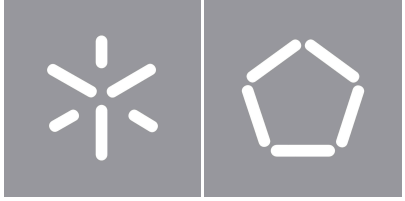


University of Minho
School of Engineering

Bruno Araújo Gomes Sá

Implementing Metabolic Transformation Algorithms and their application in Ageing-related research

december 2024



University of Minho
School of Engineering

Bruno Araújo Gomes Sá

Implementing Metabolic Transformation Algorithms and their application in Ageing-related research

Master's Dissertation in Bioinformatics

Dissertation supervised by
Miguel Francisco de Almeida Pereira da Rocha

december 2024

Copyright and Terms of Use for Third Party Work

This dissertation reports on academic work that can be used by third parties as long as the internationally accepted standards and good practices are respected concerning copyright and related rights.

This work can thereafter be used under the terms established in the license below.

Readers needing authorization conditions not provided for in the indicated licensing should contact the author through the RepositóriUM of the University of Minho.

License granted to users of this work:



CC BY-NC-SA

https://creativecommons.org/licenses/by-nc-sa/4.0/

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Miguel Rocha, for his unwavering support, guidance, and encouragement throughout this journey. His expertise and insightful feedback have been invaluable, and his mentorship has greatly enriched my academic experience.

I am also profoundly thankful to Alexandre, whose initial idea for this thesis laid the foundation of this work. His assistance and willingness to help at every step were instrumental in the completion of this work.

To my research group, BISBi, I extend my heartfelt appreciation for the camaraderie and memorable moments we shared. The dinners, retreats, and especially our lively "casino" sessions after lunch have made this journey enjoyable and unforgettable. A special mention goes to the "Biosala," the group of master students who shared their joys and challenges with me. Our time together has been a source of strength and motivation.

I would also like to thank my close group of friends, "FUGA," for the joy and balance they bring into my life. They have reminded me that life is not just about work, and their friendship has been a constant source of happiness. Additionally, I am grateful to all my music friends for keeping my passion for music alive and never letting me abandon it.

Finally, but certainly not least, I want to express my deepest appreciation to my family for their unwavering support and understanding. Even when the complexities of this process were unfamiliar to them, their encouragement never wavered, and for that, I am eternally grateful.

Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, Braga, november 2024

Bruno Araújo Gomes Sá

Abstract

Metabolic Transformation Algorithm (MTA) is a computational method that identifies potential metabolic interventions to shift a biological system from an undesired state to a desired one by targeting specific genes or reactions within a metabolic network. The Robust Metabolic Transformation Algorithm (rMTA) enhances this capability by incorporating worst-case scenario analyses and integrating the Minimization Of Metabolic Adjustment (MOMA) approach.

By transitioning these algorithms from MATLAB to Python, this work aims to enhance their accessibility and usability within the scientific community. The Python implementations were validated by successfully replicating results from the original MATLAB implementation, including gene knockouts of *RRM1* and *RRM2*.

Furthermore, the rMTA algorithm was applied to an aging-related scenario in *Caenorhabditis elegans* to identify candidate genes and pathways involved in the metabolic response to *unc-62* knockdown.

Additionally, this work integrates an Evolutionary Algorithms (EAs)-based approach to rMTA to evaluate the effects of multiple gene knockouts simultaneously. This extension optimizes the identification of metabolic targets for phenotype reversion, demonstrating the potential for more effective therapeutic strategies.

In summary, this dissertation enhances the accessibility of MTA and rMTA by implementing them in Python, applying them to an aging-related scenario, and extending their capabilities through the integration of EAs.

Keywords Metabolic Transformation Algorithm; Robust Metabolic Transformation Algorithm; Drug Identification; Python implementation; Aging research; *Caenorhabditis elegans*; Evolutionary Algorithms;

Resumo

O MTA é um método computacional que identifica potenciais alvos metabólicos de forma a alterar um sistema biológico de um estado indesejado para um estado desejado, visando genes ou reações específicas dentro de uma rede metabólica. O rMTA aprimora esta capacidade ao incorporar análise do pior cenário e ao integrar um abordagem mais robusta, o MOMA.

Ao migrar estes algoritmos do MATLAB para Python, este trabalho tem como objetivo aumentar a sua acessibilidade e usabilidade na comunidade científica. As implementações em Python foram validadas ao replicar com sucesso os resultados da implementação original em MATLAB, incluindo knockouts dos genes *RRM1* e *RRM2*.

Adicionalmente, o algoritmo rMTA foi aplicado a um caso de estudo relacionado com o envelhecimento em *Caenorhabditis elegans*, a fim de identificar genes candidatos e vias envolvidas na resposta metabólica ao knockdown do gene *unc-62*.

Para além disso, este trabalho integra uma abordagem baseada em EAs ao rMTA para avaliar os efeitos de múltiplos knockouts de genes em simultâneo. Esta extensão optimiza a identificação de alvos metabólicos para a reversão do fenótipo, demonstrando o seu potencial para estratégias mais eficazes.

Em resumo, esta dissertação melhora a acessibilidade do MTA e do rMTA ao implementá-los em Python, aplicando-os a um cenário relacionado com o envelhecimento, e ampliando as suas capacidades através da integração de EAs.

Palavras-chave Algoritmo de Transformação Metabólica; Algoritmo de Transformação Metabólica Robusto; Identificação de Fármacos; Implementação em Python; Investigação sobre o envelhecimento; *Caenorhabditis elegans*; Algoritmos Evolutivos;

Contents

I Introductory material **1**

1 Introduction **2**

1.1 Context and motivation 2

1.2 Objectives 3

1.3 Thesis Outline 4

2 State of the art **6**

2.1 Ageing-Related Research 6

2.1.1 Neurodegenerative Diseases 7

2.1.2 Model organism: *Caenorhabditis elegans* 7

2.2 Omics data 8

2.3 Genome-Scale Metabolic Models 11

2.3.1 Phenotype Simulation 14

2.3.2 Flux Sampling Methods 15

2.3.3 Human Models 16

2.3.4 *Caenorhabditis elegans* models 17

2.4 Context-specific Metabolic Models reconstruction 18

2.5 Metabolic Transformation Algorithm 22

2.6 Evolutionary Algorithms 26

II Core of the Dissertation **28**

3 rMTA/MTA Implementation and Validation **29**

3.1 Introduction 29

3.1.1 Pipeline for Drug Identification Using MTA . . . 30
3.2 Implementation . . . 31
3.3 Pipeline Validation . . . 40
3.3.1 *RRM1* Validation Study . . . 40
3.3.2 *RRM2* Validation Study . . . 48
3.3.3 Discussion . . . 49

4 Case Study: aging research using Caenorhabditis elegans model 51
4.1 Introduction . . . 51
4.2 Methods . . . 52
4.2.1 Calculation of Reference Flux of the *unc-62* Knockdown State . . . 52
4.2.2 Calculation of Differential Expressed Genes . . . 53
4.2.3 rMTA . . . 53
4.2.4 Enrichment Analysis . . . 53
4.3 Results and Discussion . . . 54
4.3.1 Differential Expressed Genes . . . 54
4.3.2 rMTA Results . . . 54
4.3.3 Enrichment . . . 57
4.4 Conclusion . . . 60

5 Evolutionary Algorithm approach to rMTA 61
5.1 Introduction . . . 61
5.2 Methods . . . 61
5.2.1 Random Sampling . . . 61
5.2.2 Evolutionary Algorithms . . . 62
5.2.3 Post-Processing . . . 64
5.3 Results . . . 65
5.3.1 Random Sampling Aproach . . . 65
5.3.2 EAs Aproach . . . 67
5.4 Discussion . . . 72
5.4.1 Random Sampling Aproach . . . 72
5.4.2 EAs Approach . . . 72
5.4.3 Biological Interpretation of Identified Genes . . . 73

5.4.4 Limitations 73
5.5 Conclusion 74

6 Conclusions and Future Work **75**
6.1 Conclusions 75
6.2 Prospects for Future Work 76

III Appendices **95**

A Details of results **96**

List of Figures

1 Conceptual representation of phenotype simulation. 14
2 Representation of the main algorithms for omics data integration. 19
3 Schematic representation of iMAT 21
4 Schematic representation of MTA: Part 1 23
5 Schematic representation of MTA: Part 2 24
6 The main steps of an Evolutionary Algorithm 26

7 `DifferentialExpression` class structure 32
8 `MTA` class 38
9 `rMTA` class 40
10 Correlation of epsilon values between Python and MATLAB implementations. 43
11 Venn diagram showing the overlap of top-ranked genes 45
12 Correlation plot of gene rankings between the original study and our replication for *RRM1*. 46
13 Comparison of rTS values between the original study and our replication for *RRM1*. . . . 47
14 Distribution of rTS values for *RRM1*. 48

15 Volcano plot of differential gene expression analysis 54
16 Enrichment analysis of the top 300 genes 58
17 Enrichment analysis of the top 50 genes 59

18 Distribution of Transformation Scores from Random Sampling 65
19 Comparison of Robust Transformation Score (rTS) distributions 66

20 Venn diagram showing the overlap of top-ranked genes 96
21 Correlation plot of gene rankings between the original study and our replication for *RRM2*. 97
22 Comparison of rTS values between the original study and our replication for *RRM2*. . . . 97
23 Distribution of rTS values for *RRM2*. 98

List of Tables

1 Summary of published Human Models 17

2 Summary of published *C.elegans* Models 18

3 Comparison of gene classifications between Python and MATLAB implementations. . . . 42

4 Comparison of reaction classifications between Python and MATLAB implementations. . 42

5 Comparison of curated reaction classifications between Python and MATLAB implementations. 42

6 Comparison of epsilon values between Python and MATLAB implementations. 43

7 Ranking positions of *RRM1* in the original study and our replication. 44

8 Top 10 ranked genes from the original study and our replication for *RRM1*. 44

9 Ranking positions of *RRM1* in the original study and our replication with only common genes. 45

10 Common top-ranked genes between the original study and our replication for *RRM1*. . . 46

11 Ranking positions of *RRM2* in the original study and our replication. 49

12 Top 10 ranked genes from the original study and our replication for *RRM2*. 49

13 Comparison of original and consistent model characteristics. 52

14 Top-ranked genes from the rMTA analysis. 55

15 Correlation matrix between different TS metrics 63

16 Example of EA output 64

17 Comparation between top random sapling solutions and top rMTA ranking 66

18 Baseline of the Initial mTS score of the top rMTA gene (*PYCR1*). 67

19 Optimal gene deletions using top 25 genes with a maximum of 2 knockouts. 67

20 Optimal gene deletions using top 25 genes with a maximum of 3 knockouts. 67

21 Optimal gene deletions using top 100 genes with a maximum of 2 knockouts. 68

22 Optimal gene deletions using top 100 genes with a maximum of 3 knockouts. 68

23 Optimal gene deletions using all genes with a maximum of 2 knockouts. 68
24 Baseline rTS score of the top rMTA gene (*PYCR1*). 69
25 Optimal gene deletion using top 25 genes with maximum solution sizes of 2 and 3 knockouts. 69
26 Optimal gene deletions using top 100 genes with a maximum of 10 knockouts. 69
27 Optimal gene deletions using top 100 genes with a maximum of 10 knockouts. 70
28 Optimal gene deletions using top 550 genes with a maximum of 2 knockouts. 70
29 Optimal gene deletions using top 550 genes with a maximum of 3 knockouts. 70
30 Summary of optimal modifications and their rTS 71
31 Top 5 most frequently occurring genes in optimal solutions. 71

Acronyms

$v^{\mathbf{ref}}$ Reference Flux Distribution.

ACHR Artificial Centering Hit-and-Run.

ATP Adenosine triphosphate.

bTS Best-case Transformation Score.

CHRR Coordinate Hit-and-Run with Rounding.

DEGs Differentially Expressed Genes.

DGE Differential Gene Expression.

dNDP deoxyribonucleotide diphosphate.

dNTP deoxyribonucleotide triphosphate.

EAs Evolutionary Algorithms.

EHMN Edinburgh Human Metabolic Network.

FBA Flux Balance Analysis.

FDR false discovery rate.

FPKM Fragments Per Kilobase Million.

GEMs Genome-scale Metabolic Models.

GEO Gene Expression Omnibus.

GIMME-like Gene Inactivity Moderated by Metabolism and Expression.

GPR Gene-Protein-Rule.

GSEA Gene Set Enrichment Analysis.

HMR2 Human Model Reaction 2.

iMAT Integrative Metabolic Analysis tool.

iMAT-like Integrative Metabolic Analysis tool.

INIT Integrative Network Inference for Tissues.

KEGG Kyoto Encyclopedia of Genes and Genomes.

log2FC log2 fold change.

log-CPM log2-counts per million.

MBA-like Model-Building Algorithm.

MILP Mixed Integer Linear Programming.

MIQP Mixed Integer Quadratic Programming.

MOMA Minimization Of Metabolic Adjustment.

MTA Metabolic Transformation Algorithm.

mTS MOMA Transformation Score.

NDP ribonucleoside diphosphate.

OPTGP Optimized General Parallel.

pFBA Parsimonious Flux Balance Analysis.

RMF Required Metabolic Function.

rMTA Robust Metabolic Transformation Algorithm.

RNA-seq RNA-sequencing.

RPKM Reads Per Kilobase Million.

RR ribonucleotide reductase.

rTS Robust Transformation Score.

TPM Transcripts Per Million.

TS Transformation Score.

wTS Worst-case Transformation Score.

Part I

Introductory material

Chapter 1

Introduction

1.1 Context and motivation

Aging is characterized by a gradual decline in biological function and an increased susceptibility to diseases. This multifaceted challenge intersects several domains of biology. The process is driven by a complex interplay of factors, culminating in a gradual decline of physiological integrity. This constitutes a primary risk factor for a wide spectrum of age-related diseases such as neurodegeneration, cardiovascular diseases, and cancer. Aging is not merely a chronological phenomenon, but is a complex interaction of various factors, partially shaped by biological pathways and biochemical processes (1; 2; 3; 4; 5).

Systems biology emerges as a crucial framework in this context. Its goal is to understand biological systems holistically by integrating complex interactions at various levels, from genes to cells to entire organisms. This approach is particularly suited to the study of aging, where single-gene or single-pathway analyses fall short in capturing the multiscale interactions that drive the aging process. In the context of aging, systems biology seeks to map the intricate networks of genes, proteins, and metabolic pathways, aiming to understand how alterations in these networks contribute to the phenotypic manifestations of aging (6; 7).

The MTA (8) and its robust version rMTA (9) have emerged as promising computational tools in this domain, particularly for identifying drug targets in metabolic disorders. Unlike traditional methods that often focus on cell destruction, as in cancer therapies and antibiotics (10; 11), MTA uniquely identifies drug targets that can shift a cell's metabolism from a diseased state to a healthier one. MTA combines Genome-scale Metabolic Models (GEMs) with Differential Gene Expression (DGE) data from diseased and healthy metabolic states to pinpoint and revert crucial differences.

Originally developed in MATLAB, these algorithms offer a novel approach by leveraging advanced computational techniques to analyze metabolic networks and identify potential drug targets. However, their wider application in the research community has been constrained by MATLAB's environment. The

adaptation and implementation of MTA and rMTA in Python represent a significant step forward in the identification of drug targets. The transition to Python, known for its versatility, open-source nature, and strong support in the scientific community, offers a more accessible, flexible, and collaborative platform for these algorithms, and provides the possibility to interact with a comprehensive set of constraint-based tools written in Python, enriching this ecosystem.

The primary goal of this dissertation is to bridge this gap by adapting the MTA and its robust version rMTA to Python, a more accessible and versatile programming language. This transition is expected to democratize the use of these algorithms, facilitating their application in aging-related phenotype research. By replicating the outcomes of the original MATLAB versions through various case studies including gene knockout perturbations, Alzheimer's disease, and prostate cancer, and comparing them with the new Python implementations, this research aims to validate the effectiveness and reliability of the Python-based MTA algorithm.

Furthermore, the application of these algorithms to study aging phenotypes in *Caenorhabditis elegans* (C. elegans) represents a novel approach to understanding and potentially reversing aging mechanisms. The insights that could be gained from this research have the potential to revolutionize the identification of therapeutic pathways in age-related diseases.

A novel aspect of this research is the enhancement of MTA's capabilities through the integration of evolutionary algorithms, aiming to set a new benchmark in disease phenotype reversion. This integration aims to optimize the reversion of disease-related phenotypes, thereby refining the precision and effectiveness of drug target identification.

1.2 Objectives

The primary aim of this dissertation is to adapt and implement the MTA and rMTA algorithms in Python, validating them through case studies, and applying them in aging research, namely to identify drug targets for reversing aging phenotypes in *C. elegans*, with an innovative integration of evolutionary algorithms for enhanced effectiveness.

To achieve this, the following objectives must be achieved:

- Literature review on neurodegenerative diseases, aging research, MTA and rMTA and their implementations and current applications.
- Implementation of the MTA and rMTA algorithms in Python and validation by comparative analyses using several case studies to ensure the reliability and accuracy of the Python-based algorithms.

This will involve replicating results from the original MATLAB versions and conducting comparative analyses using case studies that include gene knockout perturbations.

- Application of the MTA and rMTA novel implementations to identify and analyze aging-related phenotypes in *C. Elegans*, to discover critical drug targets and pathways that can reverse aging phenotypes, providing insights into potential therapeutic strategies for human age-related conditions.

- Optimization of MTA and rMTA recurring to evolutionary algorithms to optimize the reversion of disease-related phenotypes. This approach aims to refine target identification and therapeutic strategies, enhancing the algorithm's effectiveness in mitigating aging and disease processes.

1.3 Thesis Outline

This thesis is organized into several chapters, each dedicated to a specific aspect of the research conducted. The structure is as follows:

- **Chapter 1: Introduction** – This chapter provides an overview of the context and motivation behind the research, with an emphasis on aging, its biological implications, and the role of systems biology. It also introduces the MTA and its robust counterpart, rMTA, and their relevance in identifying drug targets through metabolic network manipulation. The chapter concludes with the main objectives of the dissertation.

- **Chapter 2: State of the Art** – A review of the relevant literature is presented in this chapter, covering research in aging and neurodegenerative diseases, the model organism *Caenorhabditis elegans*, omics data integration, and genome-scale metabolic models. It also explores the computational techniques involved in phenotype simulation, flux sampling, and the development of context-specific metabolic models. The chapter ends with an overview of the Metabolic Transformation Algorithm (MTA) and its variants, as well as evolutionary algorithms.

- **Chapter 3: rMTA/MTA Implementation and Validation** – This chapter details the adaptation and implementation of MTA and rMTA in Python. It includes a description of the pipeline for drug identification using MTA, the steps involved in model reconstruction, and gene knockout simulations. Validation of these implementations is carried out by comparing results with the original MATLAB versions.

- **Chapter 4: Case Study: Aging Research Using Caenorhabditis elegans Model** – A case study is conducted applying the rMTA to an aging-related research scenario using *C. elegans*. The goal is to uncover metabolic changes associated with the knockdown of the gene *unc-62*, which plays a role in aging and longevity. The chapter describes the methods used for the calculation of reference flux, gene expression analysis, rMTA execution, and enrichment analysis of the identified genes.

- **Chapter 5: Evolutionary Algorithm Approach to rMTA** – This chapter introduces evolutionary algorithms as an extension to rMTA, optimizing the reversion of disease-related phenotypes. It presents the methodology and results of integrating evolutionary algorithms with rMTA to assess multiple gene knockouts simultaneously.

- **Chapter 6: Conclusions and future work** – The final chapter summarizes the findings of the thesis, reflecting on the success of the Python implementation of MTA and rMTA, their application in aging research, and the potential for future studies that integrate these algorithms with evolutionary approaches to further advance the field of metabolic network interventions.

Chapter 2

State of the art

2.1 Ageing-Related Research

Aging is a complex and inevitable biological process, influenced by a combination of inherent genetic factors and external environmental factors. It is characterized by a gradual deterioration in physiological capabilities, manifesting the organism's reduced capacity to adapt and respond to environmental changes (5). This process is not merely a superficial deterioration, but rather a fundamental decrease in the regenerative potential of systemic tissues and organs (12).

The alterations associated with aging serve as the predominant risk factors for neurodegenerative conditions, cardiovascular disorders, and cancer. The process of aging involves complex mechanisms, such as replicative senescence and stress-induced senescence, which contribute to several cellular and molecular transformations. Prominent features of aging include genomic instability, loss of proteostasis, telomere attrition, dysregulated nutrient-sensing, mitochondrial failure, cellular senescence, depletion of stem cells, and altered intercellular communication (1). Predominant theories that explain the mechanisms underlying aging include DNA damage theory, genetic senescence theory, the telomere theory, and the free radical theory (13).

In recent years, there has been a significant augmentation in human life expectancy, attributable to advancements in medical care, nutrition, and general living conditions. This increase in longevity has contributed to a demographic shift, with a growing proportion of the global population being elderly. Consequently, there has been a rise in the prevalence of age-associated diseases. Disorders like Alzheimer's and Parkinson's are increasingly prevalent, presenting substantial global challenges (14). This event underscores the necessity for thorough research to address neurodegenerative diseases.

2.1.1 Neurodegenerative Diseases

Neurodegenerative disorders are characterized by the progressive death and loss of nerve cell structure and function in the brain, primarily affecting older adults (15). Alzheimer's and Parkinson's diseases are the most prevalent among these disorders (14). Despite their widespread occurrence globally, there remains a significant deficiency in effective treatments, with the majority of current therapies aimed at relieving symptoms rather than curing the underlying disease (16).

At the pathological level, these disorders are categorized as proteinopathies, characterized by the detrimental accumulation of misfolded proteins. This accumulation disrupts cellular function and accelerates disease progression (17). However, the pathology of neurodegenerative diseases is multifaceted, extending beyond protein aggregation. It involves a complex network of disrupted homeostatic regulation, encompassing immune response dysregulation, metabolic alterations, synaptic degradation, and neuronal loss (15).

A critical aspect of neurodegenerative diseases is the presence of variants known for their monogenic origins. These variants serve as valuable models for *in vitro*, *in vivo* and *in silico* studies, aimed at unraveling the complexities of the more prevalent and multifactorial forms of the diseases (6).

Another crucial consideration is the dynamic nature of these pathophysiological characteristics, which evolve as the disease progresses. This underscores the necessity for constructing models capable of simulating disease progression at various stages in different tissues. The most widely utilized models are those that mimic the genetic mutations found in the monogenic forms of these diseases, being transgenic mouse and *C.elegans* models the most prevalent (6)

2.1.2 Model organism: Caenorhabditis elegans

The nematode *C. elegans*, a small roundworm with a short life cycle, has become an essential model organism in biological research, particularly in the study of neurodegenerative diseases (18). This organism shares many molecular pathways and cellular mechanisms with mammals, making it an excellent subject for comparative studies (19). Remarkably, *C. elegans* is the only organism with a completely mapped connectome, offering an unparalleled resource for understanding neuronal networks (20).

The anatomical transparency of *C. elegans* is a significant advantage, allowing researchers to visualize neurons by expressing fluorescent proteins. This feature is important for dynamically studying neuronal properties throughout various developmental stages and over time (21). Leveraging these unique characteristics, researchers can rapidly and cost-effectively pinpoint gene targets and drug candidates that may

mitigate or halt neurodegenerative progression (22).

For instance, studies using *C. elegans* have elucidated the roles of specific genes in Alzheimer's and Parkinson's disease, showcasing the utility in identifying crucial biological pathways involved in neurodegeneration (23). The fully mapped connectome, anatomical transparency, and molecular similarities with mammals, position *C. elegans* as a cornerstone in neurodegenerative research.

2.2 Omics data

Omics technologies have revolutionized the field of systems biology by enabling comprehensive studies of various biological entities and processes. These high-throughput methods generate large amounts of data, analyzing biological samples at different levels, and are essential in understanding the complexity of biological systems (24).

Genomics assesses variability in the DNA sequences in the genome, while transcriptomics focuses on studying the entire set of transcripts or RNA molecules. Proteomics investigates the entire component of proteins, and metabolomics analyzes the metabolite profile. Lastly, epigenomics examines epigenetic modifications, such as DNA methylation and histone modifications (25).

Besides these typical approaches, there are other emerging technologies. Epiomics (epigenomics, epitranscriptomics, and epiproteomics) studies the complex interactions and regulatory networks that influence gene expression and cellular function. Interactomics, on the other hand, offers insights into the dynamic interactions among various biological molecules, while Immunomics combines immunology with omics technologies to study the immune system. At last, microbiomics involves the study of microbial communities in various environments, such as the human body (26).

Although all these methods provide a lot of information, the challenge lies in effectively analyzing, interpreting, and integrating the data. Advanced computational tools, bioinformatics, and systems biology approaches are essential for translating these data into a comprehensive understanding of biological systems and their dynamic behavior. The ultimate goal is to elucidate complex biological networks and their implications in health, disease, and aging.

Transcriptomics

Transcriptomics involves analyzing the full range of RNA transcripts produced by the genome under specific conditions. This field provides an essential overview of gene expression patterns and regulatory mechanisms (27). It is particularly crucial in exploring molecular-level changes during aging and in diseases

such as Alzheimer's and Parkinson's, offering insights into these intricate biological processes (28). Within transcriptomics, methodologies vary, but microarrays and RNA-sequencing (RNA-seq) stand out for their impact and prevalence. Microarrays, an earlier technique in this field, continue to offer a cost-effective means for high-throughput gene expression analysis. Especially useful in well-characterized organisms and conditions, microarrays have been essential in identifying gene expression patterns linked to aging and neurodegenerative diseases (29).

On the other hand, RNA-seq provides a more comprehensive and high-resolution view of the transcriptome. This approach not only enables quantification of gene expression, but also allows for the discovery of new transcripts, isoforms, and single-nucleotide variants. Its application in aging research has been particularly transformative, shedding light on transcriptomics alterations associated with aging and related diseases (30). The advent of single-cell transcriptomics has marked another leap forward, especially in the context of neurological studies. By enabling gene expression analysis at the individual cell level, it offers a window into tissue heterogeneity, a critical aspect in understanding complex diseases and disorders (31).

RNA-Seq is a commonly used technique to identify genes that are differentially expressed under different conditions. Before sequencing, the total number of reads or sequencing depth is standardized. The expression of mRNA transcripts is then determined by the ratio of the total read count or its abundance level. Transcripts with high expression have more mapped reads compared to those with lower expression. Transcripts with longer lengths also tend to have more mapped reads than shorter transcripts with similar expression levels. Therefore, various normalization techniques have been developed for RNA-Seq data to address variations in library size, transcript length, and GC content (32).

One prominent normalization technique is the use of Transcripts Per Million (TPM) or Fragments Per Kilobase Million (FPKM), which accounts for library size and transcript length, providing a normalized measure of gene expression, Reads Per Kilobase Million (RPKM) is often employed to normalize for transcript length (33). For this, several tools, such as SAMtools (34), DESeq2 (35) and edgeR (36), have been designed to process and interpret RNA-Seq data accurately by accounting for various biases and factors that affect the data, ensuring that differentially expressed genes are correctly identified (37).

Given the widespread availability and significance of transcriptomic data, developing robust databases was a natural progression. The Gene Expression Omnibus (GEO) (38) database exemplifies this. As a public repository, GEO archives and freely distributes extensive sets of microarray, next-generation sequencing, and other high-throughput functional genomic data. This database has become an indispensable resource for researchers delving into disease mechanisms, biomarker identification, and therapeutic target discovery. Its rich collection of aging-related datasets is a cornerstone for those investigating the transcriptional

changes accompanying aging and age-related diseases.

A noteworthy dataset is GSE15222, titled "Genetic control of human brain transcript expression in Alzheimer's disease". This dataset includes 363 cortical samples from the post-mortem brains of both Alzheimer's patients and non-affected control subjects, providing vital insights specifically into late-onset Alzheimer's disease (39).

Differential Gene Expression

DGE analysis is a fundamental aspect of transcriptomic, enabling the identification of genes whose expression levels change significantly under different biological conditions (40). This approach is crucial for identifying Differentially Expressed Genes (DEGs) involved in specific biological processes such as aging and neurodegenerative diseases (41).

Several statistical tools have been developed to identify DEGs, including DESeq2 (42), edgeR (43), and limma (44). These tools detect significant expression changes using various statistical models. For instance limma, originally designed for microarray data, has been adapted for RNA-seq data analysis. It uses the voom transformation to convert count data to log2-counts per million (log-CPM) with associated precision weights, allowing linear modeling and empirical Bayes methods for differential expression analysis (45).

Identifying DEGs requires careful selection of statistical thresholds. This involves choosing appropriate cutoffs for p-values or adjusted p-values, such as the false discovery rate (FDR), and for the magnitude of expression change represented by the log2 fold change (log2FC). The choice of these thresholds should be guided by the specific characteristics of the data and the experimental design, as there are no universal cutoff values (46). Combining both the p-value (or adjusted p-value) and log2FC criteria provides greater reliability in DEGs identification, balancing statistical significance with biological relevance (47).

To graphically visualize the DGE results, methods such as MA plots which display the relationship between the log2FC and the mean expression level—and volcano plots—which plot statistical significance against the log2FC are commonly used. These visualizations help in observing trends, patterns, and anomalies in the data, facilitating the overall identification of DEGs and aiding in assessing the quality of the results (48).

Enrichment Analysis

To interpret the results from differential gene expression studies, enrichment analysis methods are employed on the set of DEGs. These computational tools help to identify biological pathways, functions, and

processes that are significantly overrepresented in the set of the gene of interest (40).

Methods such as Gene Set Enrichment Analysis (GSEA) (49) and pathway enrichment analysis (50) play a crucial role in uncovering these changes, which is particularly important in the context of aging and neurodegenerative diseases (51). These methods help identify significantly enriched pathways present in the DEGs associated with specific conditions

Enrichment analysis allows for the identification of potential biomarkers that are overrepresented in a specific biological pathway (52), it helps in validating the biological significance of these candidate genes by analyzing their roles and interactions within the enriched pathways (53). Enrichment analysis can also help link the gene expression changes to relevant diseases, demonstrating their potential as disease markers or drug targets (54), finally, it also enables to obtain a refined list of genes with clinical relevance (55).

For example in the context of aging and neurodegenerative diseases, enrichment analysis has been used to identify biomarkers and pathways associated with several neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. It identified 101 metabolites as potential biomarkers for these conditions, with uric acid, choline, creatine, L-glutamine, alanine, creatinine, and N-acetyl-L-aspartate being common across the three diseases. Key metabolic pathways, such as those involving ATP-binding cassette transporters, arginine, proline metabolism, and alternative energy-producing routes (alanine, aspartate, glutamate, purine metabolism), were highlighted (56). These discoveries offer insights into biomarkers and potential therapeutic targets to address these diseases.

2.3 Genome-Scale Metabolic Models

Genome-scale metabolic models represent a refined computational framework aimed at simulating and predicting the complex network of metabolic interactions within an organism. These models are constructed based on the comprehensive data derived from genome annotation efforts, effectively mapping the full range of an organism's metabolic capabilities (57).

GEMs function under a set of fundamental premises that enable them to model the relevant entities and behaviors of metabolic systems, under some simplifying assumptions. The stoichiometric matrix (S), is a mathematical representation that captures the relationships between metabolites (m) and their corresponding reactions (n) within the network. Each element of this matrix indicates the stoichiometry of a metabolite participating in a particular reaction, thereby defining the network's structure (Equation 2.1).

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix} \tag{2.1}$$

An essential component of GEMs is the pseudo-steady-state assumption, which assumes the concentration of internal metabolites remains constant over time. This assumption simplifies the complexity of metabolic networks by focusing on the fluxes, rather than the absolute concentrations of metabolites. Under this assumption, the system reaches a steady-state condition, where the production and consumption of the metabolites are balanced. Additionally, GEMs rely on a mass balance assumption ensuring that for each biochemical reaction, the sum of the reactant masses equals that of the product masses (Equation 2.2).

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

Flux constraints are applied to bound the possible rates of reactions, reflecting biological capacities among other constraints (Equation 2.3). This ensures that the model's predictions remain within biologically feasible parameters, enhancing the model's reliability and applicability in various biological contexts.

$$\begin{aligned} l_1 \leq v_1 \leq u_1 \\ l_2 \leq v_2 \leq u_2 \\ \vdots \\ l_n \leq v_n \leq u_n \end{aligned} \tag{2.3}$$

Furthermore, GEMs are uniquely structured to incorporate various omics data types, including transcriptomics, enhancing their predictive capacity (58).

Gene-Protein-Reaction Rules

An essential component in GEMs is the integrations of Gene-Protein-Rule (GPR) rules. These rules establish the relationships between genes, the proteins (enzymes) they encode, and the metabolic reactions catalyzed by those enzymes. By defining how genes encode proteins that catalyze specific metabolic reactions, GPR rules enable GEMs to translate genetic information into metabolic functions, thereby bridging the gap between genotype and phenotype (59).

The presence of isoenzymes and protein complexes introduces complexity to the relationships between

genes, proteins, and reactions. Isoenzymes are different enzymes that catalyze the same reaction but are encoded by different genes. This redundancy provides flexibility and robustness in the metabolic network, allowing the organism to maintain metabolic flux under varying conditions. Protein complexes, on the other hand, are enzymes composed of multiple subunits, each encoded by different genes. All subunits must be present and correctly assembled for the enzyme to be functional, emphasizing the importance of gene co-expression and protein-protein interactions. To accurately represent these complex relationships, GPR rules utilize logical expressions involving the operators **AND** and **OR**. These operators are applied at two hierarchical levels: the gene-to-protein level and the protein-to-reaction level. At the gene-to-protein level, the AND operator indicates that multiple genes are required to produce a functional protein or protein complex, while the OR operator represents alternative genes encoding isoenzymes. At the protein-to-reaction level, similar logical operators are used to represent alternative enzymes or the necessity of multiple enzymes for a reaction (60).

For example, in an isoenzyme scenario where a reaction can be catalyzed by enzymes encoded by genes $\boldsymbol{G_1}$ or $\boldsymbol{G_2}$, the GPR rule is expressed as:

$$\boldsymbol{G_1}\ OR\ \boldsymbol{G_2}$$

This expression indicates that the presence of either gene is sufficient for the reaction to occur, reflecting redundancy in the metabolic network.

In a protein complex scenario where a reaction requires a protein complex formed by subunits encoded by genes $\boldsymbol{G_1}$, $\boldsymbol{G_2}$, and $\boldsymbol{G_3}$, the GPR rule is expressed as:

$$\boldsymbol{G_1}\ AND\ \boldsymbol{G_2}\ AND\ \boldsymbol{G_3}$$

This indicates that all the genes are necessary to form the functional enzyme complex for the reaction to proceed, emphasizing the interdependence of gene expression for complex formation.

GPR rules are a fundamental component of GEMs, providing the critical link between the genetic and metabolic layers of biological systems. By accurately representing the intricate relationships between genes, proteins, and reactions through logical expressions, GPR rules enhance the predictive power of metabolic networks. This enables the simulation of genetic perturbations and the assessment of their impact on metabolism, which is essential in fields such as metabolic engineering, drug target identification, and especially in understanding complex disease mechanisms, including aging and neurodegenerative diseases.

2.3.1 Phenotype Simulation

GEMs are essential in simulating phenotypes, offering insights into an organism's metabolic behavior under various environmental conditions. These models integrate constraints that shape the solution space of a metabolic network, ensuring that the predictions align with biological realities (61).

The objective function, which typically reflects a biological goal such as growth maximization, plays a crucial role in this context. It directs the selection process, pinpointing the most biologically pertinent flux distribution from the range of possibilities (Figure 1). Thus, while constraints form the boundaries of possible solutions, the objective function identifies the optimal solution within these boundaries (62). Despite these constraints, the complexity of metabolic networks often results in multiple feasible flux distributions.

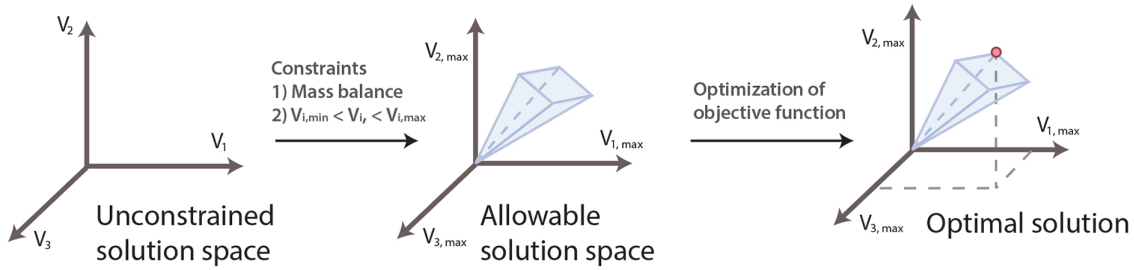


Figure 1: **Conceptual representation of phenotype simulation.** Initially, the solution space is unconstrained, encompassing all possible flux distribution solutions. Application of the steady-state assumption and definition of feasible ranges for metabolic flux further constrain the solution space. Finally, by optimizing the objective function, one of several optimal solutions is obtained. Adapted from (63)

Flux Balance Analysis (FBA) (64) stands as a widely adopted method in this domain. By solving linear equations that represent the stoichiometry of metabolic reactions, FBA predicts flux distributions that fulfill a specific objective function. This function often pertains to biomass production, Adenosine triphosphate (ATP) production, or the synthesis of certain metabolites (65). Nonetheless, the redundancy inherent in metabolic networks can lead FBA to produce multiple potential solutions.

Parsimonious Flux Balance Analysis (pFBA) enhances the FBA approach by imposing an additional optimization criterion: minimizing the total flux across all reactions. This inclusion aligns with the evolutionary principle of resource conservation and metabolic efficiency, suggesting that organisms are optimized to minimize energy and resource expenditure. Consequently, pFBA refines the solution space of FBA, providing predictions of metabolic flux distributions that are more consistent with physiological evidence (66).

Moreover, other algorithms can be used for phenotype prediction, such as MOMA. MOMA is a computational approach used to predict the metabolic behavior of cells in response to genetic mutations. Unlike FBA, which seeks an optimal flux distribution, MOMA assumes that cells attempt to preserve their pre-mutation metabolic state. It calculates the smallest possible deviation from the original flux distribution by minimizing the Euclidean distance between the fluxes of the wild-type and mutant strains. This approach is especially useful for understanding immediate cellular responses to gene knockouts (67).

2.3.2 Flux Sampling Methods

While the previous phenotype simulation methods offer valuable insight into the metabolic network by predicting flux distributions that optimize a given objective function, they typically provide a single optimal distribution. However, since metabolic networks are undetermined systems due to having more reactions than metabolites, leading to the existence of multiple feasible solutions that satisfy the network constraints, this means that several alternative flux distributions are capable of achieving the same optimal value for the objective function (68). Flux sampling methods address this limitation by exploring all the space of possible flux distributions that satisfy the metabolic network constraints (69). By generating a representative set of feasible flux distributions, flux sampling provides a more comprehensive understanding of the metabolic network.

Several algorithms have been developed for flux sampling in metabolic networks, for instance, the Artificial Centering Hit-and-Run (ACHR) algorithm (70) works by estimating in an early phase the center of the solution space and then exploring the full solutions space from there. Furthermore the Optimized General Parallel (OPTGP) algorithm (71) is considered an improved version of the ACHR since allows a more increased randomness when exploring the total solutions space and also allows a parallel sampling process. The Coordinate Hit-and-Run with Rounding (CHRR) algorithm (72) is similar to the ACHR but in this algorithm a pre-processing step occurs to round the search space, allowing for a more efficient sampling process.

In this context of aging and neurodegenerative diseases, flux sampling becomes particularly valuable. Since these conditions often involve minimal metabolic shifts rather than larger changes that a single optimal flux distribution might miss. Exploring the full space of the metabolic network, flux sampling can uncover metabolic adaptations associated with the disease providing deeper insights into the underlying metabolic alterations.

2.3.3 Human Models

Since the publication of the draft human genome in 2001 (73), and the subsequent release of the complete human genome sequences in 2004 (74), the landscape of biological research has undergone a revolutionary transformation. With this, several comprehensive databases like HumanCyc (75) and the Reactome knowledgebase (76) were created. These databases were essential in compiling an extensive collection of biochemical reactions occurring within human cells, needed for advances in metabolic modeling.

The initial effort in GEMs modeling of human metabolism resulted in the development of RECON1 (77). In parallel with this model, the Edinburgh Human Metabolic Network (EHMN) (78) was introduced. RECON 1 not only stood as a comprehensive model in its own right but also served as a template for developing several mammalian models. Based on gene orthologs, these models extended to species such as *Mus musculus* with the iMM1415 (79) model.

Over time, the field has witnessed the evolution of these models with the introduction of improved and more sophisticated versions, notable among these are the Recon2.2 (80) and Recon3D (81), which brought enhanced depth and dimensionality to the modeling of human metabolism. Similarly, HepatoNet1 (82) and the Human Model Reaction 2 (HMR2) database (83) expanded the scope, incorporating more extensive data and refining the accuracy of metabolic simulations. The most recent model, Human1 (84), represents the culmination of these continuous efforts, reflecting a more comprehensive and precise representation of human metabolic processes. An overview of these models is presented in Table 1.

Table 1: Summary of published Human Models

	Metabolites	Reactions	Genes	Reference	Year Published
Recon1	2766	3741	1905	(77)	2007
EHMN	3381	2823	2314	(78)	2007
HMR2	6006	8181	3765	(83)	2014
Recon2.2	5324	7785	1675	(80)	2016
Recon3D	5835	10600	2248	(81)	2018
Human1	8456	12995	2889	(84)	2020

2.3.4 Caenorhabditis elegans models

Several metabolic reconstructions for *C.elegans*, a key model organism for studying biological phenomena relevant to human health and disease, particularly in aging and neurodegenerative diseases, have been developed over the years.

The first reconstruction emerged from the Path2Models (85) project in 2013, employing automatic reconstruction from Kyoto Encyclopedia of Genes and Genomes (KEGG) (86) pathways and MetaCyc (87) databases, complemented by a gap-filling step. This was followed by the development of ElegCyc (88) and iCEL1273 (89) in 2016, and CeCon (90) in 2017, each introducing important improvements in modeling the organism's metabolic pathways.

To integrate different insights from these diverse models, a consensus model WormJam (91) was created, merging different preexisting models. It underwent manual curation and reconciliation, aiming to provide a comprehensive and cohesive representation of *C.elegans* metabolism by combining the most reliable and consistent data from the other models.

More recently, iCEL1314 (92), an updated version of iCEL1273, was released, offering refined metabolic insights and enhanced accuracy in modeling the organism's biochemical networks. This evolution reflects

the deeper understanding of *C.elegans* metabolism and underscores the importance of these models in decoding the complex metabolic processes in this organism. An overview of these models is provided in Table 2.

Table 2: Summary of published *C.elegans* Models

	Metabolites	Reactions	Genes	Reference	Year Published
ElegCyc	1640	1914	979	(88)	2016
iCEL1273	887	1985	1273	(89)	2016
CeCon	1394	1923	20791	(90)	2017
WormJam	1629	3632	1524	(91)	2018
iCEL1314	907	2230	1314	(92)	2020

2.4 Context-specific Metabolic Models reconstruction

The integration of omics data into GEMs plays an essential role in refining phenotype predictions, thereby advancing our comprehension of complex biological processes. Contextualization is important in enhancing the predictive accuracy of GEMs. While generic models provide broad insights, they often lack the context to capture individual-specific variations. In contrast, contextualized models provide a detailed representation of biological processes, crucial for specific treatments and understanding of specific cellular mechanisms (93).

GEMs represent metabolic networks, which are subjected to variability due to the unique characteristics of different tissues, developmental stages, cell types, and environmental conditions. Transcriptomics, with its comprehensive portrayal of gene expression, stands as a primary omics approach in creating context-specific GEMs (94). A notable application of these models is the comparison between healthy and pathological states, such as in cancer research, to deepen our understanding of both conditions. This methodology is essential in identifying key metabolic pathways that undergo alterations in disease states, thereby providing potential therapeutic targets (95).

Regarding the task of creating context-specific models from their generic counterparts, three principal algorithmic approaches have gained relevance (Figure 2): Gene Inactivity Moderated by Metabolism and Expression (GIMME-like), Integrative Metabolic Analysis tool (iMAT-like), and Model-Building Algorithm (MBA-like) methods (96).

The selection of these methods depends on whether a flux distribution is required or if a specific

Required Metabolic Function (RMF) is defined.

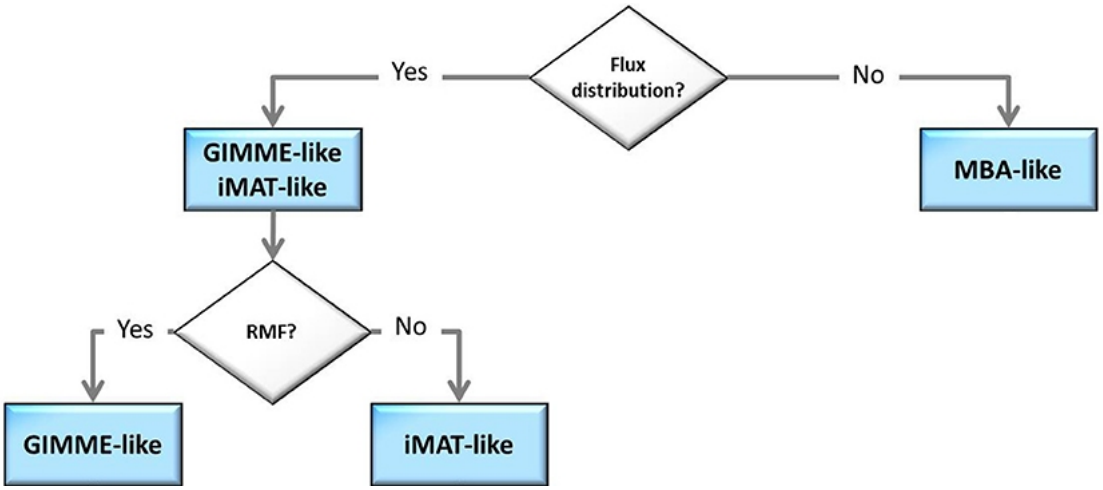


Figure 2: **Representation of the main algorithms for omics data integration.** GIMME-like and iMAT-like methods should be employed if flux distribution is needed; otherwise, MBA-like methods should be used. If the RMF is specified, GIMME-like methods should be used; in the absence of RMF, iMAT-like methods should be utilized. Adapted from (96)

GIMME-like methods prioritize maintaining a correlation between the simulated phenotype with experimental data, ensuring a defined set of active and inactive genes under specific conditions, while maintaining an RMF such as ATP production or growth (97).

In contrast, iMAT-like methods focus on achieving flux consistency with experimental data, without an imposed RMF. These methods label reactions based on gene or protein expression data and aim to maximize the similarity between predicted fluxes and their corresponding classifications (98).

MBA-like methods, on the other hand, focus on model consistency, constructing context-specific models without blocked reactions. They categorize metabolic model reactions into 'core', reactions present in the generic model that should be incorporated into the context-specific model, and 'non-core' refining the model iteratively by removing non-core reactions (99).

Each approach offers a unique view through which omics data can be integrated into GEMs, with their applicability hinging on the specific research objectives and the specific requirements of the biological system under study (100).

In the next section, we will focus only on iMAT-like methods, as they are the best choice to obtain a context-specific model and flux prediction in the absence of an RMF.

iMAT-like methods

The iMAT-like family encompasses the Integrative Metabolic Analysis tool (iMAT), the Integrative Network Inference for Tissues (INIT).A notable feature of these methods is their elimination of the need for RMF, distinguishing them from other approaches.

At the core of the iMAT-like methods is the classification of gene expression levels into three categories: high, moderate, and low. This classification plays an essential role in subsequent computational steps, which aim to minimize the reactions associated with genes expressed at low levels while maximizing those linked to highly expressed genes (98). The optimization problem underlying these methods is formulated as a Mixed Integer Linear Programming (MILP). While this approach enhances precision, it also demands substantial computational resources, particularly for extensive and complex models like those used in human metabolism (101).

- **iMAT** primarily processes gene expression data, categorizing them into low, moderate, or high tiers. Utilizing these categories, it predicts the flux activity state for each reaction in the model, minimizing the number of reactions associated with low-expressed genes while maximizing those associated with high-expressed genes. Thus, providing a comprehensive map of the active metabolic fluxes anticipated by the system (Figure 3).

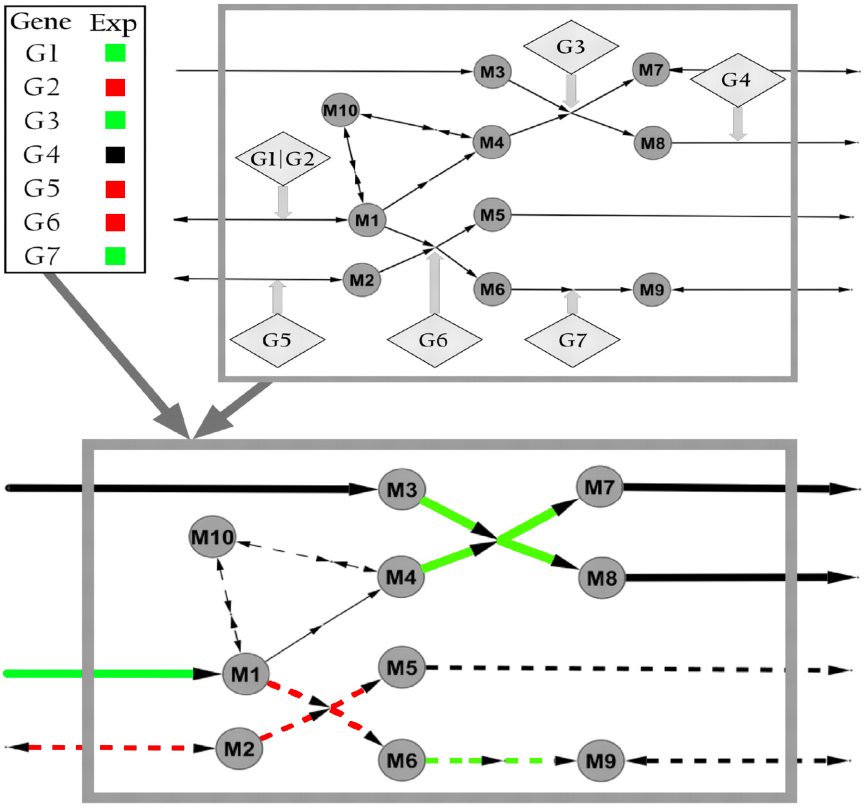


Figure 3: **Schematic representation of iMAT.** The diagram presents a simplified metabolic network model along with gene expression data, with red indicating low-expressed genes (G2, G5, and G6) and green for high-expressed ones (G1, G3, and G7). iMAT produces an optimal flux distribution aligning with the provided gene expression data, where reactions associated with highly expressed genes (green) and lowly expressed genes (red) are highlighted. Adapted from (98).

- **INIT** incorporates data directly into the objective function, including metabolomics data. It diverges from models strictly adhering to steady-state assumptions by allowing a positive net production of metabolites, thus mirroring the dynamic metabolic state under specific conditions more accurately (102).

Thus, the primary advantage of iMAT-like methods is their autonomy from RMFs. This method is particularly advantageous for constructing human metabolic models, where the determination of a universal objective function is not straightforward (96).

2.5 Metabolic Transformation Algorithm

The metabolic transformation algorithm represents a significant leap forward in computational biology, particularly for identifying drug targets in metabolic disorders. Unlike traditional methods that often focus on cell destruction, as in cancer therapies and antibiotics (10; 11), the MTA uniquely identifies drug targets that can shift a cell's metabolism from a diseased state to a healthier one. This feature is critically relevant in aging and age-related diseases, aiming to rejuvenate cells rather than eliminate them (8).

MTA combines GEMs with differential gene expression data from diseased and healthy metabolic states to pinpoint crucial differences. The algorithm follows primarily four steps:

- **Step 1: Obtain a flux description of the source metabolic state.** Initially, the flux description of the source metabolic state based on its gene expression levels is determined using iMAT. Given the potential non-uniqueness of the iMAT solution, there is a space of alternative optimal solutions. To address this, 2,000 flux distributions are sampled and the mean of these distributions was calculated, denoted as v^{ref}, which serves as an approximation of the source's metabolic state (Figure 4: Step 1).

- **Step 2: Determination of changed and unchanged reactions based on $\mathbf{v^{ref}}$, differential gene expression analysis, and gene-to-reaction mappings.** Following this, an analysis of gene expression data from both the source and target is conducted. The Student's t-test is employed to categorize genes into three groups: those with no significant change in expression between the source and target, those with a reduced expression in the target compared to the source, and those with an increased expression in the source compared to the target. After categorizing genes, the focus shifts to reactions. Genes are mapped to reactions, and the reactions are categorized based on their flux direction. Three distinct groups of reactions emerge, those with no significant flux change (R_S), those with a significant change in flux in the forward direction (R_F), and those with a significant change in flux in the backward direction (R_B) (Figure 4: Step 2).

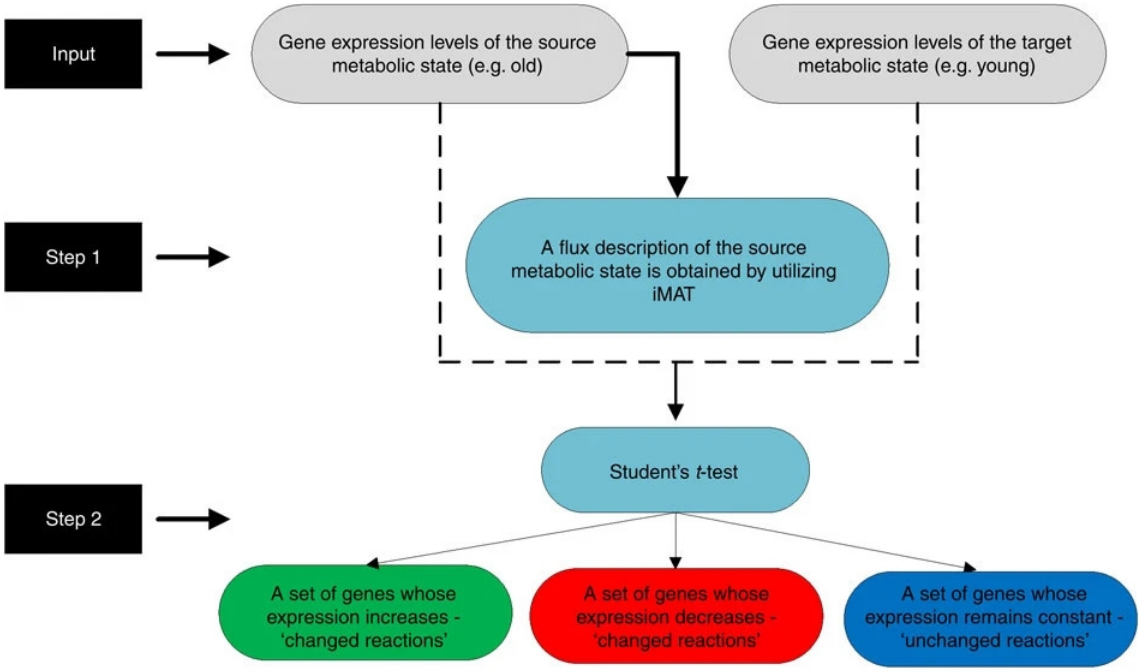


Figure 4: **Schematic representation of MTA: Part 1**. Determination of the flux description of the source metabolic state using iMAT, and determination of changed and unchanged reactions based on v^{ref}, differential gene expression analysis, and gene-to-reaction mappings. Adapted from (8).

- **Step 3: Identification of the best 'transforming' knockout reactions using a Mixed Integer Quadratic Programming (MIQP) algorithm.** To identify the most effective knockout reactions that can cause a network shift from one state to another, each reaction in the network is perturbed individually at the source state by setting their flux to zero. The network's potential to achieve the desired shift to the target state is then evaluated under each perturbation. This step uses MIQP to maximize the number of changes in the changed classes (R_F and R_B) while minimizing the changes in the unchanged class, R_S. This results in a flux distribution denoted as v^{res} (Figure 5: Step 3).

  Equation 2.4 is used to represent the objective function of the MIQP. Each reaction in the equation is uniquely identified by i. Here, v_i^{ref} represents the reference flux, v_i represents the flux of the reaction, and y_i is a binary variable that indicates whether the flux through the reaction is significantly changed or not. The parameter α is a trade-off parameter.

$$\min\ (1-\alpha)\sum_{i\in\mathcal{R}_S}\left(v_i^{\text{ref}}-v_i\right)^2+\frac{\alpha}{2}\sum_{i\in\mathcal{R}_F}y_i+\frac{\alpha}{2}\sum_{i\in\mathcal{R}_B}y_i \tag{2.4}$$

- **Step 4: Calculation of the Transformation Score (TS).** A TS is assigned to each simulated

perturbation in the previous step. This score reflects the efficiency of transforming the source state back to the target state. These perturbations are ranked based on their score (Figure 5: Step 4).

Equation 2.5 serves as the scoring function, with R_{success} and $R_{\text{unsuccess}}$ categorizing reactions based on whether they have successfully achieved the desired change in flux rate or not.

$$\frac{\sum_{i \in \mathcal{R}_{\text{success}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right] - \sum_{i \in \mathcal{R}_{\text{unsuccess}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right]}{1 \sum_{i \in \mathcal{R}_S} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right]} \tag{2.5}$$

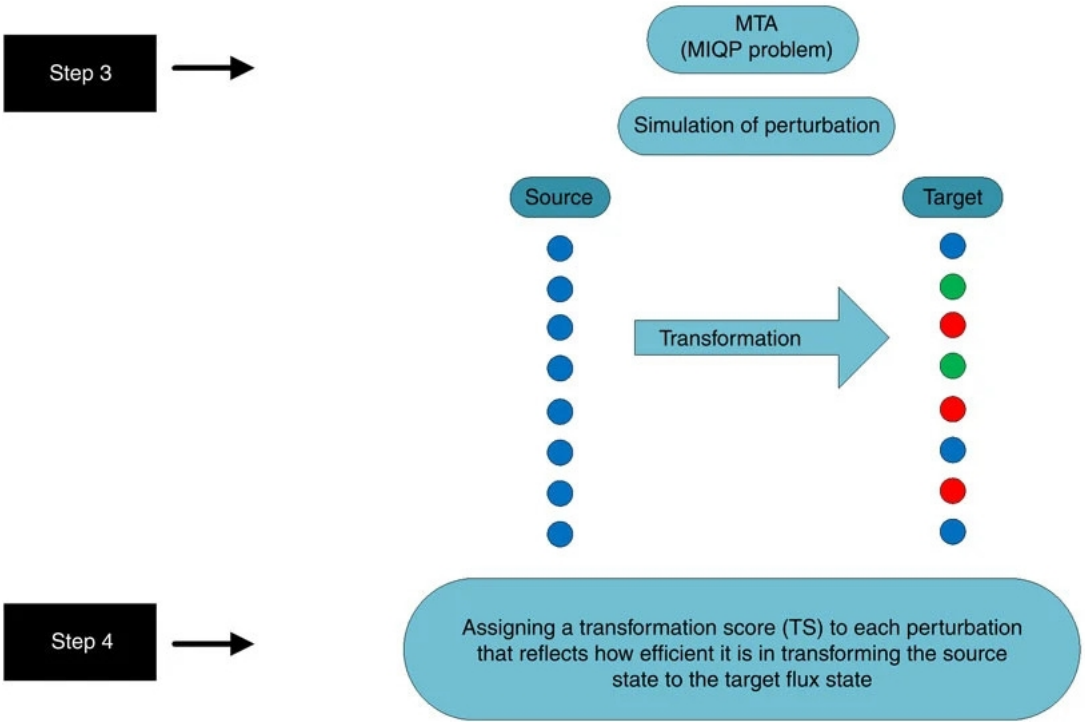


Figure 5: **Schematic representation of MTA: Part 2.** Application of MIQP algorithm to identify optimal knockout reactions by perturbing each reaction individually at the source state and calculation of their corresponding TS. Adapted from (8).

These steps enable MTA to systematically identify genetic or environmental changes capable of reversing a diseased or aged metabolic state to a healthier condition.

The algorithm's effectiveness and predictive precision are validated by various datasets, encompassing genetic and environmental perturbations in organisms like *Escherichia coli*, yeast, mice, and human cells. MTA excels in predicting gene knockout outcomes and recognizing metabolic shifts due to environmental factors, such as nutrient changes in growth media.

A noteworthy achievement of MTA is its contribution to aging research, where it identified and experimentally confirmed genes in yeast (e.g., GRE3 and ADH2) that, when knocked out, significantly extended their lifespan. This predictive power also applies to human biology, with MTA pinpointing potential drug targets to reverse aging in human muscle tissue (8).

Recently, rMTA has emerged as a potent extension of MTA. It introduces worst-case scenario analysis and incorporates the MOMA, enhancing the precision and reliability of identified therapeutic targets. rMTA not only considers the Best-case Transformation Score (bTS) but also integrates the Worst-case Transformation Score (wTS), providing a comprehensive perspective on potential perturbation effects. In the best-case scenario, bTS reflects the maximum capacity of a gene knockout to shift the metabolic flux distribution toward the target state, however, a gene knockout might also reinforce the source state. In the worst-case scenario analysis, wTS is obtained by swapping the roles of the reaction subsets representing the target and source states, assessing this way the maximum capacity to further move the flux distribution in the source direction. Incorporating MOMA into rMTA ensures a realistic evaluation of the metabolic changes required after perturbation, further improving the algorithm's predictive accuracy. To unify these metrics, rMTA introduces a rTS that combines bTS, wTS, and MOMA Transformation Score (mTS). The rTS can be calculated using the following formula:

$$rTS = mTS \cdot |k \cdot |bTS - wTS||^{a}\,, \tag{2.6}$$

The rTS effectively ranks gene knockouts by integrating their capacity to move toward the healthy state (bTS), the risk of moving toward the disease state (wTS), and the most realistic approach (mTS). This approach enhances the predictive accuracy of the algorithm, providing a more reliable identification of therapeutic targets (9).

MTA and rMTA collectively offer a powerful toolkit for drug target discovery in metabolic disorders, providing systematic and robust strategies to address cellular metabolism therapeutically. Their integrated approach, which includes best-case, worst-case, and MOMA-derived scores, confirms that the identified targets are not just theoretically promising but also practically viable. This collaborative application not only promises immense potential in drug discovery but also sets a new standard for pioneering treatments in metabolic disorders, thereby establishing a new benchmark in the field.

2.6 Evolutionary Algorithms

Evolutionary Algorithms (EAs) are computational strategies that draw inspiration from Darwinian principles of natural evolution (103). These algorithms replicate the survival of the fittest, where the most adept individuals in a population are predisposed to survive and reproduce.

EAs excel in solving optimization challenges through this biomimetic approach, representing a diverse family of population-based algorithms. These algorithms utilize principles of genetic inheritance and natural selection to explore search spaces effectively. This population-based approach ensures robust solution generation, proving their efficiency in complex, multifaceted problems (104).

The advent of EAs was initiated by the introduction of the genetic algorithm, distinguished by its binary representation of solutions, use of mutation and recombination techniques to promote population diversity (105), and parent selection mechanisms (106).

The process initiates with the definition of an initial, randomly generated pool of candidate solutions. These candidates are evaluated by a fitness function, and the most promising solutions are chosen stochastically for subsequent iterations. After this selection phase, various operators, such as recombination or mutation, are used to generate new candidates from the parents. This iterative process persists until a pre-defined termination criterion is achieved (Figure 6).

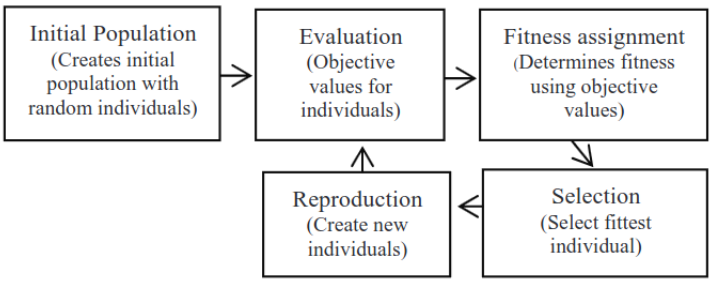


Figure 6: The main steps of an Evolutionary Algorithm. Adapted from (107)

It is crucial to recognize that although EAs aim for optimal solutions, their stochastic nature does not ensure the attainment of the absolute best solution (107)

In practical realms, EAs have been essential in the development of tools for strain optimization and drug discovery. One example is the use of variable-sized EAs, which optimizes metabolic networks by automatically determining the optimal number of gene deletions to enhance production, outperforming traditional fixed-size methods in terms of efficiency and consistency (108). In such applications, individual genes are represented as components of a solution, typically encoded as Boolean arrays. The population

comprises various strains of a singular organism, with simulation methodologies, such as FBA or MOMA utilized as fitness evaluators. These evaluators return an objective function score, correlating with the efficacy of each solution. This tool facilitates a systematic, efficient exploration of the extensive solution spaces characteristic of metabolic networks, using the versatility and applicability of EAs.

Thus, EAs gather the principles of biological evolution to solve complex optimization challenges. Their structured methodology, characterized by a population-based search paradigm, genetic operators, and fitness-centric selection, positions them as formidable tools across various domains.

Part II

Core of the Dissertation

Chapter 3

rMTA/MTA Implementation and Validation

3.1 Introduction

MTA is a computational approach used to identify potential metabolic interventions that can drive a biological system from one state to another by targeting specific genes within a metabolic network. It leverages constraint-based modeling techniques to predict the effects of genetic perturbations on metabolic flux distributions, facilitating the discovery of therapeutic targets and metabolic engineering strategies.

The rMTA extends the capabilities of MTA by incorporating worst-case scenario analyses and integrating the MOMA approach. This enhancement provides a more comprehensive evaluation of potential interventions by considering not only the best-case scenarios, but also the robustness of the metabolic network to perturbations, yielding a more realistic result.

Originally, MTA and rMTA were implemented in MATLAB, a commercial platform, as part of the COBRA Toolbox (109), which offers a wide range of functions for constraint-based metabolic modeling. While MATLAB provides a robust environment for such analyses, the increasing popularity of Python in the scientific community, due to its open-source nature and extensive libraries, has motivated the development of equivalent tools in Python. Python packages such as COBRApy (110) and MEWpy (111) have emerged to offer similar functionalities as their MATLAB counterparts, promoting accessibility and collaboration in the community.

This chapter presents the implementation and validation of MTA and rMTA algorithms in Python. The objectives are twofold:

- **Implementation**: Develop Python versions of the MTA and rMTA algorithms, ensuring that they are functionally equivalent to the original MATLAB implementations.

- **Validation**: Validate the Python implementations by replicating the results obtained from the MATLAB version tutorial. This involves comparing the outputs of key functions, verifying the consistency of optimization results, and ensuring that the algorithms correctly identify potential metabolic targets.

3.1.1 Pipeline for Drug Identification Using MTA

MTA provides a systematic framework for identifying potential drug targets by simulating how genetic interventions can redirect a metabolic network from an undesired state (e.g., disease) to a desired state (e.g., healthy). The key steps in this pipeline are:

1. **Model Reconstruction and Preprocessing:** Omics data, such as transcriptomics, are integrated into a generalized metabolic model to create a context-specific model using methods, already implemented, like iMAT. Flux sampling is performed on this model to establish a Reference Flux Distribution (v^{ref}) that represents the metabolic state of the system.

2. **Identification of Differential Expressed Reactions:** Differential gene expression analysis identifies significantly altered genes between the target and reference states. These genes are mapped onto metabolic reactions using GPR associations. Reactions are then classified into R_F, R_B, and R_S categories based on their expected flux changes (forward, backward, and unchanged).

3. **Optimization of Genetic Perturbations:** An optimization problem is formulated to simulate gene knockouts using MIQP. The objective is to maximize significant flux changes in R_F and R_B while minimizing deviations in R_S. A threshold epsilon (ϵ) defines the minimum significant flux change for reactions to be considered altered. The Gurobi optimizer (112) solves the MIQP problem for each gene knockout scenario.

4. **TS Calculation and Ranking:** The Transformation Score (TS) quantifies the effectiveness of each gene knockout in shifting the metabolic state towards the desired condition. Genes are ranked based on their TS values, highlighting the most promising targets.

5. **Enhanced Analysis with rMTA:** rMTA extends the original algorithm by incorporating worst-case scenario analyses and integrating the MOMA approach. The rTS combines the best-case, worst-case, and MOMA into a unified metric, providing a more robust assessment of potential interventions.

This pipeline enables the systematic identification and prioritization of metabolic genes that, when targeted, can induce a desired metabolic shift. By focusing on system-wide effects rather than isolated pathways, MTA aids in the discovery of effective and selective drug targets.

3.2 Implementation

Both MTA and rMTA were originally implemented in MATLAB within the COBRA Toolbox. To leverage the advantages of Python's open-source nature, and widespread adoption in the scientific community, we re-implemented these algorithms in Python version 3.7.

One of the main challenges in this reimplementation was the lack of support for MIQP problems in Python packages commonly used for metabolic modeling, such as Optlang (113). MIQP is essential for formulating the optimization problems underlying MTA and rMTA. To overcome this limitation, we directly utilized the Gurobi Python API, allowing us to develop and solve the complex optimization problems required. Gurobi is free for academic use, making it accessible for research purposes.

Identification of Differentially Expressed Genes and Reactions

This step focuses on identifying genes and reactions that are differentially expressed between the target and reference states.

`get_differentially_expressed_genes` This function identifies and categorizes differentially expressed genes using omics data previously processed with the limma package (44). Differential gene identification is based on the log-fold change (`logFC`) and adjusted p-values. The first step is to map the gene IDs from the omics data to the corresponding genes in the model.

Using the values of `logFC` and `pval`, the genes are filtered according to specified thresholds. The parameter `logFC_required` represents the minimum absolute log-fold change required for a gene to be considered differentially expressed, and `pval_required` represents the maximum p-value allowed. Genes are then classified as upregulated (`logFC` > 0) or downregulated (`logFC` < 0), assigning values to the genes in the model: 1 for upregulated, -1 for downregulated, and 0 for genes not differentially expressed.

The output of this function consists of a DataFrame (`geneFBS_df`) with each gene and its corresponding value (-1, 0, or 1).

`get_differentially_expressed_reactions_from_genes` This function identifies differentially expressed reactions within a metabolic model by mapping gene expression changes onto reactions using GPR associations. The function takes three inputs: the metabolic model, the DataFrame of differentially expressed genes (`geneFBS`), and a v^{ref}.

To map gene expression changes to reactions, MEWpy's advanced GPR handling capabilities were used. Unlike COBRApy, which lacks sophisticated GPR evaluation similar to the COBRA Toolbox, MEWpy allows for the systematic interpretation of GPR rules in the context of gene expression data.

Using the results obtained from the previous (`geneFBS_df`), the highlighted genes were divided in two groups: upregulated and downregulated genes. MEWpy evaluates the Boolean GPR expressions for each reaction by assigning truth values based on the upregulation or downregulation of genes.

For reversible reactions, further categorization is performed based on the v^{ref}. Reactions that should change in the forward direction R_F are those carrying a positive flux in the reference state and categorized as elevated, or carrying a negative flux and categorized as reduced. Conversely, reactions that should change in the backward direction R_B are determined in the opposite manner.

Ultimately, the function outputs three distinct sets of reactions: unchanged reactions R_S, reactions altered in the forward direction R_F, and reactions altered in the backward direction R_B.

Differential Expression Class

To structure the package, the `DifferentialExpression` class was developed, which encapsulates the previous methods for identifying differentially expressed genes and reactions.

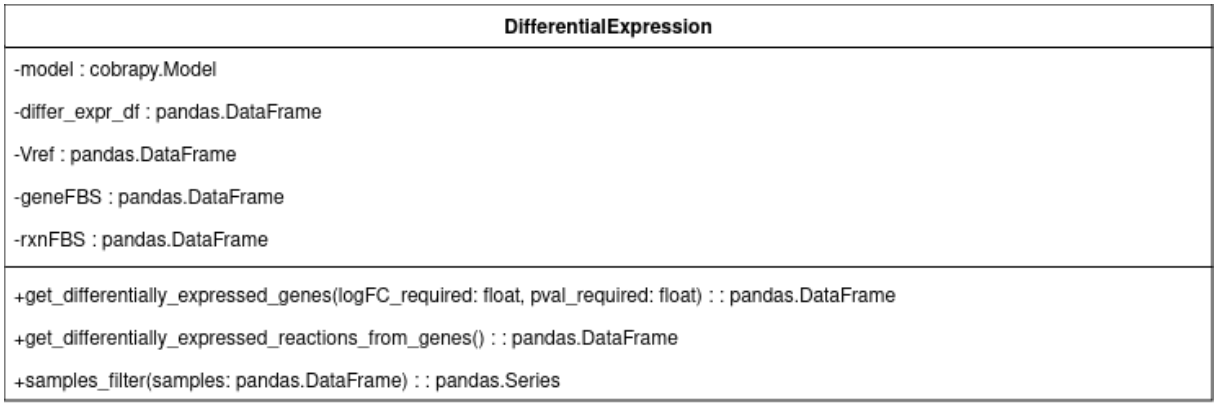


Figure 7: `DifferentialExpression` class structure

Additionally, the `samples_filter` method was added to remove inactive reactions based on the sampling data. This ensures that the list of differentially expressed reactions accurately reflects those that are functionally active under the given conditions.

The general workflow of this class is as follows:

```
diff_expr = DifferentialExpression(model, diff_express_genes, Vref)
diff_expr.get_differentially_expressed_genes()
diff_expr.get_differentially_expressed_reactions_from_genes()
diff_expr.samples_filter(samples)
rxnFBS = diff_expr.rxnFBS
```

Listing 3.1: Differential Expression Analysis Code

Formulation of the Optimization Problem

This is the main step of the MTA algorithm, where the optimization problem is formulated and solved to identify potential targets to revert the undesired phenotype.

`calculate_epsilon` The `calculate_epsilon` function is designed to compute the minimum required flux change, referred to as epsilon (ϵ), for each reaction in the metabolic model to be considered significantly changed (with a p-value of 0.05). This parameter plays a vital role in the optimization model by setting a threshold for significant changes in reaction fluxes. By defining a minimum flux change that must occur for a reaction to be considered altered, ϵ helps to filter out minor, potentially insignificant changes, thus enhancing the robustness and interpretability of the results.

Using the sampled flux distribution data from the metabolic model, the standard error of the mean flux for each reaction was calculated. Additionally, for reactions previously identified as unchanged in `rxnFBS` (with a 0 value), ϵ is set to zero. Hence, ensuring that no change is forced upon reactions that are not supposed to be changed.

`calculate_gene_ko_matrix` The `calculate_gene_ko_matrix` function computes a matrix that maps each gene knockout to the reactions it affects in the metabolic model. This matrix is essential for simulating gene knockouts in the optimization problem by adjusting the bounds of the associated reactions.

`build_mta_problem_from_model` For the optimization problem, an MIQP approach was employed to identify steady-state flux distributions that satisfy stoichiometric and thermodynamic constraints, aiming to characterize metabolic state transitions under different genetic or environmental perturbations. The MIQP formulation, revisited in Equation 2.4, is designed to achieve two primary objectives:

1. **Minimize Deviation in Unchanged Reactions (R_S):** The problem aims to keep the flux

through specific reactions R_S as close as possible to their reference values v_i^{ref}.

2. **Maximize Significant Flux Changes in Desired Reactions (R_F, R_B):** The problem maximizes the number of reactions in R_F and R_B that experience significant flux changes in the desired direction, compared to the reference flux v^{ref}.

The **objective function** for this problem is defined as follows:

$$\min\ (1-\alpha)\sum_{i\in\mathcal{R}_S}\left(v_i^{\text{ref}}-v_i\right)^2+\frac{\alpha}{2}\sum_{i\in\mathcal{R}_F}y_i+\frac{\alpha}{2}\sum_{i\in\mathcal{R}_B}y_i \tag{2.4 revisited}$$

It is important to note that the parameter α is introduced in the multi-objective function. This parameter balances the relative importance of minimizing deviations in unchanged reactions and maximizing significant changes in others. The results presented use $\alpha = 0.66$.

These objectives are subject to several constraints, formulated as follows:

$$v_i^{\text{min}} \leq v_i \leq v_i^{\text{max}}, \quad i \in R, \tag{3.1}$$

Equation 3.1 enforces the bounds on the reaction fluxes v_i within biologically feasible limits, as defined by the minimum v_i^{min} and maximum v_i^{max} allowable values. These values ensure the resulting flux distribution is consistent with thermodynamic, gene knockout, and growth medium constraints.

$$S \cdot v = 0, \tag{3.2}$$

Equation 3.2 represents the mass balance constraint ensuring that each metabolite's net production and consumption is balanced, maintaining a steady-state assumption across the metabolic network. This constraint is expressed through the stoichiometric matrix S, where each entry in the product $S \cdot v$ represents the balance for a specific metabolite.

$$v_i - y_i^F \cdot (v_i^{\text{ref}} + \epsilon_i) - y_i \cdot v_i^{\text{min}} \geq 0, \quad i \in R_F, \tag{3.3}$$

Equation 3.3 enforces that for each reaction i in the set R_F (reactions expected to change in the forward direction), the flux v_i must increase by more than a threshold ϵ_i compared to the reference flux v_i^{ref}. The binary variable y_i^F indicates whether a significant increase in the forward direction has occurred.

$$y_i^F + y_i = 1, \quad i \in R_F, \tag{3.4}$$

Equation 3.4 ensures that for each reaction i in R_F, either y_i^F (indicating a significant increase in flux) or y_i (indicating no significant change) must be active, but not both. This binary relationship enforces mutual exclusivity between the two possible states of the reaction.

$$v_i - y_i^B \cdot (v_i^{\text{ref}} - \epsilon_i) - y_i \cdot v_i^{\text{max}} \leq 0, \quad i \in R_B, \tag{3.5}$$

Equation 3.5 applies to reactions i in R_B (reactions expected to change in the backward direction). It ensures that the flux v_i decreases by more than a threshold ϵ_i compared to v_i^{ref}, with y_i^B as the binary indicator of a significant decrease.

$$y_i^B + y_i = 1, \quad i \in R_B. \tag{3.6}$$

Similar to Equation 3.5, equation 3.6 constraint ensures mutual exclusivity for backward reactions i in R_B. Either the flux significantly decreases ($y_i^B = 1$) or it does not change significantly ($y_i = 1$).

`mta_miqp` To solve the optimization problem, we utilized the Gurobi Python API (Gurobipy). The `mta_miqp` function was developed to account for specific reaction knockouts and return the optimal flux distribution. This required restructuring the original optimization model, initially designed for the MATLAB Gurobi API, to be compatible with Python.

The function begins by unpacking key components of the optimization model, including the left-hand side (LHS) constraint matrix, the quadratic matrix, the right-hand side (RHS) vector, linear coefficients, variable bounds (upper and lower), constraint senses, and the objective sense. This breakdown allows the model to be reformulated with the Gurobi solver.

Once the model components are unpacked, the specific reaction knockouts are imposed by setting the lower and upper bounds of the corresponding reactions to zero, effectively eliminating flux through those reactions. After modifying the model, solver parameters are defined to ensure precise and robust optimization. These parameters include settings for MIP gap, integer feasibility tolerance, and optimality tolerance.

The optimization problem is then formulated and executed within the Gurobipy environment. Upon completion, the solver returns the optimal flux distribution for the specified metabolic network instance. If an optimal solution is found, it is returned as the final result. Otherwise, a fallback mechanism returns a default vector of zeros, indicating failure to converge or an infeasible solution.

`mta_ts` To assess the impact of optimization results on metabolic flux distributions, we developed a function that computes the TS, which quantifies the deviation of a given solution from a reference state. The TS quantifies flux changes induced by the knockouts, focusing on whether reactions achieve desired directional changes in flux (forward or backward) and the extent of those changes. The function calculates the TS based on the flux distribution solution of the optimization model (v^{res}), a v^{ref}, and a set of reactions classified according to their directional shifts (`rxnFBS`).

The first step of the TS calculation is the identification of reactions within each category. The success of a transformation is then determined based on whether the reaction in R_{success} achieves the required flux alterations, or not, in this case being considered $R_{\text{unsuccess}}$.

Lastly, the TS is computed using equation 2.5, the numerator represents the net change in flux rates after a perturbation. This can be split in 2 mains parts: The first one is the sum of absolute flux changes for all reactions in the **Rsuccess** set, which includes reactions that experienced flux alterations in the desired direction. The larger this sum, the more successful the transformation in the correct direction. The second one is the sum of absolute flux changes for reactions in the **Runsuccess** set, which includes reactions that experienced flux alterations in the undesired direction. Subtracting this value penalizes changes in the wrong direction. Together, the numerator balances the positive effects (desired flux changes) against the negative effects (undesired flux changes). The denominator represents the total flux change in R_S, which includes the reactions that should ideally remain unchanged. This value helps to normalize the TS score, ensuring that the transformation is not only effective but also conservative (i.e., minimizing unnecessary changes in reactions that should stay unaltered).

$$\frac{\sum_{i \in R_{\text{success}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right] - \sum_{i \in R_{\text{unsuccess}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right]}{\sum_{i \in R_S} \text{abs}\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)} \qquad \text{(2.5 revisited)}$$

`evaluate_KO` This function, which was not included in the original MATLAB implementation, enables users to evaluate a specific knockout scenario without the need to execute the entire MTA algorithm. By providing a set of reactions linked to a particular gene, this method allows for the calculation of the TS for the specified case. This approach facilitates a partial execution of the MTA algorithm, optimizing the computational process.

Implementation of the MTA Algorithm

The `mta` function is the central component of the MTA, orchestrating the integration of all prior computations to identify gene knockouts that can shift the metabolic network toward a desired state. This function systematically evaluates each gene in the metabolic model, simulates its knockout, and assesses the impact on the network using the TS.

Inputs:

- Metabolic Model
- **`rxnFBS`**: A classification of reactions based on desired flux shifts—forward (R_F), backward (R_B), or unchanged (R_S).
- **`Vref`**: A reference flux distribution representing the baseline metabolic state.
- α: A weighting parameter balancing the objectives in the MIQP optimization problem.
- ϵ: A vector of minimum significant flux changes for reactions to be considered altered.

Overview of the Workflow The workflow of the `mta` function can be summarized in the following key steps:

1. **Preparation Phase**:
    - **Gene Knockout Matrix Creation**: Generate a matrix mapping gene deletions to their associated reactions.
    - **Result Initialization**: Prepare data structures to store flux distributions and TS values.
    - **Reaction Filtering**: Exclude reactions where desired flux changes are infeasible.
2. **Construction of the Optimization Model**:
    - Build the general MIQP optimization problem incorporating stoichiometric constraints, objectives, and the v^{ref}.
3. **Iterative Gene Knockout Simulation**:
    - For each gene in the model:
        - Simulate the gene knockout by adjusting reaction bounds.

- Solve the modified optimization problem.
- Calculate the TS for the resulting flux distribution.

4. **Results and Ranking**:
    - Aggregate TS values and flux distributions for all gene knockouts.
    - Rank genes based on their TS values to identify top targets.
5. **Output Generation**:
    - Provide a ranked list of potential gene targets.
    - Offer detailed flux distributions for further analysis.

The `MTA` class encapsulates the entire algorithm, providing a structured and modular implementation.

MTA
-model (cobrapy.Model): The metabolic model -rxnFBS (pandas.DataFrame): DataFrame containing the reactions names and the respective values. -Vref (pandas.DataFrame): DataFrame containing the reference flux distribution - alpha (float): Parameter for the quadratic problem. -samples (pandas.DataFrame): DataFrame containing the fluxed samples. -epsilon (pandas Series): Series containing the minimum disturbance for each reaction
+ __init__(model, rxnFBS, Vref, alpha, samples=None, epsilon=None) + calculate_gene_KO_matrix(self) + calculate_epsilon(self, samples) + build_optimization_model(self) + solve_MIQP(self, KOrxn) + transformation_score(self, v_res) + run(self) + evaluate_KO(self, KOrxn)

Figure 8: `MTA` class

Implementation of the rMTA Algorithm

The rMTA extends the standard MTA by incorporating worst-case scenario analyses and the MOMA approach. While the overall workflow of both share similarities, key differences enhance its ability to identify more robust metabolic tagets. The sections below will focus on these main differences and explain how they are integrated into this implementation.

Incorporation of Worst-Case Scenario In addition to the best-case scenario analysis performed in MTA, rMTA introduces a worst-case scenario by swapping the reactions in the forward reactions (R_F) and backward reactions (R_B) sets within the `rxnFBS`. This swapping simulates the possibility that reactions may shift in the opposite direction of what is desired, allowing for the evaluation of the worst possible outcome for each gene knockout.

Integration of MOMA rMTA integrates the MOMA approach to simulate gene knockouts, providing a more realistic prediction. MOMA works by minimizing the Euclidean distance between the flux distribution after a perturbation and the v^{ref}, which reflects the cell's natural tendency to adjust minimally from its original metabolic state.

Calculation of the Robust Transformation Score (rTS) The robust Transformation Score (rTS) is introduced in rMTA as a metric that combines the best-case TS (bTS), worst-case TS (wTS), and MOMA TS (mTS) into a single measure. This score is designed to highlight gene targets that not only perform well in the best-case scenario but also demonstrate significant differences from the worst-case scenario, reflecting robustness. The rTS is calculated using the following formula:

$$rTS = mTS \cdot |k \cdot |bTS - wTS||^a \,, \qquad \text{(2.6 revisited)}$$

Following the parameter values used in the referenced study, we set $k = 100$ and $a = 1$ when $bTS > 0$, $mTS > 0$, and $wTS < 0$.

rMTA
-model (cobrapy.Model): The metabolic model -rxnFBS (pandas.DataFrame): DataFrame containing the reactions names and the respective values. -Vref (pandas.DataFrame): DataFrame containing the reference flux distribution - alpha (float): Parameter for the quadratic problem. -samples (pandas.DataFrame): DataFrame containing the fluxed samples. -epsilon (pandas Series): Series containing the minimum disturbance for each reaction
+ __init__(model, rxnFBS, Vref, alpha, samples=None, epsilon=None) + calculate_gene_KO_matrix(self) + calculate_epsilon(self, samples) + build_optimization_model(self, direction=True) + solve_MIQP(self, KOrxn, direction=None) + transformation_score(self, v_res, direction=None) + run(self) + evaluate_KO(self, KOrxn)

Figure 9: `rMTA` class

3.3 Pipeline Validation

To validate the MTA and rMTA pipelines implemented in Python, the validation pipelines originally implemented in MATLAB by the authors of this algorithm were replicated and compared (9). This process involved three key steps: preprocessing and integration of the omics data, DGE, and finally, the MTA and rMTA algorithms themselves.

3.3.1 RRM1 Validation Study

This validation study focuses on the role of the RRM1 (ribonucleotide reductase catalytic subunit M1) gene. Using data from a controlled gene knockout experiment, this study aimed to predict how the knockdown of RRM1 drives the metabolic state from the wild-type (source state) to the RRM1-knockout (target state). This study is significant because elevated RRM1 expression has been associated with poor prognosis in multiple myeloma patients, and its knockdown has been shown to inhibit tumor growth and induce apoptosis (114).

Step 1: Preprocessing, Omics Integration, and Sampling

In the original study, the initial steps involved preprocessing the metabolic model, integrating omics data and performing flux sampling to establish a reference flux distribution. The authors used the Recon1 metabolic model (77) with the RPMI-1640 growth medium, commonly used for culturing human cells. Furthermore, blocked reactions and orphan metabolites were removed using the COBRA toolbox (109) to

eliminate unnecessary reactions, thereby decreasing computation time. As for the gene expression data, a dataset, GSE9345, comprised of Scramble control and siRRM1 knockdown samples was obtained from the GEO database. This data was processed using the limma (44) package, and genes were categorized into either high, medium, or low according to expression values, as required by the iMAT algorithm (98).

The discretized gene expression data was mapped from genes to reactions in the metabolic model using GPR associations. The iMAT algorithm was then applied to integrate the expression data into the metabolic model, generating a tissue-specific reconstruction that reflects the metabolic state under study. To capture the variability of possible metabolic states, flux sampling was performed on the reconstructed model using the ACHR method (70). A total of 2000 flux distributions were generated, providing a comprehensive representation of the feasible solution space. The mean of these sampled fluxes was calculated to obtain a v^{ref}, which serves as a baseline for subsequent analyses.

Due to the stochastic nature of flux sampling and the robust implementations of these methods in existing tools like COBRApy and Troppo, we did not replicate this step in our Python implementation. Instead, the precomputed v^{ref} data provided in the COBRA Toolbox tutorial was used.

Hence, by using the same inputs as the authors, the goal was to ensure that different results were uniquely related to the implemented methods.

Step 2: Identification of Differentially Expressed Genes and Reactions

The second step involves performing DGE and mapping differentially expressed genes to corresponding reactions. Since DGE was not implemented in this pipeline, results from the original implementation were reused. These results include key metrics such as log-fold change (`logFC`) and adjusted p-values (`pval`), which provide insights into both the magnitude of gene expression changes and their statistical significance.

After obtaining the DGE results, genes needed to be classified into either Downregulated, Upregulated or Normal expression. To do so, the `get_differentially_expressed_genes` method was used. Thresholds for `logFC` and `pval` were selected as in the original publication and used to define genes as either upregulated or downregulated. This results of this step, and subsequent comparison with the MATLAB `diffexprs2rxnFBS` method, are presented in Table 3. As shown, both implementations identified the same number of genes in each category. Moreover, the genes in each category were also the same, ensuring the accuracy of this method so far.

	Python	**MATLAB**
Normal expression	1701	1701
Downregulated	112	112
Upregulated	92	92

Table 3: Comparison of gene classifications between Python and MATLAB implementations.

Next, we mapped the differentially expressed genes onto reactions in the metabolic model using the `get_differentially_expressed_reactions_from_genes` method. The comparison between the results of both implementations showed no differences on the classification of each reaction (Table 4).

	Python	**MATLAB**
Normal expression	1754	1754
Downregulated	86	86
Upregulated	73	73

Table 4: Comparison of reaction classifications between Python and MATLAB implementations.

An additional filtering step was carried out to remove inactive reactions—those unable to carry flux under the given conditions. Although the MATLAB pipeline lacked a dedicated function for this, we employed the `samples_filter` method in Python. The curated reaction classifications are compared in Table 5. These consistent results further support the effectiveness of our filtering step and confirm the overall accuracy of the DifferentialExpression class in our implementation.

	Python	**MATLAB**
Normal expression	1837	1837
Downregulated	39	39
Upregulated	37	37

Table 5: Comparison of curated reaction classifications between Python and MATLAB implementations.

Step 3: Running the rMTA Algorithm

To validate the rMTA algorithm, the individual functions within our implementation were first verified. The `calculate_epsilon` is used to determine epsilon (ϵ) values. These represent the minimum significant flux change for reactions to be considered altered. Since epsilon is only calculated for differentially

expressed reactions, we expected a total of 76 non-zero values (39 downregulated reactions and 37 upregulated reactions). We compared the epsilon values from our Python implementation with those from the MATLAB version. As expected, the number of non-zero epsilon values matched (Table 6), and a correlation plot (Figure 10) confirmed a strong linear relationship between the values from both implementations.

Metric	**Python**	**MATLAB**
Zero values	1837	1837
Non-zero values	76	76

Table 6: Comparison of epsilon values between Python and MATLAB implementations.

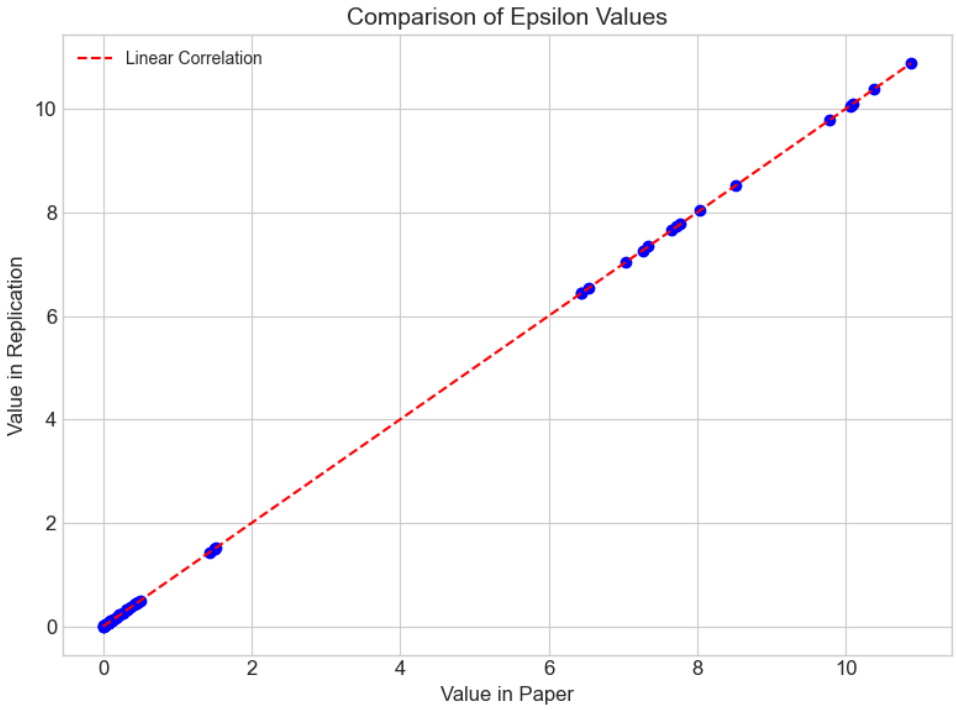


Figure 10: Correlation of epsilon values between Python and MATLAB implementations.

The `calculate_gene_ko_matrix` function was validated by ensuring that the shape of the resultant matrix corresponded correctly to the number of genes and reactions in the model. Furthermore, reactions associated with each gene were also confirmed to match between the Python and MATLAB implementations.

Functions such as `build_mta_problem_from_model`, `mta_miqp`, and `mta_ts` are essential to the algorithm's performance. Their validation is inherently tied to the overall results. Initially, ranking positions of the *RRM1* gene were compared to the results from the authors of the original rMTA implementation. The results are summarized in Table 7. Our replication yielded a significantly higher ranking

for *RRM1* in the MTA analysis (305th) compared to the original study (353rd). In the rMTA analysis, the rankings were more consistent, with *RRM1* placed at 28th in the original study and 23rd in our replication. This improvement suggests that our implementation may provide a more accurate prediction of *RRM1*'s influence on the metabolic network.

Gene name	**MTA Paper**	**MTA Replication**	**rMTA Paper**	**rMTA Replcation**
RRM1	353	305	28	23

Table 7: Ranking positions of *RRM1* in the original study and our replication.

To further assess the consistency between our replication and the original study, we compared the top 10 ranked genes identified by both analyses (Table 8). While there are differences in the exact rankings and some variation in the specific genes listed, several key genes are common to both lists. Despite this, there is notable overlap in the identified genes, indicating that the python replication captures similar metabolic targets.

Ranking	**Paper Gene**	**Replication Gene**
1	PYCR1	PRODH
2	PRODH	SUCLG2
3	SUCLG2	SUCLG1
4	SUCLG1	LDHA
5	LDHA	SUCLA2
6	SUCLA2	SLC16A1
7	SLC16A1	NT5C3A
8	NT5C3A	SLC7A5
9	MTMR3	RDH5
10	SLC7A5	APRT

Table 8: Top 10 ranked genes from the original study and our replication for *RRM1*.

To quantify this overlap, a Venn diagram was generated (Figure 11), illustrating the number of unique and shared genes between the two analyses.

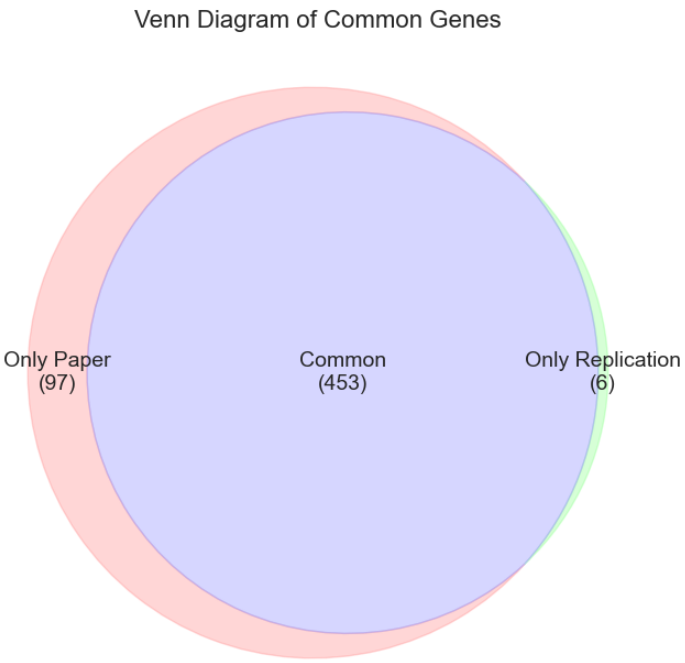


Figure 11: Venn diagram showing the overlap of top-ranked genes between the original study and our replication for *RRM1*.

In fact, on a total of 550 genes, 453 were common between both analyses, suggesting a significant level of consistency. The genes that were not common between the two analyses were excluded to allow a direct comparison of gene position rankings. Table 9 and 10 details the common genes and their rankings.

Gene name	**MTA Paper**	**MTA Replication**	**rMTA Paper**	**rMTA Replcation**
RRM1	299	303	22	23

Table 9: Ranking positions of *RRM1* in the original study and our replication with only common genes.

Ranking	Paper Gene	Replication Gene
1	PRODH	PRODH
2	SUCLG2	SUCLG2
3	SUCLG1	SUCLG1
4	LDHA	LDHA
5	SUCLA2	SUCLA2
6	SLC16A1	SLC16A1
7	NT5C3A	NT5C3A
8	SLC7A5	SLC7A5
9	RDH5	RDH5
10	APRT	APRT

Table 10: Common top-ranked genes between the original study and our replication for *RRM1*.

To evaluate the correlation between gene rankings in the original study and our replication, the rankings of common genes were plotted (Figure 12). The strong correlation at the extremes (highest and lowest ranks) indicates that both analyses agree on the most and least influential genes. However, discrepancies in mid-range rankings are apparent, likely due to small differences in TS and solver sensitivity limitations.

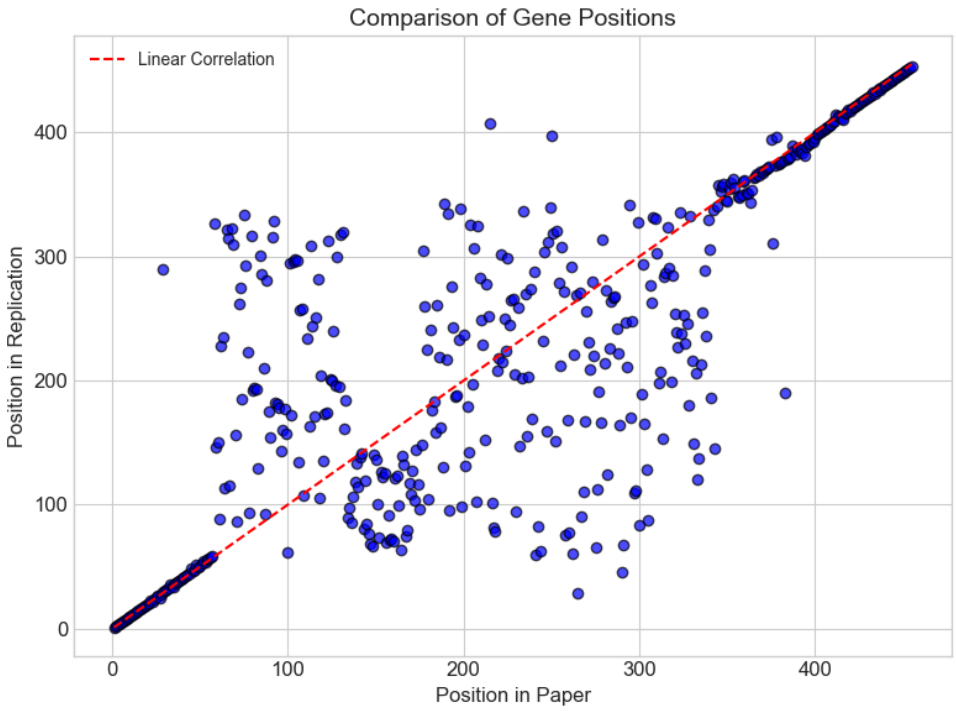


Figure 12: Correlation plot of gene rankings between the original study and our replication for *RRM1*.

rTS values obtained from both analyses were also compared (Figure 13). The similarity in the distribution of rTS values supports the conclusion that the python implementation closely replicates the original algorithm's performance.

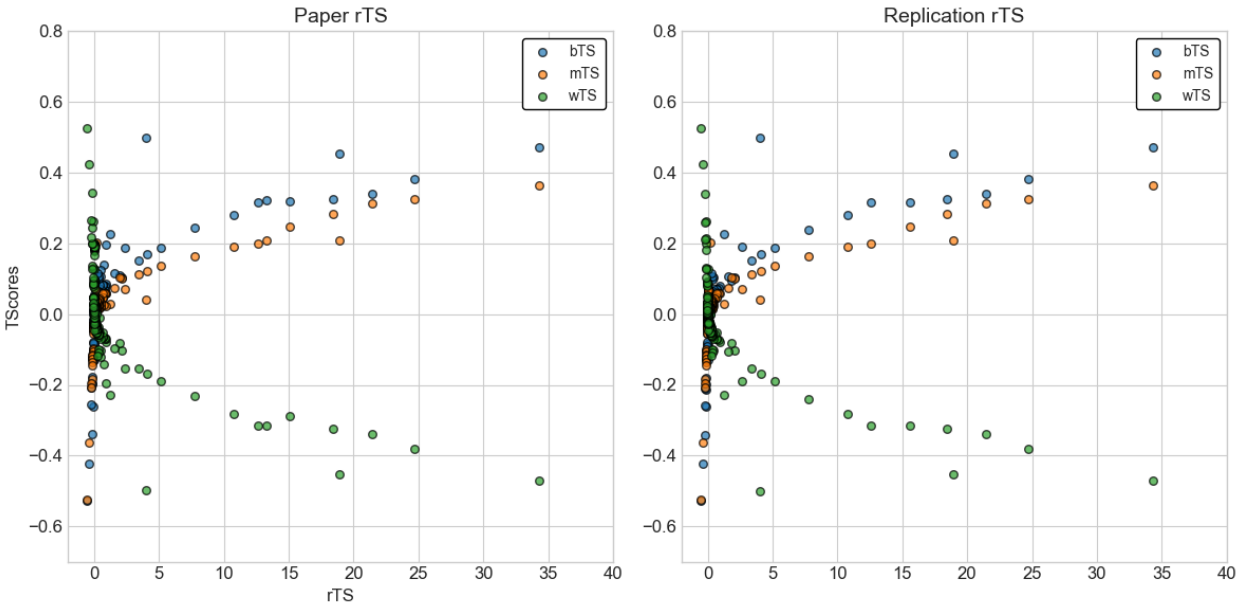


Figure 13: Comparison of rTS values between the original study and our replication for *RRM1*.

The distribution of rTS values (Figure 14) shows that many genes have scores near zero, indicating minimal impact on the metabolic state when knocked out. This clustering makes it challenging to rank genes with very similar rTS values precisely, explaining some discrepancies in mid-range rankings.

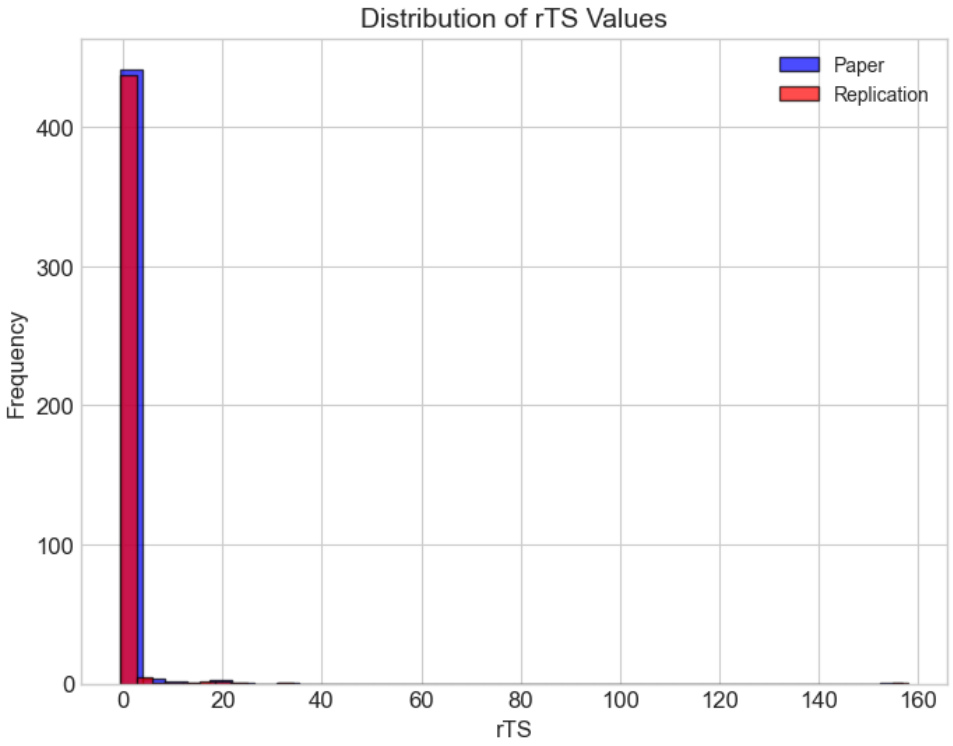


Figure 14: Distribution of rTS values for *RRM1*.

Overall, this replication demonstrates that the Python implementation of the rMTA algorithm is reliable and produces results consistent with the original MATLAB-based study. The high degree of overlap in top-ranked genes and the similarity in rTS distributions confirm that our implementation effectively identifies key metabolic targets, including *RRM1*.

3.3.2 RRM2 Validation Study

While the primary focus of our validation was on the *RRM1* gene, we conducted a similar comprehensive analysis for the *RRM2* gene to further validate our Python implementation of the rMTA algorithm. The results of the *RRM2* analysis are consistent with those obtained for *RRM1*, reinforcing the reliability and accuracy of our approach. The detailed findings for *RRM2*, including rankings, correlation plots, and rTS value distributions, are provided in the supplementary material.

The ranking positions from both the original study and the Python replication are presented in Table 11. In the MTA analysis, *RRM2* ranked similarly low in both approaches. However, in the rMTA analysis, the rankings improved significantly, with *RRM2* placed at 36^{th} in the original study and 32^{nd} in our replication. This consistency indicates that this implementation captures the metabolic impact of *RRM2*.

Comparing the top 10 genes identified in both analyses (Table 12), a considerable overlap was ob-

Gene name	MTA Paper	MTA Replication	rMTA Paper	rMTA Replcation
RRM2	372	357	36	32

Table 11: Ranking positions of *RRM2* in the original study and our replication.

served, reinforcing the reliability of our replication. While some genes appear in different rankings, several key genes such as *SLC34A3*, *SLC6A7*, *AK3*, *RDH5*, *SLC36A1*, *SLC38A5*, *SLC5A2*, *SLC16A1*, and *LDHA* are common to both analyses. This overlap underscores the consistency of our implementation in identifying significant metabolic targets.

Ranking	Paper Gene	Replication Gene
1	SLC34A3	SLC6A7
2	SLC6A7	SLC34A3
3	AK3	SLC38A5
4	RDH5	AK3
5	SLC36A1	SLC36A1
6	SLC38A5	RDH5
7	SLC5A2	SLC5A2
8	SLC16A1	SLC16A1
9	LDHA	LDHA
10	SLC43A2	SLC25A11

Table 12: Top 10 ranked genes from the original study and our replication for *RRM2*.

3.3.3 Discussion

The results from both case studies confirm that the Python implementation of the MTA and rMTA algorithms successfully replicates the findings of the original MATLAB-based study. The consistent identification of key genes and the alignment of ranking positions demonstrate its robustness and reliability. However, discrepancies were found in the rankings of both approaches. These could be related to two main factors:

- **Solver Sensitivity:** Small numerical differences between the optimization solvers used in MATLAB and Python can lead to variations in TS values, especially for genes with minimal impact.
- **Parameter Settings:** Slight differences in parameter configurations, such as tolerance levels or iteration limits, may influence optimization outcomes.

Nevertheless, the overall agreement between the analyses suggests that our implementation is valid and does not compromise the effectiveness of the original approach.

By successfully replicating the key findings, it was demonstrated that these Python-based tools can serve as reliable alternatives to the original MATLAB implementations. This enhances accessibility due to Python's open-source nature and facilitates broader collaboration within the scientific community.

Chapter 4

Case Study: aging research using Caenorhabditis elegans model

4.1 Introduction

This thesis aims to apply the rMTA to aging-related scenarios in *C. elegans*. One of the main limitations highlighted by the authors of MTA was the lack of comprehensive *C. elegans* metabolic models. However, several metabolic models have been developed and validated since then. For this case study, data from GEO series GSE39574 was used. This data was used to study the role of the developmental regulator *unc-62*/Homothorax in limiting longevity in *C. elegans* (115) using a gene knockdown approach where the gene knockout increased the lifespan. unc-62 encodes a HOX co-factor known as Homothorax (Hth) or Meis, playing essential roles in development and aging (116). Notably, reducing unc-62 expression in the intestine and hypodermis extends the median lifespan by approximately 30–40%, even when RNA interference (RNAi) is initiated in adulthood (117). This lifespan extension is linked to decreased expression of yolk protein genes (vitellogenins), reducing toxic protein aggregation in aging worms (115).

The main objective is to apply the rMTA algorithm to this case to uncover the underlying metabolic changes associated with aging processes occurring due to *unc-62* knockdown. The rMTA requires several preprocessing steps. Initially, the reference flux distribution of the source conditions, in this scenario, the *unc-62* knockdown state was obtained. Using the limma package (44), the omics data was processed to identify differentially expressed genes for the pipeline. The rMTA algorithm was run and performed post-processing using BioMart (118) to gather more information about the results. Finally, the top-scoring genes identified by the rMTA were analyzed.

4.2 Methods

4.2.1 Calculation of Reference Flux of the unc-62 Knockdown State

Model Preprocessing

The *C. elegans* metabolic model iCEL1314 was utilized for this study. To simplify and refine the model, blocked reactions were identified using the `find_blocked_reactions` function from the COBRApy package and subsequently removed. This process yielded a consistent model suitable for further analysis. The differences between the original and consistent models are summarized in Table 13.

Model Version	Metabolites	Reactions	Genes
Original Model	1,533	2,230	1,314
Consistent Model	1,353	1,936	1,271

Table 13: Comparison of original and consistent model characteristics.

Omics Dataset

To obtain the reference flux distribution of the *unc-62* knockdown state, GEO series GSE39574 was used. Specifically, samples GSM971873, GSM971874, and GSM971875 were analyzed, corresponding to *unc-62* RNA interference (RNAi) treatments in biological replicates 1, 2, and 3, respectively. The original study demonstrated that knockdown of *unc-62* in adult worms, resulted in a significant increase in lifespan, suggesting that *unc-62* may be a potential target for anti-aging interventions. Using the rMTA algorithm, we aimed to identify other metabolic changes occurring in response to *unc-62* knockdown.

Integration and Sampling

Given the biological replicates, the mean value for each Transcript ID across the replicates was calculated. Next, the omics data was mapped to the respective genes in the model, of a total of 25532 Transcript IDs only 1271 were mapped to genes in the metabolic model. For omics integration on the model, Troppo was used, using iMAT algorithm the context-specific model for the *unc-62* knockdown condition was generated. After obtaining the context-specific model using the iMAT algorithm, flux sampling was performed using the COBRApy ACHR sampler, generating 2,000 flux samples. This sampling provided a distribution of feasible fluxes under the *unc-62* knockdown condition, which served as the reference flux distribution for

the rMTA analysis.

4.2.2 Calculation of Differential Expressed Genes

Differential gene expression analysis was performed using the limma package on the omics dataset. We compared the expression levels between control worms and *unc-62* knockdown worms to identify genes that were significantly differentially expressed, as needed for the rMTA more specifically for the TS calculation.

To identify significant genes, we applied a filtering criterion based on the adjusted p-value. Similar to the approach made in the MTA paper, but to obtain the recommended number of altered reactions for the rMTA around 100 -200 changed reactions, the threshold was modified from an adjusted p-value < 0.05 to a p-value < 0.1.

4.2.3 rMTA

For the rMTA analysis, the trade-off parameter α was defined as 0.66 similar to the MTA paper, and our previous results. This value of α provides a suitable trade-off that allows the algorithm to identify genes whose perturbation would most significantly transform the metabolic state from the control to the *unc-62* knockdown condition. We then executed the rMTA algorithm, which systematically evaluates each gene in the metabolic model, simulates its knockout, and assesses the impact on the network using the rTS. The algorithm ranks genes based on their rTS values, highlighting the most promising targets for further investigation.

4.2.4 Enrichment Analysis

To identify enriched biological pathways and functional annotations associated with the results, an enrichment analysis was performed using the GSEApy package (119), specifically the Enrichr module. All gene set libraries associated with the organism *C. elegans* were selected for this analysis. This included KEGG pathways (120) and other curated databases relevant to *C. elegans* biology, ensuring a comprehensive assessment of potential biological functions and pathways. Based on the ranking provided by the rMTA, the top genes were focused on for this enrichment analysis. These genes were selected according to their rTS scores, allowing the selection of the most significant ones. Finally, the dot plot function from GSEApy was used to visualize the enrichment results. This visualization highlighted the most enriched pathways based on the adjusted p-value, facilitating the interpretation of the analysis.

4.3 Results and Discussion

4.3.1 Differential Expressed Genes

The differential gene expression analysis resulted in the identification of 129 differentially expressed genes. Among those, 74 genes were significantly upregulated, while 55 genes were significantly downregulated between the control and the *unc-62* knockdown worms as shown in Figure 15. Figure 15 also illustrates the distribution of gene expression changes based on their fold change and statistical significance. The plot highlights the genes that exhibit the most substantial changes, serving as a visual representation of the overall impact of *unc-62* knockdown on the gene expression. These differentially expressed genes provide valuable insights on the metabolic network alterations induced by the *unc-62* knockdown, thus ensuring that the rMTA can be performed.

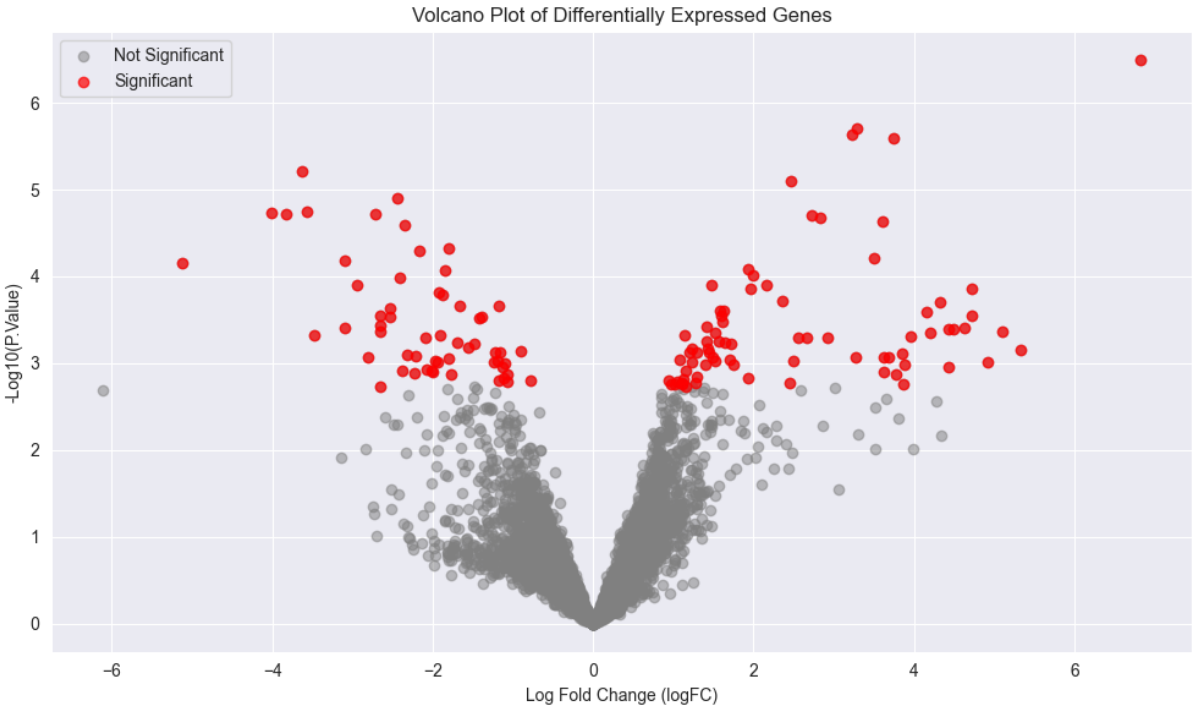


Figure 15: Volcano plot of differential gene expression analysis

4.3.2 rMTA Results

The rMTA analysis produced a ranked list of genes potentially involved in the metabolic response to *unc-62* knockdown. The top genes are presented in Table 14. These genes may play a relevant role in aging and longevity in *C. elegans*, hence should be further investigated for their potential as eventual drug targets. Each of those will be further discussed to assess their role on the biological mechanisms of aging.

Gene stable ID	Gene name	Gene description	Ranking
WBGene00000283	cah-5	Putative carbonic anhydrase 5	1
WBGene00006932	vps-34	Phosphatidylinositol 3-kinase catalytic subunit type 3	2
WBGene00016629	C44B7.7	Gamma-glutamylcyclotransferase	3
WBGene00000269	bre-4	Beta-1,4-N-acetylgalactosaminyltransferase bre-4	4
WBGene00001823	hap-1	Inosine triphosphate pyrophosphatase	5
WBGene00022393	abhd-12	2-arachidonoylglycerol hydrolase ABHD12	6
WBGene00003476	mtm-3	Myotubularin-related protein 3	7
WBGene00013301	Y57G11C.3	Putative 6-phosphogluconolactonase	8
WBGene00009138	F25H9.6	Flavoprotein domain-containing protein	9
WBGene00014203	ZK1058.3	putative galactose-1-phosphate uridylyltransferase	10
WBGene00004092	ppt-1	Palmitoyl-protein thioesterase 1	11
WBGene00019813	ger-1	GDP-L-fucose synthase	12
WBGene00016957	atic-1	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase	13

Table 14: Top-ranked genes from the rMTA analysis.

cah-5 The cah-5 gene encodes a member of the carbonic anhydrase family, enzymes that catalyze the reversible hydration of carbon dioxide, thereby playing a critical role in pH regulation, aion transport, and water balance. Using fluorescence microscopy, studies have shown that the cah-5 gene is expressed in the neurons and intestine, suggesting functional roles in the nervous system and digestive tract (121), similar to unc-62, which is prominently expressed in the intestine.

vps-34 The vps-34 encodes a class III phosphatidylinositol 3-kinase, essential for endocytic sorting, engulfment, and autophagy. Partial inhibition of this gene in adult worms improves neurotransmission and muscle integrity, ameliorating motor aging (122). This effect resembles the effect of the unc-62 knockdowns in adulthood, which also expands lifespan.

bre-4 bre-4 encodes a beta-1,4-N-acetylgalactosaminyltransferase, its expression increases during aging in wild-type worms, suggesting a role in the aging process (123). Furthermore, bre-4 has an essential role in the formation of glycolipid receptors on the plasma membrane of intestinal cells, being these receptors necessary for the binding and uptake of Cry proteins, more specifically the Cry5Ba. Inhibition of this

gene results in resistance to the toxin (124). While increased bre-4 expression correlates with aging, it is unclear if its inhibition extends lifespan. Similar to unc-62, where knockdown reduces detrimental protein accumulation, bre-4 knockdown confers resistance to toxins.

hap-1 The hap-1 ortholog in C. elegans is trak-1, associated with neuronal function, knockout mice lacking hap-1 do not survive postnatal, since this, impact in adult mice is unknown (125), but trak-1 knockout worms develop normally without significant behavioral deficits (126). There is no evidence that reducing hap-1/trak-1 expression extends lifespan.

abhd-12 abhd-12 encodes a serine hydrolase involved mainly in lipid metabolism. In human mutations in abhd-12 lead to a neurodegenerative disorder PHARC (polyneuropathy, hearing loss, ataxia, retinosis pigmentosa, and cataract) (127), in mice, similar effects occur (128). There is no information on abhd-12 in *C.elegans*, its role in lipid metabolism could influence aging, similar to human and mice models.

mtm-3 mtm-3 encodes a myotubularin-related phosphatase involved in autophagy, similar to other autophagy-defective mutants loss of mtm-3 reduces survival, indicating its role in stress response (129). Similar to unc-62 knockdown, which is letal in the worm larvae stage, mtm-3 knockdown also leads to reduced survival.

F25H9.6 F25H9.6 is the ortholog of phosphopantothenoylcysteine decarboxylase, involved in Coenzyme A biosynthesis (130), Knockdown may enhance the mitochondrial unfolded protein response (UPR^{mt}), associated with increased lifespan (131). Similar to unc-62 knockdown, F25H9.6 knockdown could also extend the worm lifespan.

ppt-1 ppt-1 encodes palmitoyl-protein thioesterase 1, involved in depalmitoylation of proteins. Mutations in human PPT1 gene cause infantile neuronal ceroid lipofuscinosis (INCL), a neurodegenerative disease (132). Furthermore, in *C.elegans* mutations resulted in mild developmental and reproductive phenotype, resulting in an abnormal mitochondrial morphology (133) and ppt-1 RNAi showed a decreased median lifespan (134).

ger-1 ger-1 encodes GDP-L-fucose synthase, essential for fucosylation processes, mutations in ger-1 impact pathogen resistance, where ger-1 prevent the *Coprinopsis cinerea* lectin 2 (CCL2) toxin binding (135). Similar to bre-4, ger-1 knockout confers resistance to toxins.

atic-1 atic-1 encodes a bifunctional enzyme in purine biosynthesis, atic-1 overexpression decreases lifespan, while inhibition increases lifespan and reduces neuronal damage (136) Similar to unc-62 knockdown, atic-1 knockdown also extends lifespan.

This analysis highlights key interactions between unc-62 and several candidate genes that share functional similarities affecting aging and lifespan in C. elegans. The knockdowns of bre-4 and ger-1 confer resistance to toxins by disrupting the formation of glycolipid and fucosylated receptors necessary for toxin binding (124; 135). This parallels the effect of unc-62 knockdown, which reduces toxic protein aggregation by decreasing yolk protein expression (117; 115). These similarities suggest that unc-62 may relate with bre-4 and ger-1 within pathways that enhance cellular defenses against internal and external toxins, contributing to improved survival and longevity.

Additionally, the inhibition of vps-34, atic-1, and F25H9.6 results in lifespan extension similar to that observed with unc-62 knockdown. Vps-34 is crucial for autophagy, and its partial inhibition ameliorates motor aging (122), aligning with unc-62's role in reducing protein aggregation. Atic-1 inhibition extends lifespan by modulating purine biosynthesis and reducing neuronal damage (136), while F25H9.6 knockdown may enhance mitochondrial unfolded protein responses associated with increased lifespan (131). These findings indicate that unc-62 may relate with these genes through shared pathways involving autophagy, metabolic regulation, and mitochondrial function, collectively contributing to the regulation of aging and the promotion of longevity.

4.3.3 Enrichment

To further analyze the results, an enrichment analysis was conducted using Enrichr (137). The analysis revealed several significantly enriched pathways, notably the Pentose Phosphate Pathway (PPP), valine, leucine, and isoleucine degradation, the sulfide oxidation pathway, and aminoacyl-tRNA synthetases (ARSs). These pathways are known to be involved in the aging process and have been linked to lifespan regulation in *C elegans*.

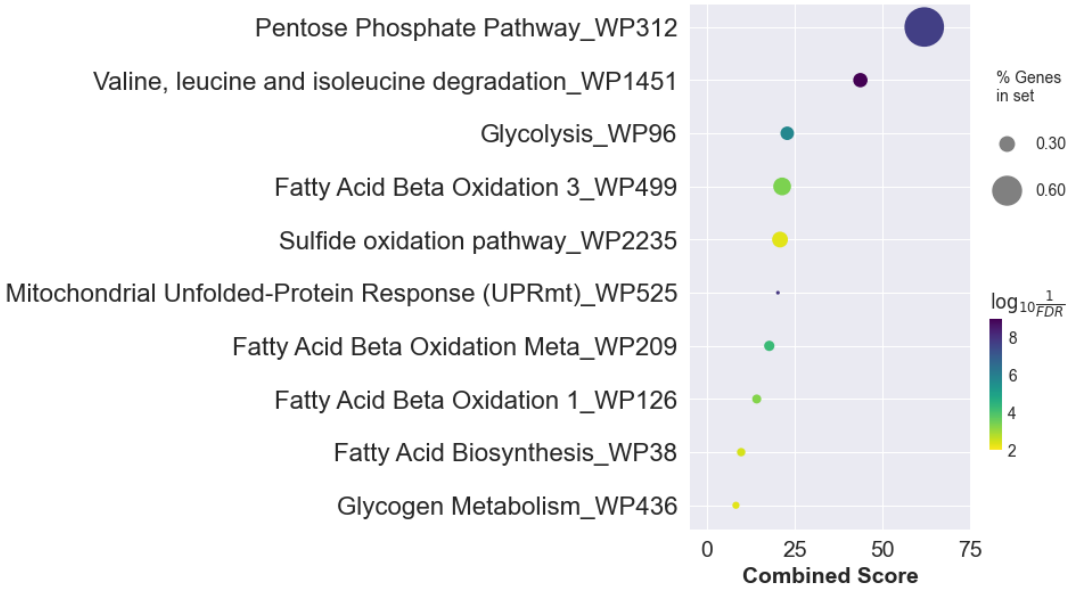


Figure 16: Enrichment analysis of the top 300 genes, highlighting significantly enriched pathways involved in aging

Pentose Phosphate Pathway (PPP)

The PPP plays a crucial role in cellular redox balance by producing nicotinamide adenine dinucleotide phosphate (NADPH), an essential reducing agent that helps combat oxidative stress. NADPH is primarily generated in the oxidative phase of the PPP and is vital for maintaining cellular homeostasis. Several studies have linked shifts in this pathway to lifespan extension in *C. elegans*. For instance, reduction of glycolysis has been shown to activate the PPP, leading to increased NADPH production and enhanced oxidative stress resistance, prolonging lifespan (138). This activation is associated with hormesis, where mild stress induces adaptive beneficial effects on the organism (139). The increased NADPH levels contribute to improved redox stress tolerance and have been reported to extend lifespan (140).

Valine, leucine and isoleucine degradation

The degradation of branched-chain amino acids (BCAAs)—valine, leucine, and isoleucine—is another pathway significantly enriched in the analysis. BCAAs are essential amino acids involved not only in protein synthesis but also in various physiological processes (141). In *C. elegans*, disruption of the branched-chain α-ketoacid dehydrogenase (BCKDH) complex, which catalyzes a critical step in BCAA catabolism, leads to developmental defects and affects lifespan (142)

Sulfide oxidation Pathway

The sulfide oxidation pathway is implicated in the regulation of lifespan through its role in redox signaling. In *C. elegans*, increased production of hydrogen sulfide (H_2S) has been linked to lifespan extension, particularly in response to germline loss (143).

H_2S acts as a signaling molecule that can induce protective cellular processes, including the activation of transcription factors such as SKN-1/Nrf2 and the (UPR^{mt}). These processes enhance stress resistance and promote longevity (144).

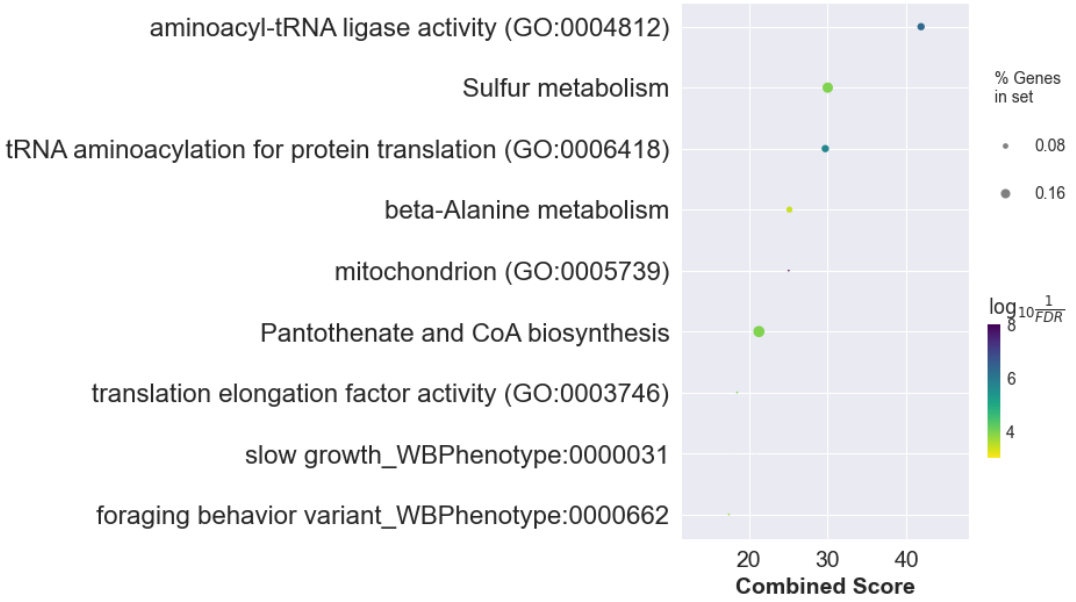


Figure 17: Enrichment analysis of the top 50 genes, highlighting significantly enriched pathways involved in aging

Aminoacyl-tRNA Synthetases (ARSs)

ARSs are essential enzymes required for protein synthesis, responsible for attaching amino acids to their corresponding tRNAs. Beyond their canonical role, ARSs have been associated with various human inherited diseases and have been implicated in aging processes (145). In *C. elegans*, mutations in ARS genes have been linked to lifespan regulation. For example, a null mutation in mitochondrial leucyl-tRNA synthetase (LARS2) is associated with increased lifespan, potentially due to impaired mitochondrial function leading to adaptive stress responses (146). Moreover, downregulation of certain ARS genes has been shown to extend lifespan by modulating pathways such as the Target of Rapamycin (TOR) signaling and the nuclear localization of transcription factors like HLH-30/TFEB (147). These findings suggest that ARSs play a complex role in aging, mediating stress responses and influencing longevity.

The enrichment analysis supports the rMTA findings by highlighting pathways that are known to influence aging and lifespan in *C. elegans*. The identification of the PPP and BCAA degradation pathways suggests that metabolic shifts affecting redox balance and amino acid metabolism are central to the aging process. The sulfide oxidation pathway's involvement underscores the importance of signaling molecules like H_2S in promoting stress resistance and longevity. The enrichment of ARSs indicates that alterations in protein synthesis machinery can have significant impacts on lifespan, possibly through effects on mitochondrial function and cellular stress responses

4.4 Conclusion

In this chapter, rMTA was applied to investigate the metabolic changes associated with aging processes due to *unc-62* knockdown in *C. elegans*. Our analysis identified several candidate genes and pathways that may contribute to lifespan extension and aging regulation. The top-ranked genes from the rMTA analysis share functional similarities with *unc-62* in pathways involving autophagy, metabolic regulation, and stress responses.

The enrichment analysis further supported these findings by highlighting significant pathways such as the Pentose Phosphate Pathway, branched-chain amino acid degradation, sulfide oxidation pathway, and aminoacyl-tRNA synthetases, all of which have been linked to aging and longevity in *C. elegans*.

Chapter 5

Evolutionary Algorithm approach to rMTA

5.1 Introduction

rMTA is a computational approach designed to identify potential metabolic interventions that can shift a biological system from one state to another by targeting specific genes within a metabolic network, particularly single-gene knockouts (9). While rMTA effectively pinpoints single-gene deletions, exploring the combined effects of multiple gene knockouts can provide deeper insights into the metabolic network's complexity and enhance the capacity to achieve the desired metabolic state more effectively.

In this chapter, an EAs-based approach to extend the rMTA will be presented, by evaluating the effects of multiple gene knockouts in the metabolic network. An EA can be used to optimize the metabolic network through several gene deletions to achieve the desired target state more effectively. This approach is similar to strain optimization methods, where multiple gene modifications are employed to achieve a certain metabolic objective. Additionally, a Random Sampling method approach was employed to assess the viability and effectiveness of randomly selected gene combinations in achieving the desired metabolic transformation.

5.2 Methods

5.2.1 Random Sampling

To assess the viability and effectiveness of multi-gene knockout solutions, a Random Sampling approach was first employed as a baseline. This method aimed to assess whether randomly selected gene combinations could achieve significant metabolic transformations, and to what extent the top-ranking genes identified by rMTA contribute to the success of multi-gene knockouts.

Random samples of gene knockouts were generated in two main scenarios:

- **Using Top rMTA Genes:** Gene knockouts were randomly selected from the top-ranking genes identified by rMTA. This was to evaluate whether combinations of these high-ranking genes could effectively shift the metabolic state when multiple knockouts are considered.
- **Using All Model Genes:** Gene knockouts were randomly selected from the entire set of genes present in the metabolic model. This broader sampling provided a comprehensive baseline to compare the effectiveness of random gene combinations against those identified by the and rMTA.

For each scenario, batches of random samples were generated with the following parameters:

- **Number of Samples:** 100 and 500 samples per batch, totaling 8,200 samples
- **Modification Size (Number of Knockouts):** The number of simultaneous gene knockouts in each sample ranged from 2 to 10. This range allowed exploration of the effects of different sizes of gene knockout combinations on the metabolic transformation.
- **Evaluation Metric:** The Transformation Score (TS) was calculated for each random sample to quantify its effectiveness in shifting the metabolic state towards the target.

The main objective of the Random Sampling approach was to determine whether the top-ranking genes from rMTA are effective when used in combination in a multi-knockout scenario.

5.2.2 Evolutionary Algorithms

To systematically explore the effects of multiple gene knockouts and identify optimal combinations, the MEWpy framework (111) was utilized. MEWpy provides a flexible platform for strain optimization using EAs, making it suitable for integrating with rMTA.

Objective Function

A critical component of the EAs is the objective function used to evaluate the fitness of each candidate solution, which in this context is a set of gene deletions. The Transformation Score (TS), previously used in MTA and rMTA, was adapted for use within the MEWpy framework to quantify how effectively a candidate solution shifts the metabolic state towards the target.

The approach was divided into two phases:

- **Initial Objective Function:** The MOMA Transformation Score (mTS), calculated using the Minimization of Metabolic Adjustment (MOMA), was initially used due to its simplicity and lower computational demand. As shown in Table 15, mTS demonstrated a strong correlation with the robust Transformation Score (rTS) from rMTA, justifying its use in preliminary evaluations.

	bTS	**mTS**	**wTS**	**rTS**
rTS	0.659174	0.776672	-0.571475	1.000000

Table 15: Correlation matrix between different TS metrics. The strong correlation between mTS and rTS supports using mTS as the objective function.

- **Robust Objective Function:** Encouraged by initial results, the more computationally intensive but reliable rTS metric was implemented. The rTS also incorporates the best-case TS (bTS) and worst-case TS (wTS) to provide a more robust fitness evaluation, aligning with the rMTA ranking method.

Adaptation of the Evaluation Functions

Custom evaluation functions were developed based on the `EvaluationFunction` class in MEWpy to integrate the TS metrics:

- `TransformationScore_mTS`: Calculates the mTS for each candidate solution.
- `TransformationScore_rTS`: Calculates the rTS, incorporating also, bTS and wTS for robust evaluation.

The Transformation Score (TS) for each scenario is defined as:

$$\frac{\sum_{i \in R_{\text{success}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right] - \sum_{i \in R_{\text{unsuccess}}} \text{abs}\left[\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)\right]}{\sum_{i \in R_S} \text{abs}\left(v_i^{\text{ref}} - v_i^{\text{res}}\right)} \qquad (\textbf{??} \text{ revisited})$$

The robust Transformation Score (rTS) is calculated as:

$$rTS = mTS \cdot |k \cdot |bTS - wTS||^a\,, \qquad (2.6 \text{ revisited})$$

Following the parameter values used in the referenced study, $k = 100$, $a = 1$ when $bTS > 0$, $mTS > 0$, and $wTS < 0$.

Selection of Candidate Genes

To focus the EAs on the most impactful gene deletions on the Top rMTA gene scenario, the gene pool was limited to the top genes identified by the rMTA, more specifically the top 25, top 100 and top 550 genes ranked by their rTS values were selected from the *RRM1* knockout scenario. Additionally, to evaluate the effectiveness of this focused approach of the top-ranking genes, a run including all the genes was conducted.

EA Parameters

The EA parameters were set as follows:

- **Population Size:** 100 (default value from MEWpy).
- **Maximum Generations:** 100 (extended to 500 when considering all genes due to the larger search space).
- **Gene Knockout Limit:** Maximum of 2, 3, and 10 simultaneous knockouts per candidate.

These parameters were selected based on preliminary experiments to balance convergence speed and solution diversity.

5.2.3 Post-Processing

To evaluate the relevance and effectiveness of the EAs solutions, the following analyses were performed:

- **Overlap with rMTA Rankings:** For each solution, the mean ranking of the genes based on their position in the rMTA results was calculated. This assessed whether the EAs favored top-ranked genes.
- **Comparison of Objective Scores:** The scores of multi-gene knockout solutions were compared against the baseline single-gene knockout scores from rMTA.

An example of the EA output is shown in Table 16.

Modification	Size	Score	Modification Index	Modification Mean
8802.1, 5625.1, 353.1	3	0.02712	[4, 2, 12]	6.0

Table 16: Example of EA output showing the selected gene modifications, their respective sizes, scores, modification indices, and the mean rank derived from gene relative positions of the rMTA results.

5.3 Results

5.3.1 Random Sampling Aproach

A total of 8,200 random samples were generated across various batches and modification sizes to assess the effectiveness of random multi-gene knockouts. The Transformation Scores (TS) of these samples were analyzed to evaluate their capacity to shift the metabolic state toward the target and to compare them against the top score of the rMTA. After removing duplicates, a total of 6,691 unique samples were obtained.

Distribution of Transformation Scores The distribution of Transformation Scores from the random samples is shown in Figure 18. The majority of random combinations resulted in TS values near zero, indicating minimal impact on shifting the metabolic state towards the target. A small number of samples achieved higher TS values

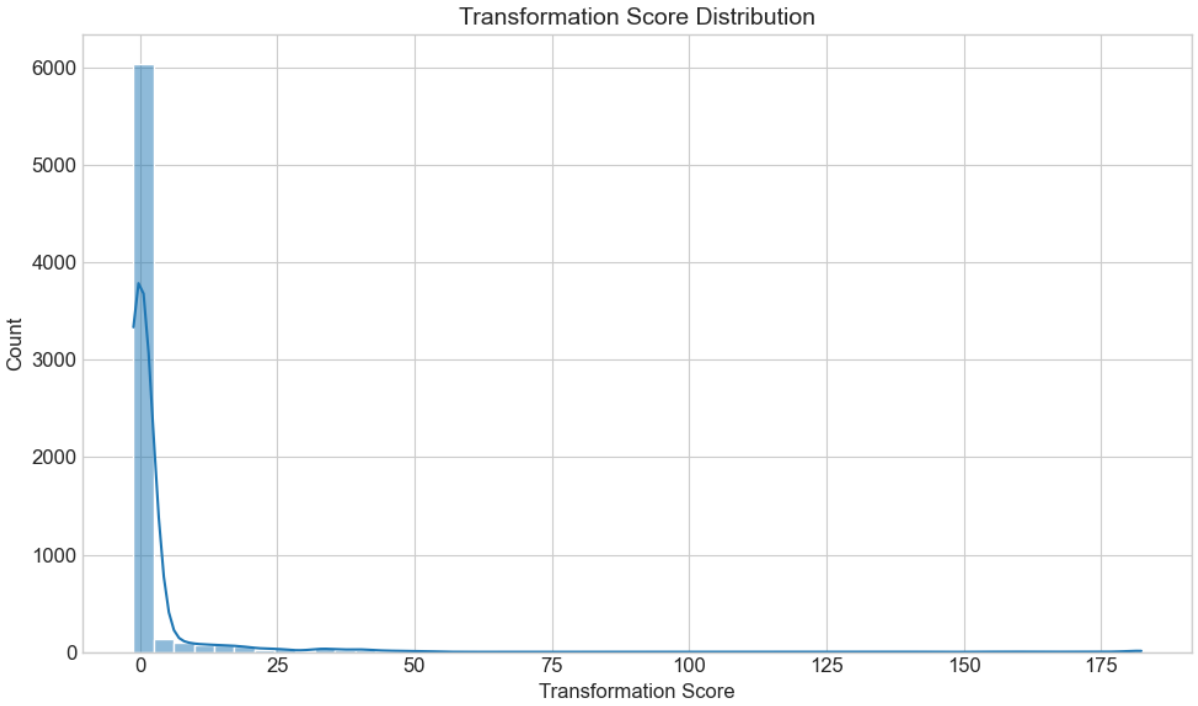


Figure 18: Distribution of Transformation Scores from Random Sampling

Comparison Between Top rMTA Genes and All Model Genes To assess whether the top-ranking genes from rMTA are more effective in multi-gene knockout scenarios, the rTS distributions of random samples using only top rMTA genes were compared with those using all model genes. For this comparison, only the top 250 results from each group were considered (Figure 19).

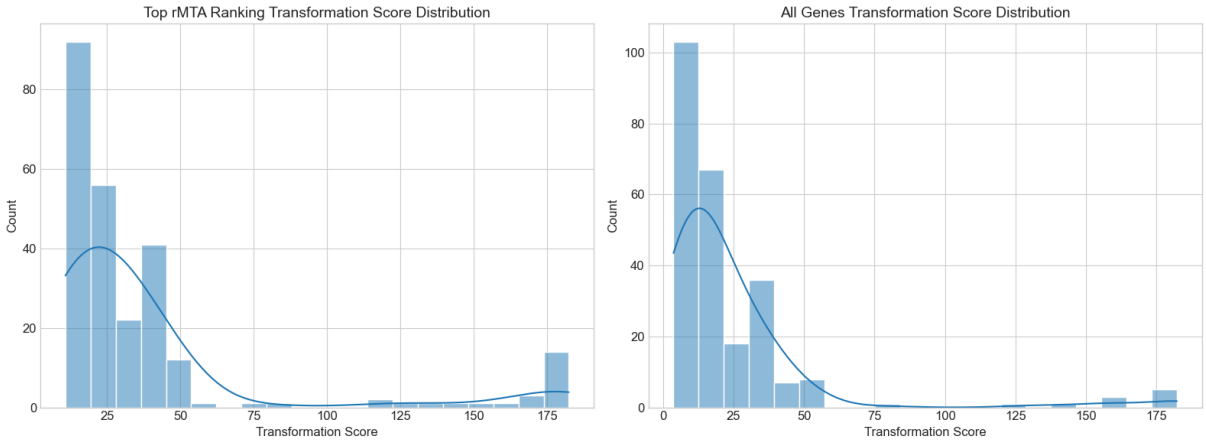


Figure 19: Comparison of rTS distributions between random samples using top rMTA genes and all model genes (top 250 results each) .

The results indicate that random combinations of top rMTA genes tended to achieve slightly higher rTS values compared to random combinations from all model genes.

Statistical Analysis A Mann-Whitney U test was performed to compare the rTS values between the two scenarios (top rMTA genes vs. all model genes). The p-value obtained was 6.59×10^{-14}, indicating a highly statistically significant difference between the two groups. This result suggests that using top rMTA genes in random combinations has a considerably better chance of achieving higher rTS values compared to all model genes.

Comparison with rMTA When comparing the highest rTS values achieved by random sampling with the top rMTA ranking, the top rTS from random sampling was higher than the rTS value from the top score of the rMTA algorithm (Table 17).

Genes	Transformation Score
3034.1, 2730.1, 5831.1	182.33
5831.1*	182.31
51477.1, 3290.1, 5625.1	182.19
5831.1, 4363.1, 4594.1	181.53

Table 17: Comparation between top random sapling solutions and top rMTA ranking. The asterisk (*) indicates the top rMTA gene

5.3.2 EAs Aproach

Initial Approach

As a reference point, the score for the top-ranked gene from the rMTA output corresponding to the *PYCR1* gene (ID: 5831.1) was used. This baseline serves to evaluate the improvements achieved through the multi-gene deletions (Table 18).

Modification	Size	Score
5831.1	1	0.026488

Table 18: Baseline of the Initial mTS score of the top rMTA gene (*PYCR1*).

Optimization of Top 25 genes

Max solutions size of 2 Using the top 25 genes and allowing a maximum of 2 gene knockouts, the EAs identified the optimal modification shown in Table 19.

Modification	Size	Score	Modification Index	Modification Mean
3417.1, 8402.1	2	0.026801	[22, 14]	18.0

Table 19: Optimal gene deletions using top 25 genes with a maximum of 2 knockouts.

Max solutions size of 3 Allowing up to 3 gene knockouts, the EAs identified the optimal modification shown in Table 20.

Modification	Size	Score	Modification Index	Modification Mean
8802.1, 5625.1, 353.1	3	0.027100	[4, 2, 12]	6.0

Table 20: Optimal gene deletions using top 25 genes with a maximum of 3 knockouts.

Optimization with Top 100 Genes

Maximum Solution Size of 2 Expanding the candidate pool to the top 100 genes and limiting solutions to 2 gene knockouts produced the result in Table 21.

Modification	Size	Score	Modification Index	Modification Mean
353.1, 2271.1	2	0.027119	[12, 46]	29.0

Table 21: Optimal gene deletions using top 100 genes with a maximum of 2 knockouts.

Maximum Solution Size of 3 Allowing 3 gene knockouts resulted in several combinations with similar scores (Table 22). Notably, the genes 353.1 and 2271.1 are present in all the optimal solutions.

Modification	Size	Score	Modification Index	Modification Mean
509.1, 353.1, 2271.1	3	0.027118	[73, 12, 46]	43.67
353.1, 2271.1	2	0.027118	[12, 46]	29.00
353.1, 2271.1, 6573.1	3	0.027118	[12, 46, 24]	27.33
178.1, 353.1, 2271.1	3	0.027118	[84, 12, 46]	47.33
7915.1, 353.1, 2271.1	3	0.027118	[49, 12, 46]	35.67
6120.1, 353.1, 2271.1	3	0.027118	[25, 12, 46]	27.67
2235.1, 353.1, 2271.1	3	0.027118	[92, 12, 46]	50.00
514.1, 353.1, 2271.1	3	0.027118	[75, 12, 46]	44.33

Table 22: Optimal gene deletions using top 100 genes with a maximum of 3 knockouts.

Optimization with All Genes To explore the entire search space, a run was conducted allowing all genes to be considered, with a maximum solution size of 2. Due to the large search space, the number of generations was extended to 500. The optimal modification identified is shown in Table 23.

Modification	Size	Score	Modification Index	Modification Mean
3081.1, 26266.1	2	0.026874	[241, 496]	368.5

Table 23: Optimal gene deletions using all genes with a maximum of 2 knockouts.

Robust Approach

Using the rTS as the objective function, the EA runs were repeated to assess the robustness of the solutions.

Baseline Reference As a reference point, the rTS score for the top-ranked gene from the rMTA output, corresponding to the *PYCR1* gene (ID: 5831.1), was used. This baseline serves to evaluate the improvements achieved through the multi-gene deletions (Table 24).

Modification	Size	Score
5831.1	1	182.31

Table 24: Baseline rTS score of the top rMTA gene (*PYCR1*).

Optimization of Top 25 genes

Max solutions size of 2 and 3 Using the top 25 genes and allowing maximum solution sizes of 2 and 3 gene knockouts, the EA consistently identified the single gene 5831.1 (*PYCR1*) as the optimal solution. This suggests that within the top 25 genes, no combination of multi-gene knockouts surpassed the effectiveness of the single-gene knockout of *PYCR1*.

Modification	Size	Score	Modification Index	Modification Mean
5831.1	1	182.31	[1]	1.0

Table 25: Optimal gene deletion using top 25 genes with maximum solution sizes of 2 and 3 knockouts.

Optimization with Top 100 Genes

Maximum Solution Size of 2 Expanding the candidate pool to the top 100 genes and limiting solutions to 2 gene knockouts produced the same result as the top 25 genes, with the single gene 5831.1 being the optimal solution. This indicates that no two-gene combinations within the top 100 genes provided a better rTS than the single-gene knockout of *PYCR1*.

Maximum Solution Size of 3 Allowing up to 3 gene knockouts resulted in a slight improvement in the rTS score. The EA identified the combination of genes 5831.1 (*PYCR1*), 81537.1, and 9415.1 as the optimal solution, achieving an rTS of 182.32, Table 26, marginally higher than the baseline.

Modification	Size	Score	Modification Index	Modification Mean
5831.1, 81537.1, 9415.1	3	182.32	[1, 100, 87]	62.667

Table 26: Optimal gene deletions using top 100 genes with a maximum of 10 knockouts.

The increase in the rTS score suggests that the addition of genes 81537.1 and 9415.1 slightly enhances the effectiveness of the intervention. However, the improvement is minimal, indicating that *PYCR1* remains the primary driver of the metabolic transformation towards the target state.

Maximum Solution Size of 10 When the maximum solution size was increased to 10, the EA identified a combination of genes that slightly improved the rTS score compared to the baseline. The optimal modification identified is shown in Table 27

Modification	Size	Score	Modification Index	Modification Mean
5831.1, 5625.1, 19.1, 81537.1, 3155.1, 9415.1, 4573.1	7	182.32	[1, 2, 99, 100, 83, 87, 90]	66.000

Table 27: Optimal gene deletions using top 100 genes with a maximum of 10 knockouts.

Although the maximum solution size was set to 10, the EA identified an optimal solution with 7 gene knockouts, furthermore, the achieved rTS score of 182.32 is higher than the baseline single-gene knockout score.

Optimization with All Genes To further explore the search space, we expanded the candidate pool to the top 550 genes from the rMTA rankings.

Maximum Solution Size of 2 With a larger gene pool and a maximum of 2 gene knockouts, the EA identified a new optimal combination: genes 5625.1 and 3155.1, achieving an rTS of 183.01. This score surpasses the baseline and previous results, indicating a more effective multi-gene knockout combination (Table 28).

Modification	Size	Score	Modification Index	Modification Mean
5625.1, 3155.1	2	183.01	[2, 226]	114.00

Table 28: Optimal gene deletions using top 550 genes with a maximum of 2 knockouts.

Maximum Solution Size of 3 Allowing up to 3 gene knockouts, the EA identified the combination of genes 5831.1, 57379.1, and 2618.1, achieving an rTS of 182.54. While this score is higher than the baseline, it does not surpass the two-gene combination identified previously.

Modification	Size	Score	Modification Index	Modification Mean
5831.1, 57379.1, 2618.1	3	182.54	[1, 347, 417]	255.00

Table 29: Optimal gene deletions using top 550 genes with a maximum of 3 knockouts.

Summary of Optimal Modifications To consolidate the findings from the robust EA approach, a summary of the optimal gene knockout combinations and their corresponding rTS is presented in Table 30. This table compares the highest-scoring solutions identified across different gene pools and maximum solution sizes, highlighting the improvements achieved over the single-gene knockout reference.

Samples	Transformation Score
Top 550, Max Solution Size of 2	183.01
Top 550, Max Solution Size of 3	182.54
Top 100, Max Solution Size of 10	182.32
Top 100, Max Solution Size of 3	182.32
Reference	182.31

Table 30: Summary of optimal modifications and their rTS

Table 30 shows that the EA runs with the larger gene pool (top 550 genes) and a maximum solution size of 2 achieved the highest rTS score of 183.01, surpassing the single-gene knockout reference score of 182.31. This indicates that certain combinations of two gene knockouts can slightly enhance the metabolic transformation towards the target state compared to single-gene interventions. The marginal improvements observed suggest that while multi-gene knockouts can be beneficial, the primary driver of the transformation remains the key gene identified by rMTA.

Genes Analysis An analysis of the most frequently occurring genes in the optimal solutions revealed that certain genes appeared multiple times across different runs, suggesting their significant role in achieving the desired metabolic transformation (Table 31).

Gene	Occurrences	rMTA Ranking Position
5831.1	5	1
9415.1	2	87
5625.1	2	2
3155.1	2	226
81537.1	2	100

Table 31: Top 5 most frequently occurring genes in optimal solutions.

These genes consistently appear across different EA runs, indicating their potential importance in the

metabolic network for achieving the desired target state. Notably, gene 5831.1 (*PYCR1*) appears in all optimal solutions, underscoring its significant impact on the metabolic transformation.

The rMTA ranking positions of these genes indicate that they are among the top candidates identified by the rMTA, further validating their significance. Gene 5625.1, ranked second in the rMTA results, appears twice in the optimal solutions, reinforcing its potential as a key target. Gene 3155.1, despite being ranked 226th, occurs twice, suggesting that its role may be more critical in combination with other genes than as a single knockout.

5.4 Discussion

Biological Target State: RRM1 RRM1 encodes the large subunit of ribonucleotide reductase (RR), a crucial enzyme for the de novo synthesis of deoxyribonucleotide triphosphate (dNTP). RR catalyzes the reduction of ribonucleoside diphosphate (NDP) to deoxyribonucleotide diphosphate (dNDP), providing essential building blocks for DNA synthesis and repair. Knocking out RRM1 directly impairs dNTP production, leading to replication stress and potential cell cycle arrest due to insufficient DNA synthesis capability. This disruption is particularly impactful in rapidly dividing cells, where high dNTP demand is critical for proliferation (114).

5.4.1 Random Sampling Aproach

The Random Sampling approach aimed to explore random points in the search space to assess the viability and effectiveness of multi-gene knockout solutions. The results demonstrated that random combinations of genes generally resulted in minimal impact on the metabolic state, as indicated by TS values near zero. However, a small subset of random combinations achieved higher TS values, even surpassing the top rMTA single-gene knockout score.

The statistically significant difference between the TS distributions of random samples using top rMTA genes versus all model genes suggests that the top-ranked genes have a higher propensity to produce effective multi-gene knockout combinations. This validates the use of rMTA rankings genes as a foundation for further optimization.

5.4.2 EAs Approach

The integration of EAs with the rMTA enabled systematic exploration of multi-gene knockout strategies. The EAs identified gene combinations that marginally improved upon single-gene knockouts found by rMTA.

A key finding was the consistent identification of gene 5831.1 (*PYCR1*) across different EAs runs, underscoring its significant impact on shifting the metabolic state towards the target. Other genes like 5625.1, 3155.1, 81537.1, and 9415.1 appeared multiple times in optimal solutions, suggesting potential synergistic effects when combined with *PYCR1*. However, the slight increase of rTS indicate that additional gene deletions contribute only incrementally to the metabolic shift.

Overall, the EAs approach demonstrated that while multi-gene knockouts can offer marginal improvements, the primary driver of the metabolic shift remains the key individual genes identified by rMTA. This emphasizes the importance of targeting these critical genes in therapeutic strategies and suggests that the added complexity of multi-gene interventions may yield limited additional benefits.

5.4.3 Biological Interpretation of Identified Genes

The genes identified in the optimal multi-gene knockout combinations are involved in pathways related to nucleotide metabolism and energy production. For instance:

APRT and FH

APRT (adenine phosphoribosyltransferase) and FH (fumarate hydratase) are interconnected in nucleotide salvage and energy metabolism.

APRT functions in the purine salvage pathway by converting adenine to adenosine monophosphate (AMP), maintaining nucleotide pools, and conserving energy. Knocking out APRT disrupts this pathway, decreasing AMP and ATP levels, forcing the cell to rely on energy-intensive de novo nucleotide synthesis (148).

FH plays a key role in the tricarboxylic acid (TCA) cycle by converting fumarate to malate, contributing to ATP production and biosynthetic precursors. An FH knockout impairs the TCA cycle, leading to reduced ATP generation and accumulation of fumarate, which can inhibit other metabolic enzymes (149).

The combined knockouts of APRT and FH significantly decrease nucleotide synthesis due to limited substrates and energy molecules, indirectly inhibiting RNR activity and mimicking the cellular conditions of an RRM1 knockout.

5.4.4 Limitations

Despite the results, several limitations were noted:

- **Computational Complexity:** The EAs approach, especially when considering larger gene pools

and higher maximum solution sizes, is computationally intensive, requiring extended run times.

- **Biological Validity:** While the EA identified optimal gene combinations based on computational metrics, the biological feasibility and potential off-target effects of these multi-gene knockouts require further experimental validation.

5.5 Conclusion

In this chapter, the capabilities of rMTA were extended by integrating an EAs approach to explore multi-gene knockout strategies for achieving desired metabolic transformations. The EAs effectively identified gene combinations that improved upon the baseline single-gene knockouts, demonstrating the potential benefits of considering multi-gene interventions.

The Random Sampling approach established that combinations of top-ranking rMTA genes have a higher likelihood of producing effective metabolic transformations compared to random combinations from all model genes. This underscores the value of rMTA rankings in guiding optimization efforts.

However, the limitations of computational resources and the marginal improvements observed highlight the need for further refinement of the optimization strategies. The dataset focused on the RRM1 single-gene knockout presents limitations, as it may not capture the full spectrum of metabolic interactions that can be leveraged through multi-gene interventions.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This dissertation has successfully adapted and implemented the MTA and its robust version, rMTA, from MATLAB to Python. The transition to Python was motivated by the need to improve accessibility and usability for the scientific community. Python's open-source nature and extensive libraries make it a more flexible and widely available platform compared to MATLAB, which is proprietary and may limit accessibility for some researchers.

The validation of the Python implementations was achieved through gene knockout case studies. These studies confirmed that the Python versions of MTA and rMTA maintain the accuracy and robustness of the original MATLAB implementations. Comparative analyses demonstrated that the results from the Python algorithms closely align with those obtained from the MATLAB versions, thereby affirming the reliability of the adapted algorithms.

The application of the rMTA algorithm in aging research was demonstrated using the model organism *Caenorhabditis elegans*. Specifically, the study focused on the knockdown of the *unc-62* gene, which is known for its significant role in longevity and aging processes. The rMTA analysis identified key metabolic changes associated with aging-related phenotypes resulting from the *unc-62* knockdown. These findings underscore the utility of rMTA in uncovering candidate genes and pathways involved in the aging process, providing valuable insights into the underlying biological mechanisms.

To enhance the capability of rMTA further, EAs were used. This allowed for the simultaneous analysis of multiple gene knockouts, optimizing the identification of metabolic targets for phenotype reversion. The use of EAs improved the efficiency of metabolic target identification by effectively navigating the complex combinatorial space of potential genetic perturbations.

In conclusion, the successful adaptation of MTA and to Python, along with their validation and application to aging research, represents a significant advancement in computational biology tools. The use of

EAs, enhances the potential of these algorithms in identifying metabolic targets for therapeutic intervention. This work contributes to systems biology by providing accessible and robust tools for aging studies and drug target discovery, potentially accelerating research in personalized medicine and complex disease treatments.

6.2 Prospects for Future Work

Based on the work done in this dissertation, several paths for future research are proposed:

Firstly, the further refinement of the EAs integration within the framework presents a promising opportunity. Investigating more advanced or hybrid EAs techniques—such as combining genetic algorithms with machine learning approaches—could enhance the precision and efficiency of metabolic target identification. These hybrid methods may improve the algorithm's ability to explore complex search spaces and identify optimal solutions more effectively.

Secondly, expanding the application of rMTA beyond the *C. elegans* model to mammalian systems is a significant next step. Applying the algorithms to mammalian models could offer deeper insights into human aging and age-related diseases, increasing the translational potential of the research. This expansion poses challenges related to scalability, as mammalian metabolic models involve larger and more complex datasets. Enhancing computational efficiency and algorithm scalability will be essential to manage these complexities effectively.

Thirdly, integrating multi-omics data—such as proteomics and epigenomics—into the rMTA framework could significantly improve prediction accuracy and provide a more comprehensive view of the metabolic and regulatory networks involved in aging and disease processes. Incorporating diverse layers of biological information would enable a holistic approach to understanding the intricate mechanisms underlying aging, potentially uncovering novel therapeutic targets and biomarkers.

Additionally, exploring the potential role of Large Language Models (LLMs) in improving the interpretive power of rMTA is an intriguing prospect. LLM can analyze vast amounts of biological literature and datasets, potentially revealing insights not readily apparent through traditional analysis methods. Integrating LLMs with could enhance the interpretation of results, leading to more informed and impactful scientific discoveries.

Finally, extending the use of MTA and rMTA to other research areas—such as metabolic disorders—could have significant implications. By applying these algorithms to different pathological conditions, we can contribute to a better understanding of disease mechanisms and the development of targeted

therapies. This would further establish MTA and rMTA as versatile tools in systems biology, capable of addressing a wide range of biomedical challenges.

In summary, the proposed improvements and applications have the potential to transform aging research and extend the impact of MTA and rMTA in other fields. By refining computational methods, expanding organismal applications, integrating multi-omics data, and leveraging advanced analytical tools like LLMs, these algorithms can contribute significantly to advancements in personalized medicine and the treatment of complex diseases.

Bibliography

[1] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The Hallmarks of Aging. *Cell*, 153(6):1194–1217, June 2013.

[2] Arianna Calcinotto, Jaskaren Kohli, Elena Zagato, Laura Pellegrini, Marco Demaria, and Andrea Alimonti. Cellular Senescence: Aging, Cancer, and Injury. *Physiological Reviews*, 99(2):1047–1078, April 2019.

[3] Jan H.J. Hoeijmakers. DNA Damage, Aging, and Cancer. *New England Journal of Medicine*, 361(15):1475–1485, October 2009.

[4] Judith Campisi. Aging, Cellular Senescence, and Cancer. *Annual Review of Physiology*, 75(1):685–705, February 2013.

[5] Alan A. Cohen, Luigi Ferrucci, Tamàs Fülöp, Dominique Gravel, Nan Hao, Andres Kriete, Morgan E. Levine, Lewis A. Lipsitz, Marcel G. M. Olde Rikkert, Andrew Rutenberg, Nicholas Stroustrup, and Ravi Varadhan. A complex systems approach to aging biology. *Nature Aging*, 2(7):580–591, July 2022.

[6] Levi B. Wood, Ashley R. Winslow, and Samantha Dale Strasser. Systems biology of neurodegenerative diseases. *Integrative Biology*, 7(7):758–775, 2015.

[7] Thomas B. L. Kirkwood. Systems biology of ageing and longevity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1561):64–70, January 2011.

[8] Keren Yizhak, Orshay Gabay, Haim Cohen, and Eytan Ruppin. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nature Communications*, 4(1):2632, October 2013.

[9] Luis V Valcárcel, Verónica Torrano, Luis Tobalina, Arkaitz Carracedo, and Francisco J Planes. rMTA: robust metabolic transformation analysis. *Bioinformatics*, 35(21):4350–4355, November 2019.

[10] Ori Folger, Livnat Jerby, Christian Frezza, Eyal Gottlieb, Eytan Ruppin, and Tomer Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology*, 7(1):501, January 2011.

[11] Arvind K. Chavali, Kevin M. D'Auria, Erik L. Hewlett, Richard D. Pearson, and Jason A. Papin. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends in Microbiology*, 20(3):113–123, March 2012.

[12] Irina M. Conboy, Michael J. Conboy, Amy J. Wagers, Eric R. Girma, Irving L. Weissman, and Thomas A. Rando. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature*, 433(7027):760–764, February 2005.

[13] Brian T. Weinert and Poala S. Timiras. Invited Review: Theories of aging. *Journal of Applied Physiology*, 95(4):1706–1716, October 2003.

[14] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, Sergi G Costafreda, Amit Dias, Nick Fox, Laura N Gitlin, Robert Howard, Helen C Kales, Mika Kivimäki, Eric B Larson, Adesola Ogunniyi, Vasiliki Orgeta, Karen Ritchie, Kenneth Rockwood, Elizabeth L Sampson, Quincy Samus, Lon S Schneider, Geir Selbæk, Linda Teri, and Naaheed Mukadam. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248):413–446, August 2020.

[15] Christopher K. Glass, Kaoru Saijo, Beate Winner, Maria Carolina Marchetto, and Fred H. Gage. Mechanisms Underlying Inflammation in Neurodegeneration. *Cell*, 140(6):918–934, March 2010.

[16] Richard N. L. Lamptey, Bivek Chaulagain, Riddhi Trivedi, Avinash Gothwal, Buddhadev Layek, and Jagdish Singh. A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics. *International Journal of Molecular Sciences*, 23(3):1851, February 2022.

[17] Christopher A Ross and Michelle A Poirier. Protein aggregation and neurodegenerative disease. *Nature Medicine*, 10(S7):S10–S17, July 2004.

[18] Kim A. Caldwell, Corey W. Willicott, and Guy A. Caldwell. Modeling neurodegeneration in *Caenorhabditis elegans*. *Disease Models & Mechanisms*, 13(10):dmm046110, October 2020.

[19] Woojin Kim, Ryan S Underwood, Iva Greenwald, and Daniel D Shaye. OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. *Genetics*, 210(2):445–461, October 2018.

[20] Gang Yan, Petra E. Vértes, Emma K. Towlson, Yee Lian Chew, Denise S. Walker, William R. Schafer, and Albert-László Barabási. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature*, 550(7677):519–523, October 2017.

[21] Rebecca A. Green, Anjon Audhya, Andrei Pozniakovsky, Alexander Dammermann, Hayley Pemble, Joost Monen, Nathan Portier, Anthony Hyman, Arshad Desai, and Karen Oegema. Expression and Imaging of Fluorescent Proteins in the C. elegans Gonad and Early Embryo. In *Methods in Cell Biology*, volume 85, pages 179–218. Elsevier, 2008.

[22] Titus Kaletta and Michael O. Hengartner. Finding function in novel targets: C. elegans as a model organism. *Nature Reviews Drug Discovery*, 5(5):387–399, May 2006.

[23] Adanna G. Alexander, Vanessa Marfil, and Chris Li. Use of Caenorhabditis elegans as a model to study Alzheimerâ™s disease and other neurodegenerative diseases. *Frontiers in Genetics*, 5, September 2014.

[24] Leroy Hood and Lee Rowen. The human genome project: big science transforms biology and medicine. *Genome Medicine*, 5(9):79, 2013.

[25] Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy, and Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. National Academies Press, Washington, D.C., August 2012. Pages: 13297.

[26] Xiaofeng Dai and Li Shen. Advances and Trends in Omics Technology Development. *Frontiers in Medicine*, 9:911861, July 2022.

[27] Rohan Lowe, Neil Shirley, Mark Bleackley, Stephen Dolan, and Thomas Shafee. Transcriptomics technologies. *PLOS Computational Biology*, 13(5):e1005457, May 2017.

[28] Julie Lake, Catherine S. Storm, Mary B. Makarious, and Sara Bandres-Ciga. Genetic and Transcriptomic Biomarkers in Neurodegenerative Diseases: Current Situation and the Road Ahead. *Cells*, 10(5):1030, April 2021.

[29] Roger Bumgarner. Overview of DNA Microarrays: Types, Applications, and Their Future. *Current Protocols in Molecular Biology*, 101(1), January 2013.

[30] Fatih Ozsolak and Patrice M. Milos. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2):87–98, February 2011.

[31] Simon Joost, Amit Zeisel, Tina Jacob, Xiaoyan Sun, Gioele La Manno, Peter Lönnerberg, Sten Linnarsson, and Maria Kasper. Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Systems*, 3(3):221–237.e9, September 2016.

[32] Xiaohong Li, Guy N. Brock, Eric C. Rouchka, Nigel G. F. Cooper, Dongfeng Wu, Timothy E. O'Toole, Ryan S. Gill, Abdallah M. Eteleeb, Liz O'Brien, and Shesh N. Rai. A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. *PLOS ONE*, 12(5):e0176185, May 2017.

[33] Yingdong Zhao, Ming-Chung Li, Mariam M. Konaté, Li Chen, Biswajit Das, Chris Karlovich, P. Mickey Williams, Yvonne A. Evrard, James H. Doroshow, and Lisa M. McShane. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *Journal of Translational Medicine*, 19(1):269, December 2021.

[34] Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, and Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2):giab008, January 2021.

[35] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, December 2014.

[36] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, January 2010.

[37] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang, and Ali Mortazavi. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1):13, December 2016.

[38] Emily Clough and Tanya Barrett. The Gene Expression Omnibus Database. In Ewy Mathé and Sean Davis, editors, *Statistical Genomics*, volume 1418, pages 93–110. Springer New York, New York, NY, 2016. Series Title: Methods in Molecular Biology.

[39] Jennifer A. Webster, J. Raphael Gibbs, Jennifer Clarke, Monika Ray, Weixiong Zhang, Peter Holmans, Kristen Rohrer, Alice Zhao, Lauren Marlowe, Mona Kaleem, Donald S. McCorquodale, Cindy Cuello, Doris Leung, Leslie Bryden, Priti Nath, Victoria L. Zismann, Keta Joshipura, Matthew J. Huentelman, Diane Hu-Lince, Keith D. Coon, David W. Craig, John V. Pearson, Christopher B. Heward, Eric M. Reiman, Dietrich Stephan, John Hardy, and Amanda J. Myers. Genetic Control of Human Brain Transcript Expression in Alzheimer Disease. *The American Journal of Human Genetics*, 84(4):445–458, April 2009.

[40] Diletta Rosati, Maria Palmieri, Giulia Brunelli, Andrea Morrione, Francesco Iannelli, Elisa Frullanti, and Antonio Giordano. Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review. *Computational and structural biotechnology journal*, 2024.

[41] Johnathan Cooper-Knock, Janine Kirby, Laura Ferraiuolo, Paul R Heath, Magnus Rattray, and Pamela J Shaw. Gene expression profiling in human neurodegenerative disease. *Nature Reviews Neurology*, 8(9):518–530, 2012.

[42] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15:1–21, 2014.

[43] Yunshun Chen, Lizhong Chen, Aaron TL Lun, Pedro L Baldoni, and Gordon K Smyth. edger 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv*, pages 2024–01, 2024.

[44] Matthew E Ritchie, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

[45] Charity W Law, Yunshun Chen, Wei Shi, and Gordon K Smyth. voom: Precision weights unlock linear model analysis tools for rna-seq read counts. *Genome biology*, 15:1–17, 2014.

[46] Franck Rapaport, Raya Khanin, Yupu Liang, Mono Pirun, Azra Krek, Paul Zumbo, Christopher E Ma-

son, Nicholas D Socci, and Doron Betel. Comprehensive evaluation of differential gene expression analysis methods for rna-seq data. *Genome biology*, 14:1–13, 2013.

[47] Mark R Dalman, Anthony Deeter, Gayathri Nimishakavi, and Zhong-Hui Duan. Fold change and p-value cutoffs significantly alter microarray interpretations. In *BMC bioinformatics*, volume 13, pages 1–4. Springer, 2012.

[48] Adam McDermaid, Brandon Monier, Jing Zhao, Bingqiang Liu, and Qin Ma. Interpretation of differential gene expression results of rna-seq data: review and integration. *Briefings in bioinformatics*, 20(6):2044–2054, 2019.

[49] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005.

[50] Jüri Reimand, Ruth Isserlin, Veronique Voisin, Mike Kucera, Christian Tannus-Lopes, Asha Rostamianfar, Lina Wadi, Mona Meyer, Jeff Wong, Changjiang Xu, et al. Pathway enrichment analysis and visualization of omics data using g: Profiler, gsea, cytoscape and enrichmentmap. *Nature protocols*, 14(2):482–517, 2019.

[51] Xinzhong Li, Haiyan Wang, Jintao Long, Genhua Pan, Taigang He, Oleg Anichtchik, Robert Belshaw, Diego Albani, Paul Edison, Elaine K Green, et al. Systematic analysis and biomarker study for alzheimer's disease. *Scientific reports*, 8(1):17394, 2018.

[52] A Siavoshi, M Taghizadeh, E Dookhe, and M Piran. Gene expression profiles and pathway enrichment analysis to identification of differentially expressed gene and signaling pathways in epithelial ovarian cancer based on high-throughput rna-seq data. *Genomics*, 114(1):161–170, 2022.

[53] Haoguang Li, Jie Zhou, Lu Zhou, Xiuling Zhang, Jingjing Shang, Xueqin Feng, Le Yu, Jie Fan, Jie Ren, Rongwei Zhang, et al. Identification of the shared gene signatures and molecular pathways in systemic lupus erythematosus and diffuse large b-cell lymphoma. *The journal of gene medicine*, 25(12):e3558, 2023.

[54] Md Rezanur Rahman, Tania Islam, Toyfiquz Zaman, Md Shahjaman, Md Rezaul Karim, Fazlul Huq, Julian MW Quinn, RM Damian Holsinger, Esra Gov, and Mohammad Ali Moni. Identification of

molecular signatures and pathways to identify novel therapeutic targets in alzheimer's disease: Insights from a systems biomedicine perspective. *Genomics*, 112(2):1290–1299, 2020.

[55] Yuejun Fang and Xiaoan Zhan. Identification of biomarkers associated with the prognoses of colorectal cancer patients. *Digestion*, 104(2):148–162, 2023.

[56] Medi Kori, Busra Aydın, Semra Unal, Kazim Yalcin Arga, and Dilek Kazan. Metabolic biomarkers and neurodegeneration: a pathway enrichment analysis of alzheimer's disease, parkinson's disease, and amyotrophic lateral sclerosis. *Omics: a journal of integrative biology*, 20(11):645–661, 2016.

[57] Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome Biology*, 20(1):121, December 2019.

[58] Cheng Zhang and Qiang Hua. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Frontiers in Physiology*, 6, January 2016.

[59] Ines Thiele and Bernhard Ø Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1):93–121, 2010.

[60] Marzia Di Filippo, Chiara Damiani, and Dario Pescini. Gpruler: Metabolic gene-protein-reaction rules automatic reconstruction. *PLoS computational biology*, 17(11):e1009550, 2021.

[61] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, April 2012.

[62] R. Mahadevan and C.H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276, October 2003.

[63] Xin Fang, Colton J. Lloyd, and Bernhard O. Palsson. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 18(12):731–743, December 2020.

[64] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, March 2010.

[65] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349, June 2010.

[66] Nathan E Lewis, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, Daniel Lopez-Ferrer, Karl K Weitz, Roland Eils, Rainer König, Richard D Smith, and Bernhard Ø Palsson. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(1):390, January 2010.

[67] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, November 2002.

[68] Helena A Herrmann, Beth C Dyson, Lucy Vass, Giles N Johnson, and Jean-Marc Schwartz. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ systems biology and applications*, 5(1):32, 2019.

[69] Jan Schellenberger and Bernhard Ø Palsson. Use of randomized sampling for analysis of metabolic networks. *Journal of biological chemistry*, 284(9):5457–5461, 2009.

[70] David E Kaufman and Robert L Smith. Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research*, 46(1):84–95, 1998.

[71] Wout Megchelenbrink, Martijn Huynen, and Elena Marchiori. optgpsampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PloS one*, 9(2):e86587, 2014.

[72] Hulda S Haraldsdóttir, Ben Cousins, Ines Thiele, Ronan MT Fleming, and Santosh Vempala. Chrr: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11):1741–1743, 2017.

[73] J Craig Venter, Mark D Adams, Eugene W Myers, Peter W Li, Richard J Mural, Granger G Sutton, Hamilton O Smith, Mark Yandell, Cheryl A Evans, Robert A Holt, et al. The sequence of the human genome. *science*, 291(5507):1304–1351, 2001.

[74] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, October 2004.

[75] Pedro Romero, Jonathan Wagg, Michelle L Green, Dale Kaiser, Markus Krummenacker, and Peter D Karp. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology*, 6(1):R2, 2004.

[76] G. Joshi-Tope. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research*, 33(Database issue):D428–D432, December 2004.

[77] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Ø. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, February 2007.

[78] Hongwu Ma, Anatoly Sorokin, Alexander Mazein, Alex Selkov, Evgeni Selkov, Oleg Demin, and Igor Goryanin. The Edinburgh human metabolic network reconstruction and its functional analysis. *Molecular Systems Biology*, 3(1):135, January 2007.

[79] Martin I Sigurdsson, Neema Jamshidi, Eirikur Steingrimsson, Ines Thiele, and Bernhard Ø Palsson. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Systems Biology*, 4(1):140, December 2010.

[80] Neil Swainston, Kieran Smallbone, Hooman Hefzi, Paul D. Dobson, Judy Brewer, Michael Hanscho, Daniel C. Zielinski, Kok Siong Ang, Natalie J. Gardiner, Jahir M. Gutierrez, Sarantos Kyriakopoulos, Meiyappan Lakshmanan, Shangzhong Li, Joanne K. Liu, Veronica S. Martínez, Camila A. Orellana, Lake-Ee Quek, Alex Thomas, Juergen Zanghellini, Nicole Borth, Dong-Yup Lee, Lars K. Nielsen, Douglas B. Kell, Nathan E. Lewis, and Pedro Mendes. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, 12(7):109, July 2016.

[81] Elizabeth Brunk, Swagatika Sahoo, Daniel C Zielinski, Ali Altunkaya, Andreas Dräger, Nathan Mih, Francesco Gatto, Avlant Nilsson, German Andres Preciat Gonzalez, Maike Kathrin Aurich, Andreas Prlić, Anand Sastry, Anna D Danielsdottir, Almut Heinken, Alberto Noronha, Peter W Rose, Stephen K Burley, Ronan M T Fleming, Jens Nielsen, Ines Thiele, and Bernhard O Palsson. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*, 36(3):272–281, March 2018.

[82] Christoph Gille, Christian Bölling, Andreas Hoppe, Sascha Bulik, Sabrina Hoffmann, Katrin Hübner, Anja Karlstädt, Ramanan Ganeshan, Matthias König, Kristian Rother, Michael Weidlich, Jörn Behre, and Herrmann-Georg Holzhütter. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology*, 6(1):411, January 2010.

[83] Adil Mardinoglu, Rasmus Agren, Caroline Kampf, Anna Asplund, Mathias Uhlen, and Jens Nielsen. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications*, 5(1):3083, January 2014.

[84] Jonathan L. Robinson, Pınar Kocabaş, Hao Wang, Pierre-Etienne Cholley, Daniel Cook, Avlant Nilsson, Mihail Anton, Raphael Ferreira, Iván Domenzain, Virinchi Billa, Angelo Limeta, Alex Hedin, Johan Gustafsson, Eduard J. Kerkhoven, L. Thomas Svensson, Bernhard O. Palsson, Adil Mardinoglu, Lena Hansson, Mathias Uhlén, and Jens Nielsen. An atlas of human metabolism. *Science Signaling*, 13(624):eaaz1482, March 2020.

[85] Finja Büchel, Nicolas Rodriguez, Neil Swainston, Clemens Wrzodek, Tobias Czauderna, Roland Keller, Florian Mittag, Michael Schubert, Mihai Glont, Martin Golebiewski, Martijn Van Iersel, Sarah Keating, Matthias Rall, Michael Wybrow, Henning Hermjakob, Michael Hucka, Douglas B Kell, Wolfgang Müller, Pedro Mendes, Andreas Zell, Claudine Chaouiya, Julio Saez-Rodriguez, Falk Schreiber, Camille Laibe, Andreas Dräger, and Nicolas Le Novère. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, 7(1):116, December 2013.

[86] M. Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, January 2000.

[87] Ron Caspi, Richard Billington, Ingrid M Keseler, Anamika Kothari, Markus Krummenacker, Peter E Midford, Wai Kit Ong, Suzanne Paley, Pallavi Subhraveti, and Peter D Karp. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research*, 48(D1):D445–D453, January 2020.

[88] Juliane Gebauer, Christoph Gentsch, Johannes Mansfeld, Kathrin Schmeißer, Silvio Waschina, Susanne Brandes, Lukas Klimmasch, Nicola Zamboni, Kim Zarse, Stefan Schuster, Michael Ristow, Sascha Schäuble, and Christoph Kaleta. A Genome-Scale Database and Reconstruction of Caenorhabditis elegans Metabolism. *Cell Systems*, 2(5):312–322, May 2016.

[89] L. Safak Yilmaz and Albertha J.M. Walhout. A Caenorhabditis elegans Genome-Scale Metabolic Network Model. *Cell Systems*, 2(5):297–311, May 2016.

[90] Li Ma, Angelo Hoi Chung Chan, Jake Hattwell, Paul R. Ebert, and Horst Joachim Schirra. Systems biology analysis using a genome-scale metabolic model shows that phosphine triggers global metabolic suppression in a resistant strain of *C. elegans*. preprint, Systems Biology, May 2017.

[91] Michael Witting, Janna Hastings, Nicolas Rodriguez, Chintan J. Joshi, Jake P. N. Hattwell, Paul R. Ebert, Michel Van Weeghel, Arwen W. Gao, Michael J. O. Wakelam, Riekelt H. Houtkooper, Abraham Mains, Nicolas Le Novère, Sean Sadykoff, Frank Schroeder, Nathan E. Lewis, Horst-Joachim Schirra, Christoph Kaleta, and Olivia Casanueva. Modeling Meets Metabolomics—The WormJam Consensus Model as Basis for Metabolic Studies in the Model Organism Caenorhabditis elegans. *Frontiers in Molecular Biosciences*, 5:96, November 2018.

[92] Lutfu Safak Yilmaz, Xuhang Li, Shivani Nanda, Bennett Fox, Frank Schroeder, and Albertha Jm Walhout. Modeling tissue-relevant *Caenorhabditis elegans* metabolism at network, pathway, reaction, and metabolite levels. *Molecular Systems Biology*, 16(10):e9649, October 2020.

[93] Sjoerd Opdam, Anne Richelle, Benjamin Kellman, Shanzhong Li, Daniel C. Zielinski, and Nathan E. Lewis. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Systems*, 4(3):318–329.e6, March 2017.

[94] Daniel Machado and Markus Herrgård. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Computational Biology*, 10(4):e1003580, April 2014.

[95] Daniel R Rhodes and Arul M Chinnaiyan. Integrative analysis of the cancer transcriptome. *Nature Genetics*, 37(S6):S31–S37, June 2005.

[96] SemidÃ¡n Robaina EstÃ©vez and Zoran Nikoloski. Generalized framework for context-specific metabolic model extraction methods. *Frontiers in Plant Science*, 5, September 2014.

[97] Scott A. Becker and Bernhard O. Palsson. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Computational Biology*, 4(5):e1000082, May 2008.

[98] Hadas Zur, Eytan Ruppin, and Tomer Shlomi. iMAT: an integrative metabolic analysis tool. *Bioinformatics*, 26(24):3140–3142, December 2010.

[99] Livnat Jerby, Tomer Shlomi, and Eytan Ruppin. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*, 6(1):401, January 2010.

[100] Anne Richelle, Austin W. T. Chiang, Chih-Chung Kuo, and Nathan E. Lewis. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLOS Computational Biology*, 15(4):e1006867, April 2019.

[101] Sangbum Lee, Chan Phalakornkule, Michael M Domach, and Ignacio E Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers & Chemical Engineering*, 24(2-7):711–716, July 2000.

[102] Rasmus Agren, Sergio Bordel, Adil Mardinoglu, Natapol Pornputtapong, Intawat Nookaew, and Jens Nielsen. Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT. *PLoS Computational Biology*, 8(5):e1002518, May 2012.

[103] Charles Darwin and Charles Darwin. *The origin of species, 1876*. Number v. 16 in The works of Charles Darwin. New York University Press, Washington Square, N.Y, 1988.

[104] Zahra Beheshti and Siti Mariyam Shamsuddin. A review of population-based meta-heuristic algorithm. *International Journal of Advances in Soft Computing and Its Applications*, 5:1–35, March 2013.

[105] Melanie Mitchell. *An introduction to genetic algorithms*. MIT Press, Cambridge, Mass., 1996. OCLC: 42854439.

[106] Lothar M. Schmitt. Theory of genetic algorithms. *Theoretical Computer Science*, 259(1-2):1–61, May 2001.

[107] Pradnya A. Vikhar. Evolutionary algorithms: A critical review and its future prospects. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 261–265, Jalgaon, India, December 2016. IEEE.

[108] Miguel Rocha, José P Pinto, Isabel Rocha, and Eugénio Ferreira. Optimization of bacterial strains with variable-sized evolutionary algorithms. In *2007 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*, pages 331–337. IEEE, 2007.

[109] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N Mendoza, Anne Richelle, Almut Heinken, Hulda S Haraldsdóttir, Jacek Wachowiak, Sarah M Keating, Vanja Vlasov, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nature protocols*, 14(3):639–702, 2019.

[110] Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. Cobrapy: constraints-based reconstruction and analysis for python. *BMC systems biology*, 7:1–6, 2013.

[111] Vítor Pereira, Fernando Cruz, and Miguel Rocha. Mewpy: a computational strain optimization workbench in python. *Bioinformatics*, 37(16):2494–2496, 2021.

[112] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024.

[113] Kristian Jensen, Joao GR Cardoso, and Nikolaus Sonnenschein. Optlang: An algebraic modeling language for mathematical optimization. *J. Open Source Softw.*, 2(9):139, 2017.

[114] Morihiko Sagawa, Hiroto Ohguchi, Takeshi Harada, Mehmet K Samur, Yu-Tzu Tai, Nikhil C Munshi, Masahiro Kizaki, Teru Hideshima, and Kenneth C Anderson. Ribonucleotide reductase catalytic subunit m1 (rrm1) as a novel therapeutic target in multiple myeloma. *Clinical Cancer Research*, 23(17):5225–5237, 2017.

[115] Eric L Van Nostrand, Adolfo Sánchez-Blanco, Beijing Wu, Andy Nguyen, and Stuart K Kim. Roles of the developmental regulator unc-62/homothorax in limiting longevity in caenorhabditis elegans. *PLoS genetics*, 9(2):e1003325, 2013.

[116] Kimberly Van Auken, Daniel Weaver, Barbara Robertson, Meera Sundaram, Tassa Saldi, Lois Edgar, Ulrich Elling, Monica Lee, Queta Boese, and William B Wood. Roles of the homothorax/meis/prep homolog unc-62 and the exd/pbx homologs ceh-20 and ceh-40 in c. elegans embryogenesis. 2002.

[117] Sean P Curran and Gary Ruvkun. Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS genetics*, 3(4):e56, 2007.

[118] Damian Smedley, Syed Haider, Benoit Ballester, Richard Holland, Darin London, Gudmundur Thorisson, and Arek Kasprzyk. Biomart–biological queries made easy. *BMC genomics*, 10:1–12, 2009.

[119] Zhuoqing Fang, Xinyuan Liu, and Gary Peltz. Gseapy: a comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics*, 39(1):btac757, 2023.

[120] Minoru Kanehisa, Miho Furumichi, Yoko Sato, Masayuki Kawashima, and Mari Ishiguro-Watanabe. Kegg for taxonomy-based analysis of pathways and genomes. *Nucleic acids research*, 51(D1):D587–D592, 2023.

[121] Teresa A Sherman, Sharath C Rongali, Tori A Matthews, Jason Pfeiffer, and Keith Nehrke. Identification of a nuclear carbonic anhydrase in caenorhabditis elegans. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1823(4):808–817, 2012.

[122] Zhongliang Hu, Yamei Luo, Yuting Liu, Yaru Luo, Liangce Wang, Shengsong Gou, Yuling Peng, Rui Wei, Da Jia, Yuan Wang, et al. Partial inhibition of class iii pi3k vps-34 ameliorates motor aging and prolongs health span. *PLoS Biology*, 21(7):e3002165, 2023.

[123] Jennifer Kremer, Cornelia Brendel, Elisabeth Karin Maria Mack, and Hildegard Isolde Dietlinde Mack. Expression of β-1, 4-galactosyltransferases during aging in caenorhabditis elegans. *Gerontology*, 66(6):571–581, 2020.

[124] Brad D Barrows, Stuart M Haslam, Larry J Bischof, Howard R Morris, Anne Dell, and Raffi V Aroian. Resistance to bacillus thuringiensis toxin in caenorhabditis elegans from loss of fucose. *Journal of Biological Chemistry*, 282(5):3302–3311, 2007.

[125] Fran Norflus, Jingnan Bu, Evon Guyton, and Claire-Anne Gutekunst. Behavioral analysis of the huntingtin-associated protein 1 ortholog trak-1 in caenorhabditis elegans. *Journal of neuroscience research*, 94(9):850–856, 2016.

[126] Kristina B Mercer, Sarah M Szlam, Erin Manning, Kim M Gernert, Walter W Walthall, Guy M Benian, and Claire-Anne Gutekunst. A c. elegans homolog of huntingtin-associated protein 1 is expressed in chemosensory neurons and in a number of other somatic cell types. *Journal of Molecular Neuroscience*, 37:37–49, 2009.

[127] Torunn Fiskerstrand, Dorra H'mida-Ben Brahim, Stefan Johansson, Abderrahim M'zahem, Bjørn Ivar Haukanes, Nathalie Drouot, Julian Zimmermann, Andrew J Cole, Christian Vedeler, Cecilie Bredrup, et al. Mutations in abhd12 cause the neurodegenerative disease pharc: an inborn error of endocannabinoid metabolism. *The American Journal of Human Genetics*, 87(3):410–417, 2010.

[128] Jacqueline L Blankman, Jonathan Z Long, Sunia A Trauger, Gary Siuzdak, and Benjamin F Cravatt. Abhd12 controls brain lysophosphatidylserine pathways that are deregulated in a murine model of the neurodegenerative disease pharc. *Proceedings of the National Academy of Sciences*, 110(4):1500–1505, 2013.

[129] Yanwei Wu, Shiya Cheng, Hongyu Zhao, Wei Zou, Sawako Yoshina, Shohei Mitani, Hong Zhang, and Xiaochen Wang. Pi 3 p phosphatase activity is required for autophagosome maturation and autolysosome formation. *EMBO reports*, 15(9):973–981, 2014.

[130] Matthew Daugherty, Boris Polanuyer, Michael Farrell, Michael Scholle, Athanasios Lykidis, Valérie de Crécy-Lagard, and Andrei Osterman. Complete reconstitution of the human coenzyme a biosynthetic pathway via comparative genomics. *Journal of Biological Chemistry*, 277(24):21431–21439, 2002.

[131] Simon Haeussler, Assa Yeroslaviz, Stephane G Rolland, Sebastian Luehr, Eric J Lambie, and Barbara Conradt. Genome-wide rnai screen for regulators of uprmt in caenorhabditis elegans mutants with defects in mitochondrial fusion. *G3*, 11(7):jkab095, 2021.

[132] Jouni Vesa, Elina Hellsten, Linda A Verkruyse, Laura A Camp, Juhani Rapola, Pirkko Santavuori, Sandra L Hofmann, and Leena Peltonen. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature*, 376(6541):584–587, 1995.

[133] Morwenna Y Porter, Mark Turmaine, and Sara E Mole. Identification and characterization of caenorhabditis elegans palmitoyl protein thioesterase1. *Journal of neuroscience research*, 79(6):836–848, 2005.

[134] Matthew J Edmonds and Alan Morgan. A systematic analysis of protein palmitoylation in caenorhabditis elegans. *BMC genomics*, 15:1–16, 2014.

[135] Katrin Stutz, Andres Kaech, Markus Aebi, Markus Künzler, and Michael O Hengartner. Disruption of the c. elegans intestinal brush border by the fungal lectin ccl2 phenocopies dietary lectin toxicity in mammals. *PLoS One*, 10(6):e0129381, 2015.

[136] Christin Riedinger, Michael Mendler, Andrea Schlotterer, Thomas Fleming, Jürgen Okun, Hans-Peter Hammes, Stephan Herzig, and Peter P Nawroth. High-glucose toxicity is mediated by aicar-transformylase/imp cyclohydrolase and mitigated by amp-activated protein kinase in caenorhabditis elegans. *Journal of Biological Chemistry*, 293(13):4845–4859, 2018.

[137] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.

[138] Tim J Schulz, Kim Zarse, Anja Voigt, Nadine Urban, Marc Birringer, and Michael Ristow. Glucose restriction extends caenorhabditis elegans life span by inducing mitochondrial respiration and increasing oxidative stress. *Cell metabolism*, 6(4):280–293, 2007.

[139] Michael Ristow and Kim Zarse. How increased oxidative stress promotes longevity and metabolic health: The concept of mitochondrial hormesis (mitohormesis). *Experimental gerontology*, 45(6):410–418, 2010.

[140] Patrick C Bradshaw. Cytoplasmic and mitochondrial nadph-coupled redox systems in the regulation of aging. *Nutrients*, 11(3):504, 2019.

[141] Christopher B Newgard. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell metabolism*, 15(5):606–614, 2012.

[142] Fan Jia, Mingxue Cui, Minh T Than, and Min Han. Developmental defects of caenorhabditis elegans lacking branched-chain α-ketoacid dehydrogenase are mainly caused by monomethyl branched-chain fatty acid deficiency. *Journal of Biological Chemistry*, 291(6):2967–2973, 2016.

[143] Christopher Hine, Eylul Harputlugil, Yue Zhang, Christoph Ruckenstuhl, Byung Cheon Lee, Lear Brace, Alban Longchamp, Jose H Treviño-Villarreal, Pedro Mejia, C Keith Ozaki, et al. Endogenous hydrogen sulfide production is essential for dietary restriction benefits. *Cell*, 160(1):132–144, 2015.

[144] Yuehua Wei and Cynthia Kenyon. Roles for ros and hydrogen sulfide in the longevity response to germline loss in caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 113(20):E2832–E2841, 2016.

[145] Peng Yao and Paul L Fox. Aminoacyl-trna synthetases in medicine and disease. *EMBO molecular medicine*, 5(3):332–343, 2013.

[146] Tianlin Zheng, Qiang Luo, Chengxuan Han, Jiejun Zhou, Jianke Gong, Lei Chun, XZ Shawn Xu, and Jianfeng Liu. Cytoplasmic and mitochondrial aminoacyl-trna synthetases differentially regulate lifespan in caenorhabditis elegans. *Iscience*, 25(11), 2022.

[147] Melissa J Silvestrini, Joseph R Johnson, Anita V Kumar, Tara G Thakurta, Karine Blais, Zachary A Neill, Sarah W Marion, Victoria St Amand, Robert A Reenan, and Louis R Lapierre. Nuclear export inhibition enhances hlh-30/tfeb activity, autophagy, and lifespan. *Cell reports*, 23(7):1915–1921, 2018.

[148] Peter J Stambrook, Michael K Dush, John J Trill, and Jay A Tischfield. Cloning of a functional human adenine phosphoribosyltransferase (aprt) gene: identification of a restriction fragment length polymorphism and preliminary analysis of dnas from aprt-deficient families and cell mutants. *Somatic cell and molecular genetics*, 10:359–367, 1984.

[149] Blake R Wilde, Nishma Chakraborty, Nedas Matulionis, Stephanie Hernandez, Daiki Ueno, Michayla E Gee, Edward D Esplin, Karen Ouyang, Keith Nykamp, Brian Shuch, et al. Fh variant pathogenicity promotes purine salvage pathway dependence in kidney cancer. *Cancer discovery*, 13(9):2072–2089, 2023.

Part III
Appendices

Appendix A
Details of results

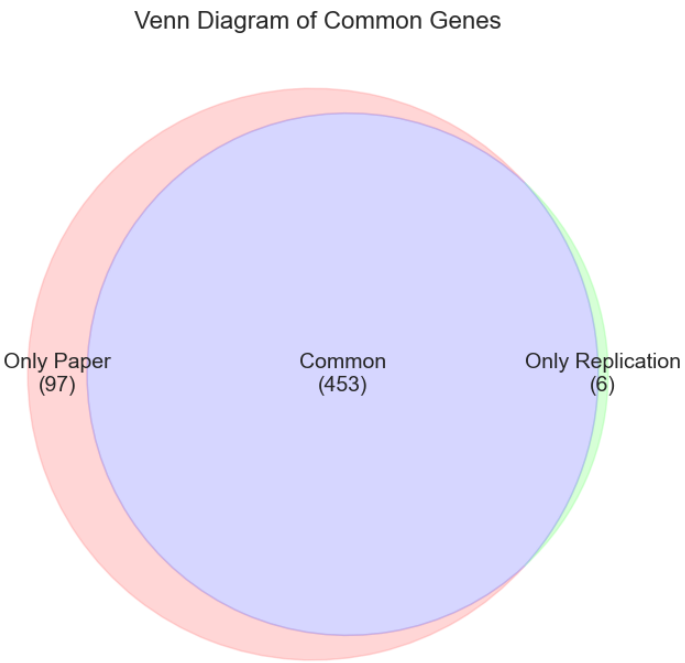


Figure 20: Venn diagram showing the overlap of top-ranked genes between the original study and our replication for *RRM2*.

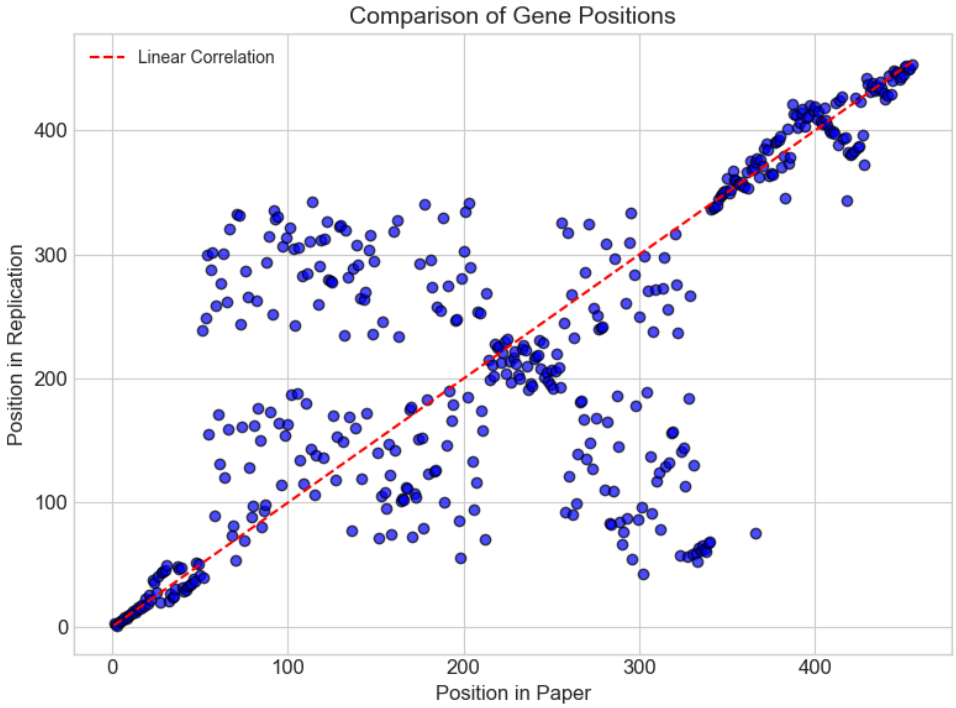


Figure 21: Correlation plot of gene rankings between the original study and our replication for *RRM2*.

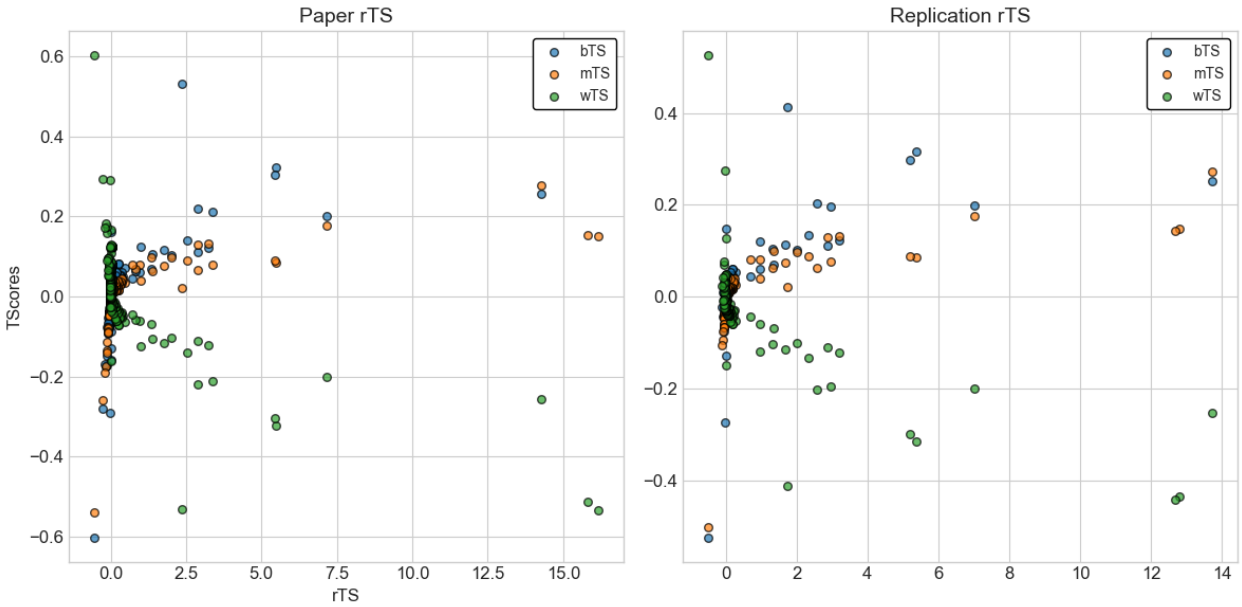


Figure 22: Comparison of rTS values between the original study and our replication for *RRM2*.

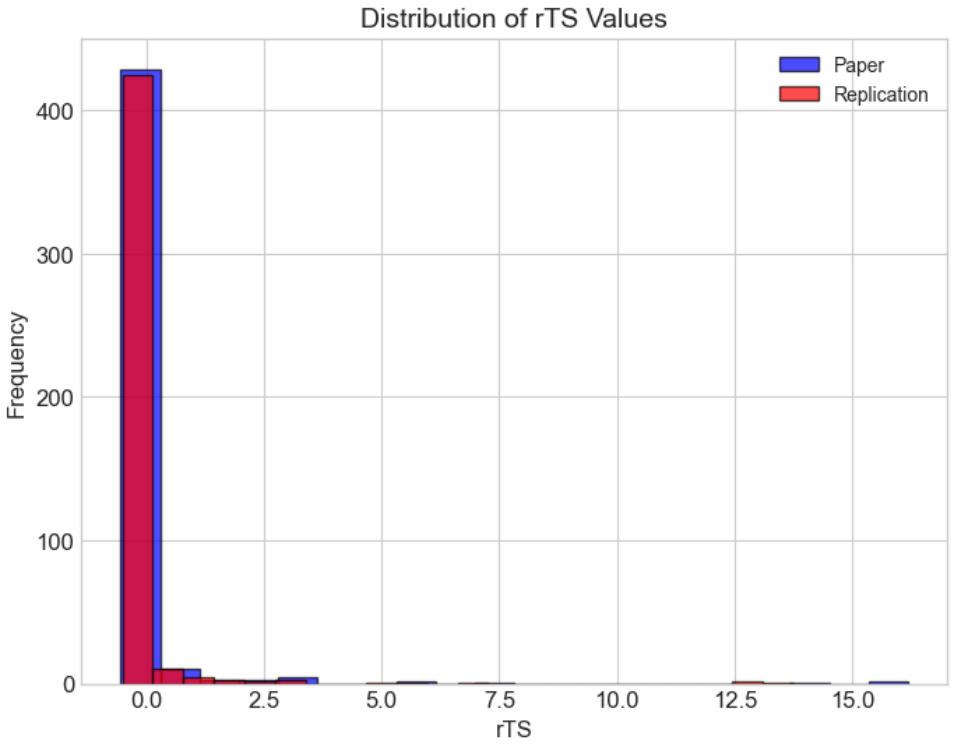


Figure 23: Distribution of rTS values for *RRM2*.