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

RECONSTRUCTION AND TRANSCRIPTOME BASED ANALYSIS OF RAT BRAIN-SPECIFIC GENOME-SCALE METABOLIC NETWORK MODEL FOR PARKINSON'S DISEASE

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

GEBZE
2021

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

RECONSTRUCTION AND TRANSCRIPTOME BASED ANALYSIS OF RAT BRAIN-SPECIFIC GENOME-SCALE METABOLIC NETWORK MODEL FOR PARKINSON'S DISEASE

ORHAN BELLUR

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

2021

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PARKİNSON HASTALIĞI İÇİN SIÇAN BEYNİNE ÖZGÜ GENOM ÖLÇEKLİ METABOLİK MODELİN OLUŞTURULMASI VE TRANSKRİPTOMA DAYALI ANALİZİ

ORHAN BELLUR

YÜKSEK LİSANS TEZİ
BİYOMÜHENDİSLİK ANABİLİM DALI
BİYOİNFORMATİK VE SİSTEM BİYOLOJİSİ PROGRAMI

DANIŞMANI
DOÇ. DR. TUNAHAN ÇAKIR

GEBZE
2021

	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 08/07/2021 tarih ve 2021/32 sayılı kararıyla oluşturulan jüri tarafından 03/08/2021 tarihinde tez savunma sınavı yapılan Orhan BELLUR'ın tez çalışması Biyomühendislik Anabilim Dalında Biyoinformatik ve Sistem Biyolojisi programında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Doç.Dr. Tunahan ÇAKIR

ÜYE : Dr. Öğr.Üyesi Pınar PİR

ÜYE : Prof.Dr. Kutlu ÜLGEN

ÜYE :

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

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

SUMMARY

Parkinson's disease (PD), affecting millions of people worldwide, is one of the most common neurodegenerative diseases. PD is mainly characterized by the neurodegeneration of the dopaminergic neurons. It is a complex multifactorial disease, and numerous molecular mechanisms have been implicated in the development of PD, including mis-folded protein formation, loss of dopaminergic neurons, and mitochondrial malfunctions. The complexity of the disease and the presence of different symptoms in each patient necessitate a systems biology approach. Constraint-based modelling is commonly used as a systems biology approach to analyze metabolism, and it requires a genome-scale metabolic model (GEM) reconstruction, which is based on comprehensive biochemical information. Many diseases in humans are intimately connected to metabolism. GEMs are widely utilized to elucidate the healthy and disease state of the cell since they provide a mechanical connection between genotype and phenotype. *Rattus norvagicus* is a commonly used model organism in clinical and pharmaceutical research. In this thesis study, six rat transcriptome datasets associated with PD were subjected to hierarhical clustering analysis by using metabolic genes to investigate the context of similarity/dissimilarity of transcriptome datasets. Later, a rat brain-specific GEM, which consists of 994 reactions and 766 genes, was reconstructed for the first time in literature by utilizing the human brain-specific GEM and mouse brain-specific GEM as templates. Then, the six rat transcriptome datasets associated with PD were used to generate context-specific models. Transcriptome datasets were integrated into rat brain-specific models through iMAT and INIT algorithms. Constraint-based modeling (FBA, MOMA, LseiFBA, modified linear LseiFBA,) on obtained context-specific metabolic networks showed that the predicted flux values are within the range of literature observations, validating the newly reconstructed rat brain-specific GEM. LseiFBA and linear modified LseiFBA gave more accurately predicted major fluxes in PD rat models.

Key Words Parkinson's disease, systems biology, genome-scale metabolic model, transcriptome, constraint-based analysis, animal models.

ÖZET

Parkinson hastalığı dünya çapında milyonlarca insanı etkileyen en yaygın nörodejeneratif hastalıklardan biridir. Hastalık dopaminerjik nöron dejenerasyonu ile karakterizedir. Parkinson çok faktörlü kompleks bir hastalık olup, yanlış protein katlanması, lewy cisimciği oluşumu, dopaminerjik nöron kaybı ve mitokondriyel fonksiyonel bozukluk gibi birçok moleküler mekanizmanın hastalığın gelişiminde rol oynadığı bilinmektedir. Hastalığın kompleks olması ve her hastada farklı semptomların varlığı sistem biyolojisi yaklaşımını zorunlu kılmaktadır. Kısıt tabanlı modelleme, yaygın olarak kullanılan bir sistem biyolojisi yaklaşımıdır. İnsan hastalıklarının çoğu metabolizma ile ilişkilendirilmektedir. Genom ölçekli metabolik modeller genotip ve fenotip arasında bağlantıyı sağladıkları için normal ve hastalıklı durumda moleküler mekanizmaları anlamada yaygın olarak kullanılmaktadır. *Rattus norvagicus,* klinik ve ilaç deneyi araştırmalarında yaygın olarak kullanılan bir model organizmadır. Bu tez çalışmasında öncelikle PD ile ilişkili altı sıçan transkriptom veri seti, aralarındaki benzersizliğin bağlamını araştırmak için metabolik genler kullanılarak hiyerarşik küme analizine tabi tutuldu. Ardından, literatürde ilk kez, insan ve fare beynine özgü mevcut olan metabolik modeller referans olarak kullanılarak sıçan beynine özgü ve 994 reaksiyon ve 766 genden oluşan bir metabolik model oluşturuldu. Koşula özgü metabolik modeller oluşturmak için bu çalışmada Parkinson hastalığı ile ilişkili altı sıçan transkriptom veriseti kullanıldı. Transkriptom verisetleri, iMAT ve INIT algoritmaları aracılığıyla sıçan beynine özgü genom ölçekli metabolik modele entegre edildi. Koşula özgü metabolik modeller kullanılarak yapılan kısıt tabanlı analizler, tahmin edilen akı değerlerinin literatür bulguları ile tutarlı olduğunu göstermiştir. Böylece, geliştirilen sıçan beynine özgü genom ölçekli metabolik modelin doğrulanması gerçekleştirilmiştir. LseiFBA ve Lineer LseiFBA yöntemlerinin hastalık durumundaki akı değerlerini daha doğru tahmin ettiği tespit edilmiştir.

Anahtar Kelimeler: Parkinson hastalığı, sistem biyolojisi, genom ölçekli metabolik model, transkriptom, kısıt tabanlı analiz, hayvan modelleri.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Assoc. Prof. Dr. Tunahan ÇAKIR for providing me with the opportunity to do research in his group. The encouragement and support I got from him provided me with the chance to get engaged in various research fields and broaden my horizons. His participation in meetings, reviewing my work, and providing expertise and inspiration with me when I needed it are all much appreciated.

I consider myself lucky to have had ongoing assistance from my lab colleagues, especially my project partner who is Ronald Regan ODONGO, and I would like to thank everyone for their contribution.

I am very grateful for giving me the pleasant working atmosphere of computer systems biology groups at the Gebze Technical University.

I also would like to thank Assoc.Prof. Dr. Mehmet Somel for him contributions and suggestions.

I would also want to express my gratitude to the Health Institutes of Turkey (TUSEB) for providing funds for this research through project code: 2019-TA-01: 3440.

My gratitude goes out to my parents for their support, understanding, and encouragement during my educational journey.

TABLE OF CONTENT

	Page
SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
LIST of ABBREVIATIONS	xi
LIST of FIGURES	xiii
LIST of TABLES	xv
1. INTRODUCTION	1
2. BACKGROUND INFORMATION	3
2.1. Energy Metabolism in the Brain	3
2.1.1. Glucose-driven Metabolic Pathways in Neurons and Astrocytes	4
2.1.2. Altered Metabolism in Neurodegenerative Disorders	7
2.2. Parkinson's Disease	9
2.2.1. The Genetics of Parkinson's Disease	10
2.2.2. Parkinson's Disease Metabolism	11
2.2.3. Experimental Models for Parkinson's Disease	14
2.3. Systems Biology	15
2.4. Transcriptomics	17
2.5. Genome-scale Metabolic Networks	19
2.5.1. Human Genome Scale Metabolic Models	20
2.5.2. Brain-specific GEMs	23
2.5.3. Genome-scale Metabolic Model Reconstruction	25
2.6. Constraint-Based Modelling	27
2.6.1. Flux Balance Analysis (FBA)	29
2.6.2. Minimization of Metabolic Adjustment (MOMA)	30
2.6.3. Least-square with Equalities and Inequalities Flux Balance Analysis (LseiFBA)	31
2.6.4. Modified Least-square with Equalities and Inequalities Flux Balance Analysis (mLseiFBA)	32
2.7. Context-specific Model Reconstruction	33

2.7.1. Integrative Metabolic Analysis Tool (iMAT) 35
2.7.2. Integrative Network Inference for Tissues (INIT) 36
3. CLUSTERING ANALYSIS 38
3.1. Datasets 38
3.2. Datasets Normalization 39
3.1.1. Datasets Preprocessing 40
3.1.2. Principal Component Analysis (PCA) 44
3.3. Hierarchical Clustering Analysis 46
4. RECONSTRUCTION OF A RAT BRAIN-SPECIFIC GENOME SCALE METABOLIC MODEL 51
4.1. Draft Reconstruction 53
4.2. Manual Curation 54
4.3. Conversion into a Mathematical Model 59
4.4. Evaluation and Analysis 60
5. RECONSTRUCTION AND CONSTRAINT-BASED ANALYSIS OF CONTEXT SPECIFIC MODELS 67
5.1. Transcriptome Data Integration Algorithms 67
5.1.1. iMAT Algorithm 70
5.1.2. INIT Algorithm 72
5.2. Constraint Based Analysis of Context Specific Models 74
5.2.1. Healthy Control Flux Prediction 74
5.2.2. Flux Distribution in Parkinson's Disease 77
6. DISCUSSION 83
6.1. Metabolically Similar Comparison Groups 83
6.2. Validation of the Newly Reconstructed iBrain766-Rn Model 84
6.3. Comparison of iMAT-like Family Algorithms 84
6.3.1. Healthy Case Flux Distribution 85
6.3.2. Parkinson's Disease Case Flux Distribution 87
6.4. Structural Comparison of GEMs 89
6.5. Comparison of Experimental PD Models 93
7. CONCLUSION 96
8. RECOMMENDATIONS 97

REFERENCES 98

BIOGRAPHY 119
APPENDİX 120

LIST of ABBREVIATIONS

Abbreviations		Descriptions
Ab	:	Amyloid-Beta
ABC	:	ATP-Binding Cassette
acetyl-CoA	:	Acetyl Coenzyme A
AD	:	Alzheimer's Disease
ALS	:	Amyotrophic Lateral Sclerosis
ATP	:	Adenosine Triphosphate
BBB	:	Blood Brain Barrier
CBF	:	Cerebral Blood Flow
CNS	:	Central Nervous System
COBRA	:	Constraint-Based Reaction and Analysis
DA	:	Dopamine
DNA	:	Deoxyribonucleic Acid
ER	:	Endoplasmic Reticulum
FADH2	:	Flavin Adenine Dinucleotide
FBA	:	Flux Balance Analysis
GEO	:	Gene Expression Omnibus
GEM	:	Genome-Scale Metabolic Model
G6P	:	Glucose-6-phosphate
Gln	:	Glutamine
Glu	:	Glutamate Dehydrogenase
GS	:	Glutamine Synthetase
GSH	:	Glutathione
HD	:	Huntington Disease
IDH1	:	Isocitrate Dehydrogenase 1
IDH2	:	Isocitrate Dehydrogenase 2
INIT	:	Integrative Network Inference for Tissues
iMAT	:	Integrative Metabolic Analysis Tool
L-DOPA	:	l-3,4-Dihydroxyphenylalanine
MOMA	:	Minimization of Metabolic Adjusment

MPTP	:	1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine
mRNA	:	Messenger RNA
mtDNA	:	Mitochondrial DNA
NADH	:	Nicotinamide Adenine Dinucleotide
NMR	:	Nuclear Magnetic Resonance
Nnt	:	Nicotine Amide Transhydrogenase
Nrf2	:	Nuclear Factor Erythroid-2-Related Factor 2
6-OHDA	:	6-Hydroxydopamine
PCA	:	Principal Component Analysis
PD	:	Parkinson's Disease
Prp	:	Prion Protein
RMA	:	Robust Multi-Array Average
RNA	:	Ribonucleic Acid
RNA-seq	:	RNA-Sequencing
SN	:	Substantia Nigra
SNpc	:	Substantia Nigra Pars Compacta
TCA	:	Tricarboxylic acid cycle
TH	:	Tyrosine Hydroxylase
tSNE	:	t-distributed Stochastic Neighbor Embedding

LIST of FIGURES

Figure No:		Page
2.1:	Nutrient transport across the blood–brain barrier.	4
2.2:	Metabolic pathways of the glucose consumption in the neurons and astrocytes.	7
2.3:	Brain regions affected in neurodegenerative diseases.	8
2.4:	Pathways involved in the development of PD are interconnected.	10
2.5:	Dopamine metabolism in neurons and microglia.	13
2.6:	A timeline depicting the reconstruction of all genome-scale metabolic models.	20
2.7:	Global human metabolic network applications.	22
2.8:	Gene-protein-reaction (GPR) rules explain genotype-phenotype relationships.	26
2.9:	Constraint-based modelling defines the space of solutions by using constraints.	29
2.10:	Families of context-specific reconstruction algorithms.	34
3.1:	Example boxplot for the GSE74382 dataset A) before and B) after quantile normalization.	42
3.2:	Example PCA for the GSE24233 dataset. A) Before removing outlier. B) After removing the outlier. Blue circle indicates the outlier sample. The x-axis corresponds to the first principle component and the y-axis corresponds to the second principle component. Percentages in parentheses refer the variance explained by each component.	45
3.3:	Gene-score distribution of the 11 PD-control comparisons.	48
3.4:	Gene-Score based hierarchical clustering analysis of datasets.	49
3.5:	Enrichment analysis of most variable genes.	50
4.1:	GEM reconstruction steps followed in this thesis study.	52
4.2:	Pathway-Reaction distribution in iBrain766-Rn.	60
5.1:	Flux-distribution predictions of iMAT-based healthy models. Each bar on the graph depicts the predicted flux ratio for a different comparison.	75

5.2: Flux-distribution predictions of INIT-based healthy control models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 76

5.3: MOMA analysis on iMAT based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 78

5.4: MOMA analysis on INIT based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 79

5.5 LseiFBA analysis on iMAT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 80

5.6: LseiFBA analysis on INIT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 80

5.7: mLseiFBA analysis on iMAT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 81

5.8: mLseiFBA analysis on INIT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison. 82

6.1: Comparison of healthy-state flux distributions between iMAT and INIT algorithms. X axis refers to the marker reactions and their experimental flux values while y axis indicates the predicted fluxes by the context-specific models. Red points indicate the outliers. 86

6.2: Flux distribution predicted for Parkinson's disease by different algorithms. Boxplots compare flux predictions by LseiFBA, mLseiFBA and MOMA approaches. Red dots refer to the outliers. 88

6.3: Comparison of reaction numbers between iMAT-based and INIT-based models. Each point represents one of 11 comparisons. 90

6.4: Reaction similarity Jaccard distance among models. 91

6.5: tSNE analysis of model reaction contents. 92

6.6: Subsystem coverage of models. 93

6.7: Hierarchical clustering analysis of PD models based on flux fold changes. 94

6.8: Most variable pathways based on flux fold changes. 95

LIST of TABLES

Table No:		Page
2.1:	Distribution of datasets present in GEO (as of July 7, 2021).	18
2.2:	Genome-scale metabolic network reconstructions of the human body.	21
3.1:	Rat transcriptome datasets.	39
3.2:	Removed outlier samples.	45
4.1:	Reactions with updated EC number.	55
4.2:	Examples of cases where a new gene is added to the Rat brain. specific GEM model reconstructed in this study.	56
4.3:	Amino acid coefficients of macromolecule synthesis reaction.	57
4.4:	Lipid coefficients of macromolecule synthesis reaction.	58
4.5:	RNA coefficients of macromolecule synthesis reaction.	58
4.6:	Carbohydrate coefficients of macromolecule synthesis reaction.	58
4.7:	Content of macromolecule synthesis reaction.	58
4.8:	Detailed information of the iBrain766-Rn model content.	59
4.9:	Experimental resting-state CMRGlucose values for rat.	61
4.10:	The constraints utilized to simulate resting state flux distributions with iBrain766-Rn model. Glucose, oxygen, amino acid, and ammonia consumption/uptake rates and macro molecule synthesis reaction rates are given in the form of μmol/g dry tissue/min. A is Astrocytes, N is Neurons.	63
4.11:	Comparison of predicted rat flux ratios with the human experimental fluxes and predicted flux ratios for human and mouse at resting state.	65
4.12:	Calculations of percentage of relative flux rates with respect to glucose uptake rate.	66
5.1:	Healthy Control and Disease State Constraints.	69
5.2:	Reaction and metabolite content of iMAT models.	71
5.3:	Reaction and metabolite content of INIT models	73

1. INTRODUCTION

The human brain is the body's most complex organ. It needs a significant amount of energy to maintain its function. At rest, the brain consumes 20% of an individual's energy despite the fact that it only accounts for 2% of total body weight. Metabolism of brain is a dynamic process and highly susceptible to energy deprivation.

The brain primarily consists of neurons and glial cells including astrocytes, oligodendrocytes and microglia [1, 2]. The neurons use 70-80 percent of the energy, while the remaining energy is used by the brain's glial cells. Glucose is considered to be the essential energy substrate of an adult brain, and it needs a continuous flow of glucose across the blood-brain barrier (BBB); but, under energetic stress, additional sources of energy may be required to maintain brain function. A cerebral blood flow (CBF) system ensures continuous nutrition and oxygen delivery and elimination of product waste [3].

Parkinson's disease (PD) is the second most prevalent neurodegenerative disease, characterized by the loss of substantia nigra dopaminergic neurons and the development of Lewy bodies. PD affects 6 million individuals worldwide [4, 5] and has a global prevalence of 1% of the population over the age of 65. Parkinson's disease is a complicated illness with a wide range of symptoms. Despite extensive studies on PD, the processes underlying the disease remain unclear, and there is no exact treatment for PD. There is growing evidence that metabolic alterations influence the development and progression of PD [3]. Studies have shown reduced glucose consumption in brain regions affected by PD.

PD animal models have been extensively utilized to study disease etiology and molecular processes. Not only do animal models offer critical information, they also enable the testing of new treatment methods [6]. Neurotoxin models and genetic models are two prominent animal models used in the investigation of PD [6]. The model category of neurotoxins comprises animal models produced utilizing chemical injection. The genetic model category includes animal models created via genetic modification.

Due to the complexity of PD, a comprehensive depiction of PD requires processing of huge amounts of data. Systems biology enables the integration of such massive amounts of data and their efficient assessment.

Human metabolism is very complex, with numerous genes and metabolites that are all closely linked with one another. Genome-scale metabolic models (GEMs) are widely employed in systems biology to show the connection between genotype and phenotype and therefore detect specific metabolic characteristics of diseases [7]. GEMs have also been utilized to model the phenotypic characteristics of cells under a range of environmental and genetic conditions, explain particular disease-related phenotypes, identify therapeutic targets, and study the relationship between different species and cell types.

Human-GEMs enable the reconstruction of other mammalian GEMs through a homology-based method. By mapping homologous genes on the human network, many mammalian GEMs have been built [8]. The human, mouse, and rat genomes all have about the same amount of genes. The rat is one of the most often utilized preclinical model organism for neurodegenerative disorders, and the interpretation of neurodegenerative diseases in rats is critical for disease diagnosis and therapy. Although brain specific human GEM and mouse GEM are present [9–11], such a model is not available for rat.

Along with this study, it was aimed to:

i) Perform hierarchical clustering analysis of transcriptome datasets from PD rat models to investigate the similarity between these datasets in terms of metabolic gene expression patterns (Chapter 3)

ii) Reconstruct the rat brain-specific genome-scale metabolic model for the first time in literature (Chapter 4)

iii) Integrate transcriptome datasets derived from PD rat models into the newly reconstructed rat brain-specific genome scale metabolic model to obtain condition-specific models, and (Chapter 5)

iv) Perform constraint-based analysis on these condition-specific models to investigate metabolic alterations associated with PD in terms of flux values (Chapter5)

2. BACKGROUND INFORMATION

2.1. Energy Metabolism in the Brain

In human nervous system brain is the major central organ. The human brain is the most complex organ of the body. It mainly comprises three regions: the cerebrum, the cerebellum, and the brainstem. There are numerous routes for inputs and outputs in the brain, and they are arranged in a somatotopic fashion, making it a complicated and diverse organ.

All brain areas are thought to be metabolically active, despite structural variability. The metabolic activity of the brain is characterized by the oxidation of glucose to carbon dioxide and water (oxygen consumption), resulting in the generation of energy rich molecule known as Adenosine Triphosphate (ATP). Neural circuit and network activities involving particular tasks control local energy demand. In other words, variations in pathway activities induce changes in metabolism to produce ATP and changes in blood flow rate to supply fuel [12]. Brain needs a substantial quantity of energy to regulate and maintain its metabolic activities. The brain utilizes 20% of an individual's energy consumption (or 25% of glucose and 20% of oxygen consumption) at rest, while accounting for just 2% of total body weight. Brain is more resistant to energy deprivation than other organs [13]. Supplying the right amount of nutrients and oxygen to the brain is important for its proper functioning. The brain uses the energy to maintain essential cellular activities such as protein, nucleotide, lipid synthesis, and signalling mechanisms. Glucose is regarded as the compulsory energy substrate of an adult brain, which requires constant blood-brain barrier (BBB) supply. Brain energy metabolism is a dynamic, tightly controlled process. Blood flow and glucose consumption rates in various brain regions are strongly linked and typically vary in parallel throughout various physiological states. The brain mainly consists of two cell types, which are neurons and glial cells [1, 2]. Glial cells include astrocytes, oligodendrocytes, and microglia. About 70-80% of the energy produced in brain is consumed by the neurons and the rest of the energy is utilized by the glial cells [1, 2].

Continuous nutrient, and oxygen supply and removal of waste products are maintained by a cerebral blood flow (CBF) mechanism in the brain [3]. Cerebral

metabolic rate (CMR) is equivalent to the cerebral blood flow rate (CBF) multiplied by the arterial–venous (A–V) difference in glucose concentration:

$$\mathrm{CMR} = \mathrm{CBF} \times (C_A^{GLC} - C_V^{GLC}) \tag{4.1}$$

The inflow of the metabolites like glucose, oxygen, ketones and amino acids is controlled by the BBB (Figure 2.1) [3]. The transport method of particular molecules across the BBB is determined by the type of the solutes being transported. There are two important families of transporters recognized in the BBB, known as the solute carrier (SLC) proteins and the ATP-binding cassette (ABC) [3].

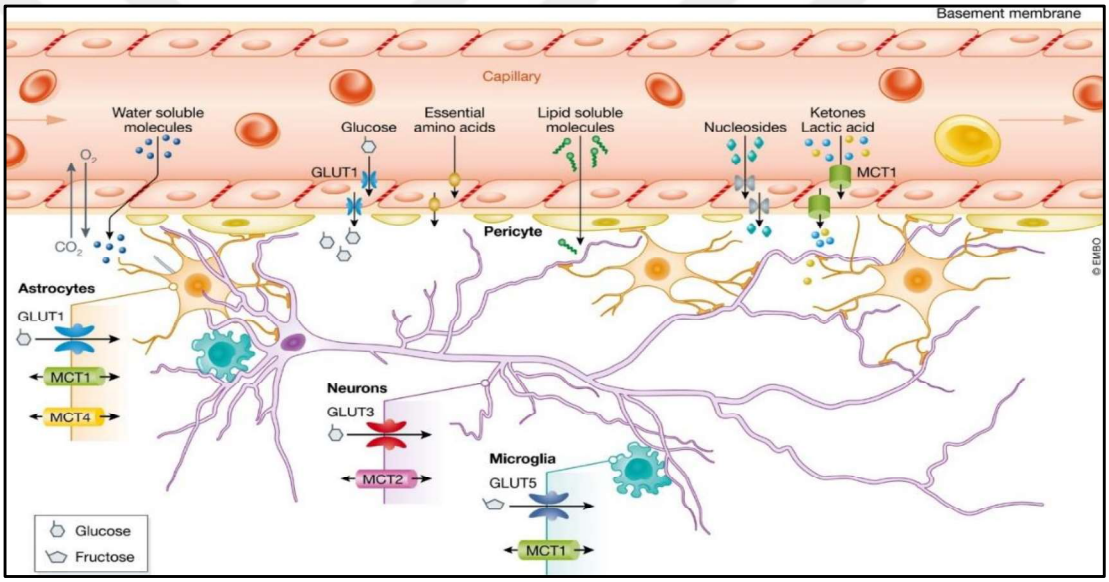

Figure 2.1: Nutrient transport across the blood–brain barrier.

2.1.1. Glucose-driven Metabolic Pathways in Neurons and Astrocytes

In the brain, cell types and selective expression of metabolic enzymes are responsible for the metabolic fate of glucose. Neurons are highly oxidative metabolic energy consumers compared to glial cells. It has been reported that neurons use more than 80 % of their energy at synapses for the maintenance of ionic gradients and restoration, the absorption and the recycling of the neurotransmitters. Neurons are assumed to predominantly metabolize glucose through the pentose phosphate pathway (PPP) to generate the reducing equivalents needed to sustain antioxidant defence by

producing nicotinamide dinucleotide phosphate (NADPH) [14]. NADPH has several functions ranging from the antioxidant defence in neuron and glial cells and mediating reactive oxygen species (ROS) formation to serving as a reductant for lipid and nucleic acid syntheses and neurotransmitter and aldehyde metabolisms. NADPH may originate from folate metabolism, malate dehydrogenase, isocitrate dehydrogenase (IDH1 and IDH2), glutamate dehydrogenase (Glu) and Nicotine amide transhydrogenase (Nnt).

Neurons do not have 6-phosphofructose-2-kinase/fructose-2,6 -bisphosphatase-3 (PFKFB3) enzyme, known as a glycolytic activator and degraded continuously by the ubiquitin-proteasome (UPS) pathway [15]. Neurons have a lower glycolytic rate than astrocytes. Since neurons have low Pfkfb3 activity [15], they are unable to increase glycolysis in response to stress [16]. Glycolysis is stimulated in astrocytes by activation of PFKFB3 through AMPK (adenosine monophosphate-activated protein kinase) [17].

Most brain synapses release glutamate as the excitatory neurotransmitter. Glutamatergic neurotransmission occurs via close metabolic interaction between glutamatergic neurons and astrocytes. The glutamate concentration is high during neurotransmission in the synaptic cleft. Astrocytes serve a critical role in the modulation of neuronal excitability by eliminating released neurotransmitters rapidly. Since glutamate receptor over-stimulation is very damaging to neurons, this phenomenon is particularly important for glutamate (Glu) functioning as the principal excitatory neurotransmitter in the brain. Glu enters into astrocytes through Na+-dependent excitatory amino acid transporters [18], and is converted to glutamine (Gln) by the astrocyte-specific enzyme Gln synthetase (GS). Glutamine also serves as the precursor for the formation of GABA in GABAergic neurons. Glutamate is the primary excitatory neurotransmitter whereas gamma aminobutyrate (GABA) is the primary inhibitory neurotransmitter in the central nervous system [19]. Glutamate and GABA homeostasis is maintained by a variety of processes, including synthesis, degradation, and transport. These systems all play a role in the preservation of the balance between excitement and inhibition, which is necessary for proper brain function. Glutamatergic and GABAergic neurons cannot restore their neurotransmitter pools unless the nearby astrocytes provide precursors. This phenomenon is known as glutamate-glutamine cycling (Glu-Gln).

Gln is delivered via glutaminase (GLS) activity to the neurons (Glu-Gln cycle) to replace the glutamate (Glu). Glu may also be utilized to produce antioxidant glutathione (GSH) [20]. In contrast to astrocytes, neurons have poor ROS-defense mechanisms. Due to the activation of the antioxidant response through the transcription factor nuclear factor erythroid-2-related factor 2 (Nrf2), astrocytes have larger quantities of the endogenous antioxidants, making them more resistant against the oxidative damage [21]. Bell et al. (2015) reported that neuronal Nrf2 is epigenetically silenced and activation of the nuclear factor Nrf2 pathway does not appear to restore its antioxidant protection [22]. Although both neurons and astrocytes may produce GSH, neurons are dependent on the availability of GSH precursors. Astrocytes also regulate the neuronal redox environment through the Glu-Gln cycle and GSH export. The NADPH works as a cofactor for the reconstruction of reduced glutathione (GSH) (Figure 2.2) and thioredoxin reductase. This process is helpful for the maintenance of antioxidant capacity of neurons.

Glycolysis is the process for the metabolism of glucose to pyruvate. A total of 30-36 molecules of ATP as well as water and carbon dioxide molecules are produced depending on the ratio of proton leakage in the mitochondria by the oxidative phosphorylation, the tricarboxylic acid (TCA) cycle and the pentose phosphate pathway (PPP) cycle (Figure 2.2) [3]. The conversion of glucose into pyruvate and then further transformation into the acetyl coenzyme A (acetyl-CoA) is maintained by the glycolytic pathway. Acetyl-CoA can be carried actively to the mitochondria. NADH (Nicotinamide adenine dinucleotide) and FADH2 (Flavin adenine dinucleotide) produced during glycolysis and the TCA cycle are subsequently reoxidized in the electron chain transport (ETC). ETC uses the energy generated by electron transfer through its numerous facilities to transport protons into intermembrane space through the inner mitochondrial membrane. The energy molecule of ATP is produced from ADP with the help of ATP synthetase enzyme at this step, which is responsible for reintroducing protons into the mitochondrial matrix.Astrocytes use glucose in a manner similar to neurons. While some Glucose-6-phosphate (G6P) is used in glycogen synthesis and PPP, its primary metabolic pathway is glycolysis, which results in the low mitochondrial oxidation rate and generation of lactate molecule[23].Astrocytes have a distinct metabolic profile as a consequence of their unique expression of enzymes and many other transporters. In addition,

phosphorylation levels of PDH under baseline circumstances are high, since they may effectively restrict conversion to Acetyl-CoA, owing to strong PDK expression [24].

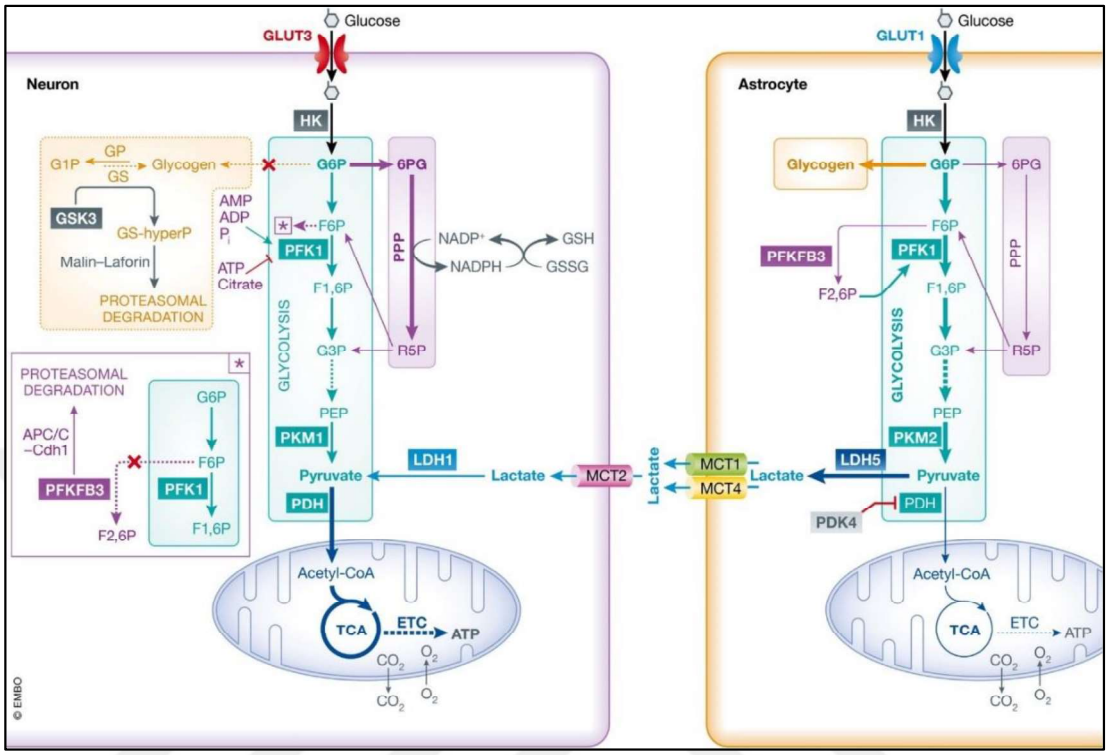

Figure 2.2: Metabolic pathways of glucose consumption in neurons and astrocytes.

2.1.2. Altered Metabolism in Neurodegenerative Disorders

Neurons are basic functional units of the brain and central nervous system (CNS). Typically, neurons cannot regenerate themselves when they are degenerated. Neurodegenerative disorders are defined by the gradual progressive loss of neurons in the central nervous system (CNS), resulting in deficiencies in specific brain activities (e.g., memory, mobility, cognition) conducted by the affected CNS region [25]. Neurodegenerative diseases leading to disruptions in the nervous system are an increasing public concern. According to World Health Organization (WHO), by mid-century, neurodegenerative disorders will become the world's second biggest cause of death, overcoming cancer [26]. Neurodegenerative diseases have been implicated as being derived from multifactorial conditions ranging from abnormal degree and aggregation of defective proteins, oxidative stress and free radical formation, impaired bioenergetics and mitochondrial dysfunction, to exposure to metal toxicity and pesticides [27]. Primary clinical features/symptoms (such as dementia, parkinsonism,

and motor disease), anatomic distribution of neurodegeneration, and molecular aberration are used to classify neurodegenerative diseases [28].

Although there are a significant number of neurodegenerative diseases, most of these can be classified based on the accumulation of amyloid-beta (Ab), tau, a-synuclein, prion protein (Prp), and transactive response DNA-binding protein 43 (TDP-43) [29]. Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD) and amyotrophic lateral sclerosis (ALS) are the most prevalent neurodegenerative diseases. These most prevalent neurodegenerative diseases differ in specific protein accumulation, affected brain regions, and the type of vulnerable neurons [30–32]. Striatal neurons in PD, hippocampal and cortical neurons in AD, and spinal motor neurons and cortical neurons in ALS are affected (Figure 2.3) [33] .

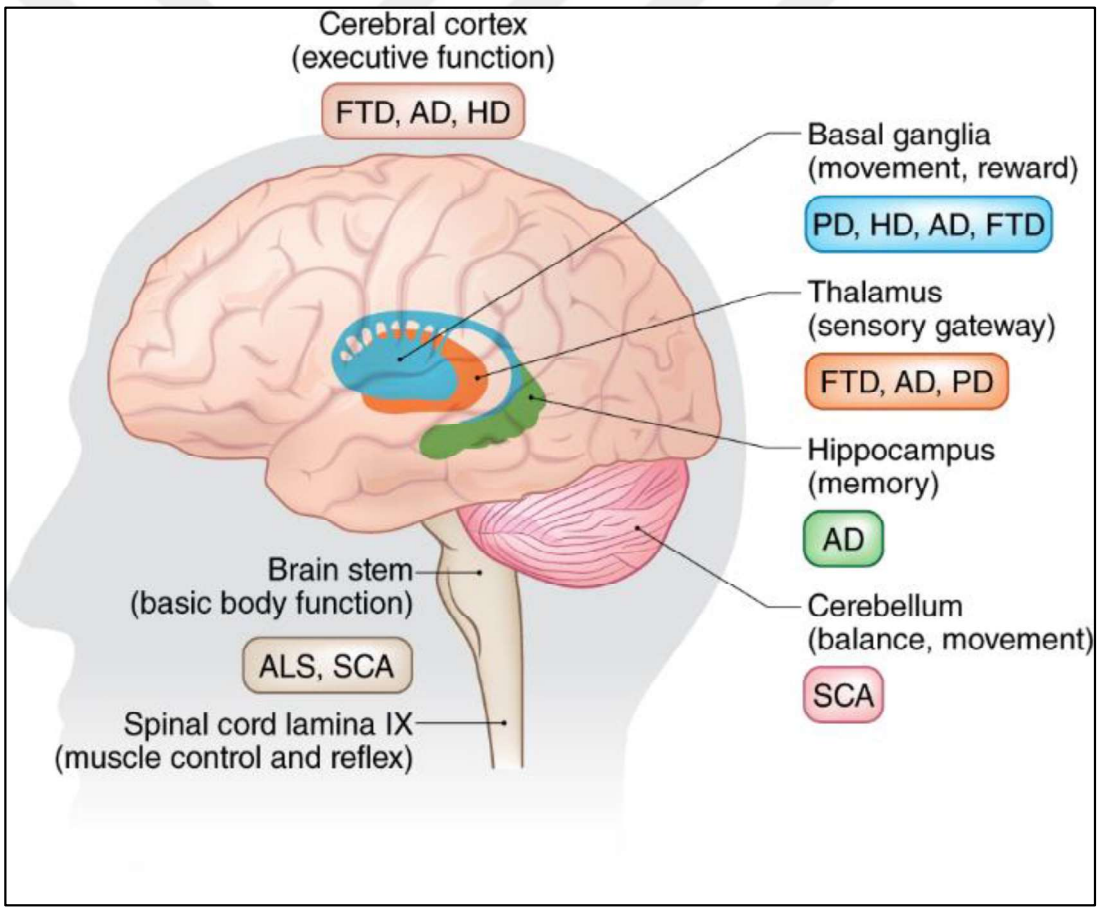

Figure 2.3: Brain regions affected in neurodegenerative diseases.

Even though each neurodegenerative disease has its own specific molecular mechanisms, there are various molecular and cellular abnormalities that are shared by neurodegenerative diseases including protein misfolding and aggregation,

mitochondrial dysfunction, the toxicity of glutamates, calcium burden, proteolytic stress, oxidative stress and free radical formation, neuroinflammation, and aging, and neuronal death [33, 34]. It is known that neurodegeneration usually begins in the subcortical regions and spread to the cortical regions as the disease progresses [35]. Mitochondrial dysfunction, protein misfolding, and oxidative stress are thought to be direct consequences of metabolic disorders/alterations [36]. It was shown that metabolic alterations have a prominent role in the initiation and progression of neurodegenerative diseases [3]. Imaging studies showed reduced amount of glucose in brain areas affected by AD, PD, ALS, and HD patients [37].

2.2. Parkinson's Disease (PD)

PD affects ~6 million people [4, 5] in the World, and it has a prevalence of 1 % of the population over 65 ages globally [38, 39]. PD is the second most common progressive age-related neurodegenerative disease, which is characterized by the loss of dopaminergic neurons in substantia nigra and the presence of Lewy bodies (misfolded a-synuclein) [38, 40]. PD is a heterogenous disorder involving motor and non-motor symptoms. It mainly occurs due to the deterioration of cells in the brain that produce dopamine [41], which regulates brain cells to have communication with each other. In the brain, dopamine-producing cells are responsible for the regulation of movement. In addition to dopaminergic neurons, other neurotransmitter systems are also affected in PD, including cholinergic, noradrenergic and serotonergic neurons [41, 42]. PD is clinically diagnosed with motor symptoms: movement disorder (bradykinesia), tremor at rest, and rigidity. Generally, non-motor symptoms such as sleep disorders, cognitive impairment, dementia, constipation, and depression accompany motor-symptoms in the later stage of PD [43]. Although there is no exact treatment for the prevention and treatment of Parkinson's disease, current treatment methods aim to reduce the symptoms of the disease [43].

2.2.1. The Genetics of Parkinson's Disease

Parkinson's disease can be classified into two types, idiopathic (sporadic) Parkinson's disease and familial Parkinson's disease, according to the etiology of the disease. Most of the PD patients (over 90 % of PD cases) are sporadic, but it is known that hereditary factors also play a role in the disease's etiology [44]. The underlying etiology of the idiopathic form of the disease is uncertain. Several factors, including genetic abnormalities and environmental situations, may contribute to the disease's development.

Mutations in numerous genes causing hereditary forms of PD (*SNCA*, *Parkin*, *PINK1*, *LRRK2*, *DJ-1*) lead to mitochondrial dysfunction [45]. Mitochondria are essential in the synthesis of cellular energy through the respiratory chain. Figure 2.4 [46] shows the PD associated pathways. Mitochondrial dysfunction in the SN (Substantia Nigra) is a critical component of Parkinson's disease pathophysiology (Figure 2.4) [47, 48].

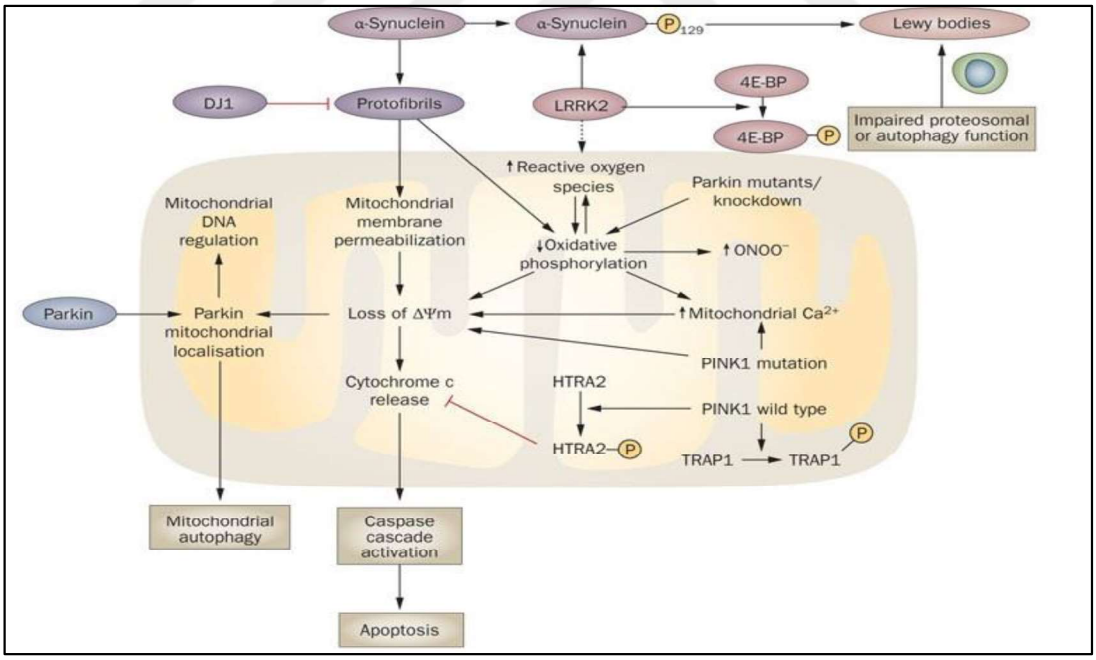

Figure 2.4: Pathways involved in the development of PD are interconnected.

The alpha-synuclein protein, encoded by *SNCA*, is involved in controlling neurotransmitter release from dopaminergic neurons' presynaptic terminals [49]. α-synuclein is a critical component of Lewy bodies, which are a characteristic of Parkinson's disease. Due to the existence of many enzymatic domains, *LRRK2* has been implicated with various physiological activities and signalling cascades, including neurite outgrowth, mitochondrial function, complex modulation retromer, and autophagy control [50, 51]. Parkin protein localizing in mitochondria controls the function and the shape of mitochondrial DNA (mtDNA). Parkin is also attracted preferentially to impaired mitochondria, resulting in the organelle's autophagy degradation (mitophagy) [52]. An N-terminal mitochondrial targeting region and a transmembrane domain are encoded by the *PINK1* gene, which is a serine/threonine protein kinase [53]. The interaction of *PINK1* with parkin was shown to promote selective autophagy of depolarized mitochondria and preservation of mitochondrial quality [54]. Mutations in the *PINK1* gene lead to decreased mtDNA levels, impaired mitochondrial calcium handling, disrupted ATP production, and increased free radical levels. As a result, the mitochondrial membrane potential decreases, and neuronal cells become more susceptible to apoptosis [55, 56]. *DJ-1* has been proven in model systems to protect DA neurons against various stressors, including mutant synuclein, rotenone, 6-hydroxydopamine, and hydrogen peroxide [57].

2.2.2. Parkinson's Disease Metabolism

There is mounting evidence that metabolic changes have a significant impact on the onset and course of neurodegenerative diseases. Studies of positron emission tomography imaging showed reduced glucose consumption in brain regions affected by AD, PD, ALS, and HD patients [58]. Moreover, decreased levels of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, the fundamental enzymes for PPP, were reported in the putamen and cerebellum of PD patients at an early stage [59].

O_2^- is produced in the mitochondria via electron leakage from Complex I [60], while in both the matrix and the inner membrane space it is produced by electron leakage from Complex III [61]. It has been implicated that dopaminergic neurons are very susceptible to Complex I inhibitors [62]. There are several investigations

indicating antioxidants protect dopaminergic cells against Complex I inhibitors' toxicity [63, 64]. However, complex I inhibition-induced dopaminergic cell death cannot be directly connected to ROS formation or entirely avoided by antioxidants [65]. Further investigations have shown that the Complex I inhibitors rotenone and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine may be the primary cause of neuronal cell death produced by these agents [66].

Dopaminergic neurons make up a small fraction of brain neurons. Nevertheless, this type of nerve cells is involved in a wide range of mental and physical activities, including behaviour and cognition, voluntary movement, motivation and reward, learning and sleep mood [67]. Dopaminergic SNpc (substantia nigra pars compacta) neurons are vulnerable to energy failure as they utilize a considerable amount of energy in order to maintain a basal dopamine (DA) tone in striatum brain region. Notably, there is no difference in the intrinsic bioenergetic capacity of dopaminergic and non-dopaminergic presynaptic synaptosomes, indicating additional factors than energy dysregulation contribute to dopaminergic cell loss [68]. Though neurons have a limited capacity to enhance glycolysis, increasing glycolysis compensates for energy deficiency and dopaminergic cell death caused by the mitochondrial dysfunction/malfunction [69]. Parkinson's disease is defined by the selective death of a subpopulation of neurons in the mesencephalon that utilize dopamine as a neurotransmitter to communicate with one another. In the case of dopaminergic neuron loss, Dopamine may cause cytotoxicity due to being highly reactive once it is not appropriately stored and metabolized.

DA levels produced in SNpc are tightly controlled by maintaining a balance between production, synaptic vesicle loading, absorption from the extracellular environment, and catabolic breakdown. Tyrosine hydroxylase (TH) and aromatic amino acid decarboxylase (AADC) take role in the synthesis of DA in cytoplasm. DA's two-step synthesis takes place in catecholaminergic neuron cytoplasm and starts with tyrosine hydroxylation of L-tyrosine to generate DOPA (Figure 2.5) [70]. This oxidation is strictly regulated and needs the cofactor tetrahydrobiopterin (BH4), produced by GTP cyclohydrolase from guanosine triphosphate (GTP) (GTPCH). Then, aromatic amino acid decarboxylase converts DOPA to DA (AADC). TH, GTPCH, and AADC enzymes associated in DA production play a crucial role in preventing excessive oxidative stress [70]. TH, GTPCH and AADC enzymes involved in DA production play a significant role in the management of DA homeostasis to

avoid from excessive oxidative stress (Figure 2.5). The cause of DAergic neuronal death is still being investigated. As a result, there are no medications available that may halt, diminish, and prolong the death of these neurons. Current therapy to replace DA loss is based primarily on the administration of l-3,4-dihydroxyphenylalanine (L-DOPA), a DA precursor. Since the administration of L-DOPA requires a specific number of active DAergic neurons to transform this metabolic precursor into DA, progression in PD negatively affects the L-DOPA treatment efficacy [71]. MAO's oxidative deamination results in the production of hydrogen peroxide as well as the reactive 3,4-dihydroxyphenylacetaldehyde (DOPAL). MAO and COMT inhibition have been employed in PD therapy to improve DOPA and DA availability [70, 72].

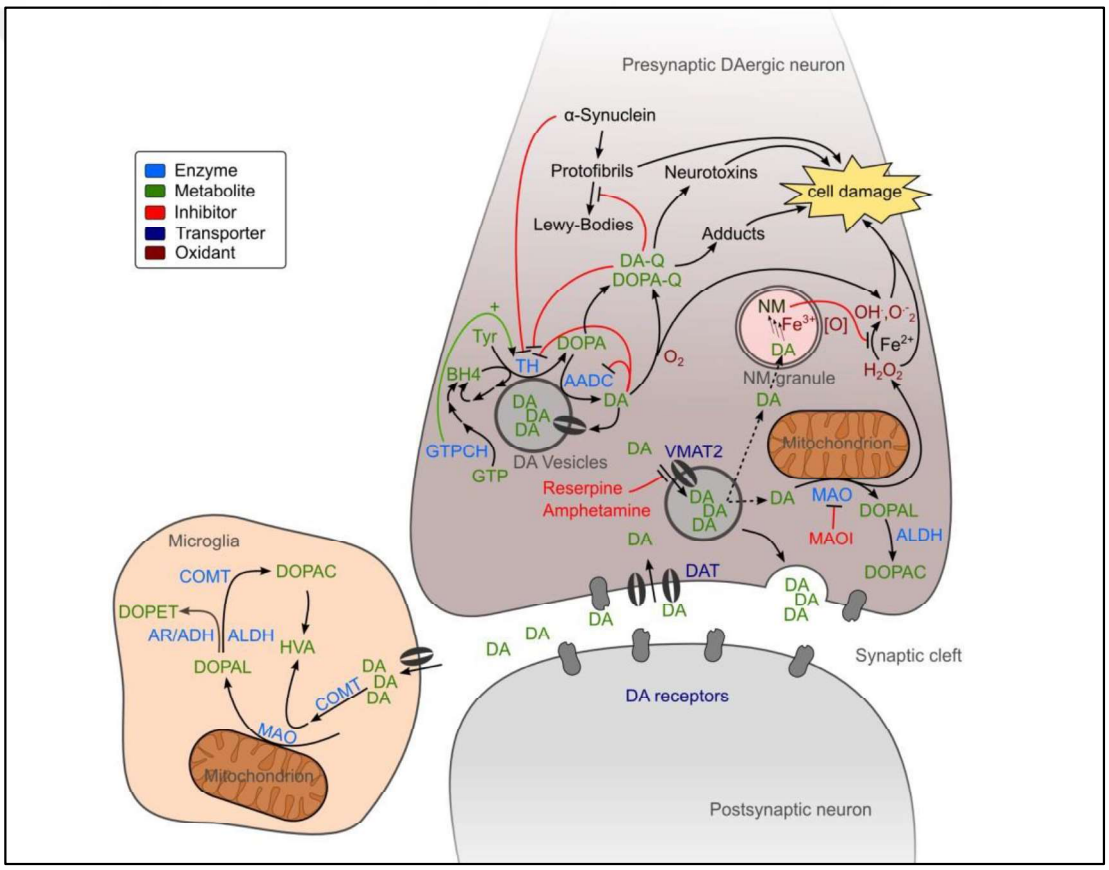

Figure 2.5: Dopamine metabolism in neurons and microglia.

The cerebral metabolic measurements indicate that the PD state alters the activity of many metabolic pathways, as well as the consumption and synthesis of a few key metabolites. Glucose and oxygen uptake is the most critical alteration. This alteration impacts basic metabolic processes such as glycolysis, TCA cycle, and oxidative phosphorylation. PET measurements show that cerebral glucose consumption

decreases by 2-32 % while oxygen consumption decreases by 6-34 % in Parkinson's disease [73]. It has also been shown that genes involved in the TCA cycle and oxidative phosphorylation are downregulated in patients with Parkinson's disease [74]. As determined by Nuclear Magnetic Resonance (NMR) spectroscopy, increased lactate synthesis in the brain is another indicator of PD state shift [75].

2.2.3. Experimental Models for Parkinson's Disease

Neurological disorders in humans may be modelled in animals utilizing standardized protocols that replicate certain pathogenic events and their behavioural consequences. Animal neurological impairment models are useful for assessing novel therapy alternatives [76]. PD animal models have been widely used to understand the pathogenesis and molecular mechanisms of the disease. Animal models provide not only essential information but also the opportunity to test novel treatment techniques [6]. Animal PD models are mainly categorized into two groups: toxin models and genetic models [77].

The toxin models referring the classic experimental models of PD aim to reproduce the clinical and behavioural changes of human PD in rodents or primates by utilizing neurotoxins, which cause nigrostriatal degeneration. Neurotoxins may be employed systemically or locally based on the type of agent and the target species [6]. Toxin based models mainly consist of 6-hydroxydopamine (6-OHDA), 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), rotenone, paraquat and methamphetamine [77].

6-OHDA is a selective neurotoxin initially discovered to create damages in rat nigrostriatal DA neurons [78] and has been demonstrated to act in other species such as mice [79]. 6-OHDA is widely employed to study motor and metabolic dysfunctions in Parkinson's disease. 6-OHDA accumulates in the cytosol and promotes the production of hydrogen peroxide, other reactive oxygen species, and quinines [80]. Since 6-OHDA cannot enter the BBB, it is necessary to administer it directly into the Substantia Nigra pars Compacta (SNpc) region in the brain.

Certain pesticides, such as rotenone, paraquat, and maneb, are utilized as neurotoxins in the development of animal models for Parkinson's disease [81–83]. Pesticide exposure is a well-established risk factor for Parkinson's disease [84]. The

lipophilic nature of rotenone allows it to pass the blood-brain barrier easily. Chronic rotenone treatment leads to the loss of striatum dopamine terminals and the degradation of SN dopaminergic neurons [85]. However, rotenone has disadvantages in terms of leading to high mortality, phenotypic heterogeneity and non-PD related symptoms [86]. Paraquat, a mild inhibitor of mitochondrial complex I, has been shown to cause dopaminergic neuron degeneration in the SN [87]. Maneb, which is a complex III inhibitor, is also employed as a neurotoxin. It has been established that high dosages of methamphetamine cause selective DAergic, serotonergic nerve terminal and SNc neuronal death in mice and primates when administered subcutaneously [88].

Advancements in transgenic technology facilitate to simulate mutations in several genes, including α-synuclein, Parkin, PINK1 and DJ-1 [89]. Several transgenic α-syn mice have been created, and numerous behavioral abnormalities have been discovered. Yet, the majority of them did not have severe nigrostriatal degeneration [90]. The Parkin knockout animal is a well-established genetic model of Parkinson's disease used by many researchers to investigate the disease. Recently, it was suggested that regulating Parkin activity might be a possible treatment method. An improvement in mitochondrial quality was detected in Parkin knockout mice with an increase in Parkin gene activity [91]. Many PINK1 knockout mice have been utilized as Parkinson's disease animal models, and none of these models have exhibited Lewy-body formation. This observation is consistent with postmortem results in Parkinson's disease patients with the PINK1 mutation [92]. A reduction in dopamine release in the striatum is seen in DJ-1 knockout mice, although there is no loss of SNc dopaminergic neurons in these animals [93].

2.3. Systems Biology

Along with completing the human genome sequence in 2003 and the development of high-throughput technologies, a paradigm shift occurred in biology, which caused the emphasis to change from a gene-centric to a system-wide approach. Systems Biology is a multidisciplinary branch of research that aims to understand the complex interactions that occur inside biological systems. It provides a framework for constructing models of biological systems using systematic observations [94].

Systems Biology's evolution was assisted by the advent of new genome sequencing techniques and other high-throughput technologies that produced so-called "omics" data like genomics [95], transcriptomics [96], proteomics [97], metabolomics [98], and fluxomics [99]. Bioinformatic analysis of the omics data enabled the development of biological networks or models capable of simulating cell activity [100].

Biological networks/models are generally classified into three types:

i) Metabolic: Chemical processes that take place inside the cell are referred to be metabolic. Metabolic networks explain the consumption/synthesis of metabolites necessary for cell survival and growth.

ii) Regulatory: The representation of regulatory interactions between regulatory elements (e.g. transcription factors, promoters) and their target genes is the goal of regulatory networks.

iii) Signalling networks/models: Signals indicate responses or "signalling events" in a network that regulates how a cell reacts to the environment via information-flow cascades, such as phosphorylation or ubiquitination.

It is important to know that all of these networks are associated with each other, and combining them at various levels enables better knowledge of cellular activities. Yet, since it increases the complexity of analysis, combined analysis limits their inference based on genome-level [101].

The human brain's structure is highly complicated, and numerous studies have been conducted to investigate the brain structure and its function. PD is a heterogenous multifactorial complex disease. Though there is intensive research on PD, the mechanisms underlying Parkinson's disease (PD) remain unknown and also there is no exact treatment for PD. Huge quantities of data and frequently heterogeneous information must be integrated into a complete, holistic representation of PD. Systems biology facilitates the integration of such huge quantities of data and the effective evaluation of integrated data. Mathematical and computational techniques are utilized in systems biology to organize thousands of information pieces into functional structures that computers can evaluate in a highly efficient manner to elucidate molecular mechanisms behind health and disease [102]

2.4. Transcriptomics

A cell's transcriptome is defined as a collection of all RNA (Ribonucleic Acid) molecules (transcripts) found in that cell at the particular time of sampling. Transcriptomics data is one of the most commonly used omics data. Through the knowledge of the transcriptome, it is possible to interpret the functional components of the genome, disclose the molecular details of cells and tissues, and comprehend diseases. Different technologies, including hybridization or sequencing-based methods, have been developed to infer and measure transcriptomes [103].

Microarray technology is utilized to evaluate expression levels of hundreds of genes concurrently. Arrays comprising hybridization probes for known genes of the organism are produced through the microarray method. Custom-made microarrays or commercial high-density oligo microarrays are often used in hybridization-based methods, which include incubating fluorescently labelled cDNA with them. Relative gene transcript levels with their matching samples are measured using fluorescence intensity (mRNAs). In contrast to microarray techniques, sequence-based techniques identify the cDNA sequence in a straightforward manner.

Microarrays of DNA (Deoxyribonucleic Acid) [104] or other methods such as RNA sequencing [103] enable the expression of the genes in cells to be quantified under various circumstances. The Gene Expression Omnibus (GEO) [105, 106] is the most widely used database for gene expression data, which includes microarray, next-generation sequencing, and other high-throughput functional genomic data. Table 2.1 shows the amount of total datasets and expression profiling datasets that have been deposited to the GEO database.

Table 2.1: Distribution of datasets present in GEO (as of July 7, 2021).

	All Organisms	***Rattus norvagicus***	***Homo Sapiens***
Number of total datasets	156372	4597	68763
Expression profiling by array	63431	2719	26169
Expression profiling by high throughput sequencing	51293	856	19574

Transcriptome data collected from Parkinson's disease patients discovered differentially expressed genes and molecular dysfunctions in PD patients [107]. The first transcriptome study of PD considering various brain regions (substantia nigra, putamen, and prefrontal cortex) was conducted by Zhang and colleagues (2005) [108]. A one-way ANOVA revealed 329 significantly altered genes in the three brain regions. Furthermore, enrichment analysis highlighted several terms related to energy metabolism, including branched-chain amino acid aminotransferase Complex I, amino acid transporter, glutamate decarboxylation, oxidative phosphorylation, and citrate metabolism. Blood and skin samples have also been utilized to investigate PD, other than brain tissues. Individual studies emphasize different genes and pathways, but the most often mentioned pathways and processes in Parkinson's disease include dopamine metabolism, mitochondrial function, oxidative stress, protein degradation, neuroinflammation, vesicular transport, and synaptic transmission [72]. Transcriptomic analysis was also applied to animal models of PD. For example, in a rat model of L-Dopa-induced dyskinesia, a transcriptome method was used to analyze differentially regulated genes in striatal tissue, their particular pathways, and functional and disease related networks [109]. TGF1 was shown to be a significantly upregulated gene in the dyskinetic animals. In another study, transcriptome analysis on a 6-hydroxidopamine (6-OHDA) rat model revealed that, 2 and 4 weeks after the neurotoxic administration, 131 genes and 698 genes were substantially changed in substantia nigra region, respectively [110]. Enrichment analysis of these genes showed that the first time point (2 weeks) genes were associated with three processes: signal transduction, modification of extracellular matrix and inflammation. On the other hand, the second time point (4 weeks) genes were associated with five processes

including synaptic transmission, neuron projections, proper functioning of the soma and dendrites, and transmission of nerve impulse.

2.5. Genome-Scale Metabolic Networks

Metabolism is the entire collection of chemical processes that take place in the cell in order to sustain life. Enzymes are the key actors in this process because they catalyze the chemical processes. The enzyme–reaction relationships can be utilized to reconstruct a metabolic model of metabolism. Human metabolism is very complicated, including many genes and metabolites, all of which are highly interconnected with one another. The monitoring of metabolic alterations necessitates a comprehensive strategy that focuses on researching the metabolic network at the genome-scale level. In this regard, Genome-Scale Metabolic Models (GEMs) are commonly used in systems biology to illustrate the association between genotype and phenotype and therefore uncover particular metabolic features of diseases [7]. GEMs use a stoichiometric matrix to describe complicated cellular metabolic networks, allowing for the advanced mathematical study of metabolism at the whole-cell level [111]. GEMs are comprised of metabolites and reactions to describe all metabolic activities in cells.

The increase in the number of novel GEM reconstructions during the past ten years (Figure 2.6) [112] demonstrates their high use for the investigation of biological systems and associated applications. GEMs also have been used to explain specific disease-related phenotypes [113], identify therapeutic targets [114], and investigate the association between various organisms and cell types [115].

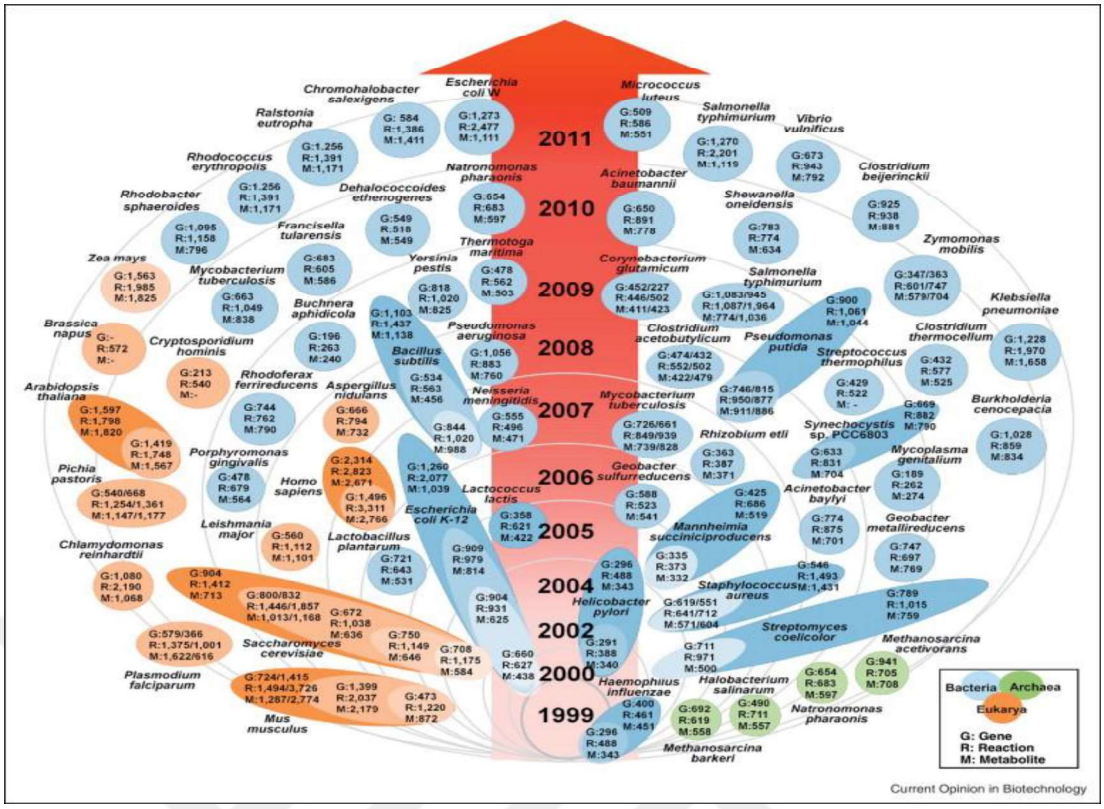

Figure 2.6: A timeline depicting the reconstruction of all genome-scale metabolic models.

2.5.1. Human Genome Scale Metabolic Models

Since many human diseases are associated with metabolic alterations, the biochemistry of the cells has been investigated through the reconstruction of the human GEMs in the context of systems biology [116]. Metabolic reconstructions enable biological information to be transformed into a mathematical representation and the following physiological state calculation [117, 118].

Recon 1, Recon 2, Recon 2.2, Recon 3D, HMR, HMR 2.0 and Human-GEM are the reconstructed major generic human GEMs so far. Because they contain information on all metabolic genes regardless of their expression level in a given tissue or cell type, generic GEMs do not address condition-specific metabolism. Recon 1, known as the first human generic GEM, was reconstructed at 2007 [119]. This metabolic model was reconsctructed based on a comprehensive collection and assessment of genomic and bibliomic data and includes 1496 genes, 2712 metabolites, and 3311 reactions, as well as metabolic and transport reactions. It separates the metabolites into seven compartments: cytosol, mitochondria, peroxisome, Golgi,

endoplasmic reticulum (ER), nucleus, lysosome. The model was validated via simulations of 288 well-characterized metabolic processes occurring in a variety of cell and tissue types. Recon 2 consists of 7440 reactions, including 4821 reactions linked with 2140 genes, and 2499 distinct metabolites spanning seven compartments [116]. Recon 2.2 [120] is an updated version of Recon 2, and it includes 5324 metabolites, 7785 reactions and 1675 genes. In comparison to the previous GEMs, Recon 3D has 10,600 reactions, with 5938 of them linked with 2248 genes, and 2797 distinct metabolites compretmentalized similar to Recon 2 with an extra compartment for the mitochondrial intermembrane space [121]. Human-GEM [122] is the most recently reconstructed global human genome-scale metabolic model. It includes 8378 metabolites, 13083 reactions and 3625 genes. Human GEMs have been extensively utilized in research including the identification of biomarkers [123], the generation of context-specific metabolic models [124] and the elucidation of the Warburg effect [125].

Table 2.2: Genome-scale metabolic network reconstructions of the human body.

Model Name	**Unique Metabolites**	**Unique Reactions**	**Genes**
Recon 1[119]	2712	3311	1496
Recon 2 [116]	2626	7440	1789
Recon 2.2 [120]	5324	7785	1675
Recon 3D [121]	4140	13543	3288
HMR [126]	3160	8100	3688
HMR2.0 [127]	3160	8181	3765
Human-GEM [122]	8378	13083	3625

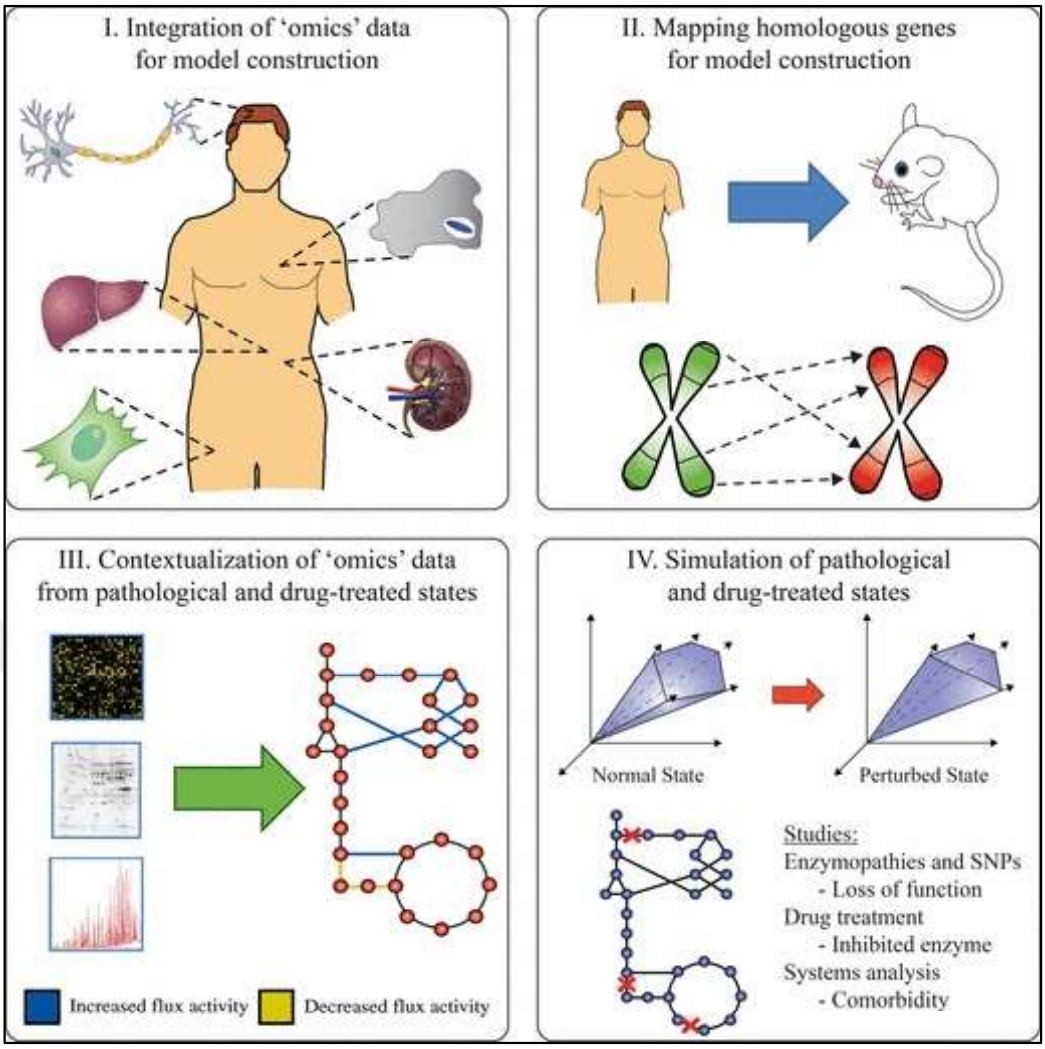

Figure 2.7: Global human metabolic network applications.

Human-GEMs provide an opportunity to reconstruct other mammalian GEMs by utilizing a homology-based approach. Several mammalian GEMs have been reconstructed by mapping homologous genes on the human network (Figure 2.7) [7]. Animal models, such as mice and rats, play an essential role in clinical and pharmaceutical research. The development of metabolic models of these model organisms, as well as the study of their metabolism, are critical in this regard [7]. There are several homologous genes between the genes in human GEM reconstructions and the animal genomes. For example, five mammalian GEMs were reconstructed including *Mus musculus*, *Rattus norvegicus*, *Canis lupus familiaris*, *Pan troglodytes* and *B.taurus* via a research of genes similar to Recon 1 in mammalian genomes [128].

Human, mouse, and rat genomes have similar number of genes. The rat is one of the widely used model organisms for neurodegenerative diseases in preclinical studies,

and the interpretation of these studies in rats is important for the diagnosis and treatment of these diseases. On the other hand, although humans and rats are in the mammalian class, there may be metabolic differences between them. It is known that there are some discrepancies between genes controlling metabolic function.

Numerous Human GEM reconstructions including thousands of genes have been published [116, 119, 126, 127] whereas only basic metabolic networks with a few of genes are known for the rat [129, 130]. iRno is the first comprehensive generic genome-scale metabolic model reconstructed for rat metabolism [131]. HMR2 was used as a template model to reconstruct the draft rat metabolic network based on orthology annotations. Then, manuel curation was performed to obtain the final iRno-GEM. iRno-GEM contains 5620 metabolites, 8268 reactions in eight compartments and 2324 genes. 14 of 8268 reactions are rat-specific reactions.

The first mouse generic-GEM was reconstrcuted by Sheikh et.al [132]. The model contains 872 metabolites, 1220 reactions, which were compartmentalized between cytosol, mitochondria, and extracellular space, and covered 473 genes. Several other mouse GEMS were reconstructed later, increasing network connection in several metabolic pathways. iMM1415 [128] , MMR [133], and iMM1865 [8] are existing generic mouse genome-scale metabolic models that were reconstructed by using human Recon1, HMR2, and Recon3D models respectively. iMM1865 is the most recently reconstructed generic mouse GEM and it includes 5839 metabolites, 10612 reactions compartmentalized between eight compartments and 1865 genes.

2.5.2. Brain-specific GEMs

Generic genome-scale metabolic models are not specific for a particular tissue, and the study of disease conditions linked with particular tissues necessitates the use of tissue-specific models. Tissue-specific genome-scale metabolic models are reconstructed by integrating generic models with high throughput data such as transcriptome data.

The interaction between neurons and astrocytic cells is important in terms of normal neuron function and brain energy metabolism. Astrocytes assist/support neuronal cells by exchanging metabolites with them, enabling them to sustain normal functions [134]. Brain-specific GEMs have been utilized to investigate brain

metabolism and neurodegenerative disorders. iNL403 [135], iMS570 [9], iBrain671-Hs [10], and iBrain674-Mm [11] are brain-specific GEMs that take into consideration two main brain cell types: astrocytes and neurons.

iNL403 is the first published human brain-specific GEM. Human Recon1 was used as the template to reconstruct iNL403. It includes 331 metabolites, 1070 reactions and 403 genes. The model was used to investigate Alzheimer's disease metabolism via transcriptome data integration and constraint-based modeling (CBM). iMS570 is a human brain-specific GEM reconstructed based on a previous brain-specific metabolic network that takes into account neuron-astrocyte compartmentation [136]. The previous model comprises 216 metabolites, and 217 reactions that occur in and between neuron and astrocyte cells. That model does not include gene information. iMS570 consists of 308 metabolites, 630 reactions, and 570 genes. iMS570 was only partially compartmentalized by separating NADH, NADPH and $FADH_2$ pools in mitochondria and cytosol. It was shown that iMS570 could correctly estimate resting state metabolic fluxes using CBM [9].

iBrain671-Hs is a human-brain specific GEM reconstructed based on the iMS570. Compared to other brain-specific GEMs, the main improvement of the model is that it separates all metabolites in the model into two compartments: cytosol, mitochondria. It contains 812 metabolites, 994 reactions and 671 genes. 470 of 994 reactions are neuronal, while 458 of them occur in astrocytes. Moreover, there are 22 reactions referring to metabolite exchange between astrocytes and neurons. It has been reported that the model could correctly simulate both resting state and Parkinson's disease state in terms of flux prediction by using CBM [10].

iBrain674-Mm, a brain-specific GEM, was reconstructed for mouse brain metabolism [11]. The reconstruction was carried out using iBrain671-Hs as a template through a homology-based approach. It also contains some mouse-specific reactions that are not present in the human reconstruction. iBrain671-Hs was used as the template also for the compartmentalization of iBrain674-Mm. The brain-specific mouse GEM consists of 814 metabolites, 992 reactions and 674 genes. There are 412 neuronal reactions and 447 reactions occurring in astrocytes.

Although there are brain-specific GEMs for human and mouse in the literature [9, 10, 137], such a model is not available for the rat.

2.5.3. Genome-Scale Metabolic Model Reconstruction

Genome-scale reconstruction of a metabolic network mainly consists of four steps [138]: i) Creation of draft model, ii) Manual curation of the draft model, iii) Conversion of curated model into a mathematical model, and iv. Evaluation and analysis.

i) Creation of draft model: The first stage in reconstructing a GEM is identifying and functionally annotating the genes in the genome that encode metabolic enzymes [139]. It is possible to get these annotations by searching through databases that use homology-based techniques to match genomic sequences to metabolic reactions. In general, many studies address the usage of these databases in GEM reconstruction processes [140, 141]. Several studies have shown that the choice of a particular database will introduce a bias on the topology of the reconstructed network [142]. This bias is due to the limited accuracy of homology-based methods [143], the possibility of false annotations in large databases [144], and the existence of many enzyme functions that do not match a specific genome sequence [145]. The variabilities due to the annotation database chosen may be reduced by merging many databases to improve annotation coverage when reconstructing a GEM [146]. A functional annotation is done on the basis of the genome sequence using information from databases such as GenBank [147]. Afterwards, information from databases such as KEGG [148], BRENDA [149], MetaCyc [150], and the literature is used to compile a list of reactions and gene-protein-reaction (GPR) relationships.

GPR connections are constructed using logical principles that define the association between genes, proteins, and reactions (Figure 2.8) [151]. This enables the incorporation of information on the transcriptional and translational levels into metabolic models via the use of references to the enzymes that catalyze the processes and the genes that encode those enzymes. A GPR association contains simply the gene/reaction matchings based on AND or OR relations to describe the gene-dependence of each reaction, as detailed below.

Most metabolic reactions, such as Reaction A in Figure 2.8 [151], are based on one gene and one enzyme association. Some reactions (Reaction B) can be catalyzed by different proteins/enzymes independently. These enzymes are called isozymes, and

isozymes are separated with the “OR” logical statement. That is, the presence of any of these enzymes (B1 OR B2) is sufficient to activate corresponding reaction. Some reactions (Reaction C) require multiple enzyme/protein complex to proceed. These enzymes are called complex enzymes, and complex enzymes are separated with “AND” logical statement. Figure 2.8D depicts a complicated GPR rule in which the gene D1 as well as either D2a or D2b are required. Fig 2.8E shows the genotype-phenotype associations in A-D as Boolean GPR rules.

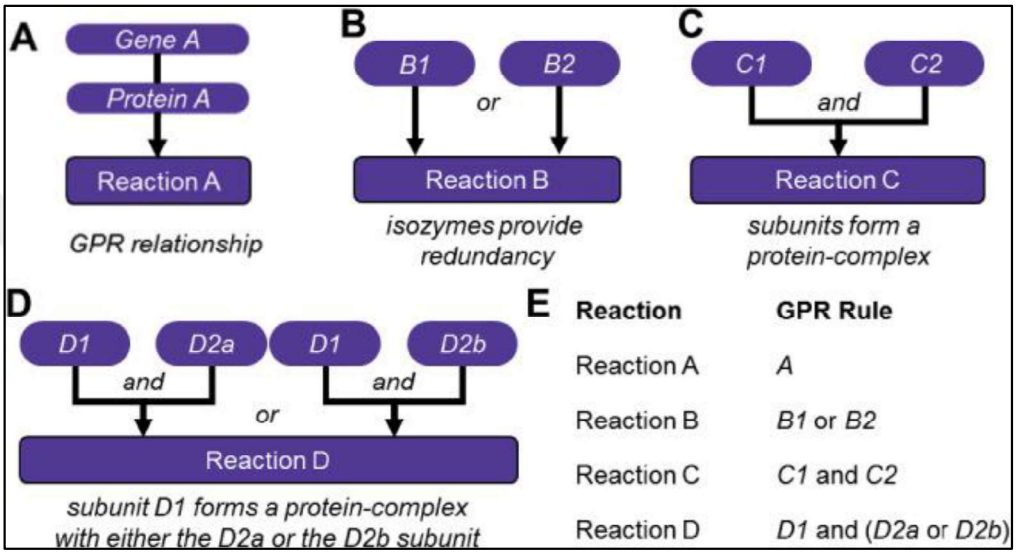

Figure 2.8: Gene-protein-reaction (GPR) rules explain genotype-phenotype relationships.

ii) Manual curation of the draft model: The information obtained from homology/orthology based approach may not be very accurate. Since the addition of non-organism-specific reactions to the model may adversely affect the model's predictive power, a manual literature review is performed to develop and validate the draft metabolic model. Based on the organism-specific literature information, some reactions can be added or removed from the draft model. At this stage, the cellular localization of the proteins is determined, and intracellular and intercellular transport reactions are added based on the cellular compartments. Biomass reaction and ATP-requirement reactions are also added at this stage. In addition, GPR rules are rearranged in line with the information obtained from literature or from template models [138].

iii) Conversion of curated model into a mathematical model: Following performing a manual literature review and improving the metabolic model; the model is converted into a mathematical form that can be analyzed with constraint-based approaches. In addition, system constraints are defined in this step, and condition-specific models can be derived from the generic model created.

iv) Evaluation and analysis: At this step, the power of the metabolic model to predict critical biological functions such as ATP synthesis, lactate release, and mitochondrial activity can be evaluated using constraint-based approaches. Validation of the draft model, conversion of the model into mathematical form, evaluation of the metabolic network, and gap-filling processes are repeated for finalizing and validating the metabolic network. These steps are repeated until the simulation results are consistent with the experimental phenotype data [152].

2.6. Constraint Based Modelling

Cell metabolism is very complicated, and it can only be understood entirely at a system-wide level via simplified mathematical representations known as models. Mathematical modelling has been utilized to investigate metabolism for at least 100 years since Michaelis and Menten developed their enzyme kinetics equation in 1913. Along with the kinetic parameters that were determined for a big enough number of enzymes, it became feasible to construct small models that could represent the fundamental metabolic activities of a live cell [153]. Yet, since the data on kinetic parameters were fragmented, the models may be susceptible to measurement errors. Indeed, enzyme kinetics are often determined from in-vitro experiments and may not always be transferable to in vivo/the living organism [154]. These parameters are also highly susceptible to external and internal variables and vary significantly across cells.

Constraint-based metabolic modeling (CBM) was developed in order to meet the demands of genome-wide models. Contrary to other modelling methods, constraint-based modelling does not need accurate information of quantities of metabolites or kinetics of enzymes [155]. CBM instead limits the solution space available by removing flow patterns that are not permitted by the system due to certain predetermined restrictions.

CBM is based on the concept of writing down differential mass balances around each intracellular metabolite. Differential mass balance equations can be represented in a matrix form:

$$\frac{\mathrm{dC}}{\mathrm{dt}} = S\,x\,v \tag{2.1}$$

$$S_{m\times n} = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m,1} & s_{m,2} & \cdots & s_{m,n} \end{bmatrix} \tag{2.2}$$

C, v, and S represent the concentration of metabolite, a vector of reaction rates (fluxes), and a stoichiometric matrix, respectively. $S_{m\ x\ n}$ is a stoichiometric matrix containing m metabolites and n reactions, with each metabolite representing one reaction. Each member $s_{i,j}$ represents the i^{th} metabolite's stoichiometric coefficient on the j^{th} reaction (Equation (2.2)). Equation (2.1) is solved using steady-state assumption in the CBM. Using the assumption that the amounts of intracellular metabolites remain constant over long periods [156], the metabolic flux distribution can be expressed as follows:

$$S\,x\,v = 0 \tag{2.3}$$

Equation (2.3) is the fundamental equation in CBM, and it is often under-determined. In an under-determined system, there are fewer equations (metabolites) than unknowns (fluxes). CBM attempts to find a solution space for the under-determined problem via using appropriate constraints. Additional constraints such as the reaction directionalities (reversibility or irreversibility) are used to constrain the unknown fluxes (Figure 2.9) [157]. The irreversible reactions have a flux range of [0, ∞] while reversible reactions have a flux range of [-∞,∞]. Mathematically, constraints on fluxes are represented as upper and lower bounds:

$$lb \leq v \leq ub \tag{2.4}$$

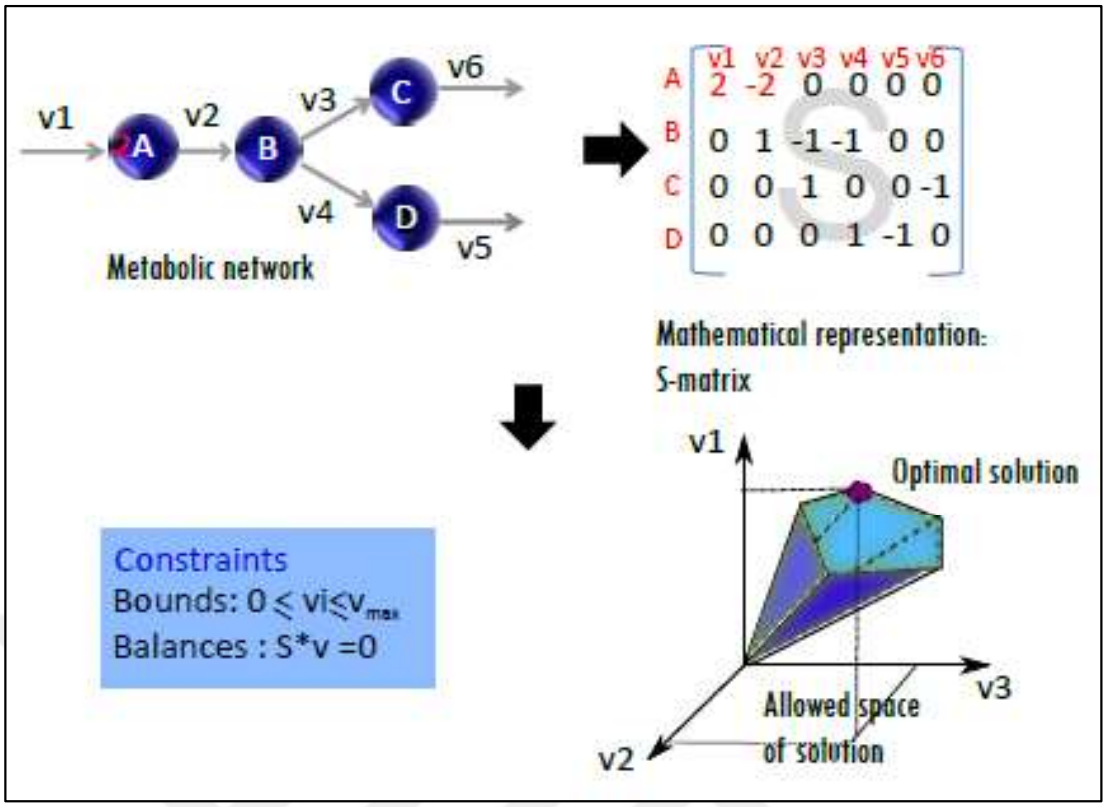

Figure 2.9: Constraint-based modelling defines the solution space by using constraints.

2.6.1. Flux Balance Analysis (FBA)

Flux balance analysis (FBA), a constraint-based modelling approach, is frequently used to analyze the metabolic state of a cell through a metabolic model [118]. FBA estimates metabolic flux distributions using in silico genome-scale metabolic models.

Constraints limit the space of possible flux distributions but do not allow for the discovery of a unique solution since the system is often under-determined. A unique solution for some fluxes can be calculated once the optimization is performed based on an objective function such as maximization of growth, minimization of ATP usage. The biological system's assumed biochemical goal and research purpose have a role in defining the objective function. The main objective function is expressed as:

$$Z = c\ x\ v \tag{2.5}$$

Where c is a row vector showing how much each reaction in v contributes to the objective function (Z). FBA finds the optimum flux distributions based on a set of criteria that satisfy the constraints that have been previously established.

FBA utilizes Equations (2.3) and (2.4) as constraints and an objective function (Equation (2.5)) to solve for unknown fluxes (Figure 2.9). It aims to identify the optimum distribution for a particular context, such as maximizing growth, from all potential flux distributions. Under a steady-state condition, the flux values of the metabolic model are computed by maximizing the objective function with the help of constraints. The majority of objective functions used in FBA are linear and solved using linear programming (LP). However, since linear objective functions have often the alternative optima problem, many solutions may have the same objective function value [158]. That is, linear objective functions results in non-unique flux distributions. The alternate problem may be addressed using a second objective function that minimizes the sum of squares of all fluxes with quadratic programming (QP) by fixing the first objective function at its optimum value. The minimization of the sum of squares method [159, 160] can be expressed mathematically as follows:

$$\min \left(\sum_{i=1}^{n} vi^2\right) \tag{2.6}$$

This approach is founded on the premise that reactions are catalyzed with minimal enzyme use, one of the fundamental characteristics of biological systems [136] .

2.6.2. Minimization of Metabolic Adjustment (MOMA)

Under unperturbed circumstances, the assumption of maximum growth is reasonable. However, genetic disturbances can occur sometimes in the organism. The living organisms are often subjected to genetic disorders yet they tend to retain their metabolism in homeostasis. Diseases can also be considered as a perturbation of healthy state. There are CBM methods to model metabolic fluxes in disturbed states. These methods assume that the disturbed organisms/cells show a minimum divergence in reaction rates from their healthy counterparts.

Minimization of Metabolic Adjustment (MOMA) is a CBM method that is used to estimate the impact of disturbances like gene deletion on metabolism [61]. MOMA assumes that a perturbed/disrupted cell will try to stay as close as possible to its reference state. MOMA utilizes the healthy state flux distribution as reference and attempts to calculate the smallest possible Euclidean distance between healthy state flux distribution (w_i) and perturbed state (disease) flux distribution (v_i). Therefore, it establishes a quadratic objective function that translates into a problem in quadratic programming (QP).

$$min \sqrt{\sum_{k=1}^{n} (w_i - v_i)^2} \tag{2.7}$$

$$(-2w_i v_i - vi^2) \tag{2.8}$$

Equation (2.7) is simplified to Equation (2.8) to be utilized in quadratic programming via eliminating the w_i^2 constant value and square root from Equation (2.7). The linear portion of the optimization is represented by $-2w_i.v_i$ while vi^2 represents the quadratic part. That is, there are two objective function terms required to solve. MOMA utilizes the quadratic programming to solve them.

2.6.3. Least-square with equalities and inequalities Flux Balance Analysis (LseiFBA)

Least-square with equalities and inequalities Flux Balance Analysis (LseiFBA) is a CBM method used to predict the metabolic changes in disease from gene expression data. Gene expression and enzymatic flux changes are assumed to be proportionate in LseiFBA algorithm [74]. Lsei-FBA offers a bioinformatic prediction of network-level changes based on trends indicated by gene expression changes. When calculating flux values in the disease state, LseiFBA assumes that the fold changes of reaction rates and gene expression fold changes are proportional between the control and disease states. In other words, if the mRNA level of the gene encoding the enzyme that catalyzes a reaction between the two conditions has increased, the rate of the

related reaction will also increase. If it has not changed, the rate of the corresponding reaction will not change either.

Lsei-FBA algorithm can be applied in four steps: (i) FBA is performed to obtain the flux prediction for healthy state (V_{rf}), (ii) Fold change (FC_{gene}) is calculated by dividing the disease gene expression value to control gene expression value for each gene (Equation (2.9)), (iii) Initial disease flux vector ($V_{ref,D}$) for disease state is obtained via multiplying the healthy state flux distribution vector with the fold change values (Equation (2.10)). $V_{ref,D}$ will not obey mass balance constraints. Therefore, the flux estimation of disease case is performed by utilizing an objective function that minimise the Euclidean distance between the initial disease flux vector and disease state flux distribution to be estimated (Equation (2.11)).

$$FC^i_{gene} = \frac{D^i_{gene2rxn}}{C^i_{gene2rxn}} \tag{2.9}$$

$$V^i_{ref,D} = FC^i_{gene} \times V^i_{rf} \tag{2.10}$$

$$min \sum \left(v^i_D - v^i_{ref,D}\right)^2 \tag{2.11}$$

2.6.4. Modified Least-square with equalities and inequalities Flux Balance Analysis (mLseiFBA)

Modified LseiFBA is the linear version of the LseiFBA [74]. Modified LseiFBA, here referred to as mLseiFBA is utilized to estimate disease state flux distributions. As in the LseiFBA method, mLseiFBA predicts the disease state flux from the reference flux distribution. It first calculates the fold change of each gene (FC_{gene}) (Equation (2.9)) and performs FBA analysis to predict the flux distribution for healthy state (V_{rf}). Fold change values of genes (FC^i_{gene}) are multiplied with healthy state flux distribution (V^i_{rf}) to obtain the initial disease flux vector ($V_{ref,D}$) (Equation (2.10)).

Unlike LseiFBA (Euclidean distance), mLseiFBA utilizes an objective function that minimize absolute flux difference between the reference flux vector and disease condition flux distribution (Equation (2.12)), subject to stoichiometric, mass balance and imposed boundary constraints. The model should be converted to an irreversible

form to apply the mLseiFBA approach since the objective function of mLseiFBA is to minimize the sum of absolute flux difference between the reference flux vector and disease condition flux distribution.

2.7. Context Specific Model Reconstruction

In multicellular organisms, cells can exhibit a wide variety of phenotypes despite sharing the same genome. GEMs have been widely used to guide the study of large omics datasets, as they offer a biological context for these data by demonstrating a mechanistic link between genotype and phenotype.

GEMs include all known metabolic processes but, increasing evidence suggests that only a fraction of these reactions are active in a specific context, such as a particular developmental stage, cell type, or environment. This is especially important when dealing with multicellular organisms such as mammalians where many cell types with specific metabolic activities coexist [162]. Numerous studies, including in silico knock-out experiments, have demonstrated that extracting subnetworks containing only active reactions improves metabolic model prediction [163].

To simulate cellular metabolism, it is essential to reconstruct a GEM tailored to the particular cell type or condition. Context-specific models are a subset of the original generic organism-specific GEMs [124]. The context-specific models are reconstructed from the generic model via extracting a subset of reactions that are active in the particular cell or condition. Several cell-type-specific models, including liver [164], brain [135], alveolar macrophage [165], and a multi-tissue model for hepatocytes, adipocytes, and myocytes have been manually built utilizing the generic models as scaffolds [166, 167]. Because not all enzymes are active in every cell type or culture environment, algorithms have been developed to construct context-specific models utilizing omics data to recapitulate the metabolism of particular cell types under specified circumstances [168].

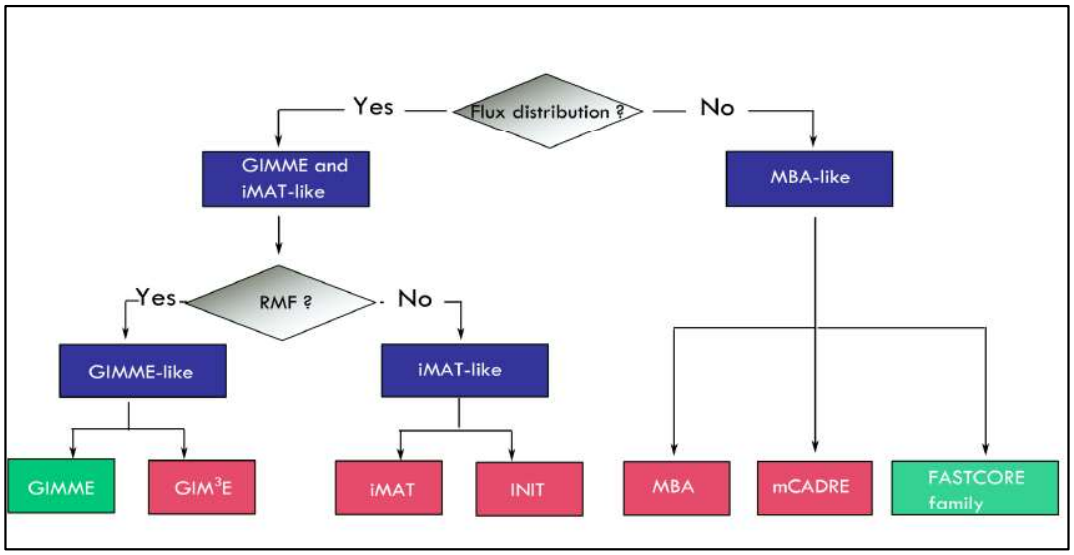

Figure 2.10: Families of context-specific reconstruction algorithms.

Regarding the objective used, context-specific reconstruction algorithms may be classified into three major families: GIMME-, iMAT-, and MBA-like families (Figure 2.10) [169].

GIMME-like family algorithms attempt to maximize the similarity between the flux phenotype and omic data/transcriptome by optimizing an objective function. MBA-like family algorithms categorize the reactions as core and non-core reactions. They tend to maximize the core-reactions and minimize the non-core reactions in the final model.

The consistency between the flux distribution and the data is optimized/maximized by the iMAT-like family of algorithms. The iMAT-like family, including iMAT [170], INIT [171] and tINIT [114] do not need to optimize for a biological function. The primary benefit of this family of techniques is not requiring an objective function. Therefore, such approaches are useful to draw up context-specific models if no particular objective function is known to dominate the situation, which is frequently the case for mammalian organisms and corresponding tissue-specific models. The disadvantage of these methods, on the other hand, is that they need a significant amount of computing power since they use MILP (Mixed Integer Linear Programming).

The major distinction between iMAT and INIT is that INIT utilizes expression-based weights or arbitrary values to favor the inclusion of reactions that are most supported by the data rather than applying rigid discretion/threshold criteria.

Additionally, INIT can impose a net positive flux via the processes that generate certain metabolites by setting non-zero limits ($b>0$ in $Sv = b$). tINIT accepts an extra parameter, which specifies the metabolic tasks that the model must be capable of doing.

2.7.1. Integrative Metabolic Analysis Tool (iMAT)

Tissue-specific metabolic models have long been developed, and the use of omics data to enhance phenotypic predictions is not a novel concept. Shlomi et al. [172] pioneered the use of the Integrative Metabolic Analysis Tool (iMAT) to predict metabolic activity in 10 human tissues: the brain, heart, kidney, liver, lung, pancreas, prostate, spleen, skeletal muscle, and thymus.

This method incorporates gene or protein expression data into the already available human metabolic network model. Since post-transcriptional regulation is not represented in gene and protein expression data, this approach uses the levels of gene expression as signals to determine the likelihood that the reactions connected with them carry metabolic flux. Genes are categorized as highly, lowly, and moderately expressed genes and then transformed to 1, -1, and 0, respectively, by utilizing the gene-protein-reaction rules. This transformation results in two sub-sets of reactions: RH (highly expressed) and RL in the model reactions (lowly expressed). iMAT uses a Mixed Integer Linear Programming (MILP) formulation to determine a steady-state metabolic flux distribution. The number of reactions associated with highly expressed genes is maximized and the number of reactions associated with lowly expressed genes is minimised. Therefore, no biological objective function is required.

The full iMAT formulation is shown below [173]:

$$
\begin{aligned}
&max\ \left(\sum_{i\in R_H}(y_i^+ + y_i^-) + \sum_{i\in R_L} y_i^+\right)\\
&s.t\ S.v = 0\\
&\quad v_{min} \le v_i \le v_{max}\\
&\quad v_i + y_i^+(v_{min\,i} - \epsilon) \ge v_{min\,i} \qquad i \in R_H\\
&\quad v_i + y_i^-(v_{max\,i} + \epsilon) \le v_{max\,i} \qquad i \in R_H\\
&\quad v_{min\,i}(1 - y_i^+) \le v_i \le v_{max\,i}(1 - y_i^+) \qquad i \in R_L\\
&\quad y_i^+, y_i^- \in [0,1]
\end{aligned}
\tag{2.13}
$$

v, S, v_{min} and v_{max} represent the flux vector, a stoichiometric matrix, lower and upper bounds of the fluxes respectively. The boolean variables y^+ and y^- indicate whether the reaction is active (in either direction), and the ϵ value represents the smallest amount of flux that must be present in order for the reaction to be classified as active.

2.7.2. Integrative Network Inference for Tissues (INIT)

In 2012, Jens Nielsen's team introduced the Integrative Tissue Network Inference (INIT) algorithm [171]. This method utilizes Human Protein Atlas (HPA) cell type expression information as a primary source of data to evaluate the existence or lack (protein evidence score) of metabolic enzyme for each human cell type.

The algorithm relies heavily on cell type-specific information from HPA to determine the presence or lack of metabolic enzymes in each of the human cell types. Additional data sources can also be utilized such as tissue-specific gene expression and Human Metabolome Database (HMDB) metabolomic data. The INIT algorithm has been developed as a MILP problem to maximize the number of scores for flux-carrying reactions. Unlike other algorithms, the INIT algorithm allows some metabolites to accumulate in small amounts (positive) instead of making a steady-state assumption for all internal metabolites. This prevents the removal of reactions involving dead-end metabolites.

INIT formulation using MILP can be specified as [171]:

$$
\begin{aligned}
& max \left(\sum_{i \in R} w_i y_i + \sum_{j \in M} x_j \right) \\
& S.v = b \\
& |v_i| \leq 1000 y_i \\
& |v_i| + 1000(1 - y_i) \geq \varepsilon \\
& v_i \geq 0 \qquad i \in irreversible \\
& b_j \leq 1000 x_i \\
& b_j + 1000(1 - x_i) \geq \varepsilon \\
& b_j \geq 0 \\
& x_j = 1 \qquad j \in present \\
& y_i, x_j \in 0, 1
\end{aligned}
\tag{2.14}
$$

S, v, b, R, M indicate stoichiometric matrix, reaction rate vector, net accumulation/consumption rate for each internal metabolite, reactions and metabolites, respectively. ε parameter is an arbitrary small positive number. y_i and x_j are binary variables and they denote the active or inactive state of a reaction and a metabolite, respectively. If the protein evidence scores are obtained from HPA, the weight scores are assigned to 20, 15, 10 or -8 values, which refer to the high, medium, low and absent evidence levels. Once transcriptome data is used, the weight for each gene is calculated using the following formula [171]:

$$w_{i,j} = 5log(\frac{Signal_{i,j}}{Average_i}) \tag{2.15}$$

$w_{i,j}$ represents the tissue of interest and i the gene of interest. Based on the formula, when the transcriptome data is used, the weight calculation for each gene is as follows: The expression value of the related gene in the relevant tissue is divided by the average of the expression value of that gene in all tissues, and when the natural logarithm X of the obtained value is taken and multiplied by 5, the weight for the relevant gene will be calculated. A positive weight value will be obtained if gene expression in a particular tissue is greater than the average of all tissues. INIT algorithm attempts to maximize the number of positive weight reactions and minimize the negative weight reactions in the final model.

3. CLUSTERING ANALYSIS

Because various rat transcriptome datasets linked with Parkinson's disease were utilized in this research, it was investigated to what extent the datasets used represented Parkinson's disease and how similar the datasets were in terms of metabolic genes by considering brain area and model type (neurotoxin model, genetic model) in this chapter. Hierarchical cluster analysis was performed to investigate the similarity between these datasets.

3.1. Datasets

The transcriptome datasets used in this study were selected from Gene Expression Omnibus (GEO) [105] based on the studies designed to compare PD with control using rat as a disease model, with at least three samples for each condition in the dataset. In this study, 6 different rat transcriptome datasets including 11 comparisons (control groups vs. Parkinson's disease groups) associated with Parkinson's disease were used. Summary of the datasets is presented in Table 3.1

Table 3.1: Rat transcriptome datasets.

Organism	GEO ID	Brain Regions	Model Type	# Samples	Platform	Comparisons
Rat	GSE24233	Striatal Brain Tissue	6-OHDA (neurotoxin model)	9	Illumina ratRef-12 v1.0	Control saline (n =4) vs 6-OHDA saline lesioned (n =5)
Rat	GSE58710	Substantia nigra	6-OHDA (neurotoxin model)	30	RaGene-1_0-st	Vehicle (n =3) vs 6OHDA (n =3) Five different time points
Rat	GSE71968	Hemi Brain (Half cerebrum + Half cerebellum + Half brainstem)	DJ-1 knockout (Genetic model)	13	Illumina HiSeq 2000	Wilde type (n =7) vs DJ-1 knockout (n =6)
Rat	GSE74382	Dorsal striatum	L-dopa (neurotoxin model)	14	Rat230_2	Sham saline saline (n =7) vs Lesion saline saline (n =7)
Rat	GSE93695	Striatum	6-OHDA (neurotoxin model)	6	RaGene-2_0-st	Normal group (n =3) vs PD-group (n = 3)
Rat	GSE150646	Frontal Cortex	Over exp-SNCA (Genetic model)	20	Illumina HiSeq 2500	Wild type (n =5) vs Transgenic rats (n = 5) Two different time points

3.2. Dataset Normalization

Since the transcriptome datasets used in this study correspond to different platforms (such as Microarray, RNA-sequencing (RNA-seq)), each dataset was read on R environment through different Bioconductor packages including affy [174] and oligo [175]. All steps involving normalization of datasets were performed in R programming environment.

3.2.1 Dataset Preprocessing

Micro-array datasets and RNA-seq datasets were processed separetely. The preprocessing of microarray datasets was performed mainly in three steps: i) Background correction (summarization of probe level data into probe-set level using RMA convolution model), ii) summarization of probe-sets to genes, and iii) quantile normalization and log2 transformation.

i) RMA correction: Affymetrix GeneChip arrays are widely utilized by researchers. Genes are probed by using oligonucleotides of 25 bp length. A probe set consisting of 11–20 probe pairs of these oligonucleotides is typically used to represent an mRNA (messenger RNA) molecule of interest in the microarray design. Each probe pair is composed of a perfect match (PM) probe and mismatch (MM) probe. Perfect match (PM) probes are used to detect specific binding to particular sections of mRNA molecules. Mismatch (MM) probes are produced by altering the middle (13th) base of the PM probes to measure non-specific binding [176]. Following scanning of the arrays hybridized to labelled RNA samples, the intensity values PM_{ij} and MM_{ij} for the arrays are calculated and recorded. The total of probe intensities for each probe set must be computed in order to get a measure of expression that represents the quantity of the associated mRNA molecules. The RMA (Robust Multi-Array Average) processing is a background normalization and probe-to-probe-set summarization approach. It employs a global background correction based on PM probes. For the RMA-based correction of "RaGene-1_0-st", "RaGene-2_0-st" platforms, "oligo" [175] and for "Rat230_2", "affy" [174] R libraries were used.

ii) Summarization of probe-sets to genes: For cross-platform analysis to be possible, expression levels need to be characterized as universal entities such as genes. As a result, probe sets should be converted to their associated gene IDs. Mostly, the relationship between probe-sets and genes is not one-to-one. That is, it is possible to have a probe-set linked with two genes, as well as two probe-sets associated with one/same gene. A probe-set corresponding to more than one gene was excluded from the dataset to avoid non-specific expression levels. In the case of multiple probe matches for a single gene, the probe with the highest mean expression value across all samples in the datasets was selected.

In different microarray-based datasets, gene information can be given in various formats (gene symbol, Entrez ID, Ensemble ID, etc.). To enable comparison of datasets, expression values in all datasets were expressed in terms of Entrez ID. Probe set-Entrez ID conversion in microarray datasets was performed/achieved using the biomaRt package [177]. Moreover, Ensembl/Ref-seq IDs were converted to Entrez ID in RNA-seq datasets through biomaRt package.

iii) Quantile Normalization and Log2 transformation: Quantile normalization is a comprehensive data analysis technique that eliminates non-biological variance between samples. The method makes sample distribution the same, but maintains differential expression across samples [178] (Figure 3.1). The *affy()* function and its default parameters were used to perform quantile normalization. Log2 transformation is often used in gene expression data analysis to make variations in gene expression levels across different genes more similar and more symmetrical [179]. It is aimed to make the data close to a normal distribution by performing the log2 transformation before applying statistical tests based on normal distribution such as t-test. Accordingly, the transformation was applied, and the distribution of mRNA expression values reported by each dataset was examined.

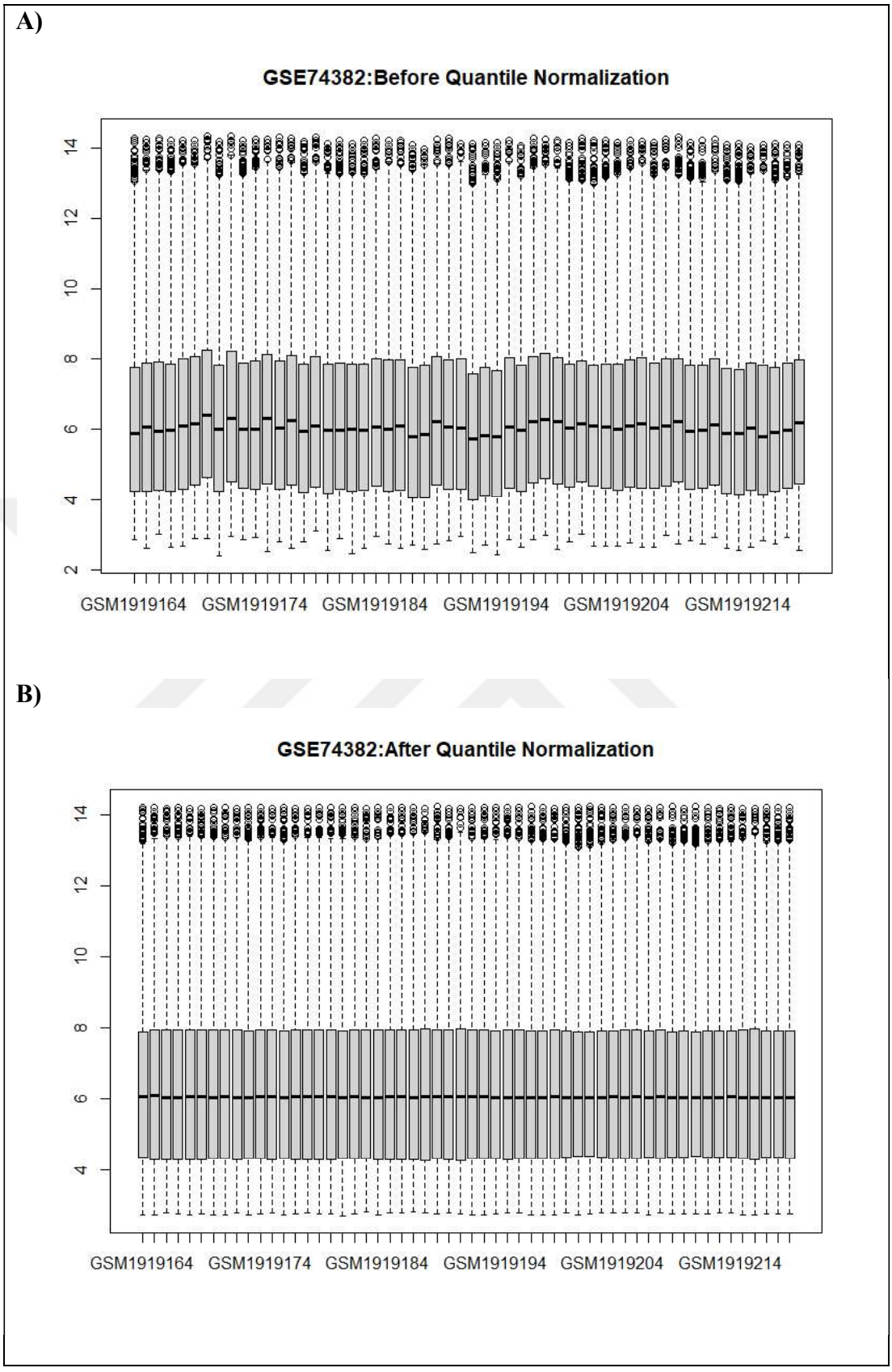

Figure 3.1: Example boxplot for the GSE74382 dataset A) before and B) after quantile normalization.

The processing of RNA-seq datasets was performed mainly in three steps: i) Converting counts to DGEList Object, ii) Filtering lowly expressed genes, and iii) Normalization.

i) Converting counts to DGEList Object: Following reading the RNA-seq count data, A DGEList class object was built using the count matrix through edgeR bioconductor package [180]. The DGEList object is used to hold all the necessary information, including the library sizes and dispersion estimation of each gene for a general linear model [181].

ii) Filtering lowly expressed genes: There is minimal indication of differential expression in genes with extremely low number counts across the libraries. Additionally, the high degree of discreteness in these counts complicates the application of certain statistical approximations later in the pipeline. In this regard, these genes should be removed before further investigation. Firstly, the count data was converted to count-per-million via *cpm()* function. Then, the highest cpm value was determined for each gene across all samples, and genes with less than one maximum-cpm-value were excluded.

iii) Normalization: RNA-seq analysis techniques generally normalize data across samples by reducing/scaling the number of reads in a particular lane or library to a shared value across the sequenced libraries [182]. Normalization of library sizes was performed via the use of the calcNormFactors function, which finds a set of scaling factors for library sizes that minimizes the log-fold differences across samples for the majority of genes. Trimmed mean of M-values (TMM), which is the default method in edgeR, was used as the normalisation method to remove compositional biases amongst libraries. This results in a number of normalizing factors that determine the effective library size of the product and the sizes of the library.

3.2.2. Principal Component Analysis (PCA)

Principal component analysis (PCA) uses a mathematical algorithm that reduces the size of the data substantially by retaining the variation in the dataset [183]. In other words, PCA is a technique used to reduce the size of datasets, increase interpretability and at the same time minimize information loss. It is also used to detect whether the samples in a dataset consisting of two groups form two separate groups as expected, and most importantly, to detect outlier samples in the dataset. PCA analysis is a statistical approach frequently used in transcriptome studies to visualize variation between samples.

Since the focus of this thesis is metabolic alterations, transcriptome data collected from in vivo rat models of Parkinson's disease were filtered using metabolic genes before PCA analysis. Metabolism associated genes were obtained from Human-GEM genome scale metabolic model containing 3628 genes [122]. Rat orthologs of 3628 Ensemble gene IDs obtained from Human-GEM were identified through biomaRt [177, 184], and 3847 rat ortholog Entrez gene IDs were obtained. Transcriptome datasets were filtered based on rat orthologous Human-GEM metabolic model genes. Following the filtering, *prcomp ()* was used in R to conduct PCA analysis to identify the outliers in each dataset, and the outliers in the datasets were excluded from the datasets (Figure 3.2). Removed samples from datasets are given in Table 3.2.

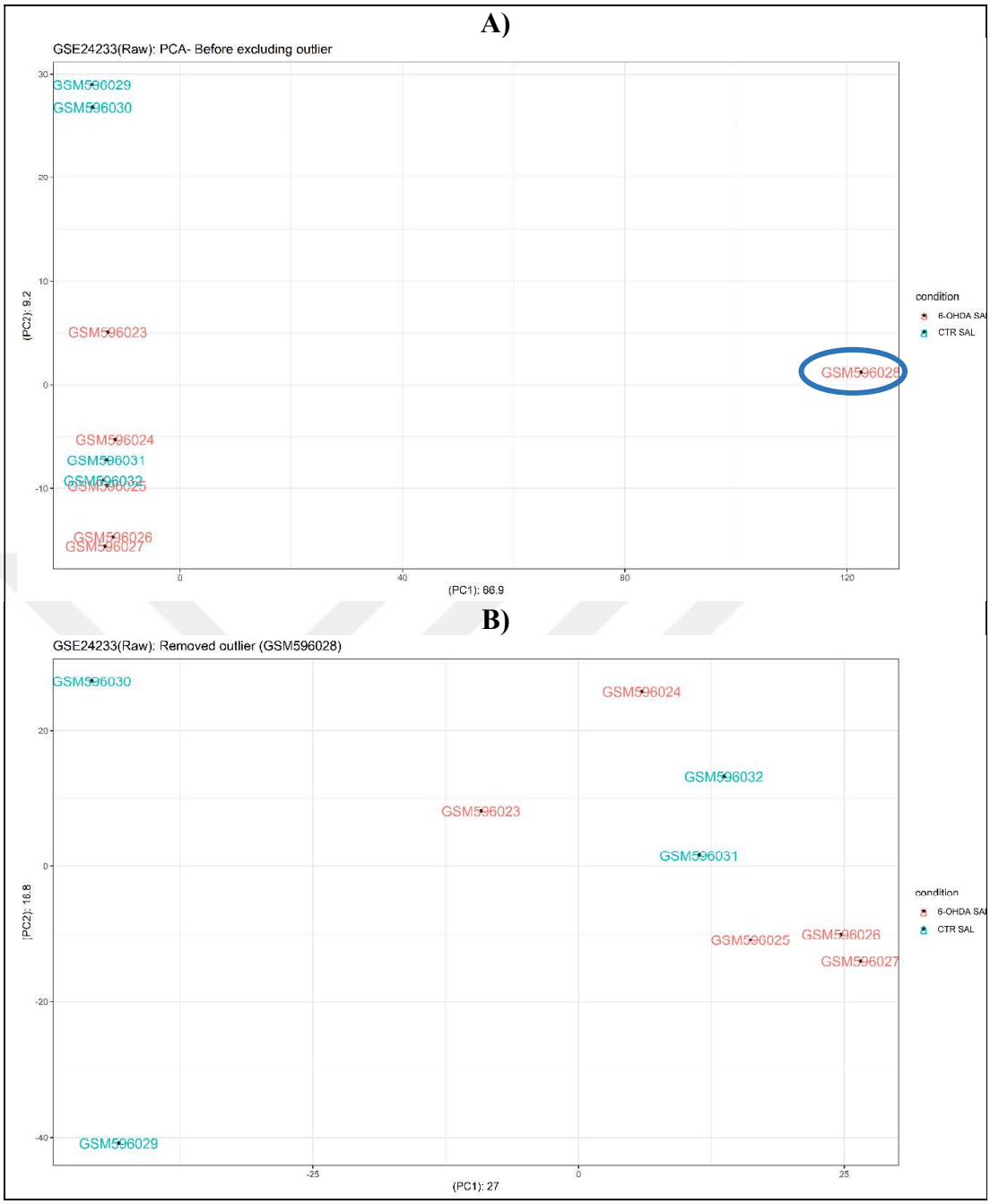

Figure 3.2: Example PCA for the GSE24233 dataset. A) Before removing outlier. B) After removing the outlier. Blue circle indicates the outlier sample. The x-axis corresponds to the first principle component and the y-axis corresponds to the second principle component. Percentages in parentheses refer the variance explained by each component.

Table 3.2: Removed outlier samples.

GEO ID	Removed Sample ID
GSE24233	GSM596028
GSE71968	GSM1848719

3.3.Hierarchical Clustering Analysis

Clustering is an unsupervised learning problem that entails dividing observations into homogeneous subsets or clusters to deduce subpopulation organization within a dataset. Numerous clustering techniques have been suggested, including hierarchical clustering [185].

Hierarchical clustering is a method that groups related items into groups named clusters. The endpoint is a collection of clusters, each unique from the others, and the items inside each cluster are more similar to each other than to other items. Agglomerative and divisive clustering algorithms are two types of methods used in hierarchical cluster analysis. A “bottom up” method is used with agglomerative clustering that directs from “the leaves” to “the root” of a cluster tree. Divisive clustering is a "top down" method that begins at the root and progresses to the leaves [186]. Agglomerative hierarchical clustering analysis is a kind of unsupervised machine learning method that is often used to identify comparable subgroups within a larger population [187].

Since rat datasets will be subjected to clustering analysis together, instead of directly using these datasets produced by different study groups, fold change and p-value for each gene in the relevant condition were used in the clustering analysis to make each dataset independent of the mRNA levels it reports. Differential expression analysis of normalized datasets was performed through the limma package [188] to calculate gene fold change and empirical Bayes p-values between disease and control groups for each of the datasets. Fold change values vary between [0-1] for down-regulated genes while for genes with increased expression, the range is [1-∞]. To compensate for the inequality in the range, the fold change values of down-regulated genes were transformed to 1/(fold-change). And, the values were multiplied with (-1) to preserve the information that they were decreased. Since bivalent information such as fold change and p-value may be critical for capturing smoother clustering patterns, scoring system called gene scoring was used to combine fold change and p-value information of the differential expression analysis results. Clustering analysis was carried out using the "gene score" method, which has been proposed in the literature [189]. The following formula was used to calculate gene score:

$$Gene\ Score\ =\ (FC)\ x\ (-log(10(p-val)) \quad (3.1)$$

To calculate gene score, the transformed fold-change values (FC) as mentioned above were obtained for 11 Control and PH comparisons from 6 datasets, and they were collected in a single matrix. The same procedure was performed for the p-values. After combining the datasets, a matrix with 3653 rows (genes) and 11 columns (comparison groups) was obtained. Among the genes with missing values in the data, there may be important metabolism genes affected in Parkinson's disease. If the number of comparisons missing a value was greater than 5% for a gene, such genes were excluded from the data. As a result of this filtering, a new data consisting of 1727 genes and 11 columns (comparisons) was obtained. Then, missing values in the data were filled in using the missForest package/library [190] in order to retain gene information in the data. missForest is a non-parametric missing value filling technique that uses the randomforest algorithm to fill in missing values that arise during the merging of various datasets. After filling in the missing values, the gene scores were calculated by using Equation (3.1). The gene score distribution of 11 comparisons is given in Figure 3.3. According to Figure 3.3, some extreme gene-score values were identified for some comparisons. Three genes with extreme gene score values were excluded from the data matrix.

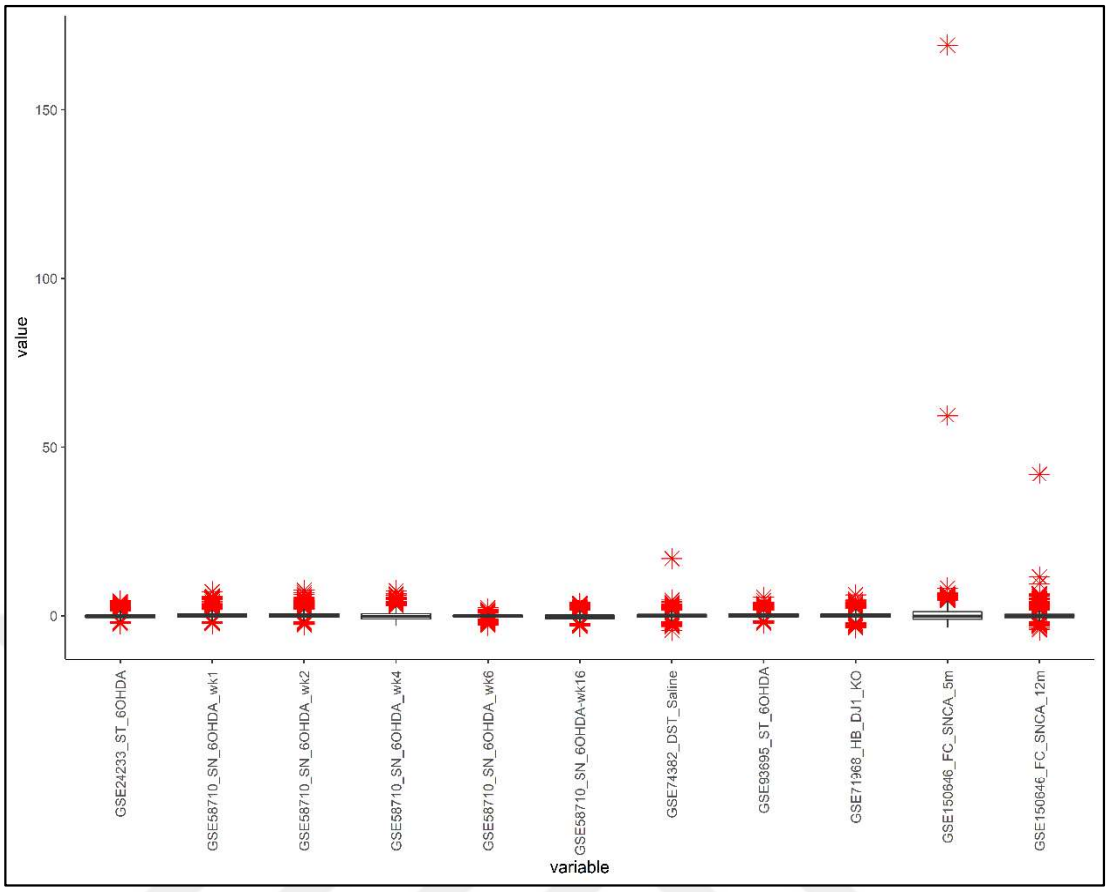

Figure 3.3: Gene-score distribution of the 11 PD-control comparisons.

After filtering out the genes with extreme gene score values, instead of utilizing all 1723 genes that remained, clustering was performed by choosing the genes that exhibited the highest variability across all circumstances. It is important to utilize genes with the highest variance in the gene score values for all data comparisons to separate the comparisons more correctly in terms of brain areas and model types (neurotoxin models, genetic models) in the hierarchical clustering analysis. Highest variable genes across the comparisons were determined using plotloading() in *PCAtools* package [191]. Here, for each of the principal components that explains 70% of variance in the gene score matrix, the genes that contribute most to the variance in that principal component was determined and selected. For this, 10% of genes with highest principle component scores and 10% of genes with lowest scores were selected. 1318 genes were obtained, and these genes were subjected to hierarchical clustering analysis. Hierarchical clustering analysis was performed using the factoextra package/library [192] Different proximity measures (Euclidean distance, Spearman correlation, Pearson correlation) were used to detect the difference between

samples through agglomerative clustering and ward.D method for hierarchical clustering analysis. Compared to other proximity measure methods, groups were better separated/clustered once the Spearman method was used (Figure 3.4).

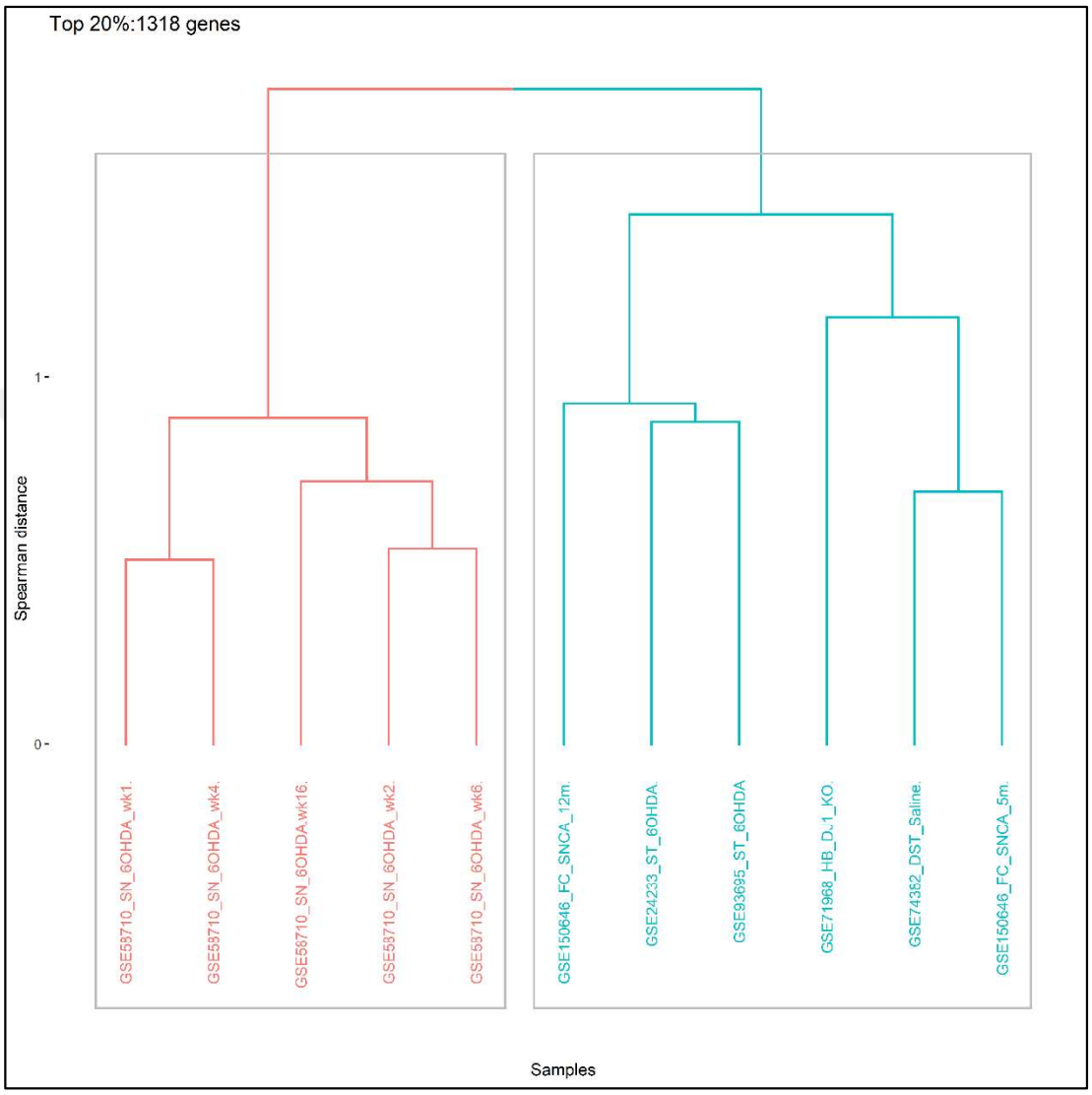

Figure 3.4: Gene-Score based hierarchical clustering analysis of datasets.

Hierarchical clustering analysis of the datasets/samples via metabolic genes showed that the samples were clustered under two main groups (Figure 3.4). Substantia nigra (SN) samples clustered in one group while the other group contained heterogeneous brain regions including striatum, dorsal striatum, hemibrain, and frontal cortex brain regions (Figure 3.4). Normally, the same/similar brain regions are expected to be under the same group. Figure 3.4 shows that the SN samples were collected under the same group and all these samples corresponded to the same dataset

(GSE58710). It was not expected that SN samples would cluster under the same group since GSE58710 is a time point dataset. This result may originate from the laboratory effect. Moreover, it was seen that striatum samples corresponding to different datasets clustered together as it was expected.

1318 genes were also subjected to enrichment analysis to investigate whether these genes are associated with the Parkinson's disease. The enrichment analysis was conducted via enrichR package [193]. Figure 3.5 indicates that there is Parkinson's disease term between enriched terms and also there are significantly enriched terms especially associated with PD including TCA-cycle metabolism [4], inositol phosphate metabolism [194], oxidative phosphorylation [4]. The results of enrichment analysis have shown that the genes subjected to hierarchical clustering analysis are metabolic genes and can be utilized to investigate the differences/similarities in the metabolism of Parkinson's disease across comparison groups.

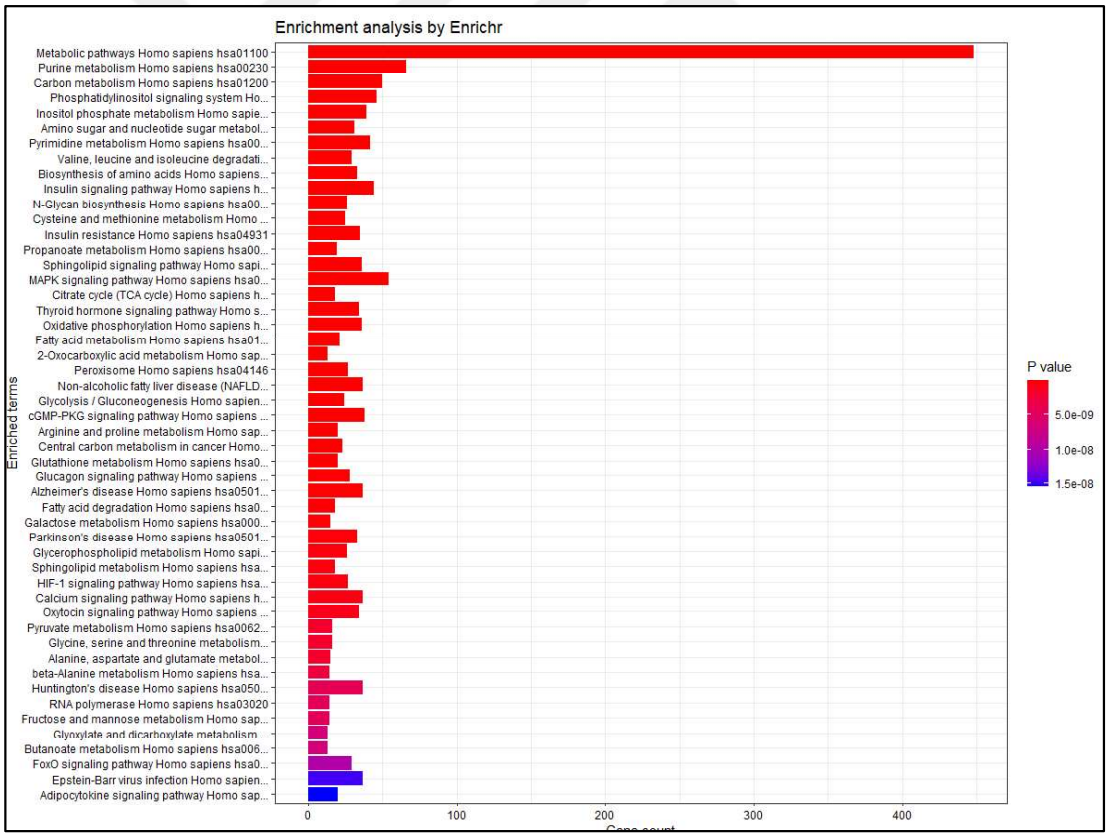

Figure 3.5: Enrichment analysis of most variable gene.

4. RECONSTRUCTION OF A RAT BRAIN-SPECIFIC GENOME SCALE METABOLIC MODEL

Rats are used as model organisms in the development of preclinical drugs and the identification of biomarkers. Although the rats and humans are genomically and physiologically similar [195], there are functional variations in nonpharmacokinetic metabolism, which may affect whether a chemical causes toxicity or increases the level of a biomarker [196]. Identifying and interpreting species-specific distinctions between rats and humans is critical in terms of the interpretation of preclinical animal research in drug development, biomarker identification, and comparative toxicogenomics [197].

Metabolic modeling of rat brain at genome scale is required to overcome the information gap in clinical and fundamental research between humans and rats. The low availability of well curated GEMs for rats and other animals has been ascribed to the significant work needed to manually build a GEM based on information from biochemical databases, genomic annotations, and literary evidence [198].

Because metabolism plays a critical function in the cell, it is essential to conduct a comprehensive study of the impact of neurodegenerative disorders on brain metabolism. Such investigations may potentially identify the common pathways disrupted by neurodegenerative illnesses and metabolism-specific disorders.

Astrocytes are critical in terms of the proper functioning of the brain's metabolism, which may lead to neurological conditions in the case of astrocyte disruption. Astrocytes and neurons have function on neurotransmitter metabolism [199, 200]. Experimental evidences have shown that the connections between neurons and adjacent astrocytes need more examination, particularly under pathological circumstances, to comprehend the neurovascular and neurometabolic linkage better [201]. Therefore, models are required that cover neuron and astrocyte metabolisms separately and couple them together.

Brain-specific genome scale metabolic models have been reconstructed based on stochiometric modelling [10, 11, 136, 165, 202]. iBrain671-Hs [10] and iBrain674-Mm[11], taking into consideration neuron-astrocye separation, are brain-specific genome scale metabolic models that have been recently reconstructed for human and

mouse organisms, respectively. All those models are two-cell models, including metabolic reactions in and between neurons and astrocytes.

There are brain-specific GEMs for both human and mouse but such a model is not available for rat. In this thesis study, a rat brain-specific metabolic model was constructed with reference to the human iBrain671-Hs and mouse iBrain674-Mm brain-specific metabolic models. The rat brain specific GEM reconsruction was performed mainly in four steps. The method of reconstructing genome-scale brain-specific metabolic networks for rats is detailed in the following sections and illustrated in Figure 4.1.

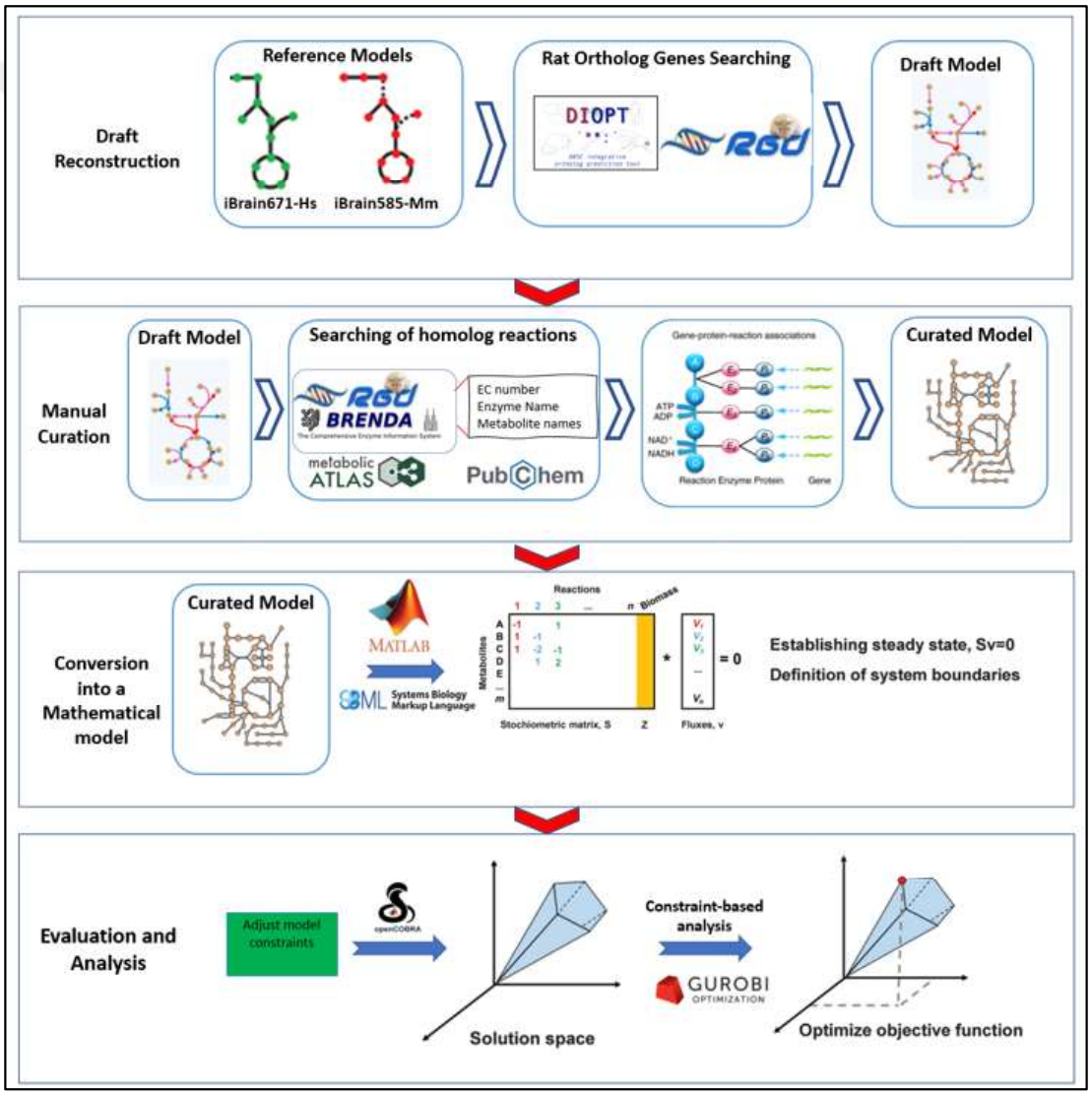

Figure 4.1: GEM reconstruction steps followed in this thesis study (Modified from [203]).

4.1. Draft Reconstruction

Homology-based approach was utilized in the reconstruction process of brain-specific genome scale metabolic model of rats [139]. Since rat brain metabolism is similar to human and mouse brain metabolisms, human iBrain671-Hs and mouse iBrain674-Mm brain-specific models were used as template models to reconstruct rat brain specific GEM.

Firstly, rat orthologs of genes responsible for each reaction in human (iBrain671-Hs) and mouse (iBrain674-Mm) models were separately identified. Since utilizing several databases for orthologous conversion may result in the introduction of erroneous genes with different functions, a single database was used for orthologous transformation.

DIOPT was used as the main source to identify orthologous gene pairs between humans and rats, as well as mouse and rats [204]. DIOPT is a simple but efficient tool for the fast ortholog identification between organisms. It integrates different approaches allowing rapid identification of orthologs among human, mouse, rat, zebrafish, *C. elegans*, Drosophila and *S. cerevisiae*. It also shows protein and domain alignments for anticipated orthology pairings, including percentage of amino acid identity. This helps users find the most suitable matches among numerous orthologs. DIOPT also offers a DIOPT score and rank referring to the orthology accuracy. The DIOPT score refers to the number of used databases for ortholog search. Since the DIOPT uses a total of 18 different databases for ortholog identification, the maximum DIOPT score is 18. DIOPT classifies the ortholog results as high, moderate or low rank. High rank indicates the best score in both ways (forward and reverse ortholog direction) with the DIOPT score ≥ 2 while moderate rank indicates the best score (forward or reverse) with the DIOPT score ≥ 2 . On the other hand, low rank refers the all others. Because several genes in the human and mouse models had multiple rat orthologs in the DIOPT database, the rat orthologous genes with the highest DIOPT score and rank labelled as high were utilized to avoid introduction of erroneous genes with different functions. On the other hand, rat orthologs of some genes were not found in DIOPT, or rat orthologs with low DIOPT scores were detected.

The Rat Genome Database (RGD) [205] was used as an additional source when rat orthologs of some genes in the human or mouse model were not detected by DIOPT

or when the ortholog pairs had a low DIOPT score. RGD is the most comprehensive rat-specific database in terms of rat genetic-genomic-phenotypic data and molecular pathway literature information.

If the rat orthologs of genes in the relevant reaction were identified in the human (iBrain671-Hs) or mouse (iBrain674-Mm) model, the reaction was added to the rat model, assuming that the reaction occurred in the rat brain. For this purpose, GPR rules in human and mouse models were converted to rat GPR rules individually by substituting human and mouse genes with known rat orthologs. Finally, GPR rules were merged to create a rat metabolic network from both the human and mouse metabolic networks, and a draft brain specific GEM was reconstructed for rat.

4.2. Manual Curation

Following a draft reconstruction of rat brain specific GEM, a complete re-evaluation and refinement of the draft reconstruction were carried out at this stage through literature information and several databases. Since the inclusion of non-organism-specific reactions in the model may have a negative impact on the model's predictive capability, the rat draft network was curated and refined to ensure the existence of all enzymes and their related genes in the rat brain.

EC number is an enzyme categorization system based on the chemical processes catalyzed. Each EC number corresponds to a suggested name for the corresponding enzyme in the system of enzyme nomenclature [206]. Firstly, all reactions of the human iBrain671-Hs and mouse iBrain674-Mm brain models were searched on RGD [205] by using Enzyme Commision numbers (E.C) of the enzymes catalyzing the reactions, enzyme names, and metabolites to identify the homolog reactions for the rat. Since the Rat Genome Database (RGD) is the most comprehensive rat-specific database, including rat genetic-genomic-phenotypic data as well as comparative disease and molecular pathway literature information for mouse and human in addition to rat, it was used as the main source in the reconstruction process.

Reactions with no rat homolog in RGD were checked in the BRENDA [198] and Metabolic Atlas [256] databases using the same method. BRENDA is the most extensive data repository that contains literature-based functional and molecular enzyme information. The Metabolic Atlas is an online service that offers extensive

information about human metabolism. Several metabolic models, including Human-GEM, are integrated into Metabolic Atlas via a GEM-browser user interface. This interface makes it easy to get information on the appropriate reaction using a particular search term (keyword) linked to the names of the gene, protein, or chemical metabolites.

The EC numbers for some reactions in the reference human and mouse metabolic models were found to be out of date, and the appropriate EC numbers were revised using the BRENDA database. Some of the reactions with updated EC numbers are given in Table 4.1.

Table 4.1: Reactions with updated EC number.

Reaction ID	**iBrain671-Hs /iBrain585 Mm**	**Updated EC in rat GEM**
RA_lys7_c (R181)	1.3.8	1.2.4.2
RA_lys9_m; RN_lys9_m	4.2.1; 4.1.1.70	1.3.8.6
RA_da5_c	1.4.3.3	1.4.3.4
RA_chol18_c	1.14.13.72	1.14.18.9

It was observed that some reactions corresponding to the relevant EC number in RGD were not exactly the same as the reactions in the reference models, and there were some differences in the metabolite names. The differences in metabolite names were confirmed by checking BRENDA and the Human Metabolic Atlas for the respective reactions. It was determined by PubChem [207] that this difference in the metabolite names of the related reactions was due to the use of synonymous metabolite names.

During the manual curation process, additional genes were added to the rat GPR rules for certain reactions. Some gene information not included in the iBrain models was added to the model based on RGD [205] and iRno-GEM [131] (Table 4.3). The existence of rat homolog reactions and GPR rules in the iRno-GEM was also investigated to ensure that such reactions occur in the metabolism of rat brain.

Table 4.2: Examples of cases where a new gene is added to the Rat brain specific GEM model reconstructed in this study.

Reaction ID	GPR in iBrain671-Hs/ iBrain674-Mm	GPR in Rat brain-specific GEM reconstructed in this study
RA_gly16_c; RN_gly16_c	LDHA | LDHAL6A | LDHAL6B | LDHB | LDHC (Human) Ldhal6b | Ldha | Ldhb | Ldhc (Mouse)	Ldhb | Ldhc | Uevld | Ldhal6b | Ldha
RA_ppp1_c; RN_ppp1_c	G6PD (Human) G6pd2 | G6pdx (Mouse)	G6pd | H6pd
RN_ggc3_m	ASNS | GLS | GLS2 (Human) Gls | Gls2 (Mouse)	Gls | Gls2 | Nadsyn1

The distribution of metabolic reactions between different compartments within the cell has a significant impact on the predictive power of the metabolic network/model. Therefore, it is important to categorize similar reactions involving the same metabolite but occurring in different compartments at the compartment level. In this context, rat homolog reactions and the genes associated with these reactions were divided into cytosolic and mitochondrial compartments based on compartment information from the reactions in the human brain model [10].

Cells tend to generate proteins, lipids, carbohydrates, and nucleic acids to perform their fundamental biological activities under all circumstances.The process of biomass (macromolecule) production may be characterized by the assembly of biomass constituents such as proteins (Table 4.3), DNAs, RNAs (Table 4.5), lipids (Table 4.4) and carbohydrates (Table 4.6) that are created from precursor metabolites, such as amino acids, nucleotides, fatty acids and sugars [208]. The biomass reaction is an abstract reaction in a GEM, which includes the biomolecules required to produce a dry-weight unit of the cell. A brain-specific macromolecule synthesis (biomass) reaction (Table 4.7) and ATP-requirement reactions (non-growth associated) were introduced to iBrain766-Rn GEM to account for these fundamental functions of brain metabolism. The GAM (growth associated ATP maintenance) coefficient in macromolecule synthesis reactions was as used as 29.04 for both astrocyte and neuron [10, 209]. On the other hand, the rates of NGAM (non-growth associated ATP maintenance) reactions were used as 20 % of maximum ATP production rate for both

neuron and astrocyte [210]. The process for macromolecule synthesis was designed based on the proportions of macromolecules in brain samples that were previously determined [10].

Table 4.3: Amino acid coefficients of macromolecule synthesis reaction.

Protein Composition	**µmol/g Protein**
Alanine	328.25
Arginine	138.96
Asparagine	33.50
Aspartate	321.27
Cysteine	45.93
Glutamine	1236.93
Glutamate	2323.97
Glycine	875.06
Histidine	45.77
Isoleucine	43.28
Leucine	85.68
Lysine	58.01
Methionine	23.03
Phenylalanine	38.17
Proline	685.02
Serine	283.11
Threonine	121.08
Tryptophan	35.28
Tyrosine	241.76
Valine	159.94
Ornithine	49.31
GABA	618.21

Table 4.4: Lipid coefficients of macromolecule synthesis reaction.

Lipid Composition	µmol/gLipid
Cholesterol	129.18
Phosphatidyl-ethanolamine	60.6
Phosphatidyl-choline	58.35
Phosphatidyl-serine	23.44
Sphingomyelin	10.72
Phosphatidyl-inositol	5.99
Cardiolipin	2.84

Table 4.5: RNA coefficients of macromolecule synthesis reaction.

RNA composition	µmol/g RNA
AMP	516.58
CMP	861.05
GMP	976.05
UMP	516.58

Table 4.6: Carbohydrate coefficients of macromolecule synthesis reaction.

Carbohydrate composition	µmol/g Carbohydrate
Glucose-6-phosphate	3844

Table 4.7: Content of macromolecule synthesis reaction.

Macromolecular Composition	µmol/g
Carbohydrates	0.062
Lipid	0.401
Proteins	0.499
RNA	0.038

4.3. Conversion into a Mathematical Model

The rat brain-specific metabolic model, which contains information about reaction sets, EC numbers of reactions, reaction GPR rules, reaction names, and associated pathways, was initially constructed in Microsoft Excel file format after following the procedures above to build it. The model created in Excel format was then converted to SBML format (Systems Biology Markup Language) [211] using a script written in MATLAB. The rat brain-specific metabolic model converted to SBML format was read in MATLAB 2021a using the COBRA ToolBox *readCbModel()* function.

COBRA (Constraint-Based Reaction and Analysis) Toolbox is frequently used in the simulations of genome-scale metabolic models [212]. It has a broad range of applications in biology, biomedicine, and biotechnology. The general method is to depict the mechanistic connection between genotype and phenotype by modeling mathematically and computationally the constraints placed on the phenotype of a biochemical system by physicochemical principles, genetics, and environment [213].

Newly reconstructed rat brain specific GEM contains 994 reactions and 766 genes associated with 71 metabolic pathways/subsystems. The metabolic model unique to the rat brain was called iBrain766-Rn based on the number of genes included in the model. Here, 766 denotes the gene number, while Rn denotes the organism name (*Rattus norvegicus*). The detailed information about the model content is represented in Table 4.8 and Figure 4.2. Figure 4.2 shows how many reactions are available in each of 71subsystems in the GEM.

Table 4.8: Detailed information of the iBrain766-Rn model content.

Type of Reactions	Number of reactions
Total	994
Metabolic	745
Transport	138
Exchange	111

iBrain766-Rn	
Reactions	994
Genes	766
Metabolites	812
Subsystem/Pathway	71

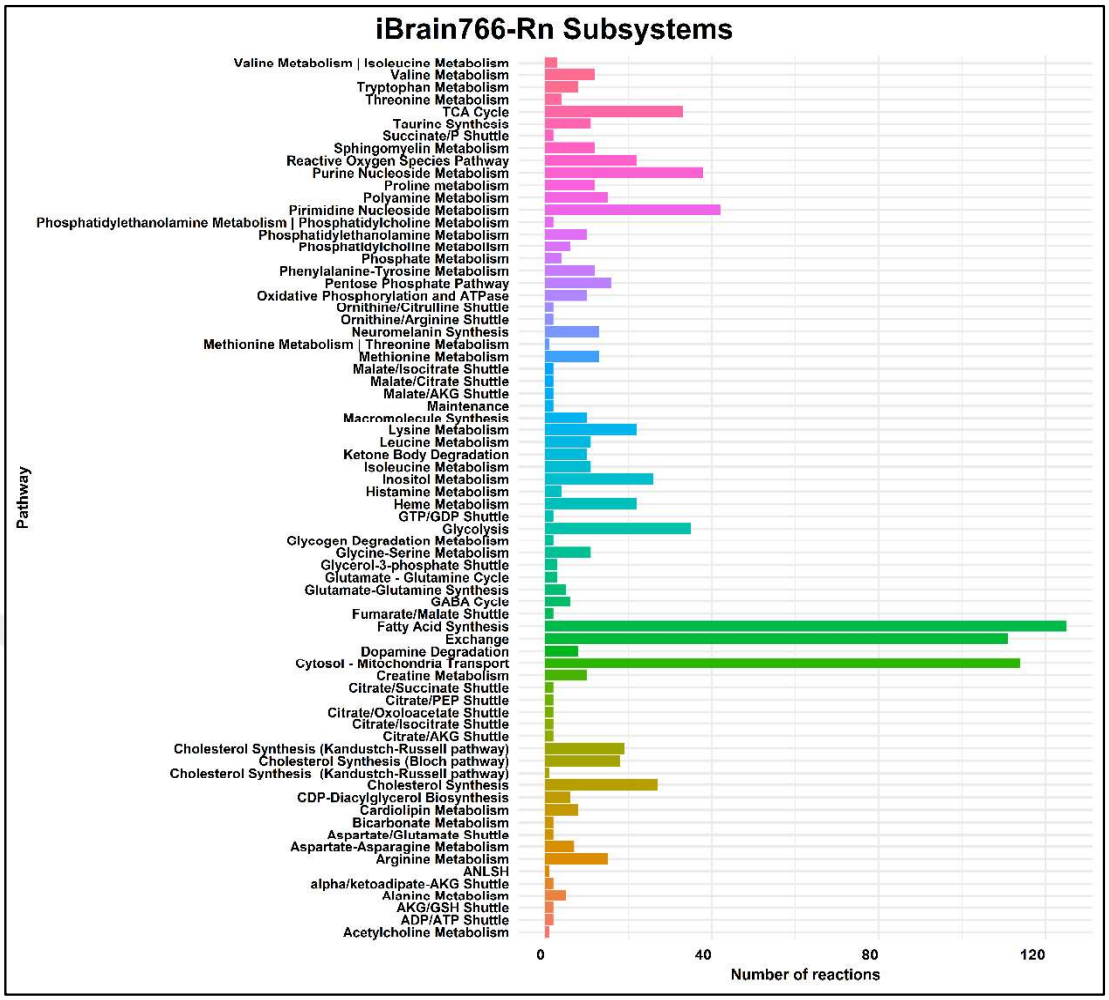

Figure 4.2: Pathway-Reaction distribution in iBrain766-Rn.

4.4. Evaluation and Analysis

After the final model is obtained, to determine whether the model can accurately simulate the brain metabolism of the relevant organism, the predicted flux values (reaction rates) obtained by simulating the model should be compared with the experimentally measured flux values [9].

FBA was used to validate newly reconstructed iBrain766-Rn model in terms of flux prediction. [118]. The validity of the prior brain-specific networks was established using flux balance analysis (FBA), and brain-specific measured fluxes in resting state were acquired from the literature. In a similar manner, FBA was used here to validate newly reconstructed iBrain766-Rn model in terms of flux prediction.

Before conducting the FBA analysis, experimentally determined constraint values were collected from the literature to reflect metabolic activity in the healthy rat

brain. Experimentally determined constraint values are based on CMR (Cerebral Metabolic Rate).

$CMR_{Glucose}$ consumption rate corresponding to healthy resting state is available in the literature for rat. $CMR_{Glucose}$ consumption rate has been reported to vary depending on the circumstances, whether the animal is awake or anesthesia-related. The measurement performed based on the anesthetic chemicals influences $CMR_{Glucose}$ value. Because various values for rat $CMR_{Glucose}$ during resting state have been reported in the literature, the $CMR_{Glucose}$ value was calculated as 0.604 μmol/g/hr by averaging these values. Resting state $CMR_{Glucose}$ values for rat are given in Table 4.9.

Table 4.9: Experimental resting-state $CMR_{Glucose}$ values for rat.

References	**$CMR_{Glucose}$**
Borgstrom et .al, (1976) [214]	0.76
Linde et. al, (1999) [215]	0.65
Madsen et. al, (1998) [216]	0.66
Du et. al, (2012) [217]	0.44
Nicola et. al, (1998) [218]	0.50

Astrocyte cells are responsible for about half of the glucose phosphorylation carried out in the blood [219]. The glucose consumption rates for each cell type (neurons and astrocyte) were therefore considered to be equal (0.302 μmol/g/hr).

CMR_{Oxygen} / $CMR_{Glucose}$ ratio is reported to be approximately 5.5 at the normal resting state [220, 221], and the CMR_{Oxygen} value was calculated as 3.322 μmol/g/hour based on this ratio. It is known that 30% of the total oxygen in the cerebral cortex is consumed by astrocytes (Gruetter et al., 2001). Oxygen consumption/uptake rates of astrocytes and neurons were calculated using this ratio as 0.997 and 2.325 respectively.

Since the respiratory coefficient (rCO2/rO2) in the brain has been reported to be in the range of 0.91-1.0 [222], the upper and lower bounds of CO2 production rates for astrocytes and neurons were calculated through this respiratory coefficient as (0.997, 0.907) and (2.325, 2.116) respectively .

Consumption rates of some amino acids in the resting state for the rat brain are available in the literature [223]. Constraints for the uptake rates of amino acids were

used for the following amino acids: cysteine, leucine, isoleucine, valine, tyrosine, tyrtophan, lysine, phenylalanine, histidine, arginine, ornithine, methionine, and threonine (Table 4.4). Ammonia constraints present in Table 4.10 were obtained from the literature [224].

The reaction rate of the pentose phosphate pathway in astrocytes was fixed to 6 % of the glucose utilization rate [225], while it was fixed to 5 % of the glucose utilization rate for neurons [226].

The rate of non-growth associated ATP maintenance (NGAM) reactions, which is needed for activities other than macromolecule synthesis, were calculated using 20% of the maximum ATP synthesis/production rate [210]. The GABA cycle rate was utilized at 25% of the glutamate-glutamine cycle rate [218, 227]. Ratio based constraints were set to the iBrain766-Rn model by introducing/adding an additional row to stoichiometric matrix. Some reactions (Glycogen metabolism reactions, β-Hydroxybutyrate, acetoacetate, bi-carbonate exchange reactions, and astrocyte-neuron lactate shuttle reaction) are known to be inactive at resting state. Thus, the upper and lower bounds of these reactions were fixed to zero to inactivate them at resting state.

The constraints used to simulate resting state reaction rates in the rat brain are given in Table 4.10.

Table 4.10: The constraints utilized to simulate resting state flux distributions with iBrain766-Rn model. Glucose, oxygen, amino acid, and ammonia consumption/uptake rates and macro molecule synthesis reaction rates are given in the form of μmol/g dry tissue/min. A is Astrocytes, N is Neurons.

Healthy Resting State		
Reaction	**Exchange Flux – lower bound (LB)**	**Exchange Flux – upper bound (UB)**
Glucose	A: 0.302 N: 0.302	A: 0.302 N: 0.302
Oxygen	A: 0.9966 N: 2.3254	A: 0.9966 N: 2.3254
Carbon dioxide	A: 0.9069 N: 2.1161	A: 0.9966 N: 2.3254
Glycogen	A: 0	A: 0
Acetoacetate	0	0
Bicarbonate	A: 0 N: 0	A: 0 N: 0
β-Hydroxybutyrate	0	0
Lactate shuttle	0	0
Macromolecule synthesis	A: 1.32×10^{-6} N: 1.32×10^{-6}	A: 1000 N: 1000
Cysteine	A: 0.009	A: 0.009
Leucine	A: 0.0145	A: 0.0145
Isoleucine	A: 0.004	A: 0.004
Valine	A: 0.0018	A: 0.0018
Arginine	A: 0.0067	A: 0.0067
Ammonia	A: 0.002	A: 0.002
Tyrosine	0.0041	0.0041
Tryptophan	0.0082	0.0082
Lysine	0.0103	0.0103
Phenylalanine	0.0132	0.0132
Histidine	0.0025	0.0025
Ornithine	0.0031	0.0031
Methionine	0.0017	0.0017
Threonine	0.0008	0.0008
Taurine	-1000	1000
ATP maintenance		20% of maximum ATP synthesis
Glutamine		25% of GABA
Pentose phosphate pathway		5 % of glucose uptake

All simulations were carried out through the COBRA Toolbox [213] and gurobi optimizer/solver [228] in MATLAB. Gurobi is a mathematical optimization problem

solver, and it can be used to handle both linear and quadratic mathematical optimization problems.

Maximizing biomass/macromolecule synthesis rate is utilized as the primary objective function in microorganisms but, the primary objective function used in mammals may vary depending on the aim of the research. Even though the primary objective function in mammals is not to maximise the amount of biomass (macromolecules), cell must maintain its fundamental biological functions such as the production of protein, lipid and nucleic acid. iBrain766-Rn GEM contains separate macromolecule synthesis reactions for astrocytes and neurons.The macromolecule synthesis reactions were maximized individually for neurons and astrocytes according to the provided Table 4.10 constraints. The maximum macromolecule reaction rates were calculated as 1.32×10^{-5} for both neurons and astrocytes. To ensure the activity of fundamental biological processes such as protein, lipid, and nucleic acid production at the basal level, lower boundaries of macromolecule synthesis reactions were fixed to 10% of the calculated maximum macromolecule synthesis rate value. The glutamate/glutamine/GABA cycle is critical for brain metabolism, and the glutamate/glutamine/GABA cycle has been utilized as a primary objective function in previous studies to predict healthy-state metabolic fluxes in human and mouse brain [9, 11, 136]. Thus, the maximization of the fluxes through glutamate/glutamine/GABA cycle was also used as the primary objective function in the simulation of resting state rat brain metabolism in this thesis study. As linear objective functions are often accompanied with an alternate optima problem, there may be many solutions with the same objective function value. In this respect, the major issue is that only one solution is provided by utilized solvers, such as gurobi, even though there is more than one solution. The minimization of the sum of squares method [159, 160] (Equation (2.6)) for all fluxes was used as a second objective function using quadratic programming to find the solution that requires the least amount of enzymes among alternate optima. After setting the constraints for the iBrain766-Rn model, linear programming was utilized to maximize the sum of the rates of three reactions corresponding the Glutamate/Glutamine/GABA cycle reactions, which were chosen as the main objective function for the rat brain-specific metabolic model. Then, the maximum value of the sum of the rates of the three reactions was used as an additional constraint by adding the indices of these reactions as an additional row to stoichiomatric matrix for the secondary optimization. Here, quadratic programming algorithm was used to

optimize a second optimization/objective function to deal with the alternate optima problem. In this manner, it is guaranteed that the model performs its biological activities with the least amount of enzyme.

Table 4.11: Comparison of predicted rat flux ratios with the human experimental fluxes and predicted flux ratios for human and mouse at resting state.

Percentage of Flux Ratio	**Prediction for Rat**	**Prediction for Mouse** [11]	**Prediction for Human** [9]	**Experimental for Human**
Lactate production flux with respect to $CMR_{Glucose}$	11.02	8.3	7.2	3-9 [222, 229–231]
Glutamate-glutamine cycle flux with respect to $CMR_{Glucose}$	74.27	73.4	73.1	40-80 [232–234]
Relative oxidative metabolism of astrocyte to total oxidative metabolism	30.07	28.8	33.9	30 [219, 232, 235]
Total PPP flux with respect to $CMR_{Glucose}$	5.5	6.0	5.5	3-6 [225, 226]
Pyruvate carboxylase flux with respect to $CMR_{Glucose}$	9.03	12.4	11.1	10 [232, 235, 236]

Rat brain fluxes (reaction rates) at resting state were estimated through the FBA approach using the constraints in Table 4.10. Since it is known that most of the cellular processes in mammals are similar, the simulation results using the rat brain-specific GEM (iBrain766-Rn) are expected to be close to the experimental human brain reaction rate values. As five reactions, which are lactate release, Glu-Gln cycle, total pentose phosphate pathway (PPP), astrocytic oxidative phosphorylation and pyruvate carboxylase reactions, are essential stages in central cellular carbon and neurotransmitter metabolisms [9], these reactions were used as reference reactions to evaluate the flux predictions at resting state/healthy state in this study. Experimental flux ratios for these reactions relative to glucose-uptake are reported to be 3-6 %, 40-80%, 3-6%, 30% and 10% respectively at resting state for human (Table 4.11). The predicted flux distributions for these reactions were converted to flux ratios relative to glucose uptake (Table 4.12), and these predicted flux ratios were compared with the human experimental flux ratios.

Table 4.12: Calculations of percentage of relative flux rates with respect to glucose uptake rate.

Percentage of Flux Ratio	**Calculation**
Lactate production flux with respect to $CMR_{Glucose}$	100 x ((n_RA_gly16_c) / (n_Ex_A_Glc + n_Ex_N_Glc))/2
Glutamate-glutamine cycle flux with respect to $CMR_{Glucose}$	100 x ((n_RAN_t3_c) / (n_Ex_A_Glc + n_Ex_N_Glc))/2
Relative oxidative metabolism of astrocyte to total oxidative metabolism	100 x ((n_RA_tca1_m) / (n_RA_tca1_m + n_RN_tca1_m))
Total PPP flux with respect to $CMR_{Glucose}$	100 x ((n_RA_ppp1_c + n_RN_ppp1_c) / (n_Ex_A_Glc + n_Ex_N_Glc))/2
Pyruvate carboxylase flux with respect to $CMR_{Glucose}$	100 x ((n_RA_gly17_m) / (n_Ex_A_Glc + n_Ex_N_Glc))/2

In comparison with human experimental flux ratios, the predicted flux distributions via the dual objective framework with primary maximisation of rates of glutamate/glutamine/GABA cycle reactions and secondary minimization of sum of squares of reaction rates were found to be within the range of experimental ratios (Table 4.11). According to Table 4.11, the predicted flux ratio for lactate is slightly higher than the experimental flux ratio. Such a slight difference between the estimated flux ratio and the experimental flux ratio is thought to be related to the rat-specific constraints (especially for glucose uptake constraint) used. The predicted flux ratios were also compared with previously reported predicted flux ratios for human and mouse models. According to Table 4.11, except for slight differences, the predicted flux ratios in the rat model are very close to the predicted flux ratios for both human and mouse models. In this regard, the reconstructed rat brain-specific GEM, iBrain766-Rn, can provide reliable predicted flux values for resting healthy rat brain once glutamate/glutamine/GABA cycle reactions are maximized.

5. RECONSTRUCTION AND CONSTRAINT-BASED ANALYSIS OF CONTEXT-SPECIFIC MODELS

5.1. Transcriptome Data Integration Algorithms

The traditional constraint-based modeling methods do not take gene regulations and protein expression mechanisms into consideration when making phenotypic predictions. Although GEMs contain all known metabolic reactions, a portion of these reactions are active in any given context, such as a specific stage of development, cell type or environment. There are many cell types having distinct functions, and it is critical to build condition-specific metabolic models in order to predict metabolic phenotypes and characterize metabolic effects of diseases [237].

Rat transcriptome datasets associated with PD (Table 3.1) were compiled and subjected to clustering analysis in Chapter 3. Here, they were utilized to reconstruct condition-specific metabolic models. The iMAT-family algorithms do not necessitate a primary objective function to build a context-specific model. Additionally, iMAT-like family techniques guarantee the highest possible degree of agreement between reaction states (active/inactive) and datasets (expressed/unexpressed). This is accomplished via the use of the MILP (Mixed Integer Linear Programming) paradigm. In this context, iMAT-like family algorithms, including iMAT [170] and INIT algorithms [171] have been used in this thesis study to integrate the transcriptome datasets with the iBrain766-Rn metabolic model.

First, iBrain766-Rn model was read in MATLAB environment through *readCbModel()*. Then, experimental constraints given in Table 5.1 for both healthy and disease (Parkinson's disease) states were set. Measured resting-state flux data of rat brain were obtained from the literature and used as constraints to simulate healthy controls. On the other hand, there are no measured fluxes for the disease (PD) situations, and only a few metabolic processes have been demonstrated to exhibit increased or decreased activity in PD, without magnitudes of reaction rates. The measurements from literature show that the consumption of cerebral glucose is reduced by 2-32%, while oxygen consumption is decreased by 6-34% in PD [73, 238]. Since glucose consumption rate is known to decline in PD [74], Because glucose

consumption is essential for central carbon metabolism, its lower limit was set to half of the value for the healthy state for both astrocytes and neurons to allow for a decrease while avoiding extreme reductions, but their upper limits remained the same. The lower bound of oxygen consumption rate was set to 1 x 10^{-2}. 10% of glucose consumption/uptake rate was taken as the upper limit for amino acids in this study based on previous studies [9], and also the lower bounds of amino acid consumption rates were constrained to 1 x 10^{-6} to prevent the amino acid uptake rates being predicted as zero. The lower bound of macromolecule synthesis reaction was set to 1 x 10^{-6} to ensure there is macromolecule synthesis at the basal level. The basal level represents the minimum level of components that the cell must produce to perform basic biological functions. After setting the constraints in Table 5.1 to iBrain766-Rn, previously normalized six rat transcriptome datasets (Table 3.1) were imported to MATLAB environment. Then, each transcriptome dataset was integrated into iBrain766-Rn model through iMAT and INIT algorithms to generate context-specific models. The lower bound (*lb)* for reactions found to be inactive in the reduced models generated by iMAT or INIT were fixed to 1 x 10^{-8} in later simulations.

Table 5.1: Healthy Control and Disease State Constraints.

Healthy Controls			Parkinson's Disease		
Reaction	**Exchange Flux – lower bound**	**Exchange Flux – upper bound**	**Reaction**	**Exchange Flux – lower bound**	**Exchange Flux – upper bound**
Glucose	A: 0.302 N: 0.302	A: 0.302 N: 0.302	Glucose	A: 0.1510 N: 0.1510	A: 0.302 N: 0.302
Oxygen	A: 0.9966 N: 2.3254	A: 0.9966 N: 2.3254	Oxygen	A: 1 x 10^{-2} N: 1 x 10^{-2}	A: 1000 N: 1000
Carbon dioxide	A: 0.9069 N: 2.3254	A: 0.9966 N: 2.1161	Carbon dioxide	A: 0 N: 0	A: 1000 N: 1000
Glycogen	A: 0	A: 0	Glycogen	A: 0	A: 0
Acetoacetate	0	0	Acetoacetate	0	0
Bicarbonate	A: 0 N: 0	A: 0 N: 0	Bicarbonate	A: 0 N: 0	A: 0 N: 0
B-Hydroxybutyrate	0	0	B-Hydroxybutyrate	0	0
Lactate shuttle	0	0	Lactate shuttle	0	0
Macromolecule synthesis	A:1.32 x 10^{-6} N:1.32 x 10^{-6}	A:1000 N:1000	Macromolecule synthesis	A:1 x 10^{-6} N:1 x 10^{-6}	A:1000 N: 1000
Cysteine	A: 0.009	A: 0.009	Cysteine	A: 1 x 10^{-6}	A:0.0604
Leucine	A: 0.0145	A: 0.0145	Leucine	A: 1 x 10^{-6}	A: 0.0604
Isoleucine	A: 0.004	A: 0.004	Isoleucine	A: 1 x 10^{-6}	A: 0.0604
Valine	A: 0.0018	A: 0.0018	Valine	A: 1 x 10^{-6}	A: 0.0604
Arginine	A: 0.0067	A: 0.0067	Arginine	A: 1 x 10^{-6}	A: 0.0604
Tyrosine	0.0041	0.0041	Tyrosine	1 x 10^{-6}	0.0604
Tryptophan	0.0082	0.0082	Tryptophan	1 x 10^{-6}	0.0604
Lysine	0.0103	0.0103	Lysine	1 x 10^{-6}	0.0604
Phenylalanine	0.0132	0.0132	Phenylalanine	1 x 10^{-6}	0.0604
Histidine	0.0025	0.0025	Histidine	1 x 10^{-6}	0.0604
Ammonia	0.002	0.002	Arginine	1 x 10^{-6}	0.0604
Arginine	0.0067	0.0067	Ornithine	1 x 10^{-6}	0.0604
Ornithine	0.0031	0.0031	Methionine	1 x 10^{-6}	0.0604
Methionine	0.0017	0.0017	Threonine	1 x 10^{-6}	0.0604
Threonine	0.0008	0.0008	Taurine	1000	-1000
Taurine	-1000	1000			
ATP maintenance		20% of maximum ATP synthesis	ATP maintenance		20% of maximum ATP synthesis
Glutamine		25% of GABA			
Pentose phosphate pathway	A: 1 x 10^{-6} N: 1 x 10^{-6}	5 % of glucose uptake			

5.1.1. iMAT Algorithm

Since the COBRA Toolbox uses a single column expression matrix to map transcriptome data into the model, absolute expression values of genes were averaged separately for the control and disease cases in each of the comparison groups. Then, the averaged gene expression data of eleven comparison groups were mapped separately onto the iBrain766-Rn model based on GPR associations using the *mapExpressionToReactions* function in the COBRA Toolbox. Each reaction in the model was scored based on gene expression levels and GPR associations. The lowest expression value among all the genes associated with an enzyme complex (AND rule) and the highest expression value among all the genes associated with an isozyme (OR rule) were used to calculate the reaction score for each reaction. As a result of the mapping, the reaction expression vectors were obtained for each of 11 comparison groups.

IMAT discretizes reactions into high, low, and moderate expressions based on a user-defined threshold. The condition-specific models including healthy control and disease cases were created using the iMAT function from Cobra Toolbox 3.0. The 25^{th} and 75^{th} percentiles of averaged absolute gene expression data for each comparison group were utilized as lower and upper thresholds in iMAT, respectively, while all other parameters were kept at their default settings. To run iMAT algorithm on the iBrain766-Rn model, reaction expression vector, which was obtained at the mapping step, and lower and upper thresholds were used as inputs in iMAT function. Finally, twenty-two iMAT-based context-specific models corresponding to eleven healthy control and eleven disease case models were obtained once transcriptome datasets were integrated to iBrain766-Rn model. Numbers of reactions and metabolites in the reduced models created by iMAT are given in Table 5.2.

Table 5.2: Reaction and metabolite content of iMAT models.

Data ID	Reaction number	Metabolite number
GSE24233_ST_Control_iMAT	781	695
GSE24233_ST_Treated_6ohda_iMAT	825	725
GSE58710_SN_wk1_control_iMAT	774	681
GSE58710_SN_wk1_6OHDA_iMAT	828	726
GSE58710_SN_wk2_control_iMAT	779	685
GSE58710_SN_wk2_6OHDA_iMAT	813	719
GSE58710_SN_wk4_control_iMAT	770	679
GSE58710_SN_wk4_6OHDA_iMAT	824	724
GSE58710_SN_wk6_control_iMAT	792	694
GSE58710_SN_wk6_6OHDA_iMAT	827	726
GSE58710_SN_wk16_control_iMAT	775	680
GSE58710_SN_wk16_6OHDA_iMAT	827	728
GSE74382_DST_control_iMAT	750	669
GSE74382_DST_lesion_iMAT	770	689
GSE93695_ST_control_iMAT	760	676
GSE93695_ST_PD_iMAT	783	692
GSE71968_HMB_wt_iMAT	757	671
GSE71968_HMB_DJ1_KO_iMAT	799	708
GSE150646_FC_5mnth_wt_iMAT	745	669
GSE150646_FC_5mnth_TG_iMAT	757	679
GSE150646_FC_12mnth_wt_iMAT	754	674
GSE150646_FC_12mnth_TG_iMAT	772	686

5.1.2. INIT Algorithm

The INIT algorithm [171] was used to integrate the rat transcriptome datasets into the model using the COBRA Toolbox *INIT()* function. Actual (absolute/unlog) expression values were used when calculating the weight scores. Absolute expression values of genes were averaged separately for the control and disease cases in each of the comparison groups. The INIT method utilizes a single-column weight vector comprising positive and negative values as input to identify condition-specific active/inactive reactions. After reading the transcriptome datasets in the MATLAB environment, weight scores for each gene in eleven comparison groups were computed using the following formula [171] for healthy and disease states:

$$Weight_{Control} = 5\log\left(\frac{Control}{Threshold}\right) \quad (5.1)$$

$$Weight_{Disease} = 5\log\left(\frac{Disease}{Threshold}\right) \quad (5.2)$$

Since only one tissue (Brain) was used in this study, INIT requires a threshold value to calculate gene weight scores. Thus, the 25th percentiles of averaged absolute gene expression data for each comparison group were utilized as threshold to calculate gene weight scores. Following the calculation of gene weight scores in each dataset, gene weight scores were mapped onto the iBrain766-Rn model reactions using the COBRA Toolbox *mapExpressionReactions() function*. The lowest weight score among all the genes associated with an enzyme complex (AND rule) and the highest weight score value among all the genes associated with an isozyme (OR rule) were used to calculate the reaction weight score for each reaction. As a result of the mapping, a reaction weight vector was obtained for each comparison. To run INIT algorithm, iBrain766-Rn model and reaction weight vector were used as inputs in INIT function. Twenty-two INIT based context-specific models corresponding to eleven healthy control and eleven disease case models were obtained. Numbers of reactions and metabolites for the INIT-derived reduced models are given in Table 5.3.

Table 5.3: Reaction and metabolite content of INIT models.

Data ID	Reaction number	Metabolite number
GSE24233_ST_Control_INIT	812	693
GSE24233_ST_Treated_6ohda_INIT	873	735
GSE58710_SN_wk1_control_INIT	805	689
GSE58710_SN_wk1_6OHDA_INIT	884	747
GSE58710_SN_wk2_control_INIT	804	688
GSE58710_SN_wk2_6OHDA_INIT	867	733
GSE58710_SN_wk4_control_INIT	805	688
GSE58710_SN_wk4_6OHDA_INIT	868	734
GSE58710_SN_wk6_control_INIT	804	688
GSE58710_SN_wk6_6OHDA_INIT	897	756
GSE58710_SN_wk16_control_INIT	803	687
GSE58710_SN_wk16_6OHDA_INIT	883	746
GSE74382_DST_control_INIT	785	683
GSE74382_DST_lesion_INIT	860	740
GSE93695_ST_control_INIT	798	693
GSE93695_ST_PD_INIT	874	738
GSE71968_HMB_wt_INIT	821	702
GSE71968_HMB_DJ1_KO_INIT	876	739
GSE150646_FC_5mnth_wt_INIT	794	688
GSE150646_FC_5mnth_TG_INIT	868	746
GSE150646_FC_12mnth_wt_INIT	778	678
GSE150646_FC_12mnth_TG_INIT	866	746

5.2. Constraint-Based Analysis of Context-Specific Models

After obtaining the context-specific models (Control healthy, and Parkinson's disease) through iMAT and INIT algorithms, constraint-based analysis was performed on the obtained context-specific models to calculate the flux distributions for healthy and disease states. The maximization of glutamate/glutamine/GABA cycle reactions as a primary objective function and the minimization of the squared sum of all internal fluxes as a secondary objective function were used to simulate healthy control states. FBA method [118] was used as a CBM method for healthy control state flux predictions whereas MOMA [161], LseiFBA [74] and mLseiFBA were used as CBM methods for Parkinson's disease state flux predictions. All optimization problems were solved using academic Gurobi solver v9.1 (Gurobi Optimization LLC, 2021) by using $changeCobraSolver$ ('$gurobi$','all') function in COBRA Toolbox [213].

As in resting state simulations, five reactions (lactate release, Glu-Gln cycle, total pentose phosphate pathway (PPP), astrocytic oxidative phosphorylation and pyruvate carboxylase reactions) were used as reference reactions to evaluate the flux predictions at different conditions including healthy control and Parkinson's disease state.

5.2.1. Healthy Control Flux Prediction

FBA analysis [118] of twenty-two context-specific models, including eleven iMAT-based control models and eleven INIT-based control models, were performed in two steps to predict healthy control flux distributions. Firstly, linear programming was utilized to maximize the summed rates of three reactions corresponding the Glutamate/Glutamine/GABA cycle reactions [136], which were chosen as the main objective function for the rat brain-specific metabolic model. Then, minimization of squared sum of internal fluxes approach [136] through quadratic programming was used to remove the alternate optima problem in FBA solution vector after fixing the primary objective value at its optimum.

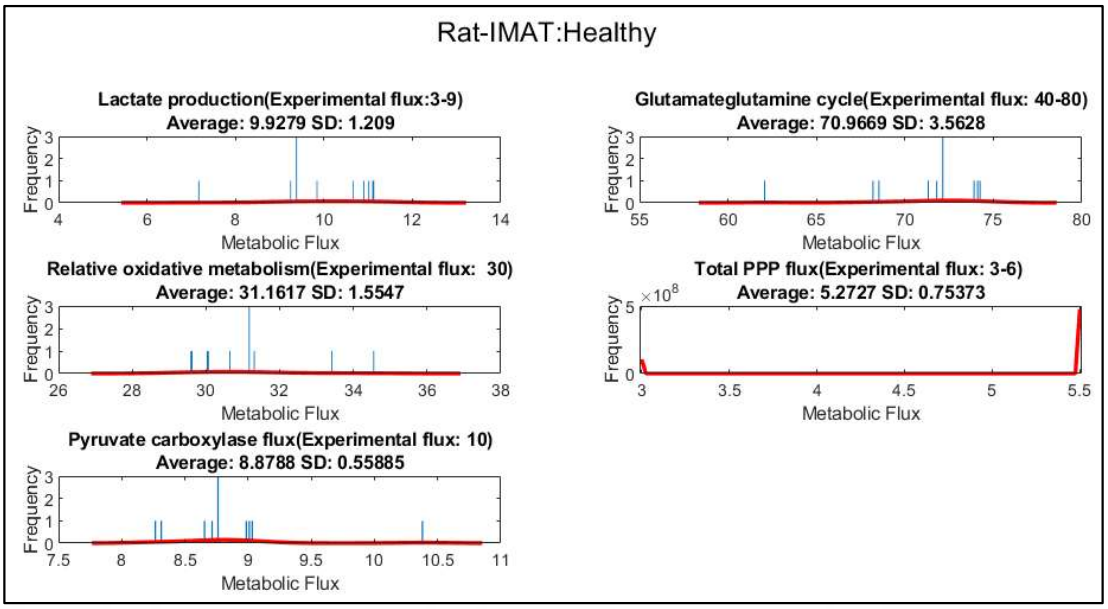

Figure 5.: Flux-distribution predictions of iMAT-based healthy models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

Figure 5.1 provides five histograms for the reference reactions, and each bar on the x axes of histograms refer to the iMAT-based predicted flux ratio for the corresponding reference reaction. Once Figure 5.1 is examined, it is observed that the majority of samples (on average) for iMAT-based healthy models have a predicted flux ratio that is consistent with the experimental flux ratios given in Table 4.11 [9, 136]. Although there are some variations in predicted flux ratios across groups, this variability is determined to be acceptable. The variability across groups in terms of flux prediction ratios may be associated with the corresponding transcriptome data.

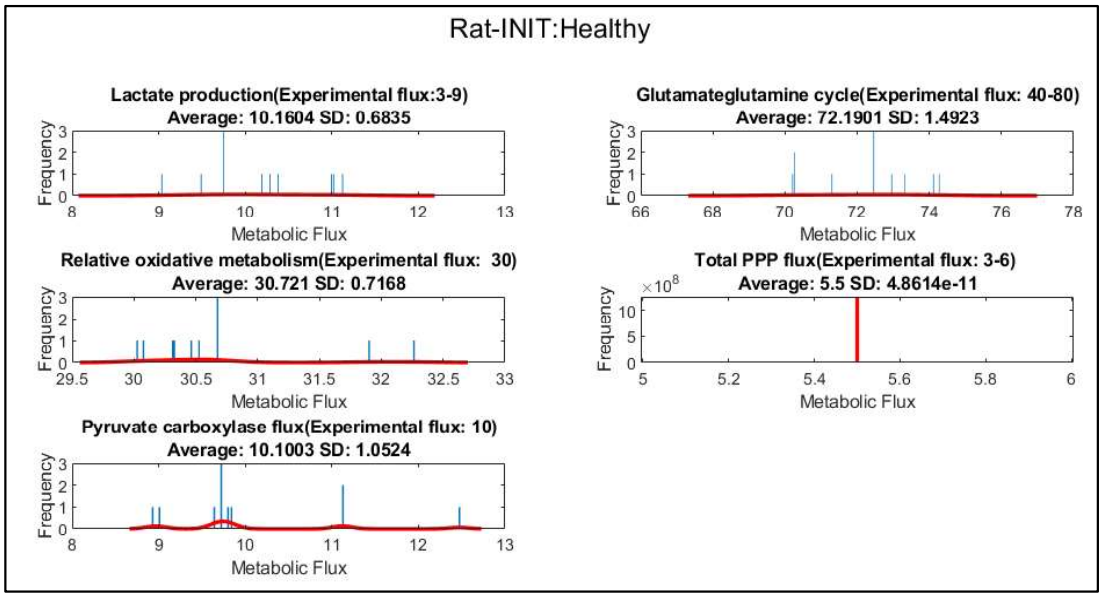

Figure 5.2: Flux-distribution predictions of INIT-based healthy control models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

Similarly, FBA and enzyme usage minimization analysis on the INIT-based healthy models revealed consistent predictions with previous observations (Figure 5.2). The variability and standard deviation across groups in terms of predicted flux ratios for INIT-based healthy models are not large, as shown in Figure 5.2. Compared to iMAT, similar flux ratios were predicted by INIT-based healthy models. However, INIT was found to give slightly better-predicted flux ratios for some reactions, especially for the glutamate/glutamine cycle and relative oxidative metabolism reactions.

The predicted flux ratios for certain reactions in both iMAT and INIT based models (Figure 5.1, Figure 5.2) were found to vary from the predicted flux ratio at resting state (Table 4.5) with no transcriptomic data incorporation, as expected. The difference in the predicted flux ratios between the healthy control state and resting state is mainly associated with the usage of transcriptome data to generate healthy control state models. The results indicate that some flux ratios were better predicted with condition-specific healthy control models, including iMAT and INIT-based models.

5.2.2. Flux Distribution Prediction in Parkinson's Disease

Minimization of metabolic adjustment (MOMA) [161], Least squares equalities and inequalities (LseiFBA) [74, 239], and a modified linear Least squares equalities and inequalities (mLseiFBA) were applied to estimate disease state flux distributions.

As the cell objective is unknown during PD, no explicit reaction can be defined as the objective function. Instead, an alternative objective function is required. MOMA has the system objective defined as the minimization of Euclidean distance flux difference between the reference (healthy) flux distribution and disease flux distribution (Equation (2.8)). Thus, healthy control flux predictions from iMAT-based and INIT-based models, which were performed in the previous section, were used as initial flux distributions in MOMA to predict flux distributions in Parkinson's disease. On the other hand, LseiFBA seeks to maximize correlation between changes in mRNA expression and fluxes. LseiFBA minimizes the Euclidean distance between initial disease flux vector ($v_{ref,D}$) and disease flux distribution (Equation (2.11)). Unlike LseiFBA, Modified linear LseiFBA (mLseiFBA) minimizes absolute flux difference between initial flux and disease distribution (Equation (2.12)).

Although there is no exact flux value information for Parkinson's disease in literature, reduced glucose and oxygen utilization and increased lactate generation in the cerebral cortex have been reported [74]. It has been shown that the PPP has a critical role in oxidative stress tolerance and the production of resources for biological synthesis. It was implicated by Dunn et al. that PPP dysregulation contributes to the development of Parkinson's disease, but they also reported enhanced NADPH generation by the PPP [59]. In a computational biology-based research, it was discovered that the flux value of the pentose phosphate pathway had increased [74].

MOMA approach was applied in two steps: (i) The upper and lower bounds for reactions found to be inactive in disease state models were fixed to $1\ x\ 10^{-8}$. (ii) Since MOMA attempts to calculate the smallest possible Euclidean distance between healthy state flux distribution and perturbed state (disease) flux distribution (Equation (2.8)), quadratic programming was used to predict disease state flux distributions. Healthy control flux predictions as initial fluxes were used to predict the Parkinson's disease flux distributions for twenty-two context-specific models (11 iMAT-models, 11 INIT-models) generated before (Table 5.2, Table 5.3).

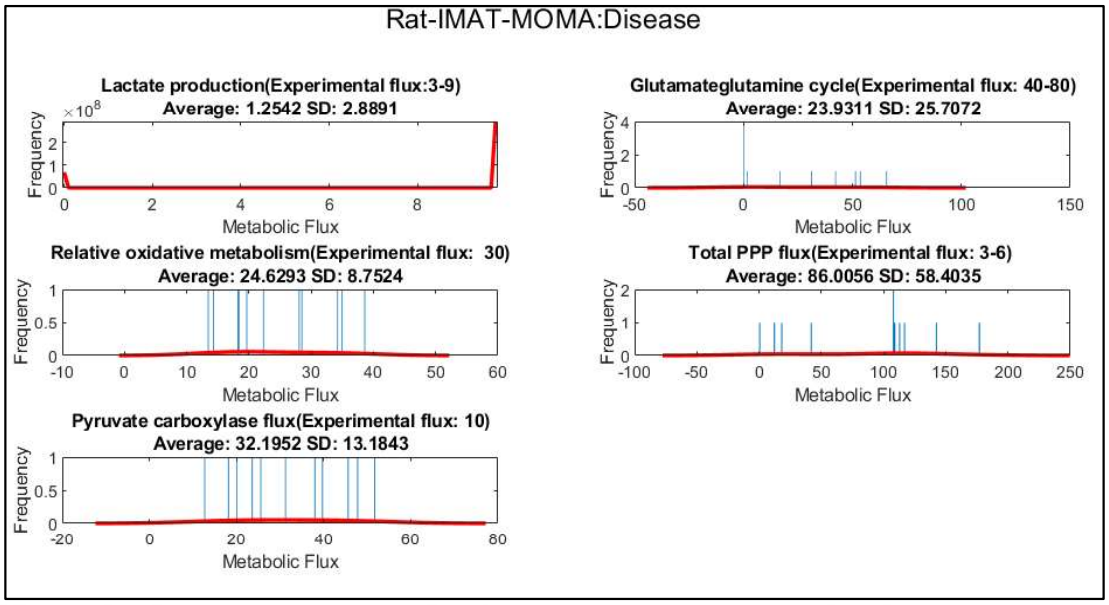

Figure 5.3: MOMA analysis on iMAT based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

MOMA analysis on eleven iMAT disease models revealed a decrease in the predicted flow rate for both the Glutamate-Glutamine cycle [240, 241] and relative oxidative metabolism compared to the resting state, consistent with previous research, as shown in Figure 5.3. However, the predicted lactate production flux ratio was expected to increase in Parkinson's disease, but there was a decrease in the predicted lactate production flux ratio. When compared to the healthy control state, the predicted total pentose phosphate pathway (PPP) flux ratio seems to be substantially increased in the Parkinson's disease state. Predicted increase in the PPP flux ratio was expected but the average of predicted PPP flux ratio was extremely high.

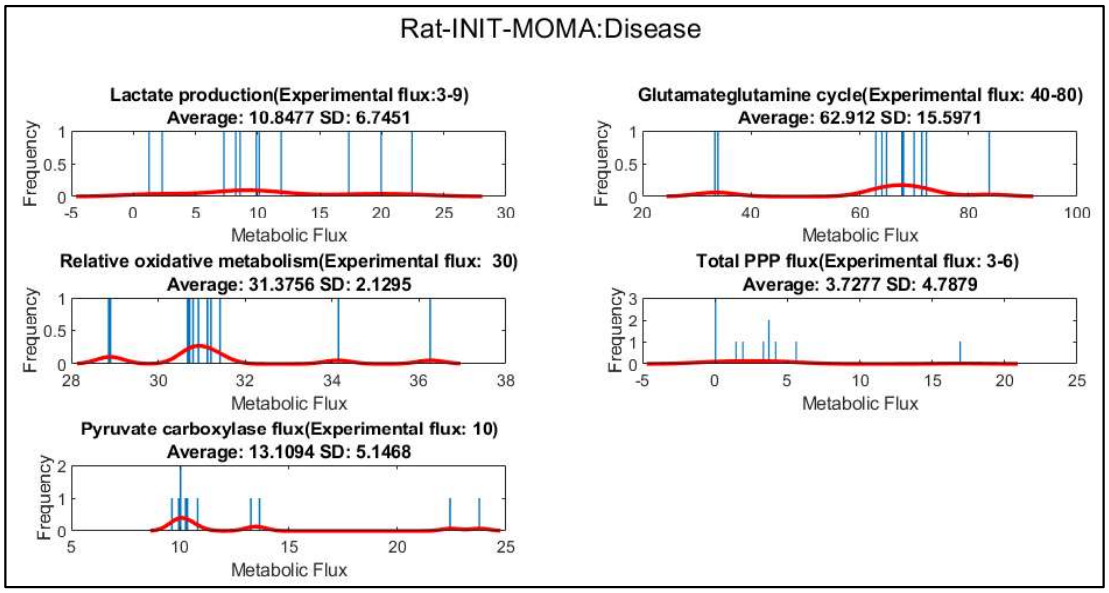

Figure 5.4: MOMA analysis on INIT based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

MOMA analysis on eleven INIT-based disease models showed that there was a slight decrease in the average of predicted flux ratio for Glutamate-glutamine cyle and total PPP while a slight increase in the average of predicted relative oxidative metabolism and pyruvate carboxylase flux ratios (Figure 5.4). Moreover, there was a slight increase in the lactate production flux ratio as it was expected [74]. On the other hand, an increase was expected in PPP flux prediction ratio [74] for the disease case, but the majority of eleven INIT disease models showed a decreased flux ratio pattern for total PPP flux ratio. These results may be affected by the used transcriptome data integration algorithm (iMAT, INIT, etc.). Compared to iMAT-MOMA results, INIT-MOMA results were found to give better flux prediction for lactate production and relative oxidative metabolism.

LseiFBA approach was conducted in four steps: (i) The upper and lower bounds for reactions found to be inactive in disease state models were fixed to $1\ x\ 10^{-8}$ (ii) Fold change value was calculated by dividing the disease GPR-mapped reaction expression vector to control reaction expression vector , (iii) Initial flux vector(s) were calculated by multiplying the healthy state flux distribution vector with the gene fold change values (Equation (2.10)), (iv) The flux estimation of the disease case was performed via utilizing an objective function that minimizes the Euclidean distance between initial disease flux vector ($V_{ref,D}$) and disease state flux distribution (Equation (2.11)) for twenty-two context-specific models (11 iMAT-models, 11 INIT-models).

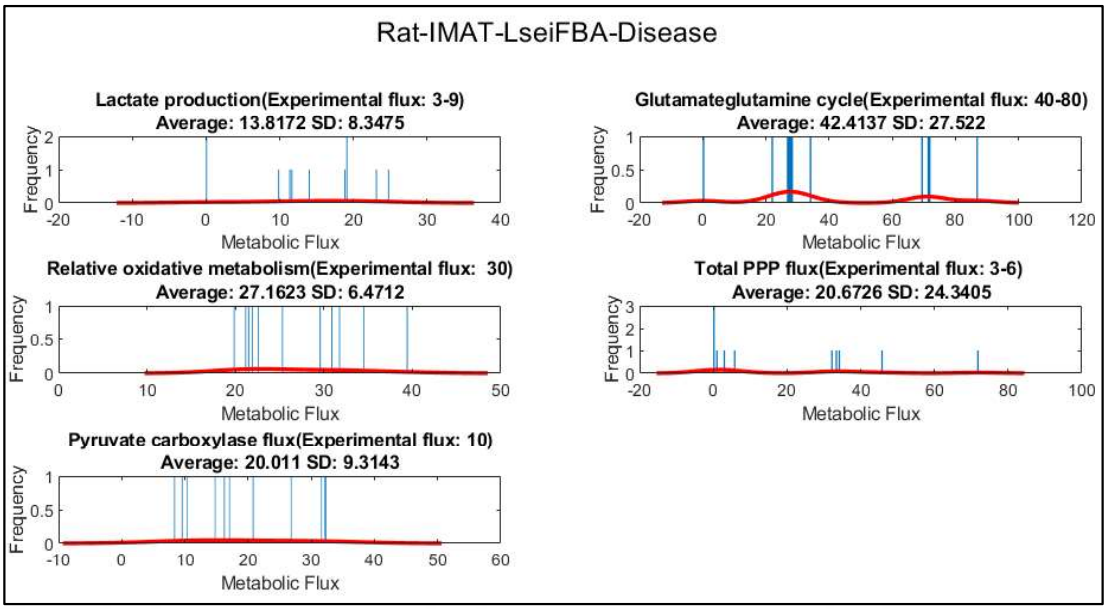

Figure 5.5: LseiFBA analysis on iMAT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

Figure 5.5 depicts consistent predicted flux ratios with previous literature observations for LseiFBA analysis based on iMAT disease models. As it was expected, there was an increase in the predicted lactate production ratio [74] and total PPP flux ratio [74], and a decrease in glutamate-glutamine cycle fluxes [240, 241].

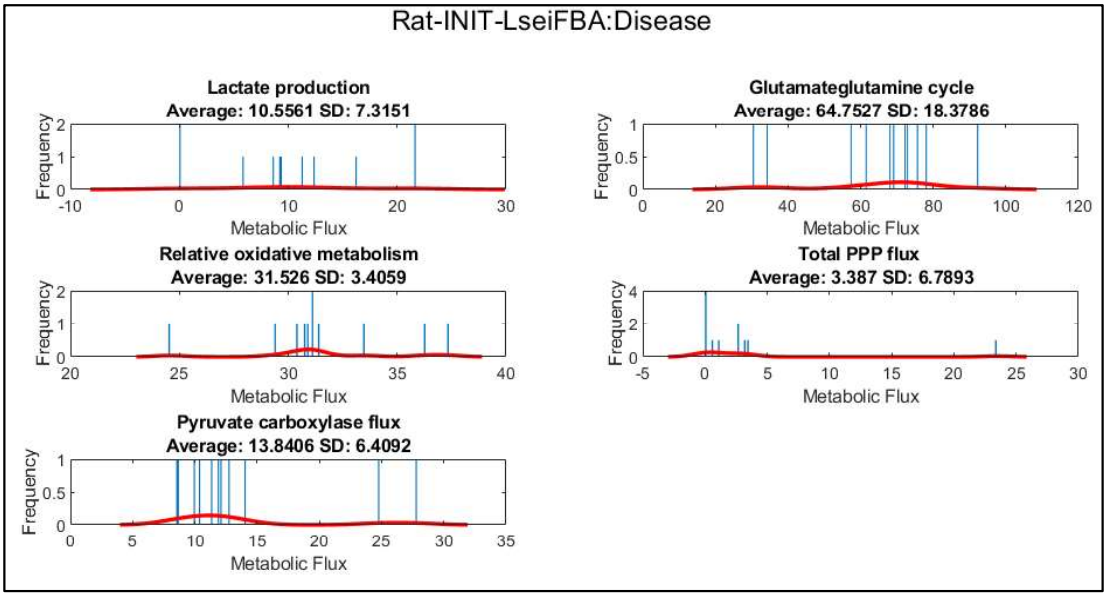

Figure 5.6: LseiFBA analysis on INIT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

Once LseiFBA analysis was conducted on eleven INIT based disease models, similar results to INIT- MOMA results (Figure 5.4) were obtained (Figure 5.6). Compared to INIT-LseiFBA, iMAT-LseiFBA models were found to give better flux predictions for some reactions including lactate production and total PPP.

Since mLseiFBA method minimizes absolute flux differences between initial flux and disease distribution, the disease model should be converted to irreversible form. Therefore, each of the disease models (11 iMAT-based and 11 INIT-based disease models) was converted to irreversible form by using *convertToIrreversible* function in COBRA Toolbox. Then, the flux estimation of the disease case was performed via utilizing an objective function that minimizes absolute flux difference between initial flux and disease distribution for twenty-two context-specific models (11 iMAT-models, 11 INIT-models).

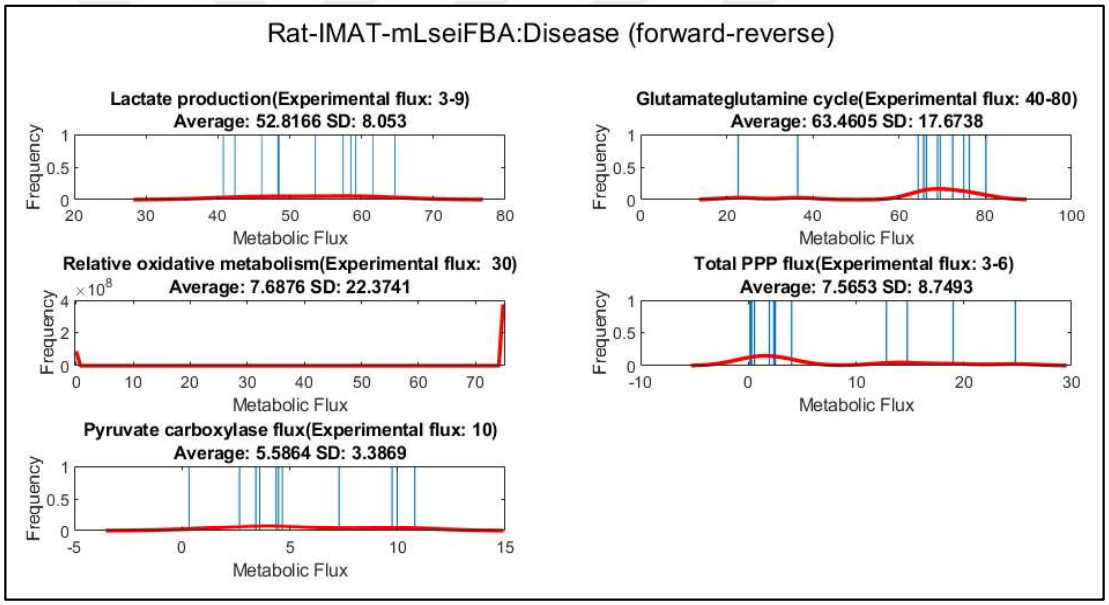

Figure 5.7: mLseiFBA analysis on iMAT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

Figure 5.7 shows that the predicted flux ratios for the reference reactions are consistent with previous studies, except for pyruvate carboxylase and relative oxidative metabolism. Furthermore, the standard deviation between groups was found to be high especially for glutamate/glutamine cycle and relative oxidative metabolism (Astrocyte TCA / Total TCA).

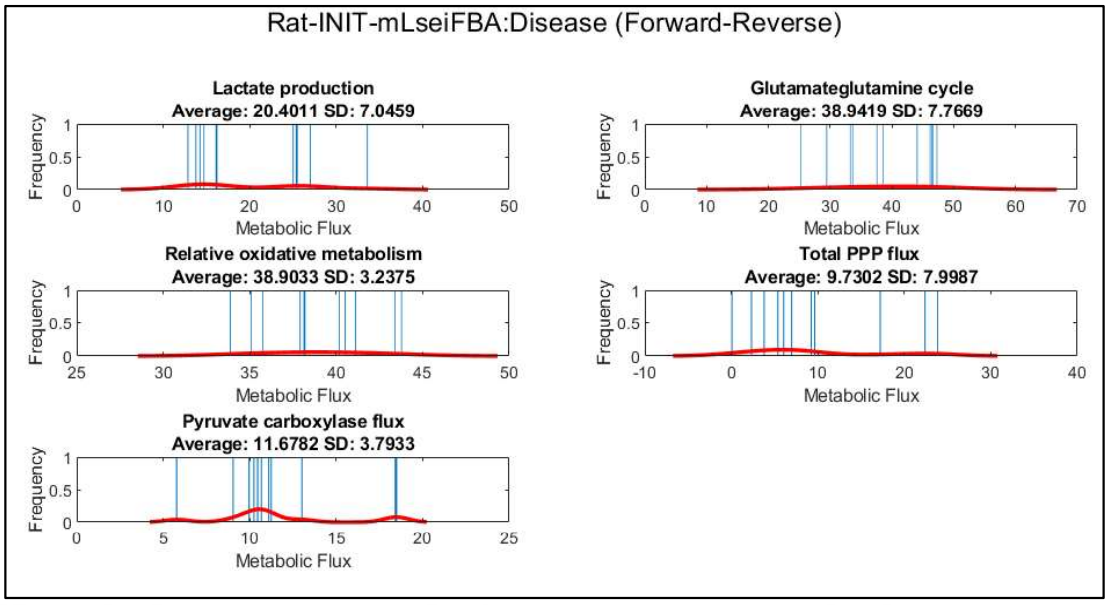

Figure 5.8: mLseiFBA analysis on INIT-based disease models. Each bar on the graph depicts the predicted flux ratio for a different comparison.

mLseiFBA analysis on INIT-based disease models showed that the predicted flux ratios for the reference reactions are consistent with previous studies (Figure 5.8).

6. DISCUSSION

6.1. Metabolically Similar Comparison Groups

Meta-analysis is a very useful technique to get a better understanding of gene expression in disease. Meta-analysis facilitates the identification of true biological signals in datasets with noisy gene expression. It can be used to investigate similarities across research by different groups or to reveal discrepancies between studies [242]. In this regard, hierarchical clustering analysis based on the gene scores of metabolic genes was performed in this thesis study (Chapter 3) to compare datasets in terms of metabolic gene expression pattern similarity.

Hierarchical clustering analysis of 11 comparisons from six transcriptome datasets via metabolic genes showed that the samples were clustered under two groups (Figure 3.3). Substantia nigra samples were clustered in one group while the other group contained heterogeneous brain regions including striatum, dorsal striatum, hemibrain, and frontal cortex brain regions (Figure 3.3). Normally, the same/similar brain regions are expected to be under the same group. However, the Substantia nigra comparisons clustered under the same group were generated in the same laboratory. Therefore, this result may originate from the laboratory effect. The dataset including SN samples is a time course dataset and corresponds to the rat neurotoxin model. On the other hand, two striatum (ST) comparisons were clustered in the same group, as expected, although they were generated by different datasets. Both of these datasets also correspond to neurotoxin models. In this context, these two comparisons were expected to cluster in the same group.

Enrichment analysis of metabolic genes that were used in hierarchical clustering analysis revealed that these genes were involved in metabolic processes including ATP metabolic process, cellular lipid metabolic process, cellular response to stress, generation of precursor metabolites and energy (Figure 3.5).

6.2. Validation of the Newly Reconstructed iBrain766-Rn Model

In this thesis study, a brain-specific genome-scale metabolic model was reconstructed for the rat (iBrain766-Rn) for the first time in the literature. iBrain766-Rn rat metabolic model consists of 994 reactions, 882 metabolites, and 766 genes. 778 of 994 reactions in the model are associated with genes. The validation of the reconstructed model was performed by comparing the predictions with the physiological data. Since the flux prediction power of the model is a measure of how accurately it reflects the relevant physiological state, the iBrain766-Rn model firstly was subjected to FBA analysis by using experimental constraints (Table 4.4) for the resting healthy state. Flux balance analysis (FBA) and enzyme usage minimization analysis on the iBrain766-Rn model showed consistent predictions with previous observations in the literature [9, 136]. Model validity was proven particularly by using five reference reactions. The estimated fluxes via five reactions with respect to the glucose uptake rate were found to be consistent with human experimental fluxes (Table 4.5). Human, mouse and rat are mammalian organisms, and the majority of metabolic processes are shared by these organisms. Consequently, comparable findings are anticipated in comparison with human brain metabolism. Considering Table 4.5, there was also a good agreement between the predicted rat fluxes and predicted human and mouse fluxes. In this regard, FBA analysis results revealed that the newly reconstructed iBrain766-Rn metabolic model can correctly simulate healthy human brain metabolism at resting state.

6.3. Comparison of iMAT-like Family Algorithms

A high percentage of gene-associated reactions in the reconstructed brain-specific rat metabolic model provides an opportunity to analyze the brain metabolism by combining the omics data, including transcriptome data, with the model. iMAT and INIT algorithms, both belonging to iMAT-like family algorithms, were used to integrate the rat-derived PD transcriptome datasets into the metabolic model in this study. iMAT-like family algorithms have an advantage since no biological objective function is required [162]. They try to keep reactions associated with highly expressed genes in the model while removing the reactions related to lowly expressed genes. The

PD transcriptome datasets were used as an additional constraint in iMAT and INIT to generate context-specific models.

Since five reactions (lactate release, Glu-Gln cycle, PPP, astrocytic oxidative phosphorylation and pyruvate carboxylase reactions) are essential stages in brain metabolism [9], these reactions were used as marker reactions to evaluate the flux predictions at different conditions in this study.

6.3.1. Healthy Case Flux Distribution

After obtaining the context-specific models, these models were subjected to FBA analysis. Specifically, predicted fluxes through five reactions relative to glucose uptake rate were consistent with literature estimates: fluxes through lactate release, Glu-Gln cycle (Glutamate-Glutamine cycle), total pentose phosphate pathway (PPP), astrocytic oxidative phosphorylation and pyruvate carboxylase, which are reported to be around 3-6%, 40-80%, 3-6%, 30% and 10% respectively in literature.

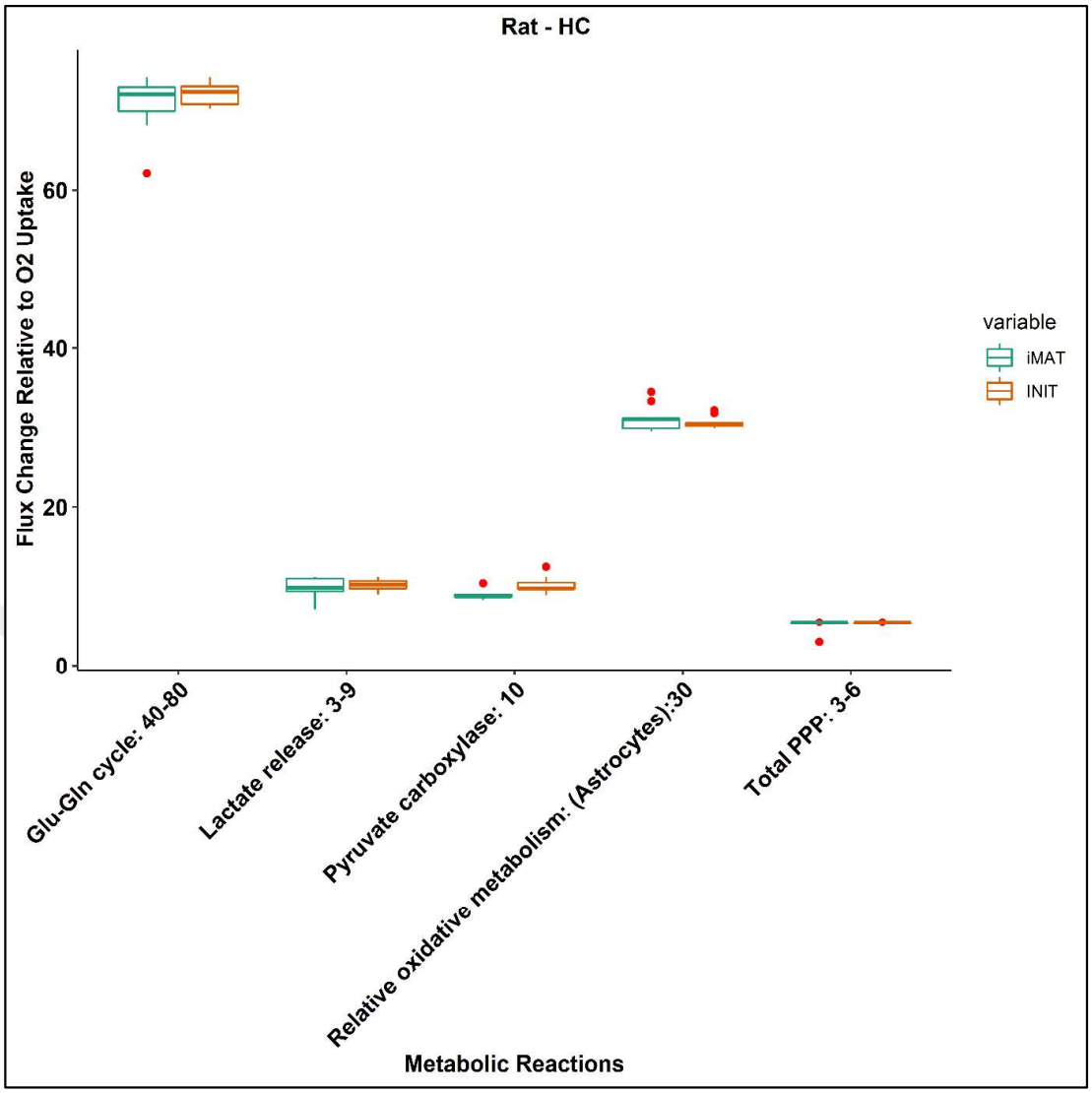

Figure 6.1: Comparison of healthy-state flux distributions between iMAT and INIT algorithms. X axis refers to the marker reactions and their experimental flux values while y axis indicates the predicted fluxes by the context-specific models. Red points indicate the outliers.

Figure 6.1 shows that the predicted fluxes for five reference reactions were not the same for all context-specific models. The predicted flux ratios were found to differ based on transcriptome data integration algorithms (iMAT and INIT) and constraint-based analysis methods (MOMA, LseiFBA). The flux prediction ratios for eleven iMAT based healthy models show some variations between 11 comparisons for some reference reactions including Glu-Gln cycle (Glutamate-glutamine cycle), lactate release and relative oxidative metabolism. These variations could be associated with the transcriptome data used since transcriptome data used corresponds to different brain regions (striatum, substantia nigra, hemibrain, dorsal striatum) and also different

disease models (neurotoxin-based or genetic-based). In this regard, variations between different comparisons were expected, and these variations were within an acceptable range. Predicted flux ratios for five reference reactions were also obtained by INIT-based healthy models similar to iMAT-based models. However, there were variations between iMAT and INIT-based models in terms of flux predictions for some reactions. Although iMAT and INIT share the basic approach, they differ in terms of how the data are handled. iMAT algorithm uses the gene expression data and upper and lower threshold values to discretize the genes into three groups as highly expressed, moderately expressed and lowly expressed. On the other hand, INIT algorithm utilizes the gene expression data to calculate reaction weights to be used in the objective function. In conclusion, the differences between context-specific models in terms of flux prediction can be attributed to the used methods to generate the context-specific model (iMAT, INIT), transcriptome data used, as well as the brain region studied. Nevertheless, the observed differences were within acceptable limits for most of the reactions by both methods.

6.3.2. Parkinson's Disease Case Flux Distribution

Three methods, MOMA, LseiFBA, and mLseiFBA with various constraints were utilized to make flux predictions in PD. The objective of MOMA was to minimize the flux difference between PD and Healthy control flux distributions (Equation (2.8)). On the other hand, the objective of LseiFBA was to minimize the difference between PD fluxes and an inital flux vector predicted for disease case from transcriptome data, given as $v_{ref,D}$. (Equation (2.11)).

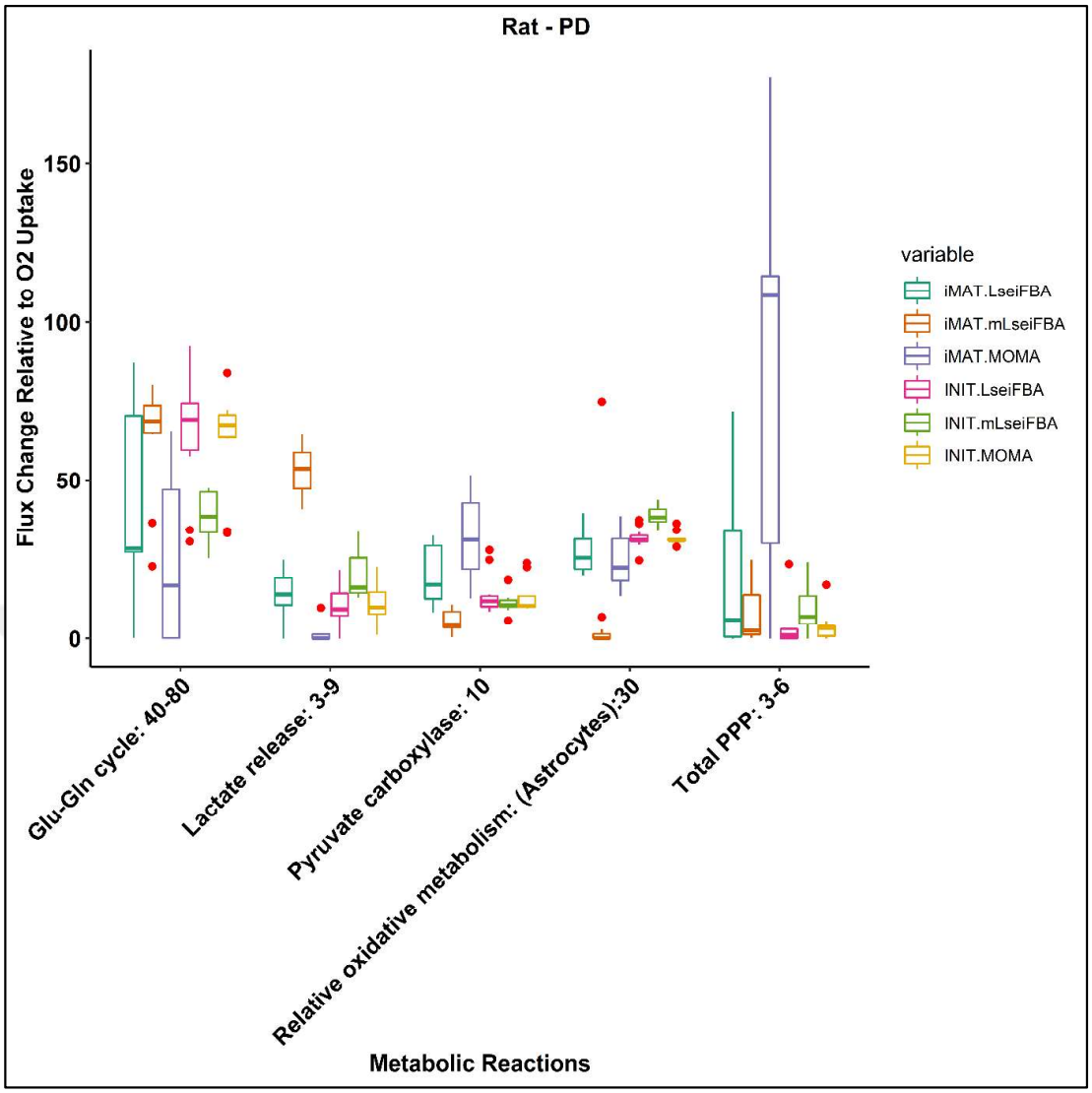

Figure 6.2: Flux distribution predicted for Parkinson's disease by different algorithms. Boxplots compare flux predictions by LseiFBA, mLseiFBA and MOMA approaches. Red dots refer to the outliers.

Figure 6.2 shows the comparison of constraint-based analysis methods (MOMA, LseiFBA, and mLseiFBA) for eleven iMAT and eleven INIT-based disease models. Metabolic alterations in Parkinson's disease including an increase in lactate production [75] and total PPP [74], a reduction in the TCA cycle, the oxygen consumption, glucose consumption, glutamine-glutamate cycle [240, 241], the malate-aspartate shuttle, and dopamine synthesis have been reported [74]. Once Figure 6.2 is evaluated in line with the literature findings, the results between CBM methods are seen to be substantially different. Figure 6.2 shows that mLseiFBA performed better in flux prediction ratios for most of reference reactions such as an increase in lactate release, which is consistent with impairments in efficient carbon metabolism and energy

production as frequently observed in Parkinson's disease through dysfunctional mitochondrial function [4, 59, 72, 74]. Moreover, compared to other models, INIT-based mLseiFBA was found to give better flux prediction for most reference reactions. Different patterns of the predicted fluxes for the total pentose phosphate pathway were detected in INIT and iMAT based models. Pentose phosphate reaction contents of the models were examined in order to determine whether this is a certain reaction in a model type or not. However, no difference was detected between the iMAT-based and INIT-based models in terms of pentose phosphate reaction content. The different patterns of the predicted total PPP fluxes between the iMAT and INIT models may be associated with the constraint-based method which was used. MOMA on the models obtained using iMAT predicted extreme flux increase for total pentose phosphate pathway (total PPP) compared to the other CBM methods. Although the literature findings support the increase of total PPP in Parkinson's diseases state [74], such a large increase is not expected. Compared to the iMAT-models, the INIT disease models have been found to have lower variability between samples/groups in flux predictions. Figure 6.2 shows that LseiFBA on the models obtained using iMAT performed better flux predictions in Parkinson's disease state compared to LseiFBA on INIT-based models. Difference in flux predictions between algorithms indicates that the model reconstruction approach has importance on the flux prediction.

6.4. Structural Comparison of GEMs

iMAT [170] and INIT are two algorithms which were utilized in this thesis study to integrate transcriptome data in order to obtain condition-specific models. These algorithms have different assumptions [162] that affect the final model content and model flux prediction capability. Figure 6.3 shows that the number of active reactions significantly differs between iMAT and INIT algorithms, as well as between healthy control case and Parkinson's disease case.

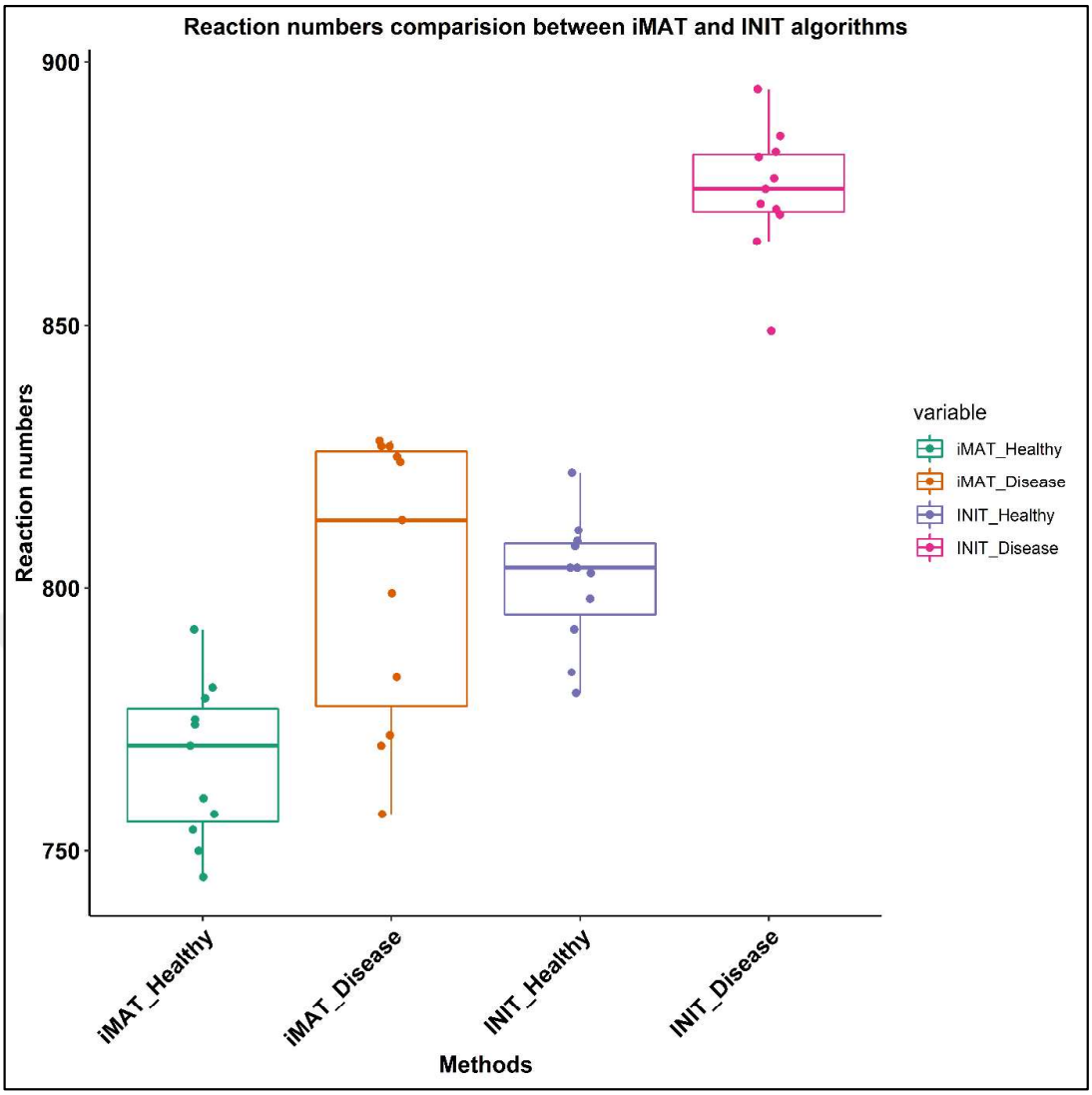

Figure 6.3: Comparison of reaction numbers between iMAT-based and INIT-based models. Each point represents one of 11 comparisons.

Several distance metrics can be used to represent variability across metabolic models at the genome-scale [243]. Jaccard index [244], hamming distance [245], and Pearson's Correlation coefficient [246] have been used to describe the model heterogeneity between the GEMs. Jaccard index was utilized in this thesis study to investigate the heterogeneity of models in terms of reaction content and subsystem coverage. Jaccard index analysis is based on dividing the intersection size of two groups with the size of the union of these groups. *CompareMultipleModels* in RAVEN Toolbox [247] was used to encode reaction content of 44 context specific models (22 iMAT, 22 INIT models) in binary vectors (one represents the existence of reactions, zero represents the lack of reactions) for Jaccard distance comparison. Models with a

small number of common reactions will have a large Jaccard index, while models with a large number of common reactions will have a small jaccard index.

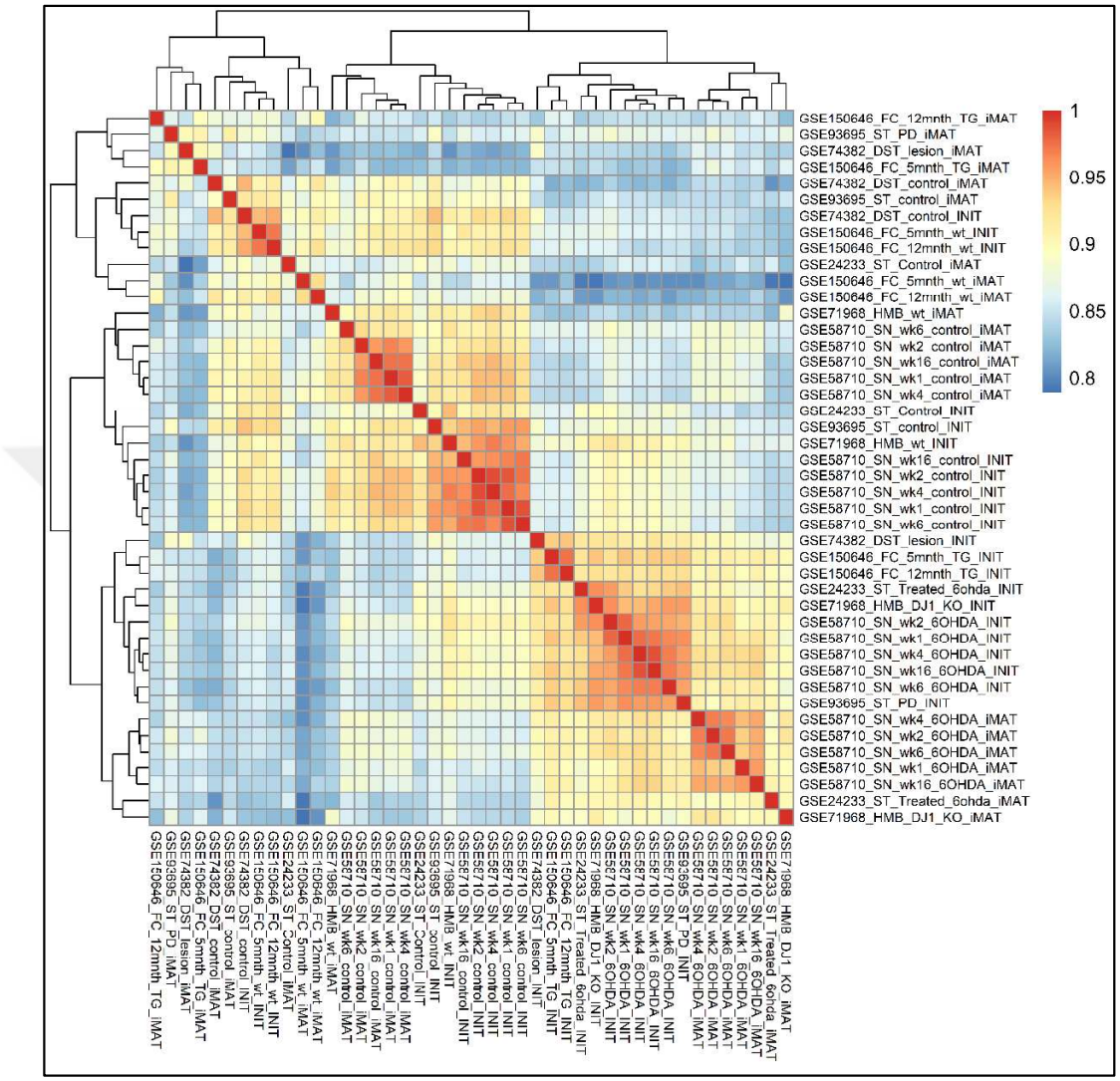

Figure 6.4: Reaction similarity Jaccard distance among models.

Model comparison result obtained by *CompareMultipleModels* function is a structure that includes StructComp field enabling to map the binary reaction vectors of models in reduced dimensions using tSNE (t-distributed Stochastic Neighbor Embedding).

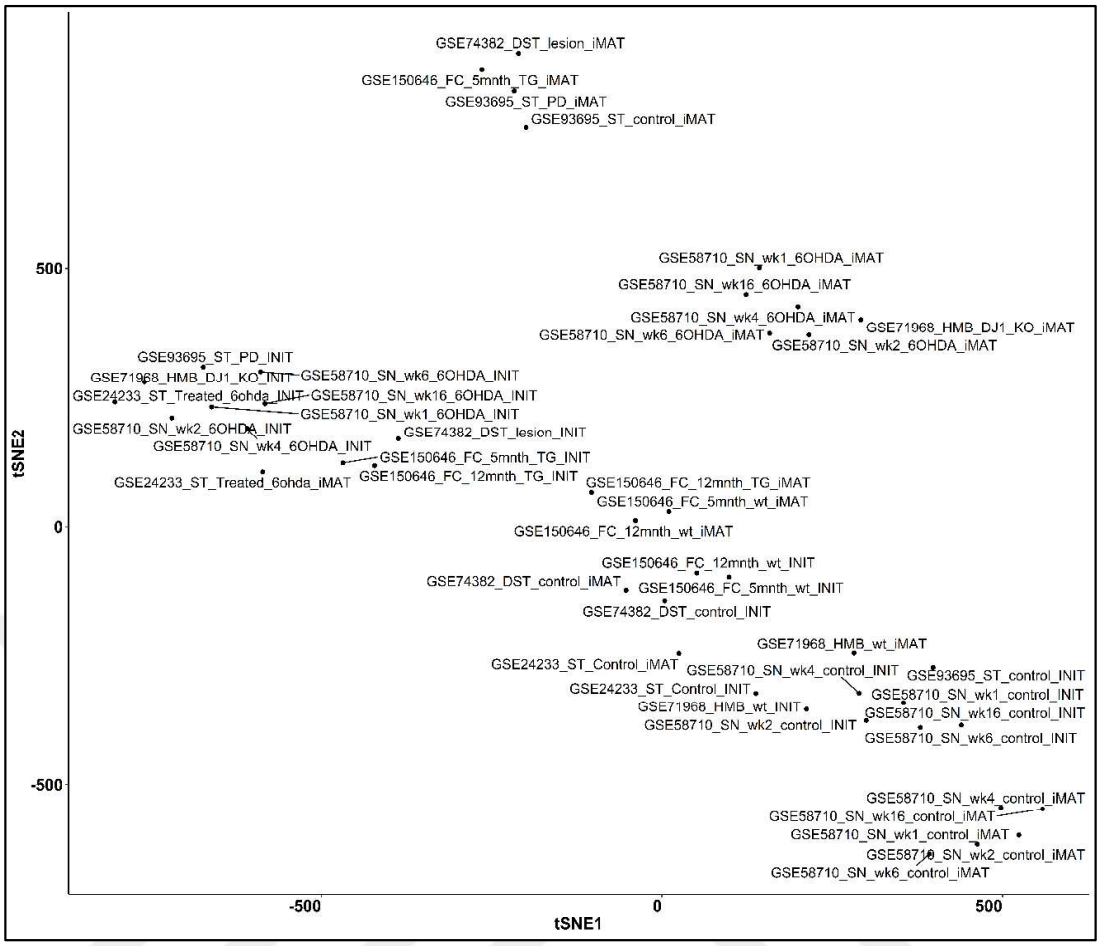

Figure 6.5: tSNE analysis of model reaction contents.

Figure 6.4 and Figure 6.5 show that INIT-disease models were clustered under one group and differed from iMAT disease models. Moreover, iMAT disease models clustered in two groups. However, INIT-healthy models were clustered with iMAT-healthy models.

The Jaccard index heatmap and tSNE projection show how similar or dissimilar models are, but not why or what they signify biologically. Therefore, reaction content in each subsystem was determined to identify which aspects of metabolism differ among GEMs. The percent subsystem coverage difference between models was calculated from the mean coverage. The difference in subsystem coverage between models was visualized by using *clustergram* function in RAVEN Toolbox. Only subsystems having at least 25 % difference in one or more GEMs were visualized to simplify the result. There were 28 subsystems having at least 25 % difference in one or more GEMs (Figure 6.6). The colour bar in Figure 6.6 shows the percent difference in subsystem coverage between the models. Subsystem coverage analysis of models

showed that iMAT and INIT models were substantially different in terms of Citrate/PEP Shuttle and tryptophan metabolism.

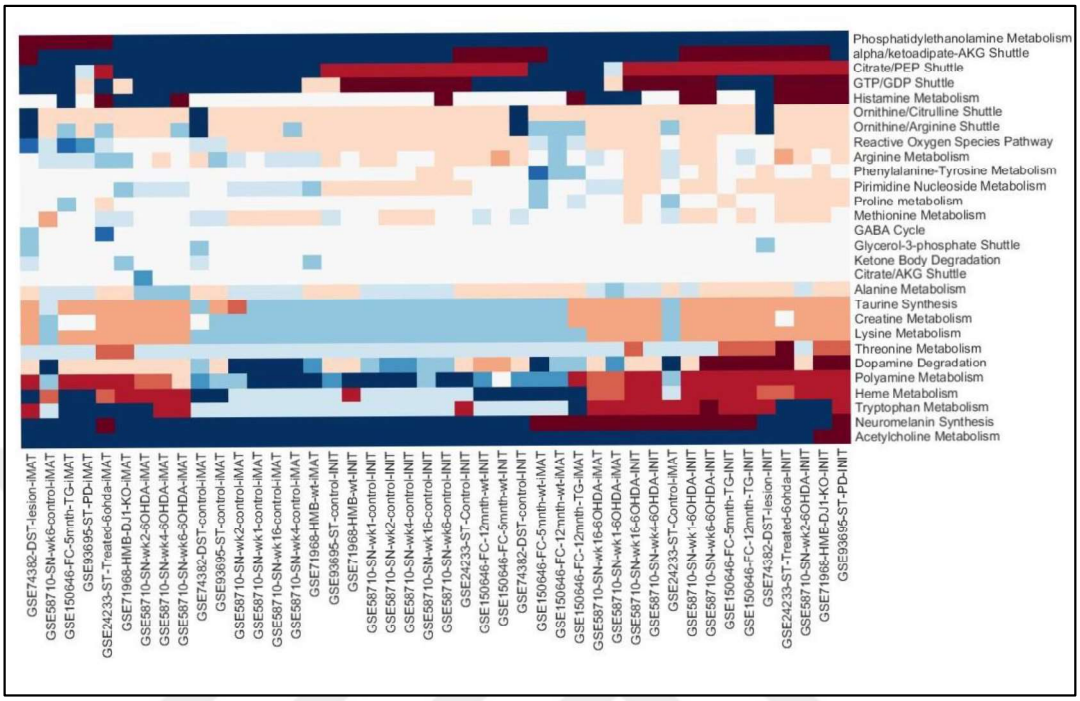

Figure 6.6: Subsystem coverage of models.

6.5. Comparison of Experimental PD Models

Hierarchical clustering was performed on flux fold changes to identify closely related groups. First, flux fold changes of reactions were calculated via dividing the predicted fluxes of Parkinson's disease case with the predicted fluxes of healthy case for each model (22 iMAT models, 22 INIT models). Then, calculated flux fold changes corresponding to iMAT-based and INIT-based models were combined. A data frame consisting of 66 columns (flux fold change values of models) and 994 rows (reactions) were obtained. In this study, 500 reactions comprising the top 10% of highly variable reactions identified by PCA from flux fold change were utilized in the hierarchical clustering analysis. Hierarchical clustering analysis of PD models based on flux fold changes revealed (Figure 6.7) that there were two main clusters, and brain regions did not cluster homogenously. Use of gene fold change alone had previously failed to capture closely related groups when similar or close brain regions were considered. Likewise, most of the groups found from all approaches did not provide consistent

patterns as previously observed in gene-score based clustering. Most of INIT-based and iMAT-based disease models were separated from each other (Figure 6.7). In addition, most of the 500 reactions were found to be related to fatty acid synthesis, TCA cycle metabolism, lipid, reactive oxygen species pathway, pentose phosphate pathway and purine nucleoside metabolism (Figure 6.8). That is, 500 reactions comprising the top 10% of highly variable reactions identified by PCA from flux fold change were found to involve pathways/subsystems mainly implicated in PD.

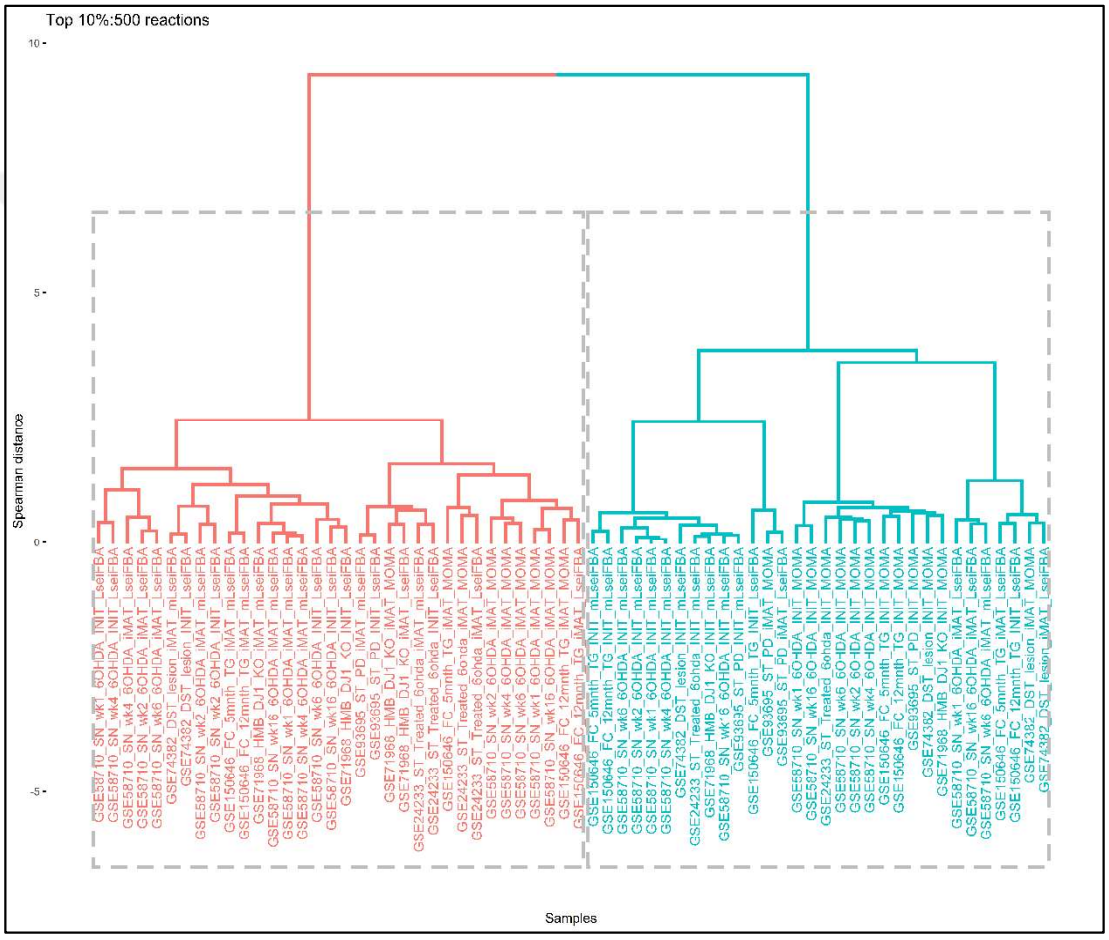

Figure 6.7: Hierarchical clustering analysis of PD models based on flux fold changes.

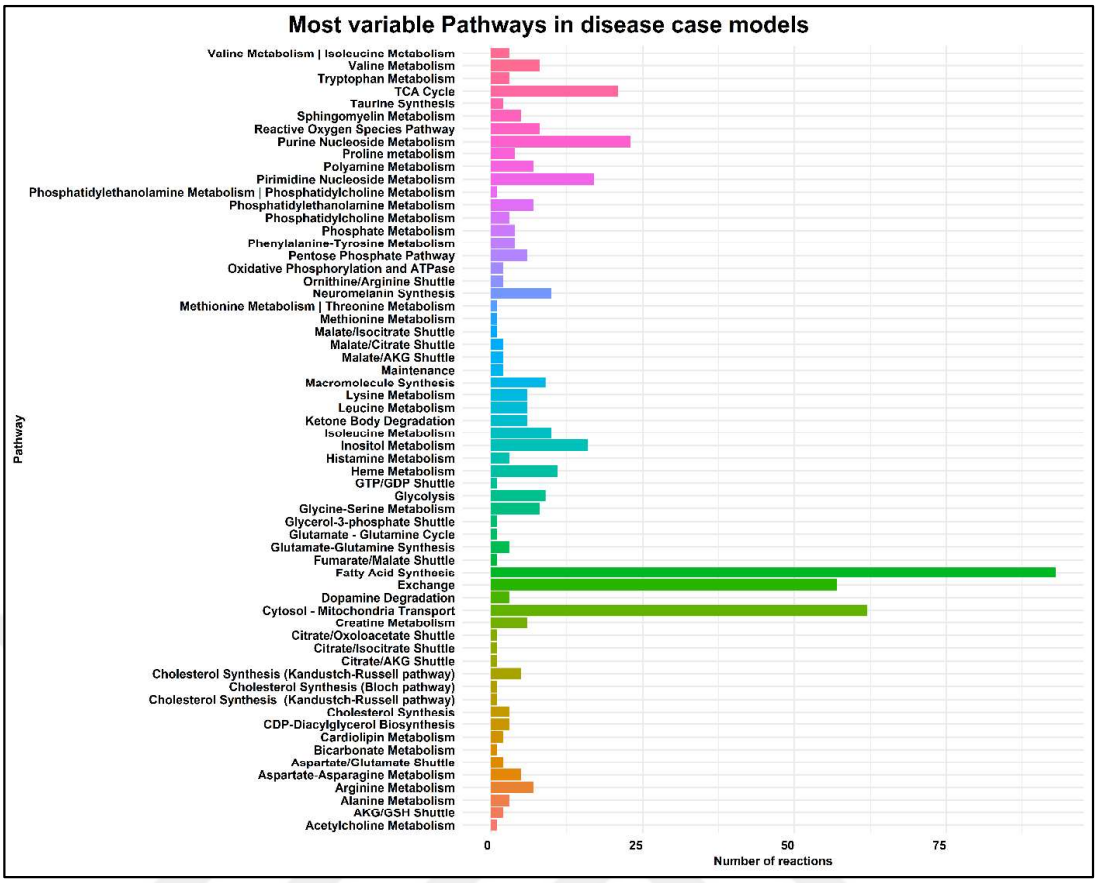

Figure 6.8: Most variable pathways based on flux fold changes.

7. CONCLUSION

GEMs are computational metabolic models that describe the complete metabolic processes and pathways of a cell, tissue, organ or organism, at genome level. In this study, a rat brain specific genome scale metabolic model, iBrain766-Rn was reconstructed for the first time in literature based on existing brain specific human and mouse models. FBA analysis on the newly reconstructed rat brain-specific metabolic model revealed that the model could correctly simulate rat brain metabolism at healthy resting state based on the flux predictions.

Eleven comparisons from six rat transcriptome datasets corresponding to Parkinson's disease models were later integrated with iBrain766-Rn by using iMAT and INIT algorithms to generate context-specific metabolic models. Although iMAT and INIT methods have similarities, they vary significantly in terms of how the data are handled. This difference in data processing between the algorithms leads to differences in the included reactions by the two algorithms, which significantly affects the flux predictions.

In addition to the effect of the employed transcriptome data integration algorithm, it was observed that differences between flux predictions is due to the used constraint-based methods. Four constraint-based methods, including FBA, MOMA, LseiFBA, and mLseiFBA were utilized to predict fluxes at different physiological states (healthy state, Parkinson's disease). The results between the CBM methods were found to be substantially different. However, different methods of constraint-based analysis on the context-specific models showed that the majority of predicted fluxes are acceptable since they are within the range of reported experimental fluxes. mLseiFBA, a CBM method, was found to give better flux prediction ratios for most of the reference reactions including an increase in lactate release and total PPP, a decrease in glutamate/glutamine cycle, which is consistent with impairments in efficient carbon metabolism and energy production as frequently observed in Parkinson's disease through dysfunctional mitochondrial function.

In conclusion, the newly reconstructed rat brain-specific GEM, iBrain766-Rn, provides an opportunity to investigate metabolic alterations in the Parkinson's disease via transcriptome data integration.

8. RECOMMENDATIONS

iBrain766-Rn model, which was reconstructed in the context of this study, was used to simulate rat brain metabolism for both resting and Parkinson's disease states. The model can also be used to study metabolic alterations in other neurodegenerative diseases.

Since iBrain766-Rn model does not include some metabolic pathways including vitamin, drug metabolism etc, these reactions can be included to the model and improved by manual curation.

Since LseiFBA and MOMA constraint-based methods are a special case of quadratic programming, they perform the enzyme usage minimization themselves without requiring the application of the enzyme usage minimization as an additional step. Unlike the LseiFBA and MOMA CBM methods, mLseiFBA requires applying the enzyme usage minimization as an additional step due to being based on a linear approach. Different assumptions employed by various transcriptome data integration algorithms affect the model content and predictive capability. Unlike other context-specific reconstruction methods including iMAT and INIT, a method such as tINIT, which ensures not only an active biomass reaction but also considers additional processes (metabolic tasks) that need to be active in the cell, may be utilized to deal with these issues. Protection of metabolic tasks between models would increase the consensus of reaction content between models and reduce the flux prediction variation between models.

Although brain region-specific constraints were not included in this study, they may be useful in enhancing the predictive ability of the models. In addition, the use of multi-omics data may be better in terms of improving the predictive ability of the models.

REFERENCES

[1] Harris J. J., Jolivet R., Attwell D., (2012), "Synaptic Energy Use and Supply". Neuron 75:762–777.

[2] Hyder F., Rothman D. L., Bennett M. R., (2013), "Cortical energy demands of signaling and nonsignaling components in brain are conserved across mammalian species and activity levels". Proc. Natl. Acad. Sci. U. S. A. 110:3549–3554.

[3] Camandola S., Mattson M. P., (2017), "Brain metabolism in health, aging, and neurodegeneration", EMBO J, 36(11), 1474–1492.

[4] Anandhan A., Jacome M. S., Lei S., Hernandez-Franco P., Pappa A., Panayiotidis M. I., Powers R., Franco R., (2017), "Metabolic Dysfunction in Parkinson's Disease: Bioenergetics, Redox Homeostasis and Central Carbon Metabolism". Brain Res. Bull. 133:12–30.

[5] Nichols E., Szoeke C. E. I., Vollset S. E., Abbasi N., Abd-Allah F., Abdela J., Aichour M. T. E., Akinyemi R. O., Alahdab F., Asgedom S. W., Awasthi A., Barker-Collo S. L., Baune B. T., Béjot Y., Belachew A. B., Bennett D. A., Biadgo B., Bijani A., Bin Sayeed M. S., Brayne C., Carpenter D. O., Carvalho F., Catalá-López F., Cerin E., Choi J. Y. J., Dang A. K., Degefa M. G., Djalalinia S., Dubey M., Duken E. E., Edvardsson D., Endres M., Eskandarieh S., Faro A., Farzadfar F., Fereshtehnejad S. M., Fernandes E., Filip I., Fischer F., Gebre A. K., Geremew D., Ghasemi-Kasman M., Gnedovskaya E. V., Gupta R., Hachinski V., Hagos T. B., Hamidi S., Hankey G. J., Haro J. M., Hay S. I., Irvani S. S. N., Jha R. P., Jonas J. B., Kalani R., Karch A., Kasaeian A., Khader Y. S., Khalil I. A., Khan E. A., Khanna T., Khoja T. A. M., Khubchandani J., Kisa A., Kissimova-Skarbek K., Kivimäki M., Koyanagi A., Krohn K. J., Logroscino G., Lorkowski S., Majdan M., Malekzadeh R., März W., Massano J., Mengistu G., Meretoja A., Mohammadi M., Mohammadi-Khanaposhtani M., Mokdad A. H., Mondello S., Moradi G., Nagel G., Naghavi M., Naik G., Nguyen L. H., Nguyen T. H., Nirayo Y. L., Nixon M. R., Ofori-Asenso R., Ogbo F. A., Olagunju A. T., Owolabi M. O., Panda-Jonas S., Passos V. M. d. A., Pereira D. M., Pinilla-Monsalve G. D., Piradov M. A., Pond C. D., Poustchi H., Qorbani M., Radfar A., Reiner R. C., Robinson S. R., Roshandel G., Rostami A., Russ T. C., Sachdev P. S., Safari H., Safiri S., Sahathevan R., Salimi Y., Satpathy M., Sawhney M., Saylan M., Sepanlou S. G., Shafieesabet A., Shaikh M. A., Sahraian M. A., Shigematsu M., Shiri R., Shiue I., Silva J. P., Smith M., Sobhani S., Stein D. J., Tabarés-Seisdedos R., Tovani-Palone M. R., Tran B. X., Tran T. T., Tsegay A. T., Ullah I., Venketasubramanian N., Vlassov V., Wang Y. P., Weiss J., Westerman R., Wijeratne T., Wyper G. M. A., Yano Y., Yimer E. M., Yonemoto N., Yousefifard M., Zaidi Z., Zare Z., Vos T., Feigin V. L., Murray C. J. L., (2019), "Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016", Lancet Neurol, 18(1), 88–106.

[6] Blandini F., Armentero M. T., (2012), "Animal models of Parkinson's disease". FEBS J. 279:1156–1166.

[7] Bordbar A., Palsson B. O., (2012), "Using the reconstructed genome-scale human metabolic network to study physiology and pathology". In: Journal of Internal Medicine. John Wiley & Sons, Ltd, pp 131–141.

[8] Khodaee S., Asgari Y., Totonchi M., Karimi-Jafari M. H., (2020), "iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model", Sci Rep, 10(1), 1–13.

[9] Sertbaş M., Ülgen K., Çakir T., (2014), "Systematic analysis of transcription-level effects of neurodegenerative diseases on human brain metabolism by a newly reconstructed brain-specific metabolic network", FEBS Open Bio, 4, 542–553.

[10] Kaynar A., (2019), "Integrative Analysis of Transcriptome Data and Genome-Scale Metabolic Networks To Identify Candidate Drug Targets and Drugs for Parkinson'S Disease".

[11] Abdik E., Çakır T., (2021), "Systematic investigation of mouse models of Parkinson's disease by transcriptome mapping on a brain-specific genome-scale metabolic network". Mol Omi. doi: 10.1039/d0mo00135j.

[12] Hertz L., Dienel G. A., (2002), "Energy metabolism in the brain". Int. Rev. Neurobiol. 51.

[13] Rolfe D. F. S., Brown G. C., (1997), "Cellular energy utilization and molecular origin of standard metabolic rate in mammals", Physiol Rev, 77(3), 731–758.

[14] Brekke E. M. F., Walls A. B., Schousboe A., Waagepetersen H. S., Sonnewald U., (2012), "Quantitative importance of the pentose phosphate pathway determined by incorporation of 13 C from 2-13 C-and 3-13 Cglucose into TCA cycle intermediates and neurotransmitter amino acids in functionally intact neurons", J Cereb Blood Flow Metab, 32(9), 1788–1799.

[15] Almeida A., Moncada S., Bolaños J. P., (2004), "Nitric oxide switches on glycolysis through the AMP protein kinase and 6-phosphofructo-2-kinase pathway", Nat Cell Biol, 6(1), 45–51.

[16] Almeida A., Almeida J., Bolaños J. P., Moncada S., (2001), "Different responses of astrocytes and neurons to nitric oxide: The role of glycolytically generated ATP in astrocyte protection", Proc Natl Acad Sci U S A, 98(26), 15294–15299.

[17] Herrero-Mendez A., Almeida A., Fernández E., Maestre C., Moncada S., Bolaños J. P., (2009), "The bioenergetic and antioxidant status of neurons is controlled by continuous degradation of a key glycolytic enzyme by APC/C-Cdh1", Nat Cell Biol, 11(6), 747–752.

[18] Bak L. K., Schousboe A., Waagepetersen H. S., (2006), "The glutamate/GABA-glutamine cycle: Aspects of transport, neurotransmitter homeostasis and ammonia transfer". J. Neurochem. 98:641–653.

[19] Waagepetersen H. S., Sonnewald U., Schousboe A., (2007), "1 Glutamine, Glutamate, and GABA: Metabolic Aspects", Handb Neurochem Mol Neurobiol Amin Acids Pept Nerv Syst, 1–21.

[20] Bélanger M., Allaman I., Magistretti P. J., (2011), "Brain energy metabolism: Focus on Astrocyte-neuron metabolic cooperation". Cell Metab. 14:724–738.

[21] Shih A. Y., Johnson D. A., Wong G., Kraft A. D., Jiang L., Erb H., Johnson J. A., Murphy T. H., (2003), "Coordinate regulation of glutathione biosynthesis and release by Nrf2-expressing glia potently protects neurons from oxidative stress", J Neurosci, 23(8), 3394–3406.

[22] Bell K. F. S., Al-Mubarak B., Martel M. A., McKay S., Wheelan N., Hasel P., Márkus N. M., Baxter P., Deighton R. F., Serio A., Bilican B., Chowdhry S., Meakin P. J., Ashford M. L. J., Wyllie D. J. A., Scannevin R. H., Chandran S., Hayes J. D., Hardingham G. E., (2015), "Neuronal development is promoted by weakened intrinsic antioxidant defences due to epigenetic repression of Nrf2", Nat Commun, 6.

[23] Itoh Y., Esaki T., Shimoji K., Cook M., Law M. J., Kaufman E., Sokoloff L., (2003), "Dichloroacetate effects on glucose and lactate oxidation by neurons and astroglia in vitro and on glucose utilization by brain in vivo", Proc Natl Acad Sci U S A, 100(8), 4879–4884.

[24] Halim N. D., Mcfate T., Mohyeldin A., Okagaki P., Korotchkina L. G., Patel M. S., Jeoung N. H., Harris R. A., Schell M. J., Verma A., (2010), "Phosphorylation status of pyruvate dehydrogenase distinguishes metabolic phenotypes of cultured rat brain astrocytes and neurons", Glia, 58(10), 1168–1176.

[25] Gao H. M., Hong J. S., (2008), "Why neurodegenerative diseases are progressive: uncontrolled inflammation drives disease progression", Trends Immunol, 29(8), 357–365.

[26] Menken M., Munsat T. L., Toole J. F., (2000), "The Global Burden of Disease Study: Implications for neurology", Arch Neurol, 57(3), 418–420.

[27] Sheikh S., Safia, Haque E., Mir S. S., (2013), "Neurodegenerative Diseases: Multifactorial Conformational Diseases and Their Therapeutic Interventions", J Neurodegener Dis, 2013, 1–8.

[28] Dugger B. N., Dickson D. W., (2017), "Pathology of neurodegenerative diseases". Cold Spring Harb. Perspect. Biol. 9.

[29] Simrén J., Ashton N. J., Blennow K., Zetterberg H., (2020), "An update on fluid biomarkers for neurodegenerative diseases: recent success and challenges ahead". Curr. Opin. Neurobiol. 61:29–39.

[30] Ilieva H., Polymenidou M., Cleveland D. W., (2009), "Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond". J. Cell Biol. 187:761–772.

[31] Marambaud P., Dreses-Werringloer U., Vingtdeux V., (2009), "Calcium signaling

in neurodegeneration". Mol. Neurodegener. 4:1–15.

[32] Kiaei M., (2013), "New hopes and challenges for treatment of neurodegenerative disorders: Great opportunities for Young neuroscientists", Basic Clin Neurosci, 4(1), 3–4.

[33] Gan L., Cookson M. R., Petrucelli L., La Spada A. R., (2018), "Converging pathways in neurodegeneration, from genetics to mechanisms". Nat. Neurosci. 21:1300–1309.

[34] Kiaei M., (2013), "New hopes and challenges for treatment of neurodegenerative disorders: Great opportunities for Young neuroscientists", Basic Clin Neurosci, 4(1), 3–4.

[35] Zhou J., Gennatas E. D., Kramer J. H., Miller B. L., Seeley W. W., (2012), "Predicting Regional Neurodegeneration from the Healthy Brain Functional Connectome", Neuron, 73(6), 1216–1227.

[36] Song J., Kim J., (2016), "Degeneration of dopaminergic neurons due to metabolic alterations and Parkinson's disease". Front. Aging Neurosci. 8:65.

[37] Alavi A., Reivich M., Ferris S., Christman D., Fowler J., MacGregor R., Farkas T., Greenberg J., Dann R., Wolf A., (1982), "Regional cerebral glucose metabolism in aging and senile dementia as determined by 18F-deoxyglucose and positron emission tomography.", Exp Brain Res, Suppl 5, 187–195.

[38] Sherer T. B., Chowdhury S., Peabody K., Brooks D. W., (2012), "Overcoming obstacles in Parkinson's disease". Mov. Disord. 27:1606–1611.

[39] Van Den Eeden S. K., Tanner C. M., Bernstein A. L., Fross R. D., Leimpeter A., Bloch D. A., Nelson L. M., (2003), "Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity", Am J Epidemiol, 157(11), 1015–1022.

[40] Coelho M., Ferreira J. J., (2012), "Late-stage Parkinson disease". Nat. Rev. Neurol. 8:435–442.

[41] Schapira A. H. V., Bezard E., Brotchie J., Calon F., Collingridge G. L., Ferger B., Hengerer B., Hirsch E., Jenner P., Novère N. Le, Obeso J. A., Schwarzschild M. A., Spampinato U., Davidai G., (2006), "Novel pharmacological targets for the treatment of Parkinson's disease". Nat. Rev. Drug Discov. 5:845–854.

[42] Brooks D. J., (2007), "Imaging non-dopaminergic function in Parkinson's disease". Mol. Imaging Biol. 9:217–222.

[43] Kalia L. V., Lang A. E., (2015), "Parkinson's disease". Lancet 386:896–912.

[44] Klein C., Westenberger A., (2012), "Genetics of Parkinson's disease", Cold Spring Harb Perspect Med, 2(1), a008888.

[45] Pickrell A. M., Youle R. J., (2015), "The roles of PINK1, Parkin, and mitochondrial fidelity in parkinson's disease". Neuron 85:257–273.

[46] Schapira A. H. V., Tolosa E., (2010), "Molecular and clinical prodrome of Parkinson disease: implications for treatment", Nat Rev Neurol 2010 66, 6(6), 309–317.

[47] Schapira A. H., (2008), "Mitochondria in the aetiology and pathogenesis of Parkinson's disease". Lancet Neurol. 7:97–109.

[48] Schapira A. H. V., Cooper J. M., Dexter D., Clark J. B., Jenner P., Marsden C. D., (1990), "Mitochondrial Complex I Deficiency in Parkinson's Disease", J Neurochem, 54(3), 823–827.

[49] Abeliovich A., Schmitz Y., Fariñas I., Choi-Lundberg D., Ho W. H., Castillo P. E., Shinsky N., Garcia Verdugo J. M., Armanini M., Ryan A., Hynes M., Phillips H., Sulzer D., Rosenthal A., (2000), "Mice lacking α-synuclein display functional deficits in the nigrostriatal dopamine system", Neuron, 25(1), 239–252.

[50] Cookson M. R., (2010), "The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease". Nat. Rev. Neurosci. 11:791–797.

[51] Wallings R., Manzoni C., Bandopadhyay R., (2015), "Cellular processes associated with LRRK2 function and dysfunction". FEBS J. 282:2806–2826.

[52] Narendra D., Tanaka A., Suen D. F., Youle R. J., (2008), "Parkin is recruited selectively to impaired mitochondria and promotes their autophagy", J Cell Biol, 183(5), 795–803.

[53] Arena G., Valente E. M., (2017), "PINK1 in the limelight: multiple functions of an eclectic protein in human health and disease". J. Pathol. 241:251–263.

[54] Seirafi M., Kozlov G., Gehring K., (2015), "Parkin structure and function". FEBS J. 282:2076–2088.

[55] Gegg M. E., Cooper J. M., Schapira A. H. V., Taanman J. W., (2009), "Silencing of PINK1 expression affects mitochondrial DNA and oxidative phosphorylation in DOPAMINERGIC cells", PLoS One, 4(3), e4756.

[56] Grünewald A., Gegg M. E., Taanman J. W., King R. H., Kock N., Klein C., Schapira A. H. V., (2009), "Differential effects of PINK1 nonsense and missense mutations on mitochondrial function and morphology", Exp Neurol, 219(1), 266–273.

[57] Wilhelmus M. M. M., Nijland P. G., Drukarch B., De Vries H. E., Van Horssen J., (2012), "Involvement and interplay of Parkin, PINK1, and DJ1 in neurodegenerative and neuroinflammatory disorders". Free Radic. Biol. Med. 53:983–992.

[58] Hoyer S., (1982), "The abnormally aged brain. Its blood flow and oxidative metabolism. A review - Part II", Arch Gerontol Geriatr, 1(3), 195–207.

[59] Dunn L., Allen G. F. G., Mamais A., Ling H., Li A., Duberley K. E., Hargreaves I. P., Pope S., Holton J. L., Lees A., Heales S. J., Bandopadhyay R., (2014), "Dysregulation of glucose metabolism is an early event in sporadic Parkinson's

disease", Neurobiol Aging, 35(5), 1111–1115.

[60] Grivennikova V. G., Vinogradov A. D., (2006), "Generation of superoxide by the mitochondrial Complex I". Biochim. Biophys. Acta - Bioenerg. 1757:553–561.

[61] Chen Q., Vazquez E. J., Moghaddas S., Hoppel C. L., Lesnefsky E. J., (2003), "Production of reactive oxygen species by mitochondria: Central role of complex III", J Biol Chem, 278(38), 36027–36031.

[62] Martinez T. N., Greenamyre J. T., (2012), "Toxin models of mitochondrial dysfunction in Parkinson's disease". Antioxidants Redox Signal. 16:920–934.

[63] Barkats M., Horellou P., Colin P., Millecamps S., Faucon-Biguet N., Mallet J., (2006), "1-Methyl-4-phenylpyridinium neurotoxicity is attenuated by adenoviral gene transfer of human Cu/Zn superoxide dismutase", J Neurosci Res, 83(2), 233–242.

[64] Kaul S., Kanthasamy A., Kitazawa M., Anantharam V., Kanthasamy A. G., (2003), "Caspase-3 dependent proteolytic activation of protein kinase Cδ mediates and regulates 1-methyl-4-phenylpyridinium (MPP+)-induced apoptotic cell death in dopaminergic cells: Relevance to oxidative stress in dopaminergic degeneration", Eur J Neurosci, 18(6), 1387–1401.

[65] Rodriguez-Rocha H., Garcia-Garcia A., Pickett C., Li S., Jones J., Chen H., Webb B., Choi J., Zhou Y., Zimmerman M. C., Franco R., (2013), "Compartmentalized oxidative stress in dopaminergic cell death induced by pesticides and complex i inhibitors: Distinct roles of superoxide anion and superoxide dismutases", Free Radic Biol Med, 61, 370–383.

[66] Bates T. E., Heales S. J. R., Davies S. E. C., Boakye P., Clark J. B., (1994), "Effects of 1-methyl-4-phenylpyridinium on isolated rat brain mitochondria: Evidence for a primary involvement of energy depletion", J Neurochem, 63(2), 640–648.

[67] Chinta S. J., Andersen J. K., (2005), "Dopaminergic neurons". Int. J. Biochem. Cell Biol. 37:942–946.

[68] Choi S. W., Gerencser A. A., Lee D. W., Rajagopalan S., Nicholls D. G., Andersen J. K., Brand M. D., (2011), "Intrinsic bioenergetic properties and stress sensitivity of dopaminergic synaptosomes", J Neurosci, 31(12), 4524–4534.

[69] Chaudhuri A. D., Kabaria S., Choi D. C., Mouradian M. M., Junn E., (2015), "MicroRNA-7 promotes glycolysis to protect against 1-methyl-4-phenylpyridinium-induced cell death", J Biol Chem, 290(19), 12425–12434.

[70] Meiser J., Weindl D., Hiller K., (2013), "Complexity of dopamine metabolism". Cell Commun. Signal. 11:34.

[71] Bisaglia M., Greggio E., Beltramini M., Bubacco L., (2013), "Dysfunction of dopamine homeostasis: Clues in the hunt for novel Parkinson's disease therapies". FASEB J. 27:2101–2110.

[72] Poewe W., Seppi K., Tanner C. M., Halliday G. M., Brundin P., Volkmann J., Schrag A. E., Lang A. E., (2017), "Parkinson disease", Nat Rev Dis Prim, 3, 1–21.

[73] Borghammer P., Chakravarty M., Jonsdottir K. Y., Sato N., Matsuda H., Ito K., Arahata Y., Kato T., Gjedde A., (2010), "Cortical hypometabolism and hypoperfusion in Parkinson's disease is extensive: Probably even at early disease stages". Brain Struct. Funct. 214:303–317.

[74] Supandi F., van Beek J. H. G. M., (2018), "Computational prediction of changes in brain metabolic fluxes during Parkinson's disease from mRNA expression", PLoS One, 13(9).

[75] Henchcliffe C., Shungu D. C., Mao X., Huang C., Nirenberg M. J., Jenkins B. G., Beal M. F., (2008), "Multinuclear magnetic resonance spectroscopy for in vivo assessment of mitochondrial dysfunction in Parkinson's disease". In: Annals of the New York Academy of Sciences. Blackwell Publishing Inc., pp 206–220.

[76] Cenci M. A., Whishaw I. Q., Schallert T., (2002), "Animal models of neurological deficits: How relevant is the rat?", Nat Rev Neurosci, 3(7), 574–579.

[77] Jagmag S. A., Tripathi N., Shukla S. D., Maiti S., Khurana S., (2016), "Evaluation of models of Parkinson's disease". Front. Neurosci. 9:503.

[78] Ungerstedt U., (1968), "6-hydroxy-dopamine induced degeneration of central monoamine neurons", Eur J Pharmacol, 5(1), 107–110.

[79] Thiele S. L., Warre R., Nash J. E., (2012), "Development of a unilaterally-lesioned 6-OHDA mouse model of Parkinson's disease", J Vis Exp, (60), e3234.

[80] Simola N., Morelli M., Carta A. R., (2007), "The 6-hydroxydopamine model of Parkinson's disease". Neurotox. Res. 11:151–167.

[81] Barrientos A., Moraes C. T., (1999), "Titrating the effects of mitochondrial complex I impairment in the cell physiology", J Biol Chem, 274(23), 16188–16197.

[82] Palmeira C. M., Moreno A. J., Madeira V. M. C., (1995), "Mitochondrial bioenergetics is affected by the herbicide paraquat", BBA - Bioenerg, 1229(2), 187–192.

[83] Domico L. M., Zeevalk G. D., Bernard L. P., Cooper K. R., (2006), "Acute neurotoxic effects of mancozeb and maneb in mesencephalic neuronal cultures are associated with mitochondrial dysfunction", Neurotoxicology, 27(5), 816–825.

[84] Cacabelos R., (2017), "Parkinson's disease: From pathogenesis to pharmacogenomics". Int. J. Mol. Sci. 18:551.

[85] Betarbet R., Sherer T. B., MacKenzie G., Garcia-Osuna M., Panov A. V., Greenamyre J. T., (2000), "Chronic systemic pesticide exposure reproduces features of Parkinson's disease", Nat Neurosci, 3(12), 1301–1306.

[86] Cicchetti F., Drouin-Ouellet J., Gross R. E., (2009), "Environmental toxins and Parkinson's disease: what have we learned from pesticide-induced animal models?". Trends Pharmacol. Sci. 30:475–483.

[87] Brooks A. I., Chadwick C. A., Gelbard H. A., Cory-Slechta D. A., Federoff H. J., (1999), "Paraquat elicited neurobehavioral syndrome caused by dopaminergic neuron loss", Brain Res, 823(1–2), 1–10.

[88] Wagner G. C., Seiden L. S., Schuster C. R., (1979), "Methamphetamine-induced changes in brain catecholamines in rats and guinea pigs", Drug Alcohol Depend, 4(5), 435–438.

[89] Smith G. A., Isacson O., Dunnett S. B., (2012), "The search for genetic mouse models of prodromal Parkinson's disease". Exp. Neurol. 237:267–273.

[90] Blesa J., Przedborski S., (2014), "Parkinson's disease: Animal models and dopaminergic cell vulnerability", Front Neuroanat, 8(DEC), 155.

[91] Kazlauskaite A., Muqit M. M. K., (2015), "PINK1 and Parkin - Mitochondrial interplay between phosphorylation and ubiquitylation in Parkinson's disease". FEBS J. 282:215–223.

[92] Poulopoulos M., Levy O. A., Alcalay R. N., (2012), "The neuropathology of genetic Parkinson's disease". Mov. Disord. 27:831–842.

[93] Goldberg M. S., Pisani A., Haburcak M., Vortherms T. A., Kitada T., Costa C., Tong Y., Martella G., Tscherter A., Martins A., Bernardi G., Roth B. L., Pothos E. N., Calabresi P., Shen J., (2005), "Nigrostriatal dopaminergic deficits and hypokinesia caused by inactivation of the familial parkinsonism-linked gene DJ-1", Neuron, 45(4), 489–496.

[94] Chuang H. Y., Hofree M., Ideker T., (2010), "A decade of systems biology". Annu. Rev. Cell Dev. Biol. 26:721–744.

[95] Morozova O., Marra M. A., (2008), "Applications of next-generation sequencing technologies in functional genomics". Genomics 92:255–264.

[96] Sorek R., Cossart P., (2009), "Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity".

[97] Martens L., Hermjakob H., Jones P., Adamsk M., Taylor C., States D., Gevaert K., Vandekerckhove J., Apweiler R., (2005), "PRIDE: The proteomics identifications database", Proteomics, 5(13), 3537–3545.

[98] Bundy J. G., Davey M. P., Viant M. R., (2009), "Environmental metabolomics: A critical review and future perspectives". Metabolomics 5:3–21.

[99] Winter G., Krömer J. O., (2013), "Fluxomics - connecting 'omics analysis and phenotypes", Environ Microbiol, 15(7), 1901–1916.

[100] Palsson B., (2015), Systems Biology. Cambridge university press.

[101] McCloskey D., Palsson B., Feist A. M., (2013), "Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli". Mol. Syst. Biol. 9.

[102] Qi Z., Voit E. O., (2013), "Systems biology provides new tools for addressing Parkinson's disease", Adv Park Dis, 02(03), 69–69.

[103] Wang Z., Gerstein M., Snyder M., (2009), "RNA-Seq: A revolutionary tool for transcriptomics". Nat. Rev. Genet. 10:57–63.

[104] Schena M., Shalon D., Davis R. W., Brown P. O., (1995), "Quantitative monitoring of gene expression patterns with a complementary DNA microarray", Science (80-), 270(5235), 467–470.

[105] Barrett T., Wilhite S. E., Ledoux P., Evangelista C., Kim I. F., Tomashevsky M., Marshall K. A., Phillippy K. H., Sherman P. M., Holko M., Yefanov A., Lee H., Zhang N., Robertson C. L., Serova N., Davis S., Soboleva A., (2013), "NCBI GEO: Archive for functional genomics data sets - Update", Nucleic Acids Res, 41(D1), D991–D995.

[106] Sarkans U., Parkinson H., Lara G. G., Oezcimen A., Sharma A., Abeygunawardena N., Contrino S., Holloway E., Rocca-Serra P., Mukherjee G., Shojatalab M., Kapushesky M., Sansone S. A., Farne A., Rayner T., Brazma A., (2005), "The ArrayExpress gene expression database: A software engineering and implementation perspective", Bioinformatics, 21(8), 1495–1501.

[107] Kelly J., Moyeed R., Carroll C., Albani D., Li X., (2019), "Gene expression meta-analysis of Parkinson's disease and its relationship with Alzheimer's disease", Mol Brain, 12(1), 1–10.

[108] Y Z., M J., FA M., RL D., (2005), "Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms", Am J Med Genet B Neuropsychiatr Genet, 137B(1), 5–16.

[109] Dyavar S. R., Potts L. F., Beck G., Dyavar Shetty B. L., Lawson B., Podany A. T., Fletcher C. V., Amara R. R., Papa S. M., (2020), "Transcriptomic approach predicts a major role for transforming growth factor beta type 1 pathway in L-Dopa-induced dyskinesia in parkinsonian rats", Genes, Brain Behav, 19(8).

[110] Shadrina M. I., Filatova E. V., Alieva A. K., Stavrovskaya A. V., Khudoerkov R. M., Limborska S. A., Illarioshkin S. N., Slominsky P. A., (2013), "Transcriptome profiling of 6-OHDA model of Parkinson's disease"2013(06), 28–35.

[111] Maarleveld T. R., Khandelwal R. A., Olivier B. G., Teusink B., Bruggeman F. J., (2013), "Basic concepts and principles of stoichiometric modeling of metabolic networks". Biotechnol. J. 8:997–1008.

[112] Kim T. Y., Sohn S. B., Kim Y. Bin, Kim W. J., Lee S. Y., (2012), "Recent advances in reconstruction and applications of genome-scale metabolic models". Curr. Opin. Biotechnol. 23:617–623.

[113] Vlassis N., Pacheco M. P., Sauter T., (2014), "Fast Reconstruction of Compact Context-Specific Metabolic Network Models", PLoS Comput Biol, 10(1), 1003424.

[114] Agren R., Mardinoglu A., Asplund A., Kampf C., Uhlen M., Nielsen J., (2014), "Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling", Mol Syst Biol, 10(3).

[115] Heinken A., Sahoo S., Fleming R. M. T., Thiele I., (2013), "Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut", Gut Microbes, 4(1), 28–40.

[116] Thiele I., Swainston N., Fleming R. M. T., Hoppe A., Sahoo S., Aurich M. K., Haraldsdottir H., Mo M. L., Rolfsson O., Stobbe M. D., Thorleifsson S. G., Agren R., Bölling C., Bordel S., Chavali A. K., Dobson P., Dunn W. B., Endler L., Hala D., Hucka M., Hull D., Jameson D., Jamshidi N., Jonsson J. J., Juty N., Keating S., Nookaew I., Le Novère N., Malys N., Mazein A., Papin J. A., Price N. D., Selkov E., Sigurdsson M. I., Simeonidis E., Sonnenschein N., Smallbone K., Sorokin A., Van Beek J. H. G. M., Weichart D., Goryanin I., Nielsen J., Westerhoff H. V., Kell D. B., Mendes P., Palsson B. O., (2013), "A community-driven global reconstruction of human metabolism", Nat Biotechnol, 31(5), 419–425.

[117] Palsson B., (2006), Systems biology: Properties of reconstructed networks. Cambridge University Press.

[118] Orth J. D., Thiele I., Palsson B. O., (2010), "What is flux balance analysis?". Nat. Biotechnol. 28:245–248.

[119] Duarte N. C., Becker S. A., Jamshidi N., Thiele I., Mo M. L., Vo T. D., Srivas R., Palsson B., (2007), "Global reconstruction of the human metabolic network based on genomic and bibliomic data", Proc Natl Acad Sci U S A, 104(6), 1777–1782.

[120] Swainston N., Smallbone K., Hefzi H., Dobson P. D., Brewer J., Hanscho M., Zielinski D. C., Ang K. S., Gardiner N. J., Gutierrez J. M., Kyriakopoulos S., Lakshmanan M., Li S., Liu J. K., Martínez V. S., Orellana C. A., Quek L. E., Thomas A., Zanghellini J., Borth N., Lee D. Y., Nielsen L. K., Kell D. B., Lewis N. E., Mendes P., (2016), "Recon 2.2: from reconstruction to model of human metabolism", Metabolomics, 12(7).

[121] Brunk E., Sahoo S., Zielinski D. C., Altunkaya A., Dräger A., Mih N., Gatto F., Nilsson A., Preciat Gonzalez G. A., Aurich M. K., Prlic A., Sastry A., Danielsdottir A. D., Heinken A., Noronha A., Rose P. W., Burley S. K., Fleming R. M. T., Nielsen J., Thiele I., Palsson B. O., (2018), "Recon3D enables a three-dimensional view of gene variation in human metabolism", Nat Biotechnol, 36(3), 272–281.

[122] Robinson J. L., Kocabaş P., Wang H., Cholley P. E., Cook D., Nilsson A., Anton M., Ferreira R., Domenzain I., Billa V., Limeta A., Hedin A., Gustafsson J., Kerkhoven E. J., Svensson L. T., Palsson B. O., Mardinoglu A., Hansson L.,

Uhlén M., Nielsen J., (2020), "An atlas of human metabolism", Sci Signal, 13(624), 1482.

[123] Shlomi T., Cabili M. N., Ruppin E., (2009), "Predicting metabolic biomarkers of human inborn errors of metabolism", Mol Syst Biol, 5.

[124] Becker S. A., Palsson B. O., (2008), "Context-specific metabolic networks are consistent with experiments", PLoS Comput Biol, 4(5), 1000082.

[125] Shlomi T., Benyamini T., Gottlieb E., Sharan R., Ruppin E., (2011), "Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect", PLoS Comput Biol, 7(3), 1002018.

[126] Mardinoglu A., Agren R., Kampf C., Asplund A., Nookaew I., Jacobson P., Walley A. J., Froguel P., Carlsson L. M., Uhlen M., Nielsen J., (2013), "Integration of clinical data with a genome-scale metabolic model of the human adipocyte", Mol Syst Biol, 9(1), 649.

[127] 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", Nat Commun, 5(1), 1–11.

[128] Sigurdsson M. I., Jamshidi N., Steingrimsson E., Thiele I., Palsson B. Ø., (2010), "A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1", BMC Syst Biol, 4, 140.

[129] C C., F B., K L., ML Y., (2003), "Metabolic flux analysis of cultured hepatocytes exposed to plasma", Biotechnol Bioeng, 81(1), 33–49.

[130] K L., F B., GN S., ML Y., (2003), "Profiling of dynamic changes in hypermetabolic livers.", Biotechnol Bioeng, 83(4), 400–415.

[131] Blais E. M., Rawls K. D., Dougherty B. V., Li Z. I., Kolling G. L., Ye P., Wallqvist A., Papin J. A., (2017), "Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions", Nat Commun, 8(1), 1–15.

[132] Sheikh K., Förster J., Nielsen L. K., (2005), "Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus", Biotechnol Prog, 21(1), 112–121.

[133] A M., S S., M B., P G., C Z., E L., F B., J N., (2015), "The gut microbiota modulates host amino acid and glutathione metabolism in mice", Mol Syst Biol, 11(10), 834.

[134] Çakır T., (2019), "Constraint-Based Modeling of Metabolic Interactions in and Between Astrocytes and Neurons"393–420.

[135] Lewis N. E., Schramm G., Bordbar A., Schellenberger J., Andersen M. P., Cheng J. K., Patel N., Yee A., Lewis R. A., Eils R., König R., Palsson B. O., (2010), "Large-scale in silico modeling of metabolic interactions between cell types in the human brain", Nat Biotechnol, 28(12), 1279–1285.

[136] Çakir T., Alsan S., Saybaşili H., Akin A., Ülgen K. Ö., (2007), "Reconstruction and flux analysis of coupling between metabolic pathways of astrocytes and neurons: Application to cerebral hypoxia", Theor Biol Med Model, 4, 48.

[137] Abdik E., (2019), "GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES RECONSTRUCTION OF A BRAIN-SPECIFIC GENOME-SCALE METABOLIC NETWORK MODEL FOR MUS MUSCULUS FOR THE INVESTIGATION OF NEURODEGENERATIVE DISEASES".

[138] Thiele I., Palsson B., (2010), "A protocol for generating a high-quality genome-scale metabolic reconstruction", Nat Protoc, 5(1), 93–121.

[139] Bernstein D. B., Sulheim S., Almaas E., Segrè D., (2021), "Addressing uncertainty in genome-scale metabolic model reconstruction and analysis". Genome Biol. 22:1–22.

[140] Feist A. M., Herrgård M. J., Thiele I., Reed J. L., Palsson B. Ø., (2008), "Reconstruction of biochemical networks in microorganisms", Nat Rev Microbiol 2009 72, 7(2), 129–143.

[141] Wang L., Dash S., Ng C. Y., Maranas C. D., (2017), "A review of computational tools for design and reconstruction of metabolic pathways", Synth Syst Biotechnol, 2(4), 243–252.

[142] Mendoza S. N., Olivier B. G., Molenaar D., Teusink B., (2019), "A systematic assessment of current genome-scale metabolic reconstruction tools", Genome Biol, 20(1), 1–20.

[143] Tian W., Skolnick J., (2003), "How well is enzyme function conserved as a function of pairwise sequence identity?", J Mol Biol, 333(4), 863–882.

[144] Lobb B., Tremblay B. J. M., Moreno-Hagelsieb G., Doxey A. C., (2020), "An assessment of genome annotation coverage across the bacterial tree of life", Microb Genomics, 6(3).

[145] Sorokina M., Stam M., Médigue C., Lespinet O., Vallenet D., (2014), "Profiling the orphan enzymes". Biol. Direct 9:1–16.

[146] Griesemer M., Kimbrel J. A., Zhou C. E., Navid A., D'Haeseleer P., (2018), "Combining multiple functional annotation tools increases coverage of metabolic annotation", BMC Genomics, 19(1), 1–11.

[147] Benson D. A., Karsch-Mizrachi I., Lipman D. J., Ostell J., Wheeler D. L., (2003), "GenBank". Nucleic Acids Res. 31:23–27.

[148] Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M., (2016), "KEGG as a reference resource for gene and protein annotation", Nucleic Acids Res, 44(D1), D457–D462.

[149] Schomburg I., Chang A., Ebeling C., Gremse M., Heldt C., Huhn G., Schomburg D., (2004), "BRENDA, the enzyme database: Updates and major new developments", Nucleic Acids Res, 32(DATABASE ISS.).

[150] Caspi R., Foerster H., Fulcher C. A., Kaipa P., Krummenacker M., Latendresse M., Paley S., Rhee S. Y., Shearer A. G., Tissier C., Walk T. C., Zhang P., Karp P. D., (2008), "The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases", Nucleic Acids Res, 36(SUPPL. 1), D623.

[151] Rawls K. D., Dougherty B. V., Blais E. M., Stancliffe E., Kolling G. L., Vinnakota K., Pannala V. R., Wallqvist A., Papin J. A., (2019), "A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools", Comput Biol Med, 105, 64–71.

[152] Oberhardt M. A., Palsson B., Papin J. A., (2009), "Applications of genome-scale metabolic reconstructions". Mol. Syst. Biol. 5:320.

[153] Othmer H. G., (1976), "The qualitative dynamics of a class of biochemical control circuits", J Math Biol, 3(1), 53–78.

[154] Bailey J. E., (2001), "Complex biology with no parameters". Nat. Biotechnol. 19:503–504.

[155] Palsson B., (2000), "The challenges of in silico biology". Nat. Biotechnol. 18:1147–1150.

[156] Varma A., Palsson B. O., (1994), "Metabolic flux balancing: Basic concepts, scientific and practical use", Bio/Technology, 12(10), 994–998.

[157] PACHECO M. I. P., (2016), "FAST RECONSTRUCTION OF COMPACT CONTEXT- SPECIFIC METABOLIC NETWORK MODELS".

[158] Lee S., Phalakornkule C., Domach M. M., Grossmann I. E., (2000), "Recursive milp model for finding all the alternate optima in lp models for metabolic networks", Comput Chem Eng, 24(2–7), 711–716.

[159] Holzhütter H. G., (2006), "The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies". In: BioSystems. pp 98–107.

[160] 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.

[161] Segrè D., Vitkup D., Church G. M., (2002), "Analysis of optimality in natural and perturbed metabolic networks", Proc Natl Acad Sci U S A, 99(23), 15112–15117.

[162] Robaina Estévez S., Nikoloski Z., (2014), "Generalized framework for context-specific metabolic model extraction methods", Front Plant Sci, 5, 491.

[163] Folger O., Jerby L., Frezza C., Gottlieb E., Ruppin E., Shlomi T., (2011), "Predicting selective drug targets in cancer through metabolic networks", Mol Syst Biol, 7.

[164] Gille C., Bölling C., Hoppe A., Bulik S., Hoffmann S., Hübner K., Karlstädt A., Ganeshan R., König M., Rother K., Weidlich M., Behre J., Holzhütter H. G., (2010), "HepatoNet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology", Mol Syst Biol, 6.

[165] Bordbar A., Lewis N. E., Schellenberger J., Palsson B., Jamshidi N., (2010), "Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions", Mol Syst Biol, 6.

[166] Mo M. L., Jamshidi N., Palsson B., (2007), "A genome-scale, constraint-based approach to systems biology of human metabolism", Mol Biosyst, 3(9), 598–603.

[167] Ma H., Sorokin A., Mazein A., Selkov A., Selkov E., Demin O., Goryanin I., (2007), "The Edinburgh human metabolic network reconstruction and its functional analysis", Mol Syst Biol, 3, 135.

[168] Opdam S., Richelle A., Kellman B., Li S., Zielinski D. C., Lewis N. E., (2017), "A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models", Cell Syst, 4(3), 318-329.e6.

[169] Pacheco M. P., (2018), "Fast recontruction of compact context-specific network models". doi: 10.31237/osf.io/8xcwa.

[170] Zur H., Ruppin E., Shlomi T., (2010), "iMAT: an integrative metabolic analysis tool", Bioinformatics, 26(24), 3140–3142.

[171] 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 Comput Biol, 8(5).

[172] Shlomi T., Cabili M. N., Herrgård M. J., Palsson B., Ruppin E., (2008), "Network-based prediction of human tissue-specific metabolism". Nat. Biotechnol. 26:1003–1010.

[173] Correia S. A. G., (2016), "A framework for the reconstruction and analysis of tissue specific genome-scale metabolic models".

[174] Gautier L., Cope L., Bolstad B. M., Irizarry R. A., (2004), "Affy - Analysis of Affymetrix GeneChip data at the probe level", Bioinformatics, 20(3), 307–315.

[175] Carvalho B. S., Irizarry R. A., (2010), "A framework for oligonucleotide microarray preprocessing", Bioinformatics, 26(19), 2363–2367.

[176] Irizarry R. A., Bolstad B. M., Collin F., Cope L. M., Hobbs B., Speed T. P., (2003), "Summaries of Affymetrix GeneChip probe level data", Nucleic Acids Res, 31(4), e15.

[177] Durinck S., Spellman P. T., Birney E., Huber W., (2009), "Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt", Nat Protoc, 4(8), 1184–1191.

[178] Bolstad B. M., Irizarry R. A., Åstrand M., Speed T. P., (2003), "A comparison

of normalization methods for high density oligonucleotide array data based on variance and bias", Bioinformatics, 19(2), 185–193.

[179] Quackenbush J., (2002), "Microarray data normalization and transformation". doi: 10.1038/ng1032.

[180] Robinson M. D., McCarthy D. J., Smyth G. K., (2009), "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data", Bioinformatics, 26(1), 139–140.

[181] Lun A. T. L., Chen Y., Smyth G. K., (2015), "It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR".

[182] Robinson M. D., Oshlack A., (2010), "A scaling normalization method for differential expression analysis of RNA-seq data", Genome Biol 2010 113, 11(3), 1–9.

[183] Ringnér M., (2008), "What is principal component analysis?". Nat. Biotechnol. 26:303–304.

[184] Durinck S., Moreau Y., Kasprzyk A., Davis S., De Moor B., Brazma A., Huber W., (2005), "BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis", Bioinformatics, 21(16), 3439–3440.

[185] Kimes P. K., Liu Y., Neil Hayes D., Marron J. S., (2017), "Statistical significance for hierarchical clustering", Biometrics, 73(3), 811–821.

[186] Zhang Z., Murtagh F., Poucke S. Van, Lin S., Lan P., (2017), "Hierarchical cluster analysis in clinical research with heterogeneous study population: Highlighting its visualization with R", Ann Transl Med, 5(4).

[187] Xu R., Wunsch D. C., (2010), "Clustering algorithms in biomedical research: A review". IEEE Rev. Biomed. Eng. 3:120–154.

[188] 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 Res, 43(7), e47.

[189] Rosario S. R., Long M. D., Affronti H. C., Rowsam A. M., Eng K. H., Smiraglia D. J., (2018), "Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas", Nat Commun, 9(1), 1–17.

[190] Stekhoven D. J., Bühlmann P., (2012), "Missforest-Non-parametric missing value imputation for mixed-type data", Bioinformatics, 28(1), 112–118.

[191] Kevin Blighe A. L., "PCAtools: everything Principal Component Analysis". https://bioconductor.org/packages/release/bioc/vignettes/PCAtools/inst/doc/PCAtools.html. Accessed 28 Oct 2020.

[192] Kassambara A., (2020), "Package "factoextra" Type Package Title Extract and Visualize the Results of Multivariate Data Analyses".

[193] Jawaid W., (2019), "Package “enrichR” Title Provides an R Interface to “Enrichr”".

[194] N K., T H., N N., C T., (1989), "Inositol 1,4,5-trisphosphate binding sites in the brain: regional distribution, characterization, and alterations in brains of patients with Parkinson’s disease", J Mol Neurosci, 1(3), 181–187.

[195] Gibbs R. A., Weinstock G. M., Metzker M. L., Muzny D. M., Sodergren E. J., Scherer S., Scott G., Steffen D., Worley K. C., Burch P. E., Okwuonu G., Hines S., Lewis L., Deramo C., Delgado O., Dugan-Rocha S., Miner G., Morgan M., Hawes A., Gill R., Holt R. A., Adams M. D., Amanatides P. G., Baden-Tillson H., Barnstead M., Chin S., Evans C. A., Ferriera S., Fosler C., Glodek A., Gu Z., Jennings D., Kraft C. L., Nguyen T., Pfannkoch C. M., Sitter C., Sutton G. G., Venter J. C., Woodage T., Smith D., Lee H. M., Gustafson E., Cahill P., Kana A., Doucette-Stamm L., Weinstock K., Fechtel K., Weiss R. B., Dunn D. M., Green E. D., Blakesley R. W., Bouffard G. G., de Jong P. J., Osoegawa K., Zhu B., Marra M., Schein J., Bosdet I., Fjell C., Jones S., Krzywinski M., Mathewson C., Siddiqui A., Wye N., McPherson J., Zhao S., Fraser C. M., Shetty J., Shatsman S., Geer K., Chen Y., Abramzon S., Nierman W. C., Gibbs R. A., Weinstock G. M., Havlak P. H., Chen R., Durbin K. J., Egan A., Ren Y., Song X. Z., Li B., Liu Y., Qin X., Cawley S., Weinstock G. M., Worley K. C., Cooney A. J., Gibbs R. A., D’Souza L. M., Martin K., Wu J. Q., Gonzalez-Garay M. L., Jackson A. R., Kalafus K. J., McLeod M. P., Milosavljevic A., Virk D., Volkov A., Wheeler D. A., Zhang Z., Bailey J. A., Eichler E. E., Tuzun E., Birney E., Mongin E., Ureta-Vidal A., Woodwark C., Zdobnov E., Bork P., Suyama M., Torrents D., Alexandersson M., Trask B. J., Young J. M., Smith D., Huang H., Fechtel K., Wang H., Xing H., Weinstock K., Daniels S., Gietzen D., Schmidt J., Stevens K., Vitt U., Wingrove J., Camara F., Albà M. M., Abril J. F., Guigo R., Smit A., Dubchak I., Rubin E. M., Couronne O., Poliakov A., Hübner N., Ganten D., Goesele C., Hummel O., Kreitler T., Lee Y. A., Monti J., Schulz H., Zimdahl H., Himmelbauer H., Lehrach H., Jacob H. J., Bromberg S., Gullings-Handley J., Jensen-Seaman M. I., Kwitek A. E., Lazar J., Pasko D., Tonellato P. J., Twigger S., Ponting C. P., Duarte J. M., Rice S., Goodstadt L., Beatson S. A., Emes R. D., Winter E. E., Webber C., Brandt P., Nyakatura G., Adetobi M., Chiaromonte F., Elnitski L., Eswara P., Hardison R. C., Hou M., Kolbe D., Makova K., Miller W., Nekrutenko A., Riemer C., Schwartz S., Taylor J., Yang S., Zhang Y., Lindpaintner K., Andrews T. D., Caccamo M., Clamp M., Clarke L., Curwen V., Durbin R., Eyras E., Searle S. M., Cooper G. M., Batzoglou S., Brudno M., Sidow A., Stone E. A., Venter J. C., Payseur B. A., Bourque G., López-Otín C., Puente X. S., Chakrabarti K., Chatterji S., Dewey C., Pachter L., Bray N., Yap V. B., Caspi A., Tesler S. D. G., Pevzner P. A., Haussler S. C. D., Roskin K. M., Baertsch R., Clawson H., Furey T. S., Hinrichs A. S., Karolchik D., Kent W. J., Rosenbloom K. R., Trumbower H., Weirauch M., Cooper D. N., Stenson P. D., Ma B., Brent M., Arumugam M., Shteynberg D., Copley R. R., Taylor M. S., Riethman H., Mudunuri U., Peterson J., Guyer M., Felsenfeld A., Old S., Mockrin S., Collins F., (2004), "Genome sequence of the Brown Norway rat yields insights into mammalian evolution", Nature, 428(6982), 493–520.

[196] Shannahan J. H., Schladweiler M. C. J., Richards J. H., Ledbetter A. D., Ghio A. J., Kodavanti U. P., (2010), "Pulmonary oxidative stress, inflammation, and dysregulated iron homeostasis in rat models of cardiovascular disease", J Toxicol

Environ Heal - Part A Curr Issues, 73(10), 641–656.

[197] Igarashi Y., Nakatsu N., Yamashita T., Ono A., Ohno Y., Urushidani T., Yamada H., (2015), "Open TG-GATEs: A large-scale toxicogenomics database", Nucleic Acids Res, 43(D1), D921–D927.

[198] Monk J., Nogales J., Palsson B. O., (2014), "Optimizing genome-scale network reconstructions". Nat. Biotechnol. 32:447–452.

[199] Deitmer J. W., "Glial strategy for metabolic shuttling and neuronal function".

[200] Hertz L., Dringen R., Schousboe A., Robinson S. R., (1999), "Mini-Review Astrocytes: Glutamate Producers for Neurons".

[201] Fields R. D., Stevens-Graham B., (2002), "Neuroscience: New insights into neuron-glia communication". Science (80-.). 298:556–562.

[202] Occhipinti R., Puchowicz M. A., Lamanna J. C., Somersalo E., Calvetti D., (2007), "Statistical analysis of metabolic pathways of brain metabolism at steady state". In: Annals of Biomedical Engineering. Springer, pp 886–902.

[203] García-Romero I., Nogales J., Díaz E., Santero E., Floriano B., (2020), "Understanding the metabolism of the tetralin degrader Sphingopyxis granuli strain TFA through genome-scale metabolic modelling", Sci Rep, 10(1), 1–14.

[204] Hu Y., Flockhart I., Vinayagam A., Bergwitz C., Berger B., Perrimon N., Mohr S. E., (2011), "An integrative approach to ortholog prediction for disease-focused and other functional studies", BMC Bioinformatics, 12, 357.

[205] Shimoyama M., De Pons J., Hayman G. T., Laulederkind S. J. F., Liu W., Nigam R., Petri V., Smith J. R., Tutaj M., Wang S. J., Worthey E., Dwinell M., Jacob H., (2015), "The Rat Genome Database 2015: Genomic, phenotypic and environmental variations and disease", Nucleic Acids Res, 43(D1), D743–D750.

[206] Boyce S., Tipton K. F., (2001), "Enzyme Classification and Nomenclature". In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd.

[207] Kim S., Thiessen P. A., Bolton E. E., Chen J., Fu G., Gindulyte A., Han L., He J., He S., Shoemaker B. A., Wang J., Yu B., Zhang J., Bryant S. H., (2016), "PubChem substance and compound databases", Nucleic Acids Res, 44(D1), D1202–D1213.

[208] Baart G. J. E., Martens D. E., (2012), "Genome-scale metabolic models: Reconstruction and analysis". In: Methods in Molecular Biology. Humana Press Inc., pp 107–126.

[209] Mookerjee S. A., Nicholls D. G., Brand M. D., (2016), "Determining Maximum Glycolytic Capacity Using Extracellular Flux Measurements", PLoS One, 11(3), e0152016.

[210] DiNuzzo M., Giove F., Maraviglia B., Mangia S., (2017), "Computational Flux Balance Analysis Predicts that Stimulation of Energy Metabolism in Astrocytes

and their Metabolic Interactions with Neurons Depend on Uptake of K+ Rather than Glutamate", Neurochem Res, 42(1), 202–216.

[211] Keating S. M., Bornstein B. J., Finney A., Hucka M., (2006), "SBMLToolbox: An SBML toolbox for MATLAB users", Bioinformatics, 22(10), 1275–1277.

[212] Schellenberger J., Que R., Fleming R. M. T., Thiele I., Orth J. D., Feist A. M., Zielinski D. C., Bordbar A., Lewis N. E., Rahmanian S., Kang J., Hyduke D. R., Palsson B., (2011), "Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0", Nat Protoc, 6(9), 1290–1307.

[213] Heirendt L., Arreckx S., Pfau T., Mendoza S. N., Richelle A., Heinken A., Haraldsdóttir H. S., Wachowiak J., Keating S. M., Vlasov V., Magnusdóttir S., Ng C. Y., Preciat G., Žagare A., Chan S. H. J., Aurich M. K., Clancy C. M., Modamio J., Sauls J. T., Noronha A., Bordbar A., Cousins B., El Assal D. C., Valcarcel L. V., Apaolaza I., Ghaderi S., Ahookhosh M., Ben Guebila M., Kostromins A., Sompairac N., Le H. M., Ma D., Sun Y., Wang L., Yurkovich J. T., Oliveira M. A. P., Vuong P. T., El Assal L. P., Kuperstein I., Zinovyev A., Hinton H. S., Bryant W. A., Aragón Artacho F. J., Planes F. J., Stalidzans E., Maass A., Vempala S., Hucka M., Saunders M. A., Maranas C. D., Lewis N. E., Sauter T., Palsson B., Thiele I., Fleming R. M. T., (2019), "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0", Nat Protoc, 14(3), 639–702.

[214] L B., AG C., BK S., (1976), "Glucose consumption in the cerebral cortex of rat during bicuculline-induced status epilipticus", J Neurochem, 27(4), 971–973.

[215] Linde R., Schmalbruch I. K., Paulson O. B., Madsen P. L., (1999), "The Kety-Schmidt technique for repeated measurements of global cerebral blood flow and metabolism in the conscious rat", Acta Physiol Scand, 165(4), 395–401.

[216] Madsen P. L., Linde R., Hasselbalch S. G., Paulson O. B., Lassen N. A., (1998), "Activation-induced resetting of cerebral oxygen and glucose uptake in the rat", J Cereb Blood Flow Metab, 18(7), 742–748.

[217] Du F., Zhang Y., Zhu X. H., Chen W., (2012), "Simultaneous measurement of glucose blood-brain transport constants and metabolic rate in rat brain using in-vivo 1 H MRS", J Cereb Blood Flow Metab, 32(9), 1778–1787.

[218] Sibson N. R., Dhankhar A., Mason G. F., Rothman D. L., Behar K. L., Shulman R. G., (1998), "Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity", Proc Natl Acad Sci U S A, 95(1), 316–321.

[219] Hertz L., Peng L., Dienel G. A., (2007), "Energy metabolism in astrocytes: High rate of oxidative metabolism and spatiotemporal dependence on glycolysis/glycogenolysis". J. Cereb. Blood Flow Metab. 27:219–249.

[220] Dienel G. A., Cruz N. F., (2004), "Nutrition during brain activation: Does cell-to-cell lactate shuttling contribute significantly to sweet and sour food for thought?", Neurochem Int, 45(2–3), 321–351.

[221] Lo A. S. Y., Liew C. T., Ngai S. M., Tsui S. K. W., Fung K. P., Lee C. Y., Waye M. M. Y., (2005), "Developmental regulation and cellular distribution of human cytosolic malate dehydrogenase (MDH1)", J Cell Biochem, 94(4), 763–773.

[222] Wahren J., Ekberg K., Fernqvist-Forbes E., Nair S., (1999), "Brain substrate utilisation during acute hypoglycaemia", Diabetologia, 42(7), 812–818.

[223] Smith Q. R., (2000), "Transport of glutamate and other amino acids at the blood-brain barrier". In: Journal of Nutrition. American Institute of Nutrition, pp 1016–1022.

[224] NR S., A D., GF M., KL B., DL R., RG S., (1997), "In vivo 13C NMR measurements of cerebral glutamine synthesis as evidence for glutamate-glutamine cycling", Proc Natl Acad Sci U S A, 94(6), 2699–2704.

[225] Wiesinger H., Hamprecht B., Dringen R., (1997), "Metabolic pathways for glucose in astrocytes", Glia, 21(1), 22–34.

[226] Ben-Yoseph O., Camp D. M., Robinson T. E., Ross B. D., (1995), "Dynamic Measurements of Cerebral Pentose Phosphate Pathway Activity In Vivo Using [1,6-13C2,6,6-2H2] Glucose and Microdialysis", J Neurochem, 64(3), 1336–1342.

[227] Patel A. B., De Graaf R. A., Mason G. F., Rothman D. L., Shulman R. G., Behar K. L., (2005), "The contribution of GABA to glutamate/glutamine cycling and energy metabolism in the rat cortex in vivo", Proc Natl Acad Sci U S A, 102(15), 5588–5593.

[228] Haggart C. R., Bartell J. A., Saucerman J. J., Papin J. A., (2011), "Whole-genome metabolic network reconstruction and constraint-based modeling". In: Methods in Enzymology. Academic Press Inc., pp 411–433.

[229] Lubow J. M., Piñón I. G., Avogaro A., Cobelli C., Treeson D. M., Mandeville K. A., Toffolo G., Boyle P. J., (2006), "Brain oxygen utilization is unchanged by hypoglycemia in normal humans: Lactate, alanine, and leucine uptake are not sufficient to offset energy deficit", Am J Physiol - Endocrinol Metab, 290(1).

[230] Madsen P. L., Cruz N. F., Sokoloff L., Dienel G. A., (1999), "Cerebral oxygen/glucose ratio is low during sensory stimulation and rises above normal during recovery: Excess glucose consumption during stimulation is not accounted for by lactate efflux from or accumulation in brain tissue", J Cereb Blood Flow Metab, 19(4), 393–400.

[231] Nybo L., Nielsen B., Blomstrand E., Møller K., Secher N., (2003), "Neurohumoral responses during prolonged exercise in humans", J Appl Physiol, 95(3), 1125–1131.

[232] Gruetter R., Seaquist E. R., Ugurbil K., (2001), "A mathematical model of compartmentalized neurotransmitter metabolism in the human brain", Am J Physiol - Endocrinol Metab, 281(1 44-1).

[233] Lebon V., Petersen K. F., Cline G. W., Shen J., Mason G. F., Dufour S., Behar K. L., Shulman G. I., Rothman D. L., (2002), "Astroglial contribution to brain energy metabolism in humans revealed by 13C nuclear magnetic resonance spectroscopy: Elucidation of the dominant pathway for neurotransmitter glutamate repletion and measurement of astrocytic oxidative metabolism", J Neurosci, 22(5), 1523–1531.

[234] Shen J., Petersen K. F., Behar K. L., Brown P., Nixon T. W., Mason G. F., Petroff O. A. C., Shulman G. I., Shulman R. G., Rothman D. L., (1999), "Determination of the rate of the glutamate/glutamine cycle in the human brain by in vivo 13C NMR", Proc Natl Acad Sci U S A, 96(14), 8235–8240.

[235] Öz G., Berkich D. A., Henry P. G., Xu Y., LaNoue K., Hutson S. M., Gruetter R., (2004), "Neuroglial metabolism in the awake rat brain: CO2 fixation increases with brain activity", J Neurosci, 24(50), 11273–11279.

[236] Aureli T., Di Cocco M. E., Calvani M., Conti F., (1997), "The entry of [1-13C]glucose into biochemical pathways reveals a complex compartmentation and metabolite trafficking between glia and neurons: A study by 13C-NMR spectroscopy", Brain Res, 765(2), 218–227.

[237] Mardinoglu A., Gatto F., Nielsen J., (2013), "Genome-scale modeling of human metabolism - a systems biology approach". Biotechnol. J. 8:985–996.

[238] P B., (2012), "Perfusion and metabolism imaging studies in Parkinson's disease.", Dan Med J, 59(6), B4466–B4466.

[239] Gavai A. K., Supandi F., Hettling H., Murrell P., Leunissen J. A. M., Van Beek J. H. G. M., (2015), "Using bioconductor package BiGGR for metabolic flux estimation based on gene expression changes in brain", PLoS One, 10(3).

[240] Plaitakis A., Shashidharan P., (2000), "Glutamate transport and metabolism in dopaminergic neurons of substantia nigra: Implications for the pathogenesis of Parkinson's disease". In: Journal of Neurology, Supplement. J Neurol, pp 25–35.

[241] M S.-W., M A., (2013), "Manganese toxicity in the central nervous system: the glutamine/glutamate-γ-aminobutyric acid cycle", J Intern Med, 273(5), 466–477.

[242] Oerton E., Bender A., (2017), "Concordance analysis of microarray studies identifies representative gene expression changes in Parkinson's disease: A comparison of 33 human and animal studies", BMC Neurol, 17(1), 58.

[243] Cabbia A., Hilbers P. A. J., van Riel N. A. W., (2020), "A Distance-Based Framework for the Characterization of Metabolic Heterogeneity in Large Sets of Genome-Scale Metabolic Models", Patterns, 1(6), 100080.

[244] Magnúsdóttir S., Heinken A., Kutt L., Ravcheev D. A., Bauer E., Noronha A., Greenhalgh K., Jäger C., Baginska J., Wilmes P., Fleming R. M. T., Thiele I., (2016), "Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota", Nat Biotechnol 2016 351, 35(1), 81–89.

[245] Bookstein A., Kulyukin V. A., Raita T., (2002), "Generalized Hamming

Distance", Inf Retr 2002 54, 5(4), 353–375.

[246] Nidelet T., Brial P., Camarasa C., Dequin S., (2016), "Diversity of flux distribution in central carbon metabolism of S. cerevisiae strains from diverse environments", Microb Cell Factories 2016 151, 15(1), 1–13.

[247] Wang H., Marcišauskas S., Sánchez B. J., Domenzain I., Hermansson D., Agren R., Nielsen J., Kerkhoven E. J., (2018), "RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor", PLoS Comput Biol, 14(10).

APPENDİX

Appendix A: Supplementary Files

Supplementary files mentioned in the thesis are provided in the electronic format and listed below:

- Supplementary File 1: iBrain766-Rn Genome Scale Metabolic Model