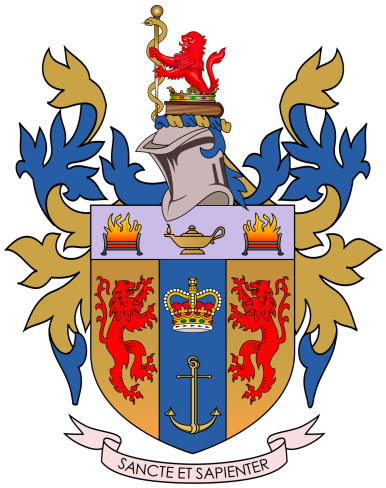
Computational models to study the interplay between diet and hepatic metabolism

Clasen, Frederick

Awarding institution:
King's College London

COMPUTATIONAL MODELS TO STUDY THE INTERPLAY BETWEEN DIET AND HEPATIC METABOLISM

FREDERICK JOHANNES CLASEN



Submitted for the Degree Doctor of Philosophy

King's College London and The Francis Crick Institute

Frederick Johannes Clasen: *Computational models to study the interplay between diet and hepatic metabolism,* Submitted for the Degree Doctor of Philosophy,

ABSTRACT

A significant increase in the incidence of liver cancer is projected over the next few decades with limited therapeutic options for patients. Like many other cancer types, liver cancer can be promoted by "Western-style" diets (WD) that are high in fat and processed sugars, such as fructose. Studies indicate that chronic WD consumption can lead to systemic dysregulation of insulin signalling and lipid metabolism that increase the risk for development of liver diseases such as non-alcoholic fatty liver disease. Liver tissue damage and the ensuing inflammation has been shown to promote mutations that can lead to oncogenic transformation in hepatocytes that lead to the formation of malignancies. A particular challenge in understanding the dietary impact on liver and tumour metabolism is how dietary nutrients fuel metabolic pathways and how this synergises with gene expression changes caused by cell-autonomous and systemic signals because of diet. In order to address this it is necessary survey metabolic fluxes on a global scale and genome scale metabolic models (GSMMs) used in systems biology provide rigorous mathematical frameworks for this purpose.

To study how diet impacts liver and liver tumour metabolism, a new mouse GSMM, Mouse Metabolic Reaction Network (MMRN), was reconstructed using orthology between mouse and human. MMRN was constrained with gene expression data from the tissues of a carcinogen-induced mouse model of liver cancer. By using *in silico* constraint-based modelling approaches, the effect of gene expression and dietary composition, alone and in combination, on liver and tumour metabolism were investigated. The WD was shown to lead to distinct metabolic phenotypes irrespective of gene expression, but also, that gene expression in the tumour drive flux through specific metabolic pathways compared to liver tissue. These observations were validated experimentally. A novel computational approach, termed Systematic Diet Composition Swap (SyDiCoS), was developed to investigate the impact of WD by swapping out individual nutrients to its corresponding composition in a CD and investigate

its impact on metabolic flux. This approach made it possible to deconvolute the effect of specific dietary nutrients on metabolic flux in a given genetic background. SyDiCoS showed that carbohydrates and lipids in WD increases glycerol and succinate production, respectively, and that both these nutrient classes are required to increase biomass production in the tumour.

Using dietary content as input for tissue GSMMs is the state of the art in metabolic modelling and has been shown to result in accurate model predictions. However, dietary nutrients are first metabolised by microorganisms in the gut, and the diet itself has a significant impact in the composition of the gut microbiome that, in turn, effects its functional metabolic capability. To this end, metagenomics data were used to reconstruct gut microbiome community GSMMs to study how the gut microbiome composition impacts gut metabolism under different dietary regiments. Using this approach together with SyDiCoS and systematic removal of specific microorganisms from the community, the interplay between diet and gut microbiome composition on metabolism was investigated. Finally, community GSMMs and tissue GSMMs were integrated into a single multi-tissue GSMM to establish a diet-microbiome-liver metabolic axis. Several novel and previously identified metabolic interactions between the microbiome and the liver as well as tumour-host metabolic interactions were identified. The diet was shown to influence similar pathways as in single-tissue GSMMs, but its integration into a multi-tissue GSMM effected exchanges of metabolites involved in these pathways. This indicated that the integration of tissues into a multi-tissue framework impacts predictions made in single-tissue GSMM predictions.

ACKNOWLEDGMENTS

The work presented in this thesis is the culmination of hundreds of hours of work and discussion between myself and my two supervisors. I will forever be indebted to you for your contribution to my scientific and personal development.

Saeed Shoaie Thank you for your constant support and encouragement as a supervisor but also as a friend. You are a true gentleman from whom I have learned a lot and I look forward to our future endeavors.

Dimitrios Anastasiou Thank you for your dedication to my development as a scientist. Your door was always open to discuss good and bad results and you taught me that both are equally important.

Thank you to the members of my **thesis committee**; Alan Hodgkinson, Jean-Paul Vincent and James Briscoe for valuable discussions that shaped my project.

Dozens of present and past **lab members** from the Shoaie and Anastasiou labs contributed to this work through discussions during lab meetings. I thank you all for your valuable contributions. I would also like to thank collaborators at the Crick, especially in the bioinformatics team, and King's that contributed directly and indirectly to this work.

The incessant support from my family was the most important piece of the puzzle to finish this work and I thank you all personally below.

I would firstly like to thank my parents. None of my achievements would have been possible without your hard work that gave me incredible opportunities. **Ferdie**, your strive to seek for life's truths is an inspiration. You are a personal and professional role model. **Lizelle**, your selfless kindness contributes more to our success than you realize. Thank you for always going out of your way to make life easier for us. The best siblings in the world, **Corneil** and **Estine**, thank you for always being curious

and interested in my work. My parents in-law, **Gerrie** and **Mercia**, you always cheered me on from afar and thank you for your ceaseless optimism in me.

The most important acknowledgement is the person that has been there since day one, by my side, no matter what: **Melissa**. Without your unconditional support, patience and love, I would never have been able to do this. Thank you for motivating and encouraging me through long days of work for the past four years. You, and Cooper, made every difficult day better and I cannot thank you enough. This thesis is as much yours as it is mine.

PUBLICATIONS

Some ideas and figures in this thesis appear in the following publications and conference outputs:

1. Clasen F, Nunes PM, Bidkhori G, Bah N, Boeing S, Shoaie S and Anastasiou D. Systematic diet composition swap in a new mouse genome-scale model reveals determinants of obesogenic diet metabolism in liver cancer. Under review: iScience.

2. Clasen F, Nunes PM, Bidkhori G, Bah N, Boeing S, Shoaie S and Anastasiou D. Systematic diet composition swap in a new mouse genome-scale model reveals determinants of obesogenic diet metabolism in liver cancer. Oral presentation at Crick Metabolism and Physiology Forum, 19 April 2022.

3. Clasen F, Nunes PM, Bidkhori G, Bah N, Boeing S, Shoaie S and Anastasiou D. Genome scale metabolic modeling reveals direct and indirect determinants of mouse liver tumour metabolism. Oral presentation at Cell Symposia: Metabolites in Signaling and Disease, 4 April 2022.

4. Clasen F, Nunes PM, Shoaie S and Anastasiou D. Unraveling metabolic reprogramming in hepatocellular carcinoma using systems biology. Poster presentation at Crick-EMBL PhD Symposia, 8 November 2019.

During my PhD I also contributed to the following publications:

1. Pellon *et al.*, Fungal infection drives metabolic reprogramming in epithelial cells via aerobic glycolysis and an alternative TCA cycle shunt. Preprint available from Research Square, doi: 10.21203/rs.3.rs-2159406/v1.

2. Ghaffari *et al.*, Detecting highly regulated signaling metabolites inside human gut ecosystem. To be submitted 2023.

3. Bidkhori *et al.*, The Reactobiome Unravels a New Paradigm in Human Gut Microbiome Metabolism . Under review: Nature Communications.

I was also involved in processing transcriptomics, metabolomics and lipodomics datasets for the Anastasiou lab that will form part of future publications. I assisted with teaching the MSc Microbiome in Health and Disease course at King's College London from 2020 to 2022 and supervised a research project dissertation in 2021 for the same course.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACC acetyl-coenzyme A carboxylase

AGORA Assembly of Gut micrOorganisms through Reconstruction and Analysis

AMP adenosine monophosphate

ATGL adipose triglyceride lipase

BA bile acid

BiGG biochemical, genetic and genomic

CBM constraint-based modelling

CD control diet

CPT carnitine palmitoyltransferase

ChREBP carbohydrate-responsive element-binding protein

DEN diethylnitrosamine

DGE differential gene expression

DNL *De novo* lipogenesis

EHMN Edinburgh Human Metabolic Network

ER endoplasmic reticulum

FA fatty acid

FAS fatty acid synthase

FBA flux balance analysis

FPKM fragments per kilobase million

FXR farnesoid X receptor

G_3P glycerol-3-phosphate

GK glucokinase

GLS Glutaminase

GLUT glucose transporter

GO Gene Ontology

GR-rules gene-reaction rules

GRCm38 Genome Reference Consortium Mouse Build 38

GSMM Genome scale metabolic model

HCC hepatocellular carcinoma

HK hexokinase

HMR Human Metabolic Reaction

HSL hormone-sensitive lipase

IM intermediate model

IMP inosine monophosphate

INIT integrative network inference for tissues

IR insulin resistance

KEGG Kyoto Encyclopedia of Genes and Genomes

KHK ketohexokinase

LP linear programming

MGS metagenomic species

MMR Mouse Metabolic Reaction

MMRN mouse metabolic reaction network

NAFLD non-alcoholic fatty liver disease

NASH nonalcoholic steatohepatitis

NEFA nonesterified FA

PCA Principal component analysis

PEP phosphoenolpyruvate

PHGDH phosphoglycerate dehydrogenase

PK pyruvate kinase

PPAR peroxisome proliferator–activated receptor

PSAT phosphoserine aminotransferase

RER respiratory exchange ratio

S-matrix stoichiometric matrix

SCFA short chain fatty acid

SREBP1 sterol regulatory element- binding protein 1

SyDiCoS Systematic Diet Composition Swap

T2DM type 2 diabetes mellitus

TAG Triacylglyceride

TME tumour microenvironment

TPM transcripts per kilobase million

VLDL very low density lipoproteins

WAT white adipose tissue

WD "Western-style" diet

WHO World Health Organisation

comGSMM community GSMM

csGSMM context-specific GSMM

hGSMM human GSMM

mGSMM mouse GSMM

mtGSMM multi-tissue GSMM

tINIT task-driven integrative network inference for tissues

1 INTRODUCTION

"The whole is greater than the sum of its parts" - Aristotle

1.1 METABOLISM OF LIVER DISEASE AND LIVER CANCER

Metabolism is the ensemble of chemical reactions that allow cells to sustain their survival and perform key functions [1]. Two of the primary objectives of metabolism are to convert dietary nutrients into energy for essential cellular processes and to use these nutrients as building blocks for proteins, lipids, carbohydrates and nucleic acids. Metabolic reactions are organised into series of successive steps, or metabolic pathways, that converts a chemical substrate into a product catalysed by a specific enzyme. Enzymes are essential regulators of the rate of metabolism, or metabolic flux, through metabolic pathways and their expression is regulated by mechanisms that operate at a cellular, tissue and organismal level.

The liver is one of the most important metabolic organs in human physiology. It performs several key metabolic functions such as maintenance of glucose homeostasis, lipid metabolism, energy biosynthesis and detoxification [2–4]. Several of these key functions are impaired in liver disease. Over recent years, we have seen an increase in the incidence of non-alcoholic fatty liver disease (NAFLD), a spectrum of progressive liver diseases that are primarily characterised by excessive retention of lipids in liver tissue [5–7] (Figure 1.1). The global prevalence of NAFLD is estimated to be 25.24% with a consistent rise in the past decade from 15% in 2005 to 25% in 2010 [5, 8]. More severe cases of increased hepatic lipid accumulation results in ensuing inflammation and hepatocyte injury, known as nonalcoholic steatohepatitis (NASH), which has seen similar trends in prevalence almost doubling from 33% to 59.1% in the same timeframe. Ultimately, severe tissue damage can lead to fibrosis and cirrhosis,

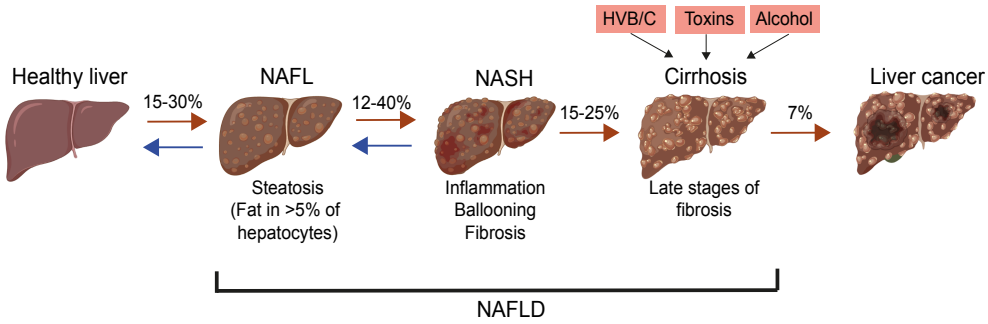


Figure 1.1: **Progression of liver disease to liver cancer**. The dysregulation of several metabolic pathways lead to the increase of fat in hepatocytes to cause steatosis. Increased tissue damage can cause persistent inflammation leading to fibrosis and cirrhosis. A large percentage of patients with cirrhosis develop liver cancer. Figure adapted from [11].

which increase the risk to develop liver cancer [7, 9]. Although there are also other environmental factors that can cause cirrhosis, NAFLD is increasingly becoming the primary cause of liver dysfunction worldwide. Liver cancer is currently one of the few cancer types that show increased incidence rates which is, in part, because of limited therapeutic options [10]. There is therefore a great need for the development of novel avenues to diagnose and treat NAFLD and liver cancer.

The exact molecular mechanisms of NAFLD pathology remain largely unknown. The establishment of steatosis as well as the progression to more advanced stages of liver disease are impacted by several intrinsic and extrinsic factors effecting liver metabolism, including diet, local and systemic metabolic reprogramming and the microbiome. Moreover, once liver tumours are established, metabolic reprogramming is also promoted by cell-autonomous factors and metabolic interactions of tumour cells with the host liver and other remote tissues.

1.1.1 *Key metabolic pathways contributing to liver disease pathogenesis*

Increased consumption of "Western-style" diet (WD) has led to a global epidemic of obesity [12]. Although different descriptions of a WD exist, it is generally characterised by a high fat and processed sugar, in particular fructose, content [13, 14]. In 2016,

the World Health Organisation (WHO) estimated that more than 1.9 billion people are overweight or obese and current predictions show that this number will rise significantly in coming years. Although obesity is a risk factor for the development of several diseases, the alarming increase of NAFLD is of particular concern [15–17]. NAFLD has a strong metabolic phenotype and is often referred to as the hepatic manifestation of metabolic disease [15]. As much as 70% of patients with type 2 diabetes mellitus (T2DM) have been shown to also have NAFLD and it is now generally accepted that the initiating events of NAFLD depend on the development of insulin resistance (IR) and obesity [18, 19]. In obese patients with T2DM, NAFLD causes increased hyperinsulinaemia, dyslipideamia and IR in hepatic and adipose tissues [20]. Ultimately, this results in an imbalance in processes that synthesise and break down lipids that increase lipid build-up and result in steatosis. Although it is difficult to holistically investigate the coordinated activity of the repertoire of metabolic pathways that synergise to promote NAFLD, there has been a significant advancement in our understanding of metabolic pathways contributing to NAFLD pathogenesis.

Steatosis is the primary clinical feature of NAFLD. Therefore, the principal metabolic features involve pathways in lipid metabolism. Carbohydrate metabolism provides several substrates for lipid metabolism and is dysregulated in IR and NAFLD. The metabolic pathways for lipid and carbohydrate metabolism as well as its regulation are discussed below.

Lipid metabolism

A complex interplay between hormones, signalling pathways and transcription factors regulates lipid handling in the liver. Triacylglyceride (TAG)s are the major class of lipids accounting for lipid accumulation in steatosis. The major pathway for TAG synthesis is through the sequential acylation of glycerol-3-phosphate (G_3P) with three fatty acid (FA)s. In the liver, FAs have three main sources; *De novo* lipogenesis (DNL), nonesterified FA (NEFA)s derived from local and peripheral lipolysis in tissues such as the white adipose tissue (WAT), and from dietary fats [7]. The former two sources are directly impacted by IR and obesity. An influx of lipids in hepatic tissue is balanced by the

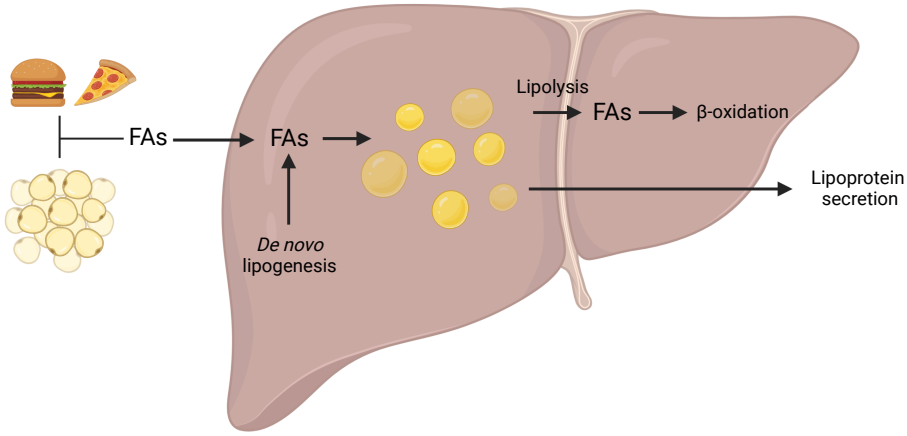


Figure 1.2: **Imbalance between lipid anabolic and catabolic pathways that can lead to steatosis.** The diet and lipolysis in adipose tissue contribute to increased circulating FAs that can be taken up by liver tissue. Together with local *de novo* lipogenesis, this contributes to an increase in fat in hepatocytes. Concomitantly, dysfunction in metabolic pathways responsible for the breakdown of lipids, lipolysis and β-oxidation, as well as secretion of lipids through lipoproteins can also increase hepatocyte lipid accumulation. Figure adapted from [21].

catabolism of lipids through β-oxidation and the secretion of lipoproteins, in particular very low density lipoproteins (VLDL), to the blood. An imbalance between these pathways results in increased lipid accumulation and the impairment of most of these pathways have been implicated in NAFLD pathogenesis (Figure 1.2) [21].

In normal physiological conditions, DNL contributes ~5% to overall hepatic lipid synthesis, but has been shown to be significantly increased in NAFLD and can be as high as 25% [22, 23]. This has been attributed to sterol regulatory element- binding protein 1 (SREBP1) and carbohydrate-responsive element-binding protein (ChREBP), the two main transcription factors that regulate DNL, that are both controlled by insulin and glucose [24–26]. These transcription factors enable the expression of key lipogenic genes such as acetyl-coenzyme A carboxylase (ACC) and fatty acid synthase (FAS). Hyperinsulinaemia and hyperglycaemia can thus synergistically stimulate DNL and promote hepatic lipid accumulation. Increased levels of several FA-binding proteins

such as FABP4, FABP5 and CD36 have been associated with the increased uptake of FAs into hepatocytes in NAFLD [27–29]. This increased flux of FAs to the liver place hepatocytes under metabolic stress that promotes lipotoxicity and endoplasmic reticulum (ER) stress.

At the same time, several mechanisms for dysfunctional clearance or breakdown of FAs in liver tissue have been proposed to contribute to NAFLD. FAs are primarily oxidised in the mitochondria in liver tissue. The carnitine shuttle is responsible for the transport of FAs from the cytosol to the mitochondria where they can be further metabolised through β-oxidation. One of the key enzymes in the shuttle, carnitine palmitoyltransferase (CPT)1, is inhibited by malonoyl-CoA, an intermediate in DNL [30]. Recently, the downregulation of CPT2, that has previously been implicated in NAFLD-related hepatocellular carcinoma (HCC) [31], was shown to be due to cell cycle regulators E2F1 and E2F2 that repress its expression [32]. Finally, the hepatic overexpression of hormone-sensitive lipase (HSL) and adipose triglyceride lipase (ATGL), two enzymes involved in lipolysis, have been shown to promotes FA oxidation and relieved steatosis in an obesogenic mouse model suggesting that impaired local lipolysis potentially also contributes to NAFLD [33].

Carbohydrate metabolism

The metabolic and signalling pathways involved in glucose metabolism have been extensively studied in normal physiology as well as in diseases such as diabetes and cancer that has characteristic dysregulated glucose metabolic profiles. The liver is a vital organ for glucose metabolism both after feeding and during fasting. In fed conditions, insulin stimulates hepatic glucose uptake through increased expression of glucose transporter (GLUT)s of which GLUT2 is the primary glucose transporter in the liver [34] (Figure 1.3). Glucose can then be used for the synthesis of glycogen or can be catabolised through glycolysis to provide carbons for storage molecules such as TAGs. The most important of these for lipid synthesis is the production of G3P through a process termed glyceroneogenesis. Transcription factors important in lipid metabolism, such as SREBP1, also control glycolysis through, for example, regulation of glucokinase (GK) expression,

the enzyme responsible for catalysing the first step in glycolysis converting glucose to glucose 6-phosphate [35, 36]. The liver is also the primary producer of glucose during fasting conditions through the metabolic processes glycogenolysis and gluconeogenesis. A primary metabolic characteristic of IR in the liver is the impairment of insulin-mediated suppression of gluconeogenesis resulting in increased glucose production and hyperglyceamia [37].

Apart from the important role of glucose in liver metabolism, several studies have implicated fructose to be a major contributor to the development of obesity and NAFLD [38, 39] (Figure 1.3). Early studies in animal models using labeled fructose and glucose showed a two to three-fold increased labelling of plasma and liver TAGs from fructose compared to glucose [40]. Interestingly, however, the overall amount of fructose that directly contribute to TAG synthesis is small, suggesting that fructose contribute to hepatic fat accumulation through different direct and indirect mechanisms [41]. In NAFLD patients with high fructose consumption, the expression levels of ketohexokinase (KHK), the first enzyme to metabolise fructose in the liver, has been shown to be increased [38]. The increased rate of fructose metabolism has also been shown to stimulate DNL [42, 43] and block β-oxidation [44] consequently increasing lipid accumulation. It has also been shown that the rapid phosphorylation of fructose to fructose 1-phosphate causes a drop in intracellular ATP activating the enzyme adenosine monophosphate (AMP) deaminase converting AMP to inosine monophosphate (IMP) that ultimately increase production of uric acid that promotes oxidative stress and inflammation [45, 46].

1.1.2 *Liver disease and gut microbiota*

There is a growing body of evidence that suggests that the gut microbiome, a term used to describe the community of microorganisms residing in the gut lumen, contributes to the development of diseases such as obesity, metabolic syndrome, and liver disease [47, 48]. The advancement in next generation sequencing has allowed quantification of the species composition of the gut. As such, dysbiosis, a term used to describe the imbalance of microflora compared to a healthy state, has been

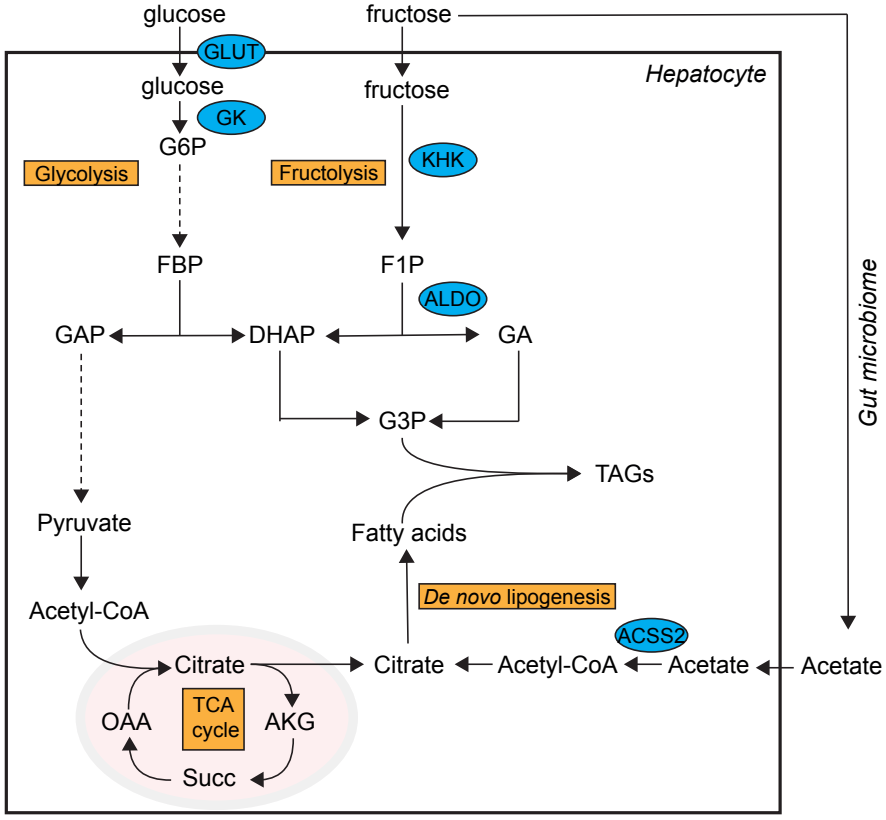


Figure 1.3: **Carbohydrate pathways supporting lipid accumulation in the liver.** The two primary carbohydrates metabolised in the liver are glucose and fructose, metabolised through glycolysis and fructolysis, respectively. This creates the necessary metabolic precursors for lipid synthesis. Fructose can also be metabolized by gut microbiota to acetate that increases hepatic acetyl-CoA production. Key enzymes involved in these pathways are depicted in blue circles. G6P: glucose 6-phosphate, FBP: fructose 1,6-bisphosphate, GAP: glyceraldehyde 3-phosphate, DHAP: dihydroxyacetone phosphate, G3P: glycerol 3-phosphate, GA: glyceraldehyde, TAGs: triacylglycerides, OAA: oxaloacetate, AKG: α-keto glutarate, Succ: succinate.

shown to contribute to functional changes of the gut in disease. There is a particular interest in studying the gut microbiome in obesity-related diseases because it metabolises dietary nutrients and produces numerous metabolites that impact host metabolism [49]. Moreover, the diet itself also impacts the composition of the gut microbiome.

The liver and the intestine are tightly linked through blood circulation via the portal vein, termed the gut-liver axis. The liver is consequently the primary recipient of gut-derived products, including several metabolites produced by the gut microbiome [48]. The most well-defined metabolic system of the gut-liver axis is that of bile acid (BA) metabolism. This involves, in short, the synthesis of primary BAs in the liver which are then further metabolised by microbes in the gut to secondary BAs that can be circulated back to the liver or secreted in faeces [50]. Several of these BAs act as ligands for important signalling molecules, such as nuclear bile acid farnesoid X receptor (FXR) and membrane G protein-coupled bile acid receptor-1 (Gpbar-1, aka TGR5), which control many carbohydrate and lipid metabolic processes in the liver. As a result, several studies have associated dysregulation of BA metabolism with liver disease [51–53].

Microbes in the gut ferment indigestible complex carbohydrates to produce short chain fatty acid (SCFA)s such as acetate, propionate, and butyrate. SCFAs produced in the gut have different downstream signalling and metabolic functions [47, 48] and current evidence for the effects of SCFAs are contradictory with positive and negative effects on liver disease pathogenesis. For example, SCFAs have the ability to alleviate some lipogenic features induced by a high-fat diet through peroxisome proliferator–activated receptor (PPAR) signalling mechanisms [54] but tissues such as the liver can also use these as substrates for synthesis of other lipid species [55]. Butyrate supplementation in mice has been shown to prevent the development of insulin resistance in obese mice suggesting a beneficial advantage of butyrate production by gut microbes [56].

More recent studies have also linked high intake of fructose in WD with changes in the gut microbiome. In a study using isotopic tracers of fructose *in vivo* the authors showed that the majority of fructose is metabolised in the small intestine rather than the liver as previously thought [57]. Interestingly, in the

same study they also showed that in high doses, fructose spills over to the liver and gut microbiota where it can also be metabolised. In a more recent study, increased consumption of fructose has been shown to increase gut permeability, known as a 'leaky gut', that can lead to endotoxins entering the blood circulation that is an important trigger in fatty liver formation [58, 59]. In some instances, microbes from the gut can also translocate to distant tissues due to gut permeability. Finally, it was also shown that dietary fructose can be metabolised by gut microbiota to acetate that, in turn, increase lipogenesis in the liver through increased acetyl-CoA production via ACSS2 [60].

1.1.3 *Liver cancer metabolism*

In the early 1900's Otto Warburg observed that cancer cells consume copious amounts of glucose to produce lactate even in the presence of oxygen, a phenomenon termed aerobic glycolysis. It was, however, only recently that we began to realise the true extent of metabolic reprogramming in cancer, and an array of metabolic pathways have now been associated with different types of cancer [61]. As with many other cancer types, the primary metabolic pathways studied in liver cancer are those involved in glucose, lipid and glutamine metabolism. There has also been an increasing interest how other amino acids, in particular serine, are metabolised in liver cancer [62].

Several tumour types show increased uptake of glucose relative to the surrounding tissue. This feature has been exploited in cancer diagnostics by using ^{18}F-2-deoxyglucose, a radioactive glucose analogue that cannot be fully metabolised, for positron emission tomography to detect cancers. The major route for glucose catabolism is through glycolysis. A key enzyme in glycolysis, hexokinase (HK)2 is highly expressed in liver cancer and its silencing has recently been shown to inhibit tumorigenesis highlighting the reliance of liver tumorigenesis on glycolysis [63]. Several other anabolic pathways use glycolytic intermediates for the synthesis of macromolecules for growth, such as nucleotides, proteins and lipids. In this regard, a widely studied enzyme in cancer cells is pyruvate kinase (PK) that catalyses the final step of glycolysis converting phosphoenolpyruvate (PEP) to pyruvate. Many cancer cells have been shown to express a spe-

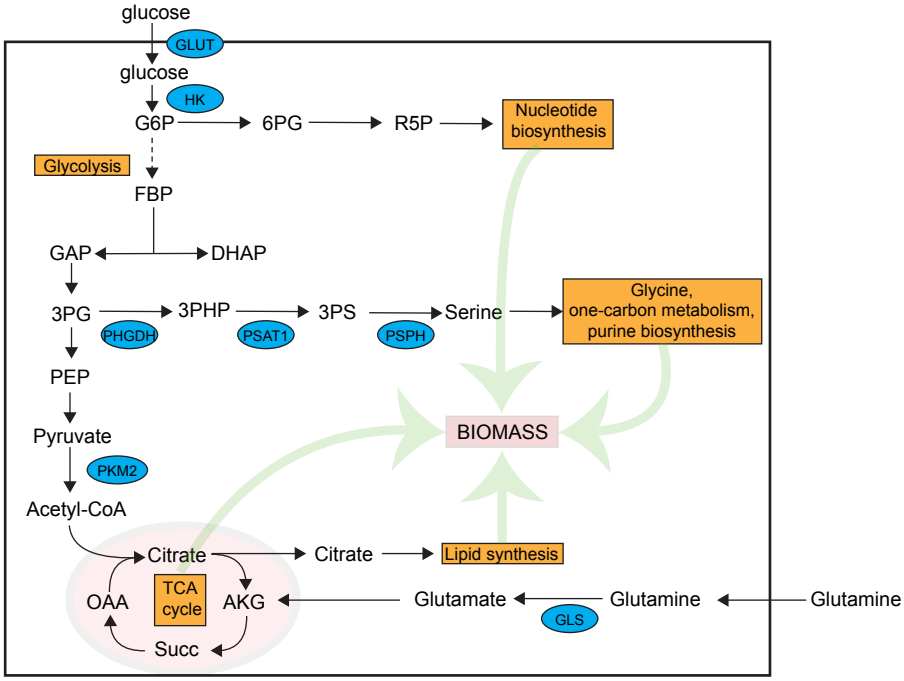


Figure 1.4: **Simplified overview of metabolic pathways implicated in cancer**. Metabolic intermediates of several metabolic pathways that are implicated in cancer contribute to biomass production.

cific splice variant, PKM2, that can switch to a dimeric state with lower activity that allow diversion of glycolytic intermediates to other pathways. In liver cancer, PKM2 is highly expressed and related to overall decreased survival of liver cancer patients [64, 65].

In addition to glucose, glutamine is a major metabolic substrate for cancer cell proliferation through TCA cycle anaplerosis [66]. Glutaminase (GLS), the first enzyme in glutaminolysis, is upregulated in liver cancer but its inhibition does not prevent tumour formation and it has recently been shown that liver tumours adapt to GLS inhibition through upregulation of amidotransferases and increased glycolysis [67]. This highlights how rapid adaptation of tumours allows them to be metabolically flexible. Serine is another amino acid that supports several metabolic processes crucial for growth and survival. *De novo* serine biosynthesis genes such as phosphoglycerate dehydrogenase (PHGDH) and phosphoserine aminotransferase (PSAT)1 have been shown to be increased in liver cancer [67].

1.1.4 *Conclusion*

A WD contributes to the pathogenesis of liver disease in a complex manner and increases the risk for the development of liver cancer, for which there are currently limited therapeutic options. Once tumours arise, several cell-autonomous factors, such as oncogenic mutations, impose metabolic rewiring that is distinct from the liver and support tumour growth and survival. It is still unclear, however, how diet-induced chronic changes manifested in gene expression patterns influence fuelling of liver cancer by dietary nutrients. Consequently, it is also unclear whether dietary changes can relieve the metabolic features associated with liver tumours. In addition, a major focus of cancer metabolism research is to study how different metabolic pathways contribute to growth in order to find potential targets to inhibit tumour cell proliferation. Metabolic pathways do not exist in isolation but operate concurrently controlled by various molecular mechanisms. Metabolic modelling used in systems biology can greatly aid in addressing such questions by providing the tools necessary to survey the simultaneous activity of different metabolic pathways *in silico*.

1.2 SYSTEMS BIOLOGY, A HOLISTIC AND INTEGRATIVE APPROACH TO STUDY METABOLISM

The reductionist approach dominated biological research in the later half of the 20th century. Using this method of dissecting biological systems into their constituent parts generated a wealth of information on individual cellular components and their functions. However, with whole genome sequencing becoming available in the 1990s the need for the development of integrative approaches to study the relationships and interactions between different biological entities became eminent and was highlighted by several researchers at the the start of the 21st century [68–71]. This sparked the birth of systems biology as an integrative approach for biological research that uses the tools of bioinformatics, mathematical modeling and *in silico* simulations to study biological systems [72, 73]. At the core of systems biology philosophy lies the understanding that the whole is greater than the sum of its individual parts. This drives the need to understand the system by integrating the information of its constituent parts into unified frameworks that can be used to study their relationships.

This approach requires a *de facto* construction of a mechanistic genotype-phenotype relationship between all the chemical components of a cell, their genetic basis and its physiological functions [72, 74] The biochemistry of metabolism has been studied for decades, but with the publication of full genome sequences it became possible, in principle, to study metabolism on a genome scale. Genome scale metabolic model (GSMM)s that result from this reconstruction process is the principle tool used in systems biology. These models provide a mathematical framework to understand genotype-phenotype relationships for metabolism by simulating metabolic flux without knowledge of metabolite concentrations or kinetic parameters. Moreover, the formulations used to reconstruct GSMMs (as described below), provides a framework for integration of high-throughput data, such as omics data, as well as other phenotypic data.

1.2.1 *Reconstruction of genome scale metabolic models*

1.2.1.1 *Bottom-up reconstruction*

Reconstructing a metabolic network

The first whole genome sequence was completed in 1995 for *Haemophilus influenzae* [75] which made it possible to reconstruct, using bottom-up methodologies, the first metabolic network on genome scale for this organism and was the genesis of the bottom-up reconstruction process for GSMMs [76]. Bottom-up reconstruction involves building a large network one reaction at a time in a step-by-step fashion starting with the genome sequence and annotation to identify candidate enzymes and associate these with metabolic reactions using databases and literature (Figure 1.5). This process is labour intensive and time-consuming and involves extensive manual curation, quality control and filling of network gaps through literature search. The resulting reconstruction is a biochemical, genetic and genomic (BiGG) knowledge-base, or k-base, that describes all the known metabolic reactions for a particular organism [77]. It became evident early on that there is a continuous need for improvement and update on existing metabolic reconstruction as our knowledge of the BiGG k-base of an organism expands. It is therefore not uncommon that multiple reconstructions are available for the metabolic network of a single species that each aim to improve and expand on previous models.

From a network to a model

The first steps in the reconstruction process aim to gather and curate the BiGG k-base for the organism of interest. This information is then converted to a computational readable mathematical format. The most important step in this process is the formulation of the stoichiometric matrix (S-matrix), a 2D matrix that describes the relationship between reactions and metabolites in the k-base (Figure 1.5) [74]. The S-matrix represents all biochemical reactions in the reconstruction and contains physiochemical and connectivity attributes of the metabolic network. The gene products that catalyse reactions are formulated as gene-reaction rules (GR-rules) and provide the basis for integra-

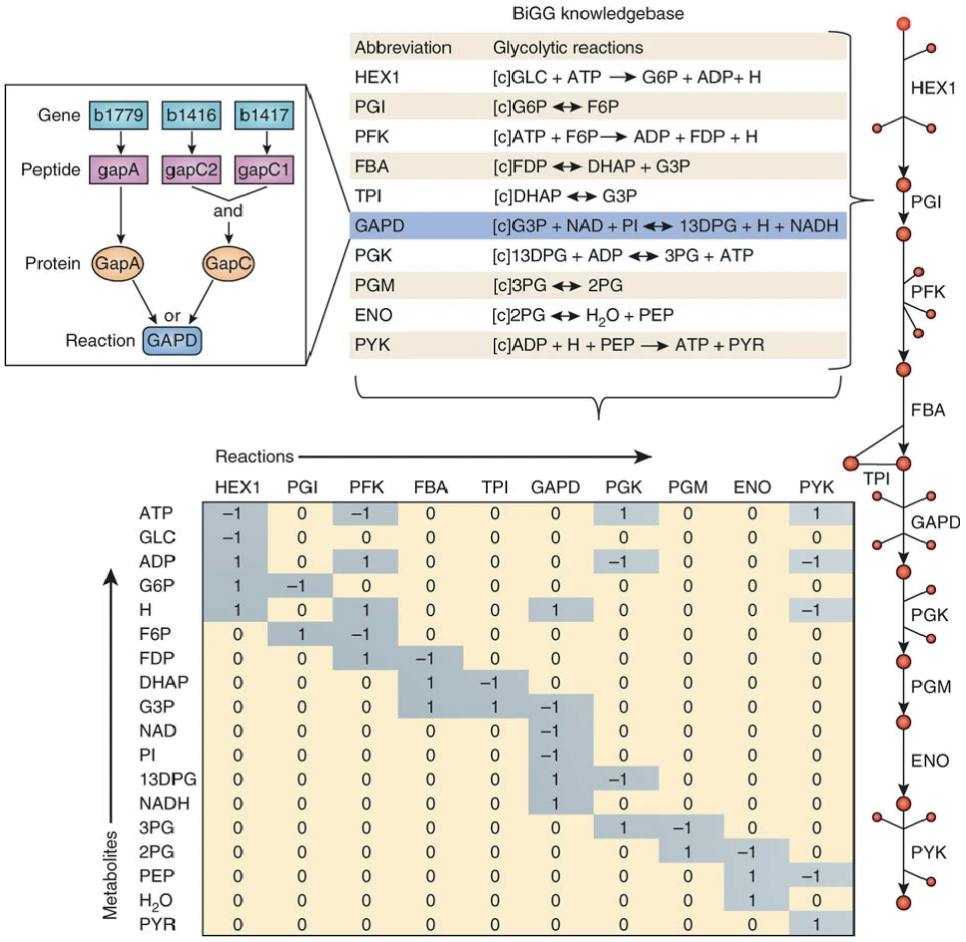


Figure 1.5: **Construction of the BiGG k-base and its conversion to a mathematical format**. The genome annotation of the organism of interest is used to identify enzymes and associate these with metabolic reactions. The result is a BiGG k-base of all metabolic reactions of interest. This can be converted to a mathematical format in the form of a matrix, the S-matrix, of metabolites *versus* reactions. In addition, gene products in the k-base are represented as GR-rules using Boolean rules. Figure from [78].

tion of phenotypic data (discussed in Section 1.2.2). Having a mathematical framework for a metabolic network reconstruction allows computation of cellular functions that result from the simultaneous activity of multiple gene products. The final step in the reconstruction process aims to evaluate and validate the computational model. This involves evaluations, such as testing if all reactions are mass and charge balanced, identifying metabolic dead-ends and blocked reactions to more involved procedures such examining if all organism specific pathways are represented, performing network gap-analysis and evaluating whether the single gene deletion phenotypes are in congruence with experimental knockouts. This evaluation is an iterative process that requires going back to the draft reconstruction for refinement. At this stage, the network can be used to start testing basic metabolic functions, for example, generating all precursor metabolites required to synthesise macromolecules for growth, evaluate whether the reconstruction support growth on substrates known that the organism can use for growth or determine if known metabolite products are secreted [74, 79].

The conversion of a BiGG k-base into a computational framework along with successive model testing as described above gives rise to a metabolic model with predictive capability. In addition, the mathematical formulations generated in this process are required for integration of data that forms the basis for the constraint-based modelling (CBM) approach [80, 81]. This approach is fundamental in GSMMs and considers biological constraints that exists on a cell, transforms this information into mathematical formulations, and brings the model as close to a specific biological contexts as possible.

1.2.1.2 *Evolution of in silico metabolic reconstructions of human and mouse*

Since the dawn of systems biology, significant effort has been made to build an exhaustive BiGG k-base of human metabolism with the purpose to study how metabolism can be used to treat disease. This led to the reconstruction of several human GSMM (hGSMM)s that have been used in CBM to study metabolic phenotypes in different contexts.

hGSMMs

The first hGSMMs, Recon1 [82] and the Edinburgh Human Metabolic Network (EHMN) [83], were reconstructed using a bottom-up systems biology approach in 2007 (Figure 1.6). These two models were used as the starting point for two model series of human metabolism; the Recon series (Recon1, Recon2 [84] and Recon3D [85]) and the Human Metabolic Reaction (HMR) series (HMR [86] and HMR2 [87]), which with each new version aimed to increase the number of metabolic reactions or improve predictive capability of the models. After the reconstruction of Recon1 and EHMN, HepatoNet1 [88] was published as the first hepatic GSMM that has been used in updated versions of GSMMs in the Recon and HMR series to augment network coverage.

In various iterations of the Recon series, BiGG information has been incorporated with the aim to organise this information in a formal, mathematical fashion. For example, in Recon2 M.2 a framework for gene-transcript-protein-reaction associations were integrated into the model that considered the effect of alternative splicing on metabolism [89] In Recon3D, three-dimensional (3D) metabolite and protein structure data were included that enable integrated analyses of metabolic functions in humans [85]. Compared to the Recon series, HMR has more comprehensive information on fatty acid metabolism that has been manually curated [87]. Models in the HMR series have also been used in the generation of various cell-type specific GSMMs, including *iAdipocytes1809* [90], *iHepatocytes2322* [87], and *iMyocyte2419* [91] that has been used to study different metabolic diseases. Although Recon and HMR reconstructions share the aim to capture all human metabolic reactions, their different formulation of metabolite, gene and reaction identifiers made direct comparison difficult. Therefore, in an attempt to unify these two model series, Recon3D and HMR2 were recently integrated into a single reconstruction, Human1 [92].

mGSMMs

In 2005, Lars Nielsen and colleagues reconstructed the first mouse GSMM (mGSMM) that comprised 1220 metabolic reactions using a bottom-up approach [93]. Thereafter, Selvarasu *et al* expanded this model by adding reactions and GR-rules from

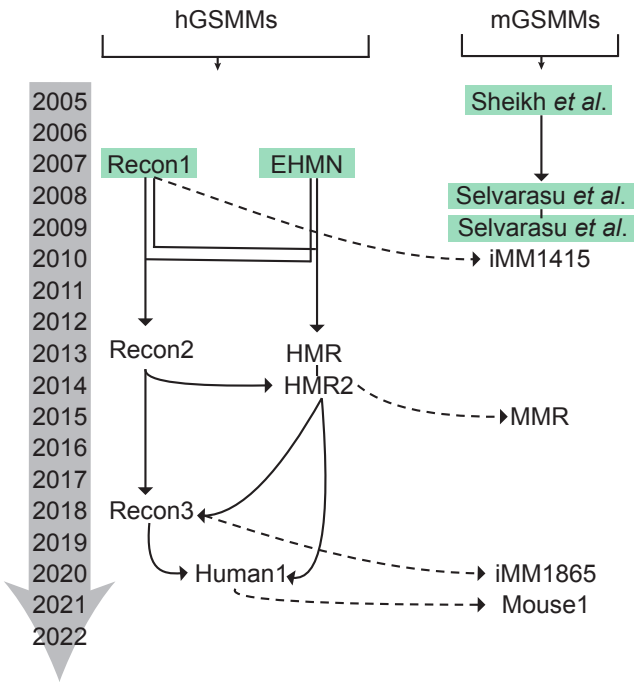


Figure 1.6: **Evolution of the reconstruction of hGSMMs and mGSMMs.** Bottom-up reconstructions are boxed in green and mGSMMs reconstructed using orthology are depicted by dotted lines. Figure adapted from [92].

literature as well as improving network connectivity of several metabolic pathways [94, 95]. The resulting model comprised 1494 reactions and three cellular compartments; cytosol, mitochondria and extracellular space. The number of reactions in these reconstructions were less compared to hGSMMs available at the time. Because the bottom-up reconstruction process is cumbersome and labor-intensive, iMM1415 was reconstructed by replacing human genes in Recon1 with their corresponding mouse ortholog [96]. This approach became widely known as the orthology-based approach for metabolic network reconstruction (Figure 1.7). The Mouse Metabolic Reaction (MMR) [97] database and iMM1865 [98] were thereafter reconstructed from on HMR2 and Recon3D, respectively, using an orthology based approach. Finally, Human1 was used as template to reconstruct GSMMs for several model organisms including the mouse, termed Mouse1 [99].

1.2.2 *Constraining GSMMs*

Modelling biological systems, and in particular metabolism, requires constraint-based thinking as opposed to theory-based thinking (Figure 1.8) [80, 81, 100]. Theory-based thinking attempts to seek a single exact solution based on the laws of physics and chemistry. In biology, however, an organism or cell can have many different behaviours based on a specific context. Moreover, over the last century there has been an increasing appreciation that the same biological function can be achieved in different ways. This is a distinction in the philosophy of mathematical modelling in biology compared to physics and engineering and the basis of constraint-based thinking.

A cell is faced with a myriad of constraints that are dynamically changing under different contexts. If we consider a solution space with infinite possible solutions, the formal imposition of constraints reduces this space into candidate functional states each defined by a set of mathematical equations that can be solved using optimisation algorithms. Constraints are represented as either *balances* or *bounds* and lie at the heart of flux balance analysis (FBA) [101], a widely used approach for studying GSMMs. This approach is described in detail below. Firstly,

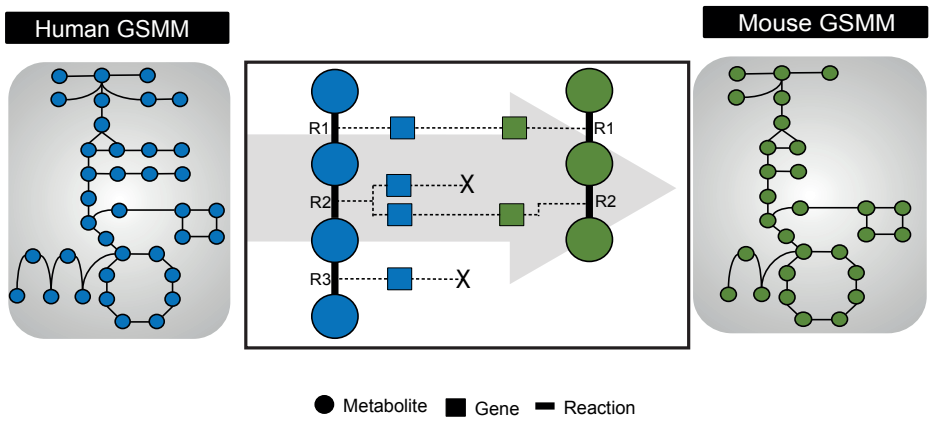


Figure 1.7: **Schematic of orthology-based approach for network reconstruction**. An extract of three reactions from a larger template hGSMM (left) is represented in blue. Genes are associated with reactions using GR-rules. The orthology mapping will have one of three different outcomes to decide whether a reaction will be included in the orthology mGSMM (right), depicted by reactions R1, R2 and R3, respectively. R1: an ortholog exists and therefore this reaction will be included. R2: a reaction catalysed by more than one gene should have at least one ortholog to be included in the mGSMM and genes without an ortholog will be omitted (shown with X). R3: no ortholog exists for this reaction and will not be included in the mGSMM.

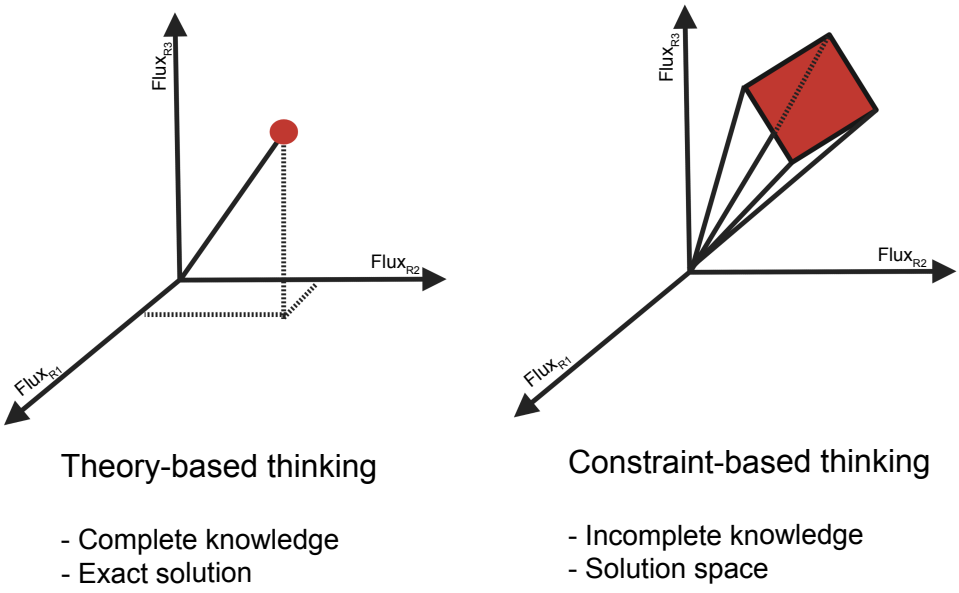


Figure 1.8: **Theory-based thinking versus constraint-based thinking**. Theory-based analysis in physics and engineering seeks to find a single exact solution that is most often based on laws of physics and chemistry. In biology, different equivalent solutions often exists for a single function and, therefore, constraint-based analysis relies on confining all possible solutions based on prior knowledge to a space that would likely satisfy the function. Figure adapted from [73].

the change in metabolite concentration over time can mathematically be described as

$$\frac{dx}{dt} = S.v$$

where x represents a concentration vector of metabolites, S the stoichiometric matrix and v a vector of fluxes. If steady-state is assumed this equation can be transformed to

$$0 = S.v$$

that ensures that the total amount of any compound being produced must be equal to the total amount being consumed. This steady state assumption is generally valid because of the fast equilibrium of concentrations of metabolites with respect to the time scale of genetic regulation. This defines a system of linear equations. Because the S-matrix is a sparse matrix that contain more reactions than metabolites, there is more than one solution to these equations. Therefore, bound constraints, represented by inequalities are often used to limit the numerical ranges of individual fluxes given by

$$a_i < v_i < b_i$$

where a and b represent lower bound and upper bound values for reaction i, respectively. These constraints therefore limit the numerical ranges of individual fluxes to reduce the possible solutions. Finally, an objective function is formulated with

$$Z = c^T v$$

where c is a weight vector indicating how much each reaction in v contributes to the objective. This set of equations can then be optimised by using optimisation algorithms such as linear programming (LP) that solves equalities and inequalities and a linear objective function. Various data types can be used for constraining reactions bounds. There has, however, been particular interest in the development of methods that use high-throughput omics data (transcriptomics, proteomics, metabolomics and fluxomics) to used in CBM to constrain different parts of the metabolic network [81, 102]. In this regard, GSMMs

also provide contextualisation of omics data by providing a framework to simultaneously integrate different data types for a more holistic understanding of potential underlying molecular mechanisms underlying phenotypes.

Three principal groups of constraints, intracellular fluxes, exchange reactions and objective functions are discussed below:

Constraining intracellular reaction fluxes

Kinetic parameters of enzymes are the most accurate data describing the capacity of individual enzymes [103, 104]. However, for most eukaryotic species these parameters are only available for a small number of enzymes and not readily obtainable by experimental measurements. In addition, the enzymes responsible for catalysing some metabolic reactions have been predicted by homology to lower organism enzymes that have been previously described without confidence that it is indeed the true catalytic enzyme. These are presented by 'or' relationships in the GR-rules and often contains families of homologous enzymes that are from the same enzyme family. Methods have therefore been developed that use omics data, in particular transcriptomics and proteomics, for constraining individual reaction bounds based on enzyme abundance [105].

Constraining uptake and release fluxes

The computational formulation of a metabolic network allows the formulation of cellular metabolism as a system [100]. Defining a system boundary and exchange reactions allow the imposition of constraints on metabolite uptake and release fluxes. For cultured cells in defined media, metabolomics measurements of media components, either at end-point or over time, can determine metabolite uptake and release rates that can be used to constrain exchange reactions [106]. For GSMMs of physiologically intact cells or tissues, this data is not readily available and most often not experimentally measurable. The diet composition has been used for this purpose in several studies, and in particular, for tissue-specific GSMMs [97, 106]. The limitation of this approach is that not all dietary nutrients are metabolised by tissues but first by the gut microbiota and intestine. In order to

overcome this, integrative whole-body GSMMs integrate microbiome GSMMs with tissue GSMMs to more accurately represent metabolic physiology [14].

Constraining GSMMs by defining objective functions

Formulation of the flux balance analysis (FBA) problem is rooted in the hypothesis that a cell is "striving" to achieve a certain metabolic objective [101, 107]. Therefore, even though the objective function is a mathematical formulation, there is also an underlying biological consideration in choosing an objective. The mathematical form of the objective is $Z = \langle w, v \rangle$ where w is a weight vector for the reactions of interest to optimise and v is the vector of network fluxes [101]. Biological objectives used in most studies are to test physiological functions, explore specific network properties or in some instances have a pertinent bioengineering application [107].

The most commonly used objective function is the production of biomass [79]. The biomass reaction is a composite reaction that is formulated to comprise of all the macromolecules necessary for growth. For many unicellular organisms, such as bacteria and yeast, GSMMs have been used for biotechnology purposes where growth often represents a desired objective. However, other cellular objectives, such as, maximal ATP production, minimisation of nutrient uptake required for growth or maximisation of the production of specific metabolites have also been used in instances with specific biological or engineering questions. Defining an objective function for differentiated eukaryotic cells or tissues has proven to be more involved and are a widely discussed topic in the metabolic modelling community. Nevertheless, several studies still consider biomass production as the objective for eukaryotic cell-specific and tissue-specific GSMMs when predicting flux distributions. In the first instance, this gives a reliable test of functionality of the model to ensure that the metabolic pathways required for the synthesis of essential macromolecules to maintain survival do indeed carry flux [108]. Furthermore, although the actual predicted fluxes under the assumption of biomass as the objective might not be close to physiological fluxes, the relative metabolic activities can

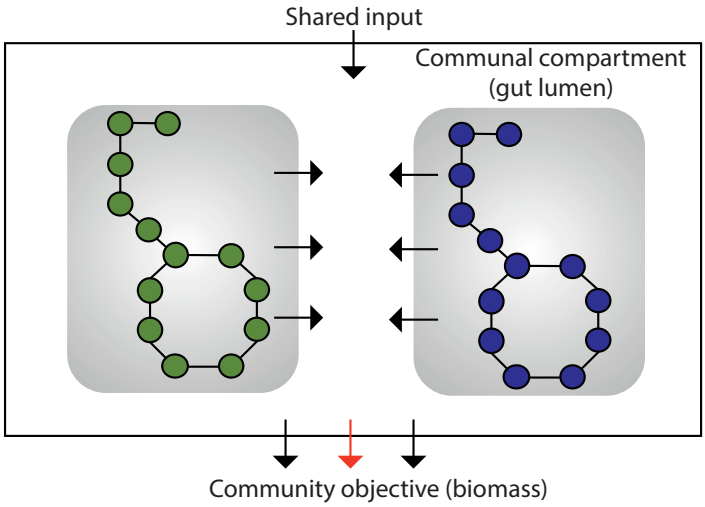


Figure 1.9: **Compartmentalization of GSMMs allows microbial community and multi-tissue modelling**. In bacterial communities, individual species interact metabolically. GSMMs can be used to model this by considering each microbe as a distinct compartment. The community can have a shared input as well as an objective function, most often community biomass production.

be compared between different pathways and under different constraining parameters.

1.2.3 *Microbial community modelling*

The metabolic interactions between microbes are essential for the functionality of microbial ecosystems and metabolic modelling has greatly advanced our understanding of these interactions. Stolyar and colleagues were the first to study mutualistic metabolic interactions between two bacterial species using GSMMs [109]. In their study they proposed to consider each individual microbe as a separate compartment and allow metabolic interaction by defining a third compartment through which metabolites could be exchanged between organisms. This allowed metabolites produced by one microbe to be taken up and metabolised by the other (Figure 1.9).

A wealth of metagenomics studies have recently suggested that the microbial community in the human gut is incredibly rich in diversity, associated with several different diseases and

that it is a major metabolic organ [49, 110]. Significant effort has therefore gone into reconstructing GSMMs for individual microbial species of the gut, but also modelling the interactions of this highly dynamic community [111]. Most recently, GSMMs for 773 bacterial species in the human gut have been reconstructed using a semi-automated computational pipeline, known as the Assembly of Gut micrOorganisms through Reconstruction and Analysis (AGORA) models, and have been used in several studies to investigate gut metabolism [112]. A current major challenge is to functionally link gut community GSMMs with host GSMMs to investigate how modulation of gut microbiota and diet potentially impact host metabolism.

1.2.4 *Multi-tissue mammalian modelling*

The approach to representing individual microbes as distinct compartments in a GSMM have also been used to study the metabolic interactions between different tissues. The first multi-tissue GSMM comprised of distinct hepatocyte, myocyte and adipocyte GSMMs connected through a blood compartment and the model was shown to be able to recapitulate known tissue metabolic cross-talk [113]. More recently, a dynamic multi-tissue GSMMs consisting of liver, muscle and adipose tissue-specific GSMMs were used to investigate a novel more physiologically accurate objective function [114]. In short, this objective was based on the assumption that metabolism aims to maintain blood concentration homeostasis and that energy might be stored in stores such as glycogen and fat to be used in times of food scarcity such as fasting. The most comprehensive integrative GSMMs are represented by two organ resolved, sex-specific GSMMs that consist of 26 organs, 13 distinct biofluid compartments and a representative gut microbiome compartment [14]. This GSMM serves as an important resource to study metabolism *in silico* in a complete physiological context.

1.2.5 *Conclusion*

Over the past three decades, genome scale metabolic modelling has seen major advances both in reconstruction and constrain-

ing methodologies. However, there is still room for improving aspects of especially mammalian GSMMs

limitations still remain and there is still room for improving several aspects of GSMMs, especially for mammalian GSMMs. Firstly, compared to constraining intracellular fluxes, studying how nutritional input in GSMMs effect metabolism has seen less attention in literature. Secondly, in the context of human diseases, GSMMs are most often focusing on individual tissues and do not explicitly model metabolic interactions between tissues. Similarly, studies that model gut microbiome metabolism lack functional integration of how these predictions impact host metabolism and mostly rely on correlative analysis to confirm predictions.

1.3 PRIMARY OBJECTIVES OF THE THESIS

1. Reconstruct a new mouse genome scale metabolic model for the integration of experimental data to model liver metabolism

2. Use constraint-based modelling to investigate the contribution of dietary nutrients to the metabolic profiles of liver and tumour tissues

3. Investigate how the gut microbiome metabolises dietary nutrients with community metabolic models

4. Use integrative modelling to understand metabolic interactions between the liver host and tumour as well as microbiome and host and tumour-host metabolic interactions

2 METHODS

"*What I cannot create, I do not understand*" - Richard Feynman

2.1 EXPERIMENTAL METHODS

Experiments were performed by the following colleagues at The Francis Crick Institute: Patricia Nunes, Albert Thommen, Antonia Kasampali and Jack Carruthers.

All the experimental procedures were conducted in conformity with public health service policy on humane care and use of laboratory animals, approved by The Francis Crick Institute's Animal Welfare and Ethical Review Body (AWERB) and comply with a license ratified by the UK Home Office.

2.1.1 *Liver cancer mouse model*

C57BL/6J mice were housed under a light-dark cycle of 12:12h with controlled temperature (22-24°C). Two-week old male mice were injected intraperitoneally with 25mg/kg of the carcinogen diethylnitrosamine (DEN). From the time of weaning, mice were fed a 'western-style' diet (WD) (TestDiet, AIN-76A). The fractional content of diet contributed by lipids and carbohydrates in WD is similar to an average European diet [13] and the injection of DEN in combination with WD-feeding has been shown to induce liver tumour formation [115]. This mouse model was used to generate data for different experimental conditions.

2.1.2 *Tissue collection*

At indicated experimental time points, mice were culled, livers were rapidly excised, tumours, where existing, were separated

from peritumoral tissue. In addition, cecum content were harvested from intestine together with fecal matter. All tissues were snap-frozen in liquid nitrogen and stored at -80°C.

2.1.3 *mRNA extraction, library preparation and RNA-sequencing*

Total liver and tumour tissue were pulverised with mortar and pestle under a liquid nitrogen atmosphere and RNA was extracted from the equivalent of 20-30mg frozen tissue in TRIzol Reagent (Thermo Fisher Scientific, UK) followed by phenol removal with chloroform. RNA was further purified with the RNeasy Mini Kit (Qiagen, UK). DNase treatment was performed to avoid genomic DNA contamination. After RNA quantification and quality controls for integrity and purity (Nanodrop, Qubit and Agilent 2100 Bioanalyzer), libraries were prepared using KAPA mRNA HyperPrep Kit (Kapa 13 Biosystems). mRNA sequencing (single-end, 40 million reads total) was performed on an Illumina HiSeq 2500 instrument.

2.1.4 *DNA extraction, library preparation and metagenomics sequencing*

Genomic DNA was extracted from frozen samples of cecum content and faeces using the NucleoSpin® Soil kit (Macherey-Nagel, UK) and following manufacturer instructions. Briefly, sample material was mechanically dissociated with beads and lysed, followed by DNA precipitation and purification. Upon elution, DNA content was determined by OD 260/280 with a Nanodrop. Samples were loaded into an agarose gel and run by electrophoresis to ascertain integrity of the DNA extracted. Libraries were prepared using Illumina TruSeq PCR-free library preparations (350bp average fragment size) from DNA samples. DNA sequencing was performed on a Illumina NovoSeq S4 instrument.

2.1.5 *Respiratory exchange ratio measurements in metabolic cages*

Eight DEN and eight nonDEN mice were individually housed in a laboratory animal monitoring system (TSE Phenomaster,

TSE Systems GmBH). After an acclimatisation period, O_2 consumption and CO_2 production, food, water intake and activity were continuously monitored for each mouse for a period of 48h. The mean O_2 consumption and CO_2 production rates measured over 48h were used to constrain context-specific genome scale metabolic models (GSMMs) – see Section 2.2 for details.

2.1.6 *Isolation of mitochondria from liver and oxygen flux measurements*

Mice were culled and the tissues of interest (liver, peritumoral liver and tumour) were quickly excised and rinsed in cold phosphate buffer before being transferred to mitochondrial isolation buffer (250mM sucrose, 10mM Hepes, 0.1% BSA fatty-acid free, pH 7.2). Tissue was homogenised on ice with a Potter-Elvehjem homogeniser before centrifugation at 800xg for 10 min at 4°C. After discarding the top lipid layer, the supernatant was centrifuged at 10000xg for 10 min at 4°C. The pellet was resuspended in washing buffer (250mM sucrose, 10mM Hepes pH 7.2) and centrifuged again at 10000xg for 10 min at 4°C, a process that was repeated twice. The resulting mitochondrial pellet was resuspended in 0.5mL of washing buffer. Mitochondrial protein quantification was performed with Pierce BCA Protein assay. Mitochondrial respiration driven by fatty acid oxidation was assessed with an oxygen electrode system (Oroboros Oxygraph-2K, Oroboros Instruments, Austria) using 1mg of mitochondrial protein in 2mL of working buffer (130mM sucrose, 50mM KCl, 5mM KH_2PO_4, 5mM $MgCl_2$, 5mM Hepes, 50µM EDTA, pH7.2 at 37°C) supplemented with 4mM ADP and 0.5mM octanoylcarnitine.

2.1.7 *DNL assessed with ^2H_2O*

The contribution of *de novo* lipogenesis (DNL) to hepatic or tumoural triacylglycerol (TAG) pool was determined *in vivo* using ^2H_2O [116, 117]. In brief, mice were intraperitoneally injected with ^2H_2O and supplied with 5% ^2H_2O drinking water while feeding *ad libitum*. 16 h later, animals were culled, tissues were collected and frozen in liquid N2. Tissues were powdered

under liquid N2 atmosphere and lipids were extracted with the Folch method [118]. TAGs were isolated using ISOLUTE® NH2 SPE (Biotage, UK). *De novo* synthesised fatty acids (FAs) and glycerol incorporated in TAGs were determined as the fraction of FA methyl or glyceryl moiety, respectively, in TAG that were labelled with ^2H, normalised for the enrichment of ^2H in H2O in the plasma of each mouse measured using ^2H NMR [119].

2.2 COMPUTATIONAL METHODS

Sequencing data was processed on either the Crick computational cluster (CAMP) or the Swedish National Infrastructure UPPMAX servers. All other analyses were performed on an Intel Core i7 MacBook Pro. Transcriptomics data were processed in R version 4.1.2 using the packages described. Metabolic modelling was performed in MATLAB R2019b using functions in the COBRA [120] and RAVEN2.0 [121] toolboxes and resulting flux distributions were further analysed in python3.10 in Jupyter Notebooks using the scipy stack. Plots were either generated using plotly, seaborn or ggplot2.

Scripts and data available at `https://zenodo.org/badge/latestdoi/536721334`.

2.2.1 *Processing of RNA-sequencing data*

Raw RNA-sequencing data were processed using an in-house analysis pipeline. Quality of raw sequencing data was checked with FastQC v0.11.7 (`http://www.bioinformatics.babraham.ac.uk/projects/fastqc`). Reference genome alignment was performed against Genome Reference Consortium Mouse Build 38 (GRCm38) with RSEM [122] and STAR [123]. RSEM was used to generate raw counts, fragments per kilobase million (FPKM) and transcripts per kilobase million (TPM) which were used for all downstream analyses. Quality control metrics were reported with picard, RSeQC [124] and RNA-SeQC [125]. A final quality control report was generated with MultiQC [126] .

2.2.2 *Differential gene expression and enrichment analysis*

The RSEM gene count matrix was used for differential gene expression (DGE) analysis with DESeq2 [127] with the *apeglm* [128] algorithm for $\log_2$-fold change shrinkage. Unless otherwise specified, gene expression differences with an absolute minimum $\log_2$-fold change of 1 and FDR-adjusted p-value 0.05 were considered statistically significant. The *clusterProfiler* [129] package was used for Gene Ontology (GO) over- representation tests using the *enrichGO* [130] function and the Benjamini-Hochberg method was used to correct for multiple testing.

2.2.3 *Processing of metagenomics data*

Raw metagenomics sequence data was processed using the quantitative metagenomic profiling software METEOR (available at `https://forgemia.inra.fr/metagenopolis/meteor`) for quality control, trimming and mapping against a catalog of mouse gut metagenome that consists of ~2.6 million nonredundant genes [131] to create gene count tables. To reduce variability, downsizing of 10 million reads was performed on the gene count tables and the MetaOMineR [110] software package was used for normalisation of gene counts for metagenomic species (MGS) abundance calculation.

2.2.4 *Reconstruction of a new mGSMM*

2.2.4.1 *Reconstruction of orthology-based mouse metabolic networks*

An orthology-based approach was used to reconstruct a new mouse genome scale metabolic model (mGSMM) (Figure 1.7). Human-to-mouse orthologs were downloaded using the online BioMart tool (`www.biomart.org`, date accessed – January 2019) and genes in HMR2 [87] were directly replaced with their corresponding mouse ortholog to generate intermediate model (IM)1. The *getBlast* in the RAVEN2.0 toolbox [121] was then used to to perform protein sequence similarity between the amino acid sequences of proteins in the GRCm38 and the amino acid sequences of proteins encoded by genes in HMR2 and Recon3D [85], respectively. The resulting BLAST structures, mouse-to-HMR2 and

mouse-to-Recon3D, were then used as input for the *getModelFromHomology* function to generate IM2 and IM3, respectively, using an e-value cutoff of 10^{-30}, a metric used to describe the number of hits one can expect to see by chance, with a lower value indicating a more 'significant' match. This function aims to replace human genes within the reference GSMM, either HMR2 or Recon3D in this case, with corresponding mouse orthologs provided there is good sequence alignment between genes.

2.2.4.2 *Integration of orthology-based networks to reconstruct a new mouse GSMM*

Two metabolic reconstructions can be merged by integration of their respective S-matrices. For this purpose, either the metabolites and/or reactions have to have the same identifiers in the two models to allow removal of duplicate reactions after the two models are merged. A metabolite map was therefore constructed that uses Kyoto Encyclopedia of Genes and Genomes (KEGG) identifiers for metabolite identifiers and metabolites in IM1-3 were renamed using this map. Finally, metabolite identifiers in MMR were also renamed to have the same nomenclature to generate IM4. This universal metabolite nomenclature allowed a step-wise integration of IM1-4 by merging their respective S-matrices with the *mergeModels* function in RAVEN2.0 to generate a new mouse GSMM, mouse metabolic reaction network (MMRN).

2.2.4.3 *Computational evaluation and benchmarking of MMRN*

Duplicate reactions, metabolites and genes in MMRN were removed. Three criteria were used to assess the resulting network; connectivity, metabolic tasks and elemental balance of metabolic reactions. The network was converted to a bipartite graph of metabolites and reactions and all nodes not connected to the single most connected component were removed as these could lead to metabolic dead-ends. The biomass reaction from HMR2 was added and the model was assessed to be able to perform 56 common metabolic growth tasks using the *checkTasks* function [132]. Finally, elemental imbalanced reactions were identified using the *getElementalBalance* function and removed in a step-wise manner ensuring fulfilment of metabolic tasks. Reactions that resulted in the failure of a metabolic task were manually curated

for mass-balance. Three mGSMMs, Mouse1 [99], MMR [97] and iMM1865 [133], and three hGSMMs, Human1 [92], HMR2 and Recon3D, were used for benchmarking of MMRN using two criteria; elemental balance and connectivity with the same method described above.

2.2.5 *Reconstruction of MMRN^{Hep}*

The maximum FPKM expression values across all mouse experimental conditions were calculated for genes in MMRN. This was used with 56 common metabolic growth tasks as input for the task-driven integrative network inference for tissues (tINIT) algorithm [132] to generate a generic hepatic GSMM, MMRN^{Hep}. The expression of genes was classified as high (FPKM > 50), medium ($10 \geq$ FPKM > 50), low ($1 \geq$ FPKM < 10) or no expression (FPKM < 1).

2.2.6 *Reconstruction of csGSMMs from MMRN^{Hep}*

An adapted version of the E-flux method [134] was used to construct csGSMMs from MMRN^{Hep}. For this purpose, a vector b was constructed for each experimental condition using expression data and GR-rules in MMRN^{Hep}. For a reaction, j, catalysed by a single gene the mean expression value across biological replicates was considered for b_j. For reactions with multiple enzymes associated with 'or' relationships the sum of the mean expression values for individual genes was considered for b_j. For reversible reactions the negative value of b_j was also imposed as a lower bound allowing these reactions to be bi-directional. In order to account for noise and in the transcriptomics data, the logarithm of the calculated bound values were used as the constraints to ensure that imposed boundaries are on the same scale. Orphan reactions, reactions without a gene association, were kept unconstrained. The respiratory exchange ratio (RER) for DEN and nonDEN mice were considered separately to constrain O_2 uptake and CO_2 production for DEN and nonDEN GSMMs, respectively. For this purpose, the mean volumes of O_2 consumption and CO_2 production for 5 mice were converted to flux values (mmol/mouse/day):

$$b_{RER} = \frac{pV}{M} \qquad \text{(Eq1)}$$

where b_{RER} is the flux value value, p the density, V the average volume of either O_2 or CO_2 and M the molecular weight of either O_2 or CO_2, respectively. The calculated flux values were used to set the lower bound of corresponding O_2-consuming and CO_2-producing reactions.

Finally, the content of the WD and the control diet (CD) were used to constrain uptake rates (or exchange reactions) of the GSMM:

$$b_{DIET} = \frac{pW}{M} \qquad \text{(Eq2)}$$

where b_{DIET} is the flux value value, p is the percentage weight of a metabolite, W the average grams of food consumed per mouse per day (3g) and M the molecular weight of the particular metabolite. The calculated flux values (mmol/mouse/day) were used to constrain the upper bounds exchange reactions and thus reflect the maximum uptake rate of a particular metabolite available to the model to perform FBA. The carbon flux (or C_{moles}) for each diet was calculated using:

$$C_{moles} = \sum_{j=1}^{n} Cb_j \qquad \text{(Eq3)}$$

where C_{moles} is the moles of carbons and b the upper bound flux for $j = 1$ to n metabolites within each diet. The same equation was used to calculate the C_{moles} for uptake and release by using the predicted flux value, v, instead of b for each metabolite.

The model setup is illustrated in Figure 2.1.

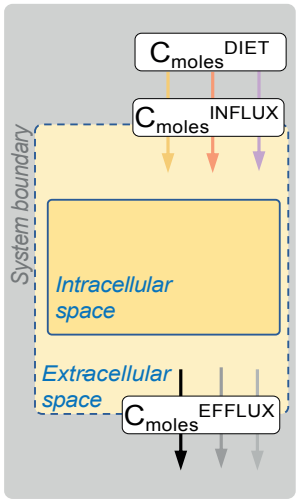


Figure 2.1: . **Schematic illustrating relation between influx and efflux C_{moles} in the context of FBA experiments**. C_{moles}^{DIET} represents the carbon flux for each metabolite that is available from the diet to $MMRN^{Hep}$ and was calculated based on the known diet composition and daily diet consumption per mouse. The dashed line represents a computational pseudo-boundary set to allow influx of metabolites from the diet into the extracellular space of the model. For a given metabolite, C_{moles}^{INFLUX} and C_{moles}^{EFFLUX} denote the flux of carbons of this metabolite taken up or produced, respectively, by $MMRN^{Hep}$.

A note on notation

The following notation is used throughout the thesis to describe csGSMMs: $[MMRN^{Hep}]_{Y}^{X}$ where X indicates the gene expression constrain (nil where no expression constrain is applied) and Y is the diet given to the model.

2.2.7 *Community modelling of microbiome*

The top 20 most abundant MGSs in each condition were used to reconstruct microbial community GSMM (comGSMM)s using the publicly available AGORA models [112]. For each MGS, if a model was not available it was checked whether a model for the corresponding genus is available and if this genus is not already represented by another MGS, the models for all members of this genus were included in the final reconstruction. The

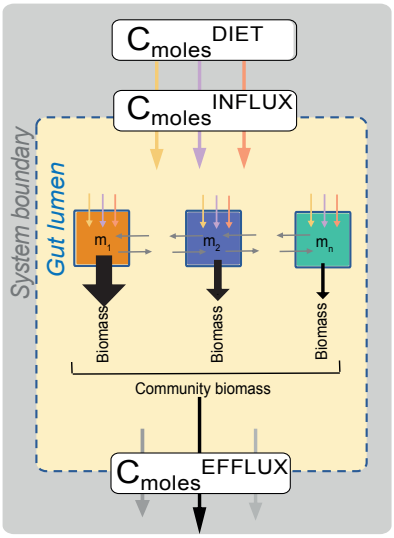


Figure 2.2: **Schematic illustrating setup of microbial community GSMMs**. The relationship between C_{moles}^{DIET}, C_{moles}^{INFLUX} and C_{moles}^{EFFLUX} are the same as described in Figure 2.1. Individual microbial models are allowed to consume nutrients from the diet and produce metabolites into a compartment representing the gut lumen which can either be further metabolised by other microbes or effluxed from the model. As described in the text, the flux boundaries of the biomass reactions of individual models were constrained to their relative abundance, and, in addition to community biomass production, set as the objective function of all simulations.

lower bound of the biomass reactions of each individual was constrained to its relative abundance so that the total biomass produced sum to 1. The objective function was set as biomass production of each individual GSMM with the community biomass as an additional objective. Each individual microbial GSMM was represented as an individual compartment within one universal extracellular compartment representing the gut lumen and the dietary content was used as uptake constraint for the community. This setup allowed individual GSMMs to exchange metabolites with each other and that the comGSMM can consume dietary nutrients and produce metabolites (Figure 2.2).

2.2.8 *Reconstruction of an integrative multi-tissue GSMM*

multi-tissue GSMM (mtGSMM)s were reconstructed for different experimental conditions by combining csGSMMs for liver and tumour tissues with the respective comGSMM of that condition. More specifically, the multi-tissue model consisted of three major compartments, a diet, gut and blood. The diet compartment represents a computational space where the uptake into the compartment is constrained using Equation 2 and dietary metabolites that are taken up into this compartment can then be consumed by either the gut or the blood. The gut compartment contained the microbiome community model and can efflux metabolites out of the model or exchange metabolites to the blood compartment. The blood compartment contained the liver and tissue csGSMMs. An additional set of exchange reactions were added between the liver and tumour to allow direct cross-talk between these two tissues and exchange to the blood compartment. The blood compartment can then efflux metabolites out of the mtGSMM. The model setup is illustrated in Figure 2.3.

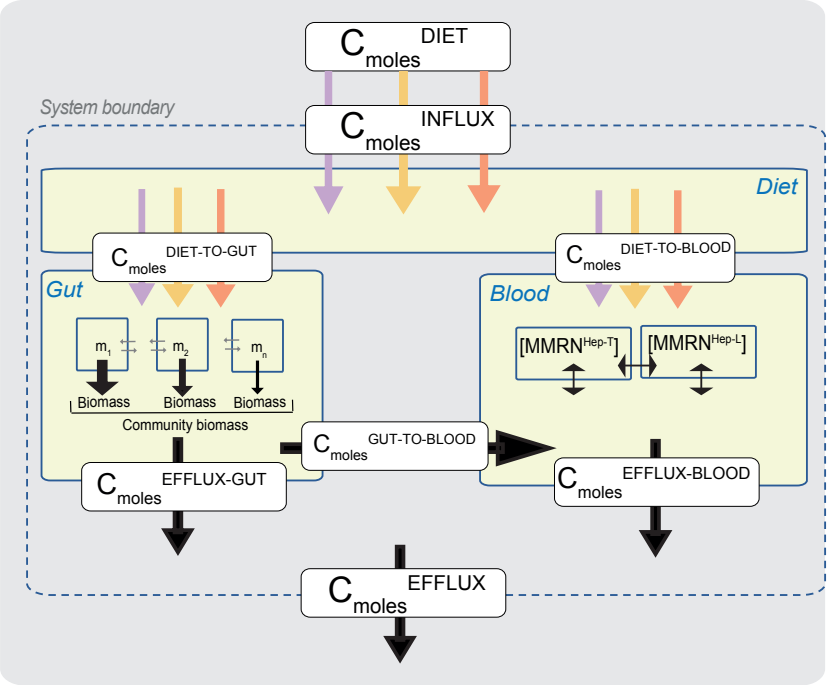


Figure 2.3: **Schematic illustrating setup of integrative mtGSMM**. The relation between C_{moles}^{DIET}, C_{moles}^{INFLUX} and C_{moles}^{EFFLUX} are the same as described in Figure 2.1. Diet is taken up into a dietary compartment and can then be exchanged either to the gut or blood compartment and is described by $C_{moles}^{DIET\text{-}TO\text{-}GUT}$ or $C_{moles}^{DIET\text{-}TO\text{-}BLOOD}$, respectively. An additional set of artificial reactions are added to allow exchange from the gut to the blood compartment described by $C_{moles}^{DIET\text{-}TO\text{-}BLOOD}$. The gut compartment is modelled according to the community model described in Section 2.2.7. In the blood compartment each individual tissue occupies a distinct compartment and can exchange metabolites into a universal blood compartment.

3

A NEW MOUSE GENOME SCALE METABOLIC MODEL

"*Pathways are concepts, networks are reality*" - Uwe Sauer

A significant effort has been made to build a comprehensive k-base of human metabolism over the last three decades [77, 92]. In parallel to this endeavour, the development of computational tools to reconstruct and constrain genome scale metabolic models (GSMMs) to study metabolism in specific biological contexts have seen similar advances [120, 121]. Comparative genomic studies have identified gene orthologs between different species which has successfully been used in systems biology to exploit the tremendous effort that has gone into reconstruction of hGSMMs to reconstruct metabolic networks of other species, including the mouse. Because the laboratory mouse is an important experimental model, mGSMMs are now widely used computational tools to study metabolism.

The Recon and HMR series of human metabolism have previously been used as templates for the reconstruction of several mGSMMs [96–98, 135]. However, at the time of conception of this thesis all mGSMMs used a single model from either the HMR and Recon series as reference and replaced the genes in the hGSMM with mouse orthologs described in orthology databases. As described in Chapter 1, any potential human-to-mouse orthologs not described in the database used will potentially result in incorrectly omitting reactions from the mGSMM.

In this chapter I describe my efforts to reconstruct a new mGSMM by using an approach that addresses both the aforementioned issues in previous mGSMMs. In the first instance, HMR2 and Recon3D are used as template hGSMMs and, secondly, known orthologs together with newly identified orthologs through protein sequence similarity are used to reconstruct intermediate orthology-based GSMMs. Orthology-based GSMMs

are then integrated to yield a new mGSMM that is evaluated using computational methods.

3.1 RESULTS

3.1.1 *A new mGSMM*

To generate a new mGSMM, the most comprehensive hGSMMs, HMR2 and Recon3D, as well as an existing mGSMM, MMR, were used as reference models to generate four intermediate models (IMs) which were subsequently integrated into a single network. IM1, reconstructed from directly replacing genes in HMR2 with its known mouse orthologs, comprised 8160 metabolic reactions, a similar number to the 8181 reactions in HMR2 . IM2 and IM3 reconstructed from HMR2 and Recon3D, respectively, using protein sequence similarity comprised of 5913 and 7235 reactions, less than IM1. The metabolite identifiers of IM1-3 were renamed to KEGG identifiers using a metabolite map constructed. Metabolite identifiers in MMR was also renamed to yield IM4 (Figure 3.1). Only 60.5% of genes and 41.9% of metabolites overlapped between IM1-4, indicating significant non-redundancy between IMs (Figure 3.2). These four networks were sequentially integrated in a step-wise manner to yield IM5 that comprised 10887 metabolic reactions. Duplicate reactions, metabolites and genes resulting from the merging of IMs were removed. In a metabolic network represented as reaction and metabolite nodes, all nodes should be connected to a single component to represent the step-by-step chemical conversion of metabolites. Therefore, reconstructing a metabolic network as a bipartite graph of reactions connected to metabolites allows the identification of subsets of reactions not connected to the primary network. A total of 647 reactions that were either unconnected to a primary network or that were elementally imbalanced were removed from IM5. The resulting network, MMRN consisted of 10349 metabolic reactions and 3461 genes, is elementally balanced and connected to a single network component (Figure 3.1).

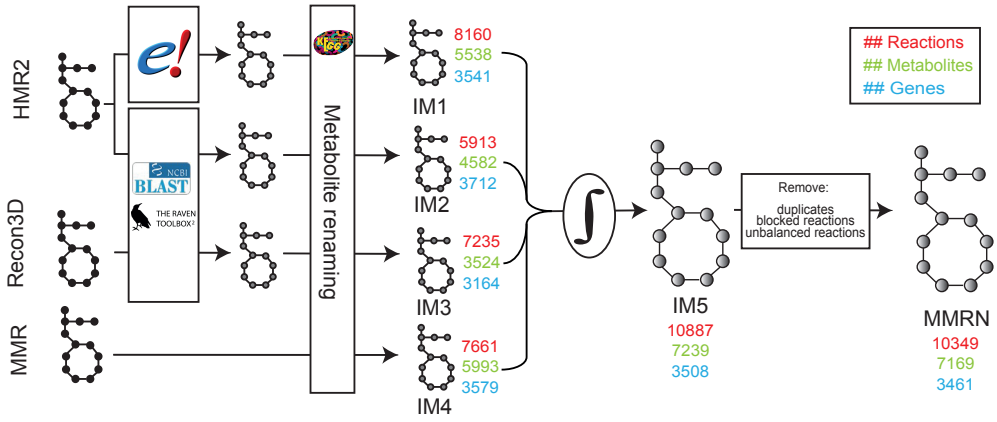


Figure 3.1: **Reconstruction of MMRN using an orthology-based approach.** The hGSMMs Recon3D and HMR2 were used a templates for the reconstruction of three orthology-based reconstructions (IM1-3). These networks were integrated together with MMR (IM4)to a single network of mouse metabolism that were computationally tested to be functional, elementally balanced and connected to single network component.

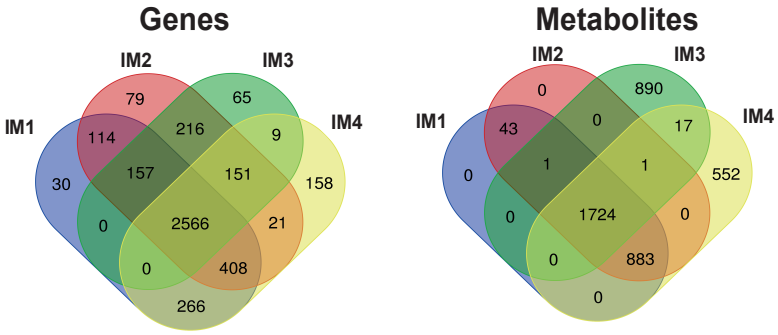


Figure 3.2: **The number of genes and metabolites in IM1-4 used to reconstruct MMRN**.

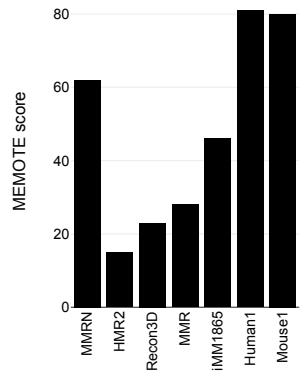


Figure 3.3: **MEMOTE score for MMRN in comparison with other mGSMMs and hGSMMs.**

3.1.2 *Evaluation and benchmarking of MMRN*

Assessing the completeness and robustness of GSMMs has been shown to be difficult [136, 137]. MMRN was therefore benchmarked against published mGSMMs and hGSMMs. In the first instance, the number of reactions, metabolites and genes in MMRN were comparable to those in other mouse GSMMs (Table 3.1). MMRN comprised fewer orphan reactions, reactions with no gene association, compared to Mouse1 and iMM1965, contained no chemically imbalanced reactions and was connected to a single component, while all other mGSMMs and hGSMMs failed these assessments. In an effort to standardise GSMM reconstructions and allow cross-validation between models, MEMOTE has recently been published as a software suit to systematically evaluate GSMMs. MMRN scored 62% which was higher than the three models used for its reconstruction, HMR2, Recon3D and MMR. iMM1865, published in 2020 after we initiated our study scored 46%. The only two models that scored higher than MMRN was Human1 and Mouse1 that was reconstructed from Human1, both recently published. MMRN was therefore an improvement on previously available mGSMMs and hGSMMs. Continued effort to improve MMRN will likely increase its comparison against more recently published mammalian models.

Table 3.1: Key metabolic network attributes for MMRN compared to other published mGSMMs and hGSMMs

	total reaction #	orphan reaction #	imbalanced reaction # (missing info #)	metabolite #	gene #	component #
mGSMMs						
MMRN	10349	2913	0 (515)	7169	3461	1
Mouse1	11413	5052	71 (8)	8382	3513	66
iMM1865	8796	4681	266 (1)	5839	1865	10
MMR	7661	2135	2 (7841)	5993	3579	7
hGSMMs						
Human1	11423	5060	66 (9)	8400	3626	64
HMR2	7681	2107	175 (266)	6015	3820	6
Recon3D	11651	5867	803 (1)	8399	3697	24

The metabolic subsystem associated with each reaction in each mGSMM and hGSMM was used to calculate the number of reactions in different major metabolic pathways (Figure 3.4). Overall, MMRN has a similar distribution of reactions across different pathways compared to other GSMMs and most comparable to the most recently published Mouse1, Human1 and Recon3D. MMRN has an increased number of reactions involved in lipid, amino acid, carbohydrate and nucleotide metabolism compared to iMM1865 and MMR.

3.2 CONCLUSION

In this Chapter I reconstructed a new mGSMM using an orthology-based approach. Genomic sequencing made it possible to computationally reconstruct metabolism for several organisms, including human, using bottom up reconstruction approaches [108]. This process is, however, time-consuming and arduous as it involves building a metabolic reaction network step-by-step from the bibliome [74]. Orthology-based methods have been used to circumvent several of the laborious tasks in this process to reconstruct mGSMMs from hGSMMs (see examples

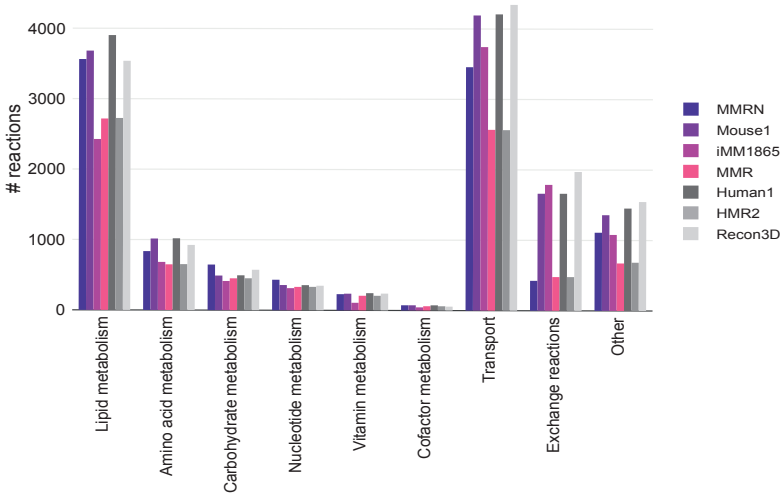


Figure 3.4: **Number of reactions for different metabolic pathways in MMRN compared to other mGSMMs and hGSMMs.**

[97, 133, 138]). This method is, however, limited by the coverage of mouse-to-human orthologs described in orthology databases. MMRN was reconstructed using a methodology that addressed these two limitations when orthology was used in previous mGSMMs, namely, using only a single hGSMM template and a fewer number of mouse-to-human orthologs.

Firstly, HMR2 and Recon3D were used for the reconstruction of MMRN. Using a template model from either one of these hGSMMs series, as in previous orthology-based mGSMMs, limits the reactions included in the mGSMM since there is not a perfect overlap in reactions in these two model series [92]. The different formulation of metabolite and reaction identifiers in Recon and HMR prevents the integration of these two models by combining their respective S-matrices. Renaming metabolites identifiers in IMs reconstructed here allowed merging the S-matrices of these networks and to include reactions from both HMR2 and Recon3D in MMRN.

Newly identified mouse-to-human orthologs were used together with known orthologs. It is clear that the completeness of the mGSMM depends on the set of orthologs when replacing genes in a template hGSMM. IMs from different hGSMMs and from protein homology show non-redundancy which indicate

that using this approach increased the coverage of genes and reactions included in MMRN.

Benchmarking of newly reconstructed GSMMs remains a challenge in systems biology. To date, MEMOTE remains the only benchmarking tool for this purpose and is still limited in its capability because of heterogenous nature of nomenclature and formulations used in GSMMs. For example, tests such as blocked reactions, dead-end metabolites and biomass production flux do not contribute to the overall MEMOTE score, which are important features of robust models. It would be beneficial to further test the validity of MMRN by performing, for example, gene deletions and comparing this against literature. However, because MMRN was not reconstructed using a bottom-up approach but rather from two published hGSMMs with the ultimate goal to reconstruct context-specific liver models, its improvement on other models at the time, and its ability to perform common metabolic tasks were sufficient for this purpose.

4

CONTEXT-SPECIFIC LIVER GSMMS REVEAL DETERMINANTS OF OBESOGENIC DIET METABOLISM IN LIVER CANCER

"*Every good scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is*" - Karl Popper

Cellular metabolic activities are predominantly determined by nutrient availability and enzyme expression [139]. The expression of metabolic enzymes is regulated by various mechanisms that operate at a cellular, tissue and organism level. Diet has been shown to have profound effects on body homeostasis and disease progression [140–142]. It is therefore of clinical interest to understand the interaction between dietary nutrient content and gene expression to get insight into how diet modifications might be used to alter the course of disease.

Liver cancer can be promoted by 'western-style' diets (WD) that are rich in fat and processed sugars, such as fructose [115, 143]. The chronic consumption of WD can lead to systemic dysregulation of insulin signalling and lipid metabolism; the ensuing inflammation and tissue damage are thought to promote mutations that lead to oncogenic transformation in hepatocytes [144]. This systemic dysregulation as well as the nutrients in WD can influence metabolic gene expression in hepatocytes. Concomitantly, metabolic rewiring in hepatic tumours is distinct from surrounding tissue and supports tumour growth and survival. It is, however, unclear how diet-induced chronic changes in gene expression in hepatic tissues influence the fuelling of liver cancer metabolism by dietary nutrients. It is also unclear whether modulation of diet composition, alone, suffices to relieve the metabolic features that promote tumour growth.

To investigate the interplay between diet composition and gene expression requires a global survey of metabolic activities because dietary nutrients simultaneously fuel multiple highly interconnected metabolic pathways. Despite the advances in *in vivo*

tracing of metabolism, the measurement of metabolic fluxes on a global scale remains challenging [145, 146]. Constraint-based modelling (CBM) offers a suitable computational framework to constrain nutrient input and expression-based intracellular fluxes to investigate the aforementioned questions. The effect of diet on metabolism has previously been studied using CBM, however, there has been no systematic analysis of how diet composition itself influences metabolic fluxes of genome scale metabolic models (GSMMs) under specific gene expression backgrounds.

In this Chapter, I used Mouse Metabolic Reaction Network (MMRN) to investigate how dietary nutrients and gene expression changes associated with chronic exposure to WD combine to influence metabolism in the liver of mice with liver cancer using a CBM framework. For this purpose, experimental data from the DEN mouse model were used together with MMRN to reconstruct a liver-specific GSMM, $MMRN^{Hep}$, that was then constrained with gene expression data and the content of a WD or a control diet (CD).

4.1 RESULTS

4.1.1 *Changes in expression of genes that control metabolic processes by a tumour-promoting western diet*

To study the interplay between diet composition and diet-induced changes in gene expression in the context of hepatic tumour development, the DEN mouse model was used with a non-injected (nonDEN) control as well as a cohort of mice that were fed a CD (Figure 4.1). In both weight and energy, the CD consisted of less lipids and carbohydrates but similar amino acids compared to the WD. Sucrose was the primary carbohydrate source in the WD and starch in the CD (Table 4.1). Consequently, the WD had increased fructose content compared to the CD.

Mice consumed either diet at similar daily rates (Figure 4.2A), however, more tumours were detectable by 25-29 weeks of age in DEN^{WD} compared to DEN^{CD} mice (Figure 4.2B). Between 30 to 36 weeks of age, tumour burden increased significantly in DEN^{WD} mice compared to DEN^{CD} mice ($p = 0.02$, paired t-test), and by 39 weeks 55% of DEN^{CD} and 100% of DEN^{WD} mice

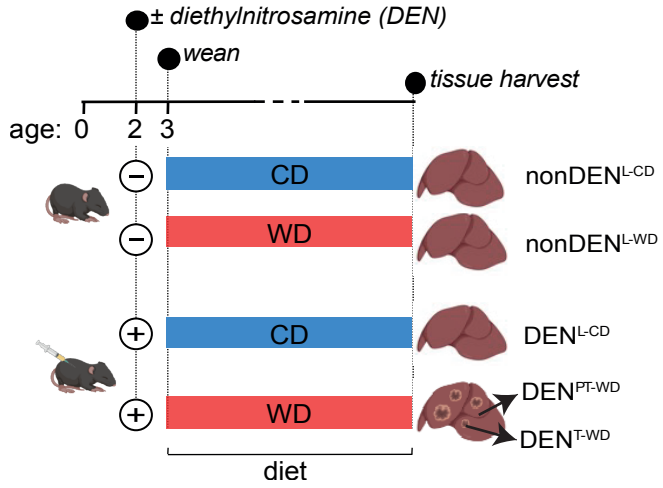


Figure 4.1: **Experimental design for the DEN mouse model with a non-injected and dietary control.** Details for the DEN mouse model are described in the text. In short, mice were injected at 2 weeks of age whereafter they were fed on either a WD or CD. L: liver, PT: peritumoral tissue, T: tumour tissue.

Table 4.1: Contributions of major dietary components (amino acids, carbohydrates and lipids) to western (WD) and control (CD) diet composition.

	Weight		Energy			
	WD	CD	WD		CD	
	%	%	%	kcal/g diet	%	kcal/g diet
Amino acids	17.4	15.5	15.5	0.697	15.9	0.618
Carbohydrates	49.9	70.1	44.4	1.995	72.2	2.806
Sucrose	34.05	10.95				
Starch	5	43.51				
Lipids	20	5.1	40.1	1.804	11.9	0.463

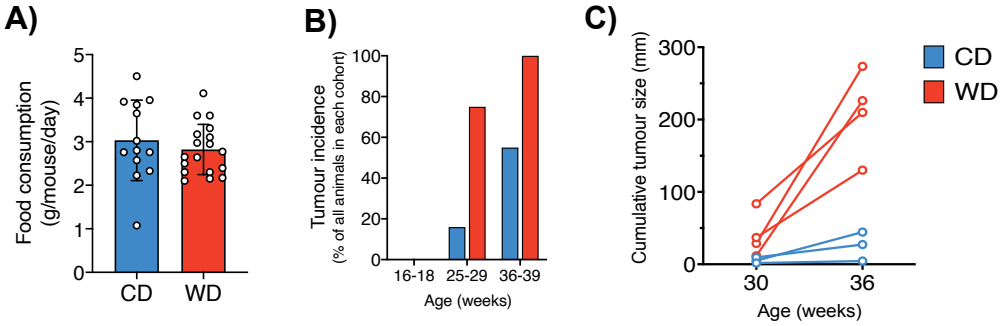


Figure 4.2: **A WD amplifies mouse liver cancer development.** A) Food consumption measurements for mice fed a CD or WD. This was used to calculate $C_{moles}{}^{DIET}$ (see chapter 2 for details). B) The tumour incidence for DEN-injected mice fed either a CD or WD at different age intervals. At time of tissue harvesting, 36-39 weeks of age, all WD mice had tumours compared to 55% of mice with tumours on a CD. C) Cumulative tumour size of three DEN^{CD} mice and four DEN^{WD} mice at 30 and 36 weeks of ages measured by magnetic resonance imaging (MRI). Tumour burden in DEN^{WD} mice increased significantly over time (paired t-test, p-value < 0.05).

had tumours (Figure 4.2B, C). These data confirmed previously observed tumour-promoting effects of WD relative to CD in the DEN mouse model [115]. Tumours in DEN^{CD} mice were too small to reliably separate from peritumoral tissue, and because processes associated with ageing may convolute comparison of heterochronous tumours, DEN^{CD} mice were not aged further in order to obtain resectable tumours. To assess gene expression changes caused by diet, carcinogen and tumour development, transcriptional profiles of available tumours and liver tissues using RNA sequencing were analysed.

Principal component analysis (PCA) of the gene expression profiles for each tissue sample revealed two major PCs: PC1 (accounting for 41% of variance) was associated with effects of DEN and PC2 (13%) with diet (Figure 4.3A). GO analysis of differentially expressed genes across all pair-wise comparisons revealed enrichment in processes related to (a) inflammation, which were broadly linked to both diet and DEN; (b) cellular proliferation and cell-cell interactions, primarily associated with tumour or peritumoral tissues, and (c) metabolic processes, sev-

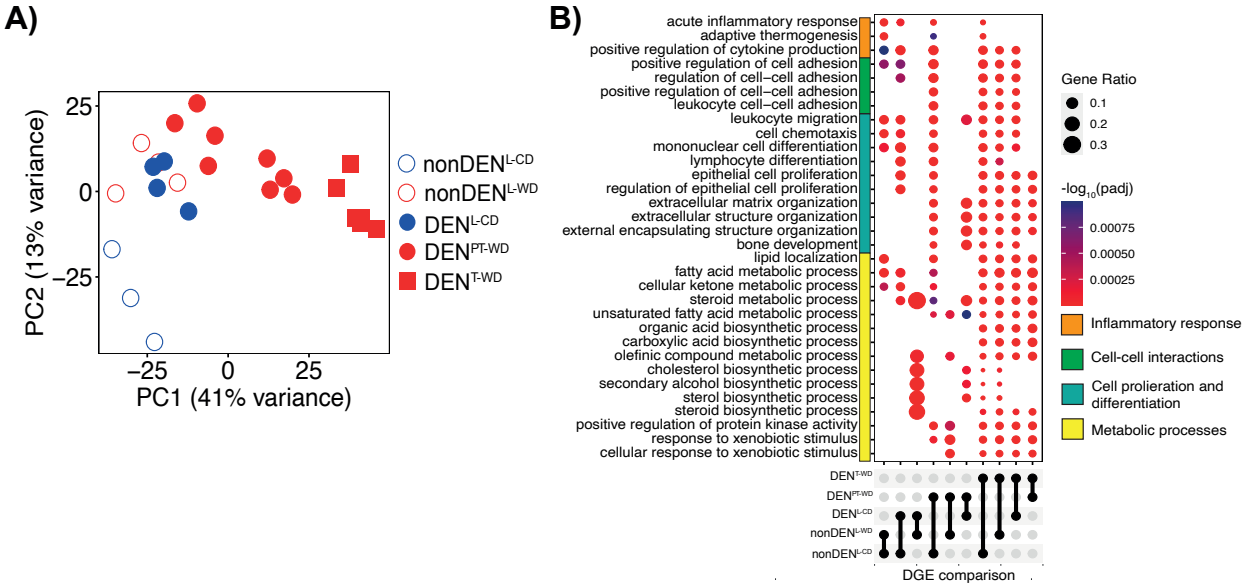


Figure 4.3: **A WD elicits distinct hepatic gene expression profile.** A) PCA of gene expression data derived from RNA-sequencing analysis of tissue samples from cohort 1 (Figure 4.1). B) GO biological process overrepresentation test for differentially expressed genes using the *enrichGO* function and visualised with the *dotplot* function from the *clusterProfiler* package. For each comparison, the bottom condition is used as baseline. Benjamini-Hochberg correction was used with a q-value cut-off of 0.01 and is represented by dot colour. Dot size represents the fractional number of genes enriched within a particular biological process compared to the total gene set size. L: liver tissue, PT: peritumoral tissue, T: tumour tissue

eral of which were related to lipid metabolism and emerged, at varying degrees, as a function of diet, DEN or the transformation state of the tissue (Figure 4.3B). These observations suggest that chronic exposure to diet is associated with changes in expression of genes that mediate metabolic processes which, in addition to dietary nutrient availability, could influence both the rate and functions of tissue metabolism.

4.1.2 *csGSMMs selectively take up nutrients for biomass production*

To model the contribution of a WD in providing nutrients and impacting gene expression liver and tumour metabolism, a gen-

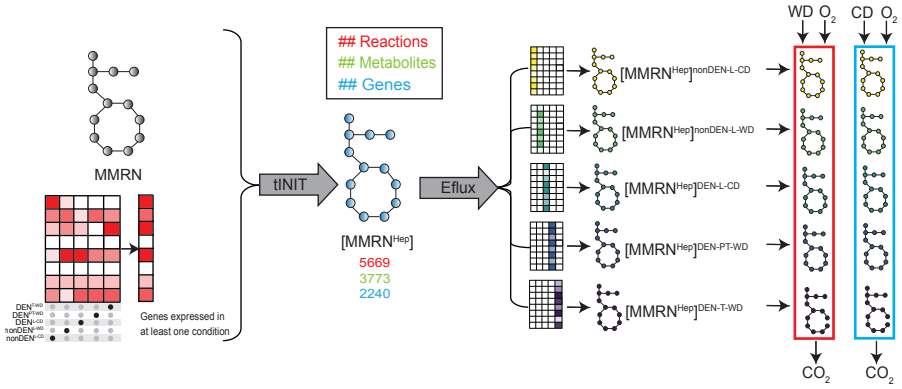


Figure 4.4: **Reconstruction of MMRN^{Hep} and csGSMMs.** MMRN was used with gene expression from cohort 1 as input for the tINIT algorithm to generate a hepatic GSMM, MMRN^{Hep}. This model was then further constrained using the Eflux method that imposes flux boundaries on individual reactions using context-specific gene expression. O_2 consumption and CO_2 production were used to constrain uptake and production bounds, respectively, and finally, the contents of the WD and CD were used to constrain the uptake of these models. L: liver tissue, PT: peritumoral tissue, T: tumour tissue

eric hepatic GSMM, MMRN^{Hep} was generated (Figure 4.4, see Methods for details). MMRN^{Hep} consisted of 5669 metabolic reactions associated with 2240 genes. MMRN^{Hep} was constrained using each experimental condition's gene expression as well as O_2 consumption and CO_2 production rates of mice measured in metabolic cages to generate five context-specific GSMMs (csGSMMs). The exchange reactions of these csGSMMs were then constrained using either the content of the WD or the CD to test how diet alter the flux distribution under different conditions (Figure 4.4).

The total flux of carbons (C_{moles}^{DIET}, Figure 4.5A) that were available from the WD was higher than that for CD. Accordingly, the flux of carbons taken up (C_{moles}^{INFLUX}) by WD-fed csGSMMs was higher than that of CD-fed csGSMMs, and higher than MMRN^{Hep} fed with either diet. However, no difference in $C_{moles}^{INFLUX}$was found between $[MMRN^{Hep}]_{WD}$ and $[MMRN^{Hep}]_{CD}$. Furthermore, C_{moles}^{INFLUX} values were lower than C_{moles}^{DIET} for both diets in all models, and the relative C_{moles}^{INFLUX} of individual diet components did not reflect the

corresponding C_{moles}^{DIET} values for those nutrients (Figure 4.5B). Together, these observations suggest that $MMRN^{Hep}$ limits the amounts and types of nutrients it takes from the diet for optimal production of biomass, irrespective of condition-specific gene expression; they further indicate that gene expression together with dietary composition, rather than nutrient availability alone, dictate nutrient uptake of models.

Consistent with a carbon-balanced model, increased C_{moles}^{INFLUX} in WD compared to CD was mirrored by higher total C_{moles}^{EFFLUX} values for all models (Figure 4.5A). An increased C_{moles}^{EFFLUX} by WD-fed models were accounted for by an increased production of glycerol and succinate since all other metabolites produced ($C_{moles}^{EFFLUX\text{-}others}$) were consistent across all csGSMMs. This prediction was confirmed by metabolomics analyses of tissues, which revealed increased levels of glycerol and succinate in both DEN^{T} and DEN^{PT} tissues (Figure 4.6A, B).

Development of Systematic Diet Composition Swap (SyDiCoS) as an approach to determine the impact of a nutrient class on metabolic flux

To determine whether particular dietary nutrient classes influenced C_{moles}^{EFFLUX}, the amount of each dietary component class in WD was swapped with the respective amount of that component in CD while keeping the remaining WD composition unchanged. The notation $WD^{X(CD)}$ refers to diets where X=WD diet component that was swapped to its respective CD value. The quantitative increase in glycerol production by WD was abrogated with $WD^{carbs(CD)}$(Figure 4.6C), while $WD^{lipid(CD)}$revealed a dependence of succinate production on the increased lipid content of the WD, which drove higher TCA cycle flux (Figure 4.6C). Further investigation of the flux distributions with $WD^{carbs(CD)}$ showed that WD-derived fructose accounted for increased glycerol production (Figure 4.6D). However, when used as the sole dietary sugar, fructose was diverted to sustain glycolysis, and glycerol production ceased. Investigating the flux distributions by reconstructing a network revealed the metabolic pathways responsible for increased glycerol and succinate production in WD using SyDiCoS and are depicted in Figure 4.6E.

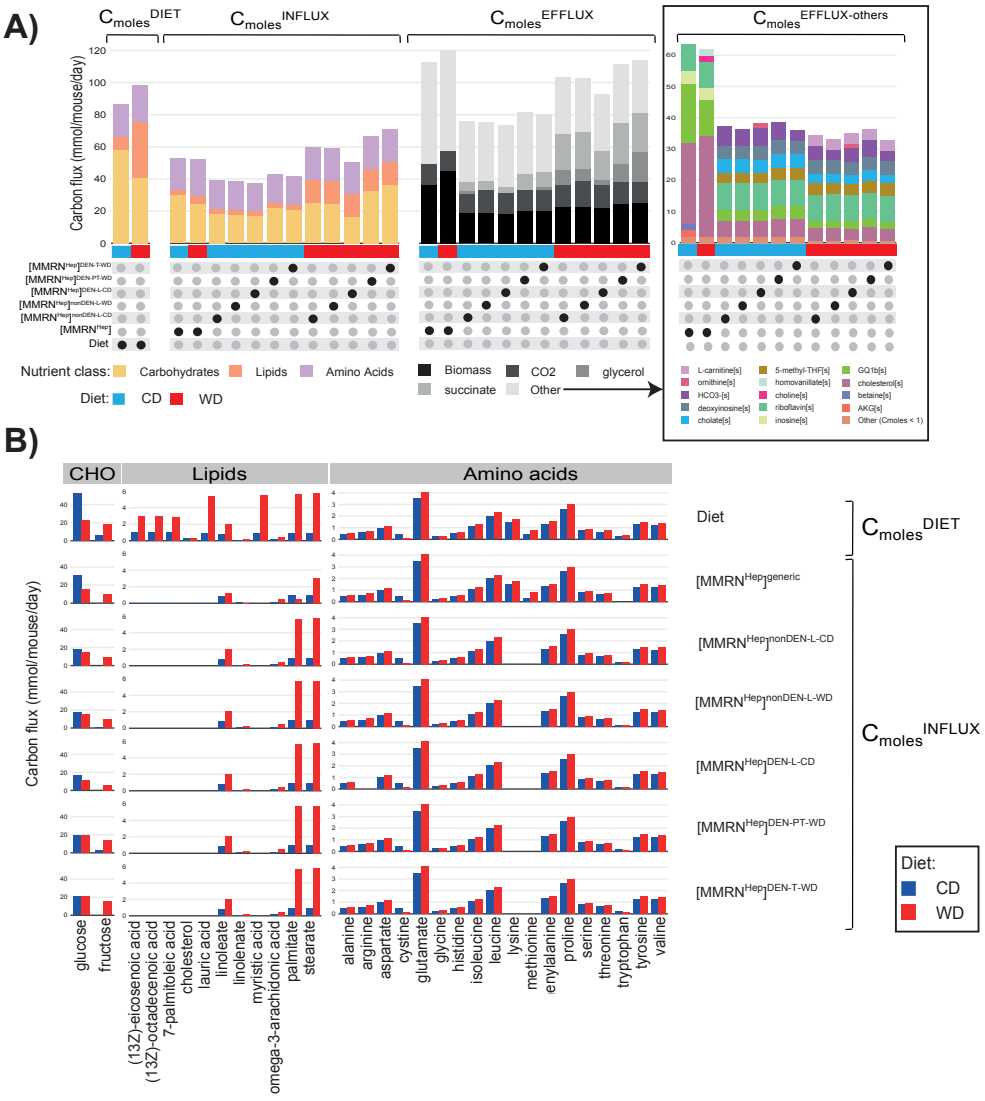


Figure 4.5: **Dietary nutrient uptake and metabolite production by csGSMMs.** A) C_{moles}^{DIET} and C_{moles}^{INFLUX}values of each of the major dietary nutrient classes (carbohydrates, lipids and amino acids) in CD and WD for $MMRN^{Hep}$ and csGSMMs. C_{moles}^{EFFLUX} for $MMRN^{Hep}$ and csGSMMs show that increased efflux in WD models is a result of increased glycerol and succinate production since $C_{moles}^{EFFLUX\text{-}others}$are consistent across all csGSMMs. B) C_{moles}^{DIET} (top) and C_{moles}^{INFLUX} for individual dietary nutrients that belong to the three major dietary component classes (carbohydrates, lipids and amino acids) for CD and WD.

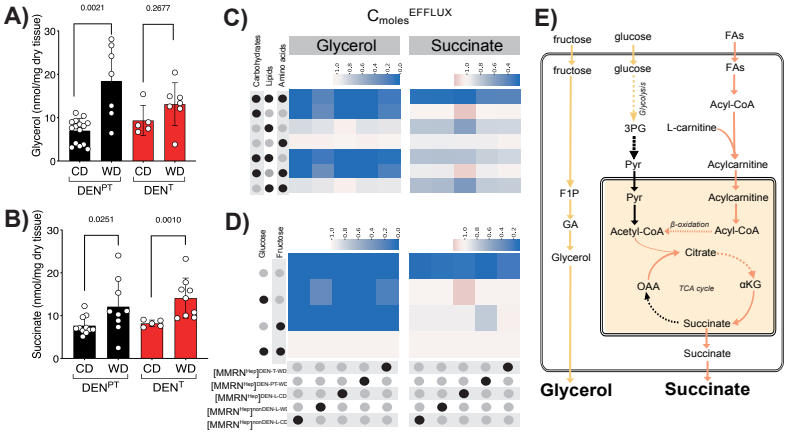


Figure 4.6: **Increased production of glycerol and succinate as a result of carbohydrate and lipid content in WD.** Amounts of glycerol (A) and succinate (B) in DEN^{PT} and DEN^{T} tissues from mice fed WD compared to respective tissues from CD-fed mice measured by GC-MS. P-values were calculated with Mann-Whitney U test and depicted on bars. C) SyDiCoS to assess the role of WD components upon glycerol and succinate production flux in csGSMMs. $C_{moles}{}^{DIET}$ values for all three major diet component classes (carbohydrates, lipids and amino acids) in the WD were swapped individually or in combination with the corresponding $C_{moles}{}^{DIET}$ values in CD while leaving the remaining dietary $C_{moles}{}^{DIET}$ values of the WD unaltered. The swapped component(s) are indicated by black dots on the left. The colour scale represents the ratio of glycerol production flux or succinate production flux in models provided with the swapped diet relative to the respective fluxes in models provided with WD, calculated for each csGSMM shown at the bottom of panel (D). D) Assessment of the role of glucose or fructose from WD on glycerol and succinate production. csGSMMs shown at the bottom were provided WD containing only glucose or fructose (using their respective $C_{moles}{}^{DIET}$ values found in WD) as indicated by the black dots on the left, or both sugars (equivalent to the original WD composition) while leaving the $C_{moles}{}^{DIET}$ values for lipids and amino acids in WD unaltered. E) Metabolic pathways that lead to increased production of glycerol and succinate in WD from fructose and FAs, respectively, derived from inspection of the flux distributions of csGMMS under various SyDiCoS conditions (panels C, D). FAs: fatty acids; 3PG: 3-phosphyglycerate; Pyr: pyruvate; F1P: fructose 1-phosphate; GA: glyceraldehyde; AKG: α-ketoglutarate; OAA: oxaloacetate.

4.1.3 *Gene expression and dietary nutrient availability dictate differential fate of FAs in tumours and peritumoral liver*

WD led to a modest increase of biomass production in all csGSMMs compared to the respective CD-fed GSMMs (Figure 4.5A) demonstrating that, while csGSMMs are robust and withstand a massive overhaul of diet composition, they are also sensitive enough to detect the resulting flux changes. However, the flux distributions of all csGSMMs differed from each other, with $[MMRN^{Hep}]_{CD}^{DEN-T-WD}$ being the most distinct (Figure 4.7A). Furthermore, while WD shifted the flux distributions of all models, it amplified the differences between the distributions of the tumour and that of non-tumour models more than for other model comparisons (Figure 4.7A). These observations indicated that similar increases in biomass production in tumour and non-tumour tissues induced by WD are associated with distinct metabolic pathway activities. Henceforth, to simplify further exploration of such pathways, the focus was on the comparisons between $[\text{MMRN}^{\text{Hep}}]^{\text{DEN-T-WD}}$ and $[\text{MMRN}^{\text{Hep}}]^{\text{DEN-PT-WD}}$.

WD caused a greater shift, relative to CD, in the flux distributions of $[\text{MMRN}^{\text{Hep}}]^{\text{DEN-T-WD}}$ than those of $[\text{MMRN}^{\text{Hep}}]^{\text{DEN-PT-WD}}$ indicating that the tumour metabolic network is primed for a greater response to WD than the peritumoral model (Figure 4.7A, B). Notably, the WD-induced increase in tumour/non-tumour divergence was completely reversed only when both lipids and carbohydrates in WD were replaced with their respective CD content (Figure 4.7A, C).

To explore specific pathways that underlie differential response of tumour and peritumoral tissue models to WD driven by carbohydrates and lipids, all metabolic subsystems that carried flux in $[MMRN^{Hep}]_{WD}^{DEN-T-WD}$ and $[MMRN^{Hep}]_{WD}^{DEN-PT-WD}$ were identified. Further analysis revealed those subsystems that were differentially engaged in the tumour and peritumoral tissue models; among them were several subsystems involved in lipid and carbohydrate metabolism (Figure 4.8).

A network comprised of all the reactions in the subsystems related to lipid and carbohydrate metabolism that showed differential flux exclusively in the T or PT in WD but not in CD was reconstructed. From this network, a single, fully connected

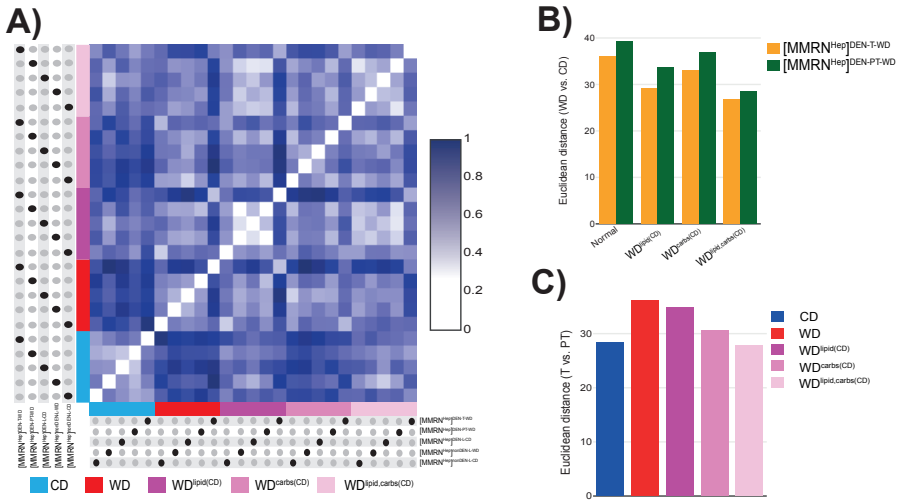


Figure 4.7: **Euclidean distances of flux distributions**. A) Effects of diet composition on flux distribution differences between csGSMMs assessed by SyDiCoS. FBA was used to calculate the flux distribution for each csGSMM provided with WD, CD, $WD^{lipids(CD)}$, $WD^{carbs(CD)}$ and $WD^{lipids,carbs(CD)}$. The colour scale represents the Euclidean distance values calculated in a pairwise manner between each of the flux distributions and plotted relative to the maximum distance value across all comparisons. B) Relative response to changes in diet composition of the flux distributions of tumoral or peritumoral models. Absolute Euclidean distances (from panel A) for either $[MMRN^{Hep}]^{DEN\text{-}T\text{-}WD}$ or $[MMRN^{Hep}]^{DEN\text{-}PT\text{-}WD}$ under different SyDiCoS conditions are plotted. C) Effect of changes in diet composition on the flux distributions differences between tumoral and peritumoral models. Absolute Euclidean distances (from panel A) between $[MMRN^{Hep}]^{DEN\text{-}T\text{-}WD}$ or $[MMRN^{Hep}]^{DEN\text{-}PT\text{-}WD}$ under different SyDiCoS conditions are plotted.

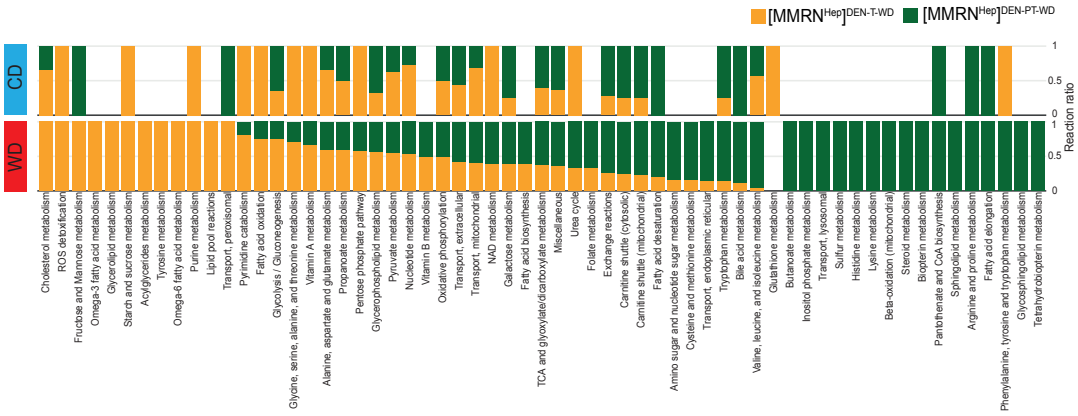


Figure 4.8: **Metabolic subsystems that include at least one reaction that carries flux in [MMRN$^{\text{Hep}}$]$^{\text{DEN-T-WD}}$ or [MMRN$^{\text{Hep}}$]$^{\text{DEN-PT-WD}}$ on either WD or CD.** In each of these subsystems, the proportion of reactions with higher flux in [MMRN$^{\text{Hep}}$]$^{\text{DEN-T-WD}}$ compared to [MMRN$^{\text{Hep}}$]$^{\text{DEN-PT-WD}}$ on either WD or CD is plotted. For each diet, a reaction ratio=1 for [MMRN$^{\text{Hep}}$]$^{\text{DEN-T-WD}}$ in a given subsystem indicates that all reactions in that subsystem have higher flux compared to [MMRN$^{\text{Hep}}$]$^{\text{DEN-PT-WD}}$

subnetwork connected to biomass production was extracted and the flux differences between $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ and $[MMRN^{Hep}]^{DEN-PT-WD}_{WD}$, and between $[MMRN^{Hep}]^{DEN-T-WD}_{CD}$ and $[MMRN^{Hep}]^{DEN-PT-WD}_{CD}$ were calculated (Figure 4.9).

Comparison of the reaction fluxes in this network revealed a distinct diet-dependent fate of dietary fatty acids (FAs) in the tumour and the peritumoral tissue models (Figure 4.9A,B). On one hand, FA uptake and FA-derived acyl-CoA synthesis fluxes were higher in WD than in CD for both the T and PT models, and equal between T and PT models in each diet. However, the models predicted increased use of FA-CoA for lipid macromolecule synthesis and decreased use for β-oxidation through the mitochondrial carnitine shuttle in $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ compared to $[MMRN^{Hep}]^{DEN-PT-WD}_{WD}$ (Figure 4.9A). These flux differences correlated with higher expression of lipid synthesis genes and lower expression of mitochondrial lipid transport and β-oxidation genes in tumours (Figure 4.10). Consistent with this observation, mitochondria purified from tumour tissue respired

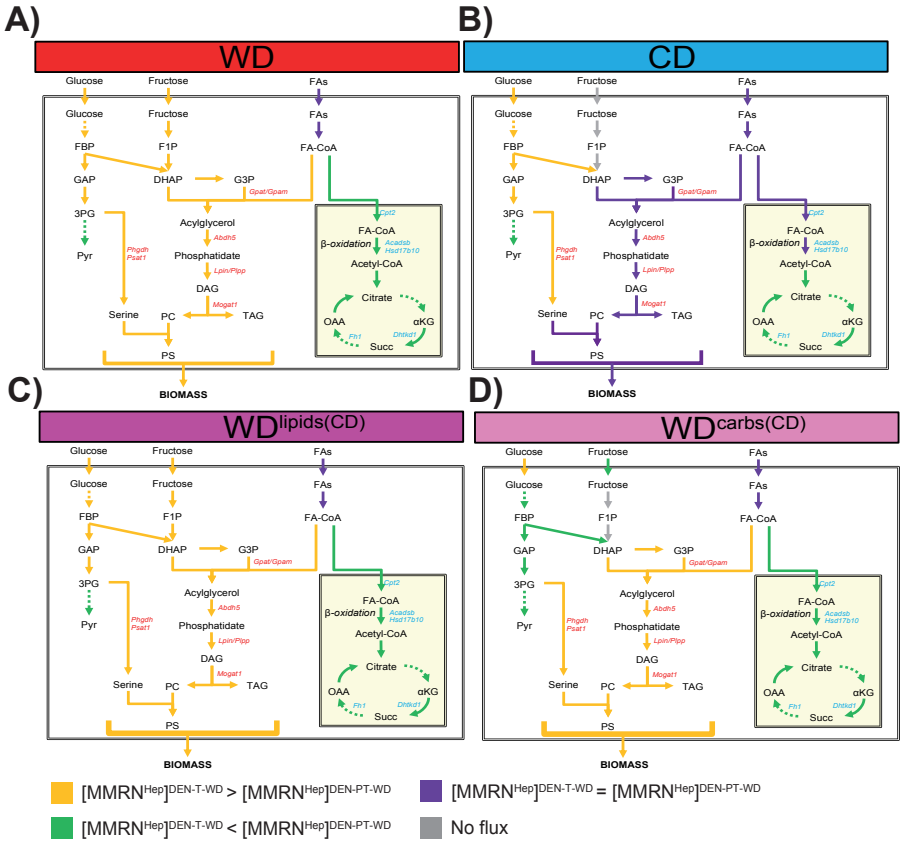


Figure 4.9: **Differential fate of FAs in tumours and peritumoral liver predicted by MMRN^{Hep}**. Flux differences between tumoral and peritumoral models fed either WD (A), CD (B), WD$^{lipid(CD)}$(C), WD$^{carbs(CD)}$(D). These networks are schematic representations of the metabolic network reconstructed from all the reactions of subsystems from panel (D) that have a reaction ratio=1 (either all reactions that carry higher flux in [MMRN^{Hep}]$^{DEN\text{-}T\text{-}WD}$ or in [MMRN^{Hep}]$^{DEN\text{-}PT\text{-}WD}$ and partake in lipid and carbohydrate metabolism. Differential fluxes for T and PT are coloured according to the legend on the bottom of these panels. FBP: fructose 1,6-bisphosphate; GAP: glyceraldehyde 3-phosphate; 3PG: 3-phosphoglycerate; Pyr: pyruvate; F1P: fructose 1-phosphate; DHAP: dihydroxyacetone phosphate; G3P: glycerol 3-phosphate; DAG: diacylglycerol; TAG: triacylglycerol; PC: phosphatidylcholine; PS: phosphatidylserine; FAs: fatty acids; FA-CoA: fatty acyl-CoA; AKG: a-ketoglutarate; Succ: succinate; OAA: oxaloacetate; Phgdh: Phosphoglycerate dehydrogenase; Psat1: Phosphoserine aminotransferase 1; Gpat: Glycerol-3-phosphate acyltransferase; Gpam: Glycerol-3-phosphate acyltransferase 1, mitochondrial; Abdh5: 1-acylglycerol-3-phosphate O-acyltransferase; Lpin: Phosphatidate phosphatase; Plpp: Pyridoxal phosphate phosphatase; Mogat1: Monoacylglycerol O-acyltransferase 1; Cpt2: Carnitine palmitoyltransferase 2; Acadsb: Acyl-CoA dehydrogenase short/branched chain; Hsd17b10: Hydroxysteroid 17-beta dehydrogenase 10; Dhtkd1: Dehydrogenase E1 and transketolase domain containing 1; Fh1: Fumarate hydratase 1.

significantly less on FAs than mitochondria purified from peritumoral tissue (Figure 4.11A).

Notably, both $[MMRN^{Hep}]^{DEN-T-WD}_{CD}$ and $[MMRN^{Hep}]^{DEN-PT-WD}_{CD}$used FA-derived acyl-CoA equally for either biomass or β-oxidation, whereas differential fate of acyl-CoA between $[MMRN^{Hep}]^{DEN\text{-}T\text{-}WD}$ and $[MMRN^{Hep}]^{DEN\text{-}PT\text{-}WD}$ persisted when these models were fed either $WD^{lipids(CD)}$ or $WD^{carbs(CD)}$(Figure 4.9C, D). This observation suggested that differential expression of β-oxidation and lipid synthesis genes does not suffice to drive divergence of FA metabolic fates in the absence of altered dietary composition of both FAs and carbohydrates.

Flux predictions showed increased fluxes of glycolysis and fructolysis, which contribute precursors for lipid synthesis, in $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ compared to $[MMRN^{Hep}]^{DEN-PT-WD}_{WD}$. Increased serine synthesis from carbohydrates persisted in $[MMRN^{Hep}]^{DEN-T-WD}_{CD}$ compared to $[MMRN^{Hep}]^{DEN-PT-WD}_{CD}$, indicating no impediment in serine synthesis that would explain attenuated lipid synthesis in CD. In contrast, incorporation of glycerol-3-phosphate (G3P) through glycerol 3-phosphate acyltransferase (encoded by *Gpat / Gpam*) into lipids ceased for both models in CD but was increased in $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ compared to $[MMRN^{Hep}]^{DEN-PT-WD}_{WD}$. In agreement with a role for G3P-fuelled esterification in driving increased lipid synthesis in T vs PT, it was found that, while both tumour and peritumoral mouse tissues synthesised FAs at comparable rates *in vivo* (Figure 4.11B, C), tumours showed increased incorporation of newly synthesised glycerol into lipids (Figure 4.11B, D).

4.1.4 *Glycerol is produced through a novel metabolic pathway*

Glycerol production as a result of WD feeding was increased in $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ compared to other WD-fed models (Figure 4.12). Although SyDiCoS revealed that this was primarily a consequence of increased fructose in WD, the model also predicted that glycerol metabolised from glucose was not from known glyceroneogenesis pathways, but used serine as substrate produced *via* the serine biosynthesis pathway (Figure 4.12). This pathway has not previously been described, however, increased

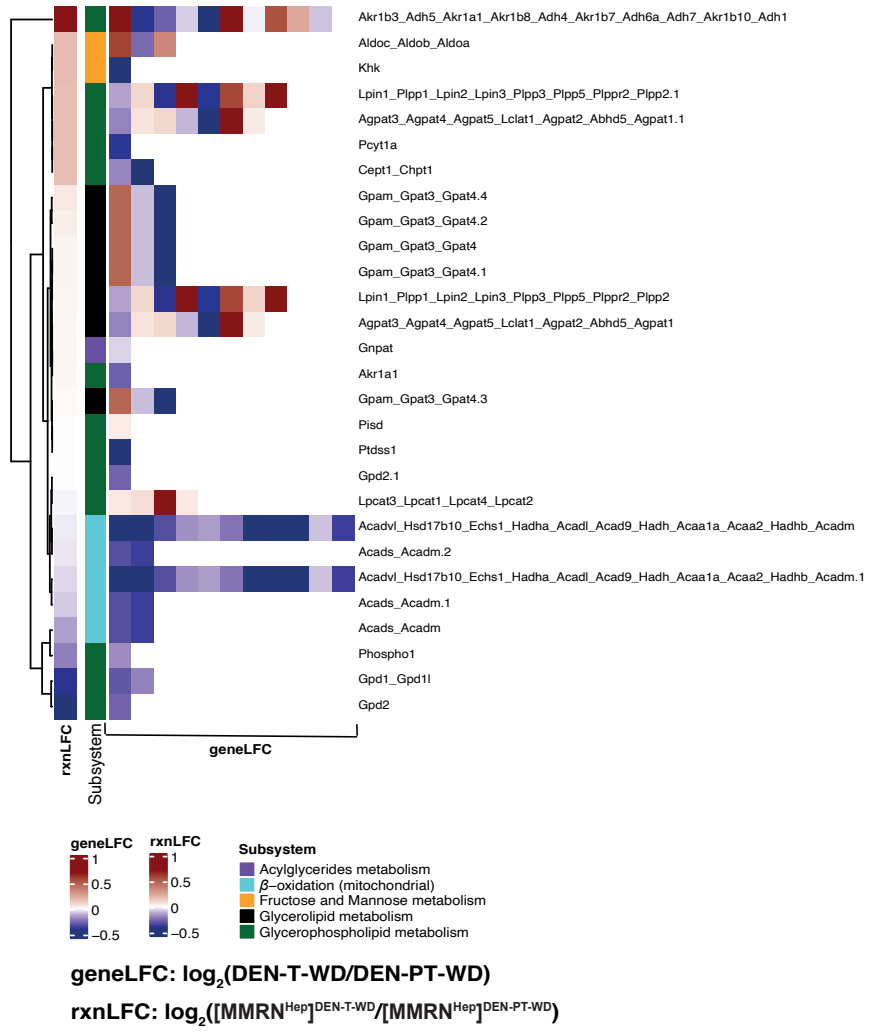


Figure 4.10: **Log_2-fold changes of genes and fluxes for the reactions in Figure 4.9.** The log_2-fold change in the expression of each gene associated with the corresponding reaction is plotted for DEN^T compared to DEN^{PT}.

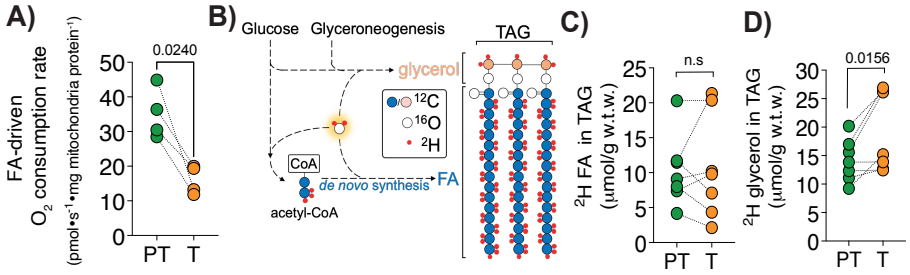


Figure 4.11: **Experimental validation of model predictions.** A) Comparison of FA-driven oxygen consumption rates in mitochondria isolated from liver tumours (T) or peritumoral (PT) tissues. Statistical significance determined by two-tailed paired t-test (n=4 different mice, each providing a paired T and PT tissue sample from which mitochondria were isolated; oxygen consumption was measured in parallel for each T/PT sample pair). B) Schematic showing metabolic routes of ^2H incorporation into the glycerol backbone and fatty-acyl chains in a triglyceride (TAG) molecule after administration of ^2H_2O to mice. C) Measurement of *de novo* synthesised fatty-acids (as outlined in B) in TAGs extracted from tumour and peritumoural tissues. D) Measurement of *de novo* synthesised glycerol (as outlined in B) in TAGs extracted from tumour and peritumoural tissues. Statistical significance in (C) and (D) determined by Wilcoxon matched-pairs signed rank test (n=7 different mice, each providing a paired T and PT tissue sample).

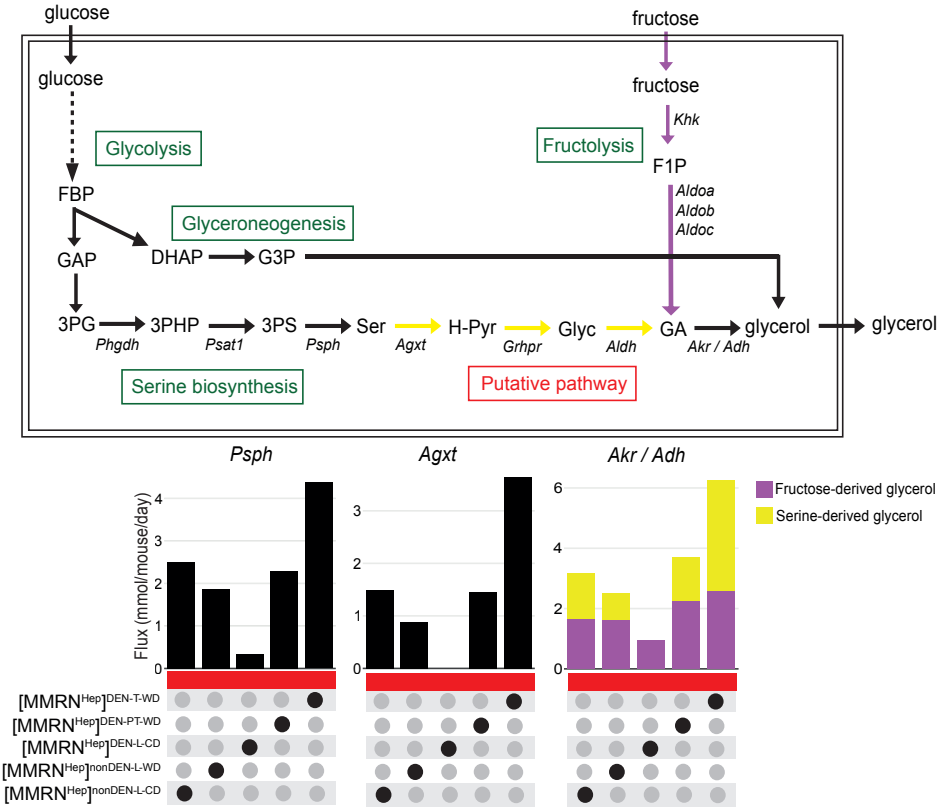


Figure 4.12: **Novel glycerol production pathway identified in MMRN^{Hep}.** Increased serine production in liver tumours has previously been shown and gives validation to the predictive capability of MMRN^{Hep}. Increased serine production in $[MMRN^{Hep}]^{DEN-T-WD}_{WD}$ was predicted to be used for glycerol synthesis by a series of reactions that was previously unknown to produce glycerol.

serine production has been shown in different cancer types, including liver cancer [62, 67].

Serine biosynthesis in cancer is primarily driven by increased expression of genes in this pathway. Genes in this pathway, *Phgdh, Psat1* and *Psph* are upregulated at mRNA level and *Phgdh* is also expressed on protein level in tumour tissue (Figure 4.13). *Agxt* is the first enzyme in the putative pathway responsible for metabolising serine to hydroxypyruvate. At mRNA level, this enzyme has a slight decreased expression in DEN$^{T\text{-}WD}$ compared to DEN$^{PT\text{-}WD}$, albeit non-significant. Interestingly, *Agxt* showed increased protein expression in DEN$^{T\text{-}WD}$. Other reactions in

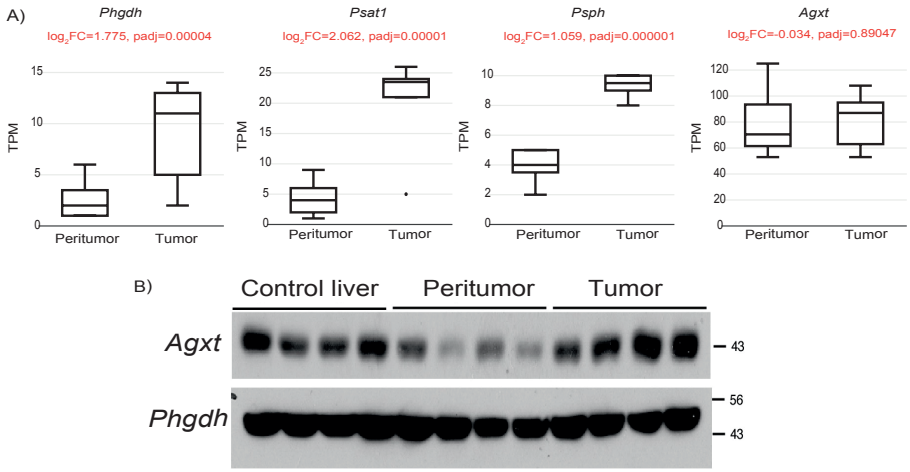


Figure 4.13: **mRNA and protein expression of genes involved in the putative glycerol producing pathway**. A) TPM expression values of *Phgdh, Psat1*, *Psph* and *Agxt* in peritumoral and tumor tissue. Differential gene expression statistics depicted on top of each panel were calculated with DESeq2 as described in Chapter 2. B) Western blot of of protein expression of *Phgdh* and *Agxt* in non-tumor bearing liver tissue (control liver), peritumoral tissue and tumor tissue.

this pathway were annotated to be catalysed by more than one enzyme.

4.2 CONCLUSION

In this Chapter I used MMRN to reconstruct a liver-specific GSMM and used this model to study how a WD impacts liver and liver tumour metabolism. The mathematical formulation of metabolism in GSMMs allows the integration of different data types to constrain the model to biological feasible solutions [80, 81]. Generic organism-specific reconstructions, such as MMRN, are indispensable tools in this process and serve as a starting point for CBM.

An ensemble of tools have been developed to extract cell- and tissue-specific GSMMs from generic reconstructions (reviewed in [147]). Although these methods often differ in their approach, their common goal is to extract a network all the reactions for which there is evidence to take place in a the cell or tissue of interest. Most of these methods therefore rely on gene or protein

expression of enzymes catalysing these reactions and use omics data, such as transcriptomics or proteomics, for this purpose. The integrative network inference for tissues (INIT) algorithm was developed to extract such networks using transcript or protein expression as evidence to include reactions for which there are sufficient expression using the GR-rules embedded in the model [86]. This algorithm was later expanded to include a metabolic task list as input [132]. The primary aim of this updated algorithm, tINIT, is to ensure that the resulting model can perform a predefined set of metabolic tasks that consequently gives the resulting model metabolic functionality. Using tINIT a generic hepatic mouse GSMM, $MMRN^{Hep}$ was reconstructed by using the maximum expression across all experimental conditions used. Then constraining $MMRN^{Hep}$ with condition-specific expression [134] allowed pair-wise comparison of flux distributions across csGSMMs because the same reactions are included in each simulation but with distinct flux distributions.

Using expression data to constrain the reaction bounds of individual reactions constrains the solution space to allow fewer possible solutions for linear optimisation. Methods such as MADE [148] and RMetD [97] rely on differential gene expression between two conditions for this purpose and thus results in a model that is constrained relative to another experimental condition. The Eflux [134] method employed here, however, directly imposes these constraints on individual reactions by using absolute gene expression data. This allowed the reconstruction of five distinct csGSMM each constrained with its corresponding transcript expression.

There are a limited number of studies that have investigated the impact of nutrient input on predicted flux distributions, particularly in mammalian GSMMs [106]. The roadmap developed here using the SyDiCoS approach allows the systematic investigation of the impact of different nutrient classes, and individual nutrients, on metabolic flux. This approach also addresses whether in the background of specific gene expression patterns, imposed by constraints, flux patterns can be impacted by changing nutrient availability. For example, SyDiCoS revealed that gene expression in lipid synthesis pathways sufficed to drive differential lipid metabolism between the tumour and peritumoral tissue when either lipids or carbohydrates were nor-

malised to CD levels, however, in a full swap to CD, differential metabolism ceased. In contrast, increased serine synthesis in tumour relative to the peritumoral tissue persisted in all diet inputs tested but in the absence of lipids, serine did not contribute to lipid biomass more than in peritumoral tissue. Even though higher glucose uptake and flux to DHAP persisted in the tumour with CD, it was not sufficient, alone, to sustain acylglycerol synthesis, even with increased *Gpat* expression, likely because fructose uptake ceased in CD.

There are limitations when using and interpreting FBA results based on the optimisation of a single objective function using LP. In the first instance, this gives only one possible solution for the given constraints and no indication of the statistical robustness of this solution or where this solution falls within the given boundaries based on the constraints. Other methods have been developed to explore the solution space through different sampling algorithms or by using flux variability analysis (FVA) that calculates the lower and upper-bound of each reaction for a given objective. To test this, key model predictions, such as succinate and glycerol production have been tested by sampling the solution space using random sampling that confirmed higher production in a WD. Furthermore, these predictions have also been confirmed by using different solver settings, and by using normal FBA, parsimonious FBA and sparse FBA. Because the primary motivation was to test how the predicted distributions of context-specific models differ under specific nutrients conditions, calculating an Euclidean distance metric provided a single metric to compare several different comparisons whereas using sampling or FVA would require comparison of individual reactions.

5

GUT MICROBIOME COMPOSITION INFLUENCES METABOLISM OF DIETARY NUTRIENTS

"*If an alien visited earth, they would likely take some note of humans, but probably spend most of their time trying to understand the dominant form of life on our planet - microorganisms like bacteria and viruses*" - Nathan Wolfe

There is an increasing appreciation that the gut microbiome plays an important role in the pathogenesis of human disease. Changes in microbial species abundance, or dysbiosis, have been linked to various human diseases, including those with a strong metabolic phenotype, such as liver disease and liver cancer [48, 110, 149]. Diet has been shown to be able to contribute to this dysbiosis which in turn affects disease progression. Furthermore, since dietary nutrients are metabolised in the gut, dysbiosis itself also changes which metabolites reach host tissues. Therefore, diet can shape microbiome composition, but also impact systemic metabolism by altered processing of dietary nutrients through the microbiome [49, 110].

In Chapter 4 the composition of the diet was used to constrain exchange reactions of tissue GSMMs. However, since the gut microbiome plays a significant role in processing dietary nutrients and changes in its composition can change its metabolic capacity, community modelling can be used to investigate this. In this chapter I use quantitative metagenomics data to calculate species abundances that are used for the reconstruction of microbial community genome scale metabolic models (comGSMMs). These models are then used to test how compositional changes of the microbiome brought about by the presence of liver tumours and fasting affect metabolism of dietary nutrients in the gut.

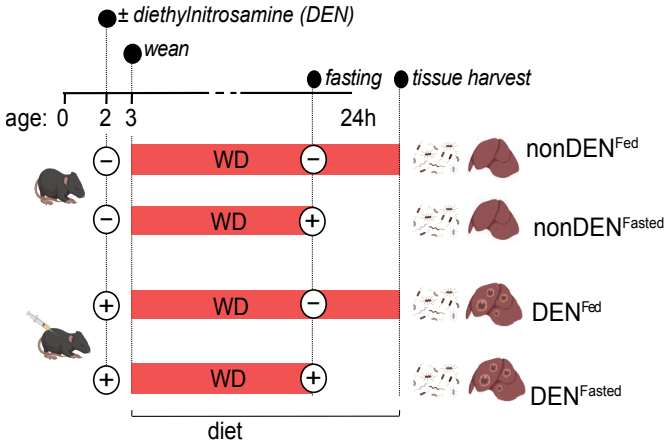


Figure 5.1: **Experimental design for the DEN mouse model with a non-injected control as well as a fasted group.**

5.1 RESULTS

5.1.1 *Tumours and fasting change the gut microbiome species composition*

The DEN mouse model was used to generate metagenomics data from fecal matter and cecum content. A nonDEN control was included in this cohort and before tissues were harvested a subset of mice were fasted for 24 hours to assess the effect of fasting on microbiome and host metabolism (Figure 5.1).

Raw reads were mapped against a gene catalog of gut metagenomes containing ~2.6 million genes. On average, 47.12% of raw sequencing reads per sample mapped against the catalog (Figure 5.2). Two samples with low percentage mapped read count were identified as outliers and removed from subsequent analysis. Outlier samples had similar total number of reads but fewer total mapped read counts compared to other samples, suggesting that low percentage read mapping was likely due to sequencing error. The gene counts for each sample were then used for calculating metagenomic species (MGS) abundances.

Consistent with previous reports, Bacteroidetes and Firmicutes were the most abundant phyla across experimental conditions (Figure 5.3). Actinobacteria, Verrucomicrobia and Proteobacteria were depleted in DEN$^{\text{Fasted}}$ mice suggesting a decrease in community diversity. Proteobacteria have been shown

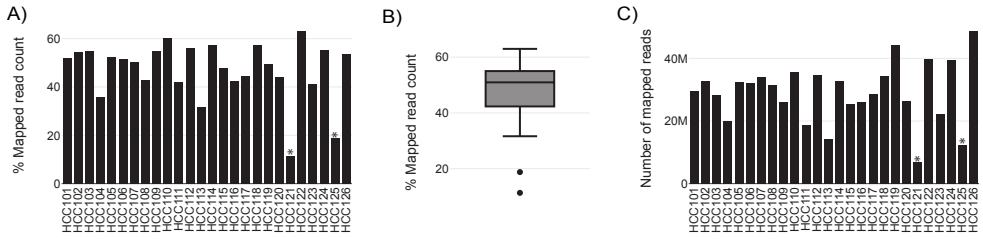


Figure 5.2: **Mapping of metagenomics data against a gene catalog of the mouse gut.** . A) Barplot showing the percentage of mapped reads against the mouse gut catalog B) Boxplot of the same data as in A) highlighting two outlier samples C) Total number of reads mapped against the mouse gut catalog. Asterisks (*) on two bars in A and B indicate outlier samples.

to use host-derived substrates for metabolism and increased in fasting. The increase in Proteobacteria, specifically *Bilophila wadsworthia*, in nonDEN$^{\text{Fasted}}$ mice but its depletion in DEN$^{\text{Fasted}}$ mice suggests that decrease in bacterial richness in DEN mice likely cause these mice to respond different to fasting (Figure 5.4). This trend was further supported by an overall increase in *Odoribacter splanchnicus* in fasting, but a decrease in DEN$^{\text{Fasted}}$ mice compared to nonDEN$^{\text{Fasted}}$ mice. *Lactobacillus johnsonni* was depleted in DEN animals compared to nonDEN animals whereas *Akkermansia muciniphila* was increased. Interestingly, *A. muciniphila* was specifically depleted in DEN$^{\text{Fasted}}$ animals. Collectively these observations indicated that the composition of the gut microbiota is impacted by both tumours and fasting.

5.1.2 *Gut microbiome composition impacts production of biomass and other metabolites in a diet specific-manner*

In order to model how compositional changes in microbiome potentially impact metabolism, comGSMMs were reconstructed (see Chapter 2 and Figure 2.2 for details). On average across conditions, the top 20 species accounted for more than 95% of total species abundances and the remaining less abundant species made up a very small proportion in each sample. Therefore, the 20 most abundant metagenomics species (MGSs) were selected as a representation of the total community composition.

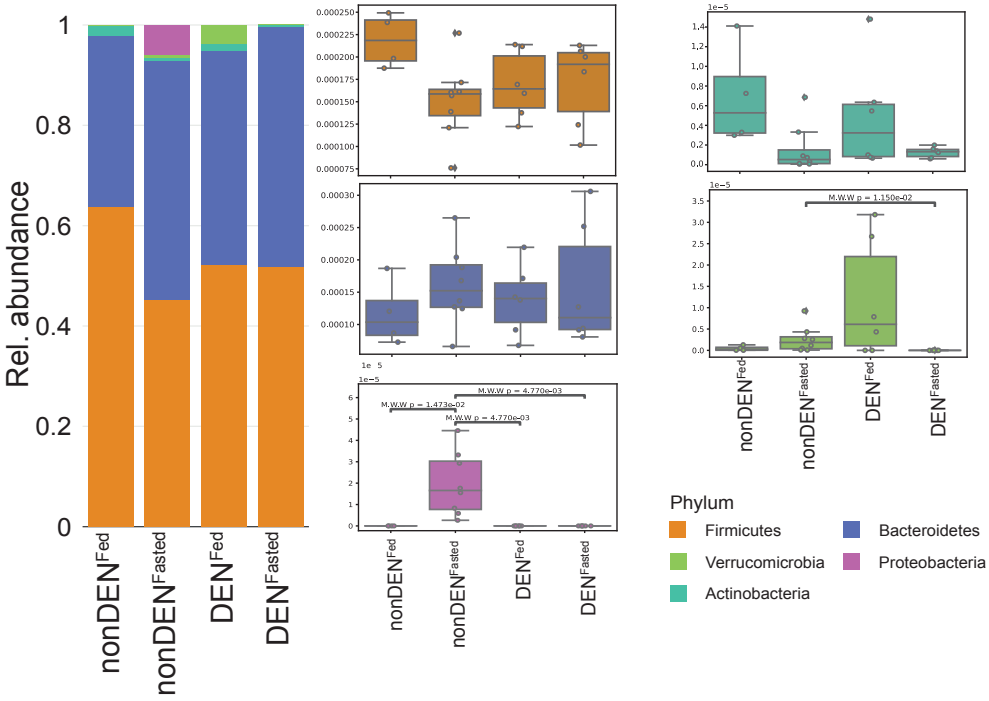


Figure 5.3: **Phylum abundances change in tumorigenic animals and upon fasting**. The gut community is dominated by species of Bacteriodetes and Firmicutes with less abundant phyla, Actinobacteria, Proteobacteria and Verrucomicrobia increased in different experimental conditions. Proteobacteria is statistically increased upon fasting in nonDEN animals but depleted in all other conditions. Verrucomicrobia is statistically decreased in DEN^{fasted} animals. M.W.W: Mann-Whitney U test

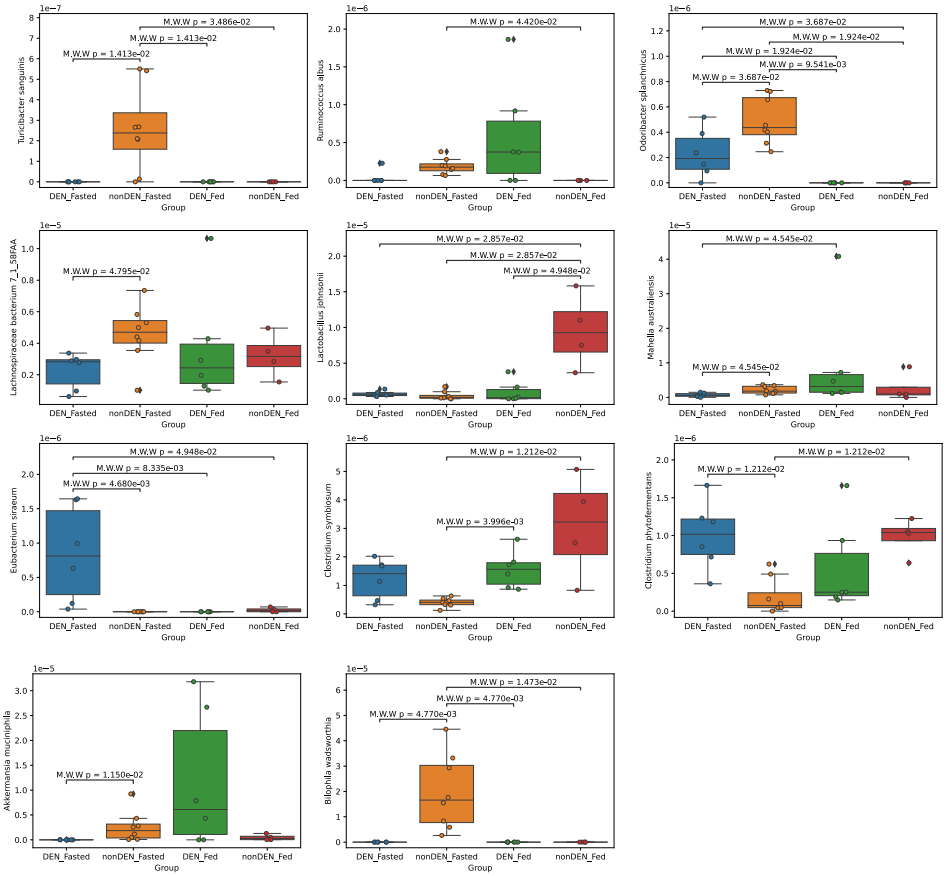


Figure 5.4: **Bacterial species significantly changing across four different experimental conditions**. The injection of DEN as well as fasting impacted the abundance of different species. M.W.W: Mann-Whitney U test

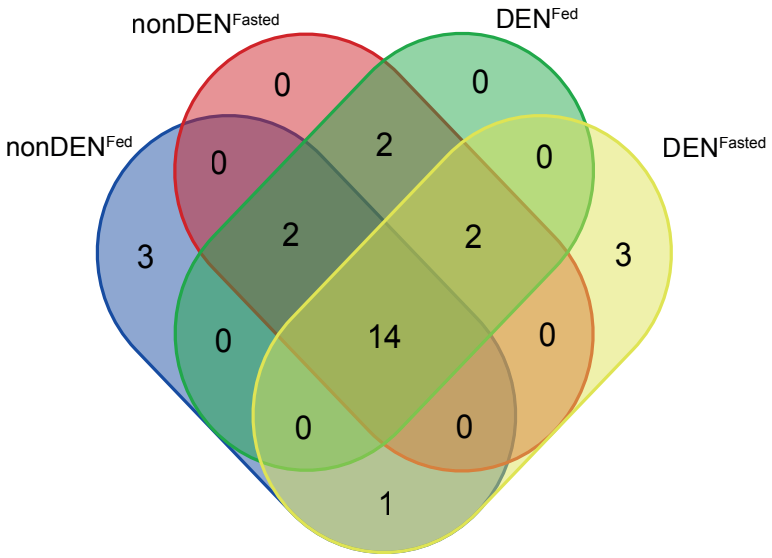


Figure 5.5: **Overlap of the top 20 most abundant species of each experimental condition selected for metabolic modeling.**

A total of 14 species were common across the four experimental conditions with only a few unique species per condition (Figure 5.5). However, individual species rank different according to their relative abundance between these conditions (Figure 5.6). Therefore, the lower bound of the biomass production reaction flux of each species GSMMs was constrained relative to abundance of that species in the community. The exchange reactions of comGSMMs were then constrained using either the content of the WD or the CD to test how diets alter the flux distribution under different conditions. The diet constraint was done identical to the tissue-specific models reconstructed in Chapter 4 to allow cross-comparison. Consistent with csGSMMs, the following notation was used for microbial models: $comGSMM_{Y}^{X}$ where X indicated the experimental condition and Y is the diet given to the model.

The total uptake of carbons by each community, $C_{moles}{}^{INFLUX}$, was less with WD compared to CD (Figure 5.7A). This is despite higher $C_{moles}{}^{DIET}$ in WD compared to CD and in contrast with simulations in csGSMMs. comGSMMs did not take up all lipids available from the diet. Lipids are primarily metabolised by tissues and not gut microbes indicating that reconstructed comGSMMs recapitulate known metabolic features of the gut.

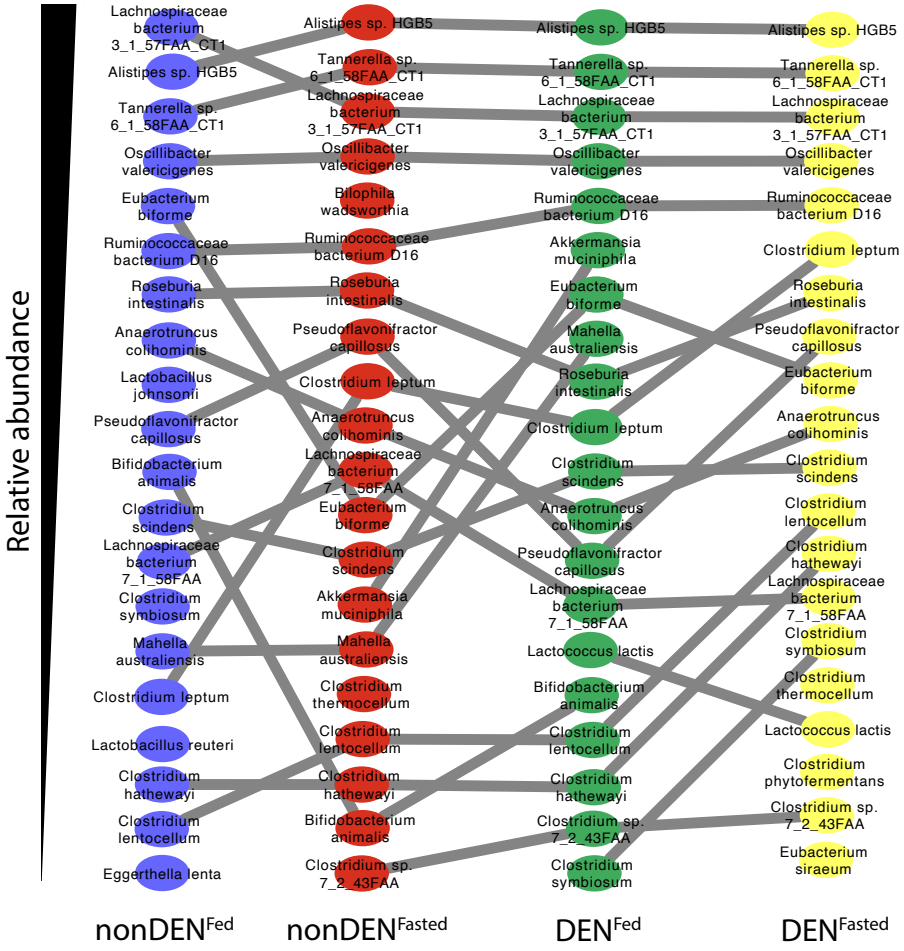


Figure 5.6: **Ranking of the top 20 most abundant MGSs of each experimental condition selected for metabolic modelling.** A total of 14 species overlap between four experimental conditions, however, the relative ranking of the top species differs between these conditions. In order to account for this, CBM was used to constrain the biomass reaction of individual species.

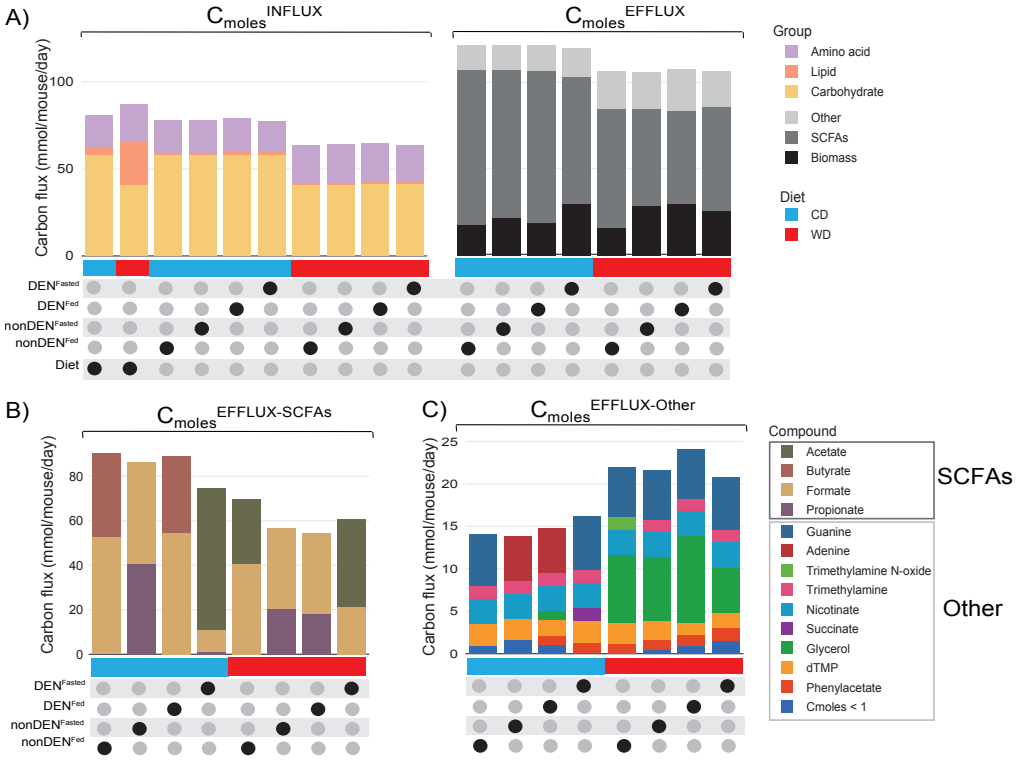


Figure 5.7: **Gut microbiome composition dictates the production of biomass and other metabolites.** A) C_{moles}^{DIET} and C_{moles}^{INFLUX} values of each of the major dietary nutrient classes (carbohydrates, lipids and amino acids) in CD and WD for comGSMMs. C_{moles}^{EFFLUX} show high production of SCFAs across comGSMMs and variable biomass between comGSMMs. B) $C_{moles}^{EFFLUX-SCFAs}$ show increased acetate production in DEN^{Fasted} mice. C) WD cause an increased production of glycerol across comGSMMs.

The proportion of nutrient classes and total C_{moles}^{INFLUX} were similar across comGSMMs for each diet, but relative proportions of C_{moles}^{EFFLUX} for biomass, short-chain fatty acids (SCFAs) and other metabolites changed across experimental conditions (Figure 5.7A). This suggested that the community composition does not dictate nutrient uptake but influences how these nutrients are metabolised.

The mean biomass production flux across conditions was 1.083 mmol/mouse/day (SD = 0.101), indicating the robustness of the constraining process. Fasting and diet composition did, however, impact biomass production (Figure

5.7A). Fasting increased biomass production in nonDEN and DEN mice when constrained with CD. Interestingly, $comGSMMM_{WD}^{DEN-Fed}$ had the overall highest biomass production, 1.18 mmol/mouse/day, and decreased with fasting while the opposite was observed in $comGSMMM_{WD}^{nonDEN-Fed}$. Furthermore, biomass production in $comGSMMM_{WD}^{nonDEN-Fasted}$ was similar to $comGSMMM_{WD}^{DEN-Fed}$, while the biomass production for these models under CD was less. It is therefore likely that the gut microbial community of DEN animals is already in a 'fasted-like' state and that fasting in these animals disrupts the community composition and cause a decrease in biomass production. As such, biomass production in $comGSMMM_{WD}^{DEN-Fed}$ is similar to $comGSMMM_{WD}^{nonDEN-Fasted}$.

To test which microbes were responsible for increased biomass production in $comGSMMM_{WD}^{DEN-Fed}$, individual species GSMMs were removed from the comGSMM. The removal of *Akkermansia muciniphila*, *Pseudoflavonifractor capillosus* or *Lachnospiraceae bacterium* stopped the community growing in $comGSMMM_{WD}^{DEN-Fed}$ (Figure 5.8). Interestingly, the removal of *L. bacterium* had the opposite effect in other comGSMMs compared to $comGSMMM_{WD}^{DEN-Fed}$ resulting in an increase in biomass production. Collectively, this further supported that biomass production in DEN^{Fed} is dependent on the composition of the community and the composition of the WD.

Systematic diet composition swap (SyDiCoS) was then used to test the dependence on diet composition for biomass production under these conditions. Increased biomass production in $nonDEN^{Fasted}$ and DEN^{Fed} in WD compared to CD, was due to a combination of all three nutrient classes and component switch for carbohydrates, lipids and amino acids reduced biomass production in these models (Figure 5.9). $comGSMMM_{WD}^{DEN-Fasted}$ had increase in biomass production flux compared to $comGSMMM_{CD}^{DEN-Fasted}$ (Figure 5.7A) and SyDiCoS showed that $comGSMMM_{WD}^{DEN-Fasted}$ can be increased by switching to the carbohydrate content of the CD, which is primarily comprised by glucose as opposed to fructose in the WD. This suggested that the increased glucose in CD are essential to sustain growth in this community.

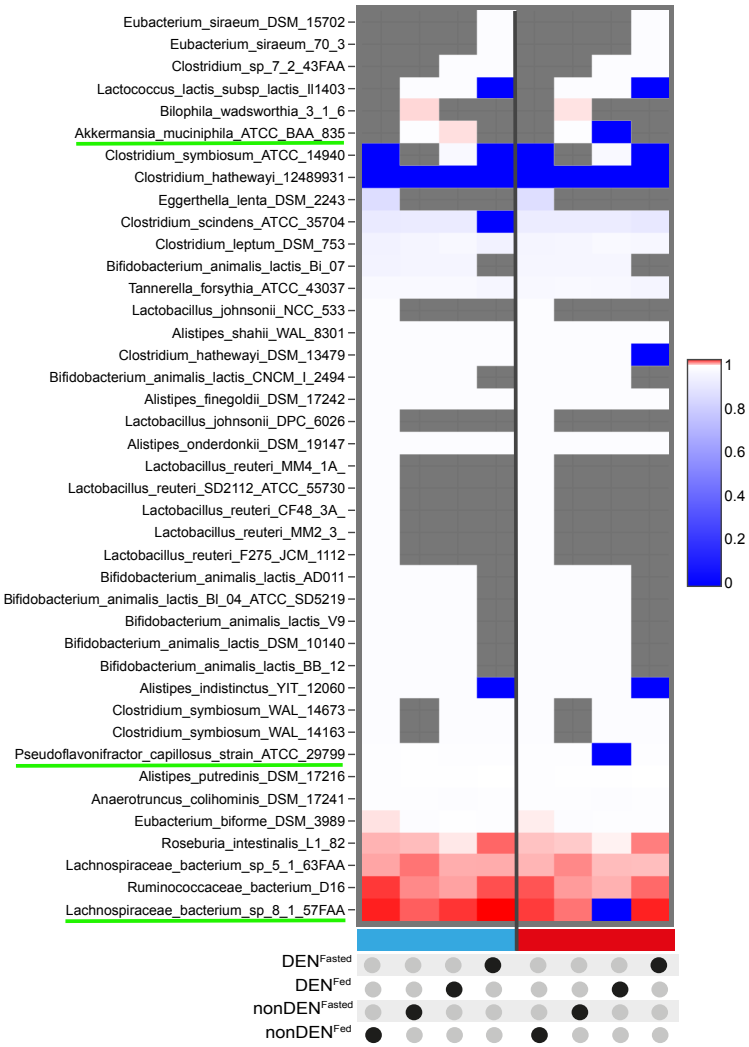


Figure 5.8: **The effect of individual species on community biomass production.** The GSMM for species on the y-axis was removed from each comGSMM on the x-axis and the effect on community biomass flux was compared to wild-type biomass flux. A reduction in biomass flux after removal of the species appears in blue and an increase in biomass flux in red relative to wild-type flux. GSMMs that are uniquely reducing biomass flux in $comGSMMM_{WD}^{DEN-Fed}$ are underlined in green.

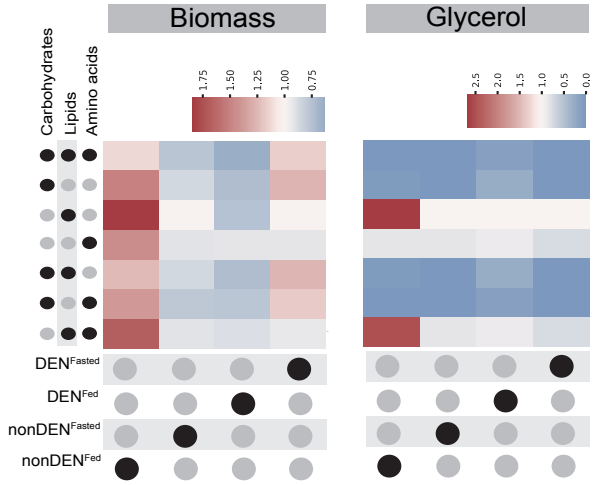


Figure 5.9: **Diet composition effects on biomass and glycerol production**. Systematic Diet Component Swap (SyDiCoS) to assess the role of WD components upon biomass and glycerol production flux in comGSMMs. $C_{moles}{}^{DIET}$ values for all three major diet component classes (carbohydrates, lipids and amino acids) in the WD were swapped individually or in combination with the corresponding $C_{moles}{}^{DIET}$ values in CD while leaving the remaining dietary $C_{moles}{}^{DIET}$ values of the WD unaltered. The swapped component(s) are indicated by black dots on the left. The colour scale represents the ratio of biomass production flux or glycerol production flux in models provided with the swapped diet relative to the respective fluxes in models provided with WD, calculated for each comGSMM shown at the bottom of the figure.

SCFAs accounted for 52.64-72.51% of C_{moles}^{EFFLUX} in all models. Formate was produced by all comGSMMs while acetate, propionate and butyrate were produced in a diet, fasting or tumour-specific manner (Figure 5.7B). Butyrate was produced only in fed-conditions and ,interestingly, acetate production was increased in DEN^{Fasted} animals irrespective of diet, suggesting that in DEN animals gut microbiome compositional changes potentially support host metabolism by providing additional FAs during fasting. $C_{moles}^{EFFLUX\text{-}Others}$ increased in WD-fed comGSMMs. This increase was mostly accounted for by an increase in glycerol production (Figure 5.7C). Glycerol production in DEN^{Fed} animals was increased compared to $nonDEN^{Fed}$ animals and in both conditions fasting decreased this but with increased magnitude in reduction in DEN animals. Using the SyDiCoS approach carbohydrates in the WD were shown to be responsible for this increased glycerol production (Figure 5.9).

5.2 CONCLUSION

The gut microbiome plays an important role in metabolising dietary nutrients. The species composition of the gut microbiome has been shown to impact its function and how it can contribute to disease. To study how gut microbiome composition and dietary nutrients impact the metabolic capability of the gut, comGSMMs have been reconstructed in this Chapter based on quantitative metagenomics data.

Different experimental conditions impact the species composition of the gut. *Odoribacter splanchnicus* that increased in fasting has previously been correlated with a healthy fasting serum lipid profile [150]. Its decrease in DEN^{Fasted} animals show that tumour-bearing animals respond different to fasting likely because of the presence of the tumour. Other differentially abundant species, such as *Lactobacillus johnsonni*, has been shown to decrease the metabolic effects of obesity [151] and its depletion in DEN animals emphasises that beneficial bacteria likely decrease in tumour-bearing animals. Several studies has previously associated *Akkermansia muciniphila* with liver disease [152–154]. The results here indicate that fasting tumour-bearing animals can potentially reduce the abundance of this species and flux analysis also showed that *A. muciniphila* is responsible

for the characteristic increased biomass production in DEN^{Fed} animals. This result suggests that fasting would have a beneficial outcome in liver cancer by decreasing the abundance of bacterial species such as *A. muciniphila* in the gut.

A WD cause an increased production of glycerol across all conditions. Glycerol is an important metabolite in carbohydrate and lipid metabolism and this prediction shows that glycerol produced in the microbiome can potentially be used for metabolism in host tissues such as the liver. Furthermore, previous studies have shown that glycerol can be used as a substrate for the synthesis of acrolein, a carcinogenic compound [155, 156]. This indicates that a WD could potentially increase production of acrolein which will be tested in future work.

The intricate interplay between gut microbiome species composition and diet have previously been studied [111, 157]. At the same time, the diversity and richness of the gut flora have been shown to correlate with metabolic function [110]. Therefore, the observation that community composition does not dictate which nutrients from the diet are used, but rather how they are metabolised is interesting. This shows that the community will most likely be able to metabolise all the nutrients of the diet but how these nutrients are metabolised are key to understanding longitudinal metabolic changes. This furthermore emphasises the importance of using mathematical models to investigate the underlying mechanisms of diet metabolism in the gut.

There are limitations in using the modelling approach as done here. In the first instance, using biomass production for optimisation using LP has similar pitfalls as in tissue-models reconstructed in Chapter 4, as this only gives one possible solution. Using this approach did, however, make comparison of nutrient distribution and efflux against tissue models possible. It was also assumed that the relative abundance of individual species correlates with biomass flux and that the most abundant species contribute to metabolism of the community.

6

MULTI-TISSUE METABOLIC MODELLING UNCOVERS NOVEL METABOLIC INTERACTIONS BETWEEN THE LIVER, TUMOUR AND MICROBIOME

"A model that proves very inadequate will be quickly rejected, without contributing much to the genesis and progression of knowledge, while a succession of adjustments to a model that is useful, though not perfect, will lead to an increasingly detailed representation of the phenomenon"
- Antoine Danchin

Studying metabolism at individual tissue level has given valuable insights into metabolic alterations in diseases, such as cancer. On an organismal level, however, metabolism involves coordinated cross-talk between individual tissues regulated by several different intrinsic and extrinsic factors. The tumour microenvironment (TME) is the ecosystem directly surrounding tumour tissue and has been shown to metabolically interact with tumour tissue in a highly dynamic manner [66]. Tumours therefore evolve mechanisms to adjust to changes in nutrient composition in the TME that allow tumour cells to grow and survive. This quick adaptation highlights the importance of studying tumour biology in appropriate physiological and nutritional contexts [158].

Cellular compartmentalisation is an important feature of metabolism. This information is included in the reconstruction of genome scale metabolic models (GSMMs) and has been used to reconstruct multi-tissue GSMMs (mtGSMMs) by representing individual tissues as different compartments in a single model [14, 113, 114, 159]. By using this approach it is possible to study inter-organ metabolic interaction and to ultimately have a more accurate representation of whole-body physiology.

In this chapter, I reconstruct a mtGSMM to study the metabolic interactions between the tumour, the liver host and the microbiome. For this purpose the tissue models reconstructed in

Chapter 4 and the microbiome models reconstructed in Chapter 5 are integrated into a single unified framework to allow investigation of metabolic interactions between these tissues. These models are then used to test how the liver host and the microbiome metabolically interact with the tumour in a diet-specific manner.

6.1 RESULTS

6.1.1 *A WD influences liver and tumour metabolic cross-talk*

To model how the liver host and microbiome potentially support tumour metabolism, a mtGSMM was reconstructed (see Chapter 2 and Figure 2.3 for details). The mtGSMM consisted of a blood and gut compartment that, within them, contained tissue GSMMs and comGSMMs, respectively. Individual tissue models were constrained with gene expression data (as described in Chapter 4) and the community model with metagenomic species abundance data (as described in Chapter 5). The objective of the gut was set to biomass production and biomass production in tumour tissue was set as an additional objective. Input into the mtGSMM was then constrained with the content of the western diet (WD) or the control diet (CD) to test how diets alter the flux distribution under different conditions. Using this modelling setup, it was possible to study potential metabolic cross-talk between the gut microbiome and host tissues, between liver host and tumour, how the diet impacts such crosstalks and, ultimately, how it potentially impacts biomass production of the tumour.

Tumour biomass production was increased in the mtGSMM compared to the tumour as a single-tissue model (Figure 6.2A). Furthermore, the differences in flux distributions were greater between single-tissue GSMMs and mtGSMMs compared to differences within these models and these differences were amplified on a WD (Figure 6.2B). This suggested that the the liver and the microbiome affect the flux distribution of the tumour. In addition, because the amplification of these differences on a WD qualitatively correlated with an increase in biomass, it shows that the liver potentially supports tumour biomass production dependent on the diet composition.

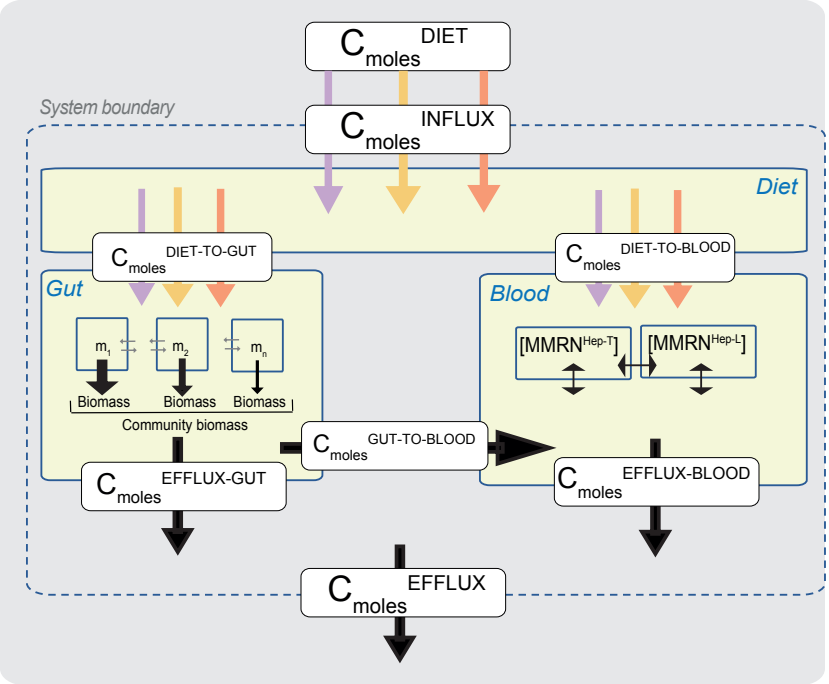


Figure 6.1: **Schematic illustrating setup of integrative mtGSMM**.

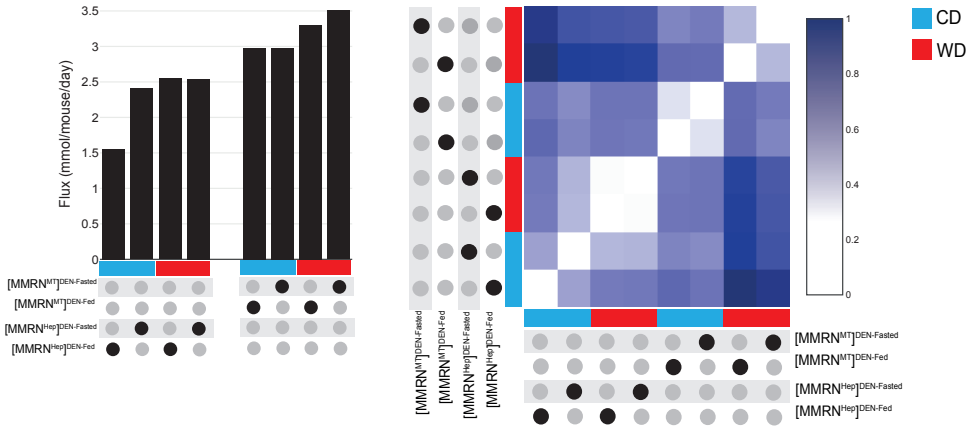


Figure 6.2: **A mtGSMM impacts tumour biomass production and flux distributions**. A) Bar plot of predicted tumour biomass production as a single-tissue GSMM and within the mtGSMM. B) Euclidean distances between flux distributions for tumour tissue in a single-tissue GSMM and within the mtGSMM under different conditions. The colour scale represents the Euclidean distance values calculated in a pairwise manner between each of the flux distributions and plotted relative to the maximum distance value across all comparisons.

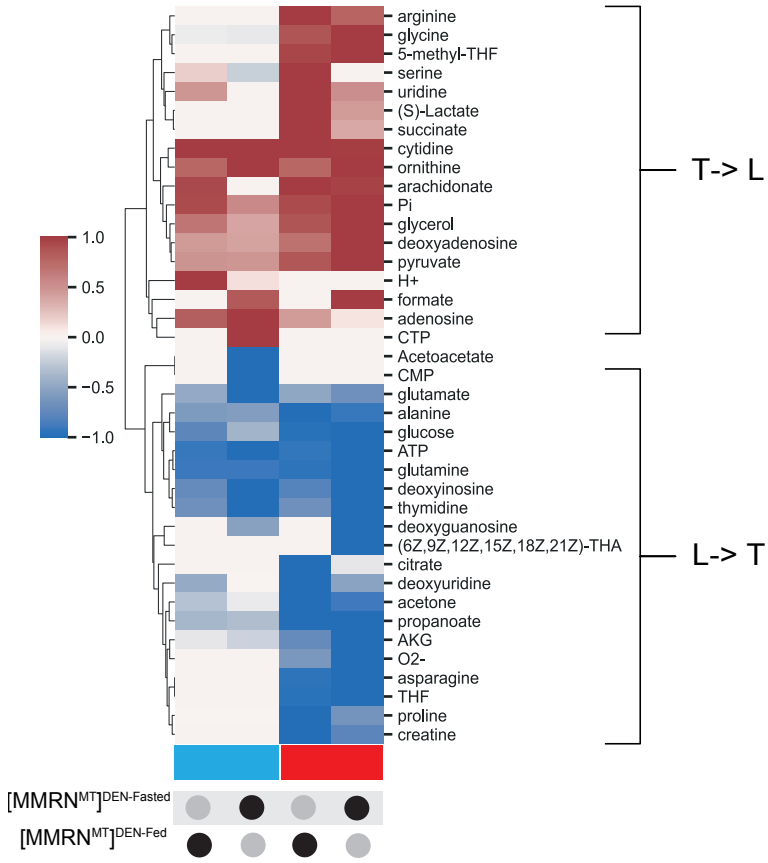


Figure 6.3: **Metabolic interaction between the tumor and liver host.** Heatmap showing metabolites exchanged from tumor (T) to liver (L) in red and from liver to tumor in blue. Each metabolite row is scaled to the maximum flux value of that metabolite with higher color intensity therefore indicating higher flux compared to other conditions for that metabolite.

The mtGSMM was reconstructed to allow exchange of metabolites between the liver and the tumour which made it possible to directly assess tumour-liver metabolic exchange. Known metabolic substrates for tumour metabolism, such as glucose and glutamine, were predicted to be exchanged from the liver to the tumour (Figure 6.3) [66]. Interestingly, some metabolites, such as creatine, propanoate and proline showed a WD-specific exchange from the liver to the tumour.

Glycerol and succinate, predicted to be produced in a WD-specific manner in single-tissue liver models, were exchanged from the tumour to the liver on WD (Figure 6.4). Further in-

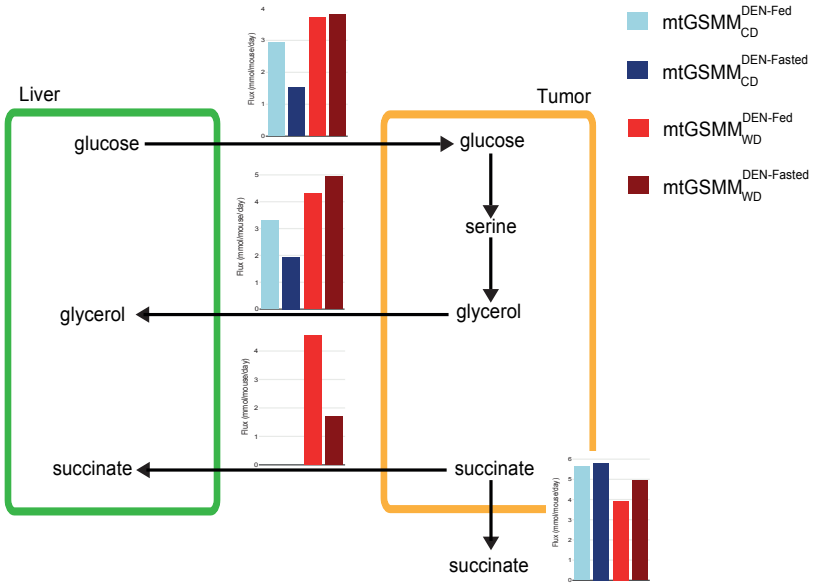


Figure 6.4: **Glycerol and succinate exchange between tumor and liver tissues.** Schematic illustration of exchange of key metabolites, glucose, glycerol and succinate with flux values depicted on reactions as bar plots.

vestigation into the flux distributions, showed that glycerol was produced from serine *via* the putative pathway identified in single-tissue GSMMs (Chapter 4) and that increased serine production in the tumour was, in part, a result of increased glucose exchange from the liver on WD. Similar to glycerol production, succinate was only produced by the tumour in the mtGSMM, however, in CD all succinate was transported from the tumour to the blood while some succinate was exchanged to the liver in WD. Interestingly, DEN^{Fed} animals in WD show the highest overall succinate production and fasting reduces this to similar levels compared to CD suggesting that a diet-switch and fasting have a similar effect on succinate production in the tumour.

6.1.2 *Acetate and propanoate produced by the microbiome are further metabolised in tissues*

Metabolite exchange from the gut to the blood increased significantly in DEN animals compared to nonDEN animals which

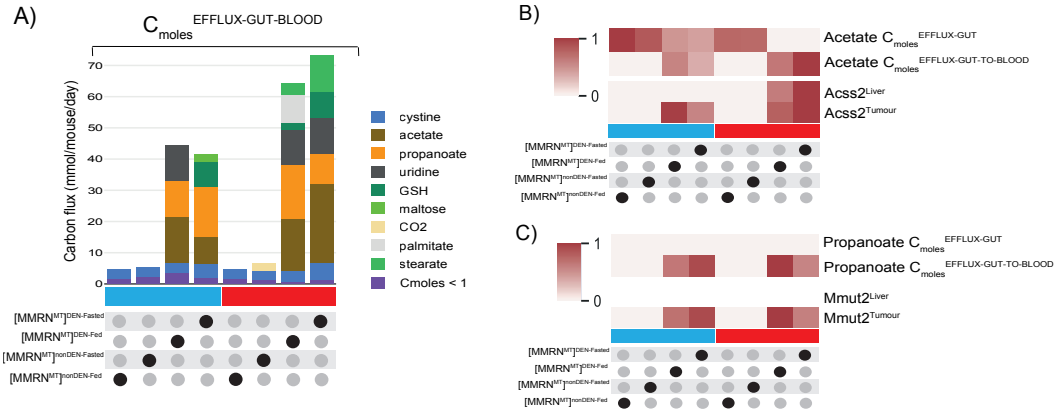


Figure 6.5: **Diet-specific acetate production by the gut is preferentially metabolized by the tumor by Acss2.** A) An increased exchange of metabolites from the gut to the blood in DEN animals accentuated by a WD. B) The relative fluxes for reactions related to acetate metabolism. Each row is scaled to the maximum flux for that row with a darker red color showing increased relative flux.

was mostly accounted for by the short chain fatty acids (SCFAs) acetate and propanoate (Figure 6.5A). Previous studies have shown that microbiome-derived acetate can increase hepatic lipogenesis through ACSS2 by increasing acetyl-CoA production. A closer investigation into the flux distribution showed microbiota-derived acetate is further metabolised in tumour tissue in CD, where less acetate is produced compared to WD, and that in WD both the liver and tumour metabolises this acetate *via* the Acss2 reaction (Figure 6.5B). Interestingly, propanoate was only produced in tumorigenic animals and was exclusively transported to the blood compartment (Figure 6.5C). Through a successive series of metabolic reactions, propanoate was metabolised in tumour tissue to succinyl-CoA. The final step in this pathway was catalysed by methylmalonyl-Coenzyme A mutase (Mmut2).

6.2 CONCLUSION

In this chapter I reconstructed a mtGSMM to investigate how the liver host and microbiome impact tumour biomass production. Cancer cells secrete different soluble factors such as cytokines

and small molecules that recruit other cells to the tumour site [66, 160]. This creates a TME with dynamic interactions including the exchange of metabolites between tumour and host cells. The cross-talk between cancer cells and cells within the TME is a widely studied research topic in cancer biology to understand how host cells support tumorigenesis and whether this can be used in therapy. Although several instances of metabolic cross-talk between tumour and host cells have been documented, the extent of this type of interactions are unknown.

The advancement of computational algorithms in metabolic modelling has led to numerous reconstructions of tissue-specific and cell-specific GSMMs. Although these GSMMs are helpful in understanding metabolism at cellular and tissue level, physiological metabolism involves coordinated metabolic exchange between different cells and tissues on an organismal level. In this regard, an effort has been made in the metabolic modelling community to try and recapitulate whole-body metabolism by reconstruction of mtGSMMs [113, 114]. The recent organ-resolved GSMMs reconstructed by Ines Thiele and colleagues are the most comprehensive reconstructions towards simulating whole-body metabolism [14]. There are, however, a very limited number of studies that modelled microbiome-host metabolic interactions [97] and none that focused on the metabolic interactions between tumour and host tissues.

WD-specific exchange of metabolites from the liver to the tumour identified here have previously been associated with cancer metabolism. For example, creatine derived from the diet or synthesised *de novo* has been shown to promote cancer metastasis in mice [161, 162]. Proline, primarily synthesised from glutamine, has also been shown to promote tumorigenesis in liver cancer [163]. Although many studies focused on the metabolism of these metabolites by the tumour itself, the flux simulations highlights that these metabolites can potentially also be derived from the host tissue that can then be used for metabolism in the tumour.

Similar to single-tissue models, there are several limitations in the reconstruction of mtGSMMs and when using these models for predicting fluxes. In the first instance, the distribution of dietary nutrients to either the microbiome or tissues are not explicitly constrained, as done in previous studies [97]. Therefore, in

conditions where the tumour is present more dietary nutrients directly be taken up to support its metabolic need rather than being first metabolised by the microbiome. This can, however, be tested with the current models in further work and will likely yield interesting results about nutrient distribution across tissues. Furthermore, although gut-to-blood exchange potentially gives a good estimate of metabolites reaching tissues, metabolites will first pass through the intestine, a major metabolic organ, which will likely change which metabolites reach the tissues. Finally, limited manual curation was done on exchange reactions between liver and tumour tissue and metabolites from the gut to systematically check whether these reactions can indeed take place. This might lead to predictions of exchanges between these tissues that is not possible.

7 DISCUSSION AND FUTURE WORK

"*Out of intense complexities, intense simplicities emerge*" - Winston Churchill

7.1 IMPROVING HGSMM-BASED MGSMMS RECONSTRUCTIONS

Reconstructing mGSMMs based on mouse-to-human orthology has become the state-of-art in mouse genome scale metabolic model (mGSMM) reconstruction over the last couple decades (examples [97, 98, 135]). This approach takes advantage of the significant effort put into reconstruction of human GSMMs (hGSMMs) to alleviate the cumbersome and time-consuming process of bottom-up reconstruction [108]. At the time of inception of this thesis, all available mGSMMs had been reconstructed from a single hGSMM, belonging to either the Recon or HMR series. Because different hGSMM series comprise non-redundant metabolite and gene components [92], resulting mGSMMs may be incomplete if a single template model is used as reference. In addition, because the orthology based approach for GSMM reconstruction is dependent on the the coverage of orthologs described between human and mouse, it is desirable to capture as many orthologs as possible.

These limitations were addressed when reconstructing Mouse Metabolic Reaction Network (MMRN) by using both HMR2 and Recon3D as template models as well as incorporating known and newly identified orthologs through protein sequence similarity. The fact that intermediate models (IMs) from different hGSMMs and from protein homology-based methods show significant non-redundancy demonstrates that integration of pre-existing models and improvement of orthology information are warranted to facilitate reconstruction of more complete model organism GSMMs based on hGSMMs.

7.2 INVESTIGATING THE INTERPLAY BETWEEN NUTRIENT AVAILABILITY AND GENE EXPRESSION ON METABOLIC FLUX

Metabolic flux is largely determined by nutrient availability and gene expression of enzymes catalysing metabolic reactions [139]. Chronic consumption of 'western-style' diets (WD) lead to the dysregulation of several homeostatic mechanisms controlling gene expression that result in distinct gene expression that contribute to altered metabolism in liver tissue. These changes contribute to the development of liver tumours and, once established, metabolic reprogramming in transformed cells is also influenced by several tumour cell-autonomous factors [141]. There has been an increasing appreciation for the use of dietary interventions [164] in cancer therapy but whether or not dietary modulation alone suffices to change metabolism in the context of concurrent gene expression changes remains largely unknown.

The systematic diet composition swap (SyDiCoS) approach developed here allows systematic investigation of how the modulation of different nutrient classes effect flux distributions. Identification of pathways that account for flux differences between simulations are challenging as it requires either sophisticated network analysis or laborious manual exploration of flux distributions [81, 100]. Therefore, models are often challenged by simulating perturbations such as *in silico* gene or reaction knockouts and studying its effect on a particular phenotype [165, 166]. Similarly, SyDiCoS allows efficient interrogation of large flux distributions by systematically interrogating which nutrients are responsible for driving predicted fluxes. In this manner it is possible to directly associate specific nutrients in the diet with predicted reaction fluxes.

Methods used to constrain intracellular fluxes through gene expression data are robust and widely available [167], however, how nutrient availability effect flux are more limited. An important distinction of SyDiCoS compared to previous investigations of model input in GSMMs is that specific nutrient classes were swapped out to physiologically relevant values in a CD. The motivation for this was rooted in the observation that in DEN^{CD}animals, tumours do not develop in the same time-frame

as in DEN^{WD} and therefore changing dietary nutrient components to corresponding values in the CD could indicate which metabolic pathways are responsible for this phenotype in WD.

There is a growing interest towards "precision nutrition" in cancer therapy [157, 168]. A detailed mechanistic understanding of the metabolic requirements of tumour cells and the modelling framework proposed here will greatly aid to uncover potential targets to attenuate cancer progression using dietary intervention. An important consideration in dietary studies should also be how feasible it is to design these diets for patients. Several mechanistic insights into diet-tumour interactions in animal studies rely on diets that, for example, have restriction in specific metabolites, such as individual amino acids, which might not be easily transferable to human diets. The results here indicated that, in some instances, changing the composition of the diet is sufficient to overcome some metabolic features of liver tumours.

7.3 MTGSMMS CHANGE PREDICTIONS MADE IN SINGLE-TISSUE GSMMS

Extracting tissue-specific GSMMs from generic organism-specific GSMMs has been done in several previous studies (examples [87, 90, 91]). Although there are a limited number of examples, mtGSMMs have also previously been reconstructed [113, 114, 169]. A pertinent question that has, however, not been addressed previously is how predictions in single-tissue GSMMs change, if at all, in mtGSMMs. GSMMs are sensitive to changes in constrains and with an ensemble of methods available in CBM, the uncertainty associated with prediction has previously been highlighted [136]. Therefore, although there are major efforts in the metabolic modelling community towards reconstruction of whole-body physiological metabolism [14], it has not been shown whether the efforts to increase GSMM size, by inclusion of additional tissues, change the predictions made in simpler models.

The work done here shows that a mtGSMM quantitatively and qualitatively change predictions made in single-tissue models. For example, tumour biomass production increased in the context of the mtGSMM compared to a single-tissue tumour

GSMM. Systematic investigation in future work will address whether this is because of increased uptake of dietary nutrients or whether the liver and microbiome provides additional substrates for biomass production. On the other hand, the tumour showed increased glycerol and succinate in WD compared to CD in both the single-tissue GSMM and mtGSMM, however, the fate of these metabolites were affected in the mtGSMM. This highlights that the inclusion of additional tissues can qualitatively change predictions made in single-tissue GSMMs. In previous mtGSMMs, constraints were introduced to the model to more accurately reflect physiology, although, in these instances, it was not shown whether it change predictions made in single-tissue models [113, 114]. In the reconstruction of the mtGSMM here, the objective functions and constraints on intracellular fluxes were kept the same as in a single-tissue setting to allow comparison between these models.

Glycerol and succinate were produced in $MMRN^{Hep}$ at higher rates in WD compared to CD. The importance of glycerol as a metabolic substrate for carbohydrate and lipid metabolism in metabolic disease has been highlighted in several studies [170–172]. Glycerol is primarily produced through glyceroneogenesis from glucose or through the lipolytic breakdown of lipid macromolecules such as triacylglycerides (TAGs). The putative pathway identified that produces glycerol from serine is interesting because it may provide a previously unknown source for liver specific glycerol production. Moreover, although serine production *via* the serine biosynthesis pathways is increased in liver cancer, the fate of increased serine is still largely under debate. Serine is used for several downstream metabolic pathways such as one-carbon metabolism, protein synthesis and the production of sphingolipids, but the production of glycerol has not previously been proposed [173]. An interesting insight from the mtGSMMs is that the glycerol produced in the tumour is exchanged to the liver as opposed to the blood, which would have otherwise have not been possible to predict without the mtGSMM. Recent evidence suggests that succinate might have immunomodulatory roles within the liver and is increased in the blood of obese individuals, although its exact source is still under debate [174, 175]. The modelling results here could therefore potentially be useful in elucidating how diet can directly and

indirectly contribute to these circulating metabolites associated with obesity.

7.4 THE INTEGRATIVE MODELLING FRAMEWORK COULD AID IN ELUCIDATING DIET-MICROBIOME-LIVER INTERACTIONS

There is an increasing realisation of the importance of the gut-liver axis in healthy and diseases states. At the same time the diet profoundly impacts the development and progression of diseases, such as liver disease, through direct and indirect mechanisms, including changing the composition of the gut microbiome. Studying the interactions between these components; diet, microbiome and liver, are important to elucidate how the interplay between them contribute to liver disease. The lowering costs in metagenomics sequencing permitted several studies to correlate gut microbiome species composition with liver disease. However, a current challenge in microbiome studies is to assign mechanistic causation to changes in species composition in disease states. The tools developed in this thesis and the modelling framework proposed will therefore greatly aid in this pursuit by providing a platform to systematically deconvolute these interactions.

7.5 FUTURE DIRECTIONS

The work presented in this thesis lays an important foundation for future computational and experimental work.

Integration of additional tissues to the mtGSMM

As highlighted in this thesis, and in several other studies, systemic metabolic reprogramming is a feature of liver disease and cancer. It is therefore important to study hepatic metabolism in the context of other tissues to gain a more holistic understanding of whole-body metabolism. The mtGSMM reconstructed here lays important foundation for this purpose to expand the model by including other tissues. For example, transcriptomics data have been generated for white adipose tissue (WAT) that will be included in the mtGSMM to investigate how, in particular, lipid metabolism in the WAT impact metabolic pathways identified here.

Glycerol metabolism

The contribution of different pathways to the production of glycerol and, in turn, how glycerol is used in liver cancer is a central focus in the Anastasiou laboratory. In short, genetic mouse models have been developed to functionally test the contribution of different glycerol producing pathways to liver metabolism. The metabolic models developed here will be used to integrate data derived from these mouse models as well as informing future experimental work. Finally, the putative pathway of glycerol synthesis identified here is currently being validated along with the contribution of dietary fructose metabolism to overall glycerol production. The GSMMs reconstructed here were therefore indispensable tools to inform experimental work.

Diet-switch experiments

The SyDiCoS analysis in $MMRN^{Hep}$ suggested that the metabolic profile of tumours, as a consequence of WD, that cause increased tumour biomass production can be reversed by switching to a diet with less lipids and fructose even in the background of gene expression changes that resulted from chronic consumption of

WD. In order to test this prediction, in a new cohort of tumour-bearing DEN^{WD} animals, mice will be switched to a chow-diet to investigate the metabolic changes associated with a diet-switch. Analysing experimental data from this cohort of animals will form an integral part in future modelling experiments to test whether a diet switch is sufficient to overcome metabolic flux predictions driven by a particular gene expression profile.

SUMMARY

Liver cancer is one of few cancer types that is increasing in incidence because of limited therapeutic options. The rise in liver cancer cases is, in part, due to the rise of liver disease as a consequence of consumption of western-style diets (WD) that contribute to obesity. The progression from early to later stages of liver disease creates a damaged tissue environment that promote the initiation of malignancies. Consequently, liver and tumour tissues distinctly differ in their metabolic and gene expression profiles. Diet modulation has increasingly been used as a therapeutic strategy for cancer patients. However, it is still unclear whether diet modulation alone is sufficient to overcome the metabolic reprogramming that occurs in the background of gene expression changes that is, in part, driven by chronic WD consumption. A particular challenge in metabolic studies is that current analytic techniques can only capture a subset of metabolites. Furthermore, measuring actual metabolic fluxes *in vivo* and *in vitro* requires sophisticated labelling experiments with stable isotopes. Therefore, to study metabolism on a global scale, genome scale metabolic models (GSMMs), the principal tools used in systems biology, have been used as mathematical frameworks for the integration of various data-types, especially omics data, using constraint-based modelling (CBM) approaches. The methods in CBM make it possible to parameterise different parts of the metabolic network to study how biological factors such as nutrient input and gene expression effect metabolic fluxes.

In this thesis, I developed a suit of GSMMs and computational tools to investigate how diet modulation affects metabolic fluxes in liver cancer and the host using data derived from a liver cancer mouse model. The principal outcome in this endeavour was a new mGSMM, Mouse Metabolic Reaction Network (MMRN). This model compared well against mammalian GSMMs published before its reconstruction and also more recently published models and is a valuable resource to study mouse metabolism *in silico*. MMRN was used to study the impact of a WD on liver

and tumour tissues and in the process a new approach was developed to systematically investigate the impact of dietary nutrients on metabolic flux, termed systematic diet composition swap (SyDiCoS). Using this approach increased glycerol and succinate production in all context-specific models were shown to be a consequence of dietary fructose and fatty acids (FAs), respectively. A deeper investigation into liver and tumour fluxes showed that there is a distinct fate of dietary FAs in these two tissues and that in tumour tissue it is used for increased biomass production. Several model predictions were experimentally validated using metabolomics and *in vivo* flux measurements.

The gut microbiome plays an important role in metabolizing dietary nutrients and, at the same time, diet has an impact on gut microbiome composition. To evaluate this, species abundances were quantified in tumorigenic and non-tumorigenic mice and community GSMMs (comGSMMs) were reconstructed using this data. Several species were identified that change in abundance as a consequence of diet, tumors or fasting and its impact on metabolism were then modeled using the reconstructed comGSMMs. The effect of specific species on community growth was also investigated by systematically removing species from the community. Finally, tissue models and microbiome models were then integrated into a single multi-tissue GSMM framework. This model was in the first instance used to test how predictions made in single-tissue models are impacted in a multi-tissue context. Metabolic interactions between the gut and tissues as well as between tumor and host tissues were then investigated using this model.

The computational approaches and resources developed in this work will be invaluable assets for the metabolic modeling community and for future experimental work. Because the diet impacts metabolism in such a complex manner, computational tools to unravel this will greatly aid in the advancement to not only understand basic metabolic research questions but also how to apply this in clinical settings.

BIBLIOGRAPHY

[1] A. Judge and M. S. Dodd. 'Metabolism.' In: *Essays Biochem* 64.4 (2020), pp. 607–647. DOI: 10.1042/EBC20190041. URL: https://www.ncbi.nlm.nih.gov/pubmed/32830223 (cit. on p. 1).

[2] E. Trefts, M. Gannon, and D. H. Wasserman. 'The liver.' In: *Curr Biol* 27.21 (2017), R1147–R1151. DOI: 10.1016/j.cub.2017.09.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/29112863 (cit. on p. 1).

[3] H. S. Han, G. Kang, J. S. Kim, B. H. Choi, and S. H. Koo. 'Regulation of glucose metabolism from a liver-centric perspective.' In: *Exp Mol Med* 48 (2016), e218. DOI: 10.1038/emm.2015.122. URL: https://www.ncbi.nlm.nih.gov/pubmed/26964834 (cit. on p. 1).

[4] J. G. Jones. 'Hepatic glucose and lipid metabolism.' In: *Diabetologia* 59.6 (2016), pp. 1098–103. DOI: 10.1007/s00125-016-3940-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/27048250 (cit. on p. 1).

[5] Z. Younossi, Q. M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, and E. Bugianesi. 'Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention.' In: *Nat Rev Gastroenterol Hepatol* 15.1 (2018), pp. 11–20. DOI: 10.1038/nrgastro.2017.109. URL: https://www.ncbi.nlm.nih.gov/pubmed/28930295 (cit. on p. 1).

[6] Q. M. Anstee, S. McPherson, and C. P. Day. 'How big a problem is non-alcoholic fatty liver disease?' In: *BMJ* 343 (2011), p. d3897. DOI: 10.1136/bmj.d3897. URL: https://www.ncbi.nlm.nih.gov/pubmed/21768191 (cit. on p. 1).

[7] T. Hardy, F. Oakley, Q. M. Anstee, and C. P. Day. 'Non-alcoholic Fatty Liver Disease: Pathogenesis and Disease Spectrum.' In: *Annu Rev Pathol* 11 (2016), pp. 451–96. DOI: 10.1146/annurev-pathol-012615-044224. URL: https:

//www.ncbi.nlm.nih.gov/pubmed/26980160 (cit. on pp. 1–3).

[8] Z. M. Younossi, A. B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, and M. Wymer. 'Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes.' In: *Hepatology* 64.1 (2016), pp. 73–84. DOI: 10.1002/hep.28431. URL: https://www.ncbi.nlm.nih.gov/pubmed/26707365 (cit. on p. 1).

[9] K. Bhaskaran, I. Douglas, H. Forbes, I. dos Santos-Silva, D. A. Leon, and L. Smeeth. 'Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults.' In: *Lancet* 384.9945 (2014), pp. 755–65. DOI: 10.1016/S0140-6736(14)60892-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/25129328 (cit. on p. 2).

[10] P. C. Valery, M. Laversanne, P. J. Clark, J. L. Petrick, K. A. McGlynn, and F. Bray. 'Projections of primary liver cancer to 2030 in 30 countries worldwide.' In: *Hepatology* 67.2 (2018), pp. 600–611. DOI: 10.1002/hep.29498. URL: https://www.ncbi.nlm.nih.gov/pubmed/28859220 (cit. on p. 2).

[11] Q. M. Anstee, H. L. Reeves, E. Kotsiliti, O. Govaere, and M. Heikenwalder. 'From NASH to HCC: current concepts and future challenges.' In: *Nat Rev Gastroenterol Hepatol* 16.7 (2019), pp. 411–428. DOI: 10.1038/s41575-019-0145-7. URL: https://www.ncbi.nlm.nih.gov/pubmed/31028350 (cit. on p. 2).

[12] B. A. Swinburn, G. Sacks, K. D. Hall, K. McPherson, D. T. Finegood, M. L. Moodie, and S. L. Gortmaker. 'The global obesity pandemic: shaped by global drivers and local environments.' In: *Lancet* 378.9793 (2011), pp. 804–14. DOI: 10.1016/S0140-6736(11)60813-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/21872749 (cit. on p. 2).

[13] K. J. Hintze, A. D. Benninghoff, C. E. Cho, and R. E. Ward. 'Modeling the Western Diet for Preclinical Investigations.' In: *Adv Nutr* 9.3 (2018), pp. 263–271. DOI:

10.1093/advances/nmy002. URL: https://www.ncbi.nlm.nih.gov/pubmed/29635305 (cit. on pp. 2, 29).

[14] I. Thiele, S. Sahoo, A. Heinken, J. Hertel, L. Heirendt, M. K. Aurich, and R. M. Fleming. 'Personalized whole-body models integrate metabolism, physiology, and the gut microbiome.' In: *Mol Syst Biol* 16.5 (2020), e8982. DOI: 10.15252/msb.20198982. URL: https://www.ncbi.nlm.nih.gov/pubmed/32463598 (cit. on pp. 2, 23, 25, 83, 89, 93).

[15] A. F. Godoy-Matos, W. S. Silva Junior, and C. M. Valerio. 'NAFLD as a continuum: from obesity to metabolic syndrome and diabetes.' In: *Diabetol Metab Syndr* 12 (2020), p. 60. DOI: 10.1186/s13098-020-00570-y. URL: https://www.ncbi.nlm.nih.gov/pubmed/32684985 (cit. on p. 3).

[16] E. Fabbrini, S. Sullivan, and S. Klein. 'Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications.' In: *Hepatology* 51.2 (2010), pp. 679–89. DOI: 10.1002/hep.23280. URL: https://www.ncbi.nlm.nih.gov/pubmed/20041406 (cit. on p. 3).

[17] Q. M. Anstee, G. Targher, and C. P. Day. 'Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis.' In: *Nat Rev Gastroenterol Hepatol* 10.6 (2013), pp. 330–44. DOI: 10.1038/nrgastro.2013.41. URL: https://www.ncbi.nlm.nih.gov/pubmed/23507799 (cit. on p. 3).

[18] R. Loomba, M. Abraham, A. Unalp, L. Wilson, J. Lavine, E. Doo, N. M. Bass, and Network Nonalcoholic Steatohepatitis Clinical Research. 'Association between diabetes, family history of diabetes, and risk of nonalcoholic steatohepatitis and fibrosis.' In: *Hepatology* 56.3 (2012), pp. 943–51. DOI: 10.1002/hep.25772. URL: https://www.ncbi.nlm.nih.gov/pubmed/22505194 (cit. on p. 3).

[19] C. D. Williams, J. Stengel, M. I. Asike, D. M. Torres, J. Shaw, M. Contreras, C. L. Landt, and S. A. Harrison. 'Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospec-

tive study.' In: *Gastroenterology* 140.1 (2011), pp. 124–31. DOI: 10.1053/j.gastro.2010.09.038. URL: https://www.ncbi.nlm.nih.gov/pubmed/20858492 (cit. on p. 3).

[20] R. Lomonaco et al. 'Metabolic Impact of Nonalcoholic Steatohepatitis in Obese Patients With Type 2 Diabetes.' In: *Diabetes Care* 39.4 (2016), pp. 632–8. DOI: 10.2337/dc15-1876. URL: https://www.ncbi.nlm.nih.gov/pubmed/26861926 (cit. on p. 3).

[21] N. L. Gluchowski, M. Becuwe, T. C. Walther, and Jr. Farese R. V. 'Lipid droplets and liver disease: from basic biology to clinical implications.' In: *Nat Rev Gastroenterol Hepatol* 14.6 (2017), pp. 343–355. DOI: 10.1038/nrgastro.2017.32. URL: https://www.ncbi.nlm.nih.gov/pubmed/28428634 (cit. on p. 4).

[22] K. L. Donnelly, C. I. Smith, S. J. Schwarzenberg, J. Jessurun, M. D. Boldt, and E. J. Parks. 'Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease.' In: *J Clin Invest* 115.5 (2005), pp. 1343–51. DOI: 10.1172/JCI23621. URL: https://www.ncbi.nlm.nih.gov/pubmed/15864352 (cit. on p. 4).

[23] J. E. Lambert, M. A. Ramos-Roman, J. D. Browning, and E. J. Parks. 'Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease.' In: *Gastroenterology* 146.3 (2014), pp. 726–35. DOI: 10.1053/j.gastro.2013.11.049. URL: https://www.ncbi.nlm.nih.gov/pubmed/24316260 (cit. on p. 4).

[24] J. Y. Cha and J. J. Repa. 'The liver X receptor (LXR) and hepatic lipogenesis. The carbohydrate-response element-binding protein is a target gene of LXR.' In: *J Biol Chem* 282.1 (2007), pp. 743–51. DOI: 10.1074/jbc.M605023200. URL: https://www.ncbi.nlm.nih.gov/pubmed/17107947 (cit. on p. 4).

[25] J. J. Repa et al. 'Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta.' In: *Genes Dev* 14.22 (2000), pp. 2819–30. DOI: 10.1101/gad.844900. URL:

https://www.ncbi.nlm.nih.gov/pubmed/11090130 (cit. on p. 4).

[26] N. Mitro, P. A. Mak, L. Vargas, C. Godio, E. Hampton, V. Molteni, A. Kreusch, and E. Saez. 'The nuclear receptor LXR is a glucose sensor.' In: *Nature* 445.7124 (2007), pp. 219–23. DOI: 10.1038/nature05449. URL: https://www.ncbi.nlm.nih.gov/pubmed/17187055 (cit. on p. 4).

[27] D. Greco et al. 'Gene expression in human NAFLD.' In: *Am J Physiol Gastrointest Liver Physiol* 294.5 (2008), G1281–7. DOI: 10.1152/ajpgi.00074.2008. URL: https://www.ncbi.nlm.nih.gov/pubmed/18388185 (cit. on p. 5).

[28] J. Westerbacka, M. Kolak, T. Kiviluoto, P. Arkkila, J. Siren, A. Hamsten, R. M. Fisher, and H. Yki-Jarvinen. 'Genes involved in fatty acid partitioning and binding, lipolysis, monocyte/macrophage recruitment, and inflammation are overexpressed in the human fatty liver of insulin-resistant subjects.' In: *Diabetes* 56.11 (2007), pp. 2759–65. DOI: 10.2337/db07-0156. URL: https://www.ncbi.nlm.nih.gov/pubmed/17704301 (cit. on p. 5).

[29] M. E. Miquilena-Colina et al. 'Hepatic fatty acid translocase CD36 upregulation is associated with insulin resistance, hyperinsulinaemia and increased steatosis in non-alcoholic steatohepatitis and chronic hepatitis C.' In: *Gut* 60.10 (2011), pp. 1394–402. DOI: 10.1136/gut.2010.222844. URL: https://www.ncbi.nlm.nih.gov/pubmed/21270117 (cit. on p. 5).

[30] J. D. McGarry and N. F. Brown. 'The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis.' In: *Eur J Biochem* 244.1 (1997), pp. 1–14. DOI: 10.1111/j.1432-1033.1997.00001.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/9063439 (cit. on p. 5).

[31] N. Fujiwara et al. 'CPT2 downregulation adapts HCC to lipid-rich environment and promotes carcinogenesis via acylcarnitine accumulation in obesity.' In: *Gut* 67.8 (2018), pp. 1493–1504. DOI: 10.1136/gutjnl-2017-315193. URL: https://www.ncbi.nlm.nih.gov/pubmed/29437870 (cit. on p. 5).

[32] F. Gonzalez-Romero et al. 'E2F1 and E2F2-Mediated Repression of CPT2 Establishes a Lipid-Rich Tumor-Promoting Environment.' In: *Cancer Res* 81.11 (2021), pp. 2874–2887. DOI: 10.1158/0008-5472.CAN-20-2052. URL: https://www.ncbi.nlm.nih.gov/pubmed/33771899 (cit. on p. 5).

[33] B. N. Reid et al. 'Hepatic overexpression of hormone-sensitive lipase and adipose triglyceride lipase promotes fatty acid oxidation, stimulates direct release of free fatty acids, and ameliorates steatosis.' In: *J Biol Chem* 283.19 (2008), pp. 13087–99. DOI: 10.1074/jbc.M800533200. URL: https://www.ncbi.nlm.nih.gov/pubmed/18337240 (cit. on p. 5).

[34] A. Chadt and H. Al-Hasani. 'Glucose transporters in adipose tissue, liver, and skeletal muscle in metabolic health and disease.' In: *Pflugers Arch* 472.9 (2020), pp. 1273–1298. DOI: 10.1007/s00424-020-02417-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/32591906 (cit. on p. 5).

[35] M. Foretz, C. Guichard, P. Ferre, and F. Foufelle. 'Sterol regulatory element binding protein-1c is a major mediator of insulin action on the hepatic expression of glucokinase and lipogenesis-related genes.' In: *Proc Natl Acad Sci U S A* 96.22 (1999), pp. 12737–42. DOI: 10.1073/pnas.96.22.12737. URL: https://www.ncbi.nlm.nih.gov/pubmed/10535992 (cit. on p. 6).

[36] S. Y. Kim, H. I. Kim, T. H. Kim, S. S. Im, S. K. Park, I. K. Lee, K. S. Kim, and Y. H. Ahn. 'SREBP-1c mediates the insulin-dependent hepatic glucokinase expression.' In: *J Biol Chem* 279.29 (2004), pp. 30823–9. DOI: 10.1074/jbc.M313223200. URL: https://www.ncbi.nlm.nih.gov/pubmed/15123649 (cit. on p. 6).

[37] M. Hatting, C. D. J. Tavares, K. Sharabi, A. K. Rines, and P. Puigserver. 'Insulin regulation of gluconeogenesis.' In: *Ann N Y Acad Sci* 1411.1 (2018), pp. 21–35. DOI: 10.1111/nyas.13435. URL: https://www.ncbi.nlm.nih.gov/pubmed/28868790 (cit. on p. 6).

[38] X. Ouyang, P. Cirillo, Y. Sautin, S. McCall, J. L. Bruchette, A. M. Diehl, R. J. Johnson, and M. F. Abdelmalek. 'Fructose consumption as a risk factor for non-alcoholic fatty liver disease.' In: *J Hepatol* 48.6 (2008), pp. 993–9. DOI: 10.1016/j.jhep.2008.02.011. URL: https://www.ncbi.nlm.nih.gov/pubmed/18395287 (cit. on p. 6).

[39] J. S. Lim, M. Mietus-Snyder, A. Valente, J. M. Schwarz, and R. H. Lustig. 'The role of fructose in the pathogenesis of NAFLD and the metabolic syndrome.' In: *Nat Rev Gastroenterol Hepatol* 7.5 (2010), pp. 251–64. DOI: 10.1038/nrgastro.2010.41. URL: https://www.ncbi.nlm.nih.gov/pubmed/20368739 (cit. on p. 6).

[40] Y. Maruhama and I. Macdonald. 'Incorporation of orally administered glucose-U-14C and fructose-U-14C into the triglyceride of liver, plasma, and adipose tissue of rats.' In: *Metabolism* 22.9 (1973), pp. 1205–15. DOI: 10.1016/0026-0495(73)90208-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/4726370 (cit. on p. 6).

[41] S. Z. Sun and M. W. Empie. 'Fructose metabolism in humans - what isotopic tracer studies tell us.' In: *Nutr Metab (Lond)* 9.1 (2012), p. 89. DOI: 10.1186/1743-7075-9-89. URL: https://www.ncbi.nlm.nih.gov/pubmed/23031075 (cit. on p. 6).

[42] D. Faeh, K. Minehira, J. M. Schwarz, R. Periasamy, S. Park, and L. Tappy. 'Effect of fructose overfeeding and fish oil administration on hepatic de novo lipogenesis and insulin sensitivity in healthy men.' In: *Diabetes* 54.7 (2005), pp. 1907–13. DOI: 10.2337/diabetes.54.7.1907. URL: https://www.ncbi.nlm.nih.gov/pubmed/15983189 (cit. on p. 6).

[43] S. Softic, D. E. Cohen, and C. R. Kahn. 'Role of Dietary Fructose and Hepatic De Novo Lipogenesis in Fatty Liver Disease.' In: *Dig Dis Sci* 61.5 (2016), pp. 1282–93. DOI: 10.1007/s10620-016-4054-0. URL: https://www.ncbi.nlm.nih.gov/pubmed/26856717 (cit. on p. 6).

[44] C. L. Cox et al. 'Consumption of fructose-sweetened beverages for 10 weeks reduces net fat oxidation and energy expenditure in overweight/obese men and women.' In:

Eur J Clin Nutr 66.2 (2012), pp. 201–8. DOI: 10.1038/ejcn.2011.159. URL: https://www.ncbi.nlm.nih.gov/pubmed/21952692 (cit. on p. 6).

[45] K. O. Raivio, A. Becker, L. J. Meyer, M. L. Greene, G. Nuki, and J. E. Seegmiller. 'Stimulation of human purine synthesis de novo by fructose infusion.' In: *Metabolism* 24.7 (1975), pp. 861–9. DOI: 10.1016/0026-0495(75)90133-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/166270 (cit. on p. 6).

[46] T. Jensen et al. 'Fructose and sugar: A major mediator of non-alcoholic fatty liver disease.' In: *J Hepatol* 68.5 (2018), pp. 1063–1075. DOI: 10.1016/j.jhep.2018.01.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/29408694 (cit. on p. 6).

[47] L. E. Olofsson and F. Backhed. 'The Metabolic Role and Therapeutic Potential of the Microbiome.' In: *Endocr Rev* (2022). DOI: 10.1210/endrev/bnac004. URL: https://www.ncbi.nlm.nih.gov/pubmed/35094076 (cit. on pp. 6, 8).

[48] A. Tripathi, J. Debelius, D. A. Brenner, M. Karin, R. Loomba, B. Schnabl, and R. Knight. 'The gut-liver axis and the intersection with the microbiome.' In: *Nat Rev Gastroenterol Hepatol* 15.7 (2018), pp. 397–411. DOI: 10.1038/s41575-018-0011-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/29748586 (cit. on pp. 6, 8, 69).

[49] N. Zmora, J. Suez, and E. Elinav. 'You are what you eat: diet, health and the gut microbiota.' In: *Nat Rev Gastroenterol Hepatol* 16.1 (2019), pp. 35–56. DOI: 10.1038/s41575-018-0061-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/30262901 (cit. on pp. 8, 25, 69).

[50] A. Wahlstrom, S. I. Sayin, H. U. Marschall, and F. Backhed. 'Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism.' In: *Cell Metab* 24.1 (2016), pp. 41–50. DOI: 10.1016/j.cmet.2016.05.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/27320064 (cit. on p. 8).

[51] S. I. Sayin et al. 'Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist.' In: *Cell Metab* 17.2 (2013), pp. 225–35. DOI: 10.1016/j.cmet.2013.01.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/23395169 (cit. on p. 8).

[52] B. Kong, L. Wang, J. Y. Chiang, Y. Zhang, C. D. Klaassen, and G. L. Guo. 'Mechanism of tissue-specific farnesoid X receptor in suppressing the expression of genes in bile-acid synthesis in mice.' In: *Hepatology* 56.3 (2012), pp. 1034–43. DOI: 10.1002/hep.25740. URL: https://www.ncbi.nlm.nih.gov/pubmed/22467244 (cit. on p. 8).

[53] C. Thomas, R. Pellicciari, M. Pruzanski, J. Auwerx, and K. Schoonjans. 'Targeting bile-acid signalling for metabolic diseases.' In: *Nat Rev Drug Discov* 7.8 (2008), pp. 678–93. DOI: 10.1038/nrd2619. URL: https://www.ncbi.nlm.nih.gov/pubmed/18670431 (cit. on p. 8).

[54] G. den Besten et al. 'Short-Chain Fatty Acids Protect Against High-Fat Diet-Induced Obesity via a PPARgamma-Dependent Switch From Lipogenesis to Fat Oxidation.' In: *Diabetes* 64.7 (2015), pp. 2398–408. DOI: 10.2337/db14-1213. URL: https://www.ncbi.nlm.nih.gov/pubmed/25695945 (cit. on p. 8).

[55] V. Singh et al. 'Microbiota-Dependent Hepatic Lipogenesis Mediated by Stearoyl CoA Desaturase 1 (SCD1) Promotes Metabolic Syndrome in TLR5-Deficient Mice.' In: *Cell Metab* 22.6 (2015), pp. 983–96. DOI: 10.1016/j.cmet.2015.09.028. URL: https://www.ncbi.nlm.nih.gov/pubmed/26525535 (cit. on p. 8).

[56] Z. Gao, J. Yin, J. Zhang, R. E. Ward, R. J. Martin, M. Lefevre, W. T. Cefalu, and J. Ye. 'Butyrate improves insulin sensitivity and increases energy expenditure in mice.' In: *Diabetes* 58.7 (2009), pp. 1509–17. DOI: 10.2337/db08-1637. URL: https://www.ncbi.nlm.nih.gov/pubmed/19366864 (cit. on p. 8).

[57] C. Jang et al. 'The Small Intestine Converts Dietary Fructose into Glucose and Organic Acids.' In: *Cell Metab* 27.2 (2018), 351–361 e3. DOI: 10.1016/j.cmet.2017.12.016.

URL: https://www.ncbi.nlm.nih.gov/pubmed/29414685 (cit. on p. 8).

[58] S. R. Taylor et al. 'Dietary fructose improves intestinal cell survival and nutrient absorption.' In: *Nature* 597.7875 (2021), pp. 263–267. DOI: 10.1038/s41586-021-03827-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/34408323 (cit. on p. 9).

[59] S. Thuy, R. Ladurner, V. Volynets, S. Wagner, S. Strahl, A. Konigsrainer, K. P. Maier, S. C. Bischoff, and I. Bergheim. 'Nonalcoholic fatty liver disease in humans is associated with increased plasma endotoxin and plasminogen activator inhibitor 1 concentrations and with fructose intake.' In: *J Nutr* 138.8 (2008), pp. 1452–5. DOI: 10.1093/jn/138.8.1452. URL: https://www.ncbi.nlm.nih.gov/pubmed/18641190 (cit. on p. 9).

[60] S. Zhao et al. 'Dietary fructose feeds hepatic lipogenesis via microbiota-derived acetate.' In: *Nature* 579.7800 (2020), pp. 586–591. DOI: 10.1038/s41586-020-2101-7. URL: https://www.ncbi.nlm.nih.gov/pubmed/32214246 (cit. on p. 9).

[61] I. Martinez-Reyes and N. S. Chandel. 'Cancer metabolism: looking forward.' In: *Nat Rev Cancer* 21.10 (2021), pp. 669–680. DOI: 10.1038/s41568-021-00378-6. URL: https://www.ncbi.nlm.nih.gov/pubmed/34272515 (cit. on p. 9).

[62] M. Yang and K. H. Vousden. 'Serine and one-carbon metabolism in cancer.' In: *Nat Rev Cancer* 16.10 (2016), pp. 650–62. DOI: 10.1038/nrc.2016.81. URL: https://www.ncbi.nlm.nih.gov/pubmed/27634448 (cit. on pp. 9, 65).

[63] D. DeWaal et al. 'Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin.' In: *Nat Commun* 9.1 (2018), p. 446. DOI: 10.1038/s41467-017-02733-4. URL: https://www.ncbi.nlm.nih.gov/pubmed/29386513 (cit. on p. 9).

[64] T. E. Li et al. 'PKM2 Drives Hepatocellular Carcinoma Progression by Inducing Immunosuppressive Microenvironment.' In: *Front Immunol* 11 (2020), p. 589997. DOI: 10.3389/fimmu.2020.589997. URL: https://www.ncbi.nlm.nih.gov/pubmed/33193421 (cit. on p. 10).

[65] W. W. Lv, D. Liu, X. C. Liu, T. N. Feng, L. Li, B. Y. Qian, and W. X. Li. 'Effects of PKM2 on global metabolic changes and prognosis in hepatocellular carcinoma: from gene expression to drug discovery.' In: *BMC Cancer* 18.1 (2018), p. 1150. DOI: 10.1186/s12885-018-5023-0. URL: https://www.ncbi.nlm.nih.gov/pubmed/30463528 (cit. on p. 10).

[66] D. Anastasiou. 'Tumour microenvironment factors shaping the cancer metabolism landscape.' In: *Br J Cancer* 116.3 (2017), pp. 277–286. DOI: 10.1038/bjc.2016.412. URL: https://www.ncbi.nlm.nih.gov/pubmed/28006817 (cit. on pp. 10, 83, 86, 89).

[67] A. Mendez-Lucas et al. 'Identifying strategies to target the metabolic flexibility of tumours.' In: *Nat Metab* 2.4 (2020), pp. 335–350. DOI: 10.1038/s42255-020-0195-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/32694609 (cit. on pp. 10, 65).

[68] B. Palsson. 'The challenges of in silico biology.' In: *Nat Biotechnol* 18.11 (2000), pp. 1147–50. DOI: 10.1038/81125. URL: https://www.ncbi.nlm.nih.gov/pubmed/11062431 (cit. on p. 12).

[69] R. Aebersold, L. E. Hood, and J. D. Watts. 'Equipping scientists for the new biology.' In: *Nat Biotechnol* 18.4 (2000), p. 359. DOI: 10.1038/74325. URL: https://www.ncbi.nlm.nih.gov/pubmed/10748470 (cit. on p. 12).

[70] B. O. Palsson. 'What lies beyond bioinformatics?' In: *Nat Biotechnol* 15.1 (1997), pp. 3–4. DOI: 10.1038/nbt0197-3. URL: https://www.ncbi.nlm.nih.gov/pubmed/9035092 (cit. on p. 12).

[71] R. C. Strohman. 'The coming Kuhnian revolution in biology.' In: *Nat Biotechnol* 15.3 (1997), pp. 194–200. DOI: 10.1038/nbt0397-194. URL: https://www.ncbi.nlm.nih.gov/pubmed/9062910 (cit. on p. 12).

[72] T. Ideker, T. Galitski, and L. Hood. 'A new approach to decoding life: systems biology.' In: *Annu Rev Genomics Hum Genet* 2 (2001), pp. 343–72. DOI: 10.1146/annurev.genom.2.1.343. URL: https://www.ncbi.nlm.nih.gov/pubmed/11701654 (cit. on p. 12).

[73] Bernhard O. Palsson. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge university press, 2015 (cit. on pp. 12, 20).

[74] I. Thiele and B. O. Palsson. 'A protocol for generating a high-quality genome-scale metabolic reconstruction.' In: *Nat Protoc* 5.1 (2010), pp. 93–121. DOI: 10.1038/nprot.2009.203. URL: https://www.ncbi.nlm.nih.gov/pubmed/20057383 (cit. on pp. 12, 13, 15, 45).

[75] R. D. Fleischmann et al. 'Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.' In: *Science* 269.5223 (1995), pp. 496–512. DOI: 10.1126/science.7542800. URL: https://www.ncbi.nlm.nih.gov/pubmed/7542800 (cit. on p. 13).

[76] J. S. Edwards and B. O. Palsson. 'Systems properties of the Haemophilus influenzae Rd metabolic genotype.' In: *J Biol Chem* 274.25 (1999), pp. 17410–6. DOI: 10.1074/jbc.274.25.17410. URL: https://www.ncbi.nlm.nih.gov/pubmed/10364169 (cit. on p. 13).

[77] J. Schellenberger, J. O. Park, T. M. Conrad, and B. O. Palsson. 'BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions.' In: *BMC Bioinformatics* 11 (2010), p. 213. DOI: 10.1186/1471-2105-11-213. URL: https://www.ncbi.nlm.nih.gov/pubmed/20426874 (cit. on pp. 13, 41).

[78] J. Schellenberger et al. 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.' In: *Nat Protoc* 6.9 (2011), pp. 1290–307. DOI: 10.1038/nprot.2011.308. URL: https://www.ncbi.nlm.nih.gov/pubmed/21886097 (cit. on p. 14).

[79] A. M. Feist and B. O. Palsson. 'The biomass objective function.' In: *Curr Opin Microbiol* 13.3 (2010), pp. 344–9. DOI: 10.1016/j.mib.2010.03.003. URL: https://www.ncbi.nlm.nih.gov/pubmed/20430689 (cit. on pp. 15, 23).

[80] N. D. Price, J. A. Papin, C. H. Schilling, and B. O. Palsson. 'Genome-scale microbial in silico models: the constraints-based approach.' In: *Trends Biotechnol* 21.4 (2003), pp. 162–9. DOI: 10.1016/S0167-7799(03)00030-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/12679064 (cit. on pp. 15, 18, 66).

[81] A. Bordbar, J. M. Monk, Z. A. King, and B. O. Palsson. 'Constraint-based models predict metabolic and associated cellular functions.' In: *Nat Rev Genet* 15.2 (2014), pp. 107–20. DOI: 10.1038/nrg3643. URL: https://www.ncbi.nlm.nih.gov/pubmed/24430943 (cit. on pp. 15, 18, 21, 66, 92).

[82] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. O. Palsson. 'Global reconstruction of the human metabolic network based on genomic and bibliomic data.' In: *Proc Natl Acad Sci U S A* 104.6 (2007), pp. 1777–82. DOI: 10.1073/pnas.0610772104. URL: https://www.ncbi.nlm.nih.gov/pubmed/17267599 (cit. on p. 16).

[83] H. Ma, A. Sorokin, A. Mazein, A. Selkov, E. Selkov, O. Demin, and I. Goryanin. 'The Edinburgh human metabolic network reconstruction and its functional analysis.' In: *Mol Syst Biol* 3 (2007), p. 135. DOI: 10.1038/msb4100177. URL: https://www.ncbi.nlm.nih.gov/pubmed/17882155 (cit. on p. 16).

[84] I. Thiele et al. 'A community-driven global reconstruction of human metabolism.' In: *Nat Biotechnol* 31.5 (2013), pp. 419–25. DOI: 10.1038/nbt.2488. URL: https://www.ncbi.nlm.nih.gov/pubmed/23455439 (cit. on p. 16).

[85] E. Brunk et al. 'Recon3D enables a three-dimensional view of gene variation in human metabolism.' In: *Nat Biotechnol* 36.3 (2018), pp. 272–281. DOI: 10.1038/nbt.4072. URL: https://www.ncbi.nlm.nih.gov/pubmed/29457794 (cit. on pp. 16, 33).

[86] R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew, and J. Nielsen. 'Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT.' In: *PLoS Comput Biol* 8.5

(2012), e1002518. DOI: 10.1371/journal.pcbi.1002518. URL: https://www.ncbi.nlm.nih.gov/pubmed/22615553 (cit. on pp. 16, 67).

[87] A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, and J. Nielsen. 'Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease.' In: *Nat Commun* 5 (2014), p. 3083. DOI: 10.1038/ncomms4083. URL: https://www.ncbi.nlm.nih.gov/pubmed/24419221 (cit. on pp. 16, 33, 93).

[88] C. Gille et al. 'HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology.' In: *Mol Syst Biol* 6 (2010), p. 411. DOI: 10.1038/msb.2010.62. URL: https://www.ncbi.nlm.nih.gov/pubmed/20823849 (cit. on p. 16).

[89] J. Y. Ryu, H. U. Kim, and S. Y. Lee. 'Framework and resource for more than 11,000 gene-transcript-protein-reaction associations in human metabolism.' In: *Proc Natl Acad Sci U S A* 114.45 (2017), E9740–E9749. DOI: 10.1073/pnas.1713050114. URL: https://www.ncbi.nlm.nih.gov/pubmed/29078384 (cit. on p. 16).

[90] A. Mardinoglu et al. 'Integration of clinical data with a genome-scale metabolic model of the human adipocyte.' In: *Mol Syst Biol* 9 (2013), p. 649. DOI: 10.1038/msb.2013.5. URL: https://www.ncbi.nlm.nih.gov/pubmed/23511207 (cit. on pp. 16, 93).

[91] L. Varemo et al. 'Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes.' In: *Cell Rep* 11.6 (2015), pp. 921–933. DOI: 10.1016/j.celrep.2015.04.010. URL: https://www.ncbi.nlm.nih.gov/pubmed/25937284 (cit. on pp. 16, 93).

[92] J. L. Robinson et al. 'An atlas of human metabolism.' In: *Sci Signal* 13.624 (2020). DOI: 10.1126/scisignal.aaz1482. URL: https://www.ncbi.nlm.nih.gov/pubmed/32209698 (cit. on pp. 16, 17, 35, 41, 46, 91).

[93] K. Sheikh, J. Forster, and L. K. Nielsen. 'Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus.' In: *Biotechnol Prog* 21.1 (2005), pp. 112–21. DOI: 10.1021/bp0498138. URL: https://www.ncbi.nlm.nih.gov/pubmed/15903248 (cit. on p. 16).

[94] S. Selvarasu, V. V. Wong, I. A. Karimi, and D. Y. Lee. 'Elucidation of metabolism in hybridoma cells grown in fed-batch culture by genome-scale modeling.' In: *Biotechnol Bioeng* 102.5 (2009), pp. 1494–504. DOI: 10.1002/bit.22186. URL: https://www.ncbi.nlm.nih.gov/pubmed/19048615 (cit. on p. 18).

[95] S. Selvarasu, I. A. Karimi, G. H. Ghim, and D. Y. Lee. 'Genome-scale modeling and in silico analysis of mouse cell metabolic network.' In: *Mol Biosyst* 6.1 (2010), pp. 152–61. DOI: 10.1039/b912865d. URL: https://www.ncbi.nlm.nih.gov/pubmed/20024077 (cit. on p. 18).

[96] M. I. Sigurdsson, N. Jamshidi, E. Steingrimsson, I. Thiele, and B. O. Palsson. 'A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1.' In: *BMC Syst Biol* 4 (2010), p. 140. DOI: 10.1186/1752-0509-4-140. URL: https://www.ncbi.nlm.nih.gov/pubmed/20959003 (cit. on pp. 18, 41).

[97] A. Mardinoglu, S. Shoaie, M. Bergentall, P. Ghaffari, C. Zhang, E. Larsson, F. Backhed, and J. Nielsen. 'The gut microbiota modulates host amino acid and glutathione metabolism in mice.' In: *Mol Syst Biol* 11.10 (2015), p. 834. DOI: 10.15252/msb.20156487. URL: https://www.ncbi.nlm.nih.gov/pubmed/26475342 (cit. on pp. 18, 22, 35, 41, 46, 67, 89, 91).

[98] S. Khodaee, Y. Asgari, M. Totonchi, and M. H. Karimi-Jafari. 'iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model.' In: *Sci Rep* 10.1 (2020), p. 6177. DOI: 10.1038/s41598-020-63235-w. URL: https://www.ncbi.nlm.nih.gov/pubmed/32277147 (cit. on pp. 18, 41, 91).

[99] H. Wang et al. 'Genome-scale metabolic network reconstruction of model animals as a platform for translational research.' In: *Proc Natl Acad Sci U S A* 118.30 (2021). DOI: 10.1073/pnas.2102344118. URL: https://www.ncbi.nlm.nih.gov/pubmed/34282017 (cit. on pp. 18, 35).

[100] N. E. Lewis, H. Nagarajan, and B. O. Palsson. 'Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.' In: *Nat Rev Microbiol* 10.4 (2012), pp. 291–305. DOI: 10.1038/nrmicro2737. URL: https://www.ncbi.nlm.nih.gov/pubmed/22367118 (cit. on pp. 18, 22, 92).

[101] J. D. Orth, I. Thiele, and B. O. Palsson. 'What is flux balance analysis?' In: *Nat Biotechnol* 28.3 (2010), pp. 245–8. DOI: 10.1038/nbt.1614. URL: https://www.ncbi.nlm.nih.gov/pubmed/20212490 (cit. on pp. 18, 23).

[102] D. R. Hyduke, N. E. Lewis, and B. O. Palsson. 'Analysis of omics data with genome-scale models of metabolism.' In: *Mol Biosyst* 9.2 (2013), pp. 167–74. DOI: 10.1039/c2mb25453k. URL: https://www.ncbi.nlm.nih.gov/pubmed/23247105 (cit. on p. 21).

[103] N. Jamshidi and B. O. Palsson. 'Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models.' In: *Biophys J* 98.2 (2010), pp. 175–85. DOI: 10.1016/j.bpj.2009.09.064. URL: https://www.ncbi.nlm.nih.gov/pubmed/20338839 (cit. on p. 22).

[104] B. J. Sanchez, C. Zhang, A. Nilsson, P. J. Lahtvee, E. J. Kerkhoven, and J. Nielsen. 'Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints.' In: *Mol Syst Biol* 13.8 (2017), p. 935. DOI: 10.15252/msb.20167411. URL: https://www.ncbi.nlm.nih.gov/pubmed/28779005 (cit. on p. 22).

[105] C. Ramon, M. G. Gollub, and J. Stelling. 'Integrating -omics data into genome-scale metabolic network models: principles and challenges.' In: *Essays Biochem* 62.4 (2018), pp. 563–574. DOI: 10.1042/EBC20180011. URL: https:

//www.ncbi.nlm.nih.gov/pubmed/30315095 (cit. on p. 22).

[106] G. Marinos, C. Kaleta, and S. Waschina. 'Defining the nutritional input for genome-scale metabolic models: A roadmap.' In: *PLoS One* 15.8 (2020), e0236890. DOI: 10.1371/journal.pone.0236890. URL: https://www.ncbi.nlm.nih.gov/pubmed/32797084 (cit. on pp. 22, 67).

[107] J. Monk, J. Nogales, and B. O. Palsson. 'Optimizing genome-scale network reconstructions.' In: *Nat Biotechnol* 32.5 (2014), pp. 447–52. DOI: 10.1038/nbt.2870. URL: https://www.ncbi.nlm.nih.gov/pubmed/24811519 (cit. on p. 23).

[108] I. Thiele and B. O. Palsson. 'A protocol for generating a high-quality genome-scale metabolic reconstruction.' In: *Nat Protoc* 5.1 (2010), pp. 93–121. DOI: 10.1038/nprot.2009.203. URL: https://www.ncbi.nlm.nih.gov/pubmed/20057383 (cit. on pp. 23, 45, 91).

[109] S. Stolyar, S. Van Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, and D. A. Stahl. 'Metabolic modeling of a mutualistic microbial community.' In: *Mol Syst Biol* 3 (2007), p. 92. DOI: 10.1038/msb4100131. URL: https://www.ncbi.nlm.nih.gov/pubmed/17353934 (cit. on p. 24).

[110] E. Le Chatelier et al. 'Richness of human gut microbiome correlates with metabolic markers.' In: *Nature* 500.7464 (2013), pp. 541–6. DOI: 10.1038/nature12506. URL: https://www.ncbi.nlm.nih.gov/pubmed/23985870 (cit. on pp. 25, 33, 69, 81).

[111] S. Shoaie et al. 'Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome.' In: *Cell Metab* 22.2 (2015), pp. 320–31. DOI: 10.1016/j.cmet.2015.07.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/26244934 (cit. on pp. 25, 81).

[112] S. Magnusdottir et al. 'Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota.' In: *Nat Biotechnol* 35.1 (2017), pp. 81–89. DOI: 10.1038/nbt.3703. URL: https://www.ncbi.nlm.nih.gov/pubmed/27893703 (cit. on pp. 25, 37).

[113] A. Bordbar, A. M. Feist, R. Usaite-Black, J. Woodcock, B. O. Palsson, and I. Famili. 'A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology.' In: *BMC Syst Biol* 5 (2011), p. 180. DOI: 10.1186/1752-0509-5-180. URL: https://www.ncbi.nlm.nih.gov/pubmed/22041191 (cit. on pp. 25, 83, 89, 93, 94).

[114] P. Martins Conde, T. Pfau, M. Pires Pacheco, and T. Sauter. 'A dynamic multi-tissue model to study human metabolism.' In: *NPJ Syst Biol Appl* 7.1 (2021), p. 5. DOI: 10.1038/s41540-020-00159-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/33483512 (cit. on pp. 25, 83, 89, 93, 94).

[115] E. J. Park, J. H. Lee, G. Y. Yu, G. He, S. R. Ali, R. G. Holzer, C. H. Osterreicher, H. Takahashi, and M. Karin. 'Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression.' In: *Cell* 140.2 (2010), pp. 197–208. DOI: 10.1016/j.cell.2009.12.052. URL: https://www.ncbi.nlm.nih.gov/pubmed/20141834 (cit. on pp. 29, 49, 52).

[116] P. M. Nunes, A. J. Wright, A. Veltien, J. J. van Asten, C. J. Tack, J. G. Jones, and A. Heerschap. 'Dietary lipids do not contribute to the higher hepatic triglyceride levels of fructose- compared to glucose-fed mice.' In: *FASEB J* 28.5 (2014), pp. 1988–97. DOI: 10.1096/fj.13-241208. URL: https://www.ncbi.nlm.nih.gov/pubmed/24500922 (cit. on p. 31).

[117] J. A. Duarte, F. Carvalho, M. Pearson, J. D. Horton, J. D. Browning, J. G. Jones, and S. C. Burgess. 'A high-fat diet suppresses de novo lipogenesis and desaturation but not elongation and triglyceride synthesis in mice.' In: *J Lipid Res* 55.12 (2014), pp. 2541–53. DOI: 10.1194/jlr.M052308. URL: https://www.ncbi.nlm.nih.gov/pubmed/25271296 (cit. on p. 31).

[118] J. Folch, M. Lees, and G. H. Sloane Stanley. 'A simple method for the isolation and purification of total lipides from animal tissues.' In: *J Biol Chem* 226.1 (1957), pp. 497–509. URL: https://www.ncbi.nlm.nih.gov/pubmed/13428781 (cit. on p. 32).

[119] J. G. Jones, R. Perdigoto, T. B. Rodrigues, and C. F. Geraldes. 'Quantitation of absolute 2H enrichment of plasma glucose by 2H NMR analysis of its monoacetone derivative.' In: *Magn Reson Med* 48.3 (2002), pp. 535–9. DOI: 10.1002/mrm.10234. URL: https://www.ncbi.nlm.nih.gov/pubmed/12210920 (cit. on p. 32).

[120] L. Heirendt et al. 'Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.' In: *Nat Protoc* 14.3 (2019), pp. 639–702. DOI: 10.1038/s41596-018-0098-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/30787451 (cit. on pp. 32, 41).

[121] H. Wang, S. Marcisauskas, B. J. Sanchez, I. Domenzain, D. Hermansson, R. Agren, J. Nielsen, and E. J. Kerkhoven. 'RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor.' In: *PLoS Comput Biol* 14.10 (2018), e1006541. DOI: 10.1371/journal.pcbi.1006541. URL: https://www.ncbi.nlm.nih.gov/pubmed/30335785 (cit. on pp. 32, 33, 41).

[122] B. Li and C. N. Dewey. 'RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.' In: *BMC Bioinformatics* 12 (2011), p. 323. DOI: 10.1186/1471-2105-12-323. URL: https://www.ncbi.nlm.nih.gov/pubmed/21816040 (cit. on p. 32).

[123] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 'STAR: ultrafast universal RNA-seq aligner.' In: *Bioinformatics* 29.1 (2013), pp. 15–21. DOI: 10.1093/bioinformatics/bts635. URL: https://www.ncbi.nlm.nih.gov/pubmed/23104886 (cit. on p. 32).

[124] L. Wang, S. Wang, and W. Li. 'RSeQC: quality control of RNA-seq experiments.' In: *Bioinformatics* 28.16 (2012), pp. 2184–5. DOI: 10.1093/bioinformatics/bts356. URL: https://www.ncbi.nlm.nih.gov/pubmed/22743226 (cit. on p. 32).

[125] D. S. DeLuca, J. Z. Levin, A. Sivachenko, T. Fennell, M. D. Nazaire, C. Williams, M. Reich, W. Winckler, and G. Getz. 'RNA-SeQC: RNA-seq metrics for quality control

and process optimization.' In: *Bioinformatics* 28.11 (2012), pp. 1530–2. DOI: 10.1093/bioinformatics/bts196. URL: https://www.ncbi.nlm.nih.gov/pubmed/22539670 (cit. on p. 32).

[126] P. Ewels, M. Magnusson, S. Lundin, and M. Kaller. 'MultiQC: summarize analysis results for multiple tools and samples in a single report.' In: *Bioinformatics* 32.19 (2016), pp. 3047–8. DOI: 10.1093/bioinformatics/btw354. URL: https://www.ncbi.nlm.nih.gov/pubmed/27312411 (cit. on p. 32).

[127] M. I. Love, W. Huber, and S. Anders. 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.' In: *Genome Biol* 15.12 (2014), p. 550. DOI: 10.1186/s13059-014-0550-8. URL: https://www.ncbi.nlm.nih.gov/pubmed/25516281 (cit. on p. 33).

[128] A. Zhu, J. G. Ibrahim, and M. I. Love. 'Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences.' In: *Bioinformatics* 35.12 (2019), pp. 2084–2092. DOI: 10.1093/bioinformatics/bty895. URL: https://www.ncbi.nlm.nih.gov/pubmed/30395178 (cit. on p. 33).

[129] G. Yu, L. G. Wang, Y. Han, and Q. Y. He. 'clusterProfiler: an R package for comparing biological themes among gene clusters.' In: *OMICS* 16.5 (2012), pp. 284–7. DOI: 10.1089/omi.2011.0118. URL: https://www.ncbi.nlm.nih.gov/pubmed/22455463 (cit. on p. 33).

[130] G. Yu, L. G. Wang, G. R. Yan, and Q. Y. He. 'DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis.' In: *Bioinformatics* 31.4 (2015), pp. 608–9. DOI: 10.1093/bioinformatics/btu684. URL: https://www.ncbi.nlm.nih.gov/pubmed/25677125 (cit. on p. 33).

[131] L. Xiao et al. 'A catalog of the mouse gut metagenome.' In: *Nat Biotechnol* 33.10 (2015), pp. 1103–8. DOI: 10.1038/nbt.3353. URL: https://www.ncbi.nlm.nih.gov/pubmed/26414350 (cit. on p. 33).

[132] R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen, and J. Nielsen. 'Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling.' In: *Mol Syst Biol* 10 (2014), p. 721. DOI: 10.1002/msb.145122. URL: https://www.ncbi.nlm.nih.gov/pubmed/24646661 (cit. on pp. 34, 35, 67).

[133] S. Khodaee, Y. Asgari, M. Totonchi, and M. H. Karimi-Jafari. 'iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model.' In: *Sci Rep* 10.1 (2020), p. 6177. DOI: 10.1038/s41598-020-63235-w. URL: https://www.ncbi.nlm.nih.gov/pubmed/32277147 (cit. on pp. 35, 46).

[134] C. Colijn et al. 'Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production.' In: *PLoS Comput Biol* 5.8 (2009), e1000489. DOI: 10.1371/journal.pcbi.1000489. URL: https://www.ncbi.nlm.nih.gov/pubmed/19714220 (cit. on pp. 35, 67).

[135] H. Wang et al. 'Genome-scale metabolic network reconstruction of model animals as a platform for translational research.' In: *Proc Natl Acad Sci U S A* 118.30 (2021). DOI: 10.1073/pnas.2102344118. URL: https://www.ncbi.nlm.nih.gov/pubmed/34282017 (cit. on pp. 41, 91).

[136] D. B. Bernstein, S. Sulheim, E. Almaas, and D. Segre. 'Addressing uncertainty in genome-scale metabolic model reconstruction and analysis.' In: *Genome Biol* 22.1 (2021), p. 64. DOI: 10.1186/s13059-021-02289-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/33602294 (cit. on pp. 44, 93).

[137] M. A. Carey, A. Drager, M. E. Beber, J. A. Papin, and J. T. Yurkovich. 'Community standards to facilitate development and address challenges in metabolic modeling.' In: *Mol Syst Biol* 16.8 (2020), e9235. DOI: 10.15252/msb.20199235. URL: https://www.ncbi.nlm.nih.gov/pubmed/32845080 (cit. on p. 44).

[138] M. I. Sigurdsson, N. Jamshidi, E. Steingrimsson, I. Thiele, and B. O. Palsson. 'A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1.' In:

BMC Syst Biol 4 (2010), p. 140. DOI: 10.1186/1752-0509-4-140. URL: https://www.ncbi.nlm.nih.gov/pubmed/20959003 (cit. on p. 46).

[139] W. Palm and C. B. Thompson. 'Nutrient acquisition strategies of mammalian cells.' In: *Nature* 546.7657 (2017), pp. 234–242. DOI: 10.1038/nature22379. URL: https://www.ncbi.nlm.nih.gov/pubmed/28593971 (cit. on pp. 49, 92).

[140] L. Zhao, C. Deng, Z. Lin, E. Giovannucci, and X. Zhang. 'Dietary Fats, Serum Cholesterol and Liver Cancer Risk: A Systematic Review and Meta-Analysis of Prospective Studies.' In: *Cancers (Basel)* 13.7 (2021). DOI: 10.3390/cancers13071580. URL: https://www.ncbi.nlm.nih.gov/pubmed/33808094 (cit. on p. 49).

[141] W. S. Yang, X. F. Zeng, Z. N. Liu, Q. H. Zhao, Y. T. Tan, J. Gao, H. L. Li, and Y. B. Xiang. 'Diet and liver cancer risk: a narrative review of epidemiological evidence.' In: *Br J Nutr* 124.3 (2020), pp. 330–340. DOI: 10.1017/S0007114520001208. URL: https://www.ncbi.nlm.nih.gov/pubmed/32234090 (cit. on pp. 49, 92).

[142] Y. Duan, L. Zeng, C. Zheng, B. Song, F. Li, X. Kong, and K. Xu. 'Inflammatory Links Between High Fat Diets and Diseases.' In: *Front Immunol* 9 (2018), p. 2649. DOI: 10.3389/fimmu.2018.02649. URL: https://www.ncbi.nlm.nih.gov/pubmed/30483273 (cit. on p. 49).

[143] G. A. Bray, S. J. Nielsen, and B. M. Popkin. 'Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity.' In: *Am J Clin Nutr* 79.4 (2004), pp. 537–43. DOI: 10.1093/ajcn/79.4.537. URL: https://www.ncbi.nlm.nih.gov/pubmed/15051594 (cit. on p. 49).

[144] M. T. Timlin, B. R. Barrows, and E. J. Parks. 'Increased dietary substrate delivery alters hepatic fatty acid recycling in healthy men.' In: *Diabetes* 54.9 (2005), pp. 2694–701. DOI: 10.2337/diabetes.54.9.2694. URL: https://www.ncbi.nlm.nih.gov/pubmed/16123359 (cit. on p. 49).

[145] M. Grima-Reyes, A. Martinez-Turtos, I. Abramovich, E. Gottlieb, J. Chiche, and J. E. Ricci. 'Physiological impact of in vivo stable isotope tracing on cancer metabolism.' In: *Mol Metab* 53 (2021), p. 101294. DOI: 10.1016/j.molmet.2021.101294. URL: https://www.ncbi.nlm.nih.gov/pubmed/34256164 (cit. on p. 50).

[146] B. Faubert, A. Tasdogan, S. J. Morrison, T. P. Mathews, and R. J. DeBerardinis. 'Stable isotope tracing to assess tumor metabolism in vivo.' In: *Nat Protoc* 16.11 (2021), pp. 5123–5145. DOI: 10.1038/s41596-021-00605-2. URL: https://www.ncbi.nlm.nih.gov/pubmed/34535790 (cit. on p. 50).

[147] S. Opdam, A. Richelle, B. Kellman, S. Li, D. C. Zielinski, and N. E. Lewis. 'A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models.' In: *Cell Syst* 4.3 (2017), 318–329 e6. DOI: 10.1016/j.cels.2017.01.010. URL: https://www.ncbi.nlm.nih.gov/pubmed/28215528 (cit. on p. 66).

[148] P. A. Jensen and J. A. Papin. 'Functional integration of a metabolic network model and expression data without arbitrary thresholding.' In: *Bioinformatics* 27.4 (2011), pp. 541–7. DOI: 10.1093/bioinformatics/btq702. URL: https://www.ncbi.nlm.nih.gov/pubmed/21172910 (cit. on p. 67).

[149] C. Ma et al. 'Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells.' In: *Science* 360.6391 (2018). DOI: 10.1126/science.aan5931. URL: https://www.ncbi.nlm.nih.gov/pubmed/29798856 (cit. on p. 69).

[150] L. K. Brahe et al. 'Specific gut microbiota features and metabolic markers in postmenopausal women with obesity.' In: *Nutr Diabetes* 5 (2015), e159. DOI: 10.1038/nutd.2015.9. URL: https://www.ncbi.nlm.nih.gov/pubmed/26075636 (cit. on p. 80).

[151] J. Xin, D. Zeng, H. Wang, X. Ni, D. Yi, K. Pan, and B. Jing. 'Preventing non-alcoholic fatty liver disease through Lactobacillus johnsonii BS15 by attenuating inflammation and mitochondrial injury and improving gut environment in obese mice.' In: *Appl Microbiol Biotechnol* 98.15

(2014), pp. 6817–29. DOI: 10.1007/s00253-014-5752-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/24811405 (cit. on p. 80).

[152] S. K. A. Raftar et al. 'The anti-inflammatory effects of Akkermansia muciniphila and its derivates in HFD/CCL4-induced murine model of liver injury.' In: *Sci Rep* 12.1 (2022), p. 2453. DOI: 10.1038/s41598-022-06414-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/35165344 (cit. on p. 80).

[153] Y. Rao et al. 'Gut Akkermansia muciniphila ameliorates metabolic dysfunction-associated fatty liver disease by regulating the metabolism of L-aspartate via gut-liver axis.' In: *Gut Microbes* 13.1 (2021), pp. 1–19. DOI: 10.1080/19490976.2021.1927633. URL: https://www.ncbi.nlm.nih.gov/pubmed/34030573 (cit. on p. 80).

[154] W. Wu, L. Lv, D. Shi, J. Ye, D. Fang, F. Guo, Y. Li, X. He, and L. Li. 'Protective Effect of Akkermansia muciniphila against Immune-Mediated Liver Injury in a Mouse Model.' In: *Front Microbiol* 8 (2017), p. 1804. DOI: 10.3389/fmicb.2017.01804. URL: https://www.ncbi.nlm.nih.gov/pubmed/29033903 (cit. on p. 80).

[155] H. N. Bell et al. 'Reuterin in the healthy gut microbiome suppresses colorectal cancer growth through altering redox balance.' In: *Cancer Cell* 40.2 (2022), 185–200 e6. DOI: 10.1016/j.ccell.2021.12.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/34951957 (cit. on p. 81).

[156] J. Zhang, S. Sturla, C. Lacroix, and C. Schwab. 'Gut Microbial Glycerol Metabolism as an Endogenous Acrolein Source.' In: *mBio* 9.1 (2018). DOI: 10.1128/mBio.01947-17. URL: https://www.ncbi.nlm.nih.gov/pubmed/29339426 (cit. on p. 81).

[157] K. L. Greathouse, M. Wyatt, A. J. Johnson, E. P. Toy, J. M. Khan, K. Dunn, D. J. Clegg, and S. Reddy. 'Diet-microbiome interactions in cancer treatment: Opportunities and challenges for precision nutrition in cancer.' In: *Neoplasia* 29 (2022), p. 100800. DOI: 10.1016/j.neo.2022.100800. URL: https://www.ncbi.nlm.nih.gov/pubmed/35500546 (cit. on pp. 81, 93).

[158] I. Elia and M. C. Haigis. 'Metabolites and the tumour microenvironment: from cellular mechanisms to systemic metabolism.' In: *Nat Metab* 3.1 (2021), pp. 21–32. DOI: 10.1038/s42255-020-00317-z. URL: https://www.ncbi.nlm.nih.gov/pubmed/33398194 (cit. on p. 83).

[159] M. Arif et al. 'Integrative transcriptomic analysis of tissue-specific metabolic crosstalk after myocardial infarction.' In: *Elife* 10 (2021). DOI: 10.7554/eLife.66921. URL: https://www.ncbi.nlm.nih.gov/pubmed/33972017 (cit. on p. 83).

[160] T. L. Whiteside. 'The tumor microenvironment and its role in promoting tumor growth.' In: *Oncogene* 27.45 (2008), pp. 5904–12. DOI: 10.1038/onc.2008.271. URL: https://www.ncbi.nlm.nih.gov/pubmed/18836471 (cit. on p. 89).

[161] L. Zhang et al. 'Creatine promotes cancer metastasis through activation of Smad2/3.' In: *Cell Metab* 33.6 (2021), 1111–1123 e4. DOI: 10.1016/j.cmet.2021.03.009. URL: https://www.ncbi.nlm.nih.gov/pubmed/33811821 (cit. on p. 89).

[162] I. Kurth et al. 'Therapeutic targeting of SLC6A8 creatine transporter suppresses colon cancer progression and modulates human creatine levels.' In: *Sci Adv* 7.41 (2021), eabi7511. DOI: 10.1126/sciadv.abi7511. URL: https://www.ncbi.nlm.nih.gov/pubmed/34613776 (cit. on p. 89).

[163] Z. Ding, R. E. Ericksen, Q. Y. Lee, and W. Han. 'Reprogramming of mitochondrial proline metabolism promotes liver tumorigenesis.' In: *Amino Acids* 53.12 (2021), pp. 1807–1815. DOI: 10.1007/s00726-021-02961-5. URL: https://www.ncbi.nlm.nih.gov/pubmed/33646427 (cit. on p. 89).

[164] M. Tajan and K. H. Vousden. 'Dietary Approaches to Cancer Therapy.' In: *Cancer Cell* 37.6 (2020), pp. 767–785. DOI: 10.1016/j.ccell.2020.04.005. URL: https://www.ncbi.nlm.nih.gov/pubmed/32413275 (cit. on p. 92).

[165] A. Paul, R. Anand, S. P. Karmakar, S. Rawat, N. Bairagi, and S. Chatterjee. 'Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models.' In: *Sci Rep* 11.1 (2021), p. 213. DOI: 10.1038/s41598-020-80561-1. URL: https://www.ncbi.nlm.nih.gov/pubmed/33420254 (cit. on p. 92).

[166] E. J. O'Brien, J. M. Monk, and B. O. Palsson. 'Using Genome-scale Models to Predict Biological Capabilities.' In: *Cell* 161.5 (2015), pp. 971–987. DOI: 10.1016/j.cell.2015.05.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/26000478 (cit. on p. 92).

[167] A. S. Blazier and J. A. Papin. 'Integration of expression data in genome-scale metabolic network reconstructions.' In: *Front Physiol* 3 (2012), p. 299. DOI: 10.3389/fphys.2012.00299. URL: https://www.ncbi.nlm.nih.gov/pubmed/22934050 (cit. on p. 92).

[168] C. Reglero and G. Reglero. 'Precision Nutrition and Cancer Relapse Prevention: A Systematic Literature Review.' In: *Nutrients* 11.11 (2019). DOI: 10.3390/nu11112799. URL: https://www.ncbi.nlm.nih.gov/pubmed/31744117 (cit. on p. 93).

[169] E. M. Maldonado, C. P. Fisher, D. J. Mazzatti, A. L. Barber, M. J. Tindall, N. J. Plant, A. M. Kierzek, and J. B. Moore. 'Multi-scale, whole-system models of liver metabolic adaptation to fat and sugar in non-alcoholic fatty liver disease.' In: *NPJ Syst Biol Appl* 4 (2018), p. 33. DOI: 10.1038/s41540-018-0070-3. URL: https://www.ncbi.nlm.nih.gov/pubmed/30131870 (cit. on p. 93).

[170] E. Bugianesi et al. 'Insulin resistance in non-diabetic patients with non-alcoholic fatty liver disease: sites and mechanisms.' In: *Diabetologia* 48.4 (2005), pp. 634–42. DOI: 10.1007/s00125-005-1682-x. URL: https://www.ncbi.nlm.nih.gov/pubmed/15747110 (cit. on p. 94).

[171] E. S. Jin, J. D. Browning, R. E. Murphy, and C. R. Malloy. 'Fatty liver disrupts glycerol metabolism in gluconeogenic and lipogenic pathways in humans.' In: *J Lipid Res* 59.9 (2018), pp. 1685–1694. DOI: 10.1194/jlr.