The Role of Metabolic Reprogramming in pH Homeostasis and Cancer Progression

by

Yi Zhou

(Under the Direction of Ying Xu)

Abstract

Metabolic reprogramming, recognized as one of the hallmarks of cancer, refers to the ability of cancer cells to rapidly adapt to microenvironment stresses via genetic mutations, transcriptomic regulations, and epigenetic modulations. In this dissertation, I have developed a framework to investigate the dysregulated metabolic reaction chains using transcriptomics data collected from cancer tissues. In the first chapter, we systematically analyzed 50 reprogrammed metabolic pathways across 14 cancer types and proposed proposed a model that these alterations are induced to neutralize a intracellular alkaline stress caused by chronic inflammation and iron overload. The close scrutiny of a specific pathway, sialic acid synthesis and utilization, confirmed our previous hypothesis of pH-stress driven metabolic reprogramming. Further analysis also identified a link between metabolic changes and tumor proliferation and metastasis. Our model provided strong evidence that sialic acid accumulation on the surface of cancer cells promotes cell-cell repulsion and ultimately drives migration. Finally, we developed a open-source tool named metabolike that can transform pre-specified systems biology models into a knowledge graph. The graph representation within Metabolike enables efficient identification and visualization of novel reaction routes, which makes it a great resource for studying complex biochemical reaction networks under diseased conditions such as cancer.

Index words: [Cancer Bioinformatics, pH homeostasis, Metabolic Reprogramming, Metabolic Graph]

The Role of Metabolic Reprogramming in pH Homeostasis and Cancer Progression

by

Yi Zhou

B.S., China Agricultural University, China, 2017
M.S., The University of Georgia, 2021

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2022

The Role of Metabolic Reprogramming in pH Homeostasis and Cancer Progression

by

Yi Zhou

Major Professor: Ying Xu

Committee: Jonathan Arnold
Liming Cai
Ping Ma

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

Acknowledgments

I would like to express my deepest gratitude to my advisor Ying Xu, who provided me with tremendous support in the past five years. Dr. Xu, thank you for your motivation, enthusiasm, patience, and guidance throughout my Ph.D. research. Your love for science has shaped who I am as a researcher today and I am extremely honored to have been a graduate student in the CSBL.

I am also thankful to my committee members Jonathan Arnold, Liming Cai, and Ping Ma. Thank you for all your time and scientific expertise offered to me during the committee and personal meetings. I am fortunate to have worked with each of you and have become a better scientist because of your help.

I would also like to thank my friends Michael Skaro and Zheng An, for enlightening when I cannot find the bug in my code, and for making life in Athens a lot more fun.

A special thanks to the trio Megan Guetter, April Mosley, and Sandra Getz, who stood by me and made graduate school a lot easier.

To my parents Binjiang Zhou and Jun Lu, thank you for your love and unshakable belief in me. This journey would not have been possible without your support.

Above all I would like to thank my wife Bufan Zhang for her love, for flying half the globe to keep me company, and for being fully supportive. Thank you for being my best friend. I owe you everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

Globally, cancer is the second leading cause of death. 1,918,030 new cancer cases and 609,360 cancer deaths are expected to occur in 2022 in the United States alone [1]. One of the reasons for the high incidence and mortality rates is that cancer is a diverse and complex disease harboring both intertumoral and intratumoral heterogeneity [2]. Cancers with different histological types and primary sites have fundamentally different initiation and progression mechanisms. Such a broad and diverse collection of diseases has prompted researchers to take a somewhat reductionist view towards cancer and summarize common features acquired by cells during neoplasm, in hope that this knowledge could help understand the mechanisms of cancer development and assist the discovery of new therapeutic interventions.

The hallmarks of cancer [3] were proposed as a set of common cellular phenotypes that are acquired by cells as they undergo carcinogenesis. Some of the most recognized hallmarks are genome instability and mutation, cell death resistance, sustained cell proliferation signaling, immune destruction invasion, and tumor-promoting inflammation [4]. These largely shared characteristics seem to contradict the heterogeneity of cancers, which prompted our pursuit of the root cause of these commonalities in cancer.

1.1 Acid-base homeostasis in cancer

One of the emerging hallmarks of cancer is its reversed pH gradient. Normal epithelial cells have an intracellular pH (pH_i) of about 7.2 and an extracellular pH (pH_e) at roughly 7.4. Cancer cells have a much more alkaline pH_i and an acidic pH_e, which has profound implications for various cellular processes [5].

For example, protein folding, protein-protein interaction, and kinetic rates of enzyme activity all directly depend on the pH in the surrounding environment [6].

The disrupted pH homeostasis plays a pivotal role in cancer progression. An increased pH_i promotes cell proliferation by increasing the rate of CDK1 activity [7] and growth factor-modulated cell cycle phase transition [8]. Decreased pH_i acts as a conserved signal for both death receptor-mediated and mitochondria-mediated apoptosis [9], and the increase of pH_i in cancer cells dampens both processes. Increased pH_i also promotes cancer cell migration by inducing cytoskeleton assembly [10] and decreasing cell-substrate adhesions [11]. The sustainably decreased pH_e, on the other hand, facilitates degradation of the extracellular matrix through the release of acid-activated proteases [10, 12]. It also modulates cell attachment to the extracellular matrix mediated by integrins [13], allowing for faster migration rates. An acidic pH_e is required for the formation and functioning of invadopodia [14], which are membrane protrusion structures essential for cancer cell invasion.

Changes in the intracellular and extracellular pH are largely attributed to the ion pumps and transporters located on the plasma membrane. Dysregulation of these proteins could overwhelm the HPO_4^{2-} / $H_2PO_4^-$-based and HCO_3^- / H_2CO_3-based buffer systems. A popular model is that Na^+/H^+ exchangers, particularly NHE1 (SLC9A1), are the main reason for the reversal of pH_i and pH_e in cancer tissues, along with monocarboxylate-H^+ efflux cotransporters MCT1 and MCT4 (SLC16A1 and SLC16A3) and carbonic anhydrase for CO_2 hydration [5]. However, we have recently proven this route inviable due to insufficient intracellular Na^+ potential or severe down-regulation of the transporters in cancer tissues [15]. Although the ion pumps and transporters could partly explain the extracellular acidification, the root cause giving rise to the intracellular alkaline stress in cancer cells remains unknown. In other words, we need to identify a cellular process that consistently produces large quantities of hydroxide.

1.2 Cytosolic Fenton reaction and alkaline stress

Elevated Fenton reaction levels has been widely observed in cancer tissues [16–18]. The Fenton reaction is an inorganic reaction that happens when both hydrogen peroxide (H_2O_2) and ferrous ion (Fe^{2+}) levels are sufficiently high [19]:

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + HO^{\bullet} + OH^{-}$$

It is noteworthy that the rates of the cytosolic Fenton reaction in cancer can quickly saturate the intracellular pH buffer system, hence driving intracellular alkalization. Our previous work has shown that all cancer tissue cells have the Fenton reaction occurring in their cytosol, mitochondria, extracellular matrix, and cell surface [20]. The reaction can be sustained by nearby reducing molecules such as superoxide ($O_2^{\bullet -}$) and NADH that can convert ferric ion (Fe^{3+}) back to Fe^{2+}. Local immune cells and mitochondria of the cancer cells generate $O_2^{\bullet -}$ [21], which is the precursor to most reactive oxygen species [22]. $O_2^{\bullet -}$ itself can drive the cytosolic Fenton reaction in the form referred to as the *Haber-Weiss reaction* with Fe^{2+} serving as a catalyst [23, 24]:

$$O_2^{\bullet -} + H_2O_2 \xrightarrow{Fe^{2+}} HO^{\bullet} + OH^{-} + O_2$$

The source of the H_2O_2 has a natural explanation. Chronic inflammation is known to be causally linked to cancer onset and development [25], giving rise to increased local concentrations of H_2O_2. We are now missing one last part to complete the puzzle: the source of the Fe^{2+} that serves either as a reactant or as a catalyst. Once the local H_2O_2 concentrations reach beyond a certain level, local red blood cells may become senescent due to the oxidation of their plasma membranes and their lack of a membrane repair mechanism [26], leading to their engulfment by macrophages [26] and local accumulation of irons released by macrophages after engulfment [27]. Under the condition of immune responses, local epithelial cells will sequester the free irons [28], leading to overload of intracellular irons. It has been widely reported that multiple chronic inflammatory diseases [29] and many cancer tissues have iron overload [30].

In summary, chronic inflammation leads to intracellular iron overload, which eventually drives elevated Fenton (or Haber-Weiss) reaction levels and overwhelms the pH buffering system. Even a slight disturbance of the intracellular pH could have profound downstream effects, which may partially explain the heterogeneity of cancers. Under our framework of cancer initiation, the genetic and transcriptomic changes of precancerous cells are all responses to the persistent microenvironment stress, among which the disruption of the pH_i may be at the root of cancer progression [20]. We postulate cancer cells utilize a suite of mechanisms to evade the intracellular alkaline stress, and as a result we observe another rising hallmark of cancer – metabolic reprogramming.

1.3 Metabolic reprogramming as a cancer hallmark

Metabolic reprogramming refers to the altered metabolisms in cancer that supports cell growth and proliferation [31]. Extensive studies have covered various aspects of biochemical changes in tumor cells, such as amino acid [32], nucleotide [33, 34], lipid [35], carbohydrate [36], vitamin [37], and sulfur [38] metabolisms. Some reprogrammed pathways are simply up-regulated or repressed canonical pathways, such as the persistent up-regulation of *de novo* biosynthesis of deoxyribonucleotides [34], and the repression of the urea cycle [39]. Another reprogramming mechanism is the coupling among normal metabolic processes, such as the simultaneous synthesis and degradation of triglyceride [40], and the activation of both glycolysis and gluconeogenesis [41]. Some cancer cells utilize truncated metabolic pathways, with the Warburg effect being the most prominent example [42]. Recently, these altered metabolisms of tumor cells have become popular drug targets [43–46].

Although widely observed in a conserved manner across multiple cancer types, the specific driving forces of metabolic reprogramming is not well established. Based on the intracellular pH stress driven cancer progression model, we hypothesize the reprogramming of the metabolic pathways serves as an escape mechanism for cancer cells to mitigate the ever-increasing pH_i. Through systematically analyzing well-established metabolically altered pathways, we revealed a pattern where genes encoding enzymes that catalyze H^+-producing reactions tend to be more up-regulated than genes encoding enzymes that catalyze

H^+-consuming reactions [47]. This indicates that increasing the overall H^+ production might be a key driver of metabolic reprogramming, which is consistent with the goal of overcoming the alkaline stress introduced by the cytosolic Fenton reaction.

Consequently, reprogrammed metabolic pathways generate various metabolites that need to be further degraded or disposed by the cell. As an example, triglyceride synthesis is a reprogrammed pathway observed in most cancer types. Some cancers degrade triglyceride into glycerol and fatty acids where a synthesis and degradation cycle is formed, while others partially degrade and release it in the form of arachidonic acid or prostaglandin [48]. Rearrangement of the downstream metabolic reactions may give rise to abnormal cellular behaviors. We took sialic acids as an example and hypothesized that the search of an exit for these metabolic by-products could be linked to various phenotypes of cancer cells. Sialic acids, negatively charged nine-carbon sugars, came to our attention during our hunt for abnormally up-regulated metabolic routes in cancer [47]. Sialic acids have long been associated with cancer development and metastasis [49, 50]. They generally serve as the capping molecules of cell-surface glycans as part of plasma membrane-embedded gangliosides [51]. Cancer cells accumulate large quantities of sialic acids on their surfaces [52]. Previous studies attributed this to the signaling roles of sialic acids. Communication of cancer and immune cells via binding of sialic acids to siglecs and selectins [53] enables cancer cells to interact with and penetrate into blood vessels [49]. We, on the other hand, argue that physical properties of these highly-charged molecules could have significant impacts in cancer migration. The accumulation of sialic acids on the cell surface could be correlated with the observed mechanical pressure within cancer tissues [54, 55], which can lead to cell deformation. The details of our investigation will be discussed in Chapter 3.

1.4 Exploring the metabolic network as a graph

The final part of this dissertation is orthogonal to the previous sections. We are now blessed with data sets of high quality and quantity thanks to the technical advancement of experimental and computational tools. Canonical metabolic pathways are discovered, annotated, and deposited to repositories such as KEGG [56], BRENDA [57], and MetaCyc [58]. However, integrating information from these sources

is not trivial due to the inconsistent ontologies, nonstandard identifiers, and duplicated components. High-throughput analyses are also hindered by the size of complexity of the networks [59].

At this point, it would be useful to have an alternative method of storing metabolic networks. The storage method should allow SQL-like queries to the unstructured data, which enables a variety of tasks that are difficult or impossible with the web interfaces provided by the resources. A graph database [60] serves the purpose since metabolic networks can be naturally represented as directed graphs consisting of interconnected nodes of diverse types. Compared to relational databases (which are used by a number of reputable services [57, 58]), a graph database is significantly more performant for querying highly-connected entities [61] and provides native support for various graph traversal algorithms. The Systems Biology Markup Language (SBML), an XML-based specification for describing biochemical networks [62], has gradually become the *de facto* standard for exchanging models between software tools and is an ideal intermediate data format between online resources and graph databases. We aim to develop a computational tool that could bridge the gap and allow users to easily explore vast metabolic reaction graphs in an efficient manner.

Another side of the story is the dynamics within the graph. The aforementioned network is mostly static, meaning that the structure does not change once new annotations are added. Researchers are often more interested in studying the transcriptomic and metabolic flux changes in the network, both varying greatly between batches of samples. In a typical RNA sequencing bioinformatics pipeline, the raw sequencing reads undergo quality control and are mapped to the reference genome. Transcript or gene level quantification measures are then computed from the mapped reads. After a series of preprocessing steps[1], a differential expression analysis is performed to identify the significantly up or down-regulated genes between sample cohorts. A gene set enrichment analysis usually follows to identify relevant biological processes and functional gene clusters. This is where domain experts come in and interpret the biological meanings of the gene set changes and reveal hidden patterns in the data.

We may view this as a suite of dimension reduction techniques, where impossible-to-interpret raw sequences are reduced into biologically meaningful gene sets. However, pathway enrichment methods

[1]The preprocessing steps usually include ID mapping, normalization, low-pass filtering, and batch effect correction.

such as GSEA [63] have their own limitations [64]. These methods are biased towards genes present in multiple pathways and cannot identify sub-networks that are not part of a predefined pathway. They are also unaware of the overall network structure, meaning that connections between pathways are not taken into account as all gene sets are tested independently. All these shortcomings can be addressed by a tool that can integrate omics data with the metabolic reaction graph and execute a route search algorithm to infer structural changes for the catabolism of a compound of interest. With a graph database available, our analyses in the previous sections could be much more automated as we can directly map omics-data onto nodes in the graph. Instead of resorting to "bag-of-genes"-based methods, it would be possible to take the structure of the reaction steps into consideration when examining differential expression data. This also enables the identification of novel metabolic routes under diseased or perturbed conditions, since we can traverse any known reaction and are not limited to the pathway-centric view of metabolism.

To this end, we developed a Python package named Metabolike that can transform metabolic models into a graph and perform various reaction route search algorithms utilizing transcriptomics data. Metabolike could be used when the researcher has collected a list of dysregulated gene sets, and is interested in the specific metabolic disruptions that occurred in the network. Instead of manually inspecting each differentially expressed metabolic pathway, researchers can use Metabolike to automatically identify the influx and outflux perturbations around a metabolite of interest.

1.5 Dissertation outline

The overarching objective of this dissertation is to characterize the role of metabolic reprogramming in mitigating intracellular alkaline stress and driving tumor progression. Specifically, the chapters were developed around the central hypothesis that metabolic reprogramming serves as a survival mechanism and links intracellular pH disruption to cancer phenotypic behaviors.

Chapter 2 is a computational study on a panel of 49 reprogrammed metabolisms in cancer. Transcriptomics data of 7011 samples spanning 14 cancer types were retrieved from the cancer genome atlas (TCGA) database, and expression profiles of amino acid, nucleotide, lipid, and sugar metabolism were analyzed. It

was found that all of the reprogrammed metabolisms tend to produce more protons than their original forms. Certain metabolic reprogramming patterns were used ubiquitously across cancer types, including nucleotide *de novo* synthesis, the Warburg effect, glycosaminoglycan synthesis, and N-linked glycosylation. A penalized linear regression analysis revealed that combinations of reprogrammed metabolisms can well-capture the cytosolic Fenton reaction levels, pointing to the possibility that metabolic reprogramming is a response to the intracellular pH disruption. The study provides a pan-cancer metabolic landscape that can be used to identify the cause of various cancer phenotypic behaviors. It proposes the idea that compounds produced in the reprogrammed metabolisms may drive the progression of cancer by introducing intracellular stress.

Chapter 3 dives into the reprogrammed metabolism of sialic acids, which are negatively charged sugars located on the terminal of cell-surface glycans. Differential expression analysis showed that sialic acid synthesis and deployment were persistently up-regulated in almost all studied cancer types. Previous studies generally attributed the increased sialic acid production to its signaling roles via binding with siglecs and selectins, but the data suggests otherwise as there was little to no co-expression. Fenton reaction levels were demonstrated to be highly correlated with sialic acid production, which confirms the predictions in the previous chapter as sialic acid production is a proton-producing process. Statistical analysis of ganglioside type preferences in advanced cancer stages also supported the proton production hypothesis. Finally, three-layer fully-connected neural network was trained to predict cancer metastasis stages from sialic acid-related features. Transcription regulation through DNA methylation and transcription factors was investigated for important features. Overall, the study provides a computational framework for cancer metastasis prediction, and links the migration of cancer cells to sialic acids accumulated on the cell surface through metabolic reprogramming.

Chapter 4 introduces a Python package for studying reprogrammed metabolic pathways. The package, Metabolike, transforms systems biology markup language (SBML) models into a graph database, and provides an interface for interacting with the graph. Using the MetaCyc database as an example, Metabolike was able to generate highly interconnected graph of pathways, reactions, compounds, and

gene products. Given a metabolite of interest and results of a differential expression analysis, Metabolike outputs novel reaction routes originating from the metabolite that are likely to occur under the given conditions. Several exploratory analyses were performed to showcase the potential of the software. The ganglioside type preference analysis in Chapter 3 was revisited using the Metabolike framework, and the results were shown to be consistent with the previous results. Repression of the urea cycle in 14 cancer types was also identified by Metabolike when supplying differential expression analysis results. This work contributes to the understanding of metabolic reprogramming by providing a automatic route search algorithm for exploring the context-specific metabolic network with omics data integration.

Chapter 2

Metabolic Reprogramming in Cancer is Induced by Cytosolic Alkalinization to Preserve Acid-Base Homeostasis

Huiyan Sun (co-first author), Yi Zhou (co-first author), Michael Skaro, Yiran Wu, Zexing Qu, Fenglou Mao, Suwen Zhao, Ying Xu. (2020) *Cancer Res*; **80** (5): 1143–1155.

Reprinted here with permission of the publisher.

2.1 Abstract

Considerable metabolic reprogramming has been observed in a conserved manner across multiple cancer types, but their true causes remain elusive. We present an analysis of around 50 such reprogrammed metabolisms (RM) including the Warburg effect, nucleotide *de novo* synthesis, and sialic acid biosynthesis in cancer. Analyses of the biochemical reactions conducted by these RMs, coupled with gene expression data of their catalyzing enzymes, in 7011 tissues of 14 cancer types, revealed that all RMs produce more H^+ than their original metabolisms. These data strongly support a model that these RMs are induced or selected to neutralize a persistent intracellular alkaline stress due to chronic inflammation and local iron overload. To sustain these RMs for survival, cells must find metabolic exits for the non-proton products of these RMs in a continuous manner, some of which pose major challenges, such as nucleotides and sialic acids, because they are electrically charged. This analysis strongly suggests that continuous cell division and other cancerous behaviors are ways for the affected cells to remove such products in a timely and sustained manner. As supporting evidence, this model can offer simple and natural explanations to a range of long-standing open questions in cancer research including the cause of the Warburg effect.

2.2 Introduction

A wide range of metabolic changes has been observed in cancer compared with matching normal tissues. Some of these changes are simple, such as persistent up- or down-regulation, whereas others involve some or substantial rewiring of the normal metabolic processes [31]. Examples of the former range from persistently increased glycosylation and fatty acid biosynthesis to decreased production and utilization of arginine; and from enhanced purine degradation but repressed pyrimidine degradation to increased *de novo* synthesis versus uptake from circulation of certain amino acids such as serine and proline. The rewiring of some metabolisms is well elucidated such as the expanded utilization of glutamine via the glutaminolysis pathway, whereas others are yet to be fully determined such as the replacement of the inhibited urea cycle for releasing NH_3, the waste of amino acid metabolism. Some reprogrammed metabolisms (RM) have

been studied extensively, such as the Warburg effect [65], lipid biosynthesis [66], and NAD^+/NADH metabolism, whereas others have not been generally considered as metabolic reprogramming issues like persistently elevated biosynthesis of sialic acids (SA) [67] and gangliosides.

The overall landscape of the RMs in cancer is probably more extensive than what has been reviewed [65, 68, 69]. A thorough literature survey, coupled with our own omic data analyses, suggests that it covers virtually every aspect of cellular metabolism, ranging from amino acid [32], nucleotide [33], lipid [35], sugar [36] to vitamin [37] and sulfur [38] metabolisms. Our analyses of cancer transcriptomics data in the Cancer Genome Atlas (TCGA) database have revealed: the RMs tend to be conserved across multiple, even possibly all cancer types; and they clearly require more than needed by cell proliferation to explain as their statistical relationship with cell proliferation ranges from being positive, negative, or independent as observed in this study.

Numerous proposals have been made regarding the possible causes for individual RMs. These largely fall into three categories: the RMs (1) provide faster ways for energy production and macromolecular biosyntheses [70]; (2) activate onco-proteins or produce onco-metabolites [71]; and (3) are results of oxidative stress [72] or hypoxia [73]. While these proposals may have offered sound explanations to some RMs, there are issues that require further thinking. Among them is whether each proposed cause is the primary or a secondary one that may benefit cancer but not necessarily be induced or selected for. There is clearly lack of data that can explain why most RMs are conserved across multiple cancer types. The extensiveness of the RMs and their consistencies across multiple cancers suggest that there might be something more fundamental than what have been proposed as common causes for majority of the RMs.

We present a computational analysis of transcriptomics data of 7011 cancer tissues of 14 cancer types in TCGA, a set of cancer types that we have been studying in our recent work [20, 74], because they each have large enough sample sizes needed for reliable statistical analyses. We have examined all the enzymatic genes with differential expressions in cancer vs. control in each cancer type, and examined: how do the altered expressions of these genes affect the intracellular pH? Throughout this paper, all the mRNAs for

each gene, including its splicing isoforms, are counted towards the expression level of the gene since we do not have a way to study isoform-specific biology here.

The main reasons that we focus on pH are: (1) cancer tissue cells are known to have an alkaline intracellular pH, which compares to a slightly acidic one in matching normal tissue cells [75] and are under alkaline stress [9, 76, 77]; (2) cancer tissue cells generally up-regulate acidifying transporters, while inhibiting the alkalizing transporters as we have previously reported [74]; and (3) all cancer tissue cells of the 14 types harbor persistent Fenton reactions: $Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + {}^{\bullet}OH + OH^-$ (or $O_2{}^{\bullet -} + H_2O_2 \longrightarrow {}^{\bullet}OH + OH^- + O_2$ if $O_2{}^{\bullet -}$ is richly available with $Fe_2{}^+$ serving as a catalyst) in multiple subcellular locations, particularly cytosol and mitochondria, at levels that can overwhelm the pH buffer quickly, hence alkalinizing the intracellular space on a persistent basis if not neutralized, as we have demonstrated [20].

2.3 Materials and Methods

2.3.1 Data

RNA-seq data of 7011 tissue samples of 14 cancer types are retrieved from TCGA, with detailed information given in Table 2.1. The 14 cancer types are selected since they are all the cancer types each with a sufficiently large number of tissue samples with RNA-seq data in TCGA. Transcripts per million are used in our differential expression analyses.

2.3.2 Methods

Identification of the number of protons consumed or produced by each enzymatic reaction For each enzymatic reaction, we count the number of H^+/CO_2 consumed or produced by each reaction as provided by HumanCyC, UniProt and KEGG.

Differential expression analyses of a pathway For each pathway under study, we identify a minimal subset of genes of the pathway, termed signature genes, whose expressions correlate with at least 80% genes of the pathway above a specified correlation coefficient, measured using Pearson correlation, across

Table 2.1: Cancer sample information from the TCGA database.

Abbreviation	Cancer type	Tumor sample count	Normal sample count
BLCA	Bladder urothelial carcinoma	414	19
BRCA	Breast invasive carcinoma	1109	113
COAD	Colon adenocarcinoma	480	41
ESCA	Esophageal carcinoma	162	11
HNSC	Head and neck squamous cell carcinoma	502	44
KICH	Kidney chromophobe	65	24
KIRC	Kidney renal clear cell carcinoma	539	72
KIRP	Kidney renal papillary cell carcinoma	289	32
LIHC	Liver hepatocellular carcinoma	374	50
LUAD	Lung adenocarcinoma	535	59
LUSC	Lung squamous cell carcinoma	502	49
PRAD	Prostate adenocarcinoma	499	52
STAD	Stomach adenocarcinoma	375	32
THCA	Thyroid carcinoma	510	58

all 14 cancer types. A pathway is considered up-regulated in a cancer (at a given stage) vs. the control if the sum of the expressions of its signature genes averaged over all cancer samples of the stage is higher than the corresponding sum overall the control samples with p-value < 0.05. Similar is defined for down-regulated.

Assessing the level of cytosolic Fenton reactions We have previously used the total expression level of proteasome genes to assess the level of cytosolic Fenton reactions, with the rationale being: the increased levels of proteasome genes can reflect the level of protein damages in cytosol predominantly due to hydroxyl radicals [20]. Specifically, it has been well established that a unique feature of hydroxyl radical-induced protein damages is to give rise to protein aggregation but not fragmentation like by other oxidizing molecules [78] and 20S proteasome is largely responsible for degrading such aggregates [79]. This information, coupled with the knowledge that only Fenton reactions can produce hydroxyl radicals inside cells, validates our use of 20S proteasome genes to gauge the level of Fenton reactions. The specific proteasome genes used for Fenton reaction assessment are given in Table S1, which are predominantly 20S genes with

a few 26S genes. The level of Fenton reaction is defined as the fold change of the combined expressions of these genes in given cancer samples over given control samples.

Regression analysis of Fenton reactions against acidifying processes We have conducted a linear regression analysis of the estimated level of Fenton reactions, Y_n, in n cancer samples of each cancer type against the levels of m RMs across the n samples, termed $X_{n,m}$ so that residual ϵ is as small as possible:

$$Y = XM + \epsilon$$

where M_m is a coefficient vector with its values to be determined through solving this optimization problem. Specifically, X represents the average expression of the signature genes for each RM in each sample; and Y is the average expression of the signature genes of Fenton reaction in each sample. To avoid using too many RMs in the regression analysis, we have also included an L_1 penalty term as follows:

$$Y = XM + \epsilon + \lambda m$$

where λ is an (adjustable) constant. We have solved the linear regression problem using the least squares method. The p-value of each free variable, representing the significance of each RM's contribution to the regression result, is calculated using the `glmnet` R package [80]. We did a two-round regression analysis. After the first-round regression, we remove all the RM terms without significant contribution to the regression result; and then conduct the second round of regression using only those selected RMs. The p-value of the obtained R^2 for each regression analysis is calculated by performing F statistics test $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$, where n is the number of observations and k is the number of variables.

Prediction of a new metabolic pathway For some metabolism, the reprogramming could be so substantial that the established pathway may not be quite correct. For each such case, we have predicted a novel RM as follows. We first identify all the pieces from the original pathway that may still be used based on co-expression patterns among the relevant genes; and then apply the following to piece together a

complete new pathway. For each metabolite produced by an enzymatic reaction, we go through all the reactions that use this metabolite as a reactant provided in KEGG. For each such enzymatic reaction, we predict a gene that may encode the enzyme based on if the gene is co-expressed with the gene whose enzymatic reaction produces the metabolite. If no gene is found, the search backtracks till all reactions are searched. The result is a pathway consisting of up-regulated genes that starts from a specified metabolite and ends with metabolites that can be secreted out of cells or known to be enriched inside cancer cells.

Validation of differentially expressed genes against protein data For each differentially expressed gene in a cancer, we have compared it against the protein abundance data in Human Protein Atlas. We consider a differentially expressed gene is validated if the matching protein has a consistent abundance pattern.

2.4 Results

We have analyzed around 50 well-established RMs, covering amino acid, nucleotide, lipid, sugar and a few other metabolisms in 14 cancer types, and have found: (1) all these RMs each produce more protons than the original metabolisms; and (2) while distinct combinations of these RMs at varying levels are used by different cancer types, all 14 types employ (i) nucleotide *de novo* syntheses, (ii) Warburg effect, (iii) syntheses of glycosaminoglycan, (iv) triglyceride biosynthesis, (v) choline production and metabolism, and (vi) N-linked glycosylation among a few others. Here we highlight a few for each type of metabolism to illustrate how our analysis is done and leave the rest in the Supplementary Results, considering the space required. Throughout this paper, a pathway is considered differentially expressed in a cancer if the average expression of its signature genes (Table 2.2) is up- or down-regulated in cancer vs. the matching control.

Table 2.2: Signature genes for each RM under study.

Pathway	Signature genes
Acidifying transporters	SLC36A1, SLC26A6, SLC47A1, SLCO2B1
Alkalizing transporters	SLC16A1, SLC16A3, SLC9A1, SLC9A6, SLC4A7

Arginine transporter	SLC7A1, SLC7A2
ATP consumption	See Figure S35
Beta oxidation	ACAD9, ACAD10
Ceramide synthesis	CERS2, DEGS1, SMPD2, SPTLC2
Choline production and metabolism	PLA2G6, CHKB, LYPLA1, LYPLA2
Chondroitin sulfate synthesis	B3GAT3, CHPF, CHST15, XYLT2
Circadian rhythm	CSNK1D, CSNK1E, NPAS2, NR1D1
Fatty acid synthesis	FASN, MCAT
Fatty acid transporter	SLC27A2, 3, 4
Gluconeogenesis	G6PC3, PCK2
Glutaminolysis	ACLY, CS, MDH2, SLC25A1
Heparan sulfate synthesis	B3GALT6, B3GAT3, B4GALT7, HS2ST1, XYLT2
Hyaluronic acid synthesis	GFPT1, GNPNAT1, PGM3
Hydroxylation enzymes	See Figure S31
Keratan sulfate synthesis	CHST1
Lysine degradation	HSD17B10, GCDH
Methylation	See Figure S30
Mevalonate metabolism	FDPS, GGPS1, MVD
NAD+ synthesis and metabolism	See Figure S28
N-linked glycosylation complex phase	FUT8, MAN2A1, MGAT4B
N-linked glycosylation initial phase	ALG3, ALG8, DPM2
O-linked glycosylation	C1GALT1, GALNT1, GCNT1
Phospholipid degradation	PLA2G6, PTGES2, PTGES3, TBXAS1
Phosphatidic acid (PA) synthesis	PTPMT1
Phosphatidylcholine (PC) synthesis	CHKB
Phosphatidylethanolamine (PE) synthesis	ETNK1, SELENOI
Phosphatidylinositol (PI) synthesis	CDIPT
Phosphatidylserine (PS) synthesis	PTDSS1
Proline synthesis	ALDH18A1, PYCR1, PYCR2
Phosphorylation/de-phosphorylation	See Figure S24
Purine dRN *de novo* synthesis	NME1, ATIC, PFAS, RRM2
Purine dRN salvage	APRT, DGUOK, GMPS, GUK1
Purine RN *de novo* synthesis	NME1, ATIC, PFAS, RRM2
Purine RN degradation	RPIA, TALDO1, TKT
Pyrimidine degradation	UPP1, CDA
Pyrimidine dRN *de novo* synthesis	CAD, CTPS1
Pyrimidine dRN salvage	UPP1, CDA
Pyrimidine RN *de novo* synthesis	CAD, CTPS1, NME1
Pyrimidine RN salvage	CTPS1, UCK1, UCK2
Retinol metabolism	CES4A, LRAT, XDH
Retinol synthesis	RBP1, RDH10
Serine synthesis	PSPH, VPS29

Sialic acid synthesis	CMAS, NANS
Triglyceride degradation	DAGLB
Triglyceride synthesis	GPAT4, LPCAT1, MBOAT7, PLPP4
Tryptophan degradation	AFMID, GCDH, HSD17B10, IDO1, TDO2
Warburg effect	PDHB, PKM

2.4.1 Reprogrammed amino acid metabolisms

Serine biosynthesis In proliferating cells, serine is used towards nucleotide *de novo* synthesis. Different from normal tissue cells, cancer cells synthesize serine from glucose and glutamate in addition to uptake from circulation via transporters SLC1A4/A5. Previous studies suggest the following as possible reasons for increased serine *de novo* synthesis in cancer, even when the amino acid is abundantly available in circulation [81, 82]: (1) increased expression of PHGDH may be essential for rapid proliferation of cancer; and (2) its uptake may deplete serine in circulation.

The overall reaction of serine biosynthesis (Figure S1) from 3-phospho-D-glycerate (3PG) can be written as:

$$3\text{PG} + \text{glutamate} + \text{NAD}^+ + \text{H}_2\text{O} \longrightarrow \text{serine} + \text{Pi} + \text{2-oxoglutarate} + \text{NADH} + \text{H}^+,$$

or as:

$$\tfrac{1}{2}\,\text{glucose} + \text{glutamate} + 2\,\text{NAD}^+ + \text{H}_2\text{O} \longrightarrow \text{serine} + \text{2-oxoglutarate} + 2\,\text{NADH} + 3\,\text{H}^+$$

if the synthesis starts from glucose. Hence, the synthesis generates (at least) three H^+ per serine while its uptake is pH neutral.

We have examined the expressions of the serine biosynthesis pathway and SLC1A4/A5. Of the three enzyme genes in the synthesis pathway, PHGDH and PSAT are known to be involved in other pathways [83]. Hence we have used the expression of PSPH to represent that of the pathway. We note: (1) PSPH is upregulated in 11 cancer types (minus KICH, PRAD, THCA); (2) the expression pattern of PSPH largely correlates with those of SLC1A4/A5 as a cancer progresses from the early to the advanced stage across all 14 cancer types, as illustrated in Figure S2 and detailed in Table S2. Overall, increased serine biosynthesis

increases proton production. The majority of cancer types utilize serine biosynthesis in addition to its uptake from circulation.

Tryptophan degradation Tryptophan is degraded via the kynurenine pathway in normal human cells to acetyl-CoA (Figure S3). We note from the relevant expression data (Table S3): this pathway is considerably altered in cancer [84]. Specifically, cancer uses a truncated rather than the whole pathway: (1) the first gene IDO1/TDO2 of the pathway is up-regulated in 13 cancer types (minus THCA); (2) KYNU is up-regulated in 11 cancer types (minus KIRP, PRAD, THCA) but HAAO, the enzyme catalyzing the next reaction, is down-regulated in 12 cancer types (minus COAD, KIRC); and (3) the expression patterns of KMO, the enzyme that converts kynurenine to 3-hydroxy-L-kynurenine, are complex, with some up-, some down-regulated and others unchanged. Putting these together, tryptophan degradation has two main products: kynurenine and 3-hydroxyanthrranliate rather than the usual acetyl-CoA.

This gene-expression data-based observation is supported by the following. Kynurenine has been found to accumulate in some cancers [85], and can be extracellularly released via the ABCC4 transporter [86], which is up-regulated in nine cancer types (minus BLCA, KIRC, LUAD, LUSC, THCA). Published studies suggest that both kynurenine and 3-hydroxyanthrranliate can promote cell survival under immune attacks [87].

The overall reactions leading to these two products can be written as follows:

$$\text{tryptophan} + H_2O + O_2 \longrightarrow \text{kynurenine} + \text{formate} + H^+, \text{ or}$$

$$\text{tryptophan} + H_2O + \text{NADPH} + 2\,O_2 \longrightarrow \text{3-hydroxyanthrranliate} + \text{alanine} + \text{formate} + \text{NADP}^+ + H^+.$$

Our interpretation of cancers using this truncated pathway is: (1) each end-product is coupled with the production of one net H^+ and can be extracellularly released by cancers; and (2) if the whole pathway were used, it would consume one H^+ and release two CO_2, hence possibly alkalizing the intracellular space[1].

[1] CO_2 can be either hydrated to HCO_3^- and H^+ catalyzed by carbonic anhydrase or released from cells; the ratio between the two depends on multiple factors such as the catalysis rate vs. the diffusion rate of CO_2.

We have also examined the reprogrammed proline biosynthesis (Figure S4, Table S4), lysine degradation (Figure S5, Table S5), glutaminolysis (Figure S6, Table S6), glycine degradation and arginine utilization in the 14 cancers (Table S7), with a similar observation: each reprogrammed metabolism produces more protons than the original one, as detailed in Supplementary Results 1(a).

2.4.2 Imbalanced nucleotide metabolisms

Imbalanced purine and pyrimidine pools Cancer cells are known to have elevated purine/pyrimidine deoxyribonucleotides (dRN) ratios compared to normal cells [34], which is supported by large-scale cancer genome analyses reporting that cancer genomes have abnormally elevated pyrimidine-to-purine transversion mutations but reduced purine-to-pyrimidine mutations [88]. Previous studies suggest this is a result of over-activation of certain oncogenes [34]. Note that the reactions given in Figures S7–S8 can be written as:

1. Purine *de novo* synthesis

    - 5-phospho-α-D-ribose-diphosphate + glycine + CO_2 + 2 10-formyltetrahydrofolate + 2 glutamine + 2 aspartate + 6 ATP + GTP + reduced thioredoxin $\longrightarrow$ dATP + 2 tetrahydrofolate + 2 glutamate + 2 fumarate + 6 ADP + GDP + 5 Pi + diphosphate + oxidized thioredoxin + 9 H^+.
    - ATP + NAD^+ + reduced thioredoxin + 2 H_2O $\longrightarrow$ dGTP+ 2 tetrahydrofolate + 3 glutamate + fumarate + 6 ADP + AMP + 4 Pi + 2 diphosphate + NADH + oxidized thioredoxin + 10 H^+.

2. Purine salvage pathway I

    - adenosine + 5-phospho-α-D-ribose 1-diphosphate + aspartate + 2 ATP + GTP + reduced thioredoxin $\longrightarrow$ dATP + NH_4^+ + α-D-ribose 1-phosphate + fumarate + 2 ADP + GDP + diphosphate + oxidized thioredoxin + H^+.

- guanosine + 5-phospho-α-D-ribose 1-diphosphate + 2 ATP + Pi + reduced thioredoxin $\longrightarrow$ dGTP + α-D-ribose 1-phosphate + 2 ADP + diphosphate + oxidized thioredoxin + H_2O.

3. Purine salvage pathway II

    - 2'-deoxyadenosine + 3 ATP $\longrightarrow$ dATP + 3 ADP + H^+.
    - 2'-deoxyguanosine + 3 ATP $\longrightarrow$ dGTP + 3 ADP + H^+.

4. Pyrimidine *de novo* synthesis

    - 5-phospho-α-D-ribose-diphosphate + 2 glutamine + aspartate + 6 ATP + FMN + reduced thioredoxin + 2 H_2O $\longrightarrow$ dCTP + 2 glutamate + 6 ADP + 4 Pi + diphosphate + $FMNH_2$ + oxidized thioredoxin + 5 H^+.
    - 5-phospho-α-D-ribose-diphosphate + glutamine + aspartate + 6 ATP + FMN + reduced thioredoxin + 5,10-methylenetetrahydrofolate + H_2O $\longrightarrow$ dTTP + glutamate + 6 ADP + 2 Pi + 2 diphosphate + $FMNH_2$ + oxidized thioredoxin + 7,8-dihydrofolate + 3 H^+.

5. Pyrimidine salvage pathway

    - 2'-deoxycytidine + 3 ATP $\longrightarrow$ dCTP + 3 ADP + H^+.
    - thymidine + 3 ATP $\longrightarrow$ dTTP + 3 ADP + H^+.
    - 2'-deoxycytidine + 3 ATP + 5,10-methylenetetrahydrofolate + H_2O $\longrightarrow$ dTTP + NH_4^+ + 3 ADP + 7,8-dihydrofolate.

Hence the production of a dATP, dGTP, dCTP and dTTP from a comparable set of molecules each generates 9, 10, 5 and 3 H^+ by *de novo* syntheses and 1, 1, 1 and 0–1 H^+ by salvage, respectively.

We have examined the expressions of the signature genes (Table 2.2) of these pathways (Tables 2.1 and S7), and noted: (1) purine *de novo* synthesis is more up-regulated than that of pyrimidine in 11 cancer types minus THCA, STAD and COAD (Figure 2.1a); (2) a similar pattern is also observed for purine salvage vs. pyrimidine salvage (Figure 2.1b); (3) purine *de novo* synthesis is more up-regulated than purine

salvage in all cancer types except for KICH and KIRC (Figure 2.1c); and (4) a similar pattern is observed for pyrimidine *de novo* synthesis vs. salvage but considerably less prominent (Figure 2.1d). We see a clear pattern here: a metabolism producing more protons is up-regulated in more cancer types.

Degradation of dRN We note from Figure S9 and Table S8 that degradation of dATP, dGTP, dCTP and dTTP each consumes 1, 0, 1, 0 H^+, and overall pyrimidine dRN degradation is slightly up-regulated than that of purine in 13 cancer types (minus THCA). Putting these and the above together, we conclude that it is the combination of elevated synthesis and reduced degradation of purine compared to those of pyrimidine that gives rise to the increased purine/pyrimidine ratio in cancer, which can be explained in terms of their level of acidification to the intracellular space.

Ribonucleotide degradation and conversion Cancer is known to have reduced pyrimidine and increased purine RN degradation and increased conversion to dRN. From Figures S10–S12, we note that degradation of guanosine, adenosine, uridine, and cytidine[2] consumes 0, 0, 2, and 3 H^+, respectively, which can explain why *DPYD* and *UBP1*, two key genes in pyrimidine degradation pathway are generally inhibited across all cancer types (Table S8) while *PGM2*, a key gene in purine degradation is expressed in all cancer types and up-regulated in 10 as its two end-products are more acidic than the two pyrimidine RNs based on their pKa values.

In addition, conversion from RN to dRN by RRM1 and RRM2 produces one H^+; and RRM1/2 are up-regulated in all cancer types (Table S8). Hence, the complex expression patterns in RN synthesis, degradation and conversion to dRN can be all be explained in terms of their production or consumption of protons. These observations are consistent with previous studies reporting that *DPYD* tends to be highly mutated in cancer [89], and *RRM1/2* are hyper-active in cancer [90].

[2] Nucleobases guanosine and adenosine are purines, whereas uridine and cytidine are pyrimidines.

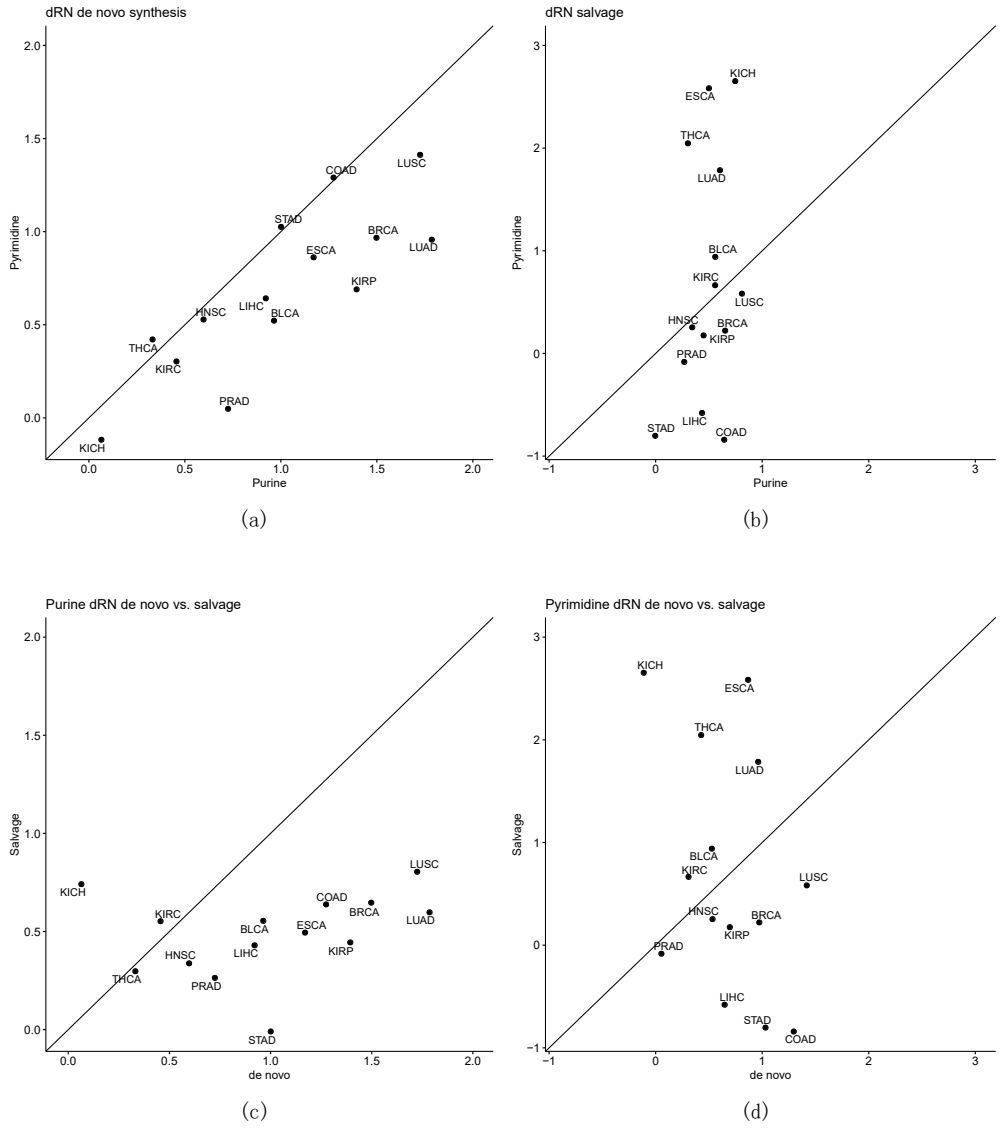


Figure 2.1: The levels of fold-changes in dRN *de novo* synthesis and salvage genes. **(a)** *De novo* syntheses of purine vs. pyrimidine. The x-axis denotes the average fold-change in expressions: $\frac{\sum FC_i(X)}{N_i}$ of *NME1*, *ATIC* and *PFAS*, marker genes of purine dRN *de novo* synthesis, with $FC_i(X)$ representing the fold-change in expression of gene X in cancer type i (one of the 14 types) vs. controls, and N_i being the number of cancer samples of type i. The y-axis is for the average fold-change in expressions of *CAD* and *CTPS1*, key genes in pyrimidine dRN *de novo* synthesis. **(b)** Salvage of purine vs. pyrimidine. The x-axis is the average fold-change of expressions of *APRT*, *DGUOK*, *GMPS* and *GUK1*, key genes in purine salvage and the y-axis is the average fold-change of expressions of *UPP1* and *CDA* in pyrimidine salvage. **(c)** Purine *de novo* synthesis vs. salvage. The x-axis is for the average fold-change in expressions of purine *de novo* synthesis genes and the y-axis is the average fold-change in expressions of purine salvage genes. **(d)** Pyrimidine *de novo* synthesis vs. salvage. The x-axis is for the average fold-change in expressions of pyrimidine *de novo* synthesis genes and the y-axis is the average fold-change in expressions of pyrimidine salvage genes.

2.4.3 Altered lipid metabolisms

Lipid metabolic reprograming is a most studied class of altered metabolisms in cancer, including: (1) increased synthesis, uptake and degradation of fatty acids [35]; (2) increased synthesis, storage, and degradation of triglycerides [91]; (3) increased synthesis and degradation of phospholipids [92]; and (4) increased synthesis of sphingolipids and derivatives. Previous authors suggest that the following may be reasons for the observed changes: (1) increased fatty-acid production is needed for anti-oxidation [93] and proliferation [94]; (2) their degradation is used towards energy production [95]; (3) elevated triglyceride synthesis is for energy reserve; (4) enhanced phospholipid synthesis is for making membranes; (5) increased degradation of phospholipid is for production of arachidonic acids, prostaglandins and leukotrienes [96]; and (6) elevated production of sphingolipids is for their roles in cell survival [97], pro-apoptosis [98], and growth signaling by their glycol-products, gangliosides.

Fatty acid uptake and synthesis Cancer cells are known to increase their uptake of fatty acids or derivatives such as lysophopholipids [99] and lipoproteins from circulation. Each of these fatty acids and derivatives has a pKa value considerably lower than the alkaline cytosolic pH of cancer cells, ranging from pKa = 1.8 for phosphatidylcholine to pKa = 4.5–4.8 for a free fatty acid. Hence each such molecule releases a H^+ once it gets inside a cancer cell. We note: the importers for fatty acid: SLC27A2–4 are collectively up-regulated in nine cancer types (Table S9).

Increased fatty acid synthesis has also been widely observed in cancer [100]. Our analysis revealed: ten cancer types have elevated fatty acid synthesis (Table S10). We note: *de novo* synthesis of a fatty acid (Figure S13) of $2n$ carbons from acetyl-CoA and malonyl-CoA consumes $2(n-1)$ (NADPH + H^+) and $(n-1)$ ATP, and produces *n* CO_2 + 2 (n-1) $NADP^+$. If the ATPs are synthesized via glycolysis (Warburg effect), hydrolysis of each ATP releases one net H^+ (see Page 26). Hence, the overall process consumes (n-1) H^+ and produces *n* CO_2. A natural question is: why are these synthesis genes up-regulated while they consume net H^+? An answer comes from the following.

Fatty acid degradation by beta-oxidation Fatty acid degradation by beta-oxidation (Figure S14) has been found to be persistently increased in some cancer types [35], whose reaction can be written as:

$$C_n\text{-acyl-CoA} + \text{FAD} + \text{NAD}^+ + \text{H}_2\text{O} + \text{CoA} \longrightarrow C_{n-2}\text{-acyl-CoA} + \text{FADH}_2 + \text{NADH} + \text{H}^+ + \text{acetyl-CoA}.$$

Hence, for a fatty acid of $2n$ carbons, the process produces $(n-1)$ H^+ when it is degraded to acetyl-CoA. By integrating this and the above subsection, we have: when a fatty acid of n carbons is synthesized and then degraded, the process produces n net CO_2. We note from Table S11 that beta-oxidation has increased expressions in six cancer types.

It is noteworthy that multiple cases of an up-regulated biosynthesis together with its up-regulated degradation have been observed in cancer, such as fatty acid synthesis and degradation or triglyceride synthesis vs. degradation. Our interpretation of such observations is that cancers utilizes a variety of means to acidify their intracellular space but to sustain, the cells must find ways to get rid of the "by-products" of each proton-producing process. For some RMs, cancer may directly secrete the by-products such as hydroxyl compounds as cancer prefers doing in its early stage. For others, direct release may not be beneficial to the cells when such products are acidic, such as fatty acids, hence their release will increase the intracellular pH.

Triglyceride synthesis and degradation Triglyceride synthesis (Figure S15(a)) is known to be up-regulated in multiple cancer types, which produces one H^+ for each round of the dephosphorylation and phosphorylation cycle that can potentially go on indefinitely so long as CTP is available. This provides an explanation to a long and perplexing observation: cancers tend to have increased CTP synthesis [101]. We note: 13 cancer types have increased CTP synthesis by *CTPS1* and *CTPS2* genes. Overall, ten cancer types have elevated expression of the dephosphorylation and phosphorylation cycle, measured using *PLPP4* (Table S12).

Degradation of a triglyceride to three fatty acids and a glycerol (Figure S15(b)) produces three H^+, where a fatty acid can be further degraded as discussed earlier. 11 cancer types have elevated triglyceride degradation (Table S13).

We have also examined the reprogrammed phospholipid biosynthesis (Figure S16, Table S14) and degradation (Figure S17, Table S15) as well as sphingolipid biosynthesis (Figure S18, Table S16) with similar observations to the above as detailed in Supplementary Results 1(c). Figure S14 summarizes the overall reprogrammed lipid metabolism studied in this section, which is integrated based on HumanCyc pathways.

2.4.4 Altered sugar metabolisms

Altered sugar metabolism was the first observed among all RMs in cancer, specifically the Warburg Effect [65], which is increased ATP production by aerobic glycolysis vs. respiration. A few other RMs of sugar have also been observed in cancer: (1) simultaneous activation of glycolysis and gluconeogenesis; (2) increased synthesis and deployment of sialic acids; (3) increased glycosylation of both O-type and N-type; (4) increased UDP-sugar synthesis; and (5) increased synthesis of glycosaminoglycan. Various proposals have been made regarding their possible reasons: (1) the Warburg effect provides a faster way for ATP production; and all proliferating cells have the Warburg effect [65]; (2) utilization of gluconeogenesis in cancer is for activating its PCK enzymes; (3) increased sialic acid production and deployment are for their influence to immunity [67] and their roles in cancer metastasis [102]; (4) increased glycosylation is to help cancer progression; (5) increased glycosaminoglycan syntheses are associated with growth factors and signaling roles [103, 104]; and (6) increased UDP-sugar production is for extracellular signals. Again, we provide one simple and consistent explanation to all these changes.

The Warburg Effect We have previously proposed [20] that the Warburg effect is induced to produce net H+, which is fundamentally different from why normal proliferating cells have this effect [74]. Specifically, glycolytic ATP generation can be written as:

$$\text{glucose} + 2\,ADP^{3-} + 2\,HPO_4^{2-} \longrightarrow 2\,\text{lactate}^- + 2\,ATP^{4-}$$

which is pH neutral, while respiration-based ATP production, written as:

$$ADP^{3-} + HPO_4^{2-} \longrightarrow ATP^{4-} + OH^-$$

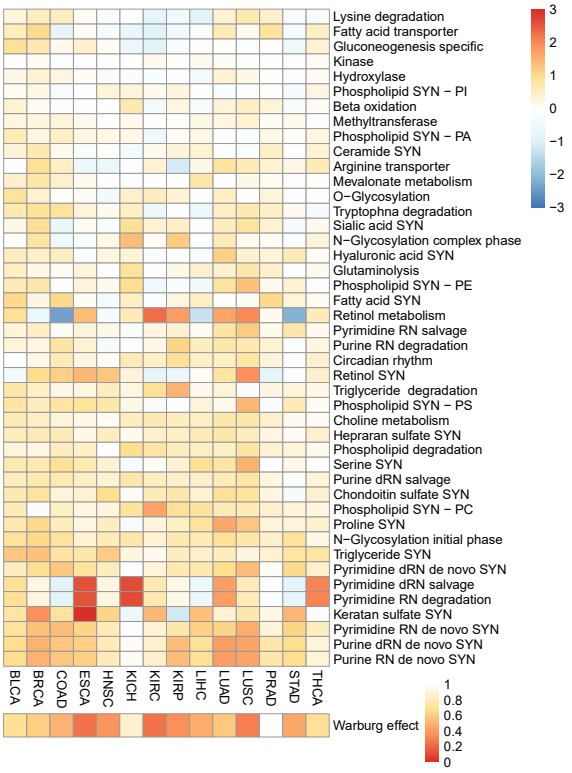
consumes one H^+ for each ATP produced [20]. And hydrolysis of an ATP: $ATP^{4-} + H_2O \longrightarrow ADP^{3-} + HPO_4^{2-} + H^+$ releases one H^+ regardless of how the ATP is produced. Hence, glycolytic ATP produces one net H^+ when it is hydrolyzed, while ATP generation by respiration is pH neutral when it is used.

We have used the fraction of the glycolytic flux going into the TCA cycle to estimate the level of the Warburg effect, estimated using the ratio between the gene expressions of *PDHB* and *PKM*; hence the smaller the ratio, the higher the Warburg effect. We note that all cancer types have the Warburg effect with PRAD having the lowest level (Figure 2.2a), which is known.

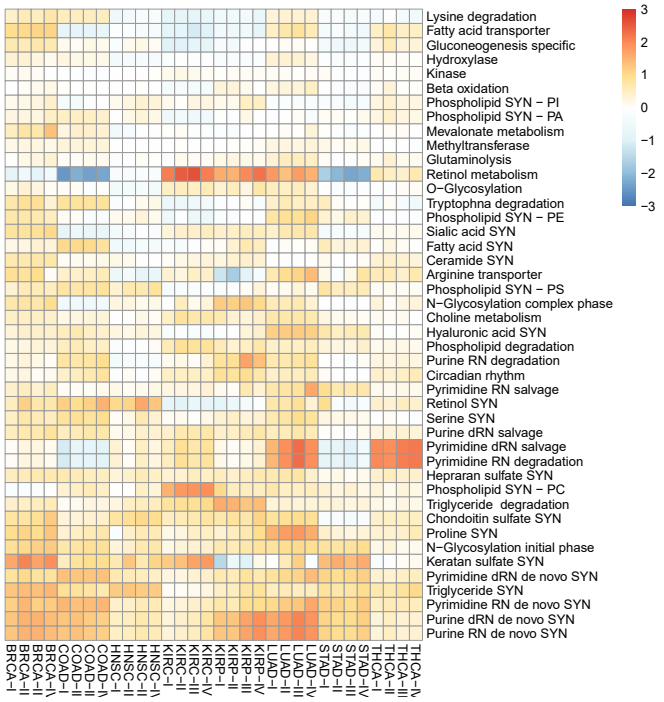
Protein glycosylation Both O-linked and N-linked glycosylation of proteins has been found to be up-regulated in a variety of cancer types [105]. The largest class of up-regulated O-glycosylation in cancer is the mucin type, which starts with the transfer of N-acetylgalactosamine (GalNAc) to a serine or threonine, generally followed by adding two consecutive structures: cores 1 and 2. Each O-glycosylation including cores produces 3–4 H^+ (Figure S19), and the signature genes involved are up-regulated in ten cancer types (Table S17).

N-glycosylation consists of three phases: the initial phase, the processing phase, and the complex synthesis phase. The initial phase produces 14 H^+ in each round; the processing phase is pH neutral; and the complex synthesis phase produces 3–10 H^+ (Figure S20). We note: the number of cancer types having a specific phase up-regulated is highly consistent with the number of H^+ produced by the phase (Table S18 and Figure 2.2).

We have also examined reprogrammed gluconeogenesis (Figure S21, Table S19), sialic acid synthesis (Figure S22, Table S20), and glycosaminoglycan syntheses (Figure S23, Table S21). We made similar observations that the number of cancer types having an RM up-regulated is highly consistent with the number of protons produced or consumed by each pathway. Details are shown in Supplementary Results 1(d).



(a) Heatmap showing the levels of differential expressions of reprogrammed metabolisms across 14 cancer types. The level of the Warburg effect is estimated using the fraction of the glycolytic flux going into the TCA cycle, measured using the ratio between the expressions of *PDHB* and *PKM* genes, using a different coloring scheme ranging from the highest level (1.0) of the Warburg effect to the lowest level (0.0).



(b) Stage-dependent heatmap for eight cancer types, representing all the cancer types among the 14 with each stage having a sufficiently large number of samples.

Figure 2.2: Levels of differential expressions of reprogrammed metabolisms. Each row is for one reprogrammed metabolic pathway. Each column is for one of the 14 cancer types, or a stage of the eight cancer types. Each entry is the $\log_2$ value of fold-change of signature genes of the relevant reprogrammed metabolism, represented using a color scheme in the figures.

2.4.5 Reprogramming of selected other metabolisms

By going through the expression data of all the metabolic genes in cancer, we have noted that metabolic reprogramming takes place in virtually every aspect of cancer biology and the level of changes is also

strongly associated with the number of net protons produced or consumed. We have examined a few additional cases to illustrate the extensiveness of RMs in cancer.

Phosphorylation Cancer tends to up-regulate kinase genes to phosphorylate a large variety of molecules, which transfers a phosphate from an ATP to a target molecule and releases a H^+. We have examined the expression data of 512 kinases encoded in the human genome and noted that kinase genes tend to be up-regulated in all cancer types (Figure S24(a)). Interestingly, some phosphatase genes, genes for dephosphorylation, are also up-regulated (Figure S24(b)). Our interpretation is: the kinase genes are up-regulated to generate H^+ and the up-regulated phosphatase genes, whose reaction is pH neural or acidifying in some cases, are used to remove the phosphoryl group to make room for re-phosphorylating the same sites, hence producing more H^+.

Choline metabolism Choline metabolism is known to be up-regulated in multiple cancer types [106]. However, the current understanding about its functions in and contributions to cancer remains fragmented [106]. While some enzymes are found to be up-regulated, the known pathways for choline synthesis and metabolism (Figure S25) do not match well with the gene-expression data. We have predicted a reprogrammed pathway for choline synthesis and metabolism, which is most consistent with the gene-expression of the 14 cancer types, as shown in Figure S26. The predicted model is largely consistent with a model for diseased tissues published in KEGG. From the figure, we can see that four H^+ are produced per round of the pathway, which is up-regulated in 11 cancer types (Table S22).

NAD^+/NADH metabolism It has been widely observed that cancer tends to be low in NAD^+ [107]. We have examined the synthesis and metabolic pathway of NAD^+/NADH (Figure S27) and the gene expression data (Figure S28). Among the three enzymes whose reactions synthesize NAD^+, NADSYN1 and QPRT are up-regulated in 14 and 9 cancer types, and their reactions each produce one H^+ and one CO_2, respectively; and NMNAT1,2,3 is up-regulated in two cancer types (LUAD and LUSC) and slightly up-regulated (fold-change within 2.0) in four additional cancer types (ESCA, HNSC, LIHC and STAD),

where their reaction each consumes one H^+. For the efflux from NAD^+, PARP10,12 both are up-regulated in 12 cancer types and each produce one H^+. In addition, the reaction from NAD^+ to NADH generally produces one H^+ and the reverse reaction consumes a H^+. Hence the equilibrium will move towards the NADH production in an alkaline environment as in cancer, which we predict is the reason for NAD^+ deficit in cancer. Previous studies have established that PARP genes play important roles in tumorigenesis in a few cancer types, hence supporting our model.

We have also examined reprogrammed retinol metabolism (Figure S29, Table S23), methylation (Figure S30), hydroxylation[3] (Figure S31), mevalonate metabolism (Figure S32, Table S24), circadian rhythm (Table S25) and ATP utilization (Figure S33) with similar observations to the above, as detailed in Supplementary Results 1(e).

Around 50 RMs are analyzed in terms of their differential expressions along with the numbers of protons that they each produce or consume derived from the relevant biochemical reactions, which is summarized in Figure 2.2(a) and Table S26. A key discovery made is: all the RMs under study each produce more protons compared to their original metabolisms, strongly suggesting that pH has a strong impact on the differential expression of each RM.

We have also examined how the average level of each RM changes as a cancer advances from stage I through stage IV[4], as shown in Figure 2.2(b), and noted that there are three major distinct patterns in the RM levels: (1) monotonically going up (or down for a small number of cases); (2) remaining at the same level; and (3) going up through the early stages (e.g., stages I–III) and then down at stage IV. We discuss the possible reasons for these observations in the following section.

2.4.6 RMs Are Responses to Fenton Reactions

We have previously reported that cancer tissue cells tend to inhibit their H^+ exporters (except for the lactic acid exporters) and up-regulate H^+ importers, which are exactly the opposite to normal proliferating cells [74]. And yet, cancer tissue cells are known to have elevated intracellular pH compared to the match-

[3]This process produces hydroxyl-compounds.

[4]Only eight cancer types are included since only those have sufficiently large number of cancer samples in each stage.

ing normal ones [75]. Hence it is only natural to assume that cancer cells have un-recognized metabolic processes that persistently produce large quantities of OH^- or equivalents. It is noteworthy to reemphasize that it has been well established that cancer tissue cells are under intracellular alkaline stress but the sources remain unsettled [9, 75–77].

We have recently discovered that all the 14 cancer types under study have Fenton reactions: $Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + {}^{\bullet}OH + OH^-$ in their cytosol [20] through mining the TCGA data and statistical modeling. It has been well established Fenton reactions are the result of two factors: (1) increased activities of the local innate immune cells, namely macrophages and neutrophils, which release large quantities of H_2O_2 and $O_2{}^{\bullet -}$, referred to as respiratory bursts on a persistent basis; and (2) local accumulation of iron, which has been reported in a wide range of cancers [108]. With the continuous availability of $O_2{}^{\bullet -}$ serving as a reducing molecule of Fe^{3+}, the reaction can be rewritten as: $O_2{}^{\bullet -} + H_2O_2 \longrightarrow {}^{\bullet}OH + OH^- + O_2$ with Fe^{2+} as a catalyst [20]. It has been established that cancer tissue cells harbor persistent Fenton reactions [109] but previous studies generally focus on the damaging effect of hydroxyl radicals while we have been focusing on the impact of the seemingly harmless but persistently produced OH^-.

We have previously demonstrated that cytosolic Fenton reactions as observed in cancer tissue cells can quickly overwhelm the cytosolic pH buffer [20], hence driving up the pH if not neutralized. Note that human cells generally have a narrow range of cytosolic pH such as $[6.5, 7.5]$ to remain viable. Our question is: are the acidifying RMs studied here responses to cytosolic Fenton reactions? To assess this possibility, we have conducted a regression analysis of the estimated cytosolic Fenton reaction levels against the predicted levels of RMs across all samples of each cancer type (see Methods).

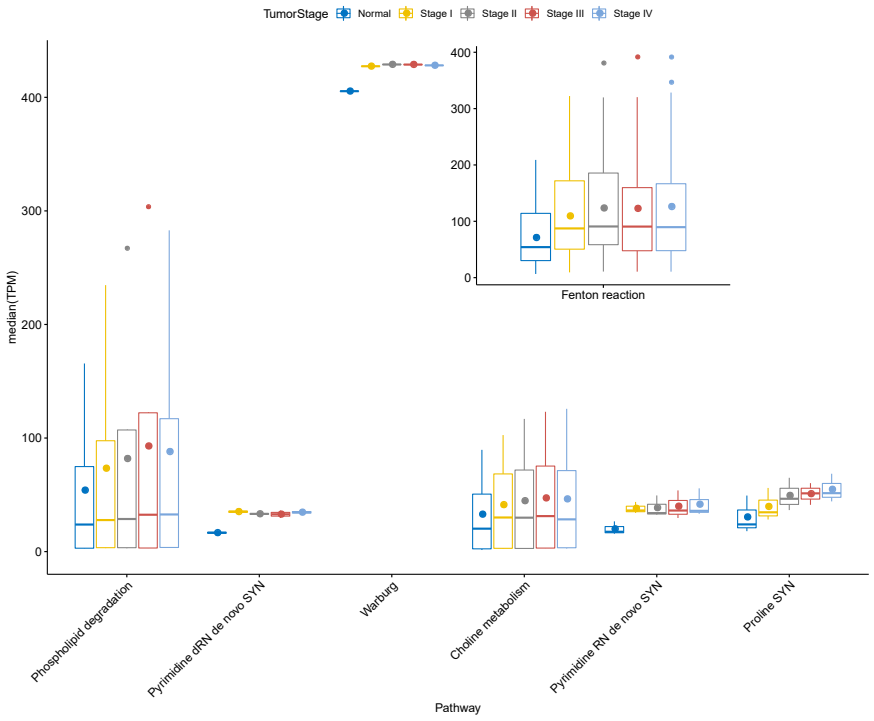
Table S27 summarizes the regression results. Clearly, all 14 cancer types have highly significant R^2 values between the predicted levels of Fenton reactions and combined RM levels plus expression levels of the acidifying and alkalizing transporters that we previously studied [74], hence indicating that Fenton reactions could be the reason for the induction of all the RMs. We note that while different combinations of RMs are selected in the regression model against Fenton reactions across different cancer types, some RMs are used by commonly by multiple cancer types, such as phospholipid degradation, sialic acid biosyn-

thesis, dRN *de novo* synthesis, Warburg effect, beta-oxidation, N-linked glycosylation, serine biosynthesis, and methylation. These results are biologically meaningful since (1) nucleotide *de novo* synthesis and the Warburg effect have been universally found in all cancers, where serine biosynthesis is used to feed into nucleotide *de novo* synthesis; (2) sialic acid has long been found to be associated with all cancer metastasis since 1960s [110]; and (3) glycosylation has been widely observed throughout cancer development; and similar has been reported about methylation and lipid degradation4. Some RMs are not widely selected by the regression models. Our examination has revealed two reasons: (1) some such RMs strongly correlate with a selected RM such as purine RN *de novo* synthesis, making them redundant and hence not used in regression models; and (2) some indeed show relatively weak correlation with Fenton reactions such as hydroxylase genes.

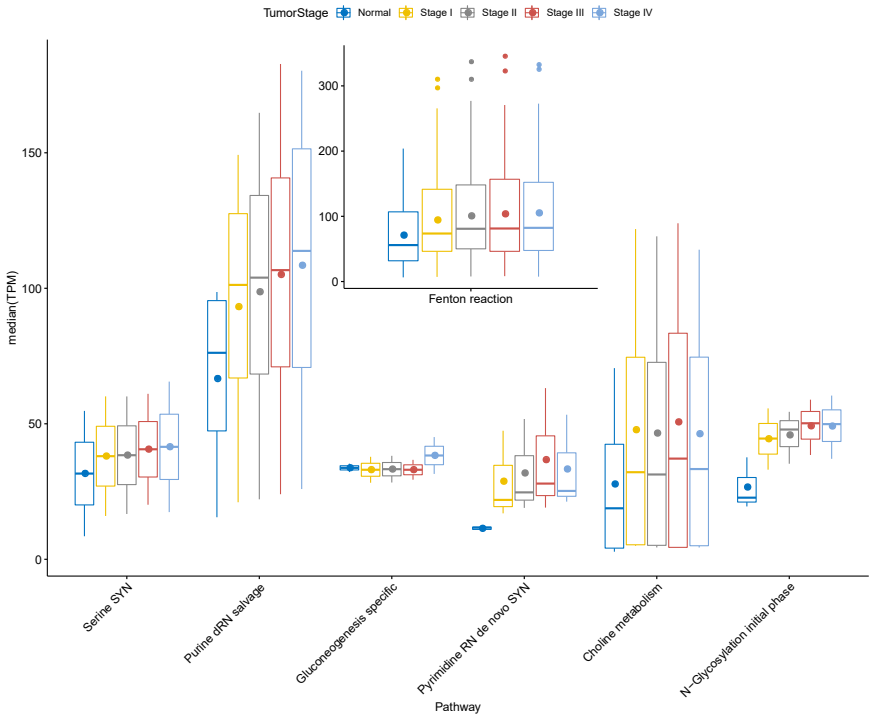
Figure 2.3 and Figure S34 show the predicted levels of Fenton reaction vs. the most significant RMs across four stages for eight cancer types, where Figure 2.3(a) and (b) are for HNSC and LUAD, respectively (see Supplementary Results 1(e)). We can see clearly the correspondence between the levels of Fenton reactions and those of the RMs selected by our regression models. It is noteworthy that the levels of Fenton reactions go down at stage IV from the peak level in more than 50% cancer types, which may explain a similar pattern observed of RMs discussed in Supplementary Results 1(e).

To provide biological evidence that each RM indeed contributes to neutralization of OH^- produced by Fenton reactions, we have conducted a comprehensive literature survey aiming to find if inhibition of some key enzymes in each of the RMs may slow down or kill cancer cells by previous studies. To our surprise, inhibition of some enzymes in every RM studied here will kill or slow down cancer cells, or sensitize them to other drugs as detailed in Table S28. Note that while we list only one enzyme for each RM, more enzymes for each RM have been generally reported as inhibitory targets in the literature.

For each cited study in Table S28 the authors made varying proposals regarding why inhibition of each target leads to the observed pharmaceutical effects. Compared to these proposals, we add one more, which is consistent across all the RMs: the inhibition of each RM enzyme reduces the neutralization power, hence making the cancer cells less viable.



(a) HNSC.



(b) LUAD.

Figure 2.3: Predicted levels of Fenton reactions versus levels of top six RMs ($P < 0.05$) that are selected as the most significant contributors (ordered from left to right in the descending order of statistical significance) to the regression result. Results for six other cancer types are shown in Figure S34.

2.5 Discussion

2.5.1 RMs and intracellular alkalizing stress

A general observation made across all the 14 cancer types is: genes whose enzymes catalyze H^+-producing reactions tend to be more up-regulated than those that consume H^+ as shown in Figure S35. Such data point to the possibility that increasing the overall net H^+ production is a key force behind all the RMs studied here. This is also supported by our observation that published small-molecular biomarkers in blood for a few cancer types generally have alkaline pKas or are hydroxylated (Table S29).

To provide further evidence regarding possible connections between RMs and intracellular alkaline stress, we have conducted co-expression analyses between the RM signature genes and ER stress genes (which reflect pH-related stress) as well as between the RM genes and DNA polymerase genes. We note from Figures S36 and S37 that (1) RM genes tend to be more co-expressed with the ER genes than with the DNA synthesis genes except for a few; and (2) the RM genes that positively correlate with DNA genes are largely nucleotide and ceramide synthesis genes, hence providing strong evidence that RMs are predominantly related to stresses rather than proliferation.

To explore how the RMs might be regulated, we have examined co-expressions among genes in the same RMs, and noted: co-expressions among genes in the same RM from ten selected RMs are considerably weaker in cancer than in controls (Figure S38) . This suggests the possibility that genes in the same RM in cancer might be individually regulated transcriptionally rather than collectively by a predefined program. Hence, we posit that an RM may be largely the result of natural selection rather than a pathway-level regulation.

2.5.2 Finding metabolic exits for RMs: secondary and disease-defining stress?

Our analyses strongly suggests that each up-regulated RM is induced to produce H^+. Clearly, the persistent up-regulation will also lead to increased production of the other (non-H^+) products of the RM. For such an acidification process to sustain, cells must find ways to consume or get rid of the other products. For

example, triglyceride synthesis produces multiple H^+ per triglyceride (see Table S26). Then cancer cells find a new way to consume them via degrading them into glycerol and fatty acids, which will be then used to synthesize triglycerides, hence forming a synthesis/degradation cycle as revealed by our analyses.

A more significant example is: purine *de novo* synthesis is considerably up-regulated in all cancer types, presumably to produce H^+ to keep up with the rates of Fenton reactions as it represents the most powerful acidifier (measured in terms of the number of H^+ per nucleotide synthesized). But cells must find ways to get rid of the purine or pyrimidine in a sustained and timely manner. It is noteworthy that removal of such nucleotides out of cells in a sustained manner is a challenge since they are negatively charged, hence their removal without co-removal of positively charged molecules (or absorption of negatively charged ones) will alter the intracellular electric neutrality, a most fundamental property that cells must maintain to stay viable!

It is foreseeable that cancerous cell-division is Fenton reaction-affected cells' way of removing the nucleotides that are continuously synthesized as a major way for neutralizing Fenton reactions! While this seems to be far-fetching, it is how proliferation of all unicellular organisms is accomplished: nutrients drive cell division. Specifically, unicellular organisms such as *Escherichia coli* or yeast are known to first synthesize ATPs when nutrients are available. Once the ATP generation rates are higher than the consumption rates, hence leading to ATP accumulation, cells will gradually switch to nucleotide synthesis and use the nucleotide-sugar concentration as cue for cell-cycle activation and cell division [111, 112]. Potentially cancer cells may have somehow activated a similar program to get rid of the continuously generated nucleotides (along with positively charged histones in yeast or similar proteins in bacteria) through cell division, to sustain nucleotide synthesis as a key acidifier to keep the affected cells alive. A recent study has shown that cancer tends to rely heavily, almost solely, on most ancient genes from unicellular organisms [113], which is clearly consistent with our prediction here.

A systematic analysis is currently under way to study connections between a few intrinsic behaviors of cancer such as cancer metastasis, reduced sodium levels in circulation, and cachexia and the RMs. Our data strongly indicate that these clinical behaviors may mostly be related to finding metabolic exits for some

RMs while maintaining intracellular electric neutrality. For example, the synthesis of SA, a negatively charged 9-carbon sugar, produces two H^+ per SA (Table S26) and is deployed on the cancer cell surface, giving rise to their highly elevated deployment on cell surface via densely packed gangliosides [114] and forming unusually long poly-SA [67]. The gradual accumulation of such negatively charged molecules on cancer cell surface will give rise to increasingly stronger cell-cell repulsion and mechanical stress due to compression-induced deformation of cell shapes, as well as migration.

We have recently developed a mouse liver disease/cancer model, induced by a carcinogen, and discovered: RMs take place way before a liver cancer is formed, initially of simple types such as increased hydroxylation, phosphorylation, and syntheses of monocarboxylic acids — all acidifying RMs; and they become more diversified as the disease evolves toward cancer formation. Based on all these and the analyses throughout the article, we have developed a model of how RMs may induce new stresses to the host cells in terms of removal of the overproduced end-products of the RMs (Figure 2.4).

2.5.3 Transcriptomic vs. metabolomics data

Our analyses are based on transcriptomic data. It is noteworthy that numerous articles have demonstrated that metabolite levels generally positively correlate with the gene expression levels of the relevant enzymes in various cancer types [115–118], hence providing strong support to the conclusions derived here.

2.5.4 Subcellular locations of pH-producing and consuming enzymes

Throughout this article, we did not distinguish the detailed subcellular compartments where an enzyme resides; instead, we consider them all intracellular, hence contributing to the intracellular pH. An implicit assumption is that cells have their internal mechanisms to coordinate the relative pH across different compartments, which has been proposed to be the case [119].

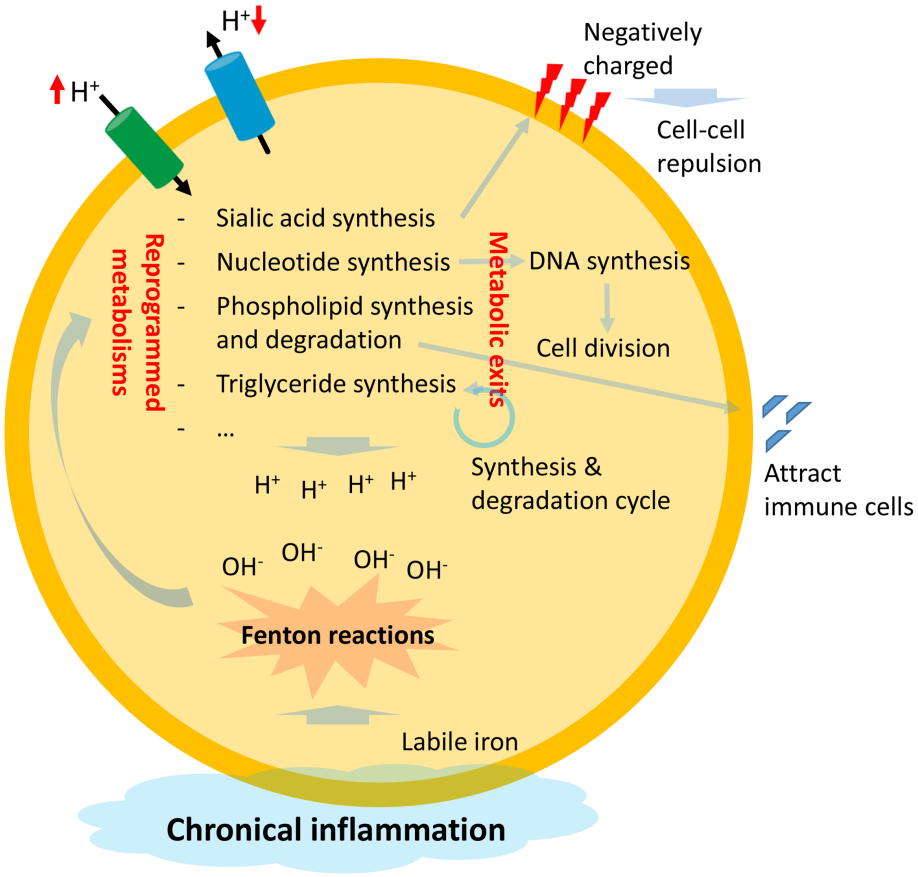


Figure 2.4: An illustration of our model for the possible causes of the RMs and their impacts on cellular behaviors. Specifically, it is Fenton reactions, due to chronic inflammation and local iron overload, that may drive the reprogramming of numerous metabolisms in a conserved manner across different cancer types. Then it is the need for finding metabolic exits for some products of the RMs that may give rise to different behaviors of cancer cells.

2.5.5 A new and improved framework for studying cancer biology?

This RM-based framework has enabled us to offer natural explanations to numerous long-standing open questions in cancer research, such as (1) the possible cause of the Warburg effect; (2) why cancer cells tend to produce nucleotides via *de novo* synthesis versus salvage; (3) why cancers produce more purine than pyrimidine; (4) why SA tend to associate with cancer metastasis, which was first reported in the 1960s; (5) why cancer cells tend to run triglyceride synthesis and degradation as well as glycolysis and gluconeogenesis

in parallel; and (6) what may determine the rate of cancer proliferation; and numerous others discussed throughout this article.

This framework may also enable to identify functional associations among mutations in cancer, which are now largely disconnected from each other functionally. For example, we noted that some mutations are selected to enable specific RMs in a sustained manner as well as to help cancer cells to overcome inhibitory functions encoded in our genome.

Throughout our analyses, we noted that cancer cells may generate H^+ at the expense of ATPs. Some other authors have made a similar observation: cancers seem to have more ATPs than matching normal tissues [120]. Our previous prediction was: cancers also have Fenton reactions in mitochondria, giving rise to a new pathway for ATP production via the ATP synthase with electrons coming from immune cells in the form of superoxide [20].

2.6 Conclusion

We have analyzed a large collection of RMs commonly observed across multiple cancer types, and discovered that they all produce more H^+ compared to their original metabolisms; or more generally cancer tends to up-regulate genes whose proteins are H+ producing and down-regulate genes whose proteins consume H^+. This, coupled with our previous studies, strongly points to the possibility that cancer is about overcoming a persistent alkalizing stress via reprogramming its metabolisms at a whole cell level and finding ways to get rid of the over-produced other end-products of RMs in a sustained manner. Numerous perplexing and long-standing open questions can be naturally explained by our model in a simple and consistent manner.

Chapter 3

Metabolic Reprogramming is Linked to Microenvironment Stress that Drives Tumor Proliferation and Metastasis

Huiyan Sun (co-first author), Yi Zhou (co-first author), Hongyang Jiang, Ying Xu. (2020) *Front. Oncol.*; **10**:401. Reprinted here with permission of the publisher.

3.1 Abstract

Sialic acids (SA), negatively charged nine-carbon sugars, have long been implicated in cancer metastasis since the 1960s, but its detailed functional roles remain elusive. We present a computational analysis of transcriptomic data of cancer vs. control tissues of eight types in TCGA, aiming to elucidate the possible reason for the increased production and utilization of SAs in cancer and their possible driving roles in cancer migration. Our analyses have revealed the synthesis and deployment enzymes of SAs are persistently up-regulated throughout the progression for all but one cancer type. Gangliosides, of which SAs are part, tend to converge to specific types that allow SAs to pack at high densities on cancer cell surface as a cancer advances for all cancer types. Statistical and modeling analyses suggest that a highly plausible reason for the increased syntheses of SAs is to produce net protons, used for neutralizing the OH^- persistently generated by elevated intracellular iron metabolism coupled with chronic inflammation in cancer tissues. The level of SA accumulation on cancer cell surface strongly correlates with the stage of cancer migration, as well as multiple migration-related characteristics such as altered cell-cell adhesion, mechanical stress, cell protrusion, and contraction. The pattern of SA deployment correlates with the five-year survival rate of a cancer type. Overall, our study provides strong evidence for that the continuous accumulation of SAs on cancer cell surface gives rise to increasingly stronger cell-cell repulsion due to their negative charges, leading to cell deformation by electrostatic force-induced mechanical compression, which is known to be able to drive cancer cell migration established by recent studies.

3.2 Introduction

It has been observed that increased syntheses of sialic acids (SAs) are associated with cancer development and metastasis since the 1960s [49, 50], where SAs are negatively charged nine-carbon sugars and generally serve as the capping molecules of cell-surface glycans as part of plasma membrane-embedded gangliosides [51]. Under physiological conditions, brain tissues have the highest concentration of SAs used for synaptogenesis. Outside of brain, red blood cells have the highest cell-surface concentration of SAs. As

of now, it remains largely unknown of why cancer cells produce unusually large numbers of SAs and accumulate them on cell surface [52]. Published studies have been mostly focused on their signaling roles with other cell types such as immune and endothelial cells, via binding to siglecs[1] and selectins, to facilitate interactions between cancer and immune cells [53] and to enable cancer cells interactions with and penetration into blood vessels [49], respectively. Very little has been established regarding if SAs may play roles in creating mechanical compression within cancer tissues. They are known to be negatively charged and increasingly placed on the cell surface, hence possibly resulting in increasingly stronger cell-cell repulsion. Mechanical pressure has been widely observed in cancer tissues but largely attributed to the confined space for growing tumors [54, 55].

We have recently studied 44 reprogrammed metabolisms widely observed in cancer, including persistent SA synthesis, and discovered that each of them produces more protons than its original metabolic route [47]. In addition, we have also discovered that all cancer tissue cells harbor Fenton reaction[2] in their cytosol, and the rates of OH^- production can saturate the intracellular pH buffer within days. If not neutralized, this increases the intracellular pH [20] and posts a major stress to the host cells. A linear regression analysis was conducted of the predicted Fenton reaction level against the predicted levels of all reprogrammed metabolisms (Figure S1), which achieves high R^2 values with statistically significant p-values for each cancer type [47]. This strongly suggests that these reprogrammed metabolisms are induced to respond to cytosolic Fenton reaction, serving as neutralizers for the OH^- persistently produced by the Fenton reaction.

In this context, we present a computational study of transcriptomic data of SA-related vs. migration-related genes in cancer tissues of eight cancer types from the TCGA database [121]. Our analyses have revealed: (1) majority of the cancer types has increased production and deployment of SAs on the cell surface, where the synthesis of a SA generates two net protons; (2) the level of SA synthesis correlates with the level of cytosolic Fenton reaction for all studied cancer types; and (3) as a cancer progresses, it tends to converge to use of specific types of gangliosides, facilitating SA packing at high densities. Further analyses

[1] Siglecs are sialic acid-binding immunoglobulin-type lectins.

[2] $Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + {}^{\bullet}OH + OH^-$.

lead to a simple model for predicting the level of SA accumulation on the cell surface that can explain cancer progression from stags N_0 through N_3 and then stage M[3]. We observed strong correlations between a range of cell migration-associated characteristics, such as increased mechanical stress, cell contraction and protrusion, and SA production and/or deployment. The detailed expression patterns of SA synthesis and degradation genes can statistically well explain the average five-year survival rates of each cancer type, hence the level of metastasis as the survival rate strongly correlates with the level of metastasis across all eight types. Overall, our analyses provide strong evidence that the SA accumulation on cancer cell surface plays key roles in mechanical compression and cell deformation in cancer tissues.

It has been observed that cancer tissues have strong mechanical pressure within [54, 55]. It was suggested that this is due to the confined space for the growing tumors. However, the "confined space" argument may not be supported by experimental data. For example, skin melanoma starts to metastasize as soon as the cancer starts to grow vertically, which is clearly not confined by space. Our analyses suggest that mechanical pressure strongly correlates with the cell-surface accumulation level of SAs. A previous study has convincingly demonstrated that mechanical compression can lead to cell deformation, which can drive the activation of actomyosin filaments and associated contractile motion, ultimately driving collective migration by cell clusters with enhanced cell-cell adhesion [122]. By integrating all these together, we have developed a model for how SA synthesis and deployment, responding to cytosolic Fenton reaction, can give rise to increasingly stronger cell-cell repulsion, further leading to mechanical compression and cell deformation, which can drive cancer cell migration.

3.3 Materials and Methods

3.3.1 Data

Cancer survival data The data given in Table 3.1 are collected from the NCI TCGA website, which provide clear and detailed clinical information of each cancer patient.

[3]Here we are using the TNM stage notation, where N_i represents cancer tissues that have metastasized to i lymph nodes and M for distant metastasis.

Table 3.1: Cancer types and their transcriptomics data sample counts used in this study.

Abbreviation	Cancer type	Tumor sample count	Normal sample count
BLCA	Bladder urothelial carcinoma	414	19
BRCA	Breast invasive carcinoma	1109	113
COAD	Colon adenocarcinoma	480	41
HNSC	Head and neck squamous cell carcinoma	502	44
LUAD	Lung adenocarcinoma	535	59
LUSC	Lung squamous cell carcinoma	502	49
STAD	Stomach adenocarcinoma	375	32
PAAD	Pancreatic adenocarcinoma	178	4

Transcriptomics data We have used all the transcriptomic data of eight cancer types in the TCGA database. Table 3.1 summarizes the relevant information. Here we used eight cancer types instead of the 14 cancer types we typically use in our recent studies [20, 47, 74] because we stratify patients using the TNM staging scheme rather than the more traditional I–IV staging approach. Since a focus of the study is on the metastasis status, the TNM staging is clearly more appropriate.

Genomics data Somatic mutation and copy number variation data in TCGA were collected for each of the four SA genes from the UCSC Xena database [123]. Gene-level non-silent mutations calls and copy number variation estimates are used in our study.

DNA methylation data The beta values generated from the Illumina Human Methylation 450K platform were collected from the UCSC Xena database. Probes in the gene body, first exon, UTRs or within 1500bp upstream the transcription start site for each of the four SA genes were used.

Transcription factor binding data ChIP-Seq validated transcription factor-target gene pairs for each of the four SA genes were collected from the ENCODE database [124], TRRUST database [125], and Marbach et al. [126].

3.3.2 Computational methods

Prediction of regulatory mechanisms via regression analysis For each of the four SA genes, samples with somatic mutations or copy number variations in the gene were filtered out as we are interested in how the relevant transcription factors and DNA methylation co-regulate the expression of the gene.

TPM values for gene expression and M values for DNA methylation were first centered and scaled to have mean zero and standard deviation one. For each target gene G, we aimed to find real-valued parameters μ, $\{\alpha_i\}$, and $\{\beta_j\}$ that minimize the following function:

$$\left(g - \left(\mu + \sum_{i=1}^{p} \alpha_i x_{ei} + \sum_{j=1}^{q} \beta_j x_{mj}\right)\right),$$

where g is the expression level of the target gene G, $\{x_{e1}, \cdots, x_{ep}\}$ are the expression levels of the transcription factors that regulate G, and $\{x_{m1}, \cdots, x_{mq}\}$ are the DNA methylation levels of the probes in or around G in the genome. The least squares method was used to solve the optimization problem with Lasso penalty using the R package glmnet [127].

The analysis of variance table was computed using ANOVA to get the sum of squares (SS) for each parameter. The SS for groups x_e and x_m were summed up to get SS_{TF} and SS_{MT} for contributions by transcription factors and DNA methylation, respectively. $\frac{SS_{MT}}{SS_{TF}+SS_{MT}}$ was used to estimate the level of contribution by DNA methylation to regulate the transcription of target gene G.

Prediction of concentrations of gangliosides We have predicted the relative concentrations of all the gangliosides based on gene expression levels of the relevant enzymes. A key simplifying assumption is the K_{cat} values for all the enzymes involved in this metabolism are approximately the same. For each metabolite M, we calculated its total influx and efflux as

$$\sum_j M_j^i \times E(g_j) \text{ and } \sum_k M_k^o \times E(g_k),$$

where $M_j^i \rightarrow M$ is the j^{th} influx reaction, $M \rightarrow M_k^o$ is the k^{th} efflux reaction, and $E(g_j)$ is the expression level of gene g_j whose encoded enzyme catalyzes the j^{th} reaction. An iterative procedure was employed to estimate these two quantities until the system converged on $\{M_j^{i,o}\}$ values. We predicted M to be intracellularly accumulated if $\sum_j M_j^i \times E(g_j)$ is significantly higher than $\sum_k M_k^o \times E(g_k)$, indicating its use in the plasma membrane. Two levels of "significance", high and moderate, are used in Table 3.3.

Neural network model for prediction of metastasis Two separate neural networks were trained, one being a multi-classification model and the other a regression model. For each cancer type and each stage $\mathbb{N}_i$, 70% samples were randomly selected for training and the remaining 30% samples were used for testing. The process was repeated ten times for cross validation.

The classification model was evaluated using the macro F_1 score, which is the arithmetic mean of all per-class F_1 scores. The F_1 score is defined as the harmonic mean of precision and recall:

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}.$$

TP, FP, and FN denote true positives, false positives, and false negatives, respectively. The regression model was evaluated using the mean squared error (MSE) between the predicted and observed metastasis status:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2,$$

where $\hat{y}$ are the predicted C values. The training process was terminated when the number of iterations reached 1,000,000 or the loss function is less than 0.1.

3.4 Results

3.4.1 Elevated synthesis of sialic acids in cancer

We have examined the key genes involved in SA synthesis (*CMAS*) and degradation (*NEU1*). *CMAS* is up-regulated in seven of the eight cancer types (except for COAD); and *NEU1* tends to correlate with *CMAS* throughout the major portion of a cancer progression for all cancer types, except for the last stage(s), where the two curves may diverge or converge for some cancer types (Figure 3.1).

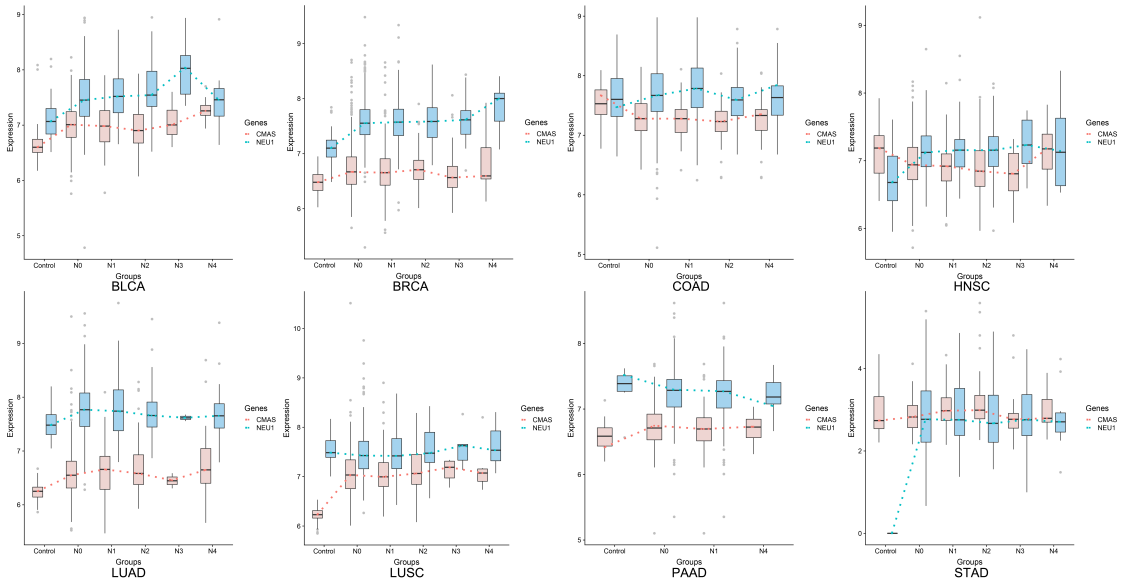


Figure 3.1: Average expression levels of *CMAS* and *NEU1*. The x-axis shows cancer-adjacent control, stages N0, N1, N2, N3, and M1 cancer tissues. The y-axis shows the average expression levels in transcripts per million.

Previous studies generally attribute the increased SA production to their signaling roles in cancer [110]. However, this is not supported by the transcriptomics data analyzed here. It is known that SAs conduct their functions through binding with siglecs and/or selectins. Our analyses of the gene expression data in bulk cancer tissues show that there is no or little correlation between *CMAS* and any siglec genes (*SIGLEC1*–*SIGLEC16*), selectin genes (*SELE*, *SELL*, and *SELP*) or their total expressions.

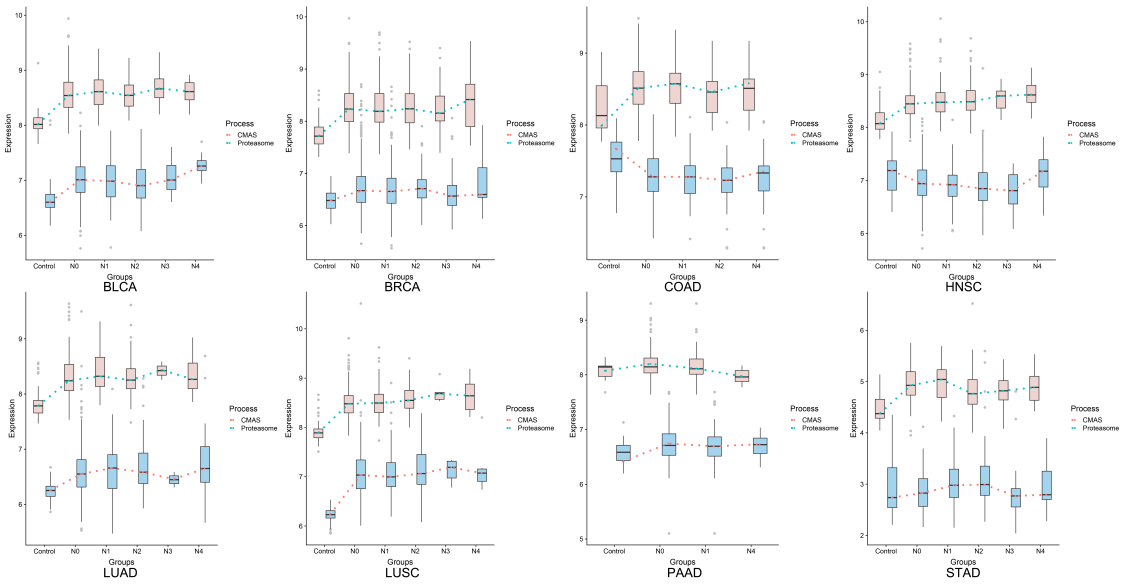


Figure 3.2: Predicted levels of cytosolic Fenton reaction levels vs. expression levels of *CMAS*. The level of Fenton reaction (orange) is estimated based on the expressions of genes related to macromolecular damages such as S20 proteasome genes, as given in Sun et al. [20].

We have previously predicted that 44 known reprogrammed metabolisms in cancer [47], including persistent production of SA, are induced to generate protons for neutralizing OH^- constantly generated by cytosolic Fenton reaction. Figure 3.2 shows the predicted levels of cytosolic Fenton reactions [20] vs. the expression level of *CMAS* across different stages of all cancer types under study. The majority of the cancers show statistically significant positive correlations, including BRCA (PCC[4] = 0.352, p-value = 3.05E-33), COAD (PCC = 0.216, p-value = 2.32E-04), LUAD (PCC = 0.343, p-value = 1.04E-15), LUSC (PCC= 0.206, p-value = 3.45E-06), STAD (PCC = 0.212, p-value = 9.96E-04). While the detailed correlation between the two curves may not always be high due to contributions by other reprogrammed metabolisms to neutralize OH^- produced by the Fenton reaction, their overall trends are generally the same. Hence, we predict the SA synthesis is induced to respond to cytosolic Fenton reactions.

[4]The Pearson correlation coefficient is calculated by $r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$.

3.4.2 Accumulation of sialic acids on the cancer cell surface

It has been established that cancer cells accumulate SAs on the cell surface, and the level of such accumulation could be considerably higher than that of red blood cells [128]. Red blood cells are known to have high levels of cell-surface SAs that prevent aggregation through repulsive forces from the negative charges [129]. Knowing that the accumulation rate of SAs depends on the rates of SA synthesis, degradation, and transfer to cell-surface glycans, we used the expression levels of the following genes to estimate the rate of such accumulation. *CMAS* was used for SA synthesis, *NEU1* for its degradation, and sialyltransferase genes *ST3GAL1/2/5*, *ST6GALNAC4/5*, and *ST8SIA1/5* for SA transfer to glycans. The following formula was used to estimate the rate of the SA placement onto glycan:

$$\sqrt{E(CMAS) \times \sum_{Y} E(Y)},$$

where $E(\cdot)$ is for the expression level of a gene, and Y represents the seven SA transferase genes. Similarly, the following formula was used to estimate the rate of SA degradation:

$$\sqrt{E(NEU1) \times E(CTSA)},$$

where *CTSA*-encoded Cathepsin A forms a complex with NEU1 to conduct the degradation function. Figure 3.3 shows the comparative levels of these two quantities across different stages of all eight cancer types. An assumption used here is that for a gene X with expression level $E(X)$, the maximum reaction rate of the enzyme encoded by X is proportional to $K_{cat} \cdot E(X)$, with K_{cat} being the reaction rate constant catalyzed by enzyme X in the Michaelis-Menten formulation[5]. Hence the reaction rates of the SA placement and degradation should be linearly proportional to the two curves for each cancer in Figure 3.3. Knowing that CMAS and NEU1 have comparable K_{cat} values [130, 131], we predict that the the two curves reflect the relative reaction rates of the two enzymes.

[5]This is essentially equivalent to the assumption that (1) the expression level of a gene is linearly proportional to its protein concentration; and (2) the reactant concentration is higher than the reaction constant KM, a common assumption used when modeling human metabolisms based on Michaelis-Menten formulation.

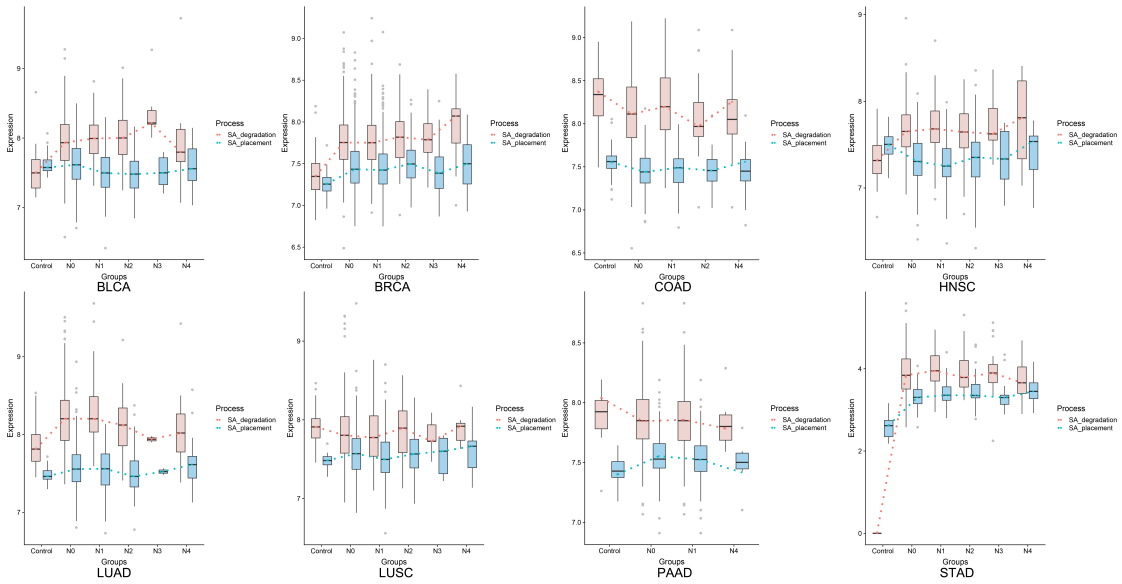


Figure 3.3: Expression levels for SA placement on cell surface (blue) and SA degradation (orange).

From Figure 3.3, we conclude that pancreatic cancer (PAAD) has by far the largest (positive) difference between the SA placement and degradation, measured using the total area between the curves of SA placement and SA degradation[6], hence giving rise to the highest level of SA accumulation and the strongest cell-cell repulsion, which is consistent with the known fact that the cancer has the highest death rate among all cancer types. More generally, cancer types with higher five-year survival rates tend to have higher SA degradation rates than their placement rates, especially towards the last stage of a cancer. Table 3.2 summarizes the average five-year survival rates of the eight cancers under study.

[6]The area is negative if the degradation curve is above the placement curve.

Table 3.2: Regression of five-year survival rates against SA-related gene expression data.

Cancer type	Survival rate	P_1	P_2
PAAD	9%	36.38011946	60.57841803
LUAD	15%	11.95926059	15.03685761
LUSC	16%	-10.21538364	4.967464172
STAD	31%	7.073508517	3.57755657
COAD	66%	1.118863508	-33.69555624
BLCA	77%	2.757848007	-23.66409384
HNSC	86%	-5.461726623	-8.198313869
BRCA	90%	-5.493377009	-0.420579177

To further demonstrate that the rates of the SA placement and degradation may have implications to a cancer's five-year survival rate, we conducted a linear regression analysis of the survival rates against changes in the rates of SA placement and degradation in the last two stages for each caner type. Specifically, let P_1 be the difference between the slopes of the SA placement curve and the slopes of SA degradation in the last two stages, and P_2 be the difference between the values of SA placement curve and degradation in the last stage of each cancer (Figure 3.3). Figure 3.4 shows a visualization of the values of these two parameters along with the corresponding survival rate for each cancer type. We can see that the survival rate can be well explained by these two parameters achieving an $R^2 = 0.876$ with p-value = 6.882E-03. LUSC was excluded since it does not fit our model, and the regression analysis was conducted only on the other seven cancer types. The fit function is given by:

$$\text{Survival} = 0.63 - 0.0725d(P_1) - 0.2279d(P_2),$$

where $d(P) = \text{sign}(P)\sqrt{|P|}$.

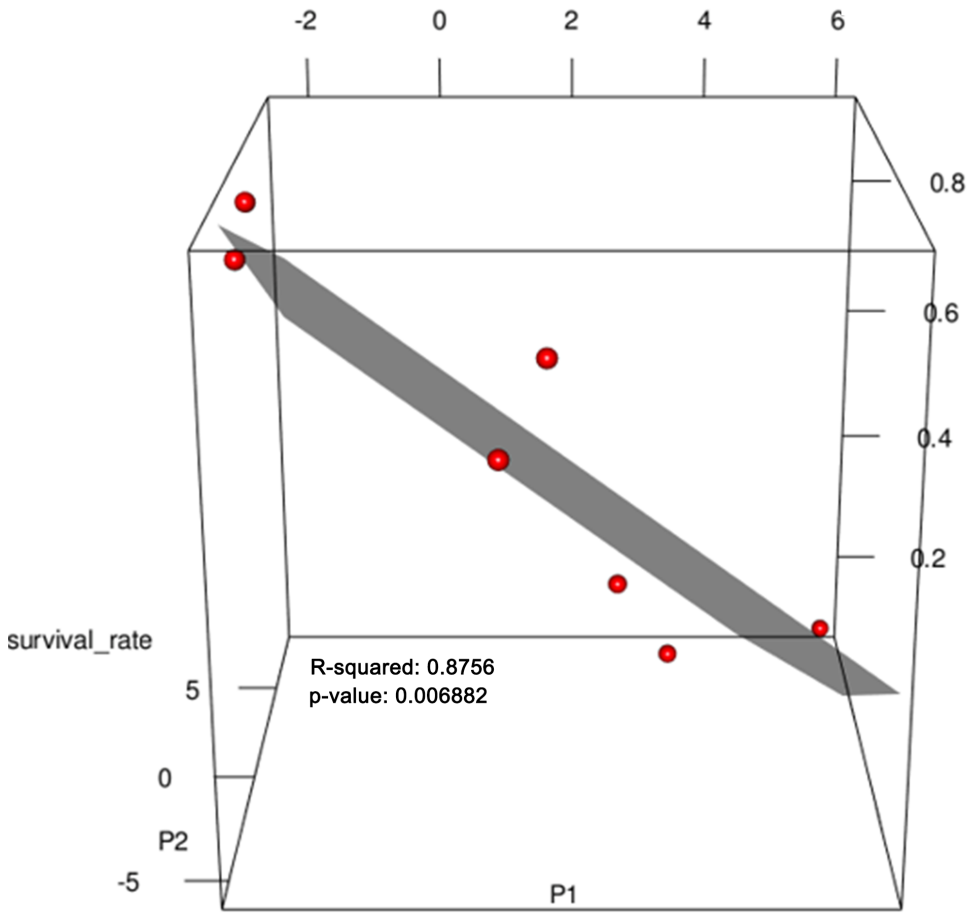


Figure 3.4: Regression of five-year survival rate against SA-related gene expression data. Each red dot represents one row in Table 3.2 excluding LUSC.

3.4.3 Ganglioside types as a cancer advances

Under physiological conditions, gangliosides as the hosts of SAs are predominantly used in brains and red blood cells. In the embryonic stage, brain cells tend to use gangliosides with simple glycan structures, and gradually switch to more complex structures, such as GM1, GD1a, GD1b, and GT1b [114], where GM is for monosialoganglioside, GD for disialoganglioside, and GT for trisialoganglioside.

It has been observed that cancers in advanced stages tend to use specific gangliosides such as GD2 and GM2, or GD3 and GM3 to a lesser extent [132]. The majority of the published studies focus on the possible signaling roles of such gangliosides like GD2 or GM2 [133, 134]. Other authors have examined

the issue from the perspective that different gangliosides enable different packing densities of gangliosides inside plasma membrane, Generally, the simpler a glycan structure is, the more gangliosides can be packed into a fixed area.

To understand these observations, we have examined the expression data of the enzyme genes in the synthesis network of gangliosides (Figure 3.5), where the "typical" relative expression levels of these genes are shown across the eight cancer types. We note that the synthesis pathways of numerous gangliosides do not quite follow the normal pathways as shown in Figure S2. Instead, they form distinct synthesis fluxes as shown in Figure 3.5 by the red or blue arrows with different widths. Specifically, the flux first goes downwards along the first column, and then travels to the second column via the ST6GALNAC4/5-catalyzed reactions more than the typical ST3GAL5 reactions. The flux then travels upwards along the second column; and from the second to the third column, the flux is relatively weak via the moderately expressed ST8SIA1.

We then examined the total expression level of the influx enzymes for each ganglioside vs. that of the efflux enzymes[7], and predicted that a ganglioside will have cellular accumulation if its influx rate is higher than its efflux rate, otherwise the ganglioside will not be accumulated. Note that such accumulation of a ganglioside should be proportional to the level of its deployment inside the plasma membrane. The reason for this simplified qualitative network flux analyses instead of the detailed kinetics-based Michaelis-Menten analyses is that we do not have the kinetic rate constants, K_{cat} and K_M, for multiple enzymes under consideration.

This analysis has revealed that advanced-stage cancer tissues generally have high accumulations of GM2 and GD2, followed by GM3 and GD3 (Table 3.3). These results are highly consistent with the published results.

[7] Influx meaning those that produce gangliosides, and efflux for those that consume the gangliosides.

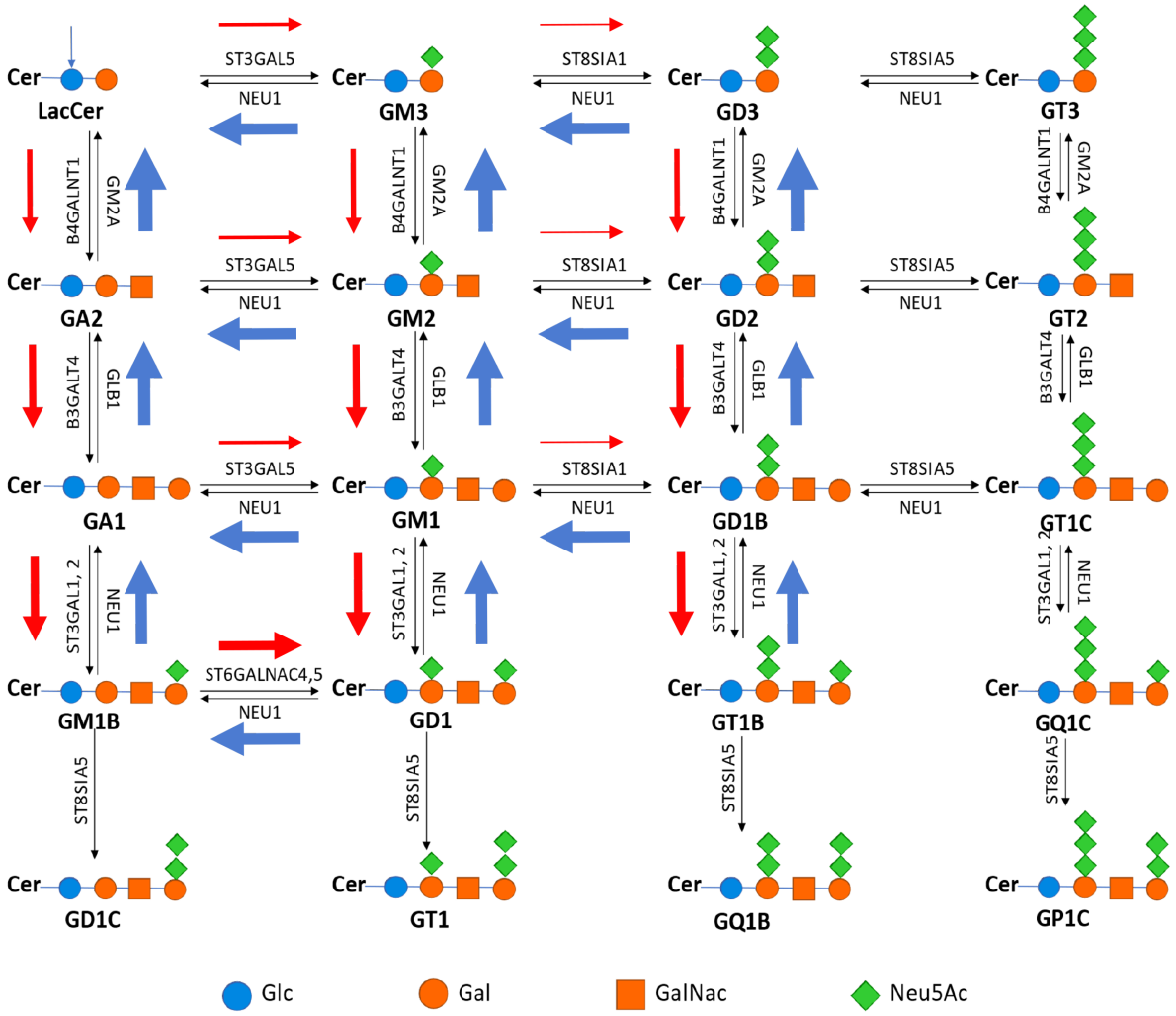


Figure 3.5: Metabolic pathway of ganglioside synthesis and metabolism. Names in bold are the names of the gangliosides above. Names next to each arrow represent the genes encoding the enzyme that catalyze the reaction represented by the arrow (adapted from Yu et al. [135]). The width of each arrow roughly represents the relative expression level of the corresponding gene, with red marking reactions that each produce one net proton and blue for pH neutral reactions. Reactions without such arrows are for those without expressions.

A key discovery made in our previous work on cancer metabolic reprogramming is cancer tends to produce as many protons as possible in each reprogrammed pathway. The altered ganglioside synthesis processes is one of the reprogrammed metabolisms studied [47]. Hence, we hypothesize that cancer may select specific ganglioside types that maximize the total number of protons produced per cell plasma membrane. We have examined the packing densities of GM2 and GM1. It has been reported that 451 GM2 molecules

Table 3.3: Estimated accumulation level of gangliosides across different stages of cancer metastasis.

Cancer type	Early stage (N_0)	Mid stages (N_1, N_2, N_3)	Late Stage (M)
BLCA	GM2 (high), GM3 and GD3 (moderate)	GM2 (high), GD3 (moderate)	GM2, GD2 (high), GM3, GM1 (moderate)
BRCA	GM2 (high), GD3, GD2, GD1 (moderate)	GM2 (high), GD3, GD2, GD1 (moderate)	GM2 (high), GD3, GD2 (moderate)
COAD	GM2, GD2 (high)	GM2, GD2 (high)	GM2, GD2 (high)
HNSC	GD3 (high), GM2, GD2 (moderate)	GD3 (high), GM2, GD2 (moderate)	GD3 (high), GM2, GM1, GD2 (moderate)
LUAD	GM1–3 (high), GD1–3 (moderate)	GM3, GM2, GM1 (high), GD3, GD2, GD1 (moderate)	GM3, GM2, GD1 (high), GM1, GD3 (moderate)
LUSC	GM2 (high), GD3, GD2 (moderate)	GM2 (high), GD3, GD2 (moderate)	GM2 (high), GM1, GD3, GD2 (moderate)
STAD	GM2, GD1 (high), GD2 (moderate)	GM2, GD1 (high), GD2 (moderate)	GM2, GD1 (high), GD2 (moderate)
PAAD	GM2, GD1 (high), GM1B, GD2 (moderate)	GM2, GD1 (high), GD2 (moderate)	GM2 (high), GD2, GD1 (moderate)

pack into a cluster with hydrodynamic radius 66.0Å, and 301 GM1 form a cluster with hydrodynamic radius 58.7Å [136], hence the ratio between the head groups per GM1 and GM2 is $(58.72/301) : (662/451)$, namely $11.45 : 9.66$. Therefore, the ratio between the numbers of GM1 and GM2 that can pack into a fixed area is $9.66 : 11.45$.

Note that the normal pathways for synthesizing GM1 and GM2 produce three and two net protons, respectively, but the altered pathways each produces four protons. In addition, the synthesis of each ganglioside requires the synthesis of some SAs, each of which produces two net protons. Figure 3.5 shows the number of SAs attached to each ganglioside. Hence for the same area in the plasma membrane, the ratio between the numbers of protons that the syntheses of GM1 and GM2 each produces is: $9.66 \times (4 + 2) : 11.45 \times (4 + 2)$, namely $9.66 : 11.45$. Therefore, more protons are produced by the synthesis and utilization of GM2 than that of GM1.

Considering that no published data regarding the packing densities in the same setting for the other gangliosides are publicly available, we extrapolate that GM3 and GM2 have a similar relationship to that between GM2 and GM1 presented above. Hence we predict that maximizing the proton production is a key determinant in a cancer's selection of utilization frequencies of GM3, GM2, and GM1 as observed above.

3.4.4 A sialic acid-based model for cancer metastatis

Under the TNM scheme, a cancer tissue is classified into stages N_0, N_1, N_2, N_3, or M, where N_i denotes the number of regional lymph nodes that contain cancer[8], and M represents distant metastasis. In the following, N_4 is an alias for M for the simplicity of presentation. Our goal is to demonstrate that for each cancer type, the average SA accumulation level in stage $N_i, 0 \leq i \leq 4$, can be calculated as $\mathbb{C}_i + C_i$, where $\mathbb{C}_i$ denotes the SA level accumulated solely in stage N_i, and C_i is a fixed positive value denoting the SA level accumulated in the previous $i-1$ stages for $i \geq 1$. By definition, we have $C_0 = 0$. Furthermore, $\mathbb{C}_i$ is a function of the rates of SA synthesis, degradation, transfer to cell-surface glycans, and the duration of stage N_i. It can be modeled based on the expression data of seven genes: *CMAS* for SA synthesis, *NEU1* for SA degradation, *ST3GAL1, 2, 5* and *B4GALNT1* for SA transfer onto cell-surface glycans, and *POLDIP2*[9] for the rate of cell cycle. Mathematically, this problem can be formulated as: for each cancer type, search for an unknown function $\mathcal{F} : \mathbb{R}^7 \to \{\mathbb{N}_i\}$ and an unknown fixed positive value C_i such that

$$\begin{aligned}\arg\min_{\mathcal{F},C_i} \|\mathbb{N}_i - \mathcal{F}(&E(CMAS), E(NEU1), E(ST3GAL1),\\ &E(ST3GAL2), E(ST3GAL5), E(B4GALNT1),\\ &E(POLDIP2), C_i)\|\end{aligned}$$

[8] $0 \leq i \leq 3$, with larger i representing more lymph nodes that have cancer.

[9] *POLDIP2* encodes DNA Polymerase δ-interacting protein 2, and the inverse of its expression is proportional to the duration of one cell cycle.

over all samples in stage $\mathbb{N}_i$, $i = C, 0, 1, [2, 3], 4$, where $\mathbb{N}_C$ is for control samples, and $\mathbb{N}_{[2,3]}$ is the union of samples in N_2 and N_3 stages[10]. $E(X)$ is the expression value of gene X in a sample. $\mathbb{N}_i$ was set to the following values: $\mathbb{N}_c = 0, \mathbb{N}_0 = 0, \mathbb{N}_1 = 1, \mathbb{N}_{2,3} = 2, \mathbb{N}_4 = 3$.

We solved this as a multi-classification and as a regression problem using two separate neural networks, each with two-hidden layers. Figure 3.6 shows the two neural network architectures. Table 3.4 shows the prediction results for each stage $\mathbb{N}_i$ of each cancer type measured by the macro F1 score. Table 3.4 also summarizes the predicted C_i values for each cancer type, which increase monotonically over stage and are logically meaningful.

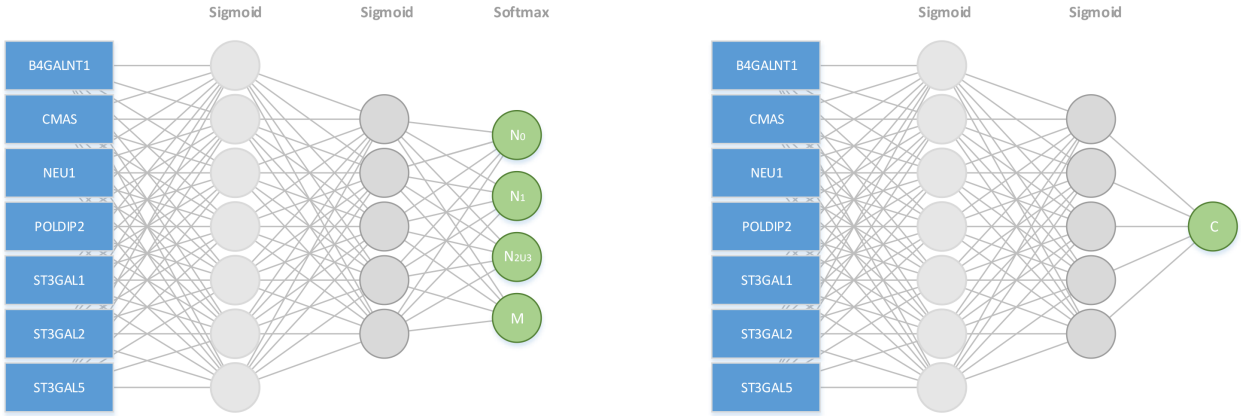


Figure 3.6: Architecture of the neural networks for predicting the stage of metastasis using SA synthesis, degradation and deployment genes *CMAS*, *NEU1*, *ST3GAL1, 2, 5*, *B4GALNT1*, and cell cycle-related gene *POLDIP2*.

From the table, we conclude that for each cancer stage, the seven genes can statistically well explain the metastasis stage, hence strongly suggesting that the SA accumulation level on cancer cell surface is a key factor in dictating the level of metastasis of a cancer tissue.

3.4.5 Supporting evidence

To provide further evidence that SA accumulation on the cancer cell surface can strongly influence the level of metastasis of a cancer, we examined the statistical relationships between the above seven genes

[10]The two stages were combined because each sample size tends to be too small.

Table 3.4: Parameters and metrics of metastasis stage prediction models. The macao F_1 scores are for the multi-classification model. Regression coefficients are estimations of the SA accumulation level prior to stage $\mathbb{N}_i$.

Cancer type	Macro average F_1 score	C_0	C_1	$C_{[2,3]}$	C_4
BLCA	0.744	16.340	18.090	26.847	58.350
BRCA	0.555	18.862	21.054	25.988	54.064
COAD	0.765	13.218	22.007	33.283	51.469
HNSC	0.649	16.701	19.715	23.493	59.880
LUAD	0.638	14.238	19.516	32.094	54.725
LUSC	0.739	14.042	19.488	26.718	59.526
PAAD	0.794	15.532	24.576	NA	60.030
STAD	0.772	11.895	17.719	30.342	59.992

and a range of characteristics uniquely associated with cancer migration, each of which is reflected by a set of marker genes given in Table 3.5. For a given set of marker genes $\{g_1, \cdots, g_k\}$ over n samples, let $X = (x_1, \cdots, x_n)^T$ be the solution that minimizes the following function:

$$\left| \frac{1}{n} \times \begin{pmatrix} \sum_{1 \le i \le n} E_i(g_1) \\ \vdots \\ \sum_{1 \le i \le n} E_i(g_k) \end{pmatrix} - \begin{pmatrix} E_1(g_1) & E_2(g_1) & \cdots & E_n(g_1) \\ E_1(g_2) & E_2(g_2) & \cdots & E_n(g_2) \\ \vdots & \vdots & \ddots & \vdots \\ E_1(g_k) & E_2(g_k) & \cdots & E_n(g_k) \end{pmatrix} X \right|$$

where $E_i(g_j)$ is the expression level of gene g_j in sample i, and X represents the feature vector of $\{g_1, \cdots, g_k\}$ over n samples, which is used to calculate the co-expression level with the SA related genes. This problem is solved as a linear regression problem.

We have examined co-expression levels between the above SA-related genes and known marker genes of compressive stress. Figure 3.7A shows that compressive stress marker genes strongly correlate with the SA genes. We have then conducted a regression analysis of the marker genes against the SA genes, with regression results shown in Table 4. It has been known that cancer cells tend to alter their cell-cell adhesion genes as cancer cells start to migrate. The co-expression and regression results are given in Figure 3.7B and

Table 3.5: Marker genes for processes related to cancer cell migration.

Migration-related process	Marker genes
Mechanical stress	*CASP3, DSG1, DSG2, DSG3*
Cell-cell adhesion	*CDH11, CDH13, CDH1, CDH2, CDH3, CDH5, CDH24, CTNNBL1, CTNND1, CTNNB1*
Cell contraction	*PTK2B, PDCD10, KCTD13, ITGB1, PHACTR1, CUL3, SRC, SRF, ARRB1, RHOA, SORBS1, TNFAIP1, ZYX, ITGB5*
Cell protrusion	*CENPB, RAB13, ZEB1, ANP32B, CORO1A, PINK1, DCC, VLDLR, FSCN1, TIAM1, MAP1B, LRP8, RELN, DCX, DCLK1, GAP43, FEZ1, CXCR4, DBN1, CTTN, ARP2, ARP3, CFL1, CFL2, LIMK1, LIMK2, WASF1, WASF2*
Cell motion and migration	*AMOT, FGF2, GPLD1, GPR124, KDR, NRP1, PTK2B, SCARB1, TDGF1, VEGFA*

Table 3.6. Similarly, results for cell contraction, cell protrusion and cell motion and migration are shown in Figure 3.7C–E and Table 3.6. From these figures and tables, we can see that each of the key migration-related characteristics can be well explained statistically by the SA genes, hence providing further evidence that SA accumulation plays key roles in driving cancer migration.

Table 3.6: Regression results of migration-related characteristics against SA-related genes.

Process	BLCA	BRCA	COAD	HNSC	LUAD	LUSC	PAAD	STAD
Cell-cell adhesion	0.562	0.495	0.797	0.754	0.581	0.529	0.578	0.299
Polarity	0.464	0.261	0.667	0.425	0.192	0.368	0.358	0.271
Mechanical stress	0.508	0.365	0.199	0.173	0.368	0.367	0.437	0.471
Protrusion	0.730	0.697	0.792	0.775	0.666	0.532	0.638	0.566
Cell contraction	0.555	0.513	0.720	0.631	0.444	0.705	0.623	0.412
Motion and migration	0.553	0.557	0.810	0.759	0.464	0.646	0.536	0.667
p-values of the calculated R^2 values								
Cell-cell adhesion	6.28E-33	7.62E-66	4.51E-61	1.60E-93	2.38E-45	4.96E-35	2.33E-15	3.33E-05
Polarity	4.02E-21	1.05E-16	1.50E-35	3.99E-22	1.18E-04	1.08E-15	1.39E-05	2.20E-04
Mechanical stress	5.63E-26	1.34E-33	4.12E-03	5.82E-04	3.76E-16	1.23E-15	2.96E-08	7.56E-13
Protrusion	2.72E-66	2.59E-156	1.25E-59	6.75E-102	1.68E-64	1.30E-35	1.24E-19	1.68E-19
Cell contraction	5.93E-32	1.27E-71	3.18E-44	2.92E-56	3.33E-24	1.29E-73	1.73E-18	1.15E-09
Motion and migration	9.10E-32	2.14E-87	2.21E-64	3.26E-95	1.05E-26	7.82E-58	6.81E-13	1.41E-29

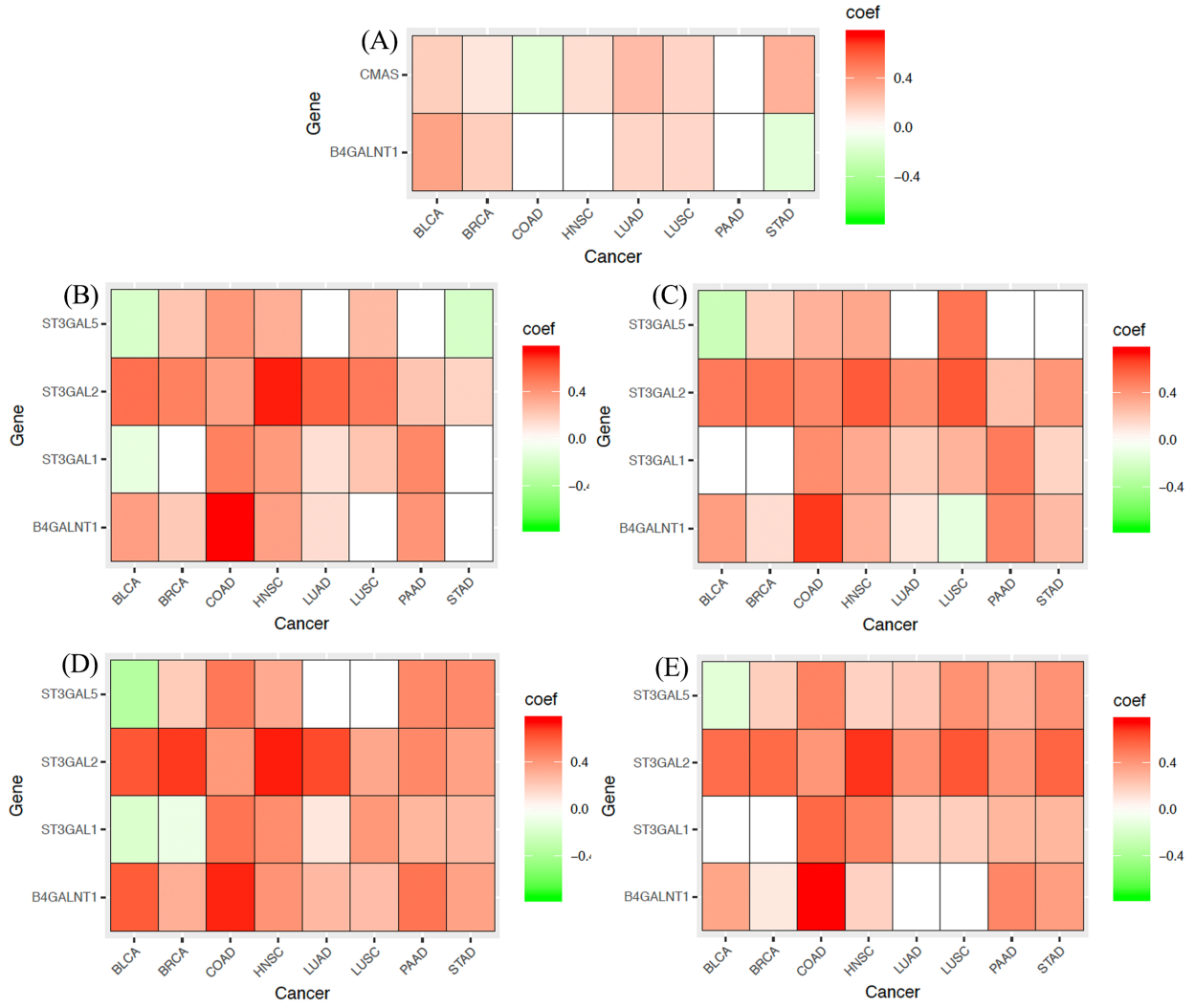


Figure 3.7: Co-expressions between SA-related genes and marker genes for multiple aspects of cancer metastasis across eight cancer types. Red, green, and white represent positive, negative, and insignificant correlations, respectively. **(A)** Mechanical compression genes. **(B)** Cell-cell adhesion genes. **(C)** Cell contraction genes. **(D)** Cell protrusion genes. **(E)** Cell motion/migration genes.

Regulation of key genes leading to cancer migration

We have conducted an integrated analysis of genomics, epigenomics, and transcriptomics data to estimate how the key SA genes, namely *CMAS*, *ST3GAL1*, *ST3GAL5*, and *NEU1*, are regulated across different cancer types. For each gene, we assess the level of contribution to its transcription regulation by transcription factors vs. DNA methylation using a regression analysis.

Table 3.7 shows the detailed regression results for all four genes. Subgroups with fewer than 20 samples were dropped and marked with "-". The percentage shown in columns DNAm and TF are relative sum of squares. From the table, we observe more DNA methylation utilization in cancer samples than in normal samples. While the level of contribution to transcription regulation by DNA methylation varies across different genes and different cancer types, DNA methylation in general makes substantial contributions to the regulation of the key SA genes.

Table 3.7: Regression results for the regulation of key SA genes by transcription factors (TF) and DNA methylation (DNAm).

Project	Stage	NEU1			ST3GAL1			ST3GAL5			CMAS		
		R^2	DNAm	TF	R^2	DNAm	TF	R^2	DNAm	TF	R^2	DNAm	TF
BLCA	Normal	0.98	0.0%	100.0%	0.66	0.0%	100.0%	1.00	79.6%	20.4%	-	-	-
BLCA	Stage II	0.51	12.7%	87.3%	0.77	48.0%	52.0%	0.89	46.3%	53.7%	0.85	10.7%	89.3%
BLCA	Stage III	0.72	44.5%	55.5%	0.70	51.0%	49.0%	0.72	39.2%	60.8%	0.75	32.6%	67.4%
BLCA	Stage IV	0.92	33.9%	66.1%	0.88	52.1%	47.9%	0.65	33.8%	66.2%	0.79	36.5%	63.5%
BRCA	Normal	0.80	0.0%	100.0%	0.96	28.9%	71.1%	0.71	0.0%	100.0%	0.92	1.2%	98.8%
BRCA	Stage I	0.76	48.1%	51.9%	0.83	74.4%	25.6%	0.64	13.5%	86.5%	0.47	0.0%	100.0%
BRCA	Stage II	0.48	21.0%	79.0%	0.63	62.3%	37.7%	0.40	40.1%	59.9%	0.60	37.4%	62.6%
BRCA	Stage III	0.37	44.8%	55.2%	0.70	61.0%	39.0%	0.63	19.9%	80.1%	0.68	13.3%	86.7%
COAD	Normal	0.91	47.6%	52.4%	0.76	0.0%	100.0%	1.00	14.1%	85.9%	0.96	21.8%	78.2%
COAD	Stage I	0.87	58.1%	41.9%	0.72	86.8%	13.2%	0.88	13.9%	86.1%	0.78	49.2%	50.8%
COAD	Stage II	0.69	52.9%	47.1%	0.73	64.4%	35.6%	0.77	30.5%	69.5%	0.82	21.4%	78.6%
COAD	Stage III	0.95	32.9%	67.1%	0.78	48.8%	51.2%	0.69	26.4%	73.6%	0.60	28.7%	71.3%
COAD	Stage IV	0.98	38.0%	62.0%	0.37	100.0%	0.0%	0.85	26.0%	74.0%	-	-	-
HNSC	Normal	1.00	53.1%	46.9%	0.96	89.6%	10.4%	0.96	46.4%	53.6%	0.84	24.9%	75.1%
HNSC	Stage I	1.00	85.9%	14.1%	0.88	68.1%	31.9%	0.96	30.8%	69.2%	0.24	100.0%	0.0%
HNSC	Stage II	0.43	57.9%	42.1%	0.66	49.2%	50.8%	0.91	39.4%	60.6%	0.84	32.4%	67.6%
HNSC	Stage III	0.84	80.3%	19.7%	0.73	37.9%	62.1%	0.84	30.7%	69.3%	0.27	100.0%	0.0%
HNSC	Stage IV	0.73	15.2%	84.8%	0.83	57.9%	42.1%	0.58	26.5%	73.5%	0.83	33.1%	66.9%
LUAD	Normal	1.00	61.5%	38.5%	1.00	43.1%	56.9%	1.00	13.7%	86.3%	0.84	0.0%	100.0%
LUAD	Stage I	0.82	42.2%	57.8%	0.76	61.5%	38.5%	0.68	33.7%	66.3%	0.67	18.5%	81.5%
LUAD	Stage II	0.88	38.4%	61.6%	0.57	41.7%	58.3%	0.67	34.9%	65.1%	0.82	15.7%	84.3%
LUAD	Stage III	0.75	82.2%	17.8%	0.90	67.8%	32.2%	0.84	34.8%	65.2%	1.00	25.3%	74.7%
LUAD	Stage IV	-	-	-	1.00	73.0%	27.0%	0.97	84.1%	15.9%	0.49	0.0%	100.0%
LUSC	Normal	-	-	-	0.95	0.0%	100.0%	1.00	0.0%	100.0%	1.00	0.0%	100.0%
LUSC	Stage I	0.84	37.5%	62.5%	0.75	59.5%	40.5%	0.79	27.6%	72.4%	0.86	13.8%	86.2%
LUSC	Stage II	0.26	37.8%	62.2%	0.73	38.3%	61.7%	0.54	59.5%	40.5%	0.52	30.1%	69.9%
LUSC	Stage III	0.64	65.4%	34.6%	0.62	73.3%	26.7%	0.84	41.3%	58.7%	-	-	-
PAAD	Normal	0.98	0.0%	100.0%	1.00	100.0%	0.0%	1.00	0.0%	100.0%	1.00	0.0%	100.0%
PAAD	Stage I	0.78	83.8%	16.2%	0.98	75.1%	24.9%	-	-	-	0.96	3.2%	96.8%
PAAD	Stage II	0.73	48.1%	51.9%	0.78	36.4%	63.6%	0.61	32.6%	67.4%	0.79	8.3%	91.7%
STAD	Stage I	0.44	0.0%	100.0%	0.57	59.1%	40.9%	0.79	34.6%	65.4%	0.47	78.0%	22.0%
STAD	Stage II	0.92	82.8%	17.2%	0.58	49.5%	50.5%	0.86	38.4%	61.6%	0.86	9.6%	90.4%
STAD	Stage III	0.90	58.3%	41.7%	0.54	81.7%	18.3%	0.77	9.2%	90.8%	0.76	62.9%	37.1%
STAD	Stage IV	1.00	78.0%	22.0%	0.84	69.4%	30.6%	0.98	51.3%	48.7%	0.99	13.2%	86.8%

3.5 Discussion

Our prediction for linking overproduction and gradual accumulation of sialic acids to cancer migration is based on identified statistical or physical connections among seemingly unrelated molecular species and cellular activities. It is the multiplicity of such chains of simple and subtle associations, which are otherwise nonexistant, that give us the statistical confidence about our prediction. Some of the detected associations are only useful for making our overall prediction while others, such as the seven-gene signature for predicting the status of cancer metastasis, could be used independently as novel markers for cancer metastasis. In addition, our prediction confidence also comes from the previous work that causally connects mechanical forces to cancer migration.

Tse et al. [122] presented an elegant study that shows external mechanical compression on cancer cells can lead to their deformation, which can give rise to enhanced cell-cell adhesion, actomyosin contraction, filopodial protrusion, and ultimately collective migration by clusters of cells. This model provides strong support to our prediction that SA accumulation on cancer cell surface will generate increasingly stronger electrostatic repulsion due to the increased densities of electric charges from SAs, leading to enhanced cell-cell adhesion, actomyosin contraction, protrusion as well as migration as established above. Compared to previous studies on SAs and cancer migration, our study provides a fundamentally novel perspective regarding the roles of SA in cancer migration, namely it is their physical property rather than the signaling functions that may play the predominant role in driving cancer metastasis.

This model can answer a few long standing open questions regarding cancer metastasis. First, while SAs have long been known to be associated with cancer metastasis, very little has been established regarding why it generally takes years for a cancer to become metastatic, knowing that the expression levels of SA synthesis and transferase genes do not go up very substantially (Figure 3.1). Our model, in conjunction with Tse et al.'s model, suggests that the gradual accumulation of the SA-associated negative charges on the cell surface will activate the migration program as discussed in Tse et al. [122] once their electrostatic repulsion reaches a critical point.

Second, very little has been established in the literature regarding why certain cancer types tend to metastasize during early stages, while other cancer types are less likely to metastasize. Our model suggests that it is the combination of the rates of SA production, degradation and transfer to cell surface glycan that determines when a cancer starts to migrate.

Third, gangliosides of certain types such as GM2 and GD2 have long been found to be associated with cancer metastasis and the current literature suggests that it is their signaling roles that may be relevant to migration [114]. Our model suggests that two factors may contribute to the selection of specific types of gangliosides: (1) the number of SAs that can be put on gangliosides per cell, which is determined by the packing density of individual ganglioside types inside the plasma membrane and the number of SAs that each ganglioside of the type can harbor; and (2) the number of protons produced by the synthesis process of individual ganglioside types. More complex gangliosides, although producing more protons per molecule through their synthesis process, have lower packing densities inside the plasma membrane, hence possibly resulting in a lower number of total protons per cell. We postulate that the selected ganglioside types are the result of tradeoff between these two factors, which ultimately enables the maximum number of protons to be produced through this combination.

Our model is clearly a statistical model based on omics data. We plan to develop a more physics-oriented model that allows us to estimate the actual density level changes as a cancer evolves, as well as to calculate the level of electrostatic repulsion in a realistic media environment, which would enable the accurate estimation of the shield effect of the electrons.

Chapter 4

Metabolike: A Unified Resource for Studying Reprogrammed Metabolic Pathways

Yi Zhou and Ying Xu. To be submitted to *Bioinformatics*.

4.1 Abstract

Metabolic models are often distributed in plain text formats, hindering efficient exploration of the network structures. To the best of our knowledge, there does not exist a tool capable of representing metabolic models as graphs and integrating omics data for downstream analyses. Metabolike fills in this gap and presents a novel method of identifying reprogrammed metabolisms under diseased cellular states such as cancer. We have developed a Python package named Metabolike that can transform metabolic models into knowledge graphs. The suite of graph traversal functions and route scoring algorithms enable the identification of context-specific metabolic reaction routes with the integration of omics data. Effectiveness of the package was validated on two established reprogrammed metabolic pathways in cancer: sialic acid synthesis and urea cycle rewiring.

4.2 Introduction

Curation of biological reactions is one of the most cumbersome and time-consuming tasks in systems biology, as it requires field experts to carefully extract the information from scientific literature and transform it into a consistent ontology. Multiple research teams have dedicated decades to construct and annotate metabolic pathways, resulting in a series of high-quality databases such as KEGG [56], BRENDA [57], and MetaCyc [58]. These databases focus on different aspects of the metabolic network. For example, BRENDA aims to be the "comprehensive enzyme information system", and holds more data on enzyme kinetic parameters and molecular properties than its peers. KEGG, on the other hand, is a slowly-evolving database that covers molecular interactions including metabolic pathways, transcription regulation, signal transduction, and various other cellular processes. The web interfaces of these databases allow users to easily query specific biochemical reactions or pathways.

With the rapid growth of annotated biochemical reactions, studying complex metabolic networks via direct inspection has become intractable. Systems biology emerged as a field in which computational and mathematical techniques can be used to study large-scale biological systems [137]. In the specific

case of metabolic networks, the gene-reaction-metabolite connections were developed into genome-scale metabolic models (GEMs) to facilitate metabolic reaction representation and simulation [138, 139]. The non-trivial task of constructing a GEM involves integrating information from multiple sources with inconsistent ontologies, nonstandard identifiers, and duplicated components. Human-GEM, a fairly recent model, is deposited in a public repository and accepts external contributions [139], thus allowing the model to be continuously refined by researchers worldwide. However, the size and complexity of current GEMs hinder the ad hoc exploration and analysis of the models [59]. Models are often disseminated in plain text formats[1], and the lack of a graphical interface makes it difficult for novices to understand the relationships between entities in the model.

Among the data formats, an XML-based specification for describing biochemical networks named the Systems Biology Markup Language (SBML) [62], has gradually become the *de facto* standard for exchanging models between software tools. Packages developed around the core SBML specification allows the integration of additional information into the model, such as gene-reaction associations, flux boundaries on reactions, and groups of related entities [141, 142]. However, as a format designed for modeling, SBML lacks support for visual inspections and structured queries, both of which are essential for exploratory analysis of metabolic networks. It would be useful at this stage to develop a tool that transforms SBML models into a more structured format and deposit them into a database. The tool should also provide interfaces for data integration and visual inspection to drive deeper insight into the network structure.

We developed a Python package named Metabolike that could transform SBML models into a comprehensive knowledge graph, and generate context-specific metabolic reaction routes given omics data[2]. The tool aligns entities collected from various databases via pattern matching techniques and refactors the information into a Neo4j graph database. The graph representation allows integration of highly flexible data formats and enables efficient route traversal and visualization, as demonstrated in the Results section.

[1] To make things worse, there are *many* competing standards for systems biology models [140]. Tools exist to convert models between these formats, but they are not officially supported and are often not up-to-date with the latest specifications.

[2] Genomics and transcriptomics data are our primary focus, but it should be fairly easy to extrapolate to other data types.

Table 4.1: Number of pathways, reactions, compounds, and gene products in the models.

Database	Pathways	Reactions	Compounds	Gene products
MetaCyc	3338	28175	25123	12604
HumanCyc	389	3067	2934	2989
MouseCyc	293	1593	1513	1740
Human-GEM	142	13069	8366	3067

The source code is publicly available on GitHub at https://github.com/y1zhou/metabolike, and the Python package can also be installed from PyPI on any system with Python 3.9 or newer. We expect the tool to facilitate the understanding of reprogrammed metabolisms in diseased cellular states such as cancer, and assist the identification of potential drug targets.

4.3 Materials and Methods

4.3.1 Data collection

SBML models MetaCyc is a curated database of metabolism that started in 1999 [143]. As of 2022, MetaCyc contains more than 18,000 biochemical reactions and over 3000 metabolic pathways. Entities in MetaCyc are extensively linked to other gene, protein, compound, and reaction databases, making MetaCyc identifiers an ideal intermediate layer for integration of heterogeneous data. Pathway databases including MetaCyc, HumanCyc, and MouseCyc were downloaded from the BioCyc database [143], version 26.0 at https://biocyc.org/download.shtml. Access to the HumanCyc and MouseCyc database files required an active subscription or a license[3], and both databases were obtained for free for research purposes. Summary statistics of the databases are shown in Table 4.1. Numbers for the MetaCyc database are a magnitude higher than those for the other databases, as it is a compilation of pathways from 3295 different organisms.

[3]The EcoCyc *Escherichia coli* database and the MetaCyc metabolic pathway database are free. MetaCyc is rather comprehensive, but misses some organism-specific reactions or pathways so we also included the proprietary HumanCyc and MouseCyc databases.

The XML format of the Human-GEM model [139] was downloaded from their GitHub release page at https://github.com/SysBioChalmers/Human-GEM. Without external data files assigning the reactions into different pathways, the "group" extension from the SBML model was extracted and used as a proxy for canonical pathways in the system.

All data in the BRENDA database are available for download in a text file upon accepting their license agreement. The February 2022 release of the file contains a total of 7608 enzymes identified by EC numbers. For each enzyme, BRENDA collected general information such as its systematic name and molecular weight, as well as organism-specific or even environment-specific information such as the K_{cat}, K_M, and optimal pH and temperature ranges. All fields were extracted into a tree-like format using the parser module provided in Metabolike.

Transcriptomics data Transcriptomics data for 14 cancer types were retrieved from the TCGA database [121], and differential expression analysis between cancer and cancer-adjacent control samples was performed using DESeq2 [144].

4.3.2 Methods

Metabolic network model parser In the core specification, an SBML model consists of lists of compartments, species, reactions, and gene products. Compartments are subcellular localizations in which the species and reactions are located or performed. Species are chemical compounds that participate in reactions defined in the model. Reactions are biochemical processes that transform one set of species into another, often with assistance from gene products. Gene products are proteins that catalyze the reactions. Sometimes, multiple proteins together form a complex to catalyze a reactions. The relationship is also not necessarily one-to-one, as some proteins/protein complexes are used in multiple reactions. The `libSBML` library [145] was used extensively for parsing SBML models. The library provides a Python interface to read, write, and manipulate SBML files, which allowed us to extract information about the constituting entities. A schema of all relevant properties of the extracted entities is given in Table 4.2. It should be noted that during our preprocessing step, we moved all reactants to the left hand side and all products to

the right hand side. This normalization procedure is related to how the compound entities are connected to the reaction and to how downstream files are parsed.

Table 4.2: Relevant properties of entities extracted from SBML models.

Property key	Property type	Description
metaId	String	Primary key matching the "metaId" in the SBML model.
name	String	Name of the entity.
comment	String	Other commentary information related to the entity.
synonyms	List[string]	Synonymous names for the entity.
canonicalId	String	Prefix of the metaId with substrates or enzymes removed.
reversible	Boolean	If the reaction is reversible.
reactionDirection	String	Direction of a reaction.
systematicName	String	Systematic name of a reaction.
reactionBalanceStatus	Boolean	If the stoichiometric coefficients are balanced in the reaction.
gibbso	Numeric	Standard Gibbs free energy change of the reaction.
smilesAtomMapping	String	SMILES representation of the reaction.
types	List[string]	Biochemical classification of a reaction or metabolic pathway.
commonName	String	Common name of the compound or pathway.
constant	Boolean	If the compound can be seen as a constant in a reaction.
chemicalFormula	String	Chemical formula of the compound.
smiles	String	SMILES representation of the compound.
monoisotopicMw	Numeric	Monoisotopic molecular weight of the compound.
inchi	String	InChI identifier of the compound.
gibbso	Numeric	Standard Gibbs free energy of formation of the compound.
molecularWeight	Numeric	Molecular weight of the compound.
ncbi	String	NCBI gene ID [146] of the annotated gene product.
keggCompound	String	KEGG compound ID [56].
chebi	String	ChEBI [147] molecular entity ID.
pubchemCompound	String	PubChem [148] chemical compound ID.
uniprot	String	UniProt protein ID [149] of the annotated gene product.

There are also cases where the SBML framework is too limiting for the description of a metabolic network. For example, with the core specification, there is no way to group reactions into metabolic pathways. This may not be a problem for metabolic flux modeling, but canonical pathways are immensely useful for a high-level understanding of the overall structure of a metabolic network. In cases like this, resource providers either utilize some existing plugin for the SBML framework or generate peripheral data files that are linked to the SBML file with shared identifiers. One of the most widely adopted SBML

plugin resolves gene products[4] that catalyze metabolic reactions. Gene products require special treatments because sometimes multiple proteins form a complex to catalyze a reaction, and sometimes a reaction can be catalyzed when any member in a set of gene products is present. These relationships are described using "And" and "Or" groups in the extended SBML specification, which allows for arbitrarily nested nodes and is useful for representing complex enzyme regulations. In our graph model, the groups were termed "GeneProductComplex" and "GeneProductSet" nodes, respectively. Only a single "GeneProduct", "GeneProductComplex", or "GeneProductSet" node can be connected to a specific reaction, thereby allowing constrained queries to the database.

For more complex models such as the BioCyc family databases, we developed custom scripts to parse the data. BioCyc offers a suite of data files that describe the entities in the SBML model in greater detail. Each file begins with a header section of metadata and the layout of the available attributes. The remainder of the file are entries separated by the "///" marker, with each entry being a list of attribute-value pairs. Metabolike supports parsing the following data files:

1. Reactions.dat containing nomenclature, literature references, associated EC numbers, reversibility, stoichiometry of reactants and products, associated gene products, and the standard Gibbs free energy change ($\Delta_r G^\circ$) of each reaction. $\Delta_r G^\circ$ is also used to canonicalize reaction directions, namely we always arrange compounds such that reversible reactions have a negative $\Delta_r G^\circ$.

2. Atom-mapping-smiles.dat containing the SMILES[5] string representation for the reactions, which is a linear notation describing the structures of the species involved in the reactions.

3. Compounds.dat containing associated compartments, chemical formulas, and various external database identifiers that are not present in the SBML model such as KEGG [56], ChEBI [147], and PubChem [148]. Linking to external databases is crucial for compounds because unlike genes which have somewhat stable identifiers [150], names and identifiers of chemical compounds vary greatly across databases.

[4] Perhaps the original specification wanted to be more inclusive with the description. In my experience all of the "gene products" I have encountered are just enzymes.

[5] SMILES stands for simplified molecular-input line-entry system.

4. Pathways.dat containing the identifiers, literature references, reaction layouts, and primary reactants and products for the manually curated metabolic pathways.

5. Pubs.dat containing the literature references for the compounds, reactions, and pathways.

6. Classes.dat containing higher-order entities defined in the BioCyc database, such as the common names for compartments, organisms, and the cryptic identifiers scattered around the other supporting documents.

To achieve idempotence, an iterative approach was taken to parse each data file that may have side effects. Taking the pathways data file as an example, the parsing process may introduce new (super-)pathway nodes, thus a single round of parsing is not sufficient to annotate all pathways. A queue was therefore implemented to store all unannotated pathways, and a pathway is added to the queue if it is not already present in the graph database. The pathways file was parsed in a loop until the queue is empty. It is ensured that all entities are resolved and annotated without duplication or ambiguity.

BRENDA text file parsing The BRENDA text file is a semi-structured file similar to the `dat` files described above with a few differences. The biggest discrepancy is that attribute-value pairs no longer live in a single line. In BRENDA, each field starts with a specific acronym prefix followed by a tab character, and could span multiple lines before the next prefix occurs. There are also fields wrapped in specific characters. For example, protein references are wrapped in "#" characters, literature citations are in "<>", commentary information are in "()", and field-specific information are in "{}". The scripts developed for the BioCyc files were not sufficient to parse a file with this complexity. Therefore, the Lark toolkit was used to develop a set of grammars that could convert the plain text into a parse-tree. We used the LALR(1) parser[6] for better performance. The extracted tree structure was mappable to the graph through shared identifiers such as the EC numbers.

[6]The LALR(1) parser comes with more constraints and fewer features than the default Earley parser, but it is much more efficient.

Graph database interface The Neo4j graph database was employed as the storage backend for the parsed metabolic network. Compared to relational databases, Neo4j has a more intuitive query syntax and offers significant performance improvement, as has been shown in the transition from a relational (MySQL) to a graph (Neo4j) database for the Reactome database [61]. Details for set-ting up the database can be found at the documentation site of Metabolike at https://metabolike.readthedocs.io/en/latest/usage/getting_started/.

Metabolic route searching Given a metabolite, we want to retrieve all possible routes that may be responsible for its biosynthesis, transportation, and consumption or degradation. Ideally, the other end of the route should be a "dead-end metabolite", meaning that the compound is not involved in any other reactions. The reaction routes that pass through highly connected intermediates like water or ATP and end with biologically irrelevant compounds should also be avoided.

Following these guidelines, the strategy of a route search is defined as follows. First, a blacklist[7] is created by retrieving compounds nodes with high degrees, meaning that compounds in the final routes cannot be connected to too many reaction nodes. The algorithm then start from the metabolite of interest, and traverse the graph in a breadth-first search, looking for repeating patterns of

$$\text{Compound} \xleftarrow{\text{hasLeft}} \text{Reaction} \xrightarrow{\text{hasRight}} \text{Compound},$$

and the branch is pruned if a blacklisted compound is encountered. To account for reversible reactions, transient nodes were added for all bidirectional reactions with their left- and right-hand sides flipped. This allows us to traverse the graph without modifying the algorithm above, and also implicitly models the actual direction of these reactions under the assumption that they will be consistent with surrounding reactions.

When the search is complete, the routes are scored based on the associated omics data metrics and the molecular similarity between the start and end metabolites. The current version of Metabolike focuses

[7] Here the blacklist only includes high-degree compounds, but it could be easily extended to also include mutated or lowly-expressed genes.

on transcriptomics data integration. By providing a table of gene identifiers and differential expression metrics, Metabolike can map the corresponding values onto the `GeneProduct` nodes in the graph. The minimum and maximum values are taken for `GeneProductComplex` and `GeneProductSet` nodes, respectively. Values are accumulated recursively until a Reaction node is met, in which case the score E is assigned to the `Reaction` node.

The molecular similarity between the start and end compounds is measured by the Tanimoto similarity. The Tanimoto similarity between the binary fingerprint of two molecules X and Y is defined as:

$$S_{X,Y} = \frac{a}{a+b+c}$$

where a denotes $x = 1, y = 1$, b denotes $x = 1, y = 0$, and c denotes $x = 0, y = 1$. Identical compounds would have a Tanimoto similarity of one, whereas compounds sharing no resemblance in their structures would have a Tanimoto similarity of zero. The Tanimoto similarity coefficient is a commonly used metric that has been shown to be robust in general [151]. It can be calculated from the chemical formulas or the SMILES strings of the two chemical compounds with the RDKit package.

The overall route score is defined as:

$$S = sign(E) \times (\alpha|E| + \beta S_{X,Y}),$$

where α and β are weight coefficients. Identified routes with high absolute values of S are more likely to be reprogrammed pathways, with positive values being active routes and negative values being suppressed routes.

An alternative search strategy is to only follow routes defined by the canonical pathways. The reaction layout section of the pathways describes the order and directionality of reactions, as well as the major substrates and products of each step. Reaction nodes belonging to different canonical pathways are also connected to avoid the need to hop through pathway nodes. This more stringent method outputs

metabolic routes that are easier to interpret, but it may fail to identify novel paths when metabolisms are heavily reprogrammed or disrupted.

4.4 Results

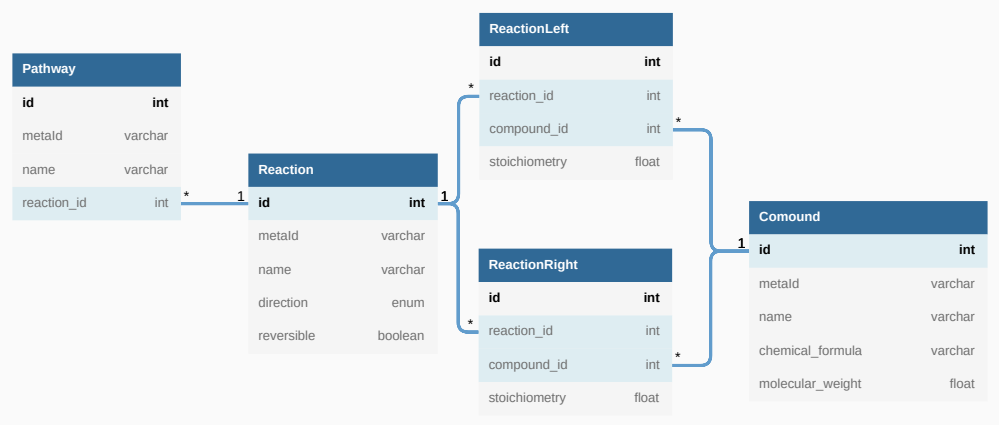
The Metabolike package was developed to transform text file representations of metabolic networks into a knowledge graph. This not only enables efficient query and visualization of the network, but also allows us to overlay omics data and automatically identify context-specific metabolic reaction routes. In the following sections, we demonstrate the performance of Metabolike on genome-scale networks and showcase its potential uses in studying cancer metabolism.

Choosing a suitable database software is an essential step that determines the stability and extensibility of downstream tasks. Various resources [57, 58] use relational databases to store information as a set of tables with some schema depicting the relationships between the entities. Here, the rationale for the choice of a graph database is described, along with demonstrations of its application to the metabolic reprogramming analysis of transcriptomics data sets.

4.4.1 Performant parser and graph database enable interactive queries

Using graph databases as the storage backend for metabolic networks has performance advantages over using relational databases. A metabolic network can be naturally represented as a directed graph consisting of interconnected pathway, reaction, metabolite, and enzyme nodes. Storing the network as a graph eliminates the need to specify multiple table schemas to accommodate for the complex structure. As a consequence of representing the graph structure in a tabular format, significant overhead is added when the relational database query involves multiple entities with many-to-many relationships. For example, a metabolic pathway would have multiple reactions associated with it, and each reaction would have substrates on its left-hand side and products on its right-hand side (Figure 4.1a). Simply querying for the reactions and compounds in a pathway involves five table joins, whereas the chain of nodes can be retrieved via a single query to the graph database (Figure 4.1b). The performance gap is further widened when

additional entities such as gene products and cellular compartments are included. Significant performance improvement has been shown in the transition from a relational (MySQL) to a graph (Neo4j) database for the Reactome [61] website.



(a) Schema of the relational database. 1 and ∗ indicate one-to-many and many-to-one relationships.

```sql
SELECT r.*, c.* FROM Reaction r
JOIN Pathway pw ON r.id = pw.reaction_id
JOIN ReactionInput ri ON r.id = ri.reaction_id
JOIN Compound c ON c.id = ri.compound_id
JOIN ReactionOutput ro ON r.id = ro.reaction_id
JOIN c ON c.id = ro.compound_id
WHERE pw.metaId = 'PWY66-398'

##########################################

MATCH (p:Pathway{metaId: 'PWY66-398'})⟶(r:Reaction)-[:hasLeft|hasRight]→(c:Compound)
RETURN r, c
```

(b) Sample SQL (top) and Cypher (bottom) queries.

Figure 4.1: Example of the pathway-reaction-metabolite graph in a relational database.

Another integral part of Metabolike is its highly optimized metabolic model parser. In short, the parser module converts a model into a tree-like structure, resolves duplicated and ambiguous entities, and maps relationships between valid entities. It should be noted that the parser and database interface modules are decoupled, thereby allowing users to easily implement parser extensions for custom datasets. The modules included in Metabolike are thoroughly tested to achieve idempotence, meaning that data integrity is preserved when the same operations are performed multiple times. We benchmarked Metabolike on SBML models acquired from HumanCyc [143] and Human-GEM [139] (see Materials and Methods for

details). With adequate thread pool management and batched transactions, transforming both models into graphs with over 50,000 nodes and 114,000 edges took less than 10 seconds. A command line tool is provided within Metabolike to simplify the configuration process and allows users to easily populate the graph database with an SBML model and peripheral data files.

The expressiveness of the graph database query language, Cypher, greatly simplifies *ad hoc* analyses of the metabolic network. For example, one can identify metabolic pathways containing a large number of reversible reactions with the following query:

```
MATCH (p:Pathway)-[:hasReaction]->(r:Reaction)
WHERE r.reactionDirection = "reversible"
RETURN p.metaId, p.commonName, COUNT(*) AS cnt
ORDER BY cnt DESC
LIMIT 3;
```

The query returns gluconeogenesis, the TCA cycle, and fatty acid β-oxidation as the top results. Another example usage of the database is to generate a stoichiometric matrix, which is a common task when preparing a model for metabolic flux analysis as networks built from scratch are often ill-defined and require missing intermediate metabolites to be identified. With subroutine support in the Cypher query language, it is straightforward to compute the overall stoichiometric changes in a pathway. The following query correctly returns the two molecules of ATP yield per glucose molecule:

```
MATCH (:Pathway {name: "glycolysis"})-[:hasReaction]->(r:Reaction)
CALL {
    WITH r
    MATCH (r)-[sub:hasLeft]->(cl:Compound)
    RETURN cl.name AS cpd, -1.0*SUM(sub.stoichiometry) AS c_change
    UNION
    WITH r
    MATCH (r)-[prod:hasRight]->(cr:Compound)
    RETURN cr.name AS cpd, SUM(prod.stoichiometry) AS c_change
}
WITH cpd, c_change
RETURN cpd, SUM(c_change) as c_changes
ORDER BY c_changes
```

In summary, Metabolike efficiently transforms metabolic network models from SBML files into an interconnected graph representation, which strikes a balance between the flexibility of plain text formats and the ease of use of programmable interfaces. Users can explore specific metabolic pathways in much finer detail, or generate summary statistics of all entities specified in the model with succinct code.

4.4.2 Integration of additional data into the graph

In certain cases, the core parser for SBML models is not sufficient to reconstruct the full metabolic network. The BioCyc family databases are good examples of this situation, where information about pathways, literature references, and various other metadata are provided in separate files. Projects like Recon2Neo4j [152] and biochem4j [153] have developed similar tools for parsing SBML models, but they lack the ability of integrating additional data sets. Metabolike was designed to support the integration of these data into the graph database through parser extensions. The graph database operation modules of Metabolike expose APIs to the parser modules, allowing users to plugin in custom parsing code designed for various data formats.

As a proof of concept, Figure 4.2 shows the schema of the MetaCyc graph augmented with metabolic pathway, literature reference, and taxonomy information. The main addition to the gene product-reaction-compound core graph are the "Pathway" and "Taxa" nodes. "Pathway" nodes group the thousands of reaction nodes into biologically meaningful clusters, and facilitate the identification of novel routes between metabolites. "Taxa" nodes partition the original graph and help identify organism-specific reactions. The extended parser also adds relationships between existing nodes in the graph. For example, it is possible to infer the order of reactions within a pathway, thus "isPrecedingEvent" edges are added between reactions.

4.4.3 Context-specific network analysis with integration of transcriptomics data

Applications of Metabolike have so far been focused on the constant parts of metabolic networks, namely the structure and annotation of biochemical reactions. In real-world scenarios, the dynamic changes in

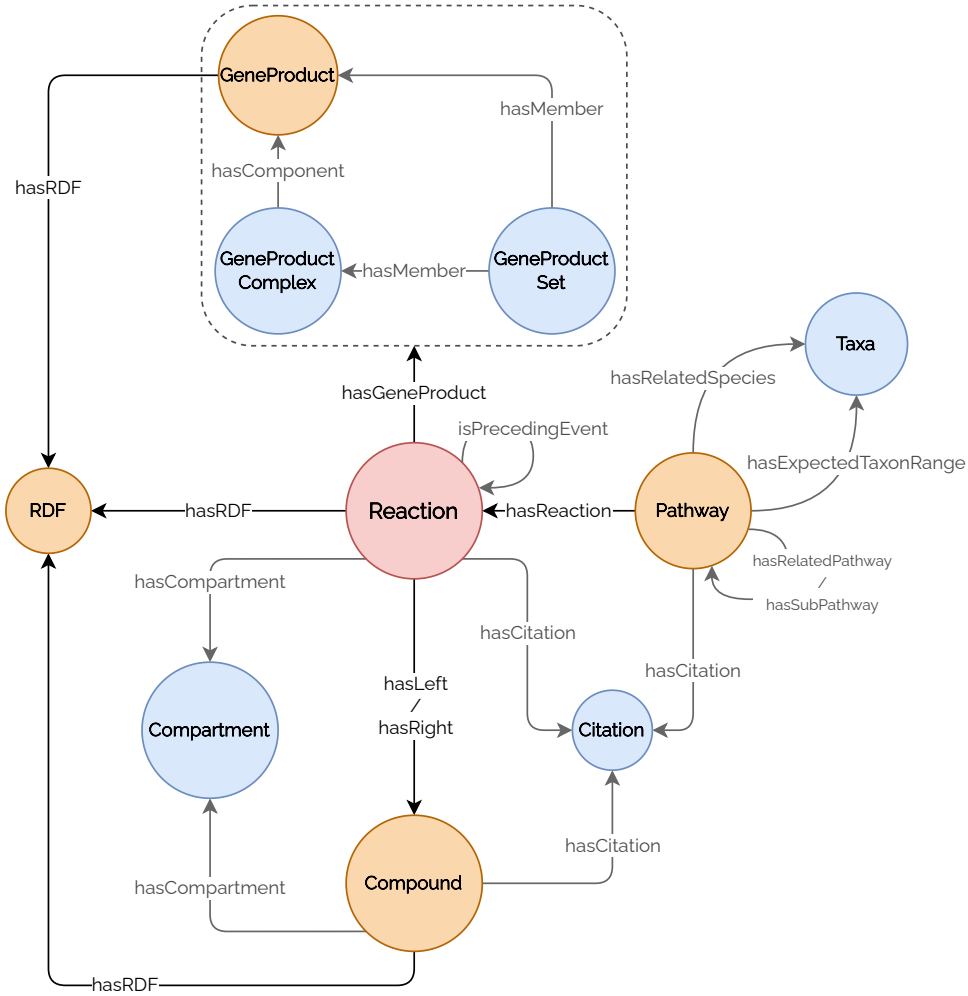


Figure 4.2: Schema of the augmented MetaCyc graph database. Light-gray edges indicate information integrated from external sources.

the network are often of higher research interest. In metabolic flux analysis, gene deletions or additional constraints are usually introduced to simulate their biological functions. In metabolic network perturbation analysis, tissue-specific transcriptional profiles can be used to interrogate the molecular targets of chemical compounds [154].

With the flexible graph representation, Metabolike is capable of mapping transcriptomics datasets or differential expression analysis results directly onto the metabolic graph by creating transient node

attributes. Metabolike can identify reprogrammed metabolic reaction routes originating from any given metabolite, presenting both activated and suppressed novel pathways (see Materials and Methods for implementation details). A graphical interface is built inside of Metabolike allowing users to: (a) upload custom datasets or use a preprocessed TCGA transcriptomics dataset; (b) identify a compound with fuzzy-matching; (c) tune various parameters such as the maximum number of reaction steps, certain compounds to ignore during the route search, and weight coefficients for the score metric; and (d) collect two tables containing the top activated and suppressed reprogrammed reaction routes, and an interactive graph of the reactions. A sample Metabolike route search web server is deployed at `https://metabolike.oy1zhou.com/`. The effectiveness of the route search algorithm is demonstrated in the following case studies, where we explored two well-known reprogrammed metabolisms in cancer.

Sialic acid metabolism in cancer Sialic acids are negatively charged sugar molecules with a nine-carbon backbone and have long been associated with tumor metastasis [50]. Gangliosides are glycosphingolipids with one or more sialic acids serving as capping molecules, thus they serve as metabolic exits for sialic acids *in vivo*. We have established in Chapter 3 that ganglioside types change as the cancer stage advances [155]. Back then, our approach was to extract marker genes from relevant literature and map their expression levels in cancer tissues onto the ganglioside synthesis network. Each reaction route was then inspected to determine its regulation state. With the help of Metabolike, most of the manual labor described above could be replaced with a few button clicks.

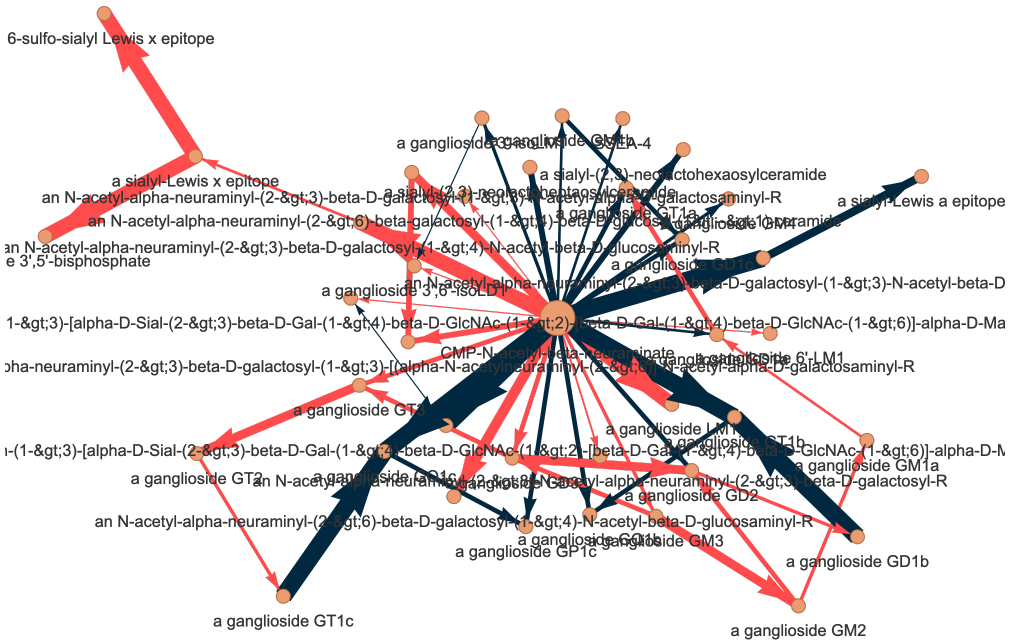


Figure 4.3: Reprogrammed metabolic reaction routes for sialic acid. Orange nodes represent compounds, and directed edges represent reactions. Red and blue edges denote activated and suppressed reactions, respectively. Width of the edges indicate reaction differential activity levels, with wider edges indicating higher differences between the cancer and normal tissues.

RNA-Seq data for 1222 breast cancer tissue samples were retrieved from the TCGA database [121], with 1109 being breast invasive carcinoma samples and 113 being tumor-adjacent normal samples. Differential expression analysis (DEA) was performed using the DESeq2 R package [144]. DEA results were mapped onto the metabolic graph constructed from BioCyc data files with Metabolike. CMP-N-acetyl-β-neuraminate, the common activated form of sialic acid, was passed to Metabolic as the starting point of the route search. Figure 4.3 shows all reprogrammed routes from sialic acid to terminal metabolites[8]. Consistent with our previous findings, gangliosides GM1a, GM2, GM3, GD1b, GD2, and GD3 all have

[8]It should be noted that the static exported image contains many overlapping compound names due to the tediously long descriptors of sugar molecules, but in the actual graphical interface the nodes are interactive and names can be easily distinguished.

elevated levels of synthesis. Metabolike also identified a cluster of sialyl-Lewis^X epitopes, which have been reported to play critical roles in cancer metastasis [156] and may be of further research interest.

Pan-cancer study of ammonia metabolism One of the most recognized reprogrammed metabolisms in cancer is nitrogen catabolism and disposal [39]. Under physiological conditions, excess nitrogen is converted to urea and disposed in the urine. A common form of intracellular nitrogen is ammonia, a toxic waste product that is usually metabolized via the urea cycle. Multiple studies have reported cancer cells reprogramming the urea cycle towards anabolic pathways [39, 157]. Partial inhibition of the urea cycle enables cancer cells to recycle ammonia and utilize it as a nitrogen source by re-assimilating it into the TCA cycle [157]. Urea cycle is also suppressed by p53, driving the inhibition of polyamine biosynthesis as the ammonia concentration increases [158]. Taken together, reprogrammed ammonia metabolism drives increased production of glutamine, proline, and aspartate by rewiring routes from urea cycle intermediates such as ornithine, citrulline, and arginino-succinate.

Transcriptomics data of 14 cancer types were retrieved from the TCGA database and DEA was subsequently performed. Without any other prior information, Metabolike identified the inhibition of the urea cycle in all cancers, and correctly discovered the rewiring of the metabolism towards the TCA cycle via glutamine (Table S1). Interestingly, elevated sarcosine levels were predicted in prostate cancer, confirming reports of its role as a biomarker [159, 160]. However, sarcosine biosynthesis appears to be significantly suppressed in all other cancer types, indicating unknown functions of the amino acid.

4.5 Discussion

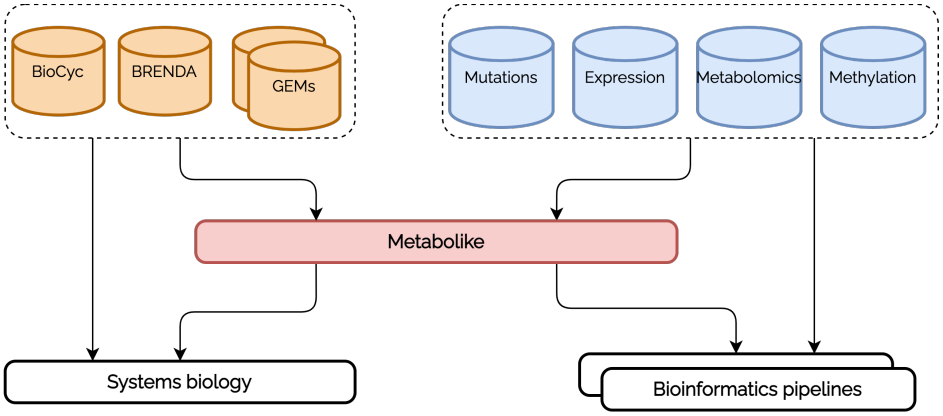


Figure 4.4: Bioinformatics pipeline with Metabolike.

We have developed a Python tool named Metabolike that can parse SBML metabolic models and transform it into a graph database. Metabolike can also map transcriptomics data directly onto the metabolic graph and identify reaction route changes for any metabolite of interest. The goal of developing Metabolike is not to replace the existing tools, but to serve as a compatibility layer that brings together omics data and the well-studied systems biology models (Figure 4.4). Typical RNA-Seq bioinformatics pipelines often end with a gene set enrichment analysis [63] that identifies relevant biological processes and functional gene clusters. However, these methods are biased towards genes present in multiple pathways and cannot identify sub-networks that are not part of a predefined pathway. They are also unaware of the overall network structure, meaning that connections between pathways are not taken into account as all gene sets are tested independently [64]. With Metabolike, we overcome these limitations by utilizing the graph structure of metabolic networks.

Currently, Metabolike could be used when the researcher has collected a list of disregulated gene sets and is interested in the specific metabolic disruptions that occurred in the network. The reaction route algorithm can take in the differential expression data, zoom-in to a specific metabolite, and return all related chains of reactions that are activated or repressed. A graphical interface is built into Metabolike to minimize the amount of programming background required for the end user. To make Metabolike

a more useful resource, we plan to support more data export formats in the next release, including stoichiometric matrices used by metabolic flux analysis software, and node-edge representation of a subgraph used by graph neural networks. We also aim to integrate gene regulatory networks and signal transduction networks into the graph database, which will make Metabolike applicable under more circumstances.

Chapter 5

Conclusion

In this dissertation, we have dissected mechanisms bridging microenvironment stresses to cancer phenotypic behaviors, particularly with regard to the disrupted intracellular pH. More specifically, we have characterized the role of metabolic reprogramming in mitigating intracellular alkaline stress and driving tumor progression.

In Chapter 2, we analyzed a large collection of commonly observed reprogrammed metabolisms and discovered their tendency of producing more H^+ compared to their original routes. Coupled with our previous studies on the pH_i stress induced by the Fenton reaction, these findings suggest the possibility that cancer is the process of overcoming a persistent alkalizing stress via reprogramming its metabolisms at a cellular level. We also hypothesized that cancer cells are forced to find metabolic exits for the overproduced metabolites of the reprogrammed metabolisms in a sustained manner. This was confirmed in Chapter 3, where we predicted a link between the accumulation of sialic acids on the cancer cell surface and the deformation and migration of cancer cells. We revealed the increased production and deployment of sialic acids in cancer, and discovered high correlation between the level of sialic acid synthesis and cytosolic Fenton reaction levels. The sialic acid-related marker genes also strongly correlated with a set of cell migration related processes, which provided further proof for our mechanical stress-based metastasis model.

In Chapter 4, we proposed an open-source Python package, Metabolike, that could parse Systems Biology Markup Language models and transform it into a graph in the Neo4j database. The software exposes a plugin interface that allows users to write custom parsers for other data types. As a proof of

concept, Metabolike included a parser for the BioCyc database and a CLI to integrate the parsed data into the graph. Apart from the core parser, Metabolike also includes a set of route searching algorithms that maps transcriptomics data onto the graph. Its ability of identifying reprogrammed metabolic pathways was showcased by revisiting pathways studied in the previous chapters and recognized reprogrammed metabolisms in cancer. In summary, our pH stress-induced cancer initiation model provides an elegant solution to the heterogeneity of cancer. We proposed a landscape of proton-producing reprogrammed metabolisms, and demonstrated their links to cancer phenotypes through statistical modeling. With the availability of the Metabolike toolkit, a more complete understanding of metabolic reprogramming could be mapped via the integration of transcriptomics data.

As a next step, we first plan to introduce more data export options to Metabolike, such as stoichiometric matrices used by metabolic flux analysis software, and node-edge representation of a subgraph used by graph neural networks (GNNs). The latter would allow researchers to infer potential reactions connecting metabolites (edge prediction), identify cellular states (graph classification), and detect disrupted reaction routes (node classification). The application of GNNs would convert our current workflow into a high-throughput pipeline, and produce rational scores of the identified novel metabolic routes. It would also take advantage of features such as enzyme functional parameters parsed from the BRENDA database and subcellular localizations of proteins, which is a step towards our ultimate goal of building a comprehensive knowledge graph of metabolic reactions.

Bibliography

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians* **72,** 7–33. ISSN: 1542-4863. doi:10.3322/caac.21708 (2022) (cit. on p. 1).

2. Dagogo-Jack, I. & Shaw, A. T. Tumour Heterogeneity and Resistance to Cancer Therapies. *Nat Rev Clin Oncol* **15,** 81–94. ISSN: 1759-4782. doi:10.1038/nrclinonc.2017.166 (2 Feb. 2018) (cit. on p. 1).

3. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674. ISSN: 0092-8674. doi:10.1016/j.cell.2011.02.013 (Mar. 4, 2011) (cit. on p. 1).

4. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12,** 31–46. ISSN: 2159-8274. doi:10.1158/2159-8290.CD-21-1059 (Jan. 12, 2022) (cit. on p. 1).

5. Webb, B. A., Chimenti, M., Jacobson, M. P. & Barber, D. L. Dysregulated pH: A Perfect Storm for Cancer Progression. *Nat Rev Cancer* **11,** 671–7. ISSN: 1474-1768 (Electronic) 1474-175X (Linking). doi:10.1038/nrc3110 (2011) (cit. on pp. 1, 2).

6. Talley, K. & Alexov, E. On the pH-optimum of Activity and Stability of Proteins. *Proteins* **78,** 2699–2706. ISSN: 0887-3585. doi:10.1002/prot.22786 (Sept. 2010) (cit. on p. 2).

7. Putney, L. K. & Barber, D. L. Na-H Exchange-dependent Increase in Intracellular pH Times G2/M Entry and Transition*. *Journal of Biological Chemistry* **278,** 44645–44649. ISSN: 0021-9258. doi:10.1074/jbc.M308099200 (Nov. 7, 2003) (cit. on p. 2).

8. Pouysségur, J., Franchi, A., L'Allemain, G. & Paris, S. Cytoplasmic pH, a Key Determinant of Growth Factor-Induced DNA Synthesis in Quiescent Fibroblasts. *FEBS Lett* **190,** 115–119. ISSN: 0014-5793. doi:10.1016/0014-5793(85)80439-7 (Oct. 7, 1985) (cit. on p. 2).

9. White, K. A., Grillo-Hill, B. K. & Barber, D. L. Cancer Cell Behaviors Mediated by Dysregulated pH Dynamics at a Glance. *J Cell Sci* **130,** 663–669. ISSN: 0021-9533. doi:10.1242/jcs.195297 (Feb. 15, 2017) (cit. on pp. 2, 13, 31).

10. Stock, C. & Schwab, A. Protons Make Tumor Cells Move like Clockwork. *Pflugers Arch - Eur J Physiol* **458,** 981–992. ISSN: 1432-2013. doi:10.1007/s00424-009-0677-8 (Sept. 1, 2009) (cit. on p. 2).

11. Srivastava, J., Barreiro, G., Groscurth, S., Gingras, A. R., Goult, B. T., *et al.* Structural Model and Functional Significance of pH-dependent Talin–Actin Binding for Focal Adhesion Remodeling. *Proceedings of the National Academy of Sciences* **105,** 14436–14441. doi:10.1073/pnas.0805163105 (Sept. 23, 2008) (cit. on p. 2).

12. Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B. & Gillies, R. J. Acid-Mediated Tumor Invasion: A Multidisciplinary Study. *Cancer Res* **66,** 5216–5223. ISSN: 0008-5472. doi:10.1158/0008-5472.CAN-05-4193 (May 15, 2006) (cit. on p. 2).

13. Stock, C., Cardone, R. A., Busco, G., Krähling, H., Schwab, A., *et al.* Protons Extruded by NHE1: Digestive or Glue? *Eur J Cell Biol* **87,** 591–599. ISSN: 0171-9335. doi:10.1016/j.ejcb.2008.01.007 (Sept. 2008) (cit. on p. 2).

14. Busco, G., Cardone, R. A., Greco, M. R., Bellizzi, A., Colella, M., *et al.* NHE1 Promotes Invadopodial ECM Proteolysis through Acidification of the Peri-Invadopodial Space. *FASEB J* **24,** 3903–3915. ISSN: 1530-6860. doi:10.1096/fj.09-149518 (Oct. 2010) (cit. on p. 2).

15. Zhou, Y., Chang, W., Lu, X., Wang, J., Zhang, C., *et al.* Acid-Base Homeostasis and Implications to the Phenotypic Behaviors of Cancer, 2022.03.04.482927. doi:10.1101/2022.03.04.482927 (Mar. 7, 2022) (cit. on p. 2).

16. Stevens, R. G. & Kalkwarf, D. R. Iron, Radiation, and Cancer. *Environ Health Perspect* **87,** 291–300. ISSN: 0091-6765 (Print) 0091-6765 (Linking). doi:10.1289/ehp.9087291 (1990) (cit. on p. 3).

17. Toyokuni, S. Role of Iron in Carcinogenesis: Cancer as a Ferrotoxic Disease. *Cancer Sci* **100,** 9–16. ISSN: 1349-7006 (Electronic) 1347-9032 (Linking). doi:10.1111/j.1349-7006.2008.01001.x (2009) (cit. on p. 3).

18. Akatsuka, S., Yamashita, Y., Ohara, H., Liu, Y. T., Izumiya, M., *et al.* Fenton Reaction Induced Cancer in Wild Type Rats Recapitulates Genomic Alterations Observed in Human Cancer. *PLoS One* **7,** e43403. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0043403 (2012) (cit. on p. 3).

19. FentonH, J. Oxidation of Tartaric Acid in Presence of Iron. *J Chem Soc, Trans* **65,** 899–910. doi:10.1039/CT8946500899 (1894) (cit. on p. 3).

20. Sun, H., Zhang, C., Cao, S., Sheng, T., Dong, N., *et al.* Fenton Reactions Drive Nucleotide and ATP Syntheses in Cancer. *J Mol Cell Biol* **10,** 448–459. ISSN: 1759-4685. doi:10.1093/jmcb/mjy039 (2018) (cit. on pp. 3, 4, 12–14, 26, 27, 31, 38, 41, 43, 47).

21. Han, D., Williams, E. & Cadenas, E. Mitochondrial Respiratory Chain-Dependent Generation of Superoxide Anion and Its Release into the Intermembrane Space. *Biochem J* **353,** 411–416. ISSN: 0264-6021. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1221585/ (2022) (Pt 2 Jan. 15, 2001) (cit. on p. 3).

22. Liou, G.-Y. & Storz, P. Reactive Oxygen Species in Cancer. *Free Radic Res* **44,** 10.3109/10715761003667554. ISSN: 1071-5762. doi:10.3109/10715761003667554 (May 2010) (cit. on p. 3).

23. Fong, K. L., McCay, P. B. & Poyer, J. L. Evidence for Superoxide-Dependent Reduction of Fe3+ and Its Role in Enzyme-Generated Hydroxyl Radical Formation. *Chem Biol Interact* **15,** 77–89. ISSN: 0009-2797 (Print) 0009-2797 (Linking). doi:10.1016/0009-2797(76)90130-7 (1976) (cit. on p. 3).

24. Kojima, N. & Bates, G. W. The Reduction and Release of Iron from Fe3+ .Transferrin.CO3(2-). *J Biol Chem* **254,** 8847–54. ISSN: 0021-9258 (Print) 0021-9258 (Linking). doi:10.1016/S0021-9258(19)86777-7 (1979) (cit. on p. 3).

25. Coussens, L. M. & Werb, Z. Inflammation and Cancer. *Nature* **420,** 860–7. ISSN: 0028-0836 (Print) 0028-0836 (Linking). doi:10.1038/nature01322 (2002) (cit. on p. 3).

26. Mohanty, J. G., Nagababu, E. & Rifkind, J. M. Red Blood Cell Oxidative Stress Impairs Oxygen Delivery and Induces Red Blood Cell Aging. *Front Physiol* **5,** 84. ISSN: 1664-042X (Print) 1664-042X (Linking). doi:10.3389/fphys.2014.00084 (2014) (cit. on p. 3).

27. Fibach, E. & Rachmilewitz, E. A. Iron Overload in Hematological Disorders. *Presse Med* **46,** e296–e305. ISSN: 2213-0276 (Electronic) 0755-4982 (Linking). doi:10.1016/j.lpm.2017.10.007 (12 Pt 2 2017) (cit. on p. 3).

28. Vigorita, V. J. & Hutchins, G. M. The Thymus in Hemochromatosis. *Am J Pathol* **93,** 661–5. ISSN: 0002-9440 (Print) 0002-9440 (Linking) (1978) (cit. on p. 3).

29. Mittal, M., Siddiqui, M. R., Tran, K., Reddy, S. P. & Malik, A. B. Reactive Oxygen Species in Inflammation and Tissue Injury. *Antioxid Redox Signal* **20,** 1126–67. ISSN: 1557-7716 (Electronic) 1523-0864 (Linking). doi:10.1089/ars.2012.5149 (2014) (cit. on p. 3).

30. Huang, X. Iron Overload and Its Association with Cancer Risk in Humans: Evidence for Iron as a Carcinogenic Metal. *Mutat Res* **533,** 153–71. ISSN: 0027-5107 (Print) 0027-5107 (Linking). doi:10.1016/j.mrfmmm.2003.08.023 (2003) (cit. on p. 3).

31. Ward, P. S. & Thompson, C. B. Metabolic Reprogramming: A Cancer Hallmark Even Warburg Did Not Anticipate. *Cancer Cell* **21,** 297–308. ISSN: 1535-6108. doi:10.1016/j.ccr.2012.02.014 (Mar. 20, 2012) (cit. on pp. 4, 11).

32. Ananieva, E. Targeting Amino Acid Metabolism in Cancer Growth and Anti-Tumor Immune Response. *World J Biol Chem* **6,** 281–9. ISSN: 1949-8454 (Print) 1949-8454 (Linking). doi:10.4331/wjbc.v6.i4.281 (2015) (cit. on pp. 4, 12).

33. Aird, K. M. & Zhang, R. Nucleotide Metabolism, Oncogene-Induced Senescence and Cancer. *Cancer Lett* **356,** 204–10. ISSN: 1872-7980 (Electronic) 0304-3835 (Linking). doi:10.1016/j.canlet.2014.01.017 (2 Pt A 2015) (cit. on pp. 4, 12).

34. Villa, E., Ali, E. S., Sahu, U. & Ben-Sahra, I. Cancer Cells Tune the Signaling Pathways to Empower de Novo Synthesis of Nucleotides. *Cancers (Basel)* **11,** 688. ISSN: 2072-6694. doi:10/gn3v8r (May 17, 2019) (cit. on pp. 4, 20).

35. Long, J., Zhang, C. J., Zhu, N., Du, K., Yin, Y. F., *et al.* Lipid Metabolism and Carcinogenesis, Cancer Development. *Am J Cancer Res* **8,** 778–791. ISSN: 2156-6976 (Print) 2156-6976 (Linking) (2018) (cit. on pp. 4, 12, 24, 25).

36. Hay, N. Reprogramming Glucose Metabolism in Cancer: Can It Be Exploited for Cancer Therapy? *Nat Rev Cancer* **16,** 635–49. ISSN: 1474-1768 (Electronic) 1474-175X (Linking). doi:10.1038/nrc.2016.77 (2016) (cit. on pp. 4, 12).

37. Mamede, A. C., Tavares, S. D., Abrantes, A. M., Trindade, J., Maia, J. M., *et al.* The Role of Vitamins in Cancer: A Review. *Nutr Cancer* **63,** 479–94. ISSN: 1532-7914 (Electronic) 0163-5581 (Linking). doi:10.1080/01635581.2011.539315 (2011) (cit. on pp. 4, 12).

38. Ramalho, R. T., Aydos, R. D., Schettert, I., Assis, P. V. & Cassino, P. C. Sulfane Sulfur Deficiency in Malignant Cells, Increasing the Inhibiting Action of Acetone Cyanohydrin in Tumor Growth. *Acta Cir Bras* **28,** 728–32. ISSN: 1678-2674 (Electronic) 0102-8650 (Linking). doi:10.1590/S0102-86502013001000007 (2013) (cit. on pp. 4, 12).

39. Keshet, R., Szlosarek, P., Carracedo, A. & Erez, A. Rewiring Urea Cycle Metabolism in Cancer to Support Anabolism. *Nat Rev Cancer* **18,** 634–645. ISSN: 1474-1768. doi:10/gfcf34 (10 Oct. 2018) (cit. on pp. 4, 80).

40. Koundouros, N. & Poulogiannis, G. Reprogramming of Fatty Acid Metabolism in Cancer. *Br J Cancer* **122,** 4–22. ISSN: 1532-1827. doi:10/gj23wg (1 Jan. 2020) (cit. on p. 4).

41. Leithner, K., Hrzenjak, A. & Olschewski, H. Gluconeogenesis in Cancer: Door Wide Open. *PNAS* **111,** E4394–E4394. ISSN: 0027-8424, 1091-6490. doi:10/gn3v8t (Oct. 21, 2014) (cit. on p. 4).

42. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does It Benefit Cancer Cells? *Trends Biochem Sci* **41,** 211–218. ISSN: 0968-0004. doi:10/gf3d86 (Mar. 2016) (cit. on p. 4).

43. Rais, R., Lemberg, K. M., Tenora, L., Arwood, M. L., Pal, A., *et al.* Discovery of DRP-104, a Tumor-Targeted Metabolic Inhibitor Prodrug. *Science Advances* **8,** eabq5925. doi:10.1126/sciadv.abq5925 (Nov. 16, 2022) (cit. on p. 4).

44. Lemberg, K. M., Gori, S. S., Tsukamoto, T., Rais, R. & Slusher, B. S. Clinical Development of Metabolic Inhibitors for Oncology. *J Clin Invest* **132.** ISSN: 0021-9738. doi:10.1172/JCI148550 (Jan. 4, 2022) (cit. on p. 4).

45. Shi, D. D., Savani, M. R., Levitt, M. M., Wang, A. C., Endress, J. E., *et al.* De Novo Pyrimidine Synthesis Is a Targetable Vulnerability in IDH Mutant Glioma. *Cancer Cell* **40,** 939–956.e16. ISSN: 1535-6108, 1878-3686. doi:10.1016/j.ccell.2022.07.011 (Sept. 12, 2022) (cit. on p. 4).

46. Pal, S., Kaplan, J. P., Nguyen, H., Stopka, S. A., Savani, M. R., *et al.* A Druggable Addiction to de Novo Pyrimidine Biosynthesis in Diffuse Midline Glioma. *Cancer Cell* **40,** 957–972.e10. ISSN: 1535-6108, 1878-3686. doi:10.1016/j.ccell.2022.07.012 (Sept. 12, 2022) (cit. on p. 4).

47. Sun, H., Zhou, Y., Skaro, M. F., Wu, Y., Qu, Z., *et al.* Metabolic Reprogramming in Cancer Is Induced to Increase Proton Production. *Cancer Res* **80,** 1143–1155. ISSN: 1538-7445 (Electronic) 0008-5472 (Linking). doi:10.1158/0008-5472.CAN-19-3392 (2020) (cit. on pp. 5, 41, 43, 47, 53).

48. Giacomini, I., Gianfanti, F., Desbats, M. A., Orso, G., Berretta, M., *et al.* Cholesterol Metabolic Reprogramming in Cancer and Its Pharmacological Modulation as Therapeutic Strategy. *Frontiers in Oncology* **11.** ISSN: 2234-943X. doi:10.3389/fonc.2021.682911 (2021) (cit. on p. 5).

49. Gasic, G. & Gasic, T. Removal of Sialic Acid from the Cell Coat in Tumor Cells and Vascular Endothelium, and Its Effects on Metastasis. *Proc Natl Acad Sci U S A* **48,** 1172–7. ISSN: 0027-8424 (Print) 0027-8424 (Linking). doi:10.1073/pnas.48.7.1172 (1962) (cit. on pp. 5, 40, 41).

50. Ohta, N., Pardee, A. B., McAuslan, B. R. & Burger, M. M. Sialic Acid Contents and Controls of Normal and Malignant Cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* **158,** 98–102. ISSN: 0304-4165. doi:10.1016/0304-4165(68)90076-7 (Apr. 16, 1968) (cit. on pp. 5, 40, 78).

51. Brocca, P., Rondelli, V., Mallamace, F., Di Bari, M. T., Deriu, A., *et al.* Water Response to Ganglioside GM1 Surface Remodelling. *Biochim Biophys Acta Gen Subj* **1861,** 3573–3580. ISSN: 0304-4165 (Print) 0304-4165. doi:10.1016/j.bbagen.2016.04.029 (1 Pt B 2017) (cit. on pp. 5, 40).

52. Sato, C. & Kitajima, K. Disialic, Oligosialic and Polysialic Acids: Distribution, Functions and Related Disease. *The Journal of Biochemistry* **154,** 115–136. ISSN: 1756-2651. doi:10.1093/jb/mvt057 (2013) (cit. on pp. 5, 41).

53. Büll, C., den Brok, M. H. & Adema, G. J. Sweet Escape: Sialic Acids in Tumor Immune Evasion. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **1846,** 238–246. ISSN: 0304-419X. doi:10.1016/j.bbcan.2014.07.005 (2014) (cit. on pp. 5, 41).

54. Cheng, G., Tse, J., Jain, R. K. & Munn, L. L. Micro-Environmental Mechanical Stress Controls Tumor Spheroid Size and Morphology by Suppressing Proliferation and Inducing Apoptosis in Cancer Cells. *PLoS one* **4,** e4632. ISSN: 1932-6203. doi:10.1371/journal.pone.0004632 (2009) (cit. on pp. 5, 41, 42).

55. Helmlinger, G., Netti, P. A., Lichtenbeld, H. C., Melder, R. J. & Jain, R. K. Solid Stress Inhibits the Growth of Multicellular Tumor Spheroids. *Nature biotechnology* **15,** 778. ISSN: 1546-1696. doi:10.1038/nbt0897-778 (1997) (cit. on pp. 5, 41, 42).

56. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28,** 27–30. ISSN: 0305-1048. doi:10/b9st54 (Jan. 1, 2000) (cit. on pp. 5, 64, 68, 69).

57. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., *et al.* BRENDA, the Enzyme Database: Updates and Major New Developments. *Nucleic Acids Research* **32,** D431–D433. ISSN: 0305-1048. doi:10.1093/nar/gkh081 (suppl_1 Jan. 1, 2004) (cit. on pp. 5, 6, 64, 73).

58. Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., *et al.* The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases. *Nucleic Acids Research* **44,** D471–D480. ISSN: 0305-1048. doi:10/f8cp7f (Jan. 4, 2016) (cit. on pp. 5, 6, 64, 73).

59. Masid, M. & Hatzimanikatis, V. Quantitative Modeling of Human Metabolism: A Call for a Community Effort. *Current Opinion in Systems Biology* **26,** 109–115. ISSN: 2452-3100. doi:10.1016/j.coisb.2021.04.008 (June 1, 2021) (cit. on pp. 6, 65).

60. Miller, J. Graph Database Applications and Concepts with Neo4j. *SAIS 2013Proceedings.* https://aisel.aisnet.org/sais2013/24 (May 18, 2013) (cit. on p. 6).

61. Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., *et al.* Reactome Graph Database: Efficient Access to Complex Pathway Data. *PLOS Computational Biology* **14,** e1005968. ISSN: 1553-7358. doi:10/gczm25 (Jan. 29, 2018) (cit. on pp. 6, 71, 74).

62. Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., *et al.* The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models. *Bioinformatics* **19,** 524–531. ISSN: 1367-4803. doi:10.1093/bioinformatics/btg015 (Mar. 1, 2003) (cit. on pp. 6, 65).

63. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., *et al.* Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proceedings of the National Academy of Sciences* **102,** 15545–15550. doi:10.1073/pnas.0506580102 (Oct. 25, 2005) (cit. on pp. 7, 81).

64. Simillion, C., Liechti, R., Lischer, H. E., Ioannidis, V. & Bruggmann, R. Avoiding the Pitfalls of Gene Set Enrichment Analysis with SetRank. *BMC Bioinformatics* **18,** 151. ISSN: 1471-2105. doi:10.1186/s12859-017-1571-6 (Mar. 4, 2017) (cit. on pp. 7, 81).

65. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324,** 1029–33. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). doi:10.1126/science.1160809 (2009) (cit. on pp. 12, 26).

66. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on Fat: The Role of Lipid Synthesis in Cancer Metabolism and Tumour Development. *Disease Models & Mechanisms* **6,** 1353–1363. ISSN: 1754-8403. doi:10.1242/dmm.011338 (Nov. 1, 2013) (cit. on p. 12).

67. Pearce, O. M. T. & Läubli, H. Sialic Acids in Cancer Biology and Immunity. *Glycobiology* **26,** 111–128. ISSN: 0959-6658. doi:10.1093/glycob/cwv097 (Feb. 1, 2016) (cit. on pp. 12, 26, 36).

68. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab* **7,** 11–20. ISSN: 1550-4131 (Print) 1550-4131 (Linking). doi:10.1016/j.cmet.2007.10.002 (2008) (cit. on p. 12).

69. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab* **23,** 27–47. ISSN: 1932-7420 (Electronic) 1550-4131 (Linking). doi:10.1016/j.cmet.2015.12.006 (2016) (cit. on p. 12).

70. Ferreira, L. M., Hebrant, A. & Dumont, J. E. Metabolic Reprogramming of the Tumor. *Oncogene* **31,** 3999–4011. ISSN: 1476-5594 (Electronic) 0950-9232 (Linking). doi:10.1038/onc.2011.576 (2012) (cit. on p. 12).

71. Zhou, Z., Ibekwe, E. & Chornenkyy, Y. Metabolic Alterations in Cancer Cells and the Emerging Role of Oncometabolites as Drivers of Neoplastic Change. *Antioxidants (Basel)* **7.** ISSN: 2076-3921 (Print) 2076-3921 (Linking). doi:10.3390/antiox7010016 (2018) (cit. on p. 12).

72. Clementino, M., Shi, X. & Zhang, Z. Oxidative Stress and Metabolic Reprogramming in Cr(VI) Carcinogenesis. *Curr Opin Toxicol* **8,** 20–27. ISSN: 2468-2934 (Print) 2468-2020 (Linking). doi:10.1016/j.cotox.2017.11.015 (2018) (cit. on p. 12).

73. Yang, C., Jiang, L., Zhang, H., Shimoda, L. A., DeBerardinis, R. J., *et al.* Analysis of Hypoxia-Induced Metabolic Reprogramming. *Methods Enzymol* **542,** 425–55. ISSN: 1557-7988 (Electronic) 0076-6879 (Linking). doi:10.1016/B978-0-12-416618-9.00022-4 (2014) (cit. on p. 12).

74. Sun, H., Chen, L., Cao, S., Liang, Y. & Xu, Y. Warburg Effects in Cancer and Normal Proliferating Cells: Two Tales of the Same Name. *Genomics Proteomics Bioinformatics* **17,** 273–286. ISSN: 2210-3244 (Electronic) 1672-0229 (Linking). doi:10.1016/j.gpb.2018.12.006 (2019) (cit. on pp. 12, 13, 26, 30, 31, 43).

75. Swietach, P., Vaughan-Jones, R. D., Harris, A. L. & Hulikova, A. The Chemistry, Physiology and Pathology of pH in Cancer. *Philos Trans R Soc Lond B Biol Sci* **369,** 20130099. ISSN: 1471-2970 (Electronic) 0962-8436 (Linking). doi:10.1098/rstb.2013.0099 (2014) (cit. on pp. 13, 31).

76. Persi, E., Duran-Frigola, M., Damaghi, M., Roush, W. R., Aloy, P., *et al.* Systems Analysis of Intracellular pH Vulnerabilities for Cancer Therapy. *Nat Commun* **9,** 2997. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). doi:10.1038/s41467-018-05261-x (2018) (cit. on pp. 13, 31).

77. McCarty, M. F. & Whitaker, J. Manipulating Tumor Acidification as a Cancer Treatment Strategy. *Altern Med Rev* **15,** 264–72. ISSN: 1089-5159 (Print) 1089-5159 (Linking) (2010) (cit. on pp. 13, 31).

78. Davies, K. J. Protein Damage and Degradation by Oxygen Radicals. I. General Aspects. *J Biol Chem* **262,** 9895–901. ISSN: 0021-9258 (Print) 0021-9258 (Linking) (1987) (cit. on p. 14).

79. Schrader, J., Henneberg, F., Mata, R. A., Tittmann, K., Schneider, T. R., *et al.* The Inhibition Mechanism of Human 20S Proteasomes Enables Next-Generation Inhibitor Design. *Science* **353,** 594–8. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). doi:10.1126/science.aaf8993 (2016) (cit. on p. 14).

80. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33,** 1–22. ISSN: 1548-7660 (Print) 1548-7660 (Linking). doi:10.18637/jss.v033.i01 (2010) (cit. on p. 15).

81. Foster, A. C., Farnsworth, J., Lind, G. E., Li, Y. X., Yang, J. Y., *et al.* D-Serine Is a Substrate for Neutral Amino Acid Transporters ASCT1/SLC1A4 and ASCT2/SLC1A5, and Is Transported by Both Subtypes in Rat Hippocampal Astrocyte Cultures. *PLoS One* **11,** e0156551. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0156551 (2016) (cit. on p. 18).

82. Newman, A. C. & Maddocks, O. D. K. Serine and Functional Metabolites in Cancer. *Trends Cell Biol* **27,** 645–657. ISSN: 1879-3088 (Electronic) 0962-8924 (Linking). doi:10.1016/j.tcb.2017.05.001 (2017) (cit. on p. 18).

83. UniProt, C. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res* **47,** D506–D515. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). doi:10.1093/nar/gky1049 (2019) (cit. on p. 18).

84. Platten, M., Wick, W. & Van den Eynde, B. J. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. *Cancer Res* **72,** 5435–40. ISSN: 1538-7445 (Electronic) 0008-5472 (Linking). doi:10.1158/0008-5472.CAN-12-0569 (2012) (cit. on p. 19).

85. Puccetti, P., Fallarino, F., Italiano, A., Soubeyran, I., MacGrogan, G., *et al.* Accumulation of an Endogenous Tryptophan-Derived Metabolite in Colorectal and Breast Cancers. *PLoS One* **10,** e0122046. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0122046 (2015) (cit. on p. 19).

86. Dankers, A. C., Mutsaers, H. A., Dijkman, H. B., van den Heuvel, L. P., Hoenderop, J. G., *et al.* Hyperuricemia Influences Tryptophan Metabolism via Inhibition of Multidrug Resistance Protein 4 (MRP4) and Breast Cancer Resistance Protein (BCRP). *Biochim Biophys Acta* **1832,** 1715–22. ISSN: 0006-3002 (Print) 0006-3002 (Linking). doi:10.1016/j.bbadis.2013.05.002 (2013) (cit. on p. 19).

87. Mbongue, J. C., Nicholas, D. A., Torrez, T. W., Kim, N. S., Firek, A. F., *et al.* The Role of Indoleamine 2, 3-Dioxygenase in Immune Suppression and Autoimmunity. *Vaccines (Basel)* **3,** 703–29. ISSN: 2076-393X (Print) 2076-393X (Linking). doi:10.3390/vaccines3030703 (2015) (cit. on p. 19).

88. Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., *et al.* Mutational Landscape and Significance across 12 Major Cancer Types. *Nature* **502,** 333–339. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). doi:10.1038/nature12634 (2013) (cit. on p. 20).

89. Jezequel, P., Joalland, M. P., Milano, G., Lanoe, D., Ricolleau, G., *et al.* Common DPYD Mutation Associated with 5-Fluorouracil Toxicity Detected by PCR-mediated Site-Directed Mutagenesis. *Clin Chem* **46,** 309–10. ISSN: 0009-9147 (Print) 0009-9147. doi:10.1093/clinchem/46.2.309 (2000) (cit. on p. 22).

90. Buj, R. & Aird, K. M. Deoxyribonucleotide Triphosphate Metabolism in Cancer and Metabolic Disease. *Front Endocrinol (Lausanne)* **9,** 177. ISSN: 1664-2392 (Print) 1664-2392 (Linking). doi:10.3389/fendo.2018.00177 (2018) (cit. on p. 22).

91. Beloribi-Djefaflia, S., Vasseur, S. & Guillaumond, F. Lipid Metabolic Reprogramming in Cancer Cells. *Oncogenesis* **5,** e189. ISSN: 2157-9024 (Print) 2157-9024 (Linking). doi:10.1038/oncsis.2015.49 (2016) (cit. on p. 24).

92. Cheng, M., Bhujwalla, Z. M. & Glunde, K. Targeting Phospholipid Metabolism in Cancer. *Front Oncol* **6,** 266. ISSN: 2234-943X (Print) 2234-943X (Linking). doi:10.3389/fonc.2016.00266 (2016) (cit. on p. 24).

93. Lin, Z., Liu, F., Shi, P., Song, A., Huang, Z., *et al.* Fatty Acid Oxidation Promotes Reprogramming by Enhancing Oxidative Phosphorylation and Inhibiting Protein Kinase C. *Stem Cell Res Ther* **9,** 47. ISSN: 1757-6512 (Electronic) 1757-6512 (Linking). doi:10.1186/s13287-018-0792-6 (2018) (cit. on p. 24).

94. Abramson, H. N. The Lipogenesis Pathway as a Cancer Target. *J Med Chem* **54,** 5615–38. ISSN: 1520-4804 (Electronic) 0022-2623 (Linking). doi:10.1021/jm2005805 (2011) (cit. on p. 24).

95. Zhang, Y. & Yang, J. M. Altered Energy Metabolism in Cancer: A Unique Opportunity for Therapeutic Intervention. *Cancer Biol Ther* **14,** 81–9. ISSN: 1555-8576 (Electronic) 1538-4047 (Linking). doi:10.4161/cbt.22958 (2013) (cit. on p. 24).

96. Wang, D. & Dubois, R. N. Eicosanoids and Cancer. *Nat Rev Cancer* **10,** 181–93. ISSN: 1474-1768 (Electronic) 1474-175X (Linking). doi:10.1038/nrc2809 (2010) (cit. on p. 24).

97. Saddoughi, S. A., Song, P. & Ogretmen, B. Roles of Bioactive Sphingolipids in Cancer Biology and Therapeutics. *Subcell Biochem* **49,** 413–40. ISSN: 0306-0225 (Print) 0306-0225 (Linking). doi:10.1007/978-1-4020-8831-5_16 (2008) (cit. on p. 24).

98. Patwardhan, G. A., Beverly, L. J. & Siskind, L. J. Sphingolipids and Mitochondrial Apoptosis. *J Bioenerg Biomembr* **48,** 153–68. ISSN: 1573-6881 (Electronic) 0145-479X (Linking). doi:10.1007/s10863-015-9602-3 (2016) (cit. on p. 24).

99. Kamphorst, J. J., Cross, J. R., Fan, J., de Stanchina, E., Mathew, R., *et al.* Hypoxic and Ras-transformed Cells Support Growth by Scavenging Unsaturated Fatty Acids from Lysophospholipids. *Proc Natl Acad Sci U S A* **110,** 8882–7. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). doi:10.1073/pnas.1307237110 (2013) (cit. on p. 24).

100. Rohrig, F. & Schulze, A. The Multifaceted Roles of Fatty Acid Synthesis in Cancer. *Nat Rev Cancer* **16,** 732–749. ISSN: 1474-1768 (Electronic) 1474-175X (Linking). doi:10.1038/nrc.2016.89 (2016) (cit. on p. 24).

101. Williams, J. C., Kizaki, H., Weber, G. & Morris, H. P. Increased CTP Synthetase Activity in Cancer Cells. *Nature* **271,** 71–3. ISSN: 0028-0836 (Print) 0028-0836 (Linking). doi:10.1038/271071a0 (1978) (cit. on p. 25).

102. Lu, J., Tan, M. & Cai, Q. The Warburg Effect in Tumor Progression: Mitochondrial Oxidative Metabolism as an Anti-Metastasis Mechanism. *Cancer Lett* **356,** 156–64. ISSN: 1872-7980 (Electronic) 0304-3835 (Linking). doi:10.1016/j.canlet.2014.04.001 (2 Pt A 2015) (cit. on p. 26).

103. Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., *et al.* Glycosaminoglycans: Key Players in Cancer Cell Biology and Treatment. *FEBS J* **279,** 1177–97. ISSN: 1742-4658 (Electronic) 1742-464X (Linking). doi:10.1111/j.1742-4658.2012.08529.x (2012) (cit. on p. 26).

104. Li, X., Lan, Y., He, Y., Liu, Y., Luo, H., *et al.* Heparan Sulfate and Chondroitin Sulfate Glycosaminoglycans Are Targeted by Bleomycin in Cancer Cells. *Cell Physiol Biochem* **43,** 1220–1234. ISSN: 1421-9778 (Electronic) 1015-8987 (Linking). doi:10.1159/000481763 (2017) (cit. on p. 26).

105. Ho, W. L., Hsu, W. M., Huang, M. C., Kadomatsu, K. & Nakagawara, A. Protein Glycosylation in Cancers and Its Potential Therapeutic Applications in Neuroblastoma. *J Hematol Oncol* **9,** 100. ISSN: 1756-8722 (Electronic) 1756-8722 (Linking). doi:10.1186/s13045-016-0334-6 (2016) (cit. on p. 27).

106. Glunde, K., Penet, M. F., Jiang, L., Jacobs, M. A. & Bhujwalla, Z. M. Choline Metabolism-Based Molecular Diagnosis of Cancer: An Update. *Expert Rev Mol Diagn* **15,** 735–47. ISSN: 1744-8352 (Electronic) 1473-7159 (Linking). doi:10.1586/14737159.2015.1039515 (2015) (cit. on p. 29).

107. Garrido, A. & Djouder, N. NAD(+) Deficits in Age-Related Diseases and Cancer. *Trends Cancer* **3,** 593–610. ISSN: 2405-8025 (Electronic) 2405-8025 (Linking). doi:10.1016/j.trecan.2017.06.001 (2017) (cit. on p. 29).

108. Chen, Y., Fan, Z., Yang, Y. & Gu, C. Iron Metabolism and Its Contribution to Cancer (Review). *Int J Oncol* **54,** 1143–1154. ISSN: 1791-2423 (Electronic) 1019-6439 (Linking). doi:10.3892/ijo.2019.4720 (2019) (cit. on p. 31).

109. Babbs, C. F. Free Radicals and the Etiology of Colon Cancer. *Free Radic Biol Med* **8,** 191–200. ISSN: 0891-5849 (Print) 0891-5849 (Linking). doi:10.1016/0891-5849(90)90091-V (1990) (cit. on p. 31).

110. Büll, C., Stoel, M. A., den Brok, M. H. & Adema, G. J. Sialic Acids Sweeten a Tumor's Life. *Cancer Research* **74,** 3199–3204. ISSN: 0008-5472. doi:10.1158/0008-5472.CAN-14-0728 (June 12, 2014) (cit. on pp. 32, 46).

111. Monahan, L. G., Hajduk, I. V., Blaber, S. P., Charles, I. G. & Harry, E. J. Coordinating Bacterial Cell Division with Nutrient Availability: A Role for Glycolysis. *MBio* **5,** e00935–14. ISSN: 2150-7511 (Electronic). doi:10.1128/mBio.00935-14 (2014) (cit. on p. 35).

112. Monahan, L. G. & Harry, E. J. You Are What You Eat: Metabolic Control of Bacterial Division. *Trends Microbiol* **24,** 181–189. ISSN: 1878-4380 (Electronic) 0966-842X (Linking). doi:10.1016/j.tim.2015.11.007 (2016) (cit. on p. 35).

113. Trigos, A. S., Pearson, R. B., Papenfuss, A. T. & Goode, D. L. Altered Interactions between Unicellular and Multicellular Genes Drive Hallmarks of Transformation in a Diverse Range of Solid Tumors. *Proc Natl Acad Sci U S A* **114,** 6406–6411. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). doi:10.1073/pnas.1617743114 (2017) (cit. on p. 35).

114. Birkle, S., Zeng, G., Gao, L., Yu, R. K. & Aubry, J. Role of Tumor-Associated Gangliosides in Cancer Progression. *Biochimie* **85,** 455–63. ISSN: 0300-9084 (Print) 0300-9084 (Linking). doi:10.1016/S0300-9084(03)00006-3 (2003) (cit. on pp. 36, 51, 62).

115. Liesenfeld, D. B., Grapov, D., Fahrmann, J. F., Salou, M., Scherer, D., *et al.* Metabolomics and Transcriptomics Identify Pathway Differences between Visceral and Subcutaneous Adipose Tissue in Colorectal Cancer Patients: The ColoCare Study. *Am J Clin Nutr* **102,** 433–43. ISSN: 1938-3207 (Electronic) 0002-9165 (Linking). doi:10.3945/ajcn.114.103804 (2015) (cit. on p. 36).

116. Heiland, D. H., Worner, J., Gerrit Haaker, J., Delev, D., Pompe, N., *et al.* The Integrative Metabolomic-Transcriptomic Landscape of Glioblastome Multiforme. *Oncotarget* **8,** 49178–49190. ISSN: 1949-2553 (Electronic) 1949-2553 (Linking). doi:10.18632/oncotarget.16544 (2017) (cit. on p. 36).

117. Terunuma, A., Putluri, N., Mishra, P., Mathe, E. A., Dorsey, T. H., *et al.* MYC-driven Accumulation of 2-Hydroxyglutarate Is Associated with Breast Cancer Prognosis. *J Clin Invest* **124,** 398–412. ISSN: 1558-8238 (Electronic) 0021-9738 (Linking). doi:10.1172/JCI71180 (2014) (cit. on p. 36).

118. Ren, S., Shao, Y., Zhao, X., Hong, C. S., Wang, F., *et al.* Integration of Metabolomics and Transcriptomics Reveals Major Metabolic Pathways and Potential Biomarker Involved in Prostate Cancer. *Mol Cell Proteomics* **15,** 154–63. ISSN: 1535-9484 (Electronic) 1535-9476 (Linking). doi:10.1074/mcp.M115.052381 (2016) (cit. on p. 36).

119. Bencina, M. Illumination of the Spatial Order of Intracellular pH by Genetically Encoded pH-sensitive Sensors. *Sensors (Basel)* **13,** 16736–58. ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). doi:10.3390/s131216736 (2013) (cit. on p. 36).

120. Pellegatti, P., Raffaghello, L., Bianchi, G., Piccardi, F., Pistoia, V., *et al.* Increased Level of Extracellular ATP at Tumor Sites: In Vivo Imaging with Plasma Membrane Luciferase. *PLoS One* **3,** e2599. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0002599 (2008) (cit. on p. 38).

121. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemp Oncol (Pozn)* **19,** A68–A77. ISSN: 1428-2526. doi:10.5114/wo.2014.47136 (1A 2015) (cit. on pp. 41, 67, 79).

122. Tse, J. M., Cheng, G., Tyrrell, J. A., Wilcox-Adelman, S. A., Boucher, Y., *et al.* Mechanical Compression Drives Cancer Cells toward Invasive Phenotype. *Proc Natl Acad Sci U S A* **109,** 911–6. ISSN: 0027-8424. doi:10.1073/pnas.1118910109 (2012) (cit. on pp. 42, 61).

123. Goldman, M., Craft, B., Brooks, A., Zhu, J. & Haussler, D. The UCSC Xena Platform for Cancer Genomics Data Visualization and Interpretation. *BioRxiv,* 326470 (2018) (cit. on p. 43).

124. Consortium, E. P. The ENCODE (ENCyclopedia of DNA Elements) Project. *Science* **306,** 636–640. ISSN: 0036-8075. doi:10.1126/science.1105136 (2004) (cit. on p. 43).

125. Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., *et al.* TRRUST v2: An Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions. *Nucleic acids research* **46,** D380–D386. ISSN: 0305-1048. doi:10.1093/nar/gkx1013 (2017) (cit. on p. 43).

126. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., *et al.* Tissue-Specific Regulatory Circuits Reveal Variable Modular Perturbations across Complex Diseases. *Nature methods* **13,** 366. ISSN: 1548-7105. doi:10.1038/nmeth.3799 (2016) (cit. on p. 43).

127. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of statistical software* **39,** 1. doi:10.18637/jss.v039.i05 (2011) (cit. on p. 44).

128. Narayanan, S. Sialic Acid as a Tumor Marker. *Annals of Clinical & Laboratory Science* **24,** 376–384. ISSN: 0091-7370 (1994) (cit. on p. 48).

129. Jan, K.-M. & Chien, S. Role of Surface Electric Charge in Red Blood Cell Interactions. *Journal of General Physiology* **61,** 638–654. ISSN: 0022-1295. doi:10.1085/jgp.61.5.638 (May 1, 1973) (cit. on p. 48).

130. Albohy, A., Li, M. D., Zheng, R. B., Zou, C. & Cairo, C. W. Insight into Substrate Recognition and Catalysis by the Human Neuraminidase 3 (NEU_3) through Molecular Modeling and Site-Directed Mutagenesis. *Glycobiology* **20,** 1127–38. ISSN: 1460-2423 (Electronic) 0959-6658 (Linking). doi:10.1093/glycob/cwq077 (2010) (cit. on p. 48).

131. Oschlies, M., Dickmanns, A., Haselhorst, T., Schaper, W., Stummeyer, K., *et al.* A C-terminal Phosphatase Module Conserved in Vertebrate CMP-sialic Acid Synthetases Provides a Tetramerization Interface for the Physiologically Active Enzyme. *J Mol Biol* **393,** 83–97. ISSN: 1089-8638 (Electronic) 0022-2836 (Linking). doi:10.1016/j.jmb.2009.08.003 (2009) (cit. on p. 48).

132. Yuyama, Y., Dohi, T., Morita, H., Furukawa, K. & Oshima, M. Enhanced Expression of GM2/GD2 Synthase mRNA in Human Gastrointestinal Cancer. *Cancer* **75,** 1273–80. ISSN: 0008-543X (Print) 0008-543x. doi:10.1002/1097-0142(19950315)75:6<1273::aid-cncr2820750609>3.0.co;2-o (1995) (cit. on p. 51).

133. Cheresh, D. A., Pytela, R., Pierschbacher, M. D., Klier, F. G., Ruoslahti, E., *et al.* An Arg-Gly-Asp-directed Receptor on the Surface of Human Melanoma Cells Exists in an Divalent Cation-Dependent Functional Complex with the Disialoganglioside GD2. *The Journal of cell biology* **105,** 1163–1173. ISSN: 0021-9525. doi:10.1083/jcb.105.3.1163 (1987) (cit. on p. 51).

134. Paller, A. S., Arnsmeier, S. L., Chen, J. D. & Woodley, D. T. Ganglioside GT1b Inhibits Keratinocyte Adhesion and Migration on a Fibronectin Matrix. *Journal of investigative dermatology* **105,** 237–242. ISSN: 0022-202X. doi:10.1111/1523-1747.ep12317572 (1995) (cit. on p. 51).

135. Yu, R. K., Tsai, Y. T., Ariga, T. & Yanagisawa, M. Structures, Biosynthesis, and Functions of Gangliosides–an Overview. *J Oleo Sci* **60,** 537–44. ISSN: 1345-8957. doi:10.5650/jos.60.537 (2011) (cit. on p. 53).

136. Sonnino, S., Mauri, L., Chigorno, V. & Prinetti, A. Gangliosides as Components of Lipid Membrane Domains. *Glycobiology* **17,** 1R–13R. ISSN: 0959-6658 (Print) 0959-6658 (Linking). doi:10.1093/glycob/cwl052 (2007) (cit. on p. 54).

137. Nielsen, J. Systems Biology of Metabolism. *Annu. Rev. Biochem.* **86,** 245–275. ISSN: 0066-4154, 1545-4509. doi:10.1146/annurev-biochem-061516-044757 (June 20, 2017) (cit. on p. 64).

138. Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., *et al.* Recon3D Enables a Three-Dimensional View of Gene Variation in Human Metabolism. *Nat Biotechnol* **36,** 272–281. ISSN: 1546-1696. doi:10.1038/nbt.4072 (3 Mar. 2018) (cit. on p. 65).

139. Robinson, J. L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., *et al.* An Atlas of Human Metabolism. *Science Signaling* **13,** eaaz1482. doi:10.1126/scisignal.aaz1482 (Mar. 24, 2020) (cit. on pp. 65, 67, 74).

140. Schreiber, F., Bader, G. D., Gleeson, P., Golebiewski, M., Hucka, M., *et al.* Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2017. *J Integr Bioinform* **15,** 20180013. ISSN: 1613-4516. doi:10.1515/jib-2018-0013 (Mar. 29, 2018) (cit. on p. 65).

141. SBML Level 3: An Extensible Format for the Exchange and Reuse of Biological Models. *Molecular Systems Biology* **16,** e9110. ISSN: 1744-4292. doi:10.15252/msb.20199110 (Aug. 2020) (cit. on p. 65).

142. Hucka, M. & Smith, L. P. SBML Level 3 Package: Groups, Version 1 Release 1. *J Integr Bioinform* **13,** 290. ISSN: 1613-4516. doi:10.2390/biecoll-jib-2016-290 (Dec. 18, 2016) (cit. on p. 65).

143. Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., *et al.* The MetaCyc Database of Metabolic Pathways and Enzymes - a 2019 Update. *Nucleic Acids Research* **48,** D445–D453. ISSN: 0305-1048. doi:10/ghg52p (Jan. 8, 2020) (cit. on pp. 66, 74).

144. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2. *Genome Biology* **15,** 550. ISSN: 1474-760X. doi:10.1186/s13059-014-0550-8 (Dec. 5, 2014) (cit. on pp. 67, 79).

145. Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. LibSBML: An API Library for SBML. *Bioinformatics* **24,** 880–881. ISSN: 1367-4803. doi:10.1093/bioinformatics/btn051 (Mar. 15, 2008) (cit. on p. 67).

146. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI Gene-Centered Resources. *Nucleic Acids Res* **29,** 137–140. ISSN: 0305-1048. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC29787/ (2022) (Jan. 1, 2001) (cit. on p. 68).

147. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., *et al.* ChEBI: A Database and Ontology for Chemical Entities of Biological Interest. *Nucleic Acids Research* **36,** D344–D350. ISSN: 0305-1048. doi:10.1093/nar/gkm791 (suppl_1 Jan. 1, 2008) (cit. on pp. 68, 69).

148. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., *et al.* PubChem Substance and Compound Databases. *Nucleic Acids Res* **44,** D1202–D1213. ISSN: 0305-1048. doi:10.1093/nar/gkv951 (Database issue Jan. 4, 2016) (cit. on pp. 68, 69).

149. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Research* **49,** D480–D489. ISSN: 0305-1048. doi:10.1093/nar/gkaa1100 (Jan. 8, 2021) (cit. on p. 68).

150. Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., *et al.* Ensembl 2022. *Nucleic Acids Research* **50,** D988–D995. ISSN: 0305-1048. doi:10.1093/nar/gkab1049 (Jan. 7, 2022) (cit. on p. 69).

151. Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M., *et al.* Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Data Sets. *J. Chem. Inf. Model.* **52,** 2884–2901. ISSN: 1549-9596. doi:10.1021/ci300261r (Nov. 26, 2012) (cit. on p. 72).

152. Balaur, I., Mazein, A., Saqi, M., Lysenko, A., Rawlings, C. J., *et al.* Recon2Neo4j: Applying Graph Database Technologies for Managing Comprehensive Genome-Scale Networks. *Bioinformatics* **33,** 1096–1098. ISSN: 1367-4803. doi:10.1093/bioinformatics/btw731 (Apr. 1, 2017) (cit. on p. 76).

153. Swainston, N., Batista-Navarro, R., Carbonell, P., Dobson, P. D., Dunstan, M., *et al.* Biochem4j: Integrated and Extensible Biochemical Knowledge through Graph Databases. *PLOS ONE* **12,** e0179130. ISSN: 1932-6203. doi:10.1371/journal.pone.0179130 (July 14, 2017) (cit. on p. 76).

154. Noh, H., Shoemaker, J. E. & Gunawan, R. Network Perturbation Analysis of Gene Transcriptional Profiles Reveals Protein Targets and Mechanism of Action of Drugs and Influenza A Viral Infection. *Nucleic Acids Research* **46,** e34. ISSN: 0305-1048. doi:10.1093/nar/gkx1314 (Apr. 6, 2018) (cit. on p. 77).

155. Sun, H., Zhou, Y., Jiang, H. & Xu, Y. Elucidation of Functional Roles of Sialic Acids in Cancer Migration. *Front Oncol* **10,** 401. ISSN: 2234-943X (Print) 2234-943X (Linking). doi:10.3389/fonc.2020.00401 (2020) (cit. on p. 78).

156. Liang, J.-X., Liang, Y. & Gao, W. Clinicopathological and Prognostic Significance of Sialyl Lewis X Overexpression in Patients with Cancer: A Meta-Analysis. *Onco Targets Ther* **9,** 3113–3125. ISSN: 1178-6930. doi:10.2147/OTT.S102389 (2016) (cit. on p. 80).

157. Spinelli, J. B., Yoon, H., Ringel, A. E., Jeanfavre, S., Clish, C. B., *et al.* Metabolic Recycling of Ammonia via Glutamate Dehydrogenase Supports Breast Cancer Biomass. *Science* **358,** 941–946. ISSN: 0036-8075. doi:10.1126/science.aam9305 (Nov. 17, 2017) (cit. on p. 80).

158. Li, L., Mao, Y., Zhao, L., Li, L., Wu, J., *et al.* P53 Regulation of Ammonia Metabolism through Urea Cycle Controls Polyamine Biosynthesis. *Nature* **567,** 253–256. ISSN: 1476-4687. doi:10.1038/s41586-019-0996-7 (Mar. 2019) (cit. on p. 80).

159. Khan, A. P., Rajendiran, T. M., Ateeq, B., Asangani, I. A., Athanikar, J. N., *et al.* The Role of Sarcosine Metabolism in Prostate Cancer Progression. *Neoplasia* **15,** 491–501. ISSN: 1522-8002. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638352/ (2022) (May 2013) (cit. on p. 80).

160. Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., *et al.* Metabolomic Profiles Delineate Potential Role for Sarcosine in Prostate Cancer Progression. *Nature* **457,** 910–914. ISSN: 1476-4687. doi:10.1038/nature07762 (7231 Feb. 2009) (cit. on p. 80).

Chapter A

Supplement for Chapter 2

Supplementary results, Figures S1–38 and Tables S25–S28 can be found online at the Supplementary data section of the paper. Tables S1–S25 can be found online at the Supplementary tables section of the paper.

Chapter B

Supplement for Chapter 3

Figures S1 and S2 can be found in the Supplementary Material section of the online paper. Alternatively, the file can be downloaded directly using this link.

Chapter C

Supplement for Chapter 4

Table C.1: Reprogrammed ammonia metabolism routes identified by Metabolike. Scores colored red and blue represent repressed and activated routes, respectively. Metabolic routes are ordered by the absolute values of the scores for each cancer type.

Cancer type	Score	Genes	Metabolic route
BLCA	-13.69650049	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methylenetetrahydrofolate -> a reduced electron-transfer flavoprotein
BLCA	-13.69650049	DMGDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
BLCA	-13.69650049	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine
BLCA	-13.69650049	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate
BLCA	-11.62670482	DMGDH, PIPOX	ammonium -> a tetrahydrofolate -> sarcosine -> formaldehyde -> glycine
BLCA	-11.59424351	DMGDH, MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
BLCA	-9.855040385	GTPBP3, DMGDH, MTO1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-taurinomethyluridine in tRNA -> a 7,8-dihydrofolate
BLCA	9.739845282	NIT2	ammonium -> (indol-3-yl)acetonitrile -> ammonium -> (indol-3-yl)acetate
BLCA	9.739845282		ammonium -> (indol-3-yl)acetonitrile
BLCA	9.739845282	NIT2	ammonium -> (indol-3-yl)acetonitrile -> (indol-3-yl)acetate
BLCA	-9.301688753	DMGDH, SARDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methenyltetrahydrofolate -> glycine -> a reduced electron-transfer flavoprotein
BLCA	-9.13100033	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> an oxidized ferredoxin [iron-sulfur] cluster -> a 5-methyltetrahydrofolate
BLCA	8.465182765	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
BLCA	8.17419121	SHMT2, GSS, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> ophthalmate
BLCA	8.17419121	SHMT2, GSS, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glutathione

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
BLCA	-8.131799716	DMGDH, MTHFD2L, MTHFD1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5,10-methenyltetrahydrofolate
BLCA	8.121593328	CPS1, CAD	ammonium -> carbamate
BLCA	8.121593328	CPS1, CAD	ammonium -> carbamoyl phosphate
BLCA	8.121593328	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
BLCA	8.121593328	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
BRCA	-86.02541445	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine
BRCA	-86.02541445	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methylenetetrahydrofolate -> a reduced electron-transfer flavoprotein
BRCA	-86.02541445	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate
BRCA	-86.02541445	DMGDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
BRCA	-57.69153125	GTPBP3, DMGDH, MTO1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-taurinomethyluridine in tRNA -> a 7,8-dihydrofolate
BRCA	-57.65758564	DMGDH, MTHFD2L, MTHFD1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5,10-methenyltetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
BRCA	-57.3502763	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> an oxidized ferredoxin [iron-sulfur] cluster -> a 5-methyltetrahydrofolate
BRCA	-57.06629057	DMGDH, MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
BRCA	-54.79339727	DMGDH, PIPOX	ammonium -> a tetrahydrofolate -> sarcosine -> formaldehyde -> glycine
BRCA	-53.432034	DMGDH, SARDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methenyltetrahydrofolate -> glycine -> a reduced electron-transfer flavoprotein
BRCA	43.39946263	GLUL, IL4I1	ammonium -> L-glutamine -> 2-oxoglutaramate -> ammonium
BRCA	43.39946263	GLUL, IL4I1	ammonium -> L-glutamine -> 2-oxoglutaramate
BRCA	38.00906396	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
BRCA	37.80249787	SHMT2, IL4I1, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glyoxylate -> ammonium
BRCA	37.63823094	MTHFD2, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5,10-methenyltetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
BRCA	29.04955211	GNPNAT1	ammonium -> alpha-D-glucosamine 6-phosphate -> N-acetyl-D-glucosamine 6-phosphate
BRCA	27.96883796		ammonium -> alpha-D-glucosamine 6-phosphate
BRCA	27.96883796	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate
BRCA	27.96883796	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate -> ammonium
BRCA	27.68214333	ASNS	ammonium -> L-asparagine
COAD	87.94875365	MTHFD1L, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate
COAD	70.85229032	MTHFD1L, GART, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> N2-formyl-N1-(5-phospho-beta-D-ribosyl)glycinamide -> a tetrahydrofolate
COAD	70.83951711	MTHFD1L, ATIC, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> 5-formamido-1-(5-phospho-D-ribosyl)-imidazole-4-carboxamide
COAD	58.85659774	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> formate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
COAD	58.85659774	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate
COAD	44.12152467	CPS1, CAD	ammonium -> carbamate
COAD	44.12152467	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
COAD	44.12152467	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
COAD	44.12152467	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
COAD	44.12152467	CPS1, CAD	ammonium -> carbamoyl phosphate
COAD	-9.563560627	DMGDH, MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
COAD	-9.211524276	MTHFR, AMT	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
COAD	-7.281050705	MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
COAD	-6.796562765	AMT, DLD	ammonium -> a tetrahydrofolate -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine -> a [glycine-cleavage complex H protein] N6-[(R)-lipoyl]-L-lysine

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
COAD	-6.319475239	DMGDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
COAD	-6.319475239	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methylenetetrahydrofolate -> a reduced electron-transfer flavoprotein
COAD	-6.319475239	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate
COAD	-6.319475239	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine
COAD	-5.791420712	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
COAD	-5.791420712	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a 5,10-methylenetetrahydrofolate
ESCA	-9.10811642	GNMT, SARDH	ammonium -> a tetrahydrofolate -> glycine -> sarcosine
ESCA	-8.505866624	GLUL	ammonium -> L-glutamine
ESCA	-7.305653044	GNMT, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> sarcosine
ESCA	6.008788202		ammonium -> alpha-D-glucosamine 6-phosphate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
ESCA	6.008788202	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate -> ammonium
ESCA	6.008788202	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate
ESCA	-5.923136	AMT	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> ammonium -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
ESCA	-5.923136		ammonium -> a tetrahydrofolate
ESCA	-5.923136		ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine
ESCA	-5.923136	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
ESCA	-5.923136	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a 5,10-methylenetetrahydrofolate -> ammonium -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
ESCA	-5.923136	AMT	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
ESCA	-5.923136	AMT	ammonium -> a tetrahydrofolate -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
ESCA	3.016468867	GNPNAT1	ammonium -> alpha-D-glucosamine 6-phosphate -> N-acetyl-D-glucosamine 6-phosphate
ESCA	2.794810163		ammonium -> (indol-3-yl)acetonitrile
ESCA	2.794810163	NIT2	ammonium -> (indol-3-yl)acetonitrile -> (indol-3-yl)acetate
ESCA	2.794810163	NIT2	ammonium -> (indol-3-yl)acetonitrile -> ammonium -> (indol-3-yl)acetate
ESCA	2.694072884		ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate
ESCA	2.694072884		ammonium -> alpha-D-glucosamine 6-phosphate -> L-glutamine
ESCA	2.503232702	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
HNSC	19.8976681 5		ammonium -> alpha-D-glucosamine 6-phosphate
HNSC	19.8976681 5	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate -> ammonium
HNSC	19.8976681 5	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
HNSC	-17.49779853	GLUL	ammonium -> L-glutamine -> L-ornithine
HNSC	-17.49779853	GLUL	ammonium -> L-glutamine -> 2-oxoglutaramate
HNSC	-14.67481455	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine
HNSC	-14.67481455	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate
HNSC	-14.67481455	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methylenetetrahydrofolate -> a reduced electron-transfer flavoprotein
HNSC	-14.67481455	DMGDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
HNSC	13.07705459	GNPNAT1	ammonium -> alpha-D-glucosamine 6-phosphate -> N-acetyl-D-glucosamine 6-phosphate
HNSC	-12.4329926	DMGDH, MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
HNSC	12.21926951	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
HNSC	12.21926951	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
HNSC	12.21926951	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
HNSC	12.21926951	CPS1, CAD	ammonium -> carbamoyl phosphate
HNSC	12.21926951	CPS1, CAD	ammonium -> carbamate
HNSC	11.84195482		ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
HNSC	-10.09676317	GTPBP3, DMGDH, MTO1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-taurinomethyluridine in tRNA -> a 7,8-dihydrofolate
HNSC	-9.7832097	DMGDH	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> an oxidized ferredoxin [iron-sulfur] cluster -> a 5-methyltetrahydrofolate
HNSC	-9.21056281	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> N-(ADP-D-ribosyl)diphthamide-[translation elongation factor 2]
KICH	-44.3427391	FTCD	ammonium -> a tetrahydrofolate -> 5-formimidoyltetrahydrofolate -> a 5,10-methenyltetrahydrofolate -> ammonium
KICH	-39.14437506	DMGDH, PIPOX	ammonium -> a tetrahydrofolate -> sarcosine -> formaldehyde -> glycine
KICH	-33.98212833	FTCD	ammonium -> a tetrahydrofolate -> 5-formimidoyltetrahydrofolate
KICH	-23.94640628	ASS1	ammonium -> L-aspartate -> L-arginino-succinate
KICH	20.38889687	ASNS	ammonium -> L-asparagine
KICH	-17.42448274	ASS1, ASL	ammonium -> L-aspartate -> L-arginino-succinate -> L-arginine -> fumarate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
KICH	16.35125335	ME3, ME2, GOT1L1, ME1, GOT1	ammonium -> L-aspartate -> oxaloacetate -> pyruvate
KICH	-14.35417027	NIT2	ammonium -> (indol-3-yl)acetonitrile -> ammonium -> (indol-3-yl)acetate
KICH	-14.35417027	NIT2	ammonium -> (indol-3-yl)acetonitrile -> (indol-3-yl)acetate
KICH	-14.35417027		ammonium -> (indol-3-yl)acetonitrile
KICH	-13.34903634	DMGDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
KICH	-13.34903634	DMGDH	ammonium -> a tetrahydrofolate -> sarcosine
KICH	11.4493634	ME2, ME1, ASPDH, ME3	ammonium -> L-aspartate -> oxaloacetate -> pyruvate
KICH	10.19444843	ASNS	ammonium -> L-asparagine -> an L-asparaginyl-[tRNAAsn]
KICH	10.18761043	GATB, QRSL1	ammonium -> an L-glutaminyl-[tRNAGln]
KICH	9.614223062	DPH6	ammonium -> a diphthamide-[translation elongation factor 2]
KICH	9.145744606	GLUL, ASNS	ammonium -> L-glutamine -> L-asparagine

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
KICH	9.145744606	GLS2, GLUL, QRSL1, GMPS, CTPS2, GATB, ASNS, CAD, CTPS1, GLS	ammonium -> L-glutamine -> ammonium
KICH	8.373606352	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> ammonium -> L-aspartate
KICH	8.373606352	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> L-aspartate
KIRC	65.41641182	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
KIRC	-60.06566445	ASS1	ammonium -> L-aspartate -> L-arginino-succinate
KIRC	43.9354094	SHMT2, IL4I1, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glyoxylate -> ammonium
KIRC	42.0390531	SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate
KIRC	42.0390531	SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> glycine
KIRC	42.0390531	SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> a 5,10-methylenetetrahydrofolate
KIRC	-41.40004214	ASS1, ASL	ammonium -> L-aspartate -> L-arginino-succinate -> L-arginine -> fumarate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
KIRC	38.84154783	BAAT, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glycocholate
KIRC	38.84154783	BAAT, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glycochenodeoxycholate
KIRC	38.81777593	DPH6, NNMT, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> nicotinamide -> 1-methylnicotinamide
KIRC	-37.93205085		ammonium -> L-aspartate
KIRC	-37.93205085	ASPDH	ammonium -> L-aspartate -> ammonium -> oxaloacetate
KIRC	-37.93205085	ASPDH	ammonium -> L-aspartate -> oxaloacetate
KIRC	-37.93205085	ASPDH	ammonium -> L-aspartate -> oxaloacetate -> ammonium
KIRC	-35.14652048	GLDC	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a [glycine-cleavage complex H protein] N6-[(R)-lipoyl]-L-lysine -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine
KIRC	34.58550454	MTHFD2, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5,10-methenyltetrahydrofolate
KIRC	-34.50914492	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate -> IDP

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
KIRC	-34.50914492	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate
KIRC	33.94436091	GLYATL2, GLYATL1, SHMT1, SHMT2, GLYAT	ammonium -> a tetrahydrofolate -> glycine -> an N-acylglycine
KIRC	-32.87313534		ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a [glycine-cleavage complex H protein] N6-[(R)-lipoyl]-L-lysine
KIRP	-31.25009136	ASS1	ammonium -> L-aspartate -> L-arginino-succinate
KIRP	27.9078433	ASNS	ammonium -> L-asparagine
KIRP	-25.28538988	ASPDH	ammonium -> L-aspartate -> ammonium -> oxaloacetate
KIRP	-25.28538988	ASPDH	ammonium -> L-aspartate -> oxaloacetate
KIRP	-25.28538988		ammonium -> L-aspartate
KIRP	-25.28538988	ASPDH	ammonium -> L-aspartate -> oxaloacetate -> ammonium
KIRP	-23.35519931	ASS1, ASL	ammonium -> L-aspartate -> L-arginino-succinate -> L-arginine -> fumarate
KIRP	-21.9288574	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate -> IDP
KIRP	-21.9288574	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
KIRP	16.45989156	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> L-aspartate
KIRP	16.45989156	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> ammonium -> L-aspartate
KIRP	-16.27305902	ME2, ME1, ASPDH, ME3	ammonium -> L-aspartate -> oxaloacetate -> pyruvate
KIRP	-14.6990561	DMGDH, PIPOX	ammonium -> a tetrahydrofolate -> sarcosine -> formaldehyde -> glycine
KIRP	14.01028358	GLS2, GLUL, QRSL1, GMPS, CTPS2, GATB, ASNS, CAD, CTPS1, GLS	ammonium -> L-glutamine -> ammonium
KIRP	14.01028358	GLUL, ASNS	ammonium -> L-glutamine -> L-asparagine
KIRP	13.95392165	ASNS	ammonium -> L-asparagine -> an L-asparaginyl-[tRNAAsn]
KIRP	13.4366564	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
KIRP	13.4366564	CPS1, CAD	ammonium -> carbamate
KIRP	13.4366564	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
KIRP	13.4366564	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
LIHC	34.98614212	CPS1, CAD	ammonium -> carbamate
LIHC	34.98614212	CPS1, CAD	ammonium -> carbamoyl phosphate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
LIHC	34.98614212	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
LIHC	34.98614212	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
LIHC	34.98614212	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
LIHC	23.39020022	GLS2, GLUL, QRSL1, GMPS, CTPS2, GATB, ASNS, CAD, CTPS1, GLS	ammonium -> L-glutamine -> ammonium
LIHC	23.39020022	GLUL, CAD	ammonium -> L-glutamine -> carbamoyl phosphate
LIHC	20.17042636	CAD	ammonium -> L-aspartate -> N-carbamoyl-L-aspartate -> (S)-dihydroorotate
LIHC	17.49307106	CPS1, CAD	ammonium -> carbamate -> ammonium
LIHC	17.49307106	CPS1, CAD	ammonium -> carbamoyl phosphate -> cyanate
LIHC	-14.02047334	NIT2	ammonium -> (indol-3-yl)acetonitrile -> (indol-3-yl)acetate
LIHC	-14.02047334	NIT2	ammonium -> (indol-3-yl)acetonitrile -> ammonium -> (indol-3-yl)acetate
LIHC	-14.02047334		ammonium -> (indol-3-yl)acetonitrile
LIHC	-13.75270696	ASS1	ammonium -> L-aspartate -> L-arginino-succinate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
LIHC	-13.2713371	DMGDH, SARDH	ammonium -> a tetrahydrofolate -> sarcosine -> a 5,10-methenyltetrahydrofolate -> glycine -> a reduced electron-transfer flavoprotein
LIHC	-12.34141703	SARDH	ammonium -> a tetrahydrofolate -> glycine
LIHC	-12.34141703	SARDH	ammonium -> a tetrahydrofolate -> a 5,10-methenyltetrahydrofolate -> glycine -> a reduced electron-transfer flavoprotein
LIHC	-12.34141703	SARDH	ammonium -> a tetrahydrofolate -> a 5,10-methenyltetrahydrofolate
LIHC	-12.34141703	SARDH	ammonium -> a tetrahydrofolate -> a reduced electron-transfer flavoprotein
LIHC	-11.91861433	SARDH, GATM	ammonium -> a tetrahydrofolate -> glycine -> guanidinoacetate -> L-ornithine
LUAD	54.28315417	CPS1, CAD	ammonium -> carbamoyl phosphate
LUAD	54.28315417	CPS1, CAD	ammonium -> carbamate
LUAD	54.28315417	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
LUAD	54.28315417	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
LUAD	42.71638654	PAICS	ammonium -> L-aspartate -> 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole
LUAD	41.71040624	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
LUAD	35.96182848	GLUL, PPAT	ammonium -> L-glutamine -> 5-phospho-beta-D-ribosylamine

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
LUAD	35.41795163	ASNS	ammonium -> L-asparagine
LUAD	30.71731534	ADSL, PAICS	ammonium -> L-aspartate -> 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole -> fumarate -> 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide
LUAD	28.96722627	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
LUAD	-7.349012524	GLUL	ammonium -> L-glutamine
LUAD	-3.674506262	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate -> L-phenylalanine
LUAD	-3.674506262	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate -> L-methionine
LUAD	-3.674506262	GLUL, KYAT1	ammonium -> L-glutamine -> L-methionine
LUAD	-3.674506262	GLUL, KYAT1	ammonium -> L-glutamine -> L-phenylalanine
LUAD	-3.674506262	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate
LUAD	-3.674506262	GLUL	ammonium -> L-glutamine -> an L-glutaminyl-[tRNAGln]
LUAD	-3.607011166	ALAS1, ALAS2, SARDH	ammonium -> a tetrahydrofolate -> glycine -> 5-aminolevulinate
LUAD	-3.313883401	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> N-(ADP-D-ribosyl)diphthamide-[translation elongation factor 2] -> nicotinamide

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
LUAD	-3.31388340l	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> nicotinamide
LUSC	65.678614	TYMS, SHMT2, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 7,8-dihydrofolate -> dTMP
LUSC	60.21878916		ammonium -> ammeline
LUSC	60.21878916	GDA	ammonium -> ammeline -> ammonium -> ammelide
LUSC	60.21878916	GDA	ammonium -> ammeline -> ammelide
LUSC	58.05123397	CPS1, CAD	ammonium -> carbamate
LUSC	58.05123397	CPS1, CAD	ammonium -> carbamoyl phosphate
LUSC	58.05123397	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
LUSC	58.05123397	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate
LUSC	58.05123397	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
LUSC	57.93290136	ASNS	ammonium -> L-asparagine
LUSC	-15.99440098	MTHFR, AMT	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
LUSC	-15.25488001	DMGDH, MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
LUSC	-15.05289127	ALAS1, ALAS2, SARDH	ammonium -> a tetrahydrofolate -> glycine -> 5-aminolevulinate
LUSC	-14.9621472	DPH6, NNMT, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> nicotinamide -> 1-methylnicotinamide
LUSC	-14.80681156	MTHFR	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-methyltetrahydrofolate
LUSC	-13.05937119	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> N-(ADP-D-ribosyl)diphthamide-[translation elongation factor 2]
LUSC	-13.05937119	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> nicotinamide
LUSC	-13.05937119	DPH6, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> N-(ADP-D-ribosyl)diphthamide-[translation elongation factor 2] -> nicotinamide
LUSC	-9.368466381	GNMT, SARDH	ammonium -> a tetrahydrofolate -> glycine -> sarcosine
LUSC	-8.656	DPH6, NAMPT, ART1	ammonium -> a diphthamide-[translation elongation factor 2] -> nicotinamide -> beta-nicotinamide D-ribonucleotide
PRAD	16.57014667	NIT2	ammonium -> (indol-3-yl)acetonitrile -> (indol-3-yl)acetate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
PRAD	16.57014667	NIT2	ammonium -> (indol-3-yl)acetonitrile -> ammonium -> (indol-3-yl)acetate
PRAD	16.57014667		ammonium -> (indol-3-yl)acetonitrile
PRAD	-15.12160934	GLUL	ammonium -> L-glutamine -> L-ornithine
PRAD	-15.12160934	GLUL	ammonium -> L-glutamine -> 2-oxoglutaramate
PRAD	14.66855469	ASNS	ammonium -> L-asparagine
PRAD	-14.38904663	GLUL	ammonium -> L-glutamine
PRAD	-13.52691739	DMGDH, PIPOX	ammonium -> a tetrahydrofolate -> sarcosine -> formaldehyde -> glycine
PRAD	-12.5031884	SARDH, GATM	ammonium -> a tetrahydrofolate -> glycine -> guanidinoacetate -> L-ornithine
PRAD	11.60285974	ADSL, PAICS	ammonium -> L-aspartate -> 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole -> fumarate -> 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide
PRAD	10.76682369	PAICS	ammonium -> L-aspartate -> 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole
PRAD	9.953648707	GLYATL2, GLYATL1, SHMT1, SHMT2, GLYAT	ammonium -> a tetrahydrofolate -> glycine -> an N-acylglycine

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
PRAD	-9.363699668	GTPBP3, MTO1, AMT	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-taurinomethyluridine in tRNA -> a 7,8-dihydrofolate
PRAD	9.354808026	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> L-aspartate
PRAD	9.354808026	ASPG, ASNS, ASRGL1	ammonium -> L-asparagine -> ammonium -> L-aspartate
PRAD	9.214342609	MTHFD1, SHMT2, MTHFD2L, SHMT1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5,10-methenyltetrahydrofolate
PRAD	-9.191191586	GTPBP3, DMGDH, MTO1	ammonium -> a tetrahydrofolate -> a 5,10-methylenetetrahydrofolate -> a 5-taurinomethyluridine in tRNA -> a 7,8-dihydrofolate
PRAD	-8.535745313	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a 5,10-methylenetetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
PRAD	-8.535745313	AMT	ammonium -> a [glycine-cleavage complex H protein] N6-aminomethyldihydrolipoyl-L-lysine -> a 5,10-methylenetetrahydrofolate -> ammonium -> a [glycine-cleavage complex H protein] N6-dihydrolipoyl-L-lysine
PRAD	-8.535745313		ammonium -> a tetrahydrofolate
STAD	-30.3444446	ASPDH	ammonium -> L-aspartate -> oxaloacetate
STAD	-30.3444446		ammonium -> L-aspartate
STAD	-30.3444446	ASPDH	ammonium -> L-aspartate -> oxaloacetate -> ammonium
STAD	-30.3444446	ASPDH	ammonium -> L-aspartate -> ammonium -> oxaloacetate
STAD	-23.91514997	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate
STAD	-23.91514997	PCK1, PCK2, ASPDH	ammonium -> L-aspartate -> oxaloacetate -> phosphoenolpyruvate -> IDP
STAD	22.80810008	MTHFD1L, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate
STAD	21.05705284	CPS1, CAD	ammonium -> carbamate -> carbamoyl phosphate
STAD	21.05705284	CPS1, CAD	ammonium -> carbamate
STAD	21.05705284	CPS1, CAD	ammonium -> carbamoyl phosphate -> carbamate
STAD	21.05705284	CPS1, CAD	ammonium -> carbamoyl phosphate -> N-carbamoyl-L-aspartate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
STAD	21.05705284	CPS1, CAD	ammonium -> carbamoyl phosphate
STAD	-20.99249996	ME2, ME1, ASPDH, ME3	ammonium -> L-aspartate -> oxaloacetate -> pyruvate
STAD	19.87776577	MTHFD1L, ATIC, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> 5-formamido-1-(5-phospho-D-ribosyl)-imidazole-4-carboxamide
STAD	19.25171725	MTHFD1L, GART, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> N2-formyl-N1-(5-phospho-beta-D-ribosyl)glycinamide -> a tetrahydrofolate
STAD	-17.44068162	GLUL	ammonium -> L-glutamine
STAD	15.53249255	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> formate
STAD	15.53249255	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate
STAD	-15.17891208	GOT1L1, GOT1	ammonium -> L-aspartate -> oxaloacetate
STAD	-15.1722223		ammonium -> L-aspartate -> adenylo-succinate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
THCA	34.84093207	MTHFD1L, ATIC, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> 5-formamido-1-(5-phospho-D-ribosyl)-imidazole-4-carboxamide
THCA	33.13883417	MTHFD1L, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate
THCA	-31.35315952	GLUL	ammonium -> L-glutamine
THCA	30.60933794		ammonium -> alpha-D-glucosamine 6-phosphate
THCA	30.60933794	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate
THCA	30.60933794	GNPDA1, GNPDA2	ammonium -> alpha-D-glucosamine 6-phosphate -> beta-D-fructofuranose 6-phosphate -> ammonium
THCA	-23.05115202	GLUL	ammonium -> L-glutamine -> 2-oxoglutaramate
THCA	-23.05115202	GLUL	ammonium -> L-glutamine -> L-ornithine
THCA	22.21508456	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate -> formate
THCA	22.21508456	ALDH1L2, MTHFD1L, ALDH1L1, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> a tetrahydrofolate

Table C.1 continued from previous page

Cancer type	Score	Genes	Metabolic route
THCA	22.07293911	MTHFD1L, GART, MTHFD1	ammonium -> a tetrahydrofolate -> an N10-formyltetrahydrofolate -> N2-formyl-N1-(5-phospho-beta-D-ribosyl)glycinamide -> a tetrahydrofolate
THCA	19.48677896	SHMT2, GSS, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> ophthalmate
THCA	19.48677896	SHMT2, GSS, SHMT1	ammonium -> a tetrahydrofolate -> glycine -> glutathione
THCA	-15.68251548	GLUL, PPAT	ammonium -> L-glutamine -> 5-phospho-beta-D-ribosylamine
THCA	-15.67657976	GLUL, KYAT1	ammonium -> L-glutamine -> L-methionine
THCA	-15.67657976	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate
THCA	-15.67657976	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate -> L-phenylalanine
THCA	-15.67657976	GLUL, KYAT1	ammonium -> L-glutamine -> 2-oxoglutaramate -> L-methionine
THCA	-15.67657976	GLUL, KYAT1	ammonium -> L-glutamine -> L-phenylalanine
THCA	-15.67657976	GLUL	ammonium -> L-glutamine -> an L-glutaminyl-[tRNAGln]