

DISS. ETH NO. 27022

**METABOLIC NETWORK ANALYSIS AND ITS
APPLICATION IN UNDERSTANDING THE
BIOLOGY OF AGING**

A dissertation submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
SUDHARSHAN RAVI

Master of Science
ETH Zurich

Born on 21 March 1991
Citizen of India

accepted on the recommendation of
Prof. Dr. Andrew deMello, examiner
Prof. Dr. Rudiyanto Gunawan, co-examiner
Prof. Dr. Paolo Arosio, co-examiner

2020

Sudharshan Ravi: *Metabolic Network Analysis and its Application in Understanding the Biology of Aging*, © 2020

DOI: [10.3929/ethz-b-000484454](https://doi.org/10.3929/ethz-b-000484454)

Everything matters.

Nothing is important.

ACKNOWLEDGEMENTS

I would like to begin by expressing my immense gratitude to Prof. Dr. Rudiyanto Gunawan for his invaluable support and guidance throughout my doctoral studies. His unwavering supervision could only be matched by his remarkable patience and understanding he had shown through some of my arduous times. I feel privileged to have had the opportunity to work closely with him. Despite the relocation from ETH Zurich, Switzerland, to the University at Buffalo, USA, I have relished the challenge. Dr. Gunawan has been pivotal in ensuring the steady progress of my understanding and knowledge in my research area. His strive for scientific perfection and fundamental research had inspired me to soldier through times when I questioned myself. Getting to know him on a personal level only strengthened my admiration for him.

I would like to sincerely thank Prof. Dr. Andrew deMello for his extraordinary kindness in welcoming me to his group. Although barely knowing me, his flexibility and understanding had ensured that I could defend my doctoral dissertation. In our occasional conversations, he instilled a sense of security, the importance of which cannot be overstated enough given the global landscape in 2020. Together, Ms. Jennifer Gassmann and Ms. Ingrid Gröbli, have collectively made the administrative and bureaucratic processes involved in my transitory phase, splitting time between ETH Zurich and University at Buffalo less forgiving, for which I am grateful. Thank you to Prof. Dr. Paolo Arosio for agreeing to co-referee my doctoral work. His swift and timely response to my emails, along with his adaptability, had kept my plans on track.

My time in CABSEL (Dr. Gunawan's Chemical and Biological Systems Engineering Laboratory) has been enriching. I had the privilege to interact and work with impeccable collaborators. Thank you, Prof. Dr. Jan Gruber and Dr. Emelyne Teo, for your consideration and trust during our many virtual conversations. Thank you, Prof. Dr. Lawrence Rajendran, for your words of support every time I interacted with you. Thank you, Prof. Dr. Stelios Andreadis, for your continued faith. Thank you, Prof. Dr. Andràs Paldi and

Romuald Parmentier, for your tolerance. Thank you Dr. Wenjuan Zha and Dr. Amit Lugade. My work would not have been possible without the timely assistance from all these teachers from whom I have learned an immense amount of knowledge and fortitude.

The last four years have been filled with highs and lows. Starting from an extreme sense of excitement when CABSEL opened its doors for me and through an agonizing year recovering from knee surgery. I am indebted to my friends in Zurich. They have quite literally supported me through the good and the bad times. I am delighted to have cultivated such numerous and sensational friendships that I would not have enough space to print every name. I am confident that those who know me know that I could not have gotten this far without your care and support. The emotional roller coaster of a ride only seemed to stretch on during my stay in Buffalo. Thank you, Mr. Burrito. Despite not being my cat, you have been my singular source of comfort through the depressive phases of my time in the USA. With uncertain times surrounding the Coronavirus pandemic and my own infection, you have single-handedly lifted my spirits.

Words could not do justice in thanking my family. Being worlds apart in more ways than we have commonalities, your unconditional love continues to be and will always be my greatest strength. To my Mum and Dad, who have sacrificed everything to enable me to succeed, your benevolence has shaped my character, for which I am eternally thankful. To my aunts and uncles, thank you for your ever-present encouragement and direction. To my cousins, thank you for your exuberance.

I can proudly say that my journey has been filled with sleepless nights, countless cups of coffee, innumerable days of self-doubt, boundless simulation trials, untold stories of personal conflicts, and finally, here, to a state of content with giving my everything. I feel fulfilled that any gaps in my work will not be for lack of effort.

Finally, and yet by no means insignificantly, to a person that had taught me love, respect, and self-worth. Thank you, Suzan! Where would I be without you?

ABSTRACT

A fundamental aim of life science is to understand the functioning of an organism and draw system-level connections between its genotype and behavior. Among the most significant types of interaction networks, cellular metabolism is a well-established descriptor of the phenotype of an organism. It relies on thousands of enzymatic reactions that are often represented as a dense and enmeshed web of biochemical conversions. Inspecting the metabolic network in its entirety is crucial in attaining a holistic understanding of the underlying biological mechanism. Colossal research efforts in the post-genomic era have enabled the curation of metabolic networks of entire organisms. Concurrently, advancements in computational strategies and algorithms have led to the inception of countless tools that utilize the essential information in genome-scale metabolic models to attain valuable insights into the physiology of an organism.

Constraint-based modeling is a widely used and tested technique to model the metabolic network. Flux balance analysis (FBA) is the most popular constraint-based approach to predict intracellular metabolic flux distributions and network capabilities in genome-scale models. They are modeled on two fundamental assumptions. Firstly, that the cellular metabolism operates at a homeostatic condition, and secondly, that the cell typically organizes its metabolism to optimize a particular cellular objective. However, these assumptions significantly impair the applicability of the traditional flux balance analysis. Where the prediction accuracy is insufficient, additional constraints from omics data sources assist in obtaining biologically relevant inference.

In this work, we showcase the addition of transcriptomic data to gain a profound understanding of the early events in the progression of Alzheimer's disease model *Caenorhabditis elegans*. Transgenically expressing human amyloid-beta recapitulated the phenotypic disease response in the worms. Our analyses of the contextualized genome-scale metabolic network curated with the integration of experimentally derived gene expression data implicated metabolic alterations in Tricarboxylic Acid (TCA) cycle following

low-level amyloid-beta expression. Along with metabolomic and enzymatic assays, we show repression in alpha-Ketoglutarate dehydrogenase. Identification of metabolic dysfunction as an early event is paramount in formulating mitigating efforts.

Formulating a metabolic objective for use in constraint-based modeling is hazy, particularly for complex multicellular organisms. In addressing such concerns, we developed a computational algorithm, Δ FBA (delta-FBA), that focuses on the differences in the metabolic distribution between a pair of conditions. By formulating the mathematical problem to optimize for maximal consistency between the inferred flux alterations and integrated gene expression changes, Δ FBA predicts metabolic rewiring as an effect of genetic or environmental perturbations. We validated our strategy in a wide range of single-gene deletion knockouts and environmental modifications in *Escherichia coli*, where Δ FBA outperforms similar methods. Furthermore, our findings of metabolic changes in human diabetic subjects show the robustness of Δ FBA.

Akin to metabolism, aging is a complex biological process that necessitates a system-level analysis in unraveling its etiology and progression. Despite numerous efforts, the aging process in humans is far from being completely understood. The Genotype-Tissue Expression (GTEx) project represents an invaluable repository of information that is ideal for studying human aging by examining the differences in the gene expression of over seven hundred individuals. Our bioinformatics and metabolic network analysis of the transcriptome associated human aging with evolutionarily conserved hallmarks of aging. Additionally, we show that the temporal changes in gene expression significantly contribute to the aging process. Our findings suggest that persistent moderators of cell fate and early repressors of cellular energetics could play a pivotal role in the progression of the aging process, one that culminates in accelerated decline.

ZUSAMMENFASSUNG

Ein grundlegendes Ziel der Biowissenschaften ist es, das Funktionieren eines Organismus zu verstehen und auf der Systemebene Verbindungen zwischen seinem Genotyp und seinem Verhalten zu knüpfen. Unter den wichtigsten Arten der Interaktionsnetzwerke ist der Zellstoffwechsel ein gut etablierter Deskriptor für den Phänotyp eines Organismus. Dieser hängt von tausenden von enzymatischen Reaktionen ab, die häufig als dichtes und verwobenes Netz von biochemischen Umwandlungen dargestellt werden. Es ist von besonderer Bedeutung, das metabolische Netzwerk in seiner Gänze zu untersuchen, um ein holistisches Verständnis der zu Grunde liegenden biologischen Mechanismen zu erhalten. Immense Forschungsbemühen in der Post-Genomischen Ära haben die Kuration metabolischer Netzwerke von gesamten Organismen ermöglicht. Gleichzeitig haben Fortschritte rechnerischer Strategien und Algorithmen zur Gründung zahlreicher Werkzeuge geführt, die wiederum von essenziellen Informationen im Genome-Scale Metabolic Model Gebrauch machen mit dem Ziel wertvolle Einsichten in die Physiologie eines Organismus zu gewinnen.

Constraint-Based Modellierung ist eine weitverbreitete und erprobte Technik, um das metabolische Netzwerk abzubilden. Flux Balance Analysis (FBA) ist der beliebteste constraint-based Ansatz, um intrazelluläre metabolische Verteilungsflüsse und Netzwerkfähigkeiten in genome-scale Modellen vorherzusagen. Sie stützen sich auf zwei grundlegende Annahmen: Erstens, dass der zelluläre Stoffwechsel nach homöostatischen Bedingungen agiert und zweitens, dass die Zelle ihren Stoffwechsel typischerweise zur Optimierung eines bestimmten zellulären Zwecks anordnet. Diese Annahmen schränken jedoch die Anwendbarkeit der traditionellen Flux balance Analysen signifikant ein. Dort, wo die Vorhersagegenauigkeit ungenügend ist, helfen zusätzliche Einschränkungen von omics Datenquellen biologisch relevante Rückschlüsse zu erhalten.

In dieser Arbeit präsentieren wir die Ergänzung von transkriptomischen Daten, um ein tiefergreifendes Verständnis der frühen Geschehnisse im Voranschreiten der Alzheimerkrankheit im Modell bei *Caenorhabditis elegans*.

gans zu gewinnen. Transgen exprimierte menschliches Amyloid-Beta führte zu einer phenotypischen Krankheitsantwort in den Würmern. Unsere Analysen des kontextualisierten Genome-Scale Metabolic Network kuratiert mit der Integration von experimentell hergeleiteten Mengenbestimmung der Genexpression wiesen auf Stoffwechselveränderungen im Tricarboxylic Acid (TCA) Zyklus nach niedriger Amyloid-Beta Expression hin. Zusammen mit metabolischen und enzymatischen Proben zeigen wir eine Unterdrückung in Alpha-Ketoglutarat-Dehydrogenase. Die Identifikation der Stoffwechseldysfunktion als ein frühes Geschehnis steht an erster Stelle um Abmilderungsbemühungen auszuformulieren.

Einen metabolischen Zweck für den Gebrauch einer constraint-based Modellierung zu formulieren ist vage, besonders für komplexe multizelluläre Organismen. Um uns solchen Bedenken zuzuwenden, haben wir einen Algorithmus namens Δ FBA (delta-FBA) entwickelt, der sich auf die Unterschiede in der metabolischen Verteilung zwischen einem Paar von Bedingungen fokussiert. Indem ein mathematisches Problem formuliert wird, das die maximale Beständigkeit zwischen vermuteter Strömungsveränderung und integrierten Veränderungen in der Genexpression optimiert, sagt Δ FBA metabolische Neuverdrahtung als einen Effekt von genetischen und umgebungsbedingten Störungen voraus. Unsere Strategie haben wir anhand eines breiten Spektrums von Single-Gene-Deletion-Knockouts und Umgebungs-Modifikationen in *Escherichia coli* validiert, bei denen Δ FBA ähnliche Methoden übertrifft. Zudem zeigen unsere Ergebnisse die Robustheit von Δ FBA bei diabetischen Probanden in Bezug auf Stoffwechselveränderungen.

Ähnlich zum Stoffwechsel ist auch das Altern ein komplizierter biologischer Prozess, der eine Analyse auf Systemlevel erfordert, um dessen Ätiologie und Voranschreiten zu entdecken. Trotz zahlreicher Bemühungen ist der menschliche Alterungsprozess noch weit von einem vollständigen Verständnis entfernt. Das Genotype-Tissue Expression (GTEx) Projekt stellt einen wertvollen Informationsspeicher dar, der ideal für das Studium des menschlichen Alterns geeignet ist, indem Unterschiede in der Genexpression von über 700 Individuen untersucht werden. Unsere bioinformatischen und metabolischen Netzwerkanalysen vom Transkriptom brachte das menschliche Altern mit evolutionär konservierten Kennzeichen des Alterns in Verbindung. Darüber hinaus zeigen wir, dass die zeitlichen Veränderungen in der Genexpression signifikant zum Altern beitragen. Unsere Ergebnisse deuten darauf hin, dass sowohl anhaltende Moderatoren des Schicksals der Zelle als

auch frühzeitige Repressoren der zellulären Energetik eine entscheidende Rolle im Voranschreiten des Alterungsprozesses spielen, dieser wiederum kulminiert in einem beschleunigten Verfall.

CONTENTS

Abstract	vii
Zusammenfassung	ix
List of Figures	xvii
List of Tables	xxi
Nomenclature	xxiii
1 INTRODUCTION	1
1.1 Background	1
1.2 Motivation	4
1.3 Objectives	8
1.4 Thesis Outline	9
2 METABOLIC NETWORK ANALYSIS IN TRANSGENIC <i>CAENORHABDITIS ELEGANS</i> STRAIN EXPRESSING HUMAN AMYLOID-BETA	11
2.1 Abstract	11
2.2 Introduction	13
2.3 Materials and Methods	14
2.3.1 Nematode maintenance and RNAi treatments	14
2.3.2 RNA extraction	15
2.3.3 RNA-Seq data analysis	15
2.3.4 Sample extraction and metabolic profiling	16
2.3.5 Metabolic flux balance analysis	17
2.3.6 Flux variability analysis and inconsistent reactions	18
2.3.7 TCA enzyme activity assay	18
	xiii

2.3.8	Lipidomics analysis	19
2.3.9	Lipidomics mass spectrometric analysis and data processing	19
2.4	Results	20
2.4.1	Alterations in amino acid and tricarboxylic acid cycle metabolism in GRU102	20
2.4.2	Reduced alpha-Ketoglutarate dehydrogenase activity in GRU102	23
2.4.3	aKGDH knockdown in controls recapitulated metabolic detriments of GRU102	25
2.5	Discussion	28
2.6	Conclusion	30
3	ΔFBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA	31
3.1	Abstract	31
3.2	Introduction	33
3.3	Methods	35
3.3.1	Method formulation	35
3.3.2	Case studies: Data and implementation	38
3.3.3	Implementation of REMI	40
3.3.4	Performance evaluation	40
3.3.5	Metabolic subsystem enrichment analysis	41
3.4	Results	42
3.4.1	<i>Escherichia coli</i> response to genetic and environmental variations	42
3.4.2	Dysregulation of skeletal muscle metabolism in type - 2 diabetes	44
3.5	Discussion	47
3.6	Conclusion	50

4	MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING	51
4.1	Abstract	51
4.2	Introduction	53
4.3	Methods	55
4.3.1	Identification of overall ‘aging’ genes	55
4.3.2	Moving window analysis of the aging transcriptome	56
4.3.3	Calculation of effect sizes	56
4.3.4	Metabolic subsystem enrichment analysis	57
4.3.5	Functional enrichment analysis	57
4.3.6	Transcription factor enrichment analysis	58
4.4	Results	58
4.4.1	Human aging signatures are associated with conserved hallmarks	58
4.4.2	The aging transcriptome in humans show a predominantly weak impact	61
4.4.3	Transcriptional roadmap of aging reveals persistent and strong changes in cell fate regulators	63
4.4.4	Regulation of metabolic genes may headline changes in early life	66
4.4.5	Progressive and accelerated decline in mid to late-life	67
4.5	Discussion	69
4.6	Conclusion	72
5	CONCLUSIONS AND OUTLOOK	73
5.1	Summary of findings	73
5.2	Future research directions	75
A	SUPPLEMENTARY INFORMATION: METABOLIC NETWORK ANALYSIS IN TRANSGENIC <i>CAENORHABDITIS ELEGANS</i> STRAIN EXPRESSING HUMAN AMYLOID-BETA	77
A.1	Nutritional and metabolic parameters used in metabolic flux balance analysis	77

A.2	List of reactions that show directional inconsistency in flux variability analysis (FVA)	78
A.3	TCA cycle enzyme activities	78
A.4	Lifespan study of GRU102 in RNAi experiments	79
B	SUPPLEMENTARY INFORMATION: Δ FBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA	81
B.1	Alternative threshold criteria of minimum flux change magnitudes	81
B.2	Supplemental figures	82
C	SUPPLEMENTARY INFORMATION: MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING	87
C.1	Functional hallmarks of ‘aging’ genes in GTEx	88
C.2	Metabolic subsystem enrichment analysis of the aging genes	89
C.3	Moving window analysis of the aging genes	91
C.4	Metabolic subsystem enrichment analysis of genes in the moving windows using Δ FBA	92
C.5	Classification of genes in the moving window analysis	93
C.6	Transcription factor enrichment analysis of persistent and early changing genes	94
C.7	Metabolic subsystem enrichment analysis of temporally changing genes	95
	Bibliography	97

LIST OF FIGURES

1 INTRODUCTION

Figure 1.1	Pictorial representation of the Δ FBA algorithm	9
Figure 1.2	Workflow in the analysis of human aging	10

2 METABOLIC NETWORK ANALYSIS IN TRANSGENIC *CAENORHABDITIS ELEGANS* STRAIN EXPRESSING HUMAN AMYLOID–BETA

Figure 2.1	Metabolomics profile of GRU102 and its transgenic controls (GRU101)	23
Figure 2.2	Metabolic flux in day 4 GRU102 modeled by metabolic flux balance analysis	24
Figure 2.3	aKGDH knockdown recapitulated metabolic detriments of GRU102	28

3 Δ FBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA

Figure 3.1	Comparison of the performance of Δ FBA and REMI in predicting <i>E. coli</i> metabolic response	44
Figure 3.2	Prediction of metabolic flux changes in <i>E. coli</i> caused by changes in the carbon source using Δ FBA	45
Figure 3.3	Enriched metabolic subsystems (FDR<0.05) among T2D patients	46
Figure 3.4	Alterations in metabolite flux throughput in T2D patients as predicted by Δ FBA	48

4 MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING

Figure 4.1	Functional annotation of the aging genes	60
Figure 4.2	Classification of the genes in the moving window analysis	64
Figure 4.3	Persistent and early altered genes	66
Figure 4.4	Temporal alterations in biological pathways in human aging	68
Figure 4.5	Transcriptional roadmap of the human aging process	70
A SUPPLEMENTARY INFORMATION: METABOLIC NETWORK ANALYSIS IN TRANSGENIC CAENORHABDITIS ELEGANS STRAIN EXPRESSING HUMAN AMYLOID-BETA		
Figure A.1	TCA cycle enzyme activities	78
Figure A.2	Lifespan study of GRU102 in RNAi experiments	79
B SUPPLEMENTARY INFORMATION: ΔFBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA		
Figure B.1	Comparison of the performance of Δ FBA's threshold variants in predicting <i>E. coli</i> metabolic response	83
Figure B.2	Comparison of the performance of Δ FBA in predicting <i>E. coli</i> metabolic response using whole-genome transcriptome data and RT-PCR mRNA data	84
Figure B.3	Comparison of the performance of Δ FBA and REMI in predicting <i>E. coli</i> metabolic response using whole-genome transcriptome data	86
C SUPPLEMENTARY INFORMATION: MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING		
Figure C.1	Functional annotation of the aging genes classified based on the hallmarks of aging	88

Figure C.2	Metabolic subsystem enrichment analysis of the aging genes	89
Figure C.3	Metabolic subsystem enrichment analysis of the aging process using Δ FBA predictions	90
Figure C.4	Metabolic subsystem enrichment analysis of the genes in moving windows using Δ FBA predictions	92
Figure C.5	Transcription factor enrichment analysis of persistent and early changing genes	94
Figure C.6	Metabolic subsystem enrichment analysis of temporally changing genes	95

LIST OF TABLES

4 MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING

Table 4.1	Differentially expressed genes in the moving window analysis	62
-----------	--	----

A SUPPLEMENTARY INFORMATION: METABOLIC NETWORK ANALYSIS IN TRANSGENIC *CAENORHABDITIS ELEGANS* STRAIN EXPRESSING HUMAN AMYLOID-BETA

Table A.1	Nutritional and metabolic parameters used in metabolic flux balance analysis	77
-----------	--	----

Table A.2	List of reactions that show directional inconsistency in flux variability analysis (FVA)	78
-----------	--	----

C SUPPLEMENTARY INFORMATION: MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING

Table C.1	Effect sizes of genes in the moving window analysis	91
-----------	---	----

Table C.3	Classification of genes in the moving window analysis	93
-----------	---	----

NOMENCLATURE

Abbreviations

GEM	Genome-scale metabolic model
GPR	Gene-Protein-Reaction
COBRA	COnstraint-Based Reconstruction and Analysis
FBA	Flux Balance Analysis
FVA	Flux Variability Analysis
pFBA	Parsimonious Flux Balance Analysis
Δ FBA	Delta Flux Balance Analysis
MOMA	Minimization of Metabolic Adjustment
ROOM	Regulatory On/Off Minimization
mCADRE	Metabolic Context-specificity Assessed by Deterministic Reaction Evaluation
GIMME	Gene Moderated by Metabolism and Expression
iNIT	Integrative Network Inference for Tissues
iMAT	Integrative Metabolic Analysis Tool
MADE	Metabolic Adjustment by Differential Expression
FDR	False Discovery Rate
MBA	Model Building Algorithm
LBFBA	Linear Bound Flux Balance Analysis
REMI	Relative Expression and Metabolic Integration
MOOMIN	Mathematical explOration of Omics data on a Metabolic Network
ME-model	Metabolism and Expression model
ETFL	Expression and Thermodynamics Flux model
NRMSE	Normalized Root Mean Square Error

AD	Alzheimer's Disease
A β	Amyloid-beta
TCA	Tricarboxylic Acid
PPP	Pentose Phosphate Pathway
BCAA	Branched Chain Amino Acids
NGM	Nematode Growth Medium
EV	Empty Vector
AA	Amino Acid
AC	Acylcarnitines
OA	Organic Acids
FA	Fatty Acid
aKG	alpha-Ketoglutarate
aKGDH	alpha-Ketoglutarate Dehydrogenase
SCS	Succinyl-CoA Synthetase
MDH	Malate Dehydrogenase
LDH	Lactate Dehydrogenase
TAG	Triacylglyceride
OCR	Oxygen Consumption Rate
T2D	Type - 2 Diabetes
BCAA	Branched Chain Amino Acids
AMP	Adenosine Monophosphate
ADP	Adenosine Diphosphate
FAD	Flavin Adenine Dinucleotide
NAD	Nicotinamide Adenine Dinucleotide
FUDR	5-fluoro-2'-deoxyuridine
BHT	Butylated Hydroxytoluene
GTE _x	Genotype-Tissue Expression
ARD	Age - Related Diseases
FDR	False Discovery Rate
LMM	Linear Mixed Model
TF	Transcription Factor

ECM	Extracellular Matrix
TNFR	Tumor Necrosis Factor Receptor
GAS6	Growth Arrest-Specific Gene 6
DDB2	DNA Binding Protein-2
IGF	Insulin Growth Factor

Notations in Δ FBA

S	Stoichiometric matrix
v^C	Control (reference) flux distribution
v^P	Perturbed flux distribution
Δv	Intracellular flux difference
w_i	Weighting coefficient for reactions
Z^U	Decision variable for upregulated reactions
Z^D	Decision variable for downregulated reactions
Z^0	Decision variable for unchanging reactions
Φ^*	Maximum consistency of Δv with gene expression changes
R^U	Set of upregulated reactions
R^D	Set of downregulated reactions
μ_i	Threshold for positive flux changes
η_i	Threshold for negative flux changes
ρ	Uncentered Pearson Correlation Coefficient

INTRODUCTION

1.1 Background

Deciphering the role of individual components in an organism is paramount in determining its behavior. However, biological systems have vast differences between each other in terms of their architecture and the functionality of their components. Moreover, the tissue-specific contributions in an organism play an essential role in bestowing versatility. Thus, examining a small group of constituents in an organism provides an incomplete explanation, therefore, warranting a systems-based approach in understanding its behavior. An overarching theme of systems biology is the investigation of complex biological processes as an integrated interaction of its many components. Such connections are often represented as a dense and enmeshed web of biochemical conversions^[1,2].

Among the significant types of biochemical networks, cellular metabolism provides a direct description of the organism's phenotype. It is widely considered as one of the best indicators of the physiological state and the behavior of a cell^[3,4]. Metabolism relies on thousands of enzymatic reactions that form a highly interconnected network involving the progressive conversion of a substrate to form a product. A holistic understanding of the metabolism will require analyzing the metabolic network in its entirety. Fortunately, numerous efforts over the past two decades have resulted in modeling the metabolic network at the genome-scale level, including humans^[5–8]. Consequently, systems-based approaches have been applied to analyze these metabolic networks, not just in explaining experimentally observed biological processes but also in hypotheses formulation and predicting peculiar cellular operations^[9–13].

Genome-scale metabolic models (GEMs) are reconstructions describing the whole set of metabolic reactions encoded by the genome of an organism, providing a mechanistic link between the genotype and the phenotype^[14]. They computationally capture the stoichiometric-based, mass balanced biochemical reactions using the Gene-Protein-Reaction (GPR) associations generated in boolean formulations based on the mode of participation of an enzyme to catalyze a given reaction^[15]. Significantly, GEMs can be leveraged for the computationally effective predictions of the entire metabolic flux distribution using linear optimization techniques, most notably, Flux Balance Analysis (FBA)^[16].

FBA is a widely used constraint-based technique that operates under the assumptions of a steady-state condition to optimizes one or more metabolic objectives of the cell^[17,18]. However, the metabolic objective of multicellular systems is unknown. Furthermore, the feasible solutions obtained by FBA models exhibit mathematical degeneracy, describing a solution space rather than a single, unique distribution, thereby curtailing its predictive power^[19]. Lastly, enzymatic expression and activities differ among cell types and extra-cellular conditions. The pitfalls in predictive prowess of the constraint-based modeling technique can be alleviated with the integration of omics data, reaction kinetics, and reaction thermodynamics that can lead to a reduced solution space for the flux distribution and more precise predictions^[6,20-24].

The advent of bioinformatics and next-generation sequencing technologies has led to increasing instances of quantitative monitoring of biological molecules in a cost-effective manner^[25]. In this regard, genome-wide mRNA abundance profiling has garnered widespread popularity over other omics data integration with GEMs. ¹³C labeled metabolic flux measurements, despite being a direct descriptor of the phenotype, remain challenging to perform for a large number of fluxes to understand systems-wide perturbations. Metabolomics remains challenging for certain classes of molecules that are easily metabolized^[26]. Proteomics, like transcriptomics, provides a static snapshot of the genome-wide abundance of molecular species. However, mRNA concentration quantification can be obtained at lower costs and higher accuracy in a highly automated fashion^[27,28].

Despite the discordance between transcript levels and metabolic fluxes, several algorithms have utilized and integrated transcriptomic data to varying degrees of success in creating a context-specific network that sufficiently

describes the state of the organism^[6,21,29–33]. Recently, integration of differential transcriptomics to predict altered metabolism between pairs of conditions has improved prediction accuracy and speed, particularly when the focus is on metabolic rewiring as a consequence of disease, treatment, gene deletion, or environmental perturbation^[34,35]. These wide ranges of methods for integrating transcriptomics and differential transcriptomics have varying strengths, and the choice of method seems to depend on the application. There is a definite discernible trend in iterative improvements in integrative metabolic network analysis and its ability to predict cellular perturbations.

Metabolism plays a pivotal role in numerous diseases with varying pathologies, including diabetes, neurodegenerative disorders, and cancer. Aging is also a major risk factor and is a common denominator in diseases that show metabolic dysregulation^[36,37]. It is a complex multifactorial process marked by a progressive functional and physiological decline in an organism, culminating in death. Along with the greying world population and falling fertility rates, aging and age-related diseases pose a pertinent concern to the global healthcare system^[38]. Hereof, a better understanding of the biology of aging has the potential to extend healthy life expectancy, delay the onset of detrimental age-related changes, and reduce the socio-economic burden.

Aging and metabolism are closely connected. In fact, the most successful aging intervention strategy in the laboratory across different model organisms by far is caloric or dietary restriction^[39,40]. Age-related changes are observed across different levels in an organism. Notably, there exists considerable variability in these changes among different species and even among different individuals of the same species. Even within an individual, the disparity in the rates of age-related changes in different tissues can be noticed. The complexity in the onset and progression of the aging process has warranted a systems-oriented approach in efforts to understand the biological processes that contribute to increasing the vulnerability in an organism. Numerous studies have postulated theories to explain the different aspects of the aging process in humans^[41–48], albeit providing an incomplete picture on its own to explain the multi-faceted nature of human aging.

1.2 Motivation

Genome-scale metabolic models are ideal platforms for integrating transcriptomic data toward predicting metabolic flux profiles in multicellular organisms. Numerous studies have relied on incorporating mRNA abundance in metabolic models of several organisms for broadly two purposes. The first category deals with creating context-specific GEMs from a global reconstruction for a specific cell type under specific environmental conditions. In comparison, the second category involves using transcript levels in reducing the solution space of constraint-based optimization techniques, such as FBA, to improve flux predictions. These methods not only differ in the manner of implementation of gene expression data into the GEM. They also vary in the amounts of datasets that need to be incorporated for a reliable outcome.

In an effort to address the poor consensus in context-specific GEMs generated for model-driven hypotheses and intracellular flux predictions from a generic GEM, Richelle *et al.*^[49] compared six of the most popular algorithms, mCADRE^[50], fastCORE^[51], GIMME^[21], INIT^[6], iMAT^[30] and MBA^[52] in *Homo sapiens*. Besides mCADRE, which requires multiple sets of data to obtain a ubiquity score for the expression of a particular gene, the other methods operate by discretizing gene expression based on a user-defined threshold to retain reactions catalyzed by highly expressed genes while trying to remove those reactions associated with unexpressed transcripts. fastCORE tries to preserve a core set of reactions that can be inferred from the gene expression data. GIMME discretizes the gene expression into unexpressed and expressed transcripts. It finds a metabolic model consistent with a given metabolic objective while minimizing the number of reactions associated with lowly expressed genes. iMAT differs from GIMME by not requiring a metabolic objective to create a context-specific model. Instead, it finds an optimal trade-off between retaining high-expressed reactions and removing low-expressed reactions.

The INIT algorithm follows a similar approach to iMAT but varies by giving weights to every reaction. The weights could be inferred from Gene-Protein-Reaction (GPR) associations but could also be user-defined to preserve particular reactions from *a priori* information on the metabolic system. MBA retains reactions with high confidence scores. Confidence scores could be user-defined based on prior knowledge, literature-derived evidence scores,

or calculated using gene expression data. Essentially, there exists considerable variability in the content of the context-specific GEM generated using these strategies^[53]. Since they use different assumptions in their approach to pruning the reaction set, they result in GEMs with vast differences in size and consequently its functionality, prediction accuracy, and ultimately biological interpretation. The choice of the algorithm in building context-specific GEM is highly application-driven.

Curtailing the solution space obtained by a traditional FBA that maximizes for a defined metabolic task to offer better predictions often relies on modification to the linear flux bounds that are modeled into the GEM. One of the earliest efforts in incorporating transcriptomic data to improve the inference of metabolic flux distribution in *Saccharomyces cerevisiae* was proposed by Åkesson *et al.*^[54]. They had constrained the fluxes of reactions that were associated with lowly expressed genes to zero. E-Flux^[31] directly incorporates the gene expression data into the upper bounds of the reactions in *Mycobacterium tuberculosis*. The rationale behind their approach was to link the levels of mRNA to approximate the maximum available protein that can be used for catalyzing a particular reaction. Linear Bound FBA (LBFBA)^[55] adds expression-derived lower and upper bounds on reactions. The reaction-specific flux bounds are a linear function of the gene expression data.

Other methods have focused on modifications to the objective function to deduce flux distributions that maximize the correlation with the gene expression data or the consistency with changes in gene expression between a pair of conditions. Lee *et al.*^[56] proposed an alternative objective function that maximized the correlation between the observed gene expression abundance and the computationally inferred metabolic fluxes in *S. cerevisiae* while maintaining metabolic productivity. GX-FBA^[57] tries to scale the internal flux predictions in a GEM of a wild-type gram-negative bacterium, *Yersinia pestis*, obtained using a traditional FBA, with the aid of transcriptional changes in response to environmental perturbations for an improved prediction of fluxes in response to these changes. The newly defined objective function maximizes the instances when a flux change is in direct proportionality to its corresponding gene's expression changes.

Despite these numerous iterative advancements in the incorporation of transcriptomic data in a GEM to improve the prediction of metabolic flux

distributions, a significant improvement over parsimonious enzyme-usage FBA (pFBA)^[58] that maximizes the metabolic objective while also minimizing the absolute flux through the network, was not clearly observable^[29]. Surprisingly, pFBA does not involve the integration of omics data. Rather, it aims to reduce the generic FBA derived flux solution space by minimizing the L1-norm. Although this could suggest that the integration of transcriptomics data could potentially hamper flux predictions, it is essential to consider the bias involved in selecting only a small number of intracellular fluxes that could be measured using ¹³C labeling for comparison and calculation of predictive accuracy. Furthermore, recent advancements in integrating differential transcriptomics in predicting differences in the metabolic network wiring between pairs of conditions have improved the predictive capabilities of differences in intracellular metabolic fluxes.

Inferring alteration in metabolism between pairs of conditions, such as gene deletions, environmental perturbations, or response to treatment, is one of the primary focuses of metabolic network analysis. In surprising contrast, only a few studies focus their attention on predicting flux variations in response to genetic or environmental changes. These differential methods could potentially overcome difficulties in needing to create context-specific GEM by directly incorporating changes in the gene expression onto a generic GEM. Additionally, the need to optimize for a metabolic objective is eliminated. Biological cellular objective is difficult to evaluate, even for microbes under varying experimental conditions, and is often unknown for complex multicellular organisms. Rather, the differential methods focus on optimizing the consistency between observed changes in the gene expression and the predicted intracellular flux alterations. Emphasis on the flux variations effectively filters out unchanged reactions. In stark contrast to the number of algorithms developed to address the integration of transcriptomic data for context-specific GEM generation and improving the intracellular flux predictions, only a small number of methods focus on the direct integration and analysis of differential gene expression changes to infer metabolic rewiring.

Relative Expression and Metabolic Integration (REMI)^[35] integrates relative expression changes in *Escherichia coli* under varying genetic and environmental perturbation to predict intracellular flux profiles for the wild type and the perturbed condition together. The algorithm formulates a mixed-integer linear optimization that maximizes the consistency of predicted

fluxes for the condition pairs and relative gene expression. Zhu *et al.* [34] used a similar approach only to differ in the definition of when a change in a particular flux is considered consistent with its associated gene expression's change. Where REMI forces a flux to change directly proportional to the mapped gene expression, Zhu *et al.* used a softer constraint in essentially only optimizing for the directionality congruence. Both these methods try and predict the intracellular flux profile of a pair of conditions together. Thereby, the flux prediction for the reference condition could suffer from the same drawbacks of a traditional FBA, potentially amplifying in error for the inference of perturbed fluxes. The lack of mathematical formulations in coalescing differential gene expression and constraint-based modeling is surprising, considering the numerous algorithms that rely on singular quantification of mRNA abundance to predict intracellular flux. For these reasons, the formulation of a strategy to overcome these difficulties and could serve as a tool for integrating differential transcriptomics data in the context of a GEM of multicellular organisms, and thereby predicting changes in metabolic flux distribution, is invaluable.

Chronological age is one of the principal effectors of change in energy metabolism. Aging, akin to metabolism, contributes to human diseases with vastly differing pathologies such as cancers, cardiovascular and neurodegenerative disorders [59–61]. Aging is a complex multifactorial process that affects every organism on our planet. It is associated with progressive physiological decline and increased vulnerability, ultimately leading to mortality. Age-related disorders have gradually become the leading global causes of non-communicable disease-associated deaths. Every aspect of the cell and its functions are affected by aging, albeit at varying rates. A comprehensive set of 'Hallmarks of Aging' have been proposed for better understanding the changes in the cellular level and the functional interconnections between them [62]. However, each of the theories provides an incomplete explanation of the aging process. Given the complexity of the aging process, it is of little wonder that an array of studies has focused on discovering dynamic transcriptomic signatures that are dysregulated with chronological age. Despite these numerous efforts, the timing of these age-related changes in biological processes and their contributions to cellular response is still hazy. Identifying the biological processes that are temporally out of balance could enhance our understanding of the aging process and have thus far been largely overlooked.

Traditional strategies in analyzing differential transcriptomic signatures in aging have primarily utilized false discovery rate (FDR) statistic as a threshold in identifying genes that show changes with age^[63,64]. Although FDR serves as a valuable statistical marker in identifying essential genes, it does not reveal information on the magnitude of the relevance of a particular gene in its contribution to aging. Complementing the statistical significance score with effect sizes has been proposed to improve confidence in identifying significant genes with a large change, particularly with the variable in question, age^[65–67]. For these reasons, deciphering temporal changes in the aging transcriptome could greatly benefit from the augmentation of effect size measurements in segregating the biological variations with timing and magnitude.

1.3 Objectives

In our efforts to understand the importance and integration of transcriptomic data with GEMs to contextualize generic models and reliably predicting intracellular fluxes, we collaborated with Dr. Jan Gruber's group in analyzing metabolic rewiring in transgenic *Caenorhabditis elegans* expressing pan-neuronal human amyloid-beta (A β). Alzheimer's disease is a chronic neurodegenerative disease characterized by the formation of A β plaques in the brain. By expressing human A β peptides in neurons, they have previously modeled A β - induced mitochondrial dysfunction in a transgenic strain of *C. elegans*. We relied on the transcriptomic data measurements of healthy and diseased worms, intending to decipher a deeper understanding of the metabolic dysfunction. Our objective was to try and shed light on how A β drives mitochondrial failure in Alzheimer's disease model worms with the help of metabolic network analysis.

Building on the knowledge and pitfalls in creating context-specific GEMs for predicting intracellular fluxes, we aimed to develop a variant of the traditional FBA, to predict changes in flux profiles between a pair of conditions. Our target was to develop a set of mathematical formulations that could improve the integration of differential transcriptomics data with a GEM while also being applicable to a number of organisms. Additionally, we envisioned an open-source set of tools that could work in harmony with the popular MATLAB based GEM analysis toolbox, the COnstraint-Based Reconstruction and Analysis (COBRA) toolbox^[68]. Figure 1.1 captures our implementation

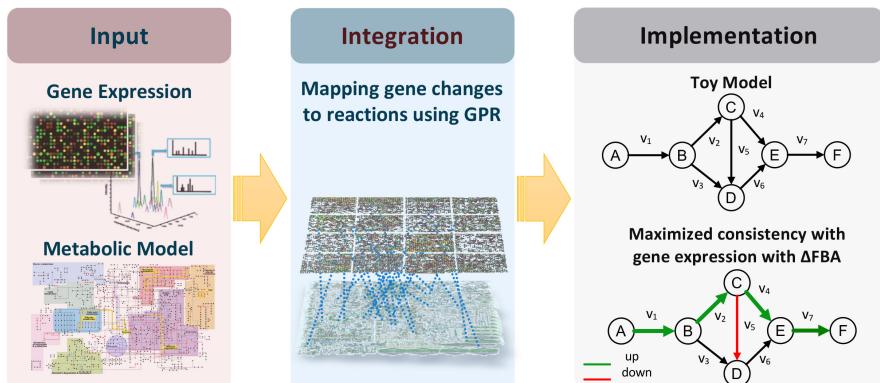


FIGURE 1.1: Pictorial representation of the Δ FBA algorithm that integrates changes in the RNA concentrations in a genome-scale metabolic model to predict flux alterations.

in focusing on inferring metabolic alterations in a genome-scale metabolic model.

In our quest to better understand the aging process, we focused on analyzing the dynamic transcriptional signatures of aging. We sought to elucidate the systems-wide biological pathways that are dysregulated as a result of the aging transcriptome derived from the Genotype-Tissue Expression (GTEx) project^[69]. Broadening our analysis, we set out to establish the mechanisms of temporal gene expression changes to explain the transcriptional roadmap of events during the aging process using a combination of bioinformatics and metabolic network analysis approaches (Figure 1.2).

1.4 Thesis Outline

This research work in this dissertation is fractioned into three parts. Firstly, in Chapter 2, we describe the transcriptomic data-driven contextualization and prediction of intracellular metabolic flux distribution in studying mitochondrial dysfunction and rewiring in transgenic *Caenorhabditis elegans* expressing pan-neuronal human amyloid-beta (A β). Our analysis sheds light on the alterations in Tricarboxylic Acid (TCA) cycle metabolism as an early event in A β -induced pathology.

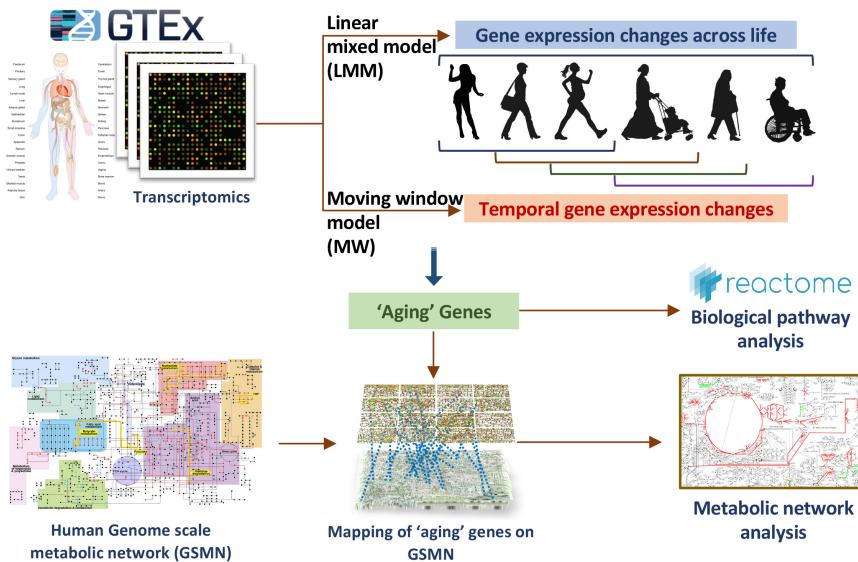


FIGURE 1.2: The workflow in the analysis of human aging leveraging on the transcriptomic data from Genotype-Tissue Expression (GTEx) project using pathway enrichment and metabolic network analysis.

Building on this platform, in Chapter 3, we present a computation algorithm, Δ FBA (delta-FBA), that was developed to improve the inferability of metabolic alterations as a response of an organism to changes in genetic or environmental perturbations. We demonstrate the ability of Δ FBA in providing a reliable prediction of changes in the metabolic network in prokaryotic and eukaryotic organisms.

Lastly, in Chapter 4, we perform bioinformatics and metabolic network analysis on the aging process in humans. We leverage the Genotype-Tissue Expression (GTEx) project data to reveal the transcriptional roadmap of events during the aging process to understand the biology behind it. Our results showed strong congruence with previous studies of aging on model organisms and are paramount in formulating mitigating strategies to delay the onset of aging.

Chapter 5 presents a summary of our findings with an optimistic look at future research directions that could further aid in understanding pressing biological questions in our time.

METABOLIC NETWORK ANALYSIS IN TRANSGENIC *CAENORHABDITIS ELEGANS* STRAIN EXPRESSING HUMAN AMYLOID-BETA

Experimental work presented in this chapter was carried out by Dr. Emelyne Teo in Prof. Dr. Jan Gruber's lab in Yale-NUS, Singapore. The results presented here have been published in *eLIFE*^[70].

2.1 Abstract

Inferring altered metabolism is at the core of life sciences and assists in understanding the effect of physiology and pathogenesis. Predicting biologically relevant intracellular metabolic flux distribution is complicated, particularly when dealing with genetic and environmental perturbations. Experimentally obtained omics data is often used to improve the accuracy of flux predictions by the inclusion of constraints in a constraint-based model. This chapter assesses the metabolic deficiencies in Alzheimer's disease (AD) model transgenic *Caenorhabditis elegans*. AD is a debilitating neurodegenerative disease affecting the elderly worldwide. Mitochondrial dysfunction has been postulated as a pivotal event in the etiology of AD. Transgenic worms expressing human amyloid-beta (A β) specifically in the neurons (GRU102) has been shown to recapitulate AD phenotype with

low-level toxicity of A β peptides. By integrating metabolomics, transcriptomics, and computational modeling, we identify alterations in Tricarboxylic Acid (TCA) cycle metabolism following even low-level A β expression. In particular, GRU102 showed reduced activity of a rate-limiting TCA cycle enzyme, alpha-Ketoglutarate dehydrogenase (aKGDH). Our results point to metabolic dysfunction as an early event in A β -induced pathology and a promising target for intervention.

2.2 Introduction

Understanding the genotype to predict cellular function is at the core of life sciences. Genome-scale metabolic (GEM) models are comprehensive repositories of biochemical reactions occurring in a cell, providing a link between the genotype and phenotype of the organism. Reconstructed GEM models contain carefully curated information on all known metabolites in an organism and its associated reactions based on genome annotations and experimentally obtained evidence^[14]. In applying these networks, the models are converted to a mathematically usable stoichiometric matrix that is central to the constraint-based modeling approaches, such as the Flux Balance Analysis (FBA)^[16]. Commonly, these approaches assume the system to be at a homeostatic state, i. e. the total production rate of every metabolite must be equal to its consumption flux. FBA identifies an optimal flux distribution given an assumed cellular objective^[17,18].

Besides the concern of unknown metabolic objectives for complex multicellular organisms, FBA often suffers from degenerate solutions due to an unconstrained solution space arising from an under-determined system^[19]. Additionally, enzyme expression and activities are tissue and condition-specific. For reasons when intracellular fluxes are insufficiently predictive with FBA, integration of additional data, such as omics data, reaction kinetics, and thermodynamics, has become popular and is continually being developed. Owing to its high availability, decreasing cost, and providing a genome-wide snapshot of mRNA abundance, transcriptomic data integration has been a significant direction of algorithms and methods development in improving the predictive power of constraint-based models^[21].

This chapter presents a case study of transcriptomic data integration in a GEM for hypothesis generation and creating a better understanding of the underlying biological mechanisms. Here, we leverage mRNA abundance quantification to uncover metabolic dysfunction in *Caenorhabditis elegans* expressing pan-neuronal human amyloid-beta (A β). Alzheimer's disease (AD) is a chronic neurodegenerative disease that is characterized by the formation of A β plaques^[71]. Although a neuronal disorder, mitochondrial dysfunction has been observed across the entire organism and not just in cells obtained from the brain^[72,73]. Recently, findings point toward a critical primary role played by oxidative stress and metabolic perturbations in the progression of AD in model organisms^[74-76]. Importantly, these changes

were observed well before the appearance of A β plaques. A β , mostly in its oligomeric form, can induce oxidative stress by inserting itself into the lipid bilayer, leading to protein oxidation^[77]. A β -induced oxidative stress arising from alterations to metabolic enzymes, particularly in the glucose metabolism, contributes to impaired energetics, terminating in synaptic dysfunction and ultimately, neuronal death^[78-80].

C. elegans is an indispensable model organism in AD understanding due to its short lifespan and a well-characterized nervous system. Moreover, the ability to create various transgenic AD strains by expressing the human A β -peptides in different cellular locations and under different promoter controls have led to worms with AD-like pathology with varying disease severity that could be useful in studying the different aspects of AD^[81-83]. A novel transgenic strain of *C. elegans* (GRU102) with a constitutive pan-neuronal expression of low-levels of human A β_{1-42} was designed to study the metabolic alterations in the early stages of AD. GRU102 displayed milder phenotypes compared to other transgenic strains expression A β in neurons, including failure in chemotaxis, neuromuscular and behavior defects. Despite the detectable presence of A β aggregates only at late-life (day 12), defects in energy metabolism were noticeable in young worms (day 4). Our findings link A β expression to Tricarboxylic acid (TCA) cycle impairment in the early life cycle of *C. elegans* (day 4).

2.3 Materials and Methods

2.3.1 Nematode maintenance and RNAi treatments

The transgenic strains, GRU101[myo-2p::yfp] and GRU102[myo-2p::yfp + unc-119p::A β_{1-42}], previously reported in Fong *et al.*^[84] were used in all experiments. These animals were raised at 20 °C throughout the experiment and transferred to nematode growth medium (NGM) plate supplemented with 250 μ M 5-fluoro-2'-deoxyuridine (FUDR; Sigma-Aldrich, St. Louis, USA) to prevent progeny production. Age-synchronized animals were obtained for all experiments via hypochlorite bleaching. Young animals refer to day four post-bleaching while old animals refer to day 12 post-bleaching. For RNAi treatment, RNAi feeding clones were obtained from the *C. elegans* ORF-RNAi Library (Vidal) from Source BioScience. L4 animals were fed with HT115(DE3) bacteria expressing either the empty vector (EV) or RNAi

clone T22B11.5 (for aKGDH knockdown), on NGM plates containing 2 mM IPTG and 100 µg ml⁻¹ Carbenicillin.

2.3.2 RNA extraction

Freshly prepared bacteria (OP50 *E. coli*) were spotted on 94 cm NGM agar plates on the previous evening and allowed to dry. Synchronized, young adult worms were transferred to fresh plates. For maintaining synchronized populations, FUdR was added to NGM media. After two days of treatment with drug or vehicle, adult worms were washed off the plates into 15 ml tubes and then washed several times with fresh buffer until a clear solution was obtained. Clean worm pellets were then frozen and later used for RNA extraction. Total RNA was isolated using Qiagen RNAeasy micro kit (Qiagen, Hilden, Germany) following the standard protocol.

2.3.3 RNA-Seq data analysis

Following RNA extraction, RNA was quantified photometrically with Nanodrop 2000 and stored at -80 °C until use. The integrity of total RNA was measured by Agilent Bioanalyzer 2100. For library preparation, an amount of 2 mg of total RNA per sample was processed using Illumina's RNA Sample Prep Kit following the manufacturer's instruction (Illumina, San Diego, CA, USA). All single drugs were processed together and multiplexed onto one lane to minimize the batch effect. The different drug combinations were processed into two multiplexed lanes. In total, we ran three lanes in parallel with 20 samples in each and we had untreated N₂ controls in each lane. Libraries were sequenced using Illumina HiSeq4000 sequencing platform (Illumina, San Diego, CA, USA) in a paired-end read approach at a read length of 150 nucleotides. Sequence data were extracted in FastQ format. The RNAseq reads from each sample were mapped to the reference *C. elegans* transcriptome (WBcel235) with kallisto (v0.44.0)^[85] and sequence-based bias correction. The estimated counts were imported from kallisto to the R environment (v3.5.1) and summarized to gene-level using the tximport package (v1.6.0)^[86]. The DESeq2 package (v1.30.0) was used to identify differentially expressed genes (DEGs) in all our analysis for a significance threshold α of 0.05 after correcting for multiple hypothesis testing through Independent Hypothesis Weighting (IHW package v1.6.0)^[87].

2.3.4 Sample extraction and metabolic profiling

Nematodes were collected in M9 buffer with two washes to remove excess bacteria. Nematode pellet was obtained by centrifugation at 4 °C, snap-frozen in liquid nitrogen and stored in -80 °C until homogenization was performed. Nematode pellet was homogenized in 50% acetonitrile, 0.3% formic acid. Methods of extraction and metabolic profiling for amino acid (AA), acylcarnitines (AC) and organic acids (OA) are performed as described in Newgard *et al.* [88], Sinha *et al.* [89], and Muoio *et al.* [90].

For AC and AA extraction, 100 µl of homogenate was extracted using methanol. The AC and AA extracts were derivatised with 3 M Hydrochloric acid in methanol or butanol (Sigma Aldrich, USA) respectively, dried and reconstituted in methanol for analysis in LC-MS. For OA extraction, 300 µl of homogenate was extracted with ethylacetate, dried and derivatized with N,O-Bis(trimethylsilyl)trifluoroacetamide, with protection of the alpha keto groups using ethoxyamine (Sigma Aldrich, USA).

AC measurements were made using flow injection tandem mass spectrometry on the Agilent 6430 Triple Quadrupole LC/MS system (Agilent Technologies, CA, USA). The sample analysis was carried out at 0.4 ml/min of 80/20 Methanol/water as mobile phase, and injection of 2 µl of sample. Data acquisition and analysis were performed on Agilent MassHunter Workstation B.06.00 Software.

Methods of AA analysis were modified from Newgard *et al.* [88] and Sinha *et al.* [89]. Briefly, AA were separated using a C8 column (Rapid Resolution HT, 4.5 × 50 mm, 1.8 µm, Zorbax SB-C8) on a Agilent 1290 Infinity LC system (Agilent Technologies, CA, USA) coupled with quadrupole-ion trap mass spectrometer (QTRAP 5500, AB Sciex, DC, USA). Mobile phase A (10/90 Water/Acetonitrile) and Mobile phase B (90/10 Water/Acetonitrile) both containing 10 µM of Ammonium formate were used for chromatography separation. The LC run was performed at a flow rate of 0.6 ml/min with initial gradient of 20% B for 0.5 min, then ramped to 100% B in 2.5 min, maintained for 0.5 min, followed by re-equilibrating the column to the initial run condition (20% B) for 2 min. All compounds were ionized in positive mode using electrospray ionization. The chromatograms were integrated using MultiQuant 3.0 software (AB Sciex, DC, USA).

Trimethylsilyl derivatives of OA were separated on a gas chromatography column (VF-1ms; 30 m x 0.25 mm x 1 μ m) by an Agilent Technologies HP 7890A and quantified by selected ion monitoring on a 5975C mass spectrometer using stable isotope dilution. The initial GC oven temperature was set at 70 °C, and ramped to 300 °C at a rate of 40 °C/min, and held for 2 min.

2.3.5 Metabolic flux balance analysis

We employed parsimonious flux balance analysis (pFBA) to evaluate the metabolic reaction fluxes in young transgenic control GRU101 and GRU102 animals. In pFBA^[91], metabolic fluxes of each metabolites are computed based on the mass balance around every metabolite in the metabolic network under the following assumptions: (1) that the metabolism is operating at steady state (i. e. no change in the metabolite concentrations over time), (2) that the organism/cell optimizes its metabolism according to a biological objective, and (3) that the total absolute flux in the metabolic network is minimized^[58]. We performed pFBA using built in functions in the COBRA toolbox^[68] in MATLAB. Given the life history of *C. elegans*, we chose the maximization of biomass production for growth and storage^[91].

We curated condition-specific metabolic network models for the GRU101 and GRU102 animals by pruning the *C. elegans* genome-scale metabolic model *iCEL* 1273^[91] based on strain-specific transcriptomic data, using the Gene Inactivity Moderated by Metabolism and Expression (GIMME)^[21] algorithm. Briefly, starting from the genome-scale metabolic network (*iCEL* 1273), the algorithm removes reactions whose mRNA transcript levels are below a pre-determined threshold. We set the threshold at the 5th percentile of the normalized transcript level for each gene. The algorithm then uses flux balance analysis (FBA) to ensure that the reduced model is able to achieve the same optimal biological objective value as the full model. In the event that the resulting model is unable to achieve the same biological objective, GIMME determines the smallest set of previously removed reactions to be added back into the reduced model to replicate the optimal biological objective value of the full model. We confirmed that changing the threshold in the range between 1 and 50th percentile does not significantly affect the results of our analysis.

Using the condition-specific metabolic models for GRU101 and GRU102 animals, we then applied pFBA to determine metabolic fluxes based on the maximization of biomass production. We started by analyzing the metabolic fluxes of the controls (GRU101), using experimentally observed bacterial consumption rate as a constraint on the bacterial intake rate^[91–93]. We used the results of the pFBA of GRU101 to set constraints on bacterial intake and oxygen consumption rates in the pFBA of GRU102. More specifically, we set the constraint on the bacterial intake rate of GRU102 by scaling the constraint of GRU101 according to the ratio of pharyngeal pumping rates between the two strains. Meanwhile, the oxygen consumption rate of GRU102 is constrained to a scalar multiple of the oxygen consumption rate of GRU101 as determined by pFBA, based on the ratio of experimentally observed oxygen uptake rates between the two strains.

2.3.6 Flux variability analysis and inconsistent reactions

We performed flux variability analysis (FVA)^[94] on the GRU101 and GRU102 metabolic models to determine the range (interval) of flux values for each reaction, within which one can vary the flux of a reaction – keeping the others constant – without changing the optimal biological objective value. We implemented the FVA using in-built functions in the COBRA toolbox. We compared the results of FVA of GRU101 and GRU102 to identify any inconsistent reactions. Here, a reaction is inconsistent between the two animals when its flux interval according to FVA is located fully to the left (negative) or right (positive) of the origin for at least one of the strains, and the flux intervals between the two animals have values of opposite signs.

2.3.7 TCA enzyme activity assay

Mitochondrial extraction was carried out as described in Goo *et al.*^[95]. Mitochondria proteins were used for alpha-Ketoglutarate dehydrogenase (aKGDH), Succinyl-CoA Synthetase (SCS), Malate dehydrogenase (MDH) activity assays while cytosolic protein was used for Lactate dehydrogenase (LDH) assay. All assays were performed using colometric assay kits (BioVision Inc, San Francisco, USA) according to manufacturer's instruction. 20 µg of mitochondrial protein was used for aKGDH assay (Cat. No K678). 10 µg of mitochondrial protein was used for SCS assay (Cat. No K597). 2 µg of

mitochondrial protein was used for MDH assay (Cat. No K654). 10 µg of cytosolic protein was used for LDH (Cat. No K726) assay.

2.3.8 Lipidomics analysis

Approximately, 1500 nematodes per sample were collected, washed with M9 buffer and transferred to 2 ml polypropylene tubes containing 250 µl lysis buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.5 mM EDTA, 5% glycerol) and left on ice for 15 min followed by homogenization using a Bead beater (Precellys, France) maintained at 4 °C. Lipids extraction from the lysed samples was carried out by Folch's extraction. In order to minimize oxidation during and after extraction, 0.5% butylated hydroxytoluene (BHT) was added to the organic solvents. Samples were spiked with known amounts of internal standards (purchased from Avanti polar lipids, Alabaster, AL, USA) corresponding to each lipid class during the extraction to control lipid-class dependent differences in extraction and ionization. Internal standard used for triacylglyceride (TAG) was d5-TAG 48:0. To achieve separation into aqueous and organic phases, the samples were vortexed and centrifuged at 3000 rpm for 5 min. The lower phase was then transferred to a fresh centrifuge tube and centrifuged in a vacuum concentrator (SpeedVac, Thermo Savant, Milford, USA) until dry. The dried lipid extract was reconstituted in 50 µl methanol and kept at -80 °C until mass spectrometry analysis.

2.3.9 Lipidomics mass spectrometric analysis and data processing

A 6490 triple Quad Mass spectrometer (QqQ; Agilent, USA) coupled to a 1260-Ultra Performance Liquid chromatography (UPLC) was used for lipid quantification. ESI was used to ionize lipids. Each lipid molecular species was analyzed using a targeted multiple reaction monitoring (MRM) approach containing transitions for known precursor/product mass-to-charge ratio and retention times. The UPLC system was equipped with a Waters AC-QUITY BEH C18 column (1.0 × 100 mm) to separate the molecular species using gradient elution. Solvent A was acetonitrile/H₂O (60:40) with 10 mM ammonium formate while solvent B was isopropanol/acetonitrile (90:10) containing 10 mM ammonium formate. The flow rate was 0.4 ml/min and the column temperature 50 °C. Solvent B was set at 70% at injection and increased linearly to 100% in 14 min, retained at this value for 3 min, de-

creased to 10% in one minute and then retained there until the end of the gradient by 14 min. The eluent was directed to the ESI source of the mass spectrometer operated in the positive ion mode. The MS conditions were as follows. For ESI: gas temperature, 250 °C; gas flow, 14 l/min; sheath gas temperature, 50 °C; sheath gas flow, 11 l/min; and capillary voltage, 3,500 V.

LC-MS data obtained on the 6490 QqQ Mass spectrometer were subjected to data processing using MassHunter software (Agilent). The identification of a lipid species was based on mass and retention times (RT). Signal intensities of the targeted species were compared with the intensities from the spiked internal standards and the retention times for the various classes were matched. Data processing, included peak smoothing and integration of areas under the curves for each measured transition. The processed data were exported to excel and normalized to protein content as well as the internal standard. Fold changes were measured by comparison of the different conditions to the control sample and finally the Student's t test was performed to determine whether differences between the samples were statistically significant. ($p < 0.05$ was considered statistically significant).

2.4 Results

2.4.1 Alterations in amino acid and tricarboxylic acid cycle metabolism in GRU102

GRU102 suffer from early and significant defects in energy metabolism^[84]. In determining which energy substrates were significantly altered in GRU102, a targeted metabolomic approach was implemented to compare intermediary metabolites involved in energy production between age-matched GRU102 and their respective transgenic controls (GRU101). Specifically, levels of different Acylcarnitines (AC), Triacylglycerides (TAG), amino acids (AA) as well as of organic acids (OA), including TCA cycle intermediates, were measured.

AC are intermediates that are typically generated during the transport of activated long-chain fatty acids (FA) from the cytosol to mitochondria for FA oxidation^[96]. Even though most AC species are derived from FA, some AC species are formed through non-FA intermediates. For example, Propionyl-

carnitine (C3) and Isovalerylcarnitine (C5) are derived from the degradation of branched-chain AA while Acetylcarnitine (C2) is a common energy precursor derived from glucose and FA metabolism^[96,97]. Significantly, levels of the common energy precursor C2, were repressed in old GRU102 compared to their age-matched GRU101 controls (Figure 2.1A). Additionally, AA-derived AC species C3-carnitine was also significantly reduced in old GRU102 (Figure 2.1A), suggesting a low energy status and altered AA metabolism in old GRU102. In contrast, FA-derived AC species remained unchanged (Figure 2.1A) and in congruence with the experimental evidence showing similar levels of storage lipid, Triacylglycerides (TAGs), between GRU102 and GRU101 (Figure 2.1B). Lowered total AA levels in old GRU102 support the repression in AA metabolism (Figure 2.1C).

Glucogenic pathways use AA as substrate for energy production by producing pyruvate, alpha-Ketoglutarate (aKG), succinyl-CoA, fumarate or oxaloacetate, where they can subsequently fuel the TCA cycle^[98]. To further explore how lowered total AA levels relate to metabolic alterations, individual AA were classified based on their glucogenic routes they use in producing distinct TCA cycle substrates (oxaloacetate, pyruvate, aKG and succinyl-CoA). We calculated the percentage of AA that forms each of these four TCA metabolites by taking the sum of individual AA that forms the particular substrate divided by the total amount of AA. For Oxaloacetate, we took the levels of Aspartate over total AA levels; for Pyruvate, we took the sum of Alanine, Serine and Glycine over total AA levels; for aKG, we took the sum of Arginine, Histidine, Glutamine, Proline and Ornithine over total AA levels; for Succinyl-CoA, we took the sum of Methionine and Valine over total AA levels. This analysis of the AA profile demonstrated that, in comparison to age-matched controls, the percentage of glucogenic AA forming alpha-Ketoglutarate (aKG) was significantly increased in old GRU102 animals, whereas the proportion of AA forming pyruvate were reduced significantly (Figure 2.1D).

From the OA profiles, we found only an insignificant trend towards higher aKG levels in old GRU102 compared to GRU101 (Figure 2.1E). However, both of the TCA cycle intermediates, fumarate and malate, were significantly reduced in old GRU102 (Figure 2.1E). These observations suggest a specific disruption in the TCA cycle that affects some but not all of the TCA intermediates.

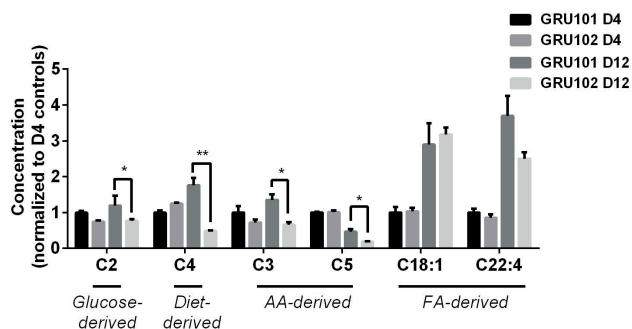
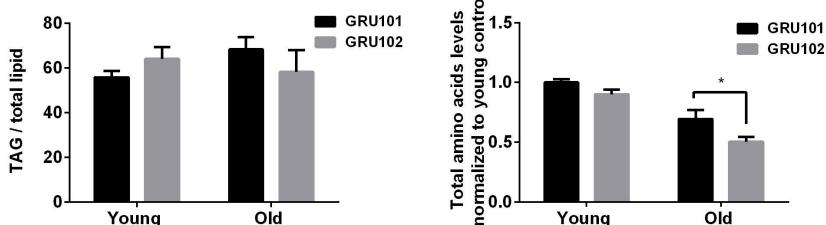
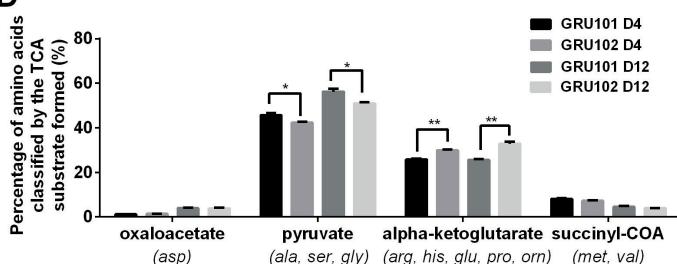
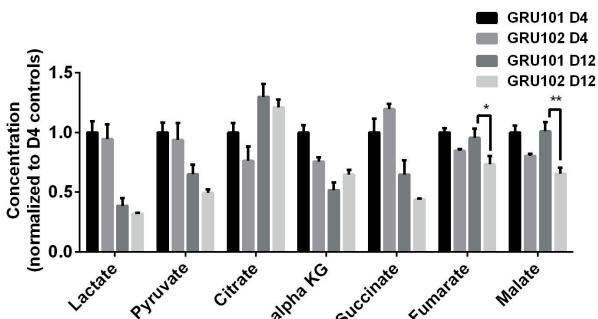
A**B****D****E**

FIGURE 2.1: Metabolomics profile of GRU102 and its transgenic controls (GRU101). (A) Acylcarnitines (AC) profile, (B) Triacylglyceride (TAG), (C) Total Amino acids (AA) level, (D) Percentage glucogenic AA, computed as [sum of AA forming the particular glucogenic substrate/total AA levels]. (E) Organic acids (OA) concentration of GRU101 and GRU102. All values were normalized to respective protein concentration and then to young GRU101 (Two-way ANOVA Dunnett's multiple comparisons test, $p < 0.05$, *; $p < 0.005$, **; $p < 0.001$, ***; $n = 3$ repeats per condition, with approximately 3000 animals per repeat collected from independent cohort).

2.4.2 Reduced alpha-Ketoglutarate dehydrogenase activity in GRU102

To help identify which reaction(s) in and around the TCA cycle were most strongly altered in GRU102, we next performed a computational metabolic flux balance analysis (FBA) using transcriptomics and metabolic data from young GRU101 and GRU102 animals as constraints. Modeling took into consideration gene expression data (RNAseq) and data related to nutritional uptake and metabolic rates, specifically pharyngeal pumping, oxygen consumption, and body size (Supplementary Table A.1). Gene expression data were used to curate *C. elegans* condition-specific metabolic networks by modifying a genome-scale metabolic model of *C. elegans* (*iCEL* 1273)^[91] using the Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm^[21]. Specifically, we performed parsimonious flux balance analysis (pFBA) using the COnstraint-Based Reconstruction and Analysis (COBRA) toolbox^[68] to determine the metabolic flux distribution corresponding to maximum biomass production (growth and storage). We subsequently carried out flux variability analysis (FVA) to determine the range of flux values for individual reactions that maintains the same optimal biomass production for GRU101 and GRU102, with the aim of identifying reactions and metabolites central to the metabolic alterations in GRU102.

Comparing the metabolic flux distributions between young GRU101 and GRU102, the FBA suggested a more than two-fold flux changes in several reactions related to aKG, including a ~3-fold reduced production of cytosolic aKG from glutamate and a ~10-fold increased production of cytosolic aKG from 3-phosphohydroxypyruvate. The model predicted a reversal in

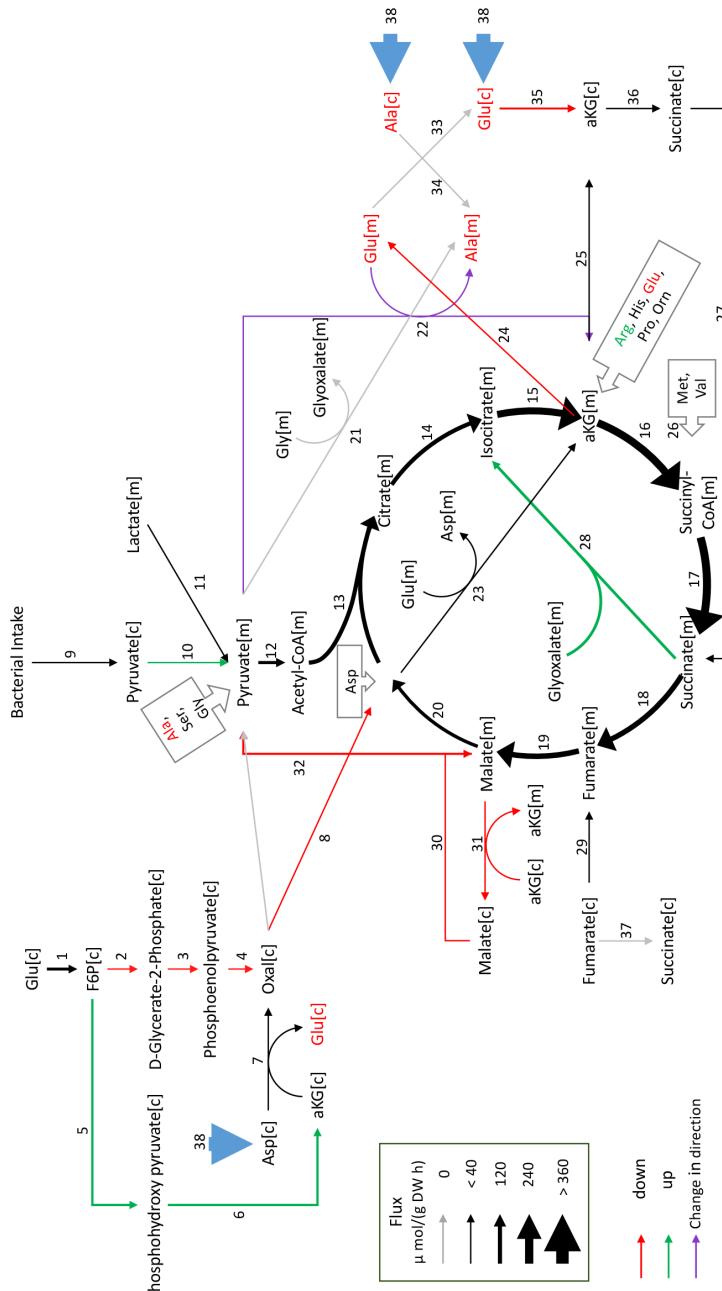


FIGURE 2.2: Metabolic flux in day 4 GRU102 modeled by metabolic flux balance analysis. Changes in flux are shown relative to GRU101 controls.

the direction of flux in the alanine-pyruvate cycle, where, in young GRU101, alanine and aKG are converted to pyruvate and glutamate, but in GRU102, pyruvate and glutamate are converted to alanine and aKG (Figure 2.2). Scanning through all results of FVA, we identified eight reactions that exhibit such directional inconsistency, where the flux intervals did not have consistent signs (direction) between GRU101 and GRU102 (Supplementary Table A.2). Seven out of the eight reactions were involved in transport, exchange, and co-factor (ADP, NADH) production, while the remaining reaction participates in the alanine-pyruvate cycle. Taken together, the pFBA and FVA results motivated us to investigate alterations specifically in the metabolism of aKG as a potential factor involved in the metabolic defects of GRU102.

A reduced pyruvate combined with a smaller fraction of AA forming pyruvate seen in GRU102 could suggest an increased Lactate Dehydrogenase (LDH) activity. The accumulation of aKG and a reduction in downstream TCA cycle metabolites (fumarate and malate) is in line with a reduced aKGDH activity or a possible alteration in enzymes downstream of aKG. Among those downstream enzymes, Succinyl-CoA Synthetase (SCS) is an important regulator of mitochondrial integrity including in neurons^[99]. Increased levels of another downstream enzyme, Malate Dehydrogenase (MDH), have previously been found in the brain of AD patients^[100]. The enzymatic activity of LDH, aKGDH, SCS and MDH were examined and evaluated for the maximum enzymatic rate in extracts of purified mitochondria derived from all cells of age-matched GRU101 and GRU102. Of the four enzymes, we only found a significant reduction in the activity of aKGDH in old GRU102 (Figure 2.3B for aKGDH, Supplementary Figure A.1 for other enzymes).

2.4.3 aKGDH knockdown in controls recapitulated metabolic detriments of GRU102

To isolate the importance and effects of repression in aKGDH activity in recapitulating the metabolic deficiencies in GRU102, a RNAi exposure against the *C. elegans* aKGDH gene, *OGDH-1*, to both GRU101 and GRU102 was performed. Their metabolic performance were analyzed using a Seahorse bioanalyzer. In order to avoid confounding errors, the RNAi treatments were initialized after the worms reached adulthood (day 3) in both GRU101 and GRU102. aKGDH knockdown significantly reduces both basal and maximal

respiratory capacity in GRU101. This response in GRU101 animals with a deficient aKGDH was similar to that of GRU102 (Figure 2.3C). The complete loss of aKGDH in GRU102 animals did not further alter its lifespan, suggesting that deficiencies in aKGDH as a result of A β toxicity is not a compensatory mechanism (Supplementary Figure A.2). In fact, knockdown of aKGDH in GRU102 aggravated the effects in terms of reduced oxygen consumption rate (OCR), confirming the impact of aKGDH deficiency in AD.

The OCR profiles of aKGDH knockdown GRU101 and GRU102 appeared to be similar following the addition of the uncoupler. We plotted the OCR of GRU102 and GRU101 with and without aKGDH knockdown to assess the similarity. The correlation coefficient, r , can be used to evaluate how well the respective conditions correlate with each other, with r value close to one being indicating a high degree of correlation. We found that the OCR during uncoupling of GRU102 correlates well with that of GRU101 fed with aKGDH RNAi ($r = 0.98, p < 0.001$) but not for animals fed EV ($r = 0.33, p = \text{ns}$, Figure 2.3D), illustrating that aKGDH knockdown causes GRU101 to experience deficits in the response to uncoupling that resembles GRU102. Additionally, GRU101 with aKGDH knockdown resembles GRU102 in the time taken to reach maximal uncoupled respiration (State 3 u) following the addition of uncoupler (FCCP). We computed the time needed for OCR to reach two standard deviations above the basal respiration rate. GRU101 showed a rapid response to the addition of FCCP. Whereas, both GRU102 and GRU101 with aKGDH knockdown suffered a delay (Figure 2.3E). Collectively, these results suggested that aKGDH knockdown is sufficient to recapitulate key metabolic deficits seen in GRU102.

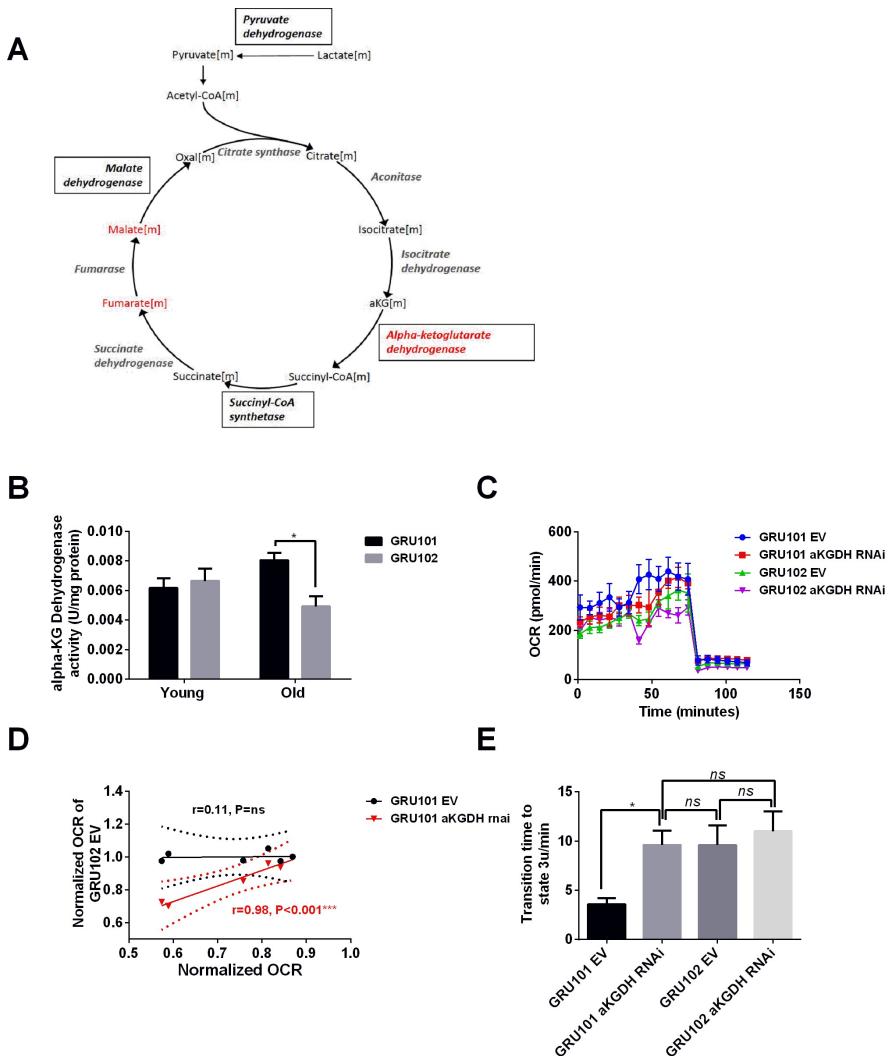


FIGURE 2.3: aKGDH knockdown recapitulated metabolic detriments of GRU102. (A) Diagram illustrating changes in key enzymes and metabolites involved in the TCA cycle. (B) aKGDH activity in GRU102 and GRU101 controls (Two-way ANOVA Dunnett's multiple comparisons test, $p < 0.05$, *; $n = 4-8$ repeats per condition, with approximately 1500 animals collected from independent cohorts). (C) Oxygen consumption rate (OCR) profiles comparing GRU101 and GRU102 fed with aKGDH RNAi or EV ($n = 6$ repeats per group, with 10 animals per repeat). (D) Scatter plot illustrating high degree of correlation in OCR timecourse of GRU102 and GRU101 fed with aKGDH RNAi. OCR values were normalized to average uncoupled values of GRU101 ($n = 6$ measurements for each condition). (E) Time taken to reach fully uncoupled respiration (state 3 u) after addition of chemical uncoupler, defined as OCR rising to a value of two standard deviation (SD) above average coupled/basal OCR ($n = 6$ repeats per group, with 10 animals per repeat, One-way ANOVA Sidak's multiple comparison, $p < 0.05$, *).

2.5 Discussion

Transgenic strain, GRU102, expressing low-levels of human A β_{1-42} has been shown to display energy deficits and deficiencies in the mitochondria well before detectable A β aggregates were observed^[84]. Here, we have shown that key energy substrates, particularly glucogenic AA that forms pyruvate and aKG, and the TCA cycle intermediates fumarate and malate, are selectively affected in GRU102. Metabolic flux balance analysis, incorporating transcriptomics and metabolic data (oxygen intake and bacterial consumption), suggested alterations in transport or metabolism of aKG to be involved in these defects. Indeed, enzymatic assays of key TCA cycle enzymes confirmed deficiency in a rate-limiting TCA cycle enzyme, aKGDH, involved in aKG metabolism.

A negatively altered aKGDH activity can explain repressed flux through the TCA cycle, energy generation, and respiratory capacity, as observed in GRU102. Inactivation of aKGDH activity can explain not only the limited TCA cycle metabolites downstream of aKG (malate and fumarate) but also the general reduction in spare respiratory capacity and metabolism. In validating the hypotheses, a aKGDH RNAi knockdown in control GRU101 re-

vealed that GRU101 showed metabolic detriments that resembled GRU102 when fed with aKGDH RNAi. A aKGDH knockdown in GRU102 animals, which already showed signs of mitochondrial dysfunction, displayed added regression of their metabolic defects. These results suggest that reduction in aKGDH activity is detrimental to metabolic capacity in adults and that it alone can recapitulate the metabolic deficits of AD in *C. elegans*.

It is fascinating that the GRU102 animals displayed such metabolic changes, given that the A β transgene was only expressed in neurons. Existing literature has suggested the activation of stress response as a non-cell-autonomous consequence in protein-misfolding diseases, such as AD^[101]. Previous research efforts in *C. elegans* have implicated mitochondrial stress response in non-neuronal compartments as a non-cell-autonomous event arising from induction of mitochondrial stress in neurons. The pan-neuronal unc-119 promoter that we use to drive A β expression in GRU102 has been suggested to have low expression in the intestine. Such expression might lead to detriments in that compartment not confined in the neurons. Therefore, in conclusion, our results suggest metabolic failure as an early event in AD pathogenesis. The normalization of metabolic dysfunction, not confined to neuronal cells, could prove valuable in formulating interventions that mitigate AD pathogenesis.

2.6 Conclusion

This chapter aimed to address and investigate the incorporation of transcriptomic data in steering intracellular metabolic flux predictions. Inferring biologically relevant flux distributions is challenging. Here, we relied on experimental culture data and mRNA quantification measurements to study the mitochondrial dysfunction in *Caenorhabditis elegans* expressing pan-neuronal human amyloid-beta (A β). Alzheimer's disease is a debilitating neurodegenerative disease characterized by synaptic loss, one that is recapitulated in worms by the expression of human A β . A novel transgenic strain with low levels of A β toxicity is ideally suited for the inspection of early events in the progression of the disease. In examining metabolic changes in *C. elegans* in the context of AD, we initially used additional constraints on a generic genome-scale metabolic model of *C. elegans* derived from experimental measurements in obtaining a suitable flux prediction for the control worms. Subsequently, we integrated transcriptomic changes and experimental data in building a disease-specific metabolic network. Our analysis shed light on the alterations in Tricarboxylic Acid (TCA) cycle metabolism as an early event in A β -induced pathology. Specifically, we identified changes surrounding alpha-Ketoglutarate from the rewired pyruvate-alanine cycle and its reduced formation of glutamate. Together with metabolomics, enzyme activity assay, and RNAi screening, we singled out alpha-Ketoglutarate dehydrogenase, whose repression is damaging to the metabolic capacity of *C. elegans*. In the next chapter, constructing on this research, we explore a computational tool built explicitly for the inference of flux alterations between a pair of conditions.

Δ FBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA

This chapter is in preparation for a publication.

3.1 Abstract

Genome-scale metabolic models (GEMs) provide a powerful framework for simulating the entire set of biochemical reactions occurring in a cell. Constraint-based modeling tools like flux balance analysis (FBA) developed for the purposes of predicting metabolic flux distribution using GEMs face considerable difficulties in estimating metabolic flux alterations between experimental conditions. Particularly, the most appropriate metabolic objective for FBA is not always obvious, is likely context-specific, and not necessarily the same between conditions. Here, we propose a new method, called Δ FBA (delta-FBA), that employs constraint-based modeling, in combination with differential gene expression data, to evaluate changes in the intracellular flux distribution between two conditions. Notably, Δ FBA does not require specifying the cellular objective to produce the flux change predictions. We showcased the performance of Δ FBA through several case studies involving

the prediction of metabolic alterations caused by genetic and environmental perturbations in *Escherichia coli* and by type-2 diabetes in human muscle.

3.2 Introduction

In the post-genomic era, there have been intense efforts directed toward the reconstruction of genome-scale models of cellular networks. An important portion of these efforts focuses on metabolic networks due to the significance of cellular metabolism for understanding diseases such as cancer^[9,102–104] as well as for metabolic engineering applications in biomanufacturing^[105]. Recent advances in high-throughput sequencing technologies, gene functional annotation, and metabolic pathway databases, and developments of algorithms for mapping gene-protein-reaction (GPR) associations and identifying missing metabolic reactions systematically (gap-filling) have enabled the reconstruction of thousands of genome-scale metabolic models (GEMs), from single-cell organisms to human^[5,8]. A GEM provides the gene-protein-reaction associations that encompass the set of metabolites and metabolic reactions in an organism as prescribed by its genome. Concurrent with these developments are the creation of efficacious algorithms that use GEMs to predict intracellular metabolic fluxes – i. e. the rates of metabolic reactions – and how these fluxes vary under different environmental, genetic, and disease conditions^[106–108].

A prominent class of algorithms based on a constrained-based modeling technique, called flux balance analysis (FBA), have flourished due to its ease of formulation and flexibility, using the stoichiometric coefficients of the metabolic reactions in a GEM, an assumed cellular objective such as maximization of biomass production, and various experimental data on metabolic capabilities and constraints of the cells, to predict metabolic fluxes^[16]. Although FBA is effective in handling large networks and predicting cell behavior in many metabolic engineering studies^[109–112], considerable uncertainty still remains about the appropriate choice of cellular objective for different conditions and cell types – a choice that requires expert knowledge of the cells and their phenotype in a given condition. Such an issue is particularly prominent for complex organisms such as human. Moreover, the flux solutions produced by applying FBA have degeneracy – that is, multiple solutions exist that give the same cellular objective value^[113]. Not to mention, standard FBA often produces biologically unrealistic flux solutions^[49,114].

Driven by the increasing ease and availability of whole-genome omics profiling data, a multitude of FBA-based algorithms have been developed to incorporate additional datasets to create context-specific metabolic net-

works and to improve flux prediction accuracy^[6,21,29–33,54]. Several of these methods, such as GIMME (Gene Inactivity Moderated by Metabolism and Expression)^[21], iMAT (integrative Metabolic Analysis Tools)^[30], and MADE (Metabolic Adjustment by Differential Expression)^[33], are based on maximizing the consistency between the predicted flux distribution and the mRNA transcript abundance of metabolic genes, where the higher the transcript level of an enzyme, the larger should the flux of the corresponding reactions. Other methods, such as E-Flux^[31], use data of transcript abundance for setting the bounds on reaction fluxes. More recent methods, such as ME-model^[24] and GECKO^[11], combine FBA with an explicit modeling of enzyme / protein expression and thus, are able to directly account for protein abundance. Thermodynamics constraints have also been integrated with the FBA to eliminate thermodynamically infeasible fluxes, and at the same time, enable the integration of metabolite concentration data, as done in recent methods such as ETFL^[23].

A number of methods focus on using differential expression data between two conditions (e.g. perturbation vs. control) to predict metabolic alterations – a particular focus of our study. The method Relative Expression and Metabolic Integration (REMI)^[115] used differential expression of transcriptome and metabolome to estimate metabolic flux profiles in *Escherichia coli* under varying dilution and genetic perturbations. The method relies on maximizing the agreement between the fold-changes of metabolic fluxes and the fold-changes of enzyme expressions between two conditions. The metabolome data, if available, are used to determine the flux directionality using reaction thermodynamics. Among the alternative flux solutions, the L1-norm minimal solution is adopted to give a representative flux distribution. Another method by Zhu *et al.*^[34] employed a softer definition when assessing consistency between the metabolic fluxes and enzyme differential expressions, where only the sign of the differences needs to agree. The method provides a qualitative determination of metabolic flux changes by determining the maximum and minimum flux through each reaction in the GEM. Both of the above methods generate metabolic flux predictions for each condition in comparison, and like standard FBA, both methods require an assumption on the cell’s metabolic objective. Generally, model prediction inaccuracy is amplified when evaluating the differences between two model predictions. Another related method, MOOMIN^[116] uses a Bayesian approach to integrate differential gene expression profiles with GEMs to

predict the qualitative change in the metabolic fluxes—increased, decreased, or no change.

In this work, we developed Δ FBA (delta-FBA) for predicting the metabolic flux difference given a GEM and differential transcriptomic data between two conditions. Δ FBA relies on a constrained-based model that governs the balance of flux difference in the GEM, while maximizing the consistency and minimizing inconsistency between the flux alterations and the gene expression changes. Δ FBA is created in MATLAB to work with the COnstraint-Based Reconstruction and Analysis (COBRA) toolbox^[117]. We applied the Δ FBA to analyze the metabolic changes of *Escherichia coli* in response to environmental and genetic perturbations using data from the studies of Ishii *et al.*^[118] and Gerosa *et al.*^[119]. We compared the performance of Δ FBA to REMI for evaluating flux differences between conditions. We also demonstrated the application of Δ FBA to human GEM, specifically evaluating the metabolic alterations associated with type-2 diabetes in skeletal muscle using myocyte-specific GEM^[120].

3.3 Methods

3.3.1 Method formulation

Δ FBA generates a prediction for metabolic flux changes between a pair of conditions, such as treated vs. untreated or mutant vs. wild-type strains. In the following, we use the superscript C to denote the control (reference) condition and P to denote the perturbed condition. Δ FBA enforces that the flux changes $\Delta\mathbf{v} = (\mathbf{v}^P - \mathbf{v}^C)$ for each metabolite are balanced, as follows:

$$\mathbf{S} \cdot \Delta\mathbf{v} = \mathbf{S} \cdot (\mathbf{v}^P - \mathbf{v}^C) = \mathbf{S} \cdot \mathbf{v}^P - \mathbf{S} \cdot \mathbf{v}^C = 0 \quad (3.1)$$

where \mathbf{S} denotes the $m \times n$ stoichiometric matrix for m metabolites that are involved in n metabolic reactions in the GEM, and \mathbf{v} and $\Delta\mathbf{v}$ denote the vector of metabolic fluxes and flux changes, respectively. The flux balance above is a direct consequence of the steady-state flux balance in each condition, i. e. $\mathbf{S} \cdot \mathbf{v} = 0$.

$$\max_{\Delta \mathbf{v}, Z^U, Z^D} \Phi = \sum_{i \in R^U} w_i (Z_i^U - Z_i^D) + \sum_{j \in R^D} w_j (Z_j^D - Z_j^U) \quad (3.2)$$

subject to:

$$\mathbf{S} \cdot \Delta \mathbf{v} = 0 \quad (3.3)$$

$$\Delta \mathbf{v}_{\min} \leq \Delta \mathbf{v} \leq \Delta \mathbf{v}_{\max} \quad (3.4)$$

$$\Delta \mathbf{v} - M Z^U \geq \mu - M \mathbf{1} \quad (3.5)$$

$$\Delta \mathbf{v} - M Z^U \leq \mu \quad (3.6)$$

$$\Delta \mathbf{v} + M Z^D \leq -\eta + M \mathbf{1} \quad (3.7)$$

$$\Delta \mathbf{v} + M Z^D \geq -\eta \quad (3.8)$$

$$\Delta \mathbf{v} + M Z^0 \leq M \mathbf{1} \quad (3.9)$$

$$\Delta \mathbf{v} - M Z^0 \geq -M \mathbf{1} \quad (3.10)$$

$$Z_k^0 + Z_k^U \geq Z_k^D, \quad (3.11)$$

$$Z_k^0 + Z_k^D \geq Z_k^U, \quad (3.12)$$

The weighting coefficients w_i allows users to prioritize certain reactions (e.g. reaction(s) corresponding to gene knock-out(s)) for (in)consistency. The set of upregulated reactions R^U and the set of downregulated reactions R^D are user-defined inputs. Typically, the sets R^U and R^D include reactions with statistically significant increase and decrease in gene expression between the perturbed condition and the control, respectively. Equation 3.4 specifies the upper and lower bounds for the flux change, which can be set based on experimental data (e.g. the change in the biomass production or growth

rates), or based on the corresponding bounds from the perturbed ($\mathbf{v}_{\min}^P, \mathbf{v}_{\max}^P$) and the control condition ($\mathbf{v}_{\min}^C, \mathbf{v}_{\max}^C$) as follows:

$$\Delta\mathbf{v}_{\min,i} = (\mathbf{v}_{\min,i}^P - \mathbf{v}_{\max,i}^C) \quad (3.13)$$

$$\Delta\mathbf{v}_{\max,i} = (\mathbf{v}_{\max,i}^P - \mathbf{v}_{\min,i}^C) \quad (3.14)$$

The MILP above will generate flux change prediction $\Delta\mathbf{v}$ whose signs (positive and negative) agree with the direction of the gene expression changes. Specifically, the binary decision variables Z^U, Z^D , and Z^0 in the MILP formulation set the sign (directionality) of the flux changes $\Delta\mathbf{v}$. When $Z_i^U = 1$, $\Delta\mathbf{v}_i$ takes a positive value above a threshold μ_i , as specified by Equations (3.5) - (3.6). Vice versa, when $Z_i^D = 1$, $\Delta\mathbf{v}_i$ takes a negative value below a threshold η_i , as specified by Equations (3.7) - (3.8). The thresholds μ_i and η_i for the positive and negative flux changes, respectively, are user-defined parameters. For example, the thresholds can be set to the same constant value ϵ , or to a value that scales with the fold-change reaction expression $e_i^{P/C}$ (Supplementary information B.1), as follows:

$$\mu_i = \epsilon e_i^{P/C} \quad (3.15)$$

$$\eta_i = \epsilon / e_i^{P/C} \quad (3.16)$$

Clearly, Z_i^U and Z_i^D cannot simultaneously be equal to 1 since no feasible $\Delta\mathbf{v}_i$ exists that simultaneously satisfy Equations (3.5)-(3.8). The decision variable Z^0 is used to force select reactions to have zero flux change value, as specified by Equations (3.9) - (3.10). Together with Equations (3.11) - (3.12) the forward and reverse halves of a reversible reaction, \mathbf{v}_k and $\Delta\mathbf{v}_k$, are prevented to simultaneously have non-zero values, so as to reduce degeneracy of the flux change solution $\Delta\mathbf{v}$.

Given the degrees of freedom in GEMs for $\Delta\mathbf{v}$, many equivalent optimal solutions often exist that give the same objective function value Φ^* as specified in Equation (3.2). By assuming parsimony in $\Delta\mathbf{v}$ – that is, $\Delta\mathbf{v}$ is minimal between the perturbed and control condition, a two-step optimization procedure is implemented in ΔFBA. The first step is to maximize consistency with gene expression changes as prescribed in Equations (3.2) - (3.12) with

the maximum objective function value denoted by Φ^* . The second step is to produce a L2-norm minimal solution for $\Delta\mathbf{v}$, as follows:

$$\max_{\Delta\mathbf{v}, \mathbf{z}^U, \mathbf{z}^D} \|\Delta\mathbf{v}\|_2^2 \quad (3.17)$$

subject to the same constraints in Equations (3.3) - (3.12) while achieving the same level of consistency Φ^* implemented by the additional constraint:

$$\sum_{i \in R^U} w_i (Z_i^U - Z_i^D) + \sum_{j \in R^D} w_j (Z_j^D - Z_j^U) = \Phi^* \quad (3.18)$$

The L2 minimization is based on the premise that the perturbed metabolic fluxes should not deviate far from the control condition, similar to the method called Minimization of Metabolic Adjustment (MOMA)^[121]. An alternative to L2-norm minimization is L1-norm minimization, which is analogous to maximizing sparsity of $\Delta\mathbf{v}$. The L1-norm minimization was previously used in the parsimonious FBA (pFBA) method, but such an approach often still leads to multiple degenerate solutions. On the other hand, the L2-norm minimization will produce a unique solution, but the mixed-integer quadratic optimization that is required to find the solution may have high computational requirements.

Δ FBA is available as MATLAB scripts and is compatible with the COBRA toolbox. Δ FBA requires Gurobi optimizer (<http://www.gurobi.com>) as a pre-requisite. Δ FBA has been tested on a Windows PC using a 6-core Intel Xeon (2146G) Processor with 16 GB RAM.

3.3.2 Case studies: Data and implementation

The first case study involved the response of *E. coli*'s metabolism to genetic (single-gene deletions) and environmental perturbations (dilution rates) performed by Ishii *et al.*^[118]. The study provided ¹³C-based flux data and RT-PCR mRNA abundances for the central carbon metabolism, pentose phosphate pathway (PPP), and the tricarboxylic acid (TCA) cycle for wild-type K12 *E. coli* culture in chemostat under different dilution rates (0.1, 0.2, 0.4, 0.5, and 0.7 h⁻¹) and for 24 single-gene perturbations along the glycolysis and PPP^[118]. The global transcriptional response was only captured for 5

of the 24 single-gene deletions (*pgm*, *pgi*, *gapC*, *zwf* and *rpe*) and two of the 4 dilution conditions (0.5 and 0.7 h⁻¹). The differential (fold-change) gene expression levels were computed with respect to the control condition, set to be wild-type K12 *E. coli* cultured at a dilution rate of 0.2 h⁻¹. The differential (fold-change) reaction expressions were subsequently evaluated based on the fold-change gene expression using the GPR Max/Min rule available in the COBRA toolbox (MATLAB)^[122]. For samples with only RT-PCR mRNA abundance data, the set of up- and downregulated reactions, R^U and R^D, included all reactions with fold-change reaction expressions higher than 1 and those with fold-change lower than 1, respectively. In the additional analyses for samples with whole-genome transcriptome data, the set of up- and downregulated reactions, R^U and R^D, respectively, were taken from the top and bottom 5th percentile of the differential reaction expressions. The differences of the measured cell specific glucose uptake rates between perturbed and control experiments were used as constraints. ΔFBA was applied using the two-step optimization with the L2-norm minimization, as described above.

The second case study came from a study of *E. coli* growth on 8 different carbon sources performed by Gerosa *et al.*^[119]. Unprocessed global transcriptomic data were obtained from ArrayExpress (E-MTAB-3392)^[123], and differential expression analyses between every pair of carbon sources were evaluated using the *Limma* package in R^[124]. As before, the fold-change reaction expressions were computed based on fold-change in the global gene expression using the Max-Min GPR rule using COBRA toolbox^[122]. The up- and downregulated set of reactions were taken from the top and bottom 5th percentile of the differential reaction expressions. In addition, cell culture data on specific growth rates were used to compute the flux change bounds for biomass production rate. The uptake rates of the carbon source changes were also incorporated as constraints. We implemented the two-step optimization of ΔFBA using L2-norm minimization.

The third case study came from two studies of skeletal muscle tissue metabolism in type-2 diabetes (T2D) patients by van Tienen *et al.*^[125] and Jin *et al.*^[126]. The microarray gene expression datasets were obtained from GEO (GSE19420^[125] and GSE25462^[126,127]) and the differential (fold-change) expression of genes for each dataset were computed using the *Limma* package in R^[124]. The fold-change reaction expressions were computed based on the differential gene expression using the Max/Min GPR rule. In the ab-

sence of additional constraints in the form of exchange fluxes or growth characteristics, we set the up- and downregulated reactions from the top and bottom 25th percentile in differential reaction expressions, rather than the 5th percentile threshold used in *E. coli* case studies above, so as to incorporate more differentially expressed transcripts. We implemented an L1-norm minimization in the second step of ΔFBA to reduce computational complexity (time) due to the large number of constraints associated with the differential reaction expressions.

3.3.3 Implementation of REMI

The method Relative Expression and Metabolomic Integrations (REMI) was developed for predicting individual flux distributions of a pair of conditions (\mathbf{v}^P and \mathbf{v}^C) using multi-omics dataset. The toolbox was downloaded from <https://github.com/EP-LCSB/remi>. The differential gene expressions in each case study were obtained as described above. The mapping from differential gene expression to the corresponding reaction expressions were done using the procedure detailed in REMI^[115]. Briefly, the authors followed the implementation of Fang *et al.*^[117] to translate gene expression ratios to obtain reaction expression ratios. When several enzyme subunits are required for a reaction, a geometric mean of expression ratios is chosen to represent the reaction ratio. In the case where multiple isozymes catalyze a reaction, the arithmetic mean of the individual expression ratios of the isozymes is used for the reaction ratio. The set of up- and downregulated reactions, R^U and R^D , were taken from the computed differential reaction expressions as in ΔFBA implementation. Unlike ΔFBA , REMI produces solutions for the metabolic fluxes of perturbed \mathbf{v}^P and control condition \mathbf{v}^C . For comparison, we evaluated the flux change predicted by REMI by taking the difference: $\Delta\mathbf{v} = (\mathbf{v}^P - \mathbf{v}^C)$.

3.3.4 Performance evaluation

The quantitative agreement between the measured flux difference $\Delta\mathbf{v}^M$ and the predicted flux changes $\Delta\mathbf{v}^*$ was assessed by using two accuracy metrics: uncentered Pearson correlation coefficient and normalized root mean square error (NRMSE). The uncentered Pearson correlation coefficient ρ was computed as follows:

$$\rho = \frac{\Delta \mathbf{v}^M \cdot \Delta \mathbf{v}^*}{\|\Delta \mathbf{v}^M\|_2 \|\Delta \mathbf{v}^*\|_2} \quad (3.19)$$

Meanwhile, the NRMSE was evaluated according to the following equation – using *tdStats* package in R:

$$\text{NRMSE} = \frac{1}{\Delta \mathbf{v}_{\max}^M - \Delta \mathbf{v}_{\min}^M} \sqrt{\frac{\|\Delta \mathbf{v}^M - \Delta \mathbf{v}^*\|_2^2}{m}} \quad (3.20)$$

Besides the quantitative agreement in flux changes, we also evaluated the qualitative agreement by comparing the signs of the flux changes between experimental measurements and predictions. To this end, we discretized the measured and predicted flux changes into +1, 0, and -1, to describe upregulated, no change, and downregulated reactions, respectively. The agreement in the direction of the flux changes was evaluated as the number of correct sign predictions divided by the total number of fluxes.

3.3.5 Metabolic subsystem enrichment analysis

The flux differences obtained from applying ΔFBA were first filtered according to the directionality of their change. The significantly altered fluxes ($|\Delta v_i| > \epsilon$) were grouped based on the subsystem to which the fluxes belong. A Fisher exact test (*fisher.test* function in the R base package) was used in determining over-represented subsystems in upregulated (positive change) and downregulated (negative change) fluxes. The statistical significance *p*-values were corrected for multiple hypothesis testing using the *p.adjust* function in R.

3.4 Results

3.4.1 *Escherichia coli* response to genetic and environmental variations

Ishii *et al.*^[118] studied the robustness of *E. coli* K12 metabolism in a chemostat in response to changes in dilution rates and gene deletions. The study generated multi-omics data, including transcriptomic, proteomic, metabolomic, and ¹³C metabolic fluxes, and demonstrated the remarkable ability of *E. coli* to reroute its metabolic fluxes to maintain metabolic homeostasis in response to environmental and genetic perturbations. Only a small fraction of variation in the measured flux ratios can be explained by the fold-change in reaction expressions, as indicated by the low coefficient of determinations R², regardless of the GPR mapping procedures in Δ FBA and REMI (Figure 3.1A). The weak agreement between reaction expressions and metabolic fluxes motivates using a system-oriented approach that considers the global network changes based on GEMs^[18]. We applied Δ FBA using *E. coli*'s iJO1366 GEM to predict the metabolic flux shifts from the control condition (wild-type K12 at 0.2 h⁻¹ dilution rate), caused by alterations in dilution rates (0.1, 0.4, 0.5, and 0.7 h⁻¹) and by 24 single-gene deletions (*galM*, *glk*, *pgm*, *pgi*, *pfkA*, *pfkB*, *fbp*, *fbaB*, *gapC*, *gpmA*, *gpmB*, *pykA*, *pykF*, *ppsA*, *zwf*, *pgl*, *gnd*, *rpe*, *rpiA*, *rpiB*, *tktA*, *tktB*, *talA*, and *talB*). We compared the predicted flux changes using Δ FBA with the differences of 46 measured metabolic fluxes along the central carbon metabolism by incorporating the enzyme expression obtained from RT-PCR. Uncentered Pearson correlations, NRMSE, and sign accuracy of the flux change predictions are depicted in Figure 3.1B and C, indicating a general agreement between the predicted flux changes and the differences in the measured fluxes. We compared the accuracy of flux change predictions by Δ FBA and REMI^[115]. As illustrated in Figure 3.1B and C, Δ FBA generally outperforms REMI in predicting the flux changes by having lower NRMSE, higher Pearson correlations, and higher sign accuracy. The results using the whole-genome gene expression profiles for a subset of perturbation experiments are comparable with those using RT-PCR data above (Supplementary Figure B.2 and Supplementary Figure B.3).

Another study, carried out by Gerosa *et al.*^[119], looked at how *E. coli*'s central carbon metabolism adapts to 8 different carbon sources: acetate, fructose, galactose, glucose, glycerol, gluconate, pyruvate, and succinate.

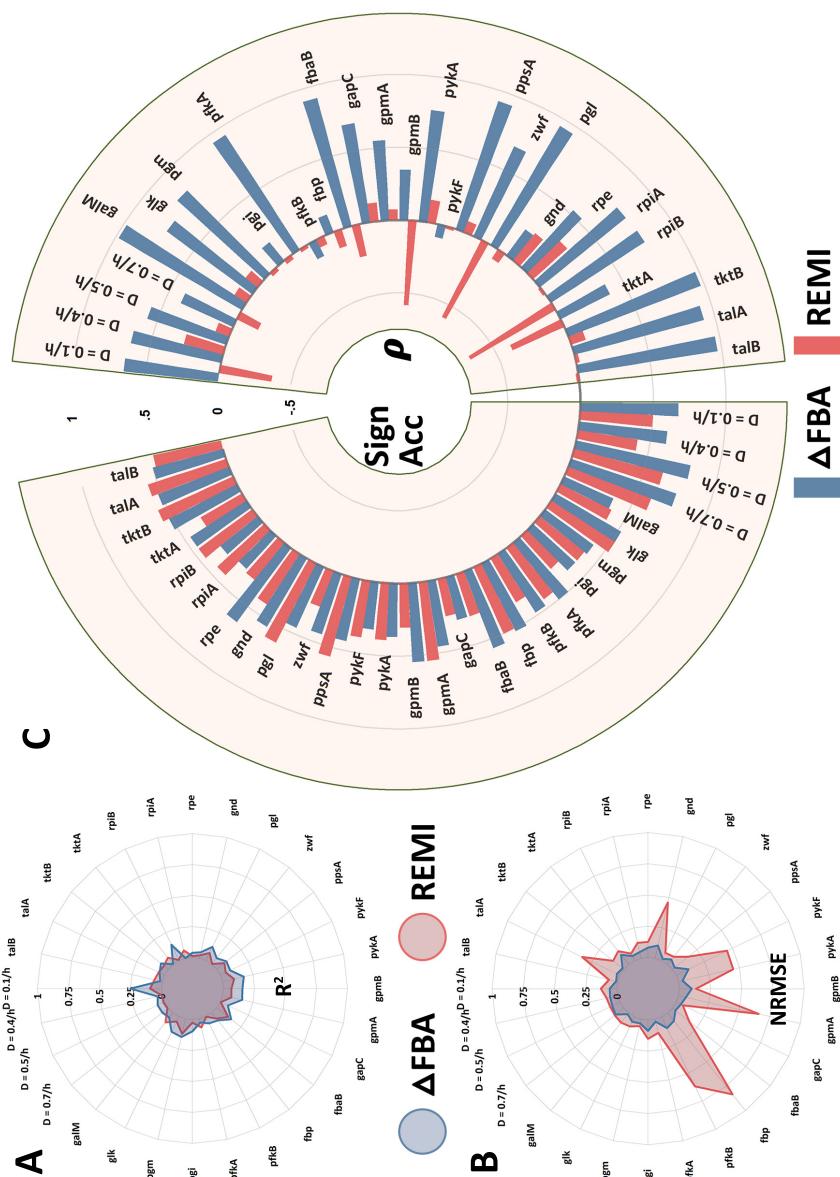


FIGURE 3.1: Comparison of the performance of Δ FBA and REMI in predicting *E. coli* metabolic response to environmental (dilution rates) and genetic (single-gene deletions) perturbations. (A) The coefficient of determination (R^2) between the measured flux ratios and the reaction expression ratios shows that differential gene expressions do not directly inform the metabolic flux changes. (B) Normalized Root Mean Square Error (NRMSE) of the predicted flux change – smaller indicates higher accuracy (mean NRMSE: Δ FBA = 0.14, REMI = 0.54). (C) Directional (Sign Accuracy) agreement and uncentered Pearson’s Correlation Coefficient (ρ) between the predicted and measured flux difference for the 46 reactions (mean sign accuracy: Δ FBA = 0.49, REMI = 0.43; mean ρ : Δ FBA = 0.61; REMI = -0.06).

The study generated ^{13}C metabolic flux, metabolite concentration, and microarray gene expression data from exponentially growing *E. coli* under each carbon source. The study found that only a small subset of the numerous transcriptome changes translates to notable shifts in the corresponding metabolic fluxes, indicating non-trivial relationships between transcriptional regulations and metabolic fluxes. We applied Δ FBA to predict flux changes between every pair of the carbon sources, treating one as the perturbation and another as the control condition. Figure 3.2 describes the good agreement between the flux change predictions by Δ FBA with the measured differences of 34 metabolic fluxes between any two carbon sources, specifically in terms of uncentered Pearson correlation (mean: 0.61), NRMSE (mean: 0.15), and sign accuracy (mean: 0.66). The findings from Ishii *et al.* and Gerosa *et al.* highlight the ability of Δ FBA in accurately predicting metabolic flux alterations using transcriptomic data for both environmental (e.g. dilution rates, carbon sources) and genetic perturbations.

3.4.2 Dysregulation of skeletal muscle metabolism in type – 2 diabetes

In this case study, we looked at metabolic alterations of human muscle using the myocyte GEM *iMyocyte2419*^[120] and gene expression datasets from two type-2 diabetes (T2D) studies, one by van Tienen *et al.*^[125] and another by Jin *et al.*^[126]. In long-term T2D patients compared to age-matched cohort, van Tienen *et al.*^[125] reported the downregulation of gene expression related to substrate transport into mitochondria, conversion of pyruvate into

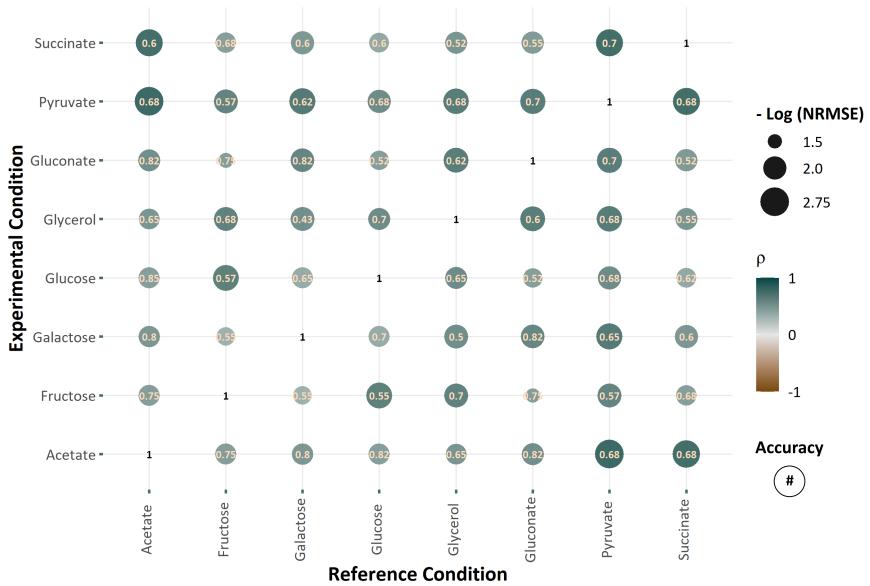


FIGURE 3.2: Prediction of metabolic flux changes in *E. coli* caused by changes in the carbon source using Δ FBA. The horizontal axis reports the reference carbon source (control), and the vertical axis shows the altered (perturbed) carbon source. The color of the markers shows the uncentered Pearson's Correlation Coefficient (ρ). The size of the markers represents NRMSE – the larger the markers, the smaller is the NRMSE. Finally, the numbers inside the markers show the directional (sign) accuracy of the flux perturbation predictions.

acetyl-CoA, aspartate-malate shuttle in mitochondria, glycolysis, TCA cycle, and electron transport chain. Similarly, Jin *et al.*^[126] reported a significant enrichment of pathways involved the oxidative phosphorylation among the downregulated genes in their T2D cohort compared to control. Jin *et al.*^[126] further identified the transcription factor SRF and its cofactor MKL1 among the top-ranking enriched gene sets with increased expression. But, the correlation between the differential gene expressions in the two studies is only modest^[120].

We applied Δ FBA to predict the flux changes based on the differential gene expressions in each of the two studies above. We grouped the reactions based on whether the predicted flux changes are positive or negative, denoted by

up- and down-reactions, respectively. We performed metabolic subsystem enrichment analysis using the subsystems defined in myocyte-specific GEM *iMyocyte2419*^[120] to identify over-represented metabolic subsystems among the up- and down-reactions. As summarized in Figure 3.3, the enrichment analysis of metabolic changes in the van Tienen *et al.* study shows a significant over-representation of β -oxidation and BCAA (branched-chain amino acids) metabolism among the down-reactions, and of extracellular transport and lipid metabolism among the up-reactions. The enrichment analysis of flux changes in the Jin *et al.* study also indicates an over-representation of lipid metabolism among the up-reactions in T2D patients, as well as an over-representation of β -oxidation pathway among the down-reactions (Figure 3.3).

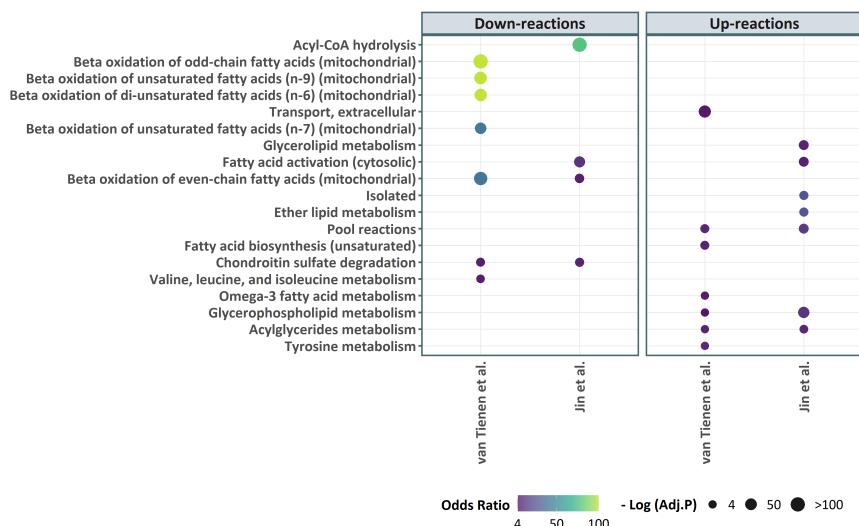


FIGURE 3.3: Enriched metabolic subsystems (FDR<0.05) among T2D patients based on flux changes predicted using Δ FBA. The flux changes were computed based on the transcriptome datasets from two T2D studies: van Tienen *et al.*^[125] (GSE19420) and Jin *et al.*^[126] (GSE25462). The size of the markers shows the statistical significance of the over-representation – larger markers have smaller adjusted *p*-values – while the color of the markers shows the odds ratio.

Furthermore, we evaluated the change in the flux throughput for every metabolite irrespective of its compartmental location, more specifically by computing the change in the total production flux of each metabolite. Metabolites with a large change in the flux throughput are of particular interest for disease biomarkers. In the following, we focus on metabolites that have a flux throughput change above a threshold ($|\Delta v_i| > \epsilon$; $\epsilon = 1$), and capture intermediary metabolites that participate in linear reaction sequences. Figure 3.4 shows the flux throughput changes predicted by Δ FBA for various metabolites. Among the metabolites with a large drop in the flux throughput in both studies are Coenzyme A (CoA), Acetyl-CoA, and AMP (Adenosine monophosphate), all of which have been previously identified as metabolite reporters of diabetes^[120,128]. Other metabolic biomarkers that have been previously proposed for T2D, such as repression of FAD (Flavin adenine dinucleotide), FADH₂ and NADH by van Tienen *et al.* study^[125] and increased glycerol by Jin *et al.* study^[126], are confirmed by Δ FBA (Figure 3.4). Väremo *et al.*^[120] had identified these markers of T2D using gene-reaction associations and consensus gene-set analysis in the GEM, *iMyocyte2419*. Besides these confirmatory observations, Δ FBA results of the two studies further suggest that arachidonate and palmitate are candidate metabolic biomarkers for T2D, both of which have a positive flux throughput change in the two T2D studies. These metabolites, although undetected by mere gene-set analysis using gene-reaction associations in the GEM, have important roles in the progression and cause of T2D^[129–131]. The results above showcase the ability of Δ FBA in elucidating metabolic flux alterations in a complex human GEM and identifying key metabolites of interest in human diseases.

3.5 Discussion

GEMs and constraint-based modeling using FBA and the myriad FBA variants have proven to be important enabling tools for establishing genotype-phenotype relationship^[20,108,132]. The increasing availability of omics data has driven the development of FBA-based strategies that are able to use such data to improve the accuracy of predictions of intracellular metabolic fluxes. In this work, we present a new FBA-based method, called Δ FBA, built for the purpose of analyzing the metabolic alterations between two conditions given data on differential gene expression. Δ FBA does not require the specification of the metabolic objective, and thus, eliminates any potential pitfalls

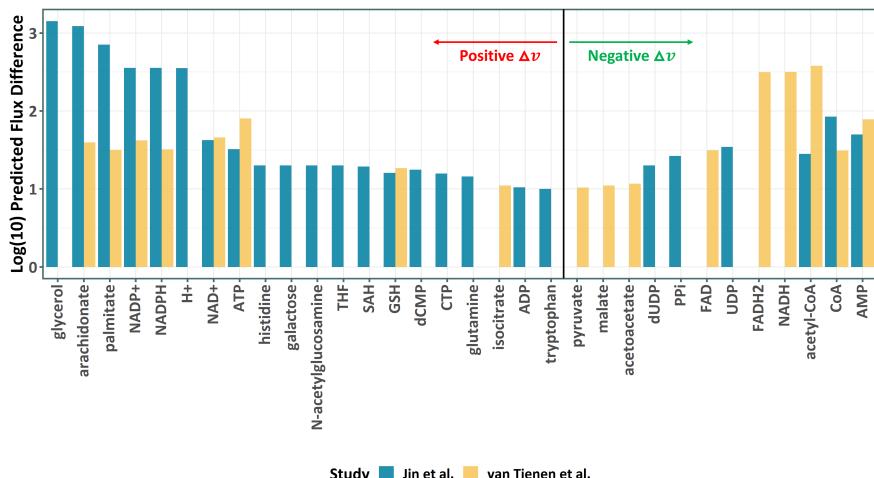


FIGURE 3.4: Alterations in metabolite flux throughput in T2D patients as predicted by Δ FBA.

that are associated with an incorrect selection of this objective. Note that Δ FBA does not generate the flux prediction for a given condition; rather, the method produces differences of metabolic fluxes between two conditions. Differential flux predictions are indispensable in formulating hypotheses and understanding the physiological response of cells to changes in the environment. Δ FBA can be easily integrated and has been tested to work with the widely popular COBRA toolbox.

We showed the applicability and performance of Δ FBA for predicting metabolic flux changes in an array of experimental perturbations and in both simple prokaryotic *E. coli* and complex multicellular human muscle cells. In comparison to a related method REMI^[115], Δ FBA shows markedly better accuracy in predicting the magnitude and direction of metabolic flux changes in *E. coli*. Further, the application of Δ FBA to two T2D studies shed light on the rewiring of muscle metabolism associated with type-2 diabetes that leads to the repression of β -oxidation and activation of glycerolphospholipids, pointing to increased lipid metabolism in the T2D patients. Interestingly, serum metabolic profiling of T2D patients showed increased glycerolphospholipids when compared to healthy controls^[133]. Besides, clinical and experimental studies have demonstrated the association between phospho-

lipids and insulin resistance^[134]. Furthermore, by looking at the changes in the flux throughput of metabolites, the results of Δ FBA suggest two fatty acids, arachidonate, and palmitate, for candidate biomarkers of T2D.

There are several limitations of Δ FBA, the most obvious of which is that the method does not produce flux predictions for the conditions under comparison (perturbed vs. control). If the values of the metabolic fluxes are desired, one can use a FBA-based method, for example parsimonious FBA^[58], to evaluate metabolic fluxes for one of the conditions (perturbed or control) – preferably one that is more well-characterized (e.g. more experimental data, more obvious metabolic objective) – and compute the metabolic fluxes for the other condition by combining this flux prediction and the flux changes from Δ FBA. Also, in the formulation and the application of Δ FBA in this work, we considered only differential gene expression data. But the method can also accommodate other omics datasets, such as proteomics, by appropriate mapping of the data to changes in reaction expressions. Metabolomics data can also be accommodated in Δ FBA via thermodynamics constraints, as done in REMI^[115], in which certain reactions can only proceed in one direction.

3.6 Conclusion

In this work, we address the challenge of studying metabolic flux alterations in organisms as a result of genetic or environmental changes. Our versatile method Δ FBA provides a set of functions for evaluating metabolic flux differences between two conditions using genome-scale metabolic models and differential gene expression data. The computational tool eliminates the need for assuming a metabolic objective by exploiting the fact that the flux differences also satisfy the same flux balance equation used in standard FBA. As demonstrated in several case studies, Δ FBA provides accurate and biologically relevant predictions of metabolic alterations caused by environmental, genetic, and disease-related perturbations. With increasing research efforts directed toward the integration of omics data with biochemical network models, tools such as Δ FBA represents a significant advancement in this direction. The MATLAB implementation of Δ FBA is freely available on <https://github.com/CABSEL/DeltaFBA>.

4

MOVING WINDOW ANALYSIS OF THE HUMAN TRANSCRIPTOME REVEALS TEMPORAL ALTERATIONS ASSOCIATED WITH AGING

This chapter is in preparation for a publication.

4.1 Abstract

Aging is a complex, multifaceted process marked by a progressive functional and physiological decline in an organism that leads to increased vulnerability and mortality. Although numerous studies of the last decade have identified genes and biological pathways that are dysregulated in multiple model organisms, the etiology and progression of the aging process is far from being completely understood. Transcriptomic data analysis presents an invaluable opportunity in obtaining snapshots of mRNA abundance across the entire body. This enormous information enables a system-oriented approach for a holistic understanding of the aging process. This study examines the aging transcriptomic signatures from over seven hundred individuals from the Genotype-Tissue Expression (GTEx) project. Our findings show that the human aging transcriptomic signatures are associated with the conserved hallmarks of aging postulated from numerous research efforts in model organisms. In further inspecting the relative importance of the age-associated

genes, we unearth a roadmap of events associated with changes in gene expression. Our analyses suggest a pivotal and continual role played by cell fate regulators. Additionally, we identify early metabolic events that precede several organism-wide perturbations observed in older ages. In summary, we identify temporal sequence of events from the analysis of gene expression that could provide valuable and novel insights into the progression of the age-associated decline. These findings could potentially be advantageous in formulating mitigating efforts in age-associated conditions.

4.2 Introduction

Aging is a complex process, one that is marked by progressive loss of physiological integrity leading to functional impairment and culminating in the death of an individual^[62,135]. Although aging is not considered a disease, it is a major risk factor for numerous age-related diseases (ARD). The risk of cancer, cardiovascular, neurodegenerative, and metabolic disorders dramatically increases with age^[36,37]. With the greying population coupled with the increase in human life expectancy and declining fertility rates, aging poses a serious threat to the health care system globally^[38]. ARDs have become the leading global cause of non-communicable disease-related deaths. In this regard, a better understanding of aging has the potential to delay the onset of ARDs, extend healthy life expectancy, and thus reduce the socioeconomic burden.

Dynamic transcriptomic changes occur during aging and are influenced by genetic factors and external stimuli. Previous studies have investigated transcriptomic signatures of aging in individual tissues and across model organisms, interestingly with minimal overlap^[136–141]. Consequently, these aging-associated transcriptomic signatures have been linked to lifespan regulation and aging-relevant pathologies. Numerous theories have been proposed to explain the multifactorial nature of aging and its link to ARDs^[41]. The narrative of accepted mechanisms of aging comprises inflammation^[42], mitochondrial dysfunction^[43,44], genomic instability^[45,46], cell cycle dysregulation (e.g. telomere shortening^[47] and accumulation of oxidative stress^[48] among many others. However, each of these theories only provides an incomplete explanation for the characteristics of aging in humans.

Despite the enormous research efforts in the past decades, the timing of these age-associated gene expression changes is incompletely understood. While many studies have identified several aging relevant genes in various tissues, the impact of these genes on cumulative changes in an organism during the aging process is unclear. When do we see these transcriptomic changes, and which of these have protective or detrimental effects on an organism? Identifying the mechanisms of temporal gene expression changes could shed light on the transcriptional roadmap of events during the aging process. Studies of human aging have largely ignored discrete temporal changes that may modulate the aging process. The assumed linearity of age-associated gene expression change is valid given the progressive events in the aging

process. Although, studies in model organisms have reported a number of genes following non-linear trajectories in worms and flies^[142,143]. Recent studies have reported large variability in gene expression in humans at middle age^[144,145]. In this context, searching for age-linked gene expression changes in temporal time stamps becomes paramount.

In screening for differential gene expression using multiple hypotheses, the assessment of power or sample size is of particular importance. Dealing with a dataset with unequal sample size distribution and very few replicates, estimating the effect size of individual genes is crucial in assessing relative importance. The number of age-associated genes that have been inferred from transcriptomic changes in previous studies can be explained by the statistical power of the study^[63]. False discovery rate (FDR) adjustment and numerous cut-off strategies have been proposed in the past. Here, we utilized a standardized effect size metric based on the proportion of variance explained by the difference in age in our regression model to estimate ‘aging’ genes. In other words, the effect size gives a measure of the importance of age differences in our model rather than purely changes in expression. Traditional measures that compute the difference between means tend to be biased toward higher expression values.

Meteoric advancements in high-throughput omics and bioinformatics technologies have led to the generation of vast amounts of data describing the panomics of an organism^[146]. The increased ease and decreasing cost of obtaining omics data combined with the extreme complexity of the aging process warrants a systems-wide study. In this work, we leveraged the comprehensive human transcriptomics data from the Genotype-Tissue Expression (GTEx) project comprising gene expression profiles from more than 50 tissues from hundreds of individuals of ages ranging from 20 to 79 years^[147]. We employed a linear mixed-effects model to identify aging gene signatures across all tissues. Functional characteristics of these ‘aging’ genes can be explained by the hallmarks of aging^[62]. We performed a moving window analysis to tease out temporal changes in expression profiles that could explain the progression of aging in humans. A comprehensive analysis of these genes can provide valuable cues in formulating intervention strategies to mitigate aging.

Our analysis of the human transcriptome from the GTEx project shows disruptions in the evolutionarily conserved hallmarks of aging, suggesting

measurable alterations during the aging process can be analyzed from the mRNA concentrations. Our findings implicate a small number of genes to play pivotal roles in the etiology and progression of the aging process, with a large number of alterations showing moderate effect sizes. Biological pathway annotation and metabolic network enrichment analysis of these genes suggests a persistent dysregulation in p53 mediated cell fate controllers and early disruptions in cellular energetics. Transcription factor enrichment analysis further implicated transcription factors that regulate cell cycle and metabolism.

4.3 Methods

4.3.1 Identification of overall ‘aging’ genes

The expressions of 56,202 transcripts in 31 human post-mortem tissue samples and 54 distinct tissue sub-categories were obtained from the GTEx project (GTEx v7 dataset). The GTEx consortium article^[147] provides all the necessary information on sample collection, RNA extraction, RNA-Seq gene expression estimation, quality control, and normalization. Genes that were not expressed in any of the samples were excluded from the analysis. We also eliminated genes with no expression in more than 95% of the tissue samples, reducing the number of gene transcripts to 34,372 from 11,551 samples in 704 subjects. We determined the genes whose expression varied significantly with aging across different tissues using a linear mixed effect model (LMM)^[69]. In the LMM analysis, the gene expression was modeled with age as a continuous covariate, tissues, and gender as fixed effects, and individuals as a random effect using the *lme* function in the R package *nlme*.

$$\mathbf{Y}_{ijk} = \beta_i + \alpha_{ij} T_j + \gamma_{ik} \text{Sex}_k + \eta_i \text{Age}_k + b_k + e_{ijk} \quad (4.1)$$

where, \mathbf{Y}_{ijk} is the expression level of a gene i in the tissue (T) j , of an individual with age k . T_j denotes the tissue sub-type effect, Age_k denotes the effect of age on gene expression and b_k denotes the random effects between individuals in the sample population. e_{ijk} signifies the cumulative error term. Both the individual effects and the errors are assumed to be normally and independently distributed. False discovery rate (FDR) adjustment was

performed using the function *p.adjust* from the R package *stats* to adjust *p*-values for multiple testing procedure using the Benjamini & Hochberg method. The same procedure for FDR adjustment was used throughout unless specified otherwise. We found 8,355 genes (FDR<0.01) with age-dependent expression across all tissues in the dataset, referred to as the aging genes. The regression constant, r_i , obtained from the LMM, was used in determining the overall directionality of change. 4,859 of the genes were downregulated with age ($r_i < 0$) and 3,496 genes were upregulated with age ($r_i > 0$).

4.3.2 Moving window analysis of the aging transcriptome

To analyze temporal changes in gene expression patterns, we started with the same 34,372 transcripts that we used for the identification of overall genes with differential expression to age. In our moving window analysis, we binned the ages of the 704 subjects into intervals of 10 years (20's, 30's, 40's, 50's, 60's, and 70's). Instead of modeling the RNA read counts with age as a covariate, we considered age as a fixed effect by contrasting pairs of age groups. We used a significance threshold for false discovery rate adjustment of 0.05. The number of differentially expressed genes for every pairwise age group comparison is shown in Table 4.1.

4.3.3 Calculation of effect sizes

To compute the relative importance of genes to the aging process, we calculated Cohen's (f^2) effect size based on the proportion of variance explained by chronological age in our regression models.

$$f^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2} \quad (4.2)$$

where B is the variable of interest in our regression model, age. A represents the set of all other variables, namely, gender, tissue sub-type, and the random error between individuals. The coefficient of determination, R^2 , captures the proportion of variance in the data accounted for by variables in the linear model. Thus, larger values of effect size signify a greater ability of age to explain the differences in the gene expression of a particular transcript.

4.3.4 Metabolic subsystem enrichment analysis

We relied on the human genome-scale metabolic model, Recon3D^[148], to identify differentially regulated reactions and subsystems using (a) GPR associations^[122] and (b) Δ FBA (delta-FBA) intracellular flux alterations (Chapter 3).

(a) Recon3D comprises Gene-Protein-Reaction (GPR) associations, which describe the relationship between genes, the proteins they encode, and the reactions the proteins catalyze^[122] in a binary fashion. Differentially expressed genes obtained from the overall linear mixed effect modeling and/or the moving window analysis were mapped using these GPR associations in elucidating the reactions which could be affected by the changes in the expression of their associated genes. Since each reaction in Recon3D is uniquely classed into one of the predefined metabolic subsystems, we performed Fisher exact tests to identify those metabolic subsystems in human Recon3D enriched in the set of up- and downregulated reactions (FDR<0.05).

(b) The regression constants (r_i) of the differentially expressed genes were mapped to their corresponding reactions using GPR association to obtain differential reaction sets, R^U and R^D . The top and bottom 25th percentile of the differentially expressed reactions were subsequently used as an input in determining altered fluxes in the metabolic model using the Δ FBA algorithm (Chapter 3). A representative solution was obtained by minimizing the L1-norm of flux deviations. Altered fluxes were used in determining enriched subsystems in positively and negatively altered fluxes (FDR<0.05).

4.3.5 Functional enrichment analysis

Functional enrichment analysis of the differentially expressed genes was performed using R Bioconductor package *ReactomePA*^[149] (v1.26.0) to extract the biological relevance of the genes that show changes in expression with age. The R Bioconductor package *biomaRt*^[150](v2.38.0) was used for gene ID conversions from Ensembl annotated genes in the GTEx database to Entrez gene IDs used in Reactome database. Functional enrichment using Reactome resulted in biological pathways significantly enriched (FDR<0.05) for the inputted gene set in question. The Reactome result and the database in general^[151] contained hierarchical relationships of biological pathways.

Every biological pathway in the Reactome database can be thought of as a node exhibiting parent-child relationships depending on its hierarchical classification. We relied on this parent-child relational database to pool together multiple pathways under the same parent node to better understand the large-scale changes (up to 3 hierarchical levels). The significance of the parent node was determined by its most significant child. Whereas the effect size of the parent biological pathway was calculated using the average of all genes that contributed to the dysregulation in significantly altering every child node under the parent.

4.3.6 Transcription factor enrichment analysis

Transcription factor (TF) that showed significant overlap between a given list of differentially expressed and previously annotated transcription factor targets were identified using transcription factor enrichment analysis (TFEA). ChEA3 is an online resource aimed at predicting TFs associated with a given gene list using experimental and literature evidence^[152]. We leveraged the web-based application to obtain the top 10 significant transcription factors associated with the list of temporally altered genes with age.

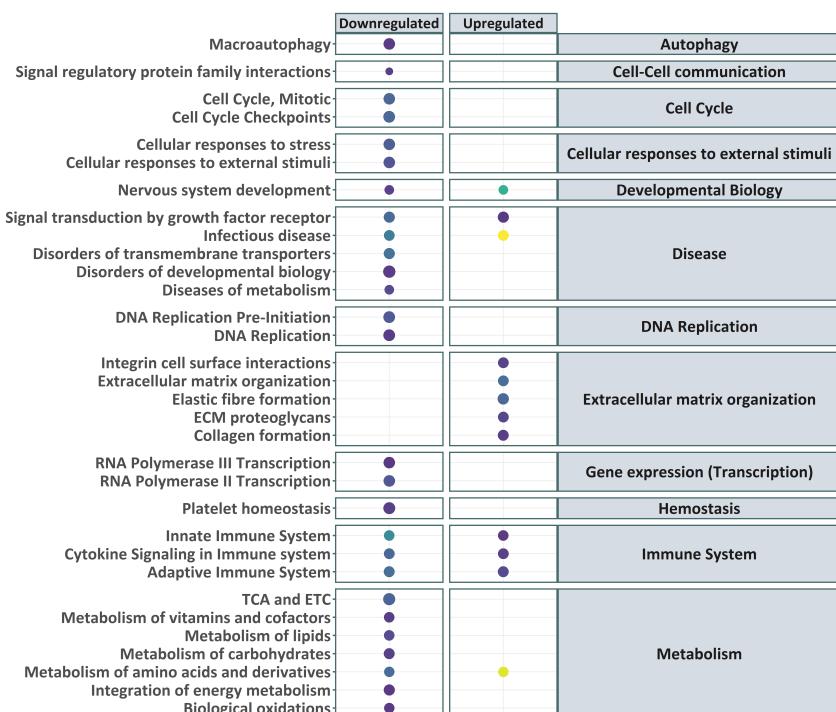
4.4 Results

4.4.1 Human aging signatures are associated with conserved hallmarks

The GTEx project provided transcriptomic data (RNA-Seq) from over 700 individuals with ages ranging from 20 to 79. We employed a linear mixed effect model correcting for variability in gender and tissue sub-type, resulting in 8,355 genes significantly differentially expressed with chronological age, 4,859 of those genes are downregulated with age, and 3,496 genes were upregulated with age. We refer to these as the ‘aging’ genes. Functional annotation of the ‘aging’ genes was derived from Reactome^[151]. The functional signatures (Figure 4.1) of the aging genes show a strong concordance with previously reported transcriptomic signatures during the aging process. Numerous studies have identified downregulation of genes involved in cellular metabolism^[138,153], cell – cycle associated processes^[154] and growth factor signaling^[155,156], induction of low-grade inflammation through the

interferon-gamma signaling pathway^[157], and upregulated stress response activation of NF-κb^[158].

We noticed a considerable overlap with other tissue-specific analyses of human aging. For example, an increased extracellular matrix (ECM) degradation capacity is synonymous with elevated levels of collagen deposition and cross-linking in cardiac aging^[159]. Age-associated repressed expression of splicing factors has been identified in multiple human tissues and in model organisms^[160-163]. Defects in splicing can result in intron retention or frameshifts, leading to nonsense-mediated decay of the improperly spliced transcripts^[63]. Age-related gene expression changes also show a repressed neuronal system. Studies of age-associated changes in the human brain have seen enrichment for repressed neuron-specific functions, such as synaptic transmission, axon guidance, and calcium signaling^[164,165].



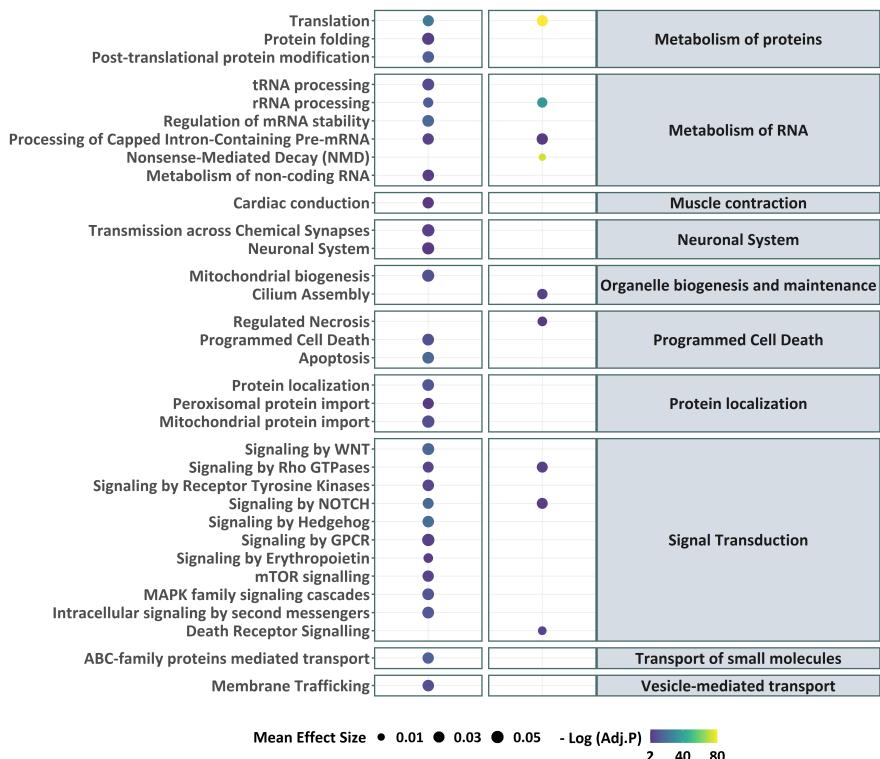


FIGURE 4.1: Functional annotation of the aging genes shows organism-wide dysregulation in metabolism, stress response, signal transduction, cell cycle, and immune system. The dot plot shows significantly altered biological pathways (FDR<0.05) for up- and downregulated genes capturing the significance and effect sizes. Downregulated genes show broader disruptions in multiple biological pathways with comparatively larger effect sizes.

López-Otin *et al.* postulated nine hallmarks of aging that represented the essential categories of cellular and molecular aging, with a particular emphasis on mammalian aging. We classified the biological pathways that were differentially regulated with age into the theorized nine hallmarks of aging and found significant changes in all nine candidates with a higher proportion of downward trends (Supplementary Figure C.1). In order to access the relative importance of the aging genes, we computed the effect size associated with every differentially expressed transcript. The functional categories of pathways that show repression with age have a larger mean effect size

from its contributing genes when compared to the set of upregulated processes signifying a more substantial explanation to the aging process from the downward trend seen in energy metabolism, loss in cell cycle control and disrupted growth factor signaling (Figure 4.1). It is crucial to keep in mind that the effect size, here, does not describe the proportion of change. Instead, it is a measure of variability in the RNA counts of each gene that the varying age of the samples can explain.

Metabolic dysfunction is a common denominator in the aging process and aging-associated pathogenesis. To this end, we performed metabolic subsystem enrichment analysis by (a) mapping the differential gene expression to a human genome-scale metabolic model, Recon3D^[148] using Gene-Protein-Reaction associations (Supplementary Figure C.2) and (b) predicting changes in the intracellular flux profiles of Recon3D using ΔFBA (Supplementary Figure C.3). We identified repression in the citric acid cycle reactions and fatty acid oxidation among the enriched subsystems. This downregulation in cellular energetics is observed not only when mapping dysregulated transcripts but also when measuring altered material flow, suggesting a significant contribution toward the repression that is not restricted to merely the changes in transcription but possibly the intracellular metabolic fluxes. A host of reactions in cellular energetics, lipid metabolism, amino acid synthesis, and degradation also showed perturbations. Collectively, these results validated the occurrence of alterations in the conserved hallmarks of aging in human gene expression changes.

4.4.2 The aging transcriptome in humans show a predominantly weak impact

Understanding the gene expression hallmarks in the aging process involves accessing the timing of these changes and their biological relevance. We utilized a moving window analysis to compute the temporal gene expression changes between pairs of age groups, binned by a decade. The number of differentially expressed genes in our moving window analysis of the GTEx dataset is shown in Table 4.1. The number of differentially regulated genes increases with larger differences in age of the compared sets, except the 70's. Is it possible that subjects in the 20's or even the 30's are more similar to the 70's than the 60's? Expression data within a population have supported heterogeneity in gene expression with age. Genetic background, lifestyle,

and environmental exposure can influence this heterogeneity. However, in this context, the median effect size of the genes changing with age is higher when compared to the 70's than the 60's. This is true for the set of variable genes and all genes irrespective of their significance to chronological age (Supplementary Table C.1). Thus, a large number of differentially expressed genes in the 60's could be due to the sample size effect. Subjects in the 60's outrank the 70's by a factor of 10 with regard to the number of samples. Although the majority of the genes are differentially expressed during the 60's, their effect sizes are relatively small, meaning a comparatively smaller proportion of variance in the regression is accounted for by variations in age.

	30-39	40-49	50-59	60-69	70-79
20-29	0	82	6256	12884	9332
30-39		0	470	7826	5988
40-49			14	8633	4487
50-59				3614	1171
60-69					1

TABLE 4.1: Differentially expressed genes in the moving window analysis. The color represents the age difference between the groups compared in sequential windows to unearth temporal changes in biological pathways.

We chose a difference of two decades (i.e., 20's vs. 40's; 30's vs. 50's; 40's vs. 60's and 50's vs. 70's) in our moving window analysis. The comparison between age groups two decades apart was large enough to have adequate statistical power, yet small enough to tease out changes to piece together a transcriptional roadmap of events during the aging process. Metabolic network analysis of the genes in these separate windows pointed toward early metabolic repression in oxidative phosphorylation followed by wider perturbations in late-life (Supplementary Figure C.4), suggesting a central role played by cellular energetics in the early progression of the aging process. However, biological pathway analysis of these selected windows required additional granularity in identifying the timing of alterations and their contributions.

In an effort to identify temporal changes and their relative importance, classes of genes that change in unison across multiple ‘windows’ were isolated. For example, a set with genes that significantly change in 20’s vs. 40’s and 30’s vs. 50’s but not in 40’s vs. 60’s nor 50’s vs. 70’s was referred to as [1,1,0,0], where 1’s stood for significant differential expression and 0’s for a non-significant outcome in a binary representation. Since the change starts in 20’s vs. 40’s, the gene set would be classified as early changers. Interestingly, a majority of the genes that were assigned as changing significantly ($FDR < 0.05$) in multiple windows had the same directionality of change in each of the windows (Figure 4.2A). We categorized the genes based on the timing of their change (as Early, Mid-life, Late or Persistent) and their cumulative average effect size profiles (as strong, moderate and weak) (Figure 4.2B). In other words, [1,1,1,1] represents a class of genes that are changing in each of the ‘windows’. Moreover, their cumulative effect size of change remains relatively large when compared to all other genes. Therefore, we label [1,1,1,1] as *Strong - Persistent*.

Our classification identifies distinct categories of genes that reflect when these genes start showing changes (Supplementary Table C.3). Most of the genes that showed alterations in expression with chronological age were predominantly weak in their effect size. This further supports the hypothesis that a large number of changes in gene expression could be attributed to heterogeneity intrinsic in populations that may not be firmly attributed to purely chronological age difference reflected by weak effect sizes. Our analysis shows that despite the significant change in a large number of transcripts, its importance in the etiology of the aging process is limited. Additionally, the genes that showed perturbation only in our LMM analysis and not in the MW analysis (LMM-MW) also had a relatively weak effect size suggesting that the genes identified to be consistently changing across all life could be of lesser significance than temporal moderators of the aging process.

4.4.3 Transcriptional roadmap of aging reveals persistent and strong changes in cell fate regulators

We focused our attention on the relatively fewer number of genes that contributed to significant alterations with large effect sizes spanning the entire age landscape from our moving window analysis. In our analysis of aging

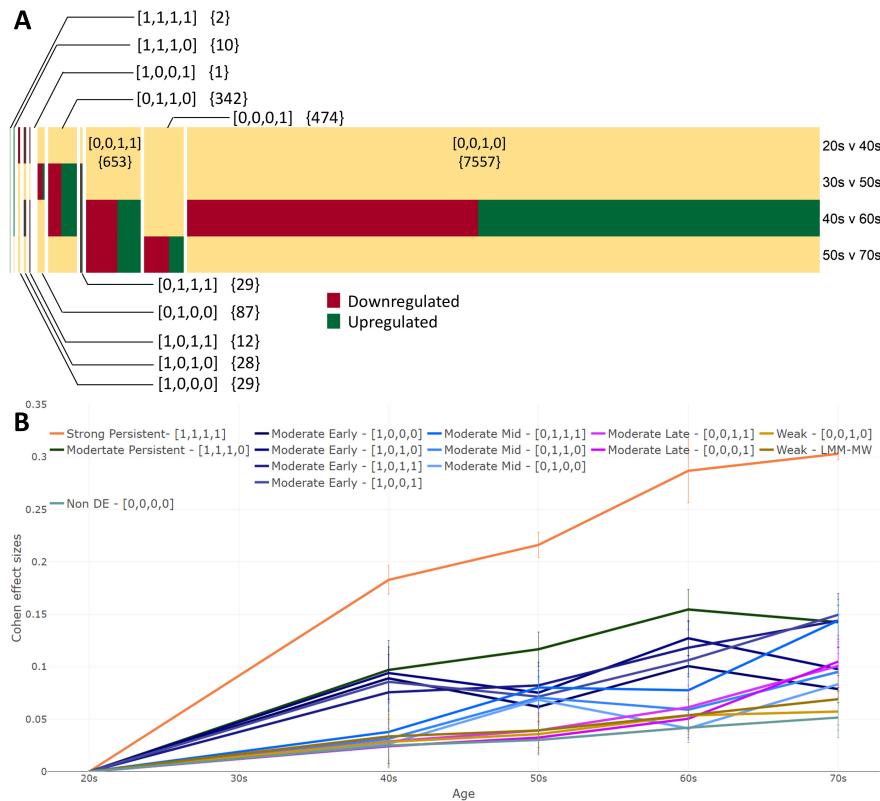


FIGURE 4.2: Classification of genes in the moving window analysis based on their timing of change (A) and mean cumulative effect size (B). The heatmap of the genes showing expression alterations in individual windows was distinctly classified, and their cumulative mean effect size showed that a large number of transcriptomic changes were weak. There also existed groups of genes that showed perturbations in a short temporal period whose effect sizes were relatively moderate.

genes, we identified transcripts that contribute to persistent and relatively large changes in the aging process. *PTCHD4* and *EDA2R* show upregulation throughout life with a significantly large effect size. *EDA2R*, a transmembrane protein of the TNFR (tumor necrosis factor receptor) superfamily, is thought to be a robust candidate biomarker for lung aging^[166]. While the exact role of the gene in lung aging remains to be explored, this transcri-

tional regulator of p53 is associated with aging in multiple tissues^[167]. The relatively large effect size of change of *EDA2R* and its persistent upregulation demands further exploration into its role in aging at an organismal level. Despite its role in regulating p53 and evidence of changes in expression with age in the thyroid^[167], *PTCHD4*'s role in the aging process is hazy. Functional enrichment analysis of these upregulated genes shows an upregulated TNFR mediated cytokine immune system pathway (Figure 4.3 and Figure 4.4).

Of the genes that show a change in the first three of the four windows ($[1,1,1,0]$), *CDKN2A* (*p16^{Ink4a}*) is one of the most significant genes in the regulation of the cell and maintaining the balance between proliferation and cell-cycle arrest (Figure 4.4). They are postulated to play a central role in the establishment of a senescent state by blocking cell-cycle progression and often tipped as the biomarker for chronological age in humans^[168,169]. Growth arrest-specific gene 6 (*GAS6*) is a ligand of receptor tyrosine kinases of the TAM (Tyro3, Axl, and Mertk) receptors that also plays a central role in cellular senescence. *GAS6* is significantly upregulated with age. Interestingly, a p53 developmental pathway gene – *ZMAT3*, a zinc finger protein involved in mRNA stability, is overexpressed.

Furthermore, reaction oxygen species have been implicated by several studies in premature senescent response. Damaged DNA binding Protein-2 (*DDB2*), a nucleotide excision repair protein, plays a crucial role in ROS regulation. The upregulation of *DDB2* is fascinating given the increase in the NADPH oxidase transcript *NOX4*. An upregulated *NOX4* was observed in cardiac tissues as a result of stress and aging in the heart, where *NOX4* could potentially be a significant source of oxidative stress^[170]. Together, these genes contribute to a dysregulation in the immune system, cell cycle, DNA repair, protein metabolism, and homeostasis (Figure 4.4).

Transcription factor enrichment analysis of the persistently upregulated genes identified *TP53*, a transcription factor regulating cell division. Together with *POU5F1* (*Oct-4*), these TFs are critical in cell fate regulation (Supplementary Figure C.5). They are essentially involved in genomic stability and have tight control over the causation of cellular senescence. Loss of differentiation potential has long been a central focus of aging studies. Interestingly, changes in the human transcriptome reveal changes in cell cycle controllers persistently and strongly throughout life.

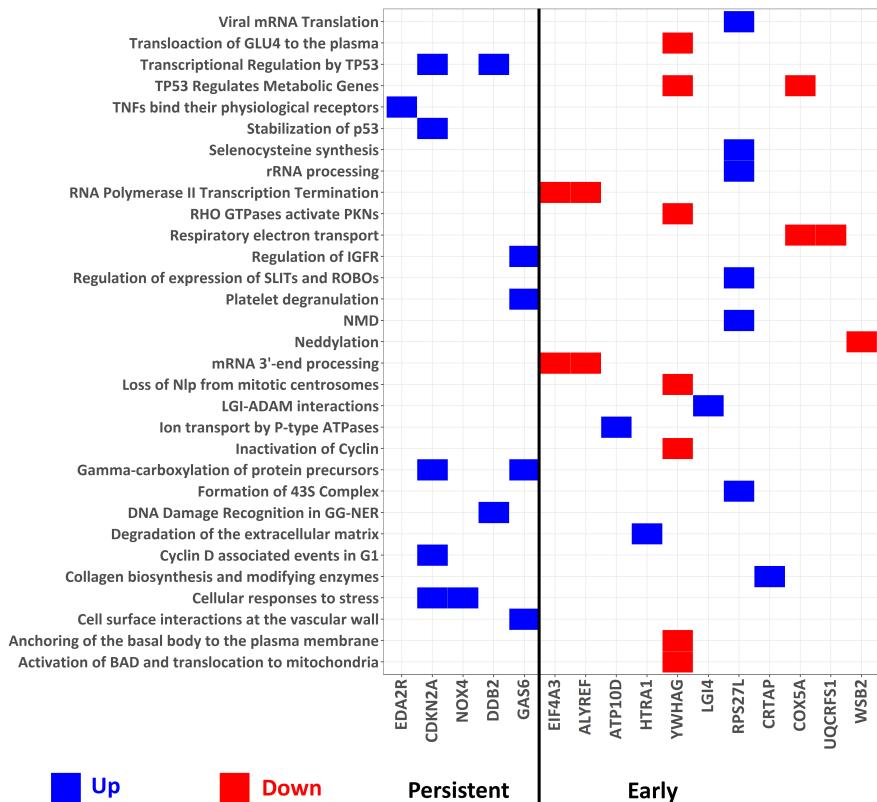


FIGURE 4.3: Genes that show persistent and early-life alterations in expression. *PTCHD4* and *EDA2R* are upregulated with relatively large effect sizes. All other genes shown here had a moderate effect size of change. Persistent alterations in gene expression perturb cell fate regulatory processes. Early changes alter gene transcriptions, RNA processing, and metabolism.

4.4.4 Regulation of metabolic genes may headline changes in early life

Among the temporal alterations that occur, capturing the earliest events is pivotal in establishing causation. Early-life mitochondrial repression has been observed in model organisms^[63]. Metabolic network analysis of the selected age-group comparison showed early life repression of cellular energetics (Supplementary Figure C.4). In our analysis of the temporally altered genes that start to show perturbations in the earliest of ages in the GTEx

set of samples, we identified metabolic events that are downregulated and continue to show downregulation in later parts of life as well ([1,0,1,1]). Metabolic network analysis of the group of genes also showed significant downregulation of oxidative phosphorylation (Supplementary Figure C.6). YWHAG is a part of a family of proteins of the 14-3-3 family highly expressed in the brain tissue and is involved in the brain development and signal transduction pathways. They have been consistently reported to be downregulated in neurodegenerative pathologies^[171,172]. They also perform essential functions in the insulin/IGF-1 signaling axis. Along with the downregulation of genes in the mitochondrial respiratory chain (*UQCRCFS1* and terminal *COX5A*), they could collectively contribute to metabolic dysfunction (Figure 4.3 and Figure 4.4).

Ribosomal protein S27-like (*RPS27L*) is a physiological regulator of p53 that activates the tumor suppressor pathway to respond to stress^[173]. A recent study of regulatory modifications in the human transcriptomic changes in the brain identified alterations in the *RPS27L* as a response to stress^[174]. Transcription factor enrichment analysis of the early changing genes implicated central roles of *FOXO3*, *KLF4*, and *E2F1*. *FOXO3* regulates stress response and has long been associated with aging and age-associated phenotypes in several model organisms^[175]. Importantly *E2F1*, a novel regulator of cellular metabolism^[176] also regulates cellular and organismal senescence by inhibiting FOXO factors^[177], suggesting a possible primary role played by metabolic regulators prior to alterations in stress response as a secondary event. Together with the findings of persistent and strong alterations in cell fate regulators, NADPH oxidase and early repression in energetics could occur prior to DNA repair and cellular response to stress which arise as a consequence of the damage.

4.4.5 Progressive and accelerated decline in mid to late-life

A host of genes and pathways begin to show alterations from mid-life (30's vs. 50's) that progressively accelerates into broader dysregulation as we progress to later stages in life (70's). Unique changes in mid-life are focused around extracellular matrix (ECM) dysfunction. ECM organization dysfunction, elevated collagen deposition, formation, and cross-linking are generally thought to be characteristics of cardiac tissue aging^[159]. Late-life modifications in biological processes are accompanied by an aggressive loss

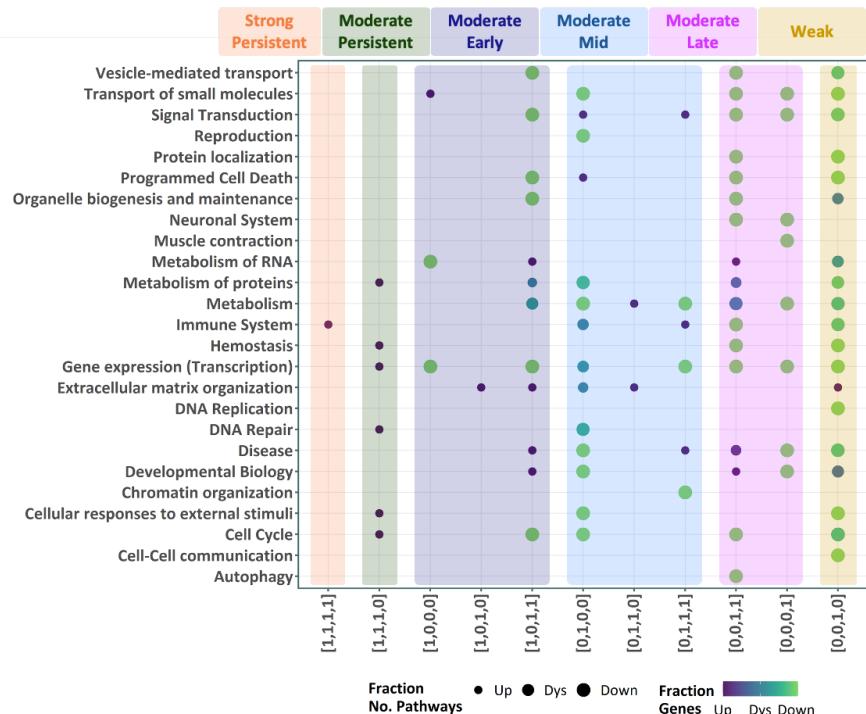


FIGURE 4.4: Temporal alterations in biological pathways in human aging. The dot plot shows significantly altered biological pathways (parent nodes) ($FDR < 0.05$) that were combined from up and downregulated genes. Individual pathways were linked to their parent nodes using the Reactome relational database. Thus a dysregulated pathway can have contributions from multiple sub-pathways and multiple genes, each of which can have different directionalities of change. The size and color represent the contribution from up- and downregulated pathways and genes, respectively.

of homeostasis in immune response, genomic integrity, cellular metabolism, and nutrient signaling. Significantly the loss of motor function and neuronal decline were only visible toward late-life suggesting neurodegenerative conditions occur as a result of accumulated damage^[178] (Figure 4.4). Metabolic network analysis of the temporal gene alterations also validate the progressive damage in late-life. Apart from the repression in oxidative phosphorylation which starts to appear in early life, dysfunction in hyaluronan metabolism, that contributes to ECM dysfunction^[179], and inositol phos-

phate metabolism, which regulates neuronal system^[180], were significantly altered in late-life (Supplementary Figure C.6). Collectively, our findings suggest an escalation of disruptions toward late-life.

4.5 Discussion

Aging is a complex, multifaceted process characterized by progressive loss of functional and mental health, leading to an increased vulnerability that culminates in death. Age-associated changes in mRNA abundance have a marked impact on an organism. Heterogeneity in gene expression makes it challenging to obtain actionable insights into the aging process. Furthermore, gene expression levels are tightly controlled by several factors making it extremely difficult to translate transcriptome alterations into modulated cellular processes. Changes in gene expression data should be treated as insights rather than direct measurement of the physiology of the cell. However, several studies have modeled changes in gene expression in humans and have hypothesized cellular functional perturbations. In this work, we analyze the human gene expression landscape to identify evolutionarily conserved hallmarks of aging. Furthermore, we evaluate the timing of these age-associated gene expression changes and their functional role in the aging process.

Our analysis of the global characteristics of the alteration in gene expression with a mixed effect modeling approach identified several changes in an organism that contributed to the aged phenotype. Categorizing these changes, the overall alterations in the gene expression showed conserved dysregulation of hallmarks of aging. In order to uncover the relative importance and timing of age-associated alterations, we analyzed the transcriptome in temporal time windows. Our results reveal that most dysregulated transcripts are associated with relatively weak effect sizes suggesting that the aging process could be initiated by a small number of central effectors that ultimately propagate to organism-wide alterations. Identifying persistent moderators of cell cycle and senescence, and early players in metabolic repression further highlights the vital role of genes with large effect sizes.

We summarize our findings on the timing of age-associated changes in Figure 4.5. Aging and metabolism are closely connected. In fact, dietary restriction in model organisms has been shown to increase lifespan. It remains one

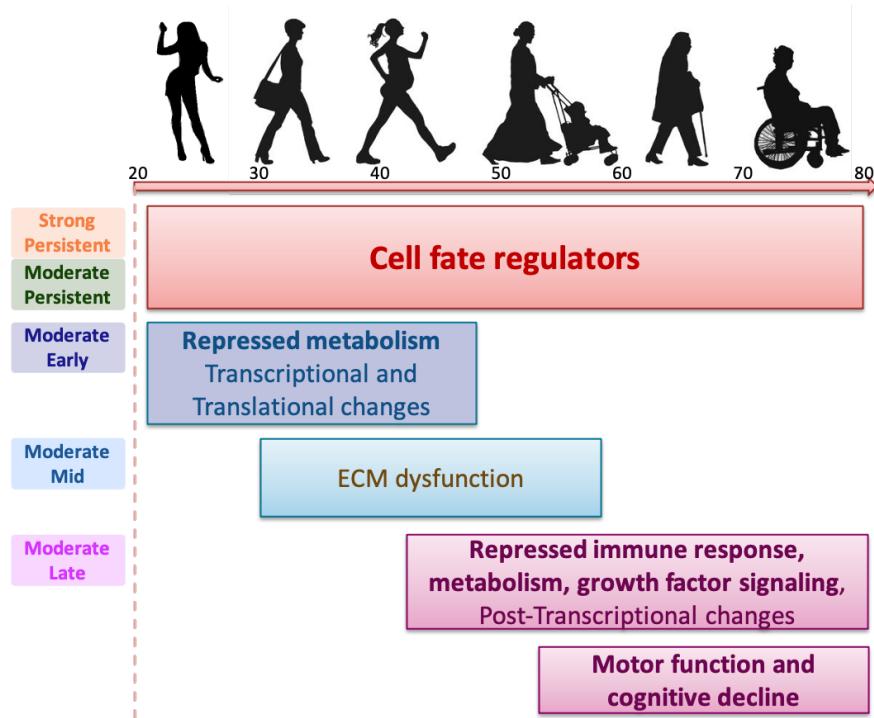


FIGURE 4.5: Transcriptional roadmap of the human aging process.

of the most successful intervention strategies in delaying age-associated modulations in an organism. Interestingly, gene expression changes in humans reveal early repression in cellular energetics that persists at older ages. Could metabolic events contribute causatively to an aged phenotype in humans? Although our analysis does not answer the question of causation, the sequence of events in the transcriptomic alterations suggests, at the very least, an early role of metabolic disruptions. Transcription factor enrichment analysis further implicates transcription factors that play a crucial role in metabolism and cell fate control. Taken together, these results warrant multi-omic analysis in further analyzing and validating the role of metabolic events and their contribution to the etiology of the aging process in humans.

Our analysis also shows aggressive disruptions in a vast array of biological pathways and metabolic subsystems that show alterations toward late-life.

Interestingly neuronal dysfunction was marked toward very late in life, suggesting a prior accumulation of damage and loss of homeostasis could contribute to the loss of function in the neuronal system. Metabolic network analysis also showed extensive alterations in enriched subsystems toward late-life. In conclusion, our bioinformatics and metabolic network analyses of temporal events in the aging process uncovered by the human transcriptomic is an important step that could aid in the study of underlying causative events in the progression of the aging process.

4.6 Conclusion

Aging is a complex process and is associated with human diseases with varying pathologies. Age-associated changes can be observed in every organism on this planet, albeit at varying rates. There also exists considerable variability within organisms of the same species. In this work, we wanted to address the unclear etiology of the aging process. The Genotype - Tissue Expression project remains an indispensable repository of transcriptomic information with subjects ranging from the 20's to the 70's, making it ideal for the study of human aging. Firstly, in our analysis of the gene expression changes and their subsequent functional annotation, we show that the human aging process is associated with evolutionarily conserved hallmarks of aging. Secondly, the temporal analysis of differentially regulated genes shows that the predominant number of transcripts that show changes has weak effect sizes. In other words, their relevance to the aging process could be feeble. Our findings show that the genes that are significantly altered ubiquitously, do so with larger effect sizes and are moderators of cell fate. Thirdly, our analysis of early alterations in the aging process implicates possible pivotal repression in cellular energetics. Enriched transcription factor, *E2F1*, a metabolic and cell fate regulator, could suggest a central role in the metabolic control of cellular events. Lastly, we identify neuronal system repression very late in life along with organism-wide alterations in a number of metabolic and biological pathways suggesting an accelerated decline.

CONCLUSIONS AND OUTLOOK

5.1 Summary of findings

In answering the fundamental questions in complex biological processes, understanding the complicated and intricate connection between the genotype and the phenotype of an organism is at the core of systems biology. The post-genomic era explosion of data in every level of the omics has clearly shifted the focus in analyzing and comprehending these enmeshed interactions to attain a meaningful insight that has the potential to delay or mitigate some of the pressing human pathologies. In this regard, metabolism is a well-established proxy for the phenotype of an organism. Additionally, cellular metabolic networks remain one of the widely studied and curated repositories of genome-wide information of a cell. The creation of computational algorithms that use information in the genome-scale reconstruction of the metabolic networks for phenotypic predictions has been progressively effective.

The depreciating cost, increased availability, and ease of obtaining transcriptomic data have ensured the integration of these invaluable resources to obtain biologically relevant predictions. mRNA concentration measurements provide a quantitative snapshot of the cell state and are widely used to analyze numerous diseases with varying pathologies. At this intersection of metabolic network analysis and the increasing use of transcriptomic resources in understanding cellular behavior, we collaborated with Dr. Jan Gruber's group in studying the metabolic dysfunction in Alzheimer's disease model nematodes. In Chapter 2, we investigate the progression of the Alzheimer's disease (AD) in transgenic *Caenorhabditis elegans*. Transgenic worms expressing human amyloid-beta (A β) specifically in the neurons (GRU102) recapitulate the pathogenesis of AD. Importantly, the novel strain had low-level toxicity enabling the study of early events in AD. With

the combination of transcriptomic data driven metabolic network analysis, metabolomics, and enzyme activity assays, we identify perturbations in the Tricarboxylic acid (TCA) cycle metabolism following A β expression. Our findings in computational predictions implicated a rewired pyruvate-alanine cycle and alteration to the flux throughput of alpha-Ketoglutarate. Together with experimental evidence of repressed downstream TCA cycle metabolites, we identify repression in the rate-limiting TCA cycle enzyme, alpha-Ketoglutarate dehydrogenase. Finally, our results point to metabolic dysfunction as an early event in A β -induced pathology and a promising target for intervention.

Building on our experience with metabolic analysis, we built a computational algorithm that eliminates some of the pitfalls of the traditional flux balance analysis. In inferring intracellular flux difference between a pair of conditions, Δ FBA (delta-FBA) eliminates the need for a cellular objective. Deducing the objective for multicellular organisms can be rather tedious and sometimes even impossible. In Δ FBA, we try and maximize the consistency of the inferred flux difference to the measured gene expression changes. By incorporating transcriptomic changes and experimental evidence, Δ FBA outperformed similar methods. We applied Δ FBA to a wide array of single-gene deletion mutants and feed dilution conditions in *E. coli* with improvements in prediction correlation and accuracy. Δ FBA was effective in predicting changes in metabolic events as a result of changes to the carbon feed sources in *E. coli*. Finally, we used Δ FBA in uncovering metabolic subsystems and key metabolites that could play a central role in type-2 diabetes. Our findings were in congruence with previous studies on diabetic patients. Δ FBA represents a versatile and robust tool that could aid in hypothesis generation and model predictions of cellular behavior.

Metabolism and aging are very closely related. The complexity of the aging process has warranted a systems-oriented approach in discovering its etiology. In Chapter 4, we detail our analysis of the human aging transcriptome. We leveraged the Genotype-Tissue Expression (GTEx) project to analyze genome-wide changes in gene expression meticulously. Our analyses of the overall gene alterations showed disruptions to the evolutionarily conserved hallmarks of aging. We show that majority of these changes are weak in their modulation. Temporal signatures reveal persistent moderators of cell cycle and senescence and early metabolic events that could play a principal role

in the progression of the aging phenotype that culminates in an accelerated decline in old life.

5.2 Future research directions

The results presented in this thesis show the untapped potential in the analysis of individual components as integrated systems when dealing with complex biological problems. Models and algorithms that could provide a reliable inference of cellular behavior are of tremendous advantage to researchers and the pharmaceutical industry. In this direction, improvements to the Δ FBA strategy could serve as a valuable tool for predicting metabolic alterations with the integration of additional levels of the omics. Firstly, we propose the incorporation of thermodynamic constraints to a generic genome-scale metabolic model. Metabolic models contain numerous reactions connecting metabolites that are thermodynamically infeasible. Pruning a generic model to include only those reactions that are feasible can drastically improve prediction accuracy. The ETFL formulation^[23] incorporates thermodynamic data to contextualize GEMs. The addition of such resources to the Δ FBA formulation would make it a versatile and robust predictor of phenotype. Additionally, we are also working on the incorporation of metabolomic measurements with Δ FBA. Secondly, we propose a graph-based metabolic network analysis to isolate central metabolites that are vital to flux alterations or to identify minimal networks^[1,115]. Discerning the principal moderators of pathology-associated metabolic changes could serve as invaluable targets. Finally, we also propose the second stage optimization for Δ FBA centered on the Regulatory on/off minimization (ROOM) algorithm^[181]. Current implementations of obtaining a singular solution to the underdetermined flux balance problem have revolved around minimizing L1 or the L2-norm. However, reduction in the cardinality of non-zero fluxes had proved to be an effective alternative for gene deletions in *E. coli*^[181]. Adopting such a strategy in Δ FBA would be possible. Z^0 in the Δ FBA formulation is a decision variable that could potentially force Δv_i to be zero. In other words, one would try and maximize consistency with the integrated omics data with as few changes as possible.

The bioinformatics and metabolic network analysis of the aging genes underline the complexity of the aging process in humans. Despite the analysis of aging genes in temporal time stamps, questions remain over the individual

contributions of various tissues. The intrinsic heterogeneity in the gene expression of a population further complicates uncovering central regulators of the aging process. Despite being a post-mortem collection of samples, GTEx is still a critical knowledge base. We process the analysis of temporal changes in the gene expression by segregating subjects based on any underlying disease condition associated with aging. Recent research efforts by Zeng *et al.* [168] revealed the difference in ‘normal’ aging patterns and ‘diseased’ aging. Together with our results, such studies and efforts could further isolate targets that can mitigate or even reverse the aging process.

A

SUPPLEMENTARY INFORMATION: METABOLIC NETWORK ANALYSIS IN TRANSGENIC *CAENORHABDITIS* *ELEGANS* STRAIN EXPRESSING HUMAN AMYLOID-BETA

A.1 Nutritional and metabolic parameters used in metabolic flux balance analysis

Parameters	GRU101	GRU102	Ratio of GRU102/GRU101
Oxygen uptake µmol/(g DWh)	734	837	1.13
ATP production µmol/(g DWh)	3966	3280	0.826
Pharyngeal pumping rate pumps/min	380	306	0.805
Body size mm	1.42	1.27	0.894
Biomass µmol/(g DWh)	68.21	52.17	0.761

TABLE A.1: Nutritional and metabolic parameters used in metabolic flux balance analysis.

A.2 List of reactions that show directional inconsistency in flux variability analysis (FVA)

Reaction Name	Subsystems
xanthine:NAD ⁺ oxidoreductase	Purine metabolism
ATP:CDP phosphotransferase	Pyrimidine metabolism
ATP:CMP phosphotransferase	Pyrimidine metabolism
L-Alanine:2-oxoglutarate aminotransferase	Alanine, aspartate and glutamate metabolism
Hydrogen peroxide transport	Transport
L-Serine transport	Transport
Coenzyme A transport	Transport
H ₂ O exchange	Exchange with the environment

TABLE A.2: List of reactions that show directional inconsistency in flux variability analysis (FVA).

A.3 TCA cycle enzyme activities

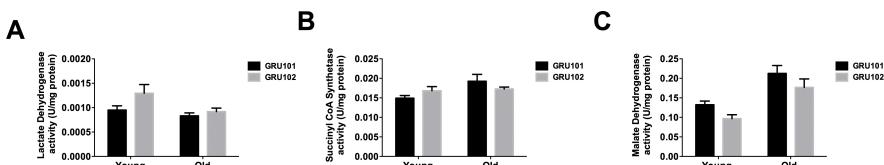


FIGURE A.1: TCA cycle enzyme activities. (A) Lactate dehydrogenase, (B) Succinyl-CoA Synthetase, and (C) Malate dehydrogenase activity in GRU102 and WT ($n = 6$ repeats per condition, with approximately 1500 animals collected from independent cohort).

A.4 Lifespan study of GRU102 in RNAi experiments

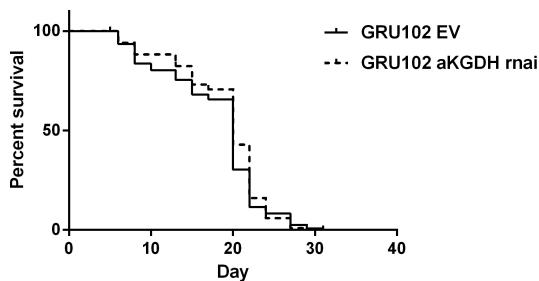


FIGURE A.2: Survival curve of GRU102 fed with empty vector (EV) or aKGDH RNAi. No significant differences were seen between the survival curves of the two conditions ($n = 120$ animals for each group).

B

SUPPLEMENTARY INFORMATION: ΔFBA — PREDICTING METABOLIC FLUX ALTERATIONS USING GENOME-SCALE METABOLIC MODELS AND DIFFERENTIAL TRANSCRIPTOMIC DATA

B.1 Alternative threshold criteria of minimum flux change magnitudes

The thresholds for the minimum magnitude of positive and negative flux changes, μ_i and η_i respectively, are user-defined parameters which can be a constant value ϵ , or a value that scales with the fold-change reaction expression $e_i^{P/C}$. If

$$\mu_i = \epsilon \tag{B.1}$$

$$\eta_i = \epsilon \tag{B.2}$$

When $Z_i^U = 1$, Δv_i should only be greater than ϵ (default = 0.1) and when $Z_i^D = 1$, Δv_i will be less than $-\epsilon$. However if,

$$\mu_i = \epsilon e_i^{P/C} \tag{B.3}$$

when $Z_i^U = 1$, Δv_i will be scaled proportional to the upregulated reaction fold change. Similarly, by setting

$$\eta_i = \epsilon / e_i^{P/C} \quad (B.4)$$

when $Z_i^D = 1$, $\Delta\mathbf{v}_i$ takes a negative value proportional to the downregulated reaction. Such scaling introduces a more stringent constraint on $\Delta\mathbf{v}_i$.

We tested the threshold prescribed by Equations (B.3) - (B.4) above for the Ishii *et al.* study, and compared the predicted flux changes $\Delta\mathbf{v}$ with those obtained with a constant threshold in Equations (B.1) - (B.2). The introduction of thresholds that scales proportionally with the fold changes of the reaction expression did not significantly alter the predictions of $\Delta\mathbf{v}$, as shown in Figure B.1. The difference in the sign accuracy and the correlation were only marginal (Mean correlation coefficient (ρ): Equations (B.1) - (B.2) threshold = 0.61, Equations (B.3) - (B.4) threshold = 0.58; Mean directional accuracy: Equations (B.1) - (B.2) threshold = 0.49, Equations (B.3) - (B.4) threshold = 0.48; Mean NRMSE: Equations (B.1) - (B.2) threshold = 0.14, Equations (B.3) - (B.4) threshold = 0.16).

B.2 Supplemental figures

Supplementary figures comparing the performance of Δ FBA's threshold variants in predicting *E. coli* metabolic response (Figure B.1), comparing the performance of Δ FBA in predicting *E. coli* metabolic response using whole-genome transcriptome data and RT-PCR mRNA data (Figure B.2) comparing the performance of Δ FBA and REMI in predicting *E. coli* metabolic response using whole-genome transcriptome data (Figure B.3).

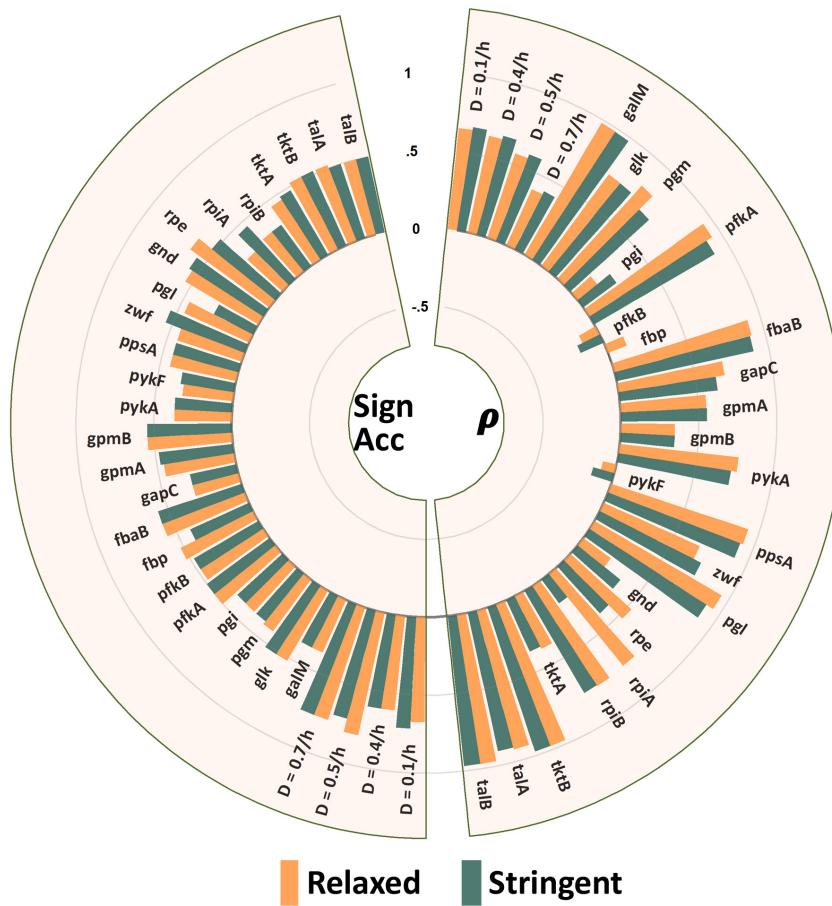


FIGURE B.1: Comparison of Δ FBA predictions of *E. coli* metabolic response to environmental and genetic variations using constant and fold-change criteria for minimum flux change magnitude. The accuracy of the flux predictions is assessed by calculating the Uncentered Pearson's Correlation Coefficient (ρ) the directional (sign) agreement between the predicted flux difference and the measured flux change for 46 reactions. The difference in flux change predictions between the two thresholds is insignificant.

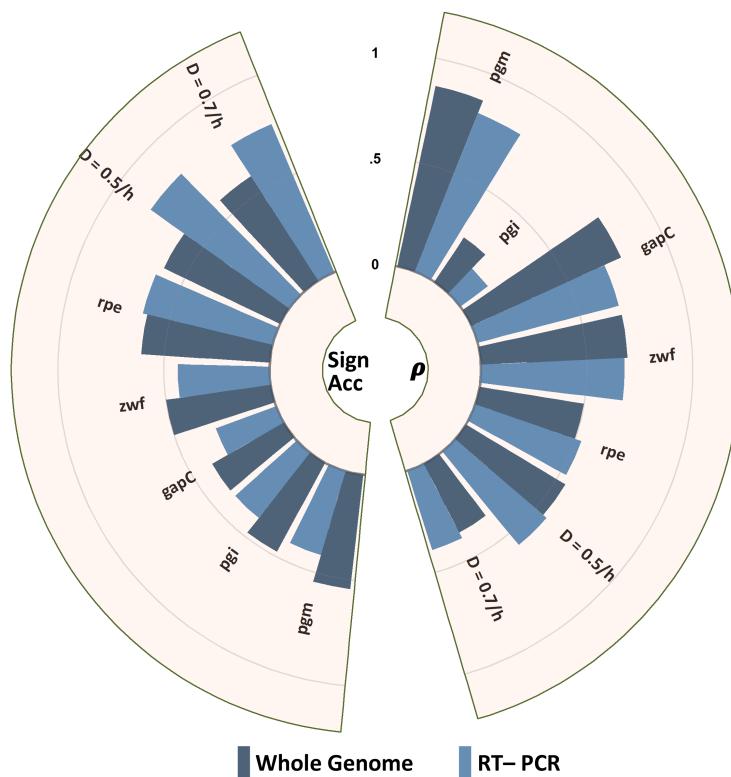


FIGURE B.2: Comparison of the performance of Δ FBA in predicting *E. coli* metabolic response to 2 dilution rate and 5 single-gene deletion perturbations using whole-genome transcriptome data and RT-PCR mRNA data for genes in the central carbon metabolism, glycolysis, PPP, and TCA cycle. Directional (Sign Accuracy) agreement and uncentered Pearson's Correlation Coefficient (ρ) between the predicted and measured flux difference for the 46 reactions have little difference between the incorporation of the two transcriptomic sources using Δ FBA (mean sign accuracy: whole-genome = 0.53; RT-PCR = 0.53; mean ρ : whole-genome = 0.57; RT-PCR = 0.54).

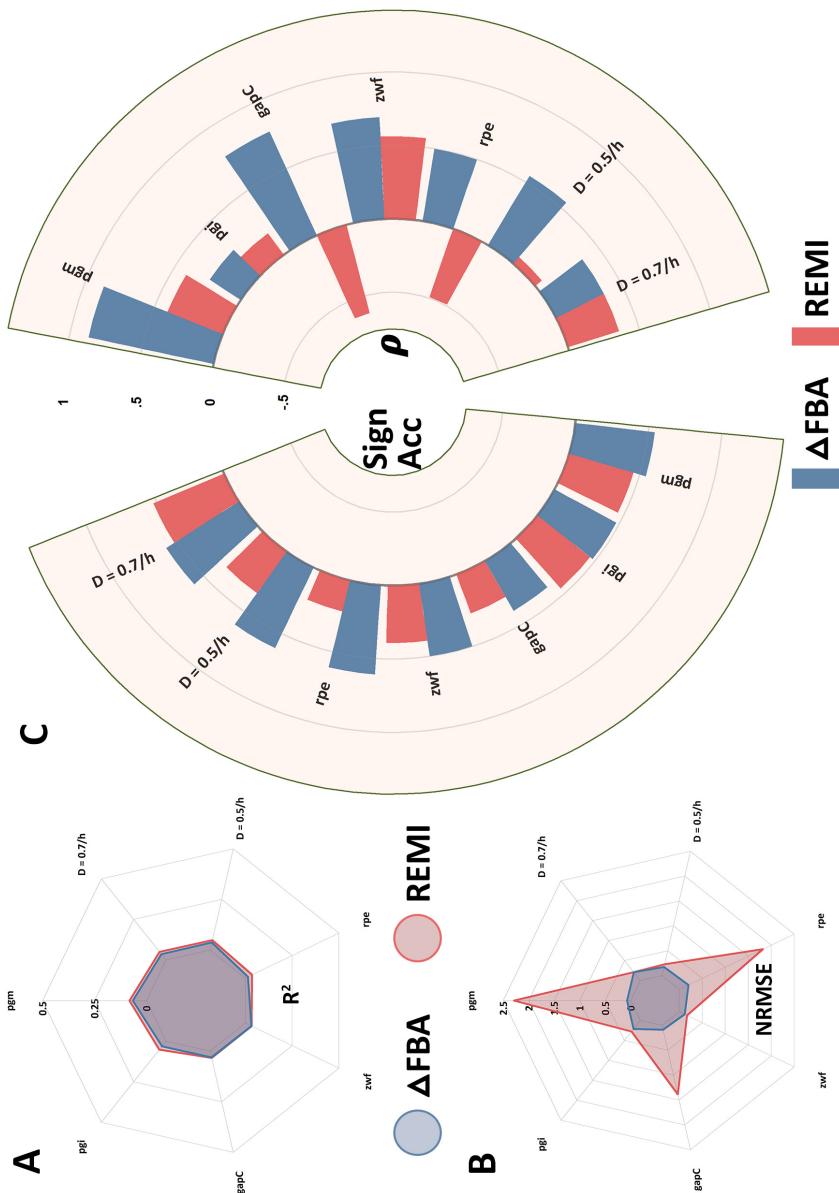


FIGURE B.3: Comparison of the performance of Δ FBA and REMI in predicting *E. coli* metabolic response to 2 dilution rate and 5 single-gene deletion perturbations using whole-genome transcriptome data. (A) The coefficient of determination (R^2) of the measured flux ratios and the reaction expression ratios remains close to zero. (B) Normalized Root Mean Square Error (NRMSE) of the flux difference normalized over the measured flux difference (Mean NRMSE: Δ FBA = 0.15; REMI = 0.91). (C) Directional (Sign Accuracy) agreement and uncentered Pearson's Correlation Coefficient (ρ) between the predicted and measured flux difference for the 46 reactions (mean sign accuracy: Δ FBA = 0.53, REMI = 0.38; mean ρ : Δ FBA = 0.57; REMI = 0.06).

C

**SUPPLEMENTARY INFORMATION:
MOVING WINDOW ANALYSIS OF THE
HUMAN TRANSCRIPTOME REVEALS
TEMPORAL ALTERATIONS
ASSOCIATED WITH AGING**

C.1 Functional hallmarks of ‘aging’ genes in GTEx

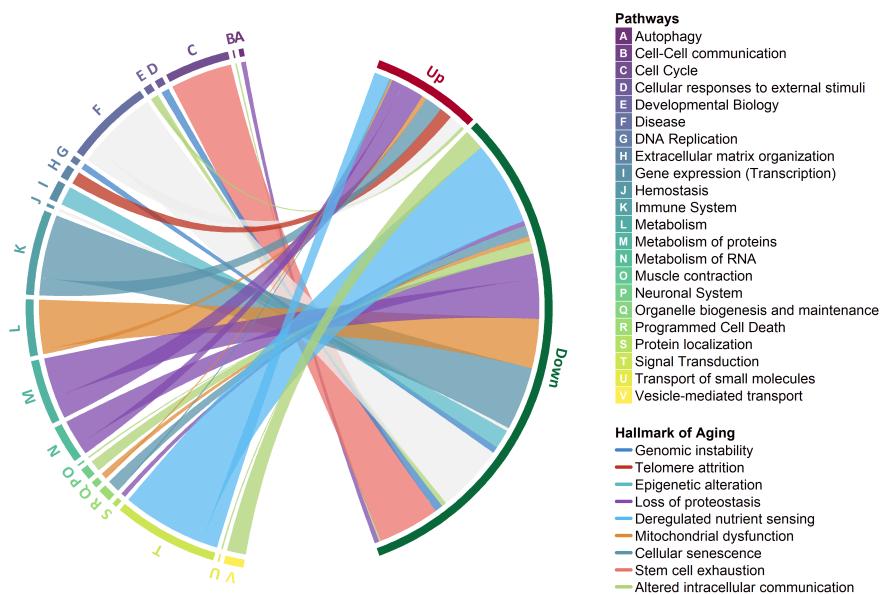


FIGURE C.1: Functional annotation of the aging genes shows organism-wide dysregulation in metabolism, stress response, signal transduction, cell cycle, and immune system in every hallmark of mammalian aging. Individual biological processes annotated using Reactome were classified based on hierarchical classifications presented in their database and their contribution to the hallmark of aging. The contribution of dysregulation from the downregulated set of genes shows a larger number of individual biological processes that are affected.

C.2 Metabolic subsystem enrichment analysis of the aging genes

Metabolic subsystem enrichment analysis of the aging genes was performed by:

- Mapping of the differential genes into a human genome-scale metabolic model, Recon3D using the Gene-Protein-Reaction (GPR) associations (Figure C.2)

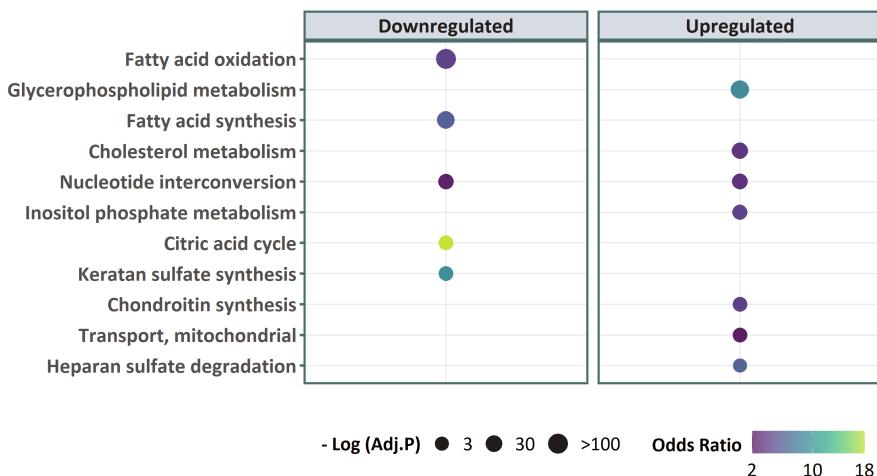


FIGURE C.2: Metabolic subsystem enrichment analysis of the aging genes shows a repressed citric acid and fatty acid metabolism and upregulation of cholesterol metabolism. Enriched subsystems ($\text{FDR} < 0.05$) are represented based on their significance (larger points have smaller adjusted p -values), and their odds ratio is conveyed using the color.

- Intracellular differential flux prediction by incorporating differential gene expression in Recon3D using Δ FBA (Figure C.3)

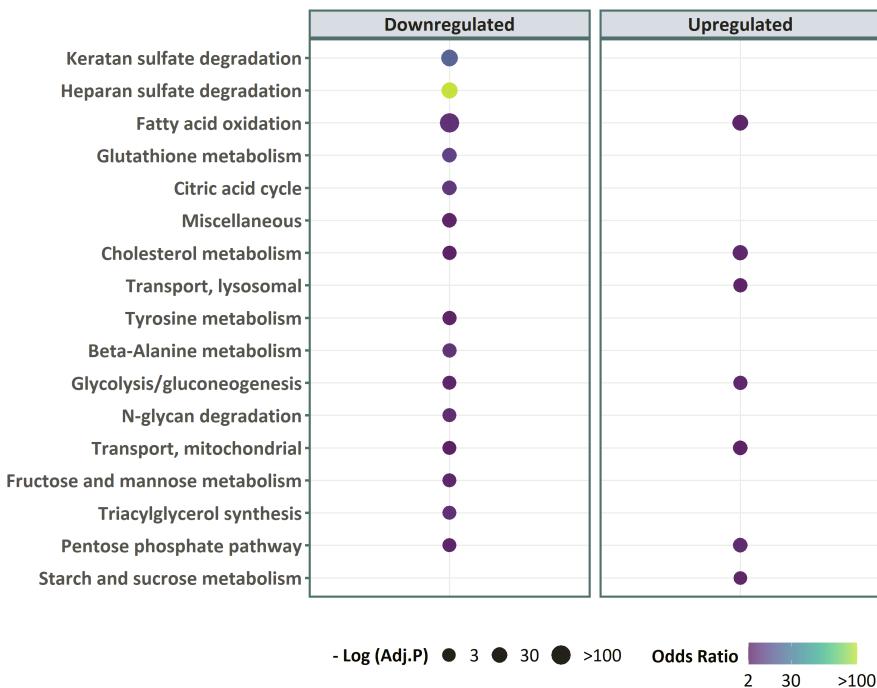


FIGURE C.3: Subsystems with significant dysregulation in material flow as predicted by Δ FBA mirror changes observed by merely mapping gene expression changes, including attenuated cellular and lipid energetics and dysfunction in cholesterol metabolism. Enriched subsystems ($FDR<0.05$) are represented based on their significance (larger points have smaller adjusted p -values), and their odds ratio is conveyed using the color.

C.3 Moving window analysis of the aging genes

		(a)				
$\times 10^{-2}$		30-39	40-49	50-59	60-69	70-79
20-29	0	8.59	4.38	3.30	4.04	
30-39		0	5.41	3.31	3.86	
40-49			3.02	2.28	5.21	
50-59				3.58	6.41	
60-69					1.57	

		(b)				
$\times 10^{-2}$		30-39	40-49	50-59	60-69	70-79
20-29	2.58	2.07	1.73	1.89	3.08	
30-39		1.81	1.58	1.73	2.81	
40-49			1.30	1.54	2.22	
50-59				1.31	1.67	
60-69					1.36	

TABLE C.1: Median effect sizes (Cohen's f^2) of (a) Differentially expressed genes with age and (b) all genes. The median effect size of differentially expressed genes is higher than the background effect sizes. Noticeably, any comparison to the 60's showed an increase in the number of differentially expressed genes and a decrease in the median effect size of these genes. This was reversed for a comparison to the 70's with the same starting point, suggest a weaker contribution of the 60's could be attributed to an increase in the statistical power due to a large sample size leading to easier identification of weakly changing transcripts.

C.4 Metabolic subsystem enrichment analysis of genes in the moving windows using Δ FBA

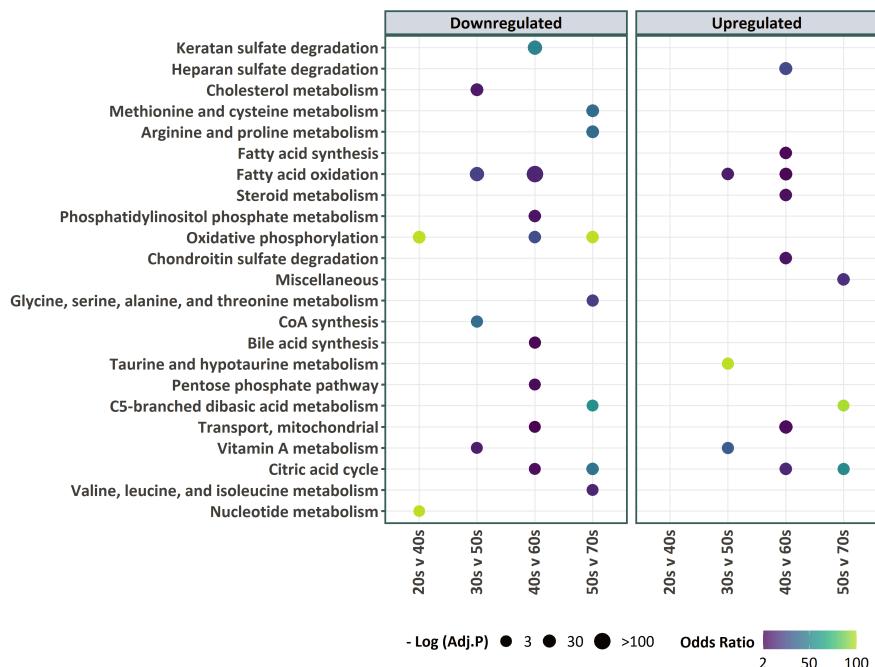


FIGURE C.4: Δ FBA predictions of genes that show significant changes in windows selected to best study the temporal alterations in the gene expression show early energetic dysregulation before a number of changes appear later in life. Enriched subsystems ($FDR < 0.05$) are represented based on their significance (larger points have smaller adjusted p -values), and their odds ratio is conveyed using the color.

C.5 Classification of genes in the moving window analysis

Classification	Gene Category	Differentially expressed genes		
		All	Upregulated	Downregulated
Strong–Persistent	[1,1,1,1]	2	2	0
Moderate–Persistent	[1,1,1,0]	10	9	1
Early–Moderate	[1,0,0,0]	29	7	22
	[1,0,1,0]	28	14	13
	[1,0,0,1]	1	0	1
	[1,0,1,1]	12	5	7
	[0,1,0,0]	87	25	60
Mid Life–Moderate	[0,1,1,0]	342	186	154
	[0,1,1,1]	29	21	7
	[0,0,1,1]	653	278	374
Late–Moderate	[0,0,0,1]	474	174	300
	[0,0,1,0]	7557	4081	3473
Non–DE	[0,0,0,0]	25508	4084	3723
LMM–MW	LMM–MW	2953	1315	1638
Zero	[1,1,0,0]	0	0	0
	[0,1,0,1]	0	0	0
	[1,1,0,1]	0	0	0

TABLE C.3: Classification of gene categories in the moving window analysis based on their timing of change and their cumulative effect size profiles.

C.6 Transcription factor enrichment analysis of persistent and early changing genes

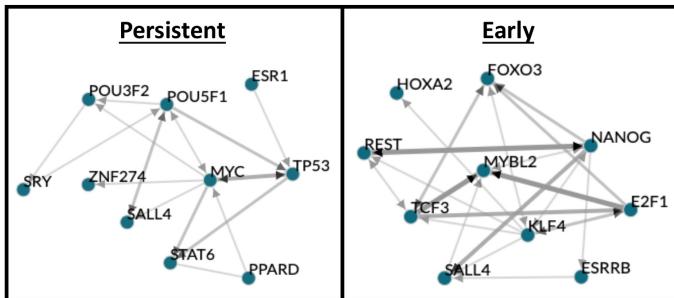


FIGURE C.5: Transcription factor enrichment analysis of persistent and early changing genes that shows top 10 TFs prioritized based on literature evidence connecting the persistent and early changing genes.

C.7 Metabolic subsystem enrichment analysis of temporally changing genes

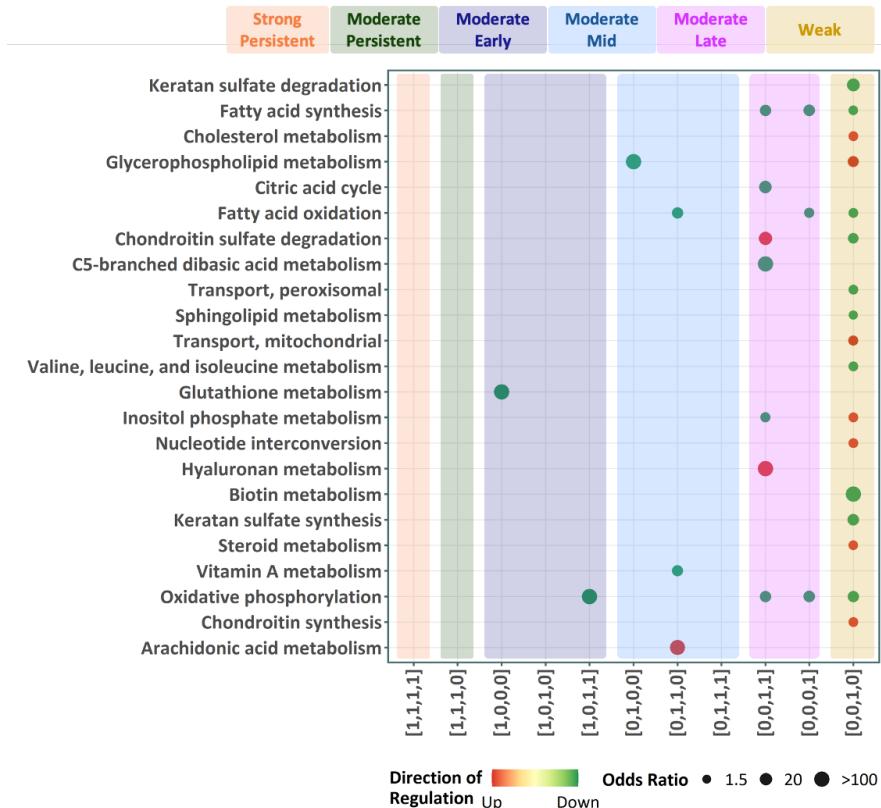


FIGURE C.6: Metabolic subsystem enrichment analysis of temporally changing genes. Moderate early repression in oxidative phosphorylation continues to late-life stages and additional repression in cellular energetics and lipid metabolism. Enriched subsystems ($\text{FDR} < 0.05$) are represented based on their significance (larger points have smaller adjusted p -values), and their color represents the overall directional contribution of individual genes.

B I B L I O G R A P H Y

1. Beguerisse-Díaz, M. *et al.* Flux-dependent graphs for metabolic networks. *NPJ Syst Biol Appl* **4**, 32 (2018).
2. Tomar, N. & De, R. K. *Comparing methods for metabolic network analysis and an application to metabolic engineering* 2013.
3. Pavlova, N. N. & Thompson, C. B. *The Emerging Hallmarks of Cancer Metabolism* 2016.
4. Piedrafita, G., Keller, M. A. & Ralser, M. The impact of non-enzymatic reactions and enzyme promiscuity on cellular metabolism during (Oxidative) stress conditions. *Biomolecules* **5**, 2101 (2015).
5. Oberhardt, M. A., Palsson, B. & Papin, J. A. *Applications of genome-scale metabolic reconstructions* 2009.
6. Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* **8**, e1002518 (2012).
7. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. *Molecular Systems Biology* (2011).
8. Gu, C. *et al.* Current status and applications of genome-scale metabolic models. *Genome Biol* **20**, 121 (2019).
9. Granata, I. *et al.* Integration of transcriptomic data in a genome-scale metabolic model to investigate the link between obesity and breast cancer. *BMC Bioinformatics* **20**, 162 (2019).
10. Wang, F.-S. *et al.* Genome-Scale Metabolic Modeling with Protein Expressions of Normal and Cancerous Colorectal Tissues for Oncogene Inference. *Metabolites* **10**, 16 (2019).
11. Sánchez, B. J. *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology* **13**, 935 (2017).
12. Zhang, C. & Hua, Q. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Front Physiol* **6**, 413 (2015).

13. Katzir, R. *et al.* The landscape of tiered regulation of breast cancer cell metabolism. *Scientific Reports* **9**, 1 (2019).
14. Çakir, T. & Khatibipour, M. J. *Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation* 2014.
15. Thiele, I. & Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* **5**, 93 (2010).
16. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* **28**, 245 (2010).
17. Palsson, B. O. *Systems biology: Constraint-based reconstruction and analysis* 1 (Cambridge University Press, 2015).
18. Angione, C. Human Systems Biology and Metabolic Modelling: A Review—From Disease Metabolism to Precision Medicine. *BioMed Research International* **2019** (2019).
19. Ramon, C., Gollub, M. G. & Stelling, J. *Integrating -omics data into genome-scale metabolic network models: Principles and challenges* 2018.
20. O'Brien, E. J., Monk, J. M. & Palsson, B. O. *Using genome-scale models to predict biological capabilities* 2015.
21. Becker, S. A. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4** (ed Sauro, H. M.) e1000082 (2008).
22. Sajitz-Hermstein, M. *et al.* iReMet-flux: constraint-based approach for integrating relative metabolite levels into a stoichiometric metabolic models. *Bioinformatics* **32**, i755 (2016).
23. Salvy, P. & Hatzimanikatis, V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun* **11**, 30 (2020).
24. O'Brien, E. J. *et al.* Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* **9**, 693 (2013).
25. Zhang, W., Li, F. & Nie, L. *Integrating multiple 'omics' analysis for microbial biology: Application and methodologies* 2010.
26. Mayr, M. *Metabolomics: ready for the prime time?* 2008.
27. Wang, Z., Gerstein, M. & Snyder, M. *RNA-Seq: A revolutionary tool for transcriptomics* 2009.
28. Malone, J. H. & Oliver, B. *Microarrays, deep sequencing and the true measure of the transcriptome* 2011.

29. Machado, D. *et al.* Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS computational biology* **10** (ed Maranas, C. D.) e1003580 (2014).
30. Zur, H., Ruppin, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics (Oxford, England)* **26**, 3140 (2010).
31. Kim, D.-K., Kim, T. H. & Lee, S.-J. Mechanisms of aging-related proteinopathies in *Caenorhabditis elegans*. *Experimental & Molecular Medicine* **48**, e263 (2016).
32. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* (2010).
33. Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* (2011).
34. Zhu, L. *et al.* A computational method using differential gene expression to predict altered metabolism of multicellular organisms. *Molecular BioSystems* **13**, 2418 (2017).
35. Pandey, V., Hadadi, N. & Hatzimanikatis, V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol* **15** (ed Patil, K. R.) e1007036 (2019).
36. Johnson, S. C. *et al.* Genetic evidence for common pathways in human age-related diseases. *Aging Cell* **14**, 809 (2015).
37. Richardson, A. & Schadt, E. The role of macromolecular damage in aging and age-related disease. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* **69** (2014).
38. Yang, J. *et al.* Discover the network mechanisms underlying the connections between aging and age-related diseases. *Scientific Reports* **6**, 32566 (2016).
39. Liao, C. Y. *et al.* Genetic variation in the murine lifespan response to dietary restriction: From life extension to life shortening. *Aging Cell* **9**, 92 (2010).
40. Harper, J. M., Leathers, C. W. & Austad, S. N. Does caloric restriction extend life in wild mice? *Aging Cell* **5**, 441 (2006).

41. Medvedev, Z. A. An attempt at a rational classification of theories of ageing. *Biological reviews of the Cambridge Philosophical Society* **65**, 375 (1990).
42. FRANCESCHI, C. *et al.* Inflamm-aging: An Evolutionary Perspective on Immunosenescence. *Annals of the New York Academy of Sciences* **908**, 244 (2006).
43. Dai, D.-F. *et al.* Mitochondrial oxidative stress in aging and healthspan. *Longevity & Healthspan* **3**, 6 (2014).
44. Wallace, D. C. Mitochondrial DNA mutations in disease and aging. *Environmental and Molecular Mutagenesis* **51**, n/a (2010).
45. Mostoslavsky, R. *et al.* Genomic Instability and Aging-like Phenotype in the Absence of Mammalian SIRT6. *Cell* **124**, 315 (2006).
46. Vijg, J. & Suh, Y. Genome Instability and Aging. *Annual Review of Physiology* **75**, 645 (2013).
47. Von Zglinicki, T. & Martin-Ruiz, C. Telomeres as biomarkers for ageing and age-related diseases. *Current Molecular Medicine* **5** (2005).
48. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239 (2000).
49. Richelle, A. *et al.* Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Computational Biology* **15** (2019).
50. Wang, Y., Eddy, J. A. & Price, N. D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol* **6**, 153 (2012).
51. Vlassis, N., Pacheco, M. P. & Sauter, T. Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLoS Computational Biology* **10**, 1003424 (2014).
52. Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol* **6**, 401 (2010).
53. Ferreira, J., Correia, S. & Rocha, M. Analysing Algorithms and Data Sources for the Tissue-Specific Reconstruction of Liver Healthy and Cancer Cells. *Interdisciplinary Sciences: Computational Life Sciences* **9**, 36 (1234).

54. Åkesson, M., Förster, J. & Nielsen, J. Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering* (2004).
55. Tian, M. & Reed, J. L. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics (Oxford, England)* **34**, 3882 (2018).
56. Lee, D. *et al.* Improving metabolic flux predictions using absolute gene expression data. *BMC Systems Biology* **6**, 1 (2012).
57. Navid, A. & Almaas, E. Genome-level transcription data of Yersinia pestis analyzed with a New metabolic constraint-based approach. *BMC Systems Biology* **6**, 1 (2012).
58. Lewis, N. E. *et al.* Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* **6**, 390 (2010).
59. Finkel, T. *The metabolic regulation of aging* 2015.
60. Catic, A. in *Progress in Molecular Biology and Translational Science* **85** (Elsevier B.V., 2018).
61. López-Otín, C. *et al.* Leading Edge Review Metabolic Control of Longevity (2016).
62. López-Otín, C. *et al.* *The Hallmarks of Aging* 2013.
63. Stegeman, R. & Weake, V. M. *Transcriptional Signatures of Aging* 2017.
64. Frenk, S. & Houseley, J. Gene expression hallmarks of cellular ageing. *Biogerontology*, 1 (2018).
65. Halsey, L. G. The reign of the p -value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters* **15**, 20190174 (2019).
66. Lin, W. J., Hsueh, H. M. & Chen, J. J. Power and sample size estimation in microarray studies. *BMC Bioinformatics* **11**, 48 (2010).
67. Matsui, S. & Noma, H. Estimating Effect Sizes of Differentially Expressed Genes for Power and Sample-Size Assessments in Microarray Experiments. *Biometrics* **67**, 1225 (2011).
68. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols* **6**, 1290 (2011).
69. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660 (2015).

70. Teo, E. *et al.* Metabolic stress is a primary pathogenic event in transgenic *Caenorhabditis elegans* expressing pan-neuronal human amyloid beta. *eLife* **8** (2019).
71. Mirra, S. S. *et al.* The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479 (1991).
72. Gibson, G. E., Sheu, K. F. & Blass, J. P. Abnormalities of mitochondrial enzymes in Alzheimer disease. *Journal of Neural Transmission* **105**, 855 (1998).
73. Müller, W. E. *et al.* *Mitochondrial dysfunction: Common final pathway in brain aging and alzheimer's disease-therapeutic aspects* in *Molecular Neurobiology* **41** (Humana Press, 2010), 159.
74. De La Monte, S. M. & Wands, J. R. *Alzheimer's disease is type 3 diabetes-evidence reviewed* 2008.
75. Yao, J. *et al.* Mitochondrial bioenergetic deficit precedes Alzheimer's pathology in female mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14670 (2009).
76. Ye, X., Tai, W. & Zhang, D. *The early events of Alzheimer's disease pathology: From mitochondrial dysfunction to BDNF axonal transport deficits* 2012.
77. Swomley, A. M. *et al.* *Abeta, oxidative stress in Alzheimer disease: Evidence based on proteomics studies* 2014.
78. Butterfield, D. A. *et al.* Redox proteomics identification of oxidatively modified hippocampal proteins in mild cognitive impairment: Insights into the development of Alzheimer's disease. *Neurobiology of Disease* **22**, 223 (2006).
79. Butterfield, D. A. & Halliwell, B. *Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease* 2019.
80. Reed, T. T. *et al.* Proteomic identification of nitrated brain proteins in early Alzheimer's disease inferior parietal lobule. *Journal of Cellular and Molecular Medicine* **13**, 2019 (2009).
81. Mccoll, G. *et al.* Utility of an improved model of amyloid-beta ($A\beta_{1-42}$) toxicity in *Caenorhabditis elegans* for drug screening for Alzheimer's disease. *Molecular Neurodegeneration* **7** (2012).

82. Link, C. D. Expression of human β -amyloid peptide in transgenic *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9368 (1995).
83. Treusch, S. *et al.* Functional links between A β toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334**, 1241 (2011).
84. Fong, S. *et al.* Energy crisis precedes global metabolic failure in a novel *Caenorhabditis elegans* Alzheimer Disease model. *Scientific Reports* **6** (2016).
85. Bray, N. L. *et al.* Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525 (2016).
86. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
87. Ignatiadis, N. *et al.* Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods* **13**, 577 (2016).
88. Newgard, C. B. *et al.* A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metabolism* **9**, 311 (2009).
89. Sinha, R. A. *et al.* Caffeine stimulates hepatic lipid metabolism by the autophagy-lysosomal pathway in mice. *Hepatology* **59**, 1366 (2014).
90. Muoio, D. M. *et al.* Muscle-specific deletion of carnitine acetyltransferase compromises glucose tolerance and metabolic flexibility. *Cell Metabolism* **15**, 764 (2012).
91. Yilmaz, L. S. *et al.* A *Caenorhabditis elegans* Genome-Scale Metabolic Network Model. *Cell Systems* **2**, 297 (2016).
92. Smith, C. Physiology of the Bacterial Cell. A Molecular Approach. *Biochemical Education* **20**, 124 (1992).
93. Voorhies, W. A. V. HEAD-TO-HEAD DEBATE ON METABOLISM AND LONGEVITY : The influence of metabolic rate on longevity in the nematode *Caenorhabditis elegans* *, 91 (2002).
94. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11** (2010).
95. Goo, C. K. *et al.* PTEN/Akt Signaling Controls Mitochondrial Respiratory Capacity through 4E-BP1. *PLoS ONE* **7** (2012).

96. Schooneman, M. G. *et al.* *Acylcarnitines: Reflecting or inflicting insulin resistance?* 2013.
97. Rinaldo, P., Cowan, T. M. & Matern, D. *Acylcarnitine profile analysis* 2008.
98. Berg, J. M., Tymoczko, J. L. & Stryer, L. *The Citric Acid Cycle* (2002).
99. Zhao, Y. *et al.* Loss of succinyl-CoA synthase ADP-forming β subunit disrupts mtDNA stability and mitochondrial dynamics in neurons. *Scientific Reports* **7**, 1 (2017).
100. Paterson, R. W. *et al.* A targeted proteomic multiplex CSF assay identifies increased malate dehydrogenase and other neurodegenerative biomarkers in individuals with Alzheimer's disease pathology. *Translational Psychiatry* **6** (2016).
101. Nussbaum-Krammer, C. I. & Morimoto, R. I. *Caenorhabditis elegans as a model system for studying non-cellautonomous mechanisms in protein-misfolding diseases* 2014.
102. Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol* **10**, 721 (2014).
103. Nilsson, A. & Nielsen, J. *Genome scale metabolic modeling of cancer* 2017.
104. Lewis, N. E. & Abdel-Haleem, A. M. The evolution of genome-scale models of cancer metabolism. *Frontiers in Physiology* **4**, 237 (2013).
105. Zhang, W. *et al.* New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biology* **16**, 202 (2015).
106. Santos, F., Boele, J. & Teusink, B. in *Methods in Enzymology* **509** (Academic Press Inc., 2011).
107. Bordbar, A. *et al.* Constraint-based models predict metabolic and associated cellular functions. *Nature reviews Genetics* **15**, 107 (2014).
108. Lewis, N. E., Nagarajan, H. & Palsson, B. O. *Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods* 2012.
109. Park, J. H. & Lee, S. Y. *Towards systems metabolic engineering of microorganisms for amino acid production* 2008.
110. Kim, T. Y. *et al.* *Strategies for systems-level metabolic engineering* 2008.

111. Nevoigt, E. Progress in Metabolic Engineering of *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **72**, 379 (2008).
112. Kim, H. U., Kim, T. Y. & Lee, S. Y. Metabolic flux analysis and metabolic engineering of microorganisms. *Molecular BioSystems* (2008).
113. Ebrahim, A. *et al.* Do genome-scale models need exact solvers or clearer standards? *Molecular Systems Biology* (2015).
114. Hyduke, D. R., Lewis, N. E. & Palsson, B. O. *Analysis of omics data with genome-scale models of metabolism* 2013.
115. Pandey, V. & Hatzimanikatis, V. Investigating the deregulation of metabolic tasks via Minimum Network Enrichment Analysis (MiNEA) as applied to nonalcoholic fatty liver disease using mouse and human omics data. *PLoS Comput Biol* **15** (ed Wallqvist, A.) e1006760 (2019).
116. Pusa, T. *et al.* MOOMIN - Mathematical explOration of 'Omics data on a Metabollc Network. *Bioinformatics* (2020).
117. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols* **14**, 639 (2019).
118. Ishii, N. *et al.* Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. **316** (2007).
119. Gerosa, L. *et al.* Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Systems* (2015).
120. Väremo, L. *et al.* Proteome- and Transcriptome-Driven Reconstruction of the Human Myocyte Metabolic Network and Its Use for Identification of Markers for Diabetes. *Cell Reports* (2015).
121. Segrè, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15112 (2002).
122. Opdam, S. *et al.* A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models Article A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Systems* **4**, 318 (2017).
123. Athar, A. *et al.* ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Research* **47**, D711 (2019).

124. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* (2015).
125. Van Tienen, F. H. *et al.* Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *Journal of Clinical Endocrinology and Metabolism* (2012).
126. Jin, W. *et al.* Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. *Journal of Clinical Investigation* (2011).
127. Lerin, C. *et al.* Defects in muscle branched-chain amino acid oxidation contribute to impaired lipid metabolism. *Molecular Metabolism* (2016).
128. Misra, P. & Chakrabarti, R. *The role of AMP kinase in diabetes* 2007.
129. Ly, L. D. *et al.* Oxidative stress and calcium dysregulation by palmitate in type 2 diabetes. *Experimental & Molecular Medicine* (2017).
130. Igoillo-Esteve, M. *et al.* Palmitate induces a pro-inflammatory response in human pancreatic islets that mimics CCL₂ expression by beta cells in type 2 diabetes. *Diabetologia* (2010).
131. Das, U. N. *Arachidonic acid in health and disease with focus on hypertension and diabetes mellitus: A review* 2018.
132. King, Z. A. *et al.* *Next-generation genome-scale models for metabolic engineering* 2015.
133. Xu, F. *et al.* Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. *Journal of Clinical Endocrinology and Metabolism* (2013).
134. *The relationship between phospholipids and insulin resistance: From clinical to experimental studies* 2019.
135. Harman, D. The aging process: major risk factor for disease and death. *88*, 5360 (1991).
136. Yang, J. *et al.* Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep* **5**, 15145 (2015).
137. Welle, S. *et al.* Skeletal muscle gene expression profiles in 20–29 year old and 65–71 year old women. *Experimental Gerontology* **39**, 369 (2004).
138. Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nature communications* **6**, 8570 (2015).

139. Yuet, K. P. *et al.* Cell-specific proteomic analysis in *< i>Caenorhabditis elegans</i>*. *Proceedings of the National Academy of Sciences* **112**, 2705 (2015).
140. Rodwell, G. E. J. G. *et al.* A Transcriptional Profile of Aging in the Human Kidney. *PLoS Biology* **2** (ed Thomas Kirkwood) e427 (2004).
141. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome biology* **14**, R75 (2013).
142. Lund, J. *et al.* Transcriptional Profile of Aging in *C. elegans*. *Current Biology* **12**, 1566 (2002).
143. Pletcher, S. D. *et al.* Genome-Wide Transcript Profiles in Aging and Calorically Restricted *Drosophila melanogaster*. *Current Biology* **12**, 712 (2002).
144. Gheorghe, M. *et al.* Major aging-associated RNA expressions change at two distinct age-positions. *BMC Genomics* **15**, 132 (2014).
145. Haustead, D. J. *et al.* Transcriptome analysis of human ageing in male skin shows mid-life period of variability and central role of NF- κ B. *Scientific reports* **6**, 26846 (2016).
146. Valdes, A., Glass, D. & Spector, T. Omics technologies and the study of human ageing. *Nature Reviews Genetics* **14** (2013).
147. Ardlie, K. G. K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648 (2015).
148. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272 (2018).
149. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Molecular BioSystems* **12**, 477 (2016).
150. Clarke, D. J. B. *et al.* eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Research* **46**, W171 (2018).
151. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Research* (2018).
152. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* **47**, W212 (2019).

153. Berchtold, N. C. *et al.* Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences* **105**, 15605 (2008).
154. Postnikoff, S. D. L. & Harkness, T. A. A. Mechanistic insights into aging, cell-cycle progression, and stress response. *Frontiers in physiology* **3**, 183 (2012).
155. Ryan, N. A. *et al.* Lower skeletal muscle capillarization and VEGF expression in aged vs. young men. *Journal of applied physiology (Bethesda, Md. : 1985)* (2006).
156. Vavken, P., Saad, F. A. & Murray, M. M. Age dependence of expression of growth factor receptors in porcine ACL fibroblasts. *Journal of Orthopaedic Research* (2010).
157. Oxenkrug, G. Interferon-gamma - Inducible Inflammation: Contribution to Aging and Aging-Associated Psychiatric Disorders. *Aging and disease* **2**, 474 (2011).
158. Tilstra, J. S. *et al.* NF- κ B in Aging and Disease. *Aging and disease* **2**, 449 (2011).
159. Meschiari, C. A. *et al.* The impact of aging on cardiac extracellular matrix. *GeroScience* **39**, 7 (2017).
160. Harries, L. W. *et al.* Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868 (2011).
161. Mazin, P. *et al.* Widespread splicing changes in human brain development and aging. *Molecular Systems Biology* (2013).
162. Blanco, F. J. & Bernabeu, C. Alternative splicing factor or splicing factor-2 plays a key role in intron retention of the endoglin gene during endothelial senescence. *Aging Cell* **10**, 896 (2011).
163. Somel, M. *et al.* MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. *Genome Research* **20**, 1207 (2010).
164. Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883 (2004).
165. Lipinski, M. M. *et al.* Genome-wide analysis reveals mechanisms modulating autophagy in normal brain aging and in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14164 (2010).

166. De Vries, M. *et al.* Lung tissue gene-expression signature for the ageing lung in COPD. *Thorax* **73**, 609 (2018).
167. Yang, S. M. *et al.* A flow-free droplet-based device for high throughput polymorphic crystallization. *Lab on a Chip* **15**, 2680 (2015).
168. Zeng, L. *et al.* Transcriptome analysis reveals the difference between “healthy” and “common” aging and their connection with age-related diseases. *Aging Cell* **19** (2020).
169. Baker, D. J., Jin, F. & Van Deursen, J. M. *The yin and yang of the Cdkn2a locus in senescence and aging* 2008.
170. Ago, T. *et al.* The NADPH oxidase Nox4 and aging in the heart. *Aging* **2**, 1012 (2010).
171. Kelly, J. *et al.* Gene expression meta-analysis of Parkinson’s disease and its relationship with Alzheimer’s disease. *Molecular Brain* **12**, 16 (2019).
172. Lanke, V. *et al.* Integrative analysis of hippocampus gene expression profiles identifies network alterations in aging and Alzheimer’s disease. *Frontiers in Aging Neuroscience* **10**, 153 (2018).
173. Xiong, X. *et al.* Ribosomal protein S27-like is a physiological regulator of p53 that suppresses genomic instability and tumorigenesis. *eLife* **3**, 1 (2014).
174. González-Velasco, O. *et al.* Transcriptomic landscape, gene signatures and regulatory profile of aging in the human brain. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1863**, 194491 (2020).
175. Sanese, P. *et al.* FOXO3 on the Road to Longevity: Lessons From SNPs and Chromatin Hubs 2019.
176. Denechaud, P. D., Fajas, L. & Giralt, A. *E2F1, a novel regulator of metabolism* 2017.
177. Xie, Q. *et al.* E2F transcription factor 1 regulates cellular and organismal senescence by inhibiting forkhead box O transcription factors. *Journal of Biological Chemistry* **289**, 34205 (2014).
178. Stein, D. & Toiber, D. *DNA damage and neurodegeneration: The unusual suspect* 2017.
179. Papakonstantinou, E., Roth, M. & Karakiulakis, G. *Hyaluronic acid: A key molecule in skin aging* 2012.
180. Hanley, M. R. *et al.* *Neural function: metabolism and actions of inositol metabolites in mammalian brain.* 1988.

181. Shlomi, T., Berkman, O. & Ruppin, E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7695 (2005).