
Genomweite Assoziationsstudien mit metabolischen Merkmalen aus Kernspinresonanz (NMR) Spektroskopie

Johannes Raffler

Dissertation
an der Fakultät für Biologie
der Ludwig-Maximilians-Universität
München

vorgelegt von
Johannes Raffler
aus Augsburg

München, den 22. Juli 2014

Erstgutachter: Prof. Dr. Karsten Suhre

Zweitgutachter: Prof. Dr. John Parsch

Tag der mündlichen Prüfung: 15. April 2015

Inhaltsverzeichnis

1. Zusammenfassung	1
2. Publikationsliste	3
3. Einleitung	5
3.1. Metabolomik	5
3.2. Genomweite Assoziationsstudien	8
3.3. Metabolomik mittels NMR-Spektroskopie	15
3.4. Ziele dieser Arbeit	17
4. Materialien und Methoden	19
4.1. Studienpopulationen	19
4.2. Genotyp-Daten	21
4.3. Metabolomik-Daten	23
4.4. Durchführung der Assoziationsstudien	35
4.5. Ergebnisinterpretation	38
5. Ergebnisse	41
5.1. Erste GWAS mit metabolischen Merkmalen im Urin	41
5.2. Zweite GWAS mit metabolischen Merkmalen im Urin	49
5.3. GWAS mit NMR-Merkmalen im Blutplasma	56
6. Diskussion	67
6.1. Die erste GWAS mit Stoffwechselprodukten im Urin	68
6.2. Genetisch beeinflusste metabolische Phänotypen im Urin und Blut . .	79
6.3. Identifikation und MS-gestützte Interpretation genetisch beeinflusster NMR-Merkmale im Blutplasma	113
7. Literaturverzeichnis	125
8. Tabellenverzeichnis	153

9. Abbildungsverzeichnis	155
10. Lebenslauf	157
11. Erklärung	159
12. Danksagungen	161
A. Anhang zur zweiten mGWAS im Urin	163
B. Anhang zur GWAS mit NMR-Merkmalen im Plasma	173

1. Zusammenfassung

Genomweite Assoziationsstudien mit metabolischen Merkmalen (mGWAS) haben in den letzten Jahren erfolgreich gezeigt, wie mit Hilfe der Metabolomik die genetischen Grundlagen des menschlichen Stoffwechsels erforscht werden können. In der vorliegenden Arbeit werden drei Studien vorgestellt, in denen anhand von Daten tausender Probanden der SHIP- und KORA-Kohortenstudien untersucht wird, inwiefern Variationen des Erbgutes im metabolischen Phänotypen (Metabotyp) sichtbar werden. In diesen mGWAS werden rund 600.000 Einzelnukleotid-Polymorphismen (*single nucleotide polymorphisms*; SNPs) auf Assoziationen zu metabolischen Merkmalen im Urin und Blutplasma, welche mittels Kernspinresonanz-Spektroskopie (*nuclear magnetic resonance*; NMR) bestimmt wurden, untersucht.

Zwei der vorgestellten Studien sind mGWAS im Urin. Der Fokus liegt hier also nicht auf der Homöostase von Metaboliten im Blut, sondern auf den ausgeschiedenen Stoffwechselendprodukten und der Nierenfunktion. Es konnten zahlreiche statistisch signifikante Assoziationen zwischen SNPs und Metabolitenkonzentrationen bzw. dem paarweisen Konzentrationsverhältnis zweier Stoffwechselprodukte (*ratios*) identifiziert werden. In der ersten Studie wurden die Daten von mehr als 800 Probanden der SHIP-0-Kohorte untersucht. Die in dieser Studie signifikant assoziierten SNPs markieren chromosomale Abschnitte (Loci), die im Bereich der Gene AGXT2, NAT2, HPD, SLC7A9 sowie SLC6A20 liegen. In der zweiten Studie wurde die Anzahl der untersuchten Probanden nahezu verfünffacht. Dadurch konnten neun weitere Loci identifiziert werden, die signifikant mit metabolischen Merkmalen im Urin assoziieren (CPS1, SLC16A9, SLC6A13, XYLB, ETNPPL, SLC6A19, DMGDH, SLC36A2, FBXL20). Bei der Interpretation der Ergebnisse lässt sich in vielen Fällen ein funktioneller Zusammenhang zwischen dem jeweiligen genetischen Locus und dem assoziierten metabolischen Merkmal herstellen. Darüber hinaus sind für die Mehrzahl der Loci Assoziationen mit klinisch relevanten Merkmalen und Krankheiten bekannt.

Die dritte in dieser Arbeit behandelte Studie ist eine GWAS mit NMR-Merkmalen im Blutplasma. Die Besonderheit dieser Studie besteht darin, dass die NMR-Spektren der Proben nicht annotiert wurden, also keine der GWAS vorausgehende Identifizierung und Quantifizierung der Metaboliten erfolgte. Es wird gezeigt, dass in einer hypothesenfreien Herangehensweise auch ohne aufwändige Spektren-Annotation signifikante genetische Assoziationen zu NMR-Merkmalen identifiziert werden können. Es wurden sieben Loci identifiziert, welche signifikant mit NMR-Ratios assoziieren (GCKR, CPS1, PYROXD2, FADS1, APOA1, LIPC, CETP). Diese NMR-Ratios können in künftigen Studien als leicht zu bestimmende metabolische Marker verwendet werden. Weiterhin wird in der Studie eine Methode vorgestellt, wie NMR-Spektren mithilfe von Korrelationen mit Metabolitenkonzentrationen, die auf anderen Plattformen gemessen wurden, interpretiert werden können. Dadurch ist ein Wissenstransfer zwischen komplementären Metabolomik-Plattformen möglich.

2. Publikationsliste

K. Suhre, H. Wallaschofski, **J. Raffler**, N. Friedrich, R. Haring, K. Michael, C. Wasner, A. Krebs, F. Kronenberg, D. Chang, C. Meisinger, H.-E. Wichmann, W. Hoffmann, H. Völzke, U. Völker, A. Teumer, R. Biffar, T. Kocher, S.B. Felix, T. Illig, H.K. Kroemer, C. Gieger, W. Römisch-Margl und M. Nauck. A genome-wide association study of metabolic traits in human urine. *Nat Genetics*, 43(6):565-9, Juni 2011

K. Suhre, S.-Y. Shin, A.-K. Petersen, R.P. Mohny, D. Meredith, B. Wägele, E. Altmaier, CARDIoGRAM, P. Deloukas, J. Erdmann, E. Grundberg, C.J. Hammond, M.H. de Angelis, G. Kastenmüller, A. Köttgen, F. Kronenberg, M. Mangino, C. Meisinger, T. Meitinger, H.-W. Mewes, M. Milburn, C. Prehn, **J. Raffler**, J.S. Ried, W. Römisch-Margl, N.J. Samani, K.S. Small, H.-E. Wichmann, G. Zhai, T. Illig, T.D. Spector, J. Adamski, N. Soranzo und C. Gieger. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 31;477(7362):54-60, August 2011

A.-K. Petersen, K. Stark, M.D. Musameh, C.P. Nelson, W. Römisch-Margl, W. Kremer, **J. Raffler**, S. Krug, T. Skurk, M.J. Rist, H. Daniel, H. Hauner, J. Adamski, M. Tomaszewski, A. Döring, A. Peters, H.-E. Wichmann, B.M. Kaess, H.R. Kalbitzer, F. Huber, V. Pfahlert, N.J. Samani, F. Kronenberg, H. Dieplinger, T. Illig, C. Hengstenberg, K. Suhre, C. Gieger und G. Kastenmüller. Genetic associations with lipoprotein subfractions provide information on their biological nature. *Hum Mol Genet*, 21(6):1433-43, März 2012

S. Krug, G. Kastenmüller, F. Stückler, M.J. Rist, T. Skurk, M. Sailer, **J. Raffler**, W. Römisch-Margl, J. Adamski, C. Prehn, T. Frank, K.H. Engel, T. Hofmann, B. Luy, R. Zimmermann, F. Moritz, P. Schmitt-Kopplin, J. Krumsiek, W. Kremer, F. Huber, U. Oeh, F.J. Theis, W. Szymczak, H. Hauner, K. Suhre und H. Daniel. The dynamic range of the human metabolome revealed by challenges. *FASEB J*, 26(6):2607-19, Juni 2012

J. Raffler, W. Römisch-Margl, A.K. Petersen, P. Pagel, F. Blöchl, C.Hengstenberg, T. Illig, C. Meisinger, K. Stark, H.-E. Wichmann, J. Adamski, C. Gieger, G. Kastenmüller und K. Suhre. Identification and MS-assisted interpretation of genetically influenced NMR signals in human plasma. *Genome Med*, 5(2):13, Februar 2013

3. Einleitung

Metaboliten sind kleine Moleküle (< 1.500 Da [1]), die als Ausgangs-, Zwischen- und Endprodukte des biochemischen Stoffwechsels in Organismen zu finden sind, also beispielsweise Zucker, Lipide oder Aminosäuren. Dabei wird zwischen endogenen Metaboliten, also jenen Molekülen, die ein Organismus eigenständig enzymatisch synthetisiert, und exogenen Metaboliten, die über die Nahrung oder über die Umwelt aufgenommen werden, unterschieden [2]. Die Gesamtheit aller Metaboliten eines biologischen Systems wird als das Metabolom bezeichnet [3]. Erklärtes Ziel der Metabolomik ist die systematische und möglichst vollständige Erfassung des Metaboloms. Dabei umfasst das Metabolom eines Menschen tausende endogene Metaboliten und zudem noch mutmaßlich zehntausende weitere nicht-endogene Substanzen [2]. So enthält die von Wishart *et al.* erstellte und verwaltete Datenbank *Human Metabolome Database* (HMDB) über 4.000 unterschiedliche Metaboliten im Blutserum sowie knapp 3.000 klassifizierte Stoffwechselprodukte im Harn, und ständig werden dieser Datenbank neue Substanzen hinzugefügt [4–7].

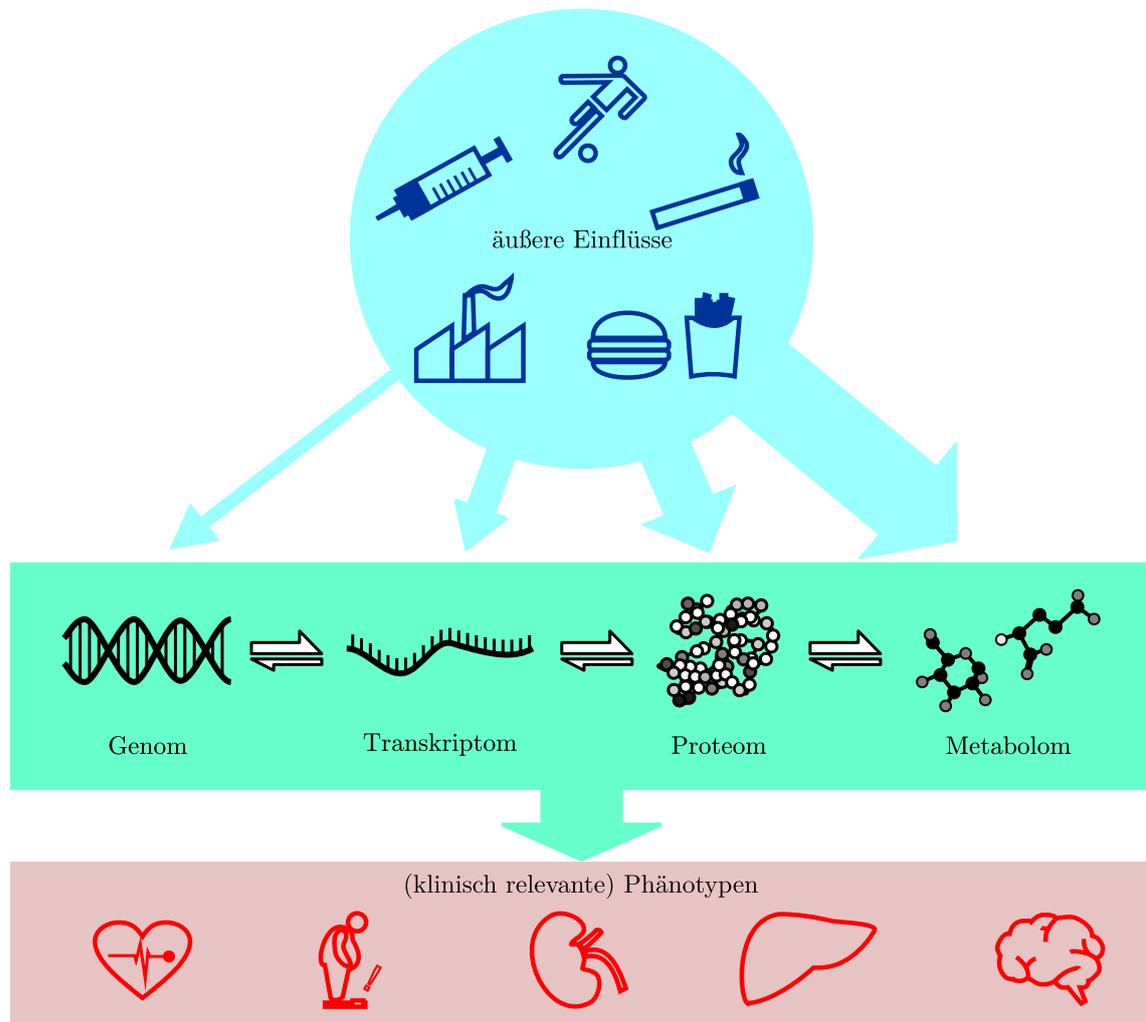
3.1. Metabolomik

Die Metabolomik ist im Vergleich zu ihren bekannteren Verwandten wie der Genomik oder Proteomik ein noch relativ junges Forschungsgebiet: die ersten Erwähnungen der Termini „Metabolom“ (engl. *metabolome*) sowie „Metabolomik“ (*metabolomics*) findet sich in um die Jahrtausendwende publizierten Artikeln, beispielsweise von Tweeddale *et al.* (1998), Nicholson *et al.* (1999) oder Fiehn *et al.* (2001) [8–10]. Auch wenn es mittels klassischer Laborchemie bereits seit geraumer Zeit möglich war, Stoffwechselprodukte in biologischen Proben (wie Blutserum oder Urin) zu identifizieren und zu quantifizieren, so eröffnete erst die in den letzten Jahren rasant fortgeschrittene Weiterentwicklung analytischer Plattformen, die auf Massenspektrometrie (MS) oder Kernspinresonanzspektroskopie (*nuclear magnetic resonance spectroscopy*, kurz

NMR) basieren, neue Möglichkeiten. Auf diesen Plattformen können heute innerhalb weniger Minuten in einer Probe eine breite Palette verschiedener chemischer Moleküle simultan identifiziert und teilweise auch quantifiziert werden. Ein hoher Durchsatz (*high throughput*) erlaubt es durch die Verkürzung der Zeitspanne, die zum Aufbereiten und Messen einer Probe nötig ist, die metabolischen Profile tausender Proben bzw. Probanden zu messen. Wie allerdings Bouatra *et al.* berechtigterweise anmerken, hinkt die Metabolomik den oben erwähnten etablierten „-omik“-Feldern gegenwärtig noch hinterher, was die Abdeckung des Metaboloms betrifft. Zum jetzigen Zeitpunkt können selbst von den modernsten analytischen Plattformen - bedingt durch die physiochemische Diversität der Metaboliten - höchstens einige wenige hundert Stoffwechselprodukte erfasst werden, was bezogen auf das Metabolom des Menschen lediglich einem niedrigen Prozentsatz aller bekannten Metaboliten entspricht [6]. Demgegenüber beträgt die Abdeckung des Genoms durch neue Sequenzierungstechniken (*next-generation sequencing*) nahezu 100 Prozent, bei ebenfalls hohem Probendurchsatz.

Trotz der Einschränkung, dass Metabolomik-Plattformen zum gegenwärtigen Zeitpunkt noch ein lückenhaftes Bild des Metaboloms liefern, bietet die systematische Untersuchung des Stoffwechsels bereits heute neue und spannende Perspektiven auf biologische Vorgänge, welche die Betrachtung des „statischen“ Genoms alleine nicht bieten kann. Oliver Fiehn bezeichnet Metaboliten als Endprodukte aller regulatorischen Prozesse eines biologischen Systems [3]. Dies bedeutet, dass auf der Ebene des Metaboloms die Auswirkungen regulatorischer Mechanismen und Interaktionen, welche in den zugrunde liegenden Ebenen wie dem Genom, dem Transkriptom oder dem Proteom stattfinden, sichtbar und damit messbar werden (Abbildung 3.1 Mitte). Das Metabolom wird jedoch nicht nur vom genetischen Unterbau, sondern zu einem großen Teil auch von äußeren Einflüssen wie Umweltbedingungen, Ernährung und Lebensstil (Abbildung 3.1 oben) sowie intrinsischen Faktoren wie Erkrankungen oder dem Lebensalter beeinflusst. Es vereint also die Antwort biologischer Prozesse auf sowohl genetische Faktoren als auch Umwelteinflüsse und ist daher laut David Wishart das wohl beste Abbild des physiologischen Zustandes eines Individuums [11]. Hinzu kommt, dass sich sowohl die inneren als auch die äußeren Einflüsse auf den Stoffwechsel innerhalb kürzester Zeit ändern können und dies rasch auf der Ebene des Metaboloms sichtbar wird, wohingegen der morphologische Phänotyp vergleichsweise träge reagiert.

Abbildung 3.1.: Innere und äußere Einflüsse auf Phänotypen



Oben: Effekt äußerer Einflüsse (u.a. Umweltbedingungen, Lebensstil, Medikation, Ernährung) auf die Prozesse im Organismus. Im Allgemeinen nimmt der (messbare) Effekt in Richtung der Metabolom-Ebene zu (angedeutet durch die Stärke der Pfeile). Mitte: Schematische Darstellung des Informationsflusses vom genetischen Code (Transkription) über Proteine (Translation) hin zu Metaboliten. Die Doppelpfeile weisen darauf hin, dass dieser Informationsfluss nicht strikt in eine Richtung erfolgt; so können beispielsweise Metaboliten regulatorische Effekte auf der Proteinebene oder der Transkriptionsebene haben. Unten: Klinisch relevante Phänotypen wie Herz-Kreislauf-Erkrankungen, Nierenleiden oder mentale Störungen. Diese Phänotypen sind das Resultat genetischer Faktoren und der damit verbundenen endogenen Reaktionen sowie äußerer Einflüsse. Abbildung angelehnt an Kaddurah-Daouk und Ranga Rama [12].

Die Bedeutung der Untersuchung metabolischer Merkmale, welche als Bindeglied zwischen der genetischen Ebene, äußeren Einflüssen und Phänotypen fungieren, nimmt daher stetig zu. Dies gilt besonders für Studien, die auf komplexe (Stoffwechsel-)Krankheiten wie Typ-2-Diabetes oder Herz-Kreislauf-Erkrankungen abzielen. Hier bietet sich einerseits die Möglichkeit, anhand metabolischer Profile unterschiedlicher Gewebetypen und Körperflüssigkeiten den (patho-)physiologischen Zustand eines Probanden zu bestimmen [13], wodurch neue Biomarker für die Diagnostik im klinischen Umfeld entwickelt werden können. Andererseits dienen Metabolomik-Studien dazu, das Zusammenspiel verschiedener Faktoren, die zur Entstehung eines klinisch relevanten Phänotypen beitragen, besser verstehen zu können [6, 14].

3.2. Genomweite Assoziationsstudien

Es gibt mehrere Ansätze, um molekulare Mechanismen, welche zur Ausprägung eines Phänotypen führen, zu identifizieren und zu charakterisieren. Einer dieser Ansätze untersucht die genetische Ebene und bedient sich dabei genomweiter Assoziationsstudien (GWAS). Nach der vollständigen Entschlüsselung des menschlichen Genoms in den ersten Jahren des 21. Jahrhunderts wurden Microarray-Chips entwickelt und kommerziell vermarktet. Durch diese „Genchips“ wurde es erstmals möglich, kostengünstig die genetische Variation des menschlichen Genoms in großen Studienpopulationen zu bestimmen, ohne dabei auf kostspielige und zeitaufwändige Sequenzierung zurückgreifen zu müssen [15]. Mit Hilfe derartiger Chips kann für jeden Probanden der Genotyp hunderttausender definierter Einzel-Nukleotid-Polymorphismen (*single nucleotide polymorphisms*, SNPs), welche relativ gleichmäßig über das gesamte Genom verteilt sind, gemessen werden. Zwar können diese SNPs durchaus funktionale Auswirkungen auf den Organismus haben, beispielsweise durch eine veränderte Sekundärstruktur von DNA-Transkripten oder durch einen Aminosäureaustausch im fertigen Protein; das Gros der durch Genchips gemessenen SNPs hat jedoch keinen (sichtbaren) direkten Effekt [16]. In GWAS, die Daten von Genchips verwenden, dienen SNPs daher im Allgemeinen lediglich als Marker für Sequenzabschnitte, die bei der meiotischen Rekombination nicht getrennt werden und dadurch am Stück vererbt werden (Haplotyp-Blöcke). Der beobachtete Effekt auf den Phänotypen wird mutmaßlich von einer genetischen Variante verursacht, welche mit dem Marker-SNP gekoppelt ist und folglich gemeinsam vererbt wird. Der innerhalb einer Population

beobachtete Grad der Kopplung zweier SNPs kann mit dem Korrelationskoeffizienten r^2 beschrieben werden: bei $r^2 = 1$ befinden sich zwei Varianten maximal im Kopplungsungleichgewicht, wobei der englische Begriff *linkage disequilibrium* (LD) auch in deutschsprachigen Publikationen weitaus geläufiger ist.

Ein SNP hat in der Regel zwei Allele: ein häufigeres, wohl das ursprüngliche Allel (*major allele*), sowie das seltenere, mutierte Allel (*minor allele*). Bezogen auf einen SNP teilt sich die Bevölkerung in drei Gruppen: aufgrund der Diploidie kann eine Person für einen SNP homozygoter Träger des häufigen bzw. seltenen Allels sein oder aber heterozygoter Träger beider Allele. Die Frequenz des seltenen Allels (*minor allele frequency*, kurz MAF) gibt dabei an, wie häufig ein SNP in der Bevölkerung auftritt [16]. Allerdings werden bei Chip-basierten GWAS zumeist nur SNPs betrachtet, die bei einer MAF von mindestens 1-5 % in der Studienpopulation vorliegen (*common variants*) [17].

In einer GWAS wird für jeden SNP untersucht, ob sich die aus den unterschiedlichen Allelkombinationen resultierenden Probandengruppen bezüglich eines definierten Phänotyps statistisch signifikant voneinander unterscheiden. In der Regel wird der Phänotyp durch eine kategoriale Variable, z.B. ein standardisierter klinischer Endpunkt in einer Fall-Kontroll-Studie, oder durch eine quantitative Variable wie beispielsweise der Körpergröße beschrieben. Während kategorial unterteilte Phänotypen üblicherweise entweder mittels Kontingenztafeln oder logistischer Regression auf genetische Assoziation hin untersucht werden, wird bei quantitativen Merkmalen zumeist die lineare Regression eingesetzt, häufig unter Verwendung linearer additiver Modelle [16]. Bei einem solchen additiven genetischen Modell wird unter optionaler Berücksichtigung zusätzlicher Kovariaten wie dem Alter oder dem Geschlecht der Probanden angenommen, dass jede Kopie des Effektallels (häufig wird hier einheitlich das seltene Allel verwendet) proportional mit einer Zu- oder Abnahme des phänotypischen Merkmals einhergeht [18]. Als Testergebnis spiegelt der Regressionskoeffizient β dabei die Größe und die Richtung des Effekts wieder. Neben dem β -Wert ist der p -Wert ein weiterer wichtiger Ergebniswert des linearen Modells. Ein Testergebnis für eine einzelne SNP-zu-Phänotyp-Assoziation gilt als statistisch signifikant, wenn der p -Wert kleiner oder gleich dem nominellen Signifikanzniveau α ist. In den meisten Studien gilt $\alpha = 0,05$. Dies bedeutet, dass bei 5 % der Tests die Null-Hypothese (kein aus dem Genotyp resultierender phänotypischer Unterschied zwischen den Gruppen) fälschlicherweise verworfen wird und somit das Testergebnis falsch-positiv ist. Allerdings wird in einer GWAS nicht nur ein Test,

sondern hunderttausende voneinander unabhängige Tests auf dem gleichen Datensatz durchgeführt, wobei jeder Test mit einer eigenen falsch-positiv-Rate einhergeht. Dieser mit dem multiplen Testen einhergehenden Alpha-Fehler-Kumulierung kann beispielsweise durch die konservative Bonferroni-Korrektur begegnet werden. Hier gilt das Ergebnis eines statistischen Test nur dann als signifikant, wenn der p-Wert kleiner oder gleich dem angepassten Signifikanzniveau $\alpha' = \alpha/n$ ist, wobei n der Anzahl der durchgeführten Tests entspricht. Bei einer GWAS, in der 1.000.000 SNPs auf Assoziation mit einem Phänotypen getestet werden, liegt folglich die Obergrenze für signifikante p-Werte bei $0,05 / 1.000.000 = 5,0 \times 10^{-8}$.

GWAS mit klinischen Phänotypen

Im Jahre 2005 wurde die erste erfolgreiche GWAS mit einem komplexen klinischen Phänotypen publiziert: Klein *et al.* identifizierten Assoziationen von Polymorphismen im *Complement Factor H*-Gen mit altersbedingter Makulardegeneration und lieferten zugleich eine biologische Erklärung für den beobachteten Effekt [16, 19]. In den folgenden Jahren wurden mittels GWAS hunderte genetische Regionen (Loci) identifiziert, welche mit den unterschiedlichsten Phänotypen wie Körpergröße [20, 21], Haut- oder Haarfarbe [22, 23], aber auch mit häufigen Krankheitsbildern wie Typ-1- und Typ-2-Diabetes [24–26] oder Koronarer Herzkrankheit [27] assoziieren. Mittlerweile umfasst der GWAS-Katalog des *National Human Genome Research Institute* (NHGRI) über 1.700 publizierte Studien (Stand November 2013) [28]. Allerdings zeigte sich, dass der Anteil der Varianz, die durch die Modelle der signifikant assoziierten SNPs erklärt wird, bei den meisten phänotypischen Merkmalen sehr klein ist: üblicherweise liegt der Wert bestenfalls im niedrigen einstelligen Prozentbereich. Einerseits ist dies darauf zurückzuführen, dass die genetischen Varianten, die in GWAS untersucht werden, wie oben erwähnt in der Bevölkerung recht häufig auftreten, deren Penetranz im Allgemeinen aber sehr niedrig ist. Im Gegensatz hierzu sind Varianten, die beispielsweise ein Protein funktionsunfähig machen und damit zu einem bestimmten Krankheitszustand führen, sehr selten. Andererseits beeinflusst vor allem bei komplexen Erkrankungen zumeist nicht nur ein einzelner genetischer Locus, sondern mehrere Loci einen Phänotypen [15, 29]. So wurden beispielsweise für Multiple Sklerose bereits über 50 assoziierte Loci identifiziert. Die in diesen Loci signifikant assoziierten SNPs erklären jedoch in Summe lediglich 10 % der phänotypischen Varianz [29]. Um dennoch derartige Loci mit geringen Effekten auf klinische Endpunkte zu identifizieren, bedarf es aufgrund der dazu notwendigen statistischen Teststärke

sehr großer Studienpopulationen. Visscher *et al.* schätzen, dass eine Verdoppelung der Studienpopulationsgröße ungefähr mit einer Verdoppelung der signifikant assoziierten genetischen Varianten einhergeht. Deshalb werden bei aktuellen und künftigen GWAS mit klinischen Phänotypen die Daten tausender bis hunderttausender Probanden herangezogen. Im Vergleich hierzu erscheint die in der oben genannte Studie von Klein *et al.* untersuchte Studienpopulation, welche ungefähr 150 Probanden umfasst [29], fast schon verschwindend klein zu sein.

GWAS mit metabolischen Merkmalen als intermediäre Phänotypen

Das geschilderte Problem geringer Effekte bei GWAS mit klinischen Phänotypen kann umgangen werden, wenn stattdessen intermediäre Phänotypen auf genetische Assoziation hin getestet werden [15], wie beispielsweise die im Blut bestimmten Konzentrationen von Triglyceriden, Cholesterin oder Glucose. Diese metabolischen Merkmale sind bekannte Risikofaktoren für klinische Endpunkte wie Diabetes oder Gefäßkrankheiten. In GWAS mit Metaboliten als untersuchtem Phänotyp zeigte sich, dass die Auswirkungen einzelner genetischer Varianten im Metabolom deutlich stärker sichtbar sind als im klinischen Phänotypen [1]. Dies mag dem Umstand geschuldet sein, dass metabolische Merkmale tendenziell durch weniger genetische Loci beeinflusst werden als klinische Phänotypen. Außerdem können metabolische Phänotypen dank der hohen Sensitivität und Spezifität moderner analytischer Plattformen definierter ermittelt werden, als dies bei vielen klinischen Endpunkten möglich ist.

Gieger *et al.* veröffentlichten 2008 die erste GWAS, in der ein breites Spektrum unterschiedlicher Metaboliten auf genetische Assoziationen untersucht wurde. Die Autoren untersuchten in dieser für weitere GWAS mit metabolischen Merkmalen (mGWAS) wegweisenden Arbeit die Konzentrationen von 363 unterschiedlichen Stoffwechselprodukten im Blutserum von 284 Teilnehmern der KORA-F3-Populationsstudie [30]. Gieger und dessen Kollegen identifizierten eine Reihe genetischer Polymorphismen im Bereich von Genen, welche für die Enzyme SCAD (alias ACADS), MCAD (alias ACADM), LIPC sowie FADS1 kodieren. Da diese Polymorphismen sich in spezifischen und klar voneinander zu unterscheidenden metabolischen Phänotypen (Metabotypen) manifestieren, führten Gieger *et al.* den Begriff des genetisch bedingten Metabotypen (*genetically determined metabotype*) ein. Da dieser Ausdruck jedoch den Einfluss der Gene auf den Stoffwechsel überbetont und dabei die Effekte von

Umwelteinflüssen und Lebensstil außer Acht lässt, wird in der vorliegenden Arbeit der mittlerweile etablierte Begriff des genetisch beeinflussten Metabotypen (*genetically influenced metabotype*, GIM) verwendet.

Die von Gieger *et al.* identifizierten SNPs erklären bis zu 12 % der beobachteten Varianz der Konzentration des jeweils assoziierten Metaboliten. Allerdings erreichte keine dieser Assoziationen genomweite Signifikanz ($\alpha' = 1,33 \times 10^{-9}$), was die Autoren auch auf die geringe Anzahl der untersuchten Probanden zurückführten. Gieger *et al.* untersuchten jedoch in ihrer GWAS zusätzlich zu den Metabolitenkonzentrationen auch die aus allen gemessenen Metaboliten berechneten paarweisen Konzentrationsverhältnisse (*ratios*). Das Konzept des hypothesenfreien Testens von Ratios wurde in einer von Altmaier *et al.* im Juli 2008 publizierten Medikationsstudie eingeführt und von Gieger *et al.* auf mGWAS übertragen [31]. Bei Verwendung der Ratios als phänotypisches Merkmal nahm die Assoziationsstärke bei p-Werten zwischen 10^{-10} und 10^{-21} um bis zu vierzehn Größenordnungen zu, und die durch einzelne SNPs erklärte Varianz metabolischer Merkmale betrug bis zu 28 %.

Es gibt mehrere mutmaßliche Gründe für die bei Ratios beobachtete Zunahme der Assoziationsstärke. Wurde ein Ratio aus zwei biochemisch verwandten Metaboliten gebildet, beispielsweise dem Substrat und Produkt einer enzymatisch gesteuerten Reaktion, kann dieser Ratio ein ungefähres Abbild der Reaktionsgeschwindigkeit liefern. Beeinflusst ein genetischer Polymorphismus diese Umwandlungsrate, so zeigt sich dies im Ratio deutlicher als in den Konzentrationen einzelner Metaboliten, da diese auch in anderen Reaktionen des Stoffwechsels metabolisiert werden können (siehe Abbildung 3.2a). Ein anschauliches Beispiel hierfür liefert die von Gieger *et al.* identifizierte Assoziation des SNPs rs174548 im FADS1-Locus zu den Glycerophosphatidylcholinen PC aa C36:3 und C36:4. Diese PCs sind modifizierte Produkte und Substrate der FADS1-gesteuerten delta-5-Desaturasereaktion. Hier äußert sich eine durch rs174548 direkt oder indirekt reduzierte katalytische Aktivität in erhöhten Substrat- (PC aa C36:3) und verringerten Produktkonzentrationen (PC aa C36:4) [30]. Das Konzentrationsverhältnis dieser PCs bildet also wahrscheinlich die FADS1-Reaktionsgeschwindigkeit ab. Die Assoziation des SNPs rs174548 zu diesem Ratio ist um mehrere Größenordnungen stärker als zu den einzelnen PCs (Abbildung 3.2b).

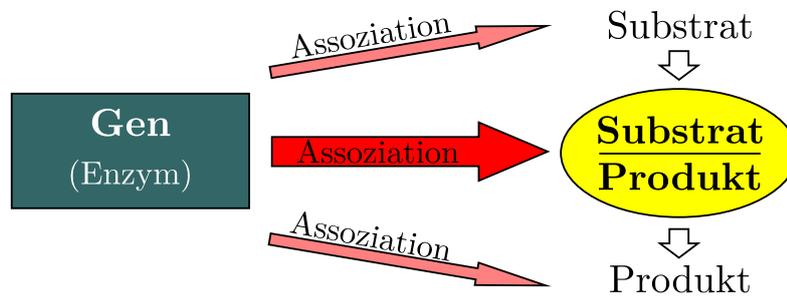
Weiterhin können Ratios einen normalisierenden Effekt auf den Datensatz haben: Schwankungen in den Stoffwechselkonzentrationen, die im Datensatz durch interindividuelle, beispielsweise durch den Lebensstil bedingte Unterschiede zwischen den

Probanden, durch Verdünnungseffekte in den Proben oder durch schwankende Qualität der Messungen bei schwer zu quantifizierenden Metaboliten auftreten, können durch Ratios teilweise ausgeglichen werden. Das „Rauschen“ im metabolischen Datensatz wird dadurch verringert [30–33].

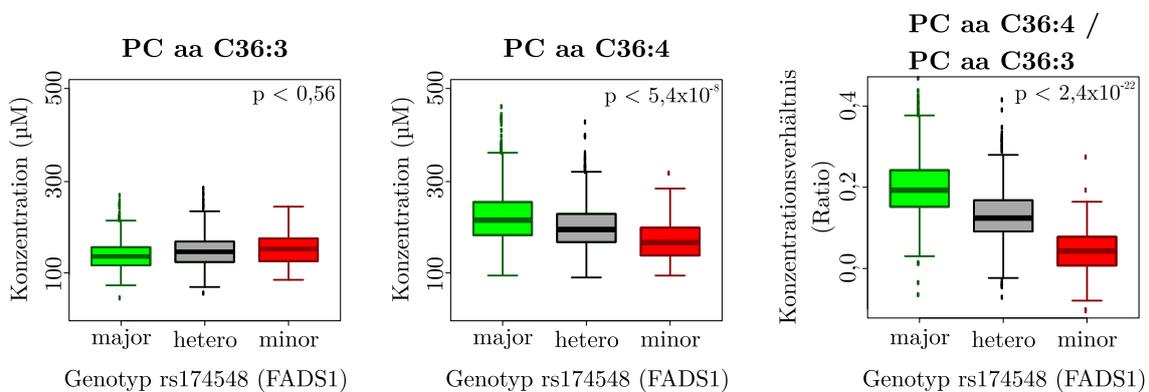
Der Studie von Gieger *et al.* folgte eine Reihe weiterer genomweiter Assoziationsstudien, in denen die Zusammenhänge zwischen genetischen Polymorphismen und Metabolitenkonzentrationen im Blutplasma und -serum tausender Probanden untersucht wurden, und in vielen dieser Studien wurde das Konzept der Ratios angewandt (z.B. [30, 34–43]). Insgesamt konnten in mGWAS weit über einhundert unterschiedliche genetisch beeinflusste Metabotypen (GIMs) identifiziert werden. Hervorzuheben sind an dieser Stelle die 2011 bzw. 2012 veröffentlichten Studien von Suhre *et al.* (37 GIMs) [35] und Kettunen *et al.* (31 GIMs) [38] sowie die im Frühjahr 2014 veröffentlichte GWAS von Shin *et al.* [43], in der 145 GIMs identifiziert wurden.

In ihrer Gesamtheit bilden GIMs die Basis für die metabolische Individualität des Menschen. Vereint man nun bereits vorhandenes Wissen über genetische Loci, die laut früheren Studien mit Krankheitsbildern assoziieren, mit den Erkenntnissen aus mGWAS, wird es möglich, zusätzliche metabolische Risikofaktoren für weit verbreitete Krankheiten zu identifizieren und neue Einsichten in deren biologischen Mechanismen zu erlangen, um dadurch Ansatzpunkte für therapeutische Maßnahmen zu finden. In vielen Fällen konnten bereits Verbindungen zwischen GIMs und häufig auftretenden Krankheiten wie Typ-2-Diabetes, Herz-Kreislauf-Erkrankungen sowie chronischem Nierenversagen hergestellt werden. So sind beispielsweise genetische Polymorphismen im oben genannten FADS1-Locus mit Morbus Crohn assoziiert [44] und stehen weiterhin unter Verdacht, durch die Assoziation mit Cholesterin und Triglyceriden eine Rolle bei Herz-Kreislauf-Erkrankungen zu spielen [45]. Weitere Beispiele sind das GCKR-Gen, welches ein Risiko-Locus für verschiedene Diabetes-relevante Merkmale ist [46], sowie genetische Varianten im CPS1-Locus, die mit Risikofaktoren für Nierenerkrankungen assoziieren [47]. Auch außerhalb des klinischen Kontexts können genetische Assoziationen zu Stoffwechselprodukten dazu beitragen, die Funktion genetischer Loci genauer zu charakterisieren und ihre Auswirkungen auf die metabolischen Individualität zu ermitteln.

Abbildung 3.2.: Prinzip von Ratios in genomweiten Assoziationsstudien mit metabolischen Merkmalen



(a) Schematische Darstellung des Ratio-Konzeptes



(b) Ratio am Beispiel einer Assoziation im FADS1-Locus

Aus paarweisen Metaboliten gebildete Konzentrationsverhältnisse (*Ratios*; gelbes Oval) zeigen in mGWAS häufig weitaus stärkere Assoziationen (dicker roter Pfeil) zu genetischen Polymorphismen als die einzelnen Metabolitenkonzentrationen, aus denen sich der jeweilige Ratio zusammensetzt. Falls diese Polymorphismen die Geschwindigkeit einer enzymatisch gesteuerten Reaktion beeinflussen, spiegelt der Ratio diese Reaktionsrate wider.

Zum Beispiel ist der bei Gieger *et al.* beobachtete Effekt sowie die Assoziationsstärke im Falle des SNPs rs174548 (FADS1-Locus) und dem Ratio, der aus den Glycerophosphatidylcholinen PC aa C36:4 und PC aa C36:3 gebildet wurde ((b) rechts), deutlich stärker als bei den einzelnen Metaboliten ((b) links und Mitte). Diese PCs sind das modifizierte Produkt (C36:4) bzw. Substrat (C36:3) der FADS1-gesteuerten delta-5-Desaturasereaktion. Der Ratio liefert also mutmaßlich eine Annäherung der Reaktionsgeschwindigkeit, welche durch den Genotyp des SNPs direkt oder indirekt beeinflusst wird. Die abgebildeten Boxplots beschreiben die Verteilung der in den Blutproben der Probanden gemessenen Metabolitenkonzentrationen ((b) links und Mitte) bzw. dem daraus errechneten Ratio ((b) rechts). Die Probanden wurden dabei anhand ihres rs174548-Genotyps kategorisiert (*major homozygotes*, *heterozygotes* und *minor homozygotes*). Allgemeine Erläuterungen zu Boxplots finden sich im Begleittext zu Abbildung 6.2(a). Abbildung (a) nach Kronenberg [33]; Abbildung (b) nach Gieger *et al.* [30].

3.3. Metabolomik mittels NMR-Spektroskopie

Wie in Abschnitt 3.2 erläutert ermöglichte erst die kostengünstige Genotypisierung ganzer Studienpopulationen mittels Genchips die Möglichkeit der Durchführung von genomweiten Assoziationsstudien. Für GWAS mit metabolischen Merkmalen ist die umfassende metabolische Charakterisierung der Probanden eine weitere Voraussetzung. Es kommen hierzu im Allgemeinen zwei Technologien zum Einsatz: Massenspektrometrie (MS) sowie NMR-Spektroskopie.

Die Massenspektrometrie beruht auf der Bestimmung des Masse-zu-Ladungsverhältnisses der in einer ionisierten Probe enthaltenen Substanzen. Zur besseren Bestimmung von Molekülen ähnlicher Masse erfolgt häufig eine Kopplung mehrerer Massenspektrometer (beispielsweise zur Fragmentierungsmuster-Analyse bei Tandem-MS) oder eine der Messung vorgeschaltete Auftrennung der Proben durch ein Chromatographieverfahren.

Bei der NMR-Spektroskopie spielt die Masse eines Moleküls keine primäre Rolle. Stattdessen wird hier die Probe einem starken Magnetfeld ausgesetzt und für bestimmte Atome (wie z.B. Wasserstoff) die Absorption von hochfrequenten elektromagnetischen Signalen gemessen. Die Stärke und Frequenz der gemessenen Signale hängt davon ab, in welcher „chemischen Umgebung“ (Wechselwirkungen mit benachbarten Atomen) sich ein NMR-sensitives Atom befindet. Somit erzeugt jedes Molekül einen mehr oder weniger eindeutigen spektralen „Fingerabdruck“ [2].

Beide Plattformen weisen jeweils Vor- und Nachteile auf. MS-basierte Analysemethoden sind ungleich sensitiver als NMR-Plattformen. Mittels MS können also Substanzen gemessen werden, deren Konzentration weit unterhalb der Detektionsgrenze von NMR-Spektrometern liegt. Somit können in der Praxis eine größere Anzahl Metaboliten in einer Probe bestimmt werden. Weiterhin kommt MS im Vergleich zu NMR mit weitaus geringen Probengrößen aus, was bei Kohortenstudien, in denen nur begrenzt Probenmaterial zur Verfügung steht, ein nicht unerheblicher Vorteil ist.

Die Vorzüge von NMR-Plattformen liegen hingegen in der einfachen Probenaufbereitung und der Robustheit der Messungen. MS reagiert äußerst empfindlich auf unterschiedlichen Probenaufbereitung und schwankende Umweltbedingungen, z.B. Temperatur und Luftfeuchtigkeit. Dies erschwert erheblich die Vergleichbarkeit von Mes-

sungen, die zu verschiedenen Zeitpunkten oder in unterschiedlichen Laboren durchgeführt wurden, selbst wenn dieselbe MS-Plattform verwendet wurde. NMR-Messungen sind hingegen weitestgehend unempfindlich gegenüber äußeren Einflüssen, wodurch eine exzellente Reproduzierbarkeit von quantitativen Messungen gegeben ist. So zeigten Viant *et al.* und Ward *et al.*, dass die Stabilität von NMR bis zu einem gewissen Grad unabhängig vom Hersteller des NMR-Spektroskops und dem Laborprotokoll ist [48, 49]. Außerdem ermöglicht NMR eine einfache absolute Quantifizierung der Metabolitenkonzentrationen, was bei MS nur mit großem Aufwand durch die Zugabe von Referenzsubstanzen zur Probe möglich ist. Und schließlich arbeitet NMR nicht-destruktiv, wodurch Proben nach der Messung weiterverwendet werden können.

Dem Anschein nach spielen die Vorzüge, die NMR-Plattformen bieten, bei mGWAS eine untergeordnete Rolle, und so wurden in bis dato veröffentlichten mGWAS überwiegend massenspektrometrisch bestimmte metabolische Merkmale untersucht (z.B. in [30, 34, 35, 39, 42, 50]). Ausnahmen bilden einige wenige Studien, beispielsweise von Nicholson *et al.* [40] sowie Kettunen *et al.* [38], in denen NMR-Spektroskopie eingesetzt wurde. Die höhere Sensitivität, die MS bietet, erlaubt es, potentiell eine höhere Anzahl unterschiedlicher GIMs innerhalb einer Kohorte zu identifizieren. Auch das Problem des eingeschränkten Transfers von Messergebnissen ist innerhalb einer Kohorte nicht wirklich gravierend: alle Proben, die für eine mGWAS herangezogen werden, werden zumeist auf einer (einheitlichen) metabolischen Plattform gemessen, und somit ist eine akzeptable Vergleichbarkeit zumindest innerhalb der Studienpopulation gegeben. Doch sobald metabolische Messungen, die in einzelnen Studienpopulationen durchgeführt wurden, in kohortenübergreifenden Studien untersucht werden sollen, ist die Vergleichbarkeit ein wichtiges Kriterium.

Auch ist es ein erklärtes Ziel von mGWAS, Marker für klinisch relevante Phänotypen zu identifizieren. Um metabolische Parameter in der alltäglichen Diagnostik zu etablieren, bedarf es robuster Messmethoden, die auch beim Blick auf das metabolische Profil eines einzelnen Patienten vergleichbare Messwerte liefern, welche als Grundlage zur Unterscheidung von gesunden und krankem Zustand dienen. Hier eignen sich NMR-Plattformen dank der hohen Stabilität der Messungen und der damit einhergehenden Vergleichbarkeit sehr gut für die routinemäßige Anwendung im klinischen Bereich.

3.4. Ziele dieser Arbeit

Der Schwerpunkt dieser Dissertation liegt in der Durchführung und Interpretation genomweiter Assoziationsstudien mit metabolischen Merkmalen, welche mittels NMR-Spektroskopie in Proben der SHIP-0- sowie der KORA-F4-Populationsstudien ermittelt wurden.

Es werden zwei genomweite Assoziationsstudien behandelt, die den Blick weg von der üblicherweise in mGWAS betrachteten Homöostase von Stoffwechselprodukten im Blut hin zu metabolischen Merkmalen richten, die mit dem Urin ausgeschieden werden. Seit altersher dient der menschliche Harn der Diagnose von Krankheiten: bereits vor über 3.000 Jahren wurde die Beobachtung gemacht, dass süß schmeckender Urin ein Indikator für eine Krankheit ist, die von dem griechischen Arzt Aretaios als *διαβήτης* (diabetes; griechisch für „Durchfluss“) bezeichnet wurde [51]. Seit dem Mittelalter war die Uroskopie ein wichtiger Bestandteil der alltäglichen Diagnostik und wurde dementsprechend systematisiert: in Schautafeln (sog. Uroskopie-Rädern) wurden Farbe, Geruch sowie Geschmack des Urins verschiedenen Körperleiden zugeordnet [52]. Noch heute spielt der Harn im medizinischen Kontext eine große Rolle, um beispielsweise Nieren- oder Harnwegeerkrankungen zu untersuchen. Desweiteren kommt der Urinausscheidung neben der Regulation des Flüssigkeitshaushaltes eine entscheidende Rolle bei der Entgiftung des Körpers zu. Als „biologisches Abfallprodukt“ enthält er unter anderem die Abbauprodukte von Nahrung, Medikamenten und (schädlichen) Umweltstoffen [6].

In den besprochenen Studien sollen genetisch beeinflusste Metabotypen im Urin identifiziert werden, die im Zusammenhang mit der Entgiftung des Körpers sowie mit chronischen Nierenerkrankungen stehen. Es soll weiterhin geprüft werden, ob die identifizierten GIMs allgemeingültig sind, also in einer unabhängigen Kohorte repliziert werden können. Die erste Studie war zum Zeitpunkt ihrer Veröffentlichung die erste GWAS, in der in einer großen Studienpopulation eine breite Palette metabolischer Merkmale im Urin untersucht wurden. Als Metabolomik-Datenbasis dienen manuell annotierte NMR-Spektren von Urinproben 862 männlicher Teilnehmer der SHIP-0-Kohorte. Die Ergebnisse wurden im Juni 2011 von Suhre *et al.* in *Nature Genetics* publiziert [53]. In einer zweiten, bislang nicht veröffentlichten mGWAS im Urin wurden die annotierten NMR-Spektren von insgesamt 3.861 männlichen und weiblichen SHIP-0-Probanden untersucht. Die Fragestellung bei dieser Studie ist, ob sich in dieser über viermal so großen Studienpopulation weitere genetisch beeinflusste

Metabotypen finden lassen, die in der Vorgängerstudie noch nicht entdeckt wurden. Zudem soll untersucht werden, inwiefern im Urin identifizierte GIMs auch im Blut sichtbar sind, weshalb die Assoziationsergebnisse im Lichte publizierter mGWAS im Serum der KORA-F4- [35] sowie der TwinsUK-Kohorte [43] betrachtet werden.

Die dritte in dieser Arbeit vorgestellte Studie hat die Identifikation genetisch beeinflusster NMR-Merkmale in NMR-Spektren von Blutplasmaproben der KORA-F4-Probanden zum Ziel. Hier soll untersucht werden, inwiefern sich in den Rohdaten, die NMR-Plattformen liefern, spektrale Merkmale identifizieren lassen, die mit genetischen Varianten assoziieren. Üblicherweise erfolgt bei NMR-basierten mGWAS eine Vorprozessierung der Spektren, bei der bestimmte Signalanteile extrahiert werden (so geschehen in den Studien von Nicholson *et al.* [40] oder Kettunen *et al.* [38]), oder es erfolgt eine Annotation der Spektren, bei der die NMR-Signale auf die Konzentrationen der darunterliegenden Metaboliten reduziert werden (wie in den in dieser Dissertation behandelten mGWAS im Urin). In dieser Studie wird jedoch einem hypothesenfreien Ansatz folgend nahezu der gesamte Spektralbereich auf genetische Assoziationen untersucht. Zudem soll gezeigt werden, dass sich das in MS-basierten GWAS als erfolgreich erwiesene Ratios-Konzept auch auf NMR-Spektraldaten übertragen lässt. Außerdem wird untersucht, wie sich NMR-Merkmale mit Hilfe von Daten aus anderen, hochsensitiven Metabolomik-Plattformen interpretieren lassen. Und schließlich soll die Frage beantwortet werden, ob sich die mittels GWAS identifizierten NMR-Merkmale als Marker für klinische Phänotypen eignen. Die Ergebnisse der Studie wurden im Februar 2013 von Raffler *et al.* in *BMC Genome Medicine* veröffentlicht [54].

4. Materialien und Methoden

4.1. Studienpopulationen

SHIP-0 und SHIP-1

Die SHIP-Studie (*Study of Health in Pomerania*) ist eine longitudinale Populationsstudie, die in der Region Greifswald in Mecklenburg-Vorpommern durchgeführt wird. Von den ungefähr 210.000 Bewohnern dieser Region wurden 6.265 erwachsene Personen im Alter von 20 bis 79 Jahren schriftlich eingeladen, an der Basisuntersuchung, der SHIP-0-Studie, teilzunehmen. Von den Angeschriebenen nahmen zwischen 1997 und 2001 letztlich 4.308 Probanden (2.193 Frauen und 2.115 Männer) an der SHIP-0-Studie teil. 3.300 dieser Probanden kamen zu den Nachuntersuchungen, die im Rahmen der SHIP-1-Studie zwischen 2002 und 2006 durchgeführt wurden [55]. Im Rahmen der Untersuchungen wurden u.a. Urinproben der Studienteilnehmer gesammelt. Weiterhin wurde jeder Proband genotypisiert.

Für die erste mGWAS im Urin wurden für die Discovery-, Replikations- und Verifikationsstudien jeweils unterschiedliche Untermengen der SHIP-0- und SHIP-1 Kohorten verwendet:

SHIP-0 males Für die Discovery-Studie wurden aus dem SHIP-0-Datensatz alle männlichen Probanden ausgewählt, deren Urinproben vor 11 Uhr gesammelt wurden. Hierbei ist zu beachten, dass es sich trotz der vormittäglichen Urinabgabe nicht wie in anderen klinischen Studien üblich um den ersten Morgenurin handelt. Insgesamt umfasst der Datensatz Metabotyp- und Genotypdaten von 922 Teilnehmern.

SHIP-0 females Für diesen zur Replikation verwendeten Datensatz wurden weibliche SHIP-0-Probanden ausgewählt. Wie im Discovery-Datensatz wurden nur Urinproben selektiert, die vor 11 Uhr gesammelt wurden. Dieser Datensatz umfasst 870 Studienteilnehmerinnen.

SHIP-0 males late Dieser Datensatz beinhaltet die Proben 126 männlicher SHIP-0-Teilnehmer und wurde zur Replikation herangezogen. Im Gegensatz zum Discovery-Datensatz wurden die Proben nach 11 Uhr vormittags gesammelt. Ein zusätzliches Kriterium zur Auswahl der Probanden erfolgte anhand des Genotyps. Diese für GWAS bislang unübliche Vorgehensweise ist angelehnt an klinische Fallstudien: die Probanden mit einer seltenen Krankheit (Fälle) werden einer vergleichbar großen Gruppe gesunder Probanden (Kontrollen) gegenübergestellt. Im hier verwendeten Datensatz besteht die Fallgruppe aus den homozygoten Trägern des seltenen Allels des zu replizierenden SNPs und die Kontrollgruppe aus den heterozygoten Trägern sowie den homozygoten Trägern des häufigen Allels, welche wiederum homozygote Träger des seltenen Allels eines der vier anderen zu replizierenden SNPs sind. Dieser Ansatz erlaubt es, die Teststärke der Replikationsstudie zu erhöhen, während gleichzeitig die Anzahl der zu annotierenden NMR-Spektren niedrig gehalten wird.

SHIP-1 verification Durch eine longitudinale Verifikationsstudie sollte gezeigt werden, dass die fünf identifizierten Assoziationen zwischen genetischen Loci und metabolischen Phänotypen über einen längeren Zeitraum hinweg stabil sind. Dazu wurden aus SHIP-1 170 männliche Teilnehmer ausgewählt, die analog zum *SHIP-0 males late*-Datensatz bezüglich genau eines der fünf SNPs homozygoter Träger des seltenen Allels sind.

In der Nachfolgestudie, also der zweiten in dieser Arbeit besprochenen mGWAS im Urin, wurde nahezu die gesamte SHIP-0 Kohorte untersucht: es lagen Genotyp- und Metabotyp-Daten für 4.105 Probanden, darunter 2.083 Frauen und 2.023 Männer, vor.

KORA F4

Die KORA-Studie (Kooperative Gesundheitsforschung in der Region Augsburg) ist eine unabhängige Populationsstudie, an der tausende zufällig ausgewählte, im Raum

Augsburg ansässige Personen teilnehmen. Im Rahmen der vierten Phase dieser Studie, KORA S4, wurden zwischen 1999 und 2001 insgesamt 4.261 Probanden (2.090 Männer und 2171 Frauen) im Alter zwischen 25 und 74 Jahren untersucht [56, 57]. Von den Probanden der S4-Studie nahmen 72,3%, also 3.080 Personen, an einer zwischen 2006 und 2008 durchgeführten Nachfolgestudie, KORA F4, teil [58]. Die bei den Untersuchungen der KORA-F4-Studie gesammelten Urinproben sowie die Genotyp-Daten bilden die Basis für den zur Replikation verwendeten Datensatz. Im Gegensatz zur SHIP-Studie waren die Teilnehmer der KORA-Studie angehalten, bis zur Abgabe ihrer Urinproben nüchtern zu bleiben, um die Variabilität der Metabolitenkonzentrationen möglichst gering zu halten.

Im Rahmen der ersten, 2011 veröffentlichten mGWAS im Urin wurden zur Replikation der Ergebnisse zufällig 992 Probanden beider Geschlechter ausgewählt, die genotypisiert waren und von deren Urinproben NMR-Spektren verfügbar waren. Für die zweite mGWAS im Urin umfasste der verwendete Replikations-Datensatz 1.814 Probanden.

Für die Durchführung der GWAS mit NMR-Merkmalen im Blutplasma standen für die gemeinsame Analyse von Genotyp und Metabotyp die Daten von 1.757 Probanden der KORA-F4-Studie zur Verfügung.

4.2. Genotyp-Daten

SHIP-0 und SHIP-1

Die Probanden aus SHIP wurden mittels Genchips vom Typ Human SNP Array 6.0 genotypisiert. Die Hybridisierung der DNA erfolgte gemäß der Vorgaben des Herstellers Affymetrix. Die Genotypen wurden mit Hilfe des Birdseed2-Algorithmus' bestimmt. Zur Qualitätskontrolle wurden zusätzlich zu den eigentlichen Proben Kontrollproben genotypisiert. Die Genchips mussten bezogen auf die Kontrollproben eine Erkennungsrate (*call rate*) von mindestens 86% aufweisen, um den Qualitätsansprüchen zu genügen. Alle Probanden mussten zu mindestens 92% genotypisiert sein, wobei die durchschnittliche Genotypisierungsrate bei 98,55% lag. Insgesamt wurden bis zu 909.508 genetische Marker (SNPs) je Proband genotypisiert. Diese Genotyp-

Daten wurden von der Ernst-Moritz-Arndt-Universität Greifswald zur Verfügung gestellt.

Um die Datenqualität zu erhöhen und somit zu vermeiden, dass es zu falschen Assoziationen kommt, wurden in der ersten mGWAS im Urin nur SNPs betrachtet, deren Genotypisierungsrate mindestens 95 % beträgt (Ausschluss von 55.876 SNPs) und die das Hardy-Weinberg-Equilibrium nicht verletzen ($p_{\text{HWE}} \geq 0,001$; Ausschluss von 5.856 SNPs). Zudem wurden nur SNPs analysiert, deren seltenes Allel mit einer Häufigkeit (*minor allele frequency*) von mindestens 5 % vorlag, wodurch insgesamt 228.303 SNPs ausgeschlossen wurden. Insgesamt erfüllen 645.249 autosomale SNPs die genannten Qualitätskriterien.

In der zweiten mGWAS im Urin wurden dieselben genotypisierten Marker untersucht, allerdings wurden in Bezug auf das Hardy-Weinberg-Equilibrium moderatere Filterkriterien implementiert. Von den ursprünglichen 909.508 genotypisierten Varianten wurden 8.623 SNPs wegen Verletzung des Hardy-Weinberg-Gleichgewichtes herausgefiltert ($p_{\text{HWE}} < 1,0 \times 10^{-6}$). 57.160 weitere SNPs wurden aufgrund zu niedriger Genotypisierungsrate ($< 95\%$) aus den Datensätzen entfernt, ebenso weitere 227.967 SNPs, bei denen das seltene Allel mit einer Häufigkeit von unter 5 % vorlag (MAF $< 5\%$). Es verblieben schließlich 643.953 autosomale und gonosomale SNPs. Für die GWAS wurden nur die 620.456 autosomalen SNPs betrachtet. Bei einem abschließenden Test wurden aus dem SHIP-Datensatz 18 Probanden entfernt, bei denen Unsicherheit bezüglich der Übereinstimmung des anhand der genetischen Daten ermittelte Geschlechts und des durch die Metadaten gegebene Geschlecht bestand.

Zur Interpretation des in der zweiten mGWAS im Urin identifizierten NAT2-Locus wurde auf imputierte genetische Daten des SHIP-0-Genotypdatensatzes zurückgegriffen. Das *pre-phasing* wurde mit der frei verfügbaren Software „SHAPEIT“ (Version 1) [59] durchgeführt. Die Imputation der Genotypen erfolgte mittels „IMPUTE2“ (Version 2.2.2) anhand des Haplotyp-Referenzdaten des *1000-Genomes-Project-Konsortiums* (*Phase 1 integrated data*, Version 3) [60, 61].

KORA F4

Analog zur SHIP-0-Kohorte wurden die Probanden der KORA-F4-Kohorte durch „Human SNP Array 6.0“ Chips genotypisiert bei einer auf die einzelnen Probanden

bezogenen Genotypisierungsrate von über 93 % (99 % über alle Probanden). Die Genotypen der SNPs wurden ebenfalls mithilfe des Birdseed2-Clustering-Algorithmus' bestimmt. Die Genotyp-Daten wurden vom Institut für Genetische Epidemiologie des Helmholtz Zentrums München bereit gestellt.

Für die Replikation der ersten mGWAS im Urin sowie die GWAS mit NMR-Merkmalen im Plasma enthielt der Genotyp-Datensatz ausschließlich SNPs, die zu mindestens 95 % genotypisiert wurden, die das Hardy-Weinberg-Equilibrium nicht verletzen ($p_{HWE} \geq 1,0 \times 10^{-6}$) und bei denen das seltene Allel mit einer Häufigkeit von mindestens 1 % vorkommt. Insgesamt erfüllen 655.658 autosomale SNPs alle drei Kriterien.

Für die Replikation der zweiten mGWAS im Urin wurden die Filterkriterien entsprechend dem SHIP-0-Datensatz angepasst: von den 906.716 genotypisierten SNPs wurden 32.033 SNPs wegen Verletzung des Hardy-Weinberg-Equilibriums entfernt ($p_{HWE} < 1,0 \times 10^{-6}$). 84.351 SNPs wurden aufgrund zu niedriger Genotypisierungsrate ($< 95\%$) aus dem Datensatz gelöscht, ebenso 224.723 SNPs mit einer MAF kleiner 5 %. Von den verbliebenen 617.068 autosomalen und gonosomalen SNPs wurden nur die 593.830 autosomalen SNPs betrachtet. Ein Proband wurde aus dem Datensatz aufgrund bestehender Unsicherheiten bezüglich des Geschlechts entfernt.

4.3. Metabolomik-Daten

Urin

Probensammlung, Probenvorbereitung und Lagerung

Bei den von den SHIP-Probanden abgegebenen Urinproben wurde vermerkt, ob die Probe zwischen 8 und 11 Uhr vormittags oder zu einem späteren Zeitpunkt gesammelt wurde. Dabei ist zu beachten, dass die Studienteilnehmer nicht dazu angehalten waren, nüchtern zur Untersuchung zu erscheinen. Im Gegensatz dazu waren alle Probanden der KORA-F4-Studie zum Zeitpunkt der morgendlichen Untersuchung und der Abgabe der Urinprobe nüchtern. Allerdings wurde nicht wie in medizinischen Untersuchungen oft üblich der Morgenurin der ersten Blasenentleerung des Tages gesammelt.

Bis zur endgültigen Analyse wurden alle Urinproben bei -80°C gelagert. Als Vorbereitung für die NMR-Spektralanalyse wurden $675\ \mu\text{l}$ Urin mit $75\ \mu\text{l}$ Phosphatpuffer versetzt, um einen pH-Wert von $7,0 \pm 0,35$ einzustellen. Der Puffer wurde auf D_2O -Basis präpariert und enthielt als Referenzsubstanz $0,5\ \text{mmol}$ Natrium-3-(trimethylsilyl)propionat-2,2,3,3- d_4 (TSP).

Aufzeichnen der NMR Spektren

Für jede Urinprobe wurden an der Universität Greifswald eindimensionale Protonen-Kernresonanzspektren ($1\text{D-}^1\text{H NMR}$) in einem Bruker DRX-400 NMR Spektrometer gemessen. Das Gerät arbeitet bei einer Frequenz von $400,13\ \text{MHz}$; gemessen wurde bei einer Temperatur von $300\ \text{K}$. Es wurde folgende $^1\text{H-NMR}$ -Pulssequenz verwendet: *Relaxation Delay* für $4\ \text{s}$ – *Non-selective Hard Pulse* (90°) für $9,4\ \mu\text{s}$ – $4\ \mu\text{s}$ – 90° -Puls für $9,4\ \mu\text{s}$ – *Mixing Time* für $100\ \text{ms}$ – 90° -Puls für $9,4\ \mu\text{s}$ – Aufzeichnen des *Free Induction Decay* (FID) für $3,96\ \text{s}$. Diese Pulssequenz dauerte also ungefähr acht Sekunden. Um das Wassersignal zu unterdrücken, diente während der *Relaxation Delay* und der *Mixing Time* ein schwacher Puls mit einer Frequenz von $25\ \text{Hz}$ zur Vorsättigung (*presaturation*).

Von jeder Probe wurden nach vier Probescans 32 FIDs aufgezeichnet, die in 32.768 Datenpunkten bei einer Spektralbreite von $20,869\ \text{ppm}$ gespeichert wurden. Um durch das Instrument bedingtes Rauschen zu verringern, wurde auf die FIDs eine Linienverbreiterung (*line broadening*) von $0,3\ \text{Hz}$ angewandt.

Annotation der NMR Spektren

Die Annotation der Spektren erfolgte mit der Chenomx NMR Suite (Version 6.1 und 7.0) der Chenomx Inc. aus Edmonton, Kanada. Die Annotation erlaubt es, aus den NMR-Spektren die absoluten Konzentrationen eines Teils der in den Proben vorhandenen Metaboliten abzuleiten. Da wechselnde Parameter wie der pH-Wert der Urinproben und Verdünnungseffekte die NMR-Signale stark beeinflussen, wurden die Spektren manuell vorprozessiert und annotiert.

Zur Vorprozessierung wurden die FIDs im Chenomx Processor Fourier-transformiert. Es erfolgte anschließend die semi-automatische Grundlinienkorrektur (*baseline cor-*

rection) sowie - falls notwendig - eine manuelle Phasenkorrektur der Signale (*phasing*). Schließlich wurde dem TSP-Referenzsignal der *Chemical Shift* $\delta = -0,016$ ppm zugewiesen und semi-automatisch die Fläche unter dem TSP-Signal bestimmt. Dadurch wurde es möglich, NMR-Signale anhand ihres *Chemical Shifts* bezogen auf den TSP-Peak zu identifizieren und aufgrund der bekannten TSP-Konzentration (0,5 mmol) zu quantifizieren. Die Annotation der NMR-Signale erfolgte im Chenomx Profiler. Hier wurden die NMR-Spektren teilweise dekonvolviert, also in diejenigen Signalbestandteile zerlegt, welche mit den Metaboliten einer Urinprobe korrespondieren. Hierzu bietet der Profiler für hunderte unterschiedlicher Substanzen an den pH-Wert sowie die Aufnahmefrequenz des Spektrometers angepasste „Idealspektren“, welche manuell über die gemessenen Spektren gelegt werden können. Durch Anpassen des Chemical Shifts einzelner Peaks dieser Idealspektren sowie deren Signalintensität erfolgt die Identifizierung sowie die Quantifizierung der Metaboliten einer Probe [62]. Überlappende Signale, wie sie in NMR-Spektren häufig vorkommen, sind mit dieser Methode jedoch nur schwer zu annotieren.

Die Annotation der NMR-Spektren des in der ersten mGWAS im Urin verwendeten Discovery-Datensatzes (*SHIP-0 males*) erfolgte in Auftragsarbeit durch Chenomx Inc. In diesen Spektren wurden die Metabolitenkonzentrationen aus einem 63 Substanzen (einschließlich Kreatinin) umfassenden Katalog bestimmt. Die zur Replikation und Verifikation herangezogenen Datensätze wurden hingegen durch den Autor dieser Dissertation annotiert. Da die manuelle Annotation sehr zeitaufwändig ist, wurden in diesen Datensätzen nur diejenigen Metaboliten quantifiziert, die bei den fünf in der Discovery-Studie identifizierten genetischen Loci die jeweils stärkste Assoziation zeigten. Bestimmt wurden also in jeder Probe, soweit möglich, die Konzentrationen von alpha-Hydroxyisobutyrat, beta-Aminoisobutyrat, Alanin, Formiat, Lysin, N,N-Dimethylglycin, Succinat, Valin sowie Kreatinin. Insgesamt wurden über 2.200 Spektren manuell präprozessiert und annotiert.

Für die Nachfolgestudie wurden die NMR-Spektren aller verfügbaren Urinproben der SHIP-0- und KORA-F4-Studien durch Chenomx Inc. annotiert, wobei eine durchgehende Neuannotation des SHIP-0-Datensatzes erfolgte, um Batch-Effekte durch unterschiedliche Annotation zu vermeiden. Für jede Probe wurden die Metabolitenkonzentrationen eines 60 Substanzen (Kreatinin eingeschlossen) umfassenden Panels bestimmt.

Normalisierung der Metabolitenkonzentrationen auf Kreatinin

Im Gegensatz zu Blutproben sind Urinproben unterschiedlich stark verdünnt, wodurch die Metabolitenkonzentrationen zwischen den Probanden nicht unmittelbar vergleichbar sind. Zur Kompensation von Verdünnungseffekten wurden alle annotierten Konzentrationen in Bezug zur ebenfalls durch Annotation bestimmten Kreatininkonzentration gesetzt (mmol/mol Kreatinin). Kreatinin ist ein Abbauprodukt des muskulären Kreatins und wird bei gesunden Menschen mit einer nahezu konstanten Tagesrate von 1 g pro 20 kg Muskelmasse mit dem Harn ausgeschieden [63]. Deshalb ist es bei metabolischen Studien gängige (aber nicht unumstrittene Praxis) im Urin gelöste Metaboliten bezüglich der Kreatininkonzentration zu normalisieren [64].

Wegen der großen Bedeutung der Kreatininkonzentration zur Kompensation von Verdünnungseffekten erfolgte bei den Proben des *SHIP-0 males* Datensatzes (Discovery-Datensatz der ersten mGWAS im Urin) zusätzlich eine biochemische Bestimmung dieses Metaboliten durch ein Siemens Dimension RXL System. Bei 12 der 922 annotierten NMR-Spektren betrug die Diskrepanz zu der per Annotation ermittelten Kreatinin-Konzentration über 10 mmol/l. Diese 12 Proben wurden aus dem Datensatz entfernt. Bei allen anderen Urin-Datensätzen entfiel die zusätzliche biochemische Bestimmung des Kreatiningehalts im Harn.

Datentransformation und Berechnung paarweiser Ratios

Alle Metabolitenkonzentrationen wurden vor der statistischen Analyse logarithmiert (dekadischer Logarithmus), da die Verteilung der \log_{10} -transformierten Datenpunkte in den meisten Fällen annähernd einer Normalverteilung entspricht. Zur Berechnung aller paarweisen Ratios zwischen Metaboliten wurden zunächst die untransformierten Konzentrationen herangezogen, bevor im Anschluss der Quotient dekadisch logarithmiert wurde: $\log_{10}(\text{Metabolit}_1/\text{Metabolit}_2)$. Durch das Logarithmieren wird zudem unerheblich, welcher Metabolit im Zähler bzw. Nenner des Bruchs steht, da gilt: $\log_{10}(\text{Metabolit}_1/\text{Metabolit}_2) = -\log_{10}(\text{Metabolit}_2/\text{Metabolit}_1)$. Somit wird die Anzahl der zu testenden Ratios halbiert [32].

Qualitätskontrolle

Um zu vermeiden, dass extreme Metabolitenkonzentrationen falsch-positive Assoziationen mit genetischen Polymorphismen hervorrufen, wurden in allen Datensätzen gleichermaßen Ausreißer gefiltert: alle normalisierten, logarithmierten Metabolitenkonzentrationen sowie Ratios, die mehr als die dreifache Standardabweichung vom über alle Proben berechneten Mittelwert abweichen, wurden entfernt. Weiterhin wurden in den Discovery-Studien der beiden GWAS im Urin nur Metaboliten bzw. paarweise Ratios mit mehr als 300 gültigen Datenpunkten als Phänotypen verwendet.

Blutplasma und Blutserum

Probensammlung, -vorbereitung und -lagerung

Im Rahmen der KORA-F4-Studie wurden Blutproben zur DNA-Extraktion und zur metabolischen Charakterisierung gesammelt. Um durch Biorhythmus und Ernährung der Testpersonen verursachte Schwankungen zu vermeiden, wurde den Probanden im nüchternen Zustand morgens zwischen 8.00 Uhr und 10.30 Uhr Blut entnommen. Ein Teil des Blutes wurde in Serum-Probenbehältnissen mit Gel-Trennhilfe gesammelt. Durch zweimaliges sanftes Kippen wurden die Proben gemischt und für 30 Minuten bei Zimmertemperatur (18° C bis 23° C) stehen gelassen, um eine vollständige Koagulation zu ermöglichen. Anschließend wurde das Probenmaterial für 10 Minuten bei 2.750 g und einer Temperatur von 15° C zentrifugiert. Das Blutserum wurde in Aliquote aufgeteilt und für längstens 6 Stunden bei 4° C zwischengelagert. Bis zur endgültigen Analyse wurden die Proben bei -80° C gelagert. Ein weiterer Teil des entnommenen Blutes wurde in mit Ethylendiamintetraacetat (EDTA) beschichteten Röhrchen gesammelt. Nach Mischen des Inhaltes durch zweimaliges sanftes Kippen wurden die Proben für höchstens fünf Minuten auf dem Sarstedt Universal Mixer belassen, um eine mechanische Hämolyse zu verhindern. Anschließend wurden die Proben für 10 Minuten bei 2.750 g und einer Temperatur von 15° C zentrifugiert. Danach wurde das Plasma abgetrennt und in Aliquots von jeweils 200 ml aufgeteilt. Bis zum Einfrieren bei -80° C betrug die Temperatur der Proben stets unter 4° C. Innerhalb von zwei Wochen erfolgte dann die endgültige Lagerung in der Gasphase flüssigen Stickstoffs bei -196° C.

Aufzeichnen der NMR Spektren und Vorprozessierung

Die ^1H -1D-NMR-Spektren von Plasmaproben wurden bei der numares GmbH (früher: LipoFIT), Regensburg, aufgezeichnet. Zur Messung diente ein Bruker Avance II+ 600 MHz Spektrometer. Jeweils $150\ \mu\text{l}$ der Plasmaproben wurden ohne weitere Vorverarbeitung mit einem Triple-Resonance-Inverse-Probenkopf bei einer Temperatur von 310 K gemessen. Da den Proben keine Referenzsubstanz wie TSP oder DSS hinzugefügt wurde, dienten die Laktat-Signale bei $\delta = 1,36\ \text{ppm}$ und $\delta = 1,37\ \text{ppm}$ als Referenzsignale zur horizontalen Ausrichtung der Spektren. Außerdem wurde seitens numares in einem semiautomatischen Verfahren die Anpassung der Phase (*Phasing*) sowie die Baseline-Korrektur vorgenommen.

Segmentierung der NMR Spektren

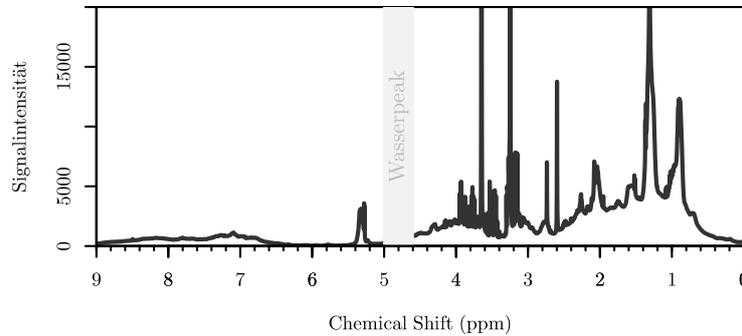
Die Spektren wurden von numares sowohl im binären FID-Format (entsprechend den Rohdaten der Messungen) als auch im Textformat zur Verfügung gestellt. Für das letztere Format wurde seitens numares die Spektren präprozessiert (inklusive Baseline- und Phasenkorrektur) sowie „gebint“, wobei die Spektren in Segmenten (*Bins*) gleicher Breite ($0,001\ \text{ppm}$) aufgeteilt werden (Abbildung 4.1b). Da ein Segment mehrere Datenpunkte des Originalspektrums beinhaltet, werden die Intensitätswerte innerhalb des Segments gemittelt. Für die weiteren im Rahmen dieser Studie durchgeführten Analysen wurden stets die segmentierten („gebinten“) Spektren verwendet.

Kompensation von Batch-Effekten

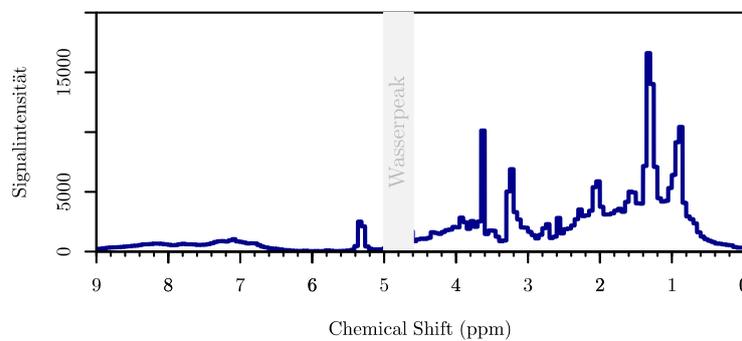
Eine zur visuellen Qualitätskontrolle durchgeführte Überlagerung der von numares gelieferten Spektren zeigte, dass die Vorprozessierung der Rohspektren, insbesondere die Baselinekorrektur, nicht einheitlich durchgeführt wurde. Abbildung 4.3a zeigt ein Überlagerungsspektrum von 100 zufällig ausgewählten Datensätzen. Daran ist ersichtlich, dass sich die Spektren bezüglich ihrer mittleren Signalintensität in zwei Gruppen einteilen lassen. Die anfängliche Vermutung, dass das Geschlecht der Probanden zu dieser Gruppenbildung beiträgt, hat sich nicht bestätigt, da die Geschlechter über die Proben zufällig verteilt sind. Die nach Messzeitpunkt bzw.

Abbildung 4.1.: Präprozessierung der NMR-Spektren

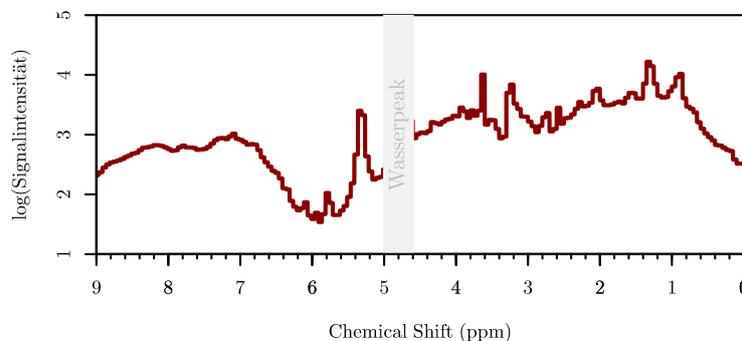
(a) Zuschnitt der Spektren und Ausblenden der Wasserregion



(b) Aufteilen des Spektrums in gleich große Segmente (Bins)



(c) Logarithmieren der Signalintensitäten



Die drei wesentlichen Vorverarbeitungsschritte, welche auf jedes NMR-Spektrum angewandt wurden. Abbildung a zeigt ein Spektrum, das auf dem Bereich zwischen $\delta = 0$ ppm und $\delta = 9$ ppm beschnitten wurde. Die Spektralregion zwischen $\delta = 4,6$ ppm und $\delta = 5,0$ ppm wird vom dominanten H_2O -Signal beeinflusst und wurde deshalb ausgeblendet. Als nächstes wird das NMR-Spektrum in gleich große Segmente (*bins*) aufgeteilt (Abbildung b; zur Verdeutlichung haben die abgebildeten Bins eine Breite von 0,05 ppm, während für die Studie tatsächlich 0,001 ppm breite Segmente verwendet wurden). Dadurch werden leichte Schwankungen (*jitter*) der NMR-Signale bezüglich ihres Chemical Shifts kompensiert. Schließlich werden die Intensitätswerte der NMR-Bins dekadisch logarithmiert (Abbildung c). Dadurch nähert sich die Verteilung der Signalintensitäten einer Normalverteilung an. Weiterhin werden durch diese Art der Datentransformation Signale mit hoher Intensität weniger stark gewichtet.

laufender ID sortierten Spektren lassen sich hingegen blockweise einer der beiden Gruppen zuordnen.

Es gelang, diesen Batch-Effekt nachträglich herauszurechnen. Dazu wurde die freie Statistik-Software R (Version 2.15.2) eingesetzt [65]. Zunächst diente eine Hauptkomponentenanalyse (*Principal Component Analysis*) über eine aus allen Spektren gebildete Datenmatrix (1.757 Proben \times 9.000 NMR-Bins (Bereich von $\delta = 0$ ppm - 9 ppm) (Zeilen \times Spalten)) dazu, einen Referenz-Bin zu ermitteln, der die durch den Batch-Effekt entstandenen Gruppen am besten separiert. Die Analyse wurde mittels der R-Funktion `prcomp` durchgeführt. Die Analyse lieferte 1.757 Hauptkomponenten (entsprechend der Anzahl der Samples), wobei die erste Hauptkomponente ungefähr 56 % der Gesamtvarianz in den Daten erklärt, gefolgt von der zweiten Hauptkomponente mit 16 % und der dritten mit 8 %. Der NMR-Bin bei $\delta = 2,643$ ppm lädt am stärksten auf die erste Hauptkomponente, weswegen diese Spektralposition als Referenz diente. Für die anschließende Batch-Effekt-Korrektur wurde für alle NMR-Bins jeweils ein lineares Modell errechnet mit den Intensitäten des zu korrigierenden Bins als abhängiger Variable und den Signalintensitäten des Referenz-Bins bei $\delta = 2,643$ ppm als unabhängige Variable: $NMR\text{-Bin} \sim Referenz\text{-NMR-Bin}$. Von den Signalintensitäten des zu korrigierenden Bins wurde nun das Produkt aus dem Regressionskoeffizienten β und der Differenz der Intensitäten des Referenz-Bins und dem Median über alle Intensitäten des Referenz-Bins abgezogen (siehe R-Code in Abbildung 4.2). Wie Abbildung 4.3b zeigt, konnte der Batch-Effekt dadurch wirksam kompensiert werden.

Zuschnitt der NMR-Spektren

Die Analyse der segmentierten Spektren beschränkte sich auf die NMR-Bins zwischen $\delta = 0,000$ ppm und $\delta = 9,000$ ppm. In den dadurch abgeschnittenen Randbereichen unterliegen die Baselines der Spektren messtechnisch bedingten Abweichungen, sodass es zu falschen Assoziationen mit Signalen in diesen Bereichen kommen kann. Zudem dient die Begrenzung dazu, den Umfang der Datenmenge zu reduzieren. (Abbildung 4.1a) Desweiteren wurde der Spektralbereich zwischen $\delta = 4,6$ ppm und $\delta = 5,0$ ppm ausgeblendet. Diese Region wird vom H₂O-Resonanzsignal dominiert, welches umliegende NutzsSignale stört und eine Interpretation dieser Signale unzuverlässig macht.

Abbildung 4.2.: Vereinfachte Darstellung des zur Korrektur des Batch-Effektes verwendeten R-Codes

```
# Datenmatrix nmr_data: Proben (Zeilen) × NMR-Bins (Spalten)
# Berechne PCA
pca_results <- prcomp(nmr_data, scale=TRUE, center=TRUE)
# Wähle den Referenz-Bin, der am stärksten auf die erste Hauptkomponente lädt
reference_chemshift <- which(pca_results$rotation[, 'PC1'] ==
max(pca_results$rotation[, 'PC1']))
# Intensitäten beim Referenzpunkt, hier  $\delta = 2,643$  ppm
reference_intensities <- nmr_data[reference_chemshift]
# Wiederhole die nächsten Schritte für alle Bins in der Datenmatrix nmr_data
# Intensitäten im zu korrigierenden Bin bei  $\delta = \text{bin\_chemshift}$  ppm
bin_intensities <- nmr_data[bin_chemshift]
# Erstelle das lineare Modell
lm_results <- lm(bin_intensities ~ reference_intensities)
# Speichere den Regressionskoeffizienten beta
beta <- summary(lm_results)$coefficients[2]
# Speichere korrigierten Bin in der Datenmatrix
nmr_data[bin_chemshift] <- (bin_intensities -
beta*(reference_intensities - median(reference_intensities)))
```

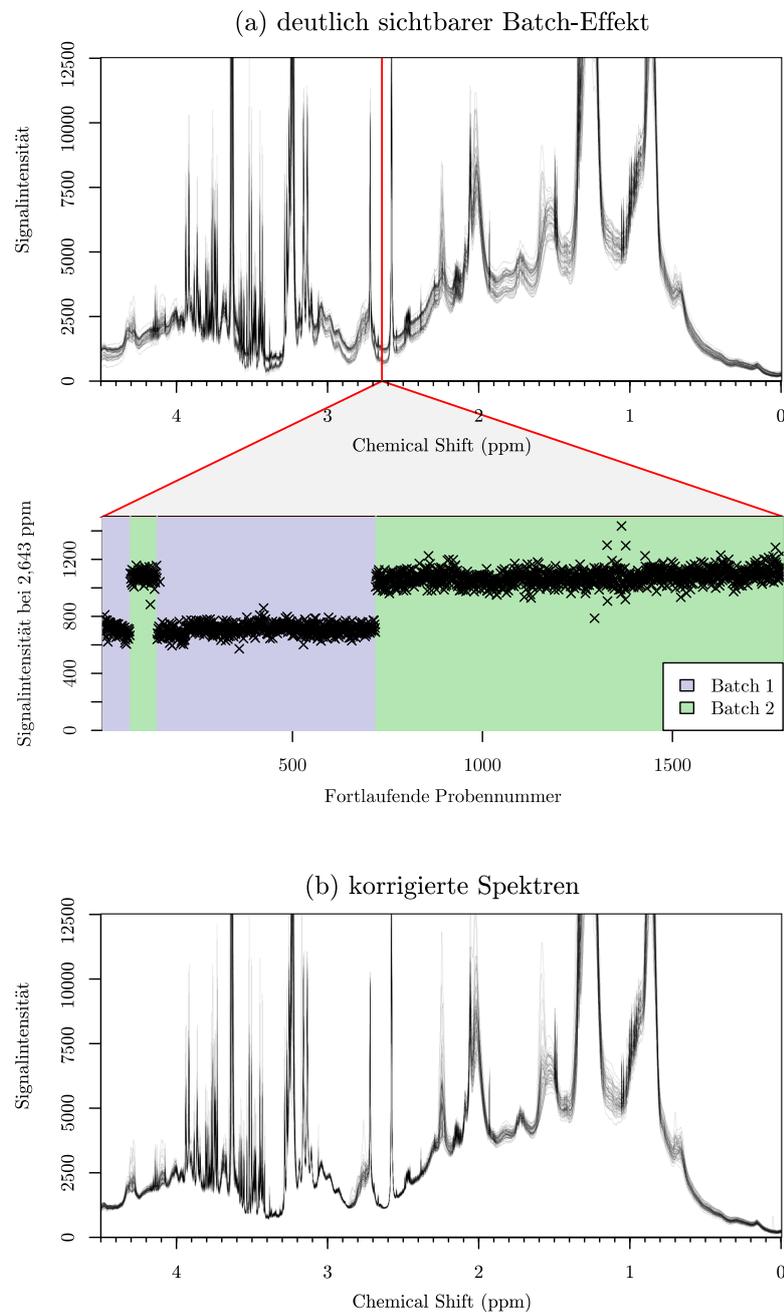
Datentransformation und Berechnung paarweiser Ratios

Da die Intensitätswerte innerhalb der Bins in der Regel nicht normalverteilt waren, wurden diese vor der statistischen Analyse dekadisch logarithmiert. Dadurch wird auch eine stärkere Gewichtung von Signalen mit geringer Intensität erreicht (Abbildung 4.1c) [66]. Analog zur oben beschriebenen Vorgehensweise wurden auch hier alle paarweisen, aus den untransformierten Bins gebildeten Ratios errechnet. Der Quotient wurde anschließend logarithmiert.

Qualitätskontrolle

Um den Effekt von Ausreißern zu verringern, wurden in einem NMR-Bin alle Intensitätswerte verworfen, die weiter als das Dreifache der Standardabweichung vom Mittelwert aller Werte innerhalb des Bins entfernt sind. Analog wurde mit NMR-Ratios verfahren.

Abbildung 4.3.: Korrektur eines Batch-Effektes, der die Probanden in zwei Gruppen unterteilt



Der obere Plot (a) zeigt die Überlagerung der NMR-Spektren von 100 zufällig ausgewählten Proben. Es ist deutlich sichtbar, dass in einem Teil der Messungen die Signalintensität durchwegs stärker ist. Dies zeigt sich im Plot an den zwei parallel verlaufenden Grundlinien. Am deutlichsten sichtbar ist dieser Effekt am NMR-Bin bei $\delta = 2,643$ ppm (rote Linie). Der mittlere Plot zeigt auf der Y-Achse die Signalstärken dieses Bins; die X-Achse zeigt die fortlaufenden Probennummern. Die deutliche, blockweise Zuordenbarkeit der Spektren in eine der zwei Gruppen (der *batch*-Effekt) rührt mutmaßlich von einer unterschiedlichen semiautomatischen Präprozessierung her. Der untere Plot (b) zeigt abermals 100 zufällig ausgewählte Spektren, bei denen eine Korrektur für diesen Effekt erfolgte.

Weitergehende Metabolomik-Messungen

Die in KORA F4 gesammelten Serum- und Plasmaproben wurden bereits im Rahmen früherer Studien auf unterschiedlichen Metabolomik-Plattformen charakterisiert. Weitergehende Informationen zu den Messmethoden und den Datensätzen sind jeweils in den am Ende referenzierten Publikationen zu finden.

Klinische Biochemie Folgende Serum-Lipide wurden in frisch entnommenen Proben auf einem Dimension RxL (Dade Behring) gemessen (siehe Petersen *et al.* [67]):

- die Gesamtmenge an Cholesterin (*total cholesterol*; TC) mittels Cholesterin-Esterase-Methode (CHOL Flex, Dade-Behring, CHOD-PAP Methode)
- HDL-Cholesterin (HDL-C) mittels AHDL Flex (Dade-Behring, CHOD-PAP Methode nach selektivem Freilassen von HDL-C)
- LDL-Cholesterin (LDL-C) mittels ALDL Flex (Dade Behring, CHOD-PAP Methode nach farblosem Verbrauch allen LDL-C)
- Triglyceride (TG) mittels TGL Flex (Dade Behring, enzymatischer, kolorimetrischer Test, GPO-PAP Methode)

Quantitative Massenspektrometrie Zur absoluten Quantifizierung eines definierten Metabolitensatzes wurde das AbsoluteIDQ-p150-Kit der Biocrates Life Sciences AG (Innsbruck, Österreich) verwendet. Die Analyse der Serumproben erfolgte auf einem API 4000 QTRAP LC/MS/MS System der Firma Applied Biosystems, welches mit einer Shimadzu Prominence LC20AD Pumpe sowie einem SIL-20AC Autosampler ausgestattet ist. Der analytische Prozess sowie die Auswertung der Rohdaten erfolgte durch die MetIQ-Software, die ein integraler Bestandteil des AbsoluteIDQ - Kits ist. Das Kit ermöglicht die Quantifizierung von bis zu 163 unterschiedlichen Metaboliten. Darunter sind 14 Aminosäuren, Hexose (H1), freies Carnitin (C0), 40 Acylcarnitine (Cx:y), Hydroxylacylcarnitine (C(OH)x:y) und Dicarboxylacylcarnitine (Cx:y-DC), 15 Sphingomyeline (SMx:y) und Hydroxylacyloylsphingosylphosphocholine (SM(OH)x:y), 77 Phosphatidylcholine (PC, aa = diacyl, ae = acyl-alkyl) und 15 Lysophosphatidylcholine. Die Zusammensetzung der Lipidseitenketten wird hierbei mit Cx:y abgekürzt. Dabei steht x für die Anzahl der Kohlenstoffatome in den Seitenketten und y für die Anzahl der vorhandenen Doppelbindungen. Details zur Messmethode und zum Datensatz finden sich in Illig *et al.* [34].

Semi-quantitative Massenspektrometrie Zur semi-quantitativen Bestimmung einer möglichst breiten Palette an Metaboliten wurde auf die analytische Plattform der Firma Metabolon Inc. (Durham, USA) zurückgegriffen. Metabolon übernahm dabei die Identifizierung und relative Quantifizierung der Metaboliten sowie die Auswertung und Qualitätskontrolle der Daten. Im analytischen Teil wurde jede Probe durch zwei UHPLC/MS/MS² (*ultrahigh-performance liquid chromatography / tandem MS*) und einer GC-MS (*gas chromatography / MS*) Methode gemessen. Die UHPLC-Injektionen wurden für basische und saure Substanzen optimiert. Die LC-MS-Analyse erfolgte durch eine Waters ACQUITY UPLC sowie ein Thermo-Finnigan LQT Massenspektrometer, das aus einer Elektrosprayionisationsquelle und einer linearen Ionenfalle als Massenanalysator bestand. Für den Teil der analytischen Plattform, der auf GC-MS basiert, wurden die Proben in einem hochauflösenden Thermo-Finnigan Trace DSQ Single-Quadrupol-Massenspektrometer vermessen. Die Annotation der resultierenden Spektren erfolgte durch Abgleich mit einer von Metabolon generierten Referenzdatenbank, welche für alle sich darin befindenden Moleküle die Retentionszeit, das Verhältnis von molekularer Masse zur Ladung (*mass-to-charge ratio*; m/z), bevorzugte Addukte, Fragmentierungsmuster sowie die zugeordneten MS/MS Spektren beinhaltet. Zur sicheren Identifikation eines Metaboliten werden mehrere dieser Parameter herangezogen. Insgesamt konnten in den KORA-F4-Proben 295 bekannte Metaboliten gemessen und identifiziert werden, hinzu kommen noch 220 Moleküle unbekannter chemischer Struktur (*unknowns*). Bei den bekannten Analyten handelt es sich unter anderem um Aminosäuren, Acylcarnitine, Sphingomyeline, Glycerophospholipide, Kohlenhydrate, Vitamine, Lipide, Nucleotide, Peptide, Steroide und Xenobiotika. Die Vermessung einer Probe dauerte 24 Minuten. Die mittlere Variabilität bei den ermittelten Konzentrationen über alle Metaboliten lag bei unter 12%. Details zur Analysemethode sowie zum dadurch generierten Datensatz finden sich in Evans *et al.* [68] und Suhre *et al.* [35].

Lipoprotein-Subklassen Die Spektren der Plasmaproben wurden bei der numares GmbH aufgezeichnet. Diese Spektren sind identisch mit den weiter oben besprochenen NMR-Spektren, die in der GWAS als Phänotypen herangezogen wurden. Durch ein patentiertes Verfahren werden aus dem Spektralbereich zwischen $\delta = 0,6$ ppm und $\delta = 1,5$ ppm insgesamt 15 Lipoprotein-Subfraktionen (L1 bis L15) abgeleitet [69]. Die nach Partikelgröße geordneten Fraktionen entsprechen dabei HDL (L1 – L4, entsprechend small, medium, large und very large), LDL (L5-L9, entsprechend very small, small, medium, large, very large), ILD (L10), VLDL (L11 und L12, entsprechend

small und large), Abbauprodukten (remnants; L13) und Chylomikronen (L14 und L15, entsprechend small und large). Details zu den Lipidsubklassen und dem Datensatz sind bei Petersen *et al.* zu finden [67].

4.4. Durchführung der Assoziationsstudien

Berechnung linearer Modelle

Die Tests auf Assoziation zwischen genetischen Polymorphismen und metabolischen Merkmalen (Metabolitenkonzentrationen sowie NMR-Bins und jeweils die paarweise davon abgeleiteten Ratios) wurden mittels linearer Modelle unter der Annahme eines additiven Effektmodells durchgeführt. Bei allen Datensätzen wurde bei der Berechnung der Modelle das Alter des jeweiligen Probanden als Kovariate einbezogen. Bei Datensätzen, in denen männliche und weibliche Studienteilnehmer gemischt vorlagen, floss außerdem das Geschlecht der Probanden als Kovariate in die Berechnung der Modelle ein.

Alle Assoziationstest wurden mit der quelloffenen Software PLINK (Version 1.07) durchgeführt [70]. Durch Kompilieren des PLINK-Quellcodes mit dem Intel C++ Compiler (Ver. 12.0.2.137) konnte die Ausführungsgeschwindigkeit gegenüber der mit dem GNU C Compiler vorkompilierten PLINK-Downloadversion um rund 10 % gesteigert werden. PLINK ist von Haus aus nicht multithreadingfähig, profitiert also nicht von CPUs mit mehr als einem Prozessorkern. Um diese Beschränkung zu umgehen, diente ein Perl-basierter Scheduler dazu, mehrere PLINK-Instanzen parallel auszuführen.

Die Berechnungen erfolgten für die Discovery- und Replikationsstudien der mGWAS im Urin sowie für die GWAS mit NMR-Bins lokal auf einer Workstation mit zwei Intel Xeon E5620 CPUs (2,4 GHz; je 4 Kerne plus 4 virtuelle Kerne) und 16 GB Arbeitsspeicher. Mit diesem System konnten pro Stunde ca. 20 metabolische Merkmale genomweit auf Assoziation überprüft werden. Im Falle der Assoziationsstudie mit NMR-Ratios ($n = 124.750$) war eine lokale Berechnung nicht praktikabel. Stattdessen wurde auf Rechenkapazität, die durch das EU-weite DEISA-Projekt (Distributed European Infrastructure for Supercomputing Applications) zur Verfügung gestellt [71] wurde, zurückgegriffen. Die Berechnungen erfolgten auf Großrechnern zweier

dem DEISA-Projekt angeschlossenen Rechenzentren: „Louhi“ des finnischen CSC (IT Center for Science) und „HECToR“ des schottischen EPCC (Edinburgh Parallel Computing Centre). Beide Rechner entspringen der XT-Modellreihe der Firma Cray (Louhi: XT4/XT5; HECToR: XT6), die aufgrund der verwendeten AMD Opteron CPUs auf x86-Rechnerarchitektur basieren. Somit konnte auf dieselben kompilierten PLINK-Binaries zurückgegriffen werden, die auch für die lokale Berechnungen verwendet wurden. Die Parallelisierung der PLINK-Instanzen erfolgte abermals durch einen Perl-Scheduler sowie durch das Batch-Job-System der Sun Grid Engine. Durch die hohe Anzahl simultaner Berechnungen (bis zu 400 parallele PLINK-Instanzen) konnten alle 124.750 Ratios innerhalb von vier Wochen getestet werden, wozu 95.000 Stunden CPU-Zeit benötigt wurden.

Identifikation signifikant assoziierter Loci

In der ersten mGWAS im Urin wurden im Discovery-Datensatz 1.720 metabolische Merkmale, davon 59 normalisierte Metabolitenkonzentrationen und 1.661 paarweise Konzentrationsverhältnisse (Ratios), auf genetische Assoziationen mit 645.249 SNPs getestet. Das Bonferroni-korrigierte Signifikanzniveau für Assoziationen liegt folglich bei $\alpha' = 0,05 / (645.249 \times 1.720) = 4,51 \times 10^{-11}$. Assoziationen mit $p \leq \alpha'$ wurden als genomweit signifikant betrachtet.

Für die zweite mGWAS im Urin gilt entsprechend der Anzahl untersuchter metabolischer Merkmale sowie genetischer Varianten ein genomweites, Bonferroni-korrigiertes Signifikanzniveau von $\alpha' = 0,05 / (620.456 \times 1.518) = 5,3 \times 10^{-11}$. Im Vergleich zu den anderen in dieser Arbeit vorgestellten GWAS galt für Assoziationen zu Metaboliten-Ratios ein zusätzliches Signifikanzkriterium. Hier musste der p-Gain größer als $1.518 / (2 \times 0,05) = 1,51 \times 10^4$ sein, wobei der p-Gain die Zunahme der Assoziationsstärke im Vergleich zu den p-Werten, die sich aus den Assoziationen des getesteten SNPs mit den einzelnen Metaboliten des Ratios ergeben, beschreibt. Die Berechnung dieser unteren p-Gain-Grenze beruht auf einer 2012 erschienenen Publikation von Petersen *et al.*, in der die statistischen Eigenschaften des p-Gains untersucht wurden. Laut Petersen *et al.* ist als konservative p-Gain-Untergrenze die Anzahl getesteter Metaboliten und Ratios geteilt durch 2α anzusetzen [32].

Die Identifikation genetischer Assoziationen zu NMR-Merkmalen (Bins und davon abgeleitete Ratios) im Blutplasma erfolgte in zwei Stufen (Abbildung 5.1). In der ers-

ten Stufe wurde eine GWAS mit 8.600 NMR-Bins und 655.658 SNPs durchgeführt. In den mGWAS im Urin sowie früheren mGWAS im KORA-Datensatz wurden stets alle möglichen paarweisen Ratios aus den gemessenen Metabolitenkonzentrationen auf genetische Assoziation getestet [30, 34, 35]. Auf die GWAS mit NMR-Merkmalen übertragen wäre es somit nötig gewesen, $3,7 \times 10^7$ NMR-Ratios zu testen (entsprechend $8600 \times 8599 / 2$ NMR-Bins), was in Anbetracht begrenzt zur Verfügung stehender Rechenzeit nicht machbar gewesen wäre. Folglich musste die Anzahl der Ratios auf ein sinnvolles Maß reduziert werden. Die Ergebnisse aus der im vorherigen Abschnitt beschriebenen GWAS mit einzelnen NMR-Bins dienten dazu, geeignete Bins zur Berechnung der Ratios zu finden. Es wurden die 500 Bins, welche am stärksten mit SNPs assoziieren, ausgewählt. In dem häufig auftretenden Fall, dass unmittelbar benachbarte Bins in ähnlicher Stärke mit dem gleichen SNP assoziieren, wurde lediglich jeweils der Bin mit dem stärksten Assoziationssignal selektiert. Dies geschah anhand einer Suche nach lokalen Minima in den Assoziationsspektren, welche den p-Wert in Abhängigkeit von der Position im Spektrum visualisieren (vergleiche Abbildung 6.19a). In einer zweiten GWAS wurden die Ratios dieser 500 Bins untersucht. Der angepasste p-Wert, unterhalb dessen die Assoziationen als genomweit signifikant betrachtet wurden, liegt bei $\alpha' = 0,05 / 655.658 / 133.350 = 5,72 \times 10^{-13}$. Dies entspricht der Korrektur für die unabhängigen Tests von 655.658 SNPs und 133.350 NMR-Merkmalen (Ratios und einzelne Bins) bei einem nominalen Signifikanzniveau von 5 %.

Für jede der in dieser Arbeit vorgestellten GWAS gilt, dass alle genomweit signifikant assoziierten SNPs innerhalb einer physischen Distanz von 1 Mb zu einem genetischen Locus zusammengefasst wurden.

Replikation

Zur Replikation und longitudinalen Verifikation der Ergebnisse der ersten mGWAS im Urin wurde von jedem der fünf Loci jeweils der SNP sowie das assoziierte metabolische Merkmal getestet, welches in der Discovery-Studie den niedrigsten p-Wert zeigte. Wiederum wurde PLINK verwendet, um lineare Modelle zu erstellen. Das Signifikanzniveau wurde für fünf unabhängige Tests angepasst ($\alpha' = 0,05 / 5 = 0,01$).

Auf vergleichbare Art wurde versucht, die Ergebnisse der zweiten mGWAS im Urin in der von SHIP-0 unabhängigen KORA-F4-Kohorte zu replizieren. Allerdings wur-

de in Fällen, bei denen die stärkste Assoziation eines Locus nicht replizierbar war, versucht, stattdessen die zweitstärkste signifikante Assoziation des jeweiligen SNPs zu replizieren, sofern diese noch genomweit signifikant war. Falls der zu testende SNP im Genotyp-Datensatz der KORA-F4-Studie nicht enthalten war, wurden stattdessen bis zu 5 weitere SNPs des KORA-Datensatzes getestet, die laut HapMap (Version 27, Phase I-III, CEU-Panel) im LD ($r^2 > 0,8$) sind. Insgesamt wurden 18 Tests durchgeführt, da bei vier der vierzehn Loci nur die zweitstärkste Assoziation repliziert werden konnte. Daraus ergibt sich ein Bonferroni-korrigiertes Signifikanzniveau $\alpha' = 0,05 / 18 = 0,0028$.

4.5. Ergebnisinterpretation

Neben der in der Diskussion der jeweiligen Loci angegebenen Literatur wurden mehrere Programme und Datenquellen verwendet, um die Ergebnisse der GWAS zu interpretieren.

Die Visualisierung der Assoziationen innerhalb eines Locus erfolgte durch Regionale Assoziationsplots. Hierzu wurde eine modifizierte Variante eines R-Scriptes verwendet, welches 2007 von Saxena *et al.* erstmals publiziert wurde [25]. Das ursprüngliche Script steht auf dieser Seite zum Download zur Verfügung: www.broadinstitute.org/diabetes/scandinav/figures.html. Die in den Plots abgebildete geschätzte Rekombinationsrate basiert auf Daten aus HapMap Release 22, abrufbar unter hapmap.ncbi.nlm.nih.gov/downloads/recombination/2008-03_re122_B36/ [72]. Die Genannotationen basieren auf Angaben aus UCSC Build 36.3 (hgdownload.cse.ucsc.edu/goldenPath/hg18/database/).

Angaben zum Kopplungsungleichgewicht (*linkage disequilibrium*; LD) stammen sofern im Text nicht anderweitig angegeben aus europäisch-stämmigen HapMap-Populationen (Release 22 bzw. Release 27, Phase I-III), abgerufen über den „SNAP“ Dienst des Broad Instituts (www.broadinstitute.org/mpg/snap) [73].

Zum Vergleich zwischen GIMs im Urin und im Blutserum wurden die Assoziationsergebnisse der von Suhre *et al.* 2011 veröffentlichten mGWAS verwendet [35]. Diese basieren auf semi-quantitativen metabolomischen Messungen im Serum der KORA-F4-Probanden (vergleiche Kapitel 4.3). Die Assoziationsdaten sind unter

der Internetadresse mips.helmholtz-muenchen.de/proj/GWAS frei verfügbar. Außerdem wurden die Ergebnisse einer 2014 von Shin *et al.* veröffentlichten GWAS, in der über 140 GIMs im Blutserum identifiziert wurden, herangezogen [43]. Diese Studie basiert auf der TwinsUK-Kohorte. Die TwinsUK-Studie ist eine Populationsstudie, die unter der Leitung des King's College in London durchgeführt wird. Dabei wurden innerhalb der letzten 20 Jahre über 7.000 ein- und zweieiige, überwiegend weibliche Zwillinge untersucht [74]. Assoziationsdaten aus GWAS mit metabolischen Markern im Blut der TwinsUK-Probanden wurden in der Vergangenheit bereits öfters herangezogen, um Ergebnisse von GWAS im KORA-F4-Datensatz wechselseitig zu replizieren [35]. Dabei ist es von Vorteil, dass die metabolischen Messungen sowohl in der TwinsUK- als auch in der KORA-F4-Population auf identischen Plattformen (Biocrates und Metabolon) durchgeführt wurden. Analog zur mGWAS im KORA-F4-Datensatz sind auch hier die Assoziationsergebnisse unter mips.helmholtz-muenchen.de/proj/GWAS zum Download verfügbar.

Genexpressionsdaten aus RNA-Sequenzierung für 16 verschiedene Gewebetypen stammen aus der *Human BodyMap 2.0 from Illumina*. Dieser Datensatz kann über den *Expression Atlas* unter www.ebi.ac.uk/gxa/experiments/E-MTAB-513 abgerufen und durchsucht werden [75].

NMR-Spektren verschiedener Metaboliten wurden aus der *Human Metabolome Database* (HMDB) abgerufen (www.hmdb.ca) [4].

Alle deskriptiven Plots (beispielsweise Boxplots, regionale Assoziationsplots, Q-Q-Plots, Korrelations- und Assoziationsspektren) wurden in R (Version 2.15.2 und 3.0.2) erstellt [65].

5. Ergebnisse

5.1. Erste GWAS mit metabolischen Merkmalen im Urin

Für die Discovery-Studie der ersten mGWAS im Urin wurden Urinproben von 862 männlichen Teilnehmern der SHIP-0-Kohorte herangezogen. Von diesen Proben wurden ^1H -NMR Spektren gemessen. Aufgrund der komplexen Zusammensetzung der NMR-Spektren (bedingt durch Signalüberlagerungen und pH-bedingte Signalverschiebungen) wurden diese mit Hilfe der Chenomx NMR Suite manuell annotiert, um die gemessenen Metaboliten zu identifizieren und zu quantifizieren. Dadurch wurden in jeder Urinprobe die Konzentrationen von bis zu 63 Stoffwechselprodukten bestimmt. Dieser Datensatz wird in dieser Arbeit mit *SHIP-0 males* bezeichnet.

5 genomweit signifikant assoziierte Loci

Die Assoziationen genetischer Polymorphismen zu metabolischen Merkmalen wurden mittels alterskorrigierter linearer Modelle untersucht. Dabei wurden über alle 862 Probanden 1.770 metabolische Phänotypen, entsprechend 59 Metabolitenkonzentrationen und 1.661 davon abgeleitete paarweisen Konzentrationsverhältnissen (Ratios), sowie 645.249 autosomale SNPs getestet. Nach der Bonferroni-Korrektur für multiples Testen von 645.249 SNPs sind 1.319 Assoziationen signifikant bei p-Werten kleiner $0,05 / 645.249 = 7,75 \times 10^{-8}$, wobei 161 SNPs mit metabolischen Merkmalen assoziierten. Diese 161 SNPs repräsentieren 61 genetische Regionen (Loci). Ein Locus ist definiert als eine Ansammlung von SNPs, die höchstens jeweils 1.000.000 Basenpaare voneinander entfernt sind. Legt man ein Signifikanzniveau zugrunde, welches neben der Anzahl untersuchter SNPs auch für die Anzahl der getesteten metabolischen Merkmale Bonferroni-korrigiert wurde, zeigen fünf genetische Loci signifi-

kante Assoziationen mit Stoffwechselprodukten oder daraus abgeleiteten Ratios bei p-Werten kleiner $0,05 / (645.249 \times 1.720) = 4,51 \times 10^{-11}$ (Tabelle 5.1). Diese fünf Loci liegen im Bereich der Gene AGXT2, NAT2, HPD, SLC7A9 sowie SLC6A20. Bei drei Loci (NAT2, SLC7A9, SLC6A20) konnte eine Steigerung der Assoziationsstärke um bis zu sechs Größenordnungen beobachtet werden, wenn anstatt der Konzentrationen einzelner Metaboliten der Ratio zweier Metaboliten getestet wurde.

Replikation und Verifikation in SHIP-0, SHIP-1 und KORA F4

Zur Replikation der jeweils stärksten Assoziation der fünf im *SHIP-0 males* Datensatz identifizieren Loci wurden unterschiedliche Teilmengen der SHIP-0-, SHIP-1- und KORA-F4-Kohorten untersucht. Während die NMR-Spektren des Discovery-Datensatzes als Auftragsarbeit durch Chenomx (Edmonton, Kanada) annotiert wurden, wurden die Spektren der Replikationsdatensätze durch den Autor dieser Dissertation annotiert. Bei allen zum Zwecke der Replikation herangezogenen Datensätzen wurden nur diejenigen Metaboliten quantifiziert, welche das jeweils stärkste Assoziationssignal zeigen, um mit vertretbarem Aufwand eine möglichst große Anzahl an Probanden für die Replikation heranzuziehen. Abbildung 5.1 zeigt die nach Genotyp des untersuchten SNPs aufgetrennte Verteilung der in den Replikationsdatensätzen annotierten Metabolitenkonzentrationen bzw. der daraus abgeleiteten Ratios; Tabelle 5.2 nennt die Ergebnisse der Regressionsanalyse.

Männliche SHIP-0 Probanden

Für den ersten Replikationsversuch wurden die annotierten NMR-Spektren der Urinproben weiterer 126 männlicher SHIP-0-Studienteilnehmer verwendet. Die Auswahl der Probanden erfolgte anhand ihres Genotyps: es wurden diejenigen ausgewählt, welche für genau einen der fünf untersuchten SNPs homozygoter Träger des seltenen Allels sind. Ein solcher von klinischen Studien inspirierter Ansatz führt dazu, dass seltene genetische Varianten selbst in einem kleinen Datensatz ausreichend repräsentiert sind. Im Gegensatz zu den Urinproben, die für die Discovery-Studie herangezogen wurden, erfolgte bei diesem Datensatz die Abgabe des Urins erst nach 11 Uhr vormittags. Bei dem als *SHIP-0 males late* bezeichneten Datensatz war daher zu erwarten, dass die durch Biorhythmus, Nahrungsaufnahme und ähnliche Faktoren verursachten

Tabelle 5.1.: Erste mGWAS im Urin: Loci mit genomweit signifikanten Assoziationen zu metabolischen Merkmalen im SHIP-0 males Datensatz

Genetische Daten							Assoziiertes metabolisches Merkmal				
Locus	SNP	Chr	Position	MIN	MAJ	MAF	Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
AGXT2	rs37369	5	35.072.872	T	C	9,2 %	beta-Aminoisobutyrat	665	2,396	$3,17 \times 10^{-75}$	
NAT2	rs4921914	8	18.316.718	C	T	23,3 %	Formiat / Succinat	854	0,2677	$5,10 \times 10^{-16}$	$3,0 \times 10^5$
WDR66 / HPD	rs830124	12	120.880.160	A	G	15,2 %	alpha-Hydroxyisobutyrat	859	-0,1547	$1,59 \times 10^{-15}$	
SLC7A9	rs8101881	19	38.056.468	C	T	35,4 %	Lysin / Valin	605	0,2853	$1,82 \times 10^{-14}$	$5,6 \times 10^5$
SLC6A20	rs17279437	3	45.789.098	A	G	11,5 %	Alanin / N,N-Dimethylglycin	831	0,3709	$2,30 \times 10^{-13}$	$2,4 \times 10^3$

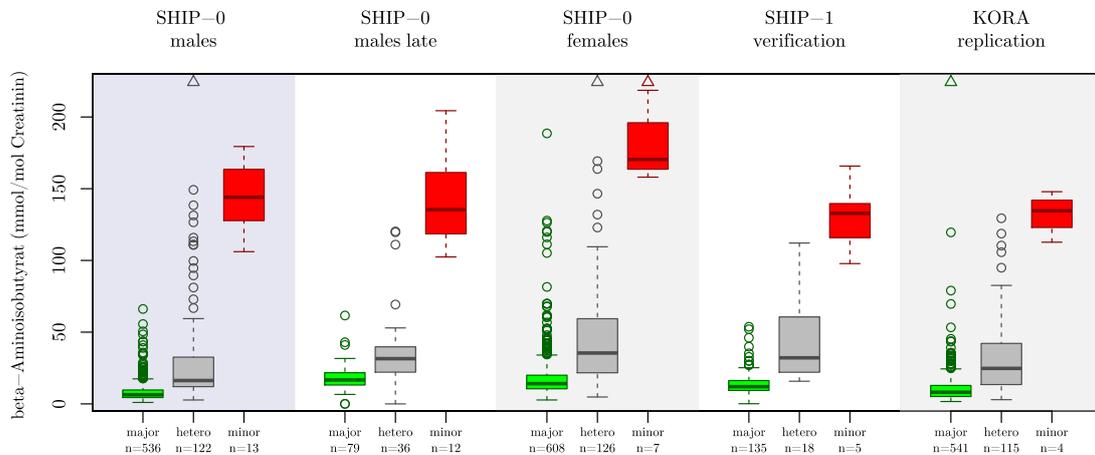
Der Locus nennt das Gen, in dessen Region der am stärksten assoziierte SNP liegt. Abkürzungen: Chr = Chromosom; MIN = seltenes (Effekt-)Allel (*minor allele*) des SNPs; MAJ = häufiges Allel (*major allele*); MAF = Häufigkeit des seltenen Allels in der Studienpopulation (*minor allele frequency*); N = Anzahl der Probanden, für die sowohl genetische als auch metabolische Daten vorlagen. β' ist die relative Effektgröße, definiert als $\beta' = 10^{\beta} - 1$, wobei β die Steigung im linearen Modell bei Verwendung von \log_{10} -skalierten Metabolitenkonzentrationen bzw. Ratios entspricht und der Genotyp des getesteten SNPs mit 0, 1 oder 2 kodiert ist (entsprechend der Anzahl der Kopien des seltenen Allels). β' repräsentiert somit den relativen Unterschied pro Kopie des seltenen Allels für unskalierte metabolische Merkmale im Vergleich zum geschätzten Mittelwert der Gruppe der homozygoten Träger des häufigen Allels. Der p-Gain ist definiert als $\min(p(M_1)/p(M_1/M_2), p(M_2)/p(M_1/M_2))$, wobei $p(M_1)$, $p(M_2)$ und $p(M_1/M_2)$ den p-Werten der Assoziationen der \log_{10} -skalierten Metabolitenkonzentrationen M_1 und M_2 bzw. daraus errechneten, \log_{10} -skalierten Ratio (M_1/M_2) entsprechen. Die chromosomale Position eines SNPs sowie die Angaben zum seltenen bzw. häufigen Allel beziehen sich stets auf den Plusstrang des menschlichen Genoms (NCBI build 36.1) [76].

Tabelle 5.2.: Erste mGWAS im Urin: Assoziationsergebnisse in den Verifikations- bzw. Replikationsdatensätzen

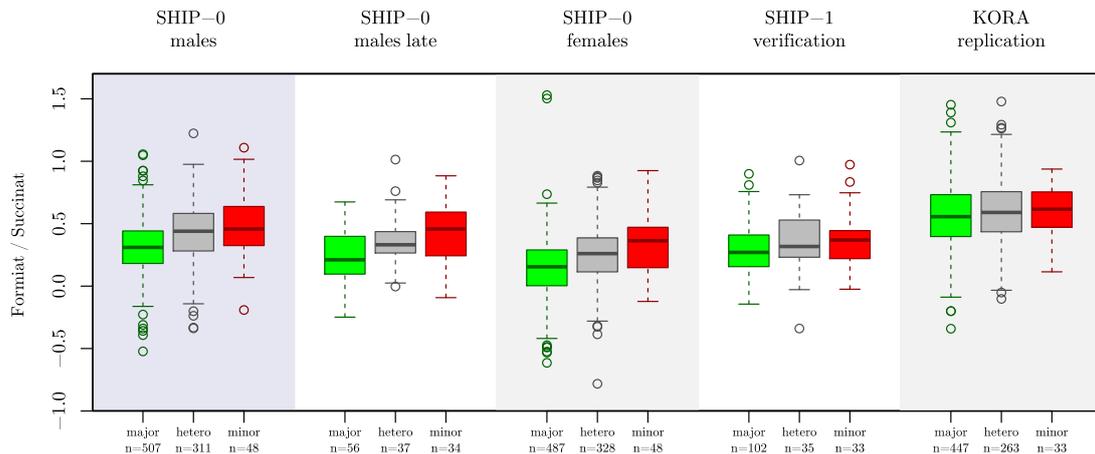
Locus	SNP	Metabolit / Ratio	SHIP-0 males				SHIP-0 females				SHIP-1 verification KORA replication			
			N	p-Wert	N	p-Wert	N	p-Wert	N	p-Wert	N	p-Wert	N	p-Wert
AGXT2	rs37369	beta-Aminoisobutytrat	665	$3,2 \times 10^{-75}$	123	$8,3 \times 10^{-27}$	724	$1,3 \times 10^{-56}$	157	$1,5 \times 10^{-23}$	660	$3,3 \times 10^{-52}$		
NAT2	rs4921914	Formiat / Succinat	854	$5,1 \times 10^{-16}$	126	$1,7 \times 10^{-4}$	860	$1,2 \times 10^{-15}$	170	0,019	738	0,0024		
HPD / WDR66	rs830124	alpha-Hydroxyisobutytrat	859	$1,6 \times 10^{-15}$	126	$8,1 \times 10^{-2}$	851	$3,4 \times 10^{-22}$	169	$1,4 \times 10^{-4}$	842	$4,8 \times 10^{-12}$		
SLC7A9	rs8101881	Lysin / Valin	605	$1,8 \times 10^{-14}$	121	0,993	521	$7,9 \times 10^{-12}$	145	$5,4 \times 10^{-6}$	689	$5,0 \times 10^{-5}$		
SLC6A20	rs17279437	Alanin / N,N-Dimethylglycin	831	$2,3 \times 10^{-13}$	126	0,183	819	$2,6 \times 10^{-5}$	167	$2,8 \times 10^{-3}$	789	$8,1 \times 10^{-5}$		

Aufgeführt sind die Assoziationsergebnisse für die Discovery-Studie, deren Datensatz (*SHIP-0 males*) die vor 11 Uhr abgegebenen Urinproben männlicher Teilnehmer der SHIP-0 Studie umfasst, sowie für die Replikationsdatensätze *SHIP-0 males late* (männliche Teilnehmer der SHIP-0 Studie; Urinabgabe nach 11 Uhr), *SHIP-0 females* (weibliche Teilnehmer der SHIP-0 Studie; Urinabgabe vor 11 Uhr), *KORA replication* (männliche und weibliche Probanden der KORA-Studie) und dem Datensatz einer longitudinalen Nachfolgestudie (*SHIP-1 verification*), der aus Probanden des *SHIP-0 males* Datensatzes besteht, welche ungefähr fünf Jahre nach der Discovery-Studie erneut untersucht wurden. *N* gibt die Anzahl der Probanden an, für die sowohl Genotyp- als auch Metabotyp-Daten in guter Qualität vorlagen.

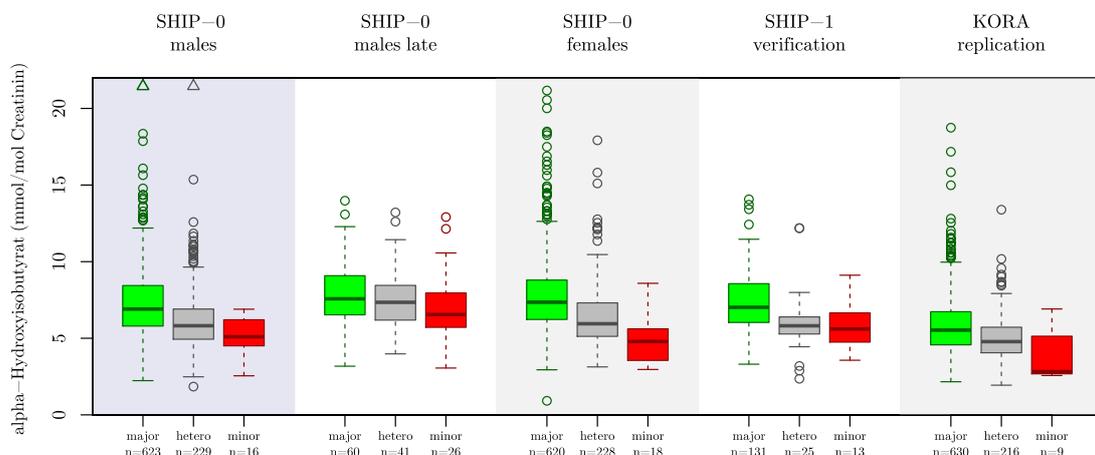
Abbildung 5.1.: Erste mGWAS im Urin: Metabolitenkonzentrationen und -ratios im Discovery-Datensatz und in den Replikationsdatensätzen



(a) beta-Aminoisobutyrat nach rs37369-Genotyp (AGXT2-Locus)

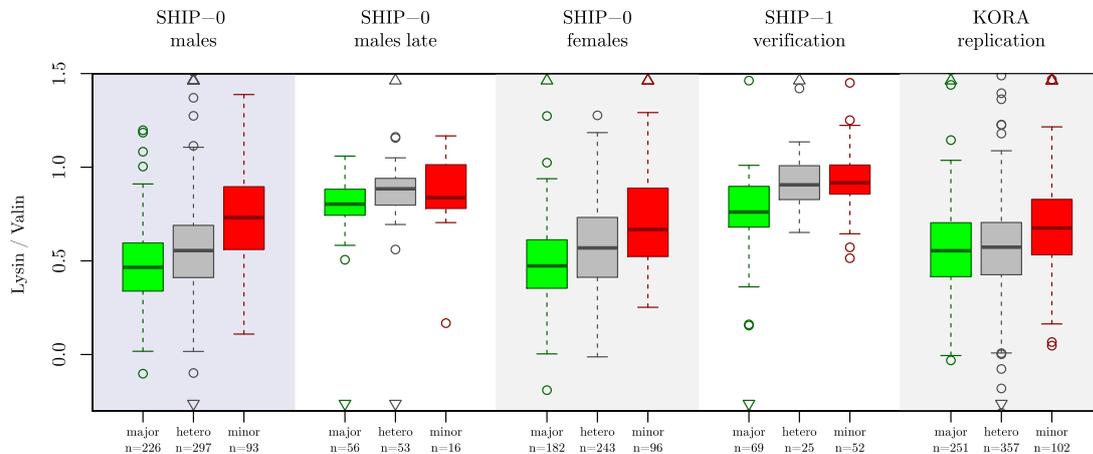


(b) Formiat/Succinat-Ratios nach rs4921914-Genotyp (NAT2-Locus)

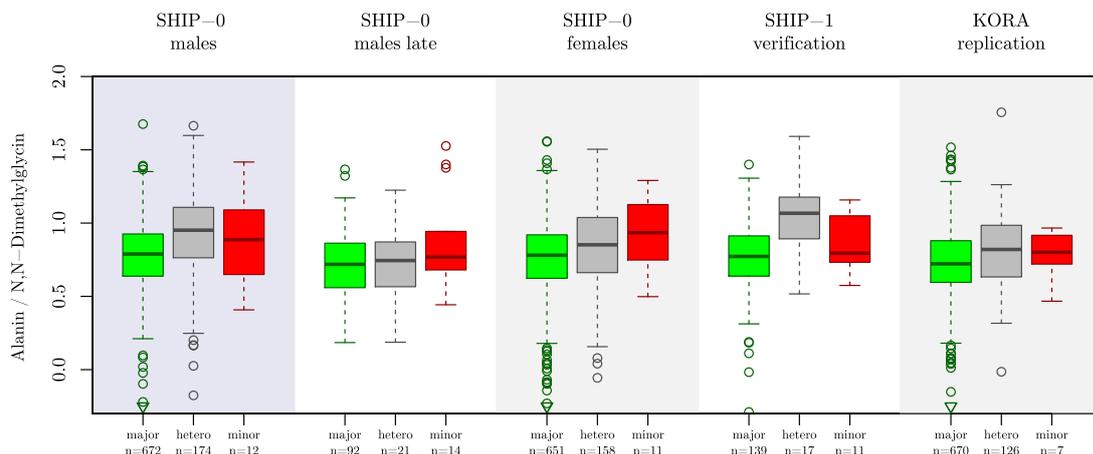


(c) alpha-Hydroxyisobutyrat nach rs8301024-Genotyp (WDR66/HPD-Locus)

Abbildung 5.1.: Metabolitenkonzentrationen und -ratios im Discovery-Datensatz und in den Replikationsdatensätzen (Forts.)



(d) Lysin/Valin-Ratios nach rs8101881-Genotyp (SLC7A9-Locus)



(e) Alanin/N,N-Dimethylglycin-Ratios nach rs17279437-Genotyp (SLC6A20-Locus)

Die Abbildungen zeigen für jeden der untersuchten Datensätze die annotierten Metabolitenkonzentrationen bzw. die daraus abgeleiteten Ratios in Abhängigkeit vom Genotyp des SNPs, welcher in der Discovery-Studie (basierend auf dem *SHIP-0 males* Datensatz) im angegebenen genetischen Locus die stärkste Assoziation zeigt. Die bei homozygoten Trägern des häufigen bzw. seltenen Allels gemessenen und auf Kreatinin normalisierten Konzentrationen oder aus paarweisen Metabolitenkonzentrationen errechneten Ratios sind in grün bzw. rot eingezeichnet, bei heterozygoten Trägern sind sie grau eingefärbt. n gibt die jeweilige Anzahl der gemessenen Samples für jeden Genotyp an. Die Boxplots beschreiben die Verteilung der Datenpunkte. Die Boxen erstrecken sich jeweils vom ersten (Q1) bis zum dritten Quartil (Q3) und umfassen somit 50% der Daten. Der Median ist durch den den dickeren horizontalen Strich gekennzeichnet. Die „Antennen“ zeigen das Maximum bzw. Minimum der Verteilung, solange die Datenpunkte nicht weiter als $1,5 \times (Q3 - Q1)$ vom jeweiligen Ende der Box entfernt sind. Ausreißer werden als einzelne Punkte dargestellt. Dreiecke am oberen bzw. unteren Rand des Plots weisen darauf hin, dass sich mindestens ein extremer Datenpunkt außerhalb des abgebildeten Wertebereiches befindet.

Schwankungen der individuellen Metabolitenkonzentrationen stärker sein würden als im Discovery-Datensatz (*SHIP-0 males*).

Aus Tabelle 5.2 wird ersichtlich, dass lediglich die Assoziationen von rs37369 (AGXT2) mit beta-Aminoisobutyrat sowie rs4921914 (NAT2) mit dem aus Formiat und Succinat gebildeten Ratio signifikant replizieren. Als Kriterium für Signifikanz dient hierbei das nach der Bonferroni-Korrektur für fünf unabhängige Tests angepasste Signifikanzniveau $\alpha' = 0,01$. Vermutlich sind die durch äußere Einflüsse verursachten Schwankungen der Metabolitenkonzentrationen sowie der geringe Umfang des Datensatzes dafür verantwortlich, dass die in der Discovery-Studie identifizierten Assoziationen hier nur teilweise repliziert werden konnten.

Weibliche SHIP-0 Probanden

Ein zweiter Replikationsversuch wurde in einem Datensatz unternommen, der bezüglich der Anzahl der Probanden sowie der Charakteristika der Urinproben dem Discovery-Datensatz sehr ähnlich ist. Erst einige Zeit nach Abschluss der Discovery-Studie waren auch die NMR-Spektren der Urinproben aller weiblichen SHIP-0-Probanden. Im Gegensatz zum ersten Replikationsdatensatz erfolgte im *SHIP-0 females*-Datensatz keine durch den Genotyp bestimmte Auswahl der Probanden. Stattdessen wurden alle Teilnehmerinnen in den Datensatz aufgenommen, deren Urinproben vormittags zwischen 8:00 Uhr und 11:00 Uhr gesammelt wurden. Insgesamt wurden 870 NMR-Spektren manuell annotiert und die resultierenden metabolischen Phänotypen zur Replikation herangezogen. In diesem Datensatz konnten alle fünf untersuchten Assoziationen bei p-Werten zwischen $2,6 \times 10^{-5}$ und $1,3 \times 10^{-56}$ repliziert werden. Diese p-Werte liegen deutlich unterhalb des Bonferroni-korrigierten Signifikanzniveaus von $\alpha' = 0,01$ und wären somit mit Ausnahme der Assoziation im SLC6A20 Locus in einer GWAS genomweit signifikant. Dies lässt vermuten, dass die im Discovery-Datensatz identifizierten GIMs nicht geschlechtsspezifisch sind.

Probanden der unabhängigen KORA-F4-Studie

Im *SHIP-0 females*-Datensatz konnten zwar die fünf in der Discovery-Studie entdeckten Loci repliziert werden, das Kriterium einer unabhängigen Replikation war damit allerdings nicht erfüllt, da sowohl die Discovery-Studie als auch die Replika-

tion in derselben Kohorte durchgeführt wurden. Für eine Veröffentlichung musste gezeigt werden, dass sich die Ergebnisse auch auf eine Population anderen geographischen Ursprungs übertragen lassen, die über einen unterschiedlichen genetischen Hintergrund verfügt und zudem möglicherweise anderen Umwelteinflüssen unterliegt als die in der SHIP-Studie untersuchte Population. Zudem sollte gezeigt werden, dass die Assoziationen auch in Daten aus anderen Studienzentren und Laboren repliziert werden können [77]. Um die gewünschte unabhängige Replikation durchzuführen, wurden die NMR-Spektren von 992 Urinproben zufällig ausgewählter Probanden der KORA-F4-Studie aufgezeichnet und anschließend manuell annotiert. Die Studienteilnehmer und -teilnehmerinnen waren bei der vormittäglichen Urinabgabe nüchtern.

Alle in der Discovery-Studie identifizierten GIMs konnten auch im KORA-F4-Datensatz repliziert werden (Tabelle 5.2). Bei zwei der fünf SNPs, rs37369 (AGXT2) sowie rs830124 (WDR66/HPD), sind die Assoziationen von vergleichbarer Stärke wie in der im Discovery-Datensatz durchgeführten GWAS. Die SNPs rs810881 (SLC7A9) und rs17279437 (SLC6A20) assoziieren zwar deutlich schwächer mit den jeweiligen metabolischen Merkmalen als in der Discovery-Studie, sind aber mit p-Werten von $5,0 \times 10^{-5}$ bzw. $8,1 \times 10^{-5}$ noch deutlich unter dem für fünf Tests angepassten Signifikanzniveau $\alpha' = 0,01$. Einen Sonderfall stellt die Assoziation zwischen rs4921914 (NAT2) und dem Formiat/Succinat-Ratio dar. Bei einem p-Wert von 0,0024 ist diese zwar gerade noch signifikant, aber dennoch um bis zu 13 Größenordnungen schwächer verglichen mit den Assoziationen im *SHIP-0 discovery*- und *SHIP-0 females*-Datensatz. Die Ursache für diesen eklatanten Unterschied ist jedoch unklar.

Longitudinale Verifikation in SHIP-1

Um zu zeigen, dass die vorgestellten Assoziationen auch über die Zeit hinweg stabil sind, wurden NMR-Spektren der Urinproben ausgewählter Teilnehmern der SHIP-1-Studie aufgezeichnet und annotiert. Die SHIP-1-Studie ist eine Follow-Up-Studie, welche fünf Jahre nach SHIP-0 durchgeführt wurde. Die Auswahl der Proben erfolgte dabei nach zwei Kriterien: Zum einen mussten die (männlichen) Probanden bereits Teilnehmer der SHIP-0-Discovery-Studie gewesen sein. Zum anderen wurden – analog zum ersten Replikationsversuch – nur Probanden ausgewählt, die für genau einen der fünf untersuchten SNPs homozygoter Träger des seltenen Allels sind. Insgesamt erfüllten 170 Probanden diese Auswahlkriterien, weshalb deren Urinproben für die longitudinale Verifikation verwendet wurden. Trotz des geringen Umfangs dieser Veri-

fikationsstudie zeigen vier der fünf genetischen Loci signifikante Assoziationen zu den jeweiligen Metabolitenkonzentrationen bzw. Ratios bei p-Werten zwischen $2,8 \times 10^{-3}$ und $1,5 \times 10^{-23}$ (Tabelle 5.2). Lediglich die Assoziation von rs4921914 (NAT2) zu dem aus Formiat und Succinat gebildeten Ratio ist mit einem p-Wert von 0,019 knapp oberhalb des Signifikanzniveaus $\alpha' = 0,01$.

Diese Verifikationsstudie zeigt, dass die im Discovery-Datensatz gemessenen metabolischen Marker zum größten Teil auch über einen längeren Zeitraum hinweg stabil genug sind, um signifikant mit genetischen Markern zu assoziieren.

5.2. Zweite GWAS mit metabolischen Merkmalen im Urin

Als Grundlage für die zweite mGWAS mit Stoffwechselprodukten im Urin dienen Genotyp- und Metabotyp-Daten von 4.087 männlichen und weiblichen Teilnehmern der SHIP-0-Studie.

14 genomweit signifikant assoziierte Loci

Analog zur Vorgängerstudie wurden lineare Modelle verwendet, um unter Annahme eines additiven Effektmodells Assoziationen zwischen SNPs und Metabolitenkonzentrationen bzw. daraus berechneten paarweisen Ratios zu identifizieren. Als Kovariaten wurden das Alter sowie das Geschlecht der Probanden in die Modellberechnungen einbezogen. In der Discovery-Studie wurden 620.456 autosomale SNPs auf Assoziation zu 1.518 im Urin bestimmten metabolische Merkmalen (55 Konzentrationen und 1.463 davon abgeleitete Ratios) untersucht. Daraus ergibt sich ein Bonferroni-korrigiertes Signifikanzniveau von $\alpha' = 0,05 / 620.456 / 1.518 = 5,3 \times 10^{-11}$. 156 SNPs zeigten Assoziationen bei p-Werten unterhalb dieser Grenze ($2,1 \times 10^{-11} < p < 3,8 \times 10^{-307}$). Sie wurden zu 14 genetischen Loci (CPS1, XYLB, SLC6A20, ETNPPL, SLC6A19, AGXT2, DMGDH, SLC36A2, NAT2, SLC16A9, SLC6A13, HPD, FBXL20, SLC7A9) zusammengefasst, wobei der paarweise Abstand der signifikanten SNPs innerhalb eines Locus' höchstens 1.000.000 Basenpaare beträgt. Tabelle 5.3 nennt die Loci, den jeweiligen SNP, der das stärkste Assoziationssignal zeigt („Top-SNP“), sowie den damit assoziierten Metabolit bzw. Ratio. Im Anhang dieser Arbeit

sind für jeden Locus die entsprechenden regionalen Assoziationsplots, die Boxplots sowie die Q-Q Plots zu finden (Abbildung A.1).

Im Falle von genetischen Assoziationen zu Ratios diene der p-Gain als ein weiteres Signifikanzkriterium. Der p-Gain ist dabei die Zunahme der Assoziationsstärke im Vergleich zu den p-Werten, die sich aus den Assoziationen des getesteten SNPs mit den einzelnen Metaboliten des Ratios ergeben. Wie Petersen *et al.* in einer 2012 erschienenen Publikation ausführen, sollte der minimale p-Gain mindestens der Anzahl getesteter Metaboliten und Ratios geteilt durch 2α [32] entsprechen. Dadurch ergibt sich für die Discovery-Studie ein p-Gain von $1.518 / (2 \times 0,05) = 1,51 \times 10^4$, oberhalb dessen genetische Assoziationen zu Ratios als signifikant gelten. Bei zehn der elf in Tabelle 5.3 aufgeführten Assoziationen zu Ratios ist der p-Gain groß genug. Lediglich im Falle der Assoziation zwischen dem SLC16A9-Locus und dem Ratio aus Betain und Citrat lag der p-Gain mit $1,19 \times 10^4$ knapp unter der Signifikanzgrenze. Da dieser Locus aber in einer früheren mGWAS in der KORA-F4-Studie signifikant mit Carnitin im Blutserum assoziiert und dort experimentell die Funktion des SLC16A9-Transporters als Carnitin-Transporter nachgewiesen wurde, wird dieser Locus dennoch in dieser Arbeit behandelt.

Replikation in KORA F4

Die Replikation der Ergebnisse erfolgte anhand des KORA-F4-Urin-Datensatzes. Dieser umfasst für 1.813 Studienteilnehmer die Genotypen von 593.830 autosomalen SNPs und 1.251 metabolischen Merkmalen im Urin (50 Konzentrationen und 1201 Ratios). Für acht der vierzehn in der Discovery-Studie identifizierten Loci konnte die jeweils beste Genotyp-Phänotyp-Assoziation direkt repliziert werden (CPS1, ETNPPL, SLC16A9, AGXT2, NAT2, SLC6A13, HPD, SLC7A9). Bei vier weiteren Loci (SLC6A20, SLC36A2, FBXL20, XYLB) konnte nicht die stärkste, jedoch die zweitstärkste genomweit signifikante Assoziation des Top-SNP zum jeweiligen metabolischen Merkmal repliziert werden. Der Top-SNP im DMGDH-Locus war aufgrund der mit der Qualitätskontrolle der Genotypdaten einhergegangenen Filterung im KORA-Datensatz nicht mehr vorhanden. Stattdessen wurde hier der SNP ausgewählt, der am stärksten im LD ($r^2 = 0,86$) mit dem ursprünglich zu testenden Top-SNP ist. Der p-Wert der Assoziation des SLC16A9-Locus lag in der Replikation deutlich oberhalb des Bonferroni-korrigierten Signifikanzniveaus von $\alpha' = 0,05 / 18 = 0,0028$, welches für 18 Tests (14 SNP/metabolisches Merkmal Paare plus vier zusätzliche Test mit

Tabelle 5.3.: Zweite mGWAS im Urin: signifikante Assoziationen genetischer Loci zu metabolischen Merkmalen im Discovery-Datensatz

Genetische Daten										Assoziiertes metabolisches Merkmal				
Locus	SNP	Chr	Position	MAJ	MIN	MAF	Metabolit oder Ratio	Top	N	beta'	p-Wert	p-Gain		
CPS1	rs2216405	2	211.325.139	A	G	19,65 %	Glycin / Glycolat	ja	3.790	0,1376	$3,72 \times 10^{-17}$	$3,38 \times 10^7$		
XYLB	rs17118	3	38.379.476	C	A	35,27 %	Glycolat / Threonin	nein	3.747	0,0770	$4,69 \times 10^{-11}$	$6,94 \times 10^6$		
SLOC6A20	rs17279437	3	45.789.098	G	A	11,49 %	N,N-Dimethylglycin / Threonin	nein	3.667	-0,2927	$1,75 \times 10^{-54}$	$2,09 \times 10^{30}$		
ETNPPL	rs7692671	4	109.933.604	G	A	39,78 %	Ethanolamin / Dimethylamin	ja	3.597	0,0656	$6,35 \times 10^{-19}$	$1,21 \times 10^{14}$		
SLOC6A19	rs7719875	5	1.243.088	A	G	36,81 %	Histidin / N,N-Dimethylglycin	ja	3.592	-0,1424	$4,75 \times 10^{-20}$	$1,40 \times 10^{11}$		
AGXT2	rs37369	5	35.072.872	C	T	8,81 %	beta-Aminoisobutyrat	ja	3.344	2,4674	$3,76 \times 10^{-307}$			
DMGDH	rs248386	5	78.365.983	C	A	18,28 %	N,N-Dimethylglycin / Glycolat	ja	3.635	0,1820	$2,85 \times 10^{-18}$	$7,15 \times 10^9$		
SLOC36A2	rs3846710	5	150.678.999	T	C	16,77 %	Glycin / Valin	nein	3.068	-0,1192	$2,11 \times 10^{-11}$	$4,86 \times 10^4$		
NAT2	rs4921914	8	18.316.718	T	C	23,46 %	Formiat / Harnstoff	ja	3.825	0,3107	$6,42 \times 10^{-64}$	$2,70 \times 10^{13}$		
SLOC16A9*	rs7094971	10	61.119.570	A	G	13,25 %	Betain / Citrat	ja	3.798	0,1853	$1,41 \times 10^{-12}$	$1,19 \times 10^4$		
SLOC6A13	rs11613331	12	221.728	A	G	46,27 %	beta-Aminoisobutyrat	ja	3.345	-0,1544	$5,61 \times 10^{-15}$			
HPD	rs4760099	12	120.803.106	G	A	15,10 %	alpha-Hydroxyisobutyrat	ja	3.853	-0,1655	$1,32 \times 10^{-77}$			
FBXL20	rs2338755	17	34.672.843	C	T	25,13 %	Histidin / Harnstoff	nein	3.720	-0,1568	$5,85 \times 10^{-23}$	$5,49 \times 10^5$		
SLOC7A9	rs8101881	19	38.056.468	T	C	35,95 %	Lysin / Valin	ja	868	0,1395	$2,06 \times 10^{-27}$	$2,67 \times 10^{11}$		

Locus = Gen, in dessen Region der SNP liegt. Chr = Chromosom. MIN = seltenes (Effekt-)Allel (*minor allele*). MAJ = häufiges Allel (*major allele*). Die chromosomale Position eines SNPs sowie die Allelangaben beziehen sich stets auf den Plusstrang des menschlichen Genoms (NCBI build 36.1) [76]. MAF = Häufigkeit des seltenen Allels (*minor allele frequency*). Bei Ratios ist der Metabolit, welcher stärker mit dem SNP assoziiert, im Zähler des Bruchs aufgeführt. Top gibt an, ob das metabolische Merkmal am stärksten mit dem in der Spalte SNP genannten Polymorphismus assoziiert. In vier Fällen zeigt die Tabelle die zweitbeste Assoziation, da nur diese im Urin der KORA-F4-Studie signifikant replizierte. N = Anzahl der Samples, für die sowohl genetische als auch metabolische Daten vorlagen. beta' ist die relative Effektgröße, definiert als $\beta' = 10^{\beta'_{\text{beta}}} - 1$, wobei beta der Steigung im linearen Modell bei Verwendung von \log_{10} -skalierten Metabolitenkonzentrationen bzw. -ratios entspricht und der Genotyp des getesteten SNPs mit 0, 1 oder 2 kodiert ist, entsprechend der Anzahl der Kopien des seltenen Allels. beta' repräsentiert somit den relativen Unterschied pro Kopie des seltenen Allels für unskalierte Konzentrationen und -ratios im Vergleich zum geschätzten Mittelwert in der Gruppe der homozygoten Träger des häufigen Allels. Bei Ratios ist der p-Gain definiert als $\min(p(M_1)/p(M_1/M_2), p(M_2)/p(M_1/M_2))$, wobei $p(M_1)$, $p(M_2)$ und $p(M_1/M_2)$ den p-Werten der Assoziationen der \log_{10} -skalierten Metabolitenkonzentrationen M_1 und M_2 bzw. daraus errechneten, \log_{10} -skalierten Ratio (M_1/M_2) entsprechen. Die Assoziation im SLOC16A9-Locus (mit * gekennzeichnet) ist aufgrund des zu kleinen p-Gains ($< 1,51 \times 10^4$) nominell nicht signifikant. Da dieser Locus in dieser Arbeit besprochen wird, ist er dennoch hier aufgeführt.

dem jeweils zweitbesten metabolischen Merkmal) angepasst wurde. Der SLC16A9-Locus konnte also nicht im Urin repliziert werden. Die Ergebnisse der Replikation sind in Tabelle 5.4 zu finden.

Tabelle 5.4.: Zweite mGWAS im Urin: Replikation der Assoziationen im KORA-F4-Urin-Datensatz

Genetische Daten		Assoziiertes metabolisches Merkmal				
Locus	SNP (LD r ²)	Metabolit oder Ratio	Tests	N	beta'	p-Wert
CPS1	rs2216405	Glycin / Glycolat	1	1.630	0,1047	1,29×10 ⁻⁵
XYLB	rs17118	Glycolat / Threonin	2	1.596	0,0568	4,04×10 ⁻⁴
SLC6A20	rs17279437	N,N-Dimethylglycin / Threonin	2	1.568	-0,2700	4,96×10 ⁻¹⁹
ETNPPL	rs7692671	Ethanolamin / Dimethylamin	1	1.535	0,0446	1,14×10 ⁻⁴
SLC6A19	rs7719875	Histidin / N,N-Dimethylglycin	1	1.506	-0,1293	2,10×10 ⁻⁸
AGXT2	rs37369	beta-Aminoisobutyrat	1	1.400	2,9021	3,18×10 ⁻¹¹⁶
DMGDH	rs631305 (0,86)	N,N-Dimethylglycin / Glycolat	1	1.547	0,1099	3,42×10 ⁻⁴
SLC36A2	rs3846710	Glycin / Valin	2	1.304	-0,1209	4,57×10 ⁻⁷
NAT2	rs4921914	Formiat / Harnstoff	1	1.613	0,0866	8,47×10 ⁻⁵
SLC16A9*	rs7094971	Betain / Citrat	1	1.573	0,0139	0,691
SLC6A13	rs11613331	beta-Aminoisobutyrat	1	1.408	-0,1444	1,90×10 ⁻⁵
HPD	rs4760099	alpha-Hydroxyisobutyrat	1	1.678	-0,1591	6,78×10 ⁻²⁶
FBXL20	rs2338755	Histidin / Harnstoff	2	1.587	-0,1166	7,53×10 ⁻⁹
SLC7A9	rs8101881	Lysin / Valin	1	169	0,2885	7,34×10 ⁻⁴

Die Erläuterungen zu Tabelle 5.3 gelten analog. Die Tests-Spalte gibt an, wie viele metabolische Merkmale auf Assoziation mit dem jeweiligen SNP untersucht wurden. In den meisten Fällen replizierte das in der Discovery-Studie am besten assoziierte Merkmal signifikant im Replikationsdatensatz; gleichwohl konnte in vier Fällen nur die zweitstärkste Assoziation repliziert werden. Im Falle des DMGDH-Locus war der im Discovery-Datensatz assoziierte SNP rs248386 im KORA-F4-Datensatz nicht enthalten. Stattdessen wurde derjenige SNP gewählt, der laut HapMap (Release 27, Phase I-III) am stärksten im LD mit rs248386 ist [72]. Die Assoziation im SLC16A9-Locus (mit * gekennzeichnet) konnte nicht repliziert werden ($p > 0,05 / 18 = 0,0028$).

Überschneidungen mit GIMs im Blut

Um zu untersuchen, welche der hier identifizierten genetischen Marker im Blutserum ebenfalls signifikante Assoziationen zeigen, wurden die Assoziationsdaten aus GWAS mit massenspektrometrisch bestimmten Metabolitenkonzentrationen aus den KORA-F4- und TwinsUK-Studien herangezogen, welche 2011 und 2014 von Suhre *et al.* sowie Shin *et al.* veröffentlicht wurden [35, 43]. Für jeden der in der Discovery-Studie identifizierte Locus wurde der Top-SNP auf Assoziation mit Metaboliten im

Serum untersucht. In Fällen, in denen ein SNP nicht im KORA-F4- bzw. TwinsUK-Datensatz enthalten war, wurden bis zu fünf weitere SNPs überprüft, die mit mindestens $r^2 > 0,7$ mit dem ursprünglich zu testenden Top-SNP gekoppelt sind (LD-Informationen basierend auf HapMap Release 27, Phase I-III [72]). Die Obergrenze für den p-Wert lag bei $4,31 \times 10^{-8}$ (KORA F4) bzw. bei $1,58 \times 10^{-8}$ (TwinsUK), entsprechend einem Bonferroni-korrigierten Signifikanzniveau von 0,05, welches für 31 (KORA) bzw. 32 (TwinsUK) getestete SNPs sowie 37.455 (KORA) bzw. 98.832 (TwinsUK) metabolische Merkmale (Konzentrationen und paarweise Ratios) korrigiert wurde. Die untere Schranke für den p-Gain war im KORA-F4-Datensatz folglich bei $3,75 \times 10^5$ bzw. bei $9,88 \times 10^5$ im TwinsUK-Datensatz angesetzt.

Im KORA-F4-Datensatz zeigen fünf genetische Loci signifikante Assoziationen (CPS1, NAT2, SLC16A9, SLC6A13, SLC7A9); in den TwinsUK-Daten konnten zusätzlich signifikante Assoziation im SLC6A20- sowie im HPD-Locus identifiziert werden (Tabelle 5.5).

Tabelle 5.5.: Zweite mGWAS im Urin: Vergleich mit Assoziationen im Serum der KORA-F4- und TwinsUK-Studien

KORA F4 [35]										TwinsUK [43]							
Genetische Daten				Assoziiertes metabolisches Merkmal						Genetische Daten				Assoziiertes metabolisches Merkmal			
Locus	Tests	SNP (LD r^2)	Metabolit / Ratio	N	beta	p-Wert	p-Gain	Tests	SNP (LD r^2)	Metabolit / Ratio	N	beta	p-Wert	p-Gain			
CPS1	1	rs2216405	Glycin	1.721	0,0413	$1,28 \times 10^{-15}$		1	rs2216405	Glycin / Betain	5.596	0,091	$1,84 \times 10^{-71}$	$2,17 \times 10^8$			
XYLB	2							1									
SLC6A20	2							1	rs17297437	Pyroglutamin / X-11315	5.575	0,091	$7,76 \times 10^{-21}$	$5,63 \times 10^8$			
ETNPPL	2							6									
SLC6A19	1							1									
AGXT2	2							4									
DMGDH	4							6									
SLC36A2	1							2									
NAT2	1	rs4921914	1-Methylxanthin / 4-Acetamido-butanoat	1.092	-0,1320	$4,03 \times 10^{-25}$	$1,79 \times 10^9$	1	rs4921913	4-Acetamido-butanoat	4.915	0,037	$5,57 \times 10^{-47}$				
SLC16A9	1	rs7094971	Carnitin	1.724	-0,0219	$1,06 \times 10^{-14}$		1	rs7094971	Carnitin	5.590	-0,028	$2,18 \times 10^{-45}$				
SLC6A13	1	rs11613331	Pyroglutamin	1.731	-0,0408	$1,15 \times 10^{-9}$		1	rs11613331	Pyroglutamin	5.590	-0,035	$7,35 \times 10^{-18}$				
HPD	6							1	rs2230681 (1,00)	alpha-Hydroxy-isobutyrat	4.558	-0,041	$1,85 \times 10^{-18}$				
FBXL20	6							5									
SLC7A9	1	rs8101881	Homocitrullin	1.381	-0,0346	$2,49 \times 10^{-8}$		1	rs8101881	Homocitrullin	2.553	-0,033	$3,45 \times 10^{-10}$				

Die Erläuterungen zu Tabelle 5.3 gelten analog. Alle in Tabelle 5.3 aufgeführten SNPs wurden auf Assoziation zu metabolischen Markern im Blutserum von Probanden der KORA-F4- und TwinsUK-Studien überprüft. Falls der Genotyp des SNP in einem Datensatz nicht vorhanden war, wurden bis zu fünf weitere SNPs getestet, die mit $r^2 > 0,7$ im LD mit dem fraglichen SNP sind (HapMap Release # 27, Phase I-III, CEU Population) [72]. Dabei wurden stärker korrelierte SNPs zuerst getestet. Die Tests-Spalte gibt an, wie viele SNPs letztendlich überprüft wurden. Gelistet ist jeweils die beste Assoziation. Die Signifikanzgrenze liegt bei $p < 0,05 / 31 / (276 + 37.179) = 4,31 \times 10^{-8}$ (KORA F4) bzw. $p < 0,05 / 32 / (486 + 98.346) = 1,58 \times 10^{-8}$ (TwinsUK). Bei Ratios muss der p-Gain größer als $3,75 \times 10^5$ (KORA F4) bzw. $9,88 \times 10^5$ sein.

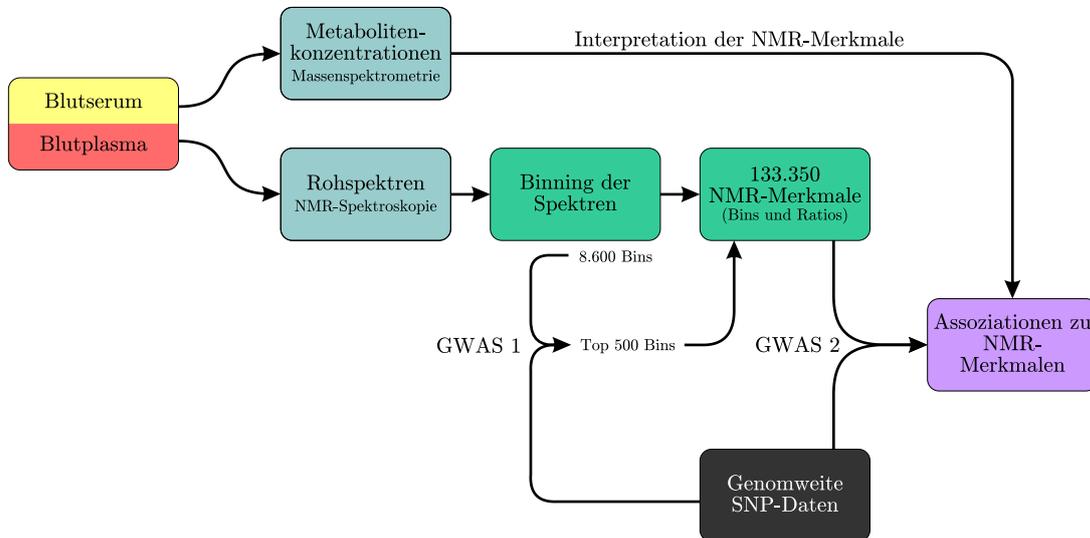
Tabelle 5.6.: GWAS mit NMR-Merkmalen im Plasma: Genetische Assoziationen zu NMR-Bins und NMR-Ratios

Locus	SNP	Chr	Position	Genetische Daten			NMR-Bins			NMR-Ratios					
				A	B	MAF	CS	N	beta'	p-Wert	CS	N	beta'	p-Wert	p-Gain
GCKR	rs780094	2	27.594.741	T	C	41,4%	1,370	1.688	0,0413	1,2×10 ⁻¹⁰	3,286 / 1,370	1.667	-0,0407	2,8×10 ⁻¹⁵	4,3×10 ⁴
CPS1	rs2216405	2	211.325.139	C	T	19,3%	3,599	1.719	0,0456	4,5×10 ⁻¹⁴	3,599 / 2,475	1.718	0,0459	1,8×10 ⁻¹⁹	2,5×10 ⁵
PYROXD2	rs4488133	10	100.149.126	A	T	34,3%	2,757	1.735	-0,0190	7,3×10 ⁻¹²	2,757 / 2,755	1.734	-0,0123	2,9×10 ⁻⁹⁴	2,5×10 ⁸²
FADS1	rs174547	11	61.327.359	C	T	30,0%	2,801	1.733	-0,0408	4,0×10 ⁻³⁵	2,801 / 2,017	1.724	-0,0449	1,1×10 ⁻⁹⁴	3,8×10 ⁵⁹
APOA1	rs3741298	11	116.162.771	C	T	21,4%	2,038	1.724	0,0363	8,4×10 ⁻¹¹	4,162 / 4,082	1.715	-0,0138	1,8×10 ⁻¹⁴	4,6×10 ³
LIPC	rs4775041	15	56.461.987	C	G	28,5%	1,283	1.671	0,0320	1,4×10 ⁻¹⁰	1,068 / 1,029	1.664	-0,0056	3,6×10 ⁻²¹	4,0×10 ¹⁰
CETP	rs247617	16	55.548.217	A	C	31,3%	3,259	1.705	0,0529	7,6×10 ⁻¹⁵	2,211 / 2,011	1.695	-0,0124	1,1×10 ⁻¹⁸	6,8×10 ³

Der Locus nennt das Gen, in dessen Region der am stärksten assoziierte SNP liegt. Abkürzungen: Chr = Chromosom; A = seltenes (Effekt-)Allel (*minor allele*); B = häufiges Allel (*major allele*); MAF = in der Studienpopulation beobachtete Häufigkeit des seltenen Allels (*minor allele frequency*); CS = Chemical Shifts der Bins bzw. Ratios in ppm; N = Anzahl der Samples, für die sowohl der Genotyp des SNPs als auch NMR-Daten vorlagen. beta' ist die relative Effektgröße, definiert als $beta' = 10^{beta} - 1$, wobei beta der Steigung im linearen Modell bei Verwendung von log₁₀-skalierten Bins bzw. Ratios entspricht und der Genotyp des getesteten SNPs mit 0, 1 oder 2 kodiert ist (entsprechend der Anzahl der Kopien des seltenen Allels). beta' repräsentiert somit den relativen Unterschied pro Kopie des seltenen Allels für unskalierte Bins bzw. Ratios im Vergleich zum geschätzten Mittelwert der Gruppe der homozygoten Träger des häufigen Allels. Der p-Gain ist definiert als $\min(p(S_1)/p(S_1/S_2), p(S_2)/p(S_1/S_2))$, wobei $p(S_1)$, $p(S_2)$ und $p(S_1/S_2)$ den p-Werten der Assoziationen der log₁₀-skalierten NMR-Bins bei den Chemical Shifts S₁ und S₂ bzw. des aus diesen Bins errechneten, log₁₀-skalierten Ratio (S₁/S₂) entsprechen. Die chromosomale Position eines SNPs sowie die Allelangaben beziehen sich stets auf den Plusstrang des Genoms (NCBI build 36.1) [76].

5.3. GWAS mit NMR-Merkmalen im Blutplasma

Abbildung 5.1.: Studiendesign



Das Diagramm zeigt die zweistufige GWAS, die auf NMR-Messungen im Blutplasma von 1.757 Probanden der KORA-F4-Studie beruht. Im ersten Schritt (GWAS 1) wurden die in 8.600 Bins aufgeteilten NMR-Signale auf genetische Assoziation untersucht. Anschließend wurden die 500 am stärksten assoziierten Bins selektiert. Aus diesen 500 NMR-Bins wurden die paarweisen Ratios errechnet, welche in einer zweiten GWAS (GWAS 2) getestet wurden. MS-bestimmte Serum-Metabolitenkonzentrationen derselben 1.757 KORA-F4-Probanden dienten der Interpretation der in der GWAS identifizierten NMR-Merkmale. Abbildung nach Inouye und Abraham [78].

Assoziationen zu NMR-Merkmalen

Die Suche nach Assoziationen zwischen genetischen Varianten und NMR-Signalen erfolgte in zwei Schritten (Abbildung 5.1):

Beim ersten Schritt wurde eine GWAS mit den Intensitätswerten der ^1H -NMR Spektren von 1.757 Blutplasma-proben durchgeführt. Zu diesem Zweck wurde jedes Spektrum auf die Spektralregion zwischen 0 ppm und 9 ppm beschränkt und in gleich große Segmente (*Bins*) von jeweils 0,001 ppm Breite aufgeteilt. Die zu einem Bin zusammengefassten Intensitätswerte wurden arithmetisch gemittelt. Bins zwischen $\delta = 4,6$ und $\delta = 5,0$ ppm wurden für die weitere Analyse verworfen, da diese Spektralregion hauptsächlich vom Wasserpeak dominiert wird, dessen markantes Signal zu Verzerrungen anderer Signale führen kann. Schließlich wurden die Intensitätswerte

der Bins logarithmiert und um extreme Ausreißer bereinigt. Dieser erste Schritt unterscheidet sich von den meisten auf NMR-Messungen basierenden GWAS dadurch, dass nahezu die kompletten und nur gering datenreduzierten Spektren als Phänotyp verwendet werden. Üblicherweise erfolgt eine Vorselektierung spektraler Merkmale entweder durch Extraktion einzelner prominenter Peaks mittels eines Peak-Picking-Algorithmus' oder durch Annotation der Spektren zur Reduzierung auf die Konzentrationen identifizierbarer Substanzen [38, 40, 53]. Letztere Herangehensweise wurde beispielsweise bei den anderen in dieser Dissertation besprochenen GWAS mit metabolischen Merkmalen im Urin angewandt.

Im zweiten Schritt wurden diejenigen 500 Bins selektiert, welche in der beim ersten Schritt durchgeführten GWAS am stärksten mit genetischen Varianten assoziierten. Dabei wurde für den Fall, dass benachbarte Bins vergleichbar starke Assoziationen mit dem selben SNP zeigten, stets der Bin mit dem niedrigsten p-Wert ausgewählt. In einer zweiten GWAS wurden dann alle möglichen paarweisen Ratios zwischen diesen 500 Bins auf Assoziationen hin getestet. Soweit dem Autor bekannt, ist dies die erste GWAS, die das Konzept der Ratios auf NMR-Rohdaten anwendet.

Für beide GWAS gilt gleichermaßen, dass Alter und Geschlecht der Probanden als Kovariaten in die Berechnung der linearen Modelle einbezogen wurden. Bei der Interpretation der Ergebnisse wurden nur SNPs betrachtet, bei denen das seltenere Allel mit einer Häufigkeit von mindestens 10 % in der Studienpopulation vorliegt ($MAF \geq 10\%$).

Insgesamt zeigen sieben genetische Loci (LIPC, CETP, FADS1, GCKR, APOA1, CPS1, PYROXD2) Assoziationssignale zu NMR-Merkmalen (Bins oder Ratios) mit p-Werten unterhalb des genomweiten Signifikanzniveaus ($\alpha' = 5,72 \times 10^{-13}$). Dies entspricht einem Signifikanzniveau von $\alpha = 0,05$ nach einer Bonferroni-Korrektur für 655.658 getestete SNPs und 133.350 aus den NMR-Spektren abgeleitete Merkmale. Letztere Zahl setzt sich zusammen aus $500 \times 499 / 2$ getesteten Ratios und 8.600 getesteten Bins. Tabelle 5.6 führt alle Loci auf, die signifikant mit NMR Bins oder Ratios assoziieren, außerdem jeweils den SNP, der in dem betreffenden Locus die stärkste Assoziation zeigt, sowie die Stärke der Assoziation und deren Effekt. Im Anhang dieser Arbeit sind für jeden Locus die entsprechenden regionalen Assoziationsplots, Boxplots sowie Q-Q Plots zu finden (Abbildung B.1).

Mit Ausnahme des PYROXD2-Locus' wurde jeder Locus in den 2010 und 2011 in KORA F4 durchgeführten GWAS mit MS-quantifizierten Metabolitenkonzentratio-

nen identifiziert und jeweils in einer unabhängigen Kohorte repliziert [34, 35]. PYROXD2 wurde erstmalig 2011 von Nicholson *et al.* mittels NMR-Metabolomik identifiziert [40]. Zwischenzeitlich konnte aber auch dieser Locus in einer GWAS mit biochemisch uncharakterisierten, per MS gemessenen Metaboliten in der KORA-F4-Population bestätigt werden [79].

Überschneidungen mit NMR-bestimmten Lipid-Subklassen

Im Jahre 2012 veröffentlichten Petersen *et al.* eine Assoziationsstudie mit aus NMR-Spektren abgeleiteten Lipoprotein-Subklassen [67]. Die Studie basiert auf denselben NMR-Rohdaten, welche auch in der hier besprochenen Studie verwendet wurden. Der Unterschied zwischen beiden Studien liegt in der Art der Auswertung der Rohdaten. In der vorliegenden Studie wurde eine *non-targeted*-Herangehensweise verwendet, bei dem die Untersuchung nicht auf bestimmte Metaboliten begrenzt ist. Daher wurden die Spektren nicht vorab annotiert. Petersen *et al.* wählten jedoch einen *targeted*-Ansatz, da sie sich auf eine vorab definierte Klasse metabolischer Merkmale beschränkten. Dabei wurden anhand der NMR-Signale der Spektralregion zwischen $\delta = 0,6$ ppm und $\delta = 1,5$ ppm die Konzentrationen von in 15 Subklassen eingeteilten Lipoproteine abgeleitet. Außerdem testeten Petersen *et al.* in ihrer Studie lediglich 101 SNPs in bekannten Lipid-Loci auf Assoziation zu den Lipoprotein-Subklassen, im Gegensatz zu dem genomweiten Ansatz, der in der hier besprochenen Studie verwendet wurde. Petersen *et al.* konnten zeigen, dass acht genetische Loci spezifisch mit einer oder mehreren Lipidklassen assoziieren.

Von diesen acht Loci konnten in der hier vorgestellten Studie fünf identifiziert werden, die mit NMR-Bins an unterschiedlichen Positionen im Spektrum assoziieren (LIPC, CETP, FADS1, GCKR, APOA1). Beim Vergleich der Assoziationsstärken fällt auf, dass die Assoziationen zu einzelnen Bins im Falle des APOA1- sowie des LIPC-Locus etwas schwächer sind als die Assoziationen zu den Lipid-Subklassen. Eine mögliche Erklärung hierfür ist, dass in die Bestimmung einer Lipid-Subklasse jeweils mehrere aus den Spektren abgeleitete Parameter einfließen. Im ersten GWAS-Durchgang der hier besprochenen Studie wird aber jeweils nur ein Parameter, nämlich ein einzelner Bin, auf genetische Assoziation hin getestet. Bei der zweiten GWAS werden hingegen bi-parametrische Werte verwendet, nämlich die aus jeweils zwei unterschiedlichen Bins berechneten Ratios. Bei vier der fünf Loci (LIPC, CETP, FADS1, GCKR) ergibt sich ein markanter Anstieg in der Assoziationsstärke: die p-Werte sind im Vergleich

zu den Assoziationen mit Lipid-Subklassen zwischen 9 und 89 Größenordnungen niedriger (Tabelle 5.7). In diesen vier Fällen scheinen also Ratios ein besseres Abbild der durch genetische Polymorphismen beeinflussten biologischen Parameter zu liefern als NMR-Bins oder Lipidklassen.

Tabelle 5.7.: GWAS mit NMR-Merkmalen im Plasma: Vergleich der Assoziationsstärken zu NMR-bestimmten Lipoprotein-Subklassen, NMR-Bins und NMR-Ratios

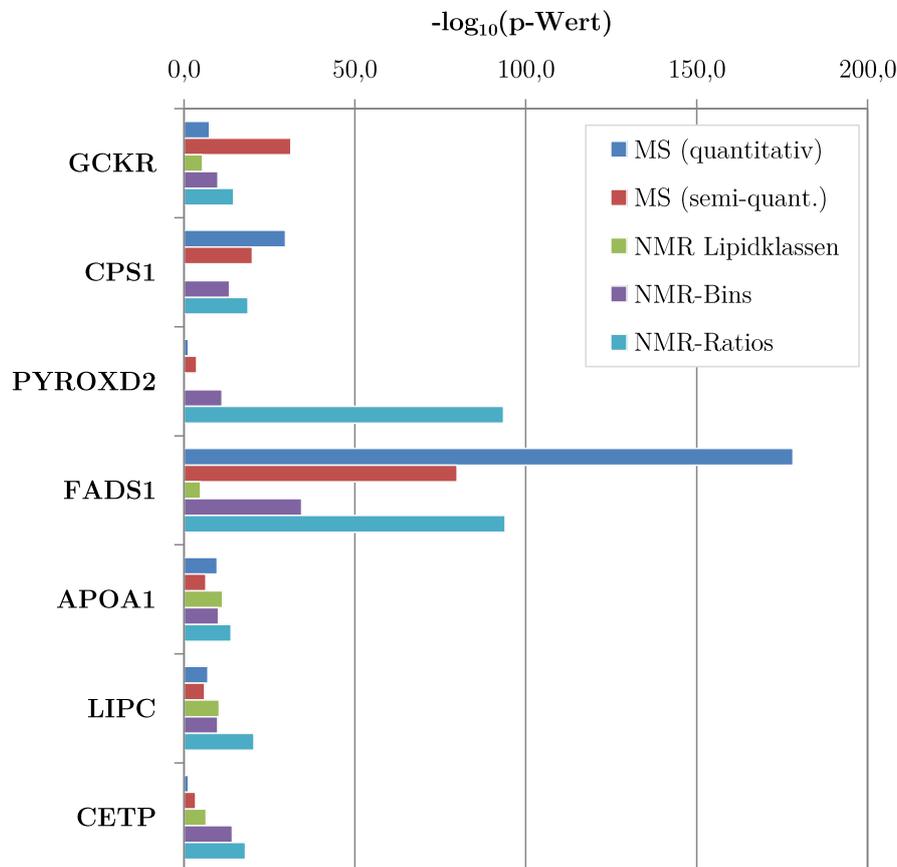
Locus		NMR Lipid-Subklassen	NMR Bins	NMR Ratios
GCKR	Merkmal	L10	1,370 ppm	3,286 ppm / 1,370 ppm
	SNP	rs1260326 (LD $r^2 = 0,93$)	rs780094	rs780094
	p-Wert	$p < 3,7 \times 10^{-6}$	$p < 1,2 \times 10^{-10}$	$p < 2,8 \times 10^{-15}$
FADS1	Merkmal	L4	2,801 ppm	2,801 ppm / 2,017 ppm
	SNP	rs174546 (LD $r^2 = 1,0$)	rs174547	rs174547
	p-Wert	$p < 1,4 \times 10^{-5}$	$p < 4,0 \times 10^{-35}$	$p < 1,1 \times 10^{-94}$
APOA1	Merkmal	L8	2,038	4,162 ppm / 4,082 ppm
	SNP	rs964184 (LD $r^2 = 0,61$)	rs3741298	rs3741298
	p-Wert	$p < 4,8 \times 10^{-12}$	$p < 8,4 \times 10^{-11}$	$p < 1,8 \times 10^{-14}$
LIPC	Merkmal	L5	1,283 ppm	1,068 ppm / 1.029 ppm
	SNP	rs1532085 (LD $r^2 = 0,55$)	rs4775041	rs4775041
	p-Wert	$p < 5,3 \times 10^{-11}$	$p < 1,4 \times 10^{-10}$	$p < 3,6 \times 10^{-21}$
CETP	Merkmal	L3	3,259 ppm	2.211 ppm / 2.011 ppm
	SNP	rs3764261 (LD $r^2 = 1,0$)	rs247617	rs247617
	p-Wert	$p < 3,6 \times 10^{-7}$	$p < 7,6 \times 10^{-15}$	$p < 1,1 \times 10^{-18}$

Aufgeführt diejenigen fünf Loci aus Petersen et al. [67], welche auch in der hier vorgestellten Studie identifiziert wurden. Petersen *et al.* testeten lediglich 101 SNPs auf genetische Assoziation, im Gegensatz zum genomweiten Ansatz der gegenwärtigen Studie. Es wurde jeweils der SNP mit dem stärksten Assoziationssignal ausgewählt, den Petersen *et al.* auf dem jeweiligen Locus identifizierten. Der LD-Wert gibt dabei die Korrelation (r^2) zu dem SNP an, der in der hier besprochenen Studie die stärkste Assoziation zeigte. Die LD-Werte basieren auf den Daten aus HapMap (Version #27, Phase I-III, CEU Population) [72].

Vergleich mit früheren Assoziationsstudien in KORA F4

Die p-Werte der Assoziationen genetischer Loci zu NMR-Merkmalen wurden mit Ergebnissen aus früheren Assoziationsstudien, die mit Daten aus komplementären Metabolomik-Plattformen in Blutproben der KORA F4 Kohorte durchgeführt wurden, verglichen. In vier von sieben Fällen (PYROXD2, APOA1, LIPC, CETP) sind die Assoziationen zu NMR-Merkmalen (Bins und/oder Ratios) stärker (Tabelle 5.8 und Abbildung 5.2). In den anderen drei Fällen (FADS1, GCKR sowie CPS1) liefern die MS-basierten Plattformen deutlich bessere Ergebnisse.

Abbildung 5.2.: Vergleich mit Assoziationsdaten aus früheren Studien mit Blutproben des KORA-F4-Datensatzes



Graphische Darstellung der p-Werte aus anderen Assoziationsstudien, die mit metabolischen Merkmalen im Blut der KORA-F4-Probanden durchgeführt wurden (siehe Tabelle 5.8). Die Balkenlänge entspricht der Stärke der in der jeweiligen Studie identifizierte Assoziation ($-\log_{10}(\text{p-Wert})$). Die Ergebnisse für quantitative MS stammen aus mGWAS von Gieger *et al.* [30] sowie Illig *et al.* [34], die für semi-quantitative MS aus einer von Suhre *et al.* [35] veröffentlichten mGWAS. Die Assoziationsdaten für NMR Lipid-Subklassen wurden von Petersen *et al.* [67] publiziert.

Tabelle 5.8.: GWAS mit NMR-Merkmalen im Plasma: Assoziationsergebnisse aus früheren Studien mit Blutproben aus KORA F4

Locus	MS (quantitativ) [30, 34]	MS (semi-quantitativ) [35]	NMR Lipid-Subklassen [67]	NMR-Bins	NMR-Ratios
GCKR	Merkmal	PC ae C34:2 / PC aa C32:2	L10	1,370 ppm	3,286 ppm / 1,370 ppm
	SNP	rs1260326 (LD $r^2 = 0,93$)	rs1260326 (LD $r^2 = 0,93$)	rs780094	rs780094
	p-Wert	$p < 3,8 \times 10^{-8}$	$p < 4,9 \times 10^{-32}$	$p < 1,2 \times 10^{-10}$	$p < 2,8 \times 10^{-15}$
CPS1	Merkmal	Glycin / PC aa C38:2	<i>keine Assoziation</i>	3,599 ppm	3,599 ppm / 2,475 ppm
	SNP	rs2216405	<i>keine Assoziation</i>	rs2216405	rs2216405
	p-Wert	$p < 1,9 \times 10^{-30}$	$p < 9,8 \times 10^{-21}$	$p < 4,5 \times 10^{-14}$	$p < 1,8 \times 10^{-19}$
PYROXD2	Merkmal	Saccharin / Threonin	<i>keine Assoziation</i>	2,757 ppm	2,757 ppm / 2,755 ppm
	SNP	<i>keine Assoziation</i>	<i>keine Assoziation</i>	rs4488133	rs4488133
	p-Wert	$p < 0,00021$		$p < 7,3 \times 10^{-12}$	$p < 2,9 \times 10^{-94}$
FADS1	Merkmal	PC aa C36:3 / PC aa C36:4	L4	2,801 ppm	2,801 ppm / 2,017 ppm
	SNP	rs174547	rs174546 (LD $r^2 = 1,0$)	rs174547	rs174547
	p-Wert	$p < 6,5 \times 10^{-179}$	$p < 1,2 \times 10^{-80}$	$p < 4,0 \times 10^{-35}$	$p < 1,1 \times 10^{-94}$
APOA1	Merkmal	PC aa C36:2 / PC aa C38:1	L8	2,038	4,162 ppm / 4,082 ppm
	SNP	rs964184 (LD $r^2 = 0,61$)	rs964184 (LD $r^2 = 0,61$)	rs3741298	rs3741298
	p-Wert	$p < 1,8 \times 10^{-10}$	$p < 4,3 \times 10^{-7}$	$p < 8,4 \times 10^{-11}$	$p < 1,8 \times 10^{-14}$
LIPC	Merkmal	PC aa C38:6	L5	1,283 ppm	1,068 ppm / 1,029 ppm
	SNP	rs4775041	rs1532085 (LD $r^2 = 0,55$)	rs4775041	rs4775041
	p-Wert	$p < 9,7 \times 10^{-8}$	$p < 8,7 \times 10^{-7}$	$p < 1,4 \times 10^{-10}$	$p < 3,6 \times 10^{-21}$
CETP	Merkmal	Guanison / Phenylacetylglutamin	L3	3,259 ppm	2,211 ppm / 2,011 ppm
	SNP	<i>keine Assoziation</i>	rs3764261 (LD $r^2 = 1,0$)	rs247617	rs247617
	p-Wert	$p < 0,00039$	$p < 3,6 \times 10^{-11}$	$p < 7,6 \times 10^{-15}$	$p < 1,1 \times 10^{-18}$

Aufgeführt sind die Assoziationsergebnisse aus vorhergehenden Assoziationsstudien für die Loci, die in der hier besprochenen Studie identifiziert wurden. Falls in einer Studie ein anderer SNP für den jeweiligen Locus publiziert wurde, ist die Korrelation (r^2) zum in dieser Studie identifizierten SNP angegeben. Diese LD-Werte basieren auf den Angaben aus HapMap (Version 27, Phase I-III, CEU Population) [72].

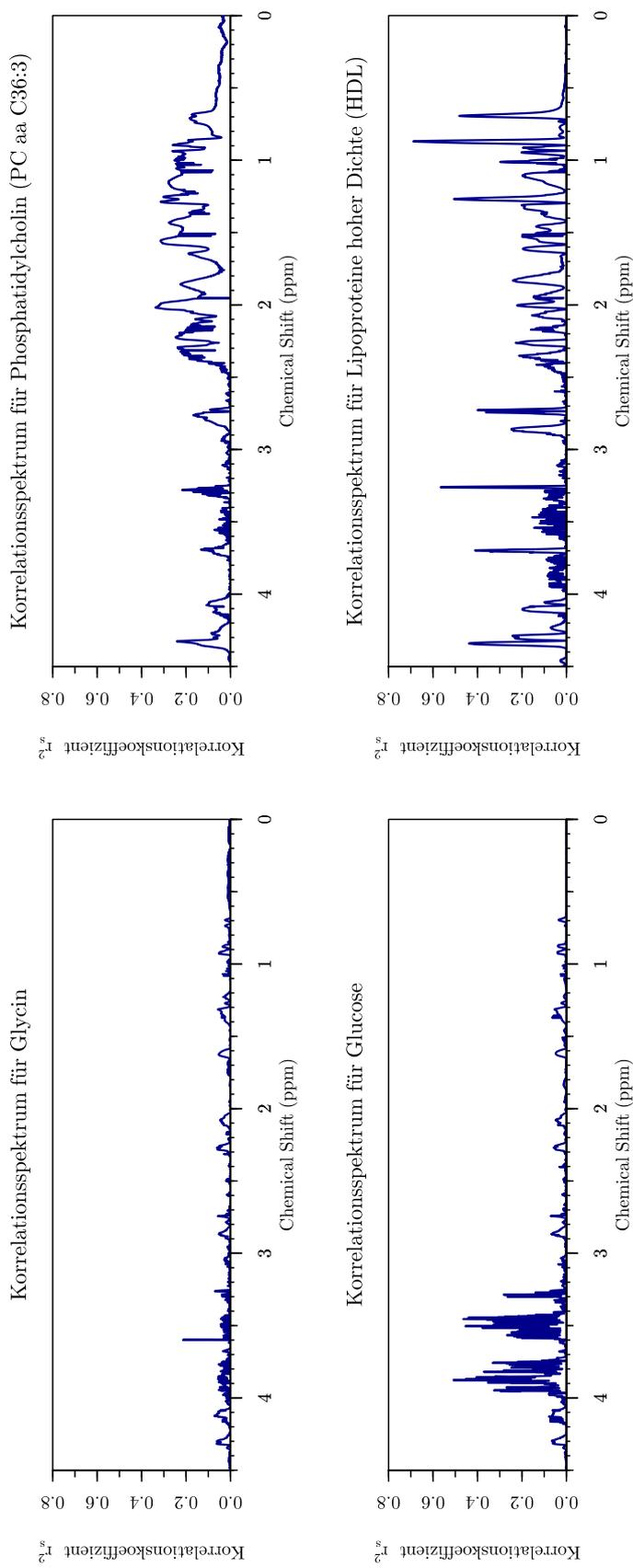
Korrelation von NMR-Signalen zu massenspektrometrisch und biochemisch bestimmten Metabolitenkonzentrationen

Das ^1H -NMR Spektrum einer biologischen Probe zeigt die Überlagerung der Resonanzspektren aller messbaren Substanzen. Prinzipbedingt kommt es dadurch an vielen Stellen im Spektrum zur Überlappung von Signalen, die unterschiedlichen Metaboliten zuzuordnen sind [80]. Dies erschwert die Annotation und Quantifizierung von einzelnen Metaboliten in NMR Spektren erheblich. In dieser Studie wurden Korrelationsspektren verwendet, um den Grad der Korrelation zwischen NMR-Signalen (über den gesamten Messbereich) und Metabolitenkonzentrationen, die auf MS-Plattformen sowie anhand klinischer Biochemie quantifiziert wurden, zu bestimmen und zu visualisieren. Ermöglicht wird dieser Ansatz dadurch, dass alle Messungen (NMR, MS und biochemisch) auf jeweils den gleichen Blutproben eines jeden der 1.757 Probanden durchgeführt wurde. Diese Korrelationsspektren stellen deshalb eine Verbindung zwischen der chemischen Identität von Metaboliten, die auf anderen Plattformen gemessen wurden, und Signalen an bestimmten Positionen in den NMR-Spektren (*Chemical Shifts*) her. Wenn ein Chemical Shift mit einem durch MS quantifizierten Metaboliten korreliert, kann dies einerseits bedeuten, dass der Chemical Shift Teil des Resonanzspektrum dieses Metaboliten ist. Andererseits ist es aber auch möglich, dass der Chemical Shift nur indirekt mit dem Metaboliten in Verbindung steht - es könnte auch ein Metabolit sein, der auf dem gleichen Stoffwechselfad liegt.

Als Beispiel zeigt Abbildung 5.3 das Korrelationsspektrum der NMR-Signale zu Glycin, dessen Konzentrationen durch MS gemessen wurde. Dieses Spektrum zeigt einen deutlichen, scharf abgegrenzten Peak bei $\delta = 3,599$ ppm. Ein von der HMDB [4] abgerufenes Spektrum lässt vermuten, dass es sich bei diesem Peak tatsächlich um das Resonanzsignal von Glycin handelt. Sicherheit gab letztlich ein zusätzliches Spiking-Experiment, bei dem zwei zufällig ausgewählte Proben erneut durch NMR vermessen wurden. Eine dieser Proben war mit Glycin versetzt. Dieses Experiment zeigte, dass das hoch mit Glycin korrelierte NMR Signal auch tatsächlich Glycin zuzuordnen ist (Abbildung 6.19).

Zudem gibt es hohe Korrelationen zwischen NMR-Signalen und per MS bestimmten Phosphatidylcholinen (PCs) und zu Lipid-Parametern, die mittels standardisierter, klinischer biochemischer Methoden bestimmt wurden. Im Einzelnen sind dies Triglyceride (TG), Lipoproteine hoher Dichte (*high-density lipoprotein*; HDL), Lipopro-

Abbildung 5.3.: Korrelationen zwischen NMR-Signalen und den massenspektrometrisch bzw. biochemisch ermittelten Konzentrationen von Glycin, PC aa C36:3, HDL und Glucose



Die Linie zeigt den quadrierten Spearman-Korrelationskoeffizienten zwischen dem jeweiligen massenspektrometrisch oder biochemisch quantifizierten Metaboliten und den NMR-Signalen zwischen $\delta = 0$ ppm und $\delta = 4,5$ ppm. Metaboliten, die in Beziehung mit dem Lipidstoffwechsel stehen, korrelieren im Allgemeinen mit breiteren Abschnitten des NMR-Spektrums, wie am Beispiel von PC aa C36:3 und HDL gezeigt. Bei anderen Metaboliten zeigt das jeweilige Korrelationspektrum scharfe, deutlich abgegrenzte Peaks (Beispiel: Glycin und Glucose), die in erster Linie den tatsächlichen NMR-Peaks der jeweiligen Reinstsubstanzen entsprechen (vgl. Abbildung 6.19).

teine niedriger Dichte (*low-density lipoprotein*; LDL), Cholesterin (*total cholesterol*; TC) sowie Glucose. Beispiele für Korrelationsspektren mit Phosphatidylcholinen, Lipoproteinen und Zuckern sind in Abbildung 5.3 zu sehen.

Zur besseren Übersichtlichkeit zeigen diese Korrelationsspektren stets den quadrierten Koeffizienten der Spearman-Korrelation. Die meisten der beobachteten Korrelationen sind positiv. Dies bedeutet, dass ein stärkeres NMR-Signal an einer gegebenen Stelle im Spektrum auch mit einer höheren Konzentration des korrelierten Metaboliten einhergeht. In den Fällen, in denen MS und NMR auf den gleichen oder eng verwandte Metaboliten abzielen, ist dies auch zu erwarten. Es gibt aber auch einige wenige Fälle, in denen die NMR-Signale negativ korreliert (antikorreliert) sind. Dies mag ein Hinweis darauf sein, dass hier das NMR-Signal indirekt mit dem durch MS gemessenen Metaboliten in Beziehung steht, beispielsweise als Produkt oder Substrat einer enzymatisch gesteuerten Reaktion.

Aus Abbildung 5.3 wird ersichtlich, dass bei Parametern mit Bezug zum Lipidstoffwechsel weite Bereiche des NMR-Spektrums mit eben diesen Metaboliten korrespondieren. Korrelationen mit Metaboliten, die nicht im direkten Zusammenhang mit Lipiden stehen, zeigen hingegen im Korrelationsspektrum einzelne, scharf abgegrenzte Peaks. In dieser Studie konnten deutliche Korrelationen zwischen solchen Peaks und den Konzentrationen von Glucose, Lactat, Prolin, Glycin und etlichen weiteren Metaboliten identifiziert werden. Der quadrierte Korrelationskoeffizient liegt dabei zwischen $r_s^2 = 0,33$ und $r_s^2 = 0,65$. In den meisten Fällen konnte bei Metaboliten dieser Art mit Hilfe von experimentell ermittelten Spektren der Reinsubstanzen aus der HMDB bestätigt werden, dass die Korrelationsspektren die jeweiligen Metaboliten repräsentieren. Wie bereits erwähnt konnte für Glycin durch ein Spiking-Experiment die biochemische Gültigkeit der Korrelation zusätzlich bestätigt werden.

Bei der Interpretation der Korrelationsspektren ist zu beachten, dass aufgrund der Tatsache, dass viele Metaboliten in metabolischen Netzwerken miteinander verknüpft sind, einige Signalkorrelationen nicht nur repräsentativ für einzelne Metaboliten, sondern auch für eine Mischung von mehreren co-regulierten Stoffwechselprodukten sein können. Auch ist es möglich, dass mehrere, nicht miteinander verwandte Metaboliten Korrelationssignale an der selben Stelle im NMR-Spektrum erzeugen, da sich in $^1\text{H-NMR}$ Spektren Signale prinzipiell überlappen können.

Im Anhang dieser Arbeit wird die jeweils stärkste Korrelation zwischen Metabolitenkonzentrationen und NMR-Bins mit $r_s^2 \geq 0,20$ aufgeführt (Tabelle B.1).

In der hier besprochenen mGWAS wurden die berechneten Korrelationsspektren zur Interpretation der Assoziationen genetischer Merkmale zur NMR-Merkmalen verwendet. Für die jeweils signifikanteste Assoziation in einem Locus wird in Tabelle 5.9 der Metabolit genannt, welcher an der jeweiligen Stelle im NMR-Spektrum am stärksten korreliert. Bei genetischen Assoziationen zu NMR-Ratios wird entweder die Korrelation zu einem einzelnen Metaboliten oder zu einem Metaboliten-Ratio genannt, je nachdem, welche Korrelation stärker ist. Hintergrund dieser Vorgehensweise ist, dass Signale an zwei unterschiedlichen Stellen im NMR-Spektrum nicht nur einen einzelnen Metaboliten, sondern auch zwei unterschiedliche Metaboliten repräsentieren können. Hier kann eine hohe Korrelation zwischen dem NMR-Ratio und dem Ratio, der aus biochemisch oder massenspektrometrisch bestimmten Metaboliten errechnet wurde, auf einen biochemischen Zusammenhang hindeuten.

Tabelle 5.9.: GWAS mit NMR-Merkmalen: Korrelationen zwischen NMR-Bins bzw. -Ratios und metabolischen Merkmalen

Locus	SNP	NMR-Bin			NMR-Ratio		
		Chemical Shift	Metabolit	r_s^2	Chemical Shifts	Metabolit oder Ratio	r_s^2
GCKR	rs780094	1,370 ppm	Triglyceride	0,62	3,286 ppm / 1,370 ppm	Glucose / Triglyceride	0,42
CPS1	rs2216405	3,599 ppm	Glycin	0,22	3,599 ppm / 2,475 ppm	Glycin / Tryptophan	0,29
PYROXD2	rs4488133	2,757 ppm	Gesamt Cholesterin	0,64	2,757 ppm / 2,755 ppm	Gesamt Cholesterin / Triglyceride	0,26
FADS1	rs174547	2,801 ppm	PC aa C38:5	0,48	2,801 ppm / 2,017 ppm	PC aa C38:5 / PC aa C36:3	0,45
APOA1	rs3741298	2,038 ppm	Triglyceride	0,73	4,162 ppm / 4,082 ppm	Lactat / Triglyceride	0,60
LIPC	rs4775041	1,283 ppm	PC aa C36:1	0,42	1,068 ppm / 1,029 ppm	4-Methyl-2- oxopentanoat / Gesamt Cholesterin	0,21
CETP	rs247617	3,259 ppm	HDL- Cholesterin	0,55	2,211 ppm / 2,011 ppm	HDL- Cholesterin	0,68

Die Spalten „Metabolit“ bzw. „Metabolit oder Ratio“ nennen den massenspektrometrisch oder biochemisch bestimmten Metaboliten oder den daraus errechneten Ratio, der am stärksten mit dem NMR-Bin bzw. dem NMR-Ratio, der in der GWAS die beste Assoziation zeigte, korreliert. r_s^2 nennt den quadrierten Spearman-Korrelationskoeffizienten.

6. Diskussion

In dieser Arbeit werden drei Studien behandelt, in denen NMR-bestimmte metabolische Merkmale im Urin und Blutplasma genomweit auf Assoziation mit genetischen Polymorphismen untersucht werden.

In zwei GWAS mit Stoffwechselprodukten im Urin konnten 14 Loci identifiziert werden, welche genomweit signifikant mit Metaboliten oder daraus abgeleiteten paarweisen Konzentrationsverhältnissen (Ratios) assoziieren. Fünf dieser Loci wurden in der ersten mGWAS im Urin identifiziert, neun weitere wurden der zweiten mGWAS im Urin identifiziert, in der die untersuchte Studienpopulation mehr als vervierfacht wurde.

In diesem Kapitel werden diese Loci vorgestellt und auf ihre Bedeutung für den Stoffwechsel und klinische Phänotypen, insbesondere die Nierenfunktion, untersucht, soweit dies aus der Literatur bekannt ist. Weiterhin wird diskutiert, inwiefern ein funktioneller Zusammenhang zwischen den Assoziationen und der Funktion der jeweiligen Genprodukte hergestellt werden kann. Die Interpretation der Ergebnisse der ersten mGWAS im Urin folgt weitestgehend der Diskussion der zugehörigen Publikation, die 2011 von Suhre *et al.* veröffentlicht wurde [53].

Bei der Interpretation der Ergebnisse der zweiten, bislang nicht publizierten Urin-GWAS ist der Fokus nicht allein darauf gerichtet, neue Assoziationen zwischen genetischen Polymorphismen und im Urin quantifizierten Stoffwechselprodukten aufzudecken; es sollen zudem Parallelen zu Assoziationsstudien mit metabolischen Merkmalen im Blutserum gezogen werden. Hierzu dienen vorrangig die Ergebnisse der 2011 von Suhre *et al.* veröffentlichten mGWAS [35] sowie die Assoziationsdaten einer 2014 von Shin *et al.* publizierten Studie [43], die auf insgesamt über 8.000 Proben aus der KORA-F4- sowie der TwinsUK-Kohorten beruht. Insbesondere soll gezeigt werden, inwiefern sich im Urin identifizierte GIMs auch im Blut sichtbar sind, oder

ob genetische Loci im Blut und im Urin mit nicht näher zueinander verwandten metabolischen Merkmalen assoziieren.

Abschließend werden die Ergebnisse einer GWAS mit aus NMR-Spektren abgeleiteten Merkmalen im Blutplasma besprochen. Insgesamt sieben Loci zeigen genomweit signifikante Assoziationen zu NMR-Bins oder daraus abgeleiteten NMR-Ratios. Die Diskussion der Loci folgt der Publikation von 2013 [54] und betont diejenigen Aspekte, welche die Besonderheiten der signifikant assoziierten NMR-Merkmale aufzeigen.

6.1. Die erste GWAS mit Stoffwechselprodukten im Urin

In der ersten mGWAS im Urin von 862 männlichen Teilnehmern der SHIP-0-Studie wurden genomweit signifikante Assoziationen von SNPs in fünf Loci entdeckt. Diese fünf Loci liegen im Bereich der Gene AGXT2, NAT2, HPD, SLC7A9 sowie SLC6A20. Bei zwei Loci liegt der jeweils am stärksten assoziierte SNP im Exon von Genen, welche für ein Enzym (AGTX2) und einen Transporter (SLC6A20) codieren. Drei Loci sind im LD mit Genen, welche einen Transporter (SLC7A9) sowie Enzyme (HPD, NAT2) kodieren. Da diese Studie zum Zeitpunkt ihrer Veröffentlichung die erste GWAS mit Stoffwechselprodukten im Urin war, gab es bis dato keine bekannten Assoziationen dieser Loci zu Metaboliten im Harn.

Ein Polymorphismus im AGXT2-Gen ist mutmaßlich die genetische Grundlage für beta-Aminoisobutansäure-Azidurie

Im Jahre 1951 wurde erstmals von Crumpler *et al.* beta-Aminoisobutansäure (beta-Aminoisobutyrate; BAIB) im menschlichen Urin identifiziert [81]. Anschließend wurde zeitnah in mehreren Studien gezeigt, dass die Ausscheidung von BAIB innerhalb der Bevölkerung stark variiert. So fanden Harris *et al.* durch Untersuchungen an 345 gesunden britischen Testpersonen heraus, dass bei ca. 10 % der Probanden eine erhöhte BAIB-Konzentration im Urin vorlag [82]. Familienstudien bestätigten, dass die Bevölkerung aufgrund unbekannter genetischer Faktoren in „Niedrigausscheider“ und „Hochausscheider“ unterteilt ist [83, 84]. Man vermutete einen autosomal rezes-

siven Erbgang für diesen Metabotypen, wobei bei homozygoten Trägern des seltenen Allels die Konzentration der mit dem Urin ausgeschiedenen BAIB am höchsten ist [83, 85, 86].

Die OMIM-Datenbank (Online Mendelian Inheritance in Man) bezeichnet im Eintrag %210100 beta-Aminoisobutansäure-Azidurie als die am häufigsten in der Bevölkerung zu beobachtende metabolische Variante, welche den Mendelschen Vererbungsregeln folgt [87]. Bislang wurde die molekulargenetische Basis für den BAIB-Polymorphismus allerdings noch nicht identifiziert.

In der hier vorgestellten mGWAS zeigt SNP rs37369 eine starke Assoziation zu BAIB bei einem p-Wert kleiner $3,17 \times 10^{-75}$ (Tabelle 5.1). Bemerkenswert ist auch die extrem starke Zunahme der BAIB-Konzentration pro Kopie des seltenen SNP-Allels ($\beta = 2,396$; Abbildung 6.3(a)). SNP rs37369 liegt auf dem 5. Chromosom im exonischen Bereich eines Gens, welches für die Alanin-Glyoxylat-Aminotransferase 2 (AGXT2) codiert (Abbildung 6.3(b)). AGXT2 ist eine mitochondriale Aminotransferase, welche laut Illumina BodyMap 2.0 hauptsächlich in der Leber und den Nieren exprimiert wird [75]. Das Enzym katalysiert unter anderem die reversible Umwandlung von BAIB mit Pyruvat zu 2-Methyl-3-oxopropanoat und L-Alanin (EC 2.6.1.40; Abbildung 6.1) [88].

Abbildung 6.1.: Enzymatische Reaktion von AGXT2 (EC 2.6.1.40)

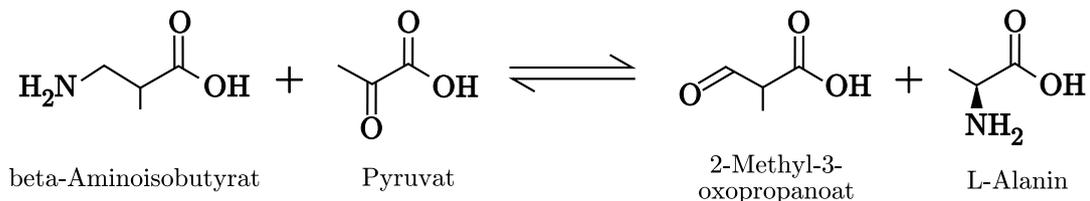


Abbildung nach KEGG Reaktion R02050 [89].

Bereits 1968 vermuteten Kakimoto *et al.*, dass eine erhöhte BAIB-Ausscheidung auf eine verminderte oder gar nicht vorhandene Funktion dieses Enzyms zurückzuführen sei [86]. Es blieb jedoch unklar, welcher genetischer Polymorphismus ursächlich für den beobachteten metabolischen Polymorphismus ist. Die hier entdeckte Assoziation von rs37369 im AGXT2-Locus zur BAIB-Konzentration im Urin passt zur experimentell bestimmten enzymatischen Funktion des AGXT2-Genprodukts. Dieser SNP hat zudem einen nicht-synonymen Austausch eines Aminosäurerestes im Protein zur Folge (V140I), wodurch mutmaßlich die enzymatische Funktion beeinträchtigt wird. Ein experimenteller Nachweis dieser Hypothese steht zwar noch aus, doch aufgrund

des Aminosäureaustausches und des außergewöhnlich starken Effekts scheint es jedoch plausibel, dass rs37369 derjenige kausative SNP ist, welcher zur BAIB-Azidurie führt.

Nur wenige Monate nach Veröffentlichung der hier besprochenen Studie publizierten Nicholson *et al.* eine GWAS basierend auf Blut- und Urinproben aus den MolTWIN- und MolOBB-Kohortenstudien [40]. Dort wurden die Metabolitenkonzentrationen ebenfalls mittels $^1\text{H-NMR}$ Spektroskopie gemessen, obgleich ein anderes Annotationsverfahren verwendet wurde. Der starke Effekt des rs37369-Polymorphismus auf die BAIB-Konzentration in Urin konnte auch dort gezeigt werden.

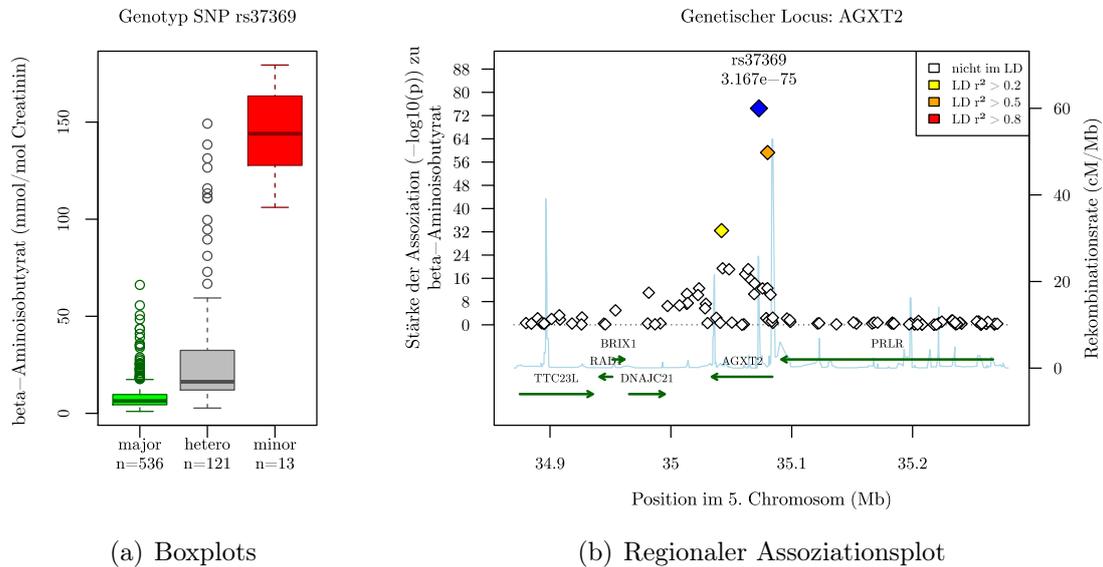
Jüngere Studien zeigen, dass das AGXT2-Genprodukt auch asymmetrisches Dimethylarginin (ADMA) metabolisiert [91–93]. ADMA spielt eine Rolle bei der Produktion von Stickstoffmonoxid, und erhöhte ADMA-Konzentrationen im Plasma assoziieren unter anderem mit Bluthochdruck [94], Herzinsuffizienz [95], chronischem Nierenleiden [96, 97] sowie Arteriosklerose [98]. In einer 2012 veröffentlichten Studie beobachteten Caplin *et al.* eine schwache, aber nicht genomweit signifikante Assoziation zwischen dem seltenen rs37369-Allel und erhöhtem diastolischem Blutdruck [99]. Kittel *et al.* zeigten 2013 in einer experimentellen Studie, dass ein Überschuss von BAIB im Blut die Katalyse von ADMA durch AGXT2 unterdrückt [92].

Sowohl der AGXT2-Locus (hier vor allem SNP rs37369) als auch das assoziierte BAIB sind also von klinischer Bedeutung und sollten daher in weiteren Studien untersucht werden.

Der Acetylierungs-Phänotyp des NAT2 Polymorphismus' korrespondiert mit den Konzentrationen von Formiat, Succinat und Acetat

In der Nähe des Gens der N-Acetyltransferase 2 (NAT2) liegt SNP rs4921914. Er ist ca. 14 kB vom 3'-Ende des Gens entfernt und ist im LD mit SNPs, welche im transkribierten Bereich des NAT2-Gens liegen (Abbildung 6.3 (b)). Im *SHIP-0 males* Datensatz zeigt rs4921914 die stärkste Assoziation zu dem Formiat/Succinat-Ratio ($p < 5,1 \times 10^{-16}$; Tabelle 5.1 sowie Abbildung 6.3). Etwas schwächer, aber dennoch genomweit signifikant sind die Assoziationen zu den Ratios Formiat/Harnstoff sowie Formiat/Acetat ($p < 1,1 \times 10^{-15}$ bzw. $9,0 \times 10^{-15}$; Tabelle 6.1).

Abbildung 6.2.: Assoziation von rs37369 (AGXT2) zur beta-Aminoisobutyrat-Konzentration



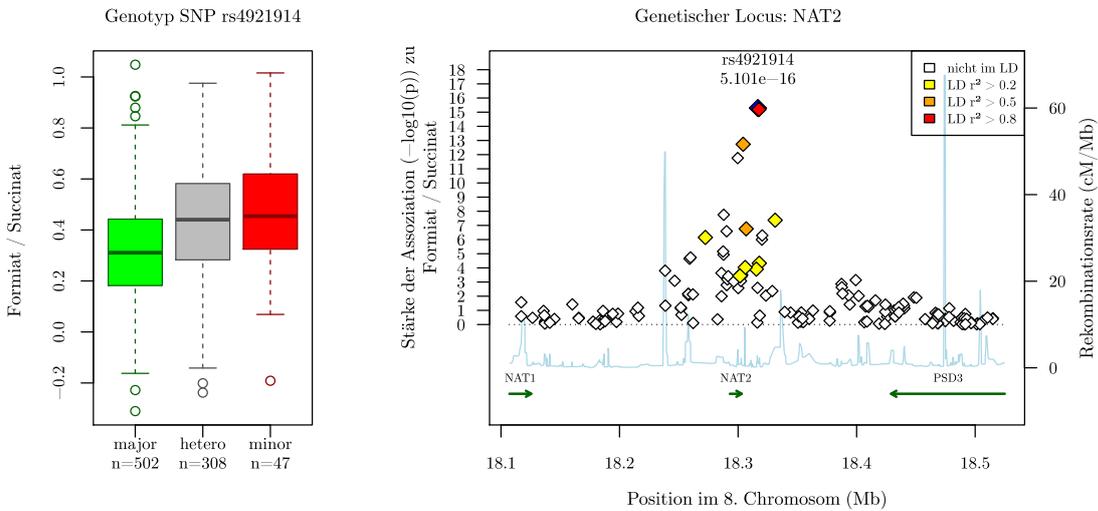
(a) Boxplots

(b) Regionaler Assoziationsplot

Abbildung 6.2(a) zeigt die auf Kreatinin normalisierte Konzentration von beta-Aminoisobutyrat in Abhängigkeit vom rs37369-Genotyp. Die bei den homozygoten Trägern des häufigen bzw. seltenen Allels ermittelten Konzentrationen sind in grün respektive rot eingezeichnet, die bei den heterozygoten Trägern gemessenen in grauer Farbe. n gibt die jeweilige Anzahl der Probanden für jeden Genotyp an. Die Boxplots beschreiben die Verteilung der zur Berechnung der linearen Modelle herangezogenen und um extreme Ausreißer (mehr als drei Standardabweichungen vom Mittel über die in allen Proben gemessene beta-Aminoisobutyrat-Konzentration entfernt) bereinigten Daten. Die Boxen erstrecken sich jeweils vom ersten (Q1) bis zum dritten Quartil (Q3) und umfassen somit 50% der Daten. Der Median ist durch die horizontale Linie gekennzeichnet, die Antennen zeigen das Maximum bzw. Minimum der Verteilung, solange die Datenpunkte nicht weiter als $1,5 \times (Q3 - Q1)$ vom oberen bzw. unteren Ende der Box entfernt sind. Datenpunkte, die weiter entfernt sind, werden als einzelne Punkte im Plot dargestellt.

Abbildung 6.2(b) zeigt den regionalen Assoziationsplot um den AGXT2-Locus. Die Assoziationsstärke (p-Wert) zwischen der Konzentration von beta-Aminoisobutyrat und den in der GWAS betrachteten SNPs ist \log_{10} -skaliert auf der Y-Achse aufgetragen. Die X-Achse zeigt die Position der SNPs sowie der Gene laut NCBI Build 36.1 des menschlichen Genoms. Der SNP mit der stärksten Assoziation, also der in Tabelle 5.1 angegebene Top-SNP, ist blau markiert. Die Farbe der übrigen SNPs gibt an, ob und wie stark sich diese im LD mit dem Top-SNP befinden (LD-Informationen basieren auf HapMap Release 22, CEU Population [72]). Des Weiteren sind auf der Y-Achse die geschätzte Rekombinationsrate in Centimorgan/Megabase (nach HapMap Release 22 [72]) sowie auf der X-Achse die Genannotation (nach RefSeq build 36.3 [76, 90]) aufgetragen.

Abbildung 6.3.: Assoziation von rs4921914 (NAT2) zum Formiat/Succinat-Ratio



(a) Boxplots

(b) Regionaler Assoziationsplot

Die Erläuterungen zu Abbildung 6.2 gelten analog.

NAT2 spielt eine Schlüsselrolle im Xenobiotika-Stoffwechsel. Es ist an der Aktivierung und Deaktivierung verschiedener Pharmazeutika, Karzinogene und anderer Fremdstoffe mit Arylamin- oder Hydralazinstruktur beteiligt [100]. So werden beispielsweise mit Zigarettenrauch aufgenommene Präkarzinogene durch NAT2 O-acetyliert (aktiviert) und somit in ultimative Karzinogene umgewandelt [100, 101]. Kürzlich wurde von Deeken *et al.* gezeigt, dass Polymorphismen im NAT2-Locus mit der Toxizität der in der Krebsbehandlung eingesetzten Wirkstoffe Docetaxel und Thalidomid assoziieren [102]. Teslovich veröffentlichten 2010 eine GWAS, in der SNPs im NAT2-Locus mit der Triglyceridekonzentration im Blutserum sowie mit koronarer Herzkrankheit assoziiert [103].

Weiterhin ist bekannt, dass NAT2-Polymorphismen die Menschheit in schnelle, mittlere und langsame „Acetylierer“ unterteilen [104, 105]. Einige Studien bringen die NAT2-bedingten Acetylierungs-Phänotypen mit dem Risiko für schwere Erkrankungen wie Blasen- und Darmkrebs in Verbindung [106, 107].

In der hier vorgestellten GWAS assoziiert wie oben erwähnt der NAT2-Locus mit den Ratios von Formiat zu Succinat, zu Harnstoff und zu Acetat. Dabei ist die Assoziation mit diesen Ratios um mindestens vier Größenordnungen stärker als mit den Konzentrationen von Formiat alleine (Tabelle 6.1).

Tabelle 6.1.: Mit SNP rs4921914 (NAT2) assoziierte Metaboliten und Ratios

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Formiat / Succinat	854	0,2677	$5,1 \times 10^{-16}$	$3,0 \times 10^5$
Formiat / Harnstoff	852	0,2972	$1,1 \times 10^{-15}$	$1,4 \times 10^5$
Formiat / Acetat	851	0,2735	$9,0 \times 10^{-15}$	$1,7 \times 10^4$
Formiat / Pyroglutamat	856	0,2162	$8,8 \times 10^{-13}$	$1,8 \times 10^2$
Formiat / Isopropanol	772	0,2764	$2,1 \times 10^{-12}$	$7,4 \times 10^1$
Formiat	859	0,2050	$1,6 \times 10^{-10}$	
Cholin / Harnstoff	465	0,2972	$4,2 \times 10^{-6}$	$5,5 \times 10^2$

Es werden alle Assoziationen zu rs4921914 aufgeführt, deren p-Wert kleiner $2,9 \times 10^{-5}$ ist (Bonferroni-Korrektur für 1.720 Tests) und - im Falle von Ratios - mit einem p-Gain größer 59, entsprechend der Anzahl der gemessenen Metaboliten. Erläuterungen zu N , $beta'$ und zum p -Gain sind im Begleittext zu Tabelle 5.1 zu finden.

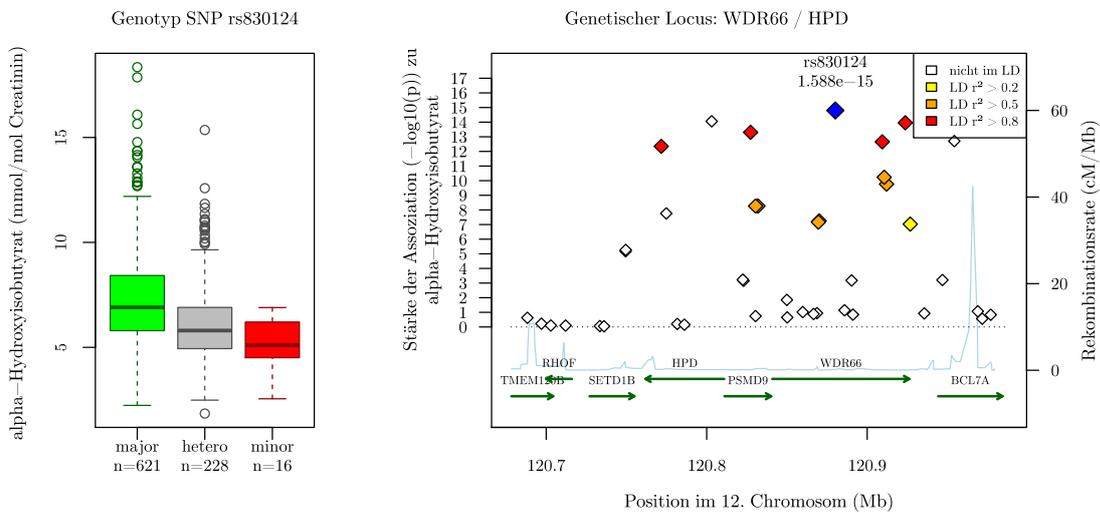
Die hier identifizierten Assoziationen legen die Vermutung nahe, dass die Konzentrationen von Succinat, Harnstoff, Acetat und weiterer Metaboliten mit Acetylierungs-Phänotypen zusammenhängen, die ihrerseits durch mit SNP rs4921914 gekoppelte Polymorphismen beeinflusst werden. Die Ratios dieser Metaboliten könnten möglicherweise in der klinischen Diagnostik als Biomarker verwendet werden, beispielsweise beim Einsatz von Krebsmedikamenten.

Der WDR66/HPD-Locus ist assoziiert mit alpha-Hydroxyisobutyrat

SNP rs830124 liegt im intronischen Bereich des WDR66-Gens, welches auf dem Plusstrang des 12. Chromosoms liegt. Dieser SNP assoziiert in der Discovery-Studie mit den Konzentrationen von alpha-Hydroxyisobutyrat (2-Hydroxyisobutyrat) bei einem p-Wert kleiner $1,6 \times 10^{-15}$ (Tabelle 5.1; Abbildung 6.4).

Das „WD Repeat-Containing Protein 66“ ist beteiligt an der Bildung von Protein-Komplexen. Soranzo *et al.* sowie Meisinger *et al.* fanden 2009 durch GWAS mit hämatologischen Parametern die Assoziation eines Polymorphismus' im WDR66-Gen mit dem mittleren Thrombozytenvolumen [108, 109]. Der dort identifizierte SNP rs7961894 liegt zwar ebenfalls im intronischen Bereich von WDR66, ist aber nicht im LD mit rs830124 ($r^2 = 0,018$).

Abbildung 6.4.: Assoziation von rs830124 (WDR66/HPD) zur alpha-Hydroxyisovalerat-Konzentration



(a) Boxplots

(b) Regionaler Assoziationsplot

Die Erläuterungen zu Abbildung 6.2 gelten analog für die hier gezeigten Plots.

Bei Betrachtung des Minusstrang des Genoms liegt rs830124 ca. 90 kB oberhalb des für die 4-Hydroxyphenylpyruvat-Dioxygenase (HPD) codierenden Gens. Wie aus Abbildung 6.4 (b) ersichtlich ist er im LD mit Varianten im HPD-Gen. Das durch HPD kodierte Enzym katalysiert im Tyrosin-Stoffwechsel die Umwandlung von 4-Hydroxyphenylpyruvat zu Homogentisat. HPD wird hauptsächlich in der Leber und in geringem Maße in den Nieren exprimiert (Illumina BodyMap 2.0 [75]). HPD ist ein bekannter Risikolocus: Mutationen im HPD-Gen führen zu Hawkinsinurie, einer autosomal dominanten Erbkrankheit (OMIM #140350) [110]. Hawkinsinurie äußert sich bei Kleinkindern durch vorübergehende metabolische Azidose und Tyrosinämie [111].

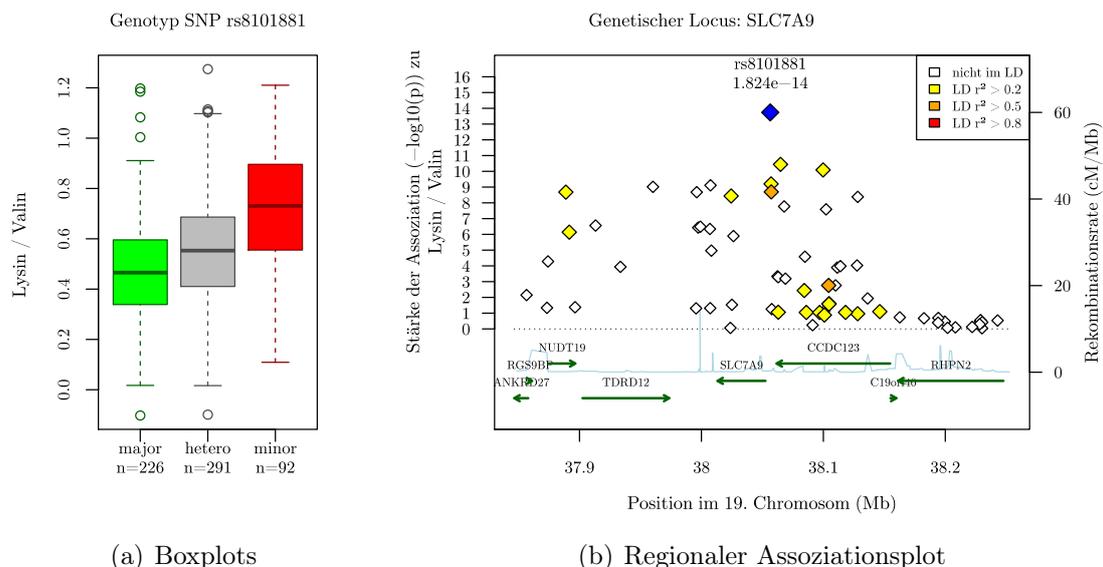
Es bleibt unklar, ob bzw. wie alpha-Hydroxyisobutyrat metabolisch mit der Funktion der Genprodukte von WDR66 und HPD in Verbindung steht. Dennoch ist diese Assoziation in Hinblick auf die Rolle von HPD für Hawkinsinurie und Tyrosinämie interessant.

Der SLC7A9-Transporter hat möglicherweise eine genotypabhängige Affinität zu Lysin

SNP rs8101881 liegt auf dem 19. Chromosom ca. 4 kB vor dem SLC7A9-Gen (Abbildung 6.6(b)), welches für einen Transporter für Cystin sowie für neutrale und dibasische Aminosäuren codiert (OMIM *604144). Mutationen in SLC7A9 können zur Cystinurie führen, einer seltenen autosomal vererbten Stoffwechselerkrankung, welche einen gestörten Transport von Cystin und strukturverwandten Aminosäuren in den Nierentubuli zur Folge hat [112]. Dadurch kann es zu einer erhöhten Cystin-Ausfällung im Urin und somit zur Bildung von Cystinsteinen kommen, welche schließlich die Harnwege verstopfen können [113, 114]. Der SLC7A9-Locus ist also ein Risikolocus für Nierenerkrankungen.

Dies zeigen auch zwei im Jahre 2010 veröffentlichte genetische Assoziationsstudien mit Nierenfunktion und Nierenkrankheiten. Köttgen *et al.* sowie Chambers *et al.* fanden Assoziationen von SNPs im SLC7A9-Locus zu Nierenfunktionsmarkern (glomeruläre Filtrationsrate und Kreatinin-Konzentration im Blutserum) sowie zu chronischer Niereninsuffizienz [47, 115]. Interessanterweise ist rs8101881 mit mehreren dieser Varianten (z.B. rs12460876) im LD und ist somit ebenfalls mit Nierenfunktion in Verbindung zu bringen.

Abbildung 6.5.: Die Assoziation von rs8101881 (SLC7A9) zum Lysin/Valin-Ratio



Die Erläuterungen zu Abbildung 6.2 gelten analog für die hier gezeigten Plots.

In der hier besprochenen Studie assoziiert rs8101881 mit Lysin bei einem p-Wert kleiner $1,0 \times 10^{-8}$, jedoch ist diese Assoziation nach der Korrektur für multiples Testen nicht genomweit signifikant. Um bis zu fünf Größenordnungen stärker und damit signifikant sind hingegen die Assoziationen zu den Ratios Lysin/Valin (Abbildung 6.5), Lysin/Asparagin sowie Lysin/Pyroglutamat (Tabelle 6.2). Die Assoziationen lassen vermuten, dass die basische Aminosäure Lysin ein Substrat des SLC7A9-Transporters ist. Die hier beobachtete tendenziell erhöhte Lysinkonzentration im Harn ist dabei im Einklang mit der in früheren Studien bei Cystinurie-Patienten beobachteten erhöhten Lysin-Ausscheidung [116]. Aufgrund der höheren Assoziationsstärke der Ratios aus Lysin zu Valin, Asparagin und Pyroglutamat ist es denkbar, dass mit rs8101881 gekoppelte Genvarianten zudem Einfluss auf den Transport dieser Metaboliten haben. Es ist jedoch auch möglich, dass diese Ratios einen normalisierenden Effekt haben, wodurch intraindividuelle Schwankungen der gemessenen Lysinkonzentration kompensiert werden.

Die hier identifizierte Assoziation des SLC7A9-Locus zu Lysin könnte dazu beitragen, die Mechanismen hinter der Entstehung von Nierenerkrankungen besser zu verstehen.

Tabelle 6.2.: Mit SNP rs8101881 (SLC7A9) assoziierte Metaboliten oder Metabolitenratios

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Lysin / Valin	605	0,2853	$1,8 \times 10^{-14}$	$5,6 \times 10^5$
Lysin / Pyroglutamat	616	0,2706	$2,4 \times 10^{-12}$	$4,2 \times 10^3$
Lysin / Asparagin	511	0,2677	$3,5 \times 10^{-12}$	$2,9 \times 10^3$
Lysin / 3-Hydroxyisovalerat	613	0,2677	$3,9 \times 10^{-11}$	$2,6 \times 10^2$
Lysin / Glutamin	612	0,2445	$5,0 \times 10^{-11}$	$2,0 \times 10^2$
Lysin	616	0,2246	$1,0 \times 10^{-8}$	

Die Erläuterungen zu Tabelle 6.1 gelten hier analog.

N,N-Dimethylglycin ist ein potentielles Substrat von SLC6A20

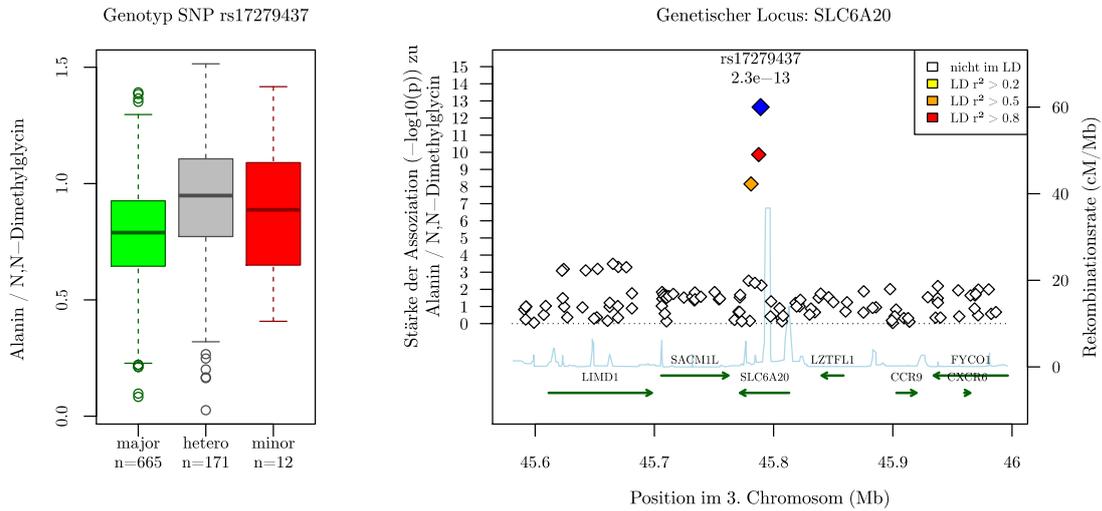
Das SLC6A20-Gen auf dem 3. Chromosom codiert ein gleichnamiges, zur Familie der Na⁺- und Cl⁻-abhängigen SLC6 Neurotransmitter-Transporter gehörendes Protein. Dieses Gen wird unter anderem in den Nieren exprimiert. Das SLC6A20-Protein ist ein Transporter für sekundäre Aminosäuren wie L-Prolin sowie für N-methylierte Aminosäuren [117].

In der hier vorgestellten GWAS assoziiert die Konzentration von N,N-Dimethylglycin bei einem p-Wert kleiner $5,5 \times 10^{-10}$ mit dem SNP rs17279437 (Abbildung 6.7(b)), welcher einen Aminosäureaustausch im SLC6A20-Protein hervorruft: an Position 199 wird Threonin durch Prolin ersetzt (T199M). Die beobachtete Assoziation scheint zur Funktion des SLC6A20-Transporters zu passen. Die Ratios von Alanin, Tyrosin sowie Threonin zu N,N-Dimethylglycin zeigen im Vergleich dazu um bis zu drei Größenordnungen stärkere Assoziationen mit rs17279437 (Tabelle 6.3; Abbildung 6.6).

Takanaga *et al.* zeigten, dass SLC6A20 für alpha-Aminosäuren wie Alanin und Lysin nur schwache bzw. keine Transporteraktivität zeigt [117]. Die Ergebnisse der GWAS lassen aber vermuten, dass die seltene Allelvariante des rs17279437-Polymorphismus' die Affinität dieses Transporters zu Alanin, Threonin und Tyrosin auf Kosten der Affinität zu N,N-Dimethylglycin verändern könnte.

Der SLC6A20-Locus liefert ein anschauliches Beispiel für Gen-zu-Gen-Interaktionen: Bröer *et al.* zeigten in einer Familienstudie, dass der Funktionsverlust des SLC36A2-Transporters in der selten auftretenden, autosomal rezessiven Erbkrankheit Iminoglycinurie resultiert. Durch diese Stoffwechselstörung wird die Reabsorption von Glycin, Prolin und Hydroxyprolin in den Nierentubuli behindert, wodurch es zu einer erhöhten Ausscheidung dieser Metaboliten kommt. Die Autoren der Studie identifizierten eine Mutation in SLC36A2 (G87V), die dem Protein noch eine schwache Transportaktivität ermöglicht und erst in Verbindung mit SNP rs17279437, respektive den dadurch bedingten T199M-Polymorphismus im SLC6A20-Transporter, zur Iminoglycinurie führt [118].

Abbildung 6.6.: Assoziation von rs17279437 (SLC6A20) zum Alanin/N,N-Dimethylglycin-Ratio



(a) Boxplots (b) Regionaler Assoziationsplot
Die Erläuterungen zu Abbildung 6.2 gelten analog für die hier gezeigten Plots.

Tabelle 6.3.: Mit SNP rs17279437 (SLC6A20) assoziierte Metaboliten oder Metabolitenratios

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Alanin / N,N-Dimethylglycin	831	0,3709	$2,3 \times 10^{-13}$	$2,4 \times 10^3$
Tyrosin / N,N-Dimethylglycin	811	0,4060	$3,2 \times 10^{-13}$	$1,7 \times 10^3$
Threonin / N,N-Dimethylglycin	827	0,3677	$5,6 \times 10^{-12}$	$9,9 \times 10^2$
N,N-Dimethylglycin	842	-0,2552	$5,5 \times 10^{-10}$	

Die Erläuterungen zu Tabelle 6.1 gelten hier analog.

6.2. Genetisch beeinflusste metabolische Phänotypen im Urin und Blut

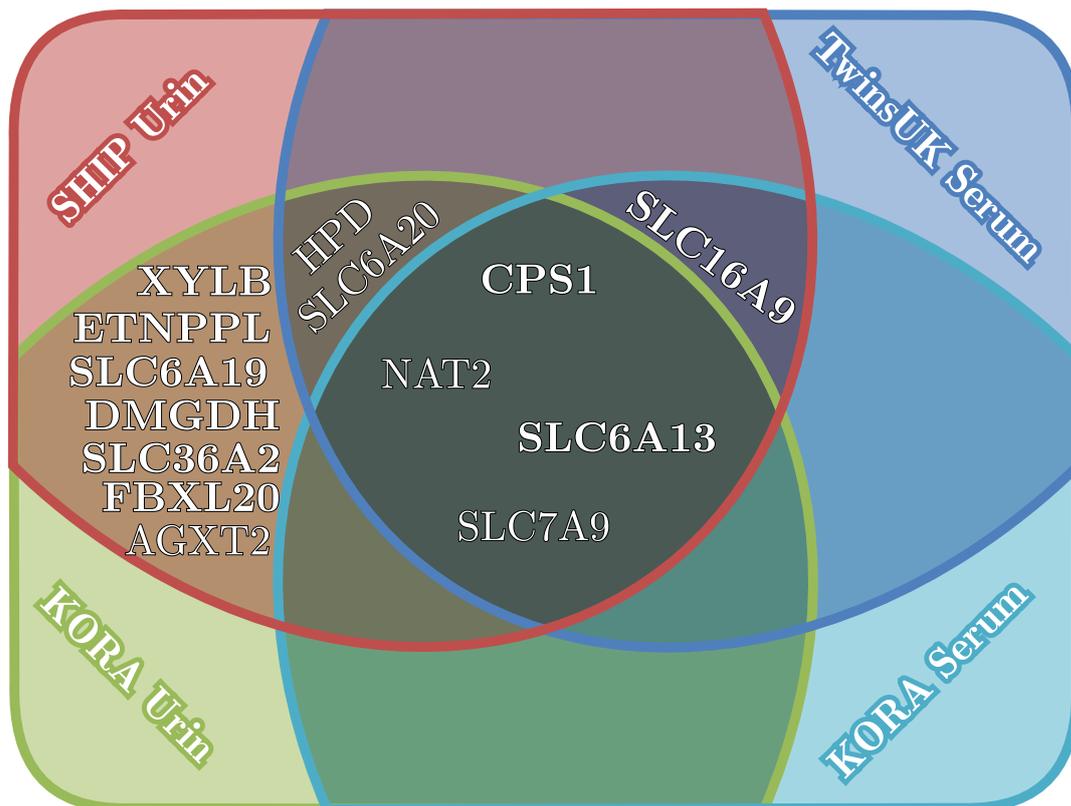
Ungefähr zwei Jahre nach Abschluss der ersten mGWAS im Urin waren die annotierten NMR-Spektren der Urinproben nahezu aller Probanden der SHIP-0-Kohorte verfügbar. Mit diesem weitaus umfangreicheren Datensatz wurde eine weitere genomweite Assoziationsstudie durchgeführt. In dieser mGWAS konnten neben den fünf bereits aus der vorherigen Studie bekannten genetisch beeinflussten Metabotypen (GIMs) neun weitere Loci identifiziert werden, welche bei genomweiter Signifikanz mit einzelnen Metaboliten oder Ratios aus Metabolitenpaaren im Urin assoziieren (CPS1, XYLB, ETNPPL, SLC6A19, DMGDH, SLC36A2, SLC16A9, SLC6A13 und FBXL20).

Wie zu Beginn dieses Kapitels erwähnt wurde bei dieser Studie zusätzlich untersucht, ob die in SHIP-0 identifizierten und in KORA F4 replizierten Loci auch mit metabolischen Merkmalen im Blut der KORA-F4- und der TwinsUK-Kohorten assoziieren. Besonders interessant wird dieser Vergleich dadurch, dass die Urin- und Blutproben aus KORA F4 von den gleichen Probanden stammen. Dieser Ansatz, GIMs im Blut und im Urin der gleichen Kohorte zu identifizieren, wurde bereits 2011 in einer von Nicholson *et al.* publizierten mGWAS verfolgt [40]. Nicholson *et al.* standen allerdings nur etwas mehr als 200 Probanden zur Verfügung, und so konnte auch nur ein Locus (PYROXD2) identifiziert werden, der mit Metaboliten sowohl im Urin (Trimethylamin) als auch im Plasma (Dimethylamin) signifikant assoziiert.

Gemäß dem Vergleich mit Assoziationsergebnissen im Blut der KORA-F4- und TwinsUK-Kohorten lassen sich die insgesamt 14 genetischen Loci, die im SHIP-0-Datensatz identifiziert wurden, in drei Kategorien einteilen (Abbildung 6.7):

- (i) für sieben Loci finden sich lediglich im Urin signifikante Assoziationen zu metabolischen Merkmalen, nicht jedoch im Blut der KORA-F4- und TwinsUK-Kohorten (SLC36A2, FBXL20, XYLB, ETNPPL, SLC6A19, AGXT2, DMGDH)
- (ii) zwei Loci assoziieren sowohl im Urin als auch im Blut mit den selben oder eng verwandten Stoffwechselprodukten (CPS1 und HPD)
- (iii) fünf Loci zeigen Assoziationen zu Metaboliten, welche sich im Urin und im Blut unterscheiden (SLC6A20, NAT2, SLC16A9, SLC7A9 und SLC6A13)

Abbildung 6.7.: Die Assoziationen der 14 identifizierten genetischen Loci in verschiedenen Medien und Studien



Das Diagramm führt die 14 genetischen Loci auf, die in der mGWAS im Urin der SHIP-0-Probanden signifikant mit metabolischen Merkmalen assoziieren. Diejenigen neun Loci, welche noch nicht in Vorgängerstudie (Kapitel 6.1) identifiziert wurden, sind fett abgedruckt. Alle Loci, mit Ausnahme des SLC16A9-Locus, zeigten in dem zur Replikation herangezogenen KORA-F4-Datensatz im Urin ebenfalls signifikante Assoziationen zu den in der SHIP-0-Studie assoziierten Metaboliten oder Ratios. Fünf Loci (CPS1, NAT2, SLC6A13, SLC7A9 und SLC16A9) sind auch im Serum der KORA-F4- und TwinsUK-Studien mit metabolischen Merkmalen signifikant assoziiert. Hinzu kommen mit HPD und SLC6A20 zwei weitere Loci, für die nur im TwinsUK-Datensatz signifikante Assoziationssignale gefunden wurden.

Für fünf der hier identifizierten Loci (XYLB, SLC6A19, SLC36A2, FBXL20 und ETNPPL) wurden laut Katalog der publizierten GWAS des National Human Genome Research Institute [28] (Stand August 2013) bislang keine Assoziationsergebnisse in mGWAS veröffentlicht. Neun der vierzehn Loci wurden in früheren GWAS mit Stoffwechselprodukten publiziert. Die Assoziationen der Loci AGXT2, SLC6A20, HPD (bzw. WDR66), SLC7A9 sowie NAT2 zu metabolischen Markern im Urin wurden erstmals im Jahre 2011 von Suhre *et al.* [53] veröffentlicht (siehe Kapitel 6.1); die Assoziationen von CPS1 sowie SLC16A9 zu Stoffwechselprodukten im Blut wurden in Studien von Suhre *et al.* [35] und anderen Gruppen (z.B [119, 120]) identifiziert. Für die Loci SLC6A13 und DMGDH veröffentlichten im Juli 2013 Rhee *et al.* Assoziationen zu beta-Aminoisobutyrat beziehungsweise Dimethylglycin in Blutplasma [42].

Im folgenden Teil soll näher auf die identifizierten Assoziationen eingegangen werden, beginnend mit den Polymorphismen im CPS1- sowie im HPD-Locus, die im Urin und Serum mit jeweils den gleichen metabolischen Merkmalen assoziieren. Anschließend werden die Loci NAT2, SLC6A20, SLC6A13, SLC7A9 sowie SLC16A9 besprochen. In diesen Loci befinden sich SNPs, die ebenfalls sowohl im Blut als auch im Urin mit Metabolitenkonzentrationen bzw. -ratios assoziieren, hier allerdings nicht mit identischen, sondern mehr oder minder metabolisch verwandten Stoffwechselprodukten. Den Abschluss bilden die Loci, bei denen nur Assoziationen zu Metaboliten im Urin beobachtet werden konnten (SLC36A2, FBXL20, XYLB, ETNPPL, AGXT2, SLC6A19 und DMGDH).

Zu jedem Locus werden alle Assoziationen ($p < 5,0 \times 10^{-4}$) zu Metabolitenkonzentrationen bzw. zu Metaboliten-Ratios ($p\text{-Gain} > 1,51 \times 10^4$) tabellarisch aufgeführt. Weiterhin finden sich im Anhang dieser Arbeit (Abbildung A.1) zu den in Tabelle 5.3 aufgeführten Assoziationen die regionalen Assoziationsplots, Boxplots sowie QQ-Plots.

Loci mit identischen GIMs im Urin und Serum

Der Carbamoylphosphat-Synthetase-1-Locus assoziiert mit den Glycin-Konzentrationen im Urin und im Serum

Der SNP rs2216405 liegt auf dem 2. Chromosom und befindet sich ungefähr 40 kB entfernt vom 3'-Ende des Carbamoylphosphat-Synthetase-1-Gens (CPS1). Das mitochondriale CPS1-Enzym katalysiert die Synthese von Carbamoylphosphat aus Ammoniak. Carbamoylphosphat ist das Eintrittsprodukt in den Harnstoffzyklus (Abbildung 6.10). In früheren Studien konnten eine nicht-synonyme Mutation (SNP rs1047891) im CPS1-Gen mit verringerter L-Arginin-Konzentration im Plasma neugeborener Kinder sowie mit ebenfalls bei Neugeborenen beobachteter pulmonaler Hypertonie in Verbindung gebracht werden [121, 122]. Der genannte SNP rs1047891 ist schwach im LD mit rs2216405 ($r^2 = 0,47$). Durch Mutationen am CPS1-Gen bedingter CPS1-Mangel (OMIM #237300) kann zu Hyperammonämie führen, die bei Neugeborenen tödlich sein kann [87]. Die Symptome können aber auch erst im Erwachsenenstadium auftreten [123].

Im Jahre 2011 identifizierten Suhre *et al.* in einer GWAS im KORA-F4-Datensatz eine signifikante Assoziation von rs2216405 mit dem Ratio aus Glycin und Kreatin [35]. Dabei ist die Blutkonzentration von Glycin bei Vorliegen des seltenen Allels tendenziell erhöht (siehe hierzu auch Kapitel 6.3). Bei der hier besprochenen mGWAS im Urin gibt es Parallelen zu den Beobachtungen im Blutserum. Es finden sich ebenfalls Assoziationen von rs2216405 zu Glycin ($p < 1,3 \times 10^{-9}$) sowie zu Kreatin ($p < 5,6 \times 10^{-11}$). Wie im Blut erhöht sich auch im Urin die Konzentration dieser Metaboliten mit der Zahl des seltenen rs2216405-Allels (Glycin: $\beta' = 0,1428$, Kreatin: $\beta' = 0,1422$). Allerdings sind die beiden eben genannten Assoziationen nicht genomweit signifikant, jedoch die Assoziation zum Glycin/Glycolat-Ratio ($p < 3,7 \times 10^{-17}$, $\beta' = 0,1376$). Die an diesem Ratio beteiligten Metaboliten sind eng verwandt: in einer zweistufigen Reaktion erfolgt die Synthese von Glycolat aus Glycin über das Zwischenprodukt Glyoxylat (Abbildung 6.8). Auf die Konzentration von Glycolat alleine hat der Genotyp des SNP rs2216405 keinen signifikanten Einfluss, und so ist es wahrscheinlich, dass durch die Verwendung dieses Ratios als Phänotyp eine Normalisierung erfolgt, wodurch der Effekt des CPS1-Polymorphismus auf die Glycin-Ausscheidung deutlicher sichtbar wird ($p\text{-Gain} > 3,4 \times 10^7$).

Abbildung 6.8.: Zweistufige Umsetzung von Glycin zu Glycolat

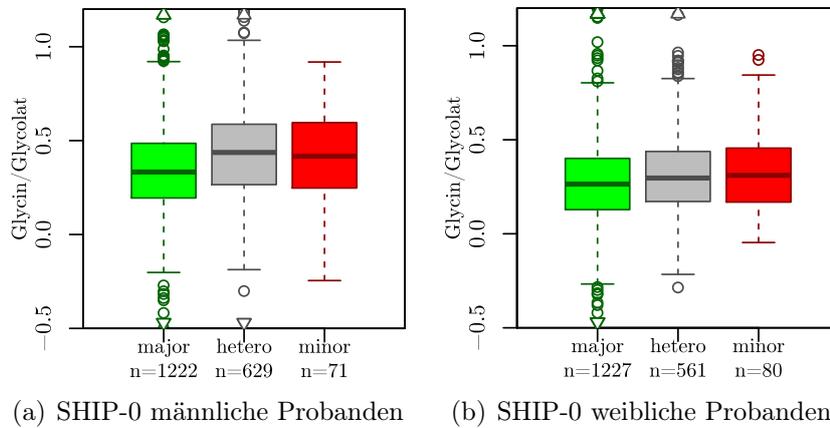


Glycin wird durch D-Aminosäureoxidase (DAO) zu Glyoxylat umgewandelt (EC 1.4.3.3), welches durch Glyoxylatreductase/Hydroxypyruvatreductase (GRHPR) zu Glycolat umgesetzt wird (EC 1.1.1.79). In der hier untersuchten GWAS hat SNP rs2216405 keinen Effekt auf die Glycolat-Konzentration ($p < 0,96$), sondern nur auf die Glycin-Konzentration ($p < 1,3 \times 10^{-9}$, $\beta' = 0,14$), was durch die gelben Pfeile angedeutet wird.

Mittelstrass *et al.* zeigten 2012, dass der Effekt von SNP rs2216405 auf die Glycin-Konzentration im Blut stark geschlechtsabhängig und bei Frauen weitaus besser zu beobachten ist [124]. Dies ist auch im Urin der SHIP-0-Probanden zu sehen: testet man nur die 1.922 weiblichen Probanden, so ist die Assoziation zum Glycin/Glycolat-Ratio trotz der halbierten Probandenzahl immer noch genomweit signifikant ($p < 1,1 \times 10^{-13}$; $\beta' = 0,1793$). Die Assoziation ist hingegen nicht mehr signifikant, wenn nur die 1.868 männlichen Studienteilnehmer getestet werden ($p < 1,4 \times 10^{-5}$; $\beta' = 0,0955$).

Einen möglichen Zusammenhang zwischen Mutationen im CPS1-Locus und den Konzentrationen von Glycin und Kreatin zeigt Abbildung 6.10: aus Arginin und Glycin wird Guanidinoacetat synthetisiert (EC 2.1.4.1), welches wiederum durch die Guanidinoacetat-N-Methyltransferase zu Kreatin umgewandelt wird (EC 2.1.1.2). Nimmt man an, dass ein mit SNP rs2116405 gekoppelter Polymorphismus mit einer Störung des Harnstoffzyklus einhergeht und dadurch die Synthese von Arginin beeinträchtigt (siehe die weiter oben erwähnten Studien an Neugeborenen [121, 122]), könnte dies zu einem Glycin-Stau sowie einer verringerten Kreatin-Synthese führen. Allerdings wird in der hier besprochenen GWAS eine erhöhte Ausscheidung von Kreatin beobachtet, wo eine Konzentrationsverringering zu erwarten wäre. Außerdem konnten in früheren Assoziationsstudien im Blutplasma erwachsener Probanden keine signifikante Assoziation zwischen der oben genannten, CPS1-Mangel verursachenden Mutation (SNP rs1047891), die mäßig im LD zu SNP rs2216405 ist ($r^2 = 0,47$), und einer verringerten Konzentration von Arginin identifiziert werden [125, 126]. Und schließlich ist, wie bereits erwähnt, im Blut der KORA-F4-Probanden

Abbildung 6.9.: Glycin/Glycolat-Ratio bei Frauen und Männern, aufgetrennt nach Genotyp des SNPs rs2216405



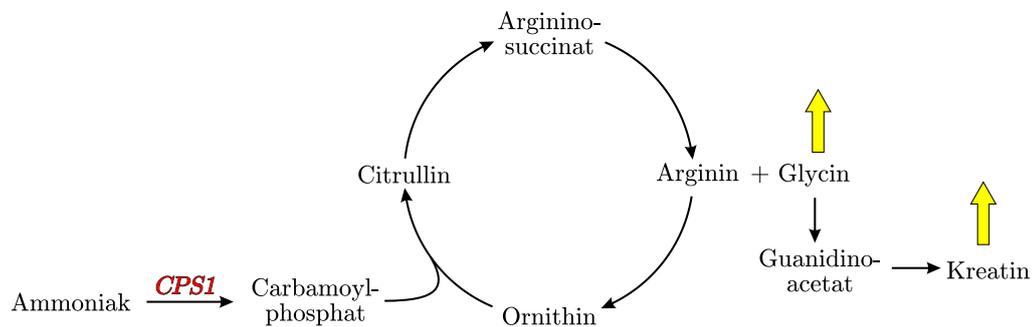
Die Assoziation des SNPs rs2216405 zum Ratio aus Glycin und Glycolat bleibt genomweit signifikant, wenn nur die weiblichen Probanden der SHIP-0-Studie getestet werden ($p < 1,1 \times 10^{-13}$). Werden nur die männlichen Studienteilnehmer getestet, ist die Assoziation hingegen nicht signifikant ($p < 1,4 \times 10^{-5}$). Die abgebildeten Boxplots zeigen den Ratio in Abhängigkeit des rs2216405-Genotyps für die nach Geschlecht aufgetrennten Probanden. Interpretationshilfen zu den Boxplots finden sich im Begleittext zu Abbildung 5.1.

die Konzentration von Kreatin bei Vorliegen des seltenen rs2216495-Allels, analog zur Beobachtung im Urin, tendenziell erhöht ($p = 9,55 \times 10^{-7}$; $\beta = 0,0394$).

Aus einem CPS1-Mangel resultierendes überschüssiges Ammoniak könnte aber durch das „*Glycine Cleavage System*“ zu Glycin umgesetzt werden (Abbildung 6.11) [127, 128]. Dieses Glycin könnte wiederum über den in Abbildung 6.10 gezeigten metabolischen Pfad zu Kreatin umgewandelt werden, wobei das zur Synthese notwendige Arginin auch einen anderen Ursprung als den Harnstoffzyklus haben kann. Durch diesen Mechanismus ließen sich sowohl die erhöhte Glycin- als auch die Kreatinkonzentration erklären.

In einer 2012 veröffentlichten, auf Serumproben der KORA-F4-Studie basierenden Publikation zeigten Wang-Sattler *et al.*, dass die Blutkonzentration von Glycin dazu geeignet ist, das Risiko für Typ-2-Diabetes vorherzusagen [129]. Die Autoren fanden heraus, dass neben anderen Metaboliten die Konzentration von Glycin bei Personen mit verringerter Glucosetoleranz signifikant niedriger ist als bei Probanden mit normaler Glucosetoleranz. Auch in anderen Studien konnten Assoziationen zwischen Glycin und dem Risiko für Typ-2-Diabetes identifiziert werden [130, 131]. Es ist aber gegenwärtig noch unklar, ob auch Mutationen im CPS1-Locus, die Assoziatio-

Abbildung 6.10.: Auswirkungen des CPS1-Mangels auf die Carbamoylphosphat-Synthese und den Harnstoffzyklus

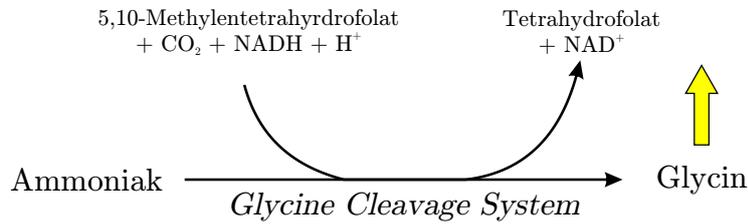


Bestimmte Polymorphismen im CPS1-Gen führen zu CPS1-Mangel, bei dem es zu einer verminderten Synthese von Carbamoylphosphat aus Ammoniak kommt [122]. In der vorliegenden GWAS assoziiert SNP rs2216405 im CPS1-Locus mit einer erhöhten Ausscheidung von Glycin und Kreatin (gelbe Pfeile). Ob allerdings ein mit diesem SNP gekoppelter Polymorphismus zu einer Beeinträchtigung des Harnstoff-Zyklus führt und dadurch Auswirkungen auf die Synthese von Kreatin aus Glycin und Arginin hat, ist angesichts der beobachteten Konzentrationserhöhungen fraglich.

nen mit der Glycin-Konzentration im Blut zeigen, ebenfalls mit einem veränderten Diabetes-Risiko einhergehen. 2013 untersuchten Xie *et al.* in 1.004 Probanden genetische Varianten im CPS1-Locus, unter anderem auch den SNP rs715, der mäßig rs2216405 gekoppelt ist ($r^2 = 0,55$) und signifikant mit Glycin assoziiert [132]. Xie *et al.* konnten in ihrer Studie allerdings keine signifikante Assoziation dieses SNPs zu Insulinresistenz identifizieren.

Zusammenfassend konnte in der hier besprochenen Studie erstmalig gezeigt werden, dass das mit einem veränderten Diabetesrisiko assoziierte Glycin bei (weiblichen) Trägern des seltenen rs2216405-Allels nicht nur im Blut vermehrt vorhanden ist, sondern auch tendenziell verstärkt mit dem Harn ausgeschieden wird. Möglicherweise kann der hier vorgestellte Glycin-Glycolat-Ratio künftig als Biomarker im Urin für CPS1-bedingte Störungen im Ammoniak-Metabolismus herangezogen werden. Über die Mechanismen, die vom CPS1-Mangel zum Glycin-Überschuss führen, kann allerdings an dieser Stelle nur spekuliert werden.

Abbildung 6.11.: Glycin-Synthese aus Ammoniak durch das Glycine Cleavage System



Schematische Darstellung der Glycin-Synthese aus Ammoniak über das *Glycine Cleavage System* [127, 128]. Der Pfeil deutet eine tendenziell erhöhte Glycin-Konzentration im Urin (und Blut) bei Vorliegen des seltenen rs2216405-Allels im CPS1-Locus an. Womöglich wird über diesen Pfad der aus einem CPS1-Mangel resultierende Ammoniak-Überschuss vermehrt in Glycin umgewandelt.

Tabelle 6.4.: Assoziationen zu rs2216405 (CPS1)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Glycin / Glycolat	3.790	0,1376	$3,72 \times 10^{-17}$	$3,38 \times 10^7$
Kreatin	3.858	0,1422	$5,64 \times 10^{-11}$	
Glycin	3.857	0,1428	$1,26 \times 10^{-9}$	

Es werden alle Assoziationen des SNPs rs2216405 mit Metaboliten und Ratios im Urin der SHIP-0-Studie gezeigt, deren p-Wert kleiner $5,0 \times 10^{-4}$ und - im Falle von Ratios - deren p-Gain größer $1,51 \times 10^4$ ist. Erläuterungen zu den Spaltenbezeichnungen sind bei Tabelle 5.3 zu finden.

alpha-Hydroxisobutyrat assoziiert sowohl im Urin als auch im Serum mit Polymorphismen im HPD-Locus

Die zweitstärkste im Discovery-Datensatz gefundene Assoziation ist zwischen dem SNP rs4760099 und der Konzentration von alpha-Hydroxyisobutyrat (α -HIBA) bei einem p-Wert kleiner $1,32 \times 10^{-77}$ (Tabelle 6.5). SNP rs4760099 liegt auf dem 12. Chromosom im intronischen Bereich des 4-Hydroxyphenylpyruvat-Dioxygenase-Gens (HPD), dessen Genprodukt im Tyrosinstoffwechsel die Umwandlung von 4-Hydroxyphenylbrenztraubensäure zu Homogentisinsäure katalysiert (EC 1.13.11.27). In Kapitel 6.1 wurde dieser Locus bereits näher beschrieben. Dort assoziiert der in der Nähe gelegene SNP rs830124 genomweit signifikant mit α -HIBA. Dabei sind die

SNPs rs830124 und der in der vorliegenden Studie identifizierte rs4760099 in Hap-Map (Release 27, Phase I-III) vollständig im LD ($r^2 = 1$).

Laut Yaneva *et al.* ist α -HIBA kein endogener, primärer Metabolit im Hauptstoffwechsel des Menschen [133]. Bei einigen Bakterien ist α -HIBA ein Produkt des Abbaus von Methyl-tert-butylether (MTBE), welches seit mehr als zwei Jahrzehnten Ottokraftstoffen als Antiklopffmittel zugesetzt wird [133, 134]. Beim Menschen wird der flüchtige MTBE über die Atemluft oder mit dem Trinkwasser aufgenommen und teilweise zu α -HIBA metabolisiert. α -HIBA wird schließlich mit dem Harn ausgeschieden (siehe hierzu auch den HMDB-Eintrag HMDB00729, Stand 14. August 2013) [135–137]. Analog hierzu wird auch der Kraftstoffzusatz Ethyl-tert-butylether zu α -HIBA verstoffwechselt [138]. Laut Kumps *et al.* wurde aber auch beobachtet, dass Milchsäure-Azidurie mit einer erhöhten Konzentration von α -HIBA im Harn koinzidiert [139]. Dies spricht dafür, dass es neben Umwelteinflüssen auch genetische Komponenten gibt, welche die α -HIBA-Ausscheidung beeinflussen.

Wie bereits weiter oben bei der Diskussion der Ergebnisse der ersten mGWAS im Urin diskutiert, ist kein direkter Zusammenhang zwischen der HPD-gesteuerten biochemischen Reaktion und α -HIBA erkennbar. Allerdings ist diese ursprünglich nur im Urin identifizierte Assoziation mittlerweile auch im Serum beobachtet worden: in der TwinsUK-Studie assoziiert rs2230681 (bei $r^2 = 1$ im LD zu rs4760099) ebenfalls mit α -HIBA ($p < 1,85 \times 10^{-18}$), wobei das seltene Allel wie im Urin mit einer verringerten Konzentration des Metaboliten im Serum einhergeht ($\beta' = -0,1655$). Somit konnte diese Assoziation mit unterschiedlichen Metabolomik-Plattformen (NMR und MS), in verschiedenen Medien (Urin und Serum) und in drei disjunkten Populationen identifiziert werden. Das Signal des HPD-Locus mit diesem speziellen Metaboliten ist also sehr stabil; die biologische Bedeutung der Assoziation ist dagegen momentan weiterhin unklar.

Tabelle 6.5.: Assoziationen zu rs4760099 (HPD)

Metabolit oder Ratio	N	beta'	p-Wert
alpha-Hydroxyisobutyrat	3.853	-0,1655	$1,32 \times 10^{-77}$
3-Hydroxyisovalerat	3.844	-0,0575	$7,60 \times 10^{-7}$

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Loci mit unterschiedlichen GIMs im Urin und Serum

NAT2-Locus: schnelle Acetylierer scheiden mehr Formiat aus

Die GWAS liefert eine Assoziation zwischen dem SNP rs4921914 und dem Ratio aus Formiat und Harnstoff bei einem p-Wert kleiner $5,85 \times 10^{-23}$. Dieser SNP liegt in der Nähe des für die N-Acetyltransferase 2 codierenden Gens, dessen Genprodukt im Xenobiotika-Stoffwechsel eine wichtige Rolle spielt. So werden beispielsweise manche Medikamente durch Acetylierung aktiviert [100]. Weiterhin ist bekannt, dass genetische Polymorphismen im NAT2-Locus die Geschwindigkeit der Acetylierung beeinflussen [104, 105].

Auf den NAT2-Locus wurde bereits in Kapitel 6.1 näher eingegangen, denn in dieser früheren Studie mit metabolischen Messungen im Urin der SHIP-0-Studienteilnehmer assoziierte ebenfalls SNP rs4921914 mit Formiat, wenn auch dort am stärksten mit dem Ratio aus Formiat und Succinat. Der in der hier besprochenen Studie assoziierte Formiat/Harnstoff-Ratio war auch in der Vorgängerstudie genomweit signifikant ($p < 1,1 \times 10^{-15}$).

Tabelle 6.6.: Assoziationen zu rs4921914 (NAT2)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Formiat / Harnstoff	3.825	0,3197	$6,42 \times 10^{-64}$	$2,72 \times 10^{13}$
Formiat / Succinat	3.821	0,2850	$2,06 \times 10^{-59}$	$8,59 \times 10^8$
Formiat / cis-Aconitat	3.620	0,2788	$4,82 \times 10^{-56}$	$3,62 \times 10^5$
Formiat	3.829	0,2954	$1,74 \times 10^{-50}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Im eben genannten Kapitel 6.1 wurde zudem ein Zusammenhang zwischen der Formiat-Konzentration im Urin und dem Acetylierungs-Phänotypen vermutet. In früheren Studien wurde gezeigt, dass es möglich ist, anhand des Genotyps mehrerer SNPs im NAT2-Locus mit sehr hoher Genauigkeit vorherzusagen, ob der Träger dieser SNPs ein langsamer, mittelschneller oder schneller Acetylierer ist. So stellten Kuznetsov *et al.* 2010 den Webserver NAT2PRED vor, der für diese Klassifizierung sechs SNPs benötigt [140]. Inzwischen wurden weitere Methoden vorgestellt, um mit zwei oder gar nur einem SNP den Acetylierungs-Phänotypen vorherzusa-

gen [141, 142]. Einer dieser „Tagging-SNPs“ ist rs1495741, der vollständig mit dem in der hier besprochenen Studie identifizierten rs4921914 im LD ist. Allerdings ist dieser Tagging-SNP weder im SHIP- noch im KORA-Datensatz genotypisiert. Um diesen SNP dennoch auf Assoziation mit Formiat hin zu testen wurde mittels Imputation der Genotyp von rs1495741 errechnet. Die Häufigkeit des seltenen Allels von rs1495741 beträgt 24,3%. Die Assoziation dieses SNPs mit Formiat ist bezüglich des p-Wertes ($7,51 \times 10^{-40}$) und des Effektes ($\beta' = 0,2885$) mit der Assoziation des SNPs rs4921914 zu Formiat vergleichbar (Tabelle 6.6). Ähnlich verhält es sich, wenn rs1495741 mit den Ratios aus Formiat/Harnstoff und Formiat/Succinat getestet wird.

Laut He *et al.* sind homozygote Träger des seltenen Allels schnelle Acetylierer; homozygote Träger des häufigen Allels demnach langsame Acetylierer [142]. Übertragen auf die Ergebnisse aus der hier besprochenen Studie (und ihrer Vorgängerstudie) bedeutet dies, dass schnelle Acetylierer tendenziell mehr Formiat ausscheiden als langsame.

In den auf Messungen in Blutserum beruhenden GWAS im KORA-F4- und TwinsUK-Datensatz assoziiert dieser SNP mit dem 1-Methylxanthin/4-Acetamidobutanoat-Ratio (KORA: $p < 4,03 \times 10^{-25}$) und der 4-Acetamidobutanoat-Konzentration (TwinsUK: $p < 5,57 \times 10^{-47}$), wobei das seltene Allel des SNPs in beiden Studien einen positiven Effekt auf die 4-Acetamidobutanoat-Konzentration und im KORA-F4-Datensatz einen negativen Effekt auf die gemessene Menge an 1-Methylxanthin hat. 1-Methylxanthin ist ein Abbauprodukt von Koffein. Seit Mitte der 1980er Jahre wird die nach definierter Koffeineinnahme zu mehreren Zeitpunkten bestimmte Konzentration von 1-Methylxanthin im Urin dazu herangezogen, um die Aktivität der an der Aktivierung bzw. Detoxifizierung von Karzinogenen und Xenobiotika (wie Koffein) beteiligten Enzyme CYP1A2, Xanthin-Oxidase und NAT2 zu bestimmen [143–145]. 4-Acetamidobutanoat ist ebenfalls mit dem Stoffwechsel von Xenobiotika verbunden. Es ist ein Produkt einer NAD-abhängigen Aldehyddehydrogenase (EC 1.2.1.3), welche durch Gene der ALDH-Familie codiert werden. Manche dieser ALDH-Enzyme sind an der xenobiotischen Bioaktivierung beteiligt [146, 147]. Der Zusammenhang zwischen 4-Acetamidobutanoat und dem NAT2-Locus ist allerdings zum gegenwärtigen Zeitpunkt noch unklar.

Ausgehend von den hier vorgestellten Assoziationen im NAT2-Locus scheint es überlegenswert, die Menge an ausgeschiedenem Formiat zur Charakterisierung des

Acetylierungs-Phänotypes zu verwenden oder zumindest zusätzlich zu dem bereits bestehenden, auf dem Koffein-Metabolismus basierenden Verfahren heranzuziehen, um eine genauere Phänotypisierung zu erreichen.

Die Assoziation des Iminosäuretransporters SLC6A20 gibt neue Hinweise auf die Struktur eines assoziierten Serum-Metaboliten

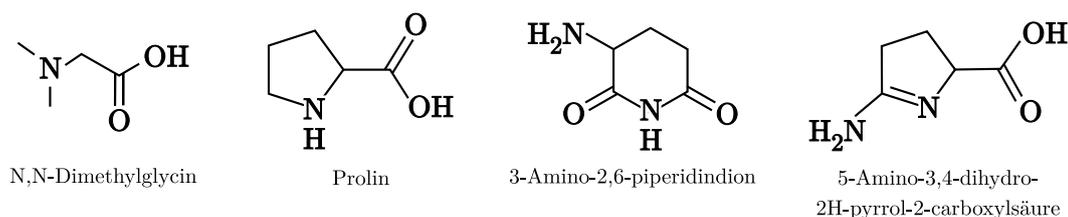
Der SNP rs17279437 befindet sich auf dem 3. Chromosom im codierenden Bereich des SLC6A20-Gens. Dieses Gen codiert den SIT1-Transporter für Iminosäuren wie die proteinogene Aminosäure Prolin sowie für N-methylierte Aminosäuren. Für nicht-methylierte Aminosäuren zeigt SIT1 nur schwache Transporteraktivität [117]. Der oben genannte SNP verursacht im SIT1-Protein einen Austausch des 199. Aminosäurerestes, wobei Threonin durch Methionin ersetzt wird.

Bei der in Kapitel 6.1 besprochenen Vorgängerstudie wurden mehrere signifikante Assoziationen von rs17297437 mit N,N-Dimethylglycin und davon abgeleiteten Ratios identifiziert. Das Konzentrationsverhältnis von N,N-Dimethylglycin zu Alanin zeigte bei einem p-Wert kleiner $2,3 \times 10^{-13}$ das stärkste Assoziationssignal, gefolgt von den Assoziationen mit den Ratios N,N-Dimethylglycin/Tyrosin ($p < 3,2 \times 10^{-13}$) und N,N-Dimethylglycin/Threonin ($p < 5,6 \times 10^{-12}$) sowie mit der Konzentration von N,N-Dimethylglycin ($p < 5,5 \times 10^{-12}$), wobei die Probanden bei Vorliegen des seltenen rs17279437 tendenziell weniger N,N-Dimethylglycin ausschieden ($\beta' = -0,2552$). Es wurde vermutet, dass der genetische Polymorphismus die Affinität des durch SLC6A20 codierten Transporters zu N-methylierten Aminosäuren verringert.

Die 2010 identifizierten Assoziationen konnten in der hier besprochenen Studie größtenteils repliziert werden (Tabelle 6.7). Die Menge des ausgeschiedenen N,N-Dimethylglycins sinkt dabei ebenfalls mit der Anzahl der seltenen Allele von rs17279437 ($\beta' = -0,2496$). Die stärkste Assoziation in der hier besprochenen Studie ist zum N,N-Dimethylglycin/Tyrosin-Ratio bei einem p-Wert kleiner $4,0 \times 10^{-58}$. Jedoch konnte die in der Vorgängerstudie als stärkste identifizierte Assoziation mit N,N-Dimethylglycin/Alanin hier nicht repliziert werden, da der SNP mit diesem Ratio schwächer assoziierte als mit N,N-Dimethylglycin alleine. Es wurde zunächst vermutet, dass dies auf eine unterschiedliche Annotation der NMR-Spektren zurückzuführen sei. Die beiden GWAS basieren zwar auf demselben NMR-Rohdatensatz, die Quantifizierung der Metabolitenkonzentrationen erfolgte aber jeweils durch verschie-

dene Annotatoren. Jedoch sind die ermittelten Konzentrationen von Alanin in beiden Annotationsversionen nahezu perfekt korreliert (Pearson-Korrelationskoeffizient $r = 0,995$), wenn die 918 Proben betrachtet werden, die sowohl in der älteren als auch in der jüngeren Annotation des SHIP-0-Datensatzes enthalten sind. Ausgehend von dieser hohen Korrelation ist ein Einfluss unterschiedlicher Annotatoren unwahrscheinlich.

Abbildung 6.12.: N,N-Dimethylglycin, Prolin, Pyroglutamin (Metabolon) und 5-Amino-3,4-dihydro-2H-pyrrol-2-carboxylsäure



Von links nach rechts: N,N-Dimethylglycin, Prolin, 3-Amino-2,6-piperidindion (Metabolon: Pyroglutamin; ChemSpider-ID 118562), 5-Amino-3,4-dihydro-2H-pyrrol-2-carboxylsäure (ChemSpider-ID 13439810)

Im Serum der TwinsUK-Studie gibt es eine Assoziation ($p < 7,76 \times 10^{-27}$, $\beta = 0,091$) zwischen rs17279437 und dem Ratio aus Pyroglutamin und X-11315, wobei Letzteres ein zwar massenspektrometrisch gemessener, aber bislang noch nicht biochemisch identifizierter Metabolit ist. Auch bei der chemischen Identität des Metaboliten, der von Metabolon als Pyroglutamin bezeichnet wird, bestehen noch Unsicherheiten. Aufgrund der per Massenspektrometrie bestimmten molekularen Masse und des spektralen Refraktionsmusters vermutet Metabolon, dass dieser Metabolit 3-Amino-2,6-piperidindion (PubChem-ID 134508 [148]) entspricht (siehe Abbildung 6.12 Mitte). Laut ChemSpider ist ein möglicher Alternativname für diese Verbindung Pyroglutamin (CSID:118562, Stand 11. April 2013) [149]. Allerdings gibt es Zweifel, ob diese Zuordnung korrekt ist. Eine Molekül mit identischer molekularer Masse ist 5-Amino-3,4-dihydro-2H-pyrrol-2-carboxylsäure (CSID:13439810, Stand 11. April 2013; Abbildung 6.12 rechts). Im Gegensatz zu Pyroglutamin weist dieses Molekül die Eigenschaften einer Iminosäure auf (vergleiche die Struktur von Prolin in Abbildung 6.12) und passt daher besser zur identifizierten Assoziation mit dem SLC6A20-Gen, das wie eingangs erwähnt einen Iminosäuretransporter codiert. Eine experimentelle Bestätigung oder Widerlegung dieser Vermutung seitens Metabolon steht noch aus. Dieses Beispiel zeigt jedoch prinzipiell, wie Assoziationen zwischen Genen mit bekannter Genprodukt-Funktion und Metaboliten mit ungewisser chemischer Struktur

dabei helfen können, die Charakterisierung von gemessenen Stoffwechselprodukten zu verbessern.

Tabelle 6.7.: Assoziationen zu rs17279437 (SLC6A20)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
N,N-Dimethylglycin / Tyrosin	3.150	-0,3112	$4,03 \times 10^{-58}$	$9,05 \times 10^{33}$
N,N-Dimethylglycin / Threonin	3.667	-0,2927	$1,75 \times 10^{-54}$	$2,09 \times 10^{30}$
N,N-Dimethylglycin / Glycolat	3.645	-0,3003	$2,46 \times 10^{-52}$	$1,48 \times 10^{28}$
N,N-Dimethylglycin / Glycin	3.705	-0,2945	$4,93 \times 10^{-50}$	$7,40 \times 10^{25}$
N,N-Dimethylglycin / Ethanolamin	3.496	-0,2890	$1,19 \times 10^{-47}$	$3,06 \times 10^{23}$
N,N-Dimethylglycin / Dimethylamin	3.708	-0,2880	$1,17 \times 10^{-46}$	$3,12 \times 10^{22}$
N,N-Dimethylglycin / Pyroglutamat	3.610	-0,2974	$2,00 \times 10^{-46}$	$1,82 \times 10^{22}$
N,N-Dimethylglycin / Valin	3.002	-0,2721	$2,54 \times 10^{-42}$	$1,43 \times 10^{18}$
N,N-Dimethylglycin / Histidin	3.590	-0,2925	$1,86 \times 10^{-41}$	$1,96 \times 10^{17}$
N,N-Dimethylglycin / Succinat	3.702	-0,2747	$4,81 \times 10^{-41}$	$7,58 \times 10^{16}$
N,N-Dimethylglycin / cis-Aconitat	3.535	-0,2774	$5,82 \times 10^{-37}$	$6,27 \times 10^{12}$
N,N-Dimethylglycin / Formiat	3.686	-0,2692	$2,18 \times 10^{-33}$	$1,67 \times 10^9$
N,N-Dimethylglycin / Lactat	3.690	-0,2909	$4,45 \times 10^{-33}$	$8,19 \times 10^8$
N,N-Dimethylglycin / Methanol	3.672	-0,2680	$4,74 \times 10^{-32}$	$7,70 \times 10^7$
N,N-Dimethylglycin / Harnstoff	3.705	-0,2700	$1,52 \times 10^{-31}$	$2,40 \times 10^7$
N,N-Dimethylglycin / N,N-Methylhistidin	2.889	-0,2759	$6,03 \times 10^{-31}$	$6,05 \times 10^6$
N,N-Dimethylglycin / Trimethylamin-N-oxid	3.695	-0,2956	$2,39 \times 10^{-28}$	$1,53 \times 10^4$
N,N-Dimethylglycin	3.708	-0,2496	$3,65 \times 10^{-24}$	

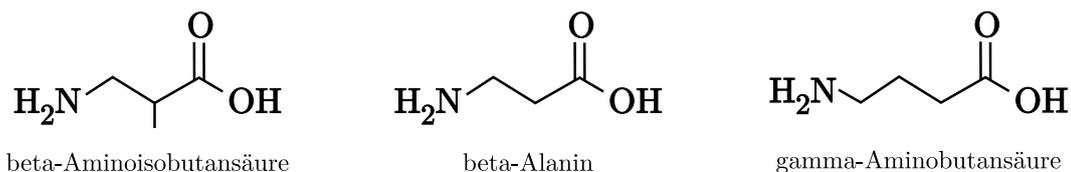
Es gelten analog die Erläuterungen zu Tabelle 6.4.

beta-Aminoisobutyrat ist möglicherweise ein Substrat des SLC6A13-Transporters

Der SNP rs11613331 liegt auf dem 12. Chromosom und befindet sich im intronischen Bereich des SLC6A13-Gens sowie ca. 30 kB entfernt vom 5'-Ende des SLC6A12-Gens. Diese Gene codieren Transporter, die eine essenzielle Rolle bei der Regulierung von Neurotransmittern spielen, indem sie den Austausch von Neurotransmittern vom extrazellulären Raum in Neuronen bzw. Gliazellen vermitteln [150]. SLC6A12 und SLC6A13 transportieren den inhibitorischen Neurotransmitter gamma-Aminobutansäure (GABA), der die neuronale Erregbarkeit im zentralen Nervensystem reguliert. SLC6A12 (in der Literatur oft als BGT1 bezeichnet) und SLC6A13 (alternativ GAT2) werden vorrangig im Gehirn, in der Leber und in der Niere exprimiert, wie beispielsweise Zhou *et al.* im Mäuseversuch experimentell zeigten [150–152]. RNA-Sequenzierungsdaten des „Illumina Human BodyMap 2.0“-Projektes, bei dem Genexpressionsprofile in 16 verschiedenen menschlichen Gewebetypen erstellt wurden, legen nahe, dass die experimentellen Daten im Mäuseversuch auch auf den Menschen übertragbar sind [153]. Die Bedeutung des SLC6A13-Transporters für die Niere zeigten Köttgen *et al.* in einer 2010 veröffentlichten GWAS. Darin wurden genetische Loci auf Assoziationen zu Nierenfunktion und chronischen Nierenerkrankungen untersucht. So assoziiert SNP rs10774021, der laut HapMap stark im LD mit rs11613331 ist ($r^2 = 0,78$), mit Kreatinin im Blutserum, einem Marker für Nierenfunktion [47].

In der vorliegenden Studie assoziiert SNP rs11613331 im SHIP-Datensatz genomweit signifikant mit der Konzentration von beta-Aminoisobutyrat (auch 3-Aminoisobutyrat, abgekürzt BAIB) bei $p < 5,6 \times 10^{-14}$, wobei der Effekt ($\beta = -0,154$) eine tendenziell niedrigere BAIB-Ausscheidung pro Kopie des seltenen rs11613331-Allels andeutet.

Abbildung 6.13.: beta-Aminoisobutansäure, beta-Alanin und gamma-Aminobutansäure



BAIB ist laut OMIM-Eintrag %210100 [87] (abgerufen am 7. Juli 2013) ein strukturelles Analogon zu dem Neurotransmitter GABA (Abbildung 6.13 rechts). Matskevitch *et al.* zeigten in einer experimentellen Studie in *Xenopus* Oocyten die Affinität des SLC6A12-Transporters zu beta-Alanin (Abbildung 6.13 Mitte), welches wie BAIB zu den beta-Aminosäuren zählt [154]. Auch für SLC6A13 wurde experimentell Transportaktivität für beta-Alanin nachgewiesen [155]. Möglicherweise verändert ein mit rs11613331 gekoppelter Polymorphismus die Affinität der SLC6A12- bzw. SLC6A13-Transporter zu Gunsten von BAIB und beeinflusst somit den Transport dieses Metaboliten.

Diese Vermutung wird gestützt von einer im Sommer 2013 durch Rhee *et al.* veröffentlichten GWAS mit Metabolitenkonzentrationen im Blutplasma von über 2000 Probanden der „Framingham Heart Study“ [42]. Die Autoren identifizierten ebenfalls eine genomweit signifikante Assoziation von rs11613331 zu BAIB ($p < 2,0 \times 10^{-12}$). Auch im Blutplasma ist bei Vorliegen des Effektallels (seltenes Allel) die BAIB-Konzentration tendenziell niedriger ($\beta = -0,23$). Im Serum des KORA-F4- sowie des TwinsUK-Datensatzes gibt es ebenfalls signifikante Assoziationen von rs11613331 zur Konzentration von Pyroglutamin ($p < 1,15 \times 10^{-9}$, $\beta = -0,0408$ in KORA und $p < 7,35 \times 10^{-18}$, $\beta = -0,035$ in TwinsUK). Aufgrund der im vorhergehenden Abschnitt (Kapitel 6.2) angesprochenen Unsicherheit über die chemische Identität des von Metabolon als Pyrolutamin bezeichneten Metaboliten ist es jedoch schwierig, ihn in Verbindung zu BAIB oder der Funktion des SLC6A12/13-Transporters zu setzen.

Tabelle 6.8.: Assoziationen zu rs11613331 (SLC6A13)

Metabolit oder Ratio	N	beta'	p-Wert
beta-Aminoisobutytrat	3.345	-0,1544	$5,61 \times 10^{-14}$

Es gelten analog die Erläuterungen zu Tabelle 6.4.

SLC7A9 zeigt Transportaktivität für die eng verwandten Metaboliten Lysin und Homocitrullin

In Kapitel 6.1 wurde gezeigt, dass ein genetischer Polymorphismus (SNP rs8101881) in der Nähe des für den SLC7A9-Transporter kodierenden Gens mit dem Ratio aus Lysin und Valin assoziiert ($p < 1,8 \times 10^{-14}$). SLC7A9 ist laut RefSeq beteiligt am Transport von Cystin sowie neutralen und dibasischen Aminosäuren (Kommentar zur Sequenz NM_014270.4, Stand Juli 2011, abgerufen am 4. September 2013) [90]. Der Transporter spielt eine wichtige Rolle bei der Reabsorption dieser Aminosäuren in die luminale Membranen des proximalen Tubulus der Niere [90, 156, 157]. Die Assoziationsergebnisse ließen vermuten, dass die dibasische Aminosäure Lysin ein Substrat von SLC7A9 ist.

In der hier besprochenen Studie repliziert diese Assoziation bei einem p-Wert kleiner $2,06 \times 10^{-27}$. Der Trend, also die beobachtete Konzentrationsänderung pro Kopie des Effektalles, geht bei beiden Studien in die gleiche Richtung und ist in der neueren Studie etwas ausgeprägter (alte Studie: $\beta = 0,2853$, neue Studie: $\beta = 0,3788$).

Obwohl im Vergleich zur Vorgängerstudie lediglich 252 zusätzliche Messwerte für Lysin/Valin zur Verfügung standen, ist die Assoziation im neu annotierten SHIP-0-Datensatz um knapp dreizehn Größenordnungen stärker. Da die Annotation des Lysin-Signals in den NMR-Spektren aufgrund von überlappenden, von anderen Metaboliten herrührenden Resonanzen aufwändig und daher potentiell fehlerbehaftet ist, könnte dies auf eine qualitativ verbesserte Annotation des neueren Datensatzes hinweisen. Allerdings ist die Korrelation der Lysin-Konzentrationen bei Proben, die sowohl in dem älteren als auch in dem neueren Datensatz annotiert wurden, nahezu perfekt (Pearson-Korrelationskoeffizient $r = 0,976$). Die Korrelation der augenscheinlich leichter annotierbaren Valin-Konzentrationen ist mit $r = 0,91$ ebenfalls sehr hoch.

Betrachtet man die Ergebnisse der GWAS, die im Serum der KORA-F4- und TwinsUK-Studie durchgeführt wurden, so zeigen sich schwache, aber dennoch signifikante Assoziationen von rs8101881 zu Homocitrullin (KORA: $p < 2,49 \times 10^{-8}$; TwinsUK: $p < 3,45 \times 10^{-10}$) (Tabelle 5.5). Der verringernde Effekt des seltenen Allels auf die Homocitrullin-Konzentration ist in beiden Studien konsistent (KORA F4: $\beta = -0,035$; TwinsUK: $\beta = -0,033$).

Abbildung 6.14.: Carbamoylierung von Lysin zu Homocitrullin (EC 2.1.3.8)

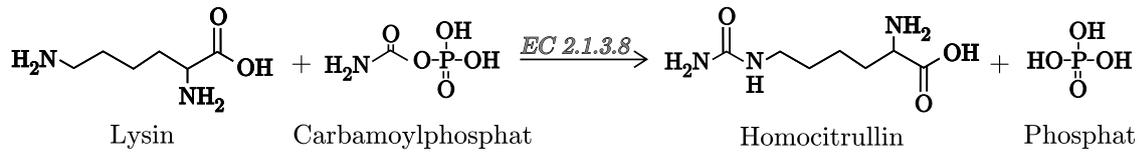


Abbildung nach KEGG Reaktion R01396 [89]

Homocitrullin entsteht aus Lysin durch Carbamoylierung (Abbildung 6.14). Bei Proteinen werden durch post-translationale Modifikation Lysin-Reste zu 6-Carbamoyllysin (Homocitrullin) umgewandelt, wodurch Proteinstruktur und -funktion verändert werden. In früheren Studien wurde vermutet, dass diese Protein-Carbamoylierung eine Schlüsselrolle bei Entzündungsreaktionen und damit auch bei der Entstehung von Autoimmunkrankheiten wie Arthritis spielt [158–160]. Hommes *et al.* vermuten, dass ein mitochondriales Enzym (Lysin-Carbamoyltransferase, EC 2.1.3.8) Lysin zu Homocitrullin umwandelt [161, 162]. Abgesehen von den Arbeiten von Hommes *et al.* wird dieses Enzym aber kaum erwähnt, und so ist bislang nicht geklärt, welches Enzym die Carbamoylierung steuert oder ob die Reaktion überhaupt enzymatisch gesteuert wird [163].

Al-Dirbashi *et al.* und Kraus *et al.* zeigten, dass bei Patienten mit Nierenerkrankungen bzw. mit Urämie die Konzentration von Homocitrullin im Harn erhöht ist [164, 165]. Patienten, die am Hyperornithinämie-Hyperammonämie-Homocitrullinurie-Syndrom leiden, zeigen stark erhöhte Homocitrullin-Werte im Urin [165]. Diese seltene, autosomal rezessive Erbkrankheit wird durch eine Mutation im SLC25A15-Gen verursacht (OMIM #238970). Durch diese Mutation ist der Transport von Ornithin ins Mitochondrium gestört. Das sich anstauende Carbamoylphosphat verbindet sich mit Lysin zu Homocitrullin, welches letztendlich über den Harn ausgeschieden wird [166, 167]. Auf ähnliche Weise könnte eine veränderte Lysin-Transportaktivität von SLC7A9 Auswirkungen auf die Homocitrullin-Homöostase im Blut haben: bei erhöhter Lysin-Ausscheidung (positiver Trend bei der Assoziation) wird entsprechend weniger Lysin carbamoyliert, was in einer verringerten Homocitrullin-Konzentration im Blut resultiert (negativer Trend).

Tabelle 6.9.: Assoziationen zu rs8101881 (SLC7A9)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Lysin / Valin	868	0,3788	$2,06 \times 10^{-27}$	$2,68 \times 10^{11}$
Lysin / Tyrosin	911	0,3431	$4,33 \times 10^{-25}$	$1,27 \times 10^9$
Lysin / Ethanolamin	912	0,3715	$1,21 \times 10^{-24}$	$4,56 \times 10^8$
Lysin / Pyroglutamat	928	0,3766	$4,49 \times 10^{-24}$	$1,23 \times 10^8$
Lysin / Succinat	935	0,4093	$5,02 \times 10^{-24}$	$1,10 \times 10^8$
Lysin / Threonin	933	0,3332	$2,74 \times 10^{-23}$	$2,01 \times 10^7$
Lysin / Glycolat	925	0,3599	$3,89 \times 10^{-23}$	$1,41 \times 10^7$
Lysin / N,N-Dimethylglycin	927	0,4233	$4,79 \times 10^{-23}$	$1,15 \times 10^7$
Lysin / Methanol	930	0,4080	$1,91 \times 10^{-22}$	$2,88 \times 10^6$
Lysin / Dimethylamin	937	0,3671	$4,80 \times 10^{-22}$	$1,15 \times 10^6$
Lysin	937	0,2918	$5,50 \times 10^{-16}$	
Succinat / Tyrosin	3.197	-0,0590	$2,41 \times 10^{-6}$	$1,68 \times 10^4$

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Der für einen Carnitin-Transporter codierende SLC16A9-Locus assoziiert mit dem strukturverwandten Osmolyten Betain

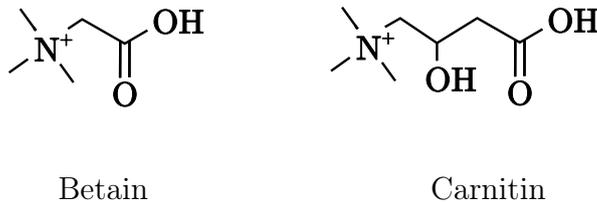
Der SNP rs7094971 liegt im intronischen Bereich des SLC16A9-Gens, welches für einen Transporter codiert. Suhre *et al.* fanden 2011 eine signifikante Assoziation dieses SNPs zur Carnitin-Konzentration im Blutserum. Mit zunehmendem Vorliegen des Effekt-Allels nimmt die Konzentration des Carnitins ab ($\beta = -0,022$) [35]. Im Datensatz der 2014 veröffentlichten TwinsUK-Studie ist diese Assoziation ebenfalls zu sehen ($p < 2,18 \times 10^{-45}$, $\beta = -0,028$). Suhre *et al.* lieferten in ihrer Studie durch Versuche mit *Xenopus* Oocyten den experimentellen Beweis, dass SLC16A9 tatsächlich ein Transporter für freies Carnitin ist.

Im SHIP-0-Datensatz der hier besprochenen Studie assoziiert rs7094971 mit dem Ratio aus Betain und Citrat bei $p < 1,41 \times 10^{-12}$; der positive Regressionskoeffizient ($\beta' = 0,183$) zeigt, dass Träger des Effektallels tendenziell mehr Betain ausscheiden. Allerdings beträgt der p-Gain des Ratios lediglich $1,19 \times 10^4$. Somit ist diese Assoziation nominell nicht signifikant, da die untere Grenze für den p-Gain bei

$1,51 \times 10^4$ liegt. Da die p-Gain-Grenze lediglich knapp unterschritten wird, wird auf diese Assoziation hier dennoch eingegangen.

Wie Sizeland *et al.* beschreiben, ist Betain in der Niere ein wichtiger Osmolyt, der aktiv intrazellulär im Nierenmark akkumuliert wird, um bei hoher extrazellulärer Konzentration von gelösten Stoffen den osmotischen Gradienten durch die Zellmembran niedrig zu halten [168]. Der Zusammenhang zwischen erhöhter Betain-Ausscheidung und verminderter Nierenfunktion wurde von Lever, Chambers, Dellow *et al.* in Studien mit Nierenkranken und Diabetikern mehrfach beschrieben [169–172]. Die Autoren dieser Studien vermuten, dass eine erhöhte Betain-Ausscheidung chronisch und gesundheitsschädigend sei [171].

Abbildung 6.15.: Betain und Carnitin



Carnitin und Betain (alias N,N,N-Trimethylglycin oder Glycylbetain) weisen aufgrund ihrer charakteristischen dreifach-methylierten Ammoniumgruppe und der Carboxygruppe strukturelle Ähnlichkeit auf, wie Abbildung 6.15 zeigt. Betain könnte also wie Carnitin ebenfalls ein Substrat des SLC16A9-Transporters sein. Ein mit rs7094971 gekoppelter Polymorphismus könnte mit einer veränderten Affinität von SLC16A9 zu Betain und damit mit dessen erhöhter Ausscheidung einhergehen. Ein Einfluss dieses SNPs auf die Menge des mit dem Harn ausgeschiedenen Carnitins konnte in den hier untersuchten SHIP-0- und KORA-F4-Proben allerdings nicht beobachtet werden.

Auch wenn die Stärke der Assoziation des SLC16A9-Locus zu der Betain-Konzentration im Harn nominell statistisch unsignifikant ist, ist der hier beschriebene genetisch beeinflusste Metabotyp gerade im Hinblick auf Nierenerkrankungen interessant, weshalb SLC16A9 in künftigen Studien als potentieller Risikolocus betrachtet werden könnte.

Tabelle 6.10.: Assoziationen zu rs7094971 (SLC16A9)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Betain / Citrat	3.798	0,1853	$1,41 \times 10^{-12}$	$1,19 \times 10^4$
Betain	3.798	0,1491	$1,68 \times 10^{-8}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Loci, die nur im Urin mit metabolischen Merkmalen assoziieren

Ein kausativer SNP im AGXT2-Locus zeigt eine enorm starke Assoziation zu beta-Aminoisobutyrat

Der SNP rs37369 liegt auf dem 5. Chromosom im codierenden Bereich des Alanin-Glyoxylat-Aminotransferase-2-Gens und hat einen nicht-synonymen Austausch des 140. Aminosäurerestes von Valin durch Isoleucin im AGTX2-Protein zur Folge. Wie in Kapitel 6.1 beschrieben, gibt es eine signifikante und replizierbare Assoziation zwischen rs37369 und der beta-Aminoisobutyrat-Konzentration im Urin. Beta-Aminoisobutyrat (auch 3-Aminoisobutyrat, abgekürzt BAIB) ist dabei ein Substrat der von AGXT2 katalysierten Reaktion (EC 2.6.1.40) [86, 88]. Aufgrund der extrem hohen Assoziationsstärke ist zu vermuten, dass rs37369 ein kausativer SNP ist, der die katalytischen Fähigkeiten von AGXT2 direkt beeinflusst und bei homozygotem Vorliegen des rezessiven Allels zu beta-Aminoisobutansäure-Azidurie führt.

In der vorliegenden GWAS konnte erneut die Assoziation zwischen dem Genotyp von rs37369 und dem metabolischen Phänotyp, also der Konzentration von beta-Aminoisobutyrat im Urin, sowohl im SHIP-0-Datensatz ($p < 3,8 \times 10^{-307}$) als auch im zur Replikation verwendeten KORA-F4-Urin-Datensatz ($p < 3,2 \times 10^{-116}$) identifiziert werden (Tabelle 6.11). Im Blut der KORA-F4- sowie der TwinsUK-Studien finden sich keine signifikanten Assoziationen des AGXT2-Locus. Jedoch wurde in diesen Datensätzen BAIB weder durch die Biocrates- noch durch die Metabolon-Plattform identifiziert bzw. quantifiziert. Bei der in Kapitel 6.2 bereits erwähnten, 2013 von Rhee *et al.* veröffentlichten GWAS mit Plasmametaboliten wurde auf einer eigenentwickelten GC/LC-MS-Plattform ein 217 Metaboliten umfassendes Panel gemessen. Laut Autoren dieser Studie beinhaltet das Panel über 100 Metaboliten, die in bisher

veröffentlichten GWAS im Blut noch nicht betrachtet wurden [42]. Neben einer signifikanten Assoziation des SLC6A13-Locus zur BAIB-Konzentration nennen Rhee *et al.* den AGXT2-Locus, der mit diesem Metaboliten assoziiert. Dabei ist die Assoziation des SNP rs37370 (laut HapMap im LD mit rs37369) bei einem p-Wert von $5,8 \times 10^{-83}$ sehr deutlich. Das seltene Allel geht mit einer starken Erhöhung der BAIB-Konzentration sowohl im Urin ($\beta' = 2,47$) als auch im Blut ($\beta = 1,02$) einher. Aufgrund einer schwachen Assoziation des AGXT2-Locus mit Triglyceriden (TG) und Cholesterinestern (CE) untersuchten Rhee *et al.* an Zebrafisch-Embryonen, ob sich ein Knockout des AGXT2-Gens auf die Lipid-Homöostase im Fisch auswirkt. Tatsächlich war eine Anreicherung von TG und eine Abreicherung von CE infolge des Knockouts messbar.

Der AGXT2-Locus ist also nicht nur hinsichtlich seiner vermutlichen Kausalität für beta-Aminoisobutansäure-Azidurie interessant, sondern könnte auch eine wichtige Rolle bei klinisch relevanten Veränderungen des Lipidhaushaltes spielen.

Tabelle 6.11.: Assoziationen zu rs37369 (AGXT2)

Metabolit oder Ratio	N	beta'	p-Wert
beta-Aminoisobutyrat	3.344	2,467	$3,76 \times 10^{-307}$
Ethanol	2.767	0,1707	$8,84 \times 10^{-6}$

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Ein nicht-synonymer Polymorphismus im XYLB-Gen könnte zu einem alternativen Xylulose-Abbaupfad führen

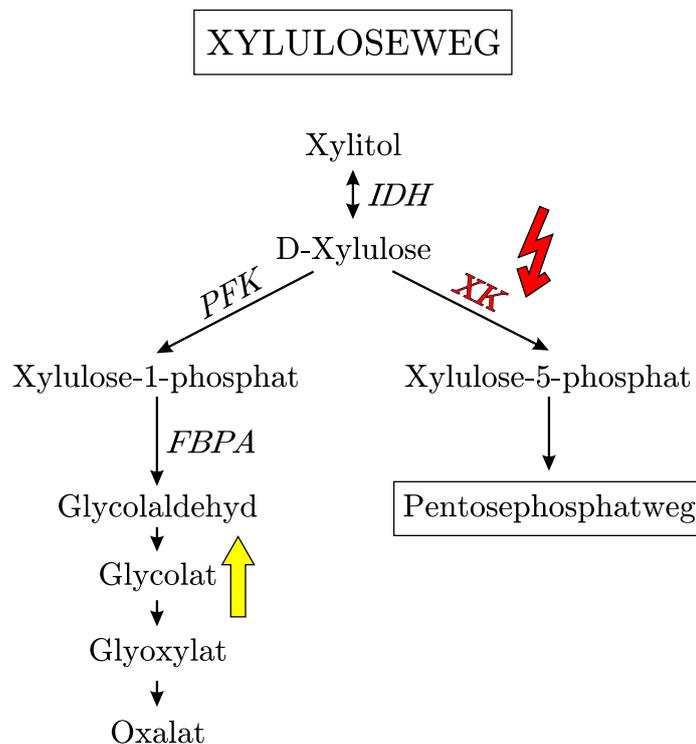
SNP rs17118 liegt auf dem 3. Chromosom im codierenden Bereich des XYLB-Gens und verursacht einen Aminosäureaustausch an Position 85 im codierten Protein, wobei Asparaginsäure durch Glutaminsäure ersetzt wird. Dieser SNP assoziiert bei einem p-Wert kleiner $4,7 \times 10^{-11}$ mit dem Ratio aus Glycolat und Threonin (Tabelle 6.12).

Aufgrund der 22-prozentigen Sequenzidentität zur Xylulokinase in *Hemophilus influenza* wurde vermutet, dass das XYLB-Gen ein zur Xylulokinase homologes Protein kodiert. Eine noch höhere Sequenzidentität besteht zu Genprodukten, die in *C. elegans* und Hefe identifiziert wurden und von denen vermutet wird, dass sie unter

anderem zur Familie der Glucokinasen, Glycerokinasen und Xylulokinasen gehören [173]. Vor kurzem konnte schließlich experimentell nachgewiesen werden, dass das XLYB-Gen tatsächlich für menschliche Xylulokinase codiert [174].

Die Xylulokinase katalysiert im Pentosephosphatweg die Phosphorylierung (EC 2.7.1.17) von D-Xylulose zu D-Xylulose-5-phosphat (X5P). Die in Leber und Niere aktive Xylulokinase ist das letzte Enzym im Glucuronat-Xylulose-Pathway und ist durch ihr Produkt X5P mit dem Pentosephosphat-Abbaupfad verbunden: X5P kann nach einer weiteren Umwandlung zu D-Glycerinaldehyd-3-phosphat (auch beta-D-Fructose-6-phosphat) in der Glycolyse verstoffwechselt werden (EC 2.2.1.1) (Abbildung 6.16) [174].

Abbildung 6.16.: Abbau von D-Xylulose über zwei Stoffwechselfade



Zwei Abbauewege für Xylulose. Rechts der übliche Weg mittels Xylulokinase, links der seltenere Pfad mittels Fructokinase. Gelber Pfeil: erhöhte Ausscheidung von Glycolat bei Trägern des seltenen Allels von rs17118 (nicht-synonymer SNP in XLYB). IDH: L-Iditoldehydrogenase, XK: Xylulokinase, PFK: Phosphofruktokinase, FBPA: Fructosebiphosphataldolase. Abbildung nach Barngrover *et al.* sowie Holmes und Assimios [175, 176].

Eine alternative Möglichkeit des Xylulose-Abbaus, die in der Vergangenheit untersucht wurde, erfolgt in der Leber durch Fructokinase (EC 2.7.1.3), wobei D-Xylulose-1-phosphat (kurz X1P) gebildet wird. Dieses wird durch Aldolase in Dihydroxyacetonphosphat und Glycolaldehyd gespalten. Letzteres wird schließlich zu dem mit rs17188 assoziierten Glycolat sowie weiter zu Glyoxylat und Oxalat (beide in dieser Studie nicht im Urin gemessen) oxidiert (Abbildung 6.16) [175–178]. Aufgrund der in ihrer Studie gemessenen enzymkinetischen Daten vermuten Bunker *et al.* allerdings, dass der Großteil an Xylulose nicht durch Fructokinase, sondern durch Xylulokinase metabolisiert wird. Bär und Oesterhelt ermittelten in einer Studie mit oral verabreichtem isotonenmarkiertem Xylitol (metabolische Vorstufe von D-Xylulose), dass aufgenommenes Xylitol nur zu ungefähr 0,007 % in Oxalat umgewandelt wird (70 mg pro 1 kg aufgenommenem Xylitol) [179]. Ein von Conyers *et al.* erstelltes Modell gibt eine ebenfalls geringe Umwandlungsrate von 0,5 % von Xylitol zu Oxalat an [180]. Ein intrazellulärer Überschuss an Oxalat kann fatale Folgen bis hin zum Organversagen mit sich bringen. Dies wurde in der Vergangenheit vereinzelt bei Patienten beobachtet, die intravenös mit Xylitol künstlich ernährt wurden [174, 178].

Die im vorliegenden Abschnitt vorgestellten Assoziationsergebnisse weisen auf die Möglichkeit hin, dass durch den nicht-synonymen und möglicherweise kausativen SNP rs17118 die katalytischen Fähigkeiten der Xylulokinase vermindert werden. Als Folge wird der Überschuss an Xylulose durch Fructokinase metabolisiert. Das sich anstauende Glycolat wird dann vermehrt mit dem Urin ausgeschieden, wie das positive beta'-Vorzeichen der Assoziation zeigt. Dies würde zur in früheren Studien gemachten Beobachtung passen, dass die intravenöse Verabreichung von Xylitol zu einer erhöhten Konzentration von Glycolat im Harn der Probanden führt [181]. Um diese These zu stützen wäre es jedoch hilfreich, die metabolischen Zwischenprodukte des Phosphofruktokinase-gesteuerten Xylulose-Abbaus (z.B. X1P oder Oxalat) im Blut und im Urin zu bestimmen.

Die Assoziation zu Ethanolamin bestätigt, dass ETNPPL für eine Ethanolaminphosphat-Phospholyase codiert

Der SNP rs7692671 liegt auf dem 4. Chromosom ca. 30 kb stromaufwärts des ETNPPL-Gens. Der SNP assoziiert mit dem Ratio aus Ethanolamin und Dimethylamin bei einem p-Wert kleiner $6,4 \times 10^{-19}$. Wie Tabelle 6.13 zeigt, gibt es schwächere, aber dennoch genomweit signifikante Assoziationen mit den Ratios von Etha-

Tabelle 6.12.: Assoziationen zu rs17118 (XYLB)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Glycolat / Glycin	3.791	0,0940	$1,46 \times 10^{-12}$	$2,24 \times 10^8$
Glycolat / Threonin	3.747	0,0770	$4,69 \times 10^{-11}$	$6,94 \times 10^6$
Glycolat / Tyrosin	3.160	0,0776	$1,20 \times 10^{-10}$	$2,71 \times 10^6$
Glycolat / Valin	3.027	0,0769	$8,94 \times 10^{-10}$	$3,64 \times 10^5$
Glycolat / Ethanolamin	3.553	0,0656	$1,98 \times 10^{-9}$	$1,65 \times 10^5$
Glycolat	3.793	0,0599	$3,26 \times 10^{-4}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

nolamin zu Pyroglutamat, zu Threonin sowie zu Glycolat. Letzteres ist sowohl mit Glyoxylat als auch über das Zwischenprodukt Glycolaldehyd (EC 1.2.1.21, 1.4.3.8) mit Ethanolamin metabolisch eng verwandt.

Dem ETNPPL-Gen wurde bis 2013 das HGNC-Gensymbol AGXT2L1 zugewiesen. Das Kürzel „L1“ (kurz für *-like 1*) im Gensymbol deutet darauf hin, dass dieses Transkript bis vor kurzem als mutmaßlich codierend für eine Alanin-Glyoxylat-Aminotransferase klassifiziert wurde, welche wie das bereits mehrmals in dieser Arbeit besprochene AGXT2-Enzym (Kapitel 6.1 und 6.2) Alanin und Glyoxylat zu Pyruvat und Glycin katalysiert. 2012 untersuchten Veiga-da-Cunha *et al.* experimentell die katalytische Aktivität des AGXT2L1-Genprodukts und kamen im Gegensatz zur bisherigen Klassifizierung zu dem Ergebnis, dass das von AGXT2L1 codierte Enzym eine Ethanolaminphosphat-Phospholyase ist [182]. Dem AGXT2L1- bzw. ETNPPL-Gen konnte somit ein Enzym zugeordnet werden, welches erstmalig 1970 von Fleshood und Pitot systematisch untersucht wurde [183]. Abbildung 6.17 zeigt die von ETNPPL katalysierte Reaktion. Nach Fleshood und Pitot findet die enzymatische Aktivität der Ethanolaminphosphat-Phospholyase bei Wirbeltieren hauptsächlich in der Leber und teilweise in der Niere statt (gemessen in verschiedenen Gewebetypen bei Kaninchen, Ratten und Ochsen); beim Menschen wird ETNPPL laut RNASeq-Daten des „Illumina Human BodyMap 2.0“-Projektes (Ensembl Release 72) vorrangig im Gehirn und in der Leber exprimiert [75].

Ethanolaminphosphat spielt eine wichtige Rolle bei der Synthese von Phospholipiden. Es wird unter anderem durch Phosphorylierung freien Ethanolamins durch Ethanolaminkinase gebildet (Abbildung 6.17) [184]. Veiga-de-Cunha *et al.* gehen da-

Abbildung 6.17.: Durch ETNPPL katalysierte Reaktion (EC 4.2.3.2)

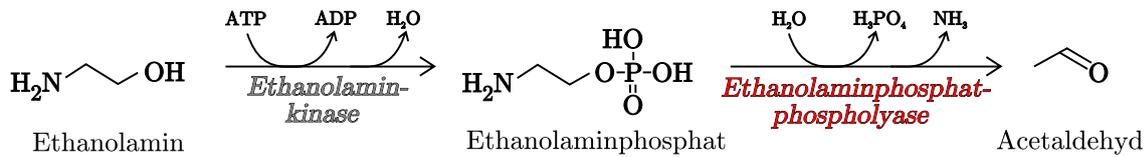


Abbildung nach KEGG Reaktion R00748, Fleshood und Pitot sowie Cunha *et al.* [89, 182, 183]

von aus, dass der Abbau von Ethanolaminphosphat durch ETNPPL dazu dient, die Konzentration dieses Metaboliten im zentralen Nervensystem konstant zu halten. Wird dieser regulatorische Mechanismus gestört, könne es zu Beeinträchtigungen im Phosphatidylethanolamin-Stoffwechsel kommen. Die Autoren folgern, dass diese Beeinträchtigungen im Zusammenhang mit psychischen Erkrankungen wie Schizophrenie stehen könnten [182]. In der Tat wurde in früheren Studien eine veränderte Genexpressionsrate von ETNPPL im Gehirn mit psychischen Störungen in Verbindung gebracht [185, 186]. Möglicherweise gibt es also tatsächlich eine Verbindung zwischen mentalen Störungen und der Ethanolaminphosphat-Phospholyaseaktivität.

In der hier vorgestellten GWAS zeigt bei zunehmender Kopienanzahl des seltenen Allels des SNP rs7692671 eine tendenziell leicht erhöhte Ausscheidung von Ethanolamin ($\beta = 0,06$). Würde der SNP bzw. ein damit gekoppelter genetischer Polymorphismus eine Beeinträchtigung der katalytischen Rate der ETNPPL-gesteuerten Reaktion verursachen, wäre aufgrund des Substrat-Staus eine erhöhte Ausscheidung von Ethanolaminphosphat zu erwarten. Ethanolaminphosphat wurde allerdings in den Urin-Datensätzen dieser Studie nicht quantifiziert. Zudem vermuten Veiga-da-Cunha *et al.*, dass ein ETNPPL-Mangel eher zu Ethanolaminurie denn zu Phosphoethanolaminurie führt, da Ethanolaminphosphat im extrazellulären Raum von alkalischer Phosphatase dephosphoryliert wird [182, 187].

Wie oben erwähnt wurde das ETNPPL-Genprodukt vor kurzem experimentell klassifiziert. Dennoch (oder gerade deshalb) stellt der hier vorgestellte Locus ein gutes Beispiel dar, wie genetische Assoziationen zu metabolischen Merkmalen dabei helfen können, die biochemische Bedeutung ungewiss klassifizierter Gene für den Stoffwechsel besser zu verstehen. Im hier vorgestellten Falle wird die von Veiga-de-Cunha *et al.* durchgeführte experimentelle Arbeit durch die „biologisch sinnvolle“ Assoziation

des ETNPPL-Locus zu Ethanolamin nochmals bestätigt. In anderen Fällen können jedoch Ergebnisse genetischer Assoziationsstudien, die scheinbar im Widerspruch zur Genklassifikation stehen, den Anstoß zur experimentellen Überprüfung der Funktion eines Genproduktes geben.

Tabelle 6.13.: Assoziationen zu rs7692671 (ETNPPL)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Ethanolamin / Dimethylamin	3.597	0,0656	$6,36 \times 10^{-19}$	$1,21 \times 10^{14}$
Ethanolamin / Pyroglutamat	3.508	0,0836	$6,84 \times 10^{-16}$	$1,13 \times 10^{11}$
Ethanolamin / Glycolat	3.553	0,0776	$2,27 \times 10^{-13}$	$3,39 \times 10^8$
Ethanolamin / Threonin	3.558	0,0813	$2,70 \times 10^{-13}$	$2,85 \times 10^8$
Ethanolamin	3.597	0,0642	$7,71 \times 10^{-5}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Ein Polymorphismus im DMGDH-Locus assoziiert mit erhöhter N,N-Dimethylglycin-Ausscheidung

Der SNP rs248386 liegt auf dem 5. Chromosom im intronischen Bereich des Dimethylglycin-Dehydrogenase-Gens (DMGDH). Das von DMGDH codierte mitochondriale Enzym spielt im Stoffwechsel eine wichtige Rolle beim Abbau von Cholin zu Glycin. DMGDH katalysiert die oxidative Demethylierung von N,N-Dimethylglycin zu Sarcosin (EC 1.5.8.4), welches in einer ähnlichen Reaktion durch die Sarcosin-Dehydrogenase zu Glycin metabolisiert wird (EC 1.5.8.3, siehe hierzu auch OMIM-Eintrag *605849, abgerufen am 4. September 2013, sowie Abbildung 6.18).

Abbildung 6.18.: Umwandlung von N,N-Dimethylglycin zu Glycin (EC 1.5.8.4 und 1.5.8.3)



Nicht eingezeichnet sind die als Elektronenüberträger an den Reaktionen beteiligten Flavoproteine.

1999 untersuchten Moolenaar *et al.* durch NMR-Spektroskopie Blutserum und Urin eines 38jährigen Probanden, der an einer DMGDH-Störung litt. Die Symptome dieser erblichen Stoffwechselkrankheit sind eine ab dem frühen Kindesalter auftretende Muskelschwäche sowie ausgeprägter Fischgeruch der Haut, der wohl durch das Ausschwitzen von Trimethylamin hervorgerufen wird [188]. Die Autoren der Studie ermittelten eine gegenüber gesunden Personen um den Faktor 20 höhere N,N-Dimethylglycin-Konzentration im Urin. Laut Moolenaar *et al.* hat reines N,N-Dimethylglycin ebenfalls einen fischartigen Geruch.

Im Jahr darauf zeigten Biznak *et al.*, dass der Proband homozygoter Träger des SNP rs121908331 im codierenden Bereich des DMGDH-Gens ist, welcher zu einem Aminosäureaustausch im Protein führt [189].

Im SHIP-0-Urin-Datensatz assoziiert rs248386 mit mehreren aus N,N-Dimethylglycin zu anderen Metaboliten gebildeten Ratios. Einen direkten Bezug zur Funktion des DMGDH-Locus hat die Assoziation zum N,N-Dimethylglycin/Glycin-Ratio

Tabelle 6.14.: Assoziationen zu rs248386 (DMGDH)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
N,N-Dimethylglycin / Glycolat	3.635	0,1819	$2,85 \times 10^{-18}$	$7,15 \times 10^9$
N,N-Dimethylglycin / Valin	2.993	0,1762	$4,17 \times 10^{-18}$	$4,88 \times 10^9$
N,N-Dimethylglycin / Glycin	3.695	0,1781	$1,18 \times 10^{-17}$	$1,72 \times 10^9$
N,N-Dimethylglycin / Taurin	3.100	0,2284	$1,88 \times 10^{-15}$	$1,09 \times 10^7$
N,N-Dimethylglycin / Histidin	3.581	0,1763	$9,71 \times 10^{-15}$	$2,10 \times 10^6$
N,N-Dimethylglycin / Tyrosin	3.143	0,1611	$1,13 \times 10^{-14}$	$1,81 \times 10^6$
N,N-Dimethylglycin / Pyroglutamat	3.600	0,1661	$2,78 \times 10^{-14}$	$7,34 \times 10^5$
N,N-Dimethylglycin / Threonin	3.657	0,1460	$8,90 \times 10^{-14}$	$2,29 \times 10^5$
N,N-Dimethylglycin	3.698	0,377	$2,04 \times 10^{-8}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

($p < 1,18 \times 10^{-17}$). Dieser Ratio entspricht dem Verhältnis von Substrat und Produkt der zweistufigen Umsetzung von N,N-Dimethylglycin zu Glycin (Abbildung 6.18). Die GWAS zeigt, dass bei Trägern des seltenen Allels tendenziell weniger N,N-Dimethylglycin zu Glycin metabolisiert wird und stattdessen vermehrt mit dem Harn ausgeschieden wird. Bezüglich der p-Werte wird diese genetische Assoziation leicht übertroffen von den Ratios aus N,N-Dimethylglycin zu Valin und N,N-Dimethylglycin zu Glycolat, wobei letzteres metabolisch eng mit Glycin verwandt ist (siehe Kapitel 6.2, Abbildung 6.8).

Die 2013 von Rhee *et al.* veröffentlichte GWAS mit Plasmametaboliten nennt ebenfalls eine signifikante Assoziation zwischen rs248386 und Dimethylglycin ($p < 1,18 \times 10^{-17}$, $\beta = 0,50$) [42]. Dies und die in diesem Kapitel aufgeführte Assoziation lässt vermuten, dass der SNP rs248386 in Verbindung mit einem oder mehreren genetischen Polymorphismen steht, welche die Funktion von DMGDH beeinträchtigen. Ob unter diesen genetischen Varianten der von Biznak *et al.* genannte SNP rs121908331 ist, kann aufgrund fehlender LD-Daten allerdings nicht überprüft werden, da dieser seltene SNP weder in HapMap (Version 27, Phase I,II,III, CEU Population) noch im *1000 Genomes Project* (Phase 1 integrated, Version 3, EUR Population) erfasst wurde [61, 72].

Ein Polymorphismus im SLC6A19-Locus verändert womöglich die Reabsorption von Histidin und Tyrosin

Der SNP rs7719875 liegt auf dem 5. Chromosom etwa 10 kb stromaufwärts des Gens für den SLC6A19-Transporter, der die Resorption neutraler Aminosäuren durch die apikale Membran von Epithelzellen der Niere und des Darms zur Aufgabe hat [190]. Mutationen im SLC6A19-Gen können zur Hartnup-Krankheit (OMIM-Eintrag #234500, abgerufen am 4. September 2013 [87]) führen, einer 1956 erstmalig beschriebenen, autosomal-rezessiv vererbten Störung des Transportes neutraler Aminosäuren in der Niere und im Verdauungstrakt [191–195]. Charakteristisch für diese Stoffwechselstörung sind unter anderem lichtbedingte Entzündungen der Haut, Verdauungsstörungen sowie Aminoazidurie, der erhöhten Ausscheidung von Aminosäuren mit dem Harn.

1969 untersuchten Halvorsen *et al.* die Ausscheidung verschiedener Aminosäuren bei Hartnup-Kranken. Sie beobachteten, dass die positiven Probanden im Vergleich zu den Kontroll-Probanden deutlich erhöhte Konzentrationen von Histidin ausschieden. Die Autoren schlossen daraus, dass bei der Hartnup-Krankheit die Resorption von Histidin gestört ist [196].

In der hier besprochenen GWAS gibt es signifikante Assoziationen zwischen rs7719875 und den Ratios von Histidin zu N,N-Dimethylglycin ($p < 4,8 \times 10^{-20}$), τ -Methylhistidin ($p < 9,3 \times 10^{-19}$), Ethanolamin ($p < 3,7 \times 10^{-18}$) und anderer Metaboliten (Tabelle 6.15). Auch Ratios von Tyrosin zu Metaboliten wie Ethanolamin, N,N-Dimethylglycin und Dimethylamin assoziieren signifikant mit diesem SNP. Dabei ist sowohl die Histidin- als auch die Tyrosin-Konzentration im Harn tendenziell umso niedriger, je häufiger das seltene Allel vorliegt.

Die Assoziation des SLC6A19-Locus zu Histidin und anderen Aminosäuren passt daher ebenso zur Funktion des Genprodukts wie zur früher bei Hartnup-Kranken beobachteten, durch Mutationen in diesem Locus hervorgerufenen Veränderungen bezüglich der Ausscheidung von Aminosäuren. Der in der Bevölkerung häufig präsente SNP rs7719875 (MAF in SHIP-0: 36,8 %) könnte sich bei Trägern des häufigeren Allels in einem metabolischen Phänotypen manifestieren, der einer milden Form der Hartnup-Krankheit gleicht. Dieser GIM ist somit ein gutes Beispiel für die bereits vor über 100 Jahren von Archibald Garrod, dem Entdecker der Alkaptonurie, aufgestellte

Vermutung, dass viele Stoffwechselkrankheiten lediglich extreme Ausprägungen der in der Bevölkerung beobachteten chemischen Variabilität sind [197].

Tabelle 6.15.: Assoziationen zu rs7719875 (SLC6A19)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Histidin / N,N-Dimethylglycin	3.592	-0,1424	$4,75 \times 10^{-20}$	$1,42 \times 10^{11}$
Histidin / τ -Methylhistidin	2.914	-0,1243	$9,26 \times 10^{-19}$	$7,28 \times 10^9$
Histidin / Ethanolamin	3.484	-0,1113	$3,70 \times 10^{-18}$	$1,82 \times 10^9$
Histidin / Glycolat	3.667	-0,1012	$6,04 \times 10^{-16}$	$1,12 \times 10^7$
Tyrosin / Ethanolamin	3.081	-0,0797	$1,12 \times 10^{-15}$	$3,02 \times 10^{11}$
Histidin / Dimethylamin	3.722	-0,1069	$5,75 \times 10^{-15}$	$1,17 \times 10^6$
Histidin / cis-Aconitat	3.514	-0,1143	$1,43 \times 10^{-14}$	$4,70 \times 10^5$
Tyrosin / N,N-Dimethylglycin	3.151	-0,1114	$2,19 \times 10^{-14}$	$1,55 \times 10^{10}$
Tyrosin / Dimethylamin	3.201	-0,0749	$4,03 \times 10^{-12}$	$8,38 \times 10^7$
Threonin / N,N-Dimethylglycin	3.669	-0,0917	$5,90 \times 10^{-11}$	$1,47 \times 10^8$
Tyrosin/ Valin	2.754	-0,0608	$8,76 \times 10^{-10}$	$3,86 \times 10^5$
Tyrosin / Methanol	3.171	-0,0862	$9,53 \times 10^{-10}$	$3,55 \times 10^5$
Tyrosin / τ -Methylhistidin	2.506	-0,0850	$1,50 \times 10^{-9}$	$2,25 \times 10^5$
Histidin	3.722	-0,0991	$6,74 \times 10^{-9}$	
Threonin / Ethanolamin	3.555	-0,0595	$3,46 \times 10^{-8}$	$3,51 \times 10^5$
Asparagin	1.492	-0,0746	$9,35 \times 10^{-6}$	
Tyrosin	3.201	-0,0536	$3,38 \times 10^{-4}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

SNP rs3846710 im SLC36A2-Locus assoziiert mit einer verringerten Glycin-Ausscheidung

Der SNP rs3846710 liegt auf dem 5. Chromosom im intronischen Bereich des SLC36A2-Gens. Aufgrund der starken Sequenzähnlichkeit des vom SLC36A2-Gens codierten Proteins zum homologen SLC36A2-Protein der Maus kann davon ausgegangen werden, dass das Genprodukt ein Transporterprotein mit hoher Affinität für kleine Aminosäuren wie Glycin, Alanin und Prolin ist [198].

2008 untersuchten Bröer *et al.* in einer Familienstudie fünf mutmaßliche Transporter für Iminosäuren und Glycin auf Assoziation mit Iminoglycinurie (IG) bzw. Hyperglycinurie (HG). Diese Transporter sind maßgeblich an der renal-tubulären Reabsorption von bestimmten Aminosäuren beteiligt. Die Autoren folgerten, dass Mutationen des SLC36A2-Gens die Hauptursache für IG und HG sind [118, 199]. Wie in Kapitel 6.1 besprochen gibt es eine Assoziation zwischen N,N-Dimethylglycin im Urin und einem genetischen Polymorphismus im SLC6A20-Gen, der zu einem Aminosäureaustausch (T199M) im Genprodukt führt. Diese Mutation allein verursacht laut Bröer *et al.* keine IG, kann aber in Verbindung mit Mutationen im SLC36A2-Transporter den Grad der IG erhöhen.

In der vorliegenden Studie konnte eine genomweit signifikante Assoziation zwischen rs3846710 und dem Glycin/Valin-Ratio identifiziert ($p < 2,1 \times 10^{-11}$) werden. Die leicht bessere Assoziation zum Glycin/Succinat-Ratio konnte hingegen im Urin der KORA-F4-Kohorte nicht repliziert werden. Die Assoziation mit Glycin passt dabei zur vermuteten Affinität des SLC36A2-Transporters. Allerdings ist rs3846710 nicht gekoppelt mit dem seltenen SNP rs77010315 ($r^2 < 0,1$; Datenbasis ist hier das *1000 Genomes Project* Phase 1, Release 3, EUR Population, da rs77010315 aufgrund der niedrigen MAF nicht in den HapMap-Daten Version 27, Phase I-III, CEU-Panel erfasst ist). Bröer *et al.* nennen diesen SNP als möglichen Verursacher für den Funktionsverlust von SLC36A2.

Desweiteren scheiden Träger des Effektallels von rs3846710 tendenziell weniger Glycin aus, was im Gegensatz zu erhöhter Glycin-Ausscheidung bei Personen mit Iminoglycinurie steht. Möglicherweise trägt der hier identifizierte SNP rs3846710 bzw. ein damit gekoppelter Polymorphismus stattdessen zu einer erhöhten Glycin-Reabsorption bei.

Tabelle 6.16.: Assoziationen zu rs3846710 (SLC36A2)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Glycin / Succinat	3.848	-0,1178	$7,18 \times 10^{-12}$	$1,43 \times 10^5$
Glycin / Valin	3.068	-0,1192	$2,11 \times 10^{-11}$	$4,86 \times 10^4$
Glycin	3.858	-0,1084	$1,02 \times 10^{-6}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

Eine Reihe von SNPs im FBXL20-Locus assoziieren mit Histidin und Tyrosin

In einem ca. 430 kb breiten Abschnitt des 17. Chromosoms zeigen eine Reihe von SNPs genomweit signifikante Assoziationen zu den Konzentrationen der aromatischen Aminosäuren Tyrosin und Histidin sowie zu Ratios, an denen diese Aminosäuren beteiligt sind. Der am stärksten assoziierende SNP auf diesem Locus ist rs2338755. Von den SNPs, die vergleichbar hohe Assoziationen wie rs2338755 zeigen, sind neun mit dem Lead-SNP im LD bei Korrelationskoeffizienten (r^2) größer 0,96. Insgesamt sind es 15 SNPs, die mit Histidin bei p-Werten zwischen $1,2 \times 10^{-16}$ und $5,3 \times 10^{-18}$ assoziieren. Von den genannten SNPs liegen zehn im intronischen Bereich, einer im 3'-untranslatierten Bereich und vier vor dem 5'-Ende des FBXL20-Gens.

FBXL20 steht für „*F-box and leucine-rich repeat protein 20*“. Dieses Gen kodiert mutmaßlich eine Substrat-Erkennungskomponente des SKP1-CUL1-FBox-Ubiquitin-Ligase-Komplexes (Uniprot-Eintrag Q96IG2, Stand 3. April 2013 [200]). Der SCF-Komplex sorgt für die Ubiquitinierung von Proteinen, die an der Steuerung des Zellzyklus, an der Signaltransduktion und an der Transkription beteiligt sind [201]. Durch Ubiquitinierung werden Proteine zum proteasomalen Abbau markiert [202].

Ein Zusammenhang zwischen dem SCF-Komplex und Histidin bzw. Tyrosin lässt sich anhand der Assoziationsergebnisse allerdings nur schwer herstellen. Möglicherweise ist die Ubiquitinierung ein Mechanismus zur Regulation eines Histidin- bzw. Tyrosin-Transporters, was letztendlich zu einer veränderten Ausscheidung der assoziierten Metaboliten führt. Andererseits könnten die zahlreichen weiteren Assoziationen des FBXL20-Locus zu einer Reihe proteinogener Aminosäuren (siehe Tabelle 6.17) darauf hinweisen, dass allgemein die Ubiquitinierung bestimmter Proteine und damit deren Abbau verändert wird.

In einer 2012 veröffentlichten Meta-Studie identifizierten und replizierten Chasman *et al.* eine genomweit signifikante Assoziation zwischen rs7208487, der im intronischen Bereich des FBXL20-Gens liegt, und der geschätzten glomerulären Filtrationsrate (eGFR), einem klinischen Marker für Nierenfunktion und chronische Niereinsuffizienz [203]. Der von Chasman *et al.* identifizierte SNP ist zwar nicht im Datensatz der hier vorgestellten Studie, er ist aber mäßig im LD ($r^2 = 0,43$) mit rs2338755 und weiteren ähnlich stark mit Histidin und Tyrosin im Urin assoziierten SNPs. Des weiteren ist rs7208487 maximal im LD ($r^2 = 1$) mit rs6503521 und fünf weiteren SNPs, die bei p-Werten kleiner $1,5 \times 10^{-13}$ im SHIP-Urin-Datensatz mit dem Histidin/ τ -Methylhistidin Ratio assoziieren. Watanabe *et al.* zeigten im Jahre 2008, dass eine verringerte Histidin-Konzentration im Blut bei Patienten mit chronischem Nierenversagen unter anderem mit einer erhöhten Sterblichkeitsrate assoziiert [204]. Somit wurden in der Vergangenheit sowohl der FBXL20-Locus als auch der am stärksten damit assoziierte Metabolit mit Nierenerkrankungen in Verbindung gebracht, was die hier vorgestellte Assoziation im Kontext dieser in der Bevölkerung zunehmend verbreiteten Krankheit interessant erscheinen lässt [205, 206].

Tabelle 6.17.: Assoziationen zu rs2338755 (FBXL20)

Metabolit oder Ratio	N	beta'	p-Wert	p-Gain
Histidin / τ -Methylhistidin	2.914	-0,1637	$3,08 \times 10^{-27}$	$1,04 \times 10^{10}$
Histidin / Harnstoff	3.720	-0,1568	$5,85 \times 10^{-23}$	$5,49 \times 10^5$
Tyrosin / Ethanolamin	3.081	-0,1009	$2,88 \times 10^{-20}$	$3,02 \times 10^9$
Tyrosin / Dimethylamin	3.201	-0,1064	$2,69 \times 10^{-19}$	$3,24 \times 10^8$
Histidin	3.723	-0,1547	$3,21 \times 10^{-17}$	
Tyrosin / τ -Methylhistidin	2.507	-0,1233	$9,70 \times 10^{-16}$	$8,99 \times 10^4$
Tyrosin	3.201	-0,1054	$8,71 \times 10^{-11}$	
Alanin	3.846	-0,0742	$2,45 \times 10^{-9}$	
Threonin	3.809	-0,0916	$1,52 \times 10^{-7}$	
Betain	3.824	-0,0815	$7,19 \times 10^{-6}$	
N,N-Dimethylglycin	3.709	-0,0787	$6,64 \times 10^{-5}$	
Glycin	3.855	-0,0743	$1,17 \times 10^{-4}$	
Asparagin	1.493	-0,0716	$1,37 \times 10^{-4}$	

Es gelten analog die Erläuterungen zu Tabelle 6.4.

6.3. Identifikation und MS-gestützte Interpretation genetisch beeinflusster NMR-Merkmale im Blutplasma

Zum Abschluss des Diskussionsteils werden die Ergebnisse einer GWAS besprochen, bei der als Phänotypen sowohl einzelne NMR-Bins als auch daraus errechnete paarweise NMR-Ratios untersucht wurden. Grundlage dieser Studie sind NMR-Spektren von Blutplasmaproben der KORA-F4-Kohorte. Einem hypothesenfreien Ansatz folgend wurde auf eine der Analyse vorausgehende Annotation der NMR-Spektren verzichtet. Insgesamt konnten sieben genetische Loci identifiziert werden, welche mit NMR-Signalen bzw. daraus abgeleiteten Merkmalen (NMR-Ratios) assoziieren (GCKR, CPS1, PYROXD2, FADS1, APOA1, LIPC, CETP). Bei vier dieser Loci war die jeweils stärkste Assoziation erst dann genomweit signifikant, als NMR-Ratios getestet wurden (Tabelle 5.6). Im Vergleich zu den genetischen Assoziationen mit einzelnen NMR-Bins sind die Assoziationen mit NMR-Ratios im Falle von CETP, GCKR, CPS1 und APOA1 mindestens um drei Zehnerpotenzen stärker. Bei den Loci FADS1 und PYROXD2 steigt die Assoziationsstärke gar um 60 bzw. 80 Größenordnungen. Es gibt verschiedene Gründe, warum genetische Assoziationen bei der Verwendung von Ratios stärker werden. Im Falle der Loci FADS1 und GCKR sind die NMR-Ratios wohl Stellvertreter für das relative Konzentrationsverhältnis zweier Metaboliten oder Metabolitenklassen. Hingegen wird am Beispiel von PYROXD2 gezeigt, dass NMR-Ratios äquivalent einer lokalen Baseline-Korrektur sein können, wobei überlappende Signale, die von anderen Metaboliten induziert werden, egalisiert werden.

Fünf der hier identifizierten Loci sind bekannte Lipid-Loci (LIPC, CETP, FADS1, GCKR, APOA1), und alle sieben Loci wurden bereits in früheren GWAS mit metabolischen Parametern identifiziert [30, 34, 35, 40, 79]. Zweifellos sind also genetische Varianten innerhalb dieser Loci in der Lage, wesentliche Veränderungen im metabolischen Phänotyp herbeizuführen. In Anbetracht der Tatsache, dass alle Loci in früheren Studien bereits intensiv auf ihre biologische und medizinische Bedeutung hin untersucht wurden, wird im Folgenden der Schwerpunkt auf diejenigen Aspekte gelegt, welche die Besonderheiten der assoziierten NMR-Merkmale hervorheben. Am Beispiel der Loci CPS1, FADS1 und GCKR wird gezeigt, wie sich durch Korrelation von NMR-Spektren mit Metabolitenkonzentrationen, welche durch komple-

mentäre Technologien wie MS bestimmt wurden, ein Wissenstransfer von sensitiven, hauptsächlich in Discovery-Studien eingesetzten Plattformen hin zu robusten NMR-Plattformen bewerkstelligen lässt. Weiterhin wird anhand einer Assoziation im PYROXD2-Locus demonstriert, wie sehr mGWAS mit NMR-Merkmalen von der Verwendung von Ratios profitieren können. Somit zeigt diese Studie, dass die Verwendung von NMR-Ratios eine einfache Methode ist, um genetisch validierte, bivariate Merkmale oder auch biologisch bzw. klinisch relevante Biomarker aus NMR-Rohdaten abzuleiten. Der hier vorgestellte Ansatz könnte auch auf die Verwendung von multivariaten Parametern erweitert werden.

Im Vergleich zu Assoziationsdaten, die auf den gleichen Proben, aber unterschiedlichen Methoden zur metabolischen Charakterisierung beruhen, und die zum Zeitpunkt der Durchführung dieser Studie verfügbar waren, konnte in der vorliegenden Studie in vier von sieben Fällen (PYROXD2, APOA1, LIPC, CETP) eine Steigerung der Assoziationsstärke festgestellt werden (Tabelle 5.8 und Abbildung 5.2). In den anderen drei Fällen (FADS1, GCKR sowie CPS1) liefern die MS-basierten Plattformen deutlich bessere Ergebnisse. Abermals wird hier die Komplementarität von MS und NMR deutlich. Ein Vorteil von NMR-Plattformen ist jedoch, dass sie aufgrund ihrer Robustheit im Prinzip besser für die klinische Diagnostik geeignet sind. Außerdem bedarf die hier vorgestellte Methode zur Ableitung bivariater Parameter aus NMR-Daten keiner aufwändigen Vorverarbeitung oder Annotierung der Spektren. Deshalb haben die hier identifizierten und genetisch validierten NMR-Ratios das Potenzial, in zukünftigen Studien als leicht zu bestimmende Marker verwendet zu werden.

Der CPS1-Locus assoziiert mit Glycin-korrelierten NMR-Peaks

Im CPS1-Locus assoziiert SNP rs2216405 signifikant mit den NMR-Signalen bei $\delta = 3,599$ ppm ($p = 4,46 \times 10^{-14}$; Abbildung 6.19a). An dieser Stelle korreliert das NMR-Signal mit den MS-bestimmten Glycinkonzentrationen bei einem Koeffizienten von $r_s^2 = 0,22$ (Abbildung 6.19b). Trotz dieser relativ schwachen Korrelation ist auf dem entsprechenden Korrelationsspektrum deutlich ein prominenter Peak sichtbar.

Ein experimentell gemessenes, aus der HMDB heruntergeladenes Reinsubstanzspektrum für Glycin zeigt einen scharf abgegrenzten Peak an $\delta = 3,54$ ppm (Abbildung 6.19c). Der Abstand zwischen dem Korrelationspeak aus dieser Studie und dem Peak aus der HMDB beträgt also ungefähr 0,06 ppm. Allerdings wurden die Spektren, die

für die hier besprochene Studie verwendet wurden, auf das Laktatsignal referenziert. Das HMDB-Spektrum wurde jedoch auf 4,4-Dimethyl-4-silapentan-1-schweflige Säure (DSS) ausgerichtet. Außerdem betrug der pH-Wert bei der HMDB-Probe 7,4, während er in den hier gemessenen Plasmaproben bei 7,0 lag. Diese experimentellen Unterschiede erklären den relativ großen Abstand zwischen dem mutmaßlichen Glycin-Peak, der in dieser Studie gefunden wurde, und dem Glycin-Peak aus der HMDB. In einem zusätzlichen Spiking-Experiment unter den gleichen experimentellen Bedingungen, wie sie auch bei den NMR-Messungen der für diese Studie untersuchten Proben vorzufinden waren, konnte schließlich nachgewiesen werden, dass die NMR-Signale bei $\delta = 3,599$ ppm in der Tat Glycin zuzuordnen sind (vergrößerte Spektralregion in Abbildung 6.19d).

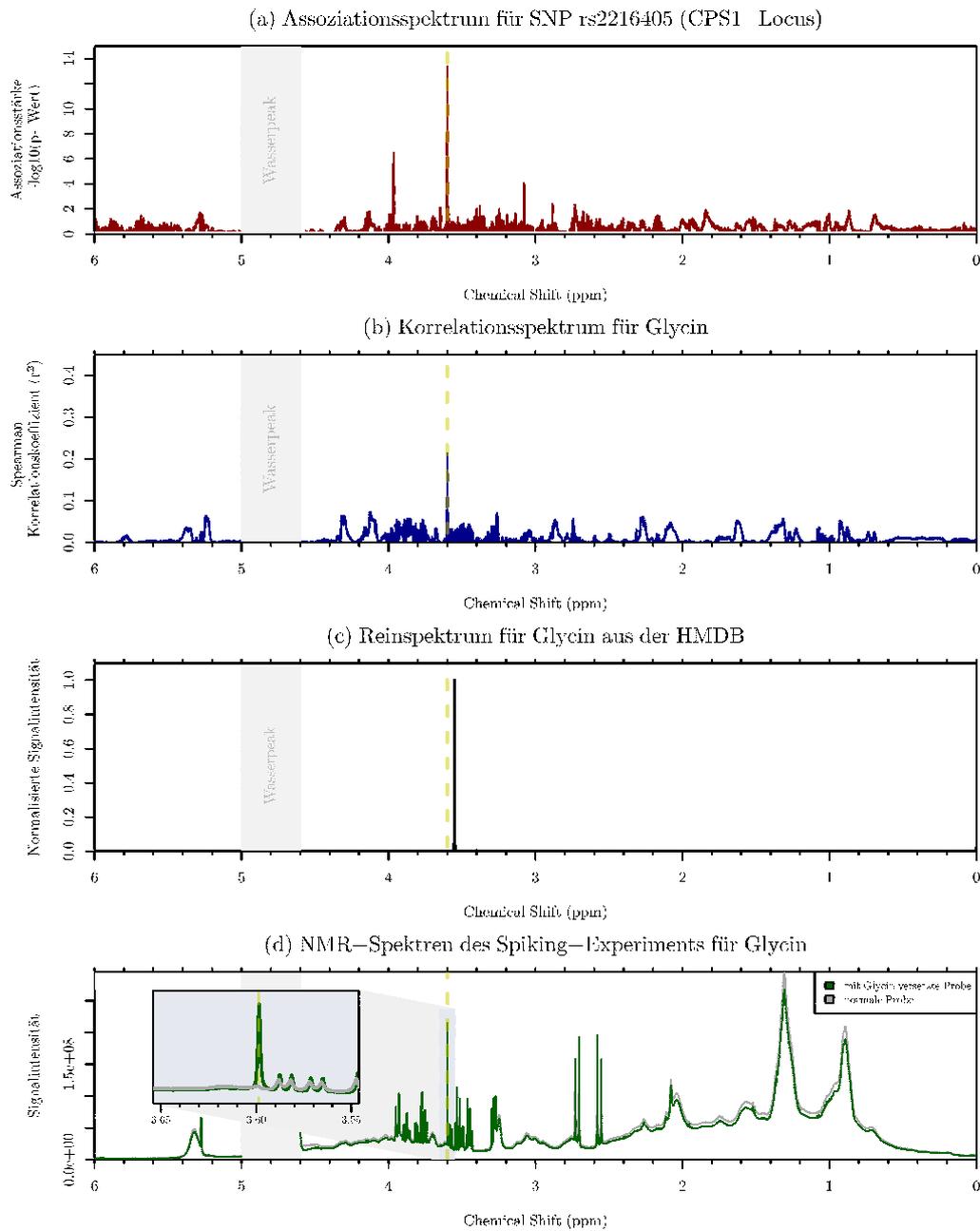
Die Carbamoylphosphat-Synthetase I (CPS1) steuert den ersten Schritt im Harnstoffzyklus. Klaus *et al.* identifizierten zwei Mutationen im CPS1-Gen, die für CPS-Mangel verantwortlich sind. Diese erbliche Stoffwechselkrankheit verursacht unter anderem Hyperammonämie [123]. In einer GWAS mit Metabolitenkonzentrationen, die durch eine quantitative MS-Plattform im KORA F4 Datensatz ermittelt wurden, haben Illig *et al.* 2010 eine signifikante Assoziation zwischen dem SNP rs2216405 und Glycinkonzentrationen bei einem p-Wert $2,59 \times 10^{-26}$ identifiziert. In einer Nachfolgestudie, bei der die Metaboliten im gleichen Datensatz mittels semiquantitativer MS gemessen wurden, assoziiert dieser SNP ebenfalls mit Glycin ($p = 1,6 \times 10^{-27}$) [35]. Glycin ist als Stoffwechselprodukt mit Carbamoylphosphat verwandt, welches durch CPS1 synthetisiert wird (siehe hierzu auch Abbildung 6.10 in Kapitel 6.2). Diese beiden Studien identifizierten also einen genetisch beeinflussten Metabotypen, der eine schwache Ausprägung des CPS-Mangels darstellen könnte.

In dem hier besprochenen Fall konnte also gezeigt werden, dass es grundsätzlich möglich ist, mit GWAS auf der Basis von nicht-annotierten NMR-Spektren vergleichbare Ergebnisse wie MS-basierten mGWAS zu erhalten, auch wenn die Assoziationsstärke bei Verwendung von NMR geringer ist.

NMR-Ratios sind repräsentativ für die Entsättigung bestimmter Fettsäuren (FADS1-Locus)

Das Assoziationsspektrum für SNP rs174547 (FADS1) zu einzelnen NMR-Bins (Abbildung 6.20a) zeigt mehrere breite Peaks, wobei die Assoziation bei $\delta = 2,801$ ppm

Abbildung 6.19.: Assoziations- und Korrelationspektren für rs2216405 (CPS1) und Glycin



Die rote Linie (a) zeigt die Stärke der Assoziation ($-\log_{10}(p)$) des SNPs rs2216405 zu den NMR-Signalen. Die stärkste Assoziation ($p < 4,5 \times 10^{-14}$) zeigen die Signale bei $\delta = 3,599$ ppm. Die blaue Linie (b) zeigt den quadrierten Koeffizienten der Spearman-Korrelation (r_s^2) zwischen MS-quantifiziertem Glycin und NMR-Signalen. Die höchste Korrelation ($r_s^2 = 0,21$) zeigt sich als scharfer Peak bei $\delta = 3,599$ ppm. Das der HMDB entnommene Reinsubstanzspektrum für Glycin (c) zeigt den Peak hingegen bei $\delta = 3,54$ ppm. Der unterste Plot (d) zeigt die NMR-Spektren zweier Proben, wobei einer Probe Glycin zugesetzt wurde (grüne Spektrallinie). Die Vergrößerung zeigt, dass die NMR-Signale bei $\delta = 3,599$ ppm tatsächlich von Glycin herrühren. Zur besseren Vergleichbarkeit ist in allen Spektren der Chemical Shift $\delta = 3,599$ ppm mit einer orangefarbenen Linie markiert.

mit einem p-Wert kleiner $3,96 \times 10^{-35}$ am stärksten ist. Interessanterweise nimmt die Assoziationsstärke um nahezu 60 Größenordnungen zu, wenn der NMR-Ratio zwischen den Signalen bei $\delta = 2,017$ ppm und $\delta = 2,801$ ppm als Phänotyp verwendet wird ($p < 1,1 \times 10^{-94}$).

Die Fettsäuren-Desaturase 1 (*fatty-acid desaturase 1*, FADS1) katalysiert die Entsättigungsreaktion von 8,11,14-Eicosatrienoyl-CoA zu Arachidonoyl-CoA (C20:3 zu C20:4; die Zahl nach dem Doppelpunkt gibt die Anzahl der Doppelbindungen in den Acylresten an). Genetische Varianten im FADS1-Locus wurden mit Morbus Crohn [44] und Risikofaktoren für Herz-Kreislauf-Erkrankungen [45] in Verbindung gebracht. Vor kurzem wurde in zwei voneinander unabhängigen Studien gezeigt, dass mit der Nahrung eingenommene langkettige, mehrfach ungesättigte Fettsäuren (wie beispielsweise C20:3 und C20:4) die Assoziation zwischen genetischen Varianten im FADS1-Locus und Blutlipidkonzentrationen modulieren und somit auch möglicherweise das Risiko für Herz-Kreislauf-Erkrankungen erhöhen [207, 208].

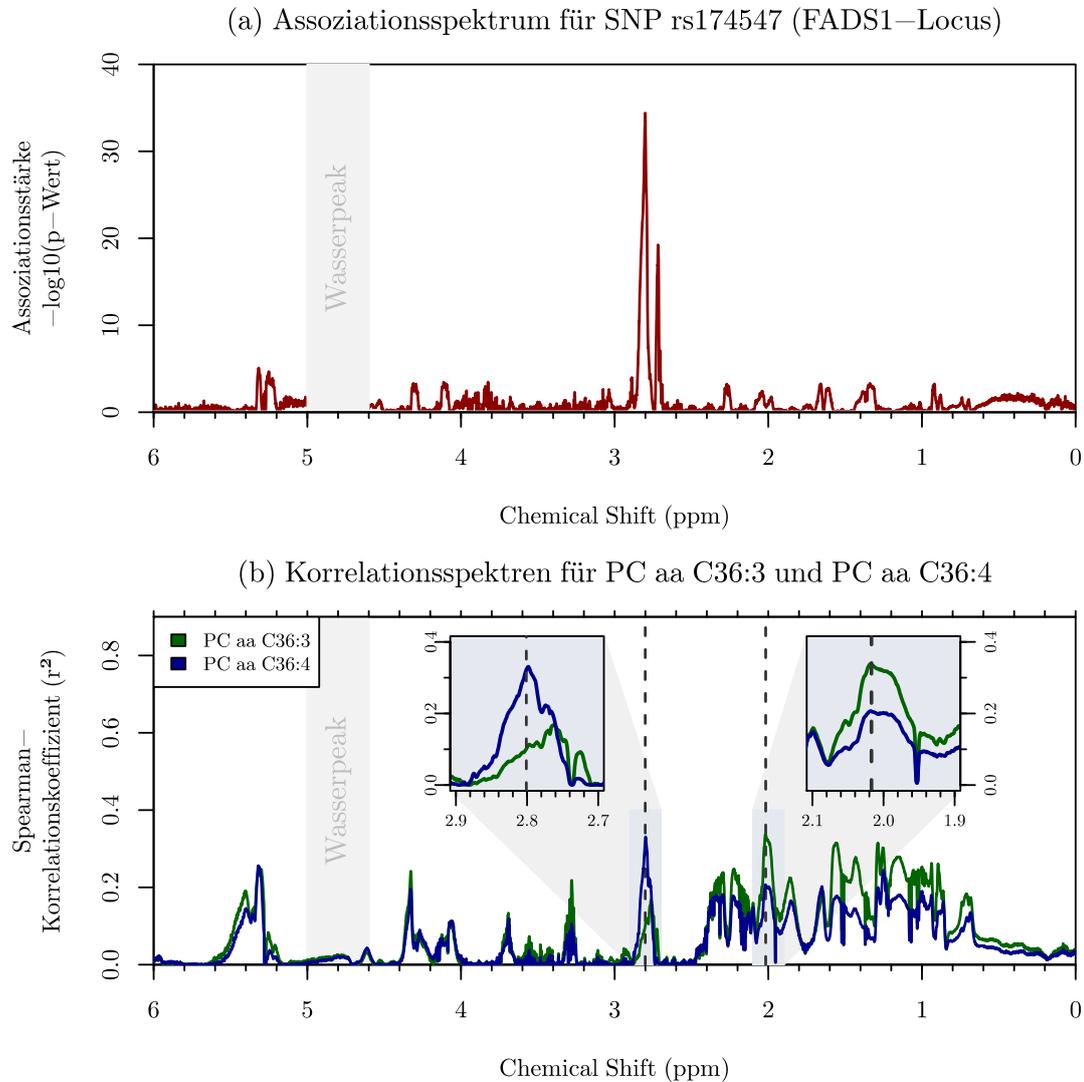
In GWAS im KORA-F4-Datensatz haben Gieger *et al.* [30] und Illig *et al.* [34] starke Assoziationen des SNPs rs174547, der im intronischen Bereich des FADS1 Gens liegt, zu etlichen Glycerophospholipiden identifiziert, von denen viele Lipidseitenketten mit mehrfach ungesättigten ω -3- und ω -6-Fettsäuren (*polyunsaturated fatty acids*, PUFAs), z.B. C20:3 und C20:4, enthalten. Die Autoren dieser Studien beobachteten einen außergewöhnlich hohen Anstieg der Assoziationsstärke, wenn paarweise Ratios zwischen PUFAs mit weniger als vier Doppelbindungen und PUFAs mit mindestens vier Doppelbindungen untersucht wurden [34]. Diese Beobachtung kann dadurch erklärt werden, dass paarweise aus Substrat und Produkt berechnete Quotienten einen Näherungswert der zugrunde liegenden enzymatischen Reaktionsrate liefern können (siehe Kapitel 3.2). Desweiteren können Ratios, die aus eng verwandten Metaboliten gebildet werden, Unterschiede in der Gesamtkonzentration von PUFAs einer Probe, die aufgrund der metabolischen Individualität der Testpersonen gegeben sind, ausgleichen und so einen normalisierenden Effekt haben [32].

In dem vorliegenden NMR-Datensatz ist anhand der Korrelationsspektren ersichtlich, dass verschiedene Spektralregionen in unterschiedlicher Stärke mit Glycerophospholipiden korrelieren, wobei dies abhängig vom Entsättigungsgrad der Lipid-Seitenketten (Abbildung 6.20b) zu sein scheint. So ist im Bereich zwischen $\delta = 0,0$ ppm und $\delta = 2,4$ ppm die Korrelation höher, wenn die Fettsäuren weniger als vier Doppelbindungen aufweisen (Vorläufer von FADS1-Substraten). Dagegen ist bei den Signalen

zwischen $\delta = 2,7$ ppm und $\delta = 2,9$ ppm die Korrelation stärker, wenn die Anzahl der Doppelbindungen in den Seitenketten gleich oder größer vier ist (Produkte bzw. Nachfolgeprodukte der FADS1-gesteuerten Reaktion).

Wie eingangs erwähnt ist das Assoziationssignal für rs174547 mit dem Ratio $\delta = 2,801 / \delta = 2,017$ ppm um knapp 60 Zehnerpotenzen stärker als mit dem NMR-Bin an $\delta = 2,801$ ppm. Das dem SNP zugehörige Assoziationsspektrum liefert keine offensichtliche Erklärung hierfür, da der Chemical Shift bei $\delta = 2,017$ ppm nur sehr schwach und weit oberhalb des Signifikanzniveaus mit rs174547 assoziiert. Der Blick auf die Korrelationsspektren hingegen offenbart, dass beide für die Berechnung dieses Ratios verwendeten Chemical Shifts stark mit drei- bzw. vierfach ungesättigten Glycerophospholipiden korrelieren. Es ist daher zu vermuten, dass der hier beschriebene NMR-Ratio das Verhältnis von dreifach- zu vierfach-ungesättigten Lipidseitenketten und daher auch die katalytische Umsetzungsrate des FADS1-Enzyms im Kontext der mit der Nahrung aufgenommenen PUFAs widerspiegelt.

Abbildung 6.20.: Assoziations- und Korrelationsspektren für rs174547 (FADS1) und PC aa C36:3 sowie PC aa C36:4



SNP rs174547 assoziiert am stärksten mit den NMR-Signalen bei $\delta = 2,801$ ppm ($p < 3,7 \times 10^{-35}$; Abbildung a). Wird dieser SNP auf Assoziation mit NMR-Ratios getestet, dann steigt die Assoziationsstärke um 59 Größenordnungen beim Ratio zwischen den NMR-Bins bei $\delta = 2,801$ ppm und $\delta = 2,017$ ppm ($p < 1,1 \times 10^{-94}$; siehe Tabelle 5.6). Ein ähnlich hoher Anstieg der Assoziationsstärke wurde auch in früheren GWAS beobachtet, wenn der FADS1-Locus mit Ratios aus Phospholipiden mit C20:3- und C20:4-Fettsäuren getestet wurde, beispielsweise PC aa 36:3 und PC aa 36:4 [30]. FADS1 codiert für die delta-5-Fettsäuren-Desaturase, deren Substrat bzw. Produkt C20:3- und C20:4-Fettsäuren sind. Der Korrelationsplot zwischen diesen Lipiden und den NMR-Signalen (Abbildung b) weist darauf hin, dass die Spektralregion um $\delta = 2,801$ ppm am stärksten mit Lipiden, welche C20:4-Fettsäuren enthalten, korreliert. Dagegen korreliert die Region bei $\delta = 2,801$ ppm stärker mit C20:3-Lipiden. Der NMR-Ratio zwischen $\delta = 2,801$ ppm und $\delta = 2,017$ ppm könnte daher das Konzentrationsverhältnis von drei- bzw. vierfach-ungesättigten Fettsäuren abbilden.

Ein Ratio aus der Triglyceride- und Glucoseregion des NMR-Spektrums ist ein potentieller pleiotroper Marker für Diabetesrisiko (GCKR-Locus)

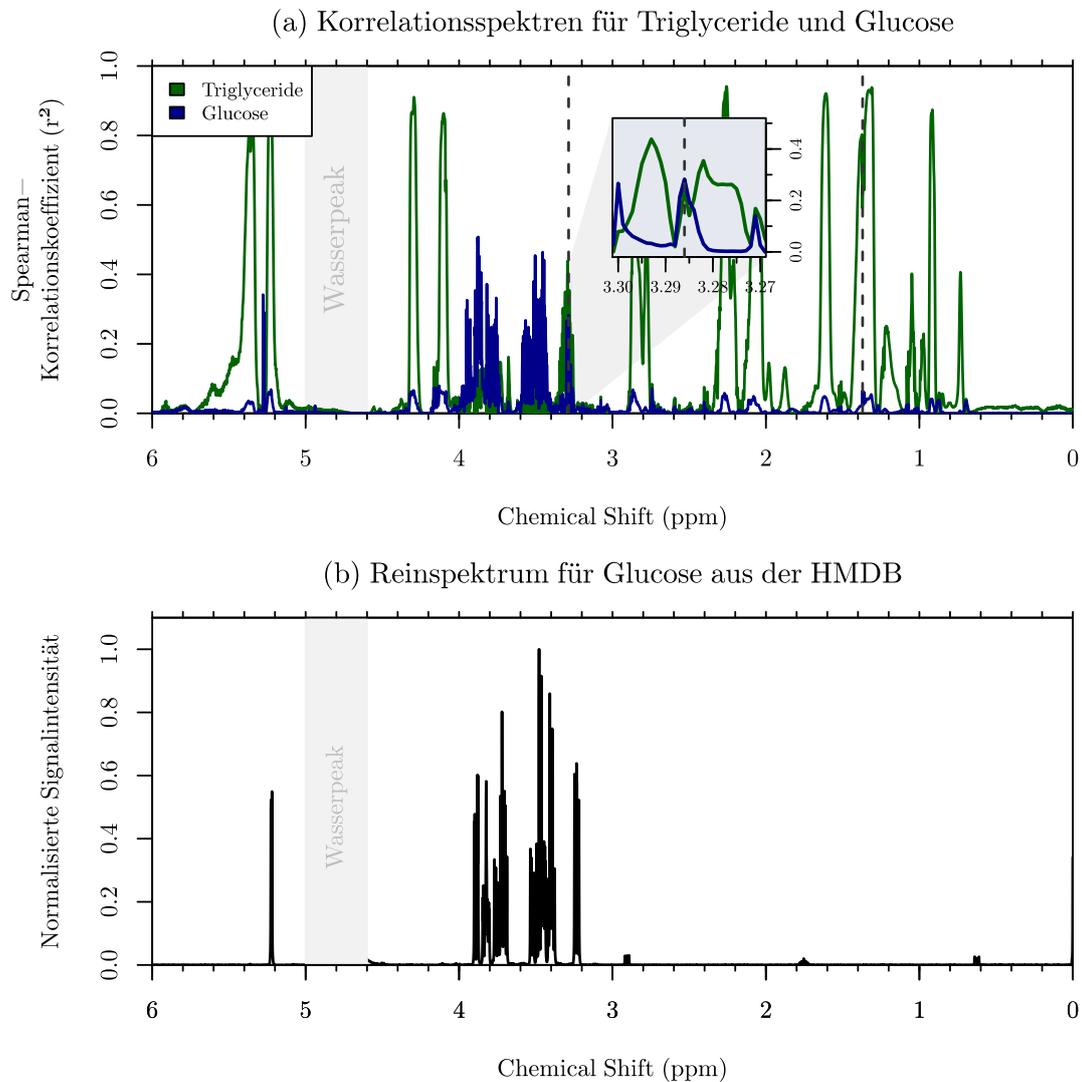
Der SNP rs780094 im GCKR-Locus assoziiert genomweit signifikant mit dem NMR-Ratio $\delta = 3,286 \text{ ppm}$ / $\delta = 1,370 \text{ ppm}$ ($p < 2,8 \times 10^{-15}$). Verglichen mit der stärksten Assoziation zu einzelnen NMR-Bins ($\delta = 1,370 \text{ ppm}$, $p < 1,2 \times 10^{-10}$) ist das Assoziationssignal im Falle des NMR-Ratios um ca. fünf Größenordnungen stärker.

Das Genprodukt des Glucokinase-Regulator-Gens (GCKR) übernimmt sowohl den Transport als auch die Regulation der Glucokinase, einem essentiellen Enzym des Glucosestoffwechsels. Der GCKR-Locus ist ein Risikolocus im Zusammenhang mit Typ-2-Diabetes [207]. In einer MS-basierten mGWAS im Blutserum identifizierten Suhre *et al.* eine Assoziation des GCKR-Locus' mit dem Mannose/Glucose-Ratio [35].

Wie aus Abbildung 6.21a ersichtlich korreliert der NMR-Bin bei $\delta = 3,286 \text{ ppm}$ mit einem Korrelationskoeffizienten von $r_s^2 = 0,28$ mit Glucose, und in der Tat wird dieser Teil des NMR-Spektrums von Glucose beeinflusst, wie das entsprechende Referenzspektrum der HMDB zeigt (Abbildung 6.21b). Die NMR-Intensitäten bei $\delta = 1,370 \text{ ppm}$ korrelieren hingegen stark ($r_s^2 = 0,61$) mit biochemisch gemessenen Triglyceriden (TG). Vergleicht man den hier vorgestellten NMR-Ratio mit dem Quotienten aus biochemisch bestimmten Glucose- bzw. TG-Konzentrationen, zeigt sich ebenfalls eine recht hohe Korrelation ($r_s^2 = 0,42$).

Aus klinischen Studien sowie früheren GWAS ist bekannt, dass GCKR ein pleiotroper Locus (beeinflusst mindestens zwei Phänotypen) für Diabetesrisiko ist: durch GCKR werden die TG- und Nüchtern-glucose-Konzentration invers reguliert. Das bedeutet, dass ein hoher Glucose-Spiegel mit einem niedrigen TG-Spiegel und umgekehrt einhergehen kann [25, 34, 209–211]. Der hier vorgestellte Ratio erlaubt es, simultan zwei Parameter zu bestimmen, die beide von großer Bedeutung bei der Bestimmung des Diabetesrisikos sind, und die unabhängig voneinander in früheren Studien mit dem GCKR Locus in Verbindung gebracht wurden. Der Ratio stellt somit möglicherweise einen Biomarker für pleiotrope biologische Prozesse mit Bezug zu Störungen in der Regulation des Glucosestoffwechsels dar.

Abbildung 6.21.: Assoziations- und Korrelationsspektren für rs780094 (GCKR) und Triglyceride sowie Glucose



In der hier vorgestellten Studie gibt es eine signifikante Assoziation ($p < 2,8 \times 10^{-15}$) von SNP rs780094 (GCKR-Locus) mit den NMR-Signalen bei $\delta = 3,286$ ppm und $\delta = 1,370$ ppm. Die Korrelationspektren (a) für biochemisch quantifizierte Glucose und Triglyceride (TG) zeigen, dass die NMR-Signale bei $\delta = 3,286$ ppm von Glucose beeinflusst werden (siehe auch das Glucose-Spektrum (b) aus der HMDB) und dass die Signale bei $\delta = 1,370$ ppm mit der TG-Konzentration korrespondieren.

Ratios aus unmittelbar benachbarten NMR-Bins können das Assoziationssignal verstärken (PYROXD2-Locus)

Die GWAS mit NMR-Bins deckt eine Assoziation zwischen SNP rs4488133 im PYROXD2-Locus (*pyridine nucleotide-disulfide oxidoreductase domain 2*), welcher für eine mutmaßliche Pyridinnukleotid-Disulfid-Oxidoreduktase kodiert, und den NMR-Intensitätswerten bei $\delta = 2,757$ ppm auf. Der p-Wert dieser Assoziation ist kleiner $7,3 \times 10^{-12}$, wodurch das genomweite Signifikanzniveau knapp verfehlt wird. In der GWAS mit NMR-Ratios findet sich hingegen ein hoch signifikantes Assoziationssignal mit dem Ratio $\delta = 2,757$ ppm / $\delta = 2,755$ ppm ($p < 2,9 \times 10^{-94}$).

In einer 2011 von Nicholson *et al.* veröffentlichten GWAS wurde erstmals die Assoziation des PYROXD2-Locus mit metabolischen Merkmalen, nämlich mit der mittels NMR gemessenen Konzentration von Dimethylamin (DMA) im Blutplasma [40], gezeigt. Im Gegensatz zu den anderen in diesem Kapitel besprochenen GIMs, für die im Vorfeld dieser Studie bereits MS-gestütztes Wissen vorhanden war, konnte der mit dem PYROXD2-Locus assoziierte Metabotyp bis zum Zeitpunkt der Durchführung der hier besprochenen GWAS nur durch die Verwendung von NMR identifiziert werden. Mittlerweile konnte aber auch dieser GIM auf einer MS-basierten Plattform im KORA-F4-Datensatz repliziert werden [79]. Allerdings ist noch ungeklärt, ob eine Verbindung zwischen diesem GIM und einem klinisch relevanten Phänotyp existiert.

Wie im Falle der FADS1-Assoziation ist auch beim PYROXD2-Locus ein außergewöhnlich hoher Anstieg der Assoziationsstärke von über 80 Größenordnungen zu beobachten, wenn NMR-Ratios anstatt einzelner NMR-Bins getestet werden. Bei der PYROXD2-Assoziation ist der NMR-Ratio jedoch aus zwei unmittelbar benachbarten Chemical Shifts errechnet ($\delta = 2,755$ ppm und $\delta = 2,757$ ppm).

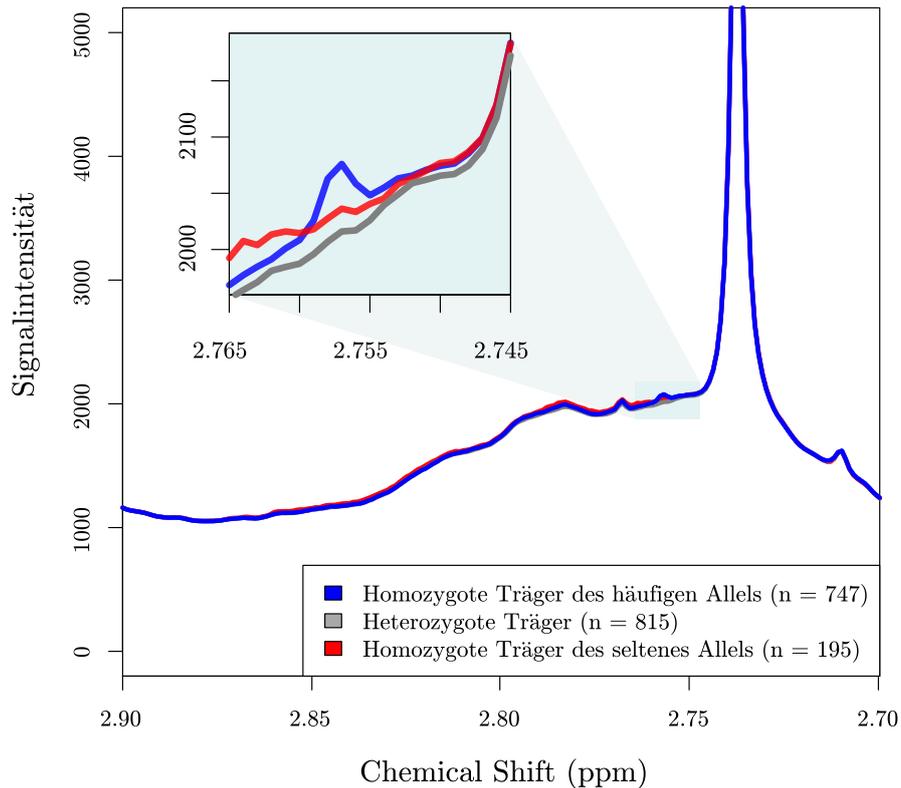
Abbildung 6.22 zeigt die nach Genotyp differenzierten NMR-Signale dieser Spektralregion; sie zeigt die Mittelwerte der Signalintensitäten der in drei Gruppen aufgeteilten Probanden. Die Unterteilung erfolgte je nachdem, ob die Testpersonen für SNP rs4488133 homozygote Träger des häufigen (in der Grafik mit einer blauen Signallinie gekennzeichnet) bzw. des seltenen (rot) Allels sind oder ob beide Allele gemischt (heterozygot; graue Linie) vorliegen. Das so differenzierte Spektrum offenbart, dass nur die Signale bei $\delta = 2,757$ ppm eine klare Abhängigkeit vom Genotyp aufweisen: ein Peak zeigt sich nur bei denjenigen Probanden, die homozygote Träger des häufigen

gen Allels sind. Dieser schwache Peak befindet sich auf der abfallenden Flanke eines darunterliegenden, sehr viel stärkeren NMR-Signals, welches mit Cholesterin korreliert. In dem hier vorgestellten Fall ist es wahrscheinlich, dass die Verwendung eines Ratios zwischen zwei benachbarten NMR-Bins, von denen einer innerhalb und der andere außerhalb eines Peaks liegt, einer lokalen Grundlinien-Korrektur (*baseline correction*) entspricht. Dadurch ist es möglich, schwache NMR-Signale aus dem Hintergrundrauschen hervorzuheben.

Die Korrelationsspektren ermöglichen keine Rückschlüsse darüber, welcher Metabolit das NMR-Signal an dieser Stelle beeinflussen könnte. In der oben erwähnten GWAS von Nicholson *et al.* assoziiert rs4488133 mit DMA [40]. Da laut HMDB der Hauptpeak für DMA in einem Bereich zwischen $\delta = 2,51$ und $\delta = 2,79$ ppm liegen kann, ist es durchaus möglich, dass es sich bei dem im vorliegenden Datensatz identifizierten Signal um DMA handelt. In der ebenfalls oben angesprochenen Studie von Krumisiek *et al.* assoziiert rs4488133 mit einem gemessenen, aber noch nicht biochemisch identifizierten Metaboliten (Metabolon-ID X-12092). Ein nachträglich errechnetes Korrelationsspektrum mit diesem unbekanntem Metaboliten zeigt allerdings für den NMR-Bin bei $\delta = 2,757$ ppm keine Korrelation ($r_s^2 = 0,04$).

Aufgrund recht guter Korrelation ($r_s^2 = 0,38$) des NMR-Ratios mit dem Ratio aus X-12092 und Cholesterin kann vermutet werden, dass der Metabolit, dessen Resonanzen die Signale bei $\delta = 2,757$ ppm beeinflussen, mit X-12092 zumindest metabolisch verwandt ist. Da der PYROXD2-Locus noch relativ uncharakterisiert ist, könnten der hier vorgestellte NMR-Ratio sowie die Ergebnisse anderer Studien und Metabolomik-Plattformen dazu dienen, dessen biologische Funktion genauer zu bestimmen.

Abbildung 6.22.: Intensität der NMR-Signale nach rs4488133-Genotyp (PYROXD2)



Der SNP rs4488133 assoziiert mit einem aus den benachbarten NMR-Bins bei $\delta = 2,757$ ppm und $\delta = 2,755$ ppm errechneten Ratio ($p < 2,9 \times 10^{-94}$), wobei der p-Gain gegenüber dem Testen des NMR-Bins bei $\delta = 2,757$ ppm mehr als 82 Größenordnungen beträgt. Die Abbildung zeigt die mittleren NMR-Signalintensitäten, aufgeteilt nach Genotyp des SNPs rs4488133. Der Peak bei $\delta = 2,757$ ppm ist dabei nur bei homozygoten Trägern des häufigen rs4488133-Allels sichtbar. Dieser Peak liegt auf einem viel größeren NMR-Signal, welches mit Cholesterin korreliert. Wahrscheinlich ist in diesem Fall die Verwendung dieses Ratios benachbarter Bins mit einer Art lokaler Grundlinien-Korrektur der Spektren gleichzusetzen, wodurch das schwache Assoziationssignal des GIMs aus dem Hintergrundrauschen hervortritt.

7. Literaturverzeichnis

- [1] J. Adamski und K. Suhre. Metabolomics platforms for genome wide association studies - linking the genome to the metabolome. *Curr Opin Biotechnol*, 24(1):39–47, Feb 2013.
- [2] D. S. Wishart. *Metabolome analysis: an introduction*, chapter 10. Wiley-Interscience series in mass spectrometry. Wiley-Interscience, Hoboken, N.J., 2007.
- [3] O. Fiehn. Metabolomics - the link between genotypes and phenotypes. *Plant Mol Biol*, 48(1-2):155–71, Jan 2002.
- [4] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res*, 35(Database issue):D521–6, Jan 2007.
- [5] N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T. L. Pedersen, S. R. Smith, F. Bamforth, R. Greiner, B. McManus, J. W. Newman, T. Goodfriend und D. S. Wishart. The human serum metabolome. *PLoS One*, 6(2):e16957, 2011.
- [6] S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring und D. S. Wishart. The human urine metabolome. *PLoS One*, 8(9):e73076, 2013.

- [7] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner und A. Scalbert. HMDB 3.0 - the human metabolome database in 2013. *Nucleic Acids Res*, 41(Database issue):D801–7, Jan 2013.
- [8] H. Tweeddale, L. Notley-McRobb und T. Ferenci. Effect of slow growth on metabolism of escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol*, 180(19):5109–16, Oct 1998.
- [9] J. K. Nicholson, J. C. Lindon und E. Holmes. 'metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological nmr spectroscopic data. *Xenobiotica*, 29(11):1181–9, Nov 1999.
- [10] O. Fiehn. Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. *Comp Funct Genomics*, 2(3):155–68, 2001.
- [11] N. Blow. Metabolomics: Biochemistry's new look. *Nature*, 455(7213):697–700, Oct 2008.
- [12] R. Kaddurah-Daouk und K. R. R. Krishnan. Metabolomics: a global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology*, 34(1):173–86, Jan 2009.
- [13] R. Kaddurah-Daouk, B. S. Kristal und R. M. Weinshilboum. Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol*, 48:653–83, 2008.
- [14] H. Wallaschofski. What will metabolomics studies mean to endocrinology? *J Endocrinol*, 215(1):1–2, Oct 2012.
- [15] K. Suhre. *Genetics meets metabolomics: from experiment to systems biology*, chapter 16. Springer, 2012.
- [16] W. S. Bush und J. H. Moore. Genome-wide association studies. *PLoS Comput Biol*, 8(12):e1002822, 2012.

- [17] J.-H. Park, M. H. Gail, C. R. Weinberg, R. J. Carroll, C. C. Chung, Z. Wang, S. J. Chanock, J. F. Fraumeni, Jr und N. Chatterjee. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A*, 108(44):18026–31, Nov 2011.
- [18] C. Gieger. *Genetics meets metabolomics: from experiment to systems biology*, chapter 4. Springer, 2012.
- [19] R. J. Klein, C. Zeiss, E. Y. Chew, J.-Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable und J. Hoh. Complement factor H polymorphism in age-related macular degeneration. *Science*, 308(5720):385–9, Apr 2005.
- [20] M. N. Weedon, G. Lettre, R. M. Freathy, C. M. Lindgren, B. F. Voight, J. R. B. Perry, K. S. Elliott, R. Hackett, C. Guiducci, B. Shields, E. Zeggini, H. Lango, V. Lyssenko, N. J. Timpson, N. P. Burtt, N. W. Rayner, R. Saxena, K. Ardlie, J. H. Tobias, A. R. Ness, S. M. Ring, C. N. A. Palmer, A. D. Morris, L. Peltonen, V. Salomaa, Diabetes Genetics Initiative, Wellcome Trust Case Control Consortium, G. Davey Smith et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet*, 39(10):1245–50, Oct 2007.
- [21] S. Sanna, A. U. Jackson, R. Nagaraja, C. J. Willer, W.-M. Chen, L. L. Bonnycastle, H. Shen, N. Timpson, G. Lettre, G. Usala, P. S. Chines, H. M. Stringham, L. J. Scott, M. Dei, S. Lai, G. Albai, L. Crisponi, S. Naitza, K. F. Doheny, E. W. Pugh, Y. Ben-Shlomo, S. Ebrahim, D. A. Lawlor, R. N. Bergman, R. M. Watanabe, M. Uda, J. Tuomilehto, J. Coresh et al. Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet*, 40(2):198–203, Feb 2008.
- [22] R. P. Stokowski, P. V. K. Pant, T. Dadd, A. Fereday, D. A. Hinds, C. Jarman, W. Filsell, R. S. Ginger, M. R. Green, F. J. van der Ouderaa und D. R. Cox. A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet*, 81(6):1119–32, Dec 2007.
- [23] P. Sulem, D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Ja-

- kobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, K. K. Aben, L. A. Kiemeny, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir und K. Stefansson. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet*, 39(12):1443–52, Dec 2007.
- [24] J. A. Todd, N. M. Walker, J. D. Cooper, D. J. Smyth, K. Downes, V. Plagnol, R. Bailey, S. Nejentsev, S. F. Field, F. Payne, C. E. Lowe, J. S. Szeszko, J. P. Hafler, L. Zeitels, J. H. M. Yang, A. Vella, S. Nutland, H. E. Stevens, H. Schuilenburg, G. Coleman, M. Maisuria, W. Meadows, L. J. Smink, B. Healy, O. S. Burren, A. A. C. Lam, N. R. Ovington, J. Allen et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*, 39(7):857–64, Jul 2007.
- [25] Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, R. Saxena, B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. W. de Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, M. J. Daly, T. E. Hughes, L. Groop, D. Altschuler, P. Almgren, J. C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Boström, B. Isomaa, G. Lettre, U. Lindblad, H. N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Råstam et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829):1331–6, Jun 2007.
- [26] R. Sladek, G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V. Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos und P. Froguel. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445(7130):881–5, Feb 2007.
- [27] N. J. Samani, J. Erdmann, A. S. Hall, C. Hengstenberg, M. Mangino, B. Mayer, R. J. Dixon, T. Meitinger, P. Braund, H.-E. Wichmann, J. H. Barrett, I. R. König, S. E. Stevens, S. Szymczak, D.-A. Tregouet, M. M. Iles, F. Pahlke, H. Pollard, W. Lieb, F. Cambien, M. Fischer, W. Ouwehand, S. Blankenberg, A. J. Balmforth, A. Baessler, S. G. Ball, T. M. Strom, I. Braenne et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*, 357(5):443–53, Aug 2007.

- [28] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins und T. A. Manolio. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, 106(23):9362–7, Jun 2009.
- [29] P. M. Visscher, M. A. Brown, M. I. McCarthy und J. Yang. Five years of GWAS discovery. *Am J Hum Genet*, 90(1):7–24, Jan 2012.
- [30] C. Gieger, L. Geistlinger, E. Altmaier, M. Hrabé de Angelis, F. Kronenberg, T. Meitinger, H.-W. Mewes, H.-E. Wichmann, K. M. Weinberger, J. Adamski, T. Illig und K. Suhre. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*, 4(11):e1000282, Nov 2008.
- [31] E. Altmaier, S. L. Ramsay, A. Graber, H.-W. Mewes, K. M. Weinberger und K. Suhre. Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication. *Endocrinology*, 149(7):3478–89, Jul 2008.
- [32] A.-K. Petersen, J. Krumsiek, B. Wägele, F. J. Theis, H.-E. Wichmann, C. Gieger und K. Suhre. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinformatics*, 13:120, 2012.
- [33] F. Kronenberg. *Genetics meets metabolomics: from experiment to systems biology*, chapter 15. Springer, 2012.
- [34] T. Illig, C. Gieger, G. Zhai, W. Römisch-Margl, R. Wang-Sattler, C. Prehn, E. Altmaier, G. Kastenmüller, B. S. Kato, H.-W. Mewes, T. Meitinger, M. H. de Angelis, F. Kronenberg, N. Soranzo, H.-E. Wichmann, T. D. Spector, J. Adamski und K. Suhre. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*, 42(2):137–41, Feb 2010.
- [35] K. Suhre, S.-Y. Shin, A.-K. Petersen, R. P. Mohny, D. Meredith, B. Wägele, E. Altmaier, CARDIoGRAM, P. Deloukas, J. Erdmann, E. Grundberg, C. J. Hammond, M. H. de Angelis, G. Kastenmüller, A. Köttgen, F. Kronenberg, M. Mangino, C. Meisinger, T. Meitinger, H.-W. Mewes, M. V. Milburn, C. Prehn, J. Raffler, J. S. Ried, W. Römisch-Margl, N. J. Samani, K. S. Small,

H.-E. Wichmann et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477(7362):54–60, Sep 2011.

- [36] Y. Kim, M. Go, C. Hu, C. Hong, Y. Kim, J. Lee, J. Hwang, J. Oh, D. Kim, N. Kim et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nature genetics*, 43(10):990–995, 2011.
- [37] C. Sabatti, S. K. Service, A.-L. Hartikainen, A. Pouta, S. Ripatti, J. Brodsky, C. G. Jones, N. A. Zaitlen, T. Varilo, M. Kaakinen, U. Sovio, A. Ruukonen, J. Laitinen, E. Jakkula, L. Coin, C. Hoggart, A. Collins, H. Turunen, S. Gabriel, P. Elliot, M. I. McCarthy, M. J. Daly, M.-R. Jarvelin, N. B. Freimer und L. Peltonen. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet*, 41(1):35–46, Jan 2009.
- [38] J. Kettunen, T. Tukiainen, A.-P. Sarin, A. Ortega-Alonso, E. Tikkanen, L.-P. Lyytikäinen, A. J. Kangas, P. Soininen, P. Würtz, K. Silander, D. M. Dick, R. J. Rose, M. J. Savolainen, J. Viikari, M. Kähönen, T. Lehtimäki, K. H. Pietiläinen, M. Inouye, M. I. McCarthy, A. Jula, J. Eriksson, O. T. Raitakari, V. Salomaa, J. Kaprio, M.-R. Jarvelin, L. Peltonen, M. Perola, N. B. Freimer et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*, 44(3):269–76, Mar 2012.
- [39] A. A. Hicks, P. P. Pramstaller, A. Johansson, V. Vitart, I. Rudan, P. Ugocsai, Y. Aulchenko, C. S. Franklin, G. Liebisch, J. Erdmann, I. Jonasson, I. V. Zorkoltseva, C. Pattaro, C. Hayward, A. Isaacs, C. Hengstenberg, S. Campbell, C. Gnewuch, A. C. W. Janssens, A. V. Kirichenko, I. R. König, F. Marroni, O. Polasek, A. Demirkan, I. Kolcic, C. Schwienbacher, W. Igl, Z. Biloglav et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet*, 5(10):e1000672, Oct 2009.
- [40] G. Nicholson, M. Rantalainen, J. V. Li, A. D. Maher, D. Malmodin, K. R. Ahmadi, J. H. Faber, A. Barrett, J. L. Min, N. W. Rayner, H. Toft, M. Krestyaninova, J. Viksna, S. G. Neogi, M.-E. Dumas, U. Sarkans, MolPAGE Consortium, P. Donnelly, T. Illig, J. Adamski, K. Suhre, M. Allen, K. T. Zondervan, T. D. Spector, J. K. Nicholson, J. C. Lindon, D. Baunsgaard, E. Holmes, M. I. McCarthy und C. C. Holmes. A genome-wide metabolic QTL analysis in Eu-

ropeans implicates two loci shaped by recent positive selection. *PLoS Genet*, 7(9):e1002270, Sep 2011.

- [41] T. Tanaka, J. Shen, G. R. Abecasis, A. Kisialiou, J. M. Ordovas, J. M. Guralnik, A. Singleton, S. Bandinelli, A. Cherubini, D. Arnett, M. Y. Tsai und L. Ferrucci. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI study. *PLoS Genet*, 5(1):e1000338, Jan 2009.
- [42] E. P. Rhee, J. E. Ho, M.-H. Chen, D. Shen, S. Cheng, M. G. Larson, A. Ghorbani, X. Shi, I. T. Helenius, C. J. O'Donnell, A. L. Souza, A. Deik, K. A. Pierce, K. Bullock, G. A. Walford, R. S. Vasan, J. C. Florez, C. Clish, J.-R. J. Yeh, T. J. Wang und R. E. Gerszten. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab*, 18(1):130–43, Jul 2013.
- [43] S.-Y. Shin, E. B. Fauman, A.-K. Petersen, J. Krumsiek, R. Santos, J. Huang, M. Arnold, I. Erte, V. Forgetta, T.-P. Yang, K. Walter, C. Menni, L. Chen, L. Vasquez, A. M. Valdes, C. L. Hyde, V. Wang, D. Ziemek, P. Roberts, L. Xi, E. Grundberg, T. M. T. H. E. R. M. Consortium, M. Waldenberger, J. B. Richards, R. P. Mohny, M. V. Milburn, S. L. John, J. Trimmer et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*, advance online publication, May 2014.
- [44] A. Franke, D. P. B. McGovern, J. C. Barrett, K. Wang, G. L. Radford-Smith, T. Ahmad, C. W. Lees, T. Balschun, J. Lee, R. Roberts, C. A. Anderson, J. C. Bis, S. Bumpstead, D. Ellinghaus, E. M. Festen, M. Georges, T. Green, T. Haritunians, L. Jostins, A. Latiano, C. G. Mathew, G. W. Montgomery, N. J. Prescott, S. Raychaudhuri, J. I. Rotter, P. Schumm, Y. Sharma, L. A. Simms et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*, 42(12):1118–25, Dec 2010.
- [45] S. Kathiresan, C. J. Willer, G. M. Peloso, S. Demissie, K. Musunuru, E. E. Schadt, L. Kaplan, D. Bennett, Y. Li, T. Tanaka, B. F. Voight, L. L. Bonnycastle, A. U. Jackson, G. Crawford, A. Surti, C. Guiducci, N. P. Burtt, S. Parish, R. Clarke, D. Zelenika, K. A. Kubalanza, M. A. Morken, L. J. Scott, H. M. Stringham, P. Galan, A. J. Swift, J. Kuusisto, R. N. Bergman et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*, 41(1):56–65, Jan 2009.

- [46] J. Dupuis, C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A. U. Jackson, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, A. L. Gloyn, C. M. Lindgren, R. Mägi, A. P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J. J. Hot-tenga, C. S. Franklin, P. Navarro, K. Song, A. Goel, J. R. B. Perry et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*, 42(2):105–16, Feb 2010.
- [47] A. Köttgen, C. Pattaro, C. A. Böger, C. Fuchsberger, M. Olden, N. L. Glazer, A. Parsa, X. Gao, Q. Yang, A. V. Smith, J. R. O’Connell, M. Li, H. Schmidt, T. Tanaka, A. Isaacs, S. Ketkar, S.-J. Hwang, A. D. Johnson, A. Dehghan, A. Teumer, G. Paré, E. J. Atkinson, T. Zeller, K. Lohman, M. C. Cornelis, N. M. Probst-Hensch, F. Kronenberg, A. Tönjes et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet*, 42(5):376–84, May 2010.
- [48] M. R. Viant, D. W. Bearden, J. G. Bundy, I. W. Burton, T. W. Collette, D. R. Ekman, V. Ezernieks, T. K. Karakach, C. Y. Lin, S. Rochfort, J. S. de Ropp, Q. Teng, R. S. Tjeerdema, J. A. Walter und H. Wu. International NMR-based environmental metabolomics intercomparison exercise. *Environ Sci Technol*, 43(1):219–25, Jan 2009.
- [49] J. L. Ward, J. M. Baker, S. J. Miller, C. Deborde, M. Maucourt, B. Biais, D. Rolin, A. Moing, S. Moco, J. Vervoort, A. Lommen, H. Schäfer, E. Humpfer und M. H. Beale. An inter-laboratory comparison demonstrates that [H]-NMR metabolite fingerprinting is a robust technique for collaborative plant metabolomic data collection. *Metabolomics*, 6(2):263–273, Jun 2010.
- [50] A. Demirkan, C. M. van Duijn, P. Ugocsai, A. Isaacs, P. P. Pramstaller, G. Liebisch, J. F. Wilson, Å. Johansson, I. Rudan, Y. S. Aulchenko, A. V. Kirichenko, A. C. J. W. Janssens, R. C. Jansen, C. Gnewuch, F. S. Domingues, C. Pattaro, S. H. Wild, I. Jonasson, O. Polasek, I. V. Zorkoltseva, A. Hofman, L. C. Karsen, M. Struchalin, J. Floyd, W. Igl, Z. Biloglav, L. Broer, A. Pfeufer et al. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet*, 8(2):e1002490, 2012.
- [51] R. Lakhtakia. The history of diabetes mellitus. *Sultan Qaboos Univ Med J*, 13(3):368–70, Aug 2013.

- [52] J. K. Nicholson und J. C. Lindon. Systems biology: Metabonomics. *Nature*, 455(7216):1054–6, Oct 2008.
- [53] K. Suhre, H. Wallaschofski, J. Raffler, N. Friedrich, R. Haring, K. Michael, C. Wasner, A. Krebs, F. Kronenberg, D. Chang, C. Meisinger, H.-E. Wichmann, W. Hoffmann, H. Völzke, U. Völker, A. Teumer, R. Biffar, T. Kocher, S. B. Felix, T. Illig, H. K. Kroemer, C. Gieger, W. Römisch-Margl und M. Nauck. A genome-wide association study of metabolic traits in human urine. *Nat Genet*, 43(6):565–9, Jun 2011.
- [54] J. Raffler, W. Römisch-Margl, A.-K. Petersen, P. Pagel, F. Blöchl, C. Hengstenberg, T. Illig, C. Meisinger, K. Stark, H.-E. Wichmann, J. Adamski, C. Gieger, G. Kastenmüller und K. Suhre. Identification and MS-assisted interpretation of genetically influenced NMR signals in human plasma. *Genome Med*, 5(2):13, Feb 2013.
- [55] H. Völzke. Study of Health in Pomerania (SHIP). Concept, design and selected results. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 55(6-7):790–4, Jun 2012.
- [56] H.-E. Wichmann, C. Gieger, T. Illig und MONICA/KORA Study Group. KORA-gen - resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, 67 Suppl 1:S26–30, Aug 2005.
- [57] R. Holle, M. Happich, H. Löwel, H. E. Wichmann und MONICA/KORA Study Group. KORA - a research platform for population based health research. *Gesundheitswesen*, 67 Suppl 1:S19–25, Aug 2005.
- [58] R. Hegar, A. Döring und A. Mielck. Einfluss des subjektiven Sozialstatus auf gesundheitliche Risiken und Gesundheitszustand - Ergebnisse der KORA-F4-Studie. *Das Gesundheitswesen*, 74(05):306–314, 2011.
- [59] O. Delaneau, J. Marchini und J.-F. Zagury. A linear complexity phasing method for thousands of genomes. *Nat Methods*, 9(2):179–81, Feb 2012.
- [60] B. N. Howie, P. Donnelly und J. Marchini. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, 5(6):e1000529, Jun 2009.

- [61] 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles und G. A. McVean. A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–73, Oct 2010.
- [62] A. M. Weljie, J. Newton, P. Mercier, E. Carlson und C. M. Slupsky. Targeted profiling: quantitative analysis of ¹H NMR metabolomics data. *Anal Chem*, 78(13):4430–42, Jul 2006.
- [63] M. Wyss und R. Kaddurah-Daouk. Creatine and creatinine metabolism. *Physiol Rev*, 80(3):1107–213, Jul 2000.
- [64] S. S. Waikar, V. S. Sabbiseti und J. V. Bonventre. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney Int*, 78(5):486–494, 09 2010.
- [65] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [66] M. Viant, C. Ludwig und U. Gunther. *1D and 2D NMR Spectroscopy: From Metabolic Fingerprinting to Profiling*. WJ Griffiths, Royal Society of Chemistry, Cambridge, UK, 2008.
- [67] A.-K. Petersen, K. Stark, M. D. Musameh, C. P. Nelson, W. Römisch-Margl, W. Kremer, J. Raffler, S. Krug, T. Skurk, M. J. Rist, H. Daniel, H. Hauner, J. Adamski, M. Tomaszewski, A. Döring, A. Peters, H.-E. Wichmann, B. M. Kaess, H. R. Kalbitzer, F. Huber, V. Pfahlert, N. J. Samani, F. Kronenberg, H. Dieplinger, T. Illig, C. Hengstenberg, K. Suhre, C. Gieger und G. Kastenmüller. Genetic associations with lipoprotein subfractions provide information on their biological nature. *Hum Mol Genet*, 21(6):1433–43, Mar 2012.
- [68] A. M. Evans, C. D. DeHaven, T. Barrett, M. Mitchell und E. Milgram. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*, 81(16):6656–67, Aug 2009.

- [69] K. W. Huber F., Kalbitzer H.R. Verfahren zur Bestimmung von Lipoproteinen in Körperflüssigkeiten und Messanordnung dafür. Patent DE 10 2004 026 903 B4, 2005.
- [70] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly und P. C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559–75, Sep 2007.
- [71] W. Gentsch, D. Girou, A. Kennedy, H. Lederer, J. Reetz, M. Riedel, A. Schott, A. Vanni, M. Vazquez und J. Wolfrat. DEISA - Distributed European Infrastructure for Supercomputing Applications. *Journal of Grid Computing*, 9(2):259–277, 2011.
- [72] International HapMap Consortium. The international HapMap project. *Nature*, 426(6968):789–96, Dec 2003.
- [73] A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O’Donnell und P. I. W. de Bakker. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, 24(24):2938–9, Dec 2008.
- [74] A. Moayyeri, C. J. Hammond, D. J. Hart und T. D. Spector. The UK adult twin registry (TwinsUK resource). *Twin Res Hum Genet*, 16(1):144–9, Feb 2013.
- [75] M. Kapushesky, I. Emam, E. Holloway, P. Kurnosov, A. Zorin, J. Malone, G. Rustici, E. Williams, H. Parkinson und A. Brazma. Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res*, 38(Database issue):D690–8, Jan 2010.
- [76] K. D. Pruitt, T. Tatusova, G. R. Brown und D. R. Maglott. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res*, 40(Database issue):D130–5, Jan 2012.
- [77] G. M. Clarke, C. A. Anderson, F. H. Pettersson, L. R. Cardon, A. P. Morris und K. T. Zondervan. Basic statistical analysis in genetic case-control studies. *Nat Protoc*, 6(2):121–33, Feb 2011.

- [78] M. Inouye und G. Abraham. Look, no hands! Spectral biomarkers from genetic association studies. *Genome Med*, 5(2):14, Feb 2013.
- [79] J. Krumsiek, K. Suhre, A. M. Evans, M. W. Mitchell, R. P. Mohney, M. V. Milburn, B. Wägele, W. Römisch-Margl, T. Illig, J. Adamski, C. Gieger, F. J. Theis und G. Kastenmüller. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet*, 8(10):e1003005, Oct 2012.
- [80] A. Ross, G. Schlotterbeck, F. Dieterle und H. Senn. *The Handbook of Metabonomics and Metabolomics*, chapter 3. Elsevier B.V., 2007.
- [81] H. R. Crumpler, C. E. Dent, H. Harris und R. G. Westall. beta-Aminoisobutyric acid (alpha-methyl-beta-alanine); a new amino-acid obtained from human urine. *Nature*, 167(4243):307–8, Feb 1951.
- [82] H. Harris. Family studies on the urinary excretion of beta-aminoisobutyric acid. *Ann Eugen*, 18(1):43–9, Jun 1953.
- [83] H. Brunschede, R. Hoffbauer und H. Goedde. Die Ausscheidung von β -Aminoisobuttersäure bei Gesunden und Kranken. *Klinische Wochenschrift*, 43:93–97, 1965.
- [84] C. Calchi-Novati, R. Cepellini, I. Biancho, E. Silvestroni und H. Harris. β -Aminoisobutyric acid excretion in urine. a family study in an italian population. *Annals of Eugenics*, 17(1):335–336, 1952.
- [85] J. Yanai, Y. Kakimoto, T. Tsujio und I. Sano. Genetic study of beta-aminoisobutyric acid excretion by Japanese. *Am J Hum Genet*, 21(2):115–32, Mar 1969.
- [86] Y. Kakimoto, K. Taniguchi und I. Sano. D-beta-aminoisobutyrate:pyruvate aminotransferase in mammalian liver and excretion of beta-aminoisobutyrate by man. *J Biol Chem*, 244(2):335–40, Jan 1969.
- [87] A. Hamosh, A. F. Scott, J. Amberger, C. Bocchini, D. Valle und V. A. McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*, 30(1):52–5, Jan 2002.

- [88] N. Tamaki, S. F. Sakata und K. Matsuda. Purification, properties, and sequencing of aminoisobutyrate aminotransferases from rat liver. *Methods Enzymol*, 324:376–89, 2000.
- [89] M. Kanehisa und S. Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, Jan 2000.
- [90] K. D. Pruitt, T. Tatusova und D. R. Maglott. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*, 33(Database issue):D501–4, Jan 2005.
- [91] J. Martens-Lobenhoffer, R. N. Rodionov und S. M. Bode-Böger. Probing AGXT2 enzyme activity in mouse tissue by applying stable isotope-labeled asymmetric dimethyl arginine as substrate. *J Mass Spectrom*, 47(12):1594–600, Dec 2012.
- [92] A. Kittel, R. Maas, J. König, M. Mieth, N. Weiss, N. Jarzebska, B. Hohenstein, J. Martens-Lobenhoffer, S. M. Bode-Böger und R. N. Rodionov. In vivo evidence that AGXT2 can regulate plasma levels of dimethylarginines in mice. *Biochem Biophys Res Commun*, 430(1):84–9, Jan 2013.
- [93] R. N. Rodionov, D. J. Murry, S. F. Vaulman, J. W. Stevens und S. R. Lentz. Human alanine-glyoxylate aminotransferase 2 lowers asymmetric dimethylarginine and protects from inhibition of nitric oxide production. *J Biol Chem*, 285(8):5385–91, Feb 2010.
- [94] M. Kayrak, A. Bacaksiz, M. A. Vatankulu, S. S. Ayhan, A. Taner, A. Unlü, M. Yazici und M. S. Ulgen. Association between exaggerated blood pressure response to exercise and serum asymmetric dimethylarginine levels. *Circ J*, 74(6):1135–41, Jun 2010.
- [95] W. H. Wilson Tang, W. Tong, K. Shrestha, Z. Wang, B. S. Levison, B. Del-fraino, B. Hu, R. W. Troughton, A. L. Klein und S. L. Hazen. Differential effects of arginine methylation on diastolic dysfunction and disease progression in patients with chronic systolic heart failure. *Eur Heart J*, 29(20):2506–13, Oct 2008.

- [96] D. Fliser, F. Kronenberg, J. T. Kielstein, C. Morath, S. M. Bode-Böger, H. Haller und E. Ritz. Asymmetric dimethylarginine and progression of chronic kidney disease: the mild to moderate kidney disease study. *J Am Soc Nephrol*, 16(8):2456–61, Aug 2005.
- [97] P. Ravani, G. Tripepi, F. Malberti, S. Testa, F. Mallamaci und C. Zoccali. Asymmetrical dimethylarginine predicts progression to dialysis and death in patients with chronic kidney disease: a competing risks modeling approach. *J Am Soc Nephrol*, 16(8):2449–55, Aug 2005.
- [98] J. Leiper, M. Nandi, B. Torondel, J. Murray-Rust, M. Malaki, B. O’Hara, S. Rossiter, S. Anthony, M. Madhani, D. Selwood, C. Smith, B. Wojciak-Stothard, A. Rudiger, R. Stidwill, N. Q. McDonald und P. Vallance. Disruption of methylarginine metabolism impairs vascular homeostasis. *Nat Med*, 13(2):198–203, Feb 2007.
- [99] B. Caplin, Z. Wang, A. Slaviero, J. Tomlinson, L. Dowsett, M. Delahaye, A. Salama, International Consortium for Blood Pressure Genome-Wide Association Studies, D. C. Wheeler und J. Leiper. Alanine-glyoxylate aminotransferase-2 metabolizes endogenous methylarginines, regulates NO, and controls blood pressure. *Arterioscler Thromb Vasc Biol*, 32(12):2892–900, Dec 2012.
- [100] R. Wolf. *Arylamin-N-Acetyltransferase 2 - genetische Polymorphismen als Suszeptibilitätsfaktoren für das Mammakarzinom?* PhD thesis, Humboldt-Universität zu Berlin, Medizinische Fakultät - Universitätsklinikum Charité, 2004.
- [101] S. Sanderson, G. Salanti und J. Higgins. Joint effects of the N-acetyltransferase 1 and 2 (NAT1 and NAT2) genes and smoking on bladder carcinogenesis: a literature-based systematic huge review and evidence synthesis. *American journal of epidemiology*, 166(7):741–751, 2007.
- [102] J. F. Deeken, T. Cormier, D. K. Price, T. M. Sissung, S. M. Steinberg, K. Tran, D. J. Liewehr, W. L. Dahut, X. Miao und W. D. Figg. A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. *Pharmacogenomics J*, 10(3):191–9, Jun 2010.

- [103] T. M. Teslovich, K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, C. T. Johansen, S. W. Fouchier, A. Isaacs, G. M. Peloso, M. Barbalić, S. L. Ricketts, J. C. Bis, Y. S. Aulchenko, G. Thorleifsson, M. F. Feitosa, J. Chambers, M. Orho-Melander, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–13, Aug 2010.
- [104] D. W. Hein, M. A. Doll, A. J. Fretland, M. A. Leff, S. J. Webb, G. H. Xiao, U. S. Devanaboyina, N. A. Nangju und Y. Feng. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev*, 9(1):29–42, Jan 2000.
- [105] A. Sabbagh, P. Darlu, B. Crouau-Roy und E. S. Poloni. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. *PLoS One*, 6(4):e18507, 2011.
- [106] D. W. Hein. N-acetyltransferase 2 genetic polymorphism: effects of carcinogen and haplotype on urinary bladder cancer risk. *Oncogene*, 25(11):1649–58, Mar 2006.
- [107] J. D. Potter, J. Bigler, L. Fosdick, R. M. Bostick, E. Kampman, C. Chen, T. A. Louis und P. Grambsch. Colorectal adenomatous and hyperplastic polyps: smoking and N-acetyltransferase 2 polymorphisms. *Cancer Epidemiol Biomarkers Prev*, 8(1):69–75, Jan 1999.
- [108] N. Soranzo, T. D. Spector, M. Mangino, B. Kühnel, A. Rendon, A. Teumer, C. Willenborg, B. Wright, L. Chen, M. Li, P. Salo, B. F. Voight, P. Burns, R. A. Laskowski, Y. Xue, S. Menzel, D. Altshuler, J. R. Bradley, S. Bumpstead, M.-S. Burnett, J. Devaney, A. Döring, R. Elosua, S. E. Epstein, W. Erber, M. Falchi, S. F. Garner, M. J. R. Ghorri et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet*, 41(11):1182–90, Nov 2009.
- [109] C. Meisinger, H. Prokisch, C. Gieger, N. Soranzo, D. Mehta, D. Rosskopf, P. Lichtner, N. Klopp, J. Stephens, N. A. Watkins, P. Deloukas, A. Greinacher, W. Koenig, M. Nauck, C. Rimbach, H. Völzke, A. Peters, T. Illig, W. H. Ouwehand, T. Meitinger, H.-E. Wichmann und A. Döring. A genome-wide

- association study identifies three loci associated with mean platelet volume. *Am J Hum Genet*, 84(1):66–71, Jan 2009.
- [110] K. Tomoeda, H. Awata, T. Matsuura, I. Matsuda, E. Ploechl, T. Milovac, A. Boneh, C. R. Scott, D. M. Danks und F. Endo. Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. *Mol Genet Metab*, 71(3):506–10, Nov 2000.
- [111] C. R. Scott. The genetic tyrosinemias. *Am J Med Genet C Semin Med Genet*, 142C(2):121–6, May 2006.
- [112] M. Barbosa, A. Lopes, C. Mota, E. Martins, J. Oliveira, S. Alves, P. De Bonis, M. d. C. Mota, C. Dias, P. Rodrigues-Santos, A. M. Fortuna, D. Quelhas, L. Lacerda, L. Bisceglia und M. L. Cardoso. Clinical, biochemical and molecular characterization of cystinuria in a cohort of 12 patients. *Clin Genet*, 81(1):47–55, Jan 2012.
- [113] A. Mattoo und D. S. Goldfarb. Cystinuria. *Semin Nephrol*, 28(2):181–91, Mar 2008.
- [114] A. P. Evan, F. L. Coe, J. E. Lingeman, Y. Shao, B. R. Matlaga, S. C. Kim, S. B. Bledsoe, A. J. Sommer, M. Grynepas, C. L. Phillips und E. M. Worcester. Renal crystal deposits and histopathology in patients with cystine stones. *Kidney Int*, 69(12):2227–35, Jun 2006.
- [115] J. C. Chambers, W. Zhang, G. M. Lord, P. van der Harst, D. A. Lawlor, J. S. Sehmi, D. P. Gale, M. N. Wass, K. R. Ahmadi, S. J. L. Bakker, J. Beckmann, H. J. G. Bilo, M. Bochud, M. J. Brown, M. J. Caulfield, J. M. C. Connell, H. T. Cook, I. Cotlarciuc, G. Davey Smith, R. de Silva, G. Deng, O. Devuyst, L. D. Dikkeschei, N. Dimkovic, M. Dockrell, A. Dominiczak, S. Ebrahim, T. Eggermann et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*, 42(5):373–5, May 2010.
- [116] M. A. Font, L. Feliubadaló, X. Estivill, V. Nunes, E. Golomb, Y. Kreiss, E. Pras, L. Bisceglia, A. P. d’Adamo, L. Zelante, P. Gasparini, M. T. Bassi, A. L. George, Jr, M. Manzoni, M. Riboni, A. Ballabio, G. Borsani, N. Reig, E. Fernández, A. Zorzano, J. Bertran, M. Palacín und International Cystinuria Consortium. Functional analysis of mutations in SLC7A9, and genotype-

- phenotype correlation in non-type I cystinuria. *Hum Mol Genet*, 10(4):305–16, Feb 2001.
- [117] H. Takanaga, B. Mackenzie, Y. Suzuki und M. A. Hediger. Identification of mammalian proline transporter SIT1 (SLC6A20) with characteristics of classical system imino. *J Biol Chem*, 280(10):8974–84, Mar 2005.
- [118] S. Bröer, C. G. Bailey, S. Kowalczyk, C. Ng, J. M. Vanslambrouck, H. Rodgers, C. Auray-Blais, J. A. Cavanaugh, A. Bröer und J. E. J. Rasko. Iminoglycinuria and hyperglycinuria are discrete human phenotypes resulting from complex mutations in proline and glycine transporters. *J Clin Invest*, 118(12):3881–92, Dec 2008.
- [119] M. Kolz, T. Johnson, S. Sanna, A. Teumer, V. Vitart, M. Perola, M. Mangino, E. Albrecht, C. Wallace, M. Farrall, A. Johansson, D. R. Nyholt, Y. Aulchenko, J. S. Beckmann, S. Bergmann, M. Bochud, M. Brown, H. Campbell, EUROSPAN Consortium, J. Connell, A. Dominiczak, G. Homuth, C. Lamina, M. I. McCarthy, ENGAGE Consortium, T. Meitinger, V. Mooser, P. Munroe et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet*, 5(6):e1000504, Jun 2009.
- [120] A. Köttgen, E. Albrecht, A. Teumer, V. Vitart, J. Krumsiek, C. Hundertmark, G. Pistis, D. Ruggiero, C. M. O’Seaghdha, T. Haller, Q. Yang, T. Tanaka, A. D. Johnson, Z. Kutalik, A. V. Smith, J. Shi, M. Struchalin, R. P. S. Middelberg, M. J. Brown, A. L. Gaffo, N. Pirastu, G. Li, C. Hayward, T. Zemunik, J. Huffman, L. Yengo, J. H. Zhao, A. Demirkan et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet*, 45(2):145–54, Feb 2013.
- [121] D. L. Pearson, S. Dawling, W. F. Walsh, J. L. Haines, B. W. Christman, A. Bazyk, N. Scott und M. L. Summar. Neonatal pulmonary hypertension: urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. *New England Journal of Medicine*, 344(24):1832–1838, 2001.
- [122] R. M. Moonen, A. D. Paulussen, N. Y. Souren, A. G. Kessels, M. E. Rubio-Gozalbo und E. Villamor. Carbamoyl phosphate synthetase polymorphisms

- as a risk factor for necrotizing enterocolitis. *Pediatric research*, 62(2):188–190, 2007.
- [123] V. Klaus, T. Vermeulen, B. Minassian, N. Israelian, K. Engel, A. M. Lund, K. Roebrock, E. Christensen und J. Häberle. Highly variable clinical phenotype of carbamylphosphate synthetase 1 deficiency in one family: an effect of allelic variation in gene expression? *Clin Genet*, 76(3):263–9, Sep 2009.
- [124] K. Mittelstrass, J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis, T. Meitinger, F. Kronenberg, S. Weidinger, H. E. Wichmann, K. Suhre, R. Wang-Sattler, J. Adamski und T. Illig. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet*, 7(8):e1002215, Aug 2011.
- [125] R. M. J. Moonen, I. Reyes, G. Cavallaro, G. González-Luis, J. A. Bakker und E. Villamor. The T1405N carbamoyl phosphate synthetase polymorphism does not affect plasma arginine concentrations in preterm infants. *PLoS One*, 5(5):e10792, 2010.
- [126] M. L. Summar, J. V. Gainer, M. Pretorius, H. Malave, S. Harris, L. D. Hall, A. Weisberg, D. E. Vaughan, B. W. Christman und N. J. Brown. Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function. *Hypertension*, 43(2):186–91, Feb 2004.
- [127] G. Kikuchi. The glycine cleavage system: composition, reaction mechanism, and physiological significance. *Mol Cell Biochem*, 1(2):169–87, Jun 1973.
- [128] G. Kikuchi, Y. Motokawa, T. Yoshida und K. Hiraga. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc Jpn Acad Ser B Phys Biol Sci*, 84(7):246–63, 2008.
- [129] R. Wang-Sattler, Z. Yu, C. Herder, A. C. Messias, A. Floegel, Y. He, K. Heim, M. Campillos, C. Holzapfel, B. Thorand, H. Grallert, T. Xu, E. Bader, C. Huth, K. Mittelstrass, A. Döring, C. Meisinger, C. Gieger, C. Prehn, W. Roemisch-Margl, M. Carstensen, L. Xie, H. Yamanaka-Okumura, G. Xing, U. Ceglarek, J. Thiery, G. Giani, H. Lickert et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol*, 8:615, 2012.

- [130] O. Fiehn, W. T. Garvey, J. W. Newman, K. H. Lok, C. L. Hoppel und S. H. Adams. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS One*, 5(12), 2010.
- [131] A. Floegel, N. Stefan, Z. Yu, K. Mühlenbruch, D. Drogan, H. G. Joost, A. Fritsche, H. U. Häring, M. Hrabě de Angelis, A. Peters, M. Roden, C. Prehn, R. Wang-Sattler, T. Illig, M. B. Schulze, J. Adamski, H. Boeing und T. Pischon. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes*, 62(2):639–648, Feb 2013.
- [132] W. Xie, A. R. Wood, V. Lyssenko, M. N. Weedon, J. W. Knowles, S. Alkayyali, T. L. Assimes, T. Quertermous, F. Abbasi, J. Paananen, H. Häring, T. Hansen, O. Pedersen, U. Smith, M. Laakso, J. M. Dekker, J. J. Nolan, L. Groop, E. Ferrannini, K. P. Adam, W. E. Gall, T. M. Frayling und M. Walker. Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. *Diabetes*, Feb 2013.
- [133] N. Yaneva, J. Schuster, F. Schäfer, V. Lede, D. Przybylski, T. Paproth, H. Harms, R. H. Müller und T. Rohwerder. Bacterial acyl-CoA mutase specifically catalyzes coenzyme B12-dependent isomerization of 2-hydroxyisobutyryl-CoA and (S)-3-hydroxybutyryl-CoA. *J Biol Chem*, 287(19):15502–15511, May 2012.
- [134] T. Rohwerder, U. Breuer, D. Benndorf, U. Lechner und R. H. Müller. The alkyl tert-butyl ether intermediate 2-hydroxyisobutyrate is degraded via a novel cobalamin-dependent mutase pathway. *Appl Environ Microbiol*, 72(6):4128–4135, Jun 2006.
- [135] A. Amberg, E. Rosner und W. Dekant. Biotransformation and kinetics of excretion of methyl-tert-butyl ether in rats and humans. *Toxicol Sci*, 51(1):1–8, Sep 1999.
- [136] J. M. Benson, E. B. Barr und J. R. Krone. MTBE inhaled alone and in combination with gasoline vapor: uptake, distribution, metabolism, and excretion in rats. *Res Rep Health Eff Inst*, (102):73–94, May 2001.

- [137] J. M. Benson, B. M. Tibbetts und E. B. Barr. The uptake, distribution, metabolism, and excretion of methyl tertiary-butyl ether inhaled alone and in combination with gasoline vapor. *J Toxicol Environ Health A*, 66(11):1029–1052, Jun 2003.
- [138] D. McGregor. Ethyl tertiary-butyl ether: a toxicological review. *Crit Rev Toxicol*, 37(4):287–312, May 2007.
- [139] A. Kumps, P. Duez und Y. Mardens. Metabolic, nutritional, iatrogenic, and artifactual sources of urinary organic acids: a comprehensive table. *Clinical chemistry*, 48(5):708–717, 2002.
- [140] I. B. Kuznetsov, M. McDuffie und R. Moslehi. A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. *Bioinformatics*, 25(9):1185–6, May 2009.
- [141] S. Selinski, M. Blaszkewicz, M.-L. Lehmann, D. Ovsianikov, O. Moormann, C. Guballa, A. Kress, M. C. Truss, H. Gerullis, T. Otto, D. Barski, G. Niegisch, P. Albers, S. Frees, W. Brenner, J. W. Thüroff, M. Angeli-Greaves, T. Seidel, G. Roth, H. Dietrich, R. Ebbinghaus, H. M. Prager, H. M. Bolt, M. Falkenstein, A. Zimmermann, T. Klein, T. Reckwitz, H. C. Roemer et al. Genotyping NAT2 with only two SNPs (rs1041983 and rs1801280) outperforms the tagging SNP rs1495741 and is equivalent to the conventional 7-SNP NAT2 genotype. *Pharmacogenet Genomics*, 21(10):673–8, Oct 2011.
- [142] Y. J. He, M. H. Shapero und H. L. McLeod. Novel tagging SNP rs1495741 and 2-SNPs (rs1041983 and rs1801280) yield a high prediction of the NAT2 genotype in HapMap samples. *Pharmacogenet Genomics*, 22(4):322–4, Apr 2012.
- [143] M. A. Butler, N. P. Lang, J. F. Young, N. E. Caporaso, P. Vineis, R. B. Hayes, C. H. Teitel, J. P. Massengill, M. F. Lawsen und F. F. Kadlubar. Determination of CYP1A2 and NAT2 phenotypes in human populations by analysis of caffeine urinary metabolites. *Pharmacogenetics*, 2(3):116–27, Jun 1992.
- [144] D. M. Grant, B. K. Tang und W. Kalow. A simple test for acetylator phenotype using caffeine. *Br J Clin Pharmacol*, 17(4):459–64, Apr 1984.

- [145] A. Nyéki, T. Buclin, J. Biollaz und L. A. Decosterd. NAT2 and CYP1A2 phenotyping with caffeine: head-to-head comparison of AFMU vs. AAMU in the urine metabolite ratios. *Br J Clin Pharmacol*, 55(1):62–7, Jan 2003.
- [146] S. A. Marchitti, C. Brocker, D. Stagos und V. Vasiliou. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol*, 4(6):697–720, Jun 2008.
- [147] V. Vasiliou, A. Pappa und T. Estey. Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. *Drug Metab Rev*, 36(2):279–99, May 2004.
- [148] E. E. Bolton, Y. Wang, P. A. Thiessen und S. H. Bryant. PubChem: integrated platform of small molecules and biological activities. *Annual reports in computational chemistry*, 4:217–241, 2008.
- [149] A. Williams. ChemSpider and its expanding web. *Chemistry International*, 30(1):30, 2008.
- [150] A. S. Kristensen, J. Andersen, T. N. Jørgensen, L. Sørensen, J. Eriksen, C. J. Loland, K. Strømgaard und U. Gether. SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol Rev*, 63(3):585–640, Sep 2011.
- [151] Y. Zhou, S. Holmseth, C. Guo, B. Hassel, G. Höfner, H. S. Huitfeldt, K. T. Wanner und N. C. Danbolt. Deletion of the γ -aminobutyric acid transporter 2 (GAT2 and SLC6A13) gene in mice leads to changes in liver and brain taurine contents. *J Biol Chem*, 287(42):35733–46, Oct 2012.
- [152] Y. Zhou, S. Holmseth, R. Hua, A. C. Lehre, A. M. Olofsson, I. Poblete-Naredo, S. A. Kempson und N. C. Danbolt. The betaine-GABA transporter (BGT1, SLC6A12) is predominantly expressed in the liver and at lower levels in the kidneys and at the brain surface. *Am J Physiol Renal Physiol*, 302(3):F316–28, Feb 2012.
- [153] P. Flicek, I. Ahmed, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, C. García-Girón, L. Gordon, T. Hourlier, S. Hunt, T. Juettemann, A. K. Kähäri, S. Keenan, M. Komorowska, E. Kulesha, I. Longden, T. Maurel, W. M. McLaren, M. Muf-

- fato, R. Nag, B. Overduin, M. Pignatelli et al. Ensembl 2013. *Nucleic Acids Res*, 41(Database issue):D48–55, Jan 2013.
- [154] I. Matskevitch, C. A. Wagner, C. Stegen, S. Bröer, B. Noll, T. Risler, H. M. Kwon, J. S. Handler, S. Waldegger, A. E. Busch und F. Lang. Functional characterization of the betaine/gamma-aminobutyric acid transporter BGT-1 expressed in *Xenopus oocytes*. *J Biol Chem*, 274(24):16709–16, Jun 1999.
- [155] M. Liu, R. L. Russell, L. Beigelman, R. E. Handschumacher und G. Pizzorno. beta-alanine and alpha-fluoro-beta-alanine concentrative transport in rat hepatocytes is mediated by GABA transporter GAT-2. *Am J Physiol*, 276(1 Pt 1):G206–10, Jan 1999.
- [156] S. Camargo, D. Bockenhauer und R. Kleta. Aminoacidurias: clinical and molecular aspects. *Kidney international*, 73(8):918–925, 2008.
- [157] K. D. Pruitt, T. Tatusova, W. Klimke und D. R. Maglott. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res*, 37(Database issue):D32–6, Jan 2009.
- [158] Z. Wang, S. J. Nicholls, E. R. Rodriguez, O. Kummu, S. Hörkkö, J. Barnard, W. F. Reynolds, E. J. Topol, J. A. DiDonato und S. L. Hazen. Protein carbamylation links inflammation, smoking, uremia and atherogenesis. *Nature medicine*, 13(10):1176–1184, 2007.
- [159] P. Mydel, Z. Wang, M. Brisslert, A. Hellvard, L. E. Dahlberg, S. L. Hazen und M. Bokarewa. Carbamylation-dependent activation of T cells: a novel mechanism in the pathogenesis of autoimmune arthritis. *J Immunol*, 184(12):6882–90, Jun 2010.
- [160] J. Shi, R. Knevel, P. Suwannalai, M. P. van der Linden, G. M. C. Janssen, P. A. van Veelen, N. E. W. Levarht, A. H. M. van der Helm-van Mil, A. Cerami, T. W. J. Huizinga, R. E. M. Toes und L. A. Trouw. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc Natl Acad Sci U S A*, 108(42):17372–7, Oct 2011.

- [161] F. A. Hommes, A. G. Eller, D. F. Scott und A. L. Carter. Separation of ornithine and lysine activities of the ornithine-transcarbamylase-catalyzed reaction. *Enzyme*, 29(4):271–7, 1983.
- [162] A. L. Carter, A. G. Eller, S. Rufo, K. Metoki und F. A. Hommes. Further evidence for a separate enzymic entity for the synthesis of homocitrulline, distinct from the regular ornithine transcarbamylase. *Enzyme*, 32(1):26–36, 1984.
- [163] Y. Li, X. Yu, J. Ho, D. Fushman, N. M. Allewell, M. Tuchman und D. Shi. Reversible post-translational carboxylation modulates the enzymatic activity of N-acetyl-L-ornithine transcarbamylase. *Biochemistry*, 49(32):6887–95, Aug 2010.
- [164] L. M. Kraus und A. P. Kraus, Jr. Carbamoylation of amino acids and proteins in uremia. *Kidney Int Suppl*, 78:S102–7, Feb 2001.
- [165] O. Y. Al-Dirbashi, Z. N. Al-Hassnan und M. S. Rashed. Determination of homocitrulline in urine of patients with HHH syndrome by liquid chromatography tandem mass spectrometry. *Anal Bioanal Chem*, 386(7-8):2013–7, Dec 2006.
- [166] T. Kato, M. Sano und N. Mizutani. Homocitrullinuria and homoargininuria in lysinuric protein intolerance. *J Inherit Metab Dis*, 12(2):157–161, 1989.
- [167] J. A. Camacho, C. Obie, B. Biery, B. K. Goodman, C. A. Hu, S. Al-mashanu, G. Steel, R. Casey, M. Lambert, G. A. Mitchell und D. Valle. Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. *Nat Genet*, 22(2):151–158, Jun 1999.
- [168] P. C. Sizeland, S. T. Chambers, M. Lever, L. M. Bason und R. A. Robson. Organic osmolytes in human and other mammalian kidneys. *Kidney Int*, 43(2):448–453, Feb 1993.
- [169] M. Lever, P. C. Sizeland, L. M. Bason, C. M. Hayman, R. A. Robson und S. T. Chambers. Abnormal glycine betaine content of the blood and urine of diabetic and renal patients. *Clin Chim Acta*, 230(1):69–79, Oct 1994.

- [170] W. J. Dellow, S. T. Chambers, M. Lever, H. Lunt und R. A. Robson. Elevated glycine betaine excretion in diabetes mellitus patients is associated with proximal tubular dysfunction and hyperglycemia. *Diabetes Res Clin Pract*, 43(2):91–99, Feb 1999.
- [171] M. Lever, W. Atkinson, P. M. George und S. T. Chambers. An abnormal urinary excretion of glycine betaine may persist for years. *Clin Biochem*, 40(11):798–801, Jul 2007.
- [172] M. Lever, S. Slow, D. O. McGregor, W. J. Dellow, P. M. George und S. T. Chambers. Variability of plasma and urine betaine in diabetes mellitus and its relationship to methionine load test responses: an observational study. *Cardio-vasc Diabetol*, 11:34, 2012.
- [173] M. Tamari, Y. Daigo, S. Ishikawa und Y. Nakamura. Genomic structure of a novel human gene (XYLB) on chromosome 3p22 > p21. 3 encoding a xylulokinase-like protein. *Cytogenetic and Genome Research*, 82(1-2):101–104, 1998.
- [174] R. D. Bunker, E. M. Bulloch, J. M. Dickson, K. M. Loomes und E. N. Baker. Structure and function of human xylulokinase, an enzyme with important roles in carbohydrate metabolism. *J Biol Chem*, 288(3):1643–1652, Jan 2013.
- [175] D. A. Barngrover, H. C. Stevens und W. L. Dills. D-Xylulose-1-phosphate: enzymatic assay and production in isolated rat hepatocytes. *Biochem Biophys Res Commun*, 102(1):75–80, Sep 1981.
- [176] R. P. Holmes und D. G. Assimos. Glyoxylate synthesis, and its modulation and influence on oxalate synthesis. *J Urol*, 160(5):1617–1624, Nov 1998.
- [177] T. Wood. Physiological functions of the pentose phosphate pathway. *Cell Biochem Funct*, 4(4):241–247, Oct 1986.
- [178] R. A. Conyers, R. Bais und A. M. Rofe. The relation of clinical catastrophes, endogenous oxalate production, and urolithiasis. *Clin Chem*, 36(10):1717–1730, Oct 1990.
- [179] A. Bär und G. Oesterhelt. Conversion of [U-13C]xylitol and D-[U-13C]glucose into urinary [1,2-13C]glycollate and [1,2-13C]oxalate in man. *Int J Vitam Nutr*

Res Suppl, 28:119–133, 1985.

- [180] R. A. Conyers, T. W. Huber, D. W. Thomas, A. M. Rofe, R. Bais und R. G. Edwards. A one-compartment model for calcium oxalate tissue deposition during xylitol infusions in humans. *Int J Vitam Nutr Res Suppl*, 28:47–57, 1985.
- [181] S. Hauschildt, R. A. Chalmers, A. M. Lawson, K. Schultis und R. W. Watts. Metabolic investigations after xylitol infusion in human subjects. *Am J Clin Nutr*, 29(3):258–273, Mar 1976.
- [182] M. Veiga-da Cunha, F. Hadi, T. Balligand, V. Stroobant und E. Van Schaftingen. Molecular identification of hydroxylysine kinase and of ammoniophospholyases acting on 5-phosphohydroxy-L-lysine and phosphoethanolamine. *J Biol Chem*, 287(10):7246–55, Mar 2012.
- [183] H. L. Fleshood und H. C. Pitot. The metabolism of O-phosphorylethanolamine in animal tissues. i. O-phosphorylethanolamine phospho-lyase: partial purification and characterization. *J Biol Chem*, 245(17):4414–20, Sep 1970.
- [184] A. Lykidis, J. Wang, M. A. Karim und S. Jackowski. Overexpression of a mammalian ethanolamine-specific kinase accelerates the CDP-ethanolamine pathway. *J Biol Chem*, 276(3):2174–9, Jan 2001.
- [185] S. Kim, K.-H. Choi, A. F. Baykiz und H. K. Gershenfeld. Suicide candidate genes associated with bipolar disorder and schizophrenia: an exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics*, 8:413, 2007.
- [186] L. Shao und M. P. Vawter. Shared gene expression alterations in schizophrenia and bipolar disorder. *Biol Psychiatry*, 64(2):89–97, Jul 2008.
- [187] K. N. Fedde und M. P. Whyte. Alkaline phosphatase (tissue-nonspecific isoenzyme) is a phosphoethanolamine and pyridoxal-5'-phosphate ectophosphatase: normal and hypophosphatasia fibroblast study. *Am J Hum Genet*, 47(5):767–75, Nov 1990.
- [188] S. H. Moolenaar, J. Poggi-Bach, U. F. Engelke, J. M. Corstiaensen, A. Heerschap, J. G. de Jong, B. A. Binzak, J. Vockley und R. A. Wevers. Defect

- in dimethylglycine dehydrogenase, a new inborn error of metabolism: NMR spectroscopy study. *Clin Chem*, 45(4):459–64, Apr 1999.
- [189] B. A. Binzak, J. G. Vockley, R. B. Jenkins und J. Vockley. Structure and analysis of the human dimethylglycine dehydrogenase gene. *Mol Genet Metab*, 69(3):181–7, Mar 2000.
- [190] A. Bröer, K. Klingel, S. Kowalczyk, J. E. J. Rasko, J. Cavanaugh und S. Bröer. Molecular cloning of mouse amino acid transport system B0, a neutral amino acid transporter related to hartnup disorder. *J Biol Chem*, 279(23):24467–76, Jun 2004.
- [191] D. N. Baron, C. E. Dent, H. Harris, E. W. Hart und J. B. Jepson. Hereditary pellagra-like skin rash with temporary cerebellar ataxia, constant renal aminoaciduria, and other bizarre biochemical features. *Lancet*, 271(6940):421–8, Sep 1956.
- [192] W. Henderson. A case of Hartnup disease. *Arch Dis Child*, 33(168):114–7, Apr 1958.
- [193] M. D. Milne. Disorders of amino-acid transport. *Br Med J*, 1(5379):327–36, Feb 1964.
- [194] H. F. Seow, S. Bröer, A. Bröer, C. G. Bailey, S. J. Potter, J. A. Cavanaugh und J. E. J. Rasko. Hartnup disorder is caused by mutations in the gene encoding the neutral amino acid transporter SLC6A19. *Nat Genet*, 36(9):1003–7, Sep 2004.
- [195] R. Kleta, E. Romeo, Z. Ristic, T. Ohura, C. Stuart, M. Arcos-Burgos, M. H. Dave, C. A. Wagner, S. R. M. Camargo, S. Inoue, N. Matsuura, A. Helip-Wooley, D. Bockenhauer, R. Warth, I. Bernardini, G. Visser, T. Eggermann, P. Lee, A. Chairoungdua, P. Jutabha, E. Babu, S. Nilwarangkoon, N. Anzai, Y. Kanai, F. Verrey, W. A. Gahl und A. Koizumi. Mutations in SLC6A19, encoding B0AT1, cause Hartnup disorder. *Nat Genet*, 36(9):999–1002, Sep 2004.
- [196] S. Halvorsen, O. Hygstedt, R. Jagenburg und O. Sjaastad. Cellular transport of L-histidine in Hartnup disease. *J Clin Invest*, 48(8):1552–9, Aug 1969.

- [197] A. Garrod. The incidence of alkaptonuria: a study in chemical individuality. *The Lancet*, 160(4137):1616–1620, 1902.
- [198] M. Boll, M. Foltz, I. Rubio-Aliaga und H. Daniel. A cluster of proton/amino acid transporter genes in the human and mouse genomes. *Genomics*, 82(1):47–56, Jul 2003.
- [199] J. M. Vanslambrouck, A. Bröer, T. Thavyogarah, J. Holst, C. G. Bailey, S. Bröer und J. E. J. Rasko. Renal imino acid and glycine transport system ontogeny and involvement in developmental iminoglycinuria. *Biochem J*, 428(3):397–407, Jun 2010.
- [200] UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res*, 42(Database issue):D191–8, Jan 2014.
- [201] K. I. Nakayama und K. Nakayama. Regulation of the cell cycle by SCF-type ubiquitin ligases. *Semin Cell Dev Biol*, 16(3):323–33, Jun 2005.
- [202] K. Suhre, S. Audic und J.-M. Claverie. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc Natl Acad Sci U S A*, 102(41):14689–93, Oct 2005.
- [203] D. I. Chasman, C. Fuchsberger, C. Pattaro, A. Teumer, C. A. Böger, K. Endlich, M. Olden, M.-H. Chen, A. Tin, D. Taliun, M. Li, X. Gao, M. Gorski, Q. Yang, C. Hundertmark, M. C. Foster, C. M. O’Seaghdha, N. Glazer, A. Isaacs, C.-T. Liu, A. V. Smith, J. R. O’Connell, M. Struchalin, T. Tanaka, G. Li, A. D. Johnson, H. J. Gierman, M. F. Feitosa et al. Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet*, 21(24):5329–43, Dec 2012.
- [204] M. Watanabe, M. E. Suliman, A. R. Qureshi, E. Garcia-Lopez, P. Bány, O. Heimbürger, P. Stenvinkel und B. Lindholm. Consequences of low plasma histidine in chronic kidney disease patients: associations with inflammation, oxidative stress, and mortality. *Am J Clin Nutr*, 87(6):1860–6, Jun 2008.
- [205] J. Coresh, E. Selvin, L. A. Stevens, J. Manzi, J. W. Kusek, P. Eggers, F. Van Lente und A. S. Levey. Prevalence of chronic kidney disease in the United States. *JAMA*, 298(17):2038–47, Nov 2007.

- [206] Q.-L. Zhang und D. Rothenbacher. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health*, 8:117, 2008.
- [207] J. Dumont, I. Huybrechts, A. Spinneker, F. Gottrand, E. Grammatikaki, N. Bevilacqua, K. Vyncke, K. Widhalm, A. Kafatos, D. Molnar, I. Labayen, M. Gonzalez-Gross, P. Amouyel, L. A. Moreno, A. Meirhaeghe, J. Dallongeville und HELENA Study group. FADS1 genetic variability interacts with dietary α -linolenic acid intake to affect serum non-HDL-cholesterol concentrations in European adolescents. *J Nutr*, 141(7):1247–53, Jul 2011.
- [208] Y. Lu, E. J. Feskens, M. E. Dollé, S. Imholz, W. M. Verschuren, M. Müller und J. M. Boer. Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem cohort study. *Am J Clin Nutr*, 92(1):258–65, Jul 2010.
- [209] K. Horvatovich, S. Bokor, N. Polgar, P. Kisfali, F. Hadarits, L. Jaromi, V. Csongei, J. Repasy, D. Molnar und B. Melegh. Functional glucokinase regulator gene variants have inverse effects on triglyceride and glucose levels, and decrease the risk of obesity in children. *Diabetes Metab*, 37(5):432–9, Nov 2011.
- [210] K. Kristiansson, M. Perola, E. Tikkanen, J. Kettunen, I. Surakka, A. S. Havulinna, A. Stancáková, C. Barnes, E. Widen, E. Kajantie, J. G. Eriksson, J. Viikari, M. Kähönen, T. Lehtimäki, O. T. Raitakari, A.-L. Hartikainen, A. Ruukonen, A. Pouta, A. Jula, A. J. Kangas, P. Soininen, M. Ala-Korpela, S. Männistö, P. Jousilahti, L. L. Bonnycastle, M.-R. Jarvelin, J. Kuusisto, F. S. Collins et al. Genome-wide screen for metabolic syndrome susceptibility loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet*, 5(2):242–9, Apr 2012.
- [211] M. Vaxillaire, C. Cavalcanti-Proença, A. Dechaume, J. Tichet, M. Marre, B. Balkau, P. Froguel und DESIR Study Group. The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. *Diabetes*, 57(8):2253–7, Aug 2008.

8. Tabellenverzeichnis

5.1. Erste mGWAS im Urin: Loci mit genomweit signifikanten Assoziationen zu metabolischen Merkmalen im SHIP-0 males Datensatz	43
5.2. Erste mGWAS im Urin: Assoziationsergebnisse in den Verifikations- bzw. Replikationsdatensätzen	44
5.3. Zweite mGWAS im Urin: signifikante Assoziationen genetischer Loci zu metabolischen Merkmalen im Discovery-Datensatz	51
5.4. Zweite mGWAS im Urin: Replikation der Assoziationen im KORA-F4-Urin-Datensatz	52
5.5. Zweite mGWAS im Urin: Vergleich mit Assoziationen im Serum der KORA-F4- und TwinsUK-Studien	54
5.6. GWAS mit NMR-Merkmalen im Plasma: Genetische Assoziationen zu NMR-Bins und NMR-Ratios	55
5.7. GWAS mit NMR-Merkmalen im Plasma: Vergleich der Assoziationsstärken zu NMR-bestimmten Lipoprotein-Subklassen, NMR-Bins und NMR-Ratios	59
5.8. GWAS mit NMR-Merkmalen im Plasma: Assoziationsergebnisse aus früheren Studien mit Blutproben aus KORA F4	61
5.9. GWAS mit NMR-Merkmalen: Korrelationen zwischen NMR-Bins bzw. -Ratios und metabolischen Merkmalen	66
6.1. Mit SNP rs4921914 (NAT2) assoziierte Metaboliten und Ratios	73
6.2. Mit SNP rs8101881 (SLC7A9) assoziierte Metaboliten oder Metabolitenratios	76
6.3. Mit SNP rs17279437 (SLC6A20) assoziierte Metaboliten oder Metabolitenratios	78
6.4. Assoziationen zu rs2216405 (CPS1)	86
6.5. Assoziationen zu rs4760099 (HPD)	87
6.6. Assoziationen zu rs4921914 (NAT2)	88
6.7. Assoziationen zu rs17279437 (SLC6A20)	92

6.8. Assoziationen zu rs11613331 (SLC6A13)	94
6.9. Assoziationen zu rs8101881 (SLC7A9)	97
6.10. Assoziationen zu rs7094971 (SLC16A9)	99
6.11. Assoziationen zu rs37369 (AGXT2)	100
6.12. Assoziationen zu rs17118 (XYLB)	103
6.13. Assoziationen zu rs7692671 (ETNPPL)	105
6.14. Assoziationen zu rs248386 (DMGDH)	107
6.15. Assoziationen zu rs7719875 (SLC6A19)	109
6.16. Assoziationen zu rs3846710 (SLC36A2)	111
6.17. Assoziationen zu rs2338755 (FBXL20)	112
B.1. Korrelation zwischen NMR-Bins und Metabolitenkonzentrationen . .	173

9. Abbildungsverzeichnis

3.1. Innere und äußere Einflüsse auf Phänotypen	7
3.2. Ratios in mGWAS	14
4.1. Präprozessierung der NMR-Spektren	29
4.2. Vereinfachte Darstellung des zur Korrektur des Batch-Effektes verwendeten R-Codes	31
4.3. Korrektur eines Batch-Effektes, der die Probanden in zwei Gruppen unterteilt	32
5.1. Erste mGWAS im Urin: Metabolitenkonzentrationen und -ratios im Discovery-Datensatz und in den Replikationsdatensätzen	45
5.1. Studiendesign	56
5.2. Vergleich mit Assoziationsdaten aus früheren Studien mit Blutproben des KORA-F4-Datensatzes	60
5.3. Korrelationspektren für Glycin, PC aa C36:3, HDL und Glucose	63
6.1. Enzymatische Reaktion von AGXT2	69
6.2. Assoziation von rs37369 (AGXT2) zur beta-Aminoisobutyrat-Konzentration	71
6.3. Assoziation von rs4921914 (NAT2) zum Formiat/Succinat-Ratio	72
6.4. Assoziation von rs830124 (WDR66/HPD) zur alpha-Hydroxyisovalerat-Konzentration	74
6.5. Die Assoziation von rs8101881 (SLC7A9) zum Lysin/Valin-Ratio	75
6.6. Assoziation von rs17279437 (SLC6A20) zum Alanin/N,N-Dimethylglycin-Ratio	78
6.7. Die Assoziationen der 14 identifizierten genetischen Loci in verschiedenen Medien und Studien	80
6.8. Zweistufige Umsetzung von Glycin zu Glycolat	83
6.9. Glycin/Glycolat-Ratio bei Männern und Frauen	84

6.10. Auswirkungen des CPS1-Mangels auf die Carbamoylphosphat-Synthese und den Harnstoffzyklus	85
6.11. Glycin-Synthese aus Ammoniak durch das Glycine Cleavage System .	86
6.12. N,N-Dimethylglycin, Prolin, Pyroglutamin (Metabolon) und 5-Amino-3,4-dihydro-2H-pyrrol-2-carboxylsäure	91
6.13. beta-Aminoisobutansäure, beta-Alanin und gamma-Aminobutansäure	93
6.14. Carbamoylierung von Lysin zu Homocitrullin (EC 2.1.3.8)	96
6.15. Betain und Carnitin	98
6.16. Abbau von D-Xylulose über zwei Stoffwechselfade	101
6.17. Durch ETNPPL katalysierte Reaktion (EC 4.2.3.2)	104
6.18. Umwandlung von N,N-Dimethylglycin zu Glycin (EC 1.5.8.4 und 1.5.8.3)	106
6.19. Assoziations- und Korrelationspektren für rs2216405 (CPS1) und Glycin	116
6.20. Assoziations- und Korrelationspektren für rs174547 (FADS1) und PC aa C36:3 sowie PC aa C36:4	119
6.21. Assoziations- und Korrelationspektren für rs780094 (GCKR) und Triglyceride sowie Glucose	121
6.22. Intensität der NMR-Signale nach rs4488133-Genotyp (PYROXD2) . .	124
A.1. Zweite mGWAS im Urin: Regionale Assoziationsplots, Boxplots und QQ-Plots	164
B.1. GWAS mit NMR-Merkmalen im Blutplasma: Regionale Assoziationsplots, Boxplots, Histogramme und QQ-Plots	177

10. Lebenslauf

Persönliche Angaben

Name: Raffler
Vornamen: Johannes Florian
Geburtsdatum: 24. Dezember 1982
Geburtsort: Augsburg
Staatsangehörigkeit: deutsch

Ausbildung und Wehrdienst

1993 - 2002: Humanistisches Gymnasium bei St. Stephan, Augsburg
2002 - 2003: Zivildienst im Studienseminar St. Joseph, Augsburg
2003 - 2009: Studium der Bioinformatik (Diplom) an der Ludwig-Maximilians-Universität (LMU) München sowie der Technischen Universität München (TUM)
seit 2009: Promotionsstudium an der Fakultät für Biologie der LMU München

Wissenschaftliche Tätigkeiten

2008 - 2009: Diplomarbeit in der AG Metabolomics des Instituts für Bioinformatik und Systembiologie (IBIS), Helmholtz Zentrum München (HMGU)
2009 - 2012: Promotionsstudent in der AG Metabolomics (IBIS, HMGU)
Stipendiat des DFG-Graduiertenkolleges RECESS (GRK 1563)
seit 2012: Wissenschaftlicher Mitarbeiter in der AG Metabolomics (IBIS, HMGU)

11. Erklärung

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 22. Juli 2014

Erklärung

Hiermit erkläre ich,

- dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
- dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
- dass ich mich mit Erfolg der Doktorprüfung im Hauptfach ... und in den Nebenfächern ... bei der Fakultät für ... der ... unterzogen habe.
- dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den 22. Juli 2014

12. Danksagungen

Ich möchte mich bei allen bedanken, die mich im Studium und während meiner Zeit als Doktorand begleitet haben.

Insbesondere danke ich ganz herzlich meinem Doktorvater, Karsten Suhre, für die intensive, unkomplizierte, freundschaftliche Betreuung und Zusammenarbeit sowie die Unterstützung während der vergangenen Jahre.

Herzlichen Dank auch an Gabi Kastenmüller, die ab 2011 die Leitung der Metabolomics-Arbeitsgruppe von Karsten Suhre übernommen hat. Sie war für mich seitdem ein verlässlicher Ansprechpartner und hat mich bei allen Projekten beraten und begleitet.

Vielen Dank an meine Kollegen der Metabolomics-Arbeitsgruppe, insbesondere Elisabeth, Matthias und Jörn, für die gute Zusammenarbeit und die lockere Arbeitsatmosphäre.

Ich möchte es auch nicht versäumen, den vielen Kollegen der SHIP- und KORA-Studien zu danken, ohne die die Durchführung der in dieser Arbeit besprochenen Studien nicht möglich gewesen wäre.

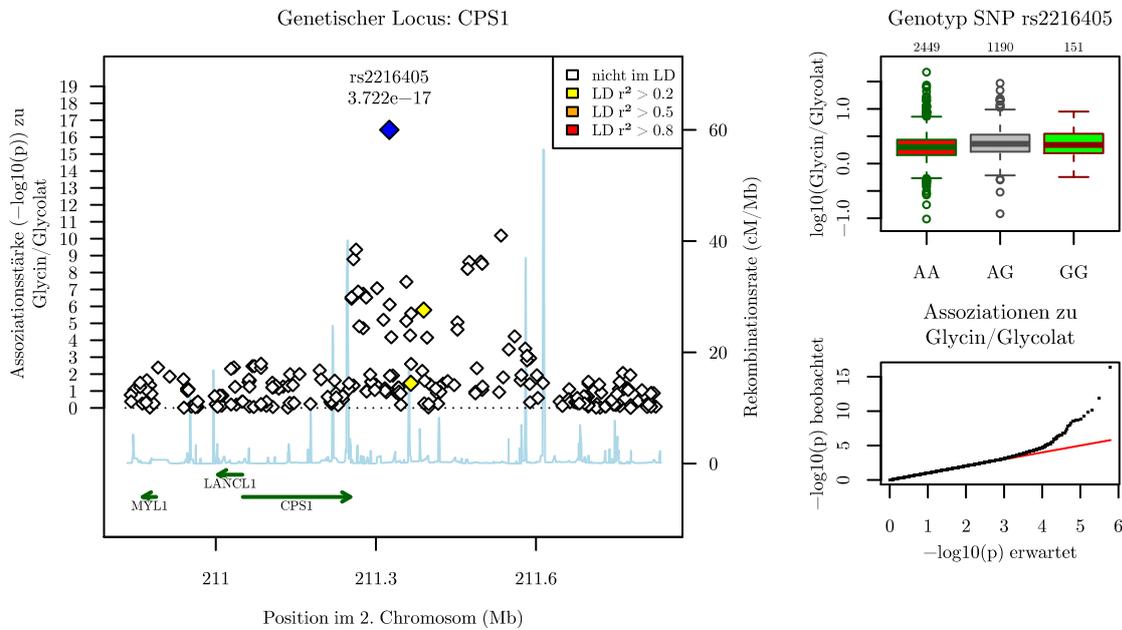
Weiterhin bedanke ich mich bei der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung sowie die wissenschaftliche Förderung im Rahmen des internationalen Graduiertenkolleges RECESS.

Ganz besonders möchte ich mich bei meinen Eltern bedanken, die mir die Aufnahme eines Studiums ermöglicht haben und damit den Grundstein für den erfolgreichen Abschluss dieser Arbeit gelegt haben.

Und nicht zuletzt möchte ich herzlich meiner Frau Franziska Danke sagen - für alles, was neben der Dissertation noch wichtig war und ist.

A. Anhang zur zweiten mGWAS im Urin

Abbildung A.1.: Zweite mGWAS im Urin: Regionale Assoziationsplots, Boxplots und QQ-Plots



Die Plots gelten für die in Tabelle 5.3 genannten genetischen Assoziationen zu metabolischen Merkmalen (Konzentrationen oder Ratios). Links: Regionaler Assoziationsplot, basierend auf den SNPs, die im für die GWAS verwendeten Datensatz nach der Qualitätskontrolle enthalten waren. Die X-Achse zeigt die Position des SNPs bzw. des Gens, die Y-Achse die negative, dekadisch logarithmierte Assoziationsstärke (p-Wert) des SNPs zum metabolischen Merkmal. Die Annotation der Gene sowie die chromosomalen Positionen basieren auf dem NCBI-Referenzgenom (hg18) sowie auf RefSeq (36.3). Die Korrelationskoeffizienten, die das Kopplungsungleichgewicht (*Linkage Disequilibrium*, kurz LD) eines SNPs zum Top-Hit (blau markiert) angeben, basieren auf Daten des HapMap-Projektes (Release 27, Phase I-III, CEU Population) [72]. Rechts oben: Boxplots der log₁₀-transformierten NMR-Signalintensitäten respektive der NMR-Ratios, aufgetrennt nach Genotyp der jeweiligen Probanden (von links nach rechts: Homozygote Träger des häufigeren Allels, heterozygote Träger sowie homozygote Träger des selteneren Allels). Die Anzahl der Probanden in den Gruppen ist oberhalb der entsprechenden Box zu finden. Weiterführende Informationen zu den regionalen Assoziationsplot und Boxplots findet sich im Begleittext zu Abbildung 6.2. Rechts unten: QQ-Plot. Der Quantile-Quantile-Plot zeigt die beobachtete gegen die theoretisch zu erwartende Verteilung der p-Werte der Assoziationen aller gestesten SNPs zu dem im Titel genannten metabolischen Merkmal.

Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)

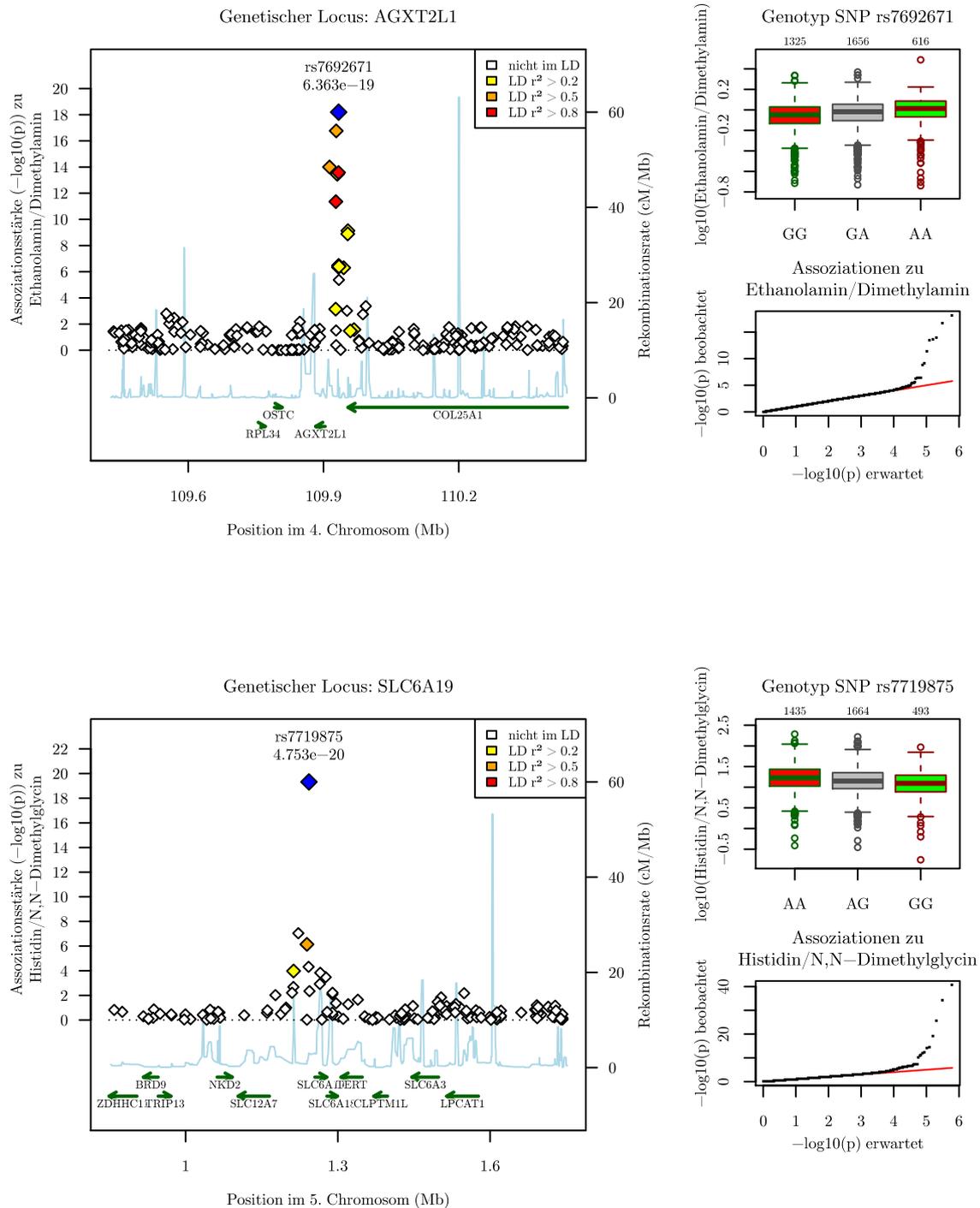


Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)

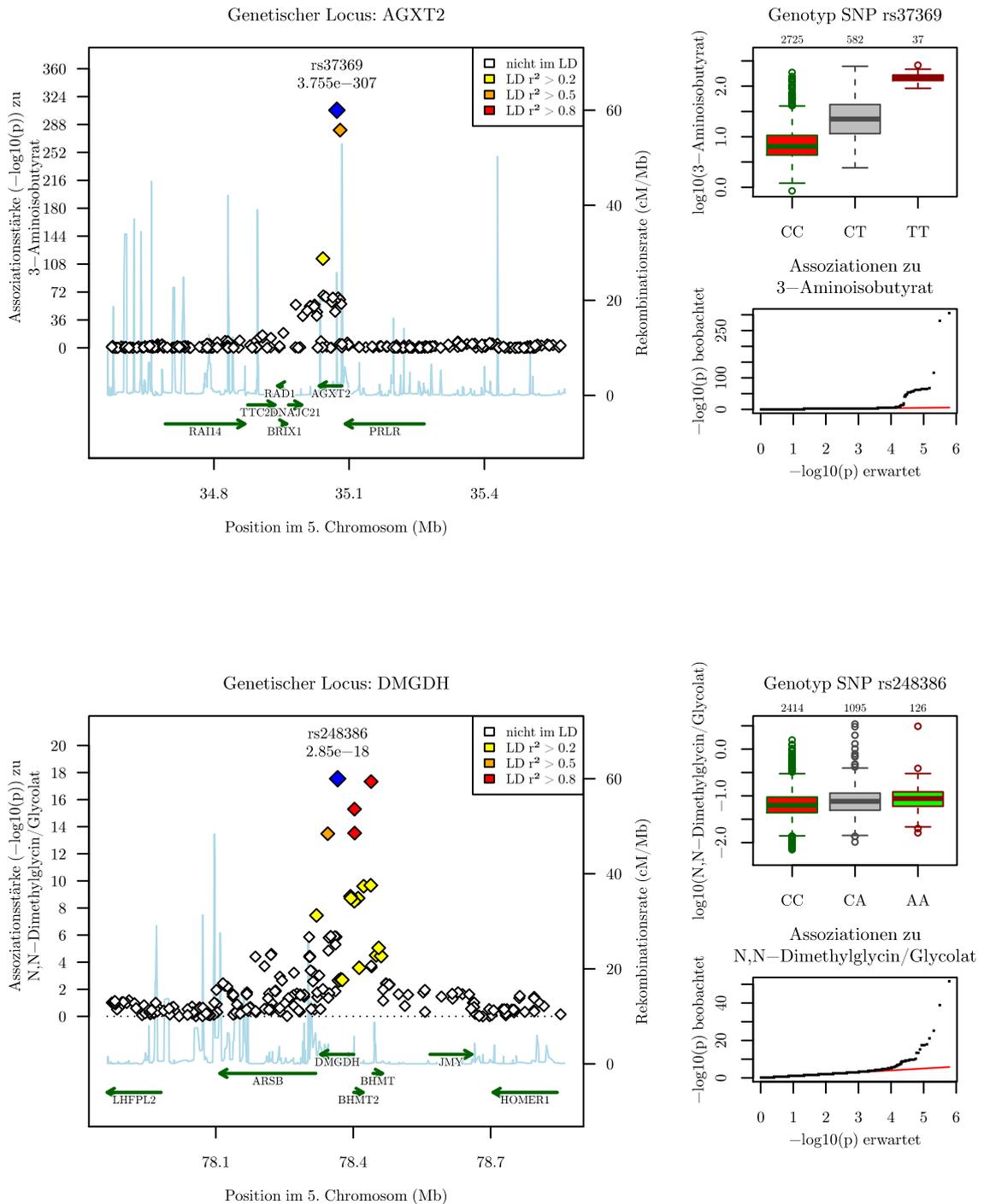


Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)

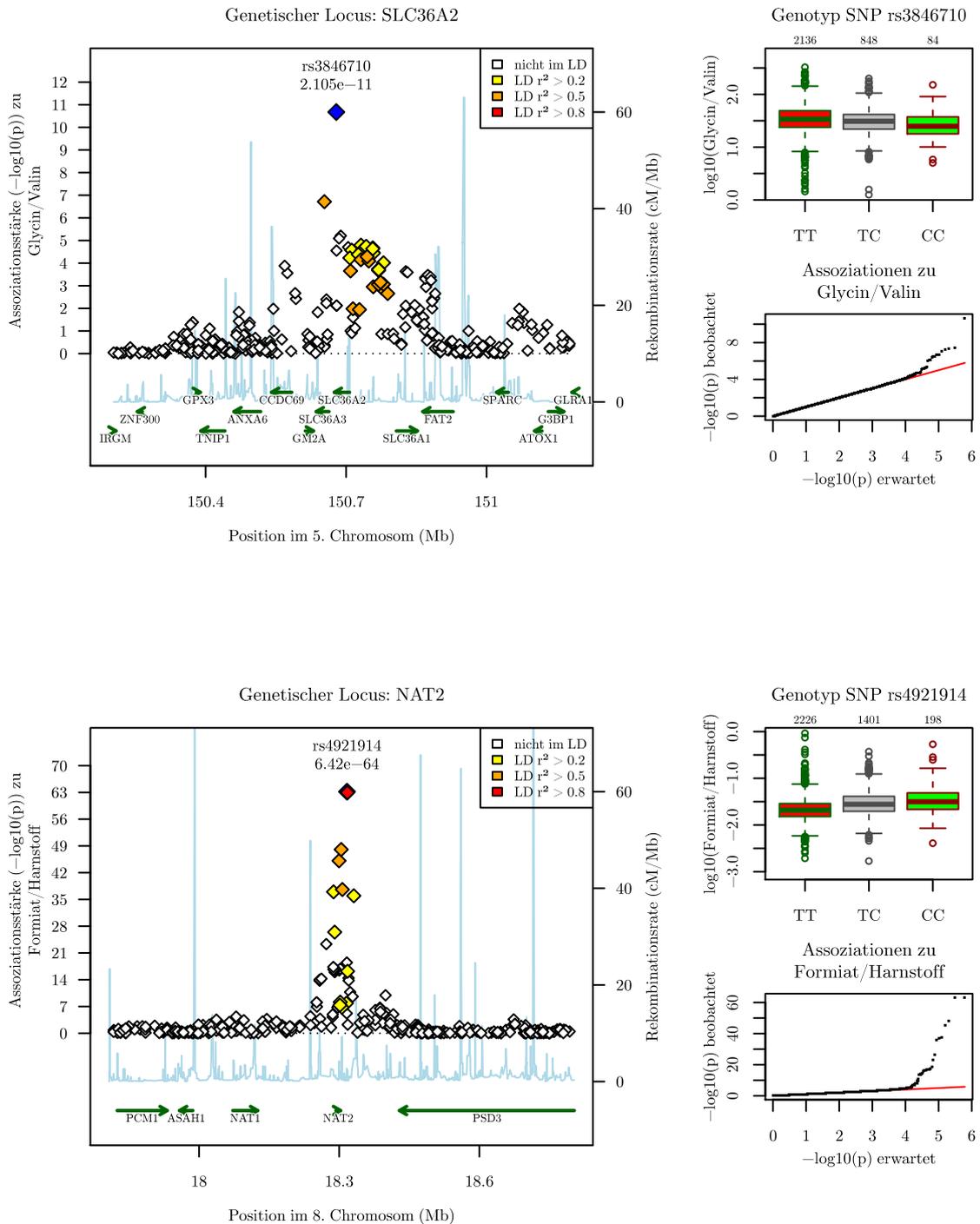


Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)

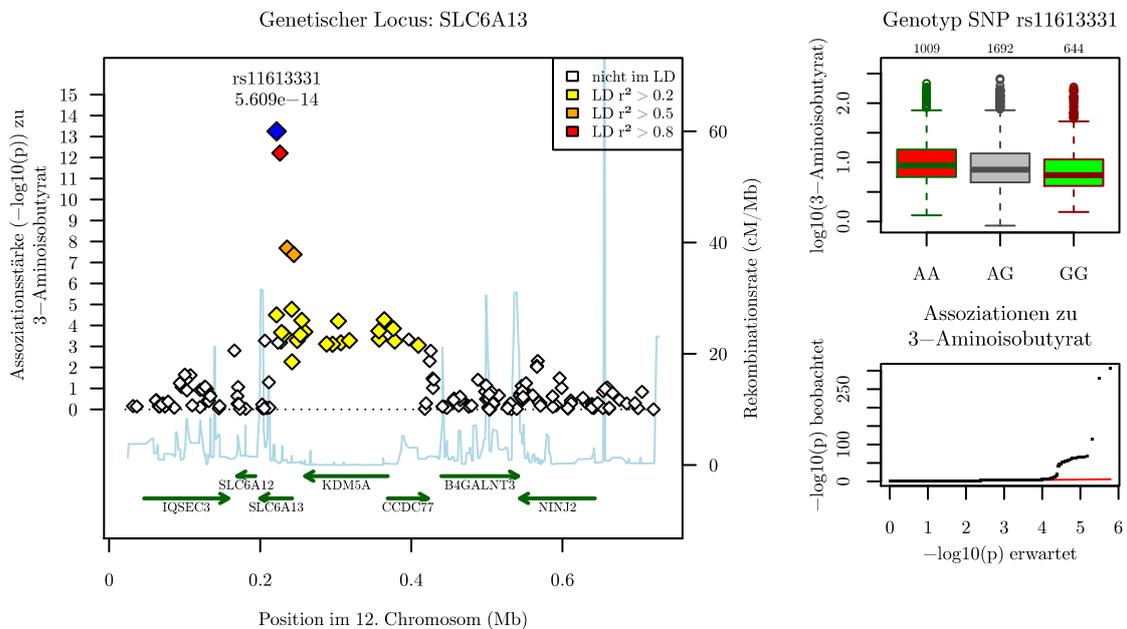
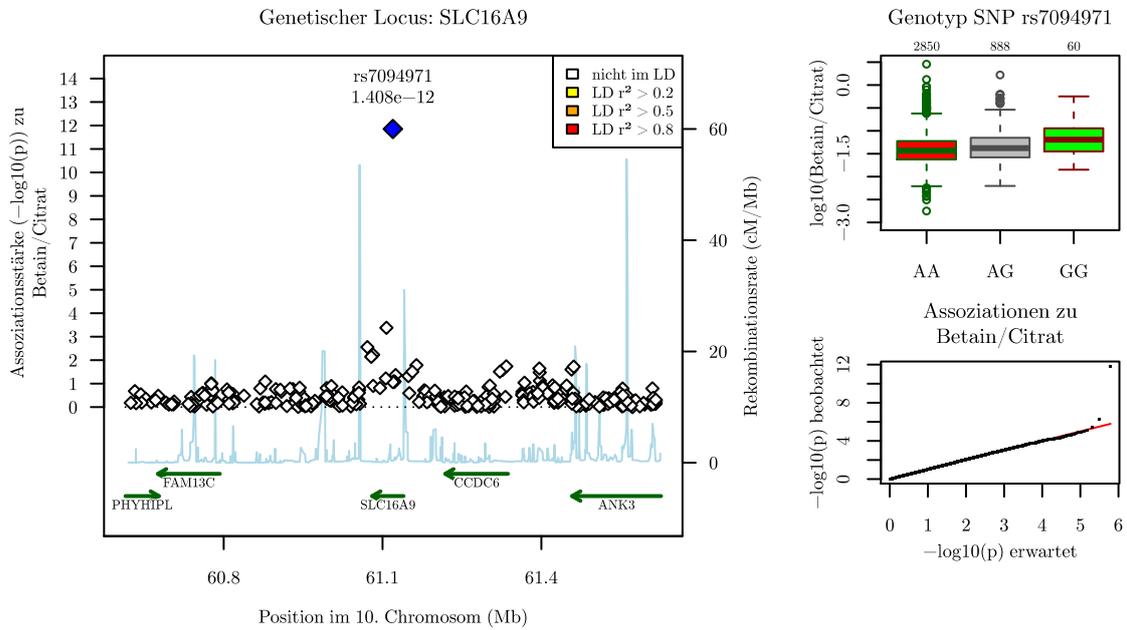


Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)

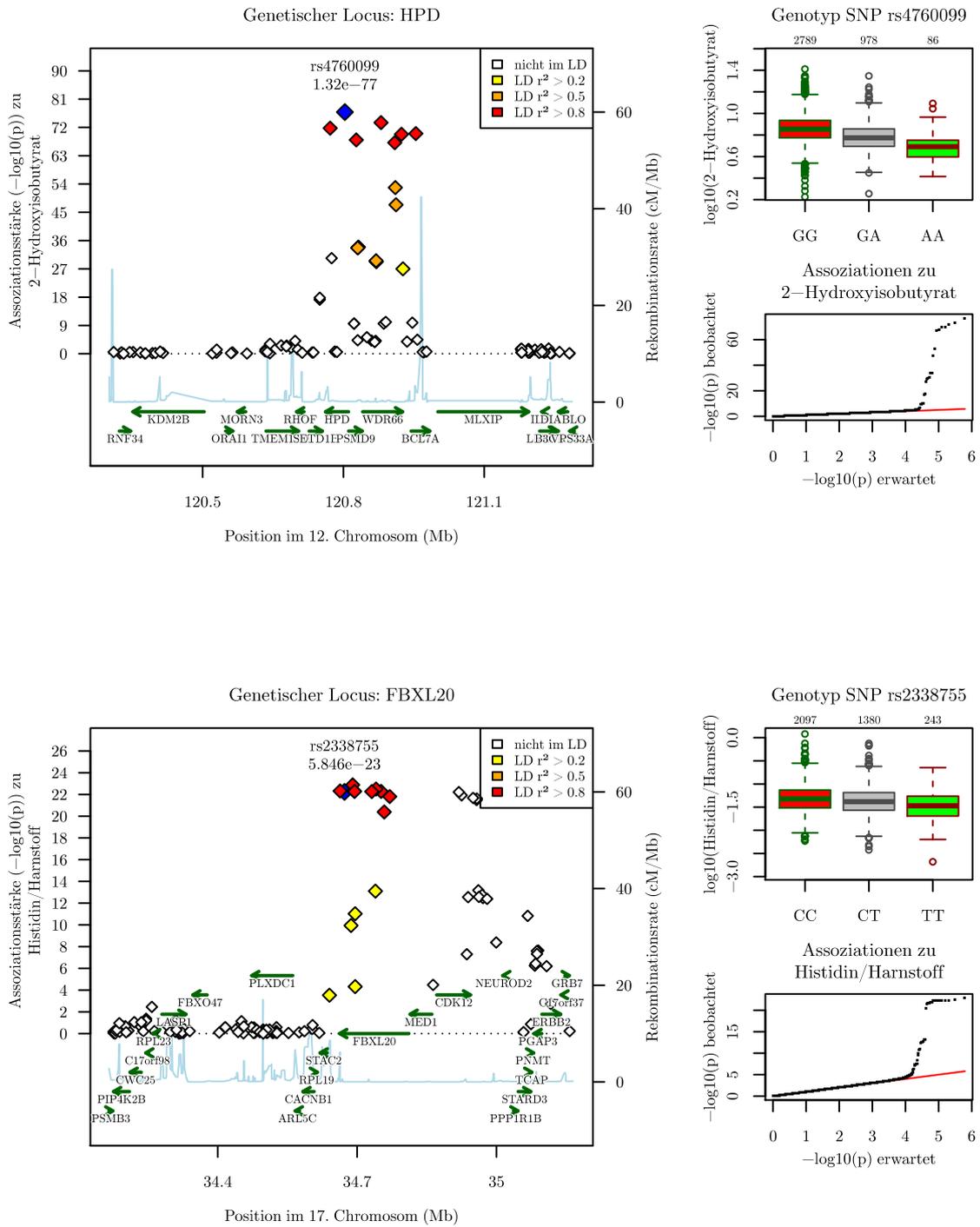
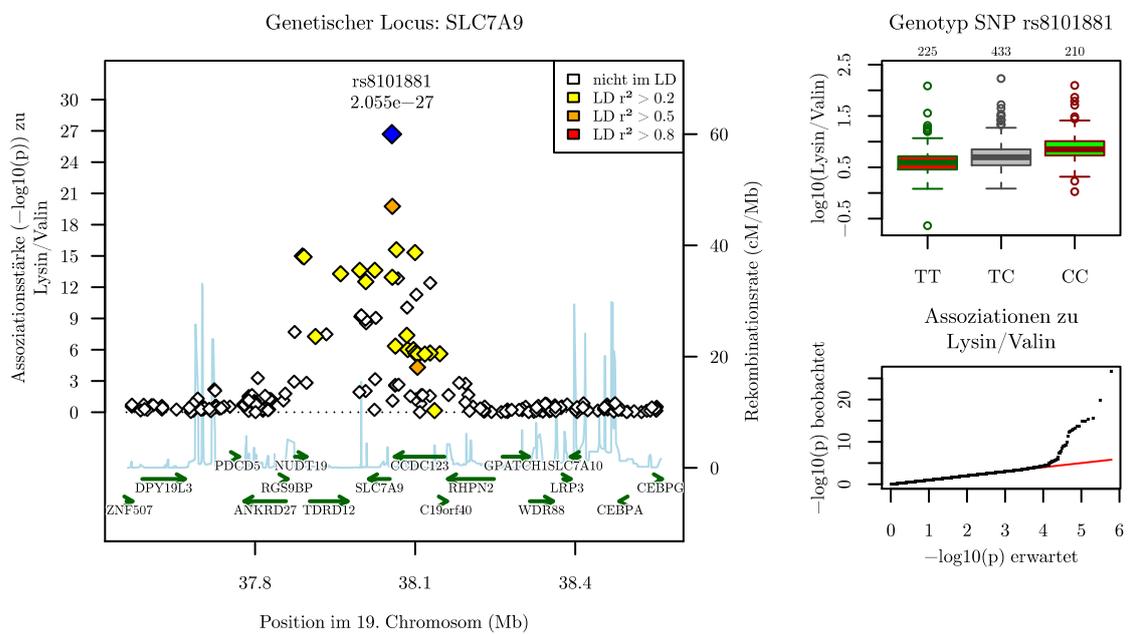


Abbildung A.1.: Regionale Assoziationsplots, Boxplots und QQ-Plots (Forts.)



B. Anhang zur GWAS mit NMR-Merkmalen im Plasma

Tabelle B.1.: Korrelation zwischen NMR-Bins und Metabolitenkonzentrationen

Chemical Shift	r_s^2		Metabolit	Plattform	N	CV
2,257 ppm	0,94	↗	Triglyceride	klinische Biochemie	1.726	
0,892 ppm	0,82	↗	Gesamtes Cholesterin	klinische Biochemie	1.742	
0,870 ppm	0,69	↗	HDL Cholesterin	klinische Biochemie	1.743	
3,386 ppm	0,61	↗	1,5-Anhydroglucitol	semi-quantitative MS	1.749	5,6%
3,276 ppm	0,56	↗	LDL Cholesterin	klinische Biochemie	1.742	
3,965 ppm	0,55	↗	Kreatin	semi-quantitative MS	1.740	5,4%
5,783 ppm	0,51	↗	Harnstoff	semi-quantitative MS	1.737	10,1%
3,875 ppm	0,51	↗	Glucose	klinische Biochemie	1.718	
4,141 ppm	0,50	↗	Lactat	semi-quantitative MS	1.725	8,6%
2,802 ppm	0,48	↗	PC aa C38:5	quantitative MS	1.732	5,6%
2,794 ppm	0,48	↗	PC aa C38:4	quantitative MS	1.728	5,3%
1,285 ppm	0,42	↗	PC aa C36:1	quantitative MS	1.724	5,2%
0,886 ppm	0,39	↗	Phosphat	semi-quantitative MS	1.748	7,8%
2,812 ppm	0,39	↗	PC aa C40:6	quantitative MS	1.728	5,0%
0,879 ppm	0,38	↗	PC ae C32:2	quantitative MS	1.733	13,4%
2,796 ppm	0,38	↗	PC aa C40:5	quantitative MS	1.723	5,2%
2,744 ppm	0,38	↗	PC ae C34:3	quantitative MS	1.735	5,7%
0,886 ppm	0,37	↗	Palmitoylsphingomyelin	semi-quantitative MS	1.741	
2,685 ppm	0,37	↗	Citrat	semi-quantitative MS	1.737	9,2%
5,307 ppm	0,37	↗	PC ae C40:1	quantitative MS	1.734	9,3%
2,806 ppm	0,36	↗	PC aa C38:6	quantitative MS	1.728	5,3%
2,803 ppm	0,36	↗	PC aa C36:5	quantitative MS	1.715	6,4%
1,419 ppm	0,36	↗	PC aa C38:3	quantitative MS	1.732	4,8%

Korrelation zwischen NMR-Bins und Metabolitenkonzentrationen

Chemical Shift	r_s^2		Metabolit	Plattform	N	CV
2,807 ppm	0,35	↗	PC aa C36:6	quantitative MS	1.726	10,6%
2,806 ppm	0,34	↗	PC ae C38:0	quantitative MS	1.728	11,9%
2,018 ppm	0,34	↗	PC aa C36:3	quantitative MS	1.729	5,5%
1,261 ppm	0,33	↗	PC aa C34:3	quantitative MS	1.730	6,5%
1,295 ppm	0,33	↗	1-Palmitoylglycerin	semi-quantitative MS	1.589	19,9%
2,798 ppm	0,33	↗	PC aa C36:4	quantitative MS	1.727	6,3%
1,251 ppm	0,33	↗	PC aa C34:4	quantitative MS	1.727	6,1%
3,476 ppm	0,32	↗	3-Hydroxybutyrat	semi-quantitative MS	1.726	8,0%
3,264 ppm	0,32	↗	PC ae C32:1	quantitative MS	1.733	7,0%
0,888 ppm	0,30	↗	PC aa C28:1	quantitative MS	1.732	10,9%
1,560 ppm	0,30	↗	PC aa C36:2	quantitative MS	1.726	5,7%
2,744 ppm	0,30	↗	PC ae C34:2	quantitative MS	1.734	5,8%
6,961 ppm	0,30	↘	Oleat	semi-quantitative MS	1.750	6,7%
2,805 ppm	0,30	↗	PC aa C42:6	quantitative MS	1.724	13,3%
5,309 ppm	0,30	↗	PC ae C42:1	quantitative MS	1.730	16,2%
0,882 ppm	0,29	↗	PC aa C32:3	quantitative MS	1.733	12,6%
2,749 ppm	0,29	↗	SM C16:1	quantitative MS	1.732	7,2%
4,323 ppm	0,29	↗	PC aa C32:1	quantitative MS	1.712	5,8%
2,748 ppm	0,29	↗	SM(OH) C22:1	quantitative MS	1.737	8,3%
5,324 ppm	0,28	↗	Alpha-tocopherol	semi-quantitative MS	1.726	19,3%
3,387 ppm	0,28	↗	Prolin	semi-quantitative MS	1.736	5,9%
6,961 ppm	0,28	↘	10-Heptadecenoat	semi-quantitative MS	1.746	10,2%
3,690 ppm	0,28	↗	Glycerin	semi-quantitative MS	1.733	8,9%
2,744 ppm	0,28	↗	PC ae C36:2	quantitative MS	1.735	6,0%
6,961 ppm	0,27	↘	Palmitoleat	semi-quantitative MS	1.740	7,5%
2,746 ppm	0,27	↗	SM(OH) C22:2	quantitative MS	1.736	7,8%
1,248 ppm	0,27	↗	PC aa C40:4	quantitative MS	1.711	5,1%
3,871 ppm	0,27	↗	Mannose	semi-quantitative MS	1.735	17,1%
1,284 ppm	0,26	↗	PC aa C32:0	quantitative MS	1.729	6,1%
1,992 ppm	0,26	↗	PC aa C34:1	quantitative MS	1.726	6,7%
2,750 ppm	0,26	↗	SM C24:0	quantitative MS	1.735	8,4%
2,744 ppm	0,26	↗	PC ae C40:3	quantitative MS	1.736	15,1%
2,744 ppm	0,26	↗	PC ae C38:2	quantitative MS	1.730	6,2%

Korrelation zwischen NMR-Bins und Metabolitenkonzentrationen

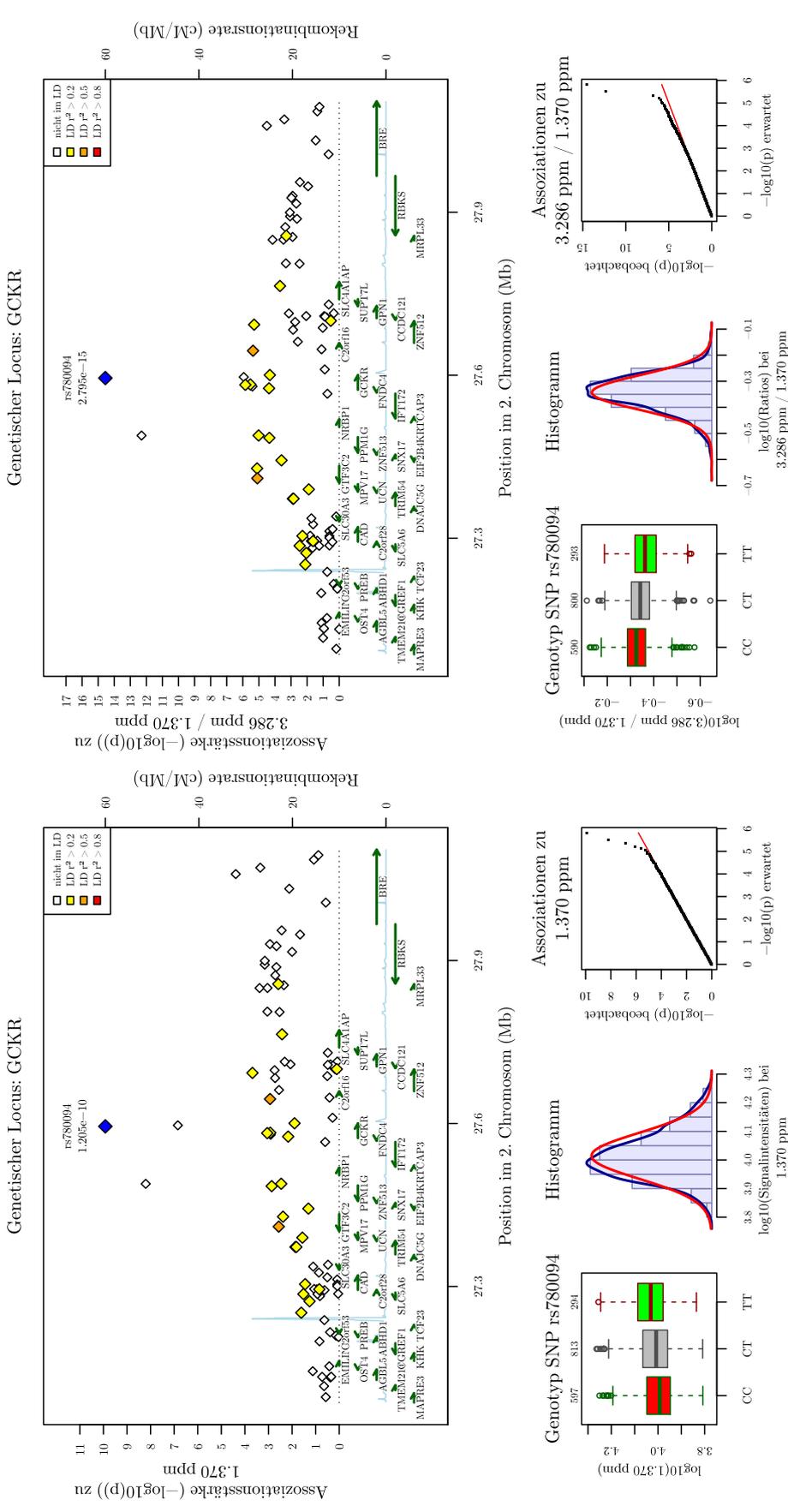
Chemical Shift	r_s^2		Metabolit	Plattform	N	CV
6,960 ppm	0,26	↘	Linoleat	semi-quantitative MS	1.745	6,1%
6,961 ppm	0,26	↘	Palmitat	semi-quantitative MS	1.751	7,8%
5,308 ppm	0,26	↗	PC aa C40:3	quantitative MS	1.719	7,3%
0,883 ppm	0,25	↗	PC ae C42:2	quantitative MS	1.739	8,4%
5,307 ppm	0,25	↗	PC ae C36:5	quantitative MS	1.733	5,7%
2,804 ppm	0,25	↗	PC ae C38:6	quantitative MS	1.731	6,4%
1,283 ppm	0,25	↗	PC aa C30:0	quantitative MS	1.729	6,2%
0,878 ppm	0,25	↗	PC ae C42:3	quantitative MS	1.733	8,6%
0,884 ppm	0,25	↗	PC ae C36:1	quantitative MS	1.729	10,1%
2,748 ppm	0,25	↗	SM C16:0	quantitative MS	1.741	7,2%
2,744 ppm	0,25	↗	PC ae C36:3	quantitative MS	1.732	5,9%
3,264 ppm	0,24	↗	PC ae C34:1	quantitative MS	1.733	6,2%
3,965 ppm	0,24	↘	Pyroglutamin	semi-quantitative MS	1.730	9,3%
6,960 ppm	0,23	↘	Myristoleat	semi-quantitative MS	1.733	7,8%
1,292 ppm	0,23	↗	Inositol-1-Phosphat	semi-quantitative MS	1.357	
2,261 ppm	0,23	↗	1-Oleoylglycerin	semi-quantitative MS	1.653	64,5%
3,297 ppm	0,23	↗	Betain	semi-quantitative MS	1.733	9,0%
6,960 ppm	0,23	↘	Myristat	semi-quantitative MS	1.745	7,1%
0,880 ppm	0,23	↗	PC ae C40:5	quantitative MS	1.732	6,0%
2,748 ppm	0,23	↗	SM(OH) C24:1	quantitative MS	1.730	10,1%
2,808 ppm	0,23	↗	PC ae C40:0	quantitative MS	1.733	3,3%
0,886 ppm	0,22	↗	PC ae C38:3	quantitative MS	1.731	8,6%
6,961 ppm	0,22	↘	Dihomo-linoleat	semi-quantitative MS	1.731	11,1%
3,599 ppm	0,21	↗	Glycin	quantitative MS	1.718	6,9%
2,805 ppm	0,21	↗	PC aa C42:5	quantitative MS	1.718	15,1%
3,263 ppm	0,21	↗	PC ae C42:4	quantitative MS	1.730	6,6%
3,331 ppm	0,21	↗	Stachydrin	semi-quantitative MS	1.725	6,4%
6,925 ppm	0,21	↗	Tyrosin	semi-quantitative MS	1.739	6,1%
1,510 ppm	0,21	↗	Alanin	semi-quantitative MS	1.745	15,1%
6,961 ppm	0,21	↘	Eicosenoat	semi-quantitative MS	1.734	12,9%
3,261 ppm	0,21	↘	Isoleucin	semi-quantitative MS	1.737	6,0%
2,749 ppm	0,20	↗	SM C18:1	quantitative MS	1.736	6,6%
2,746 ppm	0,20	↗	SM(OH) C16:1	quantitative MS	1.740	7,8%

Korrelation zwischen NMR-Bins und Metabolitenkonzentrationen

Chemical Shift	r_s^2		Metabolit	Plattform	N	CV
3,263 ppm	0,20	↗	PC ae C42:5	quantitative MS	1.728	4,7%
2,068 ppm	0,20	↗	1-Linoleoylglycerin	semi-quantitative MS	1.721	75,0%
2,750 ppm	0,20	↗	SM C18:0	quantitative MS	1.736	6,8%
2,746 ppm	0,20	↗	PC ae C40:2	quantitative MS	1.734	7,0%
0,884 ppm	0,20	↗	PC ae C34:0	quantitative MS	1.728	7,6%

Die Spalte „Chemical Shift“ zeigt die Spektralposition der Signalintensitäten an, die am stärksten ($r_s^2 \geq 0,20$) mit der Substanz korrelieren, die in der Spalte „Metabolit“ aufgeführt ist. Ein nach oben gerichteter Pfeil (↗) hinter dem quadrierten Spearman-Korrelationskoeffizienten r_s^2 zeigt, dass die NMR-Signale und die Metabolitenkonzentrationen positiv korreliert sind. Entsprechend zeigt ein nach unten gerichteter Pfeil (↘) Antikorrelation an. Die Anzahl der Proben, die zur Berechnung der Korrelation herangezogen wurden, ist in der Spalte „N“ gegeben. Falls verfügbar gibt „CV“ den Variationskoeffizienten in den zur Qualitätskontrolle gemessenen Proben.

Abbildung B.1.: GWAS mit NMR-Merkmalen im Blutplasma: Regionale Assoziationsplots, Boxplots, Histogramme und QQ-Plots



Die oben abgebildeten Plots gelten für die in Tabelle 5.6 genannten genetischen Assoziationen zu NMR-Signalen (linke Bildhälfte) bzw. NMR-Ratios (rechte Bildhälfte). Oben: Regionaler Assoziationsplot, basierend auf den SNPs, die im für die GWAS verwendeten Datensatz nach der Qualitätskontrolle enthalten waren. Die X-Achse zeigt die Position des SNPs bzw. des Gens, die Y-Achse die negative, dekadisch logarithmierte Assoziationsstärke, den p-Wert, des SNPs zum NMR-Signal oder -Ratio. Die Annotation der Gene sowie die chromosomalen Positionen basieren auf dem NCBI-Referenzgenom (hg18) sowie auf RefSeq (36.3). Die Korrelationskoeffizienten, die das Kopplungsgleichgewicht (*Linkage Disequilibrium*, kurz LD) eines SNPs zum Top-Hit angeben, basieren auf Daten des HapMap-Projektes (Version 27, Phase I-III, CEU-Panel). [72] Unten links: Boxplots der \log_{10} -transformierten NMR-Signalintensitäten respektive der NMR-Ratios, aufgetrennt nach Genotyp der jeweiligen Probanden (von links nach rechts: Homozygote Träger des häufigeren Allels, heterozygote Träger und homozygote Träger des selteneren Allels). Die Anzahl der Probanden in den Gruppen ist oberhalb jeder Box zu finden. Weiterführende Informationen zu den regionalen Assoziationsplot und Boxplots findet sich im Begleittext zu Abbildung 6.2. Unten mittig: Histogramm der NMR-Signalintensitäten oder -Ratios. Die blaue Linie zeigt die Verteilung der \log_{10} -skalierten Daten, die rote Linie zeigt eine Normalverteilung mit gleichem Mittelwert und Standardabweichung, wie sie auch in den \log_{10} -skalierten Daten zu finden sind. Unten rechts: der Quantile-Quantile-Plot (QQ-Plot) zeigt die beobachtete gegen die theoretisch zu erwartende Verteilung der p-Werte der Assoziationen aller gestesteten SNPs zu dem genannten NMR-Merkmal.

Abbildung B.1.: Regionale Assoziationsplots, Boxplots, Histogramme und QQ-Plots (Forts.)

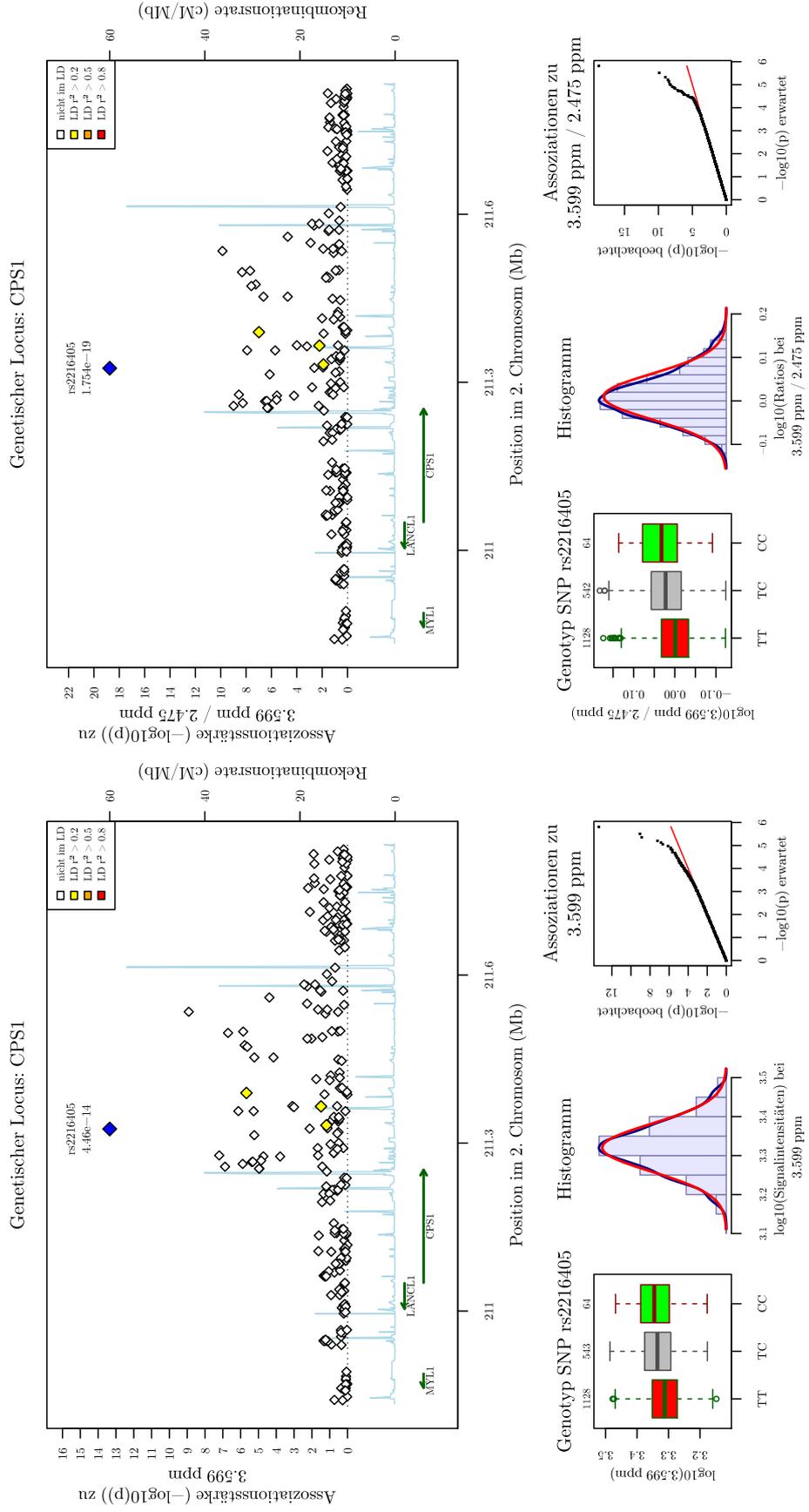


Abbildung B.1.: Regionale Assoziationsplots, Boxplots, Histogramme und QQ-Plots (Forts.)

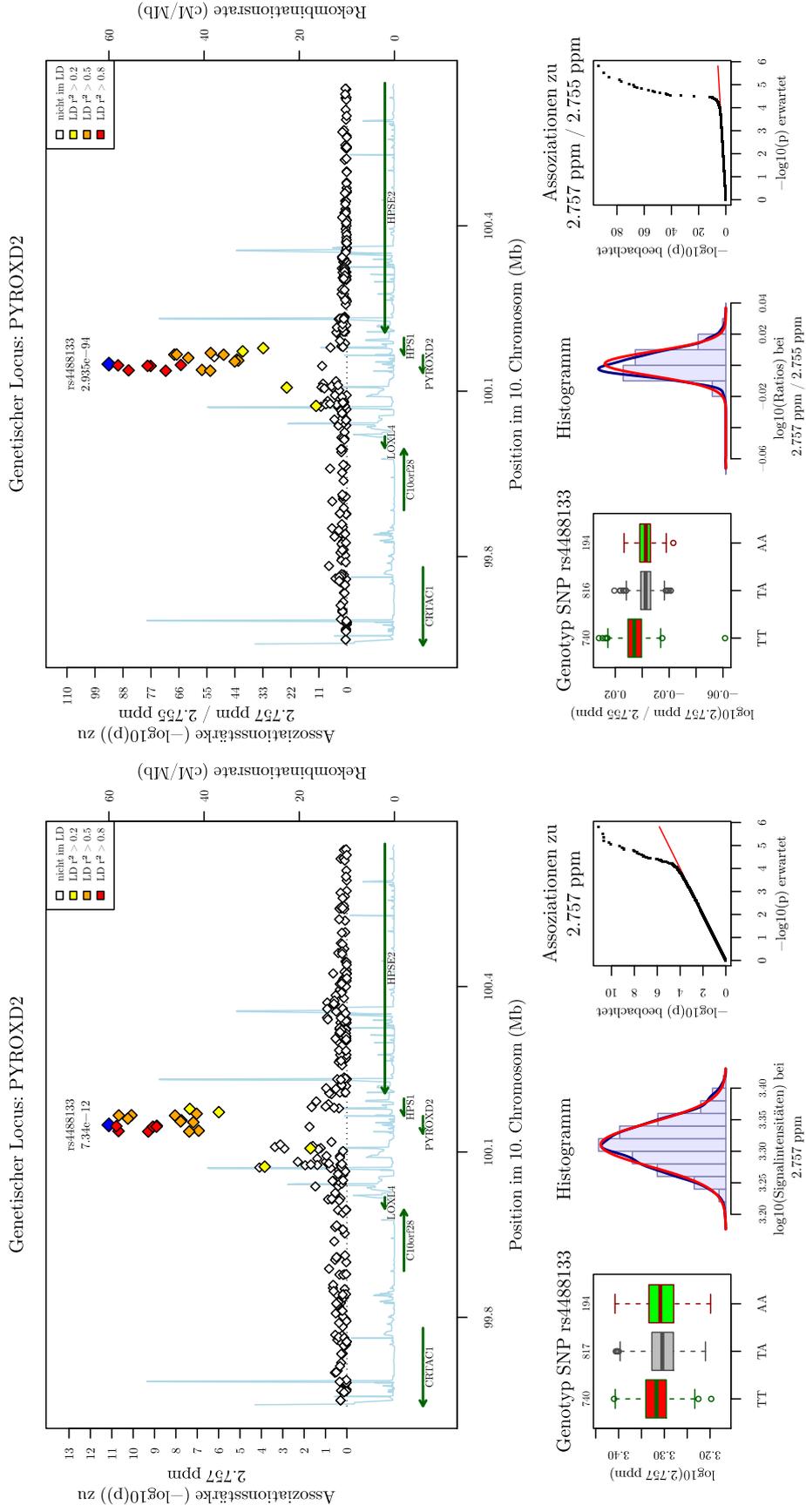


Abbildung B.1.: Regionale Assoziationsplots, Boxplots, Histogramme und QQ-Plots (Forts.)

