Institute for Molecular Medicine Finland, FIMM
University of Helsinki
Helsinki, Finland

MOLECULAR EFFECTS OF OBESITY AND RELATED METABOLIC RISK FACTORS – A TRANSCRIPTOMICS AND METABOLOMICS APPROACH

Maheswary Muniandy

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Lecture Hall 1, Haartman Institute, Haartmaninkatu 3, on 8 June 2018 at noon.

Helsinki, Finland 2018

ISBN 978-951-51-4278-8 (pbk.)
ISBN 978-951-51-4279-5 (PDF)

Unigraphia Oy
Helsinki, Finland 2018

Supervisors	Adjunct Professor Miina Ollikainen Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland Department of Public Health, University of Helsinki, Helsinki, Finland
	Professor Kirsi Pietiläinen Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland
Reviewers	Professor Marjukka Kolehmainen Public Health and Clinical Nutrition, University of Eastern Finland, Finland
	Adjunct Professor Laura Elo Research Director, Computational Biomedicine and Bioinformatics Vice Director, Turku Centre for Biotechnology University of Turku and Åbo Akademi University, Finland
Opponent	Assistant Professor Sara Hägg Docent in Molecular Epidemiology Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

“Take nothing on its looks; take everything on evidence. There's no better rule.”
— Charles Dickens, Great Expectations

For my precious little one

ABSTRACT

Obesity remains a major health problem, partly due to our limited understanding of this complex disease. Obesity carries with it the risk of many other diseases including type 2 diabetes, cardiovascular disease, hyperlipidemia and some types of cancer. The variability in the disease as well as its related comorbidities makes it a complex, multi-factorial condition that is not easily categorised and treated.

'Omics technologies and bioinformatics tools allow for the investigation of the complex biology behind obesity. These technologies enable production of complex multivariate datasets that can be investigated using bioinformatics tools to identify patterns in the data as well as associations between different features of the data. However, while advances in 'omics technologies have allowed production of large amounts of data from biological samples, extraction of useful information from the data remains a huge challenge. Choosing the correct methodology and tools to transform heterogeneous data into biological knowledge is especially difficult when different methods on the same data may yield different results, requiring further statistical or biological validation.

This thesis uses existing bioinformatics tools and methods to first combine and analyse transcriptomics and biochemical data and then, separately, metabolomics and biochemical data to gain an understanding of obesity. Body mass index (BMI)-discordant as well as BMI-concordant monozygotic (MZ) twin pairs were used to investigate the molecular effects of obesity by looking at gene expression and metabolite profiles in subcutaneous adipose tissue (SAT) and blood plasma, respectively, to gain biological insights into pathways that are associated with obesity and obesity-related clinical manifestations. The SAT was further interrogated using isolated adipocytes, to examine the transcriptomics patterns in obesity of this specific cell type. Using the blood plasma, metabolites associating with different cardiometabolic risk factors were also identified. Variations in the global profiles were also studied to assess if study participants form different subgroups of obesity according to their gene expression or metabolite profiles. Adiposity and blood biochemistry measure differences between these obesity subgroups were also examined.

In the first study, using microarray technology and within-twin pair differential analysis, downregulation of mitochondria-related pathways and upregulation of inflammation pathways in the SAT of heavy compared to lean co-twin within the twin pairs were identified. Because these within-twin pair differences are not due to genetic effects, these findings represented the effects of acquired obesity, pointing to differences in environmental effects (e.g. aspects of diet, exercise and lifestyle) between the co-twins. Three subgroups of acquired obesity were identified, each group showing distinct within-twin pair differences. Each of these groups represented different profiles of acquired obesity, with one group showing benign effects of obesity, the second group showing downregulation of mitochondrial functions and the third group showing downregulation of mitochondrial function and upregulation of inflammation in the heavy co-twin. The third group also showed significantly

higher fasting insulin and larger adipocyte diameter when comparing the heavier to leaner person within the twin pairs and hence represented the unhealthiest acquired obesity group compared to the other two groups. This study confirms that not all acquired obesities are the same and that identifying obesity subgroups and profiling them using clinical traits and gene expression is a feasible means of identifying these subgroups.

In the second study, using microarray technology and within-twin pair differential analysis in SAT and adipocytes, it was shown that most of the pathways attributed to acquired obesity in the SAT originate from the adipocytes. This study also showed that most of these pathways were mitochondria-related.

The third study, using mass spectrometry technology and linear regression analysis, investigated various adiposity and blood biochemistry measures and their associations with metabolites in the plasma. Of all the adiposity and blood biochemistry measures, high-density lipoprotein cholesterol (HDL-C) had the strongest association with associating metabolites. This finding highlights that (HDL-C)-associating metabolites are highly sensitive to even the smallest changes in HDL-C, making HDL-C a suitable measure of early changes in metabolic health. It was also confirmed that measures of SAT amount, visceral adipose tissue amount and liver fat percentage associate with metabolites that also associate with BMI and body fat percentage making BMI a suitable measure of adiposity. Lastly, two groups of people were identified according to their metabolite profiles. Out of the two, the unhealthy group showed higher levels of total cholesterol and low-density lipoprotein cholesterol (LDL-C). This study showed that metabolite profiles can be used to categorise people into different subgroups based on their metabolic health.

All three studies reveal pathways of mitochondrial downregulation and increased inflammation in obesity and together link these pathways to findings of insulin resistance, adipocyte size, total cholesterol and LDL-C. This research also confirms the variations in gene expression and metabolite profiles in obesity and suggests that mapping these profiles may help in fine-characterising obesity. This characterisation may pave the way to improved diagnostics and personalised obesity treatment. By employing existing bioinformatics methods, it was possible to first explore patterns in the data in an unrestrictive hypothesis-free manner in order to identify variations in obesity, and then identify the molecular effects of obesity using more targeted data modelling techniques.

Contents

ABSTRACT v
Contents vii
LIST OF ORIGINAL PUBLICATIONS..x
ABBREVIATIONS..xi
1. INTRODUCTION 1
2. REVIEW OF THE LITERATURE 3
2.1. Obesity.. 3
2.1.1. Metabolically healthy and unhealthy obesity 4
2.1.2. Insulin resistance 7
2.1.3. Lipid metabolism 7
2.1.4. Amino acid metabolism 8
2.2. Adipose tissue.. 9
2.3. Main fat depots in the body 10
2.4. Obesity-related adipose tissue dysfunction 13
2.4.1. Adipose tissue expansion in obesity.. 13
2.4.2. Adipocyte mitochondria in obesity 13
2.4.3. Adipose tissue inflammation in obesity 13
2.5. Twin study setting in obesity research.. 14
2.6. Transcriptomics 14
2.7. Metabolomics.. 17
2.8. Bioinformatics methods in the study of complex data 18
2.8.1. Data normalisation and transformation.. 19
2.8.2. Principal Component Analysis 20
2.8.3. Data clustering.. 20
2.8.4. Association analysis using linear mixed modelling.. 22
2.8.5. Differential analysis using moderated t-tests.. 22
2.9. Transcriptomics patterns in obesity 23
2.10. Metabolomics patterns of obesity in plasma 27
3. AIMS OF THE STUDY 30
4. MATERIALS AND METHODS.. 31
4.1. Subjects.. 31

4.1.1. FinnTwin16 and FinnTwin12 birth cohorts 31
4.1.2. Subjects in this study 32
4.2. Ethical considerations 34
4.3. Measures 34
4.3.1. Blood tests and clinical measurements 34
4.3.2. Subcutaneous adipose tissue and adipocyte gene expression 35
4.3.3. Plasma metabolite measures 36
4.4. Analysis techniques 36
4.5. Analysis methods common to the studies (Study I, II, III) 41
4.5.1. Linear and linear mixed model analysis (Study I, II, III) 41
4.5.2. Pathway Analysis (Study I, II, III) 43
4.5.3. Clinical measurements analysis (Study I, II, III) 43
5. RESULTS 44
5.1. Characteristics of the twins (I, II, III) 44
5.2. Correlations between phenotypes (I, II, III) 46
5.2.1. Correlation between the phenotype measures 46
5.3. Subcutaneous adipose tissue gene expression in acquired obesity (I) 48
5.3.1. Within-twin pair differences in transcriptomics patterns of SAT 48
5.3.2. Adipose tissue gene expression profiles in subgroups of the twin pairs (I) 49
5.3.3. Associations between cluster pathways and phenotype measures, based on the adipose tissue gene expression profile (I) 53
5.4. Differential gene expression in adipocytes and subcutaneous adipose tissue (II) 54
5.4.1. Pathways for differentially expressed genes within the BMI-discordant twin pairs, in the adipocytes (II) 54
5.4.2. Pathways for within-twin pairs differentially expressed genes in subcutaneous adipose tissue and adipocytes (II) 55
5.5. Plasma Metabolites (III) 58
5.5.1. Significant metabolite-phenotype associations (III) 58
5.5.2. Metabolite–phenotype associations confounding from shared factors in twins (III) 60
5.5.3. Shared and unique associations for metabolites with different phenotypes (III) 61
5.5.4. Effect size comparisons in metabolite-phenotype associations (III) 62
5.5.5. Plasma metabolite profiling identified two groups (III) 63
6. DISCUSSION 65
6.1. Methodological considerations 65

6.1.1. Transcriptomics data analysis 65
6.1.2. Regression analysis with metabolomics data 67
6.1.3. Linear regression and within-twin pair differential analysis 67
6.1.4. PCA before clustering in metabolomics 67
6.1.5. Clustering methods 68
6.2. Summary of the main findings 69
6.3. Adipocyte mitochondrial functions consistently implicated in obesity (Study I, II, III) 71
6.4. Branched chain amino acid consistently associated with obesity (Study I, II, III) 72
6.5. Adiposity and blood biochemistry measures associating with metabolic risk factors (Study III) 73
6.6. Biological pathways and clinical phenotypes of unhealthy groups (Study I, III) 75
7. STRENGTHS AND LIMITATIONS 79
8. CONCLUSION AND FUTURE PROSPECTS 81
ACKNOWLEDGEMENTS 83
REFERENCES 85

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I Muniandy M, Heinonen S, Yki-Jarvinen H, Hakkarainen A, Lundbom J, Lundbom N, et al. Gene expression profile of subcutaneous adipose tissue in BMI-discordant monozygotic twin pairs unravels molecular and clinical changes associated with sub-types of obesity. Int J Obes (Lond). 2017.

II Heinonen S, Muniandy M, Buzkova J, Mardinoglu A, Rodriguez A, Fruhbeck G, et al. Mitochondria-related transcriptional signature is downregulated in adipocytes in obesity: a study of young healthy MZ twins. Diabetologia. 2017; 60(1):169-81.

III Serum metabolites reveal distinct profiles associating with different metabolic risk factors in monozygotic twin pairs (submitted)
Muniandy M, Velagapudi V, Hakkarainen A, Lundbom J, Lundbom N, Kaprio J, Rissanen A, Pietiläinen K.H., Ollikainen M

The publications are referred to in the text by their Roman numerals. The original publications are reprinted at the end of this thesis with the permission of the copyright holders.

ABBREVIATIONS

^1H NMR	proton nuclear magnetic resonance
AA	amino acid
AAA	aromatic amino acid
acyl-CoA	acyl-coenzyme A
ANOVA	analysis of variance
BCAA	branched chain amino acid
BMI	body mass index
CDF	chip description file
cDNA	complementary deoxyribonucleic acid
DEXA	dual energy x-ray absorptiometry
DNA	deoxyribonucleic acid
ELISA	enzyme-linked immunosorbent assay
FA	fatty acids
FDR	False Discovery Rate
FFA	free fatty acid
GCRMA	GeneChip robust multiarray averaging
HDL-C	high density lipoprotein cholesterol
HOMA	homeostatic model assessment
hs-CRP	high-sensitivity C-reactive protein
CRP	C-reactive protein
IPA	Ingenuity® Pathway Analysis
IR	insulin resistance
LASSO	least absolute shrinkage and selection operator
LDL-C	low density lipoprotein cholesterol
LF	liver fat
MHO	metabolically healthy obese
MRI	magnetic resonance imaging
mRNA	messenger RNA
MS	mass spectrometry
MUO	metabolically unhealthy obese
MZ	monozygotic
NAFLD	non-alcoholic fatty liver disease
OGTT	oral glucose tolerance test
‘omics	areas of study in molecular biology including genomics, proteomics, transcriptomics, epigenomics, metabolomics
OXPHOS	oxidative phosphorylation system in the inner mitochondrial membrane
PCA	principal component analysis
PC	principal component
PCR	polymerase chain reaction
PLSDA	partial least square discriminate analysis
QC	quality control
RF	radio frequency
RMA	Robust Multi-array Average

RNA	ribonucleic acid
ROS	reactive oxygen species
RT-PCR	real-time polymerase chain reaction
SAT	subcutaneous adipose tissue
T2DM	type 2 diabetes mellitus
TCHOL	total cholesterol
TG	triglycerides
UPLC	ultra high performance liquid chromatography
VAT	visceral adipose tissue
VLDL	very-low-density lipoprotein

1. INTRODUCTION

Obesity increases the risk of Type 2 Diabetes Mellitus (T2DM), cardiovascular disease, cancer and mortality. However, the inter-individual variation in both obesity and the development of metabolic diseases is large. In line with this variation, researchers have identified a range of metabolically unhealthy and healthy obese phenotypes (1, 2). While exact definitions differ (3-6), it is generally agreed that metabolically unhealthy obese (MUO) people are obese with one or more of the following conditions: insulin resistance (IR), lipid disorders, hypertension, and an unfavourable inflammation profile (2, 7, 8). The underlying biology behind these different phenotypes remains unclear.

'Omics technologies and bioinformatics tools provide the means by which to investigate the complex biology behind multifactorial diseases like obesity. These technologies allow the production of complex multivariate datasets containing, for example, messenger RNA (mRNA), protein and metabolite information. Bioinformatics tools are then used to extract the maximum amount of information from these complex 'omics datasets.

Transcriptomics studies allow, for example, the detection of gene expression differences between groups, tissues, and time points, as well as different disease stages and treatments. By fitting gene expression data in regression models, it is possible to determine if any associations exist between the genes in the genome and phenotypes of interest. These associations may extend to a substantial number of genes. For easier interpretation of the results, these genes need to be analysed for biological relevance and meaning. Hence, gene expression studies are often strengthened using pathway analysis. By determining if the genes identified in the analyses are associated with a particular biological process, it would be possible to conclude that these biological processes are associated with the phenotype of interest. Although many association studies using gene expression data are carried out, replicating the results in similar studies remains a challenge, with few hits replicating.

Metabolomics is the study of small molecules that provides an end-point view into metabolism as a process. Metabolomics is now actively used to obtain a detailed mechanistic view of the pathology of metabolic diseases like obesity in order to identify metabolites as biomarkers for metabolic health. Although complex data can refer to any large dataset of multiple variables, in this thesis the term 'complex data' will be used to refer to gene expression and metabolite data that were used in this work.

This thesis is comprised of three studies that progressively build towards a deeper understanding of obesity. While previous adipose tissue gene expression studies compare obese and lean groups, or groups with different clinical health parameters, little is known about whether a hypothesis-free transcriptomics analysis can identify distinct groups of individuals by the similarity of their

subcutaneous gene expression profiles in obesity and whether these profiles associate with metabolic health. Further, the extent of the role of adipocytes in the mitochondrial transcriptomics pathways in the subcutaneous adipose tissue (SAT) in obesity has not been studied before. Additionally, because most studies focus on older individuals, there is a lack of knowledge about early predictors of metabolic health in young, healthy individuals who have not yet developed any strong symptoms that can be clinically determined. Lastly, most studies also face the problem of genetic confounding, making it difficult to study the effects of acquired obesity.

The studies in this thesis explore the associations between gene expression patterns and metabolite levels, and cardiometabolic risk factors by using monozygotic (MZ) twin pairs as: a) co-twins in discordance analyses to uncover the gene expression patterns and metabolite profiles associated with acquired obesity and b) individuals in metabolite–phenotype association studies. Clustering algorithms are employed to find patterns in the transcriptomics and metabolomics data that point to distinct subgroups of obesity. The following literature review starts by introducing the biological concepts relevant to the findings of the studies in this thesis. The already known disturbances in the body caused by obesity are briefly presented. Then, a brief introduction to transcriptomics and metabolomics technologies, as well as the various bioinformatics methods available to analyse the large datasets used in this thesis is given. The methods section covers the samples, technology and the methods used to analyze the data. The results section details the findings of the three studies. Discussion addresses the methodological considerations that were made in order to determine the selection of the bioinformatics methods and tools. It also discusses the outcome as a result of choosing these methods and tools. The remainder of the discussion discusses the meaning of the findings of all three studies. The results of this thesis in light of prior studies are also discussed. In the end, the study strengths and limitations are acknowledged and future prospects proposed.

2. REVIEW OF THE LITERATURE

This thesis focuses on the molecular (genes and metabolites) aspects of obesity and the use of existing bioinformatics methods in studying these molecular elements. The literature review starts with a review of the phenotype studied (i.e., obesity) and continues on to introduce the methods used to analyse the data.

2.1. Obesity

Obesity is a condition of excess body fat and widely accepted as exceeding 30kg/m^2 in body mass index (BMI). Obesity has been associated with several chronic conditions (9) such as cardiovascular disease, hypertension, dyslipidaemia, hyperglycaemia (10), T2DM, IR (11, 12), and cancer (13-15) as well as an increased risk of premature death (16). Heritability estimates for obesity are high at more than 0.70 (17) with both total and regional body fat being highly influenced by genetics (18). Single Nucleotide Polymorphisms (SNPs) have been able to explain only about 2% of the variation observed in BMI (19). Besides the genetic component to obesity, this complex disorder is also influenced by the complex interplay between lifestyle and the environment (20, 21), as well as epigenetics, at an interface between genes and the environment (22).

Obesity is mostly associated with an expansion of the adipose tissue which can expand up to more than 80% of one's body weight in obese people (23). The adipose tissue expansion accommodates the storage of excess nutrients as triacylglycerol in adipocytes. This expansion, in obesity, often brings with it metabolic disturbances caused by disruption to glucose, amino acid (AA) and lipid metabolism (24-26).

Figure 1 shows a general overview of the functions of the adipose tissue and changes that occur in obesity. These functions and the changes in obesity are covered in further detail in the following sections.

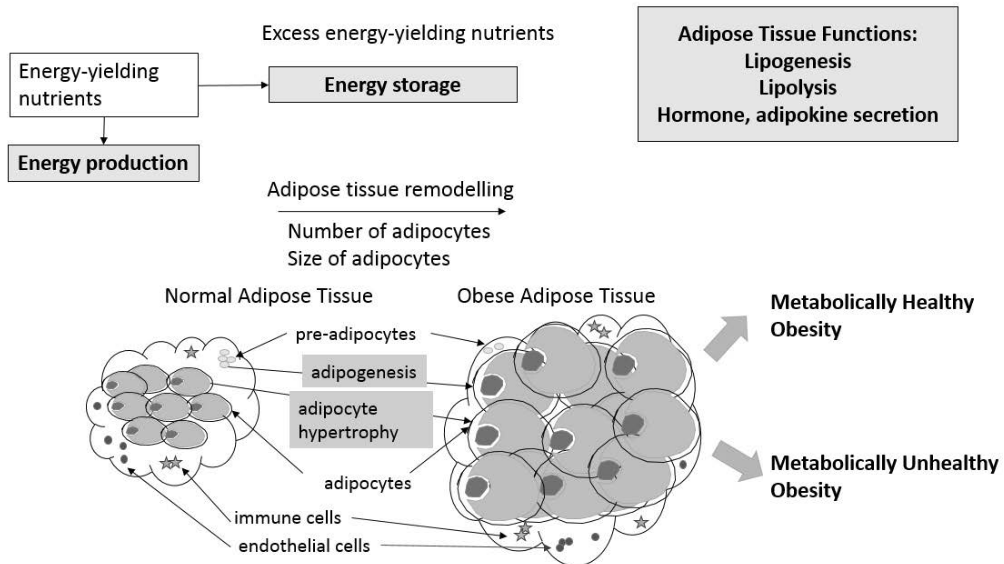

Figure 1: *General overview of adipose tissue and obesity. Excess energy-yielding nutrients that are not used for energy production are stored in the various fat depots of the body. Besides being responsible for lipolysis and lipogenesis, the adipose tissue is also an endocrine organ that secretes hormones and adipokines. In obesity, the adipose tissue undergoes remodelling with increase in number and size of adipocytes.*

2.1.1. Metabolically healthy and unhealthy obesity

Approximately 10–30% of obese people remain free from the metabolic complications associated with obesity, a condition called metabolically healthy obesity (MHO) (Figure 2) (1, 6, 8, 27, 28). Symptom-wise, MHO individuals exhibit high levels of insulin sensitivity, low values of low-density lipoprotein cholesterol (LDL-C), high values of high-density lipoprotein cholesterol (HDL-C) and low values of C-reactive protein (CRP), and they are normotensive (4). Based on 7-years of follow-up, people with MHO are not at increased risk for cardiovascular disease and all-cause mortality compared with healthy non-obese individuals (29-31).

In MUO people (Figure 2), a decreased capacity of adipose tissue to transport glucose and convert carbohydrate precursors into triglycerides is associated with adverse effects on metabolic health (32). The enlarged adipose tissue shows dysfunction in the mitochondria (21, 33), dysregulated secretion of adipokines and increased release of free fatty acids (FFA) (34). The FFAs and pro-inflammatory adipokines are transported to metabolic tissues, including skeletal muscle and the liver, and modify inflammatory responses as well as glucose and lipid metabolism, thereby contributing to metabolic syndrome (34).

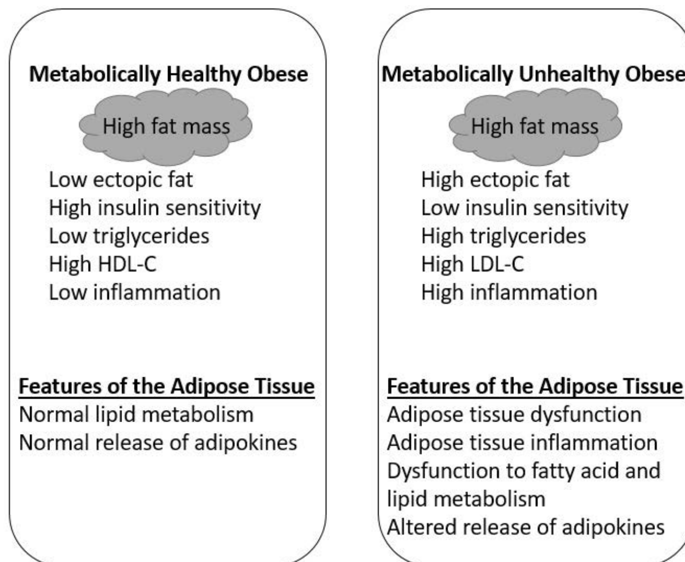

Figure 2: *Profiles of Metabolically Healthy (MHO) and Unhealthy Obese (MUO) people. MHO people show high insulin sensitivity and a better lipid profile compared to MUO people. MHO people also store less ectopic fat. In the adipose tissue of MUO people, there is dysfunction whereby fatty acid and lipid metabolism is affected. There is also more inflammation and an abnormal release of adipokines. HDL-C, high-density lipoprotein; LDL-C, low-density lipoprotein*

Several underlying reasons may exist for the difference in metabolic health in obese people. One reason could be that metabolic health in obesity is influenced by body fat distribution. There is a wide range of body fat distribution in both lean and obese adults with some individuals prone to storing subcutaneous fat, some prone to storing visceral fat and still others storing fat ectopically for example in the liver. While SAT is more protective in nature, an excess of visceral adipose tissue (VAT) associates with an increased risk for metabolic complications (35-37). This increased risk has been linked to VAT's production and release of substances that may cause metabolic abnormalities (38, 39). Individuals with high levels of intrahepatic triglyceride content (>5.5% of liver volume) also exhibit adverse metabolic health compared to individuals with normal intrahepatic triglyceride content (33, 40).

On a gene expression level, individuals with MHO, in comparison to individuals with MUO, exhibit a higher expression of genes involved in glucose uptake, lipogenesis (40, 41), and lipolysis in both VAT and SAT (42). On a metabolite level, circulating amino acids (AA), fatty acids (FA), very-low-

density lipoprotein (VLDL) and LDL-C particles, and inflammatory markers CRP and interleukin-6 in MHO individuals is lower than in MUO individuals (43-45).

2.1.2. Insulin resistance

Insulin is a hormone that regulates the metabolism of carbohydrates, fats and proteins in the body. Insulin stimulates glucose transport, triglyceride synthesis (lipogenesis) and inhibits lipolysis in mature adipocytes (46).

When carbohydrates from the diet are digested, glucose is released into the bloodstream, triggering the production of insulin. Insulin promotes the absorption of glucose from the blood into adipocytes, the liver and skeletal muscle cells (47, 48). IR is a condition in which there is a diminished ability of cells or tissues to respond to normal insulin levels, thereby resulting in a diminished ability of skeletal muscles to absorb glucose from the blood and elevated glucose production in the liver (49).

Insulin increases FA uptake from circulating lipoproteins by stimulating lipoprotein lipase activity in adipose tissue (46). Hence, increased plasma FFAs concentrations are typically linked to IR and T2DM (50-52). When there are available carbohydrates to be oxidised, carbohydrates take precedence and the body, via insulin signalling triggered by circulating glucose, supresses the oxidisation of FAs. Thus, insulin, which enhances glucose uptake in muscle and adipose tissue, inhibits release of FAs in adipose tissue, and increases esterification of FAs in adipose tissue and muscle (53).

Adipocytes are highly responsive to insulin (46), with insulin promoting the differentiation of pre-adipocytes to adipocytes (46). In obesity, adipocyte dysfunction may impair this responsiveness. Increased circulating FAs in obesity also impair the glucose FA cycle. Additionally, in obesity, increased lipid accumulation in adipocytes, muscle and liver cells, as well as disruption to adipocyte function, increase in mitochondrial oxidative stress, inflammation and circulating branched chain amino acid (BCAA) levels have all been shown to associate with IR (29, 54-60).

2.1.3. Lipid metabolism

Lipids (triglycerides and cholesterol) are ingested from food or synthesised de novo in the tissues. Chylomicrons carry the lipids from the intestine through the bloodstream to the target organs (61). In peripheral tissues, FFAs are released from the chylomicrons to be used as energy, converted to triglycerides or stored in the target tissues (62, 63). In addition to chylomicrons, the main form in which lipids are carried in the plasma is lipoproteins, which are produced by the liver. FFAs can also be circulating in the bloodstream bound to albumin (64). The lipoproteins, carrying triglycerides and cholesterol, that are produced by the liver are high-density lipoproteins (HDL), low-density lipoproteins (LDL), intermediate-density lipoproteins (IDL) or VLDL (65). Triglycerides transported in lipoproteins are lipolysed to FFA, which are then taken up by the target tissues (62). Thereby, the three main sources of FFA in the peripheral organs are chylomicrons, lipoproteins or FFAs. In the

lipid-forming tissues, such as the adipose tissue, FFAs are re-esterified to form triglycerides (63). Triglycerides can also be synthesised de novo from other carbon sources (carbohydrates and amino acids) when excess energy is available (66).

Lipid metabolism is the synthesis (lipogenesis) and degradation (lipolysis/FA oxidation) of lipids in cells. The balance between lipid synthesis and lipid breakdown determines the amount of fat accumulation. During lipogenesis, glycerol and acyl-coenzyme A (acyl-CoA) produced in the mitochondria are converted to triglycerides/triacylglycerols for storage in the adipocytes. Conversely, during lipolysis, triacylglycerols are broken down into FFAs and glycerol for the purpose of energy production.

In obesity, lipid metabolism is impaired and there is increased FA release from an expanded fat mass, resulting in increased FA concentrations (67, 68). These increased circulating FFAs have been shown to inhibit insulin action in peripheral tissues (69, 70) and impair insulin-mediated whole-body and hepatic glucose uptake (71). A increased supply of calories also leads to increased triglyceride (TG) and VLDL production in the liver, thereby promoting hypertriglyceridaemia (34).

2.1.4. Amino acid metabolism

AA metabolism is the process by which proteins in the diet are first broken down into AAs, then absorbed into the bloodstream to form new proteins. Excess AAs are converted by the liver into keto acids and urea. The keto acids can be used as an energy source or converted into glucose or lipids for storage. Urea is excreted in urine and sweat. During AA metabolism, AAs are degraded into various compounds and then ultimately oxidised to release energy. Several metabolite studies have consistently shown the association of essential AAs, BCAAs and aromatic amino acids (AAA), with obesity (26, 57, 58, 72-75). This section of the literature review will highlight these two types of AAs.

BCAAs have important roles in protein synthesis (76), glucose metabolism and oxidation (77), and leptin secretion (78). They are poorly metabolised the first time they pass through the liver (79, 80) and thus trigger a signal to the body of the AA content (81). In obesity, increased levels of BCAAs have been observed; these levels correlate with both obesity and serum insulin levels (75, 82). Accordingly, several studies have identified a downregulation of SAT genes responsible for mitochondrial BCAA catabolism in obesity, suggesting that reduced oxidation of BCAAs in tissues result in the increase in plasma BCAA levels (21, 33). Newgard et al. (2009) found that BCAAs contribute to obesity-related IR and glucose intolerance (58), suggesting a further link between IR and the adipose tissue's capacity to catabolise BCAAs. Twin studies have further confirmed that the findings of the association of both BCAA downregulation in tissue and increased circulating BCAAs in plasma with obesity are not confounded by shared environment and genetic factors (21, 83).

BCAAs can be oxidised in skeletal muscle, whereas other essential AAs are catabolised mainly in the liver (79, 84, 85). During adipogenesis (see Section 2.4.1), when mitochondrial mass is elevated, leucine catabolism and the expression of enzymes involved in BCAA catabolism is increased (86, 87).

AAAs include phenylalanine, tryptophan, and histidine, as well as tyrosine which is synthesised from phenylalanine. Circulating AAAs have been found to associate with BMI (74, 88) and are suggested to be markers of IR development. During AA metabolism (89, 90), AAAs compete with BCAAs for transport into cells by large neutral AA transporters (88, 91).

2.2. Adipose tissue

The adipose tissue is made of adipocytes, preadipocytes and a variety of other cells including fibroblasts, endothelial cells and macrophages (92, 93). Adipose tissue stores excess energy-yielding nutrients as lipids in the form of triacylglycerol; in energy deficit conditions, it supplies energy in the form of FAs to other tissues through lipolysis (94). Adipose tissue is also an active endocrine and immune organ secreting adipokines, a diverse range of protein factors and signals (92, 95) that regulate several metabolic processes in the body. Hence, it follows that adipose tissue plays a role in many functions including energy storage and homeostasis, metabolism, insulin secretion, immunity and inflammation (96-98).

In lean (BMI 22–25) people, adipose tissue makes up approximately 20% of total body weight, while in obese (BMI > 30) people, adipose tissue makes up almost half of the body weight (92). There are two types of adipose tissue in mammals: white and brown adipose tissue. The main function of brown adipose tissue is storing lipid droplets for heat production, while white adipose tissue stores excess energy as triglycerides and releases them in the form of FFAs.

The adipose tissue is mostly made up of adipocytes. Up to 85% of the weight of adipocytes is made of lipids (92). While not the only cell types that contain esterified lipids, adipocytes are unique in the quantity of lipids they can store, the rapid release of stored energy in triglycerides, and their collection of protein (99). Adipocytes serve as an energy bank with an important role in storage and release of FAs (100, 101). Adipocytes also secrete factors that include hormones, cytokines, growth factors and play an important role in the regulation of energy balance and insulin sensitivity (102).

Adipocytes are formed from pre-adipocytes during a process called adipogenesis (see Chapter 2.4.1 below). This process is made up of two phases: a first phase in which the cells are committed to triglyceride storage, and a second phase in which the cells grow and become more round due to the increase in triglyceride amount (98).

The size of adipocytes varies depending on adipose tissue location; adipocyte size associates positively with visceral and subcutaneous abdominal fat areas and negatively with lower body fat percentage (103). Adipocyte size is also positively associated with, macrophage infiltration, and secretion of pro-inflammatory adipokines, contributing to metabolic disturbances (104, 105). Adipocytes are increasingly considered to be directly linked to the pathologies associated with obesity (92). While smaller adipocytes respond to insulin by increasing lipid uptake, larger adipocytes, as seen in obesity, are less sensitive and insulin resistant (106).

For the remainder of this thesis, all text detailing adipose tissue and adipocytes are in reference to only white adipose tissue. Brown adipose tissue is not the subject of this thesis and hence will not be discussed any further.

2.3. Main fat depots in the body

Adipose tissue is distributed in multiple depots in the body, both subcutaneously and internally, and clusters of adipocytes can be found near, or embedded in, other organs such as the lymph nodes and skeletal muscle (92). The main fat depots of the human body are the SAT and VAT. SAT is found just below the skin while VAT surrounds the inner organs and can be divided into omental, mesenteric, retroperitoneal (surrounding the kidney), gonadal (attached to the uterus and ovaries in females and epididymis and testes in men) and pericardial adipose tissue (Figure 3). While SAT and VAT volume are highly correlated with total body fat (107), they have differences in structure, cellular size, and biological function (108).

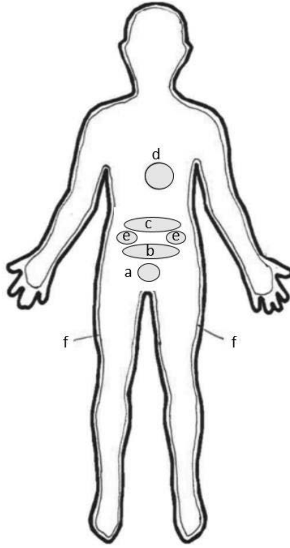

Figure 3: *Main fat depots of the body. gonadal (a), mesenteric (b), omental (c), pericardial (d), retroperitoneal (e), and SAT (f).*

When SAT cannot adequately store triglycerides, ectopic fat accumulation occurs (109, 110). Hence, when adiposity increases, fat accumulates in the SAT and VAT, but may also be deposited ectopically in individual sites such as the liver. Abnormal accumulation of fat in the liver can cause non-alcoholic fatty liver disease (NAFLD) (111-113). The fatty liver, as seen in NAFLD, overproduces glucose, VLDL particles, coagulation factors and cytokines, all of which are important in pathological cardiometabolic processes (114). Liver fat (LF) has been associated with metabolic syndrome (115, 116) and when increased in obese people, has been shown to cause continuous release of FFAs into the plasma resulting in metabolic disturbances brought on by IR (32). NAFLD can progress to non-alcoholic steatohepatitis (NASH), whereby, along with fat in the liver, there is inflammation and liver cell damage. Chronic inflammation may result in liver fibrosis, where the injured liver tissue is abnormally and continuously replaced by fibrotic tissue. Cirrhosis, the most advanced stage of liver fibrosis, is caused by the continuous replacement of liver cells by fibrotic tissue, resulting in liver damage or complications (117, 118).

Because fat depots are harder to measure than BMI and total body fat, often the latter two measurements are used in adiposity-related studies. Total body fat is a major contributor to metabolic health, with specific fat depots having different contributions to metabolic health (119).

Many studies have compared SAT, VAT and LF in their associations to metabolic risk factors, especially with regards to IR, with contradictory results as to whether they differ in function and metabolic activity (38, 40, 107, 120-122). Function-wise, VAT has higher rates of lipolysis than SAT (123, 124), thus, when comparing the same amount of VAT and SAT, VAT contributes to an increased release of circulating FFAs (37, 125). As a result, VAT has been suggested to be responsible for the whole-body IR related to higher FFA flux (37). Additionally, because lipolysis of VAT triglycerides drains these FFAs into the portal vein, delivering it to the liver, VAT has been said to be more harmful than SAT (39, 126-128), which drains FFAs into the systemic vein. The release of FFA from VAT directly into the portal vein and liver, affects glucose (129) and lipid (130) metabolism. Compared to SAT and VAT, LF is the most correlated with serum insulin and triglycerides, and the association is independent of BMI and the amount of SAT and VAT (131).

While VAT has been suggested to be the most important body fat component for metabolic risk factors (38, 121, 123, 132, 133), some studies have pointed to SAT (120) and LF (40) as playing a bigger role compared to VAT. Subjects with high LF have impaired insulin action in the liver, adipose tissue, and skeletal muscle and increased hepatic Very Low Density Lipoprotein Triglyceride (VLDL-TG) secretion rates, independent of VAT (40). One study on IR found that deep SAT amount was more correlated with IR than superficial SAT or VAT (107), while another found that SAT associated just as strongly with IR as VAT with IR (122).

On a gene-expression level, several genes, including those related to inflammation, have been found to show depot-related variations (134-137), owing to VAT containing more pro-inflammatory immune cells than SAT (138). Additionally, VAT adipocytes have a reduced capacity for lipogenesis (139) and a greater capacity for lipolysis than SAT cells (124). Marked hypertrophy of SAT compared with VAT adipocytes was observed in obese subjects (140, 141). In morbid obesity, lipogenesis and FA oxidation have been shown to be downregulated in SAT, but unchanged in VAT (142). Auguet et al. (2014) suggest that in extreme obesity, the presence of SAT but not VAT prevents further development of fat mass, thereby decreasing the expression of genes responsible for lipolysis and FA oxidation (142). On a metabolite level, SAT and VAT also differ by their metabolite content, with VAT displaying higher amounts of AAs, nucleosides, and carbohydrate metabolites than SAT (143). SAT, compared to VAT, has higher FFAs (143). These findings are in line with VAT as an active endocrine organ and SAT being more active in lipid storage and release (143).

2.4. Obesity-related adipose tissue dysfunction

2.4.1. Adipose tissue expansion in obesity

The number of preadipocytes in humans is set during childhood and adolescence and does not increase in adulthood (144). From the pool of pre-adipocytes, obese individuals generate significantly more adipocytes per year than lean individuals (144).

Two distinct mechanisms can lead to increased adipose tissue size: hyperplasia or hypertrophy. Hyperplasia refers to an increase in adipocyte cell number while hypertrophy refers to an increase in adipocyte volume (145, 146). Adipocyte hyperplasia requires the recruitment of pre-adipocytes present in the vascular stroma of adipose tissue (103), and their proliferation and differentiation through a process called adipogenesis. During adipogenesis, extra-cellular matrix (ECM) remodelling is also carried out. Along with changes in mitochondrial number and morphology, there is also higher oxidative capacity (147, 148), reactive oxygen species (ROS) levels and cell signalling (149) during adipogenesis. Adipogenesis has been suggested to be protective against lipid as well as glucose and insulin abnormalities in obesity (150). Without the recruitment of pre-adipocytes and subsequent adipocyte differentiation, an excess of adipocyte hypertrophy may occur, resulting in insulin-resistant adipocytes (151, 152). Excess adipose tissue and adipocyte hypertrophy have both been linked to metabolic disturbances, T2DM, hypertension, dyslipidaemia, cardiovascular disease, and a variety of cancers (9, 153-155).

2.4.2. Adipocyte mitochondria in obesity

Excess intake of nutrients causes an overload of FFAs, elevated ROS production and a reduction in mitochondrial biogenesis, all contributing to mitochondrial dysfunction (156). This dysfunction leads to reduced β-oxidation and ATP production and increased ROS production, as well as pro-inflammatory cytokine production (157) resulting in IR (156). Reduced mitochondrial function in obesity is thought to, in turn, impair the mitochondria's capacity to consume FAs through oxidative phosphorylation, leading to an accumulation of triglycerides (158). Reduction of the mitochondrial oxidative metabolism in SAT correlates with whole body IR and inflammation (159). Because the reduction in oxidative capacity in adipose tissue is similar in obese diabetic versus obese non-diabetic patients, it has been suggested that obesity per se impairs mitochondrial function (160).

2.4.3. Adipose tissue inflammation in obesity

Both overweight and obese persons are more likely to have elevated CRP levels (0.22 mg/dL or more) than normal-weight people (161), indicating chronic inflammation. This chronic inflammation is marked with production of an abnormal amount of adipokines and activation of pro-inflammatory signalling (56, 162-164). Inflammation in obesity has been proposed to be due to adipocyte

hypertrophy and dysfunction, oxidative stress, toxic lipolysis, and deficient intracellular matrix remodelling (165).

As the adipose tissue expands, adipocytes become hypoxic, and an inflammatory response is elicited to increase blood flow and to stimulate angiogenesis (166). During early adipose tissue expansion, a pro-inflammatory response is activated and the immune response is dominated by anti-inflammatory signals (155). Conversely, during chronic obesity, a pro-inflammatory response is triggered by adipocyte death, hypoxia, and reduced FA storage capacity in dysfunctional adipocytes (155).

Adipose tissue inflammation is now recognised as an important early event in the development of obesity complications, especially T2DM (166-169). Studies have also suggested that the inflammatory state may in fact be causal in the development of IR and the other disorders associated with obesity, such as hyperlipidaemia and metabolic syndrome (170, 171).

2.5. Twin study setting in obesity research

MZ twins share 100% of their genetic polymorphisms as well as some environmental conditions, like prenatal and early childhood family environment. However, even MZ twins growing up together have experiences that are unique to each twin (i.e. non-shared environment). These differences accumulate as the twins grow older and start to live increasingly divergent lives. Twin studies are free from genetic and shared environment confounding and highlight environmental effects not common to both twins in a pair (e.g. aspects of diet, exercise and lifestyle) as a basis to explain individual differences within MZ twin pairs (172). Comparing the co-twins of a pair to each other, especially those discordant for a disease, provides an ideal matched case versus control study. This study setting is especially useful when studying environmental effects on a disease or condition. Over the years, obesity studies using discordant MZ twin pairs have provided significant insights into acquired obesity (i.e., obesity that is due to environmental and not genetic effects) (33, 73, 159, 173-176).

2.6. Transcriptomics

The transcriptome is the entire collection of gene transcripts in a species expressed in a specific cell or tissue. Because the genes expressed in different cells and tissues vary, gene expression studies are carried out on targeted cells or tissues to ascertain the genes expressed in these specific cells or tissues. In this thesis, adipose tissue, adipocytes and plasma blood were used because as obesity develops, there are important changes occurring in the adipose tissue and blood. Over the years, transcriptomics studies have proven useful in obtaining a biological perspective into gene regulation and gene networks (177), comparing tissues and cells (178, 179), classifying sub-types of diseases (180, 181) and comparing different developmental stages (182, 183) and different species (184, 185). Historically, the study of gene expression has had to rely on technologies like Northern Blots, reverse

transcription polymerase chain reaction (PCR), expressed sequence tags (EST) and serial analysis of gene expression (SAGE). Various hybridisation- or sequence-based technologies have now been developed to identify and quantify the transcriptome.

Sequence-based technology, ribonucleic acid-sequencing (RNA-seq), allows complementary deoxyribonucleic acid (cDNA) sequencing on a larger scale (186). This technology allows cDNA fragments derived from mRNA to be sequenced in a high throughput manner, resulting in information about the transcript structures and levels (amounts) of transcripts (187). Unlike microarray technologies, RNA-seq does not rely on prior knowledge of the genome sequence, does not incur high background signals from cross-hybridisation (188, 189) and is sensitive enough to detect very low and very high amounts of transcripts (187).

Hybridisation-based approaches typically involve incubating fluorescently-labelled cDNA with microarrays; the expression of these gene transcripts are then measured based on light intensity (190). Hybridisation-based approaches are high throughput and relatively inexpensive compared to sequence-based technology (187). Amongst the more widely-used microarray platforms are the Affymetrix™ GeneChips, spotted microarrays and Agilent™ microarrays. Spotted microarrays use spotted cDNA PCR product probes and measure gene expression as ratios between signal intensities from mRNA samples and cDNA (191). These microarrays can be quite noisy due to the various processes involved in the experiment (192). Agilent™ microarrays use 60-mer long probes that are synthesised *in situ* on microarray slides (Agilent, Santa Clara, CA, USA). Each of these probes is of sufficient length to detect a single gene (193). Both spotted microarrays and Agilent™ microarrays use a 2-color scheme, which allows for the interrogation of two samples in the same microarray (193).

In a comparison study conducted by Irizarry *et al.* (2005), three types of microarrays were compared in terms of precision and accuracy. Amongst the Affymetrix™ oligo, 2-color oligo and 2-color cDNA microarrays, the Affymetrix™ microarray performed the best (194). The remainder of this section concentrates on Affymetrix™ microarrays, which is the technology employed in the study of gene expression in this thesis.

Affymetrix™ microarrays utilise short oligonucleotide probes that are of 25-mer length to represent areas of interest in the genome (190). An mRNA of interest is usually represented by a probe set made of 11-20 probe pairs of these oligonucleotides. Each probe pair is made up of a perfect match and a mismatch probe (190). RNA is extracted from samples, labelled with fluorescent dyes, hybridised to the arrays, washed and then scanned with a laser (195). Transcripts from the RNA that correspond to the oligonucleotide probes will hybridise to these probes. Gene expression is then measured as the intensity of light from the fluorescent dyes attached to these sample transcripts (195). The process flow is detailed in Figure 4.

In order to derive the expression values for each probe set (corresponding to a gene), the probe intensities are summarised (196). A popular method of summarising probe information is the log-scale robust multi-array analysis (RMA) method (197). First, background probe data is removed and probe data is normalised across arrays (197). The expression measure is then derived using a log-scale linear additive model (197). Another popular method for summarising gene expression measures is the GeneChip RMA (GC-RMA), which combines a stochastic model algorithm like the one used in RMA with physical models that predict mRNA concentrations using the sequence information of the probes (198). This model uses the same normalisation and summarisation methods as RMA and is more suitable than RMA when gene expression levels are low (198).

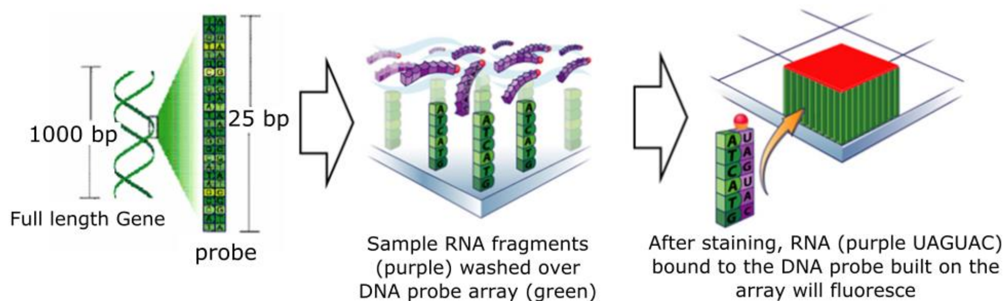

Figure 4: *Affymetrix™ microarray analysis process flow. Affymetrix™ oligonucleotide probes are 25bp long and designed to interrogate specific parts of the genome. Labelled RNA is placed on the microarray and allowed to hybridise to the probes. Complementary nucleotide sequences between the labelled RNA and probes will allow binding of the sample RNA to the probes. Scanning of the microarray provides readings of the light intensity indicative of the amount of gene expression. Adapted from http://tools.thermofisher.com/content/sfs/brochures/activity2_structure_function.pdf*

Affymetrix GeneChips™ eliminate the need to manage cDNA libraries while providing probe redundancy whereby multiple probes cover different regions of the same transcript (190). These gene chips also have an extensive range with over 50 types of arrays in the GeneChip catalogue (199). The downside to the Affymetrix™ microarrays is that shorter oligonucleotides (25-mer in this case) do not hybridise as well as longer ones (193).

The Affymetrix™ Human Genome 133 Plus array was used in the studies in this thesis (Affymetrix™, Santa Clara, CA, USA). It has probesets representative of sequences taken from Genetic Sequence Data Bank (GenBank®), database for "expressed sequence tags" (dbEST) and NCBI Reference Sequence Database (RefSeq).

2.7. Metabolomics

Metabolomics is the systematic study of small metabolite molecules (<1500 Da) in biological fluids (200-203) at a given point in time. Metabolomics provides an integrated profile of metabolism, reflecting the net results of genetic and environmental interactions (204, 205).

The human metabolome contains thousands of metabolites (206) including AAs, lipids, organic acids, nucleotides, representing a huge challenge to researchers seeking to understand not only the role of each of the metabolites but also the interplay between these metabolites in the context of biological systems as well as disease. Metabolite distributions are subjected to high temporal and spatial variability, and they are influenced by, for example, circadian fluctuations and diet (207, 208). This further adds to the complexity of metabolomics studies, requiring researchers to carefully select the experimental designs best suited for the study question.

Two key technologies to identify metabolites exist. One is proton nuclear magnetic resonance (^{1}H NMR) spectroscopy, another is mass spectrometry (MS). ^{1}H NMR, Gas Chromatography–Mass Spectrometry (GC–MS) and Liquid Chromatography–Mass Spectrometry (LC–MS) are well-established powerful analytical methods for generating metabolomics profiles (209).

In NMR technology, protons and hydrogen ions in molecules are detected based on their magnetic properties with different molecules, differentiated using their resulting spectral shapes. The signal intensities observed in the magnetic field NMR spectrum are directly proportional to the concentration (i.e., molar amount) of that molecule in the sample (210). The advantages of NMR are the minimal requirements for sample preparation and preservation of the samples used (207). However, this technology identifies only medium to high abundance metabolites.

MS has three components: an ion source, a mass analyser that measures the mass-to-charge ratio (m/z) of the ionised analytes, and a detector that quantifies the number of ions at each m/z value (211). The MS often requires a liquid or gas chromatography step in order to separate the molecules in a sample (212). MS-based techniques usually require a sample preparation step, for example, direct injection, liquid–liquid extraction (LLE), solid-phase extraction (SPE) (207). The sample extract separates when injected onto a gas chromatograph or liquid chromatograph column (212). Here, the ionised metabolites are accelerated and deflected by the magnetic field, with the amount of deflection depending on their mass and charge. MS is more sensitive than NMR, but requires a sample preparation step, which can destroy the metabolites (207).

Triple quadrupole MS (Figure 5), the technique used in this thesis, consists of two quadrupole mass analysers in series, with a (non-mass-resolving) radio frequency (RF)-only quadrupole between them as a chamber for collision-induced dissociation. Quadrupole mass analysers use oscillating electrical

fields to selectively stabilise or destabilise the paths of ions passing through an RF quadrupole field created between 4 parallel rods.

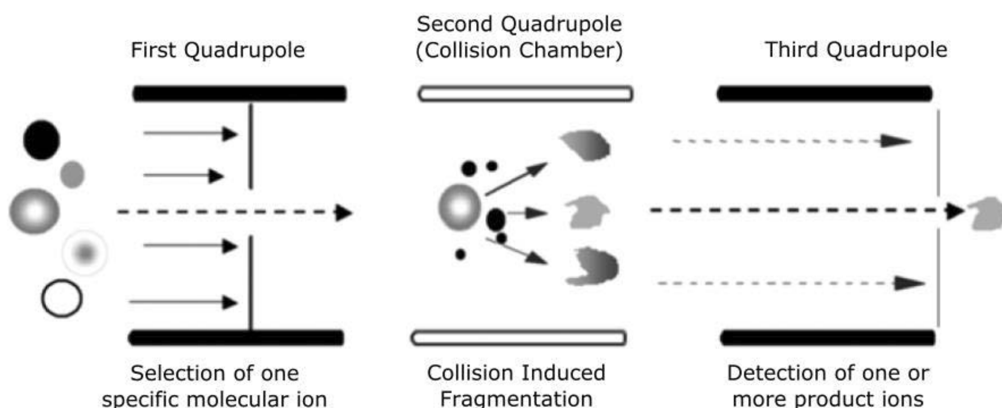

Figure 5: *Mass spectrometry experiment flow. The first mass analyser allows selection of the targeted metabolite, while the second one allows fragmentation of the metabolite. The third one measures the mass per charge of the ions produced after the collusion. Adapted from (213)*

2.8. Bioinformatics methods in the study of complex data

Bioinformatics methods employ automated, computational data processing to derive meaning from biological data. High-dimensional 'omics data like microarray and metabolite data pose a challenge because the number of covariates/features often exceed the sample size (214), complex dependencies exist between genes and between metabolites, and the data are non-normally distributed (215). These small datasets can introduce data over-fitting, which is when the model fits the data being studied well or by chance but does not perform accurately when used on another set of independent data (216, 217). Common methods to prevent over-fitting include testing the model on a different, independent dataset or performing cross-validation using several different partitions of the same dataset to train and test the dataset (218) or utilising regularisation methods like least absolute shrinkage and selection operator (LASSO) that imposes a penalty on the regression coefficients so that some coefficients can be shrunk to zero and subsequently dropped from the regression model (219).

Microarray and metabolite data also suffer from high technical variation in the experiments and high levels of noise (220). Additionally, metabolite data is usually right-skewed, those involved in central metabolism are usually more constant while those involved in secondary metabolism are more susceptible to change depending on the environmental conditions and are prone to fluctuations (220). The rest of this chapter covers the data pre-processing and analysis techniques used in this thesis.

2.8.1. Data normalisation and transformation

Technical variation in transcriptomics and metabolomics data can be introduced in a variety of ways. There can be within-instrument variation (e.g., temperature changes), variations in samples processed in different batches (e.g., duration of hybridisation on microarray) and differences introduced by human handling (e.g., sample extraction and preparation) (221, 222). Different samples can also have differences in the number of cells and concentrations of biofluid (223). These technical variations in the samples obscure the interesting biological variations in the data and need to be removed so that, as much as possible, only the biological variations remain.

Technical effects can be reduced via data normalisation, leaving all of the samples on the same measurement scale (224, 225) and allowing meaningful comparisons to be made between samples or conditions in an experiment. This step is usually undertaken after initial quality control (QC) to check the quality of the RNA, signal quality of the array, dataset homogeneity as well as comparability across the arrays (samples) (226, 227).

In this thesis, I will concentrate on quantile normalisation and rank normalisation, which were employed on the gene expression and metabolite data.

Quantile Normalisation

In quantile normalisation, each sample is given the same distribution over features (e.g., gene expression levels) (228). Values for each feature within each sample are sorted and a mean quantile over all of the samples in the microarray experiment is calculated. Subsequently, the value of the data item in the original dataset is substituted with the mean (followed by a re-sort of each sample) (228).

Quantile normalisation may suffer from false negative findings, especially at low expression levels, since this normalisation method assumes an equal distribution of expression values, which in turn may mask biological changes (196, 198).

Rank normalisation

Rank normalisation is a nonparametric normalisation technique that replaces each observation by its fractional rank (the rank divided by the total number of features) in the sample (229, 230). This procedure removes noise because it only uses the ordering of the observations (231) and is not overly affected by outliers (232). However, using only ranking can result in a loss of information, which is particularly harmful in small sample sets (229). Additionally, the rank of features may be the same in two samples but the actual gene expression or metabolite quantity may be different (229). Conversely, the gene expression or metabolite quantities may be the same across two samples but may be ranked differently in the two samples (229).

In this thesis, metabolite data was rank transformed to a standard normal distribution with a mean of zero and variance of one. Hence, data was scaled using the standard deviation. While several scaling methods have been used on metabolite data, range scaling (difference between the maximum and minimum concentrations as the scaling factor) and autoscaling (standard deviation as the scaling factor) are able to transform the data so that the ranking of the important metabolites do not depend heavily on the average concentration and the magnitude of the fold-changes between samples (220, 223).

2.8.2. Principal Component Analysis

Principal Component Analysis (PCA) is employed in 'omics data to identify a few combinations of features that best explain the total variation in the original dataset. Principal components (PC) are a set of vectors in a multidimensional vector space that decreasingly capture the variation seen in data points (233). PCA is a dimension reduction technique that finds the directions (PCs) in a multidimensional space along which the variation of the data is the maximum. PCA generates PCs with the first PC capturing more variation than the second, and so on (234). These PCs act as new variables that are linear combinations of the original variables.

PCA is often used as a first step before clustering or classification of samples because PCs are uncorrelated and may represent different aspects of samples (234). PCA is a powerful tool to reduce the dimension if the data of the subsequent biological question is related to the highest variance in the dataset (220). One consideration, though, is to decide how many and which components to use in subsequent analyses. Options include using components that correlate with a phenotype of interest (235) or using enough components to include most of the variation in the data (236). Often in microarray datasets, most of the variability can be accounted for by a small number of principal directions (237). However, the biological significance of these PCs is not directly apparent (238).

2.8.3. Data clustering

Clustering is a form of unsupervised learning used to assign similar objects into groups, thereby enabling the reduction of complex data and allowing detection of underlying patterns in the data.

The basic premise is to cluster either the samples or measured features (genes or metabolites in this thesis) based on their similarity. In the first instance, samples with similar gene expression or metabolite profiles are identified with the collection of gene expression or metabolite concentrations acting as features identifying each sample (239). In the second instance, genes or metabolites act as objects to be clustered. The purpose here is to identify groups of genes or metabolites acting correlatively on the different samples (239). Application to biological data has allowed discovery of

groups of co-expressed genes or metabolites, as well as identification of samples with similar genome-wide gene or metabolite profiles (240-242).

Several clustering techniques exist, amongst them are the distance-based techniques (using, for example, Manhattan or Euclidian distance). In this thesis, two distance-based techniques are discussed: K-means clustering (using a partitioning algorithm) and agglomerative clustering (using hierarchical algorithm). Agglomerative clustering starts with each object in its own cluster and merges iteratively similar clusters (according to the selected distance measures) until one single cluster remains. The results are visualised via dendrograms. In K-means, an initial number of cluster (K) objects are assigned and objects are randomly assigned to one of the K clusters. Iteratively, objects are moved between clusters and allowed to remain in the new cluster only if they are closer to it than to their previous cluster.

Both methods have drawbacks and no consensus has been reached as to which method is better: hierarchical clustering tends to give equal consideration to all features including less important genes and the number of clusters derived depends on at which level the dendrogram is cut (214). The K-means clustering requires initialisation and specification of the number of clusters and is sensitive to noisy data and outliers (243, 244).

While clustering genes on the basis of samples is straightforward, the clustering of samples on the basis of genes or metabolites presents a problem because the number of features exceeds the number of samples. Some ways to overcome this problem include reducing the number of genes or metabolites used in the analysis through dimension reduction techniques (e.g., factor analysis or PCA) (245, 246) and then performing clustering with the reduced dimension. Also, when calculating the similarity of samples over a large number of genes or metabolites, the similarity measures are averages which ignore subsets of genes that may be more similar to each other across different clusters (244). Genes or metabolites can be partitioned into homogenous groups before clustering is carried out separately in these groups (244).

Another option is to provide prior knowledge to the cluster. For example, if there are distinct groups of a disease in a dataset, it is possible to provide this information to the K-means clustering algorithm so that the number of clusters equals the number of subgroups in the disease. PCA can also be used to heuristically determine the number of clusters beforehand.

Clusters can be validated by bootstrapping or permutation methods (235). Other less computationally-intensive methods include checking clusters for density (variance in the cluster should be minimal) and separation from other clusters or using validity indices, for example, the partition coefficient (247).

2.8.4. Association analysis using linear mixed modelling

Association studies can be used to find associations between genes or metabolites and traits of interest in transcriptomics and metabolomics studies. In this thesis, association analyses were conducted using linear mixed modelling to identify metabolite–phenotype associations.

The linear mixed model is an extension of linear regression which allows the modelling of both fixed and random effects. The random effects are used to model variables for which there are variations between different levels of the variable. For example, measurements taken multiple times from the same subject or from siblings are more likely to be more similar than measurements taken from another subject and, hence, need to be accounted for in the model.

2.8.5. Differential analysis using moderated t-tests

Differential analysis for complex data tests differences in the data (e.g., gene expression, metabolite levels) between groups, between individuals, between different treatments and tissues as well as between different time points. Performing t-tests on microarray data is challenging because there are limited measurements for each gene. Additionally, while normalisation has been used to reduce technical variations, having a small sample size still impacts the ability to detect differentially expressed genes (231). The high number of genes also poses a multiple testing problem that can give rise to false positives (248). One upside is that genes expressed at similar levels have similar variances, and in a Bayesian modelling approach these similarities can be used as prior knowledge to estimate variances by borrowing information from genes with similar expression levels (249). Using empirical Bayes, the limma package in R-Bioconductor fits a linear model to each row of data (in this thesis, gene expression and metabolites) and shares variance information between the gene-wise models, increasing the degrees of freedom even when the number of samples is small (224, 250).

Two common methods employed in dealing with the multiple testing problem are the Bonferroni Correction and the False Discovery Rate (FDR) adjustment. The Bonferroni Correction works on the premise that if the type I error (false positive) rate for a null hypothesis is α, the study-wide error rate when n tests are carried out will be α/n. This adjustment implies that interpretations of the null hypothesis can differ according to how many tests were performed. (251). While there are several ways to implement the FDR adjustment, one commonly used way is the Benjamini Hochberg procedure which controls for the expected proportion of falsely rejected hypotheses by ordering all the p-values in the experiment in descending order and testing if each p-value is lower than $\alpha(k/n)$, where α is the error rate, n is the number of tests and k is 1,2…n. When the largest value for k is found, all the p-values calculated up to that point are rejected (252).

2.9. Transcriptomics patterns in obesity

Transcriptomics studies on human SAT have revealed upregulation of inflammation (21, 33, 253-255) and immune response (253, 254), and downregulation of mitochondrial pathways (21, 159, 254), insulin-signalling (256) and lipid metabolism (254, 257) in obesity. In a previous twin study, acquired obesity resulted in a significant reduction in transcripts responsible for mitochondrial function and an increase in inflammatory pathways in subcutaneous fat, a phenomenon that was closely correlated with disruption to whole-body insulin sensitivity (21). A list of transcriptomics studies on SAT and adipocytes (excluding weight-loss, dietary intervention studies) from the past 10 years on SAT and adipocytes can be found in Table 1. Studies investigating IR or T2DM in obesity are not included.

Most of these studies employ differential analysis to determine the differences in groups that have been defined according to clinical traits or phenotypes. Common findings across these studies show that obesity has consistently been linked to metabolic pathways, inflammation and BCAA-related pathways.

Table 1: *Previous transcriptomics studies in obesity (excluding weight-loss studies).*

Study design: **Participants** **Tissue** **Experiment method**	**Transcriptomics analysis methods**	**Results**	**References**
Within-pair comparison of heavy and lean co-twins • 17 female pairs, 9 male pairs • SAT • Affymetrix^{TM} U133 Plus 2.0 chips	• differential analysis with moderated t-tests (limma) • pathway analysis	• mitochondrial biogenesis, oxidative metabolic pathways, and OXPHOS proteins in SAT are downregulated in acquired obesity	Heinonen et al. (2015) (159)
Comparison between lean, MHO, MUO • lean healthy individuals: 5 females, 2 males, • metabolically healthy individuals: 6 females, 2 males, • metabolically unhealthy individuals: 6 females, 2 males • SAT	• differential analysis using ANOVA • pathway analysis	• genes related to branched-chain amino acid catabolism and tricarboxylic acid cycle were less downregulated in metabolically healthy obese individuals compared to metabolically unhealthy obese individuals	Badoud et al. (2014) (45)

Study design: Participants Tissue Experiment method	Transcriptomics analysis methods	Results	References
• Affymetrix™ Human Gene 2.1 ST array			
Within-pair comparison of heavy and lean • BMI-discordant twin pairs, concordant for liver fat, non-obese: 6 females, 2 males, obese: 6 females, 2 males • BMI-discordant twin pairs, discordant for liver fat, non-obese: 4 females, 4 males, obese: 4 females, 4 males • BMI-concordant pairs: lean: 4 females, 7 males, heavy: 4 females, 7 males • SAT • Affymetrix™ Human U133 plus 2 chips	• differential expression analysis using Mann–Whitney U test • pathway analysis	• maintenance of high mitochondrial transcription and lack of inflammation in SAT are associated with low liver fat and metabolically healthy obesity	Naukkarinen et al. (2013) (33)
Comparison of SAT and VAT in obese people • 8 obese females • SAT vs. VAT • Affymetrix™ Human U133 plus 2 chips	• differential analysis with moderated t-tests (limma) • pathway analysis	• 22 genes differentially expressed in SAT vs VAT	Hoggard et al. (2012) (258)
Comparison of SAT in 2 groups differing by the presence of crown-like structures indicating inflammation • 20 female, 16 male; all obese • SAT • Illumina™ HumanHT-12 v3	• differential analysis using ANOVA • pathway analysis	• genes involved in inflammation and response to inflammation were upregulated in group with crown-like structures (indicating macrophages)	Le et al. (2011) (259)

Study design: Participants Tissue Experiment method	Transcriptomics analysis methods	Results	References
expression bead chip			
Comparison of lean, overweight, obese, and obese with metabolic syndrome • 32 women (8 in each group) • SAT • Agilent™ 44k whole human genome microarrays	• repeated-measures ANOVA was used to determine group and fat depot effects • functional analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID)	• lean vs obese with metabolic syndrome – enrichment for genes related to branched-chain amino acid, fatty acid, carbohydrate, and mitochondrial energy metabolism • higher expression of genes involved in polyunsaturated fatty acid biosynthesis was a characteristic of SAT	Klimcakova et al. (2011) (260)
Comparison of morbidly obese and non-obese • 15 morbidly obese, 10 non-obese controls (all women) • SAT • custom-made cDNA Superchip I aminopropylsilane glass slides for 319 genes (PerkinElmer™)	• differential analysis using Mann-Whitney *U* test • pathway analysis	• adipocyte differentiation was increased in morbidly obese patients • expression of $PPAR_{\gamma}1$, a transcription factor that controls the final steps of pre-adipocyte conversion into mature adipocytes, was reduced	Rodriguez-Acebes et al. (2010) (261)
Comparison of SAT and VAT in obese • 53 females, 22 males; all obese • SAT vs.VAT • Illumina™ HumanHT12 BeadChips	• Wilcoxon Mann-Whitney *U* test to identify differentially expressed genes in SAT and VAT • modules of highly co-expressed genes were constructed using pair wise average-linkage cluster analysis	• 3 SAT co-expressed gene modules were inversely associated with plasma HDL-C levels • 1 SAT co-expressed gene module was inversely associated with both plasma glucose and plasma triglyceride levels • these 4 modules were enriched in immune and metabolic genes • genes upregulated in SAT (not in VAT)	Wolfs et al. (2010) (262)

Study design: Participants Tissue Experiment method	**Transcriptomics analysis methods**	**Results**	**References**
		were responsible for cell structure and vitamin metabolism	
Comparison of lean and obese • 67 females, 3 males • SAT • Agilent™ microarray	• differential analysis of lean and obese using SAM analysis (263)	• increased amount of interstitial fibrosis in obese SAT, associated with an infiltration of different types of inflammatory cells	Henegar et al. (2008) (264)
Comparison of obese and non-obese • 11 non-obese and 28 obese women • SAT • custom-made PCR amplified cDNA microarrays (produced at Stanford University)	• SAM analysis (263) to compare obese to non-obese subjects • calculation of correlations between overexpressed genes in SAT of obese subjects and BMI	• 240 genes significantly overexpressed in SAT of obese • Cathepsin S (which affects vascular structure and function) was significant and targeted for further analysis	Taleb et al. (2005) (265)
Comparison of obese and non-obese • 19 obese (10 females, 9 males) and 20 lean (10 females, 10 males) • SAT adipocyte • Affymetrix™ Human Genome U95	• differential analysis using Mann-Whitney *U* test were used to compare the mean expression level for each gene between the non-obese and obese individuals	• inflammation-related genes were upregulated in obese adipocytes	Lee et al. (2005) (255)

Only transcriptomics studies on SAT and adipocytes (excluding weight-loss, dietary intervention studies) are included in the table. ANOVA, analysis of variance; BMI; body mass index, c-DNA complementary-deoxyribonucleic acid; HDL-C, high-density lipoprotein cholesterol, IPA®, Ingenuity Pathway Analysis tool, MHO, metabolically healthy obese, MUO, metabolically unhealthy obese; OXPHOS, oxidative phosphorylation, PCA, principal component analysis; PCR-cDNA, polymerase chain reaction complementary DNA; SAM, Significance Analysis of Microarrays, SAT, subcutaneous adipose tissue; VAT, visceral adipose tissue

2.10. Metabolomics patterns of obesity in plasma

Obesity, because of its effects on the whole body, clearly involves metabolic changes, with dysfunction in different tissues and cells, representing a complex picture (177). Research into the metabolomics of obesity has seen a wealth of information regarding potential biomarkers. Previous studies on obesity consistently implicate BCAAs, acylcarnitines and AAAs (57, 58, 73, 266-269). These circulating levels of BCAAs and AAAs (isoleucine, leucine, valine, tyrosine, and phenylalanine) have now been shown to predict the risk for future T2DM, a condition closely linked to obesity (57, 58, 266, 270). A more comprehensive list of previous obesity-related plasma metabolomics studies is presented in Table 2.

Most of these studies employ association or discriminate analysis to determine metabolite–phenotype associations and find metabolites that are capable of discriminating between groups of people with different clinical phenotypes. Common findings across these studies show that obesity and other cardiometabolic risk factors have consistently been associated with BCAAs, AAAs and lipids.

Table 2: *Previous metabolomics studies (using plasma and serum) in obesity.*

Study design: Participants Experiment method	**Metabolomics analysis methods**	**Results**	**References**
Association study of metabolite profiles to cardio-metabolic phenotype • 2,383 participants,53% women • Liquid chromatography with tandem MS also known as MS/MS	• association analysis using random effects model	• BMI was associated with AAAs and BCAAs • There was considerable overlap in metabolite profiles between BMI, abdominal adiposity, insulin resistance and dyslipidaemia	Ho et al. (2016) (74)
Association study of BCAA metabolites to cardio-metabolic phenotype • 3299 participants; all females • Ultra-high performance liquid chromatography and gas chromatography MS platforms	• association analysis using regression model - utilised twins for within-twin pair analyses and as individuals	• BCAA inversely associated with insulin resistance, inflammation and blood pressure independent of shared genetic and common environmental factors	Jennings et al. (2016) (73)

Study design: Participants Experiment method	Metabolomics analysis methods	Results	References
Comparison of metabolite patterns in obese and non-obese • obese (23 females, 27 males), non-obese (27 females, 21 males), • ^{1}H NMR spectra	• Spearman's rank correlation for correlation analysis between metadata and BMI • Orthogonal partial least-squares (OPLS)-discriminant analysis to identify characteristic metabolites associated with lean and obese phenotypes	• metabolites associating with obesity included sugars, BCAAs, and lipids • the metabolites associating with obesity also associated with lower HDL-C	Saeed et al. (2016) (271)
Association study of age and adiposity to metabolites • combination of 6 studies - 739 participants: 431 females, 308 males • Tandem MS	• dimension reduction using PCA • association analysis using multivariable models	• metabolites of lipid and amino acid metabolism were associated with age and BMI	Kraus et al. (2016) (272)
Comparison of metabolite profiles of T2DM-high-BMI, T2DM-low-BMI and non-diabetic-high-BMI groups as compared to the control group (non-diabetic-low-BMI) • 128 non-diabetics, 17 pre-diabetics, 165 diabetics with each group comprised of equal numbers of both sexes • NMR	• dimension reduction using PCA • discriminate analysis using PLSDA for discriminating between the groups • t-test and ANOVA analysis for differences in groups	• 19 metabolites correlated with T2DM, irrespective of BMI	Gogna et al. (2015) (273)
Association of metabolites to BMI • 300 women • liquid chromatography triple quadrupole MS	• linear regression using BMI as a continuous variable	• 7 metabolites significantly associated with BMI: methyl succinate, asparagine, urate, kynurenic acid, glycine, glutamic acid, and serine	Zhao et al. (2015) (274)
Association study of BMI to metabolites	• linear regression using BMI as a continuous variable	• 37 metabolites (including 19 lipids, 12 amino acids)	Moore et al. (2014) (275)

Study design: Participants Experiment method	**Metabolomics analysis methods**	**Results**	**References**
• 386 females, 561 males • ultra-high performance liquid-phase chromatography and gas chromatography coupled with MS and tandem MS		significantly associated with BMI	
Comparison of lean and overweight/obese men and their metabolite profile • 60 men • ultra-performance liquid chromatography and Q-TOF MS	• PLSDA to classify overweight from obese	• lyso-phosphatidylcholine (C14:0), lyso-phosphatidylcholine (C16:0), and lyso-phosphatidylcholine (C18:1) identified as potential plasma markers for overweight/obese men • abnormal metabolism of BCAAs, AAAs, and fatty acid synthesis and oxidation found in overweight/obese men	Kim et al. (2010)
Comparison of metabolite profiles between obese and lean (blood serum used) • obese (52 females, 22 males), non-obese (38 females, 29 males,) • Tandem MS	• dimension reduction using PCA to consolidate the metabolites • Wilcoxon rank-sum testing to compare means of metabolite levels between lean and obese subjects	• the obese and lean groups differed in levels off BCAA, methionine, Glx (glutamate/glutamine), AAA, and C3 and C5 acylcarnitines	Newgard et al. (2009) (58)

Only metabolomics studies on plasma and serum (excluding time-series studies) are included in the table. ^{1}H NMR, proton nuclear magnetic resonance; AAA, aromatic amino acid; ANOVA, analysis of variance; BCAA, branched chain amino acid; MS, mass spectrometry; PCA, principal component analysis; PLSDA, Partial Least Squares Discriminant Analysis; Q-TOF MS, Quadrupole Time-of-flight Mass Spectrometry; T2DM, Type2 Diabetes Mellitus

3. AIMS OF THE STUDY

The overall aim of this thesis was to study the molecular effects of obesity in SAT and plasma and to relate them to cardiometabolic risk factors as ascertained by clinical measurements (adiposity and blood biochemistry measures). The specific aims were:

- To study the transcriptional profile of SAT in acquired obesity. (Study I)
- To determine if transcriptomics patterns in adipose tissue can be used to profile sub-types in acquired obesity. (Study I)
- To investigate the transcriptional profile of isolated adipocytes in acquired obesity. (Study II)
- To identify the blood metabolite patterns that are associated with different phenotypes including fat depot measurements and selected cardiometabolic risk factors. (Study III)

4. MATERIALS AND METHODS

The materials and methods of the studies covered in this thesis are presented in more detail in each of the original publications (I-III) and summarised here.

4.1. Subjects

4.1.1. FinnTwin16 and FinnTwin12 birth cohorts

Study participants were obtained from two population-based longitudinal birth cohorts: FinnTwin12 and FinnTwin16.

FinnTwin16 (twins born 1975–1979, n=2839 pairs) participants were assessed via five waves of questionnaires at ages 16, 17, 18.5 (276), 25 and 30 years (277). These questionnaires collected information on eating and smoking habits, anthropometry measures, physical activity, alcohol use and various health indicators (276, 277).

FinnTwin12 (twins born 1983–1987, n=2578 pairs) participants were first assessed at 11-12 years old with a follow-up at age 14, then at 17.5, with a follow-up at an average age of 24 (276, 277). Use of alcohol, smoking habits, lifestyle and age-specific items on health behaviour were assessed (276, 277).

4.1.2. Subjects in this study

From both twin cohorts, all available MZ twin pairs who were discordant for BMI (within-pair BMI difference, $\Delta BMI \geq 3$ kg/m^2) according to the last questionnaire were asked to participate in the study, and the BMI discordance was confirmed by phone interview. Additionally, BMI-concordant ($\Delta BMI<3$ kg/m^2) twin pairs were selected as controls. Selected twin pairs were invited to the Obesity Research Unit at the University of Helsinki for clinical assessment: interviews, body composition measurements, blood samples, and adipose tissue biopsies were taken. Altogether 26 BMI-discordant and 14 BMI-concordant MZ twin pairs with all data sets available were included in this thesis. In addition, 14 of the BMI discordant and 5 of the BMI concordant MZ pairs also had isolated adipocytes available for this study.

The twins were healthy, with the exception of one twin with inactive ulcerative colitis (treated with mesalazine and azathioprine), and another twin with T2DM (treated with metformin and insulin).

Figure 6 shows a brief description of the study participants.

Study I comprised 26 BMI-discordant MZ twin pairs (males n=9, females n=17 pairs). Within the 26 twin pairs, the mean BMI difference was 6.0 kg/m^2 and mean age was 30.2 years at the time of clinical study (278). A subset of these BMI-discordant twin pairs also provided the study subjects for Study II. The replication set for Study I consisted of 13 healthy BMI-discordant MZ twin pairs (males n=8, females n=5 pairs) who belonged to a previously published study. The mean BMI difference within-twin pairs was 5.25 kg/m^2 and mean age was 25.6 years (279). The dataset for the replication study was obtained from https://www.ebi.ac.uk/arrayexpress/.

Study II comprised 5 MZ BMI-concordant twin pairs (males n=1, females n=4 pairs) and 14 MZ BMI-discordant twin pairs (males n=3, females n=11 pairs).

Study III comprised 26 BMI-discordant and 14 BMI-concordant MZ twin pairs (males n=36, females n=44) (276). The mean BMI was 27.9 kg/m^2 and mean age 30.7 years at the time of clinical study. Participants from Study I and II are a subset of participants from Study III.

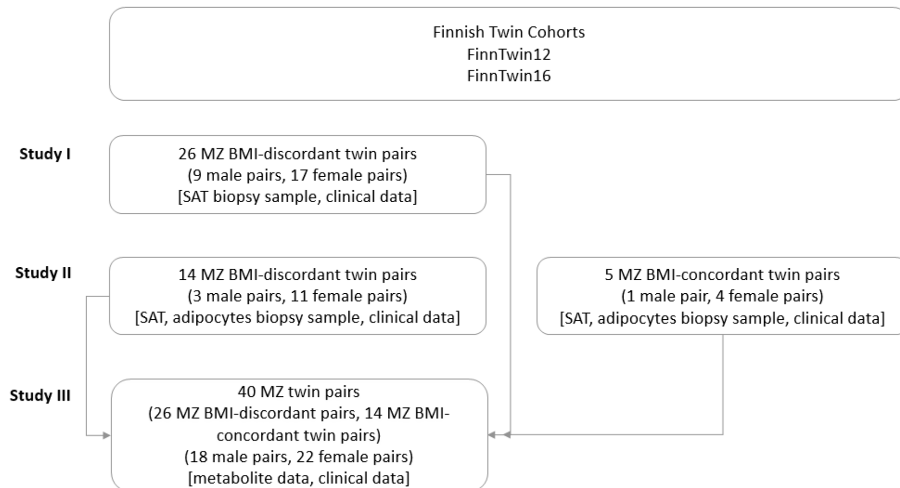

Figure 6: *Study I, II, III participants. Study I, II, III used participants from two Finnish twin cohorts. Participants from Study I and II are a subset of participants from Study III. BMI, body mass index, MZ, monozygotic; SAT, subcutaneous adipose tissue*

4.2. Ethical considerations

Written informed consent was obtained from all participants. The study protocol was designed in accordance of the principles of the Helsinki Declaration with approval from the Ethics Committee of the Helsinki University Central Hospital. The related ethics permissions are Dnro 249/E5/01 and Dnro/13/03/01/2008.

4.3. Measures

SAT gene expression and plasma metabolite data were acquired to ascertain molecular-level disturbances in the adipose tissue as well as plasma that co-occur with obesity. Clinical measurements that include anthropometric, body composition and metabolic clinical measures were obtained. Details of food intake, physical activity and smoking status were also recorded.

4.3.1. Blood tests and clinical measurements

Figure 7 below shows all of the blood tests and clinical measures taken from the study subjects.

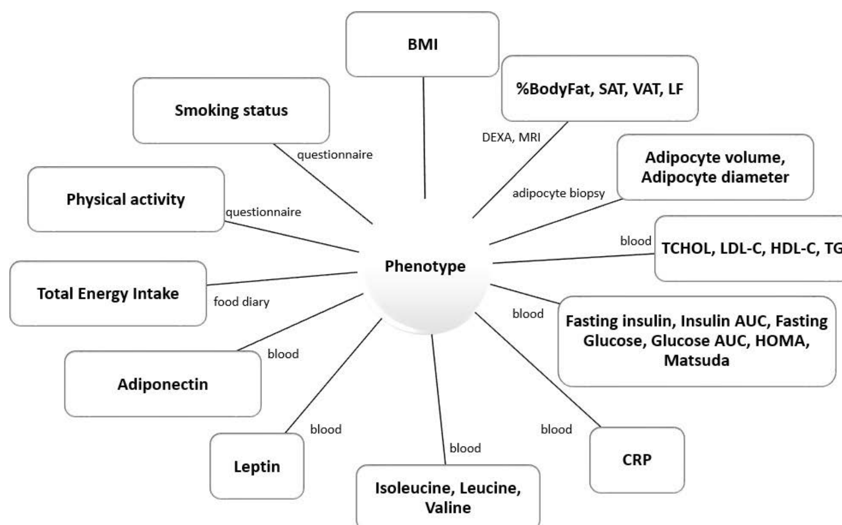

Figure 7: *Phenotype measures. AUC, area under the curve; BMI, body mass index; CRP, C-reactive protein; DEXA, dual-energy X-ray absorptiometry; HDL-C, high-density lipoprotein cholesterol; HOMA, homeostatic model assessment; LDL-C, low-density lipoprotein cholesterol; LF, liver fat; MRI, magnetic resonance imaging; SAT, subcutaneous adipose tissue; TCHOL, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue*

BMI was calculated from height and weight, which were measured from participants in their underwear after a 12-hour overnight fast. Dual-energy X-ray absorptiometry (DEXA) (280) provided body composition measures including %bodyfat while magnetic resonance imaging (MRI) determined the body fat distribution of abdominal SAT and VAT (173). Proton magnetic resonance spectroscopy (^{1}H-MRS) ascertained the percentage of LF (173).

Fasting lipids (comprising total cholesterol [TCHOL], HDL-C, LDL-C and TG) and high-sensitivity C-reactive protein (hs-CRP) were determined from blood samples. Similarly, measurements of glucose and insulin during a 2-h oral glucose tolerance test (OGTT) were taken; HOMA (homeostasis model assessment) was calculated from fasting insulin and glucose values while Matsuda was computed from the fasting insulin, fasting glucose, insulin area under the curve and glucose area under the curve values (278).

The plasma BCAAs (valine, leucine and isoleucine) were measured using a high-throughput NMR metabolomics platform (281). Plasma leptin and adiponectin were measured by enzyme-linked immunosorbent assay (ELISA) using DuoSet ELISA (R&D Systems Europe Ltd, Abindgon, UK) (174). Nutritional intake was recorded using a 3-day food diary, while physical activity was measured using Baecke indices (159, 282). Current daily smoking status for the twins was obtained via questionnaire.

4.3.2. Subcutaneous adipose tissue and adipocyte gene expression

SAT biopsies were taken after an overnight fast (12-h) via a surgical technique under local anaesthesia (159). The RNeasy Lipid Tissue Mini Kit (Qiagen, Venlo, The Netherlands) was used to extract SAT RNA (159) which was then treated with DNase I (Qiagen, Venlo, The Netherlands) and hybridised to an Affymetrix™ U133 Plus 2.0 microarray (Thermo Fischer, Santa Clara, CA, USA) (33). Scanning of the microarray provided light intensity measurements that were then converted into gene expression values (190).

QC assessment was done using the simpleaffy and affyPLM packages of Bioconductor to assess the quality of the data. Checks addressed the quality of the RNA, signal quality of the array, dataset homogeneity as well as comparability across the arrays (samples) (226, 227). In detail, the QC steps involved creating different plots to ascertain the quality of the data:

- RNA Degradation Plot to detect samples that had degraded more than others in the same dataset
- Beta-actin/Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) plot to detect any samples that may contain heavily degraded RNA

- Perfect Match (PM) intensity plot to detect if any of the arrays had problems in amplification or labelling
- Positive and negative border elements plot to detect dissimilarity that arises either from non-uniform hybridisation or gridding problems
- Array–array intensity correlation plot to test data homogeneity
- Normalised Unscaled Standard Error (NUSE) plot to detect signal variability between probes from the same probe set

The gene expression data passing the QC steps (in this case, all) was then used to perform subsequent gene expression analyses in Studies I and II.

A piece of fresh adipose tissue was digested in 2% collagenase–Dulbecco's Modified Eagle Medium/ Nutrient Mixture F-12 (DMEM/F-12) to isolate adipocytes for further processing (174). Adipocyte volume and diameter were measured (using ImageJ) from 200 SAT adipocytes per person, that were photographed with a light microscope (283, 284). Adipocytes were also subjected to RNA extraction and hybridised to an Affymetrix™ U133 Plus 2.0 microarray (Thermo Fischer, Santa Clara, CA, USA) (33), as described above, for further gene expression analysis in Study II.

4.3.3. Plasma metabolite measures

After a 12-h fast, blood samples were taken from participants. Metabolites were extracted from the plasma samples using acetonitrile (1:4, sample:solvent) and detected using ACQUITY™ Ultra Performance Liquid Chromatography (UPLC)-MS/MS system (Waters Corporation, Milford, MA, USA) (285). MassLynx 4.1 software was used for data acquisition, data handling and instrument control. Data was processed using TargetLynx software and metabolites quantified using external calibration curves.

4.4. Analysis techniques

The analysis techniques and the work flow for Studies I-III are presented in Figures 8-10 below. For analysis in R, the version used was R version 3.2.0.

In Study I (Figure 8), SAT was extracted and clinical measurements recorded from 26 BMI-discordant MZ twin pairs. The mRNA from the SAT were hybridised to Affymetrix™ U133Plus2.0 GeneChip microarrays. The resulting gene expression data was normalised (package gcrma in R) (198), annotated with the Brainarray custom chip description file (CDF) version 18 (286) and analysed further. First, differential analysis (moderated t-test with package limma in R) was conducted to identify within-twin pair differentially expressed genes. The significant genes (FDR *p*-value<0.05) were analysed further using the Ingenuity® Pathway Analysis (IPA®, Qiagen, Venlo, The Netherlands) in order to identify the pathways associated with these genes.

In Study I, the gene expression ratios (heavy/lean) across all gene transcripts for each twin pair were calculated. Because these ratios are calculated for each twin pair, they represented obesity that is acquired and not driven by genetics nor shared environmental factors in the co-twins. These ratios of gene expression values (one value per pair per gene) were used to perform average-linkage hierarchical clustering using squared Euclidean distance (package hclust in R) (287) in order to identify which twin pairs grouped together (i.e., had similar gene expression ratio profiles). Dendrograms were used to visually inspect the clusters. For each cluster identified, within-twin pair differential analysis (moderated t-test with package limma in R) was conducted. The resulting genes were analysed with IPA® to 1) identify, in each cluster, pathways which the differentially expressed genes belong to and 2) compare the enriched pathways in each cluster against the top 10 pathways identified earlier in the 26 twin pairs.

In order to determine if the within-twin pair differences in gene expression in each cluster were accompanied by within-twin pair differences in clinical measurements, one-way analysis of variance (ANOVA) with Tukey post hoc test in R software was performed for the clinical measurements. Here, the differences in clinical measurements within-twin pairs were used and the mean of the differences compared across clusters.

The replication dataset was used to verify the cluster analysis results.

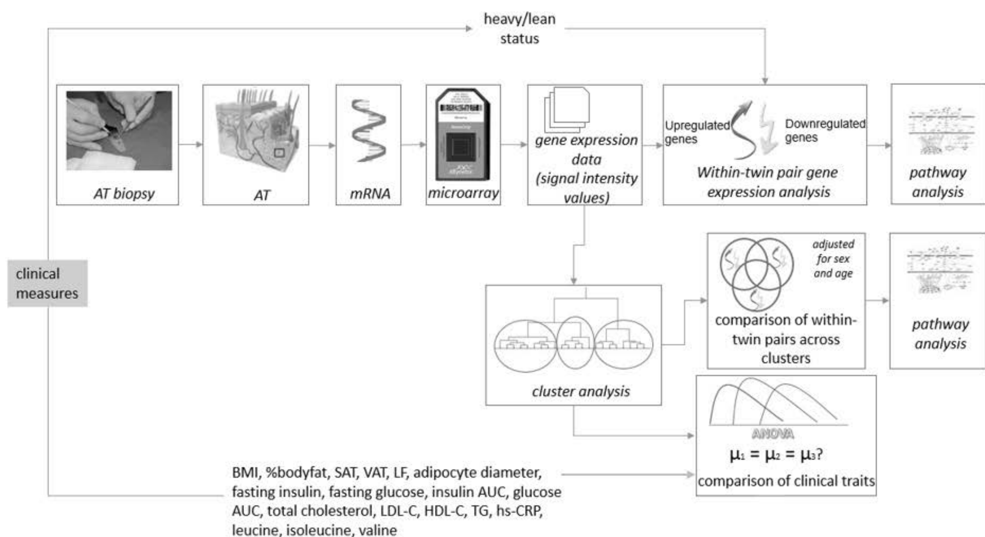

Figure 8: *Analysis flow (Study I). Subcutaneous adipose tissue gene expression was measured using microarrays. The gene expression data was then analysed for within-twin pair gene expression differences, followed by pathway analysis. The gene expression data was also used to perform clustering of the twin pairs.*

Resulting clusters were examined for within-twin pair gene expression differences. Pathway analysis was conducted for each cluster. Within-twin pair clinical measure differences was compared across the clusters. ANOVA, analysis of variance; AUC, area under the curve; AT, adipose tissue; BMI, body mass index; hs-CRP, high-sensitivity C-reactive protein; LF, liver fat; VAT, visceral adipose tissue; SAT, subcutaneous adipose tissue; LDL-C, low-density lipoprotein; HDL-C, high-density lipoprotein

In Study II (Figure 9), SAT was extracted and clinical measurements recorded from 14 BMI-discordant MZ twin pairs and 5 BMI-concordant MZ twin pairs. Additionally, adipocytes were isolated from the BMI-discordant MZ twin pairs. The mRNA from the SAT and adipocytes were hybridised to Affymetrix™ U133Plus2.0 GeneChip microarrays. The resulting gene expression data was normalised (package gcrma in R) (198), annotated with the Brainarray CDF version 18 (286) and analysed further. First, using the data from the adipocytes, differential analysis (moderated t-test with package limma in R) was conducted to identify within-twin pair differentially expressed genes. Using FDR *p*-value<0.05 as a cut-off for significance, the genes were analysed further using the IPA® analysis tool in order to identify the pathways associated with these genes. Next, using the data from the SAT, differential analysis (moderated t-test with package limma in R) was conducted to identify within-twin pair differentially expressed genes; the SAT data from the BMI-discordant twin pairs were used as controls. The significant genes (FDR *p*-value<0.05) were analysed further using the IPA® analysis tool in order to identify the pathways associated with these genes. The Pearson correlations for these pathways (both SAT and adipocytes-related) with adiposity and clinical measures were calculated. Additionally, significantly differentially expressed genes were checked against the MitoCarta database to determine which ones were mitochondria-related.

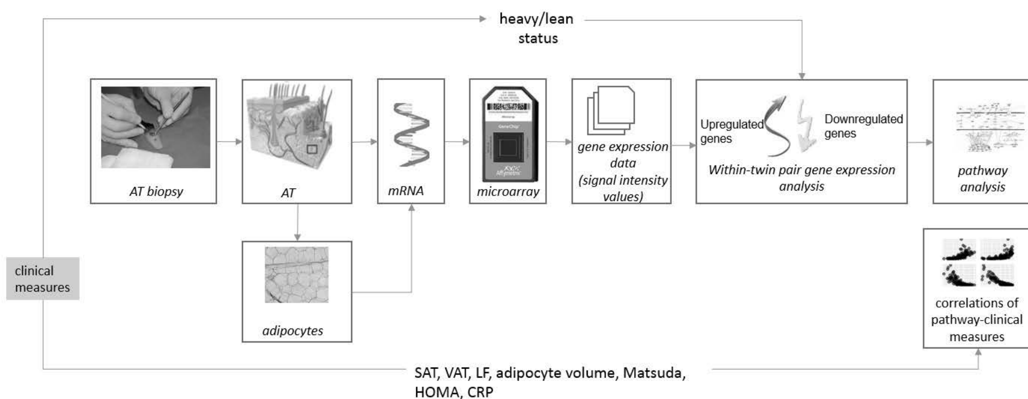

Figure 9: *Analysis flow (Study II). Subcutaneous adipose tissue and adipocyte gene expression were measured using microarrays. The gene expression data was then analysed for within-twin pair gene expression differences, followed by pathway analysis. AT, adipose tissue; LF, liver fat; VAT, visceral adipose tissue; SAT, subcutaneous adipose tissue; hs-CRP, high-sensitivity C-reactive protein; HOMA, index homeostatic model assessment*

In Study III (Figure 10), 80 MZ twins were studied as individuals. Plasma samples were analysed using the Acquity ultra high performance liquid chromatography (UPLC) system coupled to a triple quadrupole mass spectrometer. Adiposity and blood biochemistry measures were also recorded for each individual. While the previous two studies investigated acquired obesity (owing to non-shared environmental factors) by using ΔBMI>3 as the segregating factor between heavy and lean co-twins, this study looked at varying adiposity and blood biochemistry measures in order to identify associations (owing to both genetic and environmental factors) with circulating metabolites. First, associations of metabolites with these adiposity and blood biochemistry measures were identified. Then, to see which association were free from genetic and shared environment factors, within-twin pair analysis was performed.

Pearson correlations of the 80 individuals were performed using data adjusted for sex, family and smoking effects. First, the phenotype data was modelled using a linear mixed model using only sex, family and smoking as effects. Then, residual values were obtained for each phenotype. These residuals, free from the effects of sex, family and smoking, were used in the Pearson correlation calculations. The reason the residuals were used was to ensure the samples were independent, and biological effects associated with sex, family and smoking were removed before correlations were calculated.

Metabolites with more than 30% missing values across the samples were excluded from further processing. Shapiro-Wilk test for normality was conducted on the metabolite data to check for normality in the distribution. Metabolite data was rank transformed (package GenABEL in R) (288) to a standard normal distribution with a mean of zero and variance of one. The rationale for choosing rank transformation to normality was to reduce technical noise (by using rank values instead of measured values) and to ensure all the metabolites were on the same scale and hence comparable.

The associations between each phenotype (adiposity and blood biochemistry measures) and each metabolite across all individuals (n=80) were calculated using linear mixed model (package lme4 in R) regression analysis. This model was chosen so that genetic relatedness within each twin pair could be factored in using the random effects part of the model. To prevent overfitting the data, the predictor variables in the model were selected using the (LASSO) method (219). Family relatedness, sex, smoking, age and batch number were retained as independent variables. Bonferroni p-value<0.05 was considered statistically significant.

Because there is no consensus on which cardiometabolic risk factors are more important than the others (289), metabolite-phenotype associations were compared to reveal the relative importance of each of the phenotypes. For metabolites significantly associated with one or more phenotypes, the standardised beta coefficients (in units of standard deviation) were compared to determine which associations had stronger effect sizes. Additionally, within-twin pair differential analysis (moderated t-test with package limma in R) was performed to determine associations, already identified in the

association analysis using the linear mixed model, that were independent of genetic and shared environmental effects. Both regression and moderated t-test models were adjusted for family relatedness, sex, smoking, age and experiment batch number.

Lastly, in order to determine whether individuals form separate clusters based on their metabolite profiles, K-means clustering was performed. Metabolites were modelled as dependent variables using a linear mixed model with only age, sex, family and smoking status, batch number as random effects. The residual values for each metabolite in each sample was obtained and now represented metabolite concentration values with the effects of age, sex, family and smoking, batch number removed. These residual values were dimension reduced using PCA in order to capture the variance in the metabolite data while accounting for collinearity between the metabolites. Because the metabolites were rank transformed to standard normal before extraction of the residuals, the metabolites did not have differing variances; this prevents the variables with the largest variances from dominating the first few PCs. Scree plots were produced to show the fraction of total variance in the data as represented by each PC. Using enough PCs to capture 80% of the variance in the data, K-means clustering was performed in order to cluster the individuals according to their metabolite profiles.

Between-cluster differences were determined using differential analyses (package limma in R) to ascertain if metabolite concentrations between the clusters differed significantly. A Bonferroni-corrected p-value<0.05 was considered significant for the differential analyses. The significantly differing metabolites were analysed using IPA® to find pathways associated with these metabolites. In order to determine if the differences in metabolite concentration in each cluster were accompanied by significant differences in clinical measurements, Welch t-tests were conducted.

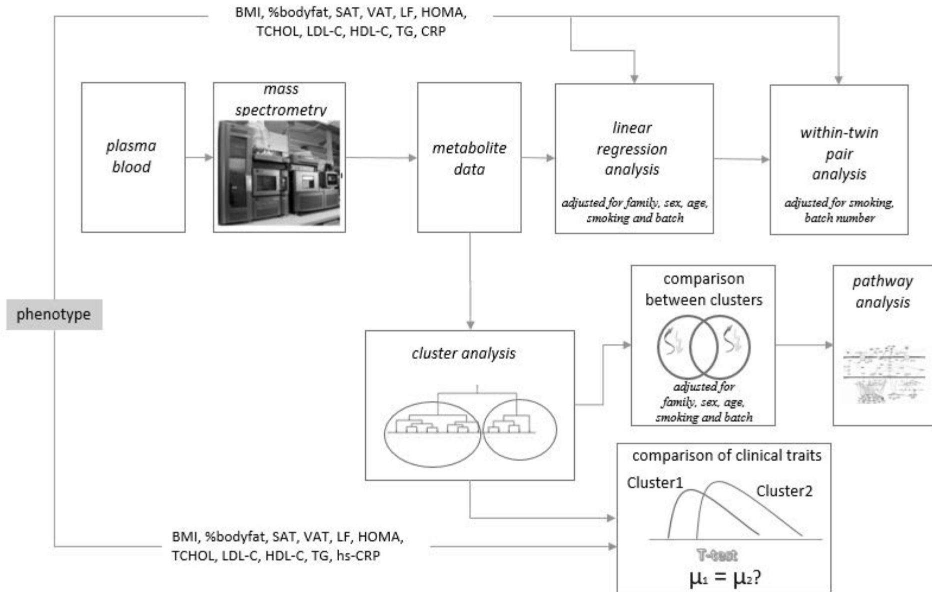

Figure 10: *Analysis flow (Study III). Plasma metabolites were extracted. The quantity of each metabolite was measured for each individual. The metabolites and phenotype measures (adiposity and blood biochemistry) were then analysed using linear regression analysis as well as within-twin pair differential analysis. The metabolite data was also used to perform clustering of the samples. Resulting clusters were examined for cluster differences, pathway analysis was conducted for the significantly different metabolites. Phenotype measure between clusters were compared. BMI, body mass index; HDL-C, high-density lipoprotein; HOMA, index homeostatic model assessment; hs-CRP, high-sensitivity C-reactive protein; LDL-C, low-density lipoprotein; LF, liver fat; SAT, subcutaneous adipose tissue; TCHOL, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue*

4.5. Analysis methods common to the studies (Study I, II, III)

4.5.1. Linear and linear mixed model analysis (Study I, II, III)

Within-twin pair analysis was performed using the linear model for microarray data (package limma in R). Association analysis with the twin individuals was performed using the linear mixed model

(package lme4 in R). For the gene expression analysis, significance was set at 0.05 for FDR correction. For the metabolite data analysis, significance was set at 0.05 for Bonferroni adjustment.

Analysis details for within-twin pair analysis (Study I, II, III):
After normalisation of the data, a design matrix (model) (package limma in R) was specified and the normalised data fit to it. Family ID was used to identify the twin individuals belonging to the same twin pair. A binary value of 1 or 2 denoted the "HeavyLean" status ("1" for the co-twin with the smaller BMI and "2" for the co-twin with the higher BMI).

In Study I and II, the gene expression was the dependent variable with HeavyLean status as explanatory variable and FamilyID as covariate in the model. In Study III, the metabolite concentration was the dependent variable with one of the adiposity or blood biochemistry measures as the explanatory variable, and FamilyID and smoking status as covariates in the model.

The gene transcripts/metabolites that had significant associations with the explanatory variable along with the fold-changes for the differential expression/concentration were then extracted.

In Study II, the within-twin pair model was first run for the discordant twin pairs and then for the concordant twin pairs using the adipocyte data. Genes that were also differentially expressed in concordant twin pairs were removed from the final list of genes differentially expressed within the BMI-discordant twin pairs. The within-twin pair analysis was repeated for adipose tissue data.

Analysis details for differential analysis in clusters (Study I, Study III):
After normalisation of the data, a design matrix (model) (package limma in R) was specified and the normalised data fit to it. ClusterNo was defined as the unique identifier for each cluster. FamilyID was used to identify the twin individuals belonging to the same twin pair. Genetic relatedness within each twin pair were calculated using the duplicateCorrelation function in limma.

In Study I, gene expression was the dependent variable with the interaction between ClusterNo and HeavyLean status as the explanatory variable, and age, sex, genetic relatedness as the covariates. In Study III, metabolite concentration was the dependent variable with age, sex, smoking, experiment batch number, ClusterNo and FamilyID as covariates.

Analysis details for association analysis (linear mixed model) with twin individuals (Study III):
After normalisation of the metabolite data, a regression model was created (package lme4 in R) and the normalised data fit to it. The outcome variable in this univariate model was the metabolite with the phenotype (adiposity or blood biochemistry measure), smoking, experiment batch number, age, sex as fixed effects, and the family identifier as the random effect. A binary value of 0 or 1 denoted

the smoking status ("0" for non-smokers and "1" for smokers).The regression coefficient for each phenotype was then extracted along with the association p-values and Bonferroni adjusted p-values.

4.5.2. Pathway Analysis (Study I, II, III)

For the gene expression data, differentially expressed genes within twin pairs along with the fold-change and adjusted p-values were uploaded to the IPA® tool. Using the IPA®'s Comparison Analysis tool, the genes differentially expressed in each cluster were compared to the original top 10 differentially expressed pathways within the 26 twin pairs. This was done to see how each cluster compares against the original top 10 pathway findings and if any one cluster was driving the original top 10 pathways. Each cluster's pathways were determined using IPA®'s Canonical Pathway analysis. A Fisher's exact test $p<0.05$ was considered significant.

For metabolite data, metabolites with significantly different concentrations between the clusters were examined using the IPA® tool. This was done to find pathways that associated with these metabolites.

4.5.3. Clinical measurements analysis (Study I, II, III)

In Study I and II, the co-twins were identified as "lean" or "heavy" depending on their BMI. In order to compare whether the remaining clinical measurements (other than BMI) within-twin pairs were significantly different, Wilcoxon signed rank tests in R (chosen because the samples were related) were performed.

For the cluster analysis using the 26 twin pairs, one-way analysis of variance (ANOVA) with Tukey post hoc test in R software compared between clusters the mean differences of clinical measures within pairs. Histogram plots showed near normal distribution for each clinical phenotype and hence the ANOVA test was considered suitable. An adjusted p-value≤0.05 was considered significant.

For the cluster analysis using the 80 twin individuals, means of the clinical measurements were compared between clusters using the Welch t-test (package psych in R). This test was considered suitable because of the unequal sample size in the two clusters. The data was adjusted for family, sex and smoking before comparison of the mean values of each clinical measure.

5. RESULTS

5.1. Characteristics of the twins (I, II, III)

Detailed characteristics of the twins (n=52 used in Study I, n=38 used in Study II, n=80 used in Study III) can be found in each publication. The twins used in each publication were from a set of 40 MZ twin pairs composed of altogether 26 BMI-discordant and 14 BMI-concordant twin pairs; Study I used 26 BMI-discordant twin pairs from this set, while Study II used 14 BMI-discordant and 5 BMI-concordant twin pairs.

Characteristics of the whole sample of 40 twin pairs (80 individuals) (Study III), BMI-discordant and BMI-concordant pairs (Study I, II) are presented in Table 3.

In the 26 BMI-discordant twin pairs, the lean and heavy co-twins were highly discordant (*p*-value<0.0001) for all measures of adiposity (%bodyfat, SAT, VAT, LF, adipocyte volume and adipocyte diameter). The heavy co-twins also had higher values of fasting insulin, insulin area under the curve (AUC) during OGTT, LDL-C, triacylglycerol, serum hs-CRP, leptin and all BCAAs. Additionally, the heavy co-twins had lower values of HDL-C and adiponectin. In the 5 BMI-concordant twin pairs, the body composition and metabolic clinical measures were similar.

Table 3: *Characteristics of the BMI-discordant and BMI -concordant twin pairs.*

	80 MZ twin individuals **(mean age = 30.73 years; males n=36, females n=44)** **Study III**	BMI-discordant pairs (ΔBMI > 3 kg/m², n=26 pairs) **(mean age = 30.2 years; males n=9, females n=17 pairs)** **Study I, II**			BMI-concordant pairs (ΔBMI < 3 kg/m², n=5 pairs) **(mean age = 29.7, males n=1, females n=4 pair(s))** **Study II**		
Variable		Lean co-twin (mean ± SE)	Heavy co-twin (mean ± SE)	*p*-value	Lean co-twin (mean ± SE)	Heavy co-twin (mean ± SE)	*p*-value
BMI (kg/m²)	27.8±0.6	25.3±0.9	31.3±1.0	<0.0001	28.2±1.9	30.4±1.	0.063
%bodyfat	34.2±1.1	32.3±1.8	41.1±1.3	<0.0001	35.7±4.85	36.9±4.0	0.63
SAT (cm³)	4511.3±280.1	3813.7±416.8	6358.9±540.4	<0.0001	4234.5±668.5	4754.7±656.9	0.13
VAT (cm³)	1147.2±113.4	790.2±178.9	1643.7±247.4	<0.0001	782.1±293.6	954.3±356.4	0.19
Liver fat (%)	2.9±0.5	1.12±0.3	4.52±0.9	<0.0001	1.32±0.6	2.6±1.8	0.81
Adipocyte diameter (μm)	85.4±1.9	80.88±2.54	95.02±2.82	<0.0001			
Adipocyte volume (pL)	440.8±28.9	355.6±3	547±59	<0.001	389.2±82.9	443.9±66.5	0.08

	80 MZ twin individuals **(mean age = 30.73 years; males n=36, females n=44)** **Study III**	BMI-discordant pairs (ΔBMI > 3 kg/m², n=26 pairs) **(mean age = 30.2 years; males n=9, females n=17 pairs)** **Study I, II**			BMI-concordant pairs (ΔBMI < 3 kg/m², n=5 pairs) **(mean age = 29.7, males n=1, females n=4 pair(s))** **Study II**		
Variable		Lean co-twin (mean ± SE)	Heavy co-twin (mean ± SE)	*p*-value	Lean co-twin (mean ± SE)	Heavy co-twin (mean ± SE)	*p*-value
fP-glucose (mmol/L)	5.30±0.05	5.1±0.1	5.3±0.1	0.17	5.3±0.24	5.28±0.2	0.85
AUC Glucose in Oral Glucose Tolerance Test (mmol/l/h)	14.34±0.3	14.16±2.7	14.85±2.8	0.35	13.97±1.4	13.49±1.6	0.813
fS-insulin (mU/L)	6.35±0.5	4.9±0.5	8.5±1.2	0.0011	6.06±1.52	6.44±0.7	0.625
AUC Insulin (mU/l/h)	93.06±0.03	87.6±8.0	129.3±24.6	0.031	68.24±18.8	80.05±24.3	0.438
Total cholesterol (mmol/L)	4.5±0.1	4.4±0.2	4.7±0.2	0.14	4.46±0.27	4.7±0.4	0.58
LDL cholesterol (mmol/L)	2.8±0.9	2.6±0.1	3.0±0.2	0.14	2.8±0.07	2.84±0.3	0.99
HDL cholesterol (mmol/L)	1.4±0.04	1.6±0.1	1.3 ± 0.1	0.00040	1.2±0.1	1.34±0.2	0.44
Triglycerides (mmol/L)	1.1±0.08	0.94±0.1	1.32 ± 0.2	0.014	1.09±0.4	1.11±0.4	0.99
fS-hs-CRP (mg/dL)	2.5±0.5	2.6±0.7	4.0 ± 1.1	0.065	0.97±0.4	1.37±0.4	0.44
fP-Adiponectin (ug/mL)	3.1±1.7	2.8±0.3	2.2±0.2	0.0023	3.9±1.4	2.7±4.1	0.13
fP-Leptin (mg/mL)	21.2±2.2	18.9±4.1	34.6±5.5	0.0015	23.8±8.4	25.6±10.4	0.63
Isoleucine (mmol/litre)		0.05±0.01	0.06±0.02	0.033			
Leucine (mmol/litre)		0.08±0.01	0.09±0.02	0.0094			
Valine (mmol/litre)		0.19±0.04	0.21±0.04	0.024			
Smokers		6 out of 26 individuals (3 smokers have a smoker twin; 3 smokers have a non-smoking heavier twin)	6 out of 26 individuals (3 smokers have a smoker twin; 3 smokers have a non-smoking leaner twin		1 individual		

Wilcoxon's rank sum test was used to compare values of the lean versus the heavy co-twin. AUC, area under the curve; fP, fasting plasma; fS, fasting serum; HDL, high-density lipoprotein; HOMA, homeostatic model assessment; hs-CRP, high-sensitivity C-reactive protein; LDL, low-density lipoprotein; OGTT, oral glucose tolerance test; SAT, subcutaneous adipose tissue; SE, standard error; VAT, visceral adipose tissue

5.2. Correlations between phenotypes (I, II, III)

5.2.1. Correlation between the phenotype measures

Pearson correlations for the 11 phenotypes (BMI, adiposity measures (%bodyfat, SAT volume, VAT volume, %LF), and blood biochemistry measures) are in Figure 11. Amongst the measures, the strongest positive correlations were between BMI and SAT (0.94) and between total cholesterol and LDL-C (0.91). The strongest negative correlations were between VAT and HDL-C (-0.54) and between BMI and HDL-C (-0.50).

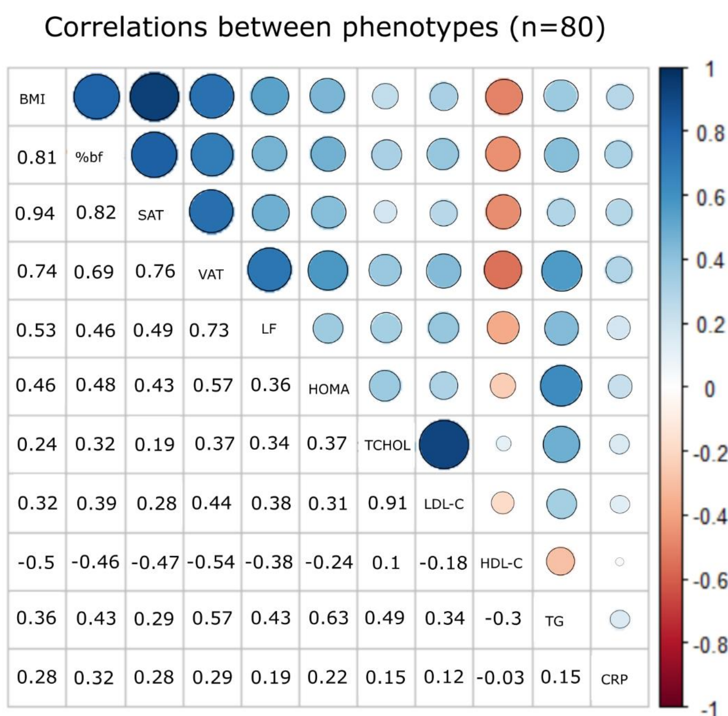

Figure 11: *Correlations between adiposity and clinical measures. Pearson correlation analysis was performed on phenotype measures from 80 individuals. Residual values from a linear mixed model were used to ensure sex and family effects do not influence the correlations. High positive correlations (>0.7) were observed between BMI and body composition measures: BMI and %bodyfat, BMI and SAT, BMI and VAT, %bodyfat and SAT, SAT and VAT and VAT and LF as well as LDL-C and Cholesterol. %bf, percentage of body fat; BMI, body mass index; CRP, C-reactive protein; HDL-C, high-density lipoprotein cholesterol;*

HOMA, homeostatic model assessment; LDL-C, low-density lipoprotein cholesterol; LF, liver fat; SAT, subcutaneous adipose tissue; TCHOL, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue

5.3. Subcutaneous adipose tissue gene expression in acquired obesity (I)

In this thesis, annotations were first done using the Affymetrix™ HGU133plus2 CDF. Due to the availability of custom CDFs that have more recent annotations, the analysis was also conducted using Brainarray CDFs. This additional analysis was done for comparison. The analysis using Brainarray CDFs yielded more gene transcripts than the Affymetrix™ CDF. However, 30% of the transcripts identified in the Affymetrix™ CDF did not exist in the list produced with Brainarray CDF. Pathway analysis done on these 30% differing gene transcripts did not yield any biological pathways that made sense regarding obesity. Hence, the results from the custom CDFs were retained.

5.3.1. Within-twin pair differences in transcriptomics patterns of SAT

A total of 2108 genes were differentially expressed (FDR p-value<0.05) between heavy and lean co-twins in the SAT of the 26 twin pairs. Amongst the top 10 (smallest adjusted p-values selected) downregulated genes in the heavy co-twins (compared to the lean co-twins) were those involved in lipid biosynthesis, lipid and FA metabolism, and cell signalling processes. Amongst the top 10 upregulated genes in the heavy co-twins (compared to the lean co-twins) were those involved in apoptosis and immune response.

Top pathways (Fisher's exact test p-value<0.05) enriched for genes differentially expressed in the SAT of the heavy co-twins compared to the lean co-twins were:

1. Oxidative Phosphorylation
2. Valine Degradation I
3. Mitochondrial Dysfunction
4. IL-8 Signalling
5. Triacylglycerol Biosynthesis
6. Glutaryl-CoA Degradation
7. Ketogenesis
8. Ketolysis
9. Isoleucine Degradation I
10. Nur77 Signalling in T Lymphocytes
11. Unfolded Protein Response

The top most significant (Fisher's exact test p-value<0.05) pathways were mitochondria-related. There was lower oxidative phosphorylation (OXPHOS) and mitochondrial function in the heavy co-twins. The Mitochondrial Dysfunction pathway represents a biological pathway that is activated when there is dysfunction of the mitochondria. The Mitochondrial Dysfunction pathway consists of the same downregulated genes as in the OXPHOS pathway; additionally, the Mitochondrial Dysfunction pathway had upregulated genes *GSR* and *CYB5R3* which have roles in reducing the damage caused

by a build-up of mitochondrial oxidative stress. Genes in the Mitochondrial Dysfunction (most of which, other than *GSR* and *CYB5R3* that were upregulated, overlapped with OXPHOS) and OXPHOS pathways, Valine Degradation I and Isoleucine Degradation I were downregulated in the heavy co-twins.

Nur77 Signalling in T Lymphocytes (with a role in inflammation) and Unfolded Protein Response (linked to endoplasmic reticulum stress) pathways were also enriched (showing both up- and downregulated genes) in the heavy co-twins.

5.3.2. Adipose tissue gene expression profiles in subgroups of the twin pairs (I)

Genome-wide gene expression differences within each twin pair were used in order to ascertain if the twin pairs cluster into separate groups. For each gene, each twin pair was represented by a within-pair gene expression ratio (heavy/lean). This collection of ratios across the genome represented the effects of acquired obesity within twin pairs. Based on their expression profiles, the twin pairs formed three distinct subgroups (clusters) of acquired obesity. These groups are referred to as Cluster 1, Cluster 2 and Cluster 3. Cluster 1 had 2 twin pairs, Cluster 2 had 19 and Cluster 3 had 5. Within the twin pairs, Cluster 1 had 413, Cluster 2 had 728, and Cluster 3 had 828 differentially expressed genes (FDR p-value<0.05).

Pathways that differed within twin pairs in Cluster 1 (Figure 12a) were enriched for lipid metabolism and signalling pathways, but did not reach significance to indicate a clear upregulation or downregulation of pathways. Cluster 2 (Figure 12b) showed enrichment for pathways involved in mitochondrial function. Both OXPHOS and valine degradation pathways showed downregulation of genes in the heavy co-twins of this cluster. Cluster 3 showed significant upregulation (Figure 12c) of the inflammation-related pathways and downregulation of valine degradation, a mitochondrial pathway, in the heavy co-twins. These cluster findings were also evident in the replication dataset (detailed in Chapter 4.1.2) in which three clusters were also found (Figure 13).

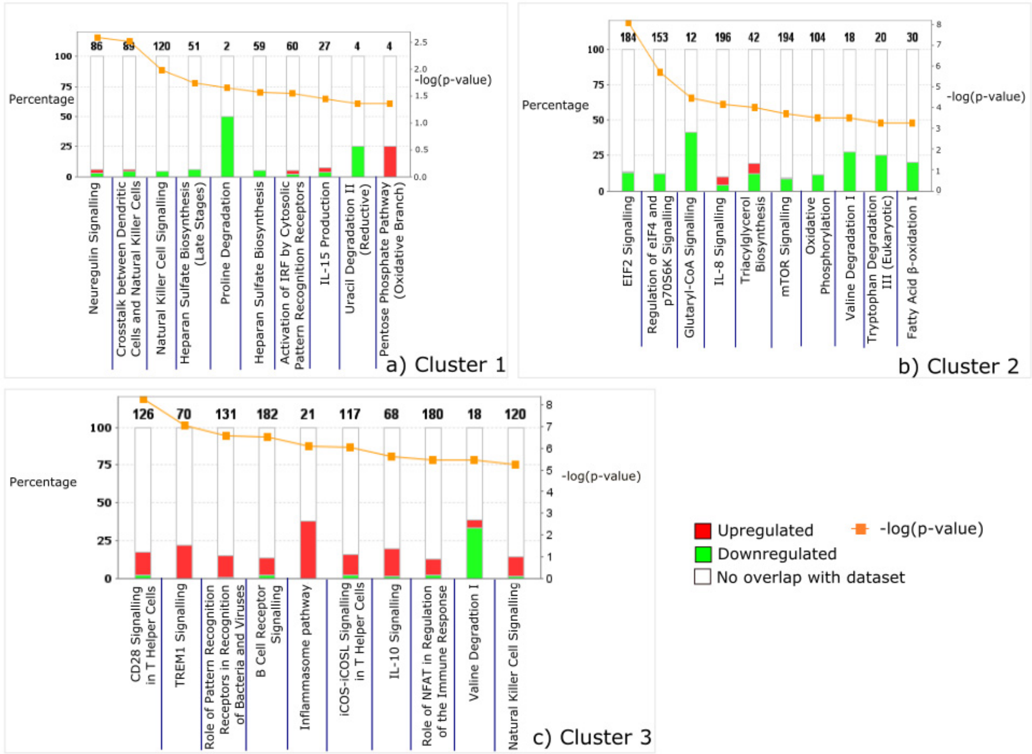

Figure 12: *The top 10 pathways of the subgroups (clusters) based on the within-twin pair differences in gene expression. Figure reprinted with the permission obtained from the International Journal of Obesity. Differentially expressed genes for each cluster (FDR p-value<0.05, n=274 for Cluster 1, n=728 for Cluster 2, n=828 for Cluster 3) were entered into IPA® to produce the pathways. (a) Cluster 1, (b) Cluster 2, and (c) Cluster 3. The y-axis displays the -log of p-values, which is calculated by Fisher's exact right-tailed test. A -log (p-value) of 1.3 is indicative of a p-value of 0.05. The percentage of upregulated (red) and downregulated (green) genes in the heavy co-twins in the dataset is represented. The white blocks represent the genes that belong to the pathway according to IPA® analysis but did not reach significance. For some pathways, IPA® was able to conclusively provide activation scores: z-scores >2 or <-2 are considered significant. For Cluster 2, IPA® predicted the following pathway to be downregulated: EIF2 Signalling (z=-2.309), and the following pathways to be upregulated: mTOR Signalling (z=0.378) and IL-8 Signalling (z=1.414). Mitochondrial pathways, oxidative phosphorylation, valine degradation had genes that were downregulated in the heavy co-twins but did not reach z-score significance. For Cluster 3, IPA® predicted the following pathways in Cluster 3 to be upregulated: CD28 Signalling in T Helper Cells (z=2.688), TREM1 Signalling (z=4.123), B Cell Receptor Signalling (z=3.128), Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses (z=4.00), Role of NFAT in regulation of the immune response (z=3.578), and iCOS-iCOSL Signalling in T Helper Cells (z=2.309). IPA®, Ingenuity Pathway Analysis*

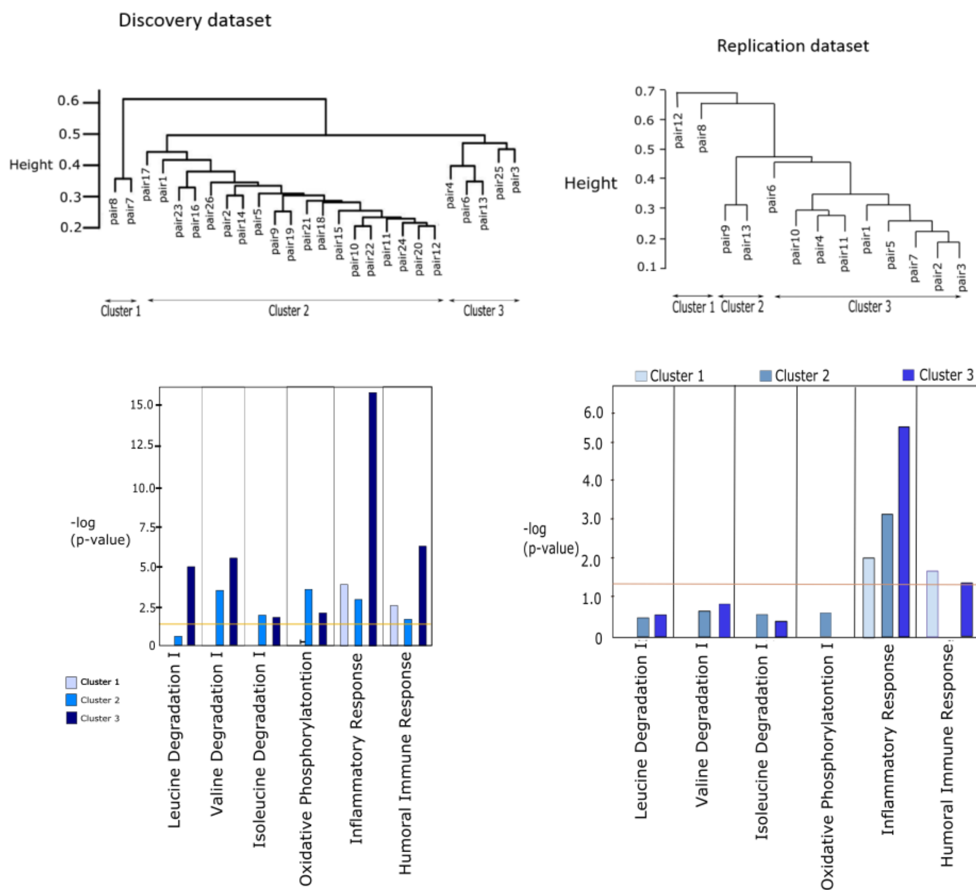

Figure 13: *Subgroups (clusters) formed in the discovery and replication dataset. Figure reprinted with the permission obtained from the International Journal of Obesity. Within-twin pair differentially expressed genes for each cluster in each dataset were entered into IPA® to produce the pathways. These pathways were compared with the IPA® Comparison Tool. The y-axis displays the -log of p-values, which is calculated by Fisher's exact right-tailed test. A -log (p-value) of 1.3 is indicative of a p-value of 0.05. IPA®, Ingenuity Pathway Analysis*

Next, the pathways enriched for differentially expressed genes within the 26 twin pairs were re-examined to determine if they were still enriched in each cluster. This was done to ascertain if any of the original pathways were driven by any specific cluster and also to have a baseline against which to compare all three clusters. IPA®'s comparison analysis revealed that compared to Cluster 1, the other two clusters had more enriched genes (differentially expressed within twin pairs) for these pathways (Figure 14). In Cluster 3, there were more genes (differentially expressed within twin pairs)

for Nur77 signalling in T lymphocytes (upregulation of inflammation) and Valine Degradation (downregulation of mitochondria) than in Cluster 1 or 2. Additionally, compared to Cluster 1 and 3, Cluster 2 had more genes (differentially expressed within twin pairs) belonging to the mitochondrial pathways. Compared to Cluster 1 and 2, Cluster 3 had more affected genes for Nur77 signalling in T lymphocytes and Valine Degradation pathways.

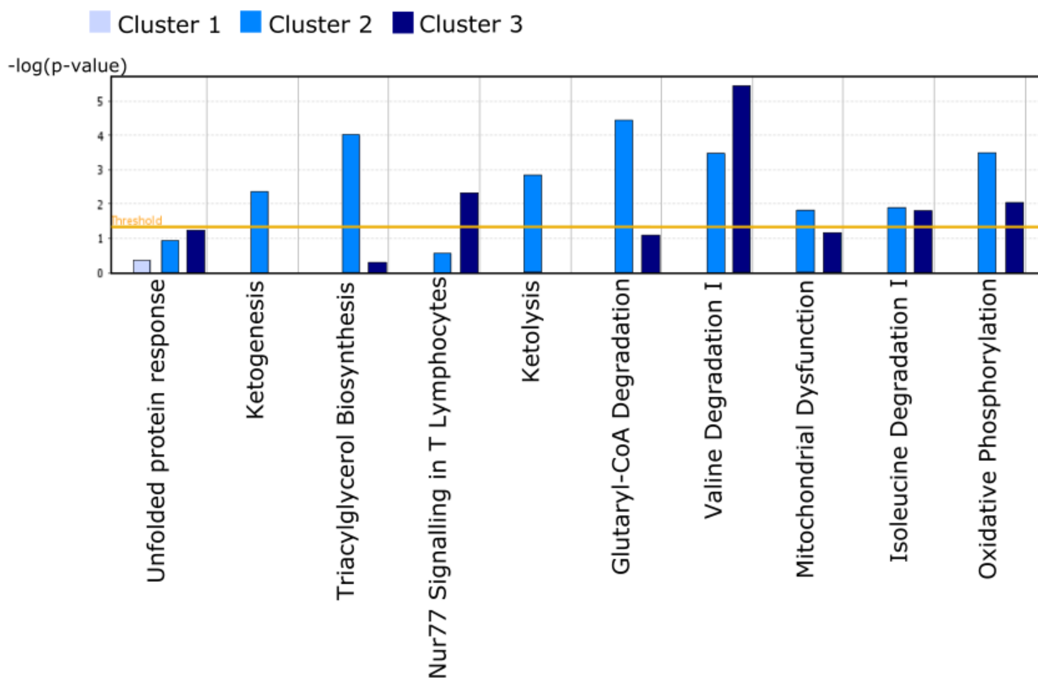

Figure 14: *Pathways of Clusters 1–3 compared for selected functions. Differentially expressed genes for each cluster (FDR p-value<0.05, n=274 for Cluster 1, n=728 for Cluster 2, n=828 for Cluster 3) were entered into IPA® to compare the selected pathways. The y-axis displays the -log of p-value which is calculated by Fisher's exact right-tailed test. A -log (p-value) of 1.3, marked by the yellow threshold line, is indicative of a p-value of 0.05.*

5.3.3. Associations between cluster pathways and phenotype measures, based on the adipose tissue gene expression profile (I)

Overall (Figure 15), Cluster 2 was characterised by mitochondrial dysfunction while Cluster 3 was characterised by mitochondrial dysfunction and inflammation in the heavy co-twins. In a comparison of all three clusters using adiposity measures and biochemistry measures, differences within-twin pairs were larger for Cluster 3 than in Cluster 2 for fasting insulin levels (adjusted p-value =0.041) and adipocyte diameter (adjusted p=0.045).

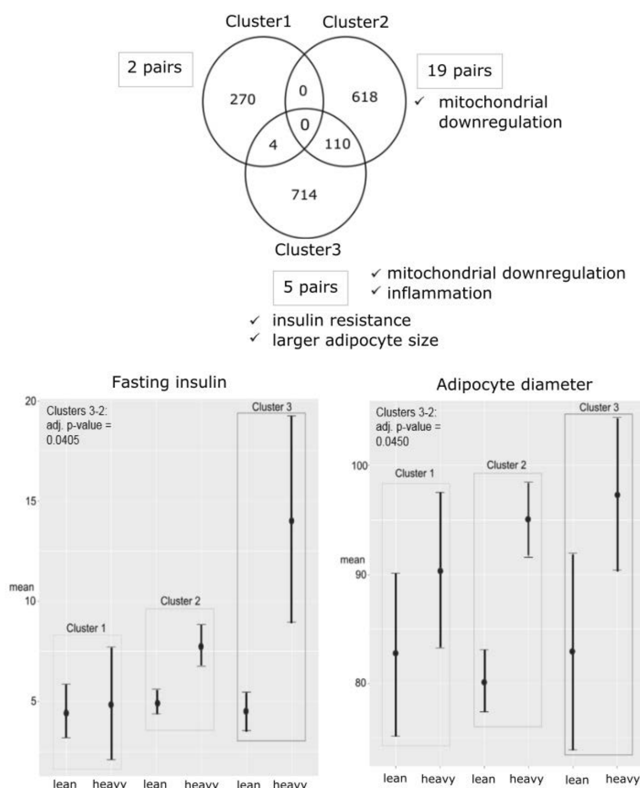

Figure 15: *Cluster differences in pathways and clinical measures. The 26 BMI-discordant twin pairs grouped into three clusters. The Venn diagram shows the number of genes differentially expressed within the twin pairs. Here, no differentially expressed genes are common to all three clusters. The genes differentially expressed for each cluster were analysed for the pathways. Cluster 1 (i.e., the "healthy" group (n=2 pairs)) showed slight differences within twin pairs in lipid metabolism and cell signalling pathways. Cluster 2 showed mitochondrial downregulation and Cluster 3 showed mitochondrial downregulation and higher inflammation within twin pairs. A comparison of clinical measures also showed there were significant differences within the twin pairs in Cluster 3 for fasting insulin and adipocyte diameters. Heavy refers to the heavier co-twin in a pair while lean refers to the leaner co-twin in a pair.*

5.4. Differential gene expression in adipocytes and subcutaneous adipose tissue (II)

5.4.1. Pathways for differentially expressed genes within the BMI-discordant twin pairs, in the adipocytes (II)

In order to investigate if the gene expression differences in the adipose tissue of the heavy co-twins (compared to the lean co-twins) were similar to the gene expression differences in the adipocytes of the heavier co-twins (compared to the lean co-twins), the adipocytes were studied in more detail. Of the 26 BMI-discordant twin pairs included in Study I, 14 pairs had isolated adipocytes available. Within-twin pair gene expression analysis was conducted for these 14 pairs; differentially expressed genes were analysed for pathway enrichment. 2538 genes were differentially expressed within the twin pairs (p-value<0.05).

The top pathways (p-value<0.001) enriched for genes differentially expressed in the adipocytes of the co-twins were:

1. Oxidative Phosphorylation
2. Valine Degradation I
3. Glucocorticoid Receptor Signalling
4. IL-8 Signalling
5. mTOR Signalling
6. Role of JAK2 in Hormone-Like Cytokine Signalling
7. Isoleucine Degradation I
8. Glutaryl-CoA Degradation
9. NRF2-mediated Oxidative Stress Response
10. Regulation of eIF4 and p70S6K Signalling

Downregulated pathways in adipocytes of the heavier co-twins compared to the leaner co-twins included OXPHOS, glutaryl-CoA degradation, mTOR signalling and BCAA catabolism (valine and isoleucine degradation), and the significantly upregulated pathways were glucocorticoid receptor and IL-8 signalling.

5.4.2. Pathways for within-twin pairs differentially expressed genes in subcutaneous adipose tissue and adipocytes (II)

Pearson correlations were calculated in order to explore if the top 10 most significant (*p*-value<0.01) pathways enriched for the genes differentially expressed within twin pairs in the SAT correlated with body composition, adipocyte volume, HOMA and CRP measures (Table 4). There were significant correlations between body composition measures, adipocyte volume, HOMA and CRP, with most of the top 10 pathways including OXPHOS, BCAA Degradation and Glutaryl-CoA Degradation pathways. Higher values of the adiposity measures (SAT, VAT, LF, adipocyte volume) as well as HOMA and CRP correlated with downregulation of the top 10 pathways. In other words, increased adiposity, higher IR and inflammation were related to downregulation of these ten pathways.

Table 4: *Pathways differentially enriched between heavy and lean subcutaneous adipose tissue.*

Pathways differentially enriched between heavy and lean adipose tissue *p*-value< 0.01 (n=38 individual twins)							
Pathways	Significant correlations between pathways to phenotype (adiposity and clinical measures) (**p*<0.025, ***p*<0.01 and ****p*<0.001)						
	SAT	VAT	LF	Adipocyte volume	Matsuda	HOMA	hs-CRP
Oxidative Phosphorylation	-0.49 *	-0.74 ***	-0.41 *	-0.55 **	0.48 **		-0.44 *
Valine Degradation I	-0.63 ***	-0.77 ***	-0.61 ***	-0.74 **	0.63 ***	-0.56 **	-0.41 *
Triacylglycerol Biosynthesis		-0.59 ***	-0.52 **				
Glutaryl-CoA Degradation	-0.68 **	-0.77 ***	-0.56 ***	-0.74 ***	0.63 ***	-0.49 **	-0.39 *
Ethanol Degradation II	-0.63 ***	-0.79 ***	-0.55 ***	-0.65 ***	0.53 ***	-0.47 ***	
Ketogenesis	-0.66 ***	-0.73 ***	-0.58 ***	-0.75 ***	0.58 **	-0.47 **	
Ketolysis	-0.65 ***	-0.68 ***	-0.57 ***	-0.73 ***	0.58 **	-0.48 **	
Lysine Degradation II	-0.51 **	-0.72 ***	-0.51 **	-0.6 ***	0.60 ***	-0.56 **	-0.52 **
Acetate Conversion to Acetyl-CoA	-0.56 **	-0.64 **	-0.49 **	-0.54 *	0.44 *	-0.35 *	
Fatty Acid β-oxidation I	-0.62 ***	-0.76 ***	-0.54 ***	-0.69 ***	0.56 ***	-0.44 **	-0.35 *

Pathways were identified in the within-twin pair analysis. Pearson correlations of these pathways with adiposity and clinical measures were calculated in individuals. Table contains summarised information from

Supplementary Table 3 of Article II. Pathways produced using the IPA® tool. HOMA, homeostatic model assessment; hs-CRP, high-sensitivity C-reactive protein; LF, liver fat; SAT, subcutaneous adipose tissue; VAT, visceral adipose tissue

Pearson correlations were also calculated, this time for the adipocyte data, in order to explore if the top 10 (according to p-value) pathways enriched for the genes differentially expressed within twin pairs in adipocytes correlated with body composition, adipocyte volume, HOMA and CRP measures (Table 5). There were significant correlations between body composition measures, adipocyte volume, HOMA with OXPHOS, BCAA degradation and glutaryl-CoA degradation pathways.

Table 5: *Correlations between pathways differentially expressed within twin pairs (for adipocyte data) with phenotype.*

	Pathways differentially enriched between heavy and lean adipocytes *p*-value< 0.01						
Pathways	Significant correlations between pathways to phenotype (adiposity and clinical measures) (n=38 individual twins) (*$p<0.025$, **$p<0.01$ and ***$p<0.001$)						
	SAT	VAT	LF	Adipocyte volume	Matsuda	HOMA	hs-CRP
Oxidative Phosphorylation		-0.70 ***	-0.55 **		0.61**		
Valine Degradation I	-0.48 *	-0.66 ***	-0.61 ***		0.68***	-0.54**	-0.46 *
Glucocorticoid Receptor Signalling		0.67 ***	0.58 **		-0.61**		0.49 *
IL-8 Signalling	0.61 **	0.84 ***	0.64 ***	0.59 *	-0.70***	0.54 **	-0.57 *
mTOR Signalling	-0.42 **	-0.46 **	-0.41 *	-0.51 **	0.39*		
Role of JAK2 in Hormone-Like Cytokine Signalling							
Isoleucine Degradation I		-0.62 **	-0.62 **		0.65***	-0.52 **	
Glutaryl-CoA Degradation	-0.54 **	-0.72 ***	-0.65 ***	-0.66 **	0.70***	-0.55 **	
NRF2-mediated Oxidative Stress Response							
Regulation of eIF4 and p70S6K Signalling							

Pathways were identified in the within-twin pair analysis. Pearson correlations of these pathways with adiposity and clinical measures were calculated in individuals. Table contains summarised information from Table 5 of Article II. Pathways produced using the Ingenuity Pathway analysis tool. HOMA, homeostatic model assessment; hs-CRP, high-sensitivity C-reactive protein; LF, liver fat; SAT, subcutaneous adipose tissue; VAT, visceral adipose tissue

5.5. Plasma Metabolites (III)

In Study I and II, within-twin pair differences in the gene expression of the heavy co-twin in comparison to the lean co-twin were analysed. These studies utilised linear association analysis to reveal how these gene expression differences associated with the differences in within-twin pair BMI values. In Study III, the aim was to study the relationships between metabolites and phenotypes (adiposity and blood biochemistry measures) in all available twin pairs, disregarding their within-twin pair discordance or concordance for BMI. Study III also investigated the effect sizes of these metabolite–phenotype relationships (i.e., how much change there is in metabolite concentration values for every one unit of change in the phenotype measure).

5.5.1. Significant metabolite-phenotype associations (III)

In order to determine associations between metabolites and different phenotypes (adiposity and blood biochemistry measures), linear regression analysis was conducted. Overall, there were 25 metabolite–phenotype associations (Bonferroni corrected p-value<0.05) involving 12 metabolites and 7 phenotypes (Table 6, Table 7). HDL-C, followed by BMI, associated with the most number of metabolites in the plasma.

Metabolites associating with adiposity measures were mostly AAs followed by acylcarnitines. Metabolites associating with %body fat also associated with BMI; however, associations with BMI were stronger. SAT, VAT, and LF associated with metabolites that also associated with BMI. Amongst the blood biochemistry measures, only TG and HDL-C associated with one or more of the metabolites. HDL-C associated with the most number of metabolites; some of these metabolites also associated with other phenotypes (both adiposity and blood biochemistry measures), but the strength of the association with HDL-C was always higher. HDL-C also associated with the most number of metabolites not associated with any other measure.

Table 6: *Metabolites associating with adiposity measures in individuals and within-twin pairs.*

	Metabolites significantly associating with phenotype in individual twins	Metabolites with concentrations that are significantly different within twin pairs	Metabolite Category
BMI			
1	Valine	Valine	Amino acids and derivatives
2	Tyrosine	Tyrosine	Amino acids and derivatives
3	Aspartate	Aspartate	Amino acids and derivatives

	Metabolites significantly associating with phenotype in individual twins	Metabolites with concentrations that are significantly different within twin pairs	Metabolite Category
4	Cysteine	-	Amino acids and derivatives
5	Propionylcarnitine	Propionylcarnitine	Acylcarnitines
6	Deoxycytidine	Deoxycytidine	Nucleosides
%bodyfat			
1	Aspartate	Aspartate	Amino acids and derivatives
2	Propionylcarnitine	Propionylcarnitine	Acylcarnitines
SAT			
1	Aspartate	Aspartate	Amino acids and derivatives
2	Propionylcarnitine	Propionylcarnitine	Acylcarnitines
3	Deoxycytidine	Deoxycytidine	Nucleosides
VAT			
1	Aspartate	Aspartate	Amino acids and derivatives
2	Propionylcarnitine	Propionylcarnitine	Acylcarnitines
LF			
1	Aspartate	Aspartate	Amino acids and derivatives

Significant associations with metabolites were found for all adiposity measures. Out of all significant metabolite–phenotype associations, except for the cysteine–BMI association, all other metabolites associated with the phenotypes in the table, independent of genetic factors. BMI, body mass index; LF, liver fat; SAT, subcutaneous adipose tissue, VAT, visceral adipose tissue

Table 7: *Metabolites associating with blood biochemistry measures in individuals and within twin pairs.*

	Metabolites significantly associating with phenotype in individual twins	Metabolites with levels that are significantly different within- twin pairs	
HDL-C			
1	Tyrosine	Tyrosine	Amino acids and derivatives
2	Alanine	Alanine	Amino acids and derivatives
3	Citrulline	Citrulline	Amino acids and derivatives

	Metabolites significantly associating with phenotype in individual twins	Metabolites with levels that are significantly different within- twin pairs	
4	Phenylalanine	Phenylalanine	Amino acids and derivatives
5	L-Kynurenine	L-Kynurenine	Amino acids and derivatives
6	Propionylcarnitine	Propionylcarnitine	Acylcarnitines
7	Kynurenic Acid	-	Urea cycle intermediates
8	Deoxycytidine	-	Nucleosides
9	Phosphoethanolamine	-	Ethanolamines
TG			
1	Aspartate	Aspartate	Amino acids and derivatives
2	S-Adenosyl-L-Homocysteine	-	Amino acids and derivatives

Out of all significant metabolite–(HDL-C) associations, except for kynurenic acid, deoxycytidine and phosphoethanolamine, all other metabolites correlated with HDL, independent of genetic factors. For all significant metabolite–TG associations, except for S-Adenosyl-L-Homocysteine*, all other metabolites correlated with TG, independent of genetic factors. No significant associations were found for HOMA, Total Cholesterol, LDL-C and CRP. HDL-C, high-density lipoprotein cholesterol; TG, triglycerides*

5.5.2. Metabolite–phenotype associations confounding from shared factors in twins (III)

To ascertain if any of the metabolite-phenotype relationships already established in the linear regression were independent of shared genetic and familial factors, within-twin pair analysis was carried out. For significant within-twin pair differences, there were differences in within-twin pair values of phenotype measures that significantly associated with metabolite concentration differences within-twin pairs. The results for the within-pair analyses are summarised in Table 6 and Table 7. Out of all significant metabolite–BMI associations in the individual twins, all metabolites except for cysteine remained associated with BMI within the pairs (i.e., independent of genetic factors). Because MZ co-twins are genetically identical, significant within-pair differences in metabolite levels would rule out shared genetics or environment and point to environmental factors for which the twin pair is discordant. Amongst the blood biochemistry measures, six out of nine associations with HDL-C and one out of two associations with TG remained free of confounding from genetic and shared environmental factors.

5.5.3. Shared and unique associations for metabolites with different phenotypes (III)

The associations identified were compared across the phenotypes to identify shared and unique associations. Four metabolites (tyrosine, propionylcarnitine, deoxycytidine and aspartate) were significantly associated with two or more phenotypes (Figure 16a). BMI, HDL-C and TG showed associations with metabolites not associating with any other phenotype (Figure 16b).

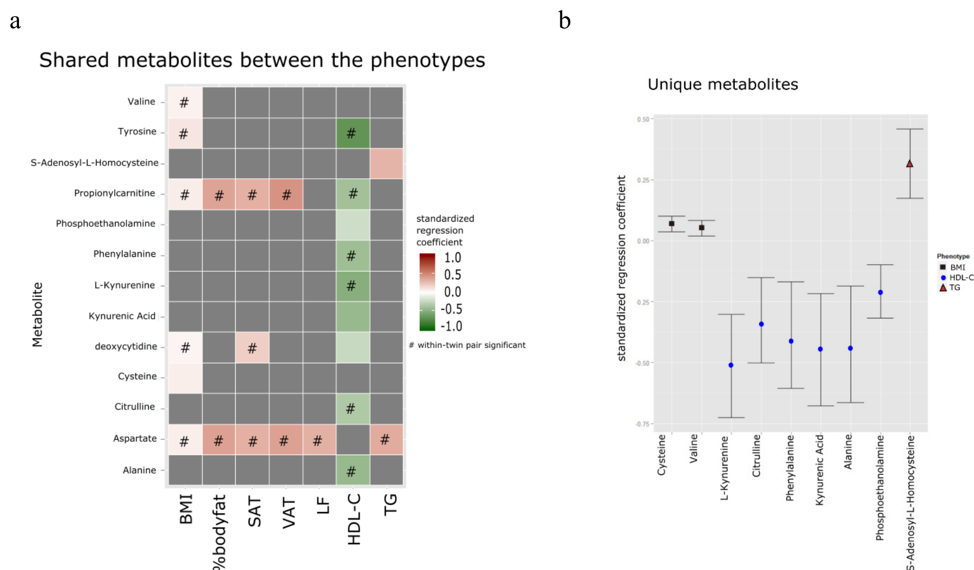

Figure 16: *Metabolite-phenotype associations. a) Metabolites associating with BMI, adiposity and blood biochemistry measures. Metabolites that associate with one or more phenotype are shown in the heatmap. Only significant Bonferroni-corrected significant associations p-value<0.05) are shown. If these associations are free from genetic confounding, the cell is marked with a hash (#) symbol. All effect sizes of the metabolite–phenotype associations have been transformed to be on the same scale (values between -1 and 1) and are comparable. b) BMI, HDL and TG showed associations with metabolites that did not associate with any other phenotype. Beta coefficients show the effect size/strength of the associations. BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; LF, liver fat; SAT, subcutaneous adipose tissue; TG, triglycerides; VAT, visceral adipose tissue. Images reproduced from manuscript #3*

5.5.4. Effect size comparisons in metabolite-phenotype associations (III)

For metabolites associating with HDL-C that also associated with another phenotype, the effect size of the metabolite-(HDL-C) relationship (compared on the same scale with other phenotype) was always the highest, except for propionylcarnitine (Table 8).

Table 8: *Comparison of effect sizes (compared on the same scale i.e., standardised beta coefficients) of metabolites which associate with HDL-C that also associate with other metabolites*

Phenotype	BMI	%bodyfat	SAT	VAT	LF	HOMA	TCHOL	LDL-C	HDL-C	TG	CRP
Alanine	0.05				0.31				-0.44	0.29	
Citrulline	0.04							0.19	-0.34		
deoxycytidine	0.04	0.19	0.20	0.22	0.13	0.14			-0.23	0.14	
Kynurenic Acid	0.05		0.24	0.33					-0.44		
L-Kynurenine	0.06		0.26	0.33	0.23			0.32	-0.51		
Phenylalanine	0.04							0.23	-0.41		
Phosphoethanolamine	0.03		0.11	0.15					-0.21		
Propionylcarnitine	0.07	0.38	0.33	0.46	0.23	0.30		0.22	-0.39	0.24	0.28
Tyrosine	0.10		0.39	0.44	0.35				-0.68		0.34

For all metabolites that associate with HDL-C, if these metabolites also associate with other phenotype, the effect size for the metabolite–(HDL-C) relationship is always larger (absolute values considered in this comparison) except for Propionylcarnitine. The actual effect sizes for each association were transformed to be on the same scale to enable comparison across the phenotypes. TCHOL did not associate with any metabolites that also associated with HDL-C; the TCHOL column is hence empty in the table. BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; HOMA, homeostatic model assessment; hs-CRP, high-sensitivity C-reactive protein; LDL-C, low-density lipoprotein cholesterol; LF, liver fat; SAT, subcutaneous adipose tissue; TCHOL, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue

5.5.5. Plasma metabolite profiling identified two groups (III)

To address the high multicollinearity in the metabolite dataset, PCs of the metabolites were used to capture the variance across all of the metabolites. It was found that 20 PCAs (Figure 17) were enough to characterise 80% of the variance in the data.

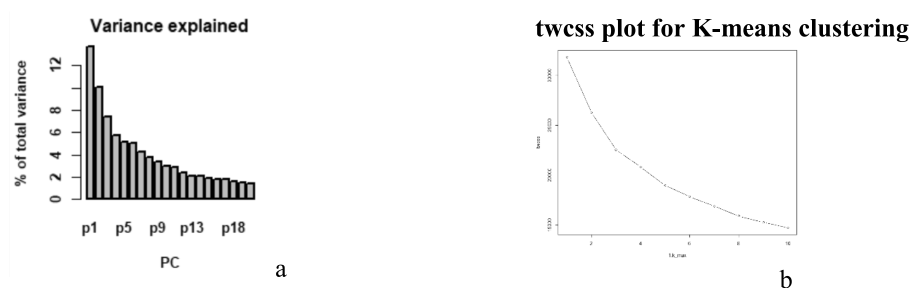

Figure 17: *a) PCA analysis of the metabolite data and b) total within sum of squares (twcss) plot pointing to an optimal n=2 clusters.*

These 20 PCs were then used to cluster the metabolite data. Two subgroups (clusters) were revealed; Cluster1 had 22 individuals and Cluster 2 had 58 individuals.

Moderated t-tests were run on these two clusters to investigate if there were any differences in metabolite concentrations between the clusters. All 32 metabolites (Table 9) had higher levels (*p*-value<0.05) in Cluster 2; these included several acylcarnitines and AAs including BCAAs. Cluster 2 also displayed significantly (*p*-value<0.05) higher values for total cholesterol and LDL-C compared to Cluster 1.

Table 9: *Differences in metabolite concentrations between Cluster 1 and Cluster 2.*

Metabolite	Classification	Cluster 2 vs Cluster 1	
		log fold change	p-value
Hexanoylcarnitine	Acylcarnitines	1.09	0.000090
Octanoylcarnitine	Acylcarnitines	0.89	0.0019
Valine	Amino acids and derivatives	0.68	0.0035
Myristoyl Carnitine	Acylcarnitines	0.69	0.0059
Dimethyl Glycine	Amino acids and derivatives	0.75	0.0060
Dodecanoyl Carnitine	Acylcarnitines	0.83	0.0061
Uracil	Nucleobases	0.83	0.0071

Cytidine	Nucleosides	0.86	0.0077
Arginine	Amino acids and derivatives	0.82	0.010
Acetylcarnitine	Acylcarnitines	0.88	0.011
Histidine	Amino acids and derivatives	0.67	0.011
Isovaleryl Carnitine	Acylcarnitines	0.61	0.011
2-deoxyuridine	Nucleosides	0.74	0.011
L-Kynurenine	Amino acids and derivatives	0.75	0.012
Propionylcarnitine	Acylcarnitines	0.68	0.012
Carnitine	Acylcarnitines	0.64	0.012
Xanthine	Nucleobases	0.71	0.014
Isoleucine	Amino acids and derivatives	0.56	0.016
Leucine	Amino acids and derivatives	0.56	0.017
Cysteine	Amino acids and derivatives	0.58	0.020
Lysine	Amino acids and derivatives	0.86	0.024
Stearoyl Carnitine	Acylcarnitines	0.69	0.025
Taurocholic Acid	Bile Acids	0.70	0.025
Decanoylcarnitine	Acylcarnitines	0.66	0.026
L-Methionine	Amino acids and derivatives	0.66	0.026
Glyceraldehyde	Central Carbon Metabolites	0.51	0.027
Phenylalanine	Amino acids and derivatives	0.63	0.029
Arachidyl Carnitine	Acylcarnitines	0.65	0.040
Deoxycytidine	Nucleosides	0.44	0.042
2-Aminoisobutyric acid	Amino acids and derivatives	0.60	0.045
Asymmetric Dimethylarginine	Amino acids and derivatives	0.46	0.048
Guanosine	Nucleosides	0.59	0.049

log fold change is the log (ratio of the metabolite quantity of Cluster 2 compared to the metabolite quantity of Cluster 1). log fold change>0 indicates that Cluster 2 has a higher quantity of the metabolite than Cluster 1.

6. DISCUSSION

6.1. Methodological considerations

Because there is a plethora of bioinformatics methods and tools available, it is challenging to select the most suitable method or tool to address the particular research question at hand. In addition to the unique dataset used in this thesis, the lack of consensus in data-pre-processing, dimension reduction and statistical analysis in other obesity-related studies make comparisons across studies particularly challenging. The tools employed in this thesis were selected based on the data and study questions employed and were largely guided by the commonly used practices in the field. The rationale for choosing the methods and tools in this thesis are explained under the relevant sub-sections in this chapter.

To take advantage of the unique dataset available, this thesis employed a powerful within-twin pair design in order to investigate the obesity phenotype related to the non-shared environmental factors between the co-twins. It is important to bear in mind that while associations may be apparent, causality cannot be established in these types of studies. MZ twins that are discordant for phenotypes/diseases being investigated provide a unique insight into diseases, independent of genetic background. Additionally, since each twin of a pair acts as the control for the other twin of the pair, they are the perfect case–control study example in humans. However, it is erroneous to use both twins of each pair in statistical models that assume sample independence (as in linear regression) without making adjustments for the possibility that outcome values from co-twins might well be more similar than values from two unrelated individuals (290, 291). Using related samples in these models may result in inaccurate standard errors and invalid confidence intervals and p-values (290).

The main analysis across the three studies focused on finding associations between BMI with gene expression data and phenotypes of interest with metabolite data. Within-twin pair and linear mixed model analyses were conducted to study these associations. Additionally, variations in gene expression and metabolite data were identified in order to see if subgroups formed based on the genome-wide profiles. Clustering the data was an important step in achieving this. The considerations in the selection for these methods are discussed below.

6.1.1. Transcriptomics data analysis

Transcriptomics data analysis is now a mature field of study. In this thesis, Affymetrix™ HGU133plus2 was chosen because its human genome chipset is very comprehensive (40,000 transcripts) and has previously been used successfully in the lab where this thesis was carried out. However, more sophisticated technology does exist and with the falling prices of sequencing technologies, RNA-seq for expression analysis seems promising. With this technology it would have

been possible to quantitatively detect all transcripts (including isoforms), even those that do not correspond to known existing genomic sequences (292). Compared to microarray technology, RNA-seq technology also provides precise measurements of transcript levels at a wider range of expression levels than is possible with microarray technology (292). Hence, because of the range of coverage in both transcripts and transcript levels, it is possible that using RNA-seq technology in this thesis might have increased the possibility of finding more true positives values for differentially expressed genes. However, the results obtained from the microarray analyses in this thesis are still valid. In a comparison of Illumina™ RNA-seq and Affymetrix™ HGU133plus2 for differential expression analysis, 87% of genes identified as differentially expressed in the microarray experiment were also identified in the RNA-seq experiment, with correlating log-fold changes found in both experiments (293). Additionally, the microarray technology used in this thesis was still the most cost-effective option at the time the experiments were conducted.

With transcriptomics data analysis, there is an abundance of tools to analyse the different types of data and a strong community of bioinformaticians and biostatisticians able to provide guidance based on prior experiences. However, no consensus has been reached on the appropriate pre-processing methods to be used. One issue is the transcript/gene annotation. Outdated annotations (286), the gene overlap between annotations, the incompleteness of annotation content, the strong connectivity/dependency among genes are amongst several problems (294). Around 10 to 40% of the original probe IDs on vendor-defined annotation packages for older GeneChips no longer match or have been retired from current annotation databases, some probes are redundant or non-specific (one probe hybridising to more than one gene transcript or to a non-coding region) (286). In this thesis, to circumvent the issue of outdated annotations, annotations were done using the Brainarray CDFs (286), which is updated on a yearly basis.

In Study I, differential analysis was conducted to determine if the three clusters identified differed in within-twin pair gene expression. Differential analyses using the limma package in R generally performs well, is robust and easy to use while supporting complex experimental designs (224, 295) and hence was chosen for differential analyses in this thesis. While some transcriptomics studies use Significance Analysis Methods (SAM), ANOVA and t-tests to analyse gene expression data, limma provides better control of false-positive rates and has increased power in analysing small sample sets (296) by borrowing information on gene expression variations across samples to deal with small sample sets (224).

6.1.2. Regression analysis with metabolomics data

Several metabolite studies on BMI and obesity have used PCA-derived factors derived from metabolite data as dependent variables in subsequent regression analysis (266, 297, 298), while others have used the individual metabolites as dependent variables (90, 299). While these factors may form biologically relevant groups or pathways, using them as dependent variables may produce results that may be difficult to envision and compare across studies. For example, a previous metabolite-related regression analysis has identified statistically different factors that are general and not specific, for example, factors comprising 'various metabolites' or 'non-esterified FAs and other AAs' (266). One study identified a factor comprising all AAs except for two as associating with several adiposity markers (297). In order to individually examine each metabolite and not groupings of metabolites, this thesis employed individual metabolites as dependent variables. Additionally, the aim of the metabolite study in this thesis was to determine the effect size of each metabolite individually and to subsequently compare them. Hence, PCA factors were not used in the regression analysis.

6.1.3. Linear regression and within-twin pair differential analysis

Study I and II utilised within-twin pair differential analysis while Study III employed regression analysis using linear mixed modelling. These two types of analyses have different objectives. The former (Study I and II) compares the differences within the twin pairs in order to derive the magnitude of the difference in gene expression for each unit difference in phenotype within the twin pairs. This study was conducted to identify the gene expression differences in acquired obesity and point to environmental effects not common to both twins in a pair (e.g. aspects of diet, exercise and lifestyle) as a basis to explain individual differences within MZ twin pairs. The latter type of analysis investigated the relationship between metabolites and phenotypes by modelling all variations in the metabolites against variations in phenotype within the whole study sample. The end result were effect sizes that showed the variation in metabolite concentration for each unit variation in phenotype value. In other words, the within-twin pair differential analysis (Study I and II) was concerned with the comparative gene expression differences between co-twins in a pair. The linear regression analysis (Study III) looks at the absolute change in metabolite levels associating with a change in phenotype.

In Study III, linear mixed modelling was applied metabolite-wise, using a separate linear model for each metabolite. Because information was not shared between metabolites as in the within-twin pair differential analysis using limma that was employed in both Study I and Study II, this analysis suffers from low degrees of freedom and low statistical power.

6.1.4. PCA before clustering in metabolomics

In this thesis, PCA was first applied on the metabolite data to extract enough PCs to model 80% of the variance in the data before K-means clustering was applied. Performing clustering with these PCs

prevented additional emphasis on highly collinear metabolites that may represent the same biological information because these highly collinear metabolites were collapsed into new, independent variables. This dimension reduction method has been employed in other metabolomics studies (58, 266, 298, 300, 301) with the resulting components used in regression (266, 298, 302) or differential analyses (58) or to identify subgroups in the data (301). The difficulties with using PCA-derived components in regression analysis is that it is not possible to directly obtain the effect size of the metabolites from the regression model. Hence, for the regression analysis in Study III, the individual metabolites and not the PCA-derived factors were used.

6.1.5. Clustering methods

In Study I, hierarchical clustering was performed to reveal if sub-types in obesity can be identified based on transcriptome profiles in adipose tissue. Hierarchical clustering has been used extensively to identify subgroups in diseases like cancer (303-305), and also in one study of subtyping obesity (306). Hierarchical clustering is widely used on gene expression data (307) and because of the ease in visually examining the results of hierarchical clustering (308), this thesis used the aforementioned method. From a biological point of view, it is very hard to choose the best cluster solution if the clinical phenotype has not been characterised completely based on prior knowledge. In this thesis, the biological findings in the obtained clusters were validated using a replication dataset. However, the results of this thesis would have benefited from statistical validation of the clusters, for example permutation testing (309). This is one limitation of this thesis.

In Study III, K-means clustering was carried out using the metabolite data. This method has been used in several metabolomics studies (310-312) and was used in this thesis because of its simplicity and ease of implementation. One challenge was to identify the actual number of clusters and cluster membership. Here, the number of clusters were identified with a total within sum of squares plot. Cluster membership was not validated; this is one area where the thesis could have been improved on.

While Study I clustered the samples based on the variations in transcriptomics profiles and Study III clustered the samples based on the variations in the metabolite profiles, the clusters themselves are not comparable. Although, 26 twins pairs from Study I are also in Study III, the findings point to different molecular effects. In the transcriptomics study, the ratios (heavy/lean) of the gene expression values were used with one value per gene per twin pair, thereby resulting in each pair being assigned to the same cluster. This clustering was done to capture the acquired obesity component, leaving out the genetic and shared environment components which are the same for the co-twins. In the metabolomics study, the individual metabolite concentration values were used (i.e. one value per metabolite per individual). This clustering was done to capture the both the genetic and acquired components of the phenotype. As a result, the differences within the twin pairs were used in the

clustering in Study I while the actual values for the individuals were used in the clustering in Study III.

6.2. Summary of the main findings

A summary of this thesis's main findings is presented in Figure 18. In brief, the transcriptomics studies showed downregulation of mitochondria-related pathways (OXPHOS, Glutaryl-CoA degradation, AAA degradation, BCAA degradation, FA β Degradation) and upregulation of inflammation pathways in the obese SAT (Study I, II). Consistent with these pathway findings in the SAT, the metabolomics study on plasma (Study III) shows higher circulating BCAAs, AAAs and acylcarnitines in obesity. Transcriptomics and metabolomics profiles were used to identify subgroups in the study participants. In both transcriptomics and metabolomics profiles, associations were found between molecular-level subgroups and the adiposity and blood biochemistry measures.

In Study I, there was a gradual worsening of the obese SAT from Cluster 1 through Cluster 2 and finally to Cluster 3. However, this was not a longitudinal study and these subgroups are not to be taken as different stages in the development of obesity. Instead, they are three subgroups of obesity profiles with different levels of adipose tissue worsening. Pathways that differed within twin pairs in Cluster 1, were enriched for lipid metabolism and signalling pathways. Cluster 2 showed worsening of mitochondrial function in the heavy co-twins. Both OXPHOS and valine degradation pathways were downregulation in the heavy co-twins of this cluster. Cluster 3 showed significant upregulation of the inflammation-related pathways and downregulation of valine degradation in the heavy co-twins. This cluster was considered the unhealthiest subgroup. In this unhealthy obese cluster that was identified from the transcriptomics profiles, there was also evidence of adipocyte diameter increase and higher fasting insulin in the heavy co-twins.

In Study II, mitochondrial pathways including OXPHOS, BCAA Degradation and Glutaryl-CoA Degradation were found downregulated in the heavy co-twins in both adipocytes and adipose tissue. There were significant correlations between body composition measures, adipocyte volume, HOMA and CRP with these pathways.

In Study III, two subgroups (i.e., Cluster 1 and Cluster 2) of individuals were identified based on their metabolite profiles. Cluster 2 showed an unhealthier metabolite profile compared to Cluster 1. In this unhealthy cluster, 32 metabolites showed higher concentration levels in comparison to Cluster 1. This unhealthy cluster also had higher LDL-C and total cholesterol in comparison to Cluster 1.

While some of the discussion on the clusters in both Study I and Study III are in the context of other studies which have studied the MHO and MUO phenotype, it should be noted that this thesis does not claim that Cluster 2 from Study I is the MHO phenotype or that Cluster 3 is the MUO phenotype. The twins in this thesis were all healthy and these clusters were formed using genome-wide gene

expression data, and not clinical classifications of the MHO and MUO phenotypes. Rather, it is recognised that of these two groups, one is healthier than the other, with one cluster appearing to be more metabolically healthy than the other. For the remainder of this thesis, for ease of reading, the healthier cluster shall be referred to as 'healthy cluster', while the less healthy cluster shall be referred to as 'unhealthy cluster'.

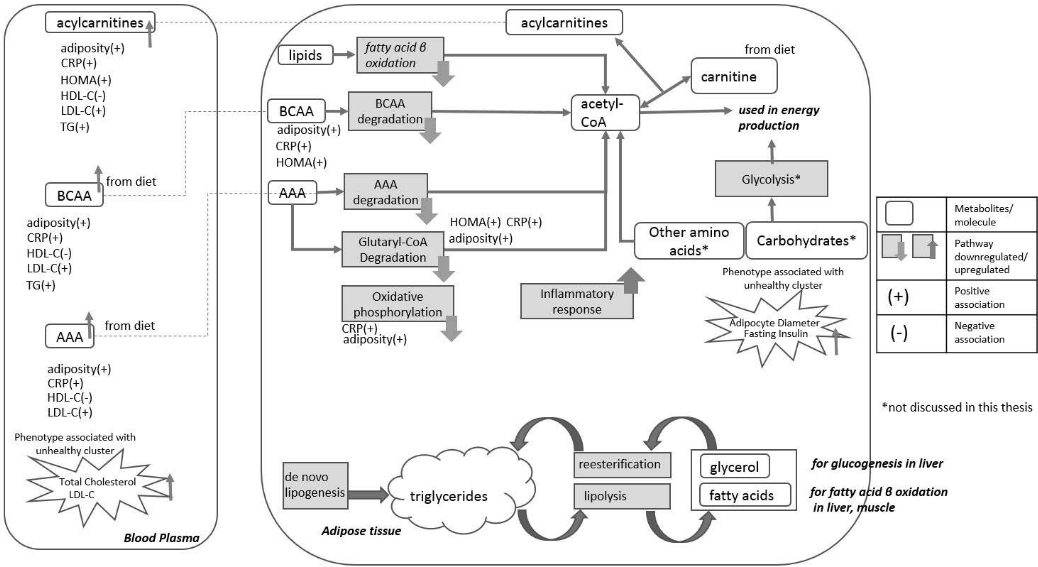

Figure 18: *Molecular changes in obesity and their associations to phenotypes. Excess nutrients are converted from fatty acids and stored as triglycerides through lipogenesis. In periods of nutrient deficiency, these triglycerides are converted to fatty acids and glycerol and used for energy production. In conditions of nutrient overload, more acetyl-CoA is converted to fatty acids and stored as triglycerides. However, downregulation of several pathways in the obese subcutaneous adipose tissue results in increased circulating amounts of BCAA, AAA and acylcarnitines in the plasma. Inflammatory response is also upregulated in obesity. Several of these pathways show significant associations with adiposity measures. Red and green arrows denote the findings from this thesis. AAA, aromatic amino acid; Acetyl-CoA, Acetyl coenzyme A; BCAA, branched chain amino acid; CRP, C-reactive protein; HDL-C, high-density lipoprotein cholesterol; HOMA, homeostatic model assessment; LDL-C, low-density lipoprotein cholesterol; LF, liver fat; SAT, subcutaneous adipose tissue; TCHOL, total cholesterol; TG, triglycerides; VAT, visceral adipose tissue*

6.3. Adipocyte mitochondrial functions consistently implicated in obesity (Study I, II, III)

In this thesis, molecular-level effects to the mitochondria-related pathways in both SAT and plasma in acquired obesity were shown. Associations of these pathways to adiposity (BMI, SAT, VAT, LF, adipocyte volume) (Study II) as well as insulin and cholesterol measures (Study I, III) were found. Pathways associated with differentially expressed genes in the heavy co-twins in both adipose tissue (Study I) and adipocytes (Study II) were similar; mitochondrial pathways were downregulated in the heavy co-twins.

Downregulation of mitochondrial function, oxidative phosphorylation and BCAA catabolism consistently found in the heavy co-twins in both Study I and Study II all pointed to adipocytes as the major contributor to the reduced mitochondrial oxidative metabolism of adipose tissue in acquired obesity. The top pathways differentially expressed in the heavy co-twins compared to the lean co-twins, namely Oxidative Phosphorylation, Glutaryl-CoA Degradation, Mitochondrial Dysfunction, BCAA Degradation, FA β Degradation, all involved the adipocyte mitochondria. The mitochondria have been recognised as the link between nutrient metabolism and oxidative respiration (156). In obesity, mitochondrial respiration is reduced in the adipocytes (313, 314). Mitochondrial dysfunction has been suggested to be due in part to this impairment in oxidative phosphorylation in addition to impairment in mitochondrial biogenesis and ATP production (156).

Results from Study I clearly show a downregulation in both oxidative phosphorylation and mitochondrial function in the heavy co-twins' SAT. The Oxidative Phosphorylation pathway encodes proteins in the five respiratory chain complexes of the mitochondrial matrix (315). The Mitochondrial Dysfunction pathway includes the same significantly downregulated genes as the Oxidative Phosphorylation pathway as well as, upregulated genes *GSR* and *CYB5R3*, which are involved in mitochondrial oxidative stress reactions. While this pathway has been termed as the mitochondrial 'dysfunction' pathway in the IPA® tool, the downregulation of this pathway effectively means reduced mitochondrial function. Adipocyte mitochondrial respiration is reduced in obesity (313, 314). Lower oxidative phosphorylation in obese people impacts normal mitochondrial function and impairs the ability of the mitochondria to function effectively (314). The majority of the genes in the BCAA pathways, Valine Degradation I and Isoleucine Degradation I also code genes that function in the mitochondria and were found to be downregulated in the heavy co-twins.

Few studies have examined the gene expression profiles of isolated adipocytes (255, 316). A comparison study of adipocytes and stromal vascular cells in lean and obese people revealed signalling-related pathways enriched in adipocytes of obese subjects (316), while another study comparing SAT adipocytes in obese and non-obese subjects found upregulation of inflammation and

immune response genes in obesity (255). Similarly, in this thesis, upregulation of inflammation was found in the heavy co-twins in both Study I and Study II. Additionally, this thesis revealed the downregulation of mitochondrial function, oxidative phosphorylation and BCAA catabolism pathways in obese adipocytes – findings that are new and important. On top of that, this thesis has highlighted that the disturbances to the adipocyte mitochondria-related pathways occur even in young, healthy adults who have increased BMI. It should be remembered that the work carried out in this thesis did not include measuring the mitochondrial pathways of stromal vascular fraction cells. Hence, no conclusion can be made regarding the differences in mitochondria-related transcriptional downregulation in adipocytes compared to the stromal vascular fraction cells of the adipose tissue. What can be concluded is that gene expression differences in acquired obesity are similar in adipocytes and adipose tissue of young MZ BMI-discordant co-twins, making adipocytes the major contributors to the downregulation of mitochondrial pathways in obese adipose tissue.

6.4. Branched chain amino acid consistently associated with obesity (Study I, II, III)

In all three studies, there was evidence of dysfunction in BCAA metabolism in obesity. Downregulation of the BCAA catabolism pathway in the SAT correlated with higher adiposity, CRP and HOMA (Study I, II), while increased plasma BCAA concentrations associated with higher BMI (Study III). In unhealthy obesity, in the plasma, higher levels of circulating BCAA correlated with LDL-C and total cholesterol, while in the SAT downregulation of BCAA degradation correlated with fasting insulin and adipocyte diameter measures. These associations also remained after controlling for both shared genetic and shared environmental factors, pointing to disturbances in BCAA catabolism as a characteristic of acquired obesity. Indeed, circulating concentrations of BCAA (valine, leucine, isoleucine) are often increased in obese, insulin-resistant states and in T2DM (58).

In this thesis, the downregulation of BCAA degradation was evident both by SAT and adipocyte gene expression, and plasma metabolite screening, confirming that the effect to reduced BCAA degradation in the adipose tissue is further reflected in the BCAA levels in circulating plasma. While Study I and III were different in that Study I investigated 26 BMI-discordant twin pairs and Study III investigated 40 twin pairs of which 26 were BMI-discordant and 14 BMI-concordant, the results from both of the studies can be integrated. Hence, this thesis shows the link between decreased BCAA degradation in the adipose tissue to higher circulating levels of BCAA in plasma as occurring in obesity. This thesis also shows the degradation of BCAA catabolism and increased circulating plasma BCAA in obesity as associating with higher LDL-C, higher total cholesterol, higher fasting insulin levels and larger adipocyte diameter. Circulating BCAAs have been consistently found to correlate with HOMA (45, 58, 317), insulin resistance (58, 298), and they are important predictors for future diabetes (90). Several studies have found higher circulating BCAAs (26, 58, 269, 275, 318) and downregulation of SAT BCAA catabolism (21, 319) in obesity. This thesis has now established the link between BCAA catabolism downregulation in SAT to the occurrence of increased circulating

BCAAs in plasma. Also, this thesis has uniquely shown this link already in increased BMI (in young, healthy people), in both Study I and III. Furthermore, the downregulation of BCAA catabolism and increased plasma BCAAs were also found in the heavy co-twins of the groups of unhealthy metabolic profiles in both of these studies (further discussed in Chapter 6.6).

Two other studies have compared circulating BCAAs between MHO and MUO (45, 266). One study (n=10 for lean and healthy, n=10 for MHO and n=10 for MUO; MHO and MUO classified based on BMI and %bodyfat, insulin, lipids and glucose values) found differences in BCAAs that were not significant between MHO and MUO, but significantly different between MHO and the lean and healthy group as well as between MUO and the lean and healthy group (45). The other study (n=119 for lean and healthy, n=114 for MHO and n=738 for MUO; MHO and MUO classified based on BMI and glucose and blood pressure values) found significant differences in BCAAs between MHO and MUO groups (266). Here, the lack of significant findings in the former study could be because the study was conducted on a small sample size and used a different criteria for MHO and MUO classification. Unlike these prior studies, the analyses in Study I and III did not aim to form subgroups based on clinical measures and then subsequently compare the gene expression and metabolite profiles. With regards to clinical measures, the unhealthy groups in Study I and Study III differed in lipid and insulin values from the healthier groups. While the definition of MUO is wide, involving several clinical measures, the generally acceptable difference between MHO and MUO fits in well with the clinical differences observed in the clusters in this thesis.

From this finding, it is now evident that as the AT function worsens in obesity; BCAA degradation is downregulated, resulting in increased plasma BCAAs. Even more interesting is that this thesis, by employing within-twin pair analyses, has shown both phenomena (decreased BCAA degradation, increased circulating BCAAs) to be free of genetic confounding.

To the best of my knowledge, this is also the first study examining the association between circulating BCAAs and all three fat depots, namely SAT, VAT and LF. Because BMI does not distinguish between fat and lean mass, associations with other adiposity measures are important to explore. Circulating BCAAs (320) have previously been found to positively associate with VAT but not SAT or BMI (320). Conversely, in this study, BCAA was found to associate with BMI but not with any of the fat depots. Although no significant associations were found for BCAA and individual fat depots in this thesis, this could be due to the relatively healthy sample set.

6.5. Adiposity and blood biochemistry measures associating with metabolic risk factors (Study III)

BMI is often used as a measure of obesity because, in addition to being easy to obtain, the general understanding is that it is a reasonable measure of adiposity (321-325), Several studies have found high correlations between BMI and adiposity measures including %bodyfat, SAT and VAT (107,

326) further supporting this understanding. However, the effectiveness of BMI, as an obesity measure, in predicting metabolic health is still under question. Some studies indicate that BMI do not perform as well as body composition measures (327, 328) in predicting metabolic health. Others suggest that BMI outperforms (329, 330) or is equal to more detailed adiposity measures in identifying metabolic risks (331). In this thesis, high correlations of BMI with %bodyfat, SAT and VAT were identified in the dataset, a finding also seen in the other studies (107, 326). On a molecular level, in Study III, it was shown that BMI, compared to %bodyfat, SAT, VAT, and LF, associated positively with the most metabolites, namely valine, tyrosine, propionylcarnitine, deoxycytidine, cysteine and aspartate. In comparison, the fat depot measures associated with three or less metabolites. Out of the six metabolites associating with BMI, three also associated with the fat depot measures: aspartate associated with all three fat depots (SAT, VAT and LF), propionylcarnitine with SAT and VAT, and deoxycytidine with SAT. The metabolite-BMI association has been studied extensively, revealing associations with various BCAAs, AAAs and acylcarnitines (58, 74, 88, 275). Additionally, valine, tyrosine and propionylcarnitine also show significant difference between lean and obese people (58, 88). The findings in this thesis of the six metabolites associating significantly with BMI with only three of them also associating with fat depot measures are in agreement with studies that propose BMI as outperforming other adiposity measures and further strengthens the possible role of these metabolites as biomarkers for obesity.

Other than Study III, to my knowledge, no other studies have compared metabolites associating with SAT, VAT and LF. A previous study comparing metabolites extracted from SAT and VAT in obese and non-obese subjects found that, in SAT, the difference between obese and non-obese associated with differences in only one metabolite, 2-ketoisocaproic acid, which is a derivative of BCAA leucine (267). In VAT, the differences between obese and non-obese subjects included methionine, threonine, lysine, serine and leucine (267). While this was a study on tissue metabolites and not plasma metabolites, it highlights that obesity associates with more VAT metabolites than SAT metabolites. Another study found significant serum metabolite associations (including BCAAs valine and isoleucine, and AAAs tryptophan and phenylalanine) for VAT only in men, while no associations were found for SAT (320). Unlike these two studies that found more metabolites associating with VAT than SAT in obesity, in Study III, SAT associated with one more metabolite than VAT, namely deoxycytidine. Additionally, LF was found to associate with plasma metabolite aspartate. This thesis has uniquely found metabolite associations across all three fat depots and by employing within-twin pair analyses to examine these associations, has shown them to be free of genetic and shared environment confounding. Importantly, this thesis has shown, for the first time that the effect sizes for SAT–, VAT– and LF–metabolite associations were larger than the associations with BMI. This suggests that these associated metabolites are more sensitive to changes in %bodyfat, SAT, VAT and LF compared to changes in BMI. While Study III did not identify any metabolites that were uniquely associated with the different fat depots, this could be due to the small sample size, as well as the limited number of metabolites studied.

While all adiposity measures had associations with one or more metabolites, only two blood biochemistry measures had significantly associating metabolites in this thesis. HDL-C and TG measures associated mostly with AAs and acylcarnitines. Across all the adiposity and blood biochemistry measures, HDL-C also associated with the most number of metabolites. Additionally, for metabolites that associate with HDL-C in addition to other phenotypes, the effect size for the association with HDL-C was always larger than the effect sizes with other phenotypes. HDL-C has been found to be a consistent risk indicator for metabolic health (332-336). The findings in this thesis that show HDL-C associating with the most number of metabolites in comparison to other phenotypes support the comprehensiveness of HDL-C as a metabolic health measure. Additionally, for the first time, this thesis uniquely highlights that HDL-C levels, due to its association with a large number of metabolites along with the large effect sizes of these association, can be used as an indirect measure of metabolite level changes indicative of metabolic health.

6.6. Biological pathways and clinical phenotypes of unhealthy groups (Study I, III)

In this thesis, genes and metabolites identified as significantly associating with obesity and metabolic risk factors were further analysed using pathway tools to understand the main biological changes that unfold in the adipose tissue and blood plasma as a response to excess body weight. Cluster analysis enabled identification of groups of individuals with similar gene expression or metabolite profiles.

Healthy and unhealthy obesity subgroups (clusters)
In Study I, three subgroups of obesity were identified. The transcriptomics profiling revealed three subgroups (referred to as Cluster 1, Cluster 2 and Cluster 3) which differ by the gene expression differences within the twin pairs. Cluster 1 comprised twin pairs with slight within-pair differences in lipid metabolism and cell signalling pathways. However, because Cluster 1 had only two twin pairs, no definitive conclusions may be drawn regarding this cluster. Cluster 2 showed lower BCAA catabolism and lower mitochondrial function in the heavy co-twins, while Cluster 3 showed even lower mitochondrial function and BCAA catabolism and higher inflammation in the heavy co-twins. Cluster 3 also exhibited higher adipocyte size and insulin resistance in the heavy co-twins. Cluster 2 thus appears to be the healthy obese cluster, while Cluster 3 is the unhealthy obese cluster. In agreement with these findings, BCAA levels have previously been noted to be lower in MHO versus MUO individuals (266). Obese individuals without pro-inflammatory adipose tissue changes also display more favourable clinical characteristics (337), suggesting that a lower inflammation state lowers the risk for cardiovascular disease in this group of people (7).

Unhealthy clusters and related phenotypes
In Study I, the unhealthy group had lower mitochondrial function, higher inflammation along with higher IR and larger adipocyte size in the heavy co-twins. In Study III, using twin individuals, the

unhealthy cluster had higher concentrations of AAs (BCAAs and AAAs) and acylcarnitines, along with higher levels of total cholesterol and LDL-C, in comparison to the healthy cluster. Figure 19 depicts the pathways and phenotypes identified.

The rest of this chapter integrates the molecular findings from the cluster analyses based on transcriptomics profiles in adipose tissue (Study I) and metabolite profiles in plasma (Study III), and discusses these in light of the clinical properties observed in the unhealthy clusters. Specifically, this part of the discussion will focus on associating the metabolic disturbances in the unhealthy obesity phenotype to adiposity measures (adipocyte size) and biochemistry measures (TCHOL, LDL-C and IR).

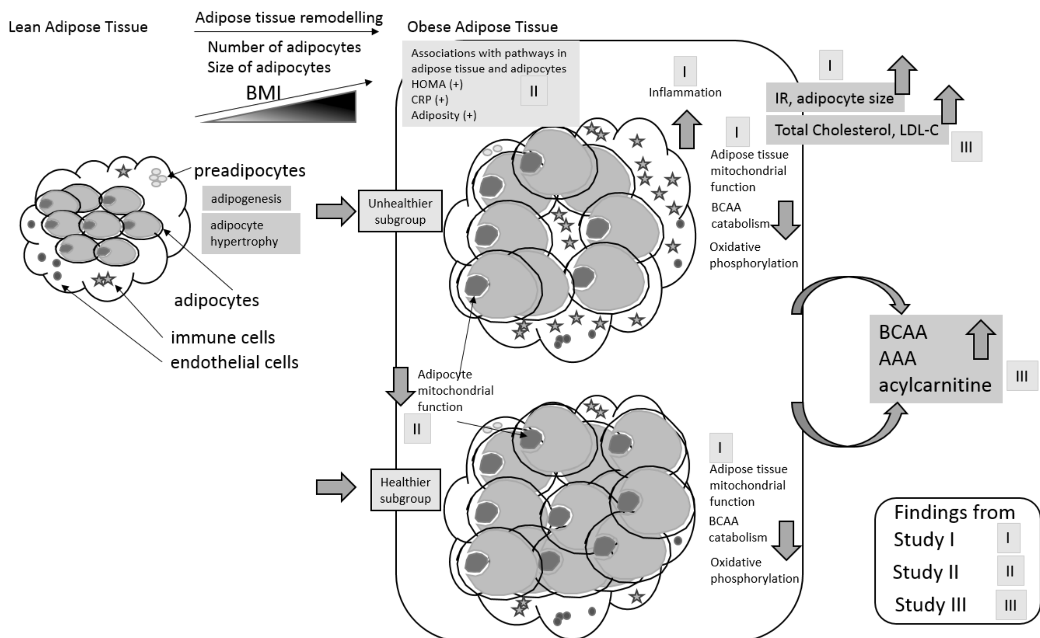

Figure 19: *Main integrated results from Study I, II and III. Two clusters were identified using metabolite profiles; higher AAA and BCAA associated with higher total cholesterol and LDL-C in the unhealthier cluster of the metabolite profile. Three clusters were identified using the gene expression profiles; in the unhealthier cluster there was mitochondrial downregulation, higher inflammation that associated with higher adipocyte size and higher insulin resistance. The downregulation of the pathways in the obese in adipocytes were similar to the ones found in adipose tissue – mitochondrial downregulation was consistently identified in the obese co-twins. These pathways identified in the adipocytes and adipose tissue associated with adiposity, HOMA and CRP values. The sequence of events from normal adipose tissue to MHO and MUO adipose tissue is not known.*

In Study III, two groups of individuals were identified based on their metabolite profiles. One group appeared to be unhealthier compared to the other group. In this unhealthier group, circulating total cholesterol and LDL-C values were higher than in the other group. High plasma cholesterol has previously been shown to correlate positively with adipocyte cholesterol (338). There is also a strong correlation between adipocyte cholesterol content and adipocyte size (339). Adipocyte size, in turn, positively correlates with the macrophage content of adipose tissue and expression of pro-inflammatory cytokines, such as Tumor Necrosis Factor-α (167). Hence, there is evidence for correlations between high cholesterol, adipocyte size and inflammation. High plasma LDL-C levels result in more LDL-C uptake by cells through the LDL receptor, which leads to a compensatory downregulation of cellular cholesterol synthesis and uptake (340), highlighting a negative association between plasma LDL-C levels and lipid metabolism. However, this negative feedback regulation of cholesterol metabolism can be overridden by inflammatory responses (340). This inflammatory response attracts monocytes/macrophages to infiltrate the adipose tissue (255). High cholesterol also disrupts secretion of adipokines that control appetite and satiety (338), potentially leading to overeating. Disruption of the cellular cholesterol homeostasis in adipocytes and occurrences of inflammation in obesity have been proposed to precede the development of T2D in obesity (340). It is proposed here that with increasing weight and nutrition, higher circulating lipids result in more lipid uptake, mitochondrial oxidation is reduced and the adipocyte dysfunctions. Results of this thesis show that mitochondrial downregulation occurs in obesity, while both mitochondrial downregulation and higher inflammation occur in more metabolically disadvantaged obesity. The dysfunctional adipocytes may cause the adipose tissue to recruit more macrophages (341), resulting in inflammation of adipose tissue (34). This inflammation impairs insulin signalling (342) by affecting phosphorylation of both the insulin receptor and its substrate, IRS-1 (343, 344). Alternatively, the dysfunctional adipocytes with higher oxidative stress and lower mitochondrial biogenesis (156) may allow the build up of ROS, disrupting insulin signalling (345, 346). The inflammatory state may in fact be causal in the development of IR and the other disorders associated with obesity, such as hyperlipidaemia and metabolic syndrome (170, 171).

While there is no consensus of the exact clinical trait criteria to distinguish between MHO and MUO (6, 8, 347, 348), prior studies comparing MHO and MUO subjects are largely in agreement with the findings of the healthy and unhealthy subgroups of this thesis. A study comparing clinical measures found that MHO individuals had significantly lower VAT, fasting insulin, TG, CRP and high HDL compared to MUO individuals (7). Lower amounts of CRP levels, despite high levels of body fat, was suggested to contribute to the favourable metabolic profile of the MHO subjects (7). Interestingly, in that study there was an overlap of CRP levels between MHO and MUO subjects (7), being in line with the findings in this thesis of no significant differences between the healthy and unhealthy clusters when comparing within-twin pair differences in CRP levels. Other studies on metabolite profile differences between MHO and MUO subjects found higher circulating BCAAs and acylcarnitines in MUO people (266, 349). Along with this difference in metabolite profile, both

serum HDL-C and TG levels, as well as fasting insulin levels, have been found to be significantly different in MHO compared to MUO people (349). These findings of different levels of HDL-C and fasting insulin were also found in Study I and Study III of this thesis. Another combined transcriptomics and metabolomics study utilising serum and adipose tissue found downregulation of BCAA catabolism and higher circulating BCAAs and acylcarnitines in MHO subjects (45). While these findings of BCAAs and acylcarnitines match the findings of this thesis, this thesis has also shown this obesity-related perturbations in the adipose tissue and plasma to be free of genetic confounding.

This thesis has now built a deeper understanding of the differences between healthy obesity and unhealthy obesity groups by showing a worsening of mitochondrial function in obesity in the unhealthy group and associating the molecular perturbations in the unhealthy phenotype to adiposity measures (adipocyte size) and biochemistry measures (TCHOL, LDL-C and IR). It has also been now shown, with the aid of within-twin pair analyses, that most of the molecular disturbances are free of genetic confounding. This is a new finding. With these analyses, it was possible to rule out the effects of genetics on the observed within-twin pair differences. The gene expression and metabolite differences between individuals with the same genotype could well be because of differences in their epigenetic profiles. The epigenetic marks react to environmental effects and thereby mediate these effects on the function of the genome. The epigenetics of obesity was not studied as part of this thesis, however. A previous epigenetic study utilising twins from the same cohort, however, revealed DNA methylation patterns in 17 genes that associate with obesity (22). Out of these 17, only 3 genes were in Study I's list of within-twin pair differentially expressed genes, hinting at perhaps some epigenetic mechanism other than DNA methylation driving the expression and metabolite differences within twin pairs in this thesis.

It should also be remembered that there is no way of knowing if both co-twins (among the BMI-discordant twin pairs) are genetically predisposed to being lean, with one subsequently becoming heavier. It is just as possible that both co-twins are genetically predisposed to being heavy. But, because of differences in lifestyle choices, one co-twin is leaner.

7. STRENGTHS AND LIMITATIONS

The strengths of these three studies include the use of a comprehensive and detailed set of clinically relevant phenotypes (both adiposity and blood biochemistry measures), along with high-dimensional metabolomics and transcriptomics data to understand obesity from both a molecular and symptomatic level. The twin pair samples allowed for the rare opportunity to study acquired obesity independent of genetic confounding. This thesis also employs bioinformatics methods developed specifically for transcriptomics and metabolomics data, capable of dealing with low numbers of samples and high numbers of observations. Regardless, the studies in this thesis do suffer from low sample numbers and relevant associations with low effect sizes may have been missed (231).

However, only 26 young adult twin pairs were BMI-discordant from the available Finnish twin cohorts comprising 5417 twin pairs. Additionally, measuring the transcript levels does not necessarily translate to proteins levels because mRNA levels may not accurately predict protein abundance (350, 351). Profiling of mRNA expression shows only the transcriptional activity of a gene, while targeted protein level analysis would allow for the detection of the physical presence and location of the respective proteins (352). Hence, to confirm protein abundance, the genes identified as differentially expressed in this thesis would need to be further tested by protein expression analyses or other functional assays in the future.

This cross-sectional study also does not allow for the determination of cause and effect; only conclusions on the level of associations can be drawn. Inferring causality would require longitudinal studies or Mendelian randomisation studies that use genetic variants to infer causality between obesity phenotypes and gene expression levels (353, 354).

In Study III, a limited number of metabolites were analysed. Because of limitations in available technology, most metabolite profiling experiments are not able to target and quantify a large number of metabolites (355). Hence, most metabolite profiling studies do not allow for a more comprehensive view of the thousands of metabolites that are involved in the metabolic pathways in the body. While the Human Metabolite Database (HMDB) has thousands of metabolites identified and categorised, technology to both identify and quantify these metabolites has not caught up or may be too expensive. The metabolomics platform used in this thesis was capable of identifying 111 metabolites. Deriving biological meaning behind the results were severely impacted by the lack of information about the metabolites that were not measured that might, in the end, play a role in the broader biological picture. In the end, understanding the metabolomics of the body as a complete system in the context of the study question becomes a difficult task. This is one limitation in Study III.

While sample clustering was applied using both gene expression and metabolomics data and all Study I participants are also in Study III, the grouping of the twin pairs and individual twins in the clusters are not comparable. In Study I, gene expression differences within the twin pairs were used and in Study III, individual metabolite concentrations were used. Hence, clusters could not be compared for sample membership. If the clusters had been comparable, it would have been possible to further

validate the findings of the obesity subgroups by determining if the same individuals group into the same clusters based on both the transcriptomics and metabolite profiles. However, the aims of Study I and III were slightly different and the focus of the clustering was to identify subgroups that point to gene expression profiles in acquired obesity (Study I) and metabolite profiles in obesity and related cardiometabolic risk factors (Study III), respectively.

The low number of subjects per cluster also provided lower statistical power. In the cluster analysis, it is not known if more data could have revealed more clusters or even different clusters. While biological validation was done in Study I to confirm the cluster findings in a replication dataset, this thesis would have benefited from additional cluster validation. For example, the clusters could have been further tested with permutation testing whereby gene expression and metabolite data can be permutated and further tested for cluster membership with within-cluster and between-cluster variance measurements used to represent the quality of the clusters (309). This is one area where clustering analyses could have been done better in this study. Studies combining other twin cohorts available globally could also be a step to address this power issue. However, to the best of my knowledge, such similar datasets (with comprehensive clinical phenotype information as in this thesis) do not exist.

In Study I, both physical activity and food intake were recorded to gain insight regarding environmental effects that might explain the BMI and associated gene expression differences within twin pairs. No significant differences were found in both physical activity and food intake. However, it has been suggested that obese individuals tend to underreport food intake by as much as 30% (356). Additionally, most obese individuals are believed to ingest more calories than lean individuals matched for exercise, to maintain their elevated weights (70). Hence, the findings may not reflect the actual difference in physical activity and food intake in the twin pairs.

8. CONCLUSION AND FUTURE PROSPECTS

In reference to the aims of this thesis, the main conclusions are as follows:

- This study confirms that acquired obesity is associated with changes in mitochondria-related gene expression in adipose tissue.
- Using gene expression profiles to identify sub-types of obesity is a feasible effort that should be further confirmed with more independent datasets.
- The pathways implicated in acquired obesity are similar for both SAT and adipocytes.
- BMI outperforms other adiposity measures including the fat depot measures. However, for SAT, VAT and LF, the effect sizes for metabolite–phenotype associations were larger than the associations of the same metabolites with BMI.
- HDL-C shows associations with the most number of metabolites with the largest effect sizes. This finding highlights HDL-C's potential as an indirect measure of metabolite level changes indicative of metabolic health.
- Downregulation of BCAA degradation in the obese adipose tissue is reflected in the increased circulating plasma BCAA.

Constant improvements in the microarray and metabolomics technologies have improved data quality, increased the number of genes and metabolites interrogated and initiated the development of countless automated statistical tools and methodology with which to study them. The studies above would benefit from improvements to the publicly available databases on the human metabolome as well as the transcriptome. Also, bearing in mind that these technologies are susceptible to technical variation, many of these findings, especially with regards to genes implicated in obesity, would benefit from validation in the laboratory via other techniques like RT-PCR. Replication cohorts would also be beneficial in confirming findings from these studies. However, as mentioned previously, to my knowledge no other BMI-discordant MZ cohort, with detailed clinical and 'omics data similar to the cohort studied in this thesis exists. However, some of the findings from this thesis can be confirmed using individuals as obese cases versus controls.

There is growing evidence that epigenetic marks associate with obesity (22, 357, 358). In MZ twin pairs, where there are no differences in gene sequences within the twin pairs, differences in gene expression could well be explained by epigenetics marks (e.g., DNA methylation or histone marks). Additionally, environmental differences (e.g., diet and exercise) within the twin pairs may result in epigenetic differences at genomic loci that regulate gene expression levels or metabolite concentrations. By integrating findings from various 'omics studies, a more detailed picture of obesity can emerge. One future prospect for this thesis would be to study the complex interaction between different 'omics layers and the clinical measures using, for example, group factor analysis (359) to interrogate the relationships between groups of 'omics data.

In conclusion, this thesis shows that by employing comprehensive methods in bioinformatics it was possible to first explore patterns in both transcriptomics and metabolomics data, perform detailed

analyses and derive biological meaning that allowed for the successful integration of findings from both types of OMICS studies.

ACKNOWLEDGEMENTS

This work was carried out at the Institute for Molecular Medicine Finland (FIMM), the Department of Public Health and the Obesity Research Unit as part of the Doctoral Program in Biomedicine (DPBM), Doctoral School of Health, Faculty of Medicine, University of Helsinki. In addition, this study has been financially supported by DPBM and the Biomedicum Helsinki Foundation, as well as the Sigrid Juselius Foundation. Present and former directors and department heads are respectfully thanked for providing excellent research facilities and a motivating work environment.

I started this journey towards a PhD in a rather unconventional manner. Recently made redundant from an organization I loved, I was looking for a fresh new start. Two amazing women gave this opportunity to me. I thank Adjunct Professor Miina Ollikainen who extended me her hand in friendship, showed me kindness, compassion and unwavering support. Miina, you have been a willing ear at all times and provided guidance in countless ways. Your invaluable input and sharp eye for what makes good research have made my work robust and wholesome. I also thank Professor Kirsi Pietiläinen who has continuously encouraged me every step of the way, taking the time to acknowledge every little effort of mine. Kirsi, you have a steadfast enthusiasm for science and it is hard not to get swept along in the euphoric wave of even the smallest achievements. Thank you for believing in me, honouring me with your appreciation and letting me be part of your important work.

My deepest gratitude goes out to Professor Jaakko Kaprio as well, for his quiet strength and focused steering of FIMM and the Kaprio group. With your foresight in setting up and nurturing the renowned Finnish Twin datasets, you have paved the way for exciting research and boundless opportunities. I also thank the pre-examiners of this thesis, Professor Marjukka Kolehmainen and Adjunct Professor Laura Elo-Uhlgren for their help in improving this thesis. Your insights made me look at several things in a new way, and as a result I have a thesis I am proud of. I would also like to thank Assistant Professor Sara Hägg for graciously accepting the position as official opponent at the public examination of this thesis. To my thesis committee advisors, Rainer Lehtonen and Sirkka-Liisa Varvio, thank you for taking time to encourage me along the way. Alyce Whipp, my language editor who jumped into action at the slightest notice, checked my text with great zeal while also having time to share the magical world of Prythian with me, to you, I am extremely grateful. I also profusely thank all my co-authors, Jana Buzkova, Gema Frühbeck, Antti Hakkarainen, Sini Heinonen, Jaakko Kaprio, Jesper Lundbom, Nina Lundbom, Adil Mardinoglu, Aila Rissanen, Amaia Rodríguez, Vidya Velagapudi and Hannele Yki-Järvinen, who contributed their time and immeasurable expertise in all my publications.

To my friends Alyce, Anu, Eeva, Elina, Emma, Jade, Jenni, Lalitha, Leonie, Pauliina, Richa, Sara, Sari and all those I do not mention here, you have each enriched my life in countless ways, not least by continuously enabling me in my weekly quest for Nepalese tea.

I also thank my parents who gave me a wonderful, stable childhood filled with the magic of books and the comfort of food laced with so much love; I never doubted that I was cherished and special. Thank you to my dad who dreamed great dreams for me, even when I was too small to dream them myself. And to my mom, who with the deepest conviction always told me that I could do anything I set out to do. I also thank Jiva, my remarkable husband, who has stood up to every challenge life has thrown his way with the steadfast belief that hope must not be abandoned. I am honoured to know you. Jiva, in you I found home! To my miracle, Roshan Vel, you are and always will be my 'great expectation' and greatest accomplishment. Akka, as big sisters go, you are the very best. Thank you

for holding my hand from the very beginning. Missy, you are as beautiful as the art you create, thank you for supporting me always. I also thank Lara and Darmik who delight me endlessly. Lastly, I thank this beautiful land of opportunities, without which I wouldn't be where I am today.

REFERENCES

1. Pajunen P, Kotronen A, Korpi-Hyövälti E, Keinänen-Kiukaanniemi S, Oksa H, Niskanen L, et al. Metabolically healthy and unhealthy obesity phenotypes in the general population: the FIN-D2D Survey. BMC public health. 2011;11(1):754.
2. Bluher M. The distinction of metabolically 'healthy' from 'unhealthy' obese individuals. Curr Opin Lipidol. 2010;21.
3. Velho S, Paccaud F, Waeber G, Vollenweider P, Marques-Vidal P. Metabolically healthy obesity: different prevalences using different criteria. European Journal Of Clinical Nutrition. 2010;64:1043.
4. Hinnouho G-M, Czernichow S, Dugravot A, Batty GD, Kivimaki M, Singh-Manoux A. Metabolically Healthy Obesity and Risk of Mortality. Does the definition of metabolic health matter? 2013;36(8):2294-300.
5. Roberson LL, Aneni EC, Maziak W, Agatston A, Feldman T, Rouseff M, et al. Beyond BMI: The "Metabolically healthy obese" phenotype & its association with clinical/subclinical cardiovascular disease and all-cause mortality -- a systematic review. BMC Public Health. 2014;14(1):14.
6. Wildman RP, Muntner P, Reynolds K, et al. The obese without cardiometabolic risk factor clustering and the normal weight with cardiometabolic risk factor clustering: Prevalence and correlates of 2 phenotypes among the us population (nhanes 1999-2004). Archives of Internal Medicine. 2008;168(15):1617-24.
7. Karelis AD, Faraj M, Bastard JP, St-Pierre DH, Brochu M, Prud'homme D, et al. The metabolically healthy but obese individual presents a favorable inflammation profile. Journal of Clinical Endocrinology & Metabolism. 2005;90(7):4145-50.
8. Stefan N, Kantartzis K, Machann J, et al. Identification and characterization of metabolically benign obesity in humans. Archives of Internal Medicine. 2008;168(15):1609-16.
9. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. JAMA. 1999;282(16):1523-9.
10. Lavie CJ, Milani RV, Ventura HO. Obesity and Cardiovascular Disease. Journal of the American College of Cardiology. 2009;53(21):1925-32.
11. Flier JS. Obesity Wars. Cell. 2004;116(2):337-50.
12. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. Nature Reviews Molecular Cell Biology. 2008;9(5):367-77.
13. Calle EE, Thun MJ. Obesity and cancer. Oncogene. 0000;23(38):6365-78.
14. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. PLOS ONE. 2013;8(1):e53916.
15. Amling CL, Riffenburgh RH, Sun L, Moul JW, Lance RS, Kusuda L, et al. Pathologic variables and recurrence rates as related to obesity and race in men with prostate cancer undergoing radical prostatectomy. J Clin Oncol. 2004;22.
16. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003;348.
17. Walley AJ, Blakemore AIF, Froguel P. Genetics of obesity and the prediction of risk for health. Human Molecular Genetics. 2006;15(suppl_2):R124-R30.
18. Malis C, Rasmussen EL, Poulsen P, Petersen I, Christensen K, Beck-Nielsen H, et al. Total and Regional Fat Distribution is Strongly Influenced by Genetic Factors in Young and Elderly Twins. Obesity Research. 2005;13(12):2139-45.
19. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature genetics. 2010;42(11):937.
20. Speakman JR. Obesity: The Integrated Roles of Environment and Genetics. The Journal of Nutrition. 2004;134(8):2090S-105S.

21. Pietilainen KH, Naukkarinen J, Rissanen A, Saharinen J, Ellonen P, Keranen H, et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. PLoS Medicine / Public Library of Science. 2008;5(3):e51.
22. Pietilainen KH, Ismail K, Jarvinen E, Heinonen S, Tummers M, Bollepalli S, et al. DNA methylation and gene expression patterns in adipose tissue differ significantly within young adult monozygotic BMI-discordant twin pairs. Int J Obes. 2016;40(4):654-61.
23. Thompson D, Karpe F, Lafontan M, Frayn K. Physical Activity and Exercise in the Regulation of Human Adipose Tissue Physiology. Physiological Reviews. 2012;92(1):157-91.
24. Saltiel AR. Insulin resistance in the defense against obesity. Cell Metabolism. 2012;15(6):798-804.
25. Sharma AM, Staels B. Peroxisome proliferator-activated receptor γ and adipose tissue—understanding obesity-related changes in regulation of lipid and glucose metabolism. The Journal of Clinical Endocrinology & Metabolism. 2006;92(2):386-95.
26. Felig P, Marliss E, Cahill GF, Jr. Plasma amino acid levels and insulin secretion in obesity. New England Journal of Medicine. 1969;281(15):811-6.
27. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and Management of the Metabolic Syndrome. An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. 2005;112(17):2735-52.
28. Geetha L, Deepa M, Anjana RM, Mohan V. Prevalence and clinical profile of metabolic obesity and phenotypic obesity in Asian Indians. J Diabetes Sci Technol. 2011;5(2):439-46.
29. Hamer M, Stamatakis E. Metabolically Healthy Obesity and Risk of All-Cause and Cardiovascular Disease Mortality. The Journal of Clinical Endocrinology & Metabolism. 2012;97(7):2482-8.
30. Kip KE, Marroquin OC, Kelley DE, Johnson BD, Kelsey SF, Shaw LJ, et al. Clinical Importance of Obesity Versus the Metabolic Syndrome in Cardiovascular Risk in Women. A Report From the Women's Ischemia Syndrome Evaluation (WISE) Study. 2004;109(6):706-13.
31. Meigs JB, Wilson PW, Fox CS, Vasan RS, Nathan DM, Sullivan LM, et al. Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease. J Clin Endocrinol Metab. 2006;91.
32. Fabbrini E, Yoshino J, Yoshino M, Magkos F, Tiemann Luecking C, Samovski D, et al. Metabolically normal obese people are protected from adverse effects following weight gain. Journal of Clinical Investigation. 2015;125(2):787-95.
33. Naukkarinen J, Heinonen S, Hakkarainen A, Lundbom J, Vuolteenaho K, Saarinen L, et al. Characterising metabolically healthy obesity in weight-discordant monozygotic twins. Diabetologia. 2014;57(1):167-76.
34. Jung UJ, Choi MS. Obesity and its metabolic complications: the role of adipokines and the relationship between obesity, inflammation, insulin resistance, dyslipidemia and nonalcoholic fatty liver disease. International Journal of Molecular Sciences. 2014;15(4):6184-223.
35. Jensen MD. Role of Body Fat Distribution and the Metabolic Complications of Obesity. The Journal of Clinical Endocrinology & Metabolism. 2008;93(11_supplement_1):s57-s63.
36. Nielsen S, Guo Z, Johnson CM, Hensrud DD, Jensen MD. Splanchnic lipolysis in human obesity. Journal of Clinical Investigation. 2004;113(11):1582-8.
37. Wajchenberg BLo. Subcutaneous and Visceral Adipose Tissue: Their Relation to the Metabolic Syndrome. Endocrine Reviews. 2000;21(6):697-738.
38. Fontana L, Eagon JC, Trujillo ME, Scherer PE, Klein S. Visceral Fat Adipokine Secretion Is Associated With Systemic Inflammation in Obese Humans. Diabetes. 2007;56(4):1010-3.
39. Bjorntorp P. "Portal" adipose tissue as a generator of risk factors for cardiovascular disease and diabetes. Arteriosclerosis.10(4):493-6.
40. Fabbrini E, Magkos F, Mohammed BS, Pietka T, Abumrad NA, Patterson BW, et al. Intrahepatic fat, not visceral fat, is linked with metabolic complications of obesity. Proceedings of the National Academy of Sciences. 2009;106(36):15430-5.
41. Hoffstedt J, Förster D, Löfgren P. Impaired subcutaneous adipocyte lipogenesis is associated with systemic insulin resistance and increased apolipoprotein B/AI ratio in men and women. Journal of Internal Medicine. 2007;262(1):131-9.

42. Dharuri H, t Hoen PA, van Klinken JB, Henneman P, Laros JF, Lips MA, et al. Downregulation of the acetyl-CoA metabolic network in adipose tissue of obese diabetic individuals and recovery after weight loss. Diabetologia. 2014;57(11):2384-92.
43. Perreault M, Zulyniak MA, Badoud F, Stephenson S, Badawi A, Buchholz A, et al. A Distinct Fatty Acid Profile Underlies the Reduced Inflammatory State of Metabolically Healthy Obese Individuals. PLOS ONE. 2014;9(2):e88539.
44. Rämö JT, Kaye SM, Jukarainen S, Bogl LH, Hakkarainen A, Lundbom J, et al. Liver Fat and Insulin Sensitivity Define Metabolite Profiles During a Glucose Tolerance Test in Young Adult Twins. The Journal of Clinical Endocrinology & Metabolism. 2017;102(1):220-31.
45. Badoud F, Lam KP, DiBattista A, Perreault M, Zulyniak MA, Cattrysse B, et al. Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese. Journal of Proteome Research. 2014;13(7):3455-66.
46. Hotamisligil GS, Spiegelman BM. Tumor Necrosis Factor α: A Key Component of the Obesity-Diabetes Link. Diabetes. 1994;43(11):1271-8.
47. Khan A, Pessin J. Insulin regulation of glucose uptake: a complex interplay of intracellular signalling pathways. Diabetologia. 2002;45(11):1475-83.
48. Sherwin RS. Role of Liver in Glucose Homeostasis. Diabetes Care. 1980;3(2):261-5.
49. Zeyda M, Stulnig TM. Obesity, inflammation, and insulin resistance--a mini-review. Gerontology. 2009;55(4):379-86.
50. Boden G, Chen X, Ruiz J, White JV, Rossetti L. Mechanisms of fatty acid-induced inhibition of glucose uptake. The Journal of Clinical Investigation. 1994;93(6):2438-46.
51. Reaven GM, Hollenbeck C, Jeng C-Y, Wu MS, Chen Y-DI. Measurement of Plasma Glucose, Free Fatty Acid, Lactate, and Insulin for 24 h in Patients With NIDDM. Diabetes. 1988;37(8):1020-4.
52. Liu A, McLaughlin T, Liu T, Sherman A, Yee G, Abbasi F, et al. Differential intra-abdominal adipose tissue profiling in obese, insulin-resistant women. Obesity Surgery. 2009;19(11):1564-73.
53. Randle PJ, Garland PB, Hales CN, Newsholme EA. THE GLUCOSE FATTY-ACID CYCLE ITS ROLE IN INSULIN SENSITIVITY AND THE METABOLIC DISTURBANCES OF DIABETES MELLITUS. The Lancet. 1963;281(7285):785-9.
54. Lowell BB, Shulman GI. Mitochondrial Dysfunction and Type 2 Diabetes. Science. 2005;307(5708):384-7.
55. Petersen KF, Shulman GI. Etiology of Insulin Resistance. The American Journal of Medicine.119(5):S10-S6.
56. Hotamisligil GS, Shargill NS, Spiegelman BM. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. Science. 1993;259(5091):87-91.
57. Newgard CB. Interplay between lipids and branched-chain amino acids in development of insulin resistance. Cell Metabolism. 2012;15(5):606-14.
58. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metabolism. 2009;9(4):311-26.
59. Fridlyand LE, Philipson LH. Reactive species and early manifestation of insulin resistance in type 2 diabetes. Diabetes, Obesity and Metabolism. 2006;8(2):136-45.
60. Fleischman A, Kron M, Systrom DM, Hrovat M, Grinspoon SK. Mitochondrial Function and Insulin Resistance in Overweight and Normal-Weight Children. The Journal of Clinical Endocrinology and Metabolism. 2009;94(12):4923-30.
61. Eisenberg S. High density lipoprotein metabolism. Journal of lipid research. 1984;25(10):1017-58.
62. Ebbert J, Jensen M. Fat Depots, Free Fatty Acids, and Dyslipidemia. Nutrients. 2013;5(2):498.
63. Klop B, Elte JW, Cabezas MC. Dyslipidemia in obesity: mechanisms and potential targets. Nutrients. 2013;5(4):1218-40.
64. Gordon RS, Cherkes A. Unesterified fatty acid in human blood plasma. The Journal of clinical investigation. 1956;35(2):206-12.
65. Nilsson-Ehle P, Garfinkel AS, Schotz MC. Lipolytic enzymes and plasma lipoprotein metabolism. Annual review of biochemistry. 1980;49(1):667-93.

66. Ameer F, Scandiuzzi L, Hasnain S, Kalbacher H, Zaidi N. De novo lipogenesis in health and disease. Metabolism. 2014;63(7):895-902.
67. Boden G, Shulman GI. Free fatty acids in obesity and type 2 diabetes: defining their role in the development of insulin resistance and β-cell dysfunction. European Journal of Clinical Investigation. 2002;32:14-23.
68. Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. Physiol Rev. 2007;87(2):507-20.
69. Shulman GI. Cellular mechanisms of insulin resistance. The Journal of Clinical Investigation. 2000;106(2):171-6.
70. Spiegelman BM, Flier JS. Obesity and the Regulation of Energy Balance. Cell.104(4):531-43.
71. Iozzo P, Lautamaki R, Geisler F, Virtanen KA, Oikonen V, Haaparanta M, et al. Non-esterified fatty acids impair insulin-mediated glucose uptake and disposition in the liver. Diabetologia. 2004;47(7):1149-56.
72. Adams SH. Emerging perspectives on essential amino acid metabolism in obesity and the insulin-resistant state. Advances in Nutrition. 2011;2(6):445-56.
73. Jennings A, MacGregor A, Pallister T, Spector T, Cassidy A. Associations between branched chain amino acid intake and biomarkers of adiposity and cardiometabolic health independent of genetic factors: A twin study. Int J Cardiol. 2016;223:992-8.
74. Ho JE, Larson MG, Ghorbani A, Cheng S, Chen MH, Keyes M, et al. Metabolomic Profiles of Body Mass Index in the Framingham Heart Study Reveal Distinct Cardiometabolic Phenotypes. PLoS One. 2016;11(2):e0148361.
75. Felig P, Wahren J, Hendler R, Brundin T. Splanchnic glucose and amino acid metabolism in obesity. The Journal of Clinical Investigation. 1974;53(2):582-90.
76. Holeček M. The BCAA–BCKA cycle: its relation to alanine and glutamine synthesis and protein balance. Nutrition. 2001;17(1):70.
77. Doi M, Yamaoka I, Nakayama M, Sugahara K, Yoshizawa F. Hypoglycemic effect of isoleucine involves increased muscle glucose uptake and whole body glucose oxidation and decreased hepatic gluconeogenesis. Am J Physiol Endocrinol Metab. 2007;292(6):E1683-93.
78. Lynch CJ, Gern B, Lloyd C, Hutson SM, Eicher R, Vary TC. Leucine in food mediates some of the postprandial rise in plasma leptin concentrations. Am J Physiol Endocrinol Metab. 2006;291(3):E621-30.
79. Ichihara A, Koyama E. Transaminase of Branched Chain Amino Acids

I. Branched Chain Amino Acids-α-Ketoglutarate Transaminase. The Journal of Biochemistry. 1966;59(2):160-9.
80. Wahren J, Felig P, Hagenfeldt L. Effect of protein ingestion on splanchnic and leg metabolism in normal man and in patients with diabetes mellitus. Journal of Clinical Investigation. 1976;57(4):987.
81. Herman MA, She P, Peroni OD, Lynch CJ, Kahn BB. Adipose Tissue Branched Chain Amino Acid (BCAA) Metabolism Modulates Circulating BCAA Levels. Journal of Biological Chemistry. 2010;285(15):11348-56.
82. Felig P, Marliss E, Cahill GFJ. Plasma Amino Acid Levels and Insulin Secretion in Obesity. New England Journal of Medicine. 1969;281(15):811-6.
83. Leskinen T, Rinnankoski-Tuikka R, Rintala M, Seppanen-Laakso T, Pollanen E, Alen M, et al. Differences in muscle and adipose tissue gene expression and cardio-metabolic risk factors in the members of physical activity discordant twin pairs. PLoS ONE [Electronic Resource]. 2010;5(9).
84. Rennie MJ. Influence of Exercise on Protein and Amino Acid Metabolism. Comprehensive Physiology: John Wiley & Sons, Inc.; 2010.
85. Harper AE, Miller RH, Block KP. Branched-Chain Amino Acid Metabolism. Annual Review of Nutrition. 1984;4(1):409-54.
86. Frerman FE, Sabran JL, Taylor JL, Grossberg SE. Leucine catabolism during the differentiation of 3T3-L1 cells. Expression of a mitochondrial enzyme system. Journal of Biological Chemistry. 1983;258(11):7087-93.

87. Kitsy A, Carney S, Vivar JC, Knight MS, Pointer MA, Gwathmey JK, et al. Effects of Leucine Supplementation and Serum Withdrawal on Branched-Chain Amino Acid Pathway Gene and Protein Expression in Mouse Adipocytes. PLOS ONE. 2014;9(7):e102615.
88. Kim JY, Park JY, Kim OY, Ham BM, Kim H-J, Kwon DY, et al. Metabolic profiling of plasma in overweight/obese and lean men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF MS). Journal of Proteome Research. 2010;9(9):4368-75.
89. Würtz P, Soininen P, Kangas AJ, Rönnemaa T, Lehtimäki T, Kähönen M, et al. Branched-Chain and Aromatic Amino Acids Are Predictors of Insulin Resistance in Young Adults. Diabetes Care. 2013;36(3):648-55.
90. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nature medicine. 2011;17(4):448-53.
91. Fernstrom JD. Branched-Chain Amino Acids and Brain Function. The Journal of Nutrition. 2005;135(6):1539S-46S.
92. Trayhurn P. Adipocyte biology. Obesity Reviews. 2007;8:41-4.
93. Saely CH, Geiger K, Drexel H. Brown versus White Adipose Tissue: A Mini-Review. Gerontology. 2012;58(1):15-23.
94. Birsoy K, Festuccia WT, Laplante M. A comparative perspective on lipid storage in animals. Journal of Cell Science. 2013;126(7):1541-52.
95. Sethi JK, Vidal-Puig AJ. Thematic review series: adipocyte biology. Adipose tissue function and plasticity orchestrate nutritional adaptation. J Lipid Res. 2007;48(6):1253-62.
96. Mora S, Pessin JE. An adipocentric view of signaling and intracellular trafficking. Diabetes Metab Res Rev. 2002;18(5):345-56.
97. Fantuzzi G. Adipose tissue, adipokines, and inflammation. J Allergy Clin Immunol. 2005;115(5):911-9; quiz 20.
98. Mariman EC, Wang P. Adipocyte extracellular matrix composition, dynamics and role in obesity. Cell Mol Life Sci. 2010;67(8):1277-92.
99. Rosen ED, MacDougald OA. Adipocyte differentiation from the inside out. Nature Reviews Molecular Cell Biology. 2006;7(12):885-96.
100. Frühbeck G, Gómez-Ambrosi J, Muruzábal FJ, Burrell MA. The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation. American Journal of Physiology - Endocrinology And Metabolism. 2001;280(6):E827-E47.
101. Trayhurn P, Beattie JH. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. Proceedings of the Nutrition Society. 2007;60(3):329-39.
102. Kershaw EE, Flier JS. Adipose Tissue as an Endocrine Organ. The Journal of Clinical Endocrinology & Metabolism. 2004;89(6):2548-56.
103. Tchoukalova YD, Koutsari C, Karpyak MV, Votruba SB, Wendland E, Jensen MD. Subcutaneous adipocyte size and body fat distribution. The American Journal of Clinical Nutrition. 2008;87(1):56-63.
104. Cinti S, Mitchell G, Barbatelli G, Murano I, Ceresi E, Faloia E, et al. Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. Journal of lipid research. 2005;46(11):2347-55.
105. Skurk T, Alberti-Huber C, Herder C, Hauner H. Relationship between Adipocyte Size and Adipokine Expression and Secretion. The Journal of Clinical Endocrinology & Metabolism. 2007;92(3):1023-33.
106. Varlamov O, Somwar R, Cornea A, Kievit P, Grove KL, Roberts CT. Single-cell analysis of insulin-regulated fatty acid uptake in adipocytes. American journal of physiology Endocrinology and metabolism. 2010;299(3):E486-96.
107. Smith SR, Lovejoy JC, Greenway F, Ryan D, deJonge L, de la Bretonne J, et al. Contributions of total body fat, abdominal subcutaneous adipose tissue compartments, and visceral adipose tissue to the metabolic complications of obesity. Metabolism: Clinical & Experimental. 2001;50(4):425-35.
108. Fruhbeck G. Overview of adipose tissue and its role in obesity and metabolic disorders. Methods in Molecular Biology. 2008;456:1-22.

109. Belfort R, Harrison SA, Brown K, Darland C, Finch J, Hardies J, et al. A Placebo-Controlled Trial of Pioglitazone in Subjects with Nonalcoholic Steatohepatitis. New England Journal of Medicine. 2006;355(22):2297-307.
110. Kim J-Y, van de Wall E, Laplante M, Azzara A, Trujillo ME, Hofmann SM, et al. Obesity-associated improvements in metabolic profile through expansion of adipose tissue. The Journal of Clinical Investigation. 2007;117(9):2621-37.
111. Söderberg C, Stål P, Askling J, Glaumann H, Lindberg G, Marmur J, et al. Decreased survival of subjects with elevated liver function tests during a 28-year follow-up. Hepatology. 2010;51(2):595-602.
112. Ludwig J, Viggiano TR, Mcgill DB, Oh B, editors. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. Mayo Clinic Proceedings; 1980.
113. Rodriguez-Gallego E, Guirro M, Riera-Borrull M, Hernandez-Aguilera A, Marine-Casado R, Fernandez-Arroyo S, et al. Mapping of the circulating metabolome reveals [alpha]-ketoglutarate as a predictor of morbid obesity-associated non-alcoholic fatty liver disease. Int J Obes. 2015;39(2):279-87.
114. Kotronen A, Yki-Jarvinen H. Fatty liver: a novel component of the metabolic syndrome. Arterioscler Thromb Vasc Biol. 2008;28(1):27-38.
115. Hwang J-H, Stein DT, Barzilai N, Cui M-H, Tonelli J, Kishore P, et al. Increased intrahepatic triglyceride is associated with peripheral insulin resistance: in vivo MR imaging and spectroscopy studies. American Journal of Physiology - Endocrinology And Metabolism. 2007;293(6):E1663-E9.
116. van der Poorten D, Milner KL, Hui J, Hodge A, Trenell MI, Kench JG, et al. Visceral fat: a key mediator of steatohepatitis in metabolic liver disease. Hepatology. 2008;48(2):449-57.
117. Wynn TA. Cellular and molecular mechanisms of fibrosis. The Journal of Pathology. 2008;214(2):199-210.
118. Farrell GC, Larter CZ. Nonalcoholic fatty liver disease: From steatosis to cirrhosis. Hepatology. 2006;43(S1):S99-S112.
119. Smith SR, Lovejoy JC, Greenway F, Ryan D, deJonge L, de la Bretonne J, et al. Contributions of total body fat, abdominal subcutaneous adipose tissue compartments, and visceral adipose tissue to the metabolic complications of obesity. Metabolism. 2001;50(4):425-35.
120. Abate N, Garg A, Peshock RM, Stray-Gundersen J, Grundy SM. Relationships of generalized and regional adiposity to insulin sensitivity in men. The Journal of clinical investigation. 1995;96(1):88-98.
121. von Eyben FE, Mouritsen E, Holm J, Montvilas P, Dimcevski G, Suciu G, et al. Intra-abdominal obesity and metabolic risk factors: a study of young adults[ast]. International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity. 0000;27(8):941-9.
122. Goodpaster BH, Leland Thaete F, Simoneau J-A, Kelley DE. Subcutaneous Abdominal Fat and Thigh Muscle Composition Predict Insulin Sensitivity Independently of Visceral Fat. Diabetes. 1997;46(10):1579-85.
123. Wajchenberg BL, Giannella-Neto D, da Silva ME, Santos RF. Depot-specific hormonal characteristics of subcutaneous and visceral adipose tissue and their relation to the metabolic syndrome. Hormone & Metabolic Research. 2002;34(11-12):616-21.
124. Arner P. Differences in Lipolysis between Human Subcutaneous and Omental Adipose Tissues. Annals of Medicine. 1995;27(4):435-8.
125. Hajer GR, van Haeften TW, Visseren FLJ. Adipose tissue dysfunction in obesity, diabetes, and vascular diseases. European Heart Journal. 2008;29(24):2959-71.
126. Klein S. The case of visceral fat: argument for the defense. The Journal of Clinical Investigation. 2004;113(11):1530-2.
127. Kanaya AM, Harris T, Goodpaster BH, Tylavsky F, Cummings SR, Health AaBCS. Adipocytokines attenuate the association between visceral adiposity and diabetes in older adults. Diabetes care. 2004;27(6):1375-80.
128. Yatagai T, Nagasaka S, Taniguchi A, Fukushima M, Nakamura T, Kuroe A, et al. Hypoadiponectinemia is associated with visceral fat accumulation and insulin resistance in Japanese men with type 2 diabetes mellitus. Metabolism: Clinical & Experimental. 2003;52(10):1274-8.

129. Després J-P, Lemieux I, Bergeron J, Pibarot P, Mathieu P, Larose E, et al. Abdominal Obesity and the Metabolic Syndrome: Contribution to Global Cardiometabolic Risk. Arteriosclerosis, Thrombosis, and Vascular Biology. 2008;28(6):1039-49.
130. Lewis GF, Uffelman KD, Szeto LW, Weller B, Steiner G. Interaction between free fatty acids and insulin in the acute control of very low density lipoprotein production in humans. The Journal of Clinical Investigation. 1995;95(1):158-66.
131. Westerbacka J, Cornér A, Tiikkainen M, Tamminen M, Vehkavaara S, Häkkinen A-M, et al. Women and men have similar amounts of liver and intra-abdominal fat, despite more subcutaneous fat in women: implications for sex differences in markers of cardiovascular risk. Diabetologia. 2004;47(8):1360-9.
132. Fox CS, Massaro JM, Hoffmann U, Pou KM, Maurovich-Horvat P, Liu C-Y, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. Circulation. 2007;116(1):39-48.
133. Grundy SM. Does a diagnosis of metabolic syndrome have value in clinical practice? The American Journal of Clinical Nutrition. 2006;83(6):1248-51.
134. Dusserre E, Moulin P, Vidal H. Differences in mRNA expression of the proteins secreted by the adipocytes in human subcutaneous and visceral adipose tissues. Biochimica et biophysica acta. 2000;1500(1):88-96.
135. Linder K, Arner P, Flores-Morales A, Tollet-Egnell P, Norstedt G. Differentially expressed genes in visceral or subcutaneous adipose tissue of obese men and women. Journal of lipid research. 2004;45(1):148-54.
136. Eriksson P, Van Harmelen V, Hoffstedt J, Lundquist P, Vidal H, Stemme V, et al. Regional variation in plasminogen activator inhibitor-1 expression in adipose tissue from obese individuals. Thrombosis and haemostasis. 2000;83(4):545-8.
137. Gabrielsson BG, Johansson JM, Lönn M, Jernås M, Olbers T, Peltonen M, et al. High Expression of Complement Components in Omental Adipose Tissue in Obese Men. Obesity Research. 2003;11(6):699-708.
138. Duffaut C, Zakaroff-Girard A, Bourlier V, Decaunes P, Maumus M, Chiotasso P, et al. Interplay Between Human Adipocytes and T Lymphocytes in Obesity. CCL20 as an Adipochemokine and T Lymphocytes as Lipogenic Modulators. 2009;29(10):1608-14.
139. Ortega FJ, Mayas D, Moreno-Navarrete JM, Catalán V, Gómez-Ambrosi J, Esteve E, et al. The Gene Expression of the Main Lipogenic Enzymes is Downregulated in Visceral Adipose Tissue of Obese Subjects. Obesity. 2010;18(1):13-20.
140. Ledoux S, Coupaye M, Essig M, Msika S, Roy C, Queguiner I, et al. Traditional Anthropometric Parameters Still Predict Metabolic Disorders in Women With Severe Obesity. Obesity. 2010;18(5):1026-32.
141. Tchernof A, Bélanger C, Morisset A-S, Richard C, Mailloux J, Laberge P, et al. Regional Differences in Adipose Tissue Metabolism in Women Minor Effect of Obesity and Body Fat Distribution2006. 1353-60 p.
142. Auguet T, Guiu-Jurado E, Berlanga A, Terra X, Martinez S, Porras JA, et al. Downregulation of lipogenesis and fatty acid oxidation in the subcutaneous adipose tissue of morbidly obese women. Obesity. 2014;22(9):2032-8.
143. Liesenfeld DB, Grapov D, Fahrmann JF, Salou M, Scherer D, Toth R, et al. Metabolomics and transcriptomics identify pathway differences between visceral and subcutaneous adipose tissue in colorectal cancer patients: the ColoCare study. American Journal of Clinical Nutrition. 2015;102(2):433-43.
144. Spalding KL, Arner E, Westermark PO, Bernard S, Buchholz BA, Bergmann O, et al. Dynamics of fat cell turnover in humans. Nature. 2008;453(7196):783-7.
145. Sun K, Kusminski CM, Scherer PE. Adipose tissue remodeling and obesity. The Journal of Clinical Investigation. 2011;121(6):2094-101.
146. Hirsch J, Batchelor B. Adipose Tissue Cellularity in Human Obesity1976. 299-311 p.
147. Wilson-Fritch L, Nicoloro S, Chouinard M, Lazar MA, Chui PC, Leszyk J, et al. Mitochondrial remodeling in adipose tissue associated with obesity and treatment with rosiglitazone. Journal of Clinical Investigation. 2004;114(9):1281-9.

148. Ducluzeau P-H, Priou M, Weitheimer M, Flamment M, Duluc L, Iacobazi F, et al. Dynamic regulation of mitochondrial network and oxidative functions during 3T3-L1 fat cell differentiation. Journal of Physiology and Biochemistry. 2011;67(3):285-96.
149. Tormos KV, Anso E, Hamanaka RB, Eisenbart J, Joseph J, Kalyanaraman B, et al. Mitochondrial complex III ROS regulate adipocyte differentiation. Cell metabolism. 2011;14(4):537-44.
150. Hoffstedt J, Arner E, Wahrenberg H, Andersson DP, Qvisth V, Lofgren P, et al. Regional impact of adipose tissue morphology on the metabolic profile in morbid obesity. Diabetologia. 2010;53(12):2496-503.
151. Virtue S, Vidal-Puig A. Adipose tissue expandability, lipotoxicity and the Metabolic Syndrome--an allostatic perspective. Biochim Biophys Acta. 2010;1801(3):338-49.
152. Medina-Gomez G, Virtue S, Lelliott C, Boiani R, Campbell M, Christodoulides C, et al. The Link Between Nutritional Status and Insulin Sensitivity Is Dependent on the Adipocyte-Specific Peroxisome Proliferator–Activated Receptor-γ2 Isoform. Diabetes. 2005;54(6):1706.
153. Kopelman PG. Obesity as a medical problem. Nature. 2000;404(6778):635-43.
154. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. The Lancet.371(9612):569-78.
155. Crewe C, An YA, Scherer PE. The ominous triad of adipose tissue dysfunction: inflammation, fibrosis, and impaired angiogenesis. The Journal of Clinical Investigation. 2017;127(1):74-82.
156. Kim J-a, Wei Y, Sowers JR. Role of Mitochondrial Dysfunction in Insulin Resistance. Circulation Research. 2008;102(4):401-14.
157. Sanyal AJ, Campbell–Sargent C, Mirshahi F, Rizzo WB, Contos MJ, Sterling RK, et al. Nonalcoholic steatohepatitis: Association of insulin resistance and mitochondrial abnormalities. Gastroenterology.120(5):1183-92.
158. Morino K, Petersen KF, Shulman GI. Molecular Mechanisms of Insulin Resistance in Humans and Their Potential Links With Mitochondrial Dysfunction. Diabetes. 2006;55(Supplement 2):S9-S15.
159. Heinonen S, Buzkova J, Muniandy M, Kaksonen R, Ollikainen M, Ismail K, et al. Impaired Mitochondrial Biogenesis in Adipose Tissue in Acquired Obesity. Diabetes. 2015;64(9):3135-45.
160. Boudina S, Graham TE. Mitochondrial function/dysfunction in white adipose tissue. Experimental Physiology. 2014;99(9):1168-78.
161. Visser M, Bouter LM, McQuillan GM, Wener MH, Harris TB. Elevated c-reactive protein levels in overweight and obese adults. JAMA. 1999;282(22):2131-5.
162. Sartipy P, Loskutoff DJ. Monocyte chemoattractant protein 1 in obesity and insulin resistance. Proceedings of the National Academy of Sciences. 2003;100(12):7265-70.
163. Fried SK, Bunkin DA, Greenberg AS. Omental and Subcutaneous Adipose Tissues of Obese Subjects Release Interleukin-6: Depot Difference and Regulation by Glucocorticoid1. The Journal of Clinical Endocrinology & Metabolism. 1998;83(3):847-50.
164. Bastard J-P, Maachi M, Tran Van Nhieu J, Jardel C, Bruckert E, Grimaldi A, et al. Adipose Tissue IL-6 Content Correlates with Resistance to Insulin Activation of Glucose Uptake both in Vivo and in Vitro2002. 2084-9 p.
165. Pasarica M, Sereda OR, Redman LM, Albarado DC, Hymel DT, Roan LE, et al. Reduced Adipose Tissue Oxygenation in Human Obesity. Evidence for Rarefaction, Macrophage Chemotaxis, and Inflammation Without an Angiogenic Response. 2009;58(3):718-25.
166. Trayhurn P, Wood IS. Adipokines: inflammation and the pleiotropic role of white adipose tissue. British Journal of Nutrition. 2007;92(3):347-55.
167. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL, Ferrante AW, Jr. Obesity is associated with macrophage accumulation in adipose tissue. Journal of Clinical Investigation. 2003;112(12):1796-808.
168. Xu H, Barnes GT, Yang Q, Tan G, Yang D, Chou CJ, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. Journal of Clinical Investigation. 2003;112(12):1821-30.
169. Wellen KE, Hotamisligil GS. Obesity-induced inflammatory changes in adipose tissue. J Clin Invest. 2003;112(12):1785-8.

170. Yudkin JS, Stehouwer CDA, Emeis JJ, Coppack SW. C-Reactive Protein in Healthy Subjects: Associations With Obesity, Insulin Resistance, and Endothelial Dysfunction. A Potential Role for Cytokines Originating From Adipose Tissue? 1999;19(4):972-8.
171. Hotamisligil G. Inflammatory pathways and insulin action. International Journal of Obesity. 2003;27(S3):S53.
172. Kaprio J, Buchsbaum M, Gottesman I, Heath A, Körner J, Kringlen E, et al., editors. What can twin studies contribute to the understanding of adult psychopathology. TJ Bouchard jr and P Propping: Twins as a tool for behavioral genetics Chichester: John Wiley & Sons, Dahlem Workshop Reports, Life Sciences Research Report; 1993.
173. Graner M, Seppala-Lindroos A, Rissanen A, Hakkarainen A, Lundbom N, Kaprio J, et al. Epicardial fat, cardiac dimensions, and low-grade inflammation in young adult monozygotic twins discordant for obesity. American Journal of Cardiology. 2012;109(9):1295-302.
174. Heinonen S, Muniandy M, Buzkova J, Mardinoglu A, Rodriguez A, Fruhbeck G, et al. Mitochondria-related transcriptional signature is downregulated in adipocytes in obesity: a study of young healthy MZ twins. Diabetologia. 2017;60(1):169-81.
175. Kaye S, Lokki AI, Hanttu A, Nissilä E, Heinonen S, Hakkarainen A, et al. Upregulation of Early and Downregulation of Terminal Pathway Complement Genes in Subcutaneous Adipose Tissue and Adipocytes in Acquired Obesity. Frontiers in Immunology. 2017;8(545).
176. Liao C, Gao W, Cao W, Lv J, Yu C, Wang S, et al. Associations of Body Composition Measurements with Serum Lipid, Glucose and Insulin Profile: A Chinese Twin Study. PLoS One. 2015;10(11):e0140595.
177. Kussmann M, Raymond F, Affolter M. OMICS-driven biomarker discovery in nutrition and health. Journal of Biotechnology. 2006;124(4):758-87.
178. Chan ET, Quon GT, Chua G, Babak T, Trochesset M, Zirngibl RA, et al. Conservation of core gene expression in vertebrate tissues. Journal of Biology. 2009;8(3):33.
179. Kai T, Williams D, Spradling AC. The expression profile of purified Drosophila germline stem cells. Developmental Biology. 2005;283(2):486-502.
180. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403:503.
181. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences. 1999;96(12):6745-50.
182. Li T, Huang J, Jiang Y, Zeng Y, He F, Zhang MQ, et al. Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. Genomics. 2009;93(3):235-42.
183. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of Drosophila melanogaster. Nature. 2011;471(7339):473.
184. Rochette A, Raymond F, Ubeda J-M, Smith M, Messier N, Boisvert S, et al. Genome-wide gene expression profiling analysis of Leishmania major and Leishmania infantum developmental stages reveals substantial differences between the two species. BMC Genomics. 2008;9(1):255.
185. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, et al. Gene-specific patterns of expression variation across organs and species. Genome Biology. 2016;17(1):151.
186. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87-98.
187. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.
188. Okoniewski MJ, Miller CJ. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics. 2006;7:276-.
189. Royce TE, Rozowsky JS, Gerstein MB. Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic Acids Research. 2007;35(15):e99-e.
190. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet.
191. Jenssen T-K, Langaas M, Kuo WP, Smith-Sørensen B, Myklebost O, Hovig E. Analysis of repeatability in spotted cDNA microarrays. Nucleic acids research. 2002;30(14):3235-44.

192. Yang YH, Dudoit S, Luu P, Speed TP. Normalization for cDNA microarray data. Microarrays: Optical Technologies and Informatics. 2001;4266.
193. Microarray platforms – comparisons and contrasts. Pharmacogenomics. 2004;5(5):487-502.
194. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, et al. Multiple-laboratory comparison of microarray platforms. Nat Meth. 2005;2(5):345-50.
195. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biology. 2011;9(1):34.
196. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4.
197. A Irizarry R, M Bolstad B, Collin F, M Cope L, Hobbs B, P Speed T. Summaries of Affymetrix GeneChip Probe Level Data2003. e15 p.
198. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association. 2004;99(468):909-17.
199. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG. [1] The Affymetrix GeneChip® Platform: An Overview. Methods in Enzymology. 410: Academic Press; 2006. p. 3-28.
200. Nicholson JK. Global systems biology, personalized medicine and molecular epidemiology. Molecular Systems Biology. 2006;2(1).
201. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov. 2002;1(2):153-61.
202. Nicholson J, Lindon JC, Holmes E. 'Metabonomics': Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli Via Multivariate Statistical Analysis of Biological Nmr Spectroscopic Data1999. 1181-9 p.
203. Lindon J, Holmes E, Nicholson J. An Overview of Metabonomics2005. 1-26 p.
204. Nicholson JK, Wilson ID. Understanding 'Global' Systems Biology: Metabonomics and the Continuum of Metabolism. Nat Rev Drug Discov. 2003;2(8):668-76.
205. Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotech. 2001;19(1):45-50.
206. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protocols. 2007;2(11):2692-703.
207. Dettmer K, Aronov PA, Hammock BD. MASS SPECTROMETRY-BASED METABOLOMICS. Mass spectrometry reviews. 2007;26(1):51-78.
208. Vigneau-Callahan KE, Shestopalov AI, Milbury PE, Matson WR, Kristal BS. Characterization of diet-dependent metabolic serotypes: analytical and biological variability issues in rats. The Journal of nutrition. 2001;131(3):924S-32S.
209. Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. Analytica Chimica Acta. 2012;750(Supplement C):82-97.
210. Pauli GF, Jaki BU, Lankin DC. Quantitative 1H NMR: Development and Potential of a Method for Natural Products Analysis. Journal of Natural Products. 2005;68(1):133-49.
211. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422(6928):198-207.
212. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. Nat Rev Cardiol. 2011;8(11):630-43.
213. Locatelli M, Melucci D, Carlucci G, Locatelli C. Recent HPLC strategies to improve sensitivity and selectivity for the analysis of complex matrices2012. 112-37 p.
214. Kapetanovic IM, Rosenfeld S, Izmirlian G. Overview of Commonly Used Bioinformatics Methods and Their Applications. Annals of the New York Academy of Sciences. 2004;1020(1):10-21.
215. Smyth G. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments2004. Article3 p.

216. Zhang X, Wei D, Yap Y, Li L, Guo S, Chen F. Mass spectrometry-based “omics” technologies in cancer diagnostics. Mass Spectrometry Reviews. 2007;26(3):403-31.
217. Ng AY, editor Preventing" overfitting" of cross-validation data. ICML; 1997.
218. Wu B. Differential gene expression detection and sample classification using penalized linear regression models. Bioinformatics. 2005;22(4):472-6.
219. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267-88.
220. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7(1):142.
221. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA. Maximum likelihood estimation of optimal scaling factors for expression array normalization. Proc SPIE BIOS. 2001;132.
222. Bolstad B, Irizarry R, Astrand M, Speed T. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19.
223. De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Tull D, et al. Normalizing and Integrating Metabolomics Data. Analytical Chemistry. 2012;84(24):10768-76.
224. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47.
225. Quackenbush J. Microarray data normalization and transformation. Nat Genet. 2002;32.
226. Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics. 2005;21(18):3683-5.
227. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80.
228. Jauhiainen A, Madhu B, Narita M, Narita M, Griffiths J, Tavaré S. Normalization of metabolomics data with applications to correlation maps. Bioinformatics. 2014;30(15):2155-61.
229. Szabo A, Boucher K, Carroll W, Klebanov L, Tsodikov A, Yakovlev A. Variable selection and pattern recognition with gene expression data generated by the microarray technology. Math Biosci. 2002;176.
230. Tsodikov A, Szabo A, Jones D. Adjustments and measures of differential expression for microarray data. Bioinformatics. 2002;18(2):251-60.
231. Qiu X, Wu H, Hu R. The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis. BMC Bioinformatics. 2013;14(1):124.
232. Tsodikov A, Szabo A, Jones D. Adjustments and measures of differential expression for microarray data. Bioinformatics. 2002;18.
233. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1901;2(11):559-72.
234. Ringner M. What is principal component analysis? Nat Biotech. 2008;26(3):303-4.
235. Landgrebe J, Wurst W, Welzl G. Permutation-validated principal components analysis of microarray data. Genome Biology. 2002;3(4):research0019.1.
236. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med. 2001;7(6):673-9.
237. Do KA, McLachlan G, Bean R, Wen S. Application of Gene Shaving and Mixture Models to Cluster Microarray Gene Expression Data2007. 25-43 p.
238. Grewal R, Das S. Microarray data analysis: Gaining biological insights2013. 996-1005 p.
239. Getz G, Levine E, Domany E. Coupled two-way clustering analysis of gene microarray data. Proceedings of the National Academy of Sciences. 2000;97(22):12079-84.
240. de Souto MC, Costa IG, de Araujo DS, Ludermir TB, Schliep A. Clustering cancer gene expression data: a comparative study. BMC Bioinformatics. 2008;9(1):497.
241. Richards AL, Holmans P, O'Donovan MC, Owen MJ, Jones L. A comparison of four clustering methods for brain expression microarray data. BMC Bioinformatics. 2008;9(1):490.

242. Wahl S, Krug S, Then C, Kirchhofer A, Kastenmüller G, Brand T, et al. Comparative analysis of plasma metabolomics response to metabolic challenge tests in healthy subjects and influence of the FTO obesity risk allele. Metabolomics. 2014;10(3):386-401.
243. De Smet F, Mathys J, Marchal K, Thijs G, De Moor B, Moreau Y. Quality-Based Clustering of Gene Expression Profiles2002. 735-46 p.
244. Sherlock G. Analysis of large-scale gene expression data. Current Opinion in Immunology. 2000;12(2):201-5.
245. McLachlan GJ, Bean RW, Peel D. A mixture model-based approach to the clustering of microarray expression data. Bioinformatics. 2002;18(3):413-22.
246. žurauskienė J, Yau C. pcaReduce: hierarchical clustering of single cell transcriptional profiles. BMC Bioinformatics. 2016;17(1):140.
247. Halkidi M, Batistakis Y, Vazirgiannis M. On Clustering Validation Techniques. Journal of Intelligent Information Systems. 2001;17(2):107-45.
248. Ge Y, Dudoit S, Speed TP. Resampling-based multiple testing for microarray data analysis. Test. 2003;12(1):1-77.
249. Hatfield GW, Hung S-p, Baldi P. Differential analysis of DNA microarray gene expression data. Molecular Microbiology. 2003;47(4):871-7.
250. Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes Analysis of a Microarray Experiment. Journal of the American Statistical Association. 2001;96(456):1151-60.
251. Perneger TV. What's wrong with Bonferroni adjustments. BMJ : British Medical Journal. 1998;316(7139):1236-8.
252. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995;57:289-300.
253. Nair S, Lee YH, Rousseau E, Cam M, Tataranni PA, Baier LJ, et al. Increased expression of inflammation-related genes in cultured preadipocytes/stromal vascular cells from obese compared with non-obese Pima Indians. Diabetologia. 2005;48(9):1784-8.
254. van Erk MJ, Pasman WJ, Wortelboer HM, van Ommen B, Hendriks HFJ. Short-term fatty acid intervention elicits differential gene expression responses in adipose tissue from lean and overweight men. Genes & Nutrition. 2008;3(3-4):127-37.
255. Lee YH, Nair S, Rousseau E, Allison DB, Page GP, Tataranni PA, et al. Microarray profiling of isolated abdominal subcutaneous adipocytes from obese vs non-obese Pima Indians: increased expression of inflammation-related genes. Diabetologia. 2005;48(9):1776-83.
256. Walley AJ, Jacobson P, Falchi M, Bottolo L, Andersson JC, Petretto E, et al. Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. International journal of obesity. 2012;36(1):137-47.
257. Badoud F, Brewer D, Charchoglyan A, Cuthbertson DJ, Mutch DM. Multi-omics Integrative Investigation of Fatty Acid Metabolism in Obese and Lean Subcutaneous Tissue. OMICS: A Journal of Integrative Biology. 2017;21(7):371-9.
258. Hoggard N, Cruickshank M, Moar K-M, Bashir S, Mayer C-D. Using Gene Expression to Predict Differences in the Secretome of Human Omental vs. Subcutaneous Adipose Tissue. Obesity. 2012;20(6):1158-67.
259. Le KA, Mahurkar S, Alderete TL, Hasson RE, Adam TC, Kim JS, et al. Subcutaneous adipose tissue macrophage infiltration is associated with hepatic and visceral fat deposition, hyperinsulinemia, and stimulation of NF-kappaB stress pathway. Diabetes. 2011;60(11):2802-9.
260. Klimcakova E, Roussel B, Marquez-Quinones A, Kovacova Z, Kovacikova M, Combes M, et al. Worsening of obesity and metabolic status yields similar molecular adaptations in human subcutaneous and visceral adipose tissue: decreased metabolism and increased immune response. Journal of Clinical Endocrinology & Metabolism. 2011;96(1):73.
261. Rodriguez-Acebes S, Palacios N, Botella-Carretero JI, Olea N, Crespo L, Peromingo R, et al. Gene expression profiling of subcutaneous adipose tissue in morbid obesity using a focused microarray: distinct expression of cell-cycle- and differentiation-related genes. BMC Medical Genomics [Electronic Resource]. 2010;3:61.

262. Wolfs MG, Rensen SS, Bruin-Van Dijk EJ, Verdam FJ, Greve JW, Sanjabi B, et al. Co-expressed immune and metabolic genes in visceral and subcutaneous adipose tissue from severely obese individuals are associated with plasma HDL and glucose levels: a microarray study. BMC Medical Genomics [Electronic Resource]. 2010;3:34.
263. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA. 2001;98.
264. Henegar C, Tordjman J, Achard V, Lacasa D, Cremer I, Guerre-Millo M, et al. Adipose tissue transcriptomic signature highlights the pathological relevance of extracellular matrix in human obesity. Genome Biology. 2008;9(1):R14.
265. Taleb S, Lacasa D, Bastard JP, Poitou C, Cancello R, Pelloux V, et al. Cathepsin S, a novel biomarker of adiposity: relevance to atherogenesis. FASEB J. 2005;19.
266. Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. Metabolism: Clinical & Experimental. 2013;62(7):961-9.
267. Hanzu FA, Vinaixa M, Papageorgiou A, Parrizas M, Correig X, Delgado S, et al. Obesity rather than regional fat depots marks the metabolomic pattern of adipose tissue: an untargeted metabolomic approach. Obesity. 2014;22(3):698-704.
268. Mai M, Tonjes A, Kovacs P, Stumvoll M, Fiedler GM, Leichtle AB. Serum levels of acylcarnitines are altered in prediabetic conditions. PLoS One. 2013;8(12):e82459.
269. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222-31.
270. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448-53.
271. Ahmad MS, Alsaleh M, Kimhofer T, Ahmad S, Jamal W, Wali SO, et al. Metabolic Phenotype of Obesity in a Saudi Population. J Proteome Res. 2017;16(2):635-44.
272. Kraus WE, Pieper CF, Huffman KM, Thompson DK, Kraus VB, Morey MC, et al. Association of Plasma Small-Molecule Intermediate Metabolites With Age and Body Mass Index Across Six Diverse Study Populations. J Gerontol A Biol Sci Med Sci. 2016;71(11):1507-13.
273. Gogna N, Krishna M, Oommen AM, Dorai K. Investigating correlations in the altered metabolic profiles of obese and diabetic subjects in a South Indian Asian population using an NMR-based metabolomic approach. Molecular Biosystems. 2015;11(2):595-606.
274. Zhao H, Shen J, Djukovic D, Daniel-MacDougall C, Gu H, Wu X, et al. Metabolomics-identified metabolites associated with body mass index and prospective weight gain among Mexican American women. Obes Sci Pract. 2016;2(3):309-17.
275. Moore SC, Matthews CE, Sampson JN, Stolzenberg-Solomon RZ, Zheng W, Cai Q, et al. Human metabolic correlates of body mass index. Metabolomics. 2014;10(2):259-69.
276. Kaprio J. Twin studies in Finland 2006. Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies. 2006;9(6):772-7.
277. Kaprio J. The Finnish Twin Cohort Study: an update. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2013;16(1):157-62.
278. Muniandy M, Heinonen S, Yki-Jarvinen H, Hakkarainen A, Lundbom J, Lundbom N, et al. Gene expression profile of subcutaneous adipose tissue in BMI-discordant monozygotic twin pairs unravels molecular and clinical changes associated with sub-types of obesity. Int J Obes (Lond). 2017.
279. Pietilainen KH, Sysi-Aho M, Rissanen A, Seppanen-Laakso T, Yki-Jarvinen H, Kaprio J, et al. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study. PLoS ONE [Electronic Resource]. 2007;2(2):e218.
280. Pietrobelli A, Formica C, Wang Z, Heymsfield SB. Dual-energy X-ray absorptiometry body composition model: review of physical concepts. American Journal of Physiology. 1996;271(6 Pt 1):941.
281. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. CirculationCardiovascular Genetics. 2015;8(1):192-206.

282. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. American Journal of Clinical Nutrition. 1982;36(5):936-42.
283. Heinonen S, Saarinen L, Naukkarinen J, Rodriguez A, Fruhbeck G, Hakkarainen A, et al. Adipocyte morphology and implications for metabolic derangements in acquired obesity. International journal of obesity. 2014;38(11):1423-31.
284. Rao JNK, Scott AJ. On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. Ann Statist. 1984;12:46-60.
285. Roman-Garcia P, Quiros-Gonzalez I, Mottram L, Lieben L, Sharan K, Wangwiwatsin A, et al. Vitamin B(1)(2)-dependent taurine synthesis regulates growth and bone mass. J Clin Invest. 2014;124(7):2988-3002.
286. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic acids research. 2005;33(20):e175.
287. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2006.
288. Team RDC. R: A Language and Environment for Statistical Computing. 2008.
289. Gomez-Ambrosi J, Silva C, Galofre JC, Escalada J, Santos S, Millan D, et al. Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity. International journal of obesity. 2012;36(2):286-94.
290. Carlin JB, Gurrin LC, Sterne JAC, Morley R, Dwyer T. Regression models for twin studies: a critical review. International Journal of Epidemiology. 2005;34(5):1089-99.
291. Sjolander A, Johansson ALV, Lundholm C, Altman D, Almqvist C, Pawitan Y. Analysis of 1 : 1 Matched Cohort Studies and Twin Studies, with Binary Exposures and Binary Outcomes. Statist Sci. 2012;27(3):395-411.
292. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics. 2009;10(1):57.
293. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome research. 2008;18(9):1509-17.
294. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research. 2009;37(1):1-13.
295. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Briefings in Bioinformatics. 2015;16(1):59-70.
296. Jeanmougin M, de Reynies A, Marisa L, Paccard C, Nuel G, Guedj M. Should We Abandon the t-Test in the Analysis of Gene Expression Microarray Data: A Comparison of Variance Modeling Strategies. PLOS ONE. 2010;5(9):e12336.
297. Boulet MM, Chevrier G, Grenier-Larouche T, Pelletier M, Nadeau M, Scarpa J, et al. Alterations of plasma metabolite profiles related to adipose tissue distribution and cardiometabolic risk. Am J Physiol Endocrinol Metab. 2015;309(8):E736-46.
298. Huffman KM, Shah SH, Stevens RD, Bain JR, Muehlbauer M, Slentz CA, et al. Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. Diabetes care. 2009;32(9):1678-83.
299. Menni C, Migaud M, Glastonbury CA, Beaumont M, Nikolaou A, Small KS, et al. Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. Obesity. 2016;24(6):1380-8.
300. Neeland IJ, Ayers CR, Rohatgi AK, Turer AT, Berry JD, Das SR, et al. Associations of visceral and abdominal subcutaneous adipose tissue with markers of cardiac and metabolic risk in obese adults. Obesity. 2013;21(9):439.
301. Böhm A, Halama A, Meile T, Zdichavsky M, Lehmann R, Weigert C, et al. Metabolic Signatures of Cultured Human Adipocytes from Metabolically Healthy versus Unhealthy Obese Individuals. PLOS ONE. 2014;9(4):e93148.
302. Newbern D, Balikcioglu PG, Balikcioglu M, Bain J, Muehlbauer M, Stevens R, et al. Sex differences in biomarkers associated with insulin resistance in obese adolescents: metabolomic profiling and principal components analysis. The Journal of clinical endocrinology and metabolism. 2014;99(12):4730.

303. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98(19):10869-74.
304. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(3):811-6.
305. Mehrian-Shai R, Shi T, J Kremen T, Horvath S, Liau L, F Cloughesy T, et al. Gene expression profiling identifies molecular subtypes of gliomas2006. 4256- p.
306. Wang S, Sparks LM, Xie H, Greenway FL, de Jonge L, Smith SR. Subtyping obesity with microarrays: implications for the diagnosis and treatment of obesity. International journal of obesity. 2009;33(4):481-9.
307. Quackenbush J. Computational analysis of microarray data. Nature Reviews Genetics. 2001;2:418.
308. Eisen MB, Brown PO. [12] DNA arrays for analysis of gene expression. Methods in enzymology. 303: Elsevier; 1999. p. 179-205.
309. Park PJ, Manjourides J, Bonetti M, Pagano M. A permutation test for determining significance of clusters with applications to spatial and gene expression data. Computational statistics & data analysis. 2009;53(12):4290-300.
310. Jacob S, Nodzenski M, Reisetter AC, Bain JR, Muehlbauer MJ, Stevens RD, et al. Targeted Metabolomics Demonstrates Distinct and Overlapping Maternal Metabolites Associated With BMI, Glucose, and Insulin Sensitivity During Pregnancy Across Four Ancestry Groups. Diabetes Care. 2017;40(7):911-9.
311. Hageman JA, van den Berg RA, Westerhuis JA, Hoefsloot HCJ, Smilde AK. Bagged K-Means Clustering of Metabolome Data. Critical Reviews in Analytical Chemistry. 2006;36(3-4):211-20.
312. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. The American Journal of Clinical Nutrition. 2011;93(2):314-21.
313. Yin X, Lanza IR, Swain JM, Sarr MG, Nair KS, Jensen MD. Adipocyte mitochondrial function is reduced in human obesity independent of fat cell size. Journal of Clinical Endocrinology & Metabolism. 2014;99(2):209.
314. Fischer B, Schottl T, Schempp C, Fromme T, Hauner H, Klingenspor M, et al. Inverse relationship between body mass index and mitochondrial oxidative phosphorylation capacity in human subcutaneous adipocytes. American Journal of Physiology - Endocrinology & Metabolism. 2015;309(4):380.
315. H A Lardy a, Ferguson SM. Oxidative Phosphorylation in Mitochondria. Annual Review of Biochemistry. 1969;38(1):991-1034.
316. Yin Z, Deng T, Peterson LE, Yu R, Lin J, Hamilton DJ, et al. Transcriptome analysis of human adipocytes implicates the NOD-like receptor pathway in obesity-induced adipose inflammation. Molecular & Cellular Endocrinology. 2014;394(1-2):80-7.
317. Shah SH, Crosslin DR, Haynes CS, Nelson S, Turer CB, Stevens RD, et al. Branched-chain amino acid levels are associated with improvement in insulin resistance with weight loss. Diabetologia. 2012;55(2):321-30.
318. Gaudet MM, Falk RT, Stevens RD, Gunter MJ, Bain JR, Pfeiffer RM, et al. Analysis of serum metabolic profiles in women with endometrial cancer and controls in a population-based case-control study. J Clin Endocrinol Metab. 2012;97(9):3216-23.
319. Mardinoglu A, Kampf C, Asplund A, Fagerberg L, Hallstrom BM, Edlund K, et al. Defining the human adipose tissue proteome to reveal metabolic alterations in obesity. Journal of Proteome Research. 2014;13(11):5106-19.
320. Schlecht I, Gronwald W, Behrens G, Baumeister SE, Hertel J, Hochrein J, et al. Visceral adipose tissue but not subcutaneous adipose tissue is associated with urine and serum metabolites. PloS one. 2017;12(4):e0175133.
321. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. Journal of Chronic Diseases. 1972;25(6):329-43.
322. S. Garrow J, Gandy J. Quetelet's Index (W/H2) as a Measure of Fatness1985. 147-53 p.
323. Smalley KJ, Knerr AN, Kendrick ZV, Colliver JA, Owen OE. Reassessment of body mass indices. The American Journal of Clinical Nutrition. 1990;52(3):405-8.

324. Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. British Journal of Nutrition. 2007;65(2):105-14.
325. Strain GW, Zumoff B. The relationship of weight-height indices of obesity to body fat content. Journal of the American College of Nutrition. 1992;11(6):715-8.
326. Boeke CE, Oken E, Kleinman KP, Rifas-Shiman SL, Taveras EM, Gillman MW. Correlations among adiposity measures in school-aged children. BMC Pediatrics. 2013;13(1):99.
327. Carroll JF, Chiapa AL, Rodriquez M, Phelps DR, Cardarelli KM, Vishwanatha JK, et al. Visceral fat, waist circumference, and BMI: impact of race/ethnicity. Obesity (Silver Spring). 2008;16(3):600-7.
328. H Lahmann P, Lissner L, Gullberg B, Berglund G. A Prospective Study of Adiposity and All-Cause Mortality: The Malm?? Diet and Cancer Study2002. 361-9 p.
329. Lee K, Song YM, Sung J. Which obesity indicators are better predictors of metabolic risk?: healthy twin study. Obesity (Silver Spring). 2008;16(4):834-40.
330. Bosy-Westphal A, Geisler C, Onur S, Korth O, Selberg O, Schrezenmeir J, et al. Value of body fat mass vs anthropometric obesity indices in the assessment of metabolic risk factors. Int J Obes (Lond). 2006;30(3):475-83.
331. Paniagua L, Lohsoonthorn V, Lertmaharit S, Jiamjarasrangsi W, Williams M. Comparison of Waist Circumference, Body Mass Index, Percent Body Fat and Other Measure of Adiposity in Identifying Cardiovascular Disease Risks among Thai Adults2008. 215-23 p.
332. Castelli WP, Garrison RJ, Wilson PF, Abbott RD, Kalousdian S, Kannel WB. Incidence of coronary heart disease and lipoprotein cholesterol levels: The framingham study. JAMA. 1986;256(20):2835-8.
333. Assmann G, Schulte H, von Eckardstein A, Huang Y. High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and pathophysiological implications for reverse cholesterol transport. Atherosclerosis.124:S11-S20.
334. Curb JD, Abbott RD, Rodriguez BL, Masaki K, Chen R, Sharp DS, et al. A prospective study of HDL-C and cholesteryl ester transfer protein gene mutations and the risk of coronary heart disease in the elderly. Journal of Lipid Research. 2004;45(5):948-53.
335. Sharrett AR, Ballantyne CM, Coady SA, Heiss G, Sorlie PD, Catellier D, et al. Coronary Heart Disease Prediction From Lipoprotein Cholesterol Levels, Triglycerides, Lipoprotein(a), Apolipoproteins A-I and B, and HDL Density Subfractions. Circulation. 2001;104(10):1108.
336. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The American Journal of Medicine.62(5):707-14.
337. Farb MG, Bigornia S, Mott M, Tanriverdi K, Morin KM, Freedman JE, et al. Reduced adipose tissue inflammation represents an intermediate cardiometabolic phenotype in obesity. Journal of the American College of Cardiology. 2011;58(3):232-7.
338. Aguilar D, Fernandez ML. Hypercholesterolemia induces adipose dysfunction in conditions of obesity and nonobesity. Advances in Nutrition. 2014;5(5):497-502.
339. Le Lay S, Ferré P, Dugail I. Adipocyte cholesterol balance in obesity. Biochemical Society Transactions. 2004;32(1):103-6.
340. Ding J, Reynolds LM, Zeller T, Muller C, Lohman K, Nicklas BJ, et al. Alterations of a Cellular Cholesterol Metabolism Network Are a Molecular Feature of Obesity-Related Type 2 Diabetes and Cardiovascular Disease. Diabetes. 2015;64(10):3464-74.
341. Suganami T, Nishida J, Ogawa Y. A Paracrine Loop Between Adipocytes and Macrophages Aggravates Inflammatory Changes: Role of Free Fatty Acids and Tumor Necrosis Factor a. Arteriosclerosis, Thrombosis, and Vascular Biology. 2005;25(10):2062-8.
342. Cai D, Yuan M, Frantz DF, Melendez PA, Hansen L, Lee J, et al. Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. Nat Med. 2005;11(2):183-90.
343. Hotamisligil GS, Budavari A, Murray D, Spiegelman BM. Reduced tyrosine kinase activity of the insulin receptor in obesity-diabetes. Central role of tumor necrosis factor-alpha. Journal of Clinical Investigation. 1994;94(4):1543.
344. Hotamisligil GS, Peraldi P, Budavari A, Ellis R, White MF, Spiegelman BM. IRS-1-Mediated Inhibition of Insulin Receptor Tyrosine Kinase Activity in TNF-α- and Obesity-Induced Insulin Resistance. Science. 1996;271(5249):665-70.

345. Nakamura S, Takamura T, Matsuzawa-Nagata N, Takayama H, Misu H, Noda H, et al. Palmitate induces insulin resistance in H4IIEC3 hepatocytes through reactive oxygen species produced by mitochondria. Journal of Biological Chemistry. 2009;284(22):14809-18.
346. Anderson EJ, Lustig ME, Boyle KE, Woodlief TL, Kane DA, Lin C-T, et al. Mitochondrial H2O2 emission and cellular redox state link excess fat intake to insulin resistance in both rodents and humans. The Journal of clinical investigation. 2009;119(3):573.
347. Karelis AD, Brochu M, Rabasa-Lhoret R. Can we identify metabolically healthy but obese individuals (MHO)? Diabetes & Metabolism. 2004;30(6):569-72.
348. Primeau V, Coderre L, Karelis A, Brochu M, Lavoie M-E, Messier V, et al. Characterizing the profile of obese patients who are metabolically health2010. 971-81 p.
349. Gao X, Zhang W, Wang Y, Pedram P, Cahill F, Zhai G, et al. Serum metabolic biomarkers distinguish metabolically healthy peripherally obese from unhealthy centrally obese individuals. Nutr Metab (Lond). 2016;13:33.
350. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. ELECTROPHORESIS. 1997;18(3-4):533-7.
351. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19(3):1720-30.
352. Hegde PS, White IR, Debouck C. Interplay of transcriptomics and proteomics. Current Opinion in Biotechnology. 2003;14(6):647-51.
353. Meng Q, Mäkinen V-P, Luk H, Yang X. Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases. Current Cardiovascular Risk Reports. 2013;7(1):73-83.
354. Katan M. Apoupoprotein e isoforms, serum cholesterol, and cancer. The Lancet. 1986;327(8479):507-8.
355. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical Applications of Metabolomics in Oncology: A Review. Clinical Cancer Research. 2009;15(2):431-40.
356. Lichtman SW, Pisarska K, Berman ER, Pestone M, Dowling H, Offenbacher E, et al. Discrepancy between Self-Reported and Actual Caloric Intake and Exercise in Obese Subjects. New England Journal of Medicine. 1992;327(27):1893-8.
357. Bays H, Scinta W. Adiposopathy and epigenetics: an introduction to obesity as a transgenerational disease. Current Medical Research & Opinion. 2015;31(11):2059-69.
358. Lavebratt C, Almgren M, Ekström T. Epigenetic regulation in obesity. International journal of obesity. 2012;36(6):757.
359. Klami A, Virtanen S, Leppäaho E, Kaski S. Group factor analysis. IEEE transactions on neural networks and learning systems. 2015;26(9):2136-47.