

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES

INTEGRATIVE ANALYSIS OF
BLADDER CANCER TRANSCRIPTOME DATA AND
GENOME-SCALE METABOLIC NETWORKS
TO UNDERSTAND DORMANCY MECHANISM

MERVE KUTAY
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF BIOENGINEERING
BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAMME

GEBZE
2019

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**INTEGRATIVE ANALYSIS OF BLADDER
CANCER TRANSCRIPTOME DATA AND
GENOME-SCALE METABOLIC
NETWORKS TO UNDERSTAND
DORMANCY MECHANISM**

MERVE KUTAY
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF BIOENGINEERING
BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAMME

THESIS SUPERVISOR
ASSOC. PROF. DR. TUNAHAN ÇAKIR

GEBZE

2019

T.R.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KANSER UYKU HALİ MEKANİZMASINI
ANLAMAK İÇİN MESANE KANSERİ
TRANSKRİPTOM VERİSİ İLE GENOM
ÖLÇEKLİ METABOLİK AĞLARIN
BÜTÜNLEŞİK ANALİZİ

MERVE KUTAY
YÜKSEK LİSANS TEZİ
BİYOMÜHENDİSLİK ANABİLİM DALI
BİYOİNFORMATİK VE SİSTEM BİYOLOJİSİ PROGRAMI

DANIŞMANI
DOÇ. DR. TUNAHAN ÇAKIR

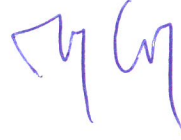
GEBZE
2019

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 03/07/2019 tarih ve 2019/30 sayılı kararıyla oluşturulan jüri tarafından 17/07/2019 tarihinde tez savunma sınavı yapılan Merve KUTAY 'ın tez çalışması BİYOMÜHENDİSLİK Anabilim Dalında Biyoinformatik ve Sistem Biyolojisi Programında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

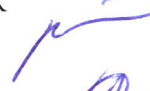
ÜYE

(TEZ DANIŞMANI) : Doç. Dr. Tunahan ÇAKIR



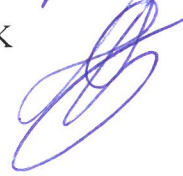
ÜYE

: Dr. Öğretim Görevlisi Pınar PİR



ÜYE

: Prof. Dr. Devrim GÖZÜAÇIK



ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve/..... sayılı kararı.

SUMMARY

Cancer is a complex disease resulting from various genetic changes that lead to disorders in the mechanisms regulating the basic functions of the cell. The metabolism of cancer cells differs from normal cells mostly because they proliferate. In addition to the changes mainly in many signaling pathways, metabolic pathways such as glycolysis, oxidative phosphorylation and macromolecule synthesis are also altered in cancer cells. One of the main problems associated with cancer is the recurrence of cancer even after a long period of treatment, known as metastasis. Cancer cells that leave the primary tumor can be disseminated to other organs and stay there in a dormant state. The metabolism of cancer dormancy should be well known to treat cancer and prevent recurrence. Transcriptome-based approaches have recently been used to understand dormancy mechanisms at the molecular level. However, studies that map transcriptome data on genome-scale metabolic networks to elucidate dormancy metabolism are missing. In this study, transcriptomic data obtained for bladder cancer and its dormancy state were mapped on two different genome-scale metabolic networks (Recon2.04 and iHsa) by two different integration methods (GIMME and E-flux). The integration methods determine the reactions catalyzed by the enzymes encoded by low-expression genes in the metabolic network to constrain them with zero or low rates and predict the rates of other reactions. These simulations were compared for both cancer and cancer dormancy, and changes in the activity of various metabolic pathways were predicted. The results will contribute to the elucidation of mechanisms of cancer dormancy.

Keywords: Cancer, Metabolic Network Models, Dormancy, Transcriptome, Constraint-based Analysis

ÖZET

Kanser, hücrenin temel işlevlerini düzenleyen mekanizmalardaki bozuklukların sonucunda ortaya çıkan ve çeşitli genetik değişikliklerden kaynaklanan kompleks bir hastalıktır. Kanser hücrelerinin metabolizması, hızlı çoğalmalarına yardımcı olacak şekilde normal hücrelerden farklılaşmaktadır. Kanserli hücrede temel sinyal yollarının etkilenmesi dışında glikoliz, oksidatif fosforilasyon ve makromolekül sentezi gibi metabolik yollarda da değişiklik görülür. Değişikliğe uğrayan metabolizmayı anlamak ve tedavi/ilaç bulabilmek adına kanser günümüzde en çok çalışılan konulardandır. Kanser çalışmasında ana sorunlardan biri, kanserin tedaviden uzun bir zaman sonra bile nüksetmesidir. Kanser tedavisi ve nüksetmesinin önlenmesi için kanser uyku halinin metabolizması iyi bilinmelidir. Bu mekanizmaların moleküler düzeyde anlaşılmasında son zamanlarda hesaplamalı biyolojisi yaklaşımından faydalanılmaya başlanılmıştır. Bu kapsamda, transkriptom verileri genom ölçekli metabolik ağlara haritalanarak kanseri biyokimyasal olarak anlamlandırmak hedeflenmektedir. Bu çalışmada kanser ve kanser uyku hali için elde edilen mesane kanseri transkriptomik verileri, literatürde mevcut iki farklı genom ölçekli metabolik ağa (Recon2.04 ve iHsa), iki farklı entegrasyon yöntemiyle (GIMME ve E-flux) haritalanmıştır. Bu yöntemler, ekspresyon seviyesi düşük genler tarafından kodlanan enzimlerin katalizlediği ve metabolik ağda yer alan tepkimeleri belirlemekte ve bu tepkimelerin inaktif olduğu/düşük hızda çalıştığı durumda diğer tepkimelerin hızlarını hesaplamaktadır. Bu simülasyonların hem kanser hem de kanser uyku hali için gerçekleştirilerek karşılaştırılması yapılmış ve çeşitli metabolik yolların aktivitesindeki değişiklikler tahmin edilmiştir. Sonuçlar, kanser uyku halinde rol alan mekanizmaların aydınlatılmasına katkı sağlayacaktır.

Anahtar Kelimeler: Kanser, Metabolik Ağ Modelleri, Uyku Hali, Transkriptom Verisi, Kısıt-bazlı Analiz

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Assoc. Prof. Tunahan akır for his never-ending patience, excellent guidance, encouragement and his support. I am so lucky to have worked under his supervision during my master study. I would like to thank the other members of my thesis committee, Assist. Prof. Pınar Pir, and Prof. Dr. Devrim Gözüaık for their comments and suggestions on my thesis.

I gratefully thank the project team for their nice collaborations and acknowledge TUBITAK for the financial support (Project Code: 216S489).

I thank the members of Computational Systems Biology Group at Gebze Technical University for their very nice collaborations and help. Especially, I would like to thank Elif Emanetci and Merve Yaşar Semiz for their friendship during my thesis. Also I would like to thank Cemal Yıldız for his never-ending patience.

I wish to express my sincere thanks to my mother Serpil Kutay, my father Salih Kutay, my brother Furkan Kutay and especially my sister Melike Kutay for supporting me. Finally, I would like to thank my housemate Serap Uuş and my close friends Ceylan Demir and Büşra Köksal for their friendship and help.

TABLE of CONTENTS

	<u>Page</u>
SUMMARY	v
ÖZET	vi
TABLE of CONTENTS	viii
LIST of ABBREVIATIONS and ACRONYMS	x
LIST of FIGURES	xii
LIST of TABLES	xiii
1. INTRODUCTION	1
2. BACKGROUND ASPECTS	3
2.1. Cancer	3
2.2. Cancer Metabolism	6
2.3. Current Challenges of Cancer Studies	9
2.3.1. Metastasis	9
2.3.2. Dormancy	10
2.3.3. New Approaches of Cancer Therapy	12
2.4. Systems Biology and Metabolic Network Models	13
2.4.1. Genome Scale Metabolic Models	15
2.4.2. Cancer Specific Metabolic Models	16
2.4.3. Reporter Pathway Analysis	18
2.4.4. Flux Balance Analysis	18
2.4.5. Flexo-FBA	21
2.5. Systems Biology and Transcriptomics	21
2.5.1. Dormancy-related Transcriptomic Data	22
2.6. Integrative Analysis of Transcriptome Data with Genome-Scale Metabolic Networks	24
2.6.1. GIMME Algorithm	25
2.6.2. E-flux Algorithm	27
3. RESULTS	30
3.1. Details of the Transcriptomic Dataset	30
3.1.1. Identification of Significantly Changed Genes	31

3.2. Reporter Pathway Analysis of the Transcriptome Data	34
3.3. Updates on Genome-scale Metabolic Network Models	36
3.3.1. Biomass Reaction and ATP Maintenance	36
3.3.2. Medium Composition	38
3.4. FBA of the Genome Scale Metabolic Networks	40
3.6. Detailed Comparison of Flux Distributions	47
3.6.1. Comparison of Pathway Activity	47
3.6.2. Functional Analysis of Genes Catalyzing the Affected Reactions	52
3.6.3. Summary Comparisons	60
REFERENCES	65
BIOGRAPHY	74
APPENDICES	75

LIST of ABBREVIATIONS and ACRONYMS

<u>Abbreviations and Acronyms</u>	<u>Explanations</u>
ALL	: Acute Lymphoblastic Leukemia
COBRA	: COntstraint-Based Reconstruction and Analysis
DGEA	: Differential Gene Expression Analysis
DMEM	: Dulbecco's Modified Eagle's Medium
DTC	: Disseminated tumor cells
ENGRO1	: Enhanced GROwth
FBA	: Flux balance analysis
FBA	: Flux balance analysis
FC	: Fold Change
FCS	: Fetal Calf Serum
GAM	: Growth-associated maintenance
GEO	: Gene Expression Omnibus
GIMME	: Gene Inactivity Moderated by Metabolism and Expression
GO	: Gene Ontology
HMR	: Human Human Metabolic Reactions
INIT	: Integrative Network Inference for Tissues
LP	: Linear Programing
MBA	: Model building algorithm
	Metabolic Context-specificity Assessed by Deterministic
mCADRE	: Reaction Evaluation
NCBI	: National Center for Biotechnology Information
PCA	: Principle Component Analysis
PEA	: Pathway Enrichment Analysis
PPIN	: Protein-protein interaction networks
PPP	: Pentose phosphate pathway
QP	: Quadratic Programing
RPA	: Reporter Pathway Analysis
S	: Stoichiometric Matrix

SIS : Small Intestine Submucosus
TCA : Tricarboxylic acid cycle

LIST of FIGURES

<u>Figure No:</u>		<u>Page</u>
2.1:	The Hallmark of Cancer.	4
2.2:	Metabolic Alterations Associated with Cancer.	8
2.3:	The Invasion-Metastasis Cascade.	10
2.4:	Dormancy.	12
2.5:	Systems Biology approach.	14
2.6:	The Formulation of Flux Balance Analysis (FBA).	19
2.7:	Prosedure of the GIMME algorithm.	26
2.8:	Prosedure of the E-flux Algorithm.	28
3.1:	The general flow of the study.	30
3.2:	PCA results for jbv cell line (P:Plastic gel, S:SISgel).	32
3.3:	PCA results for j82 cell line (P:Plastic gel, S:SISgel).	32

LIST of TABLES

<u>Table No:</u>	Page
2.1: Metabolic network models of human at genome-scale.	16
2.2: Number of cancer-related transcriptomic datasets in GEO.	22
2.3: Number of dormancy-related transcriptomic datasets in GEO.	23
3.1: Gene ontology analysis results of the significantly changed genes. Top 20 of the significantly changed terms are given.	33
3.2: The top 15 most significant pathways obtained from the PEA.	34
3.3: Top 20 most significant pathways obtained by RPA with a cut-off value of 0.05.	35
3.4: Biomass Composition used in the cancer-specific metabolic models. Coefficients are in mmol/gram dry weight.	37
3.5: Medium Compositions used in the simulations, derived from compositions of DMEM and FCS media.	40
3.6: Dimensions of the models obtained by GIMME algorithm.	43
3.7: Flux values in cancer and dormancy states, calculated by the Recon2.04 and GIMME integration.	44
3.8: Flux values in cancer and dormancy states, calculated by the Recon2.04 and E-flux integration.	45
3.9: Flux values in cancer and dormancy states, calculated by the iHsa and GIMME integration.	45
3.10: Flux values in cancer and dormancy states, calculated by the iHsa and E-flux integration.	46
3.11: Table 3.11: Integration of JBV cell line data with the Recon2.04 model using GIMME algorithm. Bold text indicates increased activity.	48
3.12: Integration of J82 cell line data with the Recon2.04 model using GIMME algorithm. Bold text indicates increased activity.	48

3.13:	Integration of JBV cell line data with the iHsa model using GIMME algorithm. *Bold text indicates increased activity.	50
3.14:	Integration of J82 cell line data with the iHsa model using GIMME algorithm. *Bold text indicates increased activity.	50
3.15:	Number of affected reactions and number of genes that catalyze affected reactions.	52
3.16:	GO Analysis results for 297 genes predicted to be affected between cancer and dormancy (jbv cell line and Recon2.04 integration). The most significant 20 GO terms are given.	53
3.17:	GO Analysis results for 280 genes predicted to be affected between cancer and dormancy (j82 cell line and Recon2.04 integration). The most significant 20 GO terms are given.	54
3.18:	GO Analysis results for 371 genes predicted to be affected between cancer and dormancy (jbv cell line and iHsa integration). The most significant 20 GO terms are given.	55
3.19:	GO Analysis results for 330 genes predicted to be affected between cancer and dormancy (j82 cell line and iHsa integration). The most significant 20 GO terms are given.	56
3.20:	PEA Analysis results for 297 genes predicted to be affected between cancer and dormancy (jbv cell line and Recon2.04 integration). The most significant 15 pathways are given.	57
3.21:	PEA Analysis results for 280 genes predicted to be affected between cancer and dormancy (j82 cell line and Recon2.04 integration). The most significant 15 pathways are given.	58
3.22:	PEA Analysis results for 371 genes predicted to be affected between cancer and dormancy (jbv cell line and iHsa integration). The most significant 15 pathways are given.	58
3.23:	PEA Analysis results for 330 genes predicted to be affected between cancer and dormancy (j82 cell line and iHsa integration). The most significant 15 pathways are given.	59
3.24:	The comparison of the significant terms in GO Analysis results for all models. The selected 10 pathways are given.	61

3.25:	The comparison of the significantly terms in PEA results for all models. The selected 10 pathways are given.	61
A1.1:	Composition of Dulbecco's Modified Eagle's medium (DMEM).	74
A1.2:	Composition of Fetal Calf serum (FCS).	75
B1.1:	Reaction Names in Tables (section 3.5).	75
C1.1:	Pathway Names in Recon2.04 model.	76
C1.2:	Pathway Names in iHsa model.	78

1. INTRODUCTION

Cancer is a complex disease that is nowadays prevalent and lethal despite all medical advances. It is affected by multiple parameters such as genetic and environmental factors. Cancer is a group of progressive diseases that are formed by cells that rapidly grow out of control. Cancer cells change their basic metabolism in order to survive in the body. With the help of the molecular changes, they can escape from the immune system, adapt to the environment, and rapidly proliferate [Hanahan and Weinberg 2011]. Changing mechanisms is one of the most studied subjects in the field of cancer treatment.

However, a major problem arises in cancer studies: metastasis [Valastyan and Weinberg, 2011]. Metastasis, known as the ability of cancer cells to spread to other organs, is an important area in the treatment of cancer. Metastasis mechanisms and metabolism of metastatic cells should be well known in cancer treatment and prevention of recurrence. A significant proportion of metastases are micrometastases, which can remain quiescent for years and are resistant to treatment. An important cause of resistance is chemotherapy and radiotherapy treatments that target dividing cells. Most micro-metastases do not divide and are called dormant cells. The cells in the dormant state are alive, not senescent, but do not divide. Inexplicably, dormant cells can wake up years after primary tumor formation and cause recurrent tumor formation in patients [Yeh and Ramaswamy, 2015]. Therefore, the molecular mechanisms behind the entry of cancer cells into the dormant state should be investigated.

Although the metabolism of cancer is well known, due to the complexity of the disease, there is no universal treatment and therefore novel approaches are developed to treat cancer. One recently common approach is to analyze genome-wide (omic) cancer data via bioinformatics approaches. Metabolic network models, one of the bioinformatic approaches, have been used to explain the causes of metabolic changes observed in cancer cells. Genome-scale metabolic networks provide a platform for the biochemical interpretation of omic data. The integration of omic data with metabolic models is used to produce tissue-specific or condition-specific metabolic network models. In short, the study of metabolic models produced for the purpose of personalized treatment and drug targeting enables researchers to analyze the disease mechanisms and determine strategies to control them. Cancer-specific genome scale

metabolic models have also been created in the last decade. On the other hand, there is not much information about dormancy metabolism in the literature, and a metabolic-model based analysis of dormancy metabolism is missing in the literature although there is a small number of dormancy-related transcriptome data available in Gene Expression Omnibus (GEO) [Clough et al 2016] database.

The main aim of this study is to understand the molecular changes between cancer and dormancy by using a computational biology approach. For this reason, a cancer-specific metabolic network is constructed here by integrating already existing human-specific genome-scale metabolic networks with transcriptome data related to cancer and dormancy. This model provides valuable insights by elucidating cancer-specific pathways, and gives clues on the mechanism of dormancy for a better interpretation of cancer metabolism.

Chapter 2 includes background information about cancer, metabolic network models, transcriptome data and computational methods for integrative analysis. Chapter 3 explains how to reconstruct and validate two different metabolic network models. Chapter 3 also presents the mapping of the transcriptomic data obtained from two different cell lines related to cancer and dormancy into metabolic models by using two different integration algorithms. Discussion of the main results are given in Chapter 4. Finally in Chapter 5, the major conclusions obtained from the analysis of transcriptome-derived cancer and dormancy specific metabolic networks are given together with recommendations on future work.

2. BACKGROUND ASPECTS

2.1. Cancer

Cancer is a complex disease that has been increasing in frequency recently and is fatal despite all medical advances. In 2018 only, 18.1 million new cancer cases and 9.6 million cancer deaths were recorded [Bray et al., 2018]. Cancer is a class of diseases characterized by reprogramming of some important processes such as growth, aging, death and dissemination. It is a complex disease resulting from defects in the mechanisms that regulate the functions of the cell such as stimulation and differentiation. Moreover, various genetic and epigenetic changes take place in cancer disease [Nilsson and Nielsen, 2016]. Genetic mutations in a cell may alter the amount or activity of proteins involved in regulating cell survival.

Tumor cells exhibit a number of different characteristics, such as excessive growth rate, local invasion, differentiation, anaplasia and metastasis, which are not observed in normal cells. In addition to the loss of control over growth in tumor cells, intercellular relationships are also impaired [Coller, 2014]. For the development of cancerous cells, signaling, apoptosis and limited cleavage mechanisms must be disrupted in the cells [Charlton and Spicer, 2015].

Carcinogenesis, the process of cancer, is a complex dynamic process that is linked to multiple variables and regulated on multiple spatial and temporal scales. [Grizzi and Chiriva-Internati 2006]. The distinctive features of cancer consist of ten biological characteristics acquired during the multi-step development of human tumors: self-sufficiency in growth signals, insensitivity to growth suppressor signals, resistance to cell death (apoptosis), unlimited potential for division, angiogenesis (vascular formation) ability, reprogramming energy metabolism, escape from immune system, genome instability and mutations, and inflammation supporting cancer (Figure 2.1) [Hanahan and Weinberg 2011]. These are known as hallmarks of cancer.

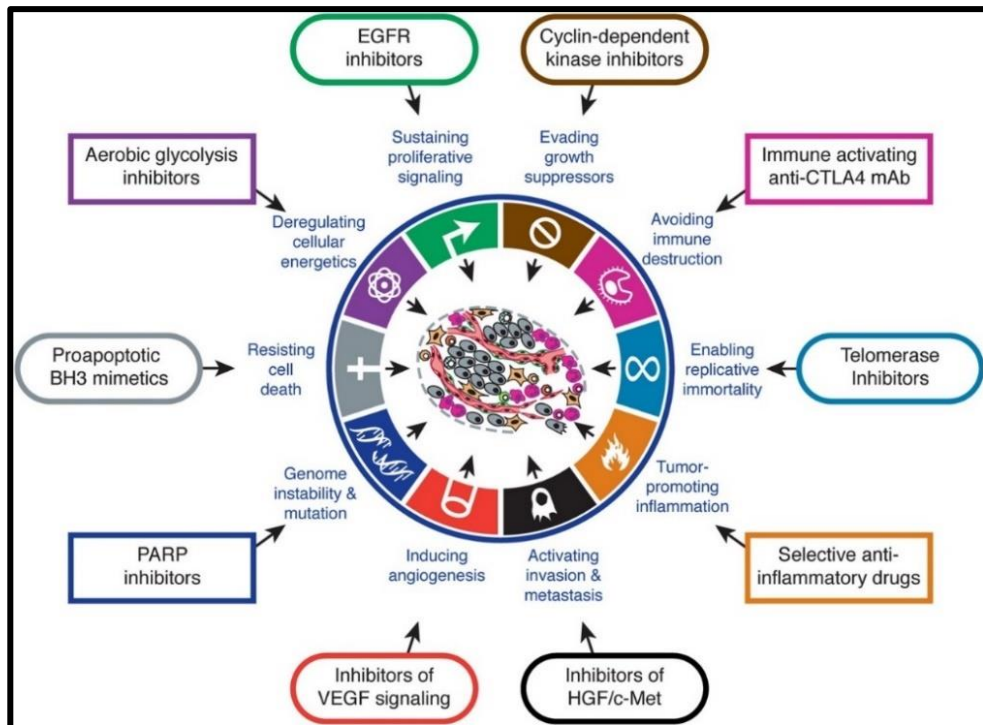


Figure 2.1: The Hallmark of Cancer.

Cancer cells have an ability to proliferate in an uncontrolled manner by maintaining proliferative signaling. Normal cells decide to proliferate or to pass into a passive state by stopping their growth when they reach a certain growth rate, by taking signals from the external environment. Cell proliferation is largely dependent on avoiding suppressor signals. Cancer cells have defects in regulatory circuits that govern normal cell proliferation and homeostasis. In general, cancer-associated genes can be divided into two groups: oncogenes and tumour suppressor genes (TSGs). Oncogenes normally function to promote cell growth/division in a controlled manner whereas tumor suppressor genes normally function to prevent abnormal proliferation and apoptosis in any DNA damage. [Harrington, 2015]. In the case of cancer, the activity of oncogenes increases greatly, while tumor suppressor genes become inactive because of DNA mutations. Mutations are observed in most cancer cells in tumor suppressor genes and in oncogenes. Tumor suppressor genes code for proteins that slow down cell growth. They can halt the cell growth cycle to stop unnecessary division or promote apoptosis if the cell DNA is damaged. [Harrington, 2015]. When DNA damage occurs in normal cells, tumor suppressor genes become functional and cell proliferation is stopped. Then, DNA repair mechanisms are activated. For the integrity of the organism, when the accumulated DNA damage in the cells becomes

irreparable, apoptosis becomes activated, which is programmed cell death. Cancer cells gain resistance to apoptosis as a result of mutations in DNA damage sensors such as the tumor suppressor p53 gene or loss of function in their proteins. In principle, the irregular proliferation program in cells is sufficient to ensure the formation of enormous cell populations that constitute macroscopic tumors [Hanahan and Weinberg 2011]. In some cases, the defect in the expression of these genes is not due to mutation, but is caused by genome instability caused by epigenetic mechanisms such as histone modifications and DNA methylation [Berdasco and Esteller 2010].

The ends of chromosomes, telomeres, consist of several thousand repetitions of short sequences of 6 bp. There is a loss of 50-100 bp of telomeric DNA at the ends of all chromosomes throughout the cell cycle. Death is inevitable for cells that cannot be further cleaved as a result of this telomere erosion [Counter et al. 1992]. Cancer cells perform this conservation by increasing the expression of telomerase enzyme, which adds six nucleotide repeats to the ends of telomeric DNA [Bryan and Cech 1999]. Keeping telomere above critical level with increased telomerase activity gives cancer cells unlimited ability to divide.

Oxygen and nutrients supplied by blood vessels are crucial for cell function and survival. The development of blood vessels and the formation of new vessels, called angiogenesis, are largely completed during embryo development. Angiogenesis is active in parts of physiological processes such as wound healing and the female reproductive cycle in adults, but only temporarily. It is, however, almost always active during tumor development [Hanahan and Weinberg 2011].

Normal cells usually maintain their position except stem cells. This changes when they reach a certain size for cancer cells. The main feature of cancer cells is the ability to migrate to other tissues and organs, called tissue invasion and metastasis. During the development of most human cancers, sooner or later, primary tumor masses produce precursor cells with mobility, which can travel to distant sites to invade neighboring tissues and establish new colonies [Hanahan and Weinberg 2011].

Foreign substances entering our bodies are recognized by the cells of defense system and destroyed by our defense system in order to prevent damage to the organism. Tumor cells can protect themselves from the detection by the immune system. They achieve this by synthesizing the cell surface proteins that normal cells possess and allowing them to pass through the control of the defense system as a normal body cell [Hanahan and Weinberg 2011].

In addition to standard chemotherapy, radiotherapy and surgical methods, some special vaccines have started to be used as gene therapy and hormonal therapy for the treatment of cancer. It is impossible to mention the existence of a single definitive treatment method because each method has its own advantages and disadvantages. Cancer is a person-specific disease, and the treatments may vary from person to person. With the advancement of technology, new personalized treatment methods are being developed in addition to existing treatments.

2.2. Cancer Metabolism

The metabolism of cancer cells differs from normal cells in terms of main cellular processes. These differences help the cell adaptation and escaping from immune system. The change in the uptake rates of glucose and amino acids, the use of glycolysis/TCA cycle intermediates for biosynthesis and NADPH production, the increase in the demand for nitrogen, and metabolic interaction with the microenvironment are among the most important metabolic processes in cancer. [Pavlova and Thompson, 2016]. The other key metabolic processes include aerobic glycolysis, reduced oxidative phosphorylation and the increased biosynthesis of macromolecules that are needed for cell growth and proliferation.

Cancer cells exhibit atypical metabolic properties. They mostly prefer aerobic glycolysis instead of oxidative phosphorylation for the generation of ATP [Nilsson and Nielsen, 2016, Damiani et al., 2017]. Rapidly proliferating tumors convert glucose to lactate in an aerobic environment, although the amount of ATP to be obtained from one molecule of glucose in the glycolysis pathway is only two at maximum. This feature of cancer cells is called the "Warburg Effect" [Danhier et al., 2017]. The Warburg effect plays a key role in most metabolic pathways present in tumor cells [Asgari et al., 2015]. It provides an advantage of survival to cancer cells due to being resistant to transient or permanent hypoxic conditions. It also contributes to the production of nucleosides and amino acids through increased glucose uptake in cancer tissues and provides a very fast way to produce energy. It has been suggested that increased uptake and preference of glucose in the direction of lactate in catabolism is intended to promote biomass accumulation and to maintain redox balance in proliferating cells. Such regulation of glucose metabolism is achieved by

overexpression of specific membrane and lactate carriers of glucose and most intermediates of the glycolytic pathway, as well as the increase in the amount of responsible enzymes involved in the conversion of glucose to lactate [Pavlova and Thompson, 2016]. Glycolysis is associated with many other metabolic pathways where various glycolytic intermediates serve as substrates. In high glucose uptake rates, the flow of glycolytic metabolites from the glycolytic pathway to the branched biosynthetic intermediates is significantly increased (Figure 2.2) [Cazzaniga et al., 2015].

Pentose phosphate pathway (PPP) is particularly important for cancer cells. PPP does not only produce pentose phosphates to meet the high rate of nucleic acid synthesis of cancer cells, but also provides NADPH, which is essential for both fatty acid synthesis and cell survival under stress conditions [Vander Heiden and DeBerardinis, 2017].

In addition to the excessive energy requirement, cancer cells accumulate building blocks for the construction of new cellular components, including nucleic acids, lipids and proteins, as well as important cofactors for maintaining cellular redox levels. The serine biosynthetic pathway in anabolic pathways constitutes an important milestone in glucose conversion [Kato et al., 2018]. Serine, which is derived from a branch of transported serine and glycolysis, can be converted to glycine, which provides carbon units for single carbon metabolism. The regulation of serine biosynthesis is directly associated with the maintenance of tumor homeostasis [Amelio et al., 2014].

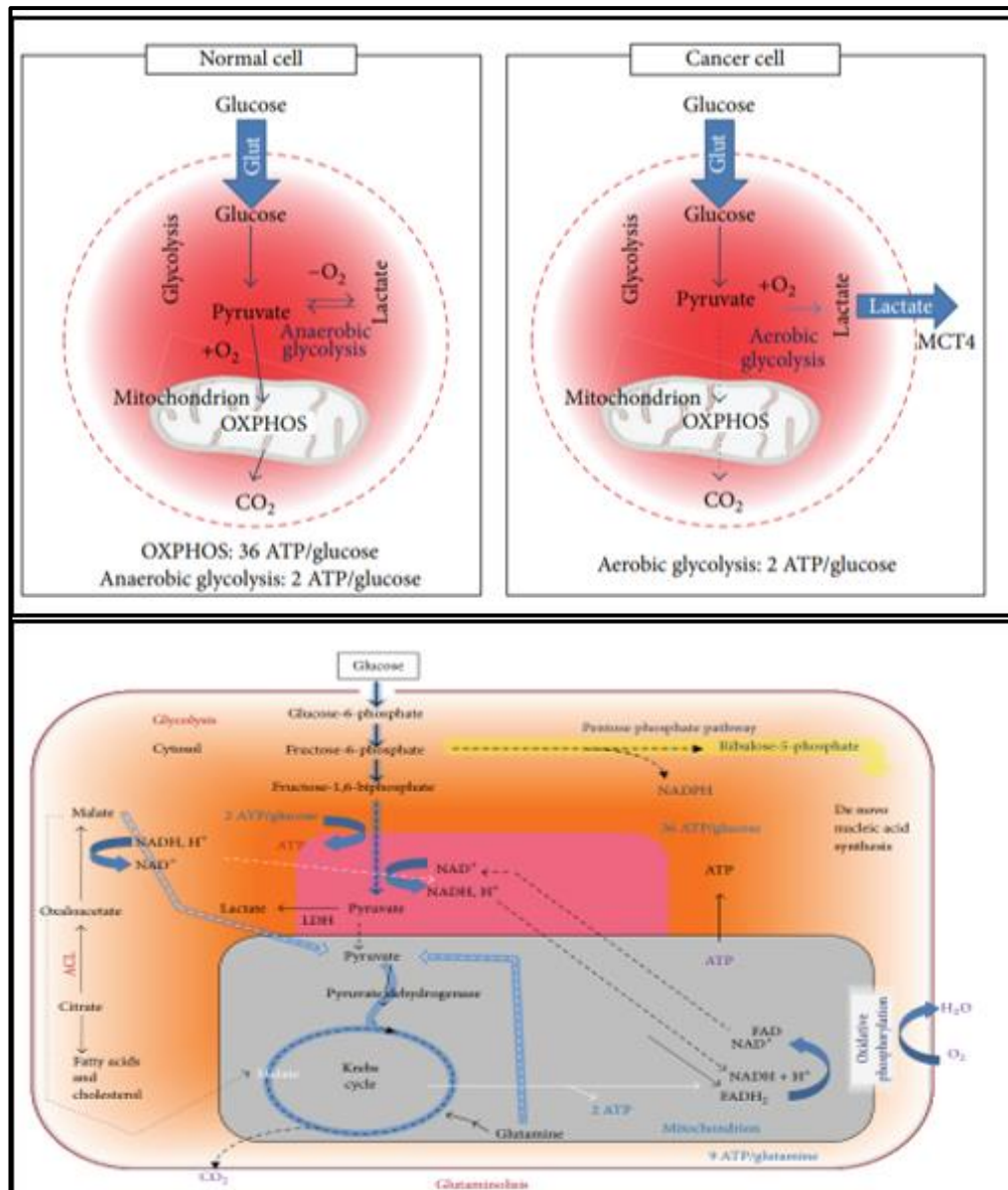


Figure 2.2: Metabolic Alterations Associated with Cancer.

Glucose and glutamine are two main sources in cancer cells, used to maintain active essential carbon flow, to support glycolysis and the tricarboxylic acid cycle (TCA). Glutamine has long been known to play more important role in the metabolism of proliferating cells as compared to other amino acids [Eagle, 1955]. Glutamine acts as a carbon source [Daye and Wellen, 2012] to contribute to continuous growth and proliferation of cancer cells by supporting ATP production, protein, lipid and nucleic acid biosynthesis (Figure 2.2). It is converted to alpha-ketoglutarate, a TCA cycle intermediate, which is used to replenish other TCA cycle metabolites, some of which act as precursors for biosynthesis. Glutamine also regulates redox homeostasis and acts on signal transduction pathway activity.

2.3. Current Challenges of Cancer Studies

Although the growing knowledge on molecular and tumor biology is increasing day by day, there are many issues that need to be resolved in the field of cancer. Cancer is affected by both genetic and environmental factors, making it difficult to study [Zugazagoitia et al., 2016]. Although it is one of the most common diseases, it has many unresolved problems such as metastasis and dormancy.

2.3.1. Metastasis

Although there is a better understanding of cancer biology today thanks to the developments in molecular biology and genetics, cancer-related deaths are still high in developed countries and account for 25% of all deaths in the community. Metastasis is responsible for 90% of cancer deaths [Spano et al., 2012]. Metastasis is the spread of cancer cells from the region of origin to different tissues and organs of the body [Poste and Fidler, 1980]. This occurs through a series of interrelated complex and multistep events, such as angiogenesis, invasion, migration-motility, extravasation, and proliferation.

In order for a metastasis to develop, tumor cells must be separated from the primary mass and complete a series of processes beginning with invasion of the surrounding tissue. The metastasis process consists of two stages; the first is migration, which includes the stages of intravasation, dissemination and extravasation, and the second is colonization (Figure 2.3) [Valastyan and Weinberg, 2011]. First, the tumor cells stimulate the formation of new vessels and then disconnect from the primary tumor tissue by severing their ties with neighboring cells. Tumor cells then migrate to the extracellular matrix where they progress and reach the surrounding tissues or enter the circulatory system and invade distant tissues [Valastyan and Weinberg, 2011]. In this way, they continue to live and proliferate.

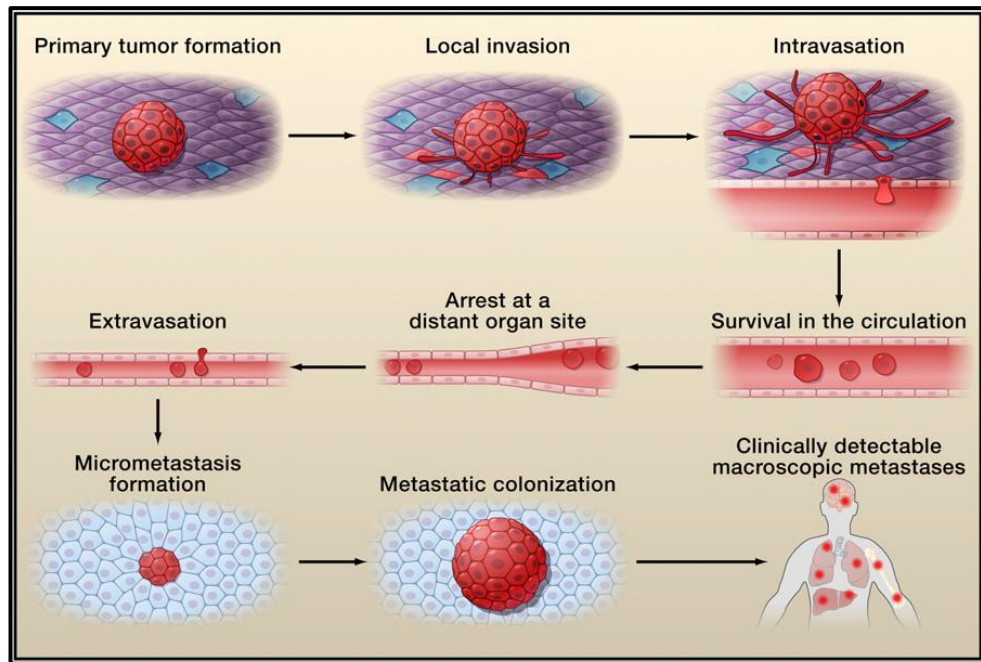


Figure 2.3: The Invasion-Metastasis Cascade.

Many molecular factors have been identified for metastasis formation. Cadherin family, integrins and cell adhesion molecules have been found as important modulators in the development of metastasis [Hart,1998]. In addition, angiogenesis and lymphangiogenesis for tumor cell migration are important biological events involved in invasion and metastasis [Valastyan and Weinberg, 2011].

Although many steps and molecular agents have been identified so far that were involved in the development of metastasis, the mechanisms of tumor metastasis have not been fully understood [Spano et al., 2012]. In recent years, even with the partial understanding of the metastasis mechanisms, therapies targeting these mechanisms have been developed.

2.3.2. Dormancy

Although cancer treatments used today are successful in destroying primary tumors, they are not curative for metastatic tumors. Many metastatic tumors can become cancerous even after many years [Wells et al., 2013]. Since the recurrence of cancer is the main cause of cancer-related deaths, the relapse process has been widely studied in recent years. Patients show no symptoms before cancer relapse, and the cell biology of this condition is not yet clear [Yeh and Ramaswamy, 2015].

Understanding the molecular, cellular, genetic and clinical mechanisms underlying metastatic progression gains importance in developing novel diagnostic and therapeutic strategies to prevent and treat metastases. Since metastases are always caused by disseminated tumor cells (DTCs), it is important to understand the biology of dormant tumor cells [Seidi et al., 2018]. DTCs are cells that are physically separated from the primary tumor mass and spread through the circulation. Cancer dormancy is the ability of DTCs to remain in low numbers and in quiescence after tumor treatment [Gomis and Gawrzak, 2016]. These cells may not be detected for many years, and they may get activated at a certain indefinite time, causing disease recurrence. Since they do not proliferate in the dormant state, they cannot be detected and killed by standard cancer treatment methods. Although it is not yet clear why and when DTCs enter the inactive state, two main approaches have been assumed: (i) the cells may remain in dormant state after separation from the primary tumor, or (ii) the cells may become dormant in the early stage of tumor development and they metastasize after reactivation [Seidi et al., 2018].

Several mechanisms have been proposed for basic cancer dormancy, such as cell cycle arrest, immune dysfunction, angiogenic insufficiency, stress-induced kinase and urokinase receptor imbalance [Aguirre-Ghiso, 2007]. It is possible to divide cancer dormancy into two main groups: tumor mass dormancy and cellular dormancy (Figure 2.4). In tumor mass dormancy, there is no net tumor development due to the balance of proliferation and cell death rates. It is divided into two classes as angiogenic and immunological dormancy [Ruppender et al., 2013]. Cellular dormancy, which is the state where the cell is quiescent by entering the G0 stage of the cell cycle, is important in metastatic studies. Since tumor mass dormancy is considered to be a more transient state of anti-stress, cancer metastasis studies have focused on cellular dormancy [Aguirre-Ghiso, 2007]. Factors such as extracellular matrix (ECM) modification, metabolic reprogramming, the effect of miRNAs, epigenetic modifications, and changes in stress are effective in cellular dormancy [Seidi et al., 2018]. It is assumed that cellular dormancy is triggered by extracellular and intracellular signals [Gao et al., 2014]. Although there are approaches to explain dormancy mechanisms, studies on dormancy metabolism are limited. Also, tumor dormancy and reactivation are important elements for metastatic relapse.

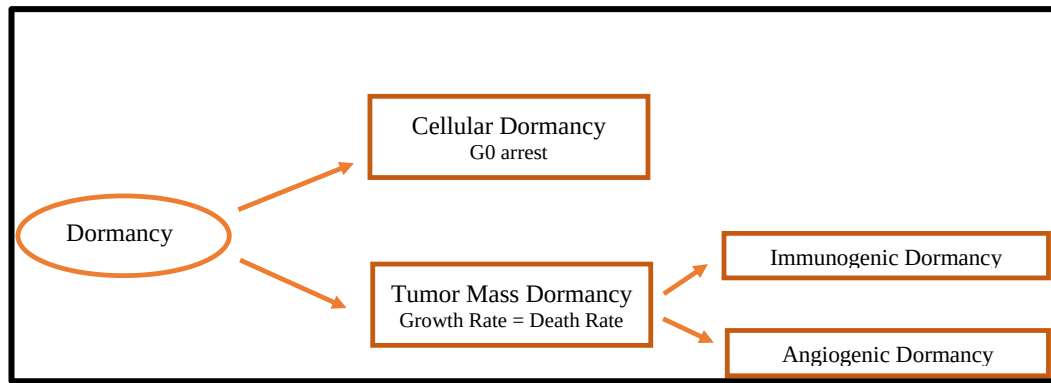


Figure 2.4. Dormancy.

In order to better understand the mechanisms of cancer dormancy state, some questions have been focused on: (i) what triggers the awakening of dormant cancer cells, (ii) the similarity between cancer stem cells and dormant cells; and (iii) how dormancy mechanisms are affected by cancer types and metastatic environment [Yeh and Ramaswamy, 2015]. It will be possible to find new therapeutic strategies targeting these dormant cancer cells by better understanding the cancer dormancy mechanisms and associated metabolic changes.

2.3.3. New Approaches of Cancer Therapy

Understanding the relationship between the metabolism of normal cells and of cancer cells can provide the benefit for the development of novel approaches targeting tumor metabolism. Targeting cancer metabolism by focusing on altered metabolic pathways is a new therapeutic window for cancer therapy [Teicher et al., 2012]. This approach can allow early cancer detection and diagnosis or be a predictive and pharmacodynamic marker of drug effects.

Although the metabolism of cancer is well known, a universal therapy of the disease has not been found yet because of the complexity of disease. For the detailed investigation of cancer, a relatively newer perspective has been computational biology approach. Recently, the study of cancer metabolism by using computational approaches has been more common [Yizhak et al., 2015& Aurich and Thiele, 2016]. The computational approaches help us to see how interacting components can produce complex system behaviors, and how changes in conditions may alter these behaviors. Metabolic network models constitute an ideal framework for the investigation of

cancer. They have been used to elucidate the cause of metabolic alterations that are observed in cancer cells [Aurich and Thiele, 2016]. Mathematics and biology can contribute together to the goal of solving cancer complexity. Modelling the growth and development of human tumors by using mathematics and biological data is a promising part of cancer research [Grizzi et al., 2006].

2.4. Systems Biology and Metabolic Network Models

Systems biology, which is a holistic approach to solving the complexity of biological systems, focuses on the understanding of cellular systems as a whole rather than focusing on its parts [Kobeissy et al., 2012]. Systems biology uses the dynamics and structure of biological systems to produce models and utilizes wide range of fields such as mathematics, physics, engineering, and computer science in addition to the biological sciences for this goal (Figure 2.5) [Kherlopian et al. 2008]. The systems biology models integrate knowledge and data obtained from experimental biology using mathematical modeling [Reichle and Vogt, 2008]. Thus, systems biology simulates complex biological networks by the aid of computational methods and numerical techniques. Here, the goal is understanding of system processes, mechanisms and principles of biology [Gonzalez-Angulo et al., 2010]. In summary, systems biology provides holistic methods for the interpretation and validation of new findings in experimental biology.

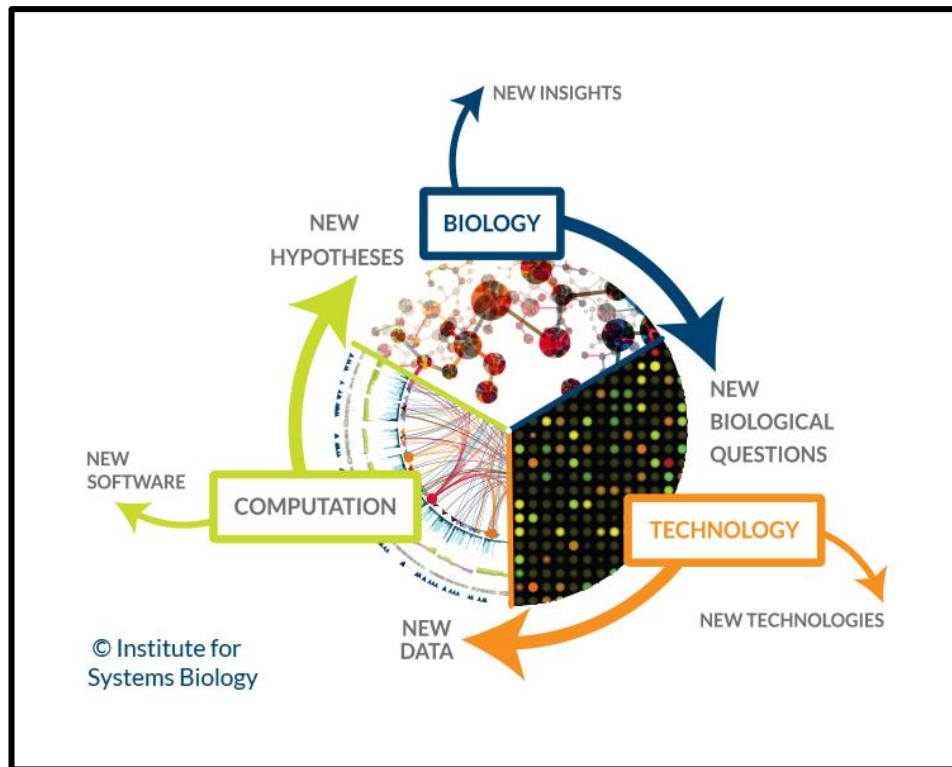


Figure 2.5: Systems Biology approach.

By means of system biology, the physiology and functioning of the cells can be understood more clearly. System biology can interpret the phenotype-genotype relationship and explain how the phenotype and genotype are affected by the environment. [Likić et al., 2010]. This approach can also be used in studies of the complex diseases such as cancer to allow for the interpretation of the detailed mechanisms and treatment of diseases [Gonzalez-Angulo et al., 2010]. There may also be preventive and personalized treatment methods for multifactorial complex diseases that can be developed by systems biology based approaches. For personalized therapies, systems biology focuses on the modeling of cancer, where high-throughput data is integrated with complex mathematical modeling [Ram et al., 2012].

Systems biology requires a thorough examination of the interaction networks between biomolecules within the cell to provide a holistic analysis. Intracellular networks can be investigated in three different categories: protein-protein interaction networks, gene regulation networks including interactions of DNA and protein, and metabolic networks. Protein-protein interaction networks (PPIN) are mathematical representations of physical contact between proteins within the cell [Safari-Alighiarloo et al., 2014]. Gene Regulatory Networks are networks that are formed by genes and

transcription factors that control the activity of genes. A transcription factor may control more than one gene and there may be more than one transcription factor controlling one gene [Emmert-Streib et al., 2014]. This type of network is important for understanding the mechanism of activity of genes in a cell under a given condition. Metabolic networks are the networks formed by the transformation of the chemical molecules necessary for the cell through enzymes [Çakır and Khatibipour, 2014]. In addition to central carbon metabolism, they consist of hundreds of reactions including amino acid, lipid and carbohydrate metabolisms. Metabolic network is a widely studied biological network type in systems biology.

Cancer research is an important area to apply systems biology methods. The high throughput data such as genomics, proteomics, and metabolomics provide a basis for new types of cancer research. New analysis techniques can be used to better understand both the dynamics of cancer and to enable the design of targeted therapeutics [Kherlopian et al., 2008].

2.4.1. Genome Scale Metabolic Models

Genome-scale metabolic network models are used to explain the cause of metabolic changes observed in cells. They constitute an ideal framework for the investigation of diseases, including cancer. Genome-scale metabolic networks provide a platform to interpret omics data in a biochemically meaningful manner [Nilsson and Nielsen, 2016]. Genome-scale metabolic modeling of human metabolism (Table 2.1) creates a reaction set that contains all metabolic reactions known to be present in human cells, and the modeling is mostly based on calculating the rates of these reactions. Human genome-scale metabolic models and their basic features are shown in Table 2.1. These analyses allow researchers to investigate disease paths and define strategies to control them. For example, the publication of Recon1, which is the first human genome-scale metabolic network, enabled the investigation of human physiology and metabolism of diseases in detail by using computational approaches. These metabolic models can be used for the integration of omics data and visualization.

Table 2.1: Metabolic network models of human at genome-scale.

Model	Genes	Reactions	Metabolites	Year- Author
Recon1	1905	3742	2766	Duarte et al. 2007
The Edinburgh Model	2322	2823	2671	Ma et al. 2007
Recon2	2194	7440	5063	Thiele et al. 2013
Recon3D	3288	13543	4140	Brunk et al. 2018
HMR2	3765	8181	6007	Mardinoğlu et al. 2013
iHsa	2315	8264	5620	Blais et al. 2017

2.4.2. Cancer Specific Metabolic Models

Genome-scale metabolic models are used to determine molecular mechanisms from the perspective of metabolic events and to identify candidate drug targets by creating condition-specific/tissue-specific metabolic models for many diseases or tissues. Different approaches have been used to create a cancer specific model. The first cancer-specific genome-scale metabolic network model was developed in 2011 [Folger et al., 2011]. It was derived from Recon1. To create the first cancer-specific model, 197 genes that have high expression levels in all cancer cell lines found in NCI-60 were first identified. Then, using the model building algorithm (MBA) [Jerby et al., 2010], which provides automated reconstruction of human tissue metabolic models, a cancer-specific model was obtained within the framework of these genes. The generated model contained 813 reactions, 665 metabolites and 696 genes. Afterwards, the model was validated and gene knock-out simulations were performed to predict candidate drug targets [Folger et al. 2011]. In 2012, condition-specific cancer metabolic models were generated by integrating omic data with HMR and Recon1 genome-scale metabolic networks. Accordingly, specific models for 16 cancer types were reconstructed using the Integrative Network Inference for Tissues (INIT) algorithm [Agren et al., 2012]. Another team reconstructed tissue-specific cancer models for 126 tissues using the metabolic Context-specificity Assessed by Deterministic Reaction Evaluation (mCADRE) algorithm [Wang et al., 2012]. None of these automatically generated models by INIT and mCADRE were used in detail for the prediction of cancer growth rates or rates of metabolic reactions. Another cancer-specific model was reported in 2014 by using a different approach, which was

an alternative to the cancer specific model developed in 2011. In the 2014 model, authors first focused on the tumor suppressor genes (81) and oncogenes (141) in Recon 1. Reactions that are controlled by the tumor suppressor genes were inactivated and the oncogenes-related reactions were activated in their flux calculation approach. The metabolites detected in the body fluid were allowed to be taken up by the cancer metabolic model in order to achieve a more realistic model in the simulation. Therefore, the uptakes for 400 metabolites, which are thought to be available in all body fluids including blood and urine, were allowed in the cancer cell simulations, with a maximum possible uptake rate of $10 \text{ mmol} / \text{gDW} / \text{h}^{-1}$. The resulting model contains 3788 reactions, 2766 metabolites and 1905 genes. The study focuses on the comparison of the predictions by 2011 and 2014 models. This new model was shown to reflect cancer metabolism more accurately than the 2011 model [Hadi and Marashi 2014]. In the following years, core models specific to cancer have been reconstructed. For example, in 2017, Zielinski et al. have developed a core metabolic model that includes 360 metabolites and 382 reactions, which are related to basic cancer metabolism. Also, the same year Damian et al. developed a core metabolic model called ENGRO1 (Enhanced GROwth) to investigate Warburg effect, which includes 67 metabolites and 84 reactions [Damian et al., 2017].

None of the mechanisms can be fully accounted for with only the generic human metabolic network model since the metabolism is also influenced by environmental conditions and external stimuli besides genetic factors. The integration of the omics data with metabolic models is therefore preferred for the generation of cell/tissue or condition specific metabolic models [Blaizer and Papin, 2012]. Moreover, cell or disease specific metabolic models have been applied for personalized cancer treatment and drug target identification [Nilsson and Nielsen, 2016]. Therefore, global understanding of metabolic networks by integrating them with omics data will allow researchers to identify and investigate cancer-associated pathways and identify strategies to control them.

2.4.3. Reporter Pathway Analysis

Metabolite-centric Reporter Pathway Analysis (RPA) can be performed in order to gain an overview of the effect of intracellular changes on metabolic pathways. It is a relatively recent bioinformatics approach that focuses on the effects of perturbations such as diseases on metabolic pathways by mapping the changes in gene expression on genome scale metabolic networks [Cakir, 2015]. The approach is helpful in drawing conclusions about which metabolic pathways are activated or inactivated in the network, based on changes in transcriptome data. For this analysis, the p-values of the genes obtained based on the comparison of expression data is required as well as a genome-scale metabolic network. The metabolites in the reactions are scored with the help of p- values of the genes catalyzing the reactions. Then, the scored metabolites for each pathway are averaged to determine metabolite-centric reporter pathways. [Cakir, 2015]. In RPA, only average change in pathways is calculated, rate values of reactions are not calculated.

2.4.4. Flux Balance Analysis

Constraint-based modeling is an approach in systems biology that analyzes phenotype in a metabolic network using biological, physical, and chemical constraints [Feist and Palsson, 2008]. The important point is that this approach assumes that the organism is at resting state (steady state).

Flux Balance Analysis (FBA) is the most popular method that uses constraint-based steady-state approach to simulate metabolism, and it is used for the prediction of metabolic fluxes [Schellenberger et al., 2011]. FBA needs some parameters such as equality constraints in the form of a list of mass balances, thermodynamic and measurement constraints (reaction reversibility information and maximum and minimum values that reaction rates can take), and an optimality assumption (objective function). It solves the optimization problem using linear programming and creates a flux distribution [Colijn et al. 2009]. Applying FBA, we can examine which reactions that create intracellular metabolic network are active in that specific condition and how pathways in cell metabolism are influenced with a condition change. [Blazier and Papin 2012]. Moreover, FBA is one of the most common methods to determine drug

targets using genome-scale metabolic networks since it can predict growth rates, leading to phenotypic information on condition-specific metabolisms [Aurich and Thiele 2016].

FBA predicts metabolic steady-state fluxes by applying mass balance constraints. The formulation of FBA includes five steps as (i) Genome-scale metabolic network reconstruction, (ii) mathematical representation of metabolic reactions and constraints, (iii) mass balance definition, (iv) definition of the objective function and (v) calculation of fluxes (Figure 2.6) [Orth et al., 2010].

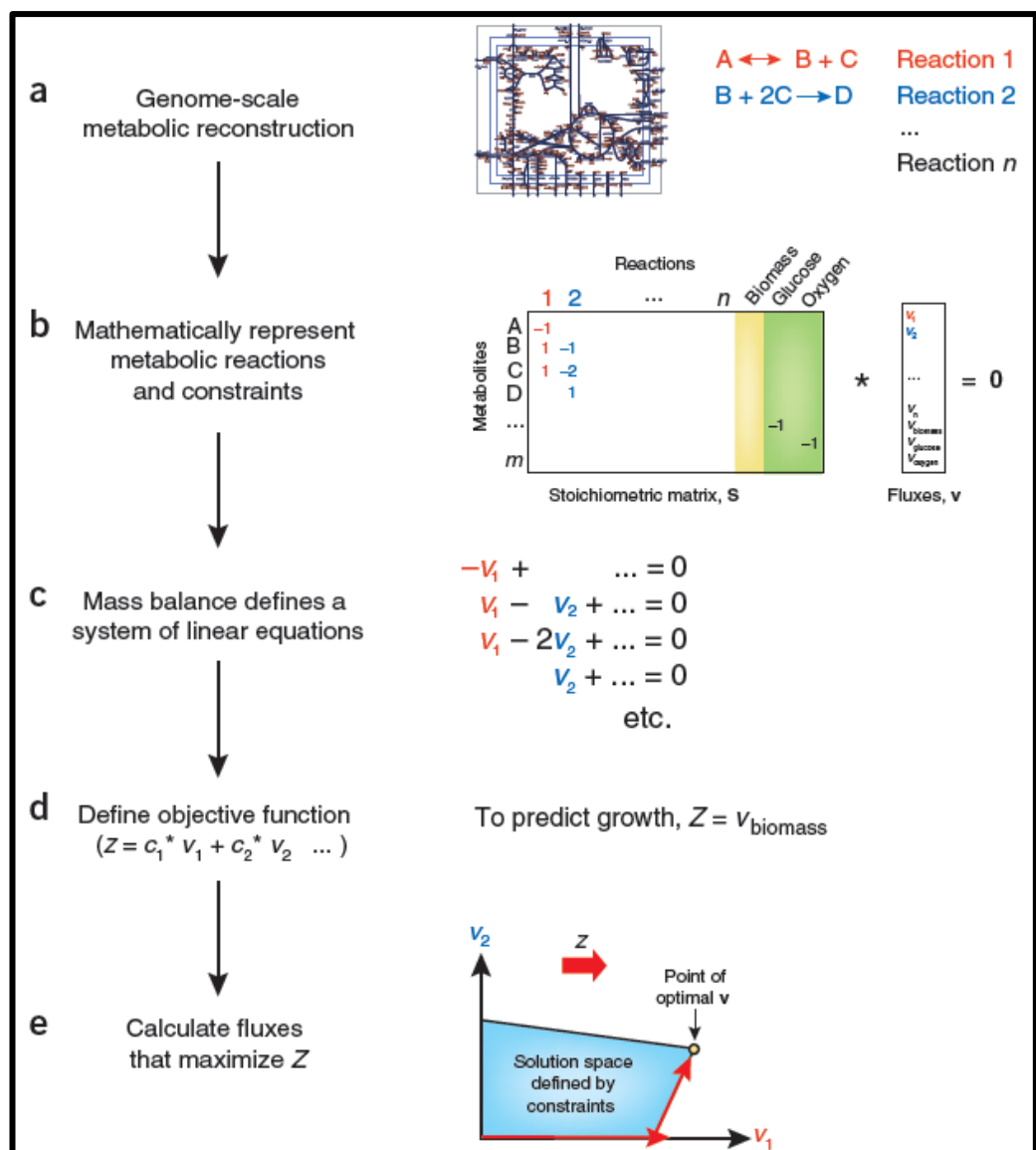


Figure 2.6: The Formulation of Flux Balance Analysis (FBA).

The first step of the FBA is definition of all metabolites and their reactions, and description of the dynamic mass balance equations around all intracellular metabolites, as demonstrated in Figure 2.6. The set of differential mass balance equations can be presented in matrix form, leading to a stoichiometric matrix, S . The stoichiometric matrix represents the relations between the rates of n reactions and mass balances around m intracellular metabolites. Each reaction is represented as a column in the matrix, and each metabolite as a row. The stoichiometric matrix (S) is a linear transformation of the flux vector (v) to a vector of time derivatives of the concentration vector (x), which represents the concentration of each metabolite as in Equation 2.1. dx / dt shows the accumulation of intracellular metabolites over time. The assumption of FBA is that the concentration pools for the intracellular metabolites will be non-changing at the steady state. FBA solves the equation system at steady state to simplify the solution. Thus, Equation 2.1 becomes simpler as in Equation 2.2.

$$\frac{dx}{dt} = S \times v \quad (2.1)$$

$$S \times v = 0 \quad (\text{steady state}) \quad (2.2)$$

In biological systems, there are many more reactions (unknown rates) than metabolites (mass balances). Since there are more unknown variables than equations in these systems, they are called under-determined systems. There is no single unique solution for an underdetermined problem. Therefore, additional constraints in the form of measured flux values need to be defined to shrink the solution space for the calculation of the flux distribution [Orth et al., 2010]. A measured constraint is represented as maximum or minimum values for the rates of reactions. These constraints are represented mathematically as the lower and the upper bounds of that flux. Every reaction is bound by two values, an upper bound (ub) and lower bound (lb), which are represented as vectors with size equal to number of reactions (unknowns) [Förster et al., 2003].

To solve the constraint-based optimization problem by using the approach of maximization or minimization, the next step in FBA is to define an objective function that is biologically meaningful. The maximization of cellular growth, the maximization of the production of a particular metabolite and the minimization of energy consumption are commonly used objective functions for FBA simulation.

These objective functions are defined as a linear equation or quadratic equation and solved with linear programming (LP) or quadratic programming (QP) to obtain a flux distribution of the reaction set. COBRA (COntstraint-Based Reconstruction and Analysis) [Schellenberger et al., 2011] toolbox, which is a free MATLAB toolbox, can be used as a solver interface for these calculations.

2.4.5. Flexo-FBA

The cells in a population are not exactly the same, they have slight differences from each other. Considering the heterogeneity of culture, not all cells can have the same maximum growth rate. A new approach, flexoFBA (flexible-optimal FBA), can be used to simulate this phenomenon [Tarlak et al., 2014]. In the FlexoFBA approach, the growth rate is calculated by obtaining a flux distribution with the primary objective function, which is the maximum growth. Then, new flux distributions are calculated with the help of the secondary objective function such that the growth rate is bounded in a certain range showing normal distribution. Thus, slight flexibility in cells around optimum values is taken into account based on the range defined by normal distribution [Tarlak et al., 2014]. Flexo-FBA assumes that the cells cannot reach the theoretical optimum, and they can be represented by a flux distribution corresponding to the average of a more flexible flux distribution of cells within a cell population. Thus, it is the flux calculation method that enables a more accurate calculation of intracellular metabolic fluxes. The mean or median of the obtained flux distributions can be compared with the classical FBA results.

2.5. Systems Biology and Transcriptomics

Omics data are large biological datasets. They are generated by the simultaneous detection of a high number of genes, mRNAs, proteins and metabolites in a specific biological sample. [Manzoni et al., 2018]. Generation, analysis, integration and interpretation of omics data require detailed techniques. Omics data covers, among others, the following fields: genomics, transcriptomics, proteomics, metabolomics.

Genomics field investigates the functions and interactions of genes in the whole genome of an organism. Transcriptomics is the simultaneous analysis of transcripts

generated by transcription from the cell genome. Proteomics focuses on the structure, location, amount, post-translational modifications and functions of all proteins present in the cell at a given time, and their interaction with other proteins and macromolecules [Manzoni et al., 2018]. Metabolism in a cell or living organism consists of enzymatic conversion of small molecules into each other. These small molecules are referred as metabolites. Metabolomics is the determination, quantification and identification of all metabolites in the cell [Clish, 2015].

Transcriptome is a term used to express all of the gene transcripts (RNA) in a cell or tissue at a given time. In other words, Transcriptomics is the simultaneous analysis of transcription-generated mRNA transcripts. It aims to measure the level of expression of all of the genes from the amount of RNA present in a sample. Whole-genome mRNA analysis methods are microarray and RNA-seq approaches.

The Gene Expression Omnibus [GEO] database includes publicly available high-throughput transcriptome data. GEO database is provided by the National Center for Biotechnology Information (NCBI). GEO records data with special accession numbers. The GSE prefix is used to define datasets. The total number of cancer related datasets identified by a search using “cancer” as the keyword in GEO is given in Table 2.2 (the search was performed in June 2019).

Table 2.2: Number of cancer-related transcriptomic datasets in GEO.

	All Organisms	Mus musculus	Homo Sapiens
Number of total datasets	24585	6095	17886
Expression profiling by array	17573	3976	13030
Expression profiling by high throughput sequencing	7012	2119	4856

2.5.1. Dormancy-related Transcriptomic Data

Since dormancy mechanisms are very crucial to elucidate in order to develop new treatment options for cancer, transcriptomic analysis of dormant cells has recently been a focus. The number of the data sets in GEO related to dormancy is very small. In GEO, the total number of all datasets found with a search using “dormancy” and “cancer” as keywords is given in Table 2.3 (the search was performed in June 2019).

Table 2.3: Number of dormancy-related transcriptomic datasets in GEO.

	All Organisms	Mus musculus	Homo Sapiens
Number of total datasets	36	11	22
Expression profiling by array	29	6	20
Expression profiling by high throughput sequencing	7	5	2

In a study, different extracellular matrices were used to create dormant cell line models [Hurst et al., 2013]. The plastic (collagen) matrix is isolated from normal tissue and symbolizes normal cancer condition. SISgel is isolated from the Small Intestine Submucosus (SIS), and it was used to create dormant models. Matrigel from mouse sarcoma was used to create metastatic model. In that study, bladder cancer cells were separately cultured in all these matrices. No growth was observed in cells in the SISgel matrix. The matrix suppressed tumor cell phenotype and inhibited tumor formation and metastasis. Therefore, SISgel is thought to be used as a model for dormancy. Two different cell lines, JBV and J82, were used in the study. JBV bladder cell line is highly metastatic, and it was derived from the J253 TCC line. J82 bladder cell line is highly invasive, and it was obtained from ATCC. With three replicates, 18 different experiments were performed in total. Transcriptome analysis of the samples from the experiments was also performed. The data is available in GEO under the accession ID of GSE35947. In their study, they did not provide any analysis result for the transcriptomic data, they instead focused on the results for a proteomic analysis, where protein levels were compared to investigate metastasis mechanisms by using samples of the SISgel and Matrigel.

In another study, the dormancy model was produced mimicking minimal residual disease, which is related to tumor relapse [Ebinger et al. 2016]. Proliferation-dependent dyes were used for the selection of dormant or cancer cells. If the dye remains stable in the cell, it was identified as dormant cell. It is the opposite in proliferative cells. There are 228 transcriptomic samples collected in the study, including single cell and bulk samples. Of those, 10 bulk samples and 50 single cell samples are related to dormancy. Transcriptomic data were obtained by the RNAseq method [Ebinger et al. 2016] The data is available in GEO under the accession ID of GSE83142. In this study, the characterization of rare, dormant and treatment resistant

cells in Acute Lymphoblastic Leukemia (ALL) was investigated. Also, the signature genes were identified for dormancy.

In another study, the reactivation of the dormant cells with osteoclasts was emphasized [Lawson et al. 2015]. The samples were classified based on the duration of the experiment. It is accepted that the cells in 7 days are in dormant phase, in 14 days they are in transitioning phase, and after 14 days they are in proliferation phase. The dataset contains a total of 12 samples collected from mice: 3 dormancy, 4 transition states and 5 proliferation states. The data is available in GEO under the accession ID of GSE57695. Consequently, it is concluded that state transitions are related with extracellular interactions, and, hence, the potential for targeting extracellular mechanisms is highlighted for cancer therapy.

2.6. Integrative Analysis of Transcriptome Data with Genome-Scale Metabolic Networks

It is important that a biomass reaction that demonstrates cell proliferation must be added into the reactions of human genome-scale metabolic network in order it to be used in the simulations of cancer cells. The biomass reaction forms all fractional contributions of biosynthetic precursors and key cofactors to form 1g of biomass. The cell growth in living systems can be simulated with the inclusion of biomass reaction into metabolic network models. By adding a biomass reaction, the first step towards creating a cancer-specific model is taken.

The integration of transcriptomic data into the genome-scale metabolic model as additional constraints is an effective method for the formation of a cancer specific model. Since metabolism is affected by environmental factors as well as by external stimuli and genetic factors, none of the mechanisms can be fully explained by the generic models of human metabolism [Yizhak et al. 2015]. The integration of omic data with metabolic models can be used to characterize the condition-specific metabolism. Also, these models can be used to find drug targets or molecular markers for the disease conditions [Aurich and Thiele 2016] and to predict system level intracellular metabolic fluxes [Kim and Lun, 2014]. Many different algorithms such as IMAT [Shlomi et al., 2008], MADE [Jensen and Papin, 2011], GIMME [Becker ve Palson, 2008], E-Flux [Colijn et al., 2009], PROM [Chandrasekaran and Price, 2010]

are used for the generation of cell specific metabolic network models from transcriptome data for personalized medicine or for the analysis of cell specific functions. When the algorithms are compared with each other, none of them are guaranteed to be the best or the most accurate method [Machado and Herrgård, 2014]. The selection of the best method depends on various factors, which include biological systems of interest, the primary objective of the study, and the availability of experimental data [Kim and Lun, 2014]. Many of these methods such as GIMME, E-flux use MATLAB as a programming and simulation environment, and the COBRA toolbox [Özcan and Çakır, 2016].

The activity of reactions in genome-scale metabolic networks varies based on the cell type and the condition. Therefore, genes that catalyze reactions should also be included in the simulations to form a condition-specific model, which could be facilitated by integrating transcriptomic data with these models. An important issue in mapping transcriptomic data on a metabolic network is how to perform the mapping for the reactions controlled by multiple genes. In the trivial state, a gene encodes an enzyme that catalyzes a single reaction. Many enzymes, however, accept several different substrates that associate the expression of a gene with several fluxes [Akesson et al., 2004]. Most methods map gene expression data in the model based on Gene-Protein-Reaction (GPR) relationships. The GPR represents functional relationships between genes/proteins and the corresponding reactions they catalyze or control. If the proteins responsible for the reaction form an enzyme complex, the minimum of expression values is assigned as the reaction score. If the responsible proteins are isozymes, which are different enzymes that independently catalyze the same reaction, the sum of expression values or maximum expression values is assigned as the reaction score.

Below, more detailed information is given for two of these approaches (GIMME and E-Flux), which are used in this study.

2.6.1. GIMME Algorithm

The most common integration algorithm is GIMME (Gene Inactivity Moderated by Metabolism and Expression) [Becker ve Palson, 2008]. The GIMME algorithm creates a context-specific metabolic model that predicts a subset of the likely active reactions of the cell under certain conditions by using the gene expression data. This

algorithm has a 2-step procedure: (i) use of FBA to find an active flux distribution of all reactions around the objective function and (ii) reducing the number of the reactions with the help of a threshold value by using the expression levels of genes that catalyze these reactions [Kim & Lun, 2014]. The GIMME algorithm uses a genome-scale metabolic network, the transcriptome dataset and at least one defined metabolic functionality or objective function. The algorithm also requires a threshold to decide which reactions are active or inactive. Given these inputs, the algorithm generates a list of reactions that are predicted to be active in the network and an inconsistency score that quantitatively classifies the discrepancy between the gene expression data and the assumed objective function. This inconsistency score allows a relative comparison of how well each gene expression data will match a particular metabolic function. Simply, reactions corresponding to mRNA transcript levels below a specified threshold are temporarily inactive. If the cell cannot provide the desired functionality without at least one of these reactions, linear optimization is used to find the most consistent set of reactions to reactivate [Blaizer and Papin, 2012]. Inconsistency scores are calculated based on the required flux and distance from the threshold for each reaction that needs to be reactivated, as shown in Figure 2.7 [Blaizer and Papin, 2012]. As a result, a functional context-specific network model is obtained.

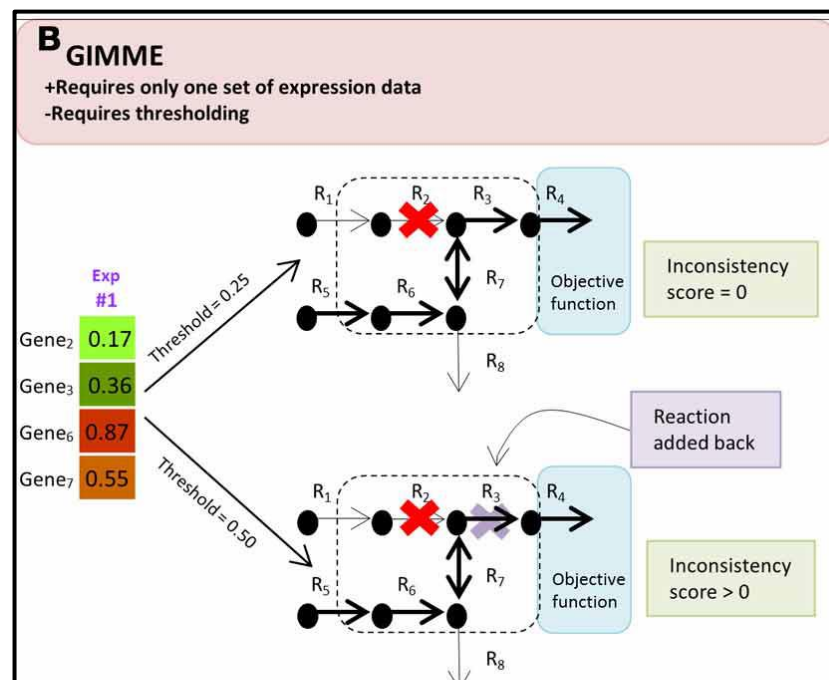


Figure 2.7: Procedure of GIMME algorithm.

Accordingly, the algorithm decreases the number of reactions and metabolites in the model and forms a smaller model. The resulting context-specific network model can be used to estimate more accurate and realistic flux distributions by FBA compared to predicted flux distributions from generic genome-scale metabolic networks without transcriptome data. The GIMME algorithm is implemented in MATLAB using the functions in the COBRA Toolbox.

The GIMME method was first applied to estimate the cellular activities of *E.coli* and human skeletal muscle cells under different conditions. The authors compared the phenotypic characteristics of the condition-specific metabolic models with the experimental results. The results showed that this method provides realistic predictions consistent with the experimental results. Compared to non-reduced genome scale metabolic networks, more consistent networks were been created with fewer active reactions [Becker ve Palson, 2008]. In another study, the cell-specific networks for three human cell types (adipocytes, hepatocytes, and myocytes), which were obtained with GIMME algorithm were manually curated and validated based on known cellular metabolic functions [Bordbar et al., 2011]. In the other study, the growth-implemented brain specific genome-scale metabolic network was integrated with glioblastoma multiforme (GBM) gene expression data by using GIMME algorithm to obtain GBM-specific models and this model was used to predict the metabolic phenotypes of the GBM and to compare with experiments [Özcan and Çakır, 2016]. The algorithm was applied for cancer [Damiani et al., 2018 & Tobalina et al., 2016]. In this study of cancer, they compared GIMME and IMAT algorithms with the single-cell Flux Balance analysis method that they developed to focus on the extracellular fluxes of the dominant single cell population [Damiani et al., 2018]. In the other article, context-specific networks for 174 cancer cell lines were reconstructed for the identification of cancer-specific essential metabolic genes by using some integration algorithms including GIMME [Tobalina et al., 2016].

2.6.2. E-flux Algorithm

Another commonly used integration method is the E-Flux method [Colijn et al., 2009]. E-flux estimates metabolic states by adjusting maximum flux constraints as a function of measured gene expression. E-Flux uses the FBA technique to incorporate gene expression data into metabolic flux calculation scheme in the form of upper-

bound constraints for reaction rates. In this way, when the expression of a particular enzyme-coding gene is low, a tighter constraint is imposed for the rate of corresponding reaction, and a more flexible constraint is imposed for the reaction rate when its expression is high. Thus, E-flux allows us to directly correlate gene expression data with metabolic fluxes [Colijn et al., 2009].

Instead of setting the upper boundary of a reaction rate to some infinite constants or zero, E-Flux limits the upper boundary of a reaction rate to the level of gene expression of interest relative to a given threshold. The expression level of each gene is normalized to the maximum gene expression level in all genes before they are used as upper bounds for the rates. The normalized expression data are entered as a constraint to the reactions controlled by those genes as maximum possible flux (Figure 2.8) [Blazier and Papin 2012]. GPR rules are taken into account in E-flux algorithm as in GIMME for gene expression assignment.

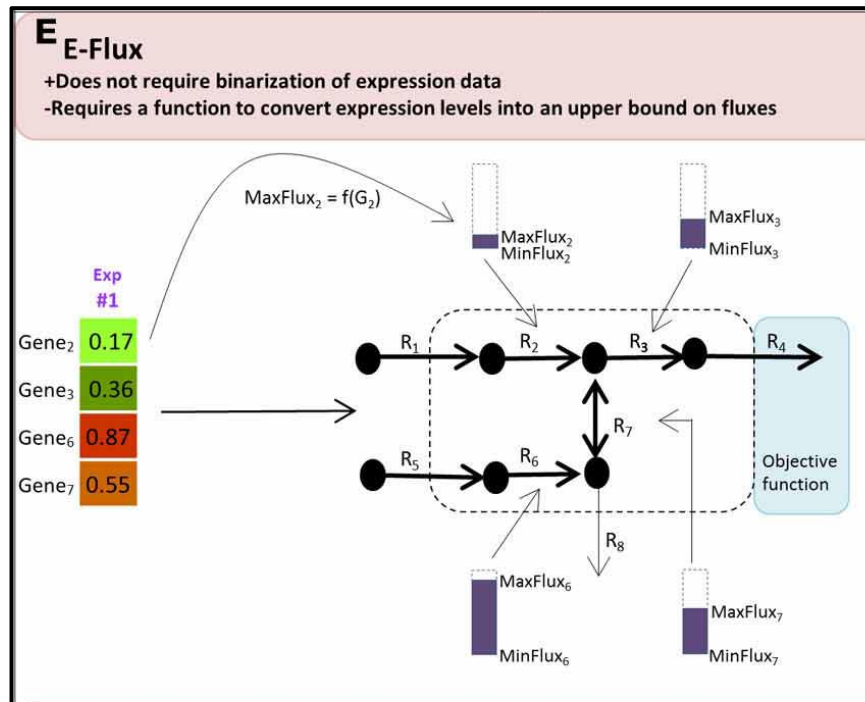


Figure 2.8: Procedure of the E-flux Algorithm.

As shown in Equation 2.3, a_j represents the lower boundary of the rate of reaction j , while b_j represents the upper boundary. The b_j value is determined based on the level of normalized gene expression data, and it constrains the reaction associated with the gene. If a reaction is reversible in a model, the flux value can be in the range of $-\infty$ and $+\infty$. the negative flux indicates that the reaction was carried out

in the opposite direction. When E-flux algorithm is applied, if the reaction is reversible, a_j is equal to the negative version of b_j . Otherwise, a_j equals zero.

$$a_j \leq V_j \leq b_j \quad (2.3)$$

The flux distribution is calculated using the flux constraints and the objective function entered, but the flux values are relative values since they are calculated without using any measured rate value. To compare with experimental flux data, the resulting distributions should be scaled with a well-known experimental flux value such as glucose uptake [Blazier and Papin 2012]. The flux distribution obtained as a result of scaling can be used for further analysis.

The E-Flux method was first applied to *Mycobacterium tuberculosis* [Colijn et al., 2009] for several first-line tuberculosis drugs by using the biosynthesis of mycolic acids as targets. The specificity of seven inhibitors for fatty acid biosynthesis and the effects of different drugs and nutrients conditions on the capacity of mycolic acids biosynthesis were correctly predicted by E-Flux [Colijn et al., 2009]. The approach was later applied to human [Asgari et al., 2015]. In this study, the gene expression of 13 different cancer cells were integrated with the genome-scale metabolic network of human (Recon1) based on the E-Flux method and analyzed by using constraint-based modeling. Results show that the distribution of metabolic changes such as Warburg effect is similar in different cancer types [Asgari et al., 2015].

3. RESULTS

In this study, gene expression data from Hurst et al. (2013) was first analyzed by basic bioinformatic analyzes, and then different computational approaches were used for detailed analyses. The general flow of the study is given in Figure 3.1.

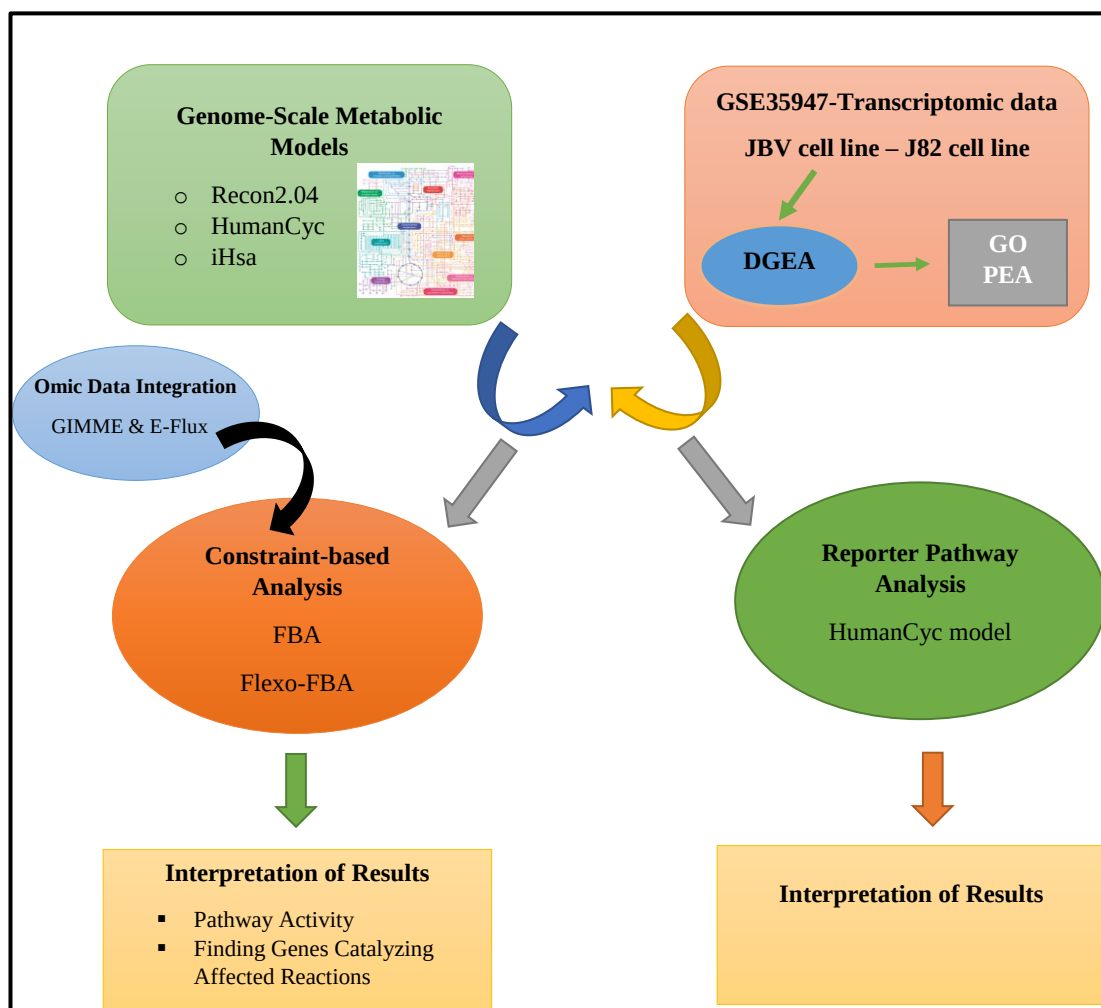


Figure 3.1. The general flow of the study.

3.1. Details of the Transcriptomic Dataset

In order to shed light on cancer dormancy mechanisms from the aspect of metabolic changes, the transcriptomic dataset of Hurst et al. (2013), available in GEO database with accession number (GSE35947), was chosen. The study focused on the comparison of the bladder cancer cell lines on Matrigel, Plastic gel and SISgel, using

the data collected on the transcriptomic and proteomic levels. Matrigel shows a profile that increases metastatic/invasive effect while SISgel suppresses cell growth. Plastic gel represents normal cancer tissue. In that study, the authors did not provide any analysis of their transcriptomic data stating that there was no significantly changed genes that changed in the same direction for both cell lines. Therefore, they studied protein levels to investigate metastasis mechanisms by using samples of the cells grown on SISgel and Matrigel. In their study, the comparison of growth on SISgel and Matrigel was used to identify re-activation mechanisms.

In this thesis, in contrast to their study, Matrigel and SISgel comparisons yielded 7 significantly changed genes with common profiles in two cell lines. Also, in contrast to their work, this thesis focuses on the comparison of plastic and SISgel samples to see how dormant status differs from normal cancer cells, rather than the mechanism of dormancy reactivation. Therefore, the samples from both cell lines (jbv and j82) grown on SISgel and Plastic gel were used for the identification of the suppression mechanisms in the dormant state.

3.1.1. Identification of Significantly Changed Genes

First of all, this dataset was analyzed by Principle Component Analysis (PCA), which is a dimension reduction method, and it was checked whether the replicate samples in the data showed a consistent grouping (Figure 3.1-3.2). PCA results show that the samples of two conditions are properly separated from each other and clustered. Therefore, this dataset was used for further statistical analysis. Differential Gene Expression Analysis (DGEA) was performed for the biological interpretation of gene expression differences between the conditions. DGEA is the procedure for performing statistical analysis to discover quantitative changes in expression levels between experimental groups [Ritchie et al. 2015, Smyth 2005]. The analysis was performed using R statistical programming language. Using the limma package, gene lists were obtained that differ significantly between dormancy and cancer status (with increased or decreased expression profile) for both cell lines. Limma is an R software package that provides an integrated solution for analyzing transcriptomic data. Over the last decade, limma has been a popular choice for gene discovery through differential expression analysis of microarray and high throughput PCR data. The

package contains particularly powerful features for reading, normalizing and analyzing this type of data [Ritchie1et al. 2015].

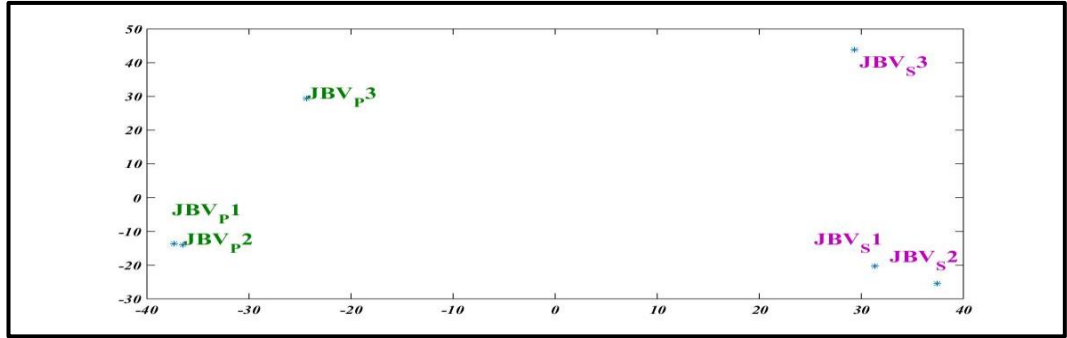


Figure 3.2: PCA results for jbv cell line (P:Plastic gel, S:SISgel). The x axis symbolizes PC1 (44.5%) and the y axis symbolizes PC2 (28.6%).

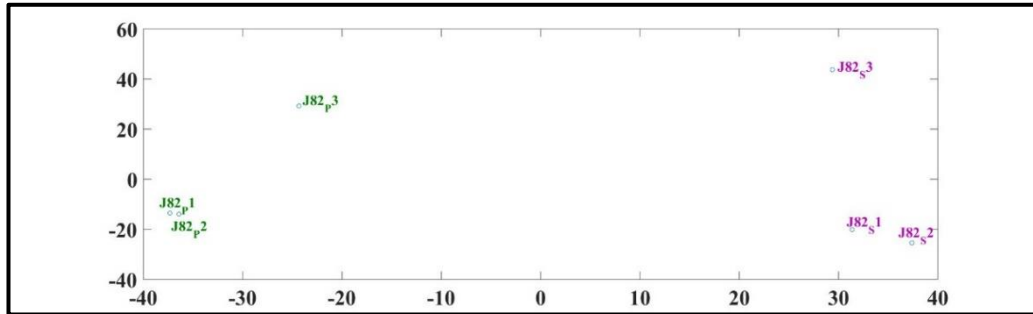


Figure 3.3: PCA results for j82 cell line (P:Plastic gel, S:SISgel). The x axis symbolizes PC1 (50.2%) and the y axis symbolizes PC2 (27.0%).

The limma package has one parameter called false discovery rate (FDR), which is used as a threshold. FDR means the expected proportion of false discoveries in the positive (significantly changed) items (genes) [Dittrich et al., 2008]. The FDR approach is one way to control for the number of false positives. FDR is closely related to the Benjamini-Hochberg procedure, which uses q-values (adjusted p-value) [Benjamini and Hochberg, 1995]. The false discovery rate is a less conservative approach to correct for multiple testing correction than the traditional methods such as Bonferroni correction. In the limma package, the cut-off value for the significance analysis was chosen as 0.05 for FDR and 1.5 for fold change. Significantly changed genes in the two cell lines were compared with each other and genes with common profiles were determined. The common profiles mean that the expression of those significantly changed genes increase or decrease in the same direction in both cell lines (i.e. they either increase in both cell lines, or decrease in both cell lines). A total of 204

significantly changed genes with the common profile were identified in 2 cell lines, 141 of which decreased and 63 increased in dormancy.

A common approach to interpret significantly changed genes is gene ontology analysis. This analysis is useful to find out whether differently expressed genes are associated with a particular biological function or molecular function. Gene Ontology (GO), which contains descriptions of genes, is a widely used database for this purpose. The analysis was performed with the help of the g:Profiler platform for 204 genes mentioned above [Reimand et al., 2016]. Biological process section of GO database was used in the analysis for the significant changed gene lists. Gene ontology analysis results are shown in Table 3.1. Top 20 of the significantly changed terms are given.

Table 3.1: Gene ontology analysis results of the significantly changed genes. Top 20 of the significantly changed terms are given.

Biological process	Adjusted p-value
Cell cycle	1.355×10^{-27}
Mitotic cell cycle	4.543×10^{-27}
Cell cycle process	6.464×10^{-26}
Mitotic cell cycle process	3.664×10^{-25}
DNA replication	1.660×10^{-20}
Chromosome segregation	4.476×10^{-18}
DNA-dependent DNA replication	5.754×10^{-18}
Cell division	2.795×10^{-16}
Mitotic cell cycle phase transition	5.180×10^{-16}
Cell cycle phase transition	6.227×10^{-16}
Cellular response to DNA damage stimulus	1.932×10^{-12}
Nuclear chromosome segregation	6.372×10^{-12}
Cell cycle G2/M phase transition	1.377×10^{-11}
DNA metabolic process	2.346×10^{-11}
Nuclear division	4.355×10^{-11}
Regulation of cell cycle process	1.710×10^{-10}
Regulation of mitotic cell cycle	2.412×10^{-10}
G2/M transition of mitotic cell cycle	2.438×10^{-10}
Chromosome organization	2.462×10^{-10}
Organelle fission	2.804×10^{-10}

In addition to GO analysis, Pathway Enrichment Analysis (PEA) was performed using Reactome database for the genes that changed significantly. The results are given in Table 3.2.

Table 3.2: The top 15 most significant pathways obtained from the PEA.

Pathway Names	Adjusted p-value
Cell Cycle, Mitotic	1.99×10^{-14}
Cell Cycle	1.99×10^{-14}
Cell Cycle Checkpoints	1.99×10^{-14}
Mitotic Prometaphase	3.21×10^{-12}
M Phase	7.25×10^{-11}
Resolution of Sister Chromatid Cohesion	1.56×10^{-10}
Mitotic Spindle Checkpoint	5.44×10^{-10}
Signaling by Rho GTPases	1.16×10^{-7}
Mitotic Metaphase and Anaphase	1.73×10^{-7}
Separation of Sister Chromatids	2.65×10^{-7}
Mitotic G1-G1/S phases	2.83×10^{-7}
Mitotic Anaphase	6.57×10^{-7}
Chromosome Maintenance	8.65×10^{-6}
G1/S Transition	9.24×10^{-6}
G0 and Early G1	9.24×10^{-6}

The results of PEA and GO analysis show that cell cycle, DNA repair mechanism, gene expression and DNA replication pathways are mostly affected between cancer and dormancy states based on the significantly changed genes. These pathways are expected to be affected in dormancy since there is no growth and proliferation in the dormant state and the overall metabolism is slowed down. Due to the promising results by DGEA, the dataset was further investigated using bioinformatics analyses based on metabolic networks. Two different approaches were used to map the dataset on metabolic networks: (i) Reporter Pathway Analysis, which enables identification of perturbed pathways by mapping significantly changed genes on reactions in the genome-scale human metabolic network, (ii) Flux Analysis, which allows the calculation of rates of reactions involved in metabolism by mapping transcriptome data on the genome-scale metabolic network. Both analyses are detailed in the following sections.

3.2. Reporter Pathway Analysis of the Transcriptome Data

First of all, Reporter Pathway Analysis (RPA) was performed in order to gain an overview of the dataset in terms of the affected metabolic pathways in dormancy. For

this analysis, p-values were obtained by limma as explained in the previous section. Then, the analysis was performed for both cell lines using HumanCyc as the genome-scale metabolic network as input as well as the list of gene symbols and p-values. HumanCyc, which includes 2887 reactions, 1929 metabolites and 3255 enzymes in human, is a manually curated metabolic network database. These reactions are grouped under 314 pathways [Trupp et al., 2010]. As a result of the reporter pathway analysis, pathway names that change significantly in both cell lines for a p-value cut-off value of 0.05 are given in Table 3.3.

Table 3.3: Top 20 most significant pathways obtained by RPA with a cut-off value of 0.05.

Pathway Names	p-value in j82 cell line	p-value in jbv cell line
Fructose 2,6-bisphosphate synthesis	0.02471	0.00002
Inositol pyrophosphates biosynthesis	0.01665	0.00002
Adenine and adenosine salvage II	0.01994	0.00005
Palmitate biosynthesis	0.00016	0.00006
Pyrimidine deoxyribonucleotides de novo biosynthesis	0.00209	0.00013
D-myo-inositol-5-phosphate metabolism	0.00177	0.00018
3-phosphoinositide degradation	0.00001	0.00043
3-phosphoinositide biosynthesis	0.00009	0.00092
Adenosine deoxyribonucleotides de novo biosynthesis	0.02799	0.00108
D-myo-inositol (1,4,5)-trisphosphate biosynthesis	0.00119	0.00121
Pyrimidine deoxyribonucleotides biosynthesis from CTP	0.00691	0.00511
D-myo-inositol (1,3,4)-trisphosphate biosynthesis	0.02212	0.00695
Heparan sulfate biosynthesis (late stages)	0.00000	0.01130
CMP phosphorylation	0.00982	0.01154
Acetone degradation I (to methylglyoxal)	0.03460	0.01594
[2Fe-2S] iron-sulfur cluster biosynthesis	0.00393	0.01655
Hydrogen sulfide biosynthesis (trans-sulfuration)	0.00545	0.02355
Thio-molybdenum cofactor biosynthesis	0.01415	0.02464
UTP and CTP de novo biosynthesis	0.04952	0.03264
Methionine salvage	0.04672	0.04654

In general, the results show that fatty acid synthesis, synthesis of secondary metabolites and pathways interacting with certain signaling pathways are affected between dormancy and cancer status. For example, the presence of pathways related to inositol [Chakraborty, 2018], which is involved in many important functions such as negative regulation of cell growth and interferon alpha / beta signaling, indicates that dormant cells differ from cancer in many pathways.

3.3. Updates on Genome-scale Metabolic Network Models

Human genome-scale metabolic network models developed by two different research groups were selected to be used in this thesis study comparatively. One of the models is Recon2.04 version, obtained by updating Recon2 and available in COBRA toolbox. The iHsa model was derived from HMR2, which is an updated version of HMR (Human Metabolic Reactions) model. Compared with other genome-scale metabolic models, iHsa model has higher number of reactions, metabolites, genes and gene rules. 169 new reactions, 1,103 manually reconciled GPR relationship rules and over 5,000 additional references to experimental literature and annotation databases were added to HMR model manually to form iHsa model [Blais et al. 2017]. In summary, while Recon2.04 model includes 2140 genes, 7440 reactions and 5063 metabolites [Thiele et al., 2013], iHsa model includes 2315 genes, 8264 reactions and 5620 metabolites [Blais et al., 2017].

3.3.1. Biomass Reaction and ATP Maintenance

The models need to be modified/updated to make them ready to use for the dormancy-specific and cancer-specific transcriptome data. Since cancer cells proliferate, a reaction to account for the synthesis of macromolecules necessary for the proliferation is needed, called biomass reaction. The biomass reaction preferred in cancer-specific genome-scale metabolic models generated by Nielsen and his team for 5 different cancer types was chosen for this study [Gatto et al. 2014]. This biomass reaction was derived from the biomass reaction in the first cancer specific genome scale metabolic model [Folger et al., 2011]. The coefficients of precursors of biomass are derived assuming that metabolic fluxes are measured in mmol/gDW/ h⁻¹. The

biomass compositions and coefficients are presented in Table 3.4. The same biomass reaction was considered in both Recon2.04 and iHsa models to be used in this study. The consumption of the metabolite 'Tyr-194 of apo-glycogenin' was added to Recon2.04 as an exchange reaction since it was required for biomass formation, and there was no reaction in the model to produce this metabolite [Hadi and Marashi 2014].

Apart from the biomass composition and network structure, there are two parameters that define the maximum in silico growth efficiency, Growth-Associated ATP Maintenance (GAM) and non- Growth-Associated ATP Maintenance (NGAM). Both are associated with non-metabolic energetic expenditure of the living cell. GAM accounts for the energy (ATP) necessary to replicate a cell. GAM is the energy (ATP) expenditure required for non-metabolic activities accompanying the synthesis of a biomass unit. NGAM is the energy (ATP) expenditure that the cell uses to maintain vital life functions other than growth. GAM is specified as ATP consumption within the biomass reaction. It was not considered in the biomass reaction obtained from the literature. Therefore, ATP was added to the biomass reaction with a coefficient value of 24 mmol / gram dry weight [Killburn et al., 1969] to account for the energy need for the growth-associated cell maintenance. Also, the reaction, known as NGAM reaction is added as Equation 3.1 into the models. The rate of this reaction was constrained as 1 mmole / gram dryweight / hour [Kilburn et al., 1969].

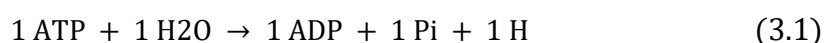


Table 3.4: Biomass Composition used in the cancer-specific metabolic models. Coefficients are in mmol/gram dry weight.

Metabolite Name	Coefficient	Metabolite Name	Coefficient
1,2-diacylglycerol-LD-TAG pool	0.00347	Lysine	0.03283
1-acylglycerol-3P-LD-TG1 pool	0.00489	Methionine	0.00657
2-lysolecithin pool	0.00354	PC-LD pool	0.03407
Alanine	0.42021	PE-LD pool	0.03384
AMP	0.01963	Phenylalanine	0.01313

Table 3.4: Continued.

Metabolite Name	Coefficient	Metabolite Name	Coefficient
Arginine	0.00328	phosphatidate-LD-TAG pool	0.02452
Asparagine	0.04268	PI pool	0.00646
Aspartate	2.4523	Proline	0.06369
Cholesterol	0.03755	PS-LD pool	0.01087
cholesterol-ester pool	0.01925	Serine	0.13132
CL pool	0.00077	SM pool	0.00706
CMP	0.03272	TAG-LD pool	0.02525
Cysteine	0.00427	Threonine	0.07222
dAMP	0.00805	Tryptophan	0.00328
dCMP	0.00537	Tyrosine	0.0197
dGMP	0.00537	UMP	0.01963
dTMP	0.00805	Valine	0.04268
Glutamate	0.59421	Glycogen	0.93989
Glutamine	0.67628	Cancer_biomass	1
Glycine	0.49244	'H+'	24
GMP	0.03709	'Pi'	24
Histidine	0.10177	'H ₂ O'	24
Isoleucine	0.01313	'ATP'	24
Leucine	0.0394	'ADP'	24

3.3.2. Medium Composition

To simulate the cell lines cultured in the Dulbecco's Modified Eagle's Medium (DMEM) and Fetal Calf Serum (FCS) media, some constraints in the models are modified [Asgari et al., 2015]. Therefore, the compositions of the media were checked in detail to allow for the consumption of the substances contained in the medium for the simulations. Firstly, all consumption reactions were inactivated in both models. Then, the composition of DMEM and FCS media were analyzed, and the consumption lists of 52 metabolites for Recon2 and iHsa were created. The full composition of DMEM and FCS media are given in Appendix A (Table A1.1-2). The consumption

lists contain glucose, oxygen, amino acids, vitamins and essential components. The medium compositions are presented in Table 3.5.

Determining the uptake rate constraints of the consumption metabolites was the next step. In a study, glucose consumption rate is measured from 59 cancer cell lines, and recorded [Zielinski et al. 2017]. In this thesis study, the glucose uptake rate was fixed to the mean of the 59 experimental values for cancer state, which is 0.81 mmol / gDW / hour [Zielinski et al. 2017]. In contrast to the cancer state, no data is available in literature for the glucose uptake value in the dormant state. Therefore, a glucose uptake ratio between cancer and dormant status was calculated based on the gene expression levels in the transcriptomic data. Genes catalyzing the reaction using glucose by converting it to glucose-6-phosphate in the glycolysis pathway were used to compare glucose uptake rate in cancer and dormancy. These genes (HK1, HK2, HK3 and GCK) show an increased expression in dormant state except GCK gene. Therefore, only GCK gene ratio was used as it had a decreasing profile in both cell lines, in agreement with the expected decrease in glucose uptake in dormant cells. The ratio of mean expression value of GCK gene between the two conditions is about 1.125 cancer over dormancy. Therefore, the glucose uptake rate of dormant state was determined using this ratio and obtained as 0.72 mmol / gDW / hour for dormant state.

Asgari et al. (2015) derived the ratio of each metabolite uptake rate to the glucose uptake rate from the medium content to use as a constraint for the uptake values of glucose and other metabolites to simulate the medium in their study [Asgari et al., 2015]. In this study, a similar approach was used for simulation. Maximum consumption rate limit was entered for each of these metabolites such that it is 1/10 of glucose consumption rate for amino acids and basic components (calcium, chloride, Fe²⁺, K⁺, sodium, ammonia, bicarbonate, phosphate and sulfate) and 1/100 of glucose consumption rate for vitamins and basic FCS contents (cholesterol ester, triacylglycerol, urea, urate, cortisol, triiodothyronine, testosterone and progesterone). Finally, only these metabolites were activated in both models. As an exception to the given rates, the maximum consumption rate of glutamine was limited to 1/4 of glucose due to high glutamine uptake rate in cancer cells [Asgari et al., 2015]. Also the maximum consumption rate for oxygen was constrained to be the same as glucose uptake rate since the measured oxygen uptake rates of cancer cell lines [Zielinski referansı buraya] were lower than the glucose uptake rates in all cancer cell lines [Zielinski et al., 2017]. In addition, although uptake rates were constrained by using

lower bound values for metabolites in exchange reactions, these metabolites were allowed to be secreted by using upper bound value. For example, the lower bound value for water was set to -1, and the upper bound value for water was set to 1000. Thus, unlimited secretion of water was allowed while limiting its uptake rate.

Table 3.5: Medium Compositions used in the simulations, derived from compositions of DMEM and FCS media.

Metabolite Names		
D-Glucose	L-serine	Ammonia
(R)-Pantothenate	L-Arginine	O ₂
Biotin	L-Aspartate	Phosphate
Choline	L-Glutamate	Sulfate
Folate	L-Isoleucine	H ₂ O
Pyridoxal	L-Leucine	Bicarbonate
Riboflavin	L-Lysine	Cholesteral
Thiamin	L-Methionine	Triacylglycerol
myo-Inositol	L-Proline	Urea
Nicotinamide	L-Threonine	Urate
Reduced glutathione	L-Tryptophan	Cortisol
Glycine	L-Tyrosine	Triiodothyronine
L-alanine	L-Valine	Testosterone
L-asparagine	Calcium	Progesterone
L-cysteine	Chloride	
L-glutamine	Fe ⁺²	
L-histidine	K ⁺	
L-phenylalanine	Sodium	

3.4. FBA of the Genome Scale Metabolic Networks

Subsequently, Flux Balance Analysis (FBA) was performed for both models to predict the phenotypes of the cancer and dormancy metabolisms by modeling. FBA is an effective method for determining which of the reactions that make up the intracellular metabolic network are conditionally active and how the pathways involved in cell metabolism change [Blazier and Papin 2012]. Maximizing biomass production has been successfully implemented as a reliable objective function for FBA in predicting flux distribution in various microorganisms [Varma and Palsson, 1994; Feist and Palsson, 2010]. Although biomass yield for microorganisms is reasonable as

a cellular objective, the objective of human or plant cells is much more complex and tissue-specific. Therefore, it is not easy to determine the objective of human cells. On the other hand, biomass production can be used as a candidate cellular objective for all cell types to reflect the cellular need to continuously regenerate metabolites for the synthesis of biosynthetic precursors to maintain the cellular functions such as lipid, protein, carbohydrate and RNA synthesis [Wagner et al., 2013]. The observation that cancer cells adapt their metabolism to achieve proliferation has been used successfully in previous studies as an objective function to calculate rates to predict characteristic changes in cancer metabolism [Price et al., 2004]. Therefore, by considering the tendency of tumor cells to proliferate, the maximization of the biomass (i.e growth rate) as an objective function was used in this study in the optimization-based determination of reaction rates to predict metabolic fluxes. The flux distributions of the reactions were calculated by FBA using the constraints explained in section 3.3.2. Cell proliferation rate was calculated as 0.0553 h⁻¹ in the Recon2.04 model and 0.0326 h⁻¹ in the iHsa model. The average growth rate of all cancer cell lines by Zielinski et al. [Zielinski et al., 2017] is 0.021 h⁻¹, with minimum value of growth rate being 0.0075 h⁻¹ and maximum value being 0.039 h⁻¹ among 59 cell lines. The growth rate predicted by iHsa model is within the reported range, implying a better prediction over Recon2.04 model.

In an optimization problem, there are many solutions that can have the same optimum value, which is called alternate optima. Therefore, additional objective functions can be added to the optimization framework to eliminate alternate optima. [Çakır et al., 2014]. Also, it is assumed that a single objective function cannot comprehensively explain the metabolic behavior of the cell. Therefore, a dual objective function framework has been used to narrow the solution space and achieve better results [Cakir et al., 2007; Tarlak et al., 2014]. Due to the problem of alternate optima, the system was subjected to a second optimization in order to eliminate the multiplicity of flux values. Because of the tendency to use minimum (optimum) amount of enzymes in biological systems to achieve cellular objectives [REF], it is biologically more meaningful to choose the solution having the minimum total flux value among alternate optimal solutions. This approach is mathematically represented as the minimization of the squared sum of all fluxes. The minimization of the squared sum of all fluxes can be used as a secondary objective function to eliminate alternative solutions with the same optimal value [Holzhütter, 2004]. The primary objective value

obtained is determined as a constraint, and then the secondary optimization, which is the minimization of the sum of the squares of all fluxes, is applied. In this case, a more reliable flux distribution with the same growth rate can be obtained to obtain optimal solution in biological system. This dual-objective approach was also used in calculating flux distributions with the metabolic models obtained with transcriptome data integration.

3.5. Transcriptomic Data Integration

Transcriptomic data can be integrated into the genome-scale metabolic models for different purposes such as characterizing condition-specific metabolism, finding drug targets or molecular markers for these conditions and providing additional-constraints for FBA simulation (to minimize the solution space for condition specific model). As previously detailed, GIMME [Aurich and Thiele 2016] and E-Flux [Colijin et al. 2009] algorithms were applied to the updated models in terms of biomass reaction and medium composition. In this section, both cancer and dormancy gene expression data of transcriptomic data with accession number GSE35947 were used. Since there are two different cell lines (jbv and j82) and two states (cancer and dormancy) in the dataset, condition-specific metabolic models were created by integrating the data from each cell line and each state with the two models. This led to 8 different flux distribution results for each integration method.

The GIMME algorithm requires two parameters: the threshold value parameter (genes with expression less than this threshold will be considered inactive), and the percentage reduction allowed in the objective function. Since the objective function in this thesis study is the maximization of tumor growth rate (biomass reaction), this corresponds to the allowed decrease in the growth rate in terms of the percentage of the maximum possible growth rate when the inactive gene-related reactions are removed from the model. For these parameters, the average of expression values of all genes in cancer and dormancy state was used as the gene expression threshold, while the limit of the objective function was allowed to be reduced by 50% in the case of cancer and 95% in the case of dormancy. The percent reduction allowed in the dormancy state is much higher because it is known that the proliferation rate of dormant cells is close to zero. As a result, 8 specific models were obtained by using

GIMME algorithm (Figure 3.6). Using the gene expression values in GSE35947 dataset, GIMME algorithm removed about 2580 reactions from Recon2 model and about 2780 reactions from the iHsa model. In order to compare the predicted flux values of the reactions between the two conditions (cancer and dormancy), common reactions were determined in cancer and dormancy models obtained by GIMME algorithm.

Table 3.6: Dimensions of the models obtained by GIMME algorithm.

Model Names	Reactions	Metabolites
Recon2.04 cancer (jbv)	4838	4113
Recon2.04 dormant (jbv)	4888	4213
Recon2.04 cancer (j82)	4832	4113
Recon2.04 dormant (j82)	4887	4210
iHsa cancer (jbv)	5562	4772
iHsa dormant (jbv)	5647	4786
iHsa cancer (j82)	5186	4579
iHsa dormant (j82)	5534	4756

E-flux does not have a preprocessing step to eliminate inactive reactions. The consumption rates based on the medium composition cannot be specified in absolute values in E-flux. Therefore, the ratio of them against each other is constrained to be constant. For example, the ratio between the glucose consumption rate and the amino acid consumption rates to be calculated by the model should be 0.10 at maximum, which is entered as an inequality constraint to the model. To achieve this, new linear equations defining these ratios are defined and added to the S matrix. Thus, these models were modified by adding new linear equations defined in Equation 3.2 a-b. In the equation, V1 indicates glucose uptake rate, V2 indicates uptake rate for another metabolite, and the coefficient n indicates the ratio.

$$\frac{V1}{V2} < n \quad (3.2a)$$

$$V1 - (n)V2 < 0 \quad (3.2b)$$

The flux distribution of the model obtained as a result of E-flux mapping should be scaled using the known flux value of glucose uptake rate to compare it with

experimental flux data. In this study, the scaling value of 0.81 was used for the cancer state and 0.72 for the dormant state.

For both algorithms, maximum biomass production and minimum enzyme usage were used as objective functions. In this case, the use of two different cell lines (jbv and j82), two different conditions (cancer and dormancy), two different metabolic models (Recon 2.04 and iHsa) and two different algorithms (GIMME and E-flux) led to 16 different state-specific models. Flux distributions of the reactions were calculated using FBA and flexoFBA in these 16 models. In FlexoFBA, firstly, the growth rate is calculated by obtaining a flux distribution with the primary objective function, which is the maximum growth. Then, the optimal growth rate is multiplied with a normally distributed (left-tailed) random number with mean 1 and standard deviation 0.05. FlexoFBA was applied with 500 repetitions, with the same parameters as normal FBA simulation to take into account culture heterogeneity. Finally, the final flux distribution was obtained using the median value of 500 flux distributions. The predicted fluxes of selected reactions are shown in the Tables 3.7 - 10. A list of the original names of the selected reactions in the corresponding models is given in Appendix B (TableB1.1).

Table 3.7: Flux values in cancer and dormancy states, calculated by the Recon2.04 and GIMME integration.

Reaction Names	Cancer (jbv)	Dormant (jbv)	Cancer (j82)	Dormant (j82)
Oxygen uptake	0.6762	0.0183	0.3960	0.2962
Lactate production	0.3425	0.8317	0.1450	0.3017
PPP	0.0049	0.0002	0.0006	0.0001
TCA	0.1150	0	0	0
Biomass	0.0546	0.0068	0.0546	0.0179
Flexo-FBA				
Oxygen uptake	0.6471	0.0206	0.3911	0.2819
Lactate production	0.3367	0.8262	0.1585	0.3170
PPP	0.0003	0	0.0002	0.0003
TCA	0.1076	0	0	0
Biomass	0.0528	0.0066	0.0528	0.0174
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.				

Table 3.8: Flux values in cancer and dormancy states, calculated by the Recon2.04 and E-Flux integration.

Reaction Names	Cancer (jbv)	Dormant (jbv)	Cancer (j82)	Dormant (j82)
Oxygen uptake	0.3565	0.2740	0.2907	0.3236
Lactate production	0.2452	0.2347	0.2102	0.1731
PPP	0.00003	0.00002	0.00002	0.00001
TCA	0	0	0	0
Biomass	0.0403	0.0336	0.0413	0.0376
Flexo-FBA				
Oxygen uptake	0.3604	0.2764	0.2941	0.3204
Lactate production	0.2391	0.2325	0.2046	0.1705
PPP	0.00002	0.00001	0.00002	0.00002
TCA	0	0	0	0
Biomass	0.0404	0.0337	0.0415	0.0375
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.				

Table 3.9: Flux values in cancer and dormancy states, calculated by the iHsa and GIMME integration.

Reaction Names	Cancer (jbv)	Dormant (jbv)	Cancer (j82)	Dormant (j82)
Oxygen uptake	0.8099	0.7200	0.8100	0.7200
Lactate production	1.6077	1.4983	1.5956	1.7792
PPP	0	0	0	0
TCA	0.0893	0.0025	0.0837	0.0116
Biomass	0.0206	0.0075	0.0217	0.0045
Flexo-FBA				
Oxygen uptake	0.81	0.72	0.81	0.719
Lactate production	1.4831	1.4262	1.5143	1.744
PPP	0	0	0	0
TCA	0.0908	0	0.0726	0.0138
Biomass	0.0199	0.0072	0.0210	0.0043
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.				

Table 3.10: Flux values in cancer and dormancy states, calculated by the iHsa and E-Flux integration.

Reaction Names	Cancer (jbv)	Dormant (jbv)	Cancer (j82)	Dormant (j82)
Oxygen uptake	0.4664	0.4225	0.4886	0.3541
Lactate production	0.0339	0.1022	0.3814	0.2604
PPP	0.0601	0.0512	0	0
TCA	0	0	0.0199	0
Biomass	0.0323	0.0292	0.0328	0.0292
Flexo-FBA				
Oxygen uptake	0.4643	0.4182	0.4440	0.3541
Lactate production	0.0218	0.0896	0.2410	0.2604
PPP	0.0610	0.0519	0.0132	0
TCA	0	0	0	0
Biomass	0.0326	0.0292	0.0326	0.0292
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.				

A decrease in the metabolic rates of dormancy is expected in the light of the information mentioned in the literature and the results obtained from the basic bioinformatics analyses in the previous section. The results of 2 different models and 2 different algorithms predict a decrease in biomass production for dormant cells. However, the results show a more noticeable decrease in dormant state in GIMME simulations and no significant decrease in E-flux simulations. Therefore, tumor growth rates predicted by GIMME results were acceptable (much lower growth rate in dormant state), but E-flux results did not yield acceptable results in this respect. Therefore, the GIMME algorithm worked better in predicting phenotypic state. A more detailed comparison of GIMME results is given in the following sections to interpret the metabolism of dormancy. The results of the Recon2 model show some differences in important fluxes for two cell lines except biomass reaction. In contrast, the iHsa model does not have a significant difference between cell lines. Model comparison shows that flux values are different from each other, especially lactate production has opposite profile. In the dormant state, lactate production decreases in the iHsa model and increases in the Recon 2.04 model.

In addition, the calculated flux values by flexoFBA approach are very close to classical FBA results for both models. There are only few differences. Comparing the classical FBA and flexoFBA results for the two cell lines in the Recon2.04 model, the

j82 cell line results show opposite profile in PPP flux. In dormant state, PPP reaction rate decreases in classical FBA results, and increases in flexoFBA results. Since flexoFBA results did not differ significantly from classical FBA results, flexoFBA-based flux distributions were not further analyzed in the following sections.

3.6. Detailed Comparison of Flux Distributions

In the obtained models, it is not sufficient to interpret only the flux values of the important reactions to interpret the metabolism of dormancy. Therefore, two approaches were used to compare 8 different flux distributions obtained by using GIMME: (i) determining affected pathways between cancer and dormancy by comparing the pathway activities, (ii) comparing rates of reactions individually between the two conditions to identify genes encoding these reactions. The two approaches are detailed below, and the results are given.

3.6.1. Comparison of Pathway Activity

When the current models for cancer and dormancy are compared in terms of flux distributions, there are 4 different comparison cases, which are (i) JBV cell line data with the Recon2.04 model, (ii) J82 cell line data with the Recon2.04 model, (iii) JBV cell line data with the iHsa model and (iv) J82 cell line data with the iHsa model. For comparison of the pathway activities, the absolute values of the fluxes of the reactions associated with each pathway were summed and recorded as the activity of that pathway. This was performed separately for flux distributions of cancer and dormancy states. While 101 pathway names (Table C1.1) are defined in Recon2.04, 46 pathway names (Table C1.2) are defined in the iHsa model. The activity value assigned to each pathway for cancer and dormancy was compared to obtain a list of significantly affected pathways. Here, simple the ratios of the pathway activity scores were calculated. A cut-off value of 2 was used for the fold change and a minimum total flux value of 0.05 was used for pathway activity. The pathway activities and their fold changes calculated for the models obtained by mapping jbv and j82 cell line samples on both Recon2.04 and iHsa model by GIMME algorithm are given in Tables 3.11-3.14.

Table 3.11: Integration of JBV cell line data with the Recon2.04 model using GIMME algorithm. Bold text indicates increased activity.

Pathway Name	Total Cancer Fluxes	Total Dormant Fluxes	Fold Change
Transport, mitochondrial	9.931	0.196	50.67
Transport, endoplasmic reticular	0.227	0.089	2.55
Tyrosine metabolism	0.080	0.005	16.00
beta-Alanine metabolism	0.204	0.001	204.00
Transport, peroxisomal	0.415	0.026	15.96
Aminosugar metabolism	0.234	0	-
Citric acid cycle	1.709	0.064	26.70
Glycolysis/gluconeogenesis	5.012	10.173	0.49
Nucleotide interconversion	5.475	0.766	7.15
Urea cycle	0.688	0.006	114.67
Alanine and aspartate metabolism	1.413	0.031	45.58
Starch and sucrose metabolism	0.103	0.013	7.92
Cholesterol metabolism	0.103	0.048	2.15
Oxidative phosphorylation	0.934	0.004	233.50
Fructose and mannose metabolism	0.364	0.108	3.37
Glyoxylate and dicarboxylate metabolism	0.065	0.003	21.67
Glutamate metabolism	0.338	0.022	15.36
Glutathione metabolism	4.395	0.062	70.89
NAD metabolism	0.024	0.079	0.30
Biomass	0.055	0.007	7.86
Uptake Tyr-194 of apo-glycogenin protein	0.051	0.006	8.50
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.			

Table 3.12: Integration of J82 cell line data with the Recon2.04 model using GIMME algorithm. Bold text indicates increased activity.

Pathway Name	Total Cancer Fluxes	Total Dormant Fluxes	Fold Change
Transport, mitochondrial	11.599	3.371	3.44
Pyrimidine catabolism	0.051	0.125	0.41
Transport, nuclear	0.027	0.218	0.12
Valine, leucine, and isoleucine metabolism	2.105	0.218	9.66
Aminosugar metabolism	0.079	0.327	0.24
Citric acid cycle	0.213	0.574	0.37
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.			

Table 3.12: Continued.

Pathway Name	Total Cancer Fluxes	Total Dormant Fluxes	Fold Change
Nucleotide interconversion	2.276	1.013	2.25
Arginine and Proline Metabolism	0.051	0.121	0.42
Pyruvate metabolism	0.316	0.876	0.36
Starch and sucrose metabolism	0.103	0.034	3.03
Oxidative phosphorylation	1.136	0.321	3.54
Fructose and mannose metabolism	1.625	0.733	2.22
Galactose metabolism	0.564	0.185	3.05
Glutathione metabolism	2.708	0.940	2.88
Bile acid synthesis	0.102	0.235	0.43
NAD metabolism	0.506	0.209	2.42
Biomass	0.055	0.018	3.06
Glycolysis/gluconeogenesis	3.907	7.535	0.52
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.			

The pathway activities in Recon2 model in Tables 3.11 and 3.12 show that there are similar activity changes in two cell lines except for a few pathways. For example, while glycolysis/gluconeogenesis and oxidative phosphorylation pathways have the same profile in the two cell lines for dormant state, citric acid cycle and NAD metabolism pathways have opposite profiles in the two cell lines for dormant state. The common ones provide a perspective for the dormant state, while the different ones may suggest different cell-specific dormant models. In general, when looking at pathway activities, in jbv cell line simulations, activities have a decreasing profile except for a few pathways, whereas in j82 cell line the situation is more variable. The analysis of pathway activities in Recon2.04 simulations show that jbv cell line is more reliable for the interpretation of metabolism in dormant state since a decrease in metabolic activities is expected for dormancy.

Table 3.13: Integration of JBV cell line data with the iHsa model using GIMME algorithm. *Bold text indicates increased activity.

Pathway Name	Total Cancer Fluxes	Total Dormant Fluxes	Fold Change
Fatty acid metabolism	2.038	0.876	2.33
Carbohydrate metabolism	0.403	0.138	2.92
Vitamin A metabolism	0.276	0.058	4.76
Glycerolipid metabolism	0.343	0.074	4.64
Vitamin B metabolism	1.297	0.289	4.49
Biomass	0.121	0.054	2.24
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.			

Table 3.14: Integration of J82cell line data with the iHsa model using GIMME algorithm. Bold text indicates increased activity.

Pathway Name	Total Cancer Fluxes	Total Dormant Fluxes	Fold Change
Fatty acid metabolism	2.18	0.54	4.04
Carbohydrate metabolism	0.06	0.34	0.18
Aromatic amino acid metabolism	0.00	0.09	0
Vitamin B metabolism	2.85	0.97	2.94
Biomass	0.09	0.02	4.5
Beta oxidation of fatty acids	0.07	0.03	2.33
*The rate of metabolic reaction flux is mmol / gDW / hour, and growth rate unit is 1 / h.			

Similarly, pathway activities were calculated using flux distributions obtained by iHsa model. For example, while fatty acid metabolism and vitamin B metabolism pathways have the same profile in the two cell lines for dormant state, carbohydrate metabolism pathway has an opposite profile in the two cell lines for dormant state. The focus should be on pathways with excessive change in activity.

According to the pathway activity results, it is possible to comment on which pathways change in the state of dormancy compared to the cancer status. Since cancer dormancy is a newly discovered phenomenon, there is no clear information in the literature to compare the results obtained in this study. In order to make sense of these results, a literature review was conducted on the state of dormancy. Normal cells are assumed to be in the dormant phase when they enter the G0 phase of the cell cycle and have low metabolism [Aguirre-Ghiso, 2007]. When the pathway activities are

analyzed in the thesis, it is seen that there is a decrease in the activity in the transition from cancer to dormant state (Table 3.11- 3.14).

Quiescence and cancer stem cell metabolism are estimated to be similar in some cases with dormancy metabolism [Cheung and Rando, 2013]. Quiescence is a reversible state in which the cell is undivided and thus retains its ability to re-enter cell proliferation [Sagot and Laporte, 2018]. Cancer Stem Cells (CSC) are a small population of cells within the tumor, with some properties such as self-regeneration, differentiation and tumorigenicity [Yu et al., 2012]. These two cellular conditions are better known and studied to understand dormant state. In the one study performed on lymphocytes, it was observed that glucose consumption and general metabolic rates decreased in quiescence. Dormancy can be thought as a long-term quiescence. Dormancy can also be defined as rare division and low metabolism compared to more frequently divided strains [Rocheteau et al. 2015]. The 2012 study reviewed metabolic changes during quiescence in multiple model systems: bacteria, yeast, lymphocytes, hematopoietic stem cells, fibroblasts, and cancer cells [Valcourt et al., 2012]. The study states that quiescence is generally associated with reduced metabolism. However, it is also possible to achieve a state of cellular quiescence in which glucose uptake and glycolysis flux are not reduced. The comparison in the study shows that, in fibroblasts unlike in lymphocytes, the quiescence state is associated with less oxidative phosphorylation and with a shift towards increased glycolysis. Also, the quiescent state is characterized by an increase in glycolysis enzymes and lower production of TCA enzymes in bacteria [Valcourt et al., 2012]. The results of the Recon2.04 simulations indicate that glycolysis activity increases in dormant state for the two cell lines (Table 3.11-12). Accordingly, the increase in glycolysis activity strengthens the similarity between dormant status and quiescence.

In the other study conducted on skeletal muscle stem cells, a hypothesis about the metabolism of quiescence was presented. According to the study, the cells decrease mitochondrial activity in dormant state and increase their glycolysis activity [Rocheteau et al. 2015]. The study associates dormancy with low metabolic activity and anaerobic metabolism. Predicted pathway activity changes point to an increase in the NAD metabolism in dormancy, and a decrease in oxidative phosphorylation (Table 3.11 - 3.14). The results obtained in this study are consistent with decreased mitochondrial activity and increased glycolysis in the studies focused above.

Furthermore, the intersection between dormancy and the cancer stem cell models may offer insight into the underlying biology in both realms but is poorly characterized in terms of metabolism.

3.6.2. Functional Analysis of Genes Catalyzing the Affected Reactions

In addition to focusing on the comparison of pathways, the reactions were also subjected to a comparison between the two conditions. In this comparison, reactions with absolute flux values greater than 0.01 mmol/gDW/hour were used, and a cut-off value of 2 was used for fold change between fluxes.

Considering these parameters, the number of affected reactions and the number of genes catalyzing these reactions are given in Table 3.15 for each of the 8 specific models obtained with GIMME.

Table 3.15: Number of affected reactions and number of genes that catalyze affected reactions.

Models	Reactions	Genes
Recon2.04 (jbv)	280	297
Recon2.04 (j82)	306	280
iHsa (jbv)	343	371
iHsa (j82)	326	330

Gene Ontology (GO) analysis and Pathway Enrichment Analysis (PEA) were performed in order to identify the functions of the genes that are predicted to be affected as a result of the comparisons. Thus, it is possible to comment on which biological processes are affected in the dormant state. g:Profiler was used for GO analysis using biological process section of GO annotations. REACTOME was used for PEA by using 2272 pathway names available in its database for *Homo sapiens*. In GO analysis, functional analysis is usually performed by using all genes in the genome as background. However, when focusing only on the metabolism in a study, it may not be possible to obtain a completely reliable result for the genes analyzed by using the whole genome as background. It is expected that the terms related to metabolism will be obtained when the genes obtained from the metabolic model are analyzed using the

whole genome. In this case, the significance of the results is a question mark. Therefore, an important point in the GO analysis for this study is that only the genes covered in the metabolic models are loaded into the database as background. Thus, the metabolic pathways affected by the genes analyzed can be understood more clearly. For this purpose, both the list of all genes found in the models and the list of obtained genes from the analyses were entered into the g:Profiler tool as inputs. The GO analysis and PEA results obtained for the models generated by mapping transcriptome data on the Recon2.04 model by using GIMME algorithm are shown in Table 3.16-19, and on the iHsa model by using GIMME algorithm are shown in Table 3.20-23.

Table 3.16: GO Analysis results for 297 genes predicted to be affected between cancer and dormancy (jbv cell line and Recon2.04 integration). The most significant 20 GO terms are given.

Biological process	Adjust p-value
Generation of precursor metabolites and energy	5.69×10^{-39}
Respiratory electron transport	3.63×10^{-35}
Nonalcoholic fatty liver disease	5.71×10^{-34}
Non-alcoholic fatty liver disease (NAFLD)	4.71×10^{-33}
ATP metabolic process	6.71×10^{-33}
Parkinson disease	1.71×10^{-32}
The citric acid (TCA) cycle and respiratory electron transport	2.71×10^{-29}
ATP synthesis coupled electron transport	6.71×10^{-29}
Mitochondrial ATP synthesis coupled electron transport	6.71×10^{-29}
Respiratory electron transport chain	1.71×10^{-27}
Huntington disease	2.71×10^{-27}
Cellular respiration	2.71×10^{-27}
Energy derivation by oxidation of organic compounds	2.71×10^{-27}
Electron Transport Chain (OXPHOS system in mitochondria)	9.71×10^{-25}
Electron transport chain	5.71×10^{-24}
Alzheimer disease	2.71×10^{-23}
Oxidative phosphorylation	1.71×10^{-22}
Thermogenesis	1.71×10^{-19}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	3.71×10^{-19}
mitochondrial respiratory chain complex assembly	1.08×10^{-15}

Table 3.17: GO Analysis results for 280 genes predicted to be affected between cancer and dormancy (j82 cell line and Recon2.04 integration). The most significant 20 GO terms are given.

Biological process	Adjust p-value
Respiratory electron transport	3.18×10^{-36}
Non-alcoholic fatty liver disease (NAFLD)	3.87×10^{-36}
Nonalcoholic fatty liver disease	4.00×10^{-36}
Parkinson disease	6.12×10^{-36}
Electron Transport Chain (OXPHOS system in mitochondria)	3.94×10^{-34}
Huntington disease	5.59×10^{-32}
ATP metabolic process	5.63×10^{-32}
The citric acid (TCA) cycle and respiratory electron transport	9.27×10^{-31}
ATP synthesis coupled electron transport	1.73×10^{-29}
Mitochondrial ATP synthesis coupled electron transport	1.73×10^{-29}
Respiratory electron transport chain	2.25×10^{-29}
Generation of precursor metabolites and energy	6.99×10^{-29}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	7.12×10^{-29}
Alzheimer disease	2.28×10^{-27}
Electron transport chain	4.77×10^{-24}
Thermogenesis	9.05×10^{-24}
Energy derivation by oxidation of organic compounds	7.72×10^{-23}
Oxidative phosphorylation	7.72×10^{-23}
Cellular respiration	1.13×10^{-20}
Mitochondrial respiratory chain complex assembly	7.69×10^{-18}

Table 3.18: GO Analysis results for 371 genes predicted to be affected between cancer and dormancy (jbv cell line and iHsa integration). The most significant 20 GO terms are given.

Biological process	Adjust p-value
ATP metabolic process	4.82×10^{-21}
Thermogenesis	1.35×10^{-17}
Parkinson disease	1.91×10^{-15}
Electron Transport Chain (OXPHOS system in mitochondria)	7.61×10^{-14}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	2.18×10^{-11}
Oxidative phosphorylation	1.71×10^{-9}
Generation of precursor metabolites and energy	1.86×10^{-8}
The citric acid (TCA) cycle and respiratory electron transport	2.69×10^{-8}
Glycolysis / gluconeogenesis	4.69×10^{-7}
Nucleotide phosphorylation	3.14×10^{-6}
Nucleoside diphosphate phosphorylation	4.03×10^{-6}
Purine ribonucleoside triphosphate biosynthetic process	2.52×10^{-5}
Purine ribonucleoside triphosphate metabolic process	2.52×10^{-5}
Nucleoside diphosphate metabolic process	3.20×10^{-5}
Purine nucleoside triphosphate biosynthetic process	1.15×10^{-4}
Respiratory electron transport	1.28×10^{-4}
Nucleoside triphosphate biosynthetic process	1.78×10^{-4}
Ribonucleoside triphosphate biosynthetic process	7.41×10^{-4}
Ribonucleotide metabolic process	3.38×10^{-3}
Fatty acid metabolic process	1.17×10^{-3}

Table 3.19: GO Analysis results for 330 genes predicted to be affected between cancer and dormancy (j82 cell line and iHsa integration). The most significant 20 GO terms are given.

Biological process	Adjust p-value
ATP metabolic process	3.9×10^{-25}
Thermogenesis	3.1×10^{-15}
Electron Transport Chain (OXPHOS system in mitochondria)	3.2×10^{-15}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	4.9×10^{-15}
The citric acid (TCA) cycle and respiratory electron transport	1.8×10^{-14}
Parkinson disease	4.7×10^{-13}
Oxidative phosphorylation	1.4×10^{-12}
Nucleoside diphosphate phosphorylation	2.3×10^{-9}
Nucleoside triphosphate biosynthetic process	1.5×10^{-7}
Nucleotide phosphorylation	1.8×10^{-7}
Purine ribonucleoside triphosphate biosynthetic process	4.2×10^{-7}
Nucleoside diphosphate metabolic process	1.9×10^{-6}
Purine ribonucleoside triphosphate metabolic process	3.4×10^{-6}
Ribonucleotide metabolic process	3.8×10^{-6}
Purine nucleotide metabolic process	4.5×10^{-6}
Respiratory electron transport	4.6×10^{-6}
Nucleoside phosphate biosynthetic process	6.8×10^{-6}
Purine ribonucleotide metabolic process	1.6×10^{-6}
Purine nucleoside triphosphate biosynthetic process	1.9×10^{-6}
Organophosphate metabolic process	4.1×10^{-5}

Coller and his team have shown that upregulated or downregulated genes that affect carbohydrate metabolism, cell cycle, metabolism, cell communication, cell proliferation, RNA metabolism, lipid metabolism, cofactor metabolism and nucleotide metabolism processes are affected in the long-term quiescence [Coller et al., 2006]. Since long-term quiescence can be approximated as dormancy [Essers and Trumpp, 2010], these processes can be checked on the metabolic profiles of dormancy. In the results of GO analysis, carbohydrates, nucleotides, fat and energy metabolism were identified to be significantly affected in dormant state (Table 3.16-3.19).

Also, GO analysis results includes processes associated with neurodegenerative diseases such as Parkinson's, Alzheimer's and Huntington's (Table 3.16-3.19). From this result, it can be concluded that the pathway related to cellular proliferation is affected in the dormant state. In neurodegenerative diseases, cell death is very common, and the opposite is expected in cancer, so it is reasonable to identify these pathways as affected. In addition, mitochondria are critical regulators of cell death, a key feature of neurodegeneration. Mutations in mitochondrial DNA and changes in oxidative reactions are effective in disease pathogenesis. Therefore, energy metabolism is the most affected pathways in these diseases [Lin and Beal, 2006]. Since mitochondrial autophagy (mitophagy) is the main destruction pathway of damaged mitochondria, autophagy is of great importance for neurodegenerative diseases. Also, autophagy is active in liver disease, neurodegeneration, Crohn's disease and metabolic syndrome [Rabinowitz and White, 2010]. Therefore, the GO results imply the presence of mitochondrial changes and autophagy in dormant state.

Table 3.20: PEA Analysis results for 297 genes predicted to be affected between cancer and dormancy (jbv cell line and Recon2.04 integration). The most significant 15 pathways are given.

Pathway Names	Adjust p-value
Complex I biogenesis	6.99×10^{-15}
Respiratory electron transport	6.99×10^{-15}
Glycolysis	6.99×10^{-15}
Glucose metabolism	6.99×10^{-15}
Pyruvate metabolism and Citric Acid (TCA) cycle	6.99×10^{-15}
The citric acid (TCA) cycle and respiratory electron transport	6.99×10^{-15}
Gluconeogenesis	6.99×10^{-15}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins,	6.99×10^{-15}
Metabolism of carbohydrates	6.99×10^{-15}
SLC-mediated transmembrane transport	6.99×10^{-15}
Metabolism	6.99×10^{-15}
Cellular hexose transport	1.90×10^{-14}
Reversible hydration of carbon dioxide	4.36×10^{-9}
Tp53 regulates metabolic genes	4.36×10^{-9}
Citric acid cycle (TCA cycle)	1.13×10^{-8}

Table 3.21: PEA Analysis results for 280 genes predicted to be affected between cancer and dormancy (j82 cell line and Recon2.04 integration). The most significant 15 pathways are given.

Pathway Names	Adjust p-value
Complex I biogenesis	6×10^{-15}
Cellular hexose transport	6×10^{-15}
Respiratory electron transport	6×10^{-15}
Glycolysis	6×10^{-15}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins,	6×10^{-15}
Glucose metabolism	6×10^{-15}
Gluconeogenesis	6×10^{-15}
The citric acid (TCA) cycle and respiratory electron transport	6×10^{-15}
Metabolism of carbohydrates	6×10^{-15}
Transport of inorganic cations/anions and amino acids/oligopeptides	6×10^{-15}
SLC-mediated transmembrane transport	6×10^{-15}
Metabolism	6×10^{-15}
Transport of small molecules	1.67×10^{-14}
Pyruvate metabolism and Citric Acid (TCA) cycle	2.07×10^{-11}
TP53 Regulates Metabolic Genes	2.04×10^{-10}

Table 3.22: PEA Analysis results for 371 genes predicted to be affected between cancer and dormancy (jbv cell line and iHsa integration). The most significant 15 pathways are given.

Pathway Names	Adjust p-value
Formation of ATP by chemiosmotic coupling	1.04×10^{-14}
Synthesis of very long-chain fatty acyl-coas	1.04×10^{-14}
Glycolysis	1.04×10^{-14}
Fatty acyl-coa biosynthesis	1.04×10^{-14}
Glucose metabolism	1.04×10^{-14}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins,	1.04×10^{-14}
The citric acid (TCA) cycle and respiratory electron transport	1.04×10^{-14}
SLC-mediated transmembrane transport	1.04×10^{-14}
Metabolism	1.04×10^{-14}
Cristae formation	1.31×10^{-13}
Triglyceride catabolism	1.12×10^{-12}
Respiratory electron transport	4.44×10^{-12}
Transport of inorganic cations/anions and amino acids/oligopeptides	1.21×10^{-11}
Triglyceride metabolism	1.72×10^{-11}
Metabolism of lipids	2.25×10^{-11}

Table 3.23: PEA Analysis results for 330 genes predicted to be affected between cancer and dormancy (j82 cell line and iHsa integration). The most significant 15 pathways are given.

Pathway Names	Adjust p-value
Formation of ATP by chemiosmotic coupling	9.99×10^{-15}
Synthesis of very long-chain fatty acyl-coas	9.99×10^{-15}
Fatty acyl-coa biosynthesis	9.99×10^{-15}
Glucose metabolism	9.99×10^{-15}
The citric acid (TCA) cycle and respiratory electron transport	9.99×10^{-15}
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins,	9.99×10^{-15}
SLC-mediated transmembrane transport	9.99×10^{-15}
Metabolism	9.99×10^{-15}
Glycolysis	1.44×10^{-14}
Cristae formation	1.44×10^{-14}
Transport of inorganic cations/anions and amino acids/oligopeptides	1.44×10^{-14}
Respiratory electron transport	3×10^{-13}
Fatty acid metabolism	2.74×10^{-12}
Pyruvate metabolism and Citric Acid (TCA) cycle	4.61×10^{-12}
Transport of small molecules	8.04×10^{-12}

In the results of PEA, carbohydrates, nucleotides, fat and energy metabolism changes in dormant state as also observed in the GO results (Table 3.20-3.23 Figure3.5-3.8). Especially, energy metabolism is more affected in the dormant state. Mitochondria are cytoplasmic organelles found in all eukaryotic cells and responsible for converting nutrients into adenosine triphosphate (ATP) for fueling cell activity. This process, known as aerobic respiration, is the reason why mitochondria are often called the power center of the cell. They also play an important role in steroid metabolism, calcium homeostasis, apoptosis and cellular proliferation [Lee et al., 2000]. Considering that cellular proliferation is minimized, a decrease in

mitochondrial activity is expected. Therefore, availability of the terms of energy metabolism in GO and PEA results encourages the reliability of the results.

In addition, based on the terms related to biomolecule synthesis such as lipid, glucose and fatty acid metabolisms, autophagy including catabolism of macromolecules can be focused. The 2012 study states that autophagy and protein turnover are critical elements of the quiescence for many quiescent model systems. Yeast, lymphocytes, hematopoietic stem cells and cancer cells all rely on autophagy to survive in their quiescent state [Valcourt et al., 2012]. In another study, it was concluded that low metabolic activity, anaerobic metabolism and active autophagy may explain the introduction of dormancy [Rocheteau et al., 2015]. Another study suggests that autophagy promotes survival of dormant cells by restricting cell macrometastasis and by providing a nutrient source [Kenific et al., 2010]. Accordingly, autophagy can be considered as active in the dormancy model, which is thought to be metabolically similar with quiescence.

In summary, dormant cells are expected to keep their metabolism at a minimal activity compared to the cancer cell. The results of this study can be used for drug target studies by focusing on the genes that control the reactions that are thought to be influenced by basic flux comparison.

3.6.3. Summary Comparisons

In summary, it can be interpreted that energy metabolism and biosynthesis pathways are the most affected in dormant state. It is shown in the summary tables (Table 3.24-25) whether the terms related to these pathways have changed significantly based on the results obtained for the cell lines and the metabolic models. In the light of the results, 10 statistically significant terms were selected from GO analysis and PEA results, based on their possible relation with dormancy-induced metabolic changes. The selected terms were then compared using the results of all models.

Table 3.24: The comparison of the significant terms in GO Analysis results for all models. The selected 10 pathways are given.

Biological Process	Recon2- jbv	Recon2- j82	iHsa- jbv	iHsa- j82
The citric acid (TCA) cycle and respiratory electron transport	+	+	+	+
Parkinson's disease	+	+	+	+
Oxidative phosphorylation	+	+	+	+
ATP metabolic process	+	+	+	+
Electron Transport Chain (OXPHOS system in mitochondria)	+	+	+	+
Glycolysis / gluconeogenesis	+	-	+	-
Generation of precursor metabolites and energy	+	+	+	+
Ribonucleotide metabolic process	+	-	+	+
Cellular respiration	+	+	-	-
Respiratory electron transport	+	+	+	+

Table 3.25: The comparison of the significantly terms in PEA results for all models. The selected 10 pathways are given.

Pathway Names	Recon2- jbv	Recon2- j82	iHsa- jbv	iHsa- j82
Glycolysis	+	+	+	+
Fatty acid metabolism	-	-	+	+
Pyruvate metabolism and Citric Acid (TCA) cycle	+	+	+	+
Glucose metabolism	+	+	+	+
Formation of ATP by chemiosmotic coupling	-	-	+	+
Respiratory electron transport	+	+	+	+
Complex I biogenesis	+	+	-	-
Metabolism of carbohydrates	+	+	+	+
Triglyceride metabolism	-	-	+	+
The citric acid (TCA) cycle and respiratory electron transport	+	+	+	+

The comparison of cell lines and models shows that important biological processes such as TCA cycle, glycolysis and oxidative phosphorylation were affected in dormant state. The terms predicted to be affected in all four simulations (Recon2.04-jbv, Recon2.04-j82, iHsa-jbv and iHsa-j82) can give crucial information about the details

of dormancy metabolism. In addition, glucose and puruvate metabolisms are significant for all four simulations. On the other hand, association of some pathways with dormancy is model dependent. For example, fatty acid metabolism and triglyceride metabolism were found to be significant only in iHsa model simulations, whereas complex1 biogenesis was significant in Recon2.04 model simulations.

4. CONCLUSIONS

In recent years, the metabolism of metastasis mechanisms has been pointed out as the cause of the biggest failure in cancer treatment [Bogenrieder and Herlyn, 2003]. The biggest problem in the cancer process is the dormant state in the metastatic cells. These are the cells that do not grow and have treatment-resistance, but can then initiate the process of cancer again [Seidi et al., 2018]. Although cancer dormancy has become a notable phenomenon lately, there is not much information about it.

With the development of new approaches, it has become easier to investigate complex diseases such as cancer. Computational approaches can be used to investigate the dormancy to provide a wider perspective. This study focuses on the differentiation of dormant state from cancer to gain insight into condition-specific metabolic changes. A genome scale metabolic model was developed to simulate cell culture (based on medium content). Transcriptome mapping methods have also been applied to this network model to simulate metabolic behaviors between the two conditions. Condition-specific models were obtained and validated by transcriptome mapping.

Transcriptomic data from two different cell lines (jbv and j82) were integrated into two different genome scale metabolic models (Recon2.04 and iHSA) with two different integration algorithms (GIMME and E-flux) for more reliable comparison. Thus, both metabolic changes between dormancy and cancer can be interpreted and the effect of model and algorithm selection can be compared.

The results of this thesis study show that GIMME algorithm provided better results in predicting metabolic behavior. GIMME simulations pointed to the fact that metabolism of dormant state differs from cancer in basic processes. Dormant cells show less growth and keep many of their reactions minimal. In the light of the results, energy metabolism and biosynthesis metabolism were the most affected cellular processes. Since there is not enough information to compare the results of dormancy metabolism, two similar mechanisms, quiescence and cancer stem cells, were reviewed from literature in terms of their effect on metabolism. Comparison of the result obtained in this study with these two similar mechanisms shows that dormancy has metabolic similarities with quiescence [Lemons et al., 2010, Rocheteau et al., 2015].

In the continuation of the study, simulations can be repeated by using different transcriptomic datasets in order to make a more reliable interpretation of metabolism. In addition to these integration algorithms, other algorithms can also be utilized. Besides the focus on the metabolic mechanisms, drug target identification studies can also be performed.

REFERENCES

GEO, 2019 <https://www.ncbi.nlm.nih.gov/geo/>. (28.06.2019)

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., & Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, 8(5), e1002518–e1002518. <https://doi.org/10.1371/journal.pcbi.1002518>

Aguirre-Ghiso, J. A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. *Nature Reviews Cancer*, 7(11), 834–846. <https://doi.org/10.1038/nrc2256>

Allan, A. L., Vantyghem, S. A., Tuck, A. B., & Chambers, A. F. (2007). Tumor dormancy and cancer stem cells: implications for the biology and treatment of breast cancer metastasis. *Breast Disease*, 26, 87–98. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17473368>

Amelio, I., Cutruzzolá, F., Antonov, A., Agostini, M., & Melino, G. (2014). Serine and glycine metabolism in cancer. *Trends in Biochemical Sciences*, 39(4), 191–198. <https://doi.org/10.1016/j.tibs.2014.02.004>

Arora, M. (2013). Cell Culture Media: A Review. *Materials and Methods*, 3. <https://doi.org/10.13070/mm.en.3.175>

Asgari, Y., Zabihinpour, Z., Salehzadeh-Yazdi, A., Schreiber, F., & Masoudi-Nejad, A. (2015). Alterations in cancer cell metabolism: The Warburg effect and metabolic adaptation. *Genomics*, 105(5–6), 275–281. <https://doi.org/10.1016/j.ygeno.2015.03.001>

Asgari, Y., Zabihinpour, Z., Salehzadeh-Yazdi, A., Schreiber, F., & Masoudi-Nejad, A. (2015). Alterations in cancer cell metabolism: The Warburg effect and metabolic adaptation. *Genomics*, 105(5–6), 275–281. <https://doi.org/10.1016/j.ygeno.2015.03.001>

Aurich, M. K., & Thiele, I. (2016). Computational Modeling of Human Metabolism and Its Application to Systems Biomedicine. In U. Schmitz & O. Wolkenhauer (Eds.), *Systems Medicine* (pp. 253–281). https://doi.org/10.1007/978-1-4939-3283-2_12

Becker, S. A., & Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, 4(5). <https://doi.org/10.1371/journal.pcbi.1000082>

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 57, No. 1, pp. 289-300

- Berdasco, M., & Esteller, M. (2010). Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. *Developmental Cell*, 19(5), 698–711. <https://doi.org/10.1016/j.devcel.2010.10.005>
- Blais, E. M., Rawls, K. D., Dougherty, B. V., Li, Z. I., Kolling, G. L., Ye, P., ... Papin, J. A. (2017). Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nature Communications*, 8, 1–15. <https://doi.org/10.1038/ncomms14250>
- Blazier, A. S., & Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*, 3 AUG(August). <https://doi.org/10.3389/fphys.2012.00299>
- Bordbar A., Feist A. M., Usaite-Black R., Woodcock J., Palsson B. O., Famili I., (2011), "A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology", *BMC Syst Biol*, 5(1), 180.
- Bogenrieder, T., & Herlyn, M. (2003). Axis of evil: molecular mechanisms of cancer metastasis. *Oncogene*, 22(42), 6524–6536. <https://doi.org/10.1038/sj.onc.1206757>
- Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Mih, N., Gatto, F., ... Group, B. (2018). Recon3D: A Resource Enabling A Three-Dimensional View of Gene Variation in Human Metabolism. *Nature Biotechnology*, 36(3), 272–281. <https://doi.org/10.1038/nbt.4072.Recon3D>
- Bryan, T. M., & Cech, T. R. (1999). Telomerase and the maintenance of chromosome ends. *Current Opinion in Cell Biology*, 11(3), 318–324. [https://doi.org/10.1016/S0955-0674\(99\)80043-X](https://doi.org/10.1016/S0955-0674(99)80043-X)
- Cazzaniga, M., & Bonanni, B. (2015). Breast Cancer Metabolism and Mitochondrial Activity: The Possibility of Chemoprevention with Metformin. In *BioMed Research International* (Vol. 2015). <https://doi.org/10.1155/2015/972193>
- Chakraborty, S., Hosen, M. I., Ahmed, M., & Shekhar, H. U. (2018). Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *BioMed Research International*, 2018(Figure 1), 1–14. <https://doi.org/10.1155/2018/9836256>
- Chandrasekaran, S., and Price, N. D.(2010).Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc.Natl.Acad.Sci.U.S.A.* 107, 17845–17850.
- Charlton, P., & Spicer, J. (2015). Targeted therapy in cancer. *Medicine (United Kingdom)*, 44(1), 34–38. <https://doi.org/10.1016/j.mpmed.2015.10.012>
- Clish, C. B. (2015). Metabolomics: an emerging but powerful tool for precision medicine. *Molecular Case Studies*, 1(1), a000588. <https://doi.org/10.1101/mcs.a000588>
- Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., ... Galagan, J. E. (2009). Interpreting expression data with metabolic flux models:

Predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Computational Biology*, 5(8). <https://doi.org/10.1371/journal.pcbi.1000489>

Coller, H. A. (2014). Is cancer a metabolic disease? 9582978472. *American Journal of Pathology*, 184(1), 4–17. <https://doi.org/10.1016/j.ajpath.2013.07.035>

Clough, E., Barret, T., (2016), The Gene Expression Omnibus database, *Methods Mol. Biol.* doi:10.1007/978-1-4939-3578-9_5.

Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., & Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO Journal*, 11(5), 1921–1929. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1582420> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC556651>

Çakır, T., Patil, K. R., Önsan, Z. I., Ülgen, K. Ö., Kirdar, B., & Nielsen, J. (2006). Integration of metabolome data with metabolic networks reveals reporter reactions. *Molecular Systems Biology*, 2, 1–11. <https://doi.org/10.1038/msb4100085>

Çakır, T., & Khatibipour, M. J. (2014). Metabolic Network Discovery by Top-Down and Bottom-Up Approaches and Paths for Reconciliation. *Frontiers in Bioengineering and Biotechnology*, 2(December), 1–11. <https://doi.org/10.3389/fbioe.2014.00062>

Çakır, T. (2015). Reporter pathway analysis from transcriptome data: Metabolite-centric versus Reaction-centric approach. *Scientific Reports*, 5(August), 1–10. <https://doi.org/10.1038/srep14563>

Damiani, C., Colombo, R., Gaglio, D., Mastroianni, F., Pescini, D., Westerhoff, H. V., ... Alberghina, L. (2017). A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The WarburQ effect. *PLoS Computational Biology*, 13(9), 1–29. <https://doi.org/10.1371/journal.pcbi.1005758>

Danhier, P., Bański, P., Payen, V. L., Grasso, D., Ippolito, L., Sonveaux, P., & Porporato, P. E. (2017). Cancer metabolism in space and time: Beyond the Warburg effect. *Biochimica et Biophysica Acta - Bioenergetics*, 1858(8), 556–572. <https://doi.org/10.1016/j.bbabo.2017.02.001>

Daye, D., & Wellen, K. E. (2012). Metabolic reprogramming in cancer: Unraveling the role of glutamine in tumorigenesis. *Seminars in Cell and Developmental Biology*, 23(4), 362–369. <https://doi.org/10.1016/j.semcdb.2012.02.002>

Dittrich, MT., Klau, G., Rosenwald, A., Dandekar, T., Müller, T., (2008), Identifying functional modules in protein–protein interaction networks: an integrated exact approach, *Bioinformatics*, <https://doi.org/10.1093/bioinformatics/btn161>

- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., ... Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6), 1777–1782. <https://doi.org/10.1073/pnas.0610772104>
- Eagle, H. (1955). Nutrition Needs of Mammalian Cells in Tissue Culture. *Science*, 122(3168), 501 LP – 504. <https://doi.org/10.1126/science.122.3168.501>
- Ebinger, S., Özdemir, E. Z., Ziegenhain, C., Tiedt, S., Castro Alves, C., Grunert, M., ... Jeremias, I. (2016). Characterization of Rare, Dormant, and Therapy-Resistant Cells in Acute Lymphoblastic Leukemia. *Cancer Cell*, 30(6), 849–862. <https://doi.org/10.1016/j.ccell.2016.11.002>
- Emmert-Streib, F., Dehmer, M., & Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology*, 2, 38. <https://doi.org/10.3389/fcell.2014.00038>
- Feist, A., & Palsson, B. (2008). The Growing Scope of Applications of Genome-scale Metabolic Reconstructions: the case of *E. coli*. *Nature Biotechnology*, 26(6), 659–667. <https://doi.org/10.1038/nbt1401>
- Folger, O., Jerby, L., Frezza, C., Gottlieb, E., Ruppin, E., & Shlomi, T. (2011). Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology*, 7(501), 1–10. <https://doi.org/10.1038/msb.2011.35>
- Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research*, 13(2), 244–253. <https://doi.org/10.1101/gr.234503>
- Gao, H., Chakraborty, G., Lee-Lim, A. P., Mavrakis, K. J., Wendel, H.-G., & Giancotti, F. G. (2014). Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation. *Proceedings of the National Academy of Sciences*, 111(46), 16532–16537. <https://doi.org/10.1073/pnas.1403234111>
- Gatto, F., Nookaew, I., & Nielsen, J. (2014). Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proceedings of the National Academy of Sciences*, 111(9), E866–E875. <https://doi.org/10.1073/pnas.1319196111>
- Gille, C., Bölling, C., Hoppe, A., Bulik, S., Hoffmann, S., Hübner, K., ... Holzhütter, H. G. (2010). HepatoNet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology*, 6(411). <https://doi.org/10.1038/msb.2010.62>
- Gomis, R. R., & Gawrzak, S. (2016). Tumor cell dormancy. *Molecular Oncology*, 11(1), 62–78. <https://doi.org/10.1016/j.molonc.2016.09.009>
- Gonzalez-Angulo, A. M., Hennessy, B. T. J., & Mills, G. B. (2010). Future of personalized medicine in oncology: a systems biology approach. *Journal of*

Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28(16), 2777–2783. <https://doi.org/10.1200/JCO.2009.27.0777>

Grizzi, F., & Chiriva-Internati, M. (2006). Cancer: Looking for simplicity and finding complexity. *Cancer Cell International*, 6, 1–7. <https://doi.org/10.1186/1475-2867-6-4>

Hadi, M., & Marashi, S. A. (2014). Reconstruction of a generic metabolic network model of cancer cells. *Molecular BioSystems*, 10(11), 3014–3021. <https://doi.org/10.1039/c4mb00300d>

Hanahan, D. (2014). Rethinking the war on cancer. *The Lancet*, 383(9916), 558–563. [https://doi.org/10.1016/S0140-6736\(13\)62226-6](https://doi.org/10.1016/S0140-6736(13)62226-6)

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>

Harrington, K. J. (2015). The biology of cancer. *Medicine (United Kingdom)*, 44(1), 1–5. <https://doi.org/10.1016/j.mpmed.2015.10.005>

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination*. London, UK: Sage Publications.

Holzhütter H.-G., (2004), "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks", *Eur J Biochem*, 271(14), 2905–2922.

Hurst, R. E., Hauser, P. J., Kyker, K. D., Heinlen, J. E., Hodde, J. P., Hiles, M. C., ... Ihnat, M. A. (2013). Suppression and Activation of the Malignant Phenotype by Extracellular Matrix in Xenograft Models of Bladder Cancer: A Model for Tumor Cell “Dormancy.” *PLoS ONE*, 8(5). <https://doi.org/10.1371/journal.pone.0064181>

Jensen, P. A., & Papin, J. A. (2011). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27(4), 541–547. <https://doi.org/10.1093/bioinformatics/btq702>

Jensen, P.A., and Papin, J.A. (2011). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* 27, 541–547.

Jerby, L., Shlomi, T., & Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Molecular Systems Biology*, 6(401), 1–9. <https://doi.org/10.1038/msb.2010.56>

Kato, Y., Maeda, T., Suzuki, A., & Baba, Y. (2018). Cancer metabolism: New insights into classic characteristics. *Japanese Dental Science Review*, 54(1), 8–21. <https://doi.org/10.1016/j.jdsr.2017.08.003>

Kherlopian, A. R., Song, T., Duan, Q., Neimark, M. A., Po, M. J., Gohagan, J. K., & Laine, A. F. (2008). A review of imaging techniques for systems biology. *BMC Systems Biology*, 2, 74. <https://doi.org/10.1186/1752-0509-2-74>

Kilburn, D. G., Lilly, M. D., & Webb, F. C. (1969). The energetics of mammalian cell growth. *Journal of Cell Science*, 4(3), 645–654. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5817088>

Kim, M. K., & Lun, D. S. (2014). Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal*, 11(18), 59–65. <https://doi.org/10.1016/j.csbj.2014.08.009>

Kobeissy, F., Alawieh, A., Mondello, S., Boustany, R.-M., & Gold, M. S. (2013). Biomarkers in psychiatry: how close are we? *Frontiers in Psychiatry*, 3, 114. <https://doi.org/10.3389/fpsyt.2012.00114>

Lawson, M. A., McDonald, M. M., Kovacic, N., Khoo, W. H., Terry, R. L., Down, J., ... Croucher, P. I. (2015). Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. *Nature Communications*, 6(May), 1–15. <https://doi.org/10.1038/ncomms9983>

Lee HC, Yin PH, Lu CY, Chi CW, Wei YH (2000) Increase of mitochondria and mitochondrial DNA in response to oxidative stress in human cells. *Biochem J* 348(Pt 2):425–432

Lemons, J. M. S., Collier, H. A., Feng, X. J., Bennett, B. D., Legesse-Miller, A., Johnson, E. L., ... Rabinowitz, J. D. (2010). Quiescent fibroblasts exhibit high metabolic activity. *PLoS Biology*, 8(10). <https://doi.org/10.1371/journal.pbio.1000514>

Likić, V. A., McConville, M. J., Lithgow, T., & Bacic, A. (2010). Systems biology: the next frontier for bioinformatics. *Advances in Bioinformatics*, 2010, 268925. <https://doi.org/10.1155/2010/268925>

Lin MT, Beal MF. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* 2006;443:787–795.

Lindl, T. (2002): "Zell- und Gewebekultur". 5th ed. Spektrum Akademischer Verlag, Heidelberg

Machado D., Herrgård M., (2014), "Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism", *PLoS Comput Biol*, 10(4), e1003580.

Manley, G. (2013). Public Access NIH Public Access. 71(2), 233–236. <https://doi.org/10.1038/mp.2011.182>.doi

Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2016). Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, 19(2), 286–302. <https://doi.org/10.1093/bib/bbw114>

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., ... Nielsen, J. (2013). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Molecular Systems Biology*, 9(1), 1–16. <https://doi.org/10.1038/msb.2013.5>

- Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications*, 5(May 2013), 1–11. <https://doi.org/10.1038/ncomms4083>
- Nilsson, A., & Nielsen, J. (2016). Genome scale metabolic modeling of cancer. *Metabolic Engineering*, 43(September 2016), 103–112. <https://doi.org/10.1016/j.ymben.2016.10.022>
- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245–248. <https://doi.org/10.1038/nbt.1614>
- Özcan, E., & Çakir, T. (2016). Reconstructed metabolic network models predict flux-level metabolic reprogramming in glioblastoma. *Frontiers in Neuroscience*, 10(APR), 1–11. <https://doi.org/10.3389/fnins.2016.00156>
- Pavlova, N. N., & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cell Metabolism*, 23(1), 27–47. <https://doi.org/10.1016/j.cmet.2015.12.006>
- Poste, G., & Fidler, I. J. (1980). The pathogenesis of cancer metastasis. *Nature*, 283(5743), 139–146. <https://doi.org/10.1038/283139a0>
- Price, N. D., Reed, J. L., & Palsson, B. Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11), 886–897. <https://doi.org/10.1038/nrmicro1023>
- Rabinowitz JD, White E. Autophagy and metabolism. *Science* 2010; 330:1344-8; PMID:21127245; [http:// dx.doi.org/10.1126/science.1193497](http://dx.doi.org/10.1126/science.1193497).
- Reichle, A., & Vogt, T. (2008). Systems biology: a therapeutic target for tumor therapy. *Cancer Microenvironment: Official Journal of the International Cancer Microenvironment Society*, 1(1), 159–170. <https://doi.org/10.1007/s12307-008-0012-5>
- Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., Vilo, J., (2016), g:Profiler—a web server for functional interpretation of gene lists, *Nucleic Acids Research*, doi: 10.1093/nar/gkw199
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Roberts, E. H. (1963), The Effects of Some Organic Growth Substances and Organic Nutrients on Dormancy in Rice Seed. *Physiologia Plantarum*, 16: 745-755. doi:10.1111/j.1399-3054.1963.tb08352.x
- Rocheteau, P., Vinet, M., & Chretien, F. (2015). Dormancy and Quiescence of Skeletal Muscle Stem Cells. 56, 215–235. <https://doi.org/10.1007/978-3-662-44608-9>

Ruppender, N.S., Morrissey, C., Lange, P.H., & Vessella, R.L. (2013). Dormancy in solid tumors: implications for prostate cancer. *Cancer and Metastasis Reviews*, 32, 501-509.

Safari-Alighiarloo, N., Taghizadeh, M., Rezaei tavirani, M., Goliaei, B., & Peyvandi, A. (2014). Protein-Protein Interaction Networks (PPI) and complex diseases. In *Gastroenterology and hepatology from bed to bench* (Vol. 7).

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., ... Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9), 1290–1307. <https://doi.org/10.1038/nprot.2011.308>

Seidi, K., Manjili, M. H., Jahanban-esfahlan, R., & Javaheri, T. (2018). Tumor cell dormancy : threat or opportunity in the fight against cancer. (October). <https://doi.org/10.20944/preprints201810.0225.v1>

Shlomi,T., Cabili,M.N., Herrgard, M.J., Palsson,B.O., and Rupp, E. (2008).Network-based predictionofhumantissue-specific metabolism. *Nat.Biotechnol.* 26, 1003–1010.

Smyth, G. K. (n.d.). limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, 397–420. https://doi.org/10.1007/0-387-29362-0_23

Spano, D., Heck, C., De Antonellis, P., Christofori, G., & Zollo, M. (2012). Molecular networks that regulate cancer metastasis. *Seminars in Cancer Biology*, 22(3), 234–249. <https://doi.org/10.1016/j.semcancer.2012.03.006>

Tarlak, F., Sadikoğlu, H., & Çakir, T. (2014). The role of flexibility and optimality in the prediction of intracellular fluxes of microbial central carbon metabolism. *Molecular BioSystems*, 10(9), 2459–2465. <https://doi.org/10.1039/c4mb00117f>

Teicher, M. H., Anderson, C. M., & Polcari, A. (2012). Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. *Proceedings of the National Academy of Sciences*, 109(9), E563 LP-E572. <https://doi.org/10.1073/pnas.1115396109>

Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., ... Palsson, B. Ø. (2013). A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, 31(5), 419–425. <https://doi.org/10.1038/nbt.2488>

Valastyan, S., & Weinberg, R. A. (2011). Tumor metastasis: molecular insights and evolving paradigms. *Cell*, 147(2), 275–292. <https://doi.org/10.1016/j.cell.2011.09.024>

Vander Heiden, M. G., & DeBerardinis, R. J. (2017). Understanding the Intersections between Metabolism and Cancer Biology. *Cell*, 168(4), 657–669. <https://doi.org/10.1016/j.cell.2016.12.039>

Valcourt JR, Lemons JM, Haley EM, Kojima M, Demuren OO, Collier HA. Staying alive: metabolic adaptations to quiescence. *Cell Cycle*. 2012;11(9):1680–1696. doi:10.4161/cc.19879

Wagner, A., Zarecki, R., Reshef, L., Gochev, C., Sorek, R., Gophna, U., & Ruppin, E. (2013). Computational evaluation of cellular metabolic costs successfully predicts genes whose expression is deleterious. *Proceedings of the National Academy of Sciences*, 110(47), 19166–19171. <https://doi.org/10.1073/pnas.1312361110>

Wang, Y., Eddy, J. A., & Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Systems Biology*, 6, 153. <https://doi.org/10.1186/1752-0509-6-153>

Yeh, A. C., & Ramaswamy, S. (2015). Mechanisms of cancer cell dormancy—another hallmark of cancer? *Cancer Research*, 75(23), 5014–5022. <https://doi.org/10.1158/0008-5472.CAN-15-1370>

Yizhak, K., Chaneton, B., Gottlieb, E., & Ruppin, E. (2015). Modeling cancer metabolism on a genome scale. *Molecular Systems Biology*, 11(6), 817–817. <https://doi.org/10.15252/msb.20145307>

Zielinski, D. C., Jamshidi, N., Corbett, A. J., Bordbar, A., Thomas, A., & Palsson, B. O. (2017). Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Scientific Reports*, 7, 1–14. <https://doi.org/10.1038/srep41241>

Zugazagoitia, J., Guedes, C., Ponce, S., Ferrer, I., Molina-Pinelo, S., & Paz-Ares, L. (2016). Current Challenges in Cancer Treatment. *Clinical Therapeutics*, 38(7), 1551–1566. <https://doi.org/10.1016/j.clinthera.2016.03.026>

BIOGRAPHY

Merve KUTAY was born in OSMANGAZI in April 11, 1992. She graduated from Gebze Technical University Department of Molecular Biology and Genetics in 2015. She is a Msc. student in Bioinformatics and Systems Biology Program under Bioengineering Department at Graduate School of Natural and Applied Sciences in the same university.

APPENDICES

Appendix A: Medium Compositions

A growth medium or culture medium comprise an appropriate source of energy and compounds that regulate the cell cycle to support the growth of cells. A typical culture medium contains amino acids, vitamins, inorganic salts, glucose, and serum [Arora, 2013]. In this thesis study, the culture medium used in the transcriptome experiments of Hurst et al. (2013) consists of Dulbecco's Modified Eagle's medium (DMEM) and Fetal Calf serum (FCS) [Lindl, 2002]. DMEM is a modification of Basal Medium Eagle (BME) that contains higher concentration of amino acids and vitamins. DMEM medium content was taken from “<https://www.sigmaaldrich.com/life-science/cell-culture/learning-center/media-formulations/dme.html>” website. Serum is a complex mix of albumins, growth factors and growth inhibitors. Related compositions are given in TableA1.1-2.

Table A1.1: Composition of Dulbecco's Modified Eagle's medium (DMEM).

Components of DMEM		
Inorganic Salts	Amino Acids	Vitamins
CaCl ₂	L-Alanyl-L-Glutamine	Choline Chloride
Fe(NO ₃) ₃ • 9H ₂ O	L-Arginine • HCl	Folic Acid
MgSO ₄	L-Cystine • 2HCl	myo-Inositol
KCl	L-Glutamine	Niacinamide
NaHCO ₃	Glycine	D-Pantothenic Acid • ½Ca
NaCl	L-Histidine • HCl • H ₂ O	Pyridoxal • HCl
NaH ₂ PO ₄	L-Isoleucine	Pyridoxine • HCl
Amino Acids	L-Leucine	Riboflavin
L-Tyrosine • 2Na • 2H ₂ O	L-Lysine • HCl	Thiamine • HCl
L-Tyrosine	L-Methionine	ADD
L-Valine	Other	NaHCO ₃
L-Threonine	D-Glucose	L-Glutamine
L-Tryptophan	HEPES	Glucose
L-Serine	Phenol Red • Na	
L-Phenylalanine	Pyruvic Acid • Na	

Table A1.2: Composition of Fetal Calf serum (FCS).

Components of FCS		
Endotoxins	Testosterone	Growth hormone
Glucose	Progesterone	Parathormone
Protein	Prolactin = Luteotropic hormone	Triiodothyronine
Albumin	Luteinizing hormone	Thyroxine
Hemoglobine	Prostaglandin E	Thyroid-stimulating hormone
Bilirubin	Prostaglandin F	Follicle-stimulating hormone
Bilirubin	Vitamine A	Sodium, Na ⁺
Urea	Vitamine E	Potassium, K ⁺
Urate	Cholesterol	Calcium, Ca ²⁺
Creatinin	Lactate-dehydrogenase	Chloride, Cl ⁻
Insulin	Alkaline Phosphatase	Phosphate, Pi
Cortisol	Aspartate-Aminotransferase	Selen

Appendix B: Names of the Important Reactions

In order to compare the results of the iHsa and Recon2.04 metabolic models, certain reactions were selected and predicted fluxes were compared (Table 3.7-10). These reactions were determined as the initial reactions of major pathways such as glycolysis, PPP and TCA cycle. The original names of the reactions in the models (Recon2.04 and iHsa) are given in Table B1.1.

Table B1.1: Reaction Names in Tables (section 3.5).

Reaction Names in Tables	Reactions in Recon2	Reactions in iHsa
Oxygen uptake	Exchange reaction for oxygen	O ₂ exchange
Production of Lactate	L-lactate dehydrogenase	(R)-Lactate:NAD ⁺ oxidoreductase
PPP	Glucose 6-phosphate dehydrogenase	D-glucose-6-phosphate:NADP ⁺ 1-oxidoreductase[c]
TCA	Citrate sythesis	Citrate lyase
Biomass	Biomass	HMRBiomass

Appendix C: All Pathways in the Models

Pathway lists of genome-scale metabolic network models (Recon2.04 and iHsa) used in the thesis study are given in the Table C1.1-2.

Table C1.1: Pathway Names in Recon2.04 model.

Recon2.04 all pathway names	
Transport, lysosomal	Blood group synthesis
Transport, mitochondrial	Glutamate metabolism
Transport, endoplasmic reticular	Pentose phosphate pathway
beta-Alanine metabolism	Cholesterol metabolism
Vitamin D metabolism	Fatty acid oxidation
Transport, extracellular	Transport, peroxisomal
Lysine metabolism	Transport, golgi apparatus
Glycine, serine, alanine and threonine metabolism	Aminosugar metabolism
Methionine and cysteine metabolism	Phosphatidylinositol phosphate metabolism
Pyrimidine catabolism	Urea cycle
Propanoate metabolism	Glycerophospholipid metabolism
Tryptophan metabolism	Citric acid cycle
Tyrosine metabolism	Pyruvate metabolism
Ubiquinone synthesis	Bile acid synthesis
Transport, nuclear	Vitamin B2 metabolism
Valine, leucine, and isoleucine metabolism	Glycolysis/gluconeogenesis
Sphingolipid metabolism	Nucleotide interconversion
Butanoate metabolism	Nucleotide salvage pathway
Purine catabolism	Selenoamino acid metabolism
Taurine and hypotaurine metabolism	Purine synthesis
Arginine and Proline Metabolism	Keratan sulfate synthesis
Cytochrome metabolism	Triacylglycerol synthesis
Steroid metabolism	Alanine and aspartate metabolism
Heme synthesis	Miscellaneous
N-glycan degradation	Eicosanoid metabolism
O-glycan synthesis	Glutathione metabolism

Table C1.1: Continued.

Recon2.04 all pathway names	
Starch and sucrose metabolism	Glyoxylate and dicarboxylate metabolism
Biotin metabolism	Galactose metabolism
R group synthesis	Nucleotide sugar metabolism
Unassigned	Chondroitin synthesis
Vitamin C metabolism	Heparan sulfate degradation
Pyrimidine synthesis	Hyaluronan metabolism
Oxidative phosphorylation	Histidine metabolism
Vitamin A metabolism	Vitamin B6 metabolism
Stilbene, coumarine and lignin synthesis	Inositol phosphate metabolism
N-glycan synthesis	CoA catabolism
Heme degradation	NAD metabolism
Limonene and pinene degradation	Phenylalanine metabolism
ROS detoxification	Thiamine metabolism
Vitamin B12 metabolism	Cysteine Metabolism
C5-branched dibasic acid metabolism	Androgen and estrogen synthesis and metabolism
Alkaloid synthesis	Lipoate metabolism
Chondroitin sulfate degradation	Arachidonic acid metabolism
D-alanine metabolism	Vitamin E metabolism
Fatty acid synthesis	Linoleate metabolism
Folate metabolism	Glycosphingolipid metabolism
Tetrahydrobiopterin metabolism	Xenobiotics metabolism
Exchange/demand reaction	Dietary fiber binding
CoA synthesis	Squalene and cholesterol synthesis
Keratan sulfate degradation	Uptake try16
Fructose and mannose metabolism	

Table C1.2: Pathway Names in iHsa model.

iHsa all pathway names	
Purine metabolism	Omega-3 fatty acid metabolism
Central carbon metabolism	Arachidonic acid metabolism
Protein metabolism	Steroid metabolism
Sphingolipid metabolism	Cholesterol metabolism
Vitamin B metabolism	Bile acid metabolism
Alanine aspartate and glutamate metabolism	Miscellaneous
Glycine serine and threonine metabolism	Glutathione metabolism
Electron transport chain	Pentose phosphate pathway
Aromatic amino acid metabolism	Glycan metabolism
Nucleotide metabolism	Serotonin and melatonin biosynthesis
Carbohydrate metabolism	Vitamin D metabolism
Fatty acid metabolism	Xenobiotics metabolism
Vitamin C metabolism	Biomass
Arginine and proline metabolism	Omega-6 fatty acid metabolism
beta-Alanine metabolism	Beta oxidation of fatty acids
Pyrimidine metabolism	Cholesterol-ester metabolism
Glycerolipid metabolism	Proteoglycan metabolism
Valine leucine and isoleucine metabolism	Porphyrim metabolism
Eicosanoid metabolism	Vitamin E metabolism
Inositol phosphate metabolism	Transport
Vitamin A metabolism	Exchange
Lysine metabolism	HMR biomass
Cysteine and methionine metabolism	ATP maintenance