

# **TECHNISCHE UNIVERSITÄT MÜNCHEN**

Lehrstuhl für analytische Lebensmittelchemie

**New prognostic and diagnostic biomarkers in diabetic nephropathy  
using comprehensive targeted and non-targeted metabolomics**

**Kilian Wörmann**

Vollständiger Abdruck der von der Fakultät für Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Naturwissenschaften**

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. E. Grill

Prüfer der Dissertation: 1. apl. Prof. Dr. Ph. Schmitt-Kopplin

2. Univ.-Prof. Dr. M. Rychlik

3. apl. Prof. Dr. R. Lehmann

(Eberhard-Karls-Universität Tübingen)

Dissertation wurde am 31.07.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 20.01.2015 angenommen.

Für meine Familie

## **Acknowledgements**

This thesis would not have been possible without the inspiration, support and encouragement of a number of wonderful individuals – my thanks and appreciation to all of them for being part of this journey and making this thesis possible.

I owe my deepest gratitude to my supervisor Prof. Dr. Philippe Schmitt-Kopplin for his guidance, encouragement and introduction into many fascinating research topics.

A special thanks to Dr. Agnese Fekete for initial aid in the first year.

I would like to thank all my colleagues first of all for the familiar atmosphere they provided and also for the support be it scientific know-how, discussions, experimental help or proof-reading.

Furthermore I am grateful to Prof. Dr. Rainer Lehmann for the scientific discussions, sample providing and his trust.

Finally, and most importantly I would like to express my gratitude to my wife Nazaré. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past years of my life have been built. I thank my parents and my sisters who gave me the strength to go my way leading to this thesis.

The dissertation has been prepared in the period of 01.02.2009 to 31.07.2012 at the research unit Analytical BioGeoChemistry at the Helmholtz Zentrum München – German Research Centre for Environmental Health in the frame of the Competence Network for Diabetes mellitus funded by the German Federal Ministry of Education and Research (BMBF, FKZ 01GI0804).

# Table of Contents

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	DIABETES MELLITUS.....	1
1.1.1	<i>Type 2 Diabetes</i> .....	2
1.1.2	<i>Diabetic nephropathy</i> .....	5
1.2	METABOLOMICS.....	15
1.2.1	<i>Metabolic network and its regulation mechanisms</i> .....	15
1.2.2	<i>Metabolomics as tool for biomarker identification</i> .....	17
1.3	OBJECTION AND OVERVIEW OF THE THESIS.....	22
<b>2</b>	<b>ANALYTICAL METHODS .....</b>	<b>25</b>
2.1	INSTRUMENTAL TECHNIQUES.....	25
2.1.1	<i>Liquid Chromatography</i> .....	25
2.1.2	<i>Mass spectrometry</i> .....	27
2.1.3	<i>Nuclear magnetic resonance spectrometry</i> .....	36
2.2	DATA ANALYSIS APPROACHES .....	38
2.2.1	<i>Unsupervised Analysis</i> .....	38
2.2.2	<i>Supervised analysis</i> .....	39
2.2.3	<i>Network calculations</i> .....	40
2.2.4	<i>Van Krevelen plot</i> .....	41
<b>3</b>	<b>IDENTIFICATION OF NEW PROGNOSTIC BIOMARKERS WITHIN THE PRE-DIABETES STUDY.....</b>	<b>43</b>
3.1	INTRODUCTION.....	43
3.2	OBJECTIVE.....	44
3.3	MATERIAL AND METHODS.....	46
3.3.1	<i>Material</i> .....	46
3.3.2	<i>Methods and methods development</i> .....	46
3.4	RESULT AND DISCUSSION .....	67
3.4.1	<i>First survey of the data</i> .....	68
3.4.2	<i>Metabolic changes through microalbuminuria</i> .....	72
3.4.3	<i>Unknown compounds</i> .....	90
3.5	CONCLUSION .....	92
<b>4</b>	<b>DISCOVERY OF NEW DIABETIC NEPHROPATHY BIOMARKERS WITHIN THE PREDICTIONS STUDY .....</b>	<b>93</b>
4.1	INTRODUCTION.....	93
4.2	OBJECTIVE.....	94
4.3	MATERIAL AND METHODS.....	94
4.3.1	<i>Comparison between instrument generations</i> .....	94

4.3.2	<i>Methods and method development</i> .....	97
4.4	RESULTS AND DISCUSSION .....	107
4.4.1	<i>First survey of the data</i> .....	107
4.4.2	<i>Metabolic alterations in patients with diabetic nephropathy</i> .....	110
4.4.3	<i>Unknown compounds</i> .....	116
4.5	ION MOBILITY ANALYSIS .....	119
4.6	CONCLUSION .....	123
<b>5</b>	<b>CONCLUDING REMARKS</b> .....	<b>124</b>
<b>6</b>	<b>APPENDIX</b> .....	<b>125</b>
<b>7</b>	<b>REFERENCES</b> .....	<b>152</b>
<b>8</b>	<b>SUMMARY</b> .....	<b>167</b>

# Abbreviations

ACN	Acetonitrile
AGE	Advanced glycation end-products
CID	Collision-induced dissociation
CKD	Chronic kidney disease
CV-ANOVA	cross-validation analysis of variance
DN	Diabetic nephropathy
ESI	Electro spray ionization
FA	Formic acid
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
GFR	Glomerular filtration rate
HCA	Hierarchical cluster analysis
HCE	Hierarchical cluster explorer
HILIC	Hydrophilic interaction liquid chromatography
IMS	Ion mobility separation
LC	Liquid chromatography
logD	Distribution coefficient
MA	Microalbuminuria
MeOH	Methanol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass to charge
NMR	Nuclear magnetic resonance spectroscopy
OPLS-DA	Orthogonal partial least square discriminant analysis
PC	Principle component
PCA	Principle component analysis
OPLS-DA	Partial least square discriminant analysis
PPE	Protein precipitation
QC	Quality control
RF	Radio frequency
ROS	Reactive oxygen species
RP	Reversed phase chromatography
RT	Retention time

SPE	solid phase extraction
T1D	Typ 1 diabetes <i>mellitus</i>
T2D	Typ 2 diabetes <i>mellitus</i>
TMAO	Trimethylamine-N-oxide
ToF	Time of flight mass spectrometer
TSP	Trimethylsilyl propanoic acid
TULIP	Tuebinger lifestyle intervention program
UHPLC	ultra-high performance liquid chromatography
WHO	World Health Organization

## List of figures

Figure 1: Estimated increase of adults aged 20-79 years with diabetes from 2010 to 2030.....	2
Figure 2: Pathophysiology of hyperglycaemia and increased circulating fatty acids in type 2 diabetes .....	3
Figure 3: Unspecific symptoms of type 2 diabetes <i>mellitus</i> .....	4
Figure 4: Diagrammatic representation of a human kidney .....	6
Figure 5: Anatomy of the nephrons .....	7
Figure 6: Different factors influencing the pathogenesis of diabetic nephropathy.....	12
Figure 7: Extrarenal and intrarenal factors involved in diabetic nephropathy.....	13
Figure 8: Complexity of the interaction between metabolic pathways.....	16
Figure 9: Connection between the different "omics" research fields and their information content. ....	19
Figure 10: Interesting biological samples for metabolomics analysis and a typical non-targeted workflow .....	21
Figure 11: Milestones in the process of biomarker discovery with non-targeted metabolomics .....	23
Figure 12: Thesis structure.....	24
Figure 13: Dependency of the theoretical plate height $H$ from the particle size $d_p$ and the velocity $v$ .....	26
Figure 14: Schematic representation of the positive mode electrospray ionisation process.....	29
Figure 15: Schematic construction and functionality of a travelling wave ion mobility cell .....	30
Figure 16: A schematic diagram of the Synapt® HDMS system .....	32
Figure 17: Schematic setting for the Bruker Apex Qe 70 FTMS .....	34
Figure 18: Generation of a mass spectrum through FT-ICR-MS instruments.....	35
Figure 19: Energetic splitting of nuclei spin states in an external magnetic field and Larmor frequency .....	37
Figure 20: PCA and HCA learning principles.....	39
Figure 21: Van Krevelen diagram of human urine FT-ICR-MS data and its expressive power.....	42



Figure 22: The four used analytical platforms to reach the main goal analysing the pre-diabetes samples .....	45
Figure 23: Overview of the applied multi-platform approach for biomarker detection	47
Figure 24: Comparison of SPE with centrifuged urine .....	49
Figure 25: Comparison of the detected signals in urine with the different tested columns .....	51
Figure 26: Development of the chromatographic method by means of detected signals .....	53
Figure 27: Schematic workflow of the data treatment and analysis procedure .....	58
Figure 28: HCE metabolic profile search .....	61
Figure 29: Filtration of the UHPLC-MS annotations by plotting the logD values against the RT. ....	63
Figure 30: Determination of the mass difference window for the UHPLC data calibration .....	65
Figure 31: Comparison of the peak shapes of three correlating peaks in ESI <sup>-</sup> data .	66
Figure 32: Overview of the result and discussion section .....	67
Figure 33: HCA of the UPLC-MS data from the TULIP study including QC's. ....	68
Figure 34: Overlay of the 13 quality control chromatograms for each detection mode .....	69
Figure 35: Van Krevelen diagram of the TULIP study FT-ICR-MS data .....	71
Figure 36: Venn diagram of the mass convergence between FT-ICR- and UHPLC-MS-Data .....	72
Figure 37: Separation of control and microalbuminuric samples with OPLS-DA after winsorisation .....	73
Figure 38: Analysis of the MS annotations within the TULIP study applying metabolic enrichment and pathway analysis.....	74
Figure 39: Box plot of creatinine in the sample urines .....	75
Figure 40: Alterations of the creatinine pathway in patients with microalbuminuria ..	76
Figure 41: Boxplot of quinic and hippuric acid in LC-MS negativ mode .....	81
Figure 42: Figure 40: Boxplot of glutamine and N $\alpha$ -acetyl-L-glutamine in LC-MS negativ mode .....	82
Figure 43: Boxplot of citric and aconitic- or isocitric acid in LC-MS negativ mode ....	84
Figure 44: Phospholipids as essential part of cell membranes .....	86
Figure 45: Boxplot of 3-hydroxycapric acid in LC-MS positive mode .....	87

Figure 46: Boxplot of ferulic acid and suberic acid identified in LC-MS ESI <sup>-</sup> mode ...	87
Figure 47: Boxplot of identified jasmonic acid in ESI <sup>+</sup> and homoveratric acid in ESI <sup>-</sup> mode.....	88
Figure 48: Boxplot of 5 $\alpha$ -dihydrotestosterone sulphate and androsterone sulphate in ESI <sup>-</sup> mode.....	89
Figure 49: Kendrick-analogous mass difference network analysis of FT-ICR-MS data .....	90
Figure 50: Comparison between the Bruker APEX Qeand the solariX <sup>TM</sup> FT-ICR-MS system .....	95
Figure 51: Comparison between the Synapt <sup>TM</sup> HDMS <sup>TM</sup> and the Synapt <sup>TM</sup> G2S HDMS <sup>TM</sup> system. ....	96
Figure 52: Comparison between different SPEs and PPE in FT-ICR-MS sample preparation .....	102
Figure 53: Reduction of false positive annotations in LC-MS by RT - logD filtering	105
Figure 54: HCA of the ESI <sup>-</sup> LC-MS data before averaging of the PREDICTIONS study .....	107
Figure 55: Overlay of the quality control runs within the LC-MS analysis .....	108
Figure 56: Van Krevelen diagram of the FT-ICR-MS data from PREDICTIONS study samples .....	109
Figure 57: Venn diagramm of the mass comparison between the platforms .....	110
Figure 58: OPLS-DA models of control and nephropathic samples.....	111
Figure 59: Metabolic enrichment and pathway analysis of significant MS annotations within the PREDICTIONS study .....	112
Figure 60: Boxplot of citric acid in LC-MS and FT-ICR-MS ESI <sup>-</sup> mod.....	114
Figure 61: Kendrick-analogous mass difference network of the PREDICTIONS study FT-ICR-MS data .....	117
Figure 62: Ion mobility drift time chromatogram in ESI <sup>+</sup> mode .....	120
Figure 63: IMS separation of two signals of the ESI <sup>+</sup> LC-MS dataset identified as significant.....	121
Figure 64: Mapping of metabolite classes by direct injection of a standard mixture with IMS.....	122

## 1 Introduction

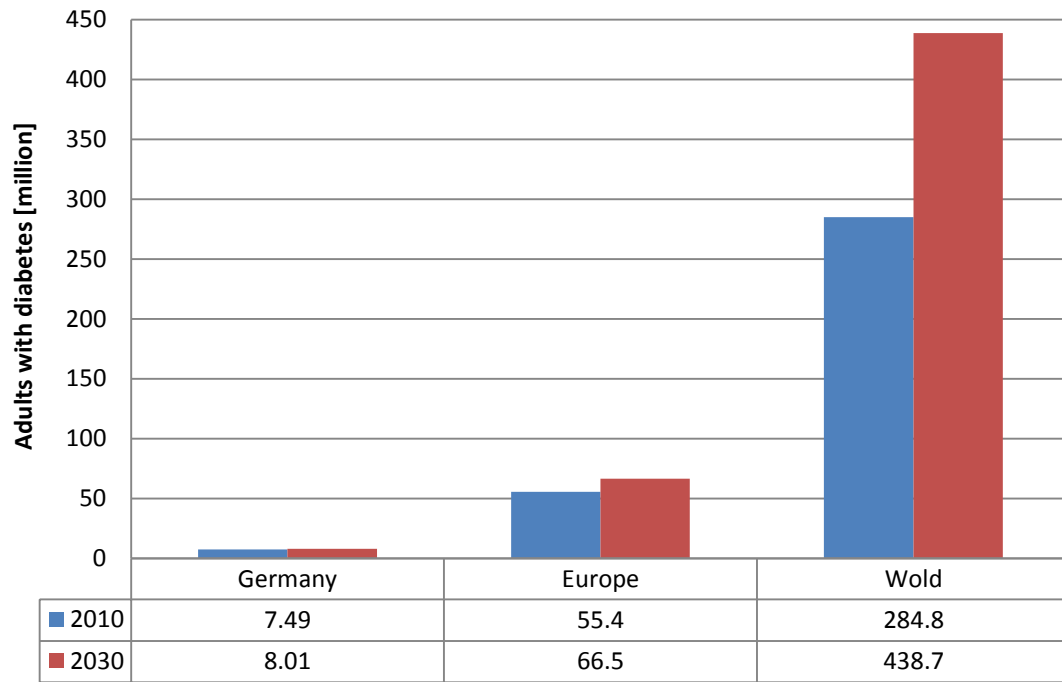
The name *Diabetes mellitus* (honey-sweet discharge; diabetes: old Greek, diabainein, διαβαίνειν, to pass through; *mellitus*: Latin, honey-sweet) was used by the British surgeon John Rollo (1749-1809) to distinguish it from *diabetes insipidus* (*insipidus*: Latin, 'without taste') [Zajac *et al.*, 2010]. It refers to the increased urination with raised amounts of sugar in the urine of patients.

The World Health Organisation (WHO) defines diabetes mellitus as a “metabolic disorder of multiple aetiology characterised by chronic hyperglycaemia [high blood sugar level] with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin action, or both. The effects of diabetes mellitus include long-term damage, dysfunction and failure of various organs” [WHO, 1999].

### 1.1 Diabetes mellitus

Already the ancient Egyptians used the sweet taste of urine as diagnostic tool. The term *diabetes* was first recognised to be used after 30 AD by Aretus of Cappadocia. Thomas Willis described as first European in the 17<sup>th</sup> century the sweet taste of urine from diabetic patients and claimed that the sugar resulted from an increase of the concentration in blood. It took another century until Matthew Dobson could prove that the blood sugar level of diabetic patients is increased. In the 19<sup>th</sup> century Apollinaire Bouchardet invented the first self-care with a copper reagent and in the beginning of the 20<sup>th</sup> century the first test-paper methods appeared [Guthrie *et al.*, 1988]. Until today quick and easy stick tests for detection of glucose in urine are available.

*Diabetes mellitus* (thereafter referred to as diabetes) is at present one of the most prevalent chronic diseases worldwide. In 2010 it was estimated that 285 million people in the age of 20-79 years were suffering from diabetes and this number is likely to increase by 50 % until 2030 (Figure 1) [Shaw *et al.*, 2010]. In Germany, the percentage of people with diabetes is expected to rise from 12 % to 13.5 % for the same period [Shaw *et al.*, 2010].



**Figure 1: Estimated increase of adults aged 20-79 years with diabetes from 2010 to 2030**

Shaw et. al analysed 133 studies from 91 countries to predict the number of adults with diabetes for the years 2010 and 2030 [Shaw *et al.*, 2010].

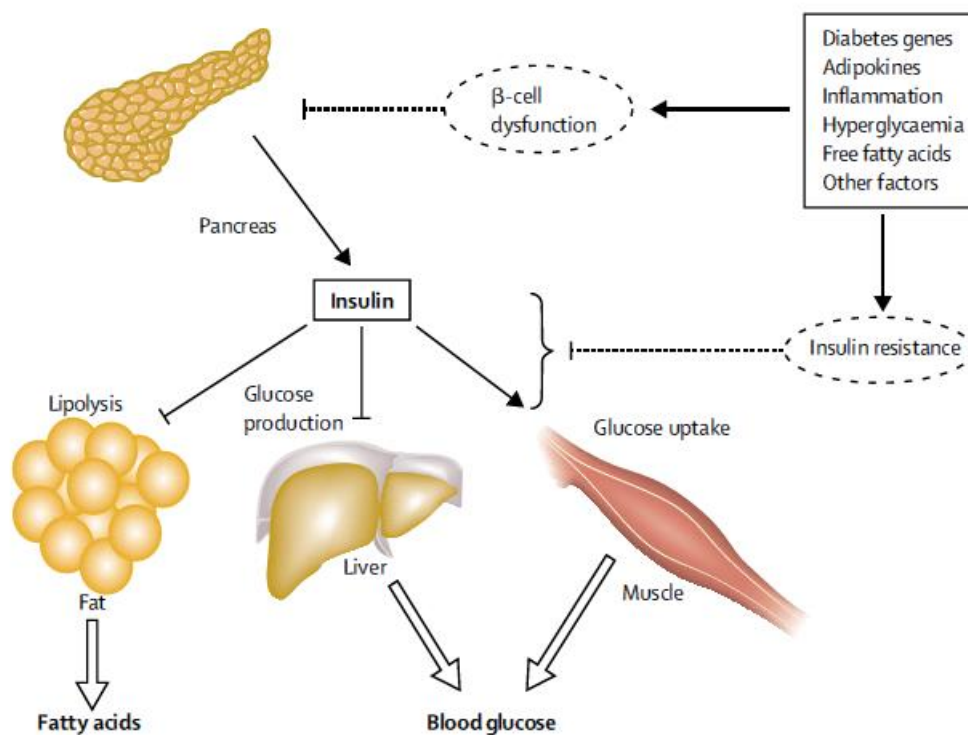
Diabetes *mellitus* is a general term for several metabolic diseases, which are characterised by hyperglycaemia. Possible causes are defects in insulin secretion, a decreased effect of insulin or both together [American Diabetes Association, 2010]. The classification of the type of diabetes is often a problem since the symptoms depending on the individual circumstances at the time of diagnosis and don't fit easily into a single class. At present four types of diabetes are differentiated. Type 2 diabetes *mellitus* (T2D) is with 90-95 % by far the most common, followed by type 1 diabetes, gestational diabetes and other specific types [American Diabetes Association, 2010].

### 1.1.1 Type 2 Diabetes

Contrary to type 1 diabetes no autoimmune destruction of the insulin producing  $\beta$ -cells takes place. T2D patients suffer from an insulin resistance, an inadequate tissue response to insulin or a  $\beta$ -cell dysfunction which lowers the insulin secretion [Rutter *et al.*, 2008]. Thus they have a relative insulin deficiency.

Within healthy people, insulin is secreted from the  $\beta$ -cells in the pancreas gland and is the most important hormone for blood glucose level control. An increased blood sugar level after a meal causes the  $\beta$ -cells to secrete insulin. It enhances the glucose uptake of muscle cells, liver and fat tissue. In the liver the storage of glucose as glycogen (gluconeogenesis) is stimulated. Insulin also reduce the fatty acid release of fat tissue and boosts their synthesis. These effects cause the blood sugar level to decrease to the regular level.

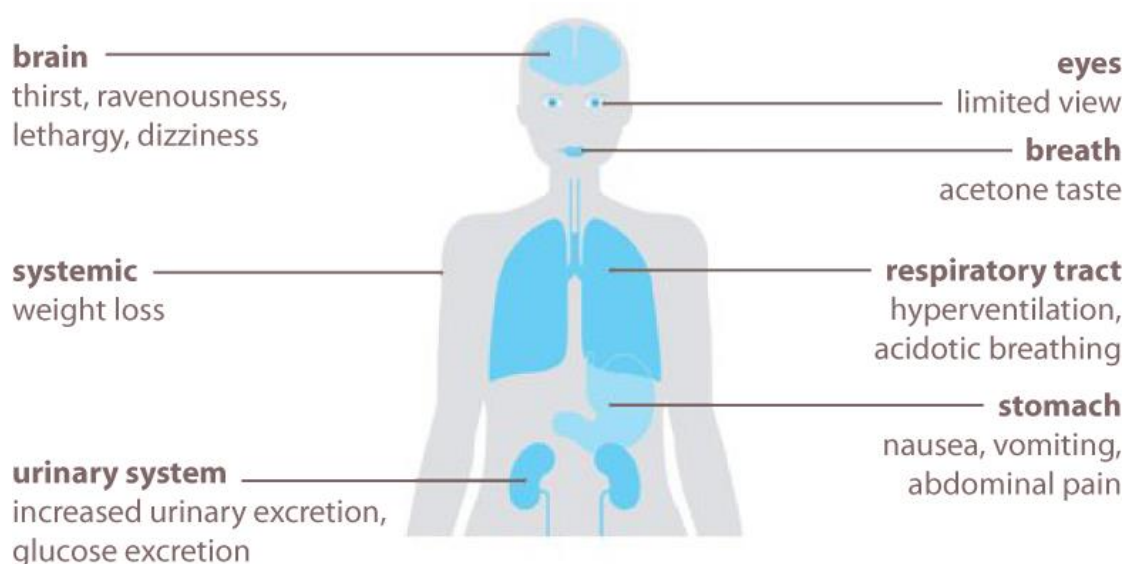
A decreased insulin secretion of the pancreas, an inadequate tissue response or the insulin resistance of the cells reduces the glucose uptake of muscles, fat tissue and liver (Figure 2).



**Figure 2: Pathophysiology of hyperglycaemia and increased circulating fatty acids in type 2 diabetes**

Insulin secretion from the pancreas normally reduces glucose output by the liver, enhances glucose uptake by skeletal muscle, and suppresses fatty acid release from fat tissue. The various factors shown that contribute to the pathogenesis of type 2 diabetes affect both insulin secretion and insulin action. Decreased insulin secretion will reduce insulin signalling in its target tissues. Insulin resistance pathways affect the action of insulin in each of the major target tissues, leading to increased circulating fatty acids and the hyperglycaemia of diabetes. In turn, the raised concentrations of glucose and fatty acids in the bloodstream will feed back to worsen both insulin secretion and insulin resistance (Reprinted from [Stumvoll *et al.*, 2005], with permission from Elsevier).

In some cases the body attempts to counterbalance insulin resistance with increased  $\beta$ -cell function and mass, hyperinsulinemia, but the blood sugar and free fatty acids levels still remain slightly elevated compared to healthy individuals [Rutter *et al.*, 2008; Stumvoll *et al.*, 2005]. The body can endure this stage for some time, which leads to a gradually pathogenesis of T2D in different stages and hamper an early diagnosis of the disease. Nevertheless hyperglycaemia and the increased free fatty acids levels are toxic over long time exposure and cause  $\beta$ -cell dysfunction,  $\beta$ -cell apoptosis, insulin gene expression inhibition and insulin gene transcription alterations first accelerating the process and finally leading to onset of frank diabetes [Rutter *et al.*, 2008]. Even then frequently for years T2D patients show no specific symptoms (Figure 3) and the disease is often diagnosed accidentally.



**Figure 3: Unspecific symptoms of type 2 diabetes mellitus**  
(Altered with permission from Helmholtz Centre Munich)

The risk factors for T2D are diverse (Figure 2). Besides age, physical inactivity, dietary influences, hypertension and obesity also the genetic background has a great impact on the pathogenesis. Family and twin studies indicate the importance of family history of diabetes in the risk assessment [Stumvoll *et al.*, 2005]. A. Doria *et al.* reviewed nicely the results of different genetic approaches. In addition to different monogenic causes for  $\beta$ -cell dysfunction and insulin resistance they also outlined some important single nucleotide polymorphisms [Doria *et al.*, 2008]. The genetically

heterogeneous character together with the other risk factors of T2D also explains the various unspecific symptoms and the individual aetiopathology, the course of this disease.

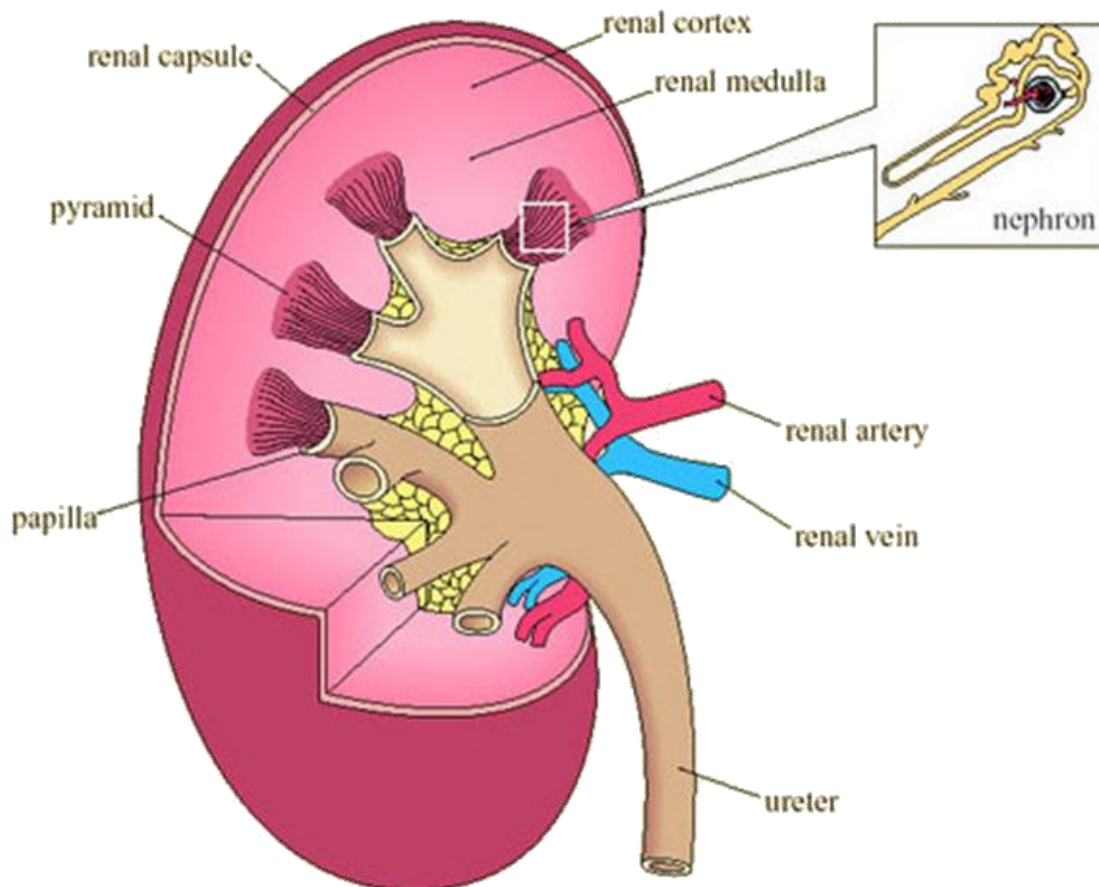
For the diagnosis and monitoring of the disease progression a number of well-established biomarkers and methods are available. Besides the determination of glucose in blood or urine the evaluation of glycated haemoglobin, of albumin in urine or other methods are used. However simple and reliable techniques for the assessment of the risk to develop diabetes or the diagnosis of the pre-diabetic state are longed for since early lifestyle interventions can prevent or at least delay the onset of T2D and its late complications. The latter can be divided into macrovascular (disease of the large blood vessels) and microvascular (small blood vessel disease) complications. Angina, myocardial infarction, stroke, peripheral artery disease and congestive heart failure are among the macrovascular and microvasculare complications are retinopathy, neuropathy, small vessel vasculopathy and diabetic nephropathy (DN) [van Dieren *et al.*, 2010].

### **1.1.2 Diabetic nephropathy**

DN is a chronic kidney disease resulting from long term diabetes and is a growing global problem. It has become the most common cause for renal replacement therapies and accounts for up to 50 % of the patients in developed countries [Rutter *et al.*, 2008]. Among dialysis patients, diabetics have the highest hospitalisation rates and mortality [Winchester *et al.*, 2010].

#### **1.1.2.1 The human kidney**

Humans have two bean shaped kidneys (Figure 4) with a weight between 120–200 g [Thews *et al.*, 1999]. They can be divided into two regions; the renal cortex is the pale fine-granulated outer part and the darker renal medulla with a fine striation [Thews *et al.*, 1999]. The medulla forms renal pyramids with papillas to direct the produced urine into the ureter which leads to the urinary bladder.



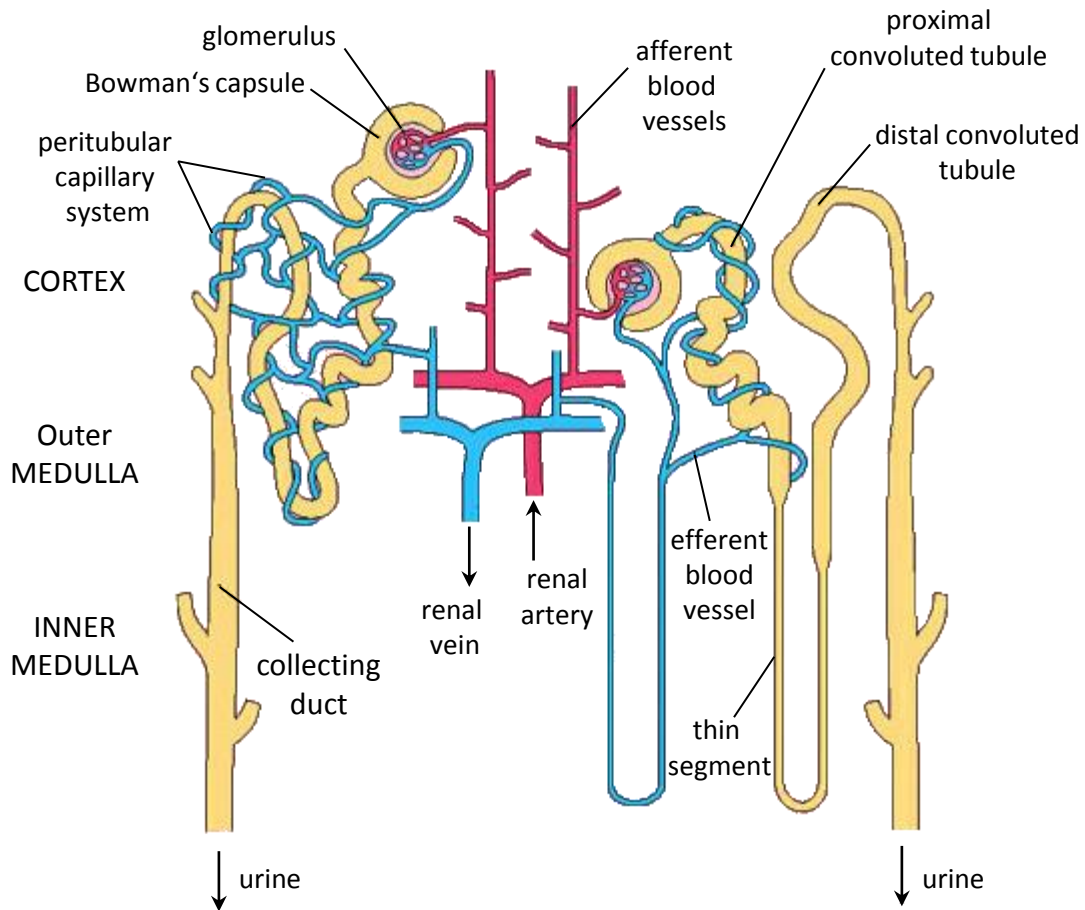
**Figure 4: Diagrammatic representation of a human kidney**

(<http://openlearn.open.ac.uk/mod/oucontent/view.php?id=398612&section=3.2>, 23.06.2012)

Several important functions are accomplished by the kidneys. They are responsible for the filtration of the blood, thereby cleaning the organism from different metabolic end products, drugs and toxins by secretion into the urine. They also maintain the equilibrium of extracellular fluid volume, the osmotic concentration and the ion balance through regulated water and electrolyte secretion [Thews *et al.*, 1999]. Together with the lungs they are responsible for the acid-base homeostasis by excretion of bicarbonates ( $\text{HCO}_3^-$ ) or protons ( $\text{H}^+$ ) into the urine. Kidneys produce hormones for regulation of different internal and external functions too. Erythropoietin is mainly released from the kidneys and promotes the production of red blood cells whereas the active form of vitamin D, calcitriol, stimulates the renal phosphate reabsorption and the intestinal absorption of calcium [Thews *et al.*, 1999]. They also regulate the arterial blood pressure through the reabsorption rate of sodium ions ( $\text{Na}^+$ ) and production of the hormone renin.



Each kidney contains about 1.2 million functional filtration units, the nephrons (Figure 5).



**Figure 5: Anatomy of the nephrons**

Two typical nephrons. The one on the right is 'long-looped' (located close to the border of the medulla and therefore called juxtamedullary) and the one on the left is 'short-looped' (or cortical). The long-looped nephron is paralleled by a loop formed by the blood capillary. A capillary network surrounds the short-looped nephron. (<http://openlearn.open.ac.uk/mod/oucontent/view.php?id=398612&section=3.2>, 23.06.2012)

Compared to other organs kidneys are well supplied with blood and nearly 25 % of the cardiac output volume is directed through them to effectively remove wastes from the blood and the organism [Thews *et al.*, 1999]. For a proper kidney function the blood filtration rate inside the glomeruli has to be constant. Thus, the blood pressure is regulated by automatically adaption of flow resistance in the afferent and efferent arteriole [Thews *et al.*, 1999].

Each nephron consists of the glomerulus, the Bowman's capsule which surrounds the glomerulus and a tubule system. The incoming blood reaches the glomerulus

through the afferent vessel. In it the blood is filtered through a capillary network before it leaves *via* the efferent vessel. The resulting ultrafiltrate or primary urine coming from the glomerular membrane is collected by the Bowman's capsule and directed into the tubule [Thews *et al.*, 1999]. The tubule is an endothelial structure with different subdivisions, the proximal tubule, the loop of Henle, the juxtaglomerular apparatus, the distal tubule, the cortical collecting tube and finally the medullary collecting duct.

Besides water, waste products and exogenous substances the ultrafiltrate also contains many useful compounds which are present in excess in the blood because of metabolism, food or beverage intake. The tubules system reabsorbs most of the solved compounds and the water. From the approximately 180 l primary urine per day filtered by the glomerulus only about 1.5 l, less than 1 %, will be excreted as urine [Thews *et al.*, 1999]. Everything essential for the organism, like electrolytes, vitamins, glucose and amino acids are reabsorbed, only excesses are excreted as urine. Proteins and Peptides which are normally present only in very small amounts are degraded to single amino acids, di- or tripeptides in the proximal tubule to resorb them [Thews *et al.*, 1999]. Different active and passive transport mechanism, partly controlled by hormones, are responsible for the reabsorption out of or the secretion of compounds into the urine [Thews *et al.*, 1999]. Na<sup>+</sup> for example is actively transported from urine to blood thereby causing chlorine ions and water to follow passively through the osmotic pressure [Thews *et al.*, 1999].

Through the regulation of the glomerular filtration rate (GFR) the kidneys are able to balance the blood pressure in medium to long term by adjusting the total blood volume [Thews *et al.*, 1999]. As mentioned above GFR of healthy kidneys is about 180 l/day for an adult man. The GFR is defined as the total volume of filtered fluid per time period in all glomerulus [Thews *et al.*, 1999]. It depends on the membrane area, the water permeability and the effective filtration pressure [Thews *et al.*, 1999]. The effective filtration pressure is influenced by hydrostatic pressure difference between the blood in the glomerulus capillaries and the ultrafiltrate inside the Bowman's capsule as well as the oncotic pressure, a form of osmotic pressure generated by the difference in amount of proteins with high molecular weight since smaller molecules can pass the membrane [Thews *et al.*, 1999].

### 1.1.2.2 Classification of chronic kidney disease

The GFR is a good measure for the kidney function and therefore also for complications like CKD. Reduced filtration rates (<60 ml/min/1.73 m<sup>2</sup>) for more than three months is an indication of CKD [Levey *et al.*, 2005]. The GFR can be estimated by different equations taking into account the serum creatinine level (SCr), body weight (m), age (a), as well as different factors for race (r) and gender (g) [Solini *et al.*, 2011].

Cockcroft-Gault formula:

$$GFR = \frac{(140 - a) \times m \times g}{72 \times SCr}$$

Levey's formula (Modification of diet in renal disease equation):

$$GFR = 186 \times SCr^{-1.154} \times a^{-0.203} \times g \times r$$

CKD epidemiology collaboration formula

$$GFR = 141 \times \left( \text{minimum} \left( \frac{SCr}{g_1} \right) \text{ or } 1 \right)^{g_2} \times \left( \text{maximum} \left( \frac{SCr}{g_1} \right) \text{ or } 1 \right)^{-1.209} \times 0.993^a \\ \times g_3 \times r$$

Creatinine results from the muscle metabolism as a break-down product of creatinine phosphate. The amount of creatinine formed depends on different factors like age, gender, muscle mass and nutrition [Rule *et al.*, 2009]. To balance the blood concentration, clearance of creatinine is carried out via the glomerular filtration and active secretion from the blood through the tubules.

These equations also have a few drawbacks and are not reliable in all circumstances. In cases when high accuracy is needed like potential kidney donors or when people with abnormal muscle mass, low body mass index, higher levels of creatine or creatinine in the diet and rapidly changing kidney function are examined, clinical

clearance measurements using exogenous filtration markers like inulin are necessary for the GFR estimation [Levey *et al.*, 2005]. On the basis of the estimated filtration rate the progression of CKD can be divided into five different stages (Table 1) [Levey *et al.*, 2005; Solini *et al.*, 2011].

CKD stage	GFR ml/min/1.73 m <sup>2</sup>	Description
1	≥ 90	Kidney damage with normal or increased GFR
2	60-89	Kidney damage with mild decreased GFR
3	30-59	Moderate decreased GFR
4	15-29	Severe decreased GFR
5	<15	Kidney failure (dialysis or kidney transplant needed)

**Table 1: Classification of chronic kidney disease**

Abbreviations: CKD, chronic kidney disease; GFR, glomerular filtration rate

Kidney damage is defined as structural or functional abnormalities of the kidney with or without decreased GFR [Levey *et al.*, 2005]. The damage can be determined by pathologic abnormalities or different markers including changes of the blood and urine composition as well as abnormalities in imaging tests.

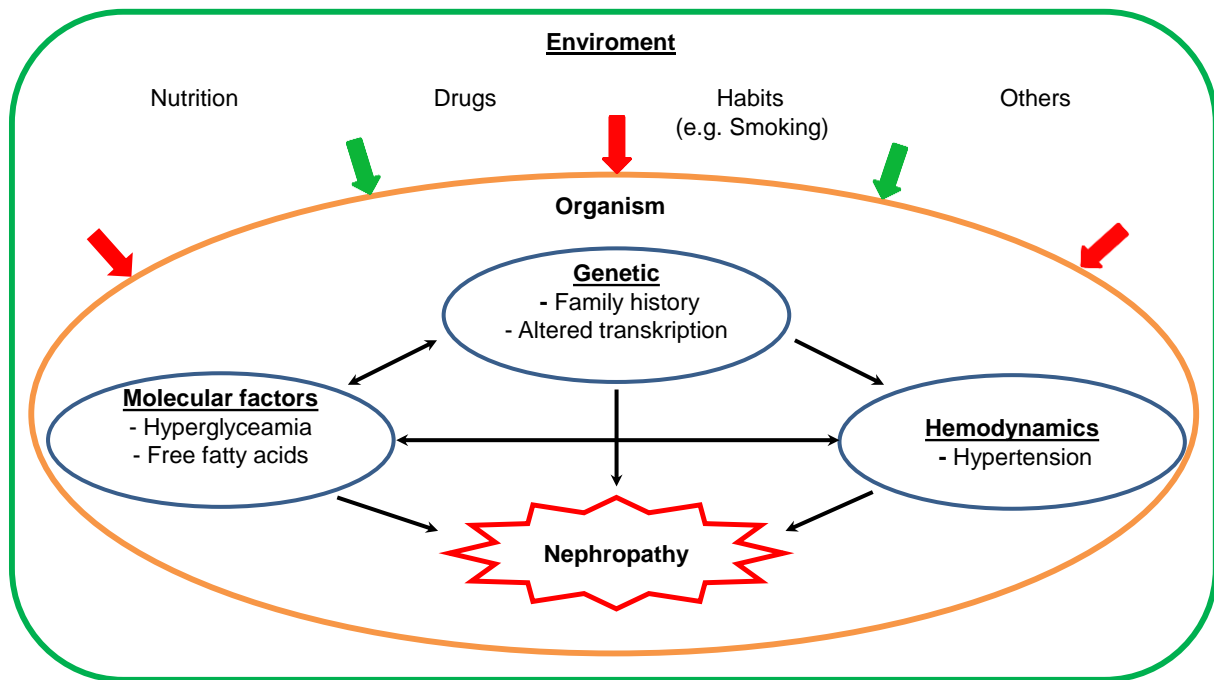
A good marker for kidney damage is proteinuria. Albumin, with a concentration of about 43 g/L, is the most abundant plasma protein and also the preferred urinary protein for kidney damage tests [Thews *et al.*, 1999]. The earliest manifestations of CKD due to diabetes or other diseases is the elevated concentration of albumin in urine [Winchester *et al.*, 2010]. A urinary excretion rate below 30 mg/24 h is regarded as normal; from 30-300 mg/24 h the albumin excretion is elevated, a so-called microalbuminuria (MA) and an excretion rate above 300 mg/24 h refers to clinical or macroalbuminuria [Molitch *et al.*, 2004].

There are three different methods for screening of albumin. The albumin concentration can be determined in a spot urine collection with the albumin to creatinine ratio, a 24 h urine collection which allows a simultaneous measurement of the creatinine clearance or a timed collection [Molitch *et al.*, 2004]. Detection of MA can be done by trained personnel with reagent tablets or dipsticks with acceptable sensitivity and specificity [Molitch *et al.*, 2004]. The measurement of albumin also has some weaknesses which can lead to false negative test results. As mentioned above, digestion of albumin happens in the proximal tubules and only exceeds are detectable. Also extreme urinary pH values, the storage conditions and the sampling container surface can influence the result [Miller *et al.*, 2009]. The problem of different albumin species and fragments in urine which are not all detectable by dipstick tests can be overcome by using polyclonal immunoassays or liquid chromatographic methods [Miller *et al.*, 2009]. But most importantly, the detection of albuminuria already indicates a damage of the kidney. Thus very early detection or better prediction of albuminuria can help to prevent or reverse the progression of CKD by suitable therapies [Ruilope *et al.*, 2010]. Still at present the measurement of albumin is a powerful predictor of renal risk in patients with T2D and the American Diabetes Association recommends yearly tests for albuminuria [Basi *et al.*, 2008].

#### 1.1.2.3 Pathogenesis of diabetic nephropathy

The pathogenesis of DN shows several distinct phases. Before clinically measurable changes can be observed, functional changes of the glomerulus in the nephron take place [Dronavalli *et al.*, 2008]. Hyperfiltration, increased glomerular filtration, hypertrophy and intrarenal hypertension are followed by thickening of the glomerular basement membrane, expansion of the mesangium and diffuse glomerulosclerosis [Dronavalli *et al.*, 2008; Wolf *et al.*, 2003]. The tubular basement membrane is thickening as well closely or parallel to the glomerular membrane and both are consequences of the extracellular matrix accumulation [Fioretto *et al.*, 2007]. These changes develop slowly over years after onset of diabetes, the non-symptomatic phase, leading in the end to proteinuria. Still the further progression from MA to macroalbuminuria or even nephropathy is not given [Dronavalli *et al.*, 2008].

Onset and progression of DN is promoted by several factors (Figure 6).



**Figure 6: Different factors influencing the pathogenesis of diabetic nephropathy.**

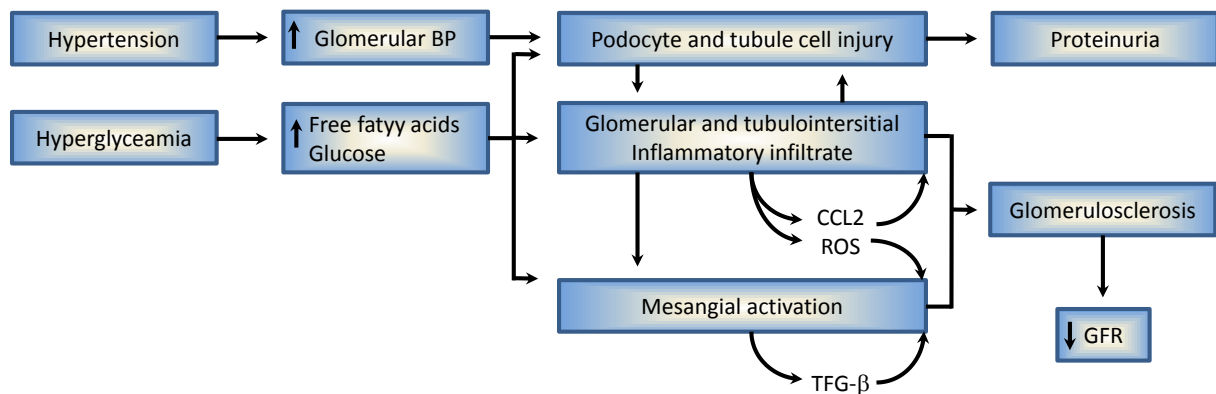
The body internal factors influence each other via different mechanisms leading to the development of diabetic nephropathy. Lifestyle and medication also affect the body internal factors in beneficial or harmful ways. This opens the possibility to prevent or at least delay the onset of the disease by controlled interventions.

## Genetic Factors

Different twin and family studies prove the importance of genetic factors concerning onset, incidence and severity of the disease. Patients with first or second degree relatives having a CKD at end stage for example have a three- to nine fold higher risk to also develop a stage 5 CKD compared to others without [Satko *et al.*, 2007]. In T2D nephropathy the carnosinase 1 gene was identified as important factor for the pathogenesis [Satko *et al.*, 2007]. The huge differences in the pathogenesis of DN favour involvement of many genes, hence multigene approaches are preferred in genetic research of DN [Dronavalli *et al.*, 2008]. The impact of family history, different genetic approaches and several important genes were reviewed in detail by Freedman *et al.* [Freedman *et al.*, 2007].

## Hemodynamic Changes

A major player in the onset and development of DN is hypertension which itself is worsening with the progression of DN [Wolf *et al.*, 2003]. Van Buren *et al.* [2011] summarize in their review the factors which contribute to an increased blood pressure. The renin-angiotensin-aldosterone system induces vasoconstriction in many tissues and increased sodium reabsorption in the kidney. The blood volume and therefore the pressure rise with the shifted sodium balance. Further the changed activities of the sympathetic nervous system, the endothelial cells and increased oxidative stress aggravate hypertension. Renal hemodynamic changes can cause an impaired blood pressure autoregulation in the glomerulus which is then directly affected by the systemic pressure. Decreased resistance of mainly the afferent but also of the efferent arteriole for example cause glomerular hyperperfusion and hyperfiltration which enables proteins like albumin to pass the glomerular capillary membranes (Figure 7) [Dronavalli *et al.*, 2008; Kanwar *et al.*, 2008]. Untreated, local inflammations, fibrosis and further injuries are following and cicatrise the membranes (glomerulosclerosis) reducing the GFR until dialysis is necessary [Van Buren *et al.*, 2011].



**Figure 7: Extrarenal and intrarenal factors involved in diabetic nephropathy**

Diabetes is characterised by endothelial dysfunction and elevated concentrations of glucose and free fatty acids, among other perturbations. Systemic hypertension increases intraglomerular pressure, which causes mechanical injury to the endothelium and podocytes. Excess glucose and free fatty acids also directly harm endothelial and epithelial components of the nephron. Mechanical and biochemical stressors converge to elicit an ongoing inflammatory response, which exacerbates injury and induces an aggravated mesangial response. The chronicity of diabetes facilitates vicious cycles of injury, inflammation and sclerosis/fibrosis, fueled by local synthesis of cytokines, chemokines (for example, CCL2) and profibrotic factors (for example, TGF- $\beta$ ). Podocyte demise and tubule epithelial dysfunction allow filtration and urinary excretion of albumin whereas progressive glomerulosclerosis gradually impairs glomerular filtration. Abbreviations: BP, blood pressure; CCL2 C-C motif chemokine 2; ROS, reactive oxygen species, TGF- $\beta$ , transforming growth factor  $\beta$  (Adapted by permission from Macmillan Publishers Ltd: [Shapiro *et al.*, 2011])

## Molecular factors

All of the aforesaid cellular and hemodynamic changes are accompanied by different molecular mechanisms. Hyperglycaemia as a result from inadequately controlled blood sugar level has a great impact in the process of DN development and its outcomes (Figure 7) [Dronavalli *et al.*, 2008; Wolf *et al.*, 2003]. A higher glucose concentration shifts different extra- and intracellular ratios. It increases the formation of advanced glycation end-products (AGE), activates the protein kinase C, alters different metabolic pathways and generates reactive oxygen species (ROS) [Dronavalli *et al.*, 2008; Kanwar *et al.*, 2008]. AGEs are formed non-enzymatically by a reaction of reducing sugars with free fatty acids, lipids, proteins and nucleic acids. By interaction with receptors and binding proteins they trigger several intracellular events like increased activity of growth factors and they also help to generate ROS [Kanwar *et al.*, 2008]. Activated by AGEs and ROS protein kinase C plays an important role in altering the blood flow, capillary permeability, local tissue inflammations and thrombotic microangiopathy, a thrombosis in the capillary caused by an endothelial injury [Kanwar *et al.*, 2008]. In addition, different pathways are deregulated in DN as well. Changes in small lipids derived from arachidonic acid, the aldose reductase pathway as well as activation of various cytokines, small cell-signalling proteins which are promoting a multiplicity of reactions from thickening of membranes to inflammations are just some examples [Dronavalli *et al.*, 2008]. All these changes have ROS as denominator or amplifier [Kanwar *et al.*, 2008]. ROS are mainly generated inside the mitochondria and normally suppressed by different enzymatic reactions or radical scavenging molecules. With abundance of glucose and fatty acids the amount of ROS generated overburdens the natural defence systems [Dronavalli *et al.*, 2008]. In the mitochondria, oxidative stress is generated during the oxidative phosphorylation via an electron loss in the electron transfer chain of the respiratory chain by forming superoxide anions [Kanwar *et al.*, 2008]. Superoxide anions can form hydrogen peroxide, hydroxyl radicals and peroxynitrite which altogether can induce renal injuries [Kanwar *et al.*, 2008]. A very detailed review with many different molecular mechanisms was published by Kanawar *et al.* [Kanwar *et al.*, 2008] and shows the multifactorial nature of the pathogenesis of DN which is still not fully understood. But even this incomplete understanding of the pathogenesis of DN allows different approaches to prevent or delay onset of DN.



DN prevention can be schematically divided into three stages: prevention of T2D, prevention/reversal of early renal changes and prevention/reversal of overt DN [Tzamaloukas *et al.*, 2005].

Onset of T2D can be avoided or delayed by different lifestyle interventions like prevention of obesity with balanced diets and sport. Prevention or reversal of early renal changes and overt DN can be achieved by strict glycaemic control, therapy of hypertension, lipid control and other measures like ceasing of smoking, protein restricting diets and avoiding of nephrotoxic influences [Tzamaloukas *et al.*, 2005].

Also, the albumin-creatinine-ratio is widely used as markers for the risk assessment of DN, the kidney is already damaged when the ratio has significantly changed. Therefore research of new biomarkers for early detection of or even prediction of DN is necessary to avoid or at least delay the development of overt DN with appropriate lifestyle interventions and treatments. In order to fully understand the pathogenesis of DN and to find new targets for risk assessment and treatment to prevent or delay kidney failure and related complications, different approaches are necessary in research.

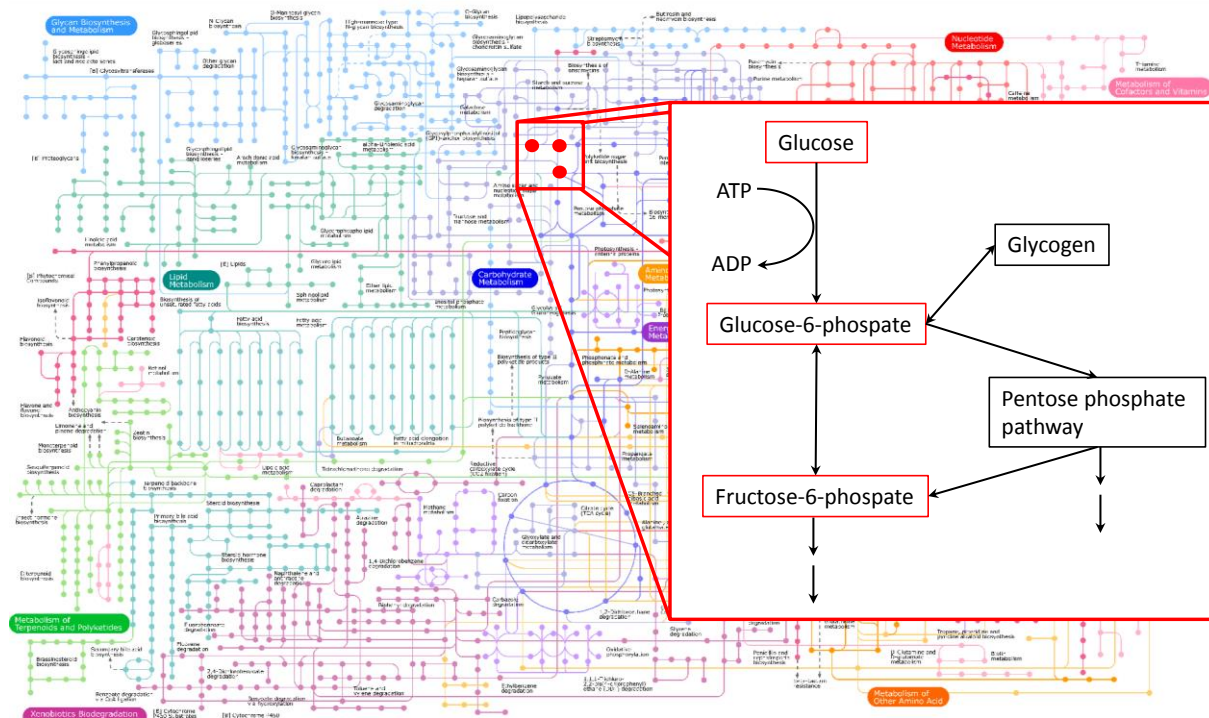
## **1.2 Metabolomics**

The general aim of metabolomics and the interchangeably used term metabonomics is to identify and quantify all metabolites in the analysed biological sample to understand biochemical mechanisms and systematic changes of the studied organism [German *et al.*, 2004; Nicholson *et al.*, 2008; Pearson, 2007].

### **1.2.1 Metabolic network and its regulation mechanisms**

Metabolites are small molecules, below 1000 Da [Holmes *et al.*, 2008], which are intermediate or end products of metabolic reactions of an organism. It is vital for every organism to constantly absorb nutrients from the environment, metabolize them and release metabolic end products in order to generate energy and building blocks for growth and survival [Soh *et al.*, 2010]. The entity of metabolites is called metabolome.

Different enzymatic reactions linked via intermediate metabolites are often assembled in pathways where the end product of one reaction is the substrate for the following. The pathways are often further classified in catabolic pathways, which degrade complex molecules to usually generate energy and anabolic pathways which usually require energy to synthesize bigger molecules. A map of these pathways consists not of separated straight sequences of reactions but form a complex metabolic network since metabolites can be substrates or end products of different reactions from different pathways (Figure 8) [Gruning *et al.*, 2010].



**Figure 8: Complexity of the interaction between metabolic pathways.**

The different pathways are formed by metabolites (points) which are linked with different (enzymatic) reactions (lines). Given that metabolites can be educts or products of different pathways and some reactions are reversible while others are not a complex network of metabolic reactions exists within cells and organisms. Abbreviations: ATP, adenosine triphosphate; ADP, adenosine diphosphate (metabolic map from <http://pathways.embl.de/iPath2.cgi>, [Yamada *et al.*, 2011])

Since one Enzyme catalyses a reaction several times, small changes in the amount of enzyme or its reactivity have big effects on the involved metabolite concentrations. Therefore a strict control of the pathways is necessary for organisms to maintain homeostasis and respond to environmental or internal changes. The metabolic conversion through a pathway is regulated by the amount of active enzymes, the activity of the enzymes and the concentrations of the substrates [Berg *et al.*, 2003].

Metabolites are not just substrates or products of enzymatic reactions; they also play an important role in regulating the pathway and the whole metabolic network. Apart from direct inhibition or activation of enzymes also the transcription and other mechanisms are influenced by some small effector molecules termed as “reporter metabolites” [Gruning *et al.*, 2010; Patil *et al.*, 2005]. They regulate the amount of enzymes by repressing or enhancing their transcription and the substrates concentration through interaction with molecular transporters or channels [Gruning *et al.*, 2010]. Depending on their level of influence reporter metabolites are further classified in specialised and global reporters. The former are metabolites regulating only particular pathways or mechanisms while global reporters affecting different pathways and mechanisms [Gruning *et al.*, 2010]. Enzymatic cofactors like ATP, ADP and NAD(H) are such global reporters and changes in their concentrations or species ratio which can be observed in T2D patients have great impact on the gene expressions and therefore the metabolic network [Zelezniak *et al.*, 2010].

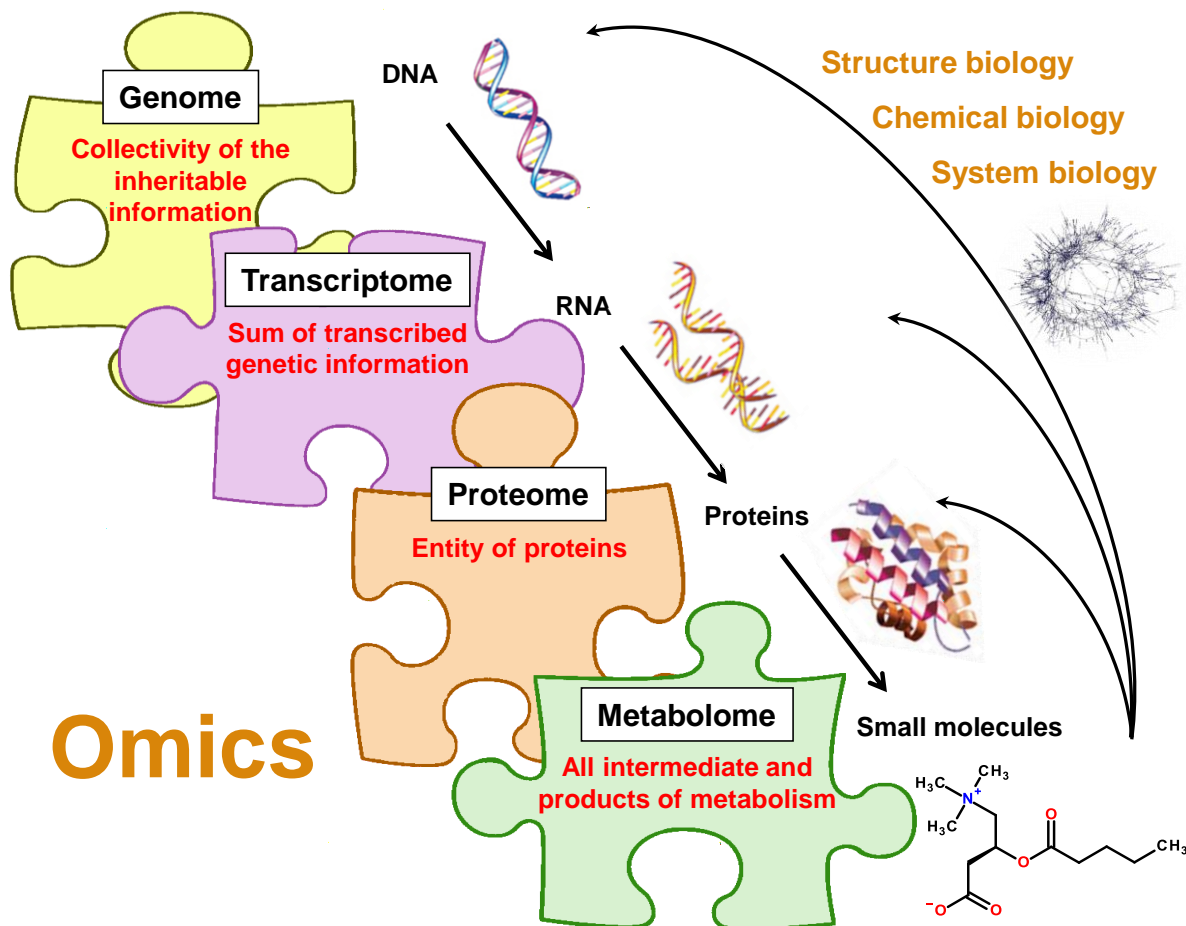
Our knowledge of the metabolic network of complex superorganisms like humans is still incomplete since at the moment only approximately 10% of all metabolites are identified. The total number of existing metabolites in mammals, plants and microbes is thought to exceed two hundred thousand or even millions and therefore the number of genes [Ott *et al.*, 2006; Schwab, 2003]. This large amount of unexplored metabolites is accompanied by many new pathways and new reporter metabolites which can also influence already known pathways. Thus, these compounds provide not only a big challenge but also a great opportunity to discover new interesting biomarkers and bioactive compounds. Anything from a single analyte to complex image modalities or multi-marker patterns that can distinguish between different biological states is referred to as biomarker [Baker, 2005]. A relatively new useful tool and scientific discipline to reveal unknown metabolites, pathways and biomarkers is metabolomics.

### **1.2.2 Metabolomics as tool for biomarker identification**

Metabolomics is a relatively new approach among the other “-omics” techniques, genomics, transcriptomics and proteomics. Although the basic idea of detectable changes in biofluids and tissues through diseases is very old the modern concept of metabolomics started in the late 90’s, with Oliver *et al.* [Oliver *et al.*, 1998] and

Nicholson et al. [Nicholson *et al.*, 1999]. At first the term metabonomics was defined by Nicholson et al. [Nicholson *et al.*, 1999] as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification”. Three years later Fiehn [Fiehn, 2002] introduced metabolomics as “a comprehensive analysis in which all the metabolites of a biological system are identified and quantified”. Since the used analytical methods are the same, the terms are often used interchangeably, but in recent years metabolomics is used more frequently.

The future aim of metabolomics is to determine the risk to develop a special disease, the health profile or the efficiency and side effects of drugs by a single analysis of e.g. a urine, plasma or exhaled breath sample of an individual in a routine health check-up [German *et al.*, 2005; Pearson, 2007]. Compared to the other “-omics” techniques metabolomics allows the global assessment of the state of an organism, taking into account the genetic regulation, enzymatic changes and environmental influences reflecting best the phenotype (Figure 9) [Dettmer *et al.*, 2007].



**Figure 9: Connection between the different "omics" research fields and their information content.**

Unlike the genome, the transcriptome, proteome and metabolome of an organism are reflecting a state at a particular time. (Translated by permission from the Thieme Publishing Group [Wörmann *et al.*, 2012])

The phenotype of an organism or if specified of a subsystem refers to all the observable characteristics and traits, from morphology over biochemical properties to behavior [Mahner *et al.*, 1997]. Through comparison of metabolic phenotypes or changes in it, metabolomics is able to discover new prognostic and diagnostic markers for diseases like T2D, and distinguish between different tissues or microorganisms. Apart from enhancing the diagnosis of diseases through discovery of new marker compounds they also forward the understanding of complex diseases. Additionally by determination and prediction of drug effects ("personalised medicine) or discovery of novel therapeutic drugs as well as drug targets, the therapy of diseases itself can be improved. These examples display the huge potential of metabolomics for biology, medicine and environmental sciences.

Unfortunately until now no single analytical technique is able to detect all metabolites of a sample in a single analysis. This is the big drawback of metabolomics and also a possible reason why the other “omics” approaches were preferred in the past [Dettmer *et al.*, 2007]. The reason for this dilemma is the huge biological diversity of metabolites both chemically and physically. The physicochemical characteristics of metabolites extend from ionic to very lipophilic species over a very broad concentration range which covers nearly 11 orders of magnitude in human blood [Psychogios *et al.*, 2011]. Therefore nowadays different analytical platforms have to be applied to draw near the goal of a complete quantification of every metabolite of a biological sample. The different analytical techniques used for this thesis with their opportunities and drawbacks are discussed in detail in Chapter 2.

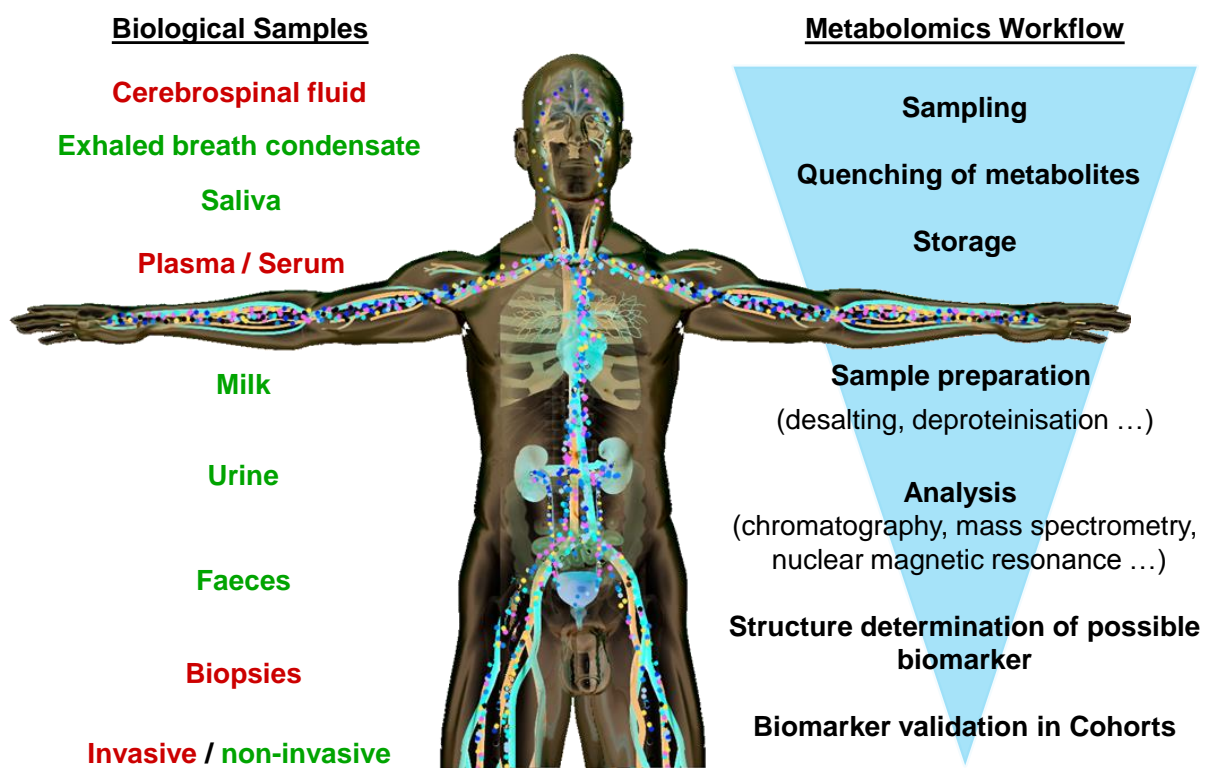
In principle, metabolomics approaches can be divided into two fractions. The targeted metabolomics approaches are quantifying a subset of up to 200 already known metabolites directly via comparison to an isotopic labelled standards or indirectly by comparison to a similar compound. In contrast the goal of untargeted approaches is to detect as many metabolites as possible to identify new unknown compounds or discover new connections between known metabolites or metabolic patterns and different biological states. In the literature different metabolomics strategies are mentioned and Table 2 gives a brief overview of the most important ones.

Strategy	Focus on	Detection techniques
Metabolite target analysis	certain metabolites	MS
Metabolite profiling	a specific group of metabolites	MS, NMR, EC
Metabolomics	entire metabolome	MS
Metabolic flux analysis	metabolites of stable isotope labeled substrates	NMR, MS
Metabolic fingerprinting/footprinting	classification of samples based on their intracellular/extracellular metabolite profile	NMR, MS, IR, Raman

**Table 2: The most popular metabolomics strategies**

To improve the detection different chromatographic methods are used before detection. Abbreviations: MS, mass spectrometry; NMR, nuclear magnetic resonance spectroscopy; EC, electrochemical array; IR, infrared spectroscopy. (Adapted with permission from [Goodacre, 2007])

Depending on the objective of the survey different strategies are recommended. For validation of a hypothesis a targeted approach may be sufficient while non-targeted strategies help to generate new theories. The next step is to consider the metabolites of interest and the sample type. Some metabolites can be found only in some biofluids or tissues and may be unstable, which has to be considered in the analytical workflow. Some examples of different biological samples from humans or animals are shown in Figure 10 along with a typical non-targeted workflow. This setup already provided some interesting results in the area of diabetes research.



**Figure 10: Interesting biological samples for metabolomics analysis and a typical non-targeted workflow**  
(Translated by permission from the Thieme Publishing Group [Wörmann *et al.*, 2012])

### 1.3 Objection and overview of the thesis

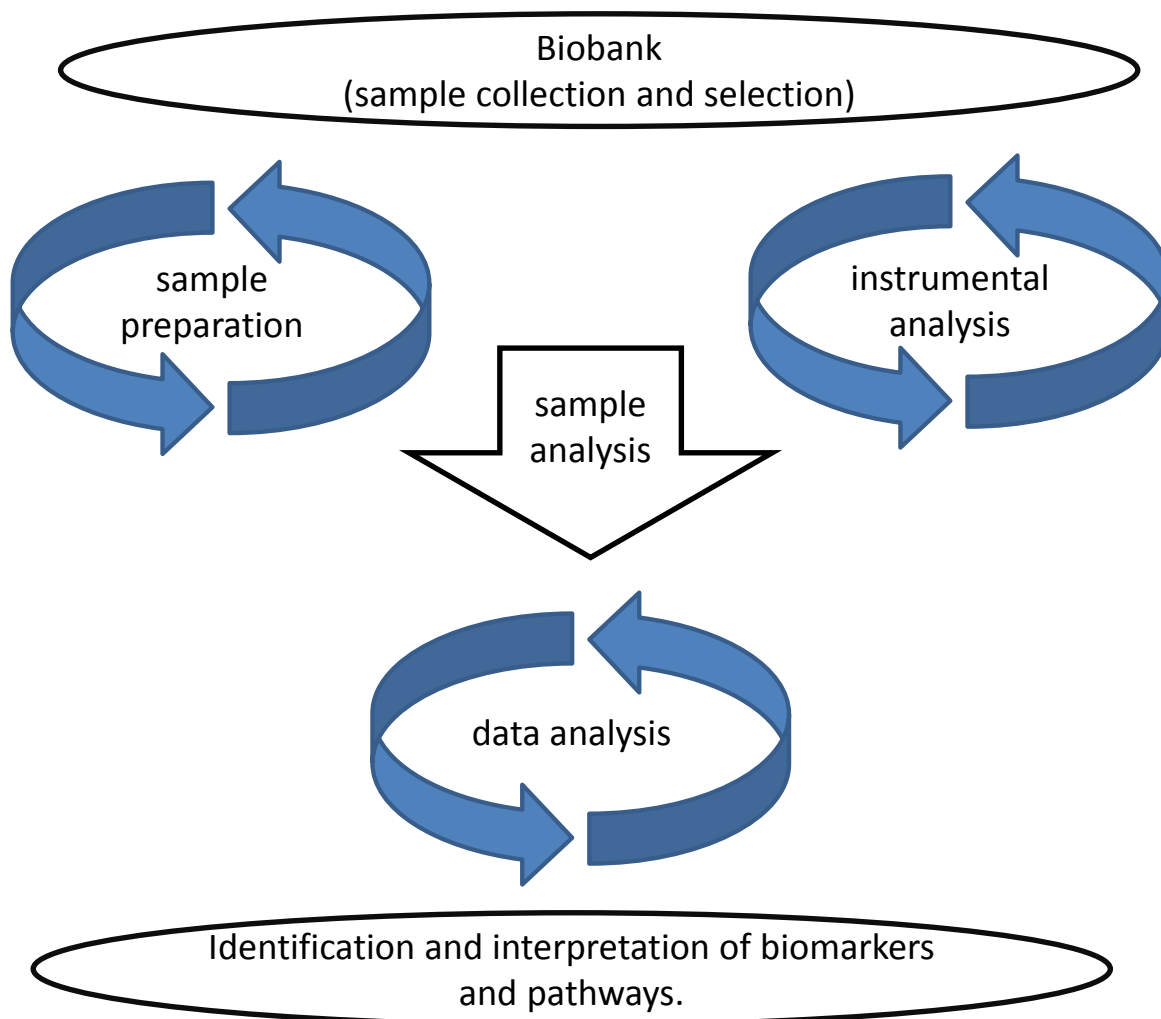
Recent progresses in instrumentation, data analysis and databases enabling metabolomics to offer tremendous possibilities to study the metabolome of superorganisms like human. An important and interesting field of this relatively new scientific discipline is the disease marker identification. Metabolomics helps to improve the understanding of the alterations diseases like T2D cause in the human metabolism and to identify novel prognostic and diagnostic biomarkers. Since T2D and DN have complex pathogenesis mechanisms with many contributing factors, it is difficult to find specific and reliable markers for diagnosis and prognosis. Focusing on DN from the various body fluids and biopsy samples, urine offers a good possibility for the biomarker detection. First of all it is directly connected with the concerned organ since the kidney is responsible for the urine production. Another advantage is the easy and non-invasive accessibility of sufficient amount of sample especially for screening of children. On the other hand one has to keep in mind that urinary metabolite concentration depends on various parameters like liquid intake for example.

Non-targeted metabolomics is a useful tool to help uncover the diverse involved factors in pathogenesis of DN. Within this work it will be used for identification of discriminative metabolic alteration and identification of biomarkers and pathways for microalbuminuria and DN. Before reaching the main objective different aspects had to be considered and sub-goals had to be met (Figure 11).

The fundament of metabolomics studies are the samples affected study design and sample selection. The next challenge was the development of a sample preparation without big metabolite losses for the different analytical techniques. Currently there is no golden analytical technique for detection of all metabolites in one go, thus several techniques are combined to cover as many compounds as possible. In the two different workflows state-of-the art nuclear magnetic resonance spectroscopy, ultra high performance liquid chromatography and direct injection ultra-high resolution mass spectrometry are used for the sample analysis. The development for data analysis and interpretation comprises data pre-treatment techniques, comparisons between different instruments, univariate and multivariate data analysis as well as a new network approach and visualisation. All these steps had to be combined for the



identification of novel biomarkers and pathways, in order to increase the biological understanding of this complex disease and its pathogenesis.



**Figure 11: Milestones in the process of biomarker discovery with non-targeted metabolomics**  
Points surrounded by blue arrows have been developed and optimised within this thesis.

The work at hand is based on modern, state-of-the-art ultra-high resolution mass spectrometry, ultra-high performance liquid chromatography hyphenated to mass spectrometry and nuclear magnetic resonance spectroscopy. The obtained data was integrated and analysed with a novel network approach and a combination of multi- and univariate statistics.

Summing up the thesis at hand combines analytical chemistry, pathology and data analysis under the structure presented by Figure 12.

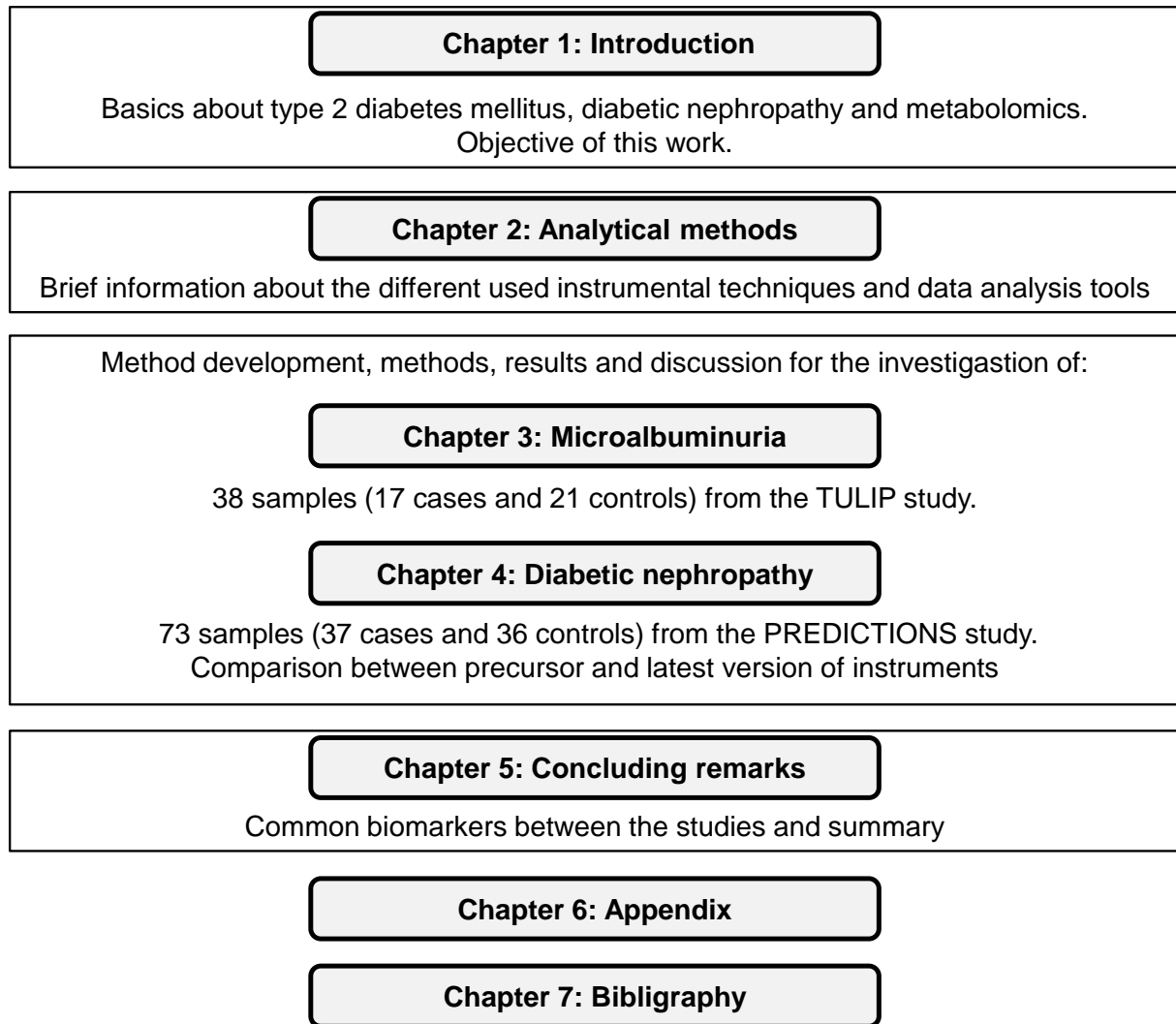


Figure 12: Thesis structur

## 2 Analytical methods

### 2.1 Instrumental techniques

#### 2.1.1 Liquid Chromatography

There are two major problems analysing complex samples with direct injection MS. On the one hand complex samples contain many isobaric compounds which are detected as one single peak. On the other hand matrix effects can influence the detection of different compounds by enhancing or preventing their ionisation and therefore their detection. Separation techniques like liquid chromatography (LC) help to overcome these problems by separating compounds on a time scale, therefore reducing the amount of different substances injected simultaneously into the MS system.

LC is a separation technique based on different interactions between solved compounds, a liquid mobile phase and a stationary solid phase. The kind and force of these interactions depend strongly on the chemical and physical characteristics of all substances. The interactions can be divided into chemical interactions like hydrogen bridging and physical interactions including size, steric structure, dipol-dipol or ion-dipol attraction or repulsion, the van-der-Waals force and hydrophobic interactions. Also the partitioning behaviour of solved substances between different mobile phases is important. All these influences on the retention behaviour combined result in a specific migration speed for each compound giving the possibility to separate them.

An effective setup is to pack the stationary phase into columns pumping the mobile phase with the analysts through it. The most important equation for the theoretical explanation and optimisation of this setup was introduced by Van Deemter. It describes the dependents of the separation efficiency, the theoretical plate height ( $H$ ), on the particle size ( $d_p$ ), velocity ( $v$ ), diffusions coefficient ( $D_m$ ) and viscosity ( $\eta$ ).

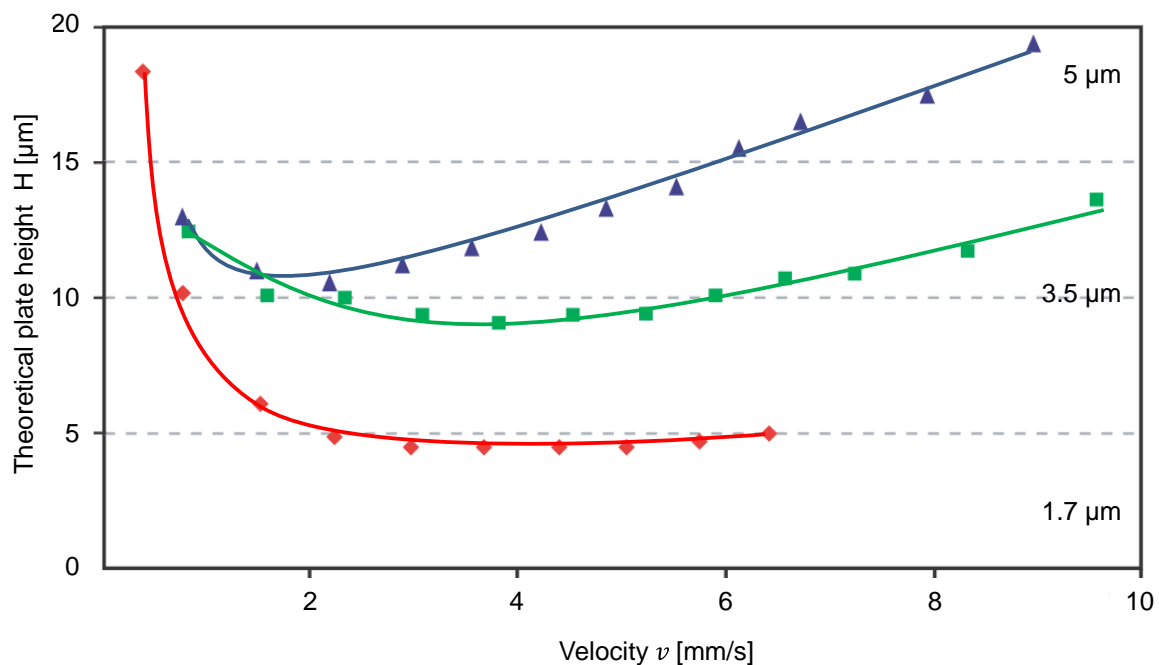
Van Deemter equation:  $H = A + \frac{B}{v} + Cv$

Lateral diffusion:  $A = 2\lambda d_p$

Longitudinal diffusion:  $B = 2\gamma D_m$

Mass transfer:  $C = \omega d_p$

Since the particle size improves two parts of the van Deemter equation, it was a main focus of development to narrow the size and the size distribution. The smaller the particle diameter, the lower is the theoretical plate height  $H$  (Figure 13) and the better is the separation.



**Figure 13: Dependency of the theoretical plate height  $H$  from the particle size  $d_p$  and the velocity  $v$**   
(Adapted with permission from [Mazzeo *et al.*, 2005])

Apart from the theoretical plate height also the optimal velocity ( $v$ ) range increases, thus allowing higher flow rates and as a result faster separation. But reducing the particle diameter creates a pressure problem due to an exponential increase of the backpressure. This was overcome in the past years by essential developments of chromatographic instrumentation and materials allowing enhanced separation with enhanced efficiency and higher speed. The development of ultra-high performance liquid chromatography (UHPLC) systems which are able to handle pressures up to 120 MPa (1200 bar) allows the use of columns with homogenous distributed and

pressure stable sub-2- $\mu\text{m}$  particles. Guillarme et al. [Guillarme *et al.*, 2008] showed the improvements of the separation and analysis time comparing a new UHPLC system with the older system.

For the LC method development various aspects can be optimised. The choice of the mobile phase, its pH and the use of additives are as important for the separation as the column material, the column dimension and the column temperature. In metabolomics analysis mostly reversed phase chromatography (RP) are performed beneath hydrophilic interaction liquid chromatography (HILIC). In RP the particle surface is coated with different lipophilic layers like with hydrocarbon chains with 18 carbon atoms while the mobile phase is polar or changes from polar to less polar or non-polar with a gradient during the separation.

In metabolomics LC is often coupled with MS to improve the result. In targeted approaches the separation of isobaric compounds is important for quantification since it is based on the peak area. In non-targeted approaches with their aim to see as many metabolites as possible, LC is a powerful tool to improve the analysis.

Within this thesis a non-targeted RP UHPLC MS method was optimised for maximisation of the detected metabolites. In the second part an existing method was converted for the analysis of a bigger sample sets in a given time frame.

### **2.1.2 Mass spectrometry**

Nowadays MS is a valuable tool in many different scientific fields apart from metabolomics, but it all started with the measurements of atomic masses and the discovery of isotopes [Griffiths, 2008]. Since the first mass spectrometers from J.J. Thomson, F.W. Aston and A.J. Demster in the early 20<sup>th</sup> century a lot has happened. The improvements of mass detectors, vacuum systems, ion lenses and magnetic fields increased the mass resolving power and mass accuracy dramatically. Especially the development of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) enables mass detection with ultra-high resolution, which is most important for the complex sample matrices metabolomics deals with.

The resolving power of an instrument is determined by the peak width at a specific peak height and the full width at half maximum definition of resolution is well-established [Gross, 2011]. However the mass resolution (R) is defined for a given mass to charge (m/z) signal as the smallest difference between neighbouring signals which can still be distinguished [Gross, 2011].

$$R = \frac{m/z}{\Delta m/z}$$

Besides the resolving power, also, the mass accuracy of an instrument is important since a high accuracy allows a better calculation of molecular formulas from the measured mass. The mass error between the measured and the exact mass is commonly calculated as relative error in parts per million (ppm).

$$\text{relative error} = \frac{\Delta m/z}{m/z} \times 10^6$$

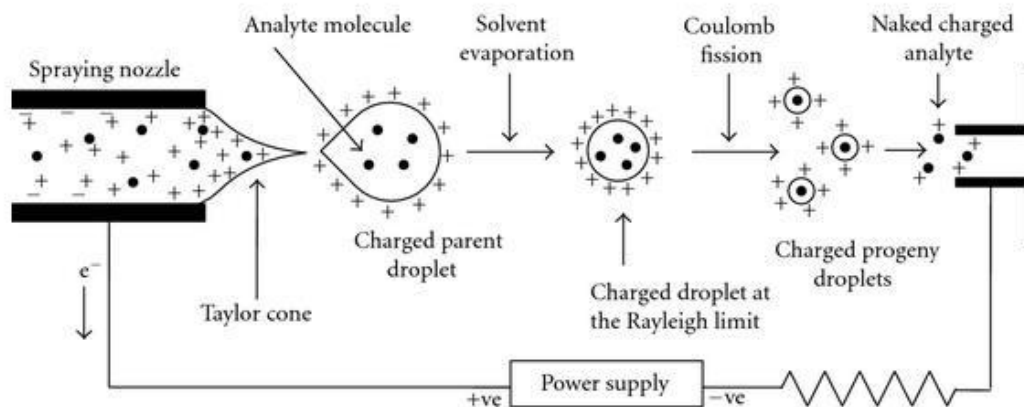
Not only improvement of the mass detection but also of the ion generation made MS to a valuable tool for metabolomics.

### 2.1.2.1 Ionisation

Especially the development of soft ionisation techniques like electrospray ionisation (ESI), atmospheric pressure photo ionisation or atmospheric pressure chemical ionisation which minimize the fragmentation of compounds during the ionisation process were an important step for MS to reach its prominence in metabolomics. Because ESI offers high ionisation efficiency for various compounds ranging from polar to hydrophilic, it is the preferred technique for chemical and bio-chemical analyses [Wilm, 2011].

Figure 14 presents the schematic process of ionisation in an ESI source. The sample liquid flows through a capillary or nozzle with an applied high voltage and is nebulised into charged droplets with the support of a nebulizer gas, usually nitrogen. The nebulizer gas can also be heated to accelerate the solvent evaporation in the droplets. The charged ions spread in the shrinking volume of the droplets until the increasing repulsive Coulomb force causes fission of the droplets into smaller ones. In the end gas-phase ions are released from the small highly charged droplets and

some of these ions are forwarded into the mass spectrometer. Kebarle et al. [Kebarle *et al.*, 2009] summarised our current knowledge about the mechanisms of ionisation in an ESI source in their interesting review.



**Figure 14: Schematic representation of the positive mode electrospray ionisation process**

The liquid sample is charged through applied high voltage at the ESI spraying nozzle releasing charged droplets. Advancing solvent evaporation aided by heated dry gas leads to droplet fission through the increasing repulsive Coulomb forces until “naked” charged analytes remain. (Picture with permission from [Banerjee *et al.*, 2012])

The ionisation can be done in both polarities and is influenced by a large number of parameters. Beneath the choice of the solvent also its pH, additives, salts, analyte concentration, concentration of other analytes, capillary current, dry gas flow and temperature are just some examples. The positive ionisation mode is good for compounds with basic groups like amines whereas in negative mode acidic groups are preferred for ionisation. Nevertheless some compounds can also be detected in both or an unexpected mode since they possess both basic and acidic groups or form complexes with additives or alkali ions like  $\text{Na}^+$ .

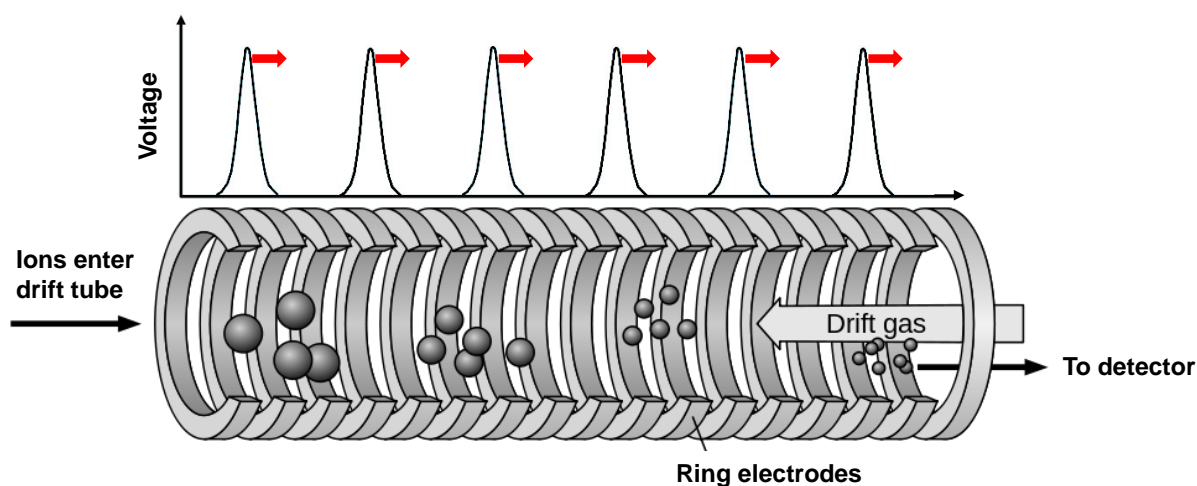
### 2.1.2.2 Ion mobility spectrometry

Ion mobility analysis is well-established for the detection of vapour phase samples to identify explosives, drugs, chemical-warfare agents and air pollution for different public and military purposes. Already in the end of the 19<sup>th</sup> century scientists started to develop instrumentations to explore ions and their mobility [Eiceman *et al.*, 2005]. Around 1960 E.W. McDaniel coupled an ion mobility separation (IMS) with mass

spectrometry for his study of ion-molecule reactions [McDaniel *et al.*, 1962]. In recent years the combination of ESI and new commercially available mass spectrometers with high-resolution IMS provides a powerful analytical tool for proteomics and metabolomics [Kanu *et al.*, 2008].

The four different ion mobility spectrometers which are drift time, aspiration, differential and travelling wave, coupled with MS are varying mainly in the gas pressure, the applied electric fields and the shape of the electrodes but the separation is based on the same principles [Kanu *et al.*, 2008]. The ions are accelerated with an electric field through a drift tube filled with gas while collisions with the gas molecules slow them down again. The individual drift velocities and therefore the drift time of different analysts are influenced by their mass, charge, molecular size and shape.

The traveling wave IMS is a new technique invented and commercially introduced some years ago by Waters [Kanu *et al.*, 2008]. Here, the buffer gas flows through a straight drift cell with ring electrodes at reduced pressure. The ions are accumulated at the entrance of the drift tube and enter it in packets from the opposite side. They are then forced with non-uniform, moving electric fields through the cell against the buffer gas flow. (Figure 15 and [http://www.youtube.com/watch?v=i\\_l2egLZ5IM](http://www.youtube.com/watch?v=i_l2egLZ5IM)).



**Figure 15: Schematic construction and functionality of a travelling wave ion mobility cell**

Ions entering the drift tube are pushed by non-uniform moving voltage pulses applied to ring electrodes through a drift gas. The ions are separated according to their individual drift velocity which depends on their mass, charge, size and shape ([http://upload.wikimedia.org/wikipedia/commons/thumb/6/66/Ion\\_mobility\\_spectrometry\\_diagram.svg/1000px-Ion\\_mobility\\_spectrometry\\_diagram.svg.png](http://upload.wikimedia.org/wikipedia/commons/thumb/6/66/Ion_mobility_spectrometry_diagram.svg/1000px-Ion_mobility_spectrometry_diagram.svg.png), modified)



The smaller and compacter ions collide less often with buffer gas molecules and are therefore faster than bigger ones, while ions with multiple charges are pushed with greater force hence their drift times are lower than singly charged ions with comparable masses. The separated ions are then forwarded to the mass detector.

IMS offers a new dimension in separation since multiple charged ions can be identified. Applying an appropriate molecular cross section calibration apart from the charge also the size of the ions can be estimated. Even without a calibration, different ions with the same mass can be distinguished if they differ in shape.

### 2.1.2.3 Time of flight mass spectrometry

In 1946 Stevens presented the first time of flight mass spectrometer (ToF) which he constructed [Gross, 2011].

In principle ToF instruments measure the time an ion needs to cover a known distance. Therefore a package of ions is accelerated by a voltage ( $U$ ) impulse and flies through a field free tube of defined length ( $s$ ). The mass ( $m$ ) to charge ( $z$ ) ratio is then proportional to the square of the flight time ( $t$ ) the ion need for this distance.

$$\frac{m}{z} = \frac{2 e U}{s^2} \times t^2$$

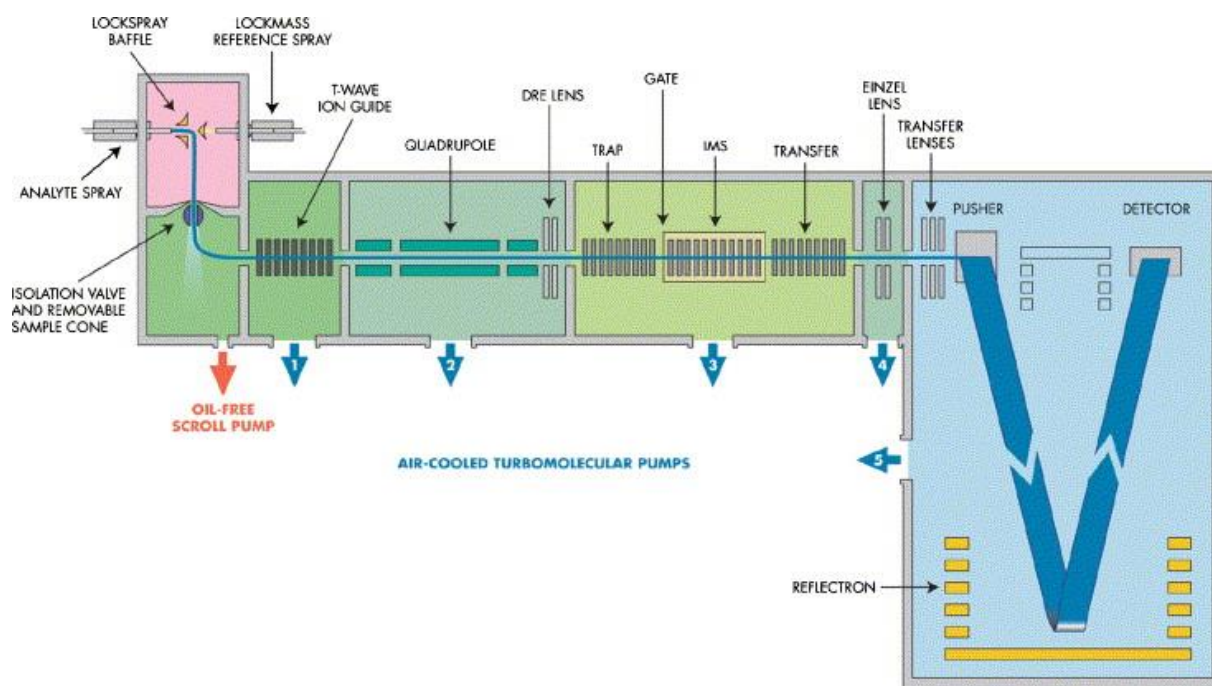
( $e$  = electron charge)

Compared to other mass spectrometers, ToF instruments offer several advantages. They offer a high detection speed and mass range with detection of all  $m/z$  for each ionisation event and are inexpensive because of their simple design and construction [Gross, 2011].

Modern ToF systems are mostly hybrid instruments where different types of mass analysers and ion guiding devices are combined in order to unite the different advantages and allowing tandem mass spectrometry. In tandem mass spectrometry ions can be isolated according to their mass by the first mass analyser. Then the selected ions are fragmented by acceleration and collusion with gas molecules. The

generated fragments are thereafter analysed by the second mass analyser. The fragmentation pattern of an ion can help to identify the molecule by comparing it with patterns stored in online databases or generated from a known standard compound.

Figure 16 shows the general setting of the in the first study (Chapter 3.3.2.2) used Synapt™ HDMS™ mass spectrometer. In the second study (Chapter 4.3.2.2) the newest follow-up system, a Synapt™ G2-S™ HDMS™ system, was used but the principle assembly is the same even if the different components were improved. Both are hybrid Q/oa-ToF dual-stage reflector instruments with a mass resolving quadrupole (Q) and an orthogonal ion acceleration (oa-) ToF mass analyser. After the quadrupole they have three traveling wave ion guides built in allowing fragmentations of ions before and after the IMS. In this work the two hybrid instruments will be generally referred to as ToF.



**Figure 16: A schematic diagram of the Synapt® HDMS system**  
(Reprinted from [Pringle *et al.*, 2007] with permission from Elsevier).

The ions generated within the ESI source enter the vacuum system through cones and are then forwarded to the mass selective quadrupole which allows a focusing on single  $m/z$  ranges. Thereafter the ions fly through three traveling wave ion guides of which the middle one is the IMS Cell and the other two guides (“TRAP” and

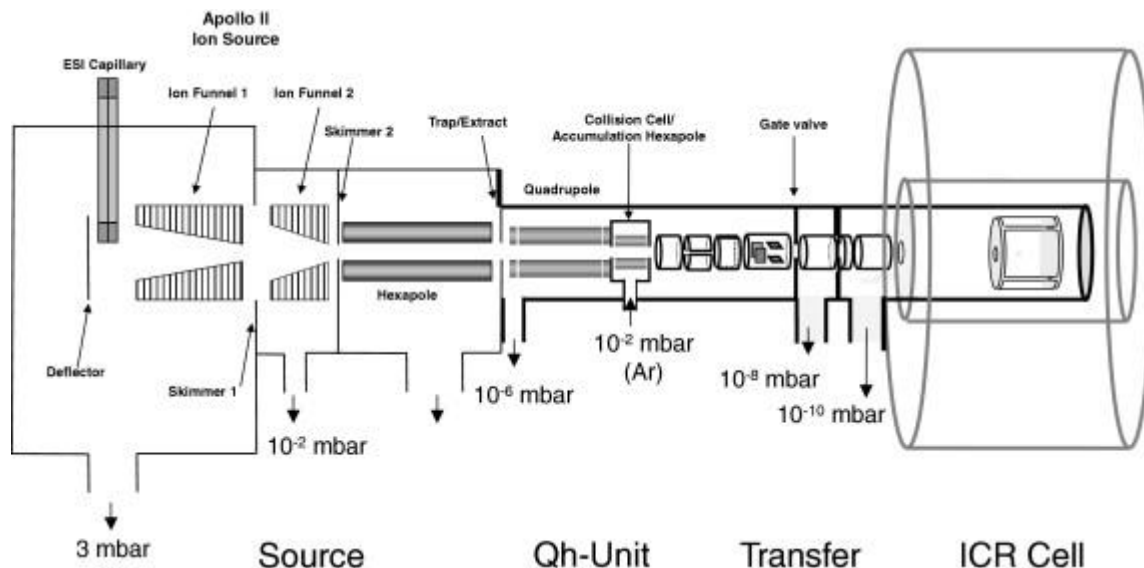
“TRANSFER”) allow fragmentation. The ion guides accelerate the ions within and fragment them through collision-induced dissociation (CID) with argon gas. The fragmentation of ions and the following analysis of the fragments are called tandem mass spectrometry (MS/MS) [McLafferty, 1981]. After refocusing the ions enter the ToF tube. There, a pulsed orthogonal acceleration is applied to ensure that all ions start the flight at almost the same time. Still the ions have different kinetic energies depending on their positions and linear velocities at the push and would spread spatial, but implementation of a single ion mirror can overcome this problem [Chernushevich *et al.*, 2001]. Ions with the same  $m/z$  but different kinetic energies penetrate the reflector to different depths thus compensating their energy discrepancy and therefore the difference in velocity which improves the mass resolution.

#### 2.1.2.4 Fourier transform ion cyclotron resonance mass spectrometry

FT-ICR-MS was first introduced 1974 by Comisarow and Marshall [Comisarow *et al.*, 1974] who developed conventional ion cyclotron resonance and Fourier transform nuclear magnetic resonance spectroscopy.

This ultra-high mass resolution technique affords a distinction of several thousand ions in a complex sample. Additionally the excellent mass accuracy below 100 ppb allows a calculation of the elemental composition for each detected ion. Even big differences of analyte concentration are detectable due to the high dynamic range of the instruments which altogether makes FT-ICR-MS a powerful tool for non-targeted metabolomics analysis. It offers the possibility to screen rapidly in large biological sample sets for similarities and dissimilarities of metabolites or classes of metabolites [Bhalla *et al.*, 2005].

Modern FT-ICR-MS instruments are mostly hybrid instruments like the in first study (Chapter 3.3.2.3) used Bruker Apex Qe FT-ICR-MS (Figure 17).



**Figure 17: Schematic setting for the Bruker Apex Qe 70 FTMS**  
(Reprinted from [Rajabi *et al.*, 2009] with permission from Elsevier)

The ions generated in the ESI source enter the system through a glass capillary. In the ion optics part they are focused through funnel systems and forwarded to the quadrupole and collision cell part where selective ion filtering and fragmentation can be done if required. Afterwards the ions are accumulated and transferred into the high vacuum part with the ICR detection cell. The ICR cell is embedded in a spatial uniform magnetic field and consists of different electrodes for excitation and detection forming a cylindrical or cubical space in between.

Moving ions under the influence of a magnetic field are subjects to the Lorentz force which affects a moving charge perpendicular to the magnetic field vector and the moving direction. Thus moving ions in the ICR cell are forced into orbits through the magnetic field [Hendrickson *et al.*, 1999]. All ions with the same  $m/z$  ratio also have the same rotation frequency, their ion cyclotron frequency ( $f$ ) which depends only on its  $m/z$  ratio and the magnetic field strength ( $B_0$ ). [Marshall *et al.*, 1998].

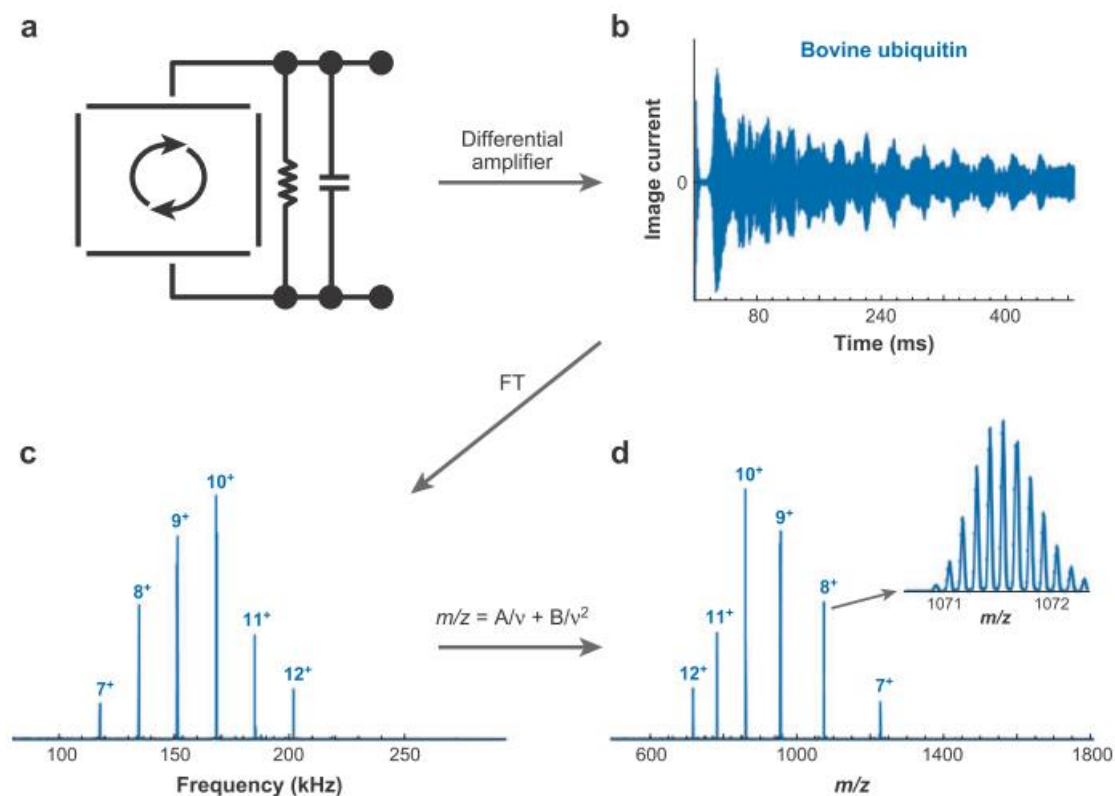
$$f = \frac{e B_0}{2\pi} \times \frac{1}{m/z}$$

( $e$  = electron charge)

For detection the ions first have to be excited with radio frequency (RF) in order to increase their orbit and bring them in phase. Ions with a cyclotron frequency at or near the RF gain energy and increase the radius of their orbits. Thus a RF with a

large bandwidth is necessary for a simultaneous excitation of all ions in the ICR cell. The rotary ions then induce a current in the detection electrodes with a frequency according to their cyclotron frequency and an intensity which corresponds to their number (Figure 18 a). The decay of the superposed currents of every ion species over time (Figure 18 b) is then transformed into the frequency spectrum (Figure 18 c) via a mathematical procedure known as the Fourier transform. The mass spectrum is then generated through a conversion of the frequencies in the corresponding  $m/z$  (Figure 18 d). In a last step, the obtained spectrum can be recalibrated with known peaks to further improve the accuracy.

A more detailed description of the FT-ICR-MS fundamentals is given by a number of excellent publication [Barrow *et al.*, 2005; Guan *et al.*, 1995; Hendrickson *et al.*, 1999; Marshall *et al.*, 1998; Marshall, 2000].



**Figure 18: Generation of a mass spectrum through FT-ICR-MS instruments**

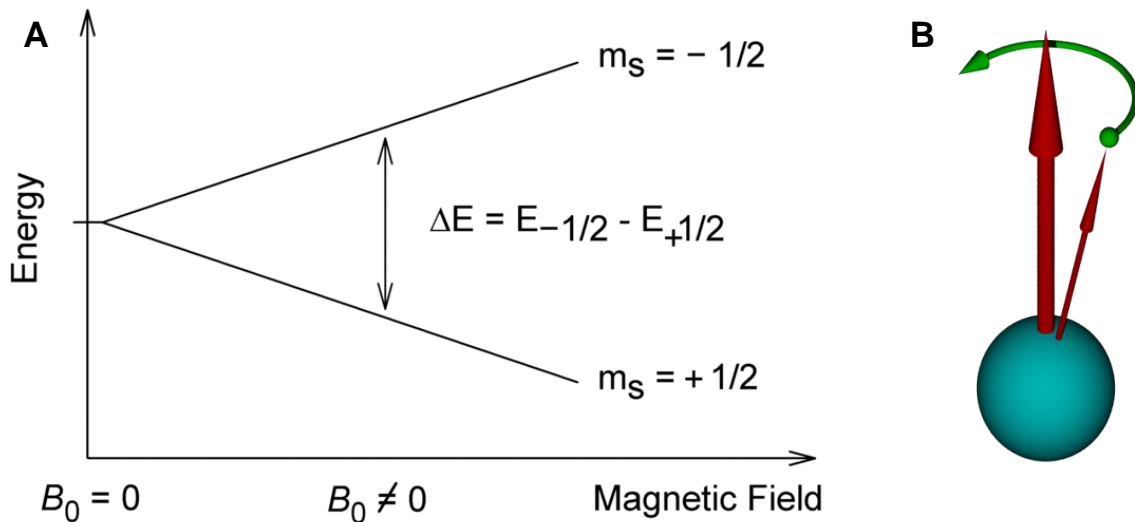
(a) Schematic representation of excited ion cyclotron rotation, (b) time-domain image-current signal from opposed detection electrodes, (c) frequency-domain spectrum obtained by fast Fourier transform of the digitised time-domain signal, and (d) Fourier transform–ion cyclotron resistance  $m/z$  spectrum obtained by calibrated frequency-to- $m/z$  conversion. (Republished with permission of Annual Reviews from [Marshall *et al.*, 2008]; permission conveyed through Copyright Clearance Center, Inc.)

### 2.1.3 Nuclear magnetic resonance spectrometry

The history of nuclear magnetic resonance spectrometry (NMR) starts in late 1945 with the independent detection of nuclear magnetic resonance in liquids and solids by F. Bloch and E.M. Purcell [Freeman, 1995]. In 1966 R.R. Ernst and W.A. Anderson published their new Fourier transform NMR technique which reduces the measurement time dramatically by recording all signals in one measurement instead of scanning through all frequencies of a spectra [Ernst *et al.*, 1966].

NMR is another powerful technique which offers a different view of a sample's content. It also allows a detection of not ionisable compounds if they provide NMR active nuclei's. The local surroundings of these nuclei's can be detected therefore allowing an identification of compounds and structural information of unknown molecules. NMR is also a deep penetrating technique without destroying or consuming the sample [Bothwell *et al.*, 2011]. Compared to MS the sensitivity is very low and only compounds with a high concentration can be detected and quantified [Dunn *et al.*, 2011].

NMR uses the magnetic properties of some nuclei, among them also hydrogen and the  $^{13}\text{C}$  isotope of carbon, for detection. A hydrogen nucleus for example can adopt one of two degenerated spin states so without an external influence the hydrogen nuclei of a sample are equal distributed between them. In NMR the sample is placed inside a homogenous magnetic field which induces an energetic difference dependent on the magnetic field strength between the two spin states (Figure 19).



**Figure 19: Energetic splitting of nuclei spin states in an external magnetic field and Larmor frequency**

A) Energy diagram of nuclei spin states in an external magnetic field. Without an external magnetic field ( $B_0$ ) the two spin states ( $m_s$ ) have the same energy, with increasing field strength the energy difference ( $\Delta E$ ) between them increases. B) Rotation of the actual nuclei spin with its characteristic Larmor frequency around its general direction.

This difference creates a small shift in numbers of nuclei in the direction of the energetic favourable orientation which is important for detection. The spins are rotating around their general direction with a characteristic Larmor frequency, depending on the magnetic field strength and the chemical neighbourhood of the atom. During a measurement all spins are stimulated by a short, broad band radio frequency pulse or a pulse sequence and their relaxation is recorded via a detector coil. After conversion of the spectra from the time domain into the frequency domain via the Fourier transformation the spectra is calibrated with a standard substance. Comparing the areas under a signal of atoms with the same chemical surroundings to the signal of the standard substance gives a relative measure of the concentration of these atoms inside the mixture. To improve the signal to noise ratio, the measurement of a sample is repeated several times and the resulting spectra are summarised. The comparability of different instruments is achieved by conversion of the frequencies into ppm to the external magnetic field strength.

A more detailed description of the NMR principles and experimental setup is given by a number of excellent publication and books [Bothwell *et al.*, 2011; Jacobsen, 2007; Keeler, 2011].

## 2.2 Data analysis approaches

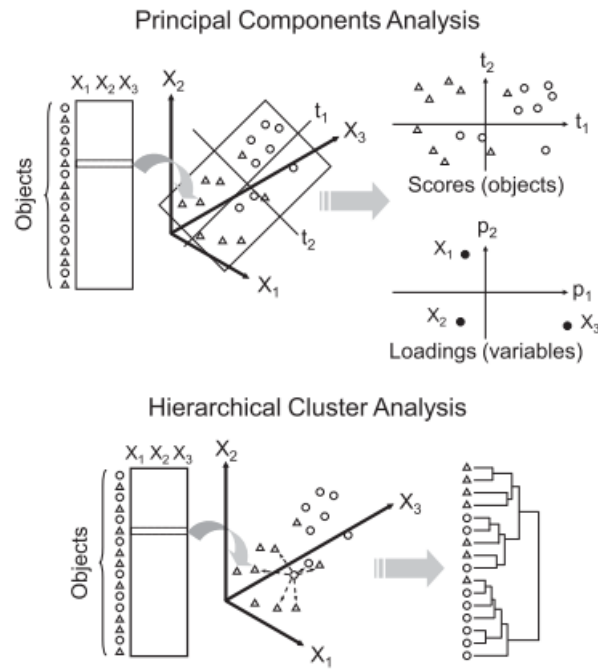
The amount of data recorded in non-targeted metabolomics approaches is usually enormous. Depending on the technique the data matrix of a sample set contains thousands to hundred thousands of signals which have to be processed. Statistical tools have to be applied in order to reduce the amount of data by filtering parts with no or not the desired information, to discover possible patterns within the data set and to visualise data or results [Boccard *et al.*, 2010; Lucio, 2008]. These tools can be divided into two groups, supervised and unsupervised methods. Since the signal intensities of different features can vary over several orders of magnitude often a data pre-treatment is important before statistical modelling. With centering, scaling and transformation of the data the overpowering influence of high abundant features on the resulting model can be suppressed. [Boccard *et al.*, 2010].

### 2.2.1 Unsupervised Analysis

Unsupervised techniques are usually used in a first step for the identification of outliers, the recognition of patterns and reduction of dimensionality of the data. They are used without a predetermined grouping of the samples.

Apart from principle component analysis (PCA) also hierarchical cluster analysis (HCA) was applied in this thesis. PCA is a strong multivariate statistical model for data reduction and visualisation which is commonly used in complex data sets of metabolomics. The model reduces the dimensionality of the data via projection of the data points on a few orthogonal axes, called the principle components, which define the maximal variations in a data set (Figure 20) [Boccard *et al.*, 2010; Lucio, 2008]. After transformation PCA visualise the samples in the scores plot with the principle components as axes (Figure 20). There, samples with comparable metabolite patterns groupe closely together while sample groups with different patterns separate from each other. This allows a fast identification of outliers or other external influences like time [Brereton, 2009].





**Figure 20: PCA and HCA learning principles**

PCA reduces the dimensionality by identification of the maximal variance, the principle components ( $t_1$ ,  $t_2$ ), of a multivariate data set. HCA calculates the similarity of objects by means of distance and linkage functions and clusters them. (Republished with permission of WILEY from [Boccard *et al.*, 2010])

HCA clusters samples according to their similarity by means of distance and linkage functions (Figure 20). The result is usually illustrated in a heat map and a dendrogram which show the expression profile and the corresponding clusters [Butte, 2002].

### 2.2.2 Supervised analysis

Supervised models require a predetermined grouping of the samples to detect distinguishing features. One of them is the orthogonal partial least square discriminant analysis (OPLS-DA) an extension of the partial least square discriminant analysis which can be understood as an extension of the PCA. Compared to PCA the partial least square discriminant analysis allows a group separation even if the variability between the groups is smaller than within the groups [Barker *et al.*, 2003]. In this supervised method the latent variables ( $X$ ), the observations, are calculated in purpose to gain a good correlation with the  $Y$  block, the given class information, via maximizing the covariance between  $X$  and  $Y$ . Thus PLS-DA extracts observations which are significantly different between the groups [Lucio, 2008]. A drawback of this model is the tendency to over-fitting, a modelling error which describes random noise

as result. Thus an index the performance of the model,  $Q^2$  (goodness of prediction), has to be calculated with a cross-validation tests.  $Q^2$  can range from 0 (no fit) to 1 (perfect fit) whereas values  $>0.5$  are desired for a good model and values  $>0.9$  are seen as excellent [Eriksson *et al.*, 2006]. To confirm the model validity and give a measure of statistical significance (p-value) a permutation test [Szymanska *et al.*, 2012] or a cross-validation analysis of variance (CV-ANOVA) [Eriksson *et al.*, 2008] can be performed. In the permutation test the p-value is obtained by repeated random permutation of the samples class affiliations, recalculation of the model and its performance. CV-ANOVA compares the variation around the global average to the models cross-validated predictive residuals and reports the p-value for the null hypothesis of equal residuals of the two models.

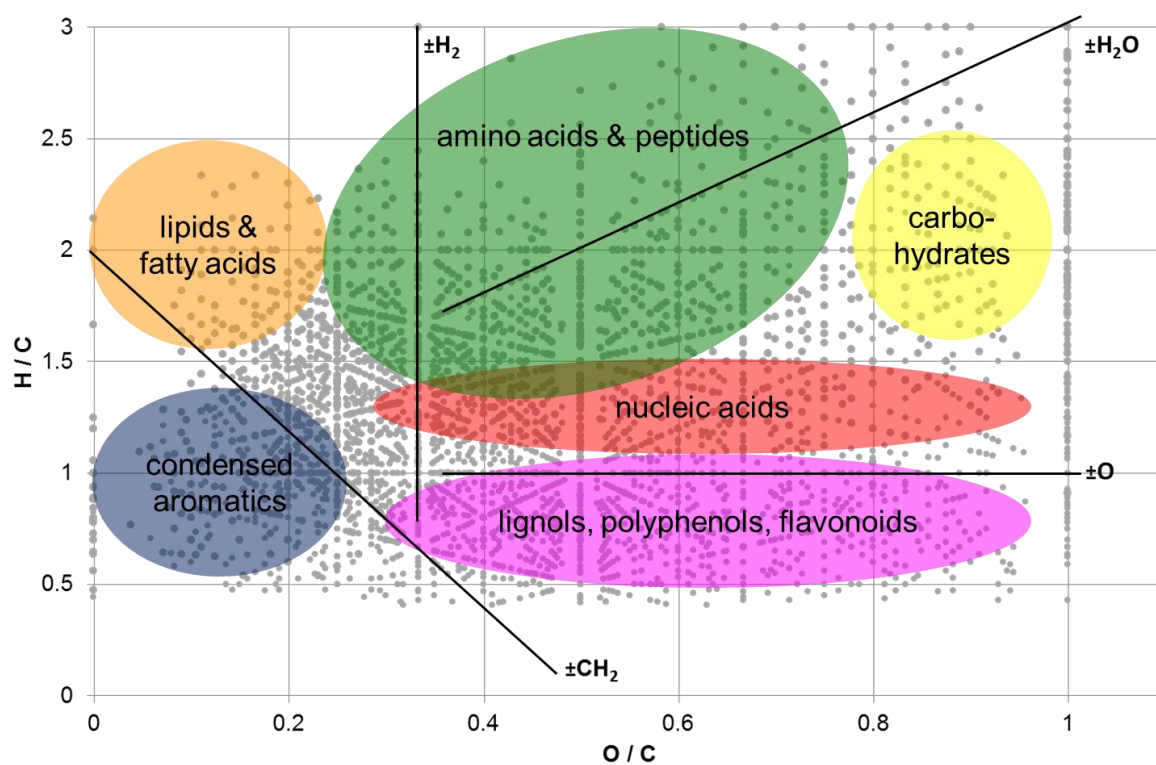
### 2.2.3 Network calculations

For the network calculation an in-house written Matlab program based on an undirected graph was used [Tziotis *et al.*, 2011]. There the nodes equals exact masses from the FT-ICR-MS measurements and the connecting lines, called edges, represent predetermined mass differences according to common biochemical transformations, like addition of an amino acid. A highly connected mass network and sub-networks are generated by a polynomial-time algorithm which compares the mass differences between the uploaded detected masses with the given ones from the biological reactions. Then, based on a single known elemental formula in a network the formulas for the other nodes are calculated through the biological transformations just by following the edges. The number of edges encountering a mass refers to its degree of connectivity. For visualisation nodes with low connectivity are arranged in the periphery while masses in the centre of the network are highly connected. In metabolic networks highly connected masses resemble important precursors, intermediates or end products that are involved in several pathways. Apart from identification of known pathways and also new metabolites whole new pathways can be discovered by following the edges and therefore the corresponding biotransformations.

### 2.2.4 Van Krevelen plot

Another visualisation technique for high complex data matrices, the Van-Krevelen-plot, was used for a first overview. Initially introduced for the visualisation of the composition of coal the Van Krevelen diagram is now widely used in naturally-occurring petroleum, crude oil and natural organic matter research [Hertkorn *et al.*, 2008; Kim *et al.*, 2003]. However it is also very useful in dealing with complex metabolomics data as reviewed by Ohta *et al.* [Ohta *et al.*, 2010].

In the Van Krevelen diagram the hydrogen carbon ration (H/C) of compounds is plotted against the oxygen carbon ration (Figure 21). Since metabolite classes are specified by different molecular formula they conglomerate in particular areas of the plot, which allows a first glance at the samples composition. Even though compounds in this area don't have to be in this metabolic class because different metabolites can have the same elemental composition. In the plot distinct lines can be observed (Figure 21) which can indicate alterations of molecules or the sample but educts-product interpretation should be judged carefully as the two metabolites could be completely independent with hundreds of mass units in between [Reemtsma, 2009].



**Figure 21: Van Krevelen diagram of human urine FT-ICR-MS data and its expressive power**

The Van Krevelen diagram was created with the FT-ICR-MS data of the TULIP study (Chapter 3). The different metabolite classes focus in distinct regions of the plot. Chemical alterations like hydrogenations (+ H<sub>2</sub>), condensations (- H<sub>2</sub>O) or oxidations (+ O) are aligned in a straight line. (Adapted with permission from Kim [Kim *et al.*, 2003] and Hertkorn [Hertkorn *et al.*, 2008], Copyright American Chemical Society and [Roullier-Gall *et al.*, 2014], Creative Commons Attribution (CC BY) license)

### **3 Identification of new prognostic biomarkers within the pre-diabetes study**

#### **3.1 Introduction**

An abnormally elevated urinary albumin level is called microalbuminuria (MA). Likely one fifth to one third of diabetes mellitus patients have MA and therefore an increased risk to develop nephropathy leading to kidney failure as a long term consequence [Ruilope *et al.*, 2010; Tobe *et al.*, 2002]. Besides the risk to develop nephropathy also cardiovascular problems are associated with MA [Schmieder *et al.*, 2012; Weir, 2007]. Therefore MA is an important sign for the early diagnosis of DN and screening for it is recommended by the National Kidney Foundation. A problem is the unreliability of current dipstick tests and therefore the need of relatively expensive laboratory tests like albumin-creatinine ratio assays [Nagrebetsky *et al.*, 2012]. Also, some years ago it was discovered that albumin is excreted as a complex mixture of immunoreactive species and brake down products of albumin lead to increased false negative results [Comper *et al.*, 2005]. Thus new biomarkers for a cheap, quick and reliable detection of MA and therefore the assessment of the risk of developing nephropathy are needed.

The human metabolom mirrors both the genetic background as well as the environmental ascendancies. Additionally some metabolites themselves play important roles in essential reactions and regulatory functions of an organism. For targeted approaches a basic idea or hypothesis of the role of a metabolite in development of MA and a further progression to DN is needed. Thus non-targeted metabolomics studies with their attempt to get a holistic picture of the phenotypes are offering an excellent opportunity to identify new biomarkers for the detection and pathological understanding of MA. Comparing the phenotypes of usually two groups, healthy controls and MA patients, this approach offers the possibility to identify new significantly different metabolites which are unsuspected or even yet unknown.

Recent publications of non-targeted metabolomics approaches conducted in the FinnDiane study for type 1 diabetes patients with normal and increased albumin excretion rates identified different novel biomarkers. The NMR analysis of serum from individuals with normal and macroalbuminuria by Mäkinen et al [Makinen *et al.*, 2012]

identified sphingomyelin, high serum triglycerides and fatty acids concentration as well as a shifted very low density lipoprotein and high density lipoprotein ratio. Van der Kloet et al [van der Kloet *et al.*, 2012] compared urine samples of patients without and patients with progression from normal to MA and identified several metabolites from three classes, acyl-carnitines, acyl-glycines and compounds from the tryptophan metabolism.

In T2D patients several studies associated oxidative stress with the progression of MA. Beneath higher blood AGE's, pentosidine and malonyldialdehyde levels were also found to be increased while the concentrations of vitamin E and radical-trapping antioxidants were decreased [Piarulli *et al.*, 2009]. Another study found a correlation between MA onset and increased total fatty acid, homocystein and H<sub>2</sub>O<sub>2</sub> concentrations [Kassab *et al.*, 2008]. Still new biomarkers are needed since the mechanism of MA development is not completely understood and an as soon as possible detection of MA can be countered with an appropriate life style intervention to maybe reverse, prevent or at least delay a further progression to DN. Compared to blood and serum parameters, urinary biomarkers offer the possibility to develop fast and easy to use dip stick tests and are thus preferred.

### 3.2 Objective

The metabolic alterations of the human metabolom in urine samples of T2D patients with MA are investigated. The urine samples were collected in the course of a pre-diabetes cohort study [Schafer *et al.*, 2007], the TUEbinger Lifestyle Intervention Program (TULIP), a long term study for obesity reduction and T2D prevention. Within this thesis a small and very carefully selected set of 21 control and 17 urine samples of T2D patients with MA is analysed and discussed. For the sample analysis four different platforms were used in order to maximize the metabolic coverage (Figure 22).

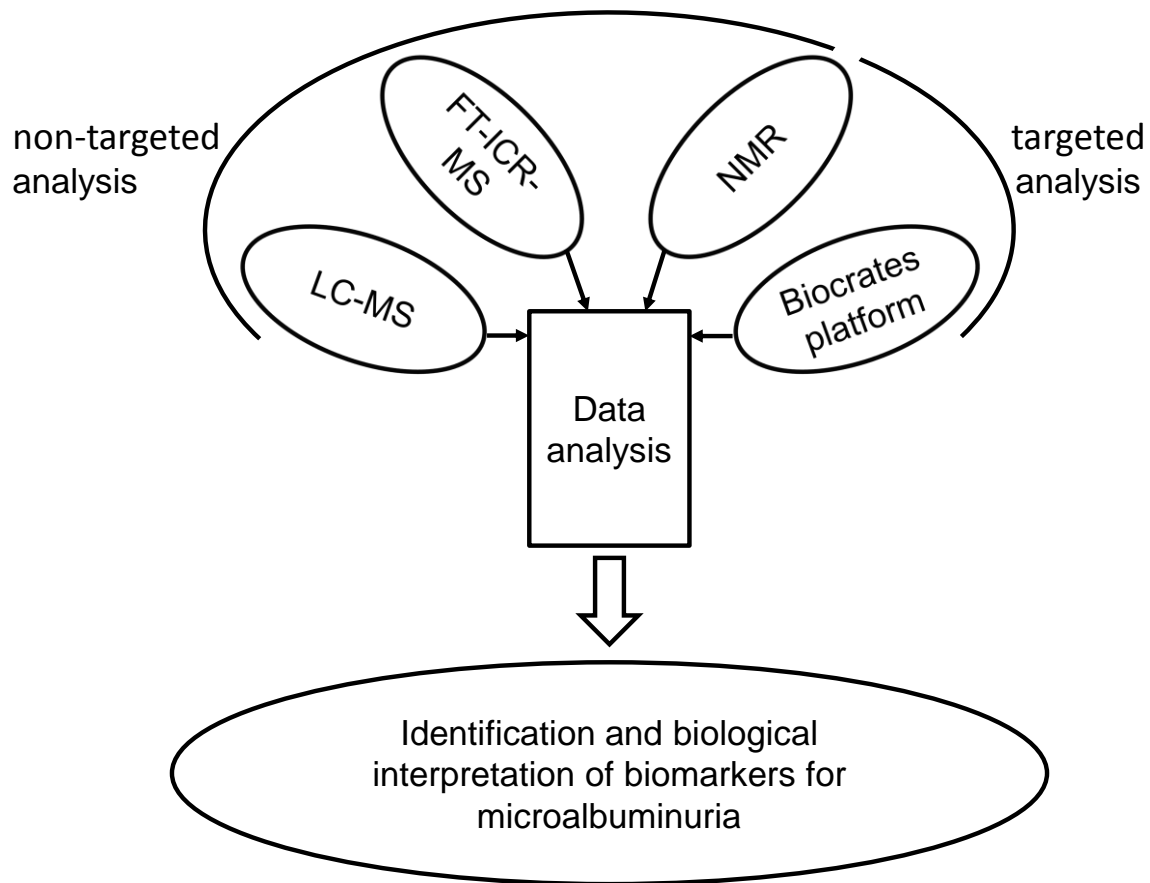


Figure 22: The four used analytical platforms to reach the main goal analysing the pre-diabetes samples

The aim of this study is the identification of new biomarkers for MA which enhance the pathological understanding, increase the reliability of the diagnosis or even allow an early prediction of MA and therefore risk assessment for DN. As basis for reaching the final goal, different objectives had to be met. An adequate sample preparation and methods for LC-MS as well as FT-ICR-MS had to be developed, followed by a data analysis strategy for discovery and annotation of discriminant signals.

### 3.3 Material and Methods

#### 3.3.1 Material

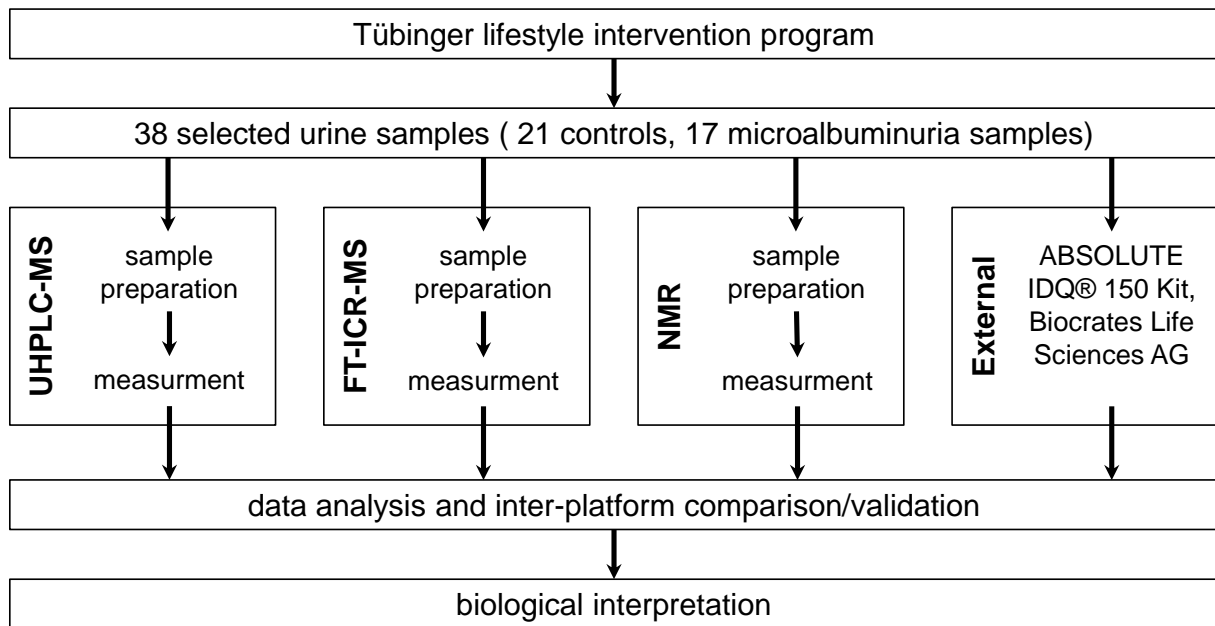
For the UPLC-MS method acetonitrile (ACN), methanol (MeOH), water (H<sub>2</sub>O) and formic acid (FA) have been purchased in ULC/MS grade from Biosolve (Valkenswaard, Netherland). Standard compounds (Table A1) and consumables (Table A2) for method development and quality control (QC) are listed in the appendix (Chapter 5). Methanol and water for the FT-ICR-MS experiments have been purchased in LC-MS quality (trade name: CROMASOLVE<sup>®</sup>) from Fluka<sup>®</sup> Analytical (Sigma-Aldrich, St. Louis, USA).

#### 3.3.2 Methods and methods development

A multi-parallel approach has been built up for the investigation of urine samples from T2D patients with and without MA. The combination of different instruments maintains the advantages of each approach while the drawbacks may be covered by another. This increases the number of detected metabolites for a more detailed view of the metabolic alterations and allows a validation of results in different platforms.

In this study the urine samples were analysed with 3 different platforms: UHPLC-MS, FT-ICR-MS and an external quantification platform from Biocrates Life Sciences AG (Figure 23). Additionally, NMR analysis was applied in order to increase the metabolic coverage of the sample. For the UHPLC-MS and FT-ICR-MS approach methods were developed to optimize the number of detected peaks. The sample preparation, measurement and data analysis for the NMR experiments were performed by Dr. Silke Heinzmann. Additionally the urine samples have been measured with the Biocrates ABSOLUTEIDQ<sup>®</sup> 150 urine Kit, a quantification platform for 158 metabolites of 4 compound classes, by Mrs. Scarpa from the Genom Analysis Center at the Helmholtz Centre Munich.





**Figure 23: Overview of the applied multi-platform approach for biomarker detection**

First, the acquired data was analysed separately for each platform. Then important metabolites and pathways were searched, aligned and compared within the data of the different platforms. The development and applied methods for sample preparation, measurement and data analysis are described in the following section divided according to the used platform.

### 3.3.2.1 Sample handling and storage

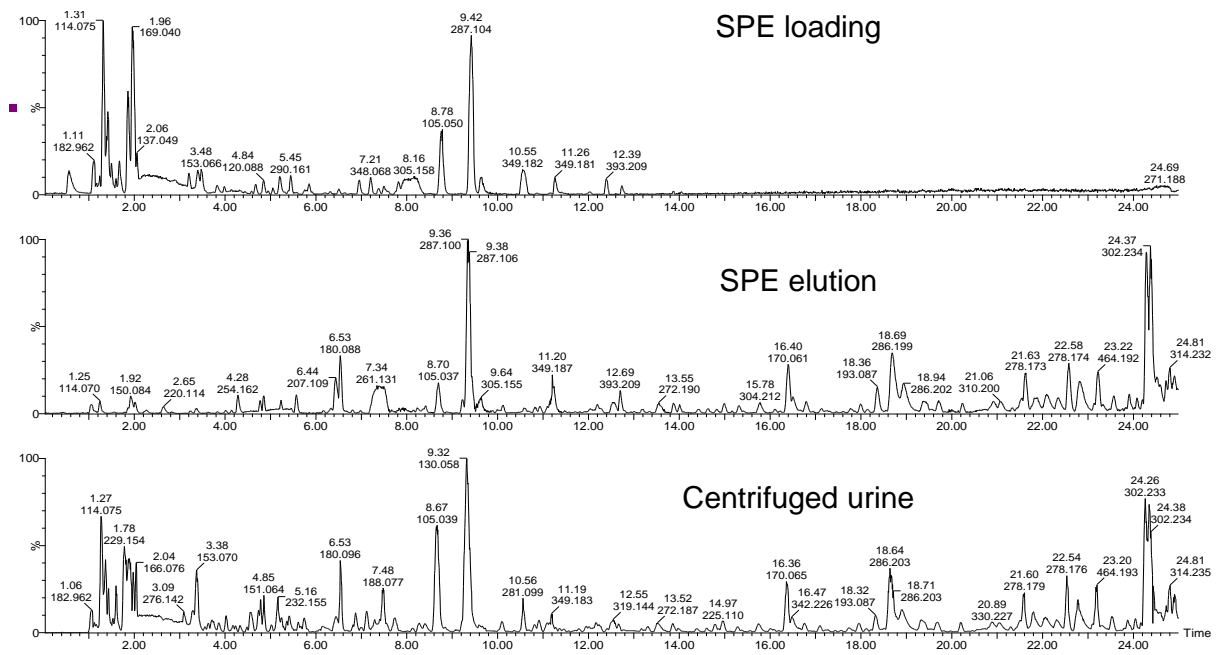
The human urine samples have been collected in the course of the TULIP study at the Tübinger university medical centre. There the 38 samples for this study were carefully selected. The control group consists of 21 urine samples with an albumin creatinine ratio below 10 mg/g while the microalbuminuria group is composed of 17 samples with a ratio above 30 mg/g. More detailed information of the urine samples is given in the appendix (Table A3). Samples were thawed slowly on ice before vortexing them for 30 s and drawing the required amount for experiments.

As QC samples spot urine from eight male subjects was collected, mixed, aliquoted in 2 ml Eppendorf caps and stored at  $-80^{\circ}\text{C}$  until use. For each analysis a fresh QC aliquot was taken and treated like the samples from the TULIP study.

### 3.3.2.2 UHPLC-MS

The UHPLC-MS approach was developed on an ACQUITY UPLC<sup>®</sup> liquid chromatography system coupled to a SYNAPT<sup>™</sup> HDMS<sup>™</sup> mass spectrometer, both from Waters GmbH (Eschborn, Germany). In order to maximize the detected signals an RP approach was developed, optimizing the sample preparation, the used columns, solvents, method and the data analysis.

First different sample preparation procedures, solid phase extraction (SPE), protein precipitation (PPE) and direct injection were compared. Sample preparation in non-targeted approaches is used to remove compounds which suppress other signals, change the analysis like through surface transformation or even affect or destroy used instruments. SPE is also a chromatographic method. While some metabolites are retained by a column material, others are flushed out during the loading of the sample onto the SPE material or the washing step. In the end remaining compounds are eluted from the material with a different solvent or buffer. Since urine contains a lot of salt which suppresses other compounds, an RP SPE was compared with a urine sample where the floating particles were removed by centrifugation. As in Figure 24 can be seen already in the loading step of an SPE a lot of compounds and therefore information is lost. Comparing the SPE elution with the centrifuged urine polar compounds which are eluting in the beginning of an RP approach are missing or detected with reduced intensity (Figure 24).



**Figure 24: Comparison of SPE with centrifuged urine**

The comparison of the total ion chromatograms (TIC) of the SPE loading, SPE elution and just centrifuged urine shows the loss of information during an SPE. Total ion chromatograms from a Grace VisionHT™ C18-HL column, 5 µl injection, Eluent: A: water + 5% MeOH + 0.1% FA B: MeOH; Flow: 0.3 ml/min; Column temperature: 40 °C; Gradient: 0% B in 0-0.5 min, 0-50% B in 0.5-22 min, 50-100% B in 22-25 min, 100 % B in 25-29 min; 0% B in 29-32 min; Ionisation mode: positive ESI

Also with PPE, where mainly proteins are removed some metabolites are lost during the process (Appendix, Figure A1). In PPE proteins are removed from the sample by adding extend of an organic solvent until the proteins lose their solubility and precipitate. During RP, proteins can be absorbed onto the surface of the stationary phase or precipitate. The absorbed proteins alter the stationary phase surface slowly which decreases the column capacity and can also change the retention behaviour of other compounds leading to peak broadening or even retention time shifts [Hagestam *et al.*, 1985]. Proteins also affect the ionisation and are able to suppress or enhance signals of other metabolites [Polson *et al.*, 2003].

Still with every additional step in sample preparation metabolites and therefore information is lost. Thus, only centrifuged urine was used in this approach. To prevent a column blockage or surface alterations of the column material a guard column was added. The guard column is a short pre-column containing the same stationary phase. Thus it has no big effect on the system pressure and the separation but it absorbs proteins and filters precipitates which would otherwise accumulate on the analytical column. If the system pressure increases too much or an alteration in

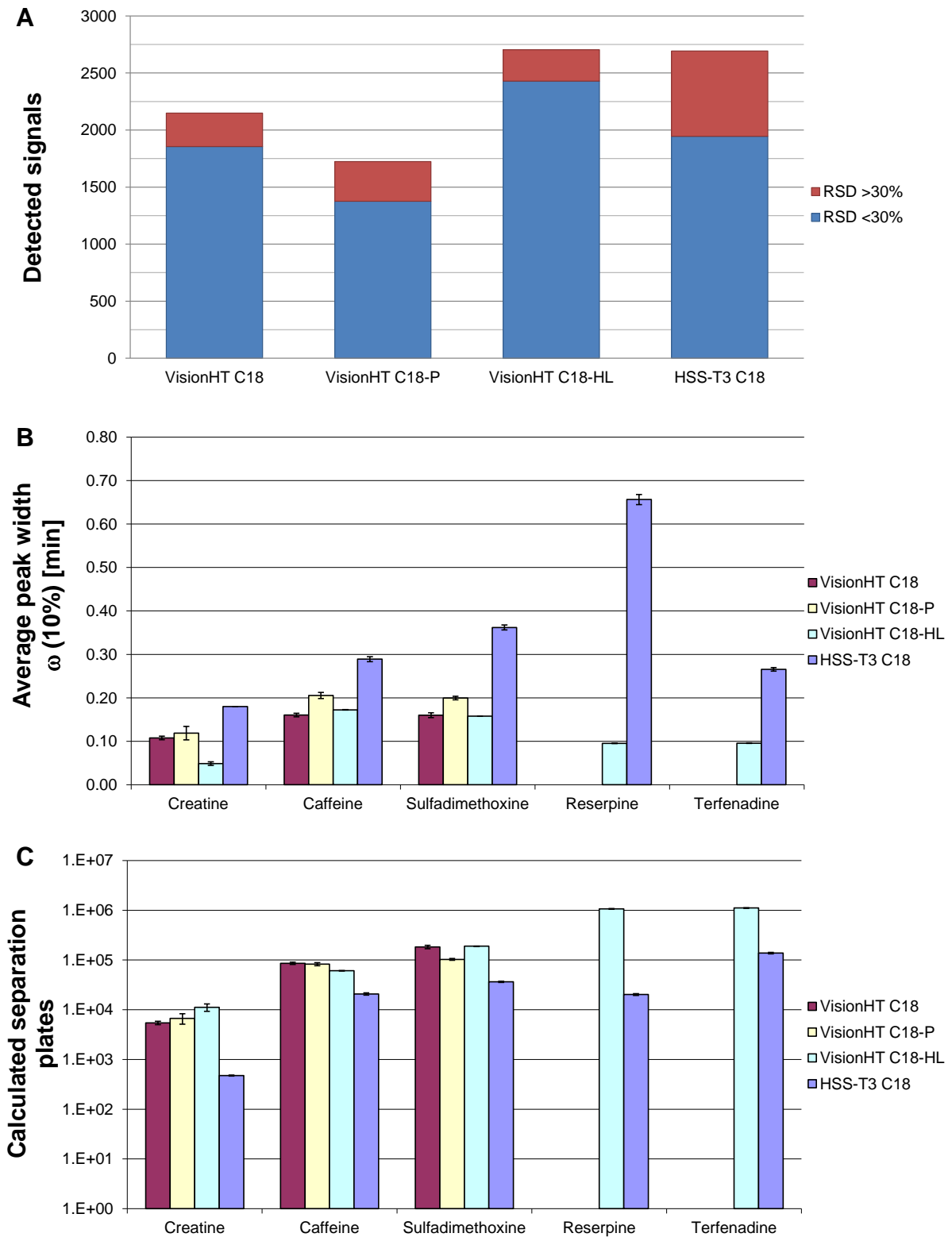
the QC chromatograms is observed only the guard column has to be exchanged and not the whole analytical column.

The optimisation of the RP was started with the selection of a water-methanol gradient for the comparison of four different analytical columns. MeOH grants a better elution profile of the mainly hydrophilic metabolites in urine and additionally supports the ionisation. Therefore a much higher amount of signals in positive (ESI+) and negative (ESI-) ESI could be detected compared to ACN as elution solvent (Appendix, Figure A2). Four different C18 columns were compared (Table 3).

Column	Dimension [mm]	Particle size [µm]	Carbon load	End capped	max. pressure [bar]	flow [ml/min]
Grace <sup>®</sup> VisionHT <sup>™</sup> C18	2,0x150	1,5	6%	X	800	0.200
Grace <sup>®</sup> VisionHT <sup>™</sup> C18-P	2,0x150	1,5	5%	-	800	0.200
Grace <sup>®</sup> VisionHT <sup>™</sup> C18-HL	2,0x150	1,5	10%	X	800	0.250
Aquity <sup>®</sup> UPLC <sup>®</sup> HSS-T3 C18	1,0x150	1,8	11%	X	1000	0.120

**Table 3: Characteristics and applied flow rate of the tested RP columns**

Each column was tested with four standard and three urine injections with 5 µl using the same gradient (Appendix, Table A4) and a column temperature of 40 °C but different flow rates which were set to reach a pressure near the maximum allowed one. The Grace<sup>®</sup> VisionHT<sup>™</sup> C18-HL and the Aquity<sup>®</sup> UPLC<sup>®</sup> HSS-T3 C18 performed best but the former showed slightly more features and more important less variability in the detected peak areas over the three urine injections (Figure 25 A).



**Figure 25: Comparison of the detected signals in urine with the different tested columns**

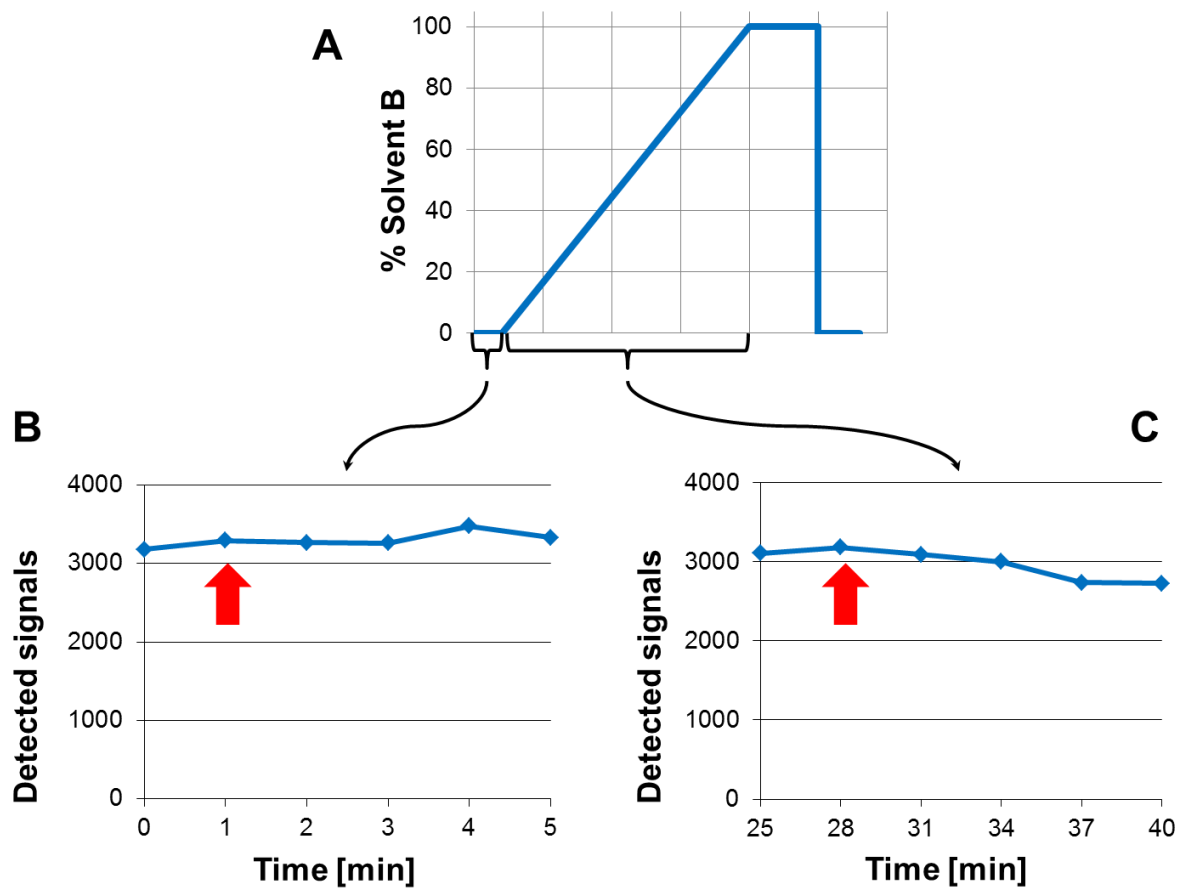
A) Detected signals with a relative standard deviation (RSD) of the peak area below 30% over the three injections are blue while the red bar shows the ones above 30% RSD. B) Comparison of the average peak width of 4 standard mixture injections at 10% peak height. The reserpine and terfenadine peak were only detectable with the C) Theoretically calculated separation plate high for the five standards

Also the peak width at 10% peak height and the number of theoretical separation plates for each standard were compared. The number of plates ( $N$ ) were estimated with the retention time ( $RT$ ) and the peak width at 10% peak height ( $\omega_b$ ) using the following equation [Poole, 2003]:

$$N = 16 \times \left( \frac{RT}{\omega_b} \right)^2$$

Thus the Grace<sup>®</sup> VisionHT<sup>™</sup> C18-HL has been used for the analysis of the TULIP samples.

The next step has been the optimisation of the chromatographic method. To prevent unexpected thermal degradations of and other reactions between metabolites of a sample during the chromatographic run the column temperature was decreased to 35 °C to be below the physiological temperature. The injection volume of urine was 5 µl since in a pre-test comparing 3 µl, 5 µl and 10 µl injections, most signals were detected with a 5 µl injection. Other injection parameters are listed in the appendix (Appendix, Table A5). As gradient a linear gradient from 0 to 100 % of the eluent was chosen since it is more reproducible and more signals could be detected compared to the stepwise gradient which was used for the column testing. Figure 26 shows the two steps which were optimised, an initial isocratic elution step after the injection (A) and the length of the gradient (B). Even though an isocratic elution step of 4 min shows the most detected signals (Figure 26, B) a length of 1 min was chosen since with longer isocratic times especially peaks at the end of the gradient have been getting broader until some were detected double. The detected features with different gradient times reach a maximum at 28 min (Figure 26, C) thus this time was chosen for the final method.



**Figure 26: Development of the chromatographic method by means of detected signals**

A) Principle shape of a linear gradient method B) Detected peaks for different isocratic elution length. A length of 1 min was chosen for the final method as peak broadening from 4 min on caused double annotations of some peaks C) Detected signals for different gradient length. A 28 min Gradient was selected for final method.

The ToF parameters were optimised with direct injection of diluted (1 equivalent urine + 4 equivalents water) urine with a flow of 0.250 ml/min. Six random selected mass peaks were observed for every ionisation mode to get the parameters with the highest average ion count (Appendix, Table A6).

According to the optimisations RP UHPLC-MS analysis of the TULIP urine samples were performed applying these conditions:

Guard column: Grace<sup>®</sup> VisionHT<sup>™</sup> C18-HL; 2.0x5.0 mm; 1.5 µm particles

Column: Grace<sup>®</sup> VisionHT<sup>™</sup> C18-HL; 2.0x150 mm; 1.5 µm particles

Temperature: 35°C

Injection volume: 5 µl

Solvent A: Water; 5% MeOH; 0.1% FA

Solvent B: MeOH; 0.1% FA

Flow rate: 0.250 ml/min

Gradient:

Time [min]	% Solvent B
0	0
1	0
29	99
32.9	99
33	0
36	0

The samples were thawed on ice, vortexed for 30 s and an aliquot of 150 µl was drawn for the UHPLC-MS experiments. The aliquots were centrifuged with >15000 g for 10 min at 4 °C on a 5804 R centrifuge with an F-45-30-11 rotor from Eppendorf AG (Hamburg, Germany). To 132.7 µl sample supernatant 2.3 µl of a standard mixture containing [13C]-Decanoic acid, Nialamid, Reserpine, Terfenandine and [D<sub>3</sub>]-



Valerylcarnitine (Appendix, Table A7) were added and vortexed for 10 s. The whole sample set was measured three times in randomised batches with one QC urine injection after every 10<sup>th</sup> sample injection for both ionisation modes.

The chromatograms were acquired and automatically calibrated with the lock mass Leucine-Enkephaline by MassLynx software (version 4.1 SCN704, Waters GmbH, Eschborn, Germany). The peak piking and data matrix generation were done by MarkerLynx program (version 4.1 SCN704) from Waters GmbH (Eschborn, Germany). Also the parameters for the peak piking and alignment algorithm were optimised, comparing different settings according to the overall detected peaks, the repeatability of the determined peak area and the double counted peaks due to artifacts. Table A8 in the appendix presents the optimised settings which were used for the generation of the data matrix. There chromatographic peaks have to have a minimum intensity of 100 counts, a mass difference of less than 0.02 Da and retention time difference between the maximum peak intensity of less than 15 s to be recognised as the same signal in the different samples. The data matrix was afterwards exported into Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA). Due to the high variability of urine unlike plasma in LC-MS analysis a normalisation is necessary [Warrack *et al.*, 2009]. The peak areas were normalised to the total peak areas of the detected signals since there has been no 24h urine volume and osmolalities available and additionally the urine creatinine values display a dependency on the excreted albumin (Results and discussion, Chapter 3.4.2.1). Also signals present in less than 6 samples were removed. This reduced data matrix then was used for data analysis.

For MS/MS experiments peaks of interest were analysed within the QC sample using the same chromatographic method. The ions were isolated with the quadropole and then fragmented in the TRAP part (Figure 16) of the ToF applying a ramped fragmentation energy from 5 to 30 eV within each scan. This fragmentation method generates a whole fragmentation pattern showing the isolated “parent” ion as well as all the fragment ions generated with the different ionisation energies. The recorded spectra were than compared with spectra stored in the online data bases METLIN ([metlin.scripps.edu](http://metlin.scripps.edu)) and MassBank ([www.massbank.jp](http://www.massbank.jp)).

### 3.3.2.3 FT-ICR-MS

FT-ICR-MS experiments were acquired on a Bruker APEX Qe with an Apollo II ESI source from Bruker Daltonics (Bremen, Germany) and a 12 Tesla super conducting magnet (magnex scientific Inc., Yarnton, United Kingdom). 200 µl thawed urine sample were diluted with 800 µl MeOH, vortexed for 30 s and then centrifuged for 10 min at >15000 g. The supernatants were taken and kept on ice until analysis. The samples were analysed randomised in ESI+ and ESI- mode by manual injection with a 250 µl Hamilton syringe and a syringe pump applying a flow of 5 µl per minute. The FT-ICR-MS instrument was calibrated prior to analysis with a 1 mg/l arginine solution reaching a mass error below 200 ppb. The data was acquired with 2 MW time domain transient and scans. The parameters were optimised with a focus for the m/z range from 150 to 500 Da (Appendix, Table A9). The resolving power of the obtained spectra is > 400000 at m/z 300 for both modes.

The spectra were elaborated in DataAnalysis 4.0 SP2 (Bruker Daltonics GmbH, Bremen, Germany). All spectra were internally calibrated using a calibration list consisting of masses from the calibrated arginine spectra which are also present in most of the sample spectra. The internal calibration reached a mass error below 400 ppb. The mass lists of each spectrum were generated and exported with a signal-to-noise ratio (S/N) of 4 for both ionisation modes. All lists were aligned with a 1 ppm mass window to generate the data matrix with Matrix generator 0.4 (developed in house, M. Lucio). Since a creatinine normalisation was not possible and all spectra had comparable maximum and summarised intensities, no normalisation was applied. For noise reduction only signals present in at least 4 out of the 38 samples were used for the further data analysis.

### 3.3.2.4 NMR

The whole NMR analysis including the data evaluation was conducted by Dr. Silke Heinzmann. NMR experiments have been executed on a 800 MHz Bruker UltraShield™ Plus NMR spectrometer from Bruker Biospin (Rheinstetten, Germany).

The sample aliquots were thawed slowly over night at 4°C and vortexed for 30 s. After centrifugation for 5 min at 4 °C and 11000 rpm, 140 µl supernatant were added to 70 µl D<sub>2</sub>O phosphate buffer pH 7,4 with 3 mM sodium azid and 1 mM sodium

3-(trimethylsilyl)propanoic acid (TSP) and the mixture was given 5 min time to acclimatize. Thereafter the supernatant of a second centrifugation step of 5 min at 4 °C and 11000 rpm was filled in 3 x 100 mm NMR sample tubes for measurement.

#### 3.3.2.5 External Biocrates quantitative metabolomics platform

The samples were delivered on dry ice to Mrs. Scarpa at the Genome Analysis Center at the Helmholtz Centre Munich where the ABSOLUTEIDQ<sup>®</sup> 150 Kit from Biocrates Life Sciences AG was performed. Sample preparation and measurement have been conducted according to their patent [Ramsay *et al.*, 2007] while the data were analysed within this thesis.

The obtained data was filtered by maintaining only metabolites with at least 4 values in the quantification range. Since it has been not possible to normalize the metabolite concentration to the total intensity or the 24 h total urine volume normalisation to the creatinine value was used knowing there is a correlation between it and MA. Additionally normalisation of the data with the dilution coefficients from the NMR experiments were applied and compared.

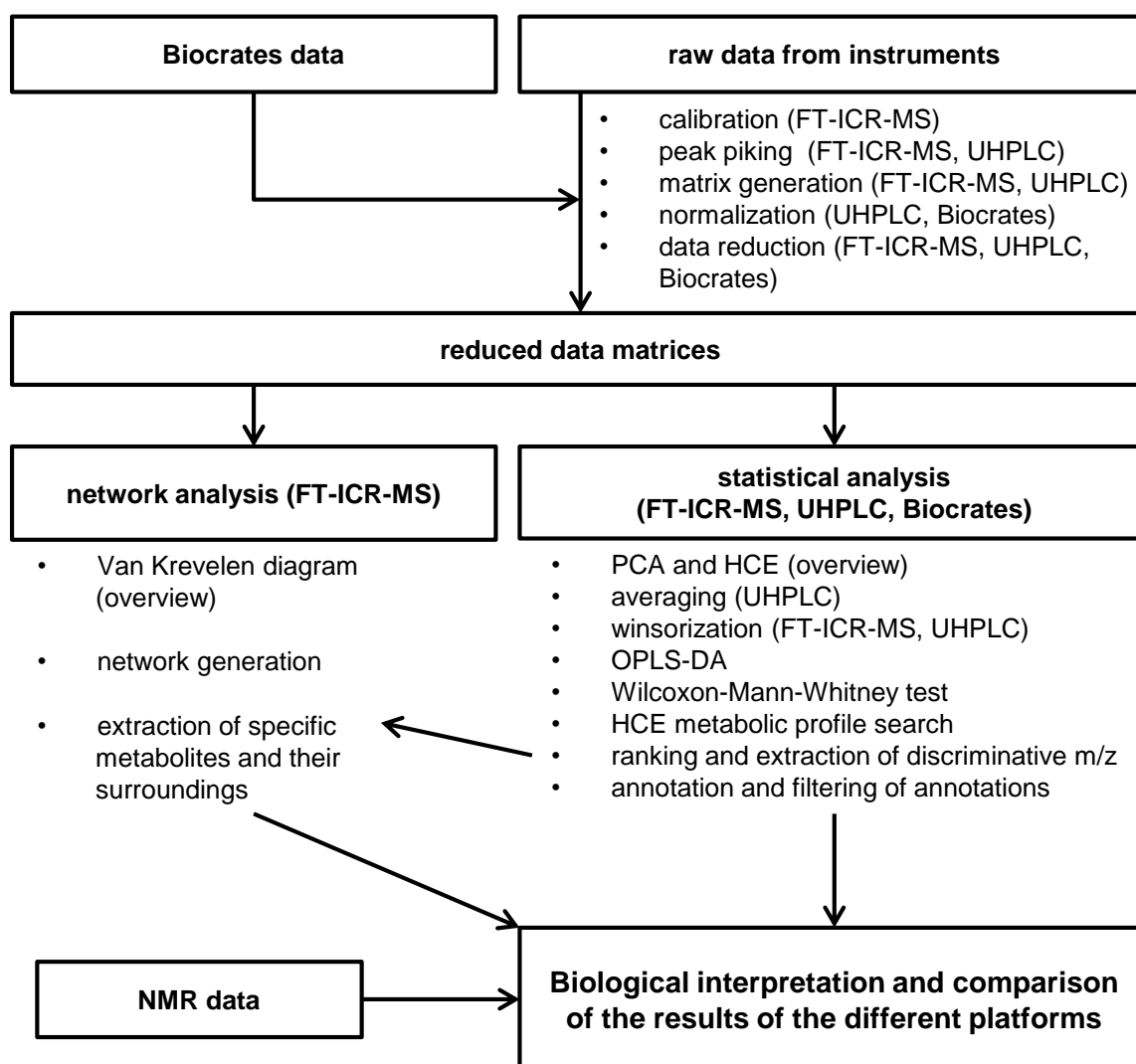
#### 3.3.2.6 Data analysis and comparison

Figure 27 presents the schematic workflow for the data pre-treatment and analysis. After the data matrix generation, normalisation and reduction described in the respective subsections above the data sets of UHPLC, FT-ICR-MS and Biocrates were first analysed separately before the results of all platforms were compared. While the UHPLC and the FT-ICR-MS data needed a complex data analysis approach, a simple univariate test was applied for the quantitative data of the Biocrates platform. The NMR data was analysed by Dr. Silke Heinzmann using only the results for the biological interpretation.

The high mass resolution and accuracy of the FT-ICR-MS allowed the plotting of a Van Krevelen and a Kendrick plot for a first overview of the data. Additionally non-supervised statistical multivariate methods (PCA and HCE) have been applied for visualisation and a first exploration. For the detailed analysis the data of the samples were divided into two groups, control and MA. Three different supervised methods

were applied and the discriminative marker ranked according to the number of tests supporting them. The FT-ICR-MS data was also analysed with a metabolic network approach.

The discriminative masses were annotated and then compared within all the applied platforms. In the end all important metabolites were merged for the biological interpretation.



**Figure 27: Schematic workflow of the data treatment and analysis procedure**

The diagram shows the principal steps for. Some procedures were only applied to specific data sets marked in brackets. Contrary to FT-ICR-MS and UHPLC-MS analysis NMR and the Biocrates platform provide quantitative data. Abbreviations: FT, FT-ICR-MS data; UHPLC, UHPLC-ToF-MS data

#### 3.3.2.6.1 Principal component analysis (PCA)

For a first data visualisation, a PCA model was calculated with the original data sets from UHPLC-MS and FT-ICR-MS using SIMCA-P version 12.0 (Umetrics, Umea, Sweden). Therefore the data was scaled using univariate scaling. The multivariate model transforms the complex data into view principle components, orthogonal variables describing the main variances, and plots the samples with the PC's as axes. Thus naturally occurring patterns, groups of similar samples or outliers can be detected.

#### 3.3.2.6.2 Hierarchical cluster analysis (HCA)

HCA is another non-supervised multivariate analysis for detection of naturally occurring clusters. Samples with similar signal profiles are clustered together with this technique. The HCA was done with the open source software Hierarchical Clustering Explorer (HCE) [Seo, 2002] version 3.5 (Human-Computer Interaction Lab, Maryland, USA) using the Pearson correlation coefficient and average linkage. Before analysis the data was mean centred and scaled using univariate scaling.

#### 3.3.2.6.3 Averaging

The three repetitions of the samples within the UHPLC analysis were averaged since they are just technical replicates but not statistically independent samples. Thus the sample set was imported into Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA) for the averaging. The peak areas of each signal within the replicates were averaged if present in at least two out of the three chromatographic runs; otherwise it was set to zero. After averaging all features present in less than four samples were deleted. The major advantage of measuring technical replicates followed by averaging is the reduction of noise and rare signals within the data set to ease the continuing analysis.

#### 3.3.2.6.4 Winsorisation

Since no valid OPLS-DA model could be generated from the original dataset a transformation had to be applied. Instead of simply excluding all signals with outliers or trimming the outliers, excluding the outliers itself, winsorisation was applied. There minimum and maximum extreme values are replaced by certain percentiles. Thus winsorisation helps to generate stable models and find significant signals but due to its intervention a careful control of the results in the original data is necessary. In this study 80% winsorisation was applied by groups. Values below the 10% quantile were exchanged with the 10% quantile as the values above the 90% quantile were replaced by the 90% quantile. The winsorisation was carried out in Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA).

#### 3.3.2.6.5 Orthogonal partial least square discriminant analysis (OPLS-DA)

After the non-supervised methods an OPLS-DA was applied by Dr. Marianna Lucio using SIMCA-P version 12 (Umetrics, Umea, Sweden). The data was scaled using univariate scaling and the resulting models were evaluated with a CV-ANOVA.

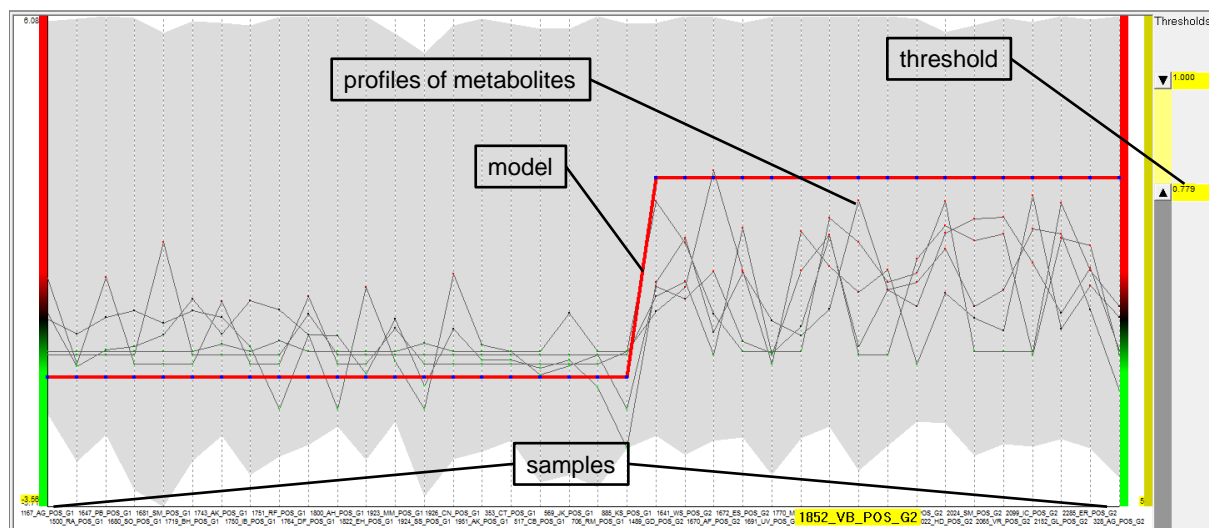
#### 3.3.2.6.6 Wilcoxon, Mann-Whitney test

As second approach to identify significant m/z's a non-parametric Wilcoxon, Mann-Whitney test was used. It doesn't assume a normal distribution, which is an advantage since with the amount of metabolites it is not assured that each of it is normally distributed. The statistical confidence of the null hypothesis, that the means of the two groups are equal, is given by the calculated p-value. Thus all metabolites with p-values below 0.05 were considered as discriminative. For execution of the Wilcoxon, Mann-Whitney test the open source program MultiExperiment Viewer [Saeed *et al.*, 2006] version 4.8.1 (Dana-Farber Cancer Institute, Boston, USA) was applied.

#### 3.3.2.6.7 HCE metabolic profile search

As third method to screen the data for significant hits the metabolic profile search (Figure 28) within the HCE [Seo, 2002] version 3.5 (Human-Computer Interaction

Lab, Maryland, USA) was applied. The data was mean centred and scaled using univariate scaling for the analysis. A model-based search mode with Pearson's correlation coefficient as distance measure as it is shown in Figure 28 was used.



**Figure 28: HCE metabolic profile search**

Within the metabolic profile search metabolites following a defined model (red line) in the samples can be filtered. The threshold defines the accuracy which metabolite profiles have to answer the model with. The figure shows an example profile search within the positive ESI UHPLC data.

Two models were applied for the search, high in MA and low in controls as well as low in MA and high in controls. The threshold for the UHPLC data was set to 0.73 for both ionisation modes. Also, with the positive ESI data of the FT-ICR-MS a threshold of 0.703 was applied while for the negative ionisation the threshold was raised to 0.736 to limit the filtered signals to the 500 most important.

### 3.3.2.6.8 Signal ranking

The signal ranking was applied for the not quantitative UHPLC-MS and FT-ICR-MS datasets. All significant  $m/z$ 's from the OPLS-DA, the Wilcoxon, Mann-Whitney test and the HCE metabolic profile search were combined and ranked according to their number of appearance as significant hits. For the further analysis only signals with at least two out of three hits were used. The ranking was performed with Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA).

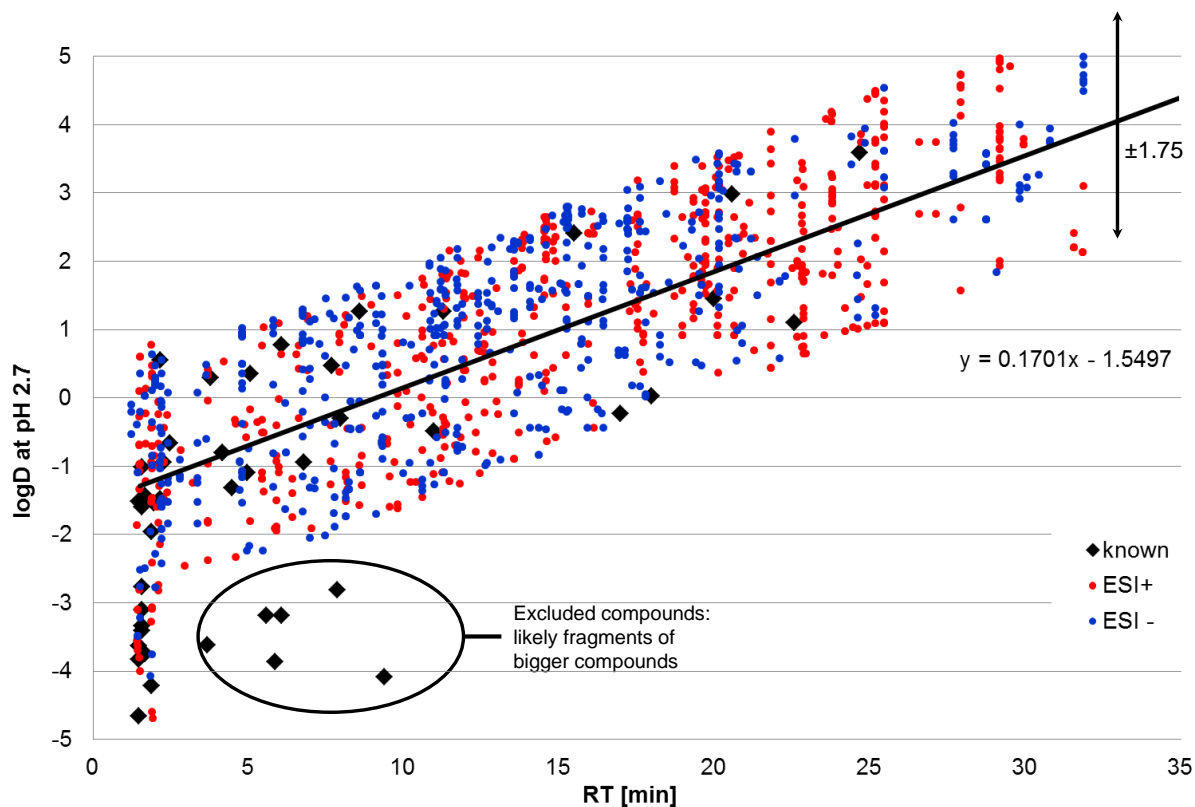
#### 3.3.2.6.9 Mass annotation

MassTRIX version 3 [Suhre *et al.*, 2008; Wagele *et al.*, 2012] (<http://masstrix3.helmholtz-muenchen.de/masstrix3/>) was applied for the mass annotation. The mass lists were uploaded onto the server

All recorded fragmentation spectra were compared with spectra stored in the online data bases METLIN ([metlin.scripps.edu](http://metlin.scripps.edu)) and MassBank ([www.massbank.jp](http://www.massbank.jp)) to identify the fragmented ion.

Since for the UHPLC-MS data a broad mass window of 0.02 Da was used for the annotation, a filter step had to be applied to reduce false positive hits. Therefore the annotations were filtered using the pH dependent distribution coefficient (logD) [Müller, 2013]. The RT of known and annotated compounds were plotted against their logD values calculated with JChem for Excel (Version 5.12.3.966, ChemAxon Ltd, Budapest, Hungary). The logD value equals the pH dependent distribution coefficient. To generate the trend line standards and via MS/MS identified compounds excluding results with a logD below -2 were used (Figure 29). Annotations were accepted if their deviation from the trend was below 1.75 logD units. The first 2 minutes are an exception. There, all compounds with a logD below the maximum acceptance limit were allowed since in this time range all highly polar compounds which are not or only hardly retained elute.





**Figure 29: Filtration of the UHPLC-MS annotations by plotting the logD values against the RT.**

The trend line was generated with known or MS/MS identified compounds excluding all with logD < -2. The limit of accepted diversion from the trend was set to 1.75 logD units (adapted from [Müller, 2013]).

### 3.3.2.6.10 MetaboAnalyst

For the interpretation and first analysis of the annotations of significant signals a web based tool was used. MetaboAnalyst (<http://www.metaboanalyst.ca>) offers various tools for metabolomics data processing of which especially the functional enrichment and the metabolic pathway analysis were used [Xia *et al.*, 2012].

From a given list of annotations it generates a plot which contains the matched pathways, their p-values and their impact values.

### 3.3.2.6.11 Network visualisation and elemental formula calculation

The high accurate FT-ICR-MS data was used to calculate a metabolic transformation network and with it the molecular formulas of the masses. Therefore the inhouse written Matlab program Netcalc [Tziotis *et al.*, 2011] was used. It generates a network from a given mass list and a transformation mass list in which each mass represent a node and an edge between two nodes corresponds to a distinct mass difference from

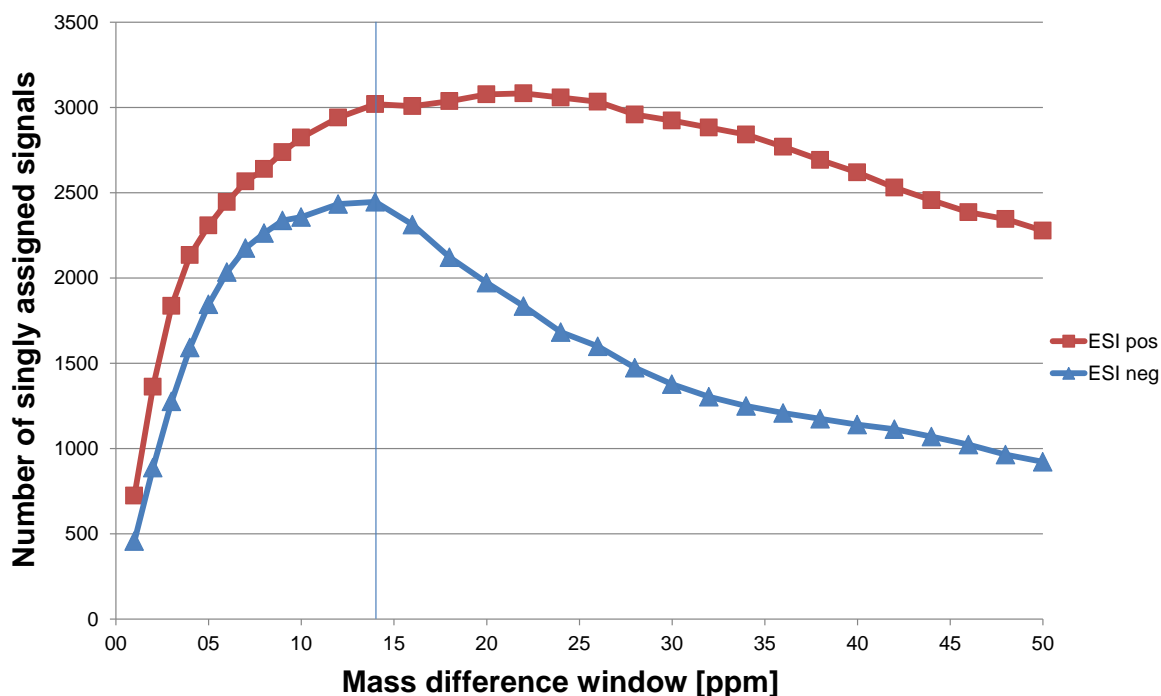
the transformation list. As transformation list mass differences from different biological conversions were used (Appendix, Table A10). As far as one formula in the network is known all other formulas can be calculated through the connections between the masses. For the formula calculation a mass error window of 0.5 ppm was applied.

For the visualisation of the calculated network Cytoscape (version 3.0, <http://www.cytoscape.org/>) was used. It is a powerful open source network visualisation tool [Shannon *et al.*, 2003; Smoot *et al.*, 2011] for the network analysis. There different layout algorithm can be applied to see the region of interest. The applied force directed layout places the nodes with high connectivity in the centre of the network while low connected nodes are placed in the border area.

#### 3.3.2.6.12 Data alignment

Since the mass accuracy of the UHPLC data is low compared to the FT-ICR-MS data an alignment method was developed to search for possible matches between the results of the two approaches. Additionally significant UHPLC signals without a corresponding significant FT-ICR-MS mass were tried to match with non-significant masses and the appropriate annotation.

Therefore the two whole data sets were compared by determining the number of features in the UHPLC data where in a given mass window only a singly assigned mass in the FT-ICR-MS data was found (Figure 30). A 14 ppm mass window performed best for the ESI- data and was a good compromise for the ESI+ data in order to keep the mass window and with it the mismatches low while a decent number of peaks were matched.



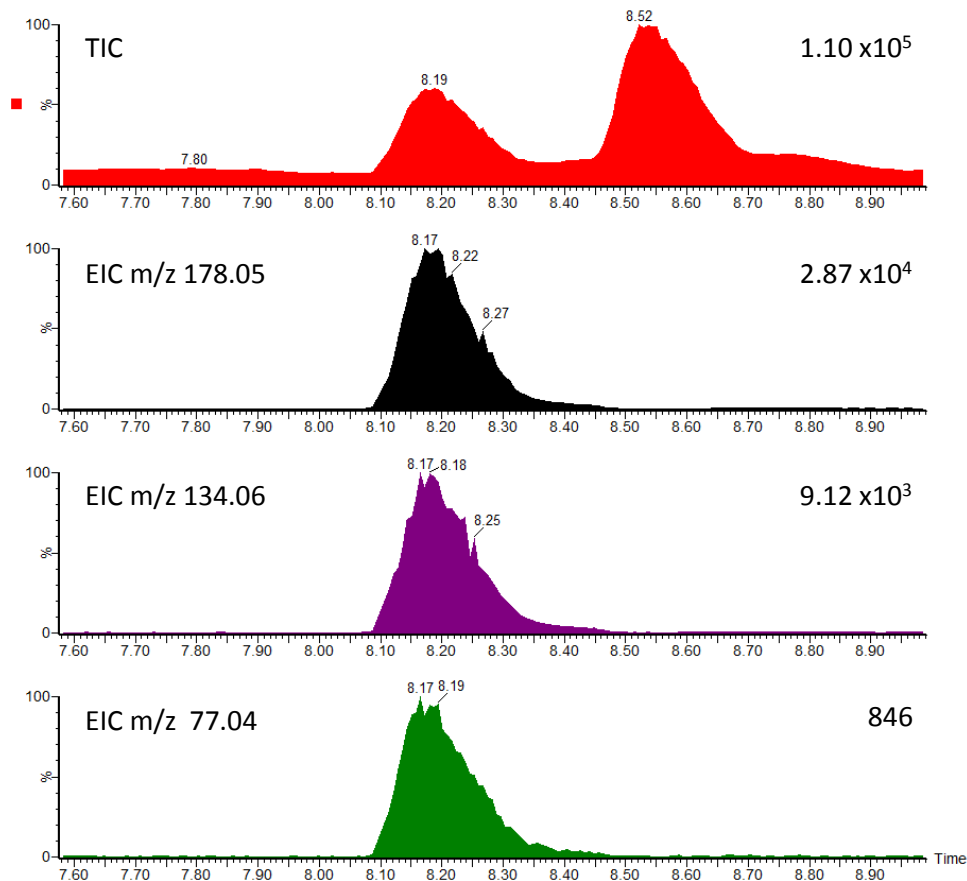
**Figure 30: Determination of the mass difference window for the UHPLC data calibration**

The number of singly assignable FT-ICR-MS masses for the UHPLC masses in a given mass difference window was determined. For the calibration of the UHPLC with FT-ICR-MS masses a 14 ppm mass error window was used.

The alignment of the UHPLC and FT-ICR-MS results was done applying a 14 ppm mass window. Since the same molecule could ionise differently in both approaches, several possible ions were calculated from the high accurate FT-ICR-MS masses to compare them with the UHPLC masses (Appendix, Table A11).

#### 3.3.2.6.13 Correlation analysis

Via acceleration of ions in the MS a given percentage of the ions fragment. Since the percentage is relatively stable over a certain concentration range correlating signals may represent fragments or other species of a compound ion. Thus correlation analysis was applied on the result signals of the LC-MS dataset to identify connection between different mass traces. In order to prohibit misleading assignments of ions from different sources a comparison of the peak shapes was done. Therefore the extracted ion chromatograms using the MassLynx software (version 4.1 SCN704, Waters GmbH, Eschborn, Germany) were compared for confirmation of a connection between the ions (Figure 31).



**Figure 31: Comparison of the peak shapes of three correlating peaks in ESI<sup>-</sup> data**

From the total ion chromatogram (TIC) the three extracted ion chromatograms (EIC) were generated by focusing on mass traces identified through correlation analysis applying a mass window of  $m/z \pm 0.02$ . The peaks were normalized to the highest peak intensity (number on the right side) in the selected window. The matching peak shapes confirming the correlation analysis results and by searching in MS/MS data bases hippuric acid could be identified.

The correlation analysis was done in Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA). Signals in a RT window of  $\pm 0.1$  min around a result mass were screened for a correlation bigger than 0.9 in order to identify possible fragments of the result mass. The matching peak shapes were checked in a couple of samples containing them.

### 3.4 Result and discussion

The result and discussion section has been structured in two main parts (Figure 32).

#### 3.4.1 First data survey

- method performance
- sample composition
- intersection between non-targeted MS platforms

#### 3.4.2 Metabolic changes through microalbuminuria

- results of the pathway analysis
- important pathways with identified metabolites:
  - creatinine pathways
  - arachidonic acid metabolism
  - caffeine metabolism
  - vitamin B<sub>6</sub> pathway
  - amino acids metabolism
  - other confirmed metabolites
- key „unknown“ metabolites

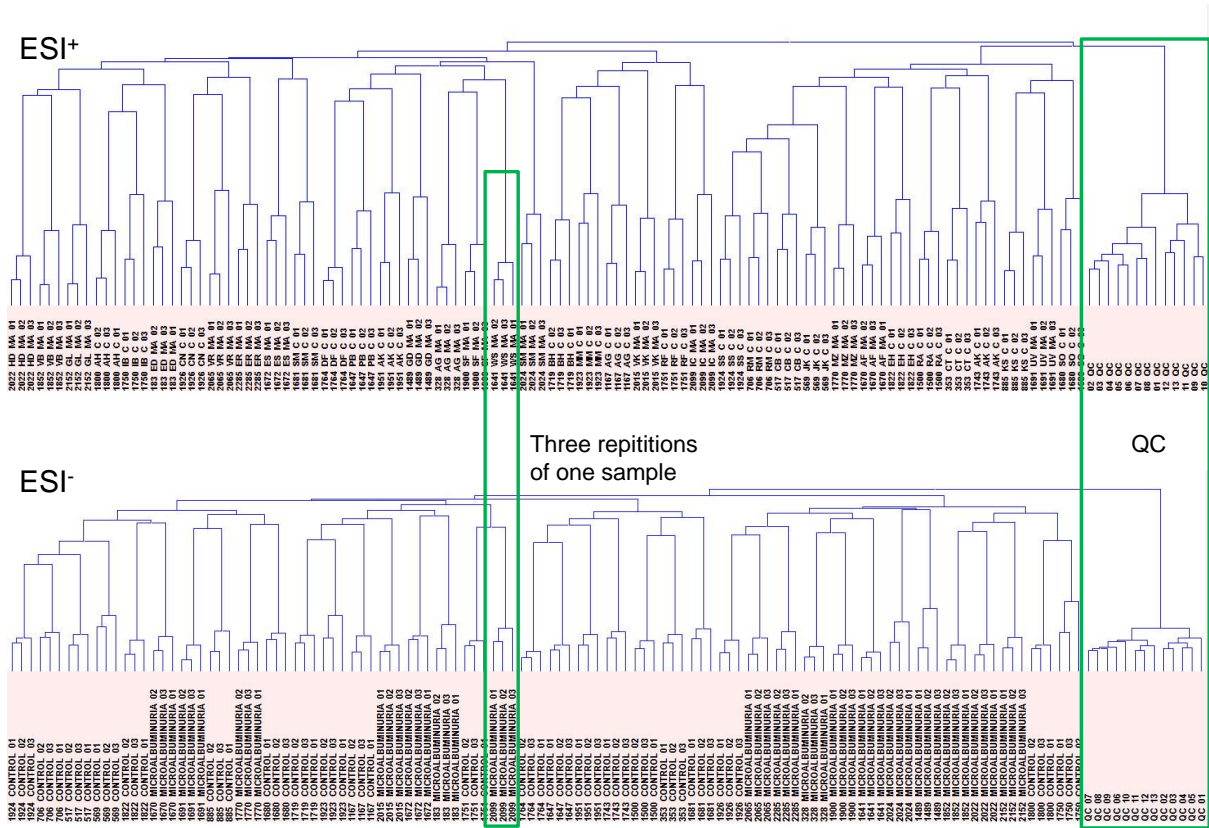
Figure 32: Overview of the result and discussion section

The first part presents an overview of data (chapter 3.4.1) obtained from the UHPLC-MS, FT-ICR-MS and Biocrates platform while the second part documents and discusses the alteration between normal and microalbuminuric T2D patients including the NMR analysis results in detail.

### 3.4.1 First survey of the data

#### 3.4.1.1 Overview

A fast and general overview over the data was created applying HCA and PCA on all data including QC's. Only peaks present in less than 10 % of the measurements were filtered from the data after normalisation. The HCA of the UPLC-MS Data is shown exemplarily (Figure 33).

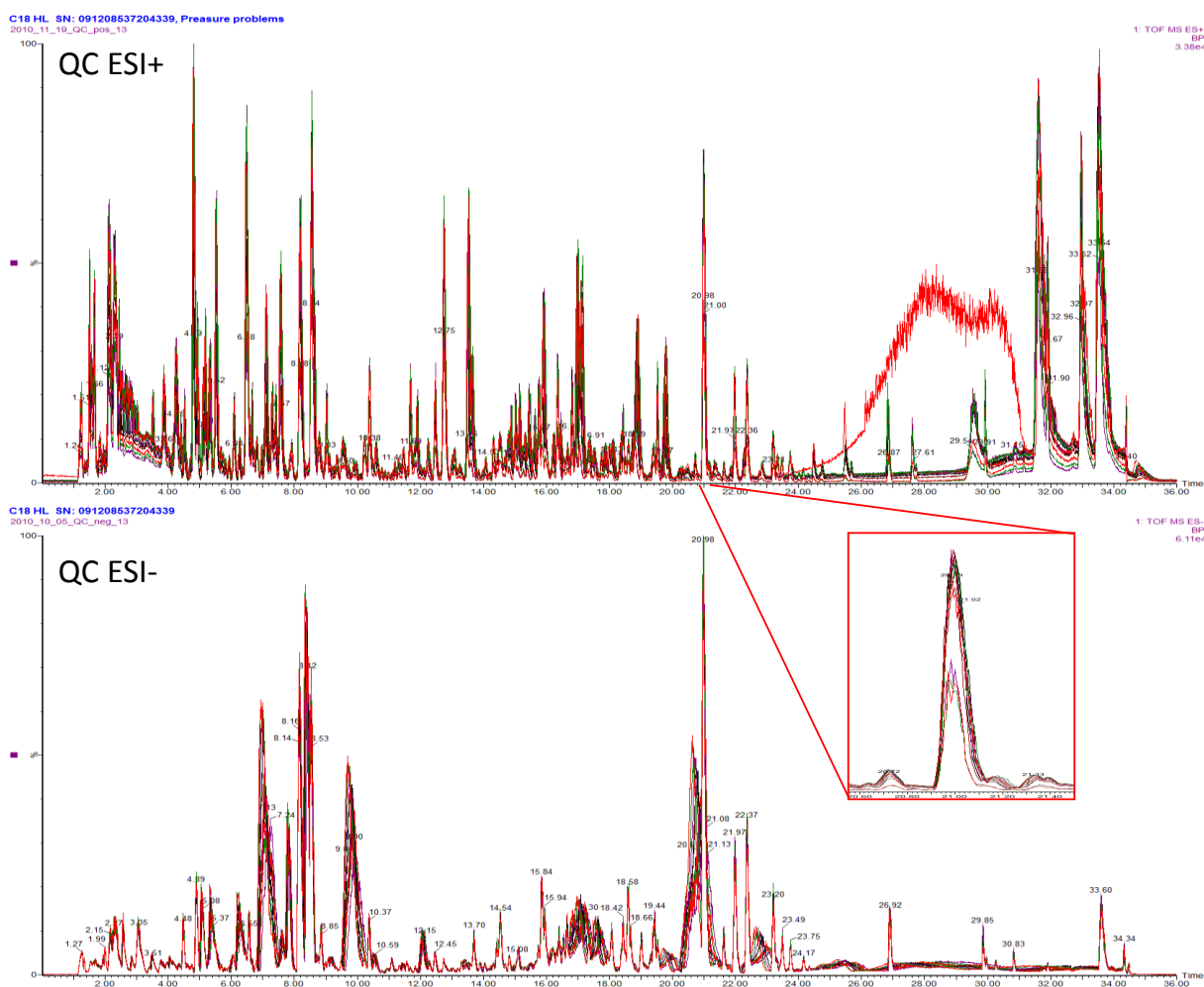


**Figure 33: HCA of the UPLC-MS data from the TULIP study including QC's.**  
 All QC's and each sample triplicate form a small cluster indicating a good reproducibility of the measurements. No separation between the control and the microalbuminuria group could be achieved.

No separation between the control and the microalbuminuria group could be achieved with HCA and PCA. The individual differences between the spot urine samples of the patients are therefore bigger than the metabolic changes caused through microalbuminuria. After the first overview a detailed data analysis was started.

### 3.4.1.2 General performance of the methods

First the reproducibility of the RP chromatography was controlled visually. An overlay of all 13 QC for each UPLC-MS ionisation mode shows the good performance of the developed method (Figure 34). The first QC chromatogram in positive mode showed an unexpected pattern in the end of the gradient which was found to have no bearing on the sample runs comparing the directly following ones with their repetitions.



**Figure 34: Overlay of the 13 quality control chromatograms for each detection mode**

The overlay of the 13 base peak intensity chromatograms shows a good repeatability. The first QC sample in positive mode shows a different behavior in the end of the gradient so the first sample runs were controlled. The zoomed peak shows the necessity of sample randomisation and averaging since the peak area decreases with time.

The internal standard peaks could be detected in every sample run and the RT shifts were below 10 seconds. The average relative standard deviation of the internal standard peak areas within the three replicates were below 25 % in positive and 10 % for the negative mode data revealing the bigger time dependency of the

detection in positive mode. This shows the necessity of repetitions of the sample analysis, randomisation and data averaging of the repetitions. The mass difference between the theoretical mass of the standards and the found mass was below 0.005 Da for positive and negative mode.

After data cleaning and averaging 10078 signals in ESI+ and 6503 in ESI- with 63 % respectively 59 % of them present in at least 50 % of the samples were forwarded to the detailed analysis.

The FT-ICR-MS chromatograms were inspected manually and calibrated. After data extraction and filtering 11151 peaks in positive and 18274 in negative were used for the detailed data analysis. Less than 20 % of the signals were detected in more than 50% of the samples. This can be explained by the higher mass accuracy and resolution of the FT-ICR-MS, the individuality of the 38 patients and the complex interactions between the ions especially with direct injection.

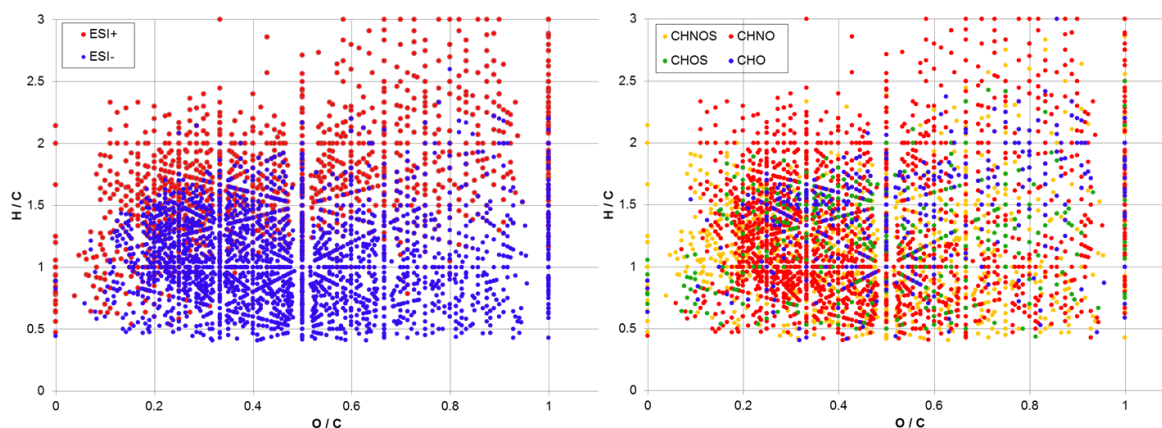
The Biocrates platform quantifies 161 metabolites and delivers additionally 27 parameters like total amounts of metabolite classes and ratios between metabolites. After removing all signals with less than 4 values in at least the semi quantitative range, 124 single metabolites and 7 other parameters were used for the further analysis. Comparing the creatinine values from the Biocrates platform with the ones measured within the TULIP study showed a good correlation of 0.94 excluding one sample as outlier (Appendix; Figure A3).

The NMR spectra were divided into 40109 sectors for the analysis via OPLS-DA. The results showed 14 discriminating signals of which 9 could be annotated with metabolites.

#### 3.4.1.3 Molecular composition of the samples

The high accuracy of the FT-ICR-MS allows a first overview of the sample composition by calculating the chemical formulas of the detected masses and plotting them in a Van Krevelen diagram (Figure 35). The figure illustrates the high complexity of the metabolic composition of urine as well as the need of both ionisation modes to cover the complexity.





**Figure 35: Van Krevelen diagram of the TULIP study FT-ICR-MS data**  
The left diagram

As expected hydrogen rich and oxygen poor metabolites are a minority in urinary samples. The organism oxidises metabolic end products to increase their solubility for an easier transportation through blood and excretion into urine. Therefore the lipid area is less crowded compared to the protein space. The ESI<sup>-</sup> data contains more unsaturated and oxidised species which can be explained by a better stabilisation of carb-anions through conjugation and inductive effects from electronegative atoms like oxygen.

By annotation with MassTriX 9 % of the detected masses in positive and 3 % in negative ionisation mode could be assigned. The calculated elemental compositions through the mass network covered 24% in ESI<sup>+</sup> and 19% in ESI<sup>-</sup> data.

#### 3.4.1.4 Intersection between the different nontargeted MS platforms

The estimation of the mass convergence between the different ionisation modes and the UHPLC and FT-ICR-MS platform has been done by using a 14 ppm mass window to compare the two methods. For the comparison between the positive and negative UHPLC-Data an RT window of 0.25 min was applied. For the comparison ions formed by protonation and deprotonation as well as sodium adducts in positive mode were considered (Figure 36). Only 680 masses are present in every dataset regardless of ionisation or platform which represent more than 10% of the smallest dataset.

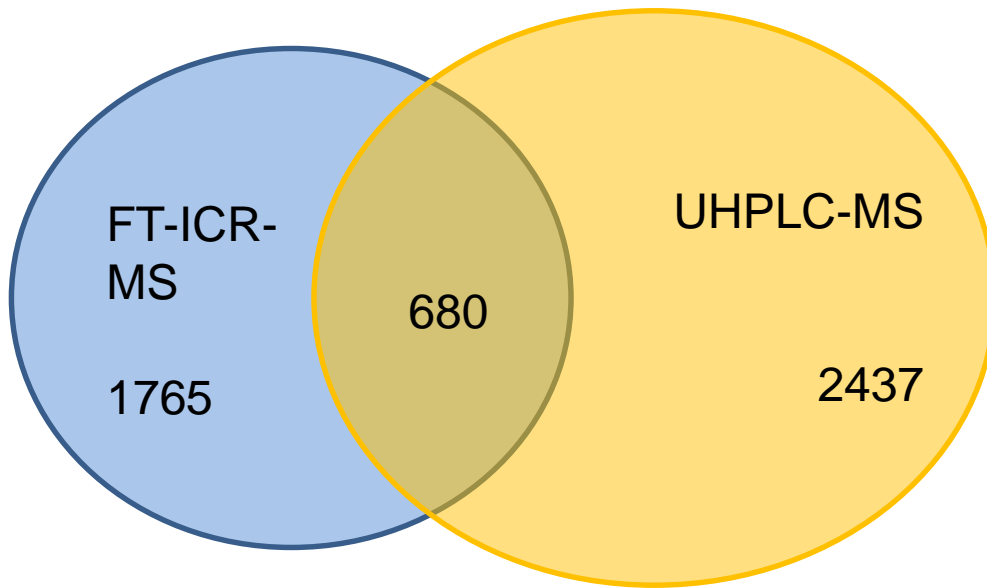
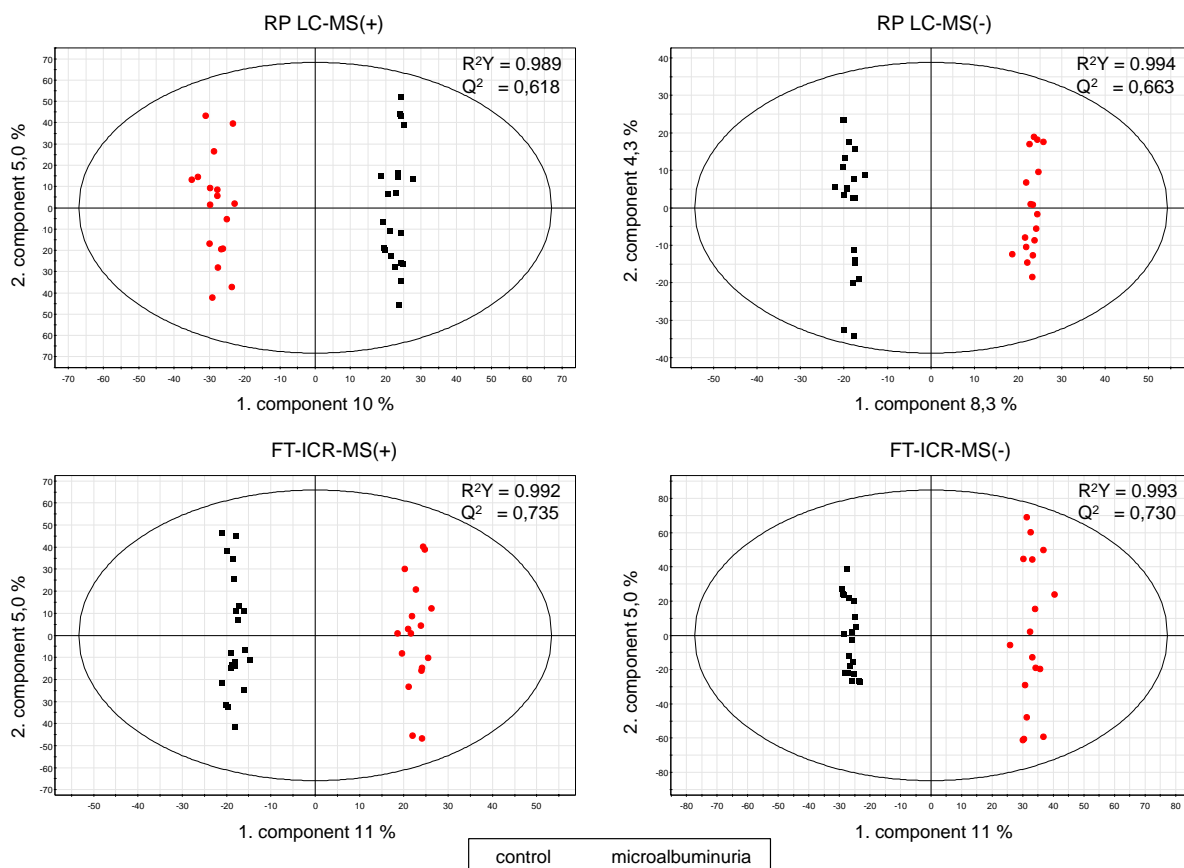


Figure 36: Venn diagram of the mass convergence between FT-ICR- and UHPLC-MS-Data

The big difference between the datasets represents the complexity of the sample and the methods. During ionisation with the different instruments different ions are formed, suppressed and detected due to many effects like in source fragmentation, source geometry or accuracy of the detection method. Also the instruments show a “preferred” ionisation mode. While the ToF detects more masses in ESI<sup>+</sup> mode the FT-ICR-MS does in ESI<sup>-</sup> mode.

### 3.4.2 Metabolic changes through microalbuminuria

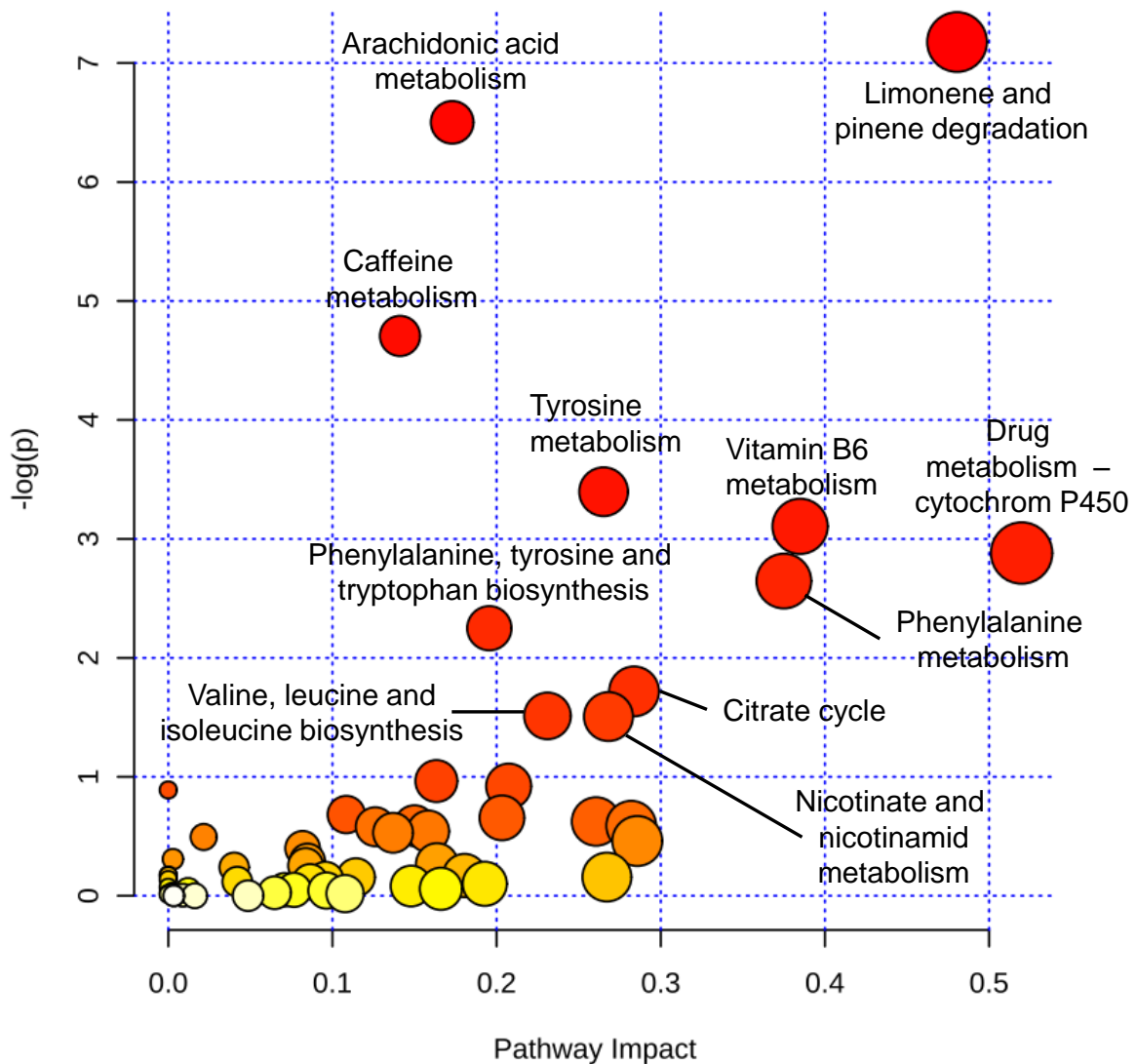
The data overview revealed the big influence of individual differences between the samples. So, supervised multivariate data analysis was used to extract significant masses out of the datasets. Since OPLS-DA was not able to create a valid model with the untreated data a winsorisation step was introduced. After the transformation good OPLS-DA models were derived using univariate scaling (Figure 37).



**Figure 37: Separation of control and microalbuminuric samples with OPLS-DA after winsorisation**

Important variables were extracted from the S-Plot. The data was additionally analysed using the HCE profile search and the non-parametric Wilcoxon, Mann-Whitney test. The results of all three statistic methods were compared and only significant signals which were supported by at least two of them were evaluated further. Thus 404 in positive respectively 312 features in negative ionisation were extracted from the UHPLC-MS data. For the FT-ICR-MS data 413 signals in  $\text{ESI}^+$  and 541  $m/z$  in  $\text{ESI}^-$  resulted from the comparison.

Since for the ToF data a relatively big mass window of 0.02 Da had to be used for peak piking a mass gets many annotations. Therefore a filter step using the logD value and the retention time had to be introduced to reduce the number of false positive annotations. In contrast the annotations of the FT-ICR-MS data are more accurate and can be analysed directly. All annotations through the KEGG database were uploaded and analysed with MetaboAnalyst [Xia *et al.*, 2012]. The results of over-representation tests and pathway topology analysis are pathways which differ between the two groups and are presented in Figure 38.



**Figure 38: Analysis of the MS annotations within the TULIP study applying metabolic enrichment and pathway analysis.**

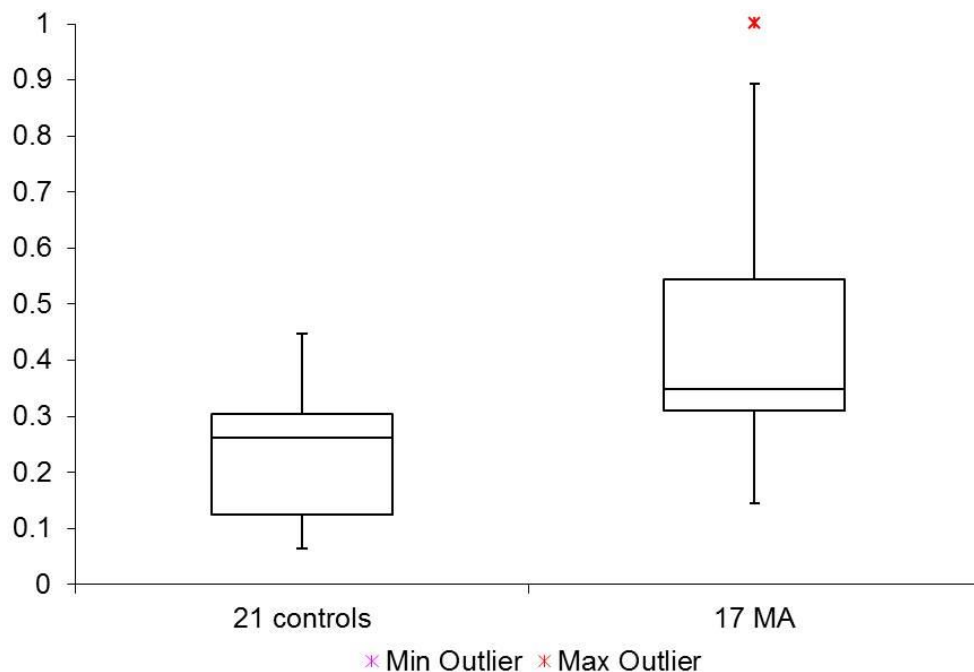
Statistically significant FT-ICR and LC-MS masses were annotated and filtered before analysing them with MetaboAnalyst. The bubble color corresponds to the p-value (y-axis) from the pathway enrichment analysis and the bubble size represents the pathway impact (x-axis) from the pathway topology analysis.

Different pathways are affected in the albuminuria group. Limonene and pinene can be found in spices, essential oils and citrus fruits whereas arachidonic acid is an important polyunsaturated fatty acid synthesised in most mammals. The cytochrom P450 drug metabolism transforms compounds to increase their solubility for excretion or bioactivation. Nicotinate is a salt of the essential vitamin B<sub>3</sub> and is converted in vivo into nicotinamide. Apart from vitamin B<sub>3</sub> also the vitamin B<sub>6</sub> metabolism, different amino acid metabolisms, the caffeine metabolism and the citrate cycle are altered.

Subsequent some pathways and identified metabolites are discussed.

### 3.4.2.1 Creatinine

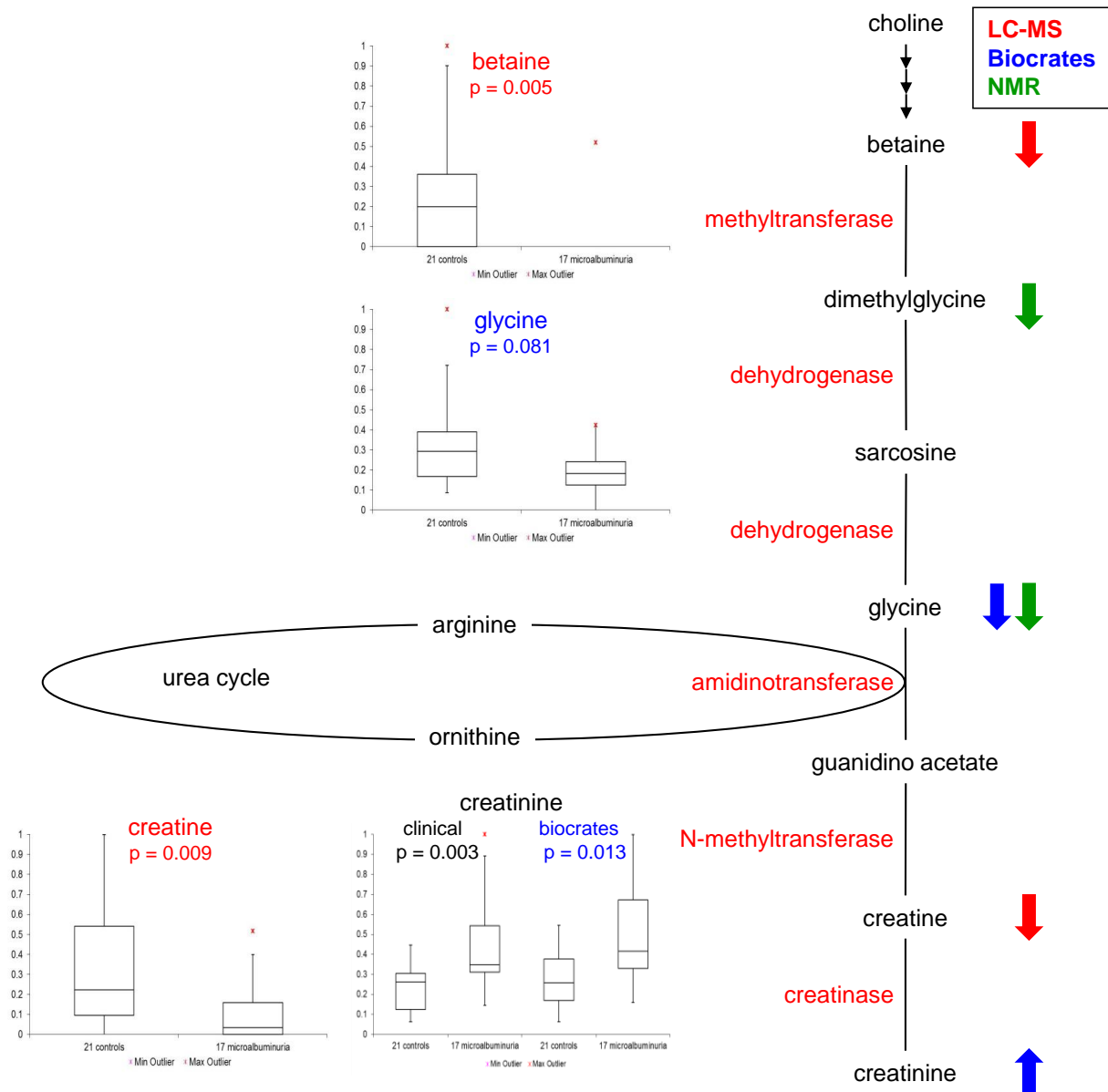
Already during the data normalisation step one of the standard compounds used for normalisation in many studies on urine samples attracted attention. Thus, in this work creatinine serves as starting point to investigate the metabolic alterations in microalbuminuria. The box plots of the creatinine value of the control and MA group shows a significant difference between those two groups with a p-value of 0.003 applying the Mann-Whitney U test (Figure 39).



**Figure 39: Box plot of creatinine in the sample urines**

The box plot shows a significant difference (Mann-Whitney U test  $p = 0.003$ ) between the two groups and therefore the infeasibility of a creatinine normalisation.

The creatinine concentration of urine samples is often deployed for normalisation of measurements to adjust the results for the sample dilution. As mentioned before the albumin-creatinine-ratio is used for the diagnosis of MA. The box plot implies a significance of the creatinine clearance alone in the early state of albuminuria. The increased creatinine concentration even conceals the increase of albuminuria by reducing the ratio. Therefore the creatinine pathway was examined and the individual intermediates or educts were searched within the different data sets (Figure 40).



**Figure 40: Alterations of the creatinine pathway in patients with microalbuminuria**

Significant alterations are shown as arrows differentiated according to the platforms. The alterations in LC-MS, biocrates and clinical data are presented with their corresponding Mann-Whitney U p-value as boxplots.

The urinary creatinine concentration is influenced by variable factors like physical activity, stress, individual differences (muscle mass, age, sex...) as well as changing GFR and disease state [Ralib *et al.*, 2012]. Ralib *et al.* observed a higher creatinine concentration in individuals with an acute kidney injury. Thus, the significant difference in urinary creatinine levels between the control and the MA group could be a result from the changed GFR and the already damaged nephrons. Additionally Hirayama *et al.* [Hirayama *et al.*, 2012] identified creatinine along with 18 other metabolites in human serum as biomarkers for DN patients with macroalbuminuria.

Increased urinary creatinine levels in patients with microalbuminuria could therefore be a balancing act of the body to maintain serum creatinine levels.

In contrast to the increased urinary creatinine concentration different other metabolites like betaine, dimethylglycine, L-glycine and creatine of the creatinine pathway are decreased in the MA group. Schartum-Hansen et al [Schartum-Hansen *et al.*, 2013] found betaine to be related with the estimated GFR. Thus the decreased betaine concentration in the MA group could indicate a lower GFR compared to the control group which would fit in the classification system of CKD. L-glycine was found to have beneficial effects on injuries to renal tubules. Weinberg et al. [Weinberg *et al.*, 1987] demonstrated the cytoprotective effect of the amino acid against hypoxic renal tubular cell injury. Several other studies confirm the antioxidant and antinecrotic effects of L-glycine on renal cells [Baines *et al.*, 1990; Miller *et al.*, 1994; Sogabe *et al.*, 1996]. The mechanisms of these effects are still not fully understood. In humans glycine has various physiological functions. Besides its importance for protein synthesis it is an essential precursor for metabolites like porphyrines, purines, glutathione, heme and creatinine. Wang et al. [Wang *et al.*, 2013] reviewed the different positive effects of glycine in detail. Glycine was additionally identified as one of three significantly altered metabolites in individuals with impaired glucose tolerance by Wang-Sattler et al. [Wang-Sattler *et al.*, 2012].

Another metabolite was found in the NMR data to be significantly decreased in the MA group. Trimethylamine N-oxide (TMAO), a breakdown product of choline is biosynthesised from trimethylamine and acts as osmolyte for urinary regulation. Bell et al. [Bell *et al.*, 1991] connected TMAO with the degree of renal damage in chronic renal failure patients. Thus, already MA patients possibly show a significant difference in their TMAO clearance levels.

#### 3.4.2.2 Arachidonic acid metabolism

MetaboAnalyst assigned several annotated metabolites in the arachidonic acid metabolism. Though annotation was done by MassTRIX with a 1 ppm mass window, a confirmation via standard compounds, correlation analysis or MS/MS is missing. Arachidonic acid is a polyunsaturated fatty acid synthesised from linoleic acid or assimilated from meat and eggs. Together with its metabolites it is involved in

regulation of several other pathways and physiological functions such as apoptosis [Brash, 2001]. The metabolites of this pathway have key mediator roles in different critical processes and are therefore researched in connection with different diseases. So, for the pathogenesis of diabetic nephropathy the different effects of arachidonic acid metabolites are researched [Dronavalli *et al.*, 2008]. Perassolo *et al.* [Perassolo *et al.*, 2003] studied the fatty acid composition in serum of T2D patients with MA and found arachidonic acid to be lower but without statistical significance. Altogether the urinary arachidonic acid metabolites may offer a new investigation goal for the prediction of albuminuria and DN.

#### 3.4.2.3 Caffeine and limonene metabolism

The top ranked pathways of limonene -, pinene- and caffeine catabolism are clearly influenced by individual habits of drinking or food intake. Still the significant differences in the annotated masses between the MA and control group contain useful information. Further research could improve life style intervention measurements and result in new dietary supplements or drugs to prevent onset of DN.

In the caffeine pathway beneath xanthine, methyl and dimethyl uric acid three more metabolites were annotated. Caffeine and its metabolites are known to trigger various effects in the human kidneys. In case of nephropathy continuous caffeine ingestion was found in several studies to reduce the performance of the kidneys by enhanced renin secretion into the plasma and potentiating renal damage [Osswald *et al.*, 2011]. Xanthine for example is a natural intermediate product in the degradation of adenosine or guanosine to uric acid. Tsujimoto *et al.* [Tsujimoto *et al.*, 2013] found that xanthine inhibits certain cytochrome P450 protein species important for degradation of drugs used for treatment of end stage renal disease. Also the uric acid serum level was found to be significantly linked with later development of macroalbuminuria or even end-stage renal disease [Hovind *et al.*, 2011; Niewczas *et al.*, 2014]. Thus, the caffeine pathway metabolites within urines of DM patient may provide also useful diagnostic markers for DN.



#### 3.4.2.4 Vitamin B<sub>6</sub> metabolism

The vitamin B<sub>6</sub> pathway is another important pathway in the MetaboAnalyst diagram (Figure 38). Vitamin B<sub>6</sub> is a cofactor for many different essential enzymes and therefore reactions in the human metabolism. It is generated from three different precursors which are present in sufficient amounts in most food.

Apart from the amino acid metabolism vitamin B<sub>6</sub> is involved in hemoglobin formation, gluconeogenesis, lipid metabolism, synthesis of neurotransmitters and histamine as well as gene expression. Vitamin B<sub>6</sub> and its metabolites play crucial roles in oxidative stress response too. Mooney et al. [Mooney *et al.*, 2009] describe the function of vitamin B<sub>6</sub> and its derivatives in detail in their review. According to them the gene expression associated with the vitamin is upregulated in cells under abiotic stress. Additionally reactions of AGE's and advanced lipoygenation end-products with lysine residues of proteins are narrowed through pyridoxamine as alternative reacting agent. Comparing the plasma levels of a control group with a CKD (stages 2-4), a hemodialysis and a renal transplant group revealed significant differences in 4-pyridoxic acid, a metabolite of vitamin B<sub>6</sub> [Busch *et al.*, 2010]. Thus the vitamin B<sub>6</sub> pathway with its strong connection to the amino acid metabolism offer an interesting pathway for further analysis in the identification of diabetic patients developing albuminuria.

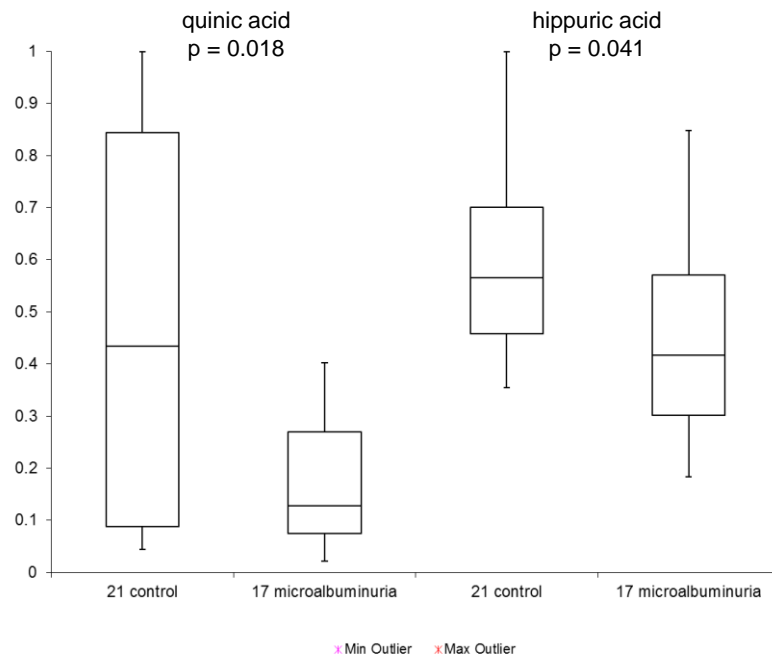
#### 3.4.2.5 Amino acid metabolism and biosynthesis

In the pathway analysis different amino acid pathways were identified. Especially from the phenylalanine and tyrosine pathways several annotated metabolites could be found. Additionally the valine, isoleucine and leucine biosynthesis were highlighted suggesting these amino acids to take part in the onset of microalbuminuria too. Wang et al [Wang *et al.*, 2011] showed the importance of leucine, valine, tyrosine and phenylalanine in the development of T2D in two independent studies. They analysed fasting urine samples with a targeted LC-MS approach for greater sensitivity and specificity. In summary a single fastening urine measurement focused on these amino acids resulted in more detailed information compared to standard risk assessment procedures like determining the fastening glucose value.

Phe, Leu, Ile and Val are essential amino acids. Tyrosine on the other hand can be biosynthesised through oxidizing Phe apart from ingesting from food. The complex Tyr and Phe catabolism use different pathways of which not all are accessible in humans.

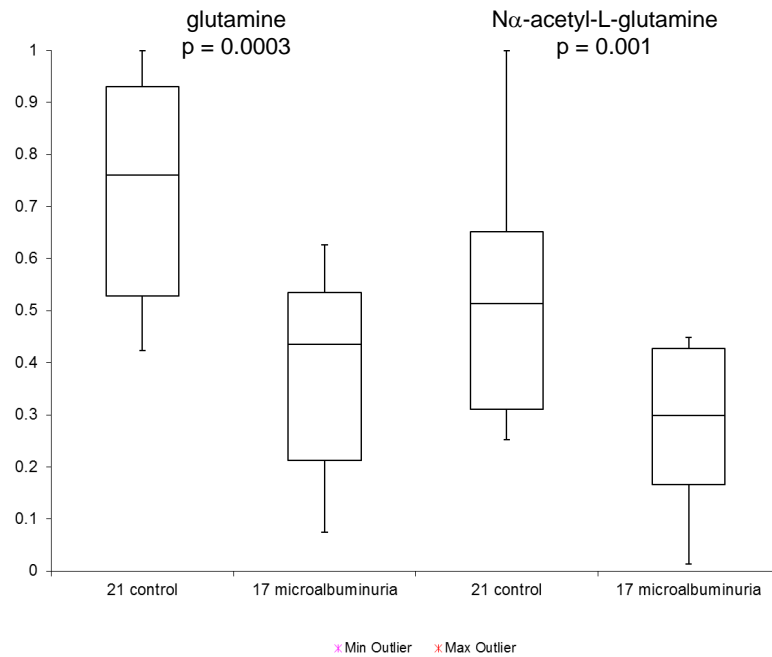
The Ratio between Tyr and Phe was found to be significantly altered in comparisons of the different CKD stages, 3 to 5 by Lundin and Weinberger [Lundin *et al.*, 2010]. Human kidneys are playing an important role in the biosynthesis and provision of Tyr by producing more Tyr than they dispose [Moller *et al.*, 2000]. Patients suffering from chronic kidney failure showed an reduced conversion of Phe to Tyr and increased Phe oxidation products in serum [Kopple, 2007]. This was also observed in the pre-diabetes samples where the Tyr/Phe ratio was significantly decreased in MA patients with a Man-Whitney U test p-value of 0,007. Thus, the metabolic products of these two amino acids may be already altered with a slightly damaged kidney as in MA patients.

From the annotated masses hippuric acid could be confirmed through correlation analysis and NMR results. Sources of hippuric acid are apart from phenolic compounds also quinic acid which was also confirmed by MS/MS experiments. Apart from plants also the human gastrointestinal tract microflora is able to transform quinic acid into hippuric acid, essential amino acids and vitamins [Pero, 2010]. Within the TULIP samples the urinary concentrations of quinic and hippuric acid are increased in the control group (Figure 41).



**Figure 41: Boxplot of quinic and hippuric acid in LC-MS negativ mode**

The altered quinic and hippuric acid levels could be based on changes in the gastrointestinal tract microbiome. Vaziri et al [Vaziri *et al.*, 2013] identified a change in the microbial population in patients with end stage renal disease. In reverse to their argument that CKD alters the composition of the gut microbiome, a changed microbiome could be involved in the development of uremia. Another cause could be a reduced clearance of hippuric acid as demonstrated in renal failure rat models [Kikuchi *et al.*, 2010; Zhao *et al.*, 2013]. There an increased hippuric acid concentration in tissue and serum was detected and explained by an accumulation of uremic toxins due to reduced renal clearance. Thus a drop in urinary hippuric acid and other uremic toxin levels could be a useful diagnostic marker for a beginning albuminuria and a good starting point for dietary interventions.



**Figure 42: Figure 40: Boxplot of glutamine and Nα-acetyl-L-glutamine in LC-MS negativ mode**

Other signals identified via correlation analysis were glutamine, Nα-acetyl-L-glutamine (Figure 42). Nα-acetyl-L-glutamine is a common urinary metabolite also used in parenteral nutrition as glutamine source because of its stability. Increased urinary levels of Nα-acetyl-L-glutamine have been associated with low estimated GFR in non-proteinuric T2D [Ng *et al.*, 2012]. The identified glutamine was one of the excluded compounds in the logD analysis (Chapter 3.3.2.6.9) because its RT didn't match with its logD. Thus, it has to be a fragment of a bigger molecule, possible a protein fragment. Also in the FT-ICR-MS results some peptides were annotated but within the LC-MS data only an Asp-Phe fragment was identified which is increased in the MA group. Since urines from MA patients have a higher protein load more protein fragments were expected but could not be confirmed. This could be explained by denaturation and precipitation of proteins and bigger fragments during the freezing which couldn't be solved again after thawing. Additionally in LC-MS the solid phase binds larger peptides on its surface which alters its separation behavior slowly. Also, their ability to form multi-charged ions because of their different functional groups complicates their detection. Still peptides are an important class of metabolites which could be important markers for the discrimination between healthy and MA individuals.

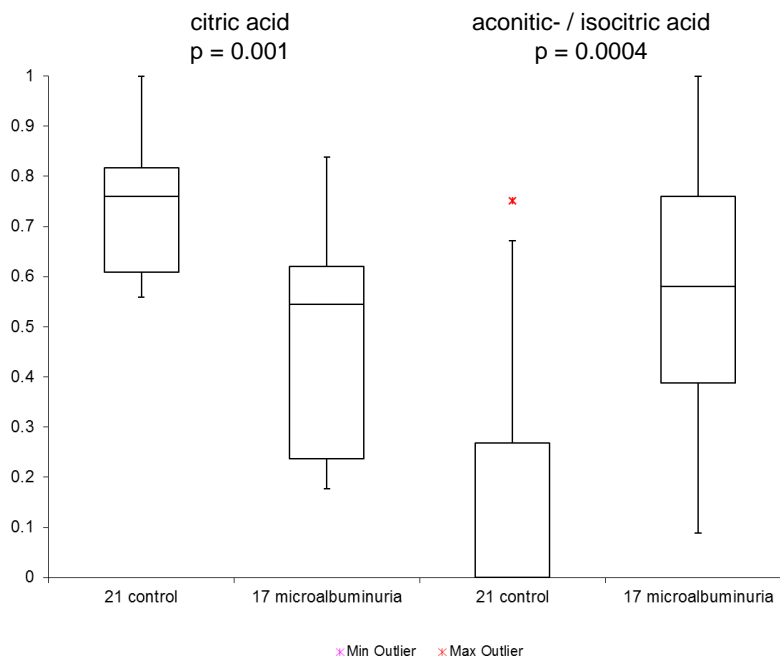
NMR analysis identified taurine, a less known sulfur amino acid, as significantly decreased in MA patients. Taurine and its metabolism are involved in several important physiological functions like the bile acid formation and in cell volume regulation as osmolyte but also in cellular dysfunction in clinical complications of diabetes e.g. nephropathy [Hansen, 2001]. In DN taurine is thought to prevent damage by reducing AGE's, reduction of glucose toxicity through Schiff's base formation with aldehydes and prevention of intracellular oxidants generation [Chesney *et al.*, 2010]. These beneficial effects will not start with onset of DN but already before. The reduced amount of taurine in the urine of MA patients could be caused by the increased demand of the body to counter renal damage. Still it has to be considered that the observed difference could also be a result of dietary intake.

#### 3.4.2.6 Citrate cycle

The citrate cycle plays an important role in the metabolism of aerobic cells. It generates energy through oxidative degradation of organic compounds and provides intermediates for different biosynthesis. The relative importance of the citrate cycle in the pathway enrichment and topography analysis (Figure 38) is based on citric, isocitric-, aconitic-, fumaric- and pyruvic acid. Only citric acid and aconitic / isocitric acid were identified through MS/MS experiments and correlation analysis. Within NMR analysis succinate and citrate were found to be increased in the control group and therefore important for the discrimination between microalbuminuric and normal subjects. Apart from ESI<sup>-</sup> LC-MS data, in positive ionisation mode several citric acid fragments were also identified and classified as significant.

Whereas citric acid is increased in the control group aconitic- or isocitric acid are in the microalbuminuria group (Figure 43). Nagai *et al* [Nagai *et al.*, 2010] observed a significant decrease of proteinuria in their T1D rat models with oral administration of citric acid. In human urine citric acid is the most abundant anion present, although more than 50% of it is reabsorbed due to its importance as substrate for renal cells [Rao *et al.*, 2011]. The reduced amount of citrate in the urine of microalbuminuria patients can be explained with the greater demand in order to counter further damage to renal cells and in this context delay progression of albuminuria. On the other hand the amount of aconitic- / isocitric acid is significantly increased in the albuminuria group. Comparison with the results of the FT-ICR-MS data could match

the LC-MS mass with aconitic acid displaying the same distribution between the two groups in FT-ICR-MS.



**Figure 43: Boxplot of citric and aconitic- or isocitric acid in LC-MS negativ mode**

Also the NMR results suggest that the amount of succinic acid is reduced in the microalbuminuria group. In their review Deen and Robben [Deen *et al.*, 2011] highlight the role of the succinate receptor in kidney. Elevated levels of succinate lead to renin release and therefore mediate hypertension which can harm the kidney. The lower concentration in patients with albuminuria seems to contradict the thesis of tissue damage caused by local RAS activation through increased succinate levels [Peti-Peterdi *et al.*, 2008]. Still the mechanisms behind succinate as signaling molecule are not yet fully understood.

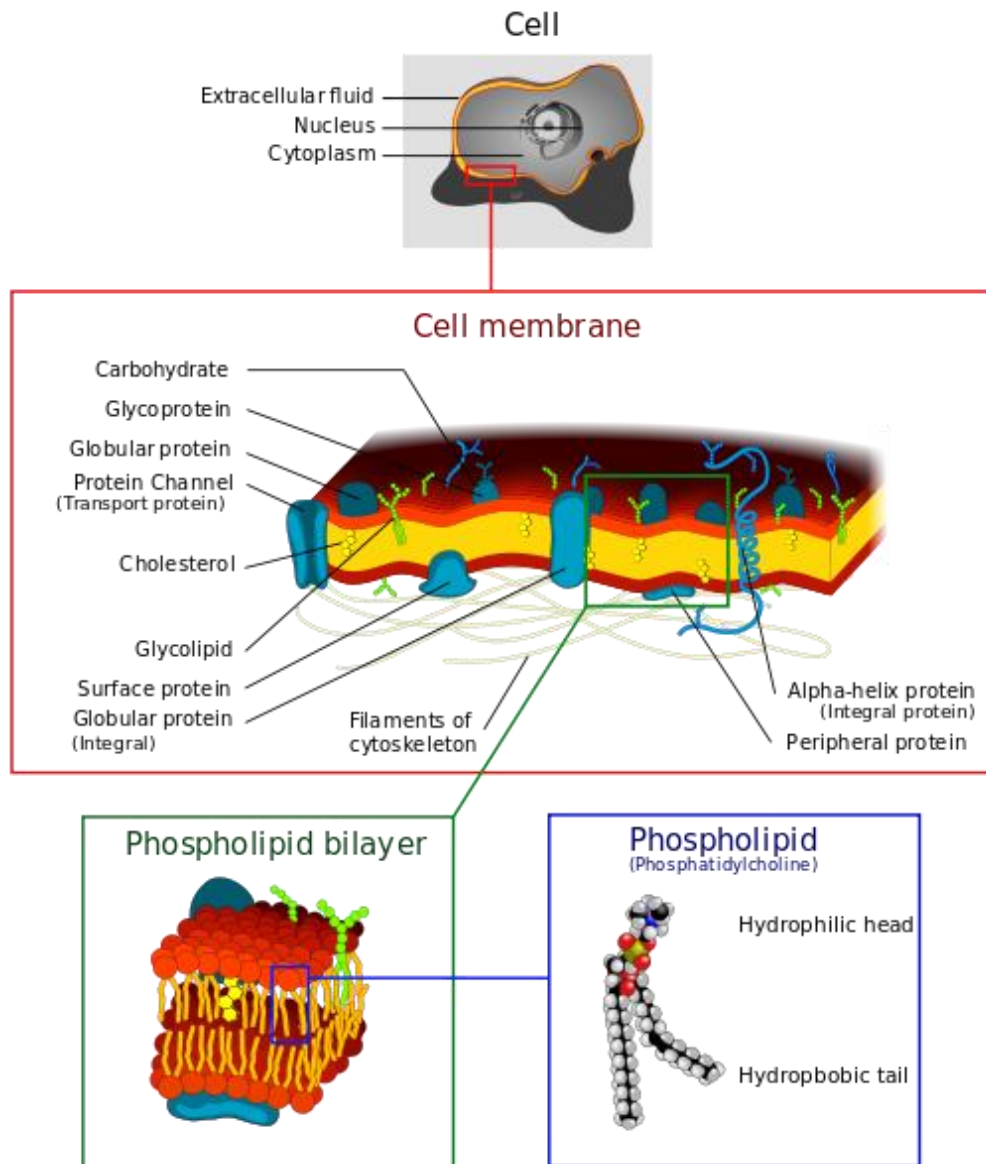
#### 3.4.2.7 Other confirmed compounds

Apart from the metabolites identified and confirmed in the pathways discussed above different others could be verified through MS/MS and correlation analysis as well as with NMR and with the biocrates kit.

Most significant metabolites found in the biocrates data and in both normalisation strategies are lipids followed by carnitines. Within the lipids the most abundant species are phosphatidylcholines besides lyso-phosphatidylcholines and sphingolipids (Appendix, Table A12). Pang et al. [Pang *et al.*, 2008] observed a change in the different blood phospholipid classes in patients with DN. Yang et al. [Yang *et al.*, 2013] suggested a connection between the altered urinary phosphatidylcholines as well as sphingolipid concentrations and proteinuria, albuminemia, lipotoxicity or damaged kidney functions. The significant difference in some phospholipid levels within the biocrates data could thus be a result of the slightly damaged nephrons indicated through MA.

Phospholipids consist of a polar head group and two hydrophobic “tails” like fatty acids. They play an important role within living organisms by performing lipid bilayers which are basis of all cell membranes (Figure 44). Cell membranes as well as intracellular membranes separate different spaces with different chemical environments from each other and allow maintaining these conditions by passive or active transportation of selected compounds through embedded proteins. An altered membrane phospholipid composition will have influence on the membrane structure and thus its function [Jones *et al.*, 2012].

Slow alterations in renal cell membrane composition could lead to the pathogenesis of MA and DN. The changing membrane properties protect less from other stress factors and contribute to renal damage.

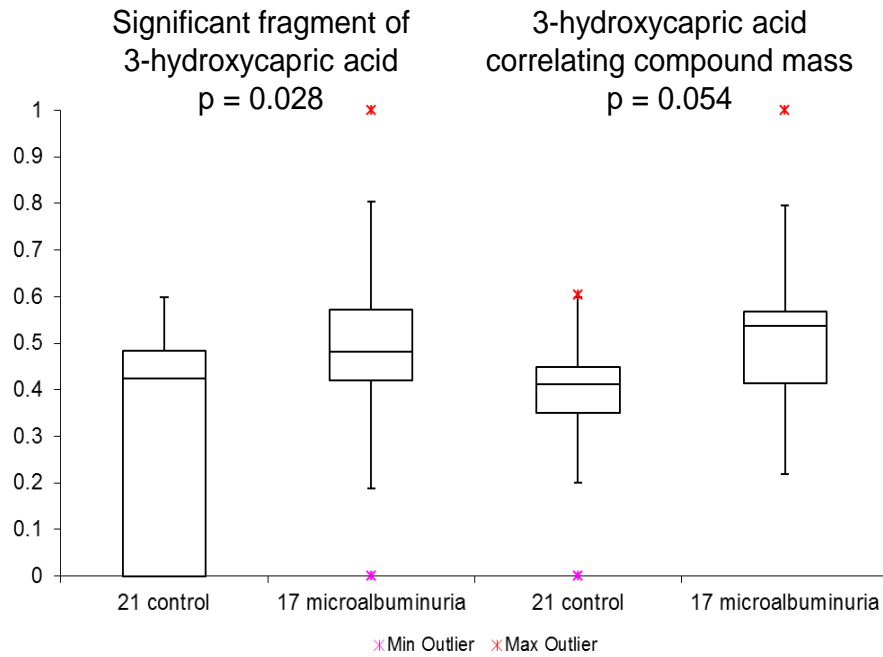


**Figure 44: Phospholipids as essential part of cell membranes**

Nearly all life depends on membranes as barriers. These membranes establish spaces within the cells for special reactions or organelle and also separate the whole cell from the surroundings. The membrane properties are dependent on the on the phospholipid composition as well as the modifications through other molecules like proteins. ([http://en.wikipedia.org/wiki/File:Cell\\_membrane\\_detailed\\_diagram\\_4.svg](http://en.wikipedia.org/wiki/File:Cell_membrane_detailed_diagram_4.svg); 02.12.2013)

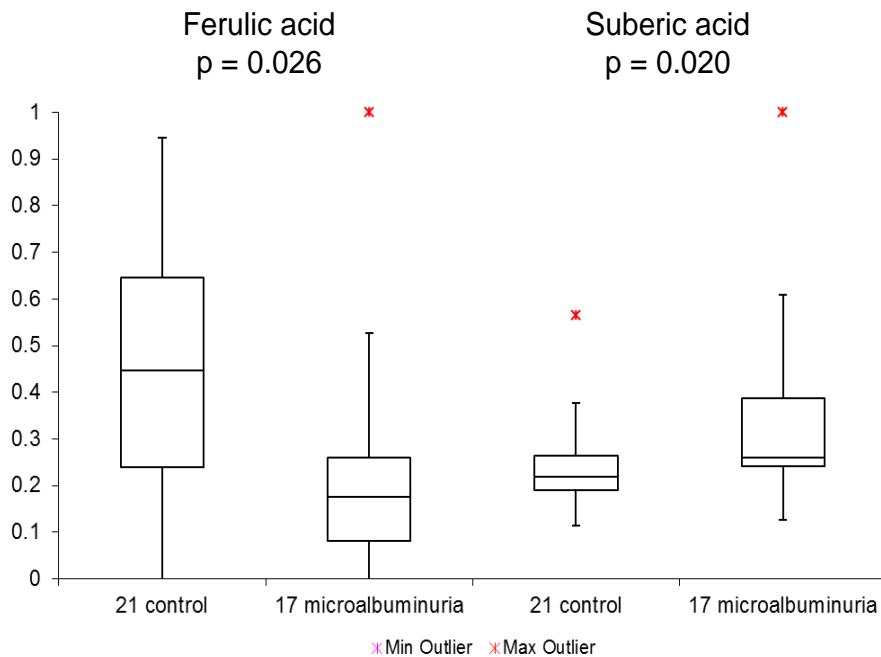
In the FT-ICR-MS data only few lipids could be annotated and also the LC-MS data showed no through MS/MS or correlation analysis identified phospholipids. One reason is MeOH as chosen solvent for the LC-MS gradient. UHPLC methods for lipid analysis use acetonitrile, isopropanol or a mixture of it for the elution. However, a fatty acid was confirmed by correlation analysis. The in the MA group significant increased mass of  $m/z$  97.06 correlates with other fragment traces and the ion mass of 3-hydroxycapric acid (Figure 45).





**Figure 45: Boxplot of 3-hydroxycapric acid in LC-MS positive mode**

Two more acids ferulic acid and suberic acid could be identified as well with comparison of correlating mass signals with MS/MS spectra databases (Figure 46).



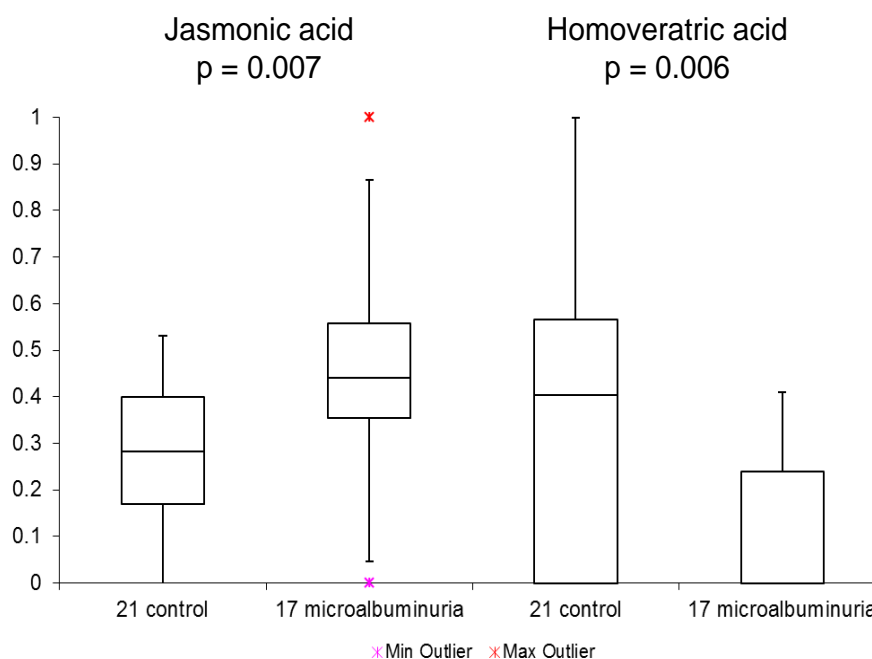
**Figure 46: Boxplot of ferulic acid and suberic acid identified in LC-MS ESI mode**

Ferulic acid is phenolic acid with antioxidant properties and can be found in various food sources like cereal products for example. Its various physiological effects like

antioxidant and anti-inflammatory as well as the different industrial capabilities are summarised by Ou and Kwok [Ou *et al.*, 2004]. Anson *et al.* [Mateo Anson *et al.*, 2009] identified the low percentage of free ferulic acid in food responsible for its low bioaccessibility. They found only less than 1% free ferulic acid in bran and aleurone while most is bound to indigestible polysaccharides. The effect of supplementary ferulic acid in a long-term therapy for CKD was researched by Peng *et al.* [Peng *et al.*, 2012] in rat models stating that ferulic acid aggravate damage caused by CKD. The decreased ferulic acid content in the MA urines could thus be a result of a different nutrition or an altered clearance rate. A reduced clearance rate increases the ferulic acid amount in blood, tilt the balance and play a part in development of albuminuria.

Suberic acid, a dicarboxylic acid, on the other hand was significantly elevated in MA patients. It was also found by Gao *et al.* [Gao *et al.*, 2012] to be elevated in urine of membranous nephropathy patients and explained by peroxisomal disorders through oxidative stress.

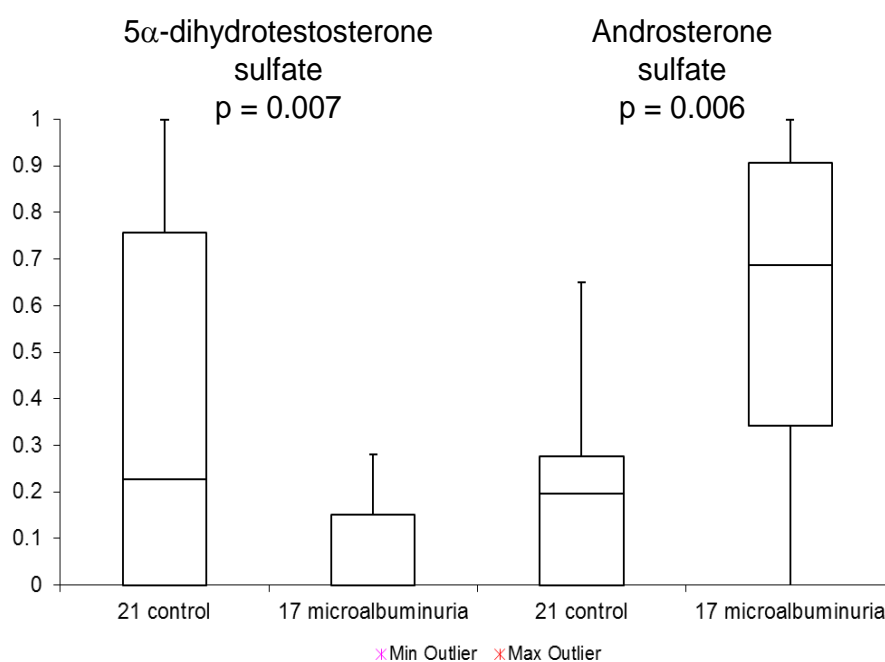
Analysis of LC-MS masses without correlating other signals through MS/MS analysis uncovered even more organic acids. Jasmonic acid is significantly increased in MA urines while homoveratric acid is reduced (Figure 47).



**Figure 47: Boxplot of identified jasmonic acid in ESI<sup>+</sup> and homoveratric acid in ESI<sup>-</sup> mode**

Jasmonic acid is a well-studied and important plant signalling hormone involved for example in the herbivore response. It was found to affect human cancer cells but not normal cells [Rotem *et al.*, 2005]. Homoveratric acid is a metabolite of 3,4-dimethoxyphenylethylamine, a methylation product of dopamine which is an important neurotransmitter. A connection with diabetic nephropathy has not been drawn for both acids, so it could be a novel target for further research.

Likewise with MS/MS experiments identified metabolites were androsterone sulphate and  $5\alpha$ -dihydrotestosterone sulphate. Their concentrations in the tested urine samples show an antidromic behaviour (Figure 48).



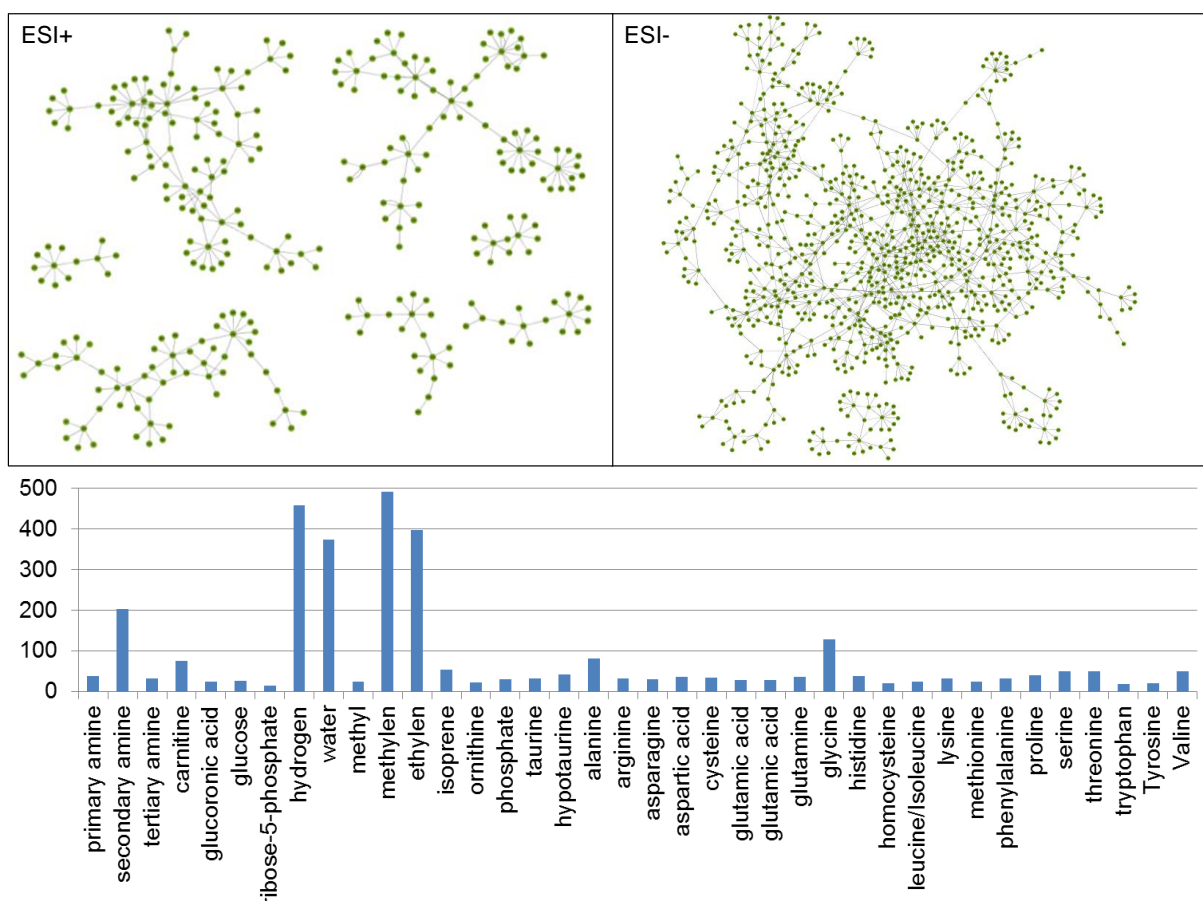
**Figure 48: Boxplot of  $5\alpha$ -dihydrotestosterone sulphate and androsterone sulphate in ESI- mode**

The MS/MS spectra of both compounds are the same and the two signals differ in their LC retention time. Since  $5\alpha$ -dihydrotestosterone sulphate, with 1.1 has a lower log D value than androsterone sulfat with 1.5, it was concluded to elute first. Both metabolites belong to the class of androgens and can be found in blood and urine of male and female.

### 3.4.3 Unknown compounds

Apart from the identified and annotated signals the majority of the significant signals are unknown. Only 9% of the significant signals could be annotated. Using Netcalc the amount of signals with a sum formula was raised to 36%. The supplementary table A13 presents the elemental compositions of significant signals with a fold change  $\geq 100\%$  between the control and the MA group. Additionally the connections and the position of a mass within the network provide additional information. A highly connected mass near the center of a network could indicate a key metabolite which takes part in several pathways.

The resulting mass-network is highly complex even by separating the data set according to the ionisation mode. For further simplification of the network analysis only direct to significant identified mass signals in the FT-ICR-MS datasets connected masses were analysed (Figure 49).



**Figure 49: Kendrick-analogous mass difference network analysis of FT-ICR-MS data**

To reduce the complexity the network was reduced to direct connections with a result signal. Detected masses are visualised as dots while the connections between two dots resemble distinct mass differences or transformations. In positive mode several distinct smaller networks were formed while in negative mode additionally a bigger network survived. Below the networks the combined frequency of the different transformations, addition or subtraction of mentioned molecules, of both ionisation modes are mapped.

The frequency of the different transformations is also contains important information. The high number of transformations depending on water can be explained by the high amount of hydroxylated compounds in urine which easily lose water. The increased amount of methylene and ethylene indicates fatty acid, free or within lipid structures with different chain length. The difference of a hydrogen mass is for example connected to hydrogenation of double bounds in unsaturated fatty acids. Even if the single amino acids have a low count, combined they reach over 800 connections. This could be explained by the increased loss of protein in MA patients which are degenerated in urine or fragmented in the ESI source. Additionally small peptides could also be found within metabolites with a high connectivity of  $\geq 10$  (Table 4).

m/z	elemental composition	mass network connectivity
447.14132	C <sub>21</sub> H <sub>24</sub> N <sub>2</sub> O <sub>9</sub>	14
254.103716	C <sub>12</sub> H <sub>17</sub> N <sub>5</sub> O <sub>5</sub>	13
449.175854	C <sub>20</sub> H <sub>33</sub> O <sub>7</sub> PS	13
447.160244	C <sub>20</sub> H <sub>31</sub> O <sub>7</sub> PS	11
232.065179	C <sub>9</sub> H <sub>15</sub> N <sub>4</sub> O <sub>4</sub> S	10
259.064878	C <sub>11</sub> H <sub>16</sub> O <sub>5</sub> S	10
292.057094	C <sub>12</sub> H <sub>11</sub> N <sub>3</sub> O <sub>6</sub>	10
311.150472	C <sub>16</sub> H <sub>24</sub> O <sub>6</sub>	10
521.142027	C <sub>24</sub> H <sub>22</sub> N <sub>6</sub> O <sub>8</sub>	10
284.061961	C <sub>11</sub> H <sub>14</sub> N <sub>3</sub> O <sub>2</sub> PS	10
374.239885	C <sub>16</sub> H <sub>31</sub> N <sub>5</sub> O <sub>5</sub>	10

**Table 4: Highly connected significant masses in the pre-diabetes study.**

Masses were filtered for a valid elemental composition and a network connectivity  $\geq 10$ .

These metabolites could connect different pathways and have therefore a regulatory function making them interesting targets for further research of the development of MA and DN.

### 3.5 Conclusion

For the metabolomics investigation of microalbuminuria 38 human urine samples from the TULIP study were analysed in a multi-parallel approach. Four different platforms UPLC-MS, FT-ICR-MS, NMR and Biocrates were applied for the data assessment and for the data analysis univariate and multivariate techniques were combined.

First a general overview of the recorded data has been given including the method performance, molecular composition of the samples and the comparability between LC-MS and FT-ICR-MS data. Thereafter in detail analysis of discriminative metabolites resulted in several important pathways altered through MA. Mainly modulated metabolite classes are small organic acids, peptides and lipids. Especially the creatinine-pathway as one highly affected was reassured through identified metabolites from different sources. The mass network analysis applied to the high accurate FT-ICR-MS masses revealed several highly connected and therefore interesting metabolites additionally delivering the corresponding chemical formula for “unknowns”.

## 4 Discovery of new diabetic nephropathy biomarkers within the PREDICTIONS study

### 4.1 Introduction

A further progression of MA to a protein excretion rate over 300 mg per day is referred to as overt nephropathy. DN represents one of the major causes of morbidity and mortality within diabetic patients and has become a leading cause of end stage renal disease [Alkhalaf *et al.*, 2010; Rossing *et al.*, 2008]. De Boer *et al.* [de Boer *et al.*, 2011] observed an increased prevalence of diabetic kidney diseases over the last two decades in the US population that wasn't influenced by the increasing use of diabetes-related medications. Since the manifestation of DN involves a complex and diverse pathology, it is still not fully understood [Valk *et al.*, 2011]. Additionally albuminuria is not necessarily a precondition of progressive DN although extensive urinary levels are associated with a higher risk of progression [Brosius *et al.*, 2013]. Therefore further research and new biomarkers are needed in order to fully understand the various pathological mechanisms and identify molecules for a practicable and reliable diagnosis of DN.

In recent years various new biomarkers were reported. MicroRNA's are at the moment the great white hope in understanding the genetic regulatory pathway alterations [Saal *et al.*, 2009]. In proteomics Alkhalaf *et al.* [Alkhalaf *et al.*, 2010] identified fragments of collagen in urine samples as biomarker for diabetes-induced damage. Also a proteomics study for differentiation between DN and nondiabetic CKD was issued by Papale *et al.* [Papale *et al.*, 2010]. In the field of metabolomics the company Biocrates Life Science AG covered sets of metabolic compounds by patent which are able to distinguish between different stages of CKD [Lundin *et al.*, 2010]. The identified biomarkers belong to various compound classes like amino acids, acyl-carnitines, biogenic amines, phosphatidylcholines, sphingomyelins, prostaglandins, sugars, bile acids and metabolites from the energy metabolism. In serum Hirayama *et al.* found 19 markers for DN including creatinine, aspartic acid, citrulline and kynurenine [Hirayama *et al.*, 2012].

## 4.2 Objective

Aim of the second part is the research of the metabolic changes caused by DN. New state-of-the-art instrumentation has been used and therefore new methods for sample preparation and measurement had to be developed for this analysis. The benefits of these next generation instruments will be shown by comparison with their predecessors from the first part (Chapter 3).

For the metabolomics analysis 74 samples from the PREDICTIONS study [Alkhalaf *et al.*, 2010] were selected and analysed. The approach described here aims to add a new dimension to the already with proteomics gathered knowledge. By identifying metabolites and pathways connected to DN, the proteomics results could be extended by an additional dimension. Thus, helping to generate new hypotheses of the pathologic background.

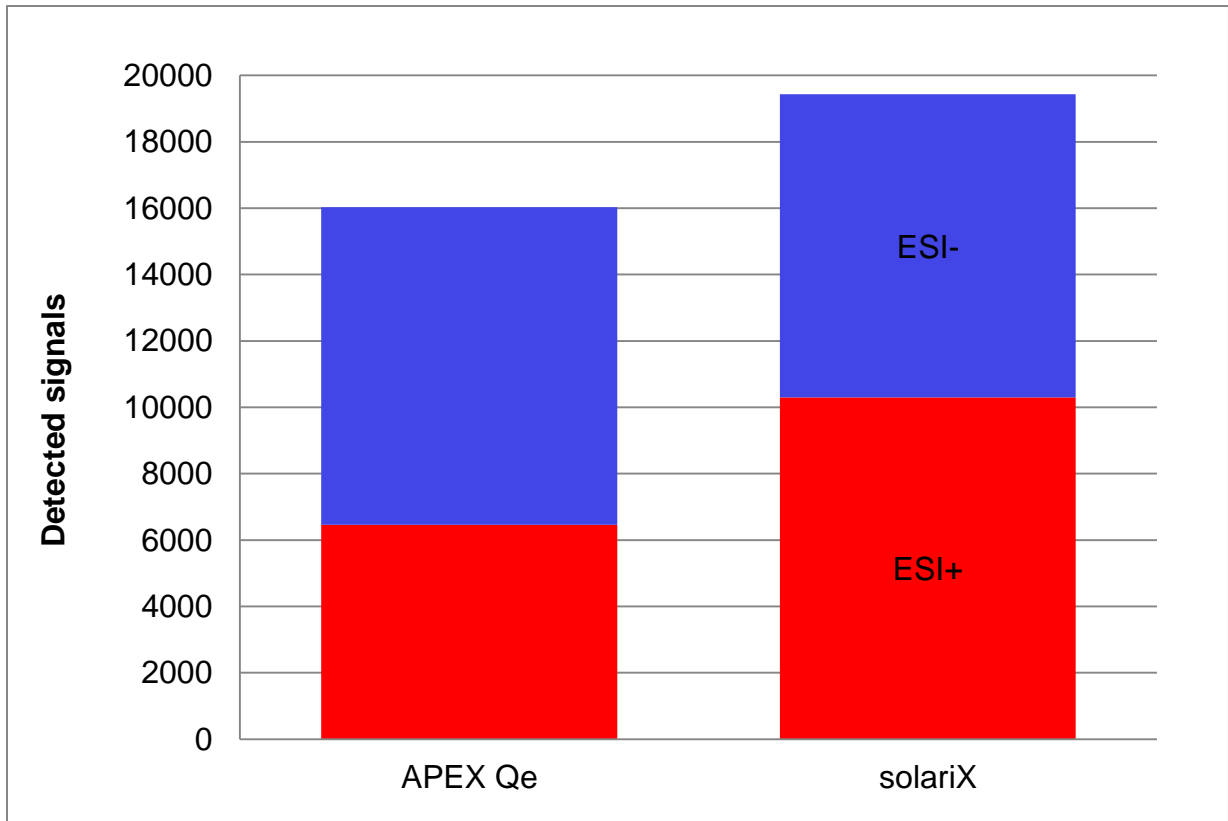
## 4.3 Material and Methods

### 4.3.1 Comparison between instrument generations

The fast evolution in the field of chromatography and mass spectrometry leads to newer instrument versions within some years. The measurements for this study were carried out on state of the art instruments which offer faster scan times by increased sensitivity and accuracy.

The new Bruker solariX<sup>TM</sup> ion cyclotron resonance Fourier transform mass spectrometer (Bruker Daltonics GmbH, Bremen Germany) was operated with the same 12 Tesla super conducting magnet (magnex scientific Inc. Yarnton, United Kingdom) as the Bruker APEX Qe system used in chapter 3. The comparison was done with urine of the same urine QC sample batch which was stored at -80°C in between. A 1:50 urine dilution with 100% MeOH for the APEX and 75 % MeOH for the solariX system was measured using 300 scans and 2 megaword. Not considering the difference in the experiment parameters and the storage time still a noticeable difference in the number of detected signals can be seen and more important the total measurement time for one sample could be reduced by more than one third (Figure 50). The truncated analysis time can be used for the analysis of more samples (high throughput) or for more scans per sample (higher resolution).





**Figure 50: Comparison between the Bruker APEX Qe and the solariX™ FT-ICR-MS system**

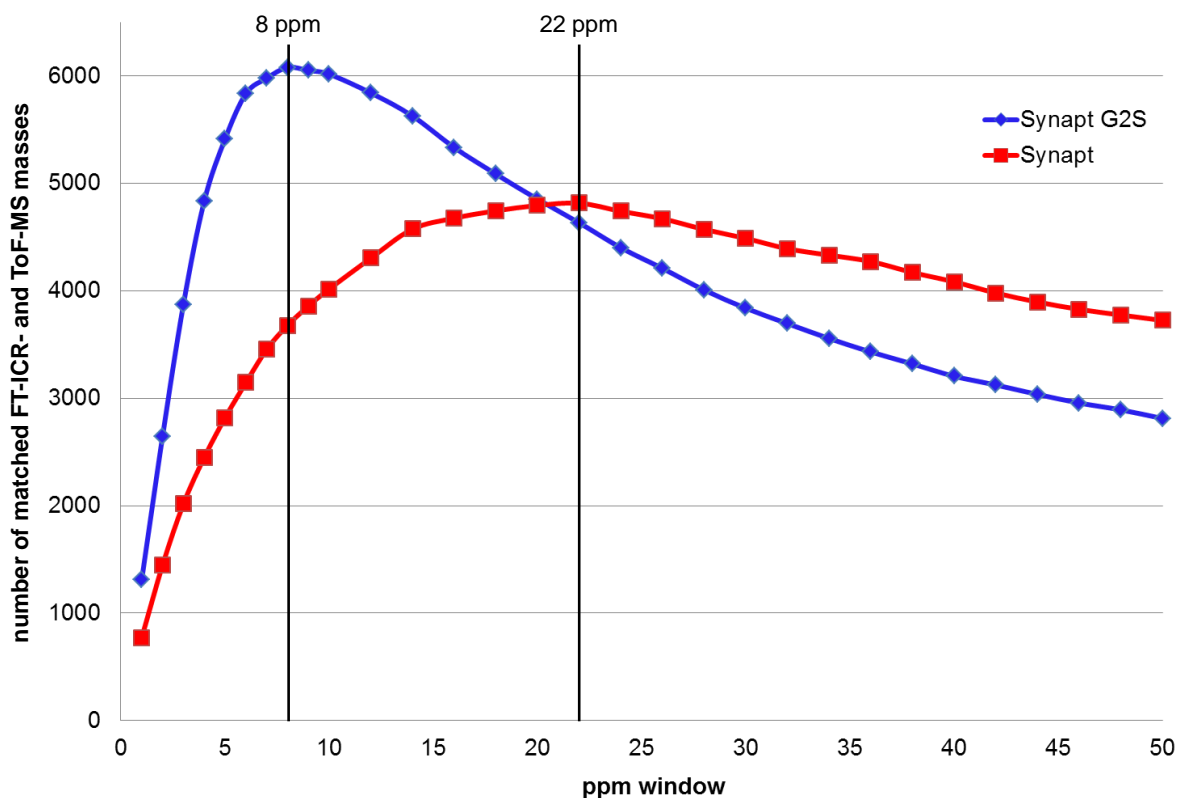
For the comparison a QC sample was diluted 1:50 and analysed with 2 megaword and 300 scans. The total analysis time for one sample was much shorter for the newer solariX™ system using the same number of scans.

Additionally the increased instrument sensitivity leads to peak deformations in the ESI<sup>+</sup> mode of the 1:50 diluted QC urine necessitating an even higher dilution, an optimisation of the method or a different sample preparation.

Also the improvements on the UHPLC-MS system are in evidence and offer new possibilities for the analysis. The major advantages of the ACQUITY UPLC® I-Class system compared to the system used for the previous study (Chapter 3) are the optimised dead volume and the injection system. A smaller system dead volume leads to sharper peaks since there is less space without a flow where substances can dwell. The new injection system directs the whole gradient through the sample needle instead of drawing sample with it in an additional sample loop and flushing it offline. This composition reduces the sample carryover effects.

The SYNAPT™ G2S HDMS™ system offers a faster scan time, improved resolution, better sensitivity and a higher accuracy compared to the SYNAPT™ HDMS™ system used in chapter 3 for the TULIP study. These developments offer the possibility of a reduced analysis time without signal loss. An increased sample set raises the

reliability of applied models and the identified significant signals. Comparing the number of matched peaks between the UHPLC-MS and the corresponding FT-ICR-MS data a big shift to smaller ppm values can be observed showing the improved accuracy (Figure 51).



**Figure 51: Comparison between the Synapt™ HDMS™ and the Synapt™ G2S HDMS™ system.**

For the comparison only ToF-MS masses in both ionisation modes were counted if with a given mass difference window only one FT-ICR-MS peak could be matched. The newer generation shows a big shift in mass accuracy, with 8 ppm compared to 22 ppm for the maximum mass pair count.

Furthermore with the G2S system the average number of detected signals in the QC samples of the two studies nearly doubled although the runtime was reduced to one-third and only one-tenth of the sample volume was injected (appendix table A14). The compared QC sample aliquotes origin from the same OC urine batch and were stored at -80 °C in between. The increased number of detected peaks can be explained with the better resolution and sensitivity. Altogether these developments offer the possibility of a reduced analysis times and therefore the analysis of bigger sample sets. Additionally the implemented optimised IMS cell offers a good opportunity to add another dimension to the UHPLC-MS approach.

#### 4.3.1.1 Material

For the UPLC-MS method acetonitrile (ACN), methanol (MeOH), water (H<sub>2</sub>O) and formic acid (FA) in ULC/MS grade from Biosolve (Valkenswaard, Netherland) have been used. Standard compounds (appendix table A15) and consumables (appendix table A16) for method development and QC are listed in the appendix (Chapter 5). Methanol and water for the FT-ICR-MS experiments have been purchased in LC-MS quality (trade name: CROMASOLVE<sup>®</sup>) from Fluka<sup>®</sup> Analytical (Sigma-Aldrich, St. Louis, USA).

#### 4.3.2 Methods and method development

For the exploration of urine samples from T2D patients with and without DN a multi-parallel approach has been used. In this study the urine samples were analysed with 2 different platforms, UHPLC-MS and FT-ICR-MS. Additionally UHPLC-IMS was applied for some samples to show the ability of this new technique.

##### 4.3.2.1 Sample preparation

The urines from the cross-sectional case-control PREDICTIONS study were collected from T2D, aged 35 – 75 years, with and without DN [Alkhalaf *et al.*, 2010]. To exclude T2D patients with non-diabetic nephropathy only patient with a albumin excretion rate greater than 300 mg per day and overt retinopathy were classified as cases [Alkhalaf *et al.*, 2010].

The analysed sample set consists of 37 case and 37 control samples of 200 µl lyophilised urines which were selected by Prof. Dr. Dr. H. Mischak. For the measurements the urine samples were resolved with 200 µl H<sub>2</sub>O containing 0.1% FA by vortexing and short sonification steps (~10 seconds). After centrifugation at 4°C for 5 min and 11000 rpm the sample were aliquoted for the different experiments. In UHPLC-MS analysis only 36 of the 37 control samples could be measured.

The QC urine samples prepared for the pre-diabetes study (see chapter 3.3.2.1) were used as QC samples. Prior to use the QC sample was removed from -80°C, thawed and stored at 4°C.

#### 4.3.2.2 UHPLC

The UHPLC-MS measurements were performed on an ACQUITY UPLC<sup>®</sup> I-Class liquid chromatography system coupled to a SYNAPT<sup>™</sup> G2S HDMS<sup>™</sup> mass spectrometer at Waters GmbH (Manchester, United Kingdom) by Dr. Cubbon. This new generation UHPLC-MS platform provides an increased sensitivity along with increased accuracy, speed and resolution compared to their precursors. Thus the total analysis time and the sample volume can be reduced still offering an equal or even better performance than the old system (Chapter 4.3.1).

The urine aliquots were sent to Manchester on dry ice and stored there at -80 °C prior to analysis. After thawing them on ice the samples were centrifuged and prepared in vials for injection.

The chromatographic conditions were:

Column:	Waters AQUITY UPLC® BEH C18; 1.0x150 mm; 1.7 µm particles
Temperature:	40°C
Injection volume:	0,5 µl
Solvent A:	Water; 5% MeOH; 0.1% FA
Solvent B:	MeOH; 0.1% FA
Flow rate:	0.100 ml/min

Time [min]	% Solvent B
0	0
0,5	0
5,5	100
9,5	100
10	0
12	0

The whole sample set was measured two times in randomised batches with one QC urine injection after every 10<sup>th</sup> sample injection for both ionisation modes.

The chromatograms were acquired and automatically calibrated with the lock mass Leucine-Enkephaline by MassLynx software (version 4.1 SCN704, Waters GmbH, Eschborn, Germany). The peak picking and data matrix generation were done with MarkerLynx (version 4.1 SCN704, Waters GmbH) at Waters GmbH (Manchester,

United Kingdom) by Dr. Cubbon. The table A17 in the appendix presents the applied settings for the peak picking algorithm and the data matrix generation.

Chromatographic peaks had to have a minimum intensity of 1000 counts, a mass difference of less than 0.02 Da, and a retention time difference of the maximum 9 s to be recognised as the same signal in the different samples. Additionally a peak had to be present in at least 25% of the samples to be included into the data matrix. The signals were normalised to the samples total peak area before exporting them into Excel 2010 (Microsoft<sup>®</sup>, Redmond, USA).

#### 4.3.2.3 FT-ICR-MS

The FT-ICR-MS measurements were carried out applying a Bruker solariX<sup>™</sup> ion cyclotron resonance Fourier transform mass spectrometer with an Apollo II electrospray source (Bruker Daltonics GmbH, Bremen Germany) and a 12 Tesla super conducting magnet (magnex scientific Inc. Yarnton, United Kingdom). The samples were kept at 6 °C and injected through a Gilson autosampler (sample changer 223, Gilson Inc., Middeltown, USA) applying a flow rate of 2.0 µl/min. The used methods were optimised for a low mass range from 150 to 600 Da. Prior to each analysis an instrument calibration using a 1 ppm arginine solution with a mass error below 100 ppm was conducted to ensure a high quality of the measurements.

As shown in chapter 4.3.1 the new Bruker solariX<sup>™</sup> with its increased sensitivity required a different sample preparation. Thus, using the QC samples different SPEs have been compared with PPE regarding the amount of detected masses.

As PPE sample a 1:50 dilution of the QC urine with 75% MeOH was prepared and centrifuged for 5 min at 4°C with 11000 rpm.

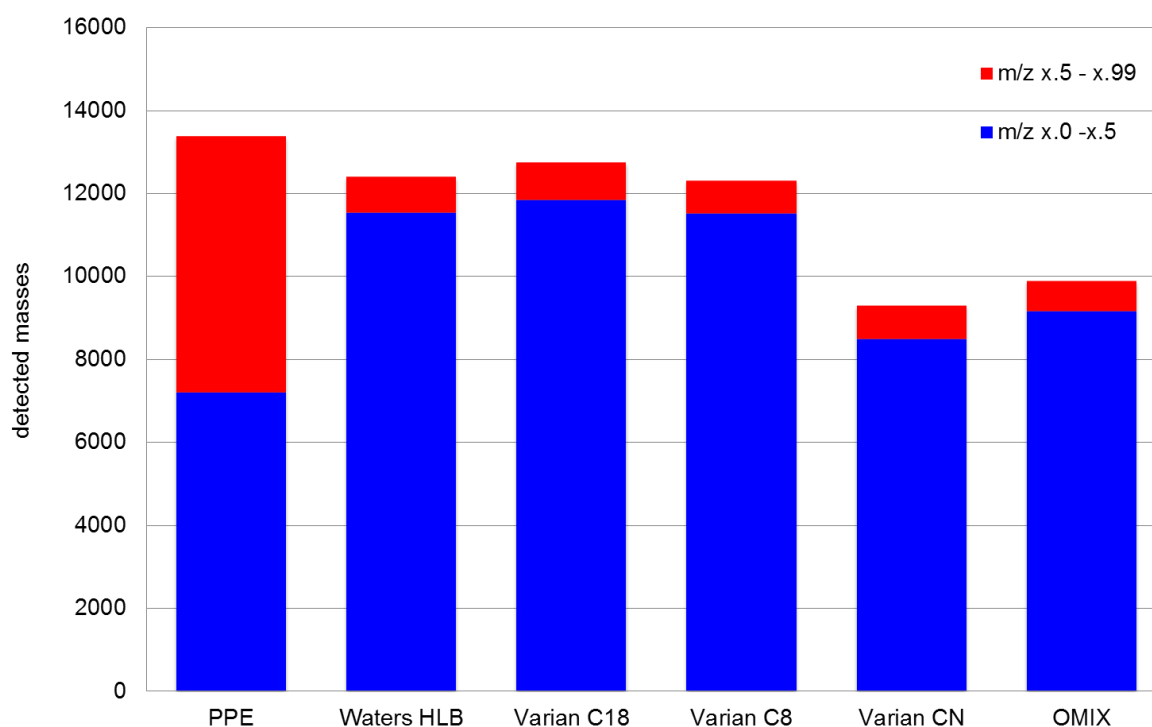
The four cartridges and the pipette tip system used for generation of the SPE samples are listed in appendix table A16. They base on different solid phase particles and extract therefore different sets of metabolites. For a better comparability a 1:10 dilution of QC urine with water was used with a simple common extraction method:

<b>Steps</b>	<b>Solvent</b>	<b>Cartridge</b>	<b>Tip</b>
1. wash	MeOH	1 ml	10 x 100 $\mu$ l
equilibration	H <sub>2</sub> O + 0.1% FA	1 ml	10 x 100 $\mu$ l
load	diluted urine	1 ml	10 x 100 $\mu$ l
wash	H <sub>2</sub> O + 0.1% FA	1 ml	10 x 100 $\mu$ l
elution	MeOH	1 ml	10 x 100 $\mu$ l

For the measurement 500  $\mu$ l of each eluat were futher diluted with 500  $\mu$ l water and 1 ml MeOH reaching a total dilution of about 1:40 and a MeOH concentration of circa 75%.

Blank samples of each sample preparation method were obtained by repeating with pure water as “sample” in order to remove the influence of leachables from the materials on the result.

All the generated samples were measured with the same FT-ICR-MS methods in positive and negative ionisation. Extracted signals found in the respective blank were excluded from the analysis.



**Figure 52: Comparison between different SPEs and PPE in FT-ICR-MS sample preparation**

The total (ESI+ and ESI-) detected m/z less the ones in the blanks of each method is shown. The amount of m/z's with decimals in the range of x.50 to x.99 which are especially in the lower mass range multicharged, or cluster ions are marked in red.

Figure 52 shows the amount of detected signals for each sample preparation in ESI<sup>+</sup> and ESI<sup>-</sup> mode combined. While with PPE most m/z's were detected nearly half of them have decimals in the range of x.50 to x.99. These masses are especially in the low mass range, below m/z's of about 500, annotated as cluster or multicharged ions as well as halogenated compounds. In order to reduce the data complexity and increase the annotation rate as well as the accuracy, the main focus was set on the decimals range from x.00 to x.50. There the three cartridges Waters HLB, Varian C18 and Varian C8 performed well with a small advance of the Varian C18. Thus the Varian C18 cartridge with the above described method was used for the analysis of the PREDICTIONS study samples.

The measurement of the samples has been randomised. The parameters were optimised with a focus for the m/z range from 150 to 600 Da (appendix table A18). The resolving power of the obtained spectra is > 300000 at m/z 300 for both modes.

The spectra were elaborated in DataAnalysis 4.0 SP2 (Bruker Daltonics GmbH, Bremen, Germany). All spectra were internally calibrated using a calibration list



consisting of masses from the calibrated arginine spectra which are also present in most of the sample spectra. The internal calibration reached a mass error below 400 ppb. The mass lists of each spectrum were generated and exported with a signal-to-noise ratio (S/N) of 4 for both ionisation modes. All lists were aligned with a 1 ppm mass window to generate the data matrix with Matrix generator 0.4 (developed in house, M. Lucio). Since the maximum intensities differ between the samples the signals were normalised using the cumulated peak intensities of the sample.

#### 4.3.2.4 Data analysis

The obtained data from the UHPLC-MS and FT-ICR-MS measurements has been similarly elaborated as in the TULIP study (Chapter 3.3.2.6). Since a good separation between the groups has been reached without winsorisation, this step was omitted.

The used parameters and software for the different methods are briefly listed in this chapter. More detailed information of the used methods can be found in the chapters 2.2 and 3.3.2.6.

##### 4.3.2.4.1 Principal component analysis (PCA)

The PCA was calculated with the original data set using SIMCA-P version 12.0 (Umetrics, Umea, Sweden) using pareto scaling.

##### 4.3.2.4.2 Hierarchical cluster analysis (HCA)

The HCA was done with the open source software Hierarchical Clustering Explorer (HCE) [Seo, 2002] version 3.5 (Human-Computer Interaction Lab, Maryland, USA) using the Pearson correlation coefficient and average linkage. Before analysis the data was mean centred and univariate scaled.

##### 4.3.2.4.3 Orthogonal partial least square discriminant analysis (OPLS-DA)

After the non-supervised methods an OPLS-DA was applied using SIMCA-P version 12.0 (Umetrics, Umea, Sweden). The resulting models were evaluated with cross-validation and permutation tests before the discriminative variables were extracted.

#### 4.3.2.4.4 Wilcoxon, Mann-Whitney test

The Wilcoxon, Mann-Whitney test was carried out within the open source program MultiExperiment Viewer version 4.8.1 (Dana-Farber Cancer Institute, Boston, USA) applying a p-value limit of 0.05.

#### 4.3.2.4.5 HCE metabolic profile search

The metabolic profile search was done with HCE [Seo, 2002] version 3.5 (Human-Computer Interaction Lab, Maryland, USA). The data was mean centred and univariate scaled for the analysis. A model-based search mode with Pearson's correlation coefficient as distance measure was used. The two models applied for the search were high in MA, low in controls and vice versa. The threshold for the UHPLC data was set to 0.762 for the positive mode and 0.802 for the negative mode data while for the FT-ICR-MS data a threshold of 0.653 was applied for both modes.

#### 4.3.2.4.6 Signal ranking

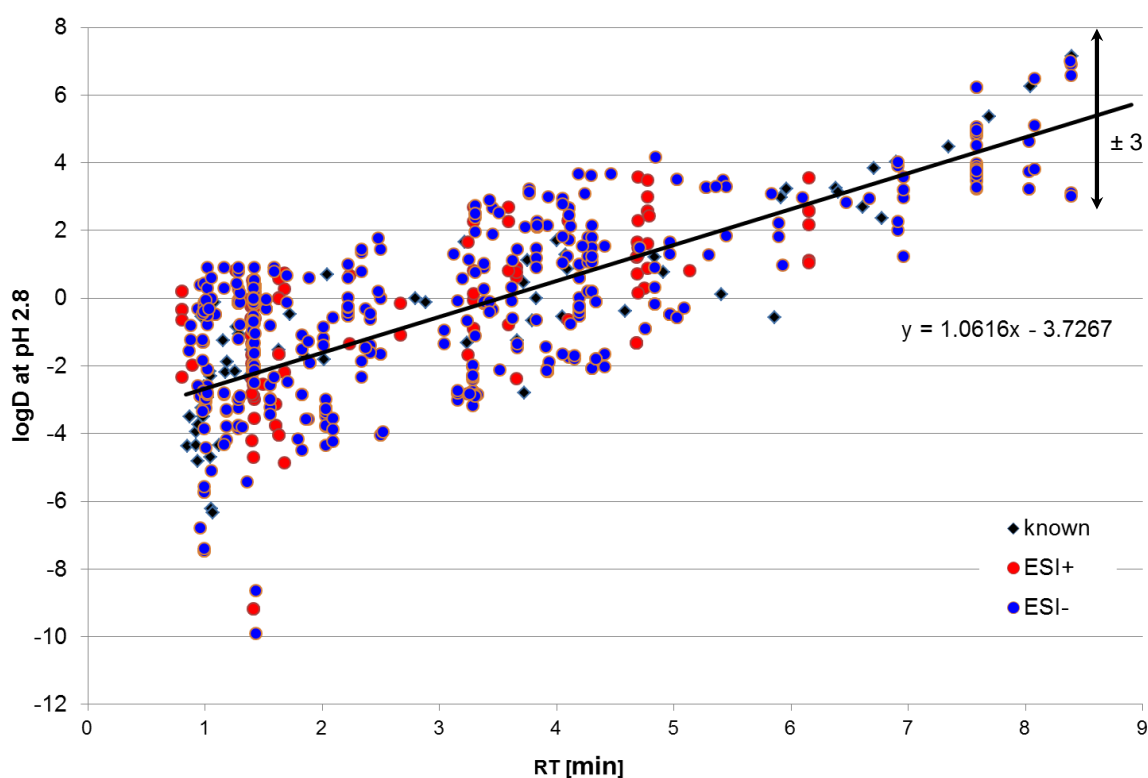
All significant m/z's from the OPLS-DA, the Wilcoxon, Mann-Whitney test and the HCE metabolic profile search were combined and ranked according to their number of appearance as significant hits. For the further analysis only signals with at least two out of three hits were used. The ranking was performed with Excel 2010 (Microsoft®, Redmond, USA).

#### 4.3.2.4.7 Mass annotation

MassTrix (<http://masstrix3.helmholtz-muenchen.de/masstrix3/>) was applied using a 1 ppm mass window in both ionisation modes for the mass annotation. Additionally manual annotation within HMDB (<http://www.hmdb.ca/>) and METLIN ([metlin.scripps.edu](http://metlin.scripps.edu)) was performed in order to compensate loss of water.

For identification of ions all correlating mass traces were compared with MS/MS spectra stored in the online data bases METLIN ([metlin.scripps.edu](http://metlin.scripps.edu)) and MassBank ([www.massbank.jp](http://www.massbank.jp)).

In order to reduce the amount of false positive annotations in LC-MS data a RT based logD window as described in chapter 3.3.2.6.9 was used (Figure 53) [Müller, 2013]. The reduction of the chromatographic runtime comes with the cost of increased standard deviation of known standards compared to the trend. Therefore, a window of  $\pm 3$  logD units around the trend was applied for filtering false positives. The calculations have been done with JChem for Excel (Version 5.12.3.966, ChemAxon Ltd, Budapest, Hungary).



**Figure 53: Reduction of false positive annotations in LC-MS by RT - logD filtering**  
(adapted from [Müller, 2013])

#### 4.3.2.4.8 MetaboAnalyst

MetaboAnalyst (<http://www.metaboanalyst.ca>) was used for the functional enrichment and the metabolic pathway analysis [Xia *et al.*, 2012].

#### 4.3.2.4.9 Network visualisation and elemental formula calculation

The high accurate FT-ICR-MS data was used to calculate a metabolic transformation network with the inhouse written Matlab program Netcalc [Tziotis *et al.*, 2011]. As transformation list mass differences from different biological conversions were used (Appendix, Table A10).

#### 4.3.2.4.10 Data alignment

The alignment of the UHPLC and FT-ICR-MS results was done applying an 8 ppm mass window (Figure 51). Since the same molecule could ionize differently in both approaches several possible ions were calculated from the high accurate FT-ICR-MS masses to compare them with the UHPLC masses.

#### 4.3.2.4.11 Correlation analysis

The correlation analysis described in chapter 3.3.2.6.13 was applied to find correlated m/z traces. As RT window  $\pm 0.05$  min were applied searching for correlations above 0.9. The matching signals were then compared with known fragmentation patterns.

## 4.4 Results and Discussion

### 4.4.1 First survey of the data

#### 4.4.1.1 Overview

After normalisation signals present in less than 10 % of the samples were removed from the dataset. Thereafter a first overview of the remaining data with HCA and PCA was generated. The QC samples and the sample repetitions in LC-MS are forming separate clusters. Thus, the individual sample differences are not superimposed by instrumental or measurement time based variations. Also a clustering between cases and controls with some exceptions can be observed in the ESI<sup>-</sup> LC-MS data (Figure 54). The other data sets show no separation with HCA and also with PCA no separation between cases and controls in FT-ICR-MS and LC-MS data was achieved.

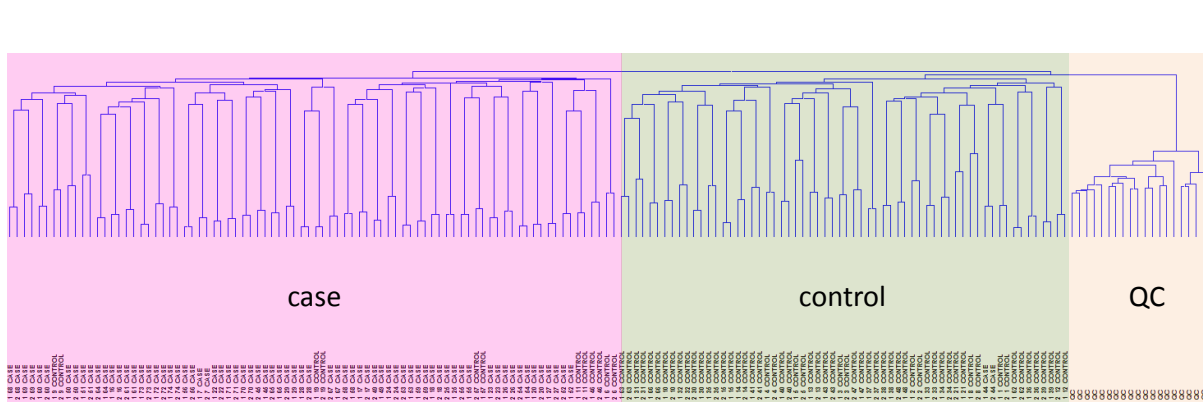
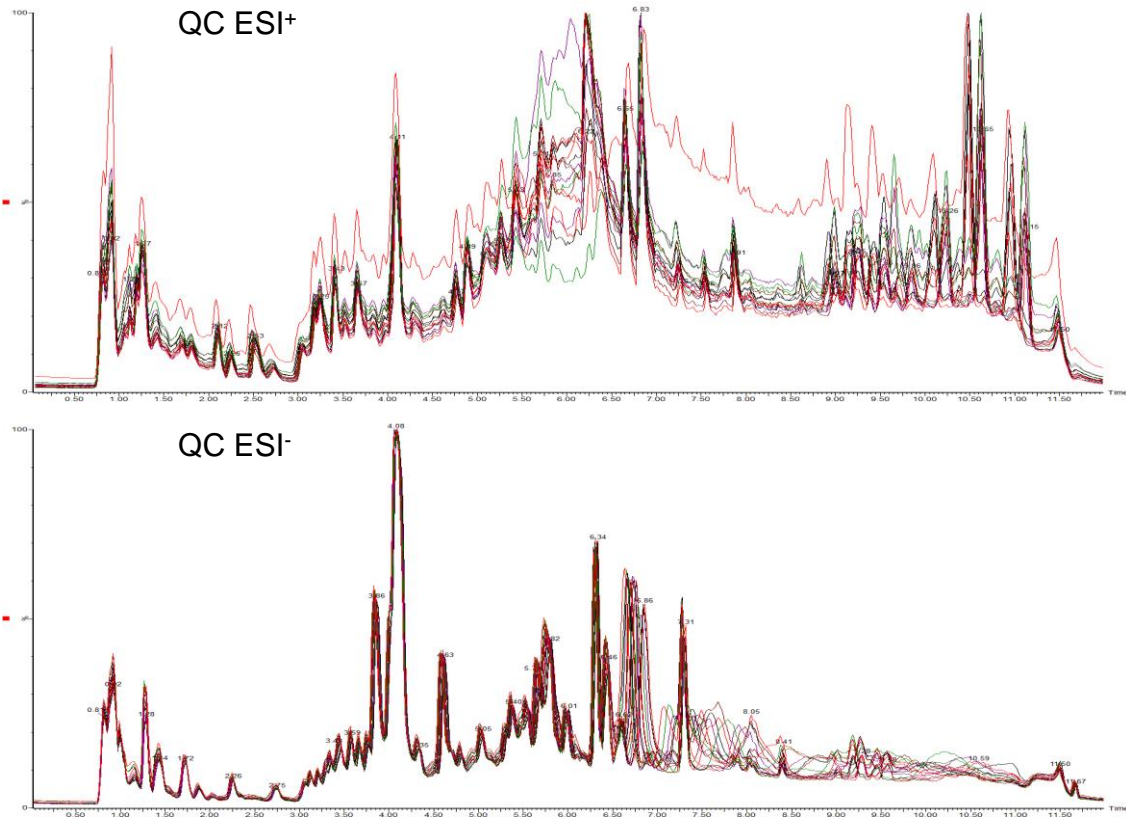


Figure 54: HCA of the ESI<sup>-</sup> LC-MS data before averaging of the PREDICTIONS study

#### 4.4.1.2 General performance of the method

The reproducibility of the externally measured LC-MS data was controlled visually by overlaying the QC sample runs of each ionisation mode (Figure 55). The first sample in positive mode shows a higher TIC but comparable peaks. In the first part of the gradient, 5 min in ESI<sup>+</sup> and 6 min in ESI<sup>-</sup>, the methods are performing well but deteriorate afterwards (Figure 55). Thus, an RT window of 0.15 min for the peak piking and alignment algorithm was chosen and resulting signals in the later part of the gradient have been examined carefully.

The average relative standard deviation of the peak area of signals detected in all QC runs are 24 % for the ESI<sup>+</sup> data and 17 % for the ESI<sup>-</sup> data. Considering all peaks present at least in 25 % of the QC measurements the relative standard deviation double in the rear part of the gradient which was less reproducible.



**Figure 55: Overlay of the quality control runs within the LC-MS analysis**

Overlay of the QC samples TIC's within the external measured data. The first QC sample in positive mode shows an increased TIC. The repeatability is good for the first 5 min in ESI<sup>+</sup> and 6 min in ESI<sup>-</sup> but decreases afterwards.

In LC-MS 15715 signals in ESI<sup>+</sup> and 7210 Signals in ESI<sup>-</sup> remained after averaging and filtration. 1469 could be matched between the two ionisation modes which accords 9 % of the positive or 20 % of the negative peaks. 44 % of the signals in positive and 36 % in negative mode respectively could be detected in more than half of the samples.

Within FT-ICR-MS after calibration, extraction and filtration 8803 signals in positive mode and 8600 in negative mode were used for the detailed data analysis. The amount of peaks present in at least 50% of the samples was increased compared to the pre-diabetes study with 29 % of the positive and 37 % of the negative signals. Still the individual differences prevail.

#### 4.4.1.3 Molecular composition of the samples

The chemical formulas can be calculated with the high accurate masses from the FT-ICR-MS measurements. 84% of the detected positive and 58 % of the negative masses could be assigned with elemental compositions through the network analysis. A first overview of the compound classes within the samples is shown by the Van Krevelen diagram (Figure 56).

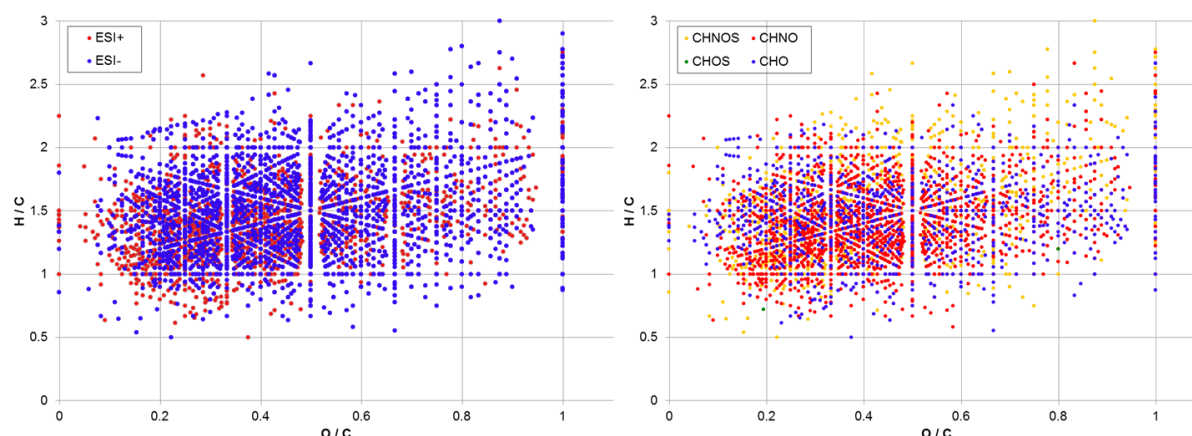


Figure 56: Van Krevelen diagram of the FT-ICR-MS data from PREDICTIONS study samples

The hydrogen rich and oxygen poor areas are, as expected, depleted. Many metabolites are physiologically oxidised in order to increase the solubility and enhance the excretion. In both positive and negative ionisation mode the protein area is very crowded. This can be explained by increased protein content in urine of nephropathy patients. Contrary to the TULIP samples (Chapter 3.4.1.3) the positive and negative formulas are equally distributed.

The mass annotation with MassTrix resulted in assignment of 23% in positive and 19 % negative mode.

#### 4.4.1.4 Intersection between the LC-MS and FT-ICR-MS data

The mass comparison between the platforms UHPLC-MS and FT-ICR-MS has been performed using a mass window of 8 ppm (see chapter 4.3.1). For the comparison of the UHPLC-MS ESI<sup>+</sup> with the ESI<sup>-</sup> dataset an RT window of 0.2 min and a mass window of 0.02 was applied. The FT-ICR-MS ionisation modes were compared applying a mass window of 1 ppm considering also sodium adducts in positive mode.

More than 19 % of the signals equalling 1395 m/z's of the smallest data set could be matched in every other dataset (Figure 57)

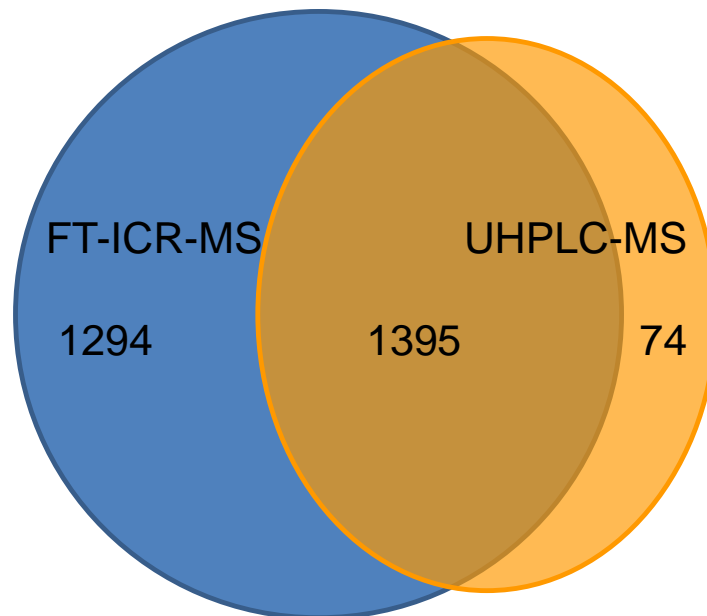


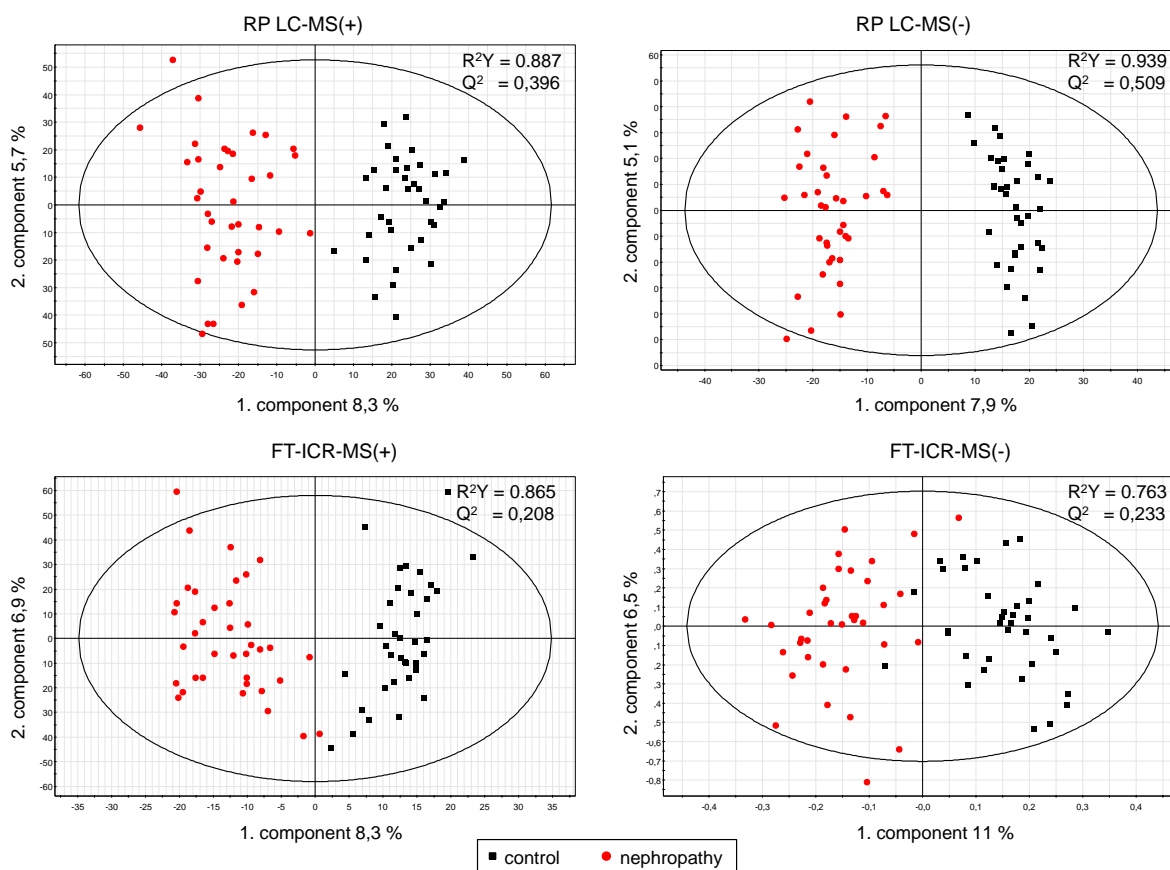
Figure 57: Venn diagramm of the mass comparison between the platforms

Still there are big differences between the datasets which can be explained by the the many influencing effects like sample preparation and ionisation behaviour due to the injection method as well as instrument geometry and many more.

#### 4.4.2 Metabolic alterations in patients with diabetic nephropathy

The individual differences between the samples seen in the overview superimpose any separation between the two groups in the first seven main components of the PCA. Still the influence of nephropathy on the urinary composition is strong enough to allow calculation of good OPLS-DA models without any further data treatment (Figure 58).

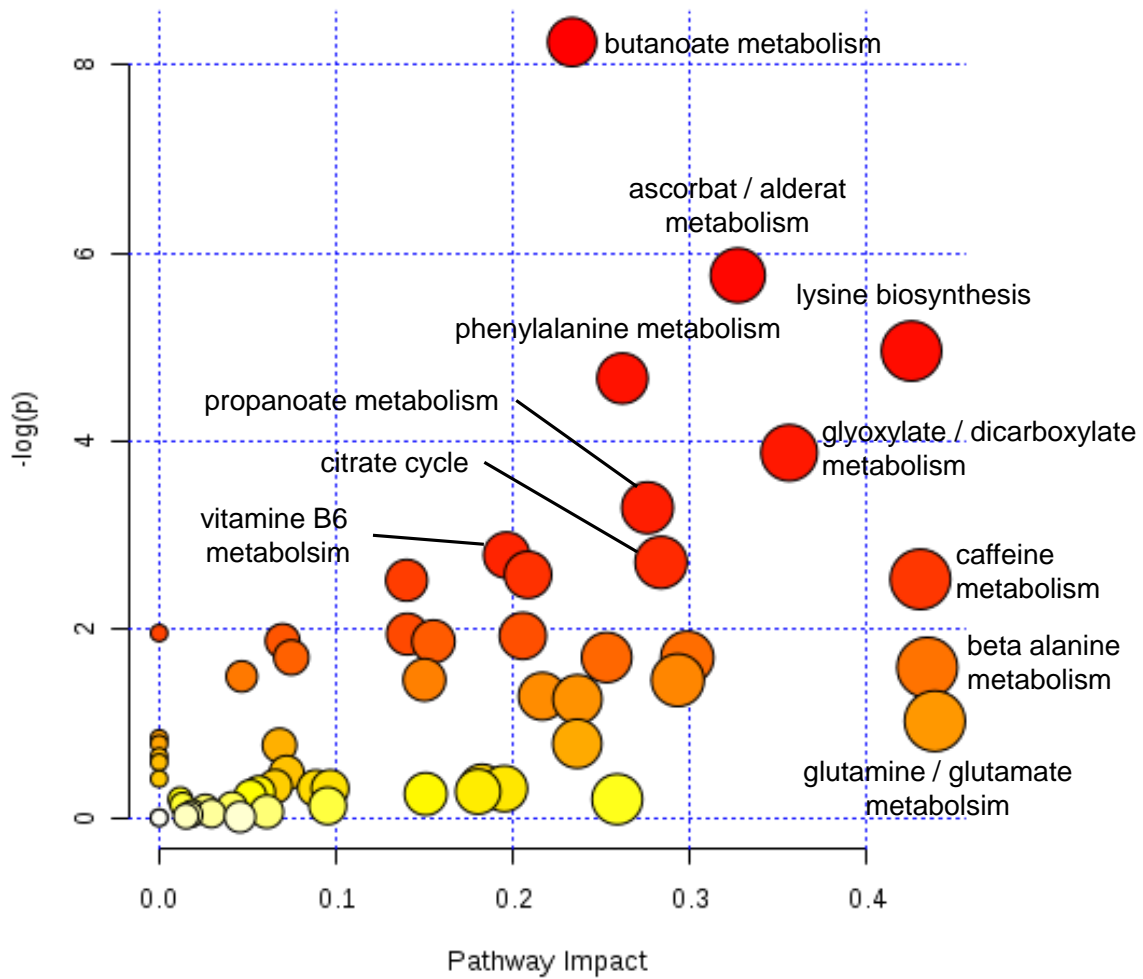




**Figure 58: OPLS-DA models of control and nephropathic samples**

Important variables from the OPLS-DA were compared with the results of a non-parametric Wilcoxon, Mann-Whitney test and the HCE profile search. In the LC-MS data 227 mass signals in positive and 535 in negative mode were supported by at least 2 statistical methods while in FT-ICR-MS it were 262 in ESI<sup>+</sup> and 209 in ESI<sup>-</sup>.

In order to confine the number of annotations for the LC-MS result a filtration step based on the logD value and retention time was introduced (see chapter 4.3.2.4.7). All annotations found within the KEGG database were forwarded to an over representation test and a pathway topology analysis using MetaboAnalyst [Xia *et al.*, 2012].



**Figure 59: Metabolic enrichment and pathway analysis of significant MS annotations within the PREDICTIONS study**

Annotations of significant FT-ICR and LC-MS masses were filtered and analysed with MetaboAnalyst. The bubble color corresponds to the p-value (y-axis) from the pathway enrichment analysis and the bubble size represents the pathway impact (x-axis) from the pathway topology analysis.

Below some of the identified pathways are examined more closely.

#### 4.4.2.1 Butanoate and propionate metabolism

Butyric and Propionic acid belong to the so-called short chain fatty acids. Butyrate was found important especially in the microbiome colon interaction and mammalian colon energy metabolism [Donohoe *et al.*, 2011]. Both pathways are directly connected to the citrate cycle, a crucial part in energy production in the human body.

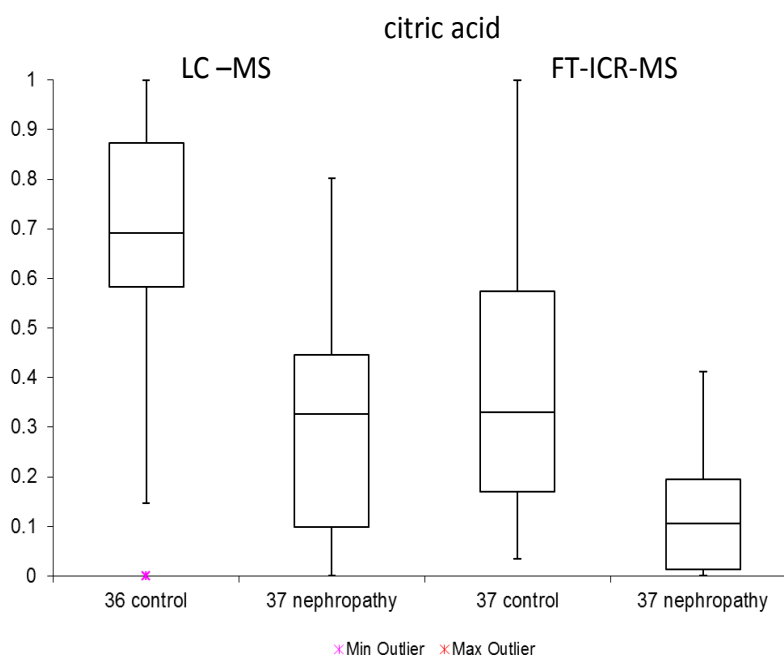
Focusing on the kidney both pathways are connected with renal cell carcinoma, one of the most common highly lethal cancer [Perroud *et al.*, 2006; Tun *et al.*, 2010].

Additionally Huang et al. [Huang *et al.*, 2006] identified significant differences in gene expression of the propanoate and the citrate pathway amongst others comparing skin fibroblasts of T1D patients with fast and very slow rates of development of DN. They concluded that the increased ROS level in nephropathy patients can be explained through the raised oxidative phosphorylation pathway activity as response to altered propanoate metabolism and citrate cycle activity. This is not restricted to T1D but is also an important issue in development of insulin resistance and therefore T2D [Houstis *et al.*, 2006].

#### 4.4.2.2 Citrate cycle

As mentioned above the butanoate and propanoate metabolism is directly connected to the citrate cycle. Through metabolising molecules from different sources to carbon dioxide and water electrons are transferred to coenzymes as energy source. Increased citrate cycle activity raises the levels of reducing agents therefore leading to increased ROS formation if they exceed a certain threshold [Huang *et al.*, 2006].

In both platforms, LC-MS and FT-ICR-MS citric acid was identified through correlation analysis. The box plot shows a decreased citric acid concentration in nephropathy patients compared to the control group (Figure 60).



**Figure 60: Boxplot of citric acid in LC-MS and FT-ICR-MS ESI mod**

This result equals the observation in the previous study (see chapter 3.4.2.6) suggesting that citric acid could be a marker for both, risk assessment and diagnosis of DN. The increased demand in DN patients as well as MA patients reflects the need of citric acid for damage prevention or at least deceleration of the DN progression.

#### 4.4.2.3 Ascorbate and aldarate pathway

Ascorbic acid is an important antioxidant vitamin and radical scavenger. It also acts as coenzyme in the biosynthesis of collagen, steroids and amino acids such as tyrosine.

Aldaric acids are a group of sugars with terminal carboxylic acids instead of alcoholic or aldehyde functionalities. They can be generated through oxidation of sugars like glucose. Thus, both pathways are directly affected by increased ROS concentrations in the human body.

Hirsch et al. [Hirsch *et al.*, 1998] observed an increased ascorbic acid clearance in DN patients and assumed a disturbed tubular reabsorption as explanation for their results.

#### 4.4.2.4 Amino acid pathways

In the pathway analysis (Figure 59) several annotations corresponding to amino acid pathways (not all are labelled) were detected. The already damaged nephrons in DN patients are not able to detain proteins as good as healthy ones. The protein degradation to single amino acids or small peptides and reabsorption in the proximal tubulus can't keep up with the high protein load, causing an increased detection of peptides and amino acids in the samples. Apart from clearance and reabsorption the kidney also plays an important role in maintaining the phenylalanine / tyrosine ratio which is impaired in humans with CKD [Kopple, 2007]. Lundin and Weinberger included different amino acids and their ratios in their method to differentiate between CKD stages 3 to 5 [Lundin *et al.*, 2010].

The annotation of the significant masses shows a high amount of small peptides and amino acids, more than in the MA study (chapter 3.4.2.5). With correlation analysis it was possible to link several mass fragments to peptides but it was not possible to identify the sequence.

Identification of peptides is especially difficult since the methods couldn't detect alteration in sequence or distinguish between amino acids with the same chemical formulas as isoleucine and leucine for example.

#### 4.4.2.5 Caffeine metabolism

The caffeine metabolism also includes, besides the caffeine degradation which depends on the individual lifestyle, xanthines and uric acids which are intermediates of physiological adenosine and guanosine degradation. Still continuous caffeine consumption in patients with nephropathy was found to further reduce the functionality of human kidneys [Osswald *et al.*, 2011]. Therefore lifestyle interventions are also an important issue in overt DN. Xia et al [Xia *et al.*, 2009] showed the clinical significance of plasma adenosine, xanthine and uric acid levels in T2D patients with DN. In the previous MA study (chapter 3.4.2.3) the caffeine metabolism was found important, too. These result suggest that the caffeine pathway might be an interesting target for the future risk assessment and DN diagnosis strategy.

#### 4.4.2.6 Vitamin B<sub>6</sub>, glyoxylate and dicarboxylate metabolism

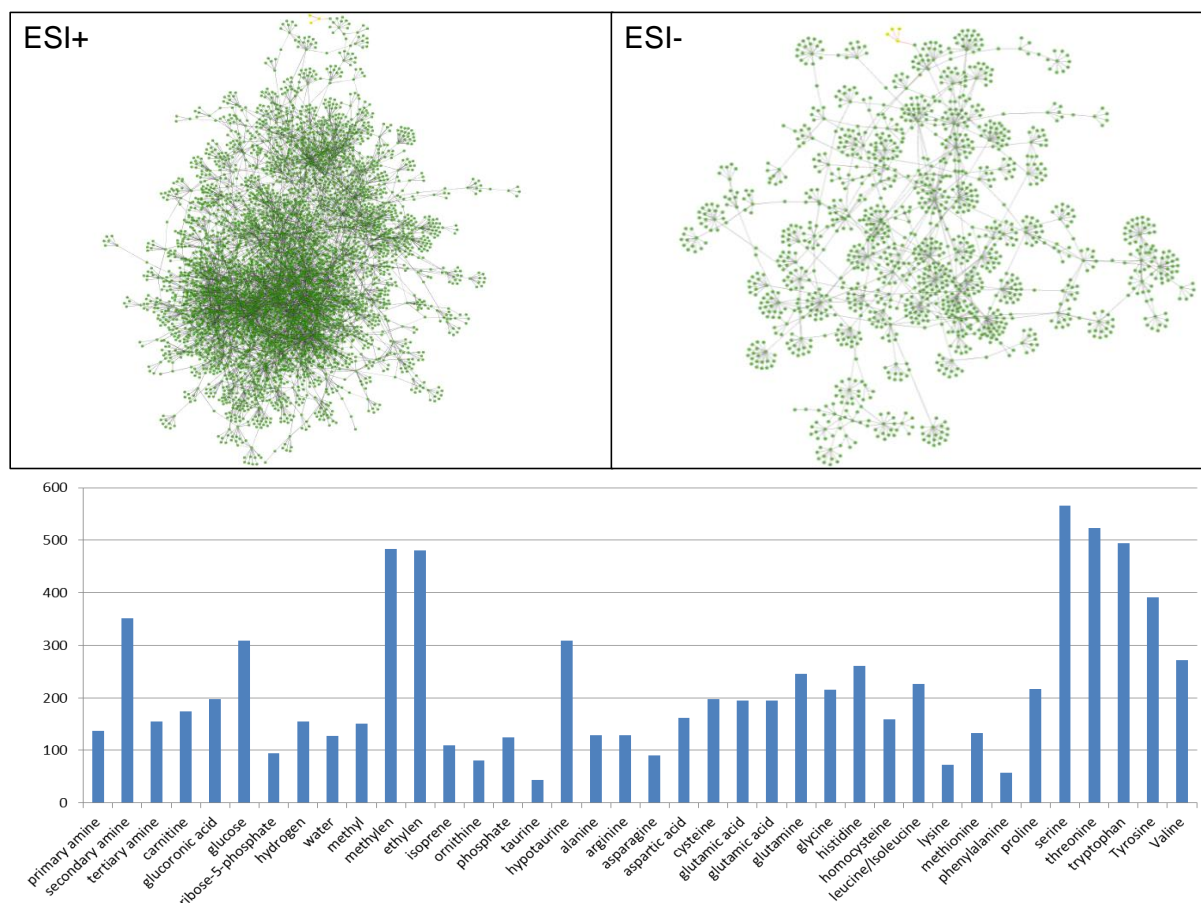
Apart from MA (chapter 3.4.2.4) the vitamin B<sub>6</sub> pathway is also important for the differentiation between T2D patients with and without DN. As mentioned above vitamin B<sub>6</sub> is an important cofactor in many other pathways such as in amino acid as well as the glyoxylate and dicarboxylate metabolism. In the latter case vitamin B<sub>6</sub> shifts the conversion of glyoxylate from oxalate to glycine thereby reducing the tubular damaging effects of high urinary oxalate concentrations [Rathi *et al.*, 2007]. Vitamin B<sub>6</sub> and its metabolite pyridoxic acid show significant changes in the development of CKD make the pathway an interesting target for research [Busch *et al.*, 2010]. In experiments with rats it could be demonstrated that the vitamin B<sub>6</sub> metabolite pyrodoxamine 5'-phosphate prevents progression of DN [Nakamura *et al.*, 2007].

#### 4.4.3 Unknown compounds

Using the highly accurate masses from the FT-ICR-MS analysis, a mass-network analysis was applied for annotation of chemical formulas to significant masses without annotation. From the 791 significant signals 37% could be annotated using online databases whereas including the formulas from the network analysis 56% of the signals could be covered. The supplementary table A19 presents the elemental compositions of all significant signals from FT-ICR-MS and UHPLC-MS analysis with a fold change  $\geq 100\%$  between the control and the MA group.

Additionally the connectivity of the signals were analysed in order to identify possible key metabolites. Highly cross-linked signals could represent an important regulatory metabolite which is part of many different pathways. Alteration in such metabolites could explain the various and therefore complicated metabolic change in DN patients.

To reduce the complexity of the mass networks only signals with a direct link to a significant mass were shown (Figure 61)



**Figure 61: Kendrick-analogous mass difference network of the PREDICTIONS study FT-ICR-MS data**

To reduce the complexity the network was reduced to direct connections with a result signal. Detected masses are visualised as dots while the connections between two dots resemble distinct mass differences or transformations. In positive mode a big network was formed while in negative a smaller one was observed. Apart from the shown main networks also smaller networks could be found. Below the networks the combined frequency of the different transformations, addition or subtraction of mentioned molecules, of both ionisation modes are mapped.

Comparing the frequency of the different transformations reveals the great impact of amino acids. Especially the amount of serine, threonine, tryptophan and tyrosine are increased. The high methylene and ethylene counts indicate presence of fatty acids in free form or as lipids but are also artefacts of certain amino acid transformations like peptides which can lose either a serine or threonine for example. Compared to the MA dataset the amount of amino acid transformation increased because of the overt proteinuria dramatically. The increased number of secondary amine transformation can also be explained by the protein load. Transformations involving sugars (glucose) were low in the MA dataset but are raised now. Higher concentrations of AGE which partly lose the sugar in the ionisation process are a plausible explanation for these findings and are backed by the also increased amounts of hypotaurine as taurine is thought a countermeasure against AGE [Chesney *et al.*, 2010].

The higher sensitivity of the Bruker solriX system compared to its predecessor the APEX system, is also visible in the increased connectivity between masses due to the increased number of detected masses. Thus, only the very highly connected masses with a connectivity  $\geq 30$  were extracted and are presented in Table 5.

<b>m/z</b>	<b>elemental composition</b>	<b>mass network connectivity</b>
359.1832	C <sub>17</sub> H <sub>22</sub> O <sub>3</sub> N <sub>6</sub>	47
432.0743	C <sub>12</sub> H <sub>21</sub> O <sub>1</sub> N <sub>3</sub> S <sub>2</sub>	46
383.1269	C <sub>17</sub> H <sub>22</sub> O <sub>6</sub> N <sub>2</sub> S	45
406.1974	C <sub>20</sub> H <sub>27</sub> N <sub>3</sub> O <sub>6</sub>	43
528.2558	C <sub>24</sub> H <sub>37</sub> O <sub>1</sub> N <sub>3</sub>	43
551.3085	C <sub>28</sub> H <sub>38</sub> N <sub>8</sub> O <sub>4</sub>	42
304.1182	C <sub>16</sub> H <sub>18</sub> O <sub>5</sub> N	39
269.1246	C <sub>11</sub> H <sub>16</sub> N <sub>4</sub> O <sub>4</sub>	38
526.2127	C <sub>27</sub> H <sub>34</sub> O <sub>9</sub>	37
479.2279	C <sub>25</sub> H <sub>34</sub> O <sub>9</sub>	35
401.2012	C <sub>41</sub> H <sub>56</sub> N <sub>8</sub> O <sub>9</sub>	34
300.1596	C <sub>18</sub> H <sub>21</sub> NO <sub>3</sub>	34
369.1614	C <sub>12</sub> H <sub>25</sub> O <sub>9</sub> N <sub>4</sub>	34
581.2611	C <sub>30</sub> H <sub>36</sub> O <sub>8</sub> N <sub>4</sub>	34
303.134	C <sub>16</sub> H <sub>19</sub> O <sub>4</sub> N <sub>2</sub>	33
522.2662	C <sub>22</sub> H <sub>39</sub> O <sub>11</sub> N <sub>3</sub>	33
645.2852	C <sub>23</sub> H <sub>45</sub> O <sub>13</sub> N <sub>6</sub> P	33
552.0955	C <sub>19</sub> H <sub>26</sub> O <sub>12</sub> N <sub>3</sub> S <sub>2</sub>	32
165.0547	C <sub>9</sub> H <sub>8</sub> O <sub>3</sub>	32
467.2503	C <sub>22</sub> H <sub>34</sub> O <sub>7</sub> N <sub>4</sub>	31



m/z	elemental composition	mass network connectivity
379.2457	C <sub>20</sub> H <sub>36</sub> O <sub>5</sub>	30
761.2089	C <sub>30</sub> H <sub>41</sub> O <sub>15</sub> N <sub>4</sub> PS	30
253.1412	C <sub>12</sub> H <sub>22</sub> O <sub>4</sub>	30
268.1042	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	30
330.06	C <sub>10</sub> H <sub>12</sub> N <sub>5</sub> O <sub>6</sub> P	30

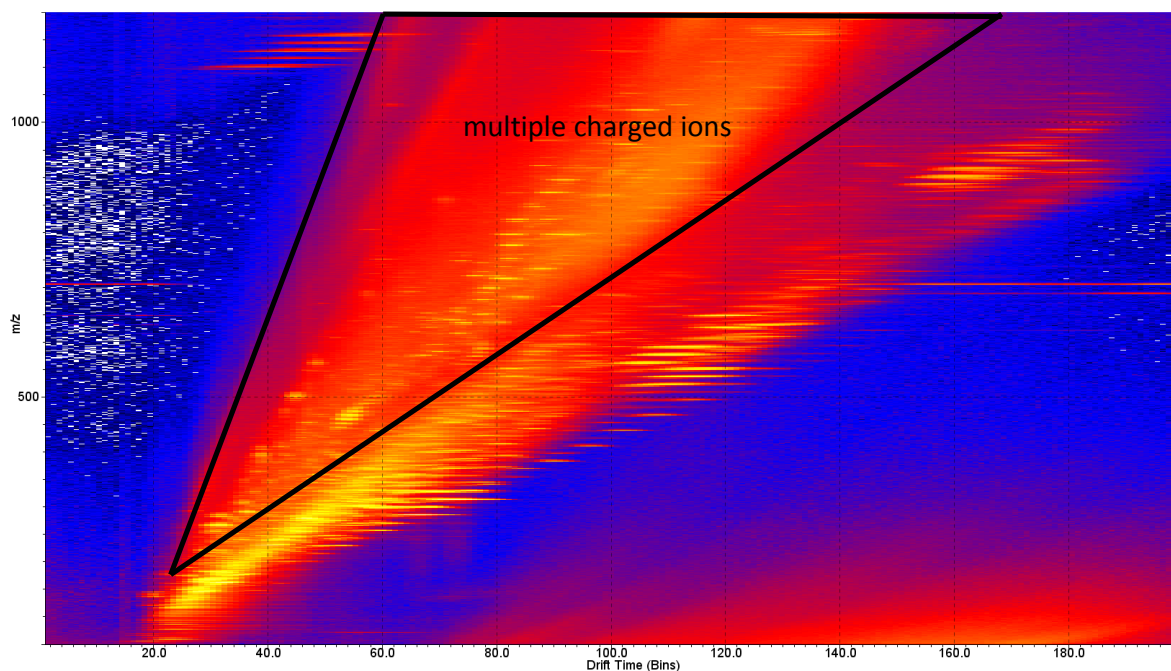
**Table 5: Highly connected significant masses in the PREDICTIONS study.**

Masses were filtered for a valid elemental composition and a network connectivity  $\geq 30$ .

Due to their high connectivity these masses could represent key metabolites connecting and regulating different pathways. Thus they are interesting targets for further research.

## 4.5 Ion mobility analysis

Since there was no automated way to extract all IMS peaks from the dataset and generate a matrix, only three random samples and a QC sample were measured with this method for evaluation of its value. **Figure 62** shows the drift time chromatogram of the QC sample and the huge amount of multiple charged ions.

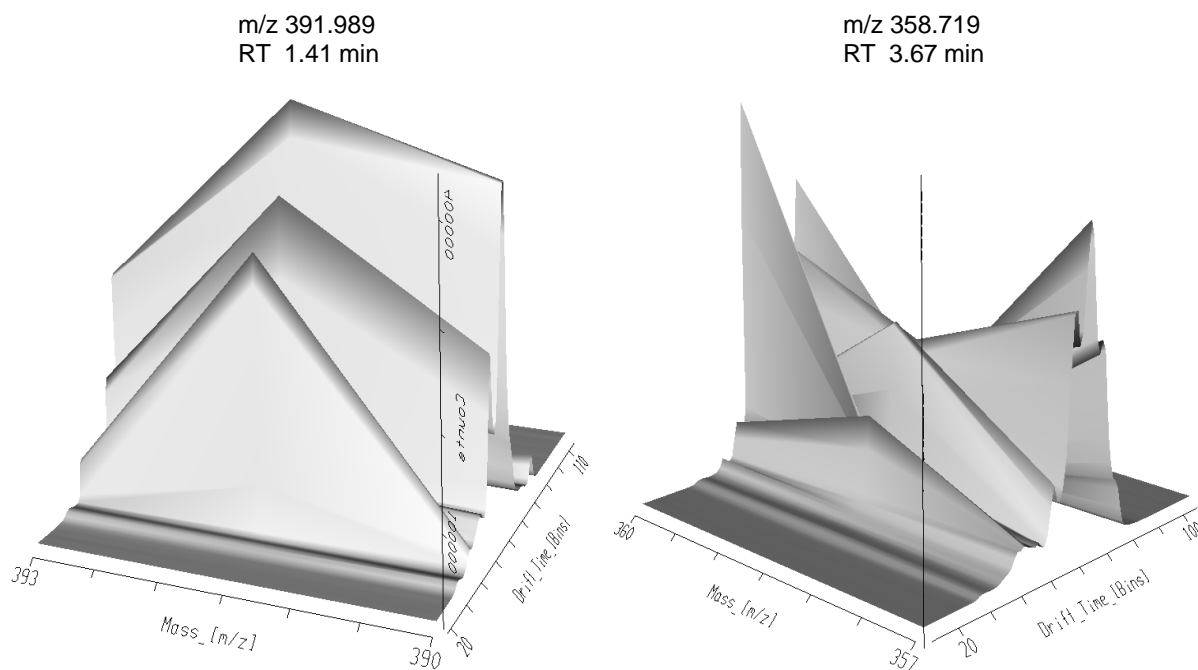


**Figure 62: Ion mobility drift time chromatogram in ESI+ mode**

Peak intensities are indicated by color, ranging from blue for low to yellow for very high peak intensities. Applying IMS analysis, the huge amount of multiple charged ions is visible.

A first big advantage of IMS is the relatively easy differentiation between single and multiple charged ions. Thus, IMS could be a useful tool for optimisation of the electrospray parameters in order to reduce multiple charges for an increase in data quality even before measurement. The biggest advantage of IMS is the separation of ions with the same mass and similar retention times but different molecular shapes.

For the search of the identified significant signals within the IMS data first a peak piking using a signal intensity threshold of 250 counts was applied before the comparison. Significant peaks of the ESI<sup>+</sup>-LC-MS data were searched within the IMS data. 11.5 % of the significant peaks show more than one peak in the drift time chromatogram within a mass window of 0.02 m/z, a RT window of 0.2 min (peak piking window) and a minimum IMS separation of 10 Bins (Figure 63). A Bin is a whole mass spectrum of a time point in the drift time chromatogram and each drift time chromatogram consists of 200 Bins in total.



**Figure 63: IMS separation of two signals of the ESI<sup>+</sup> LC-MS dataset identified as significant**

The peaks with the same  $m/z$  and RT but different drift times are most likely isomers with only slight differences in their structure but could also be metabolites from a different metabolic class with similar retention behaviour. Within IMS however the small structural differences result in a drift time separation based on their different molecular cross sections.

Using standards with different known cross sections, a calibration of the IMS system is possible allowing the calculation of cross sections for unknown compounds. These results allow narrowing down the number of annotation by removing hits with divergent calculated molecular sizes thus improving the annotation and identification process.

An experimental analysis of a mixture of standards via direct injection in both ionisation modes also shows a separation between different metabolite classes (Figure 64). Thus even if two unknown peaks have the same RT and  $m/z$ , the drift time might give an additional hint of their identity.

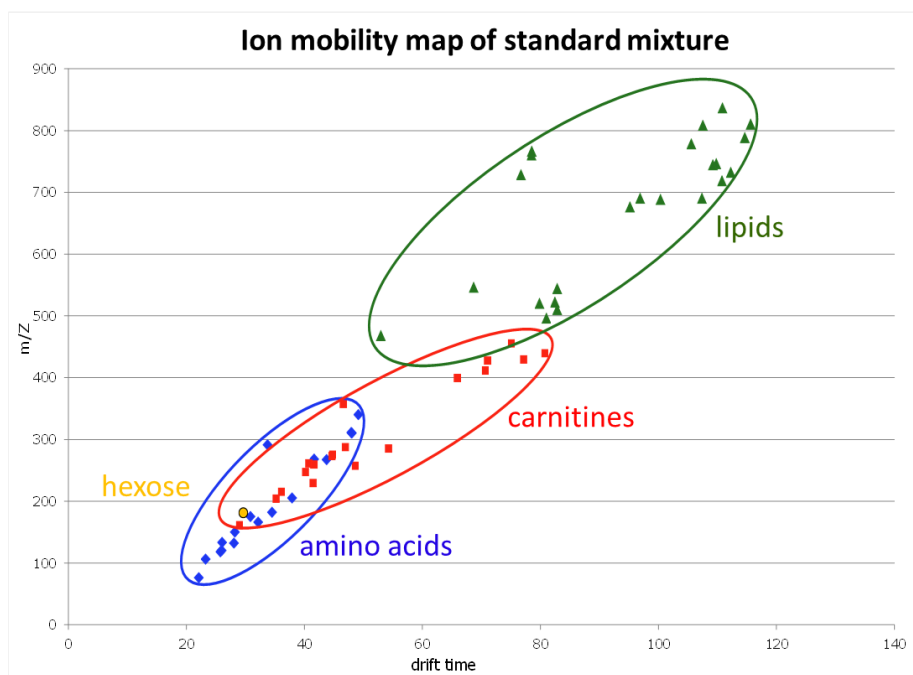


Figure 64: Mapping of metabolite classes by direct injection of a standard mixture with IMS

This result indicates the great opportunity the integration of IMS can provide. Not only metabolites with the same mass and RT behaviour can be separated if they differ in the molecular cross section, with a proper size calibration the annotation can be done more accurately. The combination of drift time, RT and m/z of standard substances in one database for comparison will further improve the annotation and identification process and is therefore a desirable objective for future analysis.

Current drawbacks of IMS are the loss of sensitivity comparing to simple MS measurements, the huge amount of data per sample as well as the lack of automated peak piking and alignment tools.

## 4.6 Conclusion

73 human urine samples of diabetes patient's origin from the PREDICTIONS study were analysed applying up to date MS platforms. The implementation of IMS offers a promising new dimension in LC-MS analysis. Still improvements especially in data processing are necessary for an implementation in high throughput platforms.

After the method development a general overview of the obtained data was presented. The four datasets obtained from LC-MS and FT-ICR-MS were compared before discriminative metabolites were annotated and forwarded to a pathway analysis filtering the most important pathways.

Through comparisons between a control and a nephropathy group already known pathways could be confirmed and new ones were found. Through correlation analysis citric acid and its fragments could be identified. Other masses could be assigned to peptides and their fragments but a clear identification was not possible. Thus combination with proteomics results of these samples could improve the picture. Non-invasively available urinary metabolites from butanoate metabolism, propanoate pathway as well as the citrate cycle and peptides could be interesting targets for the development of a simple and robust diagnostic kit.

Applying a mass network analysis on the high resolution FT-ICR-MS data identified several masses with a very high connectivity offering interesting targets for identification and future research.

## 5 Concluding remarks

The results of the TULIP and PREDICTIONS study presented in this work accent the advantages of non-targeted metabolomics approaches. Novel prognostic and diagnostic biomarkers in the pathogenesis of diabetic nephropathy were identified and gave new insights over the involved pathways. Several pathways were found important in both studies offering therefore novel targets for predicting the onset and monitoring the progression of DN. Thus, metabolomics has been proven to be a powerful tool for biomarker screening and generation of new hypothesis.

In order to approach the goal of metabolomics, the complete detection of all present metabolites, a multi-platform concept is needed to detect as many compounds as possible in a given sample. Therefore, state-of-the-art FT-ICR-MS, IMS, LC-MS and NMR techniques have been applied in the process of this work. While direct injection FT-ICR-MS with its tremendous mass accuracy allows the detection and annotation of thousands of metabolites, the UHPLC-Tof-MS offers the possibility of verification and identification of metabolites through comparison with the retention time from standards or with the MS/MS pattern from databanks. The huge amount of data generated by the different methods was handled through separate data analysis workflows for each method before integration. Applying a correlation analysis several metabolites and their fragments could be identified. Citric acid was thereby identified in both studies as a significant metabolite.

The amino acid pathways, the caffeine metabolism, the vitamin B<sub>6</sub> metabolism and the citrate cycle have been found altered in both MA and DN patients. These pathways are thus promising targets for predicting the onset and monitoring the progression of DN. Another future task will be the identification of the “unknown” metabolites as well as the confirmation of the mentioned biomarkers in large cohorts with targeted approaches. These studies might also include the monitoring of the metabolites over several stages of DN. Especially the structure elucidation of the signals without annotations offers new possibilities for both diagnostic and therapeutic applications. Additionally the integration of metabolomics, proteomics, transcriptomics and genomic data will be an important step towards completely understanding the complex pathogenesis mechanisms in T2D and DN.

## 6 Appendix

### Supplementary tables:

Compound	Producer / Distributor
Caffeine	Sigma-Aldrich, St. Luis, USA
Creatine	Sigma-Aldrich, St. Luis, USA
Decanoic acid	Sigma-Aldrich, St. Luis, USA
[C13] Decanoic acid	Sigma-Aldrich, St. Luis, USA
Leucine-Enkephaline	Waters GmbH, Eschborn, Germany
Nialamide	Sigma-Aldrich, St. Luis, USA
Reserpine	Sigma-Aldrich, St. Luis, USA
Sodium 3-(trimethylsilyl)propanoic acid	
Sulfadimethoxine	Sigma-Aldrich, St. Luis, USA
Terfenandine	Sigma-Aldrich, St. Luis, USA
Valeryl-L-carnitine + HCL	Dr. Herman J. ten Brink, VU medical center, Netherlands
[D3] Valeryl-L-carnitine + HCL	Dr. Herman J. ten Brink, VU medical center, Netherlands

Table A1: Standard Compounds for method development, validation and QC used in the TULIP study

Part	Product	Producer / Distributor
SPE Cartridge	Oasis® HLB 1cc Vac cartridge, 30 mg Sorbent,	Waters GmbH, Eschborn, Germany
Column	ACQUITY® UPLC® HSS-T3 C18 Column, 1.0 x 150 mm	Waters GmbH, Eschborn, Germany
Column	Grace® VisionHT™ C18, 2.0 x 150 mm	Alltech Grom GmbH, Worms, Germany
Column	Grace® VisionHT™ C18-HL, 2.0 x 150 mm	Alltech Grom GmbH, Worms, Germany
Column	Grace® VisionHT™ C18-P, 2.0 x 150 mm	Alltech Grom GmbH, Worms, Germany
Column	Grace® VisionHT™ C18-HL, guard column, 2.0 x 5.0 mm	Alltech Grom GmbH, Worms, Germany

Table A2: Used consumables for the method and the development in the TULIP study

<b>Nr.</b>	<b>Gender</b>	<b>Class</b>	<b>Albumin [mg/g creatinine]</b>	<b>Albumin [mg/L]</b>
1	0	control	3.2	3
2	0	control	7.5	2.26
3	0	control	9.9	2.57
4	0	control	8.6	3
5	0	control	4.7	2
6	0	control	4.2	4
7	1	control	4	7.1
8	0	control	3.3	3
9	0	control	4.1	3
10	0	control	4.7	7
11	0	control	6.7	7
12	0	control	4.2	3
13	0	control	4.5	7
14	0	control	3.5	3
15	0	control	5.9	6
16	0	control	2.9	4
17	0	control	4.7	2
18	0	control	4.2	4
19	0	control	9.1	2
20	0	control	5.3	7
21	0	control	3	4
22	0	microalbuminuria	70	84
23	0	microalbuminuria	127.5	371
24	1	microalbuminuria	77.9	155
25	1	microalbuminuria	52.2	153
26	0	microalbuminuria	49.6	58
27	0	microalbuminuria	64.9	61
28	0	microalbuminuria	54.2	58
29	0	microalbuminuria	60.9	42
30	0	microalbuminuria	68.3	235
31	0	microalbuminuria	39.9	69
32	1	microalbuminuria	149.5	154
33	1	microalbuminuria	87.7	164
34	0	microalbuminuria	51.2	62
35	0	microalbuminuria	96	48
36	0	microalbuminuria	309.6	356
37	0	microalbuminuria	46.7	64
38	0	microalbuminuria	70.8	85

Table A3: Information and classification of the 38 TULIP human urine samples



Time [min]	Solvent A (5% MeOH + 0.1% FA)	Solvent B (100% MeOH)
0	100%	0%
0.5	100%	0%
22	50%	50%
25	0%	100%
28.9	0%	100%
29	100%	0%
32	100%	0%

Table A4: Applied chromatographic method for column test

Injection parameters	
Loop volume [μl]	10
Injection volume[μl]	5.0
Injection type	PLNO
Weak wash	H2O/MeOH: 9/1
V <sub>weak wash</sub> [μl]	500
Strong wash	MeOH
V <sub>strong wash</sub> [μl]	200
Applied advanced parameters	
Syringe draw rate [μl/min]	100
Air Gaps:	
pre-aspirate [μl]	1.0
post-aspirate [μl]	0.0

Table A5: Injection parameters for the UHPLC method

Detection Parameters	Polarity	
	ESI <sup>+</sup>	ESI <sup>-</sup>
Capillary voltage [kV]	3.1	2.3
Sample cone [V]	30	35
Extraction cone [V]	4	6
Source temperature [°C]	120	
Desolvatation temperature [°C]	300	
Desolvatation gas flow [L/h]	800	
Cone gas [L/h]	50	
Detector voltage [V]	1750	

Table A6: Optimised parameters of the ToF

Name	Stocksolution (1 mg/ml)	Equivalents for standardmixture	Concentration in sample [µg/ml]
[13C]-Decanoic acid	Acetonitrile	5	5.0
Nialamide:	Acetonitrile	3	3.0
[D3] Valeryl-L-carnitine	Methanol	5	5.0
Reserpine	Acetonitrile	2	2.0
Terfenandine	Methanol	2	2.0

Table A7: Equivalents of internal standards stock solution in the standard mixture and final concentration after spiking the sample urine.

	Property	Value
	Initial retention time [min]	1
	Final retention time [min]	32
	Low mass [Da]	50
	High mass [Da]	1000
	Mass tolerance [Da]	0.05
Apex track peak parameters	Peak width 5% height [s]	auto
	Peak-to-peak baseline noise	auto
	Smoothing	Yes
Collection parameters	Intensity threshold [counts]	100
	Mass window [Da]	0.05
	RT window [min]	0.25
	Noise elimination lvl	auto
	Deisotope data	Yes

Table A8: Optimised settings for peak detection and data matrix generation within MarkerLynx™ within the TULIP study.

Parameter	ESI +	ESI -
Number of scans	300	300
Transient time domain [MW]	2	2
Source accumulation [s]	0.0	0.0
Ion accumulation time [s]	0.2	0.4
Time of flight [s]	0.0	0.0
Dwell time [s]	0.100	0.100
Capillary voltage [kV]	4.5	4.0
Dry gas flow [l/min]	2.4	2.5
Dry gas temperature [°C]	200	200
Nebulizer gas flow [l/min]	1.1	1.3
Side kick [V]	-5.0	5.0
Side kick off set [V]	-5.0	5.0
Front plate voltage [V]	0.7	-0.7
Back plate voltage [V]	0.9	-0.9
Analyzer entrance [V]	-10.0	10.0
Broadband high mass [Da]	1000	1000
Broadband low mass [Da]	122.9	122.9

Table A9: Optimised parameters for the FT-ICR-MS measurement of the TULIP samples

ID	Biological change	Change in molecular formula	mass difference [Da]
1	Phosphate	HPO3	79.96633
2	Tertiary amine	N	14.00307
3	Secondary amine	NH	15.01090
4	Primary amine	NH2	16.01872
5	Hydrogen	H2	2.01565
6	Water	H2O	18.01057
7	Methyl	CH3	15.02348
8	Methylen	CH2	14.01565
9	Ethylen	C2H4	28.03130
10	Isoprene	C5H8	68.06260
11	Ribose-5-phosphate	C5H9O7P	212.00859
12	Glucose	C6H10O5	162.05282
13	Glucuronsäure	C6H8O6	176.03209

ID	Biological change	Change in molecular formula	mass difference [Da]
14	Taurine	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> S	107.00410
15	Hypotaurine	C <sub>2</sub> H <sub>5</sub> NOS	91.00918
16	Ornithine	C <sub>5</sub> H <sub>10</sub> N <sub>2</sub> O	114.07931
17	Carnitine	C <sub>7</sub> H <sub>14</sub> NO <sub>2</sub>	160.09737
18	Glycine	C <sub>2</sub> H <sub>3</sub> NO	57.02149
19	Cysteine	C <sub>3</sub> H <sub>5</sub> NOS	103.00918
20	Homocysteine	C <sub>4</sub> H <sub>7</sub> NOS	117.02484
21	Alanine	C <sub>3</sub> H <sub>5</sub> NO	71.03711
22	Glutamic acid	C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>	129.04259
23	Arginine	C <sub>6</sub> H <sub>12</sub> N <sub>4</sub> O	156.10111
24	Asparagine	C <sub>4</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub>	114.04293
25	Aspartic acid	C <sub>4</sub> H <sub>5</sub> NO <sub>3</sub>	115.02694
27	Glutamine	C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>	128.05858
28	Histidine	C <sub>6</sub> H <sub>7</sub> NO <sub>3</sub>	137.05891
29	Leucine/Isoleucine	C <sub>6</sub> H <sub>11</sub> NO	113.08406
30	Lysine	C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> O	128.09496
31	Methionine	C <sub>5</sub> H <sub>9</sub> NOS	131.04049
32	Phenylalanine	C <sub>9</sub> H <sub>9</sub> NO	147.06841
33	Proline	C <sub>5</sub> H <sub>7</sub> NO	97.05276
34	Serine	C <sub>3</sub> H <sub>5</sub> NO <sub>2</sub>	87.03203
35	Threonine	C <sub>4</sub> H <sub>7</sub> NO <sub>2</sub>	101.04768
36	Tryptophan	C <sub>11</sub> H <sub>10</sub> N <sub>2</sub> O	186.07931
37	Tyrosine	C <sub>9</sub> H <sub>9</sub> NO <sub>2</sub>	163.06333
38	Valine	C <sub>5</sub> H <sub>9</sub> NO	99.06841

Table A10: Biological transformation mass list for Netcalc approach

Mode	start ion	calculated ion	mass modification [Da]
ESI+	M + H <sup>+</sup>	M + Na <sup>+</sup>	21.98195
	M + H <sup>+</sup>	M + NH <sub>4</sub> <sup>+</sup>	17.02654
	M + H <sup>+</sup>	M + H <sup>+</sup> - H <sub>2</sub> O	-17.00384
	M + Na <sup>+</sup>	M + H <sup>+</sup>	-21.98195
	M + Na <sup>+</sup>	M + NH <sub>4</sub> <sup>+</sup>	-4.9554
	M + NH <sub>4</sub> <sup>+</sup>	M + H <sup>+</sup>	-17.02654
	M + NH <sub>4</sub> <sup>+</sup>	M + Na <sup>+</sup>	4.9554
ESI-	M - H <sup>+</sup>	M - H <sub>2</sub> O - H <sup>+</sup>	-18.01111

Mode	start ion	calculated ion	mass modification [Da]
	M - H <sup>+</sup>	M + FA - H <sup>+</sup>	46.00548
	M - H <sub>2</sub> O - H <sup>+</sup>	M - H <sup>+</sup>	18.01111
	M - H <sub>2</sub> O - H <sup>+</sup>	M + FA - H <sup>+</sup>	64.01659
	M + FA - H <sup>+</sup>	M - H <sup>+</sup>	-46.00548
	M + FA - H <sup>+</sup>	M - H <sub>2</sub> O - H <sup>+</sup>	-64.01659

Table A11: Investigated ions and their mass difference for the matching of FT-ICR-MS with UHPLC-MS masses

Data normalisation with urinary creatinine level	Data normalisation with NMR scaling factor	
<u>Propenyl-L-carnitine</u>	Decenoyl-L-carnitine	<u>Phosphatidylcholine diacyl C 40:3</u>
<u>Hexenoyl-L-carnitine</u>	Dodecenoyl-L-carnitine	<u>Phosphatidylcholine diacyl C 42:2</u>
<u>Glutamine</u>	Hydroxytetradecadienyl-L-carnitine	<u>Phosphatidylcholine diacyl C 42:6</u>
<u>Phosphatidylcholine diacyl C 30:2</u>	Hexadecenoyl-L-carnitine	* <u>Phosphatidylcholine acyl-alkyl C 36:2</u>
* <u>Phosphatidylcholine diacyl C 34:1</u>	Hydroxyhexadecenoyl-L-carnitine	<u>Phosphatidylcholine acyl-alkyl C 38:0</u>
* <u>Phosphatidylcholine diacyl C 34:2</u>	Octadecenoyl-L-carnitine	* <u>Phosphatidylcholine acyl-alkyl C 38:4</u>
* <u>Phosphatidylcholine diacyl C 36:3</u>	Hydroxyoctadecenoyl-L-carnitine	Phosphatidylcholine acyl-alkyl C 38:5
* <u>Phosphatidylcholine diacyl C 36:4</u>	Hydroxypropionyl-L-carnitine	Phosphatidylcholine acyl-alkyl C 42:1
* <u>Phosphatidylcholine diacyl C 38:6</u>	<u>Propenyl-L-carnitine</u>	<u>Phosphatidylcholine acyl-alkyl C 44:3</u>
<u>Phosphatidylcholine diacyl C 40:3</u>	Fumaryl-L-carnitine / Hexanoyl-L-carnitine	<u>Phosphatidylcholine acyl-alkyl C 44:4</u>
<u>Phosphatidylcholine diacyl C 42:2</u>	<u>Hexenoyl-L-carnitine</u>	<u>lysoPhosphatidylcholine acyl C16:0</u>
<u>Phosphatidylcholine diacyl C 42:6</u>	<u>Glutamine</u>	lysoPhosphatidylcholine acyl C17:0
* <u>Phosphatidylcholine acyl-alkyl C 36:2</u>	Glycine	<u>lysoPhosphatidylcholine acyl C18:0</u>
<u>Phosphatidylcholine acyl-alkyl C 38:0</u>	Phosphatidylcholine diacyl C 28:1	* <u>Hydroxyshingomyeline C 22:2</u>

Data normalisation with urinary creatinine level	Data normalisation with NMR scaling factor	
<u>* Phosphatidylcholine acyl-alkyl C 38:4</u>	<u>Phosphatidylcholine diacyl C 30:2</u>	Hydroxysphingomyeline C 24:1
<u>Phosphatidylcholine acyl-alkyl C 44:3</u>	<u>* Phosphatidylcholine diacyl C 34:1</u>	
<u>Phosphatidylcholine acyl-alkyl C 44:4</u>	<u>* Phosphatidylcholine diacyl C 34:2</u>	
<u>lysoPhosphatidylcholine acyl C16:0</u>	<u>* Phosphatidylcholine diacyl C 36:3</u>	
<u>lysoPhosphatidylcholine acyl C18:0</u>	<u>* Phosphatidylcholine diacyl C 36:4</u>	
<u>* Hydroxysphingo-myeline C 22:2</u>	<u>* Phosphatidylcholine diacyl C 38:6</u>	

Table A12: Significant metabolites identified with the biocrates kit. Underlined metabolites were found significant with both normalisations. \* = elevated in patients with microalbuminuria

monoisotopic mass	elemental composition	log ratio MA/C	method
60.0211284	C2H4O2	0.78	ESI- LC-MS
60.0323614	CH4N2O	0.98	ESI- LC-MS
83.048345	C3H5N3	-0.33	ESI+ LC-MS
99.0684104	C5H9NO	-0.67	ESI+ LC-MS
105.0326958	C5H3N3	-0.33	ESI+ LC-MS
106.062991	C4H10O3	0.36	ESI+ LC-MS
109.0527612	C6H7NO	-0.33	ESI+ LC-MS
112.0272764	C4H4N2O2	-0.33	ESI+ LC-MS
115.026942	C4H5NO3	0.63	ESI+ LC-MS
117.0789746	C5H11NO2	-0.74	ESI+ LC-MS
118.062991	C5H10O3	1.01	ESI+ LC-MS
122.0480106	C6H6N2O	0.60	ESI+ LC-MS
123.032027	C6H5NO2	-1.57	ESI+ LC-MS
128.0473418	C6H8O3	-0.96	ESI- LC-MS
128.0585748	C5H8N2O2	0.33	ESI- FT-ICR-MS
128.0837252	C7H12O2	0.80	ESI- LC-MS
131.0694734	C4H9N3O2	-0.54	ESI+ LC-MS
141.0248332	C6H7NOS	-0.35	ESI+ FT-ICR-MS
143.0734954	C10H9N	0.30	ESI+ LC-MS
148.0735552	C6H12O4	-1.80	ESI+ LC-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
154.062991	C8H10O3	0.49	ESI+ LC-MS
156.0422568	C7H8O4	0.93	ESI- FT-ICR-MS
157.0738896	C7H11NO3	0.56	ESI+ LC-MS
158.0327556	C5H6N2O4	0.40	ESI+ LC-MS
159.1007718	C6H13N3O2	-0.48	ESI+ FT-ICR-MS
165.0789746	C9H11NO2	0.31	ESI+ LC-MS
170.1306728	C10H18O2	0.45	ESI+ LC-MS
172.1099386	C9H16O3	0.42	ESI+ LC-MS
174.0892044	C8H14O4	only in MA	ESI+ LC-MS
182.0439886	C6H6N4O3	-0.58	ESI+ FT-ICR-MS
182.0942894	C10H14O3	0.45	ESI+ LC-MS
182.1055224	C9H14N2O2	0.65	ESI+ FT-ICR-MS
185.0800376	C7H11N3O3	0.43	ESI- FT-ICR-MS
190.0841194	C8H14O5	0.50	ESI- FT-ICR-MS
190.0993744	C12H14O2	0.63	ESI+ LC-MS
191.0616108	C7H13NO3S	0.41	ESI- FT-ICR-MS
192.0633852	C7H12O6	-0.44	ESI- LC-MS
194.0790344	C7H14O6	0.79	ESI- FT-ICR-MS
196.0735552	C10H12O4	-0.50	ESI- LC-MS
196.1211716	C10H16N2O2	0.31	ESI+ LC-MS
198.0892044	C10H14O4	0.31	ESI+ LC-MS
198.1368208	C10H18N2O2	0.41	ESI+ FT-ICR-MS
199.0480704	C8H9NO5	0.35	ESI- FT-ICR-MS
199.1109418	C12H13N3	0.35	ESI+ LC-MS
200.0085894	C4H9O7P	0.90	ESI+ FT-ICR-MS
204.0270018	C7H8O7	-0.42	ESI+ LC-MS
206.0062676	C6H6O8	-0.33	ESI+ LC-MS
208.0007884	C9H4O6	-0.45	ESI+ LC-MS
210.1255878	C12H18O3	only in C	ESI+ FT-ICR-MS
212.1061912	C12H12N4	0.51	ESI+ LC-MS
214.0953524	C9H14N2O4	0.31	ESI- FT-ICR-MS
216.1262566	C13H16N2O	0.70	ESI+ LC-MS
219.1623056	C14H21NO	0.65	ESI+ FT-ICR-MS
221.0324212	C10H7NO5	-0.31	ESI+ LC-MS
222.0939124	C10H14N4S	0.63	ESI+ LC-MS
225.0095782	C9H7NO4S	0.60	ESI- FT-ICR-MS
225.1014406	C12H11N5	0.94	ESI+ LC-MS
226.0217356	C10H11O2PS	-0.34	ESI+ FT-ICR-MS
228.1837684	C12H24N2O2	1.00	ESI+ FT-ICR-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
229.0690176	C10H16NOPS	-0.53	ESI+ FT-ICR-MS
233.072175	C9H15NO4S	0.57	ESI- FT-ICR-MS
233.1375494	C9H19N3O4	0.69	ESI+ FT-ICR-MS
234.035063	C12H10O3S	-0.35	ESI+ LC-MS
234.1619712	C15H22O2	-0.45	ESI+ LC-MS
236.1776204	C15H24O2	0.93	ESI+ LC-MS
237.0861856	C9H11N5O3	1.21	ESI+ LC-MS
238.0841194	C12H14O5	0.48	ESI- FT-ICR-MS
239.016985	C10H10NO2PS	-0.65	ESI+ FT-ICR-MS
243.0531554	C13H9NO4	-0.35	ESI+ LC-MS
244.131067	C12H20O5	0.53	ESI+ FT-ICR-MS
244.1786834	C12H24N2O3	0.69	ESI+ FT-ICR-MS
245.100316	C11H20NOPS	-0.59	ESI+ FT-ICR-MS
245.1779548	C16H23NO	2.08	ESI+ LC-MS
249.9870696	C10H6N2O2S2	-0.84	ESI+ LC-MS
250.0841194	C13H14O5	0.53	ESI- FT-ICR-MS
250.1065854	C11H14N4O3	0.42	ESI- FT-ICR-MS
250.1317358	C13H18N2O3	0.42	ESI+ LC-MS
251.0793688	C12H13NO5	-0.45	ESI+ FT-ICR-MS
252.0189344	C12H9ClO4	0.82	ESI+ FT-ICR-MS
253.07726	C12H15NO3S	0.43	ESI+ LC-MS
253.1361094	C11H19N5S	0.87	ESI+ LC-MS
260.0718406	C11H16O5S	0.91	ESI- FT-ICR-MS
261.084848	C10H15NO7	0.54	ESI- FT-ICR-MS
262.1099558	C9H18N4O3S	0.36	ESI+ FT-ICR-MS
264.0456272	C13H12O4S	-0.93	ESI+ LC-MS
264.0667556	C10H16O6S	0.50	ESI- FT-ICR-MS
266.1167554	C15H14N4O	0.99	ESI+ LC-MS
269.0275492	C11H12NO3PS	-0.37	ESI+ FT-ICR-MS
271.1208372	C16H17NO3	0.62	ESI+ LC-MS
272.0895986	C12H16O7	0.43	ESI+ LC-MS
274.1052478	C12H18O7	0.47	ESI- FT-ICR-MS
274.1892476	C13H26N2O4	0.65	ESI+ FT-ICR-MS
277.0381922	C9H7N7O2S	-0.42	ESI+ LC-MS
277.095018	C14H15NO5	0.34	ESI- FT-ICR-MS
277.0983884	C11H19NO5S	0.37	ESI- FT-ICR-MS
280.1059166	C13H16N2O5	1.04	ESI- LC-MS
281.112399	C11H15N5O4	-0.47	ESI+ LC-MS
283.0544314	C11H14N3O2PS	-0.30	ESI+ FT-ICR-MS



monoisotopic mass	elemental composition	log ratio MA/C	method
285.1364864	C17H19NO3	0.47	ESI+ FT-ICR-MS
285.2092532	C19H27NO	0.81	ESI+ LC-MS
286.0841194	C16H14O5	0.80	ESI+ LC-MS
288.0796416	C9H21O6PS	0.67	ESI+ FT-ICR-MS
289.0983884	C12H19NO5S	0.33	ESI- FT-ICR-MS
290.0824048	C12H18O6S	0.75	ESI- FT-ICR-MS
290.1154178	C16H18O5	0.53	ESI+ LC-MS
290.224568	C19H30O2	-0.72	ESI- LC-MS
291.1082982	C8H22NO8P	-0.55	ESI+ FT-ICR-MS
291.132663	C13H25NO2S2	0.33	ESI+ LC-MS
293.0647826	C12H11N3O6	1.22	ESI- FT-ICR-MS
293.069806	C7H20NO7PS	-0.31	ESI+ FT-ICR-MS
293.1263164	C15H19NO5	0.57	ESI- FT-ICR-MS
293.9897856	C8H11N2O4PS2	-0.36	ESI+ FT-ICR-MS
294.1103328	C15H18O6	0.37	ESI- FT-ICR-MS
294.1314612	C12H22O8	0.34	ESI- FT-ICR-MS
295.0431984	C13H14NO3PS	-0.55	ESI+ FT-ICR-MS
296.1232976	C11H16N6O4	0.76	ESI+ FT-ICR-MS
296.1623654	C16H24O5	0.51	ESI+ LC-MS
298.0111886	C10H11N4OPS2	1.20	ESI+ FT-ICR-MS
298.0899928	C10H18O10	0.31	ESI- FT-ICR-MS
298.1293668	C10H23N2O6P	1.68	ESI+ FT-ICR-MS
300.0147374	C10H9N2O7P	-0.71	ESI- FT-ICR-MS
300.120897	C14H20O7	0.33	ESI+ LC-MS
301.2001462	C14H27N3O4	1.13	ESI+ FT-ICR-MS
302.0391166	C13H11N4OPS	0.85	ESI+ FT-ICR-MS
302.224568	C20H30O2	0.57	ESI+ LC-MS
303.1140376	C13H21NO5S	0.33	ESI- FT-ICR-MS
304.1569034	C12H24N4O3S	0.77	ESI+ FT-ICR-MS
305.1586778	C12H23N3O6	0.75	ESI- FT-ICR-MS
305.1660702	C14H27NO4S	0.67	ESI+ FT-ICR-MS
306.0980684	C11H19N2O6P	0.85	ESI+ FT-ICR-MS
309.1576148	C16H23NO5	-1.03	ESI+ FT-ICR-MS
309.9847006	C8H11N2O5PS2	-0.57	ESI+ FT-ICR-MS
310.1065854	C16H14N4O3	0.31	ESI- FT-ICR-MS
310.1416312	C16H22O6	0.32	ESI- FT-ICR-MS
312.120897	C15H20O7	0.34	ESI- FT-ICR-MS
313.073238	C12H15N3O5S	0.55	ESI+ LC-MS
314.1127334	C14H14N6O3	0.67	ESI+ LC-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
315.179412	C14H25N3O5	0.60	ESI+ FT-ICR-MS
316.0919992	C13H12N6O4	-0.33	ESI- FT-ICR-MS
318.0124156	C13H6N2O8	0.46	ESI- FT-ICR-MS
318.1830996	C19H26O4	0.53	ESI+ LC-MS
318.204228	C16H30O6	0.33	ESI+ LC-MS
321.1068438	C13H23NO4S2	0.35	ESI+ LC-MS
323.014616	C10H15NO5P2S	0.54	ESI+ FT-ICR-MS
324.114372	C15H20N2O4S	0.43	ESI+ LC-MS
324.1572804	C17H24O6	0.43	ESI+ FT-ICR-MS
325.079763	C14H15NO8	1.34	ESI+ LC-MS
326.1226288	C13H18N4O6	0.54	ESI+ LC-MS
326.1365462	C16H22O7	0.68	ESI+ LC-MS
326.1881846	C21H26O3	only in C	ESI+ LC-MS
327.1430286	C14H21N3O6	0.66	ESI- FT-ICR-MS
330.1903106	C14H26N4O5	0.84	ESI+ FT-ICR-MS
331.0391764	C11H14N3O5PS	0.80	ESI+ FT-ICR-MS
331.174327	C14H25N3O6	0.45	ESI+ FT-ICR-MS
334.069352	C19H15N2PS	0.81	ESI+ FT-ICR-MS
334.0929834	C12H19N2O7P	0.62	ESI+ FT-ICR-MS
335.1497666	C14H26NO6P	only in MA	ESI+ FT-ICR-MS
335.166739	C17H25N3O2S	0.33	ESI- FT-ICR-MS
338.2456964	C20H34O4	0.68	ESI+ LC-MS
339.1178782	C13H17N5O6	0.40	ESI- FT-ICR-MS
339.183434	C21H25NO3	0.56	ESI+ LC-MS
341.0283992	C15H7N3O7	0.49	ESI- FT-ICR-MS
341.0317696	C12H11N3O7S	0.56	ESI- FT-ICR-MS
341.1321898	C12H23NO10	0.39	ESI- FT-ICR-MS
341.1950612	C16H27N3O5	1.10	ESI- FT-ICR-MS
341.2062942	C15H27N5O4	0.46	ESI+ FT-ICR-MS
343.163094	C16H25NO7	0.70	ESI- FT-ICR-MS
344.0635982	C18H17O3PS	-0.57	ESI+ FT-ICR-MS
346.1416312	C19H22O6	0.54	ESI+ FT-ICR-MS
347.1368806	C18H21NO6	1.51	ESI+ LC-MS
349.1426344	C20H19N3O3	0.66	ESI+ LC-MS
350.1187882	C18H22O5S	-0.32	ESI+ LC-MS
350.1226288	C15H18N4O6	0.43	ESI- FT-ICR-MS
350.194058	C16H30O8	0.58	ESI- FT-ICR-MS
351.1708588	C21H23N2O3	0.63	ESI+ LC-MS
352.1456704	C17H24N2O4S	0.32	ESI+ LC-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
353.1967142	C15H32NO6P	1.08	ESI+ FT-ICR-MS
353.2062942	C16H27N5O4	0.35	ESI+ FT-ICR-MS
353.2314446	C18H31N3O4	0.61	ESI+ FT-ICR-MS
353.9728132	C10H12O8P2S	0.72	ESI+ FT-ICR-MS
354.1678446	C18H26O7	1.68	ESI- FT-ICR-MS
355.1119432	C15H22N3O3PS	0.32	ESI+ FT-ICR-MS
355.1508296	C12H26N3O7P	only in MA	ESI+ FT-ICR-MS
357.2303816	C22H31NO3	0.32	ESI+ LC-MS
357.98171	C13H10O8S2	0.30	ESI- FT-ICR-MS
358.1416312	C20H22O6	0.94	ESI+ LC-MS
359.2307758	C18H33NO6	0.92	ESI+ FT-ICR-MS
360.139341	C12H20N6O7	0.59	ESI+ FT-ICR-MS
360.9674582	C10H7N3O8S2	1.09	ESI- FT-ICR-MS
361.1096214	C17H19N3O4S	0.46	ESI- FT-ICR-MS
364.131753	C15H20N6O3S	0.43	ESI+ FT-ICR-MS
364.1885788	C20H28O6	-0.90	ESI- LC-MS
365.1335274	C15H19N5O6	0.50	ESI- FT-ICR-MS
366.0620648	C13H18O10S	1.26	ESI- FT-ICR-MS
366.1678446	C19H26O7	-0.31	ESI+ LC-MS
366.2406114	C21H34O5	-0.40	ESI+ FT-ICR-MS
367.2219434	C17H29N5O4	0.55	ESI+ FT-ICR-MS
368.0678186	C15H16N2O7S	0.54	ESI- FT-ICR-MS
369.1275924	C16H24N3O3PS	-0.39	ESI+ FT-ICR-MS
369.2151266	C19H31NO6	0.38	ESI+ FT-ICR-MS
370.0899928	C16H18O10	0.60	ESI+ LC-MS
370.2579924	C18H34N4O4	0.93	ESI+ FT-ICR-MS
371.0357444	C11H19NO7P2S	0.59	ESI+ FT-ICR-MS
371.0471226	C15H16Cl3N5	only in MA	ESI+ FT-ICR-MS
371.2249024	C26H29NO	0.73	ESI+ LC-MS
372.0627336	C14H16N2O8S	0.63	ESI- FT-ICR-MS
373.2325076	C16H31N5O5	0.85	ESI+ FT-ICR-MS
375.9717674	C8H14N2O7P2S2	0.53	ESI+ FT-ICR-MS
376.0794286	C18H16O9	-0.81	ESI+ FT-ICR-MS
376.1018946	C16H16N4O7	0.75	ESI- FT-ICR-MS
376.1059166	C21H16N2O5	0.46	ESI- FT-ICR-MS
380.0202022	C16H12O9S	1.03	ESI+ FT-ICR-MS
381.1552458	C15H28NO8P	only in MA	ESI+ FT-ICR-MS
381.1607804	C22H24ClN3O	only in MA	ESI+ FT-ICR-MS
383.2783922	C20H37N3O4	0.50	ESI+ LC-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
384.087884	C17H20O8S	-0.75	ESI- FT-ICR-MS
385.075741	C14H15N3O10	0.66	ESI- FT-ICR-MS
386.065117	C20H10N4O5	0.81	ESI- FT-ICR-MS
386.1212912	C17H22O10	0.51	ESI+ LC-MS
387.161654	C20H25N3O3S	0.36	ESI- FT-ICR-MS
387.2256908	C19H33NO7	1.34	ESI+ FT-ICR-MS
389.1263164	C23H19NO5	0.88	ESI- FT-ICR-MS
389.1926922	C13H32N3O8P	-0.84	ESI+ FT-ICR-MS
389.9874166	C9H16N2O7P2S2	0.97	ESI+ FT-ICR-MS
391.2722442	C23H37NO4	0.63	ESI+ FT-ICR-MS
394.1416312	C23H22O6	0.52	ESI- FT-ICR-MS
395.1633686	C25H21N3O2	0.59	ESI+ LC-MS
395.2307758	C21H33NO6	0.48	ESI- FT-ICR-MS
397.1613938	C14H28N3O8P	0.50	ESI+ FT-ICR-MS
399.1518926	C12H26N5O8P	0.51	ESI+ FT-ICR-MS
399.2409458	C24H33NO4	1.56	ESI+ LC-MS
400.115812	C21H20O8	0.51	ESI+ LC-MS
400.1885788	C23H28O6	0.84	ESI+ LC-MS
402.0950778	C20H18O9	0.68	ESI+ LC-MS
404.125982	C24H20O6	0.67	ESI- FT-ICR-MS
404.1583434	C20H24N2O7	0.47	ESI+ LC-MS
407.0300694	C11H14N5O8PS	-1.07	ESI- FT-ICR-MS
409.1055994	C16H19N5O6S	0.51	ESI- FT-ICR-MS
411.0727822	C18H22NO4PS2	-0.62	ESI+ FT-ICR-MS
416.1096196	C20H15F3N4O3	0.33	ESI- FT-ICR-MS
418.1357278	C16H28N4O3P2S	0.51	ESI+ FT-ICR-MS
418.1756456	C19H31O8P	0.71	ESI+ FT-ICR-MS
424.0906616	C21H16N2O8	0.76	ESI- FT-ICR-MS
425.1335274	C20H19N5O6	0.50	ESI- FT-ICR-MS
426.1750556	C18H26N4O8	0.48	ESI- FT-ICR-MS
426.1767086	C16H31N2O9P	-0.64	ESI+ FT-ICR-MS
432.0845136	C24H16O8	0.51	ESI- FT-ICR-MS
432.3239424	C27H44O4	0.63	ESI+ LC-MS
434.2304414	C24H34O7	0.85	ESI+ LC-MS
436.0212652	C17H12N2O10S	0.56	ESI- FT-ICR-MS
436.0906616	C22H16N2O8	0.31	ESI- FT-ICR-MS
436.1143104	C18H29O6PS2	-0.61	ESI+ FT-ICR-MS
441.2086016	C24H31N3O3S	0.38	ESI- FT-ICR-MS
446.0830264	C17H24N2O6P2S	-0.61	ESI+ FT-ICR-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
450.1936166	C <sub>20</sub> H <sub>36</sub> O <sub>7</sub> P <sub>2</sub>	0.72	ESI+ LC-MS
453.0048402	C <sub>14</sub> H <sub>17</sub> N <sub>10</sub> P <sub>2</sub> S	-0.81	ESI+ FT-ICR-MS
455.0270788	C <sub>16</sub> H <sub>13</sub> N <sub>3</sub> O <sub>11</sub> S	-0.63	ESI- FT-ICR-MS
456.2372582	C <sub>24</sub> H <sub>32</sub> N <sub>4</sub> O <sub>5</sub>	0.31	ESI- FT-ICR-MS
458.01363	C <sub>13</sub> H <sub>20</sub> N <sub>2</sub> O <sub>8</sub> P <sub>2</sub> S <sub>2</sub>	-0.46	ESI+ FT-ICR-MS
460.0760282	C <sub>12</sub> H <sub>22</sub> N <sub>4</sub> O <sub>11</sub> P <sub>2</sub>	0.83	ESI+ LC-MS
462.2617398	C <sub>26</sub> H <sub>38</sub> O <sub>7</sub>	0.81	ESI+ LC-MS
470.1172692	C <sub>19</sub> H <sub>22</sub> N <sub>2</sub> O <sub>12</sub>	0.32	ESI- FT-ICR-MS
472.1681364	C <sub>25</sub> H <sub>24</sub> N <sub>6</sub> O <sub>2</sub> S	-1.24	ESI- FT-ICR-MS
473.1658888	C <sub>20</sub> H <sub>23</sub> N <sub>7</sub> O <sub>7</sub>	-0.68	ESI+ FT-ICR-MS
477.1004528	C <sub>14</sub> H <sub>28</sub> N <sub>3</sub> O <sub>9</sub> P <sub>2</sub> S <sub>2</sub>	-0.46	ESI+ FT-ICR-MS
478.2427374	C <sub>23</sub> H <sub>34</sub> N <sub>4</sub> O <sub>7</sub>	0.30	ESI- FT-ICR-MS
483.1981714	C <sub>18</sub> H <sub>34</sub> N <sub>3</sub> O <sub>10</sub> P	0.65	ESI+ FT-ICR-MS
483.2230064	C <sub>23</sub> H <sub>29</sub> N <sub>7</sub> O <sub>5</sub>	1.24	ESI- FT-ICR-MS
484.0035072	C <sub>21</sub> H <sub>12</sub> N <sub>2</sub> O <sub>8</sub> S <sub>2</sub>	0.38	ESI- FT-ICR-MS
484.0042928	C <sub>18</sub> H <sub>13</sub> O <sub>14</sub> P	-0.70	ESI- FT-ICR-MS
487.0243012	C <sub>18</sub> H <sub>17</sub> N <sub>11</sub> S <sub>2</sub>	0.48	ESI- FT-ICR-MS
488.1568418	C <sub>21</sub> H <sub>33</sub> N <sub>2</sub> O <sub>5</sub> P <sub>2</sub> S <sub>2</sub>	-0.65	ESI+ FT-ICR-MS
489.0689414	C <sub>17</sub> H <sub>19</sub> N <sub>3</sub> O <sub>12</sub> S	0.61	ESI- FT-ICR-MS
492.2359206	C <sub>26</sub> H <sub>36</sub> O <sub>9</sub>	1.07	ESI+ LC-MS
498.0659054	C <sub>21</sub> H <sub>14</sub> N <sub>4</sub> O <sub>11</sub>	0.81	ESI- FT-ICR-MS
498.0910558	C <sub>23</sub> H <sub>18</sub> N <sub>2</sub> O <sub>11</sub>	1.02	ESI- FT-ICR-MS
504.272304	C <sub>28</sub> H <sub>40</sub> O <sub>8</sub>	0.80	ESI+ LC-MS
504.2736416	C <sub>29</sub> H <sub>36</sub> N <sub>4</sub> O <sub>4</sub>	0.72	ESI+ LC-MS
514.2930382	C <sub>30</sub> H <sub>42</sub> O <sub>7</sub>	0.92	ESI+ LC-MS
517.1631112	C <sub>23</sub> H <sub>27</sub> N <sub>5</sub> O <sub>7</sub> S	0.34	ESI+ LC-MS
522.2464848	C <sub>27</sub> H <sub>38</sub> O <sub>10</sub>	0.58	ESI+ LC-MS
533.3213178	C <sub>27</sub> H <sub>43</sub> N <sub>5</sub> O <sub>6</sub>	0.42	ESI- FT-ICR-MS
535.0774112	C <sub>17</sub> H <sub>22</sub> N <sub>5</sub> O <sub>11</sub> PS	0.65	ESI- FT-ICR-MS
535.0960366	C <sub>18</sub> H <sub>26</sub> N <sub>5</sub> O <sub>8</sub> P <sub>2</sub> S <sub>2</sub>	0.67	ESI- FT-ICR-MS
536.067167	C <sub>21</sub> H <sub>20</sub> N <sub>4</sub> O <sub>9</sub> S <sub>2</sub>	0.65	ESI- FT-ICR-MS
544.9889952	C <sub>16</sub> H <sub>12</sub> N <sub>5</sub> O <sub>13</sub> PS	0.78	ESI- FT-ICR-MS
554.2376524	C <sub>28</sub> H <sub>34</sub> N <sub>4</sub> O <sub>8</sub>	0.42	ESI- FT-ICR-MS
559.0831506	C <sub>23</sub> H <sub>21</sub> N <sub>5</sub> O <sub>8</sub> S <sub>2</sub>	0.55	ESI- FT-ICR-MS
562.2413998	C <sub>29</sub> H <sub>38</sub> O <sub>11</sub>	0.42	ESI+ LC-MS
592.1791972	C <sub>28</sub> H <sub>32</sub> O <sub>14</sub>	0.33	ESI+ LC-MS
594.3417176	C <sub>33</sub> H <sub>46</sub> N <sub>4</sub> O <sub>6</sub>	1.81	ESI+ LC-MS
597.0649194	C <sub>21</sub> H <sub>19</sub> N <sub>5</sub> O <sub>14</sub> S	-0.67	ESI- FT-ICR-MS
604.1570374	C <sub>26</sub> H <sub>29</sub> N <sub>4</sub> O <sub>11</sub> P	1.00	ESI- FT-ICR-MS

monoisotopic mass	elemental composition	log ratio MA/C	method
612.0540036	C <sub>31</sub> H <sub>16</sub> O <sub>14</sub>	0.49	ESI- FT-ICR-MS
670.1045662	C <sub>21</sub> H <sub>32</sub> N <sub>6</sub> O <sub>11</sub> P <sub>2</sub> S <sub>2</sub>	0.79	ESI+ FT-ICR-MS
687.483879	C <sub>37</sub> H <sub>70</sub> NO <sub>8</sub> P	1.04	ESI+ FT-ICR-MS

Table A13: Significant annotated metabolites with a fold change  $\geq 100\%$  between the C and MA group

	Synapt™ HDMS™	Synapt™ G2S
Total chromatographic runtime	36 min	12 min
Injected sample volume	5.0 $\mu$ l	0.5 $\mu$ l
Average detected signals in QC samples:		
ESI <sup>+</sup>	5498	8951
ESI <sup>-</sup>	4120	9276
sum	9618	18227

Table A14: Comparison between the UHPLC-MS platforms used in the two studies

Compound	Producer / Distributor
(+-) Carnitine hydrochloride	Sigma-Aldrich, St. Luis, USA
1,3,5-Benzenetricarboxylic acid	
2 Carboxycinnamic acid	
2-(4-Isobutylphenyl)propionic acid	
2-Amino-4,5-dimethoxybenzoic acid	
2-Biphenylcarboxylic acid	
4-Biphenylcarboxylic acid	
4-Hydroxyphenylacetic acid	Sigma-Aldrich, St. Luis, USA
Acetyl-L-carnitine.HCl	Dr. Herman J. ten Brink, VU medical center, Netherlands
Adipic acid	Sigma-Aldrich, St. Luis, USA
Benzoylformic acid	Sigma-Aldrich, St. Luis, USA
Caffeine	Sigma-Aldrich, St. Luis, USA
Capric acid	Sigma-Aldrich, St. Luis, USA
Caprylic acid	Sigma-Aldrich, St. Luis, USA
Citric acid	Sigma-Aldrich, St. Luis, USA
Coumarin-3-carboxylic acid	
D(-)Arabinose	Sigma-Aldrich, St. Luis, USA

<b>Compound</b>	<b>Producer / Distributor</b>
D(-)Fructose	Sigma-Aldrich, St. Luis, USA
D(+)-Galactose	Sigma-Aldrich, St. Luis, USA
D(+)-Glucose	Sigma-Aldrich, St. Luis, USA
D-(+)-Glucuronic acid	Sigma-Aldrich, St. Luis, USA
D-Alanine	Sigma-Aldrich, St. Luis, USA
Decanoyl-carnitine	Dr. Herman J. ten Brink, VU medical center, Netherlands
DL-Histidine	Sigma-Aldrich, St. Luis, USA
DL-Leucine	Sigma-Aldrich, St. Luis, USA
DL-Methionine	Sigma-Aldrich, St. Luis, USA
DL-Proline	Sigma-Aldrich, St. Luis, USA
DL-Serine	Sigma-Aldrich, St. Luis, USA
DL-Tryptophane	EGA-Chemie, Steinheim, Germany
DL-Valine	Sigma-Aldrich, St. Luis, USA
D-Phenylalanine	Sigma-Aldrich, St. Luis, USA
D-Sorbitol	Sigma-Aldrich, St. Luis, USA
Folic acid	Merck, Darmstadt, Germany
Fumaric acid	Sigma-Aldrich, St. Luis, USA
Gallic acid	Sigma-Aldrich, St. Luis, USA
Gentiobiose	Sigma-Aldrich, St. Luis, USA
Glutaconic acid	Sigma-Aldrich, St. Luis, USA
Glutaric acid	Alfa Aesar, Karlsruhe, Gerany
Glycine	Merck, Darmstadt, Germany
Glyoxilic acid	Sigma-Aldrich, St. Luis, USA
Hexadecanoyl carnitine	Dr. Herman J. ten Brink, VU medical center, Netherlands
L(-)-Fucose	Sigma-Aldrich, St. Luis, USA
L(+)-Glutamic acid	Merck, Darmstadt, Germany
L-Arginine	Sigma-Aldrich, St. Luis, USA
L-Asparagine mono hydrate	Sigma-Aldrich, St. Luis, USA
L-Aspartic acid	Sigma-Aldrich, St. Luis, USA
L-Isoleucine	Sigma-Aldrich, St. Luis, USA
L-Lysine hydrochloride	Sigma-Aldrich, St. Luis, USA
L-Rhamnose monohydrate	Sigma-Aldrich, St. Luis, USA
L-Tyrosine	Merck, Darmstadt, Germany
Malic acid	Sigma-Aldrich, St. Luis, USA
Malonic acid	Sigma-Aldrich, St. Luis, USA
Mesaconic acid	Sigma-Aldrich, St. Luis, USA
Meso-tartaric acid hydrate	Sigma-Aldrich, St. Luis, USA
Nialamide	Sigma-Aldrich, St. Luis, USA

Compound	Producer / Distributor
Nicotinamide	Sigma-Aldrich, St. Luis, USA
Nonanoic acid	Sigma-Aldrich, St. Luis, USA
Palmitic acid	Sigma-Aldrich, St. Luis, USA
Phenol	Sigma-Aldrich, St. Luis, USA
Pipecolinic acid	Sigma-Aldrich, St. Luis, USA
Reserpine	Sigma-Aldrich, St. Luis, USA
Sodium succinate dibasic hexahydrate	Sigma-Aldrich, St. Luis, USA
Stearic acid	Sigma-Aldrich, St. Luis, USA
Sulfadimethoxine	Sigma-Aldrich, St. Luis, USA
Terfenadine	Sigma-Aldrich, St. Luis, USA
Undecanoic acid	Sigma-Aldrich, St. Luis, USA
Valeryl-carnitine	Dr. Herman J. ten Brink, VU medical center, Netherlands

Table A15: Standard Compounds for method development, validation and QC used in the PREDICTIONS study

Part	Product	Producer / Distributor
SPE Cartridge	Oasis® HLB 1cc Vac cartridge, 30 mg sorbent,	Waters GmbH, Eschborn, Germany
SPE Cartridge	Varian C18 cartridge 1cc, 100 mg sorbent	Agilent, Böblingen, Germany
SPE Cartridge	Varian C8 cartridge 1cc, 100 mg sorbent	Agilent, Böblingen, Germany
SPE Cartridge	Varian CN-E cartridge 1cc, 100 mg sorbent	Agilent, Böblingen, Germany
SPE Tip	OMIX® C18 pipette tips, 10-100 µl	Agilent, Böblingen, Germany
Column	ACQUITY® UPLC® BEH C18 Column, 1.0 x 150 mm	Waters GmbH, Eschborn, Germany

Table A16: Used consumables for the method and the development in the PREDICTIONS study

	Property	ESI+	ESI-
	Initial retention time [min]	0.7	0.7
	Final retention time [min]	8.0	9.0
	Low mass [Da]	Auto	Auto
	High mass [Da]	Auto	Auto
	Mass tolerance [Da]	0.02	0.02



Apex track peak parameters	Peak width 5% height [s]	auto	auto
	Peak-to-peak baseline noise	auto	auto
	Smoothing	Yes	Yes
Collection parameters	Intensity threshold [counts]	3000	1000
	Mass window [Da]	0.02	0.02
	RT window [min]	0.15	0.15
	Noise elimination level	auto	auto
	Deisotope data?	Yes	Yes
	Replicate % Minimum	25	25

Table A17: Settings applied for peak detection and data matrix generation within MarkerLynx™ for the PREDICTIONS study

Parameter	ESI +	ESI -
Number of scans	415	385
Transient time domain [MW]	2	2
Source accumulation [s]	0.010	0.010
Ion accumulation time [s]	0.100	0.200
Ion cooling time [s]	0.050	0.050
Dwell time [s]	0.100	0.100
Capillary voltage [kV]	4.0	4.0
Dry gas flow [l/min]	4.0	4.0
Dry gas temperature [°C]	200	180
Nebulizer gas [bar]	1.0	1.0
Side kick [V]	5.0	5.0
Side kick off set [V]	-1.5	10.0
Front plate voltage [V]	0.4	-0.4
Back plate voltage [V]	0.5	-0.5
Analyzer entrance [V]	-10.0	4.0
Broadband high mass [Da]	1000	1000
Broadband low mass [Da]	147.5	147.5

Table A18: Optimised parameters for the FT-ICR-MS measurement of the PREDICTIONS study samples

monoisotopic mass	elemental composition	log ratio DN/C	method
58.04186	C3H6O	-1.46	ESI- LC-MS
68.02621	C4H4O	-0.90	ESI- LC-MS
70.00548	C3H2O2	-0.74	ESI+ LC-MS
86.03678	C4H6O2	-0.42	ESI- LC-MS
88.01604	C3H4O3	-0.39	ESI- LC-MS
88.05243	C4H8O2	only in C	ESI- LC-MS
89.04768	C3H7NO2	only in C	ESI- LC-MS
102.04293	C3H6N2O2	-0.36	ESI+ LC-MS
104.01096	C3H4O4	only in C	ESI- LC-MS
104.04734	C4H8O3	only in C	ESI- LC-MS
105.03270	C5H3N3	-1.22	ESI- LC-MS
112.01604	C5H4O3	-0.37	ESI- LC-MS
112.01604	C5H4O3	-0.66	ESI+ LC-MS
116.01096	C4H4O4	-0.59	ESI+ LC-MS
116.04734	C5H8O3	only in C	ESI- LC-MS
116.08373	C6H12O2	only in C	ESI- LC-MS
117.07897	C5H11NO2	only in C	ESI- LC-MS
117.99022	C3H2O5	-0.77	ESI- LC-MS
118.04186	C8H6O	-0.78	ESI- LC-MS
118.06299	C5H10O3	only in C	ESI- LC-MS
121.01975	C3H7NO2S	only in C	ESI- LC-MS
121.01975	C3H7NO2S	-0.71	ESI+ LC-MS
123.98303	C2H4O4S	-0.57	ESI+ LC-MS
128.02219	C4H4N2O3	-0.57	ESI+ LC-MS
128.04734	C6H8O3	only in C	ESI- LC-MS
130.02661	C5H6O4	-0.46	ESI- LC-MS
145.07389	C6H11NO3	-0.98	ESI- LC-MS
146.06914	C5H10N2O3	1.08	ESI+ LC-MS
146.98355	C3H2CIN3O2	-0.57	ESI+ LC-MS
147.05316	C5H9NO4	-1.27	ESI- LC-MS
148.03717	C5H8O5	-0.32	ESI- LC-MS
152.01567	C5H4N4S	only in C	ESI- LC-MS
152.03342	C5H4N4O2	-0.50	ESI+ LC-MS
153.02483	C7H7NOS	only in C	ESI- LC-MS
155.98238	C2H5O6P	-0.71	ESI+ LC-MS
161.06880	C6H11NO4	only in C	ESI- LC-MS
165.07897	C9H11NO2	0.61	ESI+ LC-MS
167.09462	C9H13NO2	only in C	ESI- LC-MS
170.05791	C8H10O4	only in C	ESI- LC-MS

172.14632	C10H20O2	only in C	ESI- LC-MS
174.01644	C6H6O6	-0.46	ESI- LC-MS
174.08920	C8H14O4	-1.57	ESI- LC-MS
176.09496	C10H12N2O	only in C	ESI- LC-MS
178.04774	C6H10O6	only in C	ESI- LC-MS
178.98884	C4H5NO5S	-0.62	ESI+ LC-MS
180.06339	C6H12O6	only in DN	ESI+ FT-ICR-MS
180.06472	C7H8N4O2	-0.47	ESI+ FT-ICR-MS
183.05316	C8H9NO4	-0.37	ESI+ FT-ICR-MS
183.05316	C8H9NO4	-1.14	ESI- LC-MS
183.05653	C5H13NO4S	only in C	ESI- LC-MS
188.11609	C8H16N2O3	only in C	ESI- LC-MS
190.08412	C8H14O5	only in C	ESI- LC-MS
191.98255	C6H5ClO5	-2.70	ESI- LC-MS
192.02700	C6H8O7	-0.35	ESI- LC-MS
192.02700	C6H8O7	-0.47	ESI+ FT-ICR-MS
192.02700	C6H8O7	-0.52	ESI+ LC-MS
194.11542	C8H18O5	-0.35	ESI+ FT-ICR-MS
198.03890	C6H6N4O4	only in C	ESI- LC-MS
199.05930	C7H9N3O4	-1.30	ESI- LC-MS
204.11100	C8H16N2O4	only in C	ESI- LC-MS
210.10044	C10H14N2O3	only in C	ESI- LC-MS
211.08445	C10H13NO4	-1.42	ESI- LC-MS
211.95338	C6FeN6	-0.40	ESI+ LC-MS
214.02424	C5H11O7P	only in C	ESI- LC-MS
214.13174	C10H18N2O3	1.32	ESI+ LC-MS
214.13174	C10H18N2O3	0.32	ESI+ FT-ICR-MS
217.10625	C8H15N3O4	only in C	ESI- LC-MS
218.09027	C8H14N2O5	-1.38	ESI- LC-MS
219.11067	C9H17NO5	-0.31	ESI+ FT-ICR-MS
219.11067	C9H17NO5	-0.33	ESI- FT-ICR-MS
219.11067	C9H17NO5	-0.36	ESI+ FT-ICR-MS
220.08479	C11H12N2O3	only in C	ESI- LC-MS
225.10144	C12H11N5	only in DN	ESI- FT-ICR-MS
227.22490	C14H29NO	only in DN	ESI+ FT-ICR-MS
229.05863	C9H11NO6	-1.21	ESI- LC-MS
229.10942	C9H16ClN5	only in C	ESI- LC-MS
232.10592	C9H16N2O5	-1.23	ESI- LC-MS
236.02806	C6H8N2O8	only in C	ESI- LC-MS
236.07970	C11H12N2O4	only in C	ESI- LC-MS

239.07284	C10H13N3O2S	-1.10	ESI- LC-MS
242.09027	C10H14N2O5	only in C	ESI- LC-MS
244.08816	C10H16N2O3S	-0.47	ESI- FT-ICR-MS
246.09237	C14H15ClN2	only in C	ESI- LC-MS
247.10558	C10H17NO6	only in C	ESI- LC-MS
249.08572	C10H19NO2S2	only in C	ESI- LC-MS
249.98707	C10H6N2O2S2	only in C	ESI- LC-MS
252.08585	C10H12N4O4	only in C	ESI- LC-MS
253.10625	C11H15N3O4	only in C	ESI- LC-MS
253.15789	C16H19N3	only in C	ESI- LC-MS
255.15828	C12H21N3O3	0.62	ESI+ FT-ICR-MS
258.12559	C16H18O3	-1.34	ESI- LC-MS
261.12123	C11H19NO6	only in C	ESI- LC-MS
262.22965	C18H30O	0.58	ESI+ FT-ICR-MS
263.11173	C9H17O6N3	only in DN	ESI+ FT-ICR-MS
267.14705	C14H21NO4	only in C	ESI- LC-MS
269.06880	C15H11NO4	only in C	ESI- LC-MS
272.08960	C12H16O7	only in C	ESI- LC-MS
275.11173	C10H17N3O6	1.47	ESI- LC-MS
276.07217	C11H11F3N2O3	only in C	ESI- LC-MS
277.02562	C9H11NO7S	0.37	ESI- FT-ICR-MS
277.11748	C12H15N5O3	only in C	ESI- LC-MS
279.25620	C18H33NO	1.36	ESI+ FT-ICR-MS
284.96448	C6H8ClN3O4S2	-0.63	ESI- LC-MS
285.03620	C6H12N3O8P	-0.84	ESI+ LC-MS
286.08412	C16H14O5	only in C	ESI- LC-MS
287.18450	C13H25O4N3	0.37	ESI+ FT-ICR-MS
291.21982	C18H29NO2	2.18	ESI+ LC-MS
292.09468	C15H16O6	-1.23	ESI- LC-MS
299.09397	C12H17N3O4S	only in C	ESI- LC-MS
299.18450	C14H25N3O4	-0.46	ESI+ LC-MS
300.02700	C15H8O7	-0.48	ESI+ LC-MS
301.02724	C13H13Cl2NO3	only in C	ESI- LC-MS
301.07121	C16H13O6	only in C	ESI- LC-MS
301.07976	C12H15NO8	only in C	ESI- LC-MS
301.20015	C14H27O4N3	0.35	ESI+ FT-ICR-MS
302.22457	C20H30O2	only in C	ESI- LC-MS
304.09066	C11H16N2O8	-1.53	ESI- LC-MS
305.00615	C7H15O6NS3	-0.37	ESI+ FT-ICR-MS
305.12765	C17H15N5O	-1.50	ESI- LC-MS

306.03757	C14H10O8	only in C	ESI- LC-MS
306.97054	C9H10INO3	-0.47	ESI+ LC-MS
307.06920	C14H13NO7	-0.37	ESI- FT-ICR-MS
311.12296	C12H17N5O5	0.32	ESI+ FT-ICR-MS
313.08447	C11H15O4N5S	-0.33	ESI+ FT-ICR-MS
319.01668	C16H11Cl2NO2	only in C	ESI- LC-MS
324.13615	C20H20O4	only in C	ESI- LC-MS
325.02986	C11H19NO2S4	-0.72	ESI+ LC-MS
325.07820	C18H13ClFN3	only in C	ESI- LC-MS
327.21982	C21H29NO2	2.24	ESI+ LC-MS
328.07943	C14H16O9	only in C	ESI- LC-MS
328.10189	C12H16N4O7	-1.05	ESI- LC-MS
329.12632	C18H19NO5	only in C	ESI- LC-MS
329.23144	C16H31O4N3	0.90	ESI+ FT-ICR-MS
330.00772	C12H11ClN2O5S	0.76	ESI- FT-ICR-MS
331.17835	C19H25NO4	-1.53	ESI- LC-MS
331.21071	C15H29O5N3	0.49	ESI+ FT-ICR-MS
332.07434	C13H16O10	-1.05	ESI- LC-MS
335.13286	C12H21N3O8	-1.39	ESI- LC-MS
337.16778	C21H23NO3	1.75	ESI+ LC-MS
337.17500	C15H23N5O4	0.32	ESI+ FT-ICR-MS
343.25112	C22H33NO2	3.10	ESI- LC-MS
344.14711	C16H24O8	-0.32	ESI+ FT-ICR-MS
346.06886	C17H14O8	only in C	ESI- LC-MS
346.10525	C18H18O7	only in DN	ESI+ FT-ICR-MS
349.05627	C12H16NO9P	only in C	ESI- LC-MS
349.20015	C18H27N3O4	2.12	ESI+ LC-MS
349.20015	C18H27O4N3	1.05	ESI+ FT-ICR-MS
349.20754	C20H31NO2S	2.87	ESI- LC-MS
352.07943	C16H16O9	only in C	ESI- LC-MS
352.19126	C22H26NO3	1.55	ESI+ LC-MS
353.09330	C16H19O6NS	-0.46	ESI+ FT-ICR-MS
355.10493	C11H21O8N3S	-0.67	ESI+ FT-ICR-MS
356.16237	C21H24O5	only in C	ESI- LC-MS
356.25626	C20H36O5	only in C	ESI- LC-MS
356.30790	C25H40O	0.38	ESI+ FT-ICR-MS
357.14236	C16H23NO8	only in C	ESI- LC-MS
357.19400	C21H27NO4	only in C	ESI- LC-MS
357.26274	C18H35N3O4	only in DN	ESI+ FT-ICR-MS
359.08524	C14H17NO10	-1.78	ESI- LC-MS

364.17466	C17H24O5N4	1.04	ESI+ FT-ICR-MS
364.18456	C15H28O8N2	0.53	ESI+ FT-ICR-MS
365.21887	C18H29O4N4	0.90	ESI+ FT-ICR-MS
366.20423	C20H30O6	only in DN	ESI+ LC-MS
368.02570	C10H13N2O11P	-0.49	ESI+ LC-MS
368.04852	C17H17ClO5S	only in C	ESI- LC-MS
370.18139	C19H30O5S	only in C	ESI- LC-MS
373.13794	C20H24CIN3S	only in C	ESI- LC-MS
373.19612	C15H27O6N5	only in DN	ESI+ FT-ICR-MS
374.18014	C15H26N4O7	1.64	ESI+ LC-MS
375.15943	C19H22FN3O4	only in C	ESI- LC-MS
377.25660	C22H35NO4	only in DN	ESI+ FT-ICR-MS
380.11678	C18H22O6NS	-0.32	ESI+ FT-ICR-MS
380.17360	C22H24N2O4	2.16	ESI+ LC-MS
380.17360	C22H24N2O4	2.02	ESI- LC-MS
381.17740	C17H25O6N4	0.82	ESI+ FT-ICR-MS
381.20121	C17H27O5N5	-0.32	ESI+ FT-ICR-MS
383.18450	C21H25O4N3	only in DN	ESI+ FT-ICR-MS
384.04498	C15H16N2O6S2	-0.48	ESI+ LC-MS
385.10560	C14H19N5O6S	only in C	ESI- LC-MS
386.16315	C18H34OSn	only in DN	ESI+ FT-ICR-MS
386.17293	C22H26O6	-0.34	ESI+ LC-MS
386.20931	C23H30O5	-0.77	ESI- LC-MS
391.17770	C16H29N3O6S	0.64	ESI+ FT-ICR-MS
391.17835	C24H25NO4	2.57	ESI+ LC-MS
392.21988	C22H32O6	only in C	ESI- LC-MS
401.17730	C21H27N3O3S	only in C	ESI- LC-MS
402.10631	C19H18O8N2	-0.34	ESI+ FT-ICR-MS
406.08999	C19H18O10	only in C	ESI- LC-MS
406.12638	C20H22O9	-0.51	ESI+ LC-MS
408.01196	C13H16O7N2S3	-0.45	ESI+ FT-ICR-MS
411.15828	C25H21N3O3	only in C	ESI- LC-MS
411.18458	C24H26FNO4	only in C	ESI- LC-MS
414.04008	C15H15CIN4O6S	-2.07	ESI- LC-MS
414.16784	C23H26O7	only in C	ESI- LC-MS
418.23553	C24H34O6	-0.87	ESI- LC-MS
419.96238	C6H15O15P3	-0.56	ESI+ LC-MS
427.15657	C22H25N3O4S	-1.48	ESI- LC-MS
428.07899	C18H22O8P2	only in C	ESI- LC-MS
428.19473	C23H28N2O6	-1.44	ESI- LC-MS

430.12638	C22H22O9	only in C	ESI- LC-MS
430.17801	C27H26O5	only in C	ESI- LC-MS
432.19366	C27H28O5	-1.70	ESI- LC-MS
434.23044	C24H34O7	1.82	ESI+ LC-MS
444.15821	C22H27F3O4S	only in C	ESI- LC-MS
453.04129	C16H15N5O7S2	only in C	ESI- LC-MS
454.11985	C23H22N2O6S	only in C	ESI- LC-MS
456.21479	C26H32O7	-0.65	ESI+ FT-ICR-MS
457.17097	C20H23N7O6	-1.09	ESI- LC-MS
481.03824	C13H23O12NS3	only in DN	ESI+ FT-ICR-MS
488.07369	C18H20O12N2S	0.43	ESI- FT-ICR-MS
488.28931	C21H46O9NS	-1.01	ESI+ FT-ICR-MS
490.22096	C19H34O7N6S	only in DN	ESI+ FT-ICR-MS
493.24240	C24H35O8N3	0.42	ESI+ FT-ICR-MS
493.26353	C21H39O10N3	only in DN	ESI+ FT-ICR-MS
494.08491	C25H18O11	only in C	ESI- LC-MS
511.25297	C24H37O9N3	only in DN	ESI+ FT-ICR-MS
520.06689	C15H24O14N2S2	-0.35	ESI+ FT-ICR-MS
521.25845	C22H39O11N3	only in DN	ESI+ FT-ICR-MS
522.89190	C10H12MoN5O8PS2	only in C	ESI- LC-MS
523.02705	C11H20N5O13P3	only in C	ESI- LC-MS
525.24749	C28H35N3O7	2.00	ESI+ LC-MS
526.22748	C23H34O10N4	0.85	ESI+ FT-ICR-MS
536.27202	C25H38O8N5	0.47	ESI+ FT-ICR-MS
546.17372	C27H30O12	only in C	ESI- LC-MS
550.30159	C28H38N8O4	1.25	ESI+ FT-ICR-MS
552.09579	C19H26O12N3S2	only in DN	ESI+ FT-ICR-MS
556.30360	C32H44O8	only in DN	ESI+ LC-MS
566.23765	C29H34O8N4	-0.40	ESI+ FT-ICR-MS
568.24794	C22H34O9N9	-0.51	ESI+ FT-ICR-MS
580.25330	C30H36O8N4	-0.59	ESI+ FT-ICR-MS
584.26347	C33H36N4O6	only in C	ESI- LC-MS
588.37909	C31H57O8P	1.33	ESI+ FT-ICR-MS
613.31115	C31H43O8N5	0.39	ESI+ FT-ICR-MS
664.38224	C36H56O11	1.83	ESI+ LC-MS
665.10096	C21H27N6O15P2	only in C	ESI- LC-MS
689.11526	C24H34O19PS	only in DN	ESI+ FT-ICR-MS
703.42952	C39H61NO10	-0.88	ESI+ LC-MS
760.20266	C30H41O15N4PS	-0.32	ESI+ FT-ICR-MS

Table A19: Significant annotated metabolites with a fold change  $\geq 100\%$  between the C and DN group

## Supplementary figures:

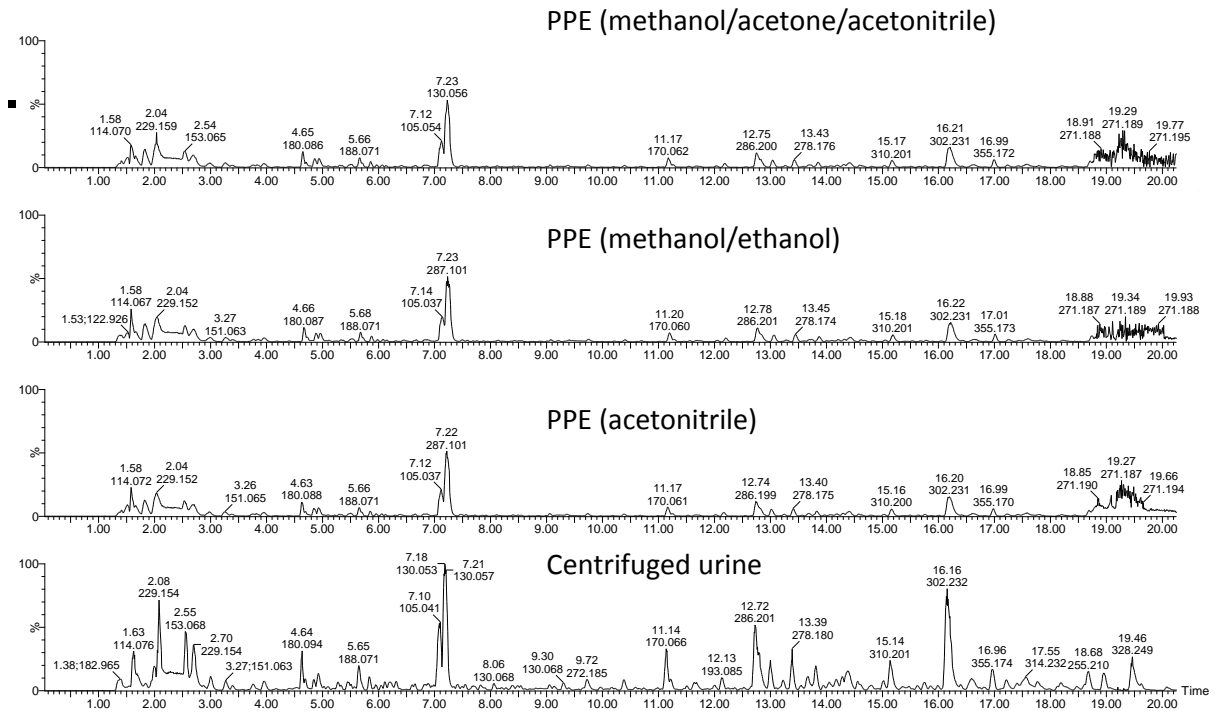


Figure A1: Comparison of protein precipitations (PPE) with different organic solvent mixtures with centrifuged urine.

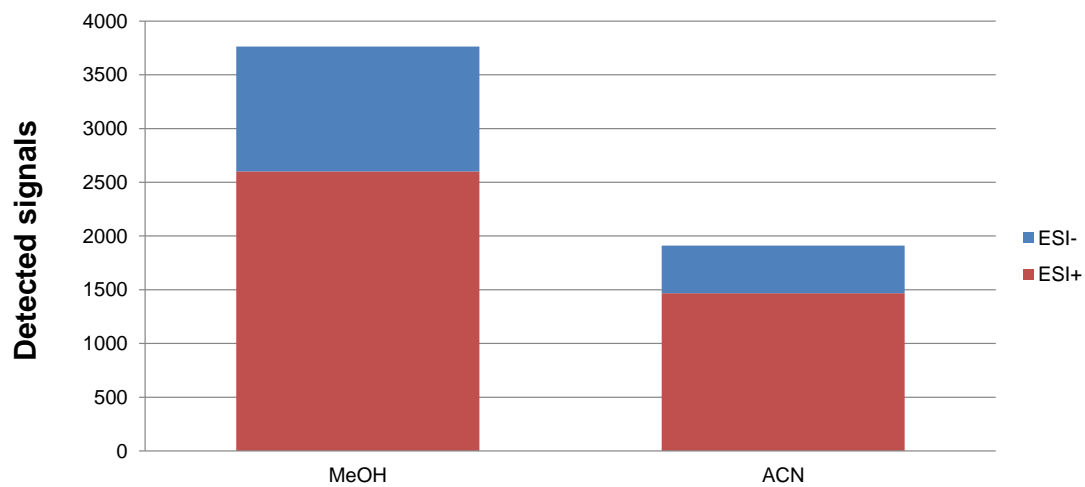


Figure A2: Comparison between MeOH and ACN as elution solvent for the RP method. Since MeOH supports the ionisation process more signals can be detected in both ionisation modes compared to ACN.



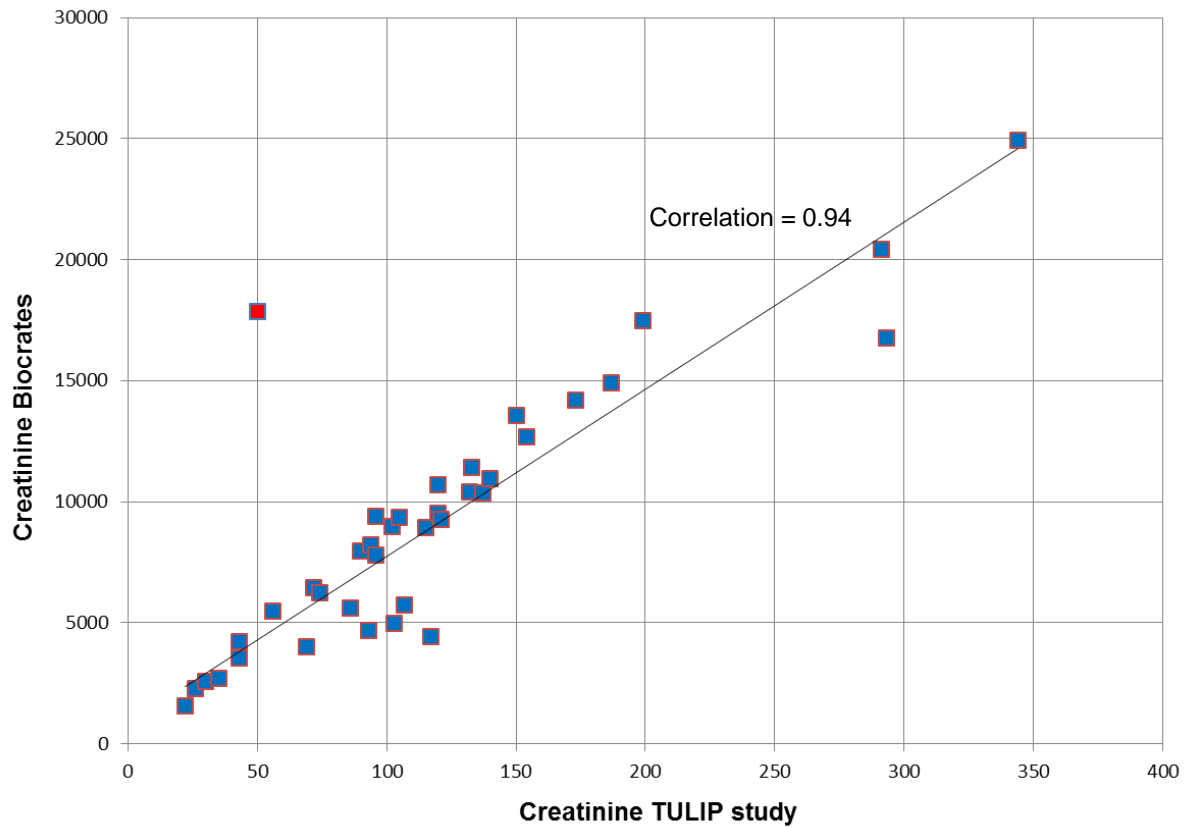


Figure A3: Comparison between the creatinine values measured within the TULIP study and the Biocrates platform. For the correlation calculation one sample (red) was excluded as outlier.

## 7 References

**Alkhalaf, A., Zurbig, P., Bakker, S. J., Bilo, H. J., Cerna, M., Fischer, C., et al. Group, P. (2010).** Multicentric validation of proteomic biomarkers in urine specific for diabetic nephropathy. *PLoS ONE* 5, e13421.

**American Diabetes Association (2010).** Diagnosis and classification of diabetes mellitus. *Diabetes Care* 33 Suppl 1, S62-69.

**Baines, A. D., Shaikh, N. & Ho, P. (1990).** Mechanisms of perfused kidney cytoprotection by alanine and glycine. *The American journal of physiology* 259, F80-87.

**Baker, M. (2005).** In biomarkers we trust? *Nat Biotechnol* 23, 297-304.

**Banerjee, S. & Mazumdar, S. (2012).** Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry* 2012, 282574.

**Barker, M. & Rayens, W. (2003).** Partial least squares for discrimination. *Journal of Chemometrics* 17, 166-173.

**Barrow, M. P., Burkitt, W. I. & Derrick, P. J. (2005).** Principles of Fourier transform ion cyclotron resonance mass spectrometry and its application in structural biology. *Analyst* 130, 18-28.

**Basi, S., Fesler, P., Mimran, A. & Lewis, J. B. (2008).** Microalbuminuria in type 2 diabetes and hypertension: a marker, treatment target, or innocent bystander? *Diabetes Care* 31 Suppl 2, S194-201.

**Bell, J. D., Lee, J. A., Lee, H. A., Sadler, P. J., Wilkie, D. R. & Woodham, R. H. (1991).** Nuclear magnetic resonance studies of blood plasma and urine from subjects with chronic renal failure: identification of trimethylamine-N-oxide. *Biochim Biophys Acta* 1096, 101-107.

**Berg, J. M., Tymoczko, J. L. & Stryer, L. (2003).** Biochemie, 5. Aufl. edn. Heidelberg [u.a.]: Spektrum Akad. Verl.

**Bhalla, R., Narasimhan, K. & Swarup, S. (2005).** Metabolomics and its role in understanding cellular responses in plants. *Plant Cell Rep* 24, 562-571.

**Boccard, J., Veuthey, J. L. & Rudaz, S. (2010).** Knowledge discovery in metabolomics: an overview of MS data handling. *J Sep Sci* 33, 290-304.

**Bothwell, J. H. & Griffin, J. L. (2011).** An introduction to biological nuclear magnetic resonance spectroscopy. *Biological reviews of the Cambridge Philosophical Society* 86, 493-510.

**Brash, A. R. (2001).** Arachidonic acid as a bioactive molecule. *J Clin Invest* 107, 1339-1345.

**Brereton, R. G. (2009).** Chemometrics for pattern recognition. Chichester, West Sussex, [u.a.]: Wiley.

**Brosius, F. C. & Pennathur, S. (2013).** How to find a prognostic biomarker for progressive diabetic nephropathy. *Kidney Int* 83, 996-998.

**Busch, M., Gobert, A., Franke, S., Ott, U., Gerth, J., Muller, A., et al. Wolf, G. (2010).** Vitamin B6 metabolism in chronic kidney disease--relation to transsulfuration, advanced glycation and cardiovascular disease. *Nephron Clinical practice* 114, c38-46.

**Butte, A. (2002).** The use and analysis of microarray data. *Nature reviews Drug discovery* 1, 951-960.

**Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. (2001).** An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 36, 849-865.

**Chesney, R. W., Han, X. & Patters, A. B. (2010).** Taurine and the renal system. *Journal of biomedical science* 17 Suppl 1, S4.

**Comisarow, M. B. & Marshall, A. G. (1974).** Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters* 25, 282-283.

**Comper, W. D. & Osicka, T. M. (2005).** Detection of urinary albumin. *Advances in chronic kidney disease* 12, 170-176.

**de Boer, I. H., Rue, T. C., Hall, Y. N., Heagerty, P. J., Weiss, N. S. & Himmelfarb, J. (2011).** Temporal trends in the prevalence of diabetic kidney disease in the United States. *JAMA* 305, 2532-2539.

**Deen, P. M. & Robben, J. H. (2011).** Succinate receptors in the kidney. *Journal of the American Society of Nephrology : JASN* 22, 1416-1422.

**Dettmer, K., Aronov, P. A. & Hammock, B. D. (2007).** Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 26, 51-78.

**Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K. & Bultman, S. J. (2011).** The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab* 13, 517-526.

**Doria, A., Patti, M. E. & Kahn, C. R. (2008).** The emerging genetic architecture of type 2 diabetes. *Cell Metab* 8, 186-200.

**Dronavalli, S., Duka, I. & Bakris, G. L. (2008).** The pathogenesis of diabetic nephropathy. *Nature clinical practice Endocrinology & metabolism* 4, 444-452.

**Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. (2011).** Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 40, 387-426.

**Eiceman, G. A., Karpas, Z. & NetLibrary Inc, (2005).** Ion mobility spectrometry, 2nd edn. Boca Raton, FL: Taylor & Francis/CRC Press.

**Eriksson, L. & Umetrics AB (2006).** Advanced applications and method extensions, 2nd rev. and enl. edn. Umeå: Umetrics.

**Eriksson, L., Trygg, J. & Wold, S. (2008).** CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics* 22, 594-600.

**Ernst, R. R. & Anderson, W. A. (1966).** Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments* 37, 93-102.

**Fiehn, O. (2002).** Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol* 48, 155-171.

**Fioretto, P. & Mauer, M. (2007).** Histopathology of diabetic nephropathy. *Semin Nephrol* 27, 195-207.

**Freedman, B. I., Bostrom, M., Daeihagh, P. & Bowden, D. W. (2007).** Genetic factors in diabetic nephropathy. *Clin J Am Soc Nephrol* 2, 1306-1316.

**Freeman, R. (1995).** A short history of NMR. *Chem Heterocycl Compd* 31, 1004-1005.

- Gao, X., Chen, W., Li, R., Wang, M., Chen, C., Zeng, R. & Deng, Y. (2012).** Systematic variations associated with renal disease uncovered by parallel metabolomics of urine and serum. *BMC Syst Biol* 6 Suppl 1, S14.
- German, J. B., Bauman, D. E., Burrin, D. G., Failla, M. L., Freake, H. C., King, J. C., et al. Zeisel, S. H. (2004).** Metabolomics in the opening decade of the 21st century: building the roads to individualized health. *J Nutr* 134, 2729-2732.
- German, J. B., Hammock, B. D. & Watkins, S. M. (2005).** Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 1, 3-9.
- Goodacre, R. (2007).** Metabolomics of a superorganism. *J Nutr* 137, 259S-266S.
- Griffiths, J. (2008).** A brief history of mass spectrometry. *Anal Chem* 80, 5678-5683.
- Gross, J. H. (2011).** Mass spectrometry : a textbook, 2. edn. Berlin [u.a.]: Springer.
- Gruning, N. M., Lehrach, H. & Ralser, M. (2010).** Regulatory crosstalk of the metabolic network. *Trends in biochemical sciences* 35, 220-227.
- Guan, S. & Marshall, A. G. (1995).** Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations. *International Journal of Mass Spectrometry and Ion Processes* 146–147, 261-296.
- Guillarme, D., Nguyen, D. T., Rudaz, S. & Veuthey, J. L. (2008).** Method transfer for fast liquid chromatography in pharmaceutical analysis: application to short columns packed with small particle. Part II: gradient experiments. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik eV* 68, 430-440.
- Guthrie, D. W. & Humphreys, S. S. (1988).** Diabetes urine testing: an historical perspective. *The Diabetes educator* 14, 521-526.
- Hagestam, I. H. & Pinkerton, T. C. (1985).** Internal surface reversed-phase silica supports for liquid chromatography. *Analytical Chemistry* 57, 1757-1763.
- Hansen, S. H. (2001).** The role of taurine in diabetes and the development of diabetic complications. *Diabetes/metabolism research and reviews* 17, 330-346.

**Hendrickson, C. L. & Emmett, M. R. (1999).** Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Annual review of physical chemistry* 50, 517-536.

**Hertkorn, N., Frommberger, M., Witt, M., Koch, B. P., Schmitt-Kopplin, P. & Perdue, E. M. (2008).** Natural organic matter and the event horizon of mass spectrometry. *Anal Chem* 80, 8908-8919.

**Hirayama, A., Nakashima, E., Sugimoto, M., Akiyama, S., Sato, W., Maruyama, S., et al.Soga, T. (2012).** Metabolic profiling reveals new serum biomarkers for differentiating diabetic nephropathy. *Anal Bioanal Chem* 404, 3101-3109.

**Hirsch, I. B., Atchley, D. H., Tsai, E., Labbe, R. F. & Chait, A. (1998).** Ascorbic acid clearance in diabetic nephropathy. *J Diabetes Complications* 12, 259-263.

**Holmes, E., Wilson, I. D. & Nicholson, J. K. (2008).** Metabolic phenotyping in health and disease. *Cell* 134, 714-717.

**Houstis, N., Rosen, E. D. & Lander, E. S. (2006).** Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* 440, 944-948.

**Hovind, P., Rossing, P., Johnson, R. J. & Parving, H. H. (2011).** Serum uric acid as a new player in the development of diabetic nephropathy. *Journal of renal nutrition : the official journal of the Council on Renal Nutrition of the National Kidney Foundation* 21, 124-127.

**Huang, C., Kim, Y., Caramori, M. L., Moore, J. H., Rich, S. S., Mychaleckyj, J. C., et al.Mauer, M. (2006).** Diabetic nephropathy is associated with gene expression levels of oxidative phosphorylation and related pathways. *Diabetes* 55, 1826-1831.

**Jacobsen, N. E. (2007).** NMR spectroscopy explained : simplified theory, applications and examples for organic chemistry and structural biology. Hoboken, NJ: Wiley.

**Jones, P. J. H. & Papamandjaris, A. A. (2012).** Lipids: Cellular Metabolism. In *Present Knowledge in Nutrition*, pp. 132-148: Wiley-Blackwell.

**Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H., Jr. (2008).** Ion mobility-mass spectrometry. *J Mass Spectrom* 43, 1-22.

**Kanwar, Y. S., Wada, J., Sun, L., Xie, P., Wallner, E. I., Chen, S., et al.Danesh, F. R. (2008).** Diabetic Nephropathy: Mechanisms of Renal Disease Progression. *Experimental Biology and Medicine* 233, 4-11.

**Kassab, A., Ajmi, T., Issaoui, M., Chaeib, L., Miled, A. & Hammami, M. (2008).** Homocysteine enhances LDL fatty acid peroxidation, promoting microalbuminuria in type 2 diabetes. *Ann Clin Biochem* 45, 476-480.

**Kebarle, P. & Verkerk, U. H. (2009).** Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev* 28, 898-917.

**Keeler, J. (2011).** Understanding NMR spectroscopy, 2. edn. Chichester, West Sussex: Wiley.

**Kikuchi, K., Itoh, Y., Tateoka, R., Ezawa, A., Murakami, K. & Niwa, T. (2010).** Metabolomic analysis of uremic toxins by liquid chromatography/electrospray ionization-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 878, 1662-1668.

**Kim, S., Kramer, R. W. & Hatcher, P. G. (2003).** Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal Chem* 75, 5336-5344.

**Kopple, J. D. (2007).** Phenylalanine and tyrosine metabolism in chronic kidney failure. *J Nutr* 137, 1586S-1590S; discussion 1597S-1598S.

**Levey, A. S., Eckardt, K. U., Tsukamoto, Y., Levin, A., Coresh, J., Rossert, J., et al. Eknoyan, G. (2005).** Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int* 67, 2089-2100.

**Lucio, M. (2008).** Datamining metabolomics: the convergence point of non-target approach and statistical investigation. In *Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt*, pp. 191: Technische Universität München.

**Lundin, U. & Weinberger, K. (2010).** New Biomarkers for Assessing Kidney Diseases. In *PCT*, pp. 68. Edited by A. Biocrates Life Sciences. AT: Biocrates Life Sciences AG.

**Mahner, M. & Kary, M. (1997).** What exactly are genomes, genotypes and phenotypes? And what about phenomes? *Journal of theoretical biology* 186, 55-63.

**Makinen, V. P., Tynkkynen, T., Soininen, P., Forsblom, C., Peltola, T., Kangas, A. J., et al. Ala-Korpela, M. (2012).** Sphingomyelin is associated with kidney disease in type 1 diabetes (The FinnDiane Study). *Metabolomics* 8, 369-375.

**Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. (1998).** Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 17, 1-35.

**Marshall, A. G. (2000).** Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. *International Journal of Mass Spectrometry* 200, 331-356.

**Marshall, A. G. & Hendrickson, C. L. (2008).** High-resolution mass spectrometers. *Annu Rev Anal Chem (Palo Alto Calif)* 1, 579-599.

**Mateo Anson, N., van den Berg, R., Havenaar, R., Bast, A. & Haenen, G. R. M. M. (2009).** Bioavailability of ferulic acid is determined by its bioaccessibility. *Journal of Cereal Science* 49, 296-300.

**Mazzeo, J. R., D. Neue, U., Kele, M. & Plumb, R. S. (2005).** Advancing LC Performance with Smaller Particles and Higher Pressure. *Analytical Chemistry* 77, 460 A-467 A.

**McDaniel, E. W., Martin, D. W. & Barnes, W. S. (1962).** Drift Tube-Mass Spectrometer for Studies of Low-Energy Ion-Molecule Reactions. *Review of Scientific Instruments* 33, 2-7.

**McLafferty, F. W. (1981).** Tandem mass spectrometry. *Science* 214, 280-287.

**Miller, G. W., Lock, E. A. & Schnellmann, R. G. (1994).** Strychnine and glycine protect renal proximal tubules from various nephrotoxicants and act in the late phase of necrotic cell injury. *Toxicol Appl Pharmacol* 125, 192-197.

**Miller, W. G., Bruns, D. E., Hortin, G. L., Sandberg, S., Aakre, K. M., McQueen, M. J., et al. National Kidney Disease Education Program, I. W. G. o. S. o. A. i. U. (2009).** Current issues in measurement and reporting of urinary albumin excretion. *Clin Chem* 55, 24-38.

**Molitch, M. E., DeFronzo, R. A., Franz, M. J., Keane, W. F., Mogensen, C. E., Parving, H. H., et al. American Diabetes, A. (2004).** Nephropathy in diabetes. *Diabetes Care* 27 Suppl 1, S79-83.

**Moller, N., Meek, S., Bigelow, M., Andrews, J. & Nair, K. S. (2000).** The kidney is an important site for in vivo phenylalanine-to-tyrosine conversion in adult humans: A metabolic role of the kidney. *Proc Natl Acad Sci U S A* 97, 1242-1246.



**Mooney, S., Leuendorf, J. E., Hendrickson, C. & Hellmann, H. (2009).** Vitamin B6: a long known compound of surprising complexity. *Molecules* 14, 329-351.

**Müller, C. (2013).**Metabolomics in host-pathogen interactions

- An investigation of Chlamydia infected human cells. In *Fakultät für Wissenschaftszentrum Weihenstephan für Ernährung*, pp. 197: TECHNISCHE UNIVERSITÄT MÜNCHEN.

**Nagai, R., Nagai, M., Shimasaki, S., Baynes, J. W. & Fujiwara, Y. (2010).** Citric acid inhibits development of cataracts, proteinuria and ketosis in streptozotocin (type 1) diabetic rats. *Biochem Biophys Res Commun* 393, 118-122.

**Nagrebetsky, A., Jin, J., Stevens, R., James, T., Adler, A., Park, P., et al. Farmer, A. (2012).** Diagnostic accuracy of urine dipstick testing in screening for microalbuminuria in type 2 diabetes: a cohort study in primary care. *Family practice*.

**Nakamura, S., Li, H., Adijiang, A., Pischetsrieder, M. & Niwa, T. (2007).** Pyridoxal phosphate prevents progression of diabetic nephropathy. *Nephrol Dial Transplant* 22, 2165-2174.

**Ng, D. P., Salim, A., Liu, Y., Zou, L., Xu, F. G., Huang, S., et al. Ong, C. N. (2012).** A metabolomic study of low estimated GFR in non-proteinuric type 2 diabetes mellitus. *Diabetologia* 55, 499-508.

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

**Nicholson, J. K. & Lindon, J. C. (2008).** Systems biology: Metabonomics. *Nature* 455, 1054-1056.

**Niewczas, M. A., Sirich, T. L., Mathew, A. V., Skupien, J., Mohny, R. P., Warram, J. H., et al. Krolewski, A. S. (2014).** Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int* 85, 1214-1224.

**Ohta, D., Kanaya, S. & Suzuki, H. (2010).** Application of Fourier-transform ion cyclotron resonance mass spectrometry to metabolic profiling and metabolite identification. *Curr Opin Biotechnol* 21, 35-44.

**Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. (1998).** Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16, 373-378.

**Osswald, H. & Schnermann, J. (2011).** Methylxanthines and the kidney. *Handbook of experimental pharmacology*, 391-412.

**Ott, M. A. & Vriend, G. (2006).** Correcting ligands, metabolites, and pathways. *BMC Bioinformatics* 7, 517.

**Ou, S. & Kwok, K.-C. (2004).** Ferulic acid: pharmaceutical functions, preparation and applications in foods. *Journal of the Science of Food and Agriculture* 84, 1261-1269.

**Pang, L. Q., Liang, Q. L., Wang, Y. M., Ping, L. & Luo, G. A. (2008).** Simultaneous determination and quantification of seven major phospholipid classes in human blood using normal-phase liquid chromatography coupled with electrospray mass spectrometry and the application in diabetes nephropathy. *J Chromatogr B Analyt Technol Biomed Life Sci* 869, 118-125.

**Papale, M., Di Paolo, S., Magistroni, R., Lamacchia, O., Di Palma, A. M., De Mattia, A., et al. Gesualdo, L. (2010).** Urine proteome analysis may allow noninvasive differential diagnosis of diabetic nephropathy. *Diabetes Care* 33, 2409-2415.

**Patil, K. R. & Nielsen, J. (2005).** Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 102, 2685-2689.

**Pearson, H. (2007).** Meet the human metabolome. *Nature* 446, 8.

**Peng, C. C., Hsieh, C. L., Wang, H. E., Chung, J. Y., Chen, K. C. & Peng, R. Y. (2012).** Ferulic acid is nephrodamaging while gallic acid is renal protective in long term treatment of chronic kidney disease. *Clinical nutrition* 31, 405-414.

**Perassolo, M. S., Almeida, J. C., Pra, R. L., Mello, V. D., Maia, A. L., Moulin, C. C., et al. Gross, J. L. (2003).** Fatty acid composition of serum lipid fractions in type 2 diabetic patients with microalbuminuria. *Diabetes Care* 26, 613-618.

**Pero, R. W. (2010).** Health consequences of catabolic synthesis of hippuric acid in humans. *Current clinical pharmacology* 5, 67-73.

**Perroud, B., Lee, J., Valkova, N., Dhirapong, A., Lin, P. Y., Fiehn, O., et al. Weiss, R. H. (2006).** Pathway analysis of kidney cancer using proteomics and metabolic profiling. *Mol Cancer* 5, 64.

**Peti-Peterdi, J., Kang, J. J. & Toma, I. (2008).** Activation of the renal renin-angiotensin system in diabetes--new concepts. *Nephrol Dial Transplant* 23, 3047-3049.

**Piarulli, F., Sartore, G., Ceriello, A., Ragazzi, E., Reitano, R., Nollino, L., et al. Lapolla, A. (2009).** Relationship between glyco-oxidation, antioxidant status and microalbuminuria in type 2 diabetic patients. *Diabetologia* 52, 1419-1425.

**Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. & Grant, R. (2003).** Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 785, 263-275.

**Poole, C. F., (2003).** The essence of chromatography. Amsterdam [u.a.]: Elsevier.

**Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., et al. Scrivens, J. H. (2007).** An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* 261, 1-12.

**Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., et al. Wishart, D. S. (2011).** The human serum metabolome. *PLoS ONE* 6, e16957.

**Rajabi, K., Easterling, M. L. & Fridgen, T. D. (2009).** Solvation of electrosprayed ions in the accumulation/collision hexapole of a hybrid Q-FTMS. *J Am Soc Mass Spectrom* 20, 411-418.

**Ralib, A. M., Pickering, J. W., Shaw, G. M., Devarajan, P., Edelstein, C. L., Bonventre, J. V. & Endre, Z. H. (2012).** Test characteristics of urinary biomarkers depend on quantitation method in acute kidney injury. *Journal of the American Society of Nephrology : JASN* 23, 322-333.

**Ramsay, S. L., Stöggli, W. M., Weinberger, K. M., Graber, A. & Guggenbichler, W. (2007).** Apparatus and method for analyzing a metabolite profile. In *PCT*, pp. 68. Edited by A. Biocrates Life Sciences. AT: Biocrates Life Sciences AG.

**Rao, N. P., Preminger, G. M. & Kavanagh, J. P., (2011).** Urinary Tract Stone Disease. London: Springer-Verlag London Limited.

**Rathi, S., Kern, W. & Lau, K. (2007).** Vitamin C-induced hyperoxaluria causing reversible tubulointerstitial nephritis and chronic renal failure: a case report. *Journal of medical case reports* 1, 155.

**Reemtsma, T. (2009).** Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry: status and needs. *J Chromatogr A* 1216, 3687-3701.

**Rossing, K., Mischak, H., Dakna, M., Zurbig, P., Novak, J., Julian, B. A., et al. Network, P. (2008).** Urinary proteomics in diabetes and CKD. *Journal of the American Society of Nephrology : JASN* 19, 1283-1290.

**Rotem, R., Heyfets, A., Fingrut, O., Blickstein, D., Shaklai, M. & Flescher, E. (2005).** Jasmonates: novel anticancer agents acting directly and selectively on human cancer cell mitochondria. *Cancer Res* 65, 1984-1993.

**Roullier-Gall, C., Lucio, M., Noret, L., Schmitt-Kopplin, P. & Gougeon, R. D. (2014).** How Subtle Is the “Terroir” Effect? Chemistry-Related Signatures of Two “Climats de Bourgogne”. *PLoS ONE* 9, e97615.

**Ruilope, L., Izzo, J., Haller, H., Waeber, B., Oparil, S., Weber, M., et al. Sowers, J. (2010).** Prevention of microalbuminuria in patients with type 2 diabetes: what do we know? *Journal of clinical hypertension* 12, 422-430.

**Rule, A. D., Bailey, K. R., Schwartz, G. L., Khosla, S., Lieske, J. C. & Melton, L. J., 3rd (2009).** For estimating creatinine clearance measuring muscle mass gives better results than those based on demographics. *Kidney Int* 75, 1071-1078.

**Rutter, G. A. & Parton, L. E. (2008).** The beta-cell in type 2 diabetes and in obesity. *Frontiers of hormone research* 36, 118-134.

**Saal, S. & Harvey, S. J. (2009).** MicroRNAs and the kidney: coming of age. *Curr Opin Nephrol Hypertens* 18, 317-323.

**Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., et al. Quackenbush, J. (2006).** TM4 microarray software suite. *Methods Enzymol* 411, 134-193.

**Satko, S. G., Sedor, J. R., Iyengar, S. K. & Freedman, B. I. (2007).** Familial clustering of chronic kidney disease. *Seminars in dialysis* 20, 229-236.

**Schafer, S., Kantartzis, K., Machann, J., Venter, C., Niess, A., Schick, F., et al. Stefan, N. (2007).** Lifestyle intervention in individuals with normal versus impaired glucose tolerance. *Eur J Clin Invest* 37, 535-543.

**Schartum-Hansen, H., Ueland, P. M., Pedersen, E. R., Meyer, K., Ebbing, M., Bleie, O., et al. Nygard, O. (2013).** Assessment of urinary betaine as a marker of diabetes mellitus in cardiovascular patients. *PLoS ONE* 8, e69454.

**Schmieder, R. E. & Bramlage, P. (2012).** Diagnostik, Bedeutung, Prävention und Therapie einer Mikroalbuminurie - Erwiderung. *Dtsch Med Wochenschr* 137, 2390-2391.

**Schwab, W. (2003).** Metabolome diversity: too few genes, too many metabolites? *Phytochemistry* 62, 837-849.

**Seo, J. (2002).** Interactively Exploring Hierarchical Clustering Results. In *Computer*, pp. 80-86. Edited by S. Ben.

**Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. Ideker, T. (2003).** Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498-2504.

**Shapiro, H., Theilla, M., Attal-Singer, J. & Singer, P. (2011).** Effects of polyunsaturated fatty acid consumption in diabetic nephropathy. *Nature reviews Nephrology* 7, 110-121.

**Shaw, J. E., Sicree, R. A. & Zimmet, P. Z. (2010).** Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 87, 4-14.

**Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. (2011).** Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431-432.

**Sogabe, K., Roeser, N. F., Venkatachalam, M. A. & Weinberg, J. M. (1996).** Differential cytoprotection by glycine against oxidant damage to proximal tubule cells. *Kidney Int* 50, 845-854.

**Soh, K. C. & Hatzimanikatis, V. (2010).** Network thermodynamics in the post-genomic era. *Curr Opin Microbiol* 13, 350-357.

**Solini, A. & Ferrannini, E. (2011).** Pathophysiology, prevention and management of chronic kidney disease in the hypertensive patient with diabetes mellitus. *Journal of clinical hypertension* 13, 252-257.

**Stumvoll, M., Goldstein, B. J. & van Haeften, T. W. (2005).** Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365, 1333-1346.

**Suhre, K. & Schmitt-Kopplin, P. (2008).** MassTRIX: mass translator into pathways. *Nucleic Acids Res* 36, W481-484.

- Szymanska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. (2012).** Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 8, 3-16.
- Thews, G., Mutschler, E. & Vaupel, P. (1999).** Anatomie, Physiologie, Pathophysiologie des Menschen, 5. edn. Stuttgart: Wiss Verl-Ges.
- Tobe, S. W., McFarlane, P. A. & Naimark, D. M. (2002).** Microalbuminuria in diabetes mellitus. *CMAJ* 167, 499-503.
- Tsujimoto, M., Sugimoto, S., Nagatomo, M., Furukubo, T., Izumi, S., Yamakawa, T., et al. Nishiguchi, K. (2013).** Possibility of Decrease in CYP1A2 Function in Patients With End-stage Renal Disease. *Therapeutic apheresis and dialysis : official peer-reviewed journal of the International Society for Apheresis, the Japanese Society for Apheresis, the Japanese Society for Dialysis Therapy*.
- Tun, H. W., Marlow, L. A., von Roemeling, C. A., Cooper, S. J., Kreinest, P., Wu, K., et al. Copland, J. A. (2010).** Pathway signature and cellular differentiation in clear cell renal cell carcinoma. *PLoS ONE* 5, e10696.
- Tzamaloukas, A. H. & Murata, G. H. (2005).** Prevention of nephropathy in patients with type 2 diabetes mellitus. *International urology and nephrology* 37, 655-663.
- Tziotis, D., Hertkorn, N. & Schmitt-Kopplin, P. (2011).** Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur J Mass Spectrom (Chichester, Eng)* 17, 415-421.
- Valk, E. J., Bruijn, J. A. & Bajema, I. M. (2011).** Diabetic nephropathy in humans: pathologic diversity. *Curr Opin Nephrol Hypertens* 20, 285-289.
- Van Buren, P. N. & Toto, R. (2011).** Hypertension in diabetic nephropathy: epidemiology, mechanisms, and management. *Advances in chronic kidney disease* 18, 28-41.
- van der Kloet, F. M., Tempels, F. W., Ismail, N., van der Heijden, R., Kasper, P. T., Rojas-Cherto, M., et al. Hankemeier, T. (2012).** Discovery of early-stage biomarkers for diabetic kidney disease using ms-based metabolomics (FinnDiane study). *Metabolomics* 8, 109-119.
- van Dieren, S., Beulens, J. W., van der Schouw, Y. T., Grobbee, D. E. & Neal, B. (2010).** The global burden of diabetes and its complications: an emerging pandemic. *European journal of cardiovascular prevention and rehabilitation : official journal of*

*the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology* 17 Suppl 1, S3-8.

**Vaziri, N. D., Wong, J., Pahl, M., Piceno, Y. M., Yuan, J., DeSantis, T. Z., et al. Andersen, G. L. (2013).** Chronic kidney disease alters intestinal microbial flora. *Kidney Int* 83, 308-315.

**Wagele, B., Witting, M., Schmitt-Kopplin, P. & Suhre, K. (2012).** MassTRIX reloaded: combined analysis and visualization of transcriptome and metabolome data. *PLoS ONE* 7, e39860.

**Wang-Sattler, R., Yu, Z., Herder, C., Messias, A. C., Floegel, A., He, Y., et al. Illig, T. (2012).** Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 8, 615.

**Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., et al. Gerszten, R. E. (2011).** Metabolite profiles and the risk of developing diabetes. *Nat Med* 17, 448-453.

**Wang, W., Wu, Z., Dai, Z., Yang, Y., Wang, J. & Wu, G. (2013).** Glycine metabolism in animals and humans: implications for nutrition and health. *Amino Acids*.

**Warrack, B. M., Hnatyshyn, S., Ott, K. H., Reily, M. D., Sanders, M., Zhang, H. & Drexler, D. M. (2009).** Normalization strategies for metabolomic analysis of urine samples. *J Chromatogr B Analyt Technol Biomed Life Sci* 877, 547-552.

**Weinberg, J. M., Davis, J. A., Abarzua, M. & Rajan, T. (1987).** Cytoprotective effects of glycine and glutathione against hypoxic injury to renal tubules. *J Clin Invest* 80, 1446-1454.

**Weir, M. R. (2007).** Microalbuminuria and cardiovascular disease. *Clin J Am Soc Nephrol* 2, 581-590.

**WHO, Dept. of Noncommunicable Disease Surveillance (1999).** Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. Geneva: World Health Organisation.

**Wilm, M. (2011).** Principles of electrospray ionization. *Mol Cell Proteomics* 10, M111 009407.

- Winchester, J. F., Feinfeld, D. A., Harbord, N. B. & Dubrow, A. (2010).** Diabetic Nephropathy. In *Principles of Diabetes Mellitus*, pp. Online Ressource (18074 KB, 18852 S.). Edited by L. Poretsky: Springer-Verlag.
- Wolf, G. & Ritz, E. (2003).** Diabetic nephropathy in type 2 diabetes prevention and patient management. *Journal of the American Society of Nephrology : JASN* 14, 1396-1405.
- Wörmann, K., Walker, A., Moritz, F., Forcisi, S., Tziotis, D., Lucio, M., et al.Schmitt-Kopplin, P. (2012).** Revolution in der Diabetesdiagnostik dank -omics - Biomarker mittels Metabolomics. *Diabetes aktuell* 10, 129-133.
- Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D. S. (2012).** MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* 40, W127-133.
- Xia, J. F., Liang, Q. L., Hu, P., Wang, Y. M., Li, P. & Luo, G. A. (2009).** Correlations of six related purine metabolites and diabetic nephropathy in Chinese type 2 diabetic patients. *Clin Biochem* 42, 215-220.
- Yamada, T., Letunic, I., Okuda, S., Kanehisa, M. & Bork, P. (2011).** iPath2.0: interactive pathway explorer. *Nucleic Acids Res* 39, W412-415.
- Yang, W. L., Bai, Q., Li, D. D., A, T. L., Wang, S., Zhao, R. S., et al.Fan, M. H. (2013).** Changes of urinary phospholipids in the chronic kidney disease patients. *Biomarkers* 18, 601-606.
- Zajac, J., Shrestha, A., Patel, P. & Poretsky, L. (2010).** The Main Events in the History of Diabetes Mellitus. In *Principles of Diabetes Mellitus*, pp. Online Ressource (18074 KB, 18852 S.). Edited by L. Poretsky. [s.l.]: Springer-Verlag.
- Zelezniak, A., Pers, T. H., Soares, S., Patti, M. E. & Patil, K. R. (2010).** Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. *PLoS computational biology* 6, e1000729.
- Zhao, Y. Y., Lei, P., Chen, D. Q., Feng, Y. L. & Bai, X. (2013).** Renal metabolic profiling of early renal injury and renoprotective effects of *Poria cocos* epidermis using UPLC Q-TOF/HSMS/MSE. *J Pharm Biomed Anal* 81-82, 202-209.



## 8 Summary

### Summary

Metabolomics relates to the research of metabolites, their pathways and interactions. The potential of metabolomics is shown in two studies aiming for identification of prognostic and diagnostic biomarkers for diabetic nephropathy. Thus human urine samples were analysed using targeted and non-targeted approaches. The development of UHPLC-MS and FT-ICR-MS methods, the identification of signals with MSMS and correlation analysis as well as a comparison between different generations of instruments are presented.

### Zusammenfassung

Die Metabolomik befasst sich mit der Erforschung der Metabolite, deren Stoffwechselwegen und Interaktionen. Das Potential der Metabolomik wird in zwei Studien zur Identifizierung von Biomarkern für die Prognose und Diagnose von diabetischer Nephropathie gezeigt. Dafür wurden Humanurinproben mittels zielgerichteter und nicht zielgerichteter Analysen untersucht. Es werden die Methodenentwicklung von UHPLC-MS und FT-ICR-MS Methoden, die Identifizierung von Signalen mittels MSMS und Korrelationsanalysen und ein Vergleich verschiedener Gerätegenerationen gezeigt.

## Scientific Contribution

### Posters and presentations:

Kilian Wörmann, Sara Forcisi, Rainer Lehmann, Agnes Fekete, Philippe Schmitt  
Kopplin HPLC 2009 Conference, June 28-July 2009, Dresden, Germany

Poster presentation

### **Study of different sample preparation and separation conditions for untargeted metabolic analysis in urine**

Kilian Wörmann

Visiting guest, October 2009, National Chromatographic R. & A. Center, Dalian  
Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

### **Oral introduction to the project**

Kilian Wörmann, Sara Forcisi, Rainer Lehmann, Agnes Fekete, Holger Franken,  
Erhan Kenar, Guowang Xu, Philippe Schmitt-Kopplin

1st International Symposium on Metabolomics and More, March 10-12, 2010,  
Freising-Weißenstephan, Germany

Poster presentation

### **UHPLC/TOF-MS method development for untargeted metabolic analysis in urine**

### Publications:

Wörmann K., Lucio M., Forcisi S., Heinzmann S.S., Kenar E., Franken H.,  
Rosenbaum L., Schmitt-Kopplin P., Kohlbacher O., Zell A., Häring H.-U., and  
Lehmann R., **Metabolomics in der Diabetesforschung**, Der Diabetologe 2012,  
1-5, Springer-Verlag, DOI:10.1007/s11428-011-0778-9.

Wörmann K., Walker A., Moritz F., Forcisi S., Tziotis D., Lucio M., Heinzmann S. S., Adamski J., Lehmann R., Häring H. U., Schmitt-Kopplin P., Revolution in der Diabetesdiagnostik dank –omics. Biomarker mittels Metabolomics

**Revolution in Diabetes Diagnostics - Metabolomics for Discovering Biomarkers**

Diabetes aktuell 2012, 10, 129-133, Georg Thieme Verlag, Stuttgart /New-York, DOI: 10.1055/s-0032-1320054.

Kenar E., Franken H., Rosenbaum L., Lehmann R., Forcisi S., Wörmann K., Lucio M., König A., Rahnenführer J., Schmitt-Kopplin P., Häring H.-U., Zell A., Kohlbacher O., **Mit bioinformatik zu biomarkern**, Medizinische Welt 2012, 63, 245-250.

Kenar E., Franken H., Forcisi S., Wörmann K., Häring H. U., Lehmann R., Schmitt-Kopplin P., Zell A., Kohlbacher O., Automated Label-Free

**Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data.**, Mol Cell Proteomics 2014,

DOI: 10.1074/mcp.M113.031278

## Lebenslauf

### Persönliche Daten:

Name: Kilian Wörmann  
Geburtsdatum: 13.12.1980  
Geburtsort: Schongau  
Nationalität : Deutschland

### Bildung:

02/2009 - 07/2012 Doktorand am Helmholtz Zentrum München

Thema: New prognostic and diagnostic biomarkers in diabetic nephropathy using comprehensive targeted and non-targeted metabolomics

Betreuung: apl. Prof. Dr. Philipp Schmitte-Kopplin (Technische Universität München, Lehrstuhl für Analytische Lebensmittelchemie, Alte Akademie, Freising, Deutschland)

01/2008 - 09/2008 Diplomarbeit im Arbeitskreis Prof.Dr. Oliver Reiser, Institut für Organische Chemie, Universität Regensburg

Thema: Untersuchungen zur Derivatisierung des a,b-ungesättigten Carbonylsystems des Chalcones Calythrospin

10/2003 - 09/2008 Diplom-Chemie Studium an der Universität Regensburg

09/2001 – 07/2003 Fachgebundene Hochschulreife Ausbildungsrichtung Technik  
Städtische Marian-Batko-Berufsoberschule München (2. Jahr)  
Staatliche Berufsoberschule Weilheim (1. Jahr)

09/1997 – 07/2000 Ausbildung zum Chemielaboranten  
Roche Diagnostics GmbH, Penzberg