Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Rebecca Madeleine Gilbert Sandtrøen

Comparative analysis of cancers and identification of novel anticancer drug targets using genome-scale metabolic modeling

Master's thesis in Industrial Chemistry and Biotechnology
Supervisor: Eivind Almaas
Co-supervisor: Vetle Simensen
August 2022

NTNU
Norwegian University of Science and Technology

Rebecca Madeleine Gilbert Sandtrøen

Comparative analysis of cancers and identification of novel anticancer drug targets using genome-scale metabolic modeling

Master's thesis in Industrial Chemistry and Biotechnology
Supervisor: Eivind Almaas
Co-supervisor: Vetle Simensen
August 2022

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Summary

Many currently used anticancer drugs aim to disrupt cellular processes in tumor tissue, inevitably affecting the same processes in healthy tissue. This results in adverse and long term side effects for the patient, as well as additional suffering and early death. Computational tools like genome-scale metabolic models (GEMs) can provide a new insights into what sets cancer cells apart from healthy cells, and how this might be exploited for developing more targeted therapies.

In this thesis GEMs were used to explore the alterations in the metabolic capabilities of five commonly occurring cancers at early and late stages of cancer progression, as well as their healthy tissue counterpart. The cancer types investigated were breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, lung adenocarcinoma, pancreatic adenocarcinoma, and prostate adenocarcinoma. The aim was to explore genetic and metabolic differences between normal cells and cancer cells, and between different cancer stages and -types.

Various gene deletion analyses were carried out *in silico*, including single and double gene deletions, with essential genes used as surrogates for potential drug targets. Several genes were found to be uniquely essential to tumors, and not to the normal tissue counterpart, which are proposed as potential target-genes for anticancer treatment. The majority of the target-genes identified in this thesis have been proposed as potential cancer drug targets for various cancers elsewhere, and therefore holds promise for future research. In addition, this thesis presents novel targets, some of which may be utilized as targets against multiple cancer types, increasing the scope of application of potential new drugs. The double gene deletion analysis performed on breast invasive carcinoma GEMs revealed a handful of genes that formed a synthetic lethal gene pair with numerous other genes, a phenomenon not observed in the normal tissue counterpart with the same genes. These genes points to metabolic alterations in the cancerous tissue that might be exploited as potential drug targets.

The findings in this thesis may provide a sense of direction to guide future cancer research, notably where it may be beneficial to allocate experimental resources, and to emphasise the potential use of GEMs as a valuable resource in this pursuit.

Sammendrag

Mange av dagens kreftmedisiner har som mål å hemme cellulære prosesser i kreftvev, noe som uungåelig også påvirker de samme prosessene i friskt vev. Dette resulterer i uønskede og langsiktige bivirkninger for pasienten, samt ytterligere lidelse og tidlig død. Dataverktøy som genom-skala metabolske modeller (GEMs) kan gi ny innsikt i hva som skiller kreftceller fra friske celler, og hvordan dette kan utnyttes for å utvikle mer målrettede behandlingsformer.

I denne oppgaven ble GEMs brukt til å utforske endringene i de metabolske egenskapene til fem vanlig forekommende kreftformer i både tidlig og sen stadie av kreftprogresjon, samt deres tilsvarende friske vev. Krefttypene som ble undersøkt var brystkreft, livmorhalskreft, lungekreft, bukspyttkjertelkreft, og prostatakreft. Målet var å utforske genetiske og metabolske forskjeller mellom normale celler og kreftceller, og mellom ulike kreftstadier og -typer.

Ulike gendelesjonsanalyser ble utført *in silico*, inkludert enkle og doble gendelesjoner, hvor essensielle gener ble brukt som surrogater for potensielle medikamentmål. Flere gener ble funnet å være essensielle kun for krefttyper, og ikke for dets tilsvarende friske vev, og disse blir derfor foreslått som potensielle mål-gener for nye kreftmedisiner. Flertallet av målgenene som ble identifisert i denne oppgaven har blitt pekt på som potensielle kreftmedisinmål for ulike kreftformer i andre studier, og framstår derfor lovende for videre forskning. I tillegg presenterer denne oppgaven nye målgener, hvorav noen kan brukes som mål mot flere krefttyper, noe som øker anvendelsesområdet for potensielle nye legemidler. Den doble gendelesjonssanalysen utført på GEMs for brystkreft avslørte en håndfull gener som dannet et syntetisk dødelig genpar med en rekke andre gener, et fenomen som ikke ble observert i det tilsvarende normale vevet med de samme genene. Disse genene peker på metabolske endringer i kreftvevet som kan utnyttes som potensielle medikamentmål.

Funnene i denne oppgaven kan bidra med å peke ut en videre retning for fremtidig kreftforskning, spesielt hvor det kan være gunstig å fokusere eksperimentelle ressurser, og med å belyse bruken av GEMs som en verdifull ressurs i dette arbeidet.

Preface

The work presented in this thesis was carried out at the Department of Biotechnology and Food Science at the Norwegian University of Science and Technology (NTNU), supervised by professor Eivind Almaas. It concludes my Master's Degree in Chemical Engineering and Biotechnology, where I have specialized in biotechnology.

I would like to thank my supervisor, professor Eivind Almaas, for his encouragement, guidance and help throughout this project. It has been an honor to have had the opportunity to be a part of this inspiring research group. I also want to express my gratitude to my co-supervisor, Vetle Simensen, for his assistance and patience with the technical parts of the project which provided me with significant challenges - his help and guidance has been invaluable for the development of many of the methods used in this project.

Lastly, I would like to thank my friends and family for supporting me in following my interests, both academically and personally, and to NTNU and NTNUI Dans for providing me with opportunities to do so. And of course a big thanks to my fiancé, Vegard, for your interesting discussions and insights, your continued support, encouragement and patience.

Rebecca Madeleine Gilbert Sandtrøen
Trondheim, August 2022

Contents

Summary i

Preface i

List of Figures iv

List of Tables vii

Abbreviations ix

1 Introduction 1

1.1 Complex systems and systems biology . . . 1

1.2 Human health as a systems problem . . . 1

1.3 Aim of this thesis . . . 3

2 Background and theory 4

2.1 Mathematical foundations for reconstructed networks . . . 4

2.1.1 From biology to mathematics . . . 4

2.1.2 Flux Balance Analysis . . . 6

2.2 Genome-scale models . . . 8

2.2.1 Reconstruction is a 4 step process . . . 9

2.2.2 Applications of a GEM . . . 10

2.2.3 Single Gene Knockouts . . . 11

2.2.4 Synthetic Lethal Gene Pairs . . . 12

2.3 Human GEMs 12
2.3.1 The emergence of human metabolic modeling 12
2.3.2 Creating tissue specific models 13
2.3.3 Choosing an objective function in human GEMs 14
2.4 Cancer 14
2.4.1 The hallmarks of cancerous disease 14
2.4.2 Cancer progression 15
2.4.3 GEMs as a tool in cancer research 16

3 Software and methods 17
3.1 Software 17
3.1.1 MATLAB 17
3.1.2 COBRA and RAVEN Toolboxes 17
3.1.3 Gurobi 18
3.1.4 Microsoft Office 18
3.1.5 Cytoscape 18
3.2 Flux Balance Analysis 18
3.3 Model retrieval and initial verification 18
3.4 Essential Gene Analyses 20
3.4.1 Essential gene analysis based on biomass production 21
3.4.2 Essential gene analysis based on metabolic tasks 21
3.5 Synthetic Lethal Gene Pair Analysis 23

4 Results and Analysis 24
4.1 Model verification 24
4.2 Essential Gene Analyses 25
4.2.1 Biomass Production essential genes 25
4.2.2 Metabolic Task essential genes 27
4.2.3 Comparing the two methods of SGD 34
4.3 Synthetic Lethal Gene Pairs 36

5 Discussion 42
5.1 Potential drug targets were discovered 42
5.2 MT-target-genes finds support in literature 44
5.3 Challenges 45
5.4 Further work 46

6 Conclusion and Outlook 48

Bibliography 50

Appendices 59

A Description of the Supplementary data files 59

B MT-target-genes supported in literature 60

C Glycolysis and gluconeogenesis 63

List of Figures

2.1 The concept of constraint-based modeling. Without any constraints, the flux distribution of a biological network may lie at any point in the solution space. Introducing constraints by the stoichiometric matrix **S** and lower and upper reaction bounds (a_i and b_i) defines an allowable solution space. Any point outside of the solution space are denied by the constraints, and thus the network may only acquire flux distributions within the solution space. FBA can identify a single optimal flux distribution within the allowable solution space that optimizes the selected objective. Figure from [17]. 7

2.2 The 96 steps required in the procedure for generating a high-quality genome-scale metabolic network reconstruction. The process is generally performed iteratively, particularly in phases 2-4, until the model demonstrates the ability to predict phenotypic features that are close to experimental data for the target organism. Figure from [27]. 10

4.1 Excerpt of the model verification results. "Pass" means the model was able to perform the metabolic task being tested. 25

4.2 Network representation of the BP-target-genes (circles) identified for the various tumor types (squares), collectively called *nodes*. Increasing size of the nodes signifies increased number of connections (*edges*). An edge between two nodes signifies that the gene is a BP-target for that tissue type. The colour gradient of the circles also represent the number of edges, with yellow signifying just one edges, green meaning a few edges, and blue representing several edges. 27

4.3 MT-essential genes overlaps between the three tissue types within the model subgroups; (a) BRCA, (b) CESC, (c) PAAD, (d) PRAD, and (e) LUAD. 29

4.4 Network representation of the MT-target-genes (circles) identified for the various tumor types (squares), collectively called *nodes*. Increased thickness of the lines (edges) represent increased weight of the MT-target-gene. Increasing size of the nodes signifies increased number of edges. The colour gradient of the circles also represent the number of edges, with yellow signifying just one edge, green meaning a few edges, and blue representing several edges. 30

4.5 The frequency of the MT-target-genes that appears in more than one tumor model. A total of 23 MT-target-genes were identified as such. 31

4.6 The weight distribution of MT-target-genes. Weight represents the number of metabolic tasks the SGD mutant model failed to perform, when testing the 57 essential metabolic tasks all human cells must be able to perform to be considered viable. The frequency depicts the total number of observations of MT-target-genes with this weight for every tumor model. 32

4.7 Number of essential genes identified with two different methods of gene knockout analysis, BP-essential (blue), and MT-essential genes (orange). 36

4.8 The number of SL gene pairs identified for each model in the double gene deletion analysis. SL-targets (orange) are SL-gene pairs that are only lethal for that tumor model, but not lethal for the normal tissue. Blue columns represent the number of SL gene pair shared with BRCA NT. 37

4.9 The number of synthetic lethal (SL) gene pairs with the various weights for BRCA TM (blue) and BRCA TP (orange). 38

4.10 Network representation of the SL-targets in BRCA TP. Increased thickness of the lines (edges) represent increased weight of the genes (circles) connected as an SL-target. Increasing size of the nodes signifies increased number of connections. The colour gradient (yellow-blue-purple) of the circles also represent number of connections, with yellow signifying the lowest number of edges and purple representing most edges. 40

4.11 Network representation of the SL-targets in BRCA TM. Increased thickness of the lines (edges) represent increased weight of the genes (circles) connected as an SL-target. Increasing size of the nodes signifies increased number of connections. The colour gradient (yellow-blue-purple) of the circles also represent number of connections, with yellow signifying the lowest number of edges and purple representing most edges. 41

C.1 Network of the human central carbohydrate metabolism. The pathways of glycolysis and gluconeogenesis are highlighted in pink. The highest-scoring SL-target, which is solely fatal for BRCA TP and BRCA TM, but not BRCA NT, is highlighted in red (TPI1) and orange (GAPDH). Figure from [110]. 64

List of Tables

3.1 List of models used in this thesis. Sample type codes: NT = solid tissue normal, TP = primary solid tumor, TM = metastatic tumor, TR = recurring solid tumor. 19

4.1 SGD analysis for the tissue models, using the ability of biomass production (BP) as the metric for cellular viability. The number of genes identified in the models as essential for biomass production (BP-essential genes) are listed. BP-target-genes are essential genes that are only essential for a tumor model, but not essential for the corresponding normal tissue model in regards to BP. 26

4.2 SGD analysis for the tissue models, with the ability to perform the 57 Metabolic Tasks (MT) essential to human cells as the indicator for cellular viability. The number of genes identified in the models as essential for completing all metabolic tasks (MT-essential genes) are listed. MT-target-genes are genes that are only essential for a tumor model, but not essential for the corresponding normal tissue model. . 28

4.3 The identified MT-target-genes with weight more than 1, and the corresponding tumor models in which they appeared as MT-target-genes. 33

4.4 The number of essential genes identified in all models by the two different methods of SGD analysis; genes essential for biomass production (BP), and genes essential for completing all 57 metabolic tasks (MT) necessary for human cell viability. Biomass production is one of the 57 MTs, thus the biomass proportion of the MT-essential genes was calculated. 35

4.5 The most highly connected genes in the SL analysis. *Italicized* genes are found as an SL-target-forming gene only in BRCA TP, all other genes are found in both BRCA tumor types as SL-target-forming genes. 38

B.1 Identified MT-target genes proposed as a drug targets for various cancer types in literature. 60

Abbreviations

GEMs	Genome-scale metabolic models
TCGA	The Cancer Genome Atlas
FBA	Flux balance analysis
COBRA	COnstraint-Based Reconstruction and Analysis
GPR	Gene-to-protein-reaction
KO	Knockout
SL	Synthetic lethal
NAFLD	Non-alcoholic fatty liver disease
tINIT	Task-driven Integrative Network Inference for Tissues
CSC	Cancer stem cells
DDR	DNA damage response
RAVEN	Reconstruction, Analysis and Visualization of Metabolic Networks
NT	Normal solid tissue
TP	Primary solid tumor
TR	Recurring solid tumor
TM	Metastatic tumor
TPM	Transcripts per million
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
LUAD	Lung adenocarcinoma
PAAD	Pancreatic adenocarcinoma
PRAD	Prostate adenocarcinoma
SGD	Single gene deletion
BP	Biomass production
MT	Metabolic tasks

Chapter 1

Introduction

1.1 Complex systems and systems biology

Intricate systems are found everywhere around us, take for example computers communicating on the Internet, airports connected by flights, and humans connected by social interactions. Such systems consisting of many components interacting in complex patterns can collectively be called *complex systems* [1]. The behaviour of complex systems is often difficult to anticipate by looking at the components alone, because interactions between them, as well as with their surroundings, can function in a multitude of ways. Furthermore, system components can constitute complete new systems, leading to systems of systems that are interdependent on one another. Prime examples of this can be found in biology; proteins work together to form living cells, numerous cells interacts to form an organ, which again interacts in a system of several organs to form a whole functioning organism, which is part of an ecosystem interacting with many other organisms and species.

To fully understand complex biological systems, it is rarely sufficient to only study individual biomolecular components and their functions, one also need information on how they interact with one another and with their environment. This realization sparked the emergence of the field of *systems biology*, which can be described as the study of the interactions between the components of a biological system, and how these interactions give rise to the functions and behaviours of the system. In other words, systems biology aims to understand the biological system as a whole, rather than piece by piece.

1.2 Human health as a systems problem

Disruptions in complex systems might create cascading effects throughout the system, resulting in changes in how the system operates and behaves. The extent of such effects, which can range from no noticeable consequences to a complete system crash, depends greatly on the robustness of the system, and the type and location

of the disruption(s) [2]. The vastly complex cellular systems of our bodies are no exception from this. Generally, most cellular disruptions are harmless and have few negative consequences, however, some can create major impacts which affects our health and life. In cases where disruptions causes diseases, medical sciences aim to find solutions to counter the harmful effects. In order to do so, it is essential to have an understanding of how the system works and to identify the mechanisms that causes the disease.

A reasonable starting point in this venture would be to study the genes and proteins of the diseased state. However, complex diseases like cancer normally has a number of genetic mutations, epigenetic changes, and variations on protein expression [3]. Although studying and cataloguing such mutations can yield a wealth of new knowledge and understanding, it is typically unclear how these changes affect and ultimately determine the phenotypic changes observed [4].

In 1986, Renato Dulbecco wrote that "One of the goals of cancer research is to ascertain the mechanisms of cancer", whilst arguing that a mechanistic understanding of cancer requires the complete sequenced human genome [5]. Only by getting a systemic overview of the genes and their function can it be possible to understand the phenotypic states of cells. Today, with an abundance of sequence data, and technologies that can perform various high-throughput experiments, we have access to a huge number of molecular measurements within tissues and cells.

This new surge of information has made it possible to create genome-scale metabolic networks for various target organisms, including humans [6, p. 2]. Such metabolic networks, consisting of all known biochemical reactions and the related genes and enzymes in an organism, can be studied as they are, or they can be converted into a mathematical format for *in silico* predictive simulations - so called genome-scale metabolic models (GEMs). The goal of a GEM is to put biochemical information in context, integrate all knowledge about the organism in a systemic way, and subsequently, in theory, be able to simulate every metabolic phenotype. For instance, gene knockout simulations can be performed to study the metabolic effects on various cellular functions, including cellular growth capabilities [7].

It has been known for some time that cancers have different metabolic patterns than healthy cells. The changes in metabolism observed as part of tumorigenesis makes metabolic modeling a suitable approach to studying cancer [8]. Additionally, recent studies utilizing human GEMs and high-throughput data has shown that also cancers differ considerably among themselves in their metabolic signature [9] [10]. For example, a lung cancer cell can be very different than a colon cancer cell, and thus the different cancers would probably respond differently to the same drugs [11]. As such, the popularized idea of finding "*the* cure for cancer" seems futile, as every cancer type needs to be viewed as separate diseases. It is therefore a growing interest in further studying the specific cancers' metabolic and genetic signature in order to better understand what sets them apart from healthy cells [11].

The anti-cancer drugs used today often have many unwanted and harmful side effects [12]. Thus, there is an apparent need for novel medications with improved therapeutic windows, meaning means the medication will target a certain cell type, such as tumour cells, while having no or little side effects on healthy cells [13]. Essential

genes can be used as a surrogate for drug targets [14], and thus any changes in gene essentiality between healthy and sick tissue has the potential to be exploited. Single and double gene knockout simulations has been able to identify growth related genes in cancer cell line GEMs, pointing to possible novel drug target which reduce the growth rate of cancer cells but not of the normal cells, of which two drugs targets were experimentally verified [15].

1.3 Aim of this thesis

The aim of this study was to investigate genetic differences between healthy tissues and cancer tissues for a selection of commonly occurring cancers. The following fundamental questions was asked: Are there genes that can be inhibited that would severely disrupt cancer cells, while not harming healthy cells? It was also investigated whether there were variances between different stages of the same cancer type, as well as whether separate cancer types shared any such metabolic vulnerabilities.

For this aim, GEMs was used to simulate different gene knockout scenarios and investigate the effects on metabolic functions. The focus in this approach was to identify genes that, alone or in pair, were essential for maintaining metabolic functions necessary for cell viability. The models used were tissue-specific GEMs created with gene expression data obtained from patient samples from the The Cancer Genome Atlas (TCGA) database. This included models of normal tissue (from where the cancer developed), primary solid tumor, and either metastatic tumor or recurring tumor for each of the cancer types that were investigated.

The goal was to discover potential drug targets within the cancer models that would, in theory, kill the cancer cell while sparing healthy cells, and thus guiding research of novel cancer medicines, as well as contributing to a deeper understanding of cancer as a complex disease.

Chapter 2

Background and theory

This chapter introduces the relevant theory and background information used in this thesis. First, the mathematical foundations for reconstructed networks will be explored. This includes the basis for how biological information can be expressed in a mathematical format, and how this can be computationally analysed in many ways. Then, the process behind creating GEMs are described, as well as the many areas of application they can have. Last is a basic overview of how cancer may be viewed as a metabolic dysfunction and why GEMs are a suitable tool for studying this disease.

2.1 Mathematical foundations for reconstructed networks

2.1.1 From biology to mathematics

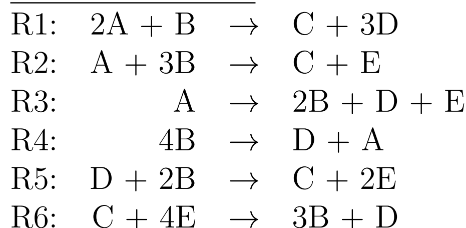
In order to create a GEM of an organism, one must start by obtaining a genome-scale view of its contents. The starting point in creating a metabolic network for a target organism is to collect information about all the chemical compounds that are found in a target cell, and how they interact. This results in a comprehensive list of chemical reactions, examples being reactions detailing how the organism creates and utilizes energy molecules, how it metabolises proteins, and how transport of molecules between cellular compartments is organized. All of these biochemical reactions can be connected to create a reaction network, also called the *metabolic (reaction) network*. The stoichiometry of the reactions details how many molecules that are consumed and/or produced in each reaction, and this information can be represented mathematically by the *stoichiometric matrix*, denoted by **S**. As such, **S** represent the total metabolic network reconstruction, detailing all known biochemical activity happening in the target cell.

$\mathbf{S} = (S_{ij})$ is of dimensions $m \times n$, where m represent the number of unique metabolites and n represent the number of unique reactions in the metabolic network. The

element S_{ij} is the stoichiometric coefficient of metabolite i produced in the reaction j. The coefficient is negative when metabolite i is consumed in reaction j, and positive when it is produced [16]. Metabolites that does not partake in reaction j has the stoichiometric coefficient of 0, and since most biochemical reactions involve only a few metabolites, **S** is a sparse matrix [17].

A hypothetical reaction network consisting of six reactions (R1-R6) involving five metabolites (A-E) is presented below, along with its corresponding stoichiometric matrix (adapted from [16]).

Reaction Network



R1:	2A + B	→	C + 3D
R2:	A + 3B	→	C + E
R3:	A	→	2B + D + E
R4:	4B	→	D + A
R5:	D + 2B	→	C + 2E
R6:	C + 4E	→	3B + D

Stoichiometric Matrix

$$\begin{array}{cccccc} R1 & R2 & R3 & R4 & R5 & R6 \end{array}$$
$$\begin{bmatrix} -2 & -1 & -1 & 1 & 0 & 0 \\ -1 & -3 & 2 & -4 & -2 & 3 \\ 1 & 1 & 0 & 0 & 1 & -1 \\ 3 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & 1 & 0 & 2 & -4 \end{bmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

The flux through the reactions are represented by the vector $\mathbf{v} = (v_1, v_2, ..., v_n)$ with length n, where v_j is the specific flux for reaction j. Likewise, the concentrations of all metabolites are represented by the vector $\mathbf{x} = (x_1, x_2, ..., x_m)$ of length m, where x_i is the specific concentration for metabolite i. The relationship between the vectors **x** and **v**, the matrix **S**, and time t, can be described as Equation (2.1).

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} \tag{2.1}$$

At a metabolic steady state, there is, per definition, no accumulation of metabolites in the system ($\mathrm{d}\mathbf{x}/\mathrm{d}t = 0$). Thus, the system is assumed to achieve a dynamic mass balance, which in turn satisfies mass balance laws. Equation (2.2) depicts this relation, and serves as the foundation for the rest of the theoretical framework depicted in this thesis.

$$\mathbf{S}\mathbf{v} = 0 \tag{2.2}$$

Any **v** that satisfies Equation (2.2) is said to be a feasible flux distribution that the network is capable of displaying within the constraints, and consequently such **v** characterizes all possible functional states of a reconstructed biochemical network. Explicitly, the flux distribution of a system are the rates at which every metabolite is consumed or produced by each reaction [18].

Equation (2.2) defines a system of linear equations, and this serves as a starting point for various mathematical analyses and computational approaches used to explore network properties [6, p. 169]. Any solution to the set of equations that satisfies all constraints is called a *feasible solution*, whereas infeasible solutions are solution in which one or more constraint is violated. The nature of such constraints will be discussed in more detail in Section 2.1.2.

2.1.2 Flux Balance Analysis

This section is adapted from previous work of the author [18].

A widely used computational approach to analyse the properties of a reconstructed biochemical network is flux balance analysis (FBA). It has shown to be especially useful for the genome-scale metabolic network reconstructions that have been constructed in the last decades [17]. In a review article from 2011, FBA/Linear programming was found to be the most frequently utilized computational analysis method in published studies regarding GEMs of *Escherichia coli* [19]. FBA can be used to assess the flux of metabolites in a metabolic network, which in turn represent the organism's functional usage of the network. This creates the opportunity for researchers to simulate growth rates or the rate at which a biotechnologically relevant molecule is produced in the target organism [17].

A key assumption made in the FBA framework is that the fluxes has entered a steady state, meaning there is no accumulation of metabolites. This assumption stems from the mass balance law, where all mass needs to be accounted for. Since accumulation of metabolites is set to 0, the input to the system (usually complex substrates) needs to be balanced against the output of the system (often set as the biomass produced). The next assumption is that the organism itself has identified and utilized the optimal solution. This assumption is supported by the hypothesis that cells are forced by evolutionary pressure to evolve optimal behavior in the given environmental conditions, be that optimal growth or conservation of resources [20]. It is further supported by studies showing that *in silico* predictions of *E. coli* metabolic capabilities have been consistent with experimental data [21] [22].

The number of reactions (n) in a biological network is typically greater than the number of compounds (m), creating a mathematical problem with more unknown variables than equations. This generates a plurality of potential flux distribution solutions, often refered to as the *solution space*. To limit the solution space, different constraints can be imposed on the system. These can include constraints imposed from thermodynamics (e.g. effective reversibility or irreversibility of reactions), enzyme or transporter capacities (e.g. maximum uptake or reaction rates), or other organism- and environmentally-specific limitations. The constraints are denoted by a_i and b_i, which represents lower and upper bounds, respectively, on the flux of the reactions [17]. Then, depending on what is being investigated, a phenotype in the form of a biological objective must be defined. This is called the objective function, termed $\mathbf{Z}$, and examples include maximizing the biomass production (as a proxy for growth rate) or the ATP production of the cell. When $\mathbf{Z}$ is defined, FBA can be used to solve the set of linear equations provided by Eq. (2.2) with the goal of

maximizing or minimizing the flux through the chosen objective reaction **Z**. This concept of so called *constraint-based modeling* can be illustrated as creating a "flux cone" representing all flux distributions that are possible given the constraints, and then using FBA to identify the optimal solution for the selected objective function (Figure 2.1). Thus, FBA computations falls into the category of COnstraint-Based Reconstruction and Analysis (COBRA) methods.

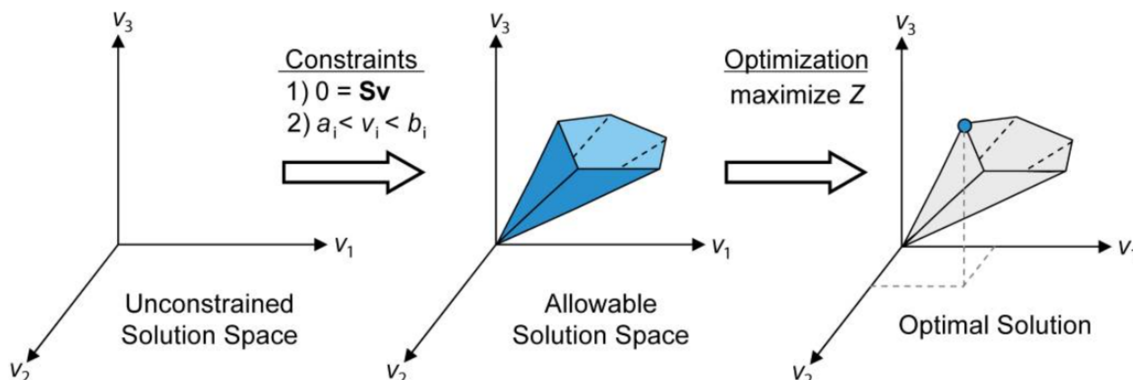


Figure 2.1: The concept of constraint-based modeling. Without any constraints, the flux distribution of a biological network may lie at any point in the solution space. Introducing constraints by the stoichiometric matrix **S** and lower and upper reaction bounds (a_i and b_i) defines an allowable solution space. Any point outside of the solution space are denied by the constraints, and thus the network may only acquire flux distributions within the solution space. FBA can identify a single optimal flux distribution within the allowable solution space that optimizes the selected objective. Figure from [17].

The output of an FBA is the particular flux distribution needed for the cell to achieve the chosen objective function. FBA can thus be used to investigate the metabolic capabilities of the cell [6, p. 145].

The incorporation of different constraints on the reaction fluxes, together with the assumption of metabolic steady state, is mainly what distinguishing the FBA framework from other theory-based models that require a large number of difficult-to-measure kinetic parameters [17]. A disadvantage of not using kinetic parameters is that the FBA model cannot predict metabolite concentrations. On the other hand, a great advantage of bypassing kinetic parameters is that FBA can be computed very quickly, even for large networks, making it possible to study a large number of different situations such as different substrates or genetic manipulations. Another limitation in the FBA framework, except in some modified forms, is that it does not account for regulatory effects happening in the cells. Regulation can affect the metabolic phenotype of the cell by means of activation or inactivation of enzymes, regulation of gene expression and more [23]. As a result, FBA predictions may not always be accurate, and thus experimental verification is always needed. On the other hand, this shortcoming has the potential to be used in an analytical sense. Standard FBA provides the widest possible solution space given the biochemical constraints, and adding additional regulatory constraints would only serve to further reduce this solution space. Therefore, if an FBA study presents the biochemically optimal flux distribution, and experimental data shows that the target cell does not adhere to

this through regulatory mechanisms, the question can be asked as to what the cell stands to gain by not being "optimal". This can shed light on what the objective function(s) of the cell might actually be and guide other biology discoveries.

Nevertheless, there are numerous applications for this theoretical framework, ranging from physiological studies, gap-filling efforts, to genome-scale synthetic biology [17]. By altering the constraints on a model, simulations of different conditions for the target organism can be performed. This allows researchers to investigate what the metabolic network can and cannot do, and hence guide biological discovery and bioprocess engineering. Environmental conditions, such as substrate availability, can be altered by changing the bounds on exchange reactions that control the transport of metabolites in and out of the system. For example, anaerobic conditions can be simulated by restricting oxygen uptake rate to 0, and similarly, growth in oxygen rich environment is simulated by setting the uptake rate arbitrarily high (often set to 1000) so that it is not a limiting factor in any way. Substrates that are not found in the media are constrained to an uptake rate of 0, and those that are present are represented in the model by allowing uptake rates within a feasible range [17]. This can be highly useful when biotechnologists want to designing microorganisms that displays overproduction of a biochemical of interest, such as biofuels, industrial chemicals, and pharmaceutical precursors. For example, FBA has been used to identify modifications to the metabolic reaction network used in fermentation processes in order to force the overproduction of succinate in *E. coli* [24]. FBA based studies can therefore be used to design optimal growth media with respect to the desired metabolic function.

As for synthetic and *in silico* biology, FBA is a useful tool to study the effect of genetic perturbations. Gene deletion studies can be performed by restricting the flux of the particular gene product to 0 and assessing the following effect on the metabolic network. This provides information on the importance of genes for growth and survival of the organism and has been used identify the lethality of single and multiple gene knock-outs [25], as well as predicting synthetic genetic interactions [26]. Sections 2.2.3 and 2.2.4 will provide more details on these topics. Consequently, metabolic modelling and analysis can be a valuable tool to generate biological hypotheses and promising candidates before initiating expensive and time-consuming experiments in the lab.

2.2 Genome-scale models

In theory, a GEM is a unified collection of all known knowledge on an organism compiled into a single model. The ultimate objective is to account for all biochemical reactions that occur in a cell and how they link to the genome, and to predict every phenotypic trait that the organism is capable of exhibiting. Evidently, this is no easy task. The basics of how such a model is created is outlined in this section.

2.2.1 Reconstruction is a 4 step process

Many different GEM building pipelines and softwares exists today to guide the construction of a GEM [27][28][29][30], with the primary difference being the starting point, which varies depending on how well studied the target organism is. A renowned and exceptionally thorough protocol that is adaptable to all GEMs, even when no prior work has been done to create a systems level view of the metabolic network of the target organism, was created by Thiele & Palsson [27]. This protocol describes the bottom-up reconstruction process needed to create a high-quality GEM, and has been divided into four main parts, as illustrated in Figure 2.2. A condensed summary of this protocol is described below, for more details the reader is referred to the original publication [27].

1. **Creating a draft reconstruction**. The first step consists of using various biochemical databases to identify all the genes present in the organism, and connecting those annotations to enzymes and metabolic reactions. This provides the basis for the gene-to-protein-reaction (GPR) assosiation rules that details the link between genes and metabolic reactions. When this comprehensive list is gathered, the reactions are connected to form a draft network reconstruction. Today, this step is largely automated [31][32].

2. **Manual refinement**. The second step consists of manually curating and re-evaluating the draft reconstruction, as it often will be incomplete, have gaps in the network, and contain content that is not organism-specific. This is an extensive and time-consuming process where primary bibliomic data on the target organism is examined to assess all the genes, enzymes, and metabolic reactions individually to ensure adequate evidence for their presence in the target organism. The result of this step is a high-quality, organism-specific reconstruction of the metabolic network of the target organism. Such a comprehensive network reconstruction is also referred to as a biochemically, genetically and genomically structured (BiGG) knowledge base, and is often made accessible to all [19].

3. **Conversion of the reconstruction into a mathematical format**. The BiGG database is then converted into a mathematical format, namely the stoichiometric matrix $\mathbf{S}$, which enables a wide range of computational analyses of the features of the metabolic network. At this stage it is common to test the ability of the network to produce biomass components, and comparing them to experimental data or primary literature. Other gap-filling analyses may be performed to add missed pathways, and remove any that have been incorrectly included.

4. **Network evaluation and validation**. The last step in this process is to evaluate, verify and validate the network. Some examples of the processes that are commonly validate at this stage are energy maintenance reactions required for both growth- and non-growth associated activities, ATP production, and any inconsistencies between the model and experimental analysis results. This is largely an iterative process that require numerous corrections

and adjustments, spanning all four steps, before the model is considered satisfactory. The resulting GEM is a computational model representing all (or most) known metabolic information about the target organism.

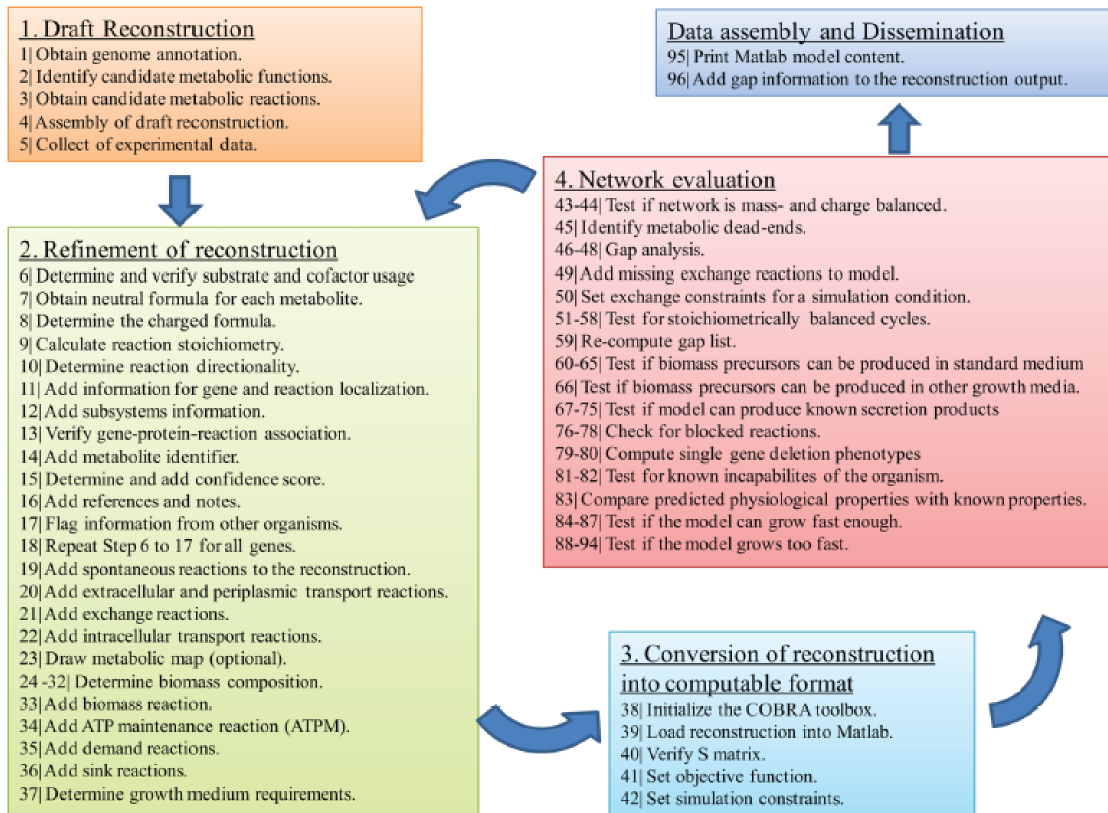


Figure 2.2: The 96 steps required in the procedure for generating a high-quality genome-scale metabolic network reconstruction. The process is generally performed iteratively, particularly in phases 2-4, until the model demonstrates the ability to predict phenotypic features that are close to experimental data for the target organism. Figure from [27].

2.2.2 Applications of a GEM

A well curated GEM can have a broad range of possible applications and serve as a valuable new[1] tool in scientific research. This includes studying the underlying metabolic mechanisms behind observable phenotypes, thus gaining biological insights on a new level. Also, since our knowledge of an organism is seldom complete, gaps in the network may appear in the reconstruction process. This serves to emphasise where our knowledge base is incomplete and, as a result, can help guide the discovery process [6, p. 45-48]. GEMs can also be used in synthetic biology design, to design media compositions or genetic mutants to optimize the performance of the metabolic network of a commercially important organism [6, p. 45-48].

[1]"New" in the sense that the first ever GEM was published in 1999 (*Haemophilus influenzae Rd*) [33], and the first GEM of human cells, Recon 1, was published in 2007 [34].

GEMs can also serve as a way to contextualise the huge amounts of data provided by various high-throughput experiments. It allows the shift from focusing research on individual genes or molecules to a broader, systems view of the network properties and functionalities. For instance, since genetic mutations take place within the context of intricate biological networks, the underlying mechanisms causing phenotypic changes are often unclear [35]. GEMs can therefore shed light on these underlying mechanisms by simulating the organisms functional use of the metabolic network in various genetic and environmental situations. Two specific types of genetic studies that can be performed using GEMs will be discussed in more detail; single gene knockouts and synthetic lethal gene pair analyses.

2.2.3 Single Gene Knockouts

The relationships between genes, proteins, and reactions, are usually listed in GEMs using logical expressions known as GPR rules. These detail the connections between what gene products catalyzes what reactions, and serve as a way to trace the network functions [6, p. 378-397]. To better understand the role of a certain gene, a method known as gene knockout (KO) can be used. In a GEM, a gene is knocked out, or *deleted*, by constraining its related metabolic reaction(s) to a flux of 0. Mathematically, the solution space is effectively shrunk by the loss of reactions that can carry flux, and thus the optimal usage of the network may change.

By optimizing for a chosen phenotypic function one can assess the capabilities of the reduced network and compare it to the full network, allowing one to study the function of the gene in question. Generally, the effects of a gene KO on the network can be classified into three categories. First, if the gene removal results in inability to satisfy the phenotype of interest, the gene is considered indispensable, or *essential*. For example, if a gene is deleted and the resulting model is incapable of producing biomass, then the gene is essential for biomass production. Secondly, the gene removal can lead to a reduction, but not completely removal, of the phenotypic capability being studied. Common ways to quantify the changes in the phenotype include calculating the relative change, for example of the growth rate of the KO strain relative to the wild-type growth rate. Lastly, the gene removal can have no consequence on the phenotypic expression, and thus the gene is considered redundant for that phenotype. [6, p. 378-397]

It should be noted that such computations are context-specific, depending on factors such as media composition, specific uptake and excretion bounds, and more. In addition, an essential gene is only essential for the objective function in question. For instance, a gene can be essential for arginine production, but *not* essential for biomass production. As such, when considering gene essentiality, the selection and formulation of the objective function is imperative.

2.2.4 Synthetic Lethal Gene Pairs

In the same way that single gene KOs can be studied using GEMs, two genes (or more) can be deleted simultaneously to create double (or triple, etc.) KO mutant strains. This can be used to study gene interactions, also called *epistatic* interactions. An epistatic interaction is when the deletion of the genes individually has no or little effect on the phenotype in question, but deleting them both simultaneously has a significant effect on the phenotype. If such a gene pair results in a lethal phenotype, they are called a *synthetic lethal* (SL) pair. [6, p. 378-397]

The number of possible gene combinations typically surpasses several thousands depending on the number of genes in the target organism. To experimentally test each gene pair for SL properties would demand an enormous amount of time and work resources, in addition to knowing that the majority of combinations are unlikely to be an SL pair. Using GEMs as a tool in SL studies has the potential to minimize this resource demand by means of screening all gene combinations *in silico*. Then, the promising SL candidates can be verified (or refuted) experimentally; a strategy that has a significantly lower demand on resources.

2.3 Human GEMs

2.3.1 The emergence of human metabolic modeling

The first ever GEM of human cells, Recon 1, was published in 2007 [34]. Since then, the model has been further curated and updated to integrate updated knowledge and improve model performance, with publications of models such as Recon 2 [36], Recon 2.2 [37], and Recon 3D [38]. A number of other human GEMs have been published in the same time frame, including HMR1 [39], HMR2 [40], and iHsa[41]. As a response to the many challenges that arose when models were updated across different model series, a unified human model was created and published in 2020; namely Human 1 [9]. Human 1 was created to be an open-source, version-controlled model framework freely available online as a Git repository at https://github.com/SysBioChalmers/Human-GEM. This allows for continuous updating and model curation, as well as a way to unify the efforts of different research groups working towards the same goal of creating an all-encompassing GEM of the human metabolism. In other words, Human 1 is the state-of-the-art of human metabolic modeling.

[2]The possible applications of human GEMs are vast, and are expected to have far reaching consequences in the years to come. Health conditions such as obesity, diabetes, hypertension, cardiovascular disease, and cancer are all associated with abnormal metabolic states in the human body [42][43]. Understanding the metabolic changes that cause these diseases at a systems level can thus aid in the development of more effective treatments and therapies. Changes in metabolite concentrations can be utilized as biomarkers for disease diagnosis and monitoring, while enzymes related to the cause of disease can be targeted for medical intervention to treat the

[2]This paragraph is partly adapted from previous work [18].

disease [44]. In recent years, GEMs have generated insights into several aspects of health related issues, including inborn errors of metabolism [45], human microbiome co-metabolism [46][47], and cancer [9][48].

A selected example of novel insight gained by using GEMs is a study done on non-alcoholic fatty liver disease (NAFLD), a disease which for long time had unknown mechanism, ultimately leading to difficulties in diagnosing, monitoring and treating the disease. Mardinoglu *et al.* constructed a GEM of hepatocytes based on transcripomic data from NAFLD patients, and several analyses of the resulting metabolic network was performed [49]. It was discovered that these patients had severely down-regulated expression levels of enzymes in the liver related to serine biosynthesis. Consequently, these enzymes were suggested as potential drug targets for the treatment of NAFLD. It was also suggested that adding serine as a dietary supplement could benefit these patients, a hypothesis that found support in some animal and human studies. Lastly, altered blood concentrations of chondroitin and heparan sulphates were also observed in these patients, and consequently suggested as novel and non-invasive diagnostic markers, in lieu of the current practice of obtaining a biopsy for diagnosis.

2.3.2 Creating tissue specific models

[3]It is important to note that human GEMs such as Human 1, and the Recon and HMR model series, are so-called generic human metabolic models, meaning the models comprise all possible metabolic reactions that take place in humans. This does not accurately represent any real human cell because each cell type utilizes only a fraction of these reactions based on their function in the body. Therefore, tissue- or cell-specific models are usually built as a reduction of a generic human GEM. Transcriptomic or proteomic data from the desired cell type is typically used to remove reactions that do not occur in the target cell from the generic GEM, the biomass function may be re-evaluated, and manual curation may be used to further include or exclude relevant information to construct a cell-specific model. In this way, human metabolic models can be used to simulate many medical situations of interest involving specific organs or cells.

A method that is commonly used to create cell-specific models is the **tINIT** (Task-driven Integrative Network Inference for Tissues) algorithm [50]. This is an automated GEM reconstruction tool that uses proteomic and/or transcriptomic data of enzyme expression to retain only the metabolic activities found in a certain cell. Furthermore, tINIT requires that all reactions in the final model be capable of carrying flux as well as performing known key metabolic reactions, assuring high model functionality. The end result is a cell-specific GEM that is an *in silico* replica of a real-life cell, although based on our current and incomplete knowledge.

[3]This paragraph is adapted from previous work [18].

2.3.3 Choosing an objective function in human GEMs

Since the selection and formulation of objective function is vital to the analysis methodology and subsequently the results, the choice must be done carefully. For microbial cells it is commonly accepted that growth (i.e., biomass production) is a reasonable cellular objective [51]. For multicellular organisms, however, this assumption falls short as a cell has to fulfill a number of functions in the body [52]. Additionally, the objective of a human cell can differ between different cell types and tissues, between different cells within a tissue, and they can even change over time for the same cell [52][53]. Therefore, it is unlikely that a maximising a single objective function, such as biomass production, captures the full extent of all combination of functions that the human cell utilizes.

As a response to this challenge, Agren *et al.* constructed a set of 57 metabolic tasks that any human cell needs to be able to carry out in order to be considered viable [50]. 56 of these are metabolic tasks that can be categorized as energy and redox provision, internal conversions processes, substrate utilization and biosynthesis of metabolites. Examples include certain protein and nucleotide synthesis reactions, respiration processes, generation of membrane potentials and ADP rephosphorylation. The 57th task is the ability for growth, by means of biomass production, represented as 'growth on Ham's media'. The complete list, as provided by Agren *et al.*, can be found in Supplementary data file **S1**.

Choosing this set of metabolic tasks to be completed instead of merely biomass production to determine gene essentiality has been found to improve the sensitivity of numerous GEM predictions when compared to experimental results [9].

2.4 Cancer

2.4.1 The hallmarks of cancerous disease

Cancer is one of the most dreaded diseases of our time, and rightfully so by accounting for nearly 10 million deaths worldwide in 2020 [54]. It is regarded as a complex disease, involving multiple dynamic changes in the genome such as oncogenes, DNA repair systems, tumor suppressors, growth and apoptosis regulation, and more [3].

[4]The formation of cancers, or more specifically neoplastic transformation, is characterized by alterations of cellular metabolism. The uncontrolled cell proliferation displayed by cancer cells results from changes in biomass synthesis, energy requirements and altered regulatory and functional properties [55]. For instance, the Warburg effect is a common phenomenon in cancer cells where the aerobic glycolysis process is elevated, leading to increased glucose uptake [56][57]. Increased glutamine uptake and metabolism to promote cellular growth, often called glutamine addiction, is also frequently observed in cancer cells [55]. In summary, it can be said that the metabolic dysregulation and altered functioning is both the cause and the consequence

[4]This paragraph partly adapted from previous work [18].

of tumorigenesis [58].

Although cancer has been extensively studied, the use of GEMs can provide new perspectives and insights on the metabolic alterations of the cancerous cells. Despite the fact that tumors may share many metabolic traits, GEM based studies has shown that cancer-induced alterations in metabolic gene expression were very variable among different tumor types, and that no uniform metabolic transformation relates to all tumors [10]. In other words, although all cancers are characterized by altered metabolism, these changes are largely dissimilar among different cancer types. This supports the notion that various cancer types would probably benefit from being treated as distinct diseases, especially in the search of uncovering novel targeted therapies.

2.4.2 Cancer progression

There are several ways to categorise the severity of the cancerous disease in a patient, with a common way being to label the *stages* of cancer. This approach specifies the tumor's size and growth type, as well as whether it has spread to other areas of the body, all of which affects the best treatment option and the prognosis. Most cancers fit into four stages, briefly summarized below. [59]

- **Stage 1**: The tumor is small and contained within the organ of its origin.
- **Stage 2**: The tumor is larger than stage 1, but has not spread to surrounding tissue. Depending on the cancer type, this stage might also include that cancer cells has spread to surrounding lymph nodes.
- **Stage 3**: The tumor is larger than stage 2, it has spread to nearby lymph nodes and may have started to spread into surrounding tissues.
- **Stage 4**: The cancer has spread from its organ of origin to other areas of the body. This is also called **metastatic** cancer.

A *primary tumor* is a term used to describe the original, or first, tumor in the body [60]. Usually, this corresponds to stages 1 and 2, although sometimes stage 3 is included. The underlying mechanisms of why some cancers metastasizes is not fully understood, although new hypotheses, such as the cancer stem cell models, are emerging and gaining support.

In recent years, the study of cancer stem cells (CSCs) and their role in tumor initiation and progression has gained focus. CSCs are a rare and distinct subpopulation of cancer cells that have been identified in most types of human cancer [61], and are thought to be responsible for cancer initiation, progression, metastasis, recurrence and drug resistance [62]. However, the metabolic adaptations underlying the properties of CSCs have been challenging to study, partly due to difficulties of isolating CSCs from other tumor cells [62]. Nevertheless, researchers have managed to identify several altered metabolic pathways of CSCs, such as the Notch, Hedgehog

and Wnt signaling pathways [62]. Accumulated evidence suggests that the elimination of CSCs is essential to treating and preventing recurring and metastatic tumors [62], and thus to gain deeper understanding of their metabolic properties could be a reasonable place to start.

2.4.3 GEMs as a tool in cancer research

The bottleneck of creating novel cancer medicines is not to create chemicals that kills cancer cells, as thousands of such chemicals have been discovered and developed the last 50 years [63]. The fundamental challenge is to discover anticancer drugs that are effective, without also causing considerable harm to the patient. Since cancer originates from our own cells, many of the currently used anticancer drugs affect targets that are shared between normal cells and cancer cells, such as enzymes involved in DNA replication [63]. Therefore, to find a target that is uniquely present for cancer cells is imperative to minimize harm to the patient.

The fact that cancer cells have an altered metabolic functioning suggests that it could be possible to find genes that are essential for cancer cells, yet not essential for normal cells, and vice versa. Essential genes, especially those that are cancer-specific, can be used as a surrogate for drug targets [14]. This means that essential gene analyses can be used to discover what enzymes are indispensable to the cell, and therefore would be fatal if inhibited by a drug.

In 2011, a GEM of generic cancer metabolism was able to predict 52 cytostatic drug targets, of which 40% was, at the time, targeted by known, approved or experimental anticancer drugs, and the rest were new [64]. Another study performed using glioblastoma (GBM) GEMs identified five genes that were uniquely essential for the cancer cells, yet non-toxic if removed from healthy brain tissue GEM [65]. Four of these were found to be experimentally verified, pointing to these as promising drug targets for the treatment of GBM.

Likewise, the concept of synthetic lethality has gained interest in the medical research community, and it has even been referred to as one of the most effective cancer therapies in the last decade. For instance, several different DNA damage response (DDR) inhibitors are being tested as a synthetic lethal targets for various cancer cells with promising results. [66]

Chapter 3

Software and methods

This chapter will outline all software and methods used to obtain the results that are presented in this thesis. The self-written scripts mentioned in this section, as well as some key functions used, are accessible at https://github.com/rebeccagsandtroen/MastersThesis.

3.1 Software

3.1.1 MATLAB

MATLAB programming language was used to write scripts that generated virtually all data presented in this thesis. Analyses using GEMs was done with a combination of self-written MATLAB code and functions from either COBRA or RAVEN Toolbox (see Section 3.1.2), or both. Any exceptions to this will be clearly stated. Processing and refining of results were done by self-written code and MATLAB integrated functions. MATLAB version R2021a [67] with an academic licence was used for these purposes.

3.1.2 COBRA and RAVEN Toolboxes

The **COBRA (The COnstraint-Based Reconstruction and Analysis) Toolbox** v3.0 [68] is a commonly used software package for constraint-based metabolic modeling, as well as analyzing and predicting metabolic phenotypes utilizing genome-scale biochemical networks.

The **RAVEN (Reconstruction, Analysis and Visualization of Metabolic Networks) Toolbox** v2.5.3 [69] is a software package utilized for reconstruction, curation, and constraint-based modeling and simulation of GEMs. It includes a range of methods for analysing and visualisation of metabolic networks and omics data.

Working with the models was primarily done by use of the COBRA Toolbox, which included exploring model features as well as doing single gene knockout analysis described in Section 3.4.1. The RAVEN Toolbox was utilised for the other analyses, including synthetic lethal gene pair analysis, and to investigate the ability of models to perform various metabolic functions in different gene deletion scenarios. The specific toolbox functions that were utilised are found in the respective analyses method description.

3.1.3 Gurobi

The Gurobi Optimizer [70], a commercial mathematical optimization solver, was used for optimization problems. Gurobi provides an interface for accessing it from MATLAB, and was it was obtained with an academic licence.

3.1.4 Microsoft Office

Microsoft Excel version 2205 [71] was used for sorting results of the gene essentiality and synthetic lethality results. This includes sorting by weight, counting the number of times different genes occurred, counting how many times various weight scores occurred, and creating figures from these data.

3.1.5 Cytoscape

Cytoscape version 3.9.1 [72] with the NetworkAnalyser plugin [73] was used to create network systems of the gene essentiality and synthetic lethal analyses, as well as identifying the degree of connections of genes in said networks.

3.2 Flux Balance Analysis

All flux balance analyses were performed using MATLAB, with Gurobi as the optimizer. The solver parameters, optimality tolerance (optTol) and feasibility tolerance (feasTol), were both set to 10^{-6}, by default. Unless otherwise stated, the objective function set to be maximized in the models were that of biomass production; reaction name `biomass_human`.

3.3 Model retrieval and initial verification

It was chosen to use models created by Robinson *et al.*, which were provided with their publication [9] as a Zenodo repository [74]. All files in this repository are available for use with Creative Commons Attribution 4.0 International licence. 15

models were chosen and retrieved, each of which was associated to one of five different cancer types (Table 3.1). Specifically, each chosen cancer type had three associated models: normal solid tissue (NT), primary solid tumor (TP), and either recurring solid tumor (TR) or metastatic tumor (TM). The three GEMs connected to a specific cancer type (NT, TP and either TM or TR), will henceforth be referred to as a model subgroup. The cancer types was selected based on factors such as high prevalence, poor prognosis following diagnosis, and gender representation.

The models were created by Robinson *et al.* through a GEM contextualization process described as follows [74]. Using the generic human GEM, Human1, as the reference model, tissue specific GEMs were generated using the tINIT algorithm, and Gurobi (version 8.0.0) as the solver. The input genetic expression data were RNA-Seq data, with an expression threshold of 1 transcripts per million (TPM), collected from The Cancer Genome Atlas (TCGA). The median expression of each gene across all patient samples was calculated for each tissue type and used as input to the algorithm. NT models represent healthy tissue from which various tumor tissues (TP, TM, or TR) arose.

Table 3.1: List of models used in this thesis. Sample type codes: NT = solid tissue normal, TP = primary solid tumor, TM = metastatic tumor, TR = recurring solid tumor.

Cancer type	Cancer abbreviation	Model name
Breast invasive carcinoma	BRCA	BRCA NT BRCA TM BRCA TP
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	CESC NT CESC TM CESC TR
Lung adenocarcinoma	LUAD	LUAD NT LUAD TP LUAD TR
Pancreatic adenocarcinoma	PAAD	PAAD NT PAAD TM PAAD TP
Prostate adenocarcinoma	PRAD	PRAD NT PRAD TM PRAD TP

Since the models were built using actual patient expression data, they should be able to realistically reproduce the metabolic skills of living cells, including the 57 metabolic tasks deemed essential for all human cells (detailed in Section 2.3.3, and found in Supplementary data file **S1**). This was verified by using RAVEN `checkTasks` function for all models, with the 57 metabolic tasks as input. The function description states that *"The tasks are defined by defining constraints on the model, and if the problem is feasible, then the task is considered successful."* Specifically, the func-

tion adds upper and lower constraints on available metabolites and/or the expected products, and on flux constraints on the reaction of the task being tested. This "forces" the model to carry flux through a reaction, or produce products that are required for the task to be regarded as complete. With these constraints in place, the model is solved with FBA, and biomass production as objective to be maximized. A model "passes" a test if there exists a feasible flux distribution solution that satisfies all given constraints. If no such solution were found without violating any constraints, the model would be considered to have "failed" to perform the metabolic tasks at hand. For example, one of the metabolic tasks is 'Aerobic rephosphorylation of ATP from glucose', represented as the following chemical equation:

$$ATP[c] + H_2O[c] \Rightarrow ADP[c] + P_i[c] + H^+[c]$$

The allowed input metabolites, O_2[s] and glucose[s], are both constricted to lower and upper uptake bounds 0 and 1000 units, respectively. Output (produced) metabolites, H_2O[s] and CO_2[s], are also both constricted to bounds of 0 to 1000 units produced. In other words, the model is allowed to take up as much oxygen and glucose as needed, as well as to excrete high amounts of water and CO_2 if necessary. The equation flux is set to lower and upper bound of 1 and 1000 units, respectively. This means that there needs to be a flux through this reaction of at least 1 unit in order to satisfy the constraints and be considered a feasible problem. If the model is capable of producing a feasible flux distribution within all these constraints, the model "passes" the test for this metabolic task. [75]

3.4 Essential Gene Analyses

Single gene deletion (SGD) analysis was performed on all 15 GEMs with two different methods, differentiated by their respective definitions of essentiality. The first method was to use the ability to produce biomass as a metric for evaluating cell viability.

As discussed in Section 2.3.3, using biomass production as the objective function alone might fail to capture the complexity of the actual objectives of real-life human cells. Therefore, a second method was chosen to investigate potential differences in results when the definition of essentiality is changed. This method was to use the ability to carry flux through all 57 essential metabolic tasks (discussed in Section 2.3.3) as a metric for cell viability.

Genes were identified by their Ensembl identifiers, as they are provided in the model structures. Direct mapping of Ensembl ID to gene names were done using the **BioTools** website, designed and maintained by Andy Saurin, accessed at https://www.biotools.fr/human/ensembl_symbol_converter. Unless otherwise specified, all instances where gene names along with gene product information are presented, this information was gathered from The Human Protein Atlas [76] at www.proteinatlas.org.

3.4.1 Essential gene analysis based on biomass production

The first essential gene analysis was done by using the cells' ability for biomass production as a predictor of cell viability. A gene was deemed as essential if its deletion caused predicted growth rate to drop to 50% or below of the wild-type cellular growth rate, a limit that have been utilized for human *in silico* models elsewhere [14] [77]. These genes will be called *BP-essential genes* (Biomass Production-essential genes).

SGD analysis were performed according to the following procedure for all 15 tissue models:

1. The model was imported into MATLAB.
2. Growth media constraints were set to only allow the model to take up metabolites present in Hams' medium, of which the composition is reported in Supplementary data file **S2**. The uptake bounds of the metabolites was set to -1000, corresponding to 'unlimited access' for the cell, with no flux limits on the exchange rates. The function `setHamsMedium` was used for this purpose, made by Robinson *et al.* [74].
3. It was verified that the objective function was set to `biomass_human` using COBRA `checkObjective` function.
4. The model was converted from a RAVEN compatible format into a COBRA compatible format using the function `ravenCobraWrapper` from the RAVEN Toolbox.
5. SGD analysis was performed using the `singleGeneDeletion` function in the COBRA Toolbox, with "FBA" as the method to be used and biomass production as the objective function to be maximized. The predicted growth rates for wild type and for each deletion strain were calculated, as well as the relative growth rate between wild-type and KO-strains.
6. Gene deletion mutants that exhibited a predicted relative growth rate lower than the set threshold were deemed *not viable*. The corresponding genes were classified as BP-essential genes.
7. BP-essential genes for tumor tissue GEMs (TP, TR and TM) were pairwise compared to their normal tissue (NT) counterpart. The BP-essential genes identified in the tumor tissue that were also found in the NT were excluded from further analysis. In other words, genes that were found to be essential just for a tumour tissue but not for its normal tissue counterpart were gathered and named *BP-target-genes*.

3.4.2 Essential gene analysis based on metabolic tasks

For this SGD analysis, the list of 57 essential metabolic tasks for human cells were utilized (provided in Supplementary data file **S1**). Due to the fact that all 57

metabolic functions are assumed to be required for cell survival, essential genes were defined as those whose deletion impaired any of the 57 metabolic tasks. The genes identified in this analysis will hereby be referred to as *MT-essential genes* (Metabolic Task-essential genes).

SGD analysis was performed according the following procedure for all 15 models. Note that no particular medium was used, as the RAVEN function `checkTasksGenes` automatically sets the specific uptake rates for the various analyses.

1. The model was imported into MATLAB.

2. To perform SGD with the 57 tasks as the metric, the RAVEN function `checkTasksGenes` was used. Inputs consisted of a model and the task list of the 57 metabolic tasks. For each gene in the model, the impact of gene deletion on each of the 57 tasks was calculated in the form of 'pass' or 'fail'. 'Pass' signifies that the reduced model is able to sustain a certain amount of flux through the reaction of the metabolic task being tested, and 'fail' means the model is incapable of carrying said flux.

3. The output of `checkTasksGenes` was a binary vector, f, that detailed the pass/fail result of every task for every deleted gene. The number of failed tasks for each gene deletion mutants were summarized as shown in Equation (3.1). In other words, each gene were given a score based on how many metabolic tasks that failed by its deletion. This score will be referred to as the *weight* of the gene, denoted by ω. E.g., a weight of 3 means that the deletion of the gene causes three of the 57 tasks to fail. Genes that did not impair any of the tasks, i.e. had a weight of 0, were deemed non-essential, and were not further studied. Genes that had a weight of at least 1 were classified as MT-essential genes.

$$\omega = \sum_{i=1}^{n} f_i \tag{3.1}$$

4. MT-essential genes for tumor tissue GEMs (TP, TR and TM) were pairwise compared to their normal tissue (NT) counterpart. The MT-essential genes identified in the tumor tissue that were also found in the NT were excluded from further analysis. In other words, genes that were found to be essential just for a tumour tissue but not for its normal tissue counterpart were gathered and named *MT-target-genes*.

5. The Human Protein Atlas, accessed at https://www.proteinatlas.org, was used to investigate each of the MT-target-genes. Information about gene name, gene product, protein function and metabolic subsystem were gathered.

6. The Drugbank database, accessed at https://go.drugbank.com/, was used to investigate whether there are any known medications that target the specific gene products of these MT-target-genes. The focus was to find inhibitory drugs, as inhibiting the gene product would, by definition of this analysis, be lethal to the tumor tissue in question but not the healthy tissue counterpart.

The number of identified MT- and BP-essential genes for each tissue model were compared in order to illustrate the difference in results by the different methods. The biomass proportion α of the MT-essential genes were calculated for each model as described in Equation (3.2). Cytoscape was used to create network representations of the BP- and MT-target genes.

$$\alpha = \frac{\text{Number of BP-essential genes}}{\text{Number of MT-essential genes}} \tag{3.2}$$

3.5 Synthetic Lethal Gene Pair Analysis

The following procedure was done for each model to identify synthetic lethal (SL) gene pairs. Due to computational limitations, the analysis was only done for the BRCA related models; BRCA NT, BRCA TM, and BRCA TP.

1. The model was imported into MATLAB.
2. The known MT-essential genes that were identified in Section 3.4.2 were removed from the gene list, while keeping their related reactions in the model.
3. Double gene deletion was then performed as follows. For every remaining gene in the model, a gene was deleted by using the function `deleteModelGenes` from COBRA Toolbox. Then the RAVEN function `checkTasksGenes` was deployed to perform the second gene deletion, and calculate its impact on the ability to carry flux through the 57 metabolic tasks. The impact of gene deletion on the individual tasks was classified as 'pass' or 'fail' if the reduced model were able to carry flux through the task in question, or not, respectively.
4. All gene pairs that, when deleted, passed all of the 57 tasks were classified as a non-lethal pair, and subsequently excluded from further study. The gene pairs that resulted in 1 or more failed tasks were deemed an *SL-pair*.
5. The weight of the SL-pair, meaning how many tasks that failed when the gene pair was deleted, was calculated using Equation (3.1).
6. SL-pairs in tumor models (TM and TP) that were also found in NT were excluded from further study. The remaining SL-pairs were named *SL-targets*, and are, by definition, lethal when pair-wise deleted in the tumor, but not lethal when deleted in the normal tissue.
7. Duplicates, in the form of GeneA+GeneB and GeneB+GeneA, were removed as to only preserve one of the two, as they, per definition, would be the identical double gene deletion scenario.
8. Cytoscape was used to create network representations of the SL-targets.

Chapter 4

Results and Analysis

This chapter presents the results from the various analyses performed on the normal- and cancer tissue GEMs. The results have been divided into two sections, the first focusing on exploring essential genes specific to tumours and the second on investigating synthetic lethal gene pairs specific to BRCA tumours.

4.1 Model verification

The first analysis to be done after the models were retrieved was to investigate whether the they, without any alterations or additionally added constraints, could perform all 57 metabolic tasks deemed essential to be a viable human cell of any kind. A failed task in this case would indicate that the model does not adequately reflect a real human cell, as the models are based on patient expression data from living, functioning cells and hence should be able pass all tests.

All of the models passed all tests, demonstrating to be capable of performing all 57 metabolic tasks. An excerpt of the results, which were identical for all models, is presented in Figure 4.1.

```
PASS: [ER] Krebs cycle NADH
PASS: [ER] Ubiquinol-to-proton
PASS: [ER] Ubiquinol-to-ATP
PASS: [SU] Beta oxidation of saturated FA
PASS: [SU] Beta oxidation of long-chain FA
PASS: [SU] Beta oxidation of odd-chain FA
PASS: [SU] Beta oxidation of unsaturated fatty acid (n-9)
PASS: [SU] Beta oxidation of unsaturated fatty acid (n-6)
PASS: [SU] Uptake and beta oxidation of all NEFAs
PASS: [SU] Choline uptake
PASS: [SU] Inositol uptake
PASS: [BS] Phosphatidylcholine de novo synthesis
PASS: [BS] Phosphatidylethanolamine de novo synthesis
PASS: [BS] Phosphatidylserine de novo synthesis
PASS: [BS] Phosphatidylinositol de novo synthesis
PASS: [BS] Thiamin phosphorylation to TPP
PASS: [BS] Coenzyme A synthesis from pantothenate
PASS: [BS] FAD synthesis from riboflavin
PASS: [BS] Heme biosynthesis
PASS: [GR] Growth on Ham's media (biomass production)
```

Figure 4.1: Excerpt of the model verification results. "Pass" means the model was able to perform the metabolic task being tested.

4.2 Essential Gene Analyses

4.2.1 Biomass Production essential genes

SGD analysis was performed for all models, with biomass production as the objective function to be maximized. Genes were deleted one at the time, followed by a calculation of the reduced model growth rate and the relative growth rate compared to the wild type model. Genes that, when deleted, caused the relative growth rate to fall below a chosen threshold were deemed biomass production essential (BP-essential) genes. The results of the analysis is presented in Table 4.1. The number of BP-essential genes differs to some extent within model subgroups, and the numbers of BP-target genes demonstrate that all tumour models have acquired a slightly different set of BP-essential genes than their normal tissue counterpart. Taken together, this implies that the tumour models have undergone some alterations in terms of biomass production-related metabolism. The complete list of BP-essential genes and BP-target-genes can be found in Supplementary data files **S3** and **S4**.

Table 4.1: SGD analysis for the tissue models, using the ability of biomass production (BP) as the metric for cellular viability. The number of genes identified in the models as essential for biomass production (BP-essential genes) are listed. BP-target-genes are essential genes that are only essential for a tumor model, but not essential for the corresponding normal tissue model in regards to BP.

Model	Number of BP-essential genes	Number of BP-target-genes
BRCA NT	191	-
BRCA TM	192	11
BRCA TP	198	13
CESC NT	197	-
CESC TM	194	11
CESC TP	197	9
LUAD NT	199	-
LUAD TP	201	4
LUAD TR	194	3
PAAD NT	192	-
PAAD TM	195	7
PAAD TP	192	6
PRAD NT	190	-
PRAD TM	193	11
PRAD TP	183	2

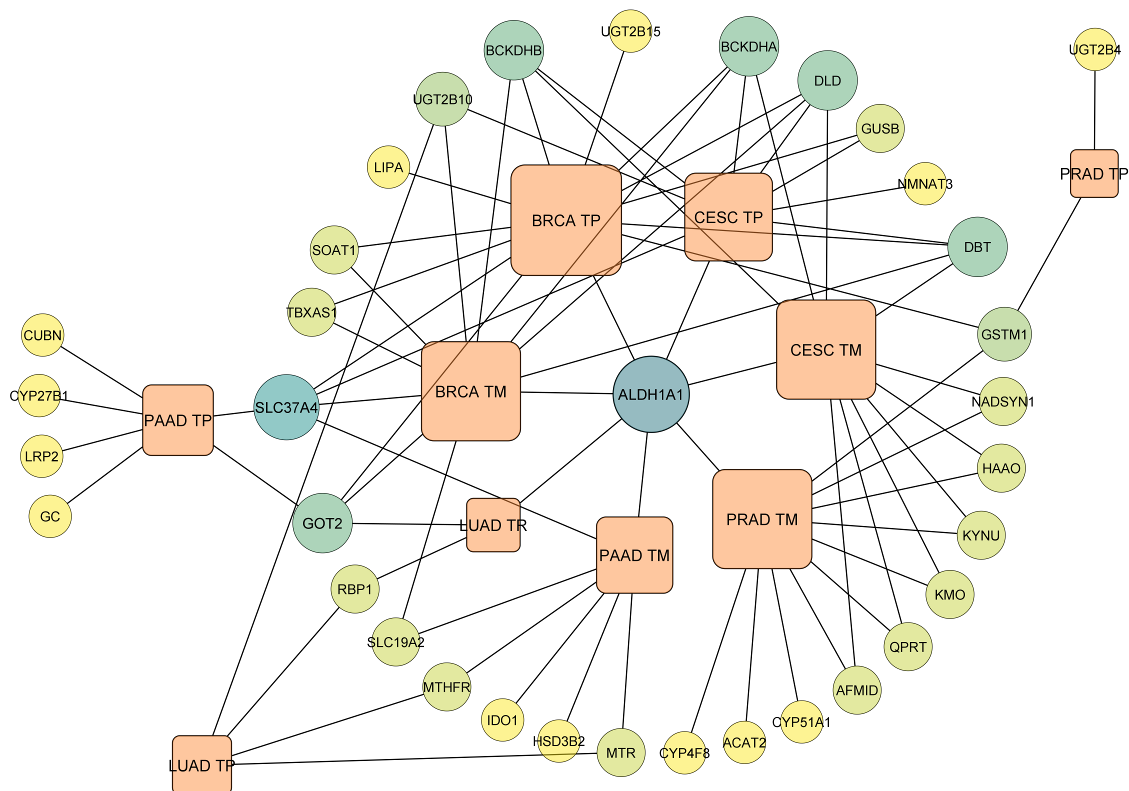


Figure 4.2: Network representation of the BP-target-genes (circles) identified for the various tumor types (squares), collectively called *nodes*. Increasing size of the nodes signifies increased number of connections (*edges*). An edge between two nodes signifies that the gene is a BP-target for that tissue type. The colour gradient of the circles also represent the number of edges, with yellow signifying just one edges, green meaning a few edges, and blue representing several edges.

From the network representation of BP-target-genes (Figure 4.2) it is observed that many BP-target-genes are connected to several tumor types. The most connected gene, ALDH1A1, is connected to seven tumor types, and the second most connected gene, SLC37A4, has five tumor connections. Out of the total of 35 unique BP-target-genes identified in this analysis, merely 13 of these are only connected to one tumor type. This means that the majority (63%) of BP-target-genes, specifically 22 genes, are BP-target-genes for two or more tumor types.

4.2.2 Metabolic Task essential genes

In this analysis, SGD was performed for all genes, for all models. The reduced models were then tested for their ability to carry flux above a set amount for all 57 metabolic tasks essential for human cell viability. If the model was able to carry said flux, the task was marked as "passed", and if not the task is deemed as "failed". Genes that, when deleted, resulted in the model failing at least one task were deemed metabolic task essential (MT-essential) genes, with the weight of that

gene signifying the number of failed tasks. MT-target-genes are the MT-essential genes that are essential for a tumor, but not for the corresponding normal tissue.

The number of MT-essential genes, as well as MT-target-genes, varies within the model subgroups, as seen in Table 4.2. Building on the analysis of BP-essential genes, the disparity observed in MT-essential genes within subgroups indicates that metabolic changes of tumours extend beyond those linked to biomass production and into various domains of cellular metabolism.

Table 4.2: SGD analysis for the tissue models, with the ability to perform the 57 Metabolic Tasks (MT) essential to human cells as the indicator for cellular viability. The number of genes identified in the models as essential for completing all metabolic tasks (MT-essential genes) are listed. MT-target-genes are genes that are only essential for a tumor model, but not essential for the corresponding normal tissue model.

Model	Number of MT-essential genes	Number of MT-target-genes
BRCA NT	302	-
BRCA TM	305	12
BRCA TP	309	15
CESC NT	306	-
CESC TM	293	15
CESC TP	304	11
LUAD NT	306	-
LUAD TP	308	5
LUAD TR	292	7
PAAD NT	300	-
PAAD TM	302	7
PAAD TP	300	6
PRAD NT	295	-
PRAD TM	300	13
PRAD TP	288	8

The number of MT-essential genes identified that overlaps within the model subgroups are illustrated in Figure 4.3. It is observed that the majority of MT-essential genes discovered are shared within the subgroups, as would be expected given that the tumours developed from the corresponding normal tissue and hence would retain a great deal of similarities. Curiously, all NT models were found to possess essential genes that were not essential to either related tumor type. However, this is outside the aim of this thesis and will not be discussed further. With the exception of the BRCA subgroup, it was found that TP and TM or TR from a model subgroup share merely 1-4 target-genes (right side middle of the venn diagrams). This points to a certain degree of metabolic difference between primary tumors and the tumors that spread (TM) or recurs after treatment (TR). The BRCA tumors shared 9 MT-target-genes, suggesting that these tumors have a somewhat higher degree of metabolic similarity than what were observed in the other model subgroups.

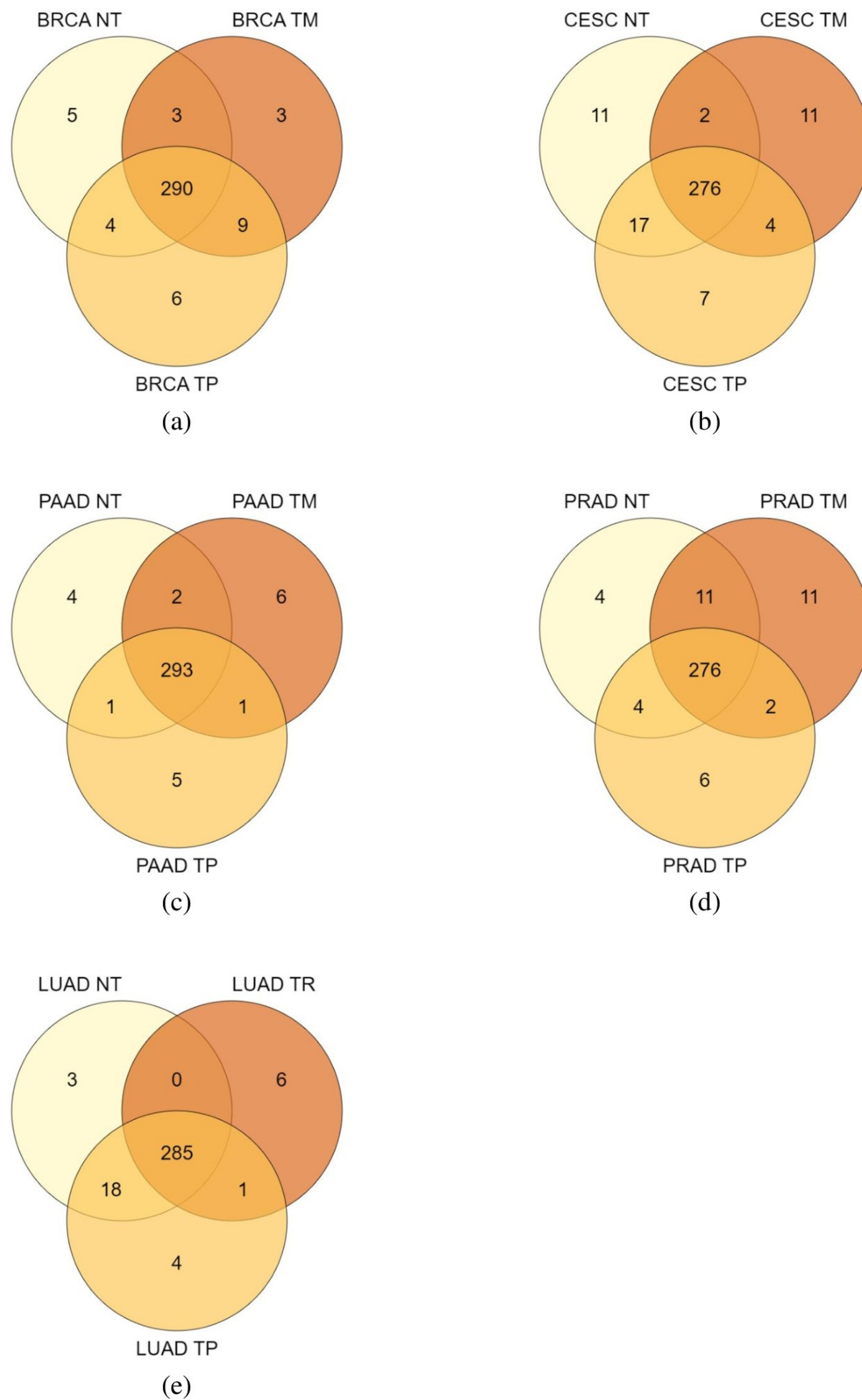


Figure 4.3: MT-essential genes overlaps between the three tissue types within the model subgroups; (a) BRCA, (b) CESC, (c) PAAD, (d) PRAD, and (e) LUAD.

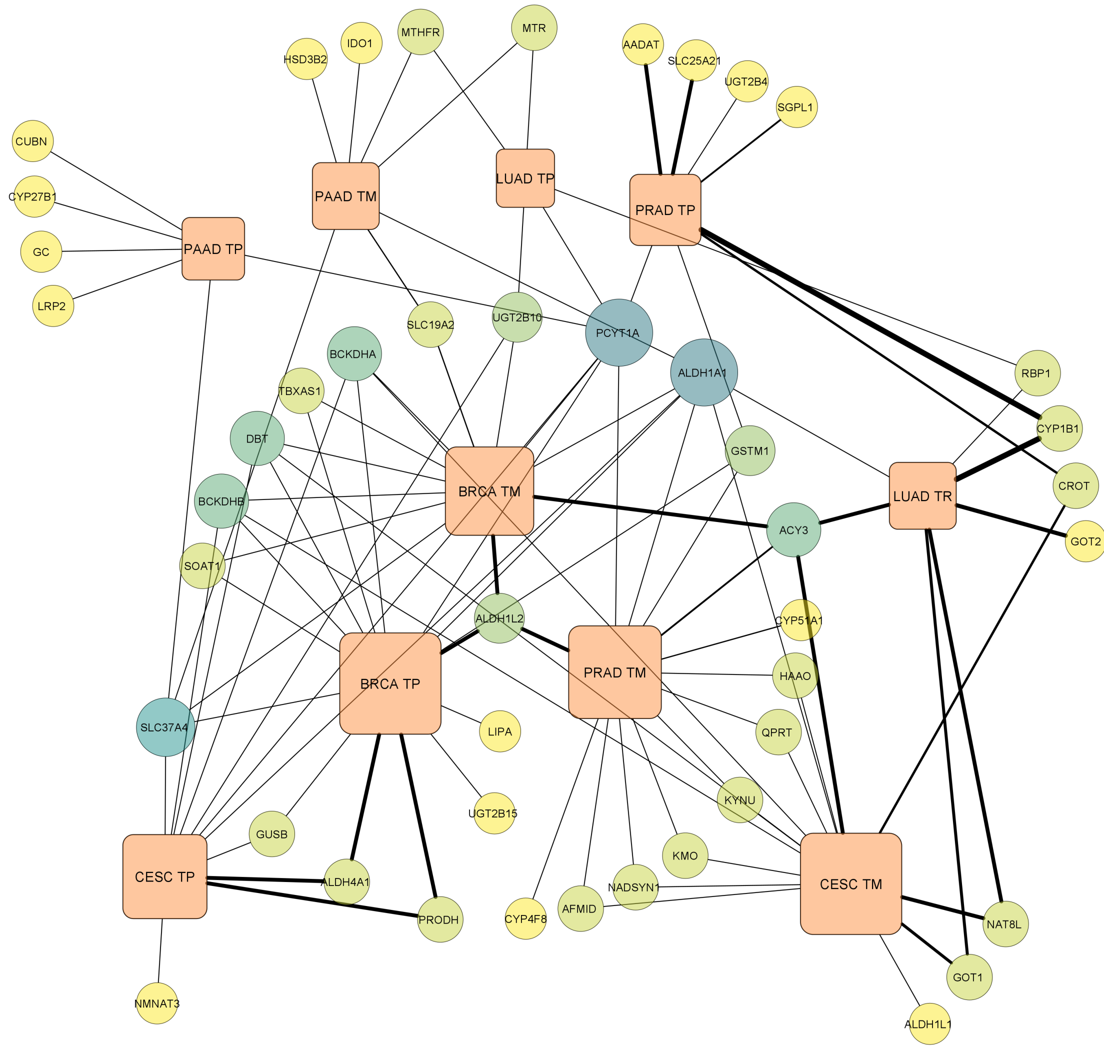


Figure 4.4: Network representation of the MT-target-genes (circles) identified for the various tumor types (squares), collectively called *nodes*. Increased thickness of the lines (edges) represent increased weight of the MT-target-gene. Increasing size of the nodes signifies increased number of edges. The colour gradient of the circles also represent the number of edges, with yellow signifying just one edge, green meaning a few edges, and blue representing several edges.

Several MT-target-genes are identified in multiple tumor types, as illustrated in Figure 4.4. Out of the 46 unique MT-target-genes identified in this analysis, 17 genes were only connected to one tumor model. Hence, the remaining 29 MT-target-genes, which makes up 63%, are connected to two or more tumor models. These highly connected genes could point to metabolic alterations that are common for several cancer types, which consequently makes for interesting drug targeting potentials. All MT-target-genes that appear in more than one tumor model can be seen in Figure 4.5, as well as for how many tumor models they appear as an MT-target-gene.

At the top of the list is the gene ALDH1A1 which was found to be a MT-target-gene

for 7 different tumor models, namely BRCA TM, BRCA TP, CESC TM, CESC TP, LUAD TR, PRAD TM, and PAAD TM. Likewise, the gene PCYT1A was identified as MT-target-gene for the following 7 models; BRCA TM, BRCA TP, CESC TP, LUAD TP, PRAD TM, PRAD TP, PAAD TP.

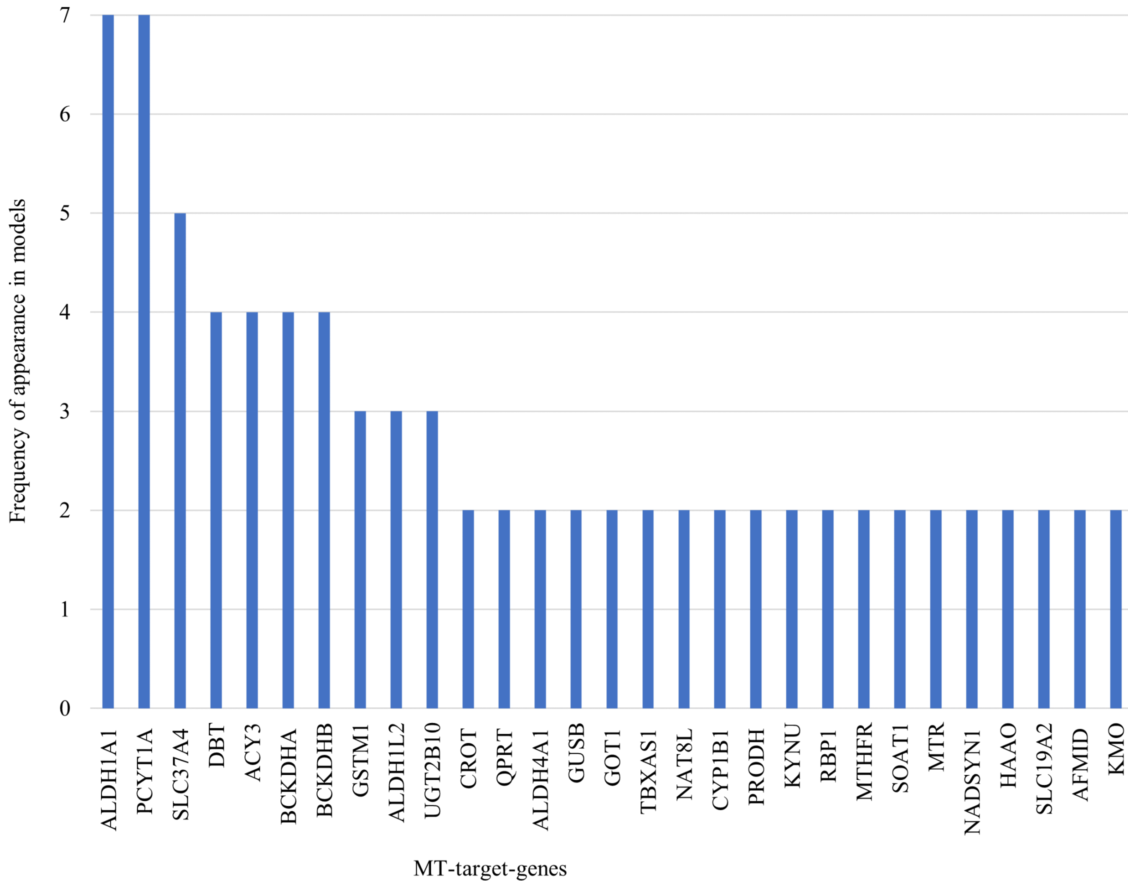


Figure 4.5: The frequency of the MT-target-genes that appears in more than one tumor model. A total of 23 MT-target-genes were identified as such.

Metabolic impact of gene deletion

Many of the reported MT-target-genes caused more than one of the 57 essential metabolic tasks to fail, with the number of failed tasks referred to as the genes' weight. The distribution of weight classes is presented in Figure 4.6. It can be observed that the majority of MT-target-genes have a weight of 1, while some genes have substantially higher weight, ranging up to 17. The heaviest MT-target-genes could be interesting to further investigate as potential drug targets since their deletion causes substantial disruption to the metabolic functioning of the cancerous cell. All MT-target-genes with weight more than 1 are presented in Table 4.3. Interestingly, several of the highest scoring MT-target-genes appear for several tumor types. Examples are CYP1B1 found in LUAD TR and PRAD TP; ACY3 in BRCA TM, CESC TM, and LUAD TR; and ALDH1L2 found as MT-target-gene for BRCA TP, BRCA TP, and PRAD TM.

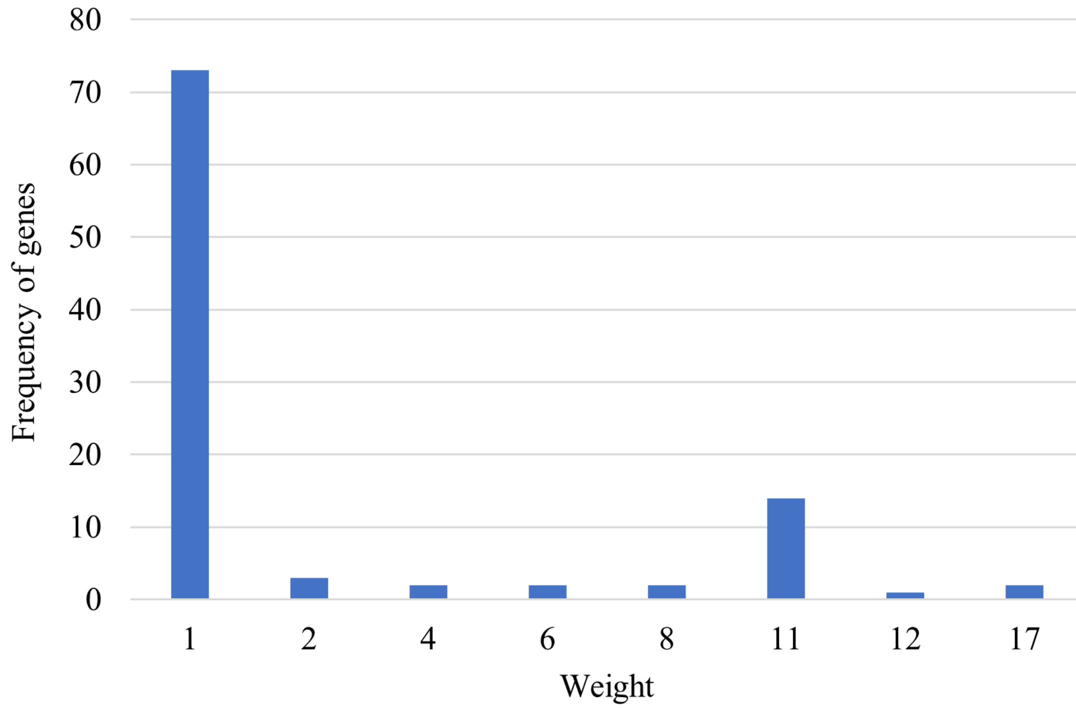


Figure 4.6: The weight distribution of MT-target-genes. Weight represents the number of metabolic tasks the SGD mutant model failed to perform, when testing the 57 essential metabolic tasks all human cells must be able to perform to be considered viable. The frequency depicts the total number of observations of MT-target-genes with this weight for every tumor model.

Table 4.3: The identified MT-target-genes with weight more than 1, and the corresponding tumor models in which they appeared as MT-target-genes.

MT-target-gene	Model	Weight
CYP1B1	LUAD TR	17
	PRAD TP	17
GOT2	LUAD TR	12
ACY3	BRCA TM	11
	CESC TM	11
	LUAD TR	11
ALDH1L2	BRCA TM	11
	BRCA TP	11
	PRAD TM	11
ALDH4A1	BRCA TP	11
	CESC TP	11
NAT8L	CESC TM	11
	LUAD TR	11
PRODH	BRCA TP	11
	CESC TP	11
SLC25A21	PRAD TP	11
AADAT	PRAD TP	11
GOT1	CESC TM	8
	LUAD TR	8
CROT	CESC TM	6
	PRAD TP	6
ACY3	PRAD TM	4
SGPL1	PRAD TP	4
CYP51A1	PRAD TM	2
SLC19A2	PAAD TM	2
	BRCA TM	2

Existing drugs for MT-target-genes
It was investigated whether the MT-target-genes had any known drugs targeting their specific gene product. A literature search was done in the DrugBank database, with focus on finding inhibitory or antagonistic drugs, and excluding results classified as 'Investigational', 'vet approved', or 'neutraceutical' only. It was found 9 unique MT-target-genes to have known inhibitory drugs, 8 of which were not related to cancer treatment and thus will not be explicitly presented here.

Only one gene was found to have known drugs used in cancer treatment, namely CYP1B1, as seen in at the top in Table 4.3, which had 15 approved inhibitory drugs,

many of which are used to treat various cancers. A chosen selection of these drugs are presented below, all satisfying the criteria of being able of inhibiting CYP1B1, being classified as 'approved', and be related to cancer treatment. All drug information below was collected from the DrugBank database at https://go.drugbank.com/.

- **Daunorubicin** - used in treatment of leukemia and other neoplasms.
- **Doxorubicin** - used to treat various cancers, including breast carcinoma, ovarian carcinoma, several leukemia types, and Kaposi's Sarcoma.
- **Mitoxantrone** - used in treatment of multiple sclerosis, several types of leukemia, metastatic breast cancer, prostate cancer, and hepatocellular carcinoma.
- **Paclitaxel** - used as first-line and subsequent therapy for the treatment of advanced carcinoma of the ovary, and other various cancers including breast and lung cancer.
- **Cannabidiol** and **Dronabinol** - used as adjunctive treatments, and symptom and pain relief of various painful conditions, including cancer.

The complete list of all MT-essential genes and MT-target-genes for all tumor models, including their respective weight in the SGD analysis, their biological function, known related drugs, and more, can be found in Supplementary data files **S3** and **S4**.

4.2.3 Comparing the two methods of SGD

The number of essential genes identified by the two methods of SGD are presented in Table 4.4, and illustrated in Figure 4.7. When compared to the MT-essential genes analysis, the BP-essential genes analysis yields just below two thirds the number of essential genes, averaging on 64,48 %. Specifically, around 300 genes (ranging from 288 to 309) were deemed MT-essential for the models, compared to around 195 genes (ranging from 190 to 201) that were classified as BP-essential. It's worth mentioning that one of the 57 metabolic tasks is "growth on Ham's media," which is equivalent to biomass production. In other words, the BP-essential findings can be regarded as a subset of the MT-essential results. This was also the reason as to why MT-related results were studied in more detail than the BP-related results.

To test this assumption, it was examined whether the same genes were discovered in both methods, particularly whether the BP-targets were a subset of the MT-targets. Except for two genes, DLD and ACAT2, all of the BP-target genes were also present among the MT-target genes. The reason behind why these two genes were not found as MT-could stem from the different definitions of essentiality of the two methods of SGD. It is probable that the two mutants exhibited a lowered growth rate to less than half that of the wild type while still producing enough biomass to "pass" the MT test. However, the MT based analysis identified 13 unique target-genes not found in the BP based analysis, which to some extent was expected due to the wider scope of the analysis method.

From Table 4.4, it can be seen that the TM or TR models displays the highest biomass proportion (α) within its model subgroup for CESC, LUAD and PAAD. In other words, the more severe tumor types may be more sensitive to disturbances in genes related to biomass production than their TP (or even NT) counterpart. However, this is not the case for BRCA and PRAD subgroups, which neither displays any obvious trends. It should be noted that the difference in biomass proportions are merely few percent points, which brings a degree of uncertainty to any possible trends observed here.

Table 4.4: The number of essential genes identified in all models by the two different methods of SGD analysis; genes essential for biomass production (BP), and genes essential for completing all 57 metabolic tasks (MT) necessary for human cell viability. Biomass production is one of the 57 MTs, thus the biomass proportion of the MT-essential genes was calculated.

Model	**BP-essential genes**	**MT-essential genes**	**Biomass proportion of MT-essential genes (α)**
BRCA NT	191	302	63,25 %
BRCA TM	192	305	62,95 %
BRCA TP	198	309	64,08 %
CESC NT	197	306	64,38 %
CESC TM	194	293	66,21 %
CESC TP	197	304	64,80 %
LUAD NT	199	306	65,03 %
LUAD TP	201	308	65,26 %
LUAD TR	194	292	66,44 %
PAAD NT	192	300	64,00 %
PAAD TM	195	302	64,57 %
PAAD TP	192	300	64,00 %
PRAD NT	190	295	64,41 %
PRAD TM	193	300	64,33 %
PRAD TP	183	288	63,54 %
		Average:	**64,48 %**

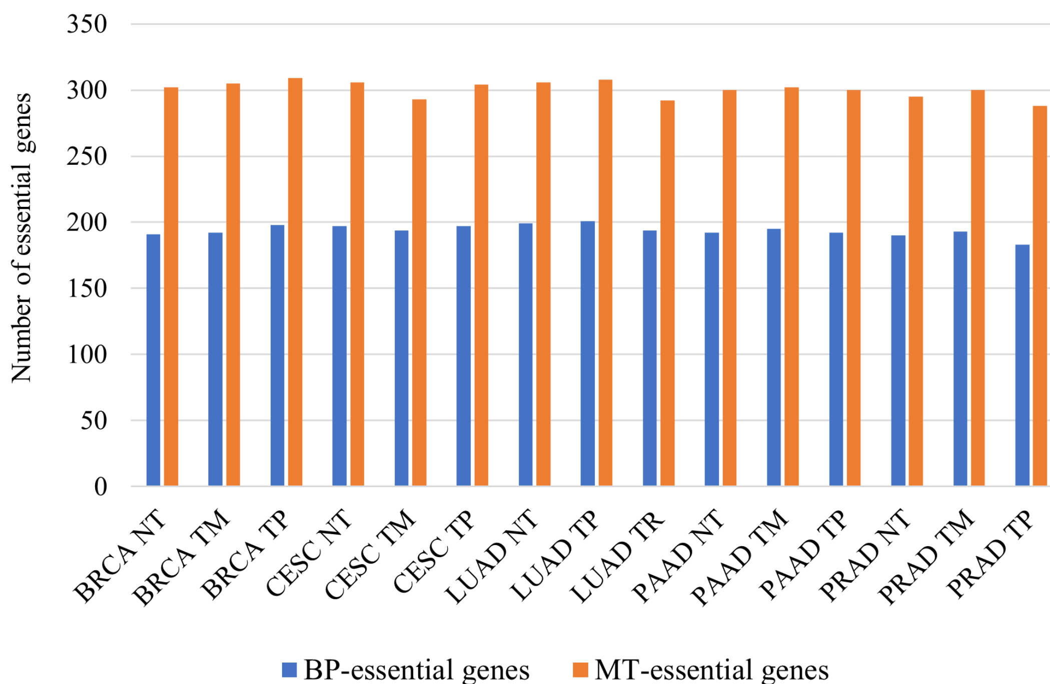


Figure 4.7: Number of essential genes identified with two different methods of gene knockout analysis, BP-essential (blue), and MT-essential genes (orange).

4.3 Synthetic Lethal Gene Pairs

Synthetic lethal (SL) gene pairs were identified by a double gene deletion analysis based on the 57 essential metabolic tasks as the metric for viability. In essence, two genes were deleted simultaneously, followed by the reduced model being tested for the ability to perform each of the 57 essential metabolic tasks for human cell viability. A gene pair that, when deleted, caused at least one of the metabolic tasks to fail were deemed an SL-pair. The analysis was only done on the BRCA model subgroup due to computational limitations.

Figure 4.8 shows the total number of SL-pairs discovered in the three BRCA models, with BRCA NT having 8808 SL-pairs, BRCA TP having 9618 SL-pairs, and BRCA TM having 7386 SL-pairs. One could hypothesise that tumour tissues, due to their genetic abnormalities and altered metabolism, would be more metabolically vulnerable to disturbances and, and as a result, might have more lethal pairs than their healthy tissue equivalent. This seems to be true for the BRCA primary tumor, but surprisingly is not the case for the BRCA metastatic tumor. However, with only these three data points, it is not possible to determine if this is a trend among different cancer types or if it is specific to BRCA.

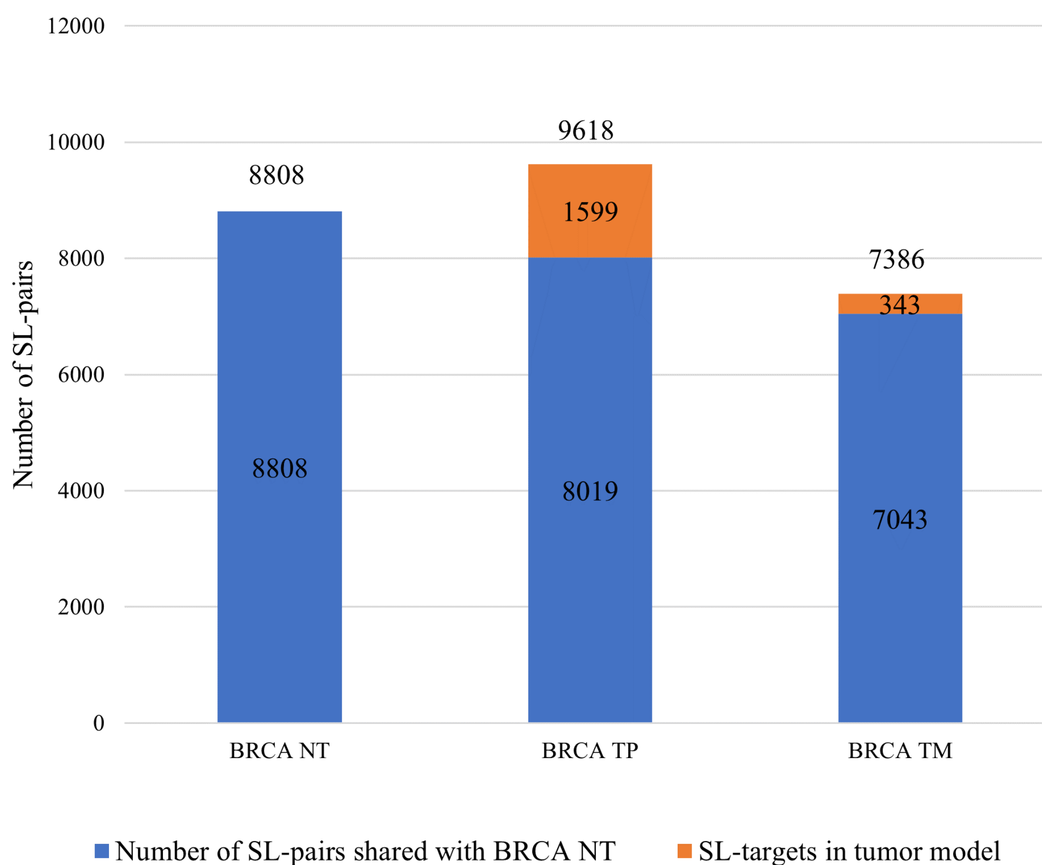


Figure 4.8: The number of SL gene pairs identified for each model in the double gene deletion analysis. SL-targets (orange) are SL-gene pairs that are only lethal for that tumor model, but not lethal for the normal tissue. Blue columns represent the number of SL gene pair shared with BRCA NT.

SL-targets for the tumor models were found by removing all SL-pairs that also appeared for the NT model. Meaning, inhibiting the genes in an SL-target will be fatal to the tumor, but not to the corresponding normal tissue. The SL-targets are illustrated as a network of epistatic interactions in Figure 4.10 and Figure 4.11, where a link (*edge*) between genes signifies them forming an SL-target. Common for both network representations is that a handful of genes are forming SL-pairs with many other genes. These can be described as having a high degree of connections, or having high *centrality* in the network. The genes ACOX3, ECH1, MCAT, MMAB, CPT2, and ABTB2 are high centrality genes in both tumour models, as well as GOT1 and CYP4V2 for BRCA TP. Several of these genes were found to be related to a high number of metabolic reactions in the BRCA models, as seen in Table 4.5.

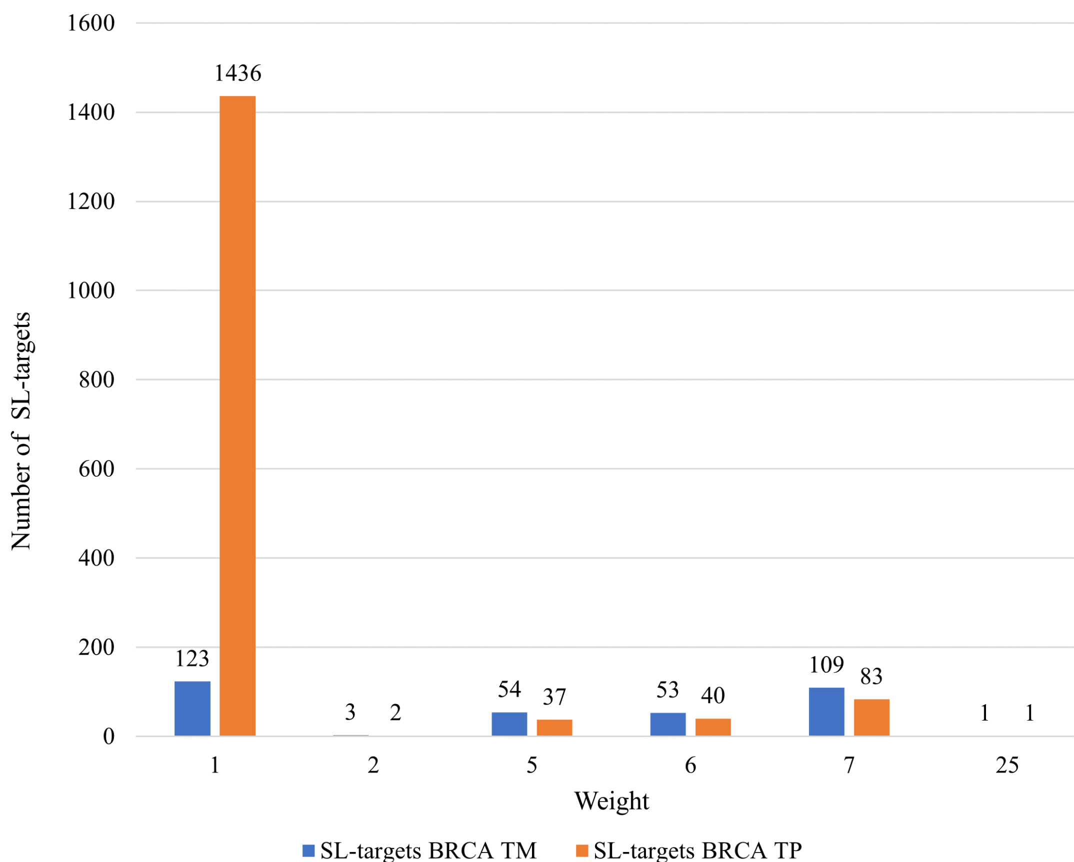


Figure 4.9: The number of synthetic lethal (SL) gene pairs with the various weights for BRCA TM (blue) and BRCA TP (orange).

Table 4.5: The most highly connected genes in the SL analysis. *Italicized* genes are found as an SL-target-forming gene only in BRCA TP, all other genes are found in both BRCA tumor types as SL-target-forming genes.

Gene	Number of related reactions
ABTB2	1
ACOX3	65
CPT2	64
CYP4V2	13
ECH1	7
GOT1	2
MCAT	1
MMAB	2

There was found to be 1599 and 343 SL-targets for BRCA TP and BRCA TM, respectively (Figure 4.8). The relatively high number of SL-targets for BRCA TP is explained by the fact that a single gene, GOT1, is a momentous centrailty point in the network, where it appears to be forming an SL-target with almost every other gene in the model (as seen in Figure 4.10). The complete list of SL-targets, as well

as high resolution versions of the network figures can be found in Supplementary data files **S5**, and **S8** and **S9**.

The weights of the SL-targets were calculated, as well as the frequency of the various weight classes, presented in Figure 4.9. It can be observed that a relatively high number of SL-targets displays a weight of 5, 6 or 7, similarly in both tumor types. However, the highest numbers of SL-targets falls into the weight category of 1, particularly for BRCA TP with 1436 such SL-targets. The single heaviest SL-target, with a weight of 25, was the same for both tumor types of BRCA; GAPDH and TPI1. Due to the uniquely heavy weight of this specific gene pair, it was chosen to investigate their metabolic role further. It was found that the gene product of both these genes are found in the cytosol, and are connected to glycolysis and gluconeogenesis, as seen in the metabolic map in Figure C.1 (Appendix C). They catalyzes an important energy-yielding step in carbohydrate metabolism, a metabolic process known to be altered in cancerous cells (as discussed in Section 2.4).

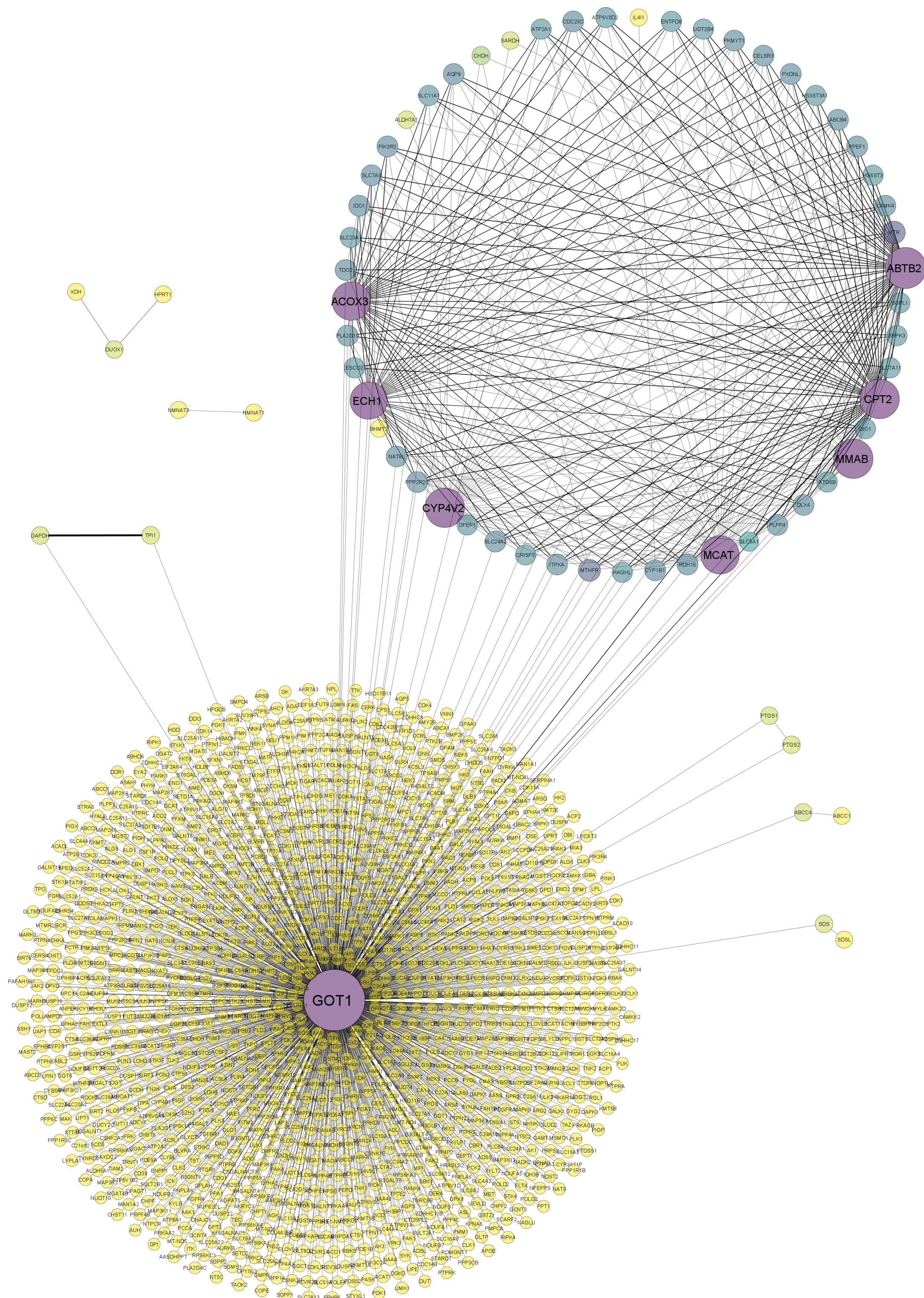


Figure 4.10: Network representation of the SL-targets in BRCA TP. Increased thickness of the lines (edges) represent increased weight of the genes (circles) connected as an SL-target. Increasing size of the nodes signifies increased number of connections. The colour gradient (yellow-blue-purple) of the circles also represent number of connections, with yellow signifying the lowest number of edges and purple representing most edges.

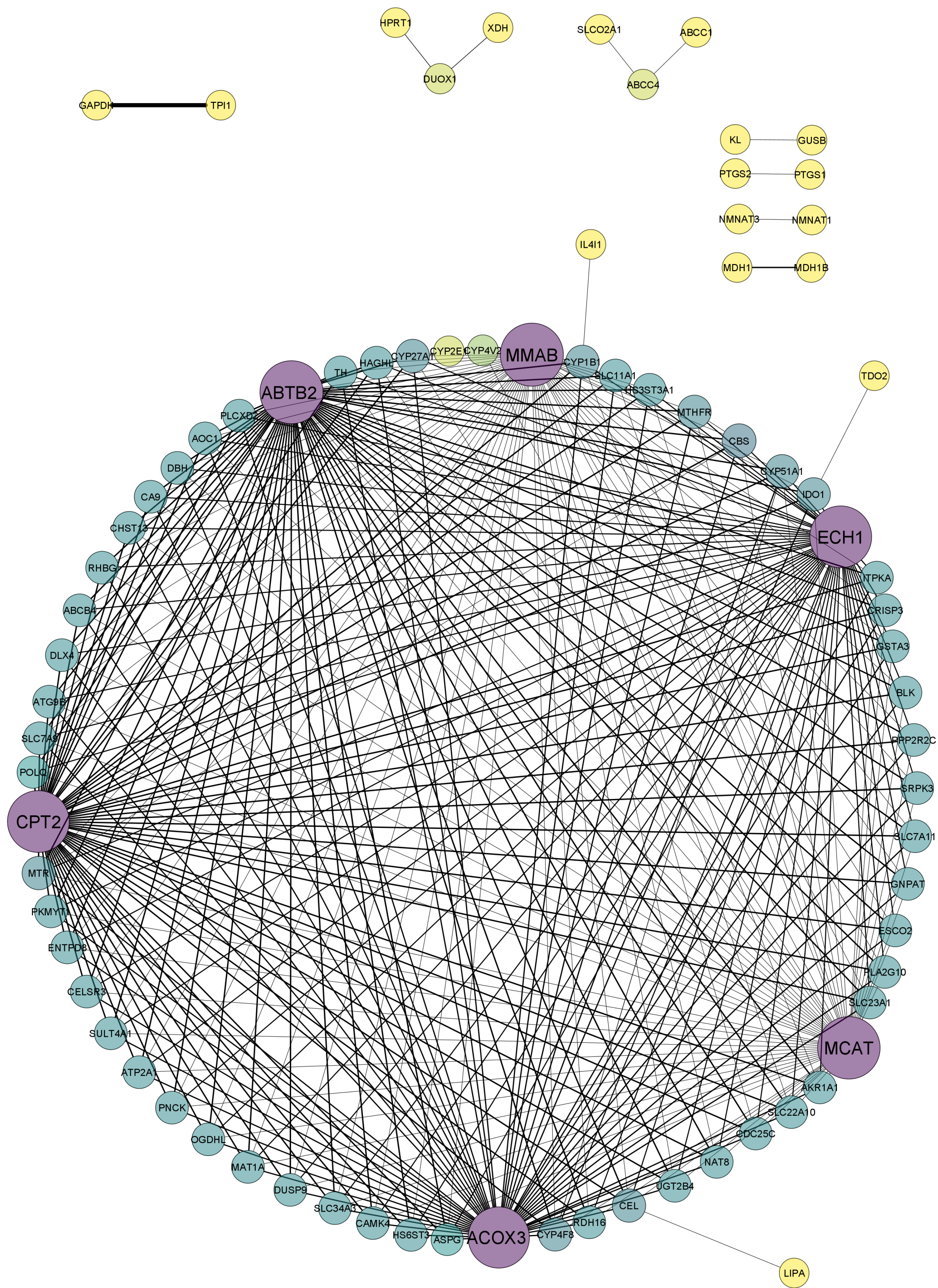


Figure 4.11: Network representation of the SL-targets in BRCA TM. Increased thickness of the lines (edges) represent increased weight of the genes (circles) connected as an SL-target. Increasing size of the nodes signifies increased number of connections. The colour gradient (yellow-blue-purple) of the circles also represent number of connections, with yellow signifying the lowest number of edges and purple representing most edges.

Chapter 5

Discussion

5.1 Potential drug targets were discovered

Although drug screens are relatively easy to perform, they can be both time and resource consuming [14]. This demand multiplies exponentially when looking for combinations of genes due to the shear number of potential combinations. There are therefore benefits to be had from conducting *in silico* gene deletion screenings, such as performed in this thesis, as to guide the allocations of resources. The over all aim of this study was to examine the genetic variations between cancer tissues and healthy tissues for a variety of frequently occurring cancers. The goal was to find genes that could be inhibited, with the effect of killing cancer cells while sparing healthy cells. The second aim was to investigate whether such gene targets were found in different cancer types, and if there was any differences between early stage and late stage cancers. GEMs of 15 different tissue types were therefore subjected to different gene KO simulations in order to explore their metabolic capabilities, including their similarities and their differences.

SGD analyses were performed for all 15 GEMs with two different methods, differentiated by their respective definition of essentiality. The first definition was the ability of BP above a certain threshold, and the second was the ability to carry flux through all 57 MTs necessary for human cell viability.

The SGD results from both methods displayed that all of the tumor types from the five different cancers investigated indeed had genes that were only lethal to the tumor, and not to the normal tissue. Such genes were named BP- or MT-target-genes, as an indication of these being possible drug targets. The same was seen in the SL analysis, where a number of SL-targets were found for both BRCA tumors.

Furthermore, the collective findings revealed that there were some differences between tumors in early (TP) and late stages (TR or TM) of the same cancer type, with some target-genes shared between the two tumours and others only identified in one of them. Although the cancer subgroups still shared the majority of the essential genes, if the intention is to target genes to kill a tumor, it would be a great advantage

to target those that are fatal to both tumor types.

The SL analysis performed on the BRCA models revealed that a handful of genes form SL-targets with numerous other genes, pointing to metabolic weaknesses that potentially could be exploited as cancer drug targets. Additionally, both BRCA tumour types shared most of these particular highly connected SL-target-genes, making a prospective SL-targeting method effective for both tumour types, which could be beneficial if there is uncertainty of the tumor type of the patient. It was also found that some target-genes across all analyses caused substantially higher degree of metabolic disruption to cancer cells (high weight) than other target-genes (low weight). Such 'heavy' target-genes or SL-targets, which represent a metabolic vulnerability present in the tumor, may be promising candidates for further research into new cancer treatments.

The two methods of performing SGD, using biomass production (Section 3.4.1) and metabolic tasks (Section 3.4.2), displayed some differences in the amount of essential genes identified, with the majority of the results of the former method also being found in the latter. This discrepancy is worth noticing, as the precision of the prediction of potential drug targets for cancer treatment relies on the definition of the objective function [14]. As opposed to using biomass production as the objective function, utilizing the 57 essential metabolic tasks for testing human cell viability have been shown to be more accurate when compared to genome wide CRISPR screenings [9]. For this reason, it was chosen to focus further investigation of the MT-target-genes compared to the BP-target-genes. Still, it can be argued that BP might be considered a reasonable objective function for a cancer cell, as they often display unregulated cellular growth similar to bacteria. Nonetheless, the discrepancy in results between these two methods of SGD analysis highlights the necessity of selecting an appropriate objective function for the specific type of cell being studied.

A general trend that was observed throughout the results was that many target-genes appeared in several tumor types, with the most connected target-genes being ALDH1A1 and PCYT1A, each with 7 tumor connections. Such "highly connected" target-genes would be interesting to investigate as a cancer drug target, as a potential medicine could be used for a number of different cancers and thus have a broader scope of application. Both of these genes have in fact been proposed as prospective cancer drug targets (see Table B.1, Appendix B). The fact that half of all the MT-target-genes identified in this thesis were connected to two or more tumor types, and seven of those were connected to three or more tumor types, suggests that tumors share certain metabolic weaknesses not found in their respective normal tissues. One could think that a target-gene connected to two tumor types would obviously be the two tumors of the same cancer, but the network representations (Figure 4.2 and Figure 4.4) shows that this is not the case. For instance, PRAD TM and CESC TM shares 7 MT-target-genes, a surprising discovery given that these tumours afflict different genders and thus would be anticipated to have somewhat different mechanisms. This provides an valuable insight into how different cancer types still can share vulnerabilities, although what these weak point are or which tumor types share them are not evident.

The differences observed between the two tumor stages, namely primary tumor

(early stage) and recurring or metastatic tumor (late stage), points to metabolic differences that could be worth exploring further. Studies has shown that a particular type of cell, called cancer stem cells (CSCs), are responsible for resistance to therapeutic agents, recurrence, relapse, metastasis of cancer [62] [61]. The available cancer treatments are believed to only be able to kill proliferative cancer cell, which CSCs manage to survive due to their ability of being dormant [78]. Therefore, it might be not only beneficial, but also completely necessary to target CSCs for the *complete* treatment of cancer [78]. A hypothesis could be proposed that the MT-target-genes that are essential for vital metabolic tasks other than growth could have the potential to kill all cancer cells, including CSCs, without the requirement of active growth. This speculation would assume that metabolic weaknesses, in the form of target-genes, are inherited from CSCs to the recurring and metastatic tumors, which would need to be investigated on a case by case basis. On the other side, it has been argued that differentiation of CSCs is required for complete cure of cancer [78], and thus perhaps a combination of differentiation therapy and inhibition of known target-genes for the relevant recurring or metastatic tumor could be an option as a treatment strategy.

This thesis suggests further research on the few genes that showed to create SL-pair with many other genes, namely ACOX3, ECH1, MCAT, MMAB, CPT2, and ABTB2 for both BRCA tumour models, as well as GOT1 and CYP4V2 for BRCA TP. The reason behind these genes being so highly connected in the SL network may lie in the fact that many of them are related to several metabolic reactions, which could explain their susceptibilities to being a SL gene (Table 4.5).

An additional interesting finding in the SL study was the single heaviest gene pair, GAPDH and TPI1, that, when deleted, inhibits 25 of the essential metabolic tasks in both BRCA tumors, but leaves the corresponding normal tissue seemingly unharmed. These genes are closely connected in the glycolysis and glyconeogenesis pathways (Figure C.1, Appendix C), which are pathways that often are observed to be altered in cancerous cells, such as through the Warburg effect [56][57]. TPI1 is classified as a 'Potential drug targets', and GAPDH is classified as 'FDA approved drug targets' by The Human Protein Atlas, thus strengthening the hypothesis that these two genes might be a valid synthetic lethal target to investigate for the treatment of BRCA.

5.2 MT-target-genes finds support in literature

It was anticipated to find more existing drugs that target the target-genes discovered in this thesis. Only one gene, CYP1B1, had existing drugs used in cancer treatment for various cancers in breast, ovaries, prostate, lung, and blood (Section 4.2.2). This gene was also found to be supported in literature as a cancer possible drug target for a number of cancers (Table B.1). This thesis argues that targeting this gene might be highly effective against primary tumor in the prostate, and in recurring lung cancer, owing to the severe metabolic impact its deletion inflicted. The latter tumor type, in particular, could be worth investigating further, since recurring lung cancer is notoriously difficult to treat [79]. The concept of using cancer drugs that

have already been approved for some tumor types on new tumor types is gaining traction in the medical community. This is the basics of the currently ongoing study IMPRESS-Norway, where the selection of anticancer drug is based on the molecular profiles of the patients, rather than the location of the cancer [80].

A significant weakness of this thesis is that the findings are only theoretical, as no experimental validation was performed. However, many of the target-genes are supported in literature as being proposed drug targets for a variety of cancers (Appendix B Table B.1). The literature review presented in Table B.1 revealed that at least 30 out of 46 MT-target-genes found in this thesis are linked to one or more cancers, and to being a prospective drug target. This strengthens the reliability of the findings of this thesis, as well as the credibility of using GEMs as a method for novel drug development.

5.3 Challenges

There have been some challenges throughout the project. The first was to create tissue specific models from a generic human GEM and transcriptomic or proteomic data. The lack of standardized gene expression units of experimental data sets from the GTEx database or the TCGA database made direct contextualization of the generic human GEM challenging. Specifically, some data sets were provided as low/medium/high/none protein or RNA expression, others used transcript per million (TPM), and others used fragments per kilo base of transcript per million mapped fragments (FPKM). In order to ensure that results from several different tissue models could be objectively compared, it was therefore decided to utilize "ready made" tissue specific models from another publication [9], which in addition assured a certain degree of model quality. Advantages that could be had with of making my own models, however, would have been that newer data sets could have been used, several different data sets, as well as data from different sources, could have been used and compared.

Another challenge was the long run time of some simulations, especially the double gene deletion script created for the synthetic lethal gene pair analysis. For roughly 2300 genes left in the models (after the known essential genes were removed), each gene was paired with every other gene, and simulated once for every 57 metabolic task. By the design of the script, this created (2300*2300*57) over 3 million cases to be solved per model. In other words, improving the methodology of performing a double gene knockout using the 57 metabolic tasks as the viability indicator could considerably reduce the time necessary to run such a study.

It was discovered that one of the RAVEN functions used in the script for the SL analysis, `checkTasksGenes`, could override the previously and manually set gene deletion constraints when setting constraints for the metabolic task being tested for a small subset of genes. However, owing to the nature of the script, all gene pairs were tested twice, first as GeneA+GeneB and then as GeneB+GeneA. Therefore, through thorough troubleshooting and manually investigating every part of the functions used, it was discovered that for the small proportion of genes affected by this script

issue, at least one of the combinations yielded results that were regarded acceptable. Nonetheless, this flaw is a potential source of error.

5.4 Further work

There are many ways of continuing the work of this thesis, and improving the methodology. The most obvious area of improvement would be to redesign the double gene KO script to better fit the use of the 57 metabolic tasks as a metric for cellular viability. This includes, but is not limited to, developing a method in which the constraints imposed while evaluating a metabolic task do not conflict with the gene deletion constraints.

Further, it could be interesting to evaluate the models' capabilities on different media compositions, as this has been shown to affect what genes are considered essential [14], as well as broadening the scope to include more tissue and tumor types for all analyses. Particularly, the SL gene pair analysis (Section 3.5) would benefit from including more tissue models, as this would allow to study potential emerging patterns of the SL phenomenon in cancer tissue.

Maybe the most important aspect of furthering the work of this thesis would be to conduct experimental validation of the data. This could be done by performing a genome-wide gene KO screen on all of the tissue types studied in this thesis, for example using CRISPR technology. The main reason behind this need is that our understanding of human metabolism is far from complete, and consequently, so are the GEMs built to simulate it. Human cells live in complex environments, and they are constantly affected by a myriad of factors such as cross-communications between cells, signalling and regulatory mechanisms that ultimately influence their metabolism [23]. GEMs are currently unable to incorporate all such cellular events and how it affects metabolism, and although studies has shown that human GEMs have proved high accuracy compared to experimental data (up to 88% in [9]), the knowledge gap emphasizes that the findings are merely theoretical until they can be proven experimentally. Nevertheless, GEM simulations such as in this thesis can provide pointers as to where it might be beneficial to allocate research.

Another area worth investigating stems from the fact that target-genes were defined as "a gene that is lethal when deleted for the tumor in question, but not lethal for the *corresponding* normal tissue", meaning it was not investigated whether other types of normal tissue would be impaired by the deletion of the target-gene. For a prospective medical use, this would imply that the target-gene must be only inhibited locally in the target tissue, or it must be deemed safe from hurting any other normal tissue. Therefore, more research would be required to evaluate the precise impacts on various normal tissues when deleting a tumor-specific target-gene to ensure that it does not harm healthy cells or significantly diminish their functioning. Such screening could be done using GEMs, but inevitably, the results must be verified experimentally.

Lastly, further network analyses of the data could yield valuable insights, and the

possibilities are many. For instance, one can create a network of the subsystems affected by the deletion of target-genes to investigate whether some subsystems are more frail than others, or to explore what subsystems that are simultaneously affected in SL-targets. Knowing from literature that many cancers share common metabolic alterations, for example in relation to energy provision, it would be reasonable to anticipate the discovery of some patterns shared amongst different cancers.

Chapter 6

Conclusion and Outlook

The aim of this project was to investigate whether there were genetic variations between cancerous tissue and healthy tissue that could be exploited for potential cancer treatment. GEMs related to five different cancer types, namely breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, lung adenocarcinoma, pancreatic adenocarcinoma, and prostate adenocarcinoma, were utilised to perform *in silico* simulations of different gene KO scenarios. Single and double gene deletions were performed, where the focus was on discovering genes that were essential only to tumor tissue, but not to the corresponding normal tissue, deeming such genes *target-genes*. In theory, such genes could be inhibited, leading to the death of the cancer cell while sparing healthy cells.

The first analysis predicted in total 46 unique target-genes that were essential for maintaining all the necessary metabolic tasks a human cell needs to survive, 30 of which have been proposed as potential drug targets in literature, and the rest were, to the authors knowledge, new. One gene (CYP1B1) that was identified to be essential to recurring lung cancer and primary tumor of the prostate, was also found to be a drug target in current cancer therapies for a variety of cancers. The synthetic lethal gene pair analysis performed on the breast cancer models revealed a handful of genes that form synthetic lethal pair with numerous other genes, revealing metabolic alterations not present in the corresponding normal tissue. As such, this thesis proposes that the target-genes, especially those connected to several tumor types and/or with the greatest adverse metabolic impact, to be further studied as potential drug targets for cancer treatment.

Although some studies have demonstrated that cancers are more similar to their tissue of origin than cancers in other organs [9] [10], this thesis argues that they nonetheless share several metabolic weaknesses that normal tissue does not have. It also provided evidence of metabolic differences between early and late stages of the same cancers. Both of these insights might be invaluable when a physician is deciding on the best treatment strategy for their patient.

Since we don't know everything about the human metabolism, GEMs built on this knowledge will not be flawless. In addition, GEMs do not incorporate signaling and

regulatory mechanisms and the likes, and thus simulations done using GEMs will not completely reflect reality. Therefore, maybe the biggest weakness of this thesis is that the findings are purely theoretical and thus must be verified experimentally. However, novel approaches and innovative methods, such as using the 57 essential metabolic tasks instead of biomass production as the objective function contributes to bridge the gap between *in silico* and *in vivo* results.

The goal of systems biology can be said to be able to explain all emerging biological functions through the understanding and integration of all parts. To reach such a goal will only be possible through an iterative process of creating models, performing simulations, testing the results experimentally, discover unknown discrepancies that guides new research, improve the models to fit new knowledge, and repeat. Thus, this thesis is a humble contribution towards this process.

The application of systems biology to gain novel insight into human health will undoubtedly benefit many fields of biological research, especially in the medical sciences, which ultimately provides opportunities for better health care. This thesis demonstrated how GEMs may be used to predict novel drug targets for cancer therapy based on metabolic variations between cancerous tissue and healthy tissue. Furthermore, it supports the notion that medications used to treat one kind of cancer may, in theory, also have the potential to treat other types of cancer. Such knowledge enables the use of current treatments in novel ways, bypassing the often costly and resource-intensive process of developing new drugs.

GEMs built using patient-specific omics data, such as in this thesis, also opens up the possibility of using GEMs in clinical medicine. A personalized and patient-specific model could be interrogated to predict the outcome of various drugs and therapies to ultimately guide the choice of treatment based on the metabolic signature of the patient. Many challenges remains to be solved before this becomes mainstream, such as establishing the range of what a healthy metabolic signature looks like, improving model performance, and more. Nonetheless, the continued development of human GEMs will undoubtedly benefit several areas of biological research, with far reaching consequences on both a general and personalized scale.

Further study is clearly needed to experimentally verify the results of this thesis and to understand the mechanisms behind their role in the metabolism of healthy and diseased cells. Nonetheless, the findings may provide a sense of direction to guide future cancer research, and contribute to a deeper insight into cancer metabolism.

Bibliography

1. De Domenico, M. & Sayama, H. *Complexity Explained (2019)* https://complexityexplained.github.io/. (Accessed: 12.05.2022).
2. Barabási, A.-L. *Network Science* http://networksciencebook.com/chapter/1. (Accessed: 02.05.2022).
3. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9 (Jan. 2000).
4. Lewis, N. E. & Abdel-Haleem, A. M. The evolution of genome-scale models of cancer metabolism. *Frontiers in Physiology* **4.** https://doi.org/10.3389/fphys.2013.00237 (2013).
5. Dulbecco, R. A Turning Point in Cancer Research: Sequencing the Human Genome. *Science* **231,** 1055–1056. https://www.science.org/doi/abs/10.1126/science.3945817 (1986).
6. Palsson, B. O. *Systems Biology: Constraint-based Reconstruction and Analysis* (Sheridan Books, Inc., 2015).
7. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using Genome-scale Models to Predict Biological Capabilities. *Cell* **161,** 971–987. https://doi.org/10.1016/j.cell.2015.05.019 (May 2015).
8. Larsson, I. *et al.* Genome-Scale Metabolic Modeling of Glioblastoma Reveals Promising Targets for Drug Development. *Frontiers in Genetics* **11.** https://doi.org/10.3389/fgene.2020.00381 (Apr. 2020).
9. Robinson, J. L. *et al.* An atlas of human metabolism. *Science Signaling* **13.** https://doi.org/10.1126/scisignal.aaz1482 (Mar. 2020).
10. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nature Biotechnology* **31,** 522–529. https://doi.org/10.1038/nbt.2530 (June 2013).
11. Jalili, M. *et al.* Exploring the Metabolic Heterogeneity of Cancers: A Benchmark Study of Context-Specific Models. *Journal of Personalized Medicine* **11,** 496. https://doi.org/10.3390/jpm11060496 (June 2021).
12. Fu, H. Y. *et al.* Protein Quality Control Dysfunction in Cardiovascular Complications Induced by Anti-Cancer Drugs. *Cardiovascular Drugs and Therapy* **31,** 109–117. https://doi.org/10.1007/s10557-016-6709-7 (Feb. 2017).

13. Paul, A. *et al.* Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. *Scientific Reports* **11.** https://doi.org/10.1038/s41598-020-80561-1 (Jan. 2021).

14. García, M. M. *et al.* Importance of the biomass formulation for cancer metabolic modeling and drug prediction. *iScience* **24,** 103110. https://doi.org/10.1016/j.isci.2021.103110 (Oct. 2021).

15. Paul, A. *et al.* Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. *Scientific Reports* **11,** 213. https://doi.org/10.1038/s41598-020-80561-1 (Jan. 2021).

16. Singh, S. *et al.* Flux-based classification of reactions reveals a functional bow-tie organization of complex metabolic networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* **87.** https://doi.org/10.1103/PhysRevE.87.052708 (May 2013).

17. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* **28,** 245–248. https://doi.org/10.1038/nbt.1614 (Mar. 2010).

18. Sandtrøen, R. M. G. *The status of genome-scale metabolic modeling of human cells* Project thesis written as a part of the masters level course "TBT4500 - Biotechnology, Specialization Project". Dec. 2021.

19. Feist, A. M. & Palsson, B. O. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. eng. *Nature biotechnology* **26,** 659–667. https://doi.org/10.1038/nbt1401 (June 2008).

20. Edwards, J. S., Covert, M. & Palsson, B. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology* **4,** 133–140. https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1462-2920.2002.00282.x (2002).

21. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology* **19,** 125–130. https://doi.org/10.1038/84379 (Feb. 2001).

22. Kumar, V. S. & Maranas, C. D. GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. *PLoS Computational Biology* **5** (ed Ouzounis, C. A.) https://doi.org/10.1371/journal.pcbi.1000308 (Mar. 2009).

23. Sethi, J. K. & Vidal-Puig, A. Wnt signalling and the control of cellular metabolism. *Biochem J* **427,** 1–17 (Mar. 2010).

24. Ranganathan, S., Suthers, P. F. & Maranas, C. D. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. eng. *PLoS computational biology* **6.** https://doi.org/10.1371/journal.pcbi.1000744 (Apr. 2010).

25. Papp, B., Pãl, C. & Hurst, L. D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429,** 661–664 (June 2004).

26. Harrison, R. *et al.* Plasticity of genetic interactions in metabolic networks of yeast. eng. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 2307–2312. https://doi.org/10.1073/pnas.0607153104 (Feb. 2007).
27. Thiele, I. & Palsson, B. O. A protocol for generating a high-quality genome-scale metabolic reconstruction. eng. *Nature protocols* **5,** 93–121. https://doi.org/10.1038/nprot.2009.203 (Jan. 2010).
28. Opdam, S. *et al.* A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Systems* **4,** 318–329. https://www.sciencedirect.com/science/article/pii/S2405471217300108 (Mar. 2017).
29. Norsigian, C. J. *et al.* A workflow for generating multi-strain genome-scale metabolic models of prokaryotes. *Nature Protocols* **15,** 1–14. https://doi.org/10.1038/s41596-019-0254-3 (Jan. 2020).
30. Norsigian, C. J. *et al.* BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research* **48.** https://doi.org/10.1093/nar/gkz1054 (Jan. 2020).
31. Karlsen, E., Schulz, C. & Almaas, E. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics* **19,** 467. https://doi.org/10.1186/s12859-018-2472-z (Dec. 2018).
32. Henry, C. S. *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* **28,** 977–982. https://doi.org/10.1038/nbt.1672 (Sept. 2010).
33. Edwards, J. S. & Palsson, B. O. Systems Properties of the Haemophilus influenzaeRd Metabolic Genotype. *Journal of Biological Chemistry* **274.** https://doi.org/10.1074/jbc.274.25.17410 (June 1999).
34. Duarte, N. C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences* **104.** https://doi.org/10.1073/pnas.0610772104 (Feb. 2007).
35. Lewis, N. E. & Abdel-Haleem, A. M. The evolution of genome-scale models of cancer metabolism. *Frontiers in Physiology* **4.** https://doi.org/10.3389/fphys.2013.00237 (2013).
36. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nature Biotechnology* **31,** 419–425. https://doi.org/10.1038/nbt.2488 (May 2013).
37. Swainston, N. *et al.* Recon 2.2: from reconstruction to model of human metabolism. eng. *Metabolomics : Official journal of the Metabolomic Society* **12,** 109–109. https://doi.org/10.1007/s11306-016-1051-4 (2016).
38. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology* **36,** 272–281 (Mar. 2018).
39. Mardinoglu, A. *et al.* Integration of clinical data with a genome-scale metabolic model of the human adipocyte. eng. *Molecular systems biology* **9,** 649–649. https://doi.org/10.1038/msb.2013.5 (2013).

40. Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications* **5.** https://doi.org/10.1038/ncomms4083 (Jan. 2014).

41. Blais, E. M. *et al.* Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nature Communications* **8.** https://doi.org/10.1038/ncomms14250 (Feb. 2017).

42. DeBerardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? eng. *Cell* **148,** 1132–1144. https://doi.org/10.1016/j.cell.2012.02.032 (Mar. 2012).

43. Ghesquiãre, B. *et al.* Metabolism of stromal and immune cells in health and disease. *Nature* **511,** 167–176. https://doi.org/10.1038/nature13312 (July 2014).

44. Mardinoglu, A. & Nielsen, J. New paradigms for metabolic modeling of human cells. *Current Opinion in Biotechnology* **34,** 91–97. https://www.sciencedirect.com/science/article/pii/S0958166914002286 (Aug. 2015).

45. Argmann, C. A. *et al.* A Next Generation Multiscale View of Inborn Errors of Metabolism. *Cell Metabolism* **23,** 13–26. https://doi.org/10.1016/j.cmet.2015.11.012 (Jan. 2016).

46. Nielsen, J. & Ji, B. New insight into the gut microbiome through metagenomics. *Advances in Genomics and Genetics,* 77. https://doi.org/10.2147/agg.s57215 (Jan. 2015).

47. Heinken, A. & Thiele, I. Systems biology of host–microbe metabolomics. *WIREs Systems Biology and Medicine* **7,** 195–219. https://doi.org/10.1002/wsbm.1301 (Apr. 2015).

48. Gatto, F., Ferreira, R. & Nielsen, J. Pan-cancer analysis of the metabolic reaction network. *Metabolic Engineering* **57,** 51–62. https://doi.org/10.1016/j.ymben.2019.09.006 (Jan. 2020).

49. Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications* **5.** https://doi.org/10.1038/ncomms4083 (Jan. 2014).

50. Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Molecular Systems Biology* **10,** 721. https://doi.org/10.1002/msb.145122 (Mar. 2014).

51. Burgard, A. P. & Maranas, C. D. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and Bioengineering* **82,** 670–677. https://doi.org/10.1002/bit.10617 (June 2003).

52. Fouladiha, H. & Marashi, S.-A. Biomedical applications of cell- and tissue-specific metabolic network models. *Journal of Biomedical Informatics* **68,** 35–49. https://www.sciencedirect.com/science/article/pii/S1532046417300394 (Apr. 2017).

53. Angione, C. Human systems biology and metabolic modelling: A review-from disease metabolism to precision medicine. *BioMed Research International* **2019.** https://doi.org/10.1155/2019/8304260 (2019).

54. Ferlay, J. *et al. Cancer Today.* Lyon: International Agency for Research on Cancer; 2020. https://gco.iarc.fr/today. (Accessed: 03.08.2022).

55. Hammoudi, N. *et al.* Metabolic alterations in cancer cells and therapeutic implications. *Chinese Journal of Cancer* **30,** 508–525. https://doi.org/10.5732/cjc.011.10267 (Aug. 2011).

56. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *Journal of General Physiology* **8,** 519–530. https://doi.org/10.1085/jgp.8.6.519 (Mar. 1927).

57. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends in Biochemical Sciences* **41,** 211–218. https://doi.org/10.1016/j.tibs.2015.12.001 (Mar. 2016).

58. Gaude, E. & Frezza, C. Defects in mitochondrial metabolism and cancer. *Cancer & Metabolism* **2.** https://doi.org/10.1186/2049-3002-2-10 (July 2014).

59. Cancer Research UK. *Stages of cancer* https://www.cancerresearchuk.org/about-cancer/what-is-cancer/stages-of-cancer. (Accessed: 05.08.2022).

60. National Cancer Institute at the National Institutes of Health. *primary tumor* https://www.cancer.gov/publications/dictionaries/cancer-terms/def/primary-tumor. (Accessed: 05.08.2022).

61. Chang, J. C. Cancer stem cells: Role in tumor growth, recurrence, metastasis, and treatment resistance. *Medicine* **95.** https://doi.org/10.1097/md.0000000000004766 (Sept. 2016).

62. Chen, K., Huang, Y.-h. & Chen, J.-l. Understanding and targeting cancer stem cells: therapeutic implications and challenges. *Acta Pharmacologica Sinica* **34,** 732–740. https://doi.org/10.1038/aps.2013.27 (May 2013).

63. Kaelin, W. G. The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nature Reviews Cancer* **5,** 689–698. https://doi.org/10.1038/nrc1691 (Aug. 2005).

64. Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology* **7,** 501. https://doi.org/10.1038/msb.2011.35 (Jan. 2011).

65. Larsson, I. *et al.* Genome-Scale Metabolic Modeling of Glioblastoma Reveals Promising Targets for Drug Development. *Frontiers in Genetics* **11,** 381. https://doi.org/10.3389/fgene.2020.00381 (2020).

66. Topatana, W. *et al.* Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *Journal of Hematology & Oncology* **13.** https://doi.org/10.1186/s13045-020-00956-5 (Sept. 2020).

67. MATLAB. *version 9.10.0.1602886 (R2021a)* (The MathWorks Inc., Natick, Massachusetts, 2021).

68. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols* **14,** 639–702. https://doi.org/10.1038/s41596-018-0098-2 (Mar. 2019).

69. Kerkhoven, E. *et al. SysBioChalmers/RAVEN: v2.5.3* version v2.5.3. Aug. 2021. https://doi.org/10.5281/zenodo.5275836.

70. Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual* 2022. https://www.gurobi.com.

71. Microsoft Corporation. *Microsoft Excel* version 2205. 1st July 2022. https://office.microsoft.com/excel.

72. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13** (2003).

73. Assenov, Y. *et al.* Computing topological parameters of biological networks. *Bioinformatics* **24,** 282–284 (Jan. 2008).

74. Robinson, J. L. *et al. Supplementary data and scripts for "An Atlas of Human Metabolism"* (Zenodo, Mar. 2019). https://doi.org/10.5281/zenodo.3583004.

75. Eduard Kerkhoven and Simonas Marcišauskas and Rasmus Ågren. *parseTaskList.m Function Description* https://github.com/SysBioChalmers/RAVEN/blob/main/core/parseTaskList.m. (Accessed: 08.08.2022).

76. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Science advances* **7.** https://doi.org/10.1126/sciadv.abh2169 (July 2021).

77. Paul, A. *et al.* Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. *Scientific Reports* **11.** https://doi.org/10.1038/s41598-020-80561-1 (Jan. 2021).

78. Prasad, S. *et al.* Cancer cells stemness: A doorstep to targeted therapy. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1866.** https://doi.org/10.1016/j.bbadis.2019.02.019 (Apr. 2020).

79. Kubouchi, Y. *et al.* Prognostic Factors for Post Recurrence Survival in Resected Pathological Stage I Non-small Cell Lung Cancer. *Yonago Acta Med* **60,** 213–219 (Dec. 2017).

80. Helland, Å. *et al.* Improving public cancer care by implementing precision medicine in Norway: IMPRESS-Norway. *Journal of Translational Medicine* **20,** 225. https://doi.org/10.1186/s12967-022-03432-5 (May 2022).

81. Tsirulnikov, K. *et al.* Aminoacylase 3 Is a New Potential Marker and Therapeutic Target in Hepatocellular Carcinoma. *J Cancer* **9,** 1–12 (2018).

82. Muralikrishnan, V., Hurley, T. D. & Nephew, K. P. Targeting Aldehyde Dehydrogenases to Eliminate Cancer Stem Cells in Gynecologic Malignancies. *Cancers (Basel)* **12** (Apr. 2020).

83. Tian, Q. *et al.* Phosphorylation of BCKDK of BCAA catabolism at Y246 by Src promotes metastasis of colorectal cancer. *Oncogene* **39** (May 2020).

84. Chen, X. *et al.* Single Cell Gene Co-Expression Network Reveals FECH/CROT Signature as a Prognostic Marker. *Cells* **8** (July 2019).

85. Wu, Y. & Xu, Y. Bioinformatics for The Prognostic Value and Function of Cubilin (CUBN) in Colorectal Cancer. *Med Sci Monit* **26** (Oct. 2020).

86. Gremel, G. *et al.* A systematic search strategy identifies cubilin as independent prognostic marker for renal cell carcinoma. *BMC Cancer* **17.** https://doi.org/10.1186/s12885-016-3030-6 (Jan. 2017).

87. Carrera, A. N., Grant, M. K. O. & Zordoky, B. N. CYP1B1 as a therapeutic target in cardio-oncology. *Clin Sci (Lond)* **134** (Nov. 2020).

88. Karlsson, S. *et al.* Inhibition of CYP27B1 and CYP24 Increases the Antiproliferative Effects of 25-Hydroxyvitamin D3 in LNCaP Cells. *Anticancer research* **41.** https://doi.org/10.21873/anticanres.15288 (Oct. 2021).
89. Vainio, P. *et al.* Arachidonic acid pathway members PLA2G7, HPGD, EPHX2, and CYP4F8 identified as putative novel therapeutic targets in prostate cancer. *Am J Pathol* **178,** 525–536 (Feb. 2011).
90. Hargrove, T. Y. *et al.* Human sterol 14 alpha-demethylase as a target for anticancer chemotherapy: towards structure-aided drug design. *J Lipid Res* **57,** 1552–1563 (Aug. 2016).
91. Thornburg, J. M. *et al.* Targeting aspartate aminotransferase in breast cancer. *Breast Cancer Res* **10** (2008).
92. Yang, S. *et al.* Mitochondrial glutamine metabolism via GOT2 supports pancreatic cancer growth through senescence inhibition. *Cell Death Dis* **9,** 55 (Jan. 2018).
93. Grobben, Y. *et al.* Targeting Indoleamine 2,3-Dioxygenase in Cancer Models Using the Novel Small Molecule Inhibitor NTRC 3883-0. *Front Immunol* **11** (2020).
94. Huang, T. T. *et al.* Kynurenine 3-monooxygenase upregulates pluripotent genes through beta-catenin and promotes triple-negative breast cancer progression. *EBioMedicine* **54** (Apr. 2020).
95. Mor, A., Tankiewicz-Kwedlo, A. & Pawlak, D. Kynurenines as a Novel Target for the Treatment of Malignancies. *Pharmaceuticals (Basel)* **14** (June 2021).
96. Al-Mansoob, M. *et al.* KYNU, a novel potential target that underpins CD44-promoted breast tumour cell invasion. *Journal of Cellular and Molecular Medicine* **25,** 2309–2314 (Mar. 2021).
97. Liu, X. *et al.* Targeting LIPA independent of its lipase activity is a therapeutic strategy in solid tumors via induction of endoplasmic reticulum stress. *Nature cancer.* https://doi.org/10.1038/s43018-022-00389-8 (June 2022).
98. Kim, Y. I. Role of the MTHFR polymorphisms in cancer risk modification and treatment. *Future Oncol* **5,** 523–542 (May 2009).
99. Stankova, J., Lawrance, A. K. & Rozen, R. Methylenetetrahydrofolate reductase (MTHFR): a novel target for cancer therapy. *Curr Pharm Des* **14,** 1143–1150 (2008).
100. Sullivan, M. R. *et al.* Methionine synthase is essential for cancer cell proliferation in physiological folate environments. *Nature metabolism* **3.** https://doi.org/10.1038/s42255-021-00486-5 (Nov. 2021).
101. Zand, B. *et al.* Role of Increased n-acetylaspartate Levels in Cancer. *J Natl Cancer Inst* **108** (Jan. 2016).
102. Yu, J. *et al.* PCYT1A suppresses proliferation and migration via inhibiting mTORC1 pathway in lung adenocarcinoma. *Biochem Biophys Res Commun* **529,** 353–361 (Aug. 2020).
103. Liu, W. & Phang, J. M. Proline dehydrogenase (oxidase) in cancer. *Biofactors* **38,** 398–406 (2012).

104. Youn, H. S. *et al.* Structural Insights into the Quaternary Catalytic Mechanism of Hexameric Human Quinolinate Phosphoribosyltransferase, a Key Enzyme in de novo NAD Biosynthesis. *Sci Rep* **6** (Jan. 2016).

105. Liu, X. *et al.* Cellular retinol binding protein-1 inhibits cancer stemness via upregulating WIF1 to suppress Wnt beta-catenin pathway in hepatocellular carcinoma. *BMC Cancer* **21.** https://doi.org/10.1186/s12885-021-08967-2 (Nov. 2021).

106. Serra, M. & Saba, J. D. Sphingosine 1-phosphate lyase, a key regulator of sphingosine 1-phosphate signaling and function. *Adv Enzyme Regul* **50,** 349–362 (2010).

107. Chen, Y. *et al.* Impacts of the SOAT1 genetic variants and protein expression on HBV-related hepatocellular carcinoma. *BMC Cancer* **21,** 615 (May 2021).

108. Li, H. *et al.* Inhibiting breast cancer by targeting the thromboxane A 2 pathway. *NPJ precision oncology* **1,** 8. https://doi.org/10.1038/s41698-017-0011-4 (2017).

109. Allain, E. P. *et al.* Emerging roles for UDP-glucuronosyltransferases in drug resistance and cancer progression. *Br J Cancer* **122** (Apr. 2020).

110. Kyoto University Bioinformatics Center. *KEGG Pathway: Glycolysis / Gluconeogenesis - Homo sapiens (human)* https://www.genome.jp/pathway/hsa00010. (Accessed: 10.06.2022).

Appendices

Appendix A

Description of the Supplementary data files

S1: The 57 metabolic tasks which must occur in all human cell types for it to be considered as viable, constructed by Agren *et al.* [50].

S2: The composition of Ham's media, extracted from **S1**.

S3: Raw data from SGD analyses presenting all essential genes found for all models, both BP-essential and MT-essential.

S4: Processed data from the SGD analyses, presenting the MT- and BP-target-genes for all models.

S5: Processed data from the SL analysis, presenting the SL-target-genes for BRCA TP and BRCA TM.

S6: High resolution image of the network representation of the BP-target-genes.

S7: High resolution image of the network representation of the MT-target-genes.

S8: High resolution image of the network representation of the epistatic gene interactions forming SL-targets for BRCA TP.

S9: High resolution image of the network representation of the epistatic gene interactions forming SL-targets for BRCA TM.

Appendix B

MT-target-genes supported in literature

Table B.1: Identified MT-target genes proposed as a drug targets for various cancer types in literature.

MT-target gene	Identified in model	Proposed drug target for cancer types	Source
ACY3	BRCA TM, CESC TM, LUAD TR, PRAD TM	Hepatocellular Carcinoma	[81]
ALDH1A1, ALDH1L2, ALDH4A1	BRCA TM, BRCA TP, CESC TM, CESC TP, LUAD TR, PAAD TM, PRAD TM	Gynecologic and other malignancies	[82]
BCKDKA, BCKDHB	BRCA TP, BRCA TM, CESC TP, CESC TM, BRCA TP, BRCA TM, CESC TP, CESC TM	Colorectal cancer, metastatic lung cancer	[83]
CROT	CESC TM, PRAD TP	Prostate cancer	[84]
CUBN	PAAD TP	Colorectal Cancer, Renal cell carcinoma (proposed as diagnostic marker)	[85], [86]

CYP1B1	LUAD TR, PRAD TP	Breast, ovarian, lung, liver, colorectal, gastric, and prostate cancers, leukemia, glioma	[87]
CYP27B1	PAAD TP	Prostate cancer	[88]
CYP4F8	PRAD TM	Prostate cancer	[89]
CYP51(A1)	PRAD TM	Lung, breast, skin cancer	[90]
GOT1	CESC TM, LUAD TR	Breast cancer	[91]
GOT2	LUAD TR	Breast cancer, pancreatic cancer	[91], [92]
IDO1	PAAD TM	Ovarian Cancer	[93]
KMO	CESC TM, PRAD TM	Breast cancer. Management of numerous neoplasms	[94], [95]
KYNU	CESC TM, PRAD TM	Breast cancer	[96]
LIPA	BRCA TP	Breast cancer	[97]
MTHFR	LUAD TP, PAAD TM	Increasing chemosensitvity of cancer cells. Decreased growth of tumors	[98], [99]
MTR	PAAD TM, LUAD TP	Slows cancer cell proliferation	[100]
NAT8L	CESC TM, LUAD TR	Ovarian cancer, melanoma, renal cell, breast, colon, and uterine cancers	[101]
PCYT1A	BRCA TP, BRCA TM, CESC TP, LUAD TP, PAAD TP, PRAD TP, PRAD TM	Lung cancer	[102]
PRODH	BRCA TP, CESC TP	General role in tumorigenesis and tumor development	[103]
QPRT	CESC TM, PRAD TM	Malignant glioma (cancer of the brain and spinal cord)	[104]
RBP1	LUAD TR, LUAD TP	Hepatocellular Carcinoma	[105]
SGPL1	PRAD TP	Modulate malignant growth, including cancer	[106]
SOAT1	BRCA TP, BRCA TM,	Hepatocellular carcinoma	[107]
TBXAS1	BRCA TP, BRCA TM,	Breast cancer	[108]

UGT2B10, UGT2B15	BRCA TM, CESC TP, LUAD TP, BRCA TP	Melanoma and hormone-related cancers	[109]

Appendix C

Glycolysis and gluconeogenesis

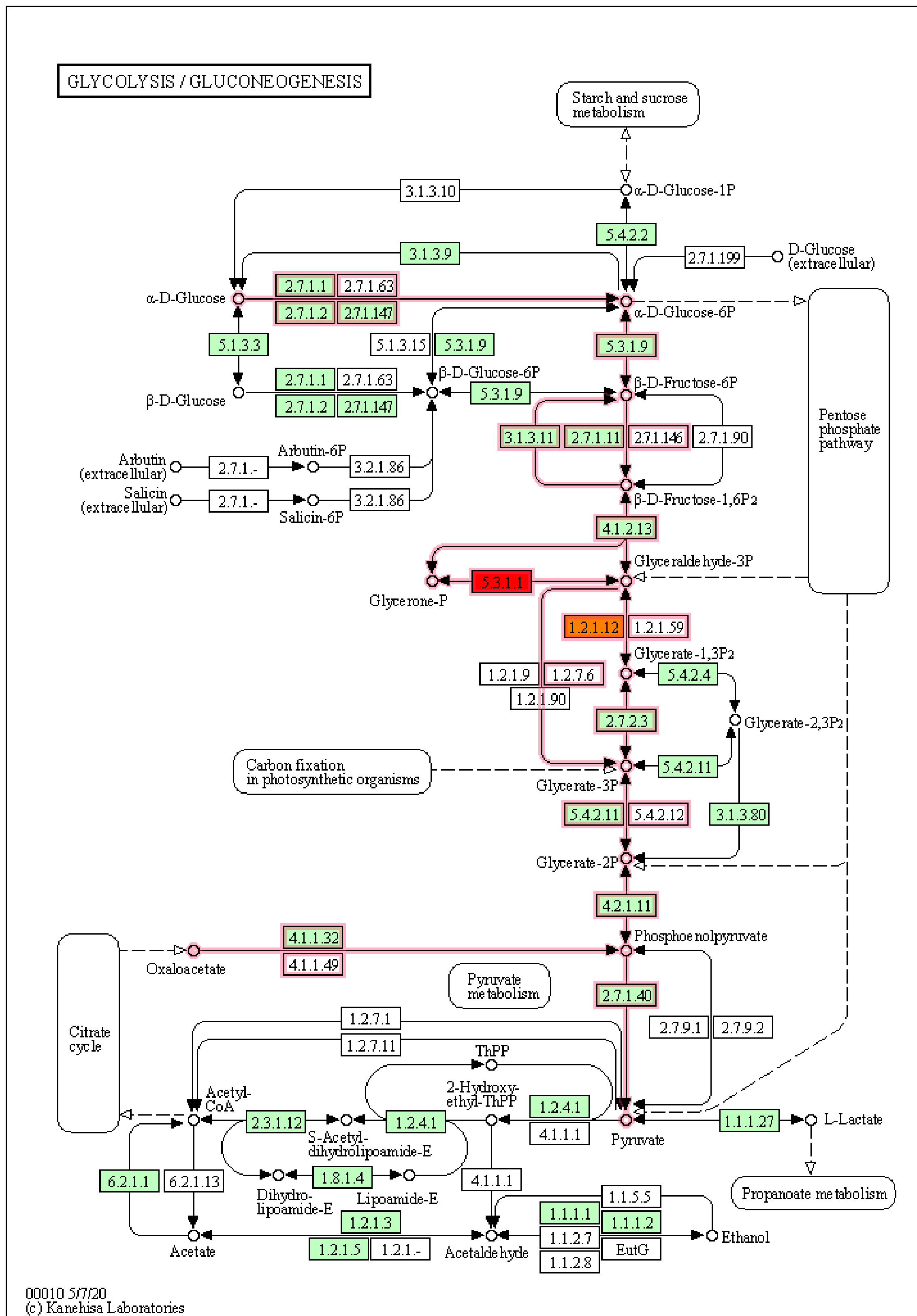
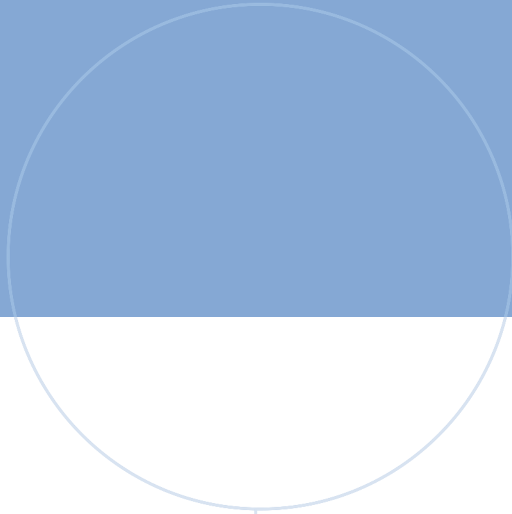
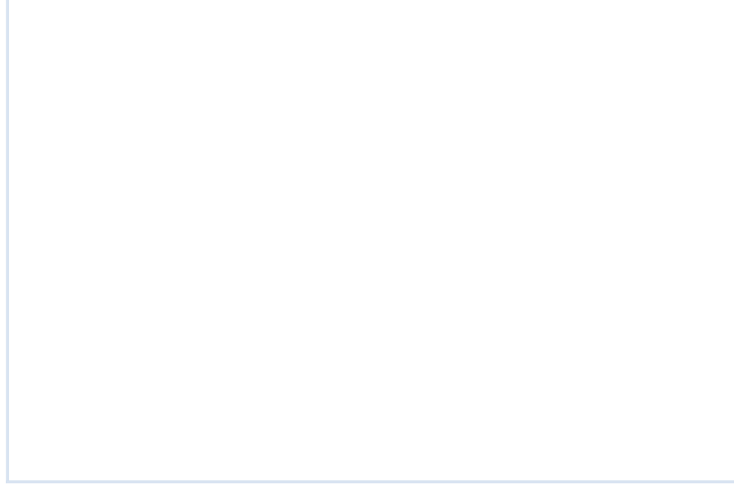


Figure C.1: Network of the human central carbohydrate metabolism. The pathways of glycolysis and gluconeogenesis are highlighted in pink. The highest-scoring SL-target, which is solely fatal for BRCA TP and BRCA TM, but not BRCA NT, is highlighted in red (TPI1) and orange (GAPDH). Figure from [110].

NTNU
Norwegian University of
Science and Technology