

## ABSTRACT

Title of Dissertation: METABOLIC MODELING OF  
ANTI-CANCER CD8 T CELLS  
AND SARS-COV-2 INFECTION

Kuoyuan Cheng  
Doctor of Philosophy, 2021

Dissertation Directed by: Dr. Eytan Ruppin, Department of Computer Science  
Prof. Philip Johnson, Department of Biology

Genome-scale metabolic models (GEMs) are mathematical network models that comprehensively encompass all cellular metabolic reactions, enzymes, and metabolites, allowing the in silico modeling and simulation of cell metabolism in a holistic manner using mathematical programming techniques. With the advent of higher-quality GEMs and effective modeling algorithms, GEM analysis has been repeatedly shown to generate accurate predictions and informative hypotheses for metabolism research. In this dissertation, I first describe the comprehensive evaluation and improvement of a GEM algorithm named metabolic transformation algorithm (MTA), which predicts the metabolic targets whose inhibition can facilitate the transformation between two arbitrary cellular metabolic states. The new MTA algorithm is shown to achieve significantly better prediction accuracy than the previous version using a large collection of validation datasets on known metabolic gene knockouts/knockdowns or metabolic protein inhibitions. Next I demonstrate the broad

applications of MTA and the GEM analysis framework in both basic and translational research.

First, MTA was used to identify metabolic processes essential to the anti-tumor function of CD8 T cells in cancer immunotherapy. T cells play a central role in cancer immunosurveillance and current cancer immunotherapies. In order to improve the patient response rate to immunotherapy, it is pivotal to understand the factors regulating T cell function. It has been recognized that metabolism can greatly affect different aspects of T cell function, including differentiation, cytokine production, longevity and exhaustion. Via the MTA analysis of adoptive cell transfer (ACT) therapy, including CAR-T therapy datasets, the mitochondrial uncoupling protein UCP2 was predicted as a key determinant of ACT therapy response and CD8 T cell anti-tumor function. GEM analysis also points to the generation of reactive oxygen species (ROS) as a mechanism underlying the effect of UCP2. The role of UCP2 in T cells and the ROS-related mechanisms were thoroughly validated with genetic knockout/overexpression and pharmacologic inhibition experiments in a Pmel-1/B16 ACT mice model.

Second, MTA was used to predict anti-SARS-CoV-2 targets that act via counteracting the virus-induced metabolic alterations. The COVID-19 pandemic caused by the SARS-CoV-2 coronavirus has been ongoing for about a year to date, with still very few therapeutic options available. Viruses, including coronaviruses, are known to hijack the host metabolism to facilitate their own proliferation, making targeting host metabolism a promising antiviral approach. I performed an integrated GEM analysis of 12 published in vitro and human patient gene expression datasets on

SARS-CoV-2 infection, revealing a wide-spread metabolic reprogramming induced by the virus. MTA successfully predicted metabolic anti-SARS-CoV-2 targets, with significant overlap with the experimentally validated targets identified from drug and genetic screens. Further, MTA was used to identify potential targets for combining with remdesivir (an approved anti-SARS-CoV-2 drug) to achieve higher efficacy.

In conclusion, I developed the improved MTA algorithm with an updated GEM workflow, which were applied to cancer immunotherapy and coronavirus infection. The integrated GEM analysis identified mitochondrial uncoupling as an essential metabolic process in anti-tumor CD8 T cells, and effectively predicted host metabolism-targeting anti-SARS-CoV-2 strategies. This demonstrates the value of GEM in a broad range of metabolism-related research problems.

METABOLIC MODELING OF ANTI-CANCER CD8 T CELLS  
AND SARS-COV-2 INFECTION

by

Kuoyuan Cheng

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2021

Advisory Committee:

Professor Philip Johnson, Chair/Advisor  
Dr. Eytan Ruppin, Co-Advisor  
Professor Stephen M. Mount  
Professor Najib M. El-Sayed  
Dr. Sridhar Hannenhalli  
Dr. Stefan Ambs

© Copyright by  
Kuoyuan Cheng  
2021

## Acknowledgments

I would start by expressing my heartfelt gratitude to Dr. Eytan Ruppin, who has been the chief advisor of my PhD. It has been a great fortune for me to join Eytan's lab and work under his mentoring, from which I have benefited a lot. His enthusiasm for science and vision for the future of our field has been a constant inspiration that motivates me to keep learning new things and taking on new challenges. More importantly, I will always be indebted to his care and kind help to me that extends beyond academic life.

I would also like to thank my committee for being extremely supportive and helpful towards my study and research. In particular, I truly appreciate Dr. Sridhar Hannenhalli for his time and patience in having many discussions with me and providing invaluable guidance to my projects. I would like to thank Prof. Philip Johnson and Prof. Stephen Mount who kindly helped me out and joined my committee later in my studies, and Prof. Najib El-Sayed, Dr. Stefan Ambs, as well as Dr. Hector Bravo (my former committee member) for their guidance and responsiveness to my questions and requests. During my PhD I have had the honor to work with many other excellent senior scientists and PIs, including Dr. Alejandro Schaffer, Dr. Michael Gertz, Dr. Cenk Sahinalp, Dr. Nicholas Restifo, Dr. Ayelet Erez, Dr. Yardena Samuels, Dr. Ze'ev Ronai, Dr. Ani Deshpande, Dr. Sumit Chanda,

Dr. Ariel Israel, Dr. Yuval Tabach, Dr. Chi-Ping Day, Dr. Max Leiserson, Dr. Yongmei Zhao, and many more. I have learned a lot from them about doing science and scientific collaboration, and I would like to thank them all.

Furthermore, I would like to thank all my lab mates and friends from NIH/CDSL, BISI/CBBG, CBCB and beyond. It is really hard to name all the people who have extended their kindness towards me and made my PhD journey a significantly better one, but I would like to mention a few including Dr. Kunal Kundu, Dr. Nathan Olson and Sushant Patkar. Additional special thanks to Sanju Sinha, Dr. Nishanth Ulhas Nair, Dr. Kun Wang, Dr. Joo Sang Lee, Dr. Madhusudhanan Sukumar, and Dr. Lipika R. Pal, whom I have worked closely with – it has been a true pleasure to learn from them, many of whom have also helped me a lot in life beyond research. I want to thank Dr. Noam Auslander and Dr. Keren Yizhak for sharing codes and answering my questions on metabolic modeling, and Dr. Laura Riva for biological experimental work on one of my thesis projects. Besides I would like to thank the staff at UMD and NIH for their kind help, including Dr. Michelle Brooks, Gwen Warman, Dr. Zakiya Whatley, Barbara Lewis, Nadia Nimley, as well as the faculties Dr. Hadiya Woodham, Prof. Charles Delwiche and Prof. Eric Haag.

Last but not least, I thank my family and my wife Jiaqi for their unconditional love and support. Throughout my journey they have been with me sharing the same joy, and the same grief from loss. I would not have reached where I am without them.



# Table of Contents

Acknowledgements . . . . .	ii
Table of Contents . . . . .	v
List of Tables . . . . .	ix
List of Figures . . . . .	x
List of Abbreviations . . . . .	xiv
Chapter 1: Introduction . . . . .	1
1.1 Cell metabolism . . . . .	1
1.1.1 Overview . . . . .	1
1.1.2 Overview of cell metabolic pathways in human . . . . .	3
1.2 Genome-scale metabolic modeling . . . . .	19
1.2.1 Genome-scale metabolic models (GEMs) . . . . .	19
1.2.2 The methodological framework of GEM . . . . .	24
1.2.3 GEM algorithms for omic data integration . . . . .	30
1.2.4 Additional GEM algorithms relevant to this thesis . . . . .	39
1.3 Metabolism of T cells in cancer and cancer immunotherapy . . . . .	42
1.3.1 Overview of cancer immunology . . . . .	42
1.3.2 Cancer immunotherapy, immune checkpoint blockade therapy and factors associated with patient response . . . . .	46
1.3.3 Adoptive cell transfer therapy . . . . .	48
1.3.4 Immunometabolism in T cell differentiation and function . . . . .	50
1.4 Metabolism in COVID-19 and host-SARS-CoV-2 interaction . . . . .	54
1.4.1 COVID-19: natural history, epidemiology and clinical features . . . . .	55
1.4.2 SARS-CoV-2 biology . . . . .	58
1.4.3 Anti-COVID-19 therapies and vaccines . . . . .	61
1.4.4 Viral regulation of host metabolism . . . . .	63
1.5 Overview of study: rationale, aims and significance . . . . .	65
Chapter 2: Analysis and improvement of metabolic transformation algorithms . . . . .	66
2.1 Introduction . . . . .	66
2.2 Results . . . . .	67
2.2.1 The MTA algorithm as in (Yizhak et al., 2013) . . . . .	68

2.2.2	Evaluating the performance of the MTA algorithm in human datasets . . . . .	73
2.2.3	First proposed alternative to the original MTA and its evaluation	76
2.2.4	Second proposed alternative to the original MTA and its evaluation . . . . .	81
2.2.5	Comparison with the rMTA algorithm and a benchmark of all MTA variants . . . . .	84
2.2.6	Comparison of the metabolic modeling-based algorithms with gene expression-based analysis . . . . .	90
2.3	Materials and Methods . . . . .	93
2.3.1	Validation datasets . . . . .	93
2.3.2	Generating the common inputs to different MTA algorithms .	93
2.3.3	Implementation of the different MTA algorithms . . . . .	94
2.3.4	Evaluation and comparison of the different MTA algorithms .	95
2.3.5	Softwares and code . . . . .	95
2.4	Discussion and Conclusion . . . . .	95
Chapter 3:	Metabolic modeling of CD8 T cell for anti-cancer immunotherapy .	100
3.1	Introduction . . . . .	100
3.2	Results . . . . .	101
3.2.1	Identification of the metabolic fluxes in CD8 T cells associated with cancer immunotherapy response . . . . .	102
3.2.2	MTA identified T cell mitochondrial uncoupling as essential for the anti-cancer function and immunotherapy response . . .	104
3.2.3	Analysis of large-scale human transcriptome datasets identified that mitochondrial uncoupling protein <i>UCP2</i> expression is associated with CD8 T cell stemness and ICB response. . .	108
3.2.4	<i>Ucp2</i> inhibition in mice T cells decreased T cell stemness marker expression and reduced anti-cancer function . . . . .	110
3.2.5	GEM modeling supports increased reactive oxygen species as a mechanism underlying the effect of <i>Ucp2</i> . . . . .	113
3.2.6	The antioxidant N-acetylcysteine suppresses the increased ROS level in mice CD8 T cell due to <i>Ucp2</i> inhibition and reversed T cell function . . . . .	116
3.3	Materials and Methods . . . . .	119
3.3.1	Differential expression and gene set enrichment analysis . . . .	119
3.3.2	Computation of metabolic fluxes from gene expression data with GEM and differential flux analysis . . . . .	120
3.3.3	Genome-scale metabolic modeling-based prediction of metabolic functions determining response to anti-CD19 CAR-T therapy	121
3.3.4	Metabolic pathway analysis of the MTA-predicted metabolic reactions . . . . .	122
3.3.5	Genome-scale metabolic modeling-based prediction of metabolic processes that possibly mediates the effect of <i>Ucp2</i> . . . . .	123

3.3.6	Analysis of the expression levels of mitochondrial uncoupling genes across human tissues or in various ICB datasets . . . . .	123
3.3.7	Analysis of the correlation between <i>UCP2</i> and several T cell function/stemness marker gene expression across TCGA cancer types . . . . .	124
3.3.8	Softwares and code . . . . .	125
3.4	Discussion and Conclusion . . . . .	125
Chapter 4:	Modeling of host metabolism during SARS-CoV-2 infection . . . . .	131
4.1	Introduction . . . . .	131
4.2	Results . . . . .	132
4.2.1	Analysis of SARS-CoV-2-induced gene expression changes across datasets . . . . .	132
4.2.2	GEM-based analysis of SARS-CoV-2-induced metabolic flux change patterns . . . . .	136
4.2.3	Prediction and validation of metabolic anti-SARS-CoV-2 targets with MTA . . . . .	141
4.2.4	Prediction of metabolic targets for anti-SARS-CoV-2 in combination with remdesivir . . . . .	145
4.3	Materials and Methods . . . . .	148
4.3.1	DE analysis . . . . .	149
4.3.2	GSEA analysis of the differential expression results . . . . .	150
4.3.3	Comparison of the differentially expressed genes and pathways across datasets . . . . .	150
4.3.4	Computation of metabolic fluxes from gene expression data with GEM . . . . .	151
4.3.5	Differential flux analysis of virus-infected vs control group in each dataset . . . . .	152
4.3.6	Analysis of reactions with consistent differential fluxes across datasets and their pathway enrichment analysis . . . . .	153
4.3.7	Differential transport flux analysis by each metabolite species . . . . .	154
4.3.8	Analysis of the consistent flux alteration patterns in different metabolic pathways . . . . .	154
4.3.9	Prediction of anti-SARS-CoV-2 target metabolic reactions with metabolic transformation algorithm . . . . .	155
4.3.10	Validation of the MTA-predicted anti-SARS-CoV-2 metabolic targets . . . . .	156
4.3.11	Defining and analyzing the consensus set of candidate anti-SARS-CoV-2 metabolic targets across datasets . . . . .	158
4.3.12	Preparation of Vero E6 cell samples with SARS-CoV-2 infection and remdesivir treatment, RNA-sequencing, and gene expression data analysis . . . . .	158
4.3.13	Genome-scale metabolic modeling of the remdesivir-treated Vero E6 cell samples and prediction of anti-SARS-CoV-2 metabolic targets in combination with remdesivir . . . . .	160

4.3.14 Softwares and code . . . . .	161
4.4 Discussion and Conclusion . . . . .	161
Conclusion . . . . .	167
Appendices . . . . .	169
Chapter 2 . . . . .	169
Chapter 3 . . . . .	175
Chapter 4 . . . . .	177
Bibliography . . . . .	187

## List of Tables

1.1	Qualitative summary of relative performances of different GEM algorithms for integrating gene expression data, in different prediction tasks or quality control assessments. The two ratings (before and after slashes) correspond to models generated with general medium and cell-specific medium, respectively. med: medium. . . . .	36
3.1	DE results of selected T cell differentiation and function marker genes in genipin-treated vs control mice Pmel-1 CD8 T cells. . . . .	111
3.2	Enrichment of top MTA predictions whose KO can reverse the effect of genipin in ROS producing reactions . . . . .	115
3.3	The predicted percent rank of several cellular antioxidation reaction in the MTA prediction of reactions whose KO can mimic the effect of genipin treatment . . . . .	115
3.4	DE results of selected T cell differentiation and function marker genes in genipin+NAC-treated vs genipin-treated mice Pmel-1 CD8 T cells.	117
4.1	Summary of the published gene expression datasets on SARS-CoV-2 infection analyzed in this study . . . . .	133
4.2	Gene expression datasets involving metabolic gene perturbation (KO/KD/protein inhibition) collected for the validation of the MTA algorithm and its variants . . . . .	169

## List of Figures

1.1	A diagram summarizing the carbohydrate metabolism pathway. This figure is taken from the book Color Atlas of Biochemistry. . . . .	6
1.2	A diagram summarizing the TCA cycle and its multiple functions. This figure is taken from the book Color Atlas of Biochemistry. . . . .	9
1.3	A diagram showing the OXPHOS, ROS generation and mitochondrial uncoupling protein. This figure is taken from Zhao et al. 2019. . . . .	11
1.4	A diagram the nucleotide de novo synthesis pathway. This figure is taken from the book Color Atlas of Biochemistry. . . . .	18
2.1	A schematic diagram as an overview of the MTA algorithm. . . . .	68
2.2	Evaluation of the prediction performance of MTA on 58 human datasets using different parameters. . . . .	75
2.3	Evaluation of the prediction performance of MeTAL on 58 human datasets using different parameters. . . . .	80
2.4	Evaluation of the prediction performance of mMTA on 58 human datasets using different parameters. . . . .	83
2.5	Evaluation of the prediction performance of mMeTAL on 58 human datasets using different parameters. . . . .	84
2.6	Evaluation of the prediction performance of rMTA on 58 human datasets using different parameters. . . . .	88
2.7	Evaluation of the prediction performance of MOMA on 58 human datasets using different parameters. . . . .	89
2.8	Summary of the predicted percentage ranks of the ground truth reactions across all the 58 datasets using the various algorithms with their optimal parameters. . . . .	90
2.9	Evaluation of the prediction performance of MOMA on 58 human datasets using the MeTAL objective function with different parameter values. . . . .	91
2.10	Summary of the predicted percentage ranks of the ground truth reactions across all the 9 datasets where the ground truth gene is not among the top 20% DE genes, using the various algorithms with their optimal parameters. . . . .	92
3.1	Metabolic pathway enrichment of reactions with consistent differential fluxes in the CAR-T cells of responders vs non-responders and in the persistent vs non-persistent T cell clones. . . . .	103

3.2	Computed metabolic flux distributions of the ETC/OXPHOS reactions in the T cells of anti-CD19 CAR-T therapy responders vs non-responders. . . . .	104
3.3	Metabolic pathway enrichment of the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy. . . . .	106
3.4	Enrichment for mitochondrial transport reactions by each metabolite by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy. . . . .	106
3.5	Comparing the MTA score ranks of the mitochondrial proton transport reactions contributing to higher mitochondrial proton influx in responders vs non-responders and those contributing to higher mitochondrial proton efflux in responders vs non-responders . . . . .	107
3.6	Total mitochondrial proton influx in the T cells of the non-responders and responders to anti-CD19 CAR-T therapy. . . . .	108
3.7	Mitochondrial uncoupling reaction rate in the T cells of the non-responders and responders to anti-CD19 CAR-T therapy. . . . .	108
3.8	Expression levels of UCP1, UCP2 and UCP3 across human tissues. . . . .	109
3.9	Correlation of UCP2 expression and the expression of several T cell stemness and memory marker genes across TCGA cancer types after controlling for tumor purity. . . . .	110
3.10	GSEA table and plots showing the enrichment of various T cell effector function and T cell exhaustion-associated pathways in genipin-treated vs control mice Pmel-1 CD8 T cells. . . . .	112
3.11	Tumor growth in B16 mice melanoma model treated by ACT of vehicle-treated CD8 T cells or genipin-treated CD8 T cells. Experimental groups: stars (untreated); red circles (WT T cells); black circles (genipin-treated T cells). Figure credit: Dr. Madhusudhanan Sukumar . . . . .	113
3.12	Total metabolic fluxes through mitochondrial and cytosol hydrogen peroxide and superoxide anion in genipin-treated Pmel-1 CD8 T cells vs vehicle-treated T cells. O <sub>2</sub> <sup>-</sup> : superoxide anion; H <sub>2</sub> O <sub>2</sub> : hydrogen peroxide; suffix '[m]' means mitochondrial, '[c]' means cytosol. The dot in each group denotes the mean of each distribution. . . . .	114
3.13	Mitochondrial ROS levels as measured by MitoSOX in genipin-treated mice Pmel-1 CD8 T cells, with or without NAC. Experimental groups: Veh (vehicle); Gen (genipin-treated); Gen+NAC (genipin+NAC-treated); NAC (NAC-treated). Figure credit: Dr. Madhusudhanan Sukumar . . . . .	117
3.14	GSEA table and plots showing the enrichment of various T cell effector function and T cell exhaustion-associated pathways in genipin+NAC-treated vs genipin-treated (without NAC) mice Pmel-1 CD8 T cells. . . . .	118
3.15	Tumor growth in B16 mice melanoma model treated by ACT of genipin-treated CD8 T cells or genipin+NAC-treated CD8 T cells. NT: non-treated; 25-Genipin: treatment with 25 micro molar genipin. Figure credit: Dr. Madhusudhanan Sukumar . . . . .	119

4.1	PCA plot of the datasets based on the inverse normal-transformed DE log fold-change values. . . . .	133
4.2	Pairwise comparison of the datasets by their top significant DE genes with Fisher’s exact tests. . . . .	134
4.3	Summary of the top consistently enriched pathways from GSEA analysis across the analyzed datasets. . . . .	136
4.4	Summary of the metabolic pathways enriched by the metabolic reactions showing the most consistent changes across the analyzed datasets.	137
4.5	Summary of the metabolites showing the most consistent transportation patterns between the extracellular space and cytosol across the analyzed datasets. . . . .	138
4.6	Visualization of the consistent flux changes in the pyrimidine biosynthesis pathway. . . . .	139
4.7	Visualization of the consistent flux changes in the fatty acid elongation pathway. Red and blue denotes increased and decreased absolute flux values, respectively; grey denotes inconsistent changes. . . . .	141
4.8	Summary of the enrichment of top MTA predictions from each dataset in the antiviral gene targets identified from CRISPR-Cas9 screens. . .	143
4.9	Summary of the enrichment of top MTA predictions from each dataset in the antiviral drugs identified from various experimental drug screens.	143
4.10	Summary of the AUROC of top MTA predictions from each dataset, based on validated positive and negative hits from CRISPR-Cas9 screens. . . . .	144
4.11	Summary of the metabolic pathway enrichment result of the consensus MTA-predicted targets across datasets. . . . .	145
4.12	PCA plot on the transcriptome data from control Vero E6 cells as well as SARS-CoV-2-infected cells with or without remdesivir treatment .	146
4.13	Summary of the metabolic pathway enrichment of the top differential fluxes between the virus+remdesivir and the control Vero E6 cells . .	148
4.14	Summary of the metabolic pathway enrichment result of the top MTA-predicted targets for combination with remdesivir . . . . .	148
4.15	Comparison of the predicted percentage ranks of the ground truth reactions across datasets from rMTA, mMTA and MOMA, with the best parameter combinations. . . . .	172
4.16	The prediction performance of rMTA across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to rMTA. . . . .	173
4.17	The prediction performance of mMeTAL across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to mMeTAL. . . . .	174
4.18	The prediction performance of MOMA across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to MOMA. . . . .	174

4.19	Computed metabolic flux distributions of the ETC/OXPHOS reactions in the 9.2-P persistent T cells vs the 9.1-NP non-persistent T cells in ACT. . . . .	175
4.20	GSEA plot showing the enrichment for the mitochondrial transport pathway by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy. . . . .	175
4.21	GSEA plot showing the enrichment for the mitochondrial proton transport reactions by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy. . . . .	175
4.22	Mitochondrial uncoupling reaction rate in the 9.2-P persistent T cells vs the 9.1-NP non-persistent T cells in ACT. . . . .	176
4.23	UCP2 expression levels in the tumor samples from responders and non-responders in various ICB datasets. . . . .	176
4.24	Flow cytometry results on the level of various T cell activation, effector function, memory and senescence gene markers in genipin-treated vs control mice Pmel-1 CD8 T cells. The gene markers are as follows: T cell activation (CD25, CD44), memory (CD62L), effector molecule (GzmB), senescence (KLRG1) and exhaustion (PD1). Figure credit: Dr. Madhusudhanan Sukumar . . . . .	177
4.25	Pairwise comparison of the datasets by their top significantly enriched pathways with Fisher's exact tests. . . . .	177
4.26	Summary of the metabolic pathway changes from GSEA analysis across the analyzed datasets. . . . .	178
4.27	Pairwise comparison of the datasets by their top positive DF reactions with Fisher's exact tests. . . . .	178
4.28	Pairwise comparison of the datasets by their top negative DF reactions with Fisher's exact tests. . . . .	179
4.29	Visualization of the consistent flux changes in the TCA cycle. . . . .	180
4.30	Visualization of the consistent flux changes in the reactions involving the cytosolic counterparts of TCA cycle metabolites. . . . .	181
4.31	Visualization of the consistent flux changes in the inositol phosphate metabolism pathway. . . . .	182
4.32	Visualization of the consistent flux changes in the glycine, serine, and threonine metabolism pathway. . . . .	183
4.33	Summary of the enrichment of top MTA predictions from each dataset in the host-virus PPI proteins. . . . .	183
4.34	An example ROC curve for the Calu-3 dataset. . . . .	184
4.35	An example ROC curve for the Liao et al. single-cell dataset. . . . .	184
4.36	Summary of the significant pathways with differentially expressed genes between the virus+remdesivir and the control Vero E6 cells . . . . .	185
4.37	PCA plot on the iMAT-computed metabolic fluxes from control Vero E6 cells as well as SARS-CoV-2-infected cells with or without remdesivir treatment . . . . .	186

## List of Abbreviations

Ac-CoA	Acetyl Coenzyme A
ACE2	Angiotensin-Converting Enzyme 2
ACT	Adoptive Cell Transfer
ACHR	Artificial Centering Hit-and-Run
ALL	Acute Lymphoblastic Leukemia
ATP	Adenosine Triphosphate
AUROC	Area Under Receiver Operating Characteristic Curve
BALF	Bronchoalveolar Lavage Fluid
CAR-T	Chimeric Antigen Receptor-T Cell
CBM	Constraint-Based Model (or Method/Modeling)
CHRR	Coordinate Hit-and-Run with Rounding
CLL	Chronic Lymphocytic Leukemia
COBRA	Constraint-Based Reconstruction and Analysis
COVID-19	Coronavirus Disease 2019
DE	Differential Expression
DLBCL	Diffuse Large B-Cell Lymphoma
DMV	Double-Membrane Vesicles
ER	Endoplasmic Reticulum
ERGIC	Endoplasmic Reticulum–Golgi Intermediate Compartment
ETC	Electron Transport Chain
EUA	Emergency Use Authorization
FAO	Fatty Acid Oxidation
FBA	Flux Balance Analysis
FDA	Food and Drug Administration
FVA	Flux Variability Analysis
GEM	Genome-Scale Metabolic Model (or Modeling)
GPR	Gene-Protein-Reaction Associations
GSEA	Gene Set Enrichment Analysis
HPI	Host-Pathogen Interaction
ICB	Immune Checkpoint Blockade
ICU	Intensive Care Unit
IFN	Interferon
IL	Interleukin
IMP	Inosine 5'-Monophosphate
iMAT	Integrative Metabolic Analysis Tool
KD	Knockdown
KO	Knockout
LP	Linear Programming

MERS	Middle East Respiratory Syndrome
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MHC	Major Histocompatibility Complexes
MIP	Mixed Integer Programming
MILP	Mixed Integer Linear Programming
MIQP	Mixed Integer Quadratic Programming
MOMA	Minimization of Metabolic Adjustment
MS	Mass Spectrometry
MTA	Metabolic Transformation Algorithm
NAC	N-Acetylcysteine
NAD <sup>+</sup> /NADH	Nicotinamide Adenine Dinucleotide
NADP <sup>+</sup> /NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NIH	National Institutes of Health
NK	Natural Killer
NLP	Non-Linear Programming
ORF	Open Reading Frame
OXPHOS	Oxidative Phosphorylation
PCA	Principle Component Analysis
PPI	Protein-Protein Interaction
PPP	Pentose Phosphate Pathway
PRPP	Phosphoribosyl Pyrophosphate
QP	Quadratic Programming
RBD	Receptor Binding Domain
RBM	Receptor Binding Motif
RdRP	RNA-Dependent RNA Polymerase
RNA-seq	RNA-Sequencing
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RTC	Replication and Transcription Complex
rMTA	Robust Metabolic Transformation Algorithm
SARS	Severe Acute Respiratory Syndrome
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
scRNA-seq	Single-Cell RNA-Sequencing
TCA	Tricarboxylic Acid
TCGA	The Cancer Genome Atlas
TCR	T Cell Receptor
THF	Tetrahydrofolate
TIL	Tumor-Infiltrating Lymphocyte
TMB	Tumor Mutational Burden
TME	Tumor Microenvironment
TRS	Transcription Regulatory Sequences
UMP	Uridine 5'-Monophosphate
WHO	World Health Organization
WT	Wild-Type

## Chapter 1: Introduction

### 1.1 Cell metabolism<sup>1</sup>

#### 1.1.1 Overview

All living organisms and cells must harness non-biological energy, convert it into biological energy and use it to drive all the biological processes needed for their survival, growth, proliferation, and other activities. Biological energy transduction obeys the laws of thermodynamics, including the conservation of energy in an isolated system (i.e. total amount of energy in the universe does not change) and the increase in disorder in an isolated system (i.e. the total entropy of the universe always increases). Under conditions of constant temperature and pressure, which (approximately) applies to most biological systems, the law of entropy increase leads to that for any process to happen spontaneously, the change in Gibbs free energy,  $\Delta G$ , must be negative. Living organisms and their cells utilize free energy, i.e. they rely on the decrease in Gibbs free energy to drive their biological activities. While some autotrophs like plants are able to utilize the energy of the radiation from the sun, other organisms including heterotrophs like animals exclusively harness

---

<sup>1</sup>a common reference for this section is the textbook Nelson, Cox, & Lehninger (2017)

the chemical energy from the molecules they consume in their food and from their environment. By maintaining and carefully regulating tens of thousands of chemical reactions in each cell, they harness the free energy release from the chemical compounds (metabolites) in a highly controlled manner, supporting their normal biological activities and the turn-over of their structural components. This highly complex and fine-tuned network of biochemical processes, the cell metabolism, is fundamental to all life without which no other biological processes or activity can happen.

Chemical reactions that the cell utilizes can happen spontaneously, i.e. they have negative  $\Delta G$ , termed thermodynamically unstable. However, although some of the biochemical reactions can readily happen without extra “assistance”, most biochemical reactions by themselves happen at very slow rates (i.e. kinetically stable). Therefore for the cell to harness these reactions efficiently, enzymes are needed for most reactions. Enzymes are biocatalysts that significantly increase the rate of chemical reactions. They are usually protein molecules (single or can be large complexes of multiple protein sub-units), which are not consumed in the reactions themselves, but act via decreasing the activation energy of the reaction they catalyze. Different enzymes can be confined within defined cellular compartments, the catalytic activities of each enzyme are usually highly specific to certain reactions, with the function levels/activities of the enzymes subject to swift and highly refined regulations. These mechanisms make it possible for the cell to regulate the entire metabolic network in a precise manner.

The phase of metabolism involving the degradation of larger organic molecules

such as carbohydrates, fats, proteins and nucleic acids into smaller products and releasing energy is called catabolism. The chemical energy harnessed during catabolism is often conserved in ATP molecules and reduced electron carriers such as NADH, NADPH and FADH<sub>2</sub>. The phase of metabolism involving the biosynthesis of smaller precursor molecules into larger biomolecules is called anabolism. Anabolism requires energy input which is usually in the form of the ATP molecules and reduced electron carriers which are produced during catabolism – more specifically, the involvement of these “energy-containing” molecules contribute to the spontaneity of the biosynthesis reactions under physiological conditions, allowing them to happen when needed. The cycle of catabolism and anabolism keeps the physical matter of life in constant turn-over accompanied by a constant free energy dissipation and entropy increase, in accordance with the physical laws of thermodynamics.

### 1.1.2 Overview of cell metabolic pathways in human

Although the entire cell metabolism has a highly intricate network-like structure, it is conventional to divide the cellular metabolic network conceptually into different parts or sub-systems, called pathways. A common way of pathway definition is based on the type of biomolecules involved, for example, pathways of carbohydrate metabolism, amino acid metabolism, lipid metabolism, nucleic acid metabolism, and so on. Additionally, a central part of cell metabolism responsible for the synthesis of ATP molecules (which acts as the “energy currency” of the cell) is usually termed Energy metabolism, this includes the mitochondrial electron

transport chain (ETC)/oxidative phosphorylation (OXPHOS) as well as the TCA cycle which is a central hub for multiple metabolic pathways and responsible for the production of reduced electron carriers which enter the ETC for ATP production. Below I provide a condensed high-level overview of the various key cellular metabolic pathways.

Glycolysis is the process of catabolizing glucose incompletely into pyruvate (via several important intermediate metabolites such as glucose 6-phosphate, 3-phosphoglycerate and phospho-enolpyruvate), producing energy in the forms of ATP and NADH. Depending on whether oxygen is available, pyruvate may be converted into lactate (during which the NADH previously generated is consumed) which is excreted and may be subsequently recycled to produce glucose via gluconeogenesis, or enter the mitochondria and be used for the synthesis of acetyl coenzyme A (Ac-CoA; which then enters the TCA cycle for further catabolism). Gluconeogenesis, on the other hand, is an anabolism pathway for the synthesis of glucose from pyruvate or other glycolytic/TCA cycle intermediates. Via these intermediates, many different types of molecules, including some amino acids and glycerol from triacylglycerols can be used for gluconeogenesis, although mammals including human cannot use fatty acids or Ac-CoA for gluconeogenesis. Gluconeogenesis overall appears to happen in the reverse direction compared to glycolysis but involve several distinct steps of reactions which ensure its irreversibility and differential regulation from glycolysis. Notably, the bypass reaction from pyruvate to phospho-enolpyruvate happens via oxaloacetate and malate in the mitochondria, allowing TCA cycle intermediates to be used for gluconeogenesis. While glucose (together with fructose) is a central

metabolic in carbohydrate metabolism, different monosaccharides can be interconverted via specific pathways. Pentose phosphate pathway (PPP) partly can act to interconvert monosaccharides with different numbers of carbon atoms (in the non-oxidative phase of the pathway), but it has another major oxidative phase which serves as an alternative pathway of glucose catabolism. PPP branches from an early upstream point of the glycolysis pathway (glucose 6-phosphate), ultimately oxidizes it completely into  $\text{CO}_2$  generating reduced electron carrier NADPH, which can not only be used for various biosynthesis processes but is also important for the cellular antioxidation and redox balance. With incomplete oxidation of glucose in PPP, ribose 5-phosphate is produced which is an important precursor for nucleotide and coenzyme synthesis. Glucose can be used to synthesize large carbohydrate molecules (polysaccharides) like glycogen (in the process of glycogenesis) for storage, and glycogen can also be broken down (glycogenolysis) into glucose; these dynamic interconversion process is regulated by hormones including adrenaline, glucagon and insulin. The carbohydrate metabolism as summarized above is illustrated in Figure 1.1 (Figure taken from (Koolman & Roehm, 2011)).

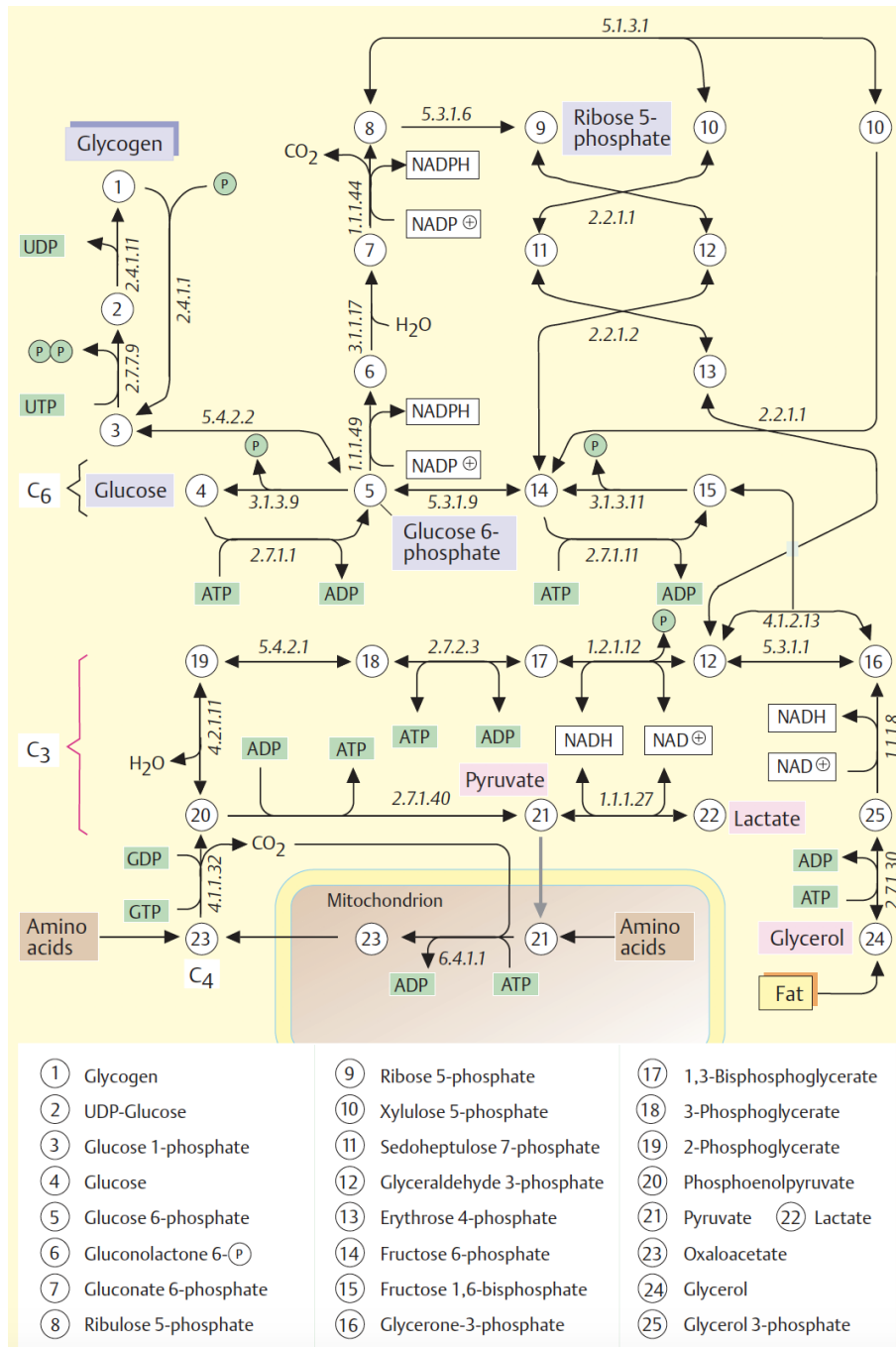


Figure 1.1: A diagram summarizing the carbohydrate metabolism pathway. This figure is taken from the book Color Atlas of Biochemistry.

As described above, the pyruvate generated in glycolysis can be oxidized to Ac-CoA and  $\text{CO}_2$  in the mitochondria, catalyzed by the pyruvate dehydrogenase complex. Ac-CoA, an important activated acetyl group donor, can then enter the

TCA cycle for further oxidation. TCA cycle, occurring in the mitochondrial matrix, is a pivotal cyclic energy metabolism pathway at the center of cell metabolism that can act as a hub connecting multiple pathways of carbohydrate, lipid and amino acid metabolism. The first step of the TCA cycle involves the reaction of Ac-CoA with oxaloacetate, forming citrate, which gives the cycle the other name (citric acid cycle). Within a cycle of about 8 steps involving via citrate, cis-aconitate, isocitrate,  $\alpha$ -ketoglutarate, succinyl-CoA, succinate, fumarate, and malate, oxaloacetate is regenerated ready to react with another molecule of Ac-CoA thus completing the cycle, while Ac-CoA is being oxidized generating NADH, GTP or ATP, FADH<sub>2</sub>, and CO<sub>2</sub>. The generated NADH and FADH<sub>2</sub> can further enter the mitochondrial ETC for the synthesis of ATP in OXPHOS. TCA pathway (together with OXPHOS) is a highly effective pathway for oxidizing glucose and generating energy in the form of ATP compared to the incomplete oxidation of glucose via glycolysis. One molecule of glucose when completely oxidized via TCA and OXPHOS can generate 30-32 molecules of ATP, while it can only generate 2 molecules of ATP (and 2 molecules of NADH, which if not consumed in lactate production can further produce about 5 molecules of ATP if entering ETC/OXPHOS) via glycolysis. The many steps and cyclic nature of TCA contributes to its important function as a hub of cell metabolism. The four- and five-carbon metabolites derived from many different catabolic pathways (e.g. those of amino acids) can be converted into and replenish TCA intermediates via the so-called anaplerotic reactions and thus enter the pathway for oxidation; many TCA intermediates can also be diverted away from the cycle and enter a wide variety of anabolic pathways (e.g. those of carbohydrates,

amino acids, fatty acids, porphyrins and heme, etc.). Therefore the TCA cycle is termed an amphibolic pathway, serving both catabolic and anabolic functions. Figure 1.2 (taken from (Koolman & Roehm, 2011)) summarizes the multiple functions of the TCA cycle.

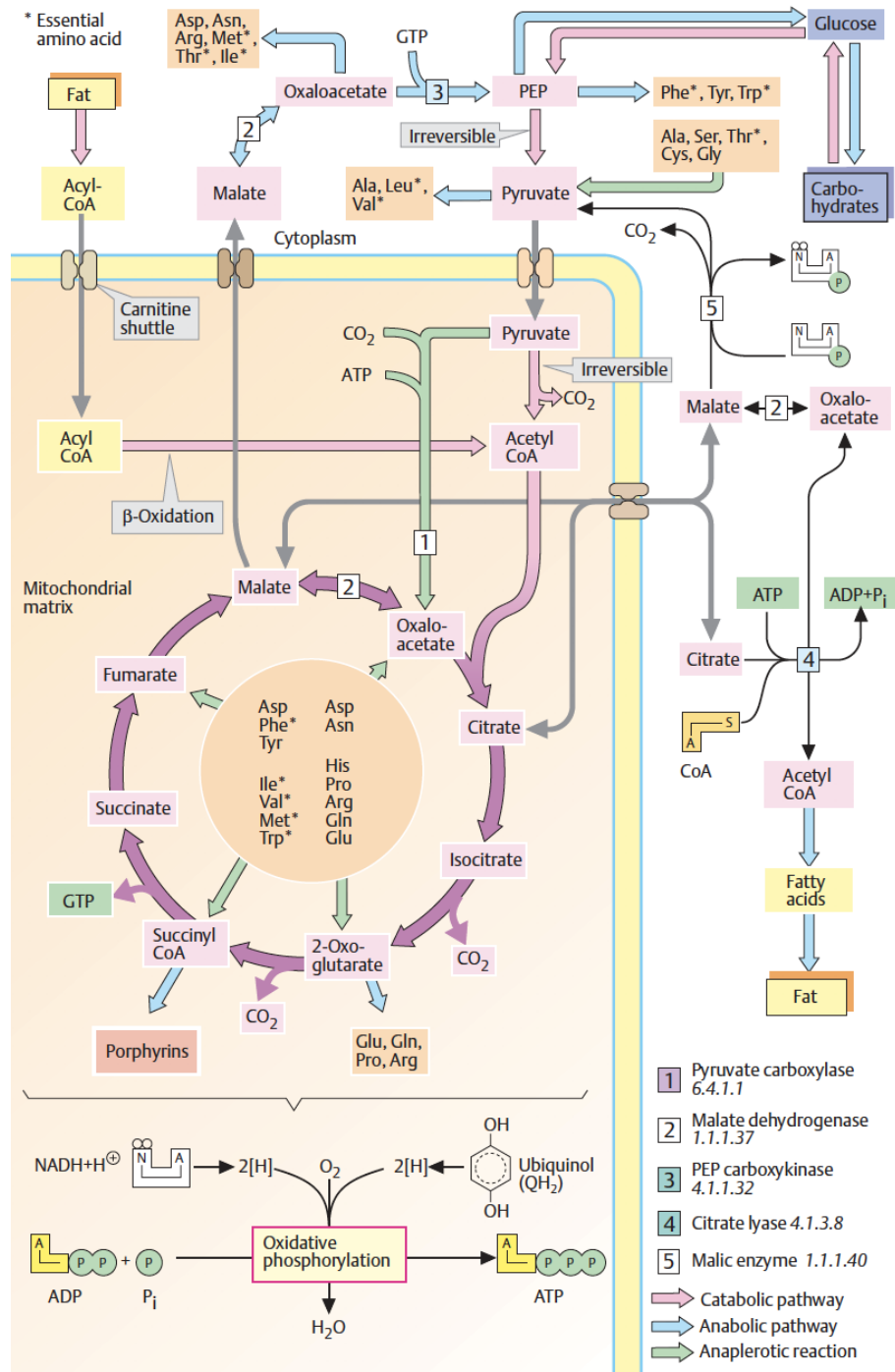


Figure 1.2: A diagram summarizing the TCA cycle and its multiple functions. This figure is taken from the book Color Atlas of Biochemistry.

The reduced electron carriers NADH and FADH<sub>2</sub> generated in the TCA cycle

and other cellular catabolic processes enter the mitochondrial ETC, also called respiratory chain, involving several large protein complexes located in the mitochondrial inner membrane, including Complex I (NADH:ubiquinone oxidoreductase), Complex II (succinate dehydrogenase, which is also part of the TCA cycle), Complex III (ubiquinone:cytochrome c oxidoreductase), and Complex IV (cytochrome oxidase). The electrons are passed through these complexes – from Complex I and II to Complex III via the mobile small molecule carrier ubiquinone (coenzyme Q), and from Complex III to Complex IV via the small protein cytochrome c. Complex IV pass the electron from cytochrome c to molecular oxygen and reduce it to water. Much of the energy released from these reactions are used by Complex I, III and IV, which are also proton pumps, to pump protons from the mitochondrial matrix into the inter-membrane space, creating a gradient in proton concentration. The chemical energy of this gradient is utilized by the ATP synthase (Complex V), which also locates in the mitochondrial inner membrane for synthesizing ATP from ADP and inorganic phosphate, coupled to the flow of proton back from the inter-membrane space into the mitochondrial matrix. This process of ATP synthesis, i.e. OXPHOS, is a pivotal metabolic process for generating the energy in the form of ATP that can be used for a wide variety of cellular processes. Mitochondrial uncoupling proteins located in the mitochondrial inner membrane can act to “leak” proton from the inter-membrane space into the matrix, thus dissipating the proton gradient created by the ETC for ATP synthesis. They have roles in regulating OXPHOS and heat generation, among others. ETC and OXPHOS is also a major source of reactive oxygen species (ROS) production in the cell, whose concentrations are regulated in

a highly refined manner as excessive ROS can cause cell damage, while ROS also serve as important signaling molecules. Mitochondrial uncoupling proteins thus also have roles in regulating ROS generation. These are illustrated in Figure 1.3 (taken from (Zhao, Jiang, Zhang, & Yu, 2019)).

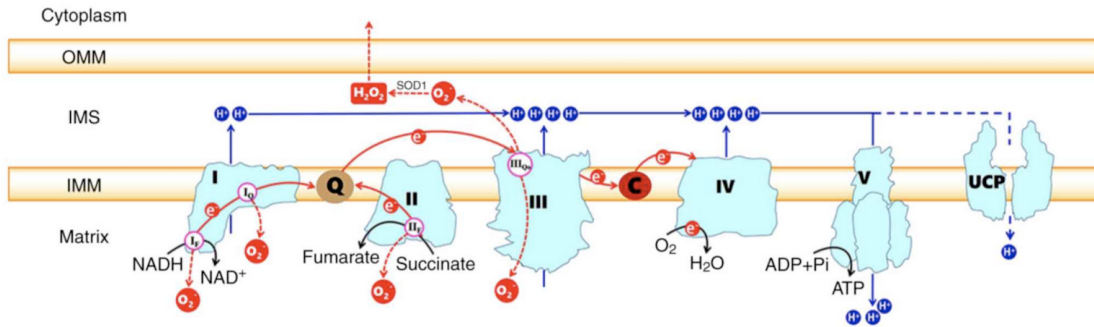


Figure 1.3: A diagram showing the OXPHOS, ROS generation and mitochondrial uncoupling protein. This figure is taken from Zhao et al. 2019.

Lipid, in the form of triacylglycerols, is broken down by lipase into glycerols and free fatty acids. While glycerol can be converted to glycerone 3-phosphate and enters glycolysis for further catabolism, fatty acids can be oxidized to generate large amounts of energy inside the mitochondria in the process called fatty acid oxidation (FAO). Free fatty acids are first activated into acyl-CoA in the cytosol, which is then transported into the mitochondria via the carnitine shuttle. Once inside the mitochondria, the acyl-CoA can undergo  $\beta$ -oxidation, which breaks down the carbon backbone of the fatty acid in a step-wise fashion (each time a two-carbon unit from the carboxyl end is removed) into Ac-CoA while producing NADH and FADH<sub>2</sub> in every step. Ac-CoA can enter the TCA cycle to produce more NADH and FADH<sub>2</sub>, which can enter ETC/OXPHOS to generate large amounts of ATP,

which also helps to produce a lot of water. FAO can produce more than two-fold the energy that can be produced by oxidizing the same weight of carbohydrate or protein. Varied reaction steps are needed in  $\beta$ -oxidation depending on the type of acyl chains (e.g. even or odd number of carbons, saturated or unsaturated, cis or trans configuration of unsaturated chains, etc.). Very-long-chain and branched-chain fatty acids, on the other hand need to be first catabolized into shorter products in the peroxisome via  $\beta$ -oxidation (which also produces  $H_2O_2$ ) or  $\alpha$ -oxidation (for branched-chain fatty acids), then transported to the mitochondria for further oxidation. In the endoplasmic reticulum (ER) exists a form of fatty acid oxidation called  $\omega$ -oxidation that breaks down fatty acid starting from the most distant carbon from the carboxyl group. For lipid anabolism, fatty acid can be biosynthesized from Ac-CoA in the cytosol (in many organisms including human), which is shuttled out of mitochondria as citrate (formed by Ac-CoA reacting with oxaloacetate, and can be converted back to Ac-CoA in the cytosol by citrate lyase consuming ATP). The overall biosynthesis is catalyzed by fatty acid synthase in a step-wise fashion, where the carbon backbone is elongated two carbons a time, similar to FAO. However, each step involves different reactions from those in FAO involving the adding of carbons from the three-carbon intermediate malonyl-CoA while consuming NADPH. The malonyl-CoA required for each step is in turn derived from Ac-CoA (via reacting with bicarbonate catalyzed by Ac-CoA carboxylase, requiring ATP). Fatty acids with longer and more complex chains are synthesized from those with simple and shorter chains in the ER, with the synthesis of unsaturated fatty acids requiring a mixed-function oxidase called fatty acyl-CoA desaturase. Some poly-unsaturated

fatty acids cannot be synthesized by mammals but are required precursors for other important products (e.g. eicosanoids), and thus need to be obtained from diet and are termed essential fatty acids. Triacylglycerols can be synthesized by conjugating glycerol 3-phosphate (derived from glycerol or from glycolysis) with acyl-CoAs (derived from fatty acids via the catalysis of acyl-CoA synthetase) through multiple steps. The metabolism of other types of lipids involve distinct pathways. For example, cholesterol can be synthesized from Ac-CoA in the mevalonate pathway (via various important metabolites including HMG-CoA, mevalonate, and squalene) ; cholesterol is the precursor for the further synthesis of many important steroid hormones. Ketone bodies refer to three metabolites including acetone, acetoacetate and D –  $\beta$ -hydroxybutyrate. These are derived from Ac-CoA in the liver and can be exported to other organs where they are converted back to Ac-CoA to be used as fuel, especially during starvation where there is increased liver FAO activity to generate Ac-CoA or in diabetes where there is deficiency in Ac-CoA oxidation via TCA due to TCA intermediates being diverted for gluconeogenesis.

Human cells contain 20 common types of amino acids (with additional amino acid derivatives). About 8 of the amino acids (methionine, valine, lysine, leucine, isoleucine, phenylalanine, tryptophan, threonine) cannot be synthesized by the human cell and need to be taken from diet, thus they are termed essential amino acids. Histidine is sometimes considered semi-essential as human body does not always require it. In mammals, tyrosine is derived from the essential amino acid phenylalanine. The rest are called non-essential amino acids. The catabolism of amino acids involves the disposition of the amino group as well as that of the carbon skeletons.

Different amino acids can often be interconverted via the process of transamination, where an amino acid and an  $\alpha$ -keto acid reacts under the catalysis of different transaminases to “swap” their carbon skeleton. Notably,  $\alpha$ -ketoglutarate can interconvert with glutamate, and oxaloacetate can interconvert with aspartate via this process. This allows the amino group in various amino acids to be transferred to glutamate or aspartate, which can then be disposed of via the urea cycle, while allowing their carbon skeletons to be catabolized as  $\alpha$ -keto acids, which ultimately enter the TCA for oxidation or be used for other biosynthesis. It needs to be noted that, however, not all amino acids can undergo transamination (exceptions include lysine, threonine and proline), and others like histidine, serine, phenylalanine and methionine are catabolized mainly via other unique pathways rather than via transamination. For urea cycle, which serves to dispose of the amino group, ammonium is first generated in the mitochondria from glutamate via oxidative deamination (requiring  $\text{NAD}^+$  or  $\text{NADP}^+$ ) or from glutamine via glutaminolysis. Ammonium joins  $\text{HCO}_3^-$  to form carbamoyl phosphate (requiring ATP), which enters the urea cycle by first reacting with ornithine to form citrulline. The rest of the urea cycle happens in the cytosol after citrulline is transported out of the mitochondria, reacting with aspartate to form arginino-succinate, which in turn converts to fumarate and arginine, arginine then breaks down to urea and ornithine, the latter being again imported into the mitochondria to complete the cycle. The urea generated is excreted in human (other species can excrete end products of different forms, e.g. birds excretes the amino group as uric acid). The urea cycle is closely coupled to the TCA cycle via the aspartate-arginino-succinate shunt, where the cytosol aspartate can be derived

from oxaloacetate from the TCA cycle which is then exported from the mitochondria, while the fumarate from arginino-succinate breakdown can be converted to malate which then enters the mitochondria to rejoin the TCA cycle. As described above, the TCA cycle is important for the metabolism of the carbon skeleton of amino acids, thus the coupling of TCA cycle and the urea cycle serves to jointly metabolize different chemical parts of amino acids. This leads me to the discussion of the catabolism of amino acid carbon skeletons, which ultimately are mostly either converted to pyruvate or Ac-CoA or enter the TCA cycle at different points. The amino acids that can be catabolized into Ac-CoA (threonine, lysine, leucine, isoleucine, tyrosine, tryptophan, phenylalanine) are termed ketogenic amino acids, as they can then be converted into ketone bodies in the liver. Others that can be converted to the other intermediates are called glucogenic amino acids, as they can be used for gluconeogenesis (some amino acids are both ketogenic and glucogenic, including isoleucine, phenylalanine, tyrosine, tryptophan, and threonine). Specifically, glycine, alanine, serine, threonine, tryptophan and cysteine can be catabolized to pyruvate; proline, glutamate, glutamine, arginine and histidine can be catabolized to  $\alpha$ -ketoglutarate; isoleucine, methionine, threonine and valine can be catabolized into succinyl-CoA; aspartate and asparagine can be catabolized to oxaloacetate; tyrosine and phenylalanine can be catabolized to fumarate. In human and in other species that are capable of synthesizing all amino acids, diverse pathways centered around glycolysis and the TCA cycle are involved in amino acid biosynthesis: serine is derived from 3-phospho-glycerate in the glycolysis pathway, which can in turn generate glycine and cysteine; alanine, valine and leucine are derived from pyruvate;

aromatic amino acids including phenylalanine, tyrosine, tryptophan and histidine are derived via the PPP. Oxaloacetate from the TCA cycle can generate aspartate, which in turn can lead to asparagine, lysine, threonine, isoleucine and methionine.  $\alpha$ -ketoglutarate from the TCA cycle can generate glutamate, which in turn generates glutamine, proline and arginine (in the urea cycle). Amino acids are used for protein synthesis in the ribosome after conjugated with tRNAs into aminoacyl-tRNAs. Conversely, cellular proteins are under constant turn-over and can be degraded in the proteasome (or lysosome) into amino acids, which can then be reused.

Nucleotides can be synthesized either *de novo* or via salvage pathways which recycle the free bases and nucleosides from nucleic acid breakdown. The *de novo* synthesis of purines involves the production of inosine 5'-monophosphate (IMP) as the first precursor with a purine ring. This process of purine ring building starts from phosphoribosyl pyrophosphate (PRPP, derived from ribose 5-phosphate from the PPP pathway), and requires glycine and two one-carbon unit donated from tetrahydrofolate (THF) for carbons and nitrogens donated from two glutamines and one aspartate. IMP can then be converted to AMP (requiring GTP) and GMP (requiring ATP), which are in turn used to synthesize A/GDP, A/GTP and the corresponding deoxyribonucleotides. The *de novo* synthesis of pyrimidines also requires PRPP leading to the formation of uridine 5'-monophosphate (UMP) as the first pyrimidine-ring containing precursor. The pyrimidine ring is built from aspartate which contributes both carbons and nitrogen atoms, glutamine which contributes another nitrogen atom, and  $\text{HCO}_3^-$  which contributes another carbon atom. UMP is in turn used to synthesize UDP and UTP. Various cytosine-containing nu-

cleotides are synthesized starting from CTP converted from UTP, whereas various thymine-containing deoxyribonucleotides are derived from dTMP, converted from dUMP. In general, the conversion of NMPs to NDPs are catalyzed by various nucleoside monophosphate kinases generally specific for a certain base but not the sugar (i.e. ribose or deoxyribose); the conversion of NDPs to NTPs are catalyzed by various nucleoside diphosphate kinases that are generally not specific for either base or the sugar. The generation of deoxyribonucleotides generally happens on the diphosphate level, converting various NDP into dNDP by ribonucleotide reductases. In the salvage pathways, free bases from the breakdown of nucleic acids can react with PRPP to form the corresponding NMP. NTPs are used to synthesize RNA e.g. during transcription via RNA polymerases, while dNTPs are used to synthesize DNA e.g. during DNA replication via DNA polymerases. The nucleotide *de novo* synthesis pathway is shown in Figure 1.4 (taken from (Koolman & Roehm, 2011)).

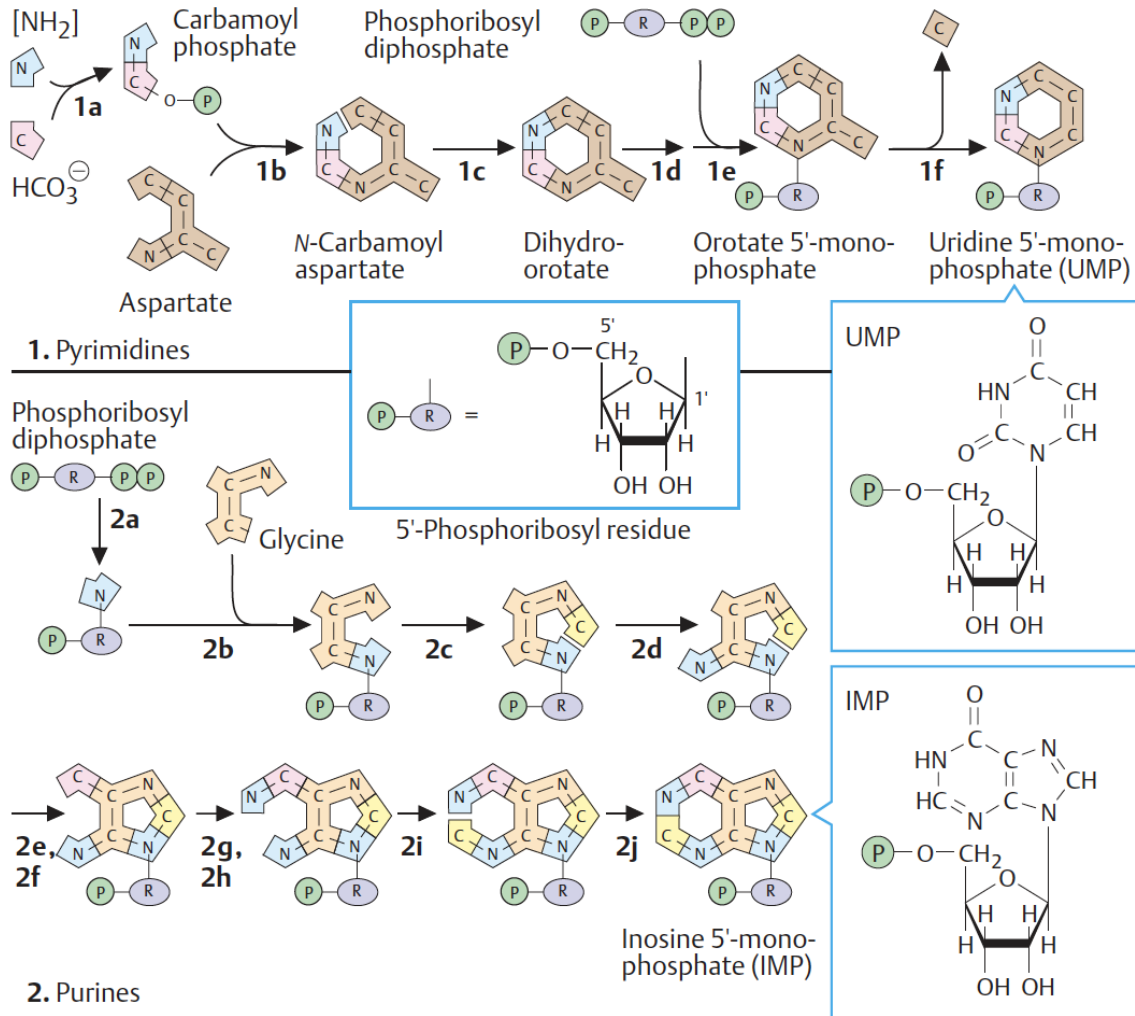


Figure 1.4: A diagram the nucleotide de novo synthesis pathway. This figure is taken from the book Color Atlas of Biochemistry.

I conclude my brief overview of human cell metabolic pathways here, with the note that this is not a complete introduction, with many important aspects of cell metabolism omitted, including but not limited to the metabolism and function of different vitamins, coenzymes/cofactors, inorganic metabolites like different mineral ions, reactive oxygen species (ROS), reactive nitrogen species, and cellular antioxidant, and the details of the metabolisms of less common types of carbohydrates, lipids, amino acid derivatives and nucleotide/base derivatives. Nevertheless, there

is no doubt that cell metabolism underlies all other aspects of cellular activities, it not only supports these other diverse processes but also involves in intricate interactions of double-way regulations with these other domains of cell biology, allowing the optimal response and normal function of cells under a wide variety of conditions.

## 1.2 Genome-scale metabolic modeling<sup>2</sup>

### 1.2.1 Genome-scale metabolic models (GEMs)

GEMs are mathematical network models that contain the information on all cellular metabolic reactions, the enzymes catalyzing the reactions, and the metabolites involved in the reactions and their stoichiometric numbers, allowing the *in silico* modeling and simulation of cell metabolism in a holistic manner using mathematical programming techniques. At its core, GEMs represent the interconnected network of  $m$  cellular metabolites and  $n$  reactions responsible for their transformation and transportation in a single matrix called the stoichiometric matrix  $\mathbf{S}_{(m \times n)}$ , where each row corresponds to a metabolite and each column corresponds to a reaction; an element in the matrix  $\mathbf{S}_{ji}$  is the stoichiometric number of the  $j$ 'th metabolite in the  $i$ 'th reaction (with positive values for products and negative values for reactants). GEM distinguishes different compartments of the cell, including for example the nucleus, mitochondria, endoplasmic reticulum (ER), Golgi apparatus, peroxisome, and extracellular space, among others (depending on the type of species); the same chemical species within different cellular compartments is represented by different

---

<sup>2</sup>a common reference for many contents throughout this section is the textbook Palsson (2015)

metabolites in the model, which are linked by exchange reactions when applicable. The metabolic reactions each has specific limits to their reaction rates (called “fluxes”, often denoted by  $\mathbf{v}$ ) due to thermodynamic and kinetic constraints, given as a vector of lower bound values  $\mathbf{v}^{lb}$  and a vector of upper bound values  $\mathbf{v}^{ub}$  in the model. Non-reversible reactions have flux lower bounds of zeros, whereas reversible reactions have negative flux lower bounds, with the negative flux value representing the reaction occurring in the reversed direction. A GEM also includes several special types of “artificial” reactions: Boundary reactions are at the system boundary representing the uptake/secretion of metabolites into/out of the entire metabolic system being modeled, for example a reaction conceptually representing metabolite exchange between the extracellular space and the environment. Sink reaction is a specific type of reversible boundary reaction added to the model to allow certain metabolites to be directly supplied into the metabolic network at certain points (e.g. can be used when the actual reaction producing the metabolite is unknown; it also allows for metabolites being diverted away from the system being modeled). Another specific type of boundary reaction is the demand reaction (or drain reaction), which irreversibly directs metabolites away from the system, conceptually representing the consumption of metabolites for other biologically processes that are not being explicitly modeled. Biomass reaction is a single demand reaction diverting a large range of essential metabolites such as ATP, different amino acids, nucleotides, lipids and carbohydrates away from the modeled system in appropriate ratios corresponding to the specific chemical composition and functionality of the cell being modeled. Conceptually the biomass reaction represents the total metabolic demand

for large molecule biosynthesis, maintenance, growth and proliferation of the cell being modeled. Relationships between the genes/enzymes and the reactions, called gene-protein-reaction associations (GPR), are represented by boolean expressions. For example, for a reaction that can be catalyzed by either one of three isozymes A, B and C, the relation is expressed as  $A|B|C$ , where as if all three genes/proteins D, E and F to form a functional holoenzyme to catalyze a reaction, then the relation is expressed as  $D\&E\&F$ . A GEM is specifically built for a certain species, or a specific strain (e.g. for unicellular organisms like *E. coli* or yeasts) or a specific tissue/cell type for a multicellular organism like human.

Traditionally, the construction of a GEM takes the so-called bottom-up approach, which involves the long and demanding process of putting together each cellular metabolic reaction, pathway and module in a step-by-step manner, requiring detailed knowledge on the metabolism of the target species or tissue, a large amount of different domains of supporting data and a wide spectrum of different expertises, usually necessitating the collaborative efforts of the research community. The process starts with a draft reconstruction, where all possible metabolic genes/proteins and functions harbored by the target organism/cell are enumerated and collected. This can involve the analysis of the annotated whole genome sequence if the species of interest to identify all open reading frames (ORFs) that possibly encode metabolic enzymes. The metabolic functions (reactions) of each gene are identified based on existing knowledge, and the GPR associations are established. Transcriptome and other omic data can also be used to utilized to refine the set of metabolic functions relevant to the specific system (e.g. cell/tissue type/organ) of

interest. The draft model is then subject to manual curation involving the survey of vast amount of literature to identify more accurate information on the metabolite specificity of the genes/reactions, the chemistry of the metabolites (most basically, their chemical formulae and electric charges), the stoichiometry of the reactions, the thermodynamic/kinetic properties of the reactions (including their reversibility and limits on reaction rates), and the cellular compartments of the metabolites and reactions. Information on the chemical composition, energy demand and growth phenotype of the cell/organism is also needed to create the biomass reaction. Such comprehensively compiled knowledge-base is converted into the computational GEM format. Naturally, the necessary information for some genes/metabolites/reactions can be missing, which then accordingly requires experimental studies, and/or theoretical computation/estimates, and the use of computational techniques such as gap-filling in order to produce a functioning model. Due to the prevalence of ambiguities due to incomplete knowledge, there is the practice of assigning confidence scores to different components (e.g. reactions, pathways, etc.) of the constructed model. The constructed model is subject to several evaluations and validations. For example, basic quality checks involve inspecting whether the model can generate all essential precursor metabolites, produce biomass on nutrients/substrates that the cell/organism is known to grow on, and maintain metabolic functions/processes known to the cell/organism of interest (e.g. the secretion of known metabolite products). Further comprehensive validation can be conducted by comparing the model predictions with available experimental data, including (but not restricted to) ATP production efficiency via oxidative phosphorylation (OXPHOS), O<sub>2</sub> consumption

and CO<sub>2</sub> production rates, and metabolic gene knockout (KO) phenotypes. The entire process of construction-curation-validation can be iterative to generate the final working model. Many different tools exist to facilitate or automate the particular steps of GEM construction, and many aim to achieve full automation of fast GEM reconstruction (Bekiaris & Klamt, 2020; Cuevas et al., 2016; Hamilton & Reed, 2014; Karlsten, Schulz, & Almaas, 2018). These methods involve the automated analysis and annotation of the genome of the species of interest, and takes advantage of databases such as MetaCyc (Caspi et al., 2010) and KEGG (Kanehisa, Goto, Sato, Furumichi, & Tanabe, 2012) for prior knowledge on metabolism. An incomplete list of examples include Model SEED (Henry et al., 2010), SuBliMinaL Toolbox (Swainston, Smallbone, Mendes, Kell, & Paton, 2011), merlin (Dias, Rocha, Ferreira, & Rocha, 2015), CarveMe (Machado, Andrejev, Tramontano, & Patil, 2018), with benchmark studies comparing the different methods available (Mendoza, Olivier, Molenaar, & Teusink, 2019). Softwares have been developed to comprehensively evaluate the quality of constructed models, notably the recent MEMOTE (Lieven et al., 2020).

Currently, GEMs are available for a wide range of organisms spanning prokaryotes, unicellular and multicellular eukaryotes. GEMs for many classical model species (e.g. the bacteria *E. coli*, the budding yeast *Saccharomyces cerevisiae*, and the house mouse *Mus musculus* (Khodaei, Asgari, Totonchi, & Karimi-Jafari, 2020; Sigurdsson, Jamshidi, Steingrimsson, Thiele, & Palsson, 2010)), pathogenic organisms (e.g. the gastric bacteria *Helicobacter pylori* and the parasite *Trypanosoma cruzi*), and microorganisms or cell types important for the fermentation/biotechnology

industry (e.g. the acetic acid bacteria *Acetobacter pasteurianu* and the Chinese hamster ovary, i.e. CHO cell) are available. Notably, there have been multiple versions of GEMs for human with several iterated updates (Ma et al., 2007; Recon 1, Duarte et al., 2007; Recon 2, Thiele et al., 2013; Recon 3D, Brunk et al., 2018; HMR1, Mardinoglu et al., 2013, HMR2, 2014; iHsa, Blais et al., 2017), with a most recent reconstruction incorporating several previous models (Human1, containing 13,417 reactions, 10,138 metabolites and 3,625 genes, Robinson et al., 2020) initiating a version-controlled and open-source model development framework based on a GitHub repository, which will potentially accelerate future research by providing real-time rolling updates to the model. Notable online databases hosting large ranges of GEMs include BiGG Models (King et al., 2016) and BioModels (Glont et al., 2018).

### 1.2.2 The methodological framework of GEM

Under the framework of systems biology modeling with GEM, the fluxes of all metabolic reactions often denoted by the vector  $v$  is the variable being modeled. Many different problems in GEM often involve the computation of optimal  $\mathbf{v}$ 's or other related objective functions subject to different types of constraints, thus termed constraint-based modeling (CBM). Together with GEM construction, this framework is also termed constraint-based reconstruction and analysis (COBRA). Different mathematical optimization techniques are used in CBM/COBRA, such as linear programming (LP), quadratic programming (QP), mixed integer program-

ming (MIP), and non-linear programming (NLP).

The most important set of constraints making the optimization problems solvable in CBM/COBRA is that imposed by the steady state assumption, which assumes that the metabolic system being modeled is in a steady state with no net change in the concentration of any metabolite. Essentially, this assumption requires that the total production flux of any one metabolite equals to its total consumption flux, thus there is no net accumulation or depletion of any metabolite within the system. Mathematically, this corresponds to the set of linear constraints expressed by  $\mathbf{S}\mathbf{v} = \mathbf{0}$ , also termed flux balance. Such an assumption is apparently seldom strictly satisfied, as it is rare, if not impossible to find a real-life biological system to be strictly under steady state. However, the steady state assumption is a quite reasonable one under biological homeostasis and many experimental conditions where the cells or system under study is in a stable biological state over a reasonably long duration of time. Besides, many biological systems can be regarded as being in quasi-steady states, which means that a system with relatively fast kinetics can be practically in steady states although its environment is changing with dynamic inputs and outputs between the system and the environment. The steady state assumption and flux balance are central and highly useful in CBM/COBRA, as without this assumption, the system would have to be modeled by kinetic modeling involving the kinetic constants of the reactions, which in many cases are unknown, rendering the system impractical to analyze. In practise, flux balance is usually assumed for all metabolites in the model, although if certain metabolites are known to not satisfy the assumption and their fluxes of net accumulation/consumption are

known, the flux balance constraints for these metabolites can be replaced by the corresponding constraints on flux production or consumption as appropriate.

The flux balance constraints restricts the viable solution space of a model comprising  $n$  reactions from  $\mathbb{R}^n$  to a subspace of  $\mathbb{R}^n$ , or specifically a hyperplane. Imposing the lower and upper bounds of the reaction fluxes further shrinks the space, i.e. one defined as:

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

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

When the reactions in the model are defined in a way that all flux values are positive (i.e. each reversible reaction  $i$  is expressed as two “half reactions” whose flux lower bounds are zero and upper bounds are  $|v_i^{\text{lb}}|$  and  $|v_i^{\text{ub}}|$ , respectively), then this solution space is a bounded subset of a convex cone. Additional constraints based on reaction kinetics and thermodynamics, metabolite concentrations, gene and protein expression, etc. can be formulated and help to further shrink the solution space, although they cannot shrink the space to a single point, namely, a single unique solution generally cannot be found only by imposing the various constraints. Therefore, many different optimization algorithms are applied to obtain a unique or at least more restricted set of solutions for detailed analysis, depending on different biologically informed assumptions, and additional techniques including sampling can be used to analyze a non-uniquely defined space of solutions.

One most basic and important optimization method in GEM to obtain unique or a more restricted set of solutions is the flux balance analysis (FBA), where cer-

tain cellular objective function is maximized (or minimized). A cellular objective function in FBA  $f_{obj}(\mathbf{v})$  is a single scalar-valued linear function of the reaction flux vector  $\mathbf{v}$  representing a certain biological process or function, a common example being the biomass reaction – In many cases, for example cultured bacteria or other actively proliferating cells in optimal (rich) growth media, it is reasonable to assume that the system being modeled is under optimal growth, represented by maximal biomass production. Maximizing the biomass production is formulated as the linear programming (LP) problem below:

$$\text{maximize } f_{obj}(\mathbf{v}) = v_{\text{biomass}}$$

subject to:

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

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

Similarly, other cellular objective functions can be specified as appropriate depending on the prior knowledge of the cell type or organism under study. FBA can also be used to study the capacities of various biological functions under different conditions by specifying the corresponding cellular objective functions, e.g. co-factors (e.g. NADH or NADPH), biosynthetic precursors or ATP production. Different biological conditions can be specified with additional constraints, for example, nutrient/metabolite availability can be specified by the flux bounds of their boundary reactions, and metabolic gene/reaction knockdowns or knockouts can be represented by decreased or zero flux bound values of the corresponding reaction.

In all cases, since both the objective function and the sets of constraints are linear, the problems can be formulated as LP problems which can be effectively solved by well-known algorithms such as the simplex algorithm.

Although the bounded LP problem in FBA is guaranteed to identify the optimal objective value, the optimal solutions  $\mathbf{v}^{\text{opt}}$  that can yield the optimal objective value may be non-unique, in which case they are termed degenerate solutions. Such cases are very common when modeling a large genome-scale model. Many techniques exist for further analysis of the optimal solution(s) of FBA and the space of degenerate solutions. For example, the sensitivity of the optimal solution can be analyzed by shadow prices and reduced costs. The shadow price of a boundary reaction is the incremental change in the optimal objective function value given the incremental flux change of the boundary reaction (i.e. the derivative of the former with respect to the latter), i.e. it measures whether or how the optimal objective value is affected by the availability of a certain metabolite. Similarly, the reduced cost is defined for an internal reaction (i.e. not boundary) and measures whether or how the change of an internal flux affects the optimal objective value. The presence of metabolites/reactions with zero shadow price/reduced cost implies the existence of degenerative solutions. These analyses of the effect of varying one parameter (the flux of the  $I$ 'th reaction) on the optimal objective value, called robustness analysis, are conducted via solving a series of LP problems with different fixed  $\theta$  values:

maximize  $f_{obj}(\mathbf{v})$

subject to:

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

$$v_I = v_{I,min} + \theta(v_{I,max} - v_{I,min}), \theta \in [0, 1]$$

$$v_i^{lb} \leq v_i \leq v_i^{ub}, i = 1, \dots, n, i \neq I$$

Similarly, the effect of varying two parameters on the optimal objective function value can be analyzed, which produces a 2D surface called phenotypic phase plane. Another more direct way of analyzing degenerate solutions is the flux variability analysis (FVA), which computes the range of possible fluxes of each reaction while fixing the cellular objective function at its optimal value  $f_{obj}^{opt}$ . FVA is formulated as follows:

maximize and minimize  $v_i$ , for each  $i = 1, \dots, n$

subject to:

$$f_{obj}(\mathbf{v}) = f_{obj}^{opt}$$

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

$$\mathbf{v}^{lb} \leq \mathbf{v} \leq \mathbf{v}^{ub}$$

Besides, the space of degenerate solutions, or any convex space of solutions in LP problems can in theory be characterized using uniform random sampling, i.e. sampling of a statistically sufficient number of solutions that lie uniformly within the solution space. These sample points reflecting the flux distribution under the

optimal objective value can then be analyzed and visualized to study the variability in the metabolic state under the optimal cellular objective function. Sampling a high-dimensional space, however, can be a challenging problem, especially that the solution space of GEM can often have irregular and elongated shape. The sampling of large GEMs can also take a very long time to converge. A well-known sampling algorithm suitable for the sampling of high-dimensional spaces with irregular shapes is called artificial centering hit-and-run (ACHR), and many variations of the algorithm have been developed (Kaufman & Smith, 1998; Megchelenbrink, Huynen, & Marchiori, 2014; Saa & Nielsen, 2016). Coordinate hit-and-run with rounding (CHRR) is another efficient algorithm shown to have fast convergence time (Fallahi, Skaug, & Alendal, 2020; Haraldsdóttir, Cousins, Thiele, Fleming, & Vempala, 2017). Rather than sampling, a recent algorithm based on expectation propagation was shown to accurately approximate the marginal distribution of high-dimensional polytopes with high computation efficiency (Braunstein, Muntoni, & Pagnani, 2017). These algorithms are valuable for more detailed analyses of the convex solution spaces arising from the CBM/COBRA of GEMs.

### 1.2.3 GEM algorithms for omic data integration

GEM modeling has proven invaluable for the *in silico* simulation of the metabolism of many different organisms, producing results with good concordance with experimental data. For unicellular organisms or cells of multicellular organisms under culture, the alteration in cell metabolism under different culture conditions can be

readily modeled, provided that a GEM tailored for the cell type under study has been constructed. However, for complex multicellular organisms like human, a large number of different tissue and cell types exist with highly diverse and distinct metabolic functions, in theory requiring a customized GEM for each. Furthermore, for the modeling of *in vivo* metabolism, the cells of different organs and tissues are within different metabolic environments that are not pre-defined. Variabilities across conditions or individuals due to other factors only indirectly related to cell metabolism are also not as straightforward to model as, say, changes in metabolite availability in the culture media or direct perturbation of metabolic genes/enzymes. Fortunately, the advent and fast progress made in different omic profiling technologies has allowed the fine-grained and high-throughput characterization of cellular molecular landscapes in any biological samples under any experimental conditions. Whole-transcriptome gene expression data from microarrays or various RNA-sequencing technologies, high-throughput protein expression data from mass spectrometry-based quantitative proteomics, and metabolomics data that quantifies the level of large numbers of metabolites in a sample are now widely available. The integration of such omic data into GEM has truly unleashed the power of GEM modeling for the study of cell metabolism in a greater range of problems. The various omic data which provides detailed high-coverage molecular profiles is often analyzed with statistical modeling techniques, which can reveal complex patterns in the data but often cannot easily establish any causal relationships. In contrast, the CBM/COBRA framework of GEM modeling, due to its capacity to perform mechanism-based *in silico* simulations, can be used to investigate causality. Thus, the integration of GEM with omic

data provides a valuable opportunity to deepen the understanding of the molecular system under study, specifically within the domain of metabolism.

Next I will focus on the discussion of integration of gene (or protein) expression data into GEM. A large number of different algorithms exist for this important problem (Jamialahmadi, Hashemi-Najafabadi, Motamedian, Romeo, & Bagheri, 2019; Kim & Lun, 2014; Opdam et al., 2017). The common idea underlying these algorithms is to either select a subset of high-confidence reactions from the general GEM (“base model”) of an organism that are supported by the additional omic data, and/or to adjust the flux bounds of the reactions or directly fit the flux values based on the notion of a rough correlation between gene/protein expression levels and the activity levels (i.e. fluxes) of the corresponding reactions, resulting in refined GEMs and/or optimal flux distributions specific to the sample or context that correspond to the external omic data. Some of the algorithms of the first type include MBA (Jerby, Shlomi, & Ruppin, 2010), mCADRE (Wang, Eddy, & Price, 2012), FASTCORE (Vlassis, Pacheco, & Sauter, 2014), FASTCORMICS (Pacheco et al., 2015), and CORDA (Schultz & Qutub, 2016), which involves the identification of “core” sets of reactions supported by external data including expression data, supplemented usually in a parsimonious fashion by additional reactions necessary to yield a consistent model without “dead-end” reactions that cannot bear fluxes. These algorithms generate customized GEMs that correspond to the external data, but do not directly compute an optimal flux distribution. Most other algorithms are of the second type, which seek to maximize the concordance of flux and gene expression in various ways. GIMME (Becker & Palsson, 2008), E-Flux (Colijn et al.,

2009), Lee et al. (Lee et al., 2012), GIM<sup>3</sup>E (Schmidt et al., 2013), E-Fmin (Song, Reifman, & Wallqvist, 2014), and TRFBA (Motamedian, Mohammadi, Shojaosadati, & Heydari, 2017) each applies different LP-based formulations to maximize the concordance between the fluxes and gene expression values (after converted to the reaction level based on the GPR boolean mapping). Also based on this idea but instead of using LP-based formulations, iMAT (Shlomi, Cabili, Herrgård, Palsson, & Rupp, 2008), INIT (Agren et al., 2012) and tINIT (Agren et al., 2014) use MILP-based formulations to maximize the number of reactions with high or low fluxes that are consistent with the corresponding high or low gene expression levels (defined based on pre-defined thresholds as parameters). RegrEx (Robaina Estévez & Nikoloski, 2015) is based on a MIQP formulation that minimizes the difference between flux and gene expression data with  $\ell_1$ -norm regularization. PRIME (Yizhak et al., 2014) requires the gene expression data of multiple samples together with the data on measured cell growth rates of the samples, which are used to identify a subset of reactions correlating with the measured cell growth. The flux bounds of these reactions are then directly adjusted according to the gene expression values. Many additional algorithms are available (e.g. Zhang et al., 2019), however, only a few attempts have been made to compare and benchmark the various algorithms in a comprehensive manner (Jamialahmadi et al., 2019; Opdam et al., 2017), and the available benchmark studies only cover a subset of the available methods. The benchmark studies mostly adopted technical evaluations (e.g. robustness to noisy data) and indirect methods (e.g. predicting known metabolic functionalities, gene essentiality, drug response, etc.) to validate the performance of the algorithms,

as large-scale data of direct flux measurements is very rare. Different algorithms can each involve numerous parameters whose varied values can affect the performance of the algorithms. For example, many algorithms requires a pre-determined gene expression level threshold to define high vs low expression, and the benchmark study of Opdam et al. that compared MBA, mCADRE, FASTCORE, iMAT, INIT and GIMME (Opdam et al., 2017) shows that in some cases, the selection of this threshold can contribute to more variation in the whether a reaction is retained in the resulting model even compared to the choice of algorithm; In gene essentiality prediction, however, the most variation appears to arise from the choice of algorithm, and using stricter gene expression thresholds as well as properly constraining the GEM based on culture media composition and exometabolomic data overall helped to improve the accuracy for gene essentiality prediction; mCADRE and MBA seemed to performed the best in the gene essentiality prediction, followed by INIT and iMAT (while these two showing high variation in their performances depending on dataset and parameters used); GIMME and FASTCORE performed worse than the other algorithms in this task. The constraining of the model based on metabolite uptake also significantly affected growth rate prediction. The study concluded that no single algorithm is a universal best performer and algorithms need to be chosen based on the available data and goal of the study (Opdam et al., 2017). The benchmark study by Jamialahmadi et al. (Jamialahmadi et al., 2019) compared mCADRE, FASTCORE, FASTCORMICS, CORDA, iMAT, INIT, PRIME, GIMME, and TRFBA based on technical assessments and a few tasks specifically based on cancer cell lines including predicting cell growth rate, metabo-

lites uptake/secretion rates, drug response and gene essentiality, and cancer driver gene prediction. This study confirms the findings in the Opdam et al. study that the algorithm performance depends on the specific task, and constraining metabolic uptake based on culture media helped to improve performance notably for growth rate prediction, although constraint models did not improve essential gene prediction much except for TRFBA. CORDA and iMAT was found to perform worse than the other algorithms in predicting growth rate. PRIME was the only algorithm found to be able to produce significant predictions on metabolite uptake/secretion rates without constraining the model based on the media, but once with proper constraining, other algorithms including GIMME, TRFBA and simply parsimonious FBA (pFBA, without expression data integration) can produce better performances. PRIME and TRFBA appeared to perform the best on drug sensitivity and gene essentiality prediction, however, both algorithms takes advantage of data on measured growth rates. In gene essentiality prediction, GIMME also had reasonable performance, but mCADRE performed worse, which is contradictory to the Opdam et al. results potentially due to the variation in the dataset and detailed algorithm parameters used, highlighting the limitations of current benchmark studies. On the other hand, FASTCORE, FASTCORMICS, INIT, and iMAT had good performances in various technical assessments including robustness to noise, lack of blocked reactions and similarity in the obtained model given similar input data. The authors further proposed a new algorithm combining TRFBA and FASTCORMICS which they called TRFBA-CORE, which does not require measured growth rate data, and showed it to have overall the most balanced performance across multiple tasks and technical

evaluations. Despite this, the authors highlighted the need for further more robust benchmarks and the choice of method based on the problem under study. The comparison of the performances of the algorithms using different tasks (according to (Jamialahmadi et al., 2019)) is summarized in Table 1.1

Table 1.1: Qualitative summary of relative performances of different GEM algorithms for integrating gene expression data, in different prediction tasks or quality control assessments. The two ratings (before and after slashes) correspond to models generated with general medium and cell-specific medium, respectively. med: medium.

Algorithms	Metabolite				
	Growth Rate	Exchange Rate	Gene Essentiality	Lack of Blocked Reactions	Robustness
GIMME	high/high	low/high	low/med	low/low	low/low
iMAT	med/med	low/med	low/low	high/high	high/high
INIT	low/low	low/low	low/low	high/high	high/high
mCADRE	med/med	low/low	low/low	high/high	low/low
FASTCORE	low/low	low/low	low/low	low/low	med/med
FASTCORE-MCS	low/low	low/low	low/low	high/high	high/high
CORDA	med/high	low/low	low/med	high/high	low/low
PRIME	high/high	med/low	high/high	high/high	med/med
TRFBA	high/high	low/high	high/high	high/high	high/high

As one example of the many gene expression data integration algorithms, and also due to the specific relevance to the work in this thesis, I next describe the

Integrative Metabolic Analysis Tool (iMAT) algorithm (Shlomi et al., 2008) in more detail. iMAT takes the transcriptome-wide gene expression data of a sample in a vector, together with the “base” GEM of the corresponding organism as input; it outputs a GEM customized to the given sample, with reaction flux bounds further constrained to fit that of the expression profile. The input vector contains the within-sample-normalized expression values of all genes in a sample (i.e. the values should ideally be comparable between any two genes), two threshold parameters are used to divide all genes into three sets: high-expression, medium-expression, and low-expression. Then the gene expression vector is mapped to a discretized vector of  $\{1, 0, -1\}$ , with high, medium, and low-expression genes mapped to 1, 0, and -1, respectively. This vector of discretized gene expression values is then converted to a vector of expected reaction activity levels, based on the boolean GPR association, but replacing the logical  $\&$  and  $|$  operators with *max* and *min* functions, respectively. The resulting reactions with values of 1 are expected to have high activity, denoted  $R_H$ , and those with values of -1 are expected to have low activity, denoted  $R_L$ . Only  $R_H$  and  $R_L$  are used in the subsequent optimization procedure, and reactions with values of 0 are not used. The subsequent optimization seeks to maximize the number of reactions bearing consistent flux levels with those expected from the expression data, while subject to the usual constraints of the metabolic network (i.e. flux balance and the flux bounds). This is formulated as an MILP problem by introducing a series of integer (actually binary) variables  $y_i^+$ 's and  $y_i^-$ 's, for each reaction  $i \in R_H$  one  $y_i^+$  and one  $y_i^-$  is introduced, and for each reaction  $i \in R_L$  only one  $y_i^+$  is introduced. The formulation is such that for a reaction  $i \in R_H$ ,  $y_i^+ = 1$

corresponds to that it has large positive flux greater than the positive parameter  $\epsilon_{hi}$  (i.e. high activity level in the positive direction),  $y_i^- = 1$  corresponds to large negative flux (smaller than  $-\epsilon_{hi}$ ), and  $y_i^+ = y_i^- = 0$  corresponds to the default flux bound constraint in the “base” model; obviously  $y_i^+$  and  $y_i^-$  cannot both be 1 for  $i \in R_H$ ; for a reaction  $i \in R_L$ ,  $y_i^+ = 1$  corresponds to that it has small flux value near zero (absolute flux value smaller than the positive parameter  $\epsilon_{lo}$ ), and  $y_i^+ = 0$  corresponds to the default flux bound constraint:

$$\text{maximize}_{\mathbf{v}, \mathbf{y}^+, \mathbf{y}^-} \left( \sum_{i \in R_H} (y_i^+ + y_i^-) + \sum_{i \in R_L} y_i^+ \right)$$

subject to:

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

$$\mathbf{v}^{lb} \leq \mathbf{v} \leq \mathbf{v}^{ub}$$

$$v_i + y_i^+(v_i^{lb} - \epsilon_{hi}) \geq v_i^{lb}, \text{ for } i \in R_H$$

$$v_i + y_i^-(v_i^{ub} + \epsilon_{hi}) \leq v_i^{ub}, \text{ for } i \in R_H$$

$$v_i + y_i^+(v_i^{lb} + \epsilon_{lo}) \geq v_i^{lb}, \text{ for } i \in R_L$$

$$v_i + y_i^+(v_i^{ub} - \epsilon_{lo}) \leq v_i^{ub}, \text{ for } i \in R_L$$

$$y_i^+, y_i^- \in \{0, 1\}$$

The original formulation of iMAT as in (Shlomi et al., 2008) set the parameter  $\epsilon_{hi}$  to 1 and does not have the  $\epsilon_{lo}$  parameter, and is equivalent to fixing  $\epsilon_{lo}$  at zero. After solving the MILP problem,  $v_i^{lb}$ ,  $i \in R_H \wedge y_i^{+,opt} = 1$  are set to  $\epsilon_{hi}$ ,  $v_i^{ub}$ ,  $i \in R_H \wedge y_i^{-,opt} = 1$  are set to  $-\epsilon_{hi}$ , and  $v_i^{ub}$ ,  $i \in R_L \wedge y_i^{+,opt} = 1$  are set to  $\epsilon_{lo}$ ,  $v_i^{lb}$ ,  $i \in R_L \wedge y_i^{+,opt} = 1$  are set to  $\max(-\epsilon_{lo}, v_i^{lb})$ . To address the issue of

possibly multiple optimal solutions, The above MILP problem is run twice for each reaction, one time forcing the reaction to be activated (having flux greater than  $\epsilon_{hi}$ ) and the second time forcing it to be inactivated (having flux lesser than  $\epsilon_{lo}$ ), the resulting optimal objective function values of the two runs,  $f_i^{obj,act}$  and  $f_i^{obj,inact}$  are compared, and if  $f_i^{obj,act} > f_i^{obj,inact}$ , reaction  $i$  is determined to be activated, if  $f_i^{obj,act} < f_i^{obj,inact}$ , reaction  $i$  is determined to be inactivated, if  $f_i^{obj,act} = f_i^{obj,inact}$ , the activity level of reaction  $i$  is undetermined and the corresponding flux bounds are not changed (left as default as in the base model). Alternatively, analogous to the FVA procedure, after solving the base MILP, an additional linear constraint can be added forcing the objective function at its optimal value, then a series of MILPs can be solved to determine the maximal and minimal fluxes of each reaction.

#### 1.2.4 Additional GEM algorithms relevant to this thesis

Many GEM algorithms have been developed for various specific purposes. In this section I describe a few additional algorithms that are relevant to the work in this dissertation.

Minimization of Metabolic Adjustment (MOMA) (Segrè, Vitkup, & Church, 2002) is an algorithm to predict the flux distribution after a metabolic perturbation, i.e. a metabolic reaction knockout (KO). Although this prediction can be also achieved with FBA after simulating the KO by constraining the corresponding reaction flux to zero (or a very low value), MOMA is based on the idea that after the KO, the preference of the system in the short term is to make only minimal ad-

adjustments and shift to a nearest point in the new feasible metabolic space. MOMA adopts the Euclidean distance measure to determine the “nearest point”, leading to a QP problem formulated as below (where  $\mathbf{v}^{\text{opt,wt}}$  is the optimal solution obtained from FBA on the wild-type (WT) model without any KO, and  $v_i = 0$  simulates the KO of the  $i$ 'th reaction):

$$\underset{\mathbf{v}}{\text{minimize}} \|\mathbf{v} - \mathbf{v}^{\text{opt,wt}}\|_2^2$$

subject to:

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

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

$$v_i = 0$$

However, to address the issue of possible degenerate solutions in the FBA, an alternative QP formulation is given below, where the WT model and the KO model are optimized together ( $f_{obj}^{\text{opt,wt}}$  is the optimal cellular objective function value obtained from FBA in the WT model):

$$\underset{\mathbf{v}^{\text{wt}}, \mathbf{v}^{\text{ko}}}{\text{minimize}} \|\mathbf{v}^{\text{wt}} - \mathbf{v}^{\text{ko}}\|_2^2$$

subject to:

$$f_{obj}(\mathbf{v}^{\text{wt}}) = f_{obj}^{opt, wt}$$

$$\mathbf{S}\mathbf{v}^{\text{wt}} = \mathbf{0}$$

$$\mathbf{S}\mathbf{v}^{\text{ko}} = \mathbf{0}$$

$$\mathbf{v}^{\text{lb, wt}} \leq \mathbf{v}^{\text{wt}} \leq \mathbf{v}^{\text{ub, wt}}$$

$$\mathbf{v}^{\text{lb, ko}} \leq \mathbf{v}^{\text{ko}} \leq \mathbf{v}^{\text{ub, ko}}$$

$$v_i^{\text{ko}} = 0$$

Metabolic Transformation Algorithm (MTA) (Yizhak, Gabay, Cohen, & Ruppin, 2013) is an algorithm for the prediction of metabolic targets whose KO can transform the cellular metabolic state from a given reference state to a given target state. For example, MTA can be used to predict therapeutic targets for reversing the disrupted cell metabolism in a metabolism-related disease. The input to MTA is the transcriptome-wide gene expression data of the reference state, in addition to the differential gene expression (DE) changes in the target state compared to the reference state, provided as two vectors; MTA outputs a score called the MTA score for each reaction, with higher MTA scores corresponding to better targets, i.e. whose KO can better transform the cell metabolism from the reference to the target state. I will describe the algorithm details of MTA in Chapter 2.

## 1.3 Metabolism of T cells in cancer and cancer immunotherapy

### 1.3.1 Overview of cancer immunology

Cancer is a group of complex diseases commonly exhibiting phenotypes of malicious cell growth and invasion, characterized and enabled by multiple hallmarks including replicative immortality, sustained proliferation signals, evasion of growth suppression signal, resistance to cell death, angiogenesis, invasion and metastasis, mutation and genome instability, cell energetics dysregulation, inflammation, and evasion of the immune system (Hanahan & Weinberg, 2011). Although evidences for the involvement of the immune system in cancer have been found from early on, it is not until recent years that the role of immune system in the development of cancer was evidently established, the detailed mechanisms of the interaction between cancers and the immune system were identified, and the principle of cancer immunology being successfully exploited for the development of novel effective anti-cancer therapies. As early as in 1863, Rudolf Virchow observed the presence of leukocytes within tumors and suggested that cancers arise at sites of chronic inflammation (Balkwill & Mantovani, 2001). Other notable reports include that in the late 19th century, William Bradley Coley notice tumor regression in a patient after *Streptococcus* infection, and hypothesized that the immune system can act to destroy tumors (Coley, 1893). He even designed and tested the now famous “Coley’s Toxins”, a mixture of bacterial products on cancer patients over the next decades, which is regarded as the first cancer immunotherapy ever tested. The hypothe-

sis of cancer immunosurveillance, namely the immune system surveys the body to eliminate any nascently transformed cells (Burnet, 1970), initially built upon many previous ideas and observations of the involvement of immune system in cancer, was supported and validated in more recent studies in mice (Dighe, Richards, Old, & Schreiber, 1994; Kaplan et al., 1998; Shankaran et al., 2001; Street, Cretney, & Smyth, 2001).

Currently, it has been established that many cancers can arise from mutations in their genomes, resulting in the expression of mutated proteins or abnormal expression of proteins that are mostly not or lowly expressed in normal adult tissues. These mutated or abnormal proteins can be degraded into short peptides and presented on the cell surface by the major histocompatibility complexes (MHC) I molecules, which can be detected by CD8 T cells that in turn activates to eliminate the mutated cells. These presented peptides arising from cancer mutations or cancer-related dysregulation in gene expressions that are recognized by CD8 T cells are called tumor antigens (Schumacher & Schreiber, 2015). Both adaptive immunity and innate immunity, with different immune cell types of the innate immunity, such as natural killer (NK) cells, have also been recognized to play important roles in cancer immunosurveillance (Waldhauer & Steinle, 2008). For example, the low and abnormal expression of MHC-I molecules by some cancer cells can lead to decreased inhibition to NK cells via its various receptors, resulting in NK cell-mediated cancer killing (Yokoyama & Kim, 2006). Studies on cancer immunosurveillance in turn also reveals that the immune system contributes to the development of tumors with reduced immunogenicity, which are ultimately able to evade the recognition and clear-

ance by the immune system, further bringing forward the closely related concept of cancer immunoediting (Dunn, Bruce, Ikeda, Old, & Schreiber, 2002; Shankaran et al., 2001). Cancer immunoediting has been conceptualized as a dynamic process involving three phases of the so-called three E's: elimination, equilibrium, and escape (Dunn, Old, & Schreiber, 2004). During elimination, i.e. the classical phase of immunosurveillance, the immune system is able to detect and specifically eliminate the emerging transformed cells via different recognition and effector mechanisms. This imposes a strong selection pressure on the cancer cells. During the equilibrium phase, some cancer cells have evolved to have reduced immunogenicity and avoided the complete destruction by the immune system, the immune system is no longer able to fully recognize and react to kill the cancer cells, reflected by a latency period. Finally during escape, the cancer cells are able to overcome the immunological constraints via multiple active mechanisms of immunosuppression, thereby achieving massive outgrowth.

Cancer cells can exploit multiple mechanisms for immunosuppression (Rabinovich, Gabrilovich, & Sotomayor, 2007). Cancers can have reduced/impaired antigen presentation, which contributes to the escape from T cell recognition. This can be due to lost expression of the antigen (Spiotto, Rowley, & Schreiber, 2004), and/or downregulation of or mutation in the antigen presentation machinery (Cornel, Mimpfen, & Nierkens, 2020). Although the downregulation of MHC-I molecules may render the cancer cells susceptible to NK cell-mediated killing, cancers have also evolved mechanisms to desensitize NK cell activation and avoid killing by NK cells (Deng et al., 2015). The cancers can also secrete immunosuppressive cytokines,

notably TGF- $\beta$  (Batlle & Massagué, 2019), which can contribute to the recruitment and promote the differentiation of immunomodulatory cell types such as regulatory T cells (Tregs) (Plitas & Rudensky, 2020) and tumor-associated macrophages (TAMs) (Mantovani, Marchesi, Malesci, Laghi, & Allavena, 2017) in the tumor microenvironment (TME), which has been regarded as immunosuppressive in general (Hinshaw & Shevde, 2019) also due to hypoxia (Ackerman & Simon, 2014), low pH (from the exported lactate generated in the tumor glycolysis, accompanied by proton) (Corbet & Feron, 2017), deficiency in essential metabolites for immune cell function (Lau & Vander Heiden, 2020) (e.g. arginine) (Geiger et al., 2016), increased ROS (Jin & Jin, 2020) and the abundance of immunosuppressive metabolites derived from tumor metabolism (e.g. adenosine (Mastelic-Gavillet et al., 2019) and kynurenine (Hornigold et al., 2020)). Importantly, cancer cells and other cell types within the TME can express and upregulate different ligands molecules that engage in inhibitory interactions with the receptors on different immune cells, called immune checkpoints. CTLA-4 represents one of the earliest studied checkpoint molecules that is mostly expressed on Treg cells, which competes with CD28 in binding the B7 costimulatory molecules on T cells with high affinity (Leach, Krummel, & Allison, 1996). PD-1 is another well-studied immune checkpoint molecule on the T cells, with its ligands including PD-L1 and PD-L2 expressed by cancer cells and other cells in the TME such as tumor-associated fibroblasts (Freeman et al., 2000; Sun, Mezzadra, & Schumacher, 2018). The identification of CTLA-4 and PD-1/PD-L1 and their role in cancer immunity led to the development of the first immune checkpoint blockade (ICB) therapies involving the targeting of these

molecules with monoclonal antibodies (Esfahani et al., 2020). Many other immune checkpoint molecules (and ligands) have been subsequently identified and under study, with the hope of the development of novel ICB therapies based on these targets (Qin et al., 2019), some examples include LAG-3, TIM-3, CD276/B7-H3, VTCN1/B7-H4, TIGIT, and VISTA (Marhelava, Pilch, Bajor, Graczyk-Jarzynka, & Zagodzón, 2019; Marin-Acevedo et al., 2018; Qin et al., 2019).

### 1.3.2 Cancer immunotherapy, immune checkpoint blockade therapy and factors associated with patient response

Although cancer immunotherapy broadly encompasses ICB, adoptive cell transfer therapy (ACT), antibody-based immunotherapies, treatment vaccines and immune modulators (NCI, 2015; Waldman, Fritz, & Lenardo, 2020), ICB with immune checkpoint inhibitors seem to have currently become one of the most active area of immunotherapy development and testing, thanks to the encouraging successes of the anti-CTLA-4 and anti-PD-1/PD-L1 ICB therapies which helped many patients to achieve persistent cancer remission unseen with previous therapies (Couzin-Frankel, 2013). Following the initial development, anti-PD-1/PD-L1 antibodies have been tested in a wide array of cancer types, resulting their approval by FDA in a total of 21 cancer types as of February 2021 (Institute, 2021). Despite long-lasting response seen in some patients, the fractions of patients responding to the therapies are overall suboptimal (around 20%) (Carretero-González et al., 2018), and many current developments have been seeking to address this issue via the identification

of biomarkers that predict response (Ganesan & Mehnert, 2020; Havel, Chowell, & Chan, 2019) and via the identification and evaluation of potential combinatory therapies with ICB (Meric-Bernstam, Larkin, Tabernero, & Bonini, 2020). Many factors have been identified to be correlated with response to ICB, including target (e.g. PD-L1) expression (Mok et al., 2019), tumor mutational burden (TMB) (Samstein et al., 2019), DNA repair deficiencies in the tumor like mismatch repair deficiency (Le et al., 2017), specific mutational signatures (Rizvi et al., 2015), neoantigen load (Van Allen et al., 2015), the presence of viral infection (Ganesan & Mehnert, 2020), MHC-I and II expression (Rodig et al., 2018), and the abundance and cytolytic function of tumor-infiltrating lymphocytes (Sade-Feldman et al., 2018; Van Allen et al., 2015). Additionally, various gene expression based signatures or predictors have been investigated for their performance in predicting response [Cristescu et al. (2018); Nirmal et al. (2018); Ott et al. (2019); Sahu et al. (2019); Auslander et al. (2018); and many others]. The gut microbiota composition has been implicated to affect response as well (Sivan et al., 2015). In terms of combinatory therapy with ICB, a wide range of different therapies, including chemotherapies, radiotherapies and targeted therapies are being actively tested, with 2900 clinical trials registered investigating 253 different potential targets as of February, 2021 (Institute, 2021). The investigation of the targeting of novel checkpoint molecules are also ongoing, with several registered clinical trials (Esfahani et al., 2020). Besides the inhibition of suppressive checkpoint molecules, molecules of co-stimulatory pathways, also called stimulatory checkpoints, are also under investigation for therapeutic strategies seeking to activate these molecules (Marhelava et al., 2019; Marin-Acevedo et al., 2018).

### 1.3.3 Adoptive cell transfer therapy

Adoptive cell transfer therapy (ACT), especially adoptive T cell transfer therapies including the T cell receptor-T cell (TCR-T) or chimeric antigen receptor-T cell (CAR-T) therapies represent another major promising development in cancer immunotherapy. This therapy involves the isolation of tumor-infiltrating lymphocytes (TILs) from tumor patients, followed by the identification of tumor-reacting lymphocytes and *ex vivo* stimulation and expansion, which are then infused back into the cancer patient to treat their tumors (Rosenberg, Restifo, Yang, Morgan, & Dudley, 2008). The efficacy of autologous TIL transfusion was first demonstrated in 1988 for the treatment of metastatic melanoma (Rosenberg et al., 1988). The treatment regimen was subsequently significantly improved, involving lympho-depletion by chemotherapy before TIL transfer (Dudley et al., 2002). The isolated TILs can also be genetically engineered allowing them to target specific tumor antigens. In TCR-T therapy, T cells reacting to tumor antigens with high avidity are identified and their TCR genes are cloned, which can then be transduced into the T cells of the patients to express the tumor antigen-reacting TCR. These engineered TCR-T cells can then be expanded and infused into the patients. The first clinical trial of TCR-T therapy was for melanoma based on a TCR reacting to the MART1 melanoma antigen, which successfully lead to regression in several metastatic melanoma patients with long-lasting response (Morgan et al., 2006). In CAR-T therapy, instead of a TCR gene, an artificially engineered chimeric antigen receptor (CAR) is transduced into the acceptor T cells. CAR comprises a single chain variable fragment

(scFv) proportion for the specific recognition of tumor antigen, which is connected to a transmembrane domain followed by intracellular domains for transducing the antigen-binding signal inside the cell for T cell activation. The intracellular domain contains a TCR CD3 $\zeta$  signaling domain, and usually multiple co-stimulating domains such as CD28 and 4-1BB (June & Sadelain, 2018). Unlike TCR-T cell, CAR-T cell does not have HLA restriction, although it can only recognize intact antigen present on the cell surface (while TCR-T cell recognized MHC-presented antigen, which can be derived from intracellular proteins) (Jafferji & Yang, 2019). CAR-T cell therapy was first investigated for targeting the B cell antigen CD19 to treat refractory chronic lymphocytic leukemia (CLL) and acute lymphoblastic leukemia (ALL), which achieved durable response and the cure of the first relapsed and refractory ALL patient (Brentjens et al., 2013; Grupp et al., 2013; Porter, Levine, Kalos, Bagg, & June, 2011). Subsequent development led to the FDA approval of anti-CD19 CAR-T products for relapsed and refractory B cell ALL and relapsed and refractory diffuse large B-cell lymphoma (DLBCL). In addition to CD19, CD20 and CD22 are being investigated as alternative or combinatory targets in CAR-T therapy for hematological malignancies (Styczyński, 2020). CAR-T therapy has so far had the greatest success in hematological malignancies, with high response rates in various leukemias and lymphomas (June & Sadelain, 2018). Both TCR-T and CAR-T therapy are under development for various solid tumors (Jiang et al., 2019), although clinical breakthroughs remain to be seen due to challenges including the selection of tumor antigen, and strategies for divert the CAR-T cells to the tumor site and infiltrate the tumor (Wagner, Wickman, DeRenzo, & Gottschalk, 2020). Ad-

ditional challenges for the further optimization of ACT therapy include addressing the potential mutation of the targeted antigen, reducing therapy-associated toxicity (esp. for CAR-T therapy), reducing the exhaustion of transferred T cell in the TME and enhancing the long-term persistence of the transferred T cells (Jiang et al., 2019).

### 1.3.4 Immunometabolism in T cell differentiation and function

In recent decades, the role of metabolism in the functional regulation of many different cell types of the immune system has been established, leading to the emergent and developing field of immunometabolism (Makowski, Chaib, & Rathmell, 2020). Among the most well studied is the metabolic underpinnings of the function and differentiation of T cells, which is central to cancer immunity (Waldman et al., 2020) and to the response in both ICB and ACT cancer immunotherapies (Fraiotta et al., 2018; Sade-Feldman et al., 2018). Different types of T cells with different differential stages are characterized by distinct metabolic profiles, which are not mere correlators of T cell states but are closely coupled to T cell signaling and can have a causal role in regulating cell function and determining cell fate (Klein Geltink, Kyle, & Pearce, 2018). Below I will describe and summarize the metabolic characteristics and their interconnection with cell signaling and cell function in different types of T cells.

Naive T cells (T<sub>h</sub>) are quiescent and relatively inactive metabolically, with a baseline level of glucose uptake which is used to generate ATP in OXPHOS (Klein

Geltink et al., 2018). The T<sub>h</sub> survival and homeostasis is partly maintained by IL-7 signaling as well as other pathways related to metabolic regulation including mTOR signaling (whose activation promotes glycolysis and drives the cell out of quiescence) (Yang, Neale, Green, He, & Chi, 2011) and extracellular adenosine signaling via its receptor (which helps to maintain IL-7R expression and decrease sensitivity to TCR activation) (Cekic, Sag, Day, & Linden, 2013).

Upon T cell activation, the TCR signaling upregulates cell metabolism via PI3K-AKT-mTOR pathway and the induction of c-Myc for subsequent transactivation of metabolic genes (Powell, Pollizzi, Heikamp, & Horton, 2012). In early activation, both glycolysis and TCA cycle/OXPHOS are increased. The increase in OXPHOS produces ROS, which is required for T cell activation (Sena et al., 2013), although an appropriate ROS level achieved by refined regulation together with antioxidation pathways is key to the proper T cell activation (Mak et al., 2017). Sufficient ATP generation via OXPHOS in early activation is also critical to ensure full activation and the differentiation of effector T cells (T<sub>eff</sub>). This increase in mitochondrial metabolic activity is supported by mitochondrial biogenesis, promoted by AMPK signaling and expression of PPAR and PGC-1 $\alpha$  (Scharping et al., 2016). Mitochondrial serine and one-carbon metabolism is also increased, which may reflect a need to synthesize glutathione for fine-tuning ROS levels (Ma et al., 2017). T cells also exhibit increased *de novo* purine synthesis and increased fatty acid synthesis early during activation (Lee et al., 2014; Ma et al., 2017). Several amino acids, including alanine, is also needed for T cell activation (Ron-Harel et al., 2019).

After full activation, CD8<sup>+</sup> T<sub>eff</sub> cells shows fast growth and proliferation, ex-

press high level of glucose transporters including GLUT1, and exhibit high levels of both glycolysis and OXPHOS (Fox, Hammerman, & Thompson, 2005), showing an aerobic glycolysis phenotype similar to that of many fast-proliferating cancer cells, famously known as the Warburg effect (Koppenol, Bounds, & Dang, 2011). Glycolysis has been shown as essential for normal Teff function, including in the anti-cancer settings (Chang et al., 2015), and the low-glucose condition in the TME can hamper effector function. Engagement of the inhibitory immune checkpoint PD-1 can inhibit glycolysis and promote FAO, negatively regulating T cell activation (Patsoukis et al., 2015). The reasons why Teff cells require glycolysis are not fully elucidated, although some studies shown that the lactate dehydrogenase in Teff is required for maintaining cellular Ac-CoA and the acetylation of IFN $\gamma$  promoter and enhancers (Peng et al., 2016). Despite the importance of glycolysis, mitochondria are also active in Teff, with high activity of mitochondrial fission (Buck et al., 2016). Glutaminolysis also has high activity in Teff, which was shown to be also used mainly for ATP instead of biomass production *in vitro* (Fox et al., 2005). Other amino acids including arginine, leucine and methionine have been shown to be important for Teff cell proliferation and effector function (Wang & Zou, 2020).

CD4<sup>+</sup> Teff cells have many established subtypes with distinct functions, and metabolism appears to be important in the CD4 differentiation into different subtypes. Activation of mTOR signaling which is a master regulator of cell metabolism promotes Th1 and Th17 differentiation; specifically mTORC2 signaling is needed for Th2 differentiation; on the other hand, inhibiting mTOR signaling shifts the differentiation towards Treg cells (Klein Geltink et al., 2018). Like CD8<sup>+</sup> Teff cells, CD4<sup>+</sup>

Teff also have active glucose import and glycolysis, together with active glutaminolysis, the inhibition of which will allow for Treg differentiation (Metzler, Gfeller, & Guinet, 2016). Besides generating ATP, glycolysis and glutaminolysis compete for substrates with the hexosamine biosynthesis pathway which synthesizes UDP-N-acetylglucosamine for O-GlcNAcylation of various proteins important for the differentiation of different CD4<sup>+</sup> T cell subtypes (Swamy et al., 2016). Fatty acid synthesis has been shown to be important for Th17 differentiation, and fatty acid synthesis deficiency can result in Treg cell differentiation (Berod et al., 2014).

As seen above, Treg cells do not have active fatty acid synthesis but in contrast use FAO, which has been shown to be important for their suppressive function (Procaccini et al., 2016). They also have active OXPHOS, consistent with the need for further oxidation of the Ac-CoA from FAO for ATP production (Gerriets et al., 2016). Glycolysis are also active in Treg cells, and appears to have a complex role on Treg function (Gerriets et al., 2016). The ability of Tregs to inhabit the acidic and nutrient-poor TME has been linked to their ability of utilize lactate in the TME, which is converted to pyruvate via the reversible lactate dehydrogenase reaction and then further used for TCA (Angelin et al., 2017). Intracellular cholesterol and cholesterol synthesis have also been shown to be important for the function of Treg cells (Zeng et al., 2013).

Memory T cells (Tm) display a distinct metabolic phenotype that is consistent with long-time survival. CD8<sup>+</sup> Tm cells have low glycolysis and remodels their mitochondria for high-efficiency OXPHOS with lower mitochondrial membrane potential (Sukumar et al., 2013). Tm cells also have increased triglyceride storage and

rely on FAO for energy production (Pearce et al., 2009). Arginine has been shown to enhance the long-term survival of Tm cells (Geiger et al., 2016).

Currently, most characterizations of T cell metabolism have been focused around the central energy metabolism (TCA/OXPHOS) and a few other common pathways like glycolysis and lipid metabolism. A more complete picture of global cell metabolic changes during T cell differentiation is still missing. Further, much of the current knowledge on T cell metabolism has been derived from *in vitro* systems, while differences can exist for the *in vivo* setting (Artyomov & Van den Bossche, 2020). The relationship between metabolic state and T cell immune functional state can be complicated, often leading to apparently paradoxical observations which requires deeper investigation to understand the underlying mechanisms and reasons for the particular metabolic regulation observed. As an example, a recent study (Vodnala et al., 2019) has shown that a nutrient deprivation state associated with high extracellular potassium concentration in the TME contributes to the dysfunction of the tumor-infiltrated T cells, but the suppression of effector function associated with such a state also corresponds to increased T cell stemness and can be harnessed to improve long-term T cell persistence and tumor clearance in ACT therapy. This study also highlights that a deep understanding of T cell metabolism can be translated into strategies for modulating T cell function and improve the efficacy of cancer immunotherapies (Makowski et al., 2020).

#### 1.4 Metabolism in COVID-19 and host-SARS-CoV-2 interaction

### 1.4.1 COVID-19: natural history, epidemiology and clinical features

The coronavirus disease 2019 (COVID-19) is a severe and highly infectious respiratory disease whose first known case appeared in late December 2019 in Wuhan, China (Zhu et al., 2020). Since then, COVID-19 has quickly evolved into a global pandemic within a few months by March 2020 (WHO, 2020c) and is still ongoing, resulting in more than 100 million infected cases and over 2 million deaths (as of March 2021) (WHO, 2020b). The pathogen for COVID-19 has been identified to be a new betacoronavirus that was never discovered before, and was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). Two related viruses of the coronavirus family (*Coronaviridae*), the severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), emerged in 2002 and 2012 respectively have also caused serious respiratory diseases (Cui, Li, & Shi, 2019), making COVID-19 the third coronavirus-related severe infectious disease discovered in the recent decades. Like SARS-CoV and MERS-CoV (Cui et al., 2019), SARS-CoV-2 is likely to be of zoonotic origin, with the first COVID-19 cases epidemiologically linked to a wet market in Wuhan selling seafood, live poultry and wildlife (Jiang, Du, & Shi, 2020). Phylogenetic studies have identified high similarities between the genomic sequences of SARS-CoV-2 and a group of betacoronaviruses of the subgenus *Sarbecovirus* previously found in bats in China (F. Wu et al., 2020; P. Zhou et al., 2020), and SARS-CoV-2-related coronaviruses have also been found in Malayan pangolins (*Manis javanica*) (Lam et al.,

2020). However, the identity of the direct progenitor of SARS-CoV-2 is still uncertain (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020), and the intermediate host responsible for the zoonotic transmission of the virus from the wild reservoir host species to human has not been confidently identified (J. Zhao et al., 2020).

Airborne transmission has been determined to be the dominant route of transmission of COVID-19 (R. Zhang et al., 2020), specifically via respiratory droplets and potentially aerosol (Doremalen et al., 2020), fecal-oral transmission can also be possible (Xu et al., 2020). COVID-19 is highly transmissible, with a basic reproduction number (the expected number of infections directly caused by transmission from one primary infected individual when all individuals are susceptible) estimated to be around 2 or 3 (Billah, Miah, & Khan, 2020; S. Zhao et al., 2020), similar to SARS and higher than e.g. the 2009 influenza and MERS (Petersen et al., 2020). The incubation period (time from infection to symptom onset) of COVID-19 was estimated to range from 1.3 to 11.3 days (Lauer et al., 2020), longer than SARS (Petersen et al., 2020). SARS-CoV-2 show a distinct kinetics of virus shedding, with high viral load when symptoms begin and a subsequent decline but persisting for a long period, making it critical to timely isolate and quarantine patients as soon as infection is detected and for sufficiently long time (Petersen et al., 2020; Wölfel et al., 2020; Zheng et al., 2020). COVID-19 is less deadly than SARS or MERS, with possibly more than 50% infected cases being asymptomatic (G. Li et al., 2020), and most symptomatic cases being mild (Z. Wu & McGoogan, 2020). However, asymptomatic patients can still spread the virus (Li et al., 2020), adding to the challenge of transmission control. The infection-fatality risk was estimated to be 1.39% in the

New York City, the first epidemic center in the USA (Yang et al., 2021), and other estimates for other regions in the world were mostly below 1% (Pastor-Barriuso et al., 2020; Streeck et al., 2020). Older age and various comorbidities including hypertension, obesity, diabetes and coronary heart disease (among others) were identified as the risk factors for SARS-CoV-2 infection and/or severe disease (Jordan & Adab, 2020; B. Mao et al., 2020).

The most common initial symptoms of COVID-19 include dry cough, fever, and fatigue, additional symptoms can include sputum production, sore throat, haemoptysis, chills, headache, chest pain, anorexia, diarrhea, and nausea/vomiting (Hu, Guo, Zhou, & Shi, 2020). Dyspnoea and pneumonia can subsequently develop within a median time of 8 days from disease onset (Z. Wu & McGoogan, 2020). Severe or critical cases (about 20% in some reports) require ventilation and intensive care unit (ICU) admission and can develop respiratory failure, septic shock and/or multiple organ dysfunction or failure (Z. Wu & McGoogan, 2020). Although being primarily a respiratory disease, multiple organs other than the lung and respiratory tract can be affected (Zaim, Chong, Sankaranarayanan, & Harky, 2020), including the heart and blood vessels (Bikdeli et al., 2020; Shi et al., 2020), liver (C. Zhang et al., 2020), kidneys (Diao et al., 2020), brain (L. Mao et al., 2020) and intestine (Xiao et al., 2020), among others. Although most patients recover from the disease, there can be long-term persisting symptoms (e.g. fatigue, shortness of breath, cough, joint pain and chest pain) (CDC, 2020) and/or abnormalities found in clinical tests (e.g. pulmonary diffusion impairment) (Cortinovis, Perico, & Remuzzi, 2021).

## 1.4.2 SARS-CoV-2 biology

SARS-CoV-2 is a positive-sense single-stranded RNA (+ssRNA) enveloped virus from the *Coronaviridae* family, *Betacoronavirus* genus, and the *Sarbecovirus* subgenus (Zhu et al., 2020). Its RNA genome has 29,891 bases, with a similar genomic organization with other betacoronaviruses (Santos, 2020). The genome encodes 29 proteins, including 6 functional open reading frames (ORFs) for the replicase (ORF1a/1b), spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins (the latter four being structural proteins), and 7 putative ORFs for accessory proteins (Chan et al., 2020). The replicase ORF encodes two large polyproteins that are proteolytically cleaved into 16 non-structural proteins (nsp1-16) (Chan et al., 2020). The spike protein is integrated into the viral envelope resulting in crown-like spikes on the surface of the virus, giving coronavirus its name (Santos, 2020), which is also responsible for the viral attachment and entry of host cells via endocytosis (J. Shang, Ye, et al., 2020; Walls et al., 2020). It has been determined that the human angiotensin-converting enzyme 2 (ACE2) protein, present on the cell membrane, is the host receptor for viral entry (R. Yan et al., 2020). The receptor binding motif (RBM) within the domain (RBD) of the viral spike protein trimer binds to the host ACE2 protein; host proteases including the serine protease TMPRSS2, cathepsin B/L and furin then cleave and “prime” the spike protein at the cell surface or in the early endosome (Hoffmann et al., 2020; J. Shang, Wan, et al., 2020), triggering a conformation change in the spike protein that subsequently promote membrane fusion and viral entry (Chambers, Yu, Valdes, & Arulanandam,

2020; Hoffmann et al., 2020).

After entry into the host cell, the virus initiates a complicated process of protein expression and genome replication. The translation of the two replicase ORFs utilizes a programmed-1 ribosomal frameshift (Finkel et al., 2021), producing two polyproteins that will be cleaved into the 16 non-structural proteins (nsp1-16) via by two cysteine protease activities present in nsp3 and nsp5 (V'kovski, Kratzel, Steiner, Stalder, & Thiel, 2020). Nsp2-16 forms the virus replication and transcription complex (RTC) (L. Yan et al., 2020), including the RNA-dependent RNA polymerase (RdRP, nsp12; with cofactors nsp7 and nsp8, as well as nsp14 which has an exonuclease domain for proofreading) for viral genome replication (Eckerle, Lu, Sperry, Choi, & Denison, 2007; Gao et al., 2020). In other coronaviruses, RTC is formed within double-membrane vesicles (DMVs) derived from endoplasmic reticulum (ER) and other host endomembranes; it is localized to specific subcellular locations involving complex interactions with host proteins to further advance the viral cell cycle (V'kovski et al., 2019). The viral genome replication first produces a full-length negative-sense RNA, which is then further replicated into positive-sense genomic RNA (V'kovski et al., 2020). More viral proteins can then be translated from the newly synthesized viral genome. As with other coronaviruses, the genome replication is characterized by a unique discontinuous transcription at transcription regulatory sequences (TRSs), producing a nested set of subgenomic RNAs (sgRNAs) with shared 5' and 3' sequences (Sola, Almazán, Zúñiga, & Enjuanes, 2015). The virion assembly involves the association between the viral RNA genome with the nucleocapsid proteins and happens on the cytoplasmic side of the ER-Golgi

intermediate compartment (ERGIC) cisternae (Klein et al., 2020). The assembled virions within membrane-bound vesicles are then released via exocytosis (Ghosh et al., 2020; Klein et al., 2020), completing the life cycle of the virus.

The host-pathogen interactions (HPI) of SARS-CoV-2 are highly complex. *In vitro* studies of SARS-CoV-2 protein expression followed by affinity purification and mass spectrometry (MS)-based protein identification has identified hundreds of host-SARS-CoV-2 protein-protein interactions (PPI) (Gordon et al., 2020). The host cell typically senses viral invasion and mounts anti-viral responses via different innate immune pathways, notably the type I/III interferon (IFN) pathway (Vabret et al., 2020), however, SARS-CoV-2 is able to suppress the activity of these pathways via different mechanisms (Vabret et al., 2020; V'kovski et al., 2020). For example, the ORF3b of the virus may effectively inhibit IFN response (Konno et al., 2020), and it has been shown that SARS-CoV-2-infected cells and patients exhibit impaired IFN response (Blanco-Melo et al., 2020; Hadjadj et al., 2020). Other viral proteins, on the other hand, may induce the activation of various inflammatory pathways, e.g. interleukin (IL)-6 and IL-8 production that can cause the symptoms associated with severe diseases (J. Li et al., 2021). It has been suggested that an imbalanced immune response involving the failure to mount an early IFN response and a dysregulated inflammatory response during SARS-CoV-2 infection drives the development of COVID-19 (Blanco-Melo et al., 2020; Giamarellos-Bourboulis et al., 2020).

### 1.4.3 Anti-COVID-19 therapies and vaccines

Governments, industries and scientific communities across the world and have responded quickly to the COVID-19 crisis in search of efficient preventative and therapeutic strategies. To date (March 2021), there are thousands of registered clinical trials on COVID-19 (WHO, 2021). Potential therapies under clinical investigation include small-molecular drugs, biologics e.g. monoclonal antibodies, plasma-based and cell-based therapies, among others (“Global Coronavirus COVID-19 Clinical Trial Tracker,” 2020; Thorlund et al., 2020). Unfortunately, Very few therapies have been established as beneficial to COVID-19 patients. Drugs proposed for COVID-19 treatment early in the pandemic including lopinavir, ritonavir and hydroxychloroquine were later shown to be ineffective (Cavalcanti et al., 2020; RECOVERY Collaborative Group, 2020). Remdesivir, a viral RdRP inhibitor represents the only drug approved by the drug regulatory authorities of several countries including the U.S. Food and Drug Administration (FDA) (Beigel et al., 2020; FDA, 2020d). However, it confers only mild clinical benefits to a subset of COVID-19 patients, and is not recommended by the World Health Organization (WHO) as a first-line treatment for COVID-19 patients (Spinner et al., 2020; WHO, 2020a; WHO Solidarity Trial Consortium et al., 2020). Janus kinase (JAK) inhibitor baricitinib (in combination with remdesivir), and virus-neutralizing antibodies bamlanivimab, and casirivimab plus imdevimab have obtained Emergency Use Authorization (EUA) from the FDA (FDA, 2020a, 2020c, 2020b). Dexamethasone and other corticosteroids have been recommended by the U.S. National Institutes of Health (NIH) for

hospitalized patients requiring supplemental oxygen (NIH, 2020; RECOVERY Collaborative Group et al., 2020). To date, one of the most promising and encouraging advances among the anti-COVID-19 efforts has been the successful development of several SARS-CoV-2 vaccines around the world, some have been approved or authorized for emergency use in different countries (Dong et al., 2020; FDA, 2021b, 2021a).

There have been a lot of preclinical efforts in identifying potential therapeutic drugs and targets for anti-SARS-CoV-2. Numerous antiviral drug repurposing screens of varied scales using libraries of existent drugs have been performed *in vitro* (e.g. Riva et al., 2020; Touret et al., 2020). Some studies have tried to identify effective drug combinations (e.g. Bobrowski et al., 2021; Nguyenla et al., 2020). A few genetic screens using e.g. the CRISPR-Cas9 technology have been reported, revealing potential gene targets for inhibiting SARS-CoV-2 infection or proliferation (Daniloski et al., 2021; Schneider et al., 2021; R. Wang et al., 2021; Wei et al., 2021). A recent study reported the application of CRISPR-Cas13 in degrading SARS-CoV-2 RNA and suppression the viral infection (Abbott et al., 2020). Different parties and research groups have created databases containing compiled resources on potential anti-SARS-CoV-2 drugs, targets and therapies (e.g. Kuleshov et al., 2020; Chen, Allot, & Lu, 2021). There are also large number of computation-driven efforts for predicting anti-SARS-CoV-2 drugs and/or targets, based on a wide variety of algorithms (e.g. Bobrowski et al., 2021; Y. Zhou, Hou, et al., 2020; Y. Zhou et al., 2020).

#### 1.4.4 Viral regulation of host metabolism

Viruses are known to “hijack” host cell metabolism to complete their own intracellular life cycle (Mayer, Stöckl, Zlabinger, & Gualdoni, 2019; Thaker, Ch’ng, & Christofk, 2019), modulating diverse pathways including carbohydrate, lipid, amino acid and nucleotide metabolism (Mayer et al., 2019; Sanchez & Lagunoff, 2015). Coronaviruses including MERS-CoV rearrange cellular lipid profiles upon infection (Yan et al., 2019; Yuan et al., 2019); as described above, coronaviruses induce the formation of complex networks of DMVs derived from ER and other intracellular membranes, in which they assemble their RTCs for viral replication and protein translation (V’kovski et al., 2019), and the formation of DMVs and RTCs are critical for the viral life cycle. Notably, counteracting the metabolic demands of viruses including MERS-CoV have been shown to abolish their ability to infect the host cells (Mayer et al., 2019; Yuan et al., 2019). Targeting the virus-induced metabolic changes has been proposed to be a promising novel antiviral strategy (Mayer et al., 2019).

Soon after the emergence of COVID-19, people have noticed that certain metabolic factors are important determinants of disease risk and severity (Marazuela, Giustina, & Puig-Domingo, 2020). Many reports have shown that diabetes patients have significantly increased risk of morbidity and mortality due to COVID-19 (e.g. Guan et al., 2020; Petrilli et al., 2020). Increase glucose level can promote SARS-COV-2 infection and proliferation in human monocytes (Codo et al., 2020). However, the full mechanisms underlying this association are yet unclear and likely

involve systemic factors such as lung function and immune response being modulated by glucose level (Lim, Bae, Kwon, & Nauck, 2021). Obesity is another risk factor for hospitalization and severe disease of COVID-19 (Petrilli et al., 2020). On the other hand, survivors of SARS may develop new-onset and long-lasting metabolic abnormalities, which may also occur in COVID-19 patients (Ayres, 2020). Recent studies have reported that SARS-CoV-2 induces metabolic changes in numerous pathways including tricarboxylic acid (TCA) cycle, oxidative phosphorylation, lipid metabolism, amino acid metabolism (e.g. tryptophan metabolism), and pentose phosphate pathway, among others in human patient samples (Bojkova, Costa, et al., 2020; Ehrlich et al., 2020; Gardinassi, Souza, Sales-Campos, & Fonseca, 2020; S. Li et al., 2021; Thomas et al., 2020). Several new studies have shown that anti-SARS-CoV-2 effect can be achieved by targeting various host metabolism pathways, for example, the PPAR $\alpha$ -agonist fenofibrate can reverse SARS-CoV-2-induced metabolic changes and block the viral replication (Ehrlich et al., 2020); DHODH inhibitors that suppress pyrimidine *de novo* synthesis showed anti-SARS-CoV-2 activity (R. Xiong et al., 2020); the inhibition of certain phosphoinositides can disrupt endocytosis and block SARS-CoV-2 cell-entry (Ou et al., 2020); and transketolase inhibitor was shown to inhibit SARS-CoV-2 in a dose-dependent manner (Bojkova, Costa, et al., 2020). These studies suggest that targeting metabolism as an antiviral strategy can also be applied to SARS-CoV-2.

## 1.5 Overview of study: rationale, aims and significance

Cell metabolism plays a pivotal role in many biological processes including T cell function in cancer and cancer immunotherapy, and in the host response to virus infection. Genome-scale metabolic modeling has been successfully applied to study cell metabolism under different contexts, and with the advent of updated GEMs and advanced modeling algorithms, metabolic modeling with GEM has been shown to have great value in generating testable hypotheses and accurate predictions especially when combined with the abundance of omic data available nowadays. It is highly desirable to closely study the metabolic underpinnings of immune cell functions and host-virus interaction utilizing the methodological framework of genome-scale metabolic modeling via the integration of experimental omic data. In this thesis I will describe efforts on algorithm development under the metabolic modeling framework and demonstrate its utility in two case studies, specifically: 1. the understanding of CD8 T cell metabolism in cancer and identification of essential metabolic processes that determine the anti-cancer function of the T cells; and 2. the investigation of host metabolic response under SARS-CoV-2 infection and the identification of potential anti-SARS-CoV-2 targets which act via counteracting the virus-induced metabolic changes. These represent some of the first studies of the application of genome-scale metabolic modeling to these fields of research and highlights the promise of the modeling-based approach in facilitate a deeper understanding of the role of cell metabolism under diverse biological contexts.

## Chapter 2: Analysis and improvement of metabolic transformation algorithms

### 2.1 Introduction

GEM algorithms have been applied to study cell metabolism under a wide variety of contexts, and have repeatedly proven to be valuable for generating accurate predictions and informative hypotheses (Gu, Kim, Kim, Kim, & Lee, 2019). An overview of GEM and its methodological framework is provided in Section 1.2. Here we focus on the MTA algorithm, initially developed by (Yizhak et al., 2013), which is used for the prediction of metabolic targets whose KO can transform the cellular metabolic state from a given reference state to a given target state. Apparently, this algorithm can be used for therapeutic target prediction for many metabolism-related diseases, by reversing the perturbed cell metabolism under the diseased state back to the normal healthy state. But due to its mechanism-based modeling approach under the GEM framework, MTA can also be more widely used to predict potential causal factors and perturbations beyond what can be identified from classical statistical inference. In the original study, MTA was used to predict lifespan-extending genes in the budding yeast *Saccharomyces cerevisiae*, and suc-

cessfully identified two novel lifespan-extending metabolic genes (*adh2* and *gre3*) whose KO's were experimentally validated to significantly prolong the chronological lifespan of yeast (Yizhak et al., 2013). In a subsequent study, MTA was used to predict metabolic cancer driver genes, resulting in a novel cancer driver gene *FUT9* in colorectal cancer (Auslander et al., 2017). A more recent study applied MTA and successfully predicted mitochondrial dihydroorotate dehydrogenase (DHODH) involved in pyrimidine nucleotide biosynthesis as an effective therapeutic target for an intractable epilepsy called Dravet syndrome (Styr et al., 2019). Despite repeated successful applications, the original algorithm was mostly validated in the bacteria *E. coli*, with only a few validations in higher organisms (2 validation datasets for mouse and 2 for human); the performance in the mice and human datasets also appeared relatively weaker than those in the *E. coli* (Yizhak et al., 2013). A recent study introduced an improvement to MTA named robust MTA (rMTA), which was shown to achieve better performance compared to the original MTA (Valcárcel, Torrano, Tobalina, Carracedo, & Planes, 2019). However, in terms of validation in higher organisms, the authors of rMTA only used the same two human and two mouse datasets as in (Yizhak et al., 2013), thus the scope of its validation is still very limited. Here I aim to perform a deeper investigation to comprehensively evaluate the performance of MTA (and rMTA) in higher organisms focusing on human, also seeking to optimize the algorithm and further improve its prediction accuracy.

## 2.2 Results

## 2.2.1 The MTA algorithm as in (Yizhak et al., 2013)

Here following the introduction to GEM in Section 1.2, I first provide a description of the algorithm details of MTA as developed in (Yizhak et al., 2013). The input to MTA is the transcriptome-wide gene expression data of the reference state, in addition to the differential gene expression (DE) changes in the target state compared to the reference state, provided as two vectors; MTA outputs a score called the MTA score for each reaction, with higher MTA scores corresponding to better targets, i.e. whose KO can better transform the cell metabolism from the reference to the target state. An high-level overview of the MTA algorithm is given in the diagram in Figure 2.1.

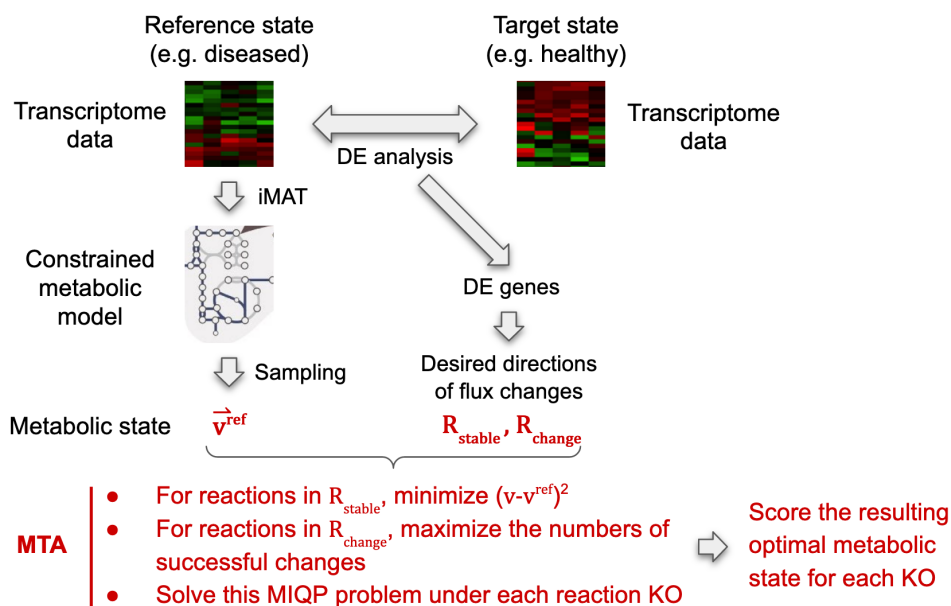


Figure 2.1: A schematic diagram as an overview of the MTA algorithm.

A first step of MTA involves applying the iMAT algorithm (Shlomi et al., 2008) to obtain a GEM customized to the reference state. Then uniform random

sampling is performed on the reference GEM, and the mean vector  $\mathbf{v}^{\text{ref}}$  of all sample points is obtained as the reference metabolic flux distribution of the reference state. Up-regulated, down-regulated and stable genes are then determined from the DE results based on certain cutoffs, which are mapped based on the boolean GPR associations to the reaction level to obtain the sets of reactions that are expected to have increased activity level (i.e. absolute flux) or decrease activity level. According to (Yizhak et al., 2013), the cutoff for determining significantly up- or down-regulated genes were partly based on considerations of computational time cost, as specifying a higher number of DE genes will increase the time to solve the MIQP optimization problem later (see below). The authors recommended picking a number of top DE genes such that they are mapped via GPR to a total of about 100 reactions that are intended to have altered activities. However this is a parameter that in principle can be further tuned. Specifically for the mapping of genes to reactions via GPR, if a reaction is expected to have increased activity level if it is catalyzed by a set of enzymes with  $\&$  relations and all of the corresponding genes are up-regulated, or if it is catalyzed by a set of enzymes with  $|$  relations and at least one of the corresponding genes are up-regulated; similarly reactions expected to have decreased activities are determined. All other cases not covered in the above description are regarded as undetermined. For the reactions with expected increase in activity and whose fluxes are positive in the reference state  $\mathbf{v}^{\text{ref}}$ , they are expected to have further increased flux value (i.e. further flux change in the forward direction, towards more positive values); similarly, for the reactions whose initial reference fluxes  $v_i^{\text{ref}}$ 's are negative and are expected to have decreased activity level, we expect that the absolute val-

ues of their fluxes to decrease, which also corresponds to their flux values increasing in the forward direction towards more positive values; this set of reactions is denoted  $R_F$ . Similarly the set of reactions whose flux value are expected to decrease (i.e. change in the backward direction, towards more negative values) can be identified and is denoted  $R_B$ . The set of reactions that are expected to have unchanged, stable activity levels is denoted  $R_S$ ; any undetermined reaction is regarded as having stable flux and is also included in  $R_S$ . Next, a series of MIQP problems under the simulated KO of each reaction  $j$  are solved to maximize the number of reactions with flux changes concordant with the expectations, while minimizing the squared sum of flux changes in reactions that are expected to have stable fluxes:

$$\underset{\mathbf{v}, \mathbf{y}}{\text{minimize}} \left( (1 - \alpha) \sum_{i \in R_S} (v_i^{ref} - v_i)^2 + \frac{\alpha}{2} \sum_{i \in R_F \cup R_B} y_i \right)$$

subject to:

$$\mathbf{Sv} = \mathbf{0}$$

$$\mathbf{v}^{lb} \leq \mathbf{v} \leq \mathbf{v}^{ub}$$

$$v_j = 0$$

$$v_i - y_i^F (v_i^{ref} + \epsilon_i) - y_i v_i^{lb} \geq 0, \text{ for } i \in R_F$$

$$y_i^F + y_i = 1, i \in R_F$$

$$v_i - y_i^B (v_i^{ref} - \epsilon_i) - y_i v_i^{ub} \leq 0, \text{ for } i \in R_B$$

$$y_i^B + y_i = 1, i \in R_B$$

$$y_i, y_i^F, y_i^B \in \{0, 1\}$$

In the formulation above, for each reaction  $i \in R_F$ , binary variables  $y_i$  and

$y_i^F$  are introduced such that one and only one of the two variables can be 1 (while the other has to be 0); similarly binary variables  $y_i$  and  $y_i^B$  are introduced for each reaction  $i \in R_B$ .  $y_i^F = 1$  or  $y_i^B = 1$  correspond to “successful” flux change that is consistent with the expected changes for reaction  $i$ , and thus  $y_i = 1$  means a failure to achieve the expected flux changes. Therefore, the integer part of the minimization problem corresponds to minimizing the number of failed flux changes. The relative weights of this part and the quadratic part (which minimizes the squared sum of flux changes in reactions that are expected to have stable fluxes) are determined by the parameter  $\alpha$ . In the original study (Yizhak et al., 2013), performance was found to be robust to the variation in this parameter, and a default value of 0.66 was used in their analyses. However, the robust analysis was only performed using several validation datasets in *E. coli* and yeast.  $\epsilon_i$  is a parameter that corresponds to the size of flux change used to define a “successfully” changed reaction, while it can be reaction-specific in principle, it was shown that a fixed value can achieve reasonably good performance, although again the evaluations that led to this conclusion were performed using *E. coli* and yeast datasets. A fixed value of  $\epsilon_i = 0.01$  was recommended as a reasonable default value (from personal correspondence with the authors).

After solving the above MIQP problem for the KO of each reaction, the optimal solution is obtained and thereby the sets of “successfully” changed reactions ( $R_{success}$ ) as well those failed to achieve the expected change ( $R_{fail}$ ) can be identified. This is mostly straightforward although two special cases requires further explanation: for the reactions whose initial reference fluxes  $v_i^{ref}$ 's are negative and

are expected to have decreased activity level, we expect that the absolute values of their fluxes to decrease, which corresponds to their flux values increasing towards more positive values, although it's not desired if their final flux value becomes larger than  $-v_i^{ref}$ . If the final flux from the optimal solution of the MIQP is larger than the corresponding  $-v_i^{ref}$ , although it contributed to the minimization of the MIQP objective function, it should be regarded as a failed case. Similarly this applies to reversible reactions with positive initial reference flux and an expected decrease in their activity level, but the final flux is smaller than  $-v_i^{ref}$ . These special "overshoot" reactions are denoted  $R_{fail,overshoot}$  (and not included in the "normal"  $R_{fail}$  set; the  $R_{fail,overshoot}$  was not explicitly explained in the original MTA (Yizhak et al., 2013) but was actually implemented as such). The MTA score for the KO is then computed as follows:

$$\frac{\sum_{i \in R_{success}} |v_i^{ref} - v_i^{opt}| - \sum_{i \in R_{fail}} |v_i^{ref} - v_i^{opt}| - \sum_{i \in R_{fail,overshoot}} \left( |v_i^{ref}| - |v_i^{opt}| \right)}{\sum_{i \in R_S} |v_i^{ref} - v_i^{opt}|}$$

The MIQP problem can also be solved using the WT model without any KO, which can also produce an MTA score that can be used as a control. Only reactions whose scores are greater than the control scores are potentially useful targets. Based on the validations mostly based on *E. coli* and yeast reported in (Yizhak et al., 2013), the 10%-20% reactions with the highest MTA scores appear to represent biologically meaningful targets in that in most of the validation datasets, the reaction corresponding to the actual gene KO (after mapped to reactions via

GPR associations) lies within top 10%-20% MTA predictions.

## 2.2.2 Evaluating the performance of the MTA algorithm in human datasets

As described above, MTA as originally reported by (Yizhak et al., 2013) has not been extensively tested in human. Given its many potential applications in human biology including the prediction of therapeutic targets for different human diseases, it is desirable to perform a more comprehensive evaluation of MTA using human benchmark datasets. Among the key points to evaluate is the robustness to the choice of the various parameters  $\alpha$  and  $\epsilon_i$ , and the determination of their best default values, since the previous analyses reported in (Yizhak et al., 2013) were mostly in *E. coli* and yeast, and the difference in the organism and its GEM, as well as the type of dataset (from different assay platforms) may influence the robustness analysis and the choice of parameters. Besides, the number of top DE genes is another parameter that can be tuned but not explored/reported in the original study (Yizhak et al., 2013). To this aim, I manually collected a total of 58 human microarray or RNA-seq transcriptome datasets involving the KD or KO or drug inhibition (by small molecules) of a metabolic gene/protein from the GEO database. Each dataset contains the expression profiles of the untreated or vehicle-treated control samples and those of the treated (i.e. KO/KD/drug inhibition) samples. The GEO dataset IDs of the collected datasets are provided in Appendix Table 4.2. I applied MTA to each of the datasets, with the gene expression profiles of the

control and treated groups as inputs, and subsequently inspected MTA score(s) and percentage rank(s) across all reactions of the “ground truth” reaction(s), i.e. the reaction(s) mapped via the GPR to the actual gene or protein being KO-ed/KD-ed/inhibited. When more than one reaction is mapped, I focused on the top ranked reaction according to MTA prediction. The effects of the parameters  $\alpha$  and  $\epsilon_i$ , as well as the number of top DE genes (which I denoted  $n$ ) on the performance are all thoroughly investigated with grid search.  $\alpha$  values ranging from 0.02 to 0.98,  $\epsilon_i$  values ranging from 1e-4 to 0.1, and  $n$  ranging from 10 to 500 were tested.

Evaluating the prediction accuracy of MTA in the 58 human datasets, I found that indeed the performance of MTA is highly dependent on the parameters chosen (Figure 2.2). As a crude measure, I evaluated the performance based on the number of datasets (out of the 58) where at least one ground truth reactions is among the top 10% MTA predictions. Overall, the number of top DE genes  $n$  used appeared to have a larger effect on the performance, with lower values of  $n$  below 100 producing markedly better performance than higher  $n$  values.  $\alpha$  values also had a large effect, with  $\alpha$  in the very high range producing noticeably better performance. Different  $\alpha$  values appeared to work relatively better with specific ranges of  $\epsilon_i$  values, with a weak trend of higher  $\alpha$  values working better with higher  $\epsilon_i$  values. Overall the best performance was achieved at  $\alpha = 0.98$ ,  $\epsilon_i = 0.05$  and  $n = 40$ , where 28 out of the 58 datasets have their ground truths predicted among the top 10%. Specifically inspecting this optimal parameter combination, 23 out of the 58 datasets have “better than top 5%” performance, and 3 datasets have “better than top 1%” performance. The performances vary given local variation in the parameter near

this optimal region, but in general yielding the “better than top 10%” performance in more than 20 datasets (Figure 2.2).

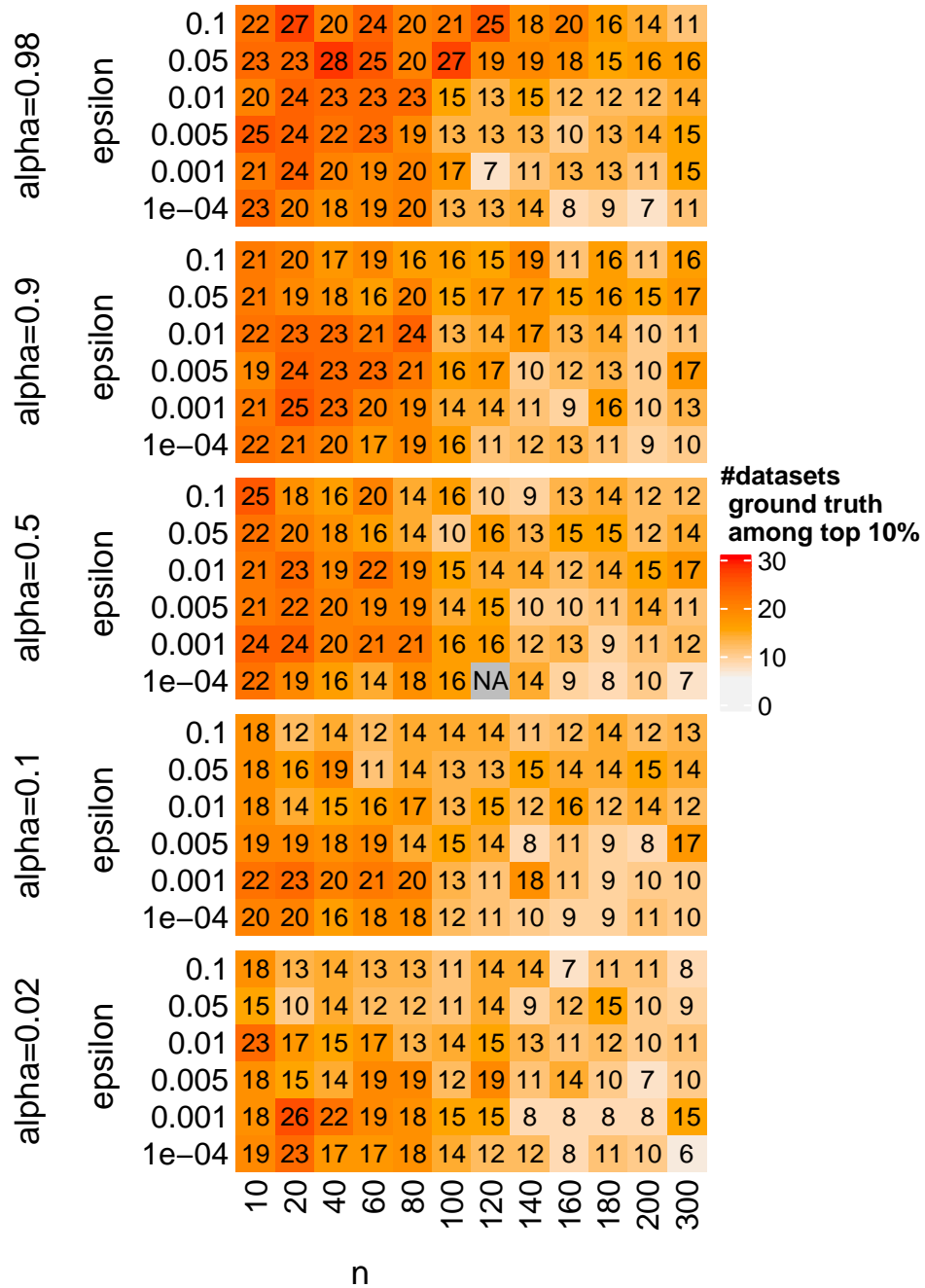


Figure 2.2: Evaluation of the prediction performance of MTA on 58 human datasets using different parameters.

### 2.2.3 First proposed alternative to the original MTA and its evaluation

One perceived drawback of the original MTA algorithm is that the objective function involves the maximization of the number of reactions that are intended to be altered, while minimizing the summed square of flux differences of the reactions that are intended to remain stable. This leads to a MIQP problem that can be difficult and time-consuming to solve. Besides, the maximization part and the minimization part of the optimization have intrinsically different biological meaning and are on different scales, thus they need to be balanced by the parameter  $\alpha$ . Further, the definition of “significant” alteration of a reaction flux is determined by the  $\epsilon_i$  parameter. As shown above, the number of top DE genes is also a parameter that can greatly affect the performance. It could be desirable to reduce the number of parameters with an alternative formulation, if reasonable performance can be retained. The original MTA formulation introduces the further complexity that the KO’s were ultimately evaluated based on yet another fraction scoring function different from the MIQP objective function. The authors argued that the scoring function cannot be optimized with off-the-shelf techniques due to its non-linear fraction form, but also due to that the calculation is dependent on the knowledge of whether a reaction is “successfully” altered as intended or not (Yizhak et al., 2013). However, it could be possible to explore the use of a linear scoring function instead of one in fraction form, that is also independent of any prior knowledge on whether a reaction is actually successfully changed.

With the aim of addressing the above drawbacks of the original MTA, I explored an alternative formulation for the same biological prediction problem with the same input. Instead of maximizing of the number of reactions that are intended to be altered, I seek to maximize the total absolute flux changes in these reactions; similarly instead of minimizing the summed square of flux differences of the reactions that are intended to remain stable, I seek to minimize their total absolute flux changes. This leads to an optimization problem involving absolute values that can be transformed into a simple LP problem, which eliminates the need for any additional parameter and a second scoring function. This new formulation is as follows:

$$\begin{aligned} & \underset{\mathbf{v}}{\text{minimize}} \frac{(1 - \alpha)}{|R_S|} \sum_{i \in R_S} |v_i - v_i^{ref}| + \\ & \frac{\alpha}{|R_{rr} \cup R_B \cup R_F|} \left( \sum_{i \in R_{rr}} (|v_i| - |v_i^{ref}|) + \sum_{i \in R_B} (v_i - v_i^{ref}) + \sum_{i \in R_F} (v_i^{ref} - v_i) \right) \end{aligned}$$

subject to:

$$\mathbf{Sv} = \mathbf{0}$$

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

$$v_j = 0$$

Here the definition of the  $R_S$  is the same as that in the original MTA.  $R_{rr}$  denotes the set of reactions that are reversible and are expected to have reduced activity levels, i.e. a decreased in the absolute values of their fluxes. The  $R_F$  and  $R_B$  sets of reactions are also defined in the same way as in the original MTA except

that the reactions in  $R_{rr}$  are excluded from these sets. Thus it can be seen that this formulation has the additional benefit of addressing the potential “overshoot” cases directly within the objective function, rather than adjusting for it in a *post hoc* fashion. The flux changes associated with the reactions that are expected to remain steady ( $R_S$ ) and the rest of the reactions that are expected to have altered activities ( $R_{rr} \cup R_B \cup R_F$ ) are normalized by the sizes of these two sets of reactions, respectively. Besides, analogous to the original MTA, the relative weights of these two parts can be further adjusted using the  $\alpha$  parameter. However, given that now both parts have the same biological meaning (i.e. flux change) and are thus on the same numerical scale,  $\alpha = 0.5$  is a very natural choice as the default value. This optimization can be transformed to the equivalent LP problem below by introducing an additional pair of variables  $p_i$  and  $n_i$  for each reaction in  $R_S$  and  $R_{rr}$ :

$$\begin{aligned} & \underset{\mathbf{v}, \mathbf{p}, \mathbf{n}}{\text{minimize}} \frac{(1 - \alpha)}{|R_S|} \sum_{i \in R_S} (p_i + n_i) + \\ & \frac{\alpha}{|R_{rr} \cup R_B \cup R_F|} \left( \sum_{i \in R_{rr}} (p_i + n_i) + \sum_{i \in R_B} v_i - \sum_{i \in R_F} v_i \right) \end{aligned}$$

subject to:

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

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

$$v_j = 0$$

$$p_i - n_i = v_i - v_i^{\text{ref}}, \text{ for } i \in R_S$$

$$p_i - n_i = v_i, \text{ for } i \in R_{rr}$$

$$p_i, n_i \geq 0$$

Given that this formulation is an LP problem, I termed it MeTAL. The prediction performance of MeTAL was again evaluated on the same 58 validation datasets, with different choices of parameters. The major parameter is  $n$ , the number of top DE genes. Although the  $\alpha$  parameter has a natural default value as described above, I still explored how the performance can vary dependent on it. Here it was found that MeTAL overall had worse performance compared to the original MTA, unfortunately (Figure 2.3). Like in MTA, larger  $\alpha$  appeared to yield relatively better performances, although the natural default of  $\alpha = 0.5$  is indeed a good choice. Global variation in the performance due to the choice of  $n$  appeared to be smaller compared to MTA, although there are some local instabilities. The best performance is achieved by the combination  $\alpha = 0.5$  and  $n = 140$ , successfully predicting

the “ground truth” reaction among the top 10% in 21 out of the 58 datasets. Although this is still largely better than random (binomial test  $P=8.3061099 \times 10^{-8}$ ), it is worse than the best-case performance of MTA. However, this performance is not statistically different from the MTA performance when tested with a Wilcoxon signed-rank test ( $P=0.2154286$ ). Upon closer inspection this appears to be due to that there exist cases where MeTAL was giving good predictions while MTA did not, and also among the cases where both MTA and MeTAL gave better than top 10% performance, the ground truth reaction was ranked higher by MeTAL than MTA. In total 34 out of the 58 datasets can have their ground truth reactions predicted among the top 10% in either MTA or MeTAL. Therefore, it appears that further efforts should seek to combine MTA and MeTAL in an informed manner to achieve even better prediction performances.

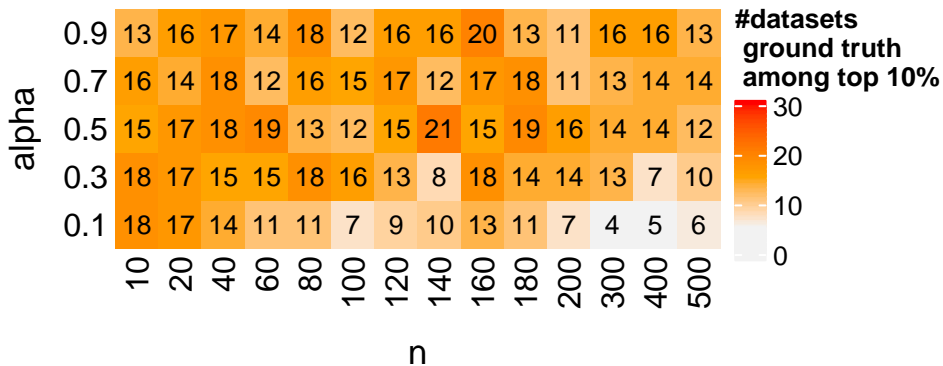


Figure 2.3: Evaluation of the prediction performance of MeTAL on 58 human datasets using different parameters.

## 2.2.4 Second proposed alternative to the original MTA and its evaluation

Another potential issue in the design of the original MTA is a more biologically informed one. Specifically, the original MTA essentially evaluates the capacity of a reaction KO in allowing or supporting the transformation in the desired direction. Nevertheless, there is no guarantee that the system will actually shift in the specific desired direction upon the reaction KO. In theory, it is possible that the same KO can support the transformation in another independent direction equally well, or even better, but MTA does not assess the capacity of the KO model in supporting the transformation in any other arbitrary and irrelevant direction. In other words, the MTA prediction does not consider the specificity of the KO in terms of realizing the desired transformation, and therefore in theory a reaction KO with a high MTA score is not necessarily a good candidate for realizing the desired transformation. This same issue also applies to the MeTAL method as described above. To address this issue, conceptually we can run MTA (or MeTAL) repeatedly using different random transformations (i.e. random sets of  $R_S$ ,  $R_F$ , and  $R_B$  reactions) for each reaction KO, then compare the resulting distribution of MTA scores (or optimal MeTAL objective function values) to that obtained using the actual desired transformation. However, to reduce the computational burden, a most simple approach is to only run an additional MTA (or MeTAL) for each KO using the opposite transformation as desired. Specifically, after mapping DE results via APR to intended directions of changes in the activity levels of the reactions, reverse the directions before using

them to define the  $R_F$ ,  $R_B$ , and  $R_{rr}$  (for MeTAL only) reaction sets. The  $R_S$  set should remain the same. Then the final “corrected” MTA score (or MeTAL objective function value) is the value obtained from the desired transformation minus that obtained from the opposite transformation. I termed this modified procedure mMTA for “multiple” MTA, or mMeTAL when applied to MeTAL.

Next, evaluating the performance of mMTA and mMeTAL on the 58 validation datasets, I found that both mMTA and mMeTAL showed large improvement compared to MTA and MeTAL, respectively. mMTA achieved “better than top 10%” performance in up to 35 out of the 58 datasets under optimal parameter combinations, which interestingly involves very small  $n$  and  $\epsilon_i$  values (Figure 2.4). mMeTAL showed particularly drastic improvement in performance compared to MeTAL, with improvement in almost all parameter combinations tested (Figure 2.5), in the best case recovering the ground truth among top 10% predictions in 34 datasets. The best parameter for mMeTAL also involves the choice of very small  $n$  value ( $n=10$ ). mMeTAL thus can achieve nearly equivalent performance with mMTA, while being more computationally efficient due to its formulation as an LP problem rather than an MIQP problem.

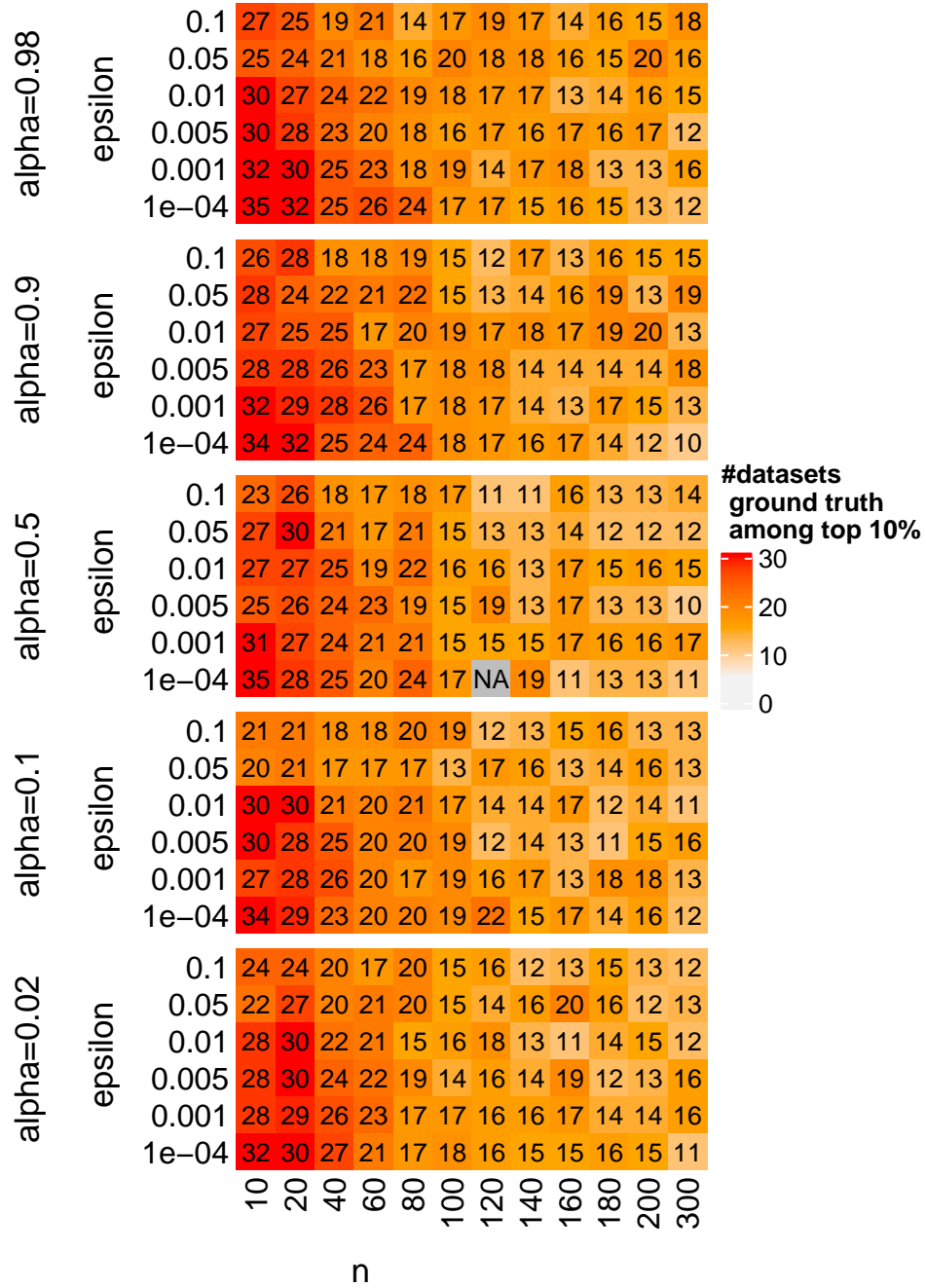


Figure 2.4: Evaluation of the prediction performance of mMTA on 58 human datasets using different parameters.

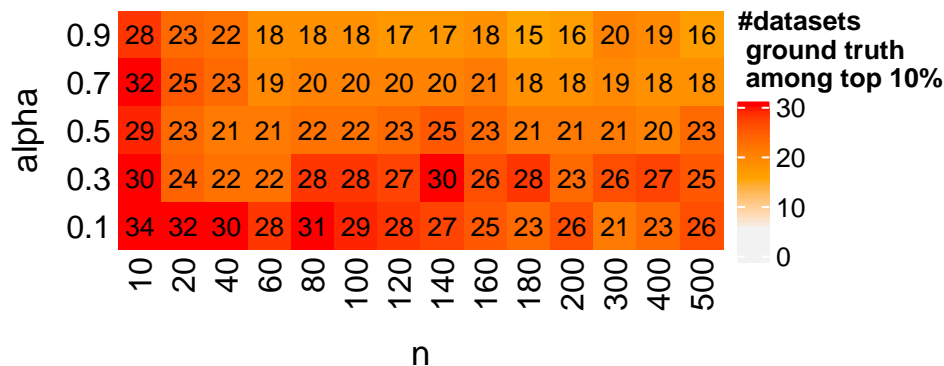


Figure 2.5: Evaluation of the prediction performance of mMeTAL on 58 human datasets using different parameters.

## 2.2.5 Comparison with the rMTA algorithm and a benchmark of all MTA variants

During my effort in evaluating and improving the original MTA algorithm, (Valcárcel et al., 2019) published another MTA improvement which they termed rMTA (for “robust MTA”) and showed that it has better predictive accuracy than the original MTA from (Yizhak et al., 2013). However, their validations were also restricted to mostly *E. coli* datasets, making it necessary to comprehensively benchmark all different MTA variants with the human datasets I collected. rMTA is based on exactly the same idea of mMTA in addressing the specificity of the KO in terms of realizing the desired transformation, where the original MTA is run twice, one in the desired direction of transformation and the other in the opposite direction, resulting in two MTA scores which they termed *bTS* and *wTS* for “best transformation score” and “worse transformation score”, respectively. Then instead of simply using  $bTS - wTS$  as the final score, they compared *bTS* and *wTS* in a

more refined manner. Specifically, they reasoned that only when a KO has  $bTS > 0$  and  $wTS < 0$  can we be certain that the KO is a good candidate, and all the other cases should be regarded as unresolved. In these unresolved cases, they proposed to use an independent MOMA-based algorithm to determine the capacity of the KO to realize the desired transformation. As described in Section 1.2.4, MOMA (Segrè et al., 2002) is an algorithm to predict the flux distribution after a metabolic reaction KO based on the assumption of minimal metabolic adjustment, i.e. the system will prefer to adapt to the new, more constraint feasible metabolic space under the KO by making only the smallest metabolic adjustment possible. Accordingly, MOMA predicts the flux distribution post-KO by minimizing the euclidean distance between the post-KO flux vector to that of the wildtype (WT) or starting condition. Adapted to this setting, the starting condition is defined by the  $\mathbf{v}^{\text{ref}}$  vector, and the MOMA optimization is as follows:

$$\underset{\mathbf{v}}{\text{minimize}} \|\mathbf{v} - \mathbf{v}^{\text{ref}}\|_2^2$$

subject to:

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

$$\mathbf{v}^{\text{lb}} \leq \mathbf{v} \leq \mathbf{v}^{\text{ub}}$$

$$v_j = 0$$

Then an MTA score corresponding to the MOMA optimal solution is computed in the same way as in the original MTA, this score was termed  $mTS$  for “MOMA transformation score”, and was used for the unresolved cases as the final score.

Besides, to ensure that the resolved cases will get higher final scores to reflect the high confidence levels associated with those predictions, MOMA was also performed for these resolved cases and the final score for these resolved cases is computed as  $k \cdot mTS \cdot (bTS - wTS)$  where  $k$  is a fixed large number such as 100 (according to the authors of (Valcárcel et al., 2019)) to ensure that these resolved reaction KO's get higher scores.

I next comprehensively evaluated rMTA on the 58 human datasets I collected, and compared it to all the other MTA variants. Consistent with the claimed superior performance, rMTA indeed performed better than the original MTA, but its performance is very similar to mMTA (as well as mMeTAL), successfully predicting the “ground truth” reaction among top 10% predictions in as many as 35 out of the 58 datasets. Surprisingly, unlike mMTA, the rMTA performance was not seen to be affected much at all by the change in the MTA parameters  $\alpha$  and  $\epsilon_i$  (Figure 2.6). Since in rMTA, the score from MOMA is independent of these MTA parameters and only dependent on the number of top DE genes ( $n$ ), this suggests that the improved performance of rMTA may be largely attributed to the adoption of the independent MOMA algorithm. I therefore evaluated the predictive performance of the MOMA algorithm, i.e. based on the  $mTS$  score alone. Indeed, MOMA achieved almost identical performance to rMTA (Figure 2.7), although with the best parameter combination in rMTA there is one more dataset with its “ground truth” reaction among top 10% predictions compared to MOMA. Comparing the predicted percentage ranks of the “ground truth” reactions across datasets, I found that the correlation between the ranks of rMTA and MOMA is almost perfect, while the cor-

relation between the ranks of rMTA and mMTA is much weaker (Appendix Figure 4.15). This suggests that the improvement of rMTA is largely due to the adoption of MOMA in its pipeline. This was not pointed out by the authors of rMTA (Valcárcel et al., 2019) and not demonstrated in their testing.

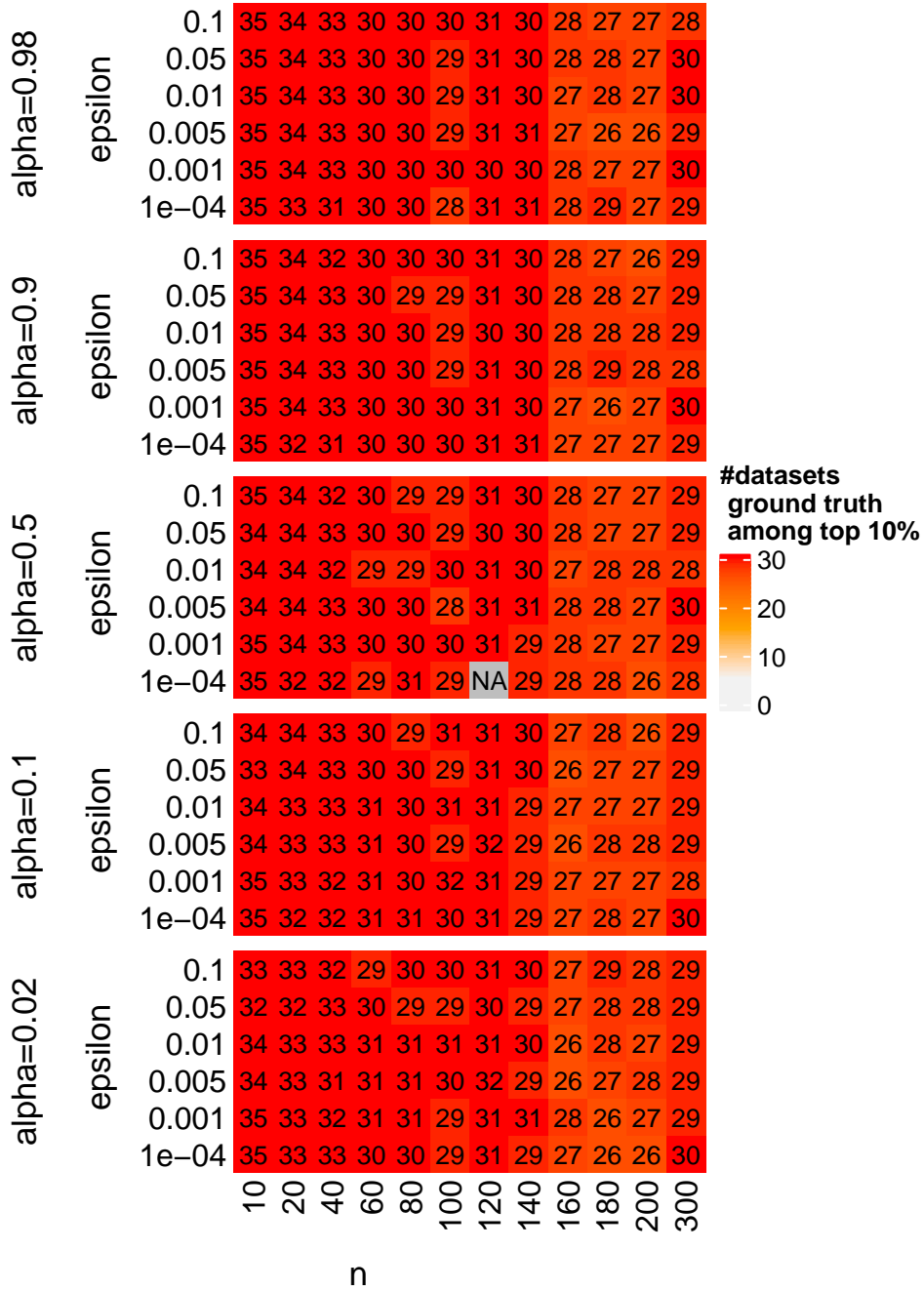


Figure 2.6: Evaluation of the prediction performance of rMTA on 58 human datasets using different parameters.

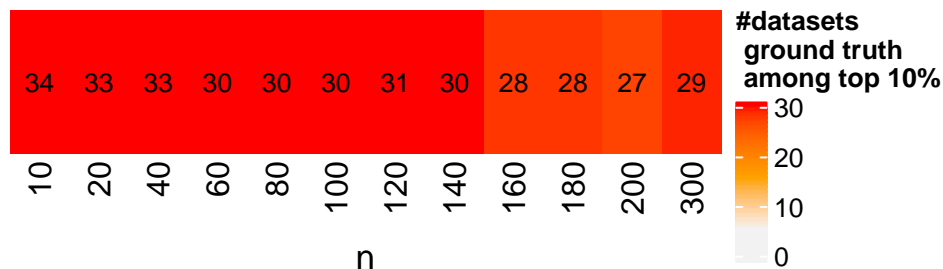


Figure 2.7: Evaluation of the prediction performance of MOMA on 58 human datasets using different parameters.

In Figure 2.8, I summarized the predicted percentage ranks of the ground truth reactions across all the 58 datasets using the various algorithms (MTA, mMTA, rMTA, MOMA, MeTAL, mMeTAL). The X-axis datasets are ordered by the predicted percentage ranks of rMTA, which is mostly highly similar to MOMA except in three cases, two where MOMA performed drastically worse while the other where MOMA performed drastically better than rMTA. We see that in most cases where rMTA/MOMA performed good (i.e. achieving “better than top 10%” performance), mMTA also achieved very similar predicted percentage ranks. mMeTAL however, had lower predicted percentage ranks in many cases, despite still being among the top 10%. MTA and MeTAL on the other hand had more fluctuation in performance, sometimes dropping well below top 25%. Interestingly though, in the datasets where rMTA performed poorly, MTA and/or MeTAL can have much better predicted percentage ranks. In the future it may be desirable to develop informed workflows for combining multiple algorithms in order to achieve even better overall predictive performance.

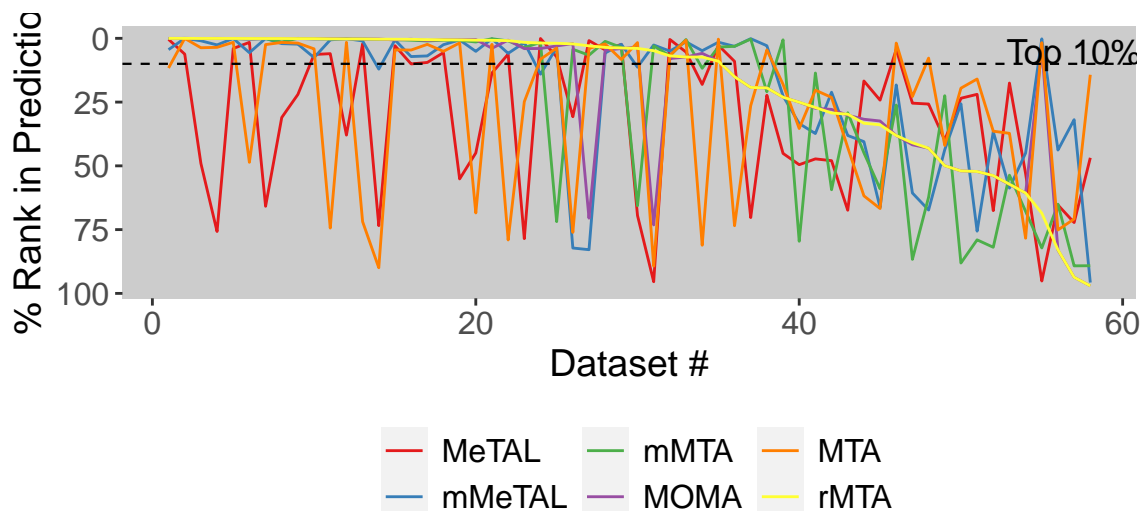


Figure 2.8: Summary of the predicted percentage ranks of the ground truth reactions across all the 58 datasets using the various algorithms with their optimal parameters.

## 2.2.6 Comparison of the metabolic modeling-based algorithms with gene expression-based analysis

From the various metabolic modeling-based prediction results, we see that they tend to produce the best performance when only the very few top DE genes (i.e. as small as  $n = 10$ ) were picked. Therefore the predictions appeared to be driven by top DE genes. When further assigning higher weight to the top changes by scoring the MOMA solution using the MeTAL objective function, while setting high  $\alpha$  values, the best-case prediction performance is increased even further (36/58 datasets have ground truth within top 10% predictions; Figure 2.9). These observations raise two important questions: 1. how well the ground truth genes (i.e. the actual genes being KO/KD-ed or proteins being inhibited) can be directly predicted from the DE results, how does it compare to the metabolic modeling-based predictions, and

whether the latter approach provides additional value; 2. whether the performances exhibited by the different MTA variants, especially rMTA/MOMA, were completely attributed to that the ground truth genes are among the top DE genes used as input to these metabolic modeling algorithms.

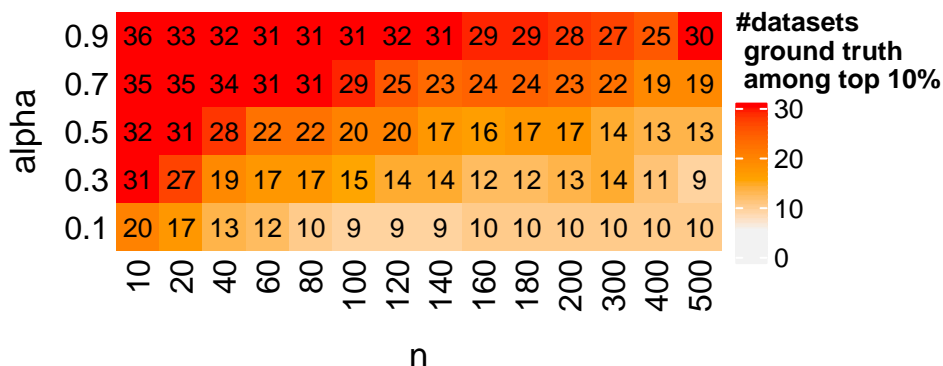


Figure 2.9: Evaluation of the prediction performance of MOMA on 58 human datasets using the MeTAL objective function with different parameter values.

To investigate the first question, I examined the percentage ranks (by DE log fold-change values) of the ground truth genes in each dataset, among only the metabolic genes present in the GEM to ensure comparability to the prediction performances obtained from the modeling algorithms. I found that in 49 out of the 58 datasets, the ground truth genes are actually among the top 10% metabolic DE genes. This number is much higher than that achieved by the various metabolic modeling-based algorithms. However, for the other 9 datasets where the ground truth gene is not among the top 10% metabolic DE genes (it happened that in these datasets the ground truth genes are all ranked >20% by DE log fold-change), the various MTA variants can successfully recover the ground truth among the top 10% or even top 5% predictions in as many as 6 out of the 9 datasets (Figure

2.10). This suggests that the modeling-based approach can identify the true perturbed gene/reaction in cases where the DE results do not provide direct clue on the ground truth, and it is in these cases that the modeling-based approach is a valuable complement to the DE-based analysis.

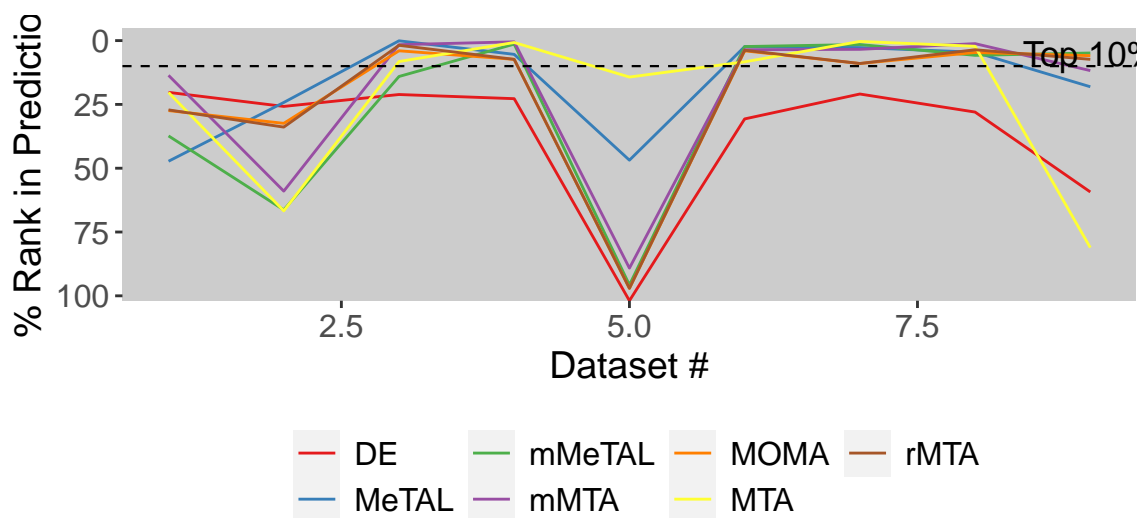


Figure 2.10: Summary of the predicted percentage ranks of the ground truth reactions across all the 9 datasets where the ground truth gene is not among the top 20% DE genes, using the various algorithms with their optimal parameters.

To investigate the second question, I performed control analysis where the ground truth gene was removed from the DE result given as the input to the various MTA variants, so that the modeling algorithms cannot directly rely on the information of the change in the ground truth gene/reaction. In this analysis, the performance of all algorithms decreased. In the optimal parameter setting, both MOMA and mMeTAL showed “better than top 10%” performance in 23 datasets, and for rMTA, 24 datasets (Appendix Figures 4.16, 4.17, 4.18). This suggests that in many datasets, the accurate prediction is dependent on the ground truth gene

present among the top DE genes passed to the metabolic modeling algorithms, although this is not needed for the algorithms to make successful predictions in still a significant number of datasets.

## 2.3 Materials and Methods

### 2.3.1 Validation datasets

A total of 58 human transcriptome datasets involving the KD or KO or drug (small molecule) inhibition of a metabolic gene/protein were manually collected from the GEO database, starting with search by keywords include “knockout”, “knock-down”, “inhibition”, “antagonist” and the name or symbol of a metabolic gene (obtained from the human GEM Recon 1 (Duarte et al., 2007) and Recon 3D (Brunk et al., 2018)). These datasets span different transcriptome profiling platforms including different microarrays and bulk RNA-seq. Each dataset contains the expression profiles of the untreated or vehicle-treated control samples and those of the treated (i.e. KO/KD/drug inhibition) samples. The detailed information on the collected datasets are provided in Table 4.2.

### 2.3.2 Generating the common inputs to different MTA algorithms

For each dataset, DE analysis between the treated and control samples were performed. The edgeR package (Robinson, McCarthy, & Smyth, 2010) was used for RNA-seq datasets and the limma package (Ritchie et al., 2015) with LOESS normalization was used for microarray datasets. log<sub>2</sub>-transformed TPM values (for

RNA-seq datasets) or log<sub>2</sub>-transformed LOESS-normalized expression values (for microarray datasets) of the control samples (usually <5 replicates) were averaged across samples (i.e. arithmetic mean) and used as the representative gene expression profile of the control group. This is passed as the input to iMAT algorithm (Shlomi et al., 2008) together with the base human GEM Recon 1 (Duarte et al., 2007), which produces a customized model of the input samples. ACHR is then used to uniformly sample the resulting model until convergence by visual inspection (this usually involves the sampling of up to 1e4 sample points or more in rare cases for Recon 1). The sample (arithmetic) mean was then computed to obtain the flux vector of all metabolic reactions in the model representative of the control condition. This flux vector as well as the DE results of treated vs control samples for each validation dataset were used as the common inputs to the different MTA algorithms described in the text above.

### 2.3.3 Implementation of the different MTA algorithms

The mathematical details of each of the algorithms were described in the text above. All algorithms were implemented in R version 3.6.3 using our in-house R package named `gembox` developed by me, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver.

### 2.3.4 Evaluation and comparison of the different MTA algorithms

For each validation dataset, the true metabolic gene being KO-ed/KD-ed (or proteins being inhibited) was mapped to metabolic reaction(s) (termed “ground truth” reactions). The different algorithms were evaluated based on whether they were able to successfully recover at least one of the mapped “ground truth” reactions among the top 10%-20% predictions across all metabolic reactions. The performance of different algorithms were compared by their predicted percentage ranks of the “ground truth” reactions across validation datasets. In cases where the gene/protein is mapped to multiple “ground truth” reactions, the percentage rank of the reaction that ranked top (i.e. highest MTA score, lowest rank value) was used.

### 2.3.5 Softwares and code

R version 3.6.3 was used for all statistical tests. P values lesser than  $2.22e-16$  may not be computed accurately and are reported as “ $P < 2.22e-16$ ”. The Benjamini-Hochberg (BH) method was used for P value adjustment. The R packages ggplot2 (Wickham, 2009) was used to create the visualizations. Implementation of the new MTA algorithm can be found in the GitHub repository: <https://github.com/ruppinlab/gembox>.

## 2.4 Discussion and Conclusion

In this chapter, I described my efforts in comprehensively validate the original MTA algorithm (Yizhak et al., 2013) in a large collection of human datasets, while

trying to develop improved algorithms for the same problem of predicting metabolic KO's that facilitate a specific metabolic transformation. Such a problem is broadly present, for example, in the prediction of therapeutic target for many metabolism-related diseases, and in the identification of causal metabolic factors underlying the difference between two biological conditions. Many such applications are in the context of human biology and medicine, and thus the development and evaluation of improved MTA algorithm specifically with human data can be highly valuable.

The original MTA method involves solving a series of MIQP problems which require the specification of multiple parameters, notably  $\alpha$  for balancing the integer and quadratic parts of the optimization, and  $\epsilon_i$  for defining altered reaction fluxes. Although default parameters showed to produce reasonable performances were recommended in (Yizhak et al., 2013), they were based on relatively limited testing and validation on mostly simpler organisms like *E. coli* and yeast. It can be expected that the optimal choice of these parameters may be dependent on the organism (i.e. GEM used) and the dataset. This indeed appeared to be the case during the validation with the 58 human datasets I collected, where the prediction performance was seen to vary a lot depending on the choice of parameters, and the optimal parameter combination obtained via a grid search procedure show deviation compared to the default values recommended in (Yizhak et al., 2013). With the validation datasets where the ground truths are known, it is possible to fine-tune the algorithm parameters, but in real datasets and application cases, the tuning can be challenging due to the lack of ground truths. Therefore, an alternative MTA formulation with minimal parameterization, or robustness with regard to the

choice of parameter can be desirable. Accordingly, I formulated the MeTAL, an efficient LP version for the same prediction problem as in MTA that does not include the  $\epsilon_i$  parameter and with a natural default choice for the  $\alpha$  parameter. MeTAL also properly handles the “overshoot” reactions within the objective function and does not require a subsequent rescoring procedure using the optimal solution like in the original MTA, thus without the issue of degenerate solutions. Unfortunately, MeTAL showed weaker prediction performance compared to MTA, although it appeared to be complementary to MTA in some datasets. MeTAL and MTA share the issue of lack of specificity in measuring the KO of a reaction to facilitate exactly the intended direction of transformation. Since they only consider the single direction corresponding to the intended transformation, the resulting scoring cannot reflect whether a KO can indeed result in the intended transformation (but not, say, another random transformation). After taking this into consideration with a very minor modification of the prediction pipeline using essentially the same algorithm, both mMTA and mMeTAL show improved performance, although at the cost of doubling of computation time. Fortunately, the LP problem in mMeTAL is very efficient to solve compared to the MIQP problem of MTA, making mMeTAL a very feasible algorithm.

Further, with the more comprehensive human-based validation, I confirmed that the new rMTA algorithm (Valcárcel et al., 2019) showed robust and stronger performance than the other MTA variants, although mMeTAL achieved equivalent performance with the best parameter. Surprisingly, the superior performance of rMTA was not due to the methodical pipeline involving an mMTA-like procedure,

but can be mostly attributed to the adoption of MOMA. Since MOMA directly predicts the likely flux distribution after a reaction KO, it elegantly avoids the need to consider non-specific transformations. Besides, MOMA doesn't contain any additional parameter. The prediction performance of MOMA alone is equivalent to rMTA.

Driven by the observation that restricting the number of top DE genes passed to the metabolic modeling algorithm, as well as assigning higher relative weight to the reactions that are intended to be changed (i.e.  $R_F$ ,  $R_B$ , and  $R_{rr}$ ) resulted in better prediction performance of many of the algorithms described, it was further investigated to what extent the performances of the algorithms were attributed to the ground truth gene being among the top DE genes. I found that although in many datasets, this is required for the achieving good prediction performance, it is not necessary for a significant number of cases – with the metabolic modeling approach, the ground truth gene/reaction can be recovered in many cases where the gene is not a top DE gene among the metabolic genes. Therefore the metabolic modeling approach can be a valuable complement to the DE analysis in identifying the relevant targets or causal genes/reactions.

Based on the benchmark results, rMTA, MOMA showed equally well performance. Although given the optimal parameter, mMeTAL also yielded the same performance in term of the number of datasets where the ground truth reaction is among the top 10% predictions, MOMA has no parameter to specify or tune other than the number of top DE genes, and its QP optimization problem only needs to be solved once. Therefore MOMA can be used as an effective algorithm for the

metabolic transformation problem. In the subsequent chapters, I will demonstrate the application of MOMA as an effective “metabolic transformation algorithm” to different contexts, showing that it can help to generate useful biological insights as well as promising candidate therapeutic targets for diseases. I will use “MTA” to refer to the algorithm for the metabolic transformation problem in general, but also the MOMA version as one such algorithm with validated good performance in human.

## Chapter 3: Metabolic modeling of CD8 T cell for anti-cancer immunotherapy<sup>1</sup>

### 3.1 Introduction

Various cancer immunotherapies, notably ICB and ACT therapies have revolutionized the landscape of cancer therapeutics, achieving unprecedented durable response. However, various challenges still remain, including the increasing the currently suboptimal response rate in ICB therapy and enhancing the long-term persistence of transfused T cells in ACT therapies (Sections 1.3.2 and 1.3.3). T cells can play a pivotal role in anti-cancer immunity, and their functional state is an important determinant of response to current cancer immunotherapies. It is therefore desirable to further elucidate the factors regulating T cell function in order to improve the success of the immunotherapies. It has been established that metabolism can greatly affect different aspects of T cell function, including cytolytic activity, differentiation, longevity and exhaustion (Section 1.3.4), with mitochon-

---

<sup>1</sup>The work of this chapter is a collaborative project with Dr. Madhusudhanan Sukumar in Dr. Nicholas Restifo's lab at NCI, who conducted the biological experiments investigating the findings based on computational predictions. All biological experiment results included are credited to Dr. Madhusudhanan Sukumar, after obtaining the permission of him and Dr. Nicholas Restifo. The contents of this chapter have been written into a manuscript on which I and Dr. Madhusudhanan Sukumar are co-first authors and will be submitted to a journal. Many contents of this chapter are taken from the above manuscript, involving direct quotes with minor rephrasing.

drial metabolism lying at the center of T cell metabolic regulation (Desdín-Micó, Soto-Heredero, & Mittelbrunn, 2018). Modulating T cell mitochondrial metabolism has further been proposed to represent a promising strategy to boost immunotherapy response (Li & Zhang, 2020). Here I aim to take advantage of the framework of GEM integrated with publicly available T cell transcriptomic data in cancer immunotherapy to understand the global metabolic features associated with T cell function and immunotherapy response. I specifically applied the MTA algorithm developed from previous algorithms (Valcárcel et al., 2019; Yizhak et al., 2013) described in Chapter 2 on various ACT datasets to predict metabolic processes that are essential for T cell function in the ACT setting. Combining with the analysis of additional ICB datasets and the cancer patient data from the TCGA project, I identified mitochondrial uncoupling and specifically the UCP2 protein as important for therapy response. Via collaboration with experimentalists, the role and mechanisms of UCP2 in regulating T cell function and immunotherapy response was further comprehensively investigated in the mice B16 melanoma model. The extensive biological experiments confirmed that UCP2 is required for T cell stemness and affects its anti-tumor function in mice, and also further elucidated an ROS and mTOR signaling-related mechanisms, echoing previous studies that established the important roles of ROS and mTOR [Klein Geltink et al. (2018); Powell et al. (2012); Section 1.3.4].

## 3.2 Results

### 3.2.1 Identification of the metabolic fluxes in CD8 T cells associated with cancer immunotherapy response

Numerous studies have performed transcriptomic profiling of patient samples in the context of cancer immunotherapy in order to identify the determinants of response. Aiming to specifically investigate the CD8 T cell metabolic profiles in immunotherapy, I mainly focused on ACT datasets where the T cells to be infused into the patients were specifically profiled instead of the bulk tumor, and applied GEM to compute the metabolic fluxes from the gene-level data. Specifically, I start by analyzing two published gene expression datasets: one dataset of CAR-T cells for anti-CD19 CAR-T therapy in CLL (Fraietta et al., 2018), and the other of KRAS(G12D)-targeting TILs for ACT in a metastatic colorectal cancer patient (Lu et al., 2019). For (Fraietta et al., 2018), I applied the GEM algorithm iMAT (Shlomi et al., 2008) with ACHR random sampling to compute the metabolic flux distributions in the CAR-T cells of the responders and non-responders to therapy, respectively, and performed differential flux analysis of the responders compared to the non-responders. In (Lu et al., 2019), the authors identified a T cell clone that persisted for up to 266 days after infusion, which they called “9.2-P”, and another clone targeting the same KRAS(G12D) epitope but did not persist for longer than 40 days, which is called “9.1-NP”. I similarly used iMAT to compute the flux distributions of these two T cell clones, and compared the fluxes in the persistent clone compared to the non-persistent clone. Metabolic pathways enriched by these two comparisons are highly similar, suggesting common metabolic processes associated

with T cell persistence and response. These common pathways are summarized in Figure 3.1. Specifically examining the fluxes corresponding to ETC/OXPHOS function, which are known to be positively associated with T cell memory, stemness and persistence (Sukumar et al., 2013), It can be seen that the iMAT-computed fluxes are overall higher (esp. for the ATP synthase reaction) in responders compared to non-responders (Figure 3.2) as well as in the persistent clones vs the non-persistent clones (Appendix Figure 4.19).

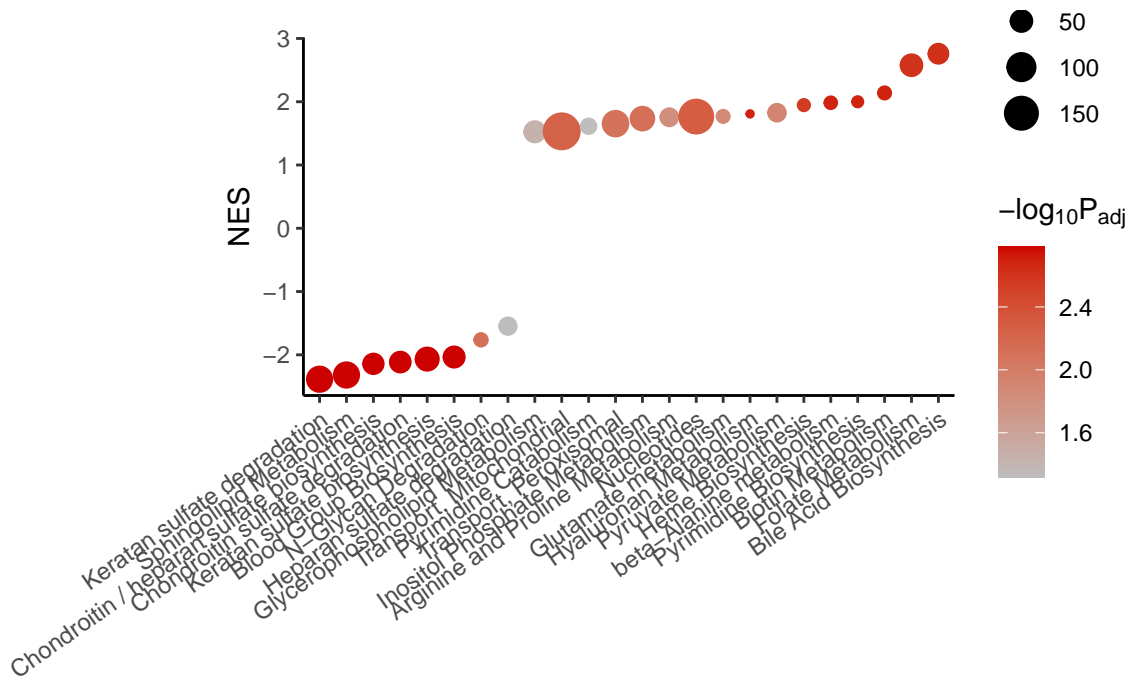


Figure 3.1: Metabolic pathway enrichment of reactions with consistent differential fluxes in the CAR-T cells of responders vs non-responders and in the persistent vs non-persistent T cell clones.

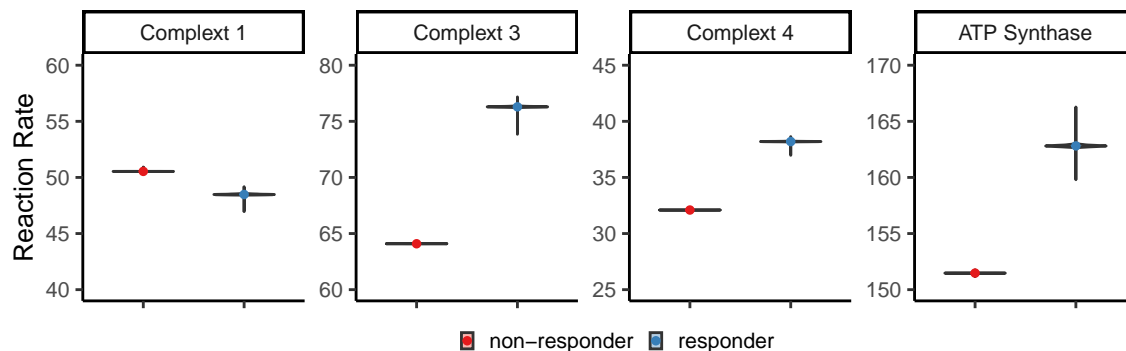


Figure 3.2: Computed metabolic flux distributions of the ETC/OXPHOS reactions in the T cells of anti-CD19 CAR-T therapy responders vs non-responders.

### 3.2.2 MTA identified T cell mitochondrial uncoupling as essential for the anti-cancer function and immunotherapy response

Although the above differential flux analysis results provide additional information than what can be obtained from only the gene expression level by integrating the metabolic network structure, the findings still represent the factors associated with T cell persistence and response, instead of causal factors. Trying to gain stronger evidence of causality and to identify the metabolic processes that are likely to causally determine the response, I applied the MTA algorithm as described and developed (based on previous algorithms (Valcárcel et al., 2019; Yizhak et al., 2013)) in the previous Chapter (Chapter 2). Here I first focused on the prediction of metabolic reactions whose KO can transform the T cell metabolic state from that of the responders to that of the non-responders using the anti-CD19 CAR-T therapy dataset (Fraiatta et al., 2018). Applying the algorithm as such, the predicted metabolic reactions represent those that are essential for T cell anti-cancer function,

since their KO's are predicted to lead to non-responsiveness to CAR-T therapy.

The metabolic pathways enriched by the top 10% MTA-predicted reactions are shown in Figure 3.3. Some of the discovered pathways are known to be important in regulating T cell function, such as nucleotide and amino acid metabolism, as well as OXPHOS (reviewed in Section 1.3.4). The most significant pathway enriched by top MTA predictions is “Transport, Mitochondrial” (Fisher’s exact test adjusted  $P < 0.05$ ). I therefore next focused on this pathway, aiming to identify the specific metabolite whose mitochondrial transportation plays a key role by performing another enrichment analysis by each metabolite (results summarized in Figure 3.4). Interestingly, the top enriched metabolite is  $H^+$  (i.e. proton; Fisher’s exact test adjusted  $P < 0.1$ ). I also confirmed the enrichment for mitochondrial transport, specifically proton transport reactions with the gene set enrichment analysis method specifically was also confirmed with the GSEA method ( $FDR < 0.1$ ; Appendix Figures 4.20 and 4.21). Moreover, I also applied MTA on the ACT data from (Lu et al., 2019) and predicted metabolic reactions whose KO transforms the metabolic state from that of the 9.2-P persistent clone to that of the 9.1-NP non-persistent clone, as an independent analysis to identify metabolic processes essential for T cell persistence. The top predictions in this latter analysis were also enriched for mitochondrial transport reactions as well as specifically mitochondrial proton transport ( $FDR < 0.1$ ; Appendix Table 4.4). These results suggest that the mitochondrial proton transport reactions are important for T cell function and CAR-T therapy response.

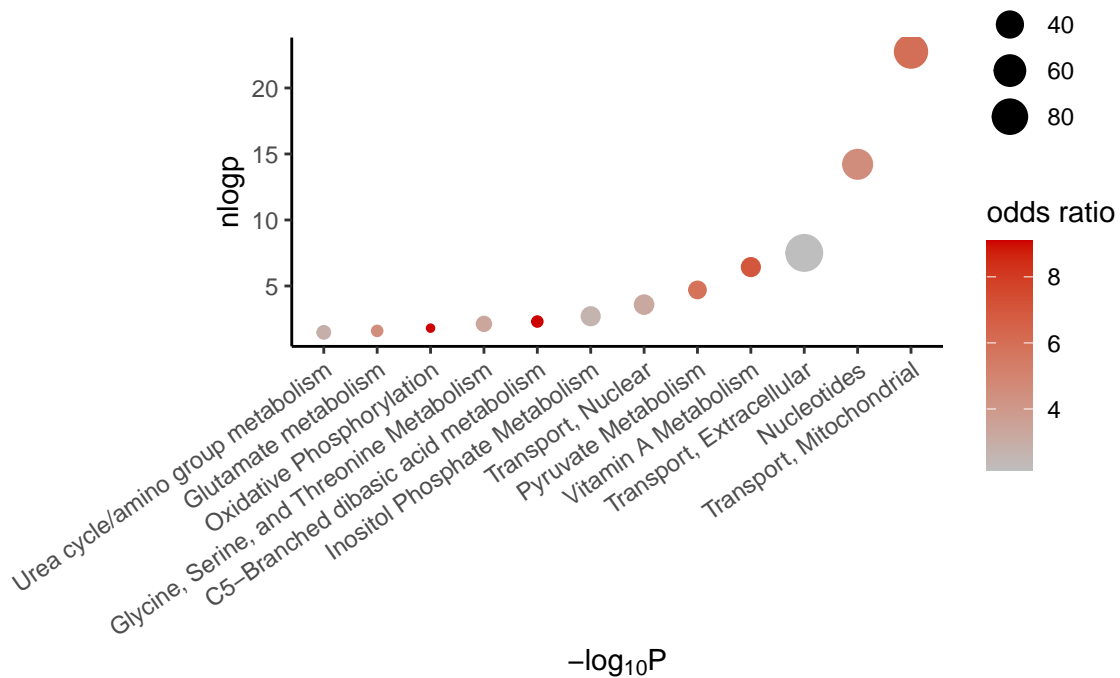


Figure 3.3: Metabolic pathway enrichment of the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy.

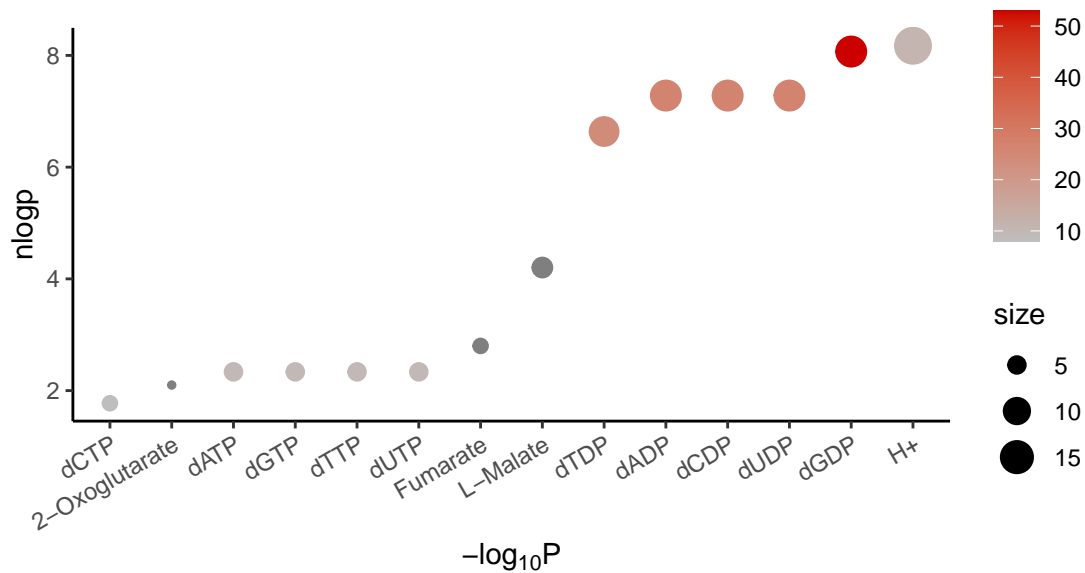


Figure 3.4: Enrichment for mitochondrial transport reactions by each metabolite by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy.

I next aimed to determine whether there exists a particular direction of mito-

chondrial proton transport (import or export) that is essential for T cell anti-cancer function. For this, I identified the reactions responsible for higher mitochondrial proton import in the anti-CD19 CAR-T responders than in non-responders, as well as the other sets of proton transport reactions showing the opposite trend. Comparing the MTA score of these two sets of reactions, I found that the former set has significantly higher MTA scores (Wilcoxon rank-sum test  $P=0.00018$ ; shown in Figure 3.5). Furthermore, the responders show both higher total mitochondrial proton influx as well as the mitochondrial uncoupling reaction flux (i.e. the key reaction for proton transport into the mitochondrial matrix), than the non-responders (Figures 3.6 and 3.7). These differences are also seen when comparing the persistent vs non-persistent T cell clones from the anti-KRAS(G12D) (Lu et al., 2019) TIL dataset (Appendix Figure 4.22). These results suggest that mitochondrial uncoupling is an important regulator of T cell function and persistence in ACT therapies.

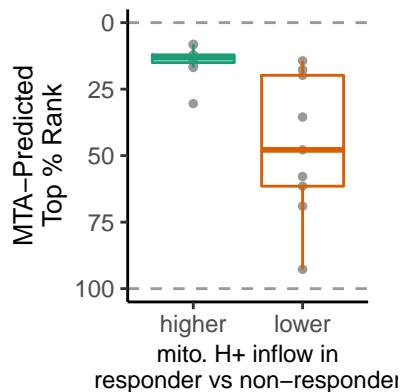


Figure 3.5: Comparing the MTA score ranks of the mitochondrial proton transport reactions contributing to higher mitochondrial proton influx in responders vs non-responders and those contributing to higher mitochondrial proton efflux in responders vs non-responders

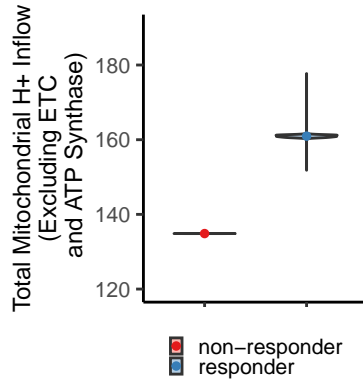


Figure 3.6: Total mitochondrial proton influx in the T cells of the non-responders and responders to anti-CD19 CAR-T therapy.

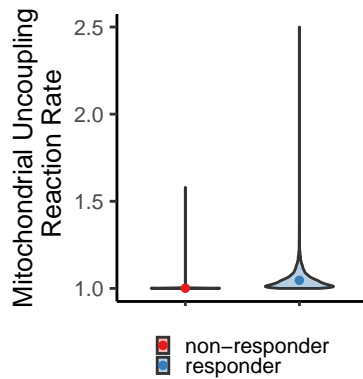


Figure 3.7: Mitochondrial uncoupling reaction rate in the T cells of the non-responders and responders to anti-CD19 CAR-T therapy.

3.2.3 Analysis of large-scale human transcriptome datasets identified that mitochondrial uncoupling protein *UCP2* expression is associated with CD8 T cell stemness and ICB response.

There are three major mitochondrial uncoupling proteins (UCPs): *UCP1*, *UCP2*, and *UCP3*, located in the inner mitochondrial membrane close to ETC protein complexes (Krauss, Zhang, & Lowell, 2005). Via the analysis of gene expression data across different types of human tissues from the Human Protein Atlas (Uhlén

et al., 2015). We see that among the three genes *UCP2* has much higher expression levels than *UCP1* and *UCP3* in the PBMC, thymus and spleen (Figure 3.8), suggesting that *UCP2* is the major form of mitochondrial uncoupling protein in immune cells. Further investigating the association between *UCP2* and T cell differentiation markers on the gene expression level in the TCGA data, I found that in many cancer types there is a positive correlation between *UCP2* expression and a several T cell memory or stemness marker genes (including *CD8A*, *TCF7*, *IL7R* and *SELL*) after controlling for tumor purity (Figure 3.9), which were not observed for *UCP1* or *UCP3*. Additionally, analyzing various transcriptome datasets on anti-PD1 and anti-CTLA-4 ICB therapies (Miao et al., 2018; Riaz et al., 2017; Van Allen et al., 2015), I found that *UCP2* expression levels tend to be higher in the responders than non-responders, although these expression data are from bulk tumor samples (Appendix Figures 4.23). These results taken together support the role of *UCP2* in the anti-cancer function of CD8 T cells.

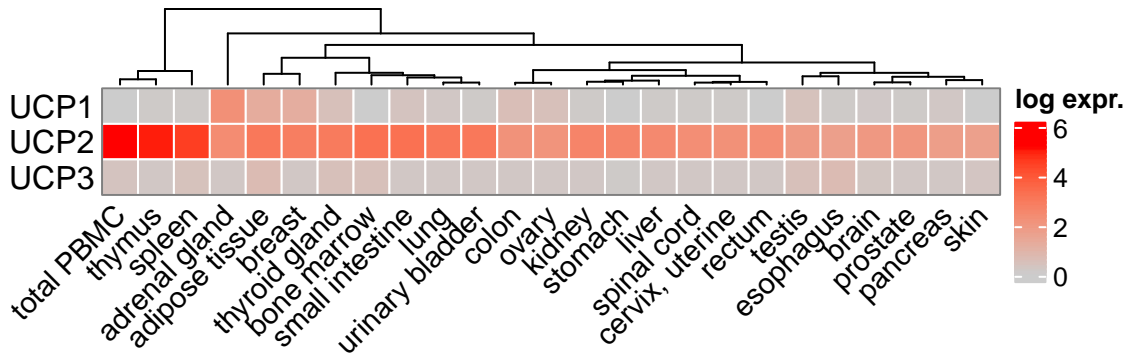


Figure 3.8: Expression levels of UCP1, UCP2 and UCP3 across human tissues.

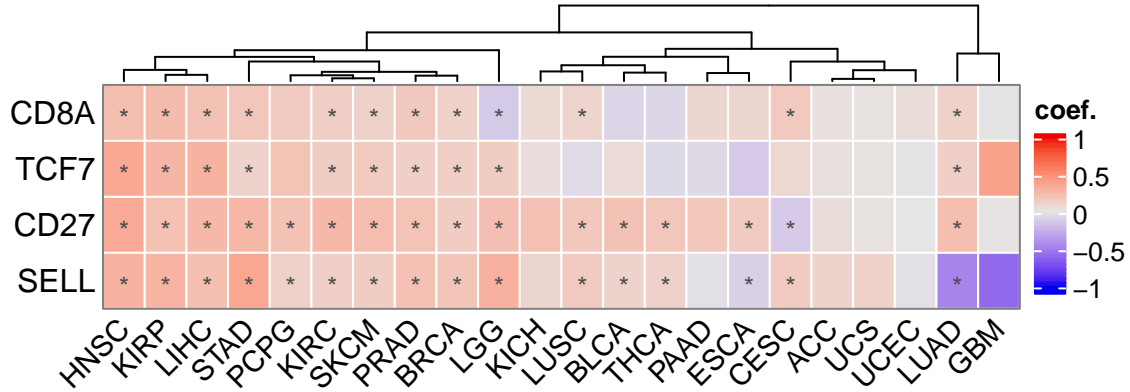


Figure 3.9: Correlation of UCP2 expression and the expression of several T cell stemness and memory marker genes across TCGA cancer types after controlling for tumor purity.

### 3.2.4 Ucp2 inhibition in mice T cells decreased T cell stemness marker expression and reduced anti-cancer function

To validate that *Ucp2* is indeed required for T cell anti-tumor function, our experimental collaborators performed experiments in mice Pmel-1 CD8<sup>+</sup> T cells and inhibited Ucp2 with the small molecule genipin (which was shown to be a Ucp2 inhibitor (Zhang et al., 2006)). With RNA-seq profiling of control T cells without genipin-treatment and the treated cells then performing DE analysis between the treated vs control groups, it was found that Ucp2 inhibition increased the expression of marker genes for effector T cell differentiation (*Eomes*) and T cell effector function (*Gzma*), but also increased the level of T cell exhaustion genes (*Havcr2* (*Tim3*), *Lag3*, *Cd244a* (*2B4*) and *Cd160*). On the other hand, T cell stemness and memory marker genes (*Tcf7*, *Lef1*, *Klf2*, *Sell*, *Il7r*, *Ccr7*) decreased in their expression. A summary of the DE results for these key genes are given in Table 3.1.

id	log.fc	log.cpm	F	pval	padj
Eomes	0.7180357	6.460565	435.995847	0.0000000	0.0000000
Gzma	0.7769064	10.329460	1037.916240	0.0000000	0.0000000
Havcr2	0.4058764	7.601680	216.363680	0.0000000	0.0000000
Lag3	-0.0853685	7.583349	8.445215	0.0076070	0.0076070
Cd244a	1.2143334	6.797390	1349.135131	0.0000000	0.0000000
Cd160	2.5506975	5.520364	1931.031535	0.0000000	0.0000000
Tcf7	-1.1238619	2.276961	63.845740	0.0000000	0.0000000
Lef1	-0.2645372	5.775450	40.509084	0.0000012	0.0000013
Klf2	-1.1470205	3.330084	141.973879	0.0000000	0.0000000
Sell	-0.6293411	6.814459	421.345849	0.0000000	0.0000000
Il7r	-1.1276077	7.027800	1489.676890	0.0000000	0.0000000
Ccr7	-2.5531972	5.990828	2932.843349	0.0000000	0.0000000

Table 3.1: DE results of selected T cell differentiation and function marker genes in genipin-treated vs control mice Pmel-1 CD8 T cells.

GSEA analysis confirmed that the genipin-treated cells (compared to the control cells) is enriched for several gene sets associated with effector T cell function and T cell exhaustion (Figure 3.10). Furthermore, the increase in effector function and exhaustion marker genes, and the decrease in T cell memory marker genes were confirmed by flow cytometry performed by our experimental collaborators (Appendix Figure 4.24). These results suggest that Ucp2 inhibition in T cells damped memory T cell differentiation and promoted T cell exhaustion.

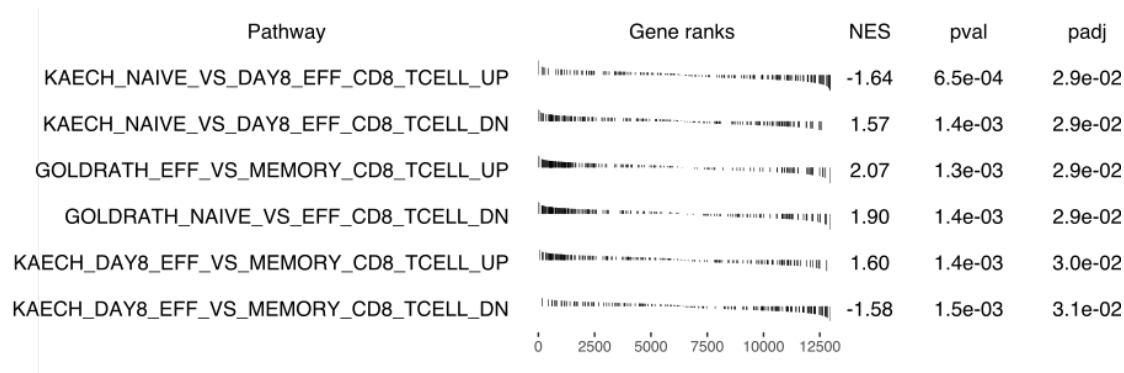


Figure 3.10: GSEA table and plots showing the enrichment of various T cell effector function and T cell exhaustion-associated pathways in genipin-treated vs control mice Pmel-1 CD8 T cells.

In order to further validate whether *Ucp2* can affect the actual anti-tumor functions of the T cells, our experimental collaborators performed ACT therapy in the B16 mice melanoma model with wildtype (WT) CD8 T cells compared to CD8 T cells with *Ucp2* inhibition. Specifically, they adoptively transferred either vehicle-treated Pmel-1 CD8<sup>+</sup> T cells or genipin-treated T cells into mice bearing subcutaneous syngeneic, large vascularized established B16 melanomas. It was found that with genipin-treated Pmel-1 CD8<sup>+</sup> T cells, the tumors had a faster growth rate compared to the tumors treated with WT T cells (Figure 3.11). This demonstrates that *Ucp2* is indeed essential for T cell function and can determine its efficacy in ACT therapy.

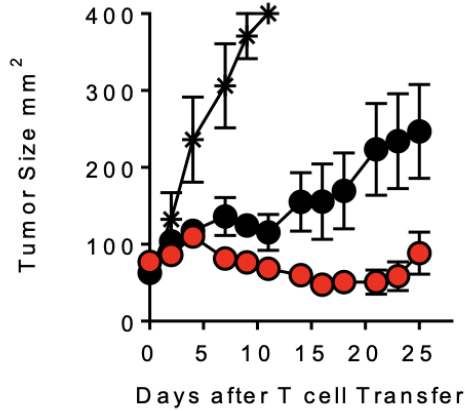


Figure 3.11: Tumor growth in B16 mice melanoma model treated by ACT of vehicle-treated CD8 T cells or genipin-treated CD8 T cells. Experimental groups: stars (untreated); red circles (WT T cells); black circles (genipin-treated T cells). Figure credit: Dr. Madhusudhanan Sukumar

### 3.2.5 GEM modeling supports increased reactive oxygen species as a mechanism underlying the effect of *Ucp2*

Next, we seek to understand the mechanisms underlying the effect of *Ucp2* on T cell function. Given that *Ucp2* is a mitochondrial uncoupling protein that regulates the ETC/OXPHOS efficiency which is also coupled to mitochondrial reactive oxygen species (ROS) production, we decided to investigate whether the effect of *Ucp2* can be mediated by ROS, as ROS has been known to play important roles in regulating T cell function (Chandel, 2015; Mehta, Weinberg, & Chandel, 2017). First starting with a computational approach, I used iMAT as described above to model the metabolic flux distributions in the genipin-treated Pmel-1 CD8 T cells and compared it to control cells. Since GEM does not explicitly model metabolite concentrations and is based on the steady state assumption, I defined the total flux

through a metabolite as the absolute value of its total production rate (which is equal to the absolute value of its total consumption rate under the steady state assumption), which measures the level of activity in the metabolism of the metabolite. I computed the total flux through each ROS species, including hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) and superoxide anion ( $\text{O}_2^-$ ) in both the mitochondria and the cytosol. Increased fluxes through both mitochondrial and cytosol  $\text{O}_2^-$  and  $\text{H}_2\text{O}_2$  were observed in genipin-treated T cells compared to control (Figure 3.12). Although not a direct measure of ROS concentrations, this suggests that the ROS production, esp. mitochondrial  $\text{O}_2^-$  that is directly coupled to ETC, is increased upon *Ucp2* inhibition, consistent with the role of *Ucp2* as a uncoupling protein.

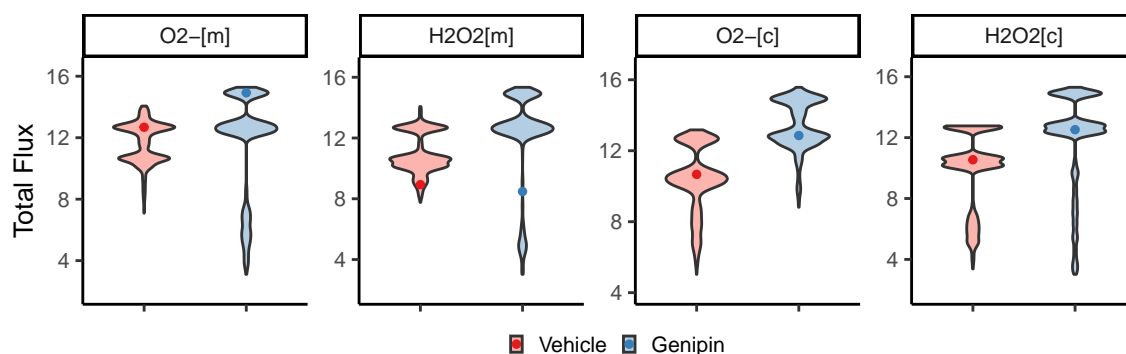


Figure 3.12: Total metabolic fluxes through mitochondrial and cytosol hydrogen peroxide and superoxide anion in genipin-treated Pmel-1 CD8 T cells vs vehicle-treated T cells. O<sub>2</sub>-: superoxide anion; H<sub>2</sub>O<sub>2</sub>: hydrogen peroxide; suffix ‘[m]’ means mitochondrial, ‘[c]’ means cytosol. The dot in each group denotes the mean of each distribution.

Furthermore, again taking advantage of the ability of GEM modeling in suggesting causality, I applied MTA to predict metabolic perturbations (KO’s) that can reverse the genipin-treated metabolic state back to the control metabolic state. Conceptually, inhibiting these reactions will counteract the effect of genipin treat-

reaction.set	odds.ratio	pval	reaction.set.size	overlap.size
ROS producing reactions	2.335643	0.0250698	30	11

Table 3.2: Enrichment of top MTA predictions whose KO can reverse the effect of genipin in ROS producing reactions

reaction	equation	top.percent
CATm	2 Hydrogen peroxide[m] $\rightarrow$ 2 H <sub>2</sub> O H <sub>2</sub> O[m] + O <sub>2</sub> O <sub>2</sub> [m]	4.401503
ASCBOX	L-Ascorbate[c] + 2 H <sup>+</sup> [c] + 2 Superoxide anion[c] $\rightarrow$ Dehydroascorbate[c] + 2 Hydrogen peroxide[c]	8.051530
GTHPi	2 Reduced glutathione[c] + Hydrogen peroxide[c] $\rightarrow$ Oxidized glutathione[c] + 2 H <sub>2</sub> O H <sub>2</sub> O[c]	8.964037

Table 3.3: The predicted percent rank of several cellular antioxidation reaction in the MTA prediction of reactions whose KO can mimic the effect of genipin treatment

ment, thus they may mediate the effect of Ucp2 inhibition by genipin. I found that the top 20% MTA predictions are enriched for the reactions that can generate ROS species (including those producing H<sub>2</sub>O<sub>2</sub> and O<sub>2</sub><sup>-</sup> in both the mitochondria and the cytosol, but excluding the reactions that catabolize O<sub>2</sub><sup>-</sup> into H<sub>2</sub>O<sub>2</sub>, e.g. superoxide dismutase reactions; Fisher’s exact test P=0.025, odds ratio=2.34; Table 3.2). Conversely, when applying MTA to predict metabolic KO’s that can mimic the effect of genipin treatment, I found that several of the key cellular antioxidation reactions (including the catalase reaction, and those involving ROS scavenging by ascorbate and reduced glutathione) are highly ranked (Table 3.3). Together these computational results suggest that the increase in ROS production may mediate the effect of Ucp2 inhibition by genipin.

### 3.2.6 The antioxidant N-acetylcysteine suppresses the increased ROS level in mice CD8 T cell due to *Ucp2* inhibition and reversed T cell function

To evaluate the role of ROS in the effect of *Ucp2* on T cell function, our experimental collaborators activated vehicle or genipin-treated mice Pmel-1 CD8<sup>+</sup> T cells in the presence or absence of the antioxidant N-acetylcysteine (NAC). The mitochondrial ROS levels in the different experimental groups were measured with the MitoSOX fluorescent probe, and it was confirmed that genipin-treatment drastically increased mitochondrial ROS levels, while NAC can suppress the genipin-induced increase in ROS to a near baseline level (Figure 3.13). RNA-seq was performed on the samples from the different experimental groups. DE analysis comparing to the genipin+NAC group to the genipin-treated group (without NAC) showed that NAC reversed the genipin-induced changes in T cell effector, stemness and memory genes (Table 3.4). Specifically, the marker genes associated with stemness and memory (*Tcf7*, *Klf2*, *Foxo1*, *Id3*) increased upon NAC treatment, and some effect and exhaustion marker genes (*Gzma*, *Gzmb*, *Cd244a*, *Cd160*) decreased. These changes are confirmed on the pathway level with GSEA analysis (results for the relevant gene sets summarized in Figure 3.14).

id	log.fc	log.cpm	F	pval	padj
Gzma	-2.7647913	9.8826336	5197.711201	0.0000000	0.0000000
Gzmb	-0.3526219	12.9988944	200.908923	0.0000000	0.0000000
Cd244a	-0.6641592	7.0051687	223.517442	0.0000000	0.0000000
Cd160	-1.3821411	5.7791014	850.073528	0.0000000	0.0000000
Tcf7	3.4544769	4.1911891	1080.386506	0.0000000	0.0000000
Klf2	0.6631848	3.0356692	37.806585	0.0000212	0.0000236
Foxo1	0.9260563	6.1033783	384.148111	0.0000000	0.0000000
Id3	0.7663993	0.5688859	8.271411	0.0118178	0.0118178
Il7r	0.9493588	6.9268148	401.398854	0.0000000	0.0000000
Ccr7	2.4381694	5.9134907	1372.733633	0.0000000	0.0000000

Table 3.4: DE results of selected T cell differentiation and function marker genes in genipin+NAC-treated vs genipin-treated mice Pmel-1 CD8 T cells.

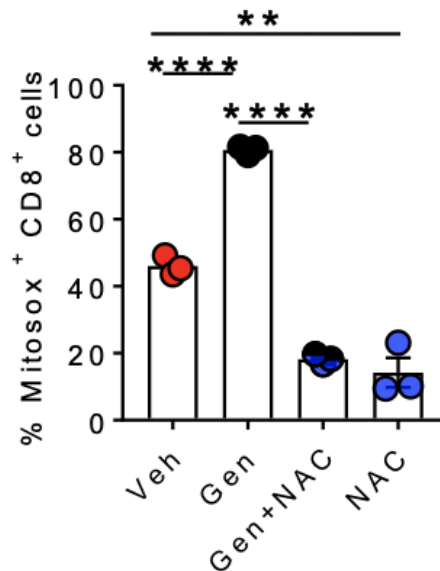


Figure 3.13: Mitochondrial ROS levels as measured by MitoSOX in genipin-treated mice Pmel-1 CD8 T cells, with or without NAC. Experimental groups: Veh (vehicle); Gen (genipin-treated); Gen+NAC (genipin+NAC-treated); NAC (NAC-treated). Figure credit: Dr. Madhusudhanan Sukumar

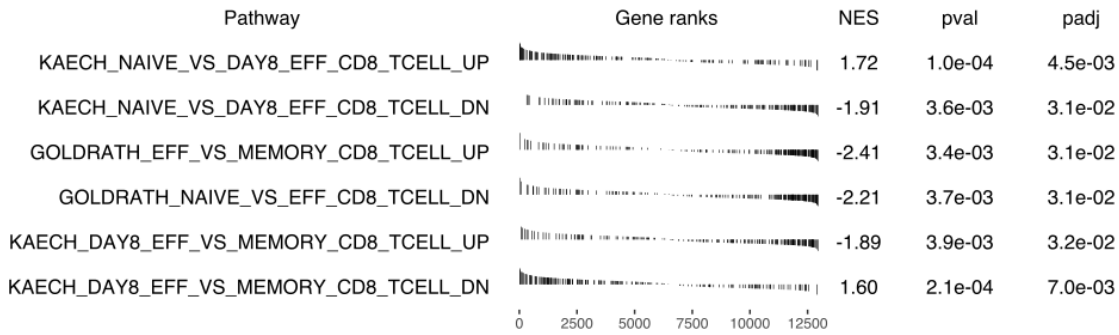


Figure 3.14: GSEA table and plots showing the enrichment of various T cell effector function and T cell exhaustion-associated pathways in genipin+NAC-treated vs genipin-treated (without NAC) mice Pmel-1 CD8 T cells.

Finally, to evaluate whether NAC can reverse the decreased anti-tumor function of the CD8 T cells with Ucp2 inhibition, our experimental collaborators adoptively transferred the genipin-treated Pmel-1 CD8<sup>+</sup> T cells cultured either with or without NAC into wild-type mice bearing subcutaneous syngeneic, large vascularized established B16 melanomas. It can be seen that NAC significantly rescued the impaired anti-tumor function of the genipin-treated T cells; NAC treatment (without genipin) also further improved the efficacy in tumor control compared to WT untreated CD8 T cells (Figure 3.15). In summary, these results consolidate that the effect of Ucp2 on T cell anti-cancer function is mediated by ROS.

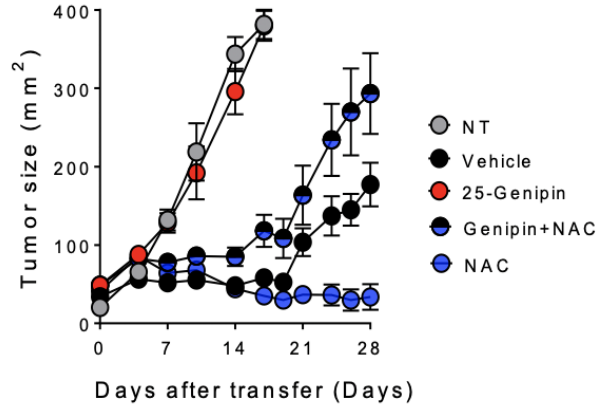


Figure 3.15: Tumor growth in B16 mice melanoma model treated by ACT of genipin-treated CD8 T cells or genipin+NAC-treated CD8 T cells. NT: non-treated; 25-Genipin: treatment with 25 micro molar genipin. Figure credit: Dr. Madhusudhanan Sukumar

### 3.3 Materials and Methods

#### 3.3.1 Differential expression and gene set enrichment analysis

Differential expression (DE) analysis of responders vs non-responders of various cancer immunotherapy datasets, as well as for the mice T cells between different experimental conditions (control, NAC and/or genipin treatment) were performed with edgeR (Robinson et al., 2010). For gene set enrichment analysis (GSEA) (Subramanian et al., 2005) from the DE log fold-change results, an implementation of the GSEA algorithm in the R package fgsea (Korotkevich et al., 2021) was used, with gene set annotations were taken from the MSigDB database (Liberzon et al., 2011). Specifically, the “C7 immunologic signature gene sets” from MSigDB v7.0 was used to analyze the differences in T cell activation or memory functions across the mice T cell samples treated by NAC and/or genipin.

### 3.3.2 Computation of metabolic fluxes from gene expression data with GEM and differential flux analysis

For each dataset, we used the GEM algorithm iMAT (Shlomi et al., 2008) to compute the metabolic flux profile from gene expression data. iMAT requires gene-length normalized expression values in the bulk RNA-seq datasets, for this we used TPM or RPKM values. Then for each dataset, we took the median expression values of the samples from each group/experimental condition as their representative expression profile, and used it as input to iMAT. We applied iMAT to model the median expression in each group or experimental condition to reduce data noise; this is also due to that iMAT is not efficient in modeling metabolic flux from the similar samples within each group. The human GEM Recon 1 (Duarte et al., 2007) was used for modeling human data in the various cancer immunotherapy datasets, while the mouse model iMM1415 (Sigurdsson et al., 2010) is used for modeling mice T cell data (i.e. genipin/NAC treatment data). The output of iMAT is a refined GEM for the each of the groups/experimental conditions in each dataset, with metabolic reaction bounds adjusted to achieve maximal concordance with the gene expression data while satisfying the stoichiometric constraints of the cellular metabolic network (Shlomi et al., 2008). Each output model defines a space rather than a single unique solution of the global metabolic flux profile, and ACHR was used to sample the metabolic space and obtain the flux distribution of each metabolic reaction in each condition and dataset. The resulting flux distributions are smooth (rather than discrete with replicates like experimentally collected data), and are meant to

be the representative flux distribution for each group/condition; the variation in the distribution reflects uncertainty in the computation and flexibility in the metabolic space, rather than variation across samples. Therefore, subsequent comparisons of the flux distributions between two groups/conditions were carried out in a qualitative manner by plotting and visually inspecting the smooth flux distributions and not by statistical tests (as these distributions are equivalent to population distributions rather than sample distributions). All GEM analyses were performed using our in-house R package named `gembox`, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver.

### 3.3.3 Genome-scale metabolic modeling-based prediction of metabolic functions determining response to anti-CD19 CAR-T therapy

Gene expression data of the anti-CD19 CAR-T cells for responders and non-responders was obtained from (Fraietta et al., 2018). The median expression profiles across all responders and all non-responders were computed respectively, then representative metabolic flux distributions of the responders were computed as described in the above section. The DE genes in the non-responders compared to the responders were also identified as described above. These were then passed as inputs for the GEM-based MTA algorithm (described in Chapter 2, based on Yizhak et al., 2013; Valcárcel et al., 2019) to predict metabolic reactions whose knock-out can transform cellular metabolic state from that of the responders to that of the non-responders. The output of MTA is a score (MTA score) for each metabolic

reaction, with higher scores corresponding to better candidates for achieving the metabolic transformation as specified above. From our previous experience (Yizhak et al., 2013), the top 10-20% MTA predictions contain promising targets, and were used for downstream analysis. The MTA algorithm implemented in our in-house R package named `gembox` was used, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver.

### 3.3.4 Metabolic pathway analysis of the MTA-predicted metabolic reactions

The “subSystems” data slot from the human Recon 1 genome-scale metabolic model (Duarte et al., 2007) was used as metabolic pathway annotations for each metabolic reaction (i.e. each “sub system” is treated as one metabolic pathway). Fisher’s exact test was used to test for significant enrichment of each metabolic pathway in the top 10% reactions predicted by MTA as described above. P values were corrected for multiple hypotheses with the Benjamini-Hochberg method. Further, to closely analyze the enrichment for mitochondrial transport reactions in a metabolite-specific manner, the transport reactions across the mitochondrial membrane for each metabolite were identified from the Recon 1 model (the model does not distinguish between the inner and outer membrane of the mitochondria), and these were used as reaction sets. The enrichment of top 10% MTA-predicted reactions in these reaction sets were tested with Fisher’s exact test similarly as above. As an independent approach, the GSEA method (Subramanian et al., 2005) was also

used to test for the enrichment of top MTA-predicted reactions in the “Transport, Mitochondrial” pathway and specifically the set of mitochondrial proton transport reactions. For this the pathway/reaction set was defined as above, and all metabolic reactions ranked by their MTA score was passed to the GSEA algorithm to compute the result.

### 3.3.5 Genome-scale metabolic modeling-based prediction of metabolic processes that possibly mediates the effect of Ucp2

The DE analysis between the control (vehicle-treated) mice Pmel-1 CD8<sup>+</sup> T cells and the genipin-treated CD8<sup>+</sup> T cells, as well as the computation of the metabolic fluxes in these two experimental conditions were performed as described above. The MTA algorithm as above (described in Chapter 2, based on Yizhak et al., 2013; Valcárcel et al., 2019) was applied to predict metabolic reactions whose knock-out can transform cellular metabolic state from that of the genipin-treated cells back to that of the control cells. The top predictions from this analysis may represent those that can mediate the effect of Ucp2 inhibition on the metabolic level.

### 3.3.6 Analysis of the expression levels of mitochondrial uncoupling genes across human tissues or in various ICB datasets

The gene expression profile data for different types of human tissues were obtained from the Human Protein Atlas database (Uhlén et al., 2015). The “RNA consensus tissue gene data” was used, and *UCP1*, *UCP2* and *UCP3* expression levels

across all the available human tissue types were examined. Several datasets on the transcriptomes of tumor samples from anti-PD-1 and anti-CTLA-4 ICB therapies were obtained from their respective studies, including (Miao et al., 2018; Riaz et al., 2017; Van Allen et al., 2015). The *UCP2* expression levels in the bulk tumor samples between the responders and non-responders to therapy were compared via the DE analysis as described above.

### 3.3.7 Analysis of the correlation between *UCP2* and several T cell function/stemness marker gene expression across TCGA cancer types

The RNA-seq data of the TCGA samples were downloaded from the UCSC Xena database (Goldman et al., 2020) as read counts, which were then normalized across samples within each cancer type with the TMM method from edgeR (Robinson et al., 2010) and log-transformed. The phenotypic data including tumor purity of each sample was also obtained. The *CD8A* gene as the marker gene for CD8<sup>+</sup> T cell and *TCF7*, *CD27*, *SELL* as marker genes of T cell stemness or memory were analyzed. Within each cancer type and for each marker gene, the association between its expression and *UCP2* expression was analyzed with a linear model with tumor purity as a covariate (independent variable).

### 3.3.8 Softwares and code

R version 3.6.3 was used for all statistical tests. P values lesser than  $2.22e-16$  may not be computed accurately and are reported as “ $P < 2.22e-16$ ”. The Benjamini-Hochberg (BH) method was used for P value adjustment. The R packages ggplot2 (Wickham, 2009) was used to create the visualizations. Code for the analyses can be found in the GitHub repository: [https://github.com/ruppinlab/Tcell\\_Ucp2](https://github.com/ruppinlab/Tcell_Ucp2).

## 3.4 Discussion and Conclusion

ACT therapy, especially CAR-T therapy has been a major promising development among the different cancer immunotherapies under clinical investigation in the recent years, achieving highly durable response in many patients of mainly hematological cancers, and in some cases essentially cured the patients (Brentjens et al., 2013; Grupp et al., 2013; Porter et al., 2011). CD8 T cell has been recognized as a major player among the various mechanisms that the immune system utilizes to control and kill cancer, and is currently at the center of ACT therapy (Styczyński, 2020). Especially, the persistence and longevity of the infused CD8 T cells in patients have been shown to determine the efficacy of therapy (Jiang et al., 2019). Besides in ACT therapy, the functional state of T cells has also been shown to be correlated with ICB therapy response (Sade-Feldman et al., 2018; Van Allen et al., 2015). Strategies for increasing T cell stemness and longevity and for reducing T cell exhaustion is highly desirable for both ACT and ICB therapies and present a current challenge for the further improvement of these cancer immunotherapies

(Jiang et al., 2019; Meric-Bernstam et al., 2020). It is therefore important to further characterize the factors that regulate T cell function and differentiation.

As reviewed in Section 1.3.4, metabolism has been recognized to be intricately coupled to T cell signaling, which can in turn modulate the fates of T cell differentiation and different T cell functions. In this study, I adopted a metabolic modeling-driven approach aiming to identify novel metabolic processes that are important for T cell function. To this aim, I mainly relied on published ACT datasets, including the CAR-T cell expression data of an anti-CD19 CAR-T therapy in CLL (Fraietta et al., 2018), and the KRAS(G12D)-targeting TIL expression data of an ACT therapy in a metastatic colorectal cancer patient (Lu et al., 2019). The expression profiling in these datasets was performed specifically for the T cells that would be infused into the patients, and the T cells are known to harbor TCR specific to certain tumor antigens presented by the patients' tumors. Therefore, the differences detected in responders vs non-responders (or persistent clones vs non-persistent clones) are more likely to represent T cell-intrinsic properties (despite that properties of the tumor can also have an impact). In comparison, the characteristics of the tumor could play a relatively larger part in determining the response in ACT therapy, and may represent extra variables to control for (Le et al., 2017; Mok et al., 2019; Samstein et al., 2019). Besides, to isolate T cell-specific expression in the tumors treated by ACT, single-cell RNA-seq is needed. Given our focus on the T cells, I therefore used the ACT datasets for the main analyses, while several ICB datasets were used to verify the findings.

A key idea of the modeling-based approach of this study is that, via the

MTA prediction of metabolic processes whose inhibition (KO) can result in a non-responder-like T cell metabolic state, metabolic properties essential to T cell anti-tumor function that are likely to have a causal role can be identified. The top pathway enriched for the MTA-predicted targets was mitochondrial transport, which echoes the prior knowledge on the important roles of mitochondrial metabolism in regulating T cell function (Desdín-Micó et al., 2018), and that modulating T cell mitochondrial function has been proposed as a promising strategy to improve immunotherapy response (Li & Zhang, 2020). In the further analysis, the mitochondrial proton transport reactions were found to be enriched by the top predictions. It needs to be noted that there are more than 20 reactions from the human Recon 1 model that involves the carrying of proton across the mitochondrial membrane (the Recon 1 does not distinguish between the inner membrane and outer membrane). Notably, all ETC reactions pump protons out from the mitochondrial matrix to build up the proton gradient for ATP synthesis by the ATP synthase, which utilizes the proton inflow back into the mitochondrial matrix. However, due to the special function of the ETC proton pumps and ATP synthase, these reactions were not included in the set of mitochondrial proton transport reactions in the analysis.

It was further determined that the inflow of proton into the mitochondrial matrix is important for T cell function. Mitochondrial proton inflow, known as mitochondrial uncoupling, is classically known for its function in dissipating the proton gradient across the mitochondrial inner membrane built up by the ETC activity to generate heat and modulate (reduce) ATP production (Demine, Renard, & Arnould, 2019), although it has further been found to have regulatory roles in

autophagy, mitophagy, cell death, ROS production and cell signaling (Demine et al., 2019). Notably, ROS is generated as a byproduct of ETC as part of the normal cell physiology, and mitochondrial uncoupling can therefore reduce ROS production due to the negative regulation of ETC function (Tahara, Navarete, & Kowaltowski, 2009), although this may be dependent on many factors of cellular conditions (Demine et al., 2019). Several uncoupling proteins are known to serve as the transporter for the proton inflow. For example, the brown adipose tissue (a major site of heat production in the body) is characterized by *UCP1* expression (Demine et al., 2019). *UCP1* has also been shown to regulate cell apoptosis during T cell development in the thymus (Adams, Kelly, & Porter, 2010). However, there has been mixed results in terms of whether the *UCP2* protein has uncoupling function (Berardi & Chou, 2014; Couplan et al., 2002), and it has been suggested that the uncoupling function of *UCP2* may be regulated and can be activated under specific conditions (Donadelli, Dando, Fiorini, & Palmieri, 2014). Regardless of its uncoupling function, nevertheless, *UCP2* has been shown to play various important roles in various cell types and contexts. For example, it can regulate mitophagy and protect the heart muscle from ischemic injury (Wu et al., 2019). It is also highly expressed in many different types of cancer and is associated with increased tumor proliferation and biosynthetic activities (Sreedhar et al., 2019). Recently, a study showed that the expression of *UCP2* in the cancer cells in melanoma sensitizes the tumor to anti-PD-1 ICB therapy (Cheng et al., 2019). There are also previous studies showing the role of *UCP2* in T cells. *UCP2* has been shown to be upregulated during T cell activation and proliferation (Krauss, Zhang, & Lowell, 2002; Rupprecht et

al., 2012). Further corroborating our findings that UCP2 has a role in regulating ROS production, it was previously shown that UCP2 inhibition with genipin in antigen-stimulated OT-1 CD8 T cells led to increased mitochondrial ROS, which was associated with increased apoptotic gene expression and reduced T cell expansion (Chaudhuri, Srivastava, Kos, & Shrikant, 2016). Interestingly, the authors also proposed that UCP2 can be modulated to boost ACT therapy efficacy, which we established in our study.

The widespread role of ROS in cell signaling and regulating cell death, senescence and differentiation has long been documented. In T cells, TCR signaling can promote mitochondrial activity and ROS generation, and ROS signaling has been shown to activate NFAT and induce IL-2 production (Sena et al., 2013). The role of ROS in T cell differentiation and On the other hand, high level of ROS, i.e. oxidative stress can lead to DNA damage, leading to cell senescence and cell death (Liochev, 2013), which has been shown in different types of stem cells to decrease their self-renewal (Chandel, Jasper, Ho, & Passegué, 2016; Suda, Takubo, & Senzema, 2011). This may also be implicated in the process of T cell differentiation (Franchina, Dostert, & Brenner, 2018), for example, it has been shown that less differentiated human CD8 T cells have higher antioxidation capacity than the more differentiated T cells, and NAC-treated stem-cell memory T cells (Tscm) exhibited improved long-term memory (Pilipow et al., 2018). In our study, such important role of ROS in T cell differentiation and stemness was also recaptured, and we further showed that the effect of UCP2 on T cell anti-cancer function was actually mediated by ROS, which is consistent with the previous reports on UCP2 as discussed above

(Chaudhuri et al., 2016). However, given the conflicting evidence on the extent of uncoupling function of UCP2, it remains to be established whether the regulation of ROS by UCP2 is mediated by the uncoupling function of the latter.

In conclusion, in this computation-driven study, I applied metabolic modeling to analyze a few datasets of CD8 T cell gene expression in the context of cancer immunotherapies, especially ACT therapy (including CAR-T). Via predicting metabolic perturbations that can transform cellular metabolic states between different samples, the MTA algorithm was able to first identify mitochondrial uncoupling as an essential process for T cell function, then predict that the underlying mechanism of the effect of mitochondrial uncoupling protein UCP2 inhibition is the production of ROS. Both predictions were then successfully validated by experiments with mice CD8 T cells *in vitro* and in the context of ACT therapy with the mice B16 melanoma model. This study demonstrates the value of a model-based approach in identifying relevant causal factors in the domain of cell metabolism, and establishes the role of UCP2 in T cell anti-tumor function via the regulation of ROS.

## Chapter 4: Modeling of host metabolism during SARS-CoV-2 infection<sup>1</sup>

### 4.1 Introduction

The COVID-19 pandemic has been ongoing for a year as of March 2021. Despite unprecedented global efforts to control the pandemic, this health crisis has not been resolved to date. While the wide deployment of vaccination is expected to significantly slow down the spread of the disease, the evolution and emergence of variant virus strains suggests a possible need to update the vaccines in the future (Z. Wang et al., 2021). Some studies predicted that there may be recurrent wintertime COVID-19 outbreaks over a longer term after the initial pandemic (Kissler, Tedijanto, Goldstein, Grad, & Lipsitch, 2020). Currently, effective therapeutic options are still very scarce, and thus the development of new treatment strategies are still urgently needed for long-term control of the pandemic (Section 1.4.3). Drug repurposing and combination may represent practical and time-efficient strategies to address this urgent unmet medical need. Given the recognized importance of host

---

<sup>1</sup>The work of this chapter has been written into a manuscript and submitted to the online preprint server bioRxiv (Cheng et al. (2021)); it has also been submitted to a journal, which is under peer-review as of Mar 2020. Many contents of this chapter are taken from the above article, involving direct quotes with minor rephrasing. This study involves biological experimental work performed by Dr. Laura Riva in Dr. Sumit K. Chanda's lab at SBP, with whom we collaborated.

metabolism in coronavirus (including SARS-CoV-2) infection and proliferation (Section 1.4.4), here I aim to identify anti-SARS-CoV-2 targets that act via counteracting the virus-induced metabolic changes, either individually or in combination with current anti-SARS-CoV-2 drugs, especially remdesivir. Such a metabolic-targeting approach may be especially valuable for anti-SARS-CoV-2 drug repurposing to address the current urgent COVID-19 crisis considering that many existing drugs are metabolism-targeting. My specific approach involves the application of GEM in an integrated analysis of multiple published *in vitro* and human patient gene expression datasets on SARS-CoV-2 infection, in order to comprehensively identify SARS-CoV-2 induced host metabolic reprogramming. Then I applied the MTA algorithm as described and developed in Chapter 2 (based on (Valcárcel et al., 2019; Yizhak et al., 2013)) to predict the metabolic anti-SARS-CoV-2 targets. The predictions were validated with the experimentally identified anti-SARS-CoV-2 drugs and targets reported in published drug or genetic screens.

## 4.2 Results

### 4.2.1 Analysis of SARS-CoV-2-induced gene expression changes across datasets

Taking advantage of multiple published studies on SARS-CoV-2 infection, I collected a total of 12 published relevant datasets involving different sample types including various cell lines, primary bronchial epithelial cells, nasopharyngeal swab and bronchoalveolar lavage fluid (BALF) from patients, and also based on different

Dataset Name	Sample Type	Sample Size	Platform	Reference
Vero	Vero E6 cell line	6	bulk RNA-seq	Riva et al. 2020
NHBE	Primary normal human bronchial epithelial cell	6	bulk RNA-seq	Blanco-Melo et al. 2020
A549	A549 human lung adenocarcinoma cell line with exogenous ACE2 expression	6	bulk RNA-seq	Blanco-Melo et al. 2020
Calu-3	Calu-3 human lung adenocarcinoma cell line	6	bulk RNA-seq	Blanco-Melo et al. 2020
293T	HEK293T human embryonic kidney cell line	12	bulk RNA-seq	Weingarten-Gabby et al. 2020
Caco-2	Caco-2 human colorectal adenocarcinoma cell line	6	MS-based proteomics	Bojkova et al. 2020
Swab.Butler	NP swab samples from human individuals	580	bulk RNA-seq	Butler et al. 2020
Swab.Lieberman	NP swab samples from human individuals	484	bulk RNA-seq	Lieberman et al. 2020
BALF	BALF from human individuals	6	bulk RNA-seq	Xiong et al. 2020
SC.Liao	BALF from human individuals (epithelial cells were used in analysis)	13	scRNA-seq	Liao et al. 2020
SC.Chua.Basal	NP and bronchial samples from human individuals (basal cells were used in analysis)	24	scRNA-seq	Chua et al. 2020
SC.Chua.Ciliated	NP and bronchial samples from human individuals (ciliated cells were used in analysis)	24	scRNA-seq	Chua et al. 2020

Table 4.1: Summary of the published gene expression datasets on SARS-CoV-2 infection analyzed in this study

assay platforms including bulk RNA-seq, scRNA-seq, and MS-based proteomics. These datasets are summarized in Table 4.1. I performed DE analysis of the SARS-CoV-2-infected samples vs the non-infected controls. For the single-cell datasets, only the airway epithelial cells are used, as they are known as the major virus-infected cell type. The PCA plot of the datasets based on the inverse normal-transformed DE log fold-change values is shown in Figure 4.1.

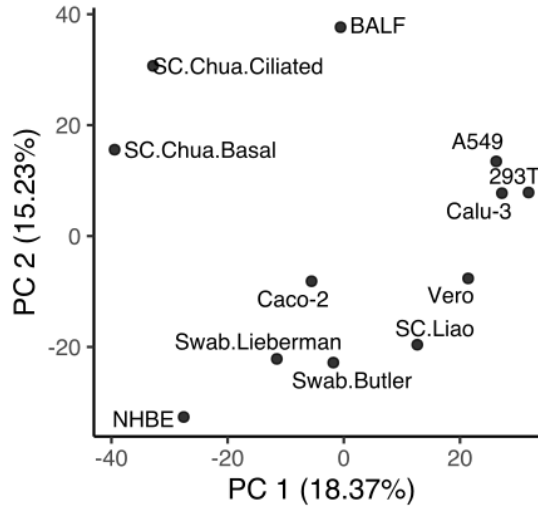


Figure 4.1: PCA plot of the datasets based on the inverse normal-transformed DE log fold-change values.

The PCA plot suggests that the cell lines tend to have distinct DE profiles from the patient samples, although different patient datasets of different sample

types and sequencing platforms exhibit considerable variation. Similar pattern is found by comparing each pair of datasets for their top significant DE genes (using Fisher’s exact tests, top n=400 DE genes from each datasets were used, all with at least FDR<0.1), as shown in Figure 4.2. Across all pairs of comparisons, the median odds ratio value is 2.18 and maximum value is 13.60 (adjusted P median 5.09e-5, minimum<2.22e-16), representing reasonable similarities across datasets.

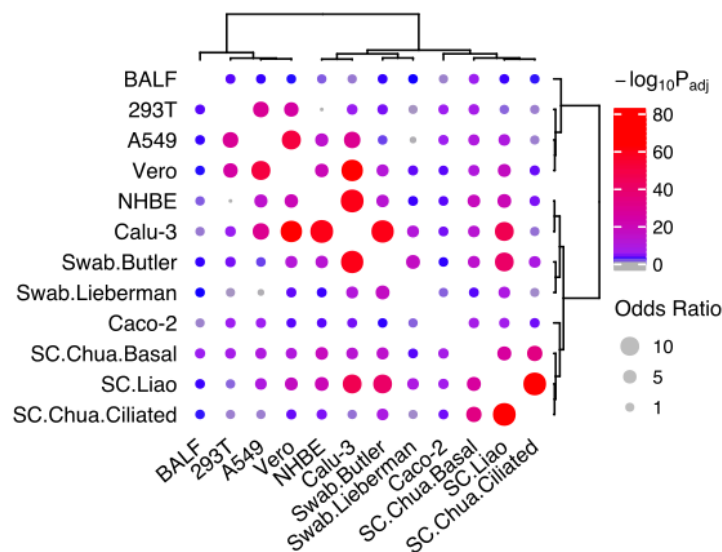


Figure 4.2: Pairwise comparison of the datasets by their top significant DE genes with Fisher’s exact tests.

I then performed GSEA analysis (Subramanian et al., 2005) based on the DE results of each dataset. The level of coherence across datasets on the pathway level appears stronger (Appendix Figure 4.25), with the median odds ratio from all pairwise comparison using Fisher’s exact test equals to 4.25 (maximum is infinity followed by 35.49), with adjusted P median 2.01e-5 (minimum<2.22e-16). No pathway is consistently significant across all 12 datasets. I identified the most consistently enriched pathways across the datasets, emphasizing the various

*in vivo* patient datasets, and the resulting pathways are visualized in Figure 4.3, Many up-regulated pathways are involved in innate immune response to viral infection, e.g. cytosolic DNA sensing and interferon signaling. Antigen presentation via MHC-I is seen coherently down-regulated in SARS-CoV-2 infection, and other down-regulated pathways span many major categories of cellular metabolism, e.g. OXPHOS (energy metabolism), valine, leucine and isoleucine degradation (amino acid metabolism), sphingolipid metabolism (lipid metabolism), and N-glycan biosynthesis, etc. These possibly reflect the specific metabolic demands of SARS-CoV-2 or are related to its pathogenic effects. These results also suggest that besides immune response, metabolic reprogramming represents one of the most robust changes induced by SARS-CoV-2 infection across various systems, consistent with the key roles of metabolism in viral infection. However, it should be noted that there is no global downregulation of metabolic activity upon SARS-CoV-2 infection, as can be seen in Appendix Figure 4.26. As metabolic fluxes cannot be directly reflected by the gene expression levels. I therefore next seek to investigate the SARS-CoV-2-induced metabolic changes using GEM methods.

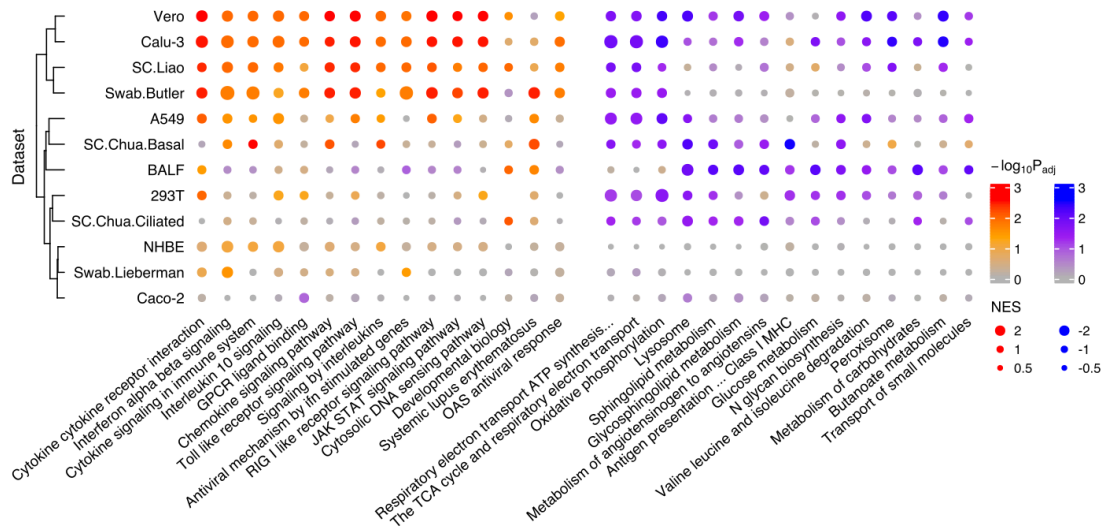


Figure 4.3: Summary of the top consistently enriched pathways from GSEA analysis across the analyzed datasets.

#### 4.2.2 GEM-based analysis of SARS-CoV-2-induced metabolic flux change patterns

I next applied the iMAT algorithm (Shlomi et al., 2008) to the median expression profiles of the control and virus-infected samples to compute the refined metabolic models representative of the two respective groups, and the flux distribution of each metabolic reaction was obtained by random sampling (ACHR method). The reactions with differential fluxes (DF) between the virus-infected samples vs controls were identified. I checked the overlap of the top DF reactions between each pair of datasets with Fisher’s exact tests to measure the level of consistency on the flux level. Like on the gene expression level, there is an overall high level of coherence of the DF reactions (odds ratio median 2.40, maximum 28.13; adjusted P value median and minimum both  $<2.22e-16$ ; see Appendix Figure 4.27 and Figure



summarized by each metabolite. Metabolites with the most consistent patterns across datasets are shown in Figure 4.5. In more than half of the datasets, several glucogenic amino acids, e.g. alanine, asparagine, and glutamate have increased cellular import from the extracellular space to the cytoplasm (or decreased export) in the virus-infected vs control samples, while nicotinamide and tetrahydrofolate show relatively consistent decrease in cellular import (or increase in export).

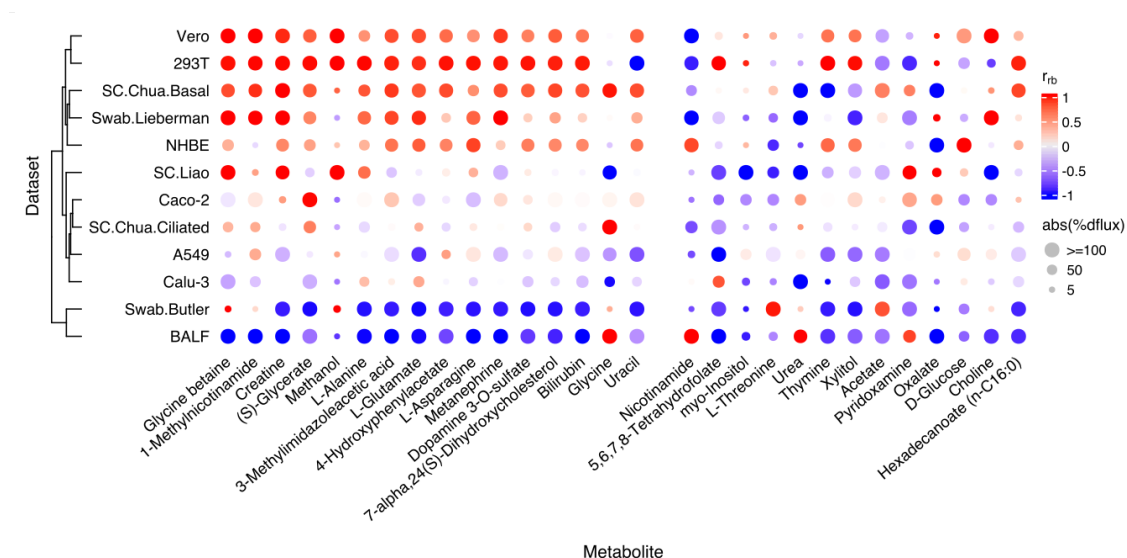


Figure 4.5: Summary of the metabolites showing the most consistent transportation patterns between the extracellular space and cytosol across the analyzed datasets.

Next, I visualized the metabolic networks for different individual pathways that are enriched for consistently changed reactions, in order to inspect the patterns of flux changes in these pathways. I show two example visualizations below, the others are given in Appendix. The pyrimidine *de novo* biosynthesis pathway has consistently increased fluxes towards the synthesis of UMP, the precursor of other pyrimidines, consistent with the nucleic acid synthesis requirement of SARS-CoV-2 (Figure 4.6, red and blue denotes increased and decreased absolute flux



phosphate metabolism pathway contains increased fluxes converging to phosphatidylinositol 4,5-bisphosphate (pail45p\_hs[c]) and inositol (inost[c]), but decreased fluxes to inositol 1-phosphate (mi1p-DASH.D[c]; Appendix Figure 4.31). In the glycine, serine, and threonine metabolism pathway, there is decreased conversion of serine to glycine but increased serine degradation to pyruvate (Appendix Figure 4.32); in the fatty acid elongation pathway, the synthesis and interconversion of different fatty acids show varied flux changes (Figure 4.7). These results show that SARS-CoV-2 can induce highly complex metabolic alterations potentially to meet their metabolic demands during their intracellular life cycle. They also demonstrate that GEM is able to add to gene expression-level analysis in recovering these fine-grained alteration patterns.

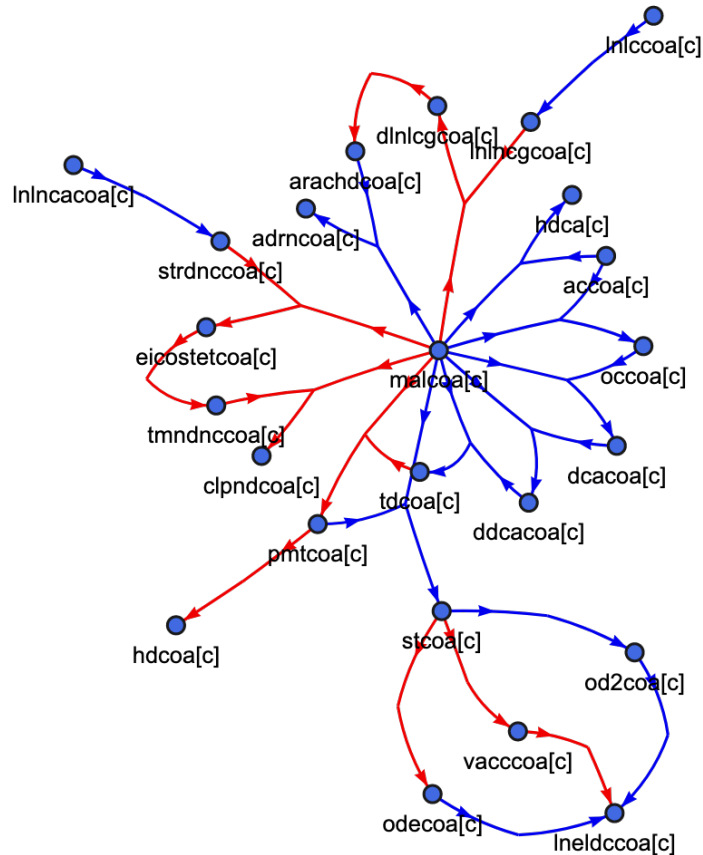


Figure 4.7: Visualization of the consistent flux changes in the fatty acid elongation pathway. Red and blue denotes increased and decreased absolute flux values, respectively; grey denotes inconsistent changes.

#### 4.2.3 Prediction and validation of metabolic anti-SARS-CoV-2 targets with MTA

Based on the proposal that targeting the virus-induced metabolic changes can be an effective antiviral strategy (Mayer et al., 2019) (Chapter 1.4.4), I next used MTA (described in Chapter 2, based on Yizhak et al., 2013; Valcárcel et al., 2019) to predict anti-SARS-CoV-2 targets, specifically those metabolic reactions whose KO can revert the cellular metabolic state from that of the virus-infected to that of the normal non-infected samples. As seen in Chapter 2, the 10-20% reactions with

the highest MTA score usually contain promising candidate targets. I inspected the overlaps of the top 10% MTA-predicted reactions based on each pair of datasets, and found overall reasonable overlaps (odds ratio median 3.42, maximum 16.47, Fisher's exact test adjusted P median and minimum both  $<2.2e-16$  across all pairwise comparison). Interestingly, there are some strong overlaps between certain cell line and patient datasets (Appendix).

I next aimed to validate these predictions using published anti-SARS-CoV-2 gene targets or drugs identified from large-scale chemical or genetic screens. First, I collected the top genes whose KO was found to inhibit SARS-CoV-2 infection in CRISPR-Cas9 genetic screens in Vero E6 cells (Wei et al., 2021) and A549 cells with exogenous ACE2 expression (Daniloski et al., 2021). The top 10% MTA-predictions from 8 out of the 12 datasets were found to have significant overlaps with these anti-viral hits from the CRISPR screens (after mapping to metabolic reactions,  $FDR < 3.95e-4$ ; Figure 4.8). Second, checking additional lists of experimentally validated drugs reported in different studies from (Kuleshov et al., 2020), there were also a few cases of significant overlap ( $FDR < 0.1$ ; Figure 4.9). Third, the top predictions from some of the datasets were also enriched for host proteins identified to interact with SARS-CoV-2 proteins from (Stukalov et al., 2020), although not those from (Gordon et al., 2020) (Appendix Figure 4.33).

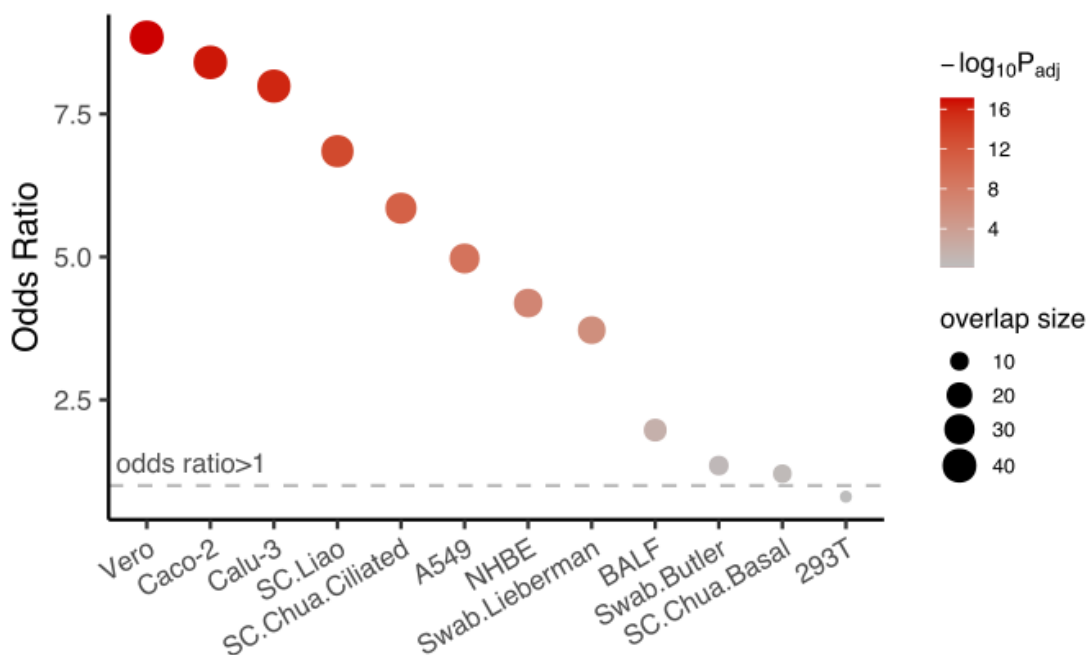


Figure 4.8: Summary of the enrichment of top MTA predictions from each dataset in the antiviral gene targets identified from CRISPR-Cas9 screens.

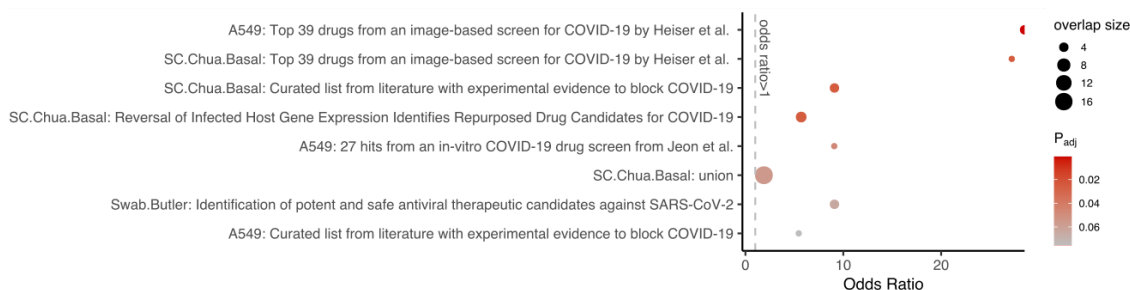


Figure 4.9: Summary of the enrichment of top MTA predictions from each dataset in the antiviral drugs identified from various experimental drug screens.

Further, based on the two genome-wide CRISPR-Cas9 screens described above, I defined comprehensive sets of positive and negative genes (i.e. genes whose KO inhibits or promotes the viral infection, respectively). These two sets are defined in a way such that they are balanced. After mapping these to the metabolic reaction level as the validation sets, I performed ROC curve analysis of the MTA predictions.

The predictions from 6 of the datasets achieved area under ROC curve (AUROC) values above 0.6 and as high as 0.81, although two of the other datasets apparently yielded AUROC significantly smaller than 0.5 (Figure 4.10). As examples, the ROC curve from the best-performing Calu-3 and SC.Liao datasets are shown in Appendix Figures 4.34 and 4.35. These results demonstrate that our metabolism-targeting strategy using the MTA algorithm is able to achieve reasonable prediction performances.

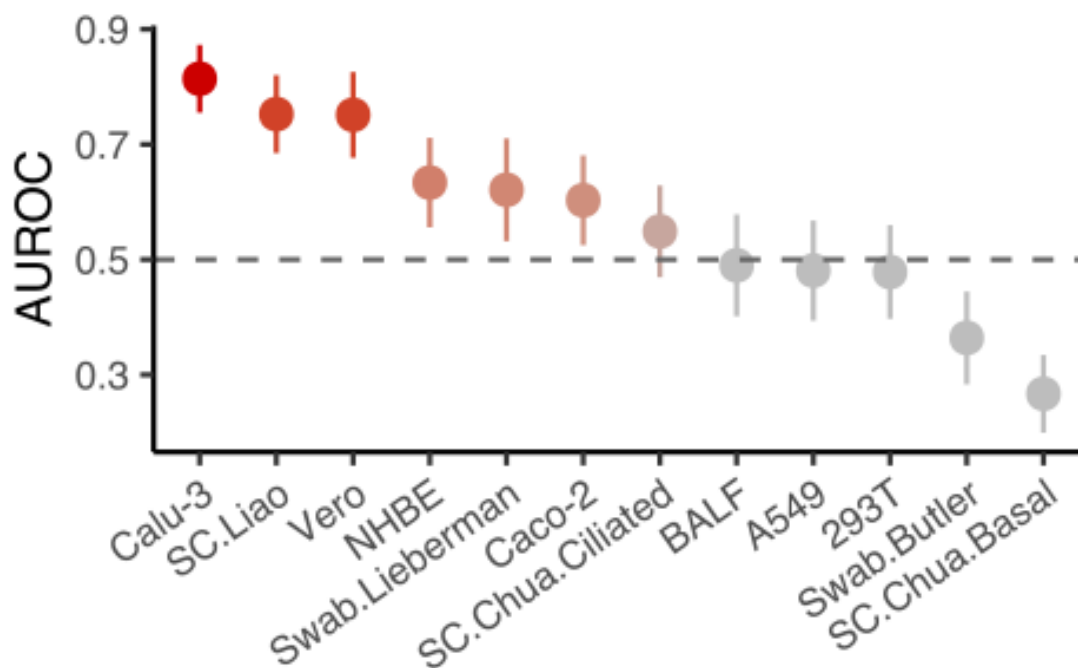


Figure 4.10: Summary of the AUROC of top MTA predictions from each dataset, based on validated positive and negative hits from CRISPR-Cas9 screens.

I next integrated the predictions from the 12 datasets into a final consensus list of high-confidence candidate targets (Methods), resulting in a final list of 66 candidate target metabolic reactions of which 59 are mapped to genes and 15 are targeted by known drugs. This final list of candidates are also strongly enriched for the posi-

tive targets identified in the two anti-SARS-CoV-2 CRISPR-Cas9 screens described above (odds ratio=26.56,  $P < 2.2e-16$ ). These candidates are enriched for cellular transport, inositol phosphate and aminosugar metabolism pathways, and within the cellular transport pathway, transporters for various amino acids, succinate, citrate and adenosine are enriched (FDR<0.1; Figure 4.11).

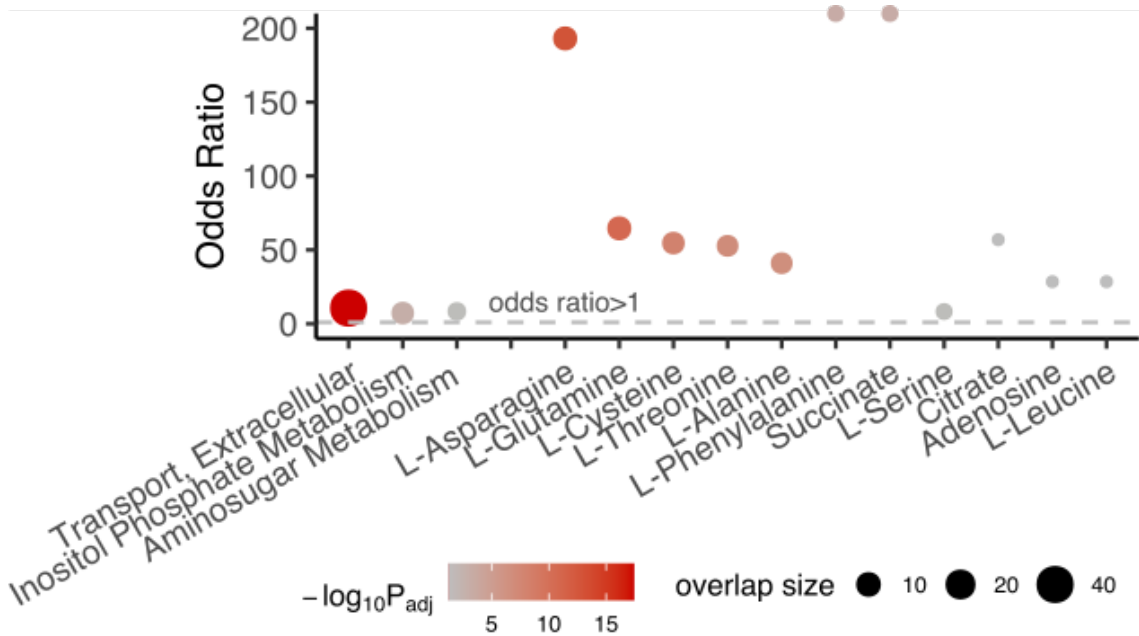


Figure 4.11: Summary of the metabolic pathway enrichment result of the consensus MTA-predicted targets across datasets.

#### 4.2.4 Prediction of metabolic targets for anti-SARS-CoV-2 in combination with remdesivir

Given that the MTA prediction of single anti-SARS-CoV-2 targets has yielded promising results, I applied the same strategy for the prediction of targets that can be combined with remdesivir to achieve higher antiviral efficacy. To this aim, we teamed up with experimental collaborators who cultured Vero E6 cells infected by

SARS-CoV-2, with or without remdesivir treatment. A control group (no viral infection or remdesivir treatment) a remdesivir-only group (no viral infection) were also included. We then performed RNA-seq on these samples. The PCA plot visualizing the gene expression of these samples are shown in Figure 4.12. From this plot it can be seen that remdesivir can indeed effectively reverse the virus-associated expression changes, but also results in additional orthogonal changes. Pathway enrichment via GSEA comparing the virus+remdesivir group to the control group confirms that there are still significant differences in many pathways, including some metabolic ones, e.g. cholesterol and steroid biosynthesis (Appendix Figure 4.36).

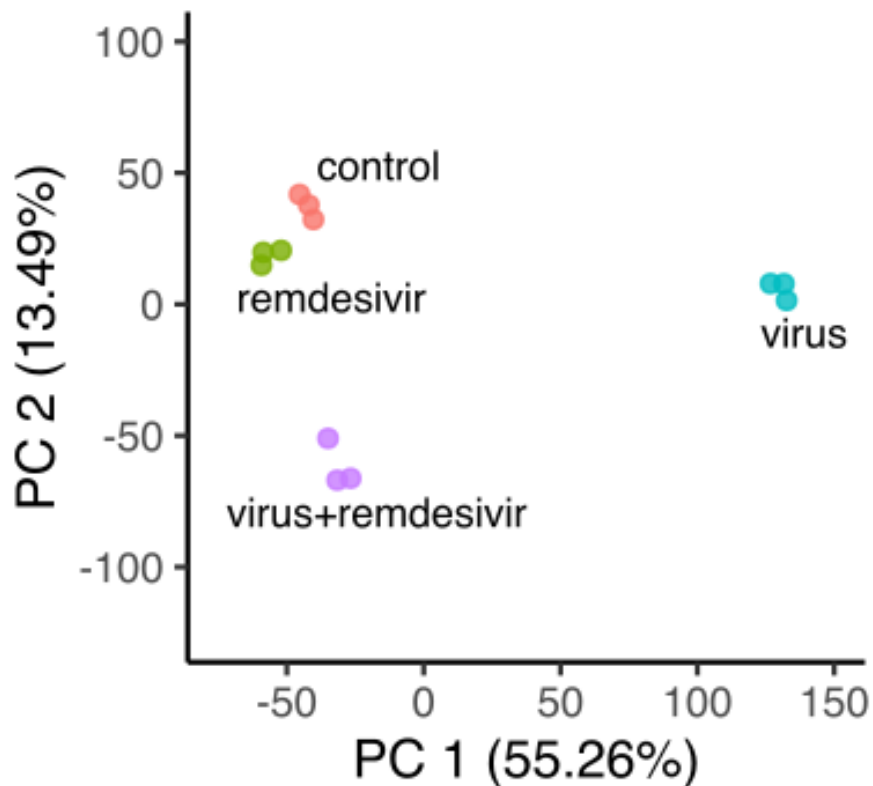


Figure 4.12: PCA plot on the transcriptome data from control Vero E6 cells as well as SARS-CoV-2-infected cells with or without remdesivir treatment

Using the iMAT algorithm to compute the metabolic fluxes of the samples as before, I confirmed that remdesivir does not fully reverse cell metabolism back to normal on the flux level (Appendix Figure 4.37). The metabolic pathways enriched by the differential fluxes between the virus+remdesivir and the control group are visualized in Figure 4.13. Based on the hypothesis that further reversing the cellular state in the virus+remdesivir group towards the healthy state may be an effective combinatorial targeting strategy to improve the antiviral efficacy of remdesivir, I applied MTA to predict targets for reversing these metabolic fluxes. There is little data available on known targets or drugs that are synergistic with remdesivir. As a preliminary validation of the predictions, I checked the enrichment between the top MTA prediction with 20 experimentally tested drugs showing synergistic anti-SARS-CoV-2 effects with remdesivir in the Calu-3 cell line reported in (Nguyenla et al., 2020). There is a significant enrichment using the top 20% MTA predictions (Fisher's exact test  $P=0.011$ , odds ratio 4.83; there is also sign of enrichment using the top 10% predictions with odds ratio 2.01, but it failed to achieve statistical significance at  $P=0.30$ ), recovering 6 of the 11 metabolic reaction targets from (Nguyenla et al., 2020). The metabolic pathways enriched by the top MTA predictions include bile acid, IMP, cholesterol and steroid metabolism pathways, and heme biosynthesis pathways, among others (Figure 4.14).

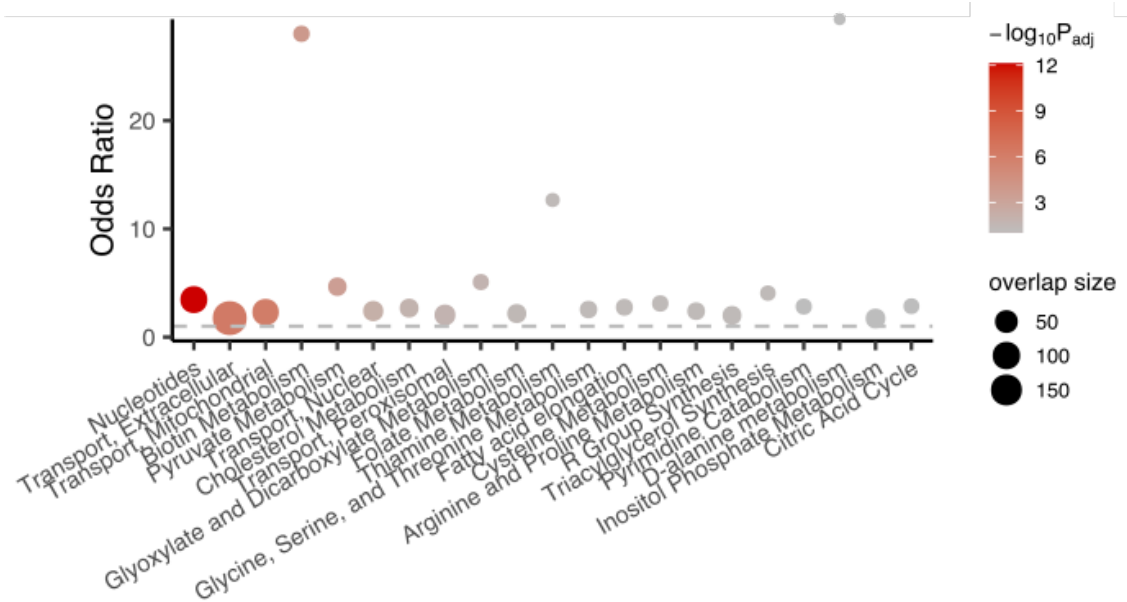


Figure 4.13: Summary of the metabolic pathway enrichment of the top differential fluxes between the virus+remdesivir and the control Vero E6 cells

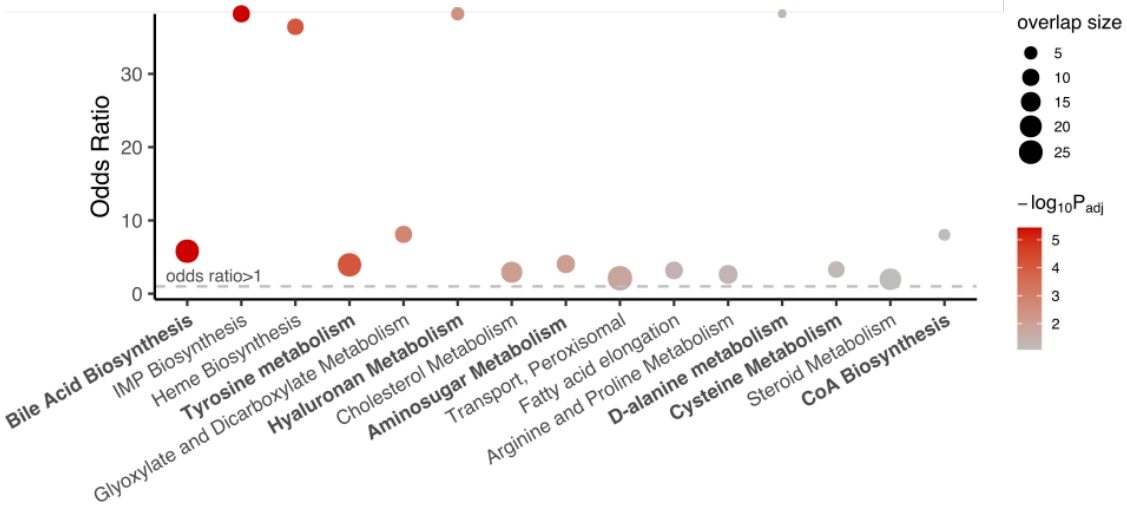


Figure 4.14: Summary of the metabolic pathway enrichment result of the top MTA-predicted targets for combination with remdesivir

### 4.3 Materials and Methods

### 4.3.1 DE analysis

We obtained each of the gene expression datasets on SARS-CoV-2 infection from the sources listed in Table 4.1 (Blanco-Melo et al., 2020; Bojkova, Klann, et al., 2020; Butler et al., 2020; Chua et al., 2020; Liao et al., 2020; Lieberman et al., 2020; Riva et al., 2020; Weingarten-Gabbay et al., 2020; Y. Xiong et al., 2020). For the bulk RNA-seq datasets whose read count data is available at the time of analysis, we performed DE analysis comparing the SARS-CoV-2-infected or positive samples to the non-infected control or negative samples with DESeq2 (Love, Huber, & Anders, 2014). For (Weingarten-Gabbay et al., 2020), a nested design for DE was needed and DESeq2 failed to run properly, and limma-voom (Law, Chen, Shi, & Smyth, 2014) was used. For (Butler et al., 2020; Y. Xiong et al., 2020), we obtained the DE results provided from the supplementary materials of the respective publication, as we were not able to obtain their gene-level expression data at the time of analysis. For (Butler et al., 2020), among their multiple versions of DE results we used the one from limma-voom with sva correction “Voom:Positive\_vs\_Negative:10M\_samples:sva\_correction\_2sv”. We also took the DE results of the proteomic data from (Bojkova, Klann, et al., 2020) as provided by the authors, and used the 24 hours post-infection data which is the latest time point available with the largest number of DE proteins). For the single-cell RNA-sequencing (scRNA-seq) datasets, The “FindMarkers” function in the R package Seurat (Stuart et al., 2019) was used to call the MAST method (Finak et al., 2015) for DE analysis in each annotated cell type, with “logfc.threshold” set to 0 to ob-

tain full results across genes. We focused on the airway epithelial cells since our major aim in this study is to investigate the changes in the cell types infected by the SARS-CoV-2 virus, these include the “Epithelial” cell type from (Liao et al., 2020) and the “Ciliated” and “Basal” cell types from (Chua et al., 2020) (other epithelial subtypes from these datasets yielded no significant DE genes).

### 4.3.2 GSEA analysis of the differential expression results

Using the DE log fold-change values from each dataset, GSEA (Subramanian et al., 2005) was performed using the implementation in the R package `fgsea` (Korotkevich et al., 2021). The gene set/pathway annotations used were the Reactome (Jassal et al., 2020) and KEGG (Kanehisa, Furumichi, Sato, Ishiguro-Watanabe, & Tanabe, 2021) subsets from the “Canonical Pathway” category in version 7.0 MSigDB database (Liberzon et al., 2011). For metabolic pathways, those under the category “Metabolism” from KEGG (Kanehisa et al., 2021) were used.

### 4.3.3 Comparison of the differentially expressed genes and pathways across datasets

The DE results across datasets were compared in a descriptive manner. As a first approach, the DE log fold-change values were inverse normal-transformed across all genes within each dataset, which preserves only the order (i.e. rank) of DE effect sizes, and then PCA was applied to the transformed data. As a second approach, top significantly DE genes or enriched pathways from each pair of datasets were

tested for significant overlap using Fisher’s exact tests. For DE genes, the number of significant genes vary greatly across datasets based on a single FDR cutoff, so we uniformly took the top  $n=400$  DE genes ranked by FDR from each datasets (all cases have at least  $FDR < 0.1$ ). For enriched pathways we simply took those with  $FDR < 0.1$  from each dataset. To identify the consistent DE changes across datasets, a formal meta-analysis of all 12 datasets is challenging given the wide range of assay platforms and DE algorithms used. So instead, we adopted a subjective criteria that give high importance to the various *in vivo* patient datasets, such that the results may be more clinically relevant: we identified pathways that are significantly ( $FDR < 0.1$ ) enriched in the consistent direction (up/down-regulation) in at least one of the bulk RNA-seq patient datasets and also at least one of the scRNA-seq datasets, while never showing significant enrichment ( $FDR < 0.1$ ) in the opposite direction in any of the datasets.

#### 4.3.4 Computation of metabolic fluxes from gene expression data with GEM

For each dataset, we used the GEM algorithm iMAT (Shlomi et al., 2008) to compute the metabolic flux profile from gene expression data. iMAT requires gene-length normalized expression values in the bulk RNA-seq datasets, for this we computed TPM values with Salmon (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) from the raw fastq files for datasets where TPM data is not provided. Then for each dataset, we took the median expression values of the control and virus-

infected samples respectively as the representative expression profile for each group, and used it as input to iMAT. The human GEM Recon 1 (Duarte et al., 2007) was used as the base model to iMAT. The main reason we used the older human GEM version was due to the significantly longer computation time and resources needed if later larger models were used instead. The output of iMAT is a refined GEM for the each of the virus-infected and control groups in each dataset, with metabolic reaction bounds adjusted to achieve maximal concordance with the gene expression data while satisfying the stoichiometric constraints of the cellular metabolic network (Shlomi et al., 2008). Each output model defines a space rather than a single unique solution of the global metabolic flux profile, and artificial centering hit-and-run (ACHR) was used to sample the metabolic space and obtain the flux distribution of each metabolic reaction in each condition (control or virus-infected) and dataset. All GEM analyses were performed using our in-house R package named gembox, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver.

#### 4.3.5 Differential flux analysis of virus-infected vs control group in each dataset

The flux distributions of the control and infected groups were compared to identify reactions with differential fluxes (DF). Since an arbitrarily large number of sample points can be sampled from the metabolic space of each group, resulting in statistical tests with arbitrarily small P values, we adopted the following effect

size-based criterion for DF reactions: absolute rank biserial correlation (an effect size measure of the difference between the two flux distributions in the control and virus-infected groups)  $>0.1$ , and absolute relative flux change (i.e. the absolute difference of the mean fluxes between the two groups over the absolute mean flux in the control group)  $>5\%$ . Positive DF reactions have flux value difference in infected vs control group  $>0$ , and vice versa for negative DF reactions. Note that for non-reversible reactions, flux values are non-negative and the sign of DF can be interpreted similarly to differential gene expression; for reversible reactions, flux values can be negative, representing reactions happening in the reverse direction, thus the sign of DF needs to be interpreted differently, e.g. negative DF represents flux shift towards the reverse direction and not necessarily decrease in absolute flux.

#### 4.3.6 Analysis of reactions with consistent differential fluxes across datasets and their pathway enrichment analysis

To compare the DF results across datasets, the DF reactions from each pair of datasets were tested for significant overlap using Fisher's exact tests (separately for positive and negative DF). Since no reaction shows fully consistent DF across all 12 datasets analyzed, similarly as with the DE analysis, we identified the DF reactions with high level of consistency especially in the in vivo patient datasets, such that the results may be more clinically relevant: we identified DF reactions in the consistent direction (positive/negative) in at least one of the bulk RNA-seq patient datasets and also at least one of the scRNA-seq datasets, while showing DF in the opposite

direction in no more than 3 datasets, resulting in 293 positive and 349 negative DF reactions. The metabolic pathway enrichment of these DF reactions was analyzed with Fisher's exact tests, with the "subSystems" slot in the metabolic model used as pathway annotation. Note that due to the special interpretation of the sign of DF values as explained above, the GSEA used for gene expression-level analysis is not appropriate for pathway enrichment analysis on the flux level.

#### 4.3.7 Differential transport flux analysis by each metabolite species

Differential transport flux analysis on the metabolite level was first performed for each dataset. Specifically, in each group (control or virus-infected), the sampled flux values on the reaction level for all transport reactions across the cell membrane were summarized by each metabolite, taking the direction of transport (i.e. import or export) into consideration to produce the net cellular transport flux of the metabolite (positive means net import, vice versa). The metabolites with differential net transport fluxes between the control and virus-infected groups were identified as in the reaction-level DF analysis described above. The consistent DF metabolites were then identified in the same way as with the DF reactions described above.

#### 4.3.8 Analysis of the consistent flux alteration patterns in different metabolic pathways

For each of the significantly enriched metabolic pathways identified in the consistent DF reaction analysis described above, we defined the "consensus" direction of

each reaction as represented by those shown in the virus-infected group from the majority (>6 out of 12) of the datasets, and also similarly for the “consensus” direction of DF for each reaction. The consensus directions of reactions and their DF were overlaid onto network diagrams of the pathways and visualized, where metabolites are represented by nodes, reactions are represented by directed (hyper) edges with edge direction corresponding to the consensus reaction direction and edge color corresponding to the consensus DF direction. Parts of the metabolic pathways where reactions are not consistently altered across datasets are greyed out or removed to increase the clarity. Upon visual inspection, potential futile loops in the network are also removed from the visualizations.

#### 4.3.9 Prediction of anti-SARS-CoV-2 target metabolic reactions with metabolic transformation algorithm

For each of the collected datasets, the DE result of virus-infected vs control samples as well as the representative flux distribution of the virus-infected group computed with iMAT (Shlomi et al., 2008) followed by ACHR sampling were used as inputs for the GEM-based MTA algorithm (described in Chapter 2, based on Yizhak et al., 2013; Valcárcel et al., 2019) to predict metabolic reactions whose knock-out can transform cellular metabolic state from that of the virus-infected to that of the control samples. The output of MTA is a score (MTA score) for each metabolic reaction, with higher scores corresponding to better candidates for achieving the metabolic transformation as specified above. From our previous experience (Yizhak

et al., 2013), the top 10-20% MTA predictions contain promising targets. The human Recon 1 GEM (Duarte et al., 2007) was used for the MTA analysis. The MTA algorithm implemented in our in-house R package named gembox was used, with the academic version of IBM ILOG CPLEX Optimization Studio 12.10 as the optimization solver. To compare the MTA predictions across datasets, the top 10% predictions from each pair of datasets were tested for significant overlap using Fisher’s exact tests.

#### 4.3.10 Validation of the MTA-predicted anti-SARS-CoV-2 metabolic targets

Multiple datasets of reported anti-SARS-CoV-2 gene targets or drugs identified from large-scale chemical or genetic screens were collected to validate our predictions. Gene-level results of two published CRISPR-Cas9 genetic screens (Daniloski et al., 2021; Wei et al., 2021) were obtained from the supplementary materials of the respective publication. For (Wei et al., 2021), gene hits with  $FDR < 0.1$  and mean  $z$  score  $> 0$  (i.e. KO inhibits the viral infection) were taken; (Daniloski et al., 2021) reported two screens with different multiplicities of infections (MOIs) and provided only single-sided FDR, so gene hits with  $FDR < 0.1$  from either screen were taken. The union set of hits from both studies were used. Lists of experimentally validated drugs reported in different studies compiled by (Kuleshov et al., 2020) were downloaded from <https://maayanlab.cloud/covid19/>, which are then mapped to the genes they inhibit using data from DrugBank v5.1.7 (Wishart et al., 2018).

Additionally, host proteins identified to interact with SARS-CoV-2 proteins were obtained from the supplementary materials of (Gordon et al., 2020) and (Stukalov et al., 2020). The genes from these validation datasets are mapped to metabolic reactions wherever applicable based on the human GEM Recon 1 (Duarte et al., 2007) data. Then, the significant overlap between the top 10% MTA-predicted targets from each dataset and each of the validation sets described above was tested with Fisher's exact tests on the reaction level. Reaction-level test is performed because multiple reactions can be mapped to the same gene, and performing Fisher's exact test on the gene-level fails to consider such multiple mapping and is thus inappropriate.

For ROC analysis, negative sets (i.e. genes whose KO promotes SARS-CoV-2 infection) were defined based on the two CRISPR-Cas9 screens described above. For (Wei et al., 2021), gene with  $FDR < 0.1$  and mean z score  $< 0$  were taken; since (Daniloski et al., 2021) provided only single-sided FDR, the log fold-change threshold corresponding to the  $FDR < 0.1$  cutoff was identified, and genes with more extreme log fold-changes in the opposite direction were taken. The union of the negative sets from both studies was used. Both the positive (described in the previous paragraph) and negative sets of genes are then mapped to metabolic reactions as described above. The negative set defined as such contains a relatively balanced number of reactions compared to the positive set (131 vs 81). The score for the reactions produced by MTA was used as the predicted value for ROC analysis. The R package pROC (Robin et al., 2011) was used to compute the AUROC values and their 95% confidence intervals (the latter computed with bootstrapping).

#### 4.3.11 Defining and analyzing the consensus set of candidate anti-SARS-CoV-2 metabolic targets across datasets

Based on top 10% MTA predictions from the 12 datasets (6 in vitro and 6 in vivo), the metabolic reaction targets that are recurrent in at least 3 of the in vitro datasets, and also in 3 of the in vivo datasets (i.e. the intersection of the two) were taken to be the final consensus candidate targets with high-confidence support across datasets. These target reactions are also mapped to genes using the human GEM Recon 1 data (Duarte et al., 2007), and further mapped to known drugs inhibiting the gene targets using data from DrugBank v5.1.7 (Wishart et al., 2018). The metabolic pathway enrichment of these targets was analyzed with Fisher’s exact tests, with the “subSystems” slot in the metabolic model used as pathway annotation. For enrichment analysis of cellular transport reactions by each metabolite species, the reactions responsible for transporting each metabolite across the cell membrane were identified from Recon 1 and used as reaction sets for the Fisher’s exact tests.

#### 4.3.12 Preparation of Vero E6 cell samples with SARS-CoV-2 infection and remdesivir treatment, RNA-sequencing, and gene expression data analysis

Vero E6 cells (ATCC® CRL-1586™) were maintained in Dulbecco’s modified eagle medium (DMEM, Gibco) supplemented with 10% heat-inactivated fetal bovine

serum (FBS, Gibco), 50 U ml<sup>-1</sup> penicillin, 50  $\mu$ g ml<sup>-1</sup> streptomycin, 1 mM sodium pyruvate (Gibco), 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Gibco), and 1 $\times$  MEM non-essential amino acids solution (Gibco). The SARS-CoV-2 USA-WA1/2020 strain was obtained from BEI Resources (NR-52281). The virus was inoculated on Vero E6 cells and the cell supernatant was collected at 72 h post-inoculation (hpi), when extensive cytopathic effects were observed. The supernatant, after clarification by centrifugation 15 min at 4 °C at 5,000 xg, was aliquoted and stored at -80 °C until use. 500,000 Vero E6 cells were seeded in 6-well plates. The following day, the cell medium was replaced with fresh medium supplemented with either DMSO or 1  $\mu$ M remdesivir (Adooq Bioscience), and cells were either mock-infected or infected with SARS-CoV-2 USA-WA1/2020 (MOI=0.3). Twenty-four hours after infection, cells were collected, and total intracellular RNA was extracted using the Qiagen® RNeasy® Plus Mini Kit. Three replicates were performed for each group, resulting in a total of six samples. The quality of the extracted RNA was assessed with Agilent® 2100 Bioanalyzer. Libraries were prepared on total RNA following ribosome RNA depletion with standard protocol according to Illumina®. Total RNA sequencing was then performed on the Illumina® NextSeq system, 150bp paired-end runs were performed and 100 million raw reads per sample were generated. STAR (Dobin et al., 2013) was used to align the reads to reference genome of the African green monkey (*Chlorocebus sabaeus*, [https://useast.ensembl.org/Chlorocebus\\_sabaeus/Info/Annotation](https://useast.ensembl.org/Chlorocebus_sabaeus/Info/Annotation)), with the SARS-CoV-2 genome ([https://www.ncbi.nlm.nih.gov/nucleotide/NC\\_045512](https://www.ncbi.nlm.nih.gov/nucleotide/NC_045512)) added to the reference genome. DESeq2 (Love et al., 2014) was used for DE analysis between

pairs of experimental groups (including virus+remdesivir vs control and virus vs control). GSEA (Subramanian et al., 2005) was performed using the implementation in the R package fgsea (Korotkevich et al., 2021). The gene set/pathway annotations used were the Reactome (Jassal et al., 2020) and KEGG (Kanehisa et al., 2021) subsets from the “Canonical Pathway” category in version 7.0 MSigDB database (Liberzon et al., 2011).

#### 4.3.13 Genome-scale metabolic modeling of the remdesivir-treated Vero E6 cell samples and prediction of anti-SARS-CoV-2 metabolic targets in combination with remdesivir

As with the metabolic modeling of the other datasets on SARS-CoV-2 infection, iMAT (Shlomi et al., 2008) together with ACHR was used to compute the metabolic flux distribution for each of the experimental groups, using the median expression TPM values of each group as the input to iMAT. Reactions with differential fluxes (DF) between groups (including virus+remdesivir vs control and virus vs control) were identified as described above, and their significant metabolic pathway enrichment was tested with Fisher’s exact tests, with pathways defined by the “subSystems” from the Recon 1 model (Duarte et al., 2007). The DE result of virus+remdesivir vs control group and the mean flux distribution of the virus+remdesivir group computed with iMAT were used as inputs for MTA to predict metabolic reactions whose knock-out can further transform the virus+remdesivir metabolic state to the normal control state. The top 10% and 20% MTA-predicted

targets were tested for significant enrichment for the targets of a list of experimentally validated synergistic drugs with remdesivir (Nguyenla et al., 2020) using Fisher’s exact test (performed on the metabolic reaction level as described above). Metabolic pathway enrichment analysis of the top MTA-predicted targets was performed as described above.

#### 4.3.14 Softwares and code

R version 3.6.3 was used for all statistical tests. P values lesser than  $2.22e-16$  may not be computed accurately and are reported as “ $P < 2.22e-16$ ” throughout the text. The Benjamini-Hochberg (BH) method was used for P value adjustment throughout the text. The R packages ggplot2 (Wickham, 2009), ComplexHeatmap (Gu, Eils, & Schlesner, 2016) and visNetwork (<https://cran.r-project.org/web/packages/visNetwork/index.html>) were used to create the visualizations. The code used for the analyses can be found in the GitHub repository: [https://github.com/ruppinlab/covid\\_metabolism](https://github.com/ruppinlab/covid_metabolism).

#### 4.4 Discussion and Conclusion

In this study, I took advantage of multiple published studies involving the gene expression profiling of SARS-CoV-2 infected samples, in order to comprehensively reanalyze these existing dataset with both gene expression-level and metabolic flux-level methods and discover recurrent virus-associated changes. The published studies involve a variety of different experimental models and/or sample types, and

although I was not able to collect and include all available datasets given their vast numbers, the selection of datasets was to cover those from both popular *in vitro* cell-based models of SARS-CoV-2 infection as well as human patients (nasopharyngeal swab and BALF samples). Another consideration is to cover datasets using different platforms. Specifically, I included a few scRNA-seq data to separate the distinctive changes within the epithelial cells from, e.g. various types of immune cells. In the current study, the main focus was on the epithelial cells, representing the major cell type infected by SARS-CoV-2, although it is also highly desirable to systematically investigate the SARS-CoV-2 induced changes (including metabolic changes) in the other cell types especially different immune cells.

For the method used for the combined analysis of multiple datasets, I have avoided a more formal effect size-based meta-analysis due to the challenge arising from the wide range of expression profiling platforms (bulk RNA-seq, scRNA-seq, MS-based proteomics) and the technical variation involved. Instead, in comparing the molecular alterations identified from across the datasets, I used the approach of comparing the significant changes (on the gene, pathway or flux level) and testing for their similarities with Fisher's exact tests, which helped to mitigate the variation across datasets due to different platforms and technical noise. In the identification of robust alterations supported by multiple datasets, due to the same reason as discussed above, a formal effect size-based analysis is challenging, and approaches for combining P values, for example the Fisher's method can be more practical. In this study, prior knowledge and subjected criteria were involved in the procedure which assigned more weights to the significant findings (genes/pathways/reactions)

from the *in vivo* patient datasets, as these may have higher clinical relevance.

Using the above integrated approach, the results show that different pathways of immune responses, including those response to virus infection, are recurrently upregulated upon SARS-CoV-2 infection across datasets (Figure 4.3), as expected. Antigen presentation via MHC-I is consistently downregulated upon the infection, consistent with the report that the ORF8 protein of SARS-CoV-2 can effectively downregulate MHC-I (Y. Zhang et al., 2020). The downregulation of antigen presentation has long been observed during the infection of many other viruses as well, as a mechanism of the viruses to evade immune detection (Hewitt, 2003). Besides immune responses, more intriguingly, metabolism is one of the cellular domains that exhibit the most coherent changes across datasets in SARS-CoV-2 infection, which echoes the previous reports that many viruses need to manipulate host metabolism for their proliferation and intracellular life cycle (Mayer et al., 2019; Thaker et al., 2019). While it appears that most of the consistent metabolic changes involves downregulated metabolic genes (Figure 4.3), it is not true that SARS-CoV-2 infection resulted in globally decreased metabolic activities, as can be seen in Appendix Figure 4.26. Besides, the pathway level changes should be carefully inspected based on the genes driving the enrichment to avoid misinterpretation of results. For example, the pathway “TCA cycle and electron transport chain” is seen consistently downregulated after SARS-CoV-2 infection, but upon closer inspection it’s only the ETC genes that are downregulated, while TCA cycle genes are not. Actually, from the GEM analysis it can be seen that the TCA cycle mostly has upregulated fluxes.

Therefore, it can also be seen that GEM can be a valuable complementary

approach to gene expression-level analysis. It is known that gene expression does not always perfectly correlate with protein level or enzyme activity (Maier, Güell, & Serrano, 2009). Besides, many metabolic reactions are reversible, while the directions of reactions are important biologically, such information is missing on the gene level. With GEM, the computed metabolic fluxes can reveal the extra complexity in SARS-CoV-2-induced metabolic reprogramming. Many of the metabolic flux changes computed by GEM modeling are consistent with prior knowledge on SARS-CoV-2 and other related viruses. For example, the highly coherent increase in pyrimidine biosynthesis corresponds to the increased need of viral genome replication and gene expression (Bojkova, Costa, et al., 2020), and pyrimidine de novo synthesis inhibitors have been shown to have anti-SARS-CoV-2 effects (R. Xiong et al., 2020). Inositol phosphate metabolism is important for the life cycle of many viruses due to the structural or signaling roles of different phosphoinositides (Beziau, Brand, & Piver, 2020), with the inhibition of certain phosphoinositides disrupting endocytosis and blocking SARS-CoV-2 cell-entry (Ou et al., 2020). Fatty acid synthesis was reported to increase in SARS-CoV-2 infection (Ehrlich et al., 2020); our results reveal a potentially more complex pattern for different types of fatty acids, which is paralleled by the results from several metabolomics studies (Barberis et al., 2020; Shen et al., 2020; Thomas et al., 2020). Both TCA cycle and OXPHOS have been shown to decrease based on gene expression during the virus infection (Ehrlich et al., 2020; Gardinassi et al., 2020), and have been implicated in the systemic syndromes of the virus (S. Li et al., 2021). In the GEM analysis of this study, the results only confirmed the decrease of OXPHOS but revealed a mostly normal-functioning TCA

cycle on the flux level, which appears to be driven by 2-oxoglutarate transported from the cytosol (Appendix Figures 4.29 and 4.30). These flux-level patterns computed by GEM modeling should be verified with isotope labeling experiments, and their biological significance in the virus infection needs to be further investigated.

Given the importance of metabolism during virus infection, targeting host metabolism has already been proposed as a promising novel antiviral strategy in general (Mayer et al., 2019). Based on this proposal, I applied the MTA framework to predict anti-SARS-CoV-2 targets that act via reverting the virus-induced metabolic perturbations. MTA is not based on supervised machine learning technique, and do not use any of the validation datasets for target prediction. Yet, the MTA predictions showed good performance during validation. It is particularly encouraging to see that in several cases, the validation data which was mostly from *in vitro* experiments correlated well with the predictions based on *in vivo* human data, for example, the predictions from the human patient scRNA-seq dataset from (Liao et al., 2020) produced AUROC of above 0.75 when validated against the CRISPR-Cas9 screen data from *in vitro* cell lines. In some datasets, however, our predictions were not successfully validated by the genetic screen data (Figures 4.8 and 4.10), one reason could be that MTA can only consider the metabolism-related effects and ignores other potential mechanisms that determine the antiviral efficacy of a target. It could also be due to biological differences between the datasets used for prediction and those used for validation. I therefore did not explicitly exclude the datasets with suboptimal performances in the validation to avoid overdependence on the limited validation sets when defining the most promising set of candidates.

Instead, I adopted a criteria of maximal support from as many datasets as possible while assigning more weights to the *in vivo* human patient datasets to achieve higher clinical relevance (as explained in Methods). In terms of the prediction for combinatory targets with remdesivir, the preliminary validation showed promising results despite that validation is much more limited due to lack of data. Follow-up studies are required to further validate the antiviral efficacy of these predicted targets.

In conclusion, it was discovered that metabolic reprogramming of the host cells is a robust characteristics of SARS-CoV-2 infection, via the combined analysis of multiple gene expression datasets on SARS-CoV-2 infection using both common gene-level bioinformatic analysis as well as metabolic modeling. The modeling approached added to the gene-level analysis in revealing fine-grained metabolic alteration patterns in the metabolic network, and also was critical in the prediction of metabolic anti-SARS-CoV-2 targets (both single and combinatory with remdesivir). Although the prediction was shown to have good performance based on validation against reported experimentally validated gene targets and drugs, these predictions should be further validated with rigorous experiments.

## Conclusion

In this thesis, I have provide an overview of the prevalent and crucial roles that cell metabolism plays in a wide spectrum of biological contexts and described the value of genome-scale metabolic modeling integrated with omic data as a generic framework for metabolic research. I first described my effort to comprehensively evaluate and further improve the metabolic transformation algorithm which can be used for causal factor/perturbation identification as well as therapeutic target prediction. Having comprehensively validated the good performance of the algorithm with a large number of validation datasets, I then demonstrated application of the genome-scale metabolic modeling framework, especially with the metabolic transformation algorithm, in two extensive case studies. In the first case, I studied the metabolic processes that regulates the anti-cancer function of CD8 T cells in cancer immunotherapies, where metabolic modeling analysis identified the mitochondrial uncoupling protein UCP2 as essential for CD8 T cell anti-cancer function. Further experimental validation confirmed the effect of Ucp2 in a mice model and showed that Ucp2 regulates CD8 T cell stemness via modulating reactive oxygen species levels. In the second study, I investigated the host metabolic alterations during SARS-CoV-2 infection and predicted anti-SARS-CoV-2 targets that act via

reversing the virus-induced host metabolic reprogramming. The predicted single targets and combinatorial targets with remdesivir were validated computationally based on published results of large-scale drug and genetic screens for anti-SARS-CoV-2 drug/target discovery. In both cases, the metabolic modeling approach was shown effective in revealing relevant and previously unknown metabolic features, which contributes to a deeper understanding of the role of cell metabolism in the respective system under study. More importantly, the findings in both cases are directly translatable to guide the development and optimization of disease therapeutics. This showcases the value and power of the genome-scale metabolic modeling framework in gaining novel and translational insights for the research in various fields related to cell metabolism.

## Appendices

Additional figures and tables for each chapter are given below.

### Chapter 2

Table 4.2: Gene expression datasets involving metabolic gene perturbation (KO/KD/protein inhibition) collected for the validation of the MTA algorithm and its variants

dataset.id	ground.truth.gene
GSE100778.1	BCAT1
GSE100778.2	BCAT1
GSE100778.3	BCAT1
GSE10289	SDHB
GSE103007	DNMT3A
GSE103960.1	BCAT1
GSE103960.2	BCAT1
GSE104123	GMDS
GSE108656	SETDB1
GSE108742	PFKFB4

GSE109240	NT5C2
GSE111760	ACOT12
GSE115590	DGKZ
GSE116162	SLC3A2
GSE117774	DGAT1, DGAT2
GSE117775	DGAT1, DGAT2
GSE13458	NMNAT1
GSE29750.1	NOS2
GSE29750.2	NOS2
GSE3699	DNMT1
GSE38367	UGT2B17
GSE39943.1	CHKA
GSE39943.2	CHKA
GSE42644	DNMT3B
GSE46538.1	PDE1C
GSE46538.2	PDE1C
GSE51395.1	ST3GAL1
GSE51395.2	ST3GAL1
GSE61286.1	ACACA
GSE61286.2	ACLY
GSE61843	AK4

GSE63746	FASN
GSE64628.1	GALNT1
GSE64628.2	GALNT1
GSE67537.1	GNPNAT1
GSE67537.2	GNPNAT1
GSE68869	PTEN
GSE70561	TYRP1
GSE70637	DNMT1
GSE72320	CPT2
GSE73456	AGL
GSE75755	GALNT2
GSE76440	SHMT1
GSE76515	TUSC3
GSE76675.1	SLC16A1
GSE76675.2	SLC16A1
GSE76675.3	SLC16A1
GSE76675.4	SLC16A1
GSE8045	ALDH1A1, ALDH3A1
GSE82096	FASN
GSE84007	MTHFD2
GSE86147	DNMT1

GSE90483	TK1
GSE90668	C1GALT1
GSE90672	C1GALT1
GSE93425.1	RRM1
GSE93425.2	RRM2
GSE94751	SLC25A10

---

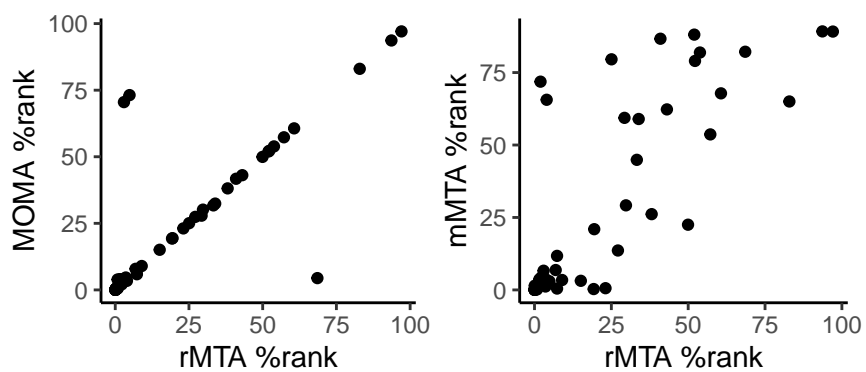


Figure 4.15: Comparison of the predicted percentage ranks of the ground truth reactions across datasets from rMTA, mMTA and MOMA, with the best parameter combinations.

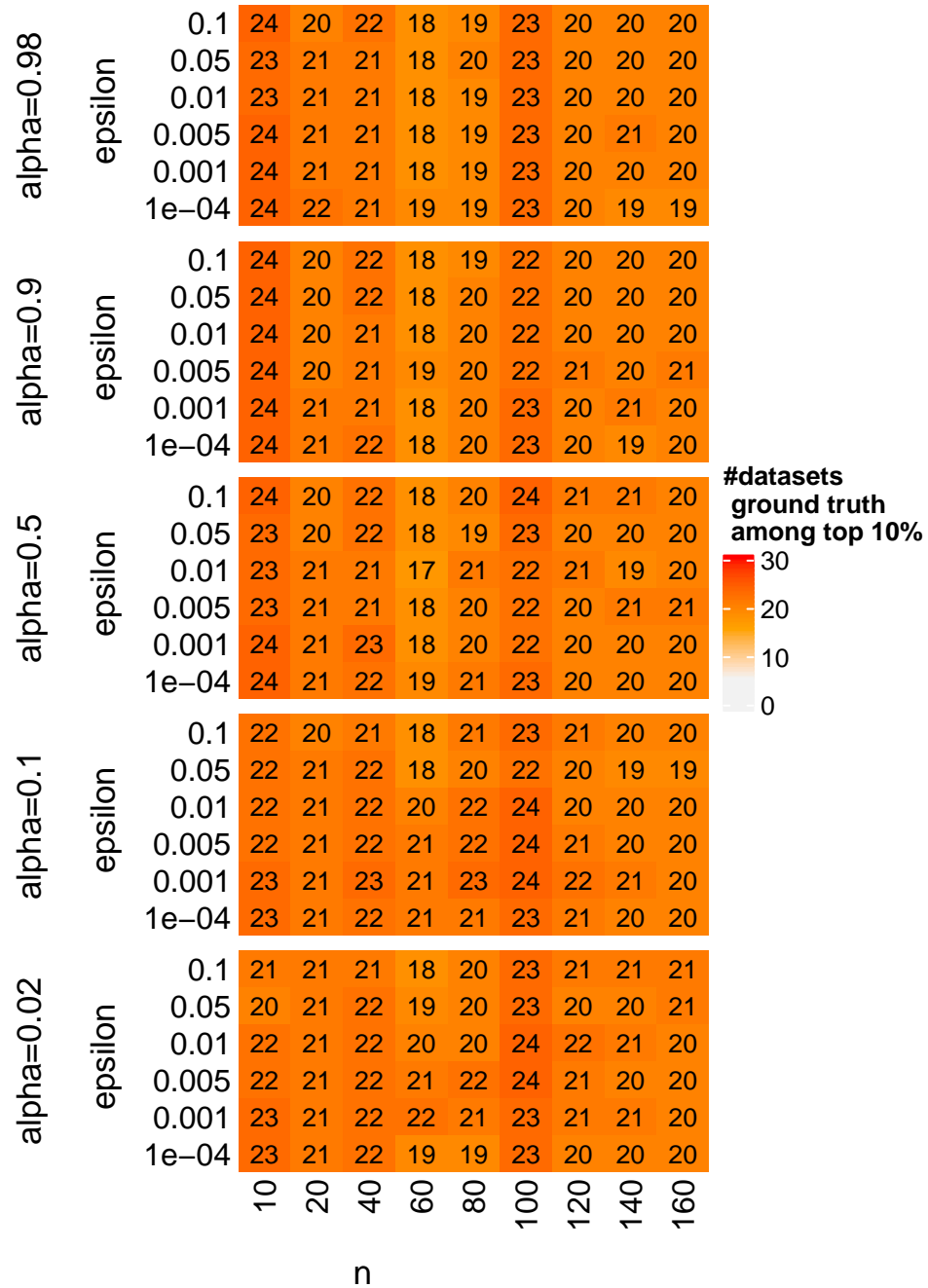


Figure 4.16: The prediction performance of rMTA across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to rMTA.

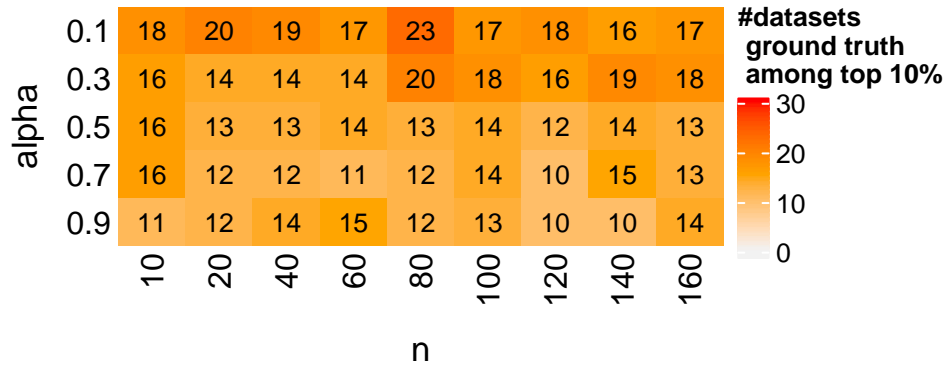


Figure 4.17: The prediction performance of mMeTAL across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to mMeTAL.

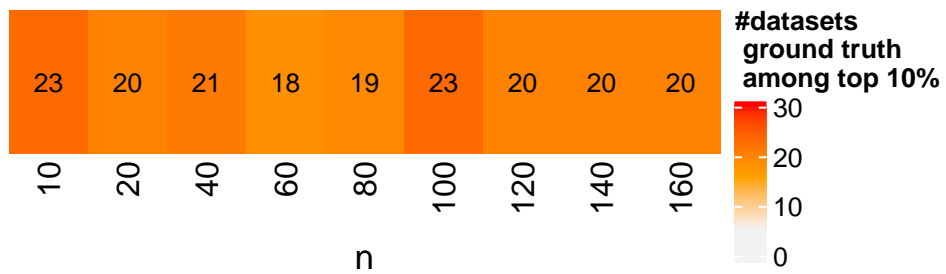


Figure 4.18: The prediction performance of MOMA across datasets with various parameters, after removing the ground truth gene from the DE result given as the input to MOMA.

## Chapter 3

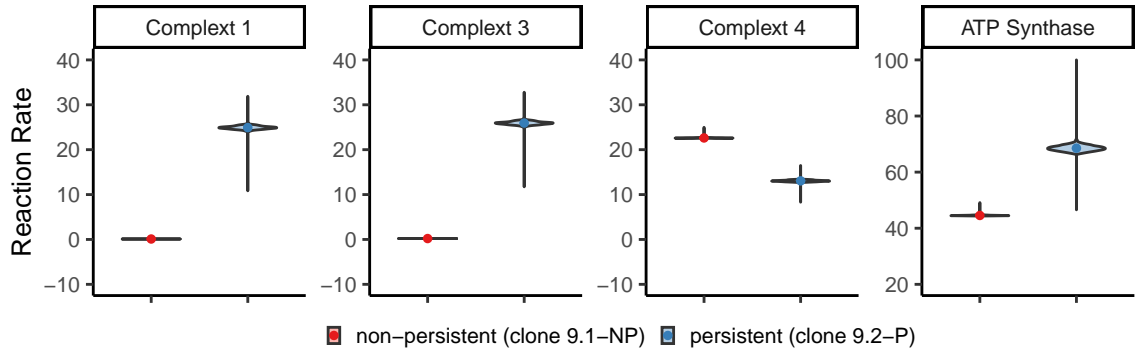


Figure 4.19: Computed metabolic flux distributions of the ETC/OXPHOS reactions in the 9.2-P persistent T cells vs the 9.1-NP non-persistent T cells in ACT.

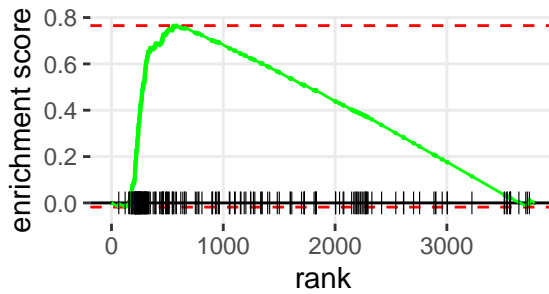


Figure 4.20: GSEA plot showing the enrichment for the mitochondrial transport pathway by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy.

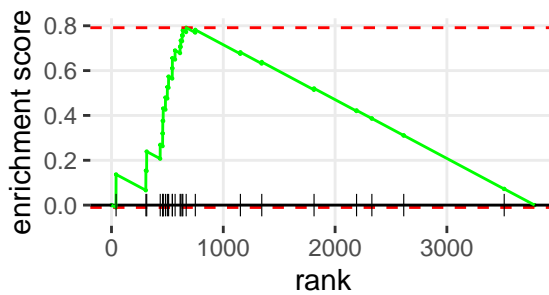


Figure 4.21: GSEA plot showing the enrichment for the mitochondrial proton transport reactions by the top MTA predictions on essential reactions to T cell anti-cancer function in CAR-T therapy.

\begin{table}

\caption{Enrichment of mitochondrial transport pathway and mitochondrial proton transport reactions by the top 10% MTA predicted reactions whose KO can reduce T cell persistence in ACT therapy.}

reaction.set	odds.ratio	pval	padj	reaction.set.size	overlap.size
Transport, Mitochondrial	4.469364	0.0000000	0.0000000	178	50
Mitochondrial Proton Transport	2.732816	0.0389527	0.089362	26	6

\end{table}

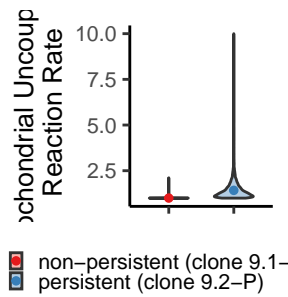


Figure 4.22: Mitochondrial uncoupling reaction rate in the 9.2-P persistent T cells vs the 9.1-NP non-persistent T cells in ACT.

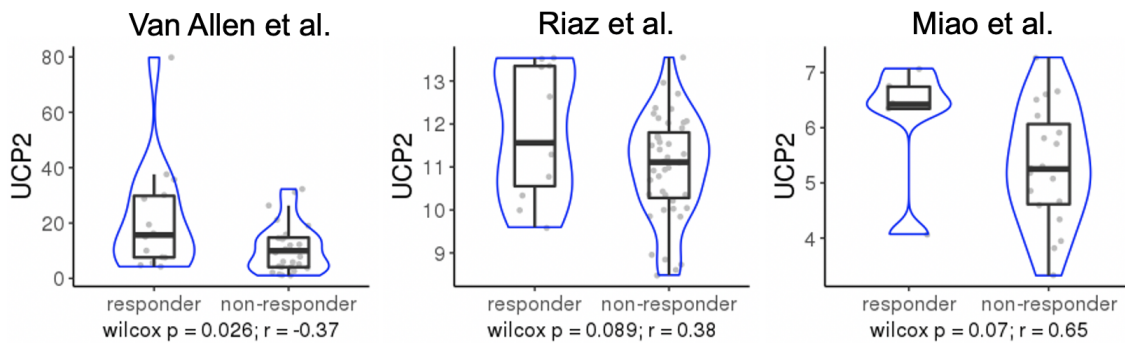


Figure 4.23: UCP2 expression levels in the tumor samples from responders and non-responders in various ICB datasets.

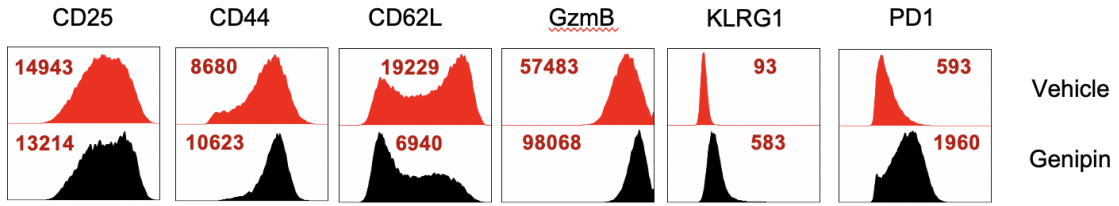


Figure 4.24: Flow cytometry results on the level of various T cell activation, effector function, memory and senescence gene markers in genipin-treated vs control mice Pmel-1 CD8 T cells. The gene markers are as follows: T cell activation (CD25, CD44), memory (CD62L), effector molecule (GzmB), senescence (KLRG1) and exhaustion (PD1). Figure credit: Dr. Madhusudhanan Sukumar

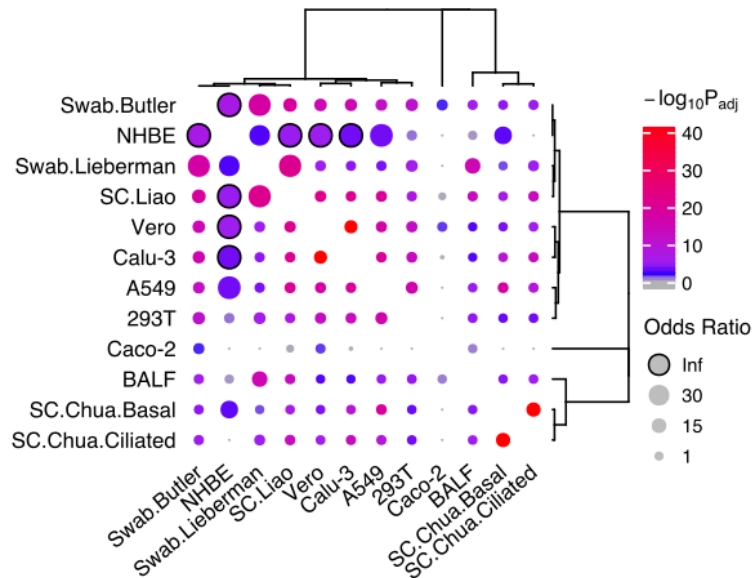


Figure 4.25: Pairwise comparison of the datasets by their top significantly enriched pathways with Fisher's exact tests.



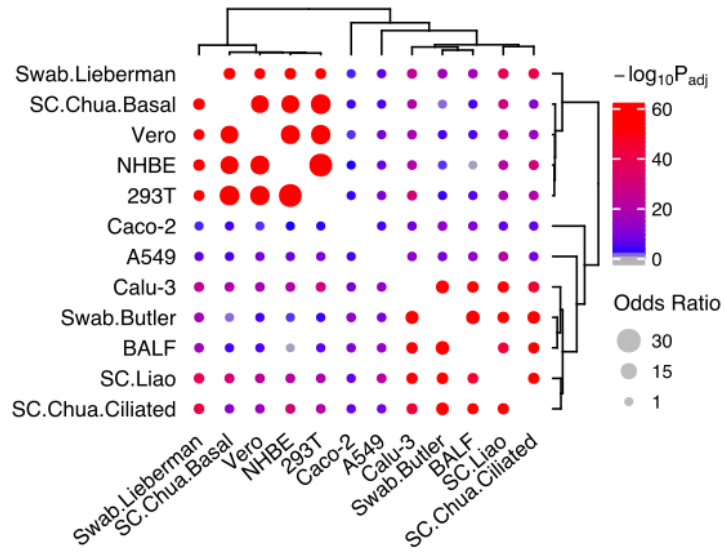


Figure 4.28: Pairwise comparison of the datasets by their top negative DF reactions with Fisher's exact tests.

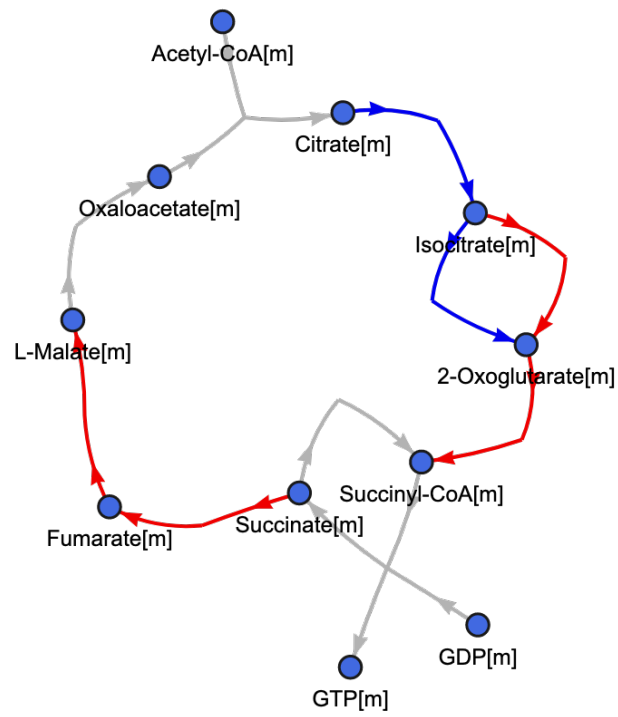


Figure 4.29: Visualization of the consistent flux changes in the TCA cycle.

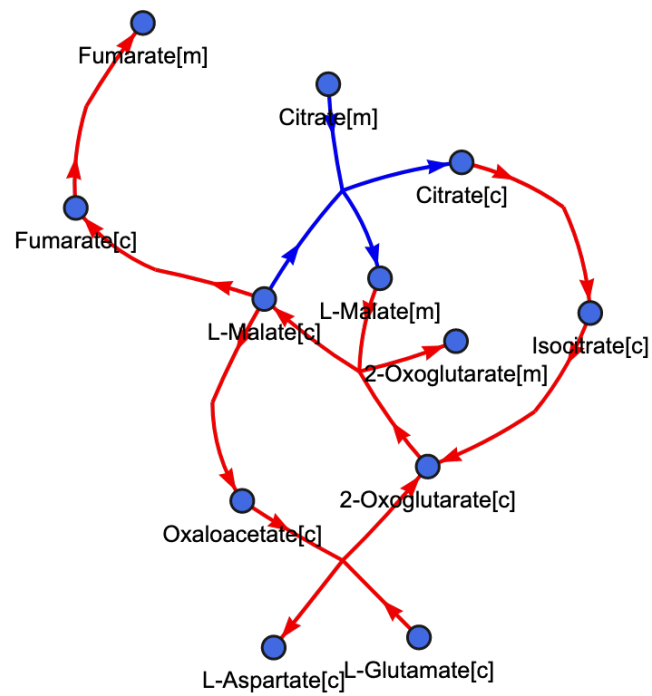


Figure 4.30: Visualization of the consistent flux changes in the reactions involving the cytosolic counterparts of TCA cycle metabolites.

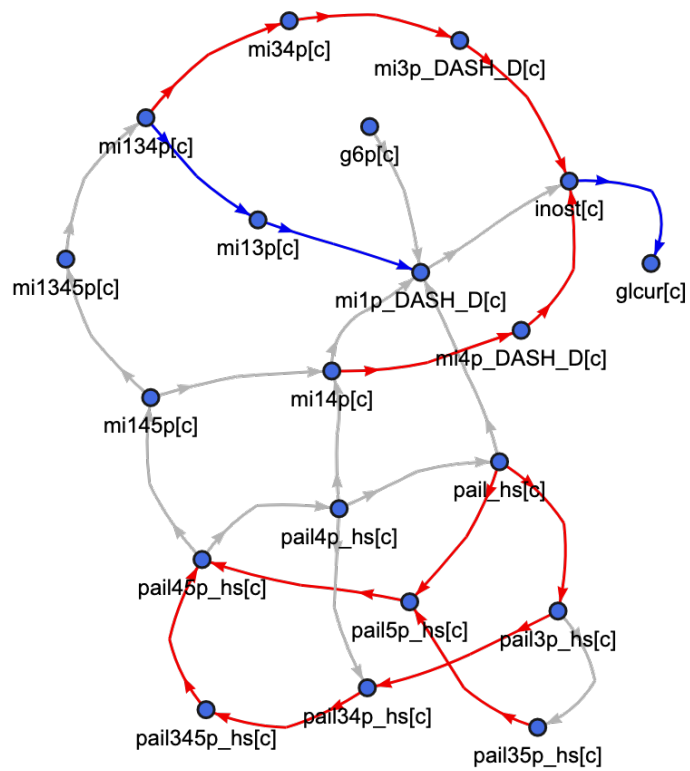


Figure 4.31: Visualization of the consistent flux changes in the inositol phosphate metabolism pathway.

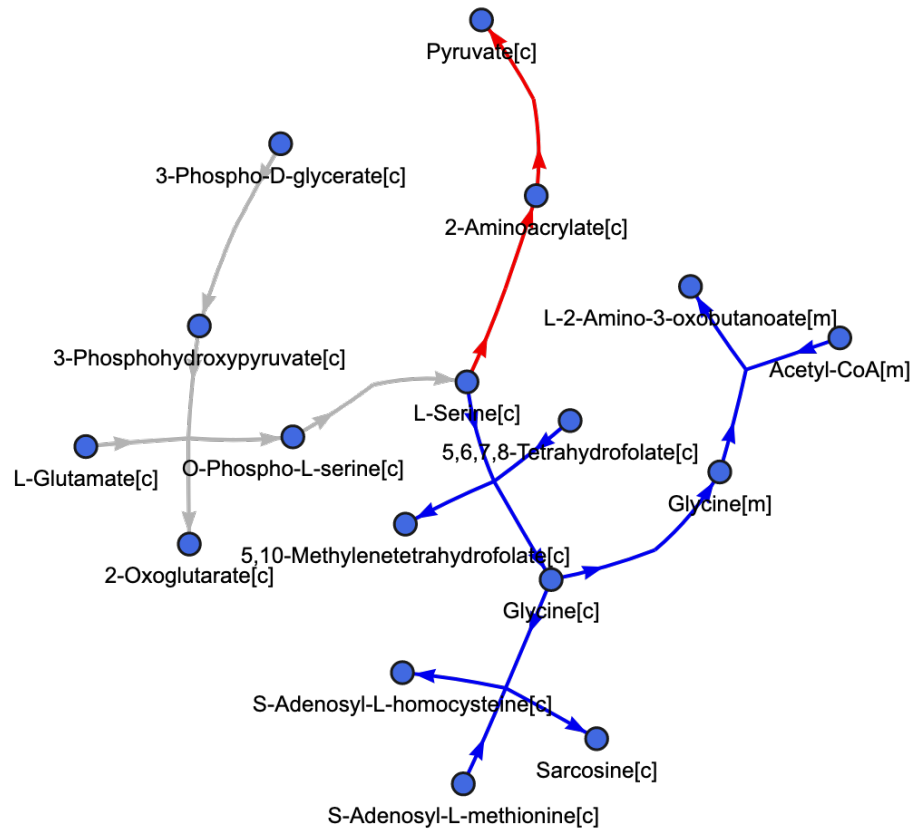


Figure 4.32: Visualization of the consistent flux changes in the glycine, serine, and threonine metabolism pathway.

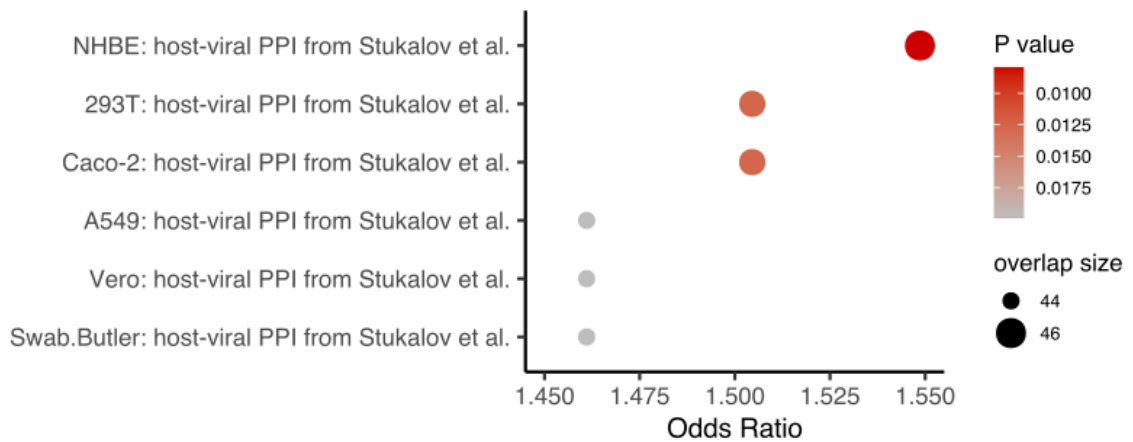


Figure 4.33: Summary of the enrichment of top MTA predictions from each dataset in the host-virus PPI proteins.

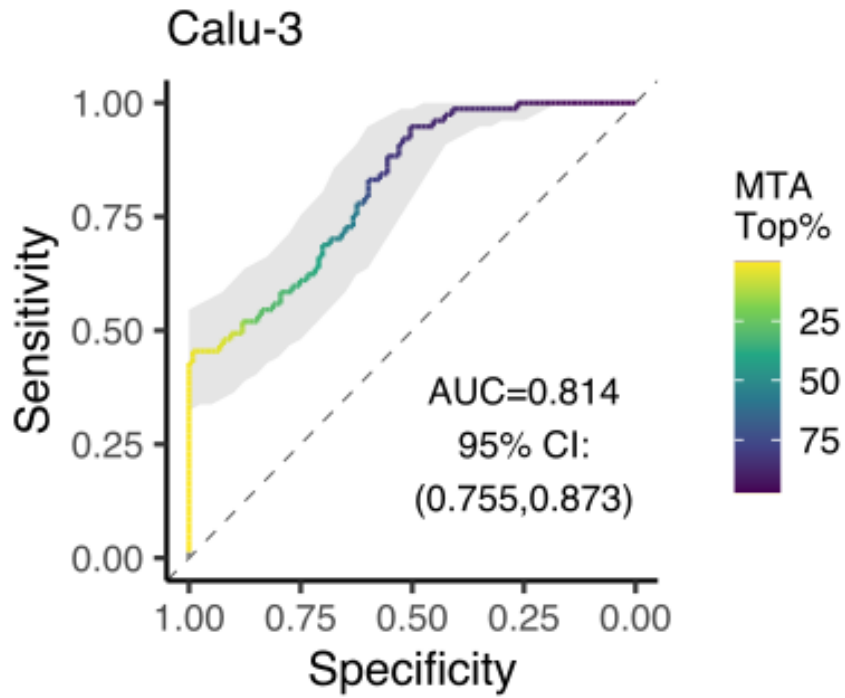


Figure 4.34: An example ROC curve for the Calu-3 dataset.

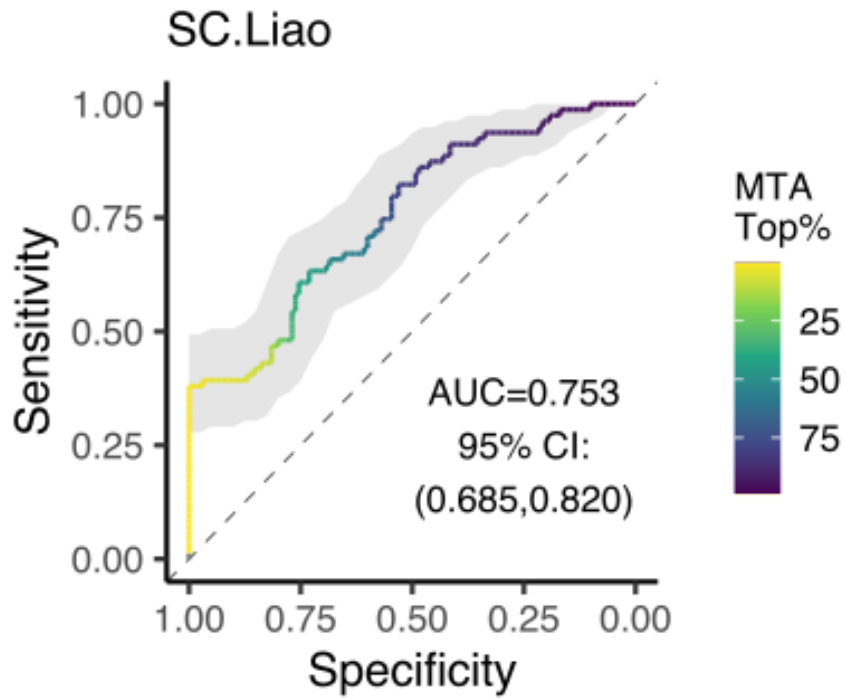


Figure 4.35: An example ROC curve for the Liao et al. single-cell dataset.

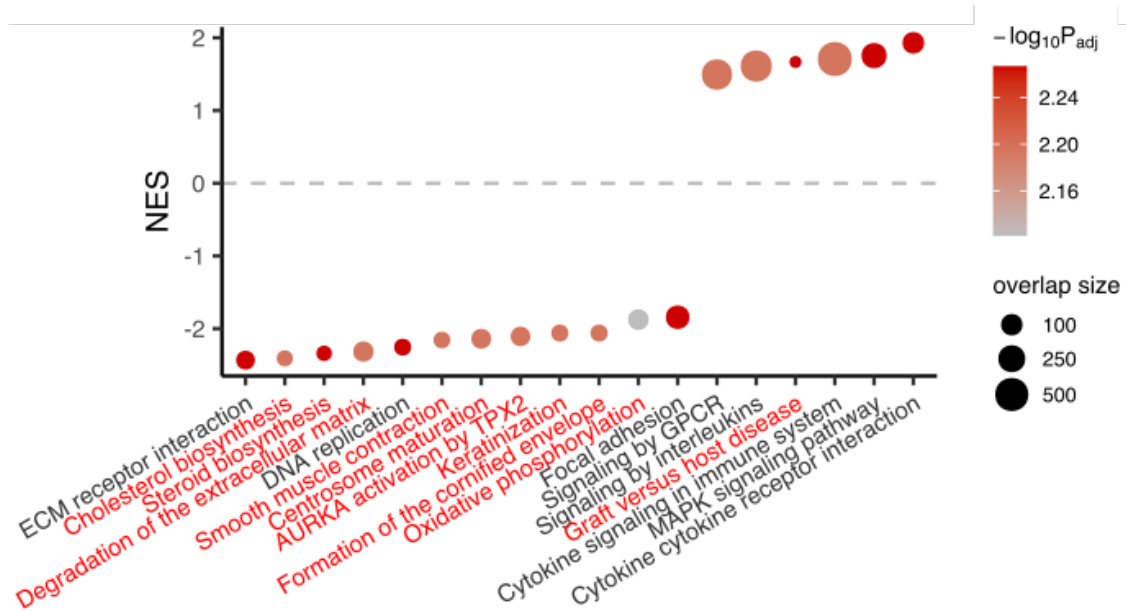


Figure 4.36: Summary of the significant pathways with differentially expressed genes between the virus+remdesivir and the control Vero E6 cells

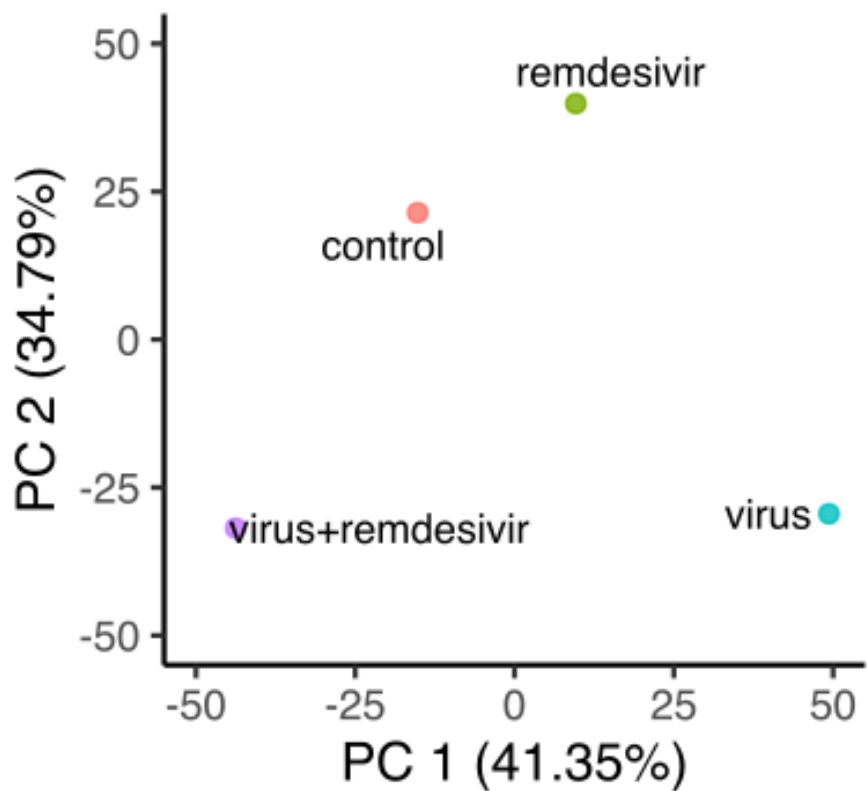


Figure 4.37: PCA plot on the iMAT-computed metabolic fluxes from control Vero E6 cells as well as SARS-CoV-2-infected cells with or without remdesivir treatment

## Bibliography

Abbott, T. R., Dhamdhare, G., Liu, Y., Lin, X., Goudy, L., Zeng, L., . . . Qi, L. S.

(2020). Development of CRISPR as an Antiviral Strategy to Combat

SARS-CoV-2 and Influenza. *Cell*, *181*(4), 865–876.e12.

<http://doi.org/10.1016/j.cell.2020.04.020>

Ackerman, D., & Simon, M. C. (2014). Hypoxia, lipids, and cancer: Surviving the

harsh tumor microenvironment. *Trends Cell Biol*, *24*(8), 472–478.

<http://doi.org/10.1016/j.tcb.2014.06.001>

Adams, A. E., Kelly, O. M., & Porter, R. K. (2010). Absence of mitochondrial

uncoupling protein 1 affects apoptosis in thymocytes, thymocyte/T-cell profile

and peripheral T-cell number. *Biochim Biophys Acta*, *1797*(6-7), 807–816.

<http://doi.org/10.1016/j.bbabi.2010.04.016>

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., & Nielsen,

J. (2012). Reconstruction of genome-scale active metabolic networks for 69

human cell types and 16 cancer types using INIT. *PLoS Comput Biol*, *8*(5),

e1002518. <http://doi.org/10.1371/journal.pcbi.1002518>

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., & Nielsen, J.

- (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol*, *10*, 721.  
<http://doi.org/10.1002/msb.145122>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat Med*, *26*(4), 450–452.  
<http://doi.org/10.1038/s41591-020-0820-9>
- Angelin, A., Gil-de-Gómez, L., Dahiya, S., Jiao, J., Guo, L., Levine, M. H., ... Beier, U. H. (2017). Foxp3 Reprograms T Cell Metabolism to Function in Low-Glucose, High-Lactate Environments. *Cell Metab*, *25*(6), 1282–1293.e7.  
<http://doi.org/10.1016/j.cmet.2016.12.018>
- Artyomov, M. N., & Van den Bossche, J. (2020). Immunometabolism in the Single-Cell Era. *Cell Metab*, *32*(5), 710–725.  
<http://doi.org/10.1016/j.cmet.2020.09.013>
- Auslander, N., Cunningham, C. E., Toosi, B. M., McEwen, E. J., Yizhak, K., Vizeacoumar, F. S., ... Ruppin, E. (2017). An integrated computational and experimental study uncovers FUT9 as a metabolic driver of colorectal cancer. *Mol Syst Biol*, *13*(12), 956. <http://doi.org/10.15252/msb.20177739>
- Auslander, N., Zhang, G., Lee, J. S., Frederick, D. T., Miao, B., Moll, T., ... Ruppin, E. (2018). Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med*, *24*(10), 1545–1549.  
<http://doi.org/10.1038/s41591-018-0157-9>

- Ayres, J. S. (2020). A metabolic handbook for the COVID-19 pandemic. *Nat Metab*, 2(7), 572–585. <http://doi.org/10.1038/s42255-020-0237-2>
- Balkwill, F., & Mantovani, A. (2001). Inflammation and cancer: Back to Virchow? *The Lancet*, 357(9255), 539–545.  
[http://doi.org/10.1016/S0140-6736\(00\)04046-0](http://doi.org/10.1016/S0140-6736(00)04046-0)
- Barberis, E., Timo, S., Amede, E., Vanella, V. V., Puricelli, C., Cappellano, G., ... Manfredi, M. (2020). Large-Scale Plasma Analysis Revealed New Mechanisms and Molecules Associated with the Host Response to SARS-CoV-2. *Int J Mol Sci*, 21(22). <http://doi.org/10.3390/ijms21228623>
- Batlle, E., & Massagué, J. (2019). Transforming Growth Factor- $\beta$  Signaling in Immunity and Cancer. *Immunity*, 50(4), 924–940.  
<http://doi.org/10.1016/j.immuni.2019.03.024>
- Becker, S. A., & Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*, 4(5), e1000082.  
<http://doi.org/10.1371/journal.pcbi.1000082>
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., ... ACTT-1 Study Group Members. (2020). Remdesivir for the Treatment of Covid-19 - Final Report. *N Engl J Med*, 383(19), 1813–1826.  
<http://doi.org/10.1056/NEJMoa2007764>
- Bekiaris, P. S., & Klamt, S. (2020). Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, 21(1), 19.

<http://doi.org/10.1186/s12859-019-3329-9>

Berardi, M. J., & Chou, J. J. (2014). Fatty acid flippase activity of UCP2 is essential for its proton transport in mitochondria. *Cell Metab*, *20*(3), 541–552.  
<http://doi.org/10.1016/j.cmet.2014.07.004>

Berod, L., Friedrich, C., Nandan, A., Freitag, J., Hagemann, S., Harmrolfs, K., ... Sparwasser, T. (2014). De novo fatty acid synthesis controls the fate between regulatory T and T helper 17 cells. *Nat Med*, *20*(11), 1327–1333.  
<http://doi.org/10.1038/nm.3704>

Beziau, A., Brand, D., & Piver, E. (2020). The Role of Phosphatidylinositol Phosphate Kinases during Viral Infection. *Viruses*, *12*(10).  
<http://doi.org/10.3390/v12101124>

Bikdeli, B., Madhavan, M. V., Jimenez, D., Chuich, T., Dreyfus, I., Driggin, E., ... Global COVID-19 Thrombosis Collaborative Group, Endorsed by the ISTH, NATF, ESVM, and the IUA, Supported by the ESC Working Group on Pulmonary Circulation and Right Ventricular Function. (2020). COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review. *J Am Coll Cardiol*, *75*(23), 2950–2973.  
<http://doi.org/10.1016/j.jacc.2020.04.031>

Billah, M. A., Miah, M. M., & Khan, M. N. (2020). Reproductive number of coronavirus: A systematic review and meta-analysis based on global level

evidence. *PLoS One*, 15(11), e0242128.

<http://doi.org/10.1371/journal.pone.0242128>

Blais, E. M., Rawls, K. D., Dougherty, B. V., Li, Z. I., Kolling, G. L., Ye, P., ...

Papin, J. A. (2017). Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*, 8, 14250.

<http://doi.org/10.1038/ncomms14250>

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., ...

tenOever, B. R. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, 181(5), 1036–1045.e9.

<http://doi.org/10.1016/j.cell.2020.04.026>

Bobrowski, T., Chen, L., Eastman, R. T., Itkin, Z., Shinn, P., Chen, C. Z., ...

Muratov, E. N. (2021). Synergistic and Antagonistic Drug Combinations against SARS-CoV-2. *Mol Ther*, 29(2), 873–885.

<http://doi.org/10.1016/j.ymthe.2020.12.016>

Bojkova, D., Costa, R., Bechtel, M., Ciesek, S., Michaelis, M., & Cinatl, J. (2020).

Targeting pentose phosphate pathway for SARS-CoV-2 therapy. *bioRxiv*, 2020.08.19.257022. <http://doi.org/10.1101/2020.08.19.257022>

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., ... Münch, C.

(2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*, 583(7816), 469–472.

<http://doi.org/10.1038/s41586-020-2332-7>

- Braunstein, A., Muntoni, A. P., & Pagnani, A. (2017). An analytic approximation of the feasible space of metabolic networks. *Nat Commun*, *8*, 14915.  
<http://doi.org/10.1038/ncomms14915>
- Brentjens, R. J., Davila, M. L., Riviere, I., Park, J., Wang, X., Cowell, L. G., ... Sadelain, M. (2013). CD19-Targeted T Cells Rapidly Induce Molecular Remissions in Adults with Chemotherapy-Refractory Acute Lymphoblastic Leukemia. *Science Translational Medicine*, *5*(177), 177ra38–177ra38.  
<http://doi.org/10.1126/scitranslmed.3005930>
- Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., ... Palsson, B. O. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*, *36*(3), 272–281.  
<http://doi.org/10.1038/nbt.4072>
- Buck, M. D., O’Sullivan, D., Klein Geltink, R. I., Curtis, J. D., Chang, C.-H., Sanin, D. E., ... Pearce, E. L. (2016). Mitochondrial Dynamics Controls T Cell Fate through Metabolic Programming. *Cell*, *166*(1), 63–76.  
<http://doi.org/10.1016/j.cell.2016.05.035>
- Burnet, F. (1970). The concept of immunological surveillance. *Progress in Experimental Tumor Research. Fortschritte Der Experimentellen Tumorforschung. Progres de La Recherche Experimentale Des Tumeurs*, *13*, 1–27.
- Butler, D. J., Mozsary, C., Meydan, C., Danko, D., Foxx, J., Rosiene, J., ...

Mason, C. E. (2020). Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique Host Responses, Viral Diversification, and Drug Interactions. *bioRxiv*.

<http://doi.org/10.1101/2020.04.20.048066>

Carretero-González, A., Lora, D., Ghanem, I., Zugazagoitia, J., Castellano, D., Sepúlveda, J. M., ... Velasco, G. de. (2018). Analysis of response rate with ANTI PD1/PD-L1 monoclonal antibodies in advanced solid tumors: A meta-analysis of randomized clinical trials. *Oncotarget*, 9(9), 8706–8715.

<http://doi.org/10.18632/oncotarget.24283>

Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., ... Karp, P. D. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, 38(Database issue), D473–479. <http://doi.org/10.1093/nar/gkp875>

Cavalcanti, A. B., Zampieri, F. G., Rosa, R. G., Azevedo, L. C. P., Veiga, V. C., Avezum, A., ... Coalition Covid-19 Brazil I Investigators. (2020).

Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *N Engl J Med*, 383(21), 2041–2052.

<http://doi.org/10.1056/NEJMoa2019014>

CDC. (2020, February). COVID-19 and Your Health. *Centers for Disease Control and Prevention*. Retrieved from

<https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html>

- Cekic, C., Sag, D., Day, Y.-J., & Linden, J. (2013). Extracellular adenosine regulates naive T cell development and peripheral maintenance. *J Exp Med*, *210*(12), 2693–2706. <http://doi.org/10.1084/jem.20130249>
- Chambers, J. P., Yu, J., Valdes, J. J., & Arulanandam, B. P. (2020). SARS-CoV-2, Early Entry Events. *J Pathog*, *2020*, 9238696. <http://doi.org/10.1155/2020/9238696>
- Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., & Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*, *9*(1), 221–236. <http://doi.org/10.1080/22221751.2020.1719902>
- Chandel, N. S. (2015). Evolution of Mitochondria as Signaling Organelles. *Cell Metab*, *22*(2), 204–206. <http://doi.org/10.1016/j.cmet.2015.05.013>
- Chandel, N. S., Jasper, H., Ho, T. T., & Passequé, E. (2016). Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat Cell Biol*, *18*(8), 823–832. <http://doi.org/10.1038/ncb3385>
- Chang, C.-H., Qiu, J., O’Sullivan, D., Buck, M. D., Noguchi, T., Curtis, J. D., . . . Pearce, E. L. (2015). Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell*, *162*(6), 1229–1241. <http://doi.org/10.1016/j.cell.2015.08.016>
- Chaudhuri, L., Srivastava, R. K., Kos, F., & Shrikant, P. A. (2016). Uncoupling

protein 2 regulates metabolic reprogramming and fate of antigen-stimulated CD8+ T cells. *Cancer Immunol Immunother*, 65(7), 869–874.

<http://doi.org/10.1007/s00262-016-1851-4>

Chen, Q., Allot, A., & Lu, Z. (2021). LitCovid: An open database of COVID-19 literature. *Nucleic Acids Res*, 49(D1), D1534–D1540.

<http://doi.org/10.1093/nar/gkaa952>

Cheng, K., Riva, L., Sinha, S., Pal, L. R., Nair, N. U., Martin-Sancho, L., ...

Ruppin, E. (2021). Genome-scale metabolic modeling reveals SARS-CoV-2-induced host metabolic reprogramming and identifies metabolic antiviral targets. *bioRxiv*, 2021.01.27.428543.

<http://doi.org/10.1101/2021.01.27.428543>

Cheng, W.-C., Tsui, Y.-C., Ragusa, S., Koelzer, V. H., Mina, M., Franco, F., ...

Ho, P.-C. (2019). Uncoupling protein 2 reprograms the tumor microenvironment to support the anti-tumor immune cycle. *Nat Immunol*,

20(2), 206–217. <http://doi.org/10.1038/s41590-018-0290-0>

Chua, R. L., Lukassen, S., Trump, S., Hennig, B. P., Wendisch, D., Pott, F., ...

Eils, R. (2020). COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol*, 38(8),

970–979. <http://doi.org/10.1038/s41587-020-0602-4>

Codo, A. C., Davanzo, G. G., Monteiro, L. de B., Souza, G. F. de, Muraro, S. P.,

Virgilio-da-Silva, J. V., ... Moraes-Vieira, P. M. (2020). Elevated Glucose

Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1 $\alpha$ /Glycolysis-Dependent Axis. *Cell Metab*, 32(3), 437–446.e5.  
<http://doi.org/10.1016/j.cmet.2020.07.007>

Coley, W. B. (1893). THE TREATMENT OF MALIGNANT TUMORS BY REPEATED INOCULATIONS OF ERYSIPELAS: WITH A REPORT OF TEN ORIGINAL CASES.1 - ProQuest. Retrieved from  
<https://search.proquest.com/openview/09fb106c24157c028c895edfa8049551/1?pq-origsite=gscholar&cbl=41361>

Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., ... Galagan, J. E. (2009). Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production. *PLoS Comput Biol*, 5(8), e1000489. <http://doi.org/10.1371/journal.pcbi.1000489>

Corbet, C., & Feron, O. (2017). Tumour acidosis: From the passenger to the driver's seat. *Nat Rev Cancer*, 17(10), 577–593.  
<http://doi.org/10.1038/nrc.2017.77>

Cornel, A. M., Mimpfen, I. L., & Nierkens, S. (2020). MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers (Basel)*, 12(7).  
<http://doi.org/10.3390/cancers12071760>

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species Severe acute respiratory syndrome-related

- coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*, 5(4), 536–544. <http://doi.org/10.1038/s41564-020-0695-z>
- Cortinovis, M., Perico, N., & Remuzzi, G. (2021). Long-term follow-up of recovered patients with COVID-19. *Lancet*, 397(10270), 173–175. [http://doi.org/10.1016/S0140-6736\(21\)00039-8](http://doi.org/10.1016/S0140-6736(21)00039-8)
- Couplan, E., Mar Gonzalez-Barroso, M. del, Alves-Guerra, M. C., Ricquier, D., Goubern, M., & Bouillaud, F. (2002). No evidence for a basal, retinoic, or superoxide-induced uncoupling activity of the uncoupling protein 2 present in spleen or lung mitochondria. *J Biol Chem*, 277(29), 26268–26275. <http://doi.org/10.1074/jbc.M202535200>
- Couzin-Frankel, J. (2013). Cancer Immunotherapy. *Science*, 342(6165), 1432–1433. <http://doi.org/10.1126/science.342.6165.1432>
- Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., ... Kaufman, D. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science*, 362(6411). <http://doi.org/10.1126/science.aar3593>
- Cuevas, D. A., Edirisinghe, J., Henry, C. S., Overbeek, R., O’Connell, T. G., & Edwards, R. A. (2016). From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model. *Front Microbiol*, 7, 907. <http://doi.org/10.3389/fmicb.2016.00907>
- Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic

coronaviruses. *Nat Rev Microbiol*, 17(3), 181–192.

<http://doi.org/10.1038/s41579-018-0118-9>

Daniloski, Z., Jordan, T. X., Wessels, H.-H., Hoagland, D. A., Kasela, S., Legut, M., . . . Sanjana, N. E. (2021). Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. *Cell*, 184(1), 92–105.e16.

<http://doi.org/10.1016/j.cell.2020.10.030>

Demine, S., Renard, P., & Arnould, T. (2019). Mitochondrial Uncoupling: A Key Controller of Biological Processes in Physiology and Diseases. *Cells*, 8(8).

<http://doi.org/10.3390/cells8080795>

Deng, W., Gowen, B. G., Zhang, L., Wang, L., Lau, S., Iannello, A., . . . Raulet, D. H. (2015). Antitumor immunity. A shed NKG2D ligand that promotes natural killer cell activation and tumor rejection. *Science*, 348(6230), 136–139.

<http://doi.org/10.1126/science.1258867>

Desdín-Micó, G., Soto-Heredero, G., & Mittelbrunn, M. (2018). Mitochondrial activity in T cells. *Mitochondrion*, 41, 51–57.

<http://doi.org/10.1016/j.mito.2017.10.006>

Diao, B., Wang, C., Wang, R., Feng, Z., Tan, Y., Wang, H., . . . Chen, Y. (2020). Human Kidney is a Target for Novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection. *medRxiv*, 2020.03.04.20031120.

<http://doi.org/10.1101/2020.03.04.20031120>

Dias, O., Rocha, M., Ferreira, E. C., & Rocha, I. (2015). Reconstructing

- genome-scale metabolic models with merlin. *Nucleic Acids Res*, 43(8), 3899–3910. <http://doi.org/10.1093/nar/gkv294>
- Dighe, A. S., Richards, E., Old, L. J., & Schreiber, R. D. (1994). Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFN gamma receptors. *Immunity*, 1(6), 447–456. [http://doi.org/10.1016/1074-7613\(94\)90087-6](http://doi.org/10.1016/1074-7613(94)90087-6)
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <http://doi.org/10.1093/bioinformatics/bts635>
- Donadelli, M., Dando, I., Fiorini, C., & Palmieri, M. (2014). UCP2, a mitochondrial protein regulated at multiple levels. *Cell. Mol. Life Sci.*, 71(7), 1171–1190. <http://doi.org/10.1007/s00018-013-1407-0>
- Dong, Y., Dai, T., Wei, Y., Zhang, L., Zheng, M., & Zhou, F. (2020). A systematic review of SARS-CoV-2 vaccine candidates. *Signal Transduct Target Ther*, 5(1), 237. <http://doi.org/10.1038/s41392-020-00352-y>
- Doremalen, N. van, Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., ... Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N Engl J Med*, 382(16), 1564–1567. <http://doi.org/10.1056/NEJMc2004973>
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., ...

- Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, *104*(6), 1777–1782. <http://doi.org/10.1073/pnas.0610772104>
- Dudley, M. E., Wunderlich, J. R., Robbins, P. F., Yang, J. C., Hwu, P., Schwartzentruber, D. J., ... Rosenberg, S. A. (2002). Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science*, *298*(5594), 850–854. <http://doi.org/10.1126/science.1076514>
- Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., & Schreiber, R. D. (2002). Cancer immunoediting: From immunosurveillance to tumor escape. *Nat Immunol*, *3*(11), 991–998. <http://doi.org/10.1038/ni1102-991>
- Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004). The three Es of cancer immunoediting. *Annu Rev Immunol*, *22*, 329–360. <http://doi.org/10.1146/annurev.immunol.22.012703.104803>
- Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L., & Denison, M. R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol*, *81*(22), 12135–12144. <http://doi.org/10.1128/JVI.01296-07>
- Ehrlich, A., Uhl, S., Ioannidis, K., Hofree, M., tenOever, B. R., & Nahmias, Y. (2020). *The SARS-CoV-2 Transcriptional Metabolic Signature in Lung Epithelium* (SSRN Scholarly Paper No. ID 3650499). Rochester, NY: Social Science Research Network. Retrieved from

<https://papers.ssrn.com/abstract=3650499>

Esfahani, K., Roudaia, L., Buhlaiga, N., Del Rincon, S., Papneja, N., & Miller, W.

(2020). A review of cancer immunotherapy: From the past, to the present, to the future. *Curr Oncol*, 27(Suppl 2), S87–S97.

<http://doi.org/10.3747/co.27.5223>

Fallahi, S., Skaug, H. J., & Alendal, G. (2020). A comparison of Monte Carlo sampling methods for metabolic network models. *PLoS One*, 15(7), e0235393.

<http://doi.org/10.1371/journal.pone.0235393>

FDA. (2020a, November). Coronavirus (COVID-19) Update: FDA Authorizes Drug Combination for Treatment of COVID-19. *FDA*. Retrieved from

<https://www.fda.gov/news-events/press-announcements/>

[coronavirus-covid-19-update-fda-authorizes-drug-combination-treatment-covid-19](https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-drug-combination-treatment-covid-19)

FDA. (2020b, November). Coronavirus (COVID-19) Update: FDA Authorizes Monoclonal Antibodies for Treatment of COVID-19. *FDA*. Retrieved from

<https://www.fda.gov/news-events/press-announcements/>

[coronavirus-covid-19-update-fda-authorizes-monoclonal-antibodies-treatment-covid-19](https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-monoclonal-antibodies-treatment-covid-19)

FDA. (2020c, November). Coronavirus (COVID-19) Update: FDA Authorizes Monoclonal Antibody for Treatment of COVID-19. *FDA*. Retrieved from

<https://www.fda.gov/news-events/press-announcements/>

[coronavirus-covid-19-update-fda-authorizes-monoclonal-antibody-treatment-covid-19](https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-monoclonal-antibody-treatment-covid-19)

FDA. (2020d, October). FDA Approves First Treatment for COVID-19. *FDA*.

Retrieved from <https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19>

FDA. (2021a). Moderna COVID-19 Vaccine. *FDA*. Retrieved from <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine>

FDA. (2021b). Pfizer-BioNTech COVID-19 Vaccine. *FDA*. Retrieved from <https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine>

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., ... Gottardo, R. (2015). MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*, *16*, 278.  
<http://doi.org/10.1186/s13059-015-0844-5>

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., ... Stern-Ginossar, N. (2021). The coding capacity of SARS-CoV-2. *Nature*, *589*(7840), 125–130.  
<http://doi.org/10.1038/s41586-020-2739-1>

Fox, C. J., Hammerman, P. S., & Thompson, C. B. (2005). Fuel feeds function: Energy metabolism and the T-cell response. *Nat Rev Immunol*, *5*(11), 844–852.  
<http://doi.org/10.1038/nri1710>

Fraietta, J. A., Lacey, S. F., Orlando, E. J., Pruteanu-Malinici, I., Gohil, M.,

- Lundh, S., ... Melenhorst, J. J. (2018). Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat Med*, *24*(5), 563–571.  
<http://doi.org/10.1038/s41591-018-0010-1>
- Franchina, D. G., Dostert, C., & Brenner, D. (2018). Reactive Oxygen Species: Involvement in T Cell Signaling and Metabolism. *Trends Immunol*, *39*(6), 489–502. <http://doi.org/10.1016/j.it.2018.01.005>
- Freeman, G. J., Long, A. J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., ... Honjo, T. (2000). Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med*, *192*(7), 1027–1034. <http://doi.org/10.1084/jem.192.7.1027>
- Ganesan, S., & Mehnert, J. (2020). Biomarkers for Response to Immune Checkpoint Blockade. *Annual Review of Cancer Biology*, *4*(1), 331–351.  
<http://doi.org/10.1146/annurev-cancerbio-030419-033604>
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., ... Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, *368*(6492), 779–782. <http://doi.org/10.1126/science.abb7498>
- Gardinassi, L. G., Souza, C. O. S., Sales-Campos, H., & Fonseca, S. G. (2020). Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. *Front Immunol*, *11*, 1636.  
<http://doi.org/10.3389/fimmu.2020.01636>

- Geiger, R., Rieckmann, J. C., Wolf, T., Basso, C., Feng, Y., Fuhrer, T., . . . Lanzavecchia, A. (2016). L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. *Cell*, *167*(3), 829–842.e13. <http://doi.org/10.1016/j.cell.2016.09.031>
- Gerriets, V. A., Kishton, R. J., Johnson, M. O., Cohen, S., Siska, P. J., Nichols, A. G., . . . Rathmell, J. C. (2016). Foxp3 and Toll-like receptor signaling balance Treg cell anabolic metabolism for suppression. *Nat Immunol*, *17*(12), 1459–1466. <http://doi.org/10.1038/ni.3577>
- Ghosh, S., Dellibovi-Ragheb, T. A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., . . . Altan-Bonnet, N. (2020).  $\beta$ -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell*, *183*(6), 1520–1535.e14. <http://doi.org/10.1016/j.cell.2020.10.039>
- Giamarellos-Bourboulis, E. J., Netea, M. G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., . . . Koutsoukou, A. (2020). Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host Microbe*, *27*(6), 992–1000.e3. <http://doi.org/10.1016/j.chom.2020.04.009>
- Global Coronavirus COVID-19 Clinical Trial Tracker. (2020). Retrieved from <https://www.covid-trials.org/>
- Glont, M., Nguyen, T. V. N., Graesslin, M., Hälke, R., Ali, R., Schramm, J., . . . Hermjakob, H. (2018). BioModels: Expanding horizons to include more modelling approaches and formats. *Nucleic Acids Res*, *46*(D1), D1248–D1253.

<http://doi.org/10.1093/nar/gkx1023>

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., ...

Haussler, D. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*, *38*(6), 675–678.

<http://doi.org/10.1038/s41587-020-0546-8>

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M.,

... Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), 459–468.

<http://doi.org/10.1038/s41586-020-2286-9>

Grupp, S. A., Kalos, M., Barrett, D., Aplenc, R., Porter, D. L., Rheingold, S. R.,

... June, C. H. (2013). Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N Engl J Med*, *368*(16), 1509–1518.

<http://doi.org/10.1056/NEJMoa1215134>

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status

and applications of genome-scale metabolic models. *Genome Biol*, *20*(1), 121.

<http://doi.org/10.1186/s13059-019-1730-3>

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and

correlations in multidimensional genomic data. *Bioinformatics*, *32*(18),

2847–2849. <http://doi.org/10.1093/bioinformatics/btw313>

Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., ... China

Medical Treatment Expert Group for Covid-19. (2020). Clinical Characteristics

of Coronavirus Disease 2019 in China. *N Engl J Med*, 382(18), 1708–1720.

<http://doi.org/10.1056/NEJMoa2002032>

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., ...

Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, 369(6504), 718–724.

<http://doi.org/10.1126/science.abc6027>

Hamilton, J. J., & Reed, J. L. (2014). Software platforms to facilitate

reconstructing genome-scale metabolic networks. *Environ Microbiol*, 16(1), 49–59. <http://doi.org/10.1111/1462-2920.12312>

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next

Generation. *Cell*, 144(5), 646–674.

<http://doi.org/10.1016/j.cell.2011.02.013>

Haraldsdóttir, H. S., Cousins, B., Thiele, I., Fleming, R. M. T., & Vempala, S.

(2017). CHRR: Coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11), 1741–1743.

<http://doi.org/10.1093/bioinformatics/btx052>

Havel, J. J., Chowell, D., & Chan, T. A. (2019). The evolving landscape of

biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*, 19(3), 133–150. <http://doi.org/10.1038/s41568-019-0116-x>

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., & Stevens,

R. L. (2010). High-throughput generation, optimization and analysis of

genome-scale metabolic models. *Nat Biotechnol*, 28(9), 977–982.

<http://doi.org/10.1038/nbt.1672>

Hewitt, E. W. (2003). The MHC class I antigen presentation pathway: Strategies for viral immune evasion. *Immunology*, 110(2), 163–169.

<http://doi.org/10.1046/j.1365-2567.2003.01738.x>

Hinshaw, D. C., & Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res*, 79(18), 4557–4566.

<http://doi.org/10.1158/0008-5472.CAN-18-3962>

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*,

181(2), 271–280.e8. <http://doi.org/10.1016/j.cell.2020.02.052>

Hornigold, N., Dunn, K. R., Craven, R. A., Zougman, A., Trainor, S., Shreeve, R., ... Vasudev, N. S. (2020). Dysregulation at multiple points of the kynurenine pathway is a ubiquitous feature of renal cancer: Implications for tumour immune evasion. *British Journal of Cancer*, 123(1), 137–147.

<http://doi.org/10.1038/s41416-020-0874-y>

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*.

<http://doi.org/10.1038/s41579-020-00459-7>

Institute, C. R. (2021). PD-1 / PD-L1 Landscape. *Cancer Research Institute*.

Retrieved from <https://www.cancerresearch.org/scientists/immuno-oncology-landscape/pd-1-pd-l1-landscape>

- Jafferji, M. S., & Yang, J. C. (2019). Adoptive T-Cell Therapy for Solid Malignancies. *Surgical Oncology Clinics of North America*, *28*(3), 465–479. <http://doi.org/10.1016/j.soc.2019.02.012>
- Jamialahmadi, O., Hashemi-Najafabadi, S., Motamedian, E., Romeo, S., & Bagheri, F. (2019). A benchmark-driven approach to reconstruct metabolic networks for studying cancer metabolism. *PLOS Computational Biology*, *15*(4), e1006936. <http://doi.org/10.1371/journal.pcbi.1006936>
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., . . . D'Eustachio, P. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res*, *48*(D1), D498–D503. <http://doi.org/10.1093/nar/gkz1031>
- Jerby, L., Shlomi, T., & Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Mol Syst Biol*, *6*, 401. <http://doi.org/10.1038/msb.2010.56>
- Jiang, S., Du, L., & Shi, Z. (2020). An emerging coronavirus causing pneumonia outbreak in Wuhan, China: Calling for developing therapeutic and prophylactic strategies. *Emerg Microbes Infect*, *9*(1), 275–277. <http://doi.org/10.1080/22221751.2020.1723441>
- Jiang, X., Xu, J., Liu, M., Xing, H., Wang, Z., Huang, L., . . . Wu, S. (2019). Adoptive CD8+ T cell therapy against cancer:Challenges and opportunities.

*Cancer Letters*, 462, 23–32. <http://doi.org/10.1016/j.canlet.2019.07.017>

Jin, M.-Z., & Jin, W.-L. (2020). The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduction and Targeted Therapy*, 5(1), 1–16. <http://doi.org/10.1038/s41392-020-00280-x>

Jordan, R. E., & Adab, P. (2020). Who is most likely to be infected with SARS-CoV-2? *Lancet Infect Dis*, 20(9), 995–996. [http://doi.org/10.1016/S1473-3099\(20\)30395-9](http://doi.org/10.1016/S1473-3099(20)30395-9)

June, C. H., & Sadelain, M. (2018). Chimeric Antigen Receptor Therapy. *New England Journal of Medicine*, 379(1), 64–73. <http://doi.org/10.1056/NEJMra1706169>

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res*, 49(D1), D545–D551. <http://doi.org/10.1093/nar/gkaa970>

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, 40(Database issue), D109–114. <http://doi.org/10.1093/nar/gkr988>

Kaplan, D. H., Shankaran, V., Dighe, A. S., Stockert, E., Aguet, M., Old, L. J., & Schreiber, R. D. (1998). Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci U S A*, 95(13), 7556–7561. <http://doi.org/10.1073/pnas.95.13.7556>

Karlsen, E., Schulz, C., & Almaas, E. (2018). Automated generation of

- genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*, 19(1), 467. <http://doi.org/10.1186/s12859-018-2472-z>
- Kaufman, D. E., & Smith, R. L. (1998). Direction Choice for Accelerated Convergence in Hit-and-Run Sampling. *Operations Research*, 46(1), 84–95. <http://doi.org/10.1287/opre.46.1.84>
- Khodaei, S., Asgari, Y., Totonchi, M., & Karimi-Jafari, M. H. (2020). iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model. *Sci Rep*, 10(1), 6177. <http://doi.org/10.1038/s41598-020-63235-w>
- Kim, M. K., & Lun, D. S. (2014). Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput Struct Biotechnol J*, 11(18), 59–65. <http://doi.org/10.1016/j.csbj.2014.08.009>
- King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., ... Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*, 44(D1), D515–522. <http://doi.org/10.1093/nar/gkv1049>
- Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493), 860–868. <http://doi.org/10.1126/science.abb5793>
- Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., ... Chlanda, P. (2020). SARS-CoV-2 structure and replication

characterized by in situ cryo-electron tomography. *Nat Commun*, 11(1), 5885.

<http://doi.org/10.1038/s41467-020-19619-7>

Klein Geltink, R. I., Kyle, R. L., & Pearce, E. L. (2018). Unraveling the Complex Interplay Between T Cell Metabolism and Function. *Annual Review of Immunology*, 36(1), 461–488.

<http://doi.org/10.1146/annurev-immunol-042617-053019>

Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., . . . Sato, K. (2020). SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Rep*, 32(12), 108185. <http://doi.org/10.1016/j.celrep.2020.108185>

Koolman, J., & Roehm, K. H. (2011). *Color Atlas of Biochemistry*. Thieme.

Koppenol, W. H., Bounds, P. L., & Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*, 11(5), 325–337. <http://doi.org/10.1038/nrc3038>

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012. <http://doi.org/10.1101/060012>

Krauss, S., Zhang, C.-Y., & Lowell, B. B. (2002). A significant portion of mitochondrial proton leak in intact thymocytes depends on expression of UCP2. *Proc Natl Acad Sci U S A*, 99(1), 118–122.

<http://doi.org/10.1073/pnas.012410699>

- Krauss, S., Zhang, C.-Y., & Lowell, B. B. (2005). The mitochondrial uncoupling-protein homologues. *Nat Rev Mol Cell Biol*, 6(3), 248–261.  
<http://doi.org/10.1038/nrm1592>
- Kuleshov, M. V., Stein, D. J., Clarke, D. J. B., Kropiwnicki, E., Jagodnik, K. M., Bartal, A., ... Ma'ayan, A. (2020). The COVID-19 Drug and Gene Set Library. *Patterns (N Y)*, 1(6), 100090.  
<http://doi.org/10.1016/j.patter.2020.100090>
- Lam, T. T.-Y., Jia, N., Zhang, Y.-W., Shum, M. H.-H., Jiang, J.-F., Zhu, H.-C., ... Cao, W.-C. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, 583(7815), 282–285.  
<http://doi.org/10.1038/s41586-020-2169-0>
- Lau, A. N., & Vander Heiden, M. G. (2020). Metabolism in the Tumor Microenvironment. *Annu. Rev. Cancer Biol.*, 4(1), 17–40.  
<http://doi.org/10.1146/annurev-cancerbio-030419-033333>
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*, 172(9), 577–582.  
<http://doi.org/10.7326/M20-0504>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*,

15(2), R29. <http://doi.org/10.1186/gb-2014-15-2-r29>

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., ... Diaz, L. A. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*, *357*(6349), 409–413.  
<http://doi.org/10.1126/science.aan6733>

Leach, D. R., Krummel, M. F., & Allison, J. P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science*, *271*(5256), 1734–1736.  
<http://doi.org/10.1126/science.271.5256.1734>

Lee, D., Smallbone, K., Dunn, W. B., Murabito, E., Winder, C. L., Kell, D. B., ... Swainston, N. (2012). Improving metabolic flux predictions using absolute gene expression data. *BMC Syst Biol*, *6*, 73.  
<http://doi.org/10.1186/1752-0509-6-73>

Lee, J., Walsh, M. C., Hoehn, K. L., James, D. E., Wherry, E. J., & Choi, Y. (2014). Regulator of fatty acid metabolism, acetyl coenzyme a carboxylase 1, controls T cell immunity. *J Immunol*, *192*(7), 3190–3199.  
<http://doi.org/10.4049/jimmunol.1302985>

Li, G., Li, W., He, X., & Cao, Y. (2020). Asymptomatic and Presymptomatic Infectors: Hidden Sources of Coronavirus Disease 2019 (COVID-19). *Clin Infect Dis*, *71*(8), 2018. <http://doi.org/10.1093/cid/ciaa418>

Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., ... Liang, Q. (2021). Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence

- Factors Influencing SARS-CoV-2 Pathogenesis. *Med (N Y)*, 2(1), 99–112.e7.  
<http://doi.org/10.1016/j.medj.2020.07.002>
- Li, S., Ma, F., Yokota, T., Garcia, G., Palermo, A., Wang, Y., ... Deb, A. (2021).  
Metabolic reprogramming and epigenetic changes of vital organs in  
SARS-CoV-2-induced systemic toxicity. *JCI Insight*, 6(2).  
<http://doi.org/10.1172/jci.insight.145027>
- Li, W., & Zhang, L. (2020). Rewiring Mitochondrial Metabolism for CD8+ T Cell  
Memory Formation and Effective Cancer Immunotherapy. *Front. Immunol.*, 11.  
<http://doi.org/10.3389/fimmu.2020.01834>
- Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., ... Zhang, Z. (2020).  
Single-cell landscape of bronchoalveolar immune cells in patients with  
COVID-19. *Nat Med*, 26(6), 842–844.  
<http://doi.org/10.1038/s41591-020-0901-9>
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., &  
Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0.  
*Bioinformatics*, 27(12), 1739–1740.  
<http://doi.org/10.1093/bioinformatics/btr260>
- Lieberman, N. A. P., Peddu, V., Xie, H., Shrestha, L., Huang, M.-L., Mears, M.  
C., ... Greninger, A. L. (2020). In vivo antiviral host transcriptional response  
to SARS-CoV-2 by viral load, sex, and age. *PLoS Biol*, 18(9), e3000849.  
<http://doi.org/10.1371/journal.pbio.3000849>

- Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., ... Zhang, C. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*, *38*(3), 272–276.  
<http://doi.org/10.1038/s41587-020-0446-y>
- Lim, S., Bae, J. H., Kwon, H.-S., & Nauck, M. A. (2021). COVID-19 and diabetes mellitus: From pathophysiology to clinical management. *Nat Rev Endocrinol*, *17*(1), 11–30. <http://doi.org/10.1038/s41574-020-00435-4>
- Liochev, S. I. (2013). Reactive oxygen species and the free radical theory of aging. *Free Radic Biol Med*, *60*, 1–4.  
<http://doi.org/10.1016/j.freeradbiomed.2013.02.011>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550.  
<http://doi.org/10.1186/s13059-014-0550-8>
- Lu, Y.-C., Jia, L., Zheng, Z., Tran, E., Robbins, P. F., & Rosenberg, S. A. (2019). Single-Cell Transcriptome Analysis Reveals Gene Signatures Associated with T-cell Persistence Following Adoptive Cell Therapy. *Cancer Immunol Res*, *7*(11), 1824–1836. <http://doi.org/10.1158/2326-6066.CIR-19-0299>
- Ma, E. H., Bantug, G., Griss, T., Condotta, S., Johnson, R. M., Samborska, B., ... Jones, R. G. (2017). Serine Is an Essential Metabolite for Effector T Cell Expansion. *Cell Metab*, *25*(2), 345–357.  
<http://doi.org/10.1016/j.cmet.2016.12.011>

- Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., & Goryanin, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*, *3*, 135.  
<http://doi.org/10.1038/msb4100177>
- Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res*, *46*(15), 7542–7553.  
<http://doi.org/10.1093/nar/gky537>
- Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett*, *583*(24), 3966–3973.  
<http://doi.org/10.1016/j.febslet.2009.10.036>
- Mak, T. W., Grusdat, M., Duncan, G. S., Dostert, C., Nonnenmacher, Y., Cox, M., ... Brenner, D. (2017). Glutathione Primes T Cell Metabolism for Inflammation. *Immunity*, *46*(4), 675–689.  
<http://doi.org/10.1016/j.immuni.2017.03.019>
- Makowski, L., Chaib, M., & Rathmell, J. C. (2020). Immunometabolism: From basic mechanisms to translation. *Immunological Reviews*, *295*(1), 5–14.  
<http://doi.org/https://doi.org/10.1111/imr.12858>
- Mantovani, A., Marchesi, F., Malesci, A., Laghi, L., & Allavena, P. (2017). Tumour-associated macrophages as treatment targets in oncology. *Nature Reviews Clinical Oncology*, *14*(7), 399–416.

<http://doi.org/10.1038/nrclinonc.2016.217>

Mao, B., Liu, Y., Chai, Y.-H., Jin, X.-Y., Lu, H.-W., Yang, J.-W., ... Xu, J.-F. (2020). Assessing risk factors for SARS-CoV-2 infection in patients presenting with symptoms in Shanghai, China: A multicentre, observational cohort study. *Lancet Digit Health*, 2(6), e323–e330.

[http://doi.org/10.1016/S2589-7500\(20\)30109-6](http://doi.org/10.1016/S2589-7500(20)30109-6)

Mao, L., Jin, H., Wang, M., Hu, Y., Chen, S., He, Q., ... Hu, B. (2020). Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol*, 77(6), 683–690.

<http://doi.org/10.1001/jamaneuro.2020.1127>

Marazuela, M., Giustina, A., & Puig-Domingo, M. (2020). Endocrine and metabolic aspects of the COVID-19 pandemic. *Rev Endocr Metab Disord*, 21(4), 495–507. <http://doi.org/10.1007/s11154-020-09569-2>

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., ... Nielsen, J. (2013). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol*, 9, 649.

<http://doi.org/10.1038/msb.2013.5>

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun*, 5,

3083. <http://doi.org/10.1038/ncomms4083>

- Marhelava, K., Pilch, Z., Bajor, M., Graczyk-Jarzynka, A., & Zagodzdzon, R. (2019). Targeting Negative and Positive Immune Checkpoints with Monoclonal Antibodies in Therapy of Cancer. *Cancers (Basel)*, *11*(11). <http://doi.org/10.3390/cancers11111756>
- Marin-Acevedo, J. A., Dholaria, B., Soyano, A. E., Knutson, K. L., Chumsri, S., & Lou, Y. (2018). Next generation of immune checkpoint therapy in cancer: New developments and challenges. *J Hematol Oncol*, *11*(1), 39. <http://doi.org/10.1186/s13045-018-0582-8>
- Mastelic-Gavillet, B., Navarro Rodrigo, B., Décombaz, L., Wang, H., Ercolano, G., Ahmed, R., . . . Vigano, S. (2019). Adenosine mediates functional and metabolic suppression of peripheral and tumor-infiltrating CD8+ T cells. *J Immunother Cancer*, *7*(1), 257. <http://doi.org/10.1186/s40425-019-0719-5>
- Mayer, K. A., Stöckl, J., Zlabinger, G. J., & Gualdoni, G. A. (2019). Hijacking the Supplies: Metabolism as a Novel Facet of Virus-Host Interaction. *Front Immunol*, *10*, 1533. <http://doi.org/10.3389/fimmu.2019.01533>
- Megchelenbrink, W., Huynen, M., & Marchiori, E. (2014). optGpSampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One*, *9*(2), e86587. <http://doi.org/10.1371/journal.pone.0086587>
- Mehta, M. M., Weinberg, S. E., & Chandel, N. S. (2017). Mitochondrial control of immunity: Beyond ATP. *Nat Rev Immunol*, *17*(10), 608–620.

<http://doi.org/10.1038/nri.2017.66>

- Mendoza, S. N., Olivier, B. G., Molenaar, D., & Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, *20*(1), 158. <http://doi.org/10.1186/s13059-019-1769-1>
- Meric-Bernstam, F., Larkin, J., Tabernero, J., & Bonini, C. (2020). Enhancing anti-tumour efficacy with immunotherapy combinations. *The Lancet*, *0*(0). [http://doi.org/10.1016/S0140-6736\(20\)32598-8](http://doi.org/10.1016/S0140-6736(20)32598-8)
- Metzler, B., Gfeller, P., & Guinet, E. (2016). Restricting Glutamine or Glutamine-Dependent Purine and Pyrimidine Syntheses Promotes Human T Cells with High FOXP3 Expression and Regulatory Properties. *J Immunol*, *196*(9), 3618–3630. <http://doi.org/10.4049/jimmunol.1501756>
- Miao, D., Margolis, C. A., Gao, W., Voss, M. H., Li, W., Martini, D. J., ... Van Allen, E. M. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science*, *359*(6377), 801–806. <http://doi.org/10.1126/science.aan5951>
- Mok, T. S. K., Wu, Y.-L., Kudaba, I., Kowalski, D. M., Cho, B. C., Turna, H. Z., ... KEYNOTE-042 Investigators. (2019). Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): A randomised, open-label, controlled, phase 3 trial. *Lancet*, *393*(10183), 1819–1830. [http://doi.org/10.1016/S0140-6736\(18\)32409-7](http://doi.org/10.1016/S0140-6736(18)32409-7)

- Morgan, R. A., Dudley, M. E., Wunderlich, J. R., Hughes, M. S., Yang, J. C., Sherry, R. M., ... Rosenberg, S. A. (2006). Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science*, *314*(5796), 126–129. <http://doi.org/10.1126/science.1129003>
- Motamedian, E., Mohammadi, M., Shojaosadati, S. A., & Heydari, M. (2017). TRFBA: An algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. *Bioinformatics*, *33*(7), 1057–1063. <http://doi.org/10.1093/bioinformatics/btw772>
- NCI. (2015, April). Immunotherapy for Cancer - National Cancer Institute. *cgvArticle*. Retrieved from <https://www.cancer.gov/about-cancer/treatment/types/immunotherapy>
- Nelson, D. L., Cox, M. M., & Lehninger, A. L. (2017). *Lehninger principles of biochemistry*. Retrieved from <http://catalog.hathitrust.org/api/volumes/oclc/986827885.html>
- Nguyenla, X., Wehri, E., Dis, E. V., Biering, S. B., Yamashiro, L. H., Stroumza, J., ... Schaletzky, J. (2020). Discovery of SARS-CoV-2 antiviral synergy between remdesivir and approved drugs in human lung cells. *bioRxiv*, 2020.09.18.302398. <http://doi.org/10.1101/2020.09.18.302398>
- NIH. (2020). Corticosteroids. *COVID-19 Treatment Guidelines*. Retrieved from <https://www.covid19treatmentguidelines.nih.gov/immune-based-therapy/immunomodulators/corticosteroids/>

- Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., & Freeman, T. C. (2018). Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunol Res*, *6*(11), 1388–1400.  
<http://doi.org/10.1158/2326-6066.CIR-18-0342>
- Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D. C., & Lewis, N. E. (2017). A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst*, *4*(3), 318–329.e6.  
<http://doi.org/10.1016/j.cels.2017.01.010>
- Ott, P. A., Bang, Y.-J., Piha-Paul, S. A., Razak, A. R. A., Bennouna, J., Soria, J.-C., . . . Luceford, J. K. (2019). T-Cell-Inflamed Gene-Expression Profile, Programmed Death Ligand 1 Expression, and Tumor Mutational Burden Predict Efficacy in Patients Treated With Pembrolizumab Across 20 Cancers: KEYNOTE-028. *J Clin Oncol*, *37*(4), 318–327.  
<http://doi.org/10.1200/JCO.2018.78.2276>
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., . . . Qian, Z. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun*, *11*(1), 1620.  
<http://doi.org/10.1038/s41467-020-15562-9>
- Pacheco, M. P., John, E., Kaoma, T., Heinäniemi, M., Nicot, N., Vallar, L., . . . Sauter, T. (2015). Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network. *BMC*

*Genomics*, 16, 809. <http://doi.org/10.1186/s12864-015-1984-4>

Palsson, B. Ø. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge: Cambridge University Press.

<http://doi.org/10.1017/CB09781139854610>

Pastor-Barriuso, R., Pérez-Gómez, B., Hernán, M. A., Pérez-Olmeda, M., Yotti, R., Oteo-Iglesias, J., ... ENE-COVID Study Group. (2020). Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: Nationwide seroepidemiological study. *BMJ*, 371, m4509.

<http://doi.org/10.1136/bmj.m4509>

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 14(4), 417–419. <http://doi.org/10.1038/nmeth.4197>

<http://doi.org/10.1038/nmeth.4197>

Patsoukis, N., Bardhan, K., Chatterjee, P., Sari, D., Liu, B., Bell, L. N., ...

Boussiotis, V. A. (2015). PD-1 alters T-cell metabolic reprogramming by inhibiting glycolysis and promoting lipolysis and fatty acid oxidation. *Nat Commun*, 6, 6692. <http://doi.org/10.1038/ncomms7692>

<http://doi.org/10.1038/ncomms7692>

Pearce, E. L., Walsh, M. C., Cejas, P. J., Harms, G. M., Shen, H., Wang, L.-S., ...

Choi, Y. (2009). Enhancing CD8 T-cell memory by modulating fatty acid metabolism. *Nature*, 460(7251), 103–107.

<http://doi.org/10.1038/nature08097>

Peng, M., Yin, N., Chhangawala, S., Xu, K., Leslie, C. S., & Li, M. O. (2016).

Aerobic glycolysis promotes T helper 1 cell differentiation through an epigenetic mechanism. *Science*, 354(6311), 481–484.

<http://doi.org/10.1126/science.aaf6284>

Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., . . .

Simonsen, L. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis*, 20(9), e238–e244.

[http://doi.org/10.1016/S1473-3099\(20\)30484-9](http://doi.org/10.1016/S1473-3099(20)30484-9)

Petrilli, C. M., Jones, S. A., Yang, J., Rajagopalan, H., O'Donnell, L., Chernyak,

Y., . . . Horwitz, L. I. (2020). Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: Prospective cohort study. *BMJ*, 369, m1966.

<http://doi.org/10.1136/bmj.m1966>

Pilipow, K., Scamardella, E., Puccio, S., Gautam, S., Paoli, F. D., Mazza, E. M.

C., . . . Lugli, E. (2018). Antioxidant metabolism regulates CD8<sup>+</sup> T memory stem cell formation and antitumor immunity. *JCI Insight*, 3(18).

<http://doi.org/10.1172/jci.insight.122299>

Plitas, G., & Rudensky, A. Y. (2020). Regulatory T Cells in Cancer. *Annual*

*Review of Cancer Biology*, 4(1), 459–477.

<http://doi.org/10.1146/annurev-cancerbio-030419-033428>

Porter, D. L., Levine, B. L., Kalos, M., Bagg, A., & June, C. H. (2011). Chimeric

antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med*,

365(8), 725–733. <http://doi.org/10.1056/NEJMoa1103849>

Powell, J. D., Pollizzi, K. N., Heikamp, E. B., & Horton, M. R. (2012). Regulation of immune responses by mTOR. *Annu Rev Immunol*, 30, 39–68.  
<http://doi.org/10.1146/annurev-immunol-020711-075024>

Procaccini, C., Carbone, F., Di Silvestre, D., Brambilla, F., De Rosa, V., Galgani, M., . . . Matarese, G. (2016). The Proteomic Landscape of Human Ex Vivo Regulatory and Conventional T Cells Reveals Specific Metabolic Requirements. *Immunity*, 44(3), 712. <http://doi.org/10.1016/j.immuni.2016.02.022>

Qin, S., Xu, L., Yi, M., Yu, S., Wu, K., & Luo, S. (2019). Novel immune checkpoint targets: Moving beyond PD-1 and CTLA-4. *Mol Cancer*, 18(1), 155.  
<http://doi.org/10.1186/s12943-019-1091-2>

Rabinovich, G. A., Gabrilovich, D., & Sotomayor, E. M. (2007). Immunosuppressive strategies that are mediated by tumor cells. *Annu Rev Immunol*, 25, 267–296.  
<http://doi.org/10.1146/annurev.immunol.25.022106.141609>

RECOVERY Collaborative Group. (2020). Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial. *Lancet*.  
[http://doi.org/10.1016/S0140-6736\(20\)32013-4](http://doi.org/10.1016/S0140-6736(20)32013-4)

RECOVERY Collaborative Group, Horby, P., Lim, W. S., Emberson, J. R., Mafham, M., Bell, J. L., . . . Landray, M. J. (2020). Dexamethasone in

Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*.

<http://doi.org/10.1056/NEJMoa2021436>

Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., ...

Chan, T. A. (2017). Tumor and Microenvironment Evolution during

Immunotherapy with Nivolumab. *Cell*, *171*(4), 934–949.e16.

<http://doi.org/10.1016/j.cell.2017.09.028>

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K.

(2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, *43*(7), e47.

<http://doi.org/10.1093/nar/gkv007>

Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., ...

Chanda, S. K. (2020). Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature*, *586*(7827), 113–119.

<http://doi.org/10.1038/s41586-020-2577-1>

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.

J., ... Chan, T. A. (2015). Cancer immunology. Mutational landscape

determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*,

*348*(6230), 124–128. <http://doi.org/10.1126/science.aaa1348>

Robaina Estévez, S., & Nikoloski, Z. (2015). Context-Specific Metabolic Model

Extraction Based on Regularized Least Squares Optimization. *PLoS One*,

*10*(7), e0131875. <http://doi.org/10.1371/journal.pone.0131875>

- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77.  
<http://doi.org/10.1186/1471-2105-12-77>
- Robinson, J. L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., Nilsson, A., ... Nielsen, J. (2020). An atlas of human metabolism. *Sci Signal*, *13*(624).  
<http://doi.org/10.1126/scisignal.aaz1482>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140.  
<http://doi.org/10.1093/bioinformatics/btp616>
- Rodig, S. J., Gusenleitner, D., Jackson, D. G., Gjini, E., Giobbie-Hurder, A., Jin, C., ... Hodi, F. S. (2018). MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci Transl Med*, *10*(450). <http://doi.org/10.1126/scitranslmed.aar3342>
- Ron-Harel, N., Ghergurovich, J. M., Notarangelo, G., LaFleur, M. W., Tsubosaka, Y., Sharpe, A. H., ... Haigis, M. C. (2019). T Cell Activation Depends on Extracellular Alanine. *Cell Rep*, *28*(12), 3011–3021.e4.  
<http://doi.org/10.1016/j.celrep.2019.08.034>
- Rosenberg, S. A., Packard, B. S., Aebersold, P. M., Solomon, D., Topalian, S. L., Toy, S. T., ... Seipp, C. A. (1988). Use of tumor-infiltrating lymphocytes and

interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N Engl J Med*, 319(25), 1676–1680.

<http://doi.org/10.1056/NEJM198812223192527>

Rosenberg, S. A., Restifo, N. P., Yang, J. C., Morgan, R. A., & Dudley, M. E.

(2008). Adoptive cell transfer: A clinical path to effective cancer immunotherapy. *Nature Reviews Cancer*, 8(4), 299–308.

<http://doi.org/10.1038/nrc2355>

Rupprecht, A., Bräuer, A. U., Smorodchenko, A., Goyn, J., Hilde, K. E.,

Shabalina, I. G., ... Pohl, E. E. (2012). Quantification of uncoupling protein 2 reveals its main expression in immune cells and selective up-regulation during T-cell proliferation. *PLoS One*, 7(8), e41406.

<http://doi.org/10.1371/journal.pone.0041406>

Saa, P. A., & Nielsen, L. K. (2016). LI-ACHRB: A scalable algorithm for sampling

the feasible solution space of metabolic networks. *Bioinformatics*, 32(15),

2330–2337. <http://doi.org/10.1093/bioinformatics/btw132>

Sade-Feldman, M., Yizhak, K., Bjorgaard, S. L., Ray, J. P., Boer, C. G. de,

Jenkins, R. W., ... Hacohen, N. (2018). Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell*, 175(4),

998–1013.e20. <http://doi.org/10.1016/j.cell.2018.10.038>

Sahu, A. D., S Lee, J., Wang, Z., Zhang, G., Iglesias-Bartolome, R., Tian, T., ...

Ruppin, E. (2019). Genome-wide prediction of synthetic rescue mediators of

resistance to targeted and immunotherapy. *Mol Syst Biol*, 15(3), e8323.

<http://doi.org/10.15252/msb.20188323>

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., ... Morris, L. G. T. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*, 51(2), 202–206. <http://doi.org/10.1038/s41588-018-0312-8>

Sanchez, E. L., & Lagunoff, M. (2015). Viral activation of cellular metabolism. *Virology*, 479-480, 609–618. <http://doi.org/10.1016/j.virol.2015.02.038>

Santos, W. G. dos. (2020). Natural history of COVID-19 and current knowledge on treatment therapeutic options. *Biomedicine & Pharmacotherapy*, 129, 110493. <http://doi.org/10.1016/j.biopha.2020.110493>

Scharping, N. E., Menk, A. V., Moreci, R. S., Whetstone, R. D., Dadey, R. E., Watkins, S. C., ... Delgoffe, G. M. (2016). The Tumor Microenvironment Represses T Cell Mitochondrial Biogenesis to Drive Intratumoral T Cell Metabolic Insufficiency and Dysfunction. *Immunity*, 45(2), 374–388. <http://doi.org/10.1016/j.immuni.2016.07.009>

Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson, B. Ø., & Hyduke, D. R. (2013). GIM3E: Condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, 29(22), 2900–2908. <http://doi.org/10.1093/bioinformatics/btt493>

Schneider, W. M., Luna, J. M., Hoffmann, H.-H., Sánchez-Rivera, F. J., Leal, A.

- A., Ashbrook, A. W., ... Poirier, J. T. (2021). Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks. *Cell*, *184*(1), 120–132.e14. <http://doi.org/10.1016/j.cell.2020.12.006>
- Schultz, A., & Qutub, A. A. (2016). Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLoS Comput Biol*, *12*(3), e1004808. <http://doi.org/10.1371/journal.pcbi.1004808>
- Schumacher, T. N., & Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science*, *348*(6230), 69–74. <http://doi.org/10.1126/science.aaa4971>
- Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *PNAS*, *99*(23), 15112–15117. <http://doi.org/10.1073/pnas.232349399>
- Sena, L. A., Li, S., Jairaman, A., Prakriya, M., Ezponda, T., Hildeman, D. A., ... Chandel, N. S. (2013). Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity*, *38*(2), 225–236. <http://doi.org/10.1016/j.immuni.2012.10.020>
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A*, *117*(21), 11727–11734. <http://doi.org/10.1073/pnas.2003138117>
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., ... Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807),

221–224. <http://doi.org/10.1038/s41586-020-2179-y>

Shankaran, V., Ikeda, H., Bruce, A. T., White, J. M., Swanson, P. E., Old, L. J., & Schreiber, R. D. (2001). IFN $\gamma$  and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature*, *410*(6832), 1107–1111. <http://doi.org/10.1038/35074122>

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., . . . Guo, T. (2020). Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*, *182*(1), 59–72.e15. <http://doi.org/10.1016/j.cell.2020.05.032>

Shi, S., Qin, M., Shen, B., Cai, Y., Liu, T., Yang, F., . . . Huang, C. (2020). Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. *JAMA Cardiol*, *5*(7), 802–810. <http://doi.org/10.1001/jamacardio.2020.0950>

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., & Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*, *26*(9), 1003–1010. <http://doi.org/10.1038/nbt.1487>

Sigurdsson, M. I., Jamshidi, N., Steingrímsson, E., Thiele, I., & Palsson, B. Ø. (2010). A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol*, *4*, 140. <http://doi.org/10.1186/1752-0509-4-140>

Sivan, A., Corrales, L., Hubert, N., Williams, J. B., Aquino-Michaels, K., Earley, Z. M., . . . Gajewski, T. F. (2015). Commensal Bifidobacterium promotes

- antitumor immunity and facilitates anti-PD-L1 efficacy. *Science*, *350*(6264), 1084–1089. <http://doi.org/10.1126/science.aac4255>
- Sola, I., Almazán, F., Zúñiga, S., & Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*, *2*(1), 265–288. <http://doi.org/10.1146/annurev-virology-100114-055218>
- Song, H.-S., Reifman, J., & Wallqvist, A. (2014). Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. *PLoS One*, *9*(11), e112524. <http://doi.org/10.1371/journal.pone.0112524>
- Spinner, C. D., Gottlieb, R. L., Criner, G. J., Arribas López, J. R., Cattelan, A. M., Soriano Viladomiu, A., ... GS-US-540-5774 Investigators. (2020). Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients With Moderate COVID-19: A Randomized Clinical Trial. *JAMA*, *324*(11), 1048–1057. <http://doi.org/10.1001/jama.2020.16349>
- Spiotto, M. T., Rowley, D. A., & Schreiber, H. (2004). Bystander elimination of antigen loss variants in established tumors. *Nat Med*, *10*(3), 294–298. <http://doi.org/10.1038/nm999>
- Sreedhar, A., Cassell, T., Smith, P., Lu, D., Nam, H. W., Lane, A. N., & Zhao, Y. (2019). UCP2 Overexpression Redirects Glucose into Anabolic Metabolic Pathways. *Proteomics*, *19*(4), e1800353. <http://doi.org/10.1002/pmic.201800353>
- Streeck, H., Schulte, B., Kümmerer, B. M., Richter, E., Höller, T., Fuhrmann, C.,

- ... Hartmann, G. (2020). Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. *Nat Commun*, 11(1), 5829.  
<http://doi.org/10.1038/s41467-020-19509-y>
- Street, S. E., Cretney, E., & Smyth, M. J. (2001). Perforin and interferon-gamma activities independently control tumor initiation, growth, and metastasis. *Blood*, 97(1), 192–197. <http://doi.org/10.1182/blood.v97.1.192>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., ... Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888–1902.e21. <http://doi.org/10.1016/j.cell.2019.05.031>
- Stukalov, A., Girault, V., Grass, V., Bergant, V., Karayel, O., Urban, C., ... Pichlmair, A. (2020). Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv*, 2020.06.17.156455.  
<http://doi.org/10.1101/2020.06.17.156455>
- Styczyński, J. (2020). A brief history of CAR-T cells: From laboratory to the bedside. *Acta Haematologica Polonica*, 51(1), 2–5.  
<http://doi.org/10.2478/ahp-2020-0002>
- Styr, B., Gonen, N., Zarhin, D., Ruggiero, A., Atsmon, R., Gazit, N., ... Slutsky, I. (2019). Mitochondrial Regulation of the Hippocampal Firing Rate Set Point and Seizure Susceptibility. *Neuron*, 102(5), 1009–1024.e8.  
<http://doi.org/10.1016/j.neuron.2019.03.045>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L.,

- Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, *102*(43), 15545–15550. <http://doi.org/10.1073/pnas.0506580102>
- Suda, T., Takubo, K., & Semenza, G. L. (2011). Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell*, *9*(4), 298–310. <http://doi.org/10.1016/j.stem.2011.09.010>
- Sukumar, M., Liu, J., Ji, Y., Subramanian, M., Crompton, J. G., Yu, Z., ... Gattinoni, L. (2013). Inhibiting glycolytic metabolism enhances CD8+ T cell memory and antitumor function. *J Clin Invest*, *123*(10), 4479–4488. <http://doi.org/10.1172/JCI69589>
- Sun, C., Mezzadra, R., & Schumacher, T. N. (2018). Regulation and Function of the PD-L1 Checkpoint. *Immunity*, *48*(3), 434–452. <http://doi.org/10.1016/j.immuni.2018.03.014>
- Swainston, N., Smallbone, K., Mendes, P., Kell, D., & Paton, N. (2011). The SuBliMinaL Toolbox: Automating steps in the reconstruction of metabolic networks. *J Integr Bioinform*, *8*(2), 186. <http://doi.org/10.2390/biecoll-jib-2011-186>
- Swamy, M., Pathak, S., Grzes, K. M., Damerow, S., Sinclair, L. V., Aalten, D. M. F. van, & Cantrell, D. A. (2016). Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy. *Nat Immunol*, *17*(6), 712–720. <http://doi.org/10.1038/ni.3439>

- Tahara, E. B., Navarete, F. D. T., & Kowaltowski, A. J. (2009). Tissue-, substrate-, and site-specific characteristics of mitochondrial reactive oxygen species generation. *Free Radic Biol Med*, *46*(9), 1283–1297.  
<http://doi.org/10.1016/j.freeradbiomed.2009.02.008>
- Thaker, S. K., Ch'ng, J., & Christofk, H. R. (2019). Viral hijacking of cellular metabolism. *BMC Biol*, *17*(1), 59.  
<http://doi.org/10.1186/s12915-019-0678-9>
- Thiele, I., Swainston, N., Fleming, R. M. T., Hoppe, A., Sahoo, S., Aurich, M. K., ... Palsson, B. Ø. (2013). A community-driven global reconstruction of human metabolism. *Nat Biotechnol*, *31*(5), 419–425.  
<http://doi.org/10.1038/nbt.2488>
- Thomas, T., Stefanoni, D., Reisz, J. A., Nemkov, T., Bertolone, L., Francis, R. O., ... D'Alessandro, A. (2020). COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight*, *5*(14). <http://doi.org/10.1172/jci.insight.140327>
- Thorlund, K., Dron, L., Park, J., Hsu, G., Forrest, J. I., & Mills, E. J. (2020). A real-time dashboard of clinical trials for COVID-19. *Lancet Digit Health*, *2*(6), e286–e287. [http://doi.org/10.1016/S2589-7500\(20\)30086-8](http://doi.org/10.1016/S2589-7500(20)30086-8)
- Touret, F., Gilles, M., Barral, K., Nougairède, A., Helden, J. van, Decroly, E., ... Coutard, B. (2020). In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci Rep*, *10*(1), 13093.

<http://doi.org/10.1038/s41598-020-70143-6>

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419.  
<http://doi.org/10.1126/science.1260419>

Vabret, N., Britton, G. J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., ... Sinai Immunology Review Project. (2020). Immunology of COVID-19: Current State of the Science. *Immunity*, *52*(6), 910–941.  
<http://doi.org/10.1016/j.immuni.2020.05.002>

Valcárcel, L. V., Torrano, V., Tobalina, L., Carracedo, A., & Planes, F. J. (2019). rMTA: Robust metabolic transformation analysis. *Bioinformatics*, *35*(21), 4350–4355. <http://doi.org/10.1093/bioinformatics/btz231>

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., ... Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*, *350*(6257), 207–211.  
<http://doi.org/10.1126/science.aad0095>

V'kovski, P., Gerber, M., Kelly, J., Pfaender, S., Ebert, N., Braga Lagache, S., ... Thiel, V. (2019). Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife*, *8*. <http://doi.org/10.7554/eLife.42037>

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2020). Coronavirus

biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol*.

<http://doi.org/10.1038/s41579-020-00468-6>

Vlassis, N., Pacheco, M. P., & Sauter, T. (2014). Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol*, *10*(1),

e1003424. <http://doi.org/10.1371/journal.pcbi.1003424>

Vodnala, S. K., Eil, R., Kishton, R. J., Sukumar, M., Yamamoto, T. N., Ha, N.-H., ... Restifo, N. P. (2019). T cell stemness and dysfunction in tumors are triggered by a common mechanism. *Science*, *363*(6434), eaau0135.

<http://doi.org/10.1126/science.aau0135>

Wagner, J., Wickman, E., DeRenzo, C., & Gottschalk, S. (2020). CAR T Cell Therapy for Solid Tumors: Bright Future or Dark Reality? *Molecular Therapy*,

*28*(11), 2320–2339. <http://doi.org/10.1016/j.ymthe.2020.09.015>

Waldhauer, I., & Steinle, A. (2008). NK cells and cancer immunosurveillance.

*Oncogene*, *27*(45), 5932–5943. <http://doi.org/10.1038/onc.2008.267>

Waldman, A. D., Fritz, J. M., & Lenardo, M. J. (2020). A guide to cancer immunotherapy: From T cell basic science to clinical practice. *Nature Reviews*

*Immunology*, *20*(11), 651–668. <http://doi.org/10.1038/s41577-020-0306-5>

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281–292.e6.

<http://doi.org/10.1016/j.cell.2020.02.058>

- Wang, R., Simoneau, C. R., Kulsuptrakul, J., Bouhaddou, M., Travisano, K. A., Hayashi, J. M., . . . Puschnik, A. S. (2021). Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. *Cell*, *184*(1), 106–119.e14. <http://doi.org/10.1016/j.cell.2020.12.004>
- Wang, W., & Zou, W. (2020). Amino Acids and Their Transporters in T Cell Immunity and Cancer Therapy. *Molecular Cell*, *80*(3), 384–395. <http://doi.org/10.1016/j.molcel.2020.09.006>
- Wang, Y., Eddy, J. A., & Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*, *6*, 153. <http://doi.org/10.1186/1752-0509-6-153>
- Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C. O., Finkin, S., . . . Nussenzweig, M. C. (2021). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *bioRxiv*, 2021.01.15.426911. <http://doi.org/10.1101/2021.01.15.426911>
- Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., . . . Wilen, C. B. (2021). Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection. *Cell*, *184*(1), 76–91.e13. <http://doi.org/10.1016/j.cell.2020.10.028>
- Weingarten-Gabbay, S., Klaeger, S., Sarkizova, S., Pearlman, L. R., Chen, D.-Y., Bauer, M. R., . . . Sabeti, P. C. (2020). SARS-CoV-2 infected cells present HLA-I peptides from canonical and out-of-frame ORFs. *bioRxiv*.

<http://doi.org/10.1101/2020.10.02.324145>

WHO. (2020a). Therapeutics and COVID-19: Living guideline. Retrieved from

<https://apps.who.int/iris/bitstream/handle/10665/337876/>

WHO-2019-nCoV-therapeutics-2020.1-eng.pdf

WHO. (2020b). WHO Coronavirus Disease (COVID-19) Dashboard. Retrieved

from <https://covid19.who.int>

WHO. (2020c). WHO Director-General's opening remarks at the media briefing on

COVID-19 - 11 March 2020. Retrieved from

<https://www.who.int/director-general/speeches/detail/>

who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-mar

WHO. (2021). International Clinical Trials Registry Platform (ICTRP). Retrieved

from <https://www.who.int/clinical-trials-registry-platform>

WHO Solidarity Trial Consortium, Pan, H., Peto, R., Henao-Restrepo, A.-M.,

Preziosi, M.-P., Sathiyamoorthy, V., ... Swaminathan, S. (2020). Repurposed

Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. *N Engl J*

*Med.* <http://doi.org/10.1056/NEJMoa2023184>

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York:

Springer-Verlag. <http://doi.org/10.1007/978-0-387-98141-3>

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ...

Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database

for 2018. *Nucleic Acids Res*, *46*(D1), D1074–D1082.

<http://doi.org/10.1093/nar/gkx1037>

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., ... Wendtner, C. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature*, *581*(7809), 465–469.

<http://doi.org/10.1038/s41586-020-2196-x>

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., ... Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. <http://doi.org/10.1038/s41586-020-2008-3>

Wu, H., Ye, M., Liu, D., Yang, J., Ding, J.-W., Zhang, J., ... Yang, J. (2019). UCP2 protect the heart from myocardial ischemia/reperfusion injury via induction of mitochondrial autophagy. *J Cell Biochem*, *120*(9), 15455–15466.

<http://doi.org/10.1002/jcb.28812>

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*, *323*(13), 1239–1242.

<http://doi.org/10.1001/jama.2020.2648>

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., & Shan, H. (2020). Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology*, *158*(6),

1831–1833.e3. <http://doi.org/10.1053/j.gastro.2020.02.055>

Xiong, R., Zhang, L., Li, S., Sun, Y., Ding, M., Wang, Y., ... Xu, K. (2020).

Novel and potent inhibitors targeting DHODH are broad-spectrum antivirals against RNA viruses including newly-emerged coronavirus SARS-CoV-2.

*Protein Cell*, 11(10), 723–739.

<http://doi.org/10.1007/s13238-020-00768-w>

Xiong, Y., Liu, Y., Cao, L., Wang, D., Guo, M., Jiang, A., ... Chen, Y. (2020).

Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*, 9(1),

761–770. <http://doi.org/10.1080/22221751.2020.1747363>

Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., ... Gong, S. (2020).

Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nat Med*, 26(4), 502–505.

<http://doi.org/10.1038/s41591-020-0817-4>

Yan, B., Chu, H., Yang, D., Sze, K.-H., Lai, P.-M., Yuan, S., ... Yuen, K.-Y.

(2019). Characterization of the Lipidomic Profile of Human

Coronavirus-Infected Cells: Implications for Lipid Metabolism Remodeling upon Coronavirus Replication. *Viruses*, 11(1).

<http://doi.org/10.3390/v11010073>

Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., ... Lou, Z. (2020).

Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat*

*Commun*, 11(1), 5874. <http://doi.org/10.1038/s41467-020-19770-1>

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis

for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), 1444–1448. <http://doi.org/10.1126/science.abb2762>

Yang, K., Neale, G., Green, D. R., He, W., & Chi, H. (2011). The tumor suppressor Tsc1 enforces quiescence of naive T cells to promote immune homeostasis and function. *Nat Immunol*, 12(9), 888–897. <http://doi.org/10.1038/ni.2068>

Yang, W., Kandula, S., Huynh, M., Greene, S. K., Van Wye, G., Li, W., . . . Shaman, J. (2021). Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: A model-based analysis. *Lancet Infect Dis*, 21(2), 203–212. [http://doi.org/10.1016/S1473-3099\(20\)30769-6](http://doi.org/10.1016/S1473-3099(20)30769-6)

Yizhak, K., Gabay, O., Cohen, H., & Ruppin, E. (2013). Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat Commun*, 4, 2632. <http://doi.org/10.1038/ncomms3632>

Yizhak, K., Gaude, E., Le Dévédec, S., Waldman, Y. Y., Stein, G. Y., Water, B. van de, . . . Ruppin, E. (2014). Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *Elife*, 3. <http://doi.org/10.7554/eLife.03641>

Yokoyama, W. M., & Kim, S. (2006). How do natural killer cells find self to achieve tolerance? *Immunity*, 24(3), 249–257.

<http://doi.org/10.1016/j.immuni.2006.03.006>

Yuan, S., Chu, H., Chan, J. F.-W., Ye, Z.-W., Wen, L., Yan, B., ... Yuen, K.-Y. (2019). SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target. *Nat Commun*, *10*(1), 120.

<http://doi.org/10.1038/s41467-018-08015-x>

Zaim, S., Chong, J. H., Sankaranarayanan, V., & Harky, A. (2020). COVID-19 and Multiorgan Response. *Curr Probl Cardiol*, *45*(8), 100618.

<http://doi.org/10.1016/j.cpcardiol.2020.100618>

Zeng, H., Yang, K., Cloer, C., Neale, G., Vogel, P., & Chi, H. (2013). mTORC1 couples immune signals and metabolic programming to establish T(reg)-cell function. *Nature*, *499*(7459), 485–490. <http://doi.org/10.1038/nature12297>

Zhang, C., Lee, S., Bidkhor, G., Benfeitas, R., Lovric, A., Chen, S., ...

Mardinoglu, A. (2019). RMetD2: A tool for integration of relative transcriptomics data into Genome-scale metabolic models. *bioRxiv*, 663096.

<http://doi.org/10.1101/663096>

Zhang, C., Shi, L., & Wang, F.-S. (2020). Liver injury in COVID-19: Management and challenges. *Lancet Gastroenterol Hepatol*, *5*(5), 428–430.

[http://doi.org/10.1016/S2468-1253\(20\)30057-1](http://doi.org/10.1016/S2468-1253(20)30057-1)

Zhang, C.-Y., Parton, L. E., Ye, C. P., Krauss, S., Shen, R., Lin, C.-T., ...

Lowell, B. B. (2006). Genipin inhibits UCP2-mediated proton leak and acutely reverses obesity- and high glucose-induced beta cell dysfunction in isolated

pancreatic islets. *Cell Metab*, 3(6), 417–427.

<http://doi.org/10.1016/j.cmet.2006.04.010>

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., & Molina, M. J. (2020). Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc Natl Acad Sci U S A*, 117(26), 14857–14863.

<http://doi.org/10.1073/pnas.2009637117>

Zhang, Y., Zhang, J., Chen, Y., Luo, B., Yuan, Y., Huang, F., ... Zhang, H.

(2020). The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv*, 2020.05.24.111823.

<http://doi.org/10.1101/2020.05.24.111823>

Zhao, J., Cui, W., & Tian, B.-P. (2020). The Potential Intermediate Hosts for SARS-CoV-2. *Front Microbiol*, 11, 580137.

<http://doi.org/10.3389/fmicb.2020.580137>

Zhao, R.-Z., Jiang, S., Zhang, L., & Yu, Z.-B. (2019). Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *Int J Mol Med*, 44(1), 3–15. <http://doi.org/10.3892/ijmm.2019.4188>

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., ... Wang, M. H.

(2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int J Infect Dis*, 92, 214–217.

<http://doi.org/10.1016/j.ijid.2020.01.050>

Zheng, S., Fan, J., Yu, F., Feng, B., Lou, B., Zou, Q., . . . Liang, T. (2020). Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: Retrospective cohort study. *BMJ*, *369*, m1443. <http://doi.org/10.1136/bmj.m1443>

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., . . . Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. <http://doi.org/10.1038/s41586-020-2012-7>

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*, *6*, 14. <http://doi.org/10.1038/s41421-020-0153-3>

Zhou, Y., Wang, F., Tang, J., Nussinov, R., & Cheng, F. (2020). Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health*, *2*(12), e667–e676. [http://doi.org/10.1016/S2589-7500\(20\)30192-8](http://doi.org/10.1016/S2589-7500(20)30192-8)

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., . . . China Novel Coronavirus Investigating and Research Team. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*, *382*(8), 727–733. <http://doi.org/10.1056/NEJMoa2001017>