The study of the effect of glucocorticoids on global and tissue-specific metabolism in humans

By

Riccardo Di Guida

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

April 2017

Abstract

Glucocorticoids are a class of steroid hormones which are highly relevant in human health and disease as they are involved in the regulation of carbohydrate, protein and fatty acid metabolism and are instrumental in the onset or progression of various diseases including those associated with glucocorticoid deficiency or excess. Their importance is strengthened because of their pharmacological use in humans including in immune-mediated diseases, for example, the use of cortisol for the treatment of arthritis. The regulation of global metabolism through changes in glucocorticoid metabolism is an important area of clinical research that has been relatively poorly investigated outside carbohydrate and fatty acid metabolism. As an example, cortisol and insulin are involved in diurnal metabolic processes but their effects and interaction on healthy subjects are not completely elucidated yet.

My PhD programme had the objectives to (1) assess and validate computational methodologies for application in untargeted metabolomic studies of healthy humans and those diagnosed with glucocorticoid-related diseases and (2) to investigate the global and tissue-specific metabolic changes induced by glucocorticoid excess and deficiency and their integrated effects with the relevant hormone insulin.

A study of data pre-processing methods applied for untargeted metabolomics including different normalisation, missing value imputation, transformation and scaling methods were investigated on an *in-silico* modified dataset. The results showed that different combinations of data pre-processing methods influenced the results and different data pre-processing methods should be applied for univariate and multivariate analysis. Specifically, for a non-targeted metabolomics study I recommend the use of probabilistic quotient normalization (PQN) for univariate analysis, the use of normalization by PQN, random forest missing value imputation and glog transformation for principal component

analysis (PCA) and the permutation PQN, k-nearest neighbour (KNN) imputation and generalized log transformation for partial least squares discriminant analysis (PLS-DA).

Untargeted metabolomic studies of biofluids were applied to investigate *in-vivo* global effects of glucocorticoids. The study of the separate and integrated effects of cortisol and insulin showed that insulin may have negating effects on the influence of cortisol and treatment with cortisol should be timed appropriately during the day to minimize the effect of insulin on the therapeutic effect of cortisol. Furthermore, cortisol administration displayed an upregulating effect on ceramides and sphingolipids. Whilst exposure to cortisol is associated to an increase in glycerophospholipids, dexamethasone shows possible activation on β-oxidation. Separately, glucocorticoid-based replacement therapies or treatments (for example, 5-alpha reductase inhibitors) have been shown to invert the biological effects induced by insulin. The studies reported here have shown the influence of the interactions between glucocorticoids and insulin across the metabolic network.

Acknowledgments

I would like to thank my supervisor Dr. Warwick Dunn, my co-supervisor Dr. Mark Viant and the whole metabolomics research team for the support during the project.

I would like to thank the Medical Research Council and College of Life and Environmental Sciences at the University of Birmingham for sponsoring the present work. I would also like to thank the Phenome Centre Birmingham for assisting in some of the biological studies and providing the facilities.

Furthermore I would like to thank the Oxford Centre for Diabetes, Endocrinology & Metabolism and in particular Jeremy Tomlinson and Jonathan Hazlehurst for the collaboration for the experimental chapters.

Finally a special thank to my family and friends for their constant support and understanding.

Contents

Table of figures

List of tables

List of Abbreviations

11β-HSD = 11β-Hydroxysteroid Dehydrogenase
5-αR = 5-α Reductase
BCAA = Branched Chain Amino Acids
BMI = Body Mass Index
BPCA = Bayesian Principal Component Analysis
DIMS = Direct Infusion Mass Spectrometry
FFA = Free Fatty Acids
GCs = Glucocorticoids
GR = Glucocorticoid Receptor
GRUs = Glucocorticoids Response Units
HC = Hydrocortisone
HCD = Higher-energy collisional dissociation
HDL = High Density Lipoproteins
HILIC = Hydrophilic Interaction Liquid Chromatography
HPA = Hypothalamopituitary-Adrenal axis
ihs = Inverse hyperbolic sine
IRS = Insulin Receptor Substrate
KNN = k-Nearest Neighbours
MD = Median Replacement
MN = Mean Replacement
MS = Mass Spectrometry
MTBE = Methyl Tert-Butyl Ether
NMR = Nuclear Magnetic Resonance
PCA = Principal Component Analysis
PLS-DA = Partial Least Square – Discriminant Analysis
PQN = Probabilistic Quotient Normalisation
QC = Quality Control
RF = Random Forest
RP-LC = Reversed Phase – Liquid Chromatography
RSD = Relative Standard Deviation
S = Saline
SIM = Selected Ion Monitoring
SOP = Standard Operating Procedure
SV = Small Value Replacement
T2D = Type 2 Diabetes
TAG = Triacylglycerols
TIC = Total Ion Current
TOF = Time Of Flight
UPLC-MS = Ultra Performance Liquid Chromatography – Mass Spectrometry
VLDL = Very Low Density Lipoproteins

Publications

- Di Guida R., Engel J., Allwood J.W., Weber R.J., Jones M.R., Sommer U., Viant M.R., Dunn W.B.; **Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling.** *Metabolomics*, 2016, 12:93. Epub 2016 Apr 15. This publication obtained the runner up award for best paper in the journal *Metabolomics* for the year 2016.
- Hazlehurst J.M., Oprescu A.I., Nikolaou N., Di Guida R., Grinbergs A.E., Davies N.P., Flintham R.B., Armstrong M.J., Taylor A.E., Hughes B.A., Yu J., Hodson L., Dunn W.B., Tomlinson J.W.; **Dual-5α-Reductase Inhibition Promotes Hepatic Lipid Accumulation in Man.** *J. Clin. Endocrinol. Metab.,* 2016, 101(1):103-13.
- Lawson T.N., Weber R.J., Jones M.R., Chetwynd A.J., Rodrıguez-Blanco G., Di Guida R., Viant M.R., Dunn W.B. msPurity: **Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics**. *Anal. Chem.* 2017, 89(4):2432-2439.
- Di Guida R., Allwood J.W., Tomlinson J.W., Dunn W.B.; **Untargeted metabolomics study of the interaction of insulin and cortisol in healthy humans.** *In preparation.*

CHAPTER 1 - Introduction

1 Glucocorticoids

Glucocorticoids (GCs) are a class of steroid metabolites that regulate a large number of biochemical pathways including glucose, fatty acid and protein metabolism and a response to inflammation. GCs promote several anabolic or catabolic mechanisms such as the glycogenolytic actions of catecholamines and gluconeogenolytic actions of glucagon[1]. GCs also interact with insulin[2-4], however this interaction may be synergic or disruptive. The loss of the regulatory mechanisms of GCs plays important roles in a number of human diseases including type 2 diabetes (T2D)[5], obesity[6], Cushing's syndrome[7], adrenal insufficiency[8] and the metabolic syndrome[9]. In fact, these disorders have been proven to be partially related to an excess or deficiency of GCs. Thus, much research has focused on mechanistic studies in order to understand their action and modulation in a detailed way. GCs have been proven to have an effect on the differentiation of pre-adipocytes into mature adipocytes[10]. This mechanism is involved in the formation of adipose deposits and obesity which is also related to insulin resistance. One of the first effects of heavy GC administration (or abnormally high levels of endogenous GCs) is Cushing's syndrome. This condition may easily degenerate into type 2 diabetes which occurs as a consequence of a long period of insulin-resistance. In fact, GCs have been directly linked to the impairment of β–cells and the inhibition of insulin secretion[11]. One of the mechanisms which causes insulin sensitivity to decrease is the suppression of adiponectin by GCs. Adiponectin is secreted by adipocytes and has a role in insulin sensitisation. In 2012, GC administration was also related to the inhibition of osteocalcin, a protein which is produced in bones and suppresses adipose tissue formation[12]. It is also true that GCs have anti-inflammatory action;

therefore they are adopted in many therapies in order to treat immune-mediated diseases (e.g. asthma, arthritis, sepsis). As a consequence, the metabolism of GCs has been highlighted as an extremely important target to investigate in order to develop efficient therapies for adrenal insufficiency[13], asthma[13], rheumatic disorders[14] and organ transplantation[14].

1.1 Corticoids and their metabolism

Steroids are a subclass of lipids formed by four condensed cycloalkane rings (three cyclohexane and one cyclopentane ring). They are characterized by methyl groups on carbons C-10 and C-13. They also often bear an alkyl group on carbon C-17. Endogenous steroids in mammalian systems can be classified as follows: glucocorticoids, mineralocorticoids, androgens, estrogens, and progestogens[15]. Glucocorticoids are a class of steroid hormones present in mammals. GCs provide different functions in living organisms regulating physiological responses through the endocrine and metabolic systems.

Androgens are steroids which are involved in the development of male sexual traits and they act as precursors for the formation of estrogens. They are characterized by a keto or hydroxyl group at carbon C-17[16]. An androgen which has a main role in steroid biosynthesis is androstenedione which is a nodal metabolite for the formation of more potent androgens or estrogens via 17β-hydroxysteroid dehydrogenase and 17β-hydroxysteroid aromatase enzymes. Conversely, estrogens are mainly involved in the regulation of female biological mechanisms such as menstrual and estrous cycles through different mechanisms including effects on plasma fibrinogen levels[17] and modulation of brain activity[18]. Progestogens are a class of steroids which share some of the functions undertaken by estrogen, being involved in pregnancy and the menstrual cycle. Moreover, progestogens are synthesised directly from cholesterol, thus acting as the main precursors for all the other classes of steroids (**Figure 1**). Corticoids are generated in the adrenal cortex and are involved in various biological processes concerning immune function, inflammation, protein synthesis and metabolism. They can be subdivided in to glucocorticoids and mineralocorticoids. The former, as stated before,

have many functions including regulation of glucose, protein and lipid metabolism as well as anti-inflammatory effects through up-regulation of specific proteins in the cell nucleus (these processes are mediated by GC receptors (GR)). Cortisol can be considered as the most important GC. It is involved in a plethora of metabolic, immunological and cardiovascular functions. For example, cortisol has been proven to regulate carbohydrate metabolism including hepatic glucose production independent of insulin concentration[19]. Immune response has been shown to be modulated by cortisol as well, displaying a weakening trend on inflammatory responses through inactivation of interleukin-2 (which produce T-cells) after cortisol administration[20]. Moreover, control of cardiovascular functions by cortisol is achieved through regulation of salt concentration (through sodium uptake) and water levels influencing blood volume and pressure[21].

It is significant that endogenous steroids do not represent unlinked biological classes, indeed progestogens, androgens, estrogens and corticoids are interconnected through articulated metabolic processes. All the metabolic reactions that modify the structure of one specific steroid into another are regulated by different enzymes, including cytochrome P450, 11β-hydroxysteroid dehydrogenase (11β-HSD) and 5-α reductase. A number of enzymes are involved in multiple metabolic reactions. The main precursor is cholesterol, which is consumed through diet and transported by low density lipoproteins, high density lipoproteins and chylomicrons; however it can also be synthesized endogenously. Cholesterol can be converted into progestogens which in turn can be transformed in to mineralocorticoids, glucocorticoids or androgens. The latter may undergo structural modifications through P450 intervention and convert into estrogens[22].

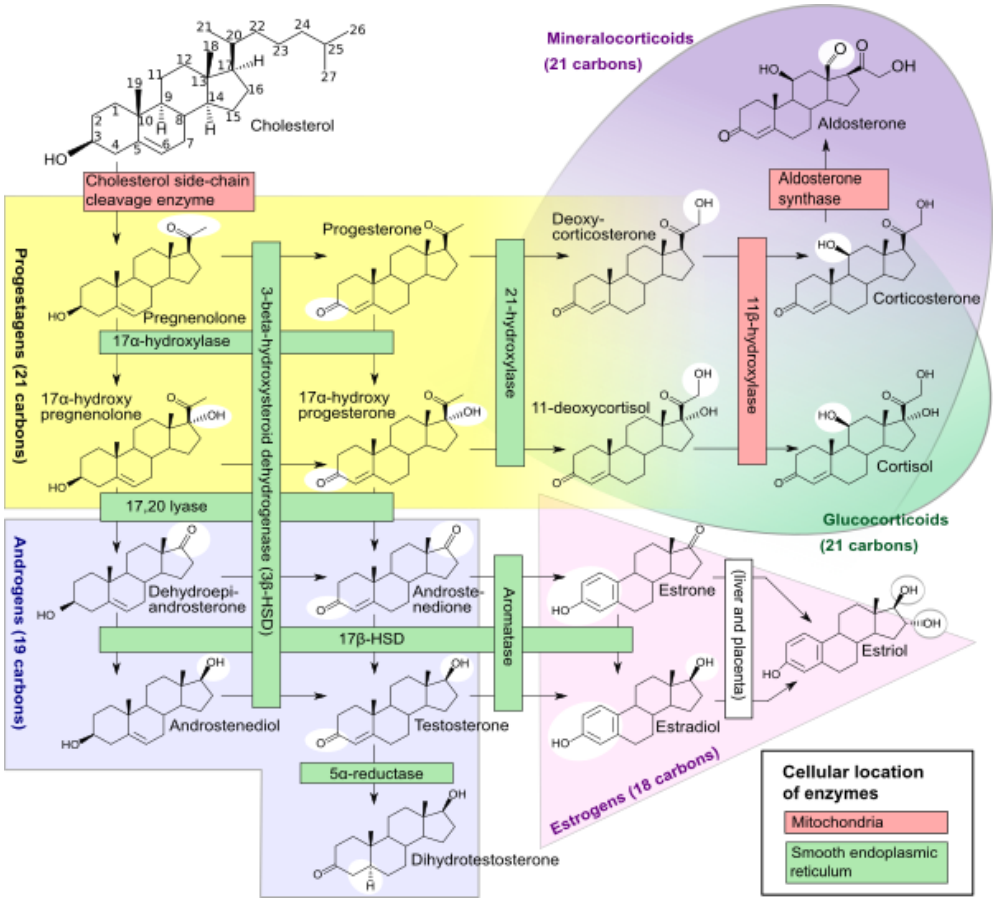


Figure 1. Overview of steroid metabolic pathways[23]. Progestogens are highlighted in yellow, androgens in blue, estrogens in pink, glucocorticoids in green and mineralocorticoids in violet.

1.2 Glucocorticoids, human health and disease

Much research has been carried out on GCs because of their anti-inflammatory properties including in order to synthesise GCs characterized by enhanced anti-inflammatory action compared to cortisol and suitable for the drug market. Common synthetic GCs include prednisone, prednisolone, dexamethasone, mometasone and beclomethasone. However, naturally produced GCs include cortisone, 11-deoxycortisol, corticosterone and cortisol. (**Figure 2**)

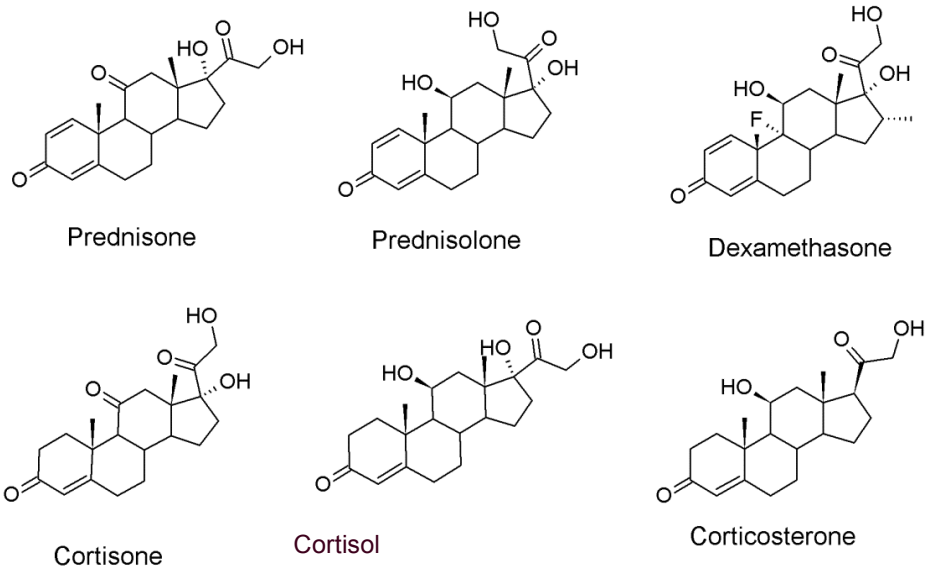


Figure 2. Chemical structure of three synthetic and three endogenous glucocorticoids[24]

As already stated, cortisol (hydrocortisone) has an important role due to its association with a variety of biological mechanisms including gluconeogenesis, immune system suppression and regulation of cardiovascular activity. The secretion of cortisol is regulated by the hypothalamopituitary-adrenal (HPA) axis[25]. Cortisol can inhibit its own release through negative feedback and is cleared through different metabolic pathways, including reduction and hydroxylation. The biological activity of GCs is a result of the presence of the hydroxyl group at the C-11 position. Indeed the inactivation of cortisol is caused by the oxidation of the hydroxyl group to the oxo form to synthesise cortisone[26].

GC concentrations are regulated through different processes: synthesis, secretion, association to globulin and negative-feedback. The free circulating fraction of GCs enter the cell through passive diffusion across phospholipid membranes. In the cytosol, there are glucocorticoid receptors (GR) which are bound to heat shock proteins and display two isoforms: α and β. Only the former binds tightly to the GCs, the latter can act as a negative regulator. Once the GCs enter the cytosol, the GR unbind proteins, bind the GCs and transport the GCs to the nucleus. Here, the interaction with

glucocorticoid response elements (GREs) regulates the activation or inactivation of specific genes through cofactors[25,27] including cAMP response element binding protein and nuclear factor-kappa B (NF-kB)[25]. The main effects obtained in this step are the up-regulation of genes involved in the synthesis of anti-inflammatory proteins such as lipocortin-1 and suppression of pro-inflammatory enzymes (e.g. phospholipase A2)[28]. Phospholipase A2 has an essential role in the synthesis of eicosanoids (the precursors of prostaglandins), a process that directly results in the inflammatory response. For this reason, GCs are largely applied in therapies as anti-inflammatory agents. Apart from cortisol (medically prescribed as hydrocortisone), other GCs with a higher activity than cortisol have been synthesized in order to provide improved efficacies; for example dexamethasone activity is eighty times greater than cortisol[29] and is prescribed routinely.

In order to take advantage of the anti-inflammatory action of GCs without incurring metabolic side effects, research has focused on the development of SElective GR Agonists (SEGRAs). These compounds trigger transcriptional activity that is selectively exerted on anti-inflammatory proteins while they have no effect on other transcriptional modifications responsible for global protein synthesis, glucose uptake, glucose transport and insulin secretion[28,30].

GCs are also administered as immunosuppressants; they inhibit two transcriptional factors which regulate the expression of mediators and proteins with major roles in the immune system: NF-kB and Activator Protein 1 (AP-1)[31]. T cell differentiation and further activation seems to be heavily affected by the presence of GCs. In fact, GR-KO mice show enhanced activation of T cells which leads to higher rates of mortality[32]. The physiological anti-inflammatory effect of GCs is not exercised only through inactivation of NF-kB and AP-1 transcriptional factors, in fact cytokine IL-10 can be induced by GCs. IL-10 is involved in strong anti-inflammatory responses and its absence has been linked with the incidence of autoimmune diseases[32]. GCs also inhibit the cytokine IL-2, involved in cell-mediated immune response[33]. Furthermore, patients with diseases that entail a lack of endogenous GCs such

as Addison's disease or Waterhouse-Friderichsen syndrome, may be administered with exogenous GCs in order to rebalance steroid homeostasis. GCs have also been linked to suppression of endoplasmic reticulum stress through activation of protein folding and up-regulation of genes controlling the degradation of misfolded proteins. Das *et al.*[34] showed that dexamethasone is able to prevent the misfolding of the protein Muc2 mucin and activate genes involved in the degradation of misfolded proteins in the endoplasmic reticulum, thus recovering from endoplasmic reticulum stress.

The conversion of cortisol to cortisone and vice versa is essential in biological systems, as the former is active and binds to the GR while the latter is inactive. The enzyme 11β-HSD accomplishes this task thanks to its two isoforms (11β-HSD-1 and 11β-HSD-2). The enzymes operate in different directions, 11β-HSD2 converts cortisol to cortisone while 11β-HSD1 catalyzes the conversion of cortisone to cortisol. In fact, 11β-HSD1 is a bidirectional enzyme, though the conversion of cortisone to cortisol is dominant in almost all tissues (**Figure 3**).

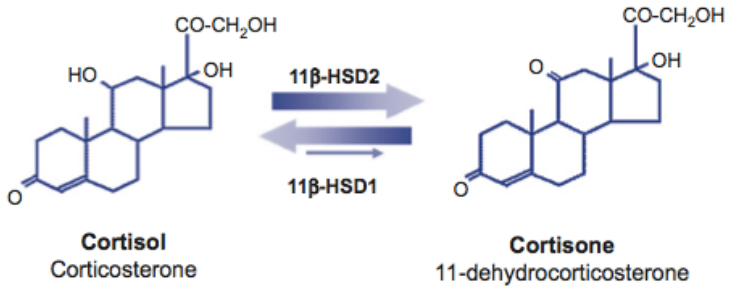


Figure 3. Conversion of cortisol in to cortisone and vice versa regulated by different isoforms of 11β-HSD[25]

Activity of isoform I is regulated by NADPH while isoform II is regulated by NADH[25]. 11β-HSD2 is mainly expressed in the kidneys, colon, salivary and sweat glands (all tissues characterized by high concentrations of aldosterone in physiological conditions) and its absence can cause hypertension as cortisol can associate and activate mineralocorticoid receptors. Indeed, their natural ligand is aldosterone but cortisol has a higher binding affinity. 11β-HSD1 is located in the liver, adipose tissue,

gonads, heart and brain (in the endoplasmatic reticulum)[35]. The activity of 11β-HSD1 can be regulated by a number of factors. For example, GCs, peroxisome proliferator-activated receptor γ agonists and pro-inflammatory cytokines enhance the enzyme activity whereas growth hormones and liver X receptor agonists suppress it[35]. Also sex steroids regulate the expression of 11β-HSD1, however their effect appears to be tissue-specific[36]. For example studies showed higher concentrations of 11β-HSD1 in adipose tissue in women compared to men[37] and suppression of 11β-HSD1 in rat liver and kidney promoted by estradiol[38].

Due to the wide range of effects GCs achieve, the metabolic response is substantial. The identification of diseases regulated by GC concentration has driven research towards the investigation of GC metabolism and the effect of GCs on global metabolism. One of the first human models to investigate GC metabolism was offered by Cushing's syndrome; subjects display high levels of GCs and develop specific symptoms including immunosuppression, myopathy, hypertension, osteoporosis and hepatic steatosis. In the 1930s Cushing[39] characterized this disease and started the investigation of GC metabolism. Long, in two separate works, stated the relationship between GCs and diabetes[40] and gluconeogenesis associated with insulin resistance[41]. During fasting, GCs increase circulating glucose and free fatty acids (FFA) through gluconeogenesis and lipolysis while in the fed state they promote glucose and lipid storage as glycogen and triacylglycerides (TAGs), respectively. Cushing's syndrome is often associated with metabolic disorders leading to type 2 diabetes. Another crucial metabolic disorder is the metabolic syndrome, which is characterized by increased abdominal circumference and high levels of TAGs and glucose in the fasting state. Subjects affected by the metabolic syndrome are likely to develop diseases such as diabetes and cardiovascular diseases. Whilst the metabolic syndrome does not present higher levels of circulating GCs in blood, experiments have shown increased metabolic activity involving GCs. Negative feedback sensitivity decreases (production of endogenous GCs is not regulated anymore by its own release leading to

possible GC concentration imbalances), urinary GC metabolites are present in higher concentrations and 11β-HSD1 activity decreases, suggesting a decrease in cortisol concentration in the hepatic region in order to retain insulin sensitivity.

1.3 The role of glucocorticoids in endogenous metabolism

GCs play important roles in lipid, protein and carbohydrate metabolism. Their role in different metabolic pathways will be discussed here.

1.3.1 Glucocorticoids and gluconeogenesis

Gluconeogenesis is a metabolic pathway which synthesises glucose starting from metabolites which are not carbohydrates[42]. It occurs only occurs in the liver (**Figure 4**). It generates glucose from pyruvate and other precursors through the same metabolic steps observed in glycolysis (which generates pyruvate from glucose). However, when compared to glycolysis the direction of the metabolic reaction of gluconeogenesis is inverted. This is possible thanks to the double functionality of the enzyme phosphofructokinase 2 (PFK2)/fructose bisphosphatase 2 (FBPase2)[43]. During glycolysis mainly PFK2 activates phosphofructokinase 1, an enzyme with glycolytic functions. When glucose levels are low (during starvation or fasting) PFK2 functions are suppressed through glucagon interaction. Importantly, other metabolites such as amino acids and glycerol may undergo several biochemical reactions before entering the gluconeogenetic pathways.

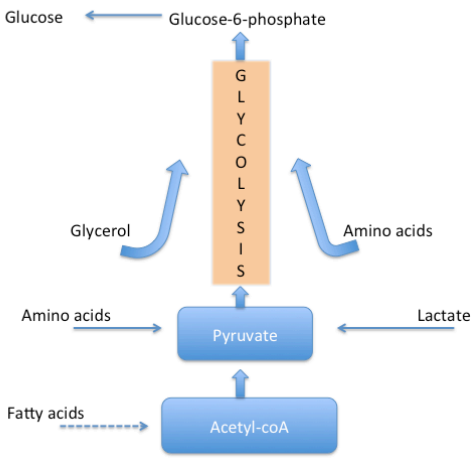


Figure 4. Schematic representation of the gluconeogenesis metabolic pathway.

GCs induce gluconeogenesis in the liver while suppressing glucose uptake in white adipose tissue and skeletal muscle. Glucose is the primary metabolic substrate for energy production in the brain and GCs are essential for maintaining high glucose levels in the brain in the fasted state and during starvation. The contribution of GCs to gluconeogenesis also involves other mechanisms: including (1) protein breakdown in skeletal muscle which provides amino acids that can be converted in to oxaloacetate and enter the pathway[44] and (2) lipolysis in white adipose tissue which produces glycerol and fatty acids[45] (which provide energy for gluconeogenesis).

The effects of GCs on gluconeogenesis are also exerted through interaction with pancreatic islets. The pancreas is able to synthesise three hormones that heavily influence glucose metabolism. In the fed state, it releases insulin from β-cells that is involved in cellular glucose uptake, glucose oxidation and inhibition of gluconeogenesis. Glucagon instead is produced in α-cells during fasting; it enhances gluconeogenesis and glycogenolysis, the conversion of glycogen to glucose. Somatostatin is secreted in δ-cells and can suppress the effects of insulin and glucagon[46]. While the interaction between GCs and β-cells (and the consequent effects on insulin secretion) are well known, it is also true that GCs are involved in α-cell mediated glucagon secretion. Indeed glucagon levels are much higher in mice

treated with GCs compared to control mice. Also α-cell size is increased and high levels of glucose fail to suppress glucagon secretion[46] at high levels of GCs. As already stated, glucagon promotes gluconeogenesis and glycogenolysis, hence chronic overexpression of glucagon eventually leads to hyperglycemia[47]. It must be noted that after administration of GCs hepatic tissues do not display any change from the physiological ratio of insulin/glucagon that are measured in the fasted and fed state.

GCs also inhibit glucose uptake and promote glycogen degradation mediated by catecholamine. Further effects on glucose are exerted on a homeostatic level since GCs favour the transfer of lipids from white adipose tissue to skeletal muscle[44].

1.3.2 Glucocorticoids and glycogen metabolism

Interaction between glycogen metabolism and GCs is highly tissue-specific. Fasted mice, for example, showed glycogen accumulation in the liver after GC administration[48]. This mechanism operates through activation of glycogen synthase by GCs[49]. Activity of glycogen synthase depends on its phosphorylation. In fact some enzymes have an effect on its activity acting through phosphorylation or dephosphorylation. Protein kinase A (PKA) and glycogen synthase kinase 3 inactivate glycogen synthase through phosphorylation[50], while protein phosphatase 1 (PP1) enhances glycogen synthase activity by removing the phosphate group from its structure[51]. This action is contrasted by glycogen phosphorylase *a* through two mechanisms. On one hand it inhibits the dephosphorylation of glycogen synthase operated by PP1 and on the other hand it breaks down glycogen in to glucose. Studies[52,53] have reported a correlation between GC levels and overexpression of PP1 although the mechanisms underlying such outcomes are still unclear.

Skeletal muscle displays a different phenotype. Indeed epinephrine, a hormone that is present in skeletal muscle is able to change PP1 activity. Adrenalectomised rats with a lack of endogenous GCs display increased PP1 activity and impaired suppression operated by epinephrine[54]. Administration of cortisol in these rats restored the original inhibiting effect of epinephrine ultimately suppressing

glycogen synthesis. Inhibition of glycogen synthesis in skeletal muscle by GCs is caused also by interference with insulin-mediated production of glycogen[55]. Notably, one study[55] reported an increase of glycogen storage despite decreased glycogen synthase activity.

Cardiac muscle displays a decrease in epinephrine-activated glycogenolysis after adrenalectomy[56]. However, administration of GCs enhances glycogen storage in this case. It seems that this effect is obtained through GC-induced overexpression of AMP-activated protein kinase which enhances glucose uptake and glycogen synthesis[57]. The overall effect is an increased level of glycogen in cardiac tissue after administration of GCs.

1.3.3 Glucocorticoids and lipid metabolism

The chemical structure of multiple classes of lipids is usually characterised by a single or multiple long carbon chain and a more polar head group. Triglycerides react with lipoprotein lipase[58] that breaks them down in to free fatty acids (FFA). These FFA can reach the cell through specific transport proteins[59]. Conversely cholesterol esters are transported to the cell through an articulate pathway involving low-density proteins and receptor signalling[60].

Triglycerides and FFA can derive from *de novo* lipogenesis, from residuals of chylomicrons and from lipolysis in adipocytes (FFA only)[61]. On the other hand, cholesterol can be obtained from lipoprotein particles that have been absorbed by the liver, from *de novo* synthesis of cholesterol and from diet[62].

Once lipids reach the cell they can enter different metabolic pathways. Triglycerides can be stored in lipoproteins in liver or adipose tissue in fat depots[60]. Fatty acids can be converted in to triglycerides or used as a precursor for β-oxidation. β-oxidation takes place in the mitochondria. Since fatty acids are negatively charged, they need prior transformation in order to traverse the mitochondrial membrane. This step is operated through carnitine acyltransferase that synthesise acyl-carnitines (and acetyl-CoA) from carnitine and acyl-CoA. Acyl-carnitines can then be transported across the

membrane. Once inside the mitochondria, the fatty acids can be oxidised and provide substrates to the citric acid cycle. Moreover fatty acids may also act as building blocks for sphingolipids, glycerophospholipids and eicosanoids[60].

GCs and regulatory hormones can interact with metabolic pathways by activating or suppressing enzyme activity[63]. GCs have a large impact on regulation of lipid metabolism and as we will explain further, they can also lead to diseases when this regulation is out of control (i.e. GCs excess). Notable examples are Cushing's syndrome and type II diabetes[64].

1.3.3.1 Glucocorticoids and lipolysis

Adipose tissue contains triacylglycerides (TAGs) that can be broken down via lipolysis in order to release fatty acids during periods of fasting. TAGs are converted to diacylglycerides (DAGs) and a fatty acid by adipose triglyceride lipase (ATGL). DAGs are then converted to monoacylglycerides (MAGs) and a fatty acid by hormone-sensitive lipase (HSL). Finally MGs are converted to glycerol and a fatty acid by monoglyceride lipase[60]. GCs are known to enhance the activity of the enzymes that catalyse lipolysis in peripheral adipose depots through interaction with gene transcription of ATGL and HSL mediated by GR. Moreover, GCs suppress the action of lipoprotein lipase in peripheral adipose depots[65,66]. However this is not true for the abdominal region where GCs have been linked to increased adipogenic activity[67].

Importantly, a study involving humans failed to detect a clear correlation between GC exposure and lipolysis; indeed individuals affected by Cushing's syndrome or administered with high levels of GCs did not show any significant change in turnover of non-esterified fatty acids[68]. Notably, another study showed no effect on β-oxidation after GC administration in the Chub-S7 adipose cell line[2]. At a biochemical level, GC administration may promote lipolysis in peripheral adipose tissue through inactivation of lipogenesis and fatty acid re-esterification with glycerol to TAGs, leading to an increase of circulating free fatty acids (FFA)[2,69]. Surprisingly, when GCs are administrated with

insulin[2,70] the opposite effect has been reported with an intensification of lipid storage through activation of lipogenesis[70].

1.3.3.2 Glucocorticoids and lipogenesis

Lipogenesis is the metabolic process that synthesises complex lipids (e.g. triglycerides). This process starts with the reaction of acetyl-CoA (which can be obtained from conversion of glucose in to pyruvate) via acetyl-CoA carboxylase (ACC) to synthesise malonyl-CoA. Malonyl-CoA is then converted into fatty acids of increasing chain length through the action of fatty acid synthase (FASN). At this point the fatty acids react with glycerol-3-phosphate to synthesise triglycerides. This process is triggered by insulin in the fed state or when the amount of nutrient intake is excessive. In this case GCs have a synergistic effect with insulin[71]; indeed lipogenesis is enhanced after GC administration and suppressed in adrenalectomised rats with high insulin levels[72]. GCs activate ACC and FASN at the genetic and transcriptional level[2,73] and genetic regions binding GRs have been located close to *loci* expressing these two enzymes[74].

GC-induced lipogenesis is displayed mostly in specific areas of the human body including the face and trunk. Lipid storage in such areas showed a significant reduction after adrenalectomy in rats but total restoration after GC replacement therapy[75]. GCs also appear to *promote* fatty acid β-oxidation in adipose tissue. The mechanism is mediated by Tribbles-homologue 3 (TRIB3), a protein which is produced during fasting periods and which has been correlated with high rates of fatty acid oxidation[76]. TRIB3 has been reported to activate degradation of ACC thus presenting a pathway which opposes the lipogenic effect of GCs previously described.

Apart from fatty acid uptake, the liver can act to provide distribution of lipids to peripheral tissues. This task is accomplished through very-low density lipoproteins (VLDL), which are regulated by various factors[77]. Dexamethasone for example, increases the availability of triglycerides transported via VLDL and the concentration of high density lipoproteins (HDL) in plasma[78]. It is possible that GC-

induced stimulation of lipogenesis (and thus increased synthesis of triglycerides) triggers the secretion of VLDL[79]. This effect seems to be assisted by an imbalance of lipoprotein lipase activity in dexamethasone-treated rats. Cushing's syndrome is characterised by increased secretion of VLDL that is not counterbalanced by enhanced peripheral clearance[80]. GCs appear to suppress the degradation of triglycerides derived from lipoproteins[81]. In fact, high levels of TAGs were reported in individuals affected by the metabolic syndrome with enhanced activity of the HPA[81]. It has been suggested that GCs may act on lipoprotein clearance through a direct interaction with receptors rather than with the TAGs.

Muscle tissue also displays lipogenesis. Chickens undergoing dexamethasone administration displayed fat accumulation on muscle tissue due to inhibition of AMPK[82]. However this effect is in contrast with *in vitro* studies. Indeed C2C12 myotubes have displayed the inhibition of genes expressing proteins involved in the lipogenic pathway after dexamethasone administration[83]. Furthermore GCs have been linked to increased β-oxidation of non-esterified fatty acids, which ultimately led to insulin resistance[84]. This outcome triggers a hyperinsulinemic response that occurs in conjunction with high levels of circulating fatty acids.

However, no study has confirmed a lipogenic mechanism operating in muscle tissue that is triggered by GCs.

A notable effect of GC administration is adipocyte differentiation. Apparently GCs can sensitise pre-adipocytes to insulin which enhances the actual production of adipose tissue. This is accomplished through activation of transcriptional adipogenic factors[85]. However it must be noted that different adipose cells respond differently to high levels of GCs. Visceral fat for example shows an increase in adipose cells while subcutaneous fat displays a marked decrease[86]. The effect on insulin resistance seems to follow the same pattern. Indeed insulin resistance is displayed in visceral fat adipose tissue

only[87]. This result has been linked to overexpression of glucocorticoid receptor in visceral fat adipose cells only.

1.3.4 Glucocorticoids: further metabolic considerations

Apart from the discussed effects on lipid and carbohydrate metabolism, other processes are heavily influenced by the presence of GCs, including glucose uptake and protein synthesis. Both of these pathways seem to be inhibited by a high concentration of GCs and both are regulated by a signal cascade that involves insulin/IGF1 receptors, PI3 kinase and Akt kinase. Their role is to promote glucose uptake (through delivery of GLUT-4 to the cell surface[88]) and protein synthesis. Separate studies proved the ability for GCs to decrease the phosphorylation rate of insulin/IGF-1 receptors[89] and reduce the activity of PI3[4] kinase and Akt kinase[90] in skeletal muscle resulting in impaired glucose uptake and protein synthesis.

Other considerations are required when investigating protein synthesis. Apart from Akt kinase signaling, GC exposure decreases the levels of free amino acids in musculoskeletal cells through uptake inhibition[91]. Protein degradation is enhanced after external administration of GCs. Corticosterone treatment administered in high doses caused higher rates of protein degradation compared to lower doses[92]. Another study, conducted on normal and adrenalectomised rats showed reduced proteolysis in adrenalectomized rats, due to the lack of endogenous GCs[93]. This effect is hard to achieve in the fed state since high levels of circulating insulin contrasts the effect on administered GCs[93].

1.4 Tissue-specific expression of 11β-HSD1

Currently, research on GC metabolism and 11β-HSD1 activity is quite controversial despite the large number of papers published. Here an overview of tissue-specific GC metabolism is presented, highlighting the acknowledged metabolic steps.

The highest concentration of the enzyme 11β-HSD1 has been observed in the liver (ten-fold higher than in adipose tissue). Hepatic lipid accumulation is a marker for metabolic syndrome and insulin resistance although such an effect is not always linked with an over-expression of 11β-HSD1. Indeed while some studies have reported an increased 11β-HSD1 concentration in liver for patients with metabolic disease[94], other studies did not report a significant difference in hepatic expression of 11β-HSD1 between healthy patients and patients affected by non-alcoholic fatty liver disease which is characterised by hepatic lipid accumulation[95]. Indeed fat accumulation imbalances insulin signaling affecting protein kinase C (PKC) and jun N-terminal kinase 1 (JNK1) pathways with a consequent decrease of phosphorylation of insulin receptor substrate 1 (IRS1)[96]. Comparisons between healthy patients and patients affected by the metabolic syndrome reported different results, sometimes a significant difference in concentrations of 11β-HSD1 and H6PDH in the liver were observed[94] and sometimes no difference was observed[95]. It has been hypothesized that changes in enzyme concentrations can occur during disease progression[26]. Overexpression of 11β-HSD1 causes hepatic steatosis, dyslipidemia and hypertension although the effects on insulin resistance are limited[97]. It has been suggested that overexpression of 11β-HSD1 in adipose tissue may be sufficient in itself to provide hepatic steatosis with the transport of cortisol to the liver possible. Activity of hepatic 11β-HSD1 has been assessed through measurement of urinary steroid metabolites and active cortisol circulating in serum[98]. This experiment highlighted a drop in activity of the enzyme in obese subjects.

11β-HSD1 is present in high concentrations in adipose tissue. While the enzyme itself does not show any particular difference in concentration between visceral and subcutaneous adipose tissue[37,99], other factors such as H6PD and GR appear to be differently expressed in various adipose tissue regions[100]. Pro-inflammatory cytokines appear to have a clear effect on the activity of 11β-HSD1 in adipose tissue[35].

The obese phenotype is unexpectedly related to low levels of circulating GCs suggesting that 11β-HSD1 plays an essential role. In fact, an overexpression of this dehydrogenase would decrease the circulating GC[101] and, at the same time, would increase GC activity locally in adipose tissue probably due to intra-adipose metabolism[26,101]. As a result, 11β-HSD1 overexpression and high levels of GC in adipose tissue induce the same phenotype of abdominal obesity, glucose intolerance and hyperglycemia.

Whilst 11β-HSD1 is present in muscle in lower concentrations compared to the liver and adipose tissue, GCs have been found to regulate metabolism in the musculoskeletal system. Patients affected by T2D showed increased levels of 11β-HSD1 in myotubes. It is acknowledged that GCs induce insulin resistance in muscular tissues. A study pointed out that phosphorylation of IRS1 was a possible cause for insulin resistance and impaired glucose uptake in muscle[102]. Moreover, high exposure to GCs as well as Cushing's syndrome has been proven to induce muscle atrophy[103].

Tissue	Effects of high levels of GCs
Liver	Glycogen synthesis ↑[48] Gluconeogenesis ↑[46]
Adipose tissue	Glucose uptake↓ (prevented glycolysis)[46]
Adipose tissue (abdominal)	Adipocyte proliferation ↑[104] Lipogenesis ↑ (cooperation with insulin) [75]
Adipose tissue (peripheral)	Lipolysis ↑[65]
Skeletal muscle	Glucose uptake ↓ (prevented glycolysis) [46] Protein breakdown ↑[44] Protein synthesis ↓[89, 90] Glycogen synthesis ↓[54]
Pancreas	Glucagon levels ↑[1]

Table 1. Summary of the metabolic effects of GCs on different tissues

The PhD research programme I undertook had an objective to assess the effects of glucocorticoid deficiency and excess on tissue-specific and global metabolism and to understand the mechanisms underlying those processes. In fact, many contrasting opinions still characterise the investigation of GC mechanisms on global and tissue-specific metabolism. The experiments performed had the objective to elucidate the effect of GC administration (at low and high insulin levels) on the human metabolome with tissue-specific considerations. In order to investigate the endogenous metabolic perturbations induced by glucocorticoid administration, non-targeted ultra performance liquid chromatography – mass spectrometry (UPLC-MS) and direct infusion mass spectrometry (DIMS) metabolomics were performed as described in chapters 4 and 5.

1.5 Interaction of glucocorticoids and insulin

While GCs inhibit glucose uptake and oxidation in skeletal muscle and white adipose tissue, insulin stimulates glucose uptake and oxidation[105]. Administration of GCs has shown suppression of insulin-induced glucose uptake through inactivation of the transport of GLUT4 to the cell membrane[106]. Conversely low levels of GCs have been shown to enhance the effect of insulin on glucose uptake in skeletal muscle (both in humans[46] and rats[107]). Mice fed on a high-fat diet displayed a recovery in insulin secretion after adrenalectomy[108]. The same outcome has been achieved through inhibition of the enzyme 11-βHSD1[107]. It seems that the effects of GCs on skeletal muscle are exerted directly on myotubes. As already mentioned insulin acts through binding to insulin receptor tyrosine kinase (IR) on the cell surface. IR then phosphorylates the insulin receptor substrate (IRS1)[109]. Hence, IRS and IRS1 trigger a series of signals that also involve phosphoinositide-3-kinase (PI3K) and Akt. Administration of GCs in mice decreased the phosphorylation levels for IR and the activity of PI3K and Akt[4,107]. This inhibition has been studied in mice C2C12 myotubes[110] and allowed the identification of the target gene Pik3r1 which also involves the synthesis of proteins regulating the insulin signal downstream.

Insulin release from β-cells in the pancreas is regulated by glucose in an eight step mechanism involving calcium uptake as a signal for insulin vesicle exocytosis. Different studies involving GCs and islet cells have been performed, revealing conflicting outcomes. GCs have been demonstrated to increase[111] and decrease[112] insulin secretion in pancreatic cells. What is for sure is that mice overexpressing GR in β-cells show reduced insulin secretion associated with hyperglycemia[113].

Due to the strong interaction between GCs and insulin it is really challenging to discern effects of GCs on pancreatic tissues from effects of GCs on insulin-mediated mechanisms. While GC-induced insulin resistance can trigger a response by β-cells which enhance insulin production in order to tackle the high levels of glucose, prolonged periods of insulin resistance cannot be counterbalanced by hyperplasia and hyperinsulinemia in β-cells, eventually leading to diseases such as Cushing's

syndrome or type 2 diabetes[114,115]. Hence the effect of GCs on β-cells is dependent on the duration of the treatment and may be reversible after short-term acute exposure[116]. β-cell activity is a direct consequence of cellular respiration as it is dependent on the glucose concentration and ADP/ATP ratios.

In vivo and *in vitro* studies on pancreatic islets with GC administration have displayed contrasting outcomes. *In vitro* reports on pancreatic islets exposed to high levels of dexamethasone show degradation of the GLUT2 transporter and impairment of insulin secretion triggered by glucose[117]. In specific cell lines dexamethasone also appears to suppress glucokinase transcription factors[118]. It must be noted that GCs have been reported to directly affect the pancreas at both the genetic[119] and transcriptional[120] levels. Indeed *in vitro* GCs administration in pancreatic islets inhibited genes associated with insulin production and transcriptional factors responsible for the actual generation and growth of β-cells. As already mentioned, these results contrast with the *in vivo* studies. Indeed pancreatic islets obtained from mice that underwent heavy GC administration displayed increased glucose uptake, improved insulin secretion and enhanced oxidative metabolism[121,122]. This outcome is likely to be caused by β-cell response after short-term high GC exposure. As stated before, GCs can stimulate β-cells proliferation. However, some cell lines (e.g. MIN6 and Langerhans) showed the opposite effect[123]. It must be noted though that presence of cAMP in high levels reduced cytotoxicity provoked by GCs[124,125]. The impairment of insulin sensitivity has been ascribed to GC-induced down-regulation of IRS1 and IRS2[126] and inhibition of GLUT-4 insulin-induced glucose delivery to the plasma membrane[127,128].

GCs act on glucose uptake and insulin sensitivity also through mediation of some lipid species. An increase in ceramides for example has been displayed in rats both in the liver and bloodstream after being exposed to GCs[129]. High levels of ceramides have been linked to insulin resistance[130]. This effect has been highlighted in studies with hyperinsulinemic mice. While GC administration alone inhibited

glucose uptake and suppressed the insulin inhibition on glucose output in skeletal muscle, myriocin (which is a known inhibitor of serine palmitoyltransferase, involved in the metabolic pathway of ceramides) administered in conjunction with GCs markedly suppressed the effects of GCs on insulin[129]. Furthermore, the presence of GCs has been linked with overexpression of enzymes directly involved in ceramide metabolism. It is possible that ceramides and other lipid mediators linked to insulin resistance (e.g. diacylglycerols) may be transported to skeletal muscle as a consequence of GC-induced lipolysis in white adipose tissue [131].

Glucose metabolism in other tissues is affected by administration of GCs. Mouse adipocytes for example showed severe impairment of insulin-mediated glucose uptake after GC administration[132]. It must be noted that human adipocytes have also been adopted for studies on the effects of GCs on glucose metabolism leading to an interesting differentiation. Omental adipocytes displayed the typical outcome of inhibited glucose uptake in conjunction with high levels of GCs, subcutaneous adipocytes showed a synergistic effect between GCs and insulin[87]. In fact GC administration enhanced insulin-induced glucose uptake in those tissues. Brain glucose levels (especially in the hypothalamus and hippocampus) appear to be regulated by the level of circulating GCs[133]. Again impaired glucose oxidation and uptake appear to be inhibited by high levels of GCs.

The effect of GC administration on impaired insulin secretion acts through β-cell dysfunction and down-regulation of the glucose type 2 transporter and glucokinase, which ultimately reduces ATP synthesis leading to imbalanced insulin secretion from β-cells. Acute effects of GC administration include inhibition of GLP1[134] which contributes to β-cell physiological function (i.e. insulin secretion). In the long term, hypercortisolism can lead to hyperinsulinaemia[116] (β-cells) and hyperglucagonaemia[135] (α-cells).

To summarise, GC administration (along with over-expression of 11-βHSD1) has been shown to inhibit insulin secretion[11] and to possibly induce diseases like Cushing's syndrome and T2D as a result

of a long period of insulin resistance. While some studies have shown opposite outcomes ascribing insulin sensitivity[87] or insulin resistance[127] to GC administration, it seems to be certain that the effect of GCs on insulin-regulated mechanisms is controlled through down-regulation of IRS1 and IRS2[126]. Moreover, insulin sensitivity has been more directly linked to expression of 11-βHSD1, which converts cortisone into cortisol. In fact this enzyme has been found to be severely unbalanced in a variety of diseases (Cushing's syndrome, metabolic syndrome and T2D), often causing secondary effects on insulin signalling and glucose intolerance[97].

Some of these diseases (e.g. T2D) have a large impact on modern health policies since they affect a significant percentage of the world's population. It is therefore beneficial to provide an overview on the mechanisms underlying such syndromes.

1.5.1 Cushing's syndrome

Cushing's syndrome encompasses a series of diseases that are all characterised by abnormally high levels of cortisol. Such levels may be reached through different pathways and be caused by different mechanisms. The expected phenotypes include weight gain, hypertension, polyuria and insulin resistance[136]. Notably, insulin resistance can be cured in the early stages of Cushing's syndrome but may later develop (20-47% of cases[86]) into T2D.

A general mechanism underlies this group of diseases although slight differences may arise depending on the specific type of disease. In physiological condition the hypothalamus produces corticotropin-releasing hormone (CRH) which in turn leads to the release of corticotropin (ACTH) from the pituitary gland[137]. ACTH then reaches the adrenal gland through the blood stream regulating the production of cortisol. A negative feedback allows the suppression of ACTH production when cortisol levels are too high[138]. Cushing's syndrome, in the broadest sense entails an impairment of this mechanism.

As already mentioned, different causes can eventually lead to the syndrome. However such causes can be eventually ascribed to exogenous or endogenous reasons. Exogenous causes are the most common[138]. In this case, high cortisol levels are reached due to cortisol prescription to patients in order to treat other diseases. If Cushing's syndrome is identified, it can be treated simply by suspending cortisol treatment. Conversely, endogenous causes are less common and can be ascribed to different mechanisms. Pituitary Cushing's is characterised by an increased production of ACTH due to the presence of a benign pituitary adenoma[136]. High levels of cortisol are also observed in adrenal Cushing's although in this case this is due to endocrine impairment of the adrenal gland (e.g. by a tumour). When tumours grow outside of the adrenal gland it is still possible to detect increased levels of ACTH, in this case the disease is known as ectopic Cushing's disease.

1.6 The study of human non-steroidal metabolism

In recent years the holistic study of metabolites in biological systems, defined as metabolomics, has been applied as an important biological research tool. Metabolites are low molecular weight biochemicals present in living organisms which have a role as precursors, intermediates or products in enzymatic reactions[146] observed in metabolism. Importantly, metabolites are studied to interrogate metabolism but also to understand the regulatory and structural roles metabolites play. In fact, many metabolites are involved in a number of biological pathways and they can act as the rate-limiting step for some of those processes. For example mammalian metabolomics often involves the detection of metabolites participating in glycolysis, TCA cycle, β-oxidation and many other metabolic pathways. Some metabolites may also hold a structural role; some cholesterol lipids for instance can merge into the phospholipid cell membrane acting through weak chemical interactions. The majority of metabolites are organic compounds (e.g. carbohydrates, amines, organic acids, fatty acids. Biochemicals involved in metabolism are defined as metabolites. The chemical structure of each metabolite allows sub-grouping of compounds into classes sharing similar biological and

physical-chemical properties. For example, hexadecanoic acid is an essential metabolite in mammalian systems and can be defined as a fatty acid. The molecular weight of metabolites ranges from 1 (proton) to greater than 1500 Dalton[147] (e.g. TAGs) and their chemical (pKa, LogP, H-bonds) and physical (boiling point, solubility) properties encompass a wide range of diversities. The total number of endogenous and exogenous metabolites in a biological system is defined as the metabolome[148]. The same organism can be constructed with different metabolomes, for example, mammals are composed of many different metabolomes associated with biofluids, cells and tissues. The Human Metabolome Database (HMDB) as of April 2017 reports that 41,514 metabolites are present in the human body[149]. Plant metabolomics report a broader variety and larger number of metabolites when compared with mammalian metabolomes[146]. Resources to define metabolomes are available including online chemical resources (e.g. PubChem[24] or ChEBI[150]) and online metabolomics resources (e.g. HMDB[151], KEGG[152], LipidMAPS[153]).

The holistic study of metabolomes is typically called “metabolomics” or “metabonomics”. Although “metabolomics” and “metabonomics” are often used without distinction, there is a small difference in their meaning. Metabolomics is defined as the “non-biased identification and quantification of all metabolites in a biological system”[146]. Metabonomics is defined as “the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification”[154]. In the early development of both disciplines, metabolomics was mainly related to microbial and plant research through the use of mass spectrometry (MS), while metabonomics studied mammalian systems applying nuclear magnetic resonance (NMR) spectroscopy.

Metabolomics has developed in the last 15 years as an acknowledged new field of study. The development of new analytical techniques has driven the field forward; for example, new developments in GC-MS during the ’60s and ‘70s led Pauling and Horning to identify the metabolic profile of human biofluids in 1968[155] and 1971[156]. The introduction of the terms “metabonomics” in

1999, focused on NMR analysis of human biofluids[154], and of "metabolomics" in 2000, focused on GC-MS analysis for metabolic studies of plants[157]. These developments occurred in conjunction with advances in analytical techniques such as mass spectrometry and NMR spectroscopy (and related computational methods and power) that allowed robust and information-rich analysis of the metabolite composition of biological extracts.

A great portion of scientific resources has been focused on other 'omics studies including genomics, transcriptomics and proteomics. Genomics is the comprehensive study of genes and their functions, transcriptomics is the study of transcription of biological information codified by genes and proteomics concerns the analysis of proteins in an organism, their structure and how this relates to their functions. There is a logical relationship between every 'omics area and a slight perturbation in one biological level may affect the others. For example, metabolites are not just end products of biological processes but can also act as active regulators of metabolism. Alpha-keto acids for instance, have a significant effect on transcriptional factors through inhibition of cyclic AMP[158]. Furthermore, a deficiency of a specific metabolite will cause the activation or inactivation of specific enzymatic processes in order to increase or maintain a specific metabolite concentration. Metabolites can interact with proteins through post-translational modifications (PTMs), chemical modifications to the amino acidic side chains or to the terminal ends of the protein. Specifically, some metabolites are covalently bound to the protein structure after enzymatic activation of the process[159]. Such enzymatic activity may be triggered by other metabolites binding to allosteric sites[160]. Therefore the role of metabolites appears to allow a rapid response to environmental changes. Given this tight interconnection, the study of a specific level, for example metabolism, can provide information about perturbations occurring in other levels such as the proteome or transcriptome. Indeed, some metabolites like glucose[161] and acetyl CoA[162] are involved in PTMs of histone and non-histone proteins. Hence, while genomics and proteomics give us the information about what may occur or is occurring, metabolomics give us a description of what is occurring or

what has just occurred. A metabolic snapshot of a biological system corresponds to its phenotype, the result of genome and environmental interactions[147]. Genetic and environmental perturbations can be monitored through metabolomics, as the metabolite concentration can be very sensitive compared to the observed transcriptional or proteomic changes[163]. Despite the fact that the global analysis of the entire metabolome is not achievable, metabolomics shows clear advantages in some studies. It takes advantage of high-throughput techniques (which are time-efficient and cost-efficient), thanks to the developments observed in the recent years involving a large number of analytical methods[164].

1.7 Metabolomics and its application

Metabolomics has been applied to study many different organisms with different levels of biological complexity. The human metabolome is one of the most complex systems to study and is typically observed with a variety of different phenotypes and with minimal control on the environment (though longitudinal studies can be performed to control the genotype variability and minimise the effect of environmental factors on changes in the studied metabolome)[147].

Metabolomics can be applied to study human disease. It may involve identification of biomarkers for disease diagnosis or risk stratification. An early diagnosis can be extremely important dealing with diseases such as heart failure or diabetes since a precautionary therapy can be adopted preventing invasive and intensive interventions. Identification of biomarkers for disease progression or treatment monitoring and phenotype characterization is also performed applying metabolomics.

Stratified or personalized medicine aims to investigate groups of patients affected by the same disease and to find multiple common traits (in this case metabolome changes) to allow stratification, which would allow treatment in a specific and personalised way. Furthermore metabolomics is an efficient instrument to investigate the pathophysiological role of metabolites in mechanisms

associated with disease onset and progression[147]. This wide area of application has led to a increasing interest in metabolomics research areas concerning endocrinology[165-167], cancer[168], diabetes[169], heart failure[170], hepatitis[168] and others. Until today, metabolomics has led to the discovery of biomarkers for jaundice syndrome[171], celiac disease[172], lung cancer[173], chronic obstructive pulmonary disease[174], hepatocellular carcinoma[169] and acute kidney injury[175] as examples. For example, Wang *et al.* reported branched chain amino acids (BCAA) and a specific ratio of TAGs as biomarkers associated with the risk of later developing type 2 diabetes[169]. Creatinine, DHEA sulfate and androsterone were used by Kim *et al.* to monitor altered metabolism induced by Pregnane X Receptor[176] and Kiss *et al.* has implemented a method to distinguish between the metabolomes of doped and non-doped athletes[177].

In optimal conditions a metabolomics experiment would allow identification and absolute quantifications of all metabolites present in the biological sample. However this is rarely the case. In order to obtain absolute quantification the calibration curves of all the metabolites involved are needed. While it is possible to obtain standards for some metabolites, this task becomes unachievable when the sample contains thousands of metabolites. Another problem that may arise is the failed detection of some metabolites. In fact, depending on the analytical method adopted, the metabolites need to undergo physico-chemical modifications in order to be detected. While some metabolites may be prone to certain modifications (e.g. ionization, vaporization), they can be resistant to others and therefore they will fail to be detected.

There are various strategies to analyze the metabolic content of samples and in general we can define three main approaches. Non-targeted analysis is defined as the holistic detection of a wide range of metabolites following minimal sample preparation. Targeted studies are applied when a limited number of metabolites of known biological importance are to be studied (**Table 2**). A new approach has emerged in recent years called semi-targeted analysis. This strategy provides

quantitative information on a larger number of metabolites (usually belonging to specific biochemical classes) compared to targeted assays but detects fewer metabolites compared to non-targeted assays. The aim of this approach is to investigate classes of compounds instead of a few selected analytes[154]. To this purpose, the company Biocrates has developed a series of kits that allow quantification of specific classes of metabolites for semi-targeted or targeted studies. These kits have been used in the study of insulin resistance[178], hepatology[179] and neurosciences[180]. This method has its own constraints; one is that the number of compounds analyzed is restricted by the number of chemical standards commercially available. For targeted and semi-targeted assays, the identity of all metabolites is known and so the conversion of analytical data to biological knowledge is simpler compared to non-targeted assays where further work to annotate or identify metabolites is required after data acquisition.

Untargeted analysis	**Targeted analysis**
Semi-quantitative	Quantitative (calibration curves required)
To generate hypothesis	To test hypothesis or for routine analysis
Thousands of metabolites analysed	Hundreds of metabolites analysed
Response is obtained as a relative intensity	Response is obtained as concentration

Table 2. Features and differences of untargeted and targeted metabolomics assays.

1.8 The metabolomics workflow

The metabolomics workflow or pipeline defines the complete scientific process performed[181]. This is shown in **Figure 5**.

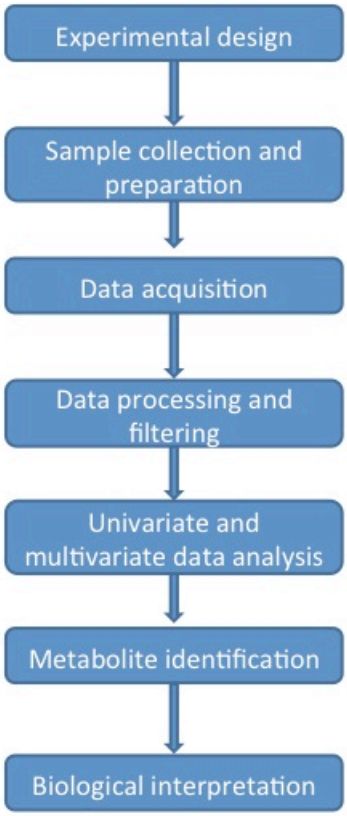


Figure 5. A generic study workflow for an untargeted metabolomics study. The instrumental setup may vary depending on the analytical platform as discussed later. Data pre-treatment and data analysis methods may follow many and diverse procedures as defined in chapter 3.

1.8.1 Experimental design

The first step to perform in metabolomics studies is the experimental design to ensure data acquired are robust and include no biases that will negatively influence the biological conclusions. Many biases can be introduced including gender, age, BMI and drug treatment. The experimental design is dependent on the objective of the experiment itself. One of the main aspects of the experimental design is the reproducibility of the data. Several procedures have been developed to contribute to this. First of all, it is recommended that the experiment is performed according to a standard operating procedure (SOP). This way, the bias that may occur due to different researchers

performing the same experiment in a different way is minimized. If the sample size is large (more than 200), then samples should be split in to batches of 100-150 units which are analysed separately. This is particularly true for mass spectrometry experiments where a decrease in sensitivity is often displayed after ~150 injections[182]. Randomisation is another important step in experimental design. It allows the minimization of unwanted correlation in the analysis. If samples are collected with a certain order related to a factor (e.g. age, sex, BMI, treatment), the variation of the instrumental response over time could be falsely associated with one of these factors leading to flawed data analysis[183]. In order to achieve better results it is often appropriate to stratify the randomization across different experimental stages (e.g. sample extraction, sample reconstitution, analysis). The use of technical replicates generated from a pooled sample has obtained widespread use in metabolomics. These replicates are called quality control samples (QC samples)[184] and are adopted to monitor the experimental variability and the instrument reproducibility. They will be discussed in more detail later in this chapter.

1.8.2 Sample collection and extraction

The sample preparation depends on the purpose of the experiment. If an untargeted study is performed, the samples will require minimal preparation so not to exclude potential analytes of interest from detection while with targeted analysis, the preparation is developed for the optimal detection of the metabolites of interest with the removal of other metabolites and matrix components[185]. Such procedures vary depending on the sample matrix since they have to take into account technical characteristics that may differ considerably between matrices.

1.8.2.1 Sample extraction in mammalian metabolomics

In mammalian and clinical studies, different samples can be analysed although they can be characterised as three major groups: biofluids, tissues and cells.

Biofluids are the most commonly applied sample when performing mammalian metabolomic studies (e.g. urine, serum and plasma). Usually samples are collected and rapidly stored at -80 °C[186]. Urine is a sample matrix that is relatively easy to extract. It can be analysed after removal of particulates (through centrifugation or filtration) and dilution[187]. Further sample treatment may include removal of proteins through the addition of methanol. However, urine has the advantage of containing low concentrations of proteins therefore simplifying preparation steps[168,176,177,188] as no quenching is required. Serum and plasma samples undergo the same extraction procedures. Due to their high protein content, they require deproteinisation through addition of methanol or acetonitrile as the most frequently applied organic solvents[189]. While monophasic extractions adopting water/methanol solutions provide a broad recovery of compounds, biphasic extractions which involve the addition of chloroform or MTBE[190], provide aliquots which are specifically tailored towards one side of the polarity spectrum (polar and non-polar fractions).

When dealing with cells, the metabolome can be collected both from the intracellular and extracellular environment. The former, known as the endometabolome, requires more preparation steps and more controlled experimental conditions while the latter, known as the exometabolome or metabolic footprint can be simply collected through centrifugation or filtration[186]. In order to sample metabolites that are present in the metabolome two main steps will be required: quenching (for samples with high enzymatic activity only) and extraction (for all samples except where direct imaging is performed). Cells undergo different quenching procedures depending on their culture mode. Suspended cells (grown in suspension) are quenched by placing in organic solvents at low or high temperatures. Examples include use of 60 % methanol solutions at -40 °C[191]. The supernatant is then removed through centrifugation or filtration. Conversely, adherent cells are washed in saline solution after removal of the growth medium and quenched in the same solution were extraction is performed[189]. The use trypsin to remove adherent cells is still debated[192]. Extraction usually involves breaking of the cell membranes through thermic or mechanical methods and the addition of

different solvents[189]. Various extraction solutions have been reported including cold methanol[193], hot ethanol[194], perchloric acid[195], and chloroform/methanol/water[194].

Tissue analysis is also performed but sampling techniques are highly demanding. The metabolites have to be released from the intracellular environment through extraction processes and blood has to be washed from the tissues to eliminate contamination. Usually blood washing is achieved through the use with saline solutions at 4 °C. The tissue needs to be homogenized in solution using homogenisers[196] or precellys tubes[197]. The sample is then extracted in a monophasic or biphasic solution.

1.9 Analytical technologies

Metabolomics studies are performed with a range of analytical techniques although the most common are Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS), typically but not always coupled with separation techniques including electrophoresis and chromatography. Direct Infusion Mass Spectrometry (DIMS), which does not involve a separation technique is applied less frequently. A PubMed search led in October 2016 querying the terms "direct infusion mass spectrometry" produced 543 results while the query "liquid chromatography mass spectrometry" produced 103081 results. Each technique has its advantages and disadvantages encompassing sensitivity, reproducibility, sample pre-treatment, range of detected compounds, analysis time and data alignment[198]. For example, querying the terms "NMR" and "metabolomics" 2315 results were retrieved while using the terms "MS" and "metabolomics" the results increased to 3513.

1.9.1 Mass Spectrometry

Mass spectrometry (MS) is a technique that provides detection of metabolites according to their mass-to-charge (*m/z*) ratio.

Mass spectrometers are in simple terms constructed of a sample inlet, ionization source, mass analyser and detection system (see **Figure 6**). One or multiple of these components operate under high vacuum pressures.

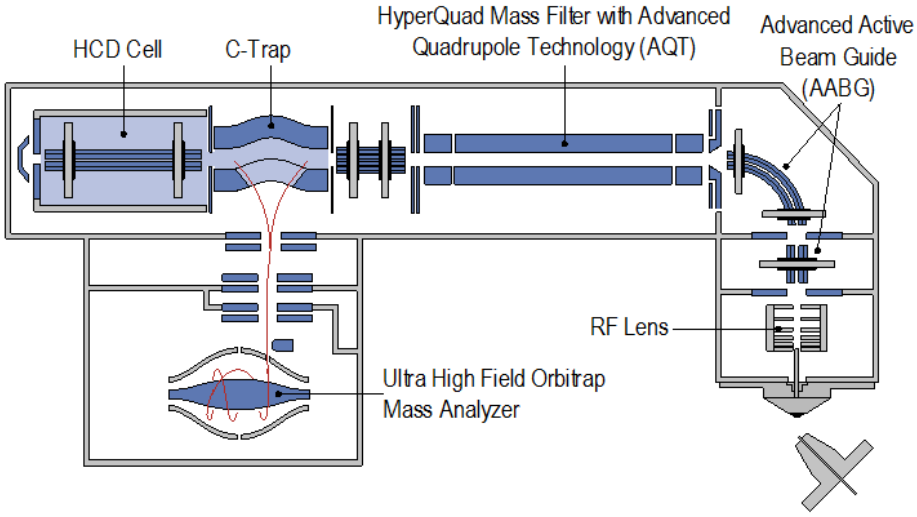


Figure 6. Schematic representation of a Q-Exactive mass spectrometer equipped with an Orbitrap mass analyser.

1.9.2 Sample Introduction

Samples can be introduced directly to the mass spectrometer or the eluent from a chromatographic system can be introduced.

When the samples are introduced directly into the source the technique is called Direct Infusion Mass Spectrometry (DIMS). This technique has successfully allowed the characterization of the metabolic fingerprint of different species of tomato using TOF analysers[199], the classifications of different types of olive oil through the use of chemometrics tools[200] and the assessments of solvent effects on protein recovery in crude cell extracts[201].

DIMS data collection is a complex process and it is characterized by numerous issues arising in the

workflow. As already stated, the metabolites are introduced in the mass spectrometer with no separation, therefore the sample matrix is often complex and multiple metabolites can be detected as a single signal. This is especially true for untargeted studies. DIMS techniques also display impediments concerning metabolite identification, as it will be later discussed.

This series of issues promoted the widespread use of chromatographic techniques in metabolomics[203-205] in order to obtain simpler data matrices and more parameters (e.g. retention time) to compare during the metabolite identification phase.

1.9.2.1 Chromatography

While the interpretation of a MS spectrum can be extremely challenging, especially for complex samples populated by many metabolites, the use of separation methods can ease the detection and identification of metabolites in a complex mixture. Chromatography can separate different compounds over a time span. The main principle involves a stream of solvent (when liquid chromatography is applied) containing the analytes (mobile phase) passing through a column packed with a solid support (stationary phase). Chromatography takes advantage of different chemical interactions between the analytes and the stationary phase whose chemical nature can vary. In this way the analytes can be retained or pass through the column with no interaction with the stationary phase. The use of different solvents in the mobile phase can vary the solubility and the chemical interaction of specific metabolites with the stationary phase, thus allowing the detection of particular classes of compounds in certain spectral areas. This technique, when coupled with MS, allows the analysis of metabolites distributed over chromatographic time (retention time). Moreover it allows the potential separation of isobaric compounds due to possible chemical structure differences. When the mobile phase is gaseous the technique is called Gas Chromatography (GC); when the mobile phase is liquid it is called Liquid Chromatography (LC).

1.9.2.2 Liquid chromatography

Liquid Chromatography (LC) applies a liquid mobile phase and a packed rather than hollow column as observed for GC. Due to its nature, LC detects many high boiling point compounds that are not detected with GC and therefore GC and LC are complementary analytical techniques. Originally, the stationary phase was silica or alumina, thus offering a hydrophilic surface that interacted with the analytes. Polar compounds were retained on the column while apolar compounds were rapidly eluted. However, starting from the '50s carbon chains, were added to the silica group in order to create a hydrophobic stationary phase[206]. The longer the carbon chains, the more hydrophobic the stationary phase. C-18 columns (18 atoms of carbon) gained much popularity in a technique that is now acknowledged as reversed phase chromatography (RP-LC). Hence RP columns are characterized by non-polar stationary phase which interacts strongly with non-polar metabolites therefore enhancing their detection. RP methods are widely used in studies that target the lipid profile of samples (i.e. lipidomics)[207]. They are particularly suited for the analysis of sample matrix containing many hydrophobic metabolites (e.g. serum, plasma, tissues). Also chromatography with hydrophilic stationary phase underwent significant advances. Nowadays the polar stationary phase are no longer silica and alumina resins, instead Hydrophilic Interaction Liquid Chromatography (HILIC)[164] proved as a more efficient alternative. HILIC columns are characterized by a cushion of water around the stationary phase providing a strong polar interaction with the mobile phase. This method is widely used for sample matrices which are rich in polar compounds (e.g. sugars, nucleotides, aminoacids). Urine for example is typically analysed with HILIC methods[208]. One of the main disadvantages of HILIC chromatography is the higher chance of column blockage and the necessity to maintain a certain pH range in solution. Hence RP and HILIC are often considered as complimentary techniques that provide a good coverage of the polarity spectrum of the metabolome. Further development in recent years has made it possible to adopt new technologies in order to decrease the diameter of packing particles, increase the flow rate and the pressure obtaining high sensitivity and enhanced resolution.

Such achievements allowed the establishment of chromatographic techniques as High Performance Liquid Chromatography (HPLC) first and Ultra Performance Liquid Chromatography (UPLC) later.

Nowadays, a typical UPLC setup include pumps that operate between 600 and 1200 bar, column lengths that vary between 30 and 150 mm, column internal diameter ranging between 2.1 and 4.5 mm, particles diameter under 2 μm and flow rates between 50 and 1000 μL/min. Depending on the method, peak separation and shape need to be optimized fine tuning numerous parameters. Solvent composition and gradient modifications are the primary instruments to optimize the chromatographic method. Flow rate also plays an important role, as a non-optimal flow rate can prove detrimental for the peak shape. Usually optimal flow rates can be calculated depending on the internal diameter of the column. UPLC techniques coupled with mass spectrometry generated an extremely successful instrumental platform which is widely used nowadays for a large variety of analysis. Taking advantage of the precision and accuracy achieved by modern mass analysers like the Orbitrap (accuracy < 5 ppm and resolving power up to 240,000 at *m/z* 400), UPLC-MS analysis turn out to be extremely reliable. As already mentioned, nowadays mass spectrometers manage to achieve high resolution. It must be noted though, that higher the resolution, slower will be the number of scans per second. Hence the analytical method is usually a matter of compromise between mass resolution and scan speed.

1.9.2.3 Ionisation

The metabolites must be ionised before being analysed as the charge is applied to manipulate the flight path of ions and separate ions based on their m/z ratio. Therefore ionization of metabolites is required and can be performed according to different techniques with two common ionization techniques being Electron ionisation (EI) and Electrospray ionization (ESI).

Electrospray ionization (ESI) is widely adopted for liquid samples including liquid chromatography-mass spectrometry applications in metabolomics. The ionization occurs in solution through a strong

electric field. The addition of acid species in solution facilitates the formation of protonated ions. Subsequently, the charged analytes flow through a charged capillary bearing a potential difference with a counterelectrode between 3000 and 4500 V[209]. Such a difference can be regulated intervening on the distance between the capillary and the electrode. The analytes exit the capillary and through the use of heat and gas stream, the solvent gradually evaporates generating nebulized charged droplets. In this way droplets containing ions with the same charge reduce their volume until exploding in a coulomb effect. Hence the analytes enter the mass analyser with no trace of solvent (**Figure 7**).

ESI mainly produces $[M+H]^+$, $[M-H]^-$ and $[M+NH_4]^+$ ions; nevertheless it is common to detect ion adducts with K^+ and Na^+ formed by thermal ion attachment. While this process can hinder the detection of the $[M+H]^+$ and $[M-H]^-$ ion forms, it can also prove useful when the analytes do not produce simple protonated or deprotonated ion forms. The use of specific salts in solution can promote the formation of specific ion adducts[210] that can be easier to detect compared to the protonated counterpart.

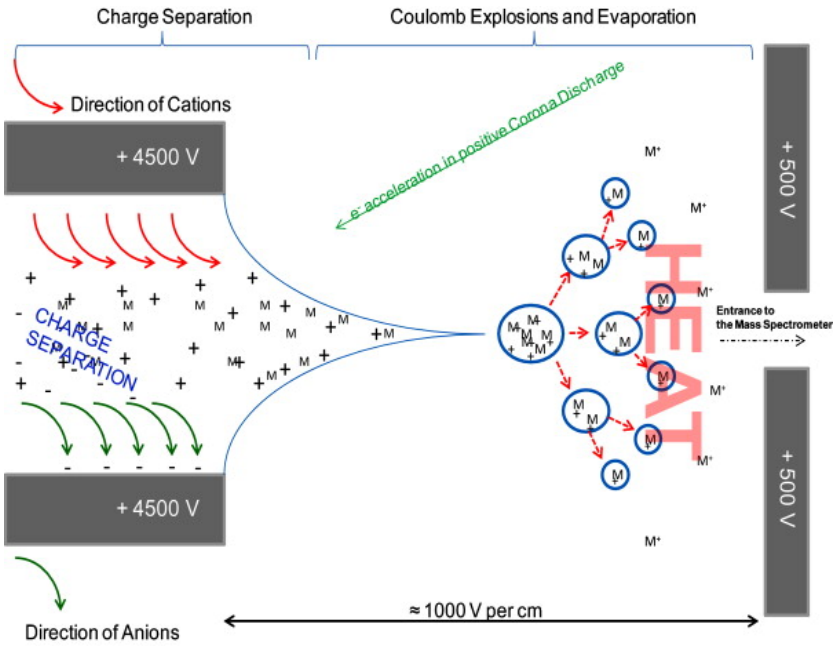


Figure 7. Schematic representation of an ESI process[209]. The ions that are present in the capillary are separated from ions with opposite charge through application of an electric field. In the central part of the picture gas steam allows the evaporation of the solution until creation of charged droplet through Coulomb effect. These droplets are ready to enter the mass spectrometer.

The ESI source may suffer from poor ionization when the solvent composition is rich in water. Ionization efficiency has also proved to be dependent on physical chemical properties as LogP and pKa. For example aromatic molecules rich in N-hetero atoms produce a larger quantity of $[M+H]^+$ ions when compared to oxidized aromatic compounds[211]. It is particularly important when using ESI sources to keep the salt concentration under a certain threshold as high salt content are known to destabilize the spray and impair the ionization.

1.9.2.4 Mass analysers

Following ionization, metabolites are transferred directly or, more commonly, via multiple ion optics to the mass analyser. For example in the Q Exactive mass spectrometer (**Figure 6**) ions pass through ion optics, a quadrupole mass analyser and the C-trap en-route to the Orbitrap mass analyser. Mass analysers separate and detect ions according to their *m/z* ratio. Different mass analysers are available

and each offers advantages and limitations. Specifications for a number of different mass analysers is shown in **Table 3**.

Mass analyser	Mass accuracy (ppm)	Resolving power (10^3)	Scan Rate (Hz)	m/z range (×10^3)	Dynamic Range
Quadrupole (Q)	Low	3-5	2-10	2-3	10^5-10^6
Ion Trap/Linear Ion Trap (IT/LIT)	Low /50-500	4-20	2-10	4-6	10^4-10^5
Orthogonal Time of Flight (oToF)	1-5	10-60	10-50	10-20	10^4-10^5
Triple Quadrupole (QqQ)	5-500	>7.5	>10	>3	10^5-10^6
Qq-TOF	1-5	>60	>100	>40	10^4-10^5
Orbitrap	1-3	100-240	>18	>4	10^3-10^4
FT-ICR	0.3-1	750–2500	0.5-2	4-10	10^4

Table 3. Specifications of widely used mass analysers[212]

Given the broad range of mass analysers that are available nowadays some trends developed in terms of untargeted and targeted metabolomics. Reproducible and sensitive mass analysers (e.g. triple quadrupole, FT-ICR) are the instruments of choice when performing targeted studies while mass analysers with higher resolving power and scan rate (e.g. Orbitrap, Q-TOF) are preferred when performing untargeted metabolomics. Untargeted studies are performed on samples that undergo minimal sample preparation; therefore the mass spectrum is often complex displaying a plethora of signals that are likely to overlap. Hence, resolving power is a fundamental feature that allows the distinction of ions with similar *m/z* ratio. The higher the resolving power, the higher is the chance to

detect different ions having the same nominal mass but different monoisotopic mass. Indeed, it is possible that even after chromatographic separation two ions having the same nominal mass are eluted at the same retention time. Overlapping peaks can in fact appear as a unique peak causing various issues. The intensity of the detected peak is generated by the sum of the intensities of two actual peaks; the peak shape can be affected thus deviating from an optimal profile; one of the two peaks can hide the other.

TOF, Q-TOF and Orbitrap mass analysers are largely used in untargeted studies due to their high mass accuracy, scan rate and resolution. In 2015 and 2016 for example, an untargeted screening for toxicants in urine was performed using an Orbitrap mass analyser[213], a GC-TOF approach detected 13 statistically significant metabolites in saliva that differentiate smokers from non-smokers[214], the complementary use of QTOF and hybrid quadruple Orbitrap (QExactive) allowed the detection of biomarkers of testosterone misuse with potential anti-doping applications[215]. In particular, the Orbitrap analyser was used extensively in the studies conducted for the present thesis hence a more exhaustive description will be provided. Orbitrap analysers are formed by two electrodes as shown in **Figure 6**. The outer electrode contains the inner one; between the two a vacuum is present in order to ensure a long mean free path for the ions. The ions that enter the analyser are attracted to the inner electrode but oscillate around it due to centrifugal force. A voltage (5 kV) is applied between the outer and inner electrodes thus creating a linear electromagnetic field[216]. The molecules are injected in the analyser with a certain tangential velocity. The radial electromagnetic field deflects the analytes towards the inner electrode but the tangential velocity provides centrifugal force. Given the linearity of the field, the molecules that are injected in the analyser are subject to an oscillating motion around the inner electrode[217] that can be described as

$$\frac{\mathrm{d}^2x}{\mathrm{d}t} = -\omega^2 x$$

that is resolved as

$$x(t) = A\sin(\omega t) + B\cos(\omega t)$$

where *t* is the independent variable and the angular frequency *ω* depends on the *m/z* ratio. While orbitrap analysers work in a pulsed fashion, they can be combined with ion sources that are continuous. The interface between ion source and analysers is then achieved through the use of ion traps that control the release of the ions to be analysed. The C-trap gained popularity as one of the most efficient ion traps to be coupled to Orbitrap analysers. Orbitrap analysers typically can reach up to resolution 100,000. However at a scan rate of 1Hz the maximal resolution is 60,000. The scan rate itself can reach 18 Hz. Orbitrap analysers are also characterized by good sensitivity and a mass range that reaches 2000 Dalton.

1.10 Conversion of raw data in to a data matrix

Once the samples are analysed data cannot be investigated *tout court.* For example, in LC-MS metabolomics experiments, the analytical output often consists of large datasets reporting data for each different sample including the intensity of all the peaks that have been detected – this is a m/z vs. retention time vs. intensity three-dimensional dataset.

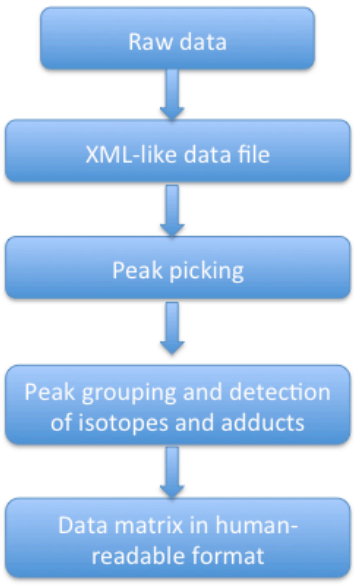


Figure 8. Data processing workflow leading from raw data to human readable datasets.

The obtained raw data requires several processing steps (**Figure 8**) prior to univariate and/or multivariate analysis and the raw data format is heavily dependent on the manufacturer of each instrument. As already stated, these data can be presented in different formats and much effort is spent to unify different formats into a single and universally accepted file format. An example is the .mzML file format which is now available in many conversion and analysis tools (e.g. ProteoWizard[218]). Once data are converted in to a standardized file format, they need to be aligned and subsequently converted into a format that may be handled and modified easily. There are specific software whose purpose is to convert different .mzML data files into a unique file (often a .csv) represented by a matrix reporting samples in rows and different metabolites in columns. Examples include freely available software like XCMS[219] , commercial software like Progenesis (Nonlinear Dynamics)[220] or software available from the instrument manufacturer like Compound Discoverer2.0 (Thermo Fisher Scientific)[221].

1.11 Assessment of data quality

The quality of data can impact on the validity of the biological conclusions. Significant work has been performed to characterise the different types of variation introduced in a metabolomics study. Sample preparation and data acquisition can introduce sources of variation. The sample physically interacts with the LC and MS instruments in LC-MS metabolomics studies and this can lead to changes in the retention time and measured response across an analytical batch. For example, research by Zelena *et al.* have shown how data for LC-MS can change across an analytical batch and defined that for LC-MS a specific number of injections should be performed and then routine instrument maintenance should be performed[182]. Earlier research demonstrated that instrument sensitivity can change significantly after routine instrument maintenance and that non-important samples should be applied to 'equilibrate' the LC-MS system prior to the analysis of biological samples[221].

Both biological and technical variation will be present in a metabolomics dataset. An ideal analytical method would consist of relevant biological variation and a low level of technical variation so that the experimental outcome would provide significant biological information even if biological variation is low. Therefore it is worthwhile to identify and quantify the various sources of technical variation that can be encountered during an experiment including factors such as sample preparation and analytical measurements. In order to quantify technical variation in LC-MS metabolomics studies, the use of Quality Control (QC) samples has been strongly recommended[184]. QC samples are usually replicates of a pooled sample produced from mixing aliquots of all the biological samples, thus producing a representation of all the qualitative variability in the study[205]. Aliquots of the pooled sample undergo the same technical procedure as the biological samples and it is assumed any technical variation introduced will be equivalent for QC and biological samples. Three reasons justify the use of QCs:

(1) the injection of a series of QCs at the beginning of the analytical batch guarantees the

conditioning of the column (in experiments involving chromatography) with a sample type which is extremely similar to the one soon to be analysed[205];

(2) the injection of a QC replicate every n^{th} injection allows the monitoring, filtering and reporting of technical variation. As QC samples are identical replicates, the removal of features with a relative standard deviation (RSD) above a certain threshold gives a quantitative measure of the instrumental stability and ability to remove features of high technical variance[182].

(3) the injection of a QC replicate every n^{th} injection allows for correction of drift in the measured response[184,222].

1.12 Normalisation, scaling and transformation

Further procedures provide the reduction of the impact of technical variation. For example, normalisation takes into account and nullifies the effect of samples of different dilution (for example, for urine). Technical variation due to instrumental nuisance may lead to the occurrence of missing values that therefore need to be imputed according to different methods. This aspect plays an important part in multivariate analysis since many procedures do not operate robustly for datasets with missing values. Successively in the workflow, scaling and/or transformation is often applied; scaling is applied to remove heteroscedasticity from the dataset (after scaling the variance for each metabolite to a constant value) while transformation is applied to minimise the differences between high abundance and low abundance metabolite features as the relative change in concentration of metabolites is more informative than their absolute concentration. This feature assumes a major role in multivariate analysis since all the variation, if the data are not transformed, will be addressed to the most abundant metabolites. In chapter 3, an assessment will be performed on LC-MS metabolomics datasets acquired on different manufacturer's platforms in order to identify the best 'fit-for-purpose' missing values imputation method and data pre-processing workflow (in this case normalisation, scaling and transformation) method to apply for metabolomic experiments. Surprisingly, limited research has been published on the most appropriate data pre-processing

steps[223], therefore a thorough analysis in this direction will fill a significant gap in LC-MS driven experiments.

1.13 Univariate and multivariate data analysis

Once the biological variation has been correctly addressed, further analysis will be performed. The most important features are identified through univariate and multivariate data analysis following the pre-processing steps discussed in the previous section.

Metabolomics data can be analyzed applying univariate and multivariate methods. The purpose of the analysis is to identify biological variation in the concentration of metabolites related to the biological question and statistical methods provide the mathematic tools to accomplish this task. Univariate methods that are commonly applied to metabolomics studies are the Student's t-test and One-Way Analysis of Variance (ANOVA)[224]. It is noteworthy that these statistics methods are parametric meaning that they can be applied only if some requirements are satisfied, e.g. the data has to follow a normal distribution. If these requirements fail then non-parametric methods should be applied, the Mann-Whitney U test and Kruskal-Wallis[225] (test for two classes and more than two classes, respectively). Non-parametric methods are more conservative and are valid in most cases. While univariate analysis considers the variation of a single variable (metabolite) at a time, it must be noted that metabolomics datasets are composed by numerous variables that are often inter-dependent. To this regard, multivariate analysis instead of focusing on the variation of a single metabolite at the time, considers the global trend of the metabolic profile (through principal components, decision trees etc.) and displays it in a synthetic fashion (e.g. two-dimensional plots)[226]. For example, metabolites detected with a weak signal may not appear as significant in univariate analysis but may contribute to the biology of the system and that would be displayed in multivariate analysis. Therefore, multivariate methods, which consider the global variation of the dataset together provide their own benefits and can be useful for three main reasons:

(1) the response variable itself (the class information) comes from an unknown function in which all the variables of the dataset are varying and some of them are inter-dependent

(2) the samples are often collected and/or analysed across different space ranges (samples collected from various areas) or time ranges (samples collected and/or analysed in different days) making it difficult for univariate analysis to show differences due to diverse conditions of sampling or analysis while a simple Principal Component Analysis (PCA) analysis may point out a difference which is induced by an external factor (e.g. samples collected at different time) and not by the biological reason of interest[224]

(3) significant variations between subjects can occur leading to differences in the metabolome which are not due to the induced variation i.e. inter-subject variability occurs.

The use of multivariate analysis can detect such variations and visually highlight them through techniques including PCA and Partial Least Squares – Discriminant Analysis (PLS-DA) [227].

For these reasons the use of multivariate analysis in metabolomics is widespread. PCA allows the visualization of all the samples (scores plot) in relation to combinations of all the variables (Principal Components) or via loading plot that shows the relationship between the variables and the principal components (loadings are computed as the eigenvectors of the covariance matrix). When we relate the principal components to a response variable a Partial Least Squares (PLS) computation is performed and through loadings plots or Variable Important in Projections (VIP) scores the discriminating variables are identified[226]. In general, methods that use the response variable to build a model are called supervised, otherwise they are called unsupervised. PCA is an unsupervised method commonly applied to visualize data and assess outliers. Supervised methods are very powerful but need to be validated since they are often extremely data-dependent. Various techniques have been applied to validate supervised methods, mainly relying on test sets including a

set for cross-validation. Recently, this aim has been pursued through machine learning algorithms using Artificial Neural Networks (ANNs)[228].

1.14 Metabolite identification

Univariate and multivariate methods identify the metabolites whose biological variation is significantly different between samples and classes. The next step is then the annotation or identification of metabolites. Metabolite annotation and identification is one of the most challenging steps in the workflow due to the complexity of the data, the number of biological isomers detected and the unavailability of a resource defining all metabolites that are expected to be present in a specific sample type.

The Metabolic Standard Initiative (MSI) is a consortium that started operating in 2005 in order to define objective standards for the research quality in the metabolomics community including metabolite identification. Four levels were proposed[229]. The levels decrease by order of confidence and are: identified metabolites (1), putatively annotated compounds (2), putatively characterized compound classes (3), unidentified compounds (4). Level 1 is reached for non-novel metabolites when a minimum of two orthogonal measurements (e.g. retention time and m/z) is collected for a standard in the same instrumental conditions and compared with the experimental data. Novel metabolites need to be fully described in elemental composition and stereochemistry hence entailing extraction, isolation and purification, accurate mass determination and ion fragmentation. Level 2 is reached when the metabolites are annotated by comparison with spectral libraries but with no standard comparison. Similarly level 3 is reached when physical chemical properties of a specific chemical class are compared against a library. At the current stage metabolite identification still actively requires the researcher's biochemical knowledge[164]. The creation of databases for definitive identification is an admirable initiative; however analysis performed on different platforms (multiple models of spectrometers and chromatography) contrast with the standardization of a unique

database. Although research is struggling in this direction, a database for definitive identification is not available as of yet.

The typical identification workflow for a UPLC-MS study uses accurate *m/z* and retention times to map the detected features against libraries. PUTMEDID is an automated tool dedicated to such task[147]. Mean or median peak area is frequently used to detect isotopic patterns. A deeper level of identification can be reached through the study of metabolite fragments. The use of subsequent ion fragmentation (MS^n where n is the number of fragmentation cycles) allows the facilitated identification of the analytes. Chemical fragments deriving from the same molecule can be combined together in order to address with more confidence the chemical structure of the precursor compound.

Occasionally these procedures do not ensure the identification of a metabolite: in such cases fractionation and isolation of the metabolite is required.

1.15 Research objectives

The application of untargeted and targeted metabolomics studies offers many advantages as discussed in this chapter. However, specific areas of the metabolomics workflow are still being developed including methods for analysis of low volume samples and for data processing. Advances in these areas will provide a higher level of applicability of metabolomics in human medical research. As shown in this chapter the impact of GC metabolism on global metabolism, including through the interaction with hormones like insulin, has been studied though requires significant further research to investigate global metabolic changes in biofluids and also in tissues. Currently, much of the research performed has focused on the application of steroid profiling[176,230-232].

Therefore this PhD programme has the following four objectives:

1. To investigate the impact of different data pre-processing methods on the quality and validity of untargeted LC-MS datasets and to define an appropriate data pre-processing workflow (see Chapter 3).

2. To investigate the effect of cortisol administration at low and high insulin levels on global metabolism in healthy human subjects through the use of UPLC-MS methods and non-targeted metabolomics. (see Chapter 4)

3. To investigate the effect of different types of cortisol administration levels on global metabolism in healthy human subjects through the use of UPLC-MS methods and non-targeted metabolomics. (see Chapter 4)

4. To investigate the effects on global and tissue-specific metabolism of two pharmacological 5-alpha reductase inhibitors (finasteride and dutasteride) and the associated influence of low and high insulin levels. (see chapter 5)

CHAPTER 2 – Materials and methods

2.1 Assessment of normalization, missing value imputation, scaling and transformation methods in LC-MS-based non-targeted metabolomics

2.1.1 MVI study

2.1.1.1 Data sets

Four UPLC-MS datasets were applied to investigate the performance of different MVI methods. These datasets and the one presented in section 2.2.1 were collected within Dr. Dunn research group, although the PhD candidate did not participate in the data collection.

(1) Mouse serum (MS) from a study of ischemia following stroke acquired in negative ion mode. The dataset consisted of 34 samples divided in to five different classes and reported 4435 metabolite features. The *m/z* range was 100-1000Da and the data were acquired applying a UPLC Accela system operating with RP chromatography coupled to an electrospray LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, UK) applying a method as previously described[183] .

(2) Placental tissue extract (PT) from a study of normal and pre-eclamptic pregnancy as published previously[233]. Data for 24 samples were acquired in negative ion mode and RP chromatography with 3412 metabolite features. The *m/z* range was 100-1000Da and data were acquired on a UPLC-MS system operating with RP chromatography (Waters Acquity UPLC system and Thermo Scientific LTQ-Orbitrap XL).

(3) Human urine samples (UR). The dataset consisted of 48 samples acquired in positive ion mode with 3823 metabolite features. The *m/z* range was 100-1000Da and data were acquired on a UPLC-

MS system operating with HILIC chromatography (Thermo Scientific Dionex Ultimate 3000 UPLC system and Thermo Scientific Q-Exactive). These data are unpublished.

(4) Intracellular metabolome dataset (FP) obtained from an arthritis study involving 88 fibroblast samples acquired in positive ion mode with 2008 metabolite features. The *m/z* range was 100-1000Da and data were acquired on a UPLC-MS system operating with RP chromatography (Thermo Scientific Dionex Ultimate 3000 UPLC system and Thermo Scientific Q-Exactive). Samples were kindly provided by Dr. Andrew Filer at University of Birmingham.

2.1.2 XCMS processing

The .RAW files produced were converted to .mzML applying ProteoWizard 2.1[218] followed by deconvolution and peak alignment applying XCMS as described previously[234]. The noise threshold applied was set to 3 and the bandwidth applied for grouping was set to 10.

2.1.3 Feature pre-treatment

Peak filtering was performed. Specifically, if a single feature in a single class was completely filled with missing values, it was replaced by a small value (defined as the half of the minimum value in the matrix). Consequently, features containing more than 20% missing values across all classes were deleted.

2.1.4 Missing values occurrence assessment

The Pearson coefficients related to missing value occurrence vs. mean abundance of the non-missing values of each metabolite feature, missing values occurrence vs. *m/z* values and missing values occurrence vs. retention time were evaluated. This step assessed whether the missing values occurred completely at random (MCAR) or did not occur at random (MNAR)[235].

2.1.5 Assessment of different MVI methods

After normalisation by sum (see section 2.3.4), six different missing value imputation methods were assessed:

1. Small value replacement (SV): the missing values were replaced by half of the minimum value of the entire dataset[236].

2. Mean replacement (MN): for every metabolite the missing values were replaced by the mean of the specific metabolite across all samples[236].

3. Median replacement (MD): for every metabolite the missing values were replaced by the median of the specific metabolite across all samples[236].

4. K-nearest neighbour imputation (KNN): the missing values are replaced by the average of the corresponding (sample specific) non-missing values in the k (here k=10) closest columns in terms of Euclidean distance of the abundances across all the samples. Therefore every imputed value is unique across the samples or across the metabolite features[236,237].

5. Bayesian principal component analysis replacement (BPCA): the missing values are replaced by the values obtained through principal component analysis regression with Bayesian method. Therefore every imputed missing value is unique across the samples or across the metabolite features[237,238].

6. Random forests (RF): missing values are iteratively imputed using as a decisional criterion the proximity matrix generated by a random forests classification computed across the total number of metabolites[239].

All the computations were performed using built-in R 3.0.2 functions except KNN which was

performed using the package "impute"[240], BPCA which was performed using the package "pcaMethods"[241] and RF which was performed using the package "missForest". "impute" and "pcaMethod" packages are freely available in Bioconductor[242] while "missForest" is downloadable from the CRAN repository[243].

Multivariate imputation by chained equation[244] was not tested since it resulted in a computationally intense method while Sangster's method[245] was not performed due to lack of technical replicates, which is a common observation in non-targeted metabolomics studies.

In order to assess the performance of the different imputation methods, all datasets were treated as follows: only the features with no missing values were retained, missing values were assigned randomly until reaching a missing value frequency of 10%. This value was acknowledged as a representative estimate of the number of missing values which are present in a UPLC-MS dataset before any data treatment stage as confirmed by other studies[223,237]. This value was consistent with six publicly available datasets collected on different UPLC-MS instruments (Agilent and Waters). The details are available in **Appendix Table 1**. Furthermore, our four datasets showed a frequency of missing values before any filtering of approximately 15%. At this stage, two datasets (for each of the four sample sets) were available, one complete and another one with missing values computationally added at random. In this case the data matrix was treated as a column vector and the missing values were imputed using the sample() function available in R. Since the corresponding original values were known, it was possible to compare the imputed values to the original values through different methods. For this purpose normalised root mean squared error (NRMSE) was calculated for every MVI method. As reported by Troyanskaya[5], NRMSE was calculated as the root mean squared difference between the original and the imputed datasets divided by the average value of the original complete dataset.

2.2 Data pre-processing study

2.2.1 Data set

One UPLC-MS dataset was used to assess the effect of different combinations of pre-processing methods, absolute response and fold change. Human serum (HS) was acquired in positive ion mode for 77 samples. The total number of detected metabolite features was 3837. The *m/z* range applied was 100-1000Da and the data were acquired on a Ultimate3000 UPLC system coupled to a LTQ-FT Ultra mass spectrometer (Thermo Scientific, UK).

2.2.2 XCMS processing

The XCMS procedure was performed as described above in section 2.1.2

2.2.3 Feature pre-treatment

Feature pre-treatment was performed as described above in section 2.1.3

2.2.4 Normalisation methods

The datasets were normalised by sum or by PQN.

- Normalisation by sum: each value in a row is divided by the total sum of the row (sample) and multiplied by 100.

- PQN: for every feature the mean is calculated across all the QC samples. A reference vector is then generated. The median between the reference vector and every sample is computed obtaining a vector of coefficients related to each sample. Each sample is then divided by its vector of coefficients. This method was adapted by Dieterle *et al.*[246]. Its purpose is to take

into account the concentration changes of some metabolites which affect just limited regions of the spectrum.

Other normalisation methods were not tested since the analytical samples contained no standard reference or internal standard and technical replicates were not available, a common observation in untargeted metabolomics studies. Therefore all normalisation methods implying external or internal standard reference or technical replicates were excluded.

2.2.5 MVI methods

All MVI methods described above in section 2.1.5 were tested.

2.2.6 Transformation methods

When a transformation was applied, three different methods[247] were assessed:

- Generalised logarithm (glog)[248]: every value is transformed according to the equation

$$z = ln(y + \sqrt{y^2 + \lambda})$$

where y is the untransformed value, z is the transformed value and λ is a parameter which is iteratively computed in order to minimise the variation

- Natural logarithm (nlog): every value is transformed in the corresponding natural logarithm

- Inverse hyperbolic sine (IHS)[249]: every value is transformed according to the equation:

$$z = ln(y + \sqrt{y^2 + 1})$$

2.2.7 Scaling methods

After transformation, the datasets were analysed without any further treatment or after processing applying three different scaling (peak-wise normalisation) approaches:

- Autoscaling: every metabolite feature is centered and divided by the standard deviation of the column. This treatment ensures the standard deviation of each metabolite is equal to 1. Autoscaling, along with range scaling, is not affected by the feature abundance[250].

- Pareto scaling: every metabolite feature is centered and divided by the square root of the standard deviation of the column[251].

- Range scaling: every metabolite feature is centered and divided by the numerical difference between maximum and minimum of the column[252].

- Vast scaling: every metabolite feature is autoscaled and divided by the coefficient of variation. It is particularly suited for metabolites bearing small fold changes[253].

All the calculations were carried out using built-in R 3.0.2 functions. All the possible permutations of normalisation, missing values imputation, transformation and scaling were explored and applied on the *in-silico* modified dataset described below.

2.2.8 Assessment of different methods

The pre-processing procedures described above were applied on an *in-silico* modified dataset in order to strictly control all the modifications occurring. The dataset was treated as follows.

The QC samples were labelled as QC. Two biological classes were randomly assigned to the other samples so that no statistical difference (applying Mann-Whitney U test) was reported between the two classes.

The dataset was normalised applying normalisation by sum and PQN according to section 2.3.4. Subsequently, after checking for consistency of fold change between the two classes (median values ranged between 0.8 and 1.2), the features were split in to three blocks based on mean abundance: specifically, the mean abundance was calculated for each metabolite feature; hence the vector of means was ordered and divided in three equal sections (blocks). 32 metabolite features were randomly selected from each of these blocks making sure that they belonged to the same response block regardless of the normalisation method. At this point 32 multiplication factors ranging from 0.1 to 10 were provided (**Table 4**). Each of the 32 features in a block was multiplied by a corresponding multiplication factor (only one class was multiplied and the other left unmodified). Multiplication factors 0.1, 0.2 and 5-10 were considered as inducing high variation; factors 0.3-0.5 and 2-4.5 were considered as inducing medium variation.

High		Medium			Medium/Low			Low			Medium/Low			
0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5

Medium						High									
2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10

Table 4. 32 metabolite features from each abundance block were multiplied (class 1 only) according to 32 different factors. These factors were classified as high, medium, medium/low and low fold change inducers.

2.2.9 Univariate and multivariate data analysis

The Shapiro-Wilk test of normality was performed on all metabolite features to assess whether the data were normally distributed or not (the null hypothesis is that the distribution does not differ from a normal distribution therefore a p-value < 0.05 means that a normal distribution is not observed).

Univariate analysis was performed applying Mann-Whitney U test and Student's t-test between the two classes. QC samples were removed from the dataset prior to univariate analysis and the Benjamini-Hochberg correction for multiple comparisons was applied[254].

Multivariate analysis was performed applying Principal Component Analysis (PCA) and Partial Least Squares – Discriminant Analysis (PLS-DA). For this purpose, the R package mixOmics and pcaMethods were adopted. Further statistical analysis consisted of t-test or Wilcoxon test performed on the PC scores in order to identify the significant clustering patterns. Prior to these procedures a Shapiro-Wilk test was carried out to assess the normality of the score matrix in order to apply parametric or non-parametric statistical methods.

For PLS-DA, the classification performance was estimated through the calculation of R^2 and Q^2 values. The optimal number of components was evaluated through a 10-fold cross validation.

2.2.10 Analysis of the univariate outcome

Mann-Whitney U test and Student's t-test between class 1 and 2 were calculated for each processing permutation (total of 280 permutations with all possible combinations of missing value imputation, normalisation, scaling and transformation). The total number of statistically significant metabolite features ($p<0.05$ after FDR correction) was recorded along with the number of statistically significant metabolite features present from the 96 imputed metabolite features. Hence an estimate of the

number of false positives was obtained by subtracting the number of imputed significant metabolite features from the total number of significant metabolite features. Here, by "false positives" we indicate metabolite features that were not *in-silico* altered, hence bearing a fold change between classes in the range 0.8 and 1.2 and not being statistically significant in the original dataset.

2.2.11 Analysis of the multivariate outcome

PCA and PLS-DA were performed on each dataset. All the metabolite features were ordered according to their loading value (absolute value) on PC1, PC2 and a combination of both for PCA and on Var1, Var2 and a combination of both for PLS-DA. Therefore a ranking position was obtained for each of the 96 imputed metabolite features on six different outcomes (three for PCA and three for PLS-DA). The top 10 positions out of the 96 metabolite features were analysed in relation to abundance and fold change. If 8 or more metabolite features in the top 10 positions belonged to the "High abundance" class then an abundance effect was reported. If 8 or more metabolite features in the top 10 positions belonged to the "High fold change" class then a fold change effect was reported. Also the 10 best performing metabolite features out of the 96 were reported and their overall ranking (against all the metabolite features in the dataset) noted.

2.3 Non-targeted UPLC-MS assessment of the metabolic interaction between insulin and cortisol or dexamethasone

Two separate studies were performed in order to investigate independently (1) the metabolic effects of the interaction of insulin and cortisol (insulin and cortisol study) and (2) the effects of dexamethasone and different modes of cortisol administration (dexamethasone and cortisol study).

2.3.1 Insulin and cortisol study

The experimental design is shown in **Figure 9**. Ten healthy subjects (five males and five females) aged between 23 and 47 were recruited and studied in two separate visits via a double-blinded, randomised, placebo controlled cross-over study. A 2-week washout period occurred between the two visits. On each visit each subject randomly underwent a cortisol (0.2mg/kg/hr) or saline infusion which lasted from 17:00 on day 1 to 14:00 on day 2. The subjects underwent a subcutaneous adipose tissue dialysis for the whole duration of the visit. At 08:00 on day 2 (15 hrs from the beginning of the infusion) a serum sample was collected. At 10:00 each patient underwent a low-dose insulin infusion (20mU/m^2/min) for two hrs after which another serum sample was collected. Afterwards, at 12:00 each patient underwent a high-dose insulin infusion (100mU/m^2/min) for two hrs after which a final serum sample was collected.

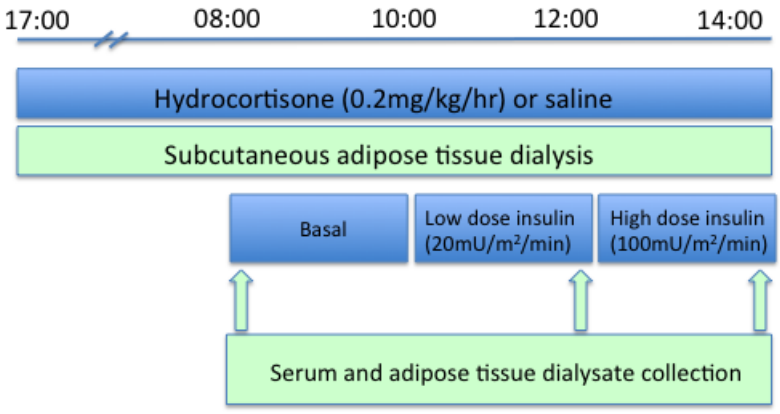


Figure 9. Experimental design for the insulin-cortisol interaction study

2.3.1.1 Serum samples

Overall, six samples were collected from each subject: cortisol infusion at basal level of insulin (HC1), cortisol infusion at low level of insulin (HC2), cortisol infusion at high level of insulin (HC3), saline

infusion at basal level of insulin (S1), saline infusion at low level of insulin (S2) and saline infusion at high level of insulin (S3).

Serum samples were thawed at room temperature. 100 µL aliquots were then transferred to a 2 mL microcentrifuge tubes (Eppendorf, Cambridge, U.K.) and placed on ice. 300 µL of HPLC grade methanol were added followed by vortex mixing for 15 s and centrifugation at 13,000 g for 15 minutes at 3 °C. 330 µL of supernatant were transferred to 2 mL microcentrifuge tubes and were dried by speed vacuum concentration (Thermo Scientific Savant SpeedVac Concentrator SPD111V) at room temperature for 4 h and stored at - 80 °C until analysis. Just before analysis, samples were reconstituted in 100 µL of HPLC grade methanol/water (50:50 v/v), vortexed for 30 s and centrifuged at 13,0 00 g for 15 minutes at 3 °C. The supernatant was then transferred to HPLC vials and analysed. Simultaneously a single pooled QC sample was prepared by combining 80 µL aliquots from each biological sample. This pooled QC sample was aliquoted into 22 different QC samples which underwent the same extraction procedure as described above.

Serum extracts were analysed on a UPLC-MS system (Dionex Ultimate 3000 RSLC UPLC system and Thermo Scientific electrospray Q-Exactive mass spectrometer, Thermo Scientific Hemel Hempstead, UK) using a C_{18} column (Thermo Scientific Hypersil GOLD 100 x 2.1 mm 1.9 µm). Solvent A was water + 0.1% formic acid and solvent B was methanol + 0.1% formic acid. The gradient elution conditions were as follows: from 0 to 1 min (A was constant at 100%), from 1 to 4 min (B increased from 0 to 100% with a concave curve gradient), from 4 to 7 min (B was kept constant at 100%), from 7 to 8 min (A increased from 0 to 100%), from 8 to 10 min (A was constant at 100%). 100% of the column eluent was introduced in to the electrospray source of the mass spectrometer. The mass spectrometer operating conditions were as follows: resolution was set to 70,000 (FWHM at *m/z* 200), AGC target to 3e6, S-lens RF = 80 and *m/z* range of 100-1000Da. Ten conditioning QCs were injected at start of each analytical run. Four of those QC samples were analysed with a HCD MS/MS data dependent analysis

(DDA) method exploring different m/z ranges (100-300Da; 300-500Da; 500-750Da; 750-1000Da) with a stepped collision energy of 30-80%[255] to help identification of unidentified metabolites of interest. After the first ten QC injections a QC sample was injected every 6th sample in order to monitor the signal and allow for further quality checks of the data to be performed[184]. Each sample was analysed separately in positive and negative ion modes.

It must be noted that the PhD candidate performed the data pre-processing and data analysis. The candidate did not perform the sample collection or the instrumental analysis.

The .RAW files produced by the software Xcalibur 3.0.63 were converted to .mzML with ProteoWizard 2.1 followed by deconvolution and peak alignment applying XCMS as described previously[234]. Metabolite features (a m/z-RT pair) showing greater than 30% missing values were removed from the dataset. If an entire class of a metabolite feature was still populated by missing values those were substituted by a small value identified as half of the minimum intensity of the dataset. Subsequently a quality assessment was performed. All of the metabolite features reporting a relative standard deviation greater than 20% on all QC samples (starting from QC 8) were removed. The dataset was normalised to total peak area for each sample and univariate analysis was performed through paired Mann-Whitney U tests. No FDR (Benjamini – Hochberg) correction was applied as no feature was statistically significant after this treatment. Subsequently missing values were imputed by random forest (R package missForest[256]) for PCA and by KNN for PLS-DA. Data were transformed applying a Glog transformation for multivariate analysis. At this point the dataset was analysed applying PCA and PLS-DA (optimal number of components was evaluated performing a 10-fold cross-validation).

Metabolite annotation was performed using the PUTMEDID_LCMS workflow[147] implemented on Taverna 2.5[257] to MSI level 2 (putatively annotated compounds)[258]. After annotation the statistically

significant metabolite features have been labelled and manually grouped according to their chemical or biological class.

2.3.1.2 Adipose tissue dialysate samples

Ten samples were collected from each subject: cortisol infusion on the evening of day 1 between 21:00 and 22:00 (fed state; HCE), cortisol infusion on the morning of day 2 between 6:00 and 8:00 (fasted state; HCM), cortisol infusion at basal level of insulin (HC1), cortisol infusion at low level of insulin (HC2), cortisol infusion at high level of insulin (HC3), saline infusion on the evening of day 1 between 21:00 and 22:00 (fed state; HCE), saline infusion on the morning of day 2 between 6:00 and 8:00 (fasted state; HCM), saline infusion at basal level of insulin (S1), saline infusion at low level of insulin (S2) and saline infusion at high level of insulin (S3).

Dialysate samples were thawed at room temperature. 2.5 µL of each sample was transferred to a 2 mL microcentrifuge tubes (Eppendorf, Cambridge, U.K.) and placed on ice. 125 µL of HPLC grade methanol/chloroform (2:1 v/v) containing 5 mM ammonium acetate was added to each sample followed by vortex mixing for 15 s and centrifugation at 16,000 g for 15 minutes at 3 °C. 12 µL of supernatant was transferred to a 384 well plate for each of four technical replicate. Simultaneously a single pooled QC sample was prepared by aliquoting and combining 6 µL from dialysate samples that were not used for the present study. This was required because of low sample volume availability. This sample was aliquoted into 19 different QC samples which underwent the same extraction procedure as described above.

Adipose tissue dialysate samples were analysed on a MS system (LTQ-Orbitrap Elite mass spectrometer, Thermo Scientific Hemel Hempstead, UK) coupled with a chip-based direct infusion nanoelectrospray ion source (nESI; Triversa, Advion Biosciences, Ithaca, NY). Electrospray conditions were as follows: gas pressure was set at 0.3 psi, the voltage applied was 1.5 kV and the amount of

sample injected was 8 µL for a duration of 3 minutes. AGC target was set to 5 x 10^5 and resolution to 240,000. Data was collected applying a SIM stitch approach[259]. The first window is a full mass range scan ranging from m/*z* 50.0 to 1005.0. The other windows are 75 Dalton wide and have an overlap of 20 Dalton between each other. In total 17 SIM windows were acquired.

The lowest quality injection out of the four replicates (calculated as the outlying total ion current) was removed from further analysis. The .RAW files produced by the software Xcalibur 3.0.63 were read through the Thermo Scientific application MSFileReader and then window stitching was operated using an in-house Matlab script. Each set of replicates was then filtered to a single peak list in order to reduce the quantity of noise peaks by including only peaks that are present in a minimum number (n=2) of all spectra. The output is a filtered peaklist with m/*z* corresponding to the mean m/*z* across the replicates and mean intensity. Metabolite features (a m/z-RT pair) bearing more than 30% of missing values were removed from the dataset. It must be noted that in this case the PhD candidate performed the sample extraction, the instrumental analysis, data preprocessing and data analysis.

The dataset was normalised to PQN and univariate analysis was performed by applying paired Mann-Whitney U tests. No FDR correction (Benjamini – Hochberg) was applied as no feature was statistically significant after such treatment. Subsequently missing values were imputed by KNN replacement and a glog transformation was applied prior to multivariate analysis[237,248]. At this point the dataset was analysed applying PCA and PLS-DA (optimal number of components was evaluated performing a 10-fold cross-validation).

Metabolite annotation was performed using the PUTMEDID_LCMS workflow[147] implemented on Taverna 2.5[257] to MSI level 2 (putatively annotated compounds)[258] and the software MI-Pack developed specifically for DIMS studies[260]. After annotation the statistically significant metabolite features have been labelled and manually grouped according to their chemical or biological class.

2.3.2 Dexamethasone and cortisol study

14 healthy male subjects (mean age 40.1 ± 5.1 years, body mass index (BMI) 29.4 ± 2.0 kg/m^2) were recruited. Each subject underwent 5 treatment arms during 5 different periods (one treatment per period) randomised across arms. A 7-day washout period occurred between different treatments. Each treatment occurred on a 2-day period and included regular meals at 13:00 (for treatment 1 only) and 19:00 on day 1 and at 08:00 and 13:00 on day 2. The different treatments are described below and illustrated in **Figure 10**:

1. Assessment of diurnal metabolic changes only; patients were admitted at 10:00 on day 1. A plasma sample was taken at 19:00 (naturally occurring low cortisol level) on day 1 (1-pre) and another at 07:00 (naturally occurring high cortisol level) on day 2 (1-post).
2. Assessment of the effect of dexamethasone administration; patients were admitted at 13:00 on day 1. 1 mg of dexamethasone was administered at 22:00 on day 1, 06:00 and 12:00 (fasted state) on day 2 in order to suppress endogenous production of cortisol. A plasma sample was taken at 07:00 (2-pre) and another at 19:00 on day 2 (2-post).
3. Assessment of the effect of hydrocortisone capsules administration; patients were admitted at 13:00 on day 1. 1 mg of dexamethasone was administered at 22:00 on day 1, 06:00 and 12:00 (fasted state) on day 2 in order to suppress endogenous production of cortisol. At 07:00 on day 2 they received a hydrocortisone capsule (20 mg). A plasma sample was taken at 7:00 (3-pre) and another at 19:00 on day 2 (3-post).
4. Assessment of the effect of hydrocortisone tablet administration; patients were admitted at 13:00 on day 1. 1 mg of dexamethasone was administered at 22:00 on day 1, 06:00 and 12:00 (fasted state) on day 2 in order to suppress endogenous production of

cortisol. At 07:00 on day 2 they received a hydrocortisone tablet (20 mg). A plasma sample was taken at 07:00 (4-pre) and another at 19:00 on day 2 (4-post).

5. Assessment of the effect of hydrocortisone injection administration; patients were admitted at 13:00 on day 1. 1 mg of dexamethasone was administered at 22:00 on day 1, 06:00 and 12:00 (fasted state) on day 2 in order to suppress endogenous production of cortisol. At 07:00 on day 2 they received an intravenous hydrocortisone injection (20 mg). A plasma sample was taken at 07:00 (5-pre) and another at 19:00 on day 2 (5-post).

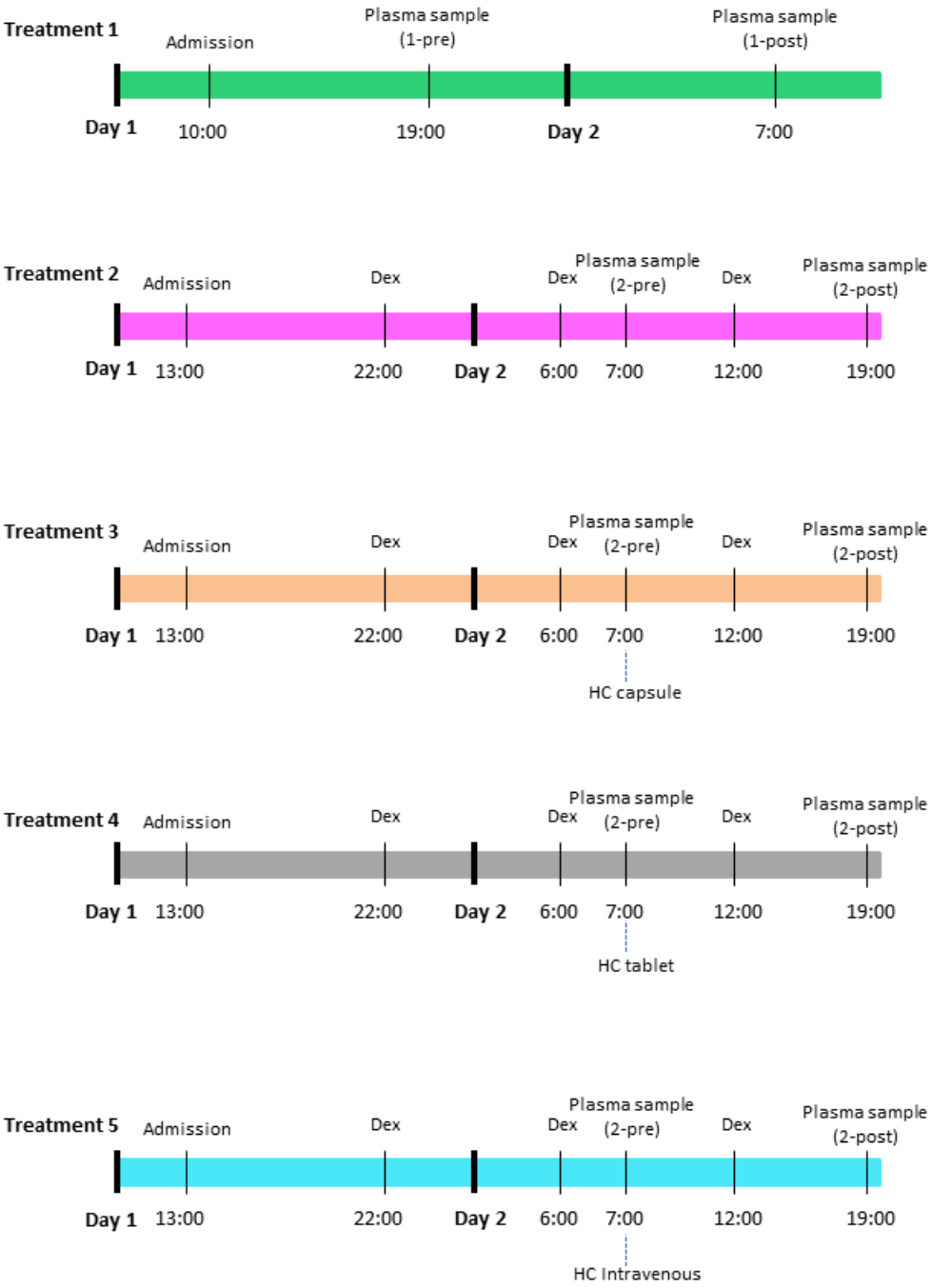


Figure 10. Experimental design for the dexamethasone and cortisol study. Dex: administration of 1 mg of dexamethasone. HC capsule: administration of 20 mg of cortisol in capsules. HC tablet: administration of 20 mg of cortisol in tablets. HC intravenous: administration of 20 mg of cortisol by intravenous infusion.

Plasma samples were extracted, analysed and processed in two batches following an identical procedure as described for the insulin and cortisol study. The two batches were integrated for statistical analysis using signal correction.

It must be noted that the PhD candidate performed the data preprocessing and data analysis. The candidate did not perform the sample collection or the instrumental analysis.

2.4 Inhibitors of 5-alpha reductase activity in steroid metabolism induce global metabolic changes

Twelve healthy male volunteers (mean age 36.3 ± 4.4 years, body mass index (BMI) 26.6 ± 1.2 kg/m^2) were recruited. The experimental design is shown in **Figure 11**.

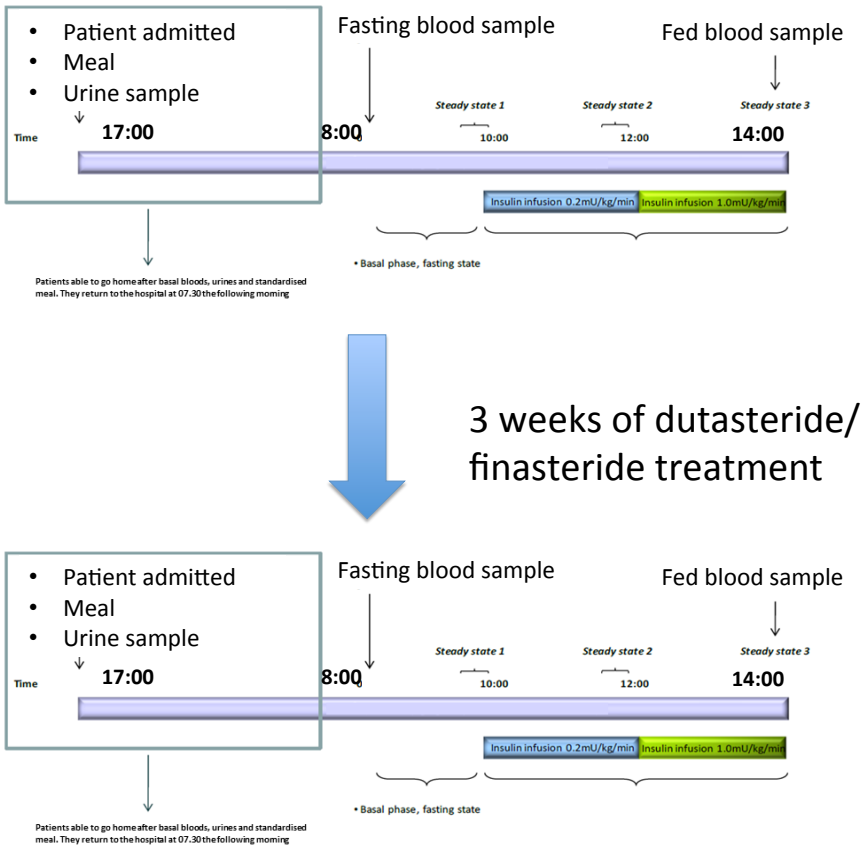


Figure 11. Experimental design for assessment of the effects of finasteride and dutasteride on the human serum metabolome.

Each subject was admitted to the research facility at 17:00 and a spot urine sample was collected. Each subject returned to the hospital at 08:00 the next day after fasting and a blood sample was collected. At 10:00 patients underwent insulin infusion (20mU/m^2/min) for two hours after which the insulin infusion was increased to 100mU/m^2/min. At 14:00 another blood sample was collected. This procedure was followed by three weeks of treatment with finasteride (6 randomly selected patients administered with 5 mg finasteride daily) or dutasteride (6 randomly selected patients administered with 0.5 mg dutasteride daily). Subsequently each subject underwent the same procedure described above. In total 4 serum samples were collected for each subject (visit 1-pre insulin infusion; visit 1-post insulin infusion; visit 2 pre-insulin infusion; visit 2-post insulin infusion) and 2 urine samples

were collected for each subject (visit 1; visit 2).

One serum vial broke during the sample preparation process leading to the removal of one subject (01GL) from the study. Serum samples were extracted following two different procedures: monophasic extraction and biphasic extraction.

The monophasic extraction was performed as follows. Samples were thawed at room temperature. 100 µL of serum was transferred to a 2 mL microcentrifuge tube (Eppendorf, Cambridge, U.K.) and placed on ice. 300 µL of HPLC grade methanol were added followed by vortex mixing for 15 s and centrifugation at 13,000 g for 15 minutes at 3 °C. 330 µL of supernatant was then transferred to a 2 mL microcentrifuge tube. Samples were dried by speed vacuum concentration (Thermo Scientific Savant SpeedVac Concentrator SPD111V) at room temperature for 4 hours and stored at -80°C until analysis. Just before analysis samples were reconstituted in 100 µL of HPLC grade methanol/water (50:50), vortexed for 30 s and centrifuged at 13,000 g for 15 minutes at 3 °C. The supernatant was then transferred to HPLC vials and analysed. Simultaneously a single pooled QC sample was prepared by aliquoting and mixing 80 µL from each biological sample. This pooled sample was aliquoted into 24 different QC samples which underwent the same extraction procedure as described above.

Biphasic extraction was performed as follows. Samples were thawed at room temperature. 200 µL was then transferred to a 2 mL microcentrifuge tube and placed on ice. 600 µL of HPLC grade methanol/chloroform (50:50, stored at -20°C) was added followed by vortex mixing for 5 minutes. 300 µL of HPLC grade water was then added to induce phase separation followed by vortex mixing for 15 seconds and centrifugation at 13,000 g for 15 minutes at 3 °C. Two aliquots of 100 µL of the lower chloroform phase were then transferred in to two 2 mL microcentrifuge tubes. Two aliquots of 300 µL of the upper methanol/water phase were transferred in to two 2 mL microcentrifuge tubes. The chloroform phases were dried under a nitrogen gas stream while methanol/water phases were dried by speed vacuum concentration at room temperature for 4 hours. Extracted samples were stored at -

80 °C until analysis. Just before analysis samples were reconstituted in 100 µL of HPLC grade methanol/water (50:50), vortexed for 30 s and centrifuged at 13,000 g for 15 minutes at 3 °C. Simultaneously, a single pooled QC sample was prepared by aliquoting and mixing 120 µL from each biological sample. This sample was aliquoted into 48 different QC samples which underwent the same extraction procedures as described above.

Urine samples were extracted through deproteinisation. Samples were thawed at room temperature. 100 µL of urine was then transferred to a 2 mL microcentrifuge tube and placed on ice. 300 µL of HPLC grade methanol were added followed by vortex mixing for 15 s and centrifugation at 13,000 g for 15 minutes at 3 °C. 330 µL of supernatant was then transferred to a 2 mL microcentrifuge tube. Samples were dried by speed vacuum concentration at room temperature for 4 h and stored at -80 °C until analysis. Just before analysis samples were reconstituted in 50 µL of HPLC grade water, vortexed for 30 s and stored at room temperature for 10 minutes. Samples were then vortexed mixed and 50 µL of HPLC grade acetonitrile was added. Subsequently samples were vortex mixed and centrifuged at 13,000 g for 15 minutes at 3 °C. The supernatant was then transferred to a HPLC vial and analysed. Simultaneously a single pooled QC sample sample was prepared by aliquoting and mixing 100 µL from each biological sample. This sample was aliquoted into 20 different QC samples which underwent the same extraction procedure as described above.

Serum monophasic extracts and biphasic extracts were analysed on a UPLC-MS system (Dionex Ultimate 3000 and Thermo Scientific Q-Exactive, Thermo Scientific Hemel Hempstead, UK) using a reversed phase C_{18} method (Thermo Scientific Hypersil GOLD 100 x 2.1 mm 1.9 µm). Solvent A was water + 0.1% formic acid and solvent B was methanol + 0.1% formic acid. The gradient elution conditions were: from 0 to 1 minute (A was constant at 100%), from 1 to 4 minutes (B increased from 0 to 100%); from 4 to 7 minutes (B was kept constant at 100%); from 7 to 8 minutes (A increased from 0 to 100%); from 8 to 10 minutes (A was constant at 100%). All of the column eluent was introduced in to the electrospray ion source of the mass spectrometer. The mass spectrometer

operating conditions were as follows: resolution was set to 70000 (FWHM at m/z 200), AGC target to 3e6, S-lens RF level to 80 and m/z range of 100-1000Da. Ten conditioning QC samples were injected at the start of the run. Four of the QC samples were analysed applying a HCD MS/MS data dependent analysis method exploring different m/z ranges (100-300Da; 300-500Da; 500-750Da; 750-1000Da) with a stepped collision energy of 30-80%. After the first ten QC samples a QC sample was injected after every 4th biological sample in order to monitor the signal and to perform a quality filtering on the data[184].

Urine extracts were analysed on the UPLC-MS system described above using a HILIC method (Thermo Scientific Accucore 150-Amide 100x2.1 mm 1.9 µm). Solvent A was 95% acetonitrile + 5mM ammonium formate at pH 3 and solvent B was water + 5mM ammonium formate at pH 3. The gradient elution conditions were: from 0 to 1 minutes (A was constant at 95%); from 1 to 12 minutes (B increased from 5 to 45%); from 12 to 15 minutes (B was kept constant at 45%); from 15 to 16 minutes (A increased from 55% to 95%); from 16 to 21 minutes (A was constant at 95%). All of the column eluent was introduced in to the electrospray source of the mass spectrometer. The mass spectrometer operating conditions were identical to the ones applied for the analysis of serum samples. Ten conditioning QC samples were injected at the start of the run. Four of those were analysed with a HCD MS/MS data dependent analysis method exploring different m/z ranges (100-300Da; 300-500Da; 500-750Da; 750-1000Da) with a stepped collision energy of 30-80%. After the first ten QC samples a QC sample was injected after every 6th biological sample.

The .RAW files produced by the software Xcalibur 3.0.63 were converted to .mzML files with ProteoWizard 2.1 followed by deconvolution and peak alignment applying XCMS as described previously[234]. Metabolite features (m/z-RT pair) bearing more than 30% missing values were removed from the dataset. If the biological class of a feature was still populated by missing values those were substituted by a small value identified as the half of the minimum intensity of the

dataset. Subsequently quality check filtering was performed, all of the metabolite features reporting a relative standard deviation for the response greater than 20% for all QC samples (starting from QC 8) were removed. The dataset was then normalised by sum and univariate analysis was performed applying paired Mann-Whitney U tests using the base stats package in R[261]. Subsequently, missing values were imputed by Random Forest (RF) replacement for PCA and by KNN replacement for PLS-DA. Glog transformation was applied before multivariate analysis. At this point the dataset was analysed applying PCA and PLS-DA (optimal number of components was evaluated performing a 10-fold cross-validation). Metabolite annotation was carried out using the PUTMEDID_LCMS workflow implemented on Taverna 1.7[147]. After annotation the statistically significant metabolite features have been labelled and grouped according to their chemical structure or biological class.

CHAPTER 3 - Assessment of normalization, missing value imputation, scaling and transformation methods in LC-MS-based non-targeted metabolomics

The present study has already been published in Metabolomics in 2016[225] being also awarded as runner-up for the Metabolomics Best Paper Award.

Mass spectrometry (MS) is, the most frequently applied analytical technique in metabolomics experiments for non-targeted analysis[164,262]. A Pubmed search for example, retrieved 1100 results using the words “metabolomics” AND “NMR” for the period from January 2013 to June 2016 and retrieved 2314 results using the words “metabolomics” AND “chromatography” for the same time period. In fact, the expansion of MS as the main technique for non-targeted metabolomics studies is witnessed by 1013 papers published in 2015 browsed on the PubMed database typing “mass spectrometry” AND “metabolomics” and 464 papers browsed typing “mass spectrometry” AND “metabolomics” AND “liquid chromatography”.

Chromatographic techniques have contributed to this prevalence with significant benefits providing enhanced sensitivity and metabolome coverage. The development of UPLC, coupled with the constant advances in MS sensitivity and mass resolution, has provided significant advantages to non-targeted metabolomics studies. Moreover, the use of chromatography allows the potential separation of isomeric compounds and provides a unique signature for every metabolite (m/*z* and retention time) that can ease the metabolite identification process. The use of ionisation techniques such as MALDI opened new research pathways allowing the imaging of samples with high spatial resolution[263]. Instruments like the triple quadrupole mass spectrometer are routinely used for targeted metabolite detection and quantification due to their enhanced sensitivity and specificity[264]

(they are acknowledged as the “gold standard” for MS targeted assays). Hence, UPLC-MS has a major role in targeted and non-targeted metabolomics studies.

As already discussed in chapter 1 there are multiple experimental and bioinformatics processes that are required when working with non-targeted UPLC-MS metabolomics datasets (**Figure 12**).

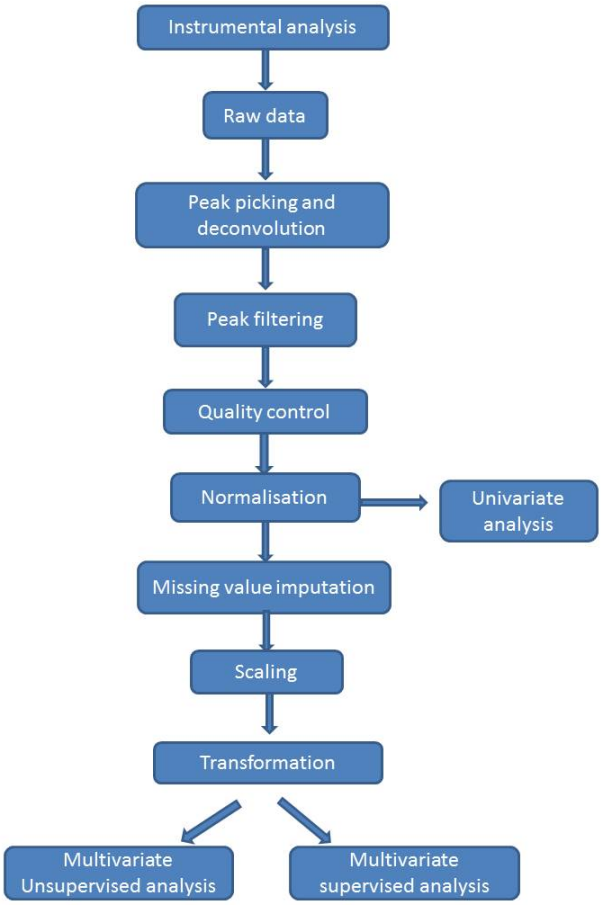


Figure 12. The common data treatment stages in a non-targeted metabolomics experiment

Once the data have been collected, they are available in a raw instrument data format. This format is usually not human-readable and includes all the information for a specific sample. While this type of data can be visualized through dedicated software, it needs to be converted to other formats (.mzML is widely adopted) in order to be human-readable and further processed. The human-readable files, which usually comply with the XML standard, can then be processed in order to identify genuine

metabolite feature peaks within and across different samples. Several software are available to perform this process: MetAlign[265], MZmine[266,267] and XCMS[268]. XCMS for example, produces a matrix containing *m/z* values, retention times and chromatographic peak intensities. The primary objective of XCMS is to detect, align and group peaks which are present in different samples. The peak alignment and grouping in XCMS can be tuned with a variety of parameters and different groups have performed studies in order to optimise the appropriate parameters for different experiments; for example, IPO, an automated tool for optimisation of metabolite detection, has also been published[269]. Alternatively, dilution series have been analysed with varying parameters and correlations between peaks have been assessed to identify the optimal conditions that lead to the integration of genuine peaks[270]. Once the data have been processed through XCMS, a matrix describing the dataset is obtained. This matrix reports all the metabolite features detected along with their retention time, *m/z* and chromatographic peak intensity. Metabolite features do not necessarily correspond to unique metabolites since they could include multiple features all generated by the same metabolite. Furthermore, multiple metabolite feature peaks may refer to the same compound as different isotopes, adducts and fragments; hence they need to be identified and integrated to a unique metabolite; this process is known as deconvolution. The high instrument sensitivity, medium throughput and complexity of samples mean the obtained datasets are often large (1000s of metabolite features and 100s of samples).

The dataset obtained can contain a high percentage of missing values depending on the threshold used to identify and report a peak across samples. Peaks with a high percentage of missing values could be interpreted as not being robustly detected and are often removed from the dataset.

The use of QC data (technical replicates) allows for the measurement and reporting of the reproducibility of the analytical processes for each peak. For example, peaks with a relative standard deviation for QC samples exceeding a certain threshold on the QCs can be removed from the

dataset[183]. Furthermore, the QC also provide an effective method to correct drifting signal during the analysis and to condition the analytical instrument with a sample that offers the same chemical properties and characteristics of the "real" samples.

Since different sample concentrations, matrix effects, errors in injected sample volume and changes in instrument sensitivity across an analytical run can change the intensity of metabolite features, data needs to be normalised prior to data analysis. Normalisation can be performed with or without an internal standard reference. Different methods involving the use of internal standards have been developed by De Livera *et al*.[271] using linear algebraic transformations. Other normalisation methods, which are not based on internal standards, use the sum, the mean or the median of the values across one sample as a normalising factor[236]. Two commonly applied methods for mass spectrometry data are (1) normalisation by total ion current where every feature in a sample is divided by the total sum of the values for that sample and (2) probabilistic quotient normalisation (PQN) where a reference sample is used to produce a normalisation coefficient for all the other samples[246].

The effect of dilution (and thus the need of normalisation) becomes more critical in cases such as the analysis of urine samples. In fact, the urine water concentration in a human subject can be very variable depending on hydration level. Moreover, urine displays different biological responses depending on the dilution factor; for instance, calcium oxalate crystallisation has been proven to occur more in concentrated urine[272]. A study by Lusczek *et al.*[273] tested seven different normalisation methods on pig urine including: two methods based on constant sum, one on internal standard normalisation, two on PQN, one on urine osmolality and one on urine output. The last two methods proved to be particularly efficient for coping with urine dilution. Osmolality normalisation proved to also be preferable to creatinine normalisation[274]. Nevertheless data on osmolality and urine output are not always available so the use of other normalisation methods may be recommended depending on the case and available metadata. It must be noted though that normalisation by creatinine concentration is still a widely adopted method in targeted assays[275].

One of the major issues encountered during data pre-treatment is the presence of missing values. Some peak processing software can impute the missing values present in the dataset although the values are usually replaced by a value representative of the noise signal. Missing values may have been produced due to different causes. In fact, their incidence can be random or systematic and in both cases the cause may be biological or technical[235,237]. Specifically:

- The analyte may be present at a concentration which is above the analytical limit of detection (LOD) but not reported by the software for technical reasons (e.g. ionisation suppression)
- The analyte may be present in the sample but at a concentration which is below the analytical LOD thus will not be detected
- The analyte may not be present at any concentration in the sample (biological reason).
- The analyte is detected by the UPLC-MS assay but is not reported by the data processing software.

It is worth noting missing values can be classified in three classes[276]: missing completely at random (MCAR) when the occurrence of missing values is not related to any other variable, missing at random (MAR) when the occurrence is somehow related to one or more variables (excluding the response) and missing not at random (MNAR) when the occurrence of missing values is related to the response itself. In biological experiments a variable may contain missing values because the analyte is not present in a class due to specific biological reasons: in this case it will be accounted as MNAR and the occurrence of missing values can be described as systematic. However, as already mentioned, a peak may be present but not detected due to technical reasons, in this case the value could be MAR or MCAR depending on the case. Different treatment methods would imply different biological interpretations and more importantly different statistical outcomes. Some studies report filtering of all features displaying more than a certain percentage of missing values, the percentage of missing values threshold varies[237,277,278]. The issue associated with missing values is also connected with the invalidity for many multivariate methods such as PCA and PLS-DA to analyse datasets containing

missing values making it necessary to find a suitable method to impute missing values. With regards to this problem, several missing value imputation (MVI) methods have been applied in the 'omics area. Probabilistic principal component analysis (PPCA) was evaluated as a valid method to process datasets with missing values since it imputes the missing values through an iterative multivariate function before computing the PCA[238]. Other multivariate methods have been performed in order to cope with the occurrence of missing values in 'omics studies[279-282] even though there are imputation methods that do not apply prior multivariate analysis[236]. Some methods have addressed the issue by imputing missing values with the same small arbitrary value[283], with the mean[284] or median[284] of a particular feature or using more complex algorithms like the k-nearest neighbour (KNN) algorithm[285], Bayesian Principal Component Analysis (BPCA)[286], multivariate imputation by chained equation[287] and Sangster's method[245]. It is still debated whether to perform missing value imputation before or after other processing steps (normalisation, scaling, transformation). A common practice is to perform imputation right after normalisation and before any data transformation. It is also important to define whether missing value imputation should be performed before univariate analysis. The present chapter discusses these issues. As the missing value imputation affects all the latter steps of the metabolomics pipeline it is extremely useful to identify the best method to perform in order to obtain the most reliable analysis. Hrydziuszko *et al.*[237] compared different commonly used missing values imputation methods for Direct Infusion Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry (DI FTICR-MS). Similarity between outcomes produced by different imputation methods applied to dummy datasets was evaluated. The results pointed out k-nearest neighbour as the most reliable imputation method for DIMS metabolomic datasets. As far as we know, this is the only work reporting a thorough comparison between different missing values imputation methods concerning DIMS metabolomics. A study by Gromski *et al.*[223] explored the effects of different missing value imputation methods on classification performance in GC-MS studies. In fact, the issue of coping with missing value imputation has been investigated in other 'omics areas such as genomics and

proteomics. Troyanskaya *et al.*[285] examined the impact of missing values in genomics studies on statistical parameter evaluation, while the effect of missing values handling on univariate and multivariate statistics was studied by Scheel et *al.*[288] for genomics and Pedreschi *et al.*[289] for proteomics, respectively.

The final step in data pre-processing of non-targeted metabolomics datasets manages the large differences in abundance of different metabolites. Indeed multivariate analysis can be heavily influenced by abundant metabolites while minor compounds (that might well have biological relevance) may not be reported as discriminant in the analysis preventing the identification of metabolite changes. Scaling and transformation methods have been developed to overcome this issue. The main point pursued by such procedures is to reduce the heteroskedasticity of the dataset. Concerning scaling procedures, autoscaling, Pareto scaling, range scaling and vast scaling have been adopted so far for data treatment[290-294]. However, transformation has almost always been performed through logarithmic equations, often adding a constant value to the argument in order to cope with near-zero values. Different scaling and transformation methods have been tested on GC-MS datasets by van Den Berg *et al.*[247] and a comparison between autoscaling and Pareto scaling was carried out on a UPLC-MS dataset by Masson P. *et al.*[295]. These studies pointed out autoscaling and range scaling as the best scaling methods to apply in metabolomics.

Despite some studies having been published regarding the best fit-for-purpose method for missing value imputation[223,237,296] (DIMS untargeted metabolomics and GC-MS untargeted metabolomics), scaling[247] (GC-MS untargeted metabolomics) and normalisation[297] (simulated data) no holistic study has ever been performed in order to investigate and propose the best data treatment pipeline involving all processes for UPLC-MS non-targeted metabolomics.

Indeed a random query of papers (a selection of 51) published in 2014 and 2015 using the search terms 'metabolomics' AND 'liquid' AND 'chromatography' revealed extreme diversity between workflows adopted by different research groups in metabolomics. However the majority of papers

completely omitted any information about the pre-processing methods adopted or did not include key information.

Processing method	**Methods applied (number of papers reported)**
Normalisation	PQN (2), normalization to IS (9), SUM (3), mean/median normalisation (3), other (3), not reported (31)
Missing value imputation	KNN (1), mean replacement (1), small value replacement (3), not reported (46)
Transformation	Natural logarithm (12), Log2(5), Log10(1), other (1), not reported (32)
Scaling	Autoscaling (11), Pareto scaling (6), level scaling (1), not reported (33)

Table 5. Different pre-processing methods applied in a selection of 51 papers retrieved through Pubmed using search terms: 'metabolomics' AND 'liquid' AND 'chromatography'. PQN: probabilistic quotient normalisation, IS: internal standard, SUM: total sum normalisation, KNN: k-nearest neighbour.

The present PhD project relies solely on UPLC-MS methods and therefore an assessment study on UPLC-MS pre-treatment routines was performed in order to define the most appropriate data processing methods to apply throughout the PhD project. The aim of this work is to provide a consistent workflow that can be adopted across different research groups enhancing the reproducibility of the experimental data and reporting standards.

3.1 Results and Discussion

3.1.1 MVI study

Firstly, the AS (mouse serum), PT (placental tissue), UR (urine) and FP (fibroblast fingerprint) datasets were filtered as described in section 2.1.3 and normalised by sum.

All datasets decreased after peak filtering: while the missing values percentage spanned the range 8.7-15.0% before filtering, after peak filtering (20% threshold on missing values) the range was 2.8-5.0%.

Dataset	Metabolites before filtering	Metabolites after filtering	Missing values before filtering (%)	Missing values after filtering (%)	Pearson coefficient (NA vs mean abundance) [p-value]	Pearson coefficient (NA vs m/z values) [p-value]	Pearson coefficient (NA vs retention time) [p-value]
AS	4435	2996	15	4.5	-0.05 [$3x10^{-4}$]	0.07 [$2x10^{-6}$]	0.02 [0.24]
PT	3412	2622	10.2	2.8	-0.12 [$4x10^{-13}$]	-0.02 [0.29]	-0.11 [$3.78x10^{-11}$]
UR	3823	2684	14	5	-0.05 [$3.2x10^{-3}$]	0.30 [$2.2x10^{-16}$]	0.03 [0.05]
FP	2008	1598	8.7	3.7	-0.08 [$4x10^{-4}$]	0.35 [$2.2x10^{-16}$]	-0.07 [0.003]

Table 6. Summary of the features related to missing values occurrence for every dataset. AS = Mouse serum; PT = Placental Tissue; UR = Urine; FP = Fibroblast fingerprint

The Pearson coefficient related to the missing values occurrence vs. the mean abundance of the metabolite features showed no particular pattern or recurrence for any of the analysed datasets (-0.05 for AS dataset, -0.12 for PT dataset, -0.05 for UR dataset and -0.08 for the FP dataset). The same outcome was observed for the relationship between missing value occurrence and *m/z* value (0.07 for AS dataset, -0.02 for PT dataset, 0.30 for UR dataset and 0.35 for the FP dataset) and missing value occurrence and retention time (0.02 for AS dataset, -0.11 for PT dataset, 0.03 for UR dataset and -0.07 for the FP dataset) (**Table 6**). Although the Pearson correlation did not highlight any particular trend, some clusters of missing values can be identified in **Figure 13** for Placental Tissue and Fibroblast Fingerprint datasets. The occurrence of missing values seems to be dependent on the chromatographic profile of the samples thus it can be interpreted as sample specific. Specifically for these two datasets, the missing values seem to lack around 400 and 200 seconds, respectively. The correlation plots for the other datasets and the correlation plots relative to mean abundance and *m/z* are available in **Appendix Figure 1**. In conclusion, the missing values were then treated as occurring completely at random in the dataset. Moreover, this outcome suggests that the occurrence of missing values is mainly due to software nuisance rather than associated with instrumental errors at high *m/z* values or sample composition.

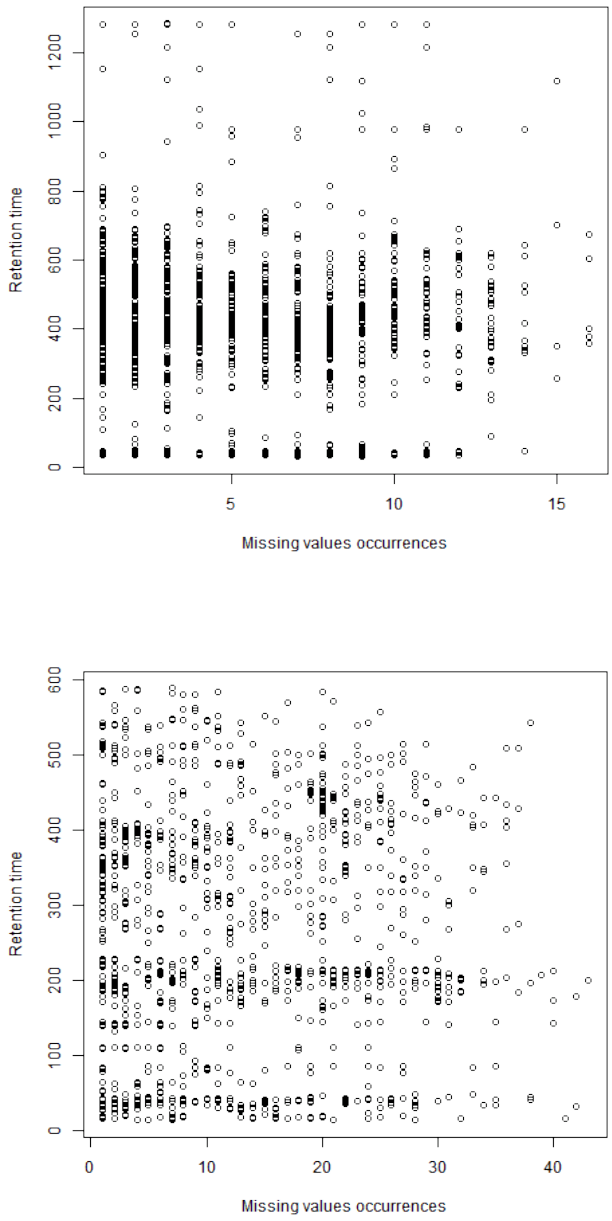


Figure 13. Plots correlating occurrences of missing values with retention time for PT dataset (top) and FP dataset (right)

After preliminary descriptive analysis, the datasets were then filtered and missing values were randomly introduced as described in section 2.1.5. Concerning the MVI performance, the results are shown in **Table 7**.

NMRSE	MS dataset	PT dataset	UR dataset	FP dataset
SV	9.99	3.64	7.66	5.53
MN	1.82	0.66	1.47	1.01
MD	1.60	0.68	1.49	1.01
KNN	1.29	0.58	1.44	0.54
BPCA	1.30	0.62	1.49	1.12
RF	0.75	0.45	1.16	0.37

Table 7. NMRSE values for every dataset with the five different imputation methods. NMRSE: Normalised mean root square error; MS: mouse serum; PT: placental tissue; UR: urine; FP: fibroblast fingerprint; SV: small value replacement; MN: mean replacement; MD: median replacement; KNN: k-nearest neighbour replacement; BPCA: Bayesian principal component analysis; RF: random forest.

The results were consistent with the conclusions reported by Hrydziuszko[293] for DIMS data. SV achieved the worst performance (highest NRMSE) among the methods tested. The reason for the poor performance of SV is to be assumed that the missing values occurring in the dataset were due uniquely to low concentration metabolites. This assumption does not reflect real experimental conditions, as missing values will occur because of technical or biological reasons. In fact the only case in which SV could perform efficiently would be in datasets in which missing values are correlated with the mean response. However the correlation data reported above excludes such a hypothesis. Conversely, RF achieved the lowest NRMSE error in all datasets proving to be the best imputation method when dealing with MCAR. This result complies with the findings shown by Gromski *et al.*[223] for GC-MS. Indeed, the KNN method generated low NMRSE values (**Table 7**) between original and predicted values. Although RF is the best imputation method tested, one of its major drawbacks is the computational time. For instance, the application of such a method in the present study took 15 minutes per dataset. Hence, when dealing with data analysis that is required to be fast, other methods should be preferred. Specifically, KNN replacement seems to perform markedly better than other methods (excluding RF) in all four datasets. Although the results showed BPCA and KNN

performances as similar, it must be highlighted that in general the computation time for BPCA is consistently higher when compared to KNN. Therefore the application of RF is strongly recommended in general case scenarios, however, KNN is recommended when the analysis is heavily dependent on time (e.g. multiple analysis in which some parameters are slightly changed). Also, the use of SV is highly discouraged. On a final note, this assessment dealt primarily with missing value imputation methods for UPLC-MS experiments affected by technical nuisance caused by software inability to detect analytes. Indeed the main assumption of the simulation is that the missing values are not caused by the actual lack of a specific metabolite; this assumption does not always hold true in real-life experiments.

3.1.2 Pre-processing study

The 46 biological samples from the dataset described in section 2.2.8 were assigned to a random class as described. This process produced 22 samples assigned to class A and 24 to class B. Before any further modification of the dataset, Mann Whitney U tests and Benjamini Hochberg correction were applied in order to identify any statistically significant difference between the two classes for each metabolite feature. Univariate analysis did not show any metabolite feature as statistically significant ($p<0.05$). PCA was performed on the dataset and showed no *a priori* class clustering (**Figure 14**). PLS-DA did not produce a validated method.

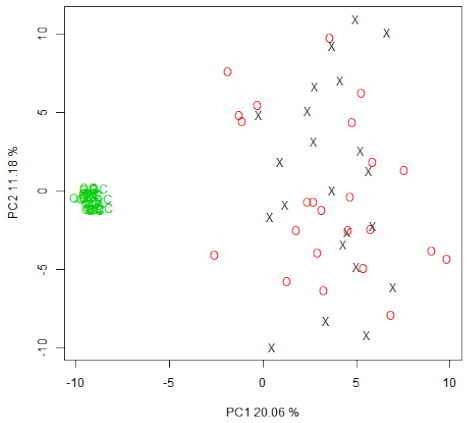


Figure 14. PCA of the dataset used for the pre-processing study before applying any modifications to the randomly assigned classes. Class 1 samples are reported as X and class 2 samples are reported as O. QC samples are in green.

The metabolite features were then ordered according to mean response and divided in to three blocks (low, medium and high response). From each block 32 metabolite features were randomly chosen. Each of the 32 metabolite features was then multiplied by a factor ranging between 0.1 and 10 (class A only). This procedure generated 96 metabolite features belonging to three different equally populated response blocks. Three metabolite features (one for each block) bore a fold change of 0.1 between class A and class B, three bore a fold change of 0.2 and so on. The dataset obtained contained 7.8 % missing values with 65.3 % of features bearing at least one missing value. The dataset underwent assessment for a normal distribution applying the Shapiro-Wilk test in order to choose the appropriate univariate method to apply (parametric or non-parametric). The high percentage of non-normally distributed metabolite features even after logarithmic transformation (**Table 9**) demonstrated that the non-parametric Mann Whitney U test is the method of choice for univariate analysis. It must be highlighted that parametric statistical methods are widely used in research; however, these methods require an underlying normal distribution of the data. This

condition is not always tested before performing parametric methods. **Table 8** shows that it is not appropriate to assume this assumption even when the sample size is not small (77 samples).

Normalisation	**KNN missing value imputation/glog transformation**	**Normally distributed metabolite features (%)**
SUM	Not performed	39.6
	Performed	44.3
PQN	Not performed	41.6
	Performed	50.7

Table 8. Percentage of normally distributed metabolite features following different pre-processing procedures

All the pre-processing methods described in 2.2 were applied to the dataset according to all the possible permutations. Considering that normalisation had two levels (PQN and SUM), missing value imputation had seven levels (none, SV, MN, MD, KNN, BPCA and RF), transformation had four levels (none, glog, ihs, nlog) and scaling had five levels (none, autoscaling, pareto scaling, range scaling and vast scaling), the permutations are 5x4x7x2=280 experiments. All of these 280 experiments were explored and the univariate and multivariate outcomes for the 96 modified metabolite features were investigated

3.1.2.1 Univariate analysis

The results are described in **Appendix Table 2** and are summarised in **Table 9**. Regardless of the data treatment permutation, the metabolite features imputed with low fold change (multiplication factor between 0.8 and 1.2) were not recorded as statistically significant while the p-values related to

medium and high fold change metabolite features gradually decreased, respectively (i.e. became more statistically significant). Although this trend was general, some variations occurred in specific cases.

Normalisation	MVI	Metabolite features out of 96 imputed	Total number of metabolite features	False positives
SUM	none	80	80	0
	SV	77	77	0
	MN	79	79	0
	MD	79	79	0
	KNN [1]	81	84	3
	KNN [2]	81	84	3
	BPCA [1]	81	82	1
	BPCA [2]	81	84	3
	BPCA [3]	81	84	3
	RF	81	81	0
PQN	none	81	81	0
	SV	77	77	0
	MN	80	80	0
	MD	80	80	0
	KNN	82	87	5
	BPCA [1]	82	88	6
	BPCA [2]	82	88	6
	RF	81	81	0

Table 9. Summary of the univariate analysis. The third column indicates the number of imputed metabolite features detected as significant ($p<0.05$ following FDR correction) while the fourth displays the total number of significantly changing metabolite features detected.

The univariate outcome was affected by the normalisation method and missing value imputation method applied but very limited effect was observed caused by scaling or transformation. Indeed the univariate analysis performed on datasets normalised by PQN (and subsequent missing value imputation methods) produced identical outcomes regardless of the transformation or scaling method applied. The only exception is represented by the BPCA MVI method which generates two

different outcomes: [1] produced after nlog transformation (regardless of scaling method applied); [2] generated by all the other transformation and scaling methods.

The results are slightly different concerning SUM normalisation. In fact, there is no scaling or transformation effect for five missing values imputation methods (none, SV, MN, MD and RF) while KNN and BPCA produced diverse outcomes dependant on transformation and scaling. Specifically, KNN imputation generates two different sets of p-values: [1] for glog and nlog transformation and [2] for all the other transformation methods. Conversely, BPCA replacement generates three different results: [1] obtained after glog and subsequent range scaling or autoscaling or vast scaling; [2] nlog transformation followed by no scaling/Pareto scaling/range scaling/autoscaling or glog and subsequent no scaling/Pareto scaling and [3] for all the other transformation methods.

Following SV, MN, MD and RF the statistically significant metabolite features were picked entirely from the 96 imputed features whereas following KNN and BPCA replacement a certain number of false positives was generated.

Overall, SUM normalisation seems to be slightly less conservative than PQN, this method results in a lower percentage of false positive results obtained through KNN and BPCA replacement. However it has been described[298] that SUM normalisation does not cope well with large differences bore by a few high intensity metabolites therefore indicating PQN as the optimal normalisation method to apply in univariate analysis.

Regarding the missing value imputation methods which do not generate false positive results (none, SV, MN, MD and RF). The highest number of correctly identified statistically significant metabolite features was achieved by no missing value imputation and RF while the lowest number was achieved through SV (**Table 9**). It is notable that following SUM normalisation, RF performed even better than no imputation (although presenting just one more true positive). In conclusion, it seems that imputation before the univariate analysis with methods different from RF could potentially

compromise the outcome. It must be noted that this outcome does not state an absolute rule while it rather offers guidance to the user since the treatment performed for univariate analysis is heavily affected by the purpose of the study.

Commonly, if a dataset presents a high percentage of features not following normal distribution it is assumed (sometimes falsely as **Table 8** shows) that logarithm transformation is able to modify the data distribution in to a normal distribution thus justifying the use of parametric statistical tests. Hence, besides Mann Whitney U test, Student's t-test was applied on the dataset and the outcome investigated (data shown in **Appendix Table 3** and summarised in **Table 10**). Although in this case the transformation step has a stronger effect (each separate transformation process produces a different univariate outcome), the general trend follows the results produced with the Mann-Whitney U test. In fact, the features with a multiplication factor between 0.8 and 1.2 were still observed to not be statistically significant in all three response blocks. Furthermore, KNN and BPCA still produced the highest number of false positives while no missing value imputation achieved 80-81 significant metabolite features for SUM normalisation and 81-82 metabolite features for PQN always while keeping the false positive rate at 0 %.

RF achieved similar results compared to no missing value imputation although the combination of SUM normalisation and nlog transformation produced a false positive. Therefore, the same conclusions described before for use of non-parametric statistical methods can be described for parametric statistics. Hence for both parametric and non-parametric univariate methods it is suggested to use no imputation or RF imputation. Following these procedures in fact, a number of metabolite features is detected as statistically significant whilst the false positive rate remains at 0%. Conversely, normalisation, scaling and transformation methods do not show effect on the univariate outcome.

Normalisation	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM
MVI	none	none	none	SV	SV	SV	MN	MN	MN	MN	MD	MD	MD
Transformation	glog	ihs	nlog	none	glog	nlog	glog	ihs	nlog	none	glog	ihs	ihs
Scaling	All	All	All	All	All	All	All	All	All	All	All	All	All
True positives	80	80	81	80	75	75	75	80	80	81	80	81	81
False positives	0	0	0	0	0	0	0	0	0	0	0	1	1

Normalisation	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM
MVI	MD	MD	KNN	KNN	KNN	BPCA	BPCA	BPCA	RF	RF	RF	RF	RF
Transformation	nlog	none	glog	ihs	nlog	glog	ihs	nlog	glog	ihs	nlog	nlog	none
Scaling	All	All	All	All	All	All	All	All	All	All	All	All	All
True positives	81	81	84	83	85	83	82	81	83	81	80	82	80
False positives	0	1	4	3	4	3	2	1	2	1	0	1	0

Normalisation	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN
MVI	none	none	none	none	SV	SV	SV	MN	MN	MN	MN	MD	MD
Transformation	glog	ihs	nlog	none	glog	ihs	none	glog	ihs	nlog	none	glog	ihs
Scaling	All	All	All	All	All	All	All	All	All	All	All	All	All
True positives	81	81	81	82	76	68	68	77	81	81	81	82	81
False positives	0	0	0	0	0	0	0	0	0	0	0	1	0

Normalisation	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN	PQN
MVI	MD	MD	KNN	KNN	KNN	BPCA	BPCA	BPCA	RF	RF	RF	RF	RF
Transformation	nlog	none	glog	ihs	nlog	glog	ihs	nlog	glog	ihs	nlog	nlog	none
Scaling	All	All	All	All	All	All	All	All	All	All	All	All	All
True positives	81	82	85	84	84	85	84	82	85	86	81	80	81
False positives	0	1	4	3	3	3	3	3	4	4	0	0	0

Table 10. Summary of the results of the t-test after different permutations

3.1.2.2 Multivariate analysis

3.1.2.2.1 PCA outcome

The overall purpose of this study was to identify data treatment methods affected by fold change alone, by abundance alone and by a combination of both. Every permutation was recorded in terms of PCA loadings ranked for the 96 imputed metabolite features, fold change effect and abundance effect on PC1 and PC2 as described in section 2.2.11. The results are shown in **Appendix Table 5**. Data treatment permutations showing an effect of fold change and no effect of abundance on PC1 and PC2 were selected. This selection reduced the 280 permutations to 32 entries. These entries were then ranked according to the amount of explained variance on the first two PCs. The first ten positions are presented in **Table 11**.

Normalisation	MVI	Transformation	Scaling	Variance on PC1	Variance on PC2	Pvalue on PC1	Pvalue on PC2
SUM	RF	glog	none	42.7	39.33	4.82E-12	2.03E-05
SUM	none	nlog	none	55.29	23.94	5.28E-08	8.47E-06
PQN	none	ihs	none	55.3	23.6	1.64E-06	9.06E-07
PQN	none	nlog	none	55.3	23.6	1.64E-06	9.06E-07
PQN	none	glog	none	54.1	23.92	1.54E-07	1.89E-06
SUM	none	glog	none	51.66	25.29	1.14E-11	5.01E-06
SUM	RF	nlog	none	44.73	28.63	1.66E-08	3.27E-05
PQN	RF	ihs	none	44.69	28.59	1.83E-07	7.44E-06
PQN	RF	nlog	none	44.69	28.59	1.83E-07	7.44E-06
PQN	RF	glog	none	43.89	28.74	1.66E-08	1.24E-05

Table 11. First ten permutations for PCA ordered by explained variance on PC1 and PC2

The treatments occupying the first positions all contained transformation followed by no further scaling independent of the normalisation method. While the first ten positions imply the use of RF imputation or no imputation at all, the positions between 11 and 20 are characterised by the use of KNN and BPCA as MVI methods. Notably SUM normalisation followed by no missing value imputation and glog achieves high values of explained variance on PC1 and PC2 and a low p-value representing separation for PC1. PCA on datasets that did not undergo missing value imputation was computed through the NIPALS algorithm. This algorithm effectively provides an alternative method for imputation, which is required for PCA. Although this method appeared to work well with PCA, it presents its own drawbacks. Firstly, computation time for PCA is much longer if compared with an already imputed dataset; secondly, it is often essential to relate a PCA model with a PLS-DA model, however PLS-DA computation on non-imputed datasets is often difficult (cross-validation presents several impediments). Hence, imputation methods like RF, KNN and BPCA, which still achieve good separation on PC1 and PC2, appear to be a valid alternative. Generally, permutations including RF as the MVI method and transformation followed by no scaling appear as the optimal combination for PCA. Notably, the only MVI method that does not appear in the first 20 positions is SV. Looking at the p-values related with separation along PC1 and PC2, glog transformation appears to achieve lower and more statistically significant p-values compared to nlog or IHS and therefore better separation on PC1. Comparing the two normalisation methods, SUM seems to generate lower p-values on PC1 while PQN is associated with lower p-values on PC2.

In order to analyse the treatments affected by abundance alone on PC1, the permutations were selected according to an effect of abundance of PCA and no effect of fold change. The resulting permutations are all characterised by PQN and range scaling (apart from a few exceptions) with no particular effect of transformation applied. Strangely the methods which are affected by abundance alone on a combination of PC1 and PC2 are different. In fact SUM normalisation followed by no

transformation and no scaling or by IHS transformation and no scaling appears to be affected by abundance only on a combination of PC1 and PC2. The same outcome is obtained through PQN followed by no transformation and no scaling. It should be considered that even when an abundance effect alone is recorded, the outcome is never totally independent from a fold change effect. Looking at the number of metabolite features in the first ten positions split in to 4 classes of fold change, the "high fold change" class contains the higher number of metabolite features meaning that even if a treatment is recorded as not affected by fold change (the "high fold change" metabolite features must be at least 8) it may still show some dependence from fold changes.

Conversely, treatments affected by concentration and fold change simultaneously on PC1 and a combination of PC1 and PC2 are all characterised by Pareto scaling. These permutations are SUM normalisation followed by IHS or no transformation and subsequent Pareto scaling or alternatively PQN followed by no transformation and subsequent Pareto scaling. Notably, these treatments also achieve a good PCA separation on both PC1 and PC2, the overall ranking first positions for PCA loadings are mainly occupied by metabolite features representing the 96 imputed metabolite features.

Another remark has to be stated. Although different MVI methods do not heavily influence the effect brought about by fold change or abundance, permutations including no missing value imputation produce the highest explained variance on PC1 and PC2. Nevertheless this outcome may still be due to computational issues related to multivariate analysis on datasets containing missing values.

Studies which are focused on metabolic changes without any particular constraint should therefore apply PCA after normalisation by SUM or PQN (there is no marked difference) and glog as transformation since this combination produces good PCA clustering and is affected by fold change only and not absolute abundance. On the other hand, studies which are focused on an effect induced

by abundance and fold change together should consider the use of normalisation by SUM or PQN (there is no marked difference) and Pareto scaling without any transformation.

3.1.2.2.2 PLS-DA outcome

After selection of data treatment permutations showing an effect of abundance or fold change on variable 1 and 2 in PLS-DA, it appears that almost all permutations are affected by fold change alone on variable 1 with some exceptions (**Appendix Table 6**).

Normalisation	MVI	Transformation	Scaling	R^2	Q^2	R^2-Q^2
PQN	BPCA	nlog	Pareto	0.63	0.47	0.16
PQN	KNN	glog	none	0.61	0.46	0.15
PQN	BPCA	nlog	none	0.59	0.44	0.15
SUM	KNN	glog	none	0.59	0.51	0.08
SUM	BPCA	nlog	none	0.57	0.40	0.16
PQN	BPCA	nlog	none	0.56	0.38	0.18
PQN	BPCA	glog	Rg	0.56	0.55	0.01
SUM	KNN	nlog	Auto	0.55	0.40	0.16

Table 12 First eight permutations for PLS-DA ordered by R2 values with R^2-Q^2 less than 0.20

In fact, PQN followed by KNN imputation and ihs or nlog (the scaling has no effect) are the only permutations being affected by fold change ratio and abundance on variable 1. Conversely, on variable 2 the majority of permutations does not seem to be affected by abundance or by fold change although some combinations present an exception displaying an abundance effect. Precisely these combinations are: all the permutations including SV replacement; SUM normalisation followed by MN replacement and IHS transformation (scaling has no effect); SUM normalisation followed by MD replacement and ihs or glog transformation (scaling has no effect).

In order to investigate the classification performance, the permutations were sorted according to R^2 values (first level) and the difference between R^2 and Q^2 being less than 0.20 (**Table 12**). It is worth mentioning that the highest R^2 and Q^2 scores are obtained by applying KNN and BPCA replacement with no particular effect of normalisation. Such models though, hardly exceed values of 0.5. It is likely that models constructed on a dataset characterised by a limited number of changing metabolite features may not produce a robust validation.

It must be noted that while the PCA outcome might point out the use of glog transformation as convenient in order to highlight the fold change effect (the first positions all contain glog transformation), the PLS-DA analysis provides the same conclusion.

The main recommendation concerning PLS-DA would be to avoid the combination of PQN, KNN and ihs (or nlog) and avoid all permutations applying SV if the experiment is designed to highlight fold changes. Conversely, the permutations applying SV, SUM, MN and ihs and SUM, MD and ihs (or glog) provide an abundance and fold change effect of the second variable of the PLS-DA model. It has to be remarked though that the effect on the first variable would still be addressed to fold change. The recommended permutation would then include SUM or PQN normalisation followed by BPCA or KNN imputation and glog transformation.

This study sheds further light on issues that are often overlooked in the metabolomics community. It is likely that the limited investigation of the topic is due to technical difficulties in reproducing the pre-processing steps (lack of statistical programming knowledge, lack of dedicated software). In order to facilitate the investigation of different pre-processing methods and promote the standardisation of a common workflow, much effort was spent in integrating the described processing steps in the online platform Galaxy.

3.2 Implementation of UPLC-MS untargeted metabolomics workflow in Galaxy

3.2.1 Galaxy and its role in the scientific community

One of the major concerns of scientific research is the reproducibility of the data processing and analysis steps and their accurate reporting. Often different research groups and sometimes even different researchers within the same group undertake different processing steps, leading eventually to different outcomes. Even more problematic is the possibility of performing a common processing step in different ways. If we consider the non-targeted metabolomics analysis of urine, at least seven different normalisation methods not involving the use of an internal standards have been reported[273]. While this limitation can somehow be overcome in some scientific areas where no complex data handling is required, it becomes absolutely necessary for 'omics studies which operate with datasets consisting of several thousands of metabolite features. Therefore the standardisation of data treatment procedures inside research groups and inside scientific communities is essential. To this extent an online platform named Galaxy has been developed[299-301]. The purpose of Galaxy is to provide a common repository to store and share computational procedures and to allow computationally intensive steps to be performed on an online or local server through dedicated routines. Once the input of a processing step is uploaded in to the system and the output has been produced, this result can be published and shared online for other researchers. In this way the experimental data and the related computation can be observed by the scientific community leading to a high level of reproducibility. Moreover, Galaxy allows the users to connect different tools one after the other to obtain a comprehensive and coherent workflow which can also be shared and repeated many times in a reproducible manner. When the user wants to upload an input file from the resulting output from a previous step, all he or she needs to do is to browse through the files that have been uploaded on the current Galaxy instance.

The main instance of Galaxy[302] is suited for general purpose computational analysis and common tasks for genomics analysis such as mapping against genome libraries and RNA analysis. Conversely, tasks more specific to proteomics or metabolomics are almost totally absent. While the main instance of Galaxy is thoroughly debugged and well presented, it is possible to run instances of Galaxy on any server (hence on local machines as well). Different groups across the globe have developed their own instance of Galaxy with specific tools. In fact it is possible to upload user-developed tools on private and public instances of Galaxy, although before a tool is approved on the public instance it must be thoroughly tested and debugged. Tools can be written in a variety of programming languages which are interpreted by Galaxy such as Python, R, Java, C and Perl. Some languages which are not open-source (sometimes called proprietary languages) can also be adopted to build tools but in this case an executable must be directly available. It must be noted that "private" instances can be made publicly available; the word "private" just denotes that such an instance is not developed by the Galaxy team.

The possibility of writing your own tools makes such platform extremely flexible and powerful at the same time. In fact numerous instances are currently being developed for specific 'omics areas, Galaxy-P[303] for example is a well-established instance for proteomics analysis. The majority of tools implemented in such instances are often shared in the Galaxy ToolShed, a database where tool scripts are grouped into different repositories and available to the entire scientific community. The scripts can therefore be downloaded, tested, modified and uploaded again leading to a continuous optimisation.

Metabolomics seems to offer much space for development towards a widely applied and accepted workflow. Workflow4Metabolomics[304] is a computational infrastructure for metabolomics which uses Galaxy to serve tools which are commonly used in LC-MS, GC-MS and NMR based metabolomics. This project is supported by a consortium of several French organisations including

MetaboHub and the Institut Française de Bioinformatique. Since the use of FTP servers is enabled, it is possible to upload relatively big files (GB sizes) on to the system. This circumstance is particularly relevant for MS-based metabolomics since the files which are produced by mass spectrometers for each spectrum usually consists of many megabytes, hence the total amount of spectra for a complete experiment easily exceeds 1 GB.

The set of tools implemented on Workflow4Metabolomics spans from the uploading of raw files up to metabolite annotation including normalisation, filtering and some statistical analysis. It proved to be particularly relevant as a model for my own implementation of a comprehensive non-targeted metabolomics workflow for both DIMS and UPLC-MS. Before proceeding further it may be useful to elucidate the different processing steps which are performed in a typical DIMS or UPLC-MS pipeline. Again, since metabolomics has not reached total agreement yet on the experimental workflows, this description will illustrate the different processing steps with a broad definition (at first) without getting in to technical details which may be changing across the metabolomics community.

3.2.2 Galaxy-M

Galaxy-M is a platform for metabolomics data processing hosted on Galaxy and developed at the University of Birmingham[305]. Its purpose is to provide a standardised workflow for both DIMS and UPLC-MS untargeted metabolomics experiments. While the main focus of the project is towards a unique processing platform which can be adopted by the entire research group, it can easily be shared with the metabolomics community. The workflow is publically available in Github[306]. Once the platform is deployed, it will remove the need for local machine-dependent computations and will provide a useful GUI for each researcher who does not have familiarity with command-line statistical tools such as R or Matlab.

The platform is hosted on a Unix interface which can be easily achieved on other operating systems through virtual machines. It relies on Python, R and a standalone application which allows the execution of Matlab binaries on machines which do not have Matlab installed. This is due to the fact that Matlab is a proprietary language; hence tools written in such languages cannot be run directly from script, instead an executable file is required. Other requirements include, MI-pack[260], XCMS[219] and Wine (which allows Windows software to run on Unix-like systems). While the DIMS pipeline was already uploaded on to the platform, my work was dedicated to the integration of UPLC-MS specific processing tools and offers further developments to the published Galaxy-M paper[305].

3.2.2.1 LC-MS workflow on Galaxy-M

The first steps of the LC-MS pipeline heavily rely on XCMS, an open source package written in R which includes numerous functions for the extraction of data from chromatographic and mass raw data files. This package accepts .mzML or .mzXML data as an input and produces a .csv output with the complete experimental data matrix. Different tasks can be performed with XCMS including peak detection, peak matching, retention time alignment, peak filling and annotation (although the integration package CAMERA is required to perform annotation).

(a) Peak detection

It is possible to detect peaks in each spectrum with three different methods.

The “matched filter” method extracts chromatograms out of the spectrum for each 0.1 m/z window. It then overlaps such chromatograms and filters the resulting peaks with a second derivative Gaussian model. It is possible to modify the full width at half-maximum (FWHM) of the model which will fit the signals. The inflection points of the model cross the x-axis approximately where the peak

will then be filtered. After this step peaks less than a defined S/N ratio are removed. In this case noise is calculated as the average signal of non-filtered peaks.

The “centwave” algorithm is best suited for data collected in centroid mode. It identifies so-called regions of interest (ROI) by detecting m/z values in successive scans with less than a defined threshold of accuracy (calculated in ppm error). Subsequently each region is analysed and the peaks are fitted with a wavelet model with variable FWHM, overcoming the issues caused by a fixed width. Eventually a S/N ratio cut-off is applied.

The “MSW” algorithm applies the same wavelet model adopted in the “centwave” method but it is designed for direct infusion data.

(b) Peak matching and grouping

After peak detection and integration, features must be grouped across all samples. The grouping algorithm implemented by XCMS is based on the division of the m/z range in overlapping bins so that groups of signals belonging to the same peak are not split. Filtering is then applied in order to remove peaks which are counted twice. Hence the algorithm calculates the peak distribution bounded by a certain chromatographic time window. This distribution is then interpolated by a Gaussian model which defines the metabolite feature. The standard deviation of the model has an effect on the retention time tolerance i.e. models with lower standard deviations will tend to group features including peaks in smaller time windows.

A parameter which is crucial when grouping features is the frequency threshold: if peaks are present in less than a certain ratio of samples, the feature is discarded. When analysing samples which are heterogeneous it is appropriate to operate the grouping algorithm with a lower threshold.

(c) Retention time alignment

This algorithm is based on the grouping achieved in the previous step. It identifies features across the dataset which have a low number of samples with missing peaks or more than one peak. These features which are usually evenly distributed across the retention time range are used as "standards". For each standard the median retention time is calculated and the deviation for every sample is noted. Hence for each sample we can plot an interpolation function based on these deviations across the retention time range. Even peaks which were not picked up as standard will then use this function to correct for any retention time drift across an analytical batch.

(d) Peak Filling

XCMS provides an additional algorithm for integrating peaks which are missing in some samples. The use of the tool is often discarded in favour of missing value imputation in subsequent steps of the workflow described here.

(e) Compound extraction

Another R package named CAMERA was developed to carry out metabolite annotation[307] and it is supposed to work coupled with XCMS. In fact, the groups and the extracted ion chromatograms calculated with XCMS are a primary input for this algorithm. First the most intense feature is located and a retention time window (generally 60% of the FWHM from the centroid) is set. This window will then include other metabolite features and will be denoted as a compound spectrum. This routine is repeated until all features are included in a compound spectrum. Subsequently, within the same compound spectrum the ^{13}C isotopic pattern is searched for using an m/z difference of 1.0033. Hence correlations between features in the same compound spectrum are calculated using extracted ion chromatograms. A score matrix is then obtained as it describes the relationship between different metabolite features. Moreover a set of rules was implemented in CAMERA in order to identify common ions or neutral losses which occur during ion fragmentation. These rules are used to create

a hypothesis on the nature of a peak and if several hypothesis are coherent the peak is finally annotated.

At this point a .csv file is created with samples as rows and features as columns. Additional files are created bearing information about m/z values collected and sample names. In order to undertake further analysis, class information becomes crucial and needs therefore to be included in the workflow.

(f) File List Manager

This step simply allows the uploading of metadata as a Python script. A list with the raw file names, class information, run order and batch number must be provided. Run order and batch data are useful for signal correction algorithms.

(g) Peak filtering

If no peak filling is performed, some of the metabolite features are likely to include missing values. Metabolite features which are completely absent in a single class of samples may indicate the actual absence of a metabolite rather than a technical fault. In this case noise filling replaces these samples with noise calculated as half of the minimum of the entire dataset. Conversely, a high percentage of missing values in a feature may denote poor performance in peak picking and suggests the exclusion of the metabolite feature. Also, a high percentage of missing values in a sample may indicate technical problems related to the sample (e.g. failed injection in the instrument, poor recovery in sample extraction etc.). It is possible then to exclude these samples from further analysis. This step allows the setting of an appropriate threshold for both feature and sample filtering.

(h) Quality assurance

As noted in paragraph 1.11, one of the reasons QC samples were adopted in UPLC-MS runs is to monitor the variability of the analytical process and subsequently the data quality. These samples are multiple replicates of the same sample injected throughout the run. To this purpose it is possible to select the QC samples across the dataset and calculate the relative standard deviation (RSD) for each feature. Before doing so any QC sample used to condition the chromatographic column must be excluded. Metabolite features which vary above a certain threshold or are detected in less than a defined fraction of samples can then be removed from the dataset.

(i) Normalisation

When analysing biological samples it is common to cope with samples characterised by different concentrations. In this case statistics might indicate metabolite features which appear significantly different simply due to a different dilution factor. In order to overcome this problem a normalisation step is often required in a metabolomics workflow. It is possible to apply two methods: normalisation by sum (also known as total ion chromatogram (TIC) normalisation) and probabilistic quotient normalisation (PQN). These procedures are described in paragraph 2.2.4.

(j) Missing value imputation

At this stage, if no peak filling was performed there are still features with missing values present. Although some multivariate algorithms are able to cope with datasets with missing values (e.g. NIPALS[308]) it is much easier to process these data prior to multivariate analysis. Furthermore it must be considered that missing values may arise due to different causes. They may be caused by a metabolite which is effectively absent, by a metabolite below the detection limit, by problems with the peak picking algorithm etc.. A correct imputation of such values is therefore desirable. The imputation methods implemented are: small value replacement, mean replacement, median

replacement, k-nearest neighbour algorithm and Bayesian principal component analysis. All these methods are described in detail in paragraph 2.1.5.

(k)Transformation

Due to the presence of metabolites characterised by different intensities it is common to encounter data heteroscedasticity (different variance for each feature). This may lead to abundant features heavily affecting multivariate models without any regard to their effective change across samples. Usually a transformation step is required. In this case the generalised logarithm transformation[248] was implemented. The operating equation can be found in paragraph 2.2.6. The algorithm first isolates the QC samples in the dataset. It then iteratively varies the λ parameter and computes the glog transformation on the filtered dataset keeping track of the sum of the squared error for each iteration. The λ parameter achieving the lowest sum of squared error is then applied to the whole dataset for the glog transformation.

Except where noted, all the tools were developed using the R programming environment.

3.3 Concluding remarks

This study has highlighted relationships between normalisation, missing value imputation, scaling and transformation methods and the validity and information provided by univariate and multivariate analysis methods. Specifically, univariate analysis (regardless of parametric or non-parametric methods) should be performed after RF imputation or no imputation at all as other methods may produce false positives or be too conservative. The choice concerning other treatments (normalisation, scaling and transformation) is not crucial as these other steps did not have an effect on the univariate outcome. Hence while normalisation (by sum or PQN) should be applied at this step, scaling and transformation can be skipped.

It has been found for PCA that SUM or PQN normalisation, in combination with RF missing value imputation, glog transformation, and no scaling highlights the metabolite features with a significant fold change between classes regardless of the metabolite feature response. For PLS-DA, KNN and BPCA missing value imputation operate more effectively than other imputation methods, including RF. The highest ranked permutations with the smallest $R2 - Q2$ difference were (1) PQN normalisation, BPCA MVI, glog transformation and range scaling and (2) SUM normalisation, KNN MVI, glog transformation and no scaling. We recommend the second of these permutations for PLS-DA analysis.

CHAPTER 4 - Non-targeted UPLC-MS assessment of the metabolic interaction between insulin and cortisol or dexamethasone

4.1 Introduction

As discussed in chapter one, the interaction between glucocorticoids (GCs) and insulin is a crucial step for understanding the biochemical mechanisms operating in a variety of metabolic diseases including the metabolic syndrome (including obesity and diabetes) and Cushing's syndrome. Schilling *et al.*[309] showed that insulin has an effect on cerebral blood flow regulating appetite on humans while the effect of GCs was not detected. Griffin and Wildenthal[310] showed that hydrocortisone can have opposite effects on cardiac protein synthesis depending on the presence or absence of insulin, while hydrocortisone infusion alone at low insulin levels displayed severe reduction of protein synthesis and hydrocortisone in conjunction with insulin increased protein synthesis rate and significantly decreased protein degradation. This suggests that endogenous hydrocortisone may have an essential role in regulation of cardiac protein synthesis. Long demonstrated that GC administration is related to lipolysis and gluconeogenesis during the fasted state (absence of insulin) and to lipid storage during the fed state (presence of insulin)[41]. In more recent years studies have focused on insulin and cortisol interaction in tissue-specific investigations. For example, Gathercole *et al.*[2] detected an increase in lipogenesis in adipose tissue caused by insulin while GC administration reduced lipogenesis. However, selective inhibition of 11-βHSD1 completely negated the lipogenesis suppression induced by GCs. Moreover, low dose dexamethasone infusion (a synthetic GC capable of suppressing cortisol levels[311]) appeared to enhance the effect of insulin on lipogenesis suggesting an active role of GCs and insulin combined.

In a separate study[312] synthetic GC administration (dexamethasone) has been linked to insulin sensitisation while high dose insulin infusion with dexamethasone appeared to induce insulin resistance. Surprisingly, when high dose insulin infusions were administered in the presence of GCs the adipose cells still displayed insulin sensitisation.

An opposite conclusion was reported by Chimin *et al.*[3]. Rat visceral adipocytes appeared to display insulin resistance and increased lipogenesis after dexamethasone infusion. The authors suggest this effect may be due to an increased availability of NADPH (involved in lipogenesis) rather than an actual overexpression of enzymes involved in the pathway. Nevertheless the controversial outcome of this study has been discussed in an editorial by Vienberg and Björnholm[313] who addressed the unexpected outcome to long exposure to GC administration.

The studies discussed have been performed by investigating specific classes of metabolites with target-specific assays (for example, free fatty acids, acyl glycerides, steroids). However, non-targeted metabolomics would be a suitable technique for the global assessment of a wide diversity of metabolic classes affected by insulin and cortisol. Furthermore, a holistic assessment may overcome disputes on contrasting outcomes since it can provide a comprehensive view of the biological mechanisms occurring following administration of GCs and insulin.

The objective of the research presented in this chapter is to determine the global (in serum) and adipose-specific metabolic changes occurring when insulin, cortisol and dexamethasone are administered to healthy human subjects as such changes occur on a daily basis. Additionally the cumulative effects of administration of two of these hormones compared to single hormones was investigated.

4.2 Results

4.2.1 Insulin and cortisol study

The serum dataset derived from the positive ion mode acquisition consisted of 7384 metabolite features which was reduced to 6177 after missing value filtering and to 5442 after quality control filtering. The dataset derived from negative ion mode acquisition consisted of 5897 metabolite features which was reduced to 5098 after missing values filtering and to 4648 after quality control filtering. PCA score plots (**Figure 15**) were investigated and showed good clustering for the QC samples both in positive and negative ionisation mode indicating reliable instrument reproducibility throughout the run.

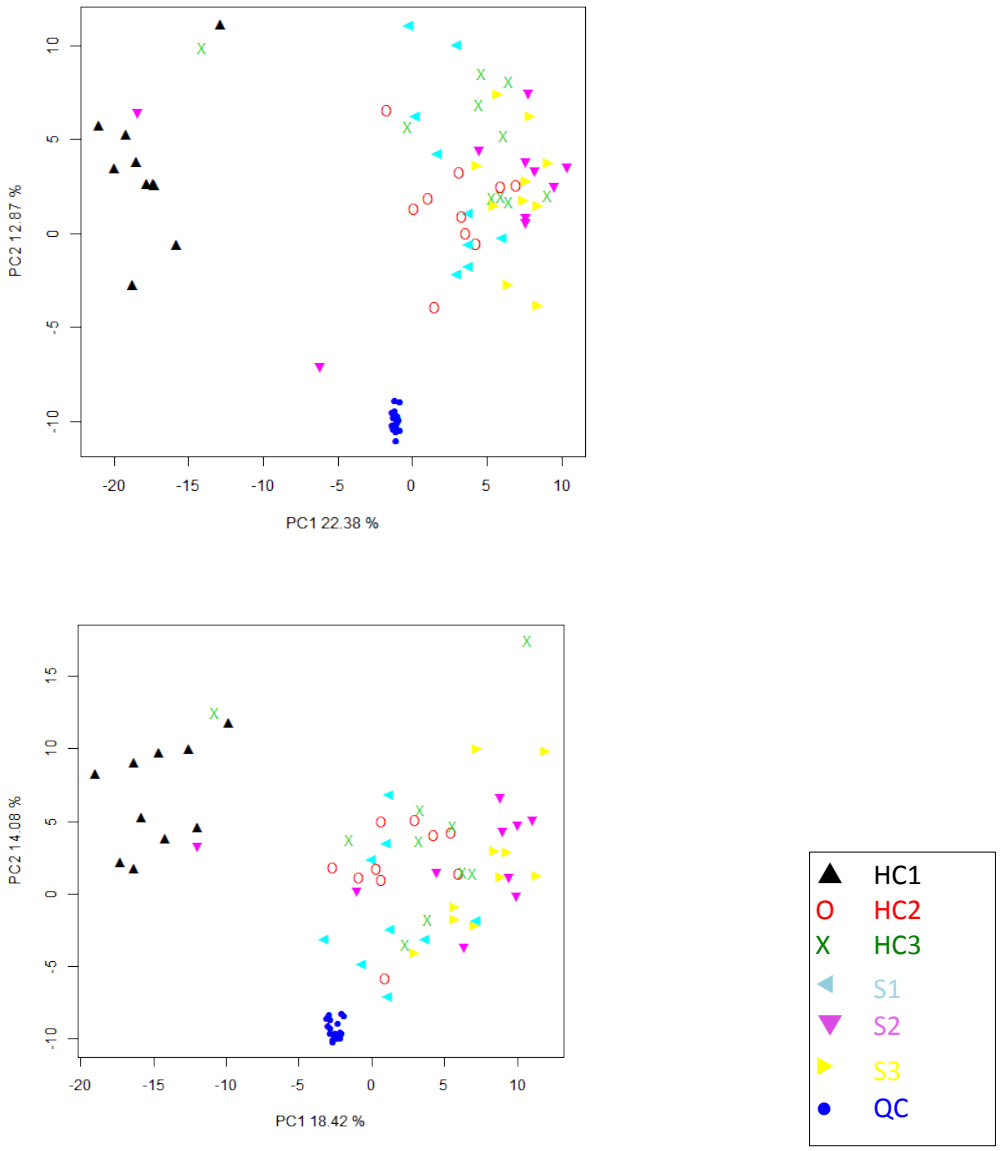


Figure 15. PCA score plots for the insulin and cortisol study in positive (top) and negative (bottom) mode. The QC samples (in dark blue circles) appear clustered indicating good instrument reproducibility.

The adipose tissue dialysate dataset obtained after positive ion mode analysis consisted of 14,486 metabolite features which was reduced to 790 features after missing value filtering. The large decrease is due to the poor recovery in the dialysate samples that produced results characterised by a high number of missing values. The dataset obtained after negative ion mode analysis consisted of

22,056 features which was reduced to 1583 features after missing value filtering. Both datasets underwent signal drift correction. PCA scores plots (**Figure 16**) for both datasets show satisfactory QC clustering.

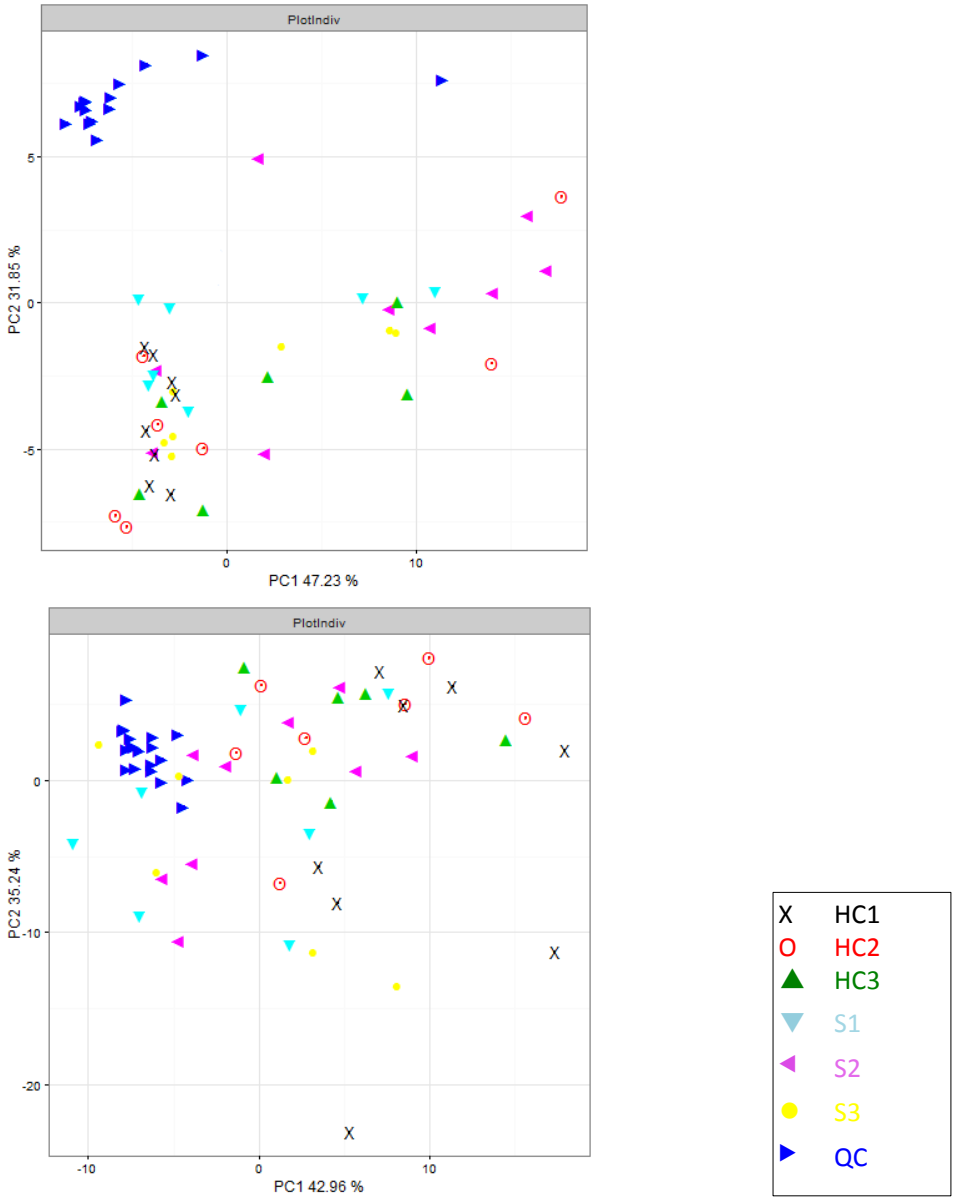


Figure 16 PCA plots score plots for the insulin and cortisol study (adipose tissue dialysate) in positive (left) and negative (right) mode. The QC samples (in dark blue triangles) appear generally clustered.

4.2.1.1 Univariate analysis – serum dataset

Five pairwise comparisons were performed in order to assess the metabolic changes occurring after insulin and cortisol administration:

- S1 – HC1 (effect of cortisol at basal level of insulin)
- S2 – HC2 (effect of cortisol at low level of insulin)
- S3 - HC3 (effect of cortisol at high level of insulin)
- S1 – S3 (effect of insulin at basal cortisol level)
- HC1 – HC3 (effect of insulin at high cortisol level)

After removing replicate metabolites features representing the same metabolite by keeping the feature with the lowest p-value comparison S1 – S3 described 347 statistically significant metabolites ($p<0.05$) changing when moving from low to high insulin at basal cortisol (**Appendix Table 7** and summarised in **Table 13**). The following results (increasing/decreasing) always refer to high levels of insulin e.g. 3 acylglycerides decreasing (at high levels of insulin). The changing metabolites included 12 acyl carnitines (decreased), 13 mono- and diacylglycerides (11 decreased), 7 triacylglycerides (5 increased), 3 amino acids (proline, methionine and phenylalanine), 2 ceramides (increased), 18 fatty acids (decreased), 27 oxidised fatty acids (24 decreased), 6 sphingolipids (5 increased), 10 steroids (8 decreased) and 60 glycerophospholipids (41 increased). Insulin induced activation of lipogenesis witnessed by fatty acids (and their oxidised forms) decreasing probably because of their usage for lipid storage (triglycerides, sphingomyelins and glycerophospholipids increasing). This outcome is consistent with the assumption that high levels of insulin promote glucose uptake and its use for lipid storage.

Comparison HC1 – HC3 investigated the effect of insulin at high levels of cortisol and described 283 statistically significant metabolites (**Appendix Table 8** and summarised in **Table 13**). The following

results (increasing/decreasing) always refer to high levels of insulin. These metabolites included 16 diacylglycerides (decreased), 5 ceramides (decreased), 57 glycerophopsholipids (46 decreased), 17 oxidised fatty acids (decreased), 7 sphingomyelins (decreased), 23 steroids (20 decreased).

Metabolite class	Upregulation (↑) or down regulation (↓) at X compared to Y (S1 – S3)	Number of metabolites changing in comparison S1-S3	Upregulation (↑) or down regulation (↓) at X compared to Y (HC1-HC3)	Number of metabolites changing in comparison HC1-HC3
Acyl carnitines	↓	12	↓	12
Mono and di-glycerides	↓	13	↓	16
Triglycerides	↑	7	↓	4
Amino acids	↓	3	MIX	5
Ceramides	↑	2	↓	5
Fatty acids	↓	18	↓	28
Oxidised fatty	↓	27	↓	17
Sphingolipids	↑	6	↓	7
Glycerophospholipids	↑	60	↓	57
Steroids	↓	10	↓	23

Table 13. Summary of statistically significant differences when comparing S1-S3 and HC1-HC3. Metabolites are grouped in classes of similar chemical structure or biological function. The arrows report up- or down-regulation of a class in a specific comparison.

When comparing the effect of insulin at low and high cortisol levels, some metabolic changes are induced by insulin regardless of the cortisol level (for example, diacylglycerides, oxidised fatty acids). However, some metabolite classes increase when insulin is administered at basal levels of cortisol and decrease at high levels of cortisol (ceramides, glycerophospholipids, sphingomyelins) (**Figure 17.** Boxplots of the glog transformed concentrations of all glycerophospholipids (left) and sphingolipids (right). Concentrations in red: basal insulin and saline infusion (S1), green: high insulin and saline infusion (S3), magenta: basal insulin and cortisol infusion (HC1), light blue: high insulin and cortisol infusion (HC3). suggesting potential inhibition of glycerophospholipid and sphingolipid synthesis occurring when insulin interacts with high cortisol levels. Interestingly, these lipid classes have been reported as mediators of insulin resistance[314]. Hence such an outcome suggests that insulin resistance, induced by GC administration, does not occur when insulin is infused in a condition characterised by high levels of GCs.

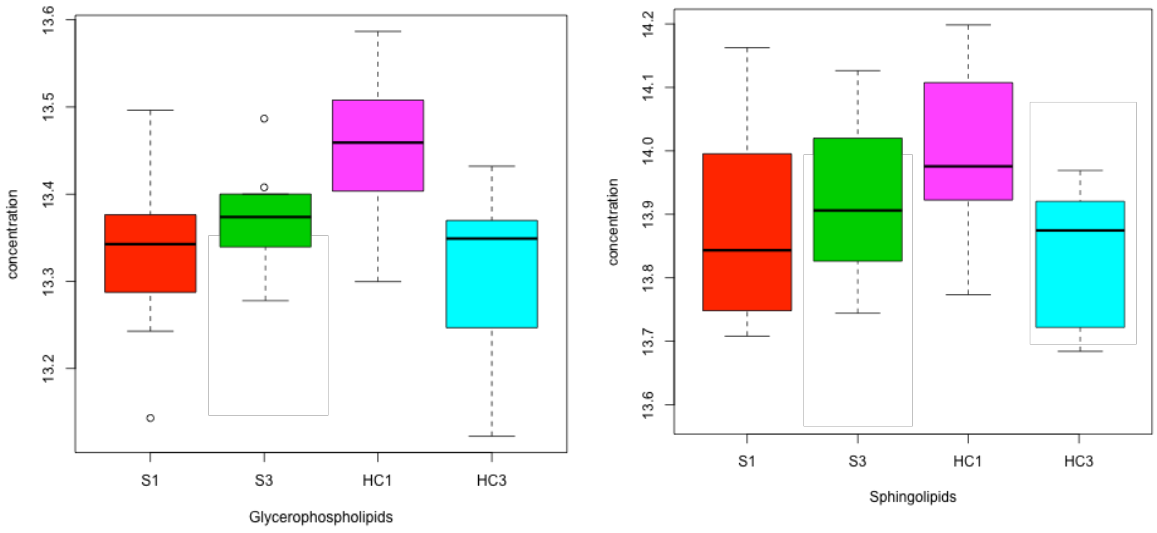


Figure 17. Boxplots of the glog transformed concentrations of all glycerophospholipids (left) and sphingolipids (right). Concentrations in red: basal insulin and saline infusion (S1), green: high insulin and saline infusion (S3), magenta: basal insulin and cortisol infusion (HC1), light blue: high insulin and cortisol infusion (HC3).

Comparison S1 - HC1 investigated the effect of cortisol infusion at basal insulin levels and resulted in 322 statistically significant metabolite features ($p<0.05$) (**Appendix Table 9**). The following results

relate to high levels of cortisol. These metabolites included: 6 acylcarnitines (increased by cortisol infusion), 24 di- and triacylglycerides (increased), 8 ceramides (6 increased), 7 fatty alcohol (increased), 62 glycerophospholipids (57 increased), 12 sphingomyelins (increased), tyrosine and dihydroxyphenyl-pyruvate (increased). The final two metabolites are involved in tyrosine metabolism which possibly highlights up-regulation of this pathway. While an increase in concentration of acylcarnitines may indicate the upregulation of lipolysis (enhancement of the long chain fatty acid carnitine shuttle), the increase of di- and triacylglycerides may suggest activation of lipogenesis (or possibly release of fatty acids as acylglycerides from adipose tissue in to blood). Although these two mechanisms indicate two complementary mechanisms operating in the opposite direction, such an outcome is expected. The study was carried out on serum samples that provide a snapshot of global metabolism. Hence, as already reported in the scientific literature, GCs induce lipolysis and fatty acid oxidation in peripheral adipose depots and induce lipogenesis in specific body regions like face and trunk. Therefore, traces of both mechanisms are expected in the global metabolism imprinted on serum. Furthermore, all lipids species reported as mediators of insulin resistance (diacylglycerols, sphingolipids, ceramides and glycerophospholipids) increase in concentration after cortisol infusion thus suggesting an immediate impairment of insulin sensitivity.

Some biological trends appear when comparing S1 – HC1, S2 – HC2 and S3 – HC3 (**Appendix Table 9**, **Appendix Table 10** and **Appendix Table 11**). The results are summarised in **Table 14**.

Metabolite class	Change in comp. S1-HC1	N° metabolites changing in comp. S1-HC1	Change in comp. S2-HC2	N° metabolites changing in comp. S2-HC2	Change in comp. S3-HC3	N° metabolites changing in comp. S3-HC3
Acyl carnitines	6	↑	13	↑	10	↑
Acyl glycerides	24	↑	13	MIX	8	MIX
Ceramides	8	↑	1	↑	3	MIX
Glycerophospholipids	62	↑	31	MIX	19	MIX
Sphingolipids	12	↑	6	MIX	4	↑

Table 14. Summary of statistically significant differences when comparing S1-HC1, S2-HC2 and S3-HC3. Metabolites are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison. MIX denotes a mixed trend.

Specifically, the number of statistically significant metabolites decrease progressively from S1 – HC1 to S3 – HC3. 322 metabolites changed in S1 – HC1 with 205 of these changing uniquely in S1 – HC1. 247 metabolites changed in S2 – HC2 with 81 of these changing uniquely in S2 –HC2. 175 metabolites changed in S3 – HC3 with 24 of these changed uniquely in S3 –HC3. This trend seems to suggest a negating effect on cortisol operated by the increase of insulin. Some biochemical classes like acyl glycerides, ceramides and sphingolipids were observed to be significantly changed in S1 – HC1 (cortisol infusion at basal insulin) but not at S2 – HC2 and S3 – HC3 (cortisol infusion at non-basal insulin). This is shown in **Figure 18**. As already noted, these classes have been reported as mediators of insulin resistance[129]. In fact GCs are known to induce insulin resistance. Hence it is possible that cortisol administration triggers mechanisms linked to insulin resistance even after single infusion. However, these differences are mitigated at non-basal levels of insulin. This is most probably due to the increasing levels of insulin that mediate the dysregulation of insulin sensitivity caused by cortisol.

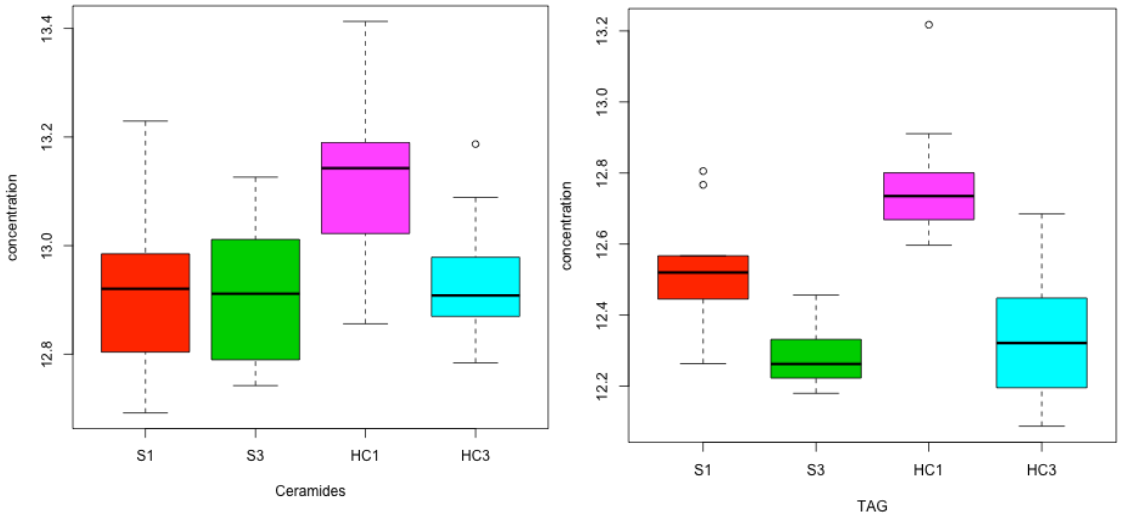


Figure 18. Boxplot of the glog transformed concentrations of all ceramides (left) and TAG (right). Concentrations in red: basal insulin and saline infusion (S1), green: high insulin and saline infusion (S3), magenta: basal insulin and cortisol infusion (HC1), light blue: high insulin and cortisol infusion (HC3).

However, other metabolites including six acyl carnitines, four steroids as well as methionine and phenylalanine increased only when cortisol was administered at non-basal levels of insulin. In fact, the increase in concentration for the acyl carnitines may indicate a cooperative effect of insulin and cortisol (insulin alone induces a drop in concentration for acyl carnitines). This is shown in **Figure 19**. High concentrations of acyl carnitines may suggest an enhancement of the fatty acid β-oxidation pathway.

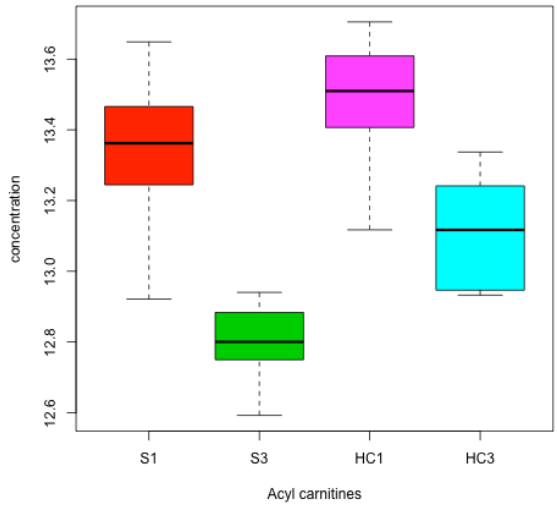


Figure 19. Boxplot of the glog transformed concentrations of all acyl carnitines. Concentrations in red: basal insulin and saline infusion (S1), green: high insulin and saline infusion (S3), magenta: basal insulin and cortisol infusion (HC1), light blue: high insulin and cortisol infusion (HC3).

Fatty acid β-oxidation is one of the four main metabolic pathways that fatty acids may undergo; the other three are sphingolipid synthesis, glycerophospholipid synthesis and acyl glycerides formation[129]. The cooperation of the effects of insulin and cortisol is likely to upregulate the fatty acid β-oxidation pathway instead of sphingolipid and glycerophospholipid synthesis. It must be noted that while the effect of increased acyl carnitines concentrations in serum is encountered when cortisol is administred in conditions of high levels of insulin (comparison S3 – HC3), the opposite effect (decreased acyl carnitine concentrationss decreasing) is displayed when insulin is administered

at non-basal levels of cortisol (comparison HC1-HC3). Therefore it is likely that the modality of cooperation between insulin and cortisol is dependent on the order of administration.

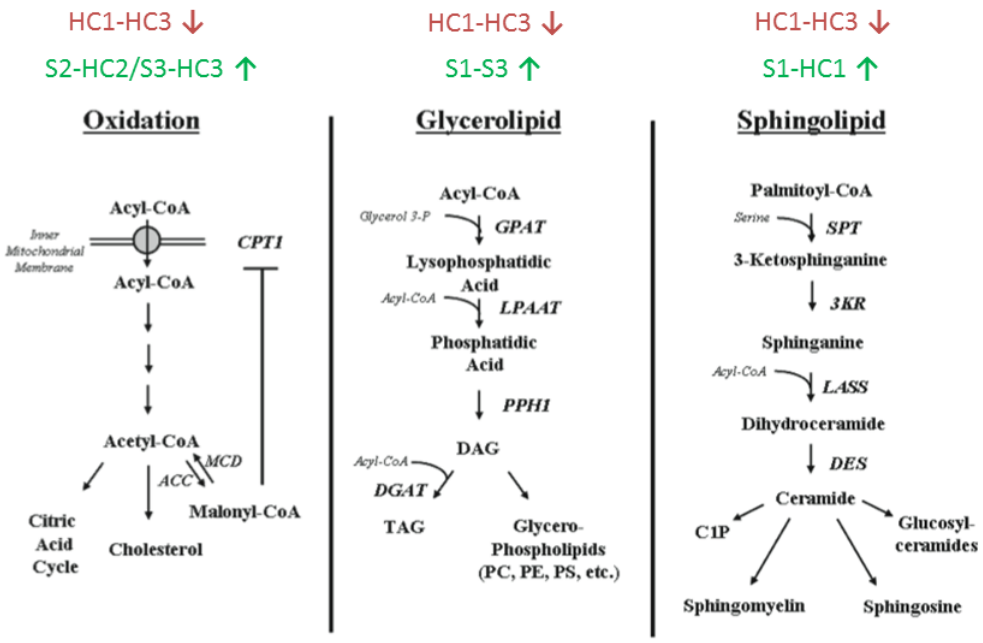


Figure 20. Overview of three metabolic pathways and their behaviour in conjunction with insulin and cortisol[129]. On the left the pathway involving fatty acid β-oxidation and the citric acid cycle, in the center acylglycerides and glycerophospholipid synthesis, on the right sphingolipid and ceramide synthesis.

The global effect of cortisol at increasing levels of insulin does not seem to highlight a precise biological mechanism. It appears that while a small number of metabolites are upregulated by the cooperative action of insulin and cortisol, the majority of changes induced by cortisol are neutralised by the administration of insulin.

4.2.1.2 Multivariate analysis – serum dataset

Comparison S1 – S3 did not produce any separation in the PCA scores plot in positive ion mode. However, the negative ion mode data showed a PCA scores plot presenting good clustering for the S1 and S3 group (**Figure 21**) and clear separation between the groups (PC1 vs PC2). Such separation is much more evident for the comparison HC1 – HC3 which is observed for positive and negative ion modes (**Figure 23**). Furthermore, PLS-DA models for the comparison HC1 – HC3 produced a good

cross-validated model (R^2 = 0.712 and Q^2=0.708 for positive ion mode and R^2 = 0.755 and Q^2=0.712 for negative ion mode) confirming the distinct biological differences between the two classes (**Figure 21** and **Figure 23**). Apparently, the effect of insulin at high cortisol levels yields a broader effect on the metabolome when compared to the effect of insulin at basal cortisol levels possibly suggesting an active interaction between hydrocortisone and insulin.

Analysing the multivariate outcomes for the effect of cortisol at increasing levels of insulin, the separation between classes decreases from comparison S1 – HC1 to S3 – HC3. In fact, the latter comparison produced PCA plots with the samples indistinguishably scattered across the scores plot.

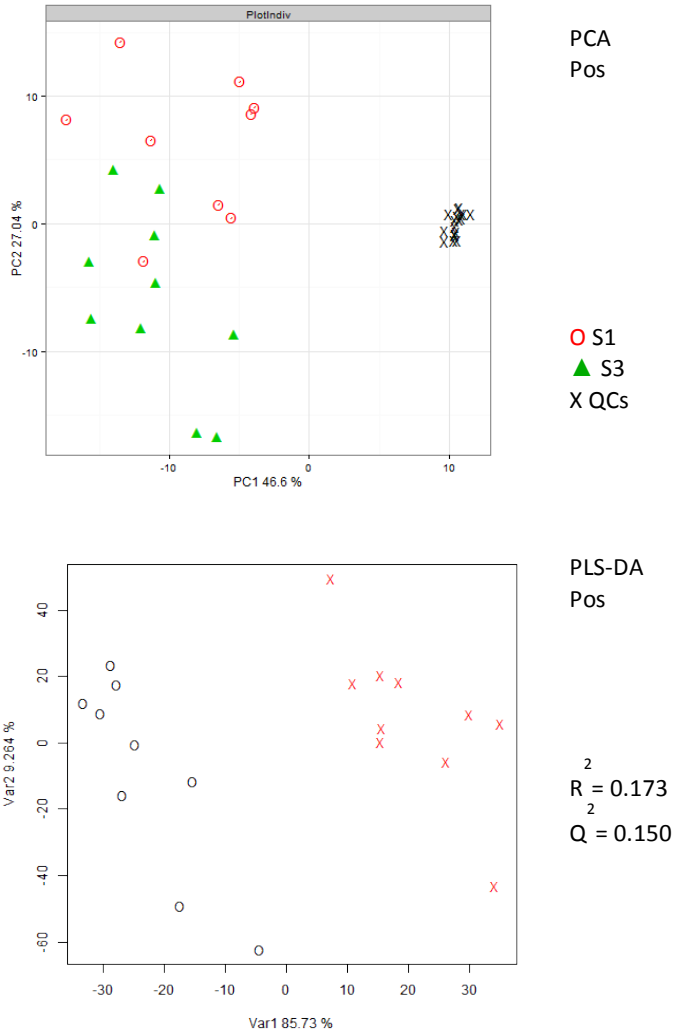
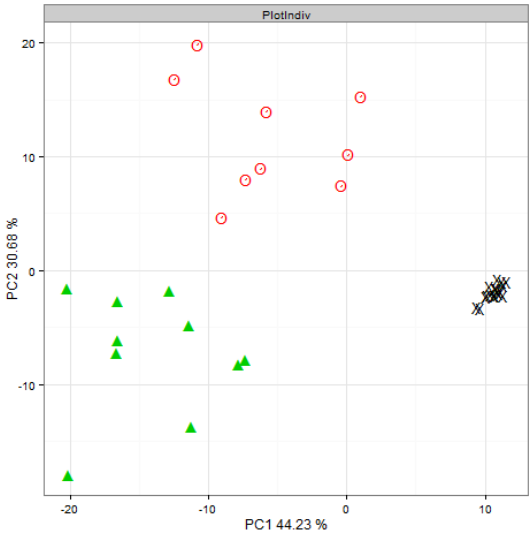
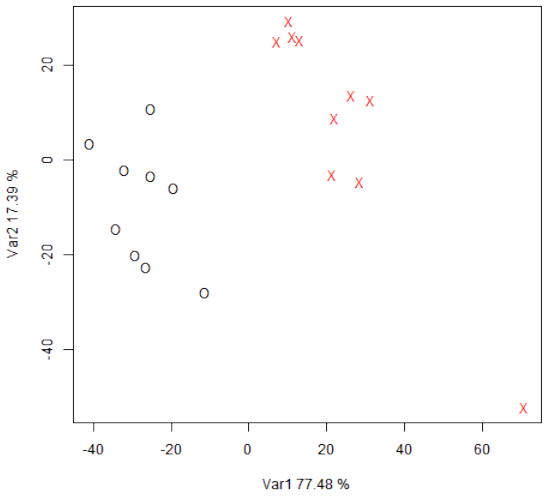


Figure 21. PCA (top) and PLS-DA score plots (bottom) for S1-S3 comparison in positive ion mode. PLS-DA plots report the cross-validation estimates. X = QC sample; O = S1; ▲ = S3



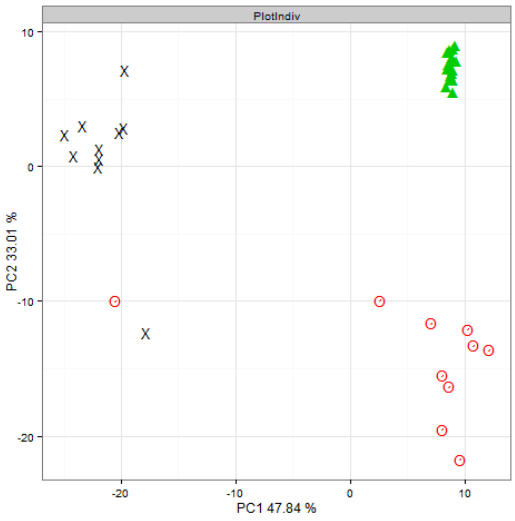
PCA
Neg



PLS-DA
Neg

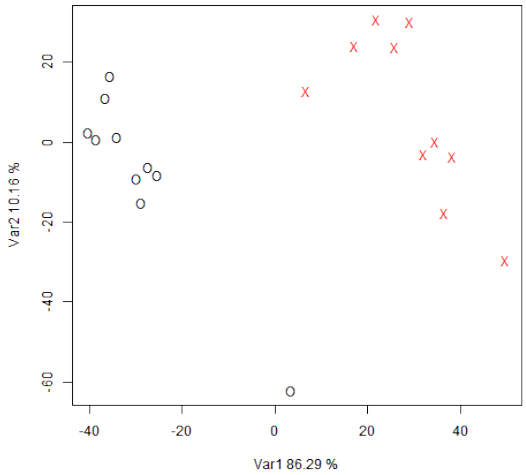
R^2 = 0.631
Q^2 = 0.347

Figure 22. PCA (top) and PLS-DA score plots (bottom) for S1-S3 comparison in negative ion mode. PLS-DA plots report the cross-validation estimates. X = QC sample; O = S1; ▲ = S3



PCA
Pos

X HC1
O HC3
▲ QCs



PLS-DA
Pos

$R^2 = 0.712$
$Q^2 = 0.708$

Figure 23. PCA (top) and PLS-DA score plots (bottom) for HC1-HC3 comparison in positive ion mode. Top: positive ion mode. Bottom: negative ion mode. PLS-DA plots report the cross-validation estimates. ▲ = QC samples; X = HC1; O = HC3

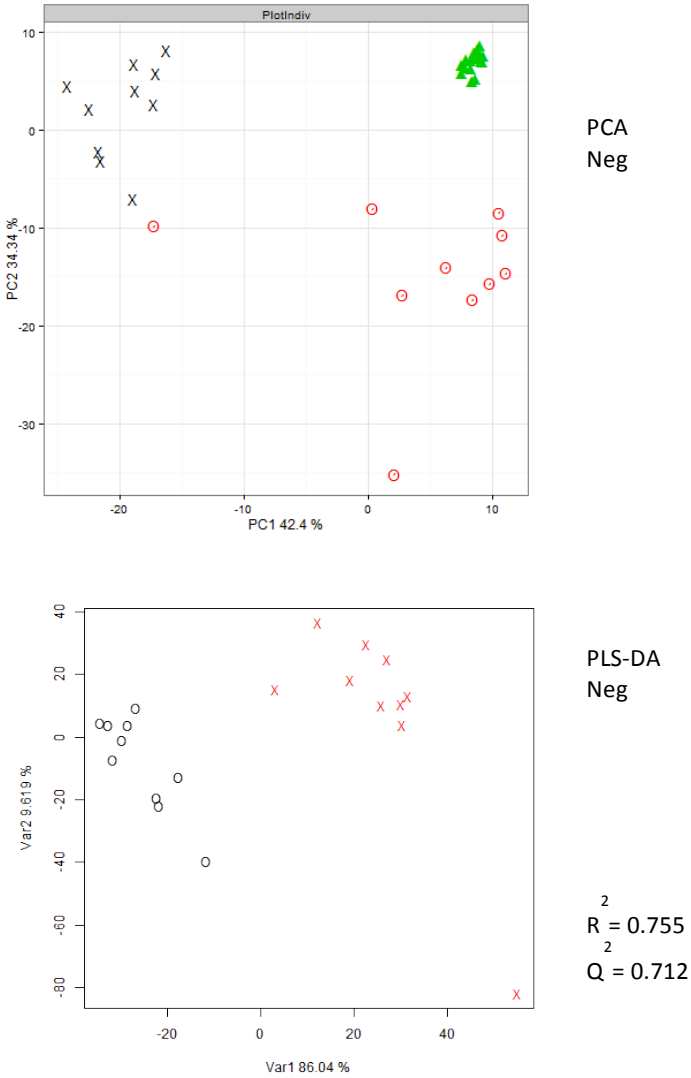
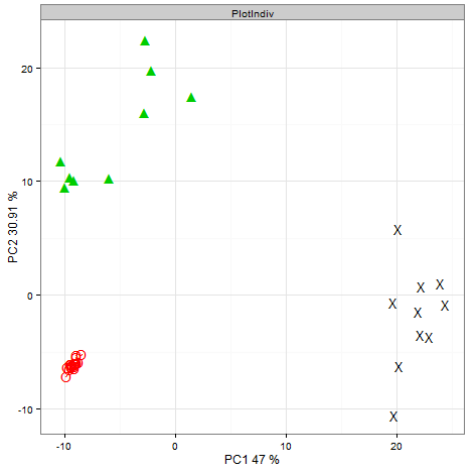


Figure 24. PCA (top) and PLS-DA score plots (bottom) for HC1-HC3 comparison in positive ion mode. Top: positive ion mode. Bottom: negative ion mode. PLS-DA plots report the cross-validation estimates. ▲ = QC samples; X = HC1; O = HC3

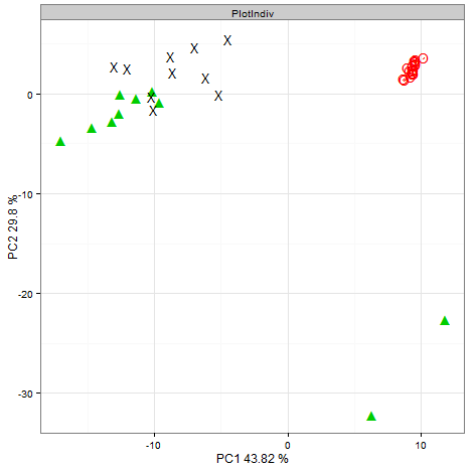
This outcome is corroborated by the PLS-DA models which produced good cross-validation estimates for comparison S1 - HC1 (R^2 and $Q^2 \approx 0.75$ for both positive and negative ion mode) but progressively worse estimates for S2–HC2 and S3–HC3 (**Figure 24**). This result may be interpreted as a confirmation of what the univariate analysis highlighted, the effect of increasing insulin seems to negate the majority of changes caused by cortisol administration.

The VIP scores for the PLS-DA models S1 – HC1 and S2 – HC2 (the only two cross-validated models) highlighted an essential role of those biochemical classes that appeared as significant in the univariate analysis.

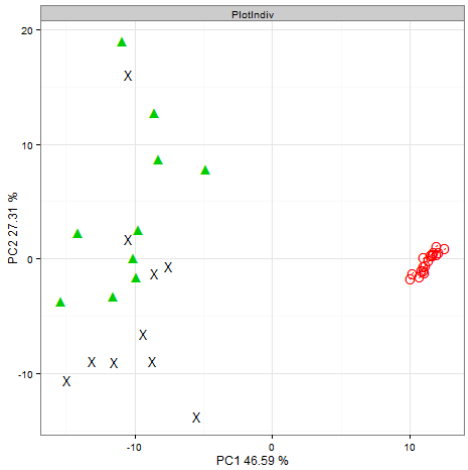


O QCs

▲ S1/X HC1

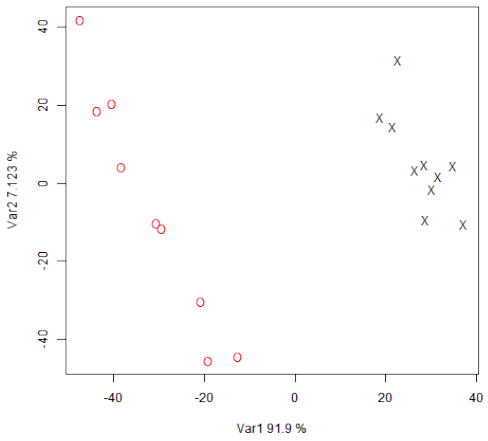


▲ S2/X HC2



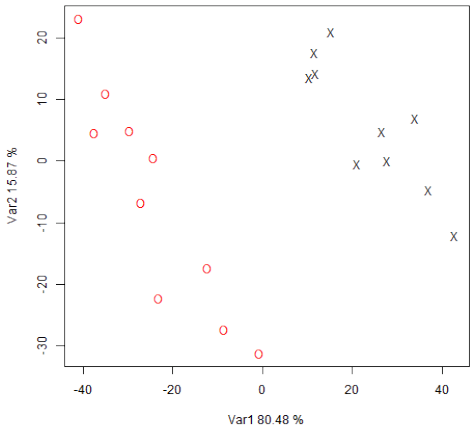
▲ S3/X HC3

Figure 25. PCA score plots of comparisons S1-HC1, S2-HC2 and S3-HC3 for the positive ion mode dataset.

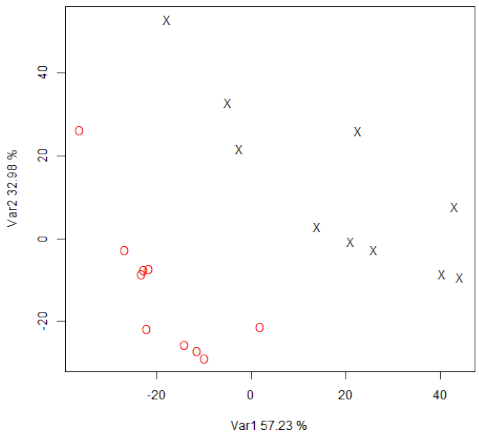


$R^2 = 0.797$ $Q^2 = 0.759$

O S/**X** HC



$R^2 = 0.586$ $Q^2 = 0.578$



$R^2 = 0.097$ $Q^2 = 0.177$

Figure 26. PLS-DA score plots of comparisons S1-HC1, S2-HC2 and S3-HC3 for the positive ion mode dataset.

After selecting the metabolite features with a VIP score > 1.5, acyl glycerides, glycerophospholipids and fatty acids appeared to drive the model's construction (**Table 15**). Notably, a higher number of acyl carnitines, which appeared as up-regulated by the interaction of insulin and cortisol, are observed in S2 - HC2 as expected. Interestingly, while glycerophospholipids drove the model for S1 – HC1, the number of statistically significant metabolite features decreases for the comparison S2 – HC2 while the number of fatty acids increase. This outcome may suggest a down-regulation of the glycerophospholipid synthesis pathway which results in an increased number of ceramides, acyl glycerides and sphingolipids being synthesised (univariate comparison S1 – HC1).

	Number of statistically significant metabolite features for comparison S1 – HC1	**Number of statistically significant metabolite features for comparison S2 – HC2**
Acyl glycerides	5	7
Acyl carnitines	2	10
Ceramides	3	0
Fatty acids	9	14
Glycerophospholipids	28	13
Sphingolipids	6	4

Table 15. Number of metabolite features with VIP score>1.5 for the PLS-DA models for comparison S1 – HC1 (left) and S2 – HC2 (right). Metabolite features are grouped according to their biochemical similarity.

4.2.1.3 Univariate analysis – adipose tissue dialysate dataset

The adipose tissue dialysate dataset was analysed according to the same comparisons that were explored for the serum dataset. Generally, the number of statistically significant metabolite features identified in the dialysate dataset is much lower when compared to the serum study. Nevertheless

some biological important trends were detected.

Comparison S1 – S3 highlighted 106 statistically significantly metabolite features ($p<0.05$) when comparing low and high insulin levels at basal cortisol. The following results refer to high insulin levels. 29 metabolite features were annotated (**Appendix Table 12** and summary in **Table 16**) including 4 amino acids and amino acid derivatives (decreasing), 1 ceramide and 1 cholesterol ester (increasing), 3 out of 4 fatty acids increasing, 1 fatty alcohol increasing, 3 oxidised fatty acids decreasing, 2 glycerophospholipids increasing and 1 sphingomyelin increasing. The trend reversal between fatty acids and oxidised fatty acids suggests an impairment of fatty acid β-oxidation and consequent accumulation of fatty acids.

Metabolite class	General trend of upregulation or downregulation in comparison S1-S3	Number of metabolite features statistically significant in comparison S1-S3	General trend of upregulation or downregulation in comparison HC1-HC3	Number of metabolite features statistically significant in comparison HC1-HC3
Amino acids	↓	3	↓	4
Ceramides	↑	1	0	0
Fatty acids	↑	3	↑	2
Oxidised fatty acids	↓	3	MIX	2
Sphingolipids	↑	1	0	0
Glycerophospholipids	↑	2	MIX	8
Steroids	↑	1	↑	3

Table 16. Summary of statistically significant differences when comparing S1-S3 and HC1-HC3 for the adipose tissue dialysate dataset. Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison. MIX denotes mixed trend

Comparison HC1 – HC3 highlighted 155 statistically significantly metabolite features ($p<0.05$) when comparing low to high insulin at high cortisol levels.

The following results refer to high insulin levels. 42 features were annotated (**Appendix Table 13**) including 4 out of 5 amino acids and amino acid derivatives decreasing, 3 steroids increasing and 8 glycerophospholipids presenting a mixed trend. Similar to what occurred in the serum dataset, insulin infusion at high cortisol levels has a different effect on some metabolic classes when compared to basal cortisol levels, including the indicators of lipogenesis that were detected in comparison S1-S3 not being observed as statistically significant or present as a mixed trend in this case. This is notable as ceramides and sphingolipids in general are known to be an indicator of insulin resistance. Here we observed a marked shift of trend, the metabolite features that were perturbed by insulin infusion at basal cortisol levels and that could indicate insulin resistance are completely absent after cortisol infusion.

Metabolite class	General trend of upregulation or downregulation in comparison S1-HC1	Number of metabolite features statistically significant in comparison S1-HC1	General trend of upregulation or downregulation in comparison S2-HC2	Number of metabolite features statistically significant in comparison S2-HC2	General trend of upregulation or downregulation in comparison S3-HC3	Number of metabolite features statistically significant in comparison S3-HC3
Amino acids	3	↓	4	↓	5	↓
Fatty acids	/	/	3	MIX	3	MIX
Oxidise fatty acids	1	↑	3	↓	3	MIX
Ceramides	/	/	1	↑	2	↓
Nucleotides	2	↓	3	↓	3	↓
Glycerophospholipids	5	↑	/	/	7	MIX
Sphingolipids	/	/	/	/	1	↓

Table 17. Summary of statistically significant differences when comparing S1-HC1, S2-HC2 and S3-HC3 for the adipose tissue dialysate dataset. Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison.

Comparison S1 - HC1 investigated the effect of cortisol infusion at basal insulin levels and described 173 statistically significant metabolite features ($p<0.05$). The following results refer to high cortisol levels. 42 features were annotated (**Appendix Table 14**) including 5 glycerophospholipids increasing, 2 nucleotides decreasing, 3 amino acids derivatives decreasing, 1 carnitine derivative decreasing and 1 steroid decreasing.

Comparison S2 – HC2 investigated the effect of cortisol infusion at low insulin levels and described 129 statistically significant metabolite features ($p<0.05$). The following results refer to high cortisol levels. 24 features were annotated (**Appendix Table 15**) including 4 out of 6 amino acids and amino acid derivatives decreasing, 3 nucleotides decreasing, 3 out of 4 oxidised fatty acids decreasing and 3 fatty acids presenting a mixed trend.

Comparison S3 – HC3 investigated the effect of cortisol infusion at high insulin levels and described 242 statistically significant metabolite features ($p<0.05$). The following results refer to high cortisol levels. 47 features were annotated (**Appendix Table 16**) including 5 out of 6 amino acids and amino acid derivatives decreasing, 2 ceramides decreasing, 3 oxidised fatty acids and 3 fatty acids presenting a mixed trend, 3 nucleotides decreasing, 7 glycerophospholipids presenting a mixed trend and 1 sphingolipid decreasing. Results are summarised in **Table 17**.

The strong effect of cortisol that was highlighted in the serum dataset is not detected in the adipose tissue dialysate dataset. While insulin seemed to diminish the effect of cortisol in serum, here it seems to cooperate at some level as observed for comparison S3-HC3. The main conclusion drawn for these three comparisons in the serum dataset was an occurrence of insulin resistance that was reduced by increasing levels of insulin. Conversely the adipose tissue dialysate dataset may demonstrate no observable phenotype of this mechanism thus complying with the previous finding of glucocorticoids failing to induce insulin resistance in adipose tissue[87].

Interestingly comparison S3-HC3 highlighted ceramides and sphingolipids as statistically significant

suggesting a deactivation of the sphingolipid formation pathway probably triggered by high levels of insulin and cortisol. This effect confirms the conclusion from the serum study. In fact the serum dataset showed an activation of fatty acid β-oxidation which is a mechanism that competes with sphingolipid synthesis for fatty acids availability. Here the data suggest that this activation may be counterbalanced by a corresponding deactivation of sphingolipid synthesis (less fatty acids are available to feed into sphingolipid synthesis).

4.2.1.4 Multivariate analysis – adipose tissue dialysate dataset

The PCA score plots did not highlight any clustering within or between the different classes, probably due to the low number of metabolite features detected. Also the PLS-DA models did not produce good cross-validation models except for the comparison HC1-HC3 in negative ion mode with R^2 = 0.50 and Q^2 = 0.58 for a model built on two components.

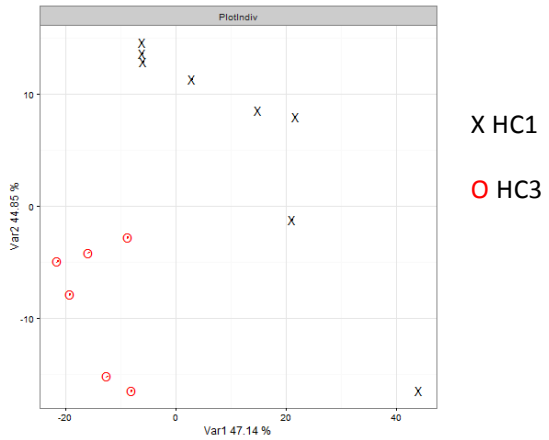


Figure 27. PLS-DA score plots for the comparison HC1-HC3 in negative mode for adipose tissue dialysate samples.

The corresponding VIP scores (variables were retained if *VIP*>1.25) showed 3 amino acid derivatives, 4 carbohydrates, 2 nucleotides and 2 phosphatidylethanolamines as the most important for model construction (**Appendix Table 17**). After comparison with the univariate analysis results, it seems that

the decrease of amino acid derivatives (induced by cortisol in every comparison) has a central role in adipose tissue. Also phosphatidylethanolamines were detected as statistically significant for comparison HC1-HC3 in univariate analysis but unfortunately different features presented opposite trends. Strangely also the two phosphatidylethanolamines that are driving the PLS-DA model present opposite trends, hence the biological interpretation is still unclear.

4.2.2 Dexamethasone and cortisol study

After signal correction and multi-batch integration the dataset derived from the positive ion mode acquisition consisted of 5747 metabolite features which were reduced to 4862 after missing value filtering and to 3860 after quality control filtering. The dataset derived from negative ion mode acquisition consisted of 5055 metabolite features which were reduced to 4222 after missing values filtering and to 3278 after quality control filtering. The PCA score plot of the datasets showed good clustering for the QC samples both in positive and negative ionisation mode (**Figure 28**) indicating acceptable instrument reproducibility throughout the two batch run.

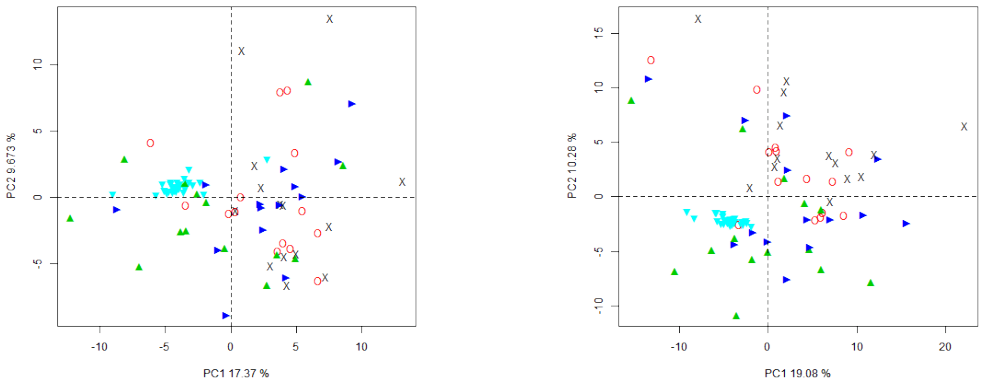


Figure 28. PCA score plots for the dexamethasone and cortisol study in positive (left) and negative (right) mode. The QC samples (in light blue triangles) are clustered indicating good instrument reproducibility.

4.2.2.1 Univariate analysis

Three main biological questions have been formulated and data were analysed accordingly.

1. In order to assess the metabolic changes occurring due to diurnal rhythms comparison of treatment 1 pre vs treatment 1 post was performed (comparison [1]). This compares each subject between 19:00 on day 1 and 7:00 on day 2 thus highlighting the changes occurring overnight related to cortisol levels increasing.
2. Changes induced by dexamethasone administration were evaluated by comparing treatment 1 pre – treatment 2 post (comparison [2]). Sample 1 pre was collected at 19:00 on day 1 of treatment 1 (no dexamethasone administration) and sample 2 post was collected at 19:00 on day 2 of treatment 2 (after three dexamethasone administrations). Since the two samples were collected at the same time during the day, the comparison should highlight the effects which are induced by dexamethasone only (and its effect on cortisol synthesis/degradation) without any effect of possible metabolic changes induced by circadian cycles.
3. The effects of cortisol administration through different modality in conjunction with dexamethasone (treatments 3-5) were assessed comparing treatment 2 post (after dexamethasone administration) vs treatment 3 or 4 or 5 post (after dexamethasone and cortisol administration). This comparison should therefore highlight metabolic changes which are triggered by cortisol. The resulting comparisons are indicated by [3], [4] and [5] respectively.

After removing replicate metabolites features representing the same metabolite by keeping the metabolite feature with the lowest p-value, comparison [1] highlighted 195 statistically significant metabolite features ($p<0.05$). Results are shown in **Appendix Table 18** and are summarised in **Table 18**. The following results refer to naturally occurring high levels of cortisol. The significant metabolite features included 11 mono-, di- and triacylglycerides (decreased after naturally increasing cortisol

levels), 5 carbohydrates (4 increased), 20 fatty acids (17 decreased), 17 glycerophospholipids (12 increased) and 23 oxidised fatty acids (decreased).

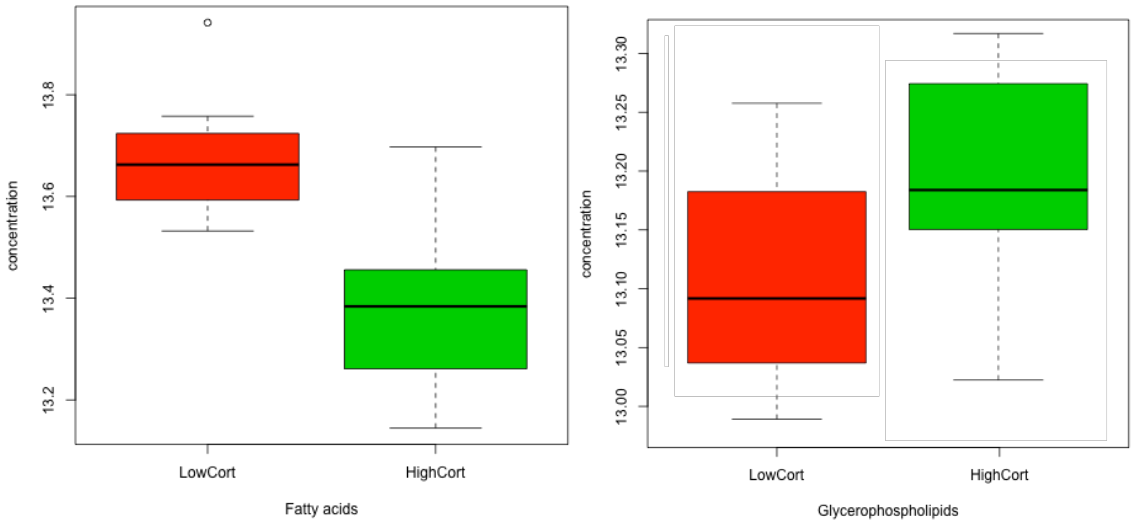


Figure 29. Boxplot of the glog transformed concentrations of all fatty acids (left) and glycerophospholipids (right). Concentrations in red: low endogenous cortisol level (1-pre), green: high endogenous cortisol level (1-post).

Metabolite class	Number of metabolite features statistically significant in comparison [1]	General trend of upregulation or downregulation in comparison [1]	Number of metabolite features statistically significant in comparison [2]	General trend of upregulation or downregulation in comparison [2]
Acyl carnitines	8	MIX	9	↑
Acyl glycerides	11	↓	14	↑
Amino acids	7	MIX	7	↓
Carbohydrates	5	↑	3	MIX
Glycerophospholipids	17	↑	22	MIX
Fatty acids	20	↓	16	↑
Oxidised fatty acids	23	↓	16	↑

Table 18. Summary of statistically significant differences when comparing [1] (changes induced by endogenous cortisol) and [2] (changes induced by dexamethasone). Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison.

3 oxidised fatty acids (hydroxy-eicosadienoic acid, hydroxy-decenoic acid/oxo-decanoic acid and oxo-tetracosenoic acid) displayed a concentration decrease of more than 4-fold. While the enhanced utilisation of fatty acids may suggest an enhancement of lipogenic mechanisms, the decrease in acyl glyceride concentrations does not suggest lipogenesis as a mechanism consuming fatty acids as cortisol levels increase. However, these results may indicate a precise mechanism in fatty acid metabolism. In fact, intracellular fatty acids can undergo three separate metabolic pathways[314]: they can be oxidised in order to produce ATP in the citric acid cycle following β oxidation feeding the resulting acetyl-CoA in to the said cycle, they can be converted to diacylglycerol which in turn may be converted to triacylglycerides or glycerophospholipids and if the carbon chain is sufficiently long they can be converted in to ceramides. Fatty acids are also involved in other metabolic pathways regulated by cyclooxygenase (COX) and lipoxygenase (LOX) enzymes. Hence, the results shown here provide evidence of an enhancement of the metabolic pathway synthesising glycerophopspholipids and suppression of fatty acid β-oxidation. This outcome complies to some extent with the results related to the comparison S1 – HC1 (effect of cortisol) presented in the insulin and cortisol study. Here 60 out of 67 statistically significant glycerophospholipids reported an increase after cortisol infusion. However, other metabolic classes demonstrate a different effect in these comparisons. Fatty acids and acyl glycerides which increased in concentration in the insulin and cortisol study moving from low to high levels of exogenous cortisol in serum samples (S1-HC1) showed a decrease in concentration in this study moving from low to high levels of endogenous cortisol (comparison [1]). It is possible that this effect is due to the different exposures to cortisol. The subjects in the insulin-cortisol study underwent a cortisol infusion, whereas in the present study the effect of the circadian rhythm of endogenous cortisol is measured. It must be noted that the exogenous levels of cortisol reached in the insulin cortisol study are much higher than the endogenous overnight peak measured in the present study in comparison [1] (1139.60 vs 405.42 nmol/L, P <.00001 vs saline)[87].

After removing replicate metabolites features representing the same metabolite by selecting the

metabolite feature with the lowest p-value, comparison [2] highlighted 197 statistically significantly metabolite features shown in **Appendix Table 19**. The following results refer to high levels of dexamethasone. These metabolite features included 9 acyl carnitines (increased), 14 acyl glycerides (increased after dexamethasone administration), 7 amino acids (decreased), 16 fatty acids (increased) and 16 oxidised fatty acids (12 increased). The outcome shown by this comparison may indicate a mechanism which is opposed to the one highlighted in comparison [1]. In fact, the up-regulation of fatty acids (and oxidised fatty acids) and acyl carnitines (**Figure 30**) suggests an enhancement of the carnitine-regulated transport of fatty acids into the mitochondrial matrix for subsequent fatty acid β-oxidation. Interestingly, the administration of a synthetic GC such as dexamethasone induces effects which are opposite to the ones induced by diurnal increase of cortisol by natural occurrence. Possibly, a reason for that could be the complete suppression of endogenous GCs induced by dexamethasone administration thus negation of all the effects induced by cortisol.

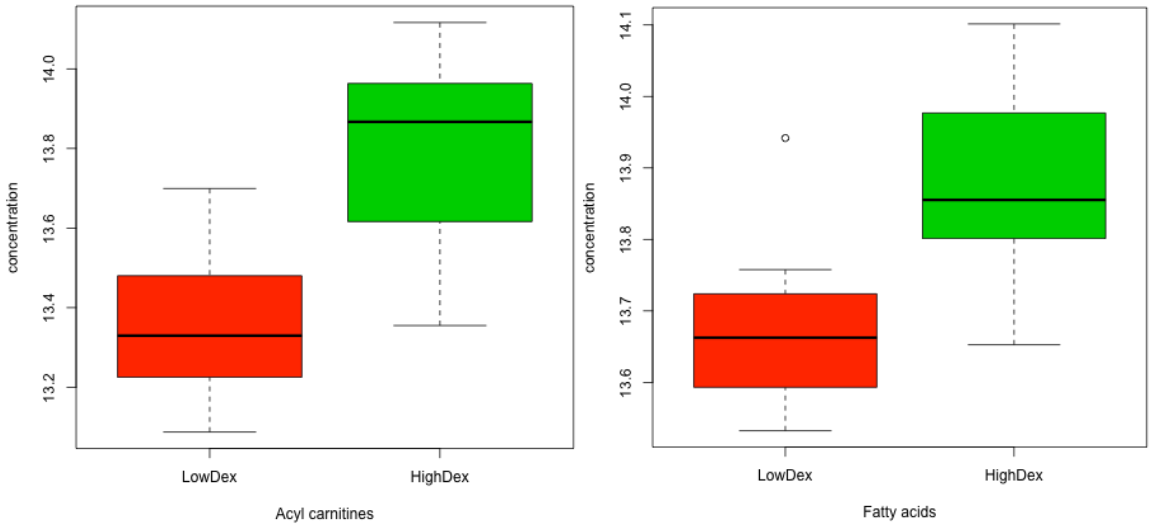


Figure 30. Boxplot of the glog transformed concentrations of all acyl carnitines (left) and fatty acids (right). Concentrations in red: low dexamethasone level (1-pre), green: high dexamethasone level (2-post).

No clear biological trends appeared in the univariate analysis results for comparisons [3], [4] and [5] (**Appendix Table 20**, **Appendix Table 21** and **Appendix Table 22**). Comparisons [3], [4] and [5]

described 71, 45 and 180 statistically significant metabolite features. Comparisons [3] and [4] do not show any particular biological class trends. However, comparison [5] describes changes of 15 acyl glycerides, 23 fatty acids, 20 oxidised fatty acids and 11 steroids all decreasing after cortisol administration. **Table 19** shows that glycerophospholipid concentrations increased (**Figure 31**) while fatty acid and acyl carnitine concentrations decreased after cortisol administration. This result would confirm a stimulation of lipogenesis rather than an enhancement of fatty acid β-oxidation. Although the interpretation of such data may be laborious and difficult it clearly appears that considering the number of statistically significant metabolite features in these comparisons, that administration of intravenous cortisol has a more striking effect compared to administration of tablets and granules. This outcome may imply a major effect which is caused by the rapid bioavailability assured by intravenous administration as opposed to the slow release provided by tablets and granule administration.

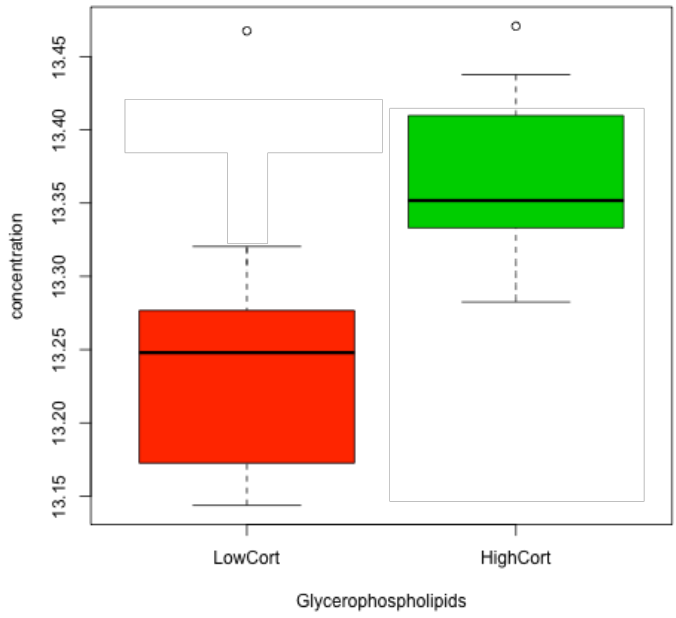


Figure 31. Boxplot of the glog transformed concentrations of all glycerophospholipids. Concentrations in red: low exogenous cortisol level (2-post), green: high exogenous cortisol level (5-post).

Metabolite class	Number of metabolite features statistically significant in comparison [3]	General trend of upregulation or downregulation in comparison [3]	Number of metabolite features statistically significant in comparison [4]	General trend of upregulation or downregulation in comparison [4]	Number of metabolite features statistically significant in comparison [5]	General trend of upregulation or downregulation in comparison [5]
Acyl glycerides	5	MIX	3	MIX	15	↓
Fatty acids	4	MIX	1	↓	20	↓
Oxidised fatty acids	1	↓	2	↓	23	↓
Steroids	4	↓	5	↓	11	↓
Glycerophospholipids	15	MIX	10	↑	16	↑
Acyl carnitines	2	MIX	2	↓	4	↓

Table 19. Summary of statistically significant differences when comparing [3],[4] and [5]. Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison.

4.2.2.2 Multivariate analysis

Comparison [1] and [2] did not produce cross-validated PLS-DA models and the PCA scores plots did not show compact clustering of the different variables according to the biological classes (**Figure 32**). No cross-validated or biologically-relevant multivariate models were constructed for comparisons [3], [4] and [5].

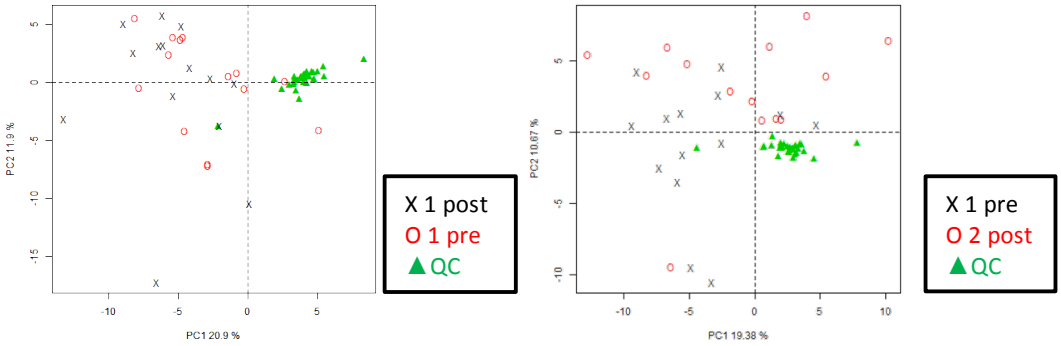


Figure 32. PCA plots (positive ion mode) relating to comparison [1] and [2]. No clear separation between samples is displayed.

4.3 Conclusions

These observations provide an interesting new approach (i.e. untargeted metabolomics) to a phenotype that was already explored with other and more specific techniques[87].

	Triglycerides	**Fatty acids**	**Ceramides**	**Glycerophospholipids**
Effect of insulin at basal levels of cortisol	Upregulation	Downregulation	Upregulation	Upregulation
Effect of insulin at high levels of cortisol	Downregulation	Downregulation	Downregulation	Downregulation

Table 20. Summary of the metabolic changes occurring after insulin administration

The two studies described highlighted that insulin is a potential inducer of triglyceride and glycerophospholipid synthesis as suggested by the decrease in fatty acid concentrations and increase in triacylglycerides and glycerophospholipid concentrations. These findings somehow hint at a similar direction pointed by studies proving insulin as a promoter of glycogenesis and suppressor of gluconeogenesis and glycogenolysis[315] in non-diabetic subjects. In fact, comparing the outcomes from the present study and the known literature, it seems that insulin administration may promote the synthesis of complex lipids and at the same time impair the catabolic formation of more simple carbohydrates.

Insulin administration in conditions of high concentrations of cortisol influences glycerophospholipid and ceramide metabolism in serum which inverts the trend shown at basal cortisol concentration. Effectively while these classes appeared as increasing after insulin administration at basal cortisol levels, they showed downregulation after insulin administration at high cortisol levels. This result

may imply an interaction between insulin and cortisol which is exerted on lipid metabolism and needs to be more thoroughly investigated.

	Acylglycerides	**Ceramides**
Endogenous cortisol	Downregulation	No effect
Exogenous cortisol	Upregulation	No effect
Cortisol at basal insulin	Upregulation	Upregulation
Cortisol at high insulin	No effect	No effect

Table 21. Summary of the metabolic changes occurring after cortisol increase

Cortisol administration was investigated through two separate studies. The effect of external cortisol infusion through comparison S1-HC1 of the insulin/cortisol branch of the study and effect of naturally occurring high levels of cortisol in comparison [1] of the cortisol/dexamethasone branch of the study was performed. It must be noted that external infusion of cortisol significantly increases the concentration of cortisol compared to the endogenous overnight peak (1139.60 vs 405.42 nmol/L, P <.00001 vs saline)[87]. While external infusion led to an increase in the concentrations of acyl carnitines, acylglycerides and glycerophospholipids, diurnal cortisol changes appeared to lead to a decrease in the concentrations of fatty acids and acylglycerides and an increase in the concentrations of carbohydrates and glycerophospholipids. None of the two effects can be clearly identified as a specific metabolic pathway which is perturbed even though (at least for diurnal changes) lipolytic activity was expected as described by Long[41] who reported suppression of lipogenesis after GC administration in the fasted state in humans. In fact, the effect which was observed in our study, seems to lead to a similar conclusion as suggested by the decrease in concentration of fatty acids. The increase in concentration of glycerophospholipids however still requires further investigation. Notably, exogenous administration of cortisol appeared to increase (fold change ≈ 1.7 for 6

ceramides and 11 sphingomyelins) some of the biochemical markers of insulin resistance immediately after cortisol infusion. This finding, albeit requiring further confirmation, offers clarification regarding the time scale of impaired insulin sensitivity. Moreover this result is supported by the trends observed in comparisons S1-HC1, S2-HC2 and S3-HC3. Specifically, while the insulin levels progressively increase, cortisol administration has a smaller impact on the global serum metabolome. Acyl glycerides, ceramides and sphingolipids (all insulin resistance mediators) are significantly increased only at basal insulin levels suggesting a negating effect of insulin on the effects of these metabolic classes. Hence, it is probable that insulin infusion inhibits the enhancement of insulin resistance triggered by high cortisol concentrations. This outcome is completely reverted in adipose tissue where the insulin resistance effect is severely blunted at high cortisol levels. This result complies with the tissue specific study performed by Halzhurst *et al.*[87] that reported insulin sensitization in subcutaneous adipose tissue. However the tissue specific study highlighted deactivation of sphingolipid synthesis triggered by cooperation between cortisol and insulin and explains the activation of fatty acid β-oxidation displayed in the serum study.

Adipose tissue dialysate also showed differences between metabolic changes observed in comparisons S1-S3 and HC1-HC3. The latter comparison does not display any statistically significant changes in ceramide and sphingolipid metabolism and the glycerophospholipids do not appear to have a definite trend (some were increased and some were decreased in concentration at high insulin). Moreover it is notable that phosphethanolamines are shown to be biologically important in both the univariate and multivariate analysis for comparison HC1-HC3 although presented different trends (again some were increased and some were decreased in high insulin). To this purpose, it would be interesting in the future to perform a targeted analysis on phosphoethanolamines to better understand the mechanism triggered by insulin infusion at high cortisol levels.

On a separate note, it must be mentioned that the metabolic effects of insulin infusion in conditions characterised by high levels of GCs do not coincide with the effects displayed after GC infusion at

high levels of insulin. Hence interaction between insulin and cortisol does not seem to operate through a trivial mechanism but is dependent on the concentrations of both insulin and cortisol.

Furthermore, it seems that amino acid metabolism in adipose tissue is particularly affected by cortisol administration regardless of the concentration of insulin. Specifically, the concentration of amino acids decreased after cortisol administration possibly hinting impaired protein breakdown or impaired amino acid synthesis. The decreased amino acid concentration may also affect later feeding into gluconeogenesis. Notably this outcome was detected through a spectroscopic analysis that is well suited to detect semi-polar and non-polar metabolites and therefore an analytical method targeted to investigate polar metabolites may describe new and biologically important discoveries.

On the other hand, dexamethasone administration, most likely through suppression of endogenous GC synthesis, led to an enhancement of fatty acid β-oxidation and carnitine-mediated fatty acid transport. Therefore the effect which is displayed by cortisol administration is not only suppressed by dexamethasone administration but completely reverted. This consideration sheds light on the effects that may arise after dexamethasone administration in conjunction with endogenous or exogenous cortisol.

For example this study highlighted that cortisol infusion, rather than tablet or granules administration may potentially display peculiar perturbations when administered with dexamethasone. It would be appropriate for future studies to assess the changes at the concentration peak of cortisol. In fact, comparisons [3], [4] and [5] included samples taken after 12 hours from cortisol administration. If this result confirms on one side the fundamental biological difference between two GCs as dexamethasone and cortisol, it also focuses on the importance of detecting the possible interactions between these two GCs. Since they are both largely adopted in clinical treatment it is of paramount importance to univocally address the possible outcomes of such interaction.

CHAPTER 5 - Inhibitors of 5-alpha reductase activity in steroid metabolism induce global metabolic changes

5.1 Introduction

5-alpha reductase (5-αR) is an enzyme which is involved in the metabolism of different steroids. 5-αR acts on different substrates including progesterone, androstenedione, aldosterone, corticosterone and cortisol through reduction of the steroidal A-ring[316]. All of the enzyme's substrates are characterised by the presence of a keto group on carbon C3[16]. The products of 5-αR reduction are known to undertake various roles in physiological conditions (**Figure 33**). Dihydroprogesterone (synthesised from progesterone) has a role in hormonal balance in fertile women[317], dihydrocortisol (synthesised from cortisol) may have a role in formation of the aqueous humour in the human eye[318], dihydroaldosterone (synthesised from aldosterone) limits the amount of sodium excreted via the kidneys in to urine[319].

One of 5-αR's main substrates in terms of concentration and biological relevance is the androgen testosterone (T) which is reduced to dihydrotestosterone (DHT). Pregnenolone and progesterone are converted into dihydroepiandrosterone and androstenedione, respectively, which can subsequently be converted into T and finally into DHT. Testosterone is the most abundant androgen present in serum[16]. Almost the totality of it (97%) is bound to albumen and globulin thus making the active fraction just 3% of the total[16] which can be converted to DHT. The structure of 5-αR bears numerous hydrophobic amino acids suggesting the location of the enzyme's active site may be situated deep inside the cell membrane[16] (microsomal enzyme).

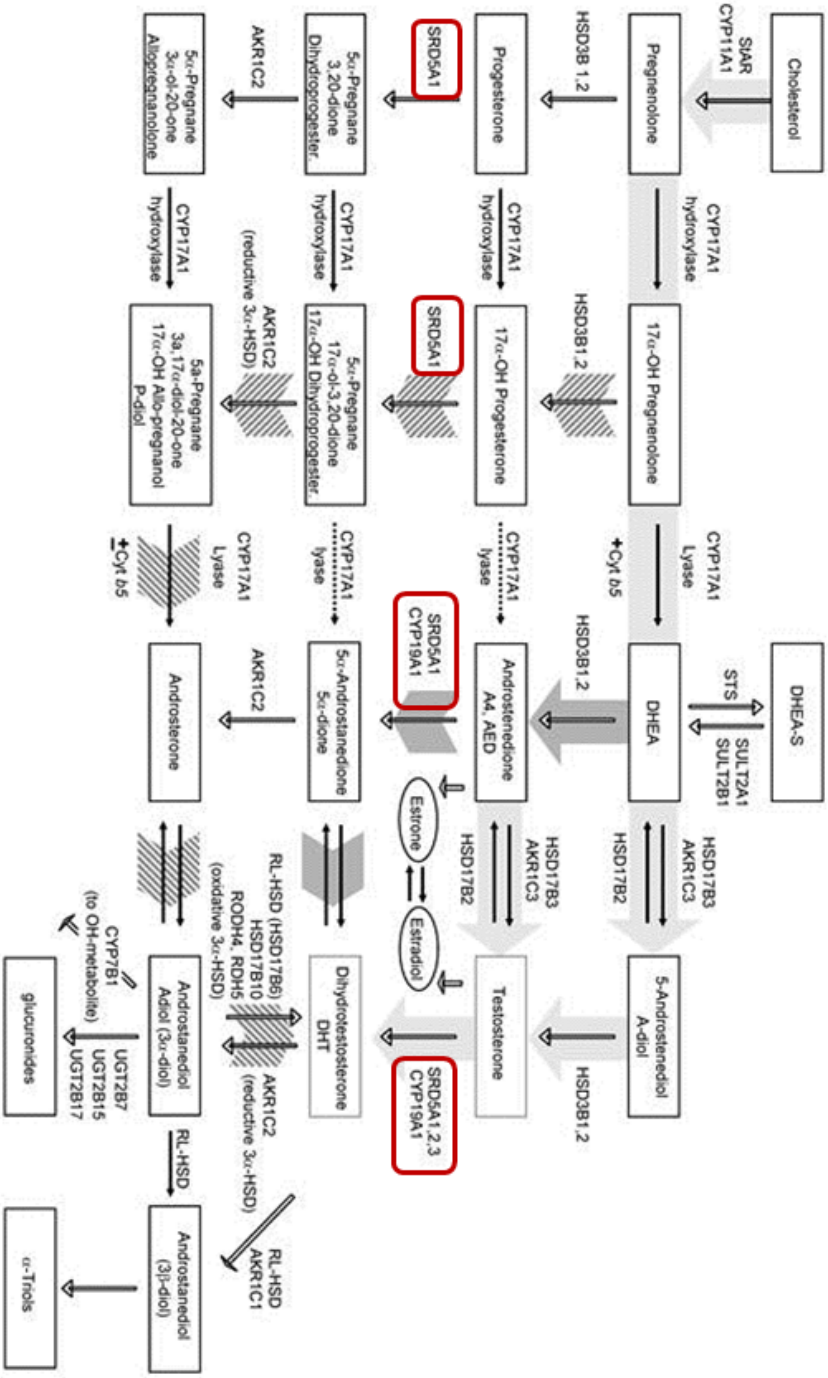


Figure 33. Overview of steroid metabolism. The reactions catalysed by 5-alpha reductase are highlighted in red[320].

While DHT in physiological conditions has a role in the development of sexual traits (male genitalia and prostate gland) it has also been proven to have a significant role in the development of benign prostatic hyperplasia and prostate cancer and therefore significant research has been dedicated to the study of its metabolism. Steroids like T and DHT bind to androgen receptors (AR) triggering several AR-regulated transcriptions. Such modifications occur through binding of AR to hormone response elements which are particular sequences of DNA or through recruitment of other transcription factors which may affect the target gene promoter region, chromatin remodelling or the structure of the AR itself[321]. AR activity has been linked for instance to up-regulation of insulin-like growth factor I receptor and sensitisation of prostate cancer cells to this protein[322]; moreover, if it interacts with serum response factor it seems to up-regulate the skeletal α-actin promoter[323]. The activity of AR has been also linked to biological mechanisms related to hyperplasia and prostate cancer[324]. One of the most important steroid metabolic pathways which involves the activity of 5-alpha reductase revolves round the reduction of T in to DHT. The obtained DHT binds more tightly to AR compared to T (up to three times lower dissociation rate[325] and ten times higher signalling induction[326]) showing an increased effect on AR-regulated transcription[326].

Three different isoforms of 5-alpha reductase have been identified. 5-αR1 is mainly present in liver and skin; 5-αR2 is present in prostate, liver and genital skin[327,328]; 5-αR3 has been recently identified as expressed in lung adenocarcinoma[329], papillary thyroid cancer and testicular seminoma[330]. All three isoforms are dependent on NADPH as a cofactor[331]. Two other enzymes (glycoprotein synaptic 2 (GPSN2) and glycoprotein synaptic 2-like (GPSN2L))[16] act through the same reaction (although this class of enzymes primarily performs fatty acid reduction rather than steroid reduction), hence they are often reported as an additional subfamily of the 5-α reductases. Different isoforms may play different roles in physiological and diseased conditions. For example, prostate cancer displays an up-regulation of 5-αR1 and a down-regulation of 5-αR2 if compared to benign prostate hyperplasia[16]. Recently isoform 5-αR3 has been reported as increased at both transcriptional[329] and protein level[330]

in prostate cancer. A disease which is related to 5-αR activity is 5-αR2 deficiency. Such impairment has been linked to potential pseudohermaphroditism, hypospadias and rudimentary prostate[332]. Moreover, lack of the 5-αR2 isoform is connected with decreased levels of circulating DHT and increased levels of T[324,332]. Diseases that affect the endocrine system often have an impact on the role of 5-αR in steroid metabolism.

Concentrations of T and DHT may vary significantly in different areas of the human body. DHT for instance is detected in the prostatic area with a concentration that is twice as high as the one observed in peripheral serum (478 ± 50 pg/ml vs. 950 ± 111 pg/ml)[333]. Conversely castrated rats displayed a 42 % drop in T concentration in prostatic tissue after 72h following castration whereas it was not possible to detect any relevant T concentration in peripheral serum after castration[334].

As defined above, T and DHT are both implicated in benign prostatic hyperplasia and prostate cancer, though DHT has a higher potency and so a reduction in the conversion of T to DHT could be applied as a treatment. Indeed treatments have largely been developed focusing on the inhibition of 5-αR through inhibitors with chemical structures that are similar to T or DHT. In fact, binding of an external compound to the active site of 5-αR would prove efficient by decreasing the synthesis of DHT. Inhibitors of 5-αR can be steroidal and non-steroidal. Since the primary substrates of 5-αR are steroids research has produced a higher number of steroidal inhibitors compared to the non-steroidal inhibitors. The mechanism of action of 5-αR involves binding of NADPH to the enzyme and later binding of this complex to the substrate. Firstly NADPH transfers a hydride anion to the carbon C-5; subsequently a proton binds to the carbon C-4 thus removing the double bond. Hence three types of competition can be achieved through use of inhibitors: they can bind to the enzyme to prevent later binding with NADPH, they can bind to the NADPH-enzyme complex thus hindering the substrate from binding or they can bind to the $NADP^+$-enzyme complex after completion of the reaction[16]. Two compounds that are widely used as a treatment in prostate cancer and prostatic hyperplasia are

finasteride and dutasteride (**Figure 34**). Both compounds are steroidal inhibitors. Finasteride is a 5-αR inhibitor which was approved by the Food and Drug Administration (FDA) in 1992 and since then has become largely applied in treatments against prostate cancer and hyperplasia. After six months of treatment with finasteride the concentration of serum DHT decreases by 71%[335]. A Pubmed search (performed in May 2016) with the term "finasteride" retrieved 2684 entries. Dutasteride was synthesised and commercialised later (late 1990s) and displays tighter protein binding than finasteride thus gaining large application in treatment of prostatic hyperplasia. A Pubmed search with the term "dutasteride" (performed in May 2016) retrieved 706 entries. The use of dutasteride offers clear advantages and efficiency which were unavailable with previous 5-αR inhibitors. Dutasteride was proven to reduce prostate volume, improve urinary flow and decrease urinary retention[336]. While finasteride binds to isoform 5-αR2 only, dutasteride is a ligand for both of the isoforms 5-αR1 and 5-αR2. Furthermore dutasteride binds to isoform 5-αR2 twice as tightly as finasteride[337]. After 6 months of treatment with dutasteride the DHT levels in serum decreased by 95% (compared to 71% for finasteride)[338]. Dutasteride also appears to inhibit isoform 5-αR3 much more efficiently than finasteride (IC_{50} = 0.33 nM and 17.4 nM respectively)[339]. Prostatic hyperplasia is dependent on mechanisms controlled by the 5-αR2 isoform.

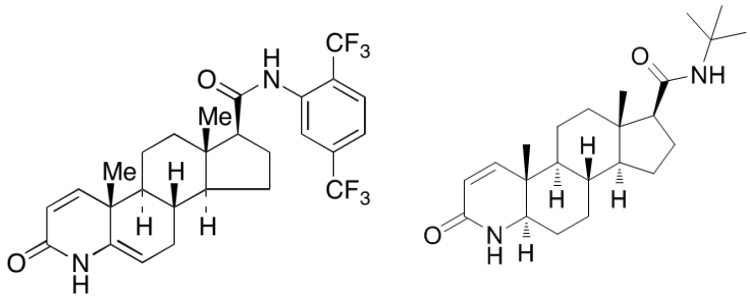


Figure 34. Dutasteride and finasteride chemical structures[340]

Prostate cancer is one of the leading causes of death in the male population (9% of deaths induced by all types of cancer) therefore diverse and numerous treatments have been applied. Administration of finasteride or dutasteride has provided positive results concerning prostate cancer prevention. Specifically finasteride decreased the risk of developing prostate cancer (as detected in biopsies) by 25%[341] while dutasteride reached 24%[336].

However, the global metabolic effects produced by 5-αR inhibitors are not thoroughly understood or described. Thompson et al.[341] performed an experiment over a 7-year period where an arm of 4368 healthy men were assigned to finasteride treatment and an arm of 4692 men were assigned to placebo. This study highlighted the lower occurrence of prostate cancer for the finasteride arm (18.4 vs. 24.4% for control $P<0.001$) counterbalanced by an incidence of higher grade tumours for the same arm (37.0 vs. 22.2% for control $P<0.001$). Other studies have demonstrated the detrimental effects of these two drugs on sexual traits. An exhaustive review published by Traish et al.[342] described the different side effects of 5-αR inhibitors, mainly focused on erectile dysfunction, decreased libido and depression. Another study highlighted dutasteride treatment as worse than finasteride in terms of incidence of side effects[343]. Incidence of erectile dysfunction, ejaculatory dysfunction and decreased libido was observed at 5.1%, 2.4% and 2.7%, respectively, for the dutasteride group compared with 2.1%, 1.8% and 1.4% for the finasteride group.

Interestingly, Upreti et al.[344] investigated the effect of finasteride and dutasteride treatment on insulin sensitivity monitoring glucose and nonesterified fatty acid disposal after insulin infusion for both treatments. While finasteride slightly increased glucose disposal (glucose disposal increased by 7.2 ± 3.0 µmol/kg fat-free mass/min), dutasteride was proven to severely impair glucose disposal (glucose disposal decreased by 5.7 ± 3.2 µmol/kg fat-free mass/min).

These findings highlight the necessity to investigate at a deeper level the effects of such inhibitors of global metabolism because 5-αR is an enzyme which is involved in a vast number of reactions in

steroid metabolism and is linked to substrates (e.g. T and DHT) which are metabolically relevant. As the effects of finasteride and dutasteride administration on global metabolism have not been investigated thoroughly, here we describe research which has taken advantage of non-targeted metabolomics techniques in order to evaluate the metabolic changes occurring during inhibition of 5-αR. Serum and urine from patients treated with finasteride and dutasteride in conjunction with insulin infusion have been analysed applying UPLC-MS followed by univariate and multivariate analysis.

5.2 Results and discussion

Finasteride and dutasteride were detected and annotated in serum (positive ionisation mode). Finasteride was detected as $[M+H]^+$ at m/z=373.2855 while dutasteride was detected as $[M+H]^+$ at m/z=529.2290. Moreover the most abundant phase I metabolite of finasteride (ω-hydroxy finasteride (M1)) was detected as $[M+H]^+$ at m/z=389.2804 as reported by Lundahl et al.[345] (**Figure 35** to **Figure 39**).

Finasteride and its main metabolite (M1) were also detected in urine (positive ionisation mode) while dutasteride was not detected as expected since it is mainly excreted in faeces[346].

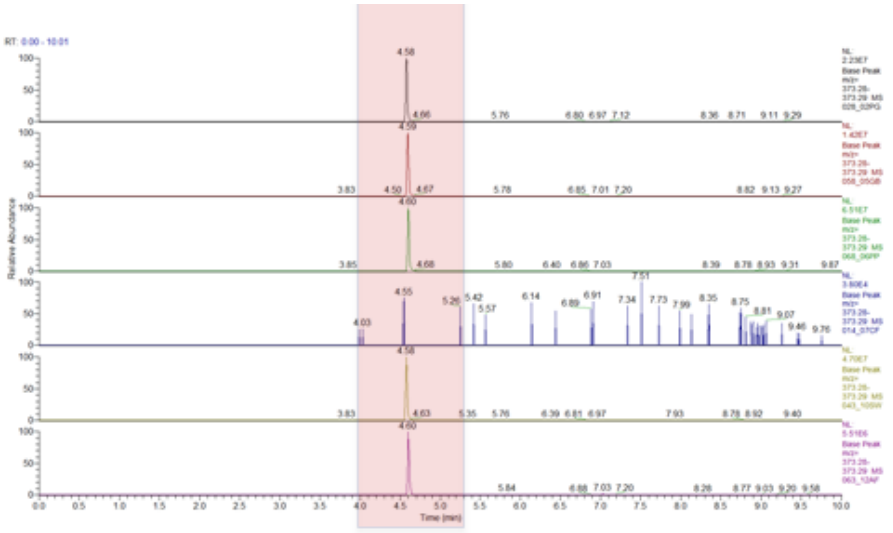


Figure 35. Detection of finasteride in serum for each of the six subjects undergoing the finasteride treatment arm.

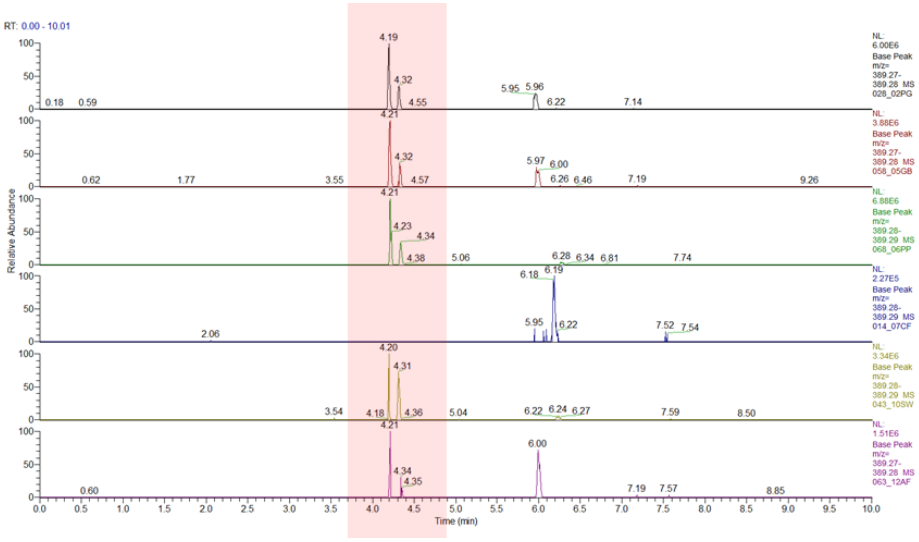


Figure 36. Detection of M1 in serum for each of the six subjects undergoing the finasteride treatment arm.

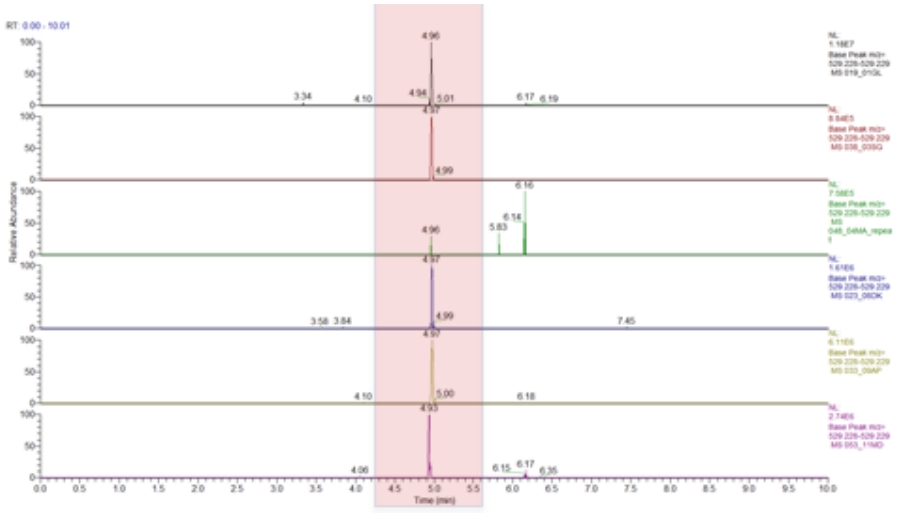


Figure 37. Detection of Dutasteride in serum for each of the six subjects undergoing the dutasteride treatment arm.

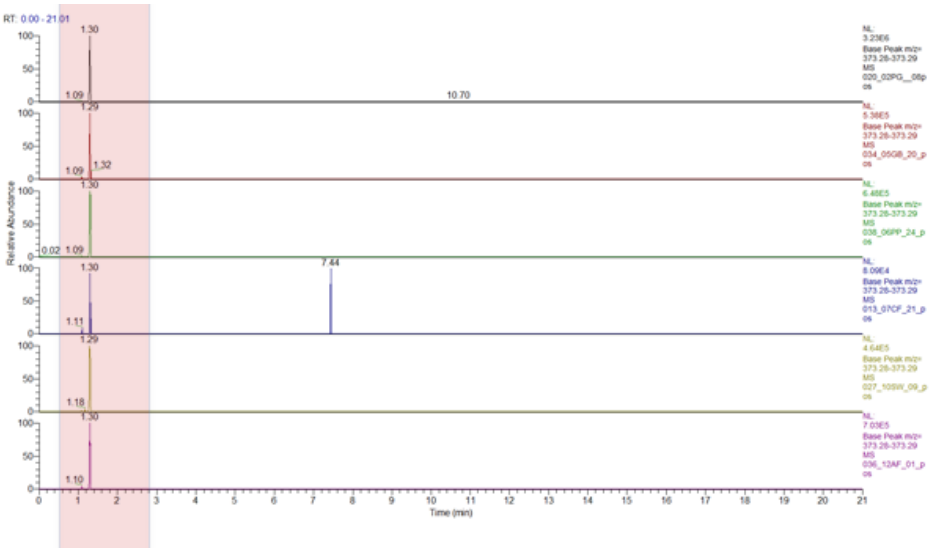


Figure 38. Detection of finasteride in urine for each of the six subjects undergoing the finasteride treatment arm.

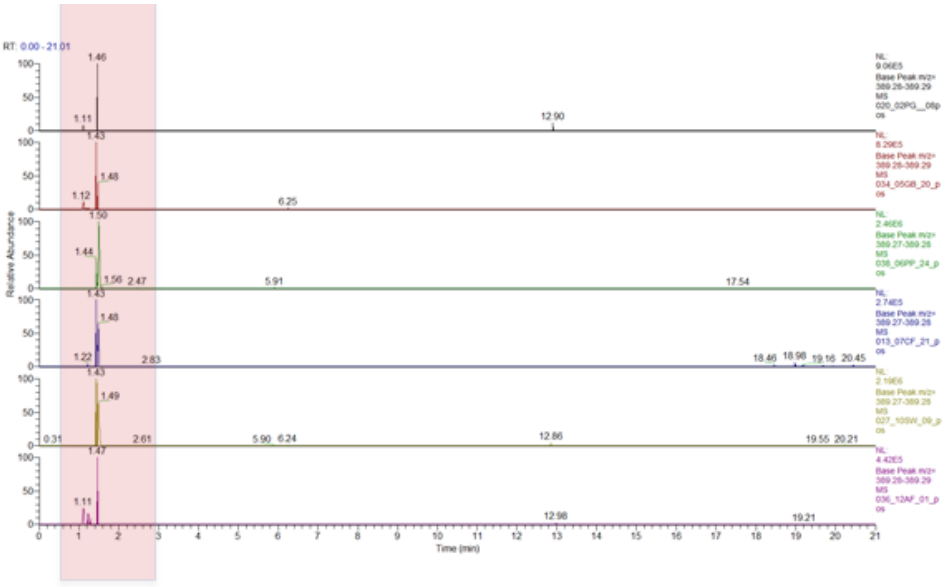


Figure 39. Detection of M1 in urine for each of the six subjects undergoing the finasteride treatment arm.

5.2.1 Univariate analysis

5.2.1.1 Serum

The effect of each drug was analysed by investigating the following comparisons: (1) before and after insulin infusion on the first visit (effect of insulin only, i.e. fasted to fed state), (2) before and after insulin infusion on the second visit (effect of drug on response to insulin, i.e. fed state), (3) before and after drug treatment pre-insulin infusion (effect of drug at low insulin levels, i.e. fasted state) and (4) before and after drug treatment post-insulin infusion (effect of drug at high insulin levels, i.e. fed state).

The serum dataset obtained by applying a monophasic extraction method and analysed in positive ion mode contained 11,278 metabolite features which were reduced to 9096 after missing value filtering and to 8198 after quality filtering. The dataset obtained by applying a biphasic extraction method and analysed in positive ion mode contained 9311 metabolite features which were reduced

to 8331 after missing value filtering and to 7486 after quality filtering. The dataset obtained by applying a monophasic extraction method and analysed in negative ion mode contained 10,516 metabolite features which were reduced to 8951 after missing value filtering and to 7257 after quality filtering. The dataset obtained by applying a biphasic extraction method and analysed in negative ion mode contained 8359 metabolite features which were reduced to 7607 after missing value filtering and to 6805 after quality filtering.

Following univariate analysis, multiple features reporting the same metabolite were reduced to a single metabolite feature (the metabolite feature recorded with the lowest p-value was retained). Finasteride showed 576 metabolite features which were statistically significant ($p<0.05$) in at least one of the four comparisons described above. The results are summarised in **Table 22**.

Metabolite class	Statistically significant changes observed in comparison (1)	Statistically significant changes observed in comparison (2)	Statistically significant changes observed in comparison (3)	Statistically significant changes observed in comparison (4)
Acyl carnitines	↓ [21]	↓ [11]	MIX [2]	↑ [1]
Mono and di-acyl glycerides	↓ [12]	↓ [2]	MIX [2]	↓ [6]
Triacylglycerides	↑ [12]	↓ [2]	↑ [1]	/
Ceramides	↓ [14]	MIX [7]	MIX [8]	↑ [2]
Fatty acids	↓ [25]	↓ [15]	↑ [2]	↓ [5]
Oxidised fatty acids	↓ [25]	↓ [11]	↑ [3]	↓ [3]
Peptides	MIX [16]	MIX [7]	MIX [3]	↑ [5]
Glycerophopsholipids	MIX [16]	MIX [33]	↑ [12]	↓ [14]

Table 22. Summary of statistically significant differences for comparison (1), (2), (3) and (4) after finasteride administration. Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison. The number of statistically significant metabolite features is reported in square brackets. MIX denotes a mixed trend.

Comparison (1) summarised the effect of insulin infusion only since at this stage the drug treatment had not started. 378 metabolite features (**Appendix Table 23**) were reported as statistically significant ($p<0.05$). The biological classes involved included 21 acyl carnitines generally decreasing at high insulin levels, 12 mono- and diacylglycerides decreasing at high insulin levels, 12 triacylglycerides generally increasing at high insulin levels, 14 ceramides generally decreasing at high insulin levels, 25 fatty acids most decreasing at high insulin levels, 5 phosphatidylcholines and 9 phosphatidylinositoles increasing at high insulin levels, 6 lysoglycerophospholipids decreasing at high insulin levels and 25 oxidised fatty acids decreasing at high insulin levels. This outcome may suggest the occurrence of increased triacylglyceride synthesis rather than lipolysis. In fact, fatty acids and oxidised fatty acids decrease in concentration at high insulin levels implying an enhancement of storage of fatty acids in adipose fat as suggested by high levels of triacylglycerides and glycerophospholipids. Specifically, it seems that metabolic pathways leading to fatty acid β-oxidation and ceramide synthesis are attenuated hence promoting fatty acids in to triacylglyceride synthesis. These results are similar to the conclusions drawn in chapter 4 for the same comparison.

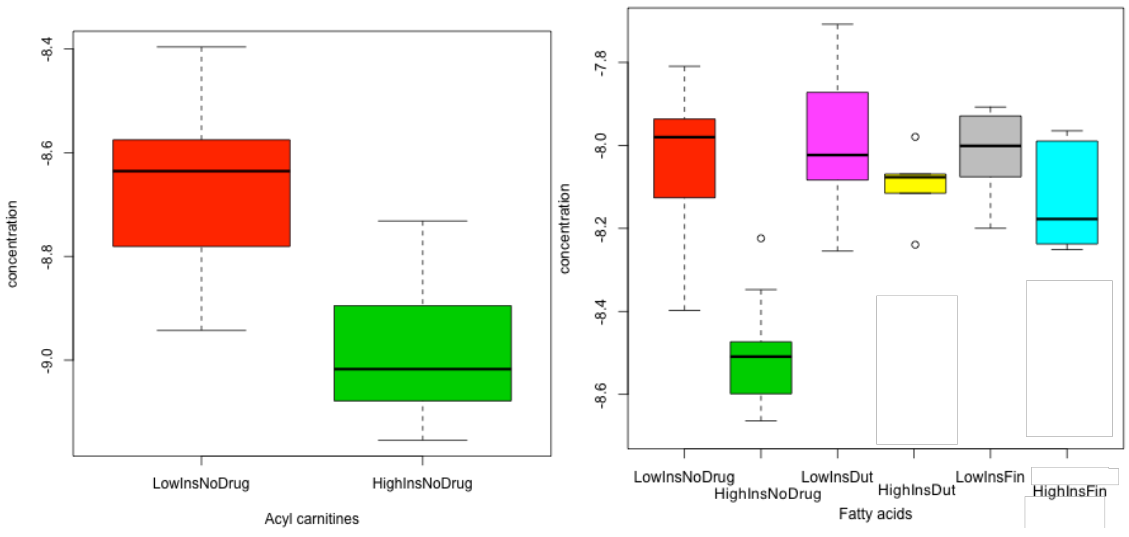


Figure 40. Boxplot of the glog transformed concentrations of all acyl carnitines (left) and fatty acids (right). Concentrations in red: low insulin level and no drug treatment, green: high insulin level and no drug treatment, magenta: low insulin level after dutasteride administration, yellow: high insulin level after dutasteride administration, grey: low insulin level after finasteride administration, light blue: high insulin level after finasteride administration.

Comparison (2) described 167 statistically significant metabolite features ($p<0.05$). The results are shown in **Appendix Table 24**.

The reduction of the number of statistically significant metabolite features observed in comparison (1) (effect of insulin only, 375 metabolite features) to comparison (2) (effect of drug on response to insulin, 167 metabolite features) may indicate impaired insulin sensitivity development after drug administration. In fact drug administration suppresses several metabolic changes observed in comparison (1). Changes significant in comparison (1) and negated after drug treatment included: 10 acyl carnitines, 14 ceramides, 10 fatty acids, 33 glycerophospholipids and 17 oxidised fatty acids. (**Figure 40**). Metabolite features which were common between comparisons (1) and (2) were investigated as these features represent the metabolite features which change due to insulin administration regardless of drug administration. 88 metabolite features were statistically significant in both comparisons including 11 acyl carnitines and 14 fatty acids. The metabolite features which were significant in comparison (1) only represent the metabolite features which change due to insulin infusion but which are suppressed by drug administration. 290 metabolite features were identified as belonging to this class. Lastly, metabolite features which were statistically significant in comparison (2) only represent metabolite features which are changing after insulin infusion but are only triggered by drug administration. 79 metabolite features appeared to be statistically significant in comparison (2) only. Therefore the drop in the number of significant metabolite features from comparison (1) only to comparison (1) and comparison (2) (metabolite features common in both comparisons) highlights the global suppression effect of finasteride on insulin sensitivity.

Comparison (3) investigated the changes provoked by the drug treatment without any effect from insulin (i.e. the effect of drug in a fasted state). The result described 68 statistically significant metabolite features ($p<0.05$; **Appendix Table 25**). 12 of 14 detected glycerophospholipids increased in concentration after treatment. Notably, dihydrotestosterone was detected as its sulfated

metabolite with more than a 4-fold decrease after drug administration for 3 weeks. This outcome is expected as the primary objective of 5-αR inhibitors is to hinder the synthesis of dihydrotestosterone from testosterone. However, testosterone was not detected suggesting either a compensation mechanism which balances the steroid concentration or an inability to detect the metabolite (which does not have chemical groups that could be ionised easily).

Comparison (4) described 96 statistically significant ($p<0.05$) metabolite features (**Appendix Table 26**) which represent the metabolic changes caused by finasteride administration in the fed state. 6 acyl glycerides decreased in concentration at high insulin levels, 14 out of 18 glycerophospholipids decreased in concentration at high insulin levels, 5 out of 6 peptides increased in concentration at high insulin levels and 7 out of 8 steroids decreased in concentration at high insulin levels. Here, while the drug administration appears to have a marked impact on steroid metabolism in general (affecting not only dihydrotestosterone), it seems that finasteride administration cooperated with insulin resulting in a negation of fatty acid β-oxidation and glycerophospholipid synthesis. These effects however, need to be confirmed in larger cohort studies.

After removal of multiple metabolites features reporting the same metabolite (the metabolite feature recorded with the lowest p-value was retained), dutasteride described 661 metabolite features which were statistically significant ($p<0.063$) in at least one of the mentioned classes. The significance threshold was set to 0.063 instead of 0.05 because a sample was lost during the sample extraction process so the statistical population reduced to $n=5$ rather than $n=6$ making it impossible to achieve a critical p-value of 0.05 with a non-parametric statistical test. A summary of the results is shown in **Table 23**.

Metabolite class	Statistically significant changes observed in comparison (1)	Statistically significant changes observed in comparison (2)	Statistically significant changes observed in comparison (3)	Statistically significant changes observed in comparison (4)
Acyl carnitines	↓ [8]	↓ [8]	↓ [1]	↓ [1]
Mono and di-acyl glycerides	↓[5]	↓[7]	↑ [2]	↓ [2]
Triacylglycerides	MIX [8]	MIX [5]	↑ [2]	↓ [5]
Ceramides	↑ [9]	↓[6]	MIX [6]	↓ [11]
Fatty acids	MIX [27]	↓ [18]	MIX [6]	↓ [6]
Oxidised fatty acids	↓ [21]	↓ [25]	MIX [7]	↓ [9]
Peptides	MIX [15]	↓ [7]	↑ [7]	↓ [5]
Glycerophopsholipids	↑ [27]	MIX [58]	↑ [26]	↓ [41]

Table 23. Summary of statistically significant differences for comparison (1), (2), (3) and (4) after dutasteride administration. Metabolite features are grouped in classes of similar chemical structure. The arrows report up- or down-regulation of a class in a specific comparison. The number of changing metabolite features is reported in square brackets. MIX indicates a mixed trend.

Comparison (1) described 302 statistically significant metabolite features (**Appendix Table 27**). The following results refer to high levels of insulin. The changing metabolite features included 8 acyl carnitines decreased, 6 amino acids increased, 9 ceramides increased, 8 lysoglycerophosphoetanolamines decreased, 27 glycerophospholipids increased and 21 out of 27 oxidised fatty acids decreased. These metabolite features may confirm the conclusion introduced by the finasteride dataset which implied an involvement of glycerophospholipid synthesis.

Comparison (2) described 271 statistically significant metabolite features (**Appendix Table 28**). As for finasteride, dutasteride negates the effect of insulin on metabolism. The following results refer to high levels of insulin. The metabolite features included 8 di- and triacylglycerides decreased, 9 ceramides decreased, 13 fatty acids decreased, 24 glycerophospholipids decreased, 11 oxidised fatty acids decreased and 11 peptides decreased. 36 metabolite features were statistically significant in comparisons (1) and (2) including 3 glycerophosphoinositols. 235 metabolite features were statistically significant in comparison (1) only and 143 metabolite features were significant in comparison (2) only including 7 acyl carnitines and 11 ceramides. Notably, ceramides have been reported as mediators of insulin resistance[314]. The difference between metabolite features significant in comparison (1) and (2) may be addressed to a general metabolic suppression induced by dutasteride administration. The nature of such negation can be observed in **Figure 41**.

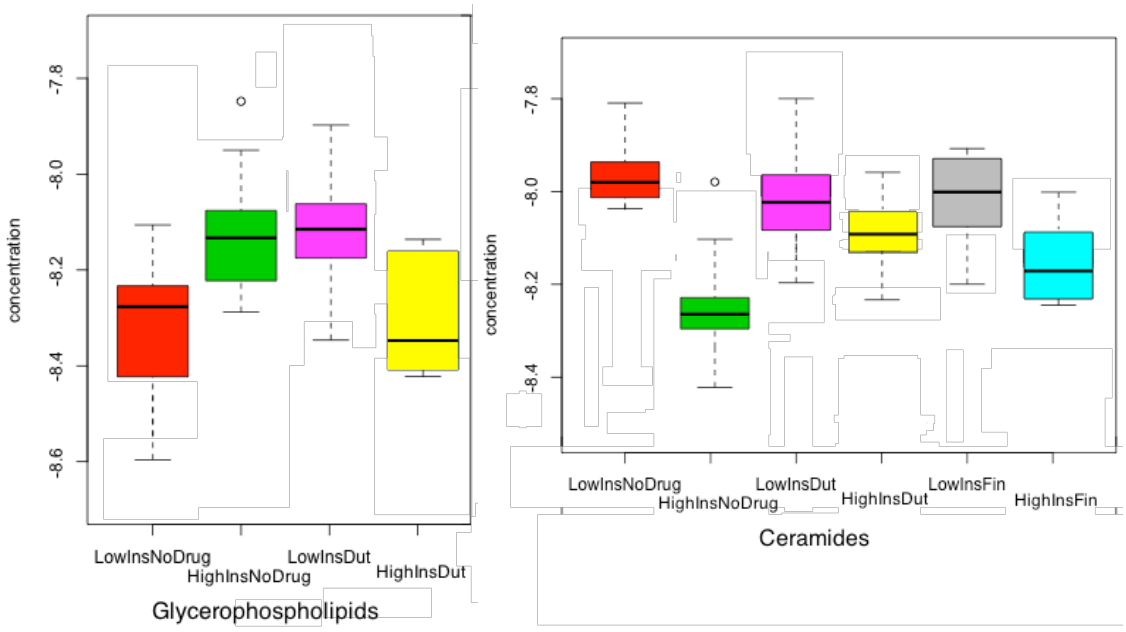


Figure 41. Boxplot of the glog transformed concentrations of all glycerophospholipids (left) and ceramides (right). Concentrations in red: low insulin level and no drug treatment, green: high insulin level and no drug treatment, magenta: low insulin level after dutasteride administration, yellow: high insulin level after dutasteride administration, grey: low insulin level after finasteride administration, light blue: high insulin level after finasteride administration.

Comparison (3) described 132 statistically significant metabolite features (**Appendix Table 29**) affected by dutasteride administration at low insulin levels (fasted state). The fact that 132 metabolite features were identified after dutasteride treatment while only 67 metabolite features were described after finasteride treatment may be linked to the inhibition of both isoforms of 5-αR by dutasteride whereas finasteride inhibits the 5-αR1 isoform only. The following results refer to the post- drug administration condition. Classes affected by dutasteride included: 4 di- and triacylglycerides increased, 4 amino acids increased, 26 glycerophospholipids increased and 7 peptides increased. Interestingly, the high number of glycerophospholipids increased in concentration corresponds to an enhancement of hepatic lipid accumulation as pointed out by magnetic resonance spectroscopy[347]. However, the metabolic effect detected in the present non-targeted study does not clearly indicate activation of lipogenesis at a global level.

Comparison (4) described 188 statistically significant ($p<0.05$) metabolite features (**Appendix Table 30**). The following results refer to the post- drug administration condition. The changing metabolite features included 7 acyl glycerides decreased, 11 out of 12 ceramides decreased, 6 fatty acids decreased, 39 out of 41 glycerophospholipids decreased, 9 out of 10 oxidised fatty acids decreased, 5 peptides decreased and 5 steroids decreased. Although the majority of changes complied with the results observed for finasteride, some differences must be noted. First, the number of significant metabolite features is much higher for dutasteride. This is probably due to the higher potency of dutasteride. Second, while peptides were increasing after finasteride treatment in a fed state, here they are decreasing, possibly suggesting a peculiar effect of dutasteride on protein degradation.

5.2.1.2 Urine

Data for only two timepoints were available for each subject and so one statistical comparison was performed (before drug administration vs. after drug administration). Finasteride and dutasteride administration reported only 62 statistically significant metabolites ($p<0.05$), 56 metabolite features were significant for finasteride only, 56 metabolite features were significant for durasteride only and 6 metabolite features were significant for both drugs (**Appendix Table 31** and **32**). Finasteride did not describe any biological trend, the biological classes detected as significant do not show uniform behaviour (no class is enriched). However, dutasteride reported several classes as being increased after administration including 5 amino acids, 7 metabolite features involved in carbohydrate metabolism, 6 glycerophopsholipids and 5 peptides. It appears that dutasteride affects global metabolism. Protein degradation appears to increase (peptides and amino acids are increased in concentration after administration) and a possible activation of gluconeogenesis or pentose-phosphate pathway is witnessed by the enhancement of carbohydrate anabolism. Another possible mechanism underlying the evidence could be an enhancement of glycolysis. This theory is not only

supported by the increase of carbohydrate metabolites but also by the increase of fatty acids. In fact, fatty acids can be synthesised from acetyl-CoA generated as an end product of the glycolysis.

Nonetheless it is evident that the number of statistically significant metabolite features in urine was much lower when compared to serum. Another reason may be the choice of a collection method (spot urine) that is not optimal when studying global metabolism; future studies may contemplate the use of a 24-hour urine collection. This recommendation reflects the need to include the array of diurnal metabolic changes in a single sample. In fact, the spot urine collection represents a single moment in the day that may not be representative of the global biological changes of interest.

5.2.2 Multivariate analysis

5.2.2.1 Serum

After QC filtering for biphasic extraction serum datasets, the PCA scores plots showed good clustering of QC samples indicating good technical reproducibility. No class separation based on insulin administration or drug treatment was observed in the PCA scores plots (**Figure 42** and **Figure 43**). Since PCA scores plots did not point out any clear biological trend, PLS-DA models have been investigated.

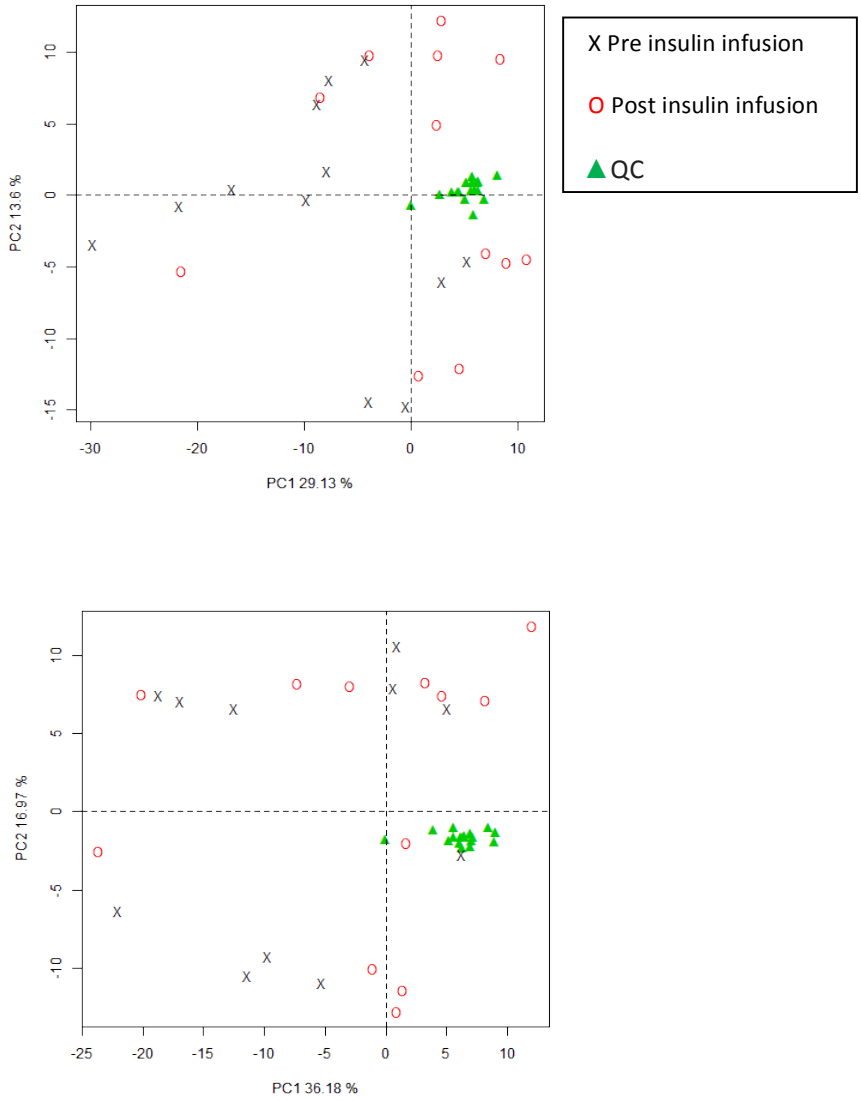


Figure 42. PCA scores plots (biphasic/pos) comparing pre and post insulin infusion for finasteride (top) and dutasteride (bottom). The QC samples are clustered

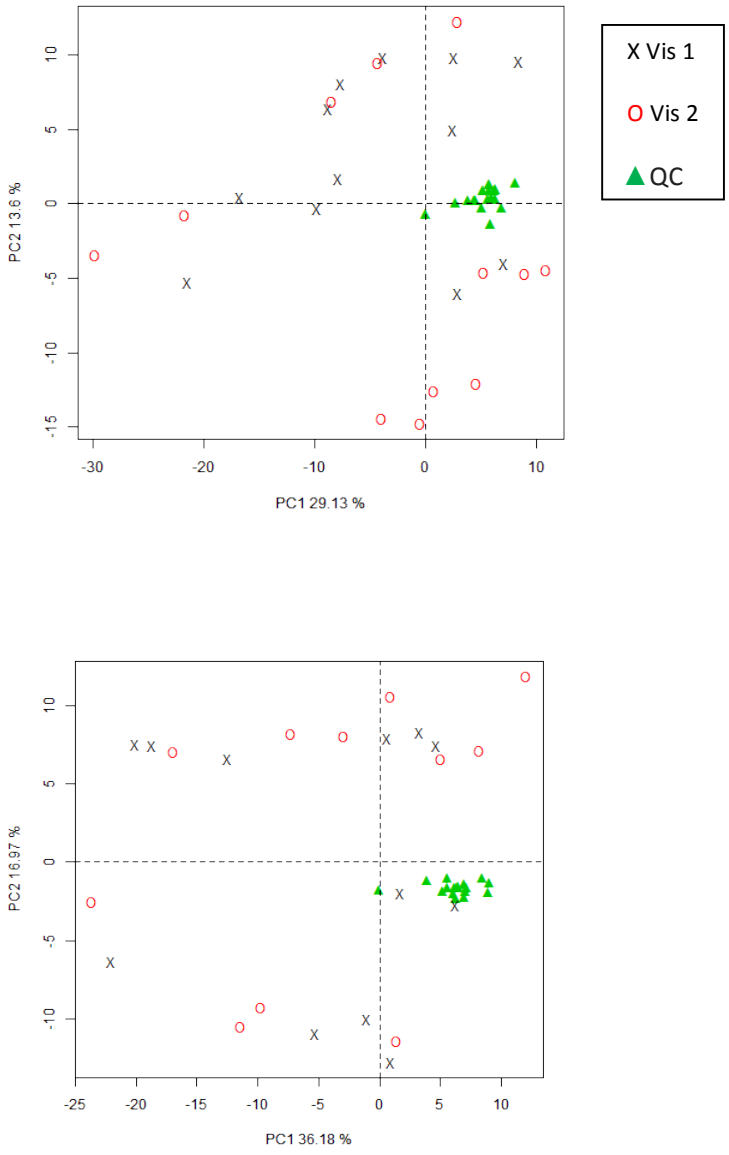


Figure 43. PCA scores plots (biphasic/pos) comparing visit 1 and visit 2 for finasteride (top) and dutasteride (bottom). The QC samples are clustered

The effect of drug administration and insulin infusion was investigated applying PLS-DA as follows. Firstly, in order to explore the effect of drug administration, PLS-DA models were implemented using drug administration as the response variable. However drug administration did not influence the multivariate outcome producing models with low R^2 and Q^2 values (≈0.1).

Subsequently, insulin infusion was adopted as the explanatory variable and PLS-DA models were constructed. Some of these models reported acceptable R^2 and Q^2 values indicating a robust cross-validated model. The biphasic/positive dataset resulted in R^2 and Q^2 of 0.6204 and 0.523, respectively while the biphasic/negative dataset resulted in 0.8126 and 0.4983, respectively (**Figure 44** and **Figure 45**). It must be remarked that these values were obtained using all of the samples present in the dataset (excluding QCs), however when the datasets were separated in to samples administered with finasteride and samples administered with dutasteride, the cross-validation values decreased possibly depending on the reduced number of samples causing the model to be poorly cross-validated.

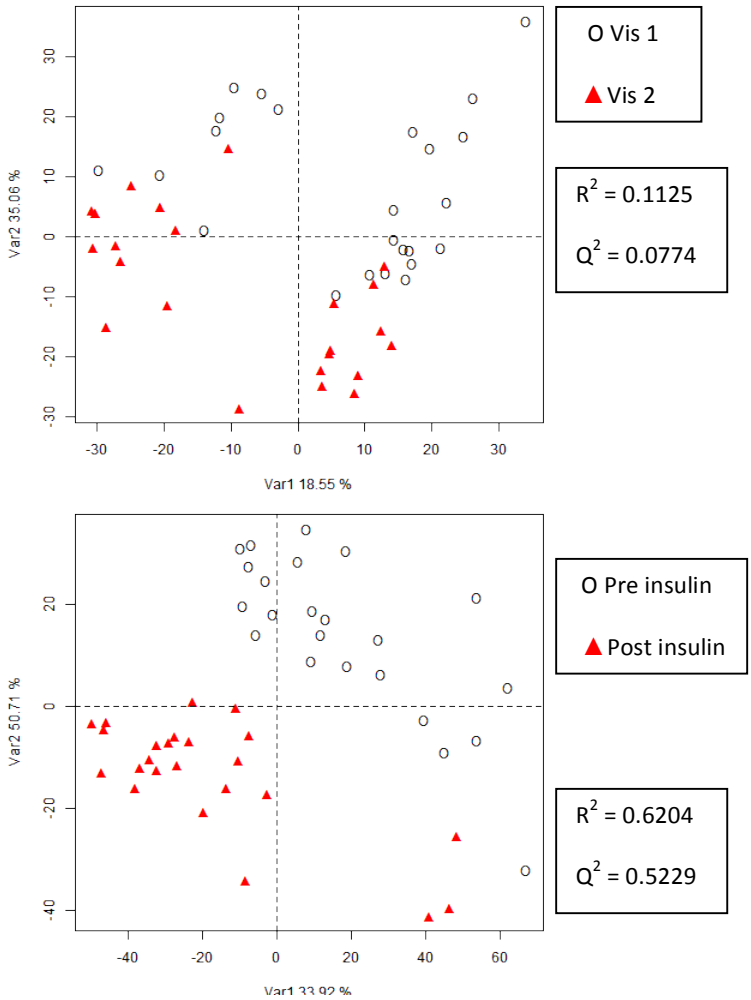


Figure 44. PLS-DA plots for biphasic extraction/positive ion mode. Top: Drug administration used as explanatory variable. Bottom: Insulin infusion used as explanatory variable.

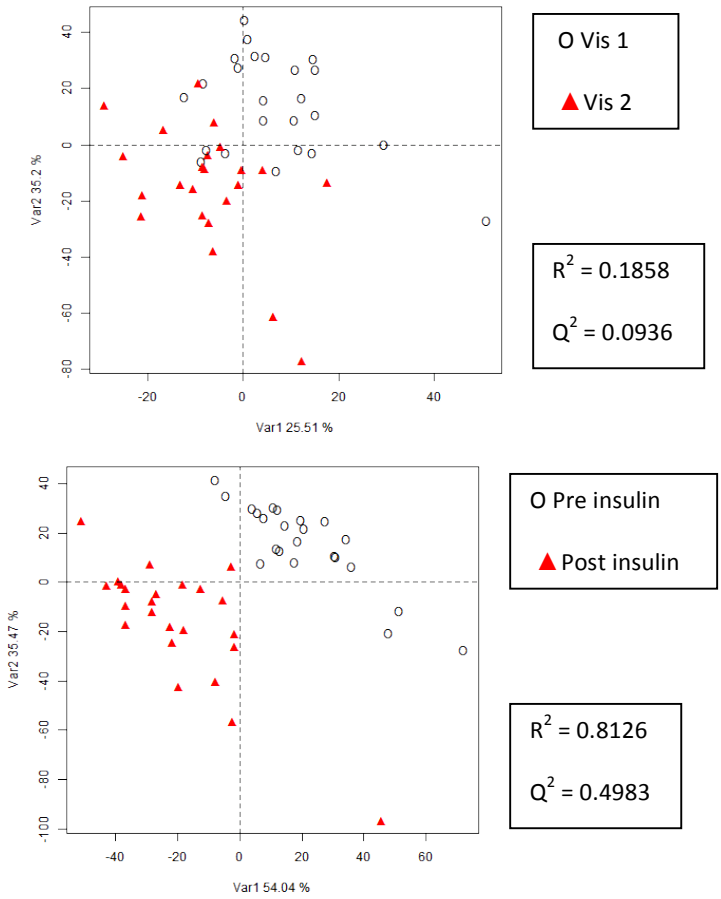


Figure 45. PLS-DA plots for biphasic extraction/negative ion mode. Top: Drug administration used as explanatory variable. Bottom: Insulin infusion used as explanatory variable.

This outcome suggests that the major effect on the global human metabolome is caused by insulin infusion rather than drug administration. In fact, univariate analysis confirms this conclusion since the majority of statistically significant metabolite features were observed when comparing before and after insulin infusion. Although this seems to be the main factor, univariate analysis revealed that drug administration had a negating effect on the changes induced by insulin which cannot be detected in the multivariate analysis results.

5.2.2.2 Urine

After QC filtering the PCA scores plots showed good clustering of QC samples representing data reliability and instrument stability. PCA scores plots did not show any particular separation of samples according to drug administration. Furthermore, PLS-DA models constructed poor cross-validatied models, probably due to the reduced sample size.

5.3 Conclusion

This study highlighted the effect of insulin infusion on human global metabolism as observed in serum, encompassing changes in acyl carnitines, acylglycerides, fatty acids, oxidised fatty acids and glycerophospholipids. Globally, these changes seem to suggest the occurrence of lipogenesis with an increased consumption rate of fatty acids involved in their storage in complex lipids. The same result was obtained in chapter 4 after analysis of samples before and after insulin infusion.

Metabolite class	N° metabolite features changing in comparison (1) (Fin.)	N° metabolite features negated in comparison (2) (Fin.)	N° metabolite features changing in comparison (1) (Dut.)	N° metabolite features negated in comparison (2) (Dut.)
Acyl carnitines	21	10	10	/
Mono- and di-acyl glycerides	12	/	6	/
Triacylglycerides	12	/	8	5
Ceramides	14	14	10	10
Fatty acids	25	10	27	13
Oxidised fatty acids	25	17	27	12
Peptides	16	/	15	11
Glycerophopsholipids	43	33	30	24

Table 24. Summary of metabolite features changing in comparison (1) and (2) grouped by biological classes. Fin.: finasteride; Dut.: dutasteride; Nº: number. A metabolite feature is changing when p-value < 0.05 for finasteride and < 0.0063 for dutasteride.

This result fits well with the evidence of insulin acting as a promoter of glycogenesis and suppressor of gluconeogenesis and glycogenolysis[315] in non-diabetic subjects. Furthermore, administration of finasteride or dutasteride appeared to reduce the influence of insulin infusion on the global metabolome.

Drug	Visit 1-pre vs Visit 1-post (insulin only)	Visit 1-pre vs Visit 2-pre (drug at low insulin only)	Visit 1-post vs Visit 2-post (drug at high insulin only)	Visit 2-pre vs Visit 2-post (insulin after drug treatment)
Fin.	378	67	36	167
Dut.	271	132	166	179

Table 25. Number of metabolite features changing in different comparisons. Fin.: finasteride; Dut.: dutasteride. A metabolite feature is changing when p-value < 0.05 for finasteride and < 0.0063 for dutasteride.

Specifically, a number of metabolite features which changed after insulin infusion did not display the same trend after drug administration and insulin infusion suggesting a potential suppression of the effect of insulin (insulin resistance) operated by 5-αR activity and following inhibition of this enzyme. Metabolite features influenced were acyl carnitines, fatty acids and glycerophopsholipids for finasteride treatment and acyl glycerides, glycerophospholipids, fatty acids and peptides for dutasteride treatment. However, finasteride and dutasteride administration on its own seemed to achieve much fewer effects on the global metabolome as witnessed by the lower number of statistically significant metabolite features. Nonetheless, glycerophospholipids appeared to be increased after drug administration independent of the drug. This outcome may possibly link to an enhancement of hepatic lipid accumulation[347], reported in the same study. It is peculiar that dihydrotestosterone, which was detected as decreasing after finasteride administration, is not observed to be statistically significant after dutasteride administration. However, dutasteride

administration caused a higher number of metabolite features to change compared to finasteride, probably due to the dual inhibition of 5-αR isoforms.

Other interesting observations on the metabolic effects of dutasteride were drawn from comparison (4) in serum and from the analysis of urine. Dutasteride appears to have an effect on protein degradation and possibly gluconeogenesis or glycolysis. While the analysis of urine highlights an enhancement of this pathway (reporting peptides and amino acids as increased in concentration), comparison (4) in serum, which denotes the effect of dutasteride in a fed state, reports peptides as decreased in concentration in serum possibly highlighting a cooperative mechanism of dutasteride with insulin to protein metabolism.

These observations might introduce more awareness concerning finasteride and dutasteride prescription since these drugs trigger mechanisms with a wide range of effects. In fact, 5-αR-related diseases proved to display potential links with insulin-resistance syndromes such as diabetes, Cushing's syndrome and metabolic syndrome. Furthermore, the occurrence of lipogenesis and reduction of fatty acid beta-oxidation triggered by insulin infusion sheds light on the actual risks which may be caused by drugs provoking insulin resistance i.e. the administration of such drugs may have a negative effect on the physiological occurrence of anabolic processes involving lipids (and possibly carbohydrates).

Concerning the analytical aspects of this study, it should be remarked that the biphasic extraction (non-polar phase) produced datasets bearing a larger number of statistically significant metabolite features when dealing with lipids and fatty acids which in this case proved to be central for the study. Therefore such an extraction method should be preferred when changes in lipid metabolism are expected. It is also true that sometimes it is required to generate a global snapshot of metabolism without focusing on particular metabolic classes. In these cases, if it is possible, it may be appropriate

to perform a biphasic extraction and integrate the data from the two extractions in a water/methanol solution for reconstitution so to have the maximum coverage of metabolites[348].

Future studies may focus on the effect of 5-αR inhibitors on specific classes of metabolites or specific indicators of insulin-resistance. For example, semi-targeted studies on steroids could shed light on the specific transformation occurring in the steroids metabolic pathway. Furthermore, correlation between DHT inhibition and insulin sensitivity impairment could be calculated showing whether there is a precise connection between the two mechanisms. Possibly, by increasing the sample size more robust statistical tests can be carried out allowing identification of more metabolites involved in 5-αR associated steroid metabolism.

CHAPTER 6 – Conclusion

This PhD thesis was aimed to shed light on some of the unclear aspects concerning the metabolism of steroids (and in particular GCs) and their effects on other metabolic pathways. In order to do so the research was structured in three parts: each one tried to answer a specific problem.

The first part investigated an array of data pre-processing methods commonly used in UPLC-MS metabolomics and detected the most efficient permutations with respect to univariate and multivariate analysis. This study pointed out Random Forest imputation as the preferred missing value imputation method since it outperforms all the other tested methods. However, its application must be evaluated case to case. In univariate analysis for example, it is advisable not to use any imputation method and no further processing methods. When dealing with multivariate statistics the conclusions change. Before performing PCA the recommended pre-processing methods to adopt would be PQN normalisation, Random Forest imputation and glog transformation. This permutation highlights the metabolite features bearing a large fold change regardless of their abundance. For PLS-DA the recommended permutation would be PQN (or SUM normalisation) followed by KNN or BPCA replacement and glog transformation. These results managed to provide guidelines that apply not exclusively to this study but to UPLC-MS untargeted metabolomics in general. The overall aim is to propose the aforementioned treatment permutations as the standard for untargeted metabolomics.

Once the computational method was tuned and set up, we proceeded to investigate the second part of the scientific question: how do glucocorticoids affect the global metabolism and how do they interact with insulin? This study highlighted the difference between administration of exogenous cortisol and naturally occurring high levels of cortisol. The two conditions produced different results; nevertheless the glycerophospholipids appeared to increase in both cases. Generally, the decrease of

fatty acids seems to suggest an impairment of lipogenesis as anticipated by Long. Cortisol also caused ceramides and sphingomyelins (mediators of insulin resistance) to increase, possibly suggesting a role on insulin sensitivity that is exerted in a short time scale. The general metabolic changes caused by cortisol also seem to blunt at high levels of insulin thus confirming a relation between cortisol induced changes and insulin sensitivity. The interaction between insulin and cortisol also caused glycerophospholipids and ceramides to decrease while they appeared as increasing after insulin administration. Furthermore, dexamethasone (commonly used as an exogenous GC in medical treatment) displayed enhanced β-oxidation and contrasting results concerning glycerophospholipids hence showing different behaviour when compared to cortisol. This part of the study not only provided further insight to the known effect of cortisol and insulin but also defined their interaction and spotted affected metabolite classes that could be investigated in dedicated studies. It also revealed that the time scale of cortisol effect on insulin sensitivity is extremely rapid (less than a day). Moreover, the different metabolic effects of dexamethasone and cortisol were evaluated thus highlighting the fact that GC replacement therapy through dexamethasone does not have the same effect on the global metabolome that cortisol has. Specifically glycerophospholipids seem to behave differently in the presence of dexamethasone or cortisol and also β-oxidation, which appeared as enhanced after dexamethasone administration, is not confirmed at high levels of cortisol.

After investigating the effects of GCs on global metabolism, the third part of the research study focused on the interaction between other steroidal metabolic pathways and the global metabolome. In order to do so, the inhibition of 5-αR was examined through administration of finasteride and dutasteride. In particular these drugs blunted the majority of the changes induced by insulin in serum. Specifically fatty acids, oxidised fatty acids and acyl carnitines were the metabolic classes that were affected. This outcome suggests that steroidal metabolic pathways and their interaction with insulin heavily influence the carnitine shuttle and β-oxidation. In fact, insulin as already known hinders gluconeogenesis and promotes glycogenesis; this behaviour is confirmed by the decrease of

fatty acids and acyl carnitines. However the 5-αR inhibitors markedly diminished this effect thus proving a strong interdependency between different steroidal pathways. Furthermore, dutasteride displayed a larger effect on the global metabolome reflecting the greater drug potency when compared to finasteride. These findings are particularly relevant considering the large application of treatments involving administration of finasteride and dutasteride. In fact, although these drugs are commonly prescribed, to our knowledge no interaction with insulin has been reported yet. Thus while the effects of GCs are investigated in relation with insulin resistance, this is still not the case for other steroidal classes (e.g. androgens). Hopefully this study will raise awareness for this aspect, motivating further research in this direction.

In conclusion, this PhD project managed to close some open topics (data treatment in untargeted metabolomics) and provide further details for some complex or obscure topics (the interaction between cortisol and insulin and the role of steroidal metabolic pathways). Some of the findings of the present work should therefore stir future investigations towards more narrow areas of the intricate field of metabolomics. Moreover, it is worth noting how metabolomics is gaining more and more importance in clinical research given its ability to examine biological aspects that are hard to investigate through conventional techniques. Hopefully, this and other works will contribute to the adoption of metabolomics techniques in routine medical practice.

References

1. Munck, A. & Naray-Fejes-Toth, A. The ups and downs of glucocorticoid physiology. Permissive and suppressive effects revisited. *Mol Cell Endocrinol* **90**, C1-4 (1992).
2. Gathercole, L.L.*, et al.* Regulation of lipogenesis by glucocorticoids and insulin in human adipose tissue. *PLoS One* **6**, e26223 (2011).
3. Chimin, P.*, et al.* Chronic glucocorticoid treatment enhances lipogenic activity in visceral adipocytes of male Wistar rats. *Acta physiologica* **211**, 409-420 (2014).
4. Saad, M.J.A., Folli, F., Kahn, J.A. & Kahn, C.R. Modulation of Insulin-Receptor, Insulin-Receptor Substrate-1, and Phosphatidylinositol 3-Kinase in Liver and Muscle of Dexamethasone-Treated Rats. *J Clin Invest* **92**, 2065-2072 (1993).
5. Cooper, M.S., Seibel, M.J. & Zhou, H. Glucocorticoids, bone and energy metabolism. *Bone* (2015).
6. Pereira, C.D., Azevedo, I., Monteiro, R. & Martins, M.J. 11beta-Hydroxysteroid dehydrogenase type 1: relevance of its modulation in the pathophysiology of obesity, the metabolic syndrome and type 2 diabetes mellitus. *Diabetes, obesity & metabolism* **14**, 869-881 (2012).
7. Guaraldi, F. & Salvatori, R. Cushing syndrome: maybe not so uncommon of an endocrine disease. *Journal of the American Board of Family Medicine : JABFM* **25**, 199-208 (2012).
8. Johannsson, G., Skrtic, S., Lennernas, H., Quinkler, M. & Stewart, P.M. Improving outcomes in patients with adrenal insufficiency: a review of current and future treatments. *Current medical research and opinion* **30**, 1833-1847 (2014).
9. Wang, M. *Nutrition & metabolism* **2**, 3 (2005).
10. Cristancho, A.G. & Lazar, M.A. Forming functional fat: a growing understanding of adipocyte differentiation. *Nature reviews. Molecular cell biology* **12**, 722-734 (2011).
11. Mazziotti, G., Gazzaruso, C. & Giustina, A. Diabetes in Cushing syndrome: basic and clinical aspects. *Trends in endocrinology and metabolism: TEM* **22**, 499-506 (2011).
12. Brennan-Speranza, T.C.*, et al.* Osteoblasts mediate the adverse effects of glucocorticoids on fuel metabolism. *The Journal of clinical investigation* **122**, 4172-4189 (2012).
13. Swartz, S.L. & Dluhy, R.G. Corticosteroids: clinical pharmacology and therapeutic use. *Drugs* **16**, 238-255 (1978).
14. Grbovic, L. & Radenkovic, M. [Therapeutic use of glucocorticoids and immunosuppressive agents]. *Srp Arh Celok Lek* **133 Suppl 1**, 67-73 (2005).
15. Sadanala, K.C.L., Jeong-Ae ; Chung, Bong-Chul ; Choi, Man-Ho. Targeted Metabolite Profiling: Sample Preparation Techniques for GC-MSBased Steroid Analysis *Mass spectrometry letters* **3**, 4-9 (2012).
16. Azzouni, F., Godoy, A., Li, Y. & Mohler, J. The 5 alpha-reductase isozyme family: a review of basic biology and their role in human diseases. *Adv Urol* **2012**, 530121 (2012).
17. Swanepoel, A.C., Lindeque, B.G., Swart, P.J., Abdool, Z. & Pretorius, E. Estrogen causes ultrastructural changes of fibrin networks during the menstrual cycle: A qualitative investigation. *Microscopy Research and Technique* **77**, 594-601 (2014).
18. Albert, K., Pruessner, J. & Newhouse, P. Estradiol levels modulate brain activity and negative responses to psychosocial stress across the menstrual cycle. *Psychoneuroendocrinology* **59**, 14-24 (2015).
19. Winnick, J.J.*, et al.* Effects of 11beta-hydroxysteroid dehydrogenase-1 inhibition on hepatic glycogenolysis and gluconeogenesis. *Am J Physiol Endocrinol Metab* **304**, E747-756 (2013).
20. Palacios, R. & Sugawara, I. Hydrocortisone abrogates proliferation of T cells in autologous mixed lymphocyte reaction by rendering the interleukin-2 Producer T cells unresponsive to

interleukin-1 and unable to synthesize the T-cell growth factor. *Scandinavian journal of immunology* **15**, 25-31 (1982).
21. Hu, X. & Funder, J.W. The evolution of mineralocorticoid receptors. *Mol Endocrinol* **20**, 1471-1478 (2006).
22. Miller, W.L. & Auchus, R.J. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev* **32**, 81-151 (2011).
23. Häggström M, R.D. Diagram of the pathways of human steroidogenesis. *Wikiversity Journal of Medicine* **1**(2014).
24. http://pubchem.ncbi.nlm.nih.gov/.
25. Tomlinson, J.W. & Stewart, P.M. Modulation of glucocorticoid action and the treatment of type-2 diabetes. *Best Pract Res Cl En* **21**, 607-619 (2007).
26. Gathercole, L.L.*, et al.* 11beta-hydroxysteroid dehydrogenase 1: translational and therapeutic aspects. *Endocr Rev* **34**, 525-555 (2013).
27. Artaza, J.N.*, et al.* Endogenous expression and localization of myostatin and its relation to myosin heavy chain distribution in C2C12 skeletal muscle cells. *Journal of cellular physiology* **190**, 170-179 (2002).
28. Rhen, T. & Cidlowski, J.A. Antiinflammatory action of glucocorticoids--new mechanisms for old drugs. *N Engl J Med* **353**, 1711-1723 (2005).
29. Leung, D.Y.*, et al.* Disease management of atopic dermatitis: an updated practice parameter. Joint Task Force on Practice Parameters. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* **93**, S1-21 (2004).
30. Schacke, H., Docke, W.D. & Asadullah, K. Mechanisms involved in the side effects of glucocorticoids. *Pharmacology & therapeutics* **96**, 23-43 (2002).
31. McKay, L.I. & Cidlowski, J.A. Molecular control of immune/inflammatory responses: Interactions between nuclear factor-kappa B and steroid receptor-signaling pathways. *Endocr Rev* **20**, 435-459 (1999).
32. Coutinho, A.E. & Chapman, K.E. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Mol Cell Endocrinol* **335**, 2-13 (2011).
33. Leung, D.Y. & Bloom, J.W. Update on glucocorticoid action and resistance. *The Journal of allergy and clinical immunology* **111**, 3-22; quiz 23 (2003).
34. Das, I.*, et al.* Glucocorticoids alleviate intestinal ER stress by enhancing protein folding and degradation of misfolded proteins. *The Journal of experimental medicine* **210**, 1201-1216 (2013).
35. Tomlinson, J.W.*, et al.* 11beta-hydroxysteroid dehydrogenase type 1: a tissue-specific regulator of glucocorticoid response. *Endocr Rev* **25**, 831-866 (2004).
36. Nixon, M.*, et al.* Salicylate downregulates 11beta-HSD1 expression in adipose tissue in obese mice and in humans, mediating insulin sensitization. *Diabetes* **61**, 790-796 (2012).
37. Paulsen, S.K., Pedersen, S.B., Fisker, S. & Richelsen, B. 11Beta-HSD type 1 expression in human adipose tissue: impact of gender, obesity, and fat localization. *Obesity* **15**, 1954-1960 (2007).
38. Gomez-Sanchez, E.P.*, et al.* Regulation of 11β-hydroxysteroid dehydrogenase enzymes in the rat kidney by estradiol. *American Journal of Physiology - Endocrinology And Metabolism* **285**, E272-E279 (2003).
39. Cushing, H. The basophil adenomas of the pituitary body and their clinical manifestations (pituitary basophilism). 1932. *Obesity research* **2**, 486-508 (1994).
40. Long, C.N. & Lukens, F.D. The Effects of Adrenalectomy and Hypophysectomy Upon Experimental Diabetes in the Cat. *The Journal of experimental medicine* **63**, 465-490 (1936).
41. Long, C.N. The endocrine control of the blood-sugar. *Lancet* **1**, 325-329 (1952).

42. Pilkis, S.J., el-Maghrabi, M.R. & Claus, T.H. Hormonal regulation of hepatic gluconeogenesis and glycolysis. *Annu Rev Biochem* **57**, 755-783 (1988).
43. Rider, M.H*., et al.* 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: Head-to-head with a bifunctional enzyme that controls glycolysis. *Biochem J* **381**, 561-579 (2004).
44. Kuo, T., Harris, C.A. & Wang, J.C. Metabolic functions of glucocorticoid receptor in skeletal muscle. *Mol Cell Endocrinol* **380**, 79-88 (2013).
45. Exton, J.H*., et al.* The hormonal control of hepatic gluconeogenesis. *Recent Prog Horm Res* **26**, 411-461 (1970).
46. Kuo, T., McQueen, A., Chen, T.C. & Wang, J.C. Regulation of Glucose Homeostasis by Glucocorticoids. *Adv Exp Med Biol* **872**, 99-126 (2015).
47. Quesada, I., Tuduri, E., Ripoll, C. & Nadal, A. Physiology of the pancreatic alpha-cell and glucagon secretion: role in glucose homeostasis and diabetes. *J Endocrinol* **199**, 5-19 (2008).
48. Exton, J.H., Miller, T.B., Harper, S.C. & Park, C.R. Carbohydrate metabolism in perfused livers of adrenalectomized and steroid-replaced rats. *Am J Physiol* **230**, 163-170 (1976).
49. Stalmans, W. & Laloux, M. Glucocorticoids and hepatic glycogen metabolism. *Monogr Endocrinol* **12**, 517-533 (1979).
50. Huang, T.S. & Krebs, E.G. Amino acid sequence of a phosphorylation site in skeletal muscle glycogen synthetase. *Biochem Biophys Res Commun* **75**, 643-650 (1977).
51. Brady, M.J., Nairn, A.C. & Saltiel, A.R. The regulation of glycogen synthase by protein phosphatase 1 in 3T3-L1 adipocytes. Evidence for a potential role for DARPP-32 in insulin action. *J Biol Chem* **272**, 29698-29703 (1997).
52. Laloux, M., Stalmans, W. & Hers, H.G. On the mechanism by which glucocorticoids cause the activation of glycogen synthase in mouse and rat livers. *Eur J Biochem* **136**, 175-181 (1983).
53. Vanstapel, F., Dopere, F. & Stalmans, W. The role of glycogen synthase phosphatase in the glucocorticoid-induced deposition of glycogen in foetal rat liver. *Biochem J* **192**, 607-612 (1980).
54. Green, G.A., Chenoweth, M. & Dunn, A. Adrenal glucocorticoid permissive regulation of muscle glycogenolysis: action on protein phosphatase(s) and its inhibitor(s). *Proc Natl Acad Sci U S A* **77**, 5711-5715 (1980).
55. Ruzzin, J., Wagman, A.S. & Jensen, J. Glucocorticoid-induced insulin resistance in skeletal muscles: defects in insulin signalling and the effects of a selective glycogen synthase kinase-3 inhibitor. *Diabetologia* **48**, 2119-2130 (2005).
56. Miller, T.B., Exton, J.H. & Park, C.R. A block in epinephrine-induced glycogenolysis in hearts from adrenalectomized rats. *J Biol Chem* **246**, 3672-3678 (1971).
57. Puthanveetil, P. & Rodrigues, B. Glucocorticoid excess induces accumulation of cardiac glycogen and triglyceride: suggested role for AMPK. *Curr Pharm Des* **19**, 4818-4830 (2013).
58. Mead, J.R., Irvine, S.A. & Ramji, D.P. Lipoprotein lipase: structure, function, regulation, and role in disease. *J Mol Med (Berl)* **80**, 753-769 (2002).
59. Williams, K.J. & Fisher, E.A. Globular warming: how fat gets to the furnace. *Nat Med* **17**, 157-159 (2011).
60. de Guia, R.M. & Herzig, S. How Do Glucocorticoids Regulate Lipid Metabolism? *Adv Exp Med Biol* **872**, 127-144 (2015).
61. Choi, S.H. & Ginsberg, H.N. Increased very low density lipoprotein (VLDL) secretion, hepatic steatosis, and insulin resistance. *Trends Endocrinol Metab* **22**, 353-363 (2011).
62. Huff, M.W. Dietary cholesterol, cholesterol absorption, postprandial lipemia and atherosclerosis. *Can J Clin Pharmacol* **10 Suppl A**, 26A-32A (2003).
63. Desvergne, B., Michalik, L. & Wahli, W. Transcriptional regulation of metabolism. *Physiol Rev* **86**, 465-514 (2006).
64. Granner, D.K., Wang, J.C. & Yamamoto, K.R. Regulatory Actions of Glucocorticoid Hormones: From Organisms to Mechanisms. *Adv Exp Med Biol* **872**, 3-31 (2015).

65. Peckett, A.J., Wright, D.C. & Riddell, M.C. The effects of glucocorticoids on adipose tissue lipid metabolism. *Metabolism* **60**, 1500-1510 (2011).
66. Yu, C.Y.*, et al.* Genome-Wide Analysis of Glucocorticoid Receptor Binding Regions in Adipocytes Reveal Gene Network Involved in Triglyceride Homeostasis. *Plos One* **5**(2010).
67. Ebbert, J.O. & Jensen, M.D. Fat Depots, Free Fatty Acids, and Dyslipidemia. *Nutrients* **5**, 498-508 (2013).
68. Johnston, D.G., Gill, A., Orskov, H., Batstone, G.F. & Alberti, K.G. Metabolic effects of cortisol in man--studies with somatostatin. *Metabolism* **31**, 312-317 (1982).
69. Olswang, Y.*, et al.* Glucocorticoids repress transcription of phosphoenolpyruvate carboxykinase (GTP) gene in adipocytes by inhibiting its C/EBP-mediated activation. *J Biol Chem* **278**, 12929-12936 (2003).
70. Wang, Y.X.*, et al.* The human fatty acid synthase gene and de novo lipogenesis are coordinately regulated in human adipose tissue. *J Nutr* **134**, 1032-1038 (2004).
71. Hillgartner, F.B., Salati, L.M. & Goodridge, A.G. Physiological and molecular mechanisms involved in nutritional regulation of fatty acid synthesis. *Physiol Rev* **75**, 47-76 (1995).
72. Williams, B.H. & Berdanier, C.D. Effects of diet composition and adrenalectomy on the lipogenic responses of rats to starvation-refeeding. *J Nutr* **112**, 534-541 (1982).
73. Xu, C.*, et al.* Direct effect of glucocorticoids on lipolysis in adipocytes. *Mol Endocrinol* **23**, 1161-1170 (2009).
74. Lu, Z., Gu, Y. & Rooney, S.A. Transcriptional regulation of the lung fatty acid synthase gene by glucocorticoid, thyroid hormone and transforming growth factor-beta 1. *Biochim Biophys Acta* **1532**, 213-222 (2001).
75. Berdanier, C.D. Role of glucocorticoids in the regulation of lipogenesis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **3**, 2179-2183 (1989).
76. Qi, L.*, et al.* TRB3 links the E3 ubiquitin ligase COP1 to lipid metabolism. *Science* **312**, 1763-1766 (2006).
77. Sundaram, M. & Yao, Z. Recent progress in understanding protein and lipid factors affecting hepatic VLDL assembly and secretion. *Nutrition & metabolism* **7**, 35 (2010).
78. Martin-Sanz, P., Vance, J.E. & Brindley, D.N. Stimulation of apolipoprotein secretion in very-low-density and high-density lipoproteins from cultured rat hepatocytes by dexamethasone. *Biochem J* **271**, 575-583 (1990).
79. Wang, C.N., Hobman, T.C. & Brindley, D.N. Degradation of apolipoprotein B in cultured rat hepatocytes occurs in a post-endoplasmic reticulum compartment. *J Biol Chem* **270**, 24924-24931 (1995).
80. Taskinen, M.R., Nikkila, E.A., Pelkonen, R. & Sane, T. Plasma lipoproteins, lipolytic enzymes, and very low density lipoprotein triglyceride turnover in Cushing's syndrome. *J Clin Endocrinol Metab* **57**, 619-626 (1983).
81. Chan, D.C., Watts, G.F., Barrett, P.H., Mamo, J.C. & Redgrave, T.G. Markers of triglyceride-rich lipoprotein remnant metabolism in visceral obesity. *Clinical chemistry* **48**, 278-283 (2002).
82. Wang, X., Wei, D., Song, Z., Jiao, H. & Lin, H. Effects of fatty acid treatments on the dexamethasone-induced intramuscular lipid accumulation in chickens. *PLoS One* **7**, e36663 (2012).
83. Morgan, S.A., Gathercole, L.L., Bujalska, I.J., Stewart, P.M. & Tomlinson, J.W. Impact of Glucocorticoids upon Lipogenesis and beta-Oxidation in Skeletal Muscle. *Endocr Rev* **31**(2010).
84. Guillaume-Gentil, C., Assimacopoulos-Jeannet, F. & Jeanrenaud, B. Involvement of non-esterified fatty acid oxidation in glucocorticoid-induced peripheral insulin resistance in vivo in rats. *Diabetologia* **36**, 899-906 (1993).

85. Tomlinson, J.J., Boudreau, A., Wu, D., Atlas, E. & Hache, R.J. Modulation of early human preadipocyte differentiation by glucocorticoids. *Endocrinology* **147**, 5284-5293 (2006).
86. Ferrau, F. & Korbonits, M. Metabolic comorbidities in Cushing's syndrome. *Eur J Endocrinol* **173**, M133-157 (2015).
87. Hazlehurst, J.M*., et al.* Glucocorticoids fail to cause insulin resistance in human subcutaneous adipose tissue in vivo. *J Clin Endocrinol Metab* **98**, 1631-1640 (2013).
88. Jiang, Z.Y*., et al.* Insulin signaling through Akt/protein kinase B analyzed by small interfering RNA-mediated gene silencing. *P Natl Acad Sci USA* **100**, 7569-7574 (2003).
89. Giorgino, F., Almahfouz, A., Goodyear, L.J. & Smith, R.J. Glucocorticoid Regulation of Insulin-Receptor and Substrate Irs-1 Tyrosine Phosphorylation in Rat Skeletal-Muscle Invivo. *J Clin Invest* **91**, 2020-2030 (1993).
90. Long, W., Barrett, E.J., Wei, L.P. & Liu, Z.Q. Adrenalectomy enhances the insulin sensitivity of muscle protein synthesis. *Am J Physiol-Endoc M* **284**, E102-E109 (2003).
91. Hasselgren, P.O. & Fischer, J.E. Counter-regulatory hormones and mechanisms in amino acid metabolism with special reference to the catabolic response in skeletal muscle. *Current opinion in clinical nutrition and metabolic care* **2**, 9-14 (1999).
92. Tomas, F.M., Munro, H.N. & Young, V.R. Effect of glucocorticoid administration on the rate of muscle protein breakdown in vivo in rats, as measured by urinary excretion of N tau-methylhistidine. *Biochem J* **178**, 139-146 (1979).
93. Wing, S.S. & Goldberg, A.L. Glucocorticoids activate the ATP-ubiquitin-dependent proteolytic system in skeletal muscle during fasting. *Am J Physiol* **264**, E668-676 (1993).
94. Torrecilla, E*., et al.* Liver upregulation of genes involved in cortisol production and action is associated with metabolic syndrome in morbidly obese patients. *Obesity surgery* **22**, 478-486 (2012).
95. Konopelska, S*., et al.* Hepatic 11beta-HSD1 mRNA expression in fatty liver and nonalcoholic steatohepatitis. *Clin Endocrinol (Oxf)* **70**, 554-560 (2009).
96. Samuel, V.T*., et al.* Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J Biol Chem* **279**, 32345-32353 (2004).
97. Paterson, J.M*., et al.* Liver-selective transgene rescue of hypothalamic-pituitary-adrenal axis dysfunction in 11beta-hydroxysteroid dehydrogenase type 1-deficient mice. *Endocrinology* **148**, 961-966 (2007).
98. Valsamakis, G*., et al.* 11beta-hydroxysteroid dehydrogenase type 1 activity in lean and obese males with type 2 diabetes mellitus. *J Clin Endocrinol Metab* **89**, 4755-4761 (2004).
99. Goedecke, J.H*., et al.* Glucocorticoid metabolism within superficial subcutaneous rather than visceral adipose tissue is associated with features of the metabolic syndrome in South African women. *Clin Endocrinol (Oxf)* **65**, 81-87 (2006).
100. Veilleux, A*., et al.* Expression of genes related to glucocorticoid action in human subcutaneous and omental adipose tissue. *The Journal of steroid biochemistry and molecular biology* **122**, 28-34 (2010).
101. Fraser, R*., et al.* Cortisol effects on body mass, blood pressure, and cholesterol in the general population. *Hypertension* **33**, 1364-1368 (1999).
102. Dimitriadis, G*., et al.* Effects of glucocorticoid excess on the sensitivity of glucose transport and metabolism to insulin in rat skeletal muscle. *Biochem J* **321**, 707-712 (1997).
103. Woods, C. & Tomlinson, J.W. The Dehydrogenase Hypothesis. *Adv Exp Med Biol* **872**, 353-380 (2015).
104. Masuzaki, H*., et al.* A transgenic model of visceral obesity and the metabolic syndrome. *Science* **294**, 2166-2170 (2001).
105. Ferrannini, E*., et al.* The disposal of an oral glucose load in patients with non-insulin-dependent diabetes. *Metabolism* **37**, 79-85 (1988).

106. Dimitriadis, G.*, et al.* Effects of glucocorticoid excess on the sensitivity of glucose transport and metabolism to insulin in rat skeletal muscle. *Biochem J* **321 (Pt 3)**, 707-712 (1997).
107. Morgan, S.A.*, et al.* 11beta-hydroxysteroid dehydrogenase type 1 regulates glucocorticoid-induced insulin resistance in skeletal muscle. *Diabetes* **58**, 2506-2515 (2009).
108. Ohshima, K., Shargill, N.S., Chan, T.M. & Bray, G.A. Effects of dexamethasone on glucose transport by skeletal muscles of obese (ob/ob) mice. *Int J Obes* **13**, 155-163 (1989).
109. Lee, Y.H. & White, M.F. Insulin receptor substrate proteins and diabetes. *Arch Pharm Res* **27**, 361-370 (2004).
110. Kuo, T.*, et al.* Genome-wide analysis of glucocorticoid receptor-binding sites in myotubes identifies gene networks modulating insulin signaling. *Proc Natl Acad Sci U S A* **109**, 11160-11165 (2012).
111. Lambillotte, C., Gilon, P. & Henquin, J.C. Direct glucocorticoid inhibition of insulin secretion. An in vitro study of dexamethasone effects in mouse islets. *J Clin Invest* **99**, 414-423 (1997).
112. Hult, M.*, et al.* Short-term glucocorticoid treatment increases insulin secretion in islets derived from lean mice through multiple pathways and mechanisms. *Mol Cell Endocrinol* **301**, 109-116 (2009).
113. Delaunay, F.*, et al.* Pancreatic beta cells are important targets for the diabetogenic effects of glucocorticoids. *J Clin Invest* **100**, 2094-2098 (1997).
114. Ogawa, A.*, et al.* Roles of insulin resistance and beta-cell dysfunction in dexamethasone-induced diabetes. *J Clin Invest* **90**, 497-504 (1992).
115. Rafacho, A., Cestari, T.M., Taboga, S.R., Boschero, A.C. & Bosqueiro, J.R. High doses of dexamethasone induce increased beta-cell proliferation in pancreatic rat islets. *Am J Physiol Endocrinol Metab* **296**, E681-689 (2009).
116. van Raalte, D.H.*, et al.* Acute and 2-week exposure to prednisolone impair different aspects of beta-cell function in healthy men. *Eur J Endocrinol* **162**, 729-735 (2010).
117. Gremlich, S., Roduit, R. & Thorens, B. Dexamethasone induces posttranslational degradation of GLUT2 and inhibition of insulin secretion in isolated pancreatic beta cells. Comparison with the effects of fatty acids. *J Biol Chem* **272**, 3216-3222 (1997).
118. Borboni, P.*, et al.* Quantitative analysis of pancreatic glucokinase gene expression in cultured beta cells by competitive polymerase chain reaction. *Mol Cell Endocrinol* **117**, 175-181 (1996).
119. Goodman, P.A., Medina-Martinez, O. & Fernandez-Mejia, C. Identification of the human insulin negative regulatory element as a negative glucocorticoid response element. *Mol Cell Endocrinol* **120**, 139-146 (1996).
120. Sharma, S.*, et al.* Hormonal regulation of an islet-specific enhancer in the pancreatic homeobox gene STF-1. *Mol Cell Biol* **17**, 2598-2604 (1997).
121. Negrato, C.A.*, et al.* Association between different levels of dysglycemia and metabolic syndrome in pregnancy. *Diabetol Metab Syndr* **1**, 3 (2009).
122. Rafacho, A.*, et al.* Morphofunctional alterations in endocrine pancreas of short- and long-term dexamethasone-treated rats. *Horm Metab Res* **43**, 275-281 (2011).
123. Reich, E., Tamary, A., Sionov, R.V. & Melloul, D. Involvement of thioredoxin-interacting protein (TXNIP) in glucocorticoid-mediated beta cell death. *Diabetologia* **55**, 1048-1057 (2012).
124. Avram, D.*, et al.* IGF-1 protects against dexamethasone-induced cell death in insulin secreting INS-1 cells independent of AKT/PKB phosphorylation. *Cell Physiol Biochem* **21**, 455-462 (2008).
125. Roma, L.P.*, et al.* N-acetylcysteine protects pancreatic islet against glucocorticoid toxicity. *Redox Rep* **16**, 173-180 (2011).
126. Caperuto, L.C.*, et al.* Distinct regulation of IRS proteins in adipose tissue from obese aged and dexamethasone-treated rats. *Endocrine* **29**, 391-398 (2006).

127. Buren, J., Liu, H.X., Jensen, J. & Eriksson, J.W. Dexamethasone impairs insulin signalling and glucose transport by depletion of insulin receptor substrate-1, phosphatidylinositol 3-kinase and protein kinase B in primary cultured rat adipocytes. *Eur J Endocrinol* **146**, 419-429 (2002).
128. Sakoda, H.*, et al.* Dexamethasone-induced insulin resistance in 3T3-L1 adipocytes is due to inhibition of glucose transport rather than insulin signal transduction. *Diabetes* **49**, 1700-1708 (2000).
129. Holland, W.L.*, et al.* Inhibition of ceramide synthesis ameliorates glucocorticoid-, saturated-fat-, and obesity-induced insulin resistance. *Cell Metab* **5**, 167-179 (2007).
130. Chavez, J.A. & Summers, S.A. A ceramide-centric view of insulin resistance. *Cell Metab* **15**, 585-594 (2012).
131. Tappy, L.*, et al.* Mechanisms of dexamethasone-induced insulin resistance in healthy humans. *J Clin Endocrinol Metab* **79**, 1063-1069 (1994).
132. Bazuine, M., Carlotti, F., Tafrechi, R.S., Hoeben, R.C. & Maassen, J.A. Mitogen-activated protein kinase (MAPK) phosphatase-1 and -4 attenuate p38 MAPK during dexamethasone-induced insulin resistance in 3T3-L1 adipocytes. *Mol Endocrinol* **18**, 1697-1707 (2004).
133. de Leon, M.J.*, et al.* Cortisol reduces hippocampal glucose metabolism in normal elderly, but not in Alzheimer's disease. *J Clin Endocrinol Metab* **82**, 3251-3259 (1997).
134. Hansen, K.B., Vilsboll, T., Bagger, J.I., Holst, J.J. & Knop, F.K. Reduced glucose tolerance and insulin resistance induced by steroid treatment, relative physical inactivity, and high-calorie diet impairs the incretin effect in healthy subjects. *J Clin Endocrinol Metab* **95**, 3309-3317 (2010).
135. Rafacho, A.*, et al.* Pancreatic alpha-cell dysfunction contributes to the disruption of glucose homeostasis and compensatory insulin hypersecretion in glucocorticoid-treated rats. *PLoS One* **9**, e93531 (2014).
136. Nieman, L.K. & Ilias, I. Evaluation and treatment of Cushing's syndrome. *Am J Med* **118**, 1340-1346 (2005).
137. Shahani, S., Nudelman, R.J., Nalini, R., Kim, H.-S. & Samson, S.L. Ectopic corticotropin-releasing hormone (CRH) syndrome from metastatic small cell carcinoma: a case report and review of the literature. *Diagnostic Pathology* **5**, 56 (2010).
138. Raff, H. & Findling, J.W. A physiologic approach to diagnosis of the Cushing syndrome. *Annals of internal medicine* **138**, 980-991 (2003).
139. Giordano, R.*, et al.* Glucose metabolism in patients with subclinical Cushing's syndrome. *Endocrine* **41**, 415-423 (2012).
140. Di Dalmazi, G.*, et al.* Progressively increased patterns of subclinical cortisol hypersecretion in adrenal incidentalomas differently predict major metabolic and cardiovascular outcomes: a large cross-sectional study. *Eur J Endocrinol* **166**, 669-677 (2012).
141. Kim, B.Y.*, et al.* Clinical Characteristics and Metabolic Features of Patients with Adrenal Incidentalomas with or without Subclinical Cushing's Syndrome. *Endocrinol Metab (Seoul)* **29**, 457-463 (2014).
142. Androulakis, II*, et al.* Patients with apparently nonfunctioning adrenal incidentalomas may be at increased cardiovascular risk due to excessive cortisol secretion. *J Clin Endocrinol Metab* **99**, 2754-2762 (2014).
143. Mancini, T., Kola, B., Mantero, F., Boscaro, M. & Arnaldi, G. High cardiovascular risk in patients with Cushing's syndrome according to 1999 WHO/ISH guidelines. *Clin Endocrinol (Oxf)* **61**, 768-777 (2004).
144. Jain, V.*, et al.* Neonatal presentation of familial glucocorticoid deficiency resulting from a novel splice mutation in the melanocortin 2 receptor accessory protein. *Eur J Endocrinol* **165**, 987-991 (2011).

145. Storr, H.L., Chan, L.F., Grossman, A.B. & Savage, M.O. Paediatric Cushing's syndrome: epidemiology, investigation and therapeutic advances. *Trends Endocrinol Metab* **18**, 167-174 (2007).
146. Fiehn, O. Metabolomics--the link between genotypes and phenotypes. *Plant molecular biology* **48**, 155-171 (2002).
147. Brown, M.*, et al.* Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. *Bioinformatics* **27**, 1108-1112 (2011).
148. Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends in biotechnology* **16**, 373-378 (1998).
149. Wishart, D.S.*, et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic acids research* **41**, D801-807 (2013).
150. http://www.ebi.ac.uk/chebi/.
151. http://www.hmdb.ca/.
152. http://www.genome.jp/kegg/.
153. http://www.lipidmaps.org/.
154. Nicholson, J.K., Lindon, J.C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica; the fate of foreign compounds in biological systems* **29**, 1181-1189 (1999).
155. Brooks, C.J., Horning, E.C. & Young, J.S. Characterization of sterols by gas chromatography-mass spectrometry of the trimethylsilyl ethers. *Lipids* **3**, 391-402 (1968).
156. Pauling, L., Robinson, A.B., Teranishi, R. & Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Natl Acad Sci U S A* **68**, 2374-2376 (1971).
157. Fiehn, O.*, et al.* Metabolite profiling for plant functional genomics. *Nature biotechnology* **18**, 1157-1161 (2000).
158. Rabinowitz, J.D. & Silhavy, T.J. Systems biology: metabolite turns master regulator. *Nature* **500**, 283-284 (2013).
159. Guengerich, F.P. Principles of covalent binding of reactive metabolites and examples of activation of bis-electrophiles by conjugation. *Arch Biochem Biophys* **433**, 369-378 (2005).
160. Bouskila, M.*, et al.* Allosteric regulation of glycogen synthase controls glycogen synthesis in muscle. *Cell Metab* **12**, 456-466 (2010).
161. Ciccarelli, M.*, et al.* Glucose-induced expression of the homeotic transcription factor Prep1 is associated with histone post-translational modifications in skeletal muscle. *Diabetologia* **59**, 176-186 (2015).
162. Kuo, M.-H. & Allis, C.D. Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20**, 615-626 (1998).
163. Kell, D. Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology* **7**, 296-307 (2004).
164. Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R. & Griffin, J.L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society reviews* **40**, 387-426 (2011).
165. McNelis, J.C.*, et al.* Dehydroepiandrosterone exerts antiglucocorticoid action on human preadipocyte proliferation, differentiation, and glucose uptake. *Am J Physiol Endocrinol Metab* **305**, E1134-1144 (2013).
166. Swali, A., Walker, E.A., Lavery, G.G., Tomlinson, J.W. & Stewart, P.M. 11beta-Hydroxysteroid dehydrogenase type 1 regulates insulin and glucagon secretion in pancreatic islets. *Diabetologia* **51**, 2003-2011 (2008).
167. Vegiopoulos, A. & Herzig, S. Glucocorticoids, metabolism and metabolic diseases. *Molecular and cellular endocrinology* **275**, 43-61 (2007).

168. Zhang, A.H.*, et al.* Ultraperformance liquid chromatography-mass spectrometry based comprehensive metabolomics combined with pattern recognition and network analysis methods for characterization of metabolites and metabolic pathways from biological data sets. *Anal Chem* **85**, 7606-7612 (2013).
169. Chen, T.*, et al.* Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Molecular & cellular proteomics : MCP* **10**, M110 004945 (2011).
170. Qi, Y.*, et al.* Metabolomics Study of Resina Draconis on Myocardial Ischemia Rats Using Ultraperformance Liquid Chromatography/Quadrupole Time-of-Flight Mass Spectrometry Combined with Pattern Recognition Methods and Metabolic Pathway Analysis. *Evidence-based complementary and alternative medicine : eCAM* **2013**, 438680 (2013).
171. Wang, X.*, et al.* Urine metabolomics analysis for biomarker discovery and detection of jaundice syndrome in patients with liver disease. *Molecular & cellular proteomics : MCP* **11**, 370-380 (2012).
172. Bertini, I.*, et al.* The metabonomic signature of celiac disease. *Journal of proteome research* **8**, 170-177 (2009).
173. Carrola, J.*, et al.* Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. *Journal of proteome research* **10**, 221-230 (2011).
174. McClay, J.L.*, et al.* (1)H nuclear magnetic resonance metabolomics analysis identifies novel urinary biomarkers for lung function. *Journal of proteome research* **9**, 3083-3090 (2010).
175. Al-Ismaili, Z., Palijan, A. & Zappitelli, M. Biomarkers of acute kidney injury in children: discovery, evaluation, and clinical application. *Pediatric nephrology* **26**, 29-40 (2011).
176. Kim, B.*, et al.* Global metabolomics and targeted steroid profiling reveal that rifampin, a strong human PXR activator, alters endogenous urinary steroid markers. *Journal of proteome research* **12**, 1359-1368 (2013).
177. Kiss, A.*, et al.* Doping control using high and ultra-high resolution mass spectrometry based non-targeted metabolomics-a case study of salbutamol and budesonide abuse. *PLoS One* **8**, e74584 (2013).
178. van den Berg, R.*, et al.* A single night of sleep curtailment increases plasma acylcarnitines: Novel insights in the relationship between sleep and insulin resistance. *Arch Biochem Biophys* **589**, 145-151 (2016).
179. Lehmann, R.*, et al.* Circulating Lysophosphatidylcholines Are Markers of a Metabolically Benign Nonalcoholic Fatty Liver. *Diabetes Care* **36**, 2331-2338 (2013).
180. Reinke, S.N.*, et al.* 1H NMR Derived Metabolomic Profile of Neonatal Asphyxia in Umbilical Cord Serum: Implications for Hypoxic Ischemic Encephalopathy. *Journal of proteome research* **12**, 4230-4239 (2013).
181. Brown, M.*, et al.* A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **1**, 39-51 (2005).
182. Zelena, E.*, et al.* Development of a Robust and Repeatable UPLC−MS Method for the Long-Term Metabolomic Study of Human Serum. *Anal Chem* **81**, 1357-1364 (2009).
183. Dunn, W.B.*, et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols* **6**, 1060-1083 (2011).
184. Dunn, W.B., Wilson, I.D., Nicholls, A.W. & Broadhurst, D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **4**, 2249-2264 (2012).
185. Dunn, W.B., Bailey, N.J. & Johnson, H.E. Measuring the metabolome: current analytical technologies. *The Analyst* **130**, 606-625 (2005).
186. William Allwood, J.*, et al.* Considerations in Sample Preparation, Collection, and Extraction Approaches Applied in Microbial, Plant, and Mammalian Metabolic Profiling. 79-118 (2012).

187. Gika, H.G., Theodoridis, G.A. & Wilson, I.D. Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *Journal of separation science* **31**, 1598-1608 (2008).
188. Taylor, N.F. Urinary steroid profiling. *Methods in molecular biology* **324**, 159-175 (2006).
189. R:, A.J.W.W.C.L.D.W.B.G. Considerations in sample preparation, collection, and extraction approaches applied in miccrobial, plant, and mammalian metabolic profiling. in *Methodologies for metabolomics* (ed. Lutz;Sweedler;Wevers) (Cambridge, 2013).
190. Chen, S.*, et al.* Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry. *J Chromatogr A* **1298**, 9-16 (2013).
191. Sellick, C.A.*, et al.* Effective quenching processes for physiologically valid metabolite profiling of suspension cultured Mammalian cells. *Anal Chem* **81**, 174-183 (2009).
192. Teng, Q., Huang, W.L., Collette, T.W., Ekman, D.R. & Tan, C. A direct cell quenching method for cell-culture based metabolomics. *Metabolomics* **5**, 199-208 (2009).
193. Yuan, J., Bennett, B.D. & Rabinowitz, J.D. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature protocols* **3**, 1328-1340 (2008).
194. Winder, C.L.*, et al.* Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. *Anal Chem* **80**, 2939-2948 (2008).
195. Theobald, U., Mailinger, W., Reuss, M. & Rizzi, M. In vivo analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Analytical biochemistry* **214**, 31-37 (1993).
196. Bradford, B.U.*, et al.* Metabolomic profiling of a modified alcohol liquid diet model for liver injury in the mouse uncovers new markers of disease. *Toxicol Appl Pharmacol* **232**, 236-243 (2008).
197. Denkert, C.*, et al.* Metabolite profiling of human colon carcinoma--deregulation of TCA cycle and amino acid turnover. *Mol Cancer* **7**, 72 (2008).
198. Metabolomics: Current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry* **24**, 285-294 (2005).
199. Dunn, W.B., Overy, S. & Quick, W.P. Evaluation of automated electrospray-TOF mass spectrometryfor metabolic fingerprinting of the plant metabolome. *Metabolomics* **1**, 137-148 (2005).
200. Goodacre, R., Vaidyanathan, S., Bianchi, G. & Kell, D.B. Metabolic profiling using direct infusion electrospray ionisation mass spectrometry for the characterisation of olive oils. *The Analyst* **127**, 1457-1462 (2002).
201. Vaidyanathan, S., O'Hagan, S. & Goodacre, R. Direct infusion electrospray ionization mass spectra of crude cell extracts for microbial characterizations: influence of solvent conditions on the detection of proteins. *Rapid communications in mass spectrometry : RCM* **20**, 21-30 (2006).
202. Southam, A.D., Payne, T.G., Cooper, H.J., Arvanitis, T.N. & Viant, M.R. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal Chem* **79**, 4595-4602 (2007).
203. Tauler, R., Gorrochategui, E., Jaumot, J. & Tauler, R. A protocol for LC-MS metabolomic data processing using chemometric tools. *Protocol Exchange* (2015).
204. Martano, G.*, et al.* Fast sampling method for mammalian cell metabolic analyses using liquid chromatography–mass spectrometry. *Nature protocols* **10**, 1-11 (2014).
205. Dunn, W.B.*, et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols* **6**, 1060-1083 (2011).

206. Dorsey, J.G. & Dill, K.A. The molecular mechanism of retention in reversed-phase liquid chromatography. *Chemical Reviews* **89**, 331-346 (1989).
207. Ogiso, H., Suzuki, T. & Taguchi, R. Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Analytical biochemistry* **375**, 124-131 (2008).
208. Spagou, K.*, et al.* HILIC-UPLC-MS for Exploratory Urinary Metabolic Profiling in Toxicological Studies. *Anal Chem* **83**, 382-390 (2011).
209. Forcisi, S.*, et al.* Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **1292**, 51-65 (2013).
210. Boutegrabet, L.*, et al.* Attachment of chloride anion to sugars: mechanistic investigation and discovery of a new dopant for efficient sugar ionization/detection in mass spectrometers. *Chemistry* **18**, 13059-13067 (2012).
211. Oss, M., Kruve, A., Herodes, K. & Leito, I. Electrospray ionization efficiency scale of organic compounds. *Anal Chem* **82**, 2865-2872 (2010).
212. Jorge, T.F.*, et al.* Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Mass Spectrometry Reviews*, n/a-n/a (2015).
213. Hsu, J.-Y.*, et al.* Urinary exposure marker discovery for toxicants using ultra-high pressure liquid chromatography coupled with Orbitrap high resolution mass spectrometry and three untargeted metabolomics approaches. *Anal Chim Acta* **939**, 73-83 (2016).
214. Mueller, D.C., Piller, M., Niessner, R., Scherer, M. & Scherer, G. Untargeted Metabolomic Profiling in Saliva of Smokers and Nonsmokers by a Validated GC-TOF-MS Method. *Journal of proteome research* **13**, 1602-1613 (2014).
215. Raro, M.*, et al.* Untargeted Metabolomics in Doping Control: Detection of New Markers of Testosterone Misuse by Ultrahigh Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. *Anal Chem* **87**, 8373-8380 (2015).
216. Michalski, A.*, et al.* Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Molecular & Cellular Proteomics* **10**, M111.011015-M011111.011015 (2011).
217. Zubarev, R.A. & Makarov, A. Orbitrap Mass Spectrometry. *Anal Chem* **85**, 5288-5296 (2013).
218. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534-2536 (2008).
219. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787 (2006).
220. http://www.nonlinear.com/progenesis/qi/.
221. https://mycompounddiscoverer.com/.
222. Veselkov, K.A.*, et al.* Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem* **83**, 5864-5872 (2011).
223. Gromski, P.S.*, et al.* Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites* **4**, 433-452 (2014).
224. Saccenti, E., Hoefsloot, H.C.J., Smilde, A.K., Westerhuis, J.A. & Hendriks, M.M.W.B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**(2013).
225. Di Guida, R.*, et al.* Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics* **12**, 93 (2016).
226. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Current Metabolomics* **1**, 92-107 (2013).

227. Hendriks, M.M.W.B.*, et al.* Data-processing strategies for metabolomics studies. *Trac-Trend Anal Chem* **30**, 1685-1698 (2011).
228. Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. & Kell, D.B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in biotechnology* **22**, 245-252 (2004).
229. Sumner, L.W.*, et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211-221 (2007).
230. Dervilly-Pinel, G.*, et al.* Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves. *Anal Chim Acta* **700**, 144-154 (2011).
231. Galuska, C.E.*, et al.* Profiling intact steroid sulfates and unconjugated steroids in biological fluids by liquid chromatography-tandem mass spectrometry (LC-MS-MS). *The Analyst* **138**, 3792-3801 (2013).
232. Rijk, J.C.*, et al.* Metabolomics approach to anabolic steroid urine profiling of bovines treated with prohormones. *Anal Chem* **81**, 6879-6888 (2009).
233. Kenny, L.C.*, et al.* Detection and identification of novel metabolomic biomarkers in preeclampsia. *Reproductive sciences* **15**, 591-597 (2008).
234. Dunn, W.B.*, et al.* Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **871**, 288-298 (2008).
235. Little, R.J.A. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* **83**, 1198-1202 (1998).
236. Xia, J. & Wishart, D.S. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nature protocols* **6**, 743-760 (2011).
237. Hrydziuszko, O. & Viant, M.R. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics* **8**, S161-S174 (2012).
238. Nyamundanda, G., Brennan, L. & Gormley, I.C. Probabilistic principal component analysis for metabolomic data. *BMC bioinformatics* **11**, 571 (2010).
239. Breiman, L. Random Forests. *Machine Learning* **45**, 5-32 (2001).
240. Trevor Hastie, R.T., Balasubramanian Narasimhan and Gilbert Chu. impute: impute: Imputation for microarray data. R package version 1.36.0.
241. Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods--a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164-1167 (2007).
242. http://www.bioconductor.org/.
243. https://cran.r-project.org/.
244. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* **45**, 1-67 (2011).
245. Sangster, T.P., Wingate, J.E., Burton, L., Teichert, F. & Wilson, I.D. Investigation of analytical variation in metabonomic analysis using liquid chromatography/mass spectrometry. *Rapid Commun Mass Sp* **21**, 2965-2970 (2007).
246. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **78**, 4281-4290 (2006).
247. van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K. & van der Werf, M.J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006).
248. Parsons, H.M., Ludwig, C., Gunther, U.L. & Viant, M.R. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC bioinformatics* **8**, 234 (2007).

249. Mak, T.D., Laiakis, E.C., Goudarzi, M. & Fornace, A.J., Jr. MetaboLyzer: A Novel Statistical Workflow for Analyzing Postprocessed LC-MS Metabolomics Data. *Anal Chem* **86**, 506-513 (2014).
250. Jackson, J. Wiley series in probability and mathematical statistics. Applied probability and statistics. A user's guide to principal components. *John Wiley & Sons, Inc.* (1991).
251. Eriksson, L.J., E; Kettaneh-Wold, N; Wold, S. Scaling. Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS). *Umetrics*, 213-225 (1999).
252. Smilde, A.K., van der Werf, M.J., Bijlsma, S., van der Werff-van der Vat, B.J.C. & Jellema, R.H. Fusion of Mass Spectrometry-Based Metabolomics Data. *Anal Chem* **77**, 6729-6736 (2005).
253. Keun, H.C*., et al.* Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Anal Chim Acta* **490**, 265-276 (2003).
254. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).
255. Mullard, G*., et al.* A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications. *Metabolomics* **11**, 1068-1080 (2014).
256. Stekhoven, D.J. & Buhlmann, P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112-118 (2011).
257. Wolstencroft, K*., et al.* The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic acids research* **41**, W557-561 (2013).
258. Salek, R.M., Steinbeck, C., Viant, M.R., Goodacre, R. & Dunn, W.B. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience* **2**(2013).
259. Weber, R.J.M., Southam, A.D., Sommer, U. & Viant, M.R. Characterization of Isotopic Abundance Measurements in High Resolution FT-ICR and Orbitrap Mass Spectra for Improved Confidence of Metabolite Identification. *Anal Chem* **83**, 3737-3743 (2011).
260. Weber, R.J.M. & Viant, M.R. MI-Pack: Increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways. *Chemometr Intell Lab* **104**, 75-82 (2010).
261. Team, R.C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2015).
262. Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**, 51-78 (2007).
263. Deutskens, F., Yang, J. & Caprioli, R.M. High spatial resolution imaging mass spectrometry and classical histology on a single tissue section. *Journal of mass spectrometry : JMS* **46**, 568-571 (2011).
264. Chen, S*., et al.* Pseudotargeted Metabolomics Method and Its Application in Serum Biomarker Discovery for Hepatocellular Carcinoma Based on Ultra High-Performance Liquid Chromatography/Triple Quadrupole Mass Spectrometry. *Anal Chem* **85**, 8326-8333 (2013).
265. Lommen, A. & Kools, H.J. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics* **8**, 719-726 (2012).
266. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics* **11**, 395 (2010).
267. Styczynski, M.P*., et al.* Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Anal Chem* **79**, 966-973 (2007).
268. Coble, J.B. & Fraga, C.G. Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *J Chromatogr A* **1358**, 155-164 (2014).

269. Libiseller, G.*, et al.* IPO: a tool for automated optimization of XCMS parameters. *BMC bioinformatics* **16**(2015).
270. Eliasson, M.*, et al.* Strategy for optimizing LC-MS data processing in metabolomics: a design of experiments approach. *Anal Chem* **84**, 6869-6876 (2012).
271. De Livera, A.M.*, et al.* Normalizing and Integrating Metabolomics Data. *Anal Chem* **84**, 10768-10776 (2012).
272. Guerra, A.*, et al.* Effects of urine dilution on quantity, size and aggregation of calcium oxalate crystals induced in vitro by an oxalate load. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **43**, 585-589 (2005).
273. Lusczek, E.R., Nelson T, Lexcen D, Witowski NE, Mulier KE. Urine Metabolomics in Hemorrhagic Shock: Normalization of Urine in the Face of Changing Intravascular Fluid Volume and Perturbations in Metabolism. *Bioanalysis & Biomedicine* **3**, 038-048 (2011).
274. Chetwynd, A.J., Abdul-Sada, A., Holt, S.G. & Hill, E.M. Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses. *J Chromatogr A* **1431**, 103-110 (2016).
275. Waikar, S.S., Sabbisetti, V.S. & Bonventre, J.V. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney International* **78**, 486-494 (2010).
276. Rubin, D.B. Inference and Missing Data. *Biometrika* **63**, 581-590 (1976).
277. Smilde, A.K., van der Werf, M.J., Bijlsma, S., van der Werff-van der Vat, B.J. & Jellema, R.H. Fusion of mass spectrometry-based metabolomics data. *Anal Chem* **77**, 6729-6736 (2005).
278. Yang, J., Zhao, X., Lu, X., Lin, X. & Xu, G. A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. *Frontiers in molecular biosciences* **2**, 4 (2015).
279. Beale, E.M.L. & Little, R.J.A. Missing Values in Multivariate-Analysis. *J Roy Stat Soc B Met* **37**, 129-145 (1975).
280. Karpievitch, Y.V., Dabney, A.R. & Smith, R.D. Normalization and missing value imputation for label-free LC-MS analysis. *BMC bioinformatics* **13 Suppl 16**, S5 (2012).
281. Webb-Robertson, B.J.*, et al.* Sequential projection pursuit principal component analysis--dealing with missing data associated with new -omics technologies. *BioTechniques* **54**, 165-168 (2013).
282. Husson, F. & Josse, J. Handling missing values in multiple factor analysis. *Food Qual Prefer* **30**, 77-85 (2013).
283. Xia, J., Psychogios, N., Young, N. & Wishart, D.S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research* **37**, W652-660 (2009).
284. Steuer, R., Morgenthal, K., Weckwerth, W. & Selbig, J. A gentle guide to the analysis of metabolomic data. *Methods in molecular biology* **358**, 105-126 (2007).
285. Troyanskaya, O.*, et al.* Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520-525 (2001).
286. Oba, S.*, et al.* A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics* **19**, 2088-2096 (2003).
287. Buuren, S.v. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations inR. *Journal of Statistical Software* **45**(2011).
288. Scheel, I.*, et al.* The influence of missing value imputation on detection of differentially expressed genes from microarray data. *Bioinformatics* **21**, 4272-4279 (2005).
289. Pedreschi, R.*, et al.* Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. *Proteomics* **8**, 1371-1383 (2008).
290. Kim, G.R.*, et al.* Combined mass spectrometry-based metabolite profiling of different pigmented rice (Oryza sativa L.) seeds and correlation with antioxidant activities. *Molecules* **19**, 15673-15686 (2014).

291. Ho, W.E*., et al.* Metabolomics Reveals Inflammatory-Linked Pulmonary Metabolic Alterations in a Murine Model of House Dust Mite-Induced Allergic Asthma. *Journal of proteome research* (2014).
292. Madala, N.E., Piater, L.A., Steenkamp, P.A. & Dubery, I.A. Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited Nicotiana tabacum and Sorghum bicolor cells. *Springerplus* **3**, 254 (2014).
293. Rothwell, J.A*., et al.* New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PLoS One* **9**, e93474 (2014).
294. Baniasadi, H., Vlahakis, C., Hazebroek, J., Zhong, C. & Asiago, V. Effect of environment and genotype on commercial maize hybrids using LC/MS-based metabolomics. *J Agric Food Chem* **62**, 1412-1422 (2014).
295. Masson, P., Spagou, K., Nicholson, J.K. & Want, E.J. Technical and biological variation in UPLC-MS-based untargeted metabolic profiling of liver extracts: application in an experimental toxicity study on galactosamine. *Anal Chem* **83**, 1116-1123 (2011).
296. Huan, T. & Li, L. Counting missing values in a metabolite-intensity data set for measuring the analytical performance of a metabolomics platform. *Anal Chem* **87**, 1306-1313 (2015).
297. Filzmoser, P. & Walczak, B. What can go wrong at the data normalization step for identification of biomarkers? *Journal of chromatography. A* **1362**, 194-205 (2014).
298. Kohl, S.M*., et al.* State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* **8**, 146-160 (2012).
299. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86 (2010).
300. Blankenberg, D*., et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* **Chapter 19**, Unit 19 10 11-21 (2010).
301. Giardine, B*., et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**, 1451-1455 (2005).
302. https://usegalaxy.org/.
303. Sheynkman, G.M*., et al.* Using Galaxy-P to leverage RNA-Seq for the discovery of novel protein variations. *BMC Genomics* **15**, 703 (2014).
304. Giacomoni, F*., et al.* Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* **31**, 1493-1495 (2015).
305. Davidson, R.L., Weber, R.J.M., Liu, H., Sharma-Oates, A. & Viant, M.R. Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. *GigaScience* **5**(2016).
306. https://github.com/Viant-Metabolomics/Galaxy-M.
307. Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T.R. & Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* **84**, 283-289 (2012).
308. Wold, H.O. Operative Aspects of Econometric and Sociological Models Current Developments of Fp (Fix-Point) Estimation and Nipals (Nonlinear Iterative Partial Least Squares) Modelling. *Econ Appl* **26**, 385-421 (1973).
309. Schilling, T.M*., et al.* Intranasal insulin increases regional cerebral blood flow in the insular cortex in men independently of cortisol manipulation. *Human brain mapping* **35**, 1944-1956 (2014).
310. Griffin, E.E. & Wildenthal, K. Regulation of cardiac protein balance by hydrocortisone: interaction with insulin. *Am J Physiol* **234**, E306-313 (1978).

311. Cole, M.A., Kim, P.J., Kalman, B.A. & Spencer, R.L. Dexamethasone suppression of corticosteroid secretion: evaluation of the site of action by receptor measures and functional studies. *Psychoneuroendocrinology* **25**, 151-167 (2000).
312. Gathercole, L.L., Morgan, S.A., Bujalska, I.J., Stewart, P.M. & Tomlinson, J.W. Short- and long-term glucocorticoid treatment enhances insulin signalling in human subcutaneous adipose tissue. *Nutrition & diabetes* **1**, e3 (2011).
313. Vienberg, S.G. & Bjornholm, M. Chronic glucocorticoid treatment increases de novo lipogenesis in visceral adipose tissue. *Acta physiologica* **211**, 257-259 (2014).
314. Holland, W.L*., et al.* Lipid mediators of insulin resistance. *Nutrition reviews* **65**, S39-46 (2007).
315. Aronoff, S.L.B., K. Glucose Metabolism and Regulation: Beyond Insulin and Glucagon. *Diabetes Spectrum* **17**(2004).
316. Dowman, J.K*., et al.* Loss of 5alpha-reductase type 1 accelerates the development of hepatic steatosis but protects against hepatocellular carcinoma in male mice. *Endocrinology* **154**, 4536-4547 (2013).
317. Milewich, L*., et al.* Progesterone and 5alpha-pregnane-3,20-dione in peripheral blood of normal young women: Daily measurements throughout the menstrual cycle. *J Clin Endocrinol Metab* **45**, 617-622 (1977).
318. Weinstein, B.I*., et al.* 5 alpha-dihydrocortisol in human aqueous humor and metabolism of cortisol by human lenses in vitro. *Investigative ophthalmology & visual science* **32**, 2130-2135 (1991).
319. Kenyon, C.J*., et al.* Antinatriuretic and kaliuretic activities of the reduced derivatives of aldosterone. *Endocrinology* **112**, 1852-1856 (1983).
320. Mostaghel, E.A. Beyond T and DHT - Novel Steroid Derivatives Capable of Wild Type Androgen Receptor Activation. *International Journal of Biological Sciences* **10**, 602-613 (2014).
321. Heemers, H.V. & Tindall, D.J. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev* **28**, 778-808 (2007).
322. Pandini, G*., et al.* Androgens up-regulate the insulin-like growth factor-I receptor in prostate cancer cells. *Cancer research* **65**, 1849-1857 (2005).
323. Vlahopoulos, S*., et al.* Recruitment of the androgen receptor via serum response factor facilitates expression of a myogenic gene. *J Biol Chem* **280**, 7786-7792 (2005).
324. Wang, K., Fan, D.D., Jin, S., Xing, N.Z. & Niu, Y.N. Differential expression of 5-alpha reductase isozymes in the prostate and its clinical implications. *Asian journal of andrology* **16**, 274-279 (2014).
325. Wilson, E.M. & French, F.S. Binding properties of androgen receptors. Evidence for identical receptors in rat testis, epididymis, and prostate. *J Biol Chem* **251**, 5620-5629 (1976).
326. Saartok, T., Dahlberg, E. & Gustafsson, J.A. Relative binding affinity of anabolic-androgenic steroids: comparison of the binding to the androgen receptors in skeletal muscle and in prostate, as well as to sex hormone-binding globulin. *Endocrinology* **114**, 2100-2106 (1984).
327. Normington, K. & Russell, D.W. Tissue distribution and kinetic characteristics of rat steroid 5 alpha-reductase isozymes. Evidence for distinct physiological functions. *J Biol Chem* **267**, 19548-19554 (1992).
328. Thigpen, A.E*., et al.* Tissue distribution and ontogeny of steroid 5 alpha-reductase isozyme expression. *J Clin Invest* **92**, 903-910 (1993).
329. Uemura, M*., et al.* Novel 5 alpha-steroid reductase (SRD5A3, type-3) is overexpressed in hormone-refractory prostate cancer. *Cancer Sci* **99**, 81-86 (2008).
330. Godoy, A*., et al.* 5alpha-reductase type 3 expression in human benign and malignant tissues: a comparative analysis during prostate cancer progression. *Prostate* **71**, 1033-1046 (2011).
331. Titus, M.A*., et al.* 5 alpha-Reductase Type 3 Enzyme in Benign and Malignant Prostate. *Prostate* **74**, 235-249 (2014).

332. Kang, H.J., Imperato-McGinley, J., Zhu, Y.S. & Rosenwaks, Z. The effect of 5 alpha-reductase-2 deficiency on human fertility. *Fertil Steril* **101**, 310-316 (2014).
333. Olsson, M.*, et al.* Correlation between circulatory, local prostatic, and intra-prostatic androgen levels. *Prostate* **71**, 909-914 (2011).
334. Kashiwagi, B.*, et al.* Changes in testosterone and dihydrotestosterone levels in male rat accessory sex organs, serum, and seminal fluid after castration: establishment of a new highly sensitive simultaneous androgen measurement method. *Journal of andrology* **26**, 586-591 (2005).
335. McConnell, J.D.*, et al.* The effect of finasteride on the risk of acute urinary retention and the need for surgical treatment among men with benign prostatic hyperplasia. Finasteride Long-Term Efficacy and Safety Study Group. *N Engl J Med* **338**, 557-563 (1998).
336. Andriole, G.L. & Kirby, R. Safety and tolerability of the dual 5 alpha-reductase inhibitor dutasteride in the treatment of benign prostatic hyperplasia. *Eur Urol* **44**, 82-88 (2003).
337. Frye, S.V. Discovery and clinical development of dutasteride, a potent dual 5alpha-reductase inhibitor. *Current topics in medicinal chemistry* **6**, 405-421 (2006).
338. Clark, R.V.*, et al.* Marked suppression of dihydrotestosterone in men with benign prostatic hyperplasia by dutasteride, a dual 5alpha-reductase inhibitor. *J Clin Endocrinol Metab* **89**, 2179-2184 (2004).
339. Yamana, K., Labrie, F. & Luu-The, V. Human type 3 5alpha-reductase is expressed in peripheral tissues at higher levels than types 1 and 2 and its activity is potently inhibited by finasteride and dutasteride. *Horm Mol Biol Clin Investig* **2**, 293-299 (2010).
340. http://www.drugbank.ca/.
341. Thompson, I.M.*, et al.* The influence of finasteride on the development of prostate cancer. *N Engl J Med* **349**, 215-224 (2003).
342. Traish, A.M., Hassani, J., Guay, A.T., Zitzmann, M. & Hansen, M.L. Adverse Side Effects of 5α-Reductase Inhibitors Therapy: Persistent Diminished Libido and Erectile Dysfunction and Depression in a Subset of Patients. *The Journal of Sexual Medicine* **8**, 872-884 (2011).
343. Kaplan, S.A., Chung, D.E., Lee, R.K., Scofield, S. & Te, A.E. A 5-year retrospective analysis of 5alpha-reductase inhibitors in men with benign prostatic hyperplasia: finasteride has comparable urinary symptom efficacy and prostate volume reduction, but less sexual side effects and breast complications than dutasteride. *International journal of clinical practice* **66**, 1052-1055 (2012).
344. Upreti, R.*, et al.* 5alpha-reductase type 1 modulates insulin sensitivity in men. *J Clin Endocrinol Metab* **99**, E1397-1406 (2014).
345. Lundahl, A., Lennernas, H., Knutson, L., Bondesson, U. & Hedeland, M. Identification of finasteride metabolites in human bile and urine by high-performance liquid chromatography/tandem mass spectrometry. *Drug metabolism and disposition: the biological fate of chemicals* **37**, 2008-2017 (2009).
346. Miller, J. & Tarter, T.H. Update on the use of dutasteride in the management of benign prostatic hypertrophy. *Clin Interv Aging* **2**, 99-104 (2007).
347. Hazlehurst, J.M.*, et al.* Dual-5alpha-Reductase Inhibition Promotes Hepatic Lipid Accumulation in Man. *J Clin Endocrinol Metab* **101**, 103-113 (2016).
348. Gehmlich, K.*, et al.* Changes in the cardiac metabolome caused by perhexiline treatment in a mouse model of hypertrophic cardiomyopathy. *Molecular bioSystems* **11**, 564-573 (2015).