not expected and could not be observed. Reductions in both synthesis and breakdown of hepatic glycogen have been found in patients with poorly controlled type 1 diabetes [Bischof et al., 2001, Regnell and Lernmark, 2011].

5.3 Increased RDH16 expression as a potential link between insulin deficiency and stimulated gluconeogenesis

The mRNA of two retinol dehydrogenase genes were found to be upregulated in the MIDY liver; RDH11 and RDH16. Simultaneous proteomic studies showed that RDH16 was the most increased protein in the MIDY pig liver. This was however not a surprise as RDH16 is negatively regulated by insulin through signaling by FOXO1 [Obrochta et al., 2015]. RDH16, which is also known as RDH1 in mouse [Zhang et al., 2001], is known to be one of the two retinol dehydrogenases responsible for creating the active form of vitamin A, all-trans retinoic acid (atRA), along with RDH10 (**Figure 5.1**) [Belyaeva et al., 2008, Napoli, 2012]. RDH10 however, is the dominant enzyme, the RDH10 null mutant is lethal whereas RDH16 is not [Rhinn et al., 2011]. Obrochta et al. [2015] showed that RDH10 and RDH16 regulate gluconeogenesis, this could imply that RDH16 is the version of retinol dehydrogenase required for hypoglycemic stress. We measured levels of retinol, retinal and atRA to verify this and found atRA to be significantly increased in the MIDY liver. Supporting the idea of an atRA enrichment in the MIDY liver is the significant increase in the CYP26A1 gene expression, the main catabolic enzyme of atRA. A null-mutant mouse for RDH1 showed a decrease in CYP26A1 suggesting the two genes have a connection in regulation [Zhang et al., 2007].

5.3.1 RDH16 was also increased in abundance in adipose tissue

After sequencing of the transcriptomes of the adipose tissues, it was found that RDH16 was also increased in these tissues, it was in fact the only transcript found significantly differentially expressed in all four sub-tissues. In three of the four tissues, the exception being mesenteric tissue, RDH16 is the most significantly increased transcript. Once again this result was confirmed in a proteomic experiment analyzing two of the tissues where RDH16 was the most significantly increased protein. This is at odds with the idea that RDH16 is mainly responsible for gluconeogenesis activation since adipose tissue does not have gluconeogenesis. atRA binds to RAR and RXR elements, motifs that when bound increase transcription of nearby genes [Marill et al., 2003]. RAR and RXR elements regulate not only metabolism, but also development and immune response. While developmental effects should be unimportant in the adult sows, immune response is found to be reduced in all MIDY tissues analyzed so far with no obvious explanation as to why. As a possibility, RDH16 could be a mechanism for counteracting this reduction. A more worrying possibility is that RDH16 is a pig specific response to diabetes, which explains why the change in expression has never been observed before. Another possibility is that insulin signaling does not take into account the tissue, and thus only one of the tissues requires the regulation of RDH16 by insulin, while in the other it is just upregulated as a side-effect. Clearly, other tissues of the MIDY biobank should be explored in order to research deeper into this issue.

Pigs are notorious for being a mammal lacking brown adipose tissue (BAT) [Trayhurn et al., 1989]. BAT generates heat by using uncoupling protein 1 (UCP1) which uses the built up electron potential in the mitochondrial membrane to generate heat instead of ATP, and which is dysfunctional in pigs [Berg et al., 2006]. RDH16 has recently been shown to regulate browning in mice [Krois et al., 2019], leading to questions whether there is an alternate heat-generating mechanism utilized in pigs that is regulated by RDH16. As discussed later on, one gene that was found upregulated in the liver is the liver-specific uncoupling gene SLC25A47, hinting that perhaps another similar gene could exit for pig adipose tissue.

5.4 Changes in amino acid metabolism

Under conditions of constant gluconeogenesis leading to an increase in amino acid catabolism, it would be expected that more than normal amounts of ammonia would be produced, thus requiring a more active urea cycle than under normal conditions. Indeed, the KEGG path-

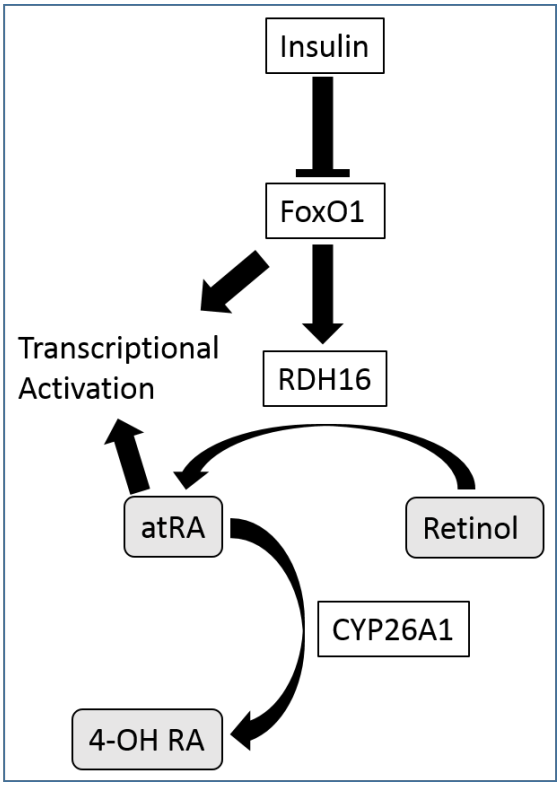

Figure 5.1: Overview of atRA metabolism in the liver. White rectangles indicate genes while grey indicates metabolites. Insulin leads to transcription of RDH16 through FoxO1. RDH16 turns Retinol into atRA, an important transcriptional activator. CYP26A1 is responsible for atRA clearance.

way for the urea cycle is significantly increased, with most enzymes showing an increase in transcription. The mRNA levels for urea cycle enzymes carbamoyl-phosphate synthase 1 (CPS1), ornithine carbamoyltransferase (OTC), argininosuccinate synthase 1 (ASS1), and arginase 1 (ARG1) were significantly increased in MIDY samples. This was corroborated by the parallel proteomics study which found increased levels of OTC and ARG1. The increase of ARG1, responsible for converting arginine to urea and ornithine [Morris Jr, 2002], could be a possible explanation as to why arginine can not be detected in the MIDY samples while it is found in all but one of the WT samples. Ornithine also has a minor increase in abundance in accordance with this theory.

One amino acid with a high increase in the MIDY pig is lysine. Lysine degradation is more complicated than most amino acids, requiring several steps. The first enzyme of the pathway, alpha-aminoadipic semialdehyde synthase (AASS), was found significantly increased in the MIDY liver, suggesting that the liver is trying to metabolize lysine. Sac-

charopine, the product of lysine metabolization by AASS has found to be toxic, suggesting there might be an alternate mechanism preserving lysine [Leandro and Houten, 2019]. As one of the two purely ketogenic amino acids, it would be expected that less lysine is used for energy than other amino acids as ketone body generation is mainly derived from fatty acid degradation. With an increase in protein degradation to fuel gluconeogenesis, an increase in lysine would then be expected. The other purely ketogenic amino acid is leucine, which as a branched-chain amino acid is mainly metabolized in adipose and skeletal muscle, explaining the difference between lysine and leucine concentrations. Lysine is interestingly the precursor of carnitine [Vaz et al., 2001], however, since carnitine levels are stable there is no indication that the increased lysine levels affect carnitine levels.

5.4.1 Methionine, serine and cysteine metabolism

Methionine is an essential amino acid that has an important role in the cell as it is necessary for methylation reactions and provides a source for the production of cysteine. The pathway leading from methionine to cysteine involves 5 enzymes, 4 of which are significantly increased in the MIDY pig liver. Nevertheless, the abundance of methionine is increased in the MIDY pig liver, leading to questions about the cause of this increased methionine content. The start of Methionine metabolism lies in the methionine cycle, a metabolic recycling mechanism for maintaining a pool of methionine [Martinov et al., 2010]. The mRNA of methionine adenosyltransferase 1A (MAT1A), the one enzyme not significantly increased, is still increased by 20%, meaning the whole pathway is increased. Glycine N-methyltransferase (GNMT), the enzyme causing S-adenosyl methionine (SAM) turnover, requires glycine in order to function properly as it transfers the methyl group of SAM onto glycine to create S-adenosyl homocysteine (SAH) and sarcosine. However, glycine levels in the MIDY liver were only slightly decreased, most likely not hindering the reaction. Homocysteine is then created by metabolizing SAH using the enzyme adenosylhomocysteinase (AHCY). Homocysteine, which is toxic to the cell, can be used to regenerate methionine by using the enzyme methionine synthase (MTR). The MTR transcript had a minor insignificant decrease, most likely not enough to be the reason for methionine enrichment.

Homocysteine can be utilized in the transsulfuration pathway to produce cysteine [Sbodio et al., 2019]. To do this homocysteine is combined with serine by the enzyme cystathionine beta-synthetase (CBS). Since serine is reduced in the MIDY liver, one reason for the methionine enrichment could be a slower than normal transsulfuration pathway. The final step of the transsulfuration pathway produces cysteine by cystathionine gamma-lyase

(CTH). Cysteine cannot be measured by the targeted metabolomics kit used, and due to the instability of thiols it cannot be measured so long after sampling. However, while cysteine itself cannot be measured its polypeptide form glutathione can be measured. In a human hepatoma cell line, it has been shown that 50% of cysteine derived from the transsulfuration pathway is incorporated into glutathione [Mosharov et al., 2000]. Glutathione is the most important antioxidant of the body and is formed by conjugating glycine and glutamate onto serine [Lu, 2013]. Total glutathione levels were not altered, however the levels of oxidized glutathione, GSSG, were changed as a trend, showing some amount of oxidative stress. It should also be noted that due to the age of the samples and as the instability of thiols apply to glutathione as well, both the ratios and total amount of glutathione could have been changed. Cysteine can also be converted into taurine, another antioxidant, starting with the enzyme cysteine dioxygenase 1 (CDO1). While Taurine levels were increased by a non-significant amount, CDO1 was markedly reduced in expression, suggesting the MIDY liver is trying to conserve cysteine residues.

Serine levels could be reduced due to several different reasons. The most probable reason is a significant increase in serine dehydratase (SDS), the enzyme responsible for serine catabolism [Xue et al., 1999]. SDS produces pyruvate from serine which can then be shuttled into the gluconeogenesis pathway. Serine is also used to produce sphingomyelins, a lipid family that was found reduced in the MIDY liver in the targeted metabolomics experiment and also together with its precursor ceramide in the lipidomics experiments (**Figure 4.13 C, D**). This reduction could just be due to the lower serine content in the first place as serine is rate-limiting for this production [Merrill Jr et al., 1988]. Another reason for serine depletion could be it being used as a fuel for one-carbon metabolism [Scott and Weir, 1998]. Serine provides the one-carbon units of tetrahydrofolate, which ultimately comes full circle as donor of methyl groups to homocysteine, regenerating it to methionine. Possibly due to the toxicity of homocysteine, it is being recycled at a higher rate by an increase in tetrahydrofolate. Under normal circumstances serine could simply be synthesized by the hepatocyte, however as the metabolite used for serine production, 3-phosphoglycerate, is located in the middle of the gluconeogenic chain, with active serine conversion into pyruvate, this essentially creates a futile cycle.

Serine can also be generated by the glycine cleavage system, which interlinks the two aminoacid pools [Locasale, 2013]. While serine can directly be converted to glycine, two glycines are required by the cleavage system, one of which will be consumed. Glycine levels were also reduced in MIDY pigs but only to a minor degree (**Figure 4.13 A**). One pressure

on the glycine concentration could be the production of glutathione.

In summary, what is going on in methionine-serine metabolism is most likely a reduction in available serine due to gluconeogenesis, which then leads to a lack of serine for the transsulfuration pathway. Instead, homocysteine will be used to regenerate methionine which requires a one-carbon unit, further putting pressure on serine levels. An increased level of methionine means more methylation substrates, possibly leading to hypermethylation. Lack of transsulfuration instead could mean less production of cysteine and thus less antioxidation potential in the cell.

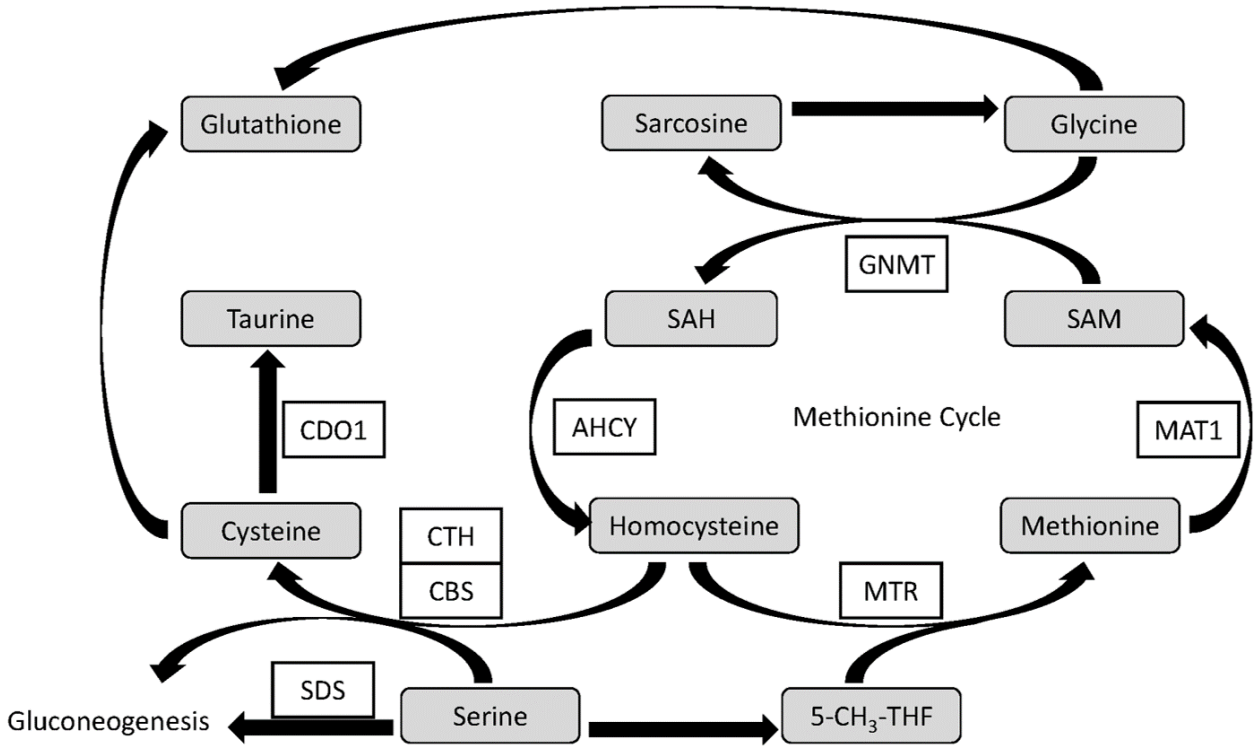

Figure 5.2: A condensed view of Methionine/Serine metabolism. White rectangles indicate genes while grey indicates metabolites. Shown genes are genes highlighted in the discussion showing especially the methionine cycle and the transsulfuration pathway.

5.5 Stimulated ketogenesis and beta-oxidation

5.5.1 Acylcarnitines

Stimulated ketogenesis could be detected in the MIDY pigs by measurement of beta-hydroxybutyrate, which was significantly increased in the MIDY pigs [Blutke et al., 2017]. In ketogenesis, acetyl-CoA is produced by the metabolism of fatty acids (reviewed in [Newman and Verdin, 2014]). In order to metabolize the fatty acids, they need to be transferred into the mitochondria across the mitochondrial membranes. As mitochondria are impermeable to acyl-CoAs, the fatty acid is instead transferred onto carnitine which can move across the membranes in the so-called carnitine shuttle. The enzyme carnitine palmitoyltransferase 1A (CPT1A) is the enzyme responsible for this conversion. The ratio of long-chain to short-chain acylcarnitines ((C16 + C18)/C0) is indicative of the activity of CPT1A, and thus also beta-oxidation. The enzymes involved in beta-oxidation are increased as a trend, but not significantly so. In addition, there was an increase in the levels of dicarboxylated acylcarnitines which is indicative of a higher level of omega-oxidation of fatty acids [Longo et al., 2016]. At first this could seem counter-intuitive as omega-oxidation is generally active when beta-oxidation is defective, however, it could also be that a higher level of acyl-CoA overloads the beta-oxidation system and thus omega-oxidation can help reduce pressure.

5.5.2 Carnitine as a buffer for free CoA

While carnitine's main function is the shuttling of fatty acids into the mitochondria, they have another role as regulators of the availability of free CoA. When excess levels of mitochondrial acyl-CoA exist, they can be transesterified with l-carnitine to form acylcarnitine and free CoA. As a proxy to measuring the ratio of mitochondrial acyl-CoA to free CoA, the extracytosolic ratio of acylcarnitine to free carnitine (C0) can be used as the ratios are reflective of each other [Reuter and Evans, 2012]. As excess acyl-CoA can mean mitochondrial disturbance, the ratio of total acylcarnitine to free carnitine can be used to measure mitochondrial disturbance. The normal ratio in human liver has been measured in two different studies to be 35%/65% and 44%/56% [Harper et al., 1993, De Sousa et al., 1990]. The first ratio is close to the values measured in the WT controls of this study which is 37%/63%. The MIDY pig had a ratio of 30%/70% due to reduced amounts of short-chain acylcarnitines. This suggests that acyl-CoAs could be efficiently processed in the MIDY

pig and that free CoA pools were maintained correctly. Acetylcarnitines could also be exported from hepatocytes to economize COA, however this was unlikely since the plasma acylcarnitines and liver acylcarnitines of this study has similar levels ($1.07 \pm 0.56\,\mu M$) and WT controls ($1.05 \pm 0.34\,\mu M$) [Blutke et al., 2017].

5.5.3 Markedly upregulated HMGCS2

The first and rate-limiting step of ketogenesis is carried out by the mitochondrial enzyme 3-hydroxy-3-methylglutaryl-CoA synthase 2 (HMGCS2). Transcript levels of HMGCS2 were highly increased, and in the parallel proteomic analysis HMGCS2 was also the most increased protein. Transcription of HMGCS2 is activated by FOXA2 [Newman and Verdin, 2014] which in turn is inactivated by insulin signaling through the PI3K-AKT pathway [Wolfrum et al., 2003]. The reduced AKT activation is thus likely the cause of increased HMGCS2 levels. Enzymes involved in ketogenesis downstream of HMGCS2 such as 3-hydroxy-3-methylglutaryl-CoA lyase (HMGCL) and 3-hydroxybutyrate dehydrogenase 1 (BDH1) were not increased in abundance in the MIDY liver samples. This supports the notion that HMGCS2 alone is sufficient for stimulated ketogenesis in the liver.

5.6 Upregulation of antioxidative mechanisms

Several antioxidative mechanisms were found in the MIDY liver, in accordance with the fact that hyperglycemia increases oxidative stress [Mohamed et al., 2016]. The transcript abundance of the solute carrier 25 A47 (SLC25A47), which encodes a liver mitochondria specific uncoupling protein that facilitates proton leak in the mitochondrial matrix and thus leads to less ATP and H2O2 production [Jin et al., 2009, Mailloux and Harper, 2011b], was significantly increased. SLC25A47 is not well described and has so far not been described in the context of diabetes. As none of the other uncoupling proteins are increased, the identification of the transcript as highly increased suggests that it's one of the main antioxidative proteins of the liver for hyperglycemic conditions. On the proteomics level glutathione S-transferase mu 2 (GSTM2), which has been found to protect against oxidative stress, was significantly increased [Hayes et al., 2005]. Furthermore, protein levels of glutathione peroxidases 1 & 4 (GPX1, GPX4) were also increased. This could explain the fact that glutathione was not found significantly oxidized as the glutathione peroxidases would counter this (**Figure B.7**).

Isocitrate dehydrogenase 1 (IDH1), the cytosolic version of isocitrate dehydrogenase,

which has antioxidant properties by producing NADPH for the regeneration of glutathione [Itsumi et al., 2015], was also increased in abundance in MIDY samples both on the transcript and protein level. IDH1 requires the presence of cytosolic isocitrate, as isocitrate mostly resides in the mitochondrial TCA-cycle. Accordingly, mRNA levels for the mitochondrial citrate/isocitrate transporter solute carrier 25 A1 (SLC25A1) were significantly increased in the MIDY pigs. In addition, the mRNA level for the plasma membrane transporter of citrate/isocitrate, SLC13A5, was significantly increased. Plasma citrate/isocitrate transporter (SLC13A5) transcripts were significantly increased as well, possibly indicating uptake of circulating plasma isocitrate. All this together shows an ongoing enrichment of cytosolic citrate/isocitrate, most likely in order to fuel IDH1 in producing NADPH.

5.7 Alterations of lipid metabolism

Both triacylglycerides (TAG) and diacylglycerides (DAG) were found to be increased in the MIDY liver (**Figure 4.13 D**), suggesting increased synthesis, storage or uptake of lipids. The increase of DAG, which is the precursor of TAG, most likely means an increase in TAG synthesis. It has been shown in rat that plasma free fatty acids levels were the main factor in the esterification of fatty acids into hepatic TAG, with insulin action being largely independent [Vatner et al., 2015]. Since the plasma concentration of TAG was slightly increased [Blutke et al., 2017] it might contribute to hepatic TAG levels, however the reverse might be true as well. While TAG was increased there was no signs of nonalcoholic fatty liver disease (NAFLD). In humans, NAFLD is found in up to 40% of adult patients of type 1 diabetes [Targher et al., 2018]. This discrepancy could be either due to the younger age of the pigs or due to the natural resistance of pigs against NAFLD [Renner et al., 2018].

5.7.1 Apolipoproteins

Apolipoproteins are protein-lipid complexes that function as transporters of lipids, mainly TAG and cholesterol, in the bloodstream. While no part of cholesterol metabolism was altered in the MIDY pig liver, the main protein of high-density lipoprotein (HDL), apolipoprotein A1 (APOA1), was found increased on the protein level. Since HDL transports cholesterol from peripheral tissues towards the liver, cholesterol would be expected to be increased in the liver, which is, the reverse of our lipidomics results (**Figure 4.13 D**). Cholesterol

is not increased in plasma either [Blutke et al., 2017] suggesting either an increase in cholesterol catabolism or a localization towards peripheral tissues. Apolipoproteins 4 & 5 (APOA4, APOA5) were found increased in the transcriptome. APOA5 is closely related to TAG transport as part of several lipoproteins and has been found to facilitate cytosolic storage of TAG in hepatocytes [Shu et al., 2010], possibly contributing to the increase of hepatic TAG.

5.8 Changes in extracellular matrix

ADAMTS17 was the most increased transcript in the MIDY pig, going from an almost non-expressed level in the WT. ADAMTS17 is not very well described, it is known to be expressed in the liver, however its upregulation in diabetes is not yet described. The members of the ADAMTS family have a number of different functions relating to the ECM as secreted zinc metalloproteases [Apte, 2009]. Several different collagens were found significantly decreased in the MIDY pig liver; COL1A1 were found both on transcriptome and proteome level, COL1A2, COL5A2, COL6A2, and COL6A3 were significant on the transcriptome level while COL6A5 and COL14A1 were significant on the proteome level. In general, the vast majority of collagens were decreased, suggesting a decrease in ECM in MIDY pigs compared to WT pigs. COL1 is specifically the collagen of scar tissue and frequent in nonalcoholic steatohepatitis (NASH) [Lefebvre et al., 2017], however, in the MIDY pigs the collagen expression is decreased instead of increased, something not previously described. AKT, which in MIDY pigs have reduced activation, is involved in activating collagen synthesis and could thus provide an explanation to the decreased collagen levels. Inflammation of the liver leads to fibrosis through the stimulation of hepatic stellate cells (HSC) [Koyama and Brenner, 2017]. HSCs are in turn activated by the hepatic macrophages, Kupffer cells. As such, volume densities of MIDY and WT Kupffer cells were compared, showing a non-significant decrease of Kupffer cells in MIDY samples (**Figure 4.16**). The variance in MIDY sample Kupffer cell mass was more narrow showing a higher variance in WT pigs which could possibly explain the non-significant but consistent decreases in individual ECM and immune system genes which can be seen as significantly downregulated pathways in the rank-based GSEA results enriched in the WT pig (**Figure 4.6**).

5.9 Reduced activation of immune and inflammatory mechanisms

Immune function was in general downregulated in the MIDY pigs, both on the mRNA and the protein levels. This was a surprise as inflammation would usually occur in diabetic livers. Multiple explanations for this behavior could exist. Since the liver is connected to the gut via the portal vein there can be a direct influence of gut-derived substances such as remains of gut-flora on the cells of the liver including immune cells [Szabo, 2015]. As small amounts of microbial remainders reach the liver they are taken up by the Kupffer cells, the resident hepatic macrophages that mainly work as scavengers. It is possible that a change in the intestinal microbiome results in less activation of the immune system. Another possibility is a difference in blood content in the organ, stemming from narrower sinusoids due to ECM remodeling for example. The possibility of fewer resident macrophages, Kupffer cells, were investigated and discussed in the previous section. As the MIDY pigs are diabetic during development, there is also a possibility that the development of their immune system is stunted. This is supported by the fact that the immune system seems to be compromised in adipose tissue as well.

5.9.1 Inflammatory mechanisms

Pattern recognition receptors (PRRs) are the activators of the innate immune system by recognizing either pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs) [Lu et al., 2016, Cui et al., 2014]. In the liver data there were several activators of immune response that were downregulated: One of the main activators of inflammation C-reactive protein (CRP), was significantly downregulated in the transcriptome. It has been reported that CRP levels increase in plasma under conditions of insulin resistance and DM2, however here we see the complete opposite in a DM1-like condition [Nesto, 2004, Grossmann et al., 2015]. It is however possible that the MIDY pig has some sight insulin resistance [Blutke et al., 2017], something shown to be occurring also in Akita mice [Hong et al., 2007], which would further add a complication to the immune response question. The DAMP high mobility group protein B1 (HMGB1) was found to be decreased in the proteome analysis. HMGB1 is an early inflammatory mediator and activator of the PRR toll-like receptor 4 (TLR4). Accordingly, the levels of several TLR4 regulated proteins [Abbas et al., 2005] were downregulated: proteasome activator complex subunit 2 (PSME2), GMP reductase 1 (GMPR), protein transport protein Sec61 subunit

beta (SEC61B), and 2-5-oligoadenylate synthetase 2 (OAS2). Other proteins connected to TLR signaling were also found downregulated: Rac family small GTPase 1 (RAC1), protein phosphatase 2 scaffold subunit A alpha (PPP2R1A), ubiquitin-conjugating enzyme E2 D2 (UBE2D2), S100 calcium-binding protein A1 (S100A1), legumain (LGMN), and mitogen-activated protein kinase 3 (MAPK3).

5.9.2 MHC molecules

One consistent finding in the transcriptome is the downregulation of major histocompatibility complex class 2 (MHC2) genes, suggesting a change in macrophage and lymphocyte number or activation. MHC2 is the MHC unique to antigen-presenting cells, i.e. macrophages and dendritic cells, and serves an important part of the adaptive immune system. Three of the pig counterparts to human HLA genes, SLA-DQA1, SLA-DQB1, and SLA-DRA are all significantly downregulated, and all other MHC2 genes are downregulated by a substantial amount. Additionally, the gene regulatory factor X5 (RFX5) which regulates the expression of MHC2 genes [Van Den Elsen, 2011] was found significantly reduced, serving as an explanation. SLA1, SLA2, and SLA3, genes of MHC1, were also all reduced, though SLA2 was the only significant one. MHC1 can be found on all cells and thus only tells about the activation levels and not the cell numbers.

5.10 Conclusion

5.10.1 Benefits of large animal models

The field of large animal model diabetes is a relatively unexplored field due to the high costs of animal maintenance and expertise required, even with the clear benefits that animals close to humans in phenotype provides. Previous to the results explained in this thesis, no other study has looked on the liver of a clinically relevant large animal model under controlled conditions. Several findings such as the role of RDH16, expression of ADAMTS17 and changes in sulfur amino acid metabolism have never been studied before and thus show the value of the model organism.

5.10.2 Benefits of biobanking

The system of using biobanks to make use of every part of an animal is not only more ethical and cost-saving in many cases, but also allows for discoveries to be made that otherwise would not be found. In the case of RDH16, it was thought to be upregulated in the liver due to its effect on energy metabolism and gluconeogenesis, however, it was discovered to be the one gene found significantly increased in all adipose tissues where gluconeogenesis does not occur. If only the liver had been studied it would not have been taken into consideration that another reason for the increase of RDH16 could exist, such as in adipose tissue. Furthermore, some tissues, such as adipose tissue and liver tissue, have a high level of crosstalk with each other and thus should ideally not be studied in isolation if a holistic picture of the disease is to be created. Biobanking also allows for an exploratory approach to be followed up with hypothesis-driven studies in the same system as ample amounts of samples exist.

5.10.3 Future prospects

As demonstrated by the numerous insights generated by only the analysis of liver tissue, much can be gained by a holistic exploratory approach to disease research. While the adipose tissue has not yet thoroughly been explored, the discovery of RDH16 suggests that much more will be gained by further analysis of other tissues and retinol levels should be checked in relevant tissues. A deeper look into the actual function of RDH16 might require an RDH16 knockout, perhaps even a MIDY/INS^{C94Y} knockout to find whether RDH16 is required for correct atRA signaling under hyperglycemic stress. The point about the alternate browning of porcine adipose tissue is also interesting to consider, even though not related to the topic of diabetes.

Two other not so well described genes were found that might be of interest in further studies: ADAMTS17 is largely undescribed, most likely due to its low expression level in WT tissue, and its function can only be guessed from its family members. As most other ECM related genes are downregulated, the upregulation of ADAMTS17 might be a countermechanism to this. Except for being the liver-specific uncoupling protein SLC25A47 is not described. The reason for upregulation in diabetes is almost certainly to reduce oxidative stress from overactive mitochondria, as the other function of uncoupling proteins, heat generation, is not an issue in the liver.

More time should also be put into investigating the origin of the immune system per-

turbations. One of the few differences between the effects of MIDY diabetes and DM1 is the timing of the onset of diabetes. As MIDY diabetes is neonatal it will affect the pigs from birth while DM1 will have an onset in childhood, leading to possible developmental problems. If pigs are indeed immune-compromised, the question becomes how early in development this perturbation happens, and how applicable this would be to young DM1 patients.

While transcriptomic RNA-Seq analyses should most likely be carried out on all tissues due to its wide-reaching approach, the effect of MIDY diabetes might not be strong enough on some tissues for a proteomics analysis to be worth it, as the sensitivity to small changes is weaker as seen here where the RNA-Seq approach finds more than twice the number of significant genes as the proteomics approach. For more metabolically active organs such as mesenteric adipose tissue and skeletal muscle, metabolomics experiments should be carried out. Lipidomics is probably not very relevant except in adipose tissue.

All in all, the project so far has met its goal in establishing the MIDY biobank and using the liver as the proof of concept of a multi-layered omics analysis. The point of tissue cross-talk is something that is only just touched upon here and will likely be the main focus in the future of the biobank.

Summary

With the increasing prevalence of diabetes across the world, research into the chronic effects of hyperglycemia is more relevant than ever. While research of single diabetic complications is common, holistic studies of diabetic organisms and tissue crosstalk can reveal novel insights otherwise hidden. This kind of research however is not possible on human patients and as such requires an animal model organism.

The mouse as a model organism, while convenient, is not suitable for translational studies of diabetes due to differences to humans in terms of metabolism and physiology. The pig as a model organism minimizes these differences with largely similar metabolism and physiology to that of humans, and as a large animal, it also provides a large amount of tissue. This makes the pig ideal for use in biobanking, the practice of storing body fluids and tissues over long periods of time and using them when it is suitable. For these reasons, the Munich MIDY Pig biobank was created, hosting over 19000 samples from nine pigs.

The MIDY pig (Mutant INS-gene-induced Diabetes of the Youth) is a transgenic insulin C94Y mutant, causing beta-cell failure due to the misfolding of proinsulin. MIDY pigs are severely diabetic and retain next to none of their beta-cell mass, requiring constant insulin treatment. To mimic a poorly compliant diabetes patient, insulin treatment was minimal. A cohort of 4 MIDY and 5 WT sows were kept under these conditions for 2 years and then euthanized. Tissues were dissected and various body fluids were recovered and then stored in -80 degrees Celsius.

As the most metabolically active organ, the liver was chosen as the first to go through molecular characterization. Using a multi-omic approach in order to carry out an in-depth exploratory study of the organ, tissue samples were used for RNA-Seq, protein mass spectrometry, targeted metabolomics, and shot-gun lipidomics. In addition, western blots of insulin receptor signaling were produced. The data were analyzed both on individual levels and integratively, using bioinformatics tools.

The results of this analysis contained both expected and unexpected findings. As an

organ in gluconeogenic mode, the liver actively metabolized amino acids into glucose and fatty acids into ketone bodies. Several mechanisms for reducing oxidative stress were identified in the MIDY pig including the increased transcription of SLC25A47, a liver unique uncoupling protein, and NADPH production for glutathione regeneration.

An interesting finding was the expression of retinol dehydrogenase 16, RDH16, in MIDY pigs, which was found increased in the liver and later in adipose tissue. Due to the role of RDH16 in all-trans Retinoic Acid (atRA) biosynthesis, and thus in energy regulation through RAR elements, retinoids were measured and it was indeed found that atRA levels were higher in the MIDY pigs.

Unexpectedly parts of the immune system and ECM components were found to be less expressed in MIDY pigs, which is contrary to the higher levels of hepatic inflammation usually seen in diabetic sufferers. Especially Toll-like receptor signaling expression, MHC gene expression, and collagen gene expression were decreased. While MHC2 genes were decreased in expression, no decrease in Kupffer cell volume could be found.

As the first biobank for deep characterization of diabetes in a clinically relevant animal model, the Munich MIDY Pig biobank has already revealed several novel findings from the liver. The unique opportunity for a controlled multiomics investigation on a diabetic liver shows the advantages of using several omics layers when characterizing a disease. The ongoing characterization of adipose and other tissues is expected to contribute to a new, holistic understanding of the molecular changes caused by diabetes.

Zusammenfassung

Bioinformatische Analyse von Multiomics-Daten aus der Munich MIDY Pig Biobank

Mit der zunehmenden Häufigkeit von Diabetes auf der ganzen Welt ist die Erforschung der chronischen Auswirkungen von Hyperglykämie wichtiger denn je. Üblicherweise werden einzelne diabetische Komplikationen erforscht, wogegen holistische Untersuchungen zur Auswirkung von Diabetes auf das gesamte Lebewesen oder auf die Wechselwirkung verschiedener Gewebe neue Einblicke ermöglichen, die ansonsten verborgen bleiben. Diese Art der Untersuchung ist jedoch am menschlichen Patienten nicht möglich und daher wird ein Tiermodell benötigt.

Die Maus als Modellorganismus ist zwar praktisch, eignet sich aber aufgrund der Unterschiede zum Menschen in Bezug auf Stoffwechsel und Physiologie nicht für translationale Untersuchungen von Diabetes. Der Modellorganismus Schwein minimiert diese Unterschiede, weil sein Stoffwechsel und seine Physiologie dem Menschen weitgehend ähnlichen sind. Zudem liefert es als Großtier auch eine große Menge an Gewebe. Dadurch ist das Schwein ideal zum Anlegen einer Biobank geeignet, in der Körperflüssigkeiten und Gewebe über einen langen Zeitraum aufbewahrt und bei Bedarf zur Verfügung gestellt werden. Aus diesen Gründen wurde die "Munich MIDY Pig Biobank“ etabliert, die über 19.000 Proben von neun Schweinen enthält.

Das MIDY-Schwein („Mutant INS-gene-induced Diabetes of the Youth“) trägt eine transgene Insulin-C94Y-Mutante, die aufgrund der Fehlfaltung von Proinsulin zu Betazell-Versagen führt. MIDY-Schweine haben daher eine stark reduzierte Betazell-Masse, sind klinisch diabetisch und benötigen deshalb eine ständige Behandlung mit Insulin. Im verwendeten Modell wurde die Insulinbehandlung minimiert, um einen unzureichend behandelten Diabetespatienten nachzuahmen. Eine Gruppe von 4 MIDY- und 5 WT-Sauen wurd zwei Jahre lang unter diesen Bedingungen gehalten und dann euthanasiert. Gewebeproben wurden entnommen und verschiedene Körperflüssigkeiten wurden gewonnen und

dann jeweils bei -80 Grad Celsius gelagert.

Als metabolisch am stärksten aktivstes Organ wurde als erstes die Leber für die molekulare Charakterisierung ausgewählt. In einem Multiomics-Ansatz wurden Gewebeproben mittels RNA Sequenzierung, Protein-Massenspektrometrie, zielgerichteter Metabolomik und „Shot-Gun"-Lipidomik untersucht, um eine detailierte explorative Studie des Organs durchzuführen. Darüber hinaus wurden Western Blots zur Untersuchung des Insulinrezeptor-Signalwegs durchgeführt. Die erzeugten Daten wurden sowohl separat als auch mit Multiomics-Ansätzen analysiert.

Diese Analysen ergaben sowohl erwartete als auch unerwartete Ergebnisse. Als Organ im glukoneogenen Modus verstoffwechselt die Leber aktiv Aminosäuren zu Glukose und Fettsäuren zu Ketonkörpern. Darüberhinaus wurden mehrere Mechanismen zur Reduzierung des oxidativen Stresses beim MIDY-Schwein identifiziert, darunter die erhöhte Transkription von SLC25A47, eines einzigartigen Entkopplungs-Proteins der Leber, und von Genen der NADPH-Produktion zur Glutathionregeneration.

Ein interessanter Befund bei MIDY-Schweinen war, dass Retinoldehydrogenase 16 (RDH16), in der Leber und auch im Fettgewebe stärker exprimiert wurde. RDH16 ist an der Biosynthese von all-trans Retinsäure (atRA) und damit an der Steuerung des Energiespiegels durch RAR-Elemente beteiligt. Eine Messung der Retinoide ergab, dass die Konzentrationen von atRA und seines Vorläufers Retinal bei den MIDY-Schweinen höher waren.

Ein unerwartetes Ergebnis war, dass Teile des Immunsystems und auch Komponenten der extrazellulären Matrix bei MIDY-Schweinen weniger stark exprimiert waren. Dies steht im Gegensatz dazu, dass bei Diabetikern häufiger entzündliche Veränderungen der Leber beobachtet werden. Insbesondere der Signalweg der Toll-like-Rezeptoren, die Expression von MHC und von Kollagen waren verringert. Während die MHC2-Gene reduziert exprimiert wurden, konnte kein Rückgang des Volumenanteils des Kupffer-Zellen festgestellt werden.

Als die erste Biobank zur tiefen Charakterisierung von Diabetes in einem klinisch relevanten Tiermodell hat die Münchner MIDY Pig Biobank bereits vielversprechende neue Erkenntnisse von Leber geliefert. Die einzigartige Möglichkeit einer kontrollierten Multiomik-Untersuchung an einer diabetischen Leber zeigt die Vorteile der Verwendung mehrerer Omik-Schichten bei der Charakterisierung einer Krankheit. Es ist zu erwarten, dass die laufenden Untersuchungen von Fett- und weiteren Geweben zu einem neuen, ganzheitlichen Verständnis der molekularen Veränderungen beitragen, die Diabetes beim Menschen verursacht.

Appendix A

Reference tables of genes and metabolites

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
ADAMTS17	ADAM metallopeptidase with thrombospondin type 1 motif 17	2.694	9.42E-27
SLC25A47	solute carrier family 25 member 47	1.829	4.35E-19
BMP8B	bone morphogenetic protein 8b	2.018	6.98E-15
SLX4IP	SLX4 interacting protein	1.133	3.82E-14
CBR4	carbonyl reductase 4	1.23	4.33E-12
GPT2	glutamic–pyruvic transaminase 2	1.469	7.3E-12
KLHL3	kelch like family member 3	0.923	5.99E-11
RDH11	retinol dehydrogenase 11	0.997	8.51E-11
GOT1	glutamic-oxaloacetic transaminase 1	1.158	8.72E-11
SHROOM3	shroom family member 3	-1.254	9.12E-11
LOC102165634	ncRNA	1.887	1.06E-10
LOC106505246	ncRNA	1.882	1.27E-10
CTH	cystathionine gamma-lyase	1.271	2.07E-10
HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2	1.795	1.45E-09
GYS2	glycogen synthase 2	0.933	1.49E-09
ELOVL6	ELOVL fatty acid elongase 6	0.942	6.55E-09
ARFGAP3	ADP ribosylation factor GTPase activating protein 3	0.928	1.06E-08
ASS1	argininosuccinate synthase 1	0.796	1.08E-08
MIGA2	mitoguardin 2	1.012	2.96E-08
SEC16A	SEC16 homolog A, endoplasmic reticulum export factor	0.783	4.15E-08
LOC106506349	ncRNA	1.609	0.000000144
IGFALS	insulin like growth factor binding protein acid labile subunit	-0.801	0.000000165
ALDOB	aldolase, fructose-bisphosphate B	0.899	0.000000186
AFMID	arylformamidase	0.884	0.000000229
KIAA1211	KIAA1211	1.565	0.000000229
AGXT	alanine–glyoxylate and serine–pyruvate aminotransferase	1.3	0.000000262
ATE1	arginyltransferase 1	0.572	0.000000262
CYP8B1	cytochrome P450 family 8 subfamily B member 1	1.494	0.000000262
MEGF9	multiple EGF like domains 9	-1.032	0.000000262
ATG2A	autophagy related 2A	0.811	0.000000288
CRP	C-reactive protein	-1.178	0.000000293
SFXN3	sideroflexin 3	1.459	0.000000565
GLS2	glutaminase 2	0.654	0.000000599
SLC4A4	solute carrier family 4 member 4	0.788	0.000000999
HGD	homogentisate 1,2-dioxygenase	0.905	0.00000235
LOC106505250	ncRNA	1.467	0.00000246
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	-1.154	0.00000305
LOC102167599	ncRNA	0.848	0.00000373
LOC110260953	ncRNA	1.452	0.00000373
NFIC	nuclear factor I C	0.632	0.00000431
UBL3	ubiquitin like 3	0.725	0.0000053
MIA3	MIA SH3 domain ER export factor 3	0.556	0.00000672

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
LDHB	lactate dehydrogenase B	1.108	0.00000694
THRB	thyroid hormone receptor beta	-0.68	0.00000763
PSEN2	presenilin 2	0.917	0.00000916
IDH1	isocitrate dehydrogenase (NADP(+)) 1, cytosolic	0.759	0.0000102
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	0.668	0.000013
HAL	histidine ammonia-lyase	0.889	0.0000197
PPARA	peroxisome proliferator activated receptor alpha	0.745	0.0000257
ARG1	arginase 1	1.051	0.0000287
SFRP1	secreted frizzled related protein 1	-0.994	0.0000305
LOC110262053	protein coding	1.293	0.0000306
FAM114A1	family with sequence similarity 114 member A1	0.859	0.0000338
PHACTR2	phosphatase and actin regulator 2	0.575	0.0000362
MTURN	maturin, neural progenitor differentiation regulator homolog	0.933	0.0000378
LOC106506856	protein coding	1.129	0.0000466
BNIP3	BCL2 interacting protein 3	0.493	0.0000502
BMP8A	bone morphogenetic protein 8a	1.319	0.0000509
ITPR2	inositol 1,4,5-trisphosphate receptor type 2	0.557	0.0000639
SLC35E2B	solute carrier family 35 member E2B	0.629	0.0000639
CGNL1	cingulin like 1	0.629	0.0000723
TMSB4X	thymosin beta 4 X-linked	-0.512	0.0000753
GIPC2	GIPC PDZ domain containing family member 2	-0.97	0.0000795
NDFIP2	Nedd4 family interacting protein 2	0.775	0.0000831
PDE4B	phosphodiesterase 4B	-0.827	0.0000831
MOB3B	MOB kinase activator 3B	-0.937	0.0000868
HNF4G	hepatocyte nuclear factor 4 gamma	0.831	0.0000901
TBC1D8B	TBC1 domain family member 8B	0.729	0.0000935
TMEM263	transmembrane protein 263	0.651	0.0000935
AASS	aminoadipate-semialdehyde synthase	0.775	0.0000939
AMPD3	adenosine monophosphate deaminase 3	-0.775	0.0000975
SDS	serine dehydratase	1.266	0.000103
GART	GAR transformylase	0.67	0.000114
LOC110260020	pseudo gene	1.071	0.000124
SGK1	serum/glucocorticoid regulated kinase 1	-0.882	0.000124
AHCY	adenosylhomocysteinase	0.599	0.000128
SLC13A5	solute carrier family 13 member 5	0.973	0.000139
VIL1	villin 1	-1.05	0.000181
EXTL2	exostosin like glycosyltransferase 2	0.739	0.000209
DMTN	dematin actin binding protein	0.844	0.00021
SETBP1	SET binding protein 1	-0.765	0.000235
TMEM181	transmembrane protein 181	0.836	0.000249
LOC110258394	immunoglobulin	-1.191	0.000251
ERBB3	erb-b2 receptor tyrosine kinase 3	0.654	0.000271
FBP1	fructose-bisphosphatase 1	0.797	0.00028
NREP	neuronal regeneration related protein	-1.061	0.000311
TMEM245	transmembrane protein 245	0.442	0.000331
OAT	ornithine aminotransferase	0.932	0.000334
KYAT1	kynurenine aminotransferase 1	0.956	0.000342
EEA1	early endosome antigen 1	0.483	0.00038
MXI1	MAX interactor 1, dimerization protein	0.663	0.00038
TRAM2	translocation associated membrane protein 2	0.658	0.00038
LOC110256685	ncRNA	-0.812	0.000387
SEC23A	Sec23 homolog A, coat complex II component	0.694	0.000401
PMM1	phosphomannomutase 1	1.153	0.000404
LURAP1L	leucine rich adaptor protein 1 like	-0.716	0.00043
EAF1	ELL associated factor 1	0.767	0.00047
NT5E	5'-nucleotidase ecto	1.022	0.000474
LOC110255180	ncRNA	-0.923	0.000511
TBX3	T-box 3	0.788	0.000511
DPH6	diphthamine biosynthesis 6	-0.712	0.000519
PEG3	paternally expressed 3	0.71	0.000519
CBR1	carbonyl reductase 1	-1.099	0.000528
ARHGAP35	Rho GTPase activating protein 35	0.443	0.000529
MGP	matrix Gla protein	-0.815	0.000544
PIGR	polymeric immunoglobulin receptor	-0.9	0.000544
SULT2A1	sulfotransferase family 2A member 1	0.77	0.000565
CYP2C34	cytochrome P450 2C34	0.988	0.000607
LDHA	lactate dehydrogenase A	0.69	0.000609
LOC110256684	ncRNA	-0.935	0.000622
HDAC4	histone deacetylase 4	1.001	0.000622
SLC25A42	solute carrier family 25 member 42	0.875	0.000661

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
SLA-DQA1	MHC class II histocompatibility antigen SLA-DQA	-0.681	0.000685
SLCO2A1	solute carrier organic anion transporter family member 2A1	1.134	0.000688
MAG	myelin associated glycoprotein	1.074	0.000703
GCA	grancalcin	0.73	0.000729
LOC110260079	protein coding	0.803	0.000744
NEB	nebulin	1.145	0.000744
ABHD15	abhydrolase domain containing 15	0.697	0.000751
ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	-0.627	0.000774
HIGD1A	HIG1 hypoxia inducible domain family member 1A	0.672	0.000774
GNMT	glycine N-methyltransferase	0.772	0.000808
SLA-DQB1	SLA-DQ beta1 domain	-0.598	0.000854
ACTN1	actinin alpha 1	-0.461	0.000954
ADD3	adducin 3	-0.517	0.00105
RHPN1	rhophilin Rho GTPase binding protein 1	1.077	0.00106
MFAP3L	microfibril associated protein 3 like	0.771	0.00106
TAGAP	T cell activation RhoGTPase activating protein	-0.882	0.00108
CYP26A1	cytochrome P450 family 26 subfamily A member 1	-0.903	0.0011
PPTC7	PTC7 protein phosphatase homolog	-0.564	0.00111
CORO1A	coronin 1A	-0.715	0.00119
MTSS1	MTSS I-BAR domain containing 1	-0.569	0.0012
MAP6D1	MAP6 domain containing 1	0.903	0.00121
LOC110261188	ncRNA	0.988	0.00134
NRP1	neuropilin 1	-0.455	0.00137
MAPK9	mitogen-activated protein kinase 9	0.53	0.00143
TUFT1	tuftelin 1	0.871	0.00152
SH3YL1	SH3 and SYLF domain containing 1	0.872	0.00156
CCDC13	coiled-coil domain containing 13	1.098	0.00158
P2RY4	pyrimidinergic receptor P2Y4	1.096	0.00162
PCMTD1	protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1	0.354	0.0017
CROT	carnitine O-octanoyltransferase	0.582	0.00176
LOC102161293	ncRNA	1.046	0.00191
CPS1	carbamoyl-phosphate synthase 1	0.567	0.00196
PRELP	proline and arginine rich end leucine rich repeat protein	-0.692	0.00199
AKAP7	A-kinase anchoring protein 7	0.629	0.00201
DAPK2	death associated protein kinase 2	-0.797	0.00201
B3GNT3	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3	-0.837	0.0021
MYBPC1	myosin binding protein C, slow type	1.079	0.0021
SEPHS1	selenophosphate synthetase 1	0.554	0.0021
SLC22A7	solute carrier family 22 member 7	-0.785	0.00229
TMEM38A	transmembrane protein 38A	0.746	0.00229
CIART	circadian associated repressor of transcription	-1.037	0.00235
CREB3L3	cAMP responsive element binding protein 3 like 3	0.652	0.00238
GPAM	glycerol-3-phosphate acyltransferase, mitochondrial	-0.797	0.00238
CD74	CD74 molecule	-0.662	0.00251
LRMDA	leucine rich melanocyte differentiation associated	-0.68	0.00251
NT5DC3	5'-nucleotidase domain containing 3	0.837	0.00269
GSTM3	glutathione S-transferase mu 3	1.057	0.00277
ADCY9	adenylate cyclase 9	0.519	0.00281
KITLG	KIT ligand	-0.727	0.00295
COL1A2	collagen type I alpha 2 chain	-0.691	0.00297
IREB2	iron responsive element binding protein 2	0.538	0.00303
CERS6	ceramide synthase 6	-0.615	0.00316
TLE3	TLE family member 3, transcriptional corepressor	-0.613	0.00316
TKFC	triokinase and FMN cyclase	0.551	0.00316
KIAA0040	KIAA0040	-0.636	0.0032
PDK4	pyruvate dehydrogenase kinase 4	0.626	0.0032
INSR	insulin receptor	0.401	0.00324
EMCN	endomucin	-0.653	0.00325
KRT8	keratin 8	-0.584	0.00325
LOC102163364	ncRNA	0.7	0.00325
RASL11B	RAS like family 11 member B	0.852	0.00325
ZBTB7B	zinc finger and BTB domain containing 7B	0.694	0.00325
APOA5	apolipoprotein A5	0.827	0.00338
DPP4	dipeptidyl peptidase 4	0.381	0.00341
COL1A1	collagen type I alpha 1 chain	-0.824	0.00347
NQO2	N-ribosyldihydronicotinamide:quinone reductase 2	0.492	0.00351
SMAP2	small ArfGAP2	0.608	0.00351
TTC38	tetratricopeptide repeat domain 38	0.49	0.00351
ZNF385B	zinc finger protein 385B	-0.748	0.00351
SEC24D	SEC24 homolog D, COPII coat complex component	0.573	0.00361

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
DAPK1	death associated protein kinase 1	-0.578	0.00363
RNF217	ring finger protein 217	0.633	0.00364
SAR1B	secretion associated Ras related GTPase 1B	0.473	0.00366
TPM2	tropomyosin 2	-0.684	0.00367
TFRC	transferrin receptor	-0.872	0.00368
CD2	CD2 molecule	-0.753	0.00386
OTC	ornithine carbamoyltransferase	0.408	0.00389
TCEA3	transcription elongation factor A3	-0.531	0.00393
TVP23B	trans-golgi network vesicle protein 23 homolog B	0.526	0.00395
SYTL5	synaptotagmin like 5	-0.978	0.00427
GADD45B	growth arrest and DNA damage inducible beta	-0.784	0.00436
LOC110258705	immunoglobulin	-1.017	0.00451
EIF4EBP2	eukaryotic translation initiation factor 4E binding protein 2	0.501	0.00467
FAM114A2	family with sequence similarity 114 member A2	0.544	0.00475
ZNF622	zinc finger protein 622	0.499	0.00475
LOC100624435	cytochrome P450 2C42-like	-0.912	0.00496
ADIPOR2	adiponectin receptor 2	0.489	0.00496
FMO5	flavin containing monooxygenase 5	0.481	0.00505
NTRK3	neurotrophic receptor tyrosine kinase 3	-0.943	0.00541
LGR4	leucine rich repeat containing G protein-coupled receptor 4	0.488	0.00558
PPP2R5A	protein phosphatase 2 regulatory subunit B'alpha	0.454	0.00567
ACAD11	acyl-CoA dehydrogenase family member 11	0.478	0.00568
APOA4	apolipoprotein A4	0.983	0.00568
CDC14B	cell division cycle 14B	0.57	0.00568
PPP2R1B	protein phosphatase 2 scaffold subunit Abeta	0.44	0.00568
ADRB2	adrenoceptor beta 2	-0.698	0.00572
ARHGDIB	Rho GDP dissociation inhibitor beta	-0.627	0.00572
NFE2L1	nuclear factor, erythroid 2 like 1	0.464	0.00572
LOC102163250	ncRNA	-0.963	0.00576
FOXN2	forkhead box N2	0.816	0.00617
AHR	aryl hydrocarbon receptor	-0.607	0.00622
LOC102167466	immunoglobulin	-0.899	0.00622
RUNX1T1	RUNX1 translocation partner 1	-0.538	0.00622
ETFRF1	electron transfer flavoprotein regulatory factor 1	0.594	0.00637
LOC106506356	ncRNA	0.791	0.00637
SMARCA1	ATP-dependent helicase SMARCA1	0.394	0.00659
NNT	nicotinamide nucleotide transhydrogenase	0.613	0.0066
SLC13A3	solute carrier family 13 member 3	-0.679	0.0066
LAD1	ladinin 1	0.875	0.00667
MCFD2	multiple coagulation factor deficiency 2	0.515	0.00674
FRY	FRY microtubule binding protein	-0.599	0.00685
NPC1L1	NPC1 like intracellular cholesterol transporter 1	0.827	0.00685
ID2	inhibitor of DNA binding 2	-0.548	0.00693
MUC16	mucin-16	0.852	0.00702
MASP2	mannan binding lectin serine peptidase 2	-0.66	0.00713
PDCL3	phosducin like 3	-0.526	0.00728
SIK2	salt inducible kinase 2	0.592	0.00733
ATPAF1	ATP synthase mitochondrial F1 complex assembly factor 1	0.449	0.00751
LOC110260960	ncRNA	0.894	0.00762
HACL1	2-hydroxyacyl-CoA lyase 1	0.459	0.00763
C2CD2	C2 calcium dependent domain containing 2	0.621	0.00786
CLUH	clustered mitochondria homolog	0.56	0.00849
AEBP1	AE binding protein 1	-0.761	0.00859
CD59	CD59 molecule (CD59 blood group)	-0.462	0.00867
RDH16	retinol dehydrogenase 16	0.642	0.00872
ISCA1	iron-sulfur cluster assembly 1	0.4	0.00879
LOC110256694	ncRNA	-0.964	0.00897
LOC110260756	ncRNA	-0.746	0.00897
ACTG1	actin gamma 1	-0.526	0.00909
LOC110262159	ncRNA	-0.901	0.00909
EIF5	eukaryotic translation initiation factor 5	0.339	0.00921
SEPSECS	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase	0.601	0.00921
SLC16A7	solute carrier family 16 member 7	-0.47	0.00956
CYP2C32	cytochrome P450 2C32	-0.792	0.00963
ADGRA2	adhesion G protein-coupled receptor A2	-0.627	0.00972
PPA1	pyrophosphatase (inorganic) 1	0.512	0.00982
FGD5	FYVE, RhoGEF and PH domain containing 5	-0.479	0.00996
BFAR	bifunctional apoptosis regulator	0.625	0.00999
GSDMD	gasdermin D	0.592	0.00999
MKKS	McKusick-Kaufman syndrome	0.574	0.00999

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
CFD	complement factor D	-0.595	0.01
RAC2	Rac family small GTPase 2	-0.595	0.0103
ACE2	angiotensin I converting enzyme 2	0.776	0.0103
PLSCR1	phospholipid scramblase 1	-0.633	0.0104
ITPKA	inositol-trisphosphate 3-kinase A	0.716	0.0107
LOC110256479	carbonyl reductase [NADPH] 1-like	0.929	0.0107
BBS12	Bardet-Biedl syndrome 12	-0.703	0.0108
POMT2	protein O-mannosyltransferase 2	0.828	0.0108
SAA3	serum amyloid A3, pseudogene	-0.922	0.0109
DHTKD1	dehydrogenase E1 and transketolase domain containing 1	0.548	0.011
BLNK	B cell linker	-0.862	0.0112
LOC106504205	ncRNA	0.782	0.0113
SEPT9	septin 9	-0.429	0.0114
LOC110259119	stabilin-2-like	-0.665	0.0117
PLBD2	phospholipase B domain containing 2	0.43	0.0117
PLB1	phospholipase B1	0.906	0.0118
ADAMTSL4	ADAMTS like 4	0.646	0.0121
AQP7	aquaporin 7	0.929	0.0122
TSPAN33	tetraspanin 33	-0.919	0.0122
NFIC	nuclear factor I C	0.452	0.0123
NAXD	NAD(P)HX dehydratase	0.412	0.0129
PLCXD3	phosphatidylinositol specific phospholipase C X domain containing 3	-0.904	0.0131
GPCPD1	glycerophosphocholine phosphodiesterase 1	-0.692	0.0133
LOC110258666	immunoglobulin	-0.901	0.0133
KIF21A	kinesin family member 21A	0.603	0.0134
SPRYD3	SPRY domain containing 3	0.512	0.0135
TECTB	tectorin beta	-0.928	0.0137
DDAH1	dimethylarginine dimethylaminohydrolase 1	0.599	0.0141
NQO1	NAD(P)H quinone dehydrogenase 1	0.705	0.0141
LOC106509766	immunoglobulin	-0.905	0.0141
EVI2B	ecotropic viral integration site 2B	-0.594	0.0144
SLA-2	MHC class I antigen 2	-0.696	0.0144
STMN1	stathmin 1	-0.684	0.0145
TMEM86B	transmembrane protein 86B	-0.535	0.0148
SULT1C2	sulfotransferase 1C2	0.735	0.0149
MEGF10	multiple EGF like domains 10	0.852	0.015
ITGA8	integrin subunit alpha 8	-0.883	0.0151
HDAC6	histone deacetylase 6	0.411	0.0151
LOC102167708	ncRNA	-0.86	0.0155
TUBA1A	tubulin alpha 1a	-0.76	0.0155
LEPR	leptin receptor	0.909	0.0156
TTC39C	tetratricopeptide repeat domain 39C	-0.375	0.0162
THEMIS	thymocyte selection associated	-0.872	0.0165
CXCL16	C-X-C motif chemokine ligand 16	-0.39	0.0167
PKP4	plakophilin 4	-0.445	0.0168
SLC16A10	solute carrier family 16 member 10	0.514	0.0168
SPX	spexin hormone	0.765	0.0168
STK17B	serine/threonine kinase 17b	-0.544	0.0168
OGDH	oxoglutarate dehydrogenase	0.474	0.017
NFXL1	nuclear transcription factor, X-box binding like 1	0.537	0.0171
ACAA1	acetyl-CoA acyltransferase 1	0.518	0.0174
SLC15A2	solute carrier family 15 member 2	0.354	0.0174
AOC2	amine oxidase copper containing 2	0.899	0.0176
NMRAL1	NmrA like redox sensor 1	0.67	0.0177
ITGB3	integrin subunit beta 3	-0.701	0.018
LOC110255435	ncRNA	0.495	0.0181
SFMBT2	Scm like with four mbt domains 2	0.497	0.0181
C6H1orf109	protein coding	0.487	0.019
TCF7	transcription factor 7	0.621	0.0191
SLC25A1	solute carrier family 25 member 1	0.55	0.0191
HAAO	3-hydroxyanthranilate 3,4-dioxygenase	0.445	0.0193
INHBC	inhibin subunit beta C	-0.643	0.0193
LOC110262160	ncRNA	-0.848	0.0193
PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	0.408	0.0193
RORC	RAR related orphan receptor C	0.62	0.0195
PGM2	phosphoglucomutase 2	-0.622	0.0199
HIST1H2AH	histone cluster 1 H2A family member H	-0.881	0.0202
GPC5	glypican 5	-0.807	0.0204
ACTB	actin beta	-0.445	0.0206
CD84	CD84 molecule	-0.608	0.0206

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
CEP192	centrosomal protein 192	0.373	0.0208
LPIN3	lipin 3	0.742	0.0208
UTP11	UTP11 small subunit processome component	0.584	0.0208
SUOX	sulfite oxidase	0.489	0.021
AMPD2	adenosine monophosphate deaminase 2	0.528	0.0211
ACOX2	acyl-CoA oxidase 2	0.49	0.0212
TOR1AIP2	torsin 1A interacting protein 2	0.469	0.0213
DLL3	delta like canonical Notch ligand 3	0.8	0.0214
ENTPD5	ectonucleoside triphosphate diphosphohydrolase 5 (inactive)	0.625	0.0214
OSTM1	osteoclastogenesis associated transmembrane protein 1	0.48	0.0214
SIT1	signaling threshold regulating transmembrane adaptor 1	-0.88	0.0214
OTULIN	OTU deubiquitinase with linear linkage specificity	0.436	0.0215
LYPD6	LY6/PLAUR domain containing 6	0.787	0.0217
GTF2A1L	general transcription factor IIA subunit 1 like	-0.764	0.0217
PREX1	phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 1	-0.485	0.022
CYR61	cellular communication network factor 1	-0.624	0.0223
LOC110257744	ncRNA	-0.868	0.0223
PKHD1	PKHD1 ciliary IPT domain containing fibrocystin/polyductin	-0.618	0.0226
BNC2	basonuclin 2	-0.656	0.0226
FAM84A	LRAT domain containing 1	0.863	0.0228
INTS3	integrator complex subunit 3	0.385	0.0228
SP100	nuclear body protein SP140-like protein	-0.437	0.0231
PAQR7	progestin and adipoQ receptor family member 7	-0.799	0.0231
LOC100524773	low-density lipoprotein receptor-related protein 2-like	-0.876	0.0232
FSTL1	follistatin like 1	-0.443	0.0236
ATP23	ATP23 metallopeptidase and ATP synthase assembly factor homolog	-0.635	0.0237
COPB1	coatomer protein complex subunit beta 1	0.474	0.0237
RNF152	ring finger protein 152	0.54	0.0237
WFS1	wolframin ER transmembrane glycoprotein	0.687	0.0237
ZFP36L2	ZFP36 ring finger protein like 2	-0.478	0.0237
SLC37A4	solute carrier family 37 member 4	0.613	0.0243
SYDE2	synapse defective Rho GTPase homolog 2	-0.432	0.0248
PSKH1	protein serine kinase H1	0.389	0.0248
PPP1R15A	protein phosphatase 1 regulatory subunit 15A	-0.619	0.0255
RFX5	regulatory factor X5	-0.477	0.0256
CLPX	caseinolytic mitochondrial matrix peptidase chaperone subunit	0.474	0.0263
IRAK4	interleukin 1 receptor associated kinase 4	0.609	0.0263
NOTCH2	notch receptor 2	-0.332	0.0263
C6H19orf12	protein coding	0.749	0.0265
DOLPP1	dolichyldiphosphatase 1	0.563	0.0265
LOC106505550	ncRNA	0.746	0.0265
LOC110262128	ncRNA	-0.852	0.0265
SLC27A1	solute carrier family 27 member 1	0.502	0.0265
CD58	CD58 molecule	0.502	0.0266
FBLN2	fibulin 2	-0.613	0.0267
CHRDL1	chordin like 1	-0.772	0.0267
HIST1H2BM	histone cluster 1 H2B family member M	-0.762	0.0269
COTL1	coactosin like F-actin binding protein 1	-0.532	0.027
SLC39A3	solute carrier family 39 member 3	-0.593	0.0273
TGM3	transglutaminase 3	-0.833	0.0273
CREB3L1	cAMP responsive element binding protein 3 like 1	-0.658	0.0273
RASSF9	Ras association domain family member 9	-0.692	0.0273
ABCD3	ATP binding cassette subfamily D member 3	0.651	0.0273
FAM160B2	family with sequence similarity 160 member B2	0.455	0.0274
DPT	dermatopontin	-0.671	0.0279
OGN	osteoglycin	-0.702	0.0279
ILDR2	immunoglobulin like domain containing receptor 2	-0.348	0.0281
WBP1L	WW domain binding protein 1 like	0.431	0.0284
USP12	ubiquitin specific peptidase 12	0.506	0.0289
SLC26A1	solute carrier family 26 member 1	-0.558	0.029
ALDH7A1	aldehyde dehydrogenase 7 family member A1	0.511	0.029
CLDN2	claudin 2	-0.751	0.029
ING1	inhibitor of growth family member 1	0.485	0.029
LOC100623670	protein coding	-0.85	0.029
GGCX	gamma-glutamyl carboxylase	0.343	0.0296
CHD4	chromodomain helicase DNA binding protein 4	0.365	0.0299
RABEP1	rabaptin, RAB GTPase binding effector protein 1	0.318	0.0299
TIMP3	TIMP metallopeptidase inhibitor 3	-0.336	0.0299
LYPD6B	LY6/PLAUR domain containing 6B	0.851	0.03
CA2	carbonic anhydrase 2	-0.464	0.03

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
PDGFC	platelet derived growth factor C	-0.448	0.03
CRISPLD2	cysteine rich secretory protein LCCL domain containing 2	-0.633	0.0301
SLC37A2	solute carrier family 37 member 2	-0.757	0.0304
MRC2	mannose receptor C type 2	-0.812	0.0307
NDFIP1	Nedd4 family interacting protein 1	0.416	0.0307
EXD3	exonuclease 3'-5' domain containing 3	-0.633	0.031
CBS	cystathionine-beta-synthase	0.488	0.0314
PLIN5	perilipin 5	0.547	0.0314
LOC100739791	ncRNA	0.747	0.0318
GLUL	glutamate-ammonia ligase	-0.714	0.0321
LOC100514282	prolyl-tRNA synthetase associated domain-containing protein 1	0.627	0.0324
SUCO	SUN domain containing ossification factor	0.45	0.0327
PALM2	paralemmin 2	-0.594	0.0327
CCDC6	coiled-coil domain containing 6	0.461	0.0328
CCDC80	coiled-coil domain containing 80	-0.653	0.0328
CDK18	cyclin dependent kinase 18	0.518	0.0328
CHMP3	charged multivesicular body protein 3	-0.595	0.0328
DCTD	dCMP deaminase	0.492	0.0328
FLNA	filamin A	-0.654	0.0328
ORM1	orosomucoid 1	-0.516	0.0328
SH3BP5	SH3 domain binding protein 5	0.545	0.0328
SLC2A3	solute carrier family 2 member 3	-0.62	0.0328
UBXN2A	UBX domain protein 2A	0.416	0.0328
CYSLTR1	cysteinyl leukotriene receptor 1	-0.745	0.033
DBP	D-box binding PAR bZIP transcription factor	-0.655	0.033
TK2	thymidine kinase 2	0.516	0.033
SGMS2	sphingomyelin synthase 2	0.624	0.0333
LOC110261663	histone H2B type 1	-0.672	0.0333
NLN	neurolysin	0.423	0.0335
OLR1	oxidized low density lipoprotein receptor 1	-0.636	0.0335
SEC31A	SEC31 homolog A, COPII coat complex component	0.358	0.0335
ELMO1	engulfment and cell motility 1	0.401	0.0336
LOC110258332	ncRNA	0.837	0.0336
LOC110256392	ncRNA	0.837	0.0337
MARCO	macrophage receptor with collagenous structure	0.724	0.0342
SMPDL3A	sphingomyelin phosphodiesterase acid like 3A	-0.634	0.0342
RMND5A	required for meiotic nuclear division 5 homolog A	0.37	0.0343
CLPB	ClpB homolog, mitochondrial AAA ATPase chaperonin	0.482	0.0344
PLEC	plectin	0.293	0.0346
RNF14	ring finger protein 14	0.47	0.0348
TMTC2	transmembrane and tetratricopeptide repeat containing 2	0.431	0.0348
EPHA3	EPH receptor A3	-0.833	0.0349
R3HCC1L	R3H domain and coiled-coil containing 1 like	0.464	0.035
C12H17orf75	protein coding	0.468	0.035
FAM131C	family with sequence similarity 131 member C	0.778	0.035
LOC102160793	ncRNA	-0.828	0.035
SLA-DRA	MHC class II DR-alpha	-0.585	0.035
ATP1B1	ATPase Na+/K+ transporting subunit beta 1	-0.428	0.035
CLMN	calmin	0.565	0.035
UBE2QL1	ubiquitin conjugating enzyme E2 Q family like 1	-0.831	0.0352
SLC43A3	solute carrier family 43 member 3	-0.809	0.0353
FAM91A1	family with sequence similarity 91 member A1	0.399	0.0353
PXN	paxillin	0.419	0.0355
TMEM176A	transmembrane protein 176A	0.81	0.0355
BPIFC	BPI fold containing family C	0.773	0.0356
LOC102157807	N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase-like	-0.75	0.0357
CA14	carbonic anhydrase 14	-0.72	0.0359
MMP7	matrix metallopeptidase 7	-0.737	0.0362
TRAPPC8	trafficking protein particle complex 8	0.405	0.0362
P4HA1	prolyl 4-hydroxylase subunit alpha 1	0.595	0.0363
COPA	coatomer protein complex subunit alpha	0.444	0.0366
RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	0.577	0.0368
PRKCA	protein kinase C alpha	0.603	0.037
FLT1	fms related tyrosine kinase 1	-0.405	0.0374
WRN	Werner syndrome RecQ like helicase	0.453	0.0374
ASB13	ankyrin repeat and SOCS box containing 13	0.523	0.0374
BASP1	brain abundant membrane attached signal protein 1	-0.499	0.0374
MADCAM1	mucosal vascular addressin cell adhesion molecule 1	-0.707	0.0374
LOC100623255	UDP-glucuronosyltransferase 2B31	0.338	0.0381
RPS2	ribosomal protein S2	0.417	0.0382

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
ALPK2	alpha kinase 2	0.789	0.0386
LOC102165887	ncRNA	-0.811	0.0386
PARVG	parvin gamma	-0.656	0.0386
POR	cytochrome p450 oxidoreductase	-0.429	0.039
ANKRD1	ankyrin repeat domain 1	-0.779	0.0392
SULT1B1	sulfotransferase family cytosolic 1B member 1	0.774	0.0392
TFG	trafficking from ER to golgi regulator	0.477	0.0393
GPR52	G protein-coupled receptor 52	0.542	0.0393
SMIM14	small integral membrane protein 14	0.377	0.0399
FCGR3A	Fc fragment of IgG receptor IIIa	-0.433	0.0399
COL6A2	collagen type VI alpha 2 chain	-0.546	0.0401
LNX2	ligand of numb-protein X 2	0.564	0.0402
PNPLA7	patatin like phospholipase domain containing 7	0.407	0.0402
LOC110262161	ncRNA	-0.734	0.0411
NLRX1	NLR family member X1	0.666	0.0414
LRPPRC	leucine rich pentatricopeptide repeat containing	0.314	0.0418
KLHL25	kelch like family member 25	0.714	0.0418
LOC110260620	ncRNA	0.804	0.042
C14H1orf198	protein coding	0.473	0.0421
LOC110256243	ncRNA	0.812	0.0421
C7H6orf62	protein coding	0.335	0.0421
LOC110262064	ncRNA	-0.59	0.0421
SACS	sacsin molecular chaperone	0.328	0.0424
EIF5A2	eukaryotic translation initiation factor 5A2	0.638	0.0428
LOC110260355	ncRNA	0.603	0.0429
ALDH1B1	aldehyde dehydrogenase 1 family member B1	-0.634	0.043
ARHGEF12	Rho guanine nucleotide exchange factor 12	0.298	0.043
BAIAP2L1	BAI1 associated protein 2 like 1	0.48	0.043
CLIP1	CAP-Gly domain containing linker protein 1	-0.291	0.043
KLF9	Kruppel like factor 9	0.441	0.043
PPIF	peptidylprolyl isomerase F	0.557	0.043
GALNT2	polypeptide N-acetylgalactosaminyltransferase 2	0.543	0.0434
HSD3B1	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	-0.331	0.0434
GSE1	Gse1 coiled-coil protein	-0.599	0.0435
YTHDC2	YTH domain containing 2	-0.431	0.0435
B3GALT2	beta-1,3-galactosyltransferase 2	-0.806	0.0442
TET2	tet methylcytosine dioxygenase 2	-0.482	0.0449
C15H2orf88	protein coding	0.726	0.045
LOC106504343	ncRNA	-0.795	0.045
SEC23IP	SEC23 interacting protein	0.349	0.0451
APOBR	apolipoprotein B receptor	-0.572	0.0453
BCL11A	BAF chromatin remodeling complex subunit BCL11A	-0.76	0.0453
HSD17B6	17-beta-hydroxysteroid dehydrogenase type 6	-0.519	0.0453
DHX40	DEAH-box helicase 40	0.397	0.0453
COPZ2	coatomer protein complex subunit zeta 2	0.462	0.0453
LOC100153139	HLA class II histocompatibility antigen, DRB1-4 beta chain-like	-0.756	0.0453
LOC110256556	ncRNA	0.798	0.0456
CFAP100	cilia and flagella associated protein 100	0.799	0.0465
PMM2	phosphomannomutase 2	0.494	0.0468
DNM3	dynamin 3	-0.468	0.047
KIAA1456	tRNA methyltransferase 9B (putative)	-0.62	0.047
PARP9	poly(ADP-ribose) polymerase family member 9	-0.575	0.047
PVRIG	PVR related immunoglobulin domain containing	-0.79	0.047
STBD1	starch binding domain 1	-0.508	0.0471
STXBP5L	syntaxin binding protein 5 like	0.777	0.0471
LMAN1	lectin, mannose binding 1	0.448	0.0472
PFKFB1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	-0.404	0.0472
RNF144A	ring finger protein 144A	0.508	0.0472
SLC25A37	solute carrier family 25 member 37	-0.533	0.0472
PIM1	Pim-1 proto-oncogene, serine/threonine kinase	0.68	0.0474
RBKS	ribokinase	0.563	0.0474
RNLS	renalase, FAD dependent amine oxidase	0.536	0.0474
TRIP11	thyroid hormone receptor interactor 11	0.395	0.0481
FNDC4	fibronectin type III domain containing 4	0.764	0.0485
IDO2	indoleamine 2,3-dioxygenase 2	-0.374	0.0485
ORAI3	ORAI calcium release-activated calcium modulator 3	0.571	0.0485
RHPN2	rhophilin Rho GTPase binding protein 2	-0.727	0.0485
IMPA1	inositol monophosphatase 1	0.616	0.0486
LEAP2	liver enriched antimicrobial peptide 2	-0.561	0.0487
RNF43	ring finger protein 43	-0.692	0.0491

Gene symbol	Full gene name	log2FoldChange	adjusted p-value
RBM3	RNA binding motif protein 3	-0.458	0.0494

Table A.1: List of all genes significantly diferentially expressed in the MIDY liver sorted by their p-value. Genes described in text are mentioned in the abbreviation section.

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	LOD [µM]	Mean above LOD
C0	6.6577	7.7013	-0.1852	0.3008	2.3709	TRUE
C2	1.0732	1.0453	0.0338	0.9342	0.1189	TRUE
C3	0.1072	0.1098	-0.0305	0.8701	0.0716	TRUE
C3-OH	0.0224	0.024	-0.0871	0.4404	0.0513	FALSE
C3:1	0.0144	0.0154	-0.0829	0.6249	0.036	FALSE
C3-DC (C4-OH)	0.0742	0.0773	-0.0525	0.6319	0.0362	TRUE
C4	0.0697	0.0662	0.0675	0.3733	0.0476	TRUE
C4:1	0.0626	0.036	0.7286	0.0058	0.0367	TRUE
C5	0.0336	0.0345	-0.0344	0.8499	0.0627	FALSE
C5-OH (C3-DC-M)	0.0474	0.0415	0.1718	0.3323	0.0697	FALSE
C5-M-DC	0.028	0.0284	-0.0204	0.8854	0.0552	FALSE
C5-DC (C6-OH)	0.015	0.0177	-0.2091	0.1794	0.0351	FALSE
C5:1	0.0294	0.0285	0.0403	0.6264	0.0627	FALSE
C5:1-DC	0.0134	0.0113	0.2195	0.1703	0.0233	FALSE
C6 (C4:1-DC)	0.0436	0.0464	-0.0812	0.4755	0.1179	FALSE
C6:1	0.0258	0.0277	-0.0894	0.3077	0.0599	FALSE
C7-DC	0.0157	0.0155	0.0157	0.9425	0.0427	FALSE
C8	0.0681	0.0656	0.0479	0.703	0.1321	FALSE
C9	0.0212	0.0213	-0.0029	0.9742	0.0355	FALSE
C10	0.1725	0.0934	0.8103	0.0041	0.095	TRUE
C10:1	0.0283	0.0289	-0.0276	0.8113	0.0869	FALSE
C10:2	0.0428	0.0402	0.0783	0.2849	0.0508	FALSE
C12	0.0432	0.0372	0.1931	0.2251	0.0928	FALSE
C12-DC	0.095	0.0965	-0.02	0.6833	0.3288	FALSE
C12:1	0.036	0.0337	0.0865	0.5555	0.0861	FALSE
C14	0.0134	0.0094	0.4619	0.0563	0.0252	FALSE
C14:1	0.0167	0.0133	0.2926	0.1107	0.01	TRUE
C14:1-OH	0.006	0.0054	0.1336	0.3714	0.0165	FALSE
C14:2	0.005	0.0043	0.2038	0.4329	0.0111	FALSE
C14:2-OH	0.0074	0.0061	0.2435	0.1715	0.0151	FALSE
C16	0.0239	0.0167	0.464	0.186	0.014	TRUE
C16-OH	0.0442	0.0381	0.1897	0.3672	0.0104	TRUE
C16:1	0.0312	0.0295	0.0721	0.6149	0.0838	FALSE
C16:1-OH	0.0049	0.0049	-0.0043	0.989	0.0116	FALSE
C16:2	0.0055	0.0042	0.3306	0.0533	0.0139	FALSE
C16:2-OH	0.0119	0.0121	-0.0233	0.8257	0.0322	FALSE
C18	0.0166	0.0099	0.6787	0.0074	0.0111	TRUE
C18:1	0.0235	0.0146	0.6309	0.2123	0.0172	TRUE
C18:1-OH	0.0102	0.0097	0.0701	0.4933	0.0217	FALSE
C18:2	0.0079	0.0053	0.5034	0.1103	0.0087	FALSE
Ala	325.6455	348.372	-0.0862	0.6355	5.9335	TRUE
Arg	208.7214	210.6499	-0.0118	0.9143	0.5	TRUE
Asn	29.4186	27.7052	0.0772	0.5175	1.5	TRUE
Asp	19.8956	23.2538	-0.1982	0.3153	1.5	TRUE
Cit	91.023	98.0928	-0.0955	0.508	1	TRUE
Gln	399.1926	446.3814	-0.1424	0.1346	1.5	TRUE
Glu	132.2972	142.6014	-0.0958	0.6193	2	TRUE
Gly	634.9126	897.1689	-0.4341	0.052	0.5	TRUE
His	93.724	86.2438	0.1071	0.4696	0.5	TRUE
Ile	194.4627	140.5584	0.4232	0.0177	1.5739	TRUE
Leu	276.4739	195.696	0.4509	0.0018	1.5	TRUE
Lys	157.0419	132.7425	0.2175	0.0315	0.5	TRUE
Met	43.4574	42.7889	0.0199	0.8946	0.1	TRUE
Orn	61.2988	51.8737	0.216	0.1482	0.5	TRUE
Phe	92.8511	78.172	0.2227	0.0161	0.1	TRUE
Pro	260.7201	241.8348	0.0968	0.2368	1	TRUE
Ser	111.0086	111.0663	-0.0007	0.9945	0.912	TRUE
Thr	156.613	134.9053	0.1929	0.2449	0.5	TRUE
Trp	77.2174	63.4342	0.2548	0.0227	0.5	TRUE
Tyr	81.4021	74.9213	0.1069	0.5844	0.5	TRUE
Val	348.9504	284.4986	0.2647	0.0107	0.5	TRUE

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	LOD [µM]	Mean above LOD
Ac-Orn	3.569	5.2537	-0.484	0.1286	0.2	TRUE
ADMA	0.958	1.0554	-0.1234	0.3018	0.08	TRUE
alpha-AAA	10.4663	11.4459	-0.1141	0.7141	0.4	TRUE
c4-OH-Pro	0	0	NA	NA	0.1	FALSE
Carnosine	18.1588	21.5567	-0.2178	0.1179	0.1	TRUE
Creatinine	162.7676	191.1159	-0.204	0.0376	1	TRUE
DOPA	0	0	NA	NA	0.2	FALSE
Dopamine	0	0	NA	NA	0.1	FALSE
Histamine	4.3084	3.5689	0.2439	0.2351	0.01	TRUE
Kynurenine	0.5452	0.6277	-0.1793	0.4004	0.1728	TRUE
Met-SO	1.2791	1.26	0.0193	0.8958	0.3	TRUE
Nitro-Tyr	0	0	NA	NA	0.3	FALSE
PEA	0	0	NA	NA	0.02	FALSE
Putrescine	0.315	0.3346	-0.0773	0.657	0.0015	TRUE
SDMA	0.3366	0.4647	-0.4055	0.0798	0.3	TRUE
Serotonin	2.809	4.0676	-0.464	0.1035	0.03	TRUE
Spermidine	0.1294	0.1411	-0.1111	0.3787	0.1611	FALSE
Spermine	0.1668	0.1854	-0.1347	0.2154	0.5	FALSE
t4-OH-Pro	23.4815	29.2584	-0.2784	0.1472	0.1	TRUE
Taurine	90.742	105.8048	-0.1952	0.0964	0.8	TRUE
total DMA	0.8804	1.1476	-0.3346	0.017	0.1	TRUE
lysoPC a C14:0	2.7803	2.5639	0.1044	0.0166	5.7537	FALSE
lysoPC a C16:0	31.0509	30.3973	0.0273	0.8568	0.0699	TRUE
lysoPC a C16:1	1.1032	0.9014	0.2619	0.0122	0.1276	TRUE
lysoPC a C17:0	1.2225	1.4121	-0.1834	0.3695	0.0336	TRUE
lysoPC a C18:0	20.3696	19.3171	0.0682	0.7459	0.1303	TRUE
lysoPC a C18:1	13.0533	10.2147	0.3185	0.0117	0.0659	TRUE
lysoPC a C18:2	19.4206	15.0328	0.3328	0.0652	0.1083	TRUE
lysoPC a C20:3	1.1283	1.4515	-0.3181	0.1445	0.1534	TRUE
lysoPC a C20:4	8.2415	6.492	0.3098	0.0917	0.0269	TRUE
lysoPC a C24:0	0.862	0.7381	0.2006	0.0314	0.8528	FALSE
lysoPC a C26:0	0.7805	0.6237	0.2911	0.1593	0.0839	TRUE
lysoPC a C26:1	0.3834	0.3379	0.163	0.3885	0.0602	TRUE
lysoPC a C28:0	1.1128	0.9961	0.143	0.3766	0.1637	TRUE
lysoPC a C28:1	1.4664	1.052	0.4331	0.0128	0.0556	TRUE
PC aa C24:0	0.9265	0.7082	0.3493	0.016	0.0526	TRUE
PC aa C26:0	2.311	1.8878	0.2622	0.111	1.2045	TRUE
PC aa C28:1	0.7794	0.6145	0.3086	0.0198	0.1739	TRUE
PC aa C30:0	1.1832	1.0984	0.0957	0.2368	0.1354	TRUE
PC aa C30:2	0.3489	0.2821	0.2754	0.0992	0.0042	TRUE
PC aa C32:0	3.4038	3.2133	0.0741	0.5653	0.0282	TRUE
PC aa C32:1	1.942	1.4438	0.3859	0.1485	0.009	TRUE
PC aa C32:2	0.7405	0.5353	0.4231	0.0148	0.009	TRUE
PC aa C32:3	0.1428	0.1121	0.3149	0.0523	0.007	TRUE
PC aa C34:1	70.2812	52.0358	0.3914	0.054	0.0705	TRUE
PC aa C34:2	105.5119	79.929	0.3612	0.0077	0.0943	TRUE
PC aa C34:3	5.4335	3.6093	0.5353	0.006	0.0128	TRUE
PC aa C34:4	0.3074	0.2062	0.5226	0.0014	0.0113	TRUE
PC aa C36:0	0.334	0.3182	0.0623	0.8061	0.113	TRUE
PC aa C36:1	41.4207	29.7946	0.4296	0.0224	0.0295	TRUE
PC aa C36:2	129.4485	94.2682	0.4133	0.0113	0.0566	TRUE
PC aa C36:3	27.952	21.5308	0.3392	0.0158	0.0129	TRUE
PC aa C36:4	56.2459	42.2574	0.3721	0.021	0.025	TRUE
PC aa C36:5	4.6895	3.0911	0.5456	0.0034	0.0106	TRUE
PC aa C36:6	0.138	0.0885	0.5823	0.0015	0.0061	TRUE
PC aa C38:0	0.5047	0.4079	0.2762	0.0505	0.0493	TRUE
PC aa C38:1	0.6732	0.535	0.2983	0.065	0.0086	TRUE
PC aa C38:3	21.7686	23.0063	-0.0707	0.6121	0.0265	TRUE
PC aa C38:4	133.7782	92.7396	0.4786	0.0174	0.0493	TRUE
PC aa C38:5	33.0113	24.2485	0.4019	0.0054	0.0216	TRUE
PC aa C38:6	9.033	5.2563	0.7125	0.001	0.0081	TRUE
PC aa C40:1	0.2632	0.2298	0.1751	0.1202	0.7127	FALSE
PC aa C40:2	0.3604	0.2751	0.3512	0.014	0.0347	TRUE
PC aa C40:3	0.4616	0.4556	0.0167	0.8211	0.0019	TRUE
PC aa C40:4	3.4628	4.0858	-0.2101	0.311	0.0258	TRUE
PC aa C40:5	23.9301	21.3692	0.146	0.1279	0.04	TRUE
PC aa C40:6	6.5201	4.7344	0.4171	0.0843	0.2097	TRUE
PC aa C42:0	0.1656	0.1368	0.2476	0.0213	0.0584	TRUE
PC aa C42:1	0.1117	0.0917	0.2562	0.0148	0.0177	TRUE
PC aa C42:2	0.1234	0.1161	0.0778	0.4356	0.0669	TRUE

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	LOD [µM]	Mean above LOD
PC aa C42:4	0.0957	0.0867	0.1284	0.1567	0.0054	TRUE
PC aa C42:5	0.1298	0.113	0.179	0.1758	0.0114	TRUE
PC aa C42:6	0.3079	0.283	0.1087	0.0839	0.4702	FALSE
PC ae C30:0	0.2465	0.2244	0.1207	0.1572	0.1838	TRUE
PC ae C30:1	0.4129	0.3443	0.2353	0.2481	0.0073	TRUE
PC ae C30:2	0.3162	0.2909	0.1076	0.1516	0.5916	FALSE
PC ae C32:1	1.0029	0.8285	0.2475	0.069	0.0053	TRUE
PC ae C32:2	0.3041	0.2678	0.164	0.0508	0.0223	TRUE
PC ae C34:0	0.4343	0.4138	0.0619	0.6062	0.0327	TRUE
PC ae C34:1	3.464	2.8613	0.2476	0.191	0.0122	TRUE
PC ae C34:2	3.6972	2.8691	0.3295	0.0642	0.0062	TRUE
PC ae C34:3	2.3024	1.69	0.4029	0.0816	0.0194	TRUE
PC ae C36:0	0.2683	0.25	0.0912	0.5969	0.0881	TRUE
PC ae C36:1	3.4558	3.2809	0.0668	0.7759	0.1183	TRUE
PC ae C36:2	6.8446	6.2993	0.1069	0.5116	0.0287	TRUE
PC ae C36:3	2.5593	1.9077	0.3825	0.0483	0.0166	TRUE
PC ae C36:4	2.2783	1.7337	0.3552	0.0298	0.0098	TRUE
PC ae C36:5	1.5037	1.099	0.4085	0.0193	0.008	TRUE
PC ae C38:0	0.6545	0.4038	0.634	0.0012	0.1125	TRUE
PC ae C38:1	0.6039	0.5143	0.2077	0.2211	0.0051	TRUE
PC ae C38:2	1.2999	1.1994	0.1036	0.4808	0.0058	TRUE
PC ae C38:3	1.2865	1.4222	-0.1279	0.401	0.0017	TRUE
PC ae C38:4	5.5341	4.9905	0.1333	0.3165	0.0092	TRUE
PC ae C38:5	2.8477	2.1015	0.3958	0.0167	0.0031	TRUE
PC ae C38:6	0.6321	0.4358	0.4859	0.0067	0.0053	TRUE
PC ae C40:1	0.9485	0.5443	0.7311	0.0144	0.0122	TRUE
PC ae C40:2	0.4461	0.4173	0.0859	0.4015	0.0038	TRUE
PC ae C40:3	0.4738	0.4755	-0.0047	0.9516	0.0061	TRUE
PC ae C40:4	1.3404	1.2677	0.0718	0.3829	0.0796	TRUE
PC ae C40:5	1.5436	1.5236	0.0168	0.8937	0.008	TRUE
PC ae C40:6	0.5829	0.4512	0.3328	0.0216	0.0081	TRUE
PC ae C42:0	0.596	0.5727	0.0513	0.1754	1.3755	FALSE
PC ae C42:1	0.5472	0.4237	0.3324	0.0398	0.1838	TRUE
PC ae C42:2	0.6215	0.4623	0.3853	0.0277	0.0125	TRUE
PC ae C42:3	0.1878	0.1348	0.4326	0.0199	0.0016	TRUE
PC ae C42:4	0.1346	0.1417	-0.0658	0.4012	0.3	FALSE
PC ae C42:5	0.8138	0.7737	0.065	0.2005	1.5056	FALSE
PC ae C44:3	0.117	0.1042	0.1501	0.4015	0.0249	TRUE
PC ae C44:4	0.0947	0.0828	0.1732	0.1008	0.1042	FALSE
PC ae C44:5	0.1151	0.0935	0.2693	0.0399	0.1166	FALSE
PC ae C44:6	0.1213	0.1094	0.1334	0.4483	0.051	TRUE
SM (OH) C14:1	2.0191	2.201	-0.1101	0.5405	0.025	TRUE
SM (OH) C16:1	2.3309	2.7009	-0.1873	0.2184	0.012	TRUE
SM (OH) C22:1	3.4258	3.2764	0.0573	0.7176	0.0312	TRUE
SM (OH) C22:2	1.3815	1.5801	-0.1709	0.3224	0.0064	TRUE
SM (OH) C24:1	0.4216	0.4175	0.0123	0.9437	0.0186	TRUE
SM C16:0	49.7431	46.8082	0.0782	0.4683	0.0293	TRUE
SM C16:1	5.2457	4.8345	0.1051	0.5065	0.011	TRUE
SM C18:0	10.7454	9.7661	0.1232	0.4526	0.07	TRUE
SM C18:1	2.9899	2.6396	0.1609	0.3365	0.0117	TRUE
SM C20:2	0.1879	0.1382	0.3995	0.0747	0.005	TRUE
SM C22:3	0.0967	0.0626	0.5706	0.3645	0.0088	TRUE
SM C24:0	8.4521	9.0901	-0.0929	0.4551	0.0382	TRUE
SM C24:1	9.519	11.2511	-0.2123	0.2248	0.0151	TRUE
SM C26:0	0.0943	0.1061	-0.1501	0.643	0.015	TRUE
SM C26:1	0.0682	0.091	-0.3625	0.5715	0.0103	TRUE
H1	17367.5027	6337.6666	1.3488	0.0032	20	TRUE
C2 / C0	0.1545	0.1366	0.1588	0.6197	NA	NA
(C2+C3) / C0	0.1706	0.1511	0.1569	0.5924	NA	NA
CPT1 ratio	0.0061	0.0036	0.6877	0.0092	NA	NA
Total AC / C0	0.3594	0.2939	0.2607	0.0928	NA	NA
Total AC-DC / Total AC	0.0726	0.0777	-0.0871	0.6341	NA	NA
Total AC-OH / Total AC	0.079	0.0814	-0.0386	0.8301	NA	NA
SFA (PC)	11.0284	9.6354	0.1744	0.0321	NA	NA
MUFA (PC)	125.9065	93.5426	0.3869	0.0447	NA	NA
MUFA (PC) / SFA (PC)	11.3841	9.7641	0.1985	0.0797	NA	NA
PUFA (PC)	601.9406	455.024	0.364	0.0023	NA	NA

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	LOD [µM]	Mean above LOD
PUFA (PC) / SFA (PC)	54.6959	47.3265	0.187	0.0489	NA	NA
PUFA (PC) / MUFA (PC)	4.8437	4.8608	-0.0045	0.9647	NA	NA
Total PC	738.8756	558.202	0.3648	0.0019	NA	NA
Total lysoPC	102.9753	91.5306	0.152	0.2442	NA	NA
Total lysoPC / Total PC	0.1398	0.1637	-0.2006	0.1175	NA	NA
Total PC aa	688.2423	515.1954	0.3769	0.0017	NA	NA
Total PC ae	50.6332	43.0066	0.2112	0.0746	NA	NA
Total SM	96.721	94.9636	0.0235	0.8381	NA	NA
Total (PC+SM)	835.5966	653.1656	0.3199	0.0036	NA	NA
Total SM / Total (SM+PC)	0.1158	0.1451	-0.2854	0.01	NA	NA
Total SM / Total PC	0.131	0.17	-0.3288	0.0114	NA	NA
Total SM-non OH	87.1422	84.7876	0.0352	0.7635	NA	NA
Total SM-OH	9.5788	10.176	-0.0773	0.544	NA	NA
Total SM-OH / Total SM-non OH	0.1104	0.1197	-0.1032	0.2019	NA	NA
Orn / Arg	0.2963	0.2487	0.2269	0.2068	NA	NA
Tyr / Phe	0.8723	0.9487	-0.1071	0.4446	NA	NA
AAA	251.4707	216.5275	0.1934	0.0956	NA	NA
BCAA	819.887	620.7529	0.3619	0.0028	NA	NA
ESAA	1440.7919	1159.0397	0.2822	0.0067	NA	NA
GAA	1071.5667	1356.6073	-0.2983	0.0919	NA	NA
KAA	1036.0622	820.4297	0.3029	0.0086	NA	NA
BCAA/AAA	3.2656	2.9018	0.1524	0.1204	NA	NA

Table A.2: List of metabolites from the MIDY pig plasma with their associated statistics

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	Mean above LOD
C2	1.0732	1.0453	0.0338	0.9342	TRUE
C3	0.1072	0.1098	-0.0305	0.8701	TRUE
C3-OH	0.0224	0.024	-0.0871	0.4404	FALSE
C3:1	0.0144	0.0154	-0.0829	0.6249	FALSE
C3-DC (C4-OH)	0.0742	0.0773	-0.0525	0.6319	TRUE
C4	0.0697	0.0662	0.0675	0.3733	TRUE
C4:1	0.0626	0.036	0.7286	0.0058	TRUE
C5	0.0336	0.0345	-0.0344	0.8499	FALSE
C5-OH (C3-DC-M)	0.0474	0.0415	0.1718	0.3323	FALSE
C5-M-DC	0.028	0.0284	-0.0204	0.8854	FALSE
C5-DC (C6-OH)	0.015	0.0177	-0.2091	0.1794	FALSE
C5:1	0.0294	0.0285	0.0403	0.6264	FALSE
C5:1-DC	0.0134	0.0113	0.2195	0.1703	FALSE
C6 (C4:1-DC)	0.0436	0.0464	-0.0812	0.4755	FALSE
C6:1	0.0258	0.0277	-0.0894	0.3077	FALSE
C7-DC	0.0157	0.0155	0.0157	0.9425	FALSE
C8	0.0681	0.0656	0.0479	0.703	FALSE
C9	0.0212	0.0213	-0.0029	0.9742	FALSE
C10	0.1725	0.0934	0.8103	0.0041	TRUE
C10:1	0.0283	0.0289	-0.0276	0.8113	FALSE
C10:2	0.0428	0.0402	0.0783	0.2849	FALSE
C12	0.0432	0.0372	0.1931	0.2251	FALSE
C12-DC	0.095	0.0965	-0.02	0.6833	FALSE
C12:1	0.036	0.0337	0.0865	0.5555	FALSE
C14	0.0134	0.0094	0.4619	0.0563	FALSE
C14:1	0.0167	0.0133	0.2926	0.1107	TRUE
C14:1-OH	0.006	0.0054	0.1336	0.3714	FALSE
C14:2	0.005	0.0043	0.2038	0.4329	FALSE
C14:2-OH	0.0074	0.0061	0.2435	0.1715	FALSE
C16	0.0239	0.0167	0.464	0.186	TRUE
C16-OH	0.0442	0.0381	0.1897	0.3672	TRUE
C16:1	0.0312	0.0295	0.0721	0.6149	FALSE
C16:1-OH	0.0049	0.0049	-0.0043	0.989	FALSE
C16:2	0.0055	0.0042	0.3306	0.0533	FALSE
C16:2-OH	0.0119	0.0121	-0.0233	0.8257	FALSE
C18	0.0166	0.0099	0.6787	0.0074	TRUE

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	Mean above LOD
C18:1	0.0235	0.0146	0.6309	0.2123	TRUE
C18:1-OH	0.0102	0.0097	0.0701	0.4933	FALSE
C18:2	0.0079	0.0053	0.5034	0.1103	FALSE
Ala	325.6455	348.372	-0.0862	0.6355	TRUE
Arg	208.7214	210.6499	-0.0118	0.9143	TRUE
Asn	29.4186	27.7052	0.0772	0.5175	TRUE
Asp	19.8956	23.2538	-0.1982	0.3153	TRUE
Cit	91.023	98.0928	-0.0955	0.508	TRUE
Gln	399.1926	446.3814	-0.1424	0.1346	TRUE
Glu	132.2972	142.6014	-0.0958	0.6193	TRUE
Gly	634.9126	897.1689	-0.4341	0.052	TRUE
His	93.724	86.2438	0.1071	0.4696	TRUE
Ile	194.4627	140.5584	0.4232	0.0177	TRUE
Leu	276.4739	195.696	0.4509	0.0018	TRUE
Lys	157.0419	132.7425	0.2175	0.0315	TRUE
Met	43.4574	42.7889	0.0199	0.8946	TRUE
Orn	61.2988	51.8737	0.216	0.1482	TRUE
Phe	92.8511	78.172	0.2227	0.0161	TRUE
Pro	260.7201	241.8348	0.0968	0.2368	TRUE
Ser	111.0086	111.0663	-0.0007	0.9945	TRUE
Thr	156.613	134.9053	0.1929	0.2449	TRUE
Trp	77.2174	63.4342	0.2548	0.0227	TRUE
Tyr	81.4021	74.9213	0.1069	0.5844	TRUE
Val	348.9504	284.4986	0.2647	0.0107	TRUE
Ac-Orn	3.569	5.2537	-0.484	0.1286	TRUE
ADMA	0.958	1.0554	-0.1234	0.3018	TRUE
alpha-AAA	10.4663	11.4459	-0.1141	0.7141	TRUE
c4-OH-Pro	0	0	NA	NA	FALSE
Carnosine	18.1588	21.5567	-0.2178	0.1179	TRUE
Creatinine	162.7676	191.1159	-0.204	0.0376	TRUE
DOPA	0	0	NA	NA	FALSE
Dopamine	0	0	NA	NA	FALSE
Histamine	4.3084	3.5689	0.2439	0.2351	TRUE
Kynurenine	0.5452	0.6277	-0.1793	0.4004	TRUE
Met-SO	1.2791	1.26	0.0193	0.8958	TRUE
Nitro-Tyr	0	0	NA	NA	FALSE
PEA	0	0	NA	NA	FALSE
Putrescine	0.315	0.3346	-0.0773	0.657	TRUE
SDMA	0.3366	0.4647	-0.4055	0.0798	TRUE
Serotonin	2.809	4.0676	-0.464	0.1035	TRUE
Spermidine	0.1294	0.1411	-0.1111	0.3787	FALSE
Spermine	0.1668	0.1854	-0.1347	0.2154	FALSE
t4-OH-Pro	23.4815	29.2584	-0.2784	0.1472	TRUE
Taurine	90.742	105.8048	-0.1952	0.0964	TRUE
total DMA	0.8804	1.1476	-0.3346	0.017	TRUE
lysoPC a C14:0	2.7803	2.5639	0.1044	0.0166	FALSE
lysoPC a C16:0	31.0509	30.3973	0.0273	0.8568	TRUE
lysoPC a C16:1	1.1032	0.9014	0.2619	0.0122	TRUE
lysoPC a C17:0	1.2225	1.4121	-0.1834	0.3695	TRUE
lysoPC a C18:0	20.3696	19.3171	0.0682	0.7459	TRUE
lysoPC a C18:1	13.0533	10.2147	0.3185	0.0117	TRUE
lysoPC a C18:2	19.4206	15.0328	0.3328	0.0652	TRUE
lysoPC a C20:3	1.1283	1.4515	-0.3181	0.1445	TRUE
lysoPC a C20:4	8.2415	6.492	0.3098	0.0917	TRUE
lysoPC a C24:0	0.862	0.7381	0.2006	0.0314	FALSE
lysoPC a C26:0	0.7805	0.6237	0.2911	0.1593	TRUE
lysoPC a C26:1	0.3834	0.3379	0.163	0.3885	TRUE
lysoPC a C28:0	1.1128	0.9961	0.143	0.3766	TRUE
lysoPC a C28:1	1.4664	1.052	0.4331	0.0128	TRUE
PC aa C24:0	0.9265	0.7082	0.3493	0.016	TRUE
PC aa C26:0	2.311	1.8878	0.2622	0.111	TRUE
PC aa C28:1	0.7794	0.6145	0.3086	0.0198	TRUE
PC aa C30:0	1.1832	1.0984	0.0957	0.2368	TRUE
PC aa C30:2	0.3489	0.2821	0.2754	0.0992	TRUE
PC aa C32:0	3.4038	3.2133	0.0741	0.5653	TRUE
PC aa C32:1	1.942	1.4438	0.3859	0.1485	TRUE
PC aa C32:2	0.7405	0.5353	0.4231	0.0148	TRUE
PC aa C32:3	0.1428	0.1121	0.3149	0.0523	TRUE
PC aa C34:1	70.2812	52.0358	0.3914	0.054	TRUE
PC aa C34:2	105.5119	79.929	0.3612	0.0077	TRUE

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	Mean above LOD
PC aa C34:3	5.4335	3.6093	0.5353	0.006	TRUE
PC aa C34:4	0.3074	0.2062	0.5226	0.0014	TRUE
PC aa C36:0	0.334	0.3182	0.0623	0.8061	TRUE
PC aa C36:1	41.4207	29.7946	0.4296	0.0224	TRUE
PC aa C36:2	129.4485	94.2682	0.4133	0.0113	TRUE
PC aa C36:3	27.952	21.5308	0.3392	0.0158	TRUE
PC aa C36:4	56.2459	42.2574	0.3721	0.021	TRUE
PC aa C36:5	4.6895	3.0911	0.5456	0.0034	TRUE
PC aa C36:6	0.138	0.0885	0.5823	0.0015	TRUE
PC aa C38:0	0.5047	0.4079	0.2762	0.0505	TRUE
PC aa C38:1	0.6732	0.535	0.2983	0.065	TRUE
PC aa C38:3	21.7686	23.0063	-0.0707	0.6121	TRUE
PC aa C38:4	133.7782	92.7396	0.4786	0.0174	TRUE
PC aa C38:5	33.0113	24.2485	0.4019	0.0054	TRUE
PC aa C38:6	9.033	5.2563	0.7125	0.001	TRUE
PC aa C40:1	0.2632	0.2298	0.1751	0.1202	FALSE
PC aa C40:2	0.3604	0.2751	0.3512	0.014	TRUE
PC aa C40:3	0.4616	0.4556	0.0167	0.8211	TRUE
PC aa C40:4	3.4628	4.0858	-0.2101	0.311	TRUE
PC aa C40:5	23.9301	21.3692	0.146	0.1279	TRUE
PC aa C40:6	6.5201	4.7344	0.4171	0.0843	TRUE
PC aa C42:0	0.1656	0.1368	0.2476	0.0213	TRUE
PC aa C42:1	0.1117	0.0917	0.2562	0.0148	TRUE
PC aa C42:2	0.1234	0.1161	0.0778	0.4356	TRUE
PC aa C42:4	0.0957	0.0867	0.1284	0.1567	TRUE
PC aa C42:5	0.1298	0.113	0.179	0.1758	TRUE
PC aa C42:6	0.3079	0.283	0.1087	0.0839	FALSE
PC ae C30:0	0.2465	0.2244	0.1207	0.1572	TRUE
PC ae C30:1	0.4129	0.3443	0.2353	0.2481	TRUE
PC ae C30:2	0.3162	0.2909	0.1076	0.1516	FALSE
PC ae C32:1	1.0029	0.8285	0.2475	0.069	TRUE
PC ae C32:2	0.3041	0.2678	0.164	0.0508	TRUE
PC ae C34:0	0.4343	0.4138	0.0619	0.6062	TRUE
PC ae C34:1	3.464	2.8613	0.2476	0.191	TRUE
PC ae C34:2	3.6972	2.8691	0.3295	0.0642	TRUE
PC ae C34:3	2.3024	1.69	0.4029	0.0816	TRUE
PC ae C36:0	0.2683	0.25	0.0912	0.5969	TRUE
PC ae C36:1	3.4558	3.2809	0.0668	0.7759	TRUE
PC ae C36:2	6.8446	6.2993	0.1069	0.5116	TRUE
PC ae C36:3	2.5593	1.9077	0.3825	0.0483	TRUE
PC ae C36:4	2.2783	1.7337	0.3552	0.0298	TRUE
PC ae C36:5	1.5037	1.099	0.4085	0.0193	TRUE
PC ae C38:0	0.6545	0.4038	0.634	0.0012	TRUE
PC ae C38:1	0.6039	0.5143	0.2077	0.2211	TRUE
PC ae C38:2	1.2999	1.1994	0.1036	0.4808	TRUE
PC ae C38:3	1.2865	1.4222	-0.1279	0.401	TRUE
PC ae C38:4	5.5341	4.9905	0.1333	0.3165	TRUE
PC ae C38:5	2.8477	2.1015	0.3958	0.0167	TRUE
PC ae C38:6	0.6321	0.4358	0.4859	0.0067	TRUE
PC ae C40:1	0.9485	0.5443	0.7311	0.0144	TRUE
PC ae C40:2	0.4461	0.4173	0.0859	0.4015	TRUE
PC ae C40:3	0.4738	0.4755	-0.0047	0.9516	TRUE
PC ae C40:4	1.3404	1.2677	0.0718	0.3829	TRUE
PC ae C40:5	1.5436	1.5236	0.0168	0.8937	TRUE
PC ae C40:6	0.5829	0.4512	0.3328	0.0216	TRUE
PC ae C42:0	0.596	0.5727	0.0513	0.1754	FALSE
PC ae C42:1	0.5472	0.4237	0.3324	0.0398	TRUE
PC ae C42:2	0.6215	0.4623	0.3853	0.0277	TRUE
PC ae C42:3	0.1878	0.1348	0.4326	0.0199	TRUE
PC ae C42:4	0.1346	0.1417	-0.0658	0.4012	FALSE
PC ae C42:5	0.8138	0.7737	0.065	0.2005	FALSE
PC ae C44:3	0.117	0.1042	0.1501	0.4015	TRUE
PC ae C44:4	0.0947	0.0828	0.1732	0.1008	FALSE
PC ae C44:5	0.1151	0.0935	0.2693	0.0399	FALSE
PC ae C44:6	0.1213	0.1094	0.1334	0.4483	TRUE
SM (OH) C14:1	2.0191	2.201	-0.1101	0.5405	TRUE
SM (OH) C16:1	2.3309	2.7009	-0.1873	0.2184	TRUE
SM (OH) C22:1	3.4258	3.2764	0.0573	0.7176	TRUE
SM (OH) C22:2	1.3815	1.5801	-0.1709	0.3224	TRUE
SM (OH) C24:1	0.4216	0.4175	0.0123	0.9437	TRUE

Metabolite	MIDY Mean [µM]	WT Mean [µM]	log2FC	p-value	Mean above LOD
SM C16:0	49.7431	46.8082	0.0782	0.4683	TRUE
SM C16:1	5.2457	4.8345	0.1051	0.5065	TRUE
SM C18:0	10.7454	9.7661	0.1232	0.4526	TRUE
SM C18:1	2.9899	2.6396	0.1609	0.3365	TRUE
SM C20:2	0.1879	0.1382	0.3995	0.0747	TRUE
SM C22:3	0.0967	0.0626	0.5706	0.3645	TRUE
SM C24:0	8.4521	9.0901	-0.0929	0.4551	TRUE
SM C24:1	9.519	11.2511	-0.2123	0.2248	TRUE
SM C26:0	0.0943	0.1061	-0.1501	0.643	TRUE
SM C26:1	0.0682	0.091	-0.3625	0.5715	TRUE
H1	17367.5027	6337.6666	1.3488	0.0032	TRUE
C2 / C0	0.1545	0.1366	0.1588	0.6197	NA
(C2+C3) / C0	0.1706	0.1511	0.1569	0.5924	NA
CPT1 ratio	0.0061	0.0036	0.6877	0.0092	NA
Total AC / C0	0.3594	0.2939	0.2607	0.0928	NA
Total AC-DC / Total AC	0.0726	0.0777	-0.0871	0.6341	NA
Total AC-OH / Total AC	0.079	0.0814	-0.0386	0.8301	NA
SFA (PC)	11.0284	9.6354	0.1744	0.0321	NA
MUFA (PC)	125.9065	93.5426	0.3869	0.0447	NA
MUFA (PC) / SFA (PC)	11.3841	9.7641	0.1985	0.0797	NA
PUFA (PC)	601.9406	455.024	0.364	0.0023	NA
PUFA (PC) / SFA (PC)	54.6959	47.3265	0.187	0.0489	NA
PUFA (PC) / MUFA (PC)	4.8437	4.8608	-0.0045	0.9647	NA
Total PC	738.8756	558.202	0.3648	0.0019	NA
Total lysoPC	102.9753	91.5306	0.152	0.2442	NA
Total lysoPC / Total PC	0.1398	0.1637	-0.2006	0.1175	NA
Total PC aa	688.2423	515.1954	0.3769	0.0017	NA
Total PC ae	50.6332	43.0066	0.2112	0.0746	NA
Total SM	96.721	94.9636	0.0235	0.8381	NA
Total (PC+SM)	835.5966	653.1656	0.3199	0.0036	NA
Total SM / Total (SM+PC)	0.1158	0.1451	-0.2854	0.01	NA
Total SM / Total PC	0.131	0.17	-0.3288	0.0114	NA
Total SM-non OH	87.1422	84.7876	0.0352	0.7635	NA
Total SM-OH	9.5788	10.176	-0.0773	0.544	NA
Total SM-OH / Total SM-non OH	0.1104	0.1197	-0.1032	0.2019	NA
Orn / Arg	0.2963	0.2487	0.2269	0.2068	NA
Tyr / Phe	0.8723	0.9487	-0.1071	0.4446	NA
AAA	251.4707	216.5275	0.1934	0.0956	NA
BCAA	819.887	620.7529	0.3619	0.0028	NA
ESAA	1440.7919	1159.0397	0.2822	0.0067	NA
GAA	1071.5667	1356.6073	-0.2983	0.0919	NA
KAA	1036.0622	820.4297	0.3029	0.0086	NA
BCAA/AAA	3.2656	2.9018	0.1524	0.1204	NA

Table A.3: List of metabolites from the MIDY pig liver with their associated statistics

Gene set name	Size	NES	p-value	FDR
amino acid metabolism	237	3.58	0.00E+00	0.00E+00
membrane transport	33	2.87	0.00E+00	8.29E-04
abc transporters	33	2.81	0.00E+00	1.11E-03
arginine biosynthesis	18	2.60	0.00E+00	2.18E-03
carbohydrate metabolism	280	2.33	0.00E+00	1.22E-02
cysteine and methionine metabolism	41	2.30	2.02E-03	1.19E-02
valine, leucine and isoleucine degradation	46	2.25	0.00E+00	1.41E-02
propanoate metabolism	29	2.25	0.00E+00	1.26E-02
butanoate metabolism	21	2.20	0.00E+00	1.57E-02
glycine, serine and threonine metabolism	35	2.16	0.00E+00	1.74E-02
peroxisome	76	2.12	2.07E-03	2.17E-02
citrate cycle (tca cycle)	29	2.10	4.16E-03	2.20E-02
alanine, aspartate and glutamate metabolism	31	2.07	6.04E-03	2.53E-02
metabolism of cofactors and vitamins	157	2.00	6.12E-03	3.61E-02
tgf-beta signaling pathway	67	1.99	4.07E-03	3.61E-02
ppar signaling pathway	57	1.97	7.74E-03	3.95E-02
erbb signaling pathway	70	1.94	5.92E-03	4.46E-02
fatty acid degradation	36	1.93	1.18E-02	4.49E-02
signaling pathways regulating pluripotency of stem cells	98	1.92	9.78E-03	4.53E-02
lysine degradation	53	1.83	1.37E-02	6.87E-02

Table A.4: Upregulated genes in the Lexogen library GSEA showing a similar result to that of the NuGEN library.

Gene set name	Size	NES	p-value	FDR
ribosome	120	-5.03	0.00E+00	0.00E+00
signaling molecules and interaction	374	-4.22	0.00E+00	0.00E+00
immune diseases	148	-3.93	0.00E+00	0.00E+00
translation	400	-3.80	0.00E+00	0.00E+00
phagosome	117	-3.45	0.00E+00	0.00E+00
infectious diseases: parasitic	210	-3.33	0.00E+00	0.00E+00
intestinal immune network for iga production	29	-3.22	0.00E+00	0.00E+00
antigen processing and presentation	50	-3.12	0.00E+00	0.00E+00
complement and coagulation cascades	73	-3.12	0.00E+00	0.00E+00
ecm-receptor interaction	59	-3.03	0.00E+00	1.37E-04
graft-versus-host disease	22	-3.02	0.00E+00	1.25E-04
transcription	185	-3.02	0.00E+00	1.14E-04
allograft rejection	24	-2.93	0.00E+00	1.97E-04
infectious diseases: bacterial	273	-2.85	0.00E+00	1.82E-04
protein digestion and absorption	50	-2.81	0.00E+00	1.70E-04
hematopoietic cell lineage	63	-2.81	0.00E+00	1.60E-04
cytokine-cytokine receptor interaction	135	-2.76	0.00E+00	2.17E-04
type i diabetes mellitus	27	-2.67	0.00E+00	5.49E-04
focal adhesion	159	-2.66	0.00E+00	5.82E-04
proteasome	40	-2.51	0.00E+00	1.48E-03

Table A.5: Upregulated genes in the Lexogen library GSEA showing a similar result to that of the NuGEN library.

Appendix B

Supplementary figures

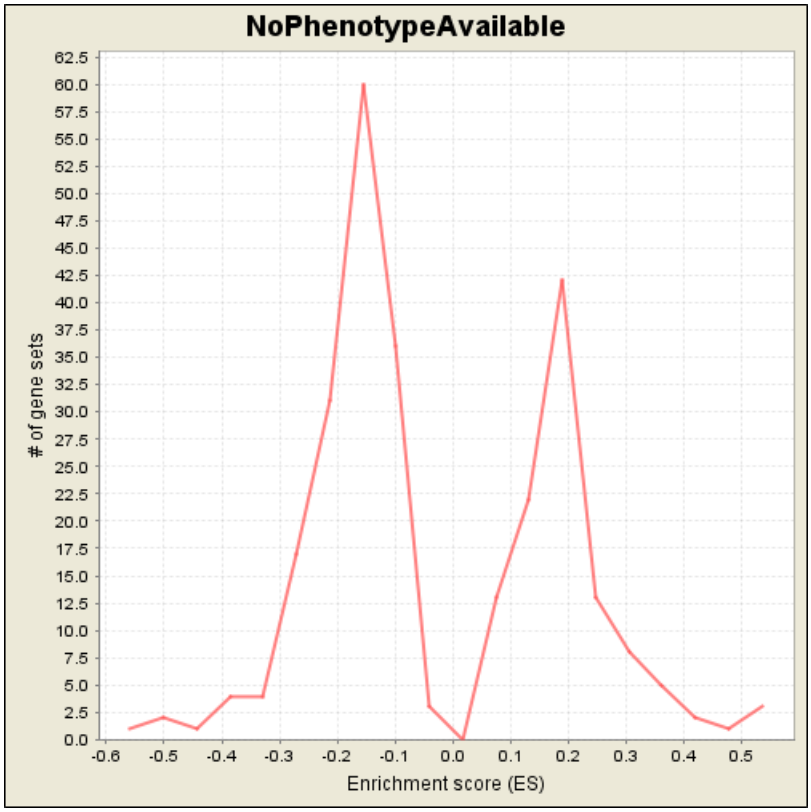

Figure B.1: Graph showing the distribution of Enrichment Scores (ES) in the MIDY liver GSEA results

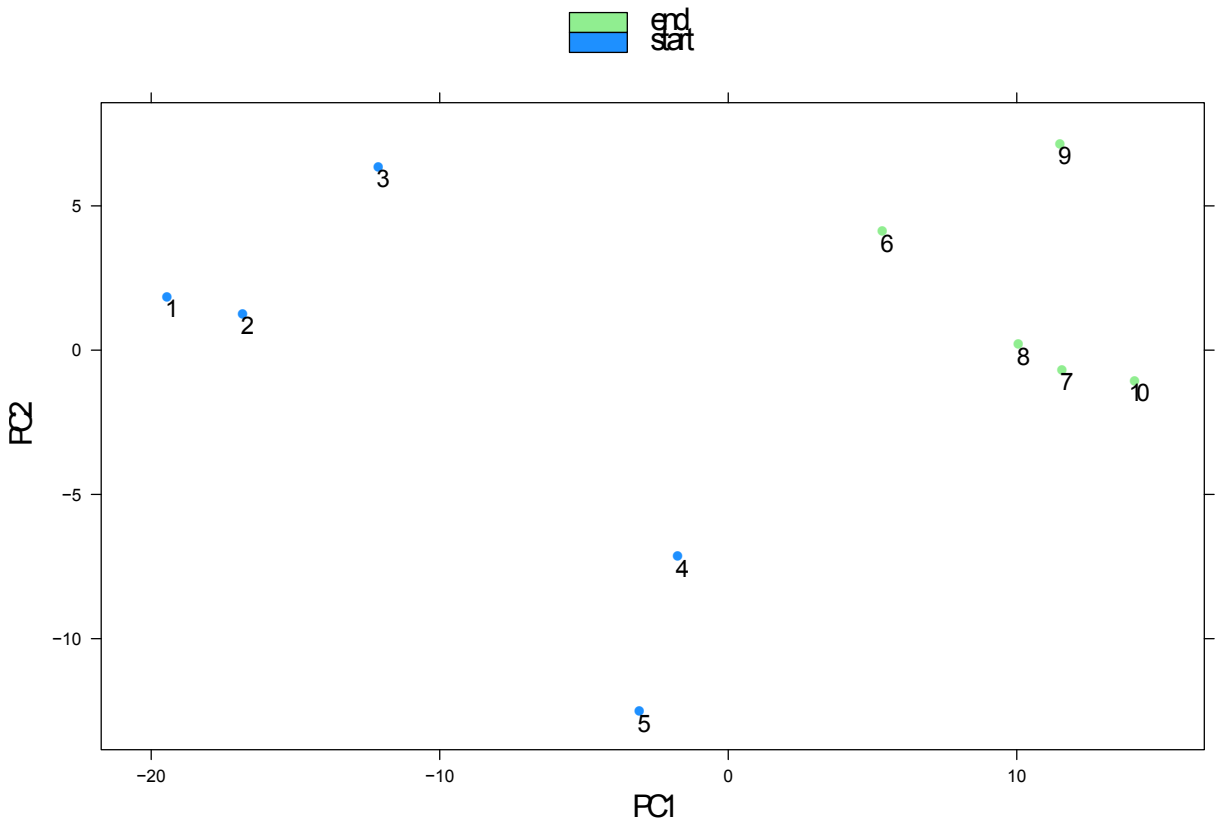

Figure B.2: PCA plot showing the sequencing of the first round of QuantSeq library preparation. PC1 correlates to the sample handling order, which is the label on the samples. Coloring indicate the two halves of samples processed.

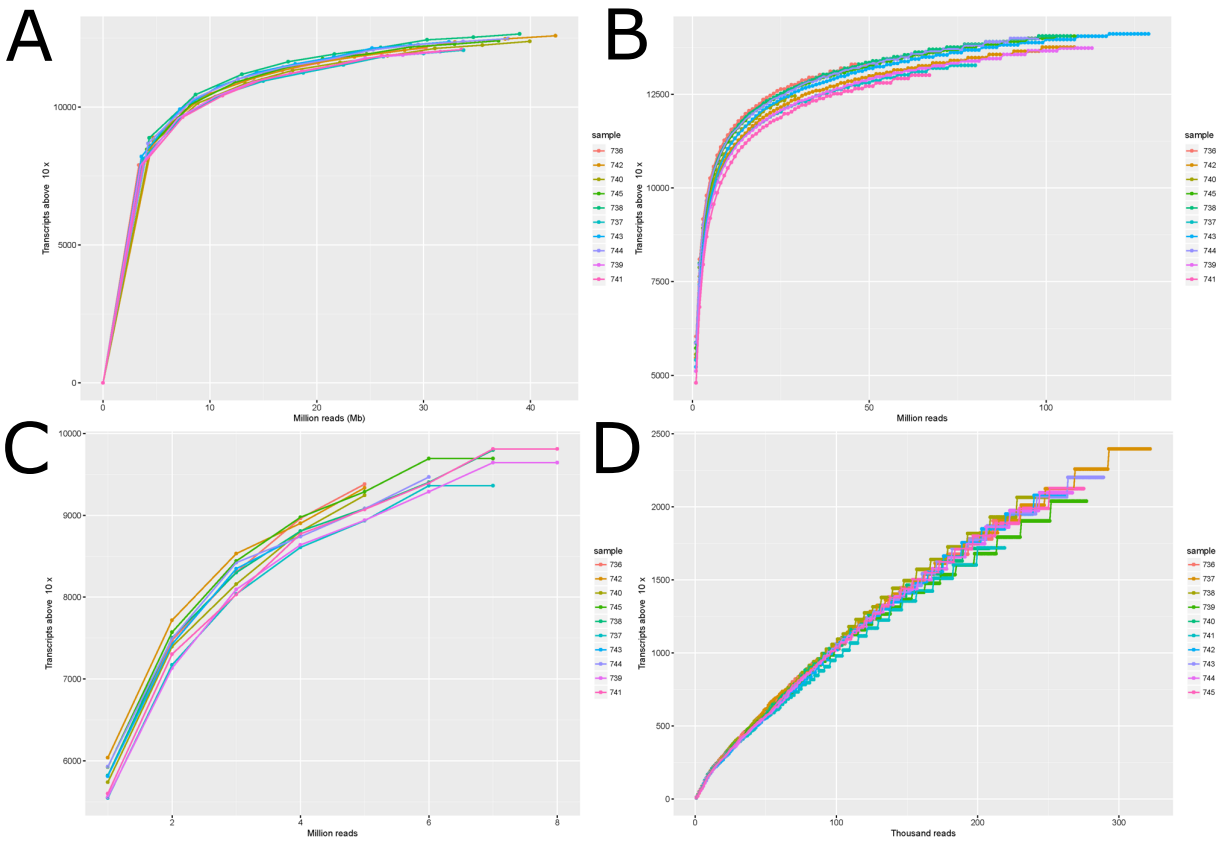

Figure B.3: Saturation plots showing how many genes are covered to a depth of 10 counts per gene. Sequencing techniques **A** NuGEN Encore complete, **B** Lexogen Sense, **C** Lexogen Quant Seq, and **SCRB Seq** are shown

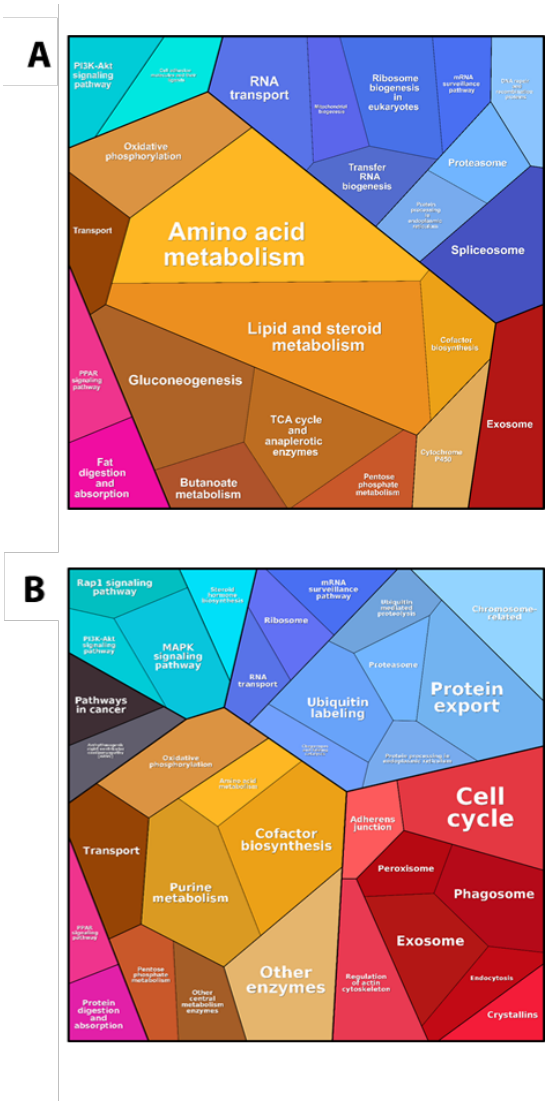

Figure B.4: Proteomaps showing **A** upregulated gene sets, and **B** downregulated gene sets. KEGG gene sets are grouped by functional relation. Each polygon represents the fraction of mass the corresponding protein family takes up in the total amount of protein.

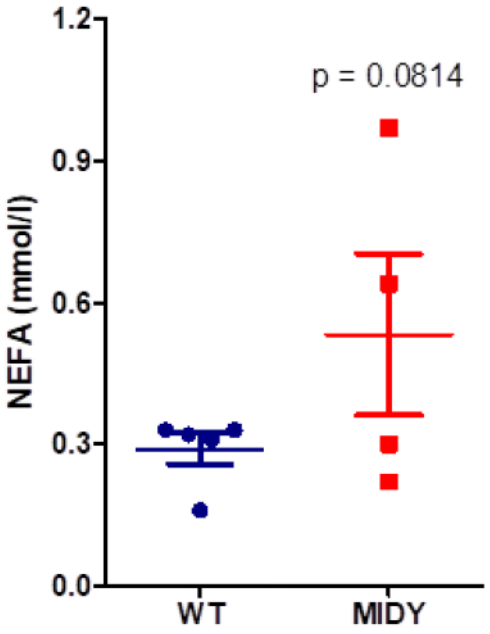

Figure B.5: Quantification of non-esterified fatty acid (NEFA) concentrations in plasma samples of INSC94Y transgenic (MIDY) pigs and wild-type (WT) controls

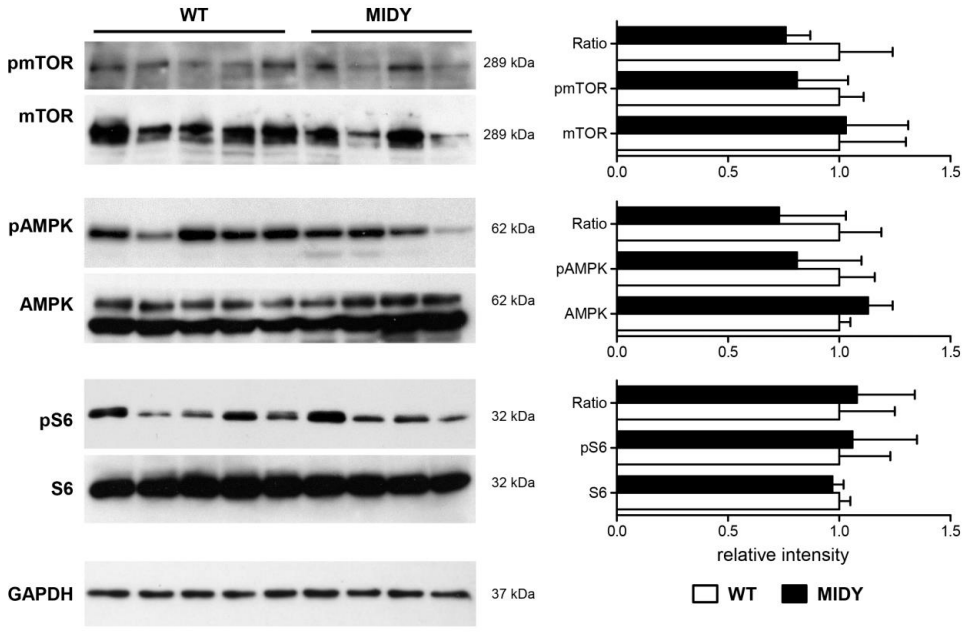

Figure B.6: Western blot analysis of abundance and phosphorylation levels of signaling molecules in liver extracts from INSC94Y transgenic (MIDY) pigs and wild-type (WT) controls.

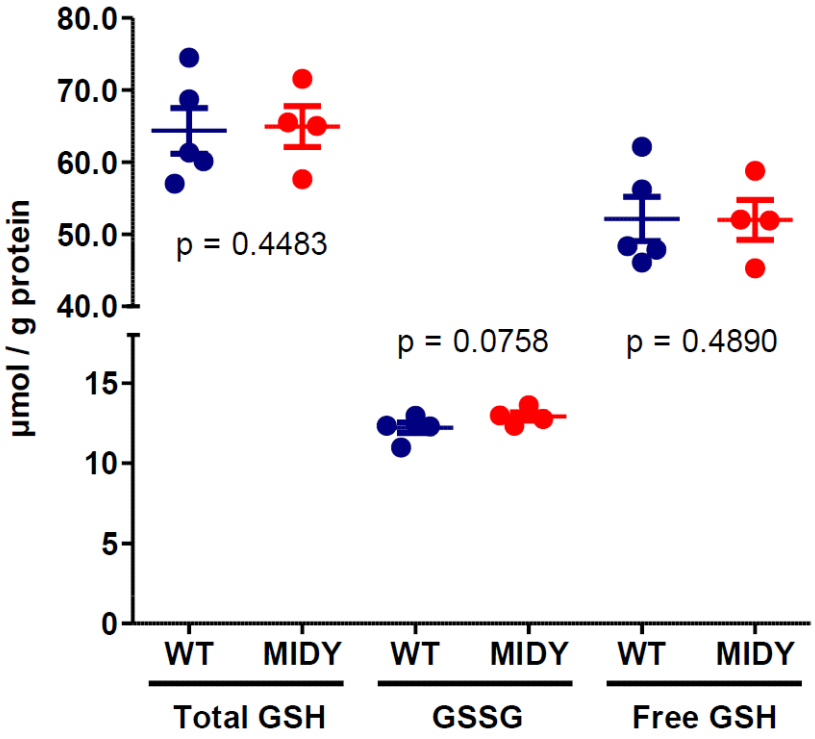

Figure B.7: Quantification of glutathione (GSH), oxidized glutathione (GSSG), and free GSH concentrations in liver samples from INSC94Y transgenic (MIDY) pigs and wildtype (WT) controls.

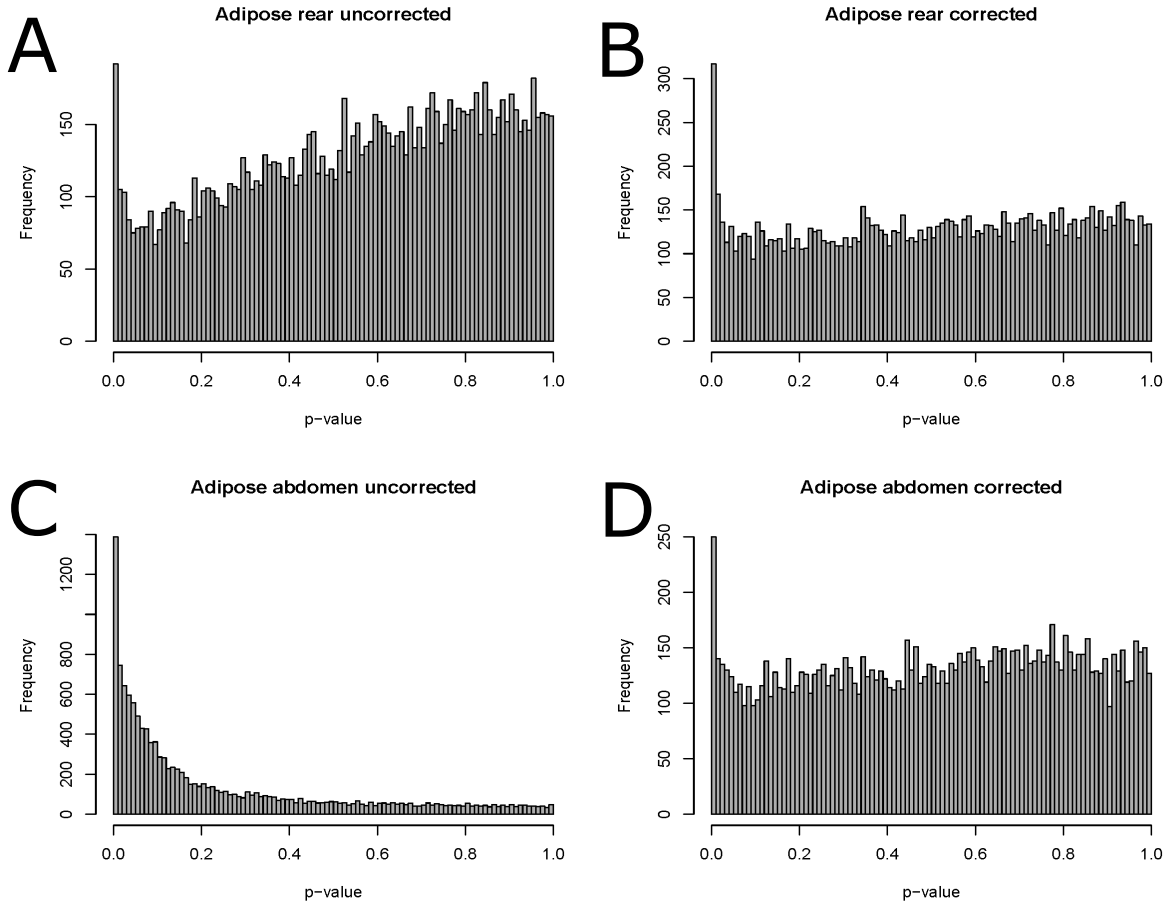

Figure B.8: Plots showing the SVA correction of the two subcutaneous adipose tissues with **A,C** showing the p-value histogram before correction while **B,D** shows the p-value histogram after correction.

Bibliography

AR Abbas, D Baldwin, Y Ma, W Ouyang, A Gurney, F Martin, S Fong, M van Lookeren Campagne, P Godowski, PM Williams, et al. Immune response in silico (iris): immune-specific genes identified from a compendium of microarray expression data. *Genes and immunity*, 6(4):319, 2005.

A Abbott. Inside the first pig biobank. *Nature*, 519(7544):397–8, 2015.

AB Adomas, F Lopez-Giraldez, TA Clark, Z Wang, and JP Townsend. Multi-targeted priming for genome-wide gene expression assays. *BMC genomics*, 11(1):477, 2010.

E Afgan, D Baker, B Batut, M Van Den Beek, D Bouvier, M Čech, J Chilton, D Clements, N Coraor, BA Grüning, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 46(W1):W537–W544, 2018.

Y Agca. Genome resource banking of biomedically important laboratory animals. *Theriogenology*, 78(8):1653–65, 2012.

A Ahamed, AG Unnikrishnan, SS Pendsey, S Nampoothiri, N Bhavani, VP Praveen, H Kumar, RV Jayakumar, V Nair, S Ellard, et al. Permanent neonatal diabetes mellitus due to a c96y heterozygous mutation in the insulin gene. a case report. *Jop*, 9(6):715–8, 2008.

H Ahsan. Diabetic retinopathy–biomolecules and multiple pathophysiology. *Diabetes Metab Syndr*, 9(1):51–4, 2015.

B Aigner, S Renner, B Kessler, N Klymiuk, M Kurome, A Wünsch, and E Wolf. Transgenic pigs as models for translational biomedical research. *J mol med*, 88(7):653–664, 2010.

B Albl, S Haesner, C Braun-Reichhart, E Streckel, S Renner, F Seeliger, E Wolf, R Wanke, and A Blutke. Tissue sampling guides for porcine biomedical models. *Toxicol Pathol*, 44 (3):414–20, 2016a.

B Albl, S Haesner, C Braun-Reichhart, E Streckel, S Renner, F Seeliger, E Wolf, R Wanke, and A Blutke. Tissue sampling guides for porcine biomedical models. *Toxicol pathol*, 44 (3):414–420, 2016b.

SF Altschul, W Gish, W Miller, EW Myers, and DJ Lipman. Basic local alignment search tool. *J mol biol*, 215(3):403–410, 1990.

S Anders, PT Pyl, and W Huber. Htseq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–9, 2015.

SS Andrali, ML Sampley, NL Vanderford, and S Ozcan. Glucose regulation of insulin gene expression in pancreatic beta-cells. *Biochem J*, 415(1):1–10, 2008.

S Andrews. Fastqc. *A quality control tool for high throughput sequence data*, 370, 2010.

BS Antharavally, KA Mallia, P Rangaraj, P Haney, and PA Bell. Quantitation of proteins using a dye–metal-based colorimetric protein assay. *Analytical biochemistry*, 385(2): 342–345, 2009.

SS Apte. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (adamts) superfamily: functions and mechanisms. *J biol chem*, 284(46): 31493–31497, 2009.

MA Ashworth, FN Leach, and RD Milner. Development of insulin secretion in the human fetus. *Arch Dis Child*, 48(2):151–2, 1973.

Diabetes Atlas. International diabetes federation. idf diabetes atlas, 7th edn. brussels, belgium: International diabetes federation, 2015.

M Backman, F Flenkenthaler, A Blutke, M Dahlhoff, E Ländström, S Renner, J Philippou-Massier, S Krebs, B Rathkolb, C Prehn, et al. Multi-omics insights into functional alterations of the liver in insulin-deficient diabetes mellitus. *Molecular Metabolism*, 26: 30–44, 2019.

M Bantscheff, S Lemeer, MM Savitski, and B Kuster. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal and bioanal chem*, 404(4):939–965, 2012.

DA Bellinger, EP Merricks, and TC Nichols. Swine models of type 2 diabetes mellitus: insulin resistance, glucose tolerance, and cardiovascular complications. *ILAR journal*, 47(3):243–258, 2006.

OV Belyaeva, MP Johnson, and NY Kedishvili. Kinetic analysis of human enzyme rdh10 defines the characteristics of a physiologically relevant retinol dehydrogenase. *J biol chem*, 283(29):20299–20308, 2008.

Y Benjamini and Y Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Stat Soc*, 57(1):289–300, 1995.

F Berg, U Gustafson, and L Andersson. The uncoupling protein 1 gene (ucp1) is disrupted in the pig lineage: a genetic explanation for poor thermoregulation in piglets. *PLoS genetics*, 2(8):e129, 2006.

V Bhargava, P Ko, E Willems, M Mercola, and S Subramaniam. Quantitative transcriptomics using designed primer-based amplification. *Scientific reports*, 3:1740, 2013.

G Bindea, B Mlecnik, H Hackl, P Charoentong, M Tosolini, A Kirilovsky, WH Fridman, F Pages, Z Trajanoski, and J Galon. Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–3, 2009.

G Bindea, J Galon, and B Mlecnik. Cluepedia cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*, 29(5):661–3, 2013.

MG Bischof, M Krssak, M Krebs, E Bernroider, H Stingl, W Waldhäusl, and M Roden. Effects of short-term improvement of insulin treatment and glycemia on hepatic glycogen metabolism in type 1 diabetes. *Diabetes*, 50(2):392–398, 2001.

H Bismuth. Surgical anatomy and anatomical surgery of the liver. *World j surg*, 6(1):3–9, 1982.

R Blomhoff and HK Blomhoff. Overview of retinoid metabolism and function. *J neurobiol*, 66(7):606–630, 2006.

A Blutke, S Renner, F Flenkenthaler, M Backman, S Haesner, E Kemter, E Landstrom, C Braun-Reichhart, B Albl, E Streckel, B Rathkolb, C Prehn, A Palladini, M Grzybek, S Krebs, S Bauersachs, A Bahr, A Bruhschwein, C A Deeg, E De Monte, M Dmochewitz, C Eberle, D Emrich, R Fux, F Groth, S Gumbert, A Heitmann, A Hinrichs, B Kessler, M Kurome, M Leipig-Rudolph, K Matiasek, H Ozturk, C Otzdorff, M Reichenbach, HD Reichenbach, A Rieger, B Rieseberg, M Rosati, MN Saucedo, A Schleicher, MR Schneider, K Simmet, J Steinmetz, N Ubel, P Zehetmaier, A Jung, J Adamski, U Coskun, M Hrabe de Angelis, C Simmet, M Ritzmann, A Meyer-Lindenberg, H Blum, GJ Arnold, T Frohlich, R Wanke, and E Wolf. The munich midy pig biobank - a unique resource for studying organ crosstalk in diabetes. *Mol Metab*, 6(8):931–940, 2017.

KN Bojsen-Møller, AM Lundsgaard, S Madsbad, B Kiens, and JJ Holst. Hepatic insulin clearance in regulation of systemic insulin concentrations—role of carbohydrate and energy availability. *Diabetes*, 67(11):2129–2136, 2018.

AM Bolger, M Lohse, and B Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–20, 2014.

D Bottomly, NAR Walter, JE Hunter, P Darakjian, S Kawane, KJ Buck, RP Searles, M Mooney, SK McWeeney, and R Hitzemann. Evaluating gene expression in c57bl/6j and dba/2j mouse striatum using rna-seq and microarrays. *PloS one*, 6(3):e17820, 2011.

F Braet and E Wisse. Structural and functional aspects of liver sinusoidal endothelial cell fenestrae: a review. *Comp hepatol*, 1(1):1, 2002.

E Breous, S Somanathan, LH Vandenberghe, and JM Wilson. Hepatic regulatory t cells and kupffer cells are crucial mediators of systemic t cell tolerance to antigens targeting murine liver. *Hepatology*, 50(2):612–621, 2009.

B Brooks-Worrell and JP Palmer. Is diabetes mellitus a continuous spectrum? *Clin Chem*, 57(2):158–61, 2011.

M Brownlee. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*, 54(6):1615–25, 2005.

Z Cao and ME Cooper. Pathogenesis of diabetic nephropathy. *J Diabetes Investig*, 2(4): 243–7, 2011.

WT Cefalu. Insulin resistance: cellular and clinical concepts. *Exp Biol Med (Maywood)*, 226(1):13–26, 2001.

AS Charonis, LA Reger, JE Dege, K Kouzi-Koliakos, LT Furcht, RM Wohlhueter, and EC Tsilibary. Laminin alterations after in vitro nonenzymatic glycosylation. *Diabetes*, 39(7):807–14, 1990.

R Chowdhury, KM Narayan, A Zabetian, S Raj, and R Tabassum. Genetic studies of type 2 diabetes in south asians: a systematic overview. *Curr Diabetes Rev*, 10(4):258–74, 2014.

G Chu, B Narasimhan, R Tibshirani, and V Tusher. Sam:"significance analysis of microarrays" users guide and technical document. 2001.

HR Chua, A Schneider, and R Bellomo. Bicarbonate in diabetic ketoacidosis-a systematic review. *Ann inten care*, 1(1):23, 2011.

PJA Cock, CJ Fields, N Goto, ML Heuer, and PM Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6):1767–1771, 12 2009. ISSN 0305-1048.

A Conesa, P Madrigal, S Tarazona, D Gomez-Cabrero, A Cervera, A McPherson, MW Szczesniak, DJ Gaffney, LL Elo, X Zhang, and A Mortazavi. A survey of best practices for rna-seq data analysis. *Genome Biol*, 17(1):13, 2016.

The Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res*, 47(D1):D330–D338, 2019.

FG Court, SA Wemyss-Holden, CP Morrison, BD Teague, PE Laws, J Kew, AR Dennison, and GJ Maddern. Segmental nature of the porcine liver and its potential as a model for experimental partial hepatectomy. *Brit j surg*, 90(4):440–444, 2003.

J Cox and M Mann. Maxquant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12):1367, 2008.

J Cox and M Mann. 1d and 2d annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC bioinformatics*, 13(16):S12, 2012.

J Cui, Y Chen, HY Wang, and RF Wang. Mechanisms and pathways of innate immune activation and regulation in health and cancer. *Hum vaccin & immunother*, 10(11): 3270–3285, 2014.

LA Cynober. Plasma amino acid levels with a note on membrane transport: characteristics, regulation, and metabolic significance. *Nutrition*, 18(9):761–766, 2002.

A Dabney, JD Storey, and GR Warnes. qvalue: Q-value estimation for false discovery rate control. *R package version*, 1(0), 2010.

K De Bock, M Georgiadou, S Schoors, A Kuchnio, BW Wong, AR Cantelmo, A Quaegebeur, B Ghesquiere, S Cauwenberghs, G Eelen, et al. Role of pfkfb3-driven glycolysis in vessel sprouting. *Cell*, 154(3):651–663, 2013.

C De Sousa, NR English, TE Stacey, and RA Chalmers. Measurement of l-carnitine and acylcarnitines in body fluids and tissues in children and in adults. *Clin Chimi Acta*, 187 (3):317–328, 1990.

K Dettmer, PA Aronov, and BD Hammock. Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, 26(1):51–78, 2007.

G Dimitriadis, P Mitrou, V Lambadiari, E Maratou, and SA Raptis. Insulin effects in muscle and adipose tissue. *Diab res and clin prac*, 93:S52–S59, 2011.

TL Dinh and A Veves. A review of the mechanisms implicated in the pathogenesis of the diabetic foot. *Intern j low extrem wounds*, 4(3):154–159, 2005.

A Dobin, CA Davis, F Schlesinger, J Drenkow, C Zaleski, S Jha, P Batut, M Chaisson, and TR Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

J Dooley, L Tian, S Schonefeldt, V Delghingaro-Augusto, JE Garcia-Perez, E Pasciuto, D Di Marino, EJ Carr, N Oskolkov, V Lyssenko, et al. Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nature genetics*, 48(5):519, 2016.

D Dufrane, M van Steenberghe, Y Guiot, RM Goebbels, A Saliez, and P Gianello. Streptozotocin-induced diabetes in large animals (pigs/primates): role of glut2 transporter and β-cell plasticity. *Transplantation*, 81(1):36–45, 2006.

S Durinck, Y Moreau, A Kasprzyk, S Davis, B De Moor, A Brazma, and W Huber. Biomart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16):3439–3440, 2005.

B Efron. Local false discovery rates, 2005.

DL Eizirik, AK Cardozo, and M Cnop. The role for endoplasmic reticulum stress in diabetes mellitus. *Endoc rev*, 29(1):42–61, 2007.

CS Ejsing, JL Sampaio, V Surendranath, E Duchoslav, K Ekroos, RW Klemm, K Simons, and A Shevchenko. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *PNAS*, 106(7):2136–2141, 2009.

G Eknoyan and J Nagy. A history of diabetes mellitus or how a disease of the kidneys evolved into a kidney disease. *Adv Chronic Kidney Dis*, 12(2):223–9, 2005.

JE Elias, W Haas, BK Faherty, and SP Gygi. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature methods*, 2(9):667, 2005.

P Ewels, M Magnusson, S Lundin, and M Käller. Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016.

J Folch, M Lees, and GHS Stanley. A simple method for the isolation and purification of total lipides from animal tissues. *J biol chem*, 226(1):497–509, 1957.

Michael J Fowler. Microvascular and macrovascular complications of diabetes. *Clinical Diabetes*, 26(2):77–82, 2008.

ES Fox, P Thomas, and SA Broitman. Comparative studies of endotoxin uptake by isolated rat kupffer and peritoneal cells. *Infection and immunity*, 55(12):2962–2966, 1987.

T Fröhlich, E Kemter, F Flenkenthaler, N Klymiuk, KA Otte, A Blutke, S Krause, MC Walter, R Wanke, E Wolf, et al. Progressive muscle proteome changes in a clinically relevant pig model of duchenne muscular dystrophy. *Scientific reports*, 6:33362, 2016.

A Geerts. History, heterogeneity, developmental biology, and functions of quiescent hepatic stellate cells. In *Seminars in liver disease*, volume 21, pages 311–336. Copyright© 2001 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ..., 2001.

WB Geho. The importance of the liver in insulin replacement therapy in insulin-deficient diabetes. *Diabetes*, 63(5):1445–1447, 2014.

WB Geho, HC Geho, JR Lau, and TJ Gana. Hepatic-directed vesicle insulin: A review of formulation development and preclinical evaluation. *J Diab Scien and Techn*, 3(6): 1451–1459, 2009.

B Giardine, C Riemer, RC Hardison, R Burhans, L Elnitski, P Shah, Y Zhang, D Blankenberg, I Albert, J Taylor, et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome research*, 15(10):1451–1455, 2005.

C Girardot, J Scholtalbers, S Sauer, SY Su, and EEM Furlong. Je, a versatile suite to handle multiplexed ngs libraries with unique molecular identifiers. *BMC bioinformatics*, 17(1):419, 2016.

JJ Goeman and A Solari. Multiple hypothesis testing in genomics. *Statistics in medicine*, 33(11):1946–1978, 2014.

S Greenland, SJ Senn, KJ Rothman, JB Carlin, C Poole, SN Goodman, and DG Altman. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur j epidem*, 31(4):337–350, 2016.

V Grossmann, VH Schmitt, T Zeller, M Panova-Noeva, A Schulz, D Laubert-Reh, C Juenger, RB Schnabel, TGJ Abt, R Laskowski, et al. Profile of the immune and inflammatory response in individuals with prediabetes and type 2 diabetes. *Diabetes Care*, 38(7):1356–1364, 2015.

HAPO Study Cooperative Research Group. The hyperglycemia and adverse pregnancy outcome (hapo) study. *Intern J Gynecol & Obstet*, 78(1):69–77, 2002.

M Guasch-Ferré, A Hruby, E Toledo, CB Clish, MA Martínez-González, J Salas-Salvadó, and FB Hu. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes care*, 39(5):833–846, 2016.

Z Han, J Guo, SM Conley, and MI Naash. Retinal angiogenesis in the ins2(akita) mouse model of diabetic retinopathy. *Invest Ophthalmol Vis Sci*, 54(1):574–84, 2013.

P Harper, C Wadström, and G Cederblad. Carnitine measurements in liver, muscle tissue, and blood in normal subjects. *Clinical chemistry*, 39(4):592–599, 1993.

Frank E Harrell Jr. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

AL Harte, NF da Silva, SJ Creely, KC McGee, T Billyard, EM Youssef-Elabd, G Tripathi, E Ashour, MS Abdalla, HM Sharada, et al. Elevated endotoxin levels in non-alcoholic fatty liver disease. *J inflam*, 7(1):15, 2010.

NL Hawley and ST McGarvey. Obesity and diabetes in pacific islanders: the current burden and the need for urgent action. *Curr Diab Rep*, 15(5):29, 2015.

JD Hayes, JU Flanagan, and IR Jowsey. Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.*, 45:51–88, 2005.

R Herzog, D Schwudke, K Schuhmann, JL Sampaio, SR Bornstein, M Schroeder, and A Shevchenko. A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome biology*, 12(1):R8, 2011.

A Hinrichs, B Kessler, M Kurome, A Blutke, E Kemter, M Bernau, AM Scholz, B Rathkolb, S Renner, S Bultmann, et al. Growth hormone receptor-deficient pigs resemble the pathophysiology of human laron syndrome and reveal altered activation of signaling cascades in the liver. *Molecular metabolism*, 11:113–128, 2018.

EG Hong, DY Jung, HJ Ko, Z Zhang, Z Ma, JY Jun, JH Kim, AD Sumner, TC Vary, TW Gardner, et al. Nonobese, insulin-deficient ins2akita mice develop type 2 diabetes phenotypes including insulin resistance and cardiac remodeling. *Amer J Physiol-Endocrinol and Metab*, 293(6):E1687–E1696, 2007.

RP Horgan and LC Kenny. 'omic'technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13(3):189–195, 2011.

JD Horton, JL Goldstein, and MS Brown. Srebps: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J clin invest*, 109(9):1125–1131, 2002.

I Horvathova, F Voigt, AV Kotrys, Y Zhan, CG Artus-Revel, J Eglinger, MB Stadler, L Giorgetti, and JA Chao. The dynamics of mrna turnover revealed by single-molecule imaging in single cells. *Molecular cell*, 68(3):615–625, 2017.

FB Hu. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes care*, 34 (6):1249–1257, 2011.

S Huang, K Chaudhary, and LX Garmire. More is better: recent progress in multi-omics data integration methods. *Frontiers in genetics*, 8:84, 2017.

XF Huang and P Arvan. Formation of the insulin-containing secretory granule core occurs within immature beta-granules. *J biol chem*, 269(33):20838–20844, 1994.

M Itsumi, S Inoue, AJ Elia, K Murakami, M Sasaki, EF Lind, D Brenner, IS Harris, IIC Chio, S Afzal, et al. Idh1 protects murine hepatocytes from endotoxin-induced oxidative stress by regulating the intracellular nadp+/nadph ratio. *Cell death and differentiation*, 22(11):1837, 2015.

H Jaeschke, MR McGill, and A Ramachandran. Oxidant stress, mitochondria, and cell death mechanisms in drug-induced liver injury: lessons learned from acetaminophen hepatotoxicity. *Drug metab rev*, 44(1):88–106, 2012.

X Jin, K Chen, Z Lv, L Zheng, Y Liu, S Chen, C Yu, X Jiang, C Zhang, Y Li, et al. Hdmcp uncouples yeast mitochondrial respiration and alleviates steatosis in l02 and hepg2 cells by decreasing atp and h2o2 levels: a novel mechanism for nafld. *J hepat*, 50 (5):1019–1028, 2009.

G Joshi-Tope, M Gillespie, I Vastrik, P D'Eustachio, E Schmidt, B de Bono, B Jassal, GR Gopinath, GR Wu, L Matthews, S Lewis, E Birney, and L Stein. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*, 33(Database issue):D428–32, 2005.

C Ju, JP McCoy, CJ Chung, MLM Graf, and LR Pohl. Tolerogenic role of kupffer cells in allergic reactions. *Chemical research in toxicology*, 16(12):1514–1519, 2003.

R Kaaja and T Ronnemaa. Gestational diabetes: pathogenesis and consequences to mother and offspring. *Rev Diabet Stud*, 5(4):194–202, 2008.

GJ Kahaly and MP Hansen. Type 1 diabetes associated autoimmunity. *Autoimmun Rev*, 15(7):644–8, 2016.

U Kampmann, L R Madsen, GO Skajaa, DS Iversen, N Moeller, and P Ovesen. Gestational diabetes: A clinical update. *World J Diabetes*, 6(8):1065–72, 2015.

KV Kandror and PF Pilch. Compartmentalization of protein traffic in insulin-sensitive cells. *Am J Physiol*, 271(1 Pt 1):E1–14, 1996.

M Kanehisa and S Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, 2000.

AM Kay, CL Simpson, and Jr. Stewart, JA. The role of age/rage signaling in diabetes-mediated vascular calcification. *J Diabetes Res*, 2016:6809703, 2016.

Y Kido, DJ Burks, D Withers, JC Bruning, CR Kahn, MF White, and D Accili. Tissue-specific insulin resistance in mice with mutations in the insulin receptor, irs-1, and irs-2. *J clin invest*, 105(2):199–205, 2000.

D Kim, G Pertea, C Trapnell, H Pimentel, R Kelley, and SL Salzberg. Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14(4):R36, 2013.

MK Kim, HS Kim, IK Lee, and KG Park. Endoplasmic reticulum stress and insulin biosynthesis: a review. *Experimental diabetes research*, 2012, 2012.

AJF King. The use of animal models in diabetes research. *Brit j pharmacol*, 166(3): 877–894, 2012.

M Kitada, Y Ogura, and D Koya. Rodent models of diabetic nephropathy: their utility and limitations. *Int J Nephrol Renovasc Dis*, 9:279–290, 2016.

PJ Klover and RA Mooney. Hepatocytes: critical for glucose homeostasis. *Intern j biochem & cell biol*, 36(5):753–758, 2004.

DC Koboldt, KM Steinberg, DE Larson, RK Wilson, and ER Mardis. The next-generation sequencing revolution and its impact on genomics. *Cell*, 155(1):27–38, 2013.

Y Koyama and DA Brenner. Liver inflammation and fibrosis. *J clin invest*, 127(1):55–64, 2017.

CR Krois, MG Vuckovic, P Huang, C Zaversnik, CS Liu, CE Gibson, MR Wheeler, KM Obrochta, JH Min, CB Herber, et al. Rdh1 suppresses adiposity by promoting brown adipose adaptation to fasting and re-feeding. *Cellular and Molecular Life Sciences*, 76(12):2425–2447, 2019.

M Kurome, B Kessler, A Wuensch, H Nagashima, and E Wolf. Nuclear transfer and transgenesis in the pig. In *Nuclear Reprogramming*, pages 37–59. Springer, 2015.

J Leandro and SM Houten. Saccharopine, a lysine degradation intermediate, is a mitochondrial toxin. *J Cell Biol*, 218(2):391–392, 2019.

Y Lee and DC Rio. Mechanisms and regulation of alternative pre-mrna splicing. *Ann Rev Biochem*, 84(1):291–323, 2015.

JT Leek and JD Storey. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*, 3(9):1724–35, 2007.

P Lefebvre, F Lalloyer, E Baugé, M Pawlak, C Gheeraert, H Dehondt, J Vanhoutte, E Woitrain, N Hennuyer, C Mazuy, et al. Interspecies nash disease activity whole-genome profiling identifies a fibrogenic role of pparα-regulated dermatopontin. *JCI insight*, 2(13), 2017.

B Li, V Ruotti, RM Stewart, JA Thomson, and CN Dewey. Rna-seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26(4):493–500, 2010.

A Liberzon, A Subramanian, R Pinchback, H Thorvaldsdóttir, P Tamayo, and JP Mesirov. Molecular signatures database (msigdb) 3.0. *Bioinformatics*, 27(12):1739–1740, 2011.

M Liu, I Hodish, L Haataja, R Lara-Lemus, G Rajpal, J Wright, and P Arvan. Proinsulin misfolding and diabetes: mutant ins gene-induced diabetes of youth. *Trends Endocrinol Metab*, 21(11):652–9, 2010.

Y Liu, A Beyer, and R Aebersold. On the dependency of cellular protein levels on mrna abundance. *Cell*, 165(3):535–550, 2016.

JW Locasale. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat Rev Cancer*, 13(8):572, 2013.

N Longo, M Frigeni, and M Pasquali. Carnitine transport and fatty acid oxidation. *Biochimica et biophysica acta (BBA)-molecular cell research*, 1863(10):2422–2435, 2016.

MI Love, W Huber, and S Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol*, 15(12):550, 2014.

L Lu, H Zhou, M Ni, X Wang, R Busuttil, J Kupiec-Weglinski, and Y Zhai. Innate immune regulations and liver ischemia reperfusion injury. *Transplantation*, 100(12):2601, 2016.

SC Lu. Glutathione synthesis. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1830(5):3143–3153, 2013.

TA Lydic and YH Goo. Lipidomics unveils the complexity of the lipidome in metabolic diseases. *Clin Transl Med*, 7(1):4, 2018.

RJ Mailloux and ME Harper. Uncoupling proteins and the control of mitochondrial reactive oxygen species production. *Free Radic Biol Med*, 51(6):1106–15, 2011a.

RJ Mailloux and ME Harper. Uncoupling proteins and the control of mitochondrial reactive oxygen species production. *Free Radical Biology and Medicine*, 51(6):1106–1115, 2011b.

P Maiuri, A Knezevich, A De Marco, D Mazza, A Kula, JG McNally, and A Marcello. Fast transcription rates of rna polymerase ii in human cells. *EMBO reports*, 12(12): 1280–1285, 2011.

RA Malik. Pathology of human diabetic neuropathy. *Handb Clin Neurol*, 126:249–59, 2014.

J Marill, N Idres, CC Capron, E Nguyen, and GG Chabot. Retinoic acid metabolism and mechanism of action: a review. *Current drug metabolism*, 4(1):1–10, 2003.

RG Marques, MJ Fontaine, and J Rogers. C-peptide: much more than a byproduct of insulin biosynthesis. *Pancreas*, 29(3):231–238, 2004.

MV Martinov, VM Vitvitsky, R Banerjee, and FI Ataullakhanov. The logic of the hepatic methionine metabolic cycle. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804(1):89–96, 2010.

T Matsumura, FC Kauffman, H Meren, and RG Thurman. O2 uptake in periportal and pericentral regions of liver lobule in perfused liver. *Amer J Physiol-Gastroint and Liv Physiol*, 250(6):G800–G805, 1986.

AV Matveyenko and PC Butler. Relationship between beta-cell mass and diabetes onset. *Diabetes Obes Metab*, 10 Suppl 4:23–31, 2008.

RC McEvoy. Changes in the volumes of the a-, b-, and d-cell populations in the pancreatic islets during the postnatal development of the rat. *Diabetes*, 30(10):813–817, 1981.

F Medici, M Hawa, A Ianari, DA Pyke, and RD Leslie. Concordance rate for type ii diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia*, 42(2):146–50, 1999.

AH Merrill Jr, E Wang, and RE Mullins. Kinetics of long-chain (sphingoid) base biosynthesis in intact lm cells: effects of varying the extracellular concentrations of serine and fatty acid precursors of this pathway. *Biochemistry*, 27(1):340–345, 1988.

RG Miller, AM Secrest, RK Sharma, TJ Songer, and TJ Orchard. Improvements in the life expectancy of type 1 diabetes: the pittsburgh epidemiology of diabetes complications study cohort. *Diabetes*, 61(11):2987–92, 2012.

J Mohamed, AHN Nafizah, AH Zariyantey, and SB Budin. Mechanisms of diabetes-induced liver damage: the role of oxidative stress and inflammation. *Sultan Qaboos University Medical Journal*, 16(2):e132, 2016.

P Morel, DB Kaufmann, AJ Matas, P Tzardis, MJ Field, JK Lloveras, and DE Sutherland. Total pancreatectomy in the pig for islet transplantation. technical alternatives. *Transplantation*, 52(1):11–15, 1991.

E Mormone, J George, and N Nieto. Molecular pathogenesis of hepatic fibrosis and current therapeutic approaches. *Chemico-biological interactions*, 193(3):225–231, 2011.

SM Morris Jr. Regulation of enzymes of the urea cycle and arginine metabolism. *Annual review of nutrition*, 22(1):87–105, 2002.

E Mosharov, MR Cranford, and R Banerjee. The quantitatively important relationship between homocysteine metabolism and glutathione synthesis by the transsulfuration pathway and its regulation by redox changes. *Biochemistry*, 39(42):13005–13011, 2000.

M Mueckler. Insulin resistance and the disruption of glut4 trafficking in skeletal muscle. *J Clin Invest*, 107(10):1211–3, 2001.

JL Napoli. Physiological insights into all-trans-retinoic acid biosynthesis. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1821(1):152–167, 2012.

SR Narum. Beyond bonferroni: less conservative analyses for conservation genetics. *Conservation genetics*, 7(5):783–787, 2006.

JA Nelder and RWM Wedderburn. Generalized linear models. *J Royal Stat Soc*, 135(3): 370–384, 1972.

R Nesto. C-reactive protein, its role in inflammation, type 2 diabetes and cardiovascular disease, and the effects of insulin-sensitizing treatment with thiazolidinediones. *Diabetic medicine*, 21(8):810–817, 2004.

JC Newman and E Verdin. Ketone bodies as signaling metabolites. *Trends in Endocrinology & Metabolism*, 25(1):42–52, 2014.

JK Nicholson and ID Wilson. Understanding'global'systems biology: metabonomics and the continuum of metabolism. *Nature Reviews Drug Discovery*, 2(8):668, 2003.

D Nishimura. Biocarta. *Comp Softw J for Scient*, 2(3):117–120, 2001.

KM Obrochta, CR Krois, B Campos, and JL Napoli. Insulin regulates retinol dehydrogenase expression and all-trans-retinoic acid biosynthesis through foxo1. *J Biol Chem*, 290 (11):7259–68, 2015.

KJ Oh, HS Han, MJ Kim, and SH Koo. Transcriptional regulators of hepatic gluconeogenesis. *Arch pharmacal res*, 36(2):189–200, 2013.

R Raghow, C Yellaturu, X Deng, EA Park, and MB Elam. Srebps: the crossroads of physiological and pathological lipid homeostasis. *Trends endocrin & metab*, 19(2):65–73, 2008.

SE Regnell and A Lernmark. Hepatic steatosis in type 1 diabetes. *The review of diabetic studies: RDS*, 8(4):454, 2011.

S Renner, A Blutke, B Dobenecker, G Dhom, TD Müller, B Finan, C Clemmensen, M Bernau, I Novak, B Rathkolb, et al. Metabolic syndrome and extensive adipose tissue inflammation in morbidly obese göttingen minipigs. *Molecular metabolism*, 16: 180–190, 2018.

SE Reuter and AM Evans. Carnitine and acylcarnitines. *Clinical pharmacokinetics*, 51(9): 553–572, 2012.

M Rhinn, B Schuhbaur, K Niederreither, and P Dollé. Involvement of retinol dehydrogenase 10 in embryonic patterning and rescue of its loss of function by maternal retinaldehyde treatment. *PNAS*, 108(40):16687–16692, 2011.

ME Ritchie, B Phipson, D Wu, Y Hu, CW Law, W Shi, and GK Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

PJ Roach, AA Depaoli-Roach, TD Hurley, and VS Tagliabracci. Glycogen and its metabolism: some new developments and old themes. *Biochem Jo*, 441(3):763–787, 2012.

JT Robinson, H Thorvaldsdóttir, W Winckler, M Guttman, ES Lander, G Getz, and JP Mesirov. Integrative genomics viewer. *Nature biotechnology*, 29(1):24, 2011.

MD Robinson and GK Smyth. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23(21):2881–2887, 09 2007. ISSN 1367-4803.

MD Robinson, DJ McCarthy, and GK Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.

F Rohart, B Gautier, A Singh, and KA Le Cao. mixomics: An r package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*, 13(11):e1005752, 2017.

KI Rother, Y Imai, M Caruso, F Beguinot, P Formisano, and D Accili. Evidence that irs-2 phosphorylation is required for insulin action in hepatocytes. *J biol chem*, 273(28): 17491–17497, 1998.

DW Russell. The enzymes, regulation, and genetics of bile acid synthesis. *Ann rev biochem*, 72(1):137–174, 2003.

JL Sampaio, MJ Gerl, C Klose, CS Ejsing, H Beug, K Simons, and A Shevchenko. Membrane lipidome of an epithelial cell line. *PNAS*, 108(5):1903–1907, 2011.

JI Sbodio, SH Snyder, and BD Paul. Regulators of the transsulfuration pathway. *Brit j pharmacol*, 176(4):583–593, 2019.

CF Schaefer, K Anthony, S Krupa, J Buchoff, M Day, T Hannay, and KH Buetow. Pid: the pathway interaction database. *Nucleic acids research*, 37(suppl_1):D674–D679, 2008.

JM Scott and DG Weir. Folic acid, homocysteine and one-carbon metabolism: a review of the essential biochemistry. *J cardiovas risk*, 5(4):223–227, 1998.

M Scott, CW Gunderson, EM Mateescu, Z Zhang, and T Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.

JP Shaffer. Multiple hypothesis testing. *Annual review of psychology*, 46(1):561–584, 1995.

P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, and T Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.

X Shu, L Nelbach, RO Ryan, and TM Forte. Apolipoprotein av associates with intrahepatic lipid droplets and influences triglyceride accumulation. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1801(5):605–608, 2010.

D Smedley, S Haider, B Ballester, R Holland, D London, G Thorisson, and A Kasprzyk. Biomart–biological queries made easy. *BMC genomics*, 10(1):22, 2009.

E Streckel, C Braun-Reichhart, N Herbach, M Dahlhoff, B Kessler, A Blutke, A Bähr, N Übel, M Eddicks, M Ritzmann, et al. Effects of the glucagon-like peptide-1 receptor agonist liraglutide in juvenile transgenic pigs modeling a pre-diabetic condition. *J translat med*, 13(1):73, 2015.

K Strimmer. fdrtool: a versatile r package for estimating local and tail area-based false discovery rates. *Bioinformatics*, 24(12):1461–2, 2008.

A Subramanian, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub, ES Lander, and JP Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–50, 2005.

YV Sun and YJ Hu. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv Genet*, 93:147–90, 2016.

G Szabo. Gut–liver axis in alcoholic liver disease. *Gastroenterology*, 148(1):30–36, 2015.

AL Tarca, G Bhatti, and R Romero. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. *PloS one*, 8(11):e79217, 2013.

G Targher, A Lonardo, and CD Byrne. Nonalcoholic fatty liver disease and chronic vascular complications of diabetes mellitus. *Nature Reviews Endocrinology*, 14(2):99, 2018.

RStudio Team et al. Rstudio: integrated development for r. *RStudio, Inc., Boston, MA*, 42:14, 2015.

AW Thomson and PA Knolle. Antigen-presenting cell function in the tolerogenic liver environment. *Nature Reviews Immunology*, 10(11):753, 2010.

H Thorvaldsdóttir, JT Robinson, and JP Mesirov. Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2):178–192, 2013.

H Tilg, AR Moschen, and M Roden. Nafld and diabetes mellitus. *Nature reviews Gastroenterology & hepatology*, 14(1):32, 2017.

P Trayhurn, NJ Temple, and JV Aerde. Evidence from immunoblotting studies on uncoupling protein that brown adipose tissue is not present in the domestic pig. *Canad j physiol and pharmac*, 67(12):1480–1485, 1989.

J Tuomilehto. The emerging global epidemic of type 1 diabetes. *Curr Diab Rep*, 13(6): 795–804, 2013.

S Tyanova, T Temu, P Sinitcyn, A Carlson, MY Hein, T Geiger, M Mann, and journal=Nature methods volume=13 number=9 pages=731 year=2016 publisher=Nature Publishing Group Cox, J. The perseus computational platform for comprehensive analysis of (prote) omics data.

B Tyrberg, A Andersson, and LAH Borg. Species differences in susceptibility of transplanted and cultured pancreatic islets to the β-cell toxin alloxan. *General and comparative endocrinology*, 122(3):238–251, 2001.

M Uhlén, L Fagerberg, BM Hallström, C Lindskog, P Oksvold, A Mardinoglu, Å Sivertsson, C Kampf, E Sjöstedt, A Asplund, IM Olsson, K Edlund, E Lundberg, S Navani, CAK Szigyarto, J Odeberg, D Djureinovic, JO Takanen, S Hober, T Alm, PE Edqvist, H Berling, H Tegel, J Mulder, J Rockberg, P Nilsson, JM Schwenk, M Hamsten, K von Feilitzen, M Forsberg, L Persson, F Johansson, M Zwahlen, G von Heijne, J Nielsen, and F Pontén. Tissue-based map of the human proteome. *Science*, 347(6220). ISSN 0036-8075.

E Van Cauter, F Mestrez, J Sturis, and KS Polonsky. Estimation of insulin secretion rates from c-peptide levels: comparison of individual and standard kinetic parameters for c-peptide clearance. *Diabetes*, 41(3):368–377, 1992.

PJ Van Den Elsen. Expression regulation of major histocompatibility complex class i and class ii encoding genes. *Frontiers in immunology*, 2:48, 2011.

TC Vary and CJ Lynch. Nutrient signaling components controlling protein synthesis in striated muscle. *J nutrition*, 137(8):1835–1843, 2007.

DF Vatner, SK Majumdar, N Kumashiro, MC Petersen, Y Rahimi, AK Gattu, M Bears, JPG Camporez, GW Cline, MJ Jurczak, et al. Insulin-independent regulation of hepatic triglyceride synthesis by fatty acids. *PNAS*, 112(4):1143–1148, 2015.

FM Vaz, R Ofman, K Westinga, JW Back, and RJA Wanders. Molecular and biochemical characterization of rat ε-n-trimethyllysine hydroxylase, the first enzyme of carnitine biosynthesis. *J biol chem*, 276(36):33512–33517, 2001.

K Vithian and S Hurel. Microvascular complications: pathophysiology and management. *Clin Med (Lond)*, 10(5):505–9, 2010.

J Wang, T Takeuchi, S Tanaka, SK Kubo, T Kayo, D Lu, K Takata, A Koizumi, and T Izumi. A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the mody mouse. *J Clin Invest*, 103(1):27–37, 1999.

M Wang, C Wang, RH Han, and X Han. Novel advances in shotgun lipidomics for biology and medicine. *Progress in lipid research*, 61:83–108, 2016.

Y Wang, J Shao, JL Zaro, and WC Shen. Proinsulin-transferrin fusion protein as a novel long-acting insulin analog for the inhibition of hepatic glucose production. *Diabetes*, 63 (5):1779–1788, 2014.

N Waseem and PH Chen. Hypoxic hepatitis: a review and clinical update. *J of clin and transl hepat*, 4(3):263, 2016.

W Weckwerth. Metabolomics in systems biology. *Annual review of plant biology*, 54(1): 669–689, 2003.

MA Weiss. Proinsulin and the genetics of diabetes mellitus. *J biol chem*, 284(29):19159–19163, 2009.

MF White. The irs-signalling system: a network of docking proteins that mediate insulin action. In *Insulin Action*, pages 3–11. Springer, 1998.

EP Widmaier, H Raff, KT Strang, and AJ Vander. *Vander's Human physiology: the mechanisms of body function.* Boston: McGraw-Hill Higher Education„ 2008.

E Wolf, C Braun-Reichhart, E Streckel, and S Renner. Genetically engineered pig models for diabetes research. *Transgenic research*, 23(1):27–38, 2014.

C Wolfrum, D Besser, E Luca, and M Stoffel. Insulin regulates the activity of forkhead transcription factor hnf-3β/foxa-2 by akt-mediated phosphorylation and nuclear/cytosolic localization. *PNAS*, 100(20):11624–11629, 2003.

G Wu, YZ Fang, S Yang, JR Lupton, and ND Turner. Glutathione metabolism and its implications for health. *J nutrition*, 134(3):489–492, 2004.

HH Xue, M Fujie, T Sakaguchi, T Oda, H Ogawa, NM Kneer, HA Lardy, and A Ichiyama. Flux of the l-serine metabolism in rat liver the predominant contribution of serine dehydratase. *J biol chem*, 274(23):16020–16027, 1999.

B Yates, B Braschi, KA Gray, RL Seal, S Tweedie, and EA Bruford. Genenames. org: the hgnc and vgnc resources in 2017. *Nucleic acids research*, page gkw1033, 2016.

MS Yoon. The emerging role of branched-chain amino acids in insulin resistance and metabolism. *Nutrients*, 8(7):405, 2016.

M Yoshioka, T Kayo, T Ikeda, and A Koizumi. A novel locus, mody4, distal to d7mit189 on chromosome 7 determines early-onset niddm in nonobese c57bl/6 (akita) mutant mice. *Diabetes*, 46(5):887–94, 1997.

M Zhang, W Chen, SM Smith, and JL Napoli. Molecular characterization of a mouse short chain dehydrogenase/reductase active with all-trans-retinol in intact cells, mrdh1. *J biol chem*, 276(47):44083–44090, 2001.

M Zhang, P Hu, CR Krois, MA Kane, and JL Napoli. Altered vitamin a homeostasis and increased size and adiposity in the rdh1-null mouse. *The FASEB Journal*, 21(11): 2886–2896, 2007.

S Zukunft, M Sorgenfrei, C Prehn, G Möller, and J Adamski. Targeted metabolomics of dried blood spot extracts. *Chromatographia*, 76(19):1295–1305, 2013.

S Zukunft, C Prehn, C Rohring, G Moller, M Hrabe de Angelis, J Adamski, and J Tokarz. High-throughput extraction and quantification method for targeted metabolomics in murine tissues. *Metabolomics*, 14(1):18, 2018.

Acknowledgements

First of all I am very grateful to Prof. Dr. Eckhard Wolf for supervising my doctoral thesis, and for his help and advice during my time in Großhadern.

Thank you to the people of the Blum lab for housing me during the last four years. Thank you especially to Dr. Helmut Blum also for your help and advice in matters of science and RNA sequencing. The same thanks goes to Dr. Stefan Krebs and Dr. Julia Philippou-Massier, who provided me with data and advice, and to Dr. Alexander Graf for his help with thesis writing and IT.

Thank you to the graduate school of Quantitative Biosciences Munich (QBM) for funding my doctoral thesis and for providing company in the form of fellow doctoral students. For creating this program I give special thanks to Prof. Dr. Ulrike Gaul. Thank you to the QBM team for organizing scientific and social gatherings, and for helping with navigating the German bureaucracy.

I want to give great thanks to all my collaborators during my thesis. Special thanks goes to Dr. Florian Flenkenthaler for a successful collaboration and for his help and experience in paper publishing. I also want to thank Erik Ländström for many philosophical and scientific discussions around our parallel projects.

Last but not least I want to thank my girlfriend Parastou and my family for the support they have given me during these four years.