Identification of novel biomarkers for predicting outcome of acute and chronic kidney disease

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Helena U. Zacharias

aus

Regensburg

im Jahr 2016

Das Promotionsgesuch wurde eingereicht am:

Die Arbeit wurde angeleitet von:
Prof. Dr. Wolfram Gronwald

Unterschrift:

Helena Zacharias

This Ph.D. thesis has in parts already been published in [Zacharias 2012, Zacharias et al. 2013a), Zacharias et al. 2013b), Zacharias et al. 2015, Hochrein et al. 2015].

Fig. 4.2, page 39, Fig. 4.3, page 40, Fig. 4.4, page 41, and Fig. 4.6, page 47, have already been published in [Zacharias 2012] and were provided by me.
A section on page 45 has already been published in a slightly altered version in [Zacharias et al. 2013b)]. Claudia Samol wrote the first draft of the manuscript and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
A section on page 49-50 has already been published in a slightly altered version in [Zacharias et al. 2013b)].
Sections on pages 52-53 and 56-58 have already been published in slightly altered versions in [Zacharias et al. 2013b)]. Dipl. Math. Jochen Hochrein wrote the first draft of the manuscript and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Section 4.3.2, page 61, has already been published in slightly altered versions in [Zacharias et al. 2013b)]. I wrote the first draft of the manuscript and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Section 5.1.2.1, pages 67-68, has already been published in [Zacharias et al. 2013a)] and [Zacharias et al. 2015] in a slightly altered version. I wrote the first drafts of the manuscripts and the published articles had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner. Patient handling as well as biofluid collection and assessment of clinical data had been conducted at the University of Erlangen-Nuremberg.
Section 5.1.2.5, page 70-71, has already been published in [Hochrein et al. 2015] in a slightly altered version, and is also part of Dipl. Math. Jochen Hochrein's Ph.D. thesis [Hochrein 2016]. The reported analyses have been performed by Prof. Dr. Wolfram Gronwald and Claudia Samol and the manuscript has been written by Prof. Dr. Wolfram Gronwald.
Sections on pages 71-72 have already been published in a slightly modified version in [Zacharias et al. 2015]. The classification/prognostication concept of nested cross-validation employed here was implemented by Dipl. Math. Jochen Hochrein and is also part of his Ph.D. thesis [Hochrein 2016]. I employed this nested cross-validation and analyzed the corresponding results.
Section 5.1.2.8, page 73, has already been published in [Zacharias et al. 2015] in a slightly modified version. I slightly modified the applied linear SVM cross-validation with optimization of the cost parameter C from a previous R-code of Dipl. Math. Jochen Hochrein.
Section 5.1.2.9, page 73, and section 5.1.3.7, pages 85-86, have already been published in [Zacharias et al. 2015] in slightly altered versions, and the presented concept as well as the corresponding algorithm was developed by Dipl. Math. Jochen Hochrein. It is also part of his Ph.D. thesis [Hochrein 2016]. I performed data analyses and wrote the first draft of [Zacharias et al. 2015] and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Parts of section 5.1.2.10, pages 74, have already been published in [Zacharias et al. 2015] in a slightly altered version. I performed both method implementation as well as data analyses. I wrote the first draft of the manuscript and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.

Section 5.1.3.1, pages 74-76 has already been published in [Zacharias et al. 2015], and [Hochrein et al. 2015] in a slightly altered version. The presented strategy to choose the appropriate data normalization method was developed by Dipl. Math. Jochen Hochrein and is also part of his Ph.D. thesis [Hochrein 2016]. I performed data analyses and wrote the first draft of [Zacharias et al. 2015], and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Section 5.1.3.2, page 76, has already been published in [Zacharias et al. 2013a)] in a slightly altered version. I performed data analyses and wrote the first draft of the manuscript, and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Parts of section 5.1.3.3, pages 77-80, have already been published in [Zacharias et al. 2015] in a slightly altered version. MS measurements and corresponding data analyses were performed by M.Sc. Franziska Vogl. I performed all measurements and data analyses corresponding to NMR spectroscopy.
Parts of section 5.1.3.4, pages 80-81, section 5.1.3.5, pages 81-82, and section 5.1.3.6, pages 83-84 have already been published in [Zacharias et al. 2015] in slightly altered versions. I performed data analyses and wrote the first draft of the manuscript and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner. I generated Fig. 5.2, page 84, and Fig. 7.3, page 146, and they were reprinted with permission from [Zacharias et al. 2015]. Copyright 2015 American Chemical Society.
Section 5.1.4, pages 87-89, has already been published in [Zacharias et al. 2015], and [Hochrein et al. 2015] in a slightly altered version. I wrote the first draft of [Zacharias et al. 2015] and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.
Appendix II section 7.2.2, page 135, has already been published in [Zacharias et al. 2015]. I wrote the first draft of [Zacharias et al. 2015] and the published article had been revised by Prof. Dr. Wolfram Gronwald and Prof. Dr. Peter Oefner.

Danksagung

Hiermit bedanke ich mich herzlichst bei allen, die mir während dieser Doktrorarbeit mit Rat und Tat zur Seite standen:

Vielen herzlichen Dank an Prof. Dr. Wolfram Gronwald, der mir die Möglichkeit gab, solch interessante Projekte während meiner Doktorarbeit zu bearbeiten sowie für seinen fortwährende, freundliche Anleitung.
Prof. Dr. Peter J. Oefner danke ich für die Möglichkeit, meine Doktorarbeit an seinem Institut zu schreiben und für seine konstruktive Anleitung.
Prof. Dr. Kai-Uwe Eckardt, Prof. Dr. Carsten Willam, Dr. Gunnar Schley, und Dr. Stephanie Titze vom Universitätsklinikum Erlangen-Nürnberg danke ich für die fruchtbaren Kollaborationen.
Außerdem möchte ich mich bei Prof. Dr. Kai-Uwe Eckardt sowie Prof. Dr. Werner Kremer für ihr Mentoring während meiner Doktorarbeit herzlichst bedanken.
Mein besonderer Dank geht an Dipl. Math. Jochen Hochrein für seine Unterstützung und Unterweisung im Bezug auf Algorithmenentwicklung sowie für die erfolgreiche Zusammenarbeit.
M.Sc. Franziska Vogl danke ich für ihre Massenspektrometriemessungen, ohne die ich vielleicht immer noch über ein einzelnes NMR-Signal grübeln würde. Herzlichsten Dank geht an Claudia Samol für ihre Hilfe im Labor und das Pipettieren wirklich unzähliger Proben. Außerdem möchte ich mich bei Dr. Matthias Klein für seine Anleitung und Hilfe bei der theoretischen und praktischen Arbeit, sowie bei Prof. Dr. Rainer Spang, Dr. Claudio Lottaz und Dr. Guisi Moffa für ihre Beratung in statistischen Fragen bedanken.
Großen Dank möchte ich meinen Bürokollegen Dipl. Math. Jochen Hochrein, M.Sc. Philipp Schwarzfischer, Dr. Matthias Klein, Dr. Claudio Lottaz, Dr. Matthias Maneck, M.Sc. Christian Kohler, Dr. Mohammad Sadeh, Dr. Paula Perez-Rubio, und Dr. Michael Altenbuchinger für die nette Arbeitsatmosphäre aussprechen.
Vor allem jedoch möchte ich mich, neben den bereits oben erwähnten, bei allen weiteren Mitgliedern des Instituts für funktionelle Genomik für die tolle Zusammenarbeit und Gemeinschaft bedanken: Dr. Martin Almstetter, Dr. Inka Appel, Dr. Nadine Aßmann, M.Sc. Raffaela Berger, Sabine Botzler, Dr. Katja Dettmer-Wilde, Lisa Ellmann, Eva Engl, Dr. Julia Engelmann, Dr. Mauritz Evers, Corinna Feuchtinger, M.Sc. Franziska Görtler, Dr. Daniela Herold, Dr. Bianca Höfelschweiger, Dr. Christian Hundsrucker, B.Sc. Elena Kremen, B.Sc. Sebastian Mehrl, Dr. Katharina Meyer, Dipl. Bioinf. Anton Moll, Sandra Nothmann, Nadine Nürnberger, Elke Perthen, Dipl. Math. Martin Pirkl, B.Sc. Sandra Rauh, M.Sc. Thorsten Rehberg, Dr. Joerg Reinders, Dr. Yvonne Reinders, Dipl. Humanbiol. Sophie Schirmer, Dr. Inga Schlecht, M.Sc. Trixi von Schlippenbach, M.Sc. Johann Simbürger, Dipl. Bioinf. Frank Stämmler, Dr. Nicholas Strieder, Dipl. Bioinf. Franziska Taruttis, Dipl. Biol. Anja Thomas,

Dr. Monica Totir, Dr. Christian Wachsmuth, Dr. Magdalena Waldhier, B.Sc. Martha Wiesmann, und Dr. Wentao Zhu.
Ein besonderer Dank geht an meine Freunde an der Universität Regensburg für den Zusammenhalt über so viele Jahre.
Ebenso möchte ich meinem Freund Michael für seine Unterstützung und sein Verständnis danken, sowie für die zahlreichen schönen Stunden in Regensburg, München, Hermannsöd, Florenz, Mailand, Brüssel, Antwerpen, Dresden, Venedig und London.
Zuletzt möchte ich meinen Eltern für ihre immerwährende Unterstützung und Liebe danken, ohne die ich wohl nicht so weit gekommen wäre.

Contents

Abbreviations

1D	one-dimensional
2D	two-dimensional
ACE	angiotensin-converting-enzyme
ACR	albumin-creatinine ratio
ADPKD	autosomal dominant polycystic kidney disease
AER	albumin excretion rate
AKI	acute kidney injury
AKIN	Acute Kidney Injury Network
AP	alkaline phosphatase
ARF	acute renal failure
a. u.	arbitrary units
AUC	area under the curve
B/H	Benjamini/Hochberg
BMI	body mass index
BSA	body surface area
CABG	coronary artery bypass grafting
CGA	Cause, GFR, and Albuminuria
ChEBI	Chemical Entities of Biological Interest
CKD	chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CKD-EPI crea	CKD-EPI formula based on SCr
CKD-EPI crea cys	CKD-EPI formula based on SCr and SCysC
CKD-EPI cys	CKD-EPI formula based on SCysC
CKD-JAC	Chronic Kidney Disease Japan Cohort
COPD	chronic obstructive pulmonary disease
COSY	Correlated Spectroscopy
CPB	cardiopulmonary bypass
CPMG	Carr-Purcell-Meiboom-Gill
CRIC	Chronic Renal Insufficiency Cohort
CRP	C-reactive protein
CysC	cystatin C
EDTA	ethylenediaminetetraacetic acid
eGFR	estimated glomerular filtration rate
$eGFR_{ckdepi\ crea}$	eGFR based on CKD-EPI crea formula
$eGFR_{ckdepi\ crea\ cys}$	eGFR based on CKD-EPI crea cys formula

$eGFR_{ckdepi\ cys}$	eGFR based on CKD-EPI cys formula
$eGFR_{mdrd4}$	eGFR based on MDRD4 formula
EPO	erythropoietin
ESA	erythropoiesis-stimulating agent
ESI	electrospray ionization
ESKD	End Stage Kidney Disease
ESRD	end-stage renal disease
FDR	false discovery rate
FFP	fresh frozen plasma
FID	free induction decay
FU2	second follow-up
GCKD	German Chronic Kidney Disease
GFR	glomerular filtration rate
GGT	γ-glutamyltranspeptidase
Hb	hemoglobin
HbA_{1c}	glycated hemoglobin
HF	hemofiltration
HMBC	heteronuclear multiple bond correlation
HMDB	Human Metabolome Database
HPLC	high-performing liquid chromatography
HSQC	heteronuclear single quantum coherence
IABP	intra-aortic balloon pump
ICU	intensive care unit
IL-18	interleukin-18
INEPT	insensitive nuclei enhanced by polarization transfer
KDIGO	Kidney Disease: Improving Global Outcomes
KIM-1	kidney injury molecule-1
LARS	least-angle regression
LASSO	least absolute shrinkage and selection operator
LC-MS	liquid chromatography-mass spectrometry
LTB_4	leukotriene B_4
MAP	mean arterial pressure
MDRD	Modification of Diet in Renal Disease
MDRD4	four-variable Modification of Diet in Renal Disease
MS	mass spectrometry
mse	mean-squared error
NAG	N-acetyl-β-D-glucosaminidase
NGAL	neutrophil-gelatinase-associated lipocalin
NMR	nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NSAID	nonsteroidal anti-inflammatory drugs
OPLS-DA	orthogonal projection to latent structures discriminant analysis

PAVD	peripheral arterial vascular disease
PC	principal component
PCA	principal component analysis
PLS-DA	Partial Least Squares Discriminant Analysis
RCC	red cell concentrate
rf	radio frequency
RF	Random Forests
RIFLE	Risk Injury Failure Loss End-Stage Renal Disease
ROC	receiver operating characteristic
RRT	renal replacement therapies
RSS	residual sum of squares
SCr	serum creatinine
SCysC	serum cystatin C
SVM	Support Vector Machine
TMS	tetramethylsilan
TOCSY	Total Correlation Spectroscopy
TOFMS	time-of-flight mass spectrometry
TSP	3-trimethylsilyl-2,2,3,3-tetradeuteropropionate
TREAT	Trial to Reduce Cardiovascular Events with Aranesp® Therapy
UO	urine output
UPLC/QTOFMS	ultra-performance reversed-phase liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer
VSN	Variance Stabilization normalization

1 Abstract

The global burden of human renal diseases continually increased in the last decades. To lower associated mortality and morbidity rates, early diagnosis as well as improved understanding of underlying biological mechanisms are essential. Here, metabolic investigations of biofluids by means of nuclear magnetic resonance (NMR) spectroscopy in the context of nephrology are presented to facilitate earlier detection and to enable new insights into renal disease manifestation.

The detection of novel low-molecular-weight factors for improved early diagnosis and patient treatment in the context of acute kidney injury (AKI) was successfully conducted in a prospective study of 85 adult patients undergoing cardiac surgery with cardiopulmonary bypass (CPB) use. One-dimensional (1D) ^{1}H NMR spectral data sets of filtered ethylenediaminetetraacetic acid (EDTA) plasma specimens collected 24 h after surgery were subjected to Random Forests based classification with *t*-test based feature filtering to prognosticate AKI. An average overall prognostication accuracy of 80 $\pm$ 0.9% with a corresponding area under the receiver-operating characteristic curve of 0.87 $\pm$ 0.01 could be obtained with, on average, 24 $\pm$ 2.8 spectral features. The set of discriminative ions and molecules included Mg^{2+}, lactate and the glucuronide conjugate of propofol, an anesthetic agent which had been administered to all patients during surgery. In AKI patients, increased levels of propofol-glucuronide seem to be a surrogate marker for reduced glomerular filtration, whereas an elevation of Mg^{2+} levels might be explained by its use for the treatment of cardiac arrythmias, and ischemic injury as well as systemic hypoperfusion present in this group might be linked to elevated lactate levels. Furthermore, this thesis presents a novel endogenous biomarker panel consisting of absolutely quantified EDTA plasma concentrations of Mg^{2+}, creatinine, and lactate, which would offer a reliable and swift diagnostic tool for the early detection of AKI after cardiac surgery with CPB use only requiring easily implementable point-of-care technologies. This biomarker panel was further employed to derive a novel Acute Kidney Injury Network (AKIN) index score, which illustrated that the metabolic profile of patients diagnosed with mildest renal injury was very similar to that of patients not developing AKI.

This study was further utilized to elucidate the importance of appropriate data normalization prior to statistical analysis, which proofed to be crucial for correct data interpretation.

The second part of this thesis presents first statistical data analysis results of 1D ^{1}H NMR spectra of EDTA plasma or urine specimens, respectively, from two large-scale clinical trials on chronic kidney disease (CKD). The German Chronic Kidney Disease (GCKD) study includes the currently world-wide largest cohort of patients suffering from CKD, which will be prospectively followed in the next ten years, and the Trial to Reduce Cardiovascular Events with Aranesp® Therapy (TREAT) study comprises a large, homogeneous cohort of patients suffering from CKD, type-2 diabetes mellitus, and concomitant anemia. Distinct differences

in metabolic fingerprints between various leading renal diseases, such as diabetic nephropathy and glomerulonephritis, in the GCKD study, or associated with adverse patient outcome in the TREAT study could be detected by t-tests in concordance with standard clinical pathologies of CKD. Additionally, the prediction of future kidney performance, which is crucial for improved patient care, with regression models based on either NMR derived EDTA plasma metabolic fingerprints or clinical parameters both assessed two years before was conducted within the GCKD study. Here, multiple regression models based on NMR fingerprints did not outperform simple regression models based on respective baseline clinical parameters. This probably reflects the fact that the renal function of most investigated CKD patients was fairly stable within these two years.

2 Zusammenfassung

Innerhalb der letzten Jahrzehnte nahm die globale Belastung aufgrund von Nierenerkrankungen im Menschen kontinuierlich zu. Um damit verbundene Sterblichkeits- und Morbiditätsraten zu verringern, sind frühe Diagnose sowie verbesserte Einsichten in zugrundeliegende biologische Mechanismen entscheidend. Diese Doktorarbeit präsentiert metabolische Untersuchungen von Körperflüssigkeiten mittels der Kernspinresonanzspektroskopie innerhalb der Nephrologie, um frühere Detektion sowie neue Erkenntnisse bezüglich klinischer Manifestation der Nierenerkrankung zu ermöglichen.
Der Nachweis neuer Komponenten mit niedrigem Molekulargewicht zur verbesserten Früherkennung und Patientenbehandlung im Kontext akuten Nierenversagens (AKI) wurde erfolgreich in einer prospektiven Studie mit 85 erwachsenen Patienten, die sich einer Herzoperationen mit Verwendung der Herz-Lungen-Maschine unterzogen hatten, durchgeführt. Eindimensionale (1D) ^{1}H Kernspinresonanzspektren gefilterter Ethylendiamintetraacetat (EDTA) Plasmaproben, die 24 Stunden nach der Operation abgenommen worden waren, wurden mittels Random Forests inklusive *t*-Test basierender Featureauswahl klassifiziert, um AKI zu prognostizieren. Bezogen auf die Gesamtkohorte, konnten mit Hilfe von, im Durchschnitt, 24 $\pm$ 2.8 spektraler Features, im Mittel 80 $\pm$ 0.9 % der Patienten richtig klassifiziert werden, was einer Fläche unter der Beobachterkennlinie von 0.87 $\pm$ 0.01 entspricht. Mg^{2+}, Laktat, und das Glucuronid-Konjugat des Propofols, das allen Patienten während der Operation als Anästhetikum verabreicht worden war, befanden sich unter den diskriminierenden Ionen und Molekülen. In AKI Patienten scheint ein erhöhter Propofol-Glukuronid-Spiegel ein surrogater Marker für reduzierte glomeruläre Filtration zu sein, wobei ein erhöhter Mg^{2+} Spiegel durch die Administration von Magnesium zur Behandlung von Herzrythmusstörungen erklärt werden könnte, und Ischämie sowie systemische Hypoperfusion in dieser Patientengruppe mit erhöhten Laktatspiegeln in Verbindung gebracht werden könnten. Außerdem präsentiert diese Doktorarbeit ein neues Set an endogenen Biomarkern bestehend aus absoluten EDTA Plasmakonzentrationen von Mg^{2+}, Kreatinin und Laktat, welches ein zuverlässiges und schnelles Diagnosewerkzeug zur AKI Früherkennung nach Herzoperationen mit Herz-Lungen-Maschine darstellen könnte. Des weiteren wurde dieses Biomarker-Set zur Ableitung eines neuen Acute Kidney Injury Network (AKIN) Scores benutzt, der die Tatsache illustrierte, dass Patienten mit geringster Nierenschädigung ein metabolisches Profil aufweisen, das sich nur gering vom metabolischen Profil von Patienten ohne AKI unterscheidet.
Zusätzlich wurde diese Studie dazu genutzt um die Bedeutung angemessener Datennormalisierung im Vorfeld von statistischen Analysen zu illustrieren, was sich als ausschlaggebend zur korrekten Dateninterpretation erwies.
Der zweite Teil dieser Doktorarbeit präsentiert erste statistische Datenauswertungen von 1D ^{1}H Kernspinresonanzspektren von EDTA Plasma- beziehungsweise Urinproben zweier großan-

gelegter klinischer Studien über chronisches Nierenversagen (CKD). Die German Chronic Kidney Disease (GCKD) Studie umfasst die derzeit weltweit größte Kohorte an Patienten mit CKD, die prospektiv über die nächsten zehn Jahre verfolgt wird, und die Trial to Reduce Cardiovascular Events with Aranesp® Therapy (TREAT) Studie schließt eine große, homogene Kohorte an Patienten mit CKD, Typ-2 Diabetes Mellitus, und begleitender Anämie ein. Ausgeprägte Unterschiede in metabolischen "Fingerprints" konnten mittels t-Tests zwischen verschiedenen führenden Nierenerkrankungen, z.B. diabetische Nephropathie und Glomerulonephritis, in der GCKD Studie, oder in Verbindung mit widrigem Krankheitsausgang in der TREAT Studie nachgewiesen werden. Diese unterschiedlichen metabolischen "Fingerprints" stimmen mit klinischen Standard-Pathogenesen chronischen Nierenversagens überein. Außerdem wurde im Rahmen der GCKD Studie die Vorhersage zukünftigen Nierenversagens, was ausschlaggebend für eine verbesserte Patientenversorgung ist, mit Regressionsmodellen entweder basierend auf metabolischen "Fingerprints" der Kernspinresonanzspektren der EDTA Plasmaproben oder basierend auf klinischen Parametern durchgeführt, wobei sowohl EDTA Plasmaproben als auch klinische Parameter zwei Jahre zuvor erhoben worden waren. Hierbei erzielten multiple Regressionsmodelle basierend auf Kernspinresonanz-"Fingerprints" keine besseren Ergebnisse im Vergleich zu einfachen Regressionsmodellen basierend auf entsprechenden klinischen Baseline-Parametern. Möglicherweise reflektiert dies die Tatsache, dass die Nierenfunktion der meisten untersuchten CKD Patienten innerhalb dieser zwei Jahre eher stabil war.

3 Introduction

3.1 Motivation: The global burden of kidney disease

The kidney is one of the vital organs in the human body due to its regulatory functions [Dörner 2013]. Maintenance of the homeostatic condition of the body is one of its main tasks [Treuting and Kowalewska 2012, Arasteh et al. 2009].
Consequently, the study and cure of kidney diseases, the primary aspect of nephrology, are of great importance for decreasing mortality and morbidity rates across the globe [O'Toole and Sedor 2014, Eckardt et al. 2013]. In the process, a large range of different illnesses is covered, mostly revealing an either acute ($\leq$3 months) or chronic (>3 months) deterioration of the kidney's performance [Eckardt et al. 2013, Kuhlmann et al. 2003].

Acute kidney injury (AKI), which is characterized by an abrupt (within one week) reduction in renal function [Eckardt et al. 2013, Mehta et al. 2007], is a subgroup of acute kidney diseases [Eckardt et al. 2013]. It comprises the whole spectrum of acute renal failure (ARF), caused by various factors including nephrotoxic drugs and complicated surgeries [Mehta et al. 2007]. The incidence of AKI constantly increased in the last decade due to, for example, augmented risk factors as well as improved diagnosis and documentation of the disease [Siew and Davenport 2015, Lameire et al. 2013].
AKI is a significant complication after cardiac surgery, leading to an increased risk of mortality and morbidity [Chawla et al. 2014, Eckardt et al. 2013, Mariscalco et al. 2011, Rosner and Okusa 2006]. Its occurrence in patients undergoing cardiac surgery approaches 30% with about 1 - 6% requiring dialysis [Mariscalco et al. 2011, Rosner and Okusa 2006]. For AKI patients requiring dialysis, the mortality rate amounts to 54% [Mariscalco et al. 2011]. Moreover, already a small decrease of renal function indicated by a reduction of the postoperative glomerular filtration rate (GFR) of equal or more than 30% is associated with a 5.9% mortality rate [Mariscalco et al. 2011]. A GFR decline of less than 30% is still associated with a mortality rate of 0.4% [Mariscalco et al. 2011]. Furthermore, a link between AKI and increased risk of long-term mortality has been reported [Hobson et al. 2009, Engoren et al. 2014]. Hence, AKI after cardiac surgery leads to increases in cost of care and length of stay in the intensive care unit (ICU) [Mariscalco et al. 2011]. The demand for early diagnosis of AKI is consequently eminent for improved patient care [Wyckoff and Augoustides 2012, Shaw 2012, Mariscalco et al. 2011].
The most common classification and staging scheme, the KDIGO (Kidney Disease: Improving Global Outcomes) criteria, make use of increases in serum creatinine (SCr) levels and decreases in GFR and urine output (UO) for diagnosis of AKI [KDIGO workgroup 2012]. Nev-

ertheless, SCr is not the ideal biomarker for AKI due to its relatively late alteration following surgery [Shaw 2012] and its modulation by nonrenal factors [Macedo and Mehta 2013, Wyckoff and Augoustides 2012, Star 1998]. Consequently, the search for alternative biomarkers is an important field in nephrology [Mariscalco et al. 2011, Parikh et al. 2011, Endre et al. 2011, Haase et al. 2010a), Haase et al. 2010b), Haase et al. 2009].
To date, none of the novel biomarkers reported for the diagnosis of AKI after cardiac surgery, e.g. neutrophil-gelatinase-associated lipocalin (NGAL) and serum cystatin C (CysC), has proven to be sufficiently predictive in heterogeneous patient cohorts with comorbidities [Endre et al. 2011, Lameire et al. 2011]. Moreover, in a clinically relevant setting, these biomarkers do not seem to clearly add any new information to the traditional approach [Lameire et al. 2011], still leaving the desire for improved diagnosis unsatisfied.

AKI is tightly connected to chronic kidney disease (CKD), with CKD being the most important risk factor for common AKI [Chawla et al. 2014, Lameire et al. 2013]. Moreover, even mild cases of AKI are associated with new-onset as well as progression to advanced stages of CKD [Chawla et al. 2014, Jha et al. 2013, Eckardt et al. 2013, Lameire et al. 2013].
CKD imposes an even larger burden on the world's health system [Jha et al. 2013], with global occurrence exceeding 10% [O'Toole and Sedor 2014, Eckardt et al. 2013] and 50% in high-risk subpopulations [Eckardt et al. 2013]. Its incidence is strongly linked to increasing age, with more than 20% of the population older than 60 years and more than 35% older than 70 years at the time of diagnosis of CKD [Eckardt et al. 2013]. In general, CKD is associated with a reduced GFR and increased albuminuria, irrespective of the cause of diminished renal function [Eckardt et al. 2013]. Along its progression, CKD leads to a large number of adverse clinical symptoms, finally ending in complete renal failure, called end-stage renal disease (ESRD) [Kuhlmann et al. 2003]. It is linked to elevated all-cause and cardiovascular mortality, AKI, cognitive decline, anemia (hemoglobin (Hb) deficiency), mineral and bone disorders, and fractures [Jha et al. 2013, KDIGO workgroup 2013, Eckardt et al. 2012, Kuhlmann et al. 2003]. Actually, CKD was ranked 18^{th} in the list of causes of total number of worldwide deaths in 2010 [Jha et al. 2013]. It is also a comorbidity of numerous chronical illnesses, e.g. cardiovascular disease, hypertension, obesity, and diabetes [Chawla et al. 2014, O'Toole and Sedor 2014, Eckardt et al. 2013], further worsening patient's prognosis [Eckardt et al. 2013]. In fact, type-2 diabetes mellitus is the leading cause of ESRD in developed countries [Jha et al. 2013, Kuhlmann et al. 2003]. Moreover, the presence of anemia in patients with type-2 diabetes mellitus and CKD further increases the rates of cardiovascular and renal events [Pfeffer et al. 2009a)].
With regard to the demanding adverse outcomes of CKD, the need for early detection is of prime interest [Eckardt et al. 2013, Jha et al. 2013]. In general, the underlying mechanisms as well as the pathophysiological and clinical consequences of CKD are still poorly understood [Eckardt et al. 2012]. Moreover, due to the overall heterogenity of CKD ethiology and pathomechanism, an urging demand for clinical studies in specific subpopulations is given [O'Toole and Sedor 2014, Eckardt et al. 2013].
The German Chronic Kidney Disease (GCKD) study was designed as a national prospective observational cohort study, involving study centers throughout Germany [Eckardt et al. 2012]. It comprises about 5000 CKD patients with a moderately reduced GFR and/or overt proteinuria

at enrollment, receiving comparable medical care [Eckardt et al. 2012]. The major goals are the characterization of burden and course of CKD patients, the identification and validation of novel risk factors and biomarkers for CKD manifestation, progression and complications as well as the achievement of an advanced understanding of the underlying pathophysiology [Eckardt et al. 2012]. Study participants are seen annually for up to ten years [Eckardt et al. 2012] and biomaterial, including urine, serum, and plasma specimens, is collected for every other year. The large size of the GCKD study cohort, which is well-characterized, and the long observation period facilitate the investigation of various hypotheses in a statistically meaningful manner.
In contrast to the GCKD study, the Trial to Reduce Cardiovascular Events with Aranesp® Therapy (TREAT) study was a randomized, multicenter, double-blind placebo-controlled, clinical trial [Pfeffer et al. 2009a), Pfeffer et al. 2009b)]. It comprised about 4000 patients with CKD, type-2 diabetes mellitus and anemia, which develops in most CKD patients as the diseased kidneys produce increasingly less erythropoietin (EPO) [Rao and Pereira 2003], and was designed to test whether the administration of darbepoetin alfa, an erythropoiesis-stimulating agent (ESA), would reduce the rates of death, cardiovascular events and ESRD [Pfeffer et al. 2009a), Pfeffer et al. 2009b)]. The large size and homogeneity of this cohort offer excellent opportunities to detect novel biomarkers associated with adverse outcomes as well as to gain novel insights into course of renal disease progression and complications in this specific patient group.

3.2 Objective: Metabolomics in the context of nephrology

Systems biology studies the behavior and development of a specific biological system under the influence of a particular perturbation [Ideker et al. 2001]. High-throughput and high-dimensional data sets are evaluated employing computational bioinformatic methods [Ideker et al. 2001]. Associated disciplines, the so-called 'omics'-sciences, comprise, among others, genomics, transcriptomics, proteomics, and metabolomics [Joyce and Palsson 2006].
The principal aim of metabolomics is the study of all small organic compounds, denoted as metabolites, present in a biological specimen [Tzoulaki et al. 2014, Kosmides et al. 2013, Nicholson and Lindon 2008, Nicholson 2006]. Their flow through bioenergetic and biosynthetic pathways is investigated in a quantitative manner [Tzoulaki et al. 2014]. Hence, the metabolome comprises the whole range of metabolites present or produced by a biological system, e.g. an organism at a defined time-point under a given set of conditions [Tzoulaki et al. 2014, Kosmides et al. 2013].
The application of metabolomics in the context of nephrology seems to be highly suitable. The kidneys' major functions comprise the excretion and also tubular secretion of metabolic waste products from the blood into the urine as well as reabsorption of essential nutritive substances [Treuting and Kowalewska 2012, Arasteh et al. 2009]. Consequently, metabolic investigations of urine, serum, and plasma specimens are predestined to facilitate new insights into pathomechanism and detection of novel biomarkers of renal diseases [Zhang et al. 2014, Weiss et al. 2011]. Moreover, changes in the metabolome due to an alteration of the renal function should be more significant and more detectable than elaborate changes in the renal proteome

or transcriptome [Wishart 2008].
Major analytical methods used in the field of metabolomics are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) [Tzoulaki et al. 2014]. Thereby, both methods facilitate the simultaneous detection as well as absolute quantification of a large range of metabolites in a specimen [Nicholson and Lindon 2008]. Hence, in comparison to the traditional, targeted approach, usually conducted in clinical studies, an untargeted approach is facilitated by measuring a so-called 'metabolic fingerprint' of the investigated specimen [Tzoulaki et al. 2014]. However, the complexity of the metabolism itself and the investigated 'metabolic fingerprint' of the sample require sophisticated bioinformatic strategies for data interpretation [Nicholson and Lindon 2008].
NMR spectroscopy, in comparison to MS spectrometry, is non-destructive, highly repeatable and requires minimum sample preparation [Tzoulaki et al. 2014, Nicholson and Lindon 2008], hence being especially well suited for the analysis of large biomaterial collections comprising several hundred to thousands of specimens. On the downside, it offers lower sensitivity as well as lower spectral resolution in comparison to MS [Tzoulaki et al. 2014, Weiss et al. 2011, Nicholson and Lindon 2008].
Biomedical studies in the field of metabolomics usually investigate biofluids, e.g. urine and blood, which are easily obtained [Tzoulaki et al. 2014]. Their clinical objectives are diverse, including detection of novel diagnostic biomarkers and determination of distinct metabolic profiles for specific clinical conditions ('metabolic fingerprinting') [Kosmides et al. 2013, Dettmer and Hammock 2004]. Consequently, they can facilitate improved or individualized patient treatment, the goal of personalized medicine [Weiss et al. 2011]. Several metabolic studies already proved the capability of NMR spectroscopy for detection of novel disease biomarkers in such diverse areas as, e.g. autosomal dominant polycystic kidney disease (ADPKD) [Gronwald et al. 2011], diary cow metabolism [Klein et al. 2012, Bertram et al. 2011, Klein et al. 2010], as well as various metabolic and renal diseases [Elliott et al. 2015, Dawiskiba et al. 2014, Deja et al. 2013, Neild et al. 1997, Holmes et al. 1997].

With regard to the diverse research questions, still pending in the field of nephrology, as depicted in section 3.1, I have formulated three specific aims concerning metabolic investigations of renal diseases by means of NMR spectroscopy for this thesis. This analytical method was selected due to its especially high suitability for the comprehensive analysis of large specimen collections as it is required here.
My first aim is the detection of metabolic biomarkers in the context of various renal diseases as alternatives to traditional clinical approaches. My second aim comprises the prediction of future kidney performance based on baseline metabolic fingerprints derived by NMR spectroscopy. General method developments and additions for NMR based metabolomics with regard to appropriate data normalization, absolute low-molecular-weight compound quantification, and NMR measurements of unfiltered plasma specimens are the third aim of my Ph.D. thesis.

The first aim of this Ph.D. thesis is the detection of metabolic biomarkers for both acute and chronic kidney diseases as alternatives to traditional clinical approaches.

In the setting of AKI diagnosis after cardiac surgery, the search and evaluation of earlier or alternative biomarkers is pursued by NMR spectroscopic fingerprinting of urine and plasma specimens collected from a heterogeneous group of AKI and non-AKI patients. Here, the detection of AKI after cardiac surgery by means of urinary biomarkers collected before, at 4 and 24 h after surgery as well as by means of plasma biomarkers collected 24 h after surgery is pursued. Furthermore, the performance of different plasma biomarker sets is evaluated and compared to each other as well as to traditional diagnostic tools. These novel metabolic markers could offer the chance of an earlier AKI detection after cardiac surgery than by monitoring changes in SCr levels or UO.
Furthermore, several thousand different plasma specimens collected at the baseline time-point of the GCKD study are measured by means of NMR spectroscopy and specific metabolic fingerprints for the discrimination of different leading renal diseases are searched and evaluated. These investigations can provide new insights into the metabolic characteristics of specific renal diseases such as vascular or glomerulonephritis.
For the generation of new insights into CKD progression and corresponding Hb responsiveness, novel biomarkers for several different clinical outcomes are scanned and appraised. About 1100 different urine specimens collected from TREAT study participants are measured by NMR spectroscopy and the following hypotheses are tested: (1) no difference exists between patients dying from any cause, and patients not dying, under the restriction that all patients within both subcohorts do not progress to ESRD, (2) no difference exists between patients progressing and not progressing to ESRD under the restriction that all patients within both subcohorts do not die and (3) no difference exists between patients with various stages of Hb responsiveness at two different time-points, respectively, whereas four different subcohorts treated with darbepoetin alfa with various stages of Hb responsiveness and one subcohort treated with a placebo compound are investigated.

The second aim, the prediction of future kidney performance based on baseline metabolic fingerprints derived by NMR spectroscopy was conducted in order to give further insights into kidney disease development and progression.
The correlation of NMR metabolic fingerprints derived from the baseline plasma specimen cohort of the GCKD study with the estimated GFR and specific renal performance markers, such as SCr and serum CysC, as well as their associated predictive performance are investigated. Here, multiple regression analyses between baseline NMR metabolic fingerprints and these clinical parameters, determined at the baseline as well as at the second follow-up time-point two years after inclusion into the study, were performed. Moreover, simple linear regression with baseline clinical parameters, i.e. SCr, serum CysC, as well as eGFR, with respect to second follow-up clinical parameters, i.e. SCr, serum CysC, as well as eGFR, respectively, were conducted. The prediction of present and future kidney performance based on these measures will be of great importance for timely interventions and improved patient care in this patient cohort.

My third aim covers method developments and improvements for NMR based metabo- lomics. Proper NMR data normalization is crucial for correct interpretation of metabolomic investigations. However, common normalization techniques, such as Quantile or Variance Stabilization

normalization, can lead to erroneous results if different investigated cohorts do not exhibit approximately equal shares of up- and down-regulated features. This thesis reports a prominent case of inappropriate NMR data normalization and provides, as well as evaluates alternative normalization methods.
The absolute quantification of the metal ions calcium and magnesium in plasma specimens by NMR spectroscopy is an addition to NMR based metabolomic methods. Here, their absolute quantification via ethylenediaminetetraacetic acid (EDTA) complexes is implemented and validated for the AKI plasma data set.
Moreover, the acquisition of 1D ^{1}H NMR spectra of unfiltered EDTA plasma specimens implies several challenges with regard to the traditional NMR reference substance 3-trimethylsilyl-2,2,3,3-tetradeuteropropionate (TSP). These challenges are reported and easily implementable solutions are presented in the context of the GCKD study.

This Ph.D. thesis has in parts already been published in [Zacharias 2012, Zacharias et al. 2013a), Zacharias et al. 2013b), Zacharias et al. 2015, Hochrein et al. 2015] and was funded by the Bavarian Genome Network (BayGene), the German Federal Ministry of Education and Research (BMBF Grant no. 01 ER 0821), the German Research Foundation (KFO 262), and the intramural funding program of the Regensburg School of Medicine.

4 Background

4.1 Introduction to nephrology

4.1.1 Renal structure and physiology

The human kidneys are the most important organs for maintaining the homeostatic balance in the body [Eckardt et al. 2013, Dörner 2013]. They are bean-shaped organs, located in the retroperitoneum [Arasteh et al. 2009].
The kidney's principal anatomy is depicted in Figure 4.1a). It is surrounded by a thin fibrous capsule and the renal parenchyma can be subdivided into the cortex and the medulla [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The medulla itself consists of so-called pyramids, which form a broad base towards the cortex and a cone end, denoted as papilla, extending into the sinus [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The spaces between the pyramids are filled by cortex tissue and are termed renal columns (of Bertin) [Treuting and Kowalewska 2012, Arasteh et al. 2009]. Each of the papillae ends in a minor calyx, whereas two to three of the minor calyces unite to form one major calyx [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The major calices themselves are discharged into the renal pelvis, located at the renal sinus [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The formed urine is conducted from the renal pelvis via the ureter to the bladder [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The kidney's blood supply is managed by the renal artery, the filtered blood is removed by the renal vein [Treuting and Kowalewska 2012, Arasteh et al. 2009]. The structure of the basic functional unit of the kidney, the nephron [Eckardt et al. 2013, Treuting and Kowalewska 2012, Arasteh et al. 2009], is given in Figure 4.1b). A single kidney comprises about one million nephrons. Every nephron can be further subdivided into a filtrating body, the glomerulus, and different tubule segments [Eckardt et al. 2013, Treuting and Kowalewska 2012, Arasteh et al. 2009], compare to Fig. 4.1b). The glomeruli are only located in the renal cortex, whereas the tubuli can additionally be found in the medulla [Treuting and Kowalewska 2012, Arasteh et al. 2009]. In the glomerulus, which is imbedded into the Bowman's capsule, incoming blood transported by the afferent arteriole, a branch from the arcuate artery, is ultrafiltrated to form the so-called primary urine or ultrafiltrate [Eckardt et al. 2013, Arasteh et al. 2009]. About 180-200 l of ultrafiltrate are generated daily, usually containing sodium, potassium, chloride, phosphate, water, glucose, amino acids, urea and proteins with a mass below 60-70 kDa [Eckardt et al. 2013, Arasteh et al. 2009]. The ultrafiltrate is released from the Bowman's capsule into the proximal tubule, whereas the filtered blood is discharged by the efferent arteriole [Eckardt et al. 2013, Arasteh et al. 2009]. During the ultrafiltrate's propagation along the proximal tubule, about 80% of the fluid is being reabsorbed into the peritubular capillaries, including about two-thirds of filtered water and salt, 100% of filtered glucose and

amino acids, as well as proteins [Eckardt et al. 2013, Arasteh et al. 2009]. At the end of the proximal tubule, organic molecules and drug metabolites are secreted into the filtrate [Eckardt et al. 2013]. The loop of Henle, subdivided into the descending and ascending limb, is responsible for concentrating the filtrate [Eckardt et al. 2013]. The associated mechanisms, namely reabsorption of water and removal of sodium from the filtrate, take place in the descending and ascending limb, respectively [Eckardt et al. 2013, Arasteh et al. 2009]. The tubular NaCl concentration is monitored at the junction between the ascending limb of the loop of Henle and the distal nephron, the macula densa [Eckardt et al. 2013, Arasteh et al. 2009]. Thereby, the glomerular blood flow is tightly autoregulated by the so called tubuloglomerular feedback mechanism [Eckardt et al. 2013, Arasteh et al. 2009]. The distal nephron, comprising the distal tubule and the collecting duct at the renal papilla, reabsorbs approximately 5% of the total amount of filtered sodium and responds to the hormons aldosterone and vasopressin, hence controlling the composition and concentration of the final urine [Eckardt et al. 2013, Treuting and Kowalewska 2012, Arasteh et al. 2009].
The main functions of the human kidney can be summarized as follows. The secretion of metabolic waste products from the blood into the urine with the possibility of an effective reabsorption mechanism enables the elimination of potentially toxic products without loosing vital nutrients [Treuting and Kowalewska 2012, Arasteh et al. 2009]. Its central role in homeostasis is furthermore reflected by the regulation of the acid-base and osmolality balance as well as the blood pressure [Treuting and Kowalewska 2012, Arasteh et al. 2009]. Finally, the human kidney produces important hormones like renin, calcitriol, the activated form of vitamin D, and EPO, which stimulates the production of red blood cells [Treuting and Kowalewska 2012, Arasteh et al. 2009].

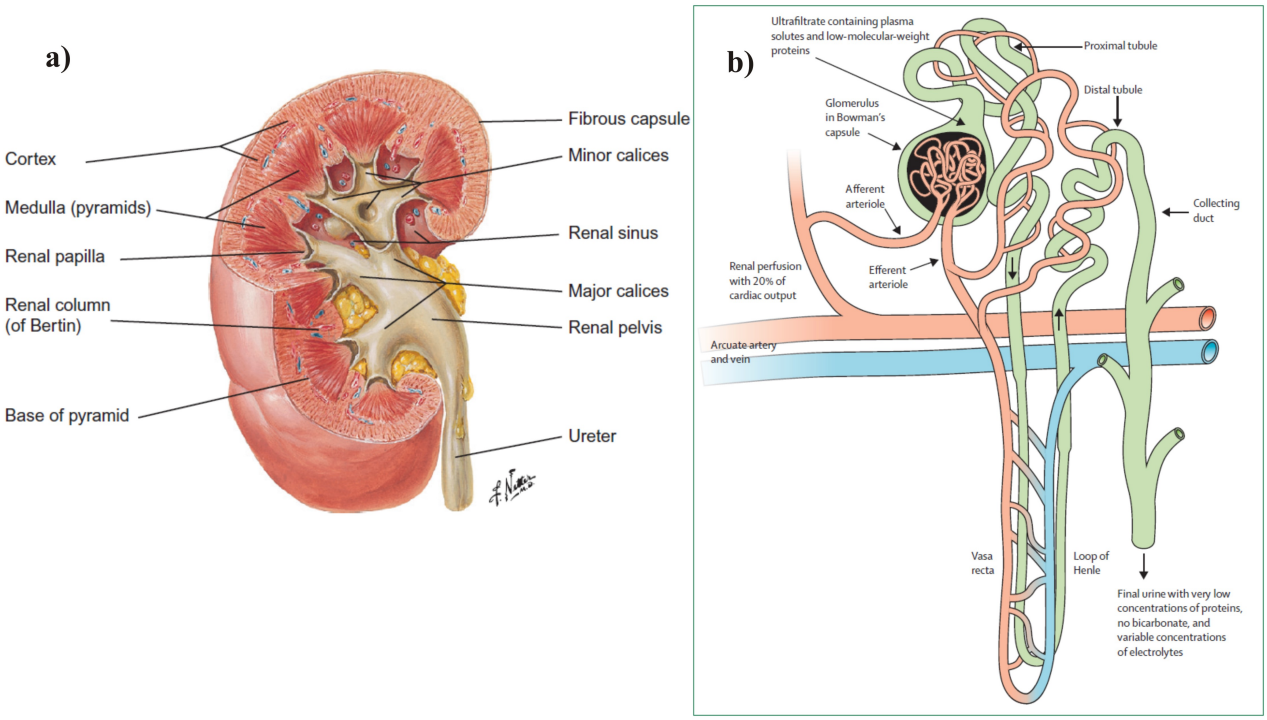


Figure 4.1: **a) Principal anatomy of the human right kidney. b) Structure of a nephron, the basic functional unit of the kidney.** a) The human kidney is surrounded by a fibrous capsule and can be further divided into the cortex and the medulla, comprising the pyramids. Their tips, the papillae, point towards the minor calices. The minor calices join in order to form the major calices, which end into the renal pelvis at the renal sinus. The renal pelvis is connected to the ureter. The cortex fills up the space between the pyramids, which is called the renal column (of Bertin) [Treuting and Kowalewska 2012, Arasteh et al. 2009]. Reprinted with permission from [Treuting and Kowalewska 2012]. b) Blood plasma, supplied by the afferent arteriole, a branch of the arcuate artery, is filtered by the glomerulus in the Bowman's capsule. The ultrafiltrate is modified along its propagation through the proximal tubule, the loop of Henle and the distal tubule until it is secreted into the collecting duct, where a final adjustment of the urine takes place [Eckardt et al. 2013, Arasteh et al. 2009]. Reprinted with permission from [Eckardt et al. 2013].

4.1.2 Clinical diagnostic tools for assessment of renal performance

With the human kidney executing essential tasks for maintaining body's homeostasis, compare to section 4.1.1, a deterioration of its performance can have hazardous consequences, as depicted in section 3.1. Therefore, monitoring renal function is vital for detection and supervision of all renal complications. As the filtration of blood and the formation of urine is strongly regulated by the kidney (compare to section 4.1.1), the study of their composition is predestined to offer comprehensive insights into renal performance.
In clinical practice, the most commonly used diagnostic methods for the detection of kidney malfunctions are based on alterations of the following parameters.
The glomerular filtration rate (GFR) is considered to be the best indicative reference for renal performance [KDIGO workgroup 2013, Stevens and Levey 2009, Stevens et al. 2006]. However, its exact determination by measuring the urinary or plasma clearance of exogenous filtration markers, e.g. inulin or EDTA, is considered to be too cumbersome and expensive for routine application [KDIGO workgroup 2013, Macedo and Mehta 2013, Stevens and Levey 2009, Stevens et al. 2006]. Therefore, the GFR is usually estimated from levels of endogenous filtration markers such as SCr, and consequently denoted as estimated GFR (eGFR) [KDIGO workgroup 2013, Cravedi and Remuzzi 2013, Jha et al. 2013, Stevens and Levey 2009]. An ideal filtration marker would be a substance that is freely filtered at the glomeruli, neither reabsorbed, secreted, synthesized, or metabolized by the tubuli, and that does not change renal function [Stevens and Levey 2009]. Various equations can be employed for the determination of eGFR, e.g. the Modification of Diet in Renal Disease (MDRD) Study or the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [KDIGO workgroup 2013, Cravedi and Remuzzi 2013, Jha et al. 2013, Stevens and Levey 2009]. However, these equations still perform poorly in comparison to direct measurement of the GFR in various critically ill patient cohorts [Bragadottir et al. 2013, Cravedi and Remuzzi 2013].
In general, SCr is well suited as an endogenous filtration marker, as it is almost completely removed from the blood by glomerular filtration and proximal tubular secretion, compare to section 4.1.1, and only marginally reabsorbed, e.g. in healthy newborns or elderly people [Dörner 2013, Arasteh et al. 2009, Musso et al. 2009]. SCr levels, however, are significantly influenced by age, sex, race, muscle mass, chronic illnesses, diet, and medications of the monitored patient [Stevens and Levey 2009, Curhan 2005]. Consequently, in order to ascertain, for example, an acute impairment of the renal function, only the alteration of the SCr level in comparison to the individual baseline level is meaningful [Stevens et al. 2006]. In healthy subjects, SCr values range between 74-110 µmol/l (0.84-1.25mg/dl) in white men and 58-96 µmol/l (0.66-1.09mg/dl) in white women [Dörner 2013]. The eGFR determined by estimating equations additionally takes age, sex, and certain races into account [KDIGO workgroup 2013, Stevens and Levey 2009, Arasteh et al. 2009, Stevens et al. 2006] and is typically adjusted for body surface area [Stevens and Levey 2009, Stevens et al. 2006], with eGFR values in young healthy whites of about 130 ml/min per 1.73m^2 for men and 120 ml/min per 1.73m^2 for women [Stevens and Levey 2009, Stevens et al. 2006].
However, the detection of renal impairment based on alterations of SCr levels and/or $eGFR_{SCr}$ can be erroneous because of several major drawbacks. First, a rise in SCr levels due to impaired

glomerular filtration does not take place until about 50% of the kidney function is already lost, hampering early detection of impaired glomerular filtration [Dörner 2013, Macedo and Mehta 2013]. Second, under non-steady state conditions, as often presented in critically ill patients, alterations in SCr levels and therefore in the $eGFR_{SCr}$ might reflect the magnitude and direction of the change in GFR, but do not accurately reflect its exact level [Macedo and Mehta 2013, Stevens and Levey 2009]. Additionally, in patients with a chronic impairment of the renal function, the kidney might adapt to the loss of nephrons and no change in the SCr levels or $eGFR_{SCr}$ is detectable, hence obscuring disease progression [Macedo and Mehta 2013]. In healthy subjects, about 10-20% of overall excreted creatinine is secreted by the proximal tubuli, but in patients with impaired glomerular filtration, up to 50% can be eliminated by proximal tubular secretion [Macedo and Mehta 2013, Curhan 2005]. This can lead to an overestimation of the eGFR based on SCr levels, which needs to be overcome by drug administration [Macedo and Mehta 2013]. Moreover, clinical measurements of SCr levels are not absolutely precise and therefore require a change in creatinine of at least 10% to exhibit significant results [Macedo and Mehta 2013]. The accurate determination of SCr levels in critically ill patients is further hampered by positive fluid balance, which leads to diluted SCr concentrations [Macedo and Mehta 2013].

An alternative to determination of eGFR by assessing SCr clearance is the measurement of serum cystatin C (SCysC) levels [Macedo and Mehta 2013, Stevens et al. 2006, Curhan 2005]. CysC is supposed to be constantly produced by all nucleated cells, it is filtered in the glomeruli, and taken up as well as degraded by the proximal tubular cells [Curhan 2005]. CysC is less affected by age, sex, muscle mass, and race of the monitored patient in comparison to SCr and faster mirrors changes in GFR [Macedo and Mehta 2013]. However, it is still significantly influenced by non-renal factors like smoking status, glucocorticoids use, and C-reactive protein (CRP) [Macedo and Mehta 2013, Stevens et al. 2006], it can be altered in specific health states such as diabetes, cancer, obesity, liver disease, and thyroid status [Macedo and Mehta 2013, Curhan 2005], and its concentration can also be affected by positive fluid balance [Macedo and Mehta 2013]. Nevertheless, CysC is also considered to be a biomarker for inflammation and a predictor of cardiovascular events and death independent of kidney function [Curhan 2005]. In healthy adults, SCysC values range between 0.54-0.94 mg/l in men and 0.48-0.82 mg/l in women [Dörner 2013].

A deterioration of the renal function is, in general, often accompanied with diminished urinary output (UO) [Dörner 2013]. Therefore, monitoring the state of UO can illustrate the renal performance in a sensitive and non-invasive way [Dörner 2013, KDIGO workgroup 2012, Macedo et al. 2011]. Daily UO for healthy adults ranges between 800-1800 ml for men and 600-1600 ml for women [Dörner 2013]. Oliguria and anuria in adults are defined as a daily UO below 400-500 ml or 100 ml, respectively [Dörner 2013, Arasteh et al. 2009]. However, the determination of UO over a fixed period of time, mostly ranging between 6-24 hrs, can be challenging and prone to errors [Dörner 2013, Macedo et al. 2011]. Moreover, the urine flow can be affected by non-renal factors, e.g. fluid intake and drug administration [Macedo et al. 2011], and oliguria or anuria can also be induced by urinary tract obstruction and total arterial or venous occlusion, diminishing their specificity for detection of renal damage [KDIGO workgroup 2012]. Furthermore, UO is usually normalized to body weight and the non-consistent use of body weight might lead

to an underestimation of UO in obese patients [KDIGO workgroup 2012].
In the course of clinical urinalysis in nephrology, several additional important parameters are assessed, e.g. the urinary protein content [Dörner 2013, Arasteh et al. 2009]. Healthy individuals usually excrete less than 150mg of protein per day into the urine [Dörner 2013, Arasteh et al. 2009], with 10-15% thereof being represented by albumin, a large-molecular-weight protein of about 67kDa [Arasteh et al. 2009]. Proteinuria is defined as a protein excretion of more than 150mg per day [Dörner 2013, Arasteh et al. 2009]. An increase of urinary protein content may be caused by elevated permeability of the glomeruli for large-molecular-weight proteins (so-called albuminuria or glomerular proteinuria), insufficient reabsorption of low-molecular-weight proteins in the tubuli (so-called tubular proteinuria) or higher concentration of low-molecular-weight proteins in the filtered plasma (so-called overproduction proteinuria) [KDIGO workgroup 2013]. With albumin representing the major part of urinary protein in most renal diseases, the assessment of urinary albumin content mainly substitutes the diagnosis of proteinuria in clinical practice [KDIGO workgroup 2013]. Drawbacks of albuminuria as a biomarker for detection and progression of impaired renal function include lack of standardized laboratory assays [Jha et al. 2013] and unreleability due to clinical treatment of albuminuria as, for example, included in clinical interventions to improve CKD outcome [Fassett et al. 2011].
The detection of increased excretion of low-molecular-weight proteins into the urine can be utilized to specify tubular dysfunction [Del Palacio et al. 2012]. These "tubular proteins" include neutrophil gelatinase-associated lipocalin (NGAL), β-2-microglobulin, retinol-binding protein, urinary CysC, and N-acetyl-β-D-glucosamini- dase (NAG) [Del Palacio et al. 2012]. Other biomarkers for tubular injury comprise glutathione S-transferases, liver-type fatty acid binding protein, kidney injury molecule-1 (KIM-1), and interleucin-18 (IL-18) [Del Palacio et al. 2012]. Kidney biopsies are one of the most specific diagnostic tools for renal malfunctions, however exhibit various drawbacks associated with the required invasive procedure [Arasteh et al. 2009, Kuhlmann et al. 2003].
In addition to the aforementioned clinical tools to monitor renal performance, numerous other parameters are typically assessed in nephrology, e.g. blood pressure, blood and urinary glucose content, urinary pH, urinary leukocyte content, urinary urea content, blood electrolyte content, etc. [Dörner 2013, Arasteh et al. 2009, Kuhlmann et al. 2003], and imaging techniques like sonography, computer tomography, or magnetic resonance imaging are applied in clinical practice [Arasteh et al. 2009, Kuhlmann et al. 2003].

4.1.3 Basic concepts of acute kidney injury after cardiac surgery

Acute kidney injury (AKI) is basically described as an abrupt decrease of the renal performance [Eckardt et al. 2013, KDIGO workgroup 2012, Mehta et al. 2007]. The most commonly used diagnostic and staging systems for AKI are the Risk Injury Failure Loss End-Stage Renal Disease (RIFLE) [Bellomo et al. 2004], the Acute Kidney Injury Network (AKIN) [Mehta et al. 2007], and the Kidney Disease: Improving Global Outcomes (KDIGO) [KDIGO workgroup 2012] criteria [Ostermann 2014], which are summarized in Table 4.1. They are all based on alterations in SCr levels and UO over certain periods of time, whereas the RIFLE criteria also consider changes in the GFR [Ostermann 2014, KDIGO workgroup 2012, Mehta et al. 2007, Bel-

lomo et al. 2004]. These are common clinical parameters assessed in nephrology, as elaborately discussed in section 4.1.2. RIFLE and AKIN criteria exhibit certain differences, compare to Table 4.1, whereas the KDIGO criteria merge both into a uniform staging system for AKI [Ostermann 2014]. One has to notice that although initiation of renal replacement therapy (RRT) is explicitly excluded in the AKIN criteria as a staging criterion for AKI [Mehta et al. 2007], patients treated with RRT are commonly classified as AKIN-stage 3 irrespective of their AKI stage at RRT initiation [Ostermann 2014].

AKI after cardiac surgery mostly arises due to different, interconnected, pathophysiological mechanisms, and main causes are patient-related factors and the use of cardiopulmonary bypass (CPB) before, during, and after the surgery [Mariscalco et al. 2011].

CPB use leads to unavoidable changes in blood flow by ischemia-reperfusion injury, low cardiac output, renal vasoconstriction, hemodilution and loss of the pulsatile blood flow [Mariscalco et al. 2011]. This leads to an imbalanced oxygen supply/demand of the kidney, whose blood circulation is usually tightly regulated, compare to section 4.1.1, resulting into significant cellular injury [Mariscalco et al. 2011]. Moreover, AKI after cardiac surgery also seems to result from hypothermia, systemic inflammatory response, cell lysis, and embolization caused by CPB use [Mariscalco et al. 2011].

Patient-related risk factors include type of surgery, sex, age, genetic AKI susceptibility, congestive heart failure, anemia, diabetes mellitus, chronic obstructive pulmonary disease, emergency status, nephrotoxic drugs and contrast agents, blood transfusions, post-surgical low cardiac output, use of post-operative intraaortic balloon pump, occurrence of sepsis after the surgery, and baseline renal performance [Mariscalco et al. 2011], with CKD, in general, being strongly associated with AKI incidence [Chawla et al. 2014, Lameire et al. 2013].

The AKI pathology is usually divided into several clinical phases [Mariscalco et al. 2011]. The early phase is determined by a vasomotor nephropathy with alterations in vasoreactivity and renal perfusion [Mariscalco et al. 2011]. Consequently, AKI is initiated and the early AKI phase is characterized by prerenal azotemia, cellular adenosine triphosphate depletion and oxidative injury [Mariscalco et al. 2011]. The extension of these symptoms in the next clinical phase leads to the activation of bone-marrow derived and endothelial cells with a subsequent proinflammatory state [Mariscalco et al. 2011]. These inflammatory cells adhere to the activated endothelium in the peritubular capillaries of the outer medulla with medullary congestion and hypoxic injury of the proximal tubule [Mariscalco et al. 2011]. The final clinical phase is characterized by a proliferation to the tubule cells and the renal function is reconstructed after their redifferentiation and repolarization [Mariscalco et al. 2011].

A typical symptom of post-operative AKI is acute tubular necrosis including urinary granular casts [Mariscalco et al. 2011].

Due to the significantly negative outcomes of AKI after cardiac surgery, as discussed in section 3.1, its prevention and/or early treatment is crucial to improve patient outcome [KDIGO workgroup 2012, Wyckoff and Augoustides 2012, Shaw 2012, Mariscalco et al. 2011]. With the use of CPB being the most important cause of AKI, its employment should be adapted in order to prevent AKI [Mariscalco et al. 2011]. In this context, pulsatile CPB proved to be superior to standard linear CPB and in general, CPB flow rates of 1.8 - 2.2 $\mathrm{l{\cdot}min^{-1}{\cdot}m^{-2}}$ (only refers to cerebral flow) with a mean arterial pressure $>$ 50 - 60 mmHg are recommended [Mariscalco et

al. 2011]. Furthermore, cardiac surgeries should be delayed beyond 24hrs of the patient's exposure to nephrotoxic contrast agents and their use should be limited [Mariscalco et al. 2011]. Moreover, drugs, which increase the renal blood flow, e.g. fenoldopam, show renal protective effects and could therefore prevent AKI [Mariscalco et al. 2011].
The detection of AKI based on SCr levels, UO, and GFR, usually takes place within 48hrs after the surgery when applying the AKIN and KDIGO criteria or within seven days after the surgery when applying the RIFLE criteria [Ostermann 2014, KDIGO workgroup 2012, Cruz et al. 2009, Mehta et al. 2007, Bellomo et al. 2004].

	SCr alterations			UO alterations		
	RIFLE	**AKIN**	**KDIGO**	**RIFLE**	**AKIN**	**KDIGO**
AKI definition		absolute increase in SCr of either ≥0.3mg/dl (≥26.4µmol/l) or percentage increase of ≥50% (1.5-fold from baseline) in 48hrs	absolute increase in SCr of either ≥0.3mg/dl (≥26.5µmol/l) within 48hrs or increase to ≥1.5 × baseline, which is known or presumed to have occurred within the prior 7ds		reduction in UO (documented oliguria of <0.5ml/kg/h for >6hrs	urine volume <0.5ml/kg/h for 6hrs
Stage 1 or Risk	increased SCr × 1.5 or GFR decrease > 25%	increased SCr of ≥0.3mg/dl (≥26.4µmol/l) or increase to ≥1.5 to 2 -fold from baseline	increased SCr of ≥0.3mg/dl (≥26.5µmol/l) or increase to 1.5 - 1.9 × baseline	UO <0.5ml/kg/h × 6hrs	UO <0.5ml/kg/h for >6hrs	UO <0.5ml/kg/h for 6-12hrs
Stage 2 or Injury	increased SCr × 2 or GFR decrease >50%	increased SCr × 2 to 3 -fold from baseline	increased SCr × 2.0 to 2.9 from baseline	UO <0.5ml/kg/h × 12hrs	UO <0.5ml/kg/h for >12hrs	UO <0.5ml/kg/h for ≥12hrs
Stage 3 or Failure	increased SCr × 3 or GFR decrease 75% or SCr ≥4mg/dl (acute rise ≥0.5mg/dl)	increased SCr >3-fold from baseline (or SCr ≥4.0mg/dl (≥354µmol/l) with acute rise of at least 0.5mg/dl (44µmol/l)	increased SCr × 3.0 from baseline or increase in SCr to ≥4.0mg/dl (≥353.6µmol/l) or initiation of RRT or in patients <18yrs decrease in eGFR to <35ml/min per 1.73m^2	UO < 0.3ml/kg/h × 24hrs or anuria × 12hrs → oliguria	UO <0.3ml/kg/h for 24hrs or anuria for 12hrs	UO <0.3ml/kg/h for ≥24hrs or anuria for ≥12hrs
Loss	persistent ARF = complete loss of kidney function >4 weeks					
ESKD	End Stage Kidney Disease (>3 months)					

Table 4.1: **AKI definition and staging systems according to RIFLE, AKIN, and KDIGO criteria.** Abbreviations: ARF, acute renal failure; GFR, glomerular filtration rate; SCr, serum creatinine; RRT, renal replacement therapy; UO, urine output. Modified from [Ostermann 2014, KDIGO workgroup 2012, Mehta et al. 2007, Bellomo et al. 2004].

4.1.4 Basic concepts of chronic kidney disease

Chronic kidney disease (CKD) in adults is broadly defined as abnormalities of the renal structure or function, which are present for more than three months and exhibit certain implications for health [KDIGO workgroup 2013]. These implications or CKD criteria are a decreased GFR <60ml/min per 1.73m^2, presence of proteinuria/albuminuria (albumin excretion rate (AER) ≥30mg for 24hrs; urinary albumin-creatinine ratio (ACR) ≥30mg/g [≥3mg/mmol]), urine sediment abnormalities, electrolyte and other abnormalities due to tubular disorders or detected by histology and imaging [KDIGO workgroup 2013]. One or more of these criteria should be present for more than three months in order to confirm CKD [KDIGO workgroup 2013]. Patients with a history of kidney transplantation are automatically defined as CKD patients, irrespective of their GFR level or presence of markers for renal damage [KDIGO workgroup 2013].
CKD patients with reduced GFR in general exhibit three significant types of complications, i.e. increased drug toxicity, metabolic and endocrine complications, e.g. anemia, acidosis, malnutrition, bone and mineral disorders, and increased risk of CVD and death [Jha et al. 2013,KDIGO workgroup 2013, Eckardt et al. 2012, Kuhlmann et al. 2003]. Other complications include infections, cognitive impairment, frailty, etc. [Jha et al. 2013, KDIGO workgroup 2013, Eckardt et al. 2012,Kuhlmann et al. 2003]. CKD renal outcomes can be generally summarized as GFR decline, albuminuria rise, AKI, and chronic kidney failure (equal to ESRD) defined as a GFR <15ml/min per 1.75m^2 [KDIGO workgroup 2013].
The classification system of CKD is called CGA (Cause, GFR, and Albuminuria) staging as it is based on cause as well as GFR and albuminuria levels [KDIGO workgroup 2013]. With CKD being broadly defined and not being a diagnosis per se, the assignment of its cause is crucial for prognostication and patient treatment [KDIGO workgroup 2013]. This assignment is based on the presence or absence of underlying systemic disease, and the location of observed or presumed pathologic-anatomic findings within the kidney [KDIGO workgroup 2013]. Consequently, CKD can be, in general, categorized into glomerular, tubulointerstitial, vascular, or cystic and congenital diseases [KDIGO workgroup 2013]. The GFR and albuminuria categories for CKD, are summarized in Table 4.2. To predict the prognosis of CKD in general, its cause and other risk factors and comorbid conditions are considered [KDIGO workgroup 2013]. However, these risk categories still differ for individual outcomes, highlighting the importance of supplementary research [KDIGO workgroup 2013, Eckardt et al. 2012].
Consequently, GFR and albuminuria should be assessed in CKD patients at least annually in order to monitor disease progression [KDIGO workgroup 2013]. Factors associated with CKD progression include, among others, cause of CKD, level of GFR and albuminuria, age, sex, race/ethnicity, elevated blood pressure, hyperglycemia, dyslipidemia, smoking, obesity, history of CVD, and ongoing exposure to nephrotoxic agents [KDIGO workgroup 2013].
Due to the complexity of CKD causes and comorbidities, treatment strategies are diverse and need to be tailored to the individual patient [Eckardt et al. 2013]. General treatment strategies to slow progression of CKD and reduce risk of mortality associated with CVD comprise blood pressure and glycemic control, especially in CKD patients with diabetes, as well as change of diet and/or lifestyle [KDIGO workgroup 2013, Jha et al. 2013]. Furthermore, with CKD

patients exhibiting an increased risk of developing AKI, appropriate precautions should be considered to prevent AKI incidence [KDIGO workgroup 2013], as, for example, outlined in section 4.1.3 in the context of cardiac surgery. Even with appropriate patient treatment to slow down CKD progression, CKD patients classified in stage G5 according to the GFR criteria (compare to Table 4.2) have established chronic kidney failure or ESRD, and the only option to avoid hazardous consequences is timed application of RRT [Kuhlmann et al. 2003]. RRT include hemodialysis and kidney transplantation [Kuhlmann et al. 2003], with favorable long-term results for transplant recipients [Purnell et al. 2013, Tonelli et al. 2011, Weitz et al. 2006].
Additionally, elaborate treatment options for CKD complications, e.g. hypertension and anemia, exist [KDIGO workgroup 2013]. Anemia in CKD patients older than 15 years is defined as a Hb concentration <13.0g/dl in males and <12.0g/dl in females [KDIGO workgroup 2013]. It represents a serious complication in CKD patients with increased risk of adverse cardiovascular and renal events [KDIGO workgroup 2013, Pfeffer et al. 2009a)]. In CKD patients, the production of EPO is mainly reduced due to impaired renal function [Rao and Pereira 2003] and resulting anemia is commonly treated by administration of ESAs, e.g. darbepoetin alfa, or iron therapy [KDIGO workgroup 2013, Pfeffer et al. 2009a)]. However, significant controversies exist regarding the target Hb concentration achieved by ESA treatment [Pfeffer et al. 2009a), Phrommintikul et al. 2007, Drüeke et al. 2006, Singh et al. 2006], as high Hb target levels are consistently associated with increased cardiovascular mortality [Strippoli et al. 2007]. Furthermore, an increased stroke risk is reported for CKD patients with diabetes mellitus 2 treated with darbepoetin alfa in the TREAT study [Skali et al. 2011]. In this context [KDIGO workgroup 2013, Singh 2010] recommend no ESA treatment for patients with mild to moderate anemia or with active or recent history of malignancy, the target Hb concentration achieved with ESA treatment should not exceed 11.5g/dl and in general, blood transfusion or treatment with short course of ESAs is only recommended as a rescue therapy for patients with Hb <9g/dl. For transplant candidates or patients with severe anemia (Hb <9g/dl) requiring regular blood transfusions, a long-term treatment with ESAs should be considered [Singh 2010].

4.1.5 Clinical study design in nephrology

Patient-oriented clinical trials are, in general medical science as well as in the specific field of nephrology, conducted in order to establish guidelines for clinical decision making [Palmer et al. 2011]. They can be subdivided into two major categories according to the investigator's role: interventional and observational studies [Thiese 2014, Palmer et al. 2011]. These categories both comprise three basic steps of research, i.e. definition and assessment of exposure in two or more patient groups, outcome measurement in these groups, and statistical analysis in order to make inferences about possible relationships between exposure and outcome [Thiese 2014].

In interventional or experimental trials, the investigator acts upon the study participants as part of the study design [Thiese 2014]. This offers the opportunity to investigate the effect of therapeutic interventions, e.g. preventive drugs [Thiese 2014]. Several different types of interventional studies exist, e.g. pre-post, non-randomized controlled, crossover randomized controlled, and randomized controlled trials, whereas the latter is commonly considered to be

GFR categories of CKD		
G1	normal or high	$\geq$90ml/min/1.73m^2
G2	mildly decreased	60-89ml/min/1.73m^2
G3a	mildly to moderately decreased	45-59ml/min/1.73m^2
G3b	moderately to severely decreased	30-44ml/min/1.73m^2
G4	severely decreased	15-29ml/min/1.73m^2
G5	kidney failure	<15ml/min/1.73m^2
Albuminuria categories of CKD		
A1	normal to mildly increased	<30mg/mmol
A2	moderately increased	3-30mg/mmol
A3	severely increased	>30mg/mmol

Table 4.2: **CKD staging system based on GFR and albuminuria criteria.** Abbreviations: GFR, glomerular filtration rate. Modified from [KDIGO workgroup 2013].

the gold standard [Thiese 2014, Concato 2013].
In randomized controlled trials, a homogeneous group of study participants is randomly split into two or more separate groups with, therefore, equal characteristics and confounders [Thiese 2014]. Then, one group is left untreated, the so-called control group, whereas a therapy is administered to the other(s) [Thiese 2014]. The analysis of differences between the treated group(s) and the control offers the possibility of making inferences about the therapy's impact on the treated group(s) [Thiese 2014]. The vigorousness of randomized controlled trials comprises balanced baseline characteristics due to randomization, collection of pertinent data due to a "prospective" infrastructure, and mainly straightforward analytical evaluation of differences between treated and control group(s) [Concato 2013]. On the other hand, different randomized controlled trials on the same topic often show contradictory results due to, e.g. differences concerning technical issues, inclusion and baseline criteria, etc., as well as comparisons between meta-analyses, i.e. combined statistical analysis of several small independent studies [Fisher and Wood 2007], and large randomized controlled trials [Concato 2013]. Furthermore, randomized controlled trials have limited generalizability due to strict inclusion criteria, can be very expensive and time consuming, often demand complex administration, and patient recruitment can be challenging [Concato 2013, Fisher and Wood 2007].
The study design of a randomized controlled trial can be enhanced by following further guidelines: Allocation concealment, guaranteed for both study participants and involved researchers, should theoretically prevent biased randomization [Thiese 2014]. In comparison, blinding represents the concealment of the group membership from the study participants and involved researchers by, e.g. treating the control group with a placebo medication [Thiese 2014]. It assures equal treatment for the control and the treated group(s) and consequently minimizes biases [Thiese 2014]. By measuring the compliance of study participants and potentially adjusting for differences in intervention adherence between control and treated group(s), erroneous results of the statistical analysis can be avoided [Thiese 2014]. Co-interventions, which might

have a different impact on the outcome than the primary treatment of the study, should be either excluded or statistically adjusted in order to prevent wrong deductions about the treatment effect [Thiese 2014]. The number of study participants, who drop out of a trial, should be counted and considered while interpreting the study results [Thiese 2014]. Deviations from random allocation can be regulated by an intention-to-treat analysis, where the data from study participants is analysed based solely on their allocated intervention, i.e. regardless whether they actually received a treatment or not [Thiese 2014]. It consequently allows deductions about the randomization benefits, but relies on data completeness [Thiese 2014]. Finally, data should be collected from each group at the same time-point and in the same way in order to reduce substantial bias [Thiese 2014].

In observational studies, also called epidemiological studies, the investigator does not perform any interventions on the study participants, but solely observes exposure, e.g. disease, therapy, etc., and outcome [Thiese 2014]. They can be, based on their measures of disease and risk as well as temporality, further subdivided into ecological, proportional mortality, case-crossover, cross-sectional, case-control, as well as retrospective and prospective cohort studies [Thiese 2014]. The gold standard for observational studies are prospective cohort studies [Thiese 2014]. For cohort studies, study participants are classified according to their exposure status and are either followed through time in order to determine their outcome (prospective) or their health data, which has been recorded prior to outcome development, is retrospectively evaluated [Thiese 2014]. Cohort studies allow the determination of incidence, point and period prevalence, as well as numerous risk measures [Thiese 2014, Fisher and Wood 2007].
In general, diagnostic accuracy studies, where a new diagnostic method is compared to the current "gold standard" in a cross-section of both diseased and healthy study participants, are also classified as observational studies, although they could be seen as a unique category [Thiese 2014].
Even though randomized controlled trials are often stated to be superior to observational studies [Fisher and Wood 2007], the latter still offer certain advantages [Concato 2013]. They can be further improved by restricted eligibility criteria and well-defined time-points of last interventions [Concato 2013]. Moreover, observational studies and corresponding randomized controlled trials proved to show similar results, and the conclusions derived from observational studies seem to be trustworthy [Concato 2013]. Additionally, observational cohort studies seem to offer greater generalizability than randomized controlled trials due to broader patient populations [Fisher and Wood 2007, Jager et al. 2007]. However, observational studies usually exhibit imbalanced baseline characteristics, the quality of the data collected with regard to the research question can be variable and therefore introduce bias, and the analytical methods can be very complex and obscure [Concato 2013].

Alternative categorization schematas for clinical trials are based on the temporal nature of the study, i.e. retrospective or prospective, the usability of the study results, i.e. basic or applied, the investigator's aim, i.e. descriptive or analytical, or the study purpose, i.e. prevention, diagnosis, or treatment [Thiese 2014].
Retrospective studies, where data on past exposures and outcomes is collected based on medi-

cal records or the participants' memories, do not offer easy deductions about temporality and can be prone to several biases depending on the record quality [Thiese 2014]. In comparison, prospective trials, where study participants are monitored forward through time, are less prone to biases and deductions about causality are easier achieved due to determined time-frames for exposure and outcome [Thiese 2014].

Which study design is to be chosen for an actual trial, crucially depends on the research question and practical aspects, e.g. costs, available infrastructure, etc. [Fisher and Wood 2007].
The clinical trials included in this thesis both comprise interventional, e.g. TREAT, and observational, e.g. GCKD, studies and the individual study designs are briefly introduced in the respective Introduction parts of section 5.

4.2 Fundamentals of nuclear magnetic resonance spectroscopy

4.2.1 The theory of nuclear magnetic resonance spectroscopy

This introduction approaches NMR spectroscopy in a classical way and does not go into quantum mechanics or quantum electro dynamics. Quantum mechanical background can be found in [Cavanagh et al. 1996, Ernst et al. 1987].
The spin angular momentum, in general, is a fundamental, quantum-mechanical property of particles such as protons, neutrons, and electrons, and can be described mathematically as a vector [Ernst et al. 1987].
The principle of NMR spectroscopy is based on the presence of a nuclear spin angular momentum $\vec{I}$ for certain nuclei and the corresponding nuclear magnetic moment $\vec{\mu}$ [Cavanagh et al. 1996]. The magnitude of $\vec{I}$ is given by

$$|\vec{I}|^2 = \hbar^2[I(I+1)]. \tag{4.1}$$

I is the angular momentum quantum number and $\hbar = 1.055 \cdot 10^{-34}$Js is Planck's constant divided by 2π. Nuclei with odd mass numbers possess half-integer angular momentum quantum numbers, the most important nuclei with $I = \frac{1}{2}$ are ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P [Cavanagh et al. 1996]. Nuclei with an even mass and an even charge number have angular momentum quantum numbers equal to zero and are therefore NMR-inactive [Cavanagh et al. 1996]. Nuclei with an even mass number and an odd charge number have integer angular momentum quantum numbers, whereas the most important nuclei with $I = 1$ are ^{2}H and ^{14}N [Cavanagh et al. 1996, Ernst et al. 1987]. As nuclei with $I > \frac{1}{2}$ additionally possess electric quadrupole moments due to nonspherical nuclear charge distributions, the lifetimes for their magnetic states in solution are usually much shorter than for nuclei with $I = \frac{1}{2}$ [Cavanagh et al. 1996]. Consequently, NMR resonance lines of such quadrupolar nuclei are broader and more difficult to resolve than resonance lines for nuclei with $I = \frac{1}{2}$ [Cavanagh et al. 1996].
Due to quantum mechanical restrictions, only one Cartesian component of $\vec{I}$, usually I_z, can be measured simultaneously with $\vec{I}^2$ [Cavanagh et al. 1996], and the magnitude of the nuclear spin

angular momentum $\vec{I}$ is always greater than its z-component I_z [Ernst et al. 1987, Cavanagh et al. 1996]:

$$I_z = \hbar m, \qquad |\vec{I}|^2 > I_z^2, \tag{4.2}$$

where m defines the magnetic quantum number with $m = (-I, -I+1, ..., I-1, I)$. Correspondingly, I_z possesses $2I+1$ possible values and the orientation of the nuclear spin angular momentum $\vec{I}$ in space is quantized [Cavanagh et al. 1996].
Nuclei with a non-zero nuclear spin angular momentum $\vec{I}$ also have a magnetic moment $\vec{\mu}$, defined by [Cavanagh et al. 1996]

$$\vec{\mu} = \gamma \vec{I}, \qquad \mu_z = \gamma I_z = \gamma \hbar m, \tag{4.3}$$

with γ being the gyromagnetic ratio, a characteristic constant for a given nucleus. The receptivity of a nucleus in NMR spectroscopy depends, in part, on the magnitude of γ, as illustrated below [Cavanagh et al. 1996]. In the absence of external fields, the quantum states of an isolated spin corresponding to $2I+1$ different values of m have the same energy and the nuclear spin angular momentum $\vec{I}$ and therefore the magnetic moment $\vec{\mu}$ have no preferred orientation [Cavanagh et al. 1996].
In presence of an external static magnetic field $\vec{B_0}$, the spin quantum states of the nucleus possess the following discrete energies [Cavanagh et al. 1996]:

$$E_m = -\vec{\mu} \cdot \vec{B_0}. \tag{4.4}$$

In case of $\vec{B_0} \parallel \hat{e}_z$, as usually postulated in an NMR spectrometer, one obtains [Cavanagh et al. 1996]

$$E_m = -\mu_z B_z = -\gamma I_z B_z = -m\hbar\gamma B_z, \tag{4.5}$$

in which B_z is the static magnetic field strength. The minimum energy E_0 equals zero, corresponding to a magnetic quantum number of $m = 0$ [Cavanagh et al. 1996].
Consequently, $2I+1$ different energy levels exist, the so called Zeeman levels, all separated from each other by equal spaces [Cavanagh et al. 1996]. The transition condition equals $\Delta m = \pm 1$ and the energy gap ΔE between two neighboring Zeeman levels is [Cavanagh et al. 1996]

$$\Delta E = \gamma \hbar B_z. \tag{4.6}$$

This energy gap ΔE corresponds, according to Planck's law to a frequency ν_0 of electromagnetic radiation required to excite a transition [Cavanagh et al. 1996],

$$\nu_0 = \frac{\Delta E}{h} = \frac{\gamma B_z}{2\pi}, \qquad \omega_0 = 2\pi\nu_0 = \gamma B_z, \tag{4.7}$$

where $h = 6.63 \cdot 10^{-34}$ Js denotes Planck's constant.
With $|\vec{I}|^2 > I_z^2$ and $\vec{\mu}$ and $\vec{I}$ being collinear, $\vec{\mu}$ and $\vec{B_0}$ are not collinear [Ernst et al. 1987]. This leads to a precession of $\vec{\mu}$ around $\vec{B_0}$ with the so-called Larmor frequency ω_0 [Ernst et al. 1987], as, for example, depicted in Figure 4.2a) for a single nucleus with an angular momentum

quantum number $I = \frac{1}{2}$ and nuclear spin angular momentum $\vec{I}$. The Larmor frequency ω_0 is equal to the required excitation frequency given in equation (4.7). $\vec{I}$ can be either parallel oriented to an external magnetic field $\vec{B_0}$, the so-called spin-up state with the energy level α, corresponding to a magnetic quantum number $m = +\frac{1}{2}$ [Ernst et al. 1987]. Or $\vec{I}$ can be oriented anti-parallel to $\vec{B_0}$, the so-called spin-down state, corresponding to $m = -\frac{1}{2}$ and a higher energy level β [Ernst et al. 1987].

In a bulk material with millions of nuclei, all magnetic moments of the single nuclei need to be added up as vectors in order to calculate the macroscopic magnetization $\vec{M}$ [Cavanagh et al. 1996]. At thermal equilibrium, the number of nuclei N_α in the spin-up state at the lower energy level, is slightly higher than the number of nuclei N_β in the spin-down state [Cavanagh et al. 1996]. This population difference can be calculated as [Ernst et al. 1987]

$$\frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} = \frac{1 - \exp^{-\Delta E/k_B T}}{1 + \exp^{-\Delta E/k_B T}} \approx \frac{\Delta E}{2k_B T} = \frac{\gamma \hbar B_z}{2k_B T}. \tag{4.8}$$

Here, $k_B = 1.38 \cdot 10^{-23}$ JK^{-1} represents Boltzmann's constant and T the temperature. The longitudinal macroscopic magnetization $\vec{M}$ is therefore parallel oriented to the external static magnetic field $\vec{B_0}$ and precesses with Larmor frequency ω_0 around the z-axis [Cavanagh et al. 1996, Ernst et al. 1987]. Note that only the population difference between energy state α and β as calculated in (4.8) contributes to the macroscopic magnetization and to the detected NMR signal [Ernst et al. 1987], as illustrated in Figure 4.2b). This population difference depends on the nucleus type and the applied static magnetic field strength as well as the temperature [Cavanagh et al. 1996] and is usually on the order of 1 in 10^5 for ^{1}H spins for static magnetic field strengths of about 11T at room temperature [Cavanagh et al. 1996]. Hence, NMR spectroscopy is quite insensitive in comparison to other spectroscopic techniques like visible or ultraviolet spectroscopy [Cavanagh et al. 1996]. Its sensitivity can be increased by choosing higher static magnetic field strengths [Ernst et al. 1987, Cavanagh et al. 1996].
If a radio frequency (rf) pulse with frequency ν_{rf} equal to the Larmor frequency ν_0 is irradiated onto the bulk material, the orientation of $\vec{M}$ can be altered [Ernst et al. 1987, Cavanagh et al. 1996]. $\vec{B}_{rf}(t)$ is the magnetic field component of this induced electromagnetic field and is, for example, linearly polarized along the x-axis [Ernst et al. 1987]. It can be fractionized into two circularly polarized magnetic fields, which rotate around the z-axis in opposite directions [Ernst et al. 1987]. $\vec{M}$ can only significantly interact with the component of $\vec{B}_{rf}$ rotating in the same sense as itself [Ernst et al. 1987]. It starts to perform an additional rotation around $\vec{B}_{rf}$, i.e. around the x-axis, and moves away from the z-axis towards the x-y-plane [Ernst et al. 1987]. At the same time, $\vec{M}(t)$ still rotates around the z-axis due to the external static magnetic field $\vec{B_0}$ leading to a helical line movement [Ernst et al. 1987]. This process is illustrated in Figure 4.3a). In the considered case of a 90° rf pulse, the rf-field is switched off when $\vec{M}(t)$ reaches the x-y-plane [Ernst et al. 1987]. $\vec{M}(t)$, now called transversal magnetization, rotates in the x-y-plane with angular frequency $\omega_0 = \gamma B_z$ [Ernst et al. 1987]. It induces, due to electromagnetic induction, the NMR signal in the receiver coil (represented as an eye), located in the x-y-plane [Bloch 1946, Ernst et al. 1987], as illustrated in Figure 4.3b). The NMR spectrometer, employed for this thesis, uses a magnetic field strength B_z of 14.1 T and consequently works

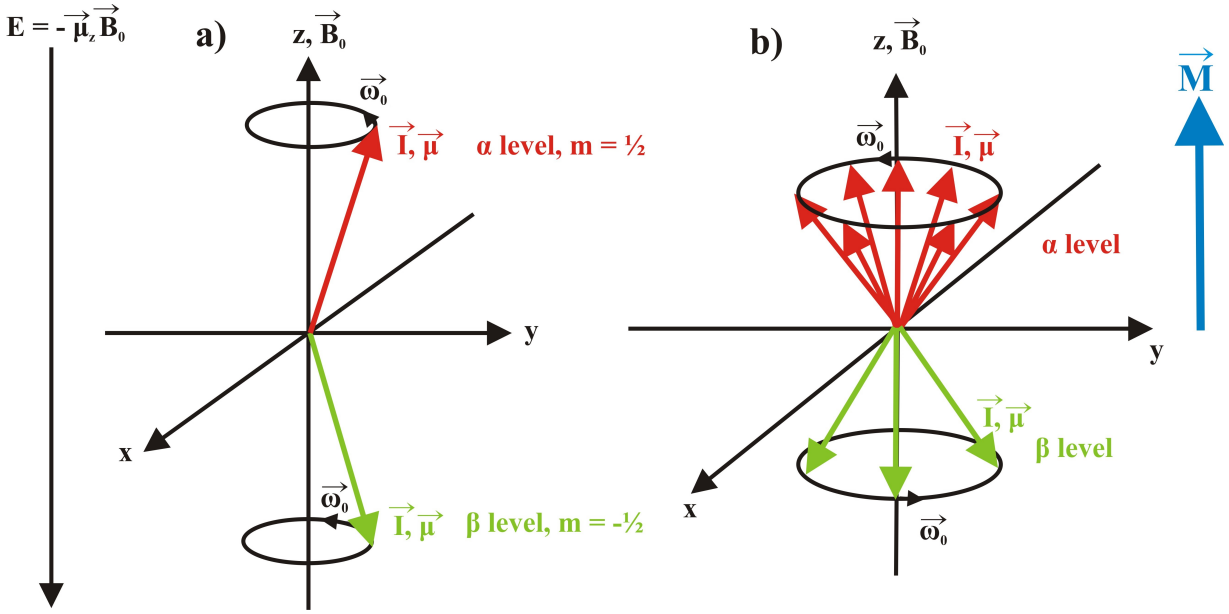


Figure 4.2: **Illustration of two different energy levels α (spin-up state) and β (spin-down state) for the nuclear spin represented by the spin angular momentum $\vec{I}$ with an angular momentum quantum number $I = \frac{1}{2}$ in an external magnetic field $\vec{B}_0$.** a) shows a single nucleus in an external magnetic field $\vec{B}_0$ and b) illustrates the longitudinal macroscopic magnetization $\vec{M}$ in bulk material at thermal equilibrium. More nuclei are in the energy state α than in the energy state β. The longitudinal macroscopic magnetization $\vec{M}$ is parallel oriented to the external magnetic field $\vec{B}_0$. The x- and y-fractions of $\vec{I}$ and $\vec{\mu}$, respectively, which rotate with Larmor frequency ω_0, compensate each other. Taken from [Zacharias 2012].

with a resonance frequency for protons of approximately 600 MHz. The time for recording the NMR signal is called the acquisition period and the signal itself is called the free induction decay (FID) [Ernst et al. 1987], compare to Figure 4.3c). The time-dependent FID, whose data points are not continually recorded, but only in distinctive time intervals facilitated by the Nyquist theorem, is transformed into the corresponding NMR spectrum as a function of frequency ν using, in most cases, a fast Fourier transformation [Ernst et al. 1987, Cavanagh et al. 1996].

Two different relaxation effects of the macroscopic magnetization $\vec{M}$ occur, which are mainly induced by interactions with surrounding electromagnetic fields [Bloch 1946, Ernst et al. 1987]. The longitudinal relaxation, also called T_1-relaxation, arises from spin-lattice-interaction and leads to a reorientation of the transversal magnetization back to a longitudinal orientation parallel to $\vec{B}_z$ carried out in the time period T_1, see Figure 4.4a) [Bloch 1946, Ernst et al. 1987]. T_1 depends on the mobility of the lattice and determines the waiting time between two acquisition periods [Ernst et al. 1987].

The transverse relaxation or T_2-relaxation describes the dephasing of the transversal magnetization due to intramolecular spin-spin or dipole-dipole-interaction in the time period T_2 [Bloch

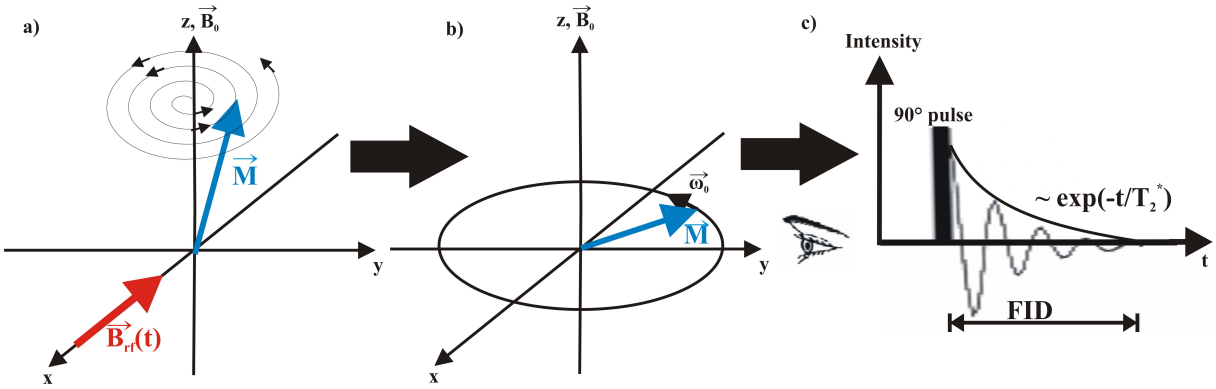


Figure 4.3: **A 90° rf pulse irradiates on a bulk material with macroscopic magnetization $\vec{M}$ parallel oriented to $\vec{B}_0$.** a) The rf pulse with the magnetic field component $\vec{B}_{rf}(t)$ causes an additional rotation of $\vec{M}$ around the x-axis leading to a helical line movement of $\vec{M}$. b) $\vec{M}$ rotates in the x-y-plane with angular frequency ω_0 and induces an electrical potential in the receiver coil (represented as an eye), which is the measured NMR signal. c) During the acquisition period, the FID, a superposition of different sine and cosine waves, is recorded as the NMR signal. It declines exponentially due to spin-spin relaxation, indicated by the time constant for inhomogeneous magnetic fields T_2^* [Ernst et al. 1987]. Taken from [Zacharias 2012].

1946, Ernst et al. 1987], see Figure 4.4b). T_2 depends consequently on the molecule's size, on the density of the interacting nuclei, the viscosity of the solvent, and the temperature and determines the useful maximum acquisition period [Ernst et al. 1987]. Taking into account local inhomogeneities of the magnetic field B_z, which arise from fluctuations of the magnetic susceptibility, one considers the effective transversal time constant T_2^* as presented in Figure 4.3c) [Homans 1995].
These two relaxation processes are mathematically described in the phenomenological Bloch equations [Ernst et al. 1987].
The energy gap ΔE, as described in (4.6), depends on the effective magnetic field at the nucleus [Cavanagh et al. 1996]. Local changes of the magnetic field at a nucleus can arise due to magnetic shielding effects caused by its electronic environment in a molecule [Cavanagh et al. 1996]. Hence, the Larmor frequency ω_0 of the electromagnetic radiation required to excite a transition between different spin states of a nucleus, as given in (4.7), is shifted and now denoted as ω_{local} [Cavanagh et al. 1996]. In general, this shift is given as the chemical shift δ in ppm and is independent of the applied static magnetic field B_0 [Cavanagh et al. 1996, Ernst et al. 1987]

$$\delta = \frac{\omega_{local} - \omega_{ref}}{\omega_0} \times 10^6. \tag{4.9}$$

ω_{ref} denotes the offset resonance frequency of a reference substance, the so called internal standard compound, for example tetramethylsilan (TMS; $Si(CH_3)_4$) or 3-trimethylsilyl-2,2,3,3-

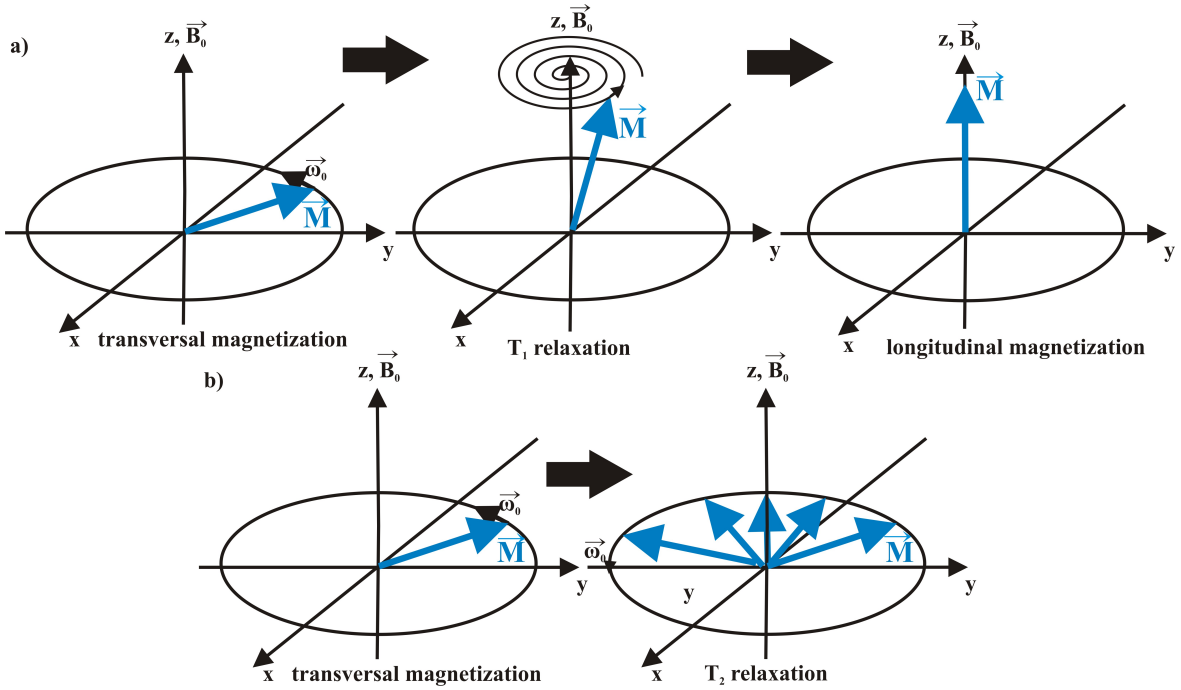


Figure 4.4: **Illustration of T_1 and T_2 relaxation.** a) Due to spin-lattice-interaction, the transversal magnetization relaxes back to the original longitudinal relaxation in the time period T_1. b) Due to spin-spin- or dipole-dipole-relaxation, the transversal magnetization dephases in the x-y-plane in the time period T_2. Taken from [Zacharias 2012].

tetradeuteropropionate (TSP), and can be calculated using quantum mechanics or the chemical shift increment system.

With the resonance frequency and consequently the chemical shift δ of a nucleus in a certain molecular surrounding being distinct, NMR spectroscopy enables the identification and validation of compounds in an investigated substance. The NMR spectrum can be seen as a molecular fingerprint of different molecular structures present in biological fluids. However, it is important to consider the influence of the solvent, the salt concentration, the pH value, the temperature and other environmental impacts of the investigated specimen on the chemical shift.

For a nucleus A with nuclear spin angular momentum $\vec{I_A}$, which is located near a second nucleus X with nuclear spin angular momentum $\vec{I_X}$, possible coupling effects need to be considered [Ernst et al. 1987]. One possible coupling mechanism is called scalar or J-coupling, also called indirect dipole-dipole coupling, which is mediated by the electrons of the chemical bonds between these two nuclei [Ernst et al. 1987]. J is the coupling constant for such a two spin system and its magnitude depends on the number and types of bonds between nuclei A and X, as well as, if applicable, the dihedral angles between them [Ernst et al. 1987, Homans 1995]. It is therefore a distinct constant for the considered spin system with fixed dihedral angle [Ernst et al. 1987]. The J-coupling mechanism changes the energy levels of a non-coupled two spin system, as illustrated in Figure 4.5. A two spin system has four possible energy levels E in a

molecular surrounding, depending on the orientation of the two nuclear spin angular momenta $\vec{I_A}$ and $\vec{I_X}$ with respect to the static magnetic field [Cavanagh et al. 1996]. In the presence of scalar coupling conveyed by the chemical bonds between the two nuclei, the energy levels are shifted [Ernst et al. 1987, Cavanagh et al. 1996], as illustrated in Figure 4.5b). Consequently, the two original NMR lines reflecting transitions between E_1 and E_3, as well as E_2 and E_4 or transitions between E_1 and E_2, as well as E_3 and E_4 (compare to Figure 4.5c)) are now split into two lines, respectively, each of them reflecting one of the four possible transitions (compare to Figure 4.5d)) [Cavanagh et al. 1996, Ernst et al. 1987]. Note that the resulting resonance lines all have the same intensity, as the small population difference between spin-up and spin-down state is neglected in this case [Ernst et al. 1987]. However, this is only true for the so-called high-field or weak coupling approximation, which assumes that the J couplings are much smaller than the chemical shifts between resonances of the coupled nuclei [Homans 1995]. Depending on the magnitude of the coupling constant J as well as the presence of additional neighboring nuclei, the resonance line pattern of a nucleus in a molecular surrounding becomes even more complex [Ernst et al. 1987]. If the magnitude of J approximates the value of the chemical shift difference between the two nuclei, the weak coupling approximation becomes invalid and one refers to strong coupling [Homans 1995]. The resulting resonance lines in Figure 4.5d) become distorted, also known as roofing-effect [Ernst et al. 1987, Homans 1995]. For magnetically equivalent nuclei, i.e. nuclei with the same chemical environment and only one coupling constant to the neighboring nuclei, no line splitting is observed [Ernst et al. 1987].
Another coupling mechanism is dipolar coupling, which is based on direct interaction between different nuclei through space [Ernst et al. 1987]. It is independent of the chemical bonds between the two nuclei, but depends on the angle between the static magnetic field and the vector connecting the two nuclei [Ernst et al. 1987]. In isotropic solution, where molecules can move freely, this angle continually changes and the dipolar coupling is averaged to zero [Ernst et al. 1987]. Therefore, line splitting due to dipolar coupling effects are, in general, not visible in the NMR spectrum [Ernst et al. 1987]. However, fast rotation of the molecules and intra-molecular motions give rise to fluctuations of their magnetic fields [Ernst et al. 1987]. A fluctuating field of one nucleus can lead to longitudinal and transversal relaxation of the neighboring nuclei [Ernst et al. 1987]. This is, in case of longitudinal relaxation, called cross relaxation [Ernst et al. 1987] and described by the Solomon equations [Solomon 1955].

Scalar or dipolar coupling mechanisms can be employed in order to transfer magnetization between different nuclei populations [Ernst et al. 1987]. This property can be utilized to transfer magnetization from nuclei, which are sensitive to external rf pulses (e.g. ^{1}H nuclei), to insensitive nuclei (e.g. ^{13}C nuclei), and vice versa, as realized, for example, in two-dimensional (2D) heteronuclear NMR experiments [Ernst et al. 1987]. An important example is the indirect measurement of ^{13}C spectra via ^{1}H spectra [Ernst et al. 1987]. 2D NMR spectra offer the possibility of effectively reducing the excessive signal overlap present in 1D NMR spectra of complex biofluids, e.g. urine, and can illustrate correlations between different nuclei in a molecule [Ernst et al. 1987]. 2D NMR experiments can usually be subdivided into four different parts, namely preparation, evolution characterized by a variable time interval t_1, mixing, and acquisition characterized by the time interval t_2 [Cavanagh et al. 1996, Ernst et al. 1987].

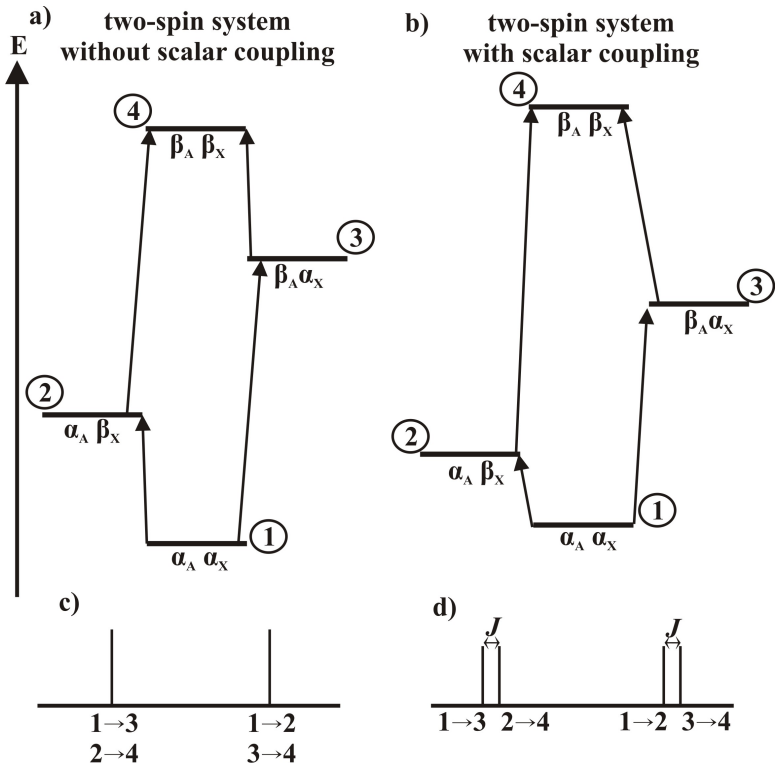


Figure 4.5: **The effect of J-coupling on the energy levels of a two-spin system.** a) A two-spin system in the absence of J-coupling has four different energy levels. The differences between energy levels 1 and 2 and energy levels 3 and 4 are equivalent, the corresponding transitions give rise to one line in the NMR spectra (same holds true for differences between energy levels 1 and 3 and energy levels 2 and 4), as illustrated in c) [Cavanagh et al. 1996]. b) In the presence of J-coupling, the four energy levels of a two-spin system are shifted, and the previous two NMR lines are split into four [Cavanagh et al. 1996], as illustrated in d) under the assumption of weak coupling. Modified from [Cavanagh et al. 1996].

By increasing t_1 m-times in a step-wise manner and thereby executing the pulse sequence and recording an FID of n digitized data points for each increment, one obtains an $m \times n$ data matrix [Cavanagh et al. 1996]. A double Fourier transformation consequently yields a 2D spectrum as a function of two frequency variables [Cavanagh et al. 1996, Ernst et al. 1987].

4.2.2 General data acquisition

This section comprises a general overview of the different data acquisition steps performed for this thesis. Details and modifications specific for certain projects are explicitly addressed in the respective Materials and Methods parts of section 5.

4.2.2.1 Sample characteristics and preparation

The main biofluids measured for this thesis are human urine and plasma. A general overview of the sample characteristics and preparation procedures for these two fluids is given here.

Urine is the most popular biofluid for metabolomic investigations [Emwas et al. 2014]. It comprises a large range of metabolites and reflects the most metabolic processes, which take place throughout the body [Emwas et al. 2014]. As urine is not constantly released by the body, but rather stored in the bladder until excretion, it represents a "time-averaged" profile of whole-body homeostasis [Kosmides et al. 2013, Maher et al. 2007]. With the kidney being responsible for its creation, compare to section 4.1.1, the urine composition can excellently reflect the renal performance, as it is utilized in traditional clinical monitoring described in section 4.1.2. Moreover, urine specimens are easily obtained in a non-invasive way and are usually available in large volume [Emwas et al. 2014]. Furthermore, in comparison to, for example, blood, urine contains fewer protein complexes and lipids [Emwas et al. 2014], whose broad NMR signals need to be attenuated in order to obtain well resolved spectra.
However, the metabolic concentration in a urine specimen of an individual crucially depends on the person's fluid intake, adding to overall data variance [Dieterle et al. 2011]. This effect is usually addressed by employing appropriate scaling and normalization techniques, which are discussed in section 4.3.1.
Further individual-related factors contributing to significant variation of the urinary metabolic concentrations are diet, age and gender effects, metabolic phenotypes, gut microflora effects, comorbidities, drug administration, and physical activity, which can be interconnected [Emwas et al. 2014]. These effects can be overcome by appropriate matching of the compared patient groups and by applying various normalization techniques, as described in section 4.3.1.
Factors related to specimen collection and storage are different sample collection time-points, presence of human or bacterial cells that might break open, different sample storage conditions, and repeated thawing and freezing [Emwas et al. 2014]. Human or bacterial cells present in urine are usually removed by a short centrifugation step prior to sample freezing [Emwas et al. 2014, Zacharias et al. 2013b)], as applied in this thesis. Differences in sample storage conditions were avoided by employing unified storage conditions for all biofluid specimens belonging to one project. The number of thawing and freezing cycles was minimized.
Furthermore, the individual pH value of the specimens as well as differences in osmolality of the specimens, ionic strength or metal ion composition lead to chemical shift changes [Emwas et al. 2014, Zacharias et al. 2013b), Ross et al. 2007]. The individual pH value of each sample was adjusted to 7.0 by adding an appropriate buffer volume [Zacharias et al. 2013b)], as described below, and general differences in chemical shifts were compensated by spectral binning, compare to section 4.2.2.3.

The second most prominent biofluid for metabolomic investigations is blood plasma or serum [Kosmides et al. 2013]. For this thesis, only plasma samples were investigated. Plasma is the liquid carrier of the blood cells and is usually obtained by adding anti-coagulants, e.g. EDTA, to the blood, centrifuging it and subsequently decanting the remaining liquid [Psychogios et al.

2011].
Blood circulates around and inside every tissue and organ of the body and carries all molecules that are secreted, excreted, or discarded during the bodies metabolism [Psychogios et al. 2011]. Therefore, the composition of blood/plasma is strongly affected by organ dysfunctions, tissue lesions, and pathological states of the body, justifying its prominent role in clinical tests [Psychogios et al. 2011]. In comparison to urine, plasma metabolic profiles rather represent an "instantaneous" picture of the whole-body homeostasis [Kosmides et al. 2013, Maher et al. 2007]. As the kidney serves as a filter for blood/plasma in the body, compare to section 4.1.1, alterations of the blood/plasma composition are strong indications of modified renal function. Consequently, traditional clinical monitoring of the renal performance is also based on the assessment of blood/plasma composition, as depicted in section 4.1.2. In general, blood/plasma is easily accessed by employing a minimally invasive procedure [Psychogios et al. 2011].
In comparison to urine, however, plasma specimens contain large amounts of macromolecules, especially proteins, which give rise to broad and unspecific NMR signals [Zacharias et al. 2013b)]. Therefore, macromolecules should be removed from plasma specimens prior to NMR data acquisition by, e.g. ultrafiltration, or their NMR signals should be attenuated employing, for instance, the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence [Zacharias et al. 2013b)], as introduced in section 4.2.2.2. For ultrafiltration, Millipore Amicon Ultra-4 (Millipore, Billerica, MA, USA) cellulose filter devices with a molecular weight cutoff of 10 kDa were used in this thesis.
[1] Filters were prewashed with 3 ml of distilled water and centrifuged at 4000 $\times$ g in a swing-bucket rotor at 22°C for 30 min in order to remove filter-preserving substances like glycerol and triethylene glycol. Spectra of blank samples of rinsing water were acquired and compared to spectra of filtered biofluid to detect and exclude signals from filter residues in subsequent data analysis. 1000 µl of plasma were placed into the filter device and centrifuged at 4000 $\times$ g at 4°C for 60 min.

Although the metabolite composition and concentration of plasma is usually tightly controlled and therefore not significantly influenced by individual fluid intake [Warrack et al. 2008], certain other patient-related factors can affect its composition, e.g. diet, age, sex, comorbidities, and drug administration, which also need to be addressed appropriately. Like urine, plasma composition can be influenced by collection and storing conditions. These collection and storing conditions were unified for all plasma specimens belonging to one project for this thesis. Furthermore, the number of thawing and freezing cycles was also minimized for plasma samples.

All biofluid specimens measured for this thesis were either immediately stored at -80°C or put on ice after collection and stored at -80°C as soon as possible until measurement. Prior to sample preparation, biofluid specimens were thawed at room temperature. Then, usually 400 µl of each biofluid sample (i.e. either urine, filtered or unfiltered plasma) was placed in an individual 5 mm NMR tube (Bruker BioSpin GmbH, Rheinstetten, Germany, or Norell Inc., Marion, USA) mixed with 200 µl of 0.1 mol/l phosphate buffer at pH 7.4 and 50 µl of 0.75%

[1] The following description of the filtering procedure was similarly published in [Zacharias et al. 2013b)].

(w) of the sodium salt of TSP (Sigma-Aldrich, Taufkirchen, Germany), solved in deuterium oxide (D_2O), as recommended in [Zacharias et al. 2013b)]. TSP serves as an internal standard and provides the reference signal, the phosphate buffer is added in order to stabilize the pH value of the biofluid samples at 7.0 and D_2O is used as a reference for the internal lock signal of the spectrometer [Zacharias et al. 2013b)]. Moreover, the buffer contained 30 mg boric acid (3.9 mmol/l) in order to inhibit bacterial growth [Zacharias et al. 2013b)]. Further information about alternative internal standards or buffer solutions is given in [Zacharias et al. 2013b)].

4.2.2.2 Experimental setup of the employed NMR spectrometer

All NMR experiments for this thesis were carried out on a 600 MHz Bruker Avance III (Bruker BioSpin GmbH, Rheinstetten, Germany) employing a triple-resonance (^1H, ^{13}C, ^{31}P, ^2H lock) cryogenic probe equipped with z-gradients and an automatic cooled sample changer (SampleJet, Bruker BioSpin GmbH, Rheinstetten, Germany). The spectrometer's operating frequency of 600 MHz for protons corresponds to a magnetic field strength of 14.1 T. This magnetic field is generated by a superconducting magnet, which is cooled by liquid helium at 4 K, whereas the helium Dewar itself is cooled by liquid nitrogen at about 77.35 K [Butler 2002]. This magnet is surrounded by a shielding superconducting magnet of inverse polarization, also located in the Dewar, to minimize the magnetic field outside the spectrometer. The probe head contains the emitter/receiver coils for the frequencies of the ^1H, ^{13}C, and ^{31}P nuclei and a separate coil for locking purposes. They are cooled by helium gas. Presaturation of the large water signal is enhanced by an additional coil generating magnetic field gradients along the z-axis. A picture of the used spectrometer is given in Figure 4.6.
The prepared samples were placed in the sample changer at 4°C. A robotic arm inserted an individual sample into the static magnetic field of the superconducting magnet through the bore. Before starting the measurements, each sample was allowed to equilibrate for 300 s in the magnet at 298 K (25°C), as recommended in [Zacharias et al. 2013b)]. The temperature unit had been previously calibrated using a deuterated methanol sample [Zacharias et al. 2013b)]. For each inserted specimen, the following steps were, if not indicated elsewise, performed automatically, employing the automated acquisition suite ICON-NMR (Bruker BioSpin GmbH, Rheinstetten, Germany) included in the TopSpin program (latest version: TopSpin 3.1) (Bruker BioSpin GmbH, Rheinstetten, Germany). First, the cryogenic probe head was "tuned and matched" by a "wobble" routine, i.e. it was correctly tuned to the observe frequency of the inserted sample, which depends on the solvent of the specimen, and the impedance of the network was correctly matched [Butler 2002]. Then, the probe head was locked onto the resonance frequency of D_2O, which had been previously added to the sample as described in section 4.2.2.1, employing an independent coil [Butler 2002]. This allowed a monitoring of possible variations of the static magnetic field strength during the measurement, which can be corrected accordingly [Butler 2002]. The automated shimming procedure ensures maximum field homogeneity, which is crucial for optimal signal resolution and sensitivity [Butler 2002]. Thereby, the currents of a set of shim coils, the so-called shim system, were adjusted to eliminate any magnetic field strength gradients along the sample [Butler 2002] starting from a standard shim file op-

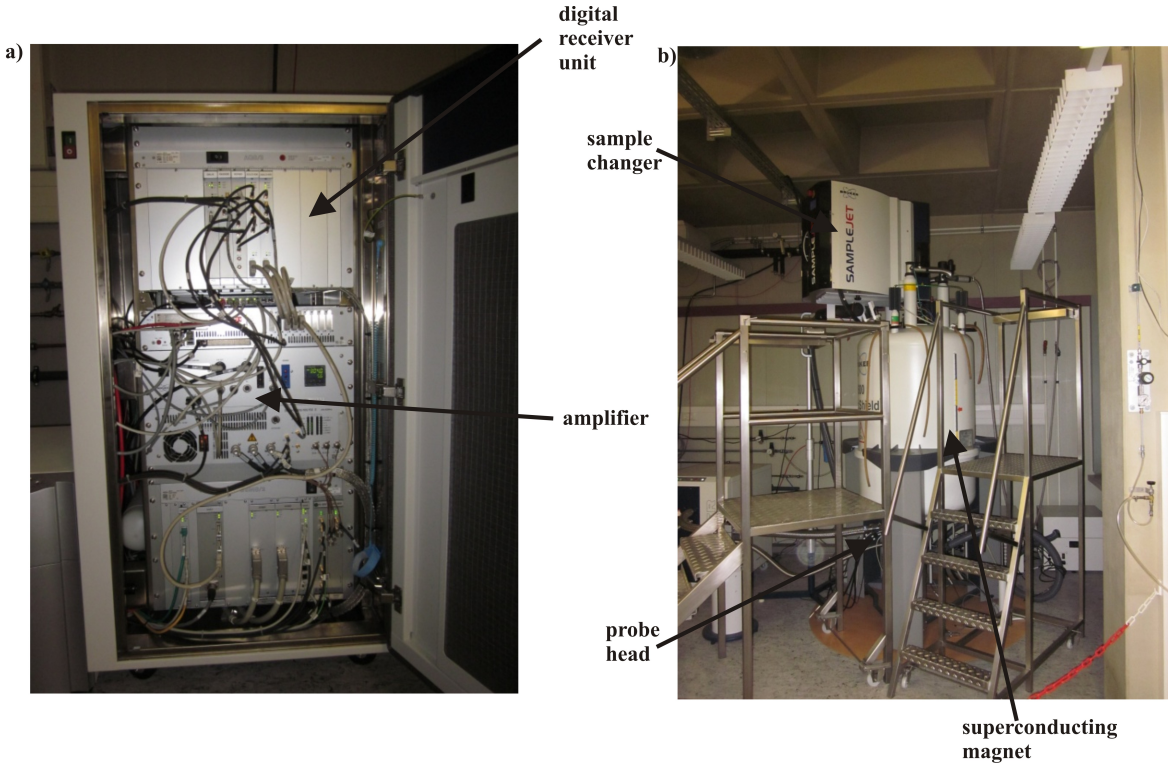


Figure 4.6: **Experimental NMR setup.** a) Console with amplifier and digital receiver unit. b) 600 MHz Bruker Avance III NMR spectrometer with cooled sample changer, superconducting magnet, and probe head. Taken from [Zacharias 2012].

timized for the respective sample matrix [Zacharias et al. 2013b)]. Finally, the pulse lengths were calibrated automatically [Zacharias et al. 2013b)]. A detailed description of the actual procedures performed for an automated tuning and matching, locking, and shimming of the probe head are beyond the scope of this thesis and can be found in [Butler 2002].
After these adjustments, the actual FIDs were recorded employing the respective pulse sequences, as detailed separately in the Materials and Methods part of each project.

In the following, some popular one-dimensional (1D) and 2D NMR experiments are briefly introduced, which were applied for this thesis.

The nuclear Overhauser enhancement spectroscopy (NOESY) experiment is very popular in NMR based metabolomic investigations of complex biofluids [Zacharias et al. 2013b), McKay 2011]. It utilizes the nuclear Overhauser effect (NOE), which describes the transfer of magnetization between different nuclei populations via cross relaxation through space [Kumar et al. 1980]. The NOE manifests itself in a fractional change in intensity of one NMR line when another resonance is perturbed [Kumar et al. 1980]. Consequently, by saturating one spin resonance, the resonance of a neighboring spin can be enhanced and the corresponding sensitivity is improved [McKay 2011]. This phenomenon is employed to effectively suppress the solvent, i.e. water, signal [McKay 2011]. Furthermore, NOESY experiments allow inference of the spatial

arrangement of nuclei in a molecule [Kumar et al. 1980]. Measurement of the first increment of the 2D NOESY pulse sequence yields a well resolved 1D spectrum with efficient water signal suppression in addition to presaturation [McKay 2011].

The Carr-Purcell-Meiboom-Gill (CPMG) [Carr and Purcell 1954, Meiboom and Gill 1958] experiment offers an effective method for the suppression of broad NMR signals arising from spins in macromolecules such as proteins [Zacharias et al. 2013b), Beckonert et al. 2007]. It employs differences in the relaxation properties of spins in macromolecules and low-molecular weight metabolites [Zacharias et al. 2013b)]. In fact, the transversal or T_2 relaxation of protons in macromolecules is faster than the T_2 relaxation of protons in small molecules due to more possible spin-spin interactions [Beckonert et al. 2007]. Consequently, after the magnetization has been transferred to the x - y - plane by an initial 90° pulse, the magnetization of protons in macromolecules and the magnetization of protons in small molecules start to dephase with different rates. During the application of appropriate 180° pulses in order to refocus the chemical shifts, the wanted magnetization of spins in small molecules relaxes with a smaller rate than the unwanted magnetization of spins in macromolecules. Consequently, the FID, acquired after an appropriate filter period, does not reflect rf signals from spins in macromolecules.

The heteronuclear single quantum coherence (HSQC) experiment correlates a nucleus with its directly attached heteronucleus via scalar coupling [Ernst et al. 1987]. Thereby, magnetization is usually transferred from sensitive ^{1}H nuclei to insensitive ^{13}C heteronuclei by scalar coupling, a process called insensitive nuclei enhanced by polarization transfer (INEPT) [Ernst et al. 1987]. The preparation period is used to establish a maximum amount of magnetization for the ^{13}C heteronuclei by employing INEPT [Ernst et al. 1987]. During the evolution period t_1, the chemical shifts of the ^{13}C heteronuclei evolve [Ernst et al. 1987]. In the mixing period, the ^{13}C magnetization is brought back to the ^{1}H nucleus by scalar coupling, which is called a reversed INEPT [Ernst et al. 1987]. The resulting ^{1}H magnetization is now recorded in the acquisition period with time t_2. During the acquisition period, the J-coupling between protons and heteronuclei is removed by applying decoupling pulses, which are longer and less powerful than excitation pulses, to avoid splitting of the acquired ^{1}H signals [Ernst et al. 1987, Butler 2002]. By varying t_1, a second time/frequency dimension is created [Ernst et al. 1987], and the resulting 2D HSQC spectrum usually shows cross peaks of the excited ^{1}H nucleus, whose resonance frequency is commonly displayed on the x-axis, and its scalar coupled ^{13}C nucleus, whose resonance frequency is commonly displayed on the y-axis [Ernst et al. 1987]. The great signal dispersion of about 140ppm in the indirect ^{13}C dimension without an increased number of signals in the 2D ^{1}H-^{13}C HSQC spectra in comparison to 1D ^{1}H spectra supports efficient metabolite identification [Zacharias et al. 2013b)].

The heteronuclear multiple bond correlation (HMBC) experiment usually transfers magnetization from sensitive ^{1}H nuclei to insensitive ^{13}C heteronuclei, which are separated from each other by more than one chemical bond [Berger and Braun 2004]. Consequently, the ^{1}H frequency of the direct proton is correlated with the ^{13}C frequency of the indirect nucleus usually separated by two to three chemical bonds [Berger and Braun 2004], similar to the HSQC experiment. Its

application for this thesis was limited to providing supplemental information about molecular structures for the identification of unknown metabolites.

The total correlation spectroscopy (TOCSY) experiment is closely related to the correlation spectroscopy (COSY) experiment. In a COSY experiment, directly coupled nuclei as well as nuclei coupled through multiple couplings are correlated with each other by scalar coupling [Berger and Braun 2004]. Cross-peaks in COSY spectra usually provide information about nuclei separated from each other by two to three chemical bonds and deductions about connectivities and even about the chemical structure of the investigated molecules can be made [Berger and Braun 2004]. In TOCSY spectra, all connected nuclei within the same spin system can be correlated via scalar coupling under the restriction that nuclei are not separated from each other by more than three bonds, since otherwise, the interactions between the nuclei become too weak to be detected [Ernst et al. 1987]. In complex biofluids like urine or plasma, compare to section 4.2.2.1, severe signal overlap is present in ^{1}H ^{1}H TOCSY spectra [Zacharias et al. 2013b)]. Therefore, ^{1}H ^{1}H TOCSY spectra have only been used as a complementary tool for metabolite identification in combination with 1D NOESY and 2D HSQC spectra in this thesis (compare to section 4.3.2).

4.2.2.3 Data preprocessing

The acquired FID is automatically Fourier transformed to yield the frequency dependent NMR spectrum employing the TopSpin program (latest version: TopSpin 3.1) (Bruker BioSpin GmbH, Rheinstetten, Germany). Prior to this, an exponential filter function with a line broadening of 0.3 Hz and zero filling to 131072 points was automatically applied for zero order phase correction for the 1D spectra [Zacharias et al. 2013b)]. For 2D HSQC spectra, squared sine functions were used as window functions for both dimensions with a shift of 90° between them. For 1D spectra, the first points of the FID were corrected using the *baseopt* option of TopSpin (latest version: TopSpin 3.1) (Bruker BioSpin GmbH, Rheinstetten, Germany) in order to obtain a flat baseline prior to Fourier transformation and to avoid first order phase distortions [Zacharias et al. 2013b)]. In the rare case of phase errors, usually by 180° mostly for highly diluted samples, spectra were manually phase corrected. Additionally, an automatic baseline correction with fifth-degree polynomial was applied employing a readily available *Python* script [Klein 2011]. 2D spectra were also preprocessed using a readily available *Python* script [Klein 2011]. Each 2D spectrum was manually phase corrected and a fifth-order polynomial baseline correction was applied in both dimensions [Klein 2011].

For subsequent statistical data analysis, as described in section 4.3.1, NMR-derived fingerprints were employed, which usually exploit variations in NMR signal position due to differences in pH, salt concentration, and/or temperature [Zacharias et al. 2013b)]. The following passage describes a procedure to account for these differences and has already been published in a slightly modified version in [Zacharias et al. 2013b)].
A widely used and robust method to compensate for the previously described effects is called binning or bucketing, whereby a spectrum is split into a number of segments called bins, buck-

ets, or features. Equal-sized buckets were used throughout this thesis, albeit other schemes such as adaptive binning have been proposed. Data points inside every bucket were summed up. The whole spectrum is then represented as a vector of bucket integrals, which are used for statistical data analysis.
The bucketing procedure for this thesis was carried out employing Amix VIEWER (latest version: Amix Viewer 3.9.13) (Bruker BioSpin GmbH, Rheinstetten, Germany). Furthermore, the region around the water artifact and the broad urea signal were excluded for both urinary and plasma NMR data. The water artifact, still remaining despite accurate water suppression, obscures all neighboring signals and might further vary across different spectra depending on the quality of the applied water suppression [Dieterle et al. 2011]. Urea, both present in urine and plasma specimens [Bouatra et al. 2013, Psychogios et al. 2011], rapidly exchanges protons with the surrounding water solvent and therefore exhibits a very broad NMR signal [Dieterle et al. 2011, Klein 2011].
Exact details about the used parameters for bucketing and NMR signal exclusion are given explicitly for each project in the respective Materials and Methods parts in section 5.

4.3 Data analysis

4.3.1 Statistical data analysis

All statistical data analyses described in this section and in section 5 were carried out with the freely available statistical analysis software *R* [R Core Team 2014], if not stated elsewise. *R* is highly accepted in the biostatistics community and provides a large range of software packages for statistical computing and graphics [James et al. 2013].
A step-by-step *R*-Code for the statistical data analysis performed for this thesis has been developed and is given in a general form in Appendix I section 7.1. The following paragraphs will provide the statistical background and refer to the specific *R*-Code in Appendix I section 7.1.

4.3.1.1 Data normalization

As already highlighted in section 4.2.2.1, metabolic fingerprints derived from urine or plasma specimens exhibit, in general, data variance across the sample cohort from three major sources: technical variation, non-intended biological variation, and intended biological variation [Zacharias et al. 2013b), Maher et al. 2007, van den Berg et al. 2006].
Technical variations, e.g. differences in sample storage, were avoided or kept to a minimum in this thesis, as outlined in section 4.2.2.1.
Investigations of the intended biological variation, i.e. disease status of the biofluid donors, are the main objectives of this thesis, as described in section 3.2. Therefore, multivariate statistics were employed, which utilize the joint distribution of the data including the variance of individual features and their joint covariance structure [Zacharias et al. 2013b), Kohl et al. 2012].
Non-intended biological variation, e.g. differences in fluid intake across investigated urine specimens, comorbidities, etc., but also inevitable measurement noise need to be appropriately

addressed as they can obscure statistical analysis of the intended biological variation [Zacharias et al. 2013b), Kohl et al. 2012, van den Berg et al. 2006]. Moreover, several-order-of-magnitude differences between metabolite concentrations of a biological specimen can lead to erroneous inferences by statistical data analyses, as the high abundance of metabolites is not necessarily proportional to their biological importance [Kohl et al. 2012, van den Berg et al. 2006]. Additionally, technical and non-intended biological variation is usually heteroscedastic, leading to secondary data structures [Kohl et al. 2012, van den Berg et al. 2006]. To overcome these confounding effects, adequate scaling and normalization techniques can be applied, which are briefly introduced in this section.
The actual scaling and/or normalization method including the respective parameters employed for the individual projects are given in the respective Materials and Methods parts in section 5. A detailed mathematical description of the available normalization and scaling techniques was omitted, as it lies beyond the scope of this thesis. It can be found in [Kohl et al. 2012, van den Berg et al. 2006].

Differences in urinary metabolic concentrations across a sample cohort due to individual fluid intake are usually overcome by a scaling of the metabolite concentrations/bucket intensities of a sample to the respective creatinine concentration [Zacharias et al. 2013b), Kohl et al. 2012, Klein 2011]. As the creatinine excretion into urine is normally constant over time [Dörner 2013], this is also general practice in traditional clinical approaches [Dörner 2013].
For this thesis, this scaling method was usually performed during the bucketing procedure, compare to section 4.2.2.3, employing Amix VIEWER (latest version: Amix Viewer 3.9.13) (Bruker BioSpin GmbH, Rheinstetten, Germany). However, note that scaling of bucket intensities to urinary creatinine is only applicable if there exist no overall differences in creatinine concentrations between the intended biological groups.

Previously, various normalization methods for NMR-based metabolomic data had been systematically compared utilizing two different urinary data sets by [Kohl et al. 2012]. With regard to overall results for classification of specimens, bias reduction, and correct detection of fold changes, Quantile [Bolstad et al. 2003], Variance Stabilization [Huber et al. 2002], and Cubic-Spline [Workman et al. 2002] normalization performed best [Zacharias et al. 2013b), Kohl et al. 2012]. For this thesis, only Quantile and Variance Stabilization normalization were employed and are briefly discussed here.
Quantile normalization removes unwanted sample-to-sample variation of metabolites/bucket intensities by introducing an equal distribution of feature intensities across all spectra [Zacharias et al. 2013b), Kohl et al. 2012]. In brief, for each spectrum, the individual buckets are sorted according to their feature intensities in ascending order [Kohl et al. 2012]. Then, the mean value for each quantile, i.e., for example, the bucket with the highest feature intensity in one spectrum, is calculated across all spectra [Kohl et al. 2012]. Now, each bucket intensity is set to the corresponding mean value of its respective quantile [Kohl et al. 2012]. Finally, the individual buckets with new values of feature intensities are brought back to the original order of the respective spectrum [Kohl et al. 2012]. Consequently, each spectrum of the normalized data cohort now exhibits the same set of feature intensities, although individually distributed

across the features/buckets [Kohl et al. 2012]. In this thesis, Quantile normalization was carried out utilizing the *R* package *affy* [Gautier et al. 2004], and the corresponding *R*-code can be found in section 7.1.3.2.
Variance Stabilization normalization (VSN) performs an inverse hyperbolic sine transformation of the data, which has logarithmic character for large, and linear character for small values [Huber et al. 2002, Kohl et al. 2012]. For unnormalized data, the coefficient of variation, i.e. the variance divided by the corresponding mean, fairly stays constant for strong and medium feature intensities under the assumption that the standard deviation is proportional to the mean [Kohl et al. 2012]. However, for small feature intensities, the variance rather stays constant, resulting into an increasing coefficient of variation for decreasing feature intensities [Kohl et al. 2012]. Consequently, after applying the VSN, the adjusted variance of different metabolites/bucket intensities across the spectra becomes fairly homoscedastic [Kohl et al. 2012]. In this thesis, VSN was carried out utilizing the *R* package *vsn* [Huber et al. 2002] and the corresponding *R*-code can be found in section 7.1.3.3. Note that the *vsn* package, in addition, linearly maps all spectra to the first spectrum [Kohl et al. 2012].
It has to be noted that Quantile and VS normalization are only applicable if a relatively small proportion of metabolites/feature intensities is regulated in approximately equal shares up and down between the intended biological groups [Hochrein et al. 2015, Kohl et al. 2012].

4.3.1.2 Unsupervised statistical data analysis

[2] For unsupervised statistical data analysis, no information about underlying groups is used. Therefore, group separations observed are purely data-driven. This renders these approaches, in contrast to supervised statistical data analysis, as described in section 4.3.1.3, insensitive to overfitting in case of small sample numbers. Unsupervised statistical data analyses are often employed initially to check for group separation prior to classification of data, in cases where too few samples are available for classification with rigid cross-validation, or if the group identities are unknown.
The following paragraphs briefly introduce the unsupervised algorithms utilized in this thesis without a detailed mathematical description, which can be found in [Abdi and Williams 2010, Eisen et al. 1998].

Principal component analysis (PCA) is a widely used unsupervised approach for easy visualization of experimental data. In case of binned NMR spectra, the data of each spectrum can be considered as one point in a multidimensional space, with each bucket representing one dimension and the bucket intensity representing the value in that dimension. PCA performs a data transformation by defining a new coordinate system within this space. The newly defined dimensions are referred to as Principal Components (PCs). The first PC is aligned along the direction of maximum variance in the data. The second PC is chosen to be orthogonal to PC1 and to have maximum variance, again. According to this scheme, PCs are defined either until a fixed number of PCs is reached or until the PCs variance exceeds a certain amount of the total

[2] This section was published in [Zacharias et al. 2013b)] in a slightly altered version.

variance of all original dimensions, e.g. 95%. Mathematically speaking, the whole procedure is based on matrix diagonalization. When plotting the spectra in the reduced PC space, e.g. showing PC1 versus PC2, a considerable amount of the variance present in the data set is visualized allowing an easy inspection of the data such as the identification of distinct groups of samples or the detection of batch effects.
In this thesis, PCA was carried out with the *R* intern PCA algorithm [Venables and Ripley 2002, Mardia et al. 1979, Becker et al. 1988], and the package *missMDA* [Husson and Josse 2013]. The corresponding *R*-code can be found in section 7.1.4.1.

The general goal of clustering algorithms is to combine observations into groups or clusters based on a distance measure, by minimizing the distances within a cluster as compared to the distances between clusters. For this, both hierarchical and non-hierarchical algorithms are used.
Hierarchical clustering is an intriguingly simple method for finding similarities between spectra. All spectra are arranged in groups called clusters. At the beginning, each cluster contains exactly one spectrum. Using a distance matrix of pairwise distances, such as Euclidean distance, Manhattan distance, Pearson's correlation coefficient, or Spearman's correlation coefficient, between clusters, with bucket values serving as coordinates in a multidimensional space, similar clusters are merged to form a new, larger cluster. This procedure is repeated iteratively, i.e. a new distance matrix is calculated, the closest clusters are joined, and so on. In case a cluster contains more than one sample, an overall coordinate for the cluster has to be defined. In average linkage, for example, the average of all data of a cluster is used. The choice of distance measure and linkage type exerts a decisive effect on the final clustering result. In the end, all spectra are contained in one cluster. Taking all intermediate steps into account, a hierarchy of clusters has been created that can be visualized as a cluster dendrogram. This tree will reveal groups of similar spectra. Ideally, spectra from predefined groups (e.g. healthy and diseased groups) should end up in different clusters. This will only work if the inter-spectra differences are dominated by intended group differences rather than noise or other disturbing factors.
Clustering in this thesis was performed by choosing the option `cluster="TRUE"` of the function `geneImager` included in the *R*-package *compdiagTools* [Held et al. 2012], as outlined in Appendix I section 7.1.4.2. This conducts hierarchical clustering employing Euclidian distances and the average linkage method [Eisen et al. 1998].

4.3.1.3 Supervised statistical data analysis

Several principle aims of this thesis, as outlined in section 3.2, deal with the detection of metabolic differences between two patient groups with intended biological inter-group variance, e.g. healthy and diseased patients. To fulfill this major objective, supervised statistical data analysis is traditionally the method of choice [Zacharias et al. 2013b)]. It requires information about the class labels of the individual specimen groups [Zacharias et al. 2013b)].

For the detection of metabolic differences between two predefined patient groups, the Student's *t*-test was employed in this thesis. The investigated null hypothesis H_0 here represents

the case that no statistically significant difference between the means of two groups exists. To be more precise, for each feature/bucket b of the NMR data set, it is tested whether H_0, i.e. that there exists no difference between the two group means $\overline{X}_{b1}$ and $\overline{X}_{b2}$, is true or not [Zacharias et al. 2013b)]. The alternative hypothesis H_A represents the case that a statistically significant difference between these two group means exists.
These hypotheses can be tested by employing a Student's t-test under the assumption that the continuous data, here the NMR bucket intensities, is normally distributed [Zacharias et al. 2013b),Livingston 2004]. The normal distribution had been previously tested by employing the Kolmogorov-Smirnov test [Klein 2011]. By conducting a Student's t-test for a specific bucket b, the test statistic T_b is calculated in order to estimate how likely the two group means are different [Livingston 2004]

$$T_b = \frac{\overline{X}_{b1} - \overline{X}_{b2}}{s_b\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \tag{4.10}$$

s_b denotes the standard deviation under the assumption that both groups show the same standard deviation and n_1 and n_2 denote the size of the different groups, respectively, which are the degrees of freedom of the t statistic. If $|T_b|$ exceeds the value determined by the aimed statistical significance, H_0 is rejected [Livingston 2004]. By repeating the random drawing of two sample sets from a normal distribution and calculating the corresponding t-values, one obtains a normal t-value distribution described by the test statistic T_b under the assumption of a limited sample size [Livingston 2004].
The corresponding p-values denote the probability to observe a given or even more extreme t-value under the assumption that H_0 is valid [Livingston 2004]. This is termed statistical significance [Livingston 2004]. The p-values correspond to the area under the curve of the Student's t distribution from inf to the specific t-value or from $-\inf$ to $-t$, respectively [Livingston 2004]. Given that p-values are, by definition, uniformly distributed, p-values indicating a rejection of H_0 can also just be obtained by chance (type I error) [Casella and Berger 2002]. For a complete NMR data set, several hundred different t-tests are usually performed [Zacharias et al. 2013b)]. Consequently, a rejection of H_0 just by chance gets even more likely [Zacharias et al. 2013b)], which is termed a multiple testing problem. Consequently, the so-called α-value is defined as the probability of rejecting H_0 when it is true or as the probability of an observed difference resulting from chance alone (type I error) [Livingston and Cassidy 2005]. In this thesis, the multiple testing problem was overcome by applying the method of Benjamini and Hochberg (B/H) [Benjamini and Hochberg 1995]. This method controls the false discovery rate (FDR), i.e. the expected proportion of falsely rejected H_0, at a given significance level [Zacharias et al. 2013b), Benjamini and Hochberg 1995]. Here, α was, in agreement with general practice [Livingston and Cassidy 2005], chosen to be 0.05, i.e. for the group of all buckets with a p-value B/H adjusted below 0.05, there exists a less than 5% probability that observed differences between the two compared groups occur because of chance alone rather than because of a true difference between the groups (i.e. chance of detecting false positives) [Livingston and Cassidy 2005]. This corresponds to an FDR below 5% [Zacharias et al. 2013b)]. Mathematical details can be found in [Benjamini and Hochberg 1995].

The β-value is defined as the probability of accepting H_0 when it is false or as the probability of concluding that no difference exists when one is present (type II error) [Livingston and Cassidy 2005]. Hence, the statistical power $\pi = 1 - \beta$ is defined as the probability of rejecting H_0 when it is false or as the probability of detecting a statistically significant difference if one exists [Livingston and Cassidy 2005]. In other words, it is the probability that a true difference between two groups is correctly detected. It can be calculated post-hoc, i.e. after executing a Student's t-test, and depends on the actual sample sizes of the two groups n_1 and n_2, the magnitude of the measured effect, i.e. the size of the actual difference between the two groups, here $\overline{X}_{b1} - \overline{X}_{b2}$, the underlying variability of the outcome measurements of interest, i.e. the standard deviation s_b, and the α-value, here $\alpha = 0.05$ [Livingston and Cassidy 2005]. The magnitude of the measured effect and the corresponding standard deviation together are called the effect size [Livingston and Cassidy 2005]. Note that in clinical research, however, one usually determines the required sample size for detecting a certain effect size (which is usually estimated based on previous studies) prior to patient recruitment [Livingston and Cassidy 2005]. The required sample size, in turn, depends on the effect size, the α-value, and the desired statistical power π, which is, as a general agreement, usually set to $\pi = 0.8$ [Livingston and Cassidy 2005]. More details about the mathematical relationship between these parameters can be found in [Livingston and Cassidy 2005]. The statistical power π was either calculated employing the freely available software G*Power Version 3.1.7 [Faul et al. 2007], or with the *R* package *pwr* [Champely 2015], whereas the corresponding effect size was calculated with the *R* package *compute.es* [Del Re 2013]. The corresponding *R*-code can be found in Appendix I section 7.1.4.2.
In this thesis, t- and p-values for two groups were calculated using the *R* package *multtest*, which employs the so called Welch-test, a modification of the two sample t-test assuming unequal and unknown variances of the basic population [Pollard et al. 2005]. The corresponding *R*-code is given in section 7.1.4.2.
If more than two groups were compared, a one-way analysis of variance (one-way ANOVA) was conducted under the assumption of normally distributed continuous data [Zacharias et al. 2013b), Casella and Berger 2002]. The one-way ANOVA additionally assumes that the samples are independent, which implies that one sample can only belong to one group, and that the observations within each sample were acquired independently [Motulsky 1995]. The corresponding H_0 is that the specimens of two or more groups are members of the same basic population [Motulsky 1995]. Therefore, a one-way ANOVA tests whether all investigated groups have equal mean and variance, which is equivalent to an F-test [Motulsky 1995]. The result of a one-way ANOVA therefore indicates the probability that at least two groups of the data set have different means, but it does not reveal the varying groups. Therefore, additional two sample t-tests have to be employed for all possible group comparisons [Klein 2011]. The utilized *R* package *limma* [Smyth 2005] furthermore fits a linear model to the data, which describes the influence of each given group characteristic (a group characteristic can be, e.g., the treatment with a specific drug or a specific sample collection time-point) on the data [Zacharias et al. 2013b)]. The respective *R*-code can be found in section 7.1.4.2.
The discriminating features, identified by the two sample t-test, were displayed in heat-map representations, which offer a convenient possibility of illustrating their respective up- and

down-regulations in each investigated spectrum [Zacharias et al. 2013b)]. They were generated employing the *R* package *compdiagTools* [Held et al. 2012], and the corresponding *R*-code is outlined in section 7.1.4.2.
In the case of comparing categorical variables between two groups, Fisher's exact test for 2×2 contingency tables was employed for this thesis utilizing the *R* intern function "fisher.test" [Agresti 1990, Agresti 2002, Fisher 1935, Fisher 1962, Fisher 1970, Mehta and Patel 1986, Clarkson et al. 1993, Patefield 1981]. It assumes random sampling as well as both independent observations and samples [Motulsky 1995].

[3] Classification of an unknown sample to known classes of disease (e.g. healthy and diseased) based on discriminating features offers a convenient way of estimating the strength of a novel disease biomarker. Here, it was performed by employing supervised techniques from machine learning.

Thus far, approaches based on Partial Least Squares Discriminant Analysis (PLS-DA) [Barker and Rayens 2003], often in the combination with orthogonal projection to latent structures (OPLS-DA) [Trygg and Wold 2002], have dominated the classification of NMR metabolomic data. Dipl. Math. Jochen Hochrein compared the performance of PLS-DA to other classification approaches commonly used in genomics on various metabolite fingerprinting data including an AKI plasma data set [Hochrein et al. 2012]. For some classifiers consistently good performance was obtained independent of the data set in question. These classifiers included Random Forests (RF) [Hochrein et al. 2012, Breiman 2001] and Support Vector Machines (SVM) [Hochrein et al. 2012, Dudoit et al. 2002, Burges 1998]. The former are particularly suited for the analysis of high-dimensional NMR data.
A RF classifier consists of a set of tree predictors, where each tree is constructed from a different bootstrap sample of the training data. At each node of the tree the splitting in branches is based on a random selection of the input features. The final class label given to a new sample is the result of a majority vote over all trees. Another advantage of RFs is the provision of different measures of variable importance, which was used for the identification of predictive subsets of spectral features [Menze et al. 2009, Bryan et al. 2008]. A schematic representation for an RF classifier is given in Figure 4.7a). The internal parameters of the RF classifier are the number of grown decision trees n_{tree} and the number of tried variables m_{try}, which represents the number of variables employed by the RF classifier for the splitting at each node of a tree [Breiman 2001, Breiman 2002].
SVMs showed good performance on high- as well as low-dimensional data sets. SVM classifiers are so-called large margin classifiers, in which a separating hyperplane is determined in a way to maximize the distance between the individual classes of the training data. This hyperplane is constructed in a high dimensional vector space defined by the individual feature levels [Zacharias et al. 2013b), Hochrein et al. 2012]. Therefore, SVMs map the data to a higher-dimensional space by employing kernel-functions [Hochrein et al. 2012]. Both linear and radial basis function kernels were used in this thesis. Figure 4.7b) illustrates the main

[3] The following two paragraphs were published in [Zacharias et al. 2013b)] in a slightly altered version.

principle of a linear SVM trained on linearly separable data in a 2D space. For both linear and radial basis function kernels, the cost parameter C, which weighs the distance between outliers and the separating hyperplane and needs to be optimized in order to avoid over- and underfitting [Burges 1998], is optimized as an internal parameter of the SVM classifier in the cross-validation procedure. For a radial basis function kernel, the Kernel parameter γ, which is an indicator for the smearing out of the data points [Varma and Simon 2006], is additionally optimized.

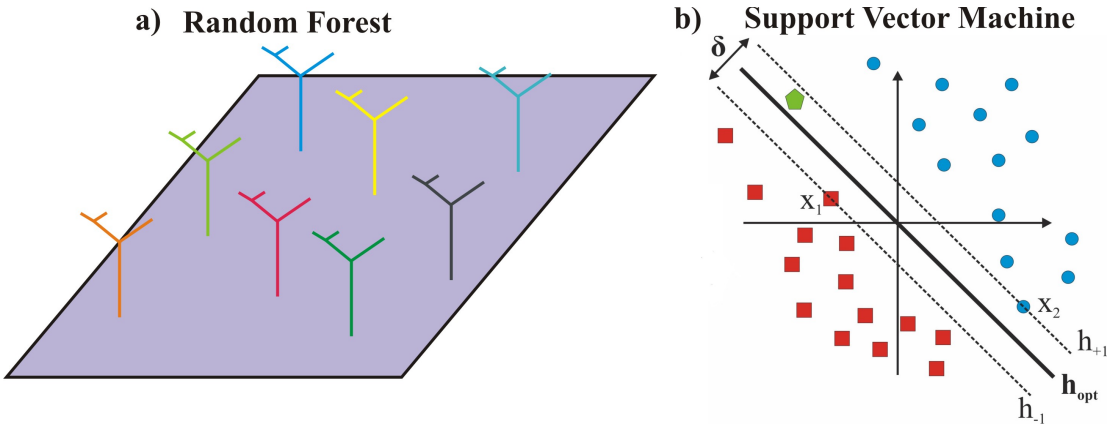


Figure 4.7: **Main principle of a RF and a linear SVM trained on linearly separable data in 2D space.** a) An RF consists of multiple decision trees, each of them independently grown during training [Hochrein et al. 2012]. The classification of a new data point with unknown group membership by the RF is based on a majority vote of all decision trees [Hochrein et al. 2012]. Modified from [Zacharias et al. 2013b)]. b) By training an SVM, an optimal hyperplane h_{opt} with maximal margin δ is generated, which is capable of separating two predefined groups (blue circles and red squares). A new data point/vector (green polygon) with unknown group membership can now be classified. δ is defined as the maximal possible distance to the two hyperplanes h_{-1} and h_{+1}. These hyperplanes h_{-1} and h_{+1} indicate the position of the data points (vectors) x_1 and x_2. As the position of h_{opt} only depends on the position of x_1 and x_2, they are called support vectors. Note that the cost parameter C, which weighs the distance between outliers and the separating hyperplane, is not considered in this case since only perfectly separable data is shown. Modified from [Zacharias 2012].

Generally, classifiers or classification algorithms are trained on a training data set where a class label for each sample is known, followed by an application of the trained algorithm on new test data. For performance evaluation, class labels of the test data also have to be known. In this thesis, performance evaluations were conducted within a cross-validation setting for small specimen cohorts (number of specimens $<$ number of features) in order to avoid overfitting, where the complete data set was iteratively split into training and test data [Varma and Simon 2006]. For larger specimen cohorts (number of specimens $>$ number of features), the complete data set was once split into a training and test set, comprising 2/3 and 1/3 of the complete set,

respectively, as recommended in [Lottaz et al. 2008]. It was ensued that no systematic differences exist between these two subcohorts. The employed classifier was subsequently trained on this specific training set and its predictive performance was assessed on the fixed test set.
Nested cross-validation schemes where parameters relevant for feature selection and the classification algorithm are optimized within the inner loops were employed for small specimen cohorts. This ensures that validation is not biased by training or parameter optimization. A schematic representation of a 3-fold nested cross-validation is given in Figure 4.8. Here, the upper bar represents the complete data set; it is split iteratively in training and test data (indicated by the green and orange bars, respectively). The training data of this loop is passed on to the middle loop, where it is again split in training and test data. The training data of the middle loop is transferred to the innermost loop of the nested cross-validation scheme where it is again split as described above. Here, parameters inherent to the classifier are tuned. In the middle loop, the sparsity of the classifier, e. g. the number of used features/buckets is optimized, while in the outmost loop validation is performed. It was ensured that all data of a specific loop are used once for testing. It has been previously shown that with a nested cross-validation approach an almost unbiased assessment of the true classification error is obtained [Varma and Simon 2006].

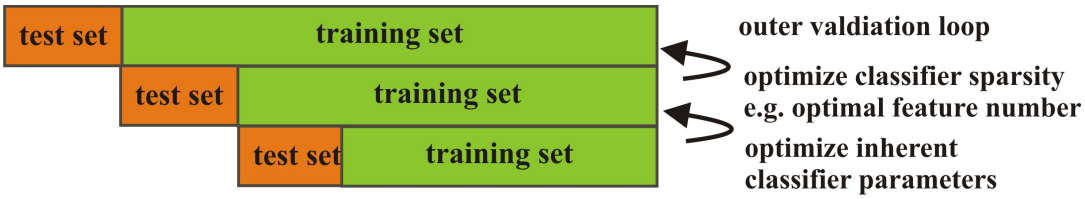


Figure 4.8: **General scheme of a three-fold nested cross-validation.** Each line, consisting of two bars, which represent test and training data set, respectively, represents one loop. The training and test data set of the upmost line comprises the complete data set. More details are given in the text. Modified from [Zacharias et al. 2013b)].

For this thesis, RF as well as SVM classification methods were combined with a t-score-based feature filtering approach [Zacharias et al. 2013b), Hochrein et al. 2012], as it performed best in [Hochrein et al. 2012].
For RF classification, the *R* package *randomForest* [Liaw and Wiener 2002], for SVM classification, the *R* package *e1071* [Dimitriadou et al. 2011] was employed. Classifier performances were evaluated utilizing the area under the respective receiver operating characteristic (ROC) curve, employing the *R* package *ROCR* [Sing et al. 2005]. The corresponding *R*-code of the cross-validation procedures with feature selection based on t-tests was implemented by Dipl. Math. Jochen Hochrein [Hochrein et al. 2012]. More details about the mathematical background of RFs and SVMs are given in [Burges 1998, Breiman 2001]. The specific parameters for the classification procedures conducted for the respective projects are detailed in the particular Materials and Methods parts in section 5.

Correlation calculations were performed in order to estimate the linear dependence between two variables x and y [Merziger et al. 2004]. In this thesis, Pearson's correlation coefficient r was calculated according to the following definition [Motulsky 1995]

$$r = \frac{\sum_{i=1}^{n} \left[\frac{(x_i - \overline{x})}{s_x}\right] \cdot \left[\frac{(y_i - \overline{y})}{s_y}\right]}{n-1}, \tag{4.11}$$

where s denotes the standard deviation and n the number of data points [Motulsky 1995]. It assumes that x and y values were sampled from populations which follow a Gaussian distribution [Motulsky 1995]. r can attain values over the interval $[-1, +1]$, where $r = -1$ and $r = 1$ indicate perfect anti-correlation or correlation of two variables, respectively [Motulsky 1995]. If $r = 0$, no correlation exists between the compared variables [Motulsky 1995]. The square of r is called the coefficient of determination R^2, and is a measure for the proportion of variance shared between the two variables [Motulsky 1995]. The R intern function "cor" [Becker et al. 1988] was employed.

The principle of regression analysis can be summarized as the description of a dependent variable by one or more explanatory variables. Therefore, a so-called regression function or model is computed, which can be used to predict dependent variables based on the corresponding explanatory variables [Dalgaard 2008]. In this thesis, simple as well as multiple linear regression analyses were performed. A number of explanatory or predictor variables x_j are employed to fit a linear model

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon, \tag{4.12}$$

where y_i is the dependent or response variable, β_0 is the intercept, β_j the so-called regression coefficients, and ϵ the disturbance parameters representing random variability, which are assumed to be independent and $N(0, \sigma^2)$ [Dalgaard 2008, Hastie et al. 2001, Motulsky 1995]. β_0, β_j, and σ are estimated on a training data set $(x_1, y_1), ..., (x_N, y_N)$, where each $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ is a vector of feature intensities for the i^{th} case [Hastie et al. 2001]. This is conducted by employing the method of least squares, i.e. minimizing the residual sum of squares (RSS) [Hastie et al. 2001]

$$RSS = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2. \tag{4.13}$$

The derived linear regression model or equation can now be employed to predict new dependent variables in an independent test set [James et al. 2013, Hastie et al. 2001]. By employing the method of least squares, however, two major problems can occur [James et al. 2013]. First, if the number of explanatory variables exceeds the number of dependent variables, the regression model derived by the method of least squares will perfectly describe the dependent variables of the training set [James et al. 2013]. However, this model is highly prone to overfitting and will therefore probably not yield satisfying predictions for an independent test set [James et al. 2013]. Second, the interpretation of such multiple regression models is not straightforward,

because they often include variables without any association with the dependent responses [James et al. 2013]. Especially in the case of NMR metabolomics data, where very large numbers of explanatory variables, i.e. numbers of features/buckets, are present, often exceeding the number of samples used for model fitting, the problems of overfitting and data interpretation become eminent.
These drawbacks can be overcome by various methods, including subset selection and regression coefficient shrinkage [James et al. 2013]. A prominent example for subset selection in multiple regression analysis is the least-angle regression (LARS) [Efron et al. 2004].
The two most popular regression coefficient shrinkage methods are ridge regression [Hoerl and Kennard 1970] and least absolute shrinkage and selection operator (LASSO) regression [Tibshirani 1996]. An upper bound t or s, respectively, is set for the l_1 or l_2 norm of the regression coefficients β_j in LASSO or ridge regression, respectively [James et al. 2013, Hastie et al. 2001]. The minimization problems can be formulated as [James et al. 2013, Hastie et al. 2001]

$$\widehat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \quad \text{subject to } \sum_{j=1}^{p} |\beta_j| \leq t, \text{or} \tag{4.14}$$

$$\widehat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \quad \text{subject to } \sum_{j=1}^{p} \beta_j^2 \leq s. \tag{4.15}$$

These minimization problems can be equivalently formulated following the Lagrangian formalism [James et al. 2013, Hastie et al. 2001]

$$\widehat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \text{or} \tag{4.16}$$

$$\widehat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}. \tag{4.17}$$

λ is the so-called tuning parameter [James et al. 2013], which is usually minimized in an internal cross-validation procedure. In the case of LASSO regression, some regression coefficient estimates $\widehat{\beta}^{lasso}$ can become exactly equal to zero, which leads to the exclusion of the corresponding explanatory variable from the regression model [James et al. 2013, Hastie et al. 2001]. In comparison, regression coefficient estimates $\widehat{\beta}^{ridge}$ are only shrunk towards zero and all explanatory variables are always included in the regression model [James et al. 2013, Hastie et al. 2001]. Note that a simple modification of the LARS algorithm implements the entire solutions of the LASSO [Efron et al. 2004].
In this thesis, regression models were trained on an exclusive training set and then evaluated in an independent test set. Both the coefficient of determination R^2 between true and predicted dependent variables as well as the mean-squared errors (*mse*) on training and test data are reported.
In this thesis, the LASSO algorithm was applied using the *R*-package *glmnet* [Friedman et al. 2010]. The corresponding *R*-code is given in section 7.1.4.3. For simple regression analysis, a linear model was fitted using the *R* function "lm" [Chambers 1992, Wilkinson and Rogers

1973]. More details about the mathematical background of regression analysis can be found in [James et al. 2013, Dalgaard 2008, Efron et al. 2004, Casella and Berger 2002, Hastie et al. 2001, Motulsky 1995, Tibshirani 1996, Hoerl and Kennard 1970].

4.3.2 Metabolite identification

[4] The assignment of features in NMR spectra of biofluids to specific metabolites can be a laborious task that is often complicated by massive signal overlap present in 1D ^{1}H spectra. This is especially true for the typically crowded region between 4.0 - 3.0 ppm.
As described in section 4.2.2.2, overlapping 1D signals may be resolved in a second dimension. Consequently, NMR signal assignment in 1D spectra conducted in this thesis was verified by corresponding 2D spectra. In case of blood samples, broad NMR peaks arising from proteins or other macromolecules would affect peak discrimination. Therefore, proteins were either removed by ultrafiltration, compare to section 4.2.2.1, or suppressed by employing the CPMG pulse sequence, as illustrated in section 4.2.2.2.
Initial assignment of distinct NMR peaks, which discriminate between two investigated groups according to a Student's *t*-test (compare to section 4.3.1.3), to metabolites was usually performed on representative 1D ^{1}H and corresponding high-resolution 2D ^{1}H-^{13}C HSQC spectra. Signals were manually identified by comparison with reference spectra of pure compounds measured ideally under the same experimental conditions. These reference spectra were downloaded from the commercially available Bruker Biofluid Reference Compound Database BBIOREFCODE that includes a large amount of reference spectra of currently almost 600 mostly naturally occurring metabolites acquired under various experimental conditions (e.g. different pH-values, solvents, etc.). The NMR analysis software suite AMIX-Viewer (latest version: Amix Viewer 3.9.13) (BrukerBioSpin GmbH, Rheinstetten, Germany) provided the interface for directly comparing acquired spectra with reference spectra from the BBIOREFCODE database. By manually overlaying reference spectra with actual NMR spectra, a considerable number of resonances were assigned. Despite the considerable number of reference spectra of metabolites and pharmaceuticals stored in the BBIOREFCODE database, coverage is far from complete in comparison to MS-based databases such as NIST [Linstrom and Mallard 2016]. Therefore, if no clear assignment of an NMR peak to a metabolite was possible, reference spectra of pure compounds, which potentially could be identified in the spectra, were acquired under conditions similar to the biofluid spectra, and subsequently manually overlaid with and compared to the biofluid NMR spectra. Additionally acquired 2D spectra further assisted in metabolite identification as well as the concomitant use of other analytical methods such as high-performing liquid chromatography (HPLC) and MS.
More details about the actual metabolite assignment procedure are given in the respective Materials and Methods parts in section 5.

[4] This section was published in [Zacharias et al. 2013b)] in a slightly altered version.

4.3.3 Metabolite quantification

Absolute quantification of metabolites and statistical data analysis based on the absolute concentrations can be regarded as a targeted profiling approach [Zacharias et al. 2013b)]. For a correct quantitative analysis of NMR spectra, two issues need to be addressed [Zacharias et al. 2013b)].
First, the complexity of biofluids, especially urine, with hundreds to thousands of different endo- and exogenous metabolites [Holmes et al. 1997] leads to massive signal overlap especially in 1D ^{1}H spectra [Zacharias et al. 2013b)]. This can induce over-quantification of compounds, whose signals are located in crowded regions, e.g. between 4.0 ppm and 3.0 ppm. A manual inspection of each signal used for quantification, as performed in this thesis, minimizes this problem in combination with the fact that only NMR signals not overlapping with other signals were used [Zacharias et al. 2013b)]. Moreover, quantification based on 2D NMR spectra was applied whenever possible, as the introduction of a second dimension significantly reduces signal overlap [Zacharias et al. 2013b), Gronwald et al. 2008].
The second issue for NMR quantification describes the fact that, in a given spectrum, two signals of equal intensity do not necessarily imply equal concentration values [Zacharias et al. 2013b), Klein et al. 2013]. Therefore, Dr. Matthias Klein developed the freely available software "MetaboQuant" [Klein et al. 2013, Klein 2011], which offers a tool for automatically calculating accurate metabolite concentrations from 1D and 2D NMR signal intensities employing individual calibration factors and different outlier detection algorithms [Klein et al. 2013, Zacharias et al. 2013b), Klein 2011]. For each NMR signal, which was used to determine the absolute concentration of a compound, individual calibration factors had been either experimentally determined by Dr. Matthias Klein and Prof. Dr. Wolfram Gronwald [Zacharias et al. 2013b), Klein 2011, Gronwald et al. 2008] or were experimentally determined in this thesis, as explicitly described in the respective Materials and Methods parts in section 5.
In this thesis, peak picking, fitting, and integration of 1D and 2D NMR signals were performed with the Analytic Profiler of AMIX-Viewer (latest version: Amix Viewer 3.9.13) (Bruker BioSpin GmbH, Rheinstetten, Germany) employing 1D ^{1}H NOESY, as well as 2D ^{1}H-^{13}C HSQC spectra. The Analytic Profiler compares the investigated spectrum to a real reference spectrum and picks the best matching peak in a specified ^{1}H range. This ^{1}H or ^{13}C spectral range had been either previously determined by Dr. Matthias Klein and Prof. Dr. Wolfram Gronwald or was determined throughout this thesis for the respective sample matrices. Afterwards, the peak integral was calculated and the resulting integral was reported, if not stated otherwise, relative to the known amount of the internal reference substance TSP [Zacharias et al. 2013b)]. Subsequently, the absolute concentrations were calculated employing "MetaboQuant" [Klein et al. 2013, Klein 2011].

5 Biomedical Applications

5.1 Acute Kidney Injury study

5.1.1 Introduction

The first and major aim of this thesis is the detection of novel metabolic biomarkers in the context of renal diseases. This objective was pursued in an NMR based study of acute kidney injury after cardiac surgery in collaboration with the University Clinic of Erlangen-Nuremberg. Small molecule markers for early diagnosis and prediction of AKI were studied in both urine and plasma specimens and both studies have already been published in [Zacharias et al. 2013a), Zacharias 2012] and [Zacharias et al. 2015]. I performed parts of the urine analyses in the context of a master thesis at the Institute of Functional Genomics [Zacharias 2012]. The corresponding results are briefly summarized in this section. This study resulted in a peer-reviewed publication [Zacharias et al. 2013a)], which was written and published during my Ph.D. time.

Acute kidney injury has already been elaborately described as a frequent and severe complication after cardiac surgery in section 3.1. Its classification and staging based on increases in SCr levels or reduction in UO is discussed in section 4.1.2 and 4.1.3. Their drawbacks for early detection of AKI after cardiac surgery including relatively late rise of SCr after renal injury and affection by non-renal factors are explicitly outlined in section 4.1.2. Nevertheless, the demand for early detection of AKI is eminent for early intervention and improved patient care, compare to section 3.1. Consequently, the identification of urinary and/or serum biomarkers for the early prognostication of AKI after cardiac surgery has become a prominent field in nephrology [Mariscalco et al. 2011, Parikh et al. 2011, Endre et al. 2011, Haase et al. 2010a), Haase et al. 2010b), Haase et al. 2009, Han et al. 2002, Westhuyzen et al. 2003, Parikh et al. 2005, Han et al. 2009, Lameire et al. 2011]. One of the most promising biomarkers in urine and serum was reported as NGAL, which predicted AKI as early as two hours after surgery in a cohort of pediatric cardiac surgery patients with area under the curve of the receiver operating characteristic (AUC-ROC) values above 0.90 [Haase et al. 2010a)]. Nevertheless, the individual or combined performance of NGAL and other novel protein biomarkers including serum CysC, γ-glutamyltranspeptidase (GGT), alkaline phosphatase (AP), kidney injury molecule-1 (KIM-1), and interleukin-18 (IL-18) in adult patients was less powerful with AUC-ROC values around 0.80 [Kidher et al. 2014, Haase-Fielitz et al. 2009, Wagener et al. 2006].

During my master thesis, I investigated urine specimens from 106 patients that had undergone cardiac surgery with cardiopulmonary bypass (CPB) use. Every patient had donated one

urine specimen each before, as well as at 4 and 24 h after surgery. Thirty-four study participants had been diagnosed with post-operative AKI, compare to section 5.1.2.1. In total, 318 1D ^{1}H NOESY urinary spectra, which had been scaled to the reference region of creatinine, and subsequently Quantile normalized, as described in section 4.3.1.1, had been investigated. Since urine specimens collected at three different time-points were available, I conducted a time-course evaluation for these specimens by means of a PCA taking all NMR features into account. The corresponding PCA plot including all urine specimens is shown in Appendix II section 7.2.4 Figure 7.1a). It revealed a clear separation of urine specimens collected before and at 4 h after surgery, whereas the specimens collected 24 h after surgery fall between the earlier time-points. Further analysis of the corresponding loadings (data not shown) indicated that spectral differences were mainly due to the presence of D-mannitol in urine collected at 4 h after surgery, presumably caused by the pre-filling of the tubes of the CPB machine with 500 ml of D-mannitol solution, as described in the CPB protocol in Appendix II section 7.2.2. This up-regulation of D-mannitol at 4 h was also apparent by visual inspection of representative urinary 1D ^{1}H spectra collected at 0 h pre-op, 4 h post-op and 24 h post-op, respectively, as depicted in Appendix II section 7.2.4 Figure 7.1c).
For each of the three urine collection time-points, I conducted classification analyses of the 34 AKI patients versus the 72 non-AKI patients using an SVM with radial basis function kernel on the 1D ^{1}H NMR data as described in section 4.3.1.3. Results were averaged over five cross-validation runs where each run started from a different random splitting of test and training data. For urine specimens collected before surgery, no satisfactory classification of patients with or without AKI could be obtained. For urine specimens collected at 4 h after surgery, a group separation was achieved with an overall prediction accuracy of 72.2 $\pm$ 2.8 % and a corresponding AUC-ROC of 0.79 $\pm$ 0.02. To obtain these results, on average, 47.4 $\pm$ 2.7 features were employed, most of which could not be assigned unambiguously to a specific solute. Exceptions were hippuric acid and 4-hydroxyhippuric acid. Hippuric acid is a conjugate of glycine and benzoic acid, which is eliminated by active tubular secretion [Geng and Pang 1999]. Benzoic acid originates mostly from the gut microbial catabolism of dietary polyphenols contained in fruits, vegetables, wine, tea, extra virgin olive oil, chocolate and other cocoa products [Selma et al. 2009]. Benzoic acid is also a common component of plasticizers, and it is added to pharmaceuticals, foods, beverages, and cleaning agents because of its anti-microbial and anti-fungal properties. The simultaneous increase in excretion of 4-hydroxyhippuric acid, the glycine conjugate of 4-hydroxybenzoic acid, which also mostly originates from dietary intake and is a major metabolite of parabens commonly found in pharmaceutical, cosmetic and nutritional products [Harvey and Everett 2004], makes it the more likely that the increased urinary levels of hippuric acid in AKI patients are due to the delayed elimination of exogenous benzoic acid in the proximal tubule, which is particularly prone to ischemia/reperfusion injury following cardiac surgery with CPB use. However, it cannot be ruled out entirely that other factors account for or contribute to the increased urinary levels of hippuric acid observed in AKI patients 4 h after surgery. Cardiac surgery with CPB use, as well as AKI may be accompanied by metabolic acidosis. Indeed, a study conducted on healthy human volunteers has found that acidification increases both synthesis of hippuric acid in liver and kidney and its subsequent excretion [Dzúrik et al. 2001].

Results of the classification analysis for the 24 h urine specimens showed an improved predictive performance with an overall accuracy of 76.0 ± 1.9 % and a corresponding AUC-ROC of 0.83 ± 0.02. These results were accomplished with an average number of 2.4 ± 0.5 features. The overall sensitivity and specificity amounted to 57.1 ± 3.7 and 88.6 ± 1.2 %, respectively. Data indicate a reliable prediction of non-AKI patients, whereas the prediction accuracy of AKI patients strongly depended on the final stage of the disease. In fact, the worse the AKIN-staging of the patient the better the prediction accuracy of the trained classifier became. Additionally performed permutation tests [Mukherjee et al. 2003, Zacharias 2012], as described in section 5.1.2.7, proved that the observed classification accuracies based on 24 h urinary NMR fingerprints had not been obtained by chance.
Moreover, spectral differences distinguishing AKI and non-AKI urinary NMR fingerprints at 24 h after surgery were investigated according to a Welch-test as described in section 4.3.1.3. The three most significant features were used by the SVM classification algorithm and included carnitine ($P_{adj} = 5.0e^{-8}$), a feature representing at least in part 2-oxoglutaric acid ($P_{adj} = 5.0e^{-8}$), and tranexamic acid ($P_{adj} = 9.2e^{-6}$). Tranexamic acid, which is a synthetic derivative of lysine, had been administered to approximately 96 % of all patients enrolled at the time of operation as an antifibrinolytic agent. On average, 1,489 ± 710 and 1,450 ± 546 mg of tranexamic acid were administered to patients with and without AKI, respectively ($P = 0.76$), compare to the CPB protocol given in Appendix II section 7.2.2. To investigate whether the classification results obtained were critically dependent on the differential excretion of tranexamic acid, I repeated tests after exclusion of all spectral regions corresponding to tranexamic acid. Results revealed an average prediction accuracy of 78.1 % and an area under the ROC curve of 0.84, indicating that the prediction accuracy does not diminish upon exclusion of tranexamic acid from analysis.
An up-regulation of tranexamic acid in urine specimens collected at 24 h from AKI patients in comparison to non-AKI patients appears to indicate reduced glomerular filtration in these patients, as tranexamic acid is eliminated by glomerular filtration with neither tubular secretion nor adsorption taking place [Eriksson et al. 1974]. This delayed excretion of exogenous compounds is further reflected in the PCA shown in Appendix II section 7.2.4 Figure 7.1a), where at 24 h the urinary specimens of the non-AKI patients (marked in red) are located closer to the specimens collected preoperatively than the specimens of the AKI patients (marked in orange).
Aside from tranexamic acid, the prediction of AKI at 24 h after surgery rested mainly on carnitine. The main function of carnitine is the transport of long-chain fatty acids into the mitochondria for subsequent beta-oxidation [Arduini et al. 2008]. In addition, carnitine is also used to transport peroxisomal β-oxidation products to the mitochondria, to export accumulating acyl-groups, and to modulate the level of free coenzyme A in different subcellular compartments. In mammals, carnitine homeostasis is maintained by endogenous synthesis from the amino acids lysine and methionine, absorption from dietary sources, and efficient (> 95%) renal tubular reabsorption [Vaz and Wanders 2002,Lohninger et al. 2005]. Hence, this finding of increased urinary concentrations of free carnitine in the non-AKI group did come as a surprise, as it appears to indicate reduced tubular reabsorption by the high-affinity sodium dependent carnitine cotransporter OCTN2 (SLC22A5), which is expressed in the brush-border membrane

of the proximal tubule. Damage of the proximal tubule is a hallmark of ischemic kidney injury. However, if renal tubular dysfunctions were the cause of increased urinary levels of free carnitine, one would expect to observe markedly higher levels in the AKI group, in which carnitine levels remained interestingly near physiological levels [Ciba-Geigy 1983] with an average value of 0.040 ± 0.073 mmol/mmol$_{\text{crea}}$, while the average value in the non-AKI group [0.083 ± 0.099 mmol/mmol$_{\text{crea}}$] was significantly higher ($P = 0.014$). In comparison, in the specimens collected before and at 4 h after surgery average urinary carnitine levels of <0.020 mmol/mmol$_{\text{crea}}$ were obtained for both groups, which is in line with values reported in the literature for healthy subjects. Other reasons for reduced renal carnitine absorption include reduced urinary sodium levels or direct inhibition of OCTN2 by drugs such as β-lactam antibiotics and the calcium channel blocker verapamil [Diao et al. 2010, Ganapathy et al. 2000]. However, there were no differences in medication administered to AKI and non-AKI patients. Animal studies of AKI have reported conflicting changes in carnitine levels: In a rat model of ischemia/reperfusion-induced AKI, reduced serum carnitine levels were observed [Liu et al. 2012], while a rat model of gentamicin-induced AKI yielded increased levels of urinary carnitine [Al-Shabanah et al. 2010]. Interestingly, the treatment of rats subjected to renal ischemia/reperfusion injury with carnitine has been shown to improve energy metabolism as evidenced by increased tissue ATP levels in comparison to untreated animals and to reduce tissue damage [Idrovo et al. 2012]. Hence, the observation of increased urinary levels of carnitine in the non-AKI group may represent a successful protective response against ischemic injury, whereby carnitine helps to restore fatty acid oxidation and facilitates the export of the accumulated long-chain acylcarnitines. Support for this hypothesis comes from the reported significant increase in urinary levels of hepcidin, a small cysteine-rich peptide that plays an important role in iron homeostasis, as early as 6 h after cardiac surgery in patients not developing AKI, whereas hepcidin levels remained unchanged for AKI patients [Haase-Fielitz et al. 2011]. Increased renal expression of hepcidin as an adaptive response to limit ischemia and reperfusion injury caused by CPB was discussed as a possible mechanism to explain its increased excretion in non-AKI patients.
More detailed information about NMR data acquisition, preprocessing, and statistical analysis of the urinary spectral data can be found in [Zacharias et al. 2013a), Zacharias 2012]. An excessive discussion of the presented results can be found in [Zacharias et al. 2013a)].

In the last months of my master thesis, the University Clinic of Erlangen-Nuremberg provided us with plasma specimens obtained from the same patient cohort. Eight of these plasma specimens had been, in the context of my master thesis, prepared for and measured by NMR spectroscopy [Zacharias 2012]. The remaining plasma specimens were prepared for and measured by NMR spectroscopy at the beginning of my Ph.D. thesis.
Encouraged by the promising prognostic performance of the urinary biomarkers for AKI investigated in my master thesis [Zacharias 2012], I extended the investigation of metabolic alterations in the context of AKI after cardiac surgery on plasma specimens obtained from the same patient cohort in my Ph.D. thesis. Moreover, I complemented the NMR study based on urine specimens with additional statistical data analyses in order to respond to suggestions from peer reviewers during the publication procedures for [Zacharias et al. 2013a)]. Consequently, this section mainly focuses on the investigation of plasma specimens by means of NMR spectroscopy

in the context of AKI after cardiac surgery, but also presents novel results for the urine specimen study as performed for my master thesis.
Parts of the plasma analyses were performed by M.Sc. Franziska Vogl. Moreover, several important method developments and improvements were conducted for this study jointly together with Dipl. Math. Jochen Hochrein. They have already been published in [Hochrein et al. 2012] and [Hochrein et al. 2015], and are also part of the Ph.D. thesis of Dipl. Math. Jochen Hochrein [Hochrein 2016].

5.1.2 Materials and Methods

5.1.2.1 Patient selection and sample collection

[1] In total, 106 patients undergoing cardiac surgery with CPB use at the University Clinic of Erlangen-Nuremberg from July 2009 to August 2010 were included in this study. Operative procedures included coronary artery bypass grafting (CABG), aortic and/or mitral valve surgery (replacement and repair), combinations of CABG and heart valve surgery, and thoracic aortic surgery. The CPB protocol is given in Appendix II section 7.2.2. For the determination of SCr levels, serum specimens were collected for each patient on the day before surgery and daily thereafter at 6:00 am. For patient classification according to the AKIN-criteria, SCr levels until the second day after surgery were taken into account. Of the 106 patients, 34 were diagnosed with AKI following surgery. In all five patients that required post-operative RRT, treatment was initiated more than 24 h after surgery. Hence, RRT did not affect metabolite levels in the simultaneously collected urine and plasma specimens. Detailed clinical characteristics, administered medication and outcome are given in Table 5.1 and Appendix II section 7.2.1 Table 7.1. Written declarations of consent had been obtained from all study participants before inclusion. Spot urine samples were collected on the day before surgery, and at 4 and 24 h after surgery. Urine was centrifuged at 1,500 rpm for 5 min and the clear supernatant was immediately frozen and stored at -80°C until NMR analysis. At 24 h, an additional EDTA plasma specimen was collected from each patient and stored at -80°C. The daily collected serum specimens were not available for NMR spectroscopy.
For the plasma study, a subcohort of 85 patients was included for whom enough EDTA plasma was available. Detailed information on clinical characteristics, administered medication and outcome for this subcohort is given in Table 5.1b) and Appendix II section 7.2.1 Table 7.1b). In total, 33 patients out of these 85 were diagnosed with post-operative AKI 48 h and 72 h after cardiac surgery, compare to Table 5.1b). While 32 patients had reached the same stage of AKI already after 48 h, one patient, being classified as AKIN 1 48 h after surgery was re-classified as AKIN 3 72 h after surgery due to a dramatic increase in SCr on the third post-operative day and was persistently classified as AKIN 3 for the purpose of the plasma analysis. This difference in AKI-classification at 48 and 72 h after surgery does play a substantial role in the analysis of intermediate AKI cases, i.e. patients clinically classified as AKIN-stage 1 patients, as described in section 5.1.2.9. In fact, the different clinically staging at 48 and 72 h after

[1] The following section has already been published in [Zacharias et al. 2013a)] and [Zacharias et al. 2015] in a slightly altered version.

surgery of this particular AKI patient was first noticeable during the analysis of intermediate AKI cases based on plasma specimens, whereas it did not fall into account during the analysis of the corresponding urinary specimens for my master thesis [Zacharias et al. 2013a), Zacharias 2012]. Since the AKI-staging for the urine analysis was only based on the clinical staging 48 h after surgery, this patient had been persistently classified as AKIN 1 in that study [Zacharias et al. 2013a), Zacharias 2012].

AKIN-stage	**0**	**1**	**2**	**3**
a) Number of patients for urine study	72	26	3	5
b) Number of patients for plasma study	52	24	3	6[a]

Table 5.1: **Classification by AKIN-criteria.** a) Number of patients included in AKI urine study [Zacharias et al. 2013a), Zacharias 2012]. AKI patients were diagnosed based on AKIN-criteria, as elaborately described in section 4.1.3. Here, serum samples collected until 6:00 am on the second post-operative day were taken into account [Zacharias et al. 2013a), Zacharias 2012]. b) Number of patients included in AKI plasma study [Zacharias et al. 2015]. [a]Here, 33 patients out of 85 were diagnosed with post-operative AKI 48 h and 72 h after cardiac surgery. While 32 patients had reached the same stage of AKI already after 48 h, one patient, who had been originally classified as AKIN 1 48 h after surgery for the AKI urine study [Zacharias et al. 2013a), Zacharias 2012], was re-classified as AKIN 3 72 h after the surgery [Zacharias et al. 2015]. This re-classification reflects the dramatic increase of SCr on the third post-operative day for this patient. Modified from [Zacharias et al. 2013a)].

5.1.2.2 NMR spectroscopy

Urine specimens were prepared for NMR measurements as described in section 4.2.2.1 [Zacharias et al. 2013a)], EDTA plasma specimens were ultrafiltrated with a cut-off of 10kD and subsequently prepared for NMR measurements as described in section 4.2.2.1 [Zacharias et al. 2013a), Zacharias et al. 2015]. Note that 0 h and 4 h urine specimens had been prepared by Caridad Louis, whereas I prepared 24 h urine and a subset of eight plasma specimens in the context of my master thesis [Zacharias 2012]. I explicitly prepared all other EDTA plasma specimens for this thesis.
1D ^{1}H NOESY as well as 2D ^{1}H-^{13}C HSQC spectra were measured for all specimens according to the standard protocols described in section 4.2.2. All urine as well as eight plasma specimens had been measured in the context of my master thesis [Zacharias 2012], whereas the remaining plasma specimens were measured during my Ph.D. thesis. For each 1D ^{1}H NOESY spectrum, 128 scans were collected into 65536 data points employing the pulse program *noesygppr1d.comp* (BrukerBioSpin GmbH, Rheinstetten, Germany) with water suppression by presaturation during relaxation and mixing. Four dummy scans were acquired prior to measurement, the spectral width was 20.55 ppm, the relaxation delay was 4 s, the acquisition time amounted to 2.66 s, and the mixing time to 0.01 s, respectively. Each 2D ^{1}H-^{13}C HSQC spectrum was acquired

employing the pulse program *r_hsqcetgppr* (BrukerBioSpin GmbH, Rheinstetten, Germany) with water suppression by presaturation during the relaxation delay. 2048 × 128 data points were collected using 8 scans per increment, an acquisition time of 0.14 s, a relaxation delay of 3 s, and 16 dummy scans, resulting into a total acquisition time of less than one hour. The spectral widths were 12.01 ppm in the ^{1}H, and 165.01 ppm in the ^{13}C direction, respectively. One representative high-resolution 2D ^{1}H-^{13}C HSQC spectrum from a non-AKI plasma specimen was acquired with 2048 × 512 data points using 40 scans per increment, and a spectral width of 12.01 ppm in the ^{1}H, and 200.00 ppm in the ^{13}C direction, respectively, in the context of my Ph.D. thesis. For the same plasma specimen, one 2D ^{1}H ^{1}H TOCSY spectrum was acquired during my Ph.D. thesis using the pulse program *mlevgpphw5* (BrukerBioSpin GmbH, Rheinstetten, Germany) with 2048 × 512 data points, 16 scans per increment, an acquisition time of 0.14 s, 32 dummy scans, a relaxation delay of 3 s, a mixing time of 60 ms, and a spectral width of 12.07 ppm. The total acquisition times of the high-resolution 2D ^{1}H-^{13}C HSQC and 2D ^{1}H ^{1}H TOCSY spectrum amounted to approximately 19.6 h and 13 h, respectively. Furthermore, a 2D ^{1}H-^{13}C HMBC spectrum was acquired for this plasma specimen during my Ph.D. thesis with 2048 × 1024 data points using 32 scans per increment. Spectral widths amounted to 13.02 ppm in the ^{1}H, and 200.00 ppm in the ^{13}C direction, respectively. The pulse program *hmbcgplpndprqf* (BrukerBioSpin GmbH, Rheinstetten, Germany) was employed with 16 dummy scans, an acquisition time of 0.13 s, a relaxation delay of 1.5 s, and spectral widths of 13.02 ppm in the ^{1}H, and 200.00 ppm in the ^{13}C direction, respectively. The complete measurement time of the 2D ^{1}H-^{13}C HMBC spectrum amounted to approximately 14 h.

5.1.2.3 Mass spectrometry

M.Sc. Franziska Vogl diluted ultrafiltered plasma samples of five AKI and five non-AKI patients each with deionized water (1:4) [Zacharias et al. 2015]. She performed metabolic fingerprinting by means of high-resolution LC-QTOF-MS as previously described [Dettmer et al. 2013]. In brief, a Thermo Scientific Dionex Ultimate 3000 UHPLC system (Idstein, Germany) coupled to a Maxis Impact QTOF-MS (Bruker Daltonics, Bremen Germany) equiped with an electrospray ionization (ESI) source was employed. She used a Kinetex™ (Phenomenex, Aschaffenburg, Germany) 2.6 µm C18 100 × 2.1 mm id column at 25°C utilizing 0.1% formic acid in (i) water and in (ii) acetonitrile as mobile phase with a flow-rate of 0.3 ml/min. For elution, an acetonitrile gradient of 0 - 40 % in 10 min, 40 - 95 % in 2 min, back to 0 % in 0.1 min, followed by equilibration for 5 min was employed. M.Sc. Franziska Vogl operated the ESI source in separate runs in both positive and negative mode. She set the source temperature and flow rate of the drying nitrogen gas to 220°C and 10 l/min, respectively. The pressure of the nebulizer nitrogen gas was set to 2.6 bar, the end plate offset to 500 V, and the capillary voltage to 4500 V. The spectral range was 50 - 1000 m/z at 5 spectra/s. M.Sc. Franziska Vogl had externally calibrated the mass spectrometer prior to data acquisition utilizing sodium formate clusters (10 mM in 50:50 v/v water/isopropanol) and internal recalibration was achieved by employing sodium formate clusters injected via a six-port valve at the beginning of each run. Automated MS/MS measurements were performed with a signal threshold of 1000 and the fragmentation voltage was ramped from 25 - 35 eV with an isolation width of 4 - 8 m/z. Feature extraction was

achieved with the signal-to-noise threshold set to 20 in the "find molecular feature" algorithm in CompassDataAnalysis 4.1 (Bruker Daltonics, Bremen, Germany) [Zacharias et al. 2015]. She employed the 64-bit beta version of Profile Analysis 2.1 (Bruker Daltonics) for feature alignment over a retention time window of 0.01 - 14 min [Zacharias et al. 2015]. Reference compounds of propofol metabolites were obtained from Toronto Research Chemicals (Toronto, Canada) [Zacharias et al. 2015].

5.1.2.4 NMR data preprocessing

All 1D ^{1}H NOESY and 2D ^{1}H-^{13}C HSQC spectra were preprocessed as described in section 4.2.2.3. Note that a subset of eight plasma specimens had been preprocessed in the context of my master thesis [Zacharias 2012]. In order to compensate slight shifts in signal positions across spectra due to small variations in sample pH, salt concentration and/or temperature, the NMR spectral data was subjected to a bucketing procedure as described in section 4.2.2.3. For plasma 1D ^{1}H NOESY spectra, the spectral region from 9.5 - 0.5 ppm was evenly split into bins of 0.01 ppm width [Zacharias et al. 2015]. The spectral region from 6.2 - 4.6 ppm containing the broad urea signal and the remaining water signal, as well as the NMR signals (3.815 - 3.76 ppm, 3.68 - 3.52 ppm, 3.23 - 3.20 ppm, and 0.75 - 0.725 ppm) corresponding to residual glycerol from the ultrafilration membrane and free EDTA, respectively, were excluded during the bucketing procedure [Zacharias et al. 2015]. Note that the unspecific urea signal was much smaller in the 1D ^{1}H spectra acquired for plasma in comparison to urine specimens [Zacharias et al. 2013a), Zacharias 2012]. A total number of 718 bins remained for each 1D ^{1}H NOESY plasma spectrum [Zacharias et al. 2015].

5.1.2.5 NMR data normalization

The general goal of data normalization can be described as a minimization of technical and undesired biological variances without reduction of the intended biological variation, as explicitly outlined in section 4.3.1.1.

[2] Prior to subsequent normalization, data of each plasma spectrum were scaled to the integral of the reference TSP signal from 0.05 ppm to - 0.05 ppm to correct for variations in spectrometer performance over time. This is especially important for larger sample sets such as the AKI data set where acquisition time of the whole data set amounts to several days. For this it is important that the pipetting error of the reference substance is smaller than the observed variations in spectrometer performance. Prof. Dr. Wolfram Gronwald and Claudia Samol analyzed this in detail by splitting a urine specimen from a healthy volunteer at the University of Regensburg in 10 different aliquots and adding 50 µl of deuterium oxide containing 0.75% (w/v) TSP to each sample. To simulate a realistic setting they defined 10 different runs in which each aliquot was measured once with identical parameters, yielding a total of 100 1D ^{1}H spectra. Results showed that for the average pipetting error, defined as signal variations of the TSP reference signal between aliquots, a relative standard deviation of 0.8% was obtained. For

[2] The following section has been published in [Hochrein et al. 2015] in a slightly altered version, and is also part of Dipl. Math. Jochen Hochrein's Ph.D. thesis [Hochrein 2016].

the spectrometer performance, defined as signal variations of the TSP reference signal within each aliquot across different measurements, an average relative standard deviation of 3.7% was determined. As a consequence it is save to conclude that scaling relative to the TSP signal helps reducing variations in spectrometer performance.
A subsequent $\log_2$ transformation was applied to all 1D ^{1}H plasma spectra to minimize heteroscedasticity [Zacharias et al. 2015], as explicitly discussed in section 5.1.3.1. The corresponding *R*-code can be found in Appendix I section 7.1.3.1.

5.1.2.6 Prognostication method

The predictive performance of potential plasma biomarkers was assessed by employing classification, compare to section 4.3.1.3. Dipl. Math. Jochen Hochrein systematically evaluated the predictive performance of six different binary classification algorithms in combination with various strategies for data-driven feature selection on five different data sets including the current AKI plasma set [Hochrein et al. 2012]. For most data sets, a RF classification algorithm combined with *t*-score-based feature filtering performed best with regard to prediction accuracy [Hochrein et al. 2012]. The combination of an SVM with radial basis function kernel and *t*-score-based feature filtering performed best with respect to AUC-ROC values for almost all employed data sets [Hochrein et al. 2012]. Nevertheless, the current AKI plasma set was best classified by an RF classification with *t*-score-based feature filtering in terms of both prediction accuracy and AUC-ROC values [Hochrein et al. 2012]. As a consequence, this classification algorithm has been chosen for the prognostication of plasma 1D ^{1}H NMR metabolic fingerprints in this thesis. Note that, in comparison to [Hochrein et al. 2012], where the AKI plasma data set had been VS-normalized, compare to section 4.3.1.1., the corresponding data set was only $\log_2$-transformed in this thesis [Zacharias et al. 2015], as explicitly discussed in section 5.1.3.1. For the AKI urinary data set, an SVM algorithm with radial basis function kernel in combination with a *t*-score-based feature selection had been employed since it performed best in a preliminary classification algorithm evaluation conducted by Dipl. Math. Jochen Hochrein during his diploma thesis [Hochrein 2011]. A performance evaluation of the RF classification algorithm for NMR derived metabolomic data sets had not yet been taken place when the investigation of the AKI urinary data set was performed in the context of my master thesis [Zacharias 2012].

[3] Prognostication of plasma specimens was performed employing an RF classifier in combination with *t*-score-based feature filtering, as described in section 4.3.1.3. This combined strategy allows fast subsequent identification of NMR signals driving the separation of cases. It also keeps the computational model relatively sparse. Prognostications were accomplished within a nested leave-five-out cross-validation scheme. To guarantee an almost unbiased estimate of the true prognostication error [Varma and Simon 2006], two nested inner loops were included for parameter selection. The number of selected features was optimized in the first inner loop

[3] The following section has already been published in a slightly modified version in [Zacharias et al. 2015]. The classification/prognostication concept of nested cross-validation employed here was implemented by Dipl. Math. Jochen Hochrein and is also part of his Ph.D. thesis [Hochrein 2016].

in steps of one from a starting value of one. The internal parameters of the RF classifier such as the number of trees (n_{tree}) and the number of tried variables (m_{try}) were calibrated in the second inner loop employing a grid search procedure, where each of the two parameters was varied over 5 different settings leading in total to 25 combinations. The parameter n_{tree} was varied in steps of 100 between 100 and 500. Since the number of variables employed by Random Forests for the splitting in each node (m_{try}) depends on the total number of input variables, i.e. the number of selected features, it has been proposed [Liaw and Wiener 2002, Breiman 2001] to start optimization of m_{try} at the square-root of the number of input variables (default) and then trying 2 times default and 0.5 times default. Dipl. Math. Jochen Hochrein complemented this sequence by 1.5 times default and 0.75 times default to cover a finer grid of the optimization space. Assuming, for example, an output of 100 variables by the feature-filtering step, the (floored) values of m_{try} to test would have been 5, 7, 10, 15, and 20, respectively. Each sample was used once as a test-sample in each RF run.
Classification performance was evaluated by analyzing ROC plots, compare to section 4.3.1.3. For each classification, the average prediction/prognostication accuracy given as the arithmetic mean ± standard deviation of the individual results and the area under the ROC curve are given. The significance of identified biomarkers that are present at different levels in AKI and non-AKI patients was assessed by the corresponding P values. Raw P values were calculated by a two-sided Welch t statistic assuming Gaussian distribution of the data, which was confirmed by means of the Kolmogorov-Smirnov test, compare to section 4.3.1.3. To adjust for multiple testing, P values were modified for controlling the FDR according to the method of Benjamini and Hochberg, as described in section 4.3.1.3.

5.1.2.7 Permutation tests

By comparing the prognostic accuracies obtained for originally non-permuted and randomly permuted data, the significance of the obtained classification results can be estimated [Zacharias et al. 2013a), Zacharias et al. 2015, Mukherjee et al. 2003]. Here, the original class-labels of the AKI patients, reflecting their clinical diagnosis of AKI/non-AKI incidence, were randomly permuted before performing an RF classification with t-score-based feature selection, as explicitly described in section 5.1.2.6. For the plasma data set, the permutation test was performed 20 times [Zacharias et al. 2015], each starting with a new random permutation of the original class-labels as well as a fresh splitting into training and test data [Zacharias et al. 2015]. The mean values and standard deviations of the averaged total prediction accuracy, the area under the ROC curve, and the sensitivity and specificity were calculated [Zacharias et al. 2015]. Moreover, the optimal n_{tree} and m_{try} parameters of the RF classifier were reported [Zacharias et al. 2015]. If the mean averaged total prediction accuracy and the mean AUC-ROC values of the RF classification for randomly permuted class-labels are considerably lower than the corresponding values of the RF classification for non-permuted class-labels, one can conclude that the latter results were not obtained by chance [Mukherjee et al. 2003].

5.1.2.8 Prognostication of AKI with selected metabolites

[4] For prognostication of AKI with selected individual metabolites, the threshold for group assignment was varied over the respective measured concentration ranges to obtain ROC curves. For prognostication with predefined sets of known metabolites, probability estimates for group assignment, reflected by the respective decision values, were first obtained in five runs of leave-five-out cross-validation by means of an SVM with a linear kernel function, compare to section 4.3.1.3, thresholds were then varied over the entire range to generate ROC curves. The cost parameter C was stepwise increased from 2^{-5} to 2^{5}.

5.1.2.9 Analysis of intermediate cases of AKI

[5] During the evaluation of the prognostic performance of plasma biomarkers based on 1D ^{1}H NMR spectra, as explicitly described in section 5.1.3.3, the poor predictive performance of the tested plasma fingerprints for AKIN-stage 1 patients became striking. Therefore, to gain more insight into the nature of AKIN 1 disease, Dipl. Math. Jochen Hochrein adapted a computational algorithm called "core-group extension", which was originally devised to derive a molecular signature of Burkitt's lymphoma from gene expression profiles [Hummel et al. 2006]. The algorithm is initially trained on a core group of certain outcome or diagnosis. Next, the trained classifier is used to calculate scores for the intermediate cases followed by a ranking of samples according to these scores. Here, the dataset was first separated according to the AKIN criteria into a "stable" core group comprising the AKIN 0 non-AKI and AKIN 2 and 3 AKI cases, while the so-called "unstable" group comprised all patients assigned as AKIN 1. Using the plasma concentrations of a specific set of biomarkers, whose choice is explicitly described in section 5.1.3.7, of the "stable" group, I optimized the cost parameter C of an SVM algorithm capable of estimating prognostication probabilities, compare to section 4.3.1.3, in a leave-one-out cross-validation, where the exponent of C was increased systematically from 2^{-40} to 2^{-10} at a step size of 1, to select the model with the lowest error-rate in the discrimination of AKI from non-AKI members of the "stable" group, and recorded for this model within the leave-one-out cross-validation the corresponding scores. The scores reflect the corresponding prognostication probabilities. Note, as parameter optimization and determination of scores was performed on the same data due to the small size of the AKIN 2/3 cohort (N=9), which precluded the meaningful application of a nested cross-validation, the score obtained for the "stable" group should be treated with care. Next, the classifier was trained using all the samples of the "stable" group as training data and the parameter settings corresponding to the optimal computational model to score the samples from the "unstable" group.

[4] The following section has been published in [Zacharias et al. 2015] in a slightly modified version. The applied linear SVM cross-validation with optimization of the cost parameter C was slightly modified from a previous R-code of Dipl. Math. Jochen Hochrein.

[5] This section has already been published in [Zacharias et al. 2015], and the presented concept as well as the corresponding algorithm was developed by Dipl. Math. Jochen Hochrein. It is also part of his Ph.D. thesis [Hochrein 2016].

5.1.2.10 Metabolite quantification

Absolute metabolite quantification was performed as explicitly described in section 4.3.3.
[6] For the quantification of Ca^{2+} and Mg^{2+} ions I made use of the fact that both ions form complexes with EDTA that give rise to distinct NMR peaks in the 1D ^{1}H and 2D ^{1}H-^{13}C HSQC NMR spectra [Barton et al. 2010, Nicholson et al. 1983, Somashekar et al. 2006]. Here I used the singlet ^{1}H NCH2CH2N NMR signals at 2.56 ppm and 2.70 ppm for Ca-EDTA^{2-} and Mg-EDTA^{2-}, respectively. For validation, spike-in experiments were performed in H_2O and pooled plasma (Appendix II section 7.2.3 Table 7.2). In water, mean recoveries of 97 ± 2.5% and 102 ± 2.1% were obtained for Ca-EDTA^{2-} and Mg-EDTA^{2-}, respectively, while the respective values for ultrafiltered human plasma were 95.0 ± 6.8% and 104.0 ± 4.9%. Individual calibration factors, lower limits of quantification, as well as ^{1}H and/or ^{13}C peak ranges (compare to section 4.3.3) for CaEDTA^{2-}, MgEDTA^{2-}, and propofol-glucuronide have been determined experimentally according to [Klein 2011].
Baseline serum creatinine concentrations prior to cardiac surgery were determined with standard techniques from clinical chemistry at the University Clinic of Erlangen-Nuremberg.

5.1.3 Results

5.1.3.1 Appropriate data normalization

[7] For all investigated urine as well as eight plasma specimens, both 1D ^{1}H and 2D ^{1}H-^{13}C HSQC NMR spectra were acquired in the context of my master thesis [Zacharias 2012]. For the remaining plasma specimens, both 1D ^{1}H and 2D ^{1}H-^{13}C HSQC NMR spectra were acquired in the context of my Ph.D. thesis. Figure 5.1 shows an exemplary subtraction spectrum obtained by subtracting the 1D ^{1}H NMR spectrum of a non-AKI plasma specimen from that of an AKIN 3 specimen, both of which had been collected 24 h after surgery. The subtraction of measured spectra generates a virtual NMR spectrum that highlights those spectral features that differ between the samples.
While investigating the AKI plasma data set, several different normalization techniques were tested, including Variance Stabilization normalization as outlined in section 4.3.1.1. VSN preprocessing yielded a significant (B/H adjusted $P = 1.2 \times 10^{-6}$) difference in the abundance of CaEDTA^{2-} between the AKI and the non-AKI group. However, the subsequent targeted quantitative analysis of CaEDTA^{2-}, compare to Table 5.3, revealed no significant difference for the absolute concentrations of CaEDTA^{2-} between non-AKI and AKI group ($P = 0.47$). In contrast, simple scaling of spectral features to the TSP reference signal followed by $\log_2$-transformation, confirmed for CaEDTA^{2-} the absence of a significant intergroup difference (B/H adjusted $P = 0.67$), but instead revealed MgEDTA^{2-}, which had not been among the discriminating features upon VSN, to be highly discriminative, compare to Appendix II section 7.2.6 Table 7.4. Since calcium levels are usually tightly regulated in the human body [Felsenfeld

[6] The following section has already been published in [Zacharias et al. 2015] in a slightly altered version.
[7] The following section has already been published in [Zacharias et al. 2015], and [Hochrein et al. 2015] in a slightly altered version. Some of the results presented here are also part of Dipl. Math. Jochen Hochrein's Ph.D. thesis [Hochrein 2016].

et al. 2013], a significant difference in the $CaEDTA^{2-}$-levels of AKI and non-AKI patients is rather unlikely, and points to an inappropriate application of VSN on 1D ^{1}H plasma NMR data. Significant differences in $MgEDTA^{2-}$ could also be confirmed by targeted quantitative analysis (compare to Table 5.3). Note that careful manual inspection of the spectra revealed that the bins corresponding to $CaEDTA^{2-}$ and $MgEDTA^{2-}$ at 2.56 ppm and 2.70 ppm, respectively, do not contain contributions from citrate, although the small citrate signals are in close proximity. Simple scaling to the TSP signal works well in this case, although it provided only a correction for differences in spectrometer performance and no adjustment for non-induced biological variances and technical biases not related to spectrometer performance.

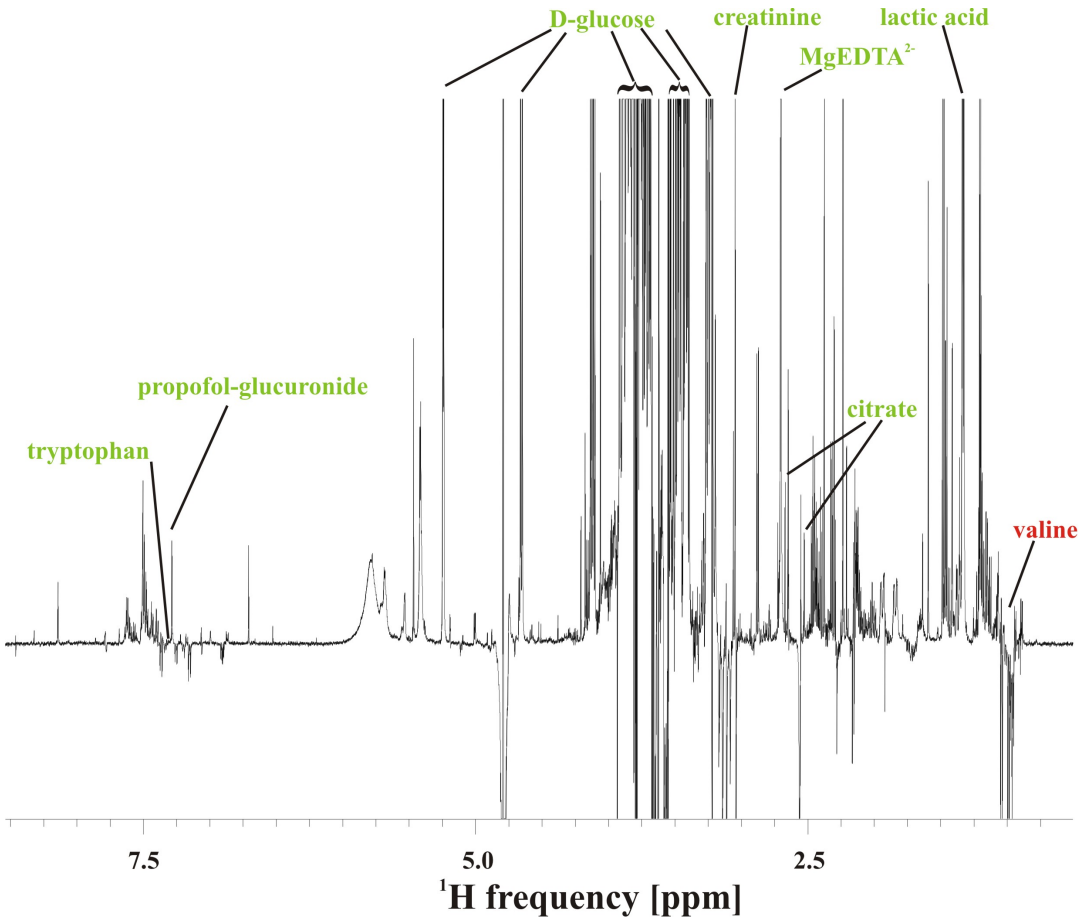


Figure 5.1: **Exemplary 1D ^{1}H NMR subtraction spectrum of plasma specimens collected 24 h after cardiac surgery.** 1D ^{1}H NMR subtraction spectrum obtained by the subtraction of a representative non-AKI plasma spectrum from an AKIN 3 spectrum and generated by Prof. Dr. Wolfram Gronwald. Metabolites significantly up- and downregulated in AKI patients compared to metabolites in unaffected patients are marked in green and red, respectively. The complete list of discriminating features is given in Appendix II section 7.2.6 Table 7.4. The ratio of the total spectral areas of the two spectra used for computing the difference in the spectra amounts to 1.97, indicating a considerably higher overall metabolite concentration in the specimen of the AKIN 3 case. Modified from [Zacharias et al. 2015].

Close inspection of Figure 5.1 indicates that the total integral of spectral features upregulated in the AKIN 3 specimen compared to that of the non-AKI specimen is much larger than that of the downregulated features. This is mainly due to the significantly ($P = 0.03$) higher levels of glucose, the most abundant plasma metabolite, in the AKI (9.72 ± 2.73 mmol/l) rather than the non-AKI group (8.38 ± 2.83 mmol/l) (Table 5.3). As already outlined in section 4.3.1.1, VSN is only applicable if a relatively small proportion of metabolites/feature intensities is regulated in approximately equal shares up and down between the intended biological groups [Hochrein et al. 2015, Kohl et al. 2012]. Therefore, significant differences in glucose levels and consequently total spectral areas between the two biological groups prohibits the application of VSN in this setting.

Encouraged by these results, Dipl. Math. Jochen Hochrein developed a strategy to choose the appropriate data normalization method for the statistical analysis of an NMR data set without explicit prior knowledge about significant inter-group inhomogeneities [Hochrein et al. 2015, Hochrein 2016]. First, the total spectral area for each NMR spectrum, excluding areas of the solvent signal and the broad urea signal, is calculated. Then, the Shapiro-Wilk test [Shapiro and Wilk 1965] is applied in order to test the set of total spectral areas for normal distribution. The corresponding null hypothesis is that the total spectral areas are normally distributed around a fixed mean, irrespective of the experimental groups. To detect single outliers within the investigated groups that may not be detected by the Shapiro-Wilk-Test, total spectral areas are also plotted in a histogram representation (data not shown). The corresponding *R*-code can be found in Appendix I section 7.1.2.
For the AKI plasma data set, the Shapiro-Wilk test yielded a P-value of 1.6×10^{-4}. This indicates the presence of significant inter- and intra-group inhomogeneities in total spectral areas, as caused in part by the significantly higher glucose levels in the AKI group than in the non-AKI group. Consequently, the complete statistical analysis for the AKI plasma data set was performed with the 1D ^{1}H NMR bucket intensities scaled to TSP and subsequently $\log_2$ transferred.

5.1.3.2 Time-course development

[8] Since the clear separation between urine specimens collected before and at 4 h after the surgery in the PCA plot shown in Appendix II section 7.2.4 Figure 7.1a), was mainly driven by the high concentration of the exogenous compound D-mannitol at 4 h after the surgery [Zacharias et al. 2013a), Zacharias 2012], compare to section 5.1.1, a second PCA was conducted excluding all spectral regions corresponding to this compound (Appendix II section 7.2.4 Fig. 7.1b)) in the context of my Ph.D. thesis [Zacharias et al. 2013a)]. As in Appendix II section 7.2.4 Fig. 7.1a), a clear separation between the different time-points is visible. The specimens collected at 24 h after surgery are located approximately in between the two other time-points, with urine specimens from non-AKI patients being located on average closer to the pre-surgical specimens, whereas those of the AKI patients group closer to the 4 h specimens. The group separation is now mainly driven by creatinine (loadings not shown).

[8] The following section has already been published in [Zacharias et al. 2013a)] in a slightly altered version.

5.1.3.3 Prognostication of AKI

Encouraged by the promising classification results obtained by training an SVM classifier with radial basis function kernel on 1D ^{1}H NMR spectra of 106 urine specimens collected at 24 h after cardiac surgery, compare to section 5.1.1, I performed the same task employing the corresponding plasma specimens collected at the same time-point, which were available for NMR spectroscopy.
[9] For RF-based prognostication of the eventual AKIN stage, $\log_2$ transformed 1D ^{1}H NMR plasma spectra were split into 718 evenly spaced bins or features, excluding chemical shifts representing water, urea, glycerol and free EDTA. Subsequent analysis of the 33 AKI and 52 non-AKI cases yielded an overall prognostication accuracy of 80.0 ± 0.9 % (compare to Table 5.2) and a corresponding area under the ROC curve of 0.87 ± 0.01. On average, the RF algorithm employed 24.0 ± 2.8 of the most discriminative features as selected by a t-test based feature selection step prior to classification. As can be seen from Table 7.4 in Appendix II section 7.2.6, the corresponding p-values of these features showed a range from 2.06e^{-8} to 8.55e^{-6}. The RF parameters m_{try} and n_{tree} were optimized to 3.0 ± 0.0 and 270 ± 24.5, respectively. The overall sensitivity and specificity amounted to 72.7 ± 1.9% and 84.6 ± 1.7%, respectively. Considering the AKIN stages separately, the sensitivity for AKIN 2 and 3 amounted to 100.0 ± 0.0% and 96.7 ± 6.7%, respectively, whereas it dropped to 63.3 ± 1.7% for AKIN 1 patients, compare to Table 5.2.

AKIN-stage	**All (1-3 and 0)**	**1**	**2**	**3**
24 h plasma prediction accuracy [%]	80.0 ± 0.9	63.3 ± 1.7	100.0 ± 0.0	96.7 ± 6.7

Table 5.2: **Classification performance depending on AKIN-stage for plasma specimens collected at 24 h post-surgery.** Given are the prediction accuracies, ordered according to diagnosis. Mean values and corresponding standard deviations are obtained from five nested cross-validation runs. Modified from [Zacharias et al. 2015].

The 85 plasma samples constituted a subsample of the original study of 106 individual urine specimens, as described in section 5.1.2.1. To allow for a fair comparison between urinary and plasma data, the 1D ^{1}H NMR urine spectra of the 85 patients, for whom plasma specimens were available, were scaled to creatinine and $\log_2$ transformed before they were subjected to a single random forest run. The obtained overall prognostication accuracy amounted to 69.4% employing 7 features with a corresponding area under the ROC curve of 0.73.
Next, permutation tests with randomly perturbed class-labels were also performed for the AKI plasma data set to exclude the possibility that the observed prognostication accuracies had been obtained by chance. After an initial RF run with the complete feature set of 718 features and permuted class-labels, which revealed 119 as the median number of selected features, the

[9] The following section has already been published in [Zacharias et al. 2015] in a slightly altered version. Parts of the analysis presented here were performed by M.Sc. Franziska Vogl.

feature selection was limited to a range of 109 to 129 features for the subsequent twenty RF runs, each of which started with a fresh permutation of the class-labels and a random splitting of test and training data. Over the twenty RF runs, I received an averaged total accuracy of 55.7 ± 5.1%, a mean area under the ROC curve of 0.48 ± 0.08, and a sensitivity and specificity of 17.1 ± 7.2% and 80.2 ± 6.3%, respectively (Appendix II section 7.2.5 Table 7.3). Results for the permuted data were in all 20 runs considerably lower than for the non-permuted data, indicating that the results for the non-permutated data were with high probability not obtained by chance ($P < 0.05$) and that the study was sufficiently powered. Two exemplary ROC curves for the permuted and non-permuted data are given in Appendix II section 7.2.5 Figures 7.2a) and 7.2b), respectively.

As described in section 5.1.2.6, *t*-statistics were used for both, feature filtering and identification of spectral features that distinguish between AKI and non-AKI plasma NMR fingerprints. After correction for multiple testing by controlling the FDR at 5%, 261 significantly differential NMR features were obtained. A heat-map representation of these features is displayed in Appendix II section 7.2.6 Figure 7.3 and a list of all significant NMR features is given in Appendix II section 7.2.6 Table 7.4. Their up- and down-regulation in the heat-map representation is color coded in yellow and blue, respectively. The patients were arranged from left to right as follows: 45 cases correctly prognosticated not to develop AKI, 7 cases falsely prognosticated to develop AKI, 9 cases of AKIN 1 falsely prognosticated not to develop AKI, and 15, 3 and 6 cases each of AKIN 1, 2 and 3, respectively, correctly prognosticated. Rows were ordered according to increasing correlation coefficients between disease status and feature intensities. As on average 24.0 ± 2.8 of the most significant features were used by the RF algorithm, the 27 most significant NMR features are indicated with red arrows in Appendix II section 7.2.6 Figure 7.3. These NMR features could only partly be assigned to known metabolites due to either massive signal overlap in some regions of the 1D spectra (see Figure 5.1) or insufficient signal intensity.

The most significant plasma feature was a well-resolved singlet signal present at 7.285 ppm (P_{adj}=2.1e^{-8}), which could be identified by neither database searches nor 2D ^{1}H ^{1}H TOCSY, ^{1}H-^{13}C HSQC, and ^{1}H-^{13}C HMBC spectra, respectively. Therefore, to facilitate assignment, M. Sc. Franziska Vogl performed metabolic fingerprinting by means of high-resolution LC-QTOF-MS on five plasma specimens each selected from the AKI and the non-AKI group, respectively. A total of 531 features were observed in positive mode and 16 in negative mode. Dipl. Math. Jochen Hochrein sorted features according to Student's *t*-tests. After controlling the FDR at the 5% level according to the method of Benjamini and Hochberg, 11 significant features remained, each of which was defined by retention time and the m/z value. By means of the Smart Formula tool (Bruker Daltonics, Bremen, Germany), M.Sc. Franziska Vogl determined molecular sum formulas to search the HMDB [Wishart et al. 2007], METLIN [Smith et al. 2005], and ChEBI (Chemical Entities of Biological Interest) metabolite databases [Hastings et al. 2013]. For the most promising hits, she analyzed commercial standards to verify identification. Furthermore, M.Sc. Franziska Vogl performed MS/MS experiments on both standards and plasma specimens for additional verification. Among the most discriminating features, she positively identified the propofol metabolites propofol-glucuronide and 4-hydroxy-propofol-1-OH-D-glucuronide.

I acquired 1D ^{1}H NMR reference spectra on these compounds and unambiguously verified the

assignment of the NMR signal at 7.285 ppm to propofol-glucuronide. Furthermore, the presence of 4-hydroxy-propofol-1-OH-D-glucuronide as another discriminating compound could be verified by the NMR data. NMR-based quantification of propofol-glucuronide showed significantly increased plasma levels (0.004 ± 0.002 mmol/l vs. 0.010 ± 0.08 mmol/l, P=0.00008) in AKI patients. However, for both the total dosage of administered propofol (2747.7 ± 1257.4 mg vs. 3313.3 ± 1896.4 mg, P=0.14) and the dosing rate (6.8 ± 3.7 mg/min vs. 5.8 ± 1.9 mg/min, P=0.09), no significant differences between non-AKI and AKI patients could be observed. However, the two groups differed significantly with regard to the duration of propofol administration (427.94 ± 185.34 min and 600.79 ± 360.69 for non-AKI and AKI patients, respectively, p=0.015) and the time elapsed between termination of propofol infusion and sample collection (1138.6 ± 201.7 min and 969.0 ± 362.7 min for non-AKI and AKI patients, respectively, p=0.02). This suggested that the increased plasma levels of propofol-glucuronide in AKI patients were a consequence of both prolonged administration and delayed excretion. This was confirmed by re-analysis of the urinary NMR fingerprints obtained for the same patients. At 4 hours after surgery, urinary levels of propofol-glucuronide amounted to 0.67 ± 0.30 mmol/mmol$_{\text{crea}}$ and 0.63 ± 0.35 mmol/mmol$_{\text{crea}}$, respectively, for non-AKI and AKI patients (p=0.55), while at 24 hours after surgery urinary levels had dropped to 0.14 ± 0.07 mmol/mmol$_{\text{crea}}$ and 0.19 ± 0.10 mmol/mmol$_{\text{crea}}$, respectively, but were significantly (p=0.02) higher in the AKI group (Table 5.3).

As can be seen from Table 7.4 in Appendix II section 7.2.6, plasma NMR features used for prognostication correspond to compounds of both endogenous origin, such as tryptophan (P_{adj}=1.1e^{-6}), myo-inositol (P_{adj}=2.3e^{-6}), hippurate (P_{adj}=2.5e^{-6}), citrate (P_{adj}=3.2e^{-6}), and creatinine (P_{adj}=3.9e^{-6}), and exogenous origin, such as propofol-glucuronide (P_{adj}=2.1e^{-8}) and the antifibrinolytic agent tranexamic acid (P_{adj}=3.2e^{-6}). A special case is Mg^{2+} (P_{adj}=2.9e^{-6}), which can be of both endogenous and exogenous origin as Mg^{2+} is often administered for the treatment of cardiac dissrhythmia.

To analyze the impact of tranexamic acid and other exogenous compounds such as D-mannitol, paracetamol-sulfate, propofol-glucuronide, 4-hydroxy-propofol-1-OH-D-glucuronide and 4-hydroxy-propofol-4-OH-D-glucuronide on plasma prognostication performance, all spectral areas corresponding to known exogenous compounds were excluded prior to data analysis. In subsequent analysis, which employed 26 features, an average prediction accuracy of 82.4% (vs. 80% including exogenous compounds) and an area under the ROC curve of 0.87 (vs. 0.87) were obtained. Overall sensitivity and specificity amounted to 75.8% and 86.5%, respectively, while the respective values before exclusion of exogenous compounds had amounted to 72.7% and 84.6%.

In addition, I investigated whether improved prognostication performance could be achieved by combining the plasma data with the corresponding 24 h urine data, which had been scaled to creatinine and also $\log_2$ transformed. The final data matrix consisted of 1419 rows representing 718 plasma and 701 urine features and 85 columns representing the 33 and 52 AKI and non-AKI patients, respectively. One RF classification run with t-test based feature selection employing a leave-five-out cross-validation was performed. Results showed an averaged prognostication accuracy of 81.2% and an area under the ROC curve of 0.87, values similar to those obtained for plasma only. Further analysis showed that prognostication was based on 25 features. Ranking of these features according to their p-values revealed, that the first 24 features were identical

to the first 24 plasma features listed in Appendix II section 7.2.6 Table 7.4. The most significant urinary feature was tranexamic acid at rank 25 (P_{adj}=1.6e^{-5}), which explains why a combination of urinary and plasma fingerprints did not outperform prognostication on plasma fingerprints alone.

5.1.3.4 Investigation of CKD influence

During the review process for [Zacharias et al. 2013a)], the question arose whether pre-existing CKD might have a significant impact on the predictive performance of urinary fingerprints for the discrimination of AKI vs. non-AKI patients. Consequently, I investigated the CKD influence on the AKI prognostication based on urinary NMR fingerprints in the context of my Ph.D. thesis. Note that, in concordance with [Zacharias et al. 2013a), Zacharias 2012], here, urinary 1D ^{1}H NMR bucket intensities had been scaled to creatinine and subsequently Quantile normalized as described in section 5.1.1.

[10] Close inspection of Table 7.1a) in Appendix II section 7.2.1, showed that from the urine study cohort, thirty-nine of the patients were suffering from non-dialysis CKD. However, employing a two-sided t-test, no significant differences between the preoperative urinary NMR spectra of patients with or without CKD could be detected (data not shown). P-values were adjusted for multiple testing by controlling the FDR at the 5% level. Furthermore, two-sided t-tests for NMR spectra of urine specimens collected at 4 and 24 h after surgery were performed separately for AKI and non-AKI patients, respectively, with regard to the presence or absence of CKD. Of the four tests conducted, significant FDR-adjusted differences in spectral intensity with an FDR $< 5\%$ were observed only between CKD and non-CKD spectra acquired for urine specimens collected 4 h after surgery from the cohort of 34 AKI patients. Of the two significant NMR features observed, one with a $P_{adj} = 0.03$ remains to be identified, while the second one with a $P_{adj} = 0.04$ was tentatively assigned to phenylacetylglutamine. Of the two urinary amino acid conjugates of phenylacetic acid reported in the literature, namely phenylacetylglycine and phenylacetylglutamine, the former is typically assigned to the observed significant feature representing the phenyl moiety in NMR studies of human urine [Kang et al. 2011]. However, based on available literature on the conjugation of phenylacetic acid in humans, man excretes exclusively the glutamine conjugate [James et al. 1972, Fukui et al. 2009]. Phenylacetic acid and its glutamine conjugate are known uremic solutes, the serum concentrations of which are increased in CKD patients due to attenuation of whole-body phenylalanine hydroxylation [Itoh et al. 2012, van de Poll et al. 2004]. However, neither distinguished AKI from non-AKI patients at 4 h after surgery.

In the plasma subcohort, twelve out of 52 non-AKI patients (23.1%) and 21 out of 33 AKI patients (63.6%), respectively, suffered from CKD, with a P-value calculated by Fisher's exact test of 0.0003 (compare to Appendix II section 7.2.1 Table 7.1b)). To test whether results obtained for AKI and non-AKI specimens had been dominated by CKD, I selected randomly 24 patients each from the AKI and non-AKI group, so that each group included 12 patients with

[10] The following section has already been published in [Zacharias et al. 2013a)] and [Zacharias et al. 2015] in a slightly altered version.

and 12 patients without CKD. A Student's t-test yielded after correction for multiple testing 73 significant features, 72 of which had been also part of the significant features obtained when all 85 samples were included (Appendix II section 7.2.6 Table 7.4). RF based prognostication with leave-two-out cross-validation obtained an averaged prognostication accuracy of 72.9%, as well as an area under the ROC curve of 0.84. On average, 17.5 features were employed by the algorithm. Sensitivity and specificity amounted to 70.8% and 75.0%, respectively.
The corresponding permutation test was performed once, with an average total accuracy of 41.7%, an area under the ROC curve of 0.44, a sensitivity of 50.0% and a specificity of 66.7%. As both groups contained the same number of patients with and without CKD, these results clearly showed that CKD incidence did not exert a major effect on prognostication of AKI based on plasma 1D ^{1}H NMR spectra.

5.1.3.5 Quantification of metabolites

[11] In addition to the analysis of the NMR fingerprints, the plasma levels of 16 organic metabolites and the dications Ca^{2+} and Mg^{2+} were quantified from the 85 1D ^{1}H NMR spectra. Mean plasma concentrations and standard deviations for the non-AKI and AKI group, respectively, as well as p-values based on two-sided t-tests are given in Table 5.3. Note that due to the relatively small number of quantified metabolites no correction for multiple testing was applied. Metabolites that differed significantly ($P < 0.05$) in concentration between non-AKI and AKI group included propofol-glucuronide, lactic acid, valine, creatinine, D-glucose, and Mg-EDTA^{2-}. The statistical power (compare to section 4.3.1.3) of differences in absolute metabolite concentrations between groups was, except for D-glucose (56.9%), above the threshold of 80% for all significantly differential metabolites (Table 5.3).

[11] The following section has already been published in [Zacharias et al. 2015] in a slightly altered version.

Metabolite	Mean value ± SD [mmol/l] non-AKI	Mean value ± SD [mmol/l] AKI	*P*-value	Statistical power[5]
[1]**3-Hydroxybutyric acid**	0.65 ± 1.30	0.40 ± 0.38	0.21	18.5%[6]
[1]**Acetic acid**	0.19 ± 0.62	0.04 ± 0.02	0.10	27.5%[6]
[1]**Acetone**	0.15 ± 0.28	0.11 ± 0.10	0.31	12.1%[6]
[1]**Acetoacetic acid**	0.22 ± 0.34	0.12 ± 0.11	0.10	35.5%[6]
[2]**Alanine**	0.24 ± 0.09	0.26 ± 0.09	0.37	16.7%
[1,3]**Ca-EDTA**$^{2-}$	1.97 ± 0.14	1.95 ± 0.12	0.49	10.5%[6]
[1,4]**Creatinine**	0.09 ± 0.03	0.14 ± 0.05	0.000005	100%[6]
[2]**D-glucose**	8.38 ± 2.83	9.72 ± 2.73	0.03	56.9%[6]
[1]**Formic acid**	0.03 ± 0.01	0.03 ± 0.01	0.94	-
[2]**Glutamine**	0.37 ± 0.06	0.42 ± 0.08	0.053	88.7%[6]
[2]**Glycine**	0.29 ± 0.36	0.25 ± 0.22	0.74	8.8%[6]
[2]**Lactic acid**	1.66 ± 0.60	2.40 ± 1.27	0.003	94.7%[6]
[1]**L-isoleucine**	0.05 ± 0.02	0.04 ± 0.02	0.07	60.3%[6]
[1,3]**Mg-EDTA**$^{2-}$	1.05 ± 0.20	1.29 ± 0.23	0.00001	99.9%[6]
[1]**Propofol-glucuronide**	0.004 ± 0.002	0.010 ± 0.08	0.00008	100.0%[6]
[1,7]**Propofol-glucuronide (urine, 4h)**	0.67 ± 0.30	0.63 ± 0.35	0.55	9.0%[6]
[1,8]**Propofol-glucuronide (urine, 24 h)**	0.14 ± 0.07	0.19 ± 0.10	0.02	84.5%[6]
[1]**Threonine**	0.08 ± 0.02	0.08 ± 0.02	0.71	-
[1]**Tyrosine**	0.05 ± 0.01	0.05 ± 0.01	0.52	-
[1]**Valine**	0.20 ± 0.04	0.17 ± 0.05	0.007	91.5%[6]

Table 5.3: **Plasma levels of 18 selected analytes 24 hours after surgery in addition to the urine levels of propofol-glucuronide 4 and 24 hours after surgery, respectively.** Data were obtained from 1D ^{1}H and 2D ^{1}H-^{13}C HSQC spectra. Given are mean values and standard deviations in mmol/l for the non-AKI and the AKI group, respectively, as well as *P*-values calculated by a two-sided heteroscedastical *t*-test employing Microsoft EXCEL and the actual power of the hypothesis test employed. [1]Determined from 1D ^{1}H spectra. [2]Determined from 2D ^{1}H-^{13}C HSQC spectra. [3]Note the recoveries given in Appendix II section 7.2.3 Table 7.2. [4]Due to massive signal overlap in the 1D NMR spectral region of creatinine of a non-AKI patient, the creatinine value for this particular patient was determined from the corresponding 2D ^{1}H-^{13}C HSQC spectrum. [5]Statistical power was calculated by employing G*Power 3.1.7. [6]Effect size calculated according to Cohen's *d* in case of unequal variances. [7]Determined in urine 4 hours past surgery, values are normalized to urinary creatinine [mmol/mmol$_{crea}$]. [8]Determined in urine 24 hours past surgery, values are normalized to urinary creatinine [mmol/mmol$_{crea}$]. Modified from [Zacharias et al. 2015].

5.1.3.6 Prognostication of AKI by a small set of metabolites

[12] To evaluate the feasibility of a reliable prognostication of patients by employing a small subset of easily quantifiable metabolites only, prognostications were repeated with selected metabolites either individually or in different combinations thereof. Selection of these metabolites was performed according to P-values (Table 5.3). In case that more than a single metabolite was used, a linear SVM algorithm was employed. Five runs of leave-five-out cross-validation gave for the combination of the plasma metabolites Mg-EDTA^{2-}, lactate and creatinine, all of which are amenable to point-of-care-testing, an overall prediction accuracy of 77.0 $\pm$ 1.0% with a corresponding area under the ROC curve of 0.84 $\pm$ 0.01 (Figure 5.2). The largest AUC value of 0.94 $\pm$ 0.01 was obtained for plasma propofol-glucuronide in combination with the difference in serum creatinine before and plasma creatinine 24 hours after surgery (Figure 5.2). Note that serum creatinine levels prior to surgery had been determined by standard methods of clinical chemistry. The corresponding prediction accuracy was 86.8 $\pm$ 0.9%. In addition to the prognostication performance of the different combinations of plasma biomarkers shown in Figure 5.2, (1) a combination of plasma creatinine obtained 24 hours past surgery, with plasma lactic acid, plasma Mg-EDTA^{2-}, and plasma propofol-glucuronide, (2) a combination of the difference in pre- and postoperative (24 hours) serum/plasma creatinine with plasma lactic acid and plasma Mg-EDTA^{2-}, (3) a combination of the difference in pre- and postoperative (24 hours) serum/plasma creatinine with plasma lactic acid, plasma Mg-EDTA^{2-} and plasma propofol-glucuronide, as well as (4) a combination of plasma creatinine obtained 24 hours past surgery with plasma lactic acid, plasma Mg-EDTA^{2-}, and urinary carnitine, which had been normalized to urinary creatininine, were analyzed giving AUC values of 0.85, 0.88, 0.93, and 0.83, respectively. Urinary carnitine was chosen, because it had been identified during the investigation of the AKI urine cohort, compare to section 5.1.1, as a highly discriminative endogenous metabolite. Furthermore, prognostication of AKI based on total time of propofol administration alone revealed an AUC of 0.66 (Figure 5.2). Note that the total time of propofol administration was available for 84 out of 85 patients.

Clinical diagnosis of AKI is routinely made utilizing increases in SCr. In this study, AKI was diagnosed according to the AKIN criteria two and three days after surgery employing SCr levels, as outlined in section 5.1.2.1. Plasma creatinine levels for a subcohort of 85 patients have been determined for both non-AKI and AKI patients, compare to Table 5.3, and amounted to 0.09 $\pm$ 0.03 mmol/l and 0.14 $\pm$ 0.05 mmol/l, respectively, with a corresponding P-value of 5e^{-6}. These data show already at 24 h after surgery a significant increase in plasma creatinine levels for patients who were diagnosed later with AKI. However, the predictive performance of creatinine alone was outperformed by propofol-glucuronide (AUC-ROC 0.85 vs. 0.87, compare to Figure 5.2) determined in 24 h plasma specimens.

[12] The following section has already been published in [Zacharias et al. 2015] in a slightly altered version.

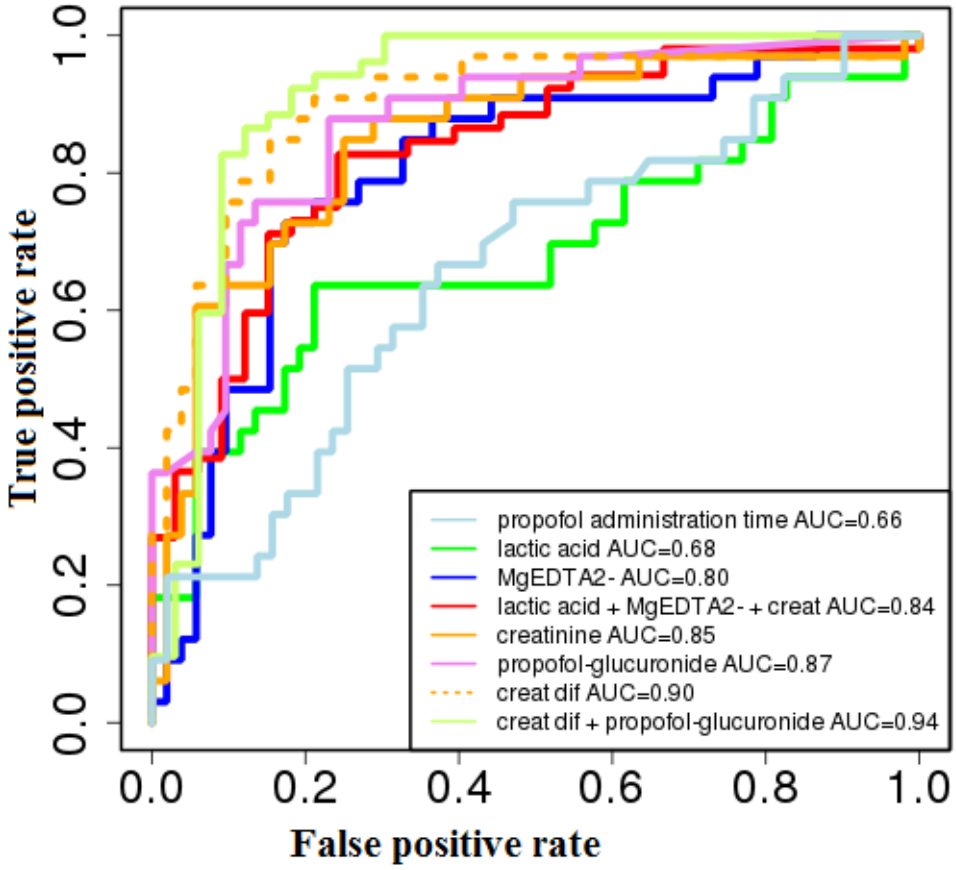


Figure 5.2: **Prognostication by different combinations of easily quantifiable biomarkers.** Prognostication performance was assessed by analysis of ROC curves. Prognostications are based on the absolute concentrations given in mmol/l of the following single metabolites or combinations of metabolites: plasma lactic acid, plasma Mg-EDTA^{2-}, plasma creatinine 24 hours past surgery, plasma propofol-glucuronide, difference in pre- and postoperative (24 hours) serum/plasma creatinine, plasma lactic acid + plasma Mg-EDTA^{2-} + plasma creatinine 24 hours past surgery, difference in pre- and postoperative (24 hours) serum/plasma creatinine + plasma propofol-glucuronide. In addition, the prognostication performance of the total time of propofol administration was evaluated. Note that the total time of propofol administration was available for 84 out of 85 patients. Reprinted with permission from [Zacharias et al. 2015]. Copyright 2015 American Chemical Society.

5.1.3.7 Analysis of intermediate cases of AKI

[13] From both the prognostication results based on plasma as well as urinary biomarkers and the heat-map representation shown in Appendix II section 7.2.6 Figure 7.3, it is obvious that the urinary and plasma metabolic profiles of AKIN 1 patients were often not in agreement with their respective staging according to the AKIN criteria. Therefore, I aimed at defining a scheme that separated them more robustly into AKI and non-AKI cases. As detailed in section 5.1.2.9, Dipl. Math. Jochen Hochrein followed a strategy originally developed by [Hummel et al. 2006] for the classification of Burkitt's lymphoma. Briefly, the original dataset was separated according to the AKIN criteria forming two groups of data denoted as the "stable" and the "unstable" group. The "stable" group comprised the AKIN 0 cases referred to as non-AKI, and the AKIN 2 and 3 cases referred to as AKI cases. The so-called "unstable" group comprised all patients of AKIN 1. To allow for potential point-of-care-testing, I selected metabolites that all could be obtained from a single sample and that have been shown to offer a good prognostication performance, namely the plasma compounds creatinine, lactate and Mg^{2+}, all determined 24 hours past surgery. Compounds of purely exogenous origin such as propofol-glucuronide were excluded, because they are not always administered and not readily amenable to bedside testing. Employing the samples of the "stable" group only, I optimized an SVM model with a linear kernel, which yielded a cost parameter of 2^{-27}. Employing this SVM model, I computed scores for both the "unstable" and the "stable" group. Note, that the score reflects an estimate of the probability that a specimen belongs to the AKI group.
An exemplary result for all 85 patients of the plasma subcohort is shown in Figure 5.3a). Not unexpectedly, most patients without AKI were assigned a score below 0.5. More interestingly the same was true for most AKIN 1 patients indicating that in these patients kidney function was largely preserved. In contrast, all AKIN 2 and 3 patients received a score greater than 0.5. The three metabolites used are also shown in a heat-map representation in Figure 5.3b). The up- and down-regulation of the plasma metabolites lactate, Mg-$EDTA^{2-}$ and creatinine with respect to the average is color coded in yellow and blue, respectively. All samples of the complete plasma data set, represented by the columns, are ordered according to the scores obtained in the scoring procedure. This scoring is congruent to the data given in Figure 5.3a) and additionally color coded in the topmost row of Figure 5.3b). A red vertical line denotes a score of 0.5 that was used for class separation into AKI (right) and non-AKI (left) cases. For reasons of comparison the scoring according to AKIN criteria in blue, dark-yellow, yellow, and light-yellow for AKIN 0, 1, 2, and 3, respectively, is included in the bottom row.

[13] The following section has already been published in [Zacharias et al. 2015] in a slightly altered version. The presented concept as well as the corresponding algorithm was developed by Dipl. Math. Jochen Hochrein. It is also part of his Ph.D. thesis [Hochrein 2016].

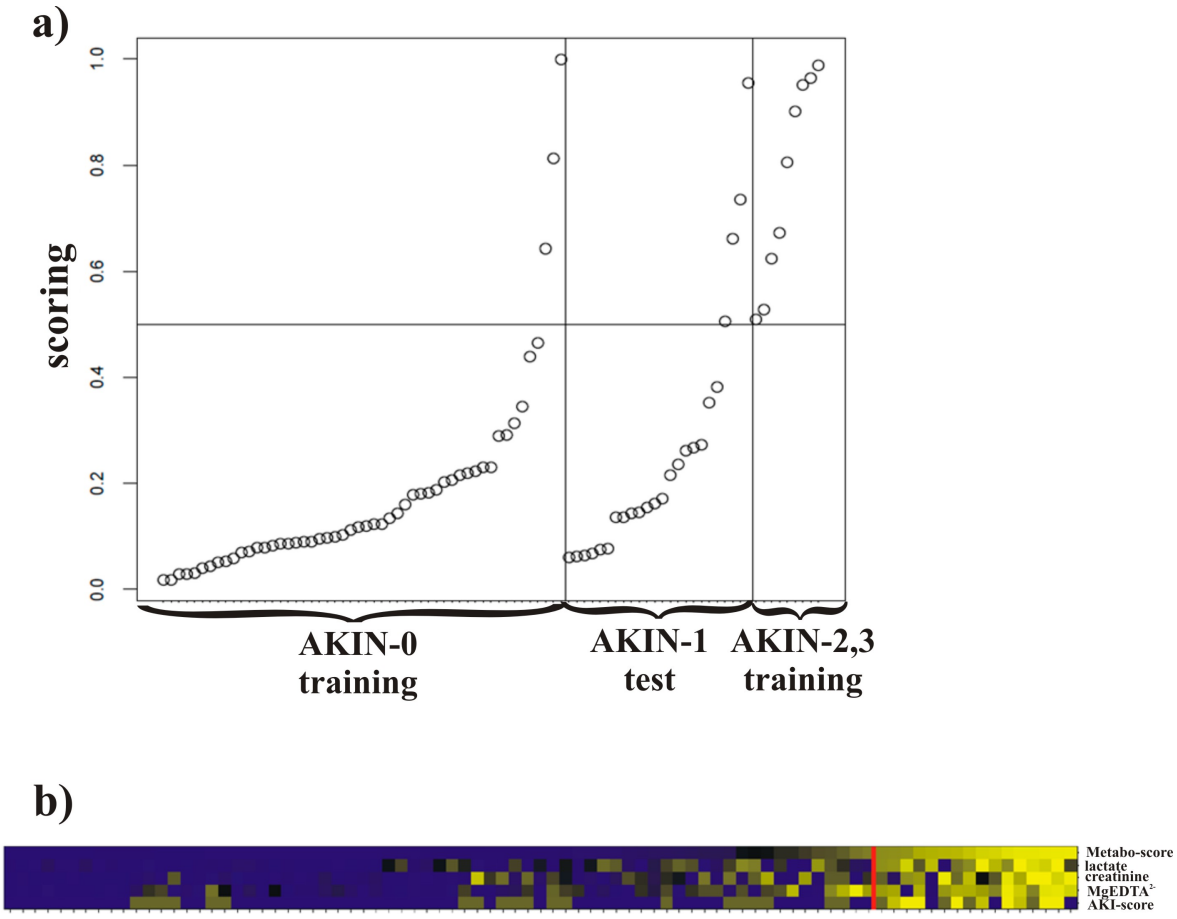


Figure 5.3: **Analysis of intermediate cases of AKIN 1 based on a model making use of the absolute concentrations of plasma lactate, magnesium, and creatinine determined for the "stable" core group that comprised all non-AKI and all AKIN 2 and 3 patients.** a) Patients are listed on the x-axis as follows from left to right: 52 non-AKI patients followed by 24 AKIN 1 patients and nine AKIN 2 and 3 patients. Corresponding scores are shown on the y-axis. The scores denote for each sample the estimate of the probability given by the SVM to belong to the class of AKI patients. Scores for the training group were obtained in a leave-one-out cross-validation. Data from one SVM run is given. Note that different runs gave very similar results. b) Heat-map representation of the investigated specimens. Displayed are the variations in concentration of plasma lactate, plasma Mg-EDTA^{2-} and plasma creatinine. These values are also summarized in Table 5.3. Samples are sorted in ascending order from left to right according to the scores obtained in the AKI-rescoring procedure. The red vertical line marks a score threshold value of 0.5. The first row shows the score values in a color-coded representation, while the next three lines correspond to the concentrations of lactate, creatinine and Mg-EDTA^{2-} that define the used model. Their up- and down-regulation with respect to the average is color coded in yellow and blue, respectively. The last row color-codes the AKIN-staging in blue, dark-yellow, yellow, and light-yellow for AKIN 0, 1, 2, and 3, respectively. Modified from [Zacharias et al. 2015].

5.1.4 Discussion

[14] This study was designed to investigate metabolic differences in plasma specimens between patients developing AKI versus patients not developing AKI after undergoing cardiac surgery with CPB use by NMR spectroscopy in combination with statistical data analysis methods.
The use of appropriate data normalization techniques is crucial for subsequent statistical data analysis, as illustrated here for the correct detection of fold changes of $CaEDTA^{2-}$ and $MgEDTA^{2-}$. The detection of possibly significant inter- and intra-group inhomogeneities in total spectral area of an NMR data set is crucial for the subsequent choice of normalization method and was conducted for all following studies.
To date, proteins have dominated the search for and validation of biomarkers of AKI. There are, however, also a few published reports on metabolites potentially prognostic of the development of AKI. The first of these metabolomic studies was performed on urine specimens obtained prior to surgery and at 4 hours and 12 hours after surgery from 40 children that underwent CPB surgery for correction of congenital cardiac defects [Beger et al. 2008]. Analysis of urine specimens of the first twenty patients enrolled by means of reverse-phase ultra-performance liquid chromatography (RP-UPLC) coupled to time-of-flight mass spectrometry (TOFMS) yielded good separation of AKI from non-AKI patients in PCA for both the 4-hour and the 12-hour urine specimens with a sensitivity and specificity of 95% each. A loading plot of PC2 versus PC3 identified an ion with a mass-to-charge (m/z ratio) of 261.01 as a potential biomarker, which was identified as the sulfate conjugate of homovanillic acid, a major metabolite of catecholamines such as epinephrine and norepinephrine that are routinely administered as inotropic agents after weaning from CPB to improve cardiac output [Gillies et al. 2005]. The subsequent determination of urinary homovanillic acid sulfate in all 40 patients enrolled yielded a cut-off value of 24 ng/µl at 12 hours after surgery that was capable of discriminating AKI from non-AKI patients with a sensitivity and specificity of 90% and 95% (AUC of 0.95), respectively. This was the more impressive as increases in SCr from baseline by $\geq$ 50% occurred as late as 48-72 hours after surgery in 11 out of 21 patients that developed AKI, thus mimicking the performance of NGAL. However, to date the validity of homovanillic acid sulfate as a prognostic marker has been neither confirmed for an independent cohort of pediatric patients nor demonstrated for adult patients. Almost all patients of the present study were treated with catecholamines (Appendix II section 7.2.1 Table 7.1). However, neither homovanillic acid nor its sulfate conjugate was detected in the 1D ^{1}H and 2D ^{1}H-^{13}C NMR plasma spectra. Also in the corresponding spectra of urine specimens collected before, 4 hours, and 24 hours after surgery these compounds could not be detected.
The utility of drugs, respectively metabolites and conjugates thereof, in prognosticating AKI is also demonstrated in the present study, which found the glucuronide conjugate of propofol to be the best prognostic indicator of acute kidney injury, outperforming even creatinine (Figure 5.2). The total administration time of propofol showed only limited prognostic value (Figure 5.2). Propofol is not known to cause AKI itself and the present dataset does not indicate otherwise. Actually, given its antioxidant and cytoprotective properties, propofol is believed

[14] The following section has already been published in [Zacharias et al. 2015], and [Hochrein et al. 2015] in a slightly altered version.

to protect the kidneys against ischemia and reperfusion injury [Snoeijs et al. 2011]. Plasma levels of propofol-glucuronide appear to serve as a surrogate marker of general kidney function similar to the antifibrinolytic agent tranexamic acid in the 24 h urinary NMR fingerprints as discussed in section 5.1.1.
A second study, which gave no details on the cause of AKI, applied ultra-performance reversed-phase liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer (UPLC /QTOFMS) to 17 serum specimens each collected from healthy subjects and patients with newly diagnosed AKI, whose serum creatinine levels at the time of enrollment had increased 1.5-9.6 times over their baseline levels [Sun et al. 2012]. In addition to AKI, patients suffered from various co-morbidities including congestive heart failure, diabetes mellitus, hypertension, coronary artery disease, hyperlipidemia, and peripheral vascular disease. Metabolites, whose serum levels were increased in comparison to the controls, included creatinine, acylcarnitines, methionine, homocysteine, pyroglutamate, asymmetric dimethylarginine, and phenylalanine, while the serum levels of several lysophosphatidyl cholines and arginine were decreased. Major limitations of that study over the present study are the lack of information on the cause of AKI, other than the absence of an obstruction of the urinary tract, the eventual stage of disease, and the time that had elapsed between the acute kidney injury and the collection of the serum specimen used for analysis. Therefore, the utility of the metabolites listed for the diagnosis and, even more so, prognostication of AKI in general and, particularly, in the context of cardiac surgery with CPB cannot be assessed.
Other than for creatinine, there is no overlap between the discriminating metabolites identified in serum by UPLC/QTOFMS and those found here in EDTA plasma by 1D ^{1}H-NMR. The use of EDTA as anticoagulant was the key to the determination of free calcium and magnesium levels, because both yielded upon complex formation with EDTA distinct signals in ^{1}H-NMR spectra. While plasma calcium levels did not differ significantly between the non-AKI (1.97 $\pm$ 0.14 mmol/l) and the AKI group (1.95 $\pm$ 0.12 mmol/l), a significant increase ($p = 1.0\text{e}^{-5}$) in plasma levels of Mg^{2+} from an average concentration of 1.05 $\pm$ 0.24 mmol/l in the non-AKI group to 1.29 $\pm$ 0.23 mmol/l in the AKI group was observed, most likely as a result of its use in treating cardiac arrhythmias.
Ischemic injury and systemic hypoperfusion, known to contribute to the pathophysiology of postoperative AKI (compare to section 4.1.3) [Rosner and Okusa 2006], may explain the significantly ($p = 3.0\text{e}^{-3}$) increased levels of plasma lactate in the AKI group. Acidosis may also be responsible for the elevated plasma and urine levels of hippuric acid in AKI patients, as discussed in section 5.1.1. Further, acidosis is also known to decrease renal excretion of citrate [Zuckerman and Assimos 2009] and may thus explain its increased plasma levels. The observed increase in tryptophan levels might not only be a consequence of reduced glomerular filtration, but also of reduced albumin binding due to the accumulation of competing solutes in plasma [Druml et al. 1994].
Prognostication based on NMR or mass spectrometric metabolite fingerprints is not feasible in a routine intensive care setting, which requires markers amenable to modern point-of-care testing to initiate swift therapeutic and preventive measures to treat and avoid complications such as AKI. Plasma creatinine, plasma Mg^{2+}, and plasma lactate are easily quantifiable by point-of-care technologies and their combined predictive accuracy of 77.0% (AUC 0.84) is only

slightly lower than that (80%, AUC 0.87) of the full metabolic fingerprinting dataset. For advanced stage disease, accuracy comes even close to 100%.
Regarding the analysis of intermediate cases of AKI, I trained a classifier only on data of patients where it was clear whether they had developed AKI or not, i.e. patients of AKIN stages 0, 2 and 3. Employing this classifier on data of AKIN 1 patients, Figure 5.3a) reveals that most of these patients received a score below 0.5 indicating that their metabolic profiles are similar to patients without AKI. This becomes also apparent by close inspection of Figure 7.3 in Appendix II section 7.2.6, which shows in the third column from the left in a heat-map representation the metabolic profiles of AKIN 1 patients that were falsely prognosticated by the Random Forest classifier. The metabolic profiles of these patients resemble those of patients, who did not develop AKI. It has been reported, that even small postoperative increases in serum creatinine of up to 0.5 mg/dl (44.2 µmol/l) are associated with a nearly threefold increase in 30-day mortality [Lassnigg et al. 2004]. For patients who suffered from AKI due to a variety of reasons it was shown that AKIN 1 was associated with an almost twofold increased risk of death [Bedford et al. 2014]. It remains to be investigated, whether patients with metabolic profiles indistinguishable from those of non-AKI patients have a better postoperative outcome than those, who were correctly classified as AKIN 1.
The most common risk factors for developing postoperative AKI include elevated preoperative serum creatinine levels and length of CPB use, as already discussed in section 4.1.3. Inspection of Table 7.1 in Appendix II section 7.2.1 shows, that length of CPB use (bypass time period) did not differ significantly between AKI and non-AKI patients for both urine and plasma cohort. Given, that prognostication of AKI based on length of CPB use yielded only an AUC of 0.55 and 0.52 for urine and plasma cohort, respectively, it is unlikely that length of CPB use was a major contributor to the development of AKI in the present study. There was, however, a significant difference in preoperative eGFR between AKI and non-AKI patients in both cohorts.

5.2 German Chronic Kidney Disease study

5.2.1 Introduction

After the successful investigation of novel urinary and plasma metabolic biomarkers for the early detection of acute kidney injury after cardiac surgery with CPB use, as detailed in section 5.1, I focused on gaining new insights into the pathophysiology and development of chronic kidney diseases.
As elaborately outlined in section 3.1, CKD is one of the largest burdens of global health [Jha et al. 2013]. The heterogeneity of its disease pattern hinders effective patient care [Titze et al. 2015], and the demand for clinical trials has not been satisfied yet [Eckardt et al. 2013, Titze et al. 2015]. Especially large-scale studies focusing on CKD patients under nephrological care are still scarce [Titze et al. 2015], although two cohorts with about 3000 CKD patients each have been recruited in the US (Chronic Renal Insufficiency Cohort (CRIC) Study) [Feldman et al. 2003] and in Japan (Chronic Kidney Disease Japan Cohort (CKD-JAC)) [Imai et al. 2010], respectively.
The German Chronic Kidney Disease (GCKD) study includes the currently worldwide largest CKD cohort with about 5200 patients enrolled from March 2010 to March 2012 with a large observation period of up to ten years [Titze et al. 2015]. It was designed as a national prospective observational cohort study with nine study centers throughout Germany (DRKS 00003971) [Eckardt et al. 2012, Titze et al. 2015]. Enrolled patients were aged between 18 and 74 years and had to exhibit either an eGFR of 30-60ml/min per 1.73m^2 or an eGRF above 60ml/min per 1.73m^2 and 'overt' albuminuria/proteinuria [Titze et al. 2015], compare to section 4.1.4. eGFR was usually determined employing the 4-variable MDRD equation [Titze et al. 2015], compare to section 4.1.2. Here, albuminuria/proteinuria is defined by either a urinary ACR above 300mg/g, an albuminuria of more than 300mg/day, a urinary protein/creatinine ratio of more than 500mg/g, or a proteinuria of more than 500mg/day [Titze et al. 2015]. Patients with solid organ or bone marrow transplantation, active malignancy within 24 months prior to screening, heart failure New York Heart Association Stage IV, i.e. patient experiences severe cardiovascular disease including severe limitations in physical activity, legal attendance or inability to provide consent, and/or non-Caucasian ethnicity have been excluded [Titze et al. 2015].
Information about numerous clinical chemistry parameters, sociodemographic factors, medical and family history, medications, quality of life, comorbidities, etc. have been collected by study teams during the patients visit to a nephrologist's practice or an outpatient unit of the regional centers. More details can be found in [Titze et al. 2015, Eckardt et al. 2012].
Detailed baseline clinical and demographic characteristics are given in [Titze et al. 2015]. In summary, the GCKD cohort comprised 60% men, the mean patient age was (60 $\pm$ 12) years, a mean eGFR of (47 $\pm$ 17)ml/min per 1.73m^2 and a median urinary ACR of 51 (9 - 392)mg/g was reported [Titze et al. 2015]. The most frequent leading causes of CKD comprised vascular nephropathy (23%), primary glomerulopathy (19%), and diabetic nephropathy (15%), whereas the leading cause was unknown in up to 20% of cases. 35% of patients additionally suffered from diabetes [Titze et al. 2015]. The prevalence of cardiovascular disease was 32% and prevalent

risk factors were identified as smoking, obesity, as well as positive family history of diabetes, cardiovascular and/or renal disease [Titze et al. 2015]. Although anti-hypertensive drugs were frequently administered, almost half of the patient cohort still exhibited an office blood pressure above 140/90mmHg [Titze et al. 2015].
Plasma, serum, blood, and spot-urine specimens were collected, processed and further distributed to the following collaborators. Synlab (Heidelberg, Germany) assessed routine clinical chemistry parameters, Central Lab (University Hospital Erlangen, Germany) performed Hb and glycated hemoglobin (HbA_{1c}) measurements, and DNA extraction as well as biofluid storage for future analyses was conducted at the Central Biobank of the University Hospital Erlangen. The latter provided us with frozen aliquots of plasma specimens for metabolic analyses performed by NMR spectroscopy as well as MS spectrometry, conducted by the group of Dr. Katja Dettmer.
The large observation time period and the huge study population enables deductions about the future disease development based on clinical data collected at the baseline and second follow-up (FU2) time-point.
For this study, I aimed at the prediction of present and future kidney performance reflected by the respective eGFR and serum creatinine/cystatin C levels by assessing multiple regression analyses models for bucket intensities derived from 1D ^{1}H NMR plasma spectra. The detection of novel biomarkers of kidney function asides from SCr and SCysC, as elaborately described in section 4.1.2, would offer improved patient care and better understanding of metabolic alterations in the context of evolving chronic kidney disease. Furthermore, I studied distinct plasma metabolic profiles of different renal diseases by employing *t*-tests. These investigations might again enable enhanced insights into the different pathophysiologies of these diseases and possibly lead to revised treatment procedures.

5.2.2 Materials and Methods

5.2.2.1 Sample preparation and NMR data acquisition

In total, 5129 plastic tubes, each automatically filled by a pipetting robot with one EDTA plasma aliquot collected at the baseline time-point, were sent to us from the University Hospital Erlangen. Each plastic tube was bar-coded by a unique sample ID. Specimens had been stored at -80°C until preparation for NMR spectroscopy.
Specimens were prepared for NMR measurements in collaboration with Claudia Samol. 400 µl of each unfiltered EDTA plasma specimen was mixed with 200 µl of 0.1 mol/l phosphate buffer at pH 7.4 and 50 µl of 0.75% (w) of the sodium salt of TSP (Sigma-Aldrich, Taufkirchen, Germany), solved in D_2O, as detailed in section 4.2.2.1. A cost-intensive filtering of the plasma specimens in order to remove proteins and lipids, as realized for the acute kidney injury study, see section 5.1.2.2, was omitted due to the large specimen cohort. Since the presence of broad protein signals might hinder the interpretation of metabolite signals, 1D ^{1}H spectra employing a CPMG pulse-sequence, compare to section 4.2.2.2, for the suppression of macromolecular signals were acquired for each plasma specimen. For each 1D ^{1}H CPMG spectrum, 128 scans were collected into 73728 data points employing the pulse program *cpmgpr1d.comp* (BrukerBioSpin

GmbH, Rheinstetten, Germany) with water suppression by presaturation during relaxation. Four dummy scans were acquired prior to measurement, the spectral width was 20.02 ppm, the relaxation delay was 4 s and the acquisition time amounted to 3.07 s. The filtering delay amounted to 0.08 s, resulting in a total acquisition time of about 16 min. I preprocessed every recorded 1D ^{1}H CPMG spectrum as described in section 4.2.2.3.

The baseline EDTA plasma specimen cohort of the GCKD study was measured by NMR spectroscopy from August 2014 until June 2015. The long-term stability of the spectrometer performance was controlled and evaluated by employing a reference plasma specimen from a healthy volunteer. This reference plasma specimen had been split into a sufficient number of aliquots immediately after decanting, compare to section 4.2.2.1, and stored at -80°C until measurement. For each weekly GCKD NMR data acquisition, a new aliquot was thawed at room temperature and 400 µl thereof were each filled into two NMR tubes. Again, 200 µl of 0.1 mol/l phosphate buffer at pH 7.4 and 50 µl of 0.75% (w) of the sodium salt of TSP (Sigma-Aldrich, Taufkirchen, Germany), solved in D_2O, were added to each NMR tube. One 1D ^{1}H CPMG spectrum for each of these freshly prepared plasma specimens were recorded at the beginning and end of the weekly NMR measurement period, respectively, typically spanning in total 69 h. We are aware that changes in metabolite concentration do occur for unfiltered plasma specimens stored at 4°C for more than 24 h [Klein 2011]. These changes are exemplarily highlighted for two reference plasma 1D ^{1}H CPMG spectra measured 69 h appart from each other in Fig. 5.4. They add to overall intragroup variance of investigated datasets and consequently can influence calculated p-values as well as statistical powers. However, no systematic differences for specific investigated specimen groups should be present in the GCKD NMR spectral data set, since NMR measurements were performed with thorough specimen randomization. Nevertheless, NMR signals, whose chemical shifts as well as intensities are significantly influenced by storage time, should be regarded with care.

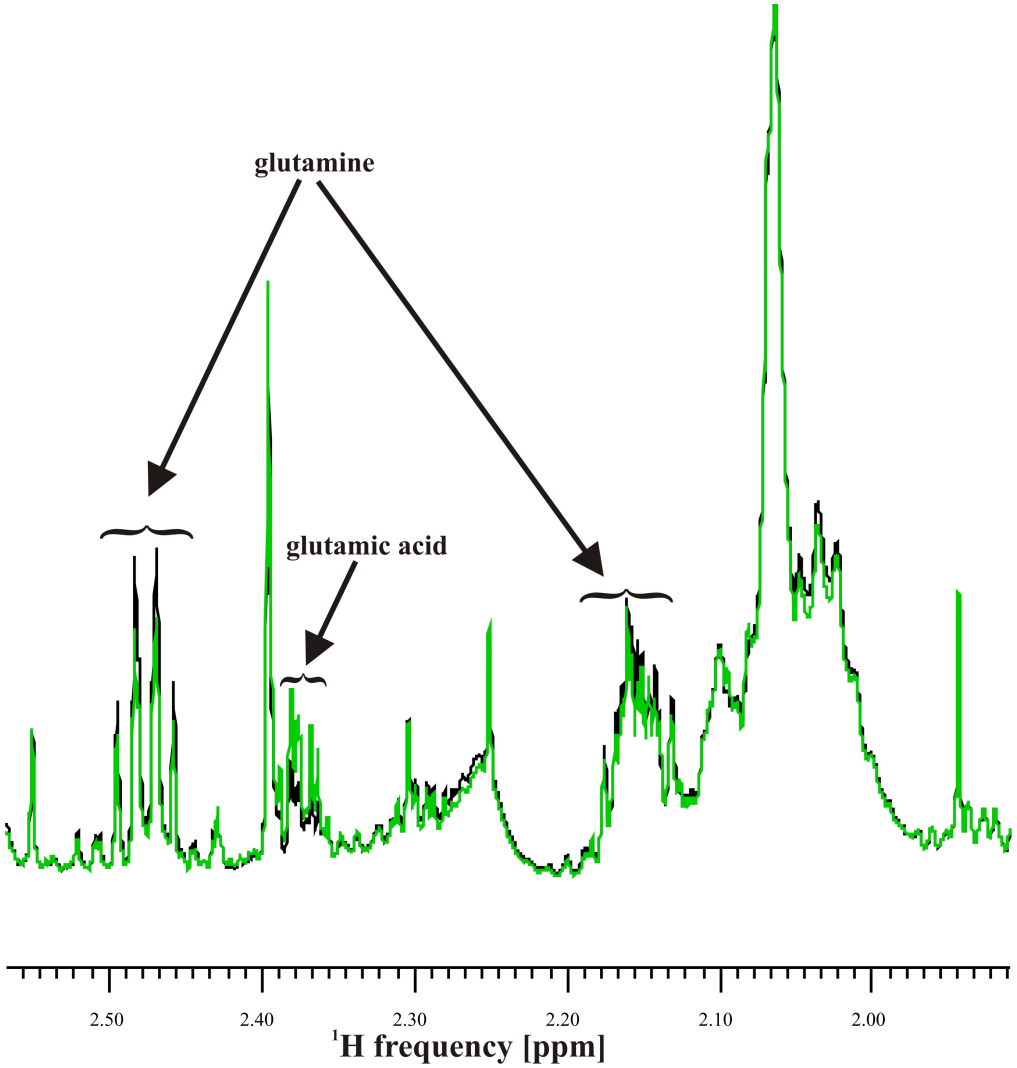


Figure 5.4: **Exemplary spectral comparison of two reference EDTA plasma specimens measured 69 h apart from each other.** The EDTA plasma 1D ^{1}H CPMG spectra acquired at the beginning and end of the measurement period are plotted in black and green, respectively. Only a small fraction of the spectra are shown with overall largest variations in metabolite concentrations observed for illustration purposes.

The conventional NMR reference substance TSP undergoes complex formation with plasma macromolecules, resulting into a significant diminishment and broadening of the TSP signal [Zacharias et al. 2013b)]. In fact, the signal variation of the TSP reference signal across all 1D ^{1}H NMR spectra of the complete GCKD EDTA plasma cohort defined as the relative standard deviation of the TSP integral amounted to approximateyl 15.0%. In comparison, the technical variability, defined as the signal variation of the TSP reference signal across all 1D ^{1}H NMR spectra of the control EDTA plasma specimen, was only 4.7%, i.e. significantly (p-value $< 10^{-16}$) lower than the TSP signal variability across different plasma specimens. This illustrates the fact that the TSP signal in 1D ^{1}H NMR spectra of unfiltered plasma is significantly

influenced by the specific macromolecule content of the respective specimen and therefore inappropriate as an internal standard for the reduction of variations in spectrometer performance, as applied for the AKI study detailed in section 5.1.2.5. Note that the technical variability here includes both the pipetting error and the observed variations in spectrometer performance over time, and is comparable to the pipetting error plus the spectrometer performance variability evaluated by Prof. Wolfram Gronwald and Claudia Samol for the AKI study, see section 5.1.2.5. As a consequence, we decided to further add 10 µl of 81.97 mmol/l formic acid as internal standard to each EDTA plasma specimen of the GCKD study cohort prior to NMR measurement. The utilization of formic acid as an alternative internal standard for NMR measurements of unfiltered plasma specimens has already been recommended in the literature, e.g. [Beckonert et al. 2007].

5.2.2.2 Spectral alignment and data normalization

Overall variations in chemical shifts were observed in some 1D ^{1}H CPMG spectra of the GCKD study cohort. They probably result from chemical shifts of the TSP reference signal induced by variable macromolecule content of the respective EDTA plasma specimen [Klein 2011]. Since the TSP signal was employed as reference signal for the spectral zero point, compare to section 4.2.2.1, shifts of the TSP NMR signal lead to overall shifts of the respective spectrum. These general shifts spanned a larger range than the usually employed bucket width of 0.01 ppm and can therefore not be compensated by this bucketing procedure. Consequently, I decided to eliminate these general inter-spectral offsets by aligning all investigated 1D ^{1}H CPMG spectra with respect to each other. This alignment procedure was automatically conducted employing *R*.
First, the NMR spectral data was subjected to a bucketing procedure as described in section 4.2.2.3. The spectral region from 9.5 - 0.5 ppm was evenly split into bins of 0.001 ppm width. As the TSP signal was significantly influenced by the specific macromolecule content of the respective unfiltered EDTA plasma specimen and therefore inappropriate as an internal standard for the reduction of variations in spectrometer performance, compare to section 5.2.2.1, I scaled all investigated spectral data relative to the formic acid signal from 8.5 - 8.46 ppm in order to reduce variations in spectrometer performance. Indeed, the signal variation of the formic acid reference signal across all 1D ^{1}H CPMG spectra acquired after the addition of formic acid as internal standard (in total 3206 spectra) defined as the relative standard deviation of the formic acid integral only amounted to 3.2%, in comparison to the relative standard deviation of the TSP integral for this specimen cohort of 15.0%. For further analysis, data were imported into *R* (Development Core Team 2009).
Now, for each of the 3206 with formic acid as internal standard acquired 1D ^{1}H CPMG spectra, the bucket with the greatest bucket intensity in the spectral region from 8.5 - 8.46 ppm, corresponding to formic acid, was picked. In the next step, all 1D ^{1}H CPMG spectra were aligned to each other with respect to the bucket with maximum intensity corresponding to formic acid. Fig. 5.5 depicts the formic acid singulet around 8.48 ppm as well as the alanine doublet around 1.49 ppm, which was chosen exemplarily, prior to (Fig. 5.5 a) and c), respectively) and after bucket adjustment (Fig. 5.5 b) and d), respectively). The NMR bucket at 8.4785 ppm was, in

most cases, i.e. 883 times, picked as the bucket with maximum intensity in the specified region. If one assumes a typical bucket width of roughly 0.01 ppm around 8.4785 ppm, the bucket with maximum peak intensity was picked 3120 times in the corresponding spectral region from 8.484 - 8.473 ppm. Consequently, only 86 of, in total 3206 1D ^{1}H CPMG spectra (i.e. about 3%) displayed an overall spectral shift greater than 0.01 ppm in comparison to the NMR spectra, for which the formic acid bucket with maximum intensity was picked in the spectral region from 8.4835 - 8.4735 ppm. One exemplary 1D ^{1}H CPMG spectrum for which the bucket at 8.4895 ppm was chosen as bucket with maximum intensity, one exemplary 1D ^{1}H CPMG spectrum with the maximum bucket intensity at 8.4785 ppm, and one exemplary 1D ^{1}H CPMG spectrum for which the bucket at 8.4635 ppm was chosen as bucket with maximum intensity are plotted in Fig. 5.5 in black, red, and green, respectively. After bucket alignment, both the formic acid singulet (Fig. 5.5 b)) as well as the alanine doublet (Fig. 5.5 d)) of all plotted spectra are aligned to each other.

After this alignment procedure, bucket intensities across ten buckets with a bucket width of 0.001 ppm each were fused together in one bucket of 0.01 ppm width by summing up the individual bucket intensities to facilitate easier data interpretation. Note that the corresponding bucket positions denote the middle of these fused buckets in ppm. The corresponding *R*-codes of bucket alignment and fusion procedure are given in Appendix I section 7.1.4.4.

Afterwards, the spectral region from 6.5 - 4.4 ppm containing the broad urea signal and the remaining water signal, as well as the NMR signals (3.69 - 3.6 ppm, 3.3 - 3.2 ppm) corresponding to free EDTA were excluded, resulting into a total number of 660 bins for each 1D ^{1}H CPMG spectrum.

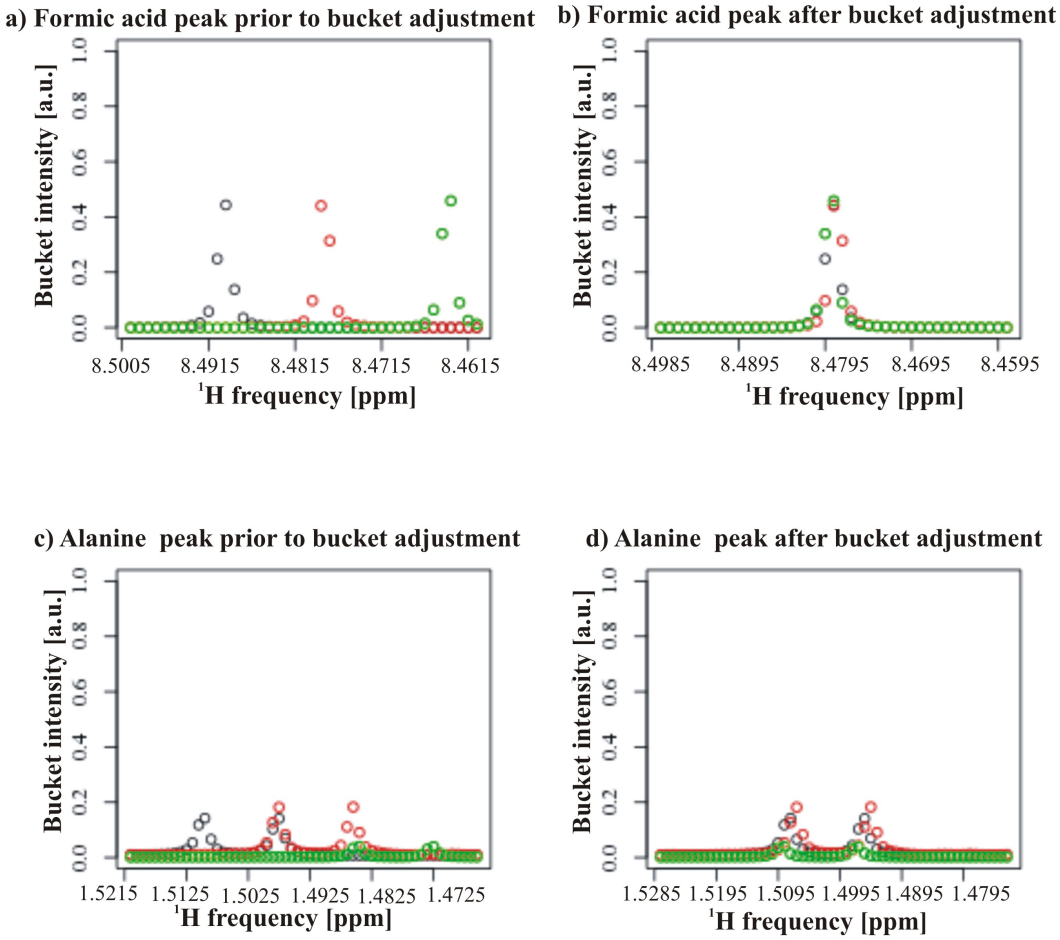


Figure 5.5: **Exemplary results of bucketing alignment of GCKD study cohort.** a) Formic acid signal of three representative 1D ^{1}H NMR bucket lists prior to bucket alignment. The overall shifts of the formic acid signal span a range larger than the usually applied bucket width of 0.01 ppm. b) Formic acid signal of three representative 1D ^{1}H NMR bucket lists after bucket alignment. The maximum of the formic acid peak is now at 8.4785 ppm for all spectra. c) Alanine doublet around 1.49 ppm of three representative 1D ^{1}H NMR bucket lists prior to bucket alignment. The overall shifts of the alanine signals span a range larger than the usually applied bucket width of 0.01 ppm. d) Alanine doublet of three representative 1D ^{1}H NMR bucket lists after bucket alignment. The alanine doublet is now between 1.50 - 1.48 ppm for all spectra. One exemplary 1D ^{1}H CPMG spectrum for which the bucket at 8.4895 ppm was chosen as bucket with maximum intensity, one exemplary 1D ^{1}H CPMG spectrum with the maximum bucket intensity at 8.4785 ppm, and one exemplary 1D ^{1}H CPMG spectrum for which the bucket at 8.4635 ppm was chosen as bucket with maximum intensity are plotted in black, red, and green, respectively.

5.2.2.3 Patient selection and characteristics

We received a clinical data file with sample and patient ID, clinical chemistry parameters, sociodemographic factors, medical and family history information, etc. collected for 5296 patients at the baseline, and for 4478 patients at the FU2 time-point, respectively, from the University Hospital Erlangen. In order to collect the required patient information for statistical analysis, the NMR sample IDs were matched with the corresponding sample IDs in the clinical data file. During the course of sample preparation for NMR measurements and sample ID match, several hundred EDTA plasma specimens and/or their respective NMR spectra had to be excluded from statistical data analysis out of various reasons, as illustrated in the flow-charts of Fig. 5.6.

54 plastic tubes did not contain sufficient, i.e. at least 100 µl, specimen material, and the respective EDTA plasma aliquots had to be excluded from NMR measurements. If plastic tubes only contained between 100 and 400 µl of EDTA plasma (3.6% of measured EDTA plasma specimens), the missing plasma volume was substituted with H_2O and the bucket intensities were multiplied with the respective dilution factors. Furthermore, 155 EDTA plasma aliquots had to be excluded due to pipetting irregularities. Consequently, NMR sample IDs for a total of 4920 individual EDTA plasma specimens collected at the baseline time-point could be compared to sample IDs in the clinical data file.

Clinical data had been collected for 5296 patients at the baseline time-point, however, for 77 patients, no sample ID had been provided and I consequently excluded these patients from the sample ID matching procedure, compare to Fig. 5.6a). By comparing the sample IDs of, in total 4920 available NMR sample IDs, with, in total 5219 reported sample IDs in the clinical baseline data file, I had to further exclude 64 NMR sample IDs and 363 patients in the clinical baseline data file, respectively, due to sample ID mismatch (Fig. 5.6a)). Consequently, a total of 4856 NMR sample IDs could be matched with their respective sample ID in the clinical baseline data file. Only NMR spectra acquired after the addition of formic acid as internal standard were included for the statistical analysis of this Ph.D. thesis. Therefore, I further excluded 1692 EDTA plasma specimens, for which a 1D ^{1}H CPMG spectra after the addition of formic acid as internal standard had not yet been available at the time-point of statistical analysis for this Ph.D. thesis. The GCKD study baseline sample cohort consequently comprises in total 3164 patients, for whom one EDTA plasma specimen collected at the baseline time-point had been measured with NMR spectroscopy and formic acid as internal standard, and their respective baseline sample ID could be matched between NMR spectrum and clinical baseline data file. Baseline patient characteristics corresponding to this sample cohort are given in Appendix III section 7.3.1 Table 7.5.

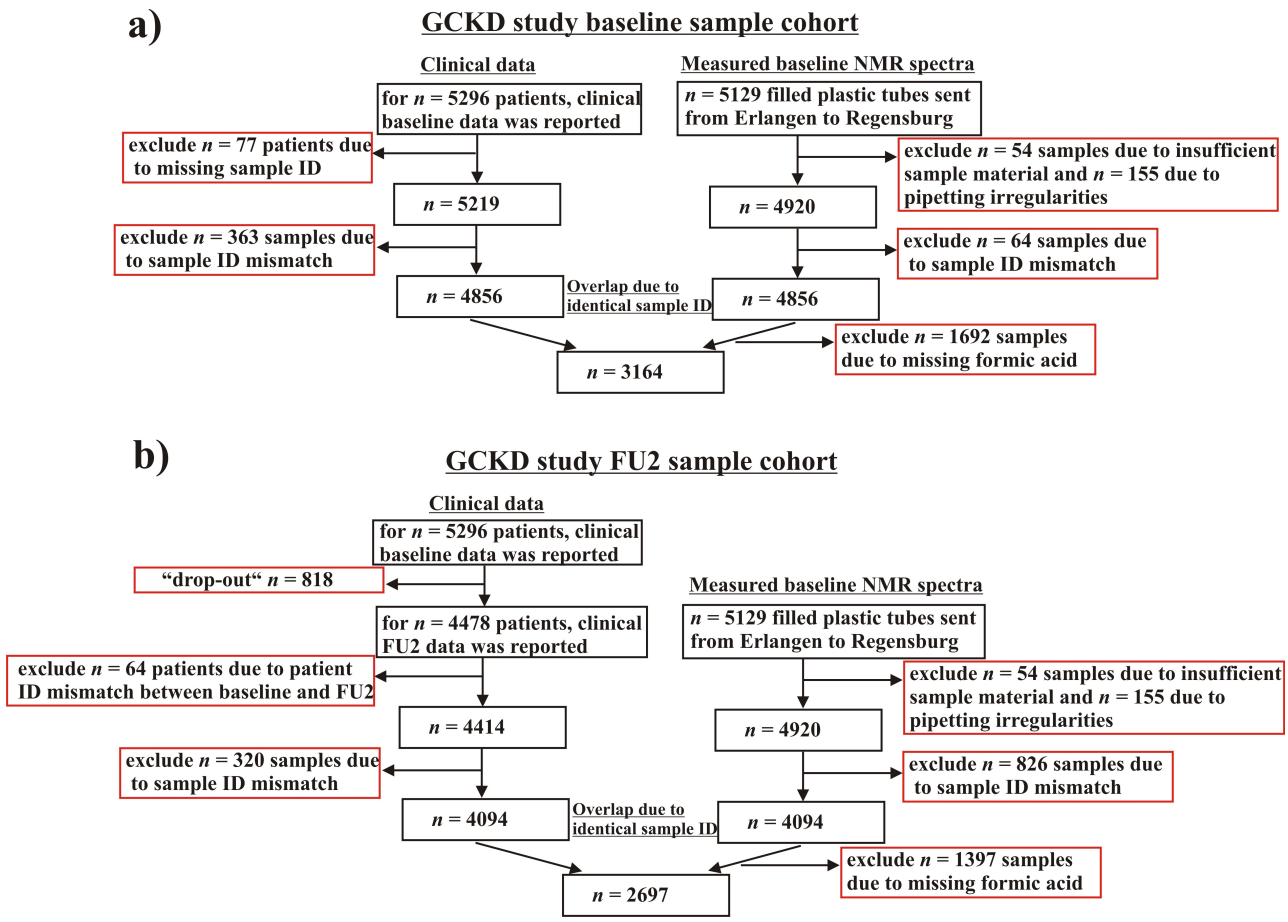


Figure 5.6: **Flow-charts illustrating sample exclusion procedures conducted during NMR sample preparation and sample ID match.** a) GCKD study sample cohort with clinical data collected at the baseline time-point. b) GCKD study sample cohort with clinical data collected at the FU2 time-point. More details are given in the text.

FU2 clinical data had been collected for, in total, 4478 patients. Consequently, a "drop-out" of 818 patients in comparison to the baseline clinical data collection has to be reported. Reasons for this "drop-out" include death and consent withdrawal. In order to match the baseline NMR sample IDs, I first had to match the respective patient IDs in the baseline and clinical FU2 data files, since the clinical FU2 data file did not contain the baseline sample IDs. Here, I had to exclude 64 patients in the FU2 clinical data file due to patient ID mismatch, compare to Fig. 5.6b). By now comparing the sample IDs of the remaining 4414 patients in the FU2 clinical data file with the 4920 baseline NMR sample IDs, I further excluded 826 NMR baseline sample IDs and 320 patients in the clinical FU2 data file, respectively, due to sample ID mismatch (Fig. 5.6b)). Moreover, for 1397 out of the matched 4094 EDTA plasma specimens a 1D ^{1}H CPMG spectrum after the addition of formic acid as internal standard had not yet been available at the time-point of statistical analysis for this Ph.D. thesis. They were consequently excluded from statistical data analysis for this Ph.D. thesis. Therefore, the GCKD study FU2 sample cohort comprises in total 2697 patients, for whom one EDTA plasma specimen collected at the

baseline time-point had been measured with NMR spectroscopy and formic acid as internal standard, and their respective baseline sample ID could be matched between NMR spectrum and clinical FU2 data file. Baseline as well as FU2 patient characteristics corresponding to this sample cohort are given in Appendix III section 7.3.1 Table 7.6.

5.2.2.4 Statistical data analysis

In order to guide decision making for the appropriate normalization technique, the for the AKI study developed strategy was utilized, compare to section 5.1.3.1. The Shapiro-Wilk normality test was applied to the GCKD study baseline and FU2 sample cohorts separately, and yielded in both cases significant p-values $< 10^{-16}$, indicating that the spectral data of both cohorts is not normally distributed. Consequently, common data normalization methods such as Quantile or VS normalization were not applied.

In order to investigate distinct metabolic differences between various leading renal diseases, I employed an ANOVA and subsequently two sample t-tests as detailed in section 4.3.1.3. For this analysis, I applied a $\log_2$ transformation to the formic acid scaled spectral bucket intensities to minimize heteroscedasticity, as explicitly discussed in section 5.1.3.1. The effect size and statistical power for all discriminating NMR features was determined with the R packages *compute.es* [Del Re 2013] and *pwr* [Champely 2015], respectively. Here, I employed the GCKD study baseline sample cohort. However, due to the fact that the leading renal disease of one patient had not been assigned, compare to Appendix III section 7.3.1 Table 7.5, I excluded this patient from the corresponding statistical analysis. For the comparison of specific leading renal diseases, I furthermore excluded patients with no leading renal disease as well as patients suffering from "other" leading renal diseases due to large group heterogenity. Note that all leading renal disease groups, which did not include at least 100 individual patients, were summarized in "other" leading renal diseases.

For the prediction of present and future kidney performance reflected by the patient's serum creatinine/cystatin C levels and the respective eGFR, I performed simple linear as well as multiple regression analyses employing the LASSO method, as introduced in section 4.3.1.3. The training of an individual regression model was performed employing a fixed training set comprising 2/3 of the complete sample cohort, compare to section 4.3.1.3. Its predictive performance was assessed on a fixed test set comprising the remaining 1/3 of the complete sample cohort. Regression analyses for the prediction of baseline and FU2 SCr, SCysC, and eGFR values were performed with the GCKD study FU2 sample cohort. Note that due to missing clinical parameters, 205 samples had to be excluded from the GCKD study FU2 sample cohort. The baseline as well as FU2 patient characteristics of training and test set for this sample cohort, compare to section 5.2.2.3, are given in Appendix III section 7.3.1 Table 7.7. Note that no significant difference of baseline and FU2 patient characteristics between training and respective test sets are reported. For the determination of multiple regression models, I employed $\log_2$ transformed bucket intensities derived from 1D ^{1}H NMR plasma spectra in order to remove heteroscedasticity. To preserve the linear relationship between $\log_2$ transformed bucket intensi-

ties and the response variables, i.e. SCr, SCysC, as well as eGFR values, I furthermore applied a $\log_2$ transformation on the respective response variables for the model training step. Note that the predicted response variables in the testing step are also $\log_2$ transformed. Therefore, the predicted response variables of the test set were subsequently inversely $\log_2$ transformed and the reported mean-squared error (*mse*) as well as the coefficient of determination R^2 are calculated between the originally not $\log_2$ transformed true and the inversely $\log_2$ transformed predicted response variables of the test set. Note that the *mse* for the training set, derived from an internal cross-validation procedure for determining the optimal λ λ_{min}, is also recorded, but refers to the $\log_2$ transformed response values of the training set. In order to compare the *mse* for the training set with the respective *mse* of the test set, I also report the *mse* between the $\log_2$ transformed true and predicted response variables for the test set. For each trained LASSO model, the number of employed NMR features with β coefficients unequal to zero is reported and the not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables are plotted in a $x-y$ - diagram, whereas a linear model fitted between these true and predicted response variables is also given in this diagram.
For the determination of simple linear regression models based on respective baseline clinical parameters, neither explanatory nor response variables were $\log_2$ transformed in the model training step. The reported coefficients of determination R^2 as well as the *mse* are calculated between the originally not $\log_2$ transformed true and predicted response variables of the test set. Again, a $x-y$ - diagram with a fitted linear model is plotted between these true and predicted response variables, as realized for the different LASSO models.

A comparison of the GFR estimation performance of the different employed estimation formulas was conducted by performing Pearson's correlation calculations, as well as applying Bland-Altman plots [Bland and Altman 1999, Bland and Altman 1986] utilizing the *R* package *MethComp* [Carstensen et al. 2015]. Metabolite identification was performed as detailed in section 4.3.2. The identification of lipid functional groups was conducted with the aid of specific peak lists provided by [Klein et al. 2011, Klein 2011].

5.2.3 Results

5.2.3.1 Data acquisition and appropriate preprocessing of 1D ^{1}H NMR spectra of unfiltered EDTA plasma

For 3206 unfiltered EDTA plasma specimens collected at the baseline time-point, 1D ^{1}H NMR spectra were acquired employing a CPMG pulse sequence to suppress broad NMR signals arising mainly from protons in proteins, whereas formic acid had been added to each investigated specimen as internal standard. An exemplary 1D ^{1}H CPMG spectrum of an unfiltered EDTA plasma specimen of the GCKD study is shown in Fig. 5.7. For reasons of clarity, only a subset of prominent compounds is assigned. The CPMG pulse sequence yields a well resolved 1D spectrum. Although the spectrum lacks broad signals mainly arising from protons in proteins, it still includes a number of broad NMR peaks from protons in lipids, especially in the region from 2.5 - 0.7 ppm. Furthermore, one has to note that the signal intensities in a 1D ^{1}H CPMG

spectrum cannot be directly compared to corresponding signal intensities in a 1D ^{1}H NOESY spectrum acquired from the same specimen, since the former are, in general, smaller than the latter due to overall NMR signal decay during the refocusing period.
As detailed in sections 5.2.2.1 and 5.2.2.2, the TSP signal proved to be inappropriate as an internal standard to account for variations in spectrometer performance due to its dependence on the macromolecule content of the individual unfiltered EDTA plasma specimen. Consequently, a scaling of bucket intensities to the spectral region of formic acid, which had been added as an alternative internal standard, was applied in order to account for variations in spectrometer performance, see section 5.2.2.2.

For about 3% of all 1D ^{1}H CPMG spectra acquired with formic acid as internal standard, global peak shifts larger than 0.01ppm were noticed. These shifts probably arise from offsets of the TSP reference signal, which was employed to denote the zero point of the spectrum. These offsets were introduced by the binding of the TSP reference substance to macromolecules present in the unfiltered EDTA plasma specimens, as already reported in [Klein 2011]. In order to compensate these global peak shifts, all 3206 1D ^{1}H NMR spectra acquired with formic acid as internal standard were properly aligned to each other prior to further statistical data analysis, as detailed in section 5.2.2.2.

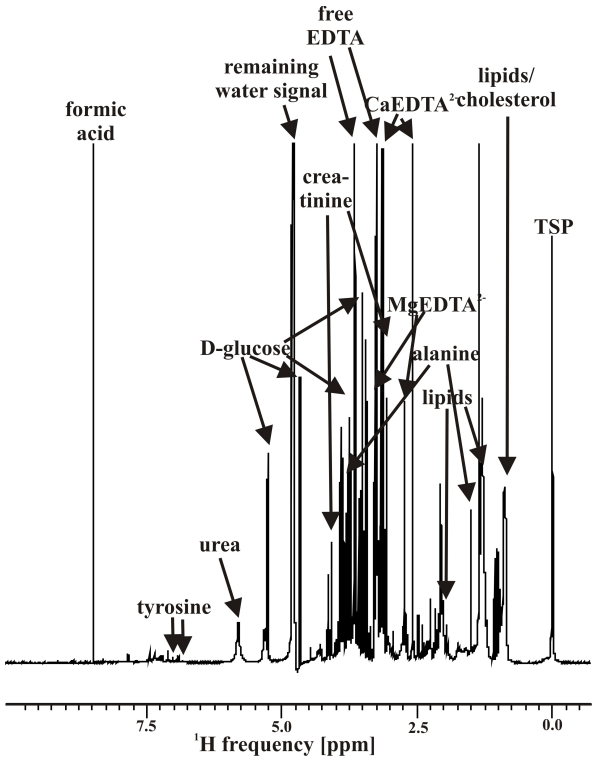


Figure 5.7: **Exemplary 1D ^{1}H CPMG NMR spectrum of an EDTA plasma specimen of the GCKD study collected at the baseline time-point.** Some prominent NMR signals are assigned to the corresponding compounds. Note that due to signal overlap, only a small subset of identified compounds is exemplarily assigned in this representative spectrum.

5.2.3.2 Specific metabolic fingerprints of various leading renal diseases

In order to detect statistically significant differences between 6 individual leading renal diseases, I performed an ANOVA including all 6 groups comprising, in total, 2305 individual 1D ^{1}H spectral data sets, as outlined in section 5.2.2.4. To remove heteroscedasticity, the NMR bucket intensities had been $\log_2$ transformed. The corresponding F-test yielded 403 buckets with a B/H-adjusted p-value < 0.05. I subsequently conducted 15 individual pairwise group comparisons by means of Student's t-tests. Table 5.4 lists the number of significantly differentiating bins (p-value B/H-adjusted < 0.05) for all conducted group comparisons. The spectral positions, fold changes, unadjusted as well as B/H-adjusted p-values, statistical powers, IDs, and corresponding identified metabolites for group comparisons yielding statistically significant B/H-adjusted p-values are listed in Appendix III section 7.3.2 Tables 7.8 to 7.21.
In summary, significant differences in D-glucose concentration of EDTA plasma specimens of patients suffering from diabetic nephropathy in comparison to patients suffering from another leading renal disease are apparent in Appendix III section 7.3.2 Tables 7.8 to 7.12. Moreover, an up-regulation of lipids and cholesterol in the EDTA plasma specimens of patients with glomerulonephritis in comparison to patients suffering from another leading renal disease is shown in Appendix III section 7.3.2 Tables 7.8 and 7.13 to 7.15. The EDTA plasma concentrations of D-glucose were however down-regulated in patients suffering from glomerulonephritis in comparison to patients suffering from hypertensive nephropathy (see Appendix III section 7.3.2 Table 7.16). An up-regulation of lipids in the EDTA plasma specimens of patients with hypertensive nephropathy is reported in comparison to patients suffering from hereditary diseases, interstitial nephropathy as well as systemic diseases, respectively.

Note that the statistical power of all highly significant metabolites was above 80%, indicating that the reported differences between these leading renal diseases are truly present in the general population.

	DN[a]	**GN**[b]	**Hereditary disease**[c]	**IN**[d]	**Systemic disease**[e]	**HN**[f]
DN[a]	-	279	320	310	350	234
GN[b]	279	-	184	47	136	130
Hereditary disease[c]	320	184	-	3	3	223
IN[d]	310	47	3	-	0	135
Systemic disease[e]	350	136	3	0	-	251
HN[f]	234	130	223	135	251	-

Table 5.4: **Numbers of significantly different NMR buckets (B/H-adjusted p-values < 0.05 according to FDR $< 5\%$) for group comparisons between 6 different leading renal diseases by means of Student's t-tests.** The individual spectral positions, IDs, log(Fold-changes), p-values both unadjusted and B/H-adjusted, statistical power, as well as correspondingly identified compounds of these discriminating NMR features are listed in Appendix III section 7.3.2. [a]DN had been assigned by our collaboration partners if patient suffered from diabetes mellitus or other diabetic nephropathies. [b]GN had been assigned by our collaboration partners if patient suffered from primary glomerulonephritis. [c]Hereditary disease had been assigned by our collaboration partners if patient suffered from ADPKD, Fabry disease, or other hereditary diseases. [d]IN had been assigned by our collaboration partners if patient suffered from interstitial nephropathy, analgesic nephropathy, or other interstitial nephropathies. [e]Systemic disease had been assigned by our collaboration partners if patient suffered from granulomatosis with polyangiitis, microscopic polyangiitis, systemic lupus erythematosus, scleroderma, hemolytic-uremic syndrome, thrombotic thrombocytopenic purpura, gout, tuberculosis, amyloidosis, sarcoidosis, or other systemic diseases. [f]HN had been assigned by our collaboration partners if patient suffered from renal artery stenosis, nephrosclerosis, kidney infarction, or other hypertensive nephropathies. Abbreviations: DN, diabetic nephropathy; GN, glomerulonephritis; HN, hypertensive nephropathy; IN, interstitial nephropathy.

5.2.3.3 Prediction of present and future kidney performance

For the prediction of baseline and FU2 kidney performance, reflected by the respective SCr, SCysC, and the GFR values estimated employing the MDRD4 ($eGFR_{mdrd4}$) [Levey et al. 1999], the CKD-EPI crea ($eGFR_{ckdepi\ crea}$) [Levey et al. 2009, Inker et al. 2012], the CKD-EPI cys ($eGFR_{ckdepi\ cys}$) [Inker et al. 2012], and the CKD-EPI crea cys ($eGFR_{ckdepi\ crea\ cys}$) [Inker et al. 2012] formula, respectively, I utilized the GCKD study FU2 sample cohort. As for 205 patients, at least one of these clinical parameters had not been assigned in the clinical data file, compare to Appendix III section 7.3.1 Table 7.6, I excluded these patients from the following statistical data analysis. Consequently, the complete specimen cohort for the prediction of present and future kidney performance comprises in total 2492 individual patients, for whom SCr, SCysC, and eGFR values were reported at both the baseline and the FU2 time-point and 1D ^{1}H CPMG spectra had been acquired after the addition of formic acid as internal standard, compare to sections 5.2.2.1, 5.2.2.2, and 5.2.2.3.

First, I compared the GFR estimation performance of the different employed estimation formulas by applying Pearson's correlation calculations, as well as Bland-Altman plots [Bland and Altman 1999, Bland and Altman 1986]. The corresponding results, including Pearson's correlation coefficient r and the coefficient of determination $R^2 = r^2$ are given in Table 5.5. $x - y$ scatter plots of the compared eGFR values with fitted simple linear regression lines and corresponding Bland-Altman plots are given in Appendix III section 7.3.3 Fig. 7.4.

a) Baseline eGFR values				
	$eGFR_{mdrd4}$	**$eGFR_{ckdepi\ crea}$**	**$eGFR_{ckdepi\ cys}$**	**$eGFR_{ckdepi\ crea\ cys}$**
$eGFR_{mdrd4}$	$r = 1$; $R^2 = 1$	$r = 0.987$; $R^2 = 0.973$	$r = 0.767$; $R^2 = 0.589$	$r = 0.922$; $R^2 = 0.849$
$eGFR_{ckdepi\ crea}$	$r = 0.987$; $R^2 = 0.973$	$r = 1$; $R^2 = 1$	$r = 0.784$; $R^2 = 0.614$	$r = 0.933$; $R^2 = 0.871$
$eGFR_{ckdepi\ cys}$	$r = 0.767$; $R^2 = 0.589$	$r = 0.784$; $R^2 = 0.614$	$r = 1$; $R^2 = 1$	$r = 0.953$; $R^2 = 0.908$
$eGFR_{ckdepi\ crea\ cys}$	$r = 0.922$; $R^2 = 0.849$	$r = 0.933$; $R^2 = 0.871$	$r = 0.953$; $R^2 = 0.908$	$r = 1$; $R^2 = 1$
b) FU2 eGFR values				
	$eGFR_{mdrd4}$	**$eGFR_{ckdepi\ crea}$**	**$eGFR_{ckdepi\ cys}$**	**$eGFR_{ckdepi\ crea\ cys}$**
$eGFR_{mdrd4}$	$r = 1$; $R^2 = 1$	$r = 0.990$; $R^2 = 0.980$	$r = 0.821$; $R^2 = 0.674$	$r = 0.938$; $R^2 = 0.880$
$eGFR_{ckdepi\ crea}$	$r = 0.990$; $R^2 = 0.980$	$r = 1$; $R^2 = 1$	$r = 0.837$; $R^2 = 0.700$	$r = 0.948$; $R^2 = 0.899$
$eGFR_{ckdepi\ cys}$	$r = 0.821$; $R^2 = 0.674$	$r = 0.837$; $R^2 = 0.700$	$r = 1$; $R^2 = 1$	$r = 0.966$; $R^2 = 0.933$
$eGFR_{ckdepi\ crea\ cys}$	$r = 0.938$; $R^2 = 0.880$	$r = 0.948$; $R^2 = 0.899$	$r = 0.966$; $R^2 = 0.933$	$r = 1$; $R^2 = 1$

Table 5.5: **Method comparison of different GFR estimation equations for a) baseline, and b) FU2 eGFR values.** Given are Pearson's correlation coefficients r and coefficients of determination $R^2 = r^2$ for the different comparisons. Corresponding $x - y$ scatter plots of the compared eGFR values including equations of the fitted simple linear regression lines and Bland-Altman plots are given in Appendix III section 7.3.3 Fig. 7.5. Abbreviations: eGFR, estimated glomerular filtration rate; $eGFR_{ckdepi\ crea}$, eGFR based on CKD-EPI crea formula; $eGFR_{ckdepi\ crea\ cys}$, eGFR based on CKD-EPI crea cys formula; $eGFR_{ckdepi\ cys}$, eGFR based on CKD-EPI cys formula; $eGFR_{mdrd4}$, eGFR based on MDRD4 formula; FU2, second follow-up.

This method comparison for the four different GFR estimation equations indicates an almost

perfect consensus between the MDRD4, CKD-EPI crea, and CKD-EPI crea cys formulas indicated by large Pearson's correlation coefficients around 0.9 and corresponding coefficients of determination around 0.87 for both time-points. Furthermore, the limits of agreement shown in the Bland-Altman plots in Appendix III section 7.3.3 Fig. 7.4 are very close to the mean difference between the respective compared methods for the aforementioned comparisons. Estimating the GFR values utilizing the CKD-EPI cys formula in comparison to the MDRD4 and the CKD-EPI crea formula, respectively, seem to, however, yield worse agreement between the compared methods for both time-points with Pearson's correlation coefficients only around 0.77 and 0.82, and corresponding coefficients of determination around 0.6 and 0.7 for baseline and FU2 time-point, respectively. This apparent deviation between eGFR values either based on SCr or SCysC values is furthermore reflected by comparable Pearson's correlation coefficients and coefficients of determination between SCr and SCysC values measured at the baseline ($r = 0.713$, $R^2 = 0.508$) and FU2 ($r = 0.804$, $R^2 = 0.647$) time-point, respectively, given in Appendix III section 7.3.3 Fig. 7.5.

In order to assess the relationship between 1D ^{1}H NMR data of baseline EDTA plasma specimens and present kidney function, I conducted multiple regression analyses employing the LASSO method with the complete NMR spectral feature set of 660 individual bucket intensities. The corresponding results are given in Table 5.6.
In order to illustrate the importance of creatinine as an explanatory variable in the derived LASSO models, I further performed multiple regression analyses employing the LASSO method after the exclusions of all creatinine buckets in the 1D ^{1}H CPMG spectra, corresponding to the spectral regions of 4.117 - 4.037 ppm and 3.117 - 3.017 ppm, leaving, in total, 641 individual NMR buckets. The LASSO method was chosen since it yields, in comparison to ridge regression, rather small regression models, from which biological interpretations can be more easily drawn, see section 4.3.1.3.
For regression analysis, I split the complete cohort of 2492 specimens into an exclusive training and test set of 1661 and 831 specimens, respectively. The baseline and FU2 patient characteristics of the training and test set are given in Appendix III section 7.3.1 Table 7.7. No significant differences of these clinical parameters exist between training and test set.
As detailed in section 5.2.2.4, NMR feature intensities as well as response variables were $\log_2$ transformed in the model training step of the multiple regression analyses to remove heteroscedasticity and preserve the linear relationship between explanatory and response variables. Since the predicted response variables in the testing step are also $\log_2$ transformed, they were subsequently inversely $\log_2$ transformed and the reported *mse* as well as the R^2 values are calculated between the originally not $\log_2$ transformed true and the inversely $\log_2$ transformed predicted response variables of the test set. Note that the *mse* for the training set refers to the $\log_2$ transformed response values. For comparison, I also report the *mse* between the $\log_2$ transformed true and predicted response variables for the test set. For each trained LASSO model, the number of employed NMR features with β coefficients unequal to zero is reported. The not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables are plotted in scatter plots displayed in Appendix III section 7.3.3 Fig. 7.6 to 7.7, including a linear model fitted between these true and predicted response variables.

The LASSO regression analyses for the prediction of baseline Synlab clinical chemistry parameters including the complete set of 660 NMR buckets showed very good performance for SCr values with a high R^2 of 0.936 (mse on test set = 0.013 mg/dl) and slightly minor performance for SCysC with R^2 of 0.739 (mse on test set = 0.052 mg/l) on the independent test set. After the exclusion of NMR creatinine signals, the coefficients of determination R^2 significantly dropped and the corresponding mse values on both training and test set significantly increased for the prediction of baseline Synlab SCr values. Furthermore, the numbers of NMR buckets with β coefficients significantly increased after the exclusion of NMR buckets corresponding to creatinine. The exclusion of all NMR creatinine signals only slightly worsened the performance of the corresponding LASSO model with respect to baseline Synlab SCysC values.
The performance of LASSO models derived from 660 individual baseline EDTA plasma 1D ^{1}H CPMG bucket intensities for the prediction of baseline eGFR values likewise showed good R^2 values ranging from 0.727 for $\text{eGFR}_{\text{ckdepi cys}}$ values to 0.860 for $\text{eGFR}_{\text{ckdepi crea}}$ values. Again, the exclusion of all NMR signals corresponding to creatinine prior to LASSO model determination yielded significantly worse prediction performances with respect to eGFR values based on SCr (e.g. increase of over 50% of the mse on both training and test set for both $\text{eGFR}_{\text{mdrd4}}$ and $\text{eGFR}_{\text{ckdepi crea}}$, respectively). For $\text{eGFR}_{\text{ckdepi cys}}$ response variables, the exclusion of all creatinine signals only yielded a marginally worse performance of the new LASSO model (increase of about 12% of the mse on both training and test set).

Baseline response variables				
	R^2	cv *mse* on training set	*mse* on test set	Number of employed NMR bins
Synlab SCr				
Multiple regression with 660 NMR bins	0.936	0.011 ± 0.002	0.013(mg/dl)^2 / 0.012	131
Multiple regression with 641 NMR bins	0.677	0.059 ± 0.003	0.068(mg/dl)^2 / 0.053	465
Synlab SCysC				
Multiple regression with 660 NMR bins	0.739	0.043 ± 0.002	0.052(mg/l)^2 / 0.042	211
Multiple regression with 641 NMR bins	0.709	0.050 ± 0.002	0.058(mg/l)^2 / 0.048	231
eGFR$_{\text{mdrd4}}$				
Multiple regression with 660 NMR bins	0.843	0.033 ± 0.002	39.27(ml/min/1.73m^2)^2 / 0.032	273
Multiple regression with 641 NMR bins	0.680	0.077 ± 0.004	80.19(ml/min/1.73m^2)^2 / 0.069	410
eGFR$_{\text{ckdepi crea}}$				
Multiple regression with 660 NMR bins	0.860	0.036 ± 0.002	41.92(ml/min/1.73m^2)^2 / 0.036	295
Multiple regression with 641 NMR bins	0.710	0.083 ± 0.004	87.00(ml/min/1.73m^2)^2 / 0.074	411
eGFR$_{\text{ckdepi cys}}$				
Multiple regression with 660 NMR bins	0.727	0.080 ± 0.004	101.10(ml/min/1.73m^2)^2 / 0.079	233
Multiple regression with 641 NMR bins	0.698	0.090 ± 0.004	112.28(ml/min/1.73m^2)^2 / 0.086	286
eGFR$_{\text{ckdepi crea cys}}$				
Multiple regression with 660 NMR bins	0.837	0.042 ± 0.002	49.08(ml/min/1.73m^2)^2 / 0.040	280
Multiple regression with 641 NMR bins	0.768	0.065 ± 0.004	70.28(ml/min/1.73m^2)^2 / 0.057	356

Table 5.6: **Results of regression analyses for prediction of baseline SCr, SCysC, and eGFR values.** Multiple regression analyses employing the LASSO method with both the complete baseline NMR spectral feature set of 660 individual bucket intensities, and with a baseline NMR spectral feature set of 641 individual bucket intensities after the exclusion of all NMR buckets corresponding to creatinine, respectively, were conducted. More details are given in the text. Abbreviations: cv, cross-validated; eGFR, estimated glomerular filtration rate; eGFR$_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; eGFR$_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; eGFR$_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; eGFR$_{\text{mdrd4}}$, eGFR based on MDRD4 formula; *mse*, mean-squared error; SCr, serum creatinine; SCysC, serum cystatin C.

The reliable prediction of future kidney performance in the context of chronic kidney disease is of crucial importance for sufficient patient treatment. Therefore, I repeated the aforementioned multiple regression analyses employing either all 660 baseline NMR buckets or 641 baseline NMR buckets after the exclusion of spectral regions belonging to creatinine, respectively, with respect to FU2 clinical parameters. Furthermore, I performed simple linear regression analyses with respect to FU2 clinical parameters now based on the respective baseline clinical parameters. This means, that, e.g. in the case of Synlab SCr, a simple linear regression model is trained with the baseline Synlab SCr values as explanatory variables and the FU2 Synlab SCr values as response variables on the training set, and consequently, its predictive performance with respect to the FU2 Synlab SCr values is assessed on the test set. Note that in this case, both explanatory as well as response variables were not $\log_2$ transformed during the model training step.

The corresponding results are given in Table 5.7 and Appendix III section 7.3.3 Fig. 7.8 to 7.10.

FU2 response variables				
	R^2	cv *mse* on training set	*mse* on test set	Number of employed NMR bins
Synlab SCr				
Multiple regression with 660 NMR bins	0.498	0.101 ± 0.005	0.204(mg/dl)^2 / 0.097	113
Multiple regression with 641 NMR bins	0.383	0.146 ± 0.006	0.252(mg/dl)^2 / 0.131	136
Simple regression with baseline Synlab SCr concentrations	0.517	-	0.200(mg/dl)^2	-
Synlab SCysC				
Multiple regression with 660 NMR bins	0.492	0.103 ± 0.005	0.198(mg/l)^2 / 0.099	102
Multiple regression with 641 NMR bins	0.477	0.113 ± 0.005	0.205(mg/l)^2 / 0.108	254
Simple regression with baseline Synlab SCysC concentrations	0.589	-	0.158(mg/l)^2	-
eGFR$_{\text{mdrd4}}$				
Multiple regression with 660 NMR bins	0.586	0.153 ± 0.007	100.97(ml/min/1.73m^2)^2 / 0.122	84
Multiple regression with 641 NMR bins	0.437	0.201 ± 0.008	139.39(ml/min/1.73m^2)^2 / 0.155	155
Simple regression with baseline eGFR$_{\text{mdrd4}}$ values	0.643	-	87.94(ml/min/1.73m^2)^2	-
eGFR$_{\text{ckdepi crea}}$				
Multiple regression with 660 NMR bins	0.612	0.167 ± 0.007	116.88(ml/min/1.73m^2)^2 / 0.133	73
Multiple regression with 641 NMR bins	0.463	0.219 ± 0.008	164.01(ml/min/1.73m^2)^2 / 0.170	149
Simple regression with baseline eGFR$_{\text{ckdepi crea}}$ values	0.670	-	99.63(ml/min/1.73m^2)^2	-
eGFR$_{\text{ckdepi cys}}$				
Multiple regression with 660 NMR bins	0.585	0.190 ± 0.008	153.66(ml/min/1.73m^2)^2 / 0.174	103
Multiple regression with 641 NMR bins	0.522	0.205 ± 0.009	176.86(ml/min/1.73m^2)^2 / 0.194	260
Simple regression with baseline eGFR$_{\text{ckdepi cys}}$ values	0.699	-	110.4(ml/min/1.73m^2)^2	-
eGFR$_{\text{ckdepi crea cys}}$				
Multiple regression with 660 NMR bins	0.633	0.164 ± 0.007	114.06(ml/min/1.73m^2)^2 / 0.140	109
Multiple regression with 641 NMR bins	0.532	0.195 ± 0.008	146.42(ml/min/1.73m^2)^2 / 0.163	184
Simple regression with baseline eGFR$_{\text{ckdepi crea cys}}$ values	0.711	-	89.80(ml/min/1.73m^2)^2	-

Table 5.7: **Results of regression analyses for prediction of FU2 SCr, SCysC, and eGFR values.** LASSO regression with both 660, as well as 641 individual baseline NMR bucket intensities after the exclusion of all creatinine signals, and simple linear regression with respective baseline clinical parameters were conducted. More details are given in the text. Abbreviations: cv, cross-validated; eGFR, estimated glomerular filtration rate; eGFR$_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; eGFR$_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; eGFR$_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; eGFR$_{\text{mdrd4}}$, eGFR based on MDRD4 formula; FU2, second follow-up; *mse*, mean-squared error; SCr, serum creatinine; SCysC, serum cystatin C.

It is apparent that simple linear regression models based on the respective baseline clinical

parameters outperformed the respective LASSO models based on baseline NMR bucket intensities for the prediction of all FU2 clinical parameters in terms of both higher R^2 and lower *mse* values on the test set.
The superior prediction performance of all simple linear regression models based on baseline clinical parameters is not surprising since the majority of the included patients in this cohort did not experience large changes in SCr, SCysC or eGFR values over the investigated time period of two years, as illustrated in histograms displayed in Appendix III section 7.3.3 Fig. 7.11. This observation is actually reflected by the fact that the prediction of SCr, SCysC, or eGFR differences between baseline and FU2 time-point based on LASSO models derived from baseline EDTA plasma 1D ^{1}H CPMG spectra did not yield any satisfactory results (data not shown).

5.2.4 Discussion

The GCKD study comprises currently the world-wide largest cohort of patients suffering from chronic kidney disease [Titze et al. 2015] and gives way to a new dimensionality in NMR based metabolomic analyses. The measurement and statistical analysis of baseline EDTA plasma fingerprints derived from 1D ^{1}H NMR measurements of this cohort posed several challenges on project management as well as NMR measurements. However, at the same time, it offers sufficient sample sizes to derive meaningful statistical inferences about the total population, e.g. by comparison of metabolite concentrations in this CKD cohort or entities thereof against reference concentrations in healthy subjects as provided, for example, by the HMDB [Wishart et al. 2007].
First, an effective way to suppress broad signals in 1D ^{1}H NMR spectra arising from protons in proteins of plasma specimens had to be chosen. Common protein removal methods like filtering of specimens with 10kDa filters, as applied for the acute kidney injury study described in section 5.1, or methanol aided protein precipitation are not feasible for large sample cohorts due to cost intensive and time consuming procedures. Therefore, we decided to measure 1D ^{1}H NMR spectra of the complete baseline EDTA plasma specimen cohort of the GCKD study without a protein removal step by employing the CPMG pulse sequence. This enabled the acquisition of well resolved 1D ^{1}H spectra without broad signals arising from protons in proteins in sufficiently low measurement times for high-throughput studies. However, broad, unspecific signals arising from protons in lipids, which would have been removed by sample filtering [Klein 2011], are still present in the CPMG spectra.
Second, significant line broadening and diminishing of the TSP reference signal, which crucially depended on the macromolecule content of the respective specimen, was observed. Consequently, the alternative reference substance formic acid [Beckonert et al. 2007] was added to the remaining, not yet measured NMR spectra, which were subsequently scaled to the respective signal intensity to reduce variations in spectrometer performance. Therefore, all statistical analyses reported in this Ph.D. thesis were only conducted with 1D ^{1}H CPMG spectra acquired after the addition of formic acid. The remaining EDTA plasma specimens are currently measured with formic acid as internal standard and will be evaluated in the future. Moreover, data cleaning procedures for correct matching between NMR sample IDs and corresponding sample

IDs in the clinical data files are currently under way at the University Clinic of Erlangen. It is therefore likely that NMR spectra, which had to be excluded from statistical data analyses due to sample ID mismatch will be correctly matched in the future and consequently included in statistical data analyses. Nevertheless, the two GCKD study specimen cohorts comprising 3164 and 2697 individual patients, respectively, showed baseline patient characteristics (Appendix III section 7.3.1) comparable to overall GCKD study characteristics (section 5.2.1), except for diabetes incidence. Moreover, statistical power calculations for the comparisons of different leading renal diseases by means of *t*-tests proved that the reported analyses were conducted with sufficient numbers of specimens. Thus, overall significant differences between the results of the statistical analyses conducted for this Ph.D. thesis and respective analyses with the complete specimen cohort of 4920 specimens are not to be expected.
Third, overall spectral shifts larger than 0.01 ppm were observed for 86 1D ^{1}H CPMG spectra, which would not be efficiently compensated by a bucketing procedure with a bucket width of 0.01 ppm. Here, I present a fast and sufficient method to align the complete cohort of 3206 1D ^{1}H CPMG spectra. A manual alignment of the 86 shifted spectra was omitted, since it is prone to human bias. Moreover, this general spectral alignment tool could also be applied to other NMR-based data sets.

Specific metabolic fingerprints of six different leading renal diseases were determined via individual comparisons by means of *t*-tests. Major significant differences, characterized by a sufficient number of NMR bins with significant B/H-adjusted *p*-values and correspondingly sufficient statistical power, are reported for, in total, 11 different comparisons. A significant up-regulation of D-glucose concentrations in EDTA plasma specimens of patients suffering from diabetic nephropathy can be deduced from the presented analyses. These observations are not surprising since the definition of diabetic nephropathy itself implements that the respective patient significantly suffers from diabetes mellitus, which itself is characterized by chronic hyperglycemia [Arasteh et al. 2009]. Patients in the GCKD study were diagnosed as suffering from diabetes mellitus if either the percentaged fraction of the glycated hemoglobin HbA_{1c} in comparison to the total amount of hemoglobin was equal to or larger than 6.5%, or if anti-diabetic medication was prescribed. HbA_{1c} is a sufficient measure for the average blood glucose concentration during the last 8 to 10 weeks, and accounts for about 4 - 6% of total hemoglobin in people without diabetes mellitus [Arasteh et al. 2009]. In fact, GCKD patients suffering from diabetic nephropathy (mean baseline HbA_{1c} percentaged fraction 7.5%) exhibited significantly higher (p-value $< 2.2e^{-16}$) percentaged fractions of HbA_{1c} in comparison to GCKD patients suffering from glomerulonephritis (mean baseline HbA_{1c} percentaged fraction 6.0%), hereditary diseases (mean baseline HbA_{1c} percentaged fraction 5.9%), interstitial nephropathy (mean baseline HbA_{1c} percentaged fraction 6.1%), systemic diseases (mean baseline HbA_{1c} percentaged fraction 5.8%), and vascular nephropathy (mean baseline HbA_{1c} percentaged fraction 6.2%), respectively, despite frequent administration of anti-diabetic medication in the diabetic nephropathy group (88.5% of patients received anti-diabetic medication) in comparison to other groups (between 4 and 18%). This supports the significant up-regulation of D-glucose in EDTA plasma specimens collected at the baseline time-point of patients suffering from diabetic nephropathy in comparison to patients suffering from other leading renal

diseases as revealed by NMR spectroscopy.
Moreover, significant higher lipid signals in EDTA plasma specimens of patients suffering from glomerulonephritis in comparison to patients suffering from other leading renal diseases were detected. Hyperlipidemia is a common characteristic of the so-called nephrotic syndrome [Arasteh et al. 2009], which itself is one of the major pathophysiological conditions of glomerulonephritis [Arasteh et al. 2009]. Therefore, the presented results are in good concordance with standard clinical definitions of diabetic nephropathy and glomerulonephritis.
Although these various group comparisons revealed interesting NMR spectral differences between various leading renal diseases, one has to keep in mind that specific metabolic fingerprints of these renal diseases could only be determined by comparing the respective 1D ^{1}H NMR spectra of one individual leading renal disease to 1D ^{1}H NMR spectra of healthy individuals. Here, one has to consider that the sample preparation procedures for the respective EDTA plasma specimens of the healthy control group have to be exactly the same as reported for the EDTA plasma GCKD specimens. This implies the acquisition of 1D ^{1}H CPMG spectra without prior ultrafiltration of the specimens. Unfortunately, appropriate NMR spectra of a suitable healthy control group were not available for this Ph.D. thesis. A comparison between single entities suffering from specific leading chronic renal diseases and a suitable healthy control group by means of classification, as, e.g. realized by [Gronwald et al. 2011] for urine specimens collected from patients suffering from ADPKD, would be able to estimate the diagnostic value of novel biomarkers.

A comparison between the four different GFR estimation equations by means of Pearson's correlation coefficients and Bland-Altman plots revealed almost perfect consensus between different eGFR formulas based on SCr, and larger deviations between eGFR formulas either based on SCr or SCysC. However, I was not able to make any deductions about over- or underestimation of GFR by the employed eGFR values, since measured GFR values were not available for the GCKD study.
The LASSO regression analyses for the prediction of baseline Synlab clinical chemistry parameters employing the complete set of 660 baseline NMR bucket intensities showed very good performance for SCr values and slightly minor performance for SCysC values on the independent test set. The same observations hold true for eGFR values either based on SCr and/or SCysC. The MDRD4, CKD-EPI crea, and CKD-EPI cys formulas are all based on sex, age, ethnicity, and either SCr or SCysC [Levey et al. 1999,Inker et al. 2012]. The CKD-EPI crea cys formula is based on sex, age, ethnicity, and both SCr as well as SCysC [Inker et al. 2012]. NMR buckets corresponding to plasma creatinine had always been included in the derived LASSO models with high absolute β coefficients (data not shown). However, LASSO models derived after the exclusions of all spectral regions corresponding to creatinine mostly showed a significant performance drop in comparison to LASSO models including creatinine buckets especially for baseline response variables based on SCr. Consequently, the inclusion of creatinine buckets for multiple regression analyses seems to be very important to obtain good predictive results with respect to baseline clinical parameters.
The reliable prediction of future kidney performance in the context of chronic kidney disease is of crucial importance for sufficient patient treatment. Therefore, I evaluated the predictive

performance of different regression analyses with respect to FU2 clinical parameters. Apparently, simple linear regression models based on the respective baseline parameters outperformed respective LASSO models based on baseline NMR bucket intensities for the prediction of FU2 Synlab clinical chemistry and all eGFR values in terms of both higher R^2 and lower mse values on the test set. These results seem to reflect the fact that the majority of the included patients in this cohort did not experience large changes in SCr, SCysC or eGFR values over the investigated time period of two years. Multiple regression analyses based on baseline NMR data with respect to future follow-up response variables might reveal different results. However, clinical parameters of, e.g., the third follow-up time-point had not been available for this Ph.D. thesis. Provided that sufficient numbers of GCKD patients experienced major changes in renal function, the described regression analyses might also be performed separately within specific CKD entities, e.g. only including patients suffering from diabetic nephropathy. Thereby, possibly different CKD progression rates in individual CKD entities might be detected and more homogeneous metabolic profiles within one entity might improve multiple regression analyses results based on NMR data.

The absolute quantification of metabolites in unfiltered EDTA plasma NMR spectra acquired by utilizing the CPMG pulse sequence is going to be evaluated by Jens Wallmeier in the context of his Ph.D. thesis. Here, one has to consider NMR signal intensity losses due to overall NMR signal decay during the filtering period of the CPMG sequence. These signal intensity losses result from T_2 relaxation, which arises, as described in section 4.2.1, from intramolecular spin-spin interactions. T_2 relaxation consequently depends on the molecule's size, the density of the interacting nuclei, the viscosity of the solvent, and the temperature, as T_2 relaxation is mainly mediated by dipolar interactions which critically depend on molecular motions. Furthermore, NMR signal intensity losses are also enhanced by increasing salt content of the solution. The impact of these various influences on recorded NMR signal intensities with respect to metabolite quantification are currently investigated by Jens Wallmeier. Analyses so far showed that NMR signal intensities recorded with a CPMG pulse sequence seem to be considerably influenced by the specific matrix composition of the investigated solution. One might be able to account for these matrix effects in NMR based absolute quantification by determining suitable calibration factors. This implies the experimental determination of calibration factors with the CPMG pulse sequence in pooled unfiltered EDTA plasma specimens from the GCKD study cohort, as currently performed by Jens Wallmeier. Provided that the matrix composition of unfiltered EDTA plasma specimens is fairly similar within the GCKD study cohort, these newly determined calibration factors will probably facilitate appropriate metabolite quantification in unfiltered EDTA plasma specimens. Absolute concentrations of plasma metabolites in this CKD cohort or entities thereof could then be compared against reference concentrations in healthy subjects as provided, for example, by the HMDB [Wishart et al. 2007]. Moreover, simple as well as multiple linear regression analyses based on absolute metabolite concentrations determined by NMR spectroscopy at the baseline or at future follow-up time-points with respect to baseline or future clinical parameters could be performed. Besides NMR analyses of EDTA plasma specimens, the Institute of Functional Genomics will furthermore investigate the composition of complementary urine specimens of the GCKD study collected at the baseline as well as future follow-up time-points. Absolute quantification of metabolites simultaneously

in EDTA plasma and matching urine specimens would facilitate the determination of their respective fractional excretions to give further insights into renal clearance of these compounds in this CKD cohort or entities thereof. Please note that the described statistical analyses in this Ph.D. thesis were all performed with NMR bucket intensities relative to the reference signal intensity of formic acid. It is obvious that the NMR signal intensity decay due to T_2 relaxation during the filtering period of the CPMG sequence also impacts bucket intensities. Since, however, statistical analyses based on Student's t-test[15] or the performance of multiple regression models employing the LASSO method[16] are invariant under shifts of $\log_2$ transformed bucket intensities, one does not have to account for these previously discussed NMR signal decays in these analyses, provided that unfiltered EDTA plasma specimens of the GCKD study yield fairly similar matrix compositions.

In addition to the already discussed research opportunities within the GCKD cohort, this study offers numerous other possibilities to derive meaningful statistical inferences about the total population. Survival analyses with respect to overall patient survival as well as time elapsed until onset of RRT by means of Cox regression [Cox 1972] with baseline or future follow-up 1D ^{1}H NMR EDTA plasma and/or urine spectra might reveal interesting metabolites associated with poor survival and/or rapid need for RRT in this CKD cohort or entities thereof. Corresponding analyses will be performed in the future at the Institute of Functional Genomics. Moreover, correlation calculations, regression, as well as classification analyses based on metabolic profiles derived from EDTA plasma or urine specimens collected at various time-points with respect to numerous clinical and quality of life parameters, sociodemographic factors, comorbidities, etc. might establish novel relationships between distinct metabolic profiles and these parameters. The GCKD study comprises over 5000 patients suffering from a huge spectrum of various renal diseases, including ADPKD. NMR fingerprinting of urine specimens, for example, already proofed to reliably differentiate ADPKD patients from those suffering from CKD for other reasons than ADPKD as well as from healthy individuals [Gronwald et al. 2011]. The urinary metabolic profiles of GCKD patients suffering from ADPKD might be compared to both these already investigated study cohorts as well as other GCKD entities suffering from other renal diseases. Thereby, the ADPKD biomarkers investigated by [Gronwald et al. 2011] might be validated in an independent study cohort. Moreover, novel EDTA plasma metabolic profiles of patients suffering from ADPKD might be established in comparison to other GCKD entities, which might complement the already investigated urinary profiles provided by [Gronwald et al. 2011].

Besides, this huge NMR spectral data set offers excellent opportunities to evaluate possible im-

[15] Consider a general shift for bucket intensity $X_\mathrm{b} \rightarrow \widetilde{X_\mathrm{b}} = X_\mathrm{b} \cdot a$. For $\log_2$ transformed bucket intensities, Student's t statistic is now calculated as $\widetilde{T_b} = \frac{\overline{\log_2(\widetilde{X}_{b1})} - \overline{\log_2(\widetilde{X}_{b2})}}{\widetilde{s_b}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{\frac{\sum_{i=1}^{n_1} \log_2(X_{b1i}\cdot a)}{n_1} - \frac{\sum_{j=1}^{n_2} \log_2(X_{b2j}\cdot a)}{n_2}}{\widetilde{s_b}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{\log_2 a + \frac{\sum_{i=1}^{n_1} \log_2 X_{b1i}}{n_1} - (\log_2 a + \frac{\sum_{j=1}^{n_2} \log_2 X_{b2j}}{n_2})}{\widetilde{s_b}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{\overline{\log_2(X_{b1})} - \overline{\log_2(X_{b2})}}{s_b\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = T_b$, as $\widetilde{s_b} = s_b$.

[16] Consider a general shift for bucket intensity $x_\mathrm{ij} \rightarrow \widetilde{x_\mathrm{ij}} = x_\mathrm{ij} \cdot a_j$. For $\log_2$ transformed bucket intensities, the linear model is now calculated as $\widetilde{y_i} = \beta_0 + \sum_{j=1}^{p} \beta_j \log_2 \widetilde{x_{ij}} + \epsilon = \beta_0 + \sum_{j=1}^{p} \beta_j \log_2(x_{ij} \cdot a_j) + \epsilon = \beta_0 + \sum_{j=1}^{p} \beta_j \log_2 a_j + \sum_{j=1}^{p} \beta_j \log_2 x_{ij} + \epsilon = \widetilde{\beta_0} + \sum_{j=1}^{p} \beta_j \log_2 x_{ij} + \epsilon$. As β_0 is not included in the penalizing term of Eq. (4.16), such a general shift does not affect regression model performance.

provements for statistical data analysis in NMR based metabolomics. B.Sc. Sebastian Mehrl is currently investigating the predictive performance of a newly developed regression method, i.e. zero-sum regression [Altenbuchinger et al. 2016], employing the GCKD EDTA plasma 1D ^{1}H NMR spectra under the guidance of M.Sc. Helena U. Zacharias and Dr. Michael Altenbuchinger in the context of a master thesis. Zero-sum regression is, by construction, reference point insensitive. It might therefore yield improved predictive performance in comparison to standard multiple regression methods if the response variable cannot be uniquely associated with a reference point, e.g. in the case of survival or diagnostic data.
In summary, this NMR analysis of GCKD EDTA plasma specimens collected at the baseline time-point with respect to both baseline as well as FU2 clinical parameters can be regarded as a pilot study for future investigations of large-scale clinical trials by means of NMR spectroscopy. The presented methods to compensate drawbacks from the traditional NMR reference substance TSP in unfiltered plasma will be easily implemented for NMR data acquisition of unfiltered EDTA plasma specimens collected at future follow-up time-points. The realization of all discussed perspectives for statistical analyses of the GCKD study would have been beyond the scope of this thesis. They will be performed in the future at the Institute of Functional Genomics.

5.3 Trial to Reduce Cardiovascular Events with Aranesp® Therapy study

5.3.1 Introduction

Besides the investigation of baseline EDTA plasma samples from the German Chronic Kidney Disease study cohort, I furthermore focused my research interests on a second large-scale clinical study in the context of CKD and anemia, compare to section 4.1.4.
The Trial to Reduce Cardiovascular Events with Aranesp® Therapy (TREAT) study (`ClinicalTrials.gov` Identifier: NCT00093015) was designed as a randomized multicenter, double-blind, placebo-controlled, clinical trial and sponsored by Amgen® [Pfeffer et al. 2009a)]. It should test the hypothesis that in patients with diabetes, CKD not requiring dialysis, and concomitant anemia, increasing Hb levels by using darbepoetin alfa would reduce the rates of death or cardiovascular morbidity and RRT [Pfeffer et al. 2009a)]. In total, 4044 patients with type-2 diabetes mellitus, CKD defined as an eGFR of 20-60 ml/min per $1.73m^2$ calculated with the use of the MDRD equation, anemia (Hb level $\leq$11.0 g/dl), and a transferrin saturation $\geq$15% have been enrolled from August 2004 until December 2007 at 623 sites in 24 countries [Pfeffer et al. 2009a)]. Patient baseline characteristics, study design procedures, and outcomes are detailed in [Pfeffer et al. 2009a), Pfeffer et al. 2009b), Lewis et al. 2011, McMurray et al. 2011, Mix et al. 2005, Rao and Pereira 2003, Skali et al. 2011, Skali et al. 2013, Solomon et al. 2010]. In brief, this study could not detect a reduced risk of either primary cardiovascular or primary renal events for patients treated with darbepoetin alfa in comparison to patients treated with a placebo compound [Pfeffer et al. 2009b)]. However, the patient group treated with darbepoetin alfa seemed to suffer from an increased risk of stroke in comparison to placebo-treated study participants [Pfeffer et al. 2009b), Skali et al. 2011]. As a consequence, clinical guidelines for the treatment of anemia in CKD patients employing ESAs were adapted [KDIGO workgroup 2013, Singh 2010], as explicitly discussed in section 4.1.4.

A subset of 1167 urine samples were measured with NMR spectroscopy and statistically analyzed in order to gain new insights into CKD progression and corresponding Hb responsiveness. Therefore, the following three hypotheses were statistically evaluated: (1) no difference in metabolic profiles exists between patients dying from any cause (code V), and patients not dying (code U), under the restriction that all patients within both subcohorts do not progress to ESRD (hypothesis 1a), (2) no difference in metabolic profiles exists between patients progressing (code P) and not progressing to ESRD (code O) under the restriction that all patients within both subcohorts do not die (hypothesis 1b) and (3) no difference in metabolic profiles exists between patients with various stages of Hb responsiveness, respectively, whereas four different subcohorts treated with darbepoetin alfa with various stages of Hb responsiveness and one subcohort treated with a placebo compound are investigated (codes A - E, hypothesis 2). This subset had been provided by Amgen® to our collaboration partners at the University Clinic of Erlangen-Nuremberg, who allocated corresponding specimen aliquots to us for metabolic investigations after prior consultation with Amgen®.

5.3.2 Materials and Methods

Table 5.8 lists the number of investigated urine specimens according to hypothesis and group membership. Note that a number of specimens belonged to two different groups and were equally investigated for both eligible hypotheses. For hypotheses 1a and 1b, urine samples were collected in the last week directly before treatment randomization (W1), i.e. neither darbepoetin alfa nor the placebo compound had been administered to the patients yet, for hypothesis 2, two urine samples were collected from each patient at two different time-points, i.e. one in W1 and one in the 49th week (W49) after treatment randomization. Code identities of patient outcome for hypotheses 1a and 1b were available, whereas code identities for hypothesis 2 are still blinded.

Hypothesis	**Hypothesis 1a**		**Hypothesis 1b**		**Hypothesis 2**									
Code	**U**	**V**	**P**	**O**	**A**		**B**		**C**		**D**		**E**	
Collection time-point	**W1**	**W1**	**W1**	**W1**	**W1**	**W49**	**W1**	**W49**	**W1**	**W49**	**W1**	**W49**	**W1**	**W49**
Number of samples	129	127	113	92	90	62	87	72	87	70	84	69	83	71

Table 5.8: **Number of investigated urine samples subdivided by corresponding hypothesis, group membership, and collection time-point.** More details are given in the text.

Urine specimens were prepared in collaboration with Claudia Samol and 1D ^{1}H NOESY as well as 2D ^{1}H-^{13}C HSQC spectra were measured according to the standard protocols described in section 4.2.2 employing the parameters given in section 5.1.2.2. One representative high-resolution 2D ^{1}H-^{13}C HSQC spectrum from a patient not progressing to ESRD (code O, collected at W1) was acquired with 2048 $\times$ 512 data points using 44 scans per increment. For the same urine specimen, one 2D ^{1}H ^{1}H TOCSY spectrum was acquired using the pulse program *mlevgpphw5* (BrukerBioSpin GmbH, Rheinstetten, Germany) with 2048 $\times$ 256 data points, 56 scans per increment, and a mixing time of 50 ms. The total acquisition times of the high-resolution 2D ^{1}H-^{13}C HSQC and 2D ^{1}H ^{1}H TOCSY spectrum amounted to approximately 19.6 h and 13 h, respectively.
1D ^{1}H NOESY and standard 2D ^{1}H-^{13}C HSQC spectra were preprocessed by Claudia Samol following the protocols described in section 4.2.2.3 and the TSP reference signal in the standard 2D ^{1}H-^{13}C HSQC spectra was manually set to zero.
I evenly split the region from 9.5 - 0.5 ppm of the 1D ^{1}H NOESY spectra, excluding the solvent area and the area containing the broad urea peak from 6.5 - 4.5 ppm, into 701 buckets with a bucket width of 0.01 ppm employing AMIX-Viewer 3.9.13 (Bruker BioSpin GmbH, Rheinstetten, Germany). The signal intensities of each bin were summed and additionally scaled to the reference region of creatinine from 3.06 - 3.028 ppm, as outlined in section 4.3.1.1. For further analysis, data were imported into *R* (Development Core Team 2009).
In order to guide decision making for the appropriate normalization technique, the for the

AKI project developed strategy was applied, compare to section 5.1.3.1. The respective data for hypothesis 1a, 1b, and 2 were not normally distributed (Shapiro-Wilk normality test p-values $<2.2e^{-16}$). Consequently, only a $\log_2$ transformation was applied for minimization of heteroscedasticity, compare to section 5.1.3.1.
Identification of discriminant compounds was performed by Welch t-tests and/or one-way ANOVA as described in section 4.3.1.3, and classification was performed employing an RF classifier in a leave-five-out cross-validation with feature selection based on Welch t-test and parameter optimization, compare to section 4.3.1.3. The number of selected features was increased in steps of one starting with four features and the maximum number was set to 50. The internal parameters of the RF classifier were optimized employing a grid search procedure, as explicitly described in section 5.1.2.6. Metabolite identification was performed as described in section 4.3.2. Hierarchical clustering employing Euclidean distances and the average linkage method was performed as detailed in section 4.3.1.2.

5.3.3 Results

Welch t-tests were employed for a statistical evaluation of differences between the respective groups of the corresponding hypotheses. 283 and 247 significantly different NMR features were identified by comparing groups U and V of hypothesis 1a and comparing groups P and O of hypothesis 1b, respectively. These NMR features were identified and are listed in Appendix IV Tables 7.22 and 7.23, respectively. For both t-tests, most of the significantly different buckets comprised broad protein and macromolecule signals. For hypothesis 1a, some significant buckets further comprised sugars like D-glucose, sucrose, etc.. The corresponding significant p-values B/H-adjusted are in the range between 0.01 and 0.05, indicating only slight metabolic differences between the investigated groups.
In total, 45 individual compounds have been identified in a representative high-resolution 2D ^{1}H-^{13}C HSQC spectrum of a urine sample from group O collected at W1 supported by a corresponding 2D ^{1}H ^{1}H TOCSY spectrum. This annotated 2D ^{1}H-^{13}C HSQC spectrum is displayed in Appendix III, Figure 7.12. Metabolite identification in 1D ^{1}H spectra was rather cumbersome due to massive signal overlap with broad NMR signals arising from proteins and macromolecules.
The discriminating features indicated by a t-test comparing groups U versus V and groups O versus P, respectively, are displayed in heat-map representations in Figure 5.8. Columns (i.e. NMR spectra) and rows (i.e. features) of the heat-maps are ordered according to the results of hierarchical cluster analyses performed with the respective spectra and significant buckets, respectively, and the corresponding dendrograms are also shown in Figure 5.8. The NMR features were split into five clusters for hypothesis 1a, and into two clusters for hypothesis 1b according to the hierarchy of the generated dendrograms. These NMR feature clusters are driven by the respective bucket intensities. For hypothesis 1a, the clusters of NMR features are driven by two main compounds, i.e. either broad protein signals or various sugars such as D-glucose and sucrose. The two clusters of NMR features for hypothesis 1b comprise NMR signals arising from proteins/macromolecules and three NMR features probably corresponding to various metabolites in combination with proteins and/or macromolecules, respectively.

For both hypotheses, NMR spectra were split into four distinct clusters according to the hierarchy of the generated dendrograms. Table 5.9 lists the numbers and proportions with respect to the total number of patients belonging to the individual group for each cluster. It is obvious that the cluster formation for both hypotheses is mainly driven by the different protein content of the individual urine samples. As can be seen in the heat-map representation for hypothesis 1a, Figure 5.8A), urine spectra grouped in clusters 1 and 4 have a high protein and sugar content. Urine spectra arranged in cluster 2 have the lowest protein and sugar levels in comparison to all other spectra and cluster 3 consists of urine spectra with intermediate protein and sugar content. Likewise for hypothesis 1b, Figure 5.8B), cluster 5 has the highest protein/macromolecule content, whereas cluster 6 has the lowest protein/macromolecule content. The two intermediate clusters 7 and 8 can also be distinguished by their different protein content with cluster 8 showing a higher protein content than cluster 7. Data from Table 5.9 indicates that clusters characterized by an up-regulated protein content of the urine samples mainly comprise V and P patients. To be more precise, patients dying from any cause and patients progressing to ESRD seem to have a higher urinary protein content in comparison to patients not dying and patients not progressing to ESRD, respectively.
Note that this cluster analysis, in general described as an unsupervised statistical analysis method (compare to section 4.3.1.2), was only performed with the statistically significant NMR bins, which had been revealed by a Welch t-test representing a supervised method (compare to section 4.3.1.3). Therefore, cluster analyses were repeated for hypotheses 1a and 1b, respectively, now including all 701 NMR buckets (data not shown). Again, NMR spectra for hypothesis 1a could be divided into four distinct clusters. NMR spectra for hypothesis 1b could be divided into three clusters, whereas, in comparison to Figure 5.8, only one intermediate cluster appears. Clusters characterized by up-regulated feature intensities, dominated by broad protein signals, again mainly consisted of patients dying from any cause and patients progressing to ESRD.
The predictive power of the discriminating features for both hypotheses has also been assessed by RF classification. For each hypothesis one RF run was performed in a nested leave-five-out cross-validation procedure with parameter optimization. For hypothesis 1a, an averaged total accuracy of 49.6% and an area under the ROC curve of 0.52 were obtained. On average, nine features were selected by the RF classifier, mostly comprising proteins. The median of the selected m_{try} parameter was two and the median of the selected number of trees n_{tree} was 300. For hypothesis 1b, an averaged total accuracy of 54.2% and an area under the ROC curve of 0.56 were obtained. This classification rested, on average, on 37 features, which were mainly identified with proteins and macromolecules. The median of the selected m_{try} parameter amounted to four and the median of the selected number of trees n_{tree} was 300.
This classification data indicates a rather poor predictive power of the discriminating NMR features for both hypotheses. However, this result does not come as a surprise, since the discriminating power of the NMR features employed by the RF classifier as indicated by their respective p-values according to the performed Welch t-tests, compare to Appendix III Tables 7.22 and 7.23, had not been very prominent.
As already outlined, proteins and macromolecules, which give rise to very broad and unspecific signals in the 1D ^{1}H NMR spectra [Klein 2011], comprise most of the significantly differing

NMR buckets for both hypothesis 1a and 1b. These broad signals might obscure the discriminating power of distinct, underlying NMR features arising from metabolites. Therefore, the statistical analysis described above for hypothesis 1b was repeated with a subcohort of, in total, 67 urine spectra with minor proteinuria (36 O patients and 31 P patients, respectively). This subcohort had been selected by eye with regard to the total amount of protein signals visible in the 1D NMR spectra. The corresponding t-test did not reveal any significant buckets (smallest p-value B/H-adjusted = 0.34). The RF classification run obtained an averaged total accuracy of 50.0% and an area under the ROC curve of 0.53. On average, 15.5 features were selected, still mostly comprising proteins and macromolecules. The median of the selected m_{try} parameter was 3.25 and the median of the selected number of trees n_{tree} was 325.

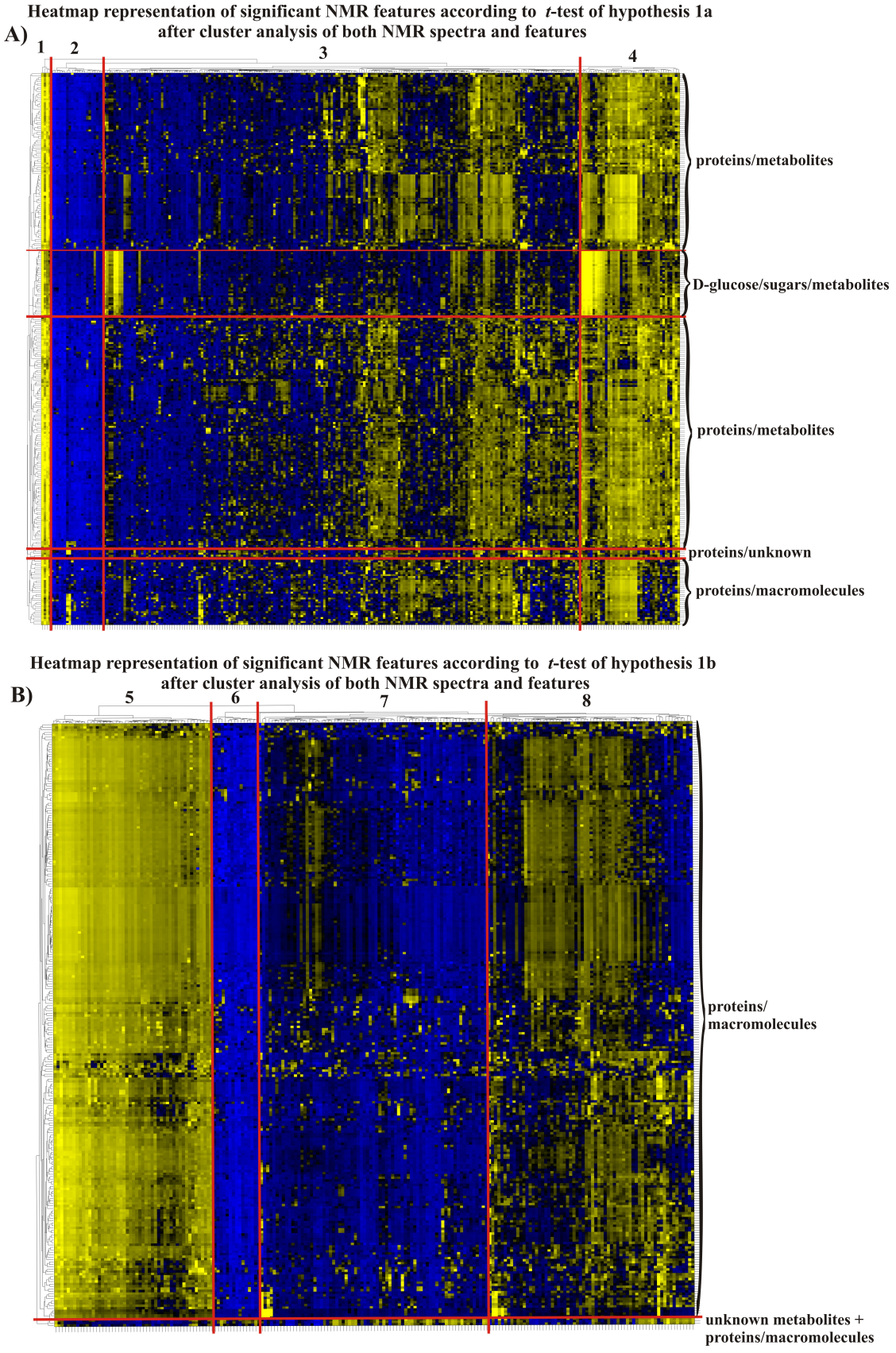


Figure 5.8: **Heat-map representation of significant features according to t-test results for hypothesis 1a (A) and 1b (B), respectively, after cluster analysis for both NMR spectra and features.** Each row corresponds to one NMR feature, each column corresponds to one urine spectrum, respectively to one patient. NMR spectra and features are ordered according to cluster analysis results. Clusters are separated from each other by red vertical/horizontal lines and were cut according to the hierarchy of the respective dendrograms, shown on the top and left side of the heat-maps for NMR spectra and features, respectively. Numbers and percentaged fractions of patients belonging to individual groups within each cluster are given in Table 5.9. Blue denotes a down-regulated feature, whereas yellow indicates an up-regulated one.

Hypothesis 1a				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Number (percentaged fraction) of patients belonging to hypothesis 1a	4 (1.6%)	21 (8.2%)	191 (74.6%)	40 (15.6%)
Number (percentaged fraction) of U patients	0 (0%/0%)	13 (10.1%/61.9%)	103 (79.8%/53.9%)	13 (10.1%/32.5%)
Number (percentaged fraction) of V patients	4 (3.2%/100%)	8 (6.3%/38.1%)	88 (69.3%/46.1%)	27 (21.3%/67.5%)
Hypothesis 1b				
	Cluster 5	**Cluster 6**	**Cluster 7**	**Cluster 8**
Number (percentaged fraction) of patients belonging to hypothesis 1b	51 (24.9%)	15 (7.3%)	73 (35.6%)	66 (32.2%)
Number (percentaged fraction) of O patients	15 (16.3%/29.4%)	11 (12%/73.3%)	34 (37%/46.6%)	32 (34.8%/48.5%)
Number (percentaged fraction) of P patients	36 (31.9%/70.6%)	4 (3.5%/26.7%)	39 (34.5%/53.4%)	34 (30.1%/51.5%)

Table 5.9: **Summary of cluster analyses results.** The total number of patients within each cluster is given as well as the percentaged fraction with respect to the total number of patients belonging to one hypothesis. The numbers as well as percentaged fractions of patients in each cluster in the heat-map representations of Figure 5.8 are given in brackets (first percentage donates percentage fraction with respect to total number of patients belonging to one group in this cluster, second percentage donates fraction of patients belonging to one group representing this cluster; for illustration purposes regard number of V patients in cluster 1: four out of, in total, 127 V patients were clustered in cluster 1 corresponding to a percentaged fraction of 3.2% with respect to the total number of V patients, four out of four patients forming this cluster are V patients, i.e. 100% of the patients forming this cluster are V patients). Hypothesis 1a: no difference exists between patients dying from any cause without progression to ESRD (code V) and patients not dying without progression to ESRD (code U). Hypothesis 1b: no difference exists between patients progressing to ESRD without dying (code P) and patients not progressing to ESRD without dying (code O). All urine samples were collected in the last week directly before treatment randomization (W1).

With regard to hypothesis 2, an ANOVA was performed including all groups and sample collection time-points belonging to this hypothesis. The corresponding F-test did not obtain any significant buckets (smallest B/H-adjusted p-value = 0.92). This result indicates a rather small probability for the detection of a significant group difference between patients with various Hb responsiveness at W1 or at W49.

5.3.4 Discussion

Results from the statistical data analyses of NMR spectra corresponding to hypothesis 1a and 1b, respectively, clearly reveal a statistically significant up-regulation of proteins and macromolecules in urine specimens collected in the week directly prior to treatment randomization

from patients who are going to die or develop ESRD in the future, respectively. As already outlined in sections 4.1.2 and 4.1.4, presence of proteinuria, i.e. mainly albuminuria, for more than three months, is one of the main criteria for the diagnosis of CKD [KDIGO workgroup 2013]. Moreover, proteinuria has proven to be a significant and independent predictor of ESRD in mass screening settings as well as in patients with type-2 diabetes mellitus and nephropathy [Iseki et al. 2003, Keane et al. 2003]. Therefore, my results are in good concordance with previous studies about ESRD.
However, the power of NMR features corresponding to broad protein signals for prediction of future ESRD was rather negligible, as indicated by both the results of the RF classification and the rather moderate B/H-adjusted p-values above 0.01.
The statistical analysis of hypothesis 1a revealed that patients dying from any cause without progressing to ESRD also exhibit a higher degree of urinary protein content in comparison to patients not dying. In general, proteinuria is also a prominent risk factor for all-cause as well as cardiovascular mortality in patients with CKD [KDIGO workgroup 2013]. Differentiating investigations for individual causes of death might reveal further interesting results, however, no information about the individual cause of death for each patient enrolled in hypothesis 1a was available.
The TREAT study had been originally designed in order to test whether the administration of darbepoetin alfa would reduce the rates of death, cardiovascular events, and ESRD in patients with type-2 diabetes mellitus, CKD not requiring dialysis, and anemia [Pfeffer et al. 2009a), Pfeffer et al. 2009b)]. No reduced risk of either primary cardiovascular or primary renal events for patients treated with darbepoetin alfa in comparison to patients treated with a placebo compound were reported [Pfeffer et al. 2009b)]. However, an increased risk of stroke in the patient group treated with darbepoetin alfa has been observed [Pfeffer et al. 2009b), Skali et al. 2011]. As an additional baseline predictor for stroke, Skali et al. identified an increased urinary protein-to-creatinine ratio, which, however, did not seem to be interconnected with the administration of darbepoetin alfa [Skali et al. 2011]. Solomon et al. further identified a poor initial hematopoietic response to the treatment with darbepoetin alfa as a risk factor for death or cardiovascular events [Solomon et al. 2010]. A higher urinary protein-to-creatinine ratio has also been identified as a significant risk factor for both the cardiovascular composite outcome, i.e. death or nonfatal cardiovascular event, as well as development of ESRD in the TREAT study itself [Pfeffer et al. 2009b), McMurray et al. 2011]. Investigations about the impact of darbepoetin alfa treatment on risk of death or progression of ESRD could not be made in my study, since for hypotheses 1a and 1b, only one urine specimen from each patient had been collected in the last week directly before treatment randomization, i.e. patients had not yet received darbepoetin alfa or the placebo compound. Furthermore, no information about the subsequent treatment assignment for these four patient groups was available. For hypothesis 2, which included four different patient groups treated with varying doses of darbepoetin alfa and one group treated with a placebo compound, two urine specimens from each patient had been collected in the last week directly before treatment randomization and in the 49^{th} week after randomization. However, no information about the group identities was available for my investigations. Moreover, the statistical analysis of hypothesis 2 itself did not reveal any promising results. Consequently, this NMR data analysis is not able to make any statements

about treatment effects related to darbepoetin alfa or placebo compound.
The identification of metabolites with statistical significance for hypothesis 1a, 1b, and/or 2 has not been successful. A possible explanation includes the concealment of probably differentiating metabolite signals by underlying broad, unspecific protein signals in the investigated 1D ^{1}H NOESY NMR spectra. A removal of this protein background by either prior ultrafiltration or acquisition of 1D ^{1}H CPMG NMR spectra, compare to section 4.2.2.1 and 4.2.2.2, was not attempted. However, statistical analysis of hypothesis 1b with a subset of specimens with minor protein content did not reveal any group separation. Furthermore, the statistical analysis of liquid chromatography-mass spectrometry (LC-MS) measurements on the same urine specimens corresponding to hypothesis 1b as for the NMR analysis, which was performed by M.Sc. Franziska Vogl, showed no clear group separation in a PCA plot and therefore supports NMR results. Hence, I would conclude that metabolic differences for hypothesis 1b are rather unlikely to be detected in this specimen cohort.
NMR analysis of specimens collected at additional time-points, individual patient information about treatment, specific outcome, clinical chemistry parameters, etc. might have facilitated further statistical investigations, but that data was not available to our lab. As already discussed for the GCKD study, compare to section 5.2.4, survival analyses with respect to overall patient survival as well as time elapsed until manifestation of ESRD by means of Cox regression [Cox 1972] with measured urinary NMR spectra of the TREAT study might have revealed interesting metabolites associated with poor survival and/or rapid onset of ESRD in this specific CKD entity. Nevertheless, the TREAT study offered a large urinary NMR spectral data set for patients belonging to a well defined CKD entity. After the acquisition of NMR spectra from urine specimens collected at the baseline and/or future follow-up time-points from the GCKD study, which will be realized in the future at the Institute of Functional Genomics, a comparison of urinary metabolic profiles from GCKD and TREAT study participants could be performed, e.g. for individual CKD entities. Moreover, urinary 1D ^{1}H NMR spectra of an apparently healthy control group, i.e. a subgroup of the German National Cohort study [German National Cohort Consortium 2014], would be already at hand [Schlecht et al. 2016]. By comparing urinary metabolic profiles of CKD patients of one study cohort, e.g. TREAT, and healthy individuals from the German National Cohort employing machine learning methods, one could determine a distinct molecular signature of CKD or entities thereof, which could be independently validated in the second CKD study cohort, i.e. GCKD or vice versa.

6 Conclusion and Perspectives

The main objective of this thesis was the application of NMR based metabolomics in the context of nephrology. This objective was both pursued in acute as well as chronic kidney diseases and could be subdivided in three specific aims.

The first aim was the detection of metabolic biomarkers for different kidney diseases as alternatives to traditional clinical approaches. In a prospective study of 85 unselected adult patients undergoing cardiac surgery with CPB, out of which 33 patients developed post-operative acute kidney injury, I identified novel low-molecular-weight factors for early detection of AKI. The most discriminative compounds included propofol-glucuronide, Mg^{2+}, and lactate 24 h after surgery. Elevated plasma levels of the glucuronide conjugate of propofol, an anesthetic agent which had been administered to all patients during surgery, in patients with AKI seem to be a surrogate marker for a general worsening of renal function even outperforming creatinine. An elevation of Mg^{2+} levels in AKI patients might be explained by its use for the treatment of cardiac arrythmias, and ischemic injury as well as systemic hypoperfusion present in AKI patients might be linked to elevated lactate levels in this group. This study had been a follow-up project of my master thesis [Zacharias 2012], where I statistically analyzed urine specimens collected before and at 4 and 24 h after cardiac surgery with CPB. Here, elevated urinary levels of tranexamic acid together with decreased urinary levels of carnitine and 2-oxoglutaric acid had shown best overall AKI prognostication at 24 h after surgery. The novel detected EDTA plasma biomarkers for the detection of AKI clearly outperformed these urinary biomarkers with respect to overall prognostication accuracy.
Moreover, I aimed at a reliable prognostication of AKI after cardiac surgery based on a small set of easily quantifiable endogenous metabolites. In this regard, a combination of plasma creatinine, plasma Mg^{2+}, and plasma lactate revealed overall best prognostication accuracy. I further employed this biomarker panel to derive a new AKIN index score, which revealed that the metabolic profiles of patients suffering from AKIN 1 disease were largely indistinguishable from those of patients not suffering from AKI. This finding might be connected to the rather mild nature of AKIN 1 disease. This novel biomarker panel would offer a reliable and swift diagnostic tool for the detection of AKI after cardiac surgery with CPB use only requiring easily implementable point-of-care technologies. This study underscores the power of NMR spectroscopy in combination with bioinformatics in identifying novel biomarkers of kidney disease and in gaining new insights into pathomechanisms. Nevertheless, larger prospective studies are required to validate the temporal development of these novel prognostic metabolites, and to further investigate how their temporal course relates to the onset, severity, and outcome of AKI. Furthermore, the prognostic validity of this biomarker panel has to be directly compared to that of other already established AKI biomarkers such as NGAL, IL-6, IL-18, KIM-1, L-FABP,

and NAG, and a combination of novel and established biomarkers might even further improve AKI prognostication.
In the context of chronic kidney diseases, specific metabolic fingerprints for the discrimination of 6 individual leading renal diseases were investigated in 1D ^{1}H CPMG spectra of EDTA plasma specimens collected at the baseline time-point of the GCKD study. This study represents the currently worldwide largest CKD patient cohort and offers an unprecedented dimensionality in NMR based metabolomic analyses. A significant up-regulation of plasma D-glucose concentrations could be reported for patients suffering from diabetic nephropathy, as well as significantly up-regulated plasma lipid signals for patients suffering from glomerulonephritis, respectively. These findings are in good concordance with standard clinical pathologies of diabetic nephropathy and glomerulonephritis. Further insights into pathologies of different renal diseases might be established with this specimen cohort in comparison to suitably acquired data from healthy individuals in the future. Comparing single CKD entities to a suitable healthy control group by means of classification, as, e.g. realized by [Gronwald et al. 2011] for urine specimens collected from patients suffering from ADPKD, might reveal as well as estimate the diagnostic value of novel biomarkers.
1D ^{1}H NOESY spectra of baseline urine specimens of a second large-scale clinical trial, the TREAT study including CKD patients with type-2 diabetes mellitus and concomitant anemia, have additionally been investigated to gain new insights into CKD progression and corresponding Hb responsiveness. A statistically significant up-regulation of proteins in urine specimens collected in the week directly prior to treatment start from patients who are going to die or develop ESRD in the future, respectively, could be reported. With proteinuria being both a diagnosis criterium for CKD and a predictor for ESRD as well as all-cause and cardiovascular mortality [KDIGO workgroup 2013, Iseki et al. 2003, Keane et al. 2003], these results are in good agreement with established pathophysiological findings in CKD patients. In the future, the urinary metabolic profiles of these patients suffering from CKD with type-2 diabetes mellitus and concomitant anemia might be compared to respective profiles from other CKD entities of the GCKD study. The acquisition of NMR spectra from urine specimens collected at the baseline and/or future follow-up time-points from the GCKD study will be realized at the Institute of Functional Genomics and urinary 1D ^{1}H NMR spectra of an apparently healthy control group [Schlecht et al. 2016] are already available. After completing the acquisition of urinary NMR spectra from the GCKD cohort, one might, for example, try to determine a distinct molecular signature for CKD or entities thereof in comparison to healthy controls employing, e.g., GCKD patients suffering from diabetic nephropathy. This molecular signature might be subsequently applied to a test set comprising urinary NMR metabolic fingerprints from TREAT study participants. Thereby, one would be able to validate this newly generated molecular signature of CKD patients suffering from diabetic nephropathy in an independent patient cohort.

The second aim of my Ph.D. thesis comprised the prediction of future kidney performance based on baseline metabolic fingerprints derived by NMR spectroscopy. The prediction of future renal performance is essential for timely interventions and improved patient care. Multiple regression analyses between NMR metabolic fingerprints derived from the baseline plasma specimen co-

hort of the GCKD study with the estimated GFR and specific renal performance markers, such as SCr and SCysC, clinically assessed both at the baseline as well as at the second follow-up time-point were performed employing the LASSO method. Furthermore, simple linear regression analyses employing baseline clinical parameters with respect to second follow-up clinical parameters were conducted. Considering baseline renal performance parameters, LASSO models derived from 1D ^{1}H NMR bucket intensities showed very good predictive performances.
The prediction of renal performance markers clinically assessed at the second follow-up time-point, however, was best achieved by simple linear regression models based on respective baseline clinical parameters. This result might be explained by the fact that the majority of the GCKD patients did not experience large changes in SCr, SCysC, or eGFR values over the investigated time period of two years. Additional regression analyses based on baseline NMR data with respect to future follow-up response variables might reveal different results. In this context, regression analyses might also be performed separately within specific CKD entities, e.g. only including patients suffering from diabetic nephropathy. As specific metabolic differences between individual leading renal disease groups do exist, such analyses might probably yield improved multiple regression analyses results based on NMR data due to more homogeneous metabolic profiles within one CKD entity. Moreover, possibly different CKD progression rates in individual CKD entities could be detected in this approach.
Additionally, survival analyses with respect to overall patient survival as well as time elapsed until onset of RRT by means of Cox regression with baseline or future follow-up 1D ^{1}H NMR EDTA plasma spectra might reveal interesting metabolites associated with poor survival and/or rapid need for RRT in this study cohort. Furthermore, the predictive performance of a newly developed reference point insensitive regression method, i.e. zero-sum regression [Altenbuchinger et al. 2016], is currently assessed at the Institute of Functional Genomics and might yield improved results.

The third aim of this Ph.D. thesis comprised general method developments and additions for NMR based metabolomics. While investigating novel compounds for the early detection of AKI after cardiac surgery with CPB use, it became apparent that the application of Variance Stabilization normalization on the EDTA plasma spectral data set yielded large intergroup differences in $CaEDTA^{2-}$ abundance by means of *t*-test based statistical analysis. This finding, however, is rather unlikely, since calcium levels are usually tightly regulated in the human body. In contrast, simple scaling of spectral features to the TSP reference signal followed by $\log_2$-transformation, confirmed for $CaEDTA^{2-}$ the absence of a significant intergroup difference. Utilizing this data preprocessing method, I could further reveal significant intergroup differences in $MgEDTA^{2-}$ levels. Thus, the choice of appropriate data normalization methods proofed to be crucial for correct data analysis and interpretation in this project and should always be thoroughly appraised. These results could be further verified by corresponding statistical analyses of the absolute concentrations of these metal ions by NMR spectroscopy.
Finally, the acquisition of 1D ^{1}H NMR spectra of EDTA plasma specimens collected from GCKD study participants posed several new challenges on both NMR data acquisition as well as analysis. To effectively suppress broad signals in 1D ^{1}H NMR spectra arising from protons in proteins, we preferred to utilize the CPMG pulse sequence over common protein removal strate-

gies. This facilitated the acquisition of well resolved 1D ^{1}H spectra in sufficiently low sample preparation and measurement times for high-throughput studies. The common NMR reference substance TSP proved to be inappropriate for unfiltered plasma specimens and alternatively, the reference substance formic acid was successfully implemented. This strategy can be easily employed for GCKD plasma specimens collected at future follow-up time-points. Moreover, a readily implementable spectral alignment tool is presented to eliminate overall spectral shifts, which can be also applied to other 1D NMR spectral data sets. Due to overall signal decay during the filtering period of the CPMG pulse sequence, which seems to be significantly influenced by the specific matrix composition of the investigated fluid, the absolute quantification of metabolites in unfiltered EDTA plasma specimens by means of 1D ^{1}H CPMG spectra is not straightforward. These matrix effects might be compensated by suitable calibration factors, which should be experimentally determined with the CPMG pulse sequence in pooled unfiltered EDTA plasma specimens from the GCKD study cohort. The corresponding analyses are currently performed by Jens Wallmeier and will be part of his Ph.D. thesis. The absolute quantification of plasma metabolites in this CKD cohort or entities thereof would facilitate further statistical analyses, including comparisons against reference concentrations in healthy subjects.

In summary, this Ph.D. thesis proofs the successful application of NMR-based metabolomics in combination with advanced bioinformatics to reveal novel markers for the early detection of renal diseases, as demonstrated in the context of acute kidney injury after cardiac surgery with cardiopulmonary bypass use. It moreover reports important method implementations for NMR-based investigations of large-scale clinical trials and presents first promising statistical analyses results for two studies on chronic kidney disease, which will be further evaluated at the Institute of Functional Genomics.

7 Appendix

7.1 Appendix I: General *R*-Code

This section provides general R commands employed for statistical data analysis of this thesis. For each command, short commentaries are given comprising a brief description of the respective command. Instead of containing specific names of R objects or R settings, the R commands displayed in this R-code are illustrated using place holder variables. These variables are printed in bold, italic type and need to be replaced by the respective R objects or settings in an actual analysis procedure. More details and an exact documentation of the used libraries and functions can be found on the specific help/documentary pages by either using the R commands `help(`***R.command***`)` or `?(`***R.command***`)`, or at `http://www.r-project.org/`.

7.1.1 Get familiar with data

```
x<-read.table("bucket.table", as.is=T, header=T, row.names=1)    # import bucket table
x1<-as.matrix(t(x))    # transpose bucket table, in x1, rows=buckets, columns=samples
kick.out<-apply(x1,1,function(z)all(z==0))    # collect all rows/buckets completely filled with zero
x2<-x1[!kick.out,]    # eliminate all rows/buckets only containing zeros
dim(x2)    # number of rows/buckets and columns/samples
colnames(x2)    # sample names
rownames(x2)    # bucket names
```

7.1.2 Choose normalization method

```
w3<-x2[,c(sample.IDs)]    # sort samples according to comparable groups
colnames(w3)    # check whether sample order is correct

areas<-apply(w3, 2, sum)    # sum over all buckets for each sample=area under the curve
plot(density(areas))    # plot areas under the curve as density plot → first indication, whether data is normally distributed or
not

shapiro.test(areas)    # Shapiro-Wilk test in order to find out whether the complete data set is normally distributed, H0:
data is normally distributed
# if normality does not show significant p-value, data can be either quantile or vs normalized (→ go to 7.1.3.2 or 7.1.3.3, respectively)

min(w3)    # indicates whether there are values <0, be aware if log2 transformation is applied
min(w3[w3>0])    # check scale of data
max(w3[w3<0])    # check scale of data
which(w3==0)    # indicates whether there are values=0, if there are values=0, then apply * before log2 transformation

# if there are values <0 in the data, take absolute values if log2 transformation is applied (compare to 7.1.3.1)
# if scale of data <10, scale data correspondingly (e.g. scale of data about 10^(-8), multiply data with 10^8 (i.e. 10^(scale))) for log2
transformation as well as quantile or vs normalization
```

7.1.3 Normalization

7.1.3.1 $\log_2$ transformation

```
w3.abs<-abs(w3)    # take absolute values in order to avoid problems with log2 transformation
w3.abs=w3.abs*10^(scale)    # scale data in order to avoid computing problems
*  w3.abs<-w3.abs+10^(- scale)    # avoid values=0 for log2 transformation, only apply if which(w3==0)!=integer(0)
w3.abs.log<-log2(w3.abs)    # log2 transformation, perform all following data analyses with this object
```

7.1.3.2 Quantile normalization

```
w3=w3*10^(scale)    # scale data in order to avoid computing problems
library(affy)    # quantile normalization library
normalize.quantile<-get("normalize.quantiles",en=asNamespace("affy"))    # quantile normalization function
quantile.w3<-normalize.quantile(w3)    # apply quantile normalization to data and obtain quantile normalized data, perform all following data analyses with this object
```

7.1.3.3 Variance Stabilization normalization

```
w3=w3*10^(scale)    # scale data in order to avoid computing problems
library(vsn)    # VSN library
vsn.model<-vsn2(w3)    # calculate VSN model
vsn.w3<-predict(vsn.model,w3)    # apply VSN model to data and obtain VS normalized data, perform all following data analyses with this object
```

7.1.4 Data analysis

all data analysis is performed with the properly normalized data set (i.e. w3.abs.log, quantile.w3 or vsn.w3), which is from now on called ***data.set***

7.1.4.1 Unsupervised machine learning methods

Principal component analysis (PCA)

```
pc<-prcomp(t(data.set))    # PCA
plot(pc$x[1:first.group.IDs,PC],pc$x[1:first.group.IDs,PC],main="name",xlab="PC",
ylab="PC",pch=symbol,col="color", xlim=c(x.margins),ylim=c(y.margins))    # plot first group
points(pc$x[first.group.IDs+1:second.group.IDs,PC],
pc$x[first.group.IDs+1:second.group.IDs,PC],pch=symbol,col="color")    # plot second group
# etc.

library(missMDA)    # required library
pca<-PCA(t(data.set),scale.unit=FALSE,axes=c(PC,PC))    # perform same PCA plot including sample IDs, loadings plot and information about percentages of variances
```

7.1.4.2 Supervised methods

***t*-test with "multtest"**

```
l1<-numeric(sample.number)    # define labels
for (i in first.group.IDs+1:second.group.IDs) {l1[i]=1}    # set labels for second group
library(multtest)    # required library
t<-mt.teststat(data.set, classlabel=l1, test="t")    # calculate t-statistics
pt<-2*pt(-abs(t),df=degrees.of.freedom)    # calculate p-values, set degrees of freedom df=sample.number-2
tmp<-rownames(data.set)    # vector with bucket names (be aware that after quantile normalization, the object quantile.w3 has no more bucket names, therefore use unnormalized object w3)
pt2<-data.frame(tmp,pt)    # create data frame combining bucket names and p-values
pt3<-pt2[order(pt2$pt),]    # order pt2 according to calculated p-values in ascending order
pAdjusted<-mt.rawp2adjp(pt,proc=c("BH"))    # calculate B/H-adjusted p-values
pt4<-data.frame(pt3$tmp,pAdjusted$adjp)    # create data frame with ordered bucket names, raw p-values and B/H-adjusted p-values
pt4[1:last.significant.bucket.ID,]    # list significant buckets
pt5 <- cbind(pt4, as.numeric(rownames(pt3)))    # get bucket IDs
pt5[1:last.significant.bucket.ID,]    # list of significant buckets with bucket IDs
```

Statistical power calculation

```
library("pwr")    # required library
library("compute.es")    # required library

m.1<-vector(mode="numeric",length=number.of.sign.buckets)
m.2<-vector(mode="numeric",length=number.of.sign.buckets)
sd.1<-vector(mode="numeric",length=number.of.sign.buckets)
sd.2<-vector(mode="numeric",length=number.of.sign.buckets)

for(i in 1:number.of.sign.buckets){
m.1[i]<-mean(data.set[which(rownames(data.set)==rownames(sign.buckets)[i]),first.group.ids])
m.2[i]<-mean(data.set[which(rownames(data.set)==rownames(sign.buckets)[i]),second.group.ids])
sd.1[i]<-sd(data.set[which(rownames(data.set)==rownames(sign.buckets)[i]),first.group.ids])
sd.2[i]<-sd(data.set[which(rownames(data.set)==rownames(sign.buckets)[i]),second.group.ids])
}

n.1<-dim(data.set[,second.group.ids])[2]
n.2<-dim(data.set[,second.group.ids])[2]
d<-list()
d.d<-vector(mode="numeric",length=number.of.sign.buckets)
d.d.p<-vector(mode="numeric",length=number.of.sign.buckets)

for(i in 1:number.of.sign.buckets){
d[[i]]<-mes(m.1[i],m.2[i],sd.1[i],sd.2[i],n.1,n.2)
d.d[i]<-d[[i]]$d
d.d.p[i]<-d[[i]]$pval.d
}

p<-list()
p.p<-vector(mode="numeric",length=number.of.sign.buckets)

for(i in 1:number.of.sign.buckets){
p[[i]]<-pwr.t2n.test(n1=n.1,n2=n.2,d=d.d[i],sig.level=0.05,power=NULL,alternative="two.sided")
p.p[i]<-round(((as.numeric(p[[i]]$power))*100),digits=2)
}
```

Heat-map for significant buckets

```
as.numeric(rownames(pt3))[1:last.significant.bucket.ID]    # get significant bucket IDs
w5<-data.set[c(significant.bucket.IDs separated by ","),]    # obtain data object from original normalized data.set with
```

all desired samples and all significant buckets ordered by p-value

```
samples<-colnames(data.set)    # sample names
samples    # check sample order

bins<-rownames(w3[c(significant.bucket.IDs separated by ","),])    # obtain bucket names
bins    # check bucket names and order

library(compdiagTools)    # required library
geneImager(w5,cluster="FALSE",dendro="FALSE",hide.colorbar=TRUE,
annotSamples=samples,annotProbes=bins,scale=TRUE, file="filename.jpg")    # plot heat-map without clustering, if clustering is desired, cluster="TRUE"
```

Correlation calculation for heat-map:

```
vx<-as.numeric(sample.number)    # vector for all samples
vcor<-as.numeric(number.of.significant.buckets)    # vector for all significant buckets
for (i in 1:first.group.IDs) {vx[i]=1}    # labels for samples
for (i in first.group.IDs+1:second.group.IDs) {vx[i]=0}    # labels for samples
for (i in 1:number.of.significant.buckets) {vcor[i]<-cor(w5[i,],vx)}    # calculate correlation coefficient for significant
```

buckets between buckets and class labels

```
z1<-data.frame(w5,vcor)    # create data frame consisting of sample × significant buckets and correlation coefficients
z2<-z1[order(z1$vcor),]    # order data frame according to correlation coefficients
```

```
sx<-data.frame(bins,vcor)    # create data frame consisting of bucket names and correlation coefficients
sx2<-sx[order(sx$vcor),]    # order data frame according to correlation coefficients
library(compdiagTools)    # required library
geneImager(z2[,1:sample.number],cluster="FALSE",dendro="FALSE",
annotSamples=samples,annotProbes=sx2$bins,hide.colorbar=TRUE,scale=TRUE, file="filename.jpg")    # plot heat-map without clustering, if clustering is desired, cluster="TRUE"
```

ANOVA/*t*-tests with "limma"

```
labels<-vector(length=sample.number)    # create vector for group labels
labels[1:first.group.IDs]=1    # labels for first group
labels[first.group.IDs+1:second.group.IDs]=2    # labels for second group
# etc.

data<-cbind(colnames(data.set),labels)    # combine sample names and labels
data    # check if labels vector is correct

library(limma)    # required library for ANOVA
design<-model.matrix( 0+factor(labels))    # create design matrix
colnames(design)<-c("first.group.name","second.group.name",etc.)    # assign colnames for design matrix according to groups
design    # check if design matrix/labels is correct

fit<-lmFit(data.set,design)    # fit linear model for data.set with design matrix
cont<-makeContrasts("first.group.name-second.group.name", "first.group.name-third.group.name",etc., levels=design)
# create contrast matrix for all group comparisons
fit2<-contrasts.fit(fit,cont)    # given a linear model, compute estimated coefficients and standard errors for a given set of contrasts
fit3<-eBayes(fit2)    # given a series of related parameter estimates and standard errors, compute moderated t-statistics, moderated F-statistic, and log-odds of differential expression by empirical Bayes shrinkage of the standard errors towards a common value
cont    # check if contrast matrix is correct
ANOVA<-topTable(fit3,coef=c(1:number.of.comparisons),p.value=0.05, adjust="BH",number=755)
ANOVA    # results of F-test
firstgroupname.secondgroupname<-topTable(fit3,coef=1,adjust="BH", number=number.of.significant.buckets)
firstgroupname.secondgroupname    # results of t-test for first comparison
firstgroupname.thirdgroupname<-topTable(fit3,coef=2,adjust="BH", number=number.of.significant.buckets)
firstgroupname.thirdgroupname    # results of t-test for second comparison
# etc.
```

7.1.4.3 Regression analysis with LASSO method

```
library(glmnet)    # required library
set.seed(2014)
LASSO.model <- cv.glmnet(t(data.train),log2(response.train))    # train LASSO model
predicted.values <- predict(LASSO.model, t(data.test),s="lambda.min")    # predict test data
coef <- predict(LASSO.model, s=LASSO.model$lambda.min, type="coefficients")    # get coefficients of LASSO model
```

7.1.5 NMR bucket alignment and bucket fusion

```
adjust.buckets <- function(data,bucket.interval)    # function for bucket alignment
{
max.bucket.list <- list()
shifts <- list()
data.temp <- matrix(NA, ncol=ncol(data),nrow=length(bucket.interval)+nrow(data))
for(j in 1:ncol(data))
{
max.bucket <- which.max(data[bucket.interval,j])    # find bucket with max. bucket intensity in bucket.interval
max.bucket.list[[j]] <- bucket.interval[max.bucket]    # record bucket with max. bucket intensity in bucket.interval
data.temp[,j] <- c(rep(NA, length(bucket.interval)-max.bucket),data[,j],rep(NA, max.bucket))    # shift spectral data of spectrum j by (length(bucket.interval)-max.bucket)
shifts[[j]] <- length(bucket.interval)-max.bucket    # record shift
}
max.bucket.list.unlist <- unlist(max.bucket.list)
```

```
names(max.bucket.list.unlist) <- colnames(data)
shifts <- unlist(shifts)
names(shifts) <- colnames(data)
feature.selected <- as.integer(names(which.max(summary(as.factor(max.bucket.list.unlist)))))    # bucket number that
is mainly selected as maximum bucket
sel.samples <- names(max.bucket.list.unlist)[max.bucket.list.unlist==feature.selected]    # spectra, where most abun-
dant bucket is selected
rownames(data.temp) <- c(paste("temp.  feature",1:shifts[sel.samples[1]]),rownames(data),paste("temp.  feature",
(1+shifts[sel.samples[1]]):(length(bucket.interval))))    # name buckets of aligned data
colnames(data.temp) <- colnames(data)    # name spectra of aligned data
return(list(data.temp,highest.bucket.in.interval=max.bucket.list.unlist))
}

reduce.buckets <- function(data,start,number.of.combined.buckets)    # function for bucket fusion
{
n <- number.of.combined.buckets
s <- start
pos <- s
data.temp <- NULL
temp.rownames <- list()
for(i in 1:floor((nrow(data)-s+1)/n)){
data.for <- data[pos:(pos+n-1),]
data.temp <- rbind(data.temp,apply(data.for,2,sum))
data.vec <- round(mean(as.numeric(gsub("X","",gsub("X........","",rownames(data.for))))),digits=4)
temp.rownames[[i]] <- paste("X........",data.vec,sep="")
pos <- pos+n
}
rownames(data.temp) <- unlist(temp.rownames)
return(data.temp)
}
```

7.2 Appendix II: Acute Kidney Injury study

7.2.1 Clinical characteristics and outcome of patients included in AKI study

	a) AKI urine study			b) AKI plasma study		
	Non-AKI patients	**AKI patients**	*P*-value	**Non-AKI patients**	**AKI patients**	*P*-value
N	72	34		52	33	
Age [years]	68 ± 10	73 ± 8	0.011[b]	68 ± 11	73 ± 8	0.008[b]
Sex, male	58 (80.6%)	23 (67.6%)	0.131[c]	43 (82.7%)	22 (66.7%)	0.11[c]
Weight [kg]	82.2 ± 13.2	82.6 ± 16.5	0.90[b]	82.8 ± 13.7	83.3 ± 16.2	0.88[b]
Height [cm]	171.2 ± 6.5	169.5 ± 7.4	0.25[b]	171.5 ± 6.5	169.2 ± 7.4	0.15[b]
BMI [kg/m^2]	28 ± 4	29 ± 5	0.481[b]	28 ± 4	29 ± 5	0.404[b]
BSA [m^2][d]	1.94 ± 0.17	1.93 ± 0.20	0.75[b]	1.95 ± 0.17	1.9 ± 0.2	0.72[b]
SCr [mg/dl]	1.16 ± 0.41	1.36 ± 0.53	0.057[b]	1.19 ± 0.44	1.35 ± 0.53	0.15[b]
eGFR 13-60 ml/min per 1.73 m^2	27	23	0.006[c]	21 (40.4%)	22 (66.7%)	0.026[c]
eGFR > 60 ml/min per 1.73 m^2	45	11	0.006[c]	31 (59.6%)	11 (33.3%)	0.026[c]
Comorbid disease, n						
Diabetes mellitus	26 (36.1%)	18 (52.9%)	0.139[c]	20 (38.5%)	18 (54.6%)	0.182[c]
Hypertension	70 (97.2%)	34 (100%)	1.000[c]	51 (98.1%)	33 (100%)	1.000[c]
Cardiac insufficiency	57 (79.2%)	30 (88.2%)	0.293[c]	38 (73.1%)	29 (87.9%)	0.172[c]
Atrial fibrillation	10 (13.9%)	8 (23.5%)	0.27[e]	9 (17.3%)	8 (24.2%)	0.579[e]
Cerebrovascular disease	24 (33.3%)	15 (44.1%)	0.29[e]	16 (30.8%)	14 (42.4%)	0.353[e]

Hypercholesterolemia	67 (93.1%)	30 (88.2%)	0.464[c]	48 (92.3%)	29 (87.9%)	0.705[c]
Acute myocardial infarction	25 (34.7%)	15 (44.1%)	0.395[c]	20 (38.5%)	15 (45.5%)	0.652[c]
COPD	12 (16.7%)	7 (20.6%)	0.601[c]	11 (21.2%)	7 (21.2%)	1.000[c]
PAVD	10 (13.9%)	6 (17.7%)	0.772[c]	7 (13.5%)	6 (18.2%)	0.554[c]
Cardiovascular disease	24 (33.3%)	15 (44.1%)	0.291[c]	16 (30.8%)	14 (42.4%)	0.353[c]
CKD	17 (23.6%)	22 (64.7%)	0.00008[c]	12 (23.1%)	21 (63.6%)	0.0003[c]
Valvular heart disease	36 (50.0%)	25 (73.5%)	0.034[c]	24 (46.2%)	24 (72.7%)	0.024[c]
cardiac surgical interventions in the past	1 (1.4%)	6 (17.7%)	0.004[e]	1 (1.9%)	6 (18.2%)	0.013[e]
AKI in the past	2 (2.8%)	3 (8.8%)	0.325[c]	2 (3.9%)	3 (9.1%)	0.372[c]
RRT in the past	2 (2.8%)	0 (0%)	1.000[c]	2 (3.9%)	0 (0%)	0.519[c]
IABP pre-op	1 (1.4%)	2 (5.9%)	0.240[b]	1 (1.9%)	2 (6.1%)	0.557[c]
Preoperative medication, n						
Statin pre-op	53 (73.6%)	25 (73.5%)	1.000[c]	34 (65.4%)	24 (72.7%)	0.633[c]
ACE inhibitor	56 (77.8%)	27 (79.4%)	1.000[c]	41 (78.9%)	27 (81.8%)	0.788[c]
Beta-blocker	53 (73.6%)	27 (79.4%)	0.631[c]	40 (76.9%)	26 (78.8%)	1.000[c]
Other antihypertensive drugs pre-op	12 (16.7%)	11 (32.4%)	0.081[c]	11 (21.2%)	11 (33.3%)	0.309[c]
Insulin pre-op	9 (12.5%)	7 (20.6%)	0.383[c]	7 (13.5%)	7 (21.2%)	0.381[c]
Oral anti-diabetic medication	15 (20.8%)	11 (32.4%)	0.230[c]	11 (21.2%)	11 (33.3%)	0.309[c]
NSAID	1 (1.4%)	0 (0%)	1.000[c]	35 (67.3%)	22 (66.7%)	1.000[c]
Type of surgery, n						
CABG	54 (75%)	16 (47.1%)	0.008[c]	40 (76.9%)	16 (48.5%)	0.01[c]
Aortic valve surgery	6 (8.3%)	7 (20.6%)	0.110[c]	6 (11.5%)	7 (21.2%)	0.354[c]
Mitral valve surgery	2 (2.8%)	2 (5.9%)	0.592[c]	1 (1.9%)	2 (6.1%)	0.557[c]
CABG + aortic valve surgery	5 (6.9%)	7 (20.6%)	0.051[c]	2 (3.9%)	6 (18.2%)	0.051[c]
CABG + mitral valve surgery	1 (1.4%)	1 (2.9%)	0.541[c]	0 (0%)	1 (0.03%)	0.388[c]
Thoracic aortic surgery	4 (5.6%)	1 (2.9%)	1.000[c]	3 (5.8%)	1 (3.0%)	1.000[c]
Surgery data						
Bypass time period [min]	83.9 ± 30.2	89.9 ± 39.4	0.44[b]	83.8 ± 29.1	90.4 ± 39.9	0.42[b]
Aortic clamping time [min]	49.7 ± 22.4	54.7 ± 21.0	0.27[b]	48.5 ± 20.4	55.0 ± 21.3	0.19[b]
Reperfusion time [min]	28.1 ± 10.1	30.4 ± 20.7	0.55[b]	28.2 ± 10.8	30.5 ± 21.0	0.56[b]
RCC, n	27 (37.5%)	26 (76.5%)	0.0003[e]	22 (42.3%)	25 (75.8%)	0.004[e]
RCC [ml]	666.7 ± 336.3	715.4 ± 530.4	0.69[b]	695.5 ± 363.2	732.0 ± 534.4	0.78[b]
Thrombocyte concentrate, n	30 (41.7%)	19 (55.9%)	0.21[e]	22 (42.3%)	18 (54.5%)	0.373[e]
Thrombocyte concentrate [ml]	253.3 ± 90.0	271.1 ± 147.5	0.64[b]	240.9 ± 68.4	275.0 ± 150.7	0.38[b]
FFP, n	13 (18.1%)	15 (44.1%)	0.009[e]	10 (19.2%)	15 (45.5%)	0.014[e]
FFP [ml]	1092.3 ± 175.4	936.7 ± 405.5	0.19[b]	1080.0 ± 193.2	936.7 ± 405.5	0.25[b]
Crystalloid solution, n	72 (100%)	34 (100%)	1.0[c]	52 (100%)	33 (100%)	1.0[c]
Crystalloid solution [ml]	1405.1 ± 487.6	1344.3 ± 393.6	0.50[b]	1503.1 ± 465.2	1354.7 ± 394.9	0.12[b]
Colloid solution, n	11 (15.3%)	2 (5.9%)	0.22[e]	9 (17.3%)	2 (6.1%)	0.190[e]
Colloid solution [ml]	500.0 ± 0.0	500.0 ± 0.0		500.0 ± 0.0	500.0 ± 0.0	
Minimum MAP [mmHg]	38.2 ± 7.5	36.5 ± 9.2	0.34[b]	38.9 ± 8.1	36.5 ± 9.3	0.24[b]
Lowest body temperature during CPB [°C]	35.3 ± 0.9	35.4 ± 0.9	0.59[b]	35.4 ± 0.7	35.4 ± 0.9	0.96[b]
Intraoperative HF, n	3 (4.2%)	3 (8.8%)	0.38[e]	2 (3.9%)	3 (9.1%)	0.372[e]
Postoperative IABP, n	1 (1.4%)	9 (26.5%)	0.0001[e]	1 (1.9%)	9 (27.3%)	0.001[e]
Need for catecholamines, n	72 (100%)	32 (94.1%)	0.10[e]	52 (100%)	31 (94.0%)	0.148[e]

Table 7.1: **Previous page: Clinical characteristics and outcome of all patients included in a) AKI urine and b) AKI plasma study.** [a]Patients with AKI were diagnosed based on AKIN-criteria. For this, serum samples of the second post-operative day were taken into account. [b]P-values calculated using two-sided t-test assuming unequal variance with Microsoft Office EXCEL 2007. [c]P-values calculated using Fisher's exact test with Microsoft Office EXCEL 2007. [d]Body surface area calculated employing the DuBois formula [DuBois 1916]. [e]P-values calculated using Fisher's exact test with R version 3.2.0. Abbreviations: ACE, angiotensin-converting-enzyme; AKI, acute kidney injury; BMI, body mass index; BSA, body surface area; CABG, coronary artery bypass grafting; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CPB, cardiopulmonary bypass; eGFR, estimated glomerular filtration rate; FFP, fresh frozen plasma; HF, hemofiltration; IABP, intra-aortic balloon pump; MAP, mean arterial pressure; NSAID, non-steroidal anti-inflammatory drug; PAVD, peripheral arterial vascular disease; RCC, red cell concentrate; RRT, renal replacement therapy. Modified from [Zacharias et al. 2013a), Zacharias et al. 2015].

7.2.2 CPB protocol

[1] The CPB circuit was prefilled with 500 ml of Ringer's solution, 500 ml of 10% Mannitol, and 500 ml of 6% Voluven supplemented with 5000 IU of Heparin and 2 g of tranexamic acid. The CPB was conducted using normothermia, non-pulsatile blood flow, and α-stat pH management.

[1]The following section has been published in [Zacharias et al. 2015].

7.2.3 Spike-In experiments for the quantification of free calcium and magnesium levels

a) Spike-In experiments in H_2O				
Added	0.50mM Ca + 1.00mM Mg	0.75mM Ca + 0.75mM Mg	1.00mM Ca + 0.50mM Mg	-
Measured	0.50mM Ca; 1.04mM Mg	0.73mM Ca; 0.77mM Mg	0.95mM Ca; 0.50mM Mg	-
Recovery	100% Ca; 104% Mg	97% Ca; 103% Mg	95% Ca; 100% Mg	-
Mean recovery	97 ± 2.5% Ca; 102 ± 2.1% Mg			
b) Spike-In experiments in human plasma				
Added	0.25mM Ca + 1.00mM Mg	0.50mM Ca + 0.75mM Mg	0.75mM Ca + 0.50mM Mg	1.00mM Ca + 0.25mM Mg
Measured	0.23mM Ca; 1.08mM Mg	0.44mM Ca; 0.81mM Mg	0.70mM Ca; 0.52mM Mg	1.06mM Ca; 0.24mM Mg
Recovery	92% Ca; 108% Mg	88% Ca; 108% Mg	93% Ca; 104% Mg	106% Ca; 96% Mg
Mean recovery	95 ± 6.8% Ca; 104 ± 4.9% Mg			

Table 7.2: **Spike-In experiments for the absolute quantification of free calcium and magnesium levels in H_2O and human plasma.** a) For H_2O spike-in experiments, 260 µl H_2O containing 2 mM EDTA were mixed with stock solutions containing Ca^{2+} and Mg^{2+} ions to give a final volume of 400 µl. The concentrations given here are the concentrations contained in 400 µl of sample. b) For plasma spike-in experiments, 350 µl of pooled EDTA plasma were mixed with stock solutions containing Ca^{2+} and Mg^{2+} ions to give a final volume of 400 µl. For plasma the concentrations given below are the added amounts in 400 µl of sample. EDTA was obtained from Carl Roth GmbH and Mg-chloride-hexahydrate and Ca-chloride-2-hydrate were purchased from Sigma-Aldrich. Modified from [Zacharias et al. 2015].

7.2.4 Time-course development

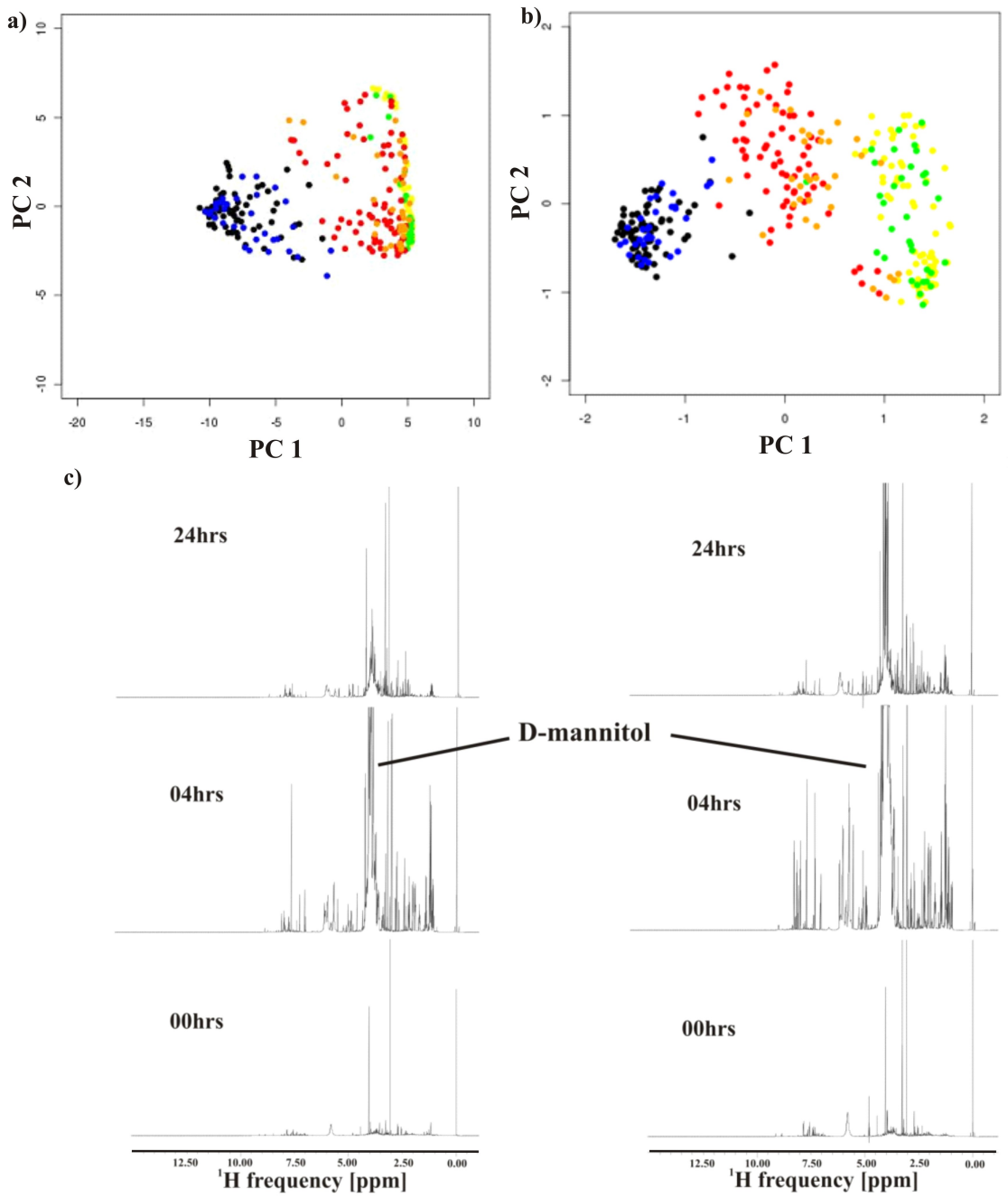


Figure 7.1: **Principal component analysis of all 318 urine specimens collected before and at 4 and 24 h after surgery.** For the time-point before surgery the data of the 72 non-AKI and 34 AKI patients are plotted in black and blue, respectively. For the first time-point after surgery, data of the non-AKI and AKI patients are shown in yellow and green, respectively, while at 24 h after surgery, data of the non-AKI and AKI patients are colored in red and orange, respectively. a) All NMR features were used. PC 1 and PC 2 explain 64.72% and 17.30% of the variance, respectively. Modified from [Zacharias et al. 2013a)]. b) Features corresponding to D-mannitol were excluded from the PCA. PC 1 and PC 2 explain 30.82% and 9.96% of the variance, respectively. The group separation is now mainly driven by creatinine. Modified from [Zacharias et al. 2013a)]. c) Representative 1D ^{1}H NMR spectra of urine specimens collected 0 h pre-op, 4 h post-op and 24 h post-op for one non-AKI patient (left side) and one AKI patient (right side). These ^{1}H urinary spectra were individually scaled by eye to creatinine for each patient. The higher abundance of D-mannitol especially at 4 h after surgery in comparison to pre-surgery urine specimens is apparent. Modified from [Zacharias 2012].

7.2.5 Results of permutation tests

Total accuracy [%]	AUC	Number of selected features	Optimized m_{try}	Optimized n_{tree}	Sensitivity [%]	Specificity [%]
55.7 ± 5.1	0.48 ± 0.08	117.1 ± 2.7	8.9 ± 1.7	270 ± 41.0	17.1 ± 7.2	80.2 ± 6.3

Table 7.3: **Classification performance on randomly permuted data.** Mean values ± standard deviations of 20 nested cross-validation runs of an RF algorithm trained on 85 plasma 1D ^{1}H NMR spectra with randomly permuted class-labels [Zacharias et al. 2015]. Given are the total accuracy, area under the ROC curve (AUC), number of selected features, optimal number of tried variables m_{try}, number of grown trees n_{tree}, sensitivity, and specificity.

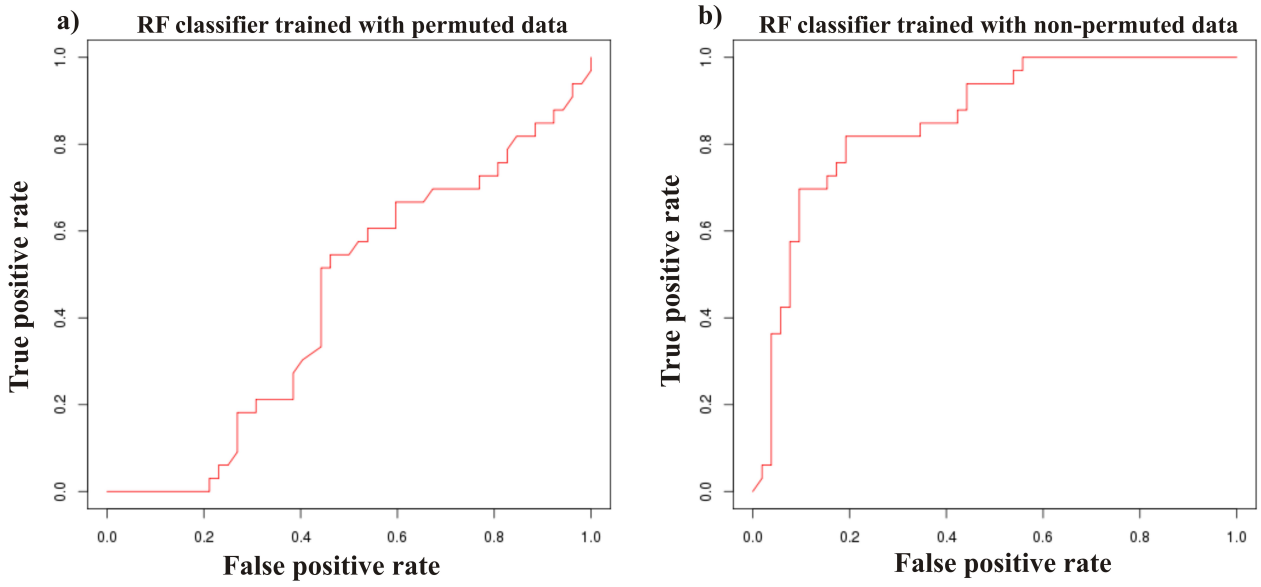


Figure 7.2: **Exemplary receiver operating characteristic (ROC) curves of the RF classifier.** a) RF trained on 85 plasma 1D ^{1}H NMR spectra with randomly permuted class-labels. b) RF trained on 85 plasma 1D ^{1}H NMR spectra with non-permuted class-labels. The x-axis denotes the false positive rate (1 - specificity), the y-axis the true positive rate (sensitivity). The area under the ROC curve is an indicator for the power of the used classifier (an area of one represents the ideal classifier). The ROC curves were obtained from nested cross-validation runs with inner cross-validation for parameter calibration. Modified from [Zacharias et al. 2015].

7.2.6 Discriminative 24 h plasma NMR features

ID	Spectral position [ppm]	*P*-value un-adjusted	*P*-value B/H-adjusted	Identified compounds
222	7.285	2.87e-11	2.06e-08	Propofol-glucuronide, tryptophan
361	4.305	3.24e-10	1.16e-07	Multiple compounds
223	7.275	4.56e-09	1.09e-06	Tryptophan, propofol-glucuronide
654	1.165	6.34e-09	1.14e-06	4-Hydroxy-propofol-1-OH-D-glucuronide, propofol-glucuronide, isopropanol[1]
494	2.765	9.76e-09	1.40e-06	Unknown
182	7.685	1.95e-08	2.34e-06	Unknown
444	3.285	2.28e-08	2.34e-06	Myo-inositol, D-glucuronic acid (?), phenylalanine, 4-hydroxy-propofol-4-OH-D-glucuronide
167	7.835	2.78e-08	2.49e-06	Hippuric acid
497	2.735	3.60e-08	2.87e-06	$MgEDTA^{2-}$, unknown
451	3.195	4.17e-08	2.99e-06	Acetyl-L-carnitine, propionylcarnitine (?), $CaEDTA^{2-}$
489	2.815	5.71e-08	3.17e-06	Unknown
506	2.645	6.01e-08	3.17e-06	Citric acid
647	1.235	6.50e-08	3.17e-06	Unknown
360	4.315	6.51e-08	3.17e-06	Multiple compounds
632	1.385	6.73e-08	3.17e-06	Tranexamic acid
362	4.295	7.07e-08	3.17e-06	Multiple compounds
504	2.665	7.58e-08	3.20e-06	Citric acid, unknown
390	4.015	8.71e-08	3.48e-06	Isopropanol[1], unknown
385	4.065	1.03e-07	3.91e-06	Creatinine, myo-inositol, choline (?)
280	6.705	2.09e-07	7.52e-06	4-Hydroxy-propofol-1-OH-D-glucuronide
662	1.085	2.39e-07	8.18e-06	Tranexamic acid, propofol-glucuronide
357	4.345	2.77e-07	8.55e-06	Unknown, 3-hydroxyglutaric acid (?), glycerophosphocholine (?)
663	1.075	2.80e-07	8.55e-06	Tranexamic acid, propofol-glucuronide
648	1.225	2.95e-07	8.55e-06	Unknown
183	7.675	3.11e-07	8.55e-06	Unknown
503	2.675	3.12e-07	8.55e-06	Citric acid, $MgEDTA^{2-}$
664	1.065	3.21e-07	8.55e-06	Unknown, tranexamic acid, propofol-glucuronide
633	1.375	4.04e-07	1.01e-05	Tranexamic acid
482	2.885	4.14e-07	1.01e-05	Tranexamic acid
635	1.355	4.23e-07	1.01e-05	2-Hydroxybutyric acid, tranexamic acid
368	4.235	4.48e-07	1.02e-05	Threonine, sucrose
358	4.335	4.55e-07	1.02e-05	Multiple compounds
631	1.395	4.90e-07	1.07e-05	Tranexamic acid
587	1.835	5.70e-07	1.20e-05	Tranexamic acid
634	1.365	6.51e-07	1.34e-05	2-Hydroxybutyric acid, tranexamic acid
384	4.075	6.95e-07	1.35e-05	Myo-inositol, D-glucuronic acid (?), choline (?)
382	4.095	6.98e-07	1.35e-05	Unknown, lactic acid, D-glucuronic acid (?)

585	1.855	8.30e-07	1.57e-05	Tranexamic acid
588	1.825	8.62e-07	1.59e-05	Tranexamic acid
586	1.845	9.56e-07	1.72e-05	Tranexamic acid
507	2.635	1.22e-06	2.08e-05	Methionine, acetyl-L-carnitine
502	2.685	1.22e-06	2.08e-05	$MgEDTA^{2-}$
492	2.785	1.43e-06	2.38e-05	Unknown
412	3.755	1.69e-06	2.75e-05	D-glucose, D-mannitol
389	4.025	1.74e-06	2.77e-05	D-gluconic acid, isopropanol[1]
549	2.215	1.82e-06	2.84e-05	Unknown
154	7.965	1.88e-06	2.86e-05	Unknown
500	2.705	1.91e-06	2.86e-05	$MgEDTA^{2-}$
418	3.695	1.96e-06	2.87e-05	D-mannitol, threitol (?), D-gluconic acid
246	7.045	2.07e-06	2.97e-05	Unknown
607	1.635	2.45e-06	3.44e-05	Tranexamic acid, unknown
417	3.705	2.50e-06	3.45e-05	D-glucose, D-mannitol, threitol, propofol-glucuronide
487	2.835	2.58e-06	3.50e-05	Methylguanidine (?)
493	2.775	2.72e-06	3.60e-05	Unknown
419	3.685	2.76e-06	3.60e-05	D-mannitol, D-gluconic acid, propofol-glucuronide
404	3.875	2.83e-06	3.61e-05	D-mannitol
465	3.055	2.87e-06	3.61e-05	Creatinine, tyrosine, unknown
405	3.865	3.49e-06	4.33e-05	D-mannitol
630	1.405	4.06e-06	4.94e-05	Tranexamic acid
410	3.815	4.20e-06	5.02e-05	D-mannitol , D-glucose, D-gluconic acid
445	3.275	4.89e-06	5.76e-05	Betaine, unknown, D-glucuronic acid (?), 4-hydroxy-propofol-4-OH-D-glucuronide
498	2.725	5.04e-06	5.84e-05	$MgEDTA^{2-}$, dimethylamine (?), unknown
566	2.045	5.15e-06	5.87e-05	N2-acetyl-L-ornithine (?), unknown, 2-hydroxyisovaleric acid
466	3.045	5.79e-06	6.50e-05	Creatinine, creatine, phosphocreatine
166	7.845	6.24e-06	6.89e-05	Hippuric acid
387	4.045	6.37e-06	6.89e-05	D-gluconic acid
264	6.865	6.43e-06	6.89e-05	Unknown, 4-hydroxyphenylacetic acid, 4-hydroxyphenyllactic acid
388	4.035	6.64e-06	7.01e-05	D-gluconic acid, isopropanol[1]
359	4.325	7.21e-06	7.50e-05	Multiple compounds
369	4.225	7.86e-06	8.06e-05	Unknown
364	4.275	8.26e-06	8.35e-05	Threonine
658	1.125	8.52e-06	8.49e-05	2-Oxoisovaleric acid, 4-hydroxy-propofol-1-OH-D-glucuronide
386	4.055	9.32e-06	9.11e-05	D-gluconic acid, creatinine, choline (?)
488	2.825	9.39e-06	9.11e-05	Unknown
483	2.875	1.02e-05	9.78e-05	Tranexamic acid
659	1.115	1.13e-05	1.07e-04	2-Oxoisovaleric acid, 4-hydroxy-propofol-1-OH-D-glucuronide
501	2.695	1.15e-05	1.07e-04	$MgEDTA^{2-}$

495	2.755	1.23e-05	1.13e-04	Unknown
472	2.985	1.24e-05	1.13e-04	Tranexamic acid, unknown
518	2.525	1.44e-05	1.29e-04	Citric acid, $CaEDTA^{2-}$
403	3.885	1.49e-05	1.31e-04	D-mannitol, paracetamol-glucuronide
575	1.955	1.49e-05	1.31e-04	Tranexamic acid, lysine
486	2.845	1.56e-05	1.35e-04	Unknown
209	7.415	1.64e-05	1.40e-04	Phenylalanine, heparin (?), unknown
395	3.965	1.71e-05	1.45e-04	Hippuric acid, isethionic acid (?), unknown
399	3.925	1.98e-05	1.65e-04	D-glucose, unknown
646	1.245	2.10e-05	1.73e-04	Unknown
363	4.285	2.39e-05	1.95e-04	Unknown, malic acid (?), pseudouridine
212	7.385	2.50e-05	2.01e-04	Unknown, phenylalanine, heparin (?)
471	2.995	2.52e-05	2.01e-04	Unknown, 2-oxoisovaleric acid
449	3.235	2.66e-05	2.10e-04	D-glucose
496	2.745	2.99e-05	2.32e-04	Unknown, $MgEDTA^{2-}$
499	2.715	3.00e-05	2.32e-04	$MgEDTA^{2-}$, unknown
220	7.305	3.30e-05	2.52e-04	Paracetamol-sulfate
406	3.855	3.47e-05	2.63e-04	D-glucose, D-mannitol
641	1.295	3.65e-05	2.73e-04	Unknown, L-isoleucine
517	2.535	3.97e-05	2.94e-04	$CaEDTA^{2-}$, citric acid
584	1.865	4.16e-05	3.05e-04	Tranexamic acid, lysine
396	3.955	4.42e-05	3.20e-04	Unknown
576	1.945	4.97e-05	3.57e-04	Tranexamic acid, lysine
470	3.005	5.12e-05	3.64e-04	Unknown, 2-oxoisovaleric acid
394	3.975	5.26e-05	3.70e-04	Unknown
397	3.945	5.42e-05	3.77e-04	Unknown
249	7.015	5.46e-05	3.77e-04	Unknown
391	4.005	5.51e-05	3.77e-04	2-Hydroxybutyric acid, unknown, phenylalanine
684	0.865	5.74e-05	3.88e-04	Unknown
211	7.395	5.88e-05	3.94e-04	Unknown, phenylalanine, heparin (?)
195	7.555	6.41e-05	4.26e-04	Unknown
560	2.105	6.65e-05	4.37e-04	Glutamine, glutamic acid, tranexamic acid, ketoleucine, 2-oxoisocaproic acid
409	3.825	6.70e-05	4.37e-04	D-glucose, D-gluconic acid, 4-hydroxy-propofol-4-OH-D-glucuronide
505	2.655	6.82e-05	4.41e-04	Unknown
464	3.065	7.25e-05	4.65e-04	Unknown
442	3.305	7.46e-05	4.74e-04	Phenylalanine, unknown
265	6.855	7.82e-05	4.93e-04	Unknown, 4-hydroxyphenylacetic acid, 4-hydroxyphenyllactic acid
603	1.675	9.43e-05	5.89e-04	Tranexamic acid
604	1.665	1.02e-04	6.31e-04	Tranexamic acid
636	1.345	1.04e-04	6.40e-04	Lactic acid, threonine, tranexamic acid
378	4.135	1.10e-04	6.66e-04	Lactic acid, D-gluconic acid
508	2.625	1.11e-04	6.70e-04	Ketoleucine, 2-oxoisocaproic acid

550	2.205	1.12e-04	6.73e-04	Unknown
605	1.655	1.16e-04	6.89e-04	Tranexamic acid
401	3.905	1.24e-04	7.27e-04	D-glucose, betaine
561	2.095	1.25e-04	7.27e-04	Tranexamic acid, ketoleucine, 2-oxoisocaproic acid
339	4.525	1.56e-04	9.06e-04	Unknown
248	7.025	1.59e-04	9.15e-04	Unknown
606	1.645	1.64e-04	9.33e-04	Tranexamic acid
210	7.405	1.72e-04	9.75e-04	Heparin (?), unknown
254	6.965	1.92e-04	1.07e-03	Unknown
253	6.975	1.96e-04	1.09e-03	Unknown
639	1.315	1.99e-04	1.09e-03	Lactic acid, threonine, unknown
279	6.715	1.99e-04	1.09e-03	4-hydroxy-propofol-1-OH-D-glucuronide
252	6.985	2.11e-04	1.15e-03	Unknown
668	1.025	2.24e-04	1.21e-03	Tranexamic acid, valine
372	4.195	2.38e-04	1.27e-03	L-pyroglutamic acid
577	1.935	2.49e-04	1.33e-03	Tranexamic acid
613	1.575	2.83e-04	1.50e-03	Unknown
247	7.035	2.87e-04	1.51e-03	Unknown
592	1.785	3.18e-04	1.65e-03	Unknown
393	3.985	3.43e-04	1.77e-03	Phenylalanine, 2-hydroxybutyric acid, serine, unknown
447	3.255	3.59e-04	1.83e-03	D-glucose, unknown, 4-hydroxy-propofol-4-OH-D-glucuronide
208	7.425	3.60e-04	1.83e-03	Phenylalanine, heparin (?), unknown, ephedrine (?)
413	3.745	3.67e-04	1.86e-03	D-glucose, D-mannitol
116	8.345	3.80e-04	1.91e-03	Unknown, S-5-adenosyl-L-homocysteine (?)
233	7.175	3.87e-04	1.93e-03	4-Hydroxyphenylacetic acid
196	7.545	3.89e-04	1.93e-03	Unknown, hippuric acid, tryptophan, heparin (?)
490	2.805	4.36e-04	2.14e-03	Unknown
567	2.035	4.39e-04	2.14e-03	Unknown, N-acetyl-L-glutamic acid (?), N-acetyl-L-glutamine (?), 2-hydroxyisovaleric acid
221	7.295	4.61e-04	2.24e-03	Tryptophan
653	1.175	4.91e-04	2.37e-03	Propofol-glucuronide, isopropanol[1]
381	4.105	4.95e-04	2.37e-03	Lactic acid
591	1.795	5.49e-04	2.61e-03	Unknown
152	7.985	6.06e-04	2.86e-03	Unknown
207	7.435	6.45e-04	3.03e-03	Phenylalanine, heparin (?), unknown, ephedrine (?)
284	6.665	6.67e-04	3.11e-03	Unknown
371	4.205	6.86e-04	3.18e-03	Unknown
441	3.315	7.14e-04	3.29e-03	Unknown
407	3.845	8.66e-04	3.96e-03	D-glucose, D-gluconic acid, 2-hydroxyisovaleric acid, 4-hydroxy-propofol-4-OH-D-glucuronide
476	2.945	9.43e-04	4.29e-03	N,N-dimethylglycine, unknown
416	3.715	9.63e-04	4.35e-03	D-glucose, threitol, propofol-glucuronide
262	6.885	1.02e-03	4.57e-03	Tyrosine
150	8.005	1.12e-03	5.01e-03	Unknown

511	2.595	1.13e-03	5.02e-03	$CaEDTA^{2-}$
392	3.995	1.16e-03	5.09e-03	2-Hydroxybutyric acid, unknown, phenylalanine
612	1.585	1.20e-03	5.26e-03	Unknown
415	3.725	1.23e-03	5.34e-03	D-glucose, leucine, threitol
443	3.295	1.24e-03	5.34e-03	Unknown, phenylalanine, myo-inositol, 4-hydroxy-propofol-4-OH-D-glucuronide
661	1.095	1.39e-03	5.98e-03	Tranexamic acid, unknown
485	2.855	1.43e-03	6.08e-03	Unknown
614	1.565	1.43e-03	6.08e-03	Unknown
164	7.865	1.45e-03	6.14e-03	Unknown
473	2.975	1.56e-03	6.56e-03	Unknown
379	4.125	1.60e-03	6.66e-03	Lactic acid, 3-hydroxybutyric acid
652	1.185	1.74e-03	7.24e-03	Propofol-glucuronide, 4-hydroxy-propofol-4-OH-D-glucuronide
193	7.575	1.76e-03	7.27e-03	Unknown
642	1.285	1.80e-03	7.39e-03	Unknown
491	2.795	1.83e-03	7.46e-03	Unknown, ephedrine (?)
197	7.535	1.86e-03	7.55e-03	Unknown, tryptophan, hippuric acid (?)
198	7.525	1.94e-03	7.81e-03	Unknown
380	4.115	1.95e-03	7.84e-03	Lactic acid
402	3.895	2.02e-03	8.04e-03	D-glucose
206	7.445	2.11e-03	8.35e-03	Phenylalanine, heparin (?), paracetamol-sulfate, ephedrine (?)
590	1.805	2.17e-03	8.56e-03	Unknown
421	3.515	2.22e-03	8.72e-03	D-glucose, propofol-glucuronide, 4-hydroxy-propofol-4-OH-D-glucuronide
408	3.835	2.24e-03	8.72e-03	D-glucose, D-gluconic acid, unknown, 4-hydroxy-propofol-4-OH-D-glucuronide
192	7.585	2.25e-03	8.72e-03	Unknown
65	8.855	2.31e-03	8.90e-03	Unknown
671	0.995	2.42e-03	9.30e-03	Valine
334	4.575	2.52e-03	9.62e-03	4-hydroxy-propofol-4-OH-D-glucuronide, glutathione, carnitine
565	2.055	2.56e-03	9.69e-03	Unknown, 2-oxoisocaproic acid, glutamic acid
672	0.985	2.56e-03	9.69e-03	Valine
628	1.425	2.60e-03	9.76e-03	Lysine, unknown
435	3.375	2.74e-03	1.03e-02	Unknown
189	7.615	2.76e-03	1.03e-02	Unknown
559	2.115	2.90e-03	1.07e-02	Glutamine, glutamic acid, tranexamic acid, ketoleucine, 2-oxoisocaproic acid
478	2.925	3.00e-03	1.10e-02	Unknown
370	4.215	3.07e-03	1.13e-02	Sucrose (?)
637	1.335	3.26e-03	1.19e-02	Lactic acid, tranexamic acid
640	1.305	3.55e-03	1.29e-02	Unknown, lactic acid
638	1.325	3.58e-03	1.29e-02	Lactic acid

153	7.975	3.59e-03	1.29e-02	Unknown
251	6.995	3.63e-03	1.30e-02	Unknown
564	2.065	3.65e-03	1.30e-02	Glutamic acid, ketoleucine, 2-oxoisocaproic acid
446	3.265	3.74e-03	1.32e-02	D-glucose, betaine, 4-hydroxy-propofol-4-OH-D-glucuronide
608	1.625	4.39e-03	1.55e-02	Tranexamic acid
267	6.835	4.54e-03	1.59e-02	Unknown
367	4.245	4.94e-03	1.72e-02	Threonine
519	2.515	5.20e-03	1.80e-02	Unknown, glutamine
563	2.075	5.21e-03	1.80e-02	Glutamic acid, ketoleucine, 2-oxoisocaproic acid
683	0.875	5.29e-03	1.82e-02	Unknown, propofol-glucuronide
509	2.615	5.41e-03	1.85e-02	Ketoleucine, 2-oxoisocaproic acid
179	7.715	5.60e-03	1.91e-02	Unknown
667	1.035	5.90e-03	2.00e-02	Valine, tranexamic acid
616	1.545	6.08e-03	2.05e-02	Unknown
558	2.125	6.34e-03	2.12e-02	Glutamine, glutamic acid, tranexamic acid, ketoleucine, 2-oxoisocaproic acid
424	3.485	6.36e-03	2.12e-02	D-glucose
353	4.385	6.37e-03	2.12e-02	Unknown
281	6.695	6.66e-03	2.20e-02	Unknown
615	1.555	6.69e-03	2.20e-02	Unknown
657	1.135	6.71e-03	2.20e-02	2-Oxoisovaleric acid, 4-hydroxy-propofol-1-OH-D-glucuronide
202	7.485	7.19e-03	2.35e-02	Unknown
474	2.965	7.27e-03	2.36e-02	Unknown, 3-aminopropionitrilefumarate (?)
203	7.475	7.32e-03	2.37e-02	Unknown, paracetamol-sulfate
184	7.665	7.41e-03	2.39e-02	Unknown
219	7.315	7.46e-03	2.39e-02	Paracetamol-sulfate
400	3.915	7.53e-03	2.40e-02	D-glucose
414	3.735	7.56e-03	2.40e-02	D-glucose, leucine
521	2.495	7.74e-03	2.45e-02	Glutamine, unknown
194	7.565	8.08e-03	2.54e-02	Unknown
626	1.445	8.11e-03	2.54e-02	Unknown, lysine
269	6.815	8.15e-03	2.54e-02	Unknown
191	7.595	8.30e-03	2.58e-02	Unknown
298	6.525	8.35e-03	2.58e-02	Fumaric acid
263	6.875	8.41e-03	2.59e-02	4-Hydroxyphenylacetic acid, 4-hydroxyphenyllactic acid
229	7.215	8.47e-03	2.60e-02	Unknown, tyrosine
686	0.845	8.80e-03	2.69e-02	2-Hydroxyisovaleric acid, unknown
469	3.015	9.11e-03	2.77e-02	Lysine, 2-oxoisovaleric acid
365	4.265	9.14e-03	2.77e-02	Threonine
589	1.815	9.57e-03	2.89e-02	Tranexamic acid
536	2.345	1.05e-02	3.15e-02	Glutamic acid
484	2.865	1.08e-02	3.24e-02	Tranexamic acid, unknown
285	6.655	1.18e-02	3.52e-02	Unknown

276	6.745	1.19e-02	3.52e-02	Unknown
142	8.085	1.23e-02	3.64e-02	Unknown
411	3.765	1.25e-02	3.68e-02	Glutamic acid, D-glucose, lysine, glutamine, D-mannitol
627	1.435	1.28e-02	3.76e-02	Lysine, unknown
623	1.475	1.29e-02	3.76e-02	Alanine, lysine
398	3.935	1.34e-02	3.90e-02	Creatine
242	7.085	1.41e-02	4.07e-02	Unknown
141	8.095	1.42e-02	4.08e-02	Unknown
448	3.245	1.54e-02	4.41e-02	D-glucose
687	0.835	1.56e-02	4.44e-02	2-Hydroxyisovaleric acid, unknown
429	3.435	1.56e-02	4.44e-02	D-glucose
431	3.415	1.60e-02	4.53e-02	D-glucose
148	8.025	1.60e-02	4.53e-02	Unknown
185	7.655	1.61e-02	4.53e-02	Unknown
218	7.325	1.64e-02	4.59e-02	Phenylalanine, paracetamol-sulfate
190	7.605	1.66e-02	4.63e-02	Unknown
174	7.765	1.71e-02	4.76e-02	Unknown
341	4.505	1.75e-02	4.86e-02	Unknown
534	2.365	1.78e-02	4.92e-02	Glutamic acid, proline, 3-hydroxyglutaric acid (?), malic acid (?)
275	6.755	1.81e-02	4.98e-02	Unknown

Table 7.4: **Spectral positions and P-values of plasma features that discriminated AKI from non-AKI patients.** The 85 plasma specimens studied were collected 24 hours post-operatively. The first 24 ± 2.8 features were used by the RF classifier. A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case more than one compound contributed to one significant bin, all possibly corresponding molecules are annotated. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. [1]Probably contamination. Adapted from [Zacharias et al. 2015].

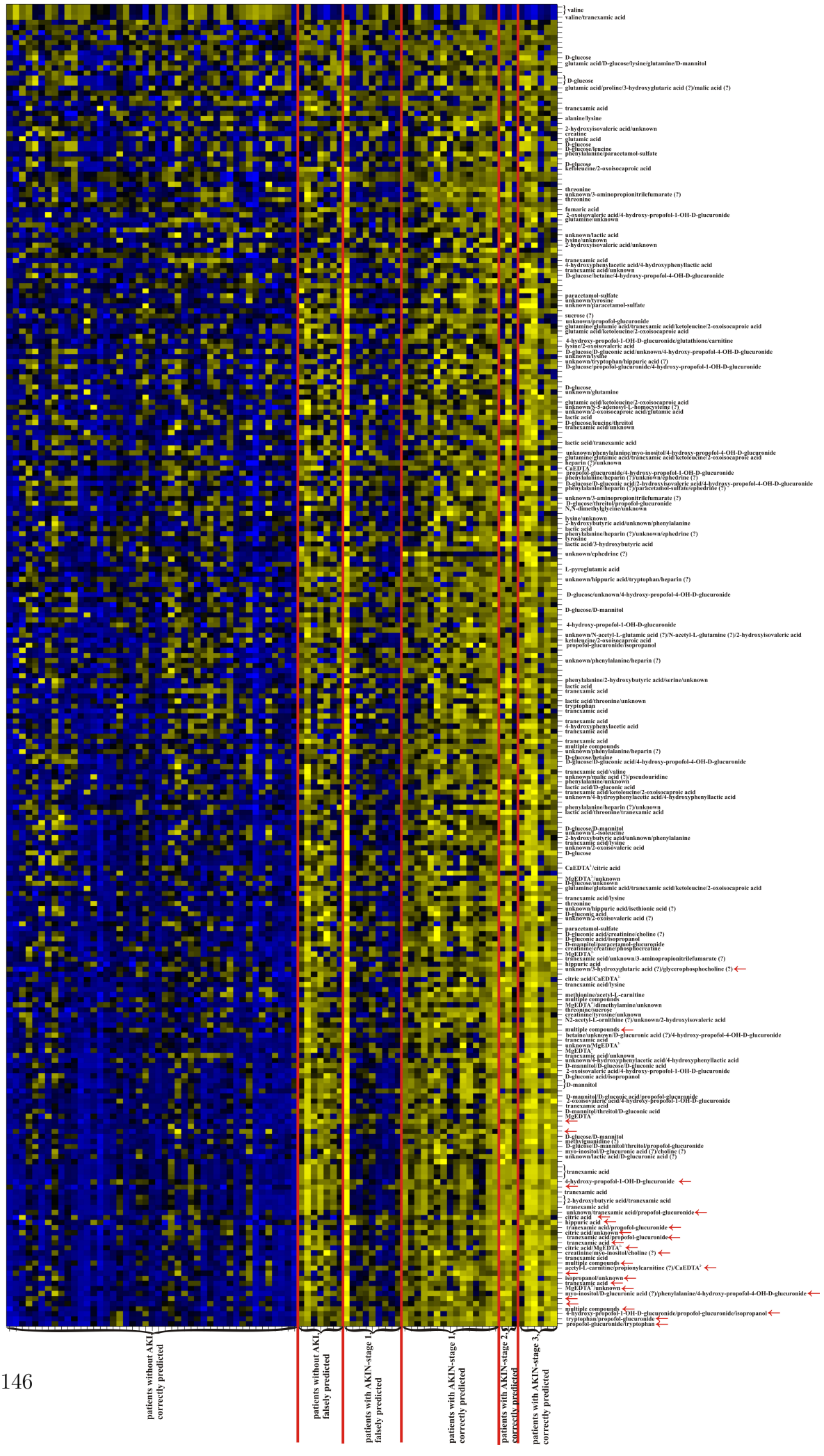

valine
valine/tranexamic acid
D-glucose
glutamic acid/D-glucose/lysine/glutamine/D-mannitol
D-glucose
glutamic acid/proline/3-hydroxyglutaric acid (?)/malic acid (?)
tranexamic acid
alanine/lysine
2-hydroxyisovaleric acid/unknown
creatine
glutamic acid
D-glucose
D-glucose/leucine
phenylalanine/paracetamol-sulfate
D-glucose
ketoleucine/2-oxoisocaproic acid
threonine
unknown/3-aminopropionitrilefumarate (?)
threonine
fumaric acid
2-oxoisovaleric acid/4-hydroxy-propofol-1-OH-D-glucuronide
glutamine/unknown
unknown/lactic acid
lysine/unknown
2-hydroxyisovaleric acid/unknown
tranexamic acid
4-hydroxyphenylacetic acid/4-hydroxyphenyllactic acid
tranexamic acid/unknown
D-glucose/betaine/4-hydroxy-propofol-4-OH-D-glucuronide
paracetamol-sulfate
unknown/tyrosine
unknown/paracetamol-sulfate
sucrose (?)
unknown/propofol-glucuronide
glutamine/glutamic acid/tranexamic acid/ketoleucine/2-oxoisocaproic acid
glutamic acid/ketoleucine/2-oxoisocaproic acid
4-hydroxy-propofol-1-OH-D-glucuronide/glutathione/carnitine
lysine/2-oxoisovaleric acid
D-glucose/D-gluconic acid/unknown/4-hydroxy-propofol-4-OH-D-glucuronide
unknown/lysine
unknown/tryptophan/hippuric acid (?)
D-glucose/propofol-glucuronide/4-hydroxy-propofol-1-OH-D-glucuronide
D-glucose
unknown/glutamine
glutamic acid/ketoleucine/2-oxoisocaproic acid
unknown/N-5-adenosyl-L-homocysteine (?)
unknown/2-oxoisocaproic acid/glutamic acid
lactic acid
D-glucose/leucine/threitol
tranexamic acid/unknown
lactic acid/tranexamic acid
unknown/phenylalanine/myo-inositol/4-hydroxy-propofol-4-OH-D-glucuronide
glutamine/glutamic acid/tranexamic acid/ketoleucine/2-oxoisocaproic acid
heparin (?)/unknown
CaEDTA2-
propofol-glucuronide/4-hydroxy-propofol-1-OH-D-glucuronide
phenylalanine/heparin (?)/unknown/ephedrine (?)
D-glucose/D-gluconic acid/2-hydroxyisovaleric acid/4-hydroxy-propofol-4-OH-D-glucuronide
phenylalanine/heparin (?)/paracetamol-sulfate/ephedrine (?)
unknown/3-aminopropionitrilefumarate (?)
D-glucose/threitol/propofol-glucuronide
N,N-dimethylglycine/unknown
lysine/unknown
2-hydroxybutyric acid/unknown/phenylalanine
lactic acid
phenylalanine/heparin (?)/unknown/ephedrine (?)
tyrosine
lactic acid/3-hydroxybutyric acid
unknown/ephedrine (?)
L-pyroglutamic acid
unknown/hippuric acid/tryptophan/heparin (?)
D-glucose/unknown/4-hydroxy-propofol-4-OH-D-glucuronide
D-glucose/D-mannitol
4-hydroxy-propofol-1-OH-D-glucuronide
unknown/N-acetyl-L-glutamic acid (?)/N-acetyl-L-glutamine (?)/2-hydroxyisovaleric acid
ketoleucine/2-oxoisocaproic acid
propofol-glucuronide/isopropanol
unknown/phenylalanine/heparin (?)
phenylalanine/2-hydroxybutyric acid/serine/unknown
lactic acid
tranexamic acid
lactic acid/threonine/unknown
tryptophan
tranexamic acid
tranexamic acid
4-hydroxyphenylacetic acid
tranexamic acid
tranexamic acid
multiple compounds
unknown/phenylalanine/heparin (?)
D-glucose/betaine
D-glucose/D-gluconic acid/4-hydroxy-propofol-4-OH-D-glucuronide
tranexamic acid/valine
unknown/malic acid (?)/pseudouridine
phenylalanine/unknown
lactic acid/D-gluconic acid
tranexamic acid/ketoleucine/2-oxoisocaproic acid
unknown/4-hydroxyphenylacetic acid/4-hydroxyphenyllactic acid
phenylalanine/heparin (?)/unknown
lactic acid/threonine/tranexamic acid
D-glucose/D-mannitol
unknown/L-isoleucine
2-hydroxybutyric acid/unknown/phenylalanine
tranexamic acid/lysine
unknown/2-oxoisovaleric acid
D-glucose
CaEDTA2-/citric acid
MgEDTA2-/unknown
D-glucose/unknown
glutamine/glutamic acid/tranexamic acid/ketoleucine/2-oxoisocaproic acid
tranexamic acid/lysine
threonine
unknown/hippuric acid/isethionic acid (?)
D-gluconic acid
unknown/2-oxoisovaleric acid (?)
paracetamol-sulfate
D-gluconic acid/creatinine/choline (?)
D-gluconic acid/isopropanol
D-mannitol/paracetamol-glucuronide
creatinine/creatine/phosphocreatine
MgEDTA2-
tranexamic acid/unknown/3-aminopropionitrilefumarate (?)
hippuric acid
unknown/3-hydroxyglutaric acid (?)/glycerophosphocholine (?)
citric acid/CaEDTA2-
tranexamic acid/lysine
methionine/acetyl-L-carnitine
multiple compounds
MgEDTA2-/dimethylamine/unknown
threonine/sucrose
creatinine/tyrosine/unknown
N2-acetyl-L-ornithine (?)/unknown/2-hydroxyisovaleric acid
multiple compounds
betaine/unknown/D-glucuronic acid (?)/4-hydroxy-propofol-4-OH-D-glucuronide
tranexamic acid
unknown/MgEDTA2-
MgEDTA2-
tranexamic acid/unknown
unknown/4-hydroxyphenylacetic acid/4-hydroxyphenyllactic acid
D-mannitol/D-glucose/D-gluconic acid
2-oxoisovaleric acid/4-hydroxy-propofol-1-OH-D-glucuronide
D-gluconic acid/isopropanol
D-mannitol
D-mannitol/D-gluconic acid/propofol-glucuronide
2-oxoisovaleric acid/4-hydroxy-propofol-1-OH-D-glucuronide
tranexamic acid
D-mannitol/threitol/D-gluconic acid
MgEDTA2-
D-glucose/D-mannitol
methylguanidine (?)
D-glucose/D-mannitol/threitol/propofol-glucuronide
myo-inositol/D-glucuronic acid (?)/choline (?)
unknown/lactic acid/D-glucaronic acid (?)
tranexamic acid
4-hydroxy-propofol-1-OH-D-glucuronide
tranexamic acid
2-hydroxybutyric acid/tranexamic acid
tranexamic acid
unknown/tranexamic acid/propofol-glucuronide
citric acid
hippuric acid
tranexamic acid/propofol-glucuronide
citric acid/unknown
tranexamic acid/propofol-glucuronide
tranexamic acid
citric acid/MgEDTA2-
creatinine/myo-inositol/choline (?)
tranexamic acid
multiple compounds
acetyl-L-carnitine/propionylcarnitine (?)/CaEDTA2-
isopropanol/unknown
tranexamic acid
MgEDTA2-/unknown
myo-inositol/D-glucuronic acid (?)/phenylalanine/4-hydroxy-propofol-4-OH-D-glucuronide
multiple compounds
4-hydroxy-propofol-1-OH-D-glucuronide/propofol-glucuronide/isopropanol
tryptophan/propofol-glucuronide
propofol-glucuronide/tryptophan
patients without AKI correctly predicted
patients without AKI falsely predicted
patients with AKIN-stage 1 falsely predicted
patients with AKIN-stage 1 correctly predicted
patients with AKIN-stage 2,3 correctly predicted
patients with AKIN-stage 3, correctly predicted

Figure 7.3: **Previous page: Heat-map representation of plasma NMR features, whose signal intensities differed significantly between AKI and non-AKI patients at 24 hours after surgery.** These features are also listed in Appendix II section 7.2.6 Table 7.4. Each row corresponds to a significant feature (as assigned on the right), each column corresponds to a single patient. Up-regulated features are indicated in yellow, while down-regulated ones are marked in blue. Rows are ordered according to increasing correlation coefficients between disease status and feature intensities. Rows that are mostly up-regulated in AKI patients and down-regulated in all other groups are shown at the bottom part of the heat-map and vice versa. The 52 patients without AKI are divided into two groups, namely correctly predicted patients ($n = 45$) and patients who were falsely classified as "AKI patients" ($n = 7$). The 33 AKI patients are listed on the right side of the heat-map; they are divided into four groups, namely falsely predicted AKIN 1 patients ($n = 9$), correctly predicted AKIN 1 patients ($n = 15$), correctly predicted AKIN 2 patients ($n = 3$) and correctly predicted AKIN 3 patients ($n = 6$). The different groups are separated from each other by red vertical bars. The 27 features that were mostly selected for Random Forest prognostication are indicated by red arrows on the right. Displayed results represent prognostication outcomes obtained in the majority of multiple RF runs. Reprinted with permission from [Zacharias et al. 2015]. Copyright 2015 American Chemical Society.

7.3 Appendix III: German Chronic Kidney Disease Study

7.3.1 Patient characteristics

Baseline patient characteristics	Reported values	Number of missing values
Number of patients	3164	-
Age [years]	60.2 ± 11.8 (18 - 75)	1
Sex, male	1945 (61.5%)	-
Smoking		12
Number of non-smokers	1265 (40.1%)	
Number of current smokers	507 (16.1%)	
Number of former smokers	1380 (43.8%)	
Waist-hip-ratio	0.94 ± 0.09 (0.64 - 1.3)	94
BMI [kg/m^2]	29.7 ± 5.8 (16.9 - 61.5)	32
Diabetes mellitus type 1	63 (2.0%)	1
Diabetes mellitus type 2	783 (24.8%)	1
Synlab baseline parameters		
SCr [mg/dl][a]	1.51 ± 0.48 (0.45 - 4.73)	17
SCysC [mg/l][b]	1.51 ± 0.48 (0.50 - 4.95)	15
Baseline eGFR values[c] **[ml/min per 1.73m^2]**		
$eGFR_{mdrd4}$	47.28 ± 16.48 (11.90 - 145.87)	18
$eGFR_{ckdepi\ crea}$	49.70 ± 18.11 (11.52 - 136.56)	18
$eGFR_{ckdepi\ cys}$	49.93 ± 19.74 (8.83 - 154.59)	16
$eGFR_{ckdepi\ crea\ cys}$	48.93 ± 18.16 (11.52 - 147.68)	18
Baseline proteinuria categories[d]		51
< 30 mg/l	1488 (47.8%)	
(30 - 300) mg/l	896 (28.8%)	
> 300 mg/l	729 (23.4%)	
Leading renal disease		1
Diabetic nephropathy	471 (14.9%)	
Glomerulonephritis	587 (18.6%)	
Hereditary disease	126 (4.0%)	
Interstitial nephropathy	138 (4.4%)	
No leading renal disease	650 (20.6%)	
Other leading renal diseases	208 (6.6%)	
Systemic disease	217 (6.9%)	
Hypertensive nephropathy	766 (24.2%)	

Table 7.5: **Previous page: Baseline patient characteristics of the GCKD study baseline sample cohort.** Data is expressed as number (percentage) for categorical, and mean ± standard deviation (range) for continuous variables, respectively. The number of missing values is reported in the last column. [a]In healthy subjects, SCr values range between 0.84-1.25mg/dl in white men and 0.66-1.09mg/dl in white women [Dörner 2013]. [b]In healthy adults, SCysC values range between 0.54-0.94mg/l in men and 0.48-0.82mg/l in women [Dörner 2013]. [c]Reference eGFR values in young healthy whites are about 130ml/min per 1.73m^2 for men and 120 ml/min per 1.73m^2 for women [Stevens and Levey 2009, Stevens et al. 2006]. [d]Healthy individuals usually excrete less than 150mg of protein per day into the urine [Dörner 2013, Arasteh et al. 2009]. Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; $eGFR_{ckdepi\ crea}$, eGFR based on CKD-EPI crea formula; $eGFR_{ckdepi\ crea\ cys}$, eGFR based on CKD-EPI crea cys formula; $eGFR_{ckdepi\ cys}$, eGFR based on CKD-EPI cys formula; $eGFR_{mdrd4}$, eGFR based on MDRD4 formula; SCr, serum creatinine; SCysC, serum cystatin C.

Baseline patient characteristics	**Reported values**	**Number of missing values**
Number of patients	2697	-
Age [years]	60.2 ± 11.7 (18 - 75)	1
Sex, male	1634 (60.6%)	-
Smoking		9
Number of non-smokers	1109 (41.3%)	
Number of current smokers	409 (15.2%)	
Number of former smokers	1170 (43.5%)	
Waist-hip-ratio	0.94 ± 0.09 (0.64 - 1.3)	71
BMI [kg/m^2]	29.7 ± 5.8 (16.9 - 61.5)	26
Diabetes mellitus type 1	50 (1.9%)	1
Diabetes mellitus type 2	642 (23.8%)	1
Synlab baseline parameters		
SCr [mg/dl]	1.50 ± 0.46 (0.45 - 4.72)	14
SCys C [mg/l]	1.50 ± 0.46 (0.53 - 4.95)	12
Baseline eGFR values [ml/min per 1.73m^2]		
$eGFR_{mdrd4}$	47.36 ± 16.00 (11.91 - 145.87)	15
$eGFR_{ckdepi\ crea}$	49.82 ± 17.62 (11.96 - 131.21)	15
$eGFR_{ckdepi\ cys}$	50.37 ± 19.26 (8.83 - 141.59)	13
$eGFR_{ckdepi\ crea\ cys}$	49.20 ± 17.60 (11.52 - 128.14)	15
Baseline proteinuria categories		41
< 30 mg/l	1281 (48.2%)	
(30 - 300) mg/l	769 (29.0%)	
> 300 mg/l	606 (22.8%)	
Leading renal disease		1
Diabetic nephropathy	386 (14.3%)	

Glomerulonephritis	510 (18.9%)	
Hereditary disease	111 (4.1%)	
Interstitial nephropathy	121 (4.5%)	
No leading renal disease	562 (20.8%)	
Other leading renal diseases	174 (6.5%)	
Systemic disease	194 (7.2%)	
Hypertensive nephropathy	639 (23.7%)	
FU2 patient characteristics	**Reported values**	**Number of missing values**
Number of patients	2697	-
Synlab FU2 parameters		
SCr [mg/dl]	1.68 ± 0.75 (0.45 - 10.54)	193
SCys C [mg/l]	1.72 ± 0.65 (0.39 - 7.00)	191
FU2 eGFR values [ml/min per 1.73m^2]		
$eGFR_{mdrd4}$	43.49 ± 16.57 (5.22 - 142.13)	193
$eGFR_{ckdepi\ crea}$	45.67 ± 18.35 (5.07 - 127.50)	193
$eGFR_{ckdepi\ cys}$	43.96 ± 19.79 (5.92 - 157.85)	191
$eGFR_{ckdepi\ crea\ cys}$	43.82 ± 18.40 (5.79 - 130.92)	193

Table 7.6: **Baseline and FU2 patient characteristics of the GCKD study FU2 sample cohort.** Data is expressed as number (percentage) for categorical, and mean ± standard deviation (range) for continuous variables, respectively. The number of missing values is reported in the last column. Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; $eGFR_{ckdepi\ crea}$, eGFR based on CKD-EPI crea formula; $eGFR_{ckdepi\ crea\ cys}$, eGFR based on CKD-EPI crea cys formula; $eGFR_{ckdepi\ cys}$, eGFR based on CKD-EPI cys formula; $eGFR_{mdrd4}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.

Baseline patient characteristics	**Complete patient set**	**Training set**	**Test set**	***p*-value**
Number of patients	2492	1661	831	-
Age [years]	60.3 ± 11.6 (18 - 75)	60.1 ± 11.7 (18 - 75)	60.7 ± 11.4 (20 - 75)	0.26[a]
Sex, male	1508 (60.5%)	997 (60.0%)	511 (61.5%)	0.49[b]
Smoking				
Number of non-smokers	1037 (41.7%)	701 (42.3%)	336 (40.6%)	0.44[b]
Number of current smokers	374 (15.0%)	238 (14.4%)	136 (16.4%)	0.19[b]
Number of former smokers	1075 (43.2%)	719 (43.4%)	356 (43.0%)	0.86[b]
Waist-hip-ratio	0.94 ± 0.09 (0.64 - 1.3)	0.94 ± 0.09 (0.67 - 1.22)	0.94 ± 0.09 (0.64 - 1.3)	1[a]
BMI [kg/m^2]	29.7 ± 5.8 (16.9 - 61.5)	29.7 ± 5.7 (17.1 - 61.5)	29.7 ± 5.8 (16.9 - 56.6)	0.93[a]
Diabetes mellitus type 1	47 (1.9%)	29 (1.7%)	18 (2.2%)	0.53[b]
Diabetes mellitus type 2	591 (23.7%)	403 (24.3%)	188 (22.6%)	0.37[b]
Synlab baseline parameters				
SCr [mg/dl]	1.49 ± 0.46 (0.45 - 4.72)	1.50 ± 0.47 (0.45 - 4.72)	1.49 ± 0.44 (0.51 - 3.85)	0.66[a]
SCysC [mg/l]	1.49 ± 0.46 (0.53 - 4.95)	1.49 ± 0.46 (0.53 - 4.95)	1.49 ± 0.45 (0.57 - 3.77)	0.82[a]
Baseline eGFR values [ml/min per 1.73m^2]				
$eGFR_{mdrd4}$	47.5 ± 16.1 (11.9 - 145.9)	47.5 ± 16.3 (11.9 - 145.9)	47.5 ± 15.8 (16.5 - 124.8)	0.95[a]
$eGFR_{ckdepi\ crea}$	49.9 ± 17.7 (12.0 - 131.2)	50.0 ± 17.9 (12.0 - 131.21)	49.9 ± 17.3 (16.8 - 119.5)	0.93[a]
$eGFR_{ckdepi\ cys}$	50.5 ± 19.3 (8.8 - 141.6)	50.6 ± 19.3 (8.8 - 141.6)	50.2 ± 19.2 (12.6 - 120.0)	0.64[a]

$eGFR_{ckdepi\ crea\ cys}$	49.3 ± 17.7 (14.5 - 128.1)	49.4 ± 17.9 (14.5 - 128.1)	49.2 ± 17.4 (14.8 - 116.1)	0.75[a]
Baseline proteinuria categories				
< 30 mg/l	1192 (48.4%)	787 (48.0%)	405 (49.5%)	0.52[b]
(30 - 300) mg/l	717 (29.2%)	478 (29.2%)	239 (29.2%)	1[b]
>300 mg/l	550 (22.4%)	375 (22.9%)	175 (21.4%)	0.41[b]
Leading renal disease				
Diabetic nephropathy	352 (14.1%)	224 (13.5%)	128 (15.4%)	0.20[b]
Glomerulonephritis	478 (19.2%)	326 (19.6%)	152 (18.3%)	0.45[b]
Hereditary disease	106 (4.3%)	71 (4.3%)	35 (4.2%)	1[b]
Interstitial nephropathy	115 (4.6%)	76 (4.6%)	39 (4.7%)	0.92[b]
No leading renal disease	517 (20.8%)	348 (21.0%)	169 (20.3%)	0.75[b]
Other leading renal diseases	162 (6.5%)	113 (6.8%)	49 (5.9%)	0.44[b]
Systemic disease	178 (7.1%)	118 (7.1%)	60 (7.2%)	0.93[b]
Hypertensive nephropathy	584 (23.4%)	385 (23.2%)	199 (24.0%)	0.69[b]
FU2 patient characteristics	**Complete patient set**	**Training set**	**Test set**	***p*-value**
Number of patients	2492	1661	831	-
Synlab FU2 parameters				
SCr [mg/dl]	1.68 ± 0.75 (0.45 - 10.54)	1.69 ± 0.80 (0.45 - 10.54)	1.66 ± 0.63 (0.5 - 6.54)	0.31[a]
SCys C [mg/l]	1.72 ± 0.65 (0.57 - 7.00)	1.72 ± 0.67 (0.57 - 7.00)	1.72 ± 0.62 (0.63 - 6.26)	0.86[a]
FU2 eGFR values [ml/min per 1.73m^2]				
$eGFR_{mdrd4}$	43.5 ± 16.6 (5.2 - 142.1)	43.6 ± 17.0 (5.2 - 142.1)	43.2 ± 15.5 (8.4 - 141.2)	0.49[a]
$eGFR_{ckdepi\ crea}$	45.7 ± 18.3 (5.1 - 127.5)	45.8 ± 18.8 (5.1 - 127.5)	45.3 ± 17.2 (7.6 - 126.7)	0.44[a]
$eGFR_{ckdepi\ cys}$	43.9 ± 19.7 (5.9 - 143.3)	44.2 ± 19.9 (5.9 - 143.3)	43.3 ± 19.1 (7.0 - 121.6)	0.24[a]
$eGFR_{ckdepi\ crea\ cys}$	43.8 ± 18.4 (5.8 - 130.9)	44.1 ± 18.8 (5.8 - 130.9)	43.3 ± 17.5 (7.4 - 128.5)	0.30[a]

Table 7.7: **Baseline and FU2 patient characteristics of the regression sample subset cohort taken from the original GCKD study FU2 sample cohort.** Given are the baseline patient characteristics for the complete regression cohort, as well as separately for training and test set. Data is expressed as number (percentage) for categorical, and mean ± standard deviation (range) for continuous variables, respectively. The last column gives the corresponding *p*-values for [a]Welch-test or [b]Fisher's exact tests between training and test set calculated with *R*. Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; $eGFR_{ckdepi\ crea}$, eGFR based on CKD-EPI crea formula; $eGFR_{ckdepi\ crea\ cys}$, eGFR based on CKD-EPI crea cys formula; $eGFR_{ckdepi\ cys}$, eGFR based on CKD-EPI cys formula; $eGFR_{mdrd4}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.

7.3.2 *t*-tests between various leading renal diseases

Spectral position [ppm]	ID	log(FC)	*P*-value un-adjusted	*P*-value B/H-adjusted	Statistical power	Identified compounds
3.782	359	0.410	3.43e-108	2.26e-105	100.00	D-glucose, alanine, glutamine, arginine
3.802	357	0.451	2.47e-107	8.14e-105	100.00	D-glucose, alanine
3.842	353	0.509	4.88e-105	1.07e-102	100.00	D-glucose, unknown
3.852	352	0.529	3.20e-103	3.74e-101	100.00	D-glucose, unknown
3.562	368	0.527	3.25e-103	3.74e-101	100.00	D-glucose
3.912	346	0.501	3.63e-103	3.74e-101	100.00	D-glucose, betaine, unknown
3.552	369	0.544	3.97e-103	3.74e-101	100.00	D-glucose, myo-inositol

3.932	344	0.512	4.94e-102	4.07e-100	100.00	D-glucose
3.872	350	0.440	6.67e-101	4.89e-99	100.00	D-glucose, unknown
3.752	362	0.535	9.80e-101	6.47e-99	100.00	D-glucose, glutamic acid
3.572	367	0.492	3.89e-99	2.34e-97	100.00	D-glucose, glycine
3.742	363	0.535	7.08e-99	3.89e-97	100.00	D-glucose, leucine
3.412	383	0.565	1.00e-98	5.08e-97	100.00	D-glucose, carnitine, taurine, proline
3.482	376	0.567	5.91e-98	2.79e-96	100.00	D-glucose
3.862	351	0.532	1.15e-97	5.04e-96	100.00	D-glucose, unknown
3.472	377	0.555	1.22e-97	5.04e-96	100.00	D-glucose
3.442	380	0.554	1.81e-97	7.01e-96	100.00	D-glucose, carnitine, taurine, proline
3.512	373	0.578	2.50e-97	9.18e-96	100.00	D-glucose
3.722	365	0.452	5.51e-97	1.91e-95	100.00	D-glucose, N,N-dimethylglycine
3.432	381	0.560	1.30e-96	4.29e-95	100.00	D-glucose, carnitine, taurine, proline
3.502	374	0.574	1.46e-96	4.57e-95	100.00	D-glucose
3.452	379	0.491	1.99e-95	5.98e-94	100.00	D-glucose, carnitine, proline
3.882	349	0.478	3.11e-95	8.91e-94	100.00	D-glucose, unknown
3.422	382	0.569	1.35e-94	3.72e-93	100.00	D-glucose, carnitine, taurine, proline
3.762	361	0.485	2.83e-94	7.47e-93	100.00	D-glucose, arginine, glutamine, glutamic acid
3.732	364	0.489	1.04e-91	2.63e-90	100.00	D-glucose, unknown
3.792	358	0.288	1.32e-91	3.24e-90	100.00	D-glucose, alanine
3.532	371	0.496	3.18e-91	7.48e-90	100.00	D-glucose
3.492	375	0.550	4.52e-90	1.03e-88	100.00	D-glucose
3.772	360	0.429	1.98e-85	4.35e-84	100.00	D-glucose, alanine, glutamine, arginine
3.812	356	0.369	2.76e-75	5.88e-74	100.00	D-glucose
3.542	370	0.412	8.07e-66	1.66e-64	100.00	D-glucose, myo-inositol
3.942	343	0.439	6.80e-63	1.36e-61	100.00	D-glucose
3.522	372	0.480	2.33e-58	4.52e-57	100.00	D-glucose
3.832	354	0.241	1.73e-57	3.27e-56	100.00	Unknown
3.402	384	0.335	1.70e-52	3.12e-51	100.00	Unknown
3.922	345	0.268	1.27e-48	2.26e-47	100.00	D-glucose, unknown
3.822	355	0.232	7.20e-42	1.25e-40	100.00	Unknown
3.962	341	0.135	3.71e-35	6.28e-34	100.00	Unknown
3.892	348	0.134	2.17e-26	3.59e-25	100.00	Unknown
3.462	378	0.217	7.03e-26	1.13e-24	100.00	D-glucose
3.972	340	0.104	4.96e-21	7.80e-20	100.00	Unknown
3.372	387	-0.167	8.72e-18	1.34e-16	100.00	Methanol, proline
7.432	205	0.136	1.60e-14	2.40e-13	100.00	Phenylalanine
4.122	325	0.204	1.84e-14	2.69e-13	100.00	Proline, lactic acid
4.152	322	0.171	2.59e-14	3.72e-13	100.00	Proline, lactic acid

4.032	334	0.130	5.67e-14	7.96e-13	100.00	Unknown
1.072	606	0.136	1.53e-13	2.11e-12	100.00	Valine
1.082	605	0.142	1.97e-13	2.65e-12	100.00	Unknown
4.142	323	0.202	9.21e-13	1.22e-11	100.00	Proline, lactic acid
3.712	366	0.116	8.66e-12	1.12e-10	100.00	Unknown
1.902	523	0.078	9.28e-12	1.18e-10	100.00	Overlap of multiple minor compounds
4.052	332	0.192	1.53e-11	1.91e-10	100.00	Unknown
1.002	613	0.125	1.83e-11	2.24e-10	100.00	Valine, lipid methyl, cholesterol (ester)
4.132	324	0.203	1.94e-11	2.33e-10	100.00	Proline, lactic acid
1.492	564	0.127	2.92e-11	3.44e-10	100.00	Alanine
1.102	603	0.133	3.58e-11	4.14e-10	100.00	Unknown
1.502	563	0.130	6.51e-11	7.41e-10	100.00	Alanine
2.382	475	0.120	6.72e-11	7.52e-10	100.00	Proline, glutamic acid
1.012	612	0.135	1.43e-10	1.55e-09	100.00	Valine, lipid methyl, cholesterol (ester)
1.892	524	0.075	1.43e-10	1.55e-09	100.00	Overlap of multiple minor compounds
4.072	330	0.164	1.45e-10	1.55e-09	99.99	Creatinine
2.392	474	0.126	1.68e-10	1.76e-09	100.00	Unknown
2.372	476	0.116	2.97e-10	3.06e-09	100.00	Proline, glutamic acid
2.552	458	0.089	3.43e-10	3.49e-09	100.00	Citric acid
1.092	604	0.126	4.48e-10	4.48e-09	100.00	Unknown
3.062	407	0.105	5.03e-10	4.96e-09	99.99	Creatinine
4.002	337	-0.071	6.02e-10	5.84e-09	100.00	Unknown
7.202	228	0.123	6.35e-10	6.08e-09	100.00	Tyrosine
1.052	608	0.181	8.22e-10	7.75e-09	100.00	Valine
7.372	211	0.091	1.19e-09	1.11e-08	99.99	Phenylalanine
3.992	338	-0.080	1.26e-09	1.15e-08	100.00	Unknown
2.362	477	0.104	1.44e-09	1.30e-08	100.00	Proline, glutamic acid
3.032	410	0.041	2.87e-09	2.56e-08	99.99	Lysine, unknown
7.382	210	0.080	3.84e-09	3.38e-08	99.99	Phenylalanine
1.912	522	0.075	8.28e-09	7.19e-08	99.99	Overlap of multiple minor compounds
2.672	446	0.106	9.59e-09	8.22e-08	100.00	Citric acid
7.342	214	0.067	1.69e-08	1.43e-07	99.99	Phenylalanine
2.572	456	0.069	1.84e-08	1.54e-07	100.00	$CaEDTA^{2-}$, citric acid
4.162	321	0.123	2.54e-08	2.09e-07	99.96	Proline, lactic acid
7.352	213	0.072	2.93e-08	2.39e-07	99.99	Phenylalanine
1.742	539	0.071	3.01e-08	2.42e-07	99.99	Leucine, lysine
1.132	600	0.098	3.27e-08	2.60e-07	99.96	Unknown
1.882	525	0.064	4.91e-08	3.86e-07	99.93	Overlap of multiple minor compounds
1.062	607	0.113	5.17e-08	4.02e-07	99.98	Valine

3.392	385	0.253	7.02e-08	5.39e-07	99.96	Methanol, proline
7.442	204	0.085	9.77e-08	7.41e-07	99.98	Phenylalanine
3.052	408	0.050	1.32e-07	9.91e-07	99.93	Creatinine
2.692	444	0.063	1.82e-07	1.35e-06	100.00	Citric acid
4.192	318	0.166	2.15e-07	1.57e-06	99.93	Unknown
3.952	342	0.106	2.52e-07	1.83e-06	99.88	Unknown
1.752	538	0.070	2.98e-07	2.13e-06	99.93	Leucine, lysine
1.732	540	0.062	3.71e-07	2.63e-06	99.93	Leucine, lysine
4.112	326	0.186	4.03e-07	2.83e-06	99.80	Proline, lactic acid
6.902	258	0.065	6.95e-07	4.81e-06	99.88	Tyrosine
2.352	478	0.069	7.00e-07	4.81e-06	99.68	Proline, glutamic acid
4.092	328	0.132	7.42e-07	5.03e-06	99.68	Unknown
1.512	562	0.101	7.46e-07	5.03e-06	99.80	Alanine
1.922	521	0.067	1.05e-06	6.97e-06	99.80	Overlap of multiple minor compounds
7.192	229	0.074	1.46e-06	9.66e-06	99.68	Tyrosine
2.302	483	0.100	3.41e-06	2.23e-05	99.48	Lipid (methylene carbonyl)
6.922	256	0.093	3.68e-06	2.38e-05	99.68	Tyrosine
7.392	209	0.062	4.20e-06	2.69e-05	99.68	Phenylalanine
2.312	482	0.088	5.14e-06	3.26e-05	99.18	Lipid (methylene carbonyl)
6.832	265	0.077	5.51e-06	3.47e-05	99.18	Unknown
3.042	409	0.029	5.61e-06	3.49e-05	99.18	Lysine, unknown
1.872	526	0.055	6.32e-06	3.90e-05	99.18	Overlap of multiple minor compounds
3.902	347	0.050	7.63e-06	4.66e-05	99.48	D-glucose, unknown
3.072	406	0.035	9.50e-06	5.75e-05	98.74	Unknown
7.402	208	0.063	1.18e-05	7.06e-05	99.48	Phenylalanine
2.292	484	0.110	1.77e-05	1.05e-04	98.74	Lipid (methylene carbonyl)
1.722	541	0.049	1.84e-05	1.08e-04	98.74	Leucine, lysine
2.402	473	0.047	2.04e-05	1.19e-04	98.11	Glutamine, carnitine
3.312	393	0.100	2.06e-05	1.20e-04	98.74	Unknown
4.062	331	0.119	2.15e-05	1.23e-04	98.74	Creatinine
1.762	537	0.056	4.25e-05	2.42e-04	98.11	Leucine, lysine
1.352	578	0.136	4.30e-05	2.43e-04	98.11	Lipid methylene, lactic acid, threonine
1.702	543	0.051	4.79e-05	2.68e-04	97.23	Unknown, arginine
1.482	565	0.075	5.31e-05	2.94e-04	97.23	Alanine
1.342	579	0.132	6.10e-05	3.35e-04	98.11	Lipid methylene, lactic acid, threonine
2.562	457	0.038	6.35e-05	3.46e-04	98.74	$CaEDTA^{2-}$
4.312	306	0.086	7.17e-05	3.88e-04	97.23	Lipid alpha-methylene to carboxyl, lipid glycerine
8.092	139	0.184	7.48e-05	4.01e-04	96.03	Trigonelline
1.032	610	0.090	1.06e-04	5.65e-04	97.23	L-isoleucine, lipid methyl, cholesterol (ester)

2.322	481	0.069	1.08e-04	5.73e-04	96.03	Lipid (methylene carbonyl)
1.232	590	0.078	1.10e-04	5.78e-04	96.03	Lipid methylene
1.022	611	0.082	1.20e-04	6.21e-04	97.23	L-isoleucine, lipid methyl, cholesterol (ester)
7.332	215	0.055	1.47e-04	7.56e-04	94.44	Phenylalanine
1.462	567	0.081	1.59e-04	8.13e-04	96.03	Lipid methylene
2.542	459	0.056	1.65e-04	8.38e-04	98.11	Unknown
6.822	266	0.064	1.78e-04	8.98e-04	96.03	Unknown
9.172	31	-0.357	1.88e-04	9.41e-04	96.03	Unknown
1.682	545	0.062	1.98e-04	9.82e-04	94.44	Unknown, arginine
1.242	589	0.083	2.02e-04	9.97e-04	96.03	Lipid methylene
2.282	485	0.119	2.13e-04	1.04e-03	96.03	Lipid (methylene carbonyl)
1.142	599	0.071	2.16e-04	1.05e-03	94.44	Unknown
3.152	398	0.027	2.21e-04	1.06e-03	86.62	$CaEDTA^{2-}$
8.102	138	0.180	2.34e-04	1.12e-03	96.03	Trigonelline
4.102	327	0.135	2.36e-04	1.12e-03	94.44	Unknown
1.932	520	0.049	2.45e-04	1.15e-03	96.03	Acetic acid
6.842	264	0.063	2.53e-04	1.18e-03	94.44	Unknown
2.342	479	0.061	2.56e-04	1.19e-03	94.44	Proline, glutamic acid
3.172	396	0.052	3.03e-04	1.40e-03	92.39	$CaEDTA^{2-}$
6.962	252	0.065	3.08e-04	1.41e-03	94.44	Unknown
7.142	234	0.072	3.27e-04	1.49e-03	94.44	Unknown
1.692	544	0.051	3.67e-04	1.66e-03	92.39	Unknown, arginine
2.112	502	0.063	4.03e-04	1.81e-03	92.39	Lipid allylic
7.182	230	0.068	4.14e-04	1.84e-03	94.44	Unknown
7.252	223	0.058	4.15e-04	1.84e-03	92.39	Unknown
1.362	577	0.126	4.46e-04	1.96e-03	94.44	Lipid methylene, lactic acid, threonine
2.872	426	0.087	4.60e-04	2.01e-03	92.39	Lipid diallylic
1.712	542	0.040	4.70e-04	2.04e-03	92.39	Leucine, lysine
1.452	568	0.084	4.89e-04	2.11e-03	92.39	Lipid methylene
6.972	251	0.078	5.47e-04	2.34e-03	94.44	Unknown
8.032	145	0.116	5.56e-04	2.37e-03	92.39	Unknown
2.652	448	0.035	5.75e-04	2.43e-03	92.39	Unknown
2.662	447	0.077	6.21e-04	2.61e-03	94.44	Citric acid
1.472	566	0.069	6.25e-04	2.61e-03	89.80	Lipid methylene
6.912	257	0.068	6.76e-04	2.81e-03	92.39	Tyrosine
2.332	480	0.062	6.87e-04	2.84e-03	89.80	Proline, glutamic acid
1.862	527	0.043	6.95e-04	2.85e-03	89.80	Unknown
7.072	241	0.065	6.99e-04	2.85e-03	89.80	Unknown
7.132	235	0.065	7.23e-04	2.93e-03	92.39	Unknown
7.242	224	0.052	7.42e-04	2.99e-03	89.80	Unknown
0.992	614	0.071	8.30e-04	3.32e-03	89.80	Leucine, lipid methyl, cholesterol (ester)
1.122	601	0.075	8.38e-04	3.33e-03	89.80	Unknown

6.542	294	0.319	8.63e-04	3.41e-03	89.80	Unknown
0.982	615	0.069	9.17e-04	3.60e-03	89.80	Leucine, lipid methyl, cholesterol (ester)
1.972	516	0.054	9.59e-04	3.75e-03	89.80	Lipid allylic
6.812	267	0.067	9.89e-04	3.84e-03	92.39	Unknown
7.562	192	-0.085	1.03e-03	3.99e-03	92.39	Unknown
1.982	515	0.061	1.05e-03	4.04e-03	89.80	Lipid allylic
2.422	471	0.061	1.11e-03	4.25e-03	89.80	Glutamine, carnitine
7.122	236	0.050	1.12e-03	4.26e-03	89.80	Unknown
4.182	319	0.141	1.15e-03	4.33e-03	89.80	Unknown
2.122	501	0.055	1.17e-03	4.38e-03	86.62	Lipid allylic
8.042	144	0.244	1.32e-03	4.90e-03	89.80	Unknown
4.042	333	0.085	1.37e-03	5.09e-03	89.80	Unknown
2.902	423	0.045	1.40e-03	5.16e-03	82.82	Unknown
1.522	561	0.053	1.41e-03	5.16e-03	86.62	Lipids (?)
7.232	225	0.048	1.42e-03	5.19e-03	86.62	Unknown
2.852	428	0.104	1.51e-03	5.49e-03	86.62	Lipid diallylic
7.262	222	0.053	1.55e-03	5.60e-03	86.62	Unknown
7.322	216	0.049	1.74e-03	6.24e-03	82.82	Unknown
2.492	464	-0.044	1.76e-03	6.27e-03	86.62	Glutamine
1.962	517	0.045	1.83e-03	6.48e-03	86.62	Lipid allylic
1.162	597	0.085	1.84e-03	6.50e-03	89.80	Lipid methylene
4.012	336	-0.037	1.93e-03	6.77e-03	86.62	Unknown
2.102	503	0.048	1.98e-03	6.91e-03	86.62	Lipid allylic
2.222	491	0.113	2.00e-03	6.93e-03	86.62	Lipid (methylene carbonyl)
4.342	303	0.058	2.05e-03	7.08e-03	86.62	Lipid alpha-methylene to carboxyl, lipid glycerine
2.212	492	0.101	2.14e-03	7.37e-03	86.62	Lipid (methylene carbonyl)
2.862	427	0.095	2.26e-03	7.74e-03	82.82	Lipid diallylic
7.862	162	0.094	2.30e-03	7.82e-03	82.82	Unknown
1.172	596	0.078	2.38e-03	8.05e-03	86.62	Lipid methylene
4.322	305	0.059	2.42e-03	8.14e-03	82.82	Lipid alpha-methylene to carboxyl, lipid glycerine
7.742	174	0.063	2.51e-03	8.40e-03	82.82	Unknown
4.302	307	0.072	2.55e-03	8.50e-03	82.82	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.432	470	0.039	2.57e-03	8.50e-03	82.82	Glutamine, carnitine
8.122	136	0.117	2.58e-03	8.50e-03	86.62	Unknown
7.872	161	0.085	2.66e-03	8.66e-03	82.82	Unknown
7.552	193	-0.084	2.66e-03	8.66e-03	86.62	Unknown
7.222	226	0.052	2.66e-03	8.66e-03	86.62	Tyrosine
3.332	391	0.200	2.73e-03	8.84e-03	86.62	Proline
1.252	588	0.074	2.82e-03	9.08e-03	82.82	Lipid methylene
7.272	221	0.047	3.02e-03	9.68e-03	78.40	Unknown
1.182	595	0.065	3.04e-03	9.69e-03	82.82	Lipid methylene

2.882	425	0.059	3.11e-03	9.87e-03	78.40	Lipid diallylic
1.992	514	0.057	3.22e-03	1.02e-02	82.82	Lipid allylic
2.132	500	0.037	3.45e-03	1.09e-02	78.40	Glutamine
2.202	493	0.076	3.57e-03	1.12e-02	82.82	Lipid (methylene carbonyl)
1.442	569	0.073	3.63e-03	1.13e-02	78.40	Lipid methylene
1.152	598	0.063	3.91e-03	1.21e-02	82.82	Lipid methylene
7.112	237	0.039	3.98e-03	1.23e-02	78.40	Unknown
7.672	181	0.152	4.17e-03	1.28e-02	78.40	Unknown
4.172	320	0.097	4.39e-03	1.34e-02	78.40	Unknown
2.272	486	0.117	4.42e-03	1.35e-02	78.40	Lipid (methylene carbonyl)
2.482	465	-0.040	4.58e-03	1.39e-02	78.40	Glutamine, carnitine
2.732	440	-0.034	4.73e-03	1.43e-02	78.40	$MgEDTA^{2-}$
2.092	504	0.070	4.89e-03	1.47e-02	78.40	Lipid allylic
7.052	243	0.055	5.05e-03	1.51e-02	78.40	Unknown
7.632	185	0.097	5.06e-03	1.51e-02	78.40	Unknown
0.972	616	0.057	5.44e-03	1.61e-02	78.40	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.532	560	0.057	5.55e-03	1.64e-02	73.37	Lipids (?)
8.112	137	0.146	5.76e-03	1.69e-02	73.37	Trigonelline
2.642	449	0.031	6.47e-03	1.89e-02	73.37	Unknown
8.082	140	0.103	6.64e-03	1.93e-02	78.40	Trigonelline
2.252	488	0.099	6.68e-03	1.93e-02	73.37	Lipid (methylene carbonyl), acetone
2.842	429	0.094	6.84e-03	1.96e-02	73.37	Lipid diallylic
2.532	460	0.040	6.86e-03	1.96e-02	73.37	Unknown
2.452	468	0.051	6.87e-03	1.96e-02	73.37	Glutamine, carnitine
2.972	416	0.019	7.21e-03	2.05e-02	73.37	Unknown
1.432	570	0.072	8.24e-03	2.33e-02	73.37	Lipid methylene
1.852	528	0.034	8.45e-03	2.38e-02	73.37	Unknown
1.672	546	0.052	8.78e-03	2.47e-02	73.37	Unknown, arginine
1.952	518	0.034	9.03e-03	2.53e-02	73.37	Acetic acid
7.152	233	0.065	9.20e-03	2.56e-02	78.40	Unknown
1.412	572	0.109	9.32e-03	2.58e-02	73.37	Lipid methylene
7.282	220	0.038	9.70e-03	2.68e-02	67.82	Unknown
2.002	513	0.051	9.89e-03	2.72e-02	67.82	Lipid allylic
7.572	191	-0.074	1.03e-02	2.81e-02	73.37	Unknown
7.172	231	0.072	1.08e-02	2.94e-02	73.37	Unknown
1.662	547	0.059	1.12e-02	3.04e-02	67.82	Lipids (?)
4.272	310	0.044	1.13e-02	3.05e-02	67.82	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
9.142	34	0.303	1.14e-02	3.06e-02	73.37	Trigonelline
3.092	404	0.017	1.15e-02	3.09e-02	67.82	$CaEDTA^{2-}$
7.302	218	0.037	1.17e-02	3.13e-02	67.82	Unknown
7.422	206	0.073	1.17e-02	3.13e-02	73.37	Phenylalanine
8.022	146	0.098	1.23e-02	3.25e-02	73.37	Unknown

1.422	571	0.085	1.23e-02	3.25e-02	67.82	Lipid methylene
0.962	617	0.057	1.26e-02	3.32e-02	67.82	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
2.892	424	0.034	1.33e-02	3.47e-02	61.84	Lipid diallylic
2.232	490	0.094	1.33e-02	3.47e-02	67.82	Lipid (methylene carbonyl)
6.762	272	0.179	1.34e-02	3.49e-02	67.82	Unknown
8.872	61	0.297	1.36e-02	3.53e-02	67.82	Trigonelline
3.182	395	0.020	1.41e-02	3.63e-02	55.55	$CaEDTA^{2-}$
7.412	207	-0.065	1.44e-02	3.71e-02	73.37	Phenylalanine
8.982	50	-0.237	1.46e-02	3.72e-02	67.82	Unknown
2.412	472	0.048	1.46e-02	3.72e-02	67.82	Glutamine, carnitine
1.402	573	0.127	1.47e-02	3.72e-02	67.82	Lipid methylene
1.542	559	0.065	1.48e-02	3.75e-02	67.82	Lipids (?)
7.992	149	0.145	1.49e-02	3.77e-02	67.82	Unknown
0.952	618	0.065	1.51e-02	3.79e-02	67.82	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
8.352	113	0.193	1.54e-02	3.84e-02	73.37	Unknown
7.292	219	0.035	1.54e-02	3.84e-02	61.84	Unknown
6.982	250	0.073	1.69e-02	4.20e-02	67.82	Unknown
1.842	529	0.030	1.70e-02	4.20e-02	61.84	Unknown
1.262	587	0.064	1.71e-02	4.21e-02	61.84	Lipid methylene
2.832	430	0.080	1.72e-02	4.21e-02	61.84	Lipid diallylic
2.192	494	0.049	1.72e-02	4.21e-02	61.84	Lipid (methylene carbonyl)
6.792	269	0.053	1.78e-02	4.34e-02	67.82	Unknown
3.142	399	0.023	1.82e-02	4.41e-02	49.13	$CaEDTA^{2-}$
3.342	390	0.068	1.85e-02	4.48e-02	61.84	Proline
7.362	212	0.036	1.88e-02	4.53e-02	67.82	Phenylalanine
6.782	270	0.095	1.95e-02	4.68e-02	61.84	Unknown
7.062	242	0.050	1.96e-02	4.68e-02	61.84	Unknown
8.392	109	0.188	2.01e-02	4.79e-02	67.82	Unknown
3.122	401	0.019	2.04e-02	4.84e-02	49.13	$CaEDTA^{2-}$
2.262	487	0.108	2.08e-02	4.93e-02	61.84	Lipid (methylene carbonyl)

Table 7.8: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from diabetic nephropathy from those suffering from glomerulonephritis.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	ID	log(FC)	*P*-value un-adjusted	*P*-value B/H-adjusted	Statistical power	Identified compounds
3.802	357	0.476	2.22e-49	1.28e-46	100.00	D-glucose, alanine
3.842	353	0.542	5.22e-49	1.28e-46	100.00	D-glucose, unknown
3.782	359	0.430	5.80e-49	1.28e-46	100.00	D-glucose, alanine, glutamine, arginine
3.882	349	0.537	8.72e-49	1.44e-46	100.00	D-glucose, unknown
3.872	350	0.472	1.12e-47	1.48e-45	100.00	D-glucose, unknown
3.862	351	0.578	4.48e-47	4.93e-45	100.00	D-glucose, unknown
3.742	363	0.576	5.53e-47	4.98e-45	100.00	D-glucose, leucine
3.442	380	0.601	6.12e-47	4.98e-45	100.00	D-glucose, carnitine, taurine, proline
3.412	383	0.608	6.79e-47	4.98e-45	100.00	D-glucose, carnitine, taurine, proline
3.562	368	0.553	9.41e-47	6.21e-45	100.00	D-glucose
3.912	346	0.525	1.28e-46	7.65e-45	100.00	D-glucose, betaine, unknown
3.852	352	0.553	1.39e-46	7.67e-45	100.00	D-glucose, unknown
3.552	369	0.569	2.04e-46	1.03e-44	100.00	D-glucose, myo-inositol
3.432	381	0.607	2.52e-46	1.19e-44	100.00	D-glucose, carnitine, taurine, proline
3.452	379	0.534	4.13e-46	1.82e-44	100.00	D-glucose, carnitine, proline
3.502	374	0.619	5.50e-46	2.24e-44	100.00	D-glucose
3.482	376	0.607	5.76e-46	2.24e-44	100.00	D-glucose
3.762	361	0.529	7.99e-46	2.93e-44	100.00	D-glucose, arginine, glutamine, glutamic acid
3.722	365	0.485	8.83e-46	3.07e-44	100.00	D-glucose, N,N-dimethylglycine
3.932	344	0.534	9.60e-46	3.17e-44	100.00	D-glucose
3.572	367	0.521	1.05e-45	3.32e-44	100.00	D-glucose, glycine
3.512	373	0.616	2.47e-45	7.42e-44	100.00	D-glucose
3.492	375	0.605	2.06e-44	5.93e-43	100.00	D-glucose
3.752	362	0.553	2.52e-44	6.93e-43	100.00	D-glucose, glutamic acid
3.532	371	0.537	1.32e-43	3.48e-42	100.00	D-glucose
3.472	377	0.573	5.73e-43	1.45e-41	100.00	D-glucose
3.772	360	0.477	6.71e-43	1.64e-41	100.00	D-glucose, alanine, glutamine, arginine
3.422	382	0.584	5.12e-41	1.21e-39	100.00	D-glucose, carnitine, taurine, proline
3.732	364	0.505	2.99e-40	6.81e-39	100.00	D-glucose, unknown
3.792	358	0.296	7.64e-40	1.68e-38	100.00	D-glucose, alanine
3.812	356	0.417	8.18e-39	1.74e-37	100.00	D-glucose
3.942	343	0.526	4.26e-36	8.79e-35	100.00	D-glucose
3.922	345	0.341	3.94e-31	7.88e-30	100.00	D-glucose, unknown
3.962	341	0.202	2.11e-30	4.11e-29	100.00	Unknown
3.522	372	0.514	3.29e-27	6.21e-26	100.00	D-glucose

3.542	370	0.396	3.82e-25	7.00e-24	100.00	D-glucose, myo-inositol
3.832	354	0.246	1.26e-24	2.25e-23	100.00	Unknown
3.402	384	0.341	2.30e-22	4.00e-21	100.00	Unknown
3.972	340	0.171	1.22e-21	2.06e-20	100.00	Unknown
3.822	355	0.257	1.01e-20	1.67e-19	100.00	Unknown
3.892	348	0.177	3.25e-18	5.23e-17	100.00	Unknown
3.462	378	0.273	1.96e-16	3.07e-15	100.00	D-glucose
4.122	325	0.345	1.21e-15	1.86e-14	100.00	Proline, lactic acid
4.142	323	0.320	2.89e-12	4.33e-11	100.00	Proline, lactic acid
4.132	324	0.340	4.41e-12	6.46e-11	100.00	Proline, lactic acid
1.102	603	0.225	5.28e-12	7.58e-11	100.00	Unknown
4.152	322	0.248	8.99e-12	1.26e-10	100.00	Proline, lactic acid
1.132	600	0.188	7.76e-11	1.07e-09	100.00	Unknown
1.232	590	0.211	9.83e-11	1.32e-09	100.00	Lipid methylene
1.092	604	0.212	1.09e-10	1.44e-09	100.00	Unknown
1.082	605	0.199	1.86e-10	2.41e-09	100.00	Unknown
2.402	473	0.112	4.01e-10	5.09e-09	100.00	Glutamine, carnitine
1.152	598	0.217	8.22e-10	1.02e-08	100.00	Lipid methylene
1.692	544	0.141	1.50e-09	1.81e-08	100.00	Unknown, arginine
2.312	482	0.189	1.51e-09	1.81e-08	100.00	Lipid (methylene carbonyl)
1.002	613	0.179	2.54e-09	2.99e-08	100.00	Valine, lipid methyl, cholesterol (ester)
1.682	545	0.160	2.83e-09	3.28e-08	100.00	Unknown, arginine
1.702	543	0.122	3.02e-09	3.44e-08	100.00	Unknown, arginine
2.382	475	0.175	4.11e-09	4.60e-08	100.00	Proline, glutamic acid
1.242	589	0.211	5.39e-09	5.93e-08	99.99	Lipid methylene
2.122	501	0.160	6.14e-09	6.65e-08	100.00	Lipid allylic
2.302	483	0.204	6.34e-09	6.75e-08	99.99	Lipid (methylene carbonyl)
1.362	577	0.338	6.80e-09	7.12e-08	99.99	Lipid methylene, lactic acid, threonine
1.072	606	0.172	7.11e-09	7.33e-08	100.00	Valine
1.342	579	0.307	9.49e-09	9.63e-08	99.99	Lipid methylene, lactic acid, threonine
4.312	306	0.202	9.67e-09	9.67e-08	99.99	Lipid alpha-methylene to carboxyl, lipid glycerine
2.362	477	0.158	1.20e-08	1.19e-07	99.99	Proline, glutamic acid
1.972	516	0.151	1.25e-08	1.21e-07	100.00	Lipid allylic
1.962	517	0.132	1.39e-08	1.33e-07	100.00	Lipid allylic
7.262	222	0.153	1.47e-08	1.38e-07	100.00	Unknown
6.832	265	0.155	1.70e-08	1.58e-07	100.00	Unknown
7.222	226	0.159	1.82e-08	1.67e-07	99.99	Tyrosine
1.142	599	0.174	1.99e-08	1.80e-07	99.98	Unknown
7.272	221	0.144	2.07e-08	1.85e-07	100.00	Unknown
2.322	481	0.162	2.52e-08	2.22e-07	99.98	Lipid (methylene carbonyl)
2.392	474	0.176	3.29e-08	2.86e-07	99.99	Unknown

1.352	578	0.299	3.55e-08	3.04e-07	99.98	Lipid methylene, lactic acid, threonine
2.652	448	0.090	4.12e-08	3.49e-07	100.00	Unknown
1.872	526	0.108	5.04e-08	4.21e-07	99.98	Overlap of multiple minor compounds
1.672	546	0.175	5.26e-08	4.34e-07	99.98	Unknown, arginine
1.712	542	0.101	5.51e-08	4.49e-07	99.98	Leucine, lysine
2.292	484	0.225	6.57e-08	5.24e-07	99.97	Lipid (methylene carbonyl)
1.842	529	0.111	6.59e-08	5.24e-07	99.98	Unknown
7.252	223	0.145	6.85e-08	5.38e-07	99.99	Unknown
1.862	527	0.111	6.98e-08	5.41e-07	99.98	Unknown
2.642	449	0.101	7.05e-08	5.41e-07	99.98	Unknown
1.982	515	0.161	7.73e-08	5.86e-07	99.98	Lipid allylic
6.842	264	0.148	9.06e-08	6.79e-07	99.99	Unknown
2.372	476	0.159	1.01e-07	7.49e-07	99.97	Proline, glutamic acid
1.222	591	0.217	1.03e-07	7.59e-07	99.99	Lipid methylene
4.112	326	0.314	1.22e-07	8.81e-07	99.98	Proline, lactic acid
7.232	225	0.129	1.31e-07	9.37e-07	99.99	Unknown
7.332	215	0.123	1.38e-07	9.77e-07	99.98	Phenylalanine
7.282	220	0.125	1.39e-07	9.77e-07	99.99	Unknown
7.292	219	0.123	1.48e-07	1.03e-06	99.99	Unknown
7.202	228	0.169	1.54e-07	1.06e-06	99.95	Tyrosine
4.342	303	0.159	1.64e-07	1.12e-06	99.95	Lipid alpha-methylene to carboxyl, lipid glycerine
3.712	366	0.144	1.70e-07	1.14e-06	99.76	Unknown
7.322	216	0.133	1.71e-07	1.14e-06	99.99	Unknown
1.062	607	0.175	1.84e-07	1.21e-06	99.95	Valine
4.322	305	0.163	2.18e-07	1.42e-06	99.94	Lipid alpha-methylene to carboxyl, lipid glycerine
7.302	218	0.125	2.20e-07	1.42e-06	99.98	Unknown
1.812	532	0.108	2.30e-07	1.47e-06	99.95	Unknown
7.212	227	0.154	2.50e-07	1.59e-06	99.97	Tyrosine
1.662	547	0.195	2.59e-07	1.63e-06	99.94	Lipids (?)
2.352	478	0.115	3.17e-07	1.97e-06	99.87	Proline, glutamic acid
7.312	217	0.130	3.19e-07	1.97e-06	99.98	Unknown
1.012	612	0.173	3.41e-07	2.09e-06	99.94	Valine, lipid methyl, cholesterol (ester)
1.882	525	0.097	3.66e-07	2.21e-06	99.94	Overlap of multiple minor compounds
1.252	588	0.205	3.73e-07	2.24e-06	99.87	Lipid methylene
6.822	266	0.141	4.09e-07	2.43e-06	99.98	Unknown
1.492	564	0.157	4.16e-07	2.44e-06	99.94	Alanine
1.482	565	0.151	4.18e-07	2.44e-06	99.94	Alanine
1.852	528	0.106	4.48e-07	2.59e-06	99.91	Unknown
2.132	500	0.105	4.54e-07	2.61e-06	99.87	Glutamine

2.282	485	0.261	5.20e-07	2.96e-06	99.91	Lipid (methylene carbonyl)
1.462	567	0.173	6.19e-07	3.49e-06	99.87	Lipid methylene
4.032	334	0.139	6.42e-07	3.59e-06	99.87	Unknown
2.332	480	0.147	6.89e-07	3.82e-06	99.87	Proline, glutamic acid
1.892	524	0.094	7.19e-07	3.93e-06	99.87	Overlap of multiple minor compounds
1.512	562	0.163	7.20e-07	3.93e-06	99.87	Alanine
1.992	514	0.155	8.10e-07	4.38e-06	99.87	Lipid allylic
3.902	347	0.090	8.79e-07	4.71e-06	99.76	D-glucose, unknown
0.962	617	0.181	9.02e-07	4.80e-06	99.82	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.452	568	0.193	9.09e-07	4.80e-06	99.82	Lipid methylene
1.502	563	0.158	9.20e-07	4.82e-06	99.87	Alanine
1.902	523	0.090	9.95e-07	5.17e-06	99.87	Overlap of multiple minor compounds
1.832	530	0.093	1.04e-06	5.31e-06	99.91	Unknown
1.122	601	0.179	1.04e-06	5.31e-06	99.87	Unknown
4.302	307	0.190	1.06e-06	5.38e-06	99.87	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.822	531	0.097	1.11e-06	5.61e-06	99.91	Unknown
2.112	502	0.140	1.14e-06	5.70e-06	99.82	Lipid allylic
1.802	533	0.112	1.26e-06	6.28e-06	99.91	Unknown
1.162	597	0.215	1.37e-06	6.77e-06	99.98	Lipid methylene
1.022	611	0.168	1.46e-06	7.13e-06	99.76	L-isoleucine, lipid methyl, cholesterol (ester)
1.172	596	0.201	1.47e-06	7.13e-06	99.98	Lipid methylene
4.332	304	0.146	1.61e-06	7.76e-06	99.76	Lipid alpha-methylene to carboxyl, lipid glycerine
2.342	479	0.130	1.87e-06	8.93e-06	99.67	Proline, glutamic acid
6.922	256	0.155	1.93e-06	9.14e-06	99.76	Tyrosine
1.952	518	0.099	2.12e-06	9.92e-06	99.82	Acetic acid
1.522	561	0.127	2.12e-06	9.92e-06	99.82	Lipids (?)
0.972	616	0.159	2.17e-06	1.01e-05	99.67	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
7.372	211	0.113	3.07e-06	1.42e-05	99.67	Phenylalanine
2.272	486	0.310	3.34e-06	1.53e-05	99.67	Lipid (methylene carbonyl)
7.242	224	0.115	3.37e-06	1.53e-05	99.87	Unknown
1.472	566	0.151	3.58e-06	1.62e-05	99.67	Lipid methylene
0.952	618	0.201	3.62e-06	1.63e-05	99.56	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
4.102	327	0.276	3.81e-06	1.70e-05	99.76	Unknown
2.002	513	0.148	3.89e-06	1.72e-05	99.67	Lipid allylic
1.722	541	0.085	4.05e-06	1.78e-05	99.56	Leucine, lysine
2.102	503	0.115	4.11e-06	1.79e-05	99.67	Lipid allylic
4.162	321	0.163	5.02e-06	2.18e-05	99.56	Proline, lactic acid

2.252	488	0.270	5.45e-06	2.34e-05	99.67	Lipid (methylene carbonyl), acetone
7.382	210	0.099	5.47e-06	2.34e-05	99.56	Phenylalanine
3.372	387	-0.142	6.07e-06	2.58e-05	98.68	Methanol, proline
1.182	595	0.160	6.10e-06	2.58e-05	99.22	Lipid methylene
1.262	587	0.197	6.37e-06	2.68e-05	99.41	Lipid methylene
0.992	614	0.156	6.71e-06	2.80e-05	99.22	Leucine, lipid methyl, cholesterol (ester)
6.912	257	0.147	7.05e-06	2.92e-05	99.56	Tyrosine
4.192	318	0.232	7.46e-06	3.08e-05	99.67	Unknown
0.982	615	0.150	7.56e-06	3.10e-05	99.22	Leucine, lipid methyl, cholesterol (ester)
1.272	586	0.208	9.38e-06	3.82e-05	99.22	Lipid methylene
1.532	560	0.145	1.24e-05	5.03e-05	99.22	Lipids (?)
1.442	569	0.178	1.34e-05	5.40e-05	99.22	Lipid methylene
1.652	548	0.200	1.46e-05	5.82e-05	99.22	Lipids (?)
1.112	602	0.138	1.46e-05	5.82e-05	98.68	Unknown
6.852	263	0.121	1.56e-05	6.17e-05	99.41	Unknown
2.262	487	0.326	1.59e-05	6.26e-05	99.22	Lipid (methylene carbonyl)
7.142	234	0.140	1.77e-05	6.88e-05	99.22	Unknown
2.632	450	0.074	1.78e-05	6.88e-05	99.41	Unknown
6.812	267	0.142	1.78e-05	6.88e-05	97.27	Unknown
1.422	571	0.236	1.86e-05	7.15e-05	98.98	Lipid methylene
2.222	491	0.254	1.88e-05	7.18e-05	99.22	Lipid (methylene carbonyl)
4.052	332	0.196	2.14e-05	8.11e-05	97.84	Unknown
1.432	570	0.189	2.21e-05	8.33e-05	98.98	Lipid methylene
4.292	308	0.160	2.29e-05	8.59e-05	98.98	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.412	572	0.288	2.34e-05	8.73e-05	98.98	Lipid methylene
7.072	241	0.131	2.38e-05	8.83e-05	98.68	Unknown
7.192	229	0.103	3.39e-05	1.25e-04	99.22	Tyrosine
1.212	592	0.197	3.54e-05	1.30e-04	99.56	Lipid methylene
7.342	214	0.080	3.69e-05	1.34e-04	98.98	Phenylalanine
1.282	585	0.209	3.92e-05	1.42e-04	98.68	Lipid methylene
2.212	492	0.220	4.00e-05	1.44e-04	98.30	Lipid (methylene carbonyl)
1.752	538	0.091	4.12e-05	1.48e-04	98.30	Leucine, lysine
4.282	309	0.126	4.22e-05	1.50e-04	98.68	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
0.942	619	0.223	4.23e-05	1.50e-04	98.30	Cholesterol, lipid methyl
1.742	539	0.085	4.60e-05	1.62e-04	98.30	Leucine, lysine
4.352	302	0.129	5.66e-05	1.99e-04	97.84	Lipid alpha-methylene to carboxyl, lipid glycerine
1.032	610	0.152	5.84e-05	2.04e-04	97.84	L-isoleucine, lipid methyl, cholesterol (ester)
1.402	573	0.339	6.12e-05	2.13e-04	97.84	Lipid methylene

2.232	490	0.246	6.58e-05	2.27e-04	98.30	Lipid (methylene carbonyl)
1.302	583	0.284	7.16e-05	2.46e-04	97.84	Lipid methylene
1.732	540	0.079	7.24e-05	2.47e-04	97.84	Leucine, lysine
2.012	512	0.136	7.26e-05	2.47e-04	98.30	Lipid allylic
2.852	428	0.210	8.12e-05	2.75e-04	97.84	Lipid diallylic
1.762	537	0.087	8.70e-05	2.93e-04	97.84	Leucine, lysine
2.822	431	0.218	8.83e-05	2.95e-04	97.84	Lipid diallylic
2.622	451	0.059	8.84e-05	2.95e-04	98.68	Unknown
1.612	552	0.309	9.06e-05	3.01e-04	97.84	Lipids (?)
1.642	549	0.216	9.59e-05	3.17e-04	97.27	Lipids (?)
1.542	559	0.169	1.00e-04	3.29e-04	97.27	Lipids (?)
1.622	551	0.282	1.13e-04	3.69e-04	97.27	Lipids (?)
2.422	471	0.116	1.31e-04	4.27e-04	97.84	Glutamine, carnitine
1.392	574	0.344	1.36e-04	4.39e-04	97.27	Lipid methylene
1.602	553	0.310	1.38e-04	4.44e-04	97.27	Lipids (?)
2.412	472	0.123	1.39e-04	4.45e-04	97.27	Glutamine, carnitine
2.242	489	0.250	1.46e-04	4.64e-04	97.27	Lipid (methylene carbonyl), acetone
2.202	493	0.162	1.46e-04	4.64e-04	96.58	Lipid (methylene carbonyl)
1.632	550	0.243	1.51e-04	4.77e-04	96.58	Lipids (?)
7.052	243	0.120	1.53e-04	4.81e-04	95.76	Unknown
0.852	628	0.099	1.56e-04	4.87e-04	96.58	Cholesterol, lipid methyl
2.082	505	0.146	1.61e-04	5.01e-04	97.27	Lipid allylic
0.842	629	0.100	1.63e-04	5.05e-04	96.58	Cholesterol, lipid methyl
1.382	575	0.340	1.80e-04	5.56e-04	96.58	Lipid methylene
1.552	558	0.202	1.90e-04	5.83e-04	96.58	Lipids (?)
2.832	430	0.204	1.91e-04	5.83e-04	96.58	Lipid diallylic
1.312	582	0.282	2.04e-04	6.19e-04	96.58	Lipid methylene
0.932	620	0.242	2.42e-04	7.32e-04	95.76	Cholesterol, lipid methyl
1.292	584	0.218	2.43e-04	7.32e-04	96.58	Lipid methylene
1.332	580	0.305	2.47e-04	7.41e-04	95.76	Lipid methylene
3.982	339	0.078	2.56e-04	7.63e-04	94.78	Unknown
2.842	429	0.207	2.61e-04	7.76e-04	95.76	Lipid diallylic
1.592	554	0.291	2.64e-04	7.81e-04	95.76	Lipids (?)
3.392	385	0.276	2.90e-04	8.54e-04	94.78	Methanol, proline
1.572	556	0.244	2.97e-04	8.71e-04	95.76	Lipids (?)
2.862	427	0.182	3.03e-04	8.83e-04	94.78	Lipid diallylic
1.912	522	0.076	3.04e-04	8.83e-04	96.58	Overlap of multiple minor compounds
1.582	555	0.269	3.08e-04	8.90e-04	95.76	Lipids (?)
1.772	536	0.085	3.11e-04	8.96e-04	95.76	Leucine, lysine
2.022	511	0.140	3.15e-04	9.03e-04	95.76	Lipid allylic
7.362	212	0.090	3.16e-04	9.04e-04	95.76	Phenylalanine
4.092	328	0.155	3.21e-04	9.14e-04	96.58	Unknown
1.562	557	0.222	3.33e-04	9.42e-04	95.76	Lipids (?)

2.032	510	0.157	3.34e-04	9.42e-04	95.76	Lipid allylic
1.792	534	0.099	3.39e-04	9.52e-04	96.58	Unknown
1.052	608	0.170	3.48e-04	9.73e-04	95.76	Valine
1.322	581	0.282	3.94e-04	1.10e-03	94.78	Lipid methylene
1.042	609	0.162	4.12e-04	1.14e-03	93.62	L-isoleucine, lipid methyl, cholesterol (ester)
1.372	576	0.320	4.24e-04	1.17e-03	94.78	Lipid methylene
2.872	426	0.140	5.10e-04	1.40e-03	93.62	Lipid diallylic
2.812	432	0.192	5.53e-04	1.52e-03	94.78	Lipid diallylic
2.432	470	0.072	5.70e-04	1.55e-03	93.62	Glutamine, carnitine
4.272	310	0.097	6.44e-04	1.75e-03	92.28	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
7.152	233	0.136	7.78e-04	2.10e-03	93.62	Unknown
2.042	509	0.155	8.56e-04	2.31e-03	92.28	Lipid allylic
0.922	621	0.234	9.22e-04	2.47e-03	90.74	Cholesterol, lipid methyl
6.902	258	0.070	9.39e-04	2.51e-03	92.28	Tyrosine
7.132	235	0.103	1.04e-03	2.77e-03	90.74	Unknown
2.072	506	0.105	1.05e-03	2.77e-03	90.74	Lipid allylic
9.232	25	0.488	1.12e-03	2.95e-03	90.74	Unknown
3.952	342	0.108	1.20e-03	3.14e-03	86.99	Unknown
0.602	653	-0.489	1.44e-03	3.77e-03	88.98	Unkown
3.032	410	0.035	1.64e-03	4.27e-03	84.76	Lysine, unknown
7.122	236	0.079	1.64e-03	4.27e-03	86.99	Unknown
6.932	255	0.103	1.66e-03	4.30e-03	90.74	Tyrosine
6.742	274	-0.213	1.70e-03	4.37e-03	88.98	Unknown
7.352	213	0.065	1.89e-03	4.86e-03	88.98	Phenylalanine
2.792	434	0.137	1.98e-03	5.06e-03	88.98	Lipid diallylic
7.112	237	0.066	2.29e-03	5.84e-03	86.99	Unknown
2.802	433	0.148	2.35e-03	5.97e-03	88.98	Lipid diallylic
1.782	535	0.088	2.47e-03	6.23e-03	88.98	Unknown
2.192	494	0.100	2.66e-03	6.70e-03	84.76	Lipid (methylene carbonyl)
4.362	301	0.104	2.73e-03	6.86e-03	84.76	Unknown
6.802	268	0.103	3.00e-03	7.51e-03	84.76	Unknown
2.962	417	-0.052	3.05e-03	7.60e-03	82.30	Unknown
4.182	319	0.208	3.17e-03	7.86e-03	86.99	Unknown
4.002	337	-0.054	3.36e-03	8.30e-03	76.66	Unknown
1.192	594	0.191	3.42e-03	8.41e-03	90.74	Lipid methylene
0.832	630	0.086	3.50e-03	8.56e-03	84.76	Cholesterol, lipid methyl
1.922	521	0.065	3.50e-03	8.56e-03	84.76	Overlap of multiple minor compounds
4.252	312	0.132	3.93e-03	9.57e-03	82.30	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.052	508	0.104	4.15e-03	1.01e-02	82.30	Lipid allylic
4.242	313	0.163	4.16e-03	1.01e-02	82.30	Unknown
9.402	8	-0.456	4.30e-03	1.04e-02	76.66	Unknown

7.182	230	0.089	4.33e-03	1.04e-02	84.76	Unknown
2.062	507	0.065	4.50e-03	1.08e-02	82.30	Lipid allylic
4.172	320	0.155	4.89e-03	1.17e-02	82.30	Unknown
0.802	633	0.118	5.32e-03	1.26e-02	79.59	Cholesterol, lipid methyl
1.202	593	0.215	5.41e-03	1.28e-02	86.99	Lipid methylene
0.862	627	0.082	5.70e-03	1.34e-02	79.59	Cholesterol, lipid methyl
2.482	465	-0.063	5.71e-03	1.34e-02	76.66	Glutamine, carnitine
2.672	446	0.082	6.08e-03	1.42e-02	86.99	Citric acid
4.062	331	0.124	6.25e-03	1.46e-02	79.59	Creatinine
4.372	300	0.115	6.34e-03	1.47e-02	76.66	Unknown
2.782	435	0.123	6.36e-03	1.47e-02	79.59	Lipid diallylic
7.012	247	0.120	6.44e-03	1.49e-02	76.66	Unknown
1.932	520	0.059	6.56e-03	1.51e-02	79.59	Acetic acid
0.912	622	0.177	6.70e-03	1.54e-02	76.66	Cholesterol, lipid methyl
2.092	504	0.107	7.68e-03	1.75e-02	79.59	Lipid allylic
4.212	316	0.150	7.73e-03	1.76e-02	79.59	Unknown
0.812	632	0.112	7.79e-03	1.77e-02	76.66	Cholesterol, lipid methyl
7.062	242	0.092	7.91e-03	1.79e-02	73.50	Unknown
0.792	634	0.099	8.94e-03	2.01e-02	76.66	Cholesterol, lipid methyl
4.232	314	0.153	9.43e-03	2.11e-02	73.50	Unknown
0.822	631	0.095	9.45e-03	2.11e-02	73.50	Cholesterol, lipid methyl
4.222	315	0.150	9.47e-03	2.11e-02	76.66	Unknown
3.022	411	0.032	9.78e-03	2.17e-02	76.66	Lysine, unknown
4.042	333	0.110	1.05e-02	2.34e-02	73.50	Unknown
2.882	425	0.082	1.11e-02	2.45e-02	70.15	Lipid diallylic
3.012	412	0.037	1.13e-02	2.49e-02	76.66	Lysine, unknown
2.182	495	0.074	1.26e-02	2.75e-02	70.15	Glutamine
3.312	393	0.095	1.26e-02	2.75e-02	66.61	Unknown
2.702	443	0.036	1.33e-02	2.89e-02	73.50	$MgEDTA^{2-}$
4.262	311	0.090	1.41e-02	3.06e-02	70.15	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.902	423	0.055	1.58e-02	3.41e-02	66.61	Unknown
4.202	317	0.125	1.62e-02	3.50e-02	70.15	Unknown
0.772	636	0.078	1.65e-02	3.54e-02	66.61	Cholesterol, lipid methyl
6.862	262	0.078	1.67e-02	3.57e-02	66.61	Unknown
6.672	281	-0.358	1.71e-02	3.66e-02	70.15	Unknown
0.762	637	0.077	1.78e-02	3.78e-02	66.61	Cholesterol, lipid methyl
0.592	654	-0.375	1.92e-02	4.07e-02	66.61	Unkown
6.962	252	0.068	1.93e-02	4.07e-02	62.91	Unknown
6.872	261	0.065	1.96e-02	4.11e-02	66.61	Unknown
7.392	209	0.051	1.96e-02	4.11e-02	66.61	Phenylalanine
2.612	452	0.031	1.99e-02	4.17e-02	70.15	$CaEDTA^{2-}$
6.892	259	0.049	2.02e-02	4.21e-02	66.61	Unknown
7.172	231	0.105	2.06e-02	4.30e-02	62.91	Unknown
2.742	439	0.056	2.08e-02	4.31e-02	66.61	Lipid diallylic

0.542	659	0.368	2.11e-02	4.36e-02	62.91	Unkown
8.352	113	0.294	2.32e-02	4.79e-02	66.61	Unknown

Table 7.9: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from diabetic nephropathy from those suffering from hereditary diseases.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
3.872	350	0.461	1.15e-48	7.56e-46	100.00	D-glucose, unknown
3.842	353	0.507	2.76e-46	9.09e-44	100.00	D-glucose, unknown
3.552	369	0.545	7.68e-46	1.20e-43	100.00	D-glucose, myo-inositol
3.782	359	0.400	8.03e-46	1.20e-43	100.00	D-glucose, alanine, glutamine, arginine
3.562	368	0.527	9.11e-46	1.20e-43	100.00	D-glucose
3.852	352	0.528	1.11e-45	1.23e-43	100.00	D-glucose, unknown
3.802	357	0.439	2.54e-45	2.39e-43	100.00	D-glucose, alanine
3.912	346	0.498	3.68e-45	3.03e-43	100.00	D-glucose, betaine, unknown
3.932	344	0.510	7.53e-45	5.52e-43	100.00	D-glucose
3.452	379	0.506	1.71e-44	1.13e-42	100.00	D-glucose, carnitine, proline
3.722	365	0.459	3.11e-44	1.86e-42	100.00	D-glucose, N,N-dimethylglycine
3.862	351	0.536	7.97e-44	4.38e-42	100.00	D-glucose, unknown
3.412	383	0.564	1.14e-43	5.80e-42	100.00	D-glucose, carnitine, taurine, proline
3.472	377	0.555	3.87e-43	1.83e-41	100.00	D-glucose
3.442	380	0.553	4.45e-43	1.85e-41	100.00	D-glucose, carnitine, taurine, proline
3.742	363	0.530	4.49e-43	1.85e-41	100.00	D-glucose, leucine
3.882	349	0.483	5.91e-43	2.30e-41	100.00	D-glucose, unknown
3.422	382	0.574	1.86e-42	6.64e-41	100.00	D-glucose, carnitine, taurine, proline
3.482	376	0.561	1.98e-42	6.64e-41	100.00	D-glucose
3.752	362	0.521	2.01e-42	6.64e-41	100.00	D-glucose, glutamic acid
3.512	373	0.573	2.26e-42	7.09e-41	100.00	D-glucose
3.432	381	0.557	3.13e-42	9.40e-41	100.00	D-glucose, carnitine, taurine, proline
3.572	367	0.480	5.33e-42	1.53e-40	100.00	D-glucose, glycine

3.532	371	0.506	7.28e-42	2.00e-40	100.00	D-glucose
3.732	364	0.498	8.45e-42	2.23e-40	100.00	D-glucose, unknown
3.502	374	0.566	1.18e-41	2.99e-40	100.00	D-glucose
3.762	361	0.485	1.27e-41	3.09e-40	100.00	D-glucose, arginine, glutamine, glutamic acid
3.492	375	0.548	1.88e-39	4.43e-38	100.00	D-glucose
3.772	360	0.410	1.06e-34	2.42e-33	100.00	D-glucose, alanine, glutamine, arginine
3.792	358	0.265	1.17e-34	2.58e-33	100.00	D-glucose, alanine
3.402	384	0.384	1.28e-29	2.72e-28	100.00	Unknown
3.542	370	0.414	3.73e-29	7.70e-28	100.00	D-glucose, myo-inositol
3.942	343	0.449	7.85e-29	1.57e-27	100.00	D-glucose
3.812	356	0.341	1.13e-28	2.20e-27	100.00	D-glucose
3.922	345	0.297	9.00e-26	1.70e-24	100.00	D-glucose, unknown
3.832	354	0.230	3.43e-23	6.29e-22	100.00	Unknown
3.962	341	0.167	6.26e-23	1.12e-21	100.00	Unknown
3.522	372	0.424	1.92e-20	3.34e-19	100.00	D-glucose
3.462	378	0.279	3.39e-18	5.73e-17	100.00	D-glucose
3.972	340	0.147	1.64e-17	2.71e-16	100.00	Unknown
3.892	348	0.159	5.27e-16	8.49e-15	100.00	Unknown
3.822	355	0.192	3.82e-13	6.00e-12	100.00	Unknown
2.402	473	0.111	1.37e-10	2.10e-09	100.00	Glutamine, carnitine
3.712	366	0.162	9.83e-10	1.48e-08	99.98	Unknown
4.112	326	0.347	1.42e-09	2.09e-08	100.00	Proline, lactic acid
2.312	482	0.179	3.18e-09	4.56e-08	99.99	Lipid (methylene carbonyl)
1.132	600	0.161	6.87e-09	9.65e-08	99.99	Unknown
2.302	483	0.196	7.11e-09	9.78e-08	99.99	Lipid (methylene carbonyl)
2.352	478	0.125	8.00e-09	1.08e-07	99.99	Proline, glutamic acid
1.702	543	0.114	9.27e-09	1.22e-07	99.99	Unknown, arginine
2.342	479	0.149	1.40e-08	1.82e-07	99.99	Proline, glutamic acid
2.322	481	0.158	2.05e-08	2.60e-07	99.98	Lipid (methylene carbonyl)
1.072	606	0.160	2.34e-08	2.92e-07	99.98	Valine
2.362	477	0.149	2.51e-08	3.07e-07	99.99	Proline, glutamic acid
1.082	605	0.165	4.33e-08	5.19e-07	99.98	Unknown
2.332	480	0.156	4.99e-08	5.89e-07	99.98	Proline, glutamic acid
1.692	544	0.122	5.39e-08	6.18e-07	99.98	Unknown, arginine
1.002	613	0.158	5.43e-08	6.18e-07	99.97	Valine, lipid methyl, cholesterol (ester)
1.232	590	0.171	5.58e-08	6.24e-07	99.97	Lipid methylene
2.382	475	0.153	9.57e-08	1.05e-06	99.98	Proline, glutamic acid
7.372	211	0.124	1.20e-07	1.30e-06	99.93	Phenylalanine
1.712	542	0.095	1.28e-07	1.36e-06	99.93	Leucine, lysine
1.682	545	0.136	1.48e-07	1.55e-06	99.95	Unknown, arginine
4.122	325	0.217	1.54e-07	1.59e-06	99.95	Proline, lactic acid
2.122	501	0.138	1.94e-07	1.97e-06	99.93	Lipid allylic

1.222	591	0.203	2.55e-07	2.55e-06	99.98	Lipid methylene
2.292	484	0.206	3.08e-07	3.03e-06	99.90	Lipid (methylene carbonyl)
7.332	215	0.115	3.68e-07	3.57e-06	99.93	Phenylalanine
1.012	612	0.166	3.87e-07	3.70e-06	99.90	Valine, lipid methyl, cholesterol (ester)
7.322	216	0.123	4.74e-07	4.47e-06	99.93	Unknown
4.092	328	0.209	4.94e-07	4.59e-06	99.95	Unknown
4.302	307	0.188	5.17e-07	4.74e-06	99.95	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
4.192	318	0.251	5.32e-07	4.81e-06	99.97	Unknown
3.982	339	0.103	5.81e-07	5.15e-06	99.86	Unknown
2.372	476	0.144	5.85e-07	5.15e-06	99.93	Proline, glutamic acid
4.312	306	0.169	6.37e-07	5.53e-06	99.93	Lipid alpha-methylene to carboxyl, lipid glycerine
4.282	309	0.148	6.69e-07	5.73e-06	99.93	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.152	598	0.169	7.16e-07	6.06e-06	99.95	Lipid methylene
1.142	599	0.146	1.04e-06	8.71e-06	99.81	Unknown
4.052	332	0.217	1.07e-06	8.82e-06	99.50	Unknown
1.962	517	0.109	1.15e-06	9.37e-06	99.86	Lipid allylic
1.212	592	0.221	1.52e-06	1.22e-05	99.97	Lipid methylene
7.242	224	0.114	1.74e-06	1.38e-05	99.86	Unknown
2.412	472	0.149	1.76e-06	1.38e-05	99.86	Glutamine, carnitine
7.262	222	0.124	1.79e-06	1.39e-05	99.86	Unknown
1.812	532	0.096	1.84e-06	1.42e-05	99.81	Unknown
7.122	236	0.115	1.89e-06	1.43e-05	99.73	Unknown
7.252	223	0.123	1.95e-06	1.45e-05	99.81	Unknown
7.312	217	0.117	1.95e-06	1.45e-05	99.81	Unknown
7.232	225	0.112	2.03e-06	1.49e-05	99.86	Unknown
4.162	321	0.163	2.17e-06	1.58e-05	99.81	Proline, lactic acid
1.102	603	0.148	2.44e-06	1.75e-05	99.63	Unknown
2.282	485	0.236	2.47e-06	1.75e-05	99.73	Lipid (methylene carbonyl)
7.292	219	0.106	2.56e-06	1.78e-05	99.81	Unknown
1.032	610	0.172	2.56e-06	1.78e-05	99.63	L-isoleucine, lipid methyl, cholesterol (ester)
1.722	541	0.084	2.69e-06	1.85e-05	99.50	Leucine, lysine
6.842	264	0.125	2.80e-06	1.90e-05	99.81	Unknown
4.292	308	0.170	2.86e-06	1.93e-05	99.81	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.972	516	0.120	2.92e-06	1.94e-05	99.73	Lipid allylic
1.092	604	0.148	2.99e-06	1.97e-05	99.63	Unknown
1.122	601	0.164	3.32e-06	2.16e-05	99.73	Unknown
1.672	546	0.144	3.33e-06	2.16e-05	99.63	Unknown, arginine
2.652	448	0.073	3.40e-06	2.18e-05	99.73	Unknown
3.902	347	0.082	3.49e-06	2.21e-05	99.33	D-glucose, unknown

4.072	330	0.185	3.52e-06	2.21e-05	99.81	Creatinine
6.822	266	0.125	3.68e-06	2.29e-05	99.73	Unknown
7.302	218	0.107	3.93e-06	2.42e-05	99.73	Unknown
2.112	502	0.128	4.16e-06	2.54e-05	99.63	Lipid allylic
4.032	334	0.124	4.31e-06	2.61e-05	99.63	Unknown
4.102	327	0.261	5.55e-06	3.33e-05	99.81	Unknown
1.662	547	0.165	6.00e-06	3.57e-05	99.50	Lipids (?)
7.272	221	0.112	6.13e-06	3.61e-05	99.63	Unknown
6.832	265	0.119	6.62e-06	3.86e-05	99.63	Unknown
1.482	565	0.130	7.17e-06	4.15e-05	99.50	Alanine
4.062	331	0.197	7.63e-06	4.38e-05	99.50	Creatinine
1.872	526	0.085	8.52e-06	4.85e-05	99.12	Overlap of multiple minor compounds
2.102	503	0.107	8.77e-06	4.95e-05	99.33	Lipid allylic
2.422	471	0.130	8.87e-06	4.96e-05	99.63	Glutamine, carnitine
1.062	607	0.143	9.56e-06	5.30e-05	99.12	Valine
2.232	490	0.262	1.06e-05	5.81e-05	99.33	Lipid (methylene carbonyl)
1.892	524	0.081	1.10e-05	6.01e-05	98.84	Overlap of multiple minor compounds
7.012	247	0.188	1.11e-05	6.02e-05	99.12	Unknown
4.152	322	0.153	1.16e-05	6.22e-05	99.33	Proline, lactic acid
7.282	220	0.101	1.19e-05	6.31e-05	99.50	Unknown
1.052	608	0.201	1.30e-05	6.86e-05	99.33	Valine
1.902	523	0.078	1.32e-05	6.89e-05	98.84	Overlap of multiple minor compounds
2.222	491	0.249	1.36e-05	7.06e-05	99.12	Lipid (methylene carbonyl)
0.982	615	0.141	1.39e-05	7.18e-05	98.84	Leucine, lipid methyl, cholesterol (ester)
1.882	525	0.080	1.44e-05	7.35e-05	98.48	Overlap of multiple minor compounds
7.382	210	0.092	1.45e-05	7.35e-05	98.84	Phenylalanine
6.852	263	0.117	1.56e-05	7.82e-05	99.33	Unknown
1.022	611	0.145	1.57e-05	7.82e-05	98.84	L-isoleucine, lipid methyl, cholesterol (ester)
1.982	515	0.125	1.59e-05	7.82e-05	99.12	Lipid allylic
2.242	489	0.274	1.59e-05	7.82e-05	99.33	Lipid (methylene carbonyl), acetone
2.252	488	0.246	1.66e-05	8.12e-05	99.50	Lipid (methylene carbonyl), acetone
0.972	616	0.139	1.78e-05	8.58e-05	98.84	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
4.272	310	0.118	1.78e-05	8.58e-05	99.33	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.842	529	0.084	1.96e-05	9.38e-05	98.84	Unknown
2.272	486	0.274	1.99e-05	9.39e-05	99.12	Lipid (methylene carbonyl)

1.162	597	0.183	1.99e-05	9.39e-05	99.86	Lipid methylene
1.172	596	0.171	2.07e-05	9.68e-05	99.81	Lipid methylene
1.862	527	0.084	2.14e-05	9.96e-05	98.84	Unknown
1.242	589	0.148	2.17e-05	1.00e-04	98.84	Lipid methylene
1.462	567	0.142	2.21e-05	1.01e-04	98.84	Lipid methylene
0.992	614	0.141	2.28e-05	1.04e-04	98.48	Leucine, lipid methyl, cholesterol (ester)
1.452	568	0.160	2.29e-05	1.04e-04	98.84	Lipid methylene
1.472	566	0.133	2.50e-05	1.12e-04	98.84	Lipid methylene
7.142	234	0.132	2.74e-05	1.22e-04	98.84	Unknown
1.182	595	0.142	2.96e-05	1.31e-04	98.04	Lipid methylene
1.442	569	0.164	3.11e-05	1.37e-04	98.48	Lipid methylene
2.132	500	0.083	3.45e-05	1.51e-04	98.48	Glutamine
2.212	492	0.213	3.71e-05	1.61e-04	98.48	Lipid (methylene carbonyl)
1.822	531	0.079	3.73e-05	1.61e-04	99.12	Unknown
1.852	528	0.083	4.04e-05	1.73e-04	98.04	Unknown
1.432	570	0.176	4.11e-05	1.75e-04	98.04	Lipid methylene
1.952	518	0.083	4.20e-05	1.78e-04	98.48	Acetic acid
2.642	449	0.074	4.38e-05	1.84e-04	98.48	Unknown
3.392	385	0.299	4.64e-05	1.94e-04	98.04	Methanol, proline
2.262	487	0.295	5.39e-05	2.24e-04	98.48	Lipid (methylene carbonyl)
7.052	243	0.124	5.52e-05	2.28e-04	97.50	Unknown
1.352	578	0.210	5.92e-05	2.43e-04	98.04	Lipid methylene, lactic acid, threonine
1.832	530	0.074	6.01e-05	2.45e-04	98.04	Unknown
4.342	303	0.117	6.73e-05	2.73e-04	98.84	Lipid alpha-methylene to carboxyl, lipid glycerine
7.042	244	0.160	7.12e-05	2.86e-04	96.02	Unknown
0.962	617	0.141	7.17e-05	2.87e-04	97.50	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
7.132	235	0.119	7.94e-05	3.16e-04	97.50	Unknown
1.422	571	0.210	8.19e-05	3.22e-04	97.50	Lipid methylene
8.152	133	0.311	8.20e-05	3.22e-04	98.04	Unknown
4.322	305	0.119	8.79e-05	3.43e-04	98.04	Lipid alpha-methylene to carboxyl, lipid glycerine
1.992	514	0.119	8.82e-05	3.43e-04	97.50	Lipid allylic
1.732	540	0.075	8.99e-05	3.47e-04	96.83	Leucine, lysine
1.412	572	0.257	9.57e-05	3.67e-04	97.50	Lipid methylene
2.082	505	0.146	9.81e-05	3.74e-04	97.50	Lipid allylic
1.582	555	0.279	1.03e-04	3.90e-04	98.04	Lipids (?)
1.342	579	0.199	1.12e-04	4.24e-04	97.50	Lipid methylene, lactic acid, threonine
2.202	493	0.158	1.15e-04	4.30e-04	96.83	Lipid (methylene carbonyl)
3.952	342	0.124	1.19e-04	4.43e-04	96.02	Unknown
2.002	513	0.119	1.21e-04	4.48e-04	96.83	Lipid allylic

1.802	533	0.085	1.30e-04	4.80e-04	98.04	Unknown
4.142	323	0.168	1.35e-04	4.95e-04	96.83	Proline, lactic acid
2.432	470	0.077	1.37e-04	4.98e-04	97.50	Glutamine, carnitine
1.572	556	0.248	1.45e-04	5.27e-04	97.50	Lipids (?)
1.652	548	0.169	1.46e-04	5.27e-04	96.83	Lipids (?)
1.592	554	0.290	1.62e-04	5.83e-04	97.50	Lipids (?)
7.192	229	0.090	1.81e-04	6.45e-04	96.83	Tyrosine
1.742	539	0.074	2.07e-04	7.33e-04	96.02	Leucine, lysine
1.402	573	0.302	2.20e-04	7.75e-04	96.02	Lipid methylene
1.752	538	0.079	2.25e-04	7.89e-04	95.05	Leucine, lysine
1.602	553	0.289	2.35e-04	8.20e-04	96.83	Lipids (?)
4.182	319	0.249	2.49e-04	8.64e-04	98.04	Unknown
3.062	407	0.096	2.55e-04	8.77e-04	93.91	Creatinine
1.382	575	0.321	2.55e-04	8.77e-04	96.02	Lipid methylene
4.132	324	0.173	2.58e-04	8.83e-04	96.02	Proline, lactic acid
7.112	237	0.077	2.66e-04	9.06e-04	95.05	Unknown
2.852	428	0.187	2.72e-04	9.22e-04	96.02	Lipid diallylic
1.502	563	0.112	2.80e-04	9.44e-04	96.02	Alanine
1.612	552	0.275	2.94e-04	9.85e-04	96.02	Lipids (?)
0.952	618	0.151	3.05e-04	1.02e-03	95.05	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
2.012	512	0.118	3.56e-04	1.18e-03	95.05	Lipid allylic
1.392	574	0.310	3.61e-04	1.19e-03	95.05	Lipid methylene
1.762	537	0.076	3.72e-04	1.22e-03	93.91	Leucine, lysine
1.642	549	0.190	3.87e-04	1.27e-03	93.91	Lipids (?)
1.562	557	0.210	4.20e-04	1.37e-03	95.05	Lipids (?)
1.622	551	0.247	4.55e-04	1.47e-03	95.05	Lipids (?)
8.142	134	0.322	4.64e-04	1.49e-03	87.17	Unknown
4.352	302	0.108	4.68e-04	1.50e-03	96.02	Lipid alpha-methylene to carboxyl, lipid glycerine
1.292	584	0.200	4.97e-04	1.58e-03	95.05	Lipid methylene
7.342	214	0.065	5.46e-04	1.73e-03	93.91	Phenylalanine
1.632	550	0.213	5.68e-04	1.79e-03	93.91	Lipids (?)
7.022	246	0.207	5.87e-04	1.84e-03	92.56	Unknown
7.062	242	0.114	6.10e-04	1.91e-03	91.00	Unknown
1.362	577	0.192	6.21e-04	1.93e-03	93.91	Lipid methylene, lactic acid, threonine
4.172	320	0.182	6.22e-04	1.93e-03	95.05	Unknown
1.522	561	0.088	6.43e-04	1.98e-03	92.56	Lipids (?)
2.072	506	0.105	6.80e-04	2.08e-03	92.56	Lipid allylic
2.482	465	-0.075	6.80e-04	2.08e-03	89.21	Glutamine, carnitine
1.532	560	0.109	7.04e-04	2.14e-03	92.56	Lipids (?)
1.372	576	0.295	7.40e-04	2.24e-03	92.56	Lipid methylene
1.282	585	0.165	7.57e-04	2.28e-03	92.56	Lipid methylene
1.302	583	0.232	7.74e-04	2.32e-03	93.91	Lipid methylene

2.022	511	0.126	8.07e-04	2.41e-03	92.56	Lipid allylic
1.552	558	0.174	8.50e-04	2.52e-03	92.56	Lipids (?)
0.942	619	0.175	8.53e-04	2.52e-03	91.00	Cholesterol, lipid methyl
4.332	304	0.098	8.62e-04	2.54e-03	93.91	Lipid alpha-methylene to carboxyl, lipid glycerine
2.192	494	0.107	9.17e-04	2.69e-03	91.00	Lipid (methylene carbonyl)
2.032	510	0.139	1.05e-03	3.06e-03	92.56	Lipid allylic
1.312	582	0.239	1.07e-03	3.11e-03	92.56	Lipid methylene
1.772	536	0.074	1.14e-03	3.31e-03	91.00	Leucine, lysine
1.542	559	0.136	1.16e-03	3.33e-03	91.00	Lipids (?)
2.632	450	0.053	1.23e-03	3.53e-03	91.00	Unknown
1.322	581	0.247	1.29e-03	3.69e-03	91.00	Lipid methylene
2.052	508	0.113	1.33e-03	3.79e-03	91.00	Lipid allylic
4.082	329	0.134	1.36e-03	3.84e-03	92.56	Creatinine
1.332	580	0.256	1.45e-03	4.08e-03	91.00	Lipid methylene
6.812	267	0.101	1.46e-03	4.08e-03	93.91	Unknown
1.792	534	0.085	1.46e-03	4.08e-03	92.56	Unknown
2.862	427	0.154	1.50e-03	4.18e-03	89.21	Lipid diallylic
1.252	588	0.123	1.51e-03	4.18e-03	87.17	Lipid methylene
1.912	522	0.064	1.63e-03	4.51e-03	89.21	Overlap of multiple minor compounds
6.862	262	0.099	1.67e-03	4.58e-03	87.17	Unknown
2.622	451	0.046	1.67e-03	4.58e-03	91.00	Unknown
0.842	629	0.080	1.68e-03	4.59e-03	87.17	Cholesterol, lipid methyl
8.402	108	0.407	1.79e-03	4.86e-03	87.17	Unknown
1.492	564	0.093	1.84e-03	4.98e-03	91.00	Alanine
1.112	602	0.095	1.85e-03	4.98e-03	84.88	Unknown
2.042	509	0.138	2.11e-03	5.67e-03	89.21	Lipid allylic
0.832	630	0.087	2.28e-03	6.09e-03	87.17	Cholesterol, lipid methyl
2.062	507	0.067	2.32e-03	6.18e-03	87.17	Lipid allylic
3.032	410	0.032	2.65e-03	7.01e-03	79.53	Lysine, unknown
3.372	387	-0.091	2.65e-03	7.01e-03	79.53	Methanol, proline
1.272	586	0.135	2.73e-03	7.17e-03	84.88	Lipid methylene
7.072	241	0.089	2.76e-03	7.24e-03	84.88	Unknown
0.932	620	0.189	2.95e-03	7.69e-03	84.88	Cholesterol, lipid methyl
2.392	474	0.091	3.11e-03	8.09e-03	84.88	Unknown
7.152	233	0.115	3.21e-03	8.32e-03	84.88	Unknown
8.352	113	0.365	3.46e-03	8.92e-03	89.21	Unknown
7.182	230	0.087	3.78e-03	9.70e-03	82.33	Unknown
2.822	431	0.154	4.08e-03	1.04e-02	82.33	Lipid diallylic
2.832	430	0.151	4.31e-03	1.10e-02	82.33	Lipid diallylic
2.842	429	0.155	4.49e-03	1.14e-02	82.33	Lipid diallylic
6.872	261	0.076	5.01e-03	1.27e-02	82.33	Unknown
2.562	457	0.041	5.12e-03	1.29e-02	87.17	$CaEDTA^{2-}$
1.782	535	0.078	5.74e-03	1.44e-02	82.33	Unknown

2.872	426	0.107	5.75e-03	1.44e-02	76.48	Lipid diallylic
8.022	146	0.169	5.84e-03	1.45e-02	76.48	Unknown
8.412	107	0.408	5.85e-03	1.45e-02	76.48	Unknown
6.882	260	0.070	5.97e-03	1.47e-02	79.53	Unknown
3.312	393	0.101	5.98e-03	1.47e-02	76.48	Unknown
1.042	609	0.121	6.07e-03	1.49e-02	76.48	L-isoleucine, lipid methyl, cholesterol (ester)
1.262	587	0.114	6.89e-03	1.68e-02	76.48	Lipid methylene
2.882	425	0.084	6.97e-03	1.70e-02	73.19	Lipid diallylic
9.342	14	0.416	8.00e-03	1.94e-02	73.19	Unknown
1.202	593	0.197	8.34e-03	2.02e-02	82.33	Lipid methylene
6.732	275	-0.317	9.86e-03	2.37e-02	73.19	Unknown
0.802	633	0.105	1.03e-02	2.46e-02	73.19	Cholesterol, lipid methyl
0.922	621	0.174	1.07e-02	2.55e-02	73.19	Cholesterol, lipid methyl
1.192	594	0.161	1.08e-02	2.57e-02	79.53	Lipid methylene
7.962	152	0.276	1.08e-02	2.57e-02	73.19	Unknown
6.802	268	0.085	1.11e-02	2.63e-02	79.53	Unknown
8.392	109	0.321	1.12e-02	2.64e-02	76.48	Unknown
8.322	116	0.289	1.17e-02	2.75e-02	76.48	Unknown
2.492	464	-0.055	1.19e-02	2.77e-02	69.69	Glutamine
7.972	151	0.275	1.19e-02	2.77e-02	69.69	Unknown
2.092	504	0.097	1.23e-02	2.86e-02	69.69	Lipid allylic
8.582	90	-0.356	1.27e-02	2.95e-02	66.00	Unknown
7.352	213	0.050	1.32e-02	3.05e-02	69.69	Phenylalanine
1.512	562	0.078	1.39e-02	3.19e-02	73.19	Alanine
2.812	432	0.132	1.40e-02	3.22e-02	69.69	Lipid diallylic
3.012	412	0.034	1.41e-02	3.23e-02	69.69	Lysine, unknown
7.422	206	0.111	1.42e-02	3.23e-02	69.69	Phenylalanine
4.372	300	0.099	1.43e-02	3.25e-02	73.19	Unknown
0.822	631	0.086	1.51e-02	3.42e-02	66.00	Cholesterol, lipid methyl
0.812	632	0.098	1.65e-02	3.71e-02	66.00	Cholesterol, lipid methyl
2.182	495	0.068	1.67e-02	3.74e-02	66.00	Glutamine
8.202	128	0.145	1.75e-02	3.92e-02	66.00	Unknown
1.922	521	0.051	1.76e-02	3.92e-02	66.00	Overlap of multiple minor compounds
8.592	89	-0.434	1.79e-02	3.97e-02	37.86	Unknown
7.362	212	0.057	1.79e-02	3.97e-02	66.00	Phenylalanine
8.172	131	0.167	1.80e-02	3.98e-02	62.14	Unknown
9.452	3	0.359	1.83e-02	4.02e-02	69.69	Unknown
8.262	122	0.089	1.90e-02	4.16e-02	62.14	Unknown
0.792	634	0.086	1.92e-02	4.20e-02	66.00	Cholesterol, lipid methyl
4.252	312	0.103	1.97e-02	4.29e-02	66.00	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
8.002	148	0.221	2.13e-02	4.63e-02	58.17	Unknown
6.892	259	0.047	2.18e-02	4.71e-02	66.00	Unknown

7.902	158	0.111	2.21e-02	4.76e-02	62.14	Unknown
4.262	311	0.081	2.22e-02	4.78e-02	66.00	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
7.612	187	0.087	2.24e-02	4.81e-02	54.10	Unknown
3.022	411	0.028	2.27e-02	4.84e-02	62.14	Lysine, unknown
2.802	433	0.106	2.32e-02	4.94e-02	66.00	Lipid diallylic

Table 7.10: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from diabetic nephropathy from those suffering from interstitial nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
3.782	359	0.446	9.04e-76	5.97e-73	100.00	D-glucose, alanine, glutamine, arginine
3.562	368	0.586	3.09e-75	8.09e-73	100.00	D-glucose
3.552	369	0.605	3.68e-75	8.09e-73	100.00	D-glucose, myo-inositol
3.802	357	0.488	1.18e-74	1.95e-72	100.00	D-glucose, alanine
3.842	353	0.557	1.81e-74	2.38e-72	100.00	D-glucose, unknown
3.852	352	0.581	1.08e-73	1.19e-71	100.00	D-glucose, unknown
3.872	350	0.487	7.48e-73	7.06e-71	100.00	D-glucose, unknown
3.752	362	0.590	4.41e-72	3.64e-70	100.00	D-glucose, glutamic acid
3.452	379	0.556	1.31e-71	9.64e-70	100.00	D-glucose, carnitine, proline
3.572	367	0.540	1.34e-70	8.85e-69	100.00	D-glucose, glycine
3.912	346	0.537	2.21e-70	1.30e-68	100.00	D-glucose, betaine, unknown
3.862	351	0.589	2.36e-70	1.30e-68	100.00	D-glucose, unknown
3.742	363	0.587	2.75e-70	1.39e-68	100.00	D-glucose, leucine
3.442	380	0.612	4.05e-70	1.91e-68	100.00	D-glucose, carnitine, taurine, proline
3.932	344	0.551	4.90e-70	2.16e-68	100.00	D-glucose
3.472	377	0.612	5.73e-70	2.36e-68	100.00	D-glucose
3.422	382	0.637	1.39e-69	5.40e-68	100.00	D-glucose, carnitine, taurine, proline
3.512	373	0.635	2.21e-69	8.09e-68	100.00	D-glucose
3.482	376	0.621	3.12e-69	1.09e-67	100.00	D-glucose
3.412	383	0.615	3.70e-69	1.22e-67	100.00	D-glucose, carnitine, taurine, proline

3.432	381	0.617	4.37e-69	1.37e-67	100.00	D-glucose, carnitine, taurine, proline
3.722	365	0.496	5.92e-69	1.78e-67	100.00	D-glucose, N,N-dimethylglycine
3.882	349	0.529	1.11e-68	3.20e-67	100.00	D-glucose, unknown
3.502	374	0.629	2.45e-68	6.73e-67	100.00	D-glucose
3.762	361	0.532	7.53e-67	1.99e-65	100.00	D-glucose, arginine, glutamine, glutamic acid
3.732	364	0.540	1.27e-65	3.22e-64	100.00	D-glucose, unknown
3.772	360	0.490	4.07e-65	9.96e-64	100.00	D-glucose, alanine, glutamine, arginine
3.492	375	0.606	2.01e-64	4.75e-63	100.00	D-glucose
3.532	371	0.538	2.15e-63	4.89e-62	100.00	D-glucose
3.792	358	0.310	1.13e-62	2.48e-61	100.00	D-glucose, alanine
3.542	370	0.482	2.78e-52	5.91e-51	100.00	D-glucose, myo-inositol
3.812	356	0.398	1.51e-51	3.11e-50	100.00	D-glucose
3.522	372	0.553	4.07e-45	8.14e-44	100.00	D-glucose
3.942	343	0.486	4.30e-45	8.34e-44	100.00	D-glucose
3.832	354	0.255	4.51e-38	8.50e-37	100.00	Unknown
3.402	384	0.364	1.82e-36	3.34e-35	100.00	Unknown
3.922	345	0.300	1.87e-35	3.34e-34	100.00	D-glucose, unknown
3.822	355	0.245	1.92e-27	3.33e-26	100.00	Unknown
3.962	341	0.155	5.55e-27	9.39e-26	100.00	Unknown
3.462	378	0.282	4.50e-25	7.43e-24	100.00	D-glucose
3.972	340	0.138	3.68e-21	5.93e-20	100.00	Unknown
1.902	523	0.142	1.32e-20	2.07e-19	100.00	Overlap of multiple minor compounds
1.082	605	0.232	1.86e-19	2.85e-18	100.00	Unknown
1.072	606	0.220	1.94e-19	2.91e-18	100.00	Valine
1.002	613	0.215	3.50e-18	5.13e-17	100.00	Valine, lipid methyl, cholesterol (ester)
1.892	524	0.132	2.56e-17	3.68e-16	100.00	Overlap of multiple minor compounds
3.892	348	0.138	9.75e-17	1.37e-15	100.00	Unknown
1.012	612	0.229	2.34e-16	3.21e-15	100.00	Valine, lipid methyl, cholesterol (ester)
1.092	604	0.221	2.43e-16	3.27e-15	100.00	Unknown
1.882	525	0.127	4.85e-16	6.41e-15	100.00	Overlap of multiple minor compounds
1.102	603	0.212	2.31e-15	3.00e-14	100.00	Unknown
2.362	477	0.178	5.31e-15	6.74e-14	100.00	Proline, glutamic acid
2.382	475	0.189	9.36e-15	1.17e-13	100.00	Proline, glutamic acid
1.132	600	0.182	1.23e-14	1.50e-13	100.00	Unknown
1.912	522	0.133	1.78e-14	2.14e-13	100.00	Overlap of multiple minor compounds
2.372	476	0.186	2.56e-14	3.01e-13	100.00	Proline, glutamic acid

3.062	407	0.171	2.72e-14	3.15e-13	100.00	Creatinine
1.232	590	0.203	3.13e-14	3.56e-13	100.00	Lipid methylene
1.702	543	0.126	5.81e-14	6.50e-13	100.00	Unknown, arginine
4.032	334	0.172	8.66e-14	9.52e-13	100.00	Unknown
1.062	607	0.204	1.24e-13	1.34e-12	100.00	Valine
2.312	482	0.189	2.09e-13	2.22e-12	100.00	Lipid (methylene carbonyl)
4.092	328	0.257	3.43e-13	3.59e-12	100.00	Unknown
2.302	483	0.208	4.68e-13	4.83e-12	100.00	Lipid (methylene carbonyl)
2.352	478	0.132	5.97e-13	6.06e-12	100.00	Proline, glutamic acid
1.242	589	0.209	1.74e-12	1.74e-11	100.00	Lipid methylene
1.052	608	0.274	2.41e-12	2.37e-11	100.00	Valine
1.682	545	0.154	2.91e-12	2.83e-11	100.00	Unknown, arginine
1.692	544	0.133	3.07e-12	2.94e-11	100.00	Unknown, arginine
7.372	211	0.138	3.66e-12	3.45e-11	100.00	Phenylalanine
1.722	541	0.105	3.72e-12	3.46e-11	100.00	Leucine, lysine
6.832	265	0.156	4.40e-12	4.04e-11	100.00	Unknown
4.072	330	0.235	4.77e-12	4.32e-11	100.00	Creatinine
1.922	521	0.126	4.98e-12	4.45e-11	100.00	Overlap of multiple minor compounds
1.022	611	0.196	6.33e-12	5.57e-11	100.00	L-isoleucine, lipid methyl, cholesterol (ester)
1.742	539	0.117	7.84e-12	6.81e-11	100.00	Leucine, lysine
1.032	610	0.212	8.00e-12	6.86e-11	100.00	L-isoleucine, lipid methyl, cholesterol (ester)
1.732	540	0.111	8.90e-12	7.47e-11	100.00	Leucine, lysine
1.482	565	0.167	8.94e-12	7.47e-11	100.00	Alanine
6.842	264	0.155	9.58e-12	7.91e-11	100.00	Unknown
2.122	501	0.154	1.02e-11	8.28e-11	100.00	Lipid allylic
1.752	538	0.122	1.95e-11	1.57e-10	100.00	Leucine, lysine
1.932	520	0.119	2.02e-11	1.60e-10	100.00	Acetic acid
2.292	484	0.227	2.67e-11	2.10e-10	100.00	Lipid (methylene carbonyl)
0.982	615	0.184	2.78e-11	2.16e-10	100.00	Leucine, lipid methyl, cholesterol (ester)
1.712	542	0.101	3.33e-11	2.55e-10	100.00	Leucine, lysine
3.712	366	0.149	3.36e-11	2.55e-10	100.00	Unknown
1.142	599	0.168	4.10e-11	3.07e-10	100.00	Unknown
1.502	563	0.174	4.23e-11	3.13e-10	100.00	Alanine
1.872	526	0.107	5.59e-11	4.10e-10	100.00	Overlap of multiple minor compounds
2.322	481	0.155	7.94e-11	5.76e-10	100.00	Lipid (methylene carbonyl)
0.992	614	0.183	1.06e-10	7.64e-10	100.00	Leucine, lipid methyl, cholesterol (ester)
4.052	332	0.243	1.28e-10	9.05e-10	100.00	Unknown
1.122	601	0.193	1.29e-10	9.05e-10	100.00	Unknown
4.112	326	0.309	2.25e-10	1.56e-09	100.00	Proline, lactic acid

0.972	616	0.174	2.36e-10	1.62e-09	100.00	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.162	597	0.231	2.61e-10	1.77e-09	100.00	Lipid methylene
1.152	598	0.182	3.20e-10	2.16e-09	100.00	Lipid methylene
1.172	596	0.213	4.18e-10	2.79e-09	100.00	Lipid methylene
6.822	266	0.142	4.60e-10	3.04e-09	100.00	Unknown
1.492	564	0.153	1.66e-09	1.08e-08	100.00	Alanine
1.972	516	0.131	1.73e-09	1.12e-08	100.00	Lipid allylic
2.282	485	0.256	1.78e-09	1.14e-08	100.00	Lipid (methylene carbonyl)
1.472	566	0.160	1.86e-09	1.18e-08	100.00	Lipid methylene
7.382	210	0.107	2.30e-09	1.45e-08	100.00	Phenylalanine
1.982	515	0.146	2.53e-09	1.58e-08	100.00	Lipid allylic
2.652	448	0.080	2.56e-09	1.58e-08	100.00	Unknown
7.192	229	0.121	2.88e-09	1.76e-08	100.00	Tyrosine
4.102	327	0.289	3.13e-09	1.89e-08	100.00	Unknown
1.182	595	0.171	3.16e-09	1.90e-08	99.99	Lipid methylene
7.252	223	0.128	6.59e-09	3.92e-08	99.99	Unknown
4.162	321	0.169	7.88e-09	4.64e-08	99.99	Proline, lactic acid
7.322	216	0.119	8.49e-09	4.96e-08	100.00	Unknown
7.262	222	0.127	8.61e-09	4.98e-08	99.99	Unknown
1.222	591	0.192	8.88e-09	5.09e-08	100.00	Lipid methylene
7.332	215	0.110	9.73e-09	5.53e-08	99.99	Phenylalanine
0.962	617	0.172	1.19e-08	6.69e-08	99.99	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
3.372	387	-0.146	1.22e-08	6.84e-08	99.97	Methanol, proline
1.462	567	0.162	1.34e-08	7.45e-08	99.99	Lipid methylene
7.232	225	0.113	1.59e-08	8.73e-08	99.99	Unknown
2.132	500	0.096	1.66e-08	9.06e-08	99.99	Glutamine
1.962	517	0.107	1.70e-08	9.21e-08	99.99	Lipid allylic
4.312	306	0.161	1.97e-08	1.06e-07	99.99	Lipid alpha-methylene to carboxyl, lipid glycerine
7.182	230	0.143	2.17e-08	1.15e-07	99.99	Unknown
2.402	473	0.082	2.17e-08	1.15e-07	99.97	Glutamine, carnitine
1.452	568	0.179	2.67e-08	1.40e-07	99.99	Lipid methylene
7.272	221	0.117	2.76e-08	1.43e-07	99.99	Unknown
1.992	514	0.143	3.10e-08	1.60e-07	99.99	Lipid allylic
2.342	479	0.123	3.49e-08	1.79e-07	99.98	Proline, glutamic acid
2.112	502	0.129	4.04e-08	2.05e-07	99.98	Lipid allylic
1.252	588	0.180	4.87e-08	2.45e-07	99.97	Lipid methylene
2.332	480	0.132	5.30e-08	2.65e-07	99.97	Proline, glutamic acid
2.552	458	0.102	5.69e-08	2.82e-07	100.00	Citric acid
2.212	492	0.235	8.53e-08	4.19e-07	99.97	Lipid (methylene carbonyl)
7.282	220	0.104	8.58e-08	4.19e-07	99.98	Unknown
2.202	493	0.186	8.90e-08	4.32e-07	99.95	Lipid (methylene carbonyl)
7.242	224	0.108	9.84e-08	4.74e-07	99.97	Unknown

7.312	217	0.111	1.10e-07	5.25e-07	99.98	Unknown
1.672	546	0.140	1.11e-07	5.25e-07	99.97	Unknown, arginine
2.252	488	0.256	1.35e-07	6.37e-07	99.95	Lipid (methylene carbonyl), acetone
7.292	219	0.101	1.36e-07	6.37e-07	99.97	Unknown
2.392	474	0.137	1.54e-07	7.17e-07	99.92	Unknown
1.952	518	0.090	1.61e-07	7.44e-07	99.97	Acetic acid
1.762	537	0.095	1.86e-07	8.50e-07	99.95	Leucine, lysine
7.302	218	0.103	1.91e-07	8.71e-07	99.97	Unknown
1.442	569	0.174	2.04e-07	9.21e-07	99.95	Lipid methylene
2.002	513	0.135	3.15e-07	1.42e-06	99.92	Lipid allylic
2.222	491	0.247	3.56e-07	1.59e-06	99.92	Lipid (methylene carbonyl)
0.842	629	0.109	4.67e-07	2.07e-06	99.82	Cholesterol, lipid methyl
1.432	570	0.183	4.87e-07	2.14e-06	99.88	Lipid methylene
0.952	618	0.177	5.76e-07	2.52e-06	99.88	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
7.342	214	0.079	5.82e-07	2.53e-06	99.82	Phenylalanine
1.522	561	0.109	5.96e-07	2.57e-06	99.92	Lipids (?)
7.032	245	0.188	6.08e-07	2.60e-06	99.82	Unknown
2.102	503	0.102	6.10e-07	2.60e-06	99.88	Lipid allylic
2.672	446	0.122	6.34e-07	2.68e-06	99.98	Citric acid
1.422	571	0.225	6.48e-07	2.73e-06	99.88	Lipid methylene
4.152	322	0.147	7.13e-07	2.98e-06	99.73	Proline, lactic acid
4.122	325	0.174	7.39e-07	3.07e-06	99.73	Proline, lactic acid
2.872	426	0.163	7.53e-07	3.11e-06	99.88	Lipid diallylic
1.412	572	0.276	7.64e-07	3.13e-06	99.88	Lipid methylene
1.212	592	0.192	7.89e-07	3.21e-06	99.98	Lipid methylene
2.272	486	0.268	8.73e-07	3.53e-06	99.88	Lipid (methylene carbonyl)
4.322	305	0.125	1.08e-06	4.35e-06	99.82	Lipid alpha-methylene to carboxyl, lipid glycerine
1.662	547	0.149	1.32e-06	5.27e-06	99.82	Lipids (?)
1.042	609	0.180	1.68e-06	6.66e-06	99.73	L-isoleucine, lipid methyl, cholesterol (ester)
6.812	267	0.129	1.75e-06	6.92e-06	99.92	Unknown
2.192	494	0.130	1.95e-06	7.67e-06	99.62	Lipid (methylene carbonyl)
7.142	234	0.126	2.10e-06	8.19e-06	99.82	Unknown
1.862	527	0.080	2.11e-06	8.19e-06	99.73	Unknown
4.302	307	0.150	2.34e-06	9.04e-06	99.82	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.402	573	0.323	3.05e-06	1.17e-05	99.73	Lipid methylene
6.852	263	0.106	3.81e-06	1.45e-05	99.62	Unknown
3.952	342	0.126	3.92e-06	1.49e-05	99.23	Unknown
1.512	562	0.124	4.66e-06	1.76e-05	99.45	Alanine
2.232	490	0.231	4.69e-06	1.76e-05	99.62	Lipid (methylene carbonyl)
7.052	243	0.119	4.74e-06	1.77e-05	98.93	Unknown

0.942	619	0.204	4.90e-06	1.82e-05	99.62	Cholesterol, lipid methyl
1.542	559	0.160	6.04e-06	2.23e-05	99.62	Lipids (?)
7.102	238	0.168	6.55e-06	2.40e-05	99.62	Unknown
7.132	235	0.115	7.65e-06	2.79e-05	99.45	Unknown
2.642	449	0.069	7.98e-06	2.89e-05	99.45	Unknown
2.852	428	0.194	8.25e-06	2.97e-05	99.45	Lipid diallylic
4.192	318	0.188	8.92e-06	3.20e-05	99.73	Unknown
4.342	303	0.110	1.00e-05	3.57e-05	99.45	Lipid alpha-methylene to carboxyl, lipid glycerine
7.352	213	0.075	1.24e-05	4.40e-05	98.53	Phenylalanine
1.532	560	0.119	1.25e-05	4.40e-05	99.23	Lipids (?)
1.552	558	0.193	1.28e-05	4.48e-05	99.23	Lipids (?)
4.042	333	0.152	1.52e-05	5.32e-05	99.23	Unknown
1.392	574	0.318	1.65e-05	5.73e-05	99.23	Lipid methylene
1.562	557	0.216	1.96e-05	6.75e-05	99.23	Lipids (?)
2.242	489	0.230	1.96e-05	6.75e-05	98.93	Lipid (methylene carbonyl), acetone
4.082	329	0.151	2.09e-05	7.16e-05	99.23	Creatinine
2.082	505	0.135	2.14e-05	7.27e-05	98.93	Lipid allylic
1.352	578	0.188	2.16e-05	7.32e-05	98.93	Lipid methylene, lactic acid, threonine
6.802	268	0.120	2.35e-05	7.91e-05	99.23	Unknown
2.262	487	0.261	2.36e-05	7.91e-05	98.93	Lipid (methylene carbonyl)
7.202	228	0.111	2.47e-05	8.25e-05	98.93	Tyrosine
3.992	338	-0.073	2.59e-05	8.59e-05	98.01	Unknown
2.862	427	0.173	2.64e-05	8.73e-05	98.93	Lipid diallylic
2.572	456	0.068	3.41e-05	1.12e-04	99.92	$CaEDTA^{2-}$, citric acid
1.262	587	0.147	3.92e-05	1.28e-04	98.01	Lipid methylene
1.852	528	0.070	3.99e-05	1.30e-04	98.01	Unknown
1.382	575	0.305	4.10e-05	1.33e-04	98.53	Lipid methylene
7.172	231	0.153	4.25e-05	1.37e-04	98.53	Unknown
4.332	304	0.101	4.68e-05	1.50e-04	98.53	Lipid alpha-methylene to carboxyl, lipid glycerine
1.572	556	0.225	4.75e-05	1.51e-04	98.53	Lipids (?)
3.032	410	0.037	4.76e-05	1.51e-04	96.51	Lysine, unknown
7.152	233	0.134	4.91e-05	1.55e-04	98.53	Unknown
2.492	464	-0.076	5.14e-05	1.62e-04	97.35	Glutamine
2.012	512	0.113	5.92e-05	1.85e-04	98.01	Lipid allylic
1.342	579	0.175	5.94e-05	1.85e-04	98.01	Lipid methylene, lactic acid, threonine
1.652	548	0.149	7.93e-05	2.46e-04	97.35	Lipids (?)
2.422	471	0.097	8.89e-05	2.74e-04	98.01	Glutamine, carnitine
2.842	429	0.181	9.67e-05	2.97e-04	97.35	Lipid diallylic
4.282	309	0.098	1.01e-04	3.09e-04	97.35	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine

0.852	628	0.083	1.06e-04	3.22e-04	96.51	Cholesterol, lipid methyl
0.932	620	0.208	1.11e-04	3.36e-04	97.35	Cholesterol, lipid methyl
1.622	551	0.230	1.18e-04	3.55e-04	97.35	Lipids (?)
4.142	323	0.143	1.31e-04	3.94e-04	95.46	Proline, lactic acid
1.612	552	0.246	1.33e-04	3.96e-04	97.35	Lipids (?)
1.842	529	0.064	1.37e-04	4.06e-04	96.51	Unknown
1.582	555	0.232	1.41e-04	4.16e-04	97.35	Lipids (?)
4.062	331	0.142	1.41e-04	4.16e-04	97.35	Creatinine
3.392	385	0.236	1.45e-04	4.26e-04	95.46	Methanol, proline
2.832	430	0.169	1.56e-04	4.56e-04	97.35	Lipid diallylic
1.272	586	0.145	1.60e-04	4.66e-04	96.51	Lipid methylene
7.122	236	0.077	1.68e-04	4.85e-04	96.51	Unknown
1.632	550	0.197	1.68e-04	4.85e-04	96.51	Lipids (?)
8.152	133	0.250	1.83e-04	5.25e-04	98.01	Unknown
1.642	549	0.169	1.88e-04	5.38e-04	96.51	Lipids (?)
4.292	308	0.115	1.98e-04	5.64e-04	96.51	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.482	465	-0.069	2.07e-04	5.86e-04	92.62	Glutamine, carnitine
1.602	553	0.246	2.11e-04	5.95e-04	96.51	Lipids (?)
2.822	431	0.167	2.31e-04	6.49e-04	96.51	Lipid diallylic
4.272	310	0.085	2.53e-04	7.08e-04	95.46	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
7.862	162	0.149	2.67e-04	7.43e-04	94.18	Unknown
7.722	176	-0.180	2.70e-04	7.49e-04	94.18	Unknown
7.432	205	0.085	2.72e-04	7.52e-04	95.46	Phenylalanine
1.372	576	0.269	2.83e-04	7.79e-04	95.46	Lipid methylene
3.052	408	0.046	2.92e-04	7.97e-04	94.18	Creatinine
1.592	554	0.236	2.92e-04	7.97e-04	95.46	Lipids (?)
7.392	209	0.065	3.01e-04	8.17e-04	94.18	Phenylalanine
1.362	577	0.172	3.11e-04	8.41e-04	95.46	Lipid methylene, lactic acid, threonine
2.882	425	0.095	3.18e-04	8.56e-04	94.18	Lipid diallylic
2.412	472	0.095	3.23e-04	8.66e-04	95.46	Glutamine, carnitine
0.832	630	0.086	3.62e-04	9.66e-04	94.18	Cholesterol, lipid methyl
1.112	602	0.092	3.89e-04	1.03e-03	94.18	Unknown
4.172	320	0.157	4.67e-04	1.24e-03	95.46	Unknown
8.122	136	0.180	4.85e-04	1.28e-03	92.62	Unknown
6.862	262	0.093	4.98e-04	1.31e-03	92.62	Unknown
3.982	339	0.060	5.59e-04	1.46e-03	92.62	Unknown
6.792	269	0.102	5.79e-04	1.51e-03	94.18	Unknown
1.282	585	0.143	5.87e-04	1.52e-03	94.18	Lipid methylene
3.312	393	0.107	6.02e-04	1.56e-03	90.76	Unknown
3.362	388	0.084	6.15e-04	1.59e-03	90.76	Proline
0.742	639	0.156	6.51e-04	1.67e-03	92.62	Unkown
2.092	504	0.112	6.85e-04	1.75e-03	92.62	Lipid allylic

2.432	470	0.058	8.22e-04	2.09e-03	90.76	Glutamine, carnitine
2.902	423	0.062	8.46e-04	2.15e-03	90.76	Unknown
2.702	443	0.040	8.61e-04	2.18e-03	90.76	$MgEDTA^{2-}$
1.332	580	0.226	8.96e-04	2.26e-03	92.62	Lipid methylene
4.132	324	0.133	8.99e-04	2.26e-03	88.58	Proline, lactic acid
2.662	447	0.098	9.36e-04	2.34e-03	92.62	Citric acid
2.452	468	0.084	9.39e-04	2.34e-03	88.58	Glutamine, carnitine
2.692	444	0.053	9.68e-04	2.40e-03	98.93	Citric acid
2.542	459	0.065	1.00e-03	2.48e-03	92.62	Unknown
7.672	181	0.229	1.12e-03	2.76e-03	88.58	Unknown
2.022	511	0.103	1.19e-03	2.92e-03	90.76	Lipid allylic
3.072	406	0.034	1.21e-03	2.96e-03	86.06	Unknown
3.342	390	0.124	1.24e-03	3.01e-03	90.76	Proline
7.062	242	0.091	1.30e-03	3.14e-03	86.06	Unknown
7.362	212	0.066	1.38e-03	3.34e-03	88.58	Phenylalanine
8.172	131	0.192	1.42e-03	3.41e-03	86.06	Unknown
8.102	138	0.207	1.42e-03	3.41e-03	86.06	Trigonelline
4.352	302	0.083	1.50e-03	3.59e-03	90.76	Lipid alpha-methylene to carboxyl, lipid glycerine
1.322	581	0.205	1.62e-03	3.87e-03	90.76	Lipid methylene
1.812	532	0.054	1.63e-03	3.87e-03	88.58	Unknown
1.312	582	0.195	1.70e-03	4.02e-03	88.58	Lipid methylene
2.632	450	0.044	1.81e-03	4.26e-03	88.58	Unknown
7.112	237	0.055	1.83e-03	4.31e-03	86.06	Unknown
1.302	583	0.180	2.04e-03	4.78e-03	88.58	Lipid methylene
2.032	510	0.110	2.10e-03	4.91e-03	88.58	Lipid allylic
6.542	294	0.388	2.29e-03	5.32e-03	86.06	Unknown
1.942	519	0.107	2.44e-03	5.64e-03	94.18	Acetic acid
4.002	337	-0.046	2.48e-03	5.72e-03	86.06	Unknown
0.752	638	0.102	2.59e-03	5.95e-03	83.18	Cholesterol, lipid methyl
1.822	531	0.049	2.81e-03	6.43e-03	86.06	Unknown
1.192	594	0.159	2.90e-03	6.61e-03	90.76	Lipid methylene
7.752	173	0.080	3.00e-03	6.84e-03	86.06	Unknown
6.982	250	0.120	3.14e-03	7.12e-03	83.18	Unknown
7.992	149	0.234	3.16e-03	7.14e-03	83.18	Unknown
1.292	584	0.143	3.19e-03	7.18e-03	86.06	Lipid methylene
7.832	165	-0.199	3.31e-03	7.44e-03	83.18	Unknown
4.182	319	0.166	3.83e-03	8.57e-03	86.06	Unknown
0.812	632	0.098	4.45e-03	9.91e-03	79.94	Cholesterol, lipid methyl
2.072	506	0.074	4.73e-03	1.05e-02	79.94	Lipid allylic
7.692	179	0.148	4.85e-03	1.07e-02	79.94	Unknown
3.172	396	0.054	4.86e-03	1.07e-02	86.06	$CaEDTA^{2-}$
0.822	631	0.084	4.87e-03	1.07e-02	79.94	Cholesterol, lipid methyl
9.382	10	0.363	4.93e-03	1.08e-02	76.35	Unknown
2.042	509	0.107	5.07e-03	1.11e-02	79.94	Lipid allylic

3.902	347	0.042	5.25e-03	1.14e-02	76.35	D-glucose, unknown
0.732	640	0.174	5.29e-03	1.15e-02	83.18	Unkown
2.142	499	0.044	5.53e-03	1.20e-02	76.35	Glutamine
8.112	137	0.194	5.57e-03	1.20e-02	76.35	Trigonelline
0.802	633	0.096	5.58e-03	1.20e-02	79.94	Cholesterol, lipid methyl
7.072	241	0.070	5.69e-03	1.22e-02	76.35	Unknown
2.812	432	0.125	5.96e-03	1.27e-02	79.94	Lipid diallylic
6.902	258	0.048	6.02e-03	1.28e-02	76.35	Tyrosine
8.092	139	0.169	6.07e-03	1.29e-02	76.35	Trigonelline
2.622	451	0.034	6.17e-03	1.31e-02	79.94	Unknown
2.562	457	0.034	6.25e-03	1.32e-02	86.06	$CaEDTA^{2-}$
8.352	113	0.287	6.64e-03	1.40e-02	86.06	Unknown
6.892	259	0.046	7.26e-03	1.52e-02	76.35	Unknown
0.922	621	0.154	7.40e-03	1.55e-02	76.35	Cholesterol, lipid methyl
3.152	398	0.026	8.58e-03	1.79e-02	86.06	$CaEDTA^{2-}$
7.442	204	0.055	9.02e-03	1.87e-02	68.24	Phenylalanine
8.652	83	-0.349	9.22e-03	1.91e-02	72.44	Unknown
8.162	132	0.186	9.47e-03	1.95e-02	68.24	Unknown
0.792	634	0.080	9.54e-03	1.96e-02	76.35	Cholesterol, lipid methyl
7.222	226	0.060	9.63e-03	1.97e-02	72.44	Tyrosine
1.202	593	0.163	9.68e-03	1.98e-02	79.94	Lipid methylene
1.772	536	0.049	1.00e-02	2.04e-02	72.44	Leucine, lysine
9.142	34	0.407	1.04e-02	2.10e-02	68.24	Trigonelline
6.922	256	0.068	1.09e-02	2.20e-02	72.44	Tyrosine
8.012	147	0.175	1.14e-02	2.30e-02	72.44	Unknown
2.182	495	0.061	1.16e-02	2.33e-02	68.24	Glutamine
8.042	144	0.248	1.36e-02	2.73e-02	72.44	Unknown
2.802	433	0.098	1.37e-02	2.75e-02	72.44	Lipid diallylic
8.982	50	-0.313	1.50e-02	2.99e-02	63.78	Unknown
2.712	442	0.031	1.55e-02	3.08e-02	68.24	$MgEDTA^{2-}$
0.762	637	0.064	1.56e-02	3.08e-02	68.24	Cholesterol, lipid methyl
7.402	208	0.046	1.60e-02	3.17e-02	63.78	Phenylalanine
2.052	508	0.071	1.67e-02	3.28e-02	68.24	Lipid allylic
8.182	130	0.159	1.75e-02	3.43e-02	59.13	Unknown
8.132	135	0.167	1.76e-02	3.44e-02	68.24	Unknown
6.962	252	0.056	1.83e-02	3.58e-02	63.78	Unknown
7.882	160	0.090	1.93e-02	3.75e-02	59.13	Unknown
6.742	274	-0.129	1.96e-02	3.80e-02	63.78	Unknown
2.792	434	0.084	2.02e-02	3.90e-02	63.78	Lipid diallylic
6.872	261	0.053	2.02e-02	3.90e-02	63.78	Unknown
0.672	646	0.126	2.14e-02	4.11e-02	49.50	Unkown
8.002	148	0.187	2.17e-02	4.16e-02	63.78	Unknown
1.832	530	0.036	2.20e-02	4.21e-02	63.78	Unknown
7.562	192	-0.078	2.24e-02	4.28e-02	59.13	Unknown
0.722	641	0.154	2.30e-02	4.37e-02	63.78	Unkown

7.742	174	0.063	2.35e-02	4.46e-02	59.13	Unknown
3.332	391	0.200	2.37e-02	4.49e-02	63.78	Proline
1.802	533	0.042	2.53e-02	4.78e-02	63.78	Unknown

Table 7.11: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from diabetic nephropathy from those suffering from systemic diseases.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
3.842	353	0.422	5.93e-83	3.91e-80	100.00	D-glucose, unknown
3.802	357	0.368	1.95e-82	6.42e-80	100.00	D-glucose, alanine
3.562	368	0.439	6.36e-82	1.40e-79	100.00	D-glucose
3.782	359	0.332	1.59e-81	2.12e-79	100.00	D-glucose, alanine, glutamine, arginine
3.552	369	0.452	1.60e-81	2.12e-79	100.00	D-glucose, myo-inositol
3.872	350	0.369	3.68e-81	4.05e-79	100.00	D-glucose, unknown
3.882	349	0.413	9.59e-81	9.04e-79	100.00	D-glucose, unknown
3.412	383	0.476	2.56e-80	2.12e-78	100.00	D-glucose, carnitine, taurine, proline
3.862	351	0.451	3.44e-80	2.52e-78	100.00	D-glucose, unknown
3.742	363	0.449	1.30e-79	8.59e-78	100.00	D-glucose, leucine
3.912	346	0.410	1.73e-79	1.04e-77	100.00	D-glucose, betaine, unknown
3.572	367	0.411	2.41e-79	1.32e-77	100.00	D-glucose, glycine
3.852	352	0.431	5.33e-79	2.71e-77	100.00	D-glucose, unknown
3.932	344	0.418	1.90e-78	8.97e-77	100.00	D-glucose
3.432	381	0.472	2.11e-78	9.30e-77	100.00	D-glucose, carnitine, taurine, proline
3.442	380	0.464	2.97e-78	1.23e-76	100.00	D-glucose, carnitine, taurine, proline
3.502	374	0.483	5.13e-78	1.99e-76	100.00	D-glucose
3.722	365	0.378	7.34e-78	2.69e-76	100.00	D-glucose, N,N-dimethylglycine
3.482	376	0.471	1.87e-77	6.36e-76	100.00	D-glucose
3.512	373	0.482	1.93e-77	6.36e-76	100.00	D-glucose
3.752	362	0.435	2.03e-76	6.37e-75	100.00	D-glucose, glutamic acid
3.762	361	0.406	1.90e-75	5.69e-74	100.00	D-glucose, arginine, glutamine, glutamic acid
3.492	375	0.471	3.14e-75	9.02e-74	100.00	D-glucose

3.532	371	0.416	1.72e-73	4.72e-72	100.00	D-glucose
3.452	379	0.402	2.20e-73	5.80e-72	100.00	D-glucose, carnitine, proline
3.472	377	0.445	2.43e-72	6.16e-71	100.00	D-glucose
3.422	382	0.461	1.89e-71	4.63e-70	100.00	D-glucose, carnitine, taurine, proline
3.732	364	0.401	1.13e-70	2.66e-69	100.00	D-glucose, unknown
3.772	360	0.364	4.98e-70	1.13e-68	100.00	D-glucose, alanine, glutamine, arginine
3.792	358	0.228	3.80e-66	8.35e-65	100.00	D-glucose, alanine
3.942	343	0.404	9.08e-60	1.93e-58	100.00	D-glucose
3.812	356	0.307	2.66e-59	5.49e-58	100.00	D-glucose
3.542	370	0.326	2.55e-47	5.09e-46	100.00	D-glucose, myo-inositol
3.922	345	0.244	2.60e-45	5.04e-44	100.00	D-glucose, unknown
3.522	372	0.379	6.51e-42	1.23e-40	100.00	D-glucose
3.832	354	0.190	3.15e-41	5.78e-40	100.00	Unknown
3.402	384	0.275	1.58e-40	2.82e-39	100.00	Unknown
3.962	341	0.124	2.04e-33	3.55e-32	100.00	Unknown
3.972	340	0.114	2.18e-27	3.69e-26	100.00	Unknown
3.822	355	0.173	7.00e-27	1.16e-25	100.00	Unknown
3.372	387	-0.184	2.93e-23	4.71e-22	100.00	Methanol, proline
3.892	348	0.117	1.17e-22	1.84e-21	100.00	Unknown
3.462	378	0.173	4.65e-19	7.13e-18	100.00	D-glucose
3.712	366	0.126	4.26e-15	6.40e-14	100.00	Unknown
7.372	211	0.097	9.92e-12	1.45e-10	100.00	Phenylalanine
2.402	473	0.070	2.99e-11	4.28e-10	100.00	Glutamine, carnitine
6.832	265	0.105	5.54e-11	7.78e-10	100.00	Unknown
7.382	210	0.079	5.47e-10	7.52e-09	100.00	Phenylalanine
6.842	264	0.100	6.52e-10	8.79e-09	100.00	Unknown
6.822	266	0.098	1.91e-09	2.53e-08	100.00	Unknown
7.332	215	0.081	3.79e-09	4.90e-08	100.00	Phenylalanine
7.232	225	0.084	4.56e-09	5.79e-08	100.00	Unknown
3.952	342	0.114	5.41e-09	6.73e-08	100.00	Unknown
7.302	218	0.082	7.15e-09	8.74e-08	100.00	Unknown
3.392	385	0.257	8.02e-09	9.45e-08	100.00	Methanol, proline
7.292	219	0.079	8.02e-09	9.45e-08	100.00	Unknown
7.142	234	0.110	8.27e-09	9.57e-08	99.99	Unknown
7.322	216	0.084	1.50e-08	1.71e-07	100.00	Unknown
4.032	334	0.092	1.70e-08	1.88e-07	99.99	Unknown
3.982	339	0.070	1.71e-08	1.88e-07	99.99	Unknown
7.312	217	0.082	2.87e-08	3.10e-07	100.00	Unknown
6.852	263	0.091	3.09e-08	3.29e-07	99.99	Unknown
3.992	338	-0.069	3.49e-08	3.65e-07	99.98	Unknown
7.242	224	0.079	4.50e-08	4.64e-07	99.99	Unknown
7.052	243	0.100	6.47e-08	6.57e-07	99.98	Unknown
1.812	532	0.064	1.65e-07	1.65e-06	99.98	Unknown

7.262	222	0.082	1.78e-07	1.75e-06	99.99	Unknown
1.872	526	0.061	1.87e-07	1.81e-06	99.96	Overlap of multiple minor compounds
4.122	325	0.130	1.94e-07	1.85e-06	99.96	Proline, lactic acid
7.282	220	0.072	2.54e-07	2.39e-06	99.98	Unknown
7.042	244	0.126	2.64e-07	2.46e-06	99.92	Unknown
1.892	524	0.057	3.47e-07	3.18e-06	99.92	Overlap of multiple minor compounds
3.062	407	0.081	3.89e-07	3.52e-06	99.96	Creatinine
7.272	221	0.076	4.21e-07	3.76e-06	99.98	Unknown
7.252	223	0.079	5.55e-07	4.88e-06	99.96	Unknown
1.802	533	0.068	5.81e-07	5.04e-06	99.96	Unknown
7.132	235	0.091	6.51e-07	5.58e-06	99.92	Unknown
7.062	242	0.100	7.66e-07	6.48e-06	99.86	Unknown
1.902	523	0.053	8.59e-07	7.18e-06	99.86	Overlap of multiple minor compounds
1.882	525	0.054	1.19e-06	9.83e-06	99.86	Overlap of multiple minor compounds
1.082	605	0.088	1.54e-06	1.25e-05	99.76	Unknown
6.862	262	0.092	1.61e-06	1.29e-05	99.86	Unknown
1.092	604	0.092	1.63e-06	1.29e-05	99.76	Unknown
4.152	322	0.101	1.65e-06	1.29e-05	99.76	Proline, lactic acid
7.072	241	0.086	2.01e-06	1.56e-05	99.86	Unknown
6.812	267	0.091	2.14e-06	1.64e-05	99.92	Unknown
1.862	527	0.056	2.75e-06	2.08e-05	99.76	Unknown
2.412	472	0.088	2.80e-06	2.10e-05	99.59	Glutamine, carnitine
4.012	336	-0.052	3.31e-06	2.46e-05	99.76	Unknown
2.352	478	0.061	3.41e-06	2.50e-05	99.76	Proline, glutamic acid
6.802	268	0.094	3.71e-06	2.69e-05	99.76	Unknown
7.122	236	0.067	4.48e-06	3.21e-05	99.76	Unknown
1.822	531	0.053	5.31e-06	3.77e-05	99.76	Unknown
4.052	332	0.122	5.66e-06	3.97e-05	99.59	Unknown
6.882	260	0.070	5.84e-06	4.06e-05	99.59	Unknown
1.702	543	0.054	6.47e-06	4.45e-05	99.59	Unknown, arginine
4.192	318	0.135	7.95e-06	5.41e-05	99.59	Unknown
4.142	323	0.118	9.43e-06	6.35e-05	99.34	Proline, lactic acid
7.872	161	0.119	9.58e-06	6.39e-05	99.34	Unknown
7.192	229	0.064	1.05e-05	6.93e-05	99.59	Tyrosine
6.872	261	0.071	1.20e-05	7.87e-05	99.34	Unknown
8.352	113	0.324	1.85e-05	1.20e-04	98.95	Unknown
1.712	542	0.046	1.96e-05	1.26e-04	98.95	Leucine, lysine
1.072	606	0.074	2.01e-05	1.28e-04	98.95	Valine
1.842	529	0.051	2.12e-05	1.33e-04	98.95	Unknown
7.152	233	0.100	2.21e-05	1.37e-04	98.95	Unknown
1.102	603	0.080	2.26e-05	1.40e-04	98.95	Unknown

1.852	528	0.052	2.41e-05	1.47e-04	98.95	Unknown
8.392	109	0.320	2.93e-05	1.77e-04	98.95	Unknown
4.162	321	0.087	3.11e-05	1.86e-04	99.34	Proline, lactic acid
2.392	474	0.077	3.39e-05	2.02e-04	98.95	Unknown
7.032	245	0.111	3.45e-05	2.03e-04	98.36	Unknown
1.692	544	0.056	3.47e-05	2.03e-04	98.95	Unknown, arginine
7.012	247	0.107	3.54e-05	2.05e-04	98.95	Unknown
6.962	252	0.070	3.97e-05	2.28e-04	97.53	Unknown
4.002	337	-0.044	4.17e-05	2.37e-04	98.36	Unknown
1.792	534	0.065	4.83e-05	2.73e-04	98.95	Unknown
7.182	230	0.074	5.00e-05	2.79e-04	98.36	Unknown
4.132	324	0.116	5.03e-05	2.79e-04	98.36	Proline, lactic acid
2.342	479	0.063	7.09e-05	3.90e-04	97.53	Proline, glutamic acid
2.642	449	0.043	8.33e-05	4.55e-04	98.36	Unknown
6.792	269	0.083	8.92e-05	4.82e-04	98.36	Unknown
3.902	347	0.041	9.61e-05	5.16e-04	97.53	D-glucose, unknown
4.112	326	0.135	1.01e-04	5.37e-04	98.36	Proline, lactic acid
7.022	246	0.141	1.02e-04	5.37e-04	97.53	Unknown
7.882	160	0.105	1.19e-04	6.24e-04	96.36	Unknown
7.992	149	0.217	1.24e-04	6.44e-04	96.36	Unknown
1.682	545	0.060	1.42e-04	7.34e-04	97.53	Unknown, arginine
7.362	212	0.056	1.48e-04	7.57e-04	97.53	Phenylalanine
7.112	237	0.048	1.56e-04	7.91e-04	97.53	Unknown
1.832	530	0.042	1.59e-04	8.00e-04	97.53	Unknown
7.342	214	0.043	1.61e-04	8.03e-04	96.36	Phenylalanine
1.722	541	0.041	1.66e-04	8.24e-04	96.36	Leucine, lysine
6.892	259	0.046	1.69e-04	8.32e-04	97.53	Unknown
1.772	536	0.051	1.94e-04	9.44e-04	96.36	Leucine, lysine
8.552	93	-0.303	1.95e-04	9.44e-04	96.36	Unknown
2.122	501	0.060	2.01e-04	9.70e-04	96.36	Lipid allylic
8.162	132	0.190	2.09e-04	9.94e-04	92.69	Unknown
1.132	600	0.062	2.09e-04	9.94e-04	96.36	Unknown
2.422	471	0.065	2.15e-04	1.02e-03	96.36	Glutamine, carnitine
1.062	607	0.072	2.23e-04	1.04e-03	96.36	Valine
2.652	448	0.035	2.62e-04	1.21e-03	96.36	Unknown
2.302	483	0.075	2.63e-04	1.21e-03	96.36	Lipid (methylene carbonyl)
2.932	420	-0.049	3.08e-04	1.41e-03	96.36	Unknown
2.332	480	0.062	3.19e-04	1.45e-03	94.78	Proline, glutamic acid
1.002	613	0.063	3.49e-04	1.58e-03	94.78	Valine, lipid methyl, cholesterol (ester)
1.782	535	0.060	4.16e-04	1.87e-03	96.36	Unknown
1.762	537	0.046	4.21e-04	1.88e-03	94.78	Leucine, lysine
2.322	481	0.059	4.63e-04	2.05e-03	92.69	Lipid (methylene carbonyl)
1.912	522	0.043	5.17e-04	2.27e-03	94.78	Overlap of multiple minor compounds

2.312	482	0.063	5.47e-04	2.39e-03	92.69	Lipid (methylene carbonyl)
8.002	148	0.200	5.61e-04	2.44e-03	92.69	Unknown
4.092	328	0.086	6.33e-04	2.73e-03	94.78	Unknown
1.752	538	0.044	6.61e-04	2.83e-03	92.69	Leucine, lysine
1.742	539	0.041	6.73e-04	2.86e-03	92.69	Leucine, lysine
8.152	133	0.162	6.82e-04	2.88e-03	92.69	Unknown
4.342	303	0.060	7.59e-04	3.19e-03	94.78	Lipid alpha-methylene to carboxyl, lipid glycerine
1.732	540	0.039	7.93e-04	3.31e-03	92.69	Leucine, lysine
2.722	441	-0.036	8.16e-04	3.37e-03	92.69	$MgEDTA^{2-}$
1.962	517	0.045	8.16e-04	3.37e-03	92.69	Lipid allylic
2.942	419	-0.026	8.25e-04	3.38e-03	94.78	N,N-dimethylglycine
6.972	251	0.070	9.74e-04	3.97e-03	92.69	Unknown
2.432	470	0.040	1.01e-03	4.08e-03	90.01	Glutamine, carnitine
1.672	546	0.060	1.27e-03	5.12e-03	90.01	Unknown, arginine
7.172	231	0.085	1.44e-03	5.76e-03	90.01	Unknown
8.192	129	0.153	1.47e-03	5.83e-03	86.69	Unknown
2.362	477	0.051	1.50e-03	5.95e-03	90.01	Proline, glutamic acid
8.282	120	0.134	1.60e-03	6.29e-03	92.69	Unknown
1.972	516	0.049	1.65e-03	6.45e-03	90.01	Lipid allylic
1.122	601	0.066	1.85e-03	7.19e-03	86.69	Unknown
7.972	151	0.205	1.94e-03	7.47e-03	86.69	Unknown
2.382	475	0.054	2.00e-03	7.68e-03	86.69	Proline, glutamic acid
7.982	150	0.187	2.04e-03	7.78e-03	86.69	Unknown
1.022	611	0.062	2.26e-03	8.58e-03	86.69	L-isoleucine, lipid methyl, cholesterol (ester)
2.632	450	0.031	2.31e-03	8.69e-03	86.69	Unknown
6.782	270	0.116	2.47e-03	9.27e-03	86.69	Unknown
7.862	162	0.088	2.73e-03	1.02e-02	82.67	Unknown
4.312	306	0.061	2.82e-03	1.05e-02	86.69	Lipid alpha-methylene to carboxyl, lipid glycerine
1.512	562	0.057	2.89e-03	1.06e-02	86.69	Alanine
7.102	238	0.079	2.93e-03	1.07e-02	82.67	Unknown
2.112	502	0.050	2.94e-03	1.07e-02	86.69	Lipid allylic
8.012	147	0.146	2.98e-03	1.08e-02	82.67	Unknown
8.092	139	0.130	2.99e-03	1.08e-02	82.67	Trigonelline
8.122	136	0.109	3.16e-03	1.13e-02	82.67	Unknown
1.952	518	0.036	3.30e-03	1.17e-02	86.69	Acetic acid
8.172	131	0.126	3.30e-03	1.17e-02	82.67	Unknown
6.902	258	0.036	3.41e-03	1.20e-02	82.67	Tyrosine
1.152	598	0.060	3.54e-03	1.24e-02	82.67	Lipid methylene
1.012	612	0.058	3.56e-03	1.24e-02	82.67	Valine, lipid methyl, cholesterol (ester)
1.142	599	0.053	3.65e-03	1.27e-02	82.67	Unknown
8.342	114	0.202	3.71e-03	1.28e-02	82.67	Unknown

1.662	547	0.063	4.02e-03	1.38e-02	82.67	Lipids (?)
2.102	503	0.042	4.04e-03	1.38e-02	82.67	Lipid allylic
8.222	126	0.063	4.42e-03	1.50e-02	77.95	Unknown
4.182	319	0.117	4.46e-03	1.51e-02	82.67	Unknown
1.922	521	0.037	4.48e-03	1.51e-02	82.67	Overlap of multiple minor compounds
1.232	590	0.054	4.75e-03	1.59e-02	82.67	Lipid methylene
4.072	330	0.068	4.79e-03	1.60e-02	86.69	Creatinine
7.352	213	0.034	4.95e-03	1.64e-02	82.67	Phenylalanine
4.282	309	0.050	5.02e-03	1.65e-02	82.67	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
8.202	128	0.104	5.02e-03	1.65e-02	77.95	Unknown
1.982	515	0.049	5.12e-03	1.67e-02	82.67	Lipid allylic
1.032	610	0.061	5.48e-03	1.78e-02	82.67	L-isoleucine, lipid methyl, cholesterol (ester)
2.132	500	0.033	5.63e-03	1.82e-02	82.67	Glutamine
2.372	476	0.048	5.81e-03	1.87e-02	77.95	Proline, glutamic acid
1.492	564	0.050	5.93e-03	1.90e-02	82.67	Alanine
4.102	327	0.096	5.97e-03	1.90e-02	82.67	Unknown
8.082	140	0.099	6.22e-03	1.97e-02	77.95	Trigonelline
4.062	331	0.073	6.24e-03	1.97e-02	77.95	Creatinine
2.922	421	-0.047	6.29e-03	1.98e-02	82.67	Unknown
4.172	320	0.087	6.56e-03	2.05e-02	77.95	Unknown
7.852	163	0.112	6.62e-03	2.06e-02	77.95	Unknown
4.082	329	0.069	6.74e-03	2.09e-02	82.67	Creatinine
9.232	25	0.236	6.89e-03	2.12e-02	77.95	Unknown
1.502	563	0.050	7.29e-03	2.24e-02	77.95	Alanine
3.032	410	0.017	7.40e-03	2.26e-02	77.95	Lysine, unknown
2.912	422	-0.036	7.88e-03	2.40e-02	77.95	Unknown
2.002	513	0.050	8.23e-03	2.49e-02	77.95	Lipid allylic
6.742	274	-0.104	8.71e-03	2.62e-02	72.57	Unknown
8.302	118	0.170	9.56e-03	2.87e-02	77.95	Unknown
8.232	125	0.056	1.01e-02	3.01e-02	72.57	Unknown
8.372	111	0.196	1.04e-02	3.08e-02	72.57	Unknown
8.292	119	0.142	1.13e-02	3.33e-02	72.57	Unknown
2.292	484	0.061	1.16e-02	3.40e-02	72.57	Lipid (methylene carbonyl)
4.302	307	0.057	1.18e-02	3.47e-02	72.57	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
8.102	138	0.116	1.24e-02	3.63e-02	72.57	Trigonelline
1.992	514	0.046	1.28e-02	3.71e-02	72.57	Lipid allylic
2.012	512	0.050	1.34e-02	3.87e-02	72.57	Lipid allylic
1.482	565	0.043	1.42e-02	4.10e-02	72.57	Alanine
4.322	305	0.045	1.47e-02	4.22e-02	72.57	Lipid alpha-methylene to carboxyl, lipid glycerine
3.332	391	0.154	1.50e-02	4.29e-02	66.61	Proline

3.072	406	0.018	1.75e-02	4.95e-02	66.61	Unknown
8.422	106	0.201	1.75e-02	4.95e-02	60.20	Unknown
8.262	122	0.054	1.76e-02	4.97e-02	66.61	Unknown

Table 7.12: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from diabetic nephropathy from those suffering from hypertensive nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
1.152	598	0.154	7.84e-06	0.00284	99.21	Lipid methylene
1.832	530	0.083	8.60e-06	0.00284	98.96	Unknown
1.232	590	0.134	2.78e-05	0.00419	98.25	Lipid methylene
1.222	591	0.163	4.59e-05	0.00419	98.65	Lipid methylene
1.802	533	0.092	4.77e-05	0.00419	96.43	Unknown
1.842	529	0.080	5.72e-05	0.00419	97.76	Unknown
1.822	531	0.078	6.13e-05	0.00419	96.43	Unknown
7.212	227	0.116	7.52e-05	0.00419	97.16	Tyrosine
1.692	544	0.090	7.99e-05	0.00419	96.43	Unknown, arginine
3.962	341	0.067	8.08e-05	0.00419	97.76	Unknown
1.812	532	0.080	8.33e-05	0.00419	95.56	Unknown
1.672	546	0.123	8.90e-05	0.00419	96.43	Unknown, arginine
2.122	501	0.105	9.41e-05	0.00419	96.43	Lipid allylic
7.272	221	0.097	1.09e-04	0.00419	94.52	Unknown
7.222	226	0.107	1.09e-04	0.00419	97.16	Tyrosine
3.972	340	0.067	1.10e-04	0.00419	96.43	Unknown
7.292	219	0.088	1.18e-04	0.00419	94.52	Unknown
7.312	217	0.096	1.20e-04	0.00419	94.52	Unknown
1.962	517	0.087	1.21e-04	0.00419	95.56	Lipid allylic
7.262	222	0.100	1.42e-04	0.00454	94.52	Unknown
2.642	449	0.070	1.44e-04	0.00454	95.56	Unknown
1.792	534	0.102	1.57e-04	0.00470	93.30	Unknown
7.282	220	0.087	1.70e-04	0.00479	94.52	Unknown
1.972	516	0.097	1.79e-04	0.00479	94.52	Lipid allylic
1.682	545	0.098	1.85e-04	0.00479	94.52	Unknown, arginine
1.362	577	0.212	1.98e-04	0.00479	95.56	Lipid methylene, lactic acid, threonine
2.402	473	0.065	1.99e-04	0.00479	94.52	Glutamine, carnitine

7.302	218	0.088	2.03e-04	0.00479	93.30	Unknown
1.662	547	0.136	2.40e-04	0.00543	94.52	Lipids (?)
1.112	602	0.114	2.47e-04	0.00543	94.52	Unknown
3.062	407	-0.097	2.75e-04	0.00579	93.30	Creatinine
1.242	589	0.129	2.81e-04	0.00579	94.52	Lipid methylene
0.852	628	0.091	3.77e-04	0.00754	94.52	Cholesterol, lipid methyl
3.042	409	-0.036	4.08e-04	0.00772	93.30	Lysine, unknown
0.842	629	0.091	4.22e-04	0.00772	94.52	Cholesterol, lipid methyl
1.702	543	0.070	4.47e-04	0.00772	91.87	Unknown, arginine
4.332	304	0.104	4.55e-04	0.00772	90.23	Lipid alpha-methylene to carboxyl, lipid glycerine
0.832	630	0.101	4.55e-04	0.00772	93.30	Cholesterol, lipid methyl
1.852	528	0.072	4.56e-04	0.00772	93.30	Unknown
0.962	617	0.125	5.53e-04	0.00912	91.87	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
2.652	448	0.055	5.70e-04	0.00918	93.30	Unknown
1.982	515	0.101	5.91e-04	0.00927	90.23	Lipid allylic
7.432	205	-0.096	6.04e-04	0.00927	90.23	Phenylalanine
4.342	303	0.101	6.35e-04	0.00937	88.36	Lipid alpha-methylene to carboxyl, lipid glycerine
1.142	599	0.104	6.39e-04	0.00937	91.87	Unknown
4.322	305	0.104	6.91e-04	0.00971	88.36	Lipid alpha-methylene to carboxyl, lipid glycerine
7.232	225	0.081	6.95e-04	0.00971	86.25	Unknown
1.862	527	0.068	7.21e-04	0.00971	90.23	Unknown
7.322	216	0.084	7.31e-04	0.00971	88.36	Unknown
4.312	306	0.116	7.49e-04	0.00971	88.36	Lipid alpha-methylene to carboxyl, lipid glycerine
4.122	325	0.141	7.50e-04	0.00971	91.87	Proline, lactic acid
1.712	542	0.061	7.84e-04	0.00995	90.23	Leucine, lysine
1.342	579	0.175	8.08e-04	0.01010	90.23	Lipid methylene, lactic acid, threonine
2.132	500	0.067	9.02e-04	0.01080	88.36	Glutamine
1.252	588	0.131	9.04e-04	0.01080	90.23	Lipid methylene
2.312	482	0.101	9.17e-04	0.01080	90.23	Lipid (methylene carbonyl)
1.782	535	0.095	9.30e-04	0.01080	88.36	Unknown
7.252	223	0.087	9.76e-04	0.01110	86.25	Unknown
2.322	481	0.093	1.10e-03	0.01230	90.23	Lipid (methylene carbonyl)
1.652	548	0.147	1.16e-03	0.01270	88.36	Lipids (?)
0.822	631	0.116	1.25e-03	0.01360	88.36	Cholesterol, lipid methyl
0.952	618	0.136	1.34e-03	0.01410	88.36	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.952	518	0.066	1.35e-03	0.01410	88.36	Acetic acid
2.632	450	0.054	1.37e-03	0.01410	86.25	Unknown
2.742	439	0.076	1.39e-03	0.01410	86.25	Lipid diallylic

1.992	514	0.098	1.42e-03	0.01420	86.25	Lipid allylic
1.132	600	0.089	1.48e-03	0.01460	88.36	Unknown
6.842	264	0.086	1.56e-03	0.01510	83.88	Unknown
0.862	627	0.092	1.59e-03	0.01520	86.25	Cholesterol, lipid methyl
1.272	586	0.143	1.83e-03	0.01720	86.25	Lipid methylene
1.262	587	0.133	1.84e-03	0.01720	86.25	Lipid methylene
0.892	624	0.127	1.89e-03	0.01730	83.88	Cholesterol, lipid methyl
2.002	513	0.097	1.95e-03	0.01750	83.88	Lipid allylic
0.972	616	0.102	1.97e-03	0.01750	83.88	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
4.302	307	0.118	1.99e-03	0.01750	83.88	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.302	583	0.216	2.04e-03	0.01770	83.88	Lipid methylene
1.212	592	0.143	2.08e-03	0.01780	88.36	Lipid methylene
3.052	408	-0.046	2.12e-03	0.01800	86.25	Creatinine
1.352	578	0.162	2.18e-03	0.01820	86.25	Lipid methylene, lactic acid, threonine
2.012	512	0.103	2.23e-03	0.01840	83.88	Lipid allylic
4.292	308	0.113	2.28e-03	0.01860	81.27	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.282	585	0.151	2.35e-03	0.01890	83.88	Lipid methylene
2.302	483	0.103	2.54e-03	0.02020	83.88	Lipid (methylene carbonyl)
1.172	596	0.123	2.59e-03	0.02040	83.88	Lipid methylene
6.852	263	0.082	2.67e-03	0.02070	83.88	Unknown
7.332	215	0.068	2.71e-03	0.02080	81.27	Phenylalanine
0.882	625	0.106	2.80e-03	0.02130	81.27	Cholesterol, lipid methyl
2.022	511	0.114	2.88e-03	0.02150	81.27	Lipid allylic
1.162	597	0.130	2.90e-03	0.02150	86.25	Lipid methylene
2.032	510	0.127	3.03e-03	0.02190	81.27	Lipid allylic
9.232	25	0.435	3.05e-03	0.02190	83.88	Unknown
2.272	486	0.193	3.05e-03	0.02190	81.27	Lipid (methylene carbonyl)
2.262	487	0.219	3.10e-03	0.02200	81.27	Lipid (methylene carbonyl)
2.622	451	0.043	3.15e-03	0.02210	78.41	Unknown
2.252	488	0.171	3.26e-03	0.02230	78.41	Lipid (methylene carbonyl), acetone
2.332	480	0.085	3.27e-03	0.02230	81.27	Proline, glutamic acid
0.802	633	0.122	3.28e-03	0.02230	81.27	Cholesterol, lipid methyl
1.292	584	0.171	3.37e-03	0.02270	78.41	Lipid methylene
0.812	632	0.121	3.52e-03	0.02350	81.27	Cholesterol, lipid methyl
6.742	274	-0.193	3.62e-03	0.02370	81.27	Unknown
6.832	265	0.078	3.63e-03	0.02370	78.41	Unknown
1.122	601	0.104	3.81e-03	0.02430	81.27	Unknown
1.642	549	0.157	3.82e-03	0.02430	81.27	Lipids (?)
1.102	603	0.092	3.87e-03	0.02430	81.27	Unknown
6.672	281	-0.425	3.89e-03	0.02430	81.27	Unknown

2.782	435	0.127	3.94e-03	0.02430	78.41	Lipid diallylic
1.312	582	0.214	3.99e-03	0.02430	78.41	Lipid methylene
0.902	623	0.150	4.01e-03	0.02430	78.41	Cholesterol, lipid methyl
0.872	626	0.097	4.03e-03	0.02430	78.41	Cholesterol, lipid methyl
0.792	634	0.107	4.05e-03	0.02430	78.41	Cholesterol, lipid methyl
2.042	509	0.130	4.17e-03	0.02480	78.41	Lipid allylic
4.352	302	0.089	4.22e-03	0.02480	75.32	Lipid alpha-methylene to carboxyl, lipid glycerine
4.132	324	0.137	4.26e-03	0.02490	81.27	Proline, lactic acid
0.922	621	0.196	4.47e-03	0.02550	78.41	Cholesterol, lipid methyl
2.792	434	0.124	4.48e-03	0.02550	78.41	Lipid diallylic
0.912	622	0.182	4.48e-03	0.02550	75.32	Cholesterol, lipid methyl
1.522	561	0.074	4.53e-03	0.02550	78.41	Lipids (?)
2.752	438	0.093	4.58e-03	0.02560	75.32	Lipid diallylic
6.932	255	0.091	4.64e-03	0.02570	75.32	Tyrosine
1.452	568	0.108	4.70e-03	0.02580	78.41	Lipid methylene
2.292	484	0.115	4.74e-03	0.02590	78.41	Lipid (methylene carbonyl)
6.822	266	0.077	4.79e-03	0.02590	75.32	Unknown
4.072	330	-0.114	4.85e-03	0.02600	68.48	Creatinine
1.612	552	0.216	5.01e-03	0.02670	75.32	Lipids (?)
2.282	485	0.142	5.07e-03	0.02670	78.41	Lipid (methylene carbonyl)
1.422	571	0.151	5.09e-03	0.02670	78.41	Lipid methylene
4.282	309	0.084	5.23e-03	0.02710	75.32	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.972	416	-0.032	5.26e-03	0.02710	78.41	Unknown
1.622	551	0.199	5.37e-03	0.02750	78.41	Lipids (?)
1.632	550	0.174	5.60e-03	0.02840	78.41	Lipids (?)
2.102	503	0.068	5.73e-03	0.02870	78.41	Lipid allylic
1.182	595	0.095	5.79e-03	0.02870	78.41	Lipid methylene
1.332	580	0.225	5.83e-03	0.02870	75.32	Lipid methylene
0.942	619	0.147	5.85e-03	0.02870	78.41	Cholesterol, lipid methyl
7.442	204	-0.069	5.86e-03	0.02870	81.27	Phenylalanine
1.322	581	0.214	5.97e-03	0.02880	75.32	Lipid methylene
2.112	502	0.077	5.98e-03	0.02880	75.32	Lipid allylic
1.602	553	0.217	6.29e-03	0.02990	72.00	Lipids (?)
1.872	526	0.053	6.30e-03	0.02990	75.32	Overlap of multiple minor compounds
1.532	560	0.088	6.47e-03	0.03050	75.32	Lipids (?)
1.462	567	0.092	6.51e-03	0.03050	75.32	Lipid methylene
2.812	432	0.147	6.86e-03	0.03190	72.00	Lipid diallylic
1.412	572	0.179	7.20e-03	0.03320	75.32	Lipid methylene
2.772	436	0.115	7.34e-03	0.03360	72.00	Lipid diallylic
1.432	570	0.116	7.49e-03	0.03380	75.32	Lipid methylene
1.092	604	0.086	7.49e-03	0.03380	75.32	Unknown
2.072	506	0.083	7.69e-03	0.03450	72.00	Lipid allylic

1.202	593	0.201	7.78e-03	0.03470	78.41	Lipid methylene
0.932	620	0.171	7.87e-03	0.03490	72.00	Cholesterol, lipid methyl
4.142	323	0.118	8.18e-03	0.03600	75.32	Proline, lactic acid
1.192	594	0.168	8.42e-03	0.03650	78.41	Lipid methylene
1.482	565	0.077	8.52e-03	0.03650	72.00	Alanine
1.772	536	0.060	8.55e-03	0.03650	72.00	Leucine, lysine
2.562	457	-0.039	8.58e-03	0.03650	75.32	$CaEDTA^{2-}$
7.242	224	0.064	8.59e-03	0.03650	68.48	Unknown
2.762	437	0.103	8.62e-03	0.03650	68.48	Lipid diallylic
4.362	301	0.089	8.67e-03	0.03650	68.48	Unknown
1.442	569	0.105	8.80e-03	0.03670	72.00	Lipid methylene
1.592	554	0.203	9.23e-03	0.03830	68.48	Lipids (?)
2.822	431	0.141	9.31e-03	0.03840	68.48	Lipid diallylic
2.552	458	-0.058	9.55e-03	0.03920	86.25	Citric acid
1.472	566	0.082	9.72e-03	0.03960	68.48	Lipid methylene
6.772	271	-0.270	9.85e-03	0.03990	75.32	Unknown
0.602	653	-0.387	9.96e-03	0.04010	78.41	Unkown
3.922	345	0.073	1.00e-02	0.04010	83.88	D-glucose, unknown
2.342	479	0.068	1.01e-02	0.04010	72.00	Proline, glutamic acid
1.382	575	0.228	1.04e-02	0.04090	68.48	Lipid methylene
1.402	573	0.212	1.04e-02	0.04090	68.48	Lipid methylene
2.802	433	0.122	1.05e-02	0.04090	68.48	Lipid diallylic
1.392	574	0.225	1.06e-02	0.04130	68.48	Lipid methylene
2.542	459	-0.060	1.07e-02	0.04140	78.41	Unknown
2.062	507	0.057	1.10e-02	0.04230	72.00	Lipid allylic
9.452	3	0.391	1.12e-02	0.04260	72.00	Unknown
1.372	576	0.225	1.13e-02	0.04300	68.48	Lipid methylene
1.042	609	0.113	1.16e-02	0.04380	68.48	L-isoleucine, lipid methyl, cholesterol (ester)
2.232	490	0.152	1.17e-02	0.04380	68.48	Lipid (methylene carbonyl)
2.082	505	0.095	1.21e-02	0.04530	68.48	Lipid allylic
1.022	611	0.085	1.23e-02	0.04570	72.00	L-isoleucine, lipid methyl, cholesterol (ester)
0.992	614	0.084	1.24e-02	0.04580	68.48	Leucine, lipid methyl, cholesterol (ester)
0.652	648	-0.283	1.26e-02	0.04610	72.00	Unkown
0.982	615	0.082	1.27e-02	0.04630	68.48	Leucine, lipid methyl, cholesterol (ester)
2.242	489	0.160	1.28e-02	0.04630	64.78	Lipid (methylene carbonyl), acetone
1.582	555	0.181	1.29e-02	0.04640	64.78	Lipids (?)
2.052	508	0.088	1.32e-02	0.04730	64.78	Lipid allylic

Table 7.13: **Previous page: Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from glomerulonephritis from those suffering from hereditary diseases.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
1.222	591	0.149	0.000113	0.0318	97.42	Lipid methylene
2.402	473	0.064	0.000144	0.0318	95.86	Glutamine, carnitine
1.212	592	0.167	0.000189	0.0318	97.42	Lipid methylene
0.832	630	0.102	0.000248	0.0318	94.84	Cholesterol, lipid methyl
4.282	309	0.106	0.000268	0.0318	93.62	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
3.982	339	0.073	0.000289	0.0318	93.62	Unknown
1.832	530	0.063	0.000395	0.0332	92.19	Unknown
3.992	338	0.071	0.000439	0.0332	95.86	Unknown
1.812	532	0.068	0.000506	0.0332	90.53	Unknown
4.292	308	0.123	0.000525	0.0332	92.19	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
7.312	217	0.082	0.000582	0.0332	86.46	Unknown
2.342	479	0.088	0.000612	0.0332	92.19	Proline, glutamic acid
1.792	534	0.088	0.000724	0.0332	88.63	Unknown
7.562	192	0.133	0.000747	0.0332	93.62	Unknown
2.332	480	0.094	0.000754	0.0332	92.19	Proline, glutamic acid
8.142	134	0.301	0.000832	0.0343	93.62	Unknown
2.972	416	-0.036	0.000894	0.0347	88.63	Unknown
2.412	472	0.100	0.000963	0.0353	90.53	Glutamine, carnitine
1.692	544	0.071	0.001170	0.0357	84.03	Unknown, arginine
7.292	219	0.071	0.001220	0.0357	84.03	Unknown
1.702	543	0.062	0.001240	0.0357	86.46	Unknown, arginine
2.322	481	0.088	0.001290	0.0357	88.63	Lipid (methylene carbonyl)
1.822	531	0.060	0.001310	0.0357	86.46	Unknown
2.122	501	0.083	0.001320	0.0357	86.46	Lipid allylic
7.012	247	0.133	0.001390	0.0357	86.46	Unknown
1.152	598	0.106	0.001410	0.0357	88.63	Lipid methylene
4.302	307	0.116	0.001510	0.0370	86.46	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine

1.712	542	0.055	0.001770	0.0397	84.03	Leucine, lysine
7.322	216	0.074	0.001840	0.0397	81.33	Unknown
2.312	482	0.091	0.001990	0.0397	86.46	Lipid (methylene carbonyl)
0.822	631	0.107	0.002050	0.0397	86.46	Cholesterol, lipid methyl
7.302	218	0.070	0.002060	0.0397	78.37	Unknown
3.342	390	-0.137	0.002090	0.0397	84.03	Proline
7.552	193	0.131	0.002110	0.0397	88.63	Unknown
8.862	62	-0.587	0.002210	0.0397	88.63	Trigonelline
7.432	205	-0.082	0.002230	0.0397	84.03	Phenylalanine
1.782	535	0.084	0.002270	0.0397	81.33	Unknown
1.672	546	0.092	0.002290	0.0397	81.33	Unknown, arginine
1.232	590	0.093	0.002360	0.0400	84.03	Lipid methylene
8.152	133	0.233	0.002520	0.0416	84.03	Unknown
1.802	533	0.065	0.002730	0.0440	81.33	Unknown
1.662	547	0.106	0.002890	0.0449	81.33	Lipids (?)
2.242	489	0.185	0.002930	0.0449	81.33	Lipid (methylene carbonyl), acetone
6.852	263	0.078	0.003100	0.0461	81.33	Unknown
1.682	545	0.075	0.003150	0.0461	78.37	Unknown, arginine
1.962	517	0.065	0.003210	0.0461	78.37	Lipid allylic
3.042	409	-0.029	0.003330	0.0467	78.37	Lysine, unknown

Table 7.14: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), *p*-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from glomerulonephritis from those suffering from interstitial nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
1.232	590	0.126	1.16e-06	0.000517	99.76	Lipid methylene
0.842	629	0.101	1.57e-06	0.000517	99.76	Cholesterol, lipid methyl
1.702	543	0.075	3.77e-06	0.000830	99.27	Unknown, arginine
2.122	501	0.099	6.15e-06	0.000954	99.27	Lipid allylic
1.692	544	0.082	8.46e-06	0.000954	98.57	Unknown, arginine
1.242	589	0.127	1.00e-05	0.000954	99.27	Lipid methylene
0.972	616	0.117	1.10e-05	0.000954	98.97	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.902	523	0.064	1.31e-05	0.000954	98.97	Overlap of multiple minor compounds

1.682	545	0.092	1.44e-05	0.000954	98.57	Unknown, arginine
0.832	630	0.101	1.51e-05	0.000954	98.97	Cholesterol, lipid methyl
0.982	615	0.115	1.59e-05	0.000954	98.97	Leucine, lipid methyl, cholesterol (ester)
7.032	245	0.156	1.82e-05	0.000963	98.97	Unknown
1.152	598	0.119	2.00e-05	0.000963	98.97	Lipid methylene
1.222	591	0.138	2.04e-05	0.000963	99.27	Lipid methylene
6.842	264	0.092	2.56e-05	0.001130	97.36	Unknown
1.882	525	0.063	3.02e-05	0.001250	98.04	Overlap of multiple minor compounds
1.712	542	0.061	3.38e-05	0.001310	98.04	Leucine, lysine
1.022	611	0.114	3.70e-05	0.001310	98.57	L-isoleucine, lipid methyl, cholesterol (ester)
1.162	597	0.145	3.78e-05	0.001310	98.97	Lipid methylene
1.172	596	0.135	4.08e-05	0.001310	98.57	Lipid methylene
1.932	520	0.070	4.30e-05	0.001310	98.04	Acetic acid
0.992	614	0.112	4.39e-05	0.001310	98.04	Leucine, lipid methyl, cholesterol (ester)
1.032	610	0.122	4.75e-05	0.001310	98.04	L-isoleucine, lipid methyl, cholesterol (ester)
2.312	482	0.101	4.85e-05	0.001310	98.04	Lipid (methylene carbonyl)
1.122	601	0.118	4.96e-05	0.001310	98.04	Unknown
0.962	617	0.116	7.53e-05	0.001850	97.36	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.142	599	0.097	7.58e-05	0.001850	97.36	Unknown
1.482	565	0.093	8.77e-05	0.002070	96.50	Alanine
2.302	483	0.108	1.03e-04	0.002350	96.50	Lipid (methylene carbonyl)
1.722	541	0.056	1.12e-04	0.002460	97.36	Leucine, lysine
1.182	595	0.107	1.31e-04	0.002790	96.50	Lipid methylene
1.002	613	0.090	1.51e-04	0.003000	96.50	Valine, lipid methyl, cholesterol (ester)
1.892	524	0.057	1.51e-04	0.003000	95.41	Overlap of multiple minor compounds
7.312	217	0.076	1.54e-04	0.003000	92.43	Unknown
2.322	481	0.086	1.99e-04	0.003740	95.41	Lipid (methylene carbonyl)
1.132	600	0.084	2.20e-04	0.003950	95.41	Unknown
4.092	328	0.126	2.26e-04	0.003950	92.43	Unknown
1.212	592	0.139	2.28e-04	0.003950	96.50	Lipid methylene
1.972	516	0.077	2.49e-04	0.004220	94.06	Lipid allylic
1.092	604	0.095	2.65e-04	0.004350	95.41	Unknown
1.082	605	0.090	2.76e-04	0.004350	94.06	Unknown
6.832	265	0.079	2.77e-04	0.004350	90.47	Unknown
0.852	628	0.075	2.91e-04	0.004360	94.06	Cholesterol, lipid methyl
1.982	515	0.086	2.95e-04	0.004360	94.06	Lipid allylic
0.822	631	0.105	2.97e-04	0.004360	94.06	Cholesterol, lipid methyl

1.042	609	0.131	3.08e-04	0.004370	94.06	L-isoleucine, lipid methyl, cholesterol (ester)
1.072	606	0.085	3.12e-04	0.004370	95.41	Valine
2.352	478	0.064	3.18e-04	0.004370	92.43	Proline, glutamic acid
4.082	329	0.123	3.46e-04	0.004640	90.47	Creatinine
7.292	219	0.066	3.61e-04	0.004640	90.47	Unknown
2.132	500	0.059	3.64e-04	0.004640	92.43	Glutamine
2.292	484	0.117	3.69e-04	0.004640	94.06	Lipid (methylene carbonyl)
1.472	566	0.092	3.72e-04	0.004640	94.06	Lipid methylene
6.822	266	0.078	4.07e-04	0.004980	90.47	Unknown
7.282	220	0.066	4.21e-04	0.005040	90.47	Unknown
7.322	216	0.071	4.27e-04	0.005040	88.16	Unknown
1.012	612	0.094	4.42e-04	0.005110	92.43	Valine, lipid methyl, cholesterol (ester)
7.262	222	0.075	4.75e-04	0.005410	88.16	Unknown
2.652	448	0.045	4.88e-04	0.005460	90.47	Unknown
1.912	522	0.058	5.42e-04	0.005960	92.43	Overlap of multiple minor compounds
1.672	546	0.088	5.55e-04	0.006010	90.47	Unknown, arginine
1.062	607	0.091	5.75e-04	0.006060	94.06	Valine
7.272	221	0.070	5.78e-04	0.006060	88.16	Unknown
1.992	514	0.086	5.91e-04	0.006090	90.47	Lipid allylic
7.302	218	0.065	6.13e-04	0.006220	88.16	Unknown
1.962	517	0.063	6.49e-04	0.006490	90.47	Lipid allylic
2.362	477	0.074	6.87e-04	0.006770	90.47	Proline, glutamic acid
1.952	518	0.056	7.03e-04	0.006830	90.47	Acetic acid
7.232	225	0.065	7.62e-04	0.007290	85.48	Unknown
1.922	521	0.059	7.99e-04	0.007530	90.47	Overlap of multiple minor compounds
2.282	485	0.138	8.19e-04	0.007600	90.47	Lipid (methylene carbonyl)
2.252	488	0.157	8.29e-04	0.007600	88.16	Lipid (methylene carbonyl), acetone
1.252	588	0.106	8.97e-04	0.008110	90.47	Lipid methylene
2.002	513	0.084	1.03e-03	0.009160	88.16	Lipid allylic
0.952	618	0.112	1.04e-03	0.009160	90.47	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.872	526	0.051	1.06e-03	0.009200	88.16	Overlap of multiple minor compounds
4.102	327	0.154	1.07e-03	0.009200	85.48	Unknown
2.202	493	0.110	1.09e-03	0.009260	88.16	Lipid (methylene carbonyl)
7.252	223	0.069	1.13e-03	0.009430	85.48	Unknown
1.422	571	0.140	1.38e-03	0.011400	88.16	Lipid methylene
0.812	632	0.107	1.45e-03	0.011800	88.16	Cholesterol, lipid methyl
2.212	492	0.134	1.62e-03	0.013000	88.16	Lipid (methylene carbonyl)
1.432	570	0.111	1.64e-03	0.013100	88.16	Lipid methylene

1.442	569	0.101	1.86e-03	0.014500	85.48	Lipid methylene
1.732	540	0.049	1.87e-03	0.014500	88.16	Leucine, lysine
1.412	572	0.167	2.00e-03	0.015300	85.48	Lipid methylene
7.102	238	0.111	2.07e-03	0.015700	88.16	Unknown
7.722	176	-0.147	2.13e-03	0.016000	88.16	Unknown
2.192	494	0.081	2.16e-03	0.016000	82.42	Lipid (methylene carbonyl)
1.102	603	0.079	2.25e-03	0.016500	85.48	Unknown
3.062	407	0.066	2.33e-03	0.016700	82.42	Creatinine
7.182	230	0.075	2.33e-03	0.016700	82.42	Unknown
1.452	568	0.094	2.35e-03	0.016700	85.48	Lipid methylene
6.852	263	0.067	2.43e-03	0.017100	82.42	Unknown
1.662	547	0.090	2.46e-03	0.017100	82.42	Lipids (?)
2.372	476	0.070	2.81e-03	0.019100	82.42	Proline, glutamic acid
2.332	480	0.070	2.81e-03	0.019100	82.42	Proline, glutamic acid
7.332	215	0.055	2.88e-03	0.019400	78.98	Phenylalanine
0.802	633	0.100	2.91e-03	0.019400	82.42	Cholesterol, lipid methyl
1.752	538	0.052	3.09e-03	0.020300	85.48	Leucine, lysine
0.942	619	0.127	3.11e-03	0.020300	82.42	Cholesterol, lipid methyl
2.382	475	0.069	3.23e-03	0.020900	82.42	Proline, glutamic acid
1.462	567	0.081	3.29e-03	0.021100	82.42	Lipid methylene
0.792	634	0.088	3.40e-03	0.021400	78.98	Cholesterol, lipid methyl
1.402	573	0.196	3.42e-03	0.021400	82.42	Lipid methylene
2.112	502	0.067	3.44e-03	0.021400	82.42	Lipid allylic
2.012	512	0.079	3.54e-03	0.021800	78.98	Lipid allylic
7.732	175	-0.121	3.61e-03	0.022100	82.42	Unknown
1.942	519	0.099	3.86e-03	0.023200	82.42	Acetic acid
7.242	224	0.057	3.87e-03	0.023200	75.18	Unknown
0.742	639	0.127	4.01e-03	0.023800	82.42	Unkown
2.272	486	0.151	4.08e-03	0.024100	78.98	Lipid (methylene carbonyl)
2.222	491	0.134	4.23e-03	0.024500	78.98	Lipid (methylene carbonyl)
2.342	479	0.062	4.23e-03	0.024500	78.98	Proline, glutamic acid
2.232	490	0.137	5.00e-03	0.028700	78.98	Lipid (methylene carbonyl)
1.392	574	0.199	5.28e-03	0.030000	78.98	Lipid methylene
1.542	559	0.095	5.39e-03	0.030400	75.18	Lipids (?)
1.742	539	0.046	5.60e-03	0.031300	78.98	Leucine, lysine
2.102	503	0.055	5.93e-03	0.032900	75.18	Lipid allylic
1.552	558	0.117	6.41e-03	0.035000	75.18	Lipids (?)
2.082	505	0.084	6.42e-03	0.035000	75.18	Lipid allylic
1.112	602	0.068	6.53e-03	0.035200	78.98	Unknown
4.312	306	0.075	6.55e-03	0.035200	75.18	Lipid alpha-methylene to carboxyl, lipid glycerine
2.242	489	0.140	7.09e-03	0.037800	71.04	Lipid (methylene carbonyl), acetone
4.322	305	0.067	7.31e-03	0.038600	71.04	Lipid alpha-methylene to carboxyl, lipid glycerine

1.522	561	0.057	7.39e-03	0.038700	71.04	Lipids (?)
1.382	575	0.192	7.45e-03	0.038700	75.18	Lipid methylene
1.562	557	0.130	7.64e-03	0.039200	71.04	Lipids (?)
0.752	638	0.087	7.66e-03	0.039200	75.18	Cholesterol, lipid methyl
8.152	133	0.172	7.77e-03	0.039500	78.98	Unknown
0.932	620	0.138	8.23e-03	0.041500	75.18	Cholesterol, lipid methyl
1.192	594	0.136	8.39e-03	0.041900	78.98	Lipid methylene
4.112	326	0.123	8.77e-03	0.043500	66.59	Proline, lactic acid
1.652	548	0.096	8.87e-03	0.043700	71.04	Lipids (?)
6.802	268	0.071	9.47e-03	0.046300	71.04	Unknown
2.262	487	0.154	1.01e-02	0.049100	71.04	Lipid (methylene carbonyl)

Table 7.15: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from glomerulonephritis from those suffering from systemic diseases.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
7.432	205	-0.111	1.23e-12	8.09e-10	100.00	Phenylalanine
3.042	409	-0.035	4.51e-10	1.49e-07	100.00	Lysine, unknown
7.202	228	-0.097	3.64e-08	8.01e-06	99.99	Tyrosine
2.552	458	-0.067	1.08e-07	1.79e-05	99.73	Citric acid
2.902	423	-0.063	4.37e-07	4.21e-05	99.85	Unknown
3.782	359	-0.079	4.80e-07	4.21e-05	99.99	D-glucose, alanine, glutamine, arginine
2.572	456	-0.055	4.86e-07	4.21e-05	99.53	$CaEDTA^{2-}$, citric acid
3.052	408	-0.042	5.24e-07	4.21e-05	99.92	Creatinine
3.792	358	-0.060	5.75e-07	4.21e-05	99.99	D-glucose, alanine
2.972	416	-0.031	7.78e-07	5.14e-05	99.85	Unknown
3.472	377	-0.110	8.84e-07	5.30e-05	99.99	D-glucose
3.802	357	-0.083	1.69e-06	9.32e-05	99.98	D-glucose, alanine
2.892	424	-0.058	1.97e-06	9.81e-05	99.73	Lipid diallylic
3.752	362	-0.101	2.08e-06	9.81e-05	99.96	D-glucose, glutamic acid
6.922	256	-0.084	2.33e-06	9.82e-05	99.73	Tyrosine
3.852	352	-0.098	2.42e-06	9.82e-05	99.96	D-glucose, unknown
7.442	204	-0.066	2.53e-06	9.82e-05	99.85	Phenylalanine
3.912	346	-0.091	3.54e-06	1.27e-04	99.96	D-glucose, betaine, unknown

3.932	344	-0.093	3.65e-06	1.27e-04	99.96	D-glucose
3.422	382	-0.108	3.89e-06	1.28e-04	99.92	D-glucose, carnitine, taurine, proline
1.492	564	-0.078	4.51e-06	1.42e-04	99.53	Alanine
1.052	608	-0.119	4.78e-06	1.43e-04	99.53	Valine
7.562	192	0.104	5.11e-06	1.43e-04	99.85	Unknown
1.792	534	0.069	5.36e-06	1.43e-04	99.53	Unknown
2.932	420	-0.058	5.41e-06	1.43e-04	99.53	Unknown
1.502	563	-0.079	6.08e-06	1.54e-04	99.53	Alanine
2.692	444	-0.048	8.10e-06	1.98e-04	97.97	Citric acid
3.452	379	-0.089	8.78e-06	2.07e-04	99.92	D-glucose, carnitine, proline
2.942	419	-0.032	1.01e-05	2.31e-04	99.21	N,N-dimethylglycine
3.842	353	-0.087	1.09e-05	2.41e-04	99.85	D-glucose, unknown
3.552	369	-0.092	1.44e-05	3.06e-04	99.85	D-glucose, myo-inositol
3.732	364	-0.088	1.57e-05	3.23e-04	99.73	D-glucose, unknown
3.562	368	-0.089	1.76e-05	3.53e-04	99.85	D-glucose
2.672	446	-0.070	2.01e-05	3.89e-04	97.97	Citric acid
4.072	330	-0.096	2.24e-05	4.23e-04	98.71	Creatinine
3.482	376	-0.096	2.64e-05	4.84e-04	99.73	D-glucose
2.372	476	-0.068	2.90e-05	5.18e-04	98.71	Proline, glutamic acid
3.542	370	-0.086	3.06e-05	5.31e-04	98.71	D-glucose, myo-inositol
1.782	535	0.066	3.23e-05	5.46e-04	98.71	Unknown
1.012	612	-0.077	3.44e-05	5.68e-04	98.71	Valine, lipid methyl, cholesterol (ester)
3.512	373	-0.096	3.86e-05	6.22e-04	99.73	D-glucose
3.572	367	-0.081	4.24e-05	6.67e-04	99.53	D-glucose, glycine
2.382	475	-0.066	4.40e-05	6.76e-04	97.97	Proline, glutamic acid
3.872	350	-0.071	4.51e-05	6.76e-04	99.53	D-glucose, unknown
3.742	363	-0.086	5.86e-05	8.46e-04	99.53	D-glucose, leucine
3.442	380	-0.090	5.89e-05	8.46e-04	99.53	D-glucose, carnitine, taurine, proline
3.722	365	-0.073	6.19e-05	8.69e-04	99.53	D-glucose, N,N-dimethylglycine
3.822	355	-0.060	6.65e-05	9.15e-04	98.71	Unknown
2.542	459	-0.052	6.89e-05	9.17e-04	96.90	Unknown
2.502	463	0.069	7.04e-05	9.17e-04	97.97	Unknown
3.762	361	-0.079	7.09e-05	9.17e-04	99.53	D-glucose, arginine, glutamine, glutamic acid
7.572	191	0.101	7.53e-05	9.55e-04	98.71	Unknown
3.502	374	-0.091	8.75e-05	1.09e-03	99.21	D-glucose
3.522	372	-0.100	9.36e-05	1.14e-03	98.71	D-glucose
3.412	383	-0.088	9.81e-05	1.18e-03	99.21	D-glucose, carnitine, taurine, proline
3.832	354	-0.051	1.02e-04	1.20e-03	97.97	Unknown
3.432	381	-0.088	1.07e-04	1.24e-03	99.21	D-glucose, carnitine, taurine, proline

3.032	410	-0.023	1.16e-04	1.31e-03	97.97	Lysine, unknown
7.402	208	-0.049	1.24e-04	1.38e-03	97.97	Phenylalanine
3.532	371	-0.079	1.31e-04	1.45e-03	99.21	D-glucose
1.072	606	-0.062	1.38e-04	1.47e-03	96.90	Valine
7.412	207	0.090	1.38e-04	1.47e-03	97.97	Phenylalanine
1.002	613	-0.062	1.53e-04	1.61e-03	96.90	Valine, lipid methyl, cholesterol (ester)
2.872	426	-0.083	1.56e-04	1.61e-03	95.39	Lipid diallylic
3.862	351	-0.081	1.67e-04	1.70e-03	99.21	D-glucose, unknown
1.802	533	0.047	1.74e-04	1.74e-03	95.39	Unknown
3.812	356	-0.063	2.75e-04	2.71e-03	97.97	D-glucose
2.662	447	-0.071	3.27e-04	3.18e-03	93.34	Citric acid
0.882	625	0.071	3.32e-04	3.18e-03	95.39	Cholesterol, lipid methyl
7.552	193	0.088	4.03e-04	3.80e-03	96.90	Unknown
4.152	322	-0.069	4.46e-04	4.11e-03	95.39	Proline, lactic acid
6.882	260	0.050	4.48e-04	4.11e-03	93.34	Unknown
3.772	360	-0.065	4.76e-04	4.30e-03	96.90	D-glucose, alanine, glutamine, arginine
0.892	624	0.079	4.84e-04	4.32e-03	93.34	Cholesterol, lipid methyl
7.312	217	0.048	5.36e-04	4.71e-03	93.34	Unknown
2.362	477	-0.052	5.51e-04	4.73e-03	93.34	Proline, glutamic acid
3.982	339	0.040	5.52e-04	4.73e-03	93.34	Unknown
7.292	219	0.044	5.79e-04	4.90e-03	93.34	Unknown
2.562	457	-0.029	6.29e-04	5.19e-03	90.64	$CaEDTA^{2-}$
2.862	427	-0.094	6.30e-04	5.19e-03	93.34	Lipid diallylic
6.912	257	-0.061	6.37e-04	5.19e-03	93.34	Tyrosine
3.492	375	-0.079	6.82e-04	5.49e-03	96.90	D-glucose
6.852	263	0.052	6.94e-04	5.52e-03	93.34	Unknown
4.142	323	-0.084	7.91e-04	6.22e-03	93.34	Proline, lactic acid
7.302	218	0.044	8.01e-04	6.22e-03	90.64	Unknown
3.882	349	-0.065	8.47e-04	6.50e-03	96.90	D-glucose, unknown
3.152	398	-0.022	9.37e-04	7.11e-03	87.22	$CaEDTA^{2-}$
7.352	213	-0.037	1.08e-03	8.09e-03	90.64	Phenylalanine
4.132	324	-0.087	1.09e-03	8.11e-03	90.64	Proline, lactic acid
7.582	190	0.080	1.24e-03	9.10e-03	90.64	Unknown
1.082	605	-0.054	1.41e-03	1.03e-02	87.22	Unknown
1.812	532	0.036	1.53e-03	1.09e-02	87.22	Unknown
3.312	393	-0.066	1.53e-03	1.09e-02	90.64	Unknown
7.032	245	0.080	1.56e-03	1.09e-02	90.64	Unknown
8.492	99	-0.246	1.57e-03	1.09e-02	90.64	Formic acid
3.402	384	-0.059	1.75e-03	1.21e-02	90.64	Unknown
1.822	531	0.034	1.78e-03	1.21e-02	87.22	Unknown
4.122	325	-0.073	1.81e-03	1.22e-02	87.22	Proline, lactic acid
2.982	415	-0.019	2.05e-03	1.37e-02	87.22	Unknown
2.912	422	-0.039	2.08e-03	1.37e-02	87.22	Unknown

1.832	530	0.032	2.20e-03	1.44e-02	87.22	Unknown
2.882	425	-0.054	2.29e-03	1.48e-02	87.22	Lipid diallylic
7.042	244	0.070	2.31e-03	1.48e-02	87.22	Unknown
3.122	401	-0.022	2.42e-03	1.54e-02	78.01	$CaEDTA^{2-}$
1.102	603	-0.053	3.01e-03	1.89e-02	83.01	Unknown
8.162	132	0.139	3.82e-03	2.38e-02	83.01	Unknown
2.682	445	-0.031	4.01e-03	2.47e-02	78.01	Citric acid
8.382	110	0.217	4.25e-03	2.60e-02	78.01	Unknown
8.452	103	0.214	4.29e-03	2.60e-02	83.01	Unknown
7.022	246	0.097	4.44e-03	2.66e-02	83.01	Unknown
7.392	209	-0.034	4.48e-03	2.67e-02	83.01	Phenylalanine
2.742	439	0.038	4.74e-03	2.79e-02	78.01	Lipid diallylic
1.912	522	-0.032	4.97e-03	2.90e-02	78.01	Overlap of multiple minor compounds
2.392	474	-0.049	5.11e-03	2.94e-02	83.01	Unknown
2.852	428	-0.081	5.13e-03	2.94e-02	78.01	Lipid diallylic
4.052	332	-0.070	5.34e-03	3.04e-02	83.01	Unknown
7.592	189	0.065	5.46e-03	3.08e-02	83.01	Unknown
8.732	75	-0.243	5.71e-03	3.19e-02	78.01	Unknown
8.212	127	0.081	5.81e-03	3.22e-02	83.01	Unknown
7.232	225	0.036	7.41e-03	4.06e-02	72.25	Unknown
8.252	123	0.052	7.43e-03	4.06e-02	78.01	Unknown
7.062	242	0.050	7.92e-03	4.29e-02	78.01	Unknown
8.232	125	0.053	8.25e-03	4.43e-02	78.01	Unknown
1.742	539	-0.030	8.46e-03	4.50e-02	72.25	Leucine, lysine
4.002	337	0.027	8.65e-03	4.57e-02	78.01	Unknown
7.052	243	0.045	8.75e-03	4.58e-02	78.01	Unknown
6.862	262	0.047	8.87e-03	4.61e-02	72.25	Unknown
2.722	441	-0.026	9.30e-03	4.80e-02	72.25	$MgEDTA^{2-}$
7.282	220	0.034	9.39e-03	4.80e-02	72.25	Unknown
2.492	464	0.032	9.45e-03	4.80e-02	78.01	Glutamine

Table 7.16: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from glomerulonephritis from those suffering from hypertensive nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	ID	log(FC)	*P*-value unadjusted	*P*-value B/H-adjusted	Statistical power	Identified compounds
6.922	256	-0.149	0.000192	0.0487	95.34	Tyrosine
7.222	226	-0.129	0.000198	0.0487	94.50	Tyrosine
6.912	257	-0.148	0.000221	0.0487	95.34	Tyrosine

Table 7.17: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), *p*-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from hereditary diseases from those suffering from interstitial nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	ID	log(FC)	*P*-value unadjusted	*P*-value B/H-adjusted	Statistical power	Identified compounds
3.062	407	0.163	9.34e-08	6.17e-05	99.96	Tyrosine
4.072	330	0.185	6.38e-05	2.11e-02	98.69	Tyrosine
4.132	324	-0.207	1.50e-04	3.30e-02	94.53	Tyrosine

Table 7.18: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), *p*-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from hereditary diseases from those suffering from systemic diseases.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	ID	log(FC)	*P*-value unadjusted	*P*-value B/H-adjusted	Statistical power	Identified compounds
4.122	325	-0.215	1.86e-07	0.000123	99.96	Proline, lactic acid
1.232	590	-0.158	4.53e-07	0.000150	99.94	Lipid methylene
1.112	602	-0.148	1.14e-06	0.000173	99.88	Unknown
7.222	226	-0.132	1.17e-06	0.000173	99.83	Tyrosine
1.242	589	-0.168	1.31e-06	0.000173	99.88	Lipid methylene

4.132	324	-0.224	1.79e-06	0.000190	99.76	Proline, lactic acid
6.922	256	-0.146	2.81e-06	0.000190	99.76	Tyrosine
3.962	341	-0.078	3.12e-06	0.000190	99.76	Unknown
1.342	579	-0.238	3.25e-06	0.000190	99.67	Lipid methylene, lactic acid, threonine
1.152	598	-0.157	3.42e-06	0.000190	99.39	Lipid methylene
7.202	228	-0.143	3.49e-06	0.000190	99.67	Tyrosine
1.102	603	-0.145	3.53e-06	0.000190	99.67	Unknown
1.362	577	-0.258	3.94e-06	0.000190	99.67	Lipid methylene, lactic acid, threonine
4.142	323	-0.202	4.03e-06	0.000190	99.67	Proline, lactic acid
1.132	600	-0.125	5.46e-06	0.000240	99.67	Unknown
1.352	578	-0.232	7.58e-06	0.000313	99.39	Lipid methylene, lactic acid, threonine
6.912	257	-0.139	8.64e-06	0.000335	99.39	Tyrosine
1.222	591	-0.173	9.92e-06	0.000364	98.60	Lipid methylene
1.252	588	-0.169	1.19e-05	0.000414	99.39	Lipid methylene
7.212	227	-0.125	1.32e-05	0.000434	99.39	Tyrosine
1.162	597	-0.182	1.97e-05	0.000600	97.03	Lipid methylene
2.382	475	-0.121	2.00e-05	0.000600	98.93	Proline, glutamic acid
4.152	322	-0.146	2.46e-05	0.000663	99.19	Proline, lactic acid
1.172	596	-0.168	2.49e-05	0.000663	97.03	Lipid methylene
2.312	482	-0.126	2.51e-05	0.000663	99.19	Lipid (methylene carbonyl)
4.312	306	-0.140	2.95e-05	0.000748	98.93	Lipid alpha-methylene to carboxyl, lipid glycerine
2.282	485	-0.206	3.61e-05	0.000883	98.60	Lipid (methylene carbonyl)
2.292	484	-0.164	4.00e-05	0.000943	98.60	Lipid (methylene carbonyl)
1.142	599	-0.122	4.20e-05	0.000957	98.60	Unknown
1.002	613	-0.117	5.06e-05	0.001110	98.60	Valine, lipid methyl, cholesterol (ester)
1.972	516	-0.102	5.47e-05	0.001160	98.93	Lipid allylic
2.362	477	-0.107	5.71e-05	0.001180	98.60	Proline, glutamic acid
0.852	628	-0.101	6.67e-05	0.001320	98.60	Cholesterol, lipid methyl
1.462	567	-0.133	6.79e-05	0.001320	98.18	Lipid methylene
2.272	486	-0.253	7.32e-05	0.001380	98.18	Lipid (methylene carbonyl)
1.452	568	-0.148	7.96e-05	0.001460	98.18	Lipid methylene
4.322	305	-0.118	8.49e-05	0.001510	98.18	Lipid alpha-methylene to carboxyl, lipid glycerine
0.842	629	-0.100	8.77e-05	0.001520	98.18	Cholesterol, lipid methyl
1.982	515	-0.112	9.20e-05	0.001560	98.18	Lipid allylic
1.682	545	-0.100	9.82e-05	0.001590	98.60	Unknown, arginine
1.962	517	-0.087	9.89e-05	0.001590	98.18	Lipid allylic
2.372	476	-0.111	1.02e-04	0.001610	97.66	Proline, glutamic acid
2.302	483	-0.129	1.19e-04	0.001780	97.66	Lipid (methylene carbonyl)

0.962	617	-0.136	1.19e-04	0.001780	97.66	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.092	604	-0.120	1.28e-04	0.001850	97.03	Unknown
1.262	587	-0.160	1.29e-04	0.001850	97.66	Lipid methylene
0.952	618	-0.159	1.33e-04	0.001860	97.03	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
2.122	501	-0.101	1.37e-04	0.001890	97.66	Lipid allylic
1.692	544	-0.085	1.45e-04	0.001960	98.18	Unknown, arginine
1.482	565	-0.109	1.48e-04	0.001960	97.66	Alanine
2.822	431	-0.201	1.63e-04	0.002080	97.03	Lipid diallylic
1.412	572	-0.246	1.64e-04	0.002080	97.03	Lipid methylene
1.402	573	-0.305	1.69e-04	0.002080	97.03	Lipid methylene
3.452	379	-0.132	1.71e-04	0.002080	98.60	D-glucose, carnitine, proline
2.862	427	-0.181	1.73e-04	0.002080	97.03	Lipid diallylic
1.082	605	-0.112	1.86e-04	0.002190	97.03	Unknown
0.942	619	-0.195	1.89e-04	0.002190	96.25	Cholesterol, lipid methyl
1.672	546	-0.115	1.95e-04	0.002210	97.03	Unknown, arginine
1.422	571	-0.197	2.02e-04	0.002260	96.25	Lipid methylene
2.322	481	-0.103	2.23e-04	0.002450	97.03	Lipid (methylene carbonyl)
2.262	487	-0.267	2.28e-04	0.002460	96.25	Lipid (methylene carbonyl)
0.972	616	-0.118	2.39e-04	0.002520	97.03	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.522	561	-0.094	2.41e-04	0.002520	97.03	Lipids (?)
2.852	428	-0.187	2.45e-04	0.002520	96.25	Lipid diallylic
3.812	356	-0.110	2.49e-04	0.002530	97.03	D-glucose
1.472	566	-0.114	2.60e-04	0.002590	96.25	Lipid methylene
4.332	304	-0.106	2.63e-04	0.002590	96.25	Lipid alpha-methylene to carboxyl, lipid glycerine
2.252	488	-0.207	2.70e-04	0.002620	96.25	Lipid (methylene carbonyl), acetone
1.992	514	-0.110	2.76e-04	0.002640	97.03	Lipid allylic
1.662	547	-0.131	2.83e-04	0.002650	96.25	Lipids (?)
1.392	574	-0.313	2.85e-04	0.002650	95.32	Lipid methylene
3.882	349	-0.124	3.02e-04	0.002710	97.66	D-glucose, unknown
1.492	564	-0.107	3.03e-04	0.002710	96.25	Alanine
0.932	620	-0.228	3.04e-04	0.002710	95.32	Cholesterol, lipid methyl
2.832	430	-0.189	3.29e-04	0.002890	95.32	Lipid diallylic
2.132	500	-0.071	3.37e-04	0.002890	97.03	Glutamine
1.212	592	-0.163	3.38e-04	0.002890	91.39	Lipid methylene
3.782	359	-0.098	3.44e-04	0.002900	97.66	D-glucose, alanine, glutamine, arginine
3.802	357	-0.108	3.47e-04	0.002900	97.66	D-glucose, alanine
1.272	586	-0.160	3.61e-04	0.002960	95.32	Lipid methylene
4.302	307	-0.133	3.63e-04	0.002960	95.32	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine

1.012	612	-0.116	3.81e-04	0.003070	95.32	Valine, lipid methyl, cholesterol (ester)
1.442	569	-0.138	4.03e-04	0.003190	95.32	Lipid methylene
3.762	361	-0.124	4.12e-04	0.003190	97.03	D-glucose, arginine, glutamine, glutamic acid
0.992	614	-0.117	4.13e-04	0.003190	95.32	Leucine, lipid methyl, cholesterol (ester)
2.872	426	-0.136	4.16e-04	0.003190	95.32	Lipid diallylic
1.432	570	-0.150	4.27e-04	0.003240	95.32	Lipid methylene
2.652	448	-0.055	4.47e-04	0.003350	95.32	Unknown
1.622	551	-0.245	4.62e-04	0.003400	94.21	Lipids (?)
1.612	552	-0.264	4.64e-04	0.003400	94.21	Lipids (?)
3.442	380	-0.137	4.73e-04	0.003410	97.03	D-glucose, carnitine, taurine, proline
1.502	563	-0.107	4.76e-04	0.003410	94.21	Alanine
2.842	429	-0.189	4.88e-04	0.003440	95.32	Lipid diallylic
3.922	345	-0.097	4.89e-04	0.003440	96.25	D-glucose, unknown
3.842	353	-0.120	5.18e-04	0.003600	97.03	D-glucose, unknown
3.772	360	-0.113	5.27e-04	0.003620	95.32	D-glucose, alanine, glutamine, arginine
1.182	595	-0.117	5.39e-04	0.003660	94.21	Lipid methylene
1.072	606	-0.098	5.51e-04	0.003660	94.21	Valine
1.382	575	-0.301	5.53e-04	0.003660	94.21	Lipid methylene
1.702	543	-0.068	5.54e-04	0.003660	95.32	Unknown, arginine
1.532	560	-0.110	5.70e-04	0.003720	94.21	Lipids (?)
2.812	432	-0.183	5.86e-04	0.003790	94.21	Lipid diallylic
4.342	303	-0.100	6.23e-04	0.003990	94.21	Lipid alpha-methylene to carboxyl, lipid glycerine
0.982	615	-0.110	6.54e-04	0.004150	94.21	Leucine, lipid methyl, cholesterol (ester)
3.742	363	-0.127	6.98e-04	0.004370	96.25	D-glucose, leucine
3.482	376	-0.136	7.01e-04	0.004370	96.25	D-glucose
3.872	350	-0.103	7.12e-04	0.004390	95.32	D-glucose, unknown
1.632	550	-0.207	7.32e-04	0.004440	92.91	Lipids (?)
3.972	340	-0.057	7.33e-04	0.004440	94.21	Unknown
3.432	381	-0.134	7.41e-04	0.004440	96.25	D-glucose, carnitine, taurine, proline
3.852	352	-0.122	7.51e-04	0.004470	95.32	D-glucose, unknown
1.512	562	-0.106	7.69e-04	0.004530	94.21	Alanine
3.862	351	-0.126	7.83e-04	0.004560	95.32	D-glucose, unknown
1.332	580	-0.268	7.88e-04	0.004560	92.91	Lipid methylene
2.902	423	-0.073	8.11e-04	0.004650	94.21	Unknown
3.502	374	-0.137	8.21e-04	0.004670	95.32	D-glucose
3.912	346	-0.115	8.49e-04	0.004760	95.32	D-glucose, betaine, unknown
1.652	548	-0.148	8.52e-04	0.004760	92.91	Lipids (?)

3.722	365	-0.106	8.94e-04	0.004920	95.32	D-glucose, N,N-dimethylglycine
1.602	553	-0.259	9.00e-04	0.004920	92.91	Lipids (?)
1.552	558	-0.172	9.03e-04	0.004920	92.91	Lipids (?)
3.412	383	-0.132	9.12e-04	0.004920	95.32	D-glucose, carnitine, taurine, proline
2.222	491	-0.188	9.17e-04	0.004920	92.91	Lipid (methylene carbonyl)
3.532	371	-0.120	9.75e-04	0.005150	95.32	D-glucose
3.492	375	-0.134	9.79e-04	0.005150	95.32	D-glucose
1.642	549	-0.175	9.84e-04	0.005150	91.39	Lipids (?)
1.372	576	-0.286	1.01e-03	0.005230	91.39	Lipid methylene
0.922	621	-0.222	1.04e-03	0.005330	91.39	Cholesterol, lipid methyl
3.472	377	-0.129	1.04e-03	0.005330	95.32	D-glucose
3.512	373	-0.134	1.05e-03	0.005330	95.32	D-glucose
0.832	630	-0.092	1.06e-03	0.005330	91.39	Cholesterol, lipid methyl
2.112	502	-0.090	1.07e-03	0.005330	91.39	Lipid allylic
3.932	344	-0.115	1.09e-03	0.005400	94.21	D-glucose
1.302	583	-0.223	1.13e-03	0.005570	91.39	Lipid methylene
1.542	559	-0.135	1.14e-03	0.005570	91.39	Lipids (?)
2.392	474	-0.099	1.15e-03	0.005590	91.39	Unknown
2.642	449	-0.058	1.21e-03	0.005800	92.91	Unknown
3.792	358	-0.068	1.21e-03	0.005800	95.32	D-glucose, alanine
2.002	513	-0.099	1.31e-03	0.006210	91.39	Lipid allylic
1.312	582	-0.233	1.32e-03	0.006210	89.64	Lipid methylene
1.122	601	-0.112	1.34e-03	0.006230	89.64	Unknown
3.822	355	-0.084	1.34e-03	0.006230	92.91	Unknown
2.792	434	-0.136	1.35e-03	0.006230	91.39	Lipid diallylic
1.062	607	-0.103	1.37e-03	0.006260	89.64	Valine
3.572	367	-0.110	1.45e-03	0.006600	94.21	D-glucose, glycine
3.752	362	-0.118	1.47e-03	0.006650	92.91	D-glucose, glutamic acid
1.022	611	-0.106	1.51e-03	0.006790	89.64	L-isoleucine, lipid methyl, cholesterol (ester)
1.322	581	-0.241	1.55e-03	0.006870	89.64	Lipid methylene
1.952	518	-0.064	1.55e-03	0.006870	89.64	Acetic acid
4.112	326	-0.180	1.57e-03	0.006870	91.39	Proline, lactic acid
3.462	378	-0.100	1.58e-03	0.006870	91.39	D-glucose
3.562	368	-0.114	1.58e-03	0.006870	92.91	D-glucose
4.102	327	-0.180	1.61e-03	0.006960	91.39	Unknown
4.292	308	-0.114	1.63e-03	0.007010	89.64	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.232	490	-0.185	1.70e-03	0.007230	89.64	Lipid (methylene carbonyl)
3.552	369	-0.117	1.73e-03	0.007330	92.91	D-glucose, myo-inositol
3.892	348	-0.060	1.80e-03	0.007560	89.64	Unknown
2.212	492	-0.160	1.86e-03	0.007760	89.64	Lipid (methylene carbonyl)
9.402	8	0.476	1.87e-03	0.007760	85.39	Unknown
1.562	557	-0.184	1.91e-03	0.007880	89.64	Lipids (?)

3.942	343	-0.122	2.03e-03	0.008260	91.39	D-glucose
1.282	585	-0.150	2.03e-03	0.008260	87.64	Lipid methylene
1.712	542	-0.055	2.08e-03	0.008420	89.64	Leucine, lysine
2.242	489	-0.194	2.14e-03	0.008610	87.64	Lipid (methylene carbonyl), acetone
1.592	554	-0.234	2.17e-03	0.008680	87.64	Lipids (?)
2.102	503	-0.073	2.21e-03	0.008780	87.64	Lipid allylic
1.842	529	-0.060	2.25e-03	0.008870	89.64	Unknown
2.082	505	-0.113	2.38e-03	0.009340	87.64	Lipid allylic
0.822	631	-0.105	2.74e-03	0.010600	85.39	Cholesterol, lipid methyl
2.332	480	-0.085	2.74e-03	0.010600	85.39	Proline, glutamic acid
1.572	556	-0.194	2.75e-03	0.010600	87.64	Lipids (?)
3.422	382	-0.123	2.77e-03	0.010600	89.64	D-glucose, carnitine, taurine, proline
3.522	372	-0.134	2.85e-03	0.010900	87.64	D-glucose
4.242	313	-0.163	2.88e-03	0.010900	87.64	Unknown
2.802	433	-0.138	2.99e-03	0.011300	87.64	Lipid diallylic
1.582	555	-0.211	3.09e-03	0.011600	85.39	Lipids (?)
4.252	312	-0.128	3.55e-03	0.013200	85.39	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
3.732	364	-0.104	3.56e-03	0.013200	89.64	D-glucose, unknown
0.862	627	-0.083	3.81e-03	0.014000	87.64	Cholesterol, lipid methyl
4.352	302	-0.088	4.17e-03	0.015300	82.88	Lipid alpha-methylene to carboxyl, lipid glycerine
1.832	530	-0.051	5.06e-03	0.018500	85.39	Unknown
3.062	407	0.073	5.11e-03	0.018500	85.39	Creatinine
2.782	435	-0.121	5.15e-03	0.018600	82.88	Lipid diallylic
1.192	594	-0.175	5.24e-03	0.018800	73.82	Lipid methylene
1.862	527	-0.055	5.47e-03	0.019500	82.88	Unknown
7.272	221	-0.068	5.50e-03	0.019500	85.39	Unknown
2.622	451	-0.040	5.66e-03	0.020000	82.88	Unknown
3.902	347	-0.048	5.75e-03	0.020200	80.11	D-glucose, unknown
7.262	222	-0.071	6.15e-03	0.021500	85.39	Unknown
1.292	584	-0.156	6.29e-03	0.021900	80.11	Lipid methylene
0.812	632	-0.110	6.49e-03	0.022400	77.08	Cholesterol, lipid methyl
4.232	314	-0.154	6.54e-03	0.022500	80.11	Unknown
2.912	422	-0.061	6.85e-03	0.023400	82.88	Unknown
1.202	593	-0.200	6.91e-03	0.023400	70.33	Lipid methylene
1.852	528	-0.054	6.91e-03	0.023400	80.11	Unknown
6.772	271	0.273	7.56e-03	0.025500	77.08	Unknown
4.362	301	-0.088	7.90e-03	0.026500	77.08	Unknown
2.042	509	-0.118	8.11e-03	0.027000	77.08	Lipid allylic
2.012	512	-0.087	8.38e-03	0.027800	77.08	Lipid allylic
2.202	493	-0.107	8.54e-03	0.028200	77.08	Lipid (methylene carbonyl)
2.632	450	-0.043	8.65e-03	0.028400	77.08	Unknown

2.032	510	-0.109	9.61e-03	0.031400	77.08	Lipid allylic
7.252	223	-0.066	9.68e-03	0.031500	80.11	Unknown
4.282	309	-0.076	1.02e-02	0.033100	73.82	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.342	479	-0.067	1.05e-02	0.033700	73.82	Proline, glutamic acid
0.912	622	-0.160	1.09e-02	0.034800	73.82	Cholesterol, lipid methyl
0.602	653	0.371	1.15e-02	0.036700	70.33	Unkown
1.722	541	-0.044	1.17e-02	0.037200	73.82	Leucine, lysine
1.872	526	-0.048	1.18e-02	0.037200	77.08	Overlap of multiple minor compounds
4.222	315	-0.140	1.19e-02	0.037300	70.33	Unknown
6.782	270	0.158	1.19e-02	0.037300	73.82	Unknown
2.352	478	-0.054	1.20e-02	0.037400	77.08	Proline, glutamic acid
1.032	610	-0.091	1.22e-02	0.037900	73.82	L-isoleucine, lipid methyl, cholesterol (ester)
2.072	506	-0.077	1.24e-02	0.038200	73.82	Lipid allylic
4.262	311	-0.087	1.26e-02	0.038600	73.82	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.882	425	-0.077	1.29e-02	0.039400	70.33	Lipid diallylic
0.802	633	-0.101	1.30e-02	0.039500	70.33	Cholesterol, lipid methyl
2.402	473	-0.043	1.31e-02	0.039500	77.08	Glutamine, carnitine
2.022	511	-0.092	1.38e-02	0.041500	73.82	Lipid allylic
3.832	354	-0.056	1.41e-02	0.042400	73.82	Unknown
8.002	148	0.231	1.54e-02	0.045900	70.33	Unknown
4.372	300	-0.097	1.57e-02	0.046800	66.64	Unknown
8.822	66	-0.366	1.63e-02	0.048200	66.64	Unknown

Table 7.19: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from hereditary diseases from those suffering from hypertensive nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
1.212	592	-0.187	1.91e-05	0.00817	97.92	Lipid methylene
1.222	591	-0.159	2.47e-05	0.00817	97.92	Lipid methylene
2.312	482	-0.116	5.72e-05	0.01240	98.40	Lipid (methylene carbonyl)
1.232	590	-0.117	9.31e-05	0.01240	98.40	Lipid methylene

4.112	326	-0.213	1.01e-04	0.01240	98.40	Proline, lactic acid
2.362	477	-0.098	1.26e-04	0.01240	97.31	Proline, glutamic acid
2.282	485	-0.181	1.56e-04	0.01240	97.31	Lipid (methylene carbonyl)
2.292	484	-0.144	1.65e-04	0.01240	96.57	Lipid (methylene carbonyl)
2.302	483	-0.121	1.69e-04	0.01240	97.31	Lipid (methylene carbonyl)
1.132	600	-0.099	1.91e-04	0.01240	96.57	Unknown
4.302	307	-0.131	2.44e-04	0.01240	96.57	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.322	481	-0.098	2.48e-04	0.01240	96.57	Lipid (methylene carbonyl)
1.162	597	-0.150	2.49e-04	0.01240	91.75	Lipid methylene
2.382	475	-0.100	2.75e-04	0.01240	95.66	Proline, glutamic acid
1.112	602	-0.106	2.97e-04	0.01240	95.66	Unknown
1.172	596	-0.139	3.01e-04	0.01240	91.75	Lipid methylene
2.242	489	-0.218	3.29e-04	0.01250	95.66	Lipid (methylene carbonyl), acetone
4.292	308	-0.125	3.41e-04	0.01250	95.66	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.232	490	-0.201	3.91e-04	0.01300	95.66	Lipid (methylene carbonyl)
2.272	486	-0.218	3.93e-04	0.01300	95.66	Lipid (methylene carbonyl)
2.372	476	-0.096	4.81e-04	0.01370	94.57	Proline, glutamic acid
3.462	378	-0.106	5.06e-04	0.01370	96.57	D-glucose
1.012	612	-0.109	5.10e-04	0.01370	94.57	Valine, lipid methyl, cholesterol (ester)
2.332	480	-0.094	5.86e-04	0.01370	94.57	Proline, glutamic acid
1.002	613	-0.095	5.88e-04	0.01370	93.28	Valine, lipid methyl, cholesterol (ester)
2.342	479	-0.086	5.94e-04	0.01370	94.57	Proline, glutamic acid
4.282	309	-0.097	5.95e-04	0.01370	94.57	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.402	573	-0.268	6.02e-04	0.01370	93.28	Lipid methylene
0.832	630	-0.093	6.03e-04	0.01370	93.28	Cholesterol, lipid methyl
1.412	572	-0.215	6.40e-04	0.01370	93.28	Lipid methylene
3.402	384	-0.108	7.06e-04	0.01370	94.57	Unknown
2.262	487	-0.235	7.30e-04	0.01370	93.28	Lipid (methylene carbonyl)
1.392	574	-0.280	7.58e-04	0.01370	93.28	Lipid methylene
2.252	488	-0.184	7.74e-04	0.01370	94.57	Lipid (methylene carbonyl), acetone
2.222	491	-0.183	7.81e-04	0.01370	91.75	Lipid (methylene carbonyl)
1.382	575	-0.281	7.90e-04	0.01370	93.28	Lipid methylene
1.432	570	-0.137	7.99e-04	0.01370	91.75	Lipid methylene
1.152	598	-0.109	8.08e-04	0.01370	89.98	Lipid methylene
2.852	428	-0.164	8.11e-04	0.01370	93.28	Lipid diallylic
1.422	571	-0.170	8.28e-04	0.01370	91.75	Lipid methylene
4.312	306	-0.108	8.71e-04	0.01400	93.28	Lipid alpha-methylene to carboxyl, lipid glycerine

1.442	569	-0.125	9.13e-04	0.01400	91.75	Lipid methylene
2.912	422	-0.071	9.14e-04	0.01400	94.57	Unknown
2.862	427	-0.154	9.38e-04	0.01410	91.75	Lipid diallylic
1.142	599	-0.094	1.04e-03	0.01510	91.75	Unknown
0.842	629	-0.080	1.05e-03	0.01510	91.75	Cholesterol, lipid methyl
0.982	615	-0.100	1.22e-03	0.01680	91.75	Leucine, lipid methyl, cholesterol (ester)
1.582	555	-0.221	1.25e-03	0.01680	91.75	Lipids (?)
0.992	614	-0.103	1.28e-03	0.01680	89.98	Leucine, lipid methyl, cholesterol (ester)
3.372	387	-0.093	1.34e-03	0.01680	89.98	Methanol, proline
1.452	568	-0.116	1.35e-03	0.01680	89.98	Lipid methylene
1.612	552	-0.231	1.46e-03	0.01680	89.98	Lipids (?)
1.472	566	-0.095	1.47e-03	0.01680	89.98	Lipid methylene
1.592	554	-0.234	1.48e-03	0.01680	89.98	Lipids (?)
1.462	567	-0.102	1.51e-03	0.01680	89.98	Lipid methylene
0.972	616	-0.098	1.52e-03	0.01680	89.98	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.032	610	-0.110	1.52e-03	0.01680	89.98	L-isoleucine, lipid methyl, cholesterol (ester)
1.702	543	-0.060	1.54e-03	0.01680	89.98	Unknown, arginine
1.602	553	-0.237	1.54e-03	0.01680	89.98	Lipids (?)
1.572	556	-0.197	1.55e-03	0.01680	89.98	Lipids (?)
1.052	608	-0.139	1.55e-03	0.01680	89.98	Valine
1.072	606	-0.086	1.59e-03	0.01700	87.95	Valine
1.482	565	-0.087	1.62e-03	0.01700	89.98	Alanine
2.082	505	-0.112	1.67e-03	0.01710	89.98	Lipid allylic
1.242	589	-0.105	1.68e-03	0.01710	89.98	Lipid methylene
3.872	350	-0.092	1.73e-03	0.01730	91.75	D-glucose, unknown
1.372	576	-0.262	1.76e-03	0.01730	89.98	Lipid methylene
1.622	551	-0.210	1.79e-03	0.01740	87.95	Lipids (?)
2.352	478	-0.064	1.88e-03	0.01770	91.75	Proline, glutamic acid
4.092	328	-0.123	1.88e-03	0.01770	91.75	Unknown
2.122	501	-0.079	1.91e-03	0.01770	89.98	Lipid allylic
1.682	545	-0.077	1.93e-03	0.01770	89.98	Unknown, arginine
2.212	492	-0.153	1.96e-03	0.01770	87.95	Lipid (methylene carbonyl)
1.692	544	-0.066	2.06e-03	0.01840	89.98	Unknown, arginine
4.072	330	-0.117	2.10e-03	0.01850	93.28	Creatinine
3.452	379	-0.104	2.17e-03	0.01880	91.75	D-glucose, carnitine, proline
1.182	595	-0.100	2.20e-03	0.01880	87.95	Lipid methylene
4.272	310	-0.080	2.36e-03	0.02000	89.98	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
1.562	557	-0.172	2.48e-03	0.02060	87.95	Lipids (?)
3.992	338	-0.059	2.50e-03	0.02060	89.98	Unknown
4.102	327	-0.166	2.53e-03	0.02060	89.98	Unknown

1.632	550	-0.178	2.63e-03	0.02110	85.65	Lipids (?)
1.962	517	-0.064	2.89e-03	0.02300	85.65	Lipid allylic
4.062	331	-0.124	3.09e-03	0.02430	85.65	Creatinine
2.112	502	-0.078	3.26e-03	0.02530	85.65	Lipid allylic
0.942	619	-0.147	3.39e-03	0.02600	83.06	Cholesterol, lipid methyl
1.662	547	-0.102	3.47e-03	0.02630	85.65	Lipids (?)
8.142	134	-0.257	3.52e-03	0.02630	80.19	Unknown
3.472	377	-0.110	3.61e-03	0.02630	89.98	D-glucose
1.972	516	-0.071	3.62e-03	0.02630	85.65	Lipid allylic
1.642	549	-0.148	3.63e-03	0.02630	83.06	Lipids (?)
1.122	601	-0.098	3.70e-03	0.02650	83.06	Unknown
1.552	558	-0.144	3.79e-03	0.02690	85.65	Lipids (?)
0.932	620	-0.175	3.90e-03	0.02740	83.06	Cholesterol, lipid methyl
1.352	578	-0.143	4.05e-03	0.02810	83.06	Lipid methylene, lactic acid, threonine
3.422	382	-0.113	4.18e-03	0.02880	87.95	D-glucose, carnitine, taurine, proline
1.332	580	-0.219	4.39e-03	0.02990	83.06	Lipid methylene
0.822	631	-0.096	4.47e-03	0.03010	83.06	Cholesterol, lipid methyl
2.102	503	-0.065	4.53e-03	0.03020	83.06	Lipid allylic
1.712	542	-0.048	4.60e-03	0.03020	83.06	Leucine, lysine
1.672	546	-0.084	4.62e-03	0.03020	83.06	Unknown, arginine
0.962	617	-0.096	4.70e-03	0.03040	83.06	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.322	581	-0.207	4.87e-03	0.03120	83.06	Lipid methylene
3.732	364	-0.097	4.93e-03	0.03130	87.95	D-glucose, unknown
2.892	424	-0.057	5.15e-03	0.03240	83.06	Lipid diallylic
2.872	426	-0.103	5.26e-03	0.03270	80.19	Lipid diallylic
3.852	352	-0.097	5.30e-03	0.03270	85.65	D-glucose, unknown
1.982	515	-0.076	5.94e-03	0.03630	80.19	Lipid allylic
1.652	548	-0.117	6.14e-03	0.03720	80.19	Lipids (?)
1.312	582	-0.191	6.24e-03	0.03750	80.19	Lipid methylene
0.952	618	-0.109	6.40e-03	0.03810	80.19	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
8.792	69	-0.407	6.79e-03	0.04000	77.05	Unknown
1.082	605	-0.078	6.97e-03	0.04070	80.19	Unknown
3.932	344	-0.091	7.23e-03	0.04190	83.06	D-glucose
2.922	421	-0.072	7.38e-03	0.04210	80.19	Unknown
3.962	341	-0.043	7.39e-03	0.04210	77.05	Unknown
2.832	430	-0.135	7.48e-03	0.04210	77.05	Lipid diallylic
2.822	431	-0.137	7.52e-03	0.04210	77.05	Lipid diallylic
2.902	423	-0.056	7.69e-03	0.04270	80.19	Unknown
2.882	425	-0.079	7.91e-03	0.04320	77.05	Lipid diallylic
2.202	493	-0.104	7.98e-03	0.04320	77.05	Lipid (methylene carbonyl)
3.912	346	-0.088	7.99e-03	0.04320	83.06	D-glucose, betaine, unknown

1.342	579	-0.130	8.06e-03	0.04320	77.05	Lipid methylene, lactic acid, threonine
2.842	429	-0.138	8.31e-03	0.04420	77.05	Lipid diallylic
3.722	365	-0.081	8.67e-03	0.04580	83.06	D-glucose, N,N-dimethylglycine
2.072	506	-0.077	9.30e-03	0.04830	77.05	Lipid allylic
3.552	369	-0.093	9.31e-03	0.04830	80.19	D-glucose, myo-inositol
3.782	359	-0.068	9.41e-03	0.04830	80.19	D-glucose, alanine, glutamine, arginine
1.302	583	-0.171	9.44e-03	0.04830	77.05	Lipid methylene
1.022	611	-0.083	9.54e-03	0.04850	73.65	L-isoleucine, lipid methyl, cholesterol (ester)
8.862	62	0.485	9.64e-03	0.04850	73.65	Trigonelline
2.132	500	-0.049	9.70e-03	0.04850	80.19	Glutamine
2.052	508	-0.086	1.00e-02	0.04960	77.05	Lipid allylic
3.352	389	-0.064	1.01e-02	0.04960	73.65	Proline
3.532	371	-0.090	1.01e-02	0.04960	80.19	D-glucose

Table 7.20: **Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from interstitial nephropathy from those suffering from hypertensive nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

Spectral position [ppm]	**ID**	**log(FC)**	***P*-value unadjusted**	***P*-value B/H-adjusted**	**Statistical power**	**Identified compounds**
1.002	613	-0.152	4.27e-11	1.73e-08	100.00	Valine, lipid methyl, cholesterol (ester)
1.012	612	-0.171	5.23e-11	1.73e-08	100.00	Valine, lipid methyl, cholesterol (ester)
1.072	606	-0.147	1.33e-10	2.92e-08	100.00	Valine
1.902	523	-0.088	4.38e-10	7.23e-08	100.00	Overlap of multiple minor compounds
2.372	476	-0.139	1.41e-09	1.86e-07	100.00	Proline, glutamic acid
1.082	605	-0.144	1.77e-09	1.86e-07	100.00	Unknown
1.232	590	-0.150	2.28e-09	1.86e-07	100.00	Lipid methylene
1.242	589	-0.166	2.39e-09	1.86e-07	100.00	Lipid methylene
2.362	477	-0.127	2.59e-09	1.86e-07	100.00	Proline, glutamic acid
2.382	475	-0.135	2.82e-09	1.86e-07	99.99	Proline, glutamic acid
1.052	608	-0.213	6.38e-09	3.83e-07	99.99	Valine

1.162	597	-0.198	7.30e-09	4.02e-07	99.96	Lipid methylene
1.172	596	-0.181	1.54e-08	7.82e-07	99.96	Lipid methylene
1.912	522	-0.090	2.63e-08	1.24e-06	99.99	Overlap of multiple minor compounds
0.982	615	-0.143	3.15e-08	1.39e-06	99.99	Leucine, lipid methyl, cholesterol (ester)
1.932	520	-0.091	4.14e-08	1.71e-06	99.99	Acetic acid
3.452	379	-0.154	4.96e-08	1.85e-06	100.00	D-glucose, carnitine, proline
0.992	614	-0.144	5.37e-08	1.85e-06	99.99	Leucine, lipid methyl, cholesterol (ester)
1.132	600	-0.120	5.40e-08	1.85e-06	99.99	Unknown
1.482	565	-0.125	5.61e-08	1.85e-06	99.99	Alanine
3.542	370	-0.156	7.54e-08	2.37e-06	99.99	D-glucose, myo-inositol
0.842	629	-0.109	8.28e-08	2.48e-06	99.96	Cholesterol, lipid methyl
3.422	382	-0.176	8.65e-08	2.48e-06	99.99	D-glucose, carnitine, taurine, proline
3.472	377	-0.167	1.06e-07	2.92e-06	99.99	D-glucose
1.102	603	-0.131	1.42e-07	3.75e-06	99.96	Unknown
4.072	330	-0.167	1.54e-07	3.92e-06	99.99	Creatinine
2.312	482	-0.126	1.70e-07	3.98e-06	99.96	Lipid (methylene carbonyl)
1.922	521	-0.089	1.75e-07	3.98e-06	99.96	Overlap of multiple minor compounds
3.782	359	-0.115	1.75e-07	3.98e-06	99.99	D-glucose, alanine, glutamine, arginine
2.292	484	-0.166	1.95e-07	4.08e-06	99.96	Lipid (methylene carbonyl)
3.752	362	-0.155	1.95e-07	4.08e-06	99.99	D-glucose, glutamic acid
1.032	610	-0.151	1.98e-07	4.08e-06	99.96	L-isoleucine, lipid methyl, cholesterol (ester)
1.892	524	-0.076	2.04e-07	4.08e-06	99.96	Overlap of multiple minor compounds
4.092	328	-0.172	2.13e-07	4.14e-06	99.98	Unknown
0.972	616	-0.133	2.22e-07	4.15e-06	99.98	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
3.852	352	-0.150	2.26e-07	4.15e-06	99.99	D-glucose, unknown
2.872	426	-0.159	2.51e-07	4.47e-06	99.96	Lipid diallylic
1.092	604	-0.129	2.79e-07	4.78e-06	99.94	Unknown
3.552	369	-0.154	2.82e-07	4.78e-06	99.99	D-glucose, myo-inositol
1.062	607	-0.132	2.95e-07	4.87e-06	99.94	Valine
3.562	368	-0.148	3.31e-07	5.32e-06	99.99	D-glucose
2.282	485	-0.201	4.65e-07	7.31e-06	99.94	Lipid (methylene carbonyl)
1.022	611	-0.134	4.95e-07	7.59e-06	99.91	L-isoleucine, lipid methyl, cholesterol (ester)
1.502	563	-0.123	5.38e-07	8.07e-06	99.91	Alanine
1.882	525	-0.073	5.76e-07	8.45e-06	99.94	Overlap of multiple minor compounds

3.802	357	-0.120	6.81e-07	9.60e-06	99.98	D-glucose, alanine
2.302	483	-0.133	6.84e-07	9.60e-06	99.94	Lipid (methylene carbonyl)
1.472	566	-0.123	7.92e-07	1.09e-05	99.91	Lipid methylene
3.842	353	-0.135	1.02e-06	1.38e-05	99.96	D-glucose, unknown
3.792	358	-0.082	1.19e-06	1.57e-05	99.96	D-glucose, alanine
1.142	599	-0.115	1.25e-06	1.62e-05	99.85	Unknown
3.872	350	-0.118	1.38e-06	1.73e-05	99.96	D-glucose, unknown
3.732	364	-0.139	1.39e-06	1.73e-05	99.96	D-glucose, unknown
3.522	372	-0.174	1.45e-06	1.75e-05	99.91	D-glucose
3.772	360	-0.126	1.46e-06	1.75e-05	99.94	D-glucose, alanine, glutamine, arginine
1.182	595	-0.129	2.02e-06	2.39e-05	99.85	Lipid methylene
1.732	540	-0.072	2.06e-06	2.39e-05	99.78	Leucine, lysine
1.742	539	-0.075	2.22e-06	2.52e-05	99.78	Leucine, lysine
1.222	591	-0.148	2.34e-06	2.62e-05	99.52	Lipid methylene
3.442	380	-0.148	2.40e-06	2.64e-05	99.94	D-glucose, carnitine, taurine, proline
3.512	373	-0.154	2.75e-06	2.98e-05	99.94	D-glucose
3.932	344	-0.132	2.82e-06	3.01e-05	99.94	D-glucose
3.572	367	-0.129	2.93e-06	3.07e-05	99.91	D-glucose, glycine
1.252	588	-0.144	3.05e-06	3.15e-05	99.78	Lipid methylene
3.482	376	-0.149	3.10e-06	3.15e-05	99.94	D-glucose
3.912	346	-0.127	3.69e-06	3.69e-05	99.91	D-glucose, betaine, unknown
1.702	543	-0.072	4.00e-06	3.94e-05	99.78	Unknown, arginine
3.742	363	-0.138	4.20e-06	4.08e-05	99.91	D-glucose, leucine
2.902	423	-0.080	4.31e-06	4.13e-05	99.78	Unknown
1.752	538	-0.078	4.49e-06	4.24e-05	99.67	Leucine, lysine
3.722	365	-0.117	4.67e-06	4.32e-05	99.91	D-glucose, N,N-dimethylglycine
1.682	545	-0.094	4.72e-06	4.32e-05	99.78	Unknown, arginine
1.722	541	-0.065	4.90e-06	4.43e-05	99.67	Leucine, lysine
3.862	351	-0.137	5.32e-06	4.75e-05	99.91	D-glucose, unknown
3.432	381	-0.145	5.54e-06	4.87e-05	99.85	D-glucose, carnitine, taurine, proline
1.462	567	-0.121	5.60e-06	4.87e-05	99.67	Lipid methylene
2.552	458	-0.080	5.92e-06	5.08e-05	96.47	Citric acid
3.762	361	-0.126	6.56e-06	5.46e-05	99.85	D-glucose, arginine, glutamine, glutamic acid
1.122	601	-0.126	6.59e-06	5.46e-05	99.52	Unknown
1.152	598	-0.122	6.62e-06	5.46e-05	99.30	Lipid methylene
0.962	617	-0.127	7.10e-06	5.79e-05	99.67	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
1.412	572	-0.234	7.76e-06	6.25e-05	99.52	Lipid methylene
3.502	374	-0.146	7.89e-06	6.25e-05	99.85	D-glucose
2.862	427	-0.172	8.00e-06	6.25e-05	99.52	Lipid diallylic
1.452	568	-0.134	8.05e-06	6.25e-05	99.52	Lipid methylene

1.402	573	-0.289	8.41e-06	6.46e-05	99.52	Lipid methylene
2.122	501	-0.094	8.52e-06	6.47e-05	99.52	Lipid allylic
1.422	571	-0.185	1.22e-05	9.18e-05	99.30	Lipid methylene
1.212	592	-0.159	1.32e-05	9.71e-05	98.06	Lipid methylene
3.412	383	-0.138	1.32e-05	9.71e-05	99.78	D-glucose, carnitine, taurine, proline
1.492	564	-0.103	1.35e-05	9.81e-05	99.30	Alanine
1.692	544	-0.077	1.61e-05	1.15e-04	99.52	Unknown, arginine
3.462	378	-0.109	1.68e-05	1.19e-04	99.67	D-glucose
1.442	569	-0.134	1.78e-05	1.24e-04	99.30	Lipid methylene
2.322	481	-0.096	1.78e-05	1.24e-04	99.30	Lipid (methylene carbonyl)
3.062	407	-0.090	1.80e-05	1.24e-04	99.67	Creatinine
2.212	492	-0.175	2.14e-05	1.44e-04	99.00	Lipid (methylene carbonyl)
2.252	488	-0.193	2.14e-05	1.44e-04	99.00	Lipid (methylene carbonyl), acetone
4.102	327	-0.194	2.22e-05	1.45e-04	99.00	Unknown
1.982	515	-0.098	2.22e-05	1.45e-04	99.30	Lipid allylic
1.432	570	-0.145	2.24e-05	1.45e-04	99.00	Lipid methylene
3.882	349	-0.116	2.25e-05	1.45e-04	99.67	D-glucose, unknown
1.112	602	-0.103	2.48e-05	1.59e-04	99.30	Unknown
2.852	428	-0.172	2.64e-05	1.66e-04	99.00	Lipid diallylic
0.942	619	-0.176	2.65e-05	1.66e-04	99.00	Cholesterol, lipid methyl
0.852	628	-0.084	2.81e-05	1.75e-04	99.00	Cholesterol, lipid methyl
2.352	478	-0.072	3.00e-05	1.84e-04	99.30	Proline, glutamic acid
3.532	371	-0.122	3.01e-05	1.84e-04	99.52	D-glucose
1.392	574	-0.287	3.24e-05	1.96e-04	99.00	Lipid methylene
3.492	375	-0.135	3.31e-05	1.99e-04	99.52	D-glucose
2.272	486	-0.212	3.44e-05	2.05e-04	99.00	Lipid (methylene carbonyl)
0.832	630	-0.092	4.34e-05	2.56e-04	98.60	Cholesterol, lipid methyl
0.952	618	-0.135	4.69e-05	2.74e-04	98.60	Leucine, L-isoleucine, lipid methyl, cholesterol (ester)
2.202	493	-0.132	5.21e-05	3.02e-04	98.60	Lipid (methylene carbonyl)
1.972	516	-0.082	5.27e-05	3.02e-04	98.60	Lipid allylic
1.992	514	-0.097	5.79e-05	3.29e-04	98.60	Lipid allylic
2.222	491	-0.182	6.40e-05	3.61e-04	98.06	Lipid (methylene carbonyl)
1.552	558	-0.163	8.17e-05	4.57e-04	98.06	Lipids (?)
2.132	500	-0.062	8.63e-05	4.79e-04	98.60	Glutamine
2.502	463	0.095	1.02e-04	5.61e-04	98.06	Unknown
0.932	620	-0.195	1.17e-04	6.38e-04	97.36	Cholesterol, lipid methyl
1.712	542	-0.055	1.21e-04	6.52e-04	98.06	Leucine, lysine
4.112	326	-0.174	1.30e-04	6.96e-04	97.36	Proline, lactic acid
1.542	559	-0.127	1.32e-04	7.04e-04	97.36	Lipids (?)
1.382	575	-0.265	1.39e-04	7.35e-04	97.36	Lipid methylene
3.812	356	-0.091	1.51e-04	7.91e-04	98.06	D-glucose
2.842	429	-0.163	1.72e-04	8.90e-04	97.36	Lipid diallylic

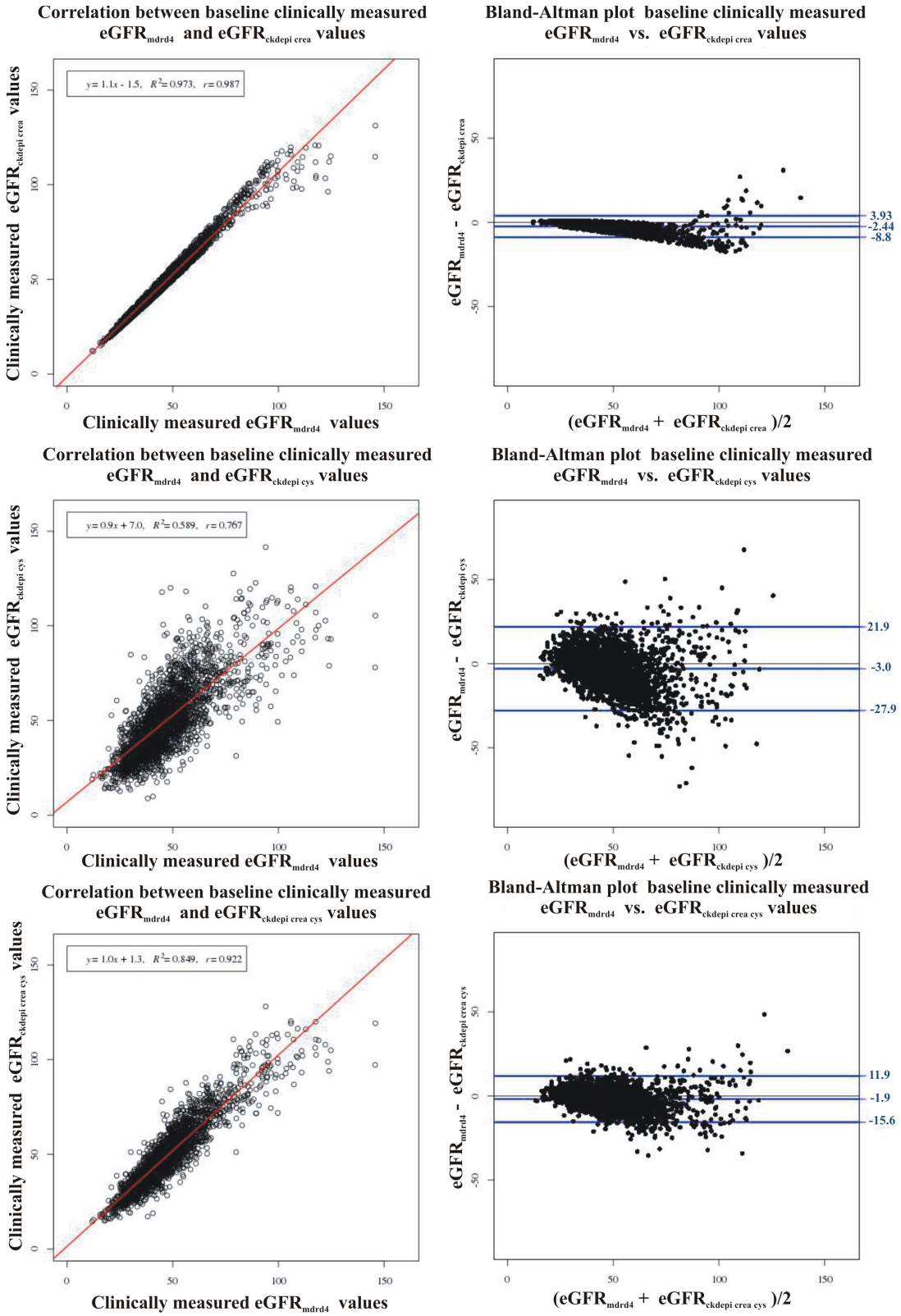
1.562	557	-0.178	1.73e-04	8.90e-04	97.36	Lipids (?)
2.672	446	-0.086	1.85e-04	9.49e-04	90.11	Citric acid
1.522	561	-0.076	1.89e-04	9.60e-04	97.36	Lipids (?)
4.312	306	-0.100	1.93e-04	9.71e-04	96.47	Lipid alpha-methylene to carboxyl, lipid glycerine
4.032	334	-0.079	2.22e-04	1.11e-03	96.47	Unknown
4.042	333	-0.122	2.25e-04	1.11e-03	96.47	Unknown
2.892	424	-0.063	2.27e-04	1.12e-03	97.36	Lipid diallylic
2.832	430	-0.154	2.54e-04	1.24e-03	96.47	Lipid diallylic
2.492	464	0.064	2.56e-04	1.24e-03	96.47	Glutamine
2.192	494	-0.093	2.73e-04	1.31e-03	96.47	Lipid (methylene carbonyl)
2.882	425	-0.090	2.74e-04	1.31e-03	96.47	Lipid diallylic
2.232	490	-0.170	3.13e-04	1.48e-03	95.33	Lipid (methylene carbonyl)
2.112	502	-0.079	3.14e-04	1.48e-03	95.33	Lipid allylic
2.652	448	-0.045	3.28e-04	1.53e-03	95.33	Unknown
0.742	639	-0.153	3.33e-04	1.55e-03	95.33	Unkown
3.832	354	-0.065	3.62e-04	1.67e-03	96.47	Unknown
2.822	431	-0.150	4.18e-04	1.91e-03	95.33	Lipid diallylic
2.912	422	-0.063	4.33e-04	1.97e-03	95.33	Unknown
2.262	487	-0.202	4.83e-04	2.18e-03	95.33	Lipid (methylene carbonyl)
2.572	456	-0.053	4.96e-04	2.22e-03	87.65	$CaEDTA^{2-}$, citric acid
1.962	517	-0.062	4.99e-04	2.22e-03	95.33	Lipid allylic
1.042	609	-0.122	5.09e-04	2.25e-03	95.33	L-isoleucine, lipid methyl, cholesterol (ester)
2.002	513	-0.085	5.53e-04	2.43e-03	93.91	Lipid allylic
7.202	228	-0.085	5.56e-04	2.43e-03	93.91	Tyrosine
1.622	551	-0.193	5.59e-04	2.43e-03	93.91	Lipids (?)
3.822	355	-0.072	5.65e-04	2.44e-03	95.33	Unknown
2.242	489	-0.174	5.82e-04	2.49e-03	93.91	Lipid (methylene carbonyl), acetone
4.052	332	-0.121	6.20e-04	2.64e-03	95.33	Unknown
2.082	505	-0.101	6.46e-04	2.73e-03	93.91	Lipid allylic
1.372	576	-0.236	7.07e-04	2.97e-03	93.91	Lipid methylene
1.942	519	-0.112	7.33e-04	3.06e-03	90.11	Acetic acid
1.572	556	-0.174	7.64e-04	3.16e-03	93.91	Lipids (?)
0.822	631	-0.095	7.69e-04	3.16e-03	92.19	Cholesterol, lipid methyl
1.952	518	-0.054	7.71e-04	3.16e-03	93.91	Acetic acid
4.322	305	-0.081	8.00e-04	3.26e-03	93.91	Lipid alpha-methylene to carboxyl, lipid glycerine
1.612	552	-0.202	8.15e-04	3.30e-03	92.19	Lipids (?)
2.662	447	-0.093	8.36e-04	3.36e-03	87.65	Citric acid
3.402	384	-0.089	8.39e-04	3.36e-03	92.19	Unknown
2.542	459	-0.061	9.04e-04	3.60e-03	87.65	Unknown
1.632	550	-0.162	9.89e-04	3.91e-03	92.19	Lipids (?)
1.262	587	-0.110	1.02e-03	4.02e-03	92.19	Lipid methylene

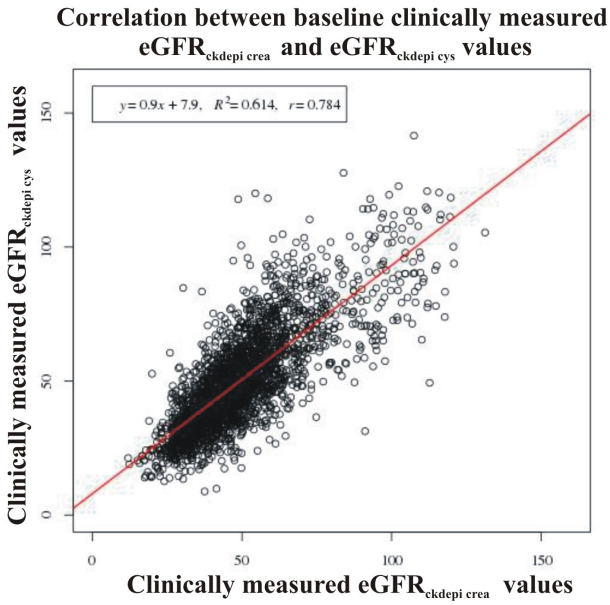
3.342	390	-0.118	1.05e-03	4.10e-03	90.11	Proline
2.712	442	-0.039	1.08e-03	4.17e-03	92.19	$MgEDTA^{2-}$
1.532	560	-0.083	1.08e-03	4.17e-03	92.19	Lipids (?)
3.362	388	-0.075	1.12e-03	4.28e-03	90.11	Proline
7.832	165	0.206	1.17e-03	4.48e-03	90.11	Unknown
1.672	546	-0.079	1.28e-03	4.84e-03	92.19	Unknown, arginine
3.052	408	-0.038	1.38e-03	5.20e-03	90.11	Creatinine
0.732	640	-0.185	1.53e-03	5.72e-03	90.11	Unkown
2.922	421	-0.071	1.54e-03	5.75e-03	90.11	Unknown
2.102	503	-0.060	1.68e-03	6.23e-03	90.11	Lipid allylic
1.602	553	-0.195	1.73e-03	6.40e-03	90.11	Lipids (?)
4.302	307	-0.093	1.76e-03	6.45e-03	90.11	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
2.332	480	-0.070	2.10e-03	7.65e-03	87.65	Proline, glutamic acid
0.752	638	-0.097	2.24e-03	8.11e-03	87.65	Cholesterol, lipid methyl
1.582	555	-0.174	2.26e-03	8.15e-03	87.65	Lipids (?)
7.562	192	0.098	2.27e-03	8.16e-03	84.80	Unknown
1.872	526	-0.046	2.35e-03	8.38e-03	87.65	Overlap of multiple minor compounds
0.882	625	0.084	2.37e-03	8.43e-03	90.11	Cholesterol, lipid methyl
1.642	549	-0.128	2.57e-03	9.05e-03	84.80	Lipids (?)
4.162	321	-0.082	2.71e-03	9.52e-03	87.65	Proline, lactic acid
7.722	176	0.138	2.73e-03	9.53e-03	84.80	Unknown
7.192	229	-0.057	2.75e-03	9.55e-03	84.80	Tyrosine
7.692	179	-0.147	2.82e-03	9.74e-03	84.80	Unknown
1.662	547	-0.086	2.90e-03	9.97e-03	84.80	Lipids (?)
0.812	632	-0.096	3.01e-03	1.03e-02	84.80	Cholesterol, lipid methyl
1.332	580	-0.189	3.08e-03	1.05e-02	84.80	Lipid methylene
2.702	443	-0.033	3.14e-03	1.06e-02	81.54	$MgEDTA^{2-}$
1.592	554	-0.180	3.30e-03	1.11e-02	84.80	Lipids (?)
1.352	578	-0.122	3.35e-03	1.12e-02	84.80	Lipid methylene, lactic acid, threonine
7.182	230	-0.069	3.77e-03	1.26e-02	81.54	Unknown
1.762	537	-0.049	3.89e-03	1.29e-02	84.80	Leucine, lysine
2.142	499	-0.042	4.03e-03	1.32e-02	84.80	Glutamine
8.262	122	0.086	4.03e-03	1.32e-02	81.54	Unknown
2.342	479	-0.060	4.10e-03	1.34e-02	84.80	Proline, glutamic acid
7.732	175	0.115	4.20e-03	1.37e-02	84.80	Unknown
1.192	594	-0.143	4.38e-03	1.42e-02	73.83	Lipid methylene
2.682	445	-0.043	5.13e-03	1.65e-02	77.88	Citric acid
7.432	205	-0.060	5.71e-03	1.83e-02	77.88	Phenylalanine
0.672	646	-0.141	6.01e-03	1.92e-02	84.80	Unkown
2.812	432	-0.116	6.38e-03	2.02e-02	81.54	Lipid diallylic
1.652	548	-0.096	6.42e-03	2.03e-02	77.88	Lipids (?)
1.272	586	-0.097	6.83e-03	2.15e-02	77.88	Lipid methylene

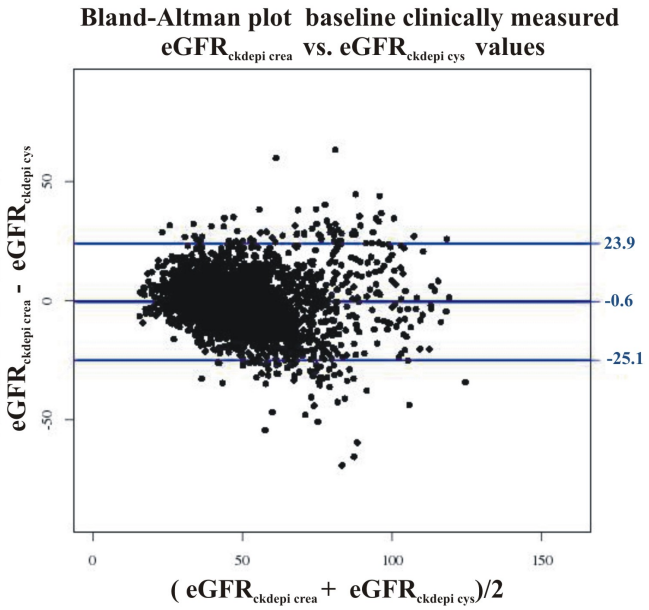
1.322	581	-0.165	7.00e-03	2.19e-02	77.88	Lipid methylene
3.042	409	-0.021	7.56e-03	2.35e-02	77.88	Lysine, unknown
0.722	641	-0.169	7.65e-03	2.37e-02	77.88	Unkown
4.332	304	-0.062	8.15e-03	2.51e-02	77.88	Lipid alpha-methylene to carboxyl, lipid glycerine
2.452	468	-0.063	8.21e-03	2.52e-02	77.88	Glutamine, carnitine
0.922	621	-0.142	8.37e-03	2.56e-02	77.88	Cholesterol, lipid methyl
1.512	562	-0.066	8.68e-03	2.64e-02	77.88	Alanine
1.342	579	-0.107	9.08e-03	2.75e-02	73.83	Lipid methylene, lactic acid, threonine
2.932	420	-0.047	9.55e-03	2.88e-02	77.88	Unknown
6.842	264	-0.055	9.87e-03	2.95e-02	77.88	Unknown
3.942	343	-0.082	9.89e-03	2.95e-02	77.88	D-glucose
7.102	238	-0.089	1.07e-02	3.19e-02	73.83	Unknown
7.082	240	-0.211	1.09e-02	3.23e-02	69.44	Unknown
7.962	152	0.219	1.10e-02	3.25e-02	69.44	Unknown
7.352	213	-0.041	1.15e-02	3.36e-02	69.44	Phenylalanine
1.312	582	-0.146	1.18e-02	3.44e-02	73.83	Lipid methylene
1.792	534	0.053	1.20e-02	3.48e-02	73.83	Unknown
1.202	593	-0.149	1.21e-02	3.50e-02	64.74	Lipid methylene
2.692	444	-0.037	1.23e-02	3.54e-02	54.72	Citric acid
3.312	393	-0.073	1.26e-02	3.63e-02	73.83	Unknown
3.922	345	-0.055	1.28e-02	3.66e-02	73.83	D-glucose, unknown
7.572	191	0.089	1.29e-02	3.66e-02	69.44	Unknown
4.082	329	-0.083	1.31e-02	3.70e-02	73.83	Creatinine
7.342	214	-0.037	1.37e-02	3.85e-02	64.74	Phenylalanine
8.232	125	0.070	1.38e-02	3.87e-02	69.44	Unknown
6.592	289	-0.284	1.38e-02	3.87e-02	69.44	Unknown
2.392	474	-0.060	1.42e-02	3.96e-02	69.44	Unknown
2.792	434	-0.083	1.43e-02	3.96e-02	69.44	Lipid diallylic
0.802	633	-0.079	1.52e-02	4.19e-02	69.44	Cholesterol, lipid methyl
2.482	465	0.043	1.53e-02	4.19e-02	64.74	Glutamine, carnitine
8.222	126	0.070	1.53e-02	4.20e-02	69.44	Unknown
3.352	389	-0.050	1.54e-02	4.21e-02	69.44	Proline
8.722	76	0.294	1.56e-02	4.23e-02	64.74	Unknown
6.832	265	-0.051	1.60e-02	4.33e-02	69.44	Unknown
2.012	512	-0.063	1.62e-02	4.35e-02	69.44	Lipid allylic
7.952	153	0.165	1.67e-02	4.49e-02	64.74	Unknown
7.252	223	-0.049	1.69e-02	4.53e-02	69.44	Unknown
4.292	308	-0.069	1.72e-02	4.56e-02	69.44	Lipid alpha-methylene to carboxyl, lipid glycerine, threonine
9.392	9	-0.292	1.72e-02	4.56e-02	69.44	Unknown
2.802	433	-0.088	1.76e-02	4.66e-02	69.44	Lipid diallylic
6.922	256	-0.059	1.82e-02	4.77e-02	64.74	Tyrosine

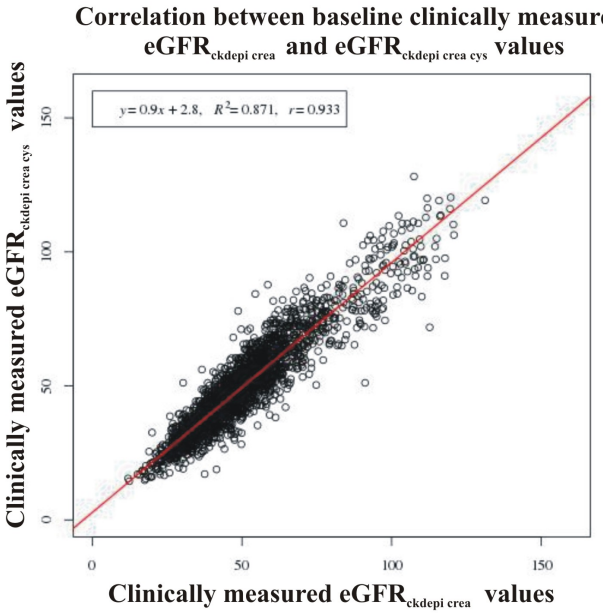
Table 7.21: **Previous page: Spectral positions given in ppm, IDs, log(Fold-change) (log(FC)), p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, statistical power in %, as well as correspondingly identified compounds of NMR features that discriminated patients suffering from systemic diseases from those suffering from hypertensive nephropathy.** A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap. The statistical power was calculated with a significance level of 0.05 and a specificity of 95%.

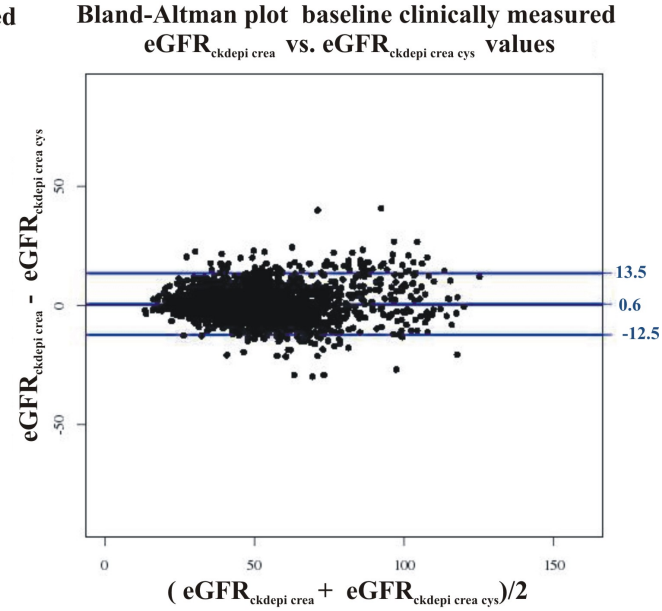
7.3.3 Prediction of present and future kidney performance

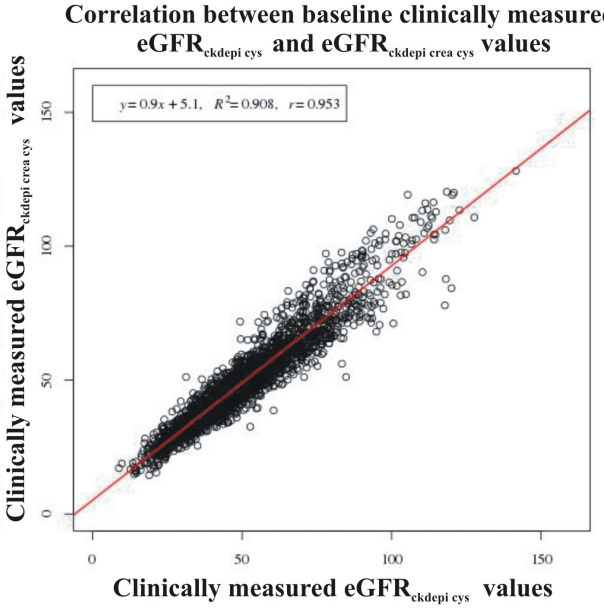


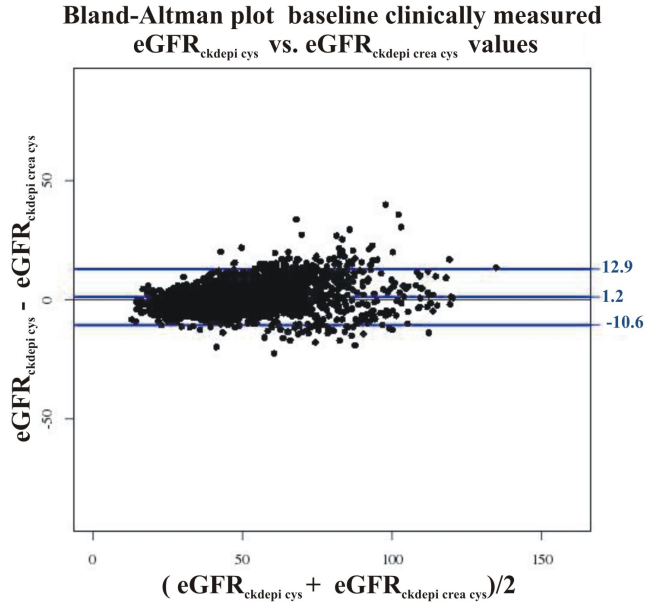

Correlation between baseline clinically measured eGFR ckdepi crea and eGFR ckdepi cys values
y = 0.9x + 7.9, R² = 0.614, r = 0.784
Clinically measured eGFR ckdepi cys values
Clinically measured eGFR ckdepi crea values
0
50
100
150

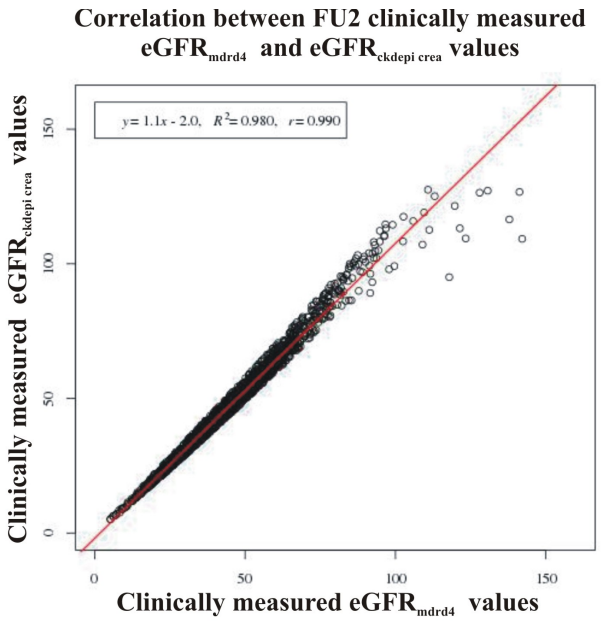

Bland-Altman plot baseline clinically measured eGFR ckdepi crea vs. eGFR ckdepi cys values
eGFR ckdepi crea - eGFR ckdepi cys
(eGFR ckdepi crea + eGFR ckdepi cys)/2
23.9
-0.6
-25.1
-50
0
50
100
150

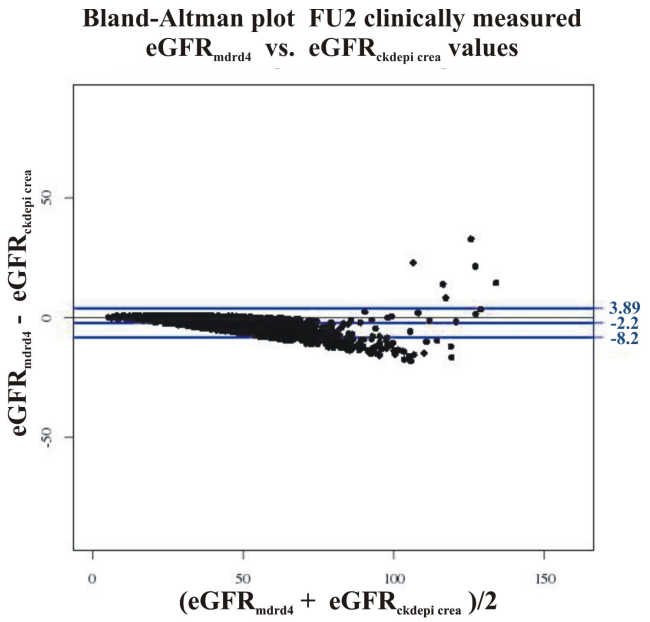

Correlation between baseline clinically measured eGFR ckdepi crea and eGFR ckdepi crea cys values
y = 0.9x + 2.8, R² = 0.871, r = 0.933
Clinically measured eGFR ckdepi crea cys values
Clinically measured eGFR ckdepi crea values
0
50
100
150

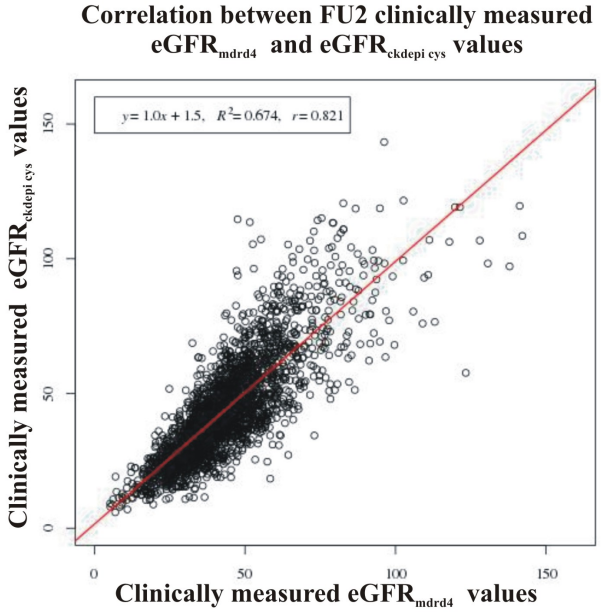

Bland-Altman plot baseline clinically measured eGFR ckdepi crea vs. eGFR ckdepi crea cys values
eGFR ckdepi crea - eGFR ckdepi crea cys
(eGFR ckdepi crea + eGFR ckdepi crea cys)/2
13.5
0.6
-12.5
-50
0
50
100
150

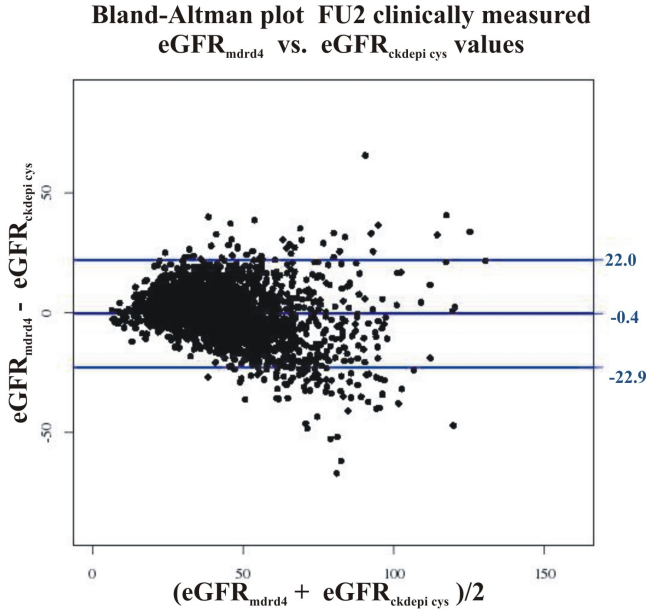

Correlation between baseline clinically measured eGFR ckdepi cys and eGFR ckdepi crea cys values
y = 0.9x + 5.1, R² = 0.908, r = 0.953
Clinically measured eGFR ckdepi crea cys values
Clinically measured eGFR ckdepi cys values
0
50
100
150

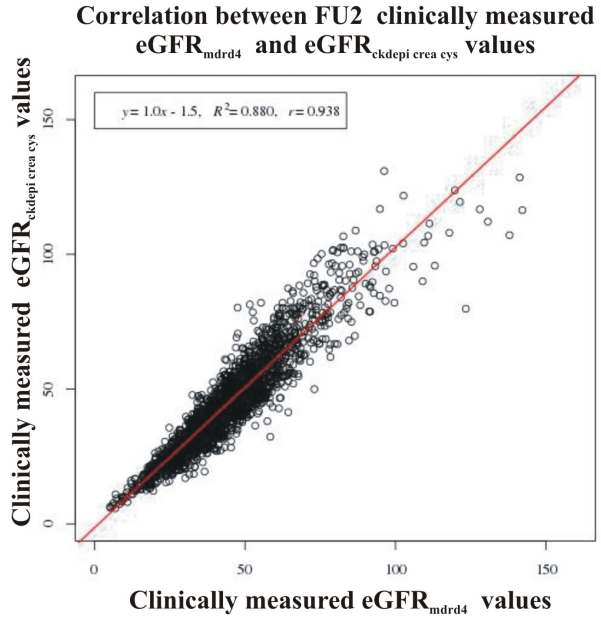

Bland-Altman plot baseline clinically measured eGFR ckdepi cys vs. eGFR ckdepi crea cys values
eGFR ckdepi cys - eGFR ckdepi crea cys
(eGFR ckdepi cys + eGFR ckdepi crea cys)/2
12.9
1.2
-10.6
-50
0
50
100
150

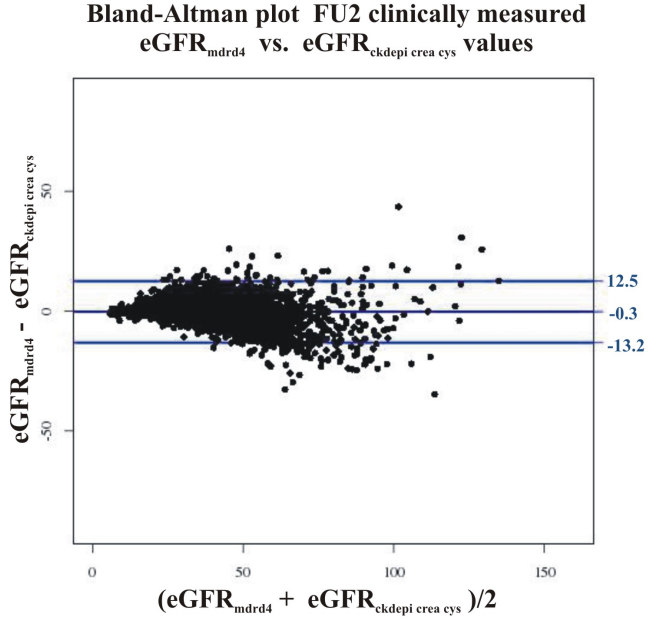

Correlation between FU2 clinically measured
eGFR mdrd4 and eGFR ckdepi crea values
y= 1.1x - 2.0, R²= 0.980, r= 0.990
Clinically measured eGFR ckdepi crea values
Clinically measured eGFR mdrd4 values
0
50
100
150

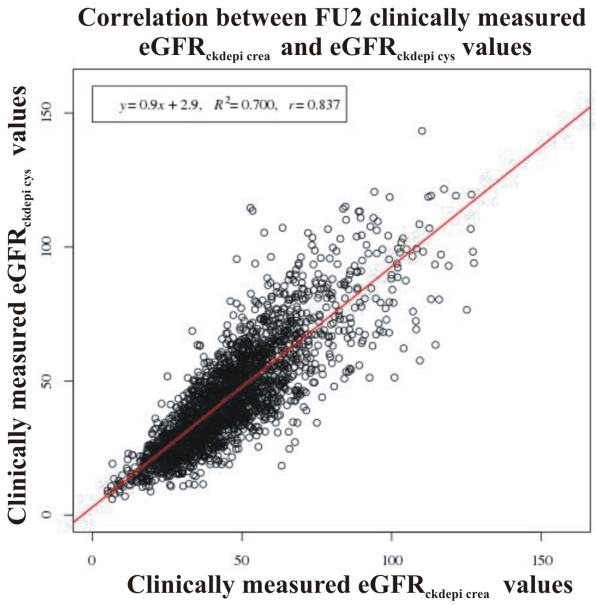

Bland-Altman plot FU2 clinically measured
eGFR mdrd4 vs. eGFR ckdepi crea values
3.89
-2.2
-8.2
eGFR mdrd4 - eGFR ckdepi crea
(eGFR mdrd4 + eGFR ckdepi crea)/2
-50
0
50
100
150

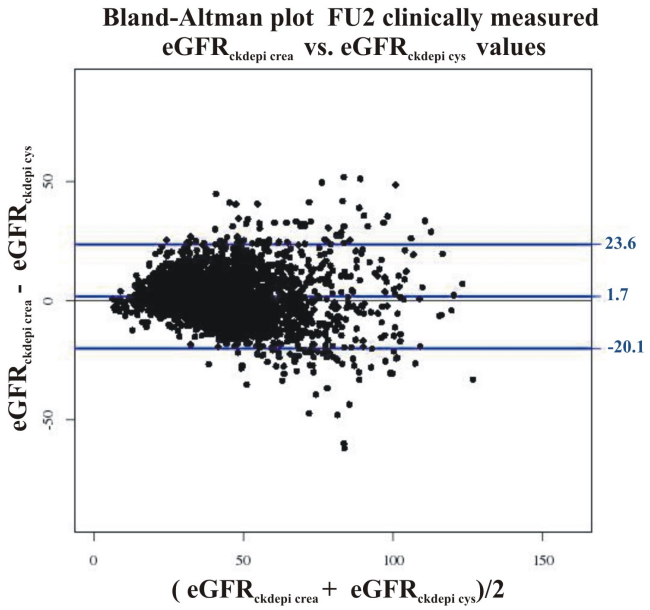

Correlation between FU2 clinically measured
eGFR mdrd4 and eGFR ckdepi cys values
y= 1.0x + 1.5, R²= 0.674, r= 0.821
Clinically measured eGFR ckdepi cys values
Clinically measured eGFR mdrd4 values
0
50
100
150

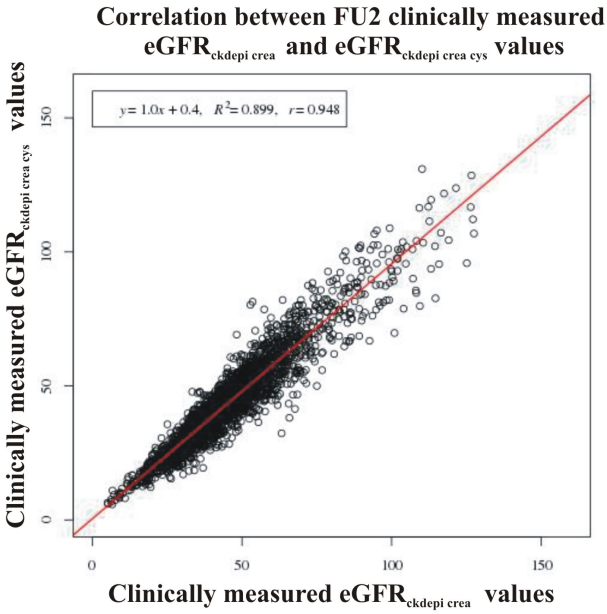

Bland-Altman plot FU2 clinically measured
eGFR mdrd4 vs. eGFR ckdepi cys values
22.0
-0.4
-22.9
eGFR mdrd4 - eGFR ckdepi cys
(eGFR mdrd4 + eGFR ckdepi cys)/2
-50
0
50
100
150

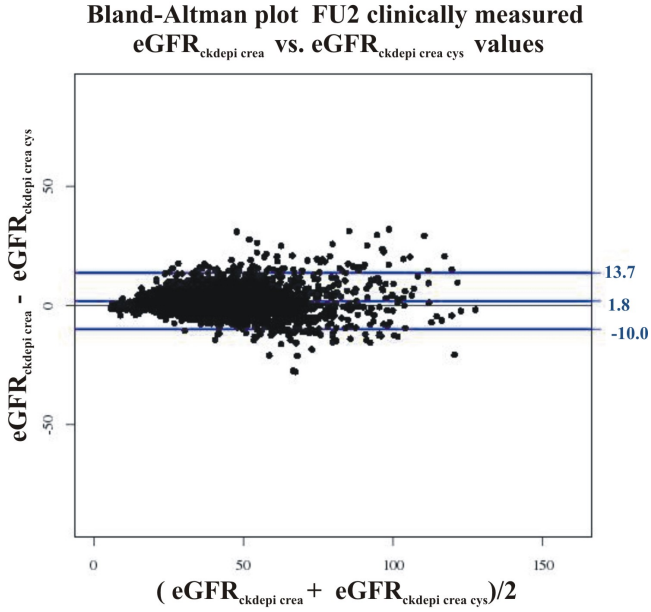

Correlation between FU2 clinically measured
eGFR mdrd4 and eGFR ckdepi crea cys values
y= 1.0x - 1.5, R²= 0.880, r= 0.938
Clinically measured eGFR ckdepi crea cys values
Clinically measured eGFR mdrd4 values
0
50
100
150

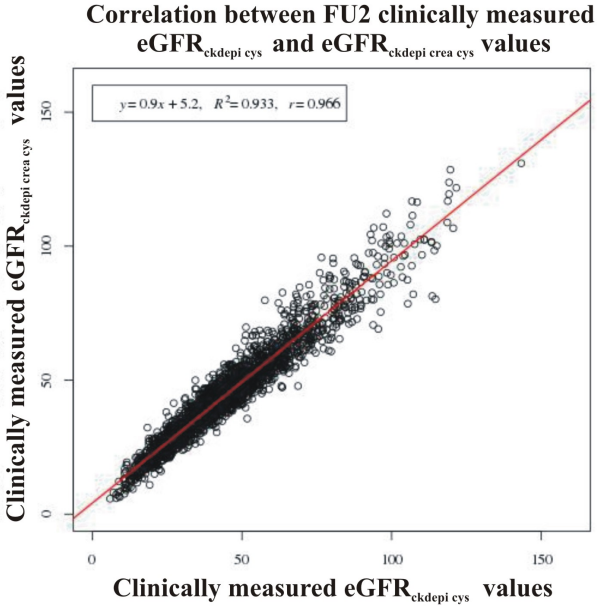

Bland-Altman plot FU2 clinically measured
eGFR mdrd4 vs. eGFR ckdepi crea cys values
12.5
-0.3
-13.2
eGFR mdrd4 - eGFR ckdepi crea cys
(eGFR mdrd4 + eGFR ckdepi crea cys)/2
-50
0
50
100
150


Correlation between FU2 clinically measured
eGFR ckdepi crea and eGFR ckdepi cys values
y = 0.9x + 2.9, R² = 0.700, r = 0.837
Clinically measured eGFR ckdepi cys values
Clinically measured eGFR ckdepi crea values
0
50
100
150


Bland-Altman plot FU2 clinically measured
eGFR ckdepi crea vs. eGFR ckdepi cys values
eGFR ckdepi crea - eGFR ckdepi cys
(eGFR ckdepi crea + eGFR ckdepi cys)/2
23.6
1.7
-20.1
-50
0
50
100
150


Correlation between FU2 clinically measured
eGFR ckdepi crea and eGFR ckdepi crea cys values
y = 1.0x + 0.4, R² = 0.899, r = 0.948
Clinically measured eGFR ckdepi crea cys values
Clinically measured eGFR ckdepi crea values
0
50
100
150


Bland-Altman plot FU2 clinically measured
eGFR ckdepi crea vs. eGFR ckdepi crea cys values
eGFR ckdepi crea - eGFR ckdepi crea cys
(eGFR ckdepi crea + eGFR ckdepi crea cys)/2
13.7
1.8
-10.0
-50
0
50
100
150


Correlation between FU2 clinically measured
eGFR ckdepi cys and eGFR ckdepi crea cys values
y = 0.9x + 5.2, R² = 0.933, r = 0.966
Clinically measured eGFR ckdepi crea cys values
Clinically measured eGFR ckdepi cys values
0
50
100
150

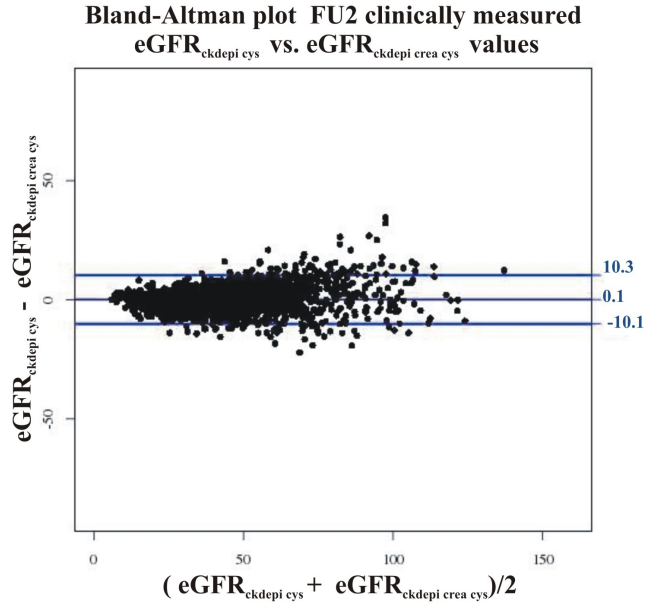

Bland-Altman plot FU2 clinically measured
eGFR ckdepi cys vs. eGFR ckdepi crea cys values
eGFR ckdepi cys - eGFR ckdepi crea cys
(eGFR ckdepi cys + eGFR ckdepi crea cys)/2
10.3
0.1
-10.1
-50
0
50
100
150

Figure 7.4: **Previous pages: Method comparison of different GFR estimation equations for baseline and FU2 eGFR values, respectively.** $x - y$ scatter plots of the compared eGFR values including equations of the fitted simple linear regression lines, Pearson's correlation coefficients r and coefficients of determination $R^2 = r^2$, as well as Bland-Altman plots are displayed, compare to section 5.2.3.3 Table 5.5. All displayed values are given in ml/min per 1.73m^2. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; FU2, second follow-up.

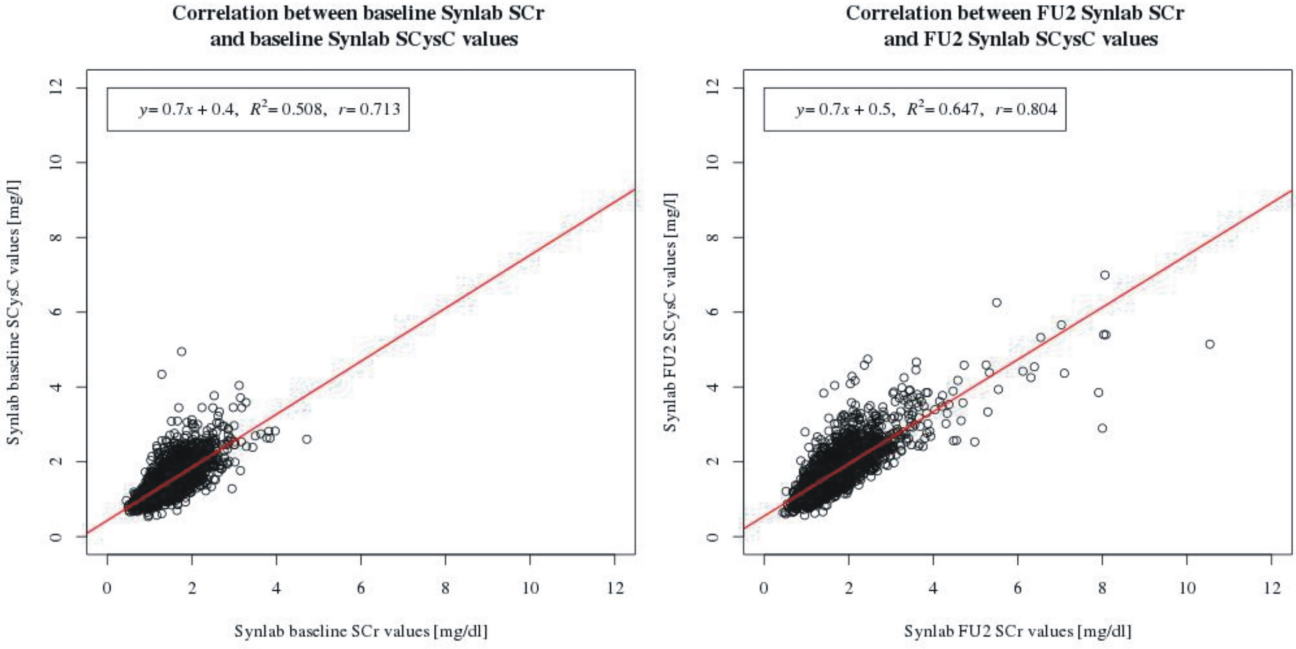
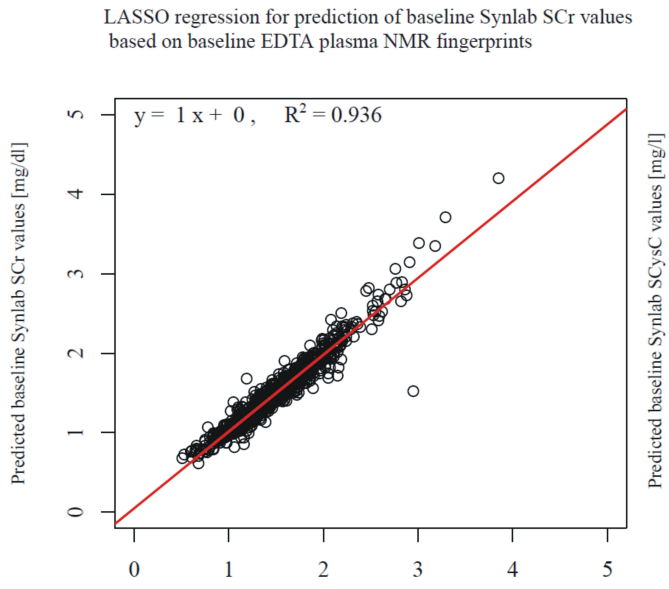
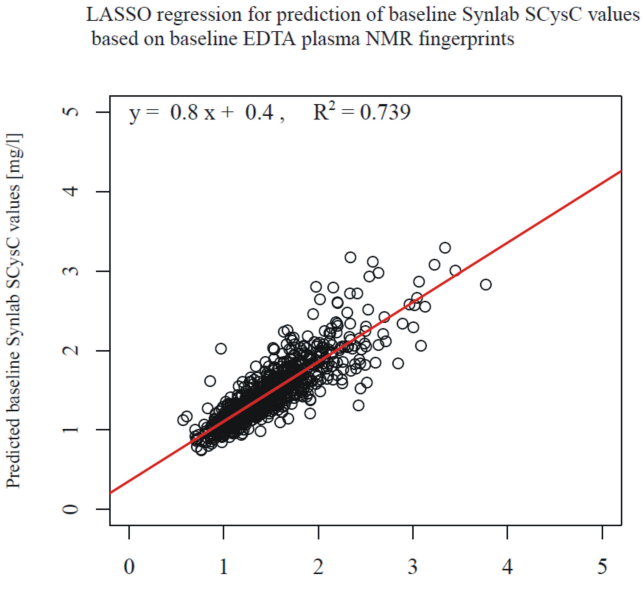
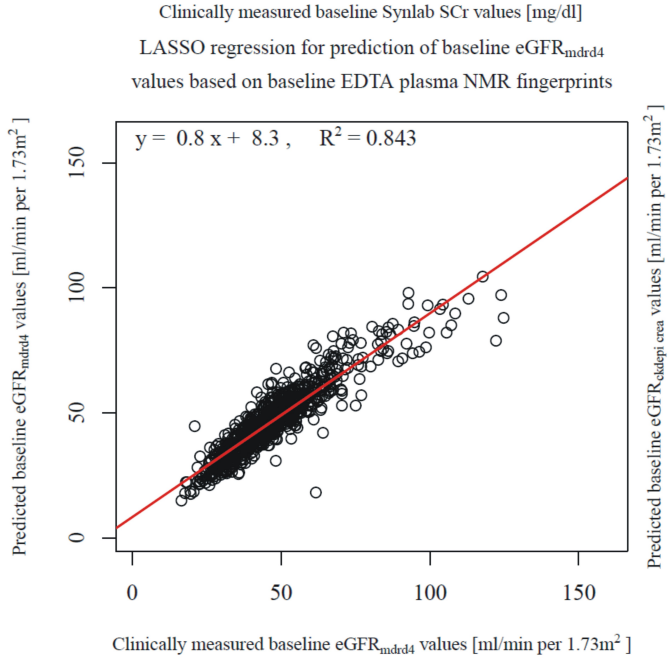
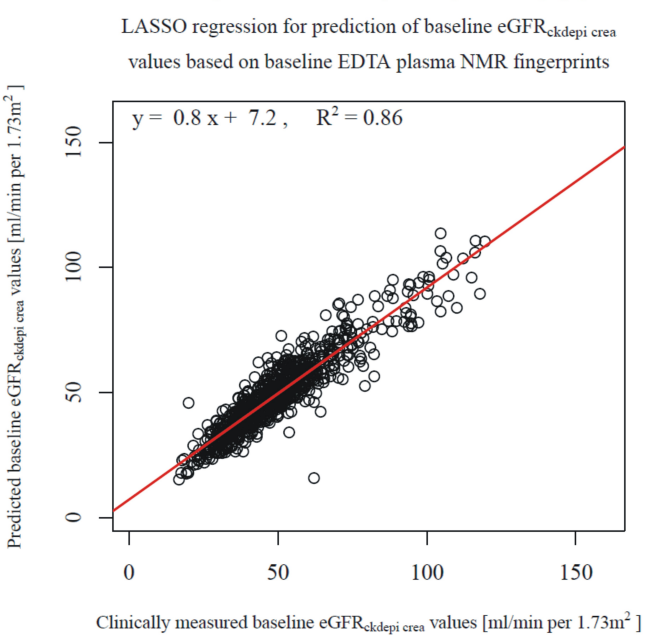


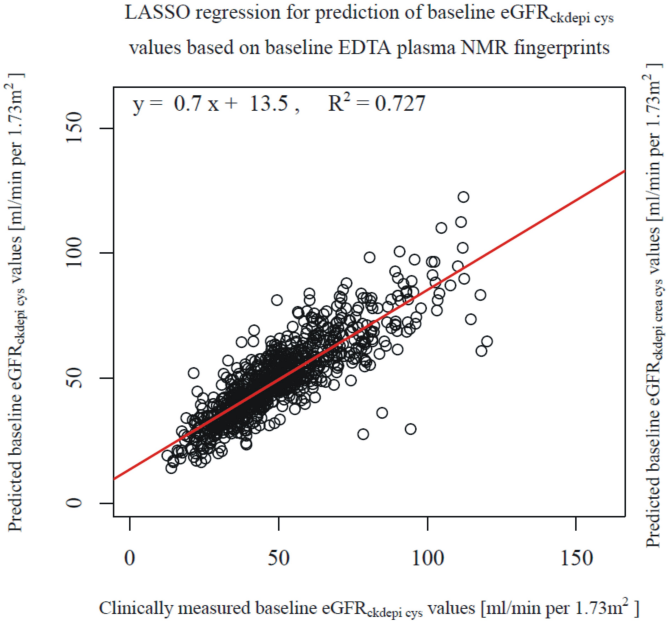
Figure 7.5: **Comparison of Synlab SCr and SCysC values at the baseline and FU2 time-point, respectively.** $x - y$ scatter plots of the compared eGFR values including equations of the fitted simple linear regression lines, Pearson's correlation coefficients r and coefficients of determination $R^2 = r^2$. Abbreviations: FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.


LASSO regression for prediction of baseline Synlab SCr values
based on baseline EDTA plasma NMR fingerprints
y = 1 x + 0 , R² = 0.936
Predicted baseline Synlab SCr values [mg/dl]
Clinically measured baseline Synlab SCr values [mg/dl]


LASSO regression for prediction of baseline Synlab SCysC values
based on baseline EDTA plasma NMR fingerprints
y = 0.8 x + 0.4 , R² = 0.739
Predicted baseline Synlab SCysC values [mg/l]
Clinically measured baseline Synlab SCysC values [mg/l]


LASSO regression for prediction of baseline eGFR mdrd4
values based on baseline EDTA plasma NMR fingerprints
y = 0.8 x + 8.3 , R² = 0.843
Predicted baseline eGFR mdrd4 values [ml/min per 1.73m²]
Clinically measured baseline eGFR mdrd4 values [ml/min per 1.73m²]


LASSO regression for prediction of baseline eGFR ckdepi crea
values based on baseline EDTA plasma NMR fingerprints
y = 0.8 x + 7.2 , R² = 0.86
Predicted baseline eGFR ckdepi crea values [ml/min per 1.73m²]
Clinically measured baseline eGFR ckdepi crea values [ml/min per 1.73m²]


LASSO regression for prediction of baseline eGFR ckdepi cys
values based on baseline EDTA plasma NMR fingerprints
y = 0.7 x + 13.5 , R² = 0.727
Predicted baseline eGFR ckdepi cys values [ml/min per 1.73m²]
Clinically measured baseline eGFR ckdepi cys values [ml/min per 1.73m²]

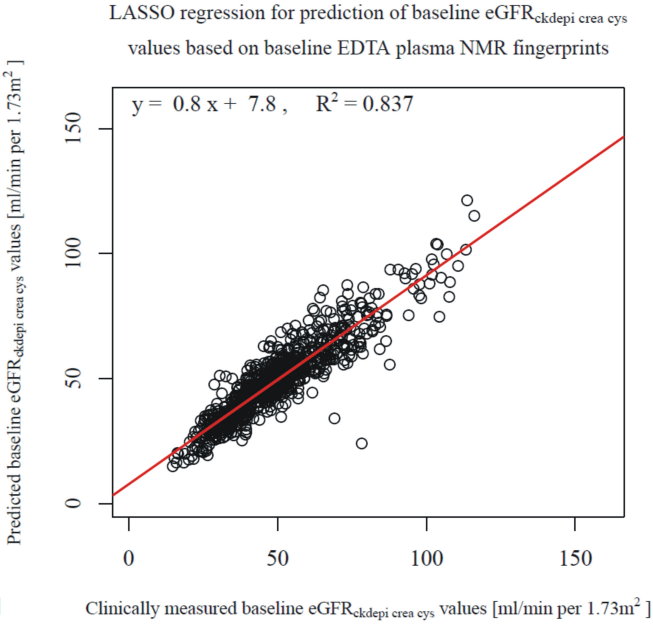
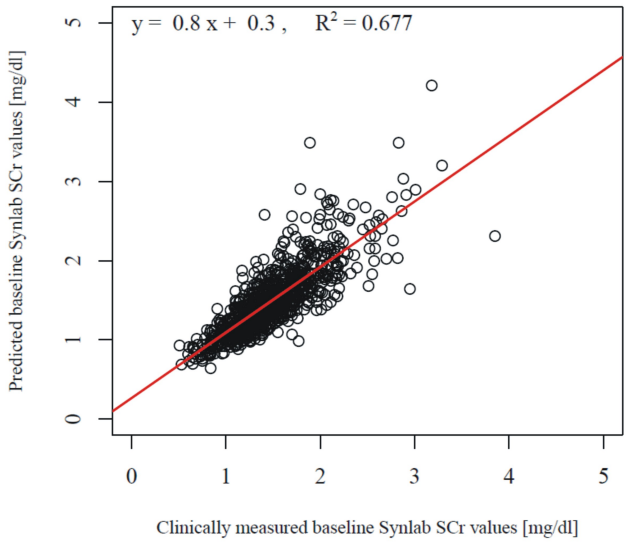
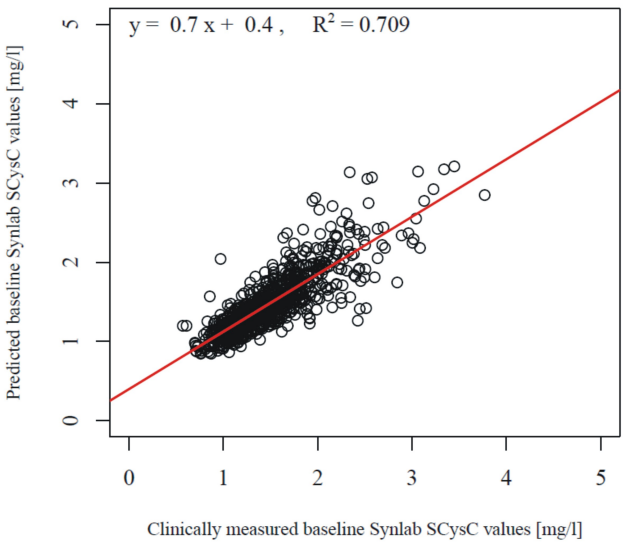

LASSO regression for prediction of baseline eGFR ckdepi crea cys
values based on baseline EDTA plasma NMR fingerprints
y = 0.8 x + 7.8 , R² = 0.837
Predicted baseline eGFR ckdepi crea cys values [ml/min per 1.73m²]
Clinically measured baseline eGFR ckdepi crea cys values [ml/min per 1.73m²]

Figure 7.6: **Previous page: Results of LASSO regression analysis for prediction of baseline SCr, SCysC, and eGFR values including all 660 NMR bins.** Displayed are scatter plots of the not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables of the test set, including a linear model fitted between these true and predicted response variables and the corresponding coefficients of determination R^2 between these variables. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; SCr, serum creatinine; SCysC, serum cystatin C.

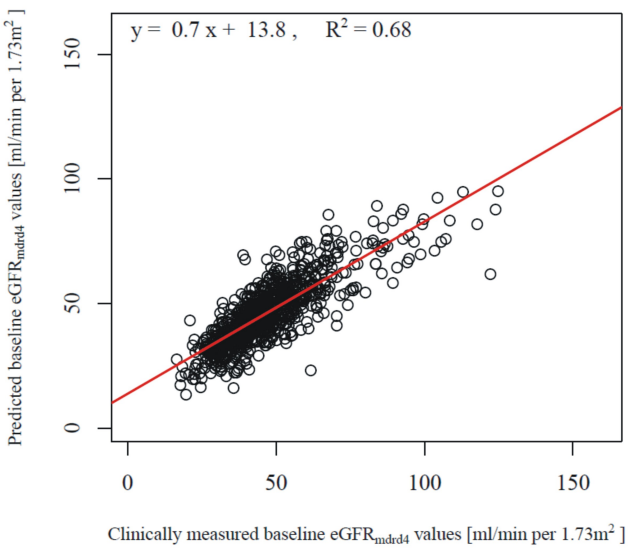
LASSO regression for prediction of baseline Synlab SCr values based on baseline EDTA plasma NMR fingerprints



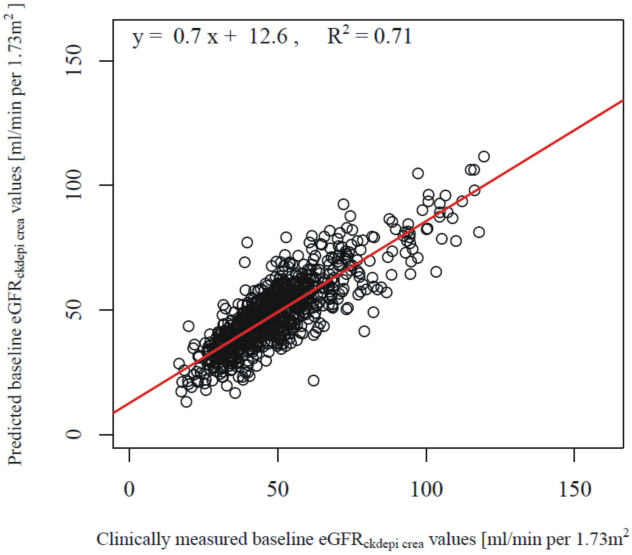
LASSO regression for prediction of baseline Synlab SCysC values based on baseline EDTA plasma NMR fingerprints



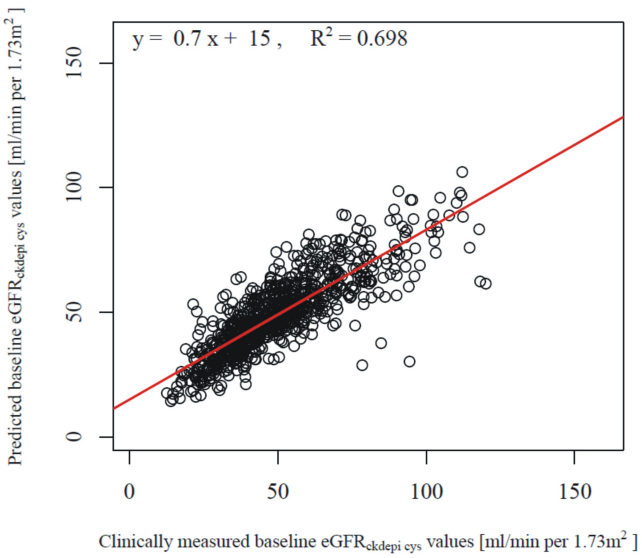
LASSO regression for prediction of baseline $eGFR_{mdrd4}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of baseline $eGFR_{ckdepi\ crea}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of baseline $eGFR_{ckdepi\ cys}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of baseline $eGFR_{ckdepi\ crea\ cys}$ values based on baseline EDTA plasma NMR fingerprints

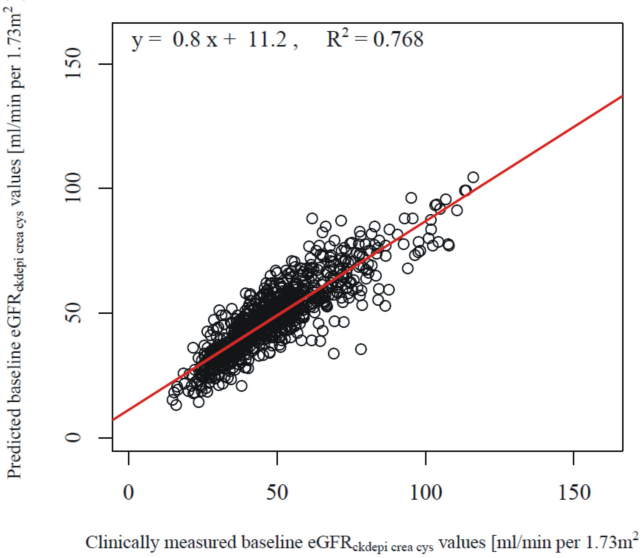
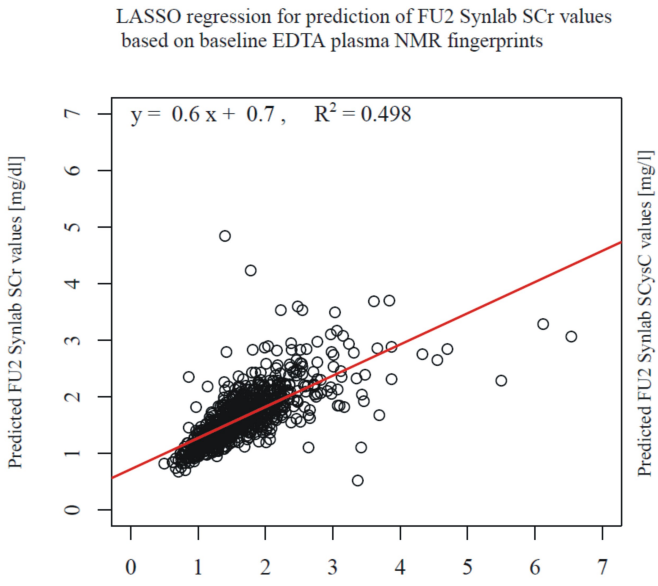
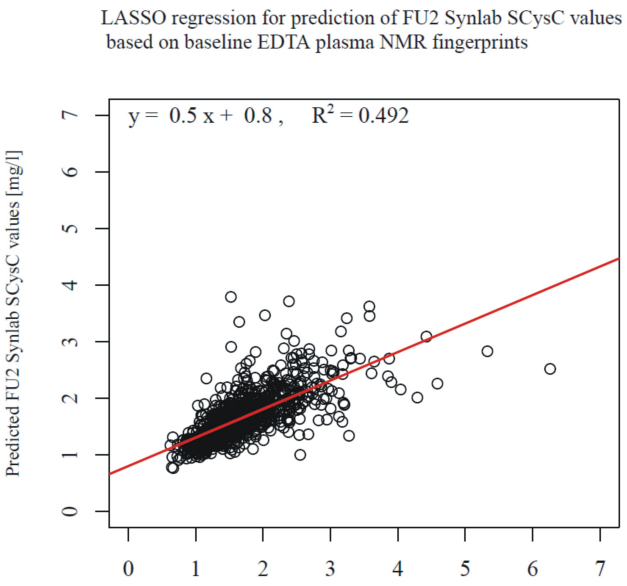
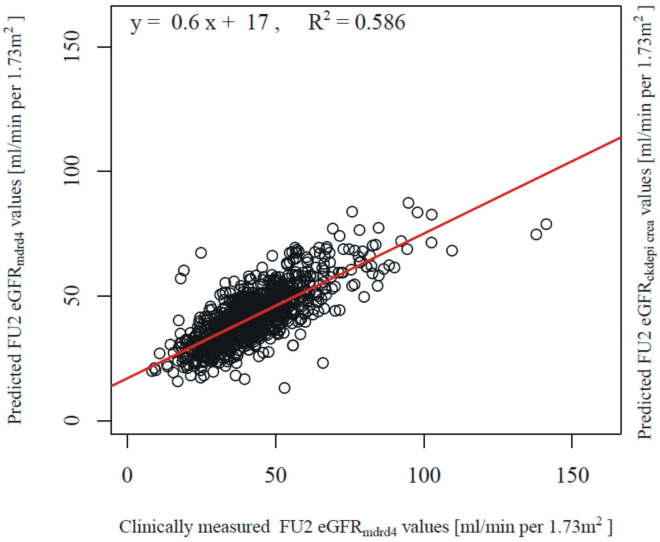
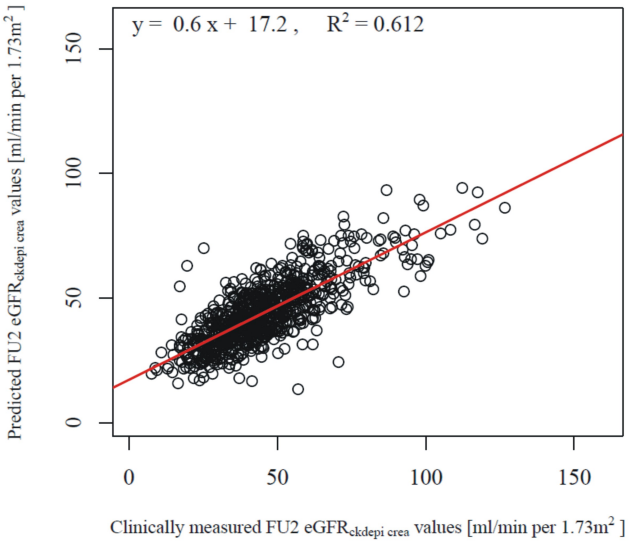


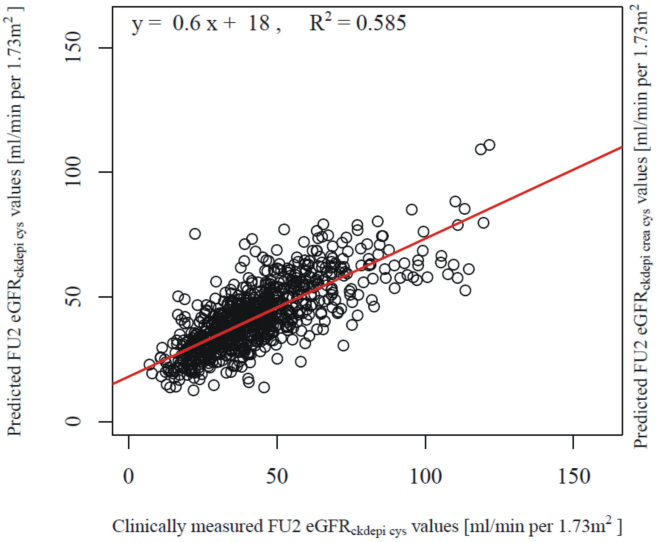
Figure 7.7: **Previous page: Results of LASSO regression analysis for prediction of baseline SCr, SCysC, and eGFR values after exclusion of all NMR buckets corresponding to creatinine.** Displayed are scatter plots of the not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables of the test set, including a linear model fitted between these true and predicted response variables and the corresponding coefficients of determination R^2 between these variables. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; SCr, serum creatinine; SCysC, serum cystatin C.


LASSO regression for prediction of FU2 Synlab SCr values
based on baseline EDTA plasma NMR fingerprints
y = 0.6 x + 0.7 , R² = 0.498
Predicted FU2 Synlab SCr values [mg/dl]
Clinically measured FU2 Synlab SCr values [mg/dl]


LASSO regression for prediction of FU2 Synlab SCysC values
based on baseline EDTA plasma NMR fingerprints
y = 0.5 x + 0.8 , R² = 0.492
Predicted FU2 Synlab SCysC values [mg/l]
Clinically measured FU2 Synlab SCysC values [mg/l]


LASSO regression for prediction of FU2 eGFR mdrd4
values based on baseline EDTA plasma NMR fingerprints
y = 0.6 x + 17 , R² = 0.586
Predicted FU2 eGFR mdrd4 values [ml/min per 1.73m²]
Clinically measured FU2 eGFR mdrd4 values [ml/min per 1.73m²]


LASSO regression for prediction of FU2 eGFR ckdepi crea
values based on baseline EDTA plasma NMR fingerprints
y = 0.6 x + 17.2 , R² = 0.612
Predicted FU2 eGFR ckdepi crea values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi crea values [ml/min per 1.73m²]


LASSO regression for prediction of FU2 eGFR ckdepi cys
values based on baseline EDTA plasma NMR fingerprints
y = 0.6 x + 18 , R² = 0.585
Predicted FU2 eGFR ckdepi cys values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi cys values [ml/min per 1.73m²]

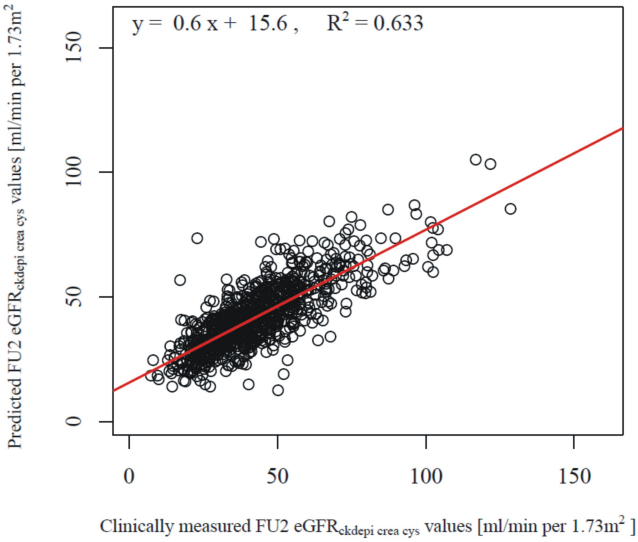
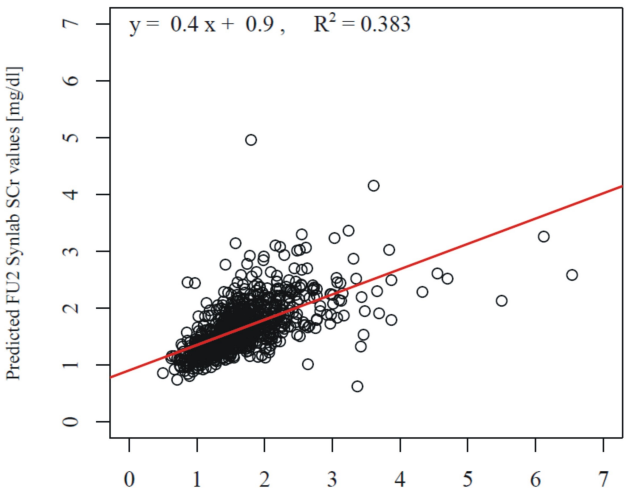
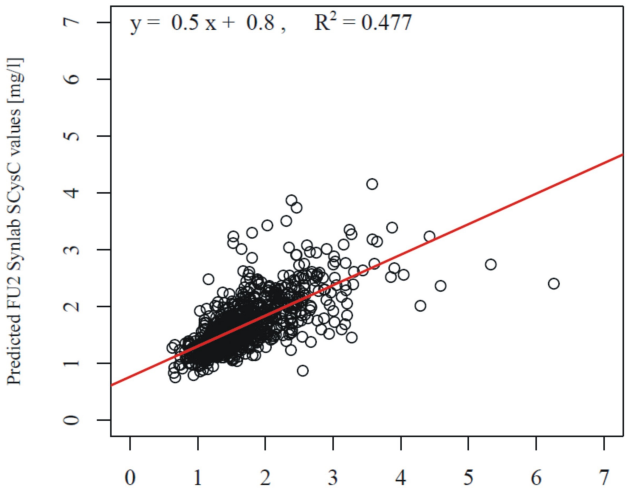

LASSO regression for prediction of FU2 eGFR ckdepi crea cys
values based on baseline EDTA plasma NMR fingerprints
y = 0.6 x + 15.6 , R² = 0.633
Predicted FU2 eGFR ckdepi crea cys values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi crea cys values [ml/min per 1.73m²]

Figure 7.8: **Previous page: Results of LASSO regression analysis for prediction of FU2 SCr, SCysC, and eGFR values.** Displayed are scatter plots of the not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables of the test set, including a linear model fitted between these true and predicted response variables and the corresponding coefficients of determination R^2 between these variables. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.

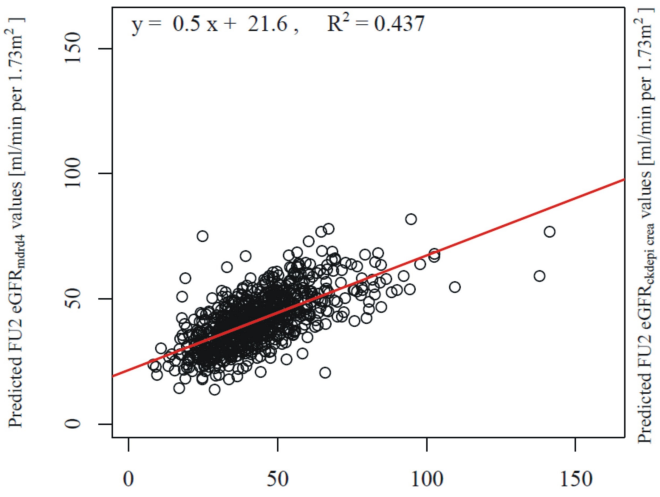
LASSO regression for prediction of FU2 Synlab SCr values based on baseline EDTA plasma NMR fingerprints



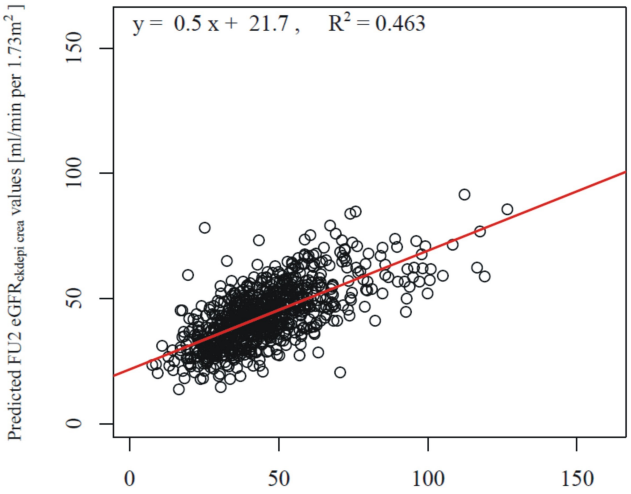
LASSO regression for prediction of FU2 Synlab SCysC values based on baseline EDTA plasma NMR fingerprints



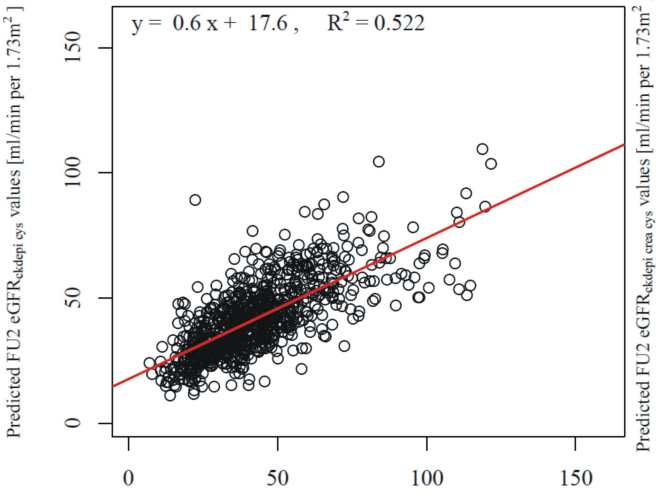
LASSO regression for prediction of FU2 $eGFR_{mdrd4}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of FU2 $eGFR_{ckdepi\ crea}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of FU2 $eGFR_{ckdepi\ cys}$ values based on baseline EDTA plasma NMR fingerprints



LASSO regression for prediction of FU2 $eGFR_{ckdepi\ crea\ cys}$ values based on baseline EDTA plasma NMR fingerprints

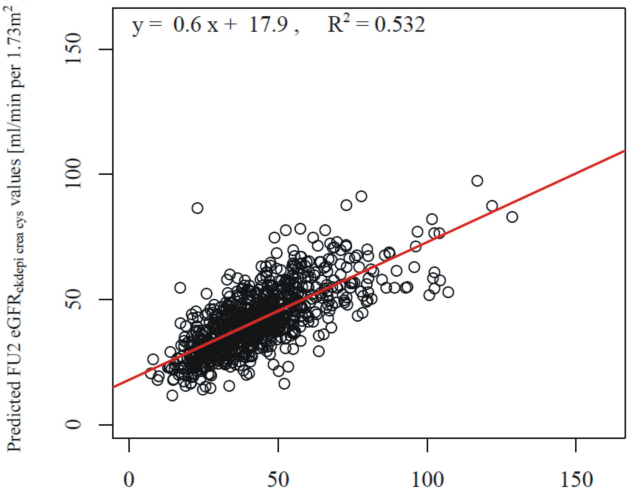
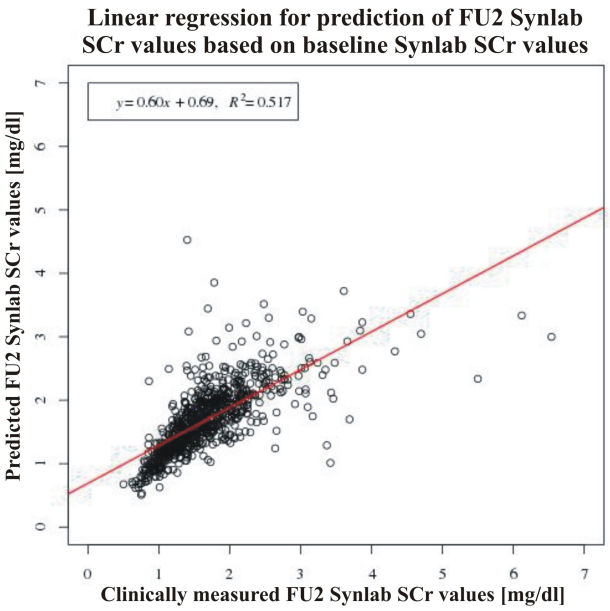
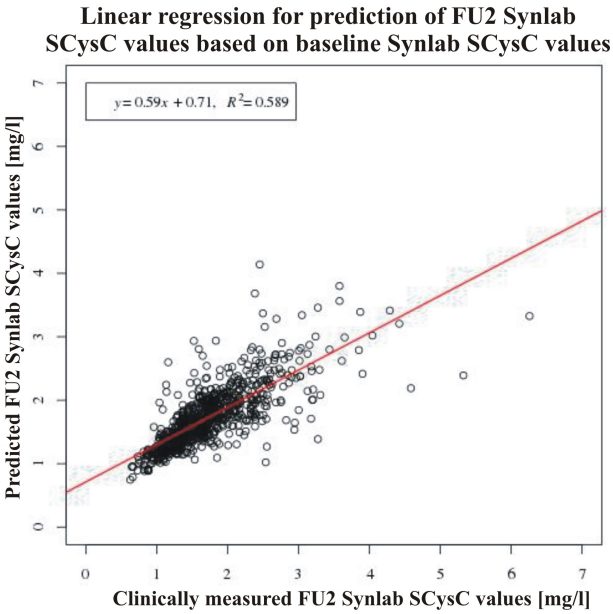
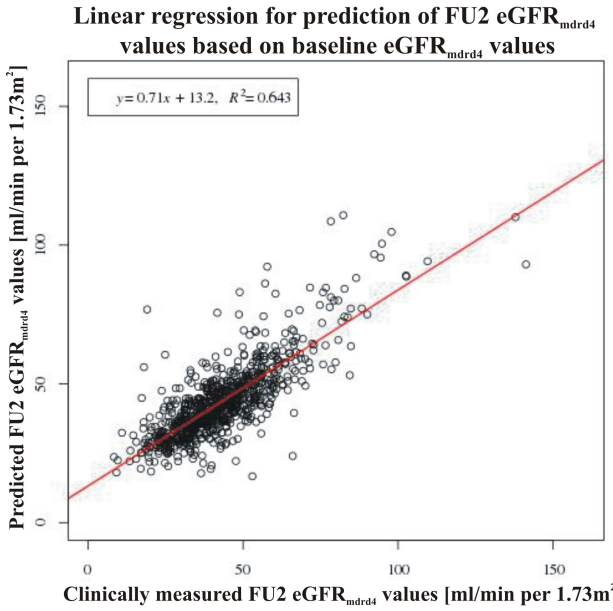
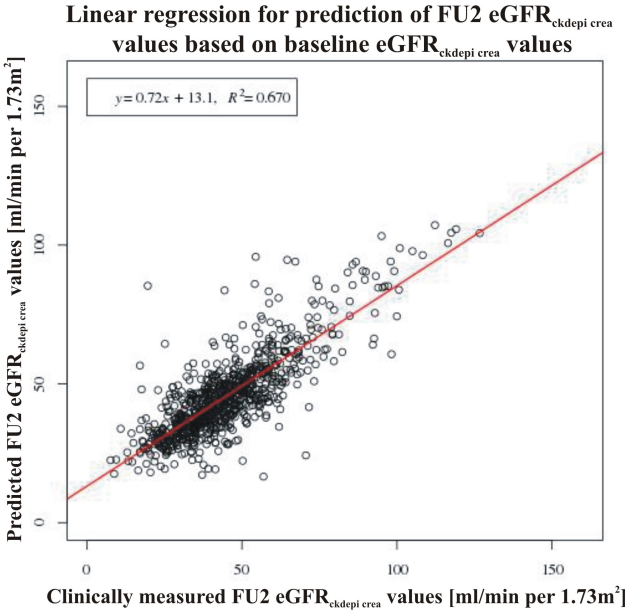


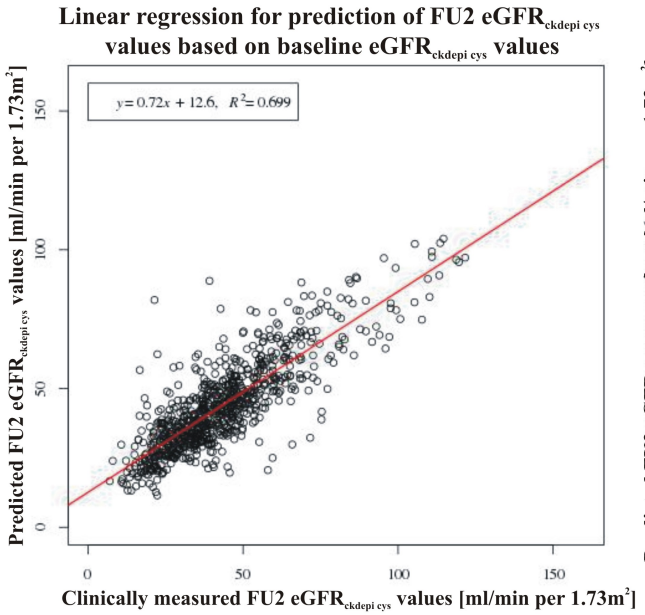
Figure 7.9: **Previous page: Results of LASSO regression analysis for prediction of FU2 SCr, SCysC, and eGFR values after exclusion of all NMR buckets corresponding to creatinine.** Displayed are scatter plots of the not $\log_2$ transformed true and inversely $\log_2$ transformed predicted response variables of the test set, including a linear model fitted between these true and predicted response variables and the corresponding coefficients of determination R^2 between these variables. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.


Linear regression for prediction of FU2 Synlab SCr values based on baseline Synlab SCr values
y = 0.60x + 0.69, R² = 0.517
Predicted FU2 Synlab SCr values [mg/dl]
Clinically measured FU2 Synlab SCr values [mg/dl]


Linear regression for prediction of FU2 Synlab SCysC values based on baseline Synlab SCysC values
y = 0.59x + 0.71, R² = 0.589
Predicted FU2 Synlab SCysC values [mg/l]
Clinically measured FU2 Synlab SCysC values [mg/l]


Linear regression for prediction of FU2 eGFR mdrd4 values based on baseline eGFR mdrd4 values
y = 0.71x + 13.2, R² = 0.643
Predicted FU2 eGFR mdrd4 values [ml/min per 1.73m²]
Clinically measured FU2 eGFR mdrd4 values [ml/min per 1.73m²]


Linear regression for prediction of FU2 eGFR ckdepi crea values based on baseline eGFR ckdepi crea values
y = 0.72x + 13.1, R² = 0.670
Predicted FU2 eGFR ckdepi crea values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi crea values [ml/min per 1.73m²]


Linear regression for prediction of FU2 eGFR ckdepi cys values based on baseline eGFR ckdepi cys values
y = 0.72x + 12.6, R² = 0.699
Predicted FU2 eGFR ckdepi cys values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi cys values [ml/min per 1.73m²]

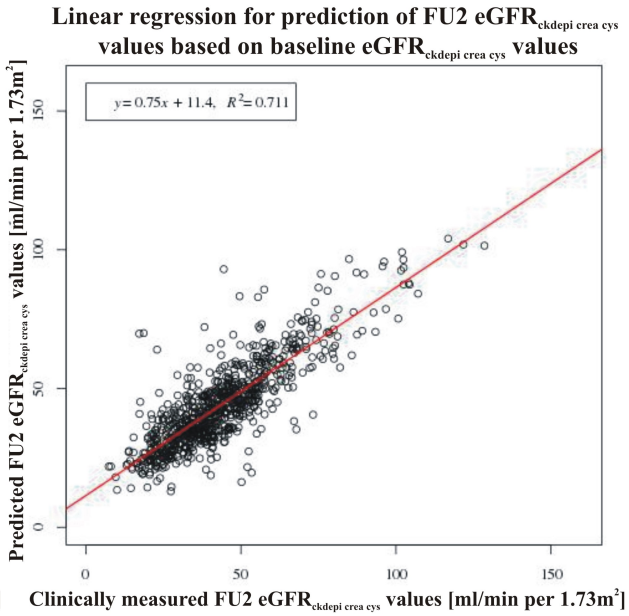
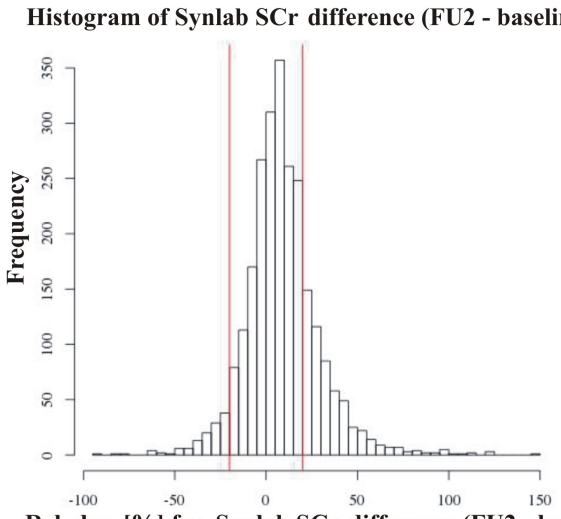
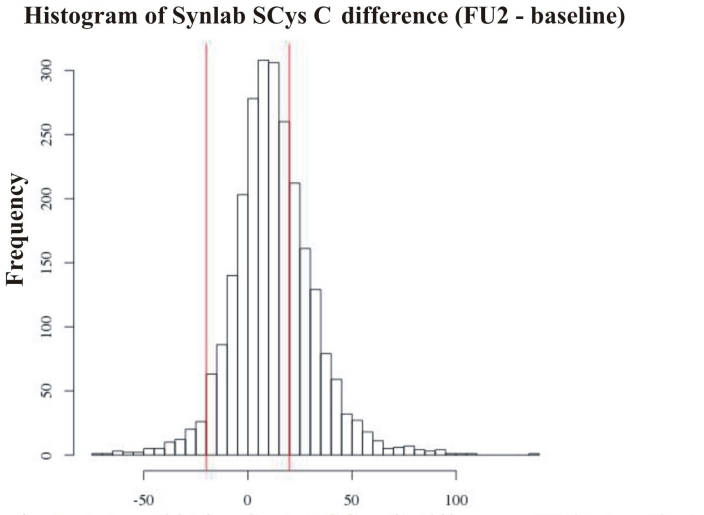
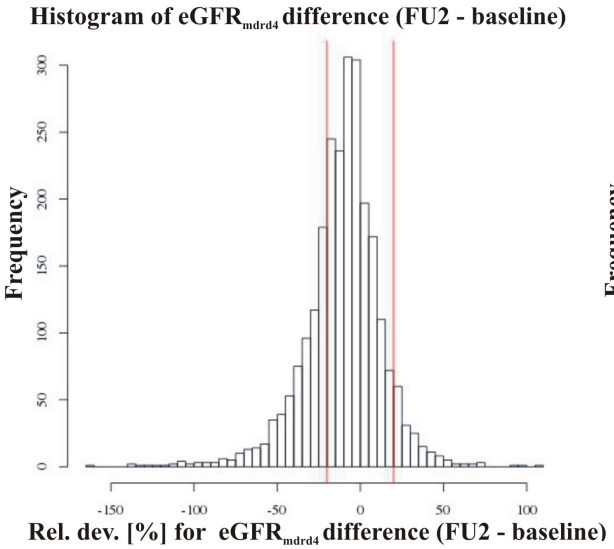
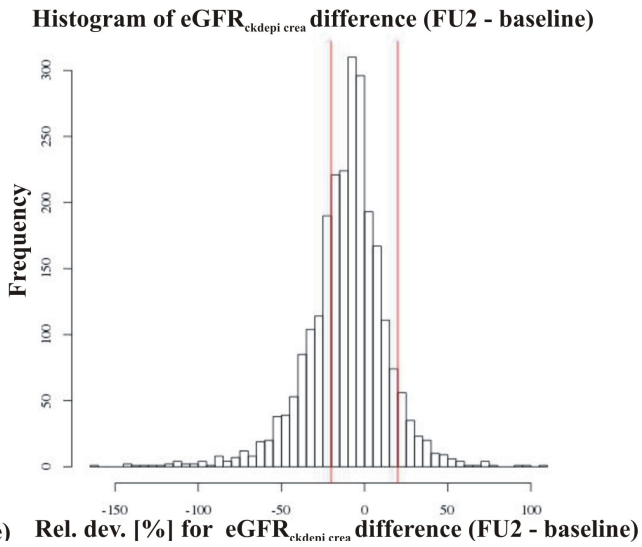

Linear regression for prediction of FU2 eGFR ckdepi crea cys values based on baseline eGFR ckdepi crea cys values
y = 0.75x + 11.4, R² = 0.711
Predicted FU2 eGFR ckdepi crea cys values [ml/min per 1.73m²]
Clinically measured FU2 eGFR ckdepi crea cys values [ml/min per 1.73m²]

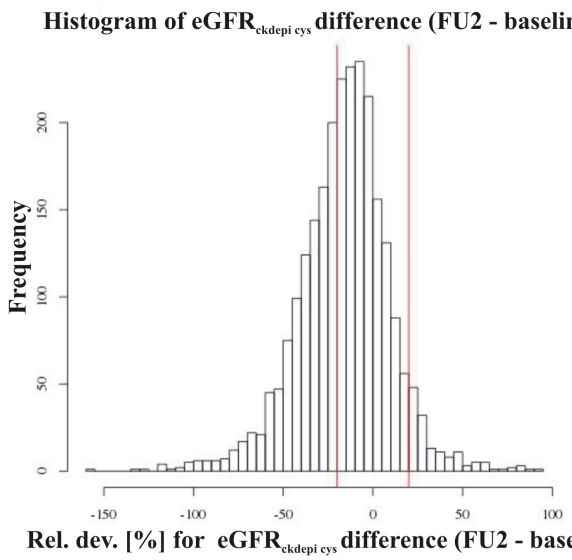
Figure 7.10: **Previous page: Results of simple linear regression analysis for prediction of FU2 SCr, SCysC, and eGFR values based on corresponding baseline values.** Displayed are scatter plots of the originally not $\log_2$ transformed true and predicted response variables of the test set, including a linear model fitted between these true and predicted response variables and the corresponding coefficients of determination R^2 between these variables. Abbreviations: eGFR, estimated glomerular filtration rate; $\text{eGFR}_{\text{ckdepi crea}}$, eGFR based on CKD-EPI crea formula; $\text{eGFR}_{\text{ckdepi crea cys}}$, eGFR based on CKD-EPI crea cys formula; $\text{eGFR}_{\text{ckdepi cys}}$, eGFR based on CKD-EPI cys formula; $\text{eGFR}_{\text{mdrd4}}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.


Histogram of Synlab SCr difference (FU2 - baseline)
Frequency
Rel. dev. [%] for Synlab SCr difference (FU2 - baseline)


Histogram of Synlab SCys C difference (FU2 - baseline)
Frequency
Rel. dev. [%] for Synlab SCys C difference (FU2 - baseline)


Histogram of eGFR mdrd4 difference (FU2 - baseline)
Frequency
Rel. dev. [%] for eGFR mdrd4 difference (FU2 - baseline)


Histogram of eGFR ckdepi crea difference (FU2 - baseline)
Frequency
Rel. dev. [%] for eGFR ckdepi crea difference (FU2 - baseline)


Histogram of eGFR ckdepi cys difference (FU2 - baseline)
Frequency
Rel. dev. [%] for eGFR ckdepi cys difference (FU2 - baseline)

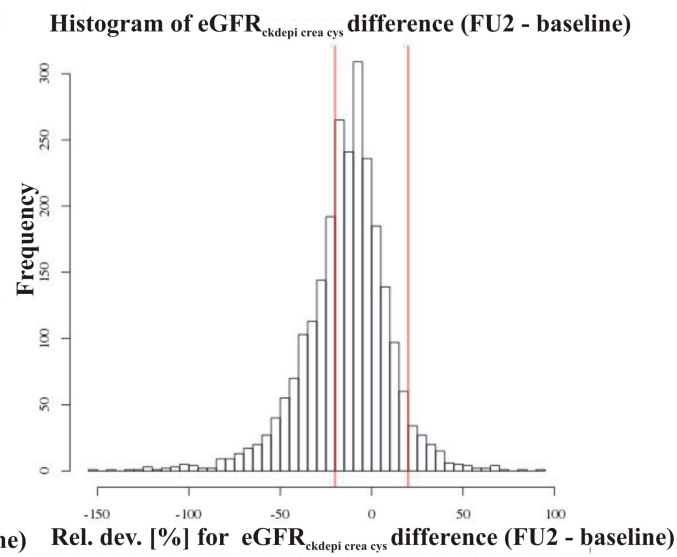

Histogram of eGFR ckdepi crea cys difference (FU2 - baseline)
Frequency
Rel. dev. [%] for eGFR ckdepi crea cys difference (FU2 - baseline)

Figure 7.11: **Previous page: Histograms of (FU2 - baseline) kidney performance clinical parameters.** Abbreviations: eGFR, estimated glomerular filtration rate; $eGFR_{ckdepi\ crea}$, eGFR based on CKD-EPI crea formula; $eGFR_{ckdepi\ crea\ cys}$, eGFR based on CKD-EPI crea cys formula; $eGFR_{ckdepi\ cys}$, eGFR based on CKD-EPI cys formula; $eGFR_{mdrd4}$, eGFR based on MDRD4 formula; FU2, second follow-up; SCr, serum creatinine; SCysC, serum cystatin C.

7.4 Appendix IV: Trial to Reduce Cardiovascular Events with Aranesp® Therapy study

Spectral position [ppm]	***P*-value un-adjusted**	***P*-value B/H-adjusted**	**ID**	**Identified compounds**
2.845	6.46e-05	0.0135	467	Proteins
2.855	6.76e-05	0.0135	466	Proteins, unknown
2.045	9.45e-05	0.0135	547	Proteins, N-acetylglycine, proline, L-pyroglutamic acid, N-acetyl-L-glutamine (?)
3.995	9.88e-05	0.0135	352	Proteins, unknown
1.885	1.09e-04	0.0135	563	Proteins, quinic acid
1.875	1.75e-04	0.0135	564	Proteins, quinic acid
2.065	1.90e-04	0.0135	545	Proteins, quinic acid, proline, L-pyroglutamic acid
3.755	2.00e-04	0.0135	376	Proteins, D-glucose, D-mannitol, leucine, sucrose, pseudouridine
4.325	2.29e-04	0.0135	319	Proteins
4.495	2.47e-04	0.0135	302	Proteins
3.685	2.59e-04	0.0135	383	Proteins, D-mannitol, L-isoleucine (?), sucrose
3.505	2.89e-04	0.0135	401	D-glucose, sucrose
3.345	3.29e-04	0.0135	417	Proteins, proline
2.055	3.54e-04	0.0135	546	Proteins, proline, quinic acid, L-pyroglutamic acid
1.675	3.79e-04	0.0135	584	Proteins, leucine, agmatine
2.085	3.91e-04	0.0135	543	Proteins, proline, N-acetyl-L-glutamine (?), N-methyl-L-proline
4.275	4.09e-04	0.0135	324	Proteins, threonine
3.865	4.21e-04	0.0135	365	D-glucose, D-mannitol, pseudouridine, sucrose
1.175	4.25e-04	0.0135	634	Proteins
2.995	4.48e-04	0.0135	452	Proteins
2.075	4.51e-04	0.0135	544	Proteins, proline, N-methyl-L-proline
2.755	4.77e-04	0.0135	476	Proteins
1.895	4.85e-04	0.0135	562	Proteins, N-acetyl-L-glutamine
4.155	5.42e-04	0.0135	336	Proteins, proline, pseudouridine, N-acetyl-L-glutamine
4.255	5.43e-04	0.0135	326	Proteins, threonine
8.465	5.60e-04	0.0135	104	Proteins, formic acid

2.015	5.66e-04	0.0135	550	Proteins, proline, L-pyroglutamic acid, L-isoleucine
4.315	5.74e-04	0.0135	320	Proteins, unknown
3.775	5.95e-04	0.0135	374	D-glucose, D-mannitol, sucrose
3.665	6.03e-04	0.0135	385	Unknown, proteins, D-mannitol, paracetamol-glucuronide
2.745	6.17e-04	0.0135	477	Proteins
1.685	6.19e-04	0.0135	583	Proteins, leucine, agmatine
2.865	6.34e-04	0.0135	465	Proteins, asparagine (?)
3.585	6.69e-04	0.0138	393	Unknown, sucrose
4.165	7.01e-04	0.0139	335	Proteins, L-pyroglutamic acid, pseudouridine, N-acetyl-L-glutamine
2.025	7.13e-04	0.0139	549	Proteins, proline, L-pyroglutamic acid, N-methyl-L-proline (?)
1.835	7.69e-04	0.0144	568	Proteins
4.455	8.02e-04	0.0144	306	Proteins, trigonelline, D-lactose (?), unknown
0.975	8.13e-04	0.0144	654	Proteins, leucine, 2-aminobutyric acid
4.425	8.42e-04	0.0144	309	Proteins, unknown
8.005	8.43e-04	0.0144	150	Proteins, unknown
3.085	8.85e-04	0.0146	443	Proteins, unknown
3.765	8.96e-04	0.0146	375	D-glucose, D-mannitol
1.905	9.37e-04	0.0149	561	Proteins, N-acetyl-L-glutamine, unknown
2.645	9.89e-04	0.0152	487	Proteins, unknown
3.855	1.02e-03	0.0152	366	D-glucose, D-mannitol, sucrose, pseudouridine
0.925	1.04e-03	0.0152	659	Proteins
4.475	1.04e-03	0.0152	304	Proteins, D-lactose
4.145	1.08e-03	0.0155	337	Proteins, proline, pseudouridine, quinic acid
1.965	1.10e-03	0.0155	555	Proteins, quinic acid, L-isoleucine, N-acetyl-L-glutamine
4.305	1.16e-03	0.0156	321	Proteins, pseudouridine, unknown
4.005	1.16e-03	0.0156	351	Proteins, quinic acid, phosphoethanolamine, ascorbic acid, unknown (?)
2.885	1.26e-03	0.0158	463	Proteins, unknown
0.985	1.29e-03	0.0158	653	Proteins, valine, 2-aminobutyric acid
3.785	1.31e-03	0.0158	373	D-glucose, sucrose, glutamine
2.975	1.31e-03	0.0158	454	Proteins, unknown
4.355	1.32e-03	0.0158	316	Proteins, cis-4-hydroxy-D-proline
4.135	1.32e-03	0.0158	338	Proline, proteins, unknown
1.825	1.34e-03	0.0158	569	Proteins
8.235	1.38e-03	0.0158	127	Proteins
4.415	1.44e-03	0.0158	310	Proteins, unknown
8.325	1.45e-03	0.0158	118	Proteins, unknown
2.805	1.46e-03	0.0158	471	Proteins, unknown
3.675	1.46e-03	0.0158	384	Proteins, D-mannitol, L-isoleucine
2.395	1.51e-03	0.0158	512	Proteins, L-pyroglutamic acid, unknown
2.035	1.51e-03	0.0158	548	Proteins, proline, L-pyroglutamic acid, N-acetyl-L-glutamine (?)

3.515	1.59e-03	0.0158	400	D-glucose
3.525	1.59e-03	0.0158	399	Proteins, D-glucose
2.765	1.63e-03	0.0158	475	Proteins, unknown
3.335	1.63e-03	0.0158	418	Proline, proteins, unknown
0.945	1.63e-03	0.0158	657	L-isoleucine, proteins
0.825	1.68e-03	0.0158	669	Proteins
0.845	1.68e-03	0.0158	667	Proteins
4.225	1.72e-03	0.0158	329	Sucrose, proteins
1.695	1.77e-03	0.0158	582	Proteins, leucine, agmatine
0.725	1.78e-03	0.0158	679	Proteins
8.695	1.78e-03	0.0158	81	Proteins
4.375	1.80e-03	0.0158	314	Proteins, cis-4-hydroxy-D-proline
4.245	1.80e-03	0.0158	327	Proteins, threonine
2.005	1.81e-03	0.0158	551	Proteins, proline, L-pyroglutamic acid, L-isoleucine
1.665	1.84e-03	0.0159	585	Proteins, leucine, agmatine
2.385	1.90e-03	0.0162	513	Proline, proteins, pyruvic acid (?)
8.245	1.94e-03	0.0163	126	Proteins
3.535	1.95e-03	0.0163	398	D-glucose
3.065	2.06e-03	0.0170	445	Unknown, proteins
4.235	2.10e-03	0.0171	328	Sucrose, proteins
1.815	2.15e-03	0.0173	570	Proteins
8.175	2.18e-03	0.0174	133	Proteins
4.335	2.21e-03	0.0174	318	Proteins, cis-4-hydroxy-D-proline
6.655	2.30e-03	0.0178	285	Proteins
4.085	2.31e-03	0.0178	343	Proteins, unknown
1.955	2.37e-03	0.0181	556	Proteins, N-acetyl-L-glutamine, L-isoleucine, quinic acid
0.815	2.41e-03	0.0182	670	Proteins
2.235	2.49e-03	0.0182	528	Acetone, proteins
8.765	2.49e-03	0.0182	74	Proteins
1.865	2.53e-03	0.0182	565	Proteins, unknown
2.225	2.53e-03	0.0182	529	Proteins
8.195	2.56e-03	0.0182	131	Proteins
8.165	2.59e-03	0.0182	134	Proteins
0.935	2.65e-03	0.0182	658	L-isoleucine, proteins, 2-oxoisocaproic acid (?)
3.815	2.67e-03	0.0182	370	D-mannitol
2.815	2.68e-03	0.0182	470	Proteins
8.185	2.68e-03	0.0182	132	Proteins
4.295	2.74e-03	0.0182	322	Pseudouridine, proteins, unknown
4.285	2.74e-03	0.0182	323	Proteins, threonine, pseudouridine
1.645	2.75e-03	0.0182	587	Proteins
2.875	2.82e-03	0.0183	464	Proteins, unknown
2.985	2.83e-03	0.0183	453	Proteins, macromolecules
1.975	2.90e-03	0.0184	554	Proteins, proline, L-isoleucine
4.385	2.90e-03	0.0184	313	Proteins, unknown
1.655	2.93e-03	0.0184	586	Proteins, leucine, unknown

0.805	2.96e-03	0.0184	671	Proteins
8.945	2.97e-03	0.0184	56	Proteins
3.445	3.03e-03	0.0185	407	D-glucose, proline
2.915	3.03e-03	0.0185	460	Proteins
0.835	3.09e-03	0.0187	668	Proteins
1.635	3.19e-03	0.0191	588	Proteins
4.265	3.27e-03	0.0194	325	Proteins, threonine
0.965	3.38e-03	0.0198	655	Proteins, leucine
3.875	3.39e-03	0.0198	364	D-mannitol
1.405	3.42e-03	0.0198	611	Proteins, macromolecules
1.525	3.49e-03	0.0200	599	Proteins, propanol
3.705	3.53e-03	0.0200	381	D-glucose
4.015	3.60e-03	0.0200	350	Pseudouridine, proteins, unknown
3.985	3.61e-03	0.0200	353	Proteins, unknown
7.595	3.64e-03	0.0200	191	Proteins
3.645	3.66e-03	0.0200	387	Paracetamol-glucuronide, myo-inositol, unknown (?)
1.185	3.67e-03	0.0200	633	Proteins
2.375	3.68e-03	0.0200	514	Proline, proteins
3.655	3.70e-03	0.0200	386	Paracetamol-glucuronide, unknown
0.855	3.79e-03	0.0203	666	Proteins
2.315	4.01e-03	0.0212	520	Proteins, valine, N-acetyl-L-glutamine
1.845	4.04e-03	0.0212	567	Proteins
3.325	4.11e-03	0.0212	419	Proline, proteins
3.925	4.11e-03	0.0212	359	D-glucose
2.245	4.14e-03	0.0212	527	Proteins, valine, acetone
0.995	4.15e-03	0.0212	652	Valine, proteins, 2-aminobutyric acid
8.285	4.17e-03	0.0212	122	Proteins
1.915	4.26e-03	0.0214	560	Proteins, N-acetyl-L-glutamine, 2-aminobutyric acid (?)
4.365	4.27e-03	0.0214	315	Proteins, cis-4-hydroxy-D-proline
0.715	4.34e-03	0.0216	680	Proteins
3.365	4.43e-03	0.0216	415	Methanol, proline, proteins, cis-4-hydroxy-D-proline, unknown
1.585	4.47e-03	0.0216	593	Proteins, propanol
6.585	4.49e-03	0.0216	292	Proteins, unknown
1.085	4.52e-03	0.0216	643	Proteins, unknown
0.785	4.53e-03	0.0216	673	Proteins
3.795	4.53e-03	0.0216	372	D-glucose, sucrose, guanidinoacetic acid, D-mannitol
2.965	4.59e-03	0.0217	455	Proteins, macromolecules
4.025	4.65e-03	0.0217	349	Pseudouridine, quinic acid, ascorbic acid, proteins
0.795	4.66e-03	0.0217	672	Proteins
8.205	4.70e-03	0.0217	130	Proteins
6.625	4.71e-03	0.0217	288	Proteins
2.325	4.80e-03	0.0218	519	Proteins, proline, N-acetyl-L-glutamine, unknown
1.215	4.85e-03	0.0218	630	Proteins, 3-hydroxybutyric acid (?)
8.255	4.85e-03	0.0218	125	Proteins, unknown

1.005	4.86e-03	0.0218	651	Proteins, valine, L-isoleucine
0.735	4.95e-03	0.0220	678	Proteins, macromolecules
3.075	4.95e-03	0.0220	444	Unknown, proteins
3.555	5.06e-03	0.0223	396	D-glucose, sucrose, myo-inositol, quinic acid (?)
2.095	5.23e-03	0.0227	542	Proline, proteins, N-acetyl-L-glutamine, N-methyl-L-proline (?)
3.175	5.24e-03	0.0227	434	Isethionic acid, proteins, N-methyl-L-proline (?), unknown
2.825	5.25e-03	0.0227	469	Methylguanidine (?), proteins
1.595	5.31e-03	0.0228	592	Proteins, unknown
0.875	5.41e-03	0.0231	664	Proteins, ibuprofen (?)
1.995	5.45e-03	0.0232	552	Proline, proteins, L-isoleucine, N-methyl-L-proline (?)
4.125	5.59e-03	0.0236	339	Proline, proteins, unknown
3.355	5.77e-03	0.0242	416	Proline, scyllo-inositol, proteins
3.825	5.84e-03	0.0244	369	D-glucose, sucrose
4.035	5.89e-03	0.0244	348	Pseudouridine, proteins
0.865	5.94e-03	0.0244	665	Proteins
4.175	5.98e-03	0.0244	334	L-pyroglutamic acid, N-acetyl-L-glutamine, proteins
0.955	5.99e-03	0.0244	656	Proteins, leucine, L-isoleucine
3.895	6.02e-03	0.0244	362	D-glucose, paracetamol-glucuronide, sucrose
4.395	6.14e-03	0.0244	312	Proteins, unknown
1.415	6.15e-03	0.0244	610	Proteins, macromolecules
1.345	6.16e-03	0.0244	617	Proteins, threonine, lactic acid (?)
4.465	6.18e-03	0.0244	305	Proteins, D-lactose
7.675	6.20e-03	0.0244	183	Pseudouridine, proteins, unknown
3.915	6.44e-03	0.0251	360	D-glucose, sucrose
2.425	6.44e-03	0.0251	509	L-pyroglutamic acid, glutamine (?), proteins, unknown
0.775	6.58e-03	0.0254	674	Proteins
2.635	6.61e-03	0.0254	488	Proteins, unknown
3.375	6.63e-03	0.0254	414	Proteins, unknown
1.205	6.68e-03	0.0254	631	Proteins, 3-hydroxybutyric acid (?), 3-aminoisobutyric acid (?), unknown
3.385	6.70e-03	0.0254	413	Proteins, unknown
8.705	6.74e-03	0.0254	80	Proteins, macromolecules
1.435	6.78e-03	0.0254	608	Proteins, unknown
2.795	6.97e-03	0.0260	472	Proteins, unknown
3.845	7.02e-03	0.0260	367	D-glucose, sucrose, pseudouridine
2.135	7.24e-03	0.0265	538	N-acetyl-L-glutamine, proteins, glutamine
1.575	7.25e-03	0.0265	594	Proteins, propanol
3.695	7.26e-03	0.0265	382	D-mannitol, proteins, unknown
1.625	7.50e-03	0.0272	589	Proteins
0.695	7.80e-03	0.0282	682	Proteins
3.475	7.92e-03	0.0285	404	D-glucose
8.145	7.97e-03	0.0285	136	Proteins
0.915	8.36e-03	0.0297	660	Proteins, propanol

2.655	8.42e-03	0.0297	486	Proteins, citric acid
8.935	8.43e-03	0.0297	57	Proteins
4.205	8.68e-03	0.0303	331	Proteins, unknown
1.705	8.68e-03	0.0303	581	Proteins, leucine, agmatine
1.285	8.88e-03	0.0308	623	Proteins, L-isoleucine
1.425	9.02e-03	0.0309	609	Proteins, unknown (?)
2.105	9.03e-03	0.0309	541	Proline, N-acetyl-L-glutamine, proteins
3.455	9.04e-03	0.0309	406	D-glucose, 4-hydroxyphenylacetic acid, unknown
1.015	9.19e-03	0.0312	650	L-isoleucine, proteins
2.895	9.24e-03	0.0312	462	Proteins, unknown
3.545	9.30e-03	0.0312	397	D-glucose, myo-inositol
1.535	9.31e-03	0.0312	598	Propanol, proteins
3.745	9.50e-03	0.0317	377	D-glucose, D-mannitol, pseudouridine, ascorbic acid (?), unknown
4.215	9.67e-03	0.0321	330	Sucrose, proteins
2.355	9.76e-03	0.0323	516	Proline, proteins, unknown (?)
8.585	9.90e-03	0.0326	92	Proteins, unknown
4.095	9.95e-03	0.0326	342	Proteins, unknown
2.405	9.98e-03	0.0326	511	L-pyroglutamic acid, proteins, succinic acid
1.235	1.01e-02	0.0328	628	L-isoleucine, proteins
1.225	1.02e-02	0.0328	629	Proteins, unknown
8.295	1.02e-02	0.0328	121	Proteins
0.705	1.03e-02	0.0328	681	Proteins
9.175	1.03e-02	0.0328	33	Proteins
8.135	1.04e-02	0.0328	137	Proteins, unknown (?)
3.495	1.04e-02	0.0328	402	D-glucose, sucrose
1.715	1.07e-02	0.0335	580	Proteins, leucine, agmatine
3.835	1.08e-02	0.0337	368	D-glucose, sucrose, pseudouridine
2.115	1.09e-02	0.0340	540	N-acetyl-L-glutamine, proteins, glutamine
7.995	1.10e-02	0.0340	151	Proteins
1.245	1.11e-02	0.0343	627	L-isoleucine, proteins, unknown
1.295	1.13e-02	0.0348	622	L-isoleucine, proteins, unknown
3.575	1.16e-02	0.0353	394	Sucrose, propanol, unknown
8.155	1.16e-02	0.0353	135	Proteins, macromolecules
3.245	1.17e-02	0.0353	427	D-glucose, agmatine
2.725	1.17e-02	0.0353	479	Dimethylamine
0.655	1.18e-02	0.0353	686	Proteins
2.945	1.18e-02	0.0353	457	N-methyl-L-proline (?), proteins, unknown
8.415	1.19e-02	0.0355	109	Proteins
8.895	1.21e-02	0.0360	61	Proteins
1.935	1.23e-02	0.0362	558	N-acetyl-L-glutamine, proteins
2.515	1.23e-02	0.0362	500	L-pyroglutamic acid, proteins
1.395	1.24e-02	0.0364	612	Proteins, macromolecules
3.395	1.28e-02	0.0373	412	D-glucose
0.905	1.30e-02	0.0376	661	Proteins, propanol

2.785	1.30e-02	0.0376	473	Proteins, unknown
9.265	1.36e-02	0.0393	24	Proteins
1.795	1.37e-02	0.0393	572	Proteins
8.915	1.39e-02	0.0396	59	Proteins
8.865	1.39e-02	0.0396	64	Proteins
1.455	1.40e-02	0.0397	606	L-isoleucine, proteins
3.255	1.41e-02	0.0399	426	D-glucose, agmatine
7.585	1.42e-02	0.0399	192	Proteins
6.685	1.44e-02	0.0405	282	Proteins, unknown
6.535	1.45e-02	0.0405	297	Proteins
1.025	1.48e-02	0.0412	649	L-isoleucine, proteins
3.805	1.49e-02	0.0412	371	D-mannitol, proteins
0.665	1.50e-02	0.0412	685	Proteins
1.615	1.50e-02	0.0412	590	Proteins
6.695	1.52e-02	0.0416	281	Proteins
1.145	1.54e-02	0.0420	637	Proteins
6.675	1.55e-02	0.0422	283	Proteins, unknown
3.025	1.57e-02	0.0425	449	Agmatine, 2-oxoglutaric acid, 3-aminoisobutyric acid
3.725	1.60e-02	0.0430	379	D-glucose, leucine, pseudouridine
1.195	1.61e-02	0.0430	632	Proteins, 3-aminoisobutyric acid
8.125	1.61e-02	0.0430	138	Proteins, macromolecules
0.625	1.68e-02	0.0447	689	Proteins
1.725	1.68e-02	0.0447	579	Proteins, leucine, agmatine
1.565	1.69e-02	0.0448	595	Proteins, propanol
9.095	1.70e-02	0.0448	41	Proteins
1.135	1.72e-02	0.0450	638	Proteins, unknown
1.805	1.76e-02	0.0461	571	Proteins
2.335	1.79e-02	0.0468	518	Proline, N-acetyl-L-glutamine, proteins
3.485	1.81e-02	0.0468	403	D-glucose, sucrose
4.435	1.82e-02	0.0468	308	Proteins, unknown
0.645	1.82e-02	0.0468	687	Proteins
2.415	1.82e-02	0.0468	510	L-pyroglutamic acid, succinic acid, proteins
8.715	1.86e-02	0.0473	79	Proteins, macromolecules
4.485	1.87e-02	0.0473	303	Proteins, unknown
3.465	1.87e-02	0.0473	405	D-glucose, sucrose
1.545	1.88e-02	0.0473	597	Propanol, proteins
0.685	1.88e-02	0.0473	683	Proteins
9.105	1.89e-02	0.0473	40	Proteins
0.765	1.89e-02	0.0473	675	Proteins
3.735	1.93e-02	0.0481	378	D-glucose, leucine
3.955	1.96e-02	0.0486	356	Isethionic acid, proteins, unknown
4.115	2.01e-02	0.0498	340	Proteins, unknown

Table 7.22: **Previous page: Spectral positions given in ppm, *p*-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, NMR IDs, as well as correspondingly identified compounds of NMR features that discriminated groups U and V of hypothesis 1a, i.e. patients dying from any cause versus patients not dying from any cause under the restriction that both groups do not progress to end-stage renal disease (ESRD).** The 256 urine specimens studied were collected in the last week directly before treatment randomization (W1). A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap.

Spectral position [ppm]	***P*-value unadjusted**	***P*-value B/H-adjusted**	**ID**	**Identified compounds**
0.955	1.19e-04	0.0178	656	Proteins, leucine (?)
8.585	1.44e-04	0.0178	92	Macromolecules, proteins, unknown (only visible in single spectra)
8.875	1.99e-04	0.0178	63	Macromolecules, proteins, 1-methylnicotinamide
0.965	2.30e-04	0.0178	655	Proteins, leucine (?)
9.315	2.58e-04	0.0178	19	Macromolecules, proteins
1.175	2.97e-04	0.0178	634	Proteins
0.945	2.97e-04	0.0178	657	Proteins, unknown
1.035	3.16e-04	0.0178	648	Proteins, valine
9.305	3.59e-04	0.0178	20	Macromolecules, proteins
9.265	3.99e-04	0.0178	24	Macromolecules, proteins
0.935	4.42e-04	0.0178	658	Proteins, unknown
1.025	4.45e-04	0.0178	649	Proteins, unknown
8.105	5.37e-04	0.0178	140	Macromolecules/proteins
9.235	5.48e-04	0.0178	27	Macromolecules/proteins
9.295	5.60e-04	0.0178	21	Macromolecules/proteins
8.665	5.65e-04	0.0178	84	Macromolecules/proteins
8.655	5.87e-04	0.0178	85	Macromolecules/proteins
9.355	5.97e-04	0.0178	15	Macromolecules/proteins
0.775	6.01e-04	0.0178	674	Proteins
8.595	6.49e-04	0.0178	91	Macromolecules/proteins
9.165	6.60e-04	0.0178	34	Macromolecules/proteins
0.535	7.06e-04	0.0178	698	Proteins
8.755	7.11e-04	0.0178	75	Macromolecules/proteins
9.405	7.97e-04	0.0178	10	Macromolecules/proteins
0.785	8.42e-04	0.0178	673	Proteins
9.005	8.65e-04	0.0178	50	Macromolecules/proteins
8.375	8.72e-04	0.0178	113	Macromolecules/proteins, unknown
0.545	8.74e-04	0.0178	697	Proteins

1.015	9.18e-04	0.0178	650	Proteins, unknown
9.455	9.28e-04	0.0178	5	Macromolecules/proteins
9.015	9.38e-04	0.0178	49	Macromolecules/proteins
9.345	9.77e-04	0.0178	16	Macromolecules/proteins
8.135	9.82e-04	0.0178	137	Macromolecules/proteins
8.605	9.84e-04	0.0178	90	Macromolecules/proteins
0.795	9.88e-04	0.0178	672	Proteins
1.005	1.01e-03	0.0178	651	Proteins, valine
7.865	1.01e-03	0.0178	164	Macromolecules/proteins
1.045	1.05e-03	0.0178	647	Proteins, unknown
0.555	1.15e-03	0.0178	696	Proteins
8.865	1.16e-03	0.0178	64	Macromolecules/proteins
0.765	1.19e-03	0.0178	675	Proteins, unknown
8.615	1.25e-03	0.0178	89	Macromolecules/proteins
0.615	1.26e-03	0.0178	690	Proteins
9.195	1.32e-03	0.0178	31	Macromolecules/proteins
0.515	1.34e-03	0.0178	700	Proteins
0.525	1.36e-03	0.0178	699	Proteins
8.645	1.37e-03	0.0178	86	Macromolecules/proteins
9.215	1.37e-03	0.0178	29	Macromolecules/proteins
0.925	1.45e-03	0.0178	659	Proteins, 3-methylglutaric acid (?)
9.145	1.47e-03	0.0178	36	Macromolecules/proteins
0.595	1.50e-03	0.0178	692	Proteins
1.695	1.55e-03	0.0178	582	Proteins
0.565	1.58e-03	0.0178	695	Proteins
0.655	1.58e-03	0.0178	686	Proteins
0.845	1.59e-03	0.0178	667	Proteins, unknown
0.605	1.59e-03	0.0178	691	Proteins
8.765	1.59e-03	0.0178	74	Macromolecules/proteins
1.165	1.62e-03	0.0178	635	Proteins
1.635	1.69e-03	0.0178	588	Proteins
0.825	1.70e-03	0.0178	669	Proteins, unknown
0.715	1.71e-03	0.0178	680	Proteins
8.695	1.72e-03	0.0178	81	Macromolecules/proteins
4.355	1.72e-03	0.0178	316	Proteins
1.645	1.73e-03	0.0178	587	Proteins
0.645	1.74e-03	0.0178	687	Proteins
1.085	1.74e-03	0.0178	643	Proteins
0.835	1.81e-03	0.0178	668	Proteins, unknown
9.105	1.84e-03	0.0178	40	Macromolecules/proteins
8.625	1.84e-03	0.0178	88	Macromolecules/proteins
8.635	1.85e-03	0.0178	87	Macromolecules/proteins
9.375	1.86e-03	0.0178	13	Macromolecules/proteins
0.625	1.86e-03	0.0178	689	Proteins
0.505	1.86e-03	0.0178	701	Proteins

8.745	1.89e-03	0.0178	76	Macromolecules/proteins
9.185	1.90e-03	0.0178	32	Macromolecules/proteins
9.175	2.02e-03	0.0185	33	Macromolecules/proteins
0.665	2.04e-03	0.0185	685	Proteins
0.805	2.06e-03	0.0185	671	Proteins
8.675	2.16e-03	0.0191	83	Macromolecules/proteins
0.585	2.22e-03	0.0195	693	Proteins
0.695	2.30e-03	0.0196	682	Proteins
0.685	2.31e-03	0.0196	683	Proteins
0.725	2.32e-03	0.0196	679	Proteins
0.705	2.35e-03	0.0196	681	Proteins
8.935	2.38e-03	0.0197	57	Macromolecules/proteins
3.065	2.53e-03	0.0206	445	Proteins
1.685	2.57e-03	0.0206	583	Proteins
9.435	2.61e-03	0.0206	7	Macromolecules/proteins
0.995	2.62e-03	0.0206	652	Proteins
9.085	2.73e-03	0.0209	42	Proteins
0.815	2.76e-03	0.0209	670	Proteins
8.115	2.84e-03	0.0209	139	Macromolecules/proteins
1.705	2.85e-03	0.0209	581	Proteins
7.785	2.85e-03	0.0209	172	Macromolecules/proteins
1.655	2.86e-03	0.0209	586	Proteins
0.635	2.90e-03	0.0209	688	Proteins
0.675	2.93e-03	0.0209	684	Proteins
1.595	2.94e-03	0.0209	592	Proteins
8.095	2.95e-03	0.0209	141	Macromolecules/proteins
1.885	3.01e-03	0.0210	563	Proteins
9.255	3.03e-03	0.0210	25	Macromolecules/proteins
9.415	3.10e-03	0.0211	9	Macromolecules/proteins
1.125	3.10e-03	0.0211	639	Proteins
8.125	3.26e-03	0.0219	138	Macromolecules/proteins
8.145	3.31e-03	0.0219	136	Macromolecules/proteins
0.975	3.33e-03	0.0219	654	Proteins, leucine (?)
0.855	3.37e-03	0.0219	666	Proteins
7.875	3.38e-03	0.0219	163	Macromolecules/proteins
1.345	3.43e-03	0.0219	617	Proteins, unknown
1.585	3.44e-03	0.0219	593	Proteins
8.235	3.49e-03	0.0220	127	Macromolecules/proteins
1.055	3.52e-03	0.0220	646	Proteins, valine
9.205	3.61e-03	0.0224	30	Macromolecules/proteins
0.735	3.67e-03	0.0226	678	Proteins
1.615	3.72e-03	0.0227	590	Proteins
0.985	3.87e-03	0.0230	653	Proteins, valine
4.345	3.89e-03	0.0230	317	Proteins, unknown
3.205	3.91e-03	0.0230	431	Unknown

4.075	3.91e-03	0.0230	344	Unknown
1.755	4.11e-03	0.0234	576	Proteins, leucine (?)
8.915	4.24e-03	0.0234	59	Macromolecules/proteins
1.605	4.24e-03	0.0234	591	Proteins
2.955	4.25e-03	0.0234	456	Proteins
0.915	4.27e-03	0.0234	660	Proteins, unknown
9.095	4.33e-03	0.0234	41	Macromolecules/proteins
1.245	4.33e-03	0.0234	627	Proteins
1.725	4.33e-03	0.0234	579	Proteins
9.075	4.35e-03	0.0234	43	Macromolecules/proteins
8.685	4.36e-03	0.0234	82	Macromolecules/proteins
1.135	4.37e-03	0.0234	638	Proteins
8.565	4.41e-03	0.0234	94	Macromolecules/proteins
1.715	4.42e-03	0.0234	580	Proteins, leucine (?)
8.945	4.62e-03	0.0241	56	Macromolecules/proteins
1.895	4.66e-03	0.0241	562	Proteins
9.225	4.67e-03	0.0241	28	Macromolecules/proteins
9.025	4.67e-03	0.0241	48	Macromolecules/proteins
1.095	4.76e-03	0.0243	642	Proteins
8.245	4.89e-03	0.0249	126	Macromolecules/proteins
8.215	4.96e-03	0.0250	129	Macromolecules/proteins
9.245	5.05e-03	0.0252	26	Macromolecules/proteins
0.865	5.12e-03	0.0252	665	Proteins
9.475	5.12e-03	0.0252	3	Macromolecules/proteins
9.335	5.14e-03	0.0252	17	Macromolecules/proteins
1.875	5.28e-03	0.0256	564	Proteins
1.765	5.33e-03	0.0256	575	Proteins
8.735	5.34e-03	0.0256	77	Macromolecules/proteins
8.435	5.41e-03	0.0256	107	Macromolecules/proteins
4.365	5.41e-03	0.0256	315	Proteins
9.465	5.47e-03	0.0258	4	Macromolecules/proteins
1.625	5.68e-03	0.0264	589	Proteins
8.985	5.69e-03	0.0264	52	Macromolecules/proteins
9.395	5.92e-03	0.0273	11	Macromolecules/proteins
4.255	5.95e-03	0.0273	326	Proteins, unknown
6.515	6.03e-03	0.0275	299	Macromolecules/proteins
8.475	6.10e-03	0.0275	103	Macromolecules/proteins
3.075	6.15e-03	0.0275	444	Proteins
9.495	6.15e-03	0.0275	1	Macromolecules/proteins
8.465	6.21e-03	0.0275	104	Macromolecules/proteins
8.205	6.34e-03	0.0278	130	Macromolecules/proteins
8.805	6.38e-03	0.0278	70	Macromolecules/proteins
4.135	6.40e-03	0.0278	338	Proteins, unknown
0.575	6.49e-03	0.0281	694	Proteins
1.575	6.58e-03	0.0283	594	Proteins, unknown

1.075	6.61e-03	0.0283	644	Proteins, unknown
0.755	6.69e-03	0.0284	676	Proteins
7.775	6.73e-03	0.0284	173	Macromolecules/proteins, unknown
8.295	6.77e-03	0.0284	121	Macromolecules/proteins
1.675	6.98e-03	0.0290	584	Proteins
8.155	6.99e-03	0.0290	135	Macromolecules/proteins
9.055	7.09e-03	0.0292	45	Macromolecules/proteins
9.445	7.25e-03	0.0297	6	Macromolecules/proteins
2.975	7.45e-03	0.0301	454	Proteins, unknown
9.035	7.46e-03	0.0301	47	Macromolecules/proteins
4.375	7.49e-03	0.0301	314	Proteins
1.565	7.54e-03	0.0301	595	Proteins, unknown
4.245	7.55e-03	0.0301	327	Proteins
9.485	7.68e-03	0.0303	2	Macromolecules/proteins
1.115	7.71e-03	0.0303	640	Proteins, unknown
0.905	7.80e-03	0.0306	661	Proteins, unknown
1.235	7.88e-03	0.0307	628	Proteins
7.885	8.19e-03	0.0317	162	Macromolecules/proteins
1.735	8.29e-03	0.0319	578	Proteins, leucine (?)
1.255	8.38e-03	0.0321	626	Proteins
1.275	8.51e-03	0.0324	624	Proteins, unknown
9.385	8.64e-03	0.0327	12	Macromolecules/proteins
4.235	8.81e-03	0.0332	328	Proteins
1.185	8.88e-03	0.0333	633	Proteins
1.265	9.00e-03	0.0335	625	Proteins, unknown
8.955	9.03e-03	0.0335	55	Macromolecules/proteins
1.805	9.16e-03	0.0336	571	Proteins
1.425	9.16e-03	0.0336	609	Proteins
8.575	9.63e-03	0.0352	93	Macromolecules/proteins
2.345	9.72e-03	0.0352	517	Proteins, unknown
6.595	9.77e-03	0.0352	291	Macromolecules/proteins, unknown
2.945	9.83e-03	0.0352	457	Proteins
9.325	9.88e-03	0.0352	18	Macromolecules/proteins
9.425	9.88e-03	0.0352	8	Macromolecules/proteins
8.445	9.97e-03	0.0352	106	Macromolecules/proteins
9.155	1.00e-02	0.0352	35	Macromolecules/proteins
2.245	1.01e-02	0.0355	527	Proteins, unknown
7.205	1.02e-02	0.0357	230	Proteins, unknown
8.705	1.03e-02	0.0357	80	Macromolecules/proteins
9.065	1.03e-02	0.0357	44	Macromolecules/proteins
0.745	1.07e-02	0.0365	677	Proteins
9.365	1.07e-02	0.0365	14	Macromolecules/proteins
6.895	1.08e-02	0.0365	261	Proteins, unknown
1.065	1.08e-02	0.0365	645	Proteins, unknown
1.905	1.09e-02	0.0365	561	Proteins, unknown

1.795	1.09e-02	0.0365	572	Proteins
6.505	1.12e-02	0.0373	300	Macromolecules/proteins
8.365	1.18e-02	0.0391	114	Macromolecules/proteins, unknown
6.525	1.18e-02	0.0391	298	Macromolecules/proteins, unknown
6.535	1.20e-02	0.0393	297	Macromolecules/proteins, unknown
6.585	1.22e-02	0.0398	292	Macromolecules/proteins
6.735	1.23e-02	0.0398	277	Macromolecules/proteins
8.815	1.23e-02	0.0398	69	Macromolecules/proteins
8.355	1.27e-02	0.0410	115	Macromolecules/proteins
6.935	1.30e-02	0.0419	257	Macromolecules/proteins
7.105	1.35e-02	0.0434	240	Macromolecules/proteins, unknown
8.925	1.37e-02	0.0435	58	Macromolecules/proteins
8.455	1.38e-02	0.0437	105	Macromolecules/proteins, unknown
6.875	1.39e-02	0.0439	263	Macromolecules/proteins
3.085	1.41e-02	0.0442	443	Proteins
6.825	1.41e-02	0.0442	268	Macromolecules/proteins
8.395	1.42e-02	0.0443	111	Macromolecules/proteins
1.535	1.43e-02	0.0443	598	Proteins, unknown
7.275	1.44e-02	0.0446	223	Unknown, macromolecules/proteins
2.075	1.45e-02	0.0447	544	Proteins, unknown
8.195	1.49e-02	0.0450	131	Macromolecules/proteins
2.985	1.49e-02	0.0450	453	Proteins
1.195	1.49e-02	0.0450	632	Proteins, unknown
3.095	1.49e-02	0.0450	442	Proteins
9.135	1.50e-02	0.0450	37	Unknown, macromolecules/proteins
2.315	1.52e-02	0.0454	520	Proteins, unknown
1.225	1.55e-02	0.0461	629	Proteins, unknown
1.775	1.56e-02	0.0463	574	Proteins
9.045	1.57e-02	0.0464	46	Macromolecules/proteins
8.715	1.58e-02	0.0465	79	Macromolecules/proteins
7.925	1.59e-02	0.0465	158	Macromolecules/proteins
7.095	1.59e-02	0.0465	241	Macromolecules/proteins, unknown
1.105	1.60e-02	0.0466	641	Proteins, unknown
1.435	1.62e-02	0.0469	608	Proteins, unknown
1.325	1.63e-02	0.0470	619	Proteins, unknown
6.865	1.63e-02	0.0470	264	Macromolecules/proteins, unknown
4.225	1.67e-02	0.0478	329	Proteins, unknown
1.555	1.70e-02	0.0485	596	Proteins, unknown
6.565	1.74e-02	0.0495	294	Macromolecules/proteins

Table 7.23: **Previous page: Spectral positions given in ppm, p-values both unadjusted and Benjamini and Hochberg (B/H)-adjusted, NMR IDs, as well as correspondingly identified compounds of NMR features that discriminated groups P and O of hypothesis 1b, i.e. patients progressing to ESRD without dying and patients not progressing to ESRD without dying.** The 205 urine specimens studied were collected in the last week directly before treatment randomization (W1). A false discovery rate (FDR) below 5% was applied. The FDR was adjusted according to the method of Benjamini and Hochberg (B/H). In case that more than one compound contributed to a significant bin, all possible assignments are given. A question mark denotes ambiguous signal assignments, mostly due to severe signal overlap.

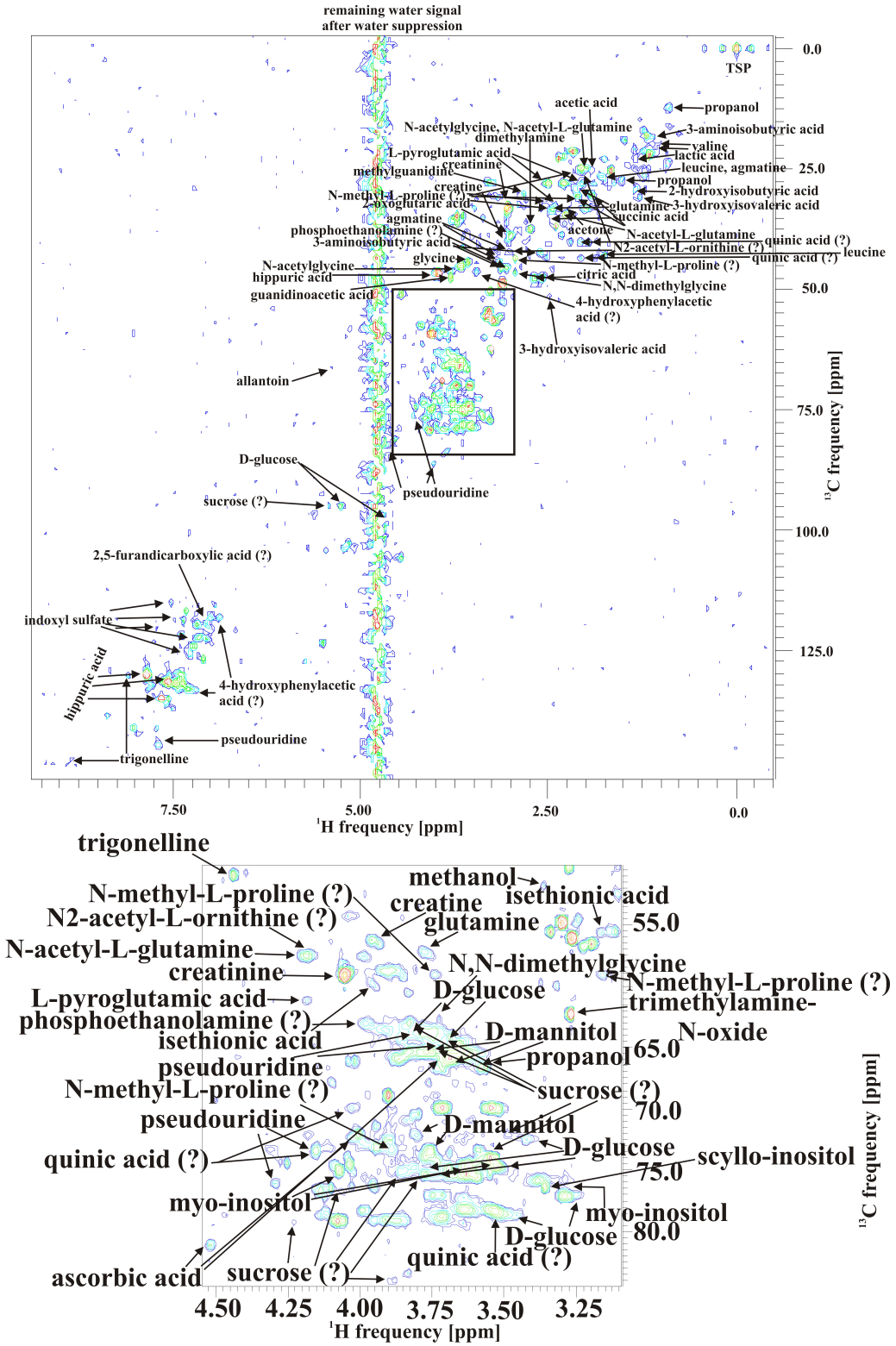


Figure 7.12: **A representative high-resolution 2D ^{1}H-^{13}C HSQC spectrum of a urine specimen from a study participant of the TREAT study.** The spectrum was recorded with 2048 × 512 data points using 44 scans per increment. The following compounds were identified supported by a corresponding 2D ^{1}H ^{1}H TOCSY spectrum, which was recorded with 2048 × 256 data points, 56 scans per increment: 2,5-furandicarboxylic acid (?), 2-hydroxyisobutyric acid, 2-oxoglutaric acid, 3-aminoisobutyric acid, 3-hydroxyisovaleric acid, 4-hydroxyphenylacetic acid (?), acetic acid, acetone, agmatine, allantoin, ascorbic acid, citric acid, creatine, creatinine, D-glucose, dimethylamine, D-mannitol (?), glutamine, glycine, guanidinoacetic acid, hippuric acid, indoxyl sulfate, isethionic acid, L-pyroglutamic acid, lactic acid, leucine, methanol, methylguanidine, myo-inositol, N2-acetyl-L-ornithine (?), N-acetylglycine, N-acetyl-L-glutamine, N-methyl-L-proline (?), N,N-dimethylglycine, phosphoethanolamine (?), propanol, pseudouridine, quinic acid (?), scyllo-inositol, succinic acid, sucrose (?), TSP, trigonelline, trimethylamine-N-oxide, valine. The crowded sugar region from 3.10 - 4.55 ppm in the ^{1}H direction and from 50.0 - 84.5 ppm in the ^{13}C direction has been enlarged and is displayed below. A question mark denotes ambiguous signal assignments, mostly due to signal overlap. Note that due to the fact that some signal intensities for few compounds, e.g. trigonelline, were below detection limit, not all peaks could be annotated.

8 About the author

8.1 Curriculum Vitae

Name	Helena Ursula Zacharias
Date of birth	January, 22nd 1988
Place of birth	Regensburg, Germany
Nationality	German

Education

2012 - present	Ph.D. student at the biology faculty of the University of Regensburg, Regensburg, Germany, Institute of Functional Genomics
2010 - 2012	studied physics (M.Sc.) at the University of Regensburg, Regensburg, Germany, master thesis *Investigation of Acute Kidney Injury after Cardiac Surgery by NMR Spectroscopy and Machine Learning Methods* at the Institute of Functional Genomics (Advisor: Prof. Dr. Elmar Lang)
2007 - 2010	studied physics (B.Sc.) at the University of Regensburg, Regensburg, Germany
1998 - 2007	attended Albrecht-Altdorfer-Gymnasium, Regensburg, Germany

8.2 Publications

Zacharias, H. U., *Investigation of Acute Kidney Injury after Cardiac Surgery by NMR Spectroscopy and Machine Learning Methods*, M.Sc. thesis in physics, Institute of Functional Genomics, University of Regensburg, March 2012.

Hochrein, J., Klein, M. S., Zacharias, H. U., Li, J., Wijffels, G., Schirra, H. J., Spang, R., Oefner, P. J., Gronwald, W., *Performance Evaluation of Algorithms for the Classification of Metabolic ^{1}H NMR Fingerprints*, Journal of Proteome Research, 2012, 11:6242-6251.

Zacharias, H. U., Schley, G., Hochrein, J., Klein, M. S., Köberle, C., Eckardt, K.-U., Willam, C., Oefner, P. J., Gronwald, W., *Analysis of human urine reveals metabolic changes related to the development of acute kidney injury following cardiac surgery*, Metabolomics, 2013, 9(3):697-707.

Zacharias, H. U., Hochrein, J., Klein, M. S., Samol, C., Oefner, P. J., Gronwald, W., *Current Experimental, Bioinformatic and Statistical Methods used in NMR Based Metabolomics*, Current Metabolomics, 2013, 1(3):253-268(16).

Zacharias, H. U., Hochrein, J., Vogl, F. C., Schley, G., Mayer, F., Jeleazcov, C., Eckardt, K.-U., Willam, C., Oefner, P. J., Gronwald, W., *Identification of Plasma Metabolites Prognostic of Acute Kidney Injury after Cardiac Surgery with Cardiopulmonary Bypass*, Journal of Proteome Research, 2015, 14(7):2897-2905.

Hochrein, J., Zacharias, H. U., Taruttis, F., Samol, C., Engelmann, J., Spang, R., Oefner, P. J., Gronwald, W., *Data Normalization of ^{1}H-NMR Metabolite Fingerprinting Datasets in the Presence of Unbalanced Metabolite Regulation*, Journal of Proteome Research, 2015, 14(8):3217-3228.

Schlecht, I., Gronwald, W., Behrens, G., Baumeister, S., Hertel, J., Hochrein, J., Zacharias, H. U., Fischer, B., Oefner, P. J., Leitzmann, M. F., *Visceral adipose tissue but not subcutaneous adipose tissue is associated with urine and serum metabolites*, 2016, *submitted to Int. J. of Obesity.*

8.3 Poster Presentations

FGMR Discussion Meeting & Joint Conference of the German, Italian and Slovenian Magnetic Resonance Societies, Frauenchiemsee, Germany, 2013, *NMR analysis of human biofluids reveals metabolic changes related to the development of acute kidney injury following cardiac surgery.*

Discussion Meeting of the GDCh-Division of "Magnetic Resonance", Berlin, Germany, 2014, *Prediction of Acute Kidney Injury after cardiac surgery with cardiopulmonary bypass use employing NMR urine and plasma fingerprints.*

Associate seminar of the Regensburg International Graduate School of Life Sciences - RIGeL, section Cellular Biochemistry and Biophysics, Kostenz, Germany, 2014, *Prediction of Acute Kidney Injury after cardiac surgery with cardiopulmonary bypass use employing NMR urine and plasma fingerprints.*

8.4 Conference Talks

Associate seminar of the Bavarian Genome Research Network (BayGene), Grosshadern/ Martinsried, Germany, 2013, *Analysis of human urine reveals metabolic changes related to the development of acute kidney injury following cardiac surgery.*

Associate seminar of the Regensburg International Graduate School of Life Sciences - RIGeL, section Cellular Biochemistry and Biophysics, Sulzbürg, Germany, 2013, *Analysis of human urine reveals metabolic changes related to the development of acute kidney injury following cardiac surgery.*

9 Bibliography

[Abdi and Williams 2010] Abdi, H., Williams, L. J., *Principal Component Analysis*, Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4):433-459.

[Agresti 1990] Agresti, A., *Categorical data analysis*, New York: Wiley, 1990, pp. 59-66.

[Agresti 2002] Agresti, A., *Categorical data analysis*, New York: Wiley, second edition, pp. 91-101.

[Al-Shabanah et al. 2010] Al-Shabanah, O. A., Aleisa, A. M., Al-Yahya, A. A., Al-Rejaie, S. S., Bakheet, S. A., Fatani, A. G., Sayed-Ahmed, M. M., *Increased urinary losses of carnitine and decreased intramitochondrial coenzyme A in gentamicin-induced acute renal failure in rats*, Nephrology Dialysis Transplantation, 2010, 25:69-76.

[Altenbuchinger et al. 2016] Altenbuchinger, M., Rehberg, T., Zacharias, H. U., Stämmler, F., Dettmer, K., Weber, D., Hiergeist, A., Gessner, A., Holler, E., Oefner, P. J., Spang, R., *Reference point insensitive molecular data analysis*, 2016, *in preparation.*

[Arasteh et al. 2009] Arasteh, K., Baenkler, H.-W., Bieber, C., Brandt, R., Chatterjee, T., Dill, T., Ditting, T., Eich, W., Ernst, S., Fritze, D., Füeßl, H. S., Goeckenjan, G., Hahn, J.-M., Hamm, C. W., Harenberg, J., Hengstmann, J. H., Herzog, W., Hofmann, T., Holzapfel, N., Huck, K., Kähler, J., Keller, M., Klingmüller, D., Köster, R., Kowol, S., Kuck, K.-H., Löwe, B., Matzdorff, A., Müller-Tasch, T., Nienaber, C. A., Nikendei, C., Pausch, J., Petzsch, M., Rösch, W., Sauer, N., Schlehofer, B., Schmidt, M., Schneider, H., Schuchert, A., Schwab, M., Schweikert, H.-U., Stern, H., Teschner, A., Träder, C., Usadel, K.-H., Veelken, R., Wahl, P., Zastrow, A., Ziegler, R., Zipfel, S., *Duale Reihe Innere Medizin*, Thieme Press, Stuttgart 2009 (2nd edition).

[Arduini et al. 2008] Arduini, A., Bonomini, M., Savica, V., Amato, A., Zammit, V., *Carnitine in metabolic disease: Potential for pharmacological intervention*, Pharmacology & Therapeutics, 2008, 120:149-156.

[Barker and Rayens 2003] Barker, M., Rayens, W., *Partial Least Squares for Discrimination*, J. Chemometrics, 2003, 17:166-173.

[Barton et al. 2010] Barton, R. H., Waterman, D., Bonner, F. W., Holmes, E., Clarke, R., Nicholson, J. K., Lindon, J. C., *The influence of EDTA and citrate anticoagulant addition to human plasma on information recovery from NMR-based metabolic profiling studies*, Mol.BioSyst., 2010, 6:215-224.

[Becker et al. 1988] Becker, R. A., Chambers, J. M., Wilks, A. R., *The New S Language*, Wadsworth & Brooks/Cole, 1988.

[Beckonert et al. 2007] Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C., Nicholson, J. K., *Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts*, Nature Protocols, 2007, 2(11):2692-703.

[Bedford et al. 2014] Bedford, M., Stevens, P. E., Wheeler, T. W. K., Farmer, C. K. T., *What is the real impact of acute kidney injury?*, BMC nephrology., 2014, 15:95.

[Beger et al. 2008] Beger, R. D., Holland, R. D., Sun, J., Schnackenberg, L. K., Moore, P. C., Dent, C. L., Devarajan, P., Portilla, D., *Metabonomics of acute kidney injury in children after cardiac surgery*, Pediatr. Nephrol., 2008, 23:977-984.

[Bellomo et al. 2004] Bellomo, R., Ronco, C., Kellum, J. A., Mehta, R. L., Palevsky, P., ADQI workgroup, *Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group*, Critical Care, 2004, 8:R204-R212.

[Benjamini and Hochberg 1995] Benjamini, Y., Hochberg, Y., *Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing*, Journal of the Royal Statistical Society. Series B (Methodological), 1995, 57(1):289-300.

[Berger and Braun 2004] Berger, S., Braun, S., *200 and More NMR Experiments*, Wiley-VCH, Weinheim, 2004.

[Bertram et al. 2011] Bertram, H. C., Yde, C. C., Zhang, X., Kristensen, N. B., *Effect of Dietary Nitrogen Content on the Urine Metabolite Profile of Dairy Cows Assessed by Nuclear Magnetic Resonance (NMR)-Based Metabolomics*, J. Agric. Food Chem., 2011, 59:12499-12505.

[Bland and Altman 1986] Bland, J. M., Altman, D. G., *Statistical methods for assessing agreement between two methods of clinical measurement*, Lancet, 1986, 1:307-10.

[Bland and Altman 1999] Bland, J. M., Altman, D. G., *Measuring agreement in method comparison studies*, Stat Methods Med Res., 1999, 8:135-60.

[Bloch 1946] Bloch, F., *Nuclear Induction*, Physical Review, 1946, 70(7-8):460-474.

[Bolstad et al. 2003] Bolstad, B. M., Irizarry, R. A., Astrand, M., Speed, T. P., *A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias*, Bioinformatics, 2003, 19:185-193.

[Bouatra et al. 2013] Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., Bjorndahl, T. C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z. T., Poelzer, J., Huynh, J., Yallou, F. S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C., Wishart, D. S., *The Human Urine Metabolome*, PLoS One, 2013, 8(9):e73076.

[Bragadottir et al. 2013] Bragadottir, G., Redfors, B., Ricksten, S.-E., *Assessing glomerular filtration rate (GFR) in critically ill patients with acute kidney injury - true GFR versus urinary creatinine clearance and estimating equations*, Critical Care, 2013, 17:R108.

[Breiman 2001] Breiman, L., *Random Forests*, Mach. Learn., 2001, 45(1):5-32.

[Breiman 2002] Breiman, L., *Manual on setting up, using, and understanding random forests v3.1*, 2002, `http://oz.berkely.edu/users/breiman/Using_random_forests_V3.1.pdf`.

[Bryan et al. 2008] Bryan, K., Brennan, L., Cunningham, P., *MetaFIND: a feature analysis tool for metabolomics data*, BMC Bioinformatics, 2008, 9:470.

[Burges 1998] Burges, C. J. C., *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, 1998, 2:121-167.

[Butler 2002] Butler, E., *AVANCE SGU Based Frequency Generation Beginners Guide*, Bruker BioSpin, Rheinstetten, 2002, version 001.

[Carr and Purcell 1954] Carr, H. Y., Purcell, E. M., *Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments*, Physical Review, 1954, 94(3):630-638.

[Carstensen et al. 2015] Carstensen, B., Gurrin, L., Ekstrom, C., Figurski, M., *MethComp: Functions for Analysis of Agreement in Method Comparison Studies*, 2015, *R* package version 1.22.2, `http://CRAN.R-project.org/package=MethComp`.

[Casella and Berger 2002] Casella, G., Berger, R. L., *Statistical Inference*, Duxbury Press, 2002, Pacific Grove, USA, 2nd edition.

[Cavanagh et al. 1996] Cavanagh, J., Fairbrother, W. J., Palmer, A. G. III, Skelton, N. J., *Protein NMR Spectroscopy*, Academic Press, San Diego, 1996.

[Chambers 1992] Chambers, J. M., *Linear models*, chapter 4 of Statistical Models in S, eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole, 1992.

[Champely 2015] Champely, S., *pwr: Basic Functions for Power Analysis*, 2015, *R* package version 1.1-3, `http://CRAN.R-project.org/package=pwr`.

[Chawla et al. 2014] Chawla, L. S., Eggers, P. W., Star, R. A., Kimmel, P. L., *Acute Kidney Injury and Chronic Kidney Disease as Interconnected Syndromes*, N. Engl. J. Med., 2014, 371:58-66.

[Ciba-Geigy 1983] Ciba-Geigy, *Wissenschaftliche Tabellen Geigy*, Basel: Ciba-Geigy, 1983.

[Clarkson et al. 1993] Clarkson, D. B., Fan, Y., Joe, H., *A Remark on Algorithm 643: FEXACT: An Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables*, ACM Transactions on Mathematical Software, 1993, 19:484-488.

[Concato 2013] Concato, J., *Study Design and "Evidence" in Patient-oriented Research*, Am. J. Respir. Crit. Care Med., 2013, 187(11):1167-1172.

[Cox 1972] Cox, D. R., *Regression Models and Life-Tables*, Journal of the Royal Statistical Society Series B, 1972, 34(2):187-220.

[Cravedi and Remuzzi 2013] Cravedi, P., Remuzzi, G., *Pathophysiology of proteinuria and its value as an outcome measure in chronic kidney disease*, British Journal of Clinical Pharmacology, 2013, 76(4):516-523.

[Cruz et al. 2009] Cruz, D. N., Ricci, Z., Ronco, C., *Clinical review: RIFLE and AKIN - time for reappraisal*, Critical Care, 2009, 13(3):211.

[Curhan 2005] Curhan, G., *Cystatin C: A Marker of Renal Function or Something More?*, Clinical Chemistry, 2005, 51(2):293-294.

[Dalgaard 2008] Dalgaard, P., *Introductory Statistics with R*, Springer Statistics and Computing, New York, USA, 2008.

[Dawiskiba et al. 2014] Dawiskiba, T., Deja, S., Mulak, A., Ząbek, A., Jawień, E., Pawełka, D., Banasik, M., Mastalerz-Migas, A., Balcerzak, W., Kaliszewski, K., Skóra, J., Barć, P., Korta, K., Pormańczuk, K., Szyber, P., Litarski, A., Młynarz, P., *Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases*, World Journal of Gastroenterology, 2014, 20(1):163-174.

[Deja et al. 2013] Deja, S., Dawiskiba, T., Balcerzak, W., Orczyk-Pawiłowicz, M., Głód, M., Pawełka, D., Młynarz, P., *Follicular Adenomas Exhibit a Unique Metabolic Profile.* ^{1}*H NMR Studies of Thyroid Lesions*, PLoS One, 2013, 8(12):e84637.

[Del Palacio et al. 2012] Del Palacio, M., Romero, S., Casado, J. L., *Proximal Tubular Renal Dysfunction or Damage in HIV-Infected Patients*, AIDS Rev., 2012, 14:179-87.

[Del Re 2013] Del Re, A. C., *compute.es: Compute Effect Sizes*, 2013, *R* package version 0.2-2, `http://cran.r-project.org/web/packages/compute.es`.

[Dettmer and Hammock 2004] Dettmer, K., Hammock, B. D., *Metabolomics - A New Exciting Field within the "omics" Sciences*, Environmental Health Perspectives, 2004, 112(7):A396-A397.

[Dettmer et al. 2013] Dettmer, K., Vogl, F. C., Ritter, A. P., Zhu, W., Nürnberger, N., Kreutz, M., Oefner, P. J., Gronwald, W., Gottfried, E., *Distinct metabolic differences between various human cancer and primary cells*, Electrophoresis, 2013, 34:2836-2847.

[Diao et al. 2010] Diao, L., Ekins, S., Polli, J. E., *Quantitative structure activity relationship for inhibition of human organic cation/carnitine transporter*, Molecular Pharmaceutics, 2010, 7:2120-2131.

[Dieterle et al. 2011] Dieterle, F., Riefke, B., Schlotterbeck, G., Ross, A., Senn, H., Amberg, A., *NMR and MS Methods for Metabonomics*, in *Drug Safety Evaluation: Methods and Protocols*, Gautier, J.-C. (ed.), Springer (Methods in Molecular Biology), 2011, pp. 385-415.

[Dimitriadou et al. 2011] Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., *e1071: Misc Functions for the Department of Statistics (e1071)*, TU Wien, 2011.

[Dörner 2013] Dörner, K., *Niere* in Dörner, K. (Ed.), *Klinische Chemie und Hämatologie*, Thieme Press, Stuttgart 2013 (8th edition).

[Drüeke et al. 2006] Drüeke, T. B., Locatelli, F., Clyne, N., Eckardt, K.-U., Macdougall, I.C., Tsakiris, D., Burger, H.-U., Scherhag, A., for the CREATE Investigators, *Normalization of Hemoglobin Level in Patients with Chronic Kidney Disease and Anemia*, N. Engl. J. Med., 2006, 355(20):2071-2084.

[Druml et al. 1994] Druml, W., Fischer, M., Liebisch, B., Lenz, K., Roth, E., *Elimination of amino acids in renal failure*, Am. J. Clin. Nutr., 1994, 60:418-423.

[DuBois 1916] DuBois, E., *A formula to estimate the approximate surface area if height and weight be known*, Arch. Int. Med. 1916, 17:863-871.

[Dudoit et al. 2002] Dudoit, S., Fridlyand, J., Speed, T. P., *Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data*, J. Am. Stat. Assoc., 2002, 97:77-87.

[Dzúrik et al. 2001] Dzúrik, R., Spustová, V., Krivosíková, Z., Gazdíková, K., *Hippurate participates in the correction of metabolic acidosis*, Kidney International Supplement, 2001, 78:S278-S281.

[Eckardt et al. 2012] Eckardt, K.-U., Bärthlein, B., Baid-Agrawal, S., Beck, A., Busch, M., Eitner, F., Ekici, A. B., Floege, J., Gefeller, O., Haller, H., Hilge, R., Hilgers, K. F., Kielstein, J. T., Krane, V., Köttgen, A., Kronenberg, F., Oefner, P., Prokosch, H.-U., Reis, A., Schmid, M., Schaeffner, E., Schultheiss, U. T., Seuchter, S. A., Sitter, T., Sommerer, C., Walz, G., Wanner, C., Wolf, G., Zeier, M., Titze, S., *The German Chronic Kidney Disease (GCKD) study: design and methods*, Nephrol Dial Transplant, 2012, 27:1454-1460.

[Eckardt et al. 2013] Eckardt, K.-U., Coresh, J., Devuyst, O., Johnson, R. J., Köttgen, A., Levey, A. S., Levin, A., *Evolving importance of kidney disease: from subspecialty to global health burden*, Lancet, 2013, 382:158-169.

[Efron et al. 2004] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., *Least Angle Regression*, The Annals of Statistics, 2004, 32(2):407-499.

[Eisen et al. 1998] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., *Cluster analysis and display of genome-wide expression patterns*, PNAS, 1998, 95:14863-14868.

[Elliott et al. 2015] Elliott, P., Posma, J. M., Chan, Q., Garcia-Perez, I., Wijeyesekera, A., Bictash, M., Ebbels, T. M. D., Ueshima, H., Zhao, L., van Horn, L., Daviglus, M., Stamler, J., Holmes, E., Nicholson, J. K., *Urinary metabolic signatures of human adiposity*, Science Translational Medicine, 2015, 7(285):285ra62.

[Emwas et al. 2014] Emwas, A.-H., Luchinat, C., Turano, P., Tenori, L., Roy, R., Salek, R. M., Ryan, D., Merzaban, J. S., Kaddurah-Daouk, R., Zeri, A. C., Gowda, G. A. N., Raftery, D., Wang, Y., Brennan, L., Wishart, D. S., *Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review*, Metabolomics, 2014, DOI: 10.1007/s11306-014-0746-7.

[Endre et al. 2011] Endre, Z. H., Pickering, J. W., Walker, R. J., Devarajan, P., Edelstein, C. L., Bonventre, J. V., Frampton, C. M., Bennett, M. R., Ma, Q., Sabbisetti, V. S., Vaidya, V. S., Walcher, A. M., Shaw, G. M., Henderson, S. J., Nejat, M., Schollum, J. B. W., George, P. M., *Improved performance of urinary biomarkers of acute kidney injury in the critically ill by stratification for injury duration and baseline renal function*, Kidney International, 2011, 79:1119-1130.

[Engoren et al. 2014] Engoren, M., Habib, R. H., Arslanian-Engoren, C., Kheterpal, S., Schwann, T. A., *The Effect of Acute Kidney Injury and Discharge Creatinine Level on Mortality Following Cardiac Surgery*, Critical Care Medicine, 2014, 42(9):2069-2074.

[Ernst et al. 1987] Ernst, R. R., Bodenhausen, G., Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford, 1987.

[Eriksson et al. 1974] Eriksson, O., Kjellman, H., Pilbrant, A., Schannong, M., *Pharmacokinetics of tranexamic acid after intravenous administration to normal volunteers*, European Journal of Clinical Pharmacology, 1974, 7:375-380.

[Fassett et al. 2011] Fassett, R. G., Venuthurupalli, S. K., Gobe, G. C., Coombes, J. S., Cooper, M. A., Hoy, W. E., *Biomarkers in chronic kidney disease: a review*, Kidney International, 2011, 80:806-821.

[Faul et al. 2007] Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences*, Behavior Research Methods, 2007, 39:175-191.

[Feldman et al. 2003] Feldman, H. I., Appel, L. J., Chertow, G. M., Cifelli, D., Cizman, B., Daugirdas, J., Fink, J. C., Franklin-Becker, E. D., Go, A. S., Hamm, L. L., He, J., Hostetter, T., Hsu, C.-Y., Jamerson, K., Joffe, M., Kusek, J. W., Landis, J. R., Lash, J. P., Miller, E. R., Mohler, E. R. III, Muntner, P., Ojo, A. O., Rahman, M., Townsend, R. R., Wright, J. T., and the Chronic Renal Insufficiency Cohort (CRIC) Study Investigators, *The Chronic Renal Insufficiency Cohort (CRIC) Study: Design and Methods*, J. Am. Soc. Nephrol., 2003, 14(suppl 2):S148-S153.

[Felsenfeld et al. 2013] Felsenfeld, A., Rodriguez, M., Levine, B., *New insights into regulation of calcium homeostasis*, Curr. Opin. Nephrol. Hypertens., 2013, 22:371-376.

[Fisher 1935] Fisher, R. A., *The logic of inductive inference*, Journal of the Royal Statistical Society Series A, 1935, 98(1):39-54.

[Fisher 1962] Fisher, R. A., *Confidence limits for a cross-product ratio*, Australian Journal of Statistics, 1962, 4(1):41.

[Fisher 1970] Fisher, R. A., *Statistical Methods for Research Workers*, Oliver & Boyd, 1970.

[Fisher and Wood 2007] Fisher, C. G., Wood, K. B., *Introduction to and Techniques of Evidence-Based Medicine*, SPINE, 2007, 32(19S):S66-S72.

[Friedman et al. 2010] Friedman, J., Hastie, T., Tibshirani, R., *Regularization Paths for Generalized Linear Models via Coordinate Descent*, Journal of Statistical Software, 2010, 33(1):1-22.

[Fukui et al. 2009] Fukui, Y., Kato, M., Inoue, Y., Matsubara, A., Itoh, K., *A metabonomic approach identifies human urinary phenylacetylglutamine as a novel marker of interstitial cystitis*, Journal of Chromatography B, 2009, 877:3806-3812.

[Ganapathy et al. 2000] Ganapathy, M. E., Huang, W., Rajan, D. P., Carter, A. L., Sugawara, M., Iseki, K., Leibach, F. H., Ganapathy, V., *Beta-lactam antibiotics as substrates for OCTN2, an organic cation/carnitine transporter*, Journal of Biological Chemistry, 2000, 275:1699-1707.

[Gautier et al. 2004] Gautier, L., Cope, L., Bolstad, B. M., Irizarry, R. A., *affy — analysis of Affymetrix GeneChip data at the probe level*, Bioinformatics, 2004, 20(3):307-315.

[Geng and Pang 1999] Geng, W., Pang, K. S., *Differences in excretion of hippurate, as a metabolite of benzoate and as an administered species, in the single-pass isolated perfused rat kidney explained*, Journal of Pharmacology and Experimental Therapeutics, 1999, 288:597-606.

[German National Cohort Consortium 2014] German National Cohort (GNC) Consortium, *The German National Cohort: aims, study design and organization*, Eur. J. Epidemiol., 2014, 29:371-382.

[Gillies et al. 2005] Gillies, M., Bellomo, R., Doolan, L., Buxton, B., *Bench-to-bedside review: Inotropic drug therapy after adult cardiac surgery – a systematic literature review*, Crit Care., 2005, 9:266-279.

[Gronwald et al. 2008] Gronwald, W., Klein, M. S., Kaspar, H., Fagerer, S. R., Nürnberger, N., Dettmer, K., Bertsch, T., Oefner, P. J., *Urinary Metabolite Quantification employing 2D NMR spectroscopy*, Analytical Chemistry, 2008, 80(23):9288-9297.

[Gronwald et al. 2011] Gronwald, W., Klein, M. S., Zeltner, R., Schulze, B.-D., Reinhold, S. W., Deutschmann, M., Immervoll, A.-K., Böger, C. A., Banas, B., Eckardt, K.-U., Oefner, P. J., *Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine*, Kidney International, 2011, 79:1244-1253.

[Haase et al. 2009] Haase, M., Bellomo, R., Devarajan, P., Schlattmann, P., Haase-Fielitz, A., *Accuracy of Neutrophil Gelatinase-Associated Lipocalin (NGAL) in Diagnosis and Prognosis in Acute Kidney Injury: A Systematic Review and Meta-analysis*, American Journal of Kidney Diseases, 2009, 54(6):1012-1024.

[Haase et al. 2010a)] Haase, M., Bellomo, R., Haase-Fielitz, A., *Neutrophil gelatinase-associated lipocalin*, Current Opinion in Critical Care, 2010, 16:526-632.

[Haase et al. 2010b)] Haase, M., Bellomo, R., Haase-Fielitz, A., *Novel Biomarkers, Oxidative Stress, and the Role of Labile Iron Toxicity in Cardiopulmonary Bypass-Associated Acute Kidney Injury*, Journal of the American College of Cardiology, 2010, 55(19):2024-2033.

[Haase-Fielitz et al. 2009] Haase-Fielitz, A., Bellomo, R., Devarajan, P., Story, D., Matalanis, G., Dragun, D., Haase, M., *Novel and conventional serum biomarkers predicting acute kidney injury in adult cardiac surgery - A prospective cohort study*, Crit. Care Med., 2009, 37(2):553-560.

[Haase-Fielitz et al. 2011] Haase-Fielitz, A., Mertens, P. R., Plaß, M., Kuppe, H., Hetzer, R., Westerman, M., Ostland, V., Prowle, J. R., Bellomo, R., Haase, M., *Urine hepcidin has additive value in ruling out cardiopulmonarybypass-associated acute kidney injury: An observational cohort study*, Critical Care, 2011, 15(4):R186.

[Han et al. 2002] Han, W. K., Bailly, V., Abichandani, R., Thadhani, R., Bonventre, J. V., *Kidney Injury Molecule-1 (KIM-1): a novel biomarker for human renal proximal tubule injury*, Kidney international, 2002, 62(1):237-244.

[Han et al. 2009] Han, W. K., Wagener, G., Zhu, Y., Wang, S., Lee, H. T., *Urinary biomarkers in the early detection of acute kidney injury after cardiac surgery*, Clinical Journal of the American Society of Nephrology, 2009, 4(5):873-882.

[Harvey and Everett 2004] Harvey, P. W., Everett, D. J., *Significance of the detection of esters of P-hydroxybenzoic acid (parabens) in human breast tumours*, Journal of Applied Toxicology, 2004, 24:1-4.

[Hastie et al. 2001] Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning*, Springer (Series in Statistics), New York 2001.

[Hastings et al. 2013] Hastings, J., Matos, P. de, Dekker, A., Ennis, M., Marsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C., *The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013*, Nucleic Acids Res., 2013, 41:D456-63.

[Held et al. 2012] Held, M., Bentink, S., Kostka, D., Lottaz, C., Scheid, S., Jaeger, J., Kohler, C., *compdiagTools: Toolbox for performing and illustrating microarray data analyses*, *R* package version 1.8.2, 2012.

[Hobson et al. 2009] Hobson, C. E., Yavas, S., Segal, M. S., Schold, J. D., Tribble, C. G., Layon, A. J., Bihorac, A., *Acute Kidney Injury Is Associated With Increased Long-Term Mortality After Cardiothoracic Surgery*, Circulation, 2009, 119:2444-2453.

[Hochrein 2011] Hochrein, J., *Klassifikatoren in der Metabolomik und Proteomik*, diploma thesis, Institute of Functional Genomics, University of Regensburg, April 2011.

[Hochrein et al. 2012] Hochrein, J., Klein, M. S., Zacharias, H. U., Li, J., Wijffels, G., Schirra, H. J., Spang, R., Oefner, P. J., Gronwald, W., *Performance Evaluation of Algorithms for the Classification of Metabolic ^{1}H NMR Fingerprints*, Journal of Proteome Research, 2012, 11:6242-6251.

[Hochrein et al. 2015] Hochrein, J., Zacharias, H. U., Taruttis, F., Samol, C., Engelmann, J., Spang, R., Oefner, P. J., Gronwald, W., *Data Normalization of ^{1}H-NMR Metabolite Fingerprinting Datasets in the Presence of Unbalanced Metabolite Regulation*, Journal of Proteome Research, 2015, 14(8):3217-3228.

[Hochrein 2016] Hochrein, J., *Ph.D. thesis in preparation*, Ph.D. thesis, Fakultät für Biologie und Vorklinische Medizin, University of Regensburg, 2016.

[Hoerl and Kennard 1970] Hoerl, A. E., Kennard, R. W., *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 1970, 12(1):55-67.

[Holmes et al. 1997] Holmes, E., Foxall, P. J. D., Spraul, M., Farrant, R. D., Nicholson, J. K., Lindon, J. C., *750 MHz ^{1}H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease*, Journal of Pharmaceutical and Biomedical Analysis, 1997, 15:1647-1659.

[Homans 1995] Homans, S. W., *A Dictionary of Concepts in NMR*, Oxford Science Publications, Oxford 1995.

[Huber et al. 2002] Huber, W., Heydebreck, A. V., Sültmann, H., Poustka, A., Vingron, M., *Variance Stabilisation Applied to Microarray Data Calibration and to the Quantification of Differential Expression*, Bioinformatics, 2002, 18:S96-S104.

[Hummel et al. 2006] Hummel, M., Bentink, S., Berger, H., Klapper, W., Wessendorf, S., Barth, T. F., Bernd, H.-W., Cogliatti, S. B., Dierlamm, J., Feller, A. C., Hansmann, M.-L., Haralambieva, E., Harder, L., Hasenclever, D., Kühn, M., Lenze, D., Lichter, P., Martin-Subero, J. I., Möller, P., Müller-Hermelink, H.-K., Ott, G., Parwaresch, R. M., Pott, C., Rosenwald, A., Rosolowski, M., Schwaenen, C., Stürzenhofecker, B., Szczepanowski, M., Trautmann, H., Wacker, H.-H., Spang, R., Loeffler, M., Trümper, L., Stein, H., Siebert, R., *A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling*, N. Engl. J. Med., 2006, 354:2419-2430.

[Husson and Josse 2013] Husson, F., Josse, J., *missMDA: Handling missing values with/in multivariate data analysis (principal component methods)*, 2013, `http://CRAN.R-project.org/package=missMDA`.

[Ideker et al. 2001] Ideker, T., Galitski, T., Hood, L., *A New Approach to Decoding Life: Systems Biology*, Annu. Rev. Genomics Hum. Genet., 2001, 2:343-72.

[Idrovo et al. 2012] Idrovo, J.-P., Yang, W.-L., Matsuda, A., Nicastro, J., Coppa, G. F., Wang, P., *Post-treatment with the combination of 5-aminoimidazole-4-carboxyamide ribonucleoside and carnitine improves renal function after ischemia/reperfusion injury*, Shock, 2012, 37:39-46.

[Imai et al. 2010] Imai, E., Matsuo, S., Makino, H., Watanabe, T., Akizawa, T., Nitta, K., Iimuro, S., Ohashi, Y., Hishida, A., *Chronic Kidney Disease Japan Cohort study: baseline characteristics and factors associated with causative diseases and renal function*, Clin. Exp. Nephrol., 2010, 14:558-570.

[Inker et al. 2012] Inker, L. A., Schmid, C. H., Tighiouart, H., Eckfeldt, J. H., Feldman, H. I., Greene, T., Kusek, J. W., Manzi, J., Van Lente, F., Zhang, Y. L., Coresh, J., Levey, A., S., *Estimating Glomerular Filtration Rate from Serum Creatinine and Cystatin C*, New England Journal of Medicine, 2012, 367(1):20-29.

[Iseki et al. 2003] Iseki, K., Ikemiya, Y., Iseki, C., Takishita, S., *Proteinuria and the risk of developing end-stage renal disease*, Kidney International, 2003, 63:1468-1474.

[Itoh et al. 2012] Itoh, Y., Ezawa, A., Kikuchi, K., Tsuruta, Y., Niwa, T., *Protein-bound uremic toxins in hemodialysis patients measured by liquid chromatography/tandem mass spectrometry and their effects on endothelial ROS production*, Analytical and Bioanalytical Chemistry, 2012, 403:1841-1850.

[Jager et al. 2007] Jager, K. J., Stel, V. S., Wanner, C., Zoccali, C., Dekker, F. W., *The valuable contribution of observational studies to nephrology*, Kidney International, 2007, 72:671-675.

[James et al. 1972] James, M. O., Smith, R. L., Williams, R. T., Reidenberg, M., *The conjugation of phenylacetic acid in man, sub-human primates and some non-primate species*, Proceedings of the Royal Society of London B: Biological Sciences, 1972, 182:25-35.

[James et al. 2013] James, G., Witten, D., Hastie, T., Tibshirani, R., *An Introduction to Statistical Learning*, Springer, New York 2013.

[Jha et al. 2013] Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., Saran, R., Wang, A. Y.-M., Yang, C.-W., *Chronic kidney disease: global dimension and perspectives*, Lancet, 2013, 382:260-272.

[Joyce and Palsson 2006] Joyce, A. R., Palsson, B. Ø., *The model organism as a system: integrating 'omics' data sets*, Nature Reviews | Molecular Cell Biology, 2006, 7:198-210.

[Kang et al. 2011] Kang, S. M., Park, J. C., Shin, M. J., Lee, H., Oh, J., Hwang, G. S., Chung, J. H., ^{1}*H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure*, Clinical Biochemistry, 2011, 44(4):293-299.

[KDIGO workgroup 2012] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group, *KDIGO Clinical Practice Guideline for Acute Kidney Injury*, Kidney International, Suppl. 2012, 2:1-138.

[KDIGO workgroup 2013] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, *KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease*, Kidney International, Suppl. 2013, 3:1-150.

[Keane et al. 2003] Keane, W. F., Brenner, B. M., de Zeeuw, D., Grunfeld, J.-P., McGill, J., Mitch, W. E., Ribeiro, A. B., Shahinfar, S., Simpson, R. L., Snapinn, S. M., Toto, R., for the RENAAL Study Investigators, *The risk of developing end-stage renal disease in patients with type 2 diabetes and nephropathy: The RENAAL Study*, Kidney International, 2003, 63:1499-1507.

[Kidher et al. 2014] Kidher, E., Harling, L., Ashrafian, H., Naase, H., Chukwuemeka, A., Anderson, J., Francis, D. P., Athanasiou, T., *Pulse wave velocity and neutrophil gelatinase-associated lipocalin as predictors of acute kidney injury following aortic valve replacement*, J Cardiothorac Surg, 2014, 9(1):89.

[Klein et al. 2010] Klein, M. S., Almstetter, M. F., Schlamberger, G., Nürnberger, N., Dettmer, K., Oefner, P. J., Meyer, H. H. D., Wiedemann, S., Gronwald, W., *Nuclear magnetic resonance and mass spectrometry-based milk metabolomics in dairy cows during early and late lactation*, J Dairy Sci, 2010, 93:1539-1550.

[Klein 2011] Klein, M. S., *Analysis of Metabolic Disorders of Dairy Cows employing multidimensional and multinuclear NMR Spectroscopy*, Ph.D. thesis, Fakultät für Biologie und Vorklinische Medizin, University of Regensburg, 2011.

[Klein et al. 2011] Klein, M. S., Dorn, C., Saugspier, M., Hellerbrand, C., Oefner, P. J., Gronwald, W., *Discrimination of steatosis and NASH in mice using nuclear magnetic resonance spectroscopy*, Metabolomics, 2011, 7(2):237-246.

[Klein et al. 2012] Klein, M. S., Buttchereit, N., Miemczyk, S., Immervoll, A.-K., Louis, C., Wiedemann, S., Junge, W., Thaller, G., Oefner, P. J., Gronwald, W., *NMR Metabolic Analysis of Dairy Cows Reveals Milk Glycerophosphocholine to Phosphocholine Ratio as Prognostic Biomarker for Risk of Ketosis*, J Prot Res., 2012, 11(2):1373-1381.

[Klein et al. 2013] Klein, M. S., Oefner, P. J., Gronwald, W., *MetaboQuant: a tool combining individual peak calibration and outlier detection for accurate metabolite quantification in 1D ^{1}H and ^{1}H-^{13}C HSQC NMR spectra*, BioTechniques, 2013, 54:251-256.

[Kohl et al. 2012] Kohl, S. M., Klein, M. S., Hochrein, J., Oefner, P. J., Spang, R., Gronwald, W., *State-of-the art data normalization methods improve NMR-based metabolomic analysis*, Metabolomics, 2012, 8(1):146-160.

[Kosmides et al. 2013] Kosmides, A. K., Kamisoglu, K., Calvano, S. E., Corbett, S. A., Androulakis, O. P., *Metabolomic Fingerprinting: Challenges and Opportunities*, Crit Rev Biomed Eng., 2013, 41(3):205-221.

[Kuhlmann et al. 2003] Kuhlmann, U., Walb, D., Luft, F. D., (Ed.), *Nephrologie*, Thieme Publisher, Stuttgart 2003, 4th edition.

[Kumar et al. 1980] Kumar, A., Ernst, R. R., Wüthrich, K., *A two-dimensional Nuclear Overhauser Enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules*, Biochemical and Biophysical Research Communications, 1980, 95(1):1-6.

[Lameire et al. 2011] Lameire, N. H., Vanholder, R. C., Van Biesen, W. A., *How to use biomarkers efficiently in acute kidney injury*, Kidney International, 2011, 79:1047-1050.

[Lameire et al. 2013] Lameire, N. H., Bagga, A., Cruz, D., De Maeseneer, J., Endre, Z., Kellum, J. A., Liu, K. D., Mehta, R. L., Pannu, N., Van Biesen, W., Vanholder, R., *Acute kidney injury: an increasing global concern*, Lancet, 2013, 382:170-179.

[Lassnigg et al. 2004] Lassnigg, A., Schmidlin, D., Mouhieddine, M., Bachmann, L. M., Druml, W., Bauer, P., Hiesmayr, M., *Minimal changes of serum creatinine predict prognosis in patients after cardiothoracic surgery: a prospective cohort study*, J Am Soc Nephrol., 2004, 15:1597-1605.

[Levey et al. 1999] Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N., Roth, D., for MDRD Study Group, *A More Accurate Method to Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation*, Ann. Intern. Med., 1999, 130(6):461-470.

[Levey et al. 2009] Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., Coresh, J., for the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), *A New Equation to Estimate Glomerular Filtration Rate*, Ann. Intern. Med., 2009, 150(9):604-612.

[Lewis et al. 2011] Lewis, E. F., Pfeffer, M. A., Feng, A., Uno, H., McMurray, J. J. V., Toto, R., Gandra, S. R., Solomon, S. D., Moustafa, M., Macdougall, I. C., Locatelli, F., Parfrey, P. S. for the TREAT Investigators, *Darbepoetin Alfa Impact on Health Status in Diabetes Patients with Kidney Disease: A Randomized Trial*, Clin. J. Am. Soc. Nephrol. 2011, 6:845-855.

[Liaw and Wiener 2002] Liaw, A., Wiener, M., *Classification and Regression by randomForest*, R. News, 2002, 2(3):18-22.

[Linstrom and Mallard 2016] P.J. Linstrom and W.G. Mallard (eds.), *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg MD, 20899, `http://webbook.nist.gov`, (retrieved March 8, 2016).

[Liu et al. 2012] Liu, Y., Yan, S., Ji, C., Dai, W., Hu, W., Zhang, W., Mei, C., *Metabolomic changes and protective effect of (L)-carnitine in rat kidney ischemia/reperfusion injury*, Kidney and Blood Pressure Research, 2012, 35:373-381.

[Livingston 2004] Livingston, E. H., *Who Was Student and Why Do We Care So Much about His* t*-Test?*, Journal of Surgical Research, 2004, 118:58-65.

[Livingston and Cassidy 2005] Livingston, E. H., Cassidy, L., *Statistical Power and Estimation of the Number of Required Subjects for a Study Based on the* t*-Test: A Surgeon's Primer*, Journal of Surgical Research, 2005, 126:149-159.

[Lohninger et al. 2005] Lohninger, A., Pittner, G., Pittner, F., *L-Carnitine: New aspects of a known compound-a brief survey*, Monatshefte Chemie, 2005, 136:1255-1268.

[Lottaz et al. 2008] Lottaz, C., Kostka, D., Markowetz, F., Spang, R., *Computational Diagnostics with Gene Expression Profiles*, in Keith, Jonathan M. (ed.), *Bioinformatics: Structure, Function and Applications*, Humana Press, Totowa, NJ, USA, 2008, pp. 281-296.

[Macedo et al. 2011] Macedo, E., Malhotra, R., Claure-Del Granado, R., Fedullo, P., Mehta, R. L., *Defining urine output criterion for acute kidney injury in critically ill patients*, Nephrol. Dial. Transplant., 2011, 26:509-515.

[Macedo and Mehta 2013] Macedo, E., Mehta, R. L., *Measuring renal function in critically ill patients: tools and strategies for assessing glomerular filtration rate*, Curr. Opin. Crit. Care, 2013, 19:560-566.

[Maher et al. 2007] Maher, A. D., Zirak, S. F. M., Holmes, E., Nicholson, J. K., *Experimental and Analytical Variation in Human Urine in ^{1}H NMR Spectroscopy-Based Metabolic Phenotyping Studies*, Anal. Chem., 2007, 79:5204-5211.

[Mardia et al. 1979] Mardia, K. V., Kent, J. T., Bibby, J. M., *Multivariate Analysis*, Academic Press, London 1979.

[Mariscalco et al. 2011] Mariscalco, G., Lorusso, R., Dominici, C., Renzulli, A., Sala, A., *Acute Kidney Injury: A Relevant Complication after Cardiac Surgery*, Society of Thoracic Surgeons, 2011, 92(4):1539-1547.

[McKay 2011] McKay, R. T. *How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution*, Concepts Magn. Reson., 2011, 38A:197-220.

[McMurray et al. 2011] McMurray, J. J. V., Uno, H., Jarolim, P., Desai, A. S., de Zeeuw, D., Eckardt, K.-U., Ivanovich, P., Levey, A. S., Lewis, E. F., McGill, J. B., Parfrey, P., Parving, H.-H., Toto, R. M., Solomon, S. D., Pfeffer, M. A., *Predictors of fatal and nonfatal cardiovascular events in patients with type 2 diabetes mellitus, chronic kidney disease, and anemia: An analysis of the Trial to Reduce cardiovascular Events with Aranesp (darbepoetin-alfa) Therapy (TREAT)*, Am. Heart J., 2011, 162:748-755.e3.

[Mehta and Patel 1986] Mehta, C. R., Patel, N. R., *Algorithm 643. FEXACT: A Fortran subroutine for Fisher's exact test on unordered r*c contingency tables*, ACM Transactions on Mathematical Software, 1986, 12:154-161.

[Mehta et al. 2007] Mehta, R. L., Kellum, J. A., Shah, S. V., Molitoris, B. A., Ronco, C., Warnock, D. G., Levin, A., Acute Kidney Injury Network, *Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury*, Critical Care, 2007, 11(2):43-51.

[Meiboom and Gill 1958] Meiboom, S., Gill, D., *Modified Spin-Echo Method for Measuring Nuclear Relaxation Times*, The Review of Scientific Instruments, 1958, 29(8):688-691.

[Menze et al. 2009] Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F. A., *A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data*, BMC Bioinformatics, 2009, 10:213.

[Merziger et al. 2004] Merziger, G., Mühlbach, G., Wille, D., Wirth, T., *Formeln + Hilfen zur höheren Mathematik*, Binomi Press, Hannover 2004.

[Mix et al. 2005] Mix, T.-C. H., Brenner, R. M., Cooper, M. E., de Zeeuw, D., Ivanovich, P., Levey, A. S., McGill, J. B., McMurray, J. J. V., Parfrey, P. S., Parving, H.-H., Pereira, B. J. G., Remuzzi, G., Singh, A. K., Solomon, S. D., Stehman-Breen, C., Toto, R. D., Pfeffer, M. A., *Rationale - Trial to Reduce Cardiovascular Events with Aranesp Therapy (TREAT): Evolving the management of cardiovascular risk in patients with chronic kidney disease*, Am. Heart J., 2005, 149:408-413.

[Motulsky 1995] Motulsky, H., *Intuitive Biostatistics*, Oxford University Press, New York, Oxford, 1995.

[Mukherjee et al. 2003] Mukherjee, S., Golland, P., Panchenko, P., *Permutation tests for classification*, Cambridge: Massachusetts Institute of Technology, 2003, Al-Memo, 2003-19.

[Musso et al. 2009] Musso, C. G., Michelángelo, H., Vilas, M, Reynaldi, J., Martinez, B., Algranati, L., Núñez, J. F. M., *Creatinine reabsorption by the aged kidney*, Int. Urol. Nephrol., 2009, 41(3):727-731.

[Neild et al. 1997] Neild, G. H., Foxall, P. J. D., Lindon, J. C., Holmes, E. C., Nicholson, J. K., *Uroscopy in the 21st Century: high-field NMR spectroscopy*, Nephrol. Dial. Transplant., 1997, 12:404-417.

[Nicholson et al. 1983] Nicholson, J. K., Buckingham, M. J., Sadler, P. J., *High resolution 1H n.m.r. studies of vertebrate blood and plasma*, Biochem. J., 1983, 211:605-615.

[Nicholson 2006] Nicholson, J. K., *Global systems biology, personalized medicine and molecular epidemiology*, Molecular Systems Biology, 2006, 2:52, doi:10.1038/msb4100095.

[Nicholson and Lindon 2008] Nicholson, J. K., Lindon, J. C., *Metabonomics*, Nature, 2008, 455(23):1054-1056.

[Ostermann 2014] Ostermann, M., *Diagnosis of acute kidney injury: Kidney Disease Improving Global Outcomes criteria and beyond*, Curr. Opin. Crit. Care, 2014, 20:581-587.

[O'Toole and Sedor 2014] O'Toole, J. F., Sedor, J. R., *Kidney disease: new technologies translate mechanisms to cure*, J. Clin. Invest., 2014, 124(6):2294-2298.

[Palmer et al. 2011] Palmer, S. C., Sciancalepore, M., Strippoli, G. F. M., *Trial Quality in Nephrology: How Are We Measuring Up?*, AM. J. Kidney Dis., 2011, 58(3):335-337.

[Parikh et al. 2005] Parikh, C. R., Abraham, E., Ancukiewicz, M., Edelstein, C. L., *Urine IL-18 is an early diagnostic marker for acute kidney injury and predicts mortality in the intensive care unit*, Journal of the American Society of Nephrology, 2005, 16(10):3046-3052.

[Parikh et al. 2011] Parikh, C. R., Devarajan, P., Zappitelli, M., Sint, K., Thiessen-Philbrook, H., Li, S., Kim, R. W., Koyner, J. L., Coca, S. G., Edelstein, C. L., Shlipak, M. G., Garg, A. X., Krawczeski, C. D., *Postoperative Biomarkers Predict Acute Kidney Injury and Poor Outcomes after Pediatric Cardiac Surgery*, Journal of the American Society of Nephrology, 2011, 22:1737-1747.

[Patefield 1981] Patefield, W. M., *Algorithm AS159. An efficient method of generating r × c tables with given row and column totals*, Applied Statistics, 1981, 30:91-97.

[Pfeffer et al. 2009a)] Pfeffer, M. A., Burdmann, E. A., Chen, C.-Y., Cooper, M. E., de Zeeuw, D., Eckardt, K.-U., Ivanovich, P., Kewalramani, R., Levey, A. S., Lewis, E. F., McGill, J., McMurray, J. J. V., Parfrey, P., Parving, H.-H., Remuzzi, G., Singh, A. K., Solomon, S. D., Toto, R., Uno, H., on behalf of the TREAT investigators, *Baseline Characteristics in the Trial to Reduce Cardiovascular Events With Aranesp Therapy (TREAT)*, American Journal of Kidney Diseases, 2009, 54(1):59-69.

[Pfeffer et al. 2009b)] Pfeffer, M. A., Burdmann, E. A., Chen, C.-Y., Cooper, M. E., de Zeeuw, D., Eckardt, K.-U., Feyzi, J. M., Ivanovich, P., Kewalramani, R., Levey, A. S., Lewis, E. F., McGill, J. B., McMurray, J. J. V., Parfrey, P., Parving, H.-H., Remuzzi, G., Singh, A. K., Solomon, S. D., Toto, R., on behalf of the TREAT investigators, *A Trial of Darbepoetin Alfa in Type 2 Diabetes and Chronic Kidney Disease*, New England Journal of Medicine, 2009, 361(21):2019-2032.

[Phrommintikul et al. 2007] Phrommintikul, A., Haas, S. J., Elsik, M., Krum, H., *Mortality and target haemoglobin concentrations in anaemic patients with chronic kidney disease treated with erythropoietin: a meta-analysis*, Lancet, 2007, 369(9559):381-388.

[Pollard et al. 2005] Pollard, K. S., Dudoit, S., van der Laan, M. J., *Multiple Testing Procedures: R multtest Package and Applications to Genomics*, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Gentleman, R., Carey, V., Huber, W., Irizarry, R., Dudoit, S. (ed.), Springer (Statistics for Biology and Health Series), 2005, pp. 251-272.

[Psychogios et al. 2011] Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T., Wishart, D. S., *The Human Serum Metabolome*, PLoS ONE, 2011, 6(2):e16957.

[Purnell et al. 2013] Purnell, T. S., Auguste, P., Crews, D. C., Lamprea-Montealegre, J., Olufade, T., Greer, R., Ephraim, P., Sheu, J., Kostecki, D., Powe, N. R., Rabb, H., Jaar, B., Boulware, L. E., *Comparison of Life Participation Activities Among Adults Treated by Hemodialysis, Peritoneal Dialysis, and Kidney Transplantation: A Systematic Review*, Am J Kidney Dis, 2013, 62(5):953-973.

[Rao and Pereira 2003] Rao, M., Pereira, B. J. G., *Prospective trials on anemia of chronic disease: The Trial to Reduce Cardiovascular Events with Aranesp Therapy (TREAT)*, Kidney International, 2003, 64(Supplement 87):S12-S19.

[R Core Team 2014] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014, `http://www.R-project.org`.

[Rosner and Okusa 2006] Rosner, M. H., Okusa, M. D., *Acute Kidney Injury Associated with Cardiac Surgery*, Clin. J. Am. Soc. Nephrol., 2006, 1:19-32.

[Ross et al. 2007] Ross, A., Schlotterbeck, G., Dieterle, F., Senn, H., *NMR Spectroscopy Techniques for Application to Metabonomics*, in Lindon, J. C., Nicholson, J. K., Holmes, E., (eds.),

NMR Spectroscopy Techniques for Application to Metabonomics, Elsevier BV: Amsterdam, The Netherlands, 2007, pp. 55-112.

[Schlecht et al. 2016] Schlecht, I., Gronwald, W., Behrens, G., Baumeister, S., Hertel, J., Hochrein, J., Zacharias, H. U., Fischer, B., Oefner, P. J., Leitzmann, M. F., *Visceral adipose tissue but not subcutaneous adipose tissue is associated with urine and serum metabolites*, 2016, *submitted to Int. J. of Obesity.*

[Selma et al. 2009] Selma, M. V., Espín, J. C., Tomás-Barberán, F. A., *Interaction between phenolics and gut microbiota: Role in human health*, Journal of Agricultural and Food Chemistry, 2009, 57:6485-6501.

[Shapiro and Wilk 1965] Shapiro, S. S., Wilk, M. B., *An Analysis of Variance Test for Normality (Complete Samples)*, Biometrika, 1965, 52(3/4):591-611.

[Shaw 2012] Shaw, A., *Update on acute kidney injury after cardiac surgery*, The Journal of Thoracic and Cardiovascular Surgery, 2012, 143:676-681.

[Siew and Davenport 2015] Siew, E. D., Davenport, A., *The growth of acute kidney injury: a rising tide or just closer attention to detail?*, Kidney International, 2015, 87:46-61.

[Sing et al. 2005] Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., *ROCR: visualizing classifier performance in R*, Bioinformatics, 2005, 21(20):7881.

[Singh et al. 2006] Singh, A. K., Szczech, L., Tang, K. L., Barnhart, H., Sapp, S., Wolfson, M., Donal Reddan, D., for the CHOIR Investigators, *Correction of Anemia with Epoetin Alfa in Chronic Kidney Disease*, N. Engl. J. Med., 2006, 355:2085-2098.

[Singh 2010] Singh, A. K., *Does TREAT Give the Boots to ESAs in the Treatment of CKD Anemia?*, J. Am. Soc. Nephrol., 2010, 21:2-6.

[Skali et al. 2011] Skali, H., Parving, H.-H., Parfrey, P. S., Burdmann, E. A., Lewis, E. F., Ivanovich, P., Keithi-Reddy, S. R., McGill, J. B., McMurray, J. J. V., Singh, A. K., Solomon, S. D., Uno, H., Pfeffer, M. A., on behalf of the TREAT Investigators, *Stroke in Patients With Type 2 Diabetes Mellitus, Chronic Kidney Disease, and Anemia Treated With Darbepoetin Alfa*, Circulation, 2011, 124:2903-2908.

[Skali et al. 2013] Skali, H., Lin, J., Pfeffer, M. A., Chen, C.-Y., Cooper, M. E., McMurray, J. J. V., Nissenson, A. R., Remuzzi, G., Rossert, J., Parfrey, P. S., Scott-Douglas, N. W., Singh, A. K., Toto, R., Uno, H., Ivanovich, P., *Hemoglobin Stability in Patients With Anemia, CKD, and Type 2 Diabetes: An Analysis of the TREAT (Trial to Reduce Cardiovascular Events with Aranesp Therapy) Placebo Arm*, Am. J. Kidney Dis., 2013, 61(2):238-46.

[Smith et al. 2005] Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., *METLIN: a metabolite mass spectral database*, Ther. Drug. Monit., 2005, 27:747-751.

[Smyth 2005] Smyth, G. K., *Limma: Linear models for microarray data*, in Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., Huber, W. (eds) *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Springer, New York, 2005, pp. 397-420.

[Snoeijs et al. 2011] Snoeijs, M. G. J., Vaahtera, L., de Vries, E. E., Schurink, G. W. H, Haenen, G. R. M. M., Peutz-Kootstra, C. J., Buurman, W. A., van Heurn, L. W. E., Parkkinen, J., *Addition of a water-soluble propofol formulation to preservation solution in experimental kidney transplantation*, Transplantation, 2011, 92:296-302.

[Solomon 1955] Solomon, I., *Relaxation processes in a system of two spins*, Physical Review, 1955, 99(2): 559.

[Solomon et al. 2010] Solomon, S. D., Uno, H., Lewis, E. F., Eckardt, K.-U., Lin, J., Burdmann, E. A., de Zeeuw, D., Ivanovich, P., Levey, A. S., Parfrey, P., Remuzzi, G., Singh, A. K., Toto, R., Huang, F., Rossert, J., McMurray, J. J., Pfeffer, M. A., for the Trial to Reduce Cardiovascular Events with Aranesp Therapy (TREAT) Investigators, *Erythropoietic Response and Outcomes in Kidney Disease and Type 2 Diabetes*, N. Engl. J. Med., 2010, 363(12):1146-55.

[Somashekar et al. 2006] Somashekar, B. S., Ijare, O. B., Nagana Gowda, G. A., Ramesh, V., Gupta, S., Khetrapal, C. L., *Simple pulse-acquire NMR methods for the quantitative analysis of calcium, magnesium and sodium in human serum*, Spectrochim. Acta A Mol. Biomol. Spectrosc., 2006, 65:254-260.

[Star 1998] Star, R. A., *Treatment of acute renal failure*, Kidney International, 1998, 54:1817-1831.

[Stevens et al. 2006] Stevens, L. A., Coresh, J., Greene, T., Levey, A. S., *Assessing Kidney Function - Measured and Estimated Glomerular Filtration Rate*, New England Journal of Medicine, 2006, 354(23):2473-2483.

[Stevens and Levey 2009] Stevens, L. A., Levey, A. S., *Measured GFR as a Confirmatory Test for Estimated GFR*, J. Am. Soc. Nephrol., 2009, 20:2305-2313.

[Strippoli et al. 2007] Strippoli, G. F. M., Tognoni, G., Navaneethan, S. D., Nicolucci, A., Craig, J. C., *Haemoglobin targets: we were wrong, time to move on*, Lancet, 2007, 369(9559):346-50.

[Sun et al. 2012] Sun, J., Shannon, M., Ando, Y., Schnackenberg, L. K., Khan, N. A., Portilla, D., Beger, R. D., *Serum Metabolomic Profiles from Patients with Acute Kidney Injury: A Pilot Study*, J.Chromatogr.B., 2012, 893:107-113.

[Thiese 2014] Thiese, M. S., *Observational and interventional study design types; an overview*, Biochemia Medica, 2014, 24(2):199-210.

[Tibshirani 1996] Tibshirani, R., *Regression Shrinkage and Selection via the LASSO*, Journal Royal Statistical Society B, 1996, 58(1):267-288.

[Titze et al. 2015] Titze, S., Schmid, M., Köttgen, A., Busch, M., Floege, J., Wanner, C., Kronenberg, F., Eckardt, K. U., Zacharias, H., Gronwald, W., Oefner, P. J., et al., *Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort*, Nephrol. Dial. Transplant., 2015, 30:441-451.

[Tonelli et al. 2011] Tonelli, M., Wiebe, N., Knoll, G., Bello, A., Browne, S., Jadhav, D., Klarenbach, S., Gill, J., *Systematic Review: Kidney Transplantation Compared With Dialysis in Clinically Relevant Outcomes*, American Journal of Transplantation, 2011, 11:2093-2109.

[Treuting and Kowalewska 2012] Treuting, P. M., Kowalewska, J., *Urinary System* in Treuting, P. M., Dintzis, S. M., (Ed.), *Comparative Anatomy and Histology*, Academic Press Elsevier, London, Waltham, San Diego, 2012 (1st edition).

[Trygg and Wold 2002] Trygg, J., Wold, S., *Orthogonal Projections to Latent Structures*, J. Chemometrics, 2002, 16:119-128.

[Tzoulaki et al. 2014] Tzoulaki, I., Ebbels, T. M. D., Valdes, A., Elliott, P., Ioannidis, J. P. A., *Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on -Omic Technologies*, American Journal of Epidemiology, 2014, 180(2): 129-139.

[van den Berg et al. 2006] van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., van der Werf, M. J., *Centering, scaling, and transformations: improving the biological information content of metabolomics data*, BMC Genomics, 2006, 7:142.

[van de Poll et al. 2004] van de Poll, M. C. G., Soeters, P. B., Deutz, N. E. P., Fearon, K. C. H., Dejong, C. H. C., *Renal metabolism of amino acids: Its role in interorgan amino acid exchange*, American Journal of Clinical Nutrition, 2004, 79:185-197.

[Varma and Simon 2006] Varma, S., Simon, R., *Bias in Error Estimation when Using Cross-Validation for Model Selection*, BMC-Bioinformatics, 2006, 7:91.

[Vaz and Wanders 2002] Vaz, F. M., Wanders, R. J. A., *Carnitine biosynthesis in mammals*, Biochemical Journal, 2002, 361:417-429.

[Venables and Ripley 2002] Venables, W. N., Ripley, B. D., *Modern Applied Statistics with S*, Springer-Verlag, 2002.

[Wagener et al. 2006] Wagener, G., Jan, M., Kim, M., Mori, K., Barasch, J. M., Sladen, R. N., Lee, H. T., *Association between increases in urinary neutrophil gelatinase-associated lipocalin and acute renal dysfunction after adult cardiac surgery*, The Journal of the American Society of Anesthesiologists, 2006, 105(3):485-491.

[Warrack et al. 2008] Warrack, B. M., Hnatyshyn, S., Ott, K.-H., Reily, M. D., Sanders, M., Zhang, H., Drexler, D. M., *Normalization strategies for metabonomic analysis of urine samples*, Journal of Chromatography B, 2009, 877:547-552.

[Weiss et al. 2011] Weiss, R. H., Kim, K., *Metabolomics in the study of kidney diseases*, Nature Reviews | Nephrology, 2011, 8(1):22-33.

[Weitz et al. 2006] Weitz, J., Koch, M., Mehrabi, A., Schemmer, P., Zeier, M., Beimler, J., Büchler, M., Schmidt, J., *Living-donor kidney transplantation: risks of the donor - benefits of the recipient*, Clin Transplant, 2006, 20(Suppl. 17):13-16.

[Westhuyzen et al. 2003] Westhuyzen, J., Endre, Z. H., Reece, G., Reith, D. M., Saltissi, D., Morgan, T. J., *Measurement of tubular enzymuria facilitates early detection of acute renal impairment in the intensive care unit*, Nephrology Dialysis Transplantation, 2003, 18(3):543-551.

[Wilkinson and Rogers 1973] Wilkinson, G. N. and Rogers, C. E., *Symbolic descriptions of factorial models for analysis of variance*, Applied Statistics, 1973,*22*:392-9.

[Wishart et al. 2007] Wishart, D. S., Tzur, D., Knox, C., Querengesser, L., *HMDB: The Human Metabolome Database*, Nucl. Acids Res., 2007, 35, D521.

[Wishart 2008] Wishart, D. S., *Metabolomics: A Complementary Tool in Renal Transplantation*, in Thongboonkerd, V. (ed.), *Proteomics in Nephrology - Towards Clinical Applications*, Contrib. Nephrol., Karger Press, Basel, 2008, vol. 160:76-87.

[Workman et al. 2002] Workman, C., Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielser, H. B., Saxild, H. H., Nielsen, C., Brunak, S., Knudsen, S., *A New Non-Linear Normalization Method for Reducing Variability in DNA Microarray Experiments*, Genome Biol., 2002, 3, research0048.

[Wyckoff and Augoustides 2012] Wyckoff, T., Augoustides, J. G. T., *Advances in Acute Kidney Injury Associated with Cardiac Surgery: The Unfolding Revolution in Early Detection*, Journal of Cardiothoracic and Vascular Anesthesia, 2012, 26(2): 340-345.

[Zacharias 2012] Zacharias, H. U., *Investigation of Acute Kidney Injury after Cardiac Surgery by NMR Spectroscopy and Machine Learning Methods*, M.Sc. thesis in physics, Institute of Functional Genomics, University of Regensburg, March 2012.

[Zacharias et al. 2013a)] Zacharias, H. U., Schley, G., Hochrein, J., Klein, M. S., Köberle, C., Eckardt, K.-U., Willam, C., Oefner, P. J., Gronwald, W., *Analysis of human urine reveals metabolic changes related to the development of acute kidney injury following cardiac surgery*, Metabolomics, 2013, 9(3):697-707.

[Zacharias et al. 2013b)] Zacharias, H. U., Hochrein, J., Klein, M. S., Samol, C., Oefner, P. J., Gronwald, W., *Current Experimental, Bioinformatic and Statistical Methods used in NMR Based Metabolomics*, Current Metabolomics, 2013, 1(3):253-268(16).

[Zacharias et al. 2015] Zacharias, H. U., Hochrein, J., Vogl, F. C., Schley, G., Mayer, F., Jeleazcov, C., Eckardt, K.-U., Willam, C., Oefner, P. J., Gronwald, W., *Identification of Plasma Metabolites Prognostic of Acute Kidney Injury after Cardiac Surgery with Cardiopulmonary Bypass*, Journal of Proteome Research, 2015, 14(7):2897-2905.

[Zhang et al. 2014] Zhang, A., Sun, H., Qiu, S., Wang, X., *Metabolomics insights into pathophysiological mechanisms of nephrology*, Int. Urol. Nephrol., 2014, 46:1025-1030.

[Zuckerman and Assimos 2009] Zuckerman, J. M., Assimos, D. G., *Hypocitraturia: pathophysiology and medical management*, Rev Urol., 2009, 11:134-144.