



Alliance on Systems Biology

HelmholtzZentrum münchen

German Research Center for Environmental Health



TECHNISCHE
UNIVERSITÄT
MÜNCHEN

Network-based analysis of genetic and nutritional effects on human metabolism

Ferdinand Stückler

2014

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl M12 (Mathematische Modellierung biologischer Systeme)

Network-based analysis of genetic and nutritional effects on human metabolism

Ferdinand Stephan Nepomuk Stückler

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr. M. Hrabě de Angelis

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Dr. F. J. Theis
2. Univ.-Prof. Dr. H.-W. Mewes

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

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während meiner Doktorarbeit begleitet und unterstützt haben:

Zuallererst gebührt mein Dank Fabian Theis für die außerordentliche Betreuung, Förderung und uneingeschränkte Unterstützung, sowie die Möglichkeit, immer etwas neues zu lernen.

Bei Hans-Werner Mewes möchte ich mich bedanken, der mein Interesse an der Bioinformatik und Systembiologie geweckt hat und meine Arbeit stets begleitet hat.

Mein besonderer Dank gilt Jan Krumsiek, der mir schon durch die Betreuung meiner Masterarbeit einen guten Start für die Doktorarbeit ermöglicht hat. Vielen Dank für die tatkräftige Unterstützung bei allen Projekten und die Hilfestellung bei sämtlichen Fragen.

Danke auch an Gabi Kastenmüller für die vielen Stunden intensiver Arbeit an Projekten, die einen großen Teil meiner Doktorarbeit ausmachen. Vielen Dank auch dem gesamten Systems Metabolomics Team für jegliche Hilfe und die fruchtbaren Diskussionen rund um mein Promotionsthema.

Ich möchte meinen Kooperationspartnern Kerstin Ehlers, Helmut Laumen, Susanne Krug, Hans Hauner, Hannelore Daniel, Simone Wahl, Daniela Much, Christian Gieger und Karsten Suhre für die gemeinsame Arbeit an vielen spannenden Projekte danken.

Meinen Zimmerkolleginnen Sabine und Sabrina möchte ich für die vielen gemeinsamen Bürostunden und alle Mathethilfen danken, sowie allen Kollegen und Freunden von CMB und ICB für die tolle Arbeitsatmosphäre, die ich während meiner Doktorarbeit genießen konnte.

Bei der Studienstiftung des deutschen Volkes bedanke ich mich für die finanzielle Unterstützung meiner Doktorarbeit.

Zuletzt geht mein Dank an meine Familie und meine Freunde, die immer für mich da waren und mich unterstützt haben. Vor allem herzlichen Dank an meine Eltern, die mir all dies ermöglicht haben.

Und natürlich vielen Dank Sophia für Deine liebevolle Unterstützung und Dein grenzenloses Verständnis. Danke!

Abstract

Metabolism is highly variable between individuals, even though the underlying key processes follow the same physicochemical laws and biological principles. Metabolic phenotypes are influenced by different genetic and environmental factors such as nutrition. For a better understanding of genetic and nutritional influences on human metabolism, we have analyzed metabolite profiles on different biological scales, ranging from cross-sectional population data over time-resolved in vivo physiological challenging results to in vitro experiments using genetically modified cell lines.

Modern high-throughput methods allow for the simultaneous quantification of hundreds of metabolite levels as readouts for metabolic functions. Yet the analysis and interpretation of the multivariate measurements remains challenging. Biochemical research has provided detailed knowledge about the relationship of metabolites. Statistical methods additionally allow for a data-driven reconstruction of metabolic dependencies. In this thesis we used both resources to build biochemical networks displaying the interplay of metabolites. In a network-based approach we combined this information with experimental data for an improved analysis of genetic and nutritional effects on human metabolism.

We applied this network-based approach first to study the general effects of genetic variation on metabolism. Here we developed a method to select biochemically related metabolites for a given metabolic network. Ratios of metabolite pairs selected in this way were tested for significant associations with single nucleotide polymorphism (SNP). We evaluated this approach both on in silico data derived from simulated reaction networks and data from genome-wide association studies. The network-based ratio method increased the statistical power, lowered computational demands, facilitated the functional characterization of ratio-SNP associations and allowed for the prediction of new associations.

Second, for a better understanding of metabolic phenotypes under specific nutritional and physiological challenges, we studied metabolite profiles of 15 healthy volunteers under fasting conditions. In order to analyze the fasting-induced interindividual variation of metabolite levels, we developed a model of the fatty acid beta-oxidation (FAO) path-

way. Based on the FAO model, volunteer-specific conversion rate parameters were derived from ratios of FAO intermediate metabolite concentrations. Investigating the relationship between phenotypic and metabolic profiles revealed that, compared to absolute metabolite concentrations, metabolite ratios as readouts for the individual metabolic capacity facilitated the characterization of distinct metabolic phenotypes.

Third, we additionally analyzed genotype-dependent effects on fatty acid oxidation pathway dynamics in a human liver cell line using a partial knockdown of the enzyme ACADS. This enzyme plays an important role in the breakdown of short-chain fatty acids. We extended the FAO pathway model in order to describe the knockdown-specific, time-resolved measurements of fatty acid intermediate metabolites. Based on the reaction rates inferred from our model and experimental data we compared the dynamical changes between wild-type and ACADS knockdown conditions statistically. Model parameters showed decreased reaction kinetics for short-chain fatty acids resulting from the ACADS knockdown, and, as a compensatory effect, increased medium- and long-chain fatty acid-related reaction rates.

The combination of established knowledge about metabolic networks and biochemical pathways with computational models facilitated the analysis of multidimensional data for different study designs. As biochemical processes with many coupled reactions can be studied at a system level, the network-based analysis is a promising approach to obtain deeper insights into the interplay between genetic effects, nutrition and metabolism.

Zusammenfassung

Der Metabolismus jedes Menschen beruht auf den gleichen physikalisch-chemischen Gesetzmäßigkeiten und biologischen Prinzipien. Dennoch ist der Stoffwechsel jedes Einzelnen unterschiedlich. Genetische Faktoren und Umwelteinflüsse prägen den metabolischen Phänotyp. Einer dieser prägenden Faktoren ist beispielsweise die Ernährung. Um ein besseres Verständnis über den Einfluss genetischer und ernährungsbedingter Faktoren auf den menschlichen Metabolismus zu erhalten, wurden Metabolitenprofile auf unterschiedlichen biologischen Skalen untersucht. Diese Skala reichte von Populationsstudien über *in vivo*-Studien bis hin zu *in vitro*-Experimenten mit genetisch modifizierten Zellen.

Moderne high throughput Methoden ermöglichen die quantitative Bestimmung hunderter Stoffwechselprodukte als Signal für die zugrunde liegenden metabolischen Prozesse. Die Analyse und Interpretation dieser multivariaten Messwerte gestaltet sich jedoch als komplex. Biochemische Grundlagenforschung liefert detailliertes Wissen über den metabolische Prozesse. Mit modernen statistischen Methoden können zudem datenbasiert die Verbindungen zwischen Metaboliten rekonstruiert werden. Darauf aufbauend wurden in meiner Arbeit biochemische Netzwerke entwickelt, die das Zusammenspiel von Metaboliten des menschlichen Stoffwechsels beschreiben. Methodisch wurde in einem netzwerk-basierten Ansatz diese Beschreibung mit experimentellen Daten verknüpft, um die Analyse zu den Auswirkungen von Genotyp und Ernährung auf den menschlichen Stoffwechsel zu verbessern.

Mit dem netzwerk-basierten Ansatz wurde zunächst der allgemeine Einfluß unterschiedlicher genetischer Ausprägungen auf den Stoffwechsel untersucht. Dazu wurde eine Methode entwickelt, biochemisch relevante Metaboliten an Hand eines metabolischen Netzwerkes zu identifizieren. Auf diese Weise wurden Ratios von Metabolitenpaaren ausgewählt und auf deren Assoziation zu Einzelnukleotid-Polymorphismen (SNP) getestet. Dieser Ansatz wurde einerseits an Hand von Daten aus *in silico* Simulationsreaktionen und andererseits mit Ergebnissen aus genomweiten Assoziationsstudien evaluiert. Als Ergebnis kann festgestellt werden, dass die netzwerk-basierte Ratio-Methode die statistische Power erhöht, die erforderliche Berechnungsdauer verringert, die funktionale

Charakterisierung von Ratio-SNP Assoziationen erleichtert und die Vorhersage neuer möglicher Assoziationen erlaubt.

Um ein vertieftes Verständnis über unterschiedliche metabolische Phänotypen zu erlangen, wurden in einem zweiten Schritt die Metabolitenprofile von 15 gesunden Teilnehmern einer Challenge-Studie im Fastenzustand untersucht. Dazu wurde zunächst ein Modell des oxidativen Abbaus der Fettsäuren (FAO) entwickelt. Auf Basis dieses Modells wurden aus den Ratios der Konzentrationen von Zwischenprodukten Teilnehmer-spezifische Reaktionsraten ermittelt. Bei der Untersuchung der Zusammenhänge von Phänotyp und Metabolitenprofil konnte gezeigt werden, dass im Vergleich zu absoluten Metabolitkonzentrationen die Ratios von Metabolitenpaaren besser die individuelle Stoffwechselaktivität der Teilnehmer widerspiegeln.

Als Drittes wurden Einflüsse des Genotyps auf die Dynamik des oxidativen Abbaus von Fettsäuren in einer humanen Leber-Zelllinie mit Hilfe eines partiellen knockdowns des Enzyms ACADS untersucht. ACADS spielt beim Abbau kurzkettiger Fettsäuren eine wichtige Rolle. Das FAO Modell wurde erweitert und angepasst, um die Zeitverläufe der knockdown-spezifischen Konzentration von Metabolit-Zwischenprodukten zu beschreiben. Mittels der Reaktionsraten, die durch das Modell auf Basis experimenteller Daten geschätzt wurden, konnten die dynamischen Änderungen beim oxidativen Abbau unter Wildtyp- beziehungsweise knockdown-Bedingungen statistisch verglichen werden. An Hand der Modell-Parameter konnte gezeigt werden, dass durch den knockdown von ACADS die Rate für Umwandlungsreaktionen kurzkettiger Fettsäuren verringert wird. Als einen möglichen Kompensationsmechanismus schlug das Modell die Erhöhung der Reaktionsraten für mittel- und langkettige Fettsäuren vor.

Die vorliegende Arbeit zeigt, dass Kenntnisse über biochemische Stoffwechselwege und metabolische Netzwerke in Verbindung mit rechnergestützten Modellen die multidimensionale Analyse der Ergebnisse unterschiedlich angelegter Studien ermöglichen und erleichtern. Beim Studium biochemischer Prozesse und deren zahlreichen Neben- und Folgereaktionen ist diese netzwerk-basierte Analyse eine vielversprechende Methode, um auf molekularer Ebene zu einem vertieften Verständnis über die Zusammenhänge zwischen genetischen Effekten, Ernährung und Stoffwechsel zu gelangen.

Contents

1	Introduction	1
1.1	Metabolism, metabolomics and nutrition	2
1.2	Genome-wide association studies with metabolic traits	9
1.3	Metabolic networks	13
1.4	Model-based analysis of biochemical systems	16
1.5	Research question	24
1.6	Overview of this thesis	25
2	Network-based metabolite ratios (NBRs)	31
2.1	Methods	36
2.2	NBRs on simulated reaction networks	41
2.3	GWAS results in the context of metabolic networks	46
2.4	NBRs on GWAS results	50
2.5	Discussion	58
3	HuMet: A Human challenging study	63
3.1	Methods	66
3.2	Challenge-induced metabolite changes	72

3.3	Modeling mitochondrial beta-oxidation	77
3.4	Association between model readouts and phenotypes	78
3.5	Discussion	81
4	Modeling an <i>in vitro</i> ACADS knockdown	85
4.1	Methods	88
4.2	ACADS knockdown phenotype	95
4.3	Model-based analysis	100
4.4	Extracellular acylcarnitine profiles	104
4.5	Discussion	105
5	Summary and outlook	111

Chapter 1

Introduction

Metabolism is highly variable between individuals [10], even though the underlying key processes follow the same physicochemical laws and biological principles. Metabolic reactions are influenced by all layers of cellular and physiological processes, ranging from genetic regulation over protein modifications to hormone signaling. Moreover, diet, age, life-style, but also disease state, drug treatment or the gut microbiome determine the metabolic phenotype. Despite those internal and external influences, homeostasis can be achieved for most cases. However, there is a fine line between natural benign variation of metabolism [48] and cases when the organism cannot respond in a robust way to perturbations, which may lead to pathophysiological metabolic conditions [97].

For disease diagnostics, chemical compounds in body fluids have been quantified for more than 100 years, pioneered by Garrod's work on inborn errors of metabolism [74]. Recent advances in analytical chemistry allow for detecting large numbers of metabolite molecules, providing a unique fingerprint of the metabolic state. Metabolites resemble the final entities of the biochemical information flow from genes to transcripts to proteins to metabolic reaction compounds. For this reason, metabolic profiling in bio fluids, tissues and cells can be a valuable indicator of an organism's phenotype, as metabolite levels display endpoints of characteristic biological reactions. The upcoming challenge is to analyze this data and interpret the subsequent results at a system level. Biochemical research has provided detailed information about individual enzymatic reactions and

insight into the assembly and interaction between different parts in the network of metabolism, allowing for a network-based analysis of metabolic phenotypes with respect to the interplay of metabolites, enzymes and transport proteins [13, 102].

In this thesis, we will address the question how to integrate the knowledge about biochemical networks with large-scale metabolite data for the investigation of genetic and nutritional influencing factors of metabolism. We will apply this network-based analysis at different biological scales for distinct experimental setups, ranging from cross-sectional population data (Chapter 2) over time-resolved *in vivo* challenging results (Chapter 3) to *in vitro* experiments using genetically modified human liver cells (Chapter 4). In Chapter 2 we describe a network-based metabolite ratio selection method that we developed for an improved analysis of genome-wide association studies with metabolic traits. In Chapter 3 and Chapter 4 we present model-based methods for the investigation of fatty acid metabolism. For each scenario, the key task is to develop appropriate models for the analysis of metabolomics data as the main readout of biological functions, allowing for a better understanding of the complex interplay between genetics, nutrition and metabolism.

Before all of this we will provide in the following a short primer on metabolism, metabolic networks and metabolite measurement techniques. Furthermore an overview about genome-wide association studies on metabolic phenotypes is given. In addition we will briefly describe modeling methods which are used to analyze biochemical systems.

1.1 Metabolism, metabolomics and nutrition

Metabolism defines the set of all biochemical reactions that take place in a living organism [19]. In a linked series of metabolic reactions, substrate metabolites are converted into product compounds. The collection of interdependent biochemical steps, which include all reactions needed for fulfilling a specific metabolic function, can be defined as *biochemical pathways* [167]. These building blocks of metabolism are essential for sustaining life and can be divided into energy producing (*catabolic*) and energy requiring (*anabolic*) pathways. For instance, in the catabolic pathway glycolysis the cell yields energy by converting glucose into pyruvate in a sequence of interconnected reaction steps. Anabolic reactions then use this energy from catabolic breakdown of fuel compounds for

synthesizing complex molecules or macromolecules. A *metabolic network* then represents the collection of all biochemical pathways and their individual interactions [90].

Metabolomics: measuring the molecular fingerprint of metabolism

Recent advantages in analytical chemistry now allow for measuring the molecular fingerprint of metabolism, also referred to as *metabolomics* [180, 245]. In this large-scale and system-wide approach one tries to analyze ideally the entire metabolite pool in biological samples such as cells, tissues or body fluids. The term *metabolite* is used to describe small molecules which are intermediates or products of metabolic reactions, with usually molecular weights of 2000 Da or below [267]. The collection of all endogenous metabolites as a whole, also referred to as the *metabolome*, consists for instance of amino acids, lipids, carbohydrates, but also of hormones, vitamins, nutritional organic compounds and drugs [22, 176]. Yet a clear definition of the human metabolome is not straightforward. Human metabolomics measurements will be a mixture of endobiotic and xenobiotic metabolites, for instance due to exogenous metabolite sources like nutrition or bacteria which metabolize food compounds. Also commonly used biofluid samples like blood or urine will not cover all metabolic compounds from the individual metabolism of organs and body compartments. Up to now, the Human Metabolome Database (HMDB) [267] lists more than 5,000 detected metabolites and 20,000 endogenous and exogenous compounds, which are expected to be found in human biological samples. This number will increase within the next years due to better analytical techniques and specific experimental setups, but also by considering metabolites derived from exogenous sources like nutrition, drug treatment and microbiome [82, 111].

The set of measured small molecules for a specific condition in individual samples is often termed *metabolic profile* [186]. This metabolite composition is assumed to be characteristic for a specific biological state, as it results from biochemical reactions which are closely linked to distinct physiological states [173]. Metabolite levels in bio fluids, tissues and cells display endpoints of characteristic biological processes. Metabolic intermediates resemble the final entities of the biochemical information flow from genes to transcripts to proteins to metabolic reaction compounds (see Figure 1.1). Metabolomics thus reveals a detectable link between the observed phenotype and the determining

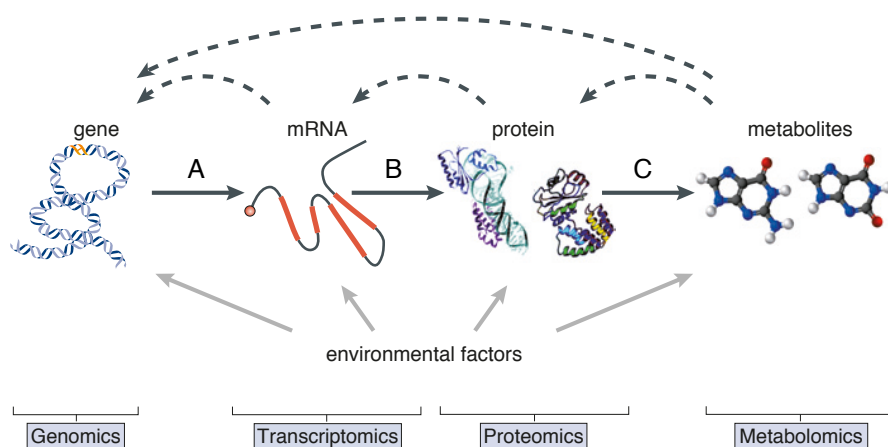


Figure 1.1: The flow of biochemical information. Genes are transcribed to messenger RNA (mRNA) molecules (**A**), which are then translated to proteins (**B**). Transporter molecules and enzymes in metabolic pathways influence in turn the levels of metabolites (**C**). Dashed arrows represent possible feedback mechanism. For instance, metabolites function as inhibitors of metabolic enzymes, alter gene functions by directly interacting with transcription factors or induce epigenetic changes. Gray arrows denote the effect of environmental factors like nutrition, life style, health state or environmental exposure. Blue boxes display large-scale approaches for measuring the respective biomolecules (Figure adapted from [184]).

genotype [69]. For instance, nowadays metabolic profiling is a standard procedure in newborn screenings for the diagnosis of inherited diseases [145]. Metabolomics measurements summarize the impact of endogenous influences like genetic variation, epigenetic and regulatory effects, transporter protein and enzyme concentrations, but also of extrinsic or environmental factors such as nutrition, exposure to harmful substances or drug treatments [87]. Therefore, metabolite profiles display an integrated response to environmental and genetic factors [174]. Compared to other large-scale approaches like proteomics and transcriptomics, metabolomics can reveal more direct information about the investigated biologic state and function in a system-wide fashion [5, 83]. Thus the concept of metabolic phenotypes or *metabotypes* has been defined as a “probabilistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids or tissues” [75].

In addition, the substantial knowledge about biochemical pathways facilitates the interpretation of metabolomics results. As experimental techniques now allow for large-scale quantification of metabolites as intermediate phenotypes, the upcoming challenge will

be how to integrate existing fundamental knowledge about biochemical reactions and metabolic networks in order to better understand the involved biological functions and their contribution to observed phenotypes.

There are two methodologically distinct approaches for the system-level analysis of metabolite profiles, namely *targeted* and *untargeted metabolomics* [58]. The first approach quantifies a defined set of metabolites, for instance in a specific biochemical pathway, with high accuracy and degree of resolution. We used this approach for measuring selected fatty acid compounds in two studies about nutritional, physiological and genetic effects on the mitochondrial breakdown of fatty acids (see Chapter 3 and Chapter 4). A more global view on metabolism offers untargeted metabolomics. The aim of this approach is to measure “as many metabolites as possible from biological samples without bias” [184]. This global metabolomics data acquisition method was applied for example to measure human plasma metabolite levels in a population study on genotype-dependent metabolic variation (see Chapter 2). Often metabolite abundances for untargeted approaches can only be given as relative numbers, whereas targeted profiling yields metabolite concentrations. For both approaches data measurement is performed using either nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS), often coupled with a preceding separation step using liquid or gas chromatography [147, 212]. The application of both modern high-throughput bioanalytical techniques allows now for the investigation of system-wide metabolite panels for a large number of samples.

Nutrition and metabolism

The human metabolism is influenced by the complex interplay of various endogenous and exogenous factors [79]. For each individual there is a substantial variation from the average metabolome [282]. Age, gender and genotype for instance contribute intrinsically to individual differences in metabolic profiles [104, 279]. Inherited molecular defects in catalyzing enzymes can cause metabolism-related disorders. Besides, bacterial gut microflora composition, physical activity or drug treatment factors influence the human metabolism extrinsically [38, 244, 264]. During the entire course of life, environmental exposures and lifestyle shape individual-specific and condition-dependent metabolic phenotypes or metabotypes [206, 266]. A major determining factor is nutrition. Malnu-

trition can have severe effects on metabolic functions. For instance, the lack of vitamin C leads to impaired collagen synthesis in scurvy disease [63]. Unbalanced diet and excessive consumption of nutrients can cause metabolic disorders [207]. The easy access to energy-dense food has led to an increased incidence and prevalence of obesity and diabetes in developed countries. In order to study the effects of diet and food compounds, nutritional challenging studies have been conducted to assess the inter-individual variation in metabolic phenotypes of healthy and diseased subjects [52, 189, 228, 284].

Food compounds and metabolites directly interact with biochemical processes (see Figure 1.1); by sensing the amount of glucose, the body regulates the production of insulin. This hormone of the endocrine system plays a crucial role in the regulation of anabolic reactions for maintaining balanced blood sugar levels. Moreover, nutrients can alter gene functions by directly interacting with transcription factors or by causing epigenetic changes of DNA structure [282]. Studies using animal models showed nutritional and tissue-dependent regulation of obesity risk genes under fat feeding and fasting conditions [277]. As modern bioanalytic methods allow for taking a snapshot of the individual metabolic state, metabolomics is a key tool to understand influencing factors of metabolism. Recent studies have analyzed nutrient-gene interactions in humans by combining metabolomics with genotyping and epigenetic profiling [191, 237]. The complex interplay between genetics, aging, pathophysiology, microbiome, environmental exposure and nutrition contributes to metabolic intra- and interindividual differences. Understanding the underlying mechanisms will allow for elucidating genetic and environmental effects contributing to metabolic disorders and might lead to improved personalized treatment suggestions [111].

Fatty acid metabolism

Fat is an efficient energy source and the major fuel reserve in the human body [59]. Fat molecules belong to the group of lipids which also include for instance sterols, phospholipids, monoglycerides, diglycerides, triglycerides and fatty acids. Lipids have different biochemical functions. First, they are used in the form of triglycerides to build up a long-term energy reservoir in fat cells and adipose tissue, providing an efficient and highly concentrated way of storing metabolic energy. Triglycerides consist of a single glycerol

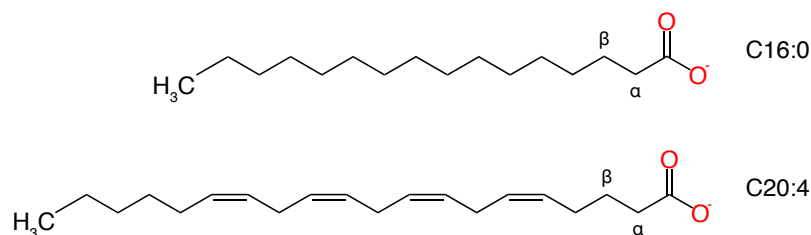


Figure 1.2: Palmitic acid (C16:0) and arachidonic acid (C20:4) as an example for the structure of saturated and unsaturated fatty acids. Fatty acid carbon atoms are numbered starting at the carboxyl terminus (red). Carbon atoms next to the carboxyl terminus are denoted by α and β . During fatty acid beta-oxidation, the acyl chain is shortened between the α and β carbon atom.

molecule which is esterified with three fatty acids. Besides being fuel molecules for providing energy to anabolic processes, lipids are building blocks of biological membranes [19], serve as tagging or modification factors of proteins or as hormones and intracellular messengers in biological signal transduction [30, 204, 247].

During fasting and sustained exercise the catabolic breakdown of fatty acids (FA) in the mitochondrial *beta-oxidation* pathway is a major energy source for cellular organisms. In Chapter 3 and 4 we describe mathematical representations for this degradation pathway and how we used these models to analyze metabolomics data from nutritional challenging studies. As a primer we will therefore introduce briefly the biochemical properties of this major physiological process.

Fatty acids are hydrophobic lipid molecules consisting of long hydrocarbon or alkyl chains with terminal carboxylate groups (see Figure 1.2). Alkyl chains differ in the number of carbon atoms and degree of saturation. Saturated fatty acids contain no double bonds, while unsaturated fatty acids have one or more double bonds. A short notation for fatty acids with n carbon atoms and d double bonds is $Cn:d$. For instance, palmitic acid (C16:0) is an abundant fatty acid in the human body with 16 carbon chain atoms. It is also quite common to omit the number of double bonds for saturated fatty acids in the notation, i.e. C16 instead of C16:0.

Dietary fatty acid molecules are mostly ingested in the form of triglycerides. When not degraded immediately, these molecules are packed into lipid transport particles and located to adipose tissue for storage [94]. During physiological conditions like fasting,

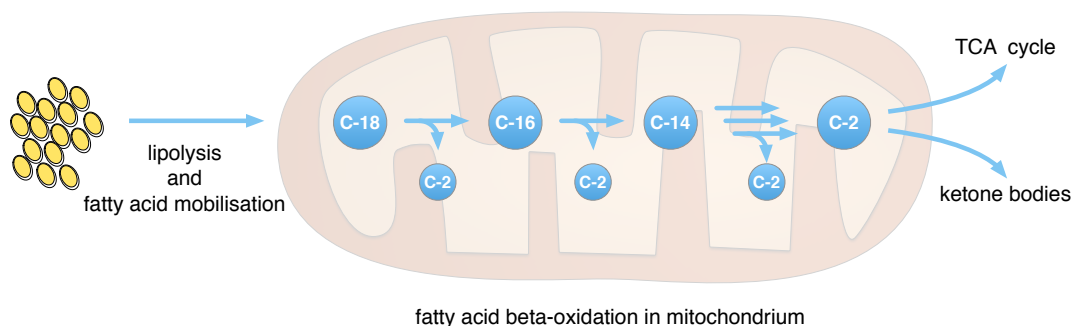


Figure 1.3: Schematic representation of fatty acid beta-oxidation. Mobilized fatty acids from adipose tissue are subsequently degraded in a linear pathway. During each reaction step the carbon chain is shortened by two carbon atoms. Resulting acetyl-CoA (C-2) is further used in the tricarboxylic acid (TCA) cycle for energy production or consumed during formation of ketone bodies.

catabolic stress and sustained exercise when sugar-based fuel molecules have been exhausted already, the lipid energy reserves are mobilized. Free fatty acids are transported to peripheral tissues like muscle or liver. The energy producing breakdown of the fatty acid carbon chains mainly occurs within the mitochondria of cells, for which a carnitine shuttle system is needed. First enzymes called carnitine acyltransferases catalyze the conversion of activated fatty acids (acyl-CoAs) to acylcarnitines. These transport molecules are then translocated to the inner mitochondrial matrix [116], where acylcarnitines are converted back to acyl-CoAs. This active channeling allows for the modulation and regulation of general beta-oxidation activity [59].

Within the mitochondria the fatty acids are degraded in a recurring sequence of four enzymatic steps: oxidation, hydration, a second oxidation and thiolysis. In each reaction cycle the alkyl chain is shortened by two carbon atoms, leading to the production of acetyl-CoA, molecules with high energy potential (see Figure 1.3). The first oxidation step is catalyzed by four *acyl-CoA dehydrogenase* enzymes with different specificities for short (C4 and C6, ACADS), medium (C4 to C12, ACADM), long (C8 to C20, ACADL) and very long carbon chain lengths (C12 to C24, ACADVL) [16]. The sequential shortening by two carbon atoms generates acetyl-CoA, which enters the tricarboxylic acid (TCA) cycle for the production of energy transfer molecules adenosine triphosphate (ATP). ATP provides energy for physiological and cellular processes such as growth, cell division or

anabolic production of biomolecules. Fatty acid synthesis is essentially the reverse process to oxidation in basic chemical reactions, but regulated in an antagonistic fashion. The production of fatty acids is for instance used to store nutritional energy, but also for synthesizing cell signaling molecules [85].

Inherited molecular defects in fatty acid catalyzing enzymes can cause beta-oxidation related disorders. The resulting impaired breakdown of fatty acids and disruption of energy homeostasis can lead to symptoms muscle weakness, hypotonia or heart failure [260]. For instance short-chain acyl-coenzyme A dehydrogenase deficiency (SCADD), a rare fatty acid oxidation disorder, can result from alterations in the ACADS gene [108, 169]. The impaired fatty acid beta-oxidation activity due to SCADD leads to an accumulation of byproducts of fatty acid metabolites. Increased levels of butyrylcarnitine and ethylmalonic acid in plasma and urine are therefore biomarkers for the diagnosis of SCADD [8, 20]. In Chapter 4 we will investigate impaired ACADS function in an *in vitro* knock-down using a mathematical model of the beta-oxidation pathway. Other inherited disorders of fatty acid metabolism are related to impaired transport of long-chain fatty acid into mitochondria due to defects in enzymes of the carnitine shuttle [214].

Imbalanced lipid metabolism in general is a risk factor for the pathogenesis of systemic metabolic disorders. The accumulation of fat contributes, for instance, to the development of insulin resistance and cardiovascular diseases in obese patients [131, 207]. Though detailed knowledge about molecular malfunctions has been established recently, the complex interactions between genetics, life-style and clinical phenotypes have not been fully explained yet [100]. Within the last years genome-wide association studies with metabolic traits have been carried out to address especially the genetic part of this question.

1.2 Genome-wide association studies with metabolic traits

Genome-wide association studies (GWAS) analyze the genetic contribution to an observed phenotype on a population scale [54]. GWAS have been applied initially to identify genetic factors which influence the susceptibility and etiology of common diseases [262].

On a population level, associations between variants of genes and specific phenotypes or quantitative traits are studied. Single nucleotide changes within the DNA sequence of a gene (*single nucleotide polymorphism*, SNP), differences in gene copy numbers, but also structural changes in the DNA sequence lead to genetic variation between individuals. Traits for GWAS can be for instance blood pressure, weight, frequency of diseases, but also molecular data about levels of protein, metabolites or transcripts [1]. In contrast to gene candidate-driven studies, GWAS aim at a hypothesis-free investigation of the entire genome [239]. To this end, the genetic variation in a population of many individuals (usually a few thousands) is characterized using SNP arrays. For case-control settings, participants are divided into two groups (for instance healthy and disease). A genetic variant which is more frequent in one group than expected then is considered to be associated with the phenotype of interest. Also quantitative traits as continuous variables (for instance weight or blood pressure) can be correlated with genetic polymorphism [93].

In recent years GWAS have provided new discoveries regarding genes and pathways which play a critical role in common diseases [258]. Despite the overall success of GWAS, there are several limitations which need to be kept in mind. The initial assignment of sequence positions for SNP-genotyping can introduce a bias to specific genetic regions, as for instance the causal variants cannot be measured [42]. Yet with declining DNA sequencing costs, exome sequencing [36, 171] or even whole genome sequencing [41] will provide detailed genetic information, but also imposing new statistical challenges for association tests [12]. Population structure or stratification due to the presence of subgroups from different ancestries within the study group can also introduce substantial bias [93] and several methods have been developed to correct for these effects [196]. Besides, many identified genetic risk loci only explain a small proportion of the observed phenotypic variation [151]. Especially for diseases such as cancer or diabetes, the complex interplay of many genetic, but also non-hereditary factors leads to the onset and progression of the disease [273]. A clinical classification of multifactorial diseases with heterogeneous symptoms and conditions is also not straightforward, for instance in the case of mental disorders [154]. When performing a GWAS analysis for a specific disease one might in reality investigate several unrelated diseases with similar symptoms.

Moreover, it is still under debate whether rare alleles, common alleles or a combination of both contribute to the phenotypes of common diseases [258]. Several rare alleles with

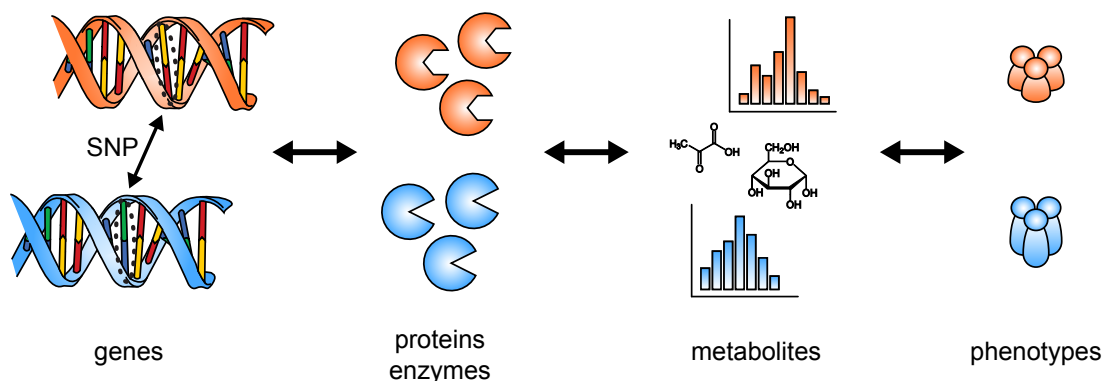


Figure 1.4: Genetic variation (single nucleotide polymorphism, SNP) in genes coding for metabolic enzymes and transporter molecules lead to differences in metabolite levels (metabotypes). Metabolic profiles as endophenotypes are often more related to the observed (exo-)phenotypes, allowing for a better understanding of the influence of genetic factors on biological mechanism.

high effects could increase the susceptibility to diseases, but on the other hand also the combination of many common alleles with only minor effects [224]. In addition, risk markers which are located in intergenic or intronic regions pose a challenge for the mechanistic interpretation of the findings [92]. Often these markers are not the causal genes or variants, but will be credited in a GWAS setting with common alleles to the signal of rare, but unmeasured variants, which in reality cause the phenotype. Such synthetic associations [55] will then complicate the functional interpretation of risk-associated SNPs. Until the detection of these rare variants by whole-genome sequencing is affordable for population studies, subgroups with medium to high effect sizes could resemble potential sequencing candidates, for instance members of families with many affected individuals [209]. Similarly, only extreme cases of a specific phenotype (e.g. very high and very low blood pressure) may be selected for sequencing, also referred to as extreme-trait sequencing [15].

GWAS have identified risk associated SNPs for specific diseases. A functional interpretation of the results remains challenging and only few associations have been functionally validated [80]. Effects which cause a specific phenotype are often a mixture of many underlying processes. For a better understanding of disease mechanisms, further insights into the underlying biological processes and the role of altered functions in disease related genes is required [7]. For this reason, recent GWAS approaches linked genetic variation with quantitative molecular phenotypes. These *quantitative trait loci* (QTL) analyses

included gene expression profiles (eQTL) or metabolite concentrations (mQTL). Studies in plants or fungi for instance linked gene expression and metabolic traits in order to explain the molecular basis of physiologic phenotypes [118, 126, 263, 285]. The eQTL analysis of human transcription profiles revealed genetic factors which underpin individual differences in gene expression levels [43].

For the investigation of biochemical processes, metabolite levels can be used as quantitative GWAS traits. As described in Section 1.1, metabolic profiles can be seen as readouts of intermediate processes which are related to the underlying biological mechanism (see Figure 1.4). Inborn variation of metabolism often leads to altered individual levels of metabolites as reaction-related enzymes are not functional or missing [166]. Human GWAS studies with metabolic traits (mGWAS) analyzed genetic variants which explain variation in metabolite concentrations also referred to as *genetically-influenced metabolotypes* (GIM) [236]. Several mGWAS studies [81, 98, 104, 106, 117, 172, 230] revealed direct associations between metabolic traits and genetic variants located near to genes encoding metabolite-specific enzymes or transporters. For example, a single-nucleotide polymorphism in the N-acetyltransferase 8 (NAT8) locus was reported to associate with N-acetylornithine [237]. Variants in the *short chain acyl-coenzyme A dehydrogenase* (ACADS) gene locus were found to be associated to levels of butyrylcarnitine, the transport form of the fatty acid beta-oxidation product butyryl-CoA [104]. ACADS, which catalyzes the first and committing reaction step in the degradation cycle of fatty acids, is a key enzyme in the beta-oxidation pathway (see Section 1.1) with high substrate specificity for short carbon chain acyl-CoA molecules like butyryl-CoA [77]. Interestingly, butyrylcarnitine is also an established biomarker for short-chain acyl-CoA dehydrogenase deficiency in newborn screenings [169]. Linking mQTL information with established genetic risk loci can therefore allow for a better understanding of the pathophysiology of diseases [1, 2].

Using metabolite concentration ratios as quantitative traits in addition to single metabolite concentrations further improved the results and interpretation of SNP-metabolite associations. It was shown that ratios between metabolite concentrations pairs reduced the overall biological variability in population data and resulted in robust statistical associations [190, 263]. For instance, the level of a nutritional metabolite, but also the respective breakdown products, might be elevated in specific subjects. The ratio of the

two metabolites accounts for this interindividual variation. In a biochemical interpretation, the ratio between product-substrate metabolite pairs can be interpreted as a proxy of the corresponding enzymatic reaction rate [236]. For example, Suhre *et al.* [237] reported that the association of a genetic variant in the FADS1 locus and the ratio between fatty acids 20:3 and 20:4 is much stronger compared to the association with the respective single metabolite levels. The FADS1 locus encodes for a fatty acid delta-5 desaturase with fatty acids 20:3 and 20:4 as substrate and product, respectively. The increase in association strength due to the ratio between reaction substrate-product pairs thus matches the biological function of the enzyme [139].

1.3 Metabolic networks

Individual reactions have been studied and characterized in great detail in the field of biochemistry. This large knowledge is now collected in reaction databases, which also connect the intermediate steps trying to give an overall picture of metabolism. In the following we will give a short overview about different metabolic pathway resources. We will also address methods for data-driven reconstruction of biochemical networks based on large-scale metabolomics measurements, which yields knowledge about metabolite dependencies without the need for database information.

Databases for biochemical pathways and metabolic networks

A variety of databases exists for the description of metabolic reactions. They differ with respect to data assembly, pathway definition and coverage of genes, enzymes and metabolites. Databases such as KEGG [114], MetaCyc [31] or Reactome [47] combine the knowledge about individual reactions in order to reveal the connectivity of metabolic pathways. In addition, resources such as BRENDA [223] or SABIO-RK [268] offer detailed information about reaction-specific kinetic parameters found in primary literature. Often these parameters have been determined for *in vitro* settings or are highly dependent on the experimental setup.

Instead of collecting all known reactions, recent approaches try to reconstruct metabolic networks based on genetic information [241]. These networks are based on metabolism-related genome annotations such as presence of metabolic enzymes and transporter proteins. Often several iterations of manual curation using primary literature, textbooks, review articles and validation experiments are needed to assure the reconstruction quality. Databases for manually reconstructed human metabolic networks are for instance EHMN [149], BiGG [219] or Recon 1 [57] and 2 [243].

Comparing several databases revealed that only parts of them are consistent with respect to content overlap. The general level of agreement is quite low due to different ways of generating, curating and interpreting specific reaction entries [234, 235]. Therefore efforts have been made to combine the heterogeneous information in order to construct consensus models of organism-specific metabolism aiming for a complete characterization of all metabolic processes [242, 243]. Here the description of organism-, tissue- and cell type-specific reactions is critical for understanding physiological effects in a multicellular organism. While the amount and quality of annotations for biochemical processes have increased substantially within the last years, all databases still contain many potential wrong or missing annotations. The collected knowledge also shows a bias towards common pathways [234]. As the major biochemical reactions have been studied for a long time, they are also represented in more detail. These issues need to be kept in mind when using metabolic pathway database information for the analysis of experimental data and the interpretation of knowledge-based results.

Data-driven reconstruction of metabolic networks

Many measured metabolites have not been annotated in the above-mentioned pathway databases. In addition, untargeted metabolomics approaches allow for the reliable detection and quantification of compounds that have not been characterized chemically [135]. Here, data-driven reconstruction methods can complement the knowledge-based analysis and interpretation of large-scale metabolomics data. Various inference approaches on genome-scale have been developed in order to reconstruct gene regulatory networks based on high-throughput expression data [152]. Similarly - based on large-scale metabolomics data - the wiring of the underlying metabolic network can be predicted [261]. Repeated

metabolomics measurements show a considerable amount of biological variation due to stochastic fluctuations of metabolite concentrations and individual variability in reaction properties such as kinetic rates or enzyme and transporter concentrations [29, 233]. This can be observed for various biological scales ranging from samples of cellular experiments to data from cross-sectional human population studies such as blood plasma metabolite profiles. Especially for the latter case also external sources like current metabolic state, nutrition supply and environmental factors contribute to the observed variability.

Correlation-based methods provide a straightforward way to obtain pairwise associations between metabolites given substantial variation in the biological samples [232]. Especially partial correlations based on *Gaussian graphical models* (GGM) have been shown to be a valuable tool for the unbiased reconstruction of metabolic reactions from large-scale human blood serum metabolomics data [136]. Metabolites within metabolic pathways show high correlations, even if they are not part of the same biochemical reactions. For example in the fatty acid beta-oxidation cascade (see Section 1.1), the import of new fatty acids into the mitochondrion will lead to an increase of all intermediate metabolite concentrations as well. This results in indirect associations between metabolites which are not directly connected (see Figure 1.5). The GGM approach tries to remove such indirect effects. For calculating the partial pairwise correlation between two metabolites in GGMs, indirect effects are removed by conditioning against the associations with all remaining variables [218]. GGMs are based on full-order partial correlation coefficients. To remove indirect associations each pairwise correlation is corrected against all remaining variables. GGMs for data with more samples than variables can be calculated directly by a matrix inversion operation of the covariance matrix and a normalization step [50, 140]. For data with less samples than variables regularized versions of partial correlation coefficients can be estimated, for instance using the R-package GeneNet [181]. This method yields also for cases with more samples than variables robust estimates of partial correlation coefficients.

Removing indirect connections in the reconstructed metabolic networks facilitates the interpretation in terms of the underlying biochemical reactions. It needs to be noted that any correlation-based association between two factors does not imply direct causation. Moreover, only a subset of all confounders in biochemical systems will be measured. If the confounding variables have not been quantified, also partial correlation methods

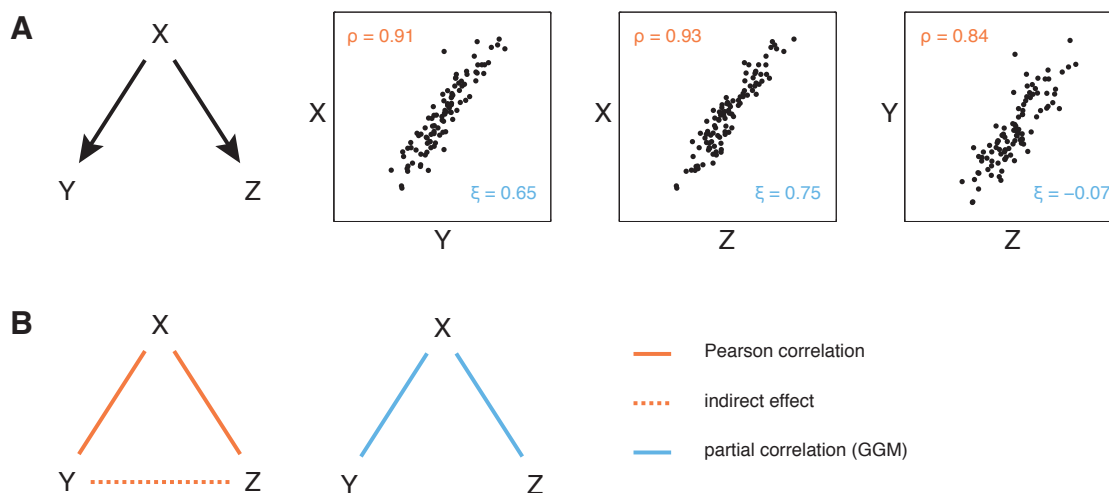


Figure 1.5: Example of indirect effects for correlation-based reconstruction of metabolic networks. **A:** Compound X affects both Y and Z. Therefore also Y and Z are highly correlated. **B:** Data-driven reconstruction of the network structure using Pearson correlation (orange) and partial correlation (Gaussian graphical model, GGM, blue). Pearson correlation fails to detect the indirect association between Y and Z (dashed line). In GGMs, indirect effects between Y and Z are removed by conditioning against the association with X. ρ and ξ : Pearson and partial correlation coefficients between the respective compounds.

cannot correct for indirect effects. For confounders with non-gaussian properties other methods than GGM can be applied [229].

1.4 Model-based analysis of biochemical systems

Knowing the interrelation of compounds within metabolic pathways, as described in Section 1.3, provides the basis for modeling the underlying biochemical reaction networks. In computational and systems biology, mathematical models are used as an abstract representation of complex processes. In a mathematical framework, these models summarize established knowledge about biological systems and describe the relationship between molecular compounds that are involved. This systems biology framework allows to examine the structure and dynamics of biological function by considering not only the characteristics of isolated parts of the system, but the system as a whole. The systemic

view helps to unravel the internal nature and dynamics of biochemical reactions and to predict the behavior for simulated conditions [122, 123].

The following sections give a short introduction to modeling methods for biochemical systems. First we will give an overview about methods which analyze the structural properties of metabolic networks without having knowledge about kinetic details. We will then explain how ordinary differential equations (ODEs) can be used as a tool for the description of biochemical reaction dynamics.

Constraint-based structural analysis of metabolic networks

Often only limited kinetic parameters or metabolite and enzyme concentrations are available for biochemical pathways, but the general structure (e.g. stoichiometry and connection between compounds) is known. In a metabolic network consisting of n compounds and m reactions, the stoichiometric coefficients are summarized in the *stoichiometry matrix*

$$S := \{s_{ij}\} \quad \text{for } i = 1, \dots, n \quad \text{and} \quad j = 1, \dots, m. \quad (1.1)$$

Each column of S represents one reaction and each row one compound. Negative stoichiometric coefficients in one column denote educts and positive coefficients products of one specific reaction. Compounds which do not take part in the specific reactions have zero entries in this column. The time evolution in the network is described by S and a reaction flux vector $v(x)$, with x being the concentration state vector for all compounds $x_1, \dots, x_n$.

Based on the structure of the metabolic network a *steady-state* analysis can be performed [123]. Here a constant equilibrium of metabolites is assumed ($Sv = 0$), i.e. the change of compound concentrations over time remains zero [142]. While the actual concentrations do not change, there is still a metabolic *flux* through the system. Flux here denotes the number of molecules flowing through each reaction per unit of time. This assumption generally holds for metabolite levels, as biochemical reactions reach equilibrium much faster compared to the time scales of upstream regulatory processes or external reactions which are connected to the system [143]. The stoichiometry of each

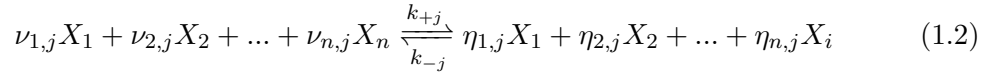
reaction, i.e. how many quantities of each compound take part in the reaction, already impose constraints on the flow of metabolites through the system [182]. For instance, due to mass conservation, the total number of any metabolite being consumed must be equal the total amount being produced under steady-state conditions. Additional constraints such as maximal biomass production or enzyme capacity using lower and upper bounds for specific reaction fluxes can be incorporated in the model as well [197]. Points in the flux solution space, which satisfy all constraints, are obtained from the topology of the metabolic network using extreme pathway [220] or elementary flux mode analysis [225]. Constraint-based analysis of metabolic networks has been applied successfully for microorganism in metabolic engineering and drug discovery studies [24, 115]. Several methods have been developed to combine metabolomics data and large-scale reaction networks for refining the constraint-based solution space. These approaches focus on fundamental properties of structure and dynamics in metabolic networks to fill the gap between system-wide, constraint-based models and detailed kinetic description of specific metabolic pathways [64, 107, 203].

Dynamic modeling using ordinary differential equations

The rate of a biochemical reaction, i.e. the change of its substrate and product concentrations per time, is determined by different factors, e.g. the amount of catalyzing enzymes, concentration of substrates or the presence of activating or inhibiting modifiers. Also biophysical properties, like the number of chemical steps required to convert substrate to product molecules or the molecular substructures of the catalytic center, may alter the reaction speed. The relationship between all mentioned factors and their respective impact on the reaction rate is described by *kinetic rate laws*.

A common way of describing biochemical reactions is the law of *mass action*, formulated by Guldberg and Waage in the nineteenth century [88]. The law states that the reaction rate is proportional to the collision probability of the reactants. This probability in turn is proportional to the concentration of the reaction compounds to the power of the *stoichiometric coefficients*, i.e. how many molecules of one compound are part of the reaction.

For the most general reaction R_j



or

$$\sum_{i=1}^n \nu_{i,j}X_i \xrightleftharpoons[k_{-j}]{k_{+j}} \sum_{i=1}^n \eta_{i,j}X_i \quad (1.3)$$

with $\nu_{i,j}$ and $\eta_{i,j}$ being the stoichiometric coefficients and $k = k_{+1}, k_{-1}, \dots, k_{+m}, k_{-m}$ reaction parameters or rate constants, the reaction flux assuming mass action kinetics reads

$$v_j(x) = k_{+j} \prod_{i=1}^n x_i^{\nu_{i,j}} - k_{-j} \prod_{i=1}^n x_i^{\eta_{i,j}}, \quad (1.4)$$

in which x_i is the concentration of X_i . Using the stoichiometric matrix (see Equation (1.1))

$$S_{ij} = \nu_{i,j} - \eta_{i,j} \quad (1.5)$$

the reaction rate equation or *ordinary differential equation* (ODEs) model reads

$$\frac{dx}{dt} = \dot{x} = Sv(x). \quad (1.6)$$

For the irreversible reaction



as an example with compound concentrations x_A, x_B, x_C the change of compound C over time then reads according to the law of mass action

$$v_C = kx_A^2x_B, \quad (1.8)$$

with rate constant k as a proportionality factor. The stoichiometric coefficients of x_A and x_B are 2 and 1, respectively.

The dynamics of metabolic reactions can be modeled in general by a set of *ordinary differential equations* (ODEs)

$$\frac{dx_i}{dt}(t) = \dot{x}_i(t) = f_i(x_1(t), \dots, x_n(t), \theta, t), \quad i = 1, \dots, n. \quad (1.9)$$

An ODE describes the change of compound $x_i(t)$ over time t [270]. In a metabolic setting, $x_i(t)$ represents the concentration of a molecule that is metabolized by a specific enzyme, and θ the reaction parameters, e.g. kinetic rate constants or amount of enzymes. The functions f_i are often determined by the contribution of production and degradation reactions in a biochemical system. For simplicity, a spatial homogeneity, meaning a “well-stirred” molecular environment, and no time dependencies for rate parameters are usually assumed [127]. In many cases, solutions for ODE systems cannot be found analytically. Hence numerical methods are applied in order to obtain approximate solutions [49].

Dynamic models of metabolic reactions can be used to generate *in silico* metabolite data. Krumsiek *et al.* for instance used metabolic reaction systems to simulate large-scale metabolomics data for an evaluation of data-driven network reconstruction methods (see Section 1.3). In Chapter 2 we will also apply a forward simulation approach to generate genotype-dependent metabolite levels for the evaluation of our network-based metabolite ratio approach.

Parameter estimation

For metabolic systems the model structure (e.g. stoichiometry and the rate laws) can be retrieved from biochemical knowledge about compounds and catalyzing enzymes in metabolic pathways. Yet the model parameters θ (e.g. kinetic constants) are mostly unknown as they cannot be measured experimentally. For this reason, the unknown model parameters need to be estimated from experimental data. For the *parameter estimation*, also called *parameter inference*, model observables $y_i(t_k)$ (for instance predicted metabolite concentration time courses) are compared to measured data $\bar{y}_i(t_k)$ for all n compounds ($i = 1, \dots, n$) and N time points t_k ($k = 1, \dots, N$). Note that in our experi-

mental setting we can compare the observables directly with the measured data and do not need a link function ($y_i(t_k) = x_i(t_k)$).

The quality of how well the model predictions for a given parameter set θ match the experimental data can be assessed using a likelihood function $L(\theta)$. Experimental data often include measurement noise denoted by ϵ_{ik} . For additive normally distributed measurement noise ($\bar{y}_{i,k} = y_i(t_k) + \epsilon_{ik}$ with $\epsilon_{ik} \sim \mathcal{N}(0, \sigma_{ik}^2)$) the likelihood function reads

$$L(\theta) = \prod_{k=1}^N \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma_{ik}} \exp \left(-\frac{1}{2} \left(\frac{\bar{y}_i(t_k) - y_i(t_k)}{\sigma_{ik}} \right)^2 \right). \quad (1.10)$$

In case of multiplicative log-normally distributed measurement noise ($\bar{y}_{i,k} = y_i(t_k) \cdot \nu_{ik}$) with $\nu_{ik} \sim \log \mathcal{N}(0, \sigma_{ik}^2)$ the likelihood is computed by

$$L(\theta) = \prod_{k=1}^N \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma_{ik}} \exp \left(-\frac{1}{2} \left(\frac{\log \bar{y}_i(t_k) - \log y_i(t_k, \theta)}{\sigma_{ik}} \right)^2 \right). \quad (1.11)$$

During the model calibration or fitting process, the parameters are adjusted in iterative steps for maximizing the likelihood, i.e. for an improved agreement between model predictions and experimental data (*maximum-likelihood estimation* [141]). If parameters have different orders of magnitude, parameter values can be fitted on log scale for an efficient search in the parameter space. For parameter estimation of biochemical pathway models, deterministic and stochastic optimization strategies can be applied [165, 253]. Software tools for the simulation of biochemical systems and for parameter estimation are for instance Gepasi [157] and its successor COPASI [99], CellDesigner [71], SBtoolbox2 for Matlab [222] or Data2Dynamics [202].

Parameter identifiability

For parameter estimation it is important to consider the uncertainties in the model parameters with respect to incomplete data and experimental measurement noise [35]. The

uncertainty analysis can be performed for instance by calculating parameter confidence intervals [156] or evaluating the identifiability of parameters [201]. A parameter is identifiable if the confidence intervals are finite. Parameters can be structurally and practically non-identifiable. Structural non-identifiability may occur for example if two parameters θ_1 and θ_2 appear only as a product in the model. Increasing θ_1 while decreasing θ_2 or vice versa will result in the same likelihood value. Thus only the product of the two parameters, but not the individual values can be determined. One might resolve this structural non-identifiability by replacing the product $\theta_1\theta_2$ with a new parameter θ_3 . Yet often it is not straightforward to detect and resolve structurally non-identifiable parameters. In contrast to structural non-identifiability, which solely depends on the model structure and is independent of the data, parameters can be practically non-identifiable if the wrong observables have been measured or too few data is available. This issue can be resolved by adapting the experimental setup. Profile likelihood approaches [201] can be applied to assess if parameter confidence intervals are finite, meaning the parameters are identifiable. Here the likelihood landscape space is explored for each parameter individually. Non-identifiable parameters are characterized by flat profile likelihoods. The Data2Dynamics software package [202] can be used to calculate profile likelihoods for assessing the parameter identifiability in metabolic models.

Model selection

Often for the description of biochemical pathways there are several alternative models possible, depending for instance on the number of model compounds or the detailed description of enzymatic reactions. During the modeling process it is often required to strike a balance between model complexity and simplicity. Complex models might include many biological details and explain the experimental data better compared to more simple models. Adding additional parameters to the model will often increase the likelihood but might result in overfitting [121]. On the other hand, simple models facilitate the interpretation of the model prediction with respect to the underlying biological processes. Therefore it is important to identify too complicated or biologically incorrect models. This can be achieved by selecting - based on statistical criteria - those models which are best supported by the experimental data [78]. Often these criteria account

for both the model complexity (e.g. the number of free parameters) and the agreement between model output and data. A likelihood ratio test can be performed when choosing between two nested models [259]. Further selection criteria are the Akaike information criterion (AIC, [3]) and the Bayesian information criterion (BIC, [226]). Both criteria include a penalty term for the number of free parameters. By minimizing the AIC or BIC score one tries to find a model which describes the data best with a minimal set of parameters, as a lower AIC or BIC value results either from fewer parameters, better fit to the data or both. A sophisticated Bayesian model selection method is the Bayes factor which includes prior information about for instance the distribution of the parameters [221].

1.5 Research question

The main goal of this thesis is to evaluate how fundamental knowledge about biochemical pathways can be combined with experimental measurements on different biological scales for a model-based analysis of genetic and nutritional effects on metabolism. Impaired enzymatic function or unbalanced nutrition can lead to pathophysiological metabolic conditions, reflected by altered metabolite levels. For this reason, metabolic profiling is for instance used in newborn screenings for the diagnosis of inherited diseases. For a better understanding of the relationship between metabolic processes and genetic, physiological and nutritional factors, research has been performed for a broad spectrum of experimental setups on different biological scales, ranging from cross-sectional population data over time-resolved *in vivo* physiological challenging results to *in vitro* experiments using genetically modified cell lines. Metabolomics profiling was applied to measure metabolite levels as readouts for metabolic functions. Although metabolic systems are complex, biochemical research has provided fundamental knowledge about the underlying pathways and their interrelations. In this thesis, I will demonstrate how information about metabolic networks and biochemical pathways can be utilized for an improved analysis of large-scale and multivariate experimental data. In particular, I will discuss how network-based analysis approaches for steady-state and dynamic data can be tailored to the specific experimental setups and biological questions related to genetic and nutritional effects on metabolism.

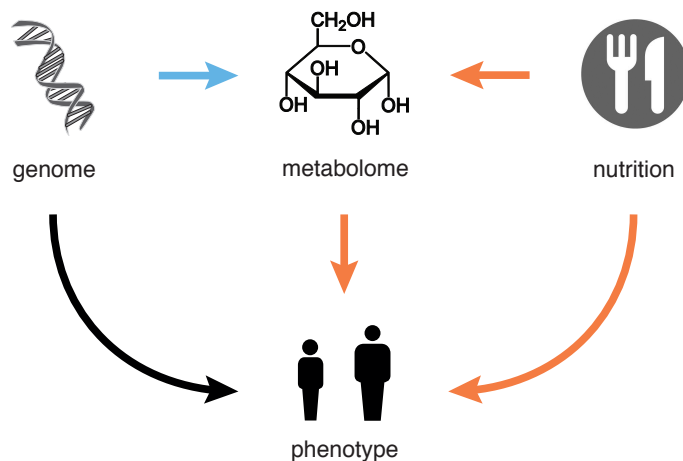


Figure 1.6: Overview of the thesis. Intrinsic factors like gene products and environmental factors like nutrition influence metabolism and define the individual phenotype (black arrow). In this thesis we analyzed the genetic and nutritional effects on metabolism for specific experimental setups on different biological scales. In Chapter 2 we investigated the relationship between genetic variation and metabolite levels as intermediate phenotypes (blue arrow). Chapter 3 presents our analysis of interindividual variation in metabolic profiles that was induced by nutritional challenges (orange arrows). In Chapter 4 we experimentally changed the expression of an enzyme gene to evaluate the impact of enzyme levels on a specific metabolic pathway (blue arrow).

1.6 Overview of this thesis

In the following we will give a short outline of the content of this thesis. For three biological scenarios, we will discuss how we developed network-based methods and computational models of metabolic pathways for the analysis of genetic and nutritional effects on metabolism (see Figure 1.6).

In **Chapter 2**, we present a system-wide analysis of the interplay between genetics and metabolism based on results from genome-wide association studies (GWAS). We describe a network-based method that we developed for the selection of biologically meaningful metabolite ratios as GWAS traits in order to overcome statistical and computational challenges. We evaluate this network-based approach both on *in silico* data derived from simulated reaction networks and published results from genome-wide association studies. Our findings demonstrate that combining biochemical networks with large-scale genetic

and metabolic phenotyping data provides a valuable approach for studying the general effects of genetic variation on metabolism at population level.

Chapter 3 introduces a model-based analysis of time-resolved metabolomics profiles from a human challenging study. For a better understanding of metabolic response under specific nutritional and physiological challenges, we study metabolite profiles of 15 healthy volunteers under fasting conditions. In order to analyze the challenge-induced interindividual variation of metabolite levels, we developed a model of the fatty acid beta-oxidation (FAO) pathway to derive conversion rate parameters which describe the individual metabolic capacity. We further show how model readouts can be related to anthropometric and biochemical parameters for a better explanation of the observed interindividual variation in metabolic profiles.

In **Chapter 4** we analyze GWAS-derived associations between butyrylcarnitine, the transport form of a short-chain fatty acid, and genetic variants in the locus of the mitochondrial fatty acid beta-oxidation enzyme ACADS. We extend the FAO pathway model presented in Chapter 3 to describe the knockdown-specific, time-resolved measurements of fatty acid intermediate metabolites. Based on the reaction rates inferred from our model and experimental data, we compare the dynamical changes between wild-type and ACADS knockdown conditions statistically and discuss possible compensation mechanisms in the FAO pathway which are predicted by the model.

The final **Chapter 5** summarizes the methods and results presented in this thesis and discusses possible extensions as well further biological scenarios for a network-based analysis of metabolomics data.

Scientific publications

The results presented in this thesis have led to the following contributions that have been published or are currently within the publication process (sorted by the corresponding Chapter):

Chapter 2:

Stückler F, Krumsiek J, Suhre K, Gieger C, Kastenmüller G, Theis FJ (2014): Network-based metabolite ratios for an improved functional characterization of genome-wide association study results. *Submitted*.

This publication describes a network-based metabolite ratio selection method that I developed for an improved analysis of genome-wide association studies with metabolic traits.

Chapter 3:

Krug S*, Kastenmüller G*, **Stückler F***, Rist MJ*, Skurk T*, Sailer M, Raffler J, Römisch-Margl W, Adamski J, Prehn C, Frank T, Engel KH, Hofmann T, Luy B, Zimmermann R, Moritz F, Schmitt-Kopplin P, Krumsiek J, Kremer W, Huber F, Oeh U, Theis FJ, Szymczak W, Hauner H, Suhre K, Daniel H (2012): The dynamic range of the human metabolome revealed by challenges. *FASEB J.* 26: 260719.

This publication is a joint first author work and the content of this paper is also part of another thesis by Susanne Krug [132]. My contribution to the publication is the development of a model to analyze the metabolomics data, as well the statistical analysis of model readouts and their interpretation.

Krumsiek J, **Stückler F**, Kastenmüller G, Theis FJ (2012): Systems Biology Meets Metabolism. In: Suhre K, editor. *Genetics Meets Metabolomics: from Experiment to Systems Biology*. New York, NY: Springer New York. pp. 281-313.

Chapter 4:

Ehlers K*, **Stückler F***, Hastreiter M, Pfeiffer L, Reischl E, Kastenmüller G, Daniel H, Ensenaer R, Krumsiek J, Hauner H, Theis FJ, Laumen H (2014): In vitro modeling

and dynamic analysis of a metabolic quantitative trait locus implies novel features of ACADS function in fatty acid oxidation. *in submission*.

This publication is a joint first author work and the content of this paper is also part of another thesis by Kerstin Ehlers, who also performed the experiments [62]. My contribution to this work is the development of a dynamic pathway model, the statistical analysis of the *in vitro* metabolomics data, as well the biological interpretation of the model prediction.

Further scientific projects

Besides the contributions presented in this thesis, I was involved in further projects, which resulted in the following publications:

Ried JS, Baurecht H, **Stückler F**, Krumsiek J, Gieger C, Heinrich J, Kabesch M, Prehn C, Peters A, Rodriguez E, Schulz H, Strauch K, Suhre K, Wang-Sattler R, Wichmann HE, Theis FJ, Illig T, Adamski J, Weidinger S (2013): Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma. *Allergy*. 68: 62936.

Brand T*, Kondofersky I*, Ehlers K, Römisch-Margl W, **Stückler F**, Krumsiek J, Bangert A, Artati A, Prehn C, Adamski J, Kastenmüller G, Fuchs C, Theis FJ, Laumen H, Hauner H (2014): Effect of dietary standardization on the metabolomic response to a defined meal challenge in healthy individuals. *Submitted*.

* = equal contributions

Chapter 2

Network-based metabolite ratios for an improved functional characterization of genome-wide association study results

In this chapter, we present an approach which utilizes metabolic network information for the selection of biologically meaningful metabolite ratios as quantitative traits in genome-wide associations studies (GWAS, see Section 1.2). Recent genome-wide association studies with population-based metabolomics datasets (mGWAS) reported high associations between genetic variants in regions of metabolic enzymes and transporters and the respective substrate and product metabolite levels [81, 104, 117, 172, 230, 237]. Using metabolite concentration ratios as quantitative traits in addition to single metabolite concentrations further improved the statistical significance of the associations. For example, Suhre *et al.* [237] reported that the association of a genetic variant in the FADS1 locus and the ratio between fatty acids 20:3 and 20:4 is much stronger compared to the association with the respective single metabolite levels (see Figure 2.1). FADS1 (fatty acid delta-5 desaturase) is an enzyme which catalyzes the desaturase reaction of fatty acids 20:3 to 20:4. The increase in association strength due to the ratio between

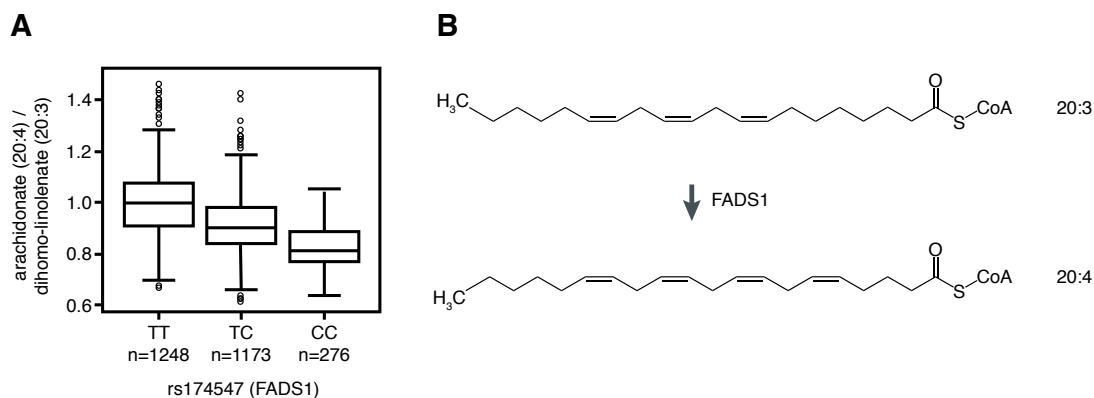


Figure 2.1: Association between the ratio of arachidonate (20:4) and dihomolinolenate (20:3) metabolite plasma levels with SNP rs174547 in the FADS1 locus as reported in [237]. **A:** Box plots show ratio levels grouped by genotype: major allele homozygotes (TT), heterozygotes (TC) and minor allele homozygotes (CC). The number of samples per group is denoted by n. Metabolite ratio data is shown on a lognormal scale. The association p-value of the ratio arachidonate / dihomolinolenate ($3.6 \cdot 10^{-101}$) is much stronger compared to the association with the single metabolite levels arachidonate ($1.7 \cdot 10^{-30}$) and dihomolinolenate ($3.3 \cdot 10^{-9}$). **B:** The FADS1 locus encodes for the enzyme fatty acid delta-5 desaturase with fatty acids 20:3 and 20:4 as substrate and product, respectively.

reaction substrate-product pairs points to the biological function of the enzyme [139]. For this analysis, the authors considered all possible ratios for all measured metabolites. This “all ratios” approach revealed new insights into the genetic basis for human metabolic individuality.

However, taking all possible metabolite ratio combinations into account can be challenging from a statistical, computational and interpretational point-of-view, since inevitably many biochemically unrelated metabolite pairs are tested. In addition, the effect of a specific genetic variant on metabolites that are within a pathway might often be quite similar. Thus, conventional multiple testing approaches (like Bonferroni correction) might be too stringent for GWAS and possibly reduce the statistical power. The increasing number of measured variables will lead to a quadratic increase in the number of tests when dealing with ratios (see also Figure 2.14 for an estimation of the number of ratio candidates). Current methods allow for the detection of a few hundred metabolites, but this number will increase rapidly [179]. Also modern sequencing techniques will provide more data about genetic variants [158]. On a practical level it will be becoming computationally

and statistically demanding to conduct association tests for millions of SNPs or other genetic variants from several thousand individual genomes in multiple cohorts across millions of metabolite ratio combinations. For instance, in a study measuring all expected 20,000 human metabolites [267], testing 1 million SNPs for all ratio combinations (e.g. $20,000^2$ cases) would result in a significance threshold after Bonferroni correction of $p < 2.5 \cdot 10^{-15}$. While for large GWAS cohorts the limited power problem might be only an issue for small effect sizes, it can be crucial for the design of case-control studies with small sample numbers [188]. The functional interpretation of the high-dimensional and complex data is also challenging for such study designs due to the vast amount of results produced. Preselecting meaningful ratio candidates based on biological network information thus is crucial to address the above-mentioned limitations.

In order to overcome these challenges, we developed a network-based method which includes metabolic pathway information for an improved selection of ratio candidates in mGWAS (*network-based metabolite ratios*, NBR, see Figure 2.2). Incorporating network information into the analysis of biological data has shown to be successful for example in proteomics and genomics applications [14, 28, 102, 109]. For the analysis of mGWAS studies, the metabolite dependencies can be obtained, for example, from metabolic pathways (see Section 1.3), which are available from various sources such as KEGG, BiGG, EHMN and MetaCyc [31, 113, 149, 219]. However, many measured metabolites are not annotated and the derived pathway information might be incomplete [119]. Statistical approaches like correlation-based methods, which purely rely on measured metabolomics data, can provide network information for all detected metabolites [9, 29, 233]. As metabolites in population data are highly correlated, we used a previously reported network reconstruction method based on partial correlations from Gaussian Graphical models (GGM, [136], see also Section 1.3). Here indirect interactions between metabolites are removed before using the data-driven network for selecting ratio candidates.

In the following, we will show how metabolic networks can be used to analyze the mGWAS data and facilitate the functional characterization of the respective results. The chapter is organized as follows: First we test the NBR approach on simulated population data. We then evaluate whether ratio-SNP association hits reflect metabolic pathway reactions based on mGWAS results of a human population cohort. To this end, we test if metabolites that have a significant ratio-SNP association are more closely connected in

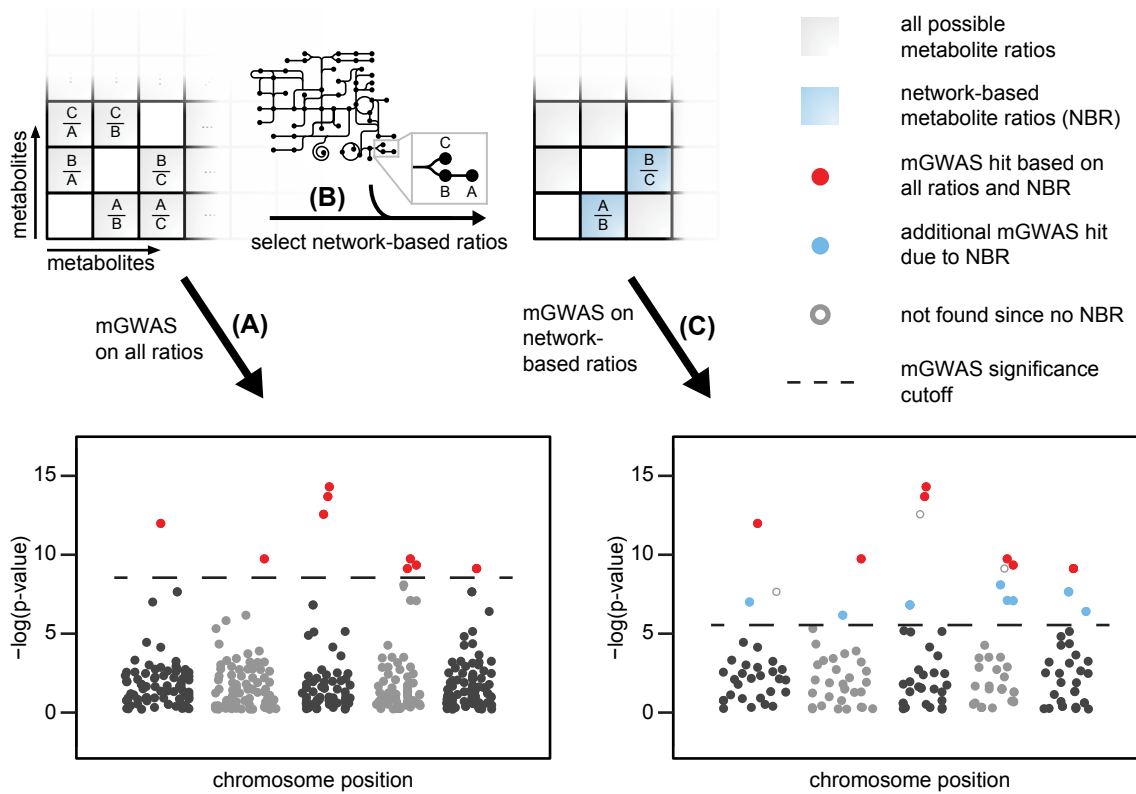


Figure 2.2: Network-based metabolite ratios (NBR) for an improved analysis and interpretation of genome-wide association studies with metabolic traits (mGWAS). Considering all possible metabolite ratios (red dots) as traits in mGWAS (A) has proven valuable in finding new functional insights about underlying biological processes [237]. Selecting network-based metabolite ratios instead of all possible ratios reduces the number of association tests and therefore results in a less stringent significance threshold after correction for multiple testing (B). Testing only for selected NBR (C) reveals new significant associations (blue dots). Due to incomplete metabolic network information or unknown complex regulatory effects some hits might not be found (grey circles).

metabolic networks reconstructed using Gaussian Graphical modeling (see Section 1.3). In addition, we compare mGWAS results that were obtained using the all-ratio approach with our NBR method. For this comparison we first consider NBRs that were selected based on metabolic networks from databases (pathway-based NBRs, PW-NBRs). Since such derived networks are incomplete due to missing annotations, we also use networks reconstructed in a purely data-driven fashion from metabolomics measurements using Gaussian graphical modeling (GGM-NBRs). Furthermore we discuss newly predicted associations and analyze associations which are not found using NBR in the context of pathway-related metabolites that are all affected by the same genetic locus. In addition we show that our NBR approach based on data-driven networks can also be applied for the analysis of further mGWAS results from different study cohorts.

The work presented in this chapter has been performed in collaboration with the group of Gabi Kastenmüller and Karsten Suhre. The results are summarized in:

- **Stückler F**, Krumsiek J, Suhre K, Gieger C, Kastenmüller G, Theis FJ (2014): Network-based metabolite ratios for an improved functional characterization of genome-wide association study results. *Submitted*.

2.1 Methods

In silico simulation of SNP effects on metabolic reaction networks

Metabolic reaction networks were simulated using mass-action kinetics (see Section 1.4). The change of all metabolites x in the reaction network can be described by

$$\frac{dx}{dt} = \dot{x} = Sv(x)$$

The topology of the reaction network is represented by the stoichiometric matrix S , while the dynamical properties are determined by the reaction flux vector $v(x)$ with reaction-specific rate parameters. Steady state metabolite concentrations are obtained either by solving the system of linear equations (i.e. setting all equations to zero) for linear systems or by simulating the dynamical system until it has reached its equilibrium.

Using this framework, synthetic data for different reaction networks topologies, altered reaction rates due to SNP effects, allele frequencies and population size were created. The allele frequencies were calculated following Hardy-Weinberg equilibrium model with chosen minor allele frequency between 0.05 and 0.5. To account for variability in the simulated population, the reaction rates for each individual were randomly and independently drawn from log-normal distributions with mean 3 and standard deviation 1. The effect of a specific SNP was modeled by adding the effect size once (heterozygote case) or twice (homozygote minor allele case) to the mean log-normal parameter of the affected reaction rate. For example $\Delta ES = 0.4$ indicates that the rates for the major allele homozygote, heterozygote and minor allele homozygote case are drawn from log-normal distributions with mean 3.0, 3.4 and 3.8, respectively. The population size was chosen to be between 50 and 10,000. We used a multiplicative error model in order to take technical noise into account, which arises during the measurement of metabolomics samples. To this end, calculated metabolite steady state concentrations were multiplied with random factors drawn from a log-normal distribution with mean 1 and standard deviation 0.05. As a result, we obtained all metabolite concentrations for each individual in the population depending on its genetic background.

Reconstruction of metabolic networks using Gaussian Graphical modeling

Gaussian graphical models (GGM) are calculated using full-order partial correlation coefficients, i.e. each pairwise correlation is corrected against all remaining ($n-2$) variables to remove indirect effects. For data with more samples than variables, full-order partial correlations can be calculated by a matrix inversion operation [140]. A more detailed description of partial correlations can be found in Section 1.3. See also [135, 136] for GGM calculation based on metabolomics data. Since in our simulated data there are for some cases less samples than variables (metabolites), we used the R-package GeneNet [181] which calculates a regularized version of partial correlation coefficients. This method yields also for cases with more samples than variables robust estimates of partial correlation coefficients. All computations were performed on log-transformed metabolite concentrations, as testing for normality revealed that for most cases the log-transformed concentrations were closer to a normal distribution than the untransformed values [237].

Association between metabolite ratios and SNP effects for simulated reactions networks

For all association tests, metabolite ratio candidates were selected by three methods (see Figure 2.3): 1) all possible ratio combinations between each metabolite in the underlying reaction network (all ratios), 2) only ratios between neighboring metabolites in the reaction network (PW-NBR) and 3) only ratios between neighboring metabolites in the network structure reconstructed from metabolomics data using GGMs (GGM-NBR). The third approach represents a purely data-driven approach that does not require known pathway interactions as input and is thus independent of functional annotations of the measured molecules (see description above and Section 1.3). To test for the association between metabolite ratios and genetic background in the simulated concentrations we used a linear additive regression model, as previously reported in several GWAS studies [6, 104, 237], with genotype as independent variable and the respective metabolite ratio as response. Simulated metabolite concentrations were log-transformed for statistical analysis. Regression coefficients were tested for significant deviation from zero. To

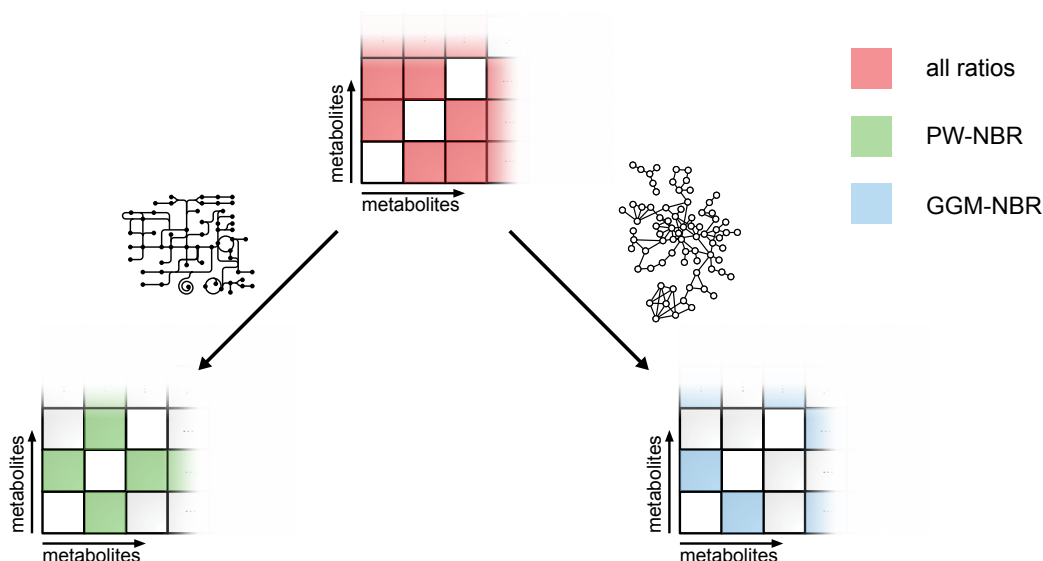


Figure 2.3: Selection of metabolite ratio candidates for association tests. Usually all possible ratio combinations are selected (red). Using our network-based metabolite ratio approach, only ratios between neighboring metabolites in a known network (green, pathway-based NBR, PW-NBR) or in a network which was reconstructed from metabolomics data using Gaussian graphical modeling (blue, GGM-NBR).

account for the number of tests for each ratio candidate set, Bonferroni correction was applied. If the best SNP-ratio association (lowest p-value) matched the underlying reaction networks, the test was counted as true positive. In this case the simulated SNP was affecting the direct reaction between the two ratio metabolites. The fraction of truly predicted associations (%TP, number true positive cases divided by all cases) was used to assess the quality of each ratio candidate selection method.

NBR on genome-wide association data

We evaluated our network-based ratio approach on reported data from the two population cohorts, KORA and TwinsUK. Details about the sample acquisition, metabolomics measurements and genotyping can be found in [237]. Briefly, we used metabolomics measurements of 295 metabolites and genotyping data for 655,658 SNPs from 1,768 fasting serum samples of the KORA study and from 1,052 individuals of the TwinsUK cohort. For the estimation of GGMs a full metabolite concentration data matrix without missing

values is required. We therefore excluded first metabolites with more than 20% missing values and after that samples with more than 10% missing values, resulting in a filtered data matrix of 1764 samples with 218 metabolites [135]. Remaining missing values were imputed with the R package “mice” [248]. Note that we only analyzed the reported ratio-SNP associations [237] for these 218 metabolites.

Since all metabolomics data were transformed to log-scale for further statistical tests, we calculated metabolite ratios by taking the difference between the log-transformed concentrations, yielding 23,653 metabolite ratios for the all ratios case. As we wanted to focus our analysis on association hits at genetic locus level, we combined ratio-SNP associations that were within linkage disequilibrium of 0.8 or higher, based on LD data from HapMap derived from the SNAP server [110]. For cases where several SNPs within one locus were associated to the same metabolite ratio we only used the most significant association. No evidence of population stratification could be found in the population cohorts. Lambda values [53] ranged from 0.965 to 1.024 (median 1.006) in KORA, and from 0.940 to 1.013 (median 0.985) in TwinsUK [104, 237].

For the selection of ratio candidates based on network information we used two metabolic network sources (see Figure 2.3): a pathway-based network (PW-NBR) and a GGM-based network (GGM-NBR). The first network was constructed by combining metabolite reaction information from three independent databases: 1) *Homo sapiens* Recon 1 from the BiGG databases (confidence score of at least 4) [219], 2) the Edinburgh Human Metabolic Network reconstruction [149] and 3) the KEGG PATHWAY database [113]. Due to missing annotations, only 122 out of 218 measured metabolites were found in the combined pathway-based network and further used for the PW-NBR analysis. The GGM-NBRs were selected based on the GGM for all 218 metabolites. In the GGM network only metabolite pairs were chosen that showed an absolute partial correlation score of 0.1 or higher, corresponding to an partial correlation p-value cutoff of $p < 8.1 \cdot 10^{-5}$ (see Figure 2.10 for an evaluation of GGM-NBR parameter settings). In order to account for missing metabolic connections in the networks, we chose metabolites that were connected via one or two steps as ratio candidates for GGM-NBR and PW-NBR, resulting in 3,786 and 879 metabolite ratios, respectively.

As described above, a linear additive regression model was used to test for the association between metabolite ratios and genetic background using log10-scaled metabolic traits. Genotype is coded as 0-1-2 for major-hetero-minor genotype. The model was adjusted for age and gender as covariates. Beta (the slope of the linear model) and p-value have been reported in [237]. We applied Bonferroni correction to account for the large number of association tests. The p-value threshold was calculated by $0.05/(\text{number of selected ratios} \cdot \text{number of SNPs})$. Thus the adjusted threshold for genome-wide significance for the all ratios, GGM-NBR and PW-NBR analysis was $p < 3.2241 \cdot 10^{-12}$, $p < 2.0142 \cdot 10^{-11}$ and $p < 8.6757 \cdot 10^{-11}$, respectively. For SNP-ratio associations that were not discovered using the GGM-NBR approach we checked whether these effects could be explained by related metabolites. Based on the edge weights of the underlying GGM network we calculated shortest paths between the two metabolites of the ratio pair [276]. On these paths we checked if there are other ratio pairs which are associated to any SNP in close genetic distance to the original SNP, that was not found using GGM-NBR.

Analysis of metabolic distances in Gaussian Graphical models

The pairwise distance $d_{i,j}$ between metabolites M_i and M_j in the GGM was calculated based on the respective partial correlation coefficient $\zeta_{i,j}$, which was transformed by $d_{i,j} = \exp(-\zeta_{i,j})$. Closely connected metabolites with high partial correlation coefficients have small distances. Based on this distance measure we calculated shortest paths between all metabolite pairs. We compared the distribution of all pairwise metabolite shortest path lengths with the distribution of shortest path lengths between metabolite pairs whose ratio was significantly associated to at least one SNP (adjusted Bonferroni threshold $p < 3.2241 \cdot 10^{-12}$). ROC analysis [67] was used to quantify the separation of the two distributions. To assess the significance of this observed AUC score, we performed graph randomization by edge rewiring on the distance-weighted graph as described in [89]. During each randomization step the target nodes of two randomly chosen edges are exchanged. In order to achieve sufficient graph randomization, the exchange step is repeated five times the number of edges in the graph, as suggested in [271]. For the empirical p-value calculation we performed the distance-based ROC analysis for 10^7 randomized graphs.

Estimating the number of ratio candidates for the all ratios and GGM-NBR approach

The number of ratio candidates that are tested for association using the all ratios (N_{all}) or GGM-NBR (N_{NBR}) approach depends on the number of measured metabolites n . The amount of all ratios candidates was calculated by $N_{all} = \binom{n}{2}$, since in the mGWAS setting all metabolomics data were transformed to log-scale for further statistical tests. Therefore, the metabolite ratios are derived by taking the difference between the log-transformed concentrations. For estimating the number of GGM-NBR ratio candidates (N_{NBR}), we randomly choose n metabolites out of the KORA dataset and calculated all partial correlations between the metabolites, resulting in a Gaussian Graphical model. We filtered out partial correlations of 0.1 and below. Metabolites that were still connected via one or two steps in this filtered network are GGM-NBR ratio candidates for further association tests. For each metabolite number n we repeated this procedure 1000 times to obtain an estimation of mean and standard deviation for N_{NBR} .

2.2 Network-based metabolite ratios improve GWAS analysis of simulated reaction networks

Simulated reaction networks are useful tools to investigate the properties of biological systems and to examine new approaches in a well-defined setup [124, 136]. We used such a framework to address whether selecting network-based metabolite ratios improves the SNP-ratio associations results, compared to taking all possible ratios. To this end, we computationally generated metabolomics measurements resembling features of a real population. Figure 2.4 depicts an example of genotype-specific metabolic traits in a simulated reaction network. Our model incorporates genetic variation that has an effect on the respective enzyme activities. Such variation has been reported for example in a GWAS study that found an association between several SNPs in the ACE structural gene and ACE activity [40]. The reactions that we studied followed mass-action kinetic rate laws and were implemented as ordinary differential equations (see Section 1.4). In order to account for variation between individuals, each reaction rate was drawn from a log-

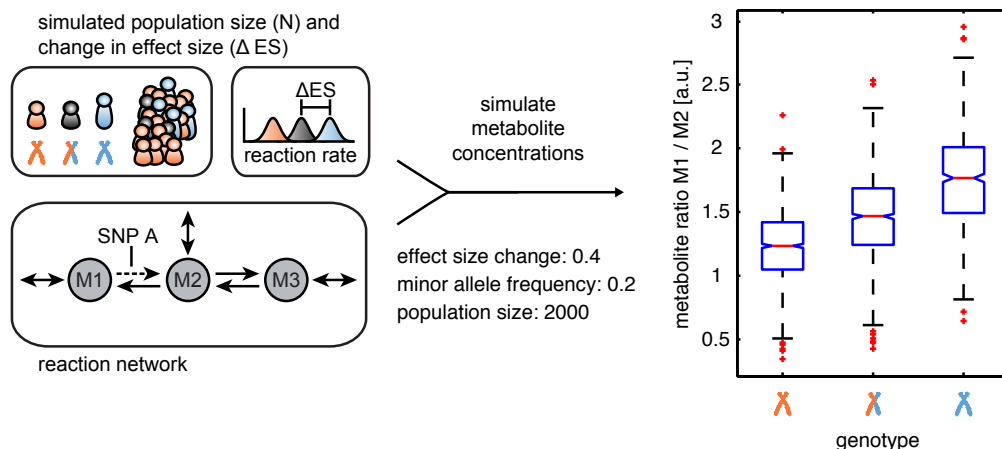


Figure 2.4: Genotype-specific metabolic traits in simulated reaction networks. The ratio on log-scale of metabolites M1 and M2 which are involved in the SNP-affected reaction A in the reaction network shows a genotype-specific association with linear dependency. For the *in silico* simulation, minor allele frequency was set to 0.2, effect size change to 0.4 and population size to 2000 (see the methods section 2.1 for a detailed description of the model parameters).

normal distribution [146] and then used to calculate individual steady state metabolite concentrations. Metabolites involved in SNP-affected reactions and their corresponding ratios showed genotype-specific levels with linear dependency (see Figure 2.4), which is in accordance with previous studies on real data [6, 81, 104, 237, 238].

Based on the simulated metabolic traits, we evaluated the NBR approach for mGWAS settings (see Figure 2.5A for a schematic representation of the evaluation workflow). For all association tests, metabolite ratio candidates were selected by three different approaches: 1) all possible ratio combinations between all metabolites (all ratios), 2) only ratios between connected metabolites in the network assuming that we know all true pathway reactions (network-based metabolite ratios from pathway information, PW-NBR) and 3) only ratios between neighboring metabolites in the network reconstructed from simulated metabolomics data using Gaussian Graphical modeling (network-based metabolite ratios from GGM information, GGM-NBR) [136]. Results of the evaluation of approach 2 for different network topologies can be found in Figure 2.5 and Figure 2.6, results for approach 3 are depicted in Figure 2.7.

We tested for SNP-ratio associations using a linear model with genotype as independent variable and the respective metabolite ratio as response. All SNP-ratio associations were

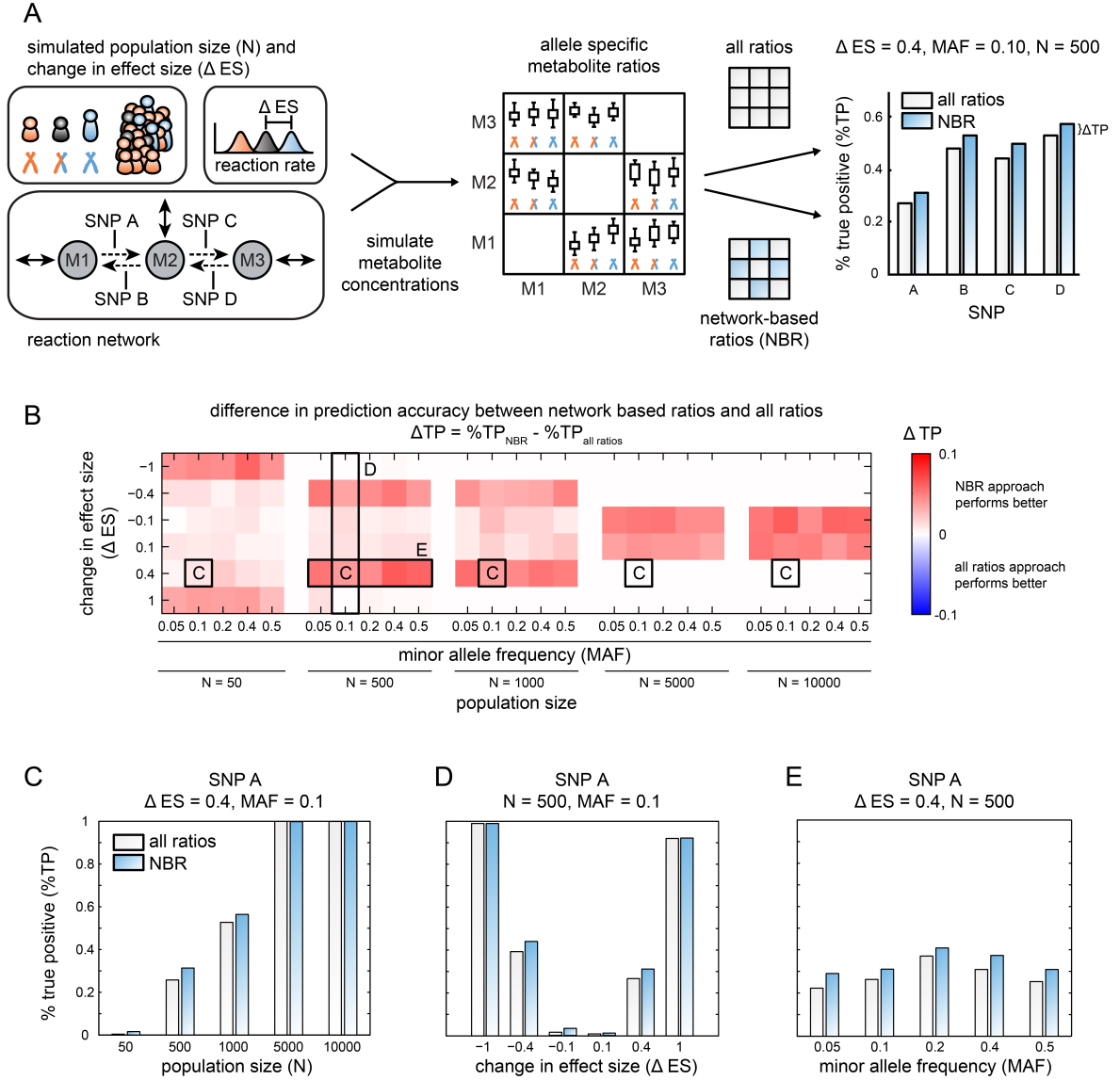


Figure 2.5: Network-based metabolite ratios (NBR) on simulated reaction networks. The NBR approach improves the analysis of ratio-SNP associations. **A:** Scheme of *in silico* simulation of SNP effects in metabolic reaction networks. For specific population sizes (N), minor allele frequencies (MAF), SNP effect sizes (ES) and reaction network topologies with different SNPs, steady state metabolite concentrations were simulated. Based on selected metabolite ratio sets (all ratios or PW-NBR) SNP-ratio associations were calculated. For true positive prediction, the best association hit matches the underlying reaction in the network. The fraction of truly predicted associations (%TP) was evaluated from 500 iterations. Since less association tests are needed using NBR, this approach is more sensitive, reflected by higher %TP values. **B:** Differences in %TP between PW-NBR and all ratios analysis (ΔTP). The simulation was based on the reaction network depicted in **A** with one SNP-affected reaction (SNP A) between M1 and M2. A more detailed view of specific scenarios is given in subfigures **C**, **D** and **E**. Especially for small sample numbers in combination with small effect sizes the NBR approach improves the association analysis. **C**, **D**, **E:** Simulation results for selected scenarios as marked in **B**, with varying population size (**C**), effect size differences (**D**) or minor allele frequencies (**E**).

PW-NBR compared to all ratios approach

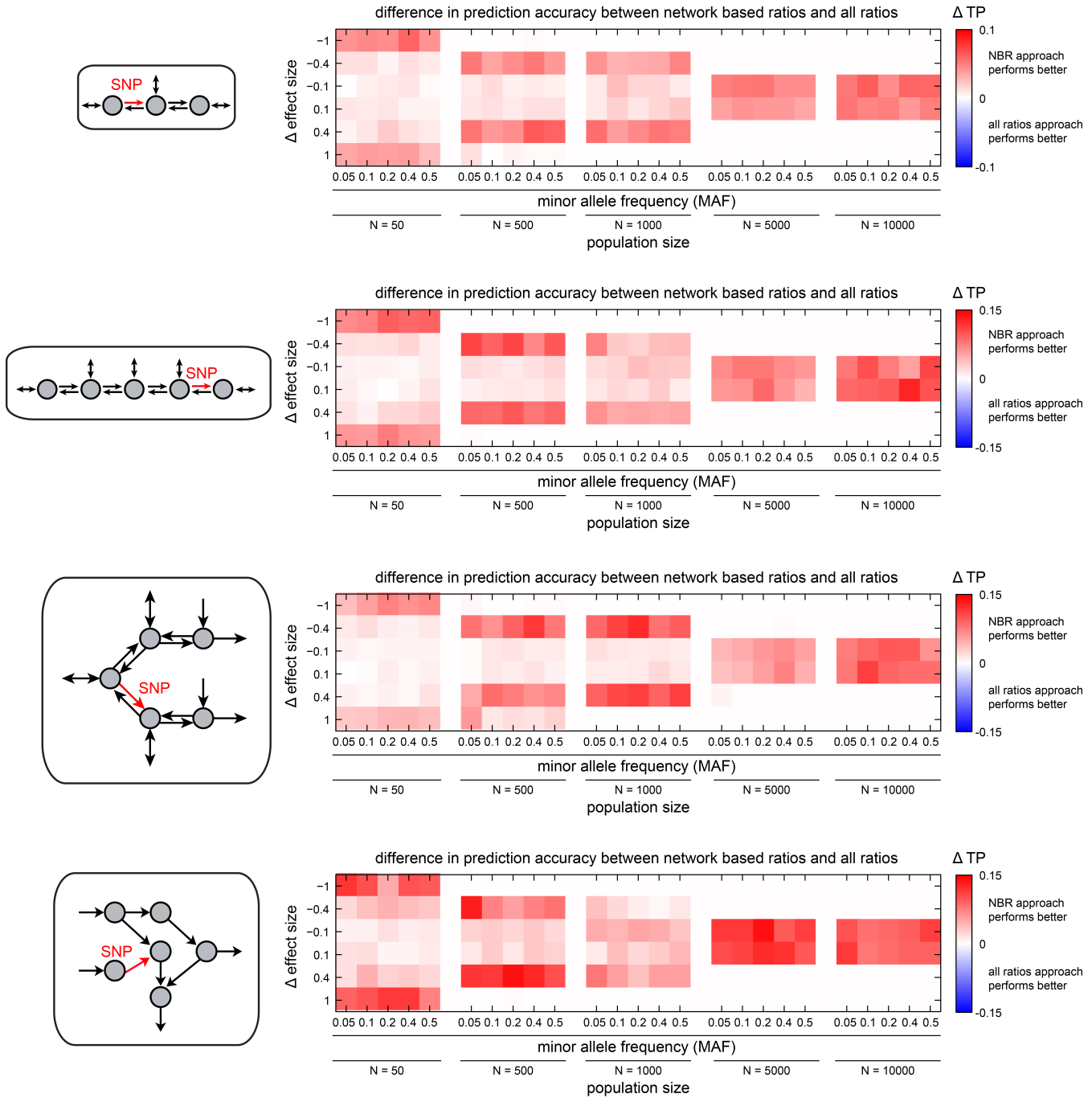


Figure 2.6: Comparing the prediction accuracy between the all ratios approach and the network-based metabolite ratio approach (ratios are selected on the basis of the known network topology, PW-NBR).

adjusted for multiple testing using Bonferroni correction based on the approach-specific number of ratio candidates. A ratio was counted as true positive if the best SNP-ratio association hit (lowest p-value) matched the underlying network, meaning that the simulated SNP was affecting the direct reaction between the two ratio metabolites. For instance, if the ratio M1/M2 in the metabolic network as depicted in Figure 2.4 shows the best association to SNP A, this SNP-ratio association is true positive, since SNP A directly affects the reaction between M1 and M2. As a quality measure the fraction of truly predicted associations (%TP) compared to all predictions was calculated.

Using this simulation framework, we tested several network topologies with different SNP-affected reactions and varying population sizes (N), minor allele frequencies (MAF) and SNP effect sizes (ΔES). The schematic workflow is depicted in Figure 2.5A for a linear reaction network consisting of three metabolites connected by reversible reactions. For each species we introduced exchange reactions, reflecting interactions with other metabolic pathways. The overview of all integrated scenario results in Figure 2.5B reveals that the network-based metabolite ratio approach performs equally well or even better compared to the all ratios approach. NBR improves the prediction of SNP-reaction associations especially for scenarios with small effect sizes ($\Delta ES = 0.4$) in combinations with sample numbers of 500 and 1000. In order to detect small effects usually one has to increase the sample size, which is often a limiting factor. The improvement of the results is based on the different choice of ratio candidate sets. By only taking ratios of connected metabolites into account for the linear association model, we reduce the number of tests and increase the power of our analysis. Further examples for which the results of approach 1 and approach 2 are compared for different network topologies can be found in Figure 2.6.

Figure 2.5 shows the results for the NBR analysis with the given network structure from the simulated model (PW-NBR, see approach 2 above). As we are using a simulation framework, we know the underlying metabolic network and can easily determine neighboring metabolites for ratio selection. Since in reality most metabolic networks are not fully annotated and the PW-NBR approach may not be applicable, we also tested the third method (GGM-NBR) using reconstructed networks on the basis of the simulated steady state metabolite concentrations. The results for different network topologies are shown in Figure 2.7. For linear cascades, PW-NBR and GGM-NBR show similar perfor-

mance results. For more complex, branched reaction networks the prediction accuracy of PW-NBR (Figure 2.6) is slightly better compared to GGM-NBR (Figure 2.7). Estimating the network-based on the metabolomics data as done in the GGM-NBR approach here still performs better than the all ratio approach. For some cases with negative effect size changes ($\Delta ES = -1$) and small to medium sample sizes ($N = 50, 500$ and 1000), Gaussian Graphical models were not able to reconstruct the right underlying reaction network. Therefore the affected metabolite ratio was not part of the candidate set, resulting in a weaker prediction accuracy.

The simulation study shows that preselecting ratio candidates based on metabolic network information in most of the cases only improves the ratio-SNP predictions. The overview of *in silico* results in Figures 2.6 and 2.7 reveals that using network-based metabolite ratios improves the prediction of SNP-reaction associations especially for scenarios with small effect sizes ($\Delta ES = 0.1$ and 0.4). The NBR improvements can especially facilitate the detection of small effects in studies with small sample numbers, when increasing the sample size is a limiting factor.

2.3 Metabolite ratios significantly associated to specific SNPs are also closely connected in metabolic networks

We have shown on simulated metabolomics data that using network information about metabolite dependencies improves the analysis of genetically-influenced metabolotypes. Next we tested the NBR approach on metabolomics and genotyping data from the German population study KORA [95] (Kooperative Gesundheitsforschung in der Region Augsburg), previously published in a genome-wide association study [237]. After quality control and stringent filtering the dataset contained measurements of 218 metabolites and 655,658 genetic variants.

As discussed above, metabolite pairs whose ratio is significantly associated to a SNP should be closely connected in metabolic networks. In order to test this, we decided not to use networks based on pathways from databases due to missing or incomplete pathway annotations of many metabolites (122 out of the 218 metabolites could be mapped to

GGM-NBR compared to all ratios approach

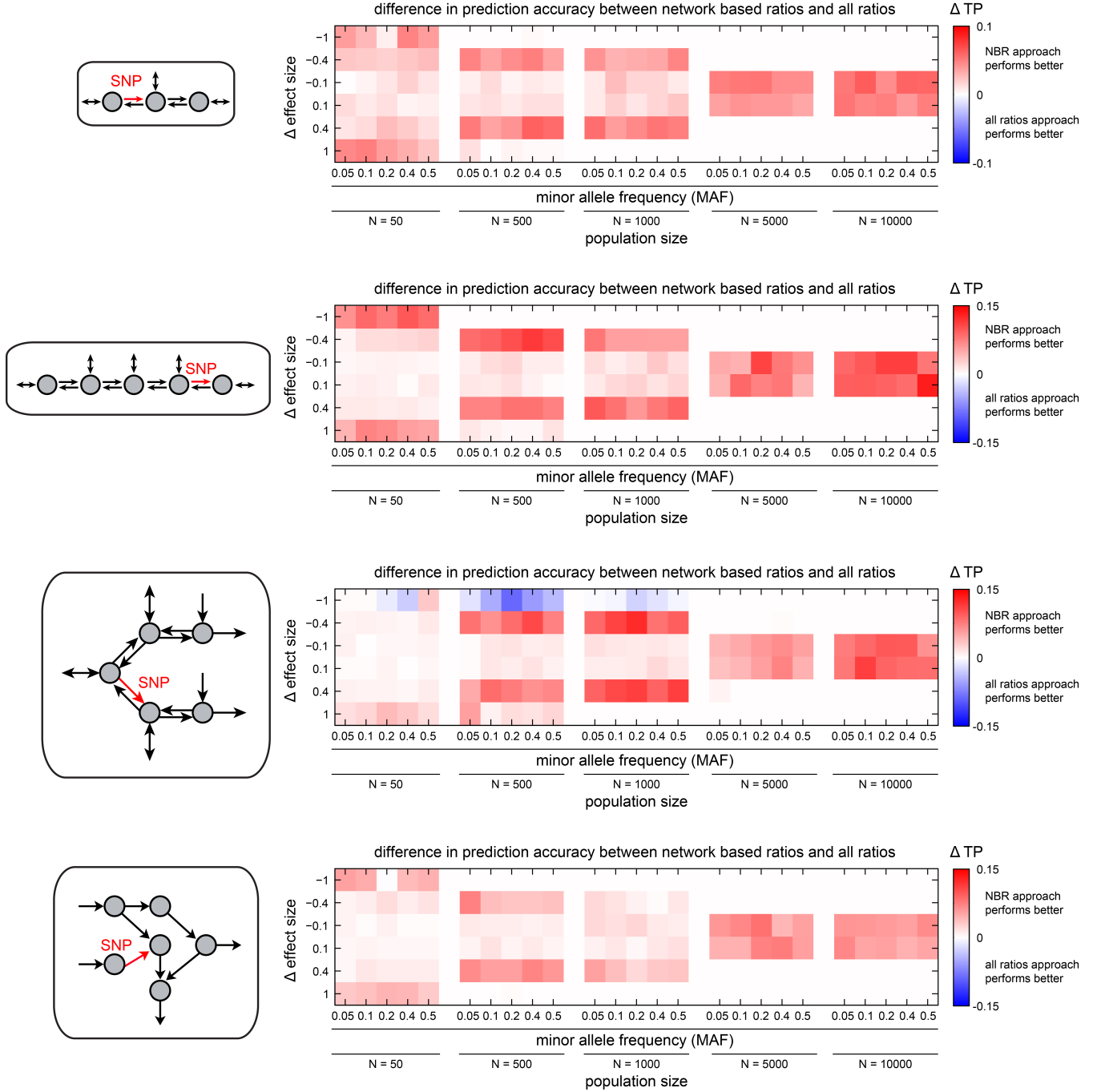


Figure 2.7: Comparing the prediction accuracy between the all ratios approach and the network-based metabolite ratio approach (ratios are selected on the basis of reconstructed metabolic networks using Gaussian Graphical modeling, GGM-NBR).

KEGG, BIGG and EHMN [113, 149, 219]). Instead we used Gaussian Graphical modeling (GGM) which allows to infer a metabolite network for all 218 measured compounds. Briefly, each edge in the network corresponds to a partial correlation coefficient above a certain threshold. Partial correlations represent pairwise correlations between metabolites after the confounding effects of all other metabolites and covariables have been removed. This approach has previously been shown to reconstruct pathways from blood serum metabolomics data in the same cohort [134, 136]. Further information about the procedure, the metabolomics dataset and the obtained GGM can be found in [134].

Figure 2.8A shows the network representation of partial correlations in the GGM. Here metabolites which belong to a significant ratio pair are marked red. We observe a clear grouping of pairs of red nodes in the network. For instance, the amino acids leucine, valine and glutamine are closely connected within the GGM, and are also part of ratios which are significantly associated to a SNP. We further asked whether metabolite pairs, which are both affected by the same genetic variant, are also closely connected in the metabolic network. To address this question, we compared the distribution of all pairwise metabolite shortest path lengths with the distribution of shortest path lengths between metabolite pairs whose ratio was significantly associated to a SNP (Figure 2.8B). To calculate the shortest paths, partial correlation coefficients were transformed to distance measures such that high partial correlation values then have low distances, meaning they are closely connected, and low partial correlation values are far apart (see methods section 2.1).

Significantly associated metabolite pairs tend to have smaller shortest path distances, i.e. higher partial correlation coefficients, compared to all shortest path distances in the GGM. The mean distance between all metabolite pairs is 1, reflecting in our distance measure that most metabolites are not interconnected and have partial correlation coefficients close to 0. On the other hand, the average distance for metabolite pairs with significant ratio-SNP associations is 0.9 and thus more closely connected. ROC analysis [67] was used to quantify this separation, resulting in an AUC score of 0.84. In order to test whether this finding was only observed by chance or does indeed depend on the metabolite network structure, we compared our results to results obtained from randomized networks yielding AUC scores of 0.55 ± 0.04 (empirical p-value $< 10^{-7}$). The ROC-analysis results for the original and randomized GGM network are shown in

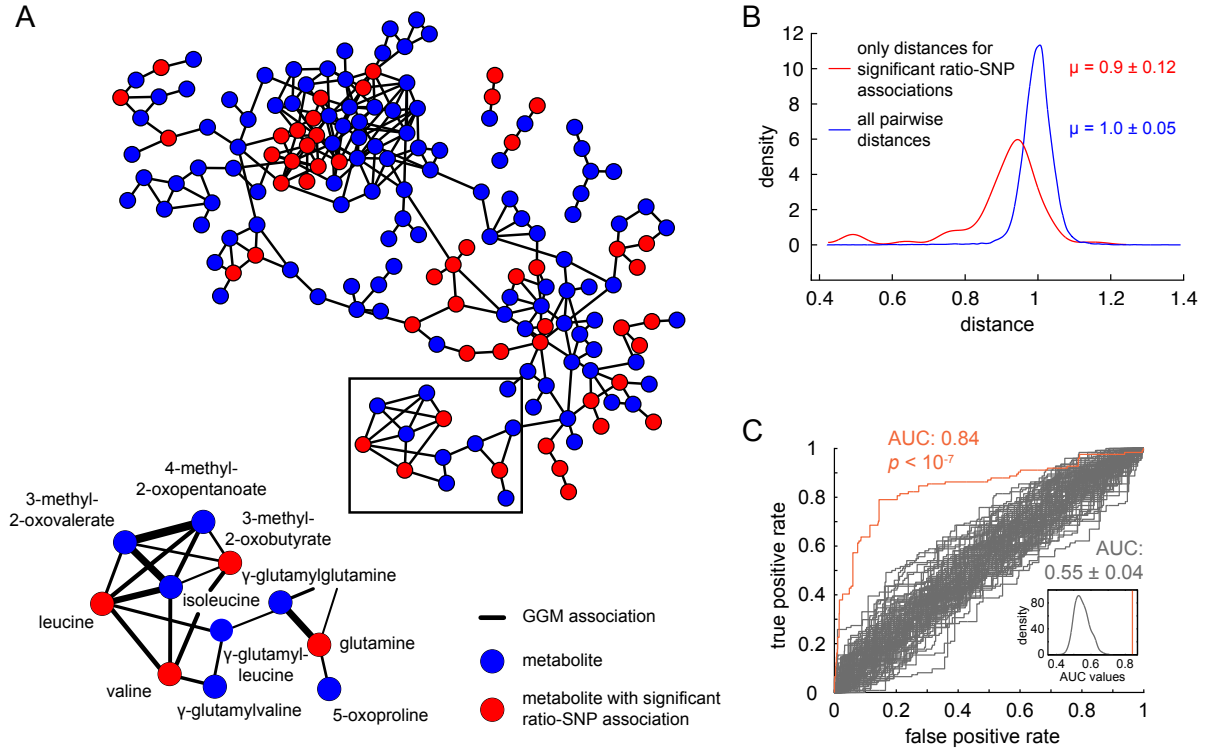


Figure 2.8: Analysis of GWAS with metabolite traits in the context of metabolic networks. Metabolite ratios that are significantly associated with specific SNPs are also closely connected in reconstructed metabolic networks. **A:** Network representation of a Gaussian Graphical model (GGM) reconstructed from large-scale metabolomics data as shown in [134, 136]. Nodes represent metabolites and edges represent partial correlation values higher than 0.15. Zooming into the network reveals that the reconstruction puts metabolically related metabolites in a network context. Metabolites which belong to a ratio pair that is significantly associated to a SNP as reported in [237] are colored red. Line widths represent partial correlation strengths. **B:** Metabolite pairs, which are both affected by the same genetic variant, are also closely connected in the metabolic network. This can be seen using partial-correlation based shortest path distances between metabolites in the GGM. Compared to all distances in the GGM, significantly associated metabolite pairs tend to have smaller distances, i.e. higher partial correlation coefficients. **C:** ROC analysis of the distance separation seen in **B** with an area under the curve (AUC) of 0.84 (orange line). Compared to random networks (grey lines), the distance separation is highly significant (empirical p-value $< 10^{-7}$, see also the small histogram) and thus depends on the underlying GGM network used for the distance analysis.

Figure 2.8C. This highly significant non-random AUC shows that most of the significantly associated metabolite pairs are in close distance. Our findings further suggest that we can use metabolic network information to preselect metabolite pairs for association studies of genetically-influenced metabolotypes.

2.4 Network-based metabolite ratios facilitate the analysis of GWAS results by integrating genomic and metabolomics network information

The results from the toy simulation (see Section 2.2) and the overall analysis of reported metabolite ratio-SNP associations (see Section 2.3) demonstrated that network information can be used to select biologically meaningful ratio candidates for genetic association studies. In the following, we will address the question how to use this information in order to improve the analysis and interpretation of genetically-influenced metabolotypes from mGWAS data. Since we want to focus our analysis on association hits at genetic locus level, we combined ratio-SNP associations that are within linkage disequilibrium of 0.8 or higher and only report the strongest hit (see methods section 2.1). Analogously to our simulation study, we used three approaches to select metabolite ratios for further association tests (see Figure 2.3): all ratios, PW-NBR and GGM-NBR (see Figure 2.9A for a comparison of the results).

For the selection of meaningful network-based metabolite ratios we first used a pathway-based network (PW-NBR) that was built by combining information from KEGG, BIGG and EHMN [113, 149, 219]. Since not all metabolites are annotated in these databases, the network contains only 122 out of the 218 originally measured metabolites. Contrary to the *in silico* simulation study shown in Figure 2.5, we do not have the full information about the true underlying reaction network to apply the PW-NBR approach for the mGWAS data set. We accounted for possible missing network connections by considering not only directly connected metabolites as ratio candidates, but also those with a network distance of one or two steps. The PW-NBR approach reveals only few significant SNP-

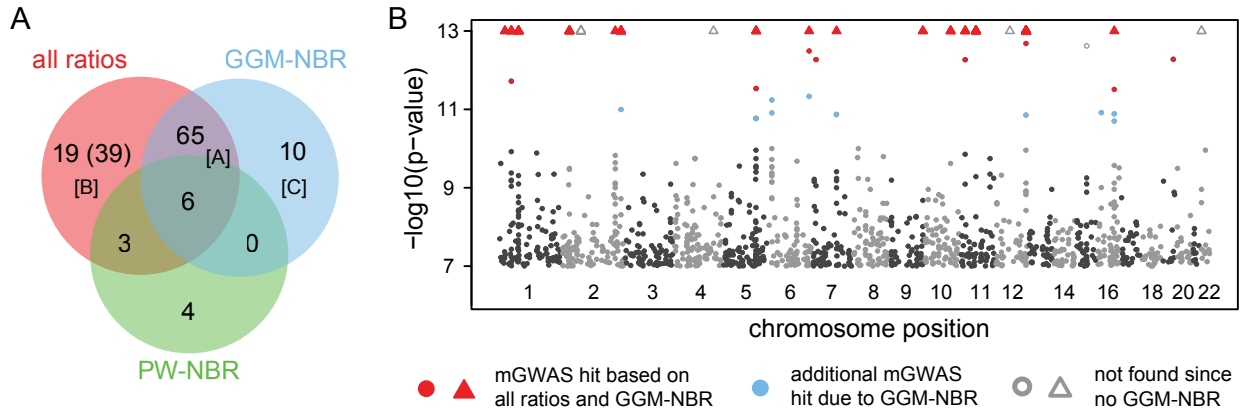


Figure 2.9: Comparison between results for different ratio candidate sets. **A:** Network-based ratio analysis yields similar associations compared to the all ratios approach. Using reconstructed networks (GGM-NBR) performs much better compared to pathway-based networks (PW-NBR) due to the incomplete annotation of many metabolites. [A]: The overlap between the all-ratio and GGM-NBR approach associations is remarkably high. [B]: 20 out of 39 hits not identified by the network approach can be explained by pathway analysis of the underlying GGM network (see also Figure 2.11 for two examples). [C]: Due to the reduced number of association tests and the resulting less stringent Bonferroni significance level, GGM-NBR reveals additional associations (see also Figure 2.12). **B:** Manhattan plot of the results revealed by all ratios and GGM-NBR approach. The strength of association for metabolite ratios is indicated as the negative logarithm of the p-value of the linear model. Only ratio-SNP associations with p-values below 10^{-7} are plotted. Triangles represent ratio-SNP associations with p-values below 10^{-13} . Same ratio-SNP associations that are within linkage disequilibrium of 0.8 or higher are combined and only the strongest hit is shown. Significant mGWAS hits that were found by the all ratios and GGM-NBR approach are marked in red (threshold after Bonferroni correction $\alpha = 3.22 \cdot 10^{-12}$). Associations which are not found by GGM-NBR are colored in grey, while additional GGM-NBR results are marked as blue dots (threshold $\alpha = 2.01 \cdot 10^{-11}$).

ratio associations (13) and the overlap with the all ratios approach is rather small (9 out of 113).

The GGM network on the other hand is purely data-driven and has the advantage of obtaining network dependencies for all 218 measured metabolites. The network was built by taking only metabolite pairs into account that showed an absolute partial correlation score of 0.1 or higher, corresponding to a cutoff of $p < 8.1 \cdot 10^{-5}$ for the p-value of the partial correlation. Similar to the PW-NBR analysis, we also accounted for missing connections by selecting metabolites as ratio candidates that were connected in the network via one or two steps. The GGM-NBR approach reveals 81 significant SNP-ratio associations, which highly overlap with the results from the approach of taking all possible ratios. We also tested other partial correlation cutoffs (see Figure 2.10) to assess the impact of the GGM-NBR parameters. Other parameter settings either revealed less overlap between the associations found by all ratios and GGM-NBR or this overlap was not significant when compared to a distribution of overlaps derived from random networks.

The comparison between all ratios, PW-NBR and GGM-NBR is shown in Figure 2.9A. The GGM-NBR approach yields considerably more significant ratio-SNP associations compared to the PW-NBR (81 vs. 13). This results from incomplete or missing annotations in pathway databases for almost 100 metabolites. In contrast, the full network information can be obtained for all measured metabolites using Gaussian Graphical modeling. Though the overlap between all ratios and GGM-NBR results is remarkably high (71 cases, set [A] in Figure 2.9A), there are some associations which are not observed using GGM-NBR (set [B], 39 cases). We inspected these cases in more detail by asking whether we could explain these effects by other effects of related metabolites. We hypothesized that in many cases the same underlying factor (e.g. genetic variation in one enzyme) influences metabolites in subsequent or neighboring reaction paths. To test this hypothesis, we calculated the shortest paths within the GGM network between the two metabolites of a not observed ratio-SNP association. On these paths we checked if there are other ratio pairs associated to SNPs in close genetic distance to the original ratio-SNP association. Two examples are shown in Figure 2.11, one related to fatty acid metabolism and one related to sugar metabolism. Two SNP variants in the locus of ACADM, an enzyme of mitochondrial fatty acid beta-oxidation, for example are associated to different

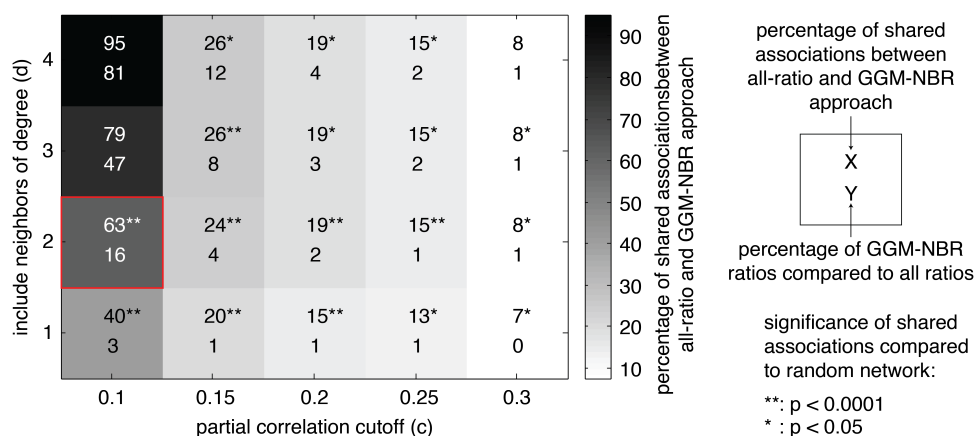


Figure 2.10: Evaluation of GGM-NBR parameter settings. Different parameter values for the partial correlation cutoff (c) and neighbor degree (d) are evaluated for the selection of ratio candidates from GGM reconstructed metabolic networks based on KORA metabolomics data. A neighbor degree d of 2 implies that metabolites, which are connected via one or two steps in the network, are chosen as ratio candidates. Depending on parameter values for c and d , different metabolite ratio candidates are selected from the GGM network. The overlap between these selected GGM-NBR candidates and all possible ratios is denoted by the top number in each box (X). The statistical significance of this overlap was assessed by comparing it to the distribution of overlaps derived from random networks. The bottom number in each box (Y) displays the fraction of GGM-NBR candidates compared to all ratios. The red box marks the parameter set (partial correlation cutoff 0.1, neighbor degree 2) which was used for all presented GGM-NBR results in Section 2.4. For this setting, the overlap between all ratios and GGM-NBR associations of 63% was highly significant ($p < 0.0001$), while the GGM-NBR candidate set contained only 16% of all possible metabolite ratios. Other parameter settings either revealed less overlap between the associations found by all ratios and GGM-NBR or this overlap was not significant.

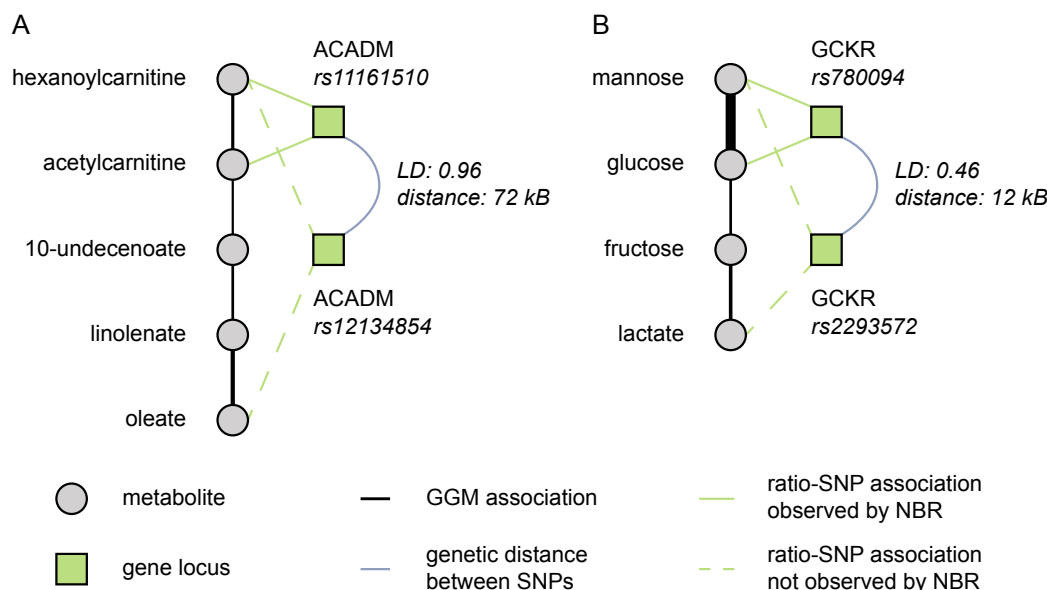


Figure 2.11: Metabolic network information reveals similar effects for associations within pathways. Analyzing ratio-SNP associations within GGM-based metabolic paths helps to group results with respect to their biological mechanism. Two examples are shown for fatty acid metabolism (**A**) and sugar metabolism (**B**). Genetic variants often affect related metabolite pairs. Two SNP variants in the ACADM locus for example are associated with different metabolite ratios. The respective metabolites are closely connected in the GGM network. Genetic effects within the ACADM locus thus have an impact on several metabolite concentrations within certain metabolic pathways. Line widths represent strength of partial correlations in GGM networks.

metabolite ratios (hexanoylcarnitine/acetylcarnitine and hexanoylcarnitine/oleate). On the basis of the data-driven reconstructed network, we can see that these metabolites are closely related. The effect of a genetic variant might affect specific metabolites, but also alter other metabolite concentrations within a pathway, both detected in mGWAS results. Using the GGM network context thus helps to group and interpret these pathway effects, especially if no pathway dependencies for the metabolites of interest can be obtained from databases. In total we found 20 associations that could be grouped with other associations within the same metabolic pathway, reducing the number of all ratios only associations from 39 to 19. Hence by taking the network connections of metabolites into account we do not miss hits, but rather find the more direct associations, which point to the underlying biological mechanism. The network-based analysis thus allows for a systematic evaluation of association results within metabolic pathways.

GGM-NBR also reveals 10 new ratio-SNP associations that were not found using the all ratios approach (set [C] in Figure 2.9A and Manhattan plot in Figure 2.9B, see also Table 2.1). This results from a higher Bonferroni significance level ($3.22 \cdot 10^{-12}$ for all ratios compared to $2.01 \cdot 10^{-11}$ for GGM-NBR), since fewer association tests are performed in the NBR-GGM case. One example for a genetically-influenced metabotype in the leucine metabolism is shown in Figure 2.12. Using both the all ratios and GGM-NBR approach, the ratio isovalerylcarnitine/isovalerate was found to be associated to the OCTN2/SLC22A5 locus, which codes for an organic cation transporter with short-chain acyl esters of carnitine as substrates [128]. GGM-NBR analysis additionally revealed an association between isovalerylcarnitine/leucine and a SNP in the ACSL6 locus. ACSL6 catalyzes the formation of acyl-CoA species, possibly also isovaleryl-CoA, which is a degradation product of leucine but was not measured in the mGWAS study. While these associations might be based on indirect effects as isovalerylcarnitine is the transport form of isovaleryl-CoA, the network context in the GGM-NBR analysis helps to understand the interplay between different association loci and metabolite ratios. Table 2.1 displays further associations that were only found using GGM-NBR.

Evaluation of the NBR approach on independent GWAS results

As described above, the metabolomics data from the KORA study was used for the mGWAS analysis and also for the reconstruction of metabolic networks in the process of the GGM-NBR analysis. We asked whether we could use study-specific NBR candidates also for the analysis of mGWAS results from a different study cohort. We addressed this question by evaluating our NBR approach on independent mGWAS results from 1,052 participants in the British TwinsUK study, published in [237]. Here the published GGM-network information was based on the metabolomics measurements in the KORA study [134], that we used for previous NBR analyses, while the mGWAS ratio-genotype associations were derived from the TwinsUK data. A comparison of the analysis based on different ratio sets is shown in Figure 2.13. PW-NBR only reveals three association hits, while both the all ratios and GGM-NBR approach find 41 and 27, respectively. Again the GGM-NBR matches the initial all ratios results as it was also shown for the KORA mGWAS data (see Figure 2.9A) and the overlap between all ratios and GGM-

Locus SNP id	Metabolite ratio	p-value	beta	Functional interpretation
ZNF655 rs1581492	androsterone sulfate/ dehydroisoandrosterone sulfate	$1.35 \cdot 10^{-11}$	-0.119	gene encodes for a zinc finger protein; potential link to regulatory elements; ratio also associated to AKR1C isoforms (involved in androgen metabolism)
HEATR7B1 rs10203853	bilirubin (E;E)/oleoylcarnitine	$1.01 \cdot 10^{-11}$	-0.047	potential link to regulatory elements; ratio also associated to UGT1A, which has bilirubin as a substrate
COX6A1 rs2076022	butyrylcarnitine/propionylcarnitine	$1.40 \cdot 10^{-11}$	-0.046	COX6A1 is a terminal oxidase in mitochondrial electron transport; fatty acids are transported into mitochondria as acylcarnitines; ratio also associated to ACADS locus (observed by both all ratio and GGM-NBR)
ELOVL2* rs9393903 and rs3734398	docosahexaenoate (DHA; 22:6n3)/ eicosapentaenoate (EPA; 20:5n3)	$1.23 \cdot 10^{-11}$ and $5.76 \cdot 10^{-12}$	-0.030 and -0.026	EPA is substrate of ELOVL2, DHA biochemically related by desaturase reactions
SLC7A6 rs6499172	acetylcarnitine/glutaroyl carnitine	$1.30 \cdot 10^{-11}$	-0.04	SLC7A6 is involved in the transport of amino acids and also associated to glutaroyl carnitine/lysine ratio; possible pathway interaction
PRMT7 rs2863978	acetylcarnitine/glutaroyl carnitine	$1.99 \cdot 10^{-11}$	-0.034	PRMT7 is a arginine methyltransferase for protein modification; potential link to regulatory elements
SLC22A1 rs456598	gamma-glutamylvaline/ isobutyrylcarnitine	$4.65 \cdot 10^{-12}$	0.064	SLC22A1 is a transporter for many organic cations; SNP associated to serum concentrations of total cholesterol and low-density lipoprotein cholesterol [240]
ACSL6 rs10040809	isovalerylcarnitine/leucine	$1.70 \cdot 10^{-11}$	-0.033	ACSL6 catalyzes the formation of acyl-CoA species; isovaleryl-CoA is involved in leucine metabolism (see also Figure 2.12)
PDXDC1 rs7200543	1-eicosatrienoylglycerophosphocholine/ 1-linoleoylglycerophosphocholine	$1.22 \cdot 10^{-11}$	-0.035	Association suggests that PDXDC1 is involved in the glycerophosphocholine metabolism

Table 2.1: List of all additional SNP-ratio associations found using the GGM-NBR approach. The associations between SNP rs9393903 and DHA/EPA ratio, as well between rs7200543 and 1 eicosatrienoylglycerophosphocholine/1-linoleoylglycerophosphocholine have been reported previously. However, this result was only obtained after increasing the sample size by combining two cohorts in a meta-analysis of two GWAS studies. Beta and p-value as reported in [237] from a linear regression model based on additive genetic effects (using log10-scaled metabolic traits, genotype is coded as 0-1-2, major-hetero-minor genotype with beta being the slope of the linear model). *: The two reported SNPs are both in the ELOVL2 locus, but with linkage disequilibrium smaller than 0.8 (0.398) and therefore not combined. See also Figure 2.12 for an example of a newly predicted association.

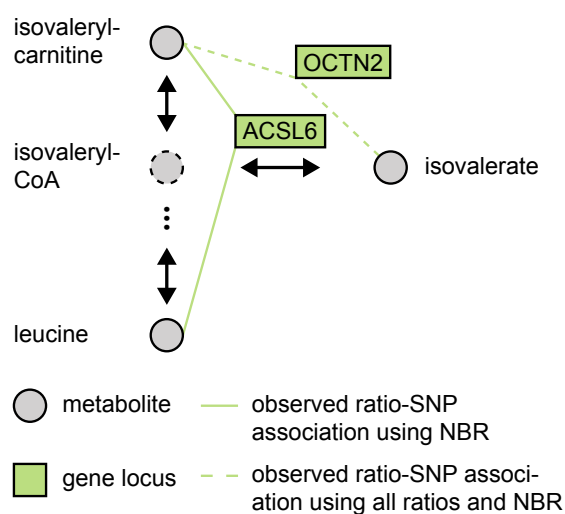


Figure 2.12: Example of an additional association found by the NBR approach. Both the all ratios and NBR approach find an association between isovalerylcarnitine/isovalerate and the OCTN2/SLC22A5 locus (rs274570), which codes for an organic cation transporter. Additionally, NBR-GGM analysis revealed an association between isovalerylcarnitine/leucine and a SNP in the ACSL6 locus (rs10040809). ACSL6 catalyzes the formation of acyl-CoA species like isovaleryl-CoA, which is a degradation product of leucine but was not measured in the mGWAS study (dashed circle). Metabolite relationships, which are obtained from known biochemical pathways or GGM networks, allow for a better understanding and interpretation of indirect effects and observed ratio-SNP associations. See Table 2.1 for a full list of all additional associations found by the NBR approach.

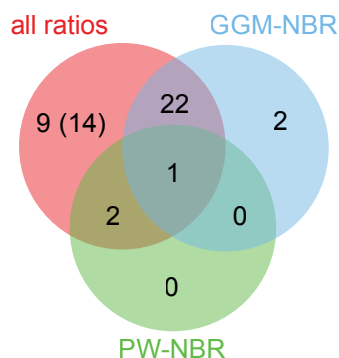


Figure 2.13: Comparison of results for different ratio sets based on Twins UK association data. The KORA GGM network was used to analyze mGWAS results of the Twins UK study, published in [237]. Similar to the results of KORA association data presented in Figure 2.9A, the overlap between all ratios and GGM-NBR results is rather high and GGM-NBR reveals two additional hits. Pathway analysis on the underlying GGM network revealed that 5 out of 14 hits, which are not identified by the GGM-NBR approach, can be explained by other associations within the same metabolic pathway.

NBR results is rather high. Due to the reduced number of association tests GGM-NBR additionally reveals two significant ratio-SNP associations. Using the pathway analysis of the underlying GGM network showed that 5 out of the 14 hits, which were not identified by the GGM-NBR approach, can be explained by other associations within the same metabolic pathway. The GGM network context points to the more direct biological mechanism and facilitates the interpretation of mGWAS results. It is important to note at this point that for our evaluation on TwinsUK association data we did not use TwinsUK raw metabolite concentration data for the network reconstruction but relied on the published KORA GGM. Our findings suggest that these data-driven networks might be used in a general context for the analysis of further metabolomics studies about genetically-influenced metabolotypes.

2.5 Discussion

Genome-wide association studies with metabolite ratios as quantitative traits have deepened our understanding of the complex relationship between genetic variants and observed phenotypes. It has been shown that ratios between metabolite concentrations pairs reduce the overall biological variability in population data resulting in robust statistical

mQTL associations [190]. In previous studies, metabolite ratios were either manually selected with respect to specific enzymatic reactions [117] or all possible ratio combinations were used [81, 104, 237]. Especially due to the large number of all possible ratios for studies with many metabolites it is important to narrow down the number of association tests. We argue that the proper selection of ratio candidates based on metabolic network information will improve the analysis of association studies with metabolic traits from a statistical, computational and interpretational point-of-view.

In this thesis we propose to choose biologically meaningful metabolite ratios based on metabolic networks for further association tests. Before applying network-based metabolite ratios (NBRs) on human population GWAS data, we used simulated reaction networks. Our model simulates differences in metabolite levels, which result from genetic variation affecting enzyme activities. Such effects have been reported for several SNPs in the ACE structural gene and the ACE activity [40]. It is important to acknowledge at this point that the *in silico* model is obviously an oversimplified model of gene-metabolite interactions. The primary goal of the presented analysis was to test our hypothesis in a well-defined and comprehensible environment before going to noisy experimental data. Our *in silico* results show that the NBR approach is applicable for small sample size studies and, even more important for practical applications, for genetic variants with small effect sizes.

We further analyzed mGWAS results from two different study cohorts as an additional evaluation of the NBR approach. The main focus of this validation was not to strictly replicate observed genotypic effects, but rather to show that the NBR method is not restricted to a specific data set. Initially the mGWAS results were obtained by using all possible metabolite ratio combinations as traits. We compared the associations detected using all ratios with associations observed after testing only ratio candidates derived from metabolic networks. Data-driven metabolic networks (GGM-NBR) gave similar results as the all ratios approach, while networks obtained from pathway annotations in the literature (PW-NBR) could not reveal many associations. The limited results from PW-NBR are certainly based on the sparseness of metabolite annotations and network information in databases like KEGG, BiGG and EHMN. Data-driven reconstruction methods provide the opportunity to measure relations between all measured metabolites. It is important to acknowledge that results from statistical inference methods like

Gaussian Graphical models should not be confused with a perfect reconstruction of metabolic pathways [21]. For instance, if intermediate metabolites cannot be detected, connections in reconstructed networks do not necessarily represent direct biochemical pathway reactions. With advanced metabolite detection techniques and network reconstruction methods, both the annotation and the data-driven pathway information will further improve. Since our NBR approach relies on the informative content of metabolic networks, the two network sources can be combined to enhance our understanding of metabolism by finding more genetically-influenced metabotypes.

Narrowing down the size of ratio candidate sets is also important for small, phenotype-specific studies. For example, studies investigating rare variants or small effect sizes often have to deal with small case numbers. For such studies it is essential to reduce the number of tests in order to improve the statistical power and lower computational demands. Moreover, advanced metabolomics methods will soon allow for the detection of several thousand metabolites. At this point it will not be feasible anymore to test all possible ratio combinations against genetic variants in order to find genetically-influenced metabotypes. Figure 2.14 displays the relationship between the number of measured metabolites and the number of ratios that have to be tested using the all ratios and the NBR approach. The number of ratios for association tests increases much faster for the all ratios case compared to the GGM-NBR approach. For upcoming mGWAS studies with larger metabolomics panels, the significance cutoff for multiple testing correction will be more stringent, allowing only the detection of rather strong signals. Selecting ratios based on metabolic networks (GGM-NBR) substantially reduces the number of tests to be made. In addition, we could show that genetic effects in one locus have an impact on the concentration levels of biochemically related metabolites. It is important to understand at this point that measuring the distance between metabolites is not straightforward, both in reconstructed and literature-based metabolic networks [65, 192]. Nevertheless, the network context helps to group associations which are related to the same underlying biological mechanism. For instance markers, which have different distributions within the population but which are associated to the same causal, yet unknown variant, will show associations to various metabolic traits. With the network context these associations can be grouped together on pathway levels which allows for a systematic evaluation and interpretation of association results.

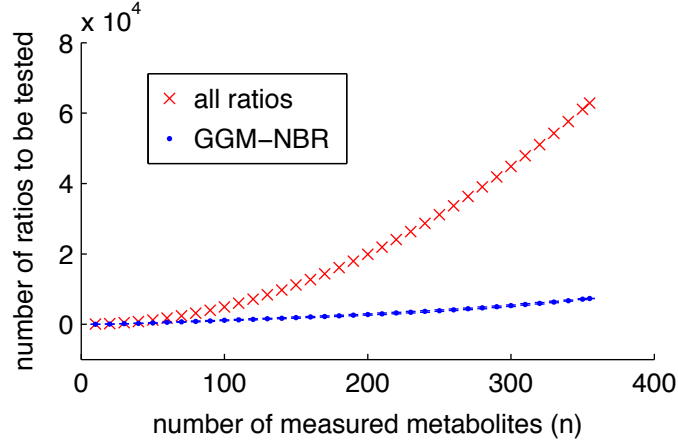


Figure 2.14: Estimating the number of ratio candidates for the all ratios and GGM-NBR approach. The amount of ratio candidates that are tested for association depends on the number of measured metabolites (n). The number of all ratios (red crosses) is calculated from n , NBR ratio candidate numbers (blue dots) were estimated based on KORA metabolomics data (mean values out of 1000 repeated estimations, see Methods section 2.1).

The NBR approach may also be combined with methods accounting for the inherent correlation between SNPs due to linkage disequilibrium [73, 178], thus reducing the number of both ratios and SNPs for multiple testing correction. The presented approach is not restricted to association studies with metabolic traits and can be extended to other quantitative *omics* data, also in case-control studies. For such studies the sample size is usually much smaller and our preselection of ratios could improve the statistical power. Moreover, NBRs can be used for other quantitative biomolecular data such as gene expression measurements or epigenetic modifications [191]. Here the interactions between gene products can be both inferred from data or obtained from biological pathways, well-established protein-protein or gene-regulatory networks [14, 28, 34].

GWAS with metabolic traits have deepened our understanding of genetic effects on metabolic functions. For such large-scale data, taking metabolic network information into account can be of great benefit for the analysis and interpretation of association results. Our NBR approach reveals nearly the same associations compared to the “all ratios” approach, while lowering computational demands. Using NBR additionally allows for the detection of weaker effects, since considering only biologically meaningful ratio candidates increases the statistical power. For upcoming studies with large-scale metabolomics data and small sample numbers, our NBR approach provides a valuable

tool to increase the statistical power, lower computational demands and facilitate the interpretation of the results. Network-based analysis will then help to better understand the complex interplay between individual phenotypes, genetics and metabolic profiles.

The network-based metabolite ratio approach is not restricted to genetic association data. For instance, in Chapter 3 we will present a study on the dynamics of human metabolism, where we analyzed subject-specific, time-resolved metabolomics data from a fasting period of 36 hours. For this analysis we apply the ratio approach on the level of a single pathway and use a fatty acid beta-oxidation model to infer metabolite ratios reflecting enzymatic activity. We show that, compared to single metabolite levels, model-driven metabolite ratios are better markers for the individual's metabolic capacity and facilitate the explanation of the observed interindividual variation in metabolomics profiles.

Chapter 3

Modeling metabolic pathways under steady state conditions in a human challenging study

In the previous chapter, we discussed the interplay between genetic variants and individual metabolic profiles. Based on population data from genome-wide association studies, we analyzed the link between genes coding for enzymes and the levels of corresponding substrate or product metabolites. Besides genetics, a variety of intrinsic and extrinsic factors defines metabolic phenotypes (see Chapter 1.1). For instance, age [279], gender [163], physical exercise [37, 193, 275] and diet [96] influence human metabolic profiles. Metabolite data can also be used as biomarkers for diseases like cancer or diabetes [84, 97, 112, 138]. Selecting markers for distinct pathophysiological states can be complex, but it is even more difficult to determine the biological boundaries between health and disease [282].

Challenging protocols like oral glucose tolerance tests (OGTT) can unmask alterations in metabolism related to early states of chronic diseases. In such a test, the individual flexibility and response capacity to perturbations from environmental and physiological stimuli like diet and exercise is assessed. To date, most metabolomics research has

been done for samples representing fasting conditions and only few studies report time-resolved measurements of individual reactions to external and internal challenges. For instance, a study with prolonged fasting conditions revealed new markers in human plasma and urine samples for this catabolic state [213]. Shaham *et al.* used mass spectrometry methods to analyze plasma samples of healthy and pre-diabetic volunteers undergoing an OGTT [228]. In this challenging setup, metabolites like bile acids and urea cycle intermediates were found to be significantly altered, indicating a role in the context of glucose homeostasis which has not been described before. In a study on the intervention effects of a mild anti-inflammatory drug in overweight human participants, OGTT challenging increased the statistical power and allowed for the detection of subtle metabolic changes [272].

For a better understanding of the human metabolome dynamics in response to environmental stresses and perturbations, a four-day human challenging study (HuMet) has been conducted, in which 15 healthy male volunteers were submitted to six different physiological and nutritional challenges. For analyzing the relationship between phenotypic and metabolite profiles under specific physiological conditions, samples from blood, urine, exhaled air and breath condensate were collected for each volunteer on up to 56 time-points. Metabolic traits like lipids, amino acids and acylcarnitines were measured using mass spectrometry and nuclear magnetic resonance methods. The HuMet study thus provides metabolomics data for more than 2100 individual samples under varying physiological conditions, measured in different body fluids on different analytical platforms.

In collaboration with the groups of Hannelore Daniel, Hans Hauner, Karsten Suhre and Gabi Kastenmüller I contributed to the analysis of the interindividual variation in healthy and phenotypically similar volunteers. In the following, we will give a brief overview about the HuMet study design and discuss volunteer specific metabolite changes induced by physiological and nutritional challenges, which could not be observed on baseline metabolite concentrations. Due to the broad coverage of lipid compounds in the metabolite panel, the focus of our analysis lied on the fasting challenge. Under such conditions mitochondrial beta-oxidation, the catabolic breakdown of fatty acids, is the main physiological process which provides energy to the cell (see Section 1.1 for a detailed description of the pathway). We investigated the individual response to

the fasting challenge by combining time-resolved metabolomics data with knowledge about metabolite connections in the fatty acid oxidation (FAO) pathway. To this end, we developed a mathematical model which approximates the beta-oxidation pathway as a linear cascade of subsequent, irreversible first-order reactions. Furthermore, we will discuss how model-derived parameters resembling the individual metabolic capacity can be used to investigate the relationship between interindividual variation in metabolic profiles and phenotypic characteristics of the study volunteers.

The results reported in this chapter have been published in collaboration with the groups of Hannelore Daniel, Hans Hauner, Karsten Suhre and Gabi Kastenmüller in the following publications:

- Krug S*, Kastenmüller G*, **Stückler F***, Rist MJ*, Skurk T*, Sailer M, Raffler J, Römisch-Margl W, Adamski J, Prehn C, Frank T, Engel KH, Hofmann T, Luy B, Zimmermann R, Moritz F, Schmitt-Kopplin P, Krumsiek J, Kremer W, Huber F, Oeh U, Theis FJ, Szymczak W, Hauner H, Suhre K, Daniel H (2012): The dynamic range of the human metabolome revealed by challenges. *FASEB J.* 26: 260719
This publication is a joint first author work and the content of this paper is also part of another thesis [132]. My contribution to the publication is the development of a model to analyze the metabolomics data, as well the statistical analysis of model readouts and their interpretation.
- Krumsiek J, **Stückler F**, Kastenmüller G, Theis FJ (2012): Systems Biology Meets Metabolism. In: Suhre K, editor. *Genetics Meets Metabolomics: from Experiment to Systems Biology*. New York, NY: Springer New York. pp. 281-313.

* = equal contributions

3.1 Methods

HuMet study design

The HuMet study was designed to assess on metabolome level individual responses to physiological and nutritional challenges. The 15 male study participants were metabolically healthy, with an age ranging between 22 and 33 years (27.8 ± 2.9) and a body mass index (BMI) between 20 and 25 kg/m² (23.1 ± 1.8). The study consisted of two test periods, each lasting two days and two nights (see Figure 3.1), with a break interval of several weeks between study day 2 and 3. A highly-controlled experimental design minimized environmental influences by admitting volunteers to the study unit on the evening before each test block. Participants received standardized meals before and during the test periods. The energy content of each individual meal was adjusted to 1/3 of the volunteer's resting metabolic rate (RMR), which was measured by indirect calorimetry (DeltatracTM Metabolic Monitor, DatexOhmeda, Helsinki, Finland). The four study days included a sequence of six different challenges: a prolonged fasting period of 36 hours (FASTING), a standard liquid diet (SLD), an oral glucose tolerance test (OGTT), an oral lipid tolerance test (OLTT), a physical activity test (PAT), and a cold pressure stress test (STRESS). Samples from plasma, urine, breath gas and exhaled breath condensate (EBC) were collected for each volunteer between every one and four hours. For short-time challenges such as physical exercise, stress and OGTT the sampling intervals were between 15 and 30 minutes.

As the main analysis presented in this thesis covers the fasting period on day 1, we will briefly provide details about this catabolic challenge and the acquired metabolite data. Further information about the overall study design, challenging protocols, data acquisition and quality control can be found in the original publication [133]. Volunteers received a standardized evening meal at 7 a.m. on the day before day 1. Afterwards subjects fasted until 8 a.m. of day 2. During the total fasting period of 36 hours, participants received 2.7 liters of mineral water according to a predefined drinking schedule. For each volunteer, plasma samples were collected every 2 hours between 8 a.m. and midnight. As the main energy source during sustained fasting is the degradation of fatty acids, we focused our analysis on the metabolite profiles of acylcarnitines, which are transport

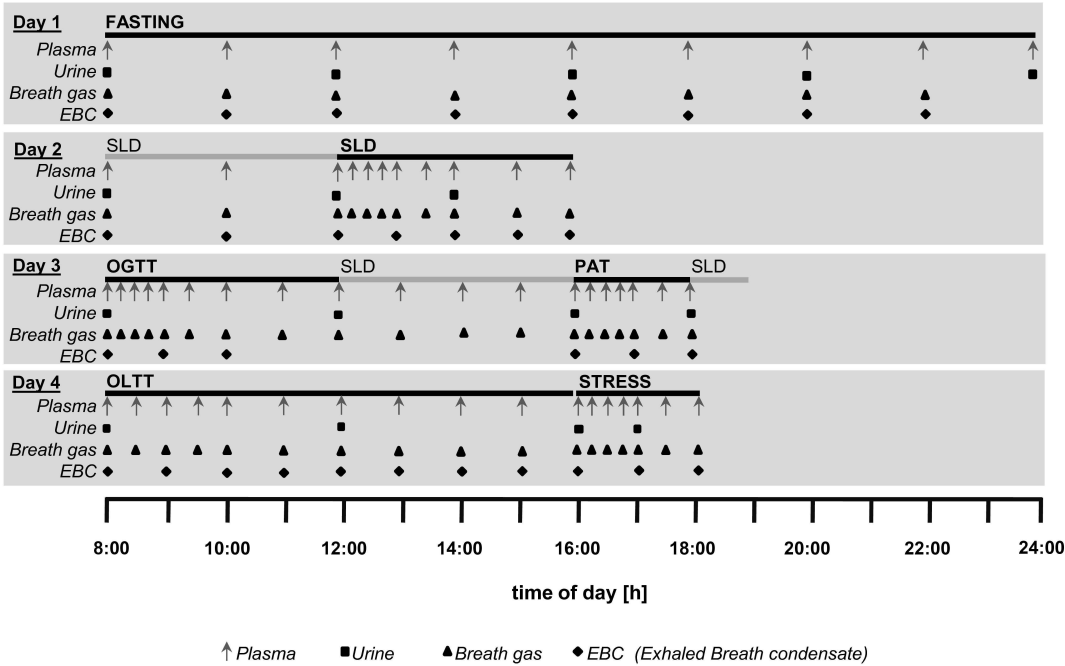


Figure 3.1: Study design of the HuMet study. Within four days, volunteers underwent six physiological and nutritional challenges: fasting, standard liquid diet (SLD), oral glucose tolerance test (OGTT), physical activity test (PAT), oral lipid tolerance (OLT), and a cold pressure stress test. Solid bars indicate duration of each challenge test. Symbols indicate time points of biofluid collection.

forms of fatty acids intermediates. Plasma acylcarnitine concentrations were used as a surrogate for acyl-CoA concentrations in the beta-oxidation pathway [160, 177, 250].

Data acquisition

Metabolites were quantified using targeted tandem mass spectrometry (MS/MS), nuclear magnetic resonance (NMR) and proton transfer reaction mass spectrometry (PTR-MS). MS/MS targeted profiling of metabolites in plasma was performed using the AbsoluteIDQ kit (Biocrates life sciences AG, Innsbruck, Austria) as described previously [6, 104, 210]. The kit allows for the quantification of mainly amino acids and lipid derivatives. The panel contains 14 amino acids; hexose (H1); free carnitine (C0); 40 acylcarnitines, hydroxylacylcarnitines and dicarboxylacylcarnitines; 15 sphingomyelins; 77 phosphatidylcholines and 15 lyso-phosphatidylcholines. Further details about the quantification of metabolites in plasma, urine and breath air using NMR and PTR-MS, as well the measurement of standard clinical chemistry parameters like venous plasma glucose, lactate, insulin and non-esterified fatty acids (NEFA) can be found in the original publication [133]. Quality of the data was evaluated by repeated measurements on different run days. Metabolites which showed a coefficient of variation (CV) $> 25\%$ and compounds with a CV greater than 20% and a significant association (Kendall correlation) to the run day of the measurement were excluded.

Principal component analysis

The metabolite panel of each plasma sample for further analysis consisted of 163 metabolites (132 traits measured by MS/MS, 28 by NMR and 3 clinical chemistry parameters). For evaluating challenge-induced effects, principal component analysis (PCA) was performed on all 163 plasma metabolic traits. Each metabolite variable was scaled to mean 0 and standard deviation 1 for comparing different concentration levels. PCA analysis was performed for two data subsets. To assess the general challenge effects on metabolic traits at each of the 56 sampling time points the metabolite data was averaged over the

15 volunteers (data matrix 56×163). To assess the interindividual variation, all samples for each volunteer and each time point (840×163) were included in the PCA analysis.

Mathematical model for the beta-oxidation cascade

The degradation of fatty acid chains in the mitochondrial beta-oxidation pathway can be seen as a linear cascade model of irreversible first-order reactions, as illustrated in Figure 3.2. This description is based on the linearity of subsequent, central steps in the degradation cascade [16] (see also Section 1.1 for a more detailed description of this biochemical pathway). The four sequentially coupled reactions of each beta-oxidation cycle (oxidation, hydration, a second oxidation and thiolysis) are combined into one central reaction assumed to be irreversible under the given fasting condition. No simultaneous production of fatty acids is assumed due to the antagonistic regulation of catabolic and anabolic pathways [200].

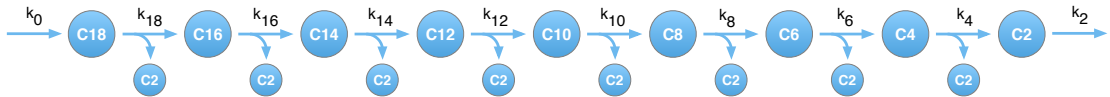


Figure 3.2: Schematic model of the beta-oxidation cascade. Mobilized fatty acids are subsequently degraded in a linear pathway. During each reaction step the carbon chain is shortened by two carbon atoms and C_2 is produced.

The fatty acid C_{18} with 18 carbon atoms is supplied to the beta-oxidation pathway. During each reaction step the carbon chain is shortened by two carbon atoms. This results also in the production of C_2 . According to the law of mass action (see Section 1.4), we can describe each reaction with differential equations. The change of the first metabolite in the chain (C_{18}) over time depends on the supply rate k_0 and the conversion of C_{18} to C_{16} with rate parameter k_{18} :

$$\dot{C}_{18}(t) = k_0 - k_{18} \cdot C_{18}(t)$$

where $\dot{C}_{18}$ represents the time derivative $\frac{dC_{18}}{dt}$. $C_{18}(t)$ is the concentration of C_{18} at time t . Equations for intermediate metabolites C_{16} to C_4 read

$$\begin{aligned}
\dot{C}_{16}(t) &= k_{18} \cdot C_{18}(t) - k_{16} \cdot C_{16}(t) \\
\dot{C}_{14}(t) &= k_{16} \cdot C_{16}(t) - k_{14} \cdot C_{14}(t) \\
\dot{C}_{12}(t) &= k_{14} \cdot C_{14}(t) - k_{12} \cdot C_{12}(t) \\
\dot{C}_{10}(t) &= k_{12} \cdot C_{12}(t) - k_{10} \cdot C_{10}(t) \\
\dot{C}_8(t) &= k_{10} \cdot C_{10}(t) - k_8 \cdot C_8(t) \\
\dot{C}_6(t) &= k_8 \cdot C_8(t) - k_6 \cdot C_6(t) \\
\dot{C}_4(t) &= k_6 \cdot C_6(t) - k_4 \cdot C_4(t)
\end{aligned} \tag{3.1}$$

The change of each intermediate metabolite $C_{16}, C_{14}, \dots, C_4$ in the beta-oxidation cascade (for instance $\dot{C}_{14}$) thus depends on the production term by shortening the preceding metabolite (first part on the right hand side of the equation, e.g. $+k_{16} \cdot C_{16}$) and the conversion to the subsequent metabolite in the chain (second part, e.g. $-k_{14} \cdot C_{14}$).

The last metabolite in the chain, C_2 , is produced during each shortening reaction of $C_{18}, C_{16}, \dots, C_6$, but also after splitting up C_4 . The change of C_2 can be described by the following differential equation:

$$\dot{C}_2(t) = k_{18} \cdot C_{18}(t) + k_{16} \cdot C_{16}(t) + \dots + 2k_4 \cdot C_4(t) - k_2 \cdot C_2(t) \tag{3.2}$$

Biochemical reactions are on timescales between milliseconds and minutes, while upstream physiological processes can be much slower. For systems with two separated time-scales the dynamics of the faster part may run into a steady-state, also referred to as quasi steady-state of the overall system [227]. The quasi steady-state assumption has been applied for instance to approximate the solution of Michaelis-Menten kinetics [19, 127]. Metabolite levels are supposed to be in a quasi-steady state compared to the timescales of upstream regulatory processes due to perturbation events [143]. As the sampling interval in the fasting period was two hours, we assume that for each sampling time point the system has attained a stationary state measured by equilibrium plasma metabolite concentrations.

Under steady state conditions each equation above is set to 0, i.e. metabolite concentrations do not change over time. Due to the coupling in the cascade all single equations in Equation (3.1) can then be reformulated (e.g. $k_{16} \cdot C_{16}(t) = k_{14} \cdot C_{14}(t)$ and $k_{14} \cdot C_{14}(t) = k_{12} \cdot C_{12}(t)$). Substituting the corresponding reformulated equations in Equation (3.2) yields for a specific measurement time point τ

$$k_i(\tau) = \frac{C_2(\tau)}{C_i(\tau)} \cdot \frac{k_2(\tau)}{9}, \quad i = 4, 6, 8, \dots, 18. \quad (3.3)$$

For instance, the rate k_{16} of the conversion $C_{16} \rightarrow C_{14}$ can be deduced from the concentration ratio C_2/C_{16} , multiplied by the factor $k_2/9$, representing the removal rate of C_2 from the system. Note that we obtain for each measurement time point τ an estimation of the reaction rate. The system of differential equations is underdetermined, as it includes 10 reactions with unknown reaction rates $k_0, k_{18}, k_{16}, \dots, k_2$ and 9 equations for the measured compounds $C_{18}, C_{16}, \dots, C_2$. For this reason, we cannot derive a unique solution for each reaction rate, but the ratios of metabolites are proportional to reaction-specific rates, normalized with respect to the removal rate of C_2 .

Correlation analysis between model readouts and phenotypic parameters

The relationship between metabolic traits (both single metabolite levels and metabolite ratios) and anthropometric or biochemical parameters (ABP) was evaluated using Spearman's rank correlation statistics. The correlations between biochemical parameters like blood sugar and insulin concentrations were calculated using metabolite levels of the fasting period. For anthropometric parameters, rank correlation was obtained using the mean metabolite concentrations of the fasting period. The association p-values were corrected for multiple testing by controlling the *false discovery rate* (FDR) at a global significance level of 0.05 [18]. In order to identify model-driven ratios that provide stronger statistical associations than single metabolites, *p-gain* statistics [190] were calculated as

$$\text{p-gain} \left(\frac{M_1}{M_2}, X \right) = \frac{\min(\text{p}(M_1, X), \text{p}(M_2, X))}{\text{p} \left(\frac{M_1}{M_2}, X \right)}$$

with metabolites M_1 and M_2 and their respective model-driven ratio M_1/M_2 , parameter X (e.g. BMI) and $\text{p}(A, B)$ being the FDR-corrected p-value of Spearman's rank

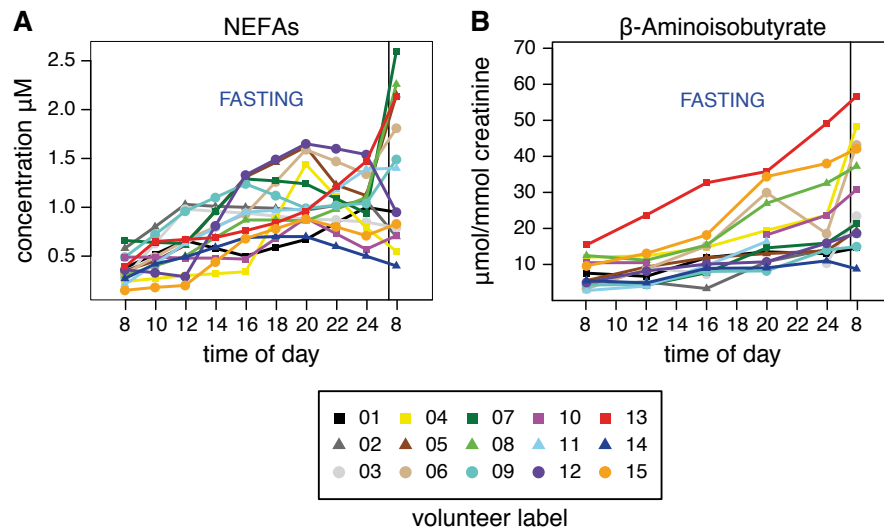


Figure 3.3: Concentration changes of the catabolic parameters non-esterified fatty acids (NEFAs, **A**), measured in plasma, and β -aminoisobutyrate (**B**), measured in urine, represent metabolic alterations caused by strict fasting.

correlation between variable A and B . Cases where the model-based analysis improves statistical correlations yield p -gains greater than 1. All analysis steps were performed using Matlab Version 7.11 (MathWorks, Natick, MA).

3.2 Challenge-induced metabolite changes and interindividual variation

Nutritional and physiological challenges induce changes in time-resolved metabolic profiles

The different challenges in this study induced catabolic and anabolic responses, also reflected in timecourse metabolite profiles. As expected, prolonged fasting for instance led to the mobilization of fatty acids from adipose tissue, reflected by elevated plasma concentrations of non-esterified fatty acids (NEFAs) and an increase of catabolic compounds like β -aminoisobutyrate measured in urine (see Figure 3.3). Plasma concentrations of insulin and glucose were low during fasting and increased during food intake and OGTT

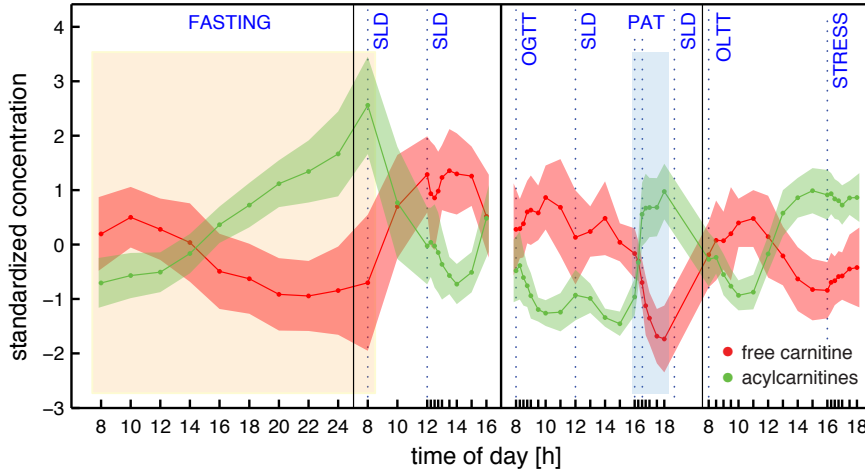


Figure 3.4: Metabolic response to challenges of plasma acylcarnitine levels. Mean plasma concentrations of the 15 subjects at each sampling time point of free carnitine (C0, red) and the sum of all acylcarnitines including acetylcarnitine (green) are significantly anticorrelated (Pearson correlation, $\rho = -0.66$, $p\text{-value} = 2.4 \cdot 10^{-8}$), reflecting switching between anabolic and catabolic metabolic states induced by specific challenges. For a better visualization, metabolite levels were scaled to mean 0 and standard deviation 1. Red and green shaded areas denote standard deviations for the timecourses of 15 subjects. Yellow and blue shaded areas define the fasting and exercise period which is analyzed in more detail in Figure 3.5

(see Figure 3.6). The switching between catabolic and anabolic states is also illustrated by the mirror-like dynamics of free carnitine (C0) and the sum of all acylcarnitines including acetylcarnitine, illustrated in Figure 3.4. During extended exercise and fasting conditions, fatty acids from adipose tissue are mobilized by lipolysis and transported to organs like muscle and liver for the production of metabolic energy [175]. C0 is required for this transport of lipid species across cellular membranes into the mitochondrium. Free fatty acids and free carnitine are converted to acylcarnitines and then translocated to the inner mitochondrial matrix [116]. Thus, C0 levels decrease during fasting and physical exercise, while acylcarnitines concentrations are elevated. This observation is also consistent with a previously reported hypothesis about the role of acylcarnitines as a buffer for beta-oxidation intermediates [177]. Due to increased fatty acid degradation, the spill-over of intracellular acetyl- and acyl-CoA compounds is buffered by the release of respective acylcarnitines into the blood plasma. During anabolic conditions, for instance SLD and OGTT, the opposite dynamics for C0 and acylcarnitines are observed, while markers for anabolism like plasma insulin levels are elevated.

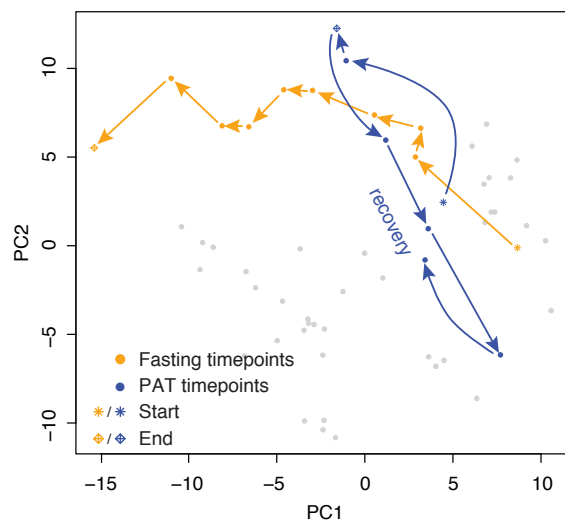


Figure 3.5: Challenge-induced changes in the entire plasma metabolite profile. Principal component analysis (PCA) was performed on the mean plasma concentrations of all 15 volunteers at each of the 56 time points. Physical activity test (PAT) time points (blue) started at 4 p.m. with sampling at 0, 15, and 30 min during cycling and in the recovery phase at 15, 30, 60, and 90 min after cycling. Time points of the fasting challenge (orange) started with a sample taken at 8 a.m. (after an overnight fast), followed by samples taken after further 2, 4, 6, 8, 10, 12, 14, 16, and 24 hours (see also color-shaded areas in Figure 3.4). Specific for the different challenges, the samples are located in a time-dependent trajectory in the metabolic space spanned by the first two principal components (PC1 and PC2).

The different metabolic changes induced by challenges were also studied using multivariate analysis of all plasma metabolites. Principal component analysis (PCA) on the mean plasma concentrations of all 15 volunteers at each of the 56 time points showed challenge-specific trajectories in the time-dependent metabolomics fingerprints. As depicted in Figure 3.5, time-dependent challenge effect were especially apparent for the fasting and exercise (PAT) challenge. The PAT trajectory also illustrates the recovery phase, after which metabolite profiles almost returned to the starting state before the challenge.

Interindividual variation in metabolomics profiles

The individual metabolic responses induced by challenges were highly diverse within the study group. Despite the fact that the volunteers were all male, metabolically healthy Caucasians within a narrow range of age and BMI, especially catabolic challenges like

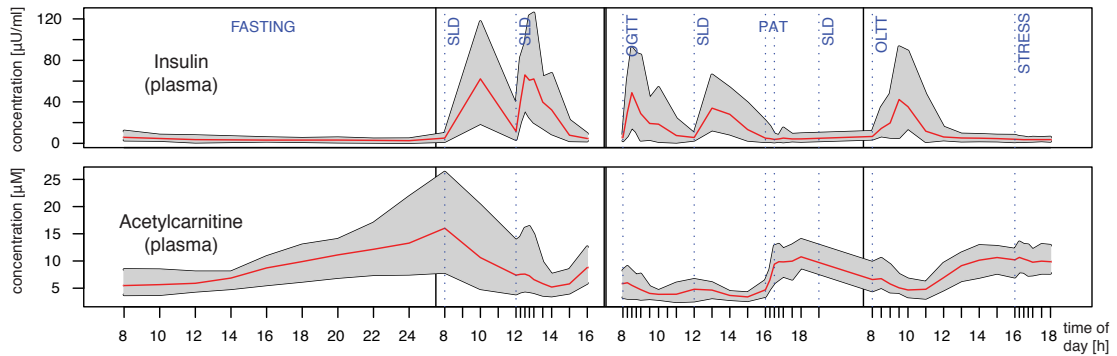


Figure 3.6: Challenge-induced interindividual variation in metabolite profiles. Depending on the challenge type, the interindividual variation in metabolite concentrations is increased or decreased. Red curves denote mean plasma concentration of insulin and acetylcarnitine (C2). The gray-shaded area depicts the range between the minimal and the maximal concentration observed in any participant. Insulin shows strong variation during food challenges, while differences in acetylcarnitine levels are increased during catabolic conditions like fasting.

fasting and physical exercise induced highly different physiological responses observable in plasma metabolomics levels. For instance, while differences in insulin plasma levels were low during fasting or exercise, nutritional conditions (e.g. OGTT, OLTT and SLD) showed large intersubject variation in concentrations during the postprandial state (Figure 3.6). Interestingly, principal component analysis based on the plasma metabolite profiles of all volunteers at all timepoints showed that subject specific samples are grouped together despite large intraindividual variation over the various challenges (see Figure 3.7A and B). Especially the 36 hours fasting period revealed a broad range of plasma acetylcarnitine concentrations between minimal and maximal concentration levels. For instance volunteer 14 (V14) and volunteer 13 (V13) mark the low and high extremes of the study group for acetylcarnitine concentrations in plasma (see Figure 3.7C). At the end of the fasting period, C2 levels of V13 were more than three-fold higher compared to the levels of V14 (25.8 μM vs. 8.1 μM). This increased between-subject variation was also prominent for related metabolites and other sample types like urine and breath air. For instance, metabolic signatures of V13 (e.g. plasma concentrations of non-esterified free fatty acids and β -aminoisobutyrate or acetone levels in urine, see Figure 3.3) indicated also a strong response to fasting (see Figure 3.7C).

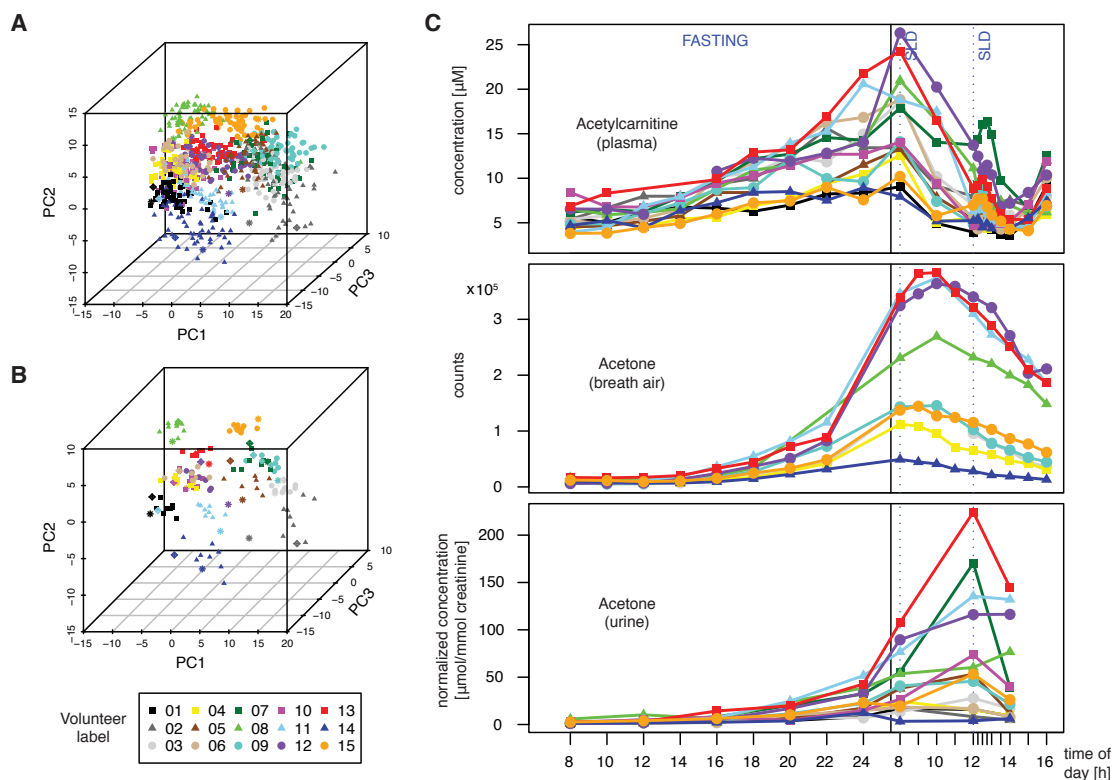


Figure 3.7: Interindividual variation of metabolite concentrations in the context of challenges and sample types. **A:** PCA scores based on the plasma metabolic profiles of all volunteers at all timepoints. **B:** Only samples of the fasting challenge are depicted in the PCA plot. Despite intraindividual variations induced by diverse challenges and sampling times, the samples of each subject are grouped together. **C:** For day 1 and 2, the individual acetylcarnitine (C2) dynamics in plasma are shown for each subject (top). The quantities of acetone in breath air (determined by PTR-MS; middle panel) and urine (determined by NMR; bottom panel) match the observations seen for C2 (determined by MS/MS) in plasma. The large differences in the concentrations between subject 13 (red squares) and subject 14 (dark blue triangles) are consistent across the different sample types.

3.3 Modeling mitochondrial beta-oxidation

The highest intersubject range was observed for the fasting challenge. As the study protocol included the best possible standardization methods for volunteer treatment, sample collection and data acquisition, we hypothesized that the marked variation in challenge response could be caused by differences in physiological characteristics such as muscle or fat mass composition. Yet a correlation analysis between anthropometric measures and plasma metabolite concentrations quantified during the fasting period showed only weak associations (see Table 3.1). Since metabolite concentrations could not explain the differences between subjects, we asked if the individual's capacity for fatty acid metabolism might explain the varying responses to the fasting challenge. We therefore developed a simplified mathematical model for the degradation of fatty acids in the beta-oxidation cascade (see Section 3.1).

We used the plasma concentrations of acylcarnitines with fatty acid chain lengths ranging from 2 to 18, which were measured during the fasting period, as input variables for the model, assuming that these fatty acid derivatives reflect beta-oxidation intermediates. This assumption is based on replicated statistical associations between gene variants in mitochondrial acyl-CoA dehydrogenase beta-oxidation enzymes and plasma acylcarnitine concentrations, reported in several genome wide association studies [81, 98, 104, 172, 237]. Findings from studies on inborn errors in beta-oxidation also show genetically associated changes of plasma acylcarnitine levels [150, 199]. Therefore we based our model-based analysis on plasma acylcarnitines as surrogate markers being in equilibrium with mitochondrial beta-oxidation intermediates, similar to recent models of fatty acid metabolism [164, 250]. Under steady state conditions parameter readouts of the linear model can be estimated by using ratios of condition-specific acylcarnitine concentrations. These parameters or *model-driven ratios* are then proportional to reaction rates in the beta-oxidation cascade and provide a surrogate marker of metabolic capacities. For instance, the rate k_{16} of the conversion $C_{16} \rightarrow C_{14}$ is proportional to the concentration ratio C_2/C_{16} . The model-driven ratios thus allow us to estimate the individual metabolic capacity of each subject for each time point during the fasting challenge (see Figure 3.8).

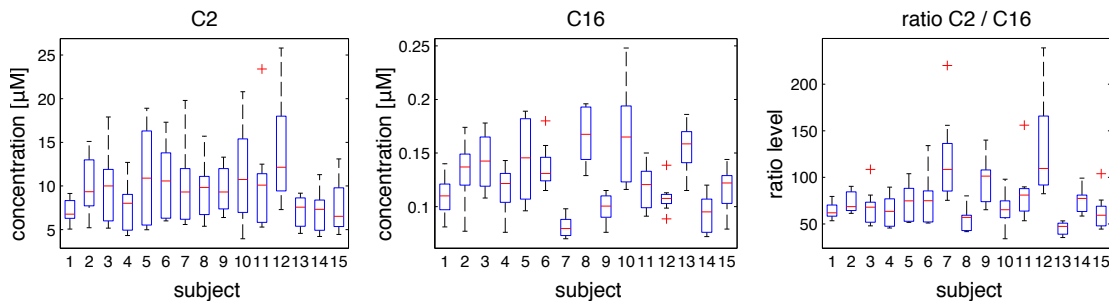


Figure 3.8: Interindividual variation in metabolic profiles and model-driven metabolite ratios. Boxplots display subject-specific metabolite levels of C2, C16 and the model-driven ratio C2/C16 for all 10 time points during the fasting challenge.

3.4 Association between model readouts and phenotypic parameters

Based on metabolite ratios as model readouts for metabolic capacities, we compared the individual response to fasting conditions with volunteer-specific phenotypes. The relationship between both metabolite concentrations and model-driven metabolite ratios and anthropometric or biochemical parameters (ABP) was evaluated using Spearman’s rank correlation statistics (see Section 3.1). Using *p-gain* statistics [190] model-driven ratios that provide stronger statistical associations than single metabolites were identified (*p-gains* greater than 1).

We correlated all levels of even-numbered, saturated plasma acylcarnitines (C2 - C18) measured during the fasting period with anthropometric and biochemical parameters. In addition, we performed all statistical analyses using model-driven ratios instead of single acylcarnitine concentrations (see Table 3.1). Metabolite ratios as readouts from the beta-oxidation model provided stronger associations with the individual’s phenotype than absolute metabolite levels. For instance, the ratio between acetylcarnitine and palmitoylcarnitine (C2/C16, corresponding to the reaction rate k_{16} , see Figure 3.2) revealed stronger rank correlation to anthropometric measures (for instance muscle-fat-ratio, fat mass and BMI) than absolute C2 and C16 plasma levels. The model-driven ratio approach improved statistical correlations, expressed as *p-gain* values greater than 1, yielding for instance 7.3 for correlation with muscle-fat-ratio, 8.8 with BMI and up to

p-gain	Cx	ABP	Rank correlation					
			Ratio C2/Cx vs. ABP		C2 vs. ABP		Cx vs. ABP	
			p-value	ρ	p-value	ρ	p-value	ρ
$7.5 \cdot 10^6$	C16	Sum of hexoses (p)	$1.9 \cdot 10^{-15}$	-0.61	$1.4 \cdot 10^{-8}$	-0.48	0.632	0.07
$3.5 \cdot 10^5$	C16	β -Aminoiso-butyrate (u)	$1.6 \cdot 10^{-16}$	0.77	$5.7 \cdot 10^{-11}$	0.67	0.901	0.03
$2.9 \cdot 10^3$	C6 ^a	Free carnitine (p)	$5.1 \cdot 10^{-7}$	-0.41	0.001	-0.29	0.918	0.02
221.5	C18	Glucose (p)	$1.4 \cdot 10^{-5}$	-0.46	0.003	-0.35	0.520	0.11
87.1	C4	Free carnitine (p)	$2.5 \cdot 10^{-6}$	-0.39	0.001	-0.29	$2.1 \cdot 10^{-4}$	0.33
51.4	C18	Hydroxyiso-butyrate (u)	$4.1 \cdot 10^{-9}$	0.59	$2.1 \cdot 10^{-7}$	0.55	0.914	0.03
18.1	C16	Fat mass	0.015	-0.68	0.731	-0.19	0.267	0.44
10.9	C18	Creatinine (p)	0.028	-0.62	0.405	-0.35	0.301	0.42
8.8	C16	BMI	0.002	-0.77	0.944	-0.04	0.014	0.70
8.5	C16	Body fat percentage	0.042	-0.58	0.798	-0.14	0.354	0.39
7.3	C16	Muscle-fat-ratio	0.037	0.60	0.850	0.11	0.271	-0.43
5.3	C4	Insulin (p)	0.018	-0.28	0.093	-0.23	0.200	0.19

Table 3.1: Correlations between anthropometric and biochemical parameters (ABPs) with metabolite concentrations (C4, C6, C16, C18 and C2) and model-driven ratios derived from the beta-oxidation model. Biochemical parameters and acylcarnitine concentrations were measured during the fasting period of study day 1. Rank correlation p-values were corrected for multiple testing using FDR. Bold values indicate cases for which only model-driven ratios, but not single metabolite levels, were significantly correlated with ABPs. Model-driven ratios reflecting biological processes improve statistical correlations with ABPs of energy metabolism when compared to the correlations with single metabolite concentrations, resulting in values of p-gain > 1. Abbreviations: Cx, acylcarnitine with chain length x; ρ , Spearman's rank correlation coefficient; p, parameter concentration determined in blood plasma; u, parameter concentration determined in urine.

18 for total fat mass. For fat mass, muscle-fat-ratio and other ABPs (bold values in Table 3.1) only metabolite ratios, but not single metabolite levels, provided significant associations with phenotypic parameters. As depicted in Figure 3.9 the fasting challenge leads to different responses in metabolic profiles. Some volunteers showed a strong increase in metabolite levels compared to others, but no clear relationship between the anthropometric parameters fat mass and body mass index (BMI) and absolute C2 and C16 plasma levels is observed. Considering metabolite ratios as a surrogate of the individual's metabolic capacity in contrast revealed significant correlations. Volunteers with high body fat mass and BMI showed decreased levels of acylcarnitine ratios, while ratios for subjects with low fat mass and BMI were increased.

Besides anthropometric measures, also time-resolved biochemical parameters (e.g. fasting insulin, glucose or creatinine plasma levels) showed higher correlation coefficients with model-driven ratios than absolute plasma levels. Figure 3.10 compares the analysis of baseline metabolite samples with a model-based analysis of the fasting challenge.

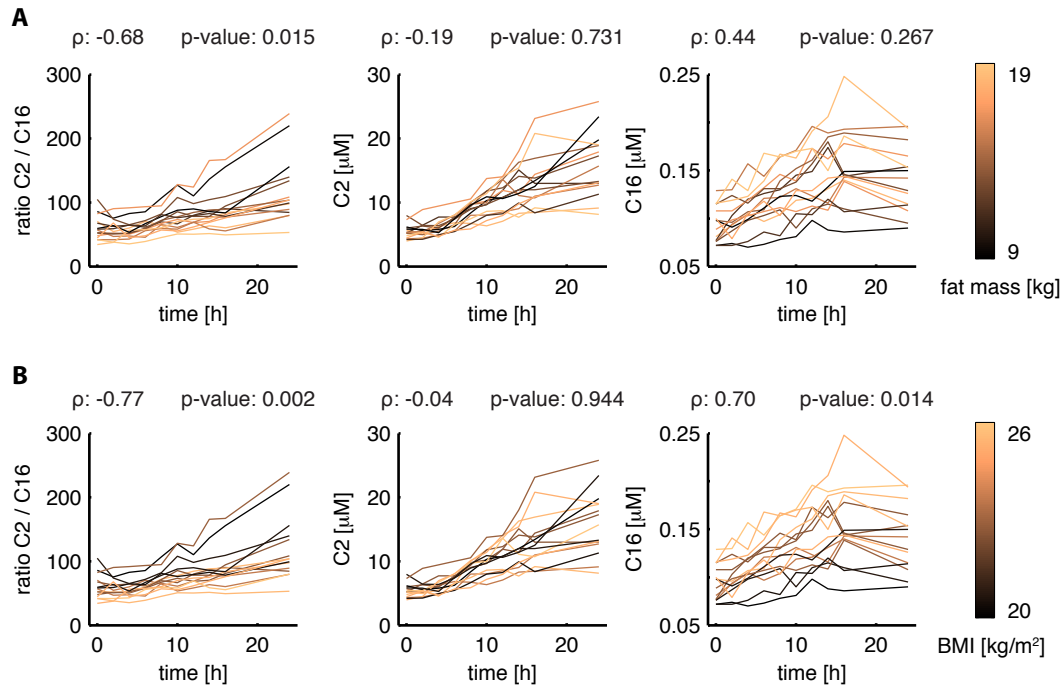


Figure 3.9: Fasting metabolite profiles correlate with anthropometric measures. Timecourses of acetylcarnitine (C2), palmitoylcarnitine (C16) and the model-driven ratio C2/C16 during the fasting challenge are shown. Coloring indicates subject-specific fat mass (panel **A**) and body mass index (BMI, panel **B**). For comparing metabolite levels Spearman's rank correlation statistics (ρ , p -value) was calculated between anthropometric measures and mean metabolite or ratio levels. Compared to single metabolite levels, model-driven ratios improved statistical correlations for fat mass and BMI, resulting in p -gain values of 18.1 and 8.8, respectively (see Table 3.1 for additional results).

Baseline samples were obtained from three resting states at 8 a.m. of day 1, day 3 and day 4, and resemble snap-shot sampling conditions without challenges which are typical for cross-sectional study designs. Correlation analysis between baseline hexose and acylcarnitine plasma concentrations showed weak signals. Note that the hexose concentration is the sum of all monosaccharides with six carbon atoms, mainly made up by glucose in blood plasma samples. Compared to the baseline measurements, the fasting challenge revealed stronger association between glucose and acylcarnitine plasma levels. Considering the individual's metabolic capacity using model-driven ratios increased the signal even further (p-gain of $7.5 \cdot 10^6$). Modeling the beta-oxidation in order to obtain parameters for the individual metabolic capacities thus improved the analysis of the challenge-induced interindividual variation.

3.5 Discussion

Studying the individual response to different nutritional and physiological challenges, metabolite profiling revealed high variation between subjects. Searching for the origin of the inter-subject variability, which was especially pronounced during the fasting challenge, we asked if model-driven metabolite ratios as readouts for metabolic capacity could explain the inter-subject variability better than absolute metabolite levels. Several models for lipid metabolism and fatty acid oxidation have been reported [129, 164, 250], which differ in terms of complexity and detailed description of biochemical reactions. Parameter values of models, which were developed on cellular data or *in vitro* settings, are not readily applicable for the description of plasma metabolite dynamics. In order to obtain a beta-oxidation cascade model which is suitable for the analysis of blood metabolite profiles, we reduced the model complexity by combining subsequent reactions. This was motivated by the fact that few intermediate compounds were quantified, as they are often immediately converted to other metabolites [60, 125, 168]. We excluded fatty acid synthesis reactions due to the antagonistic regulation of production and degradation pathways [200]. The import of fatty acids into the cells and efflux of acylcarnitines are supposed to be not rate-limiting. This facilitated the calibration of the model, as less parameters needed to be identified. Since our main objective was to describe metabolic

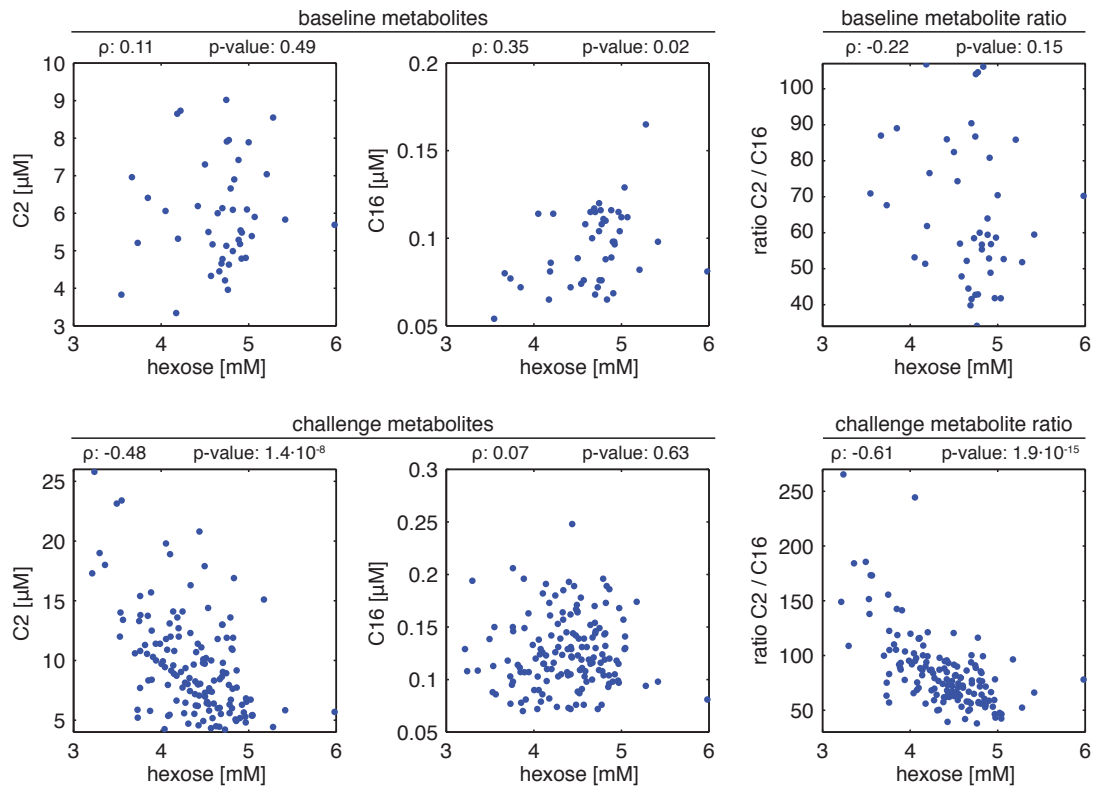


Figure 3.10: Comparison of baseline metabolite, challenge metabolite and metabolite ratio levels with hexose concentration. Under non-challenging conditions (baseline) metabolites show no (acetylcarnitine, C2) or only weak (palmitoylcarnitine, C16) correlation with hexose levels (top panels). Under challenge (fasting) conditions, C2 is significantly correlated with hexose (bottom panels). Taking the metabolite ratio C2/C16 instead, greatly increases this association for the challenge metabolites (p-gain of $7.5 \cdot 10^6$). For comparing metabolite levels Spearman's rank correlation statistics (ρ , p-value) was calculated. Volunteer baseline measurements were derived from the three resting state samples at 8 a.m. of day 1, day 3 and day 4.

capacity rather than exact kinetics, the simplified modeling approach was well suited for our analysis of interindividual variation.

Ratios of metabolites were obtained as read-outs from the model under steady state conditions. These parameters are proportional to reaction rates in the fatty acid degradation cascade and provide surrogate markers of metabolic capacities. The advantage of metabolite ratios is that they inherently account for subject-specific variations in metabolite plasma concentrations [190]. The concept of metabolite ratios as proxies for the description of enzyme reaction rates has been applied successfully in a couple of genome-wide associations studies in order to link metabolic reactions with genetic variants [81, 104, 172, 236, 237]. For our model-based analysis, we assumed that acylcarnitines levels in plasma resemble intracellular beta-oxidation intermediates [150, 199]. The role of extracellular acylcarnitines as surrogates for intracellular fatty acid oxidation intermediates has also been addressed in previous studies [131, 144, 159, 269, 284]. In Chapter 4 we additionally present metabolomics results from a liver cell model to examine the connection between measured intra- and extracellular acylcarnitines after fatty acid loading.

Though subjects were matched by age and BMI, the experimental setup did not account for genetic effects and long-term environmental factors such as dietary habits and life style. As we have discussed in Chapter 2, metabolite ratios are associated to genetic variation on a population level. Future studies could be improved by considering further experimental data to account also for genetic effects. Measuring the epigenetic markup of subjects might reveal effects resulting from environmental conditioning (e.g. from lifestyle or dietary habits), which contribute to the observed variation in metabolic response [191]. During exercise and fasting conditions when beta-oxidation flux is higher than respiratory chain and tricarboxylic acid cycle capacities, ketone bodies are produced [155], which are also detected in our metabolomics panel. Accounting for ketogenic reactions in the model might improve our results, especially for catabolic challenges. Extending the correlation analysis to a functional evaluation of metabolite timecourses using methods from functional data analysis could additionally enhance our understanding of the observed dynamics in metabolism [231].

Taken together, we suggest in this work to use metabolite ratios, which are derived from models of biochemical pathways, for characterizing distinct metabolic phenotypes. We showed that this model-driven analysis of metabolic systems under perturbations (e.g. fatty acid beta-oxidation under fasting conditions) allows for a better investigation of the relationship between individual physiological phenotypes and biological pathways. The coherence between model read-outs and specific parameters influencing biochemical reactions during fasting has proven the feasibility of this approach. Future study designs should consider the challenging approach and a model-based analysis in order to detect metabolic differences which are not observable under baseline conditions, for instance for an improved characterization of human metabotypes in genotype-phenotype association studies (see Chapter 2).

Chapter 4

Quantitative modeling of an *in vitro* enzyme knockdown in the fatty acid beta-oxidation pathway

In Chapter 2 we have assessed genetic effects on metabolism at a system level using population data from genome-wide association studies (GWAS) with metabolic traits. The mechanistic interpretation of these results is challenging, as the observed phenotypes often result from a mixture of related, yet unknown processes, or identified loci are located in non-coding regions [92, 170]. In most cases GWAS can only give an overall picture about the biological processes that link either the associated variant or the affected gene to the observed phenotype [28]. In this chapter we will evaluate how quantitative modeling of *in vitro* experimental data can be used to translate GWAS results into the functional characterization of the underlying biochemical processes.

In order to determine the molecular function of GWAS results, there is a strong need for cellular models representing the observed genetic background. Using specific models such as *Xenopus laevis* oocytes or mammalian cell lines allowed for translating associations of genetic loci with serum metabolite levels into the functional characterization of transporter proteins [32, 237]. We therefore asked if quantitative modeling of *in vitro*

enzyme knockdown experiments, which reflect the observed impaired biomolecular functions, could also be used as a tool for molecular and functional studies in order to assess the genotype-dependent impact on metabolic pathways.

To this end, we focused on a previously reported association between levels of butyrylcarnitine (C4:0-acylcarnitine), the transport form of C4:0-acyl-CoA, and the single nucleotide polymorphism (SNP) rs2014355, which is located in close proximity to the *short chain acyl-coenzyme A dehydrogenase* (ACADS) gene locus [81, 98, 104, 172, 237]. This association was also found using the network-based ratio approach presented in Chapter 2. The ACADS locus encodes for an enzyme that catalyzes the initial step of the mitochondrial fatty acid beta-oxidation (FAO) pathway with the major substrate C4:0-acyl-CoA [77]. Of note, rs2014355 tags a larger region of correlated variants including the non-synonymous-coding SNP rs1799958 which encodes for an ACADS protein variant with reduced catalytic activity and thermostability [44, 45, 86, 187]. Short-chain acyl-coenzyme A dehydrogenase deficiency (SCADD), which is a rare autosomal recessive FAO disorder, can result from alterations in the ACADS gene [108, 169]. The symptoms of SCADD, which generally appear early in life, are heterogeneous including developmental delay, epilepsy, hyper- and hypotonia or ketotic hypoglycemia [251, 252]. SCADD-related impaired FAO activity leads to an accumulation of byproducts of fatty acid metabolites. For this reason, increased levels of butyrylcarnitine and ethylmalonic acid in plasma and urine are biochemical markers for the diagnosis of SCADD [8, 20].

In order to investigate the effects of altered ACADS protein levels on FAO metabolite concentrations during beta-oxidation at cellular level, an *in vitro* model for the gradual knockdown of endogenous ACADS was generated in the Huh7 human liver cell line. Using this *in vitro* model of liver cells which are known to show high FAO activity, intra- and extracellular time-resolved acylcarnitine concentrations were measured after incubating the cells with palmitic acid to induce FAO. This allowed for the quantification of the effect of ACADS gene expression alterations on levels of FAO intermediate metabolites. As an extension of the *in vivo* modeling of fatty acid oxidation in Chapter 3, the knockdown setup allows to analyze the change of specific enzymatic rates in a controlled environment.

We analyzed the observed metabolite dynamics using a model of the FAO pathway to capture the dynamics and interactions of individual metabolites at a system level. Quantitative dynamical modeling has been applied successfully to many biological questions, ranging from cellular signaling to metabolic pathways [122, 123], thus helping to understand complex biological data. As we were interested in the overall differences of the metabolite dynamics, we developed a mathematical model of the FAO pathway which describes the fundamental reactions during the breakdown of fatty acids. Based on the *in silico* reaction rates inferred from our model and experimental data from human liver cells, we compared the dynamical changes between wild-type and ACADS knockdown conditions statistically.

The chapter is organized as follows: First we describe a regulated knockdown of ACADS in a liver cell line with high FAO activity for evaluating the influence of expression levels on intra- and extracellular acylcarnitine levels, that may contribute to the observed GWAS signals. The experimental work has been performed by Kerstin Ehlers and Helmut Laumen at the Chair of Nutritional Medicine from Technische Universität München. Next we describe the development of a FAO pathway model that we used to analyze the knockdown-specific, time-resolved metabolite measurements. We further discuss the interpretation of baseline GWAS results with respect to the observed dynamics for the *in vitro* ACADS knockdown and the *in silico* model readouts.

The work presented in this chapter has been performed in collaboration with Kerstin Ehlers, Helmut Laumen and the group Hans Hauner. The results are summarized in:

- Ehlers K*, **Stückler F***, Hastreiter M, Pfeiffer L, Reischl E, Kastenmüller, Daniel H, Ensenaer R, Krumsiek J, Hauner H, Theis FJ, Laumen H (2014): In vitro modeling and dynamic analysis of a metabolic quantitative trait locus implies novel features of ACADS function in fatty acid oxidation. *in submission*.

This publication is a joint first author work and the content of this paper is also part of another thesis by Kerstin Ehlers, who also performed the experiments [62]. My contribution to this work is the development of a dynamic pathway model, the statistical analysis of the *in vitro* metabolomics data, as well the biological interpretation of the model prediction.

4.1 Methods

In this section we will briefly describe the main experimental methods for generating the experimental data and introduce the mathematical model of the fatty acid oxidation pathway. The experimental work has been performed by Kerstin Ehlers in the group of Helmut Laumen at the Chair of Nutritional Medicine from Technische Universität München. Further details about the experimental setup and biochemical characterization methods (for instance western blot, qRT-PCR, metabolite extraction) have been described previously by Kerstin Ehlers [62].

ACADS knockdown in Huh7 cells

The human hepatoma cell line Huh7 was cultivated in DMEM medium (Gibco, Karlsruhe, Germany) containing 10% FBS and 1% penicillin-streptomycin (PAA Laboratories GmbH, Pasching, Germany) at 37°C in a humidified atmosphere at 5% CO₂. Cells were passaged twice a week. The optimal shRNA sequence for the ACADS knockdown system was identified using the pVal shRNA Validation Platform RNAiONE (Sirion Biotech, Martinsried, Germany). shRNA revealing the best knockdown efficiency and a non-target shRNA were cloned into the One-Vector inducible shmir platform (Sirion Biotech). Stable Huh7 cell pools (shACADS Huh7 for an inducible shRNA ACADS knockdown and shNTC Huh7 with a non-targeting control shRNA) were generated by packaging, transduction and stable integration of the tet-on expression vector and subsequent antibiotic selection by Sirion Biotech. Stable Huh7 cells transduced with the tet-inducible shRNA for down-regulation of ACADS or control shRNA were incubated with 0 and 10 ng/ml doxycycline (Sigma-Aldrich, Steinheim, Germany) for 5 days and with 5 ng/ml for 3 days. Medium was changed after 3 days. Knockdown efficiency was examined on RNA and protein level using qRT-PCR and western blot, respectively. The gradual ACADS knockdown was established for three doxycycline (dox) concentrations: 0 ng/ml (shACADS^{null}), 5 ng/ml (shACADS^{med}) and 10 ng/ml doxycycline (shACADS^{max}).

Loading of Huh7 cells with palmitic acid

Cells treated with doxycycline were seeded in 6-well plates at 250,000 cells/well and grown in 37°C, 5% CO₂ incubator. After two days, growth medium (DMEM medium containing 10% FBS, 1% penicillin-streptomycin and 0 ng/ml, 5 ng/ml or 10 ng/ml doxycycline) was changed to assay medium (111 mM NaCl, 4.7 mM KCl, 2 mM MgSO₄, 1.2 mM Na₂HPO₄, 0.5 mM carnitine; all components from Sigma-Aldrich, Steinheim, Germany) at a final volume of 1.8 ml. 60 min later, 0.2 ml 2 mM palmitic acid-BSA was added. At baseline and after 7, 14, 21 and 28 min 20 µl supernatant was given on a 6 mm filter paper punch for preparing metabolite extraction. Cells were washed with PBS and harvested by scraping in 300 µl ice-cold 100% methanol. Both, supernatant and cells were shock frozen in liquid nitrogen. The validity of the cell system as a model for FAO was confirmed by assessing the oxygen consumption rate in Huh7 cells after palmitic acid loading [62].

Acylcarnitine measurement in supernatant and Huh7 cell extracts

Harvested cells were lysed in an ultrasonic bath. Cell debris was spun down by full speed (10 min, 4°C) centrifugation (Eppendorf 5417 R, Hamburg, Germany) and supernatant was collected. Filter paper punches soaked with 20 µl supernatant were vacuum dried in a speed vac (Savant SPD 111V SpeedVac Concentrator, Thermo Scientific, Dreieich, Germany) for approximately 45 min. 100 µl 5 mM NH₄Ac containing internal standard was added to the dried filter paper punches with supernatant which were shaken for 30 min at full speed and room temperature (Thermomixer comfort, Eppendorf, Hamburg, Germany). Supernatant was transferred into a Millipore filter plate and filtered by centrifugation at 1,500 g for 20 min. Flow-through was collected in glass vials and stored at -80°C until measurement. Dried cell pellets were resuspended in 100 µl 5 mM NH₄Ac containing internal standards (Chromsystems, Gräfelfing, Germany) and filtered through a Millipore filter plate (Billerica, MA, USA) by centrifugation at 1,500 g for 20 min. Flow-through was collected in glass vials (Chromacol, Herts, UK) and stored at -80°C until measurement. Quantitative levels of acylcarnitine metabolites (C16, C14, C12, C10, C8, C6, C4, C2, C16:1, C14:1, C12:1) in supernatant and cell

extracts of Huh7 cells were measured using chromatographic separation on a ZIC-HILIC column (Merck, Darmstadt, Germany) and triple-quadrupole tandem mass spectrometry (QTRAP 5500, AB Sciex, Framingham, MA, USA) with Turbo V spray electron spray interface in positive ion mode for detection. Acylcarnitines measurements were normalized to the protein amount/well.

Statistical analysis

All data are expressed as mean $\pm$ standard deviation, if not stated otherwise. Differences of gene expression and metabolite levels in Huh7 cells were assessed by two-tailed, one sample t-tests. All statistical analysis was performed using MATLAB (R2012a, The Mathworks Inc., Natick, MA).

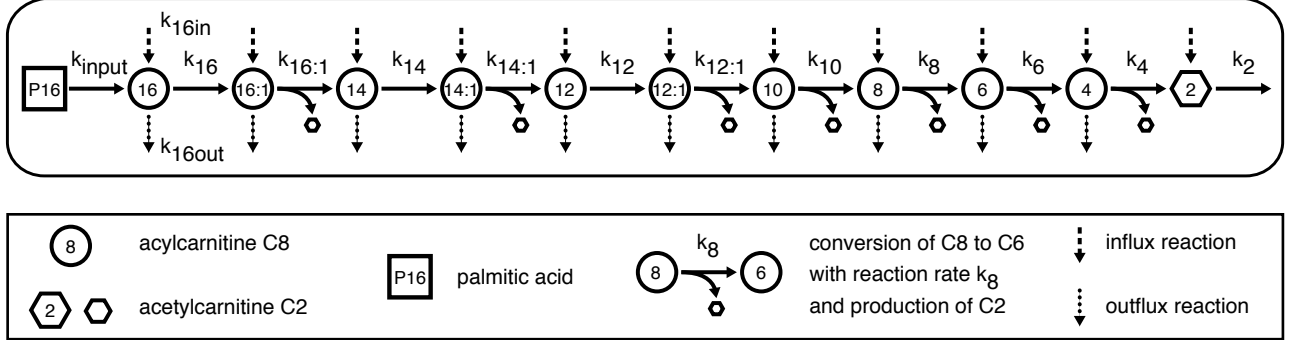


Figure 4.1: Schematic representation of the fatty acid oxidation model. The general model includes influx and outflux reactions for each compound.

Description of the fatty acid oxidation model

The degradation of palmitic acid (P16) in the fatty acid oxidation (FAO) model was described as a linear cascade of subsequent, irreversible first-order reactions using ordinary differential equations (ODEs) with mass action kinetics (see Figure 4.1). Measured intracellular acylcarnitine concentrations were used as a reflection for acyl-CoA concentrations [159, 177, 250], described as observables $C_{16}, C_{16:1}, C_{14}, \dots, C_2$ in the model. Since only few intermediate products of a FAO cycle were measured, we combined the four FAO cycle reaction steps (i.e. dehydrogenation, hydration, oxidation, thiolysis, see Section 1.1) into one fundamental reaction step (e.g. $C_6 \xrightarrow{k_6} C_4 + C_2$), with the reaction rate k_6). During each cascade step the carbon chain is shortened by two carbon atoms, resulting in the production of C_2 . Only if respective acylcarnitine species with one double bond were measured, the dehydrogenation step was added to the model (e.g. $C_{16} \xrightarrow{k_{16}} C_{16:1} \xrightarrow{k_{16:1}} C_{14} + C_2$). We took also into account exchange reactions between the FAO pathway and other active biochemical pathways by adding acylcarnitine-specific influx (e.g. $\xrightarrow{k_{4in}} C_4$) and outflux reactions (e.g. $C_4 \xrightarrow{k_{4out}}$). The concentration change of each observable over time was modeled by systems of ODEs (e.g. $\frac{dC_4}{dt}(t) = \dot{C}_4(t) = k_{4in} - k_{4out}C_4(t) + k_6C_6(t) - k_4C_4(t)$). The change of C_4 thus depends on the concentrations of C_4 and C_6 , the reaction rates of the influx and outflux reactions (k_{4in}, k_{4out}), and the rates (k_6, k_4) of conversion reactions $C_6 \xrightarrow{k_6} C_4 + C_2$ and $C_4 \xrightarrow{k_4} C_2 + C_2$. The uptake of palmitic acid (P16) into the beta-oxidation cascade ($P16 \xrightarrow{k_{input}} C_{16}$) was modeled by a first order reaction. The general FAO model with intermediate compounds C_{16} to

C_2 and reaction rates $k_{input}, k_{16}, k_{16:1}, \dots, k_2$ (see Figure 4.1) includes influx and outflux reactions for each compound $k_{16in}, k_{16out}, \dots, k_{4in}, k_{4out}, k_{2in}$.

The ODE system of the general fatty acid oxidation model reads

$$\begin{aligned}
\dot{P16}(t) &= P16loading(t) - k_{input} \cdot P16(t) \\
\dot{C}_{16}(t) &= k_{16in} - k_{16out} \cdot C_{16}(t) + k_{input} \cdot P16(t) - k_{16} \cdot C_{16}(t) \\
\dot{C}_{16:1}(t) &= k_{16:1in} - k_{16:1out} \cdot C_{16:1}(t) + k_{16} \cdot C_{16}(t) - k_{16:1} \cdot C_{16:1}(t) \\
\dot{C}_{14}(t) &= k_{14in} - k_{14out} \cdot C_{14}(t) + k_{16:1} \cdot C_{16:1}(t) - k_{14} \cdot C_{14}(t) \\
\dot{C}_{14:1}(t) &= k_{14:1in} - k_{14:1out} \cdot C_{14:1}(t) + k_{14} \cdot C_{14}(t) - k_{14:1} \cdot C_{14:1}(t) \\
\dot{C}_{12}(t) &= k_{12in} - k_{12out} \cdot C_{12}(t) + k_{14:1} \cdot C_{14:1}(t) - k_{12} \cdot C_{12}(t) \\
\dot{C}_{12:1}(t) &= k_{12:1in} - k_{12:1out} \cdot C_{12:1}(t) + k_{12} \cdot C_{12}(t) - k_{12:1} \cdot C_{12:1}(t) \\
\dot{C}_{10}(t) &= k_{10in} - k_{10out} \cdot C_{10}(t) + k_{12:1} \cdot C_{12:1}(t) - k_{10} \cdot C_{10}(t) \\
\dot{C}_8(t) &= k_{8in} - k_{8out} \cdot C_8(t) + k_{10} \cdot C_{10}(t) - k_8 \cdot C_8(t) \\
\dot{C}_6(t) &= k_{6in} - k_{6out} \cdot C_6(t) + k_8 \cdot C_8(t) - k_6 \cdot C_6(t) \\
\dot{C}_4(t) &= k_{4in} - k_{4out} \cdot C_4(t) + k_6 \cdot C_6(t) - k_4 \cdot C_4(t) \\
\dot{C}_2(t) &= k_{2in} + 2k_4 \cdot C_4(t) + k_6 \cdot C_6(t) + k_8 \cdot C_8(t) + k_{10} \cdot C_{10}(t) + \\
&\quad + k_{12:1} \cdot C_{12:1}(t) + k_{14:1} \cdot C_{14:1}(t) + k_{16:1} \cdot C_{16:1}(t) - k_2 \cdot C_2(t)
\end{aligned} \tag{4.1}$$

The initial conditions for each compound are described by additional parameters ($P16_{init}, C_{16init}, \dots, C_{2init}$). An input function for palmitic acid (P16) was used to describe the loading of Huh7 cells with palmitic acid during the experiment. Adding palmitic acid to the cells is described using a scaled normal distribution curve as input function for P16. This means that the state variable of P16 changes its value by a certain amount at a specific timepoint t during the simulation. We chose as input function:

$$P16loading(t) = P16total \cdot \frac{1}{\sqrt{2\pi}d^2} \exp \left\{ -\frac{(t - tp)^2}{2d^2} \right\}$$

with $P16_{total}$ being the total palmitic acid amount (i.e. area under the curve), tp the timepoint of palmitic acid loading (at timepoint 1 [min] of the experiment) and d being the duration of palmitic acid loading (set to 0.5 [min]). The initial amount of palmitic acid which was taken up by the cell and channeled to the FAO cascade cannot be determined. Yet it can be reasoned that the palmitic acid concentration is rather high in comparison to the measured acylcarnitine species. For this reason, we set the total amount of palmitic acid ($P16_{total}$) to 10000 [nmol/g protein], being about 100 times higher than the maximal intracellular concentration of palmitoyl-carnitine (C_{16}).

Model-based comparison between experimental conditions

In order to assess the differences in FAO dynamics between two experimental conditions (e.g. between $shACADS^{null}$ and $shACADS^{max}$), we used two different FAO models M1 and M2. The ODE system of the first model (M1) is represented as in Equation (4.1). For the second model (M2) we introduced for all cascade reactions rates ($k_{input}, k_{16}, k_{16:1}, \dots, k_2$) condition-specific prefactors ($\alpha_{input}, \alpha_{16}, \alpha_{16:1}, \dots, \alpha_2$). As an example, the change of C_4 over time is described in the first model (e.g. $shACADS^{null}$) by

$$\dot{C}_4^{M1}(t) = k_{4in} - k_{4out} \cdot C_4(t) + k_6 \cdot C_6(t) - k_4 \cdot C_4(t)$$

and in the second model (e.g. $shACADS^{max}$) by

$$\dot{C}_4^{M2}(t) = k_{4in} - k_{4out} \cdot C_4(t) + \alpha_6 \cdot k_6 \cdot C_6(t) - \alpha_4 \cdot k_4 \cdot C_4(t)$$

with initial conditions C_{4init}^{M1} and C_{4init}^{M2} . An α_4 -value of 1 then denotes that the reaction rate between the two models is not different. Note that the influx and outflux rates k_{4in} and k_{4out} are the same for both models. In order to reduce the model complexity we assumed that influx and outflux reactions should be independent of the knockdown. The respective reaction rates (e.g. k_{4in} and k_{4out}) are therefore the same for M1 and M2. The model simulations were compared to the time course acylcarnitine data on log10-scale obtained from the knockdown experiments. This comparison was performed on log10-scale to account for log-normally distributed measurement noise [72]. We used maximum likelihood estimation to obtain model parameters which describe the measured

data best (see Section 1.4). The parameter fitting was carried out on log10-scale to ensure efficient estimates for values being potentially different by orders of magnitude. A profile likelihood approach was used to check for parameter identifiability and to compute confidence intervals for parameter values (see [201] and 1.4). All quantitative dynamical modeling was performed using MATLAB (R2012a, The Mathworks Inc., Natick, MA) and the Data 2 Dynamics software package [202].

Model selection for influx and outflux reactions

We considered not all intermediate species of FAO to be in an exchange with other biochemical pathways. To test this we performed a model selection for the influx and outflux reactions (see also Section 1.4). To this end all models with all possible influx and outflux combinations were fitted individually to the control (shACADS^{null}) and knockdown (shACADS^{max}) data using model M1 and M2. During the model selection we set rates of excluded influx and outflux reactions to 0. The ODE system in Equation (4.1) reduces accordingly. As adding additional parameters to the model will increase the likelihood and might result in overfitting [121], we used for model comparison the Bayesian information criterion (BIC, [226]), which includes a penalty term for the number of model parameters:

$$BIC = -2 \cdot \log(\hat{L}) + k \cdot \log(n)$$

Here $\hat{L}$ is the maximized likelihood of the model, k the number of model parameters and n the amount of measured data points. A lower BIC value results either from fewer parameters, better data fitting or both.

4.2 *In vitro* modeling of reduced ACADS expression in Huh7 hepatocytes reflects the genotype-dependent metabolic C4:0-acylcarnitine phenotype

For a better understanding of the influence of reduced ACADS expression levels on intra- and extracellular acylcarnitine concentrations in a cell type [217] with fatty acid oxidation activity we established a regulated ACADS knockdown in Huh7 hepatocytes [281] as an *in vitro* model system. The validity of the cell system as a model for FAO was confirmed by assessing the oxygen consumption rate in Huh7 cells after palmitic acid loading [62]. Huh7 cells were stably infected using a lentiviral expression system to generate a doxycycline-(dox)-inducible shRNA ACADS knockdown (shACADS). Evaluating the efficiency of the dox-inducible ACADS knockdown by RT-qPCR we found a significant reduction of ACADS mRNA levels in the intermediate (shACADS^{med}, dox 5 ng/ml) and maximal (shACADS^{max}, dox 10 ng/ml) shACADS Huh7 cells (82% and 84%, respectively, $p < 0.001$, Figure 4.2B) as compared to the null knockdown (shACADS^{null}, dox 0 ng/ml). In western blot analyses we measured a gradual 70% and 93% reduction of ACADS protein levels for shACADS^{med} and shACADS^{max} Huh7 cells, respectively (Figure 4.2A). Dox-treatment of Huh7 cells stably transduced with a non-targeting control shRNA (shNTC) revealed no effect on both, protein and mRNA levels (Figure 4.2A and B). Proving that the lentiviral shRNA construct is solely active upon dox treatment without any leakiness, we found no differences when comparing mRNA expression levels in shACADS versus shNTC cells without dox treatment ($p = 0.96$).

Analysis of baseline intracellular acylcarnitine levels in the shACADS Huh7 cell model revealed a knockdown-dependent accumulation of C4:0-acylcarnitine (Figure 4.2C). Maximal knockdown in the shACADS^{max} cells revealed a significant 2.6-fold increase of baseline C4:0-acylcarnitine ($p < 0.05$) as compared to shACADS^{null}. Thus, the shACADS Huh7 hepatocyte knockdown model reflects the rs2014355 genotype-associated C4:0-acylcarnitine phenotype reported in GWAS [81, 98, 104, 172]. No significant differences were observed in shNTC cells with a non-targeting control shRNA upon dox-treatment, proving ACADS-specificity of the observed effect (Figure 4.2C). We found no ACADS-dependent differences in baseline levels of any other measured acylcarnitine in both

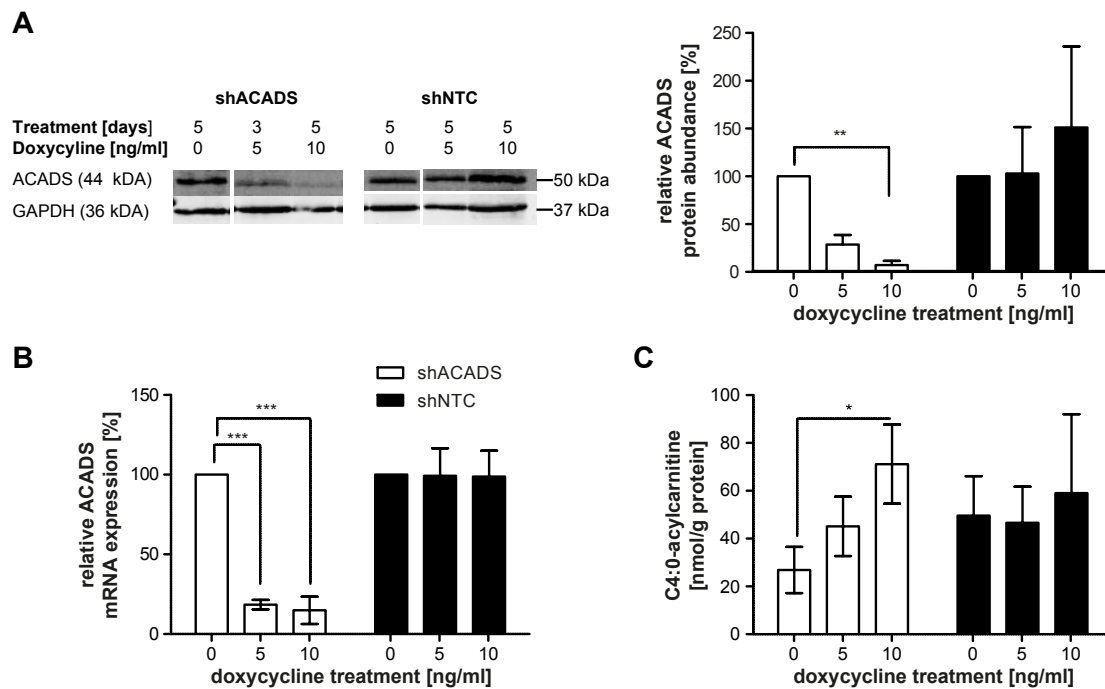


Figure 4.2: Phenotypes of the shACADS and shNTC Huh7 knockdown cells. Doxycycline-induced knockdown of ACADS resulted in a long-term decrease of mRNA expression and protein abundance, and an intracellular accumulation of C4:0-acylcarnitine, which reflects the ACADS substrate C4:0-CoA. **A:** Western blot analysis of ACADS protein in shACADS and shNTC knockdown Huh7 cell lysates, one of four experiments is depicted exemplarily. **B:** RT-qPCR analysed mRNA expression of four independent experiments. Protein (**A**) and mRNA (**B**) were harvested after 3 or 5 days of treatment with 0, 5, and 10 ng/ml doxycycline, respectively. **C:** Intracellular C4:0-acylcarnitine measurement in stably transduced Huh7 cells after 1h. Values of four independent experiments are expressed as mean $\pm$ SD. ACADS = short chain acyl-CoA dehydrogenase; NTC = non-target control; GAPDH = glyceraldehyde-3-phosphate dehydrogenase (control). *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

cellular models (time point $t=0$, Figure 4.3 and 4.4). Moreover, reflecting the results from GWAS with metabolite ratios (see Chapter 2), we found a significant decrease of the intracellular C3:0-/C4:0-acylcarnitine ratio ($p < 0.05$) in shACADS^{max} cells as compared to shACADS^{null} cells.

To assess the effect of reduced ACADS enzyme expression levels on FAO kinetics, we leveraged the shACADS model and measured short-term time courses of both intra- and extracellular acylcarnitine levels after incubation of cells with palmitic acid, a major substrate of the FAO pathway. As palmitic acid is an even-numbered, saturated fatty acid, we focused our analysis on even-numbered intermediate metabolites of the FAO cascade. In all cell models intracellular acylcarnitine concentrations increased after loading with palmitic acid, indicating an increase in FAO (Figures 4.3 and 4.4). Comparing time courses of metabolite levels in shACADS^{null}, shACADS^{med} and shACADS^{max} cells suggest a gradual inhibition of ACADS expression, i.e. with increasing ACADS knockdown efficiency, metabolite levels are either decreasing (e.g. C16:0-, C12:0-, and C10:0-acylcarnitines) or increasing (e.g. C4:0-acylcarnitine). For intermediate knockdown cells (shACADS^{med}) we find intermediate metabolite levels, most obvious for C10:0-, C8:0-, and C6:0-acylcarnitines (Figure 4.3). Pairwise comparison of metabolite levels at each time point between shACADS^{null} and shACADS^{max} cells revealed for few compounds at late time points (C6, C8, C10, C12) significant differences, while C4 showed significant differences at the beginning of the FAO assay (see Figure 4.3).

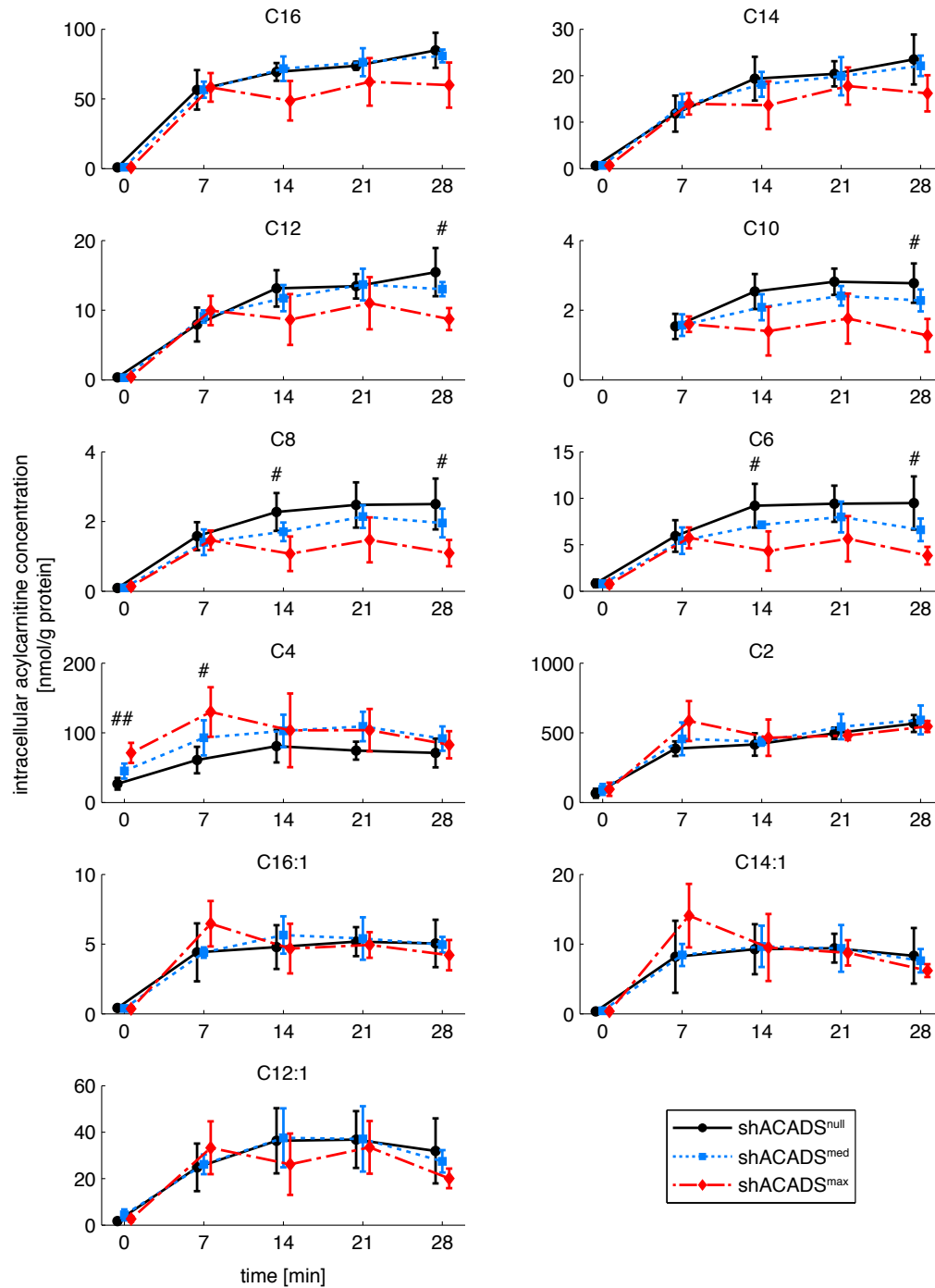


Figure 4.3: Intracellular acylcarnitines after palmitic acid loading in ACADS knockdown cells. Palmitic acid was added to induce fatty acid oxidation in shACADS Huh7 cells (shACADS^{null}, shACADS^{med} and shACADS^{max}: cells treated with 0, 5 and 10 ng/ml doxycycline, respectively). Intracellular acylcarnitines were extracted and measured before palmitic acid loading and after 7, 14, 21 and 28 minutes. Measurements are shifted slightly on the x-axis for a better visualization. Values of four independent experiments are expressed as mean $\pm$ SD. ##: $p < 0.01$, #: $p < 0.05$; time point specific comparison between shACADS^{null} and shACADS^{max}.

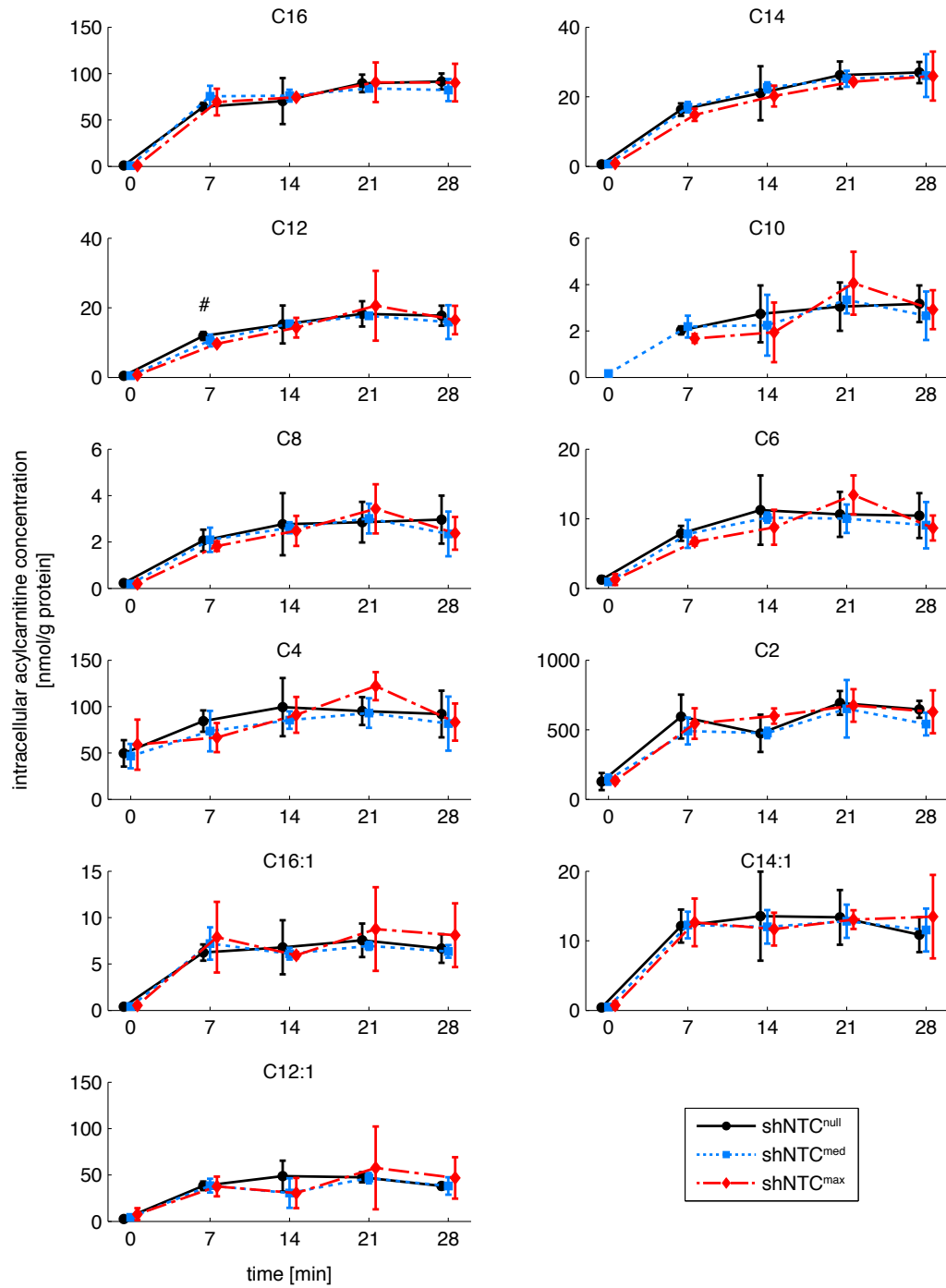


Figure 4.4: Intracellular acylcarnitines after palmitic acid loading in Huh7 cells with non-target shRNA. Palmitic acid was added to induce fatty acid oxidation in cells transduced with a non-target shRNA (shNTC^{null}, shNTC^{med}, shNTC^{max}; expression induced by 0, 5 and 10 ng/ml doxycycline, respectively). Intracellular acylcarnitines were extracted and measured before palmitic acid loading and after 7, 14, 21 and 28 minutes. Measurements are shifted slightly on the x-axis for a better visualization. Values of four independent experiments are expressed as mean $\pm$ SD. #: $p < 0.05$; time point specific comparison between shNTC^{null} and shNTC^{max}.

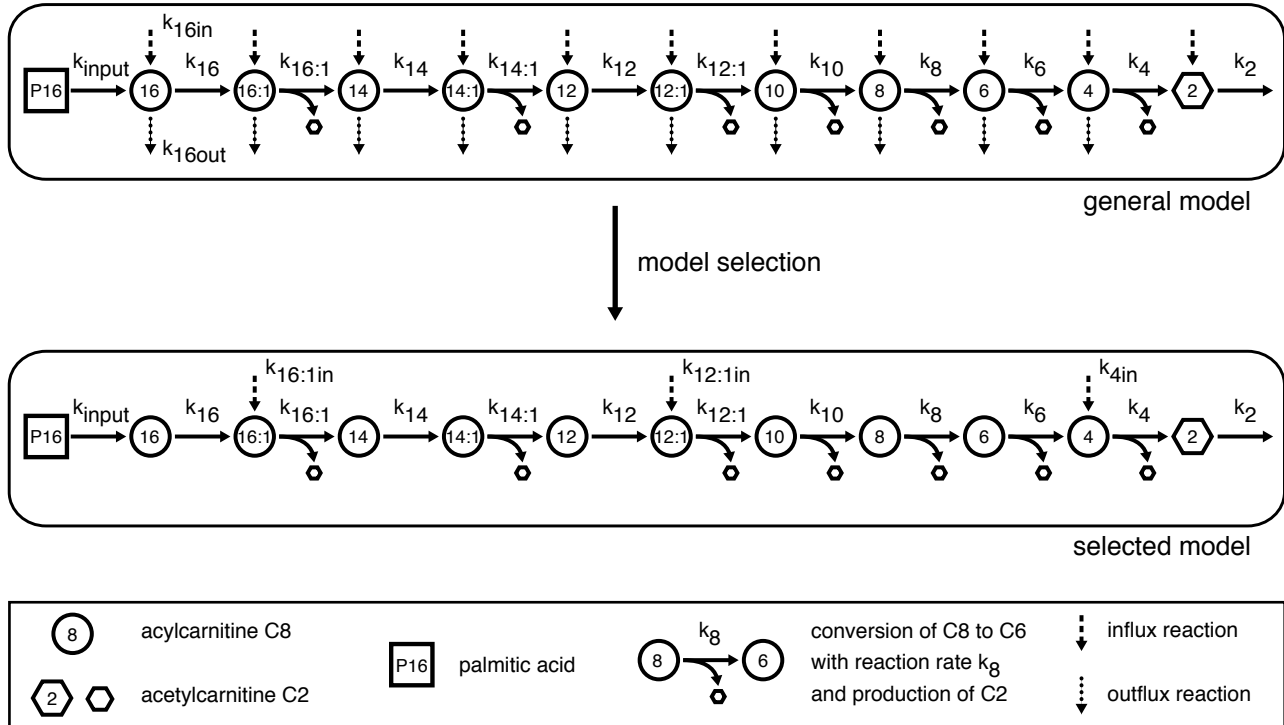


Figure 4.5: Schematic representation of the fatty acid oxidation model. The general model includes influx and outflux reactions for each compound. After data-driven model selection the selected model contains only influx reactions for C16:1-, C12:1- and C4:0-acylcarnitines.

4.3 Model-based analysis of time-dependent intracellular acylcarnitine levels in palmitic acid loaded shACADS

Next, we used a linear model of the FAO pathway to analyze the intracellular metabolite profiles in the Huh7 cell model. Similarly to the analysis of interindividual variation in plasma levels of metabolites which we discussed in Chapter 3, quantitative modeling allows to assess the dynamics of the here measured time-dependent metabolite profiles in the context of the FAO pathway (Figure 4.5). Intracellular acylcarnitine concentrations were used as proxies for FAO intermediate metabolite concentrations [159, 177, 250]. By including acylcarnitine-specific influx and outflux reactions into the model, we were able to account for a possible exchange of FAO metabolites with other biochemical pathways. As we considered not all intermediate species of FAO to undergo exchange and in order to reduce the model complexity, we performed a model selection for the influx and outflux

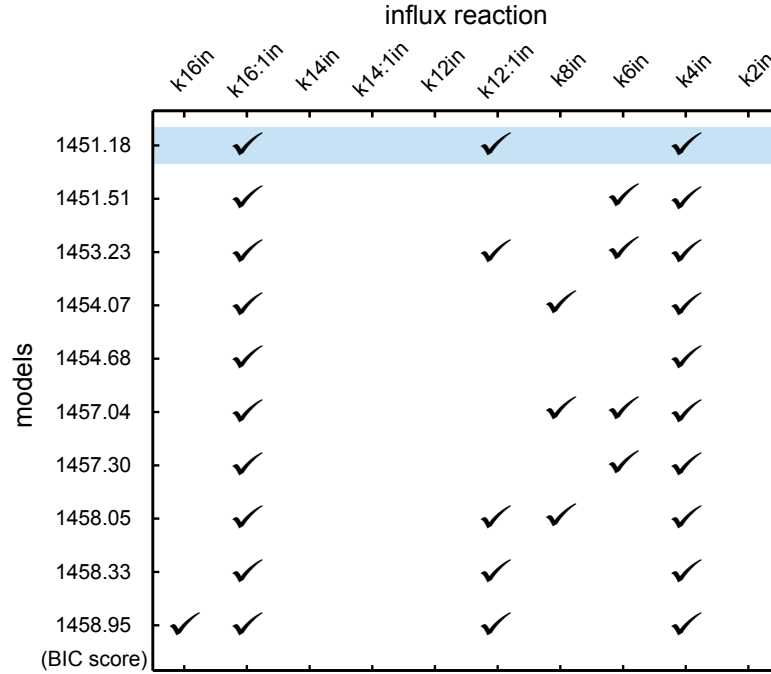


Figure 4.6: Comparison of the 10 best models according to Bayesian information criterion (BIC) score. Each row correspond to a single model. Influx reactions present in the respective model are denoted by ticks. Outflux reactions are not shown as none were present in the best 10 models. The model in the first row (blue shaded area) with the lowest BIC score was chosen for further analysis.

reactions based on the combined $\text{shACADS}^{\text{null}}$ and $\text{shACADS}^{\text{max}}$ data. To this end all models with all possible influx and outflux combinations were fitted individually to the experimental data (see methods section 4.1). Figure 4.6 shows the result for the best 10 models. The selected model only contains influx reactions for C16:1, C12:1 and C4 and no outflux out of the system except for the reaction $C_2 \xrightarrow{k_2}$ with reaction rate k_2 (see Figure 4.5).

Model-based comparison of knockdown conditions

To identify ACADS knockdown-dependent kinetic changes we fitted the selected model to the time course data from $\text{shACADS}^{\text{null}}$ and $\text{shACADS}^{\text{max}}$ experiments (see Figure 4.7A). The reaction rates $k_{16}, \dots, k_4$ in the null and maximal knockdown were compared using the reaction specific prefactor α ($\alpha_{16}, \dots, \alpha_2$). α -values of 1 denote equal,

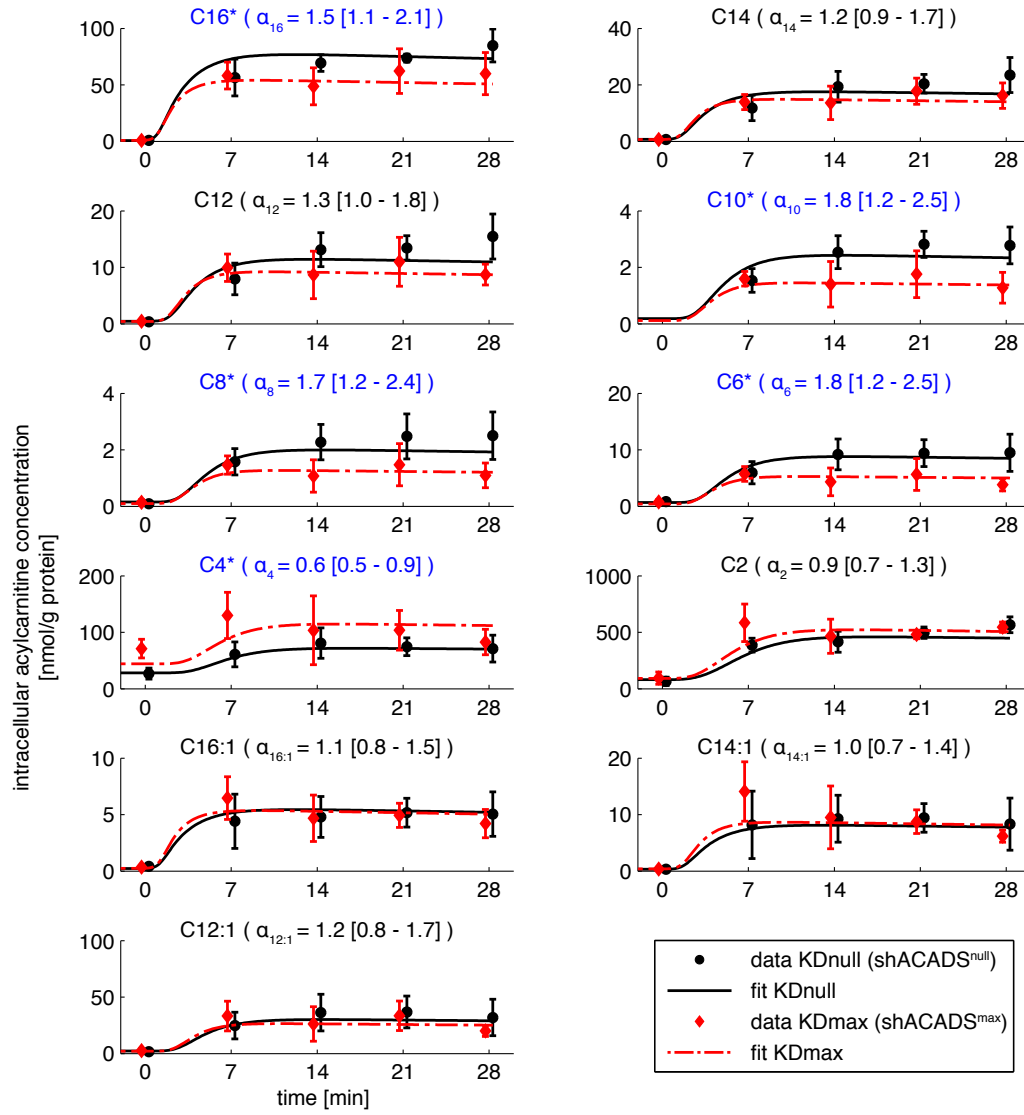
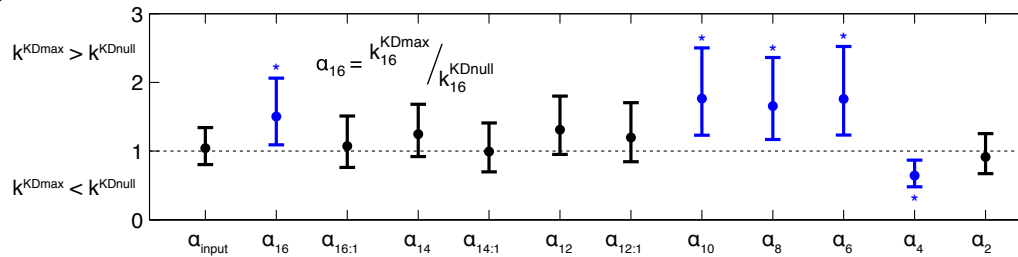
A**B**

Figure 4.7: Model-based analysis of intracellular acylcarnitine time course data in ACADS knockdown cells. **A:** Results from fitting the selected model as shown in Figure 4.5 to the null (KDnull, shACADS^{null}) and maximal ACADS knockdown (KDmax, shACADS^{max}) data (values of four independent experiments are expressed as mean $\pm$ SD). **B:** Comparison of reaction rates in the null and maximal knockdown model. Reaction-specific α -values of 1 denote no difference between the shACADS^{null} and shACADS^{max} experiments. Best solutions for α -values are represented as dots with corresponding 95% confidence intervals. *: Compared to the null knockdown, in the maximal knockdown reaction rate k_4 is significantly decreased ($\alpha_4 = 0.6$, $p < 0.05$), while reaction rates k_{16} , k_{10} , k_8 and k_6 are significantly increased.

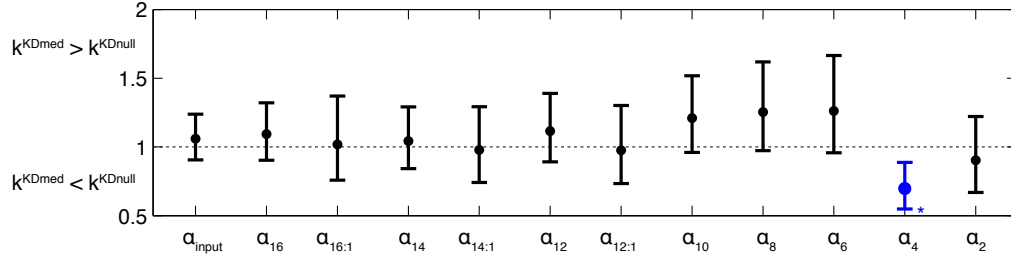


Figure 4.8: Comparison of reaction rates between the null (KDnull, shACADS^{null}) and intermediate ACADS knockdown (KDmed, shACADS^{med}) based on fitting the FAO model to experimental data. Reaction-specific α -values of 1 denote no difference between the shACADS^{null} and shACADS^{med} experiments. Best solutions for α -values are represented as dots with corresponding 95% confidence intervals. *: Compared to the null knockdown, in the intermediate knockdown reaction rate k_4 is significantly decreased ($\alpha_4 = 0.7$, $p < 0.05$).

α -values greater than 1 predict increased reaction rates in shACADS^{max} as compared to shACADS^{null}. The model-based comparison predicts for the maximal ACADS knockdown results a significant decrease of the reaction rate k_4 , i.e. the conversion rate of C4:0 to C2:0 ($p < 0.05$, $\alpha_4 = 0.6$, Figure 4.7B), reflecting the predominant role of the ACADS enzyme for this conversion reaction [16] and the observed increase of C4:0 levels due to the knockdown (Figure 4.3). Moreover, a significant increase of the reaction rates k_{16} , k_{10} , k_8 and k_6 upon maximal ACADS knockdown ($p < 0.05$) is predicted, Figure 4.7B), reflecting the observed decreased levels of C16:0-, C10:0-, C8:0- and C6:0-acylcarnitine in shACADS^{max} cells (Figure 4.3). Notably, also intermediate ACADS knockdown was sufficient to decrease specifically the k_4 reaction rate ($p < 0.05$, $\alpha_4 = 0.7$, Figure 4.8), whereas all other reaction rates were not affected, supporting the specificity of the regulated shACADS cell model. C2:0-acylcarnitine levels were not altered upon ACADS knockdown (Figure 4.7). We interpret the predicted increased reaction rates k_{16} , k_{10} , k_8 and k_6 as a compensation mechanism, which however was not explained by increased protein levels of acyl-CoA dehydrogenases ACADM, ACADL and ACADVL, measured by western blot (see Figure 4.9).

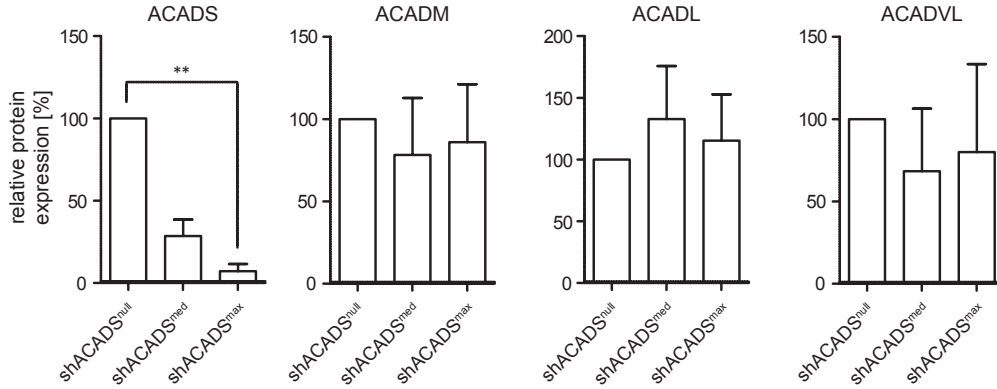


Figure 4.9: Western blot analysis of shACADS Huh7 cell lysates reveals no effect of ACADS knockdown on the protein concentrations of acyl-CoA dehydrogenases ACADM, ACADL and ACADVL. Error bars denote standard deviation from four experiments. **: $p < 0.01$.

4.4 Extracellular acylcarnitine profiles of shACADS cells suggest a direct contribution to the plasma mQTL phenotype

Metabolites related to fatty acid oxidation are measured in blood [265] and show strong GWAS signals for specific enzymatic reactions. As it remains unclear to which extent plasma metabolites reflect intracellular processes, we assessed extracellular acylcarnitine levels in the cell culture medium supernatant (we note that only a subset of acylcarnitines was detectable in the medium supernatant, see Figure 4.11). At baseline we found no ACADS-dependent differences in the levels of any measured extracellular acylcarnitine in shACADS^{null} and shACADS^{max} cells (time point $t = 0$; Figure 4.11). Moreover, in both, shACADS^{null} and shNTC cells (Figure 4.10) we observed no effect of palmitic acid loading on C4:0-acylcarnitine levels, despite the observed intracellular increase (Figure 4.3). Strikingly, in shACADS^{max} cells we found a significant increase of extracellular C4:0-acylcarnitine levels as compared to shACADS^{null} cells at each of the four measured time points after palmitic acid loading ($p < 0.05$ or $p < 0.01$; Figure 4.10). We found no increase of extracellular C6:0- and C16:0-acylcarnitines and solely a not significant trend for C2:0- and C14:0-acylcarnitines in the shACADS^{max} cells (Figure 4.11). The ACADS dependent, specific increase of extracellular C4:0-acylcarnitine (Figure 4.10) reflects the increase of intracellular (Figure 4.3) C4:0-acylcarnitine levels. Notably, in-

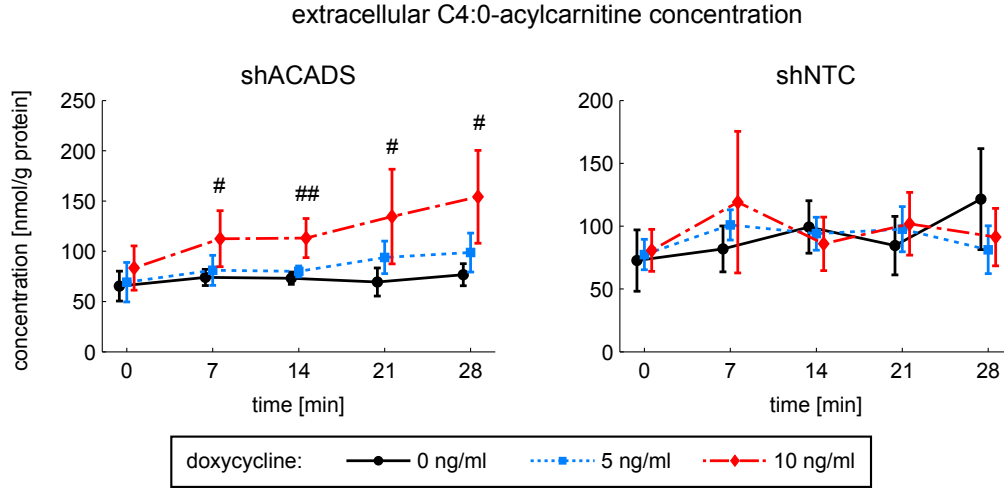


Figure 4.10: Time courses of extracellular C4:0-acylcarnitine (C4:0) after palmitic acid loading. Palmitic acid was added to induce FAO in shACADS and shNTC Huh7 cells. C4:0, which accumulates within the cell due to the ACADS knockdown (shACADS), also accumulates in the supernatant. Gradual expression of shRNA was induced using 0, 5 and 10 ng/ml doxycycline. C4:0 was measured in supernatants before palmitic acid loading and after 7, 14, 21 and 28 minutes. Values of four independent experiments are expressed as mean $\pm$ SD. Further time courses of extracellular acylcarnitines can be found in Figure 4.11. ##: $p < 0.01$, #: $p < 0.05$; comparison between shACADS^{null} (dox 0 ng/ml) and shACADS^{max} (dox 10 ng/ml) knockdown.

tracellular C4:0-acylcarnitine level in shACADS^{max} decreased during the time course as compared to the first measured time point, whereas extracellular C4:0-acylcarnitine levels increased throughout the entire time course.

4.5 Discussion

Numerous GWAS identified disease- and quantitative trait-associated, non-coding genetic variants, but in most cases the affected molecular mechanisms remain elusive. Population based metabolite QTL data found the common ACADS locus to be associated with plasma C4:0-acylcarnitine levels [81, 98, 104, 172], supposing a direct effect on FAO by affecting the ACADS enzyme with the major substrate C4:0-acyl-CoA [77]. We chose this obvious candidate gene to go beyond association data. To unveil the physiological relevance of functional metabolic phenotypes assessed in GWAS, experimental verification in appropriate cell types is needed. However such analysis is often hampered

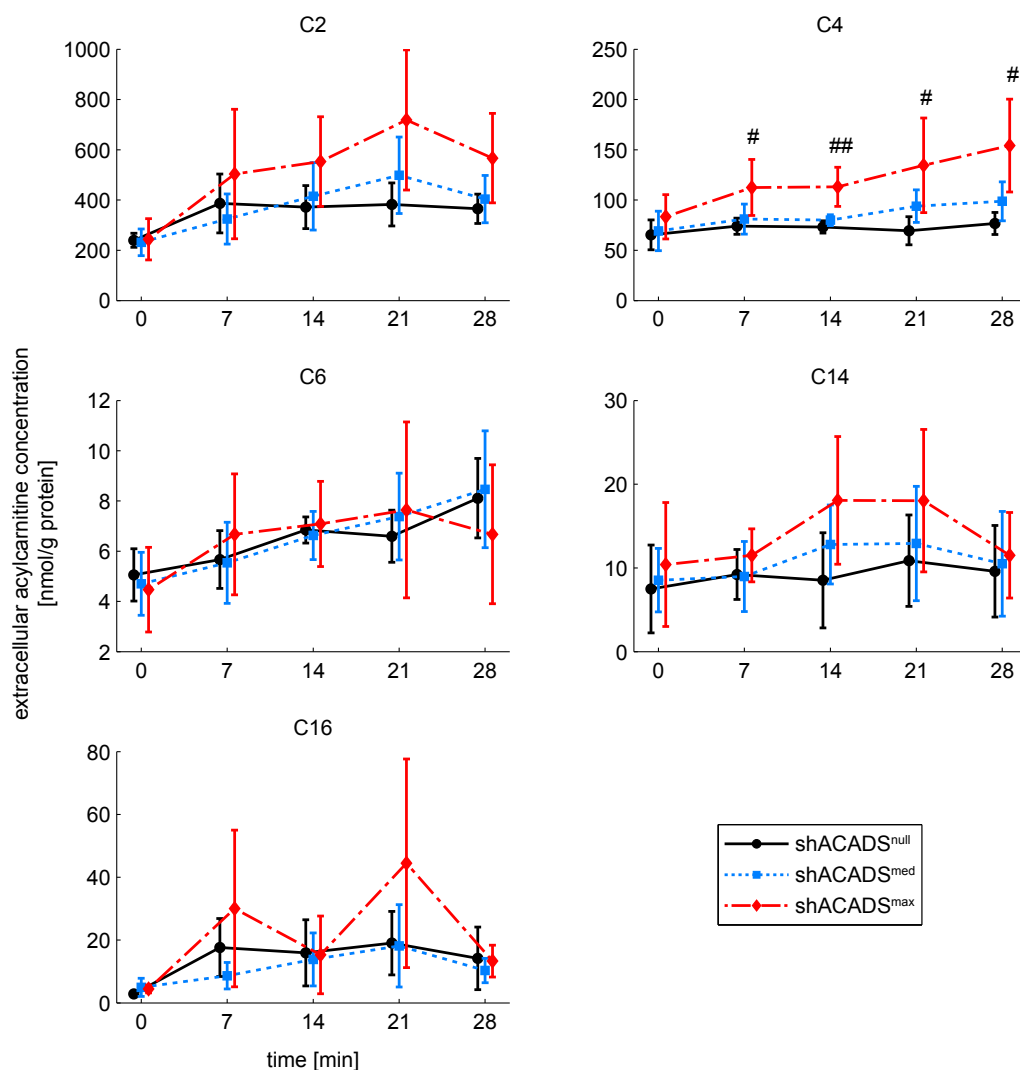


Figure 4.11: Time courses of extracellular acylcarnitines after palmitic acid loading in shACADS Huh7 cells. Palmitic acid was added to induce fatty acid oxidation in shACADS knockdown cells treated with 0, 5 and 10 ng/ml doxycycline, respectively, for shRNA induction (shACADS^{null}, shACADS^{med}, shACADS^{max}). Acylcarnitines were measured in supernatants before palmitic acid loading and after 7, 14, 21 and 28 minutes. Values of four independent experiments are expressed as mean $\pm$ SD. ##: $p < 0.01$, #: $p < 0.05$; comparison between shACADS^{null} and shACADS^{max}.

by limited accessibility of primary human cells. For this reason we generated a lentiviral-based, gradual knockdown of endogenous ACADS protein levels in the human Huh7 liver cell line for modeling the genotype-dependent reduction of ACADS expression at cellular level. Reduced ACADS mRNA and protein levels resulted in a baseline accumulation of C4:0-acylcarnitine, which resembles the SNP associated phenotype found in population studies [81, 98, 104, 172].

So far it remains unknown how reduced ACADS expression affects the overall kinetics of fatty acid oxidation. For this reason we assessed metabolite dynamics during FAO in a hepatocyte ACADS knockdown model, i.e. measured short-term time courses of intra- and extracellular acylcarnitine levels after incubation with palmitic acid, a major substrate of the FAO pathway [94]. Besides individual analysis of metabolite time courses, we developed a mathematical model of FAO tailored specifically for our experimental data. Diverse models have been supposed to study the dynamic properties of FAO [129, 164, 250], largely differing in the level of complexity and the detailed description of biochemical reactions. Direct transfer of models is limited due to different metabolite specimen measured. Here, to facilitate model calibration by reducing the model complexity, we combined subsequent reaction steps. Our model enables capturing the dynamics and interactions of individual metabolites within the FAO pathway at a system level. To account for possible in- and outflux reactions [144, 255–257] which are essential for explaining the observed metabolite dynamics, we performed a model selection using the Bayesian information criterion (BIC) score [121]. The selected model with the smallest BIC score includes three influx reactions with rates $k_{16:1in}$, $k_{12:1in}$ and k_{4in} , i.e. influx at the level of C16:1, C12:1 and C4:0, in addition to the fundamental FAO cascade reactions. This model is a simplified representation of the FAO pathway, developed to find knockdown-specific differences in metabolite dynamics, but not exact kinetic rates. A similar strategy to assess beta-oxidation activity was successfully applied to metabolomics data of the human challenging study [133] presented in Chapter 3.

In a model-based comparison of the dynamic changes resulting from ACADS knock-down we find a significant decrease of C4-related reaction rates, reflecting the reported increased plasma C4:0-acylcarnitine levels in GWAS with metabolic traits [81, 98, 104, 172]. Moreover, the model predicts increased medium- and long-chain fatty acid-related reactions rates, which we interpret as a compensatory effect of the impaired C4 turnover.

An effect on medium- and long-chain acyl dehydrogenases (ACADM, ACADL and ACADVL) expression level based on our predicted increased reaction rates was excluded by western-blot analysis, suggesting a direct effect on enzyme activity rather than protein level. However, the molecular basis for such mechanism remains elusive and requires future studies. Potential short-term effects on FAO metabolites need to be evaluated in more detail by shorter sampling intervals immediately after the palmitic acid loading. For an improved description of FAO pathway dynamics, the presented model for the human liver cell data may be combined with a recently reported model of fatty acid beta-oxidation that was evaluated on data from isolated mitochondria [250]. This will allow for the incorporation of further pathway enzymes and respective substrate specificities. As metabolic pathways are highly connected, the FAO model may also be linked with published models of glucose metabolism, citric acid cycle and fatty acid biosynthesis [33, 120, 274] and evaluated on measurements of carbohydrate metabolism.

Of note, we found a concurrent increase of C4:0-acylcarnitine levels in the supernatant of cells along with the intracellular accumulation of C4:0-acylcarnitine specifically in ACADS knockdown cells. The extracellular increase was highly specific for C4:0 and not observed for any other measured acylcarnitine, reflecting the specific GWAS association of the rs2014355 genotype with C4 and with no medium- and long-chain fatty acid substrates [104]. The efflux of C4:0-acylcarnitine solely occurs in cells with reduced ACADS expression levels which represent the genotype-specific effect of the ACADS locus. This finding suggests a cellular transport mechanism contributing to the high plasma concentrations of the FAO intermediate acylcarnitine-specimen, which leads to strong GWAS signals related to enzymatic reactions. In future studies, the here presented cell model may serve as a versatile tool for the identification of so far unknown transporter mechanisms responsible for acylcarnitine efflux [51, 256]. Combining this information with mechanistic models might lead to further insights about the interplay between intracellular processes and molecular signals measured in blood.

In conclusion, we demonstrated an altered metabolic rate of C4:0-aclycaritine catalyzed by ACADS in cellular models with an allele-dependent decrease of ACADS gene expression and resulting lower protein levels. For the first time we present data on the effect of altered ACADS expression levels on the immediate short-term acylcarnitine concentrations after palmitic acid loading, reflecting the *in vivo* FAO. Evidence is provided that

the alteration of one component of the complex interaction system of the FAO leads to regulation at multiple steps in the pathway cascade. Regarding the limited availability of relevant cell models, i.e. genotyped primary human cells [185] accessible to functional assays, the here presented cellular model may further help to assess the functional consequences of moderate effects on gene expression driven by common genetic variants.

Chapter 5

Summary and outlook

Metabolism differs substantially between individuals, as metabolic phenotypes are influenced by various intrinsic and extrinsic factors. For a better understanding of genetic and nutritional influences on human metabolism, we have analyzed metabolite profiles on different biological scales, ranging from cross-sectional population data over time-resolved *in vivo* physiological challenging results to *in vitro* experiments using genetically modified cell lines. Modern high-throughput methods allow for the simultaneous quantification of hundreds of metabolite levels as readouts for metabolic functions. Yet the analysis and interpretation of the multivariate measurements remains challenging. Although biological systems are complex, there is fundamental knowledge about the underlying biochemical properties and principles of biomolecular organization [90]. For a better understanding of biological functions at a system level, experimental measurements thus can be combined with network information about the interplay of individual components in biological systems, in order to improve the analysis of high-dimensional, large-scale *omics* data [122, 123].

In a network-based approach, we therefore combined experimental data with established biological information in terms of biochemical pathways for an improved analysis of genetic and nutritional effects on human metabolism. In this thesis, I have applied and evaluated this strategy for three specific biological scenarios (Chapter 2 - 4). The analysis was performed on different biological scales for distinct experimental setups, ranging

from human population data (Chapter 2) over challenge-induced *in vivo* results (Chapter 3) to *in vitro* experiments using genetically modified human liver cells (Chapter 4). For each scenario, the key task was to develop appropriate models for the analysis of metabolomics data as the main readout of biological functions.

Network-based metabolite ratios for an improved functional characterization of genome-wide association study results

For better understanding of genetic effects on metabolism at a system level, we applied the network-based approach for the selection of biochemically related metabolite ratios in genome-wide association studies with metabolic traits (mGWAS) in Chapter 2. It was shown for mGWAS that metabolite ratios reduced the overall biological variability in population data and resulted in robust statistical associations. In a biochemical interpretation, the ratio between product-substrate metabolite pairs can be interpreted as a proxy of the corresponding enzymatic reaction rate. Usually all possible ratio combinations are selected for association tests in mGWAS. However, with more metabolites being detectable, the increasing number of possible ratios becomes challenging from a statistical, computational and interpretational point-of-view. We therefore suggested a network-based approach by selecting only closely connected metabolites in a given metabolic network. Input networks for ratio selection were derived from public pathway databases or reconstructed from metabolomics data. The feasibility of this approach was first tested on *in silico* data derived from simulated reaction networks. Especially for small genetic effect sizes of single-nucleotide polymorphism (SNP), network-based metabolite ratios (NBRs) improved the ratio-SNP association results compared to the “all ratios” approach. We further evaluated the NBR approach on published mGWAS association results and compared reported “all ratio”-SNP hits with results obtained by selecting only NBRs as candidates for association tests. NBR-candidates accounted for more than 80% of all significant ratio-SNP associations. Moreover, the NBR analysis predicted 10 new associations between genetic and metabolic phenotypes.

Taken together, we have shown in Chapter 2 that the network-based ratio approach increases the statistical power, lowers computational demands, facilitates the functional characterization of mGWAS results and allows for the identification of new associations.

Modeling metabolic pathways under steady state conditions in a human challenging study

In order to assess nutritional and physiological effects on human metabolism, we analyzed time-resolved metabolomics profiles from a human challenging study in Chapter 3. For the analysis of individual metabolite levels, we applied the concept of network-based selection of metabolite ratios to a specific biochemical pathway. A 36h fasting period induced subject-specific responses in metabolic profiles, which were highly different between the individuals. In order to analyze the interindividual variation, we developed a linear model of mitochondrial fatty acid beta-oxidation. Under steady-state conditions, compound conversion rates can be derived from concentration ratios of pathway metabolites. Investigating the relationship between individual phenotypic parameters and metabolic profiles revealed that model-driven ratios as readouts for the individual metabolic capacity facilitate the characterization of distinct metabolic phenotypes. Compared to absolute metabolite concentrations, model-driven ratios substantially improved statistical correlations with physiological parameters like blood sugar and free carnitine, but also for anthropometric parameters like body mass index and total fat mass. To sum up, combining time-resolved metabolomics data with established knowledge about metabolite connections in biochemical pathways allows for an improved analysis of metabolic systems under challenging perturbations and a better understanding of individual physiological and metabolic phenotypes.

Quantitative modeling of an *in vitro* enzyme knockdown in the fatty acid beta-oxidation pathway

In Chapter 4, we analyzed *in vitro* metabolomics data to understand the impact of genetic variation on metabolic phenotypes at cellular level. In particular, we studied mGWAS-derived associations between butyrylcarnitine (C4), the transport form of a short-chain fatty acid, and genetic variants in the locus of the mitochondrial fatty acid beta-oxidation (FAO) enzyme ACADS. Using a human liver cell line as an *in vitro* model for the gradual knockdown of endogenous ACADS, we measured time-resolved concentrations of FAO intermediate metabolites after fatty acid loading. This allowed us to quantify the effect of ACADS protein levels on FAO metabolite concentrations during beta-oxidation activity. In order to simultaneously capture the dynamics and interactions of individual metabolites at a system level, we developed a mathematical model of

the FAO pathway, which describes the fundamental reactions during the breakdown of fatty acids. Based on the reaction rates inferred from our model and experimental data, we compared the dynamical changes between wild-type and ACADS knockdown conditions statistically. A knockdown-specific decrease in C4-related reactions is in accordance with reported mGWAS results. In addition, the model infers increased medium- and long-chain fatty acid-related reactions rates, which we interpreted as a compensatory effect related to the impaired C4 turnover. Such findings have not been reported in association studies before. A gene regulatory effect on the expression levels of specific beta-oxidation enzymes which catalyze medium- and long-chain fatty acid conversion reactions could be ruled out by western-blot analysis, suggesting a direct effect on enzyme activity.

In summary, the quantitative modeling of an *in vitro* enzyme knockdown allows to translate population-based statistical associations from GWAS into the functional characterization of the underlying biochemical processes in mitochondrial fatty acid oxidation and can be applied in general to analyze effects of genotype-metabolite interactions.

Extensions and future directions for network-based analysis of metabolomics data

Metabolomics and other high-throughput profiling methods provide large-scale measurements of molecular parameters as readouts of biological functions for various experimental setups in biomedical research. The network-based approaches presented in this thesis can be applied for a system-wide analysis of these multivariate measurements. In the following part, we will discuss possible steps for extending the approach to address upcoming methodological and biological questions.

Metabolite network information. Our analysis of metabolomics data is based on metabolic networks. The relationship between metabolites is obtained both from pathway databases and from data-driven reconstructed networks. In addition to the three databases KEGG [114], BiGG [219] and EHMN [149] that we used, further pathway resources can be included, for instance Recon 2 [243], Reactome [47] and MetaCyc [31]. Even if the general level of agreement between databases is still quite low [234, 235], integrating the different resources will lead to a better representation of cellular metabolism. Accounting for cell- or tissue-specific reactions will in addition allow for a better description of metabolic pathways at organ levels [23]. A more involved approach could be to combine experimentally-driven metabolic connections with established knowledge from databases. The reconstructed relations between compounds, which have not been annotated, and compounds, for which pathway information is available, will then lead to an improved quality of all resources for biochemical networks.

Data-driven reconstruction of metabolic networks. Many measured metabolites of current metabolomics data are not annotated in the above-mentioned pathway databases. In addition, untargeted detection methods allow to quantify compounds which have not been characterized chemically [117, 135, 254]. For these cases, data-driven reconstruction of metabolic networks can complement the knowledge-based information. In our research we used Gaussian graphical modeling to select metabolite ratios. Depending on the specific properties of the measured data and the experimental setup, other network inference methods from different research fields can be applied to obtain metabolite networks [17, 153, 205]. Several of these approaches have been evaluated and

compared on synthetic metabolomics data [27] and could be used for experimental metabolomics data.

For the data-driven reconstruction of biochemical pathways, it can be beneficial to include directional information of edges in order to account for irreversible reaction steps. Several methods have been proposed to estimate directional edges for graphical models [70, 181, 280]. The statistically-derived relation between metabolites and the direction of an network edge needs to be distinguished from causality. For this reason, results have to be evaluated carefully with respect to the biological implications [21]. The *in silico* method for generating population mGWAS data, which we developed in this thesis, provides a suitable evaluation framework for the suggested methodological extensions, as we can examine new approaches in a well-defined setup.

Marker for metabolic phenotypes. The ratio between concentrations of product and substrate metabolites in biochemical reactions can be “viewed as a proxy for the reaction rate” [236]. In order to account for reactions involving more than two metabolites, the pairwise ratio concept needs to be extended to higher-order interactions, for instance to ratios of sums of concentrations. Moreover, the use of non-linear relationships between metabolites besides ratios might be considered. In addition to the selection of connected metabolite pairs as surrogates for metabolic reactions, we may also choose a set of metabolites as biochemical markers based on modular subgroups in databases or reconstructed networks [39, 162] and combine individual signals. These markers could also be derived from differential biological networks [103, 246]. For instance, using specific metabolic networks for healthy and diseased condition, single compounds or groups of metabolites reflecting biochemical pathways could then be used as markers for metabolic phenotypes.

Identification of causal variants. Even further improvements for GWAS study designs might not solve the problem of finding the underlying causal genetic variant in all association results. The mechanistic interpretation of detected markers in intergenic and intronic regions [61, 92] and the detection of possible synthetic associations [55] will remain challenging. Also potential genotype-environment interactions due to epigenetics or genotype-genotype interactions mask the underlying biological mechanism of complex diseases [68, 130]. Especially for common diseases it is under debate whether common alleles, rare alleles or a combination of both contribute to the observed pheno-

type [258]. Just increasing the sample size might not allow for the identification of rare variants [80]. Family-based [209] or extreme-trait [15] exome or whole-genome sequencing provide strategies to detect rare alleles, but also impose new statistical challenges for association tests due to multiparametric data for small sample sizes [12].

Multiple testing correction and small sample numbers. For studies with small sample numbers, low statistical power can be a limiting factor. In this work, we addressed statistical challenges for mGWAS due to multiple testing correction by selecting biologically meaningful metabolite ratios. As association tests are performed for each ratio-SNP pair, the statistical significance also depends on the number of SNPs. Several approaches have been proposed to correct for multiple testing in GWAS [26, 249], such as conservative Bonferroni correction, controlling the false discovery rate [18] or permutation testing [25, 183]. Other methods aim to predict the effective number of independent statistical tests by considering the linkage disequilibrium among SNPs [73, 178]. The NBR approach may also be combined with methods accounting for the inherent correlation between SNPs, thus reducing the number of both ratios and SNPs for multiple testing correction.

In Chapter 2, we evaluated the NBR approach on metabolite-gene associations from GWAS data, but the concept of network-based metabolite ratios is not necessarily restricted to genetic study designs. For upcoming studies, for instance with case-control design, large-scale metabolomics data and small sample numbers, the NBR approach can increase the statistical power, lower computational demands and facilitate the functional interpretation of association results. For instance, we identified metabolites in a population study that associate with asthma risk loci and asthma disease status [208]. Considering pathway connections between metabolites, we can select specific ratios for association analyses, which might allow for a better annotation of genetic functions related to asthma.

Experimental studies and measured data. Results from genome-wide association studies with metabolic traits showed on a population level strong variation in metabolite profiles between individuals. Several longitudinal studies have been performed to analyze this variation with respect to short and long time intervals [10, 48, 216, 278], but also as a response to diet and environmental factors [66, 91, 96, 133]. As we have presented in Chapter 3, especially catabolic conditions like fasting or exercise induce strong individual

metabolic responses. For the analysis of metabolism-related genes, the use of catabolic or nutritional challenging in the experimental setup will allow for studying the genetic effects under different physiological conditions [56, 76, 277, 282].

Based on reported GWAS results biological hypotheses about the function of genes can be generated [4], but for a mechanistic characterization, cellular, tissue or *in vivo* models are required. For example, a study using nine genes, which were found in a GWAS to be associated to serum metabolomic traits, revealed high mRNA expression in liver tissue from healthy donors [161]. Several of these nine genes are involved in lipid metabolism, including the gene coding for the short-chain acyl-coenzyme A dehydrogenase ACADS. The impact of expression levels of these genes and their functional role for hepatic metabolism could be tested in an *in vitro* liver cell model. Besides, a recent study suggested a functional role of fatty acid oxidation in acetylation of mitochondrial proteins [194]. These posttranslational modifications are involved in the regulation of mitochondrial metabolism. The analysis of acetylation profiles during fatty acid loading in liver cells will uncover potential regulatory links between protein functions and metabolism, that could also be used to extend *in silico* models of mitochondrial metabolism with a regulatory layer and feedback control [11]. In addition, data on microRNAs, which are involved in the regulation of metabolism [137, 148, 195, 198, 211] or metagenomics data about the role of microbial communities in metabolism [101] could be incorporated.

Fatty acid beta-oxidation model. For an improved description of FAO pathway dynamics, the presented human liver cell data may be analyzed with a recently reported model of fatty acid beta-oxidation that was evaluated on isolated mitochondrial data [250]. This will allow for the incorporation of specific pathway enzymes and respective substrate specificities. As metabolic pathways are highly connected, FAO models may also be linked with published models of fatty acid biosynthesis, glucose metabolism or citric acid cycle [33, 120, 274]. Kinetic parameters of enzymes for specific experimental conditions then need to be compared to published results from primary literature, which is collected in databases like BRENDA [223] or SABIO-RK [268]. It is important to acknowledge at this point that any model is merely an approximate description of biological reality and may be refined to explain the observed data. Experimental results will allow for extending the model, for instance by elucidating active metabolic pathways *in vitro*. For the analysis of time-resolved metabolomics data we relied on well-established

pathway knowledge, allowing for the development of dynamic models to describe the time-dependent metabolite changes. Besides mechanistic models, approaches like functional data analysis [231] can be applied to analyze the metabolite timecourses, which are provided by both *in vivo* and *in vitro* study designs.

Fatty acid oxidation footprints are measured in blood [265] and show strong GWAS signals related to enzymatic reactions. As it remains unclear to which extent metabolites in blood reflect intracellular processes in general, future models need to incorporate extracellular compartments and also include information about membrane transporters [215]. Tracer studies using stable isotope-assisted metabolomics can be used to assess transport reactions between compartments [46]. In addition, upcoming metabolomics techniques can detect biochemical compounds at single-cell resolution [283], which will provide further details about metabolite transport and exchange reactions. Combining this information with mechanistic models might lead to further insights about the interplay between intracellular processes and molecular signals measured in blood.

In order to understand the impact of molecular, cellular, physiological and environmental factors on an organism's phenotype, it is important to investigate the affected players across all layers of biological function, i.e. genes, transcripts, proteins and metabolic reaction compounds. Modern high-throughput techniques now provide genomics, transcriptomics, proteomics and metabolomics measurements [102]. Future approaches need to provide tools for selecting the most informative biological readouts out of the multiparametric data from all layers of biological functions across multiple scales of sampling (e.g. populations, individuals, organs, cells). Considering also dynamical data (steady-state, time-resolved) will improve our understanding of biological functions in the context of dynamic biomolecular organization [105, 106, 206]. It will be challenging to address the above-mentioned extensions for a network-based analysis of *omics* data across different functional layers, but will allow us to better understand metabolic processes in general.

Conclusion

Metabolism is highly variable between individuals, even though the underlying key processes follow the same physicochemical laws and biological principles. Metabolic reactions are influenced by genetic and physiological factors, which can be measured on different biological scales, ranging from population data to cellular models. The combination of established knowledge about biochemical pathways with computational models facilitated the analysis of multivariate data for different study designs. As biochemical processes with many coupled reactions can be studied at a system level, a model-based analysis is a promising approach to obtain deeper insights into the interplay between genetic effects, nutrition and metabolism.

Bibliography

- [1] Adamski, J. Genome-wide association studies with metabolomics. *Genome Med*, 4(4):34, 2012.
- [2] Adamski, J. and Suhre, K. Metabolomics platforms for genome wide association studies-linking the genome to the metabolome. *Curr Opin Biotechnol*, pages 1–9, 2012.
- [3] Akaike, H. A new look at the statistical model identification. *IEEE Trans Automat Contr*, 19(6):716–723, 1974.
- [4] Ala-korpela, M., Kangas, A.J., and Soininen, P. Quantitative high-throughput metabolomics: a new era in epidemiology and genetics. *Genome Med*, 4(4):36, 2012.
- [5] Allen, J., Davey, H.M., Broadhurst, D., Heald, J.K., Rowland, J.J., Oliver, S.G., and Kell, D.B. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol*, 21(6):692–6, 2003.
- [6] Altmaier, E., Ramsay, S.L., Graber, A., Mewes, H.W., Weinberger, K.M., and Suhre, K. Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication. *Endocrinology*, 149(7):3478–89, 2008.
- [7] Altshuler, D., Daly, M.J., and Lander, E.S. Genetic mapping in human disease. *Science*, 322(5903):881–8, 2008.
- [8] Amendt, B.A., Greene, C., Sweetman, L., Cloherty, J., Shih, V., Moon, A., Teel, L., and Rhead, W.J. Short-chain acyl-coenzyme A dehydrogenase deficiency. Clinical and biochemical studies in two patients. *J Clin Invest*, 79(5):1303–9, 1987.
- [9] Arkin, A. A Test Case of Correlation Metric Construction of a Reaction Pathway from Measurements. *Science (80-)*, 277(5330):1275–1279, 1997.
- [10] Assfalg, M., Bertini, I., Colangiuli, D., Luchinat, C., Schäfer, H., Schütz, B., and Spraul, M. Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci U S A*, 105(5):1420–4, 2008.
- [11] Atkinson, D.E. Biological feedback control at the molecular level. *Science*, 150(3698):851–7, 1965.
- [12] Bansal, V., Libiger, O., Torkamani, A., and Schork, N.J. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet*, 11(11):773–85, 2010.

- [13] Barabási, A.L., Gulbahce, N., and Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*, 12(1):56–68, 2011.
- [14] Baranzini, S.E., Galwey, N.W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., Wu, W., Uitdehaag, B.M.J., Kappos, L., Polman, C.H., Matthews, P.M., Hauser, S.L., Gibson, R.A., Oksenberg, J.R., and Barnes, M.R. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet*, 18(11):2078–90, 2009.
- [15] Barnett, I.J., Lee, S., and Lin, X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet Epidemiol*, 37(2):142–51, 2013.
- [16] Bartlett, K. and Eaton, S. Mitochondrial beta-oxidation. *Eur J Biochem*, 271(3):462–469, 2004.
- [17] Barzel, B. and Barabási, A.L. Network link prediction by global silencing of indirect correlations. *Nat Biotechnol*, 31:720–5, 2013.
- [18] Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*, 57(1):289–300, 1995.
- [19] Berg, J.M., Tymoczko, J.L., and Stryer, L. *Biochemistry*. W.H. Freeman, New York, 2007.
- [20] Bhala, A., Willi, S.M., Rinaldo, P., Bennett, M.J., Schmidt-Sommerfeld, E., and Hale, D.E. Clinical and biochemical characterization of short-chain acyl-coenzyme A dehydrogenase deficiency. *J Pediatr*, 126(6):910–5, 1995.
- [21] Blair, R.H., Kliebenstein, D.J., and Churchill, G.A. What can causal networks tell us about metabolic pathways? *PLoS Comput Biol*, 8(4):e1002458, 2012.
- [22] Blow, N. Metabolomics: Biochemistry’s new look. *Nature*, 455(7213):697–700, 2008.
- [23] Bordbar, A., Feist, A.M., Usaite-Black, R., Woodcock, J., Palsson, B.O., and Famili, I. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst Biol*, 5(1):180, 2011.
- [24] Bordbar, A., Monk, J.M., King, Z.A., and Palsson, B.O. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet*, 15(2):107–20, 2014.
- [25] Browning, B.L. PRESTO: rapid calculation of order statistic distributions and multiple-testing adjusted P-values via permutation for one and two-stage genetic association studies. *BMC Bioinformatics*, 9:309, 2008.
- [26] Bush, W.S. and Moore, J.H. Chapter 11: Genome-wide association studies. *PLoS Comput Biol*, 8(12):e1002822, 2012.
- [27] Cakr, T., Hendriks, M.M.W.B., Westerhuis, J.A., and Smilde, A.K. Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics*, 5(3):318–329, 2009.
- [28] Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet*, 44(8):841–7, 2012.

- [29] Camacho, D., Fuente, A., and Mendes, P. The origin of correlations in metabolomics data. *Metabolomics*, 1(1):53–63, 2005.
- [30] Cao, H., Gerhold, K., Mayers, J.R., Wiest, M.M., Watkins, S.M., and Hotamisligil, G.S. Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell*, 134(6):933–44, 2008.
- [31] Caspi, R., Altman, T., Dreher, K., Fulcher, C.a., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., and Karp, P.D. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, 40(Database issue):D742–53, 2012.
- [32] Caulfield, M.J., Munroe, P.B., O’Neill, D., Witkowska, K., Charchar, F.J., Doblado, M., Evans, S., Eyheramendy, S., Onipinla, A., Howard, P., Shaw-Hawkins, S., Dobson, R.J., Wallace, C., Newhouse, S.J., Brown, M., Connell, J.M., Dominiczak, A., Farrall, M., Lathrop, G.M., Samani, N.J., Kumari, M., Marmot, M., Brunner, E., Chambers, J., Elliott, P., Kooner, J., Laan, M., Org, E., Veldre, G., Viigimaa, M., Cappuccio, F.P., Ji, C., Iacone, R., Strazzullo, P., Moley, K.H., and Cheeseman, C. SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med*, 5(10):e197, 2008.
- [33] Chalhoub, E., Hanson, R.W., and Belovich, J.M. A computer model of gluconeogenesis and lipid metabolism in the perfused liver. *Am J Physiol Endocrinol Metab*, 293(6):E1676–86, 2007.
- [34] Chasman, D.I., Fuchsberger, C., Pattaro, C., Teumer, A., Böger, C.A., Endlich, K., Olden, M., Chen, M.H., Tin, A., Taliun, D., Li, M., Gao, X., Gorski, M., Yang, Q., Hundertmark, C., Foster, M.C., O’Seaghdha, C.M., Glazer, N., Isaacs, A., Liu, C.T., Smith, A.V., O’Connell, J.R., Struchalin, M., Tanaka, T., Li, G., Johnson, A.D., Gierman, H.J., Feitosa, M.F., Hwang, S.J., Atkinson, E.J., Lohman, K., Cornelis, M.C., Johansson, A., Tönjes, A., Dehghan, A., Lambert, J.C., Holliday, E.G., Sorice, R., Kutalik, Z., Lehtimäki, T., Esko, T.o., Deshmukh, H., Ulivi, S., Chu, A.Y., Murgia, F., Trompet, S., Imboden, M., Coassin, S., Pistis, G., Harris, T.B., Launer, L.J., Aspelund, T., Eiriksdottir, G., Mitchell, B.D., Boerwinkle, E., Schmidt, H., Cavalieri, M., Rao, M., Hu, F., Demirkan, A., Oostra, B.A., de Andrade, M., Turner, S.T., Ding, J., Andrews, J.S., Freedman, B.I., Giulianini, F., Koenig, W., Illig, T., Meisinger, C., Gieger, C., Zgaga, L., Zemunik, T., Boban, M., Minelli, C., Wheeler, H.E., Igl, W., Zaboli, G., Wild, S.H., Wright, A.F., Campbell, H., Ellinghaus, D., Nöthlings, U., Jacobs, G., Biffar, R., Ernst, F., Homuth, G., Kroemer, H.K., Nauck, M., Stracke, S., Völker, U., Völzke, H., Kovacs, P., Stumvoll, M., Mägi, R., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., Aulchenko, Y.S., Polasek, O., Hastie, N., Vitart, V., Helmer, C., Wang, J.J., Stengel, B., Ruggiero, D., Bergmann, S., Kähönen, M., Viikari, J., Nikopensius, T., Province, M., Ketkar, S., Colhoun, H., Doney, A., Robino, A., Krämer, B.K., Portas, L., Ford, I., Buckley, B.M., Adam, M., Thun, G.A., Paulweber, B., Haun, M., Sala, C., Mitchell, P., Ciullo, M., Kim, S.K., Vollenweider, P., Raitakari, O., Metspalu, A., Palmer, C., Gasparini, P., Pirastu, M., Jukema, J.W., Probst-Hensch, N.M., Kronenberg, F., Toniolo, D., Gudnason, V., Shuldiner, A.R., Coresh, J., Schmidt, R., Ferrucci, L., Siscovick, D.S., van Duijn, C.M., Borecki, I.B., Kardia, S.L.R., Liu, Y., Curhan, G.C., Rudan, I., Gyllenstein, U., Wilson, J.F., Franke, A., Pramstaller, P.P., Rettig, R., Prokopenko, I., Witteman, J., Hayward, C., Ridker, P.M., Parsa, A., Bochud, M.,

- Heid, I.M., Kao, W.H.L., Fox, C.S., and Köttgen, A. Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet*, 21(24):5329–43, 2012.
- [35] Chis, O.T., Banga, J.R., and Balsa-Canto, E. Structural identifiability of systems biology models: a critical comparison of methods. *PLoS One*, 6(11):e27755, 2011.
- [36] Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkalolu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., and Lifton, R.P. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*, 106(45):19096–101, 2009.
- [37] Chorell, E., Moritz, T., Branth, S., Antti, H., and Svensson, M.B. Predictive metabolomics evaluation of nutrition-modulated metabolic stress responses in human blood serum during the early recovery phase of strenuous physical exercise. *J Proteome Res*, 8(6):2966–77, 2009.
- [38] Chorell, E., Svensson, M.B., Moritz, T., and Antti, H. Physical fitness level is reflected by alterations in the human plasma metabolome. *Mol Biosyst*, 8(4):1187, 2012.
- [39] Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D., and Ideker, T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*, 3(140):140, 2007.
- [40] Chung, C.M., Wang, R.Y., Chen, J.W., Fann, C.S.J., Leu, H.B., Ho, H.Y., Ting, C.T., Lin, T.H., Sheu, S.H., Tsai, W.C., Chen, J.H., Jong, Y.S., Lin, S.J., Chen, Y.T., and Pan, W.H. A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. *Pharmacogenomics J*, 10(6):537–44, 2010.
- [41] Cirulli, E.T. and Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*, 11(6):415–25, 2010.
- [42] Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res*, 15(11):1496–502, 2005.
- [43] Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. Mapping complex disease traits with global gene expression. *Nat Rev Genet*, 10(3):184–94, 2009.
- [44] Corydon, M.J., Gregersen, N., Lehnert, W., Ribes, A., Rinaldo, P., Kmoch, S., Christensen, E., Kristensen, T.J., Andresen, B.S., Bross, P., Winter, V., Martinez, G., Neve, S., Jensen, T.G., Bolund, L., and Kølvrå, S. Ethylmalonic aciduria is associated with an amino acid variant of short chain acyl-coenzyme A dehydrogenase. *Pediatr Res*, 39(6):1059–66, 1996.
- [45] Corydon, M.J., Vockley, J., Rinaldo, P., Rhead, W.J., Kjeldsen, M., Winter, V., Riggs, C., Babovic-Vuksanovic, D., Smeitink, J., De Jong, J., Levy, H., Sewell, A.C., Roe, C., Matern, D., Dasouki, M., and Gregersen, N. Role of common gene variations in the molecular pathogenesis of short-chain acyl-CoA dehydrogenase deficiency. *Pediatr Res*, 49(1):18–23, 2001.
- [46] Creek, D.J., Chokkathukalam, A., Jankevics, A., Burgess, K.E.V., Breitling, R., and Barrett, M.P. Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal Chem*, 84(20):8442–7, 2012.

- [47] Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D’Eustachio, P., and Stein, L. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*, 39(Database issue):D691–7, 2011.
- [48] Dallmann, R., Viola, A.U., Tarokh, L., Cajochen, C., and Brown, S.A. The human circadian metabolome. *Proc Natl Acad Sci U S A*, 109(7):2625–9, 2012.
- [49] De Vries, G., Hillen, T., Lewis, M., Schönfish, B., and Muller, J. *A course in mathematical biology: quantitative modeling with mathematical and computational methods*. Society for Industrial Mathematics, 2006.
- [50] Dempster, A.P. Covariance Selection. *Biometrics*, 28(1):157–175, 1972.
- [51] den Besten, G., van Eunen, K., Groen, A.K., Venema, K., Reijngoud, D.J., and Bakker, B.M. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res*, 54(9):2325–40, 2013.
- [52] Deo, R.C., Hunter, L., Lewis, G.D., Pare, G., Vasan, R.S., Chasman, D., Wang, T.J., Gerszten, R.E., and Roth, F.P. Interpreting metabolomic profiles using unbiased pathway models. *PLoS Comput Biol*, 6(2):e1000692, 2010.
- [53] Devlin, B. and Roeder, K. Genomic Control for Association Studies. *Biometrics*, 55(4):997–1004, 1999.
- [54] DeWan, A.T. Five classic articles in genetic epidemiology. *Yale J Biol Med*, 83(2):87–90, 2010.
- [55] Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol*, 8(1):e1000294, 2010.
- [56] Diekman, E.F., van Weeghel, M., Wanders, R.J.A., Visser, G., and Houten, S.M. Food withdrawal lowers energy expenditure and induces inactivity in long-chain fatty acid oxidation-deficient mouse models. *FASEB J*, 2014.
- [57] Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., and Palsson, B.O. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6):1777–82, 2007.
- [58] Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R., and Griffin, J.L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*, 40(1):387–426, 2011.
- [59] Eaton, S. Control of mitochondrial beta-oxidation flux. *Prog Lipid Res*, 41(3):197–239, 2002.
- [60] Eaton, S., Bhuiyan, A.K., Kler, R.S., Turnbull, D.M., and Bartlett, K. Intramitochondrial control of the oxidation of hexadecanoate in skeletal muscle. A study of the acyl-CoA esters which accumulate during rat skeletal-muscle mitochondrial beta-oxidation of [U-14C]hexadecanoate and [U-14C]hexadecanoyl-carnitine. *Biochem J*, 289 (Pt 1):161–168, 1993.

- [61] Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*, 93(5):779–97, 2013.
- [62] Ehlers, K. *From GWAS to functionality: association of rs2014355 in the ACADS gene locus with acylcarnitine ratio and postprandial metabolic and inflammatory activation of human PBMC*. Ph.D. thesis, Technical University Munich, 2014.
- [63] Fairfield, K.M. and Fletcher, R.H. Vitamins for chronic disease prevention in adults: scientific review. *JAMA*, 287(23):3116–26, 2002.
- [64] Famili, I., Mahadevan, R., and Palsson, B.O. k-Cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys J*, 88(3):1616–25, 2005.
- [65] Faust, K., Croes, D., and van Helden, J. Metabolic pathfinding using RPAIR annotation. *J Mol Biol*, 388(2):390–414, 2009.
- [66] Favé, G., Beckmann, M., Lloyd, A.J., Zhou, S., Harold, G., Lin, W., Tailliant, K., Xie, L., Draper, J., and Mathers, J.C. Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics*, 7(4):469–484, 2011.
- [67] Fawcett, T. An introduction to ROC analysis. *Pattern Recognit Lett*, 27(8):861–874, 2006.
- [68] Feinberg, A.P. Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447(7143):433–40, 2007.
- [69] Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol*, 48(1-2):155–71, 2002.
- [70] Freudenberg, J., Wang, M., Yang, Y., and Li, W. Partial correlation analysis indicates causal relationships between GC-content, exon density and recombination rate in the human genome. *BMC Bioinformatics*, 10 Suppl 1:S66, 2009.
- [71] Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., and Kitano, H. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proc IEEE*, 96(8):1254–1265, 2008.
- [72] Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T., and Kaneko, K. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics (Oxf)*, 1:25–31, 2005.
- [73] Gao, X., Starmer, J., and Martin, E.R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*, 32(4):361–9, 2008.
- [74] Garrod, A. The incidence of alkaptonuria: A study in chemical individuality. *Lancet*, 160(4137):1616–1620, 1902.
- [75] Gavaghan, C.L., Holmes, E., Lenz, E., Wilson, I.D., and Nicholson, J.K. An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett*, 484(3):169–74, 2000.

- [76] German, J.B., Bauman, D.E., Burrin, D.G., Failla, M.L., Freake, H.C., King, J.C., Klein, S., Milner, J.A., Peltó, G.H., Rasmussen, K.M., and Zeisel, S.H. Metabolomics in the opening decade of the 21st century: building the roads to individualized health. *J Nutr*, 134(10):2729–32, 2004.
- [77] Ghisla, S. and Thorpe, C. Acyl-CoA dehydrogenases. A mechanistic overview. *Eur J Biochem*, 271(3):494–508, 2004.
- [78] Ghosh, J.K. and Samanta, T. Model selection An overview. *Curr Sci*, 80(9), 2001.
- [79] Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, B., and van Ommen, B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr*, 82(3):497–503, 2005.
- [80] Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet*, 13(2):135–45, 2011.
- [81] Gieger, C., Geistlinger, L., Altmaier, E., Hrabé de Angelis, M., Kronenberg, F., Meitinger, T., Mewes, H.W., Wichmann, H.E., Weinberger, K.M., Adamski, J., Illig, T., and Suhre, K. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet*, 4(11):e1000282, 2008.
- [82] Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778):1355–9, 2006.
- [83] Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., and Kell, D.B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol*, 22(5):245–52, 2004.
- [84] Gowda, G.A.N., Zhang, S., Gu, H., Asiago, V., Shanaiah, N., and Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn*, 8(5):617–33, 2008.
- [85] Graber, R., Sumida, C., and Nunez, E.A. Fatty acids and cell signal transduction. *J Lipid Mediat Cell Signal*, 9(2):91–116, 1994.
- [86] Gregersen, N., Winter, V.S., Corydon, M.J., Corydon, T.J., Rinaldo, P., Ribes, A., Martinez, G., Bennett, M.J., Vianey-Saban, C., Bhala, A., Hale, D.E., Lehnert, W., Kmoch, S., Roig, M., Riudor, E., Eiberg, H., Andresen, B.S., Bross, P., Bolund, L.A., and Kølvrå, S. Identification of four new mutations in the short-chain acyl-CoA dehydrogenase (SCAD) gene in two patients: one of the variant alleles, 511C–T, is present at an unexpectedly high frequency in the general population, as was the case for 625G–A, together. *Hum Mol Genet*, 7(4):619–627, 1998.
- [87] Griffin, J.L. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philos Trans R Soc Lond B Biol Sci*, 361(1465):147–61, 2006.
- [88] Guldberg, C.M. and Waage, P. Über die chemische Affinität. *J Prakt Chem*, 127:69–114, 1879.

- [89] Hartsperger, M.L., Blöchl, F., Stümpflen, V., and Theis, F.J. Structuring heterogeneous biological information using fuzzy clustering of k-partite graphs. *BMC Bioinformatics*, 11(1):522, 2010.
- [90] Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–52, 1999.
- [91] Heinzmann, S.S., Merrifield, C.A., Rezzi, S., Kochhar, S., Lindon, J.C., Holmes, E., and Nicholson, J.K. Stability and Robustness of Human Metabolic Phenotypes in Response to Sequential Food Challenges. *J Proteome Res*, 11(2):643–655, 2012.
- [92] Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. 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, 2009.
- [93] Hirschhorn, J.N. and Daly, M.J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2):95–108, 2005.
- [94] Hodson, L., Skeaff, C.M., and Fielding, B.a. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog Lipid Res*, 47(5):348–80, 2008.
- [95] Holle, R., Happich, M., Löwel, H., Wichmann, H.E., and Others. KORA—a research platform for population based health research. *Gesundheitswesen*, 67:S19, 2005.
- [96] Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., Daviglus, M.L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J.K., and Elliott, P. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, 453(7193):396–400, 2008.
- [97] Holmes, E., Wilson, I.D., and Nicholson, J.K. Metabolic phenotyping in health and disease. *Cell*, 134(5):714–7, 2008.
- [98] Hong, M.G., Karlsson, R., Magnusson, P.K.E., Lewis, M.R., Isaacs, W., Zheng, L.S., Xu, J., Grönberg, H., Ingelsson, E., Pawitan, Y., Broeckling, C., Prenni, J.E., Wiklund, F., and Prince, J.a. A genome-wide assessment of variability in human serum metabolism. *Hum Mutat*, 34(3):515–24, 2013.
- [99] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. COPASI—a COMplex PATHway SIMulator. *Bioinformatics*, 22(24):3067–74, 2006.
- [100] Houten, S.M. and Wanders, R.J.a. A general introduction to the biochemistry of mitochondrial fatty acid beta-oxidation. *J Inherit Metab Dis*, 2010.
- [101] Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., Jones, P., Leinonen, R., McAnulla, C., Maguire, E., Maslen, J., Mitchell, A., Nuka, G., Oisel, A., Pesseat, S., Radhakrishnan, R., Rocca-Serra, P., Scheremetjew, M., Sterk, P., Vaughan, D., Cochrane, G., Field, D., and Sansone, S.A. EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res*, 42(Database issue):D600–6, 2014.

- [102] Ideker, T., Dutkowski, J., and Hood, L. Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell*, 144(6):860–3, 2011.
- [103] Ideker, T. and Krogan, N.J. Differential network biology. *Mol Syst Biol*, 8(565):565, 2012.
- [104] Illig, T., Gieger, C., Zhai, G., Römisch-Margl, W., Wang-Sattler, R., Prehn, C., Altmaier, E., Kastenmüller, G., Kato, B.S., Mewes, H.W., Meitinger, T., de Angelis, M.H., Kronenberg, F., Soranzo, N., Wichmann, H.E., Spector, T.D., Adamski, J., and Suhre, K. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*, 42(2):137–41, 2010.
- [105] Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L.S., Hämäläinen, E., Jousilahti, P., Kangas, A.J., Männistö, S., Savolainen, M.J., Jula, A., Leiviskä, J., Palotie, A., Salomaa, V., Perola, M., Ala-Korpela, M., and Peltonen, L. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol*, 6(441), 2010.
- [106] Inouye, M., Ripatti, S., Kettunen, J., Lyytikäinen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J., Kähönen, M., Perola, M., Salomaa, V., Raitakari, O., Lehtimäki, T., Taskinen, M.R., Järvelin, M.R., Ala-Korpela, M., Palotie, A., and de Bakker, P.I.W. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet*, 8(8):e1002907, 2012.
- [107] Jamshidi, N. and Palsson, B.O. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, 4(171):171, 2008.
- [108] Jethva, R., Bennett, M.J., and Vockley, J. Short-chain acyl-coenzyme A dehydrogenase deficiency. *Mol Genet Metab*, 95(4):195–200, 2008.
- [109] Jia, P., Wang, L., Fanous, A.H., Pato, C.N., Edwards, T.L., and Zhao, Z. Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol*, 8(7):e1002587, 2012.
- [110] Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O’Donnell, C.J., and de Bakker, P.I.W. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, 24(24):2938–9, 2008.
- [111] Jones, D.P., Park, Y., and Ziegler, T.R. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr*, 32:183–202, 2012.
- [112] Kaddurah-Daouk, R., Kristal, B.S., and Weinshilboum, R.M. Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol*, 48:653–83, 2008.
- [113] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, 40(Database issue):D109–14, 2012.
- [114] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*, 42(1):D199–205, 2014.

- [115] Karlebach, G. and Shamir, R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10):770–80, 2008.
- [116] Kerner, J. and Hoppel, C. Fatty acid import into mitochondria. *Biochim Biophys Acta*, 1486(1):1–17, 2000.
- [117] Kettunen, J., Tukiainen, T., Sarin, A.P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.P., Kangas, A.J., Soininen, P., Würtz, P., Silander, K., Dick, D.M., Rose, R.J., Savolainen, M.J., Viikari, J., Kähönen, M., Lehtimäki, T., Pietiläinen, K.H., Inouye, M., McCarthy, M.I., Jula, A., Eriksson, J., Raitakari, O.T., Salomaa, V., Kaprio, J., Järvelin, M.R., Peltonen, L., Perola, M., Freimer, N.B., Ala-Korpela, M., Palotie, A., and Ripatti, S. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*, 44(3):269–76, 2012.
- [118] Keurentjes, J.J.B., Fu, J., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. The genetics of plant metabolism. *Nat Genet*, 38(7):842–9, 2006.
- [119] Khatri, P., Sirota, M., and Butte, A.J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*, 8(2):e1002375, 2012.
- [120] Kim, J., Saidel, G.M., and Kalhan, S.C. A computational model of adipose tissue metabolism: evidence for intracellular compartmentation and differential activation of lipases. *J Theor Biol*, 251(3):523–40, 2008.
- [121] Kirk, P., Thorne, T., and Stumpf, M.P. Model selection in systems and synthetic biology. *Curr Opin Biotechnol*, 24(4):767–74, 2013.
- [122] Kitano, H. Computational systems biology. *Nature*, 420(6912):206–10, 2002.
- [123] Kitano, H. Systems biology: a brief overview. *Science*, 295(5560):1662–4, 2002.
- [124] Kitano, H. International alliances for quantitative modeling in systems biology. *Mol Syst Biol*, 1:2005.0007, 2005.
- [125] Kler, R.S., Jackson, S., Bartlett, K., Bindoff, L.A., Eaton, S., Pourfarzam, M., Frerman, F.E., Goodman, S.I., Watmough, N.J., and Turnbull, D.M. Quantitation of acyl-CoA and acylcarnitine esters accumulated during abnormal mitochondrial fatty acid oxidation. *J Biol Chem*, 266(34):22932–22938, 1991.
- [126] Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T. Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics*, 159(1):359–70, 2001.
- [127] Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., and Herwig, R. *Systems Biology: A Textbook*. Wiley-VCH, 1 edition, 2009.
- [128] Koepsell, H. and Endou, H. The SLC22 drug transporter family. *Pflugers Arch*, 447(5):666–76, 2004.
- [129] Kohn, M.C. and Garfinkel, D. Computer simulation of metabolism in palmitate-perfused rat heart. I. Palmitate Oxidation. *Ann Biomed Eng*, 11(6):533–49, 1983.

- [130] Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A., Frigge, M.L., Gylfason, A., Olason, P.I., Gudjonsson, S.A., Sverrisson, S., Stacey, S.N., Sigurgeirsson, B., Benediktsdottir, K.R., Sigurdsson, H., Jonsson, T., Benediktsson, R., Olafsson, J.H., Johannsson, O.T., Hreidarsson, A.B., Sigurdsson, G., Ferguson-Smith, A.C., Gudbjartsson, D.F., Thorsteinsdottir, U., and Stefansson, K. Parental origin of sequence variants associated with complex diseases. *Nature*, 462(7275):868–74, 2009.
- [131] Koves, T.R., Ussher, J.R., Noland, R.C., Slentz, D., Mosedale, M., Ilkayeva, O., Bain, J., Stevens, R., Dyck, J.R.B., Newgard, C.B., Lopaschuk, G.D., and Muoio, D.M. Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab*, 7(1):45–56, 2008.
- [132] Krug, S. *The human metabolome under nutritional challenges in young men and subjects carrying risk alleles for obesity and diabetes mellitus type 2*. Ph.D. thesis, Technical University Munich, 2013.
- [133] Krug, S., Kastenmüller, G., Stücker, F., Rist, M.J., Skurk, T., Sailer, M., Raffler, J., Römisch-Margl, W., Adamski, J., Prehn, C., Frank, T., Engel, K.H., Hofmann, T., Luy, B., Zimmermann, R., Moritz, F., Schmitt-Kopplin, P., Krumsiek, J., Kremer, W., Huber, F., Oeh, U., Theis, F.J., Szymczak, W., Hauner, H., Suhre, K., and Daniel, H. The dynamic range of the human metabolome revealed by challenges. *FASEB J*, 26(6):2607–19, 2012.
- [134] Krumsiek, J., Stücker, F., Kastenmüller, G., and Theis, F.J. Systems Biology Meets Metabolism. In K. Suhre (editor), *Genet. Meets Metabolomics from Exp. to Syst. Biol.*, pages 281–313. Springer New York, New York, NY, 2012.
- [135] Krumsiek, J., Suhre, K., Evans, A.M., Mitchell, M.W., Mohny, R.P., Milburn, M.V., Wägele, B., Römisch-Margl, W., Illig, T., Adamski, J., Gieger, C., Theis, F.J., and Kastenmüller, G. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet*, 8(10):e1003005, 2012.
- [136] Krumsiek, J., Suhre, K., Illig, T., Adamski, J., and Theis, F.J. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*, 5(1):21, 2011.
- [137] Krützfeldt, J. and Stoffel, M. MicroRNAs: a new class of regulatory genes affecting metabolism. *Cell Metab*, 4(1):9–12, 2006.
- [138] Langley, R.J., Tsalik, E.L., van Velkinburgh, J.C., Glickman, S.W., Rice, B.J., Wang, C., Chen, B., Carin, L., Suarez, A., Mohny, R.P., Freeman, D.H., Wang, M., You, J., Wulff, J., Thompson, J.W., Moseley, M.A., Reisinger, S., Edmonds, B.T., Grinnell, B., Nelson, D.R., Dinwiddie, D.L., Miller, N.A., Saunders, C.J., Soden, S.S., Rogers, A.J., Gazourian, L., Fredenburgh, L.E., Massaro, A.F., Baron, R.M., Choi, A.M.K., Corey, G.R., Ginsburg, G.S., Cairns, C.B., Otero, R.M., Fowler, V.G., Rivers, E.P., Woods, C.W., and Kingsmore, S.F. An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci Transl Med*, 5(195):195ra95, 2013.
- [139] Lattka, E., Illig, T., Koletzko, B., and Heinrich, J. Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism. *Curr Opin Lipidol*, 21(1):64–9, 2010.

- [140] Lauritzen, S.L. *Graphical models*. Oxford University Press, 1996.
- [141] Lawrence, N.D., Girolami, M., and Rattray, M. *Learning and inference in computational systems biology*. The MIT Press, 2010.
- [142] Lee, J.M., Gianchandani, E.P., and Papin, J.A. Flux balance analysis in the era of metabolomics. *Brief Bioinform*, 7(2):140–50, 2006.
- [143] Lee, J.M., Min Lee, J., Gianchandani, E.P., Eddy, J.A., and Papin, J.A. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol*, 4(5):e1000086, 2008.
- [144] Lehmann, R., Zhao, X., Weigert, C., Simon, P., Fehrenbach, E., Fritsche, J., Machann, J., Schick, F., Wang, J., Hoene, M., Schleicher, E.D., Häring, H.U., Xu, G., and Niess, A.M. Medium chain acylcarnitines dominate the metabolite pattern in humans under moderate intensity exercise and support lipid oxidation. *PLoS One*, 5(7):e11519, 2010.
- [145] Levy, P.A. An overview of newborn screening. *J Dev Behav Pediatr*, 31(7):622–31, 2010.
- [146] Liebermeister, W. and Klipp, E. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. *Theor Biol Med Model*, 3:42, 2006.
- [147] Ludwig, C. and Viant, M.R. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochem Anal*, 21(1):22–32, 2010.
- [148] Lynn, F.C. Meta-regulation: microRNA regulation of glucose and lipid metabolism. *Trends Endocrinol Metab*, 20(9):452–9, 2009.
- [149] Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., and Goryanin, I. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*, 3(135):135, 2007.
- [150] Maier, E.M., Liebl, B., Röschinger, W., Nennstiel-Ratzel, U., Fingerhut, R., Olgemöller, B., Busch, U., Krone, N., v Kries, R., and Roscher, A.A. Population spectrum of ACADM genotypes correlated to biochemical phenotypes in newborn screening for medium-chain acyl-CoA dehydrogenase deficiency. *Hum Mutat*, 25(5):443–52, 2005.
- [151] Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittmore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–53, 2009.
- [152] Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., and Stolovitzky, G. Wisdom of crowds for robust gene network inference. *Nat Methods*, 9(8):796–804, 2012.
- [153] Markowetz, F. and Spang, R. Inferring cellular networks—a review. *BMC Bioinformatics*, 8 Suppl 6:S5, 2007.
- [154] McClellan, J.M., Susser, E., and King, M.C. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry*, 190:194–9, 2007.

- [155] McGarry, J.D. and Foster, D.W. Regulation of hepatic fatty acid oxidation and ketone body production. *Annu Rev Biochem*, 49:395–420, 1980.
- [156] Meeker, W.Q. and Escobar, L.A. Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimation. *Am Stat*, 49(1):48–53, 1995.
- [157] Mendes, P. and Kell, D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–83, 1998.
- [158] Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet*, 11(1):31–46, 2010.
- [159] Mihalik, S.J., Goodpaster, B.H., Kelley, D.E., Chace, D.H., Vockley, J., Toledo, F.G.S., and Delany, J.P. Increased Levels of Plasma Acylcarnitines in Obesity and Type 2 Diabetes and Identification of a Marker of Glucolipotoxicity. *Obesity (Silver Spring)*, (December):1–6, 2010.
- [160] Mihalik, S.J., Michaliszyn, S.F., de Las Heras, J., Bacha, F., Lee, S., Chace, D.H., Dejesus, V.R., Vockley, J., and Arslanian, S.A. Metabolomic Profiling of Fatty Acid and Amino Acid Metabolism in Youth With Obesity and Type 2 Diabetes: Evidence for enhanced mitochondrial oxidation. *Diabetes Care*, 2012.
- [161] Mirkov, S., Myers, J.L., Ramírez, J., and Liu, W. SNPs affecting serum metabolomic traits may regulate gene transcription and lipid accumulation in the liver. *Metabolism*, 61(11):1523–7, 2012.
- [162] Mitra, K., Carvunis, A.R., Ramesh, S.K., and Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*, 14(10):719–32, 2013.
- [163] Mittelstrass, K., Ried, J.S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F.J., Meitinger, T., Kronenberg, F., Weidinger, S., Wichmann, H.E., Suhre, K., Wang-Sattler, R., Adamski, J., and Illig, T. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet*, 7(8):e1002215, 2011.
- [164] Modre-Osprian, R., Osprian, I., Tilg, B., Schreier, G., Weinberger, K.M., and Graber, A. Dynamic simulations on the mitochondrial fatty acid beta-oxidation network. *BMC Syst Biol*, 3:2, 2009.
- [165] Moles, C.G., Mendes, P., and Banga, J.R. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13(11):2467–74, 2003.
- [166] Mootha, V.K. and Hirschhorn, J.N. Inborn variation in metabolism. *Nat Genet*, 42(2):97–8, 2010.
- [167] Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., and Kanehisa, M. Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J Chem Inf Model*, 53(3):613–22, 2013.
- [168] Nada, M.A., Rhead, W.J., Sprecher, H., Schulz, H., and Roe, C.R. Evidence for intermediate channeling in mitochondrial beta-oxidation. *J Biol Chem*, 270(2):530–5, 1995.

- [169] Nagan, N., Kruckeberg, K.E., Tauscher, A.L., Bailey, K.S., Rinaldo, P., and Matern, D. The frequency of short-chain acyl-CoA dehydrogenase gene variants in the US population and correlation with the C(4)-acylcarnitine concentration in newborn blood spots. *Mol Genet Metab*, 78(4):239–46, 2003.
- [170] Newgard, C.B. and Attie, A.D. Getting biological about the genetics of diabetes. *Nat Med*, 16(4):388–91, 2010.
- [171] Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.a., and Shendure, J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261):272–6, 2009.
- [172] Nicholson, G., Rantalainen, M., Li, J.V., Maher, A.D., Malmudin, D., Ahmadi, K.R., Faber, J.H., Barrett, A., Min, J.L., Rayner, N.W., Toft, H., Krestyaninova, M., Viksna, J., Neogi, S.G., Dumas, M.E., Sarkans, U., Donnelly, P., Illig, T., Adamski, J., Suhre, K., Allen, M., Zondervan, K.T., Spector, T.D., Nicholson, J.K., Lindon, J.C., Baunsgaard, D., Holmes, E., McCarthy, M.I., and Holmes, C.C. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet*, 7(9):e1002270, 2011.
- [173] Nicholson, J.K. and Lindon, J.C. Systems biology: Metabonomics. *Nature*, 455(7216):1054–6, 2008.
- [174] Nicholson, J.K. and Wilson, I.D. Opinion: understanding ‘global’ systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov*, 2(8):668–76, 2003.
- [175] Nieuwenhoven, F.A., Vusse, G.J., and Glatz, J.F.C. Membrane-associated and cytoplasmic fatty acid-binding proteins. *Lipids*, 31(1):S223–S227, 1996.
- [176] Nobeli, I. and Thornton, J.M. A bioinformatician’s view of the metabolome. *Bioessays*, 28(5):534–45, 2006.
- [177] Noland, R.C., Koves, T.R., Seiler, S.E., Lum, H., Lust, R.M., Ilkayeva, O., Stevens, R.D., Hegardt, F.G., and Muoio, D.M. Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control. *J Biol Chem*, 284(34):22840–52, 2009.
- [178] Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*, 74(4):765–9, 2004.
- [179] Ohta, D., Kanaya, S., and Suzuki, H. Application of Fourier-transform ion cyclotron resonance mass spectrometry to metabolic profiling and metabolite identification. *Curr Opin Biotechnol*, 21(1):35–44, 2010.
- [180] Oliver, S.G., Winson, M.K., Kell, D.B., and Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol*, 16(9):373–8, 1998.
- [181] Opgen-Rhein, R. and Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol*, 1:37, 2007.
- [182] Orth, J.D., Thiele, I., and Palsson, B.O. What is flux balance analysis? *Nat Biotechnol*, 28(3):245–8, 2010.

- [183] Pahl, R. and Schäfer, H. PERMORY: an LD-exploiting permutation test algorithm for powerful genome-wide association testing. *Bioinformatics*, 26(17):2093–100, 2010.
- [184] Patti, G.J., Yanes, O., and Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*, 13(4):263–9, 2012.
- [185] Paul, D.S., Nisbet, J.P., Yang, T.P., Meacham, S., Rendon, A., Hautaviita, K., Tallila, J., White, J., Tijssen, M.R., Sivapalaratnam, S., Basart, H., Trip, M.D., Göttgens, B., Soranzo, N., Ouwehand, W.H., and Deloukas, P. Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. *PLoS Genet*, 7(6):e1002139, 2011.
- [186] Pearson, H. Meet the human metabolome. *Nature*, 446(7131):8, 2007.
- [187] Pedersen, C.B., Bross, P., Winter, V.S., Corydon, T.J., Bolund, L., Bartlett, K., Vockley, J., and Gregersen, N. Misfolding, degradation, and aggregation of variant proteins. The molecular pathogenesis of short chain acyl-CoA dehydrogenase (SCAD) deficiency. *J Biol Chem*, 278(48):47449–58, 2003.
- [188] Pe’er, I., Yelensky, R., Altshuler, D., and Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32(4):381–5, 2008.
- [189] Pellis, L., van Erk, M.J., van Ommen, B., Bakker, G.C.M., Hendriks, H.F.J., Cnubben, N.H.P., Kleemann, R., van Someren, E.P., Bobeldijk, I., Rubingh, C.M., and Wopereis, S. Plasma metabolomics and proteomics profiling after a postprandial challenge reveal subtle diet effects on human metabolic status. *Metabolomics*, 8(2):347–359, 2012.
- [190] Petersen, A.K., Krumsiek, J., Wägele, B., Theis, F.J., Wichmann, H.E., Gieger, C., and Suhre, K. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinformatics*, 13:120, 2012.
- [191] Petersen, A.K., Zeilinger, S., Kastenmüller, G., Römisch-Margl, W., Brugger, M., Peters, A., Meisinger, C., Strauch, K., Hengstenberg, C., Pagel, P., Huber, F., Mohny, R.P., Grallert, H., Illig, T., Adamski, J., Waldenberger, M., Gieger, C., and Suhre, K. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet*, 23(2):534–545, 2013.
- [192] Pey, J., Prada, J., Beasley, J.E., and Planes, F.J. Path finding methods accounting for stoichiometry in metabolic networks. *Genome Biol*, 12(5):R49, 2011.
- [193] Pohjanen, E., Thysell, E., Jonsson, P., Eklund, C., Silfver, A., Carlsson, I.B., Lundgren, K., Moritz, T., Svensson, M.B., and Antti, H. A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum. *J Proteome Res*, 6(6):2113–20, 2007.
- [194] Pougovkina, O., Te Brinke, H., Ofman, R., van Cruchten, A.G., Kulik, W., Wanders, R.J.A., Houten, S.M., and de Boer, V.C.J. Mitochondrial protein acetylation is driven by acetyl-CoA from fatty acid oxidation. *Hum Mol Genet*, 2014.
- [195] Poy, M.N., Spranger, M., and Stoffel, M. microRNAs and the regulation of glucose and lipid metabolism. *Diabetes Obes Metab*, 9 Suppl 2:67–73, 2007.

- [196] Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*, 11(7):459–63, 2010.
- [197] Price, N.D., Reed, J.L., and Palsson, B.O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–97, 2004.
- [198] Ramírez, C.M., Goedeke, L., Rotllan, N., Yoon, J.H., Cirera-Salinas, D., Mattison, J.A., Suárez, Y., de Cabo, R., Gorospe, M., and Fernández-Hernando, C. MicroRNA 33 regulates glucose metabolism. *Mol Cell Biol*, 33(15):2891–902, 2013.
- [199] Rashed, M.S. Clinical applications of tandem mass spectrometry: ten years of diagnosis and screening for inherited metabolic diseases. *J Chromatogr B Biomed Sci Appl*, 758(1):27–48, 2001.
- [200] Rasmussen, B.B., Holmbäck, U.C., Volpi, E., Morio-Liondore, B., Paddon-Jones, D., and Wolfe, R.R. Malonyl coenzyme A and the regulation of functional carnitine palmitoyltransferase-1 activity and fat oxidation in human skeletal muscle. *J Clin Invest*, 110(11):1687–93, 2002.
- [201] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–9, 2009.
- [202] Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B.D., Theis, F.J., Klingmüller, U., and Timmer, J. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One*, 8(9):e74335, 2013.
- [203] Reed, J.L. Shrinking the Metabolic Solution Space Using Experimental Datasets. *PLoS Comput Biol*, 8(8):e1002662, 2012.
- [204] Resh, M.D. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim Biophys Acta (BBA)-Molecular Cell Res*, 1451(1):1–16, 1999.
- [205] Reshef, D.N., Reshef, Y.a., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., and Sabeti, P.C. Detecting novel associations in large data sets. *Science*, 334(6062):1518–24, 2011.
- [206] Rezzi, S., Ramadan, Z., Fay, L.B., and Kochhar, S. Nutritional metabonomics: applications and perspectives. *J Proteome Res*, 6(2):513–25, 2007.
- [207] Riccardi, G., Giacco, R., and Rivellese, A.A. Dietary fat, insulin sensitivity and the metabolic syndrome. *Clin Nutr*, 23(4):447–56, 2004.
- [208] Ried, J.S., Baurecht, H., Stücker, F., Krumsiek, J., Gieger, C., Heinrich, J., Kabesch, M., Prehn, C., Peters, A., Rodriguez, E., Schulz, H., Strauch, K., Suhre, K., Wang-Sattler, R., Wichmann, H.E., Theis, F.J., Illig, T., Adamski, J., and Weidinger, S. Integrative genetic and metabolite profiling analysis suggests altered phosphatidylcholine metabolism in asthma. *Allergy*, 68(5):629–36, 2013.
- [209] Roach, J.C., Glusman, G., Smit, A.F.A., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L.B., Hood, L., and Galas, D.J. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, 328(5978):636–9, 2010.

- [210] Römisch-Margl, W., Prehn, C., Bogumil, R., Röhring, C., Suhre, K., and Adamski, J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics*, 8(1):133–142, 2011.
- [211] Rottiers, V. and Näär, A.M. MicroRNAs in metabolism and metabolic disorders. *Nat Rev Mol Cell Biol*, 13(4):239–50, 2012.
- [212] Roux, A., Lison, D., Junot, C., and Heilier, J.F. Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clin Biochem*, 44(1):119–35, 2011.
- [213] Rubio-Aliaga, I., Roos, B., Duthie, S.J., Crosley, L.K., Mayer, C., Horgan, G., Colquhoun, I.J., Gall, G., Huber, F., Kremer, W., Rychlik, M., Wopereis, S., Ommen, B., Schmidt, G., Heim, C., Bouwman, F.G., Mariman, E.C., Mulholland, F., Johnson, I.T., Polley, A.C., Elliott, R.M., and Daniel, H. Metabolomics of prolonged fasting in humans reveals new catabolic markers. *Metabolomics*, 7(3):375–387, 2010.
- [214] Rubio-Gozalbo, M.E., Bakker, J.A., Waterham, H.R., and Wanders, R.J.A. Carnitine-acylcarnitine translocase deficiency, clinical, biochemical and genetic aspects. *Mol Aspects Med*, 25(5-6):521–32, 2004.
- [215] Sahoo, S., Aurich, M.K., Jonsson, J.J., and Thiele, I. Membrane transporters in a human genome-scale metabolic knowledgebase and their implications for disease. *Front Physiol*, 5:91, 2014.
- [216] Sampson, J.N., Boca, S.M., Shu, X.O., Stolzenberg-Solomon, R.Z., Matthews, C.E., Hsing, A.W., Tan, Y.T., Ji, B.T., Chow, W.H., Cai, Q., Liu, D.K., Yang, G., Xiang, Y.B., Zheng, W., Sinha, R., Cross, A.J., and Moore, S.C. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol Biomarkers Prev*, 22(4):631–40, 2013.
- [217] Sandor, A., Cseko, J., Kispal, G., and Alkonyi, I. Surplus acylcarnitines in the plasma of starved rats derive from the liver. *J Biol Chem*, 265(36):22313–6, 1990.
- [218] Schäfer, J. and Strimmer, K. Learning large-scale graphical Gaussian models from genomic data. In *AIP Conf. Proc.*, pages 263–276. 2005.
- [219] Schellenberger, J., Park, J.O., Conrad, T.M., and Palsson, B.O. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11:213, 2010.
- [220] Schilling, C.H., Schuster, S., Palsson, B.O., and Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog*, 15(3):296–303, 1999.
- [221] Schmidl, D., Hug, S., Li, W.B., Greiter, M.B., and Theis, F.J. Bayesian model selection validates a biokinetic model for zirconium processing in humans. *BMC Syst Biol*, 6:95, 2012.
- [222] Schmidt, H. and Jirstrand, M. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics*, 22(4):514–5, 2006.

- [223] Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., and Schomburg, D. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res*, 41(Database issue):D764–72, 2013.
- [224] Schork, N.J., Murray, S.S., Frazer, K.A., and Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*, 19(3):212–9, 2009.
- [225] Schuster, S., Dandekar, T., and Fell, D.A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2):53–60, 1999.
- [226] Schwarz, G. Estimating the Dimension of a Model. *Ann Stat*, 6(2):461–464, 1978.
- [227] Segel, L.A. and Slemrod, M. The Quasi-Steady-State Assumption: A Case Study in Perturbation. *SIAM Rev*, 31(3):446–477, 1989.
- [228] Shaham, O., Wei, R., Wang, T.J., Ricciardi, C., Lewis, G.D., Vasan, R.S., Carr, S.A., Thadhani, R., Gerszten, R.E., and Mootha, V.K. Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Mol Syst Biol*, 4(214):214, 2008.
- [229] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P.O., and Bollen, K. DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model. *J Mach Learn Res*, 12:1225–1248, 2011.
- [230] Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., Walter, K., Menni, C., Chen, L., Vasquez, L., Valdes, A.M., Hyde, C.L., Wang, V., Ziemek, D., Roberts, P., Xi, L., Grundberg, E., Waldenberger, M., Richards, J.B., Mohny, R.P., Milburn, M.V., John, S.L., Trimmer, J., Theis, F.J., Overington, J.P., Suhre, K., Brosnan, M.J., Gieger, C., Kastenmüller, G., Spector, T.D., and Soranzo, N. An atlas of genetic influences on human blood metabolites. *Nat Genet*, 2014.
- [231] Smilde, A.K., Westerhuis, J.A., Hoefsloot, H.C.J., Bijlsma, S., Rubingh, C.M., Vis, D.J., Jellema, R.H., Pijl, H., Roelfsema, F., and van der Greef, J. Dynamic metabolomic data analysis: a tutorial review. *Metabolomics*, 6(1):3–17, 2010.
- [232] Steuer, R. Review: on the analysis and interpretation of correlations in metabolomic data. *Brief Bioinform*, 7(2):151–8, 2006.
- [233] Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*, 19(8):1019–1026, 2003.
- [234] Stobbe, M.D., Houten, S.M., Jansen, G.A., van Kampen, A.H.C., and Moerland, P.D. Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *BMC Syst Biol*, 5(1):165, 2011.
- [235] Stobbe, M.D., Swertz, M.a., Thiele, I., Rengaw, T., van Kampen, A.H.C., and Moerland, P.D. Consensus and conflict cards for metabolic pathway databases. *BMC Syst Biol*, 7(1):50, 2013.

- [236] Suhre, K. and Gieger, C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet*, (October), 2012.
- [237] Suhre, K., Shin, S.Y., Petersen, A.K., Mohny, R.P., Meredith, D., Wägele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C.J., de Angelis, M.H., Kastenmüller, G., Köttgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H.W., Milburn, M.V., Prehn, C., Raffler, J., Ried, J.S., Römisch-Margl, W., Samani, N.J., Small, K.S., Wichmann, H.E., Zhai, G., Illig, T., Spector, T.D., Adamski, J., Soranzo, N., and Gieger, C. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477(7362):54–60, 2011.
- [238] Suhre, K., Wallaschofski, H., Raffler, J., Friedrich, N., Haring, R., Michael, K., Wasner, C., Krebs, A., Kronenberg, F., Chang, D., Meisinger, C., Wichmann, H.E., Hoffmann, W., Völzke, H., Völker, U., Teumer, A., Biffar, R., Kocher, T., Felix, S.B., Illig, T., Kroemer, H.K., Gieger, C., Römisch-Margl, W., and Nauck, M. A genome-wide association study of metabolic traits in human urine. *Nat Genet*, 43(6):565–569, 2011.
- [239] Tabor, H.K., Risch, N.J., and Myers, R.M. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet*, 3(5):391–7, 2002.
- [240] Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., Johansen, C.T., Fouchier, S.W., Isaacs, A., Peloso, G.M., Barbalic, M., Ricketts, S.L., Bis, J.C., Aulchenko, Y.S., Thorleifsson, G., Feitosa, M.F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin Go, M., Jin Kim, Y., Lee, J.Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R., Croteau-Chonka, D.C., Lange, L.A., Smith, J.D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R.Y.L., Wright, A.F., Witteman, J.C.M., Wilson, J.F., Willemsen, G., Wichmann, H.E., Whitfield, J.B., Waterworth, D.M., Wareham, N.J., Waeber, G., Vollenweider, P., Voight, B.F., Vitart, V., Uitterlinden, A.G., Uda, M., Tuomilehto, J., Thompson, J.R., Tanaka, T., Surakka, I., Stringham, H.M., Spector, T.D., Soranzo, N., Smit, J.H., Sinisalo, J., Silander, K., Sijbrands, E.J.G., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruukonen, A., Rudan, I., Rose, L.M., Roberts, R., Rieder, M., Psaty, B.M., Pramstaller, P.P., Pichler, I., Perola, M., Penninx, B.W.J.H., Pedersen, N.L., Pattaro, C., Parker, A.N., Pare, G., Oostra, B.A., O'Donnell, C.J., Nieminen, M.S., Nickerson, D.A., Montgomery, G.W., Meitinger, T., McPherson, R., McCarthy, M.I., McArdle, W., Masson, D., Martin, N.G., Marroni, F., Mangino, M., Magnusson, P.K.E., Lucas, G., Luben, R., Loos, R.J.F., Lokki, M.L., Lettre, G., Langenberg, C., Launer, L.J., Lakatta, E.G., Laaksonen, R., Kyvik, K.O., Kronenberg, F., König, I.R., Khaw, K.T., Kaprio, J., Kaplan, L.M., Johansson, A., Jarvelin, M.R., Janssens, A.C.J.W., Ingelsson, E., Igl, W., Kees Hovingh, G., Hottenga, J.J., Hofman, A., Hicks, A.A., Hengstenberg, C., Heid, I.M., Hayward, C., Havulinna, A.S., Hastie, N.D., Harris, T.B., Haritunians, T., Hall, A.S., Gyllenstein, U., Guiducci, C., Groop, L.C., Gonzalez, E., Gieger, C., Freimer, N.B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K.G., Döring, A., Dominiczak, A.F., Demissie, S., Deloukas, P., de Geus, E.J.C., de Faire, U., Crawford, G., Collins, F.S., Chen, Y.d.I., Caulfield, M.J., Campbell, H., Burt, N.P., Bonnycastle, L.L., Boomsma, D.I., Boekholdt, S.M., Bergman, R.N., Barroso, I., Bandinelli, S., Ballantyne, C.M., Assimes, T.L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T.Y., Tai, E.S., Feranil, A.B., Kuzawa, C.W., Adair, L.S., Taylor, H.A., Borecki, I.B., Gabriel, S.B., Wilson, J.G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R.M., Mohlke, K.L., Ordovas, J.M., Munroe, P.B., Kooner, J.S.,

- Tall, A.R., Hegele, R.A., Kastelein, J.J.P., Schadt, E.E., Rotter, J.I., Boerwinkle, E., Strachan, D.P., Mooser, V., Stefansson, K., Reilly, M.P., Samani, N.J., Schunkert, H., Cupples, L.A., Sandhu, M.S., Ridker, P.M., Rader, D.J., van Duijn, C.M., Peltonen, L., Abecasis, G.R., Boehnke, M., and Kathiresan, S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–13, 2010.
- [241] Thiele, I. and Palsson, B.O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93–121, 2010.
- [242] Thiele, I. and Palsson, B.O. Reconstruction annotation jamborees: a community approach to systems biology. *Mol Syst Biol*, 6:361, 2010.
- [243] Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., Thorleifsson, S.G., Agren, R., Bölling, C., Bordel, S., Chavali, A.K., Dobson, P., Dunn, W.B., Endler, L., Hala, D., Hucka, M., Hull, D., Jameson, D., Jamshidi, N., Jonsson, J.J., Juty, N., Keating, S., Nookaew, I., Le Novère, N., Malys, N., Mazein, A., Papin, J.A., Price, N.D., Selkov, E., Sigurdsson, M.I., Simeonidis, E., Sonnenschein, N., Smallbone, K., Sorokin, A., van Beek, J.H.G.M., Weichart, D., Goryanin, I., Nielsen, J., Westerhoff, H.V., Kell, D.B., Mendes, P., and Palsson, B.O. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*, 31(5):419–25, 2013.
- [244] Trupp, M., Zhu, H., Wikoff, W.R., Baillie, R.A., Zeng, Z.B., Karp, P.D., Fiehn, O., Krauss, R.M., and Kaddurah-Daouk, R. Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment. *PLoS One*, 7(7):e38386, 2012.
- [245] Tweeddale, H., Notley-McRobb, L., and Ferenci, T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol*, 180(19):5109–5116, 1998.
- [246] Valcárcel, B., Ebbels, T.M.D., Kangas, A.J., Soininen, P., Elliot, P., Ala-Korpela, M., Järvelin, M.r., and de Iorio, M. Genome metabolome integrated network analysis to uncover connections between genetic variants and complex traits: an application to obesity. *J R Soc Interface*, 11(94):20130908, 2014.
- [247] van Bilsen, M., van der Vusse, G.J., and Reneman, R.S. Transcriptional regulation of metabolic processes: implications for cardiac metabolism. *Physiologists Arch*, 437(1):2–14, 1998.
- [248] van Buuren, S. and Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*, 45(3):1–67, 2011.
- [249] van den Oord, E.J.C.G. Controlling false discoveries in genetic studies. *Am J Med Genet B Neuropsychiatr Genet*, 147B(5):637–44, 2008.
- [250] van Eunen, K., Simons, S.M.J., Gerding, A., Bleeker, A., den Besten, G., Touw, C.M.L., Houten, S.M., Groen, B.K., Krab, K., Reijngoud, D.J., and Bakker, B.M. Biochemical Competition Makes Fatty-Acid β -Oxidation Vulnerable to Substrate Overload. *PLoS Comput Biol*, 9(8):e1003186, 2013.

- [251] van Maldegem, B.T., Duran, M., Wanders, R.J.A., Niezen-Koning, K.E., Hogeveen, M., Ijlst, L., Waterham, H.R., and Wijburg, F.A. Clinical, biochemical, and genetic heterogeneity in short-chain acyl-coenzyme A dehydrogenase deficiency. *JAMA*, 296(8):943–52, 2006.
- [252] van Maldegem, B.T., Wanders, R.J.A., and Wijburg, F.A. Clinical aspects of short-chain acyl-CoA dehydrogenase deficiency. *J Inherit Metab Dis*, 33(5):507–11, 2010.
- [253] Villaverde, A.F. and Banga, J.R. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J R Soc Interface*, 11(91):20130505, 2014.
- [254] Vinayavekhin, N., Homan, E.A., and Saghatelian, A. Exploring disease through metabolomics. *ACS Chem Biol*, 5(1):91–103, 2010.
- [255] Violante, S., Ijlst, L., Ruiter, J., Koster, J., van Lenthe, H., Duran, M., de Almeida, I.T., Wanders, R.J.A., Houten, S.M., and Ventura, F.V. Substrate specificity of human carnitine acetyltransferase: Implications for fatty acid and branched-chain amino acid metabolism. *Biochim Biophys Acta*, 1832(6):773–9, 2013.
- [256] Violante, S., Ijlst, L., Te Brinke, H., Tavares de Almeida, I., Wanders, R.J.A., Ventura, F.V., and Houten, S.M. Carnitine palmitoyltransferase 2 and carnitine/acylcarnitine translocase are involved in the mitochondrial synthesis and export of acylcarnitines. *FASEB J*, 27(5):2039–44, 2013.
- [257] Violante, S., Ijlst, L., van Lenthe, H., de Almeida, I.T., Wanders, R.J., and Ventura, F.V. Carnitine palmitoyltransferase 2: New insights on the substrate specificity and implications for acylcarnitine profiling. *Biochim Biophys Acta*, 1802(9):728–32, 2010.
- [258] Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. Five years of GWAS discovery. *Am J Hum Genet*, 90(1):7–24, 2012.
- [259] Vuong, Q.H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307, 1989.
- [260] Wanders, R.J., Vreken, P., den Boer, M.E., Wijburg, F.A., van Gennip, A.H., and Ijlst, L. Disorders of mitochondrial fatty acyl-CoA beta-oxidation. *J Inherit Metab Dis*, 22(4):442–487, 1999.
- [261] Weckwerth, W. and Fiehn, O. Can we discover novel pathways using metabolomic analysis? *Curr Opin Biotechnol*, 13(2):156–60, 2002.
- [262] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–78, 2007.
- [263] Wentzell, A.M., Rowe, H.C., Hansen, B.G., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet*, 3(9):1687–701, 2007.
- [264] Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A*, 106(10):3698–703, 2009.

- [265] Wilcken, B., Wiley, V., Hammond, J., and Carpenter, K. Screening newborns for inborn errors of metabolism by tandem mass spectrometry. *N Engl J Med*, 348(23):2304–12, 2003.
- [266] Wild, C.P. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*, 14(8):1847–50, 2005.
- [267] Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorn Dahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., and Scalbert, A. HMDB 3.0–The Human Metabolome Database in 2013. *Nucleic Acids Res*, 41(Database issue):D801–7, 2013.
- [268] Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I., and Müller, W. SABIO-RK–database for biochemical reaction kinetics. *Nucleic Acids Res*, 40(Database issue):D790–6, 2012.
- [269] Wolf, M., Chen, S., Zhao, X., Scheler, M., Irmeler, M., Staiger, H., Beckers, J., Hrabé de Angelis, M., Fritsche, A., Häring, H.U., Schleicher, E.D., Xu, G., Lehmann, R., and Weigert, C. Production and Release of Acylcarnitines by Primary Myotubes Reflect the Differences in Fasting Fat Oxidation of the Donors. *J Clin Endocrinol Metab*, (C):1–6, 2013.
- [270] Wolkenhauer, O. *Systems biology*, volume 45. Portland Pr, 45 edition, 2008.
- [271] Wong, P., Althammer, S., Hildebrand, A., Kirschner, A., Pagel, P., Geissler, B., Smialowski, P., Blöchl, F., Oesterheld, M., Schmidt, T., Strack, N., Theis, F.J., Ruepp, A., and Frishman, D. An evolutionary and structural characterization of mammalian protein complex organization. *BMC Genomics*, 9:629, 2008.
- [272] Wopereis, S., Rubingh, C.M., van Erk, M.J., Verheij, E.R., van Vliet, T., Cnubben, N.H.P., Smilde, A.K., van der Greef, J., van Ommen, B., and Hendriks, H.F.J. Metabolic Profiling of the Response to an Oral Glucose Tolerance Test Detects Subtle Metabolic Changes. *PLoS One*, 4(2):e4525, 2009.
- [273] Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., and Visscher, P.M. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*, 14(7):507–15, 2013.
- [274] Wu, F., Yang, F., Vinnakota, K.C., and Beard, D.A. Computer modeling of mitochondrial tricarboxylic acid cycle, oxidative phosphorylation, metabolite transport, and electrophysiology. *J Biol Chem*, 282(34):24525–37, 2007.
- [275] Yan, B., A, J., Wang, G., Lu, H., Huang, X., Liu, Y., Zha, W., Hao, H., Zhang, Y., Liu, L., Gu, S., Huang, Q., Zheng, Y., and Sun, J. Metabolomic investigation into variation of endogenous metabolites in professional athletes subject to strength-endurance training. *J Appl Physiol*, 106(2):531–8, 2009.
- [276] Yen, J. Finding the K Shortest Loopless Paths in a Network. *Manage Sci*, 17(11):712–716, 1971.

- [277] Yoganathan, P., Karunakaran, S., Ho, M.M., and Clee, S.M. Nutritional regulation of genome-wide association obesity genes in a tissue-dependent manner. *Nutr Metab (Lond)*, 9(1):65, 2012.
- [278] Yousri, N.a., Kastenmüller, G., Gieger, C., Shin, S.Y., Erte, I., Menni, C., Peters, A., Meisinger, C., Mohny, R.P., Illig, T., Adamski, J., Soranzo, N., Spector, T.D., and Suhre, K. Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics*, 2014.
- [279] Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., Römisch-Margl, W., Lattka, E., Gieger, C., Soranzo, N., Heinrich, J., Standl, M., Thiering, E., Mittelstraß, K., Wichmann, H.E., Peters, A., Suhre, K., Li, Y., Adamski, J., Spector, T.D., Illig, T., and Wang-Sattler, R. Human serum metabolic profiles are age dependent. *Aging Cell*, 11(6):960–7, 2012.
- [280] Yuan, Y., Li, C.T., and Windram, O. Directed partial correlation: inferring large-scale gene regulatory network through induced topology disruptions. *PLoS One*, 6(4):e16835, 2011.
- [281] Zamule, S.M., Strom, S.C., and Omiecinski, C.J. Preservation of hepatic phenotype in lentiviral-transduced primary human hepatocytes. *Chem Biol Interact*, 173(3):179–86, 2008.
- [282] Zeisel, S.H. Nutrigenomics and metabolomics will change clinical nutrition and public health practice: insights from studies on dietary requirements for choline. *Am J Clin Nutr*, 86(3):542–8, 2007.
- [283] Zenobi, R. Single-cell metabolomics: analytical and biological perspectives. *Science*, 342(6163):1243259, 2013.
- [284] Zhao, X., Fritsche, J., Wang, J., Chen, J., Rittig, K., Schmitt-Kopplin, P., Fritsche, A., Häring, H.U., Schleicher, E.D., Xu, G., and Lehmann, R. Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics*, 6(3):362–374, 2010.
- [285] Zhu, J., Sova, P., Xu, Q., Dombek, K.M., Xu, E.Y., Vu, H., Tu, Z., Brem, R.B., Bumgarner, R.E., and Schadt, E.E. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol*, 10(4):e1001301, 2012.