Ruprecht-Karls-Universität Heidelberg
Fakultät für Biowissenschaften
Bachelorstudiengang Molekulare Biotechnologie

Extracting mechanistic insights from multi-omics data of lung cancer using the method COSMOS

Bachelorarbeit

Von **Dustin Schilling**
aus **Heilbronn**

27. Juli 2022

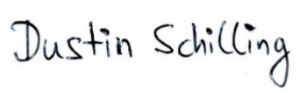
Die vorliegende Bachelorarbeit wurde im Institute for Computational Biomedicine der Medizinischen Fakultät der Universität Heidelberg in der Zeit von 16.05.2022 bis 27.07.2022 angefertigt.

Gutachter der Arbeit:	Prof. Dr. Ing. Julio Saez-Rodriguez Head of Saez-Rodriguez group Director of Institute for Computational Biomedicine

Ich erkläre hiermit ehrenwörtlich, dass:

1. ich die vorliegende Bachelorarbeit selbständig unter Anleitung verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;
2. die Übernahme wörtlicher Zitate aus der Literatur/Internet sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet wurde;
3. ich meine Bachelorarbeit bei keiner anderen Prüfung vorgelegt habe.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Ort, Datum	Unterschrift
Heidelberg, den 27.07.2022	

Acknowledgments

I want to take this opportunity to thank you, Aurélien, for the great guidance during my bachelor internship. All the advice has been extremely helpful and will certainly be kept in mind for the future! I appreciate the time you invested for me.

I would also like to thank Prof. Dr. Saez-Rodriguez for giving me the chance to write my bachelor thesis on this interesting matter. Additionally, I am grateful for the friendly welcome by the work group during this brief time.

Zudem möchte ich meiner Familie danken, die mir zu jedem Zeitpunkt von Herzen das Beste wünscht. Dabei gilt mein Dank auch meinen Freunden aus meiner Heimat und Heidelberg, ich blicke auf weitere erlebnisvolle Studienjahre voraus.

Abstract

The method COSMOS (Causal Oriented Search of Multi-Omics Space) is a multi-omics analysis approach for generating mechanistic insights into diseases and experimental measurements. It leverages prior knowledge to integrate multiple omics data sources, such as proteomics, metabolomics and transcriptomics. Prior knowledge of COSMOS is given by the combination of multiple curated databases that involve interactions of biomolecules within and between layers of signaling, metabolism and gene regulation. Based on these interactions, the computational tool CARNIVAL constructs causal links between highly deregulated players and displays their activity states in the form of networks. Networks enable the extraction of hypothetical disease processes, which need to be validated experimentally. COSMOS aims to facilitate the search for therapeutical targets and clinical biomarkers. However, limitations of COSMOS remain, as not all regulatory interactions are known and input data is incomplete.

This study applies COSMOS to differentially abundant proteins and metabolites of a lung adenocarcinoma cohort consisting of ten patients. Additionally, a pathway enrichment analysis is conducted on the hallmark gene set collection to focus on key metabolite and protein actors within the generated COSMOS networks. This work unveiled a crosstalk of dysfunctional tumor suppressor TP53 to increased SLC38A2 expression, a solute-carrier protein. This leads to elevated metabolism regarding the uptake of amino acids. Up-regulation of SLC38A2 poses a possible target for drug intervention. Furthermore, it presents good diagnostical value in detecting the malignancy via a radiotracer named [^{11}C]-MeAIB. In addition, COSMOS highlights the oncogenic activation of MYC due to upstream casein kinase II. This suggests the application of casein kinase specific allosteric inhibitors on lung adenocarcinoma.

Zusammenfassung

Die Methode COSMOS (engl. Causal Oriented Search of Multi-Omics Space) dient zur Gewinnung mechanistischer Erkenntnisse über Krankheiten und beschreibt eine Multi-Omics Analyse. COSMOS nutzt Vorwissen, um mehrere Omics-Datensätze, wie Proteomik, Metabolomik und Transkriptomik, miteinander zu integrieren. Das Vorwissen von COSMOS beruht auf mehreren stets aktualisierten Datenbanken, welche Interaktionen innerhalb und zwischen Signalübertragung, Stoffwechsel und Generegulation abbilden. Auf der Grundlage dieser Wechselwirkungen konstruiert das Rechenprogramm CARNIVAL kausale Verbindungen zwischen stark deregulierten Biomolekülen und stellt ihre Aktivitäten innerhalb des Prozess-Weg in Form von Netzwerken dar. Diese Netzwerke ermöglichen die Extraktion von hypothetischen Krankheitsprozessen, welche jedoch experimentell validiert werden müssen. Dadurch vereinfacht COSMOS die Suche nach therapeutisch angreifbaren Molekülen und Biomarkern für die Anwendung in der Medizin. COSMOS weist dennoch Limitationen auf, welche einerseits den Fakt betreffen, dass nicht alle regulatorischen Zusammenhänge bekannt sind. Andererseits sind angewandte Daten nicht vollständig.

In dieser Studie wird eine Kohorte von zehn Patienten, die an einem Adenokarzinom der Lunge erkrankt sind, untersucht. Dabei wird COSMOS auf stark differentiell vorhandene Proteine und Metabolite angewandt. Zudem wird eine Analyse der stark angereicherten Signalwege (pathway enrichment analysis) durchgeführt. Diese zeigt Schlüssel-Metabolite und -Proteine auf, deren Rolle in den COSMOS-Netzwerken hervorgehoben wird. Die vorliegende Arbeit enthüllt eine Wechselwirkung zwischen dem dysfunktionalen Tumorsuppressor TP53 und erhöhter Expression des Transportporteins SLC38A2. Dies resultiert in erhöhtem Metabolismus in Bezug auf verstärkter Aufnahme von Aminosäuren. Stärkere Expression von Transportproteinen ermöglicht therapeutisches Eingreifen. In Verbindung mit dem radioaktiven Tracer [^{11}C]-MeAIB bieten die Transportproteine gute diagnostische Erkennung des Tumors. COSMOS zeigte zusätzlich eine onkogene Aktivierung von MYC an, die durch eine Casein-Kinase verursacht werden könnte. Dies deutet auf eine potenzielle therapeutische Wirkung von spezifisch gegen Casein-Kinase gerichtete allosterische Inhibitoren bei Adenokarzinomen der Lunge hin.

Table of contents

List of abbreviations

AA	amino acid
AR	androgen receptor
autoSP3	single-pot solid-phase enhanced sample preparation using automated processing
CAD	carbamoyl-phosphate synthase
CARNIVAL	Causal Reasoning for Network identification using Integer VALue programming
COSMOS	Causal Oriented Search of Multi-Omics Space
CSNK2A1	alpha subunit of casein kinase II
EGFR	epidermal growth factor receptor
ILP	Integer Linear Programming
log-FC	logarithmic fold-change
LUAD	lung adenocarcinoma
MeAIB	α-methylaminoisobutyric acid
MS	mass-spectrometry
MTBE	methyl-tert-butylether
mTOR	mammalian target of rapamycin
norm-wmean	normalized weighted mean
NSCLC	non-small cell lung cancer
PEA	pathway enrichment analysis
PET	positron emission tomography
PKN	prior knowledge network
SLC	solute-carrier (transportprotein)
TF	transcription factor
UPR	unfolded protein response

List of figures and tables

1. Introduction

1.1. Advancements in multi-omics promise huge potential

In recent years, generating omics data, such as proteomics, metabolomics, and transcriptomics from single patients has become increasingly more cost-efficient and less time-consuming. Next-Generation-Sequencing methods, like RNA-seq, can be applied to determine the abundance of thousands of transcripts, while mass spectrometry (MS) approaches allow for the detection of hundreds of proteins or metabolites. Studying abnormal cellular mechanisms in diseases, particularly in cancer, requires tools that manage the high dimensionality of diverse omics data. In this respect, integrating at least two layers of biological regulations into a biomedical analysis refers to a “multi-omics analysis” (Hasin *et al.*, 2017).

Performing analyses across multiple omics layers aims to understand perturbed cellular mechanisms better. This consideration of the multi-omics space and the interconnection of different omics allows us to hypothesize about broader mechanisms that drive certain diseases. Thereby yielding greater molecular insights than each individual omic layer may achieve (Huang *et al.*, 2017). Thus, multi-omics analyses need to be reliable. This is especially important for oncology as it is crucial to precisely detect, predict the progression of cancers, and to identify suitable therapeutic targets.

1.2. COSMOS – a network-based multi-omics analysis approach

Due to the technological advancements in data sampling and already recorded successes in this field, efforts to incorporate multi-omics data into clinical research have increased (Krassowski *et al.*, 2020). The multi-omics field is just beginning to flourish. More and more multi-omics analysis approaches are emerging, which are expected to provide new insights into unknown pathology mechanisms of complex diseases. This could ultimately contribute to the development of new therapeutics for diseases that are difficult to treat. Furthermore, key actors of disrupted pathways identified by multi-omics analyses might serve as meaningful biomarkers for pathology identification and characterization (Yan *et al.*, 2018).

A current hurdle for gaining such mechanistic insights is that methods reducing the high dimensionality of multi-omics data still demand additional processing steps.

Therefore, a platform that allows efficient and direct extraction of hypotheses about disease mechanisms would facilitate multi-omics analysis. This study focuses on applying one such method, namely COSMOS (Causal Oriented Search of Multi-Omics Space), which was designed by Dugourd *et al.* (Dugourd *et al.*, 2021). COSMOS aims to investigate abnormal cellular mechanisms by including crosstalks between signaling, gene regulation and metabolism. The method describes a network-based multi-omics analysis approach that integrates several types of omics data into a causal network. It thereby creates an interconnection between multiple omics layers. This allows for the extraction of broader mechanistic insights into deregulated mechanisms of diseases (Dugourd *et al.*, 2021). In this study, a cohort of lung adenocarcinoma patients is analyzed.

To generate multi-omics networks, tools can either be data-driven, whereas other methods like COSMOS rely on prior information. For this, COSMOS leverages databases for the computation of network topologies. This so-called prior knowledge network (PKN) of COSMOS connects data from multiple resources to reassemble various complex interactions in the multi-omics space. In detail, these resources include protein-protein interactions, such as kinase to transcription factor (TF) or TF to metabolic enzymes. Furthermore, it contains interactions between proteins and metabolites, characterized by the enzymatic conversion of a metabolite or the allosteric regulation of an enzyme by a metabolite. The PKN features these causal links as signed and directed actions a molecule poses downstream, which can be activation or inhibition (Dugourd *et al.*, 2021). An overview of resources used for the PKN of COSMOS is shown in figure 1, which depicts possible biomolecular interactions of these databases.

The basic framework of COSMOS is a causal reasoning method named CARNIVAL (CAusal Reasoning for Network identification using Integer VALue programming). It leverages the PKN of COSMOS and identifies upstream regulatory paths. Using Integer Linear Programming (ILP), CARNIVAL tries to connect as many deregulated biological actors in the smallest network as technically feasible. For the contextualization of the networks, it poses an ILP optimization strategy for which it calculates a minimized fitting error and model size (Liu *et al.*, 2019). Specifically in COSMOS, CARNIVAL computes a "forward" run, which causally connects signaling nodes as input, such as kinases, phosphatases and TFs, to metabolites. On the

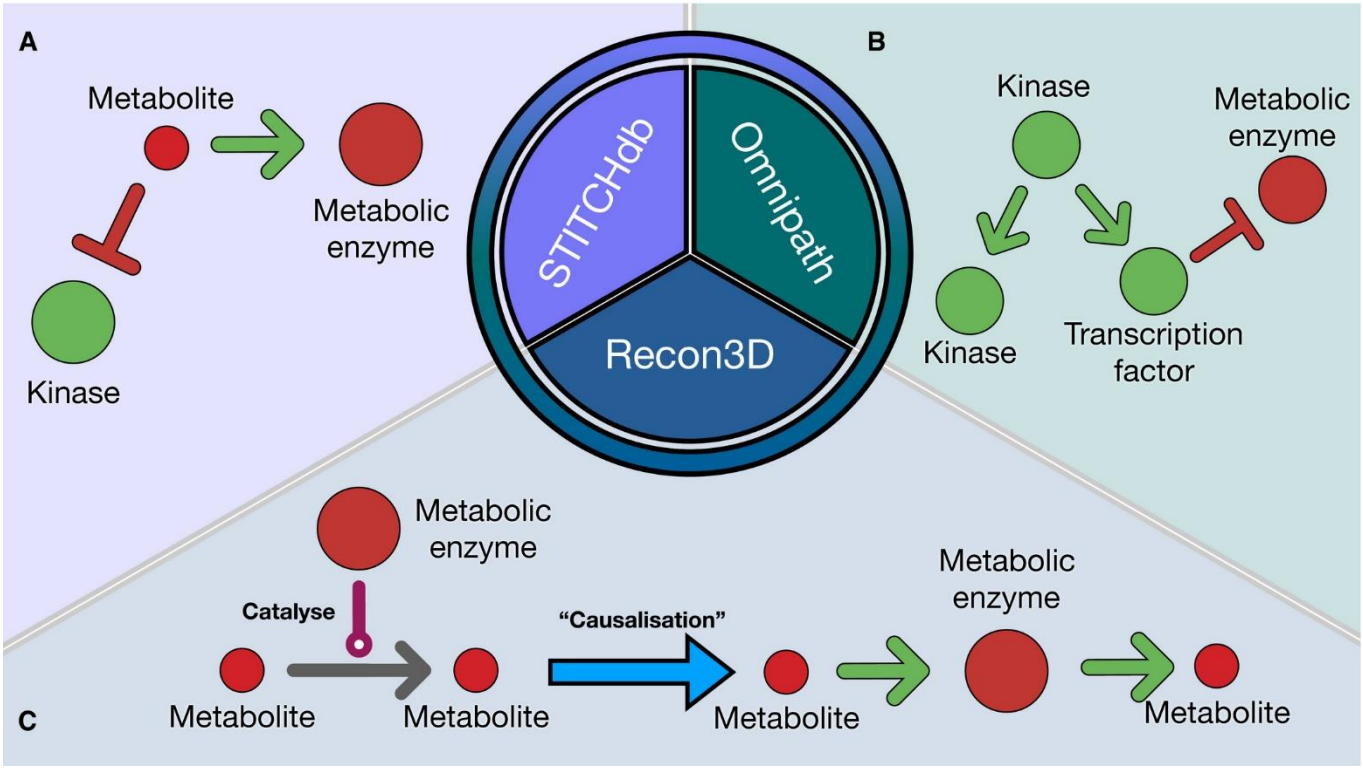


Figure 1: Overview of the collection of resources used in the PKN of COSMOS (figure taken from Dugourd *et al.*, 2021).
The PKN consists of three curated resources, all showing signed biomolecular interactions. Green arrow represents activation, while the red "T"-sign stands for inhibition. (**A**) STITCHdb displays signed protein-chemical interactions (Szklarczyk *et al.*, 2016). (**B**) OmniPath consists of protein-protein interactions, such as kinase-TF interactions in a signaling pathway (Türei *et al.*, 2016). (**C**) Recon3D represents interactions concerning human metabolism (Brunk *et al.*, 2018). The interactions were adapted by Dugourd *et al.*, so that the conversion of a metabolite is indicated by a reactant activating a metabolic enzyme which in turn up-regulates the product metabolite. This makes it clearly representable in networks.

contrary, a "backward" run calculates the connection of metabolites as input to upstream signaling nodes. For both runs, the biological input measure is linked to its upstream regulator in the pathway. The "forward" and "backward" are contextualized into a network. Additionally, the networks can be merged. As a result, the "full" network connects paths ranging from signaling nodes, such as kinases, to metabolites and *vice versa*. In general, computed networks represent causal connections as hypothetical interactions that need to be proven experimentally. The novelty of COSMOS is its ability to depict deregulated mechanisms in the multi-omics space. Consequently, insights into these formerly unknown mechanisms can be extracted, facilitating the search for biomolecular targets for treatment and disease prognosis (Dugourd *et al.*, 2021).

COSMOS requires some form of quantifiable information over deregulated processes to generate networks. Particularly suitable for this are changes in the

activities of kinases, phosphatases and TFs, as well as metabolite abundance changes (Dugourd *et al.*, 2021). These activities are inferred via footprint-based methods. In this respect, the term “activity” describes the abundance of molecules (footprints) considered to be directly downstream of the examined protein or enzyme (Dugourd and Saez-Rodriguez, 2019). However, any form of data depicting deregulation can be used as input for COSMOS. This study thereby utilizes the differential change in the abundance of proteins and metabolites of lung adenocarcinoma patients.

1.3. The pathology of lung adenocarcinoma

This study's subject is a cohort covering ten patients of stage I to II lung adenocarcinoma (LUAD). Among all cancers, lung cancer has the highest morbidity in men. Its prevalence is slightly lower for women, still having the third-highest cancer incidence. Thereby, lung cancer is the leading cause of cancer-related deaths worldwide (Travis, 2011). LUAD, classified as a sub-type of non-small lung cancer (NSCLC), makes up 40 % of all lung cancer cases in the United States. The pathology is strongly associated with smoking. However, LUAD is the most common sub-type for never smokers. Diagnosis of LUAD usually occurs late in life, with an average age of 71. Moreover, diagnosis usually becomes apparent at late stages of the disease, whereby prognosis is very poor due to frequent metastasis of the tumor. The 5-year survival rate is alarmingly low at about 15 % (Myers and Wallen, 2022).

Although advancements in the last two decades regarding treatment were achieved, the field strives to improve the clinical outcome of patients. Clinical progress is attributed to the development of various targeted therapies, as well as the application of immunotherapies for individual cases with late-stage NSCLC. Before, treatment of NSCLC consisted mainly of cytotoxic therapy, which is now mostly applied to advanced tumor stages. Furthermore, for NSCLC, radiotherapy is usually performed after the tumor has been surgically removed to reduce the risk of relapse. In this respect, treatment has shifted towards personalized approaches, regarding each case according to genetic alterations of the tumor. Discovering new potential biomarkers for predicting the course of the disease and early diagnosis and screening would be highly beneficial for applying the right treatment strategy. In addition, gaining a better understanding of deregulated mechanisms in LUAD facilities the search for novel therapeutical biomolecular targets. A remaining concern of the malignancy that needs

to be considered is intra-tumoral heterogeneity. This describes cell populations within the cancerous growth with distinct molecular features. In this respect, tumors with increased intra-tumoral heterogeneity are prone to relapse after surgical removal because they are more likely to develop metastasis early on. Furthermore, some subpopulations within the tumor may acquire resistance mechanisms against targeted therapies and become dominant (Herbst *et al.*, 2018).

Tyrosine kinase inhibitors form one branch of targeted therapies in LUAD. These target mutated EGFR, ALK and ROS1. Apart from the most common alteration concerning EGFR, another direct cause of tumorigenesis in LUAD is mutated KRAS. Moreover, in about half of all cases, the tumor suppressor protein p53, encoded by TP53, becomes dysfunctional (Herbst *et al.*, 2018). Aberrant p53 expression commonly starts in early stages of LUAD (Ahrendt *et al.*, 2003).

1.4. Pathway enrichment analysis

Due to the high dimensionality of multi-omics data, methodologies such as COSMOS offer high amounts of possible mechanistic hypotheses that can be extracted and acquired. Furthermore, the generated networks usually comprise many biological players. For this, setting the focus on distinct elements helps to guide the direction of network analysis. A fitting approach for identifying important biological actors and their involvement in deregulated processes is to conduct a pathway enrichment analysis (PEA). This portrays pathways that appear to be significantly dysfunctional within a specific disease (Khatri *et al.*, 2012). PEA leverages any measurement statistic, such as fold-changes or t-values, which displays differential expression between two conditions. Furthermore, based on this measurement, the enrichment analysis finds sets, in this case pathways, whose expression is more extreme. It thereby computes a score for each pathway indicating its up- or down-regulation for positive or negative values, respectively. As for other omics analyses, enrichment scores can be calculated in a data-driven manner or again by taking advantage of a PKN, which constitutes interactions of molecules in a pathway (Dugourd and Saez-Rodriguez, 2019).

1.5. Aim of this study

This study applies COSMOS to differentially abundant proteins and metabolites in LUAD. The analysis is purely exploratory. Hereby, broader mechanisms that have become dysfunctional in the subclass of lung cancer are depicted. PEA helps to identify key actors among the deregulated mechanisms, whose role is further investigated in subnetworks created by COSMOS. Alongside these biological actors, the respective mechanisms are subjected to the context of the disease. The subnetworks represent the multi-omics nature of LUAD, whereby important crosstalks between the metabolomic and proteomic layer can be shown. Finally, the potential use of important biomolecules as therapeutical targets or biomarkers will be discussed. The findings are evaluated considering therapeutical agents and biomarkers, which are already established or in clinical trials. The disease mechanisms involving these biological actors are compared with supporting or disproving evidence from the literature.

2. Methods

The PEA and COSMOS network generation are part of this work and applied scripts are available at: https://github.com/DustinSchilling/application_cosmos_bsc_thesis. Data generation and statistical testing were performed externally.

2.1. Acquisition of samples and the multi-omics data

Samples stem from a cohort of ten individuals with adenocarcinoma lung cancer ranging from stage I to II. At the point of surgical tissue removal, the patients had not yet received systemic therapy. Tumorous and adjacent non-tumorous tissue were removed for each patient. These samples have been acquired by the laboratories of Prof. Dr. Krijgsveld and Prof. Dr. Hell as part of collaborative research within the SMART-CARE consortium and respective data was allowed to be used for this study. Proteomic and metabolomic data were aliquoted from the same sample. Hereby, metabolomic data was separately sampled, in which a bi-phasic extraction with methyl-tert-butylether (MTBE) and 75% ethanol was performed. Afterward, metabolites and proteins were automatically processed using single-pot solid-phase enhanced sample preparation (autoSP3). Subsequently, the absolute abundances of the purified proteins and metabolites were analyzed via liquid chromatography-MS (Gegner *et al.*, 2022; Müller *et al.*, 2020).

2.2. Differentially abundance analysis

The quality of proteomic and metabolomic datasets was examined and filtered using the tool MatrixQCvis (Naake and Huber, 2021), whereby low-quality samples were removed. Additionally, the determined intensities of MS were log2-transformed. While for the proteomic data, 1679 proteins were available, the metabolomic dataset consists of only 405 metabolites.

On this data, a cohort-wide statistical test between all respective cancerous versus healthy samples has been performed using limma testing. The function *lmFit* of the LIMMA R package (version 3.50.1) achieves this by fitting multiple linear models in a least-squares approach. Afterward, moderated t-statistics were computed by empirical Bayes moderation (Ritchie *et al.*, 2015). This results in differentially abundant proteins and metabolites. To amount for higher False-Discovery-Rates in multiple testing, p-

values were adjusted according to the Benjamini-Hochberg-correction with an α-value of 0.05 (Benjamini and Hochberg, 1995).

2.3. Pathway enrichment analysis using weighted normalized means

The analysis platform of PEA was decoupleR (version 2.1.10). This tool offers a framework for applying enrichment analysis with the incorporation of prior knowledge (Badia-I-Mompel *et al.*, 2022). As prior knowledge, the hallmark gene set collection (v7.1) of MSigDB was used (Liberzon *et al.*, 2015). This comprises a collection of genes that constitute as part of 50 pathways conveying specific biological states and processes. For instance, members involved in cellular mechanisms, such as angiogenesis or apoptosis. The native hallmark set was used as prior knowledge to conduct PEA on the proteomics dataset. Additional measures had to be taken to be able to perform PEA on metabolomic data. For this, the PKN of COSMOS was leveraged to map interactions of members of the hallmarks set to metabolites. These metabolites, which are directly connected to genes of the hallmarks set, were then merged with the respective pathways. In the end, all added metabolites from the PKN of COSMOS and native genes of the hallmark set were assigned to a respective pathway. Afterward, the complete set was divided into metabolites and proteins only. These two separate sets were then used as prior knowledge for metabolomic and proteomic data, respectively.

With the consideration of prior knowledge, the statistical enrichment analysis algorithm normalized weighted mean (norm-wmean) was performed on the t-values of differentially abundant proteins and metabolites, separately. This estimates pathway scores representing the deregulation of each biological process.

2.4. Network generation via COSMOS

Differentially abundant proteins with an absolute t-value > 6 and differentially abundant metabolites with an absolute t-value > 3 were chosen as input for cosmosR (version 1.3.4). To adjust for the fact that more proteins are differentially abundant compared to metabolites, the t-value threshold for metabolites was chosen to be lower. Both thresholds account for proteins and metabolites with an adjusted p-value < 0.05 and are thereby considered deregulated.

Methods

CARNIVAL performs the network optimization via an ILP algorithm. Based on the PKN of COSMOS and the t-value input measures, sets of constraints are generated. The ILP solver IBM CPLEX (version 22.1.0.0) tries to find optimal solutions. Within the set of constraints, it minimizes the discrepancy between simulated values for kinases, TFs, and metabolites and the corresponding values acquired from differentially abundance analysis. Simultaneously, the size of the solution network is minimized. After a specified run time, it returns a pool of optimal solution networks. It averages the activity of each node, depending on the frequency the node was either not found at all or found to be up- or down-regulated in the pool of solution networks (Dugourd *et al.*, 2021). These averaged activities, alongside with the computed edges connecting deregulated proteins, TFs and metabolites, form the interactions and attributes of the solution network. In the end, networks are displayed using the open-source program cytoscape, commonly used to visualize biomolecular interactions (Shannon *et al.*, 2003).

The following table 1 gives an overview of the amounts of nodes and edges displayed in the networks, the solution run time for the ILP solver and the gap in the solution network. The parameter gap describes the deviation of the resulting network from the optimal in percent. It should be smaller than 5.0 % since high gap values attribute to incoherent network topologies. A further important parameter is the network depth. Across all generated networks, this was set to a maximum depth of 4, which means that CARNIVAL connects input measures within a maximum range of four steps downstream. Nodes that do not contribute any informational value, such as duplicated nodes, were removed in cytoscape. As part of the following discussion, subnetworks of either the “forward” or “backward” run are manually identified to decrease overall network size for visibility reasons. The complete network of the “forward” and the “backward” run can be seen in the appendix figure S1 and figure S2, respectively.

Table 1: Characteristics of generated networks used for subnetwork extraction.

network	nodes	edges	solution run time	gap
“forward” run	45	60	165.23 sec	0.00 %
“backward” run	133	181	14442.66 sec	0.49 %

3. Results

3.1. Enriched pathways in LUAD

LUAD data was examined in a PEA using statistical enrichment norm-wmean. The aim was to identify deregulated pathways, which might drive the disease. Furthermore, key actors of the respective pathways are later extracted and considered as important players in the analysis of networks generated with COSMOS. The pathways were ordered after their absolute enrichment score. The top five most deregulated pathways for metabolomics (left panel) and proteomics (right panel) of LUAD were selected and are shown in figure 2.

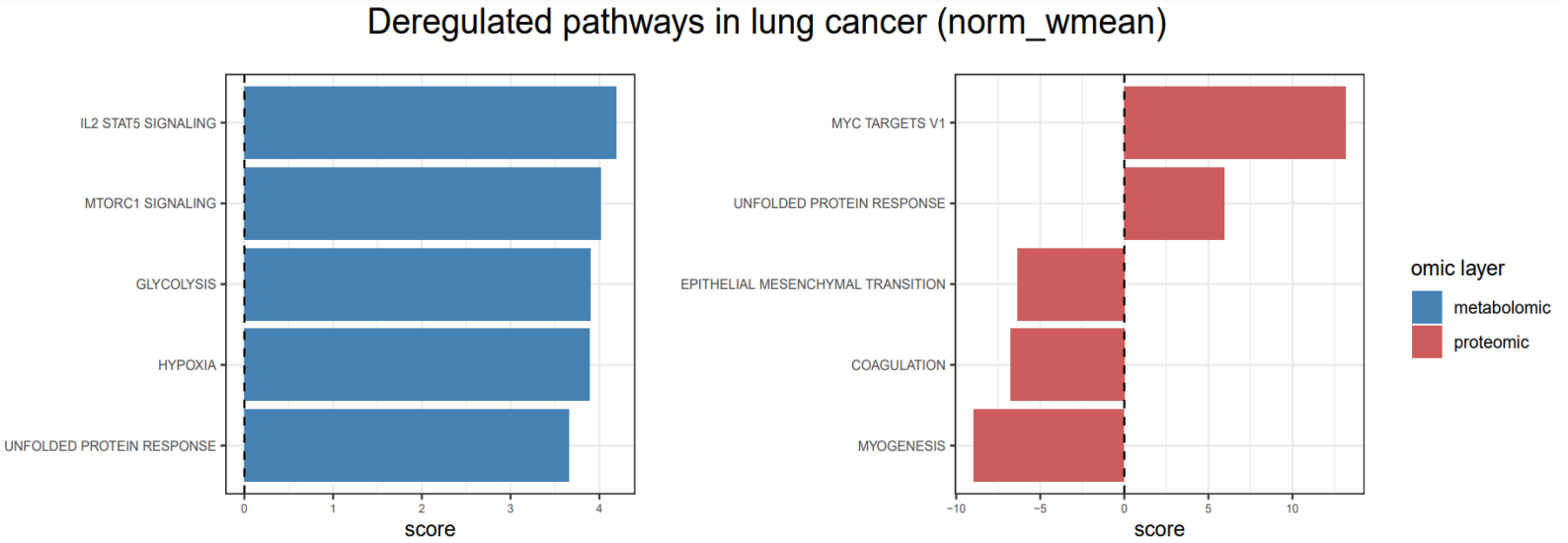


Figure 2: Most deregulated pathways in LUAD regarding the metabolomic and proteomic layer.
Pathway enrichment analysis has been performed on both proteomic and metabolomic data using the hallmark gene set collection of MSigDB (Liberzon *et al.*, 2015) as prior knowledge. As for the enrichment analysis algorithm, normalized weighted mean was applied. The left panel (blue) shows PEA conducted on metabolomic data. It is noticeable that metabolism is elevated for all out of the top five most deregulated pathways. One example is the up-regulated metabolism of glycolysis in LUAD in respect to healthy patients. The IL2-STAT5 signaling pathway thereby is the highest enriched pathway. The right panel (red) depicts the top five most deregulated pathways for proteomic data. Here, the gene set collection MYC targets V1, subgroup of genes regulated by MYC, forms the most deregulated pathway.

The PEA revealed that the biological process unfolded protein response (UPR) is elevated for both metabolomics and proteomics. According to the hallmark set annotation, this pathway depicts a cellular stress response. Deregulation, regarding the up- or down-regulation of pathways, is coherent across both omics layers in this case. This might also be traced back to the generation of the metabolomic prior knowledge set. Hereby, the PKN of COSMOS was used to identify metabolites that are directly linked to proteins of each respective pathway in the hallmark set. Thus,

PEA reflects the close relationship between signaling regulation and metabolism. Similar trends can also be observed for other pathways, whereby PEA indicates comparable enrichment scores between the proteomic and metabolomic layer. However, computing the correlation of pathway scores between the proteomic and metabolomic layer emits a coefficient of 0.43 (= Pearson correlation). This shows that the enrichment of the pathways in the two omics layers coincides to some extent. Again, this correlation becomes relatively low when considering the direct link between the prior knowledge sets used for the PEA of proteomic and metabolomic data.

3.2. Identification of key actors in deregulated pathways of LUAD

Enriched pathways were linked to their respective deregulated proteins and metabolites in the following analysis step. For this, members of the aforementioned most deregulated pathways were sorted according to their t-values. The top five actors with the highest absolute t-values of each pathway were chosen. These selected members are referred to as key metabolite and protein actors since they make up the most deregulated players in the respective pathway. Thus, indicating their importance among the dysfunctional pathways. Disrupted regulation for each actor is displayed in the form of volcano plots in figure 3 and the metabolomic and proteomic layer are shown, respectively. The plots display the logarithmic fold-change (log-FC) of the abundance between the tumorous and healthy cohort samples for every key actor. Moreover, the significance of this abundance change is shown by the adjusted p-value on a negative decadic logarithmic scale.

3.3. Key actors contribute differently to the enrichment of the pathway

The higher a protein or metabolite appears in the volcano plot, the more significant the alteration of its abundance. It is noticeable that the extracted key actors do not exactly correspond to the overall most significantly altered proteins and metabolites. However, these key actors exhibit strong deregulation, particularly for proteomics data. This highlights their importance in dysfunctional processes concerning signaling regulation and metabolism. Hereby, the volcano plots illustrate that some actors contribute more while others contribute less to the enrichment of each pathway. This is especially noticeable in the metabolomic layer, whereby many actors reside near a

Results

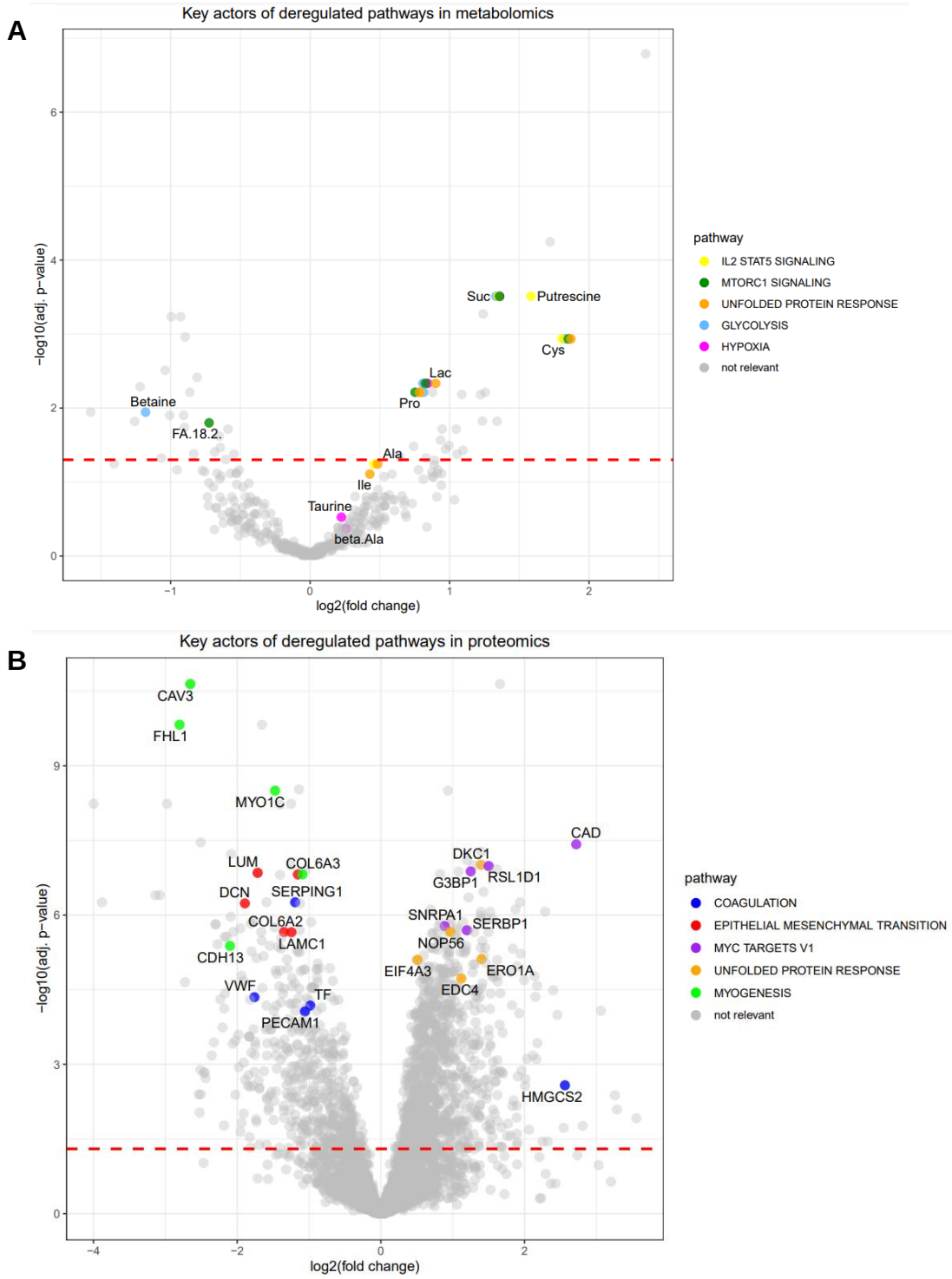


Figure 3: Volcano plots of key actors in the most deregulated pathways in the metabolomic and proteomic layer of LUAD.
Volcano plots show the log2-FC of the abundance of the actors between cancerous and healthy samples. Positive log-FC indicates up-regulation of an actor's abundance in LUAD, whereas negative states its down-regulation. The y-axis displays the negative log10-FC of the adjusted p-value. For both volcano plots, actors surpassing the red dotted line are considered significantly deregulated (adj. p-value < 0.05). (**A**) Five members of each of the top five most deregulated pathways in metabolomics were assigned according to highest absolute t-values of differentially abundance analysis. These metabolite actors are highlighted in respect to their featuring pathway stated in the legend on the right side. To increase visibility of actors attributed to several pathways, the function *jitter* of ggplot2 (version 3.3.6) adds small variation to the points. Thereby, all pathways the actor is involved in become visible. (**B**) For proteomics, the five main actors across top five most deregulated pathways are visualized in the same way as mentioned in (A).

log-FC of zero alongside with low significance. Low significance is indicated by the red dotted line with an adjusted p-value > 0.05. This includes metabolomic actors like beta-alanine (beta.Ala), taurine, isoleucine (Ile) and alanine (Ala). In theory, it is highly unlikely that insignificant deregulated molecules are part of the networks generated with COSMOS, as they do not surpass the input threshold. Nevertheless, key actors exhibiting significant deregulation do not necessarily need to end up in the network either, since CARNIVAL may not find a causal connection to other nodes. Still, it is plausible that a high proportion of the identified key actors end up as regulators in the generated networks. These will then be used as important building blocks to manually create subnetworks out of the generated networks. As identified in figure 3, significantly altered players used for further analysis are metabolite actors, such as butanedioic acid (suc for succinic acid), cysteine (cys), proline (pro) and linoleic acid (FA.18.2). All identified protein actors exhibit significant perturbation. Among the deregulated proteins that can be found in following extracted subnetworks are CAD, DKC1, CAV3, EIF4A3, NOP56 and SERBP1.

3.4. Analysis of subnetworks generated by COSMOS

COSMOS was performed to connect signaling with metabolite input nodes in the “forward” run and *vice versa* for the “backward” run. Thereby, COSMOS considered 194 deregulated proteins (absolute t-value > 6) and 46 deregulated metabolites (absolute t-value > 3) as input for CARNIVAL. Here, the network depth was set to 4, describing the maximum number of steps used to connect input nodes. This was chosen relatively low because COSMOS proposed a large network for the “backward” run. These two networks offer many possible mechanistic insights into LUAD. Such insights consist of directed, signed interactions and important crosstalks between proteomics and metabolomics. Alterations in molecular mechanisms can be seen. The following analyses focus on subnetworks either extracted from the “forward” or “backward" run. To assist in the manual extraction of such subnetworks, identified key actors of PEA are used as start- and endpoints, as well as principal components.

3.4.1. Increased amino acid uptake mediated by inactive TP53

The respective subnetwork shown in figure 4 was obtained from the “forward” run (see full version appendix figure S1). Nodes that do not contribute informational value were omitted to increase visibility in the subnetwork. It thereby comprises 27 nodes and 31 interactions. This COSMOS subnetwork proposes a link between the inactivation of the tumor suppressor gene TP53 to elevated metabolism regarding the uptake of amino acids (AA). COSMOS highlights the inactivity of TP53, which leads to the tumor suppressor gene being unable to inhibit the expression of SLC38A2, a solute-carrier protein (SLC). This SLC protein is a neutral AA-sodium symporter that transports neutrally charged AA over the plasma membrane into the cell. For instance, proline and cysteine were also identified as key actors involved in the glycolysis pathway. The glycolysis pathway was shown to be positively enriched, see figure 2. Furthermore, the volcano plot in figure 3A indicated elevated metabolite levels of cysteine, proline, and lactic acid, which is coherent with the observed regulation in the extracted subnetwork. COSMOS suggests that low activity of the key mediator TP53 is the direct cause of inactive androgen receptor (AR), which lays upstream of it. As depicted in the figure AR activity corresponds to the expression of the kinases PRKCD and MAPK3, which are direct activators of TP53. This part of the subnetwork is however completely down-regulated. Low AR activity can be traced back to strong inhibition mediated by PARP1, a poly-ADP-ribose-polymerase. Lactic acid levels are also increased due to a similar signaling regulation path. Thereby, the subnetwork considers PARP1 as the upstream source. Proline is converted to the furthest downstream target butanedioic acid through abundant P4HA1 in LUAD. A further interaction COSMOS considered essential is the strong down-regulation of the player CAV1 as a downstream node of inactive TP53. The transmembrane protein CAV1 exhibits the same deregulation as CAV3, which can be seen in data of differentially abundance analysis. In this respect, CAV3 forms a highly deregulated player in LUAD, as shown in the volcano plot in figure 3B.

To summarize the topology of this subnetwork, it represents the crosstalk between decreased signaling activity, which in turn causes enhanced metabolism concerning members of the glycolysis pathway.

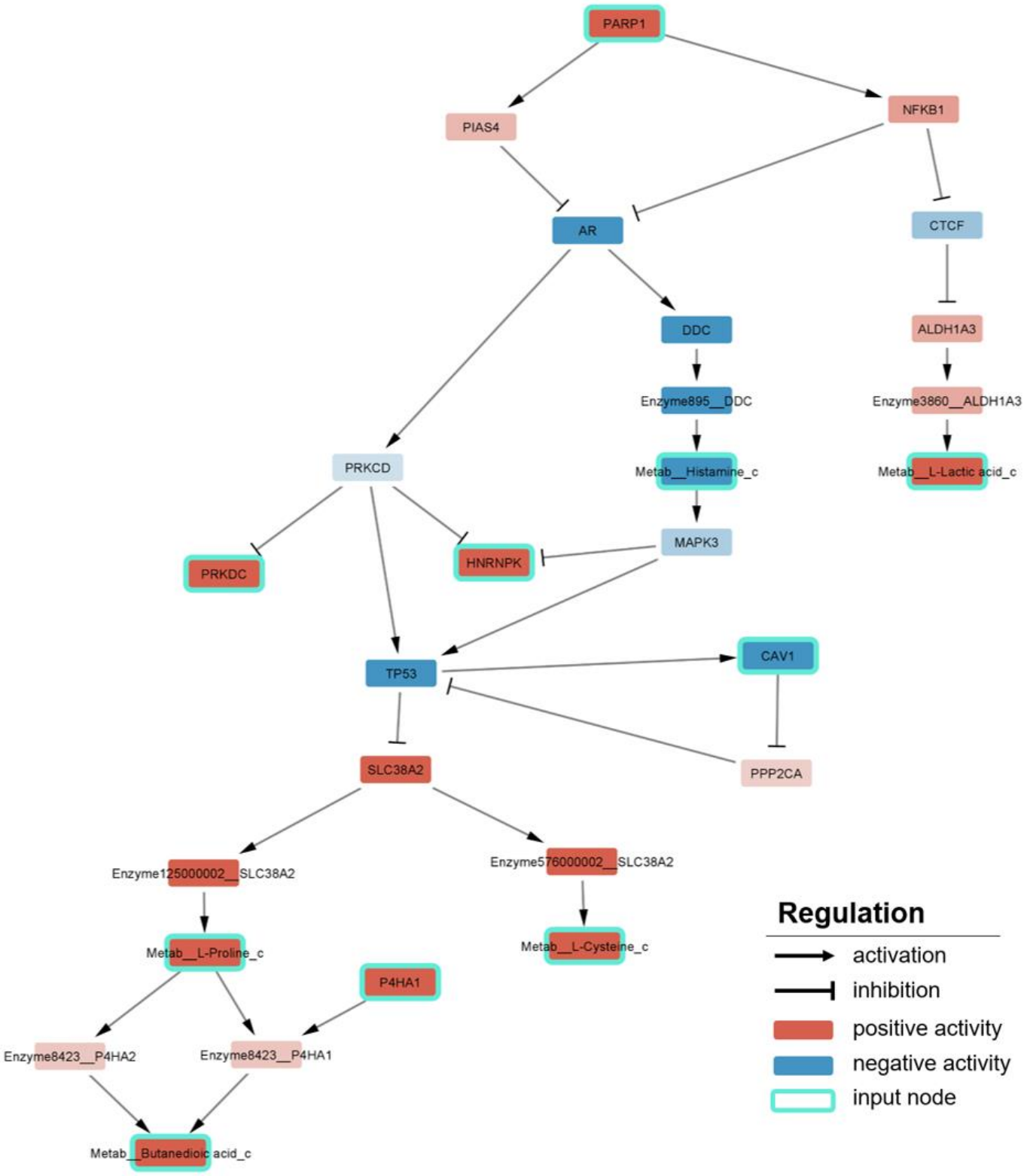


Figure 4: Subnetwork generated by COSMOS with TP53 as mediator between signaling and metabolism.
The legend on the right side shows the mode of regulations that are depicted in the network. Interactions indicate activation or inhibition of the successor node and are marked by an arrow or "T"-sign, respectively. Estimated positive activity is shown by red coloring of the node, while blue coloring displays negative activity. Fading of the color represents lower absolute activity. Activity is here defined as the regulation a node poses downstream. Finally, COSMOS input nodes have turquoise borders. These are proteins and metabolites, which sufficiently surpassed the imposed t-value threshold to be deemed as deregulated. Metabolites are marked by Metab__ at the beginning of the node name. Their cellular location is also indicated. Hereby, metabolites with the addition _c are present in the cytosol, while _m stands for mitochondria. Interactions with nodes designated with Enzyme__ point towards a process that is catalyzed by the respective enzyme. For instance, the enzyme P4HA1 catalyzes the conversion of proline to butanedioic acid.

3.4.2. Targets of MYC exhibit strong expression in LUAD

Enrichment analysis unveiled “MYC targets” of the hallmark gene set as the highest deregulated pathway of the proteomic layer. This results in high deregulation of various proteins, which have MYC as their common regulator. The respective subnetwork extracted from the “backward” run (see full network appendix figure S2) is shown in figure 5 and provides an overview of the altered regulation of MYC in LUAD. It comprises 28 interactions connecting 23 signaling nodes. Upstream regulators of MYC try to explain the strong deregulation observed for targets of MYC. The influence of metabolites as upstream regulators of MYC was omitted because a direct connection does not seem plausible in the context of the “backward” run. This is because metabolomic nodes are very distant regulators of MYC, about six steps within the network. Compared to the “forward” run, the “backward” run mostly revolves around signaling regulation as it only includes three metabolites. Crosstalks between the metabolomic and proteomic layer are hardly observable “backward” run and omitted in the here extracted subnetwork.

Figure 5 displays MYC as a key regulator of several downstream targets. Here, the elevated levels of these proteins are linked to MYC as a direct footprint. MYC acts as a positive regulator, which leads to the co-expression of for example CAD, NOP56 and EIF4A3. These listed examples were particularly found as highly deregulated actors of the most enriched pathways in LUAD. The overly active state of MYC in LUAD is coherent with the results of PEA, in which the targets of MYC account for the strongest positively enriched pathway. Among the targets of MYC, there is only one indirect accounted down-regulation. This is due to the downstream kinase CDK4 phosphorylating MEF2D inhibiting DES. COSMOS explains MYC's highly active state due to upstream signal transduction pathways, including elevated kinase activities of MAPK3 and CSNK2A1, and up-regulated TFs, such as HNF4A, YY1 and TCF7L2 in LUAD.

There are some observable incoherence concerning the regulatory states of players in the here featured COSMOS subnetwork. This is that the kinase MAPK3 depicts a positive upstream regulator of MYC. However, figure 4 indicates the down-regulation of MAPK3. Comparing this inconsistency to the differential abundance analysis of the proteomic layer, a t-value of t = -4.38 is given for MAPK3. Thereby, the activity state of MAPK3, as shown in figure 5, is not reasonable. Furthermore, there is a discrepancy

between the activity state of MYC itself between the “forward” and “backward” run. The “forward” run proposes down-regulation of MYC linked to metabolism (see appendix figure S1), whereas the here featured “backward” run, as well as PEA, indicate an over-active state of MYC.

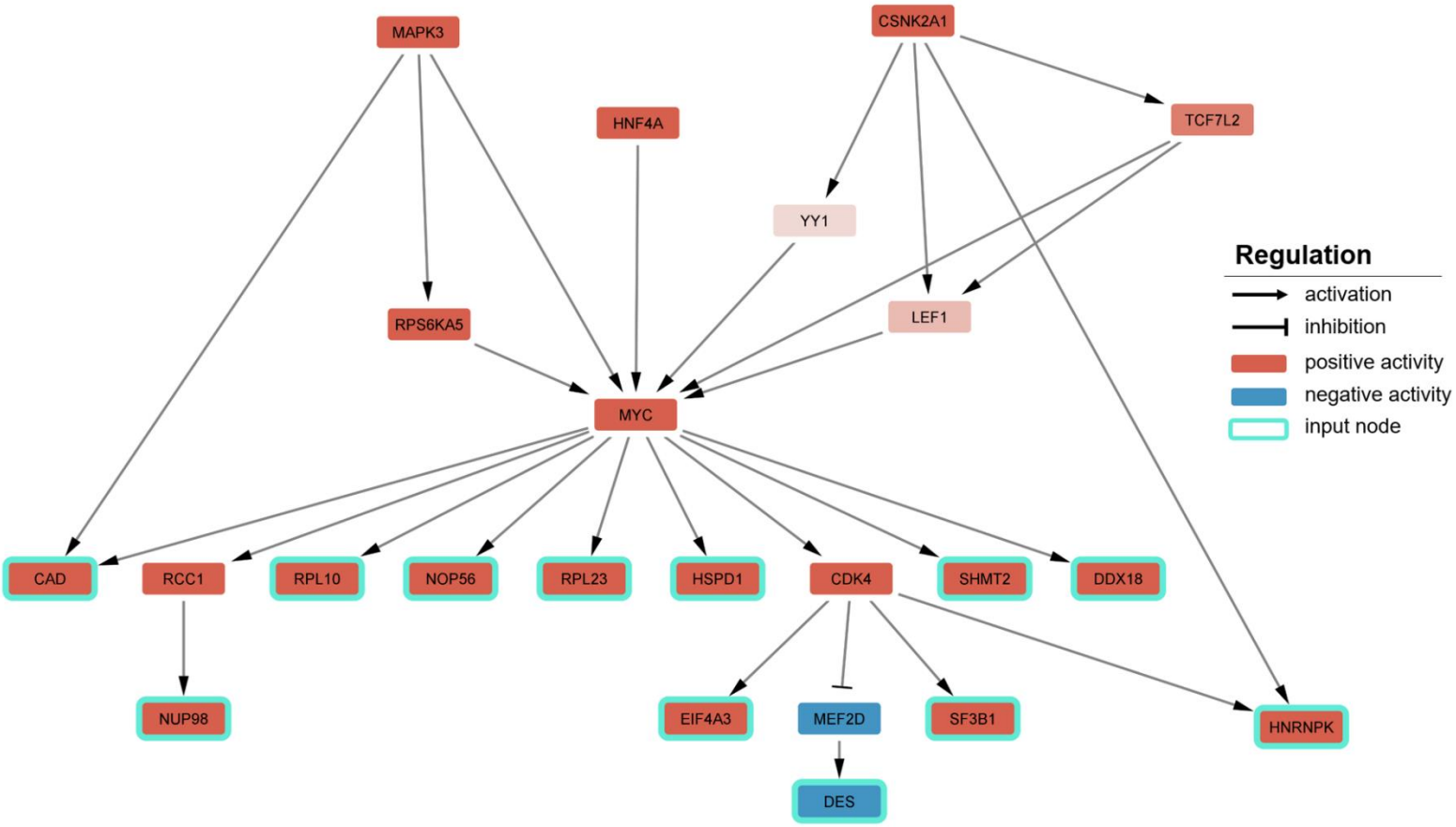


Figure 5: MYC is responsible for elevated levels of various downstream targets.
The mode of regulations for the respective subnetwork are given in the legend on the right side. COSMOS proposes a direct link between upstream kinases and TFs to the up-regulation of MYC. Thereby, the kinases MAPK3, as well as CSNK2A1 and the TF HNF4A modulate the activation of MYC. Downstream targets of MYC, such as CAD, NOP56, EIF4A3 were already identified as highly deregulated players in PEA.

3.4.3. Low linoleic acid levels initiate signaling regulation

A further disease mechanism described by COSMOS was also extracted from the “backward” run. It comprises 22 nodes that are involved in 21 interactions. The subnetwork shown in figure 6 depicts a crosstalk from metabolism to signaling regulation. The observed mechanism starts with strong down-regulation of linoleic acid levels in LUAD. Linoleic acid (FA.18.2) forms a strongly down-regulated metabolite actor, which can be seen in figure 3B. As a consequence of low abundance of the fatty acid, the glycoprotein CD36 does not contribute to calcitriol uptake in LUAD cells. COSMOS suggests that due to low cellular calcitriol levels, translation of CASP3 is not repressed in LUAD. Hereby, CASP3 describes the gene encoding for the protease caspase-3. In this respect, the protease up-regulates TOP1 and SRP72 expression while inhibiting the expression of SPTAN1 and ADH4. Besides regulating the expression of proteins directly, high CASP3 activity also initiates a signal transduction pathway. Thereby, COSMOS proposes the up-regulation of the downstream kinase PRKDC, which in turn enhances the activity of the TF USF1. As a result of the observed signaling mechanism, the expression of P4HA1, SERBP1, DKC1 and EI2FA are elevated in LUAD. ITGA1 expression, on the other hand, is decreased.

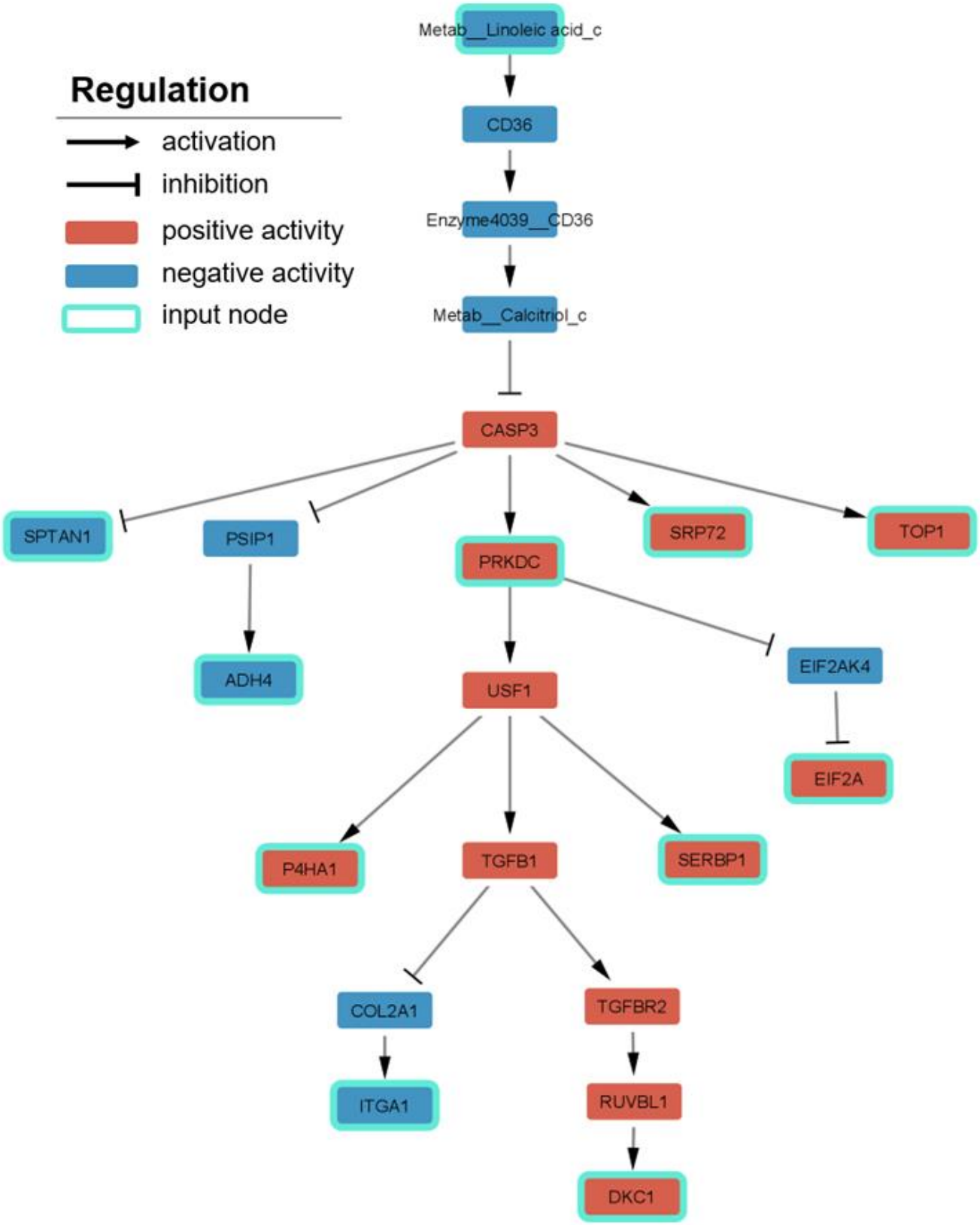


Figure 6: Signaling cascade initiated by low linoleic acid levels in LUAD.
The respective subnetwork extracted from the "backward" run connects the metabolic player linoleic acid as the starting point of the given signaling cascade. The signaling cascade comprises a transduction pathway spanning over highly active CASP3 and PRKDC to elevated levels of already detected player, such as DKC1 and SERBP1. Inhibition of downstream targets on the other hand is also partly observable for players, such as for example ITGA1 and ADH4.

4. Discussion

In this study, COSMOS was applied to differentially abundant proteins and metabolites. This approach generates multi-omics networks that illustrate disease mechanisms of LUAD. With indications of enriched pathways, three subnetworks were obtained. The first presented one focuses on up-regulation of metabolism, in the sense of increased AA uptake and conversion, due to absent regulation by inactive TP53. As the most enriched pathway of the proteomic layer suggested the over-active state of MYC, the second network emphasizes this respective regulation. Furthermore, another subnetwork has been identified conveying a crosstalk between low cellular levels of linoleic acid over a signal transduction pathway to various deregulated proteins.

Firstly, the discussion puts the mechanisms behind enriched pathways in the context of the pathology of LUAD. Hereby, occurring sources of error of the PEA and its informational value are discussed. Then the extracted hypothetical disease mechanisms from the three presented subnetworks of LUAD are evaluated according to proven physiological LUAD disease mechanisms, as well as to supporting evidence of relevant displayed interactions. Lastly, the usage of deregulated players appearing in the subnetworks is assessed with respect to established therapeutical targets and biomarkers, as well as targets and clinical markers proposed by literature.

4.1. Enriched pathways indicate crosstalks between the omics layers

This study aims to discover important crosstalks between the metabolomic and proteomic layer of LUAD as proposed by COSMOS. Enrichment analysis of the hallmark gene set already indicated a mild relationship within the multi-omics data. Looking at the Pearson correlation of r=0.43, many pathways of the metabolomic and proteomic layer appear to be deregulated simultaneously. For instance, the process UPR is among the most positively enriched pathways of the multi-omics data set. UPR describes a mechanism whereby un- and misfolded proteins accumulate in the endoplasmic reticulum due to physiological stress, such as oncogene activation or oxygen- and nutrient deprivation. To endure this stress and escape apoptosis, a signaling response is initiated, restoring normal cellular function by up-regulating the synthesis of chaperones and halting protein translation. Cancer might even take advantage of this elevated homeostasis for its survival and proliferation (Madden *et al.*,

2019). This enriched pathway reflects the high cellular stress observed in LUAD, as actors of the signaling pathway, such as the translation initiation factor EIF4A3, appear to be strongly up-regulated. It also represents a close relationship to other pathways. For instance, intrinsic stress might also be associated with the over-active state of the oncogene MYC. Its effect on downstream targets is observed to be the most deregulated pathway of the proteomic layer, as seen in figure 2 (MYC targets V1). This remark was utilized to model MYC's upstream regulators and downstream targets in a subnetwork.

Elevated metabolism concerning glycolysis is a further disrupted cellular process extracted from the PEA, which occurs for nearly all cancers. In this respect, change in metabolism concerning energy generation forms a hallmark of cancer. Most cancer cells exhibit the Warburg effect, whereby energy is produced predominantly through aerobic glycolysis. Moreover, rewiring metabolism supports cell growth and proliferation by providing building block for anabolism (Liberti and Locasale, 2016). Studies show the promising use of members constituting the glycolysis pathway as a biomarker for predicting metastasis and survival of LUAD patients (Zhang *et al.*, 2019).

Generally, such findings offer valuable insights into dysfunctional biological processes of LUAD, such as disrupted cell signaling or elevated glycolysis. However, results of PEA need to be treated with caution. This is due to the small sample size (10 patients) used in the differentially abundance analysis. Results achieved in this analysis might not be reproducible with larger cohorts and pathways could reflect different modes of regulations. In addition to the sample size, there is bias in selecting the most enriched pathways for the metabolomic layer. Individual metabolites that appear to be strongly deregulated are multi-functional (Maleki *et al.*, 2020). This can be seen in figure 3A, in which some metabolites, such as proline, succinic acid and lactic acid, have even been attributed to several pathways. One possible explanation can be traced back to the approach, how prior knowledge of the PEA has been set up for the metabolomic layer. Hereby, each appearing protein of the hallmark gene set was linked to metabolites displaying interactions in the PKN of COSMOS. Since metabolites are involved as substrates or allosteric regulators of multiple proteins, such as enzymes, kinases or TFs, they appear to be multi-functional. To add to this, the figure also shows that some of the metabolite actors exhibit a low log-FC and are less significant concerning their perturbation. One observable example is the hypoxia

pathway, in which metabolites like taurine and beta-alanine are not significantly deregulated (adjusted p-value > 0.05). This demonstrates a further flaw of PEA, as some players contribute much more to the enrichment of a pathway than others. Nevertheless, these effects can be neglected since PEA is basically used to identify important deregulated players in following analyses. It also helps to reveal their biological involvement in these networks.

4.2. Dysfunctional TP53 increases cellular levels of amino acids

COSMOS linked the inactivity of TP53 to increased uptake of amino acids proline and cysteine, as seen in figure 4. The subnetwork proposes a mechanism in which down-regulation of upstream kinases causes the inactivity of the regulator TP53. In brief, the mechanism conveys a significant crosstalk between down-regulated signaling transduction and elevated metabolism. COSMOS highlights this crosstalk due to the inactive state of p53 being unable to repress the expression of the solute-carrier SLC38A2. Hereby, the tumor suppressor protein p53 is encoded by the observed TP53 gene. This particular interplay is also evident in literature (Bhutia and Ganapathy, 2016). This change in metabolism concerns high abundant SLC38A2, which ensures sodium-coupled influx transport of neutral AA. The expression of the SLC protein is probably needed to keep up with the increased nutrient demand in proliferating cancer cells.

Interestingly, the mTORC1 signaling pathway, which forms a complex of several proteins including mammalian target of rapamycin (mTOR), was reported to be activated by an AA sensing signaling cascade (Jewell *et al.*, 2013). In this study, strong positive enrichment of the signaling process mTORC1 has been detected in PEA, which is also associated with increased AA abundance of proline, cysteine, and metabolite abundance of lactic acid and succinic acid (seen in the volcano plot of figure 3A). Similar to this cellular mechanism, Morotti *et al.* demonstrated that knockdown of SLC38A2 inhibits the mTROC1 signaling pathway in breast cancer cell lines, leading to reduced tumor growth. For cancer, they linked SLC38A2 specifically to the depletion of glutamine, which is not observed as a deregulated actor in the proposed subnetwork and metabolomic data (Morotti *et al.*, 2021). Nevertheless, these studies are consistent with the here observed findings. This strongly suggests investigating the development of a drug to specifically target SLC38A2 in LUAD. Problems could arise concerning

sensitivity towards cancerous cells, as targeting transport proteins might pose an immense risk to other tissues in the body. Still, there is evidence that the inhibition of AA transporters displays an anti-tumor effect, which is even applicable for NSCLC (Imai *et al.*, 2010; Marshall *et al.*, 2016). One potential therapeutical agent, MeAIB (α-methylaminoisobutyric acid), targets a range of SLC proteins, such as the here proposed SLC38A2, by functioning as a transporter selective substrate analog (Morotti *et al.*, 2021). However, its application is more promising as a clinical marker in detecting tumor size and metastasis, whereby clinical trials were completed in 2018. In this regard, positron emission tomography (PET) using MeAIB as a radiotracer can be performed, which identifies the malignancy and helps to evaluate further treatment strategies for LUAD. It was shown that [^{11}C]-MeAIB-PET (MeAIB labeled with isotope carbon-11) exhibits higher diagnostic accuracy in lung cancer and even performs better than the established [^{18}F]-fluoro-2-deoxy-D-glucose-PET ([^{18}F]-FDG-PET) in respect to oncologic imaging (Nishii *et al.*, 2013).

To refer back to the regulator TP53, COSMOS suggests its inactive state due to down-regulation of upstream kinases, which traces back to low AR activity in LUAD. This particular interplay of AR attributing to the regulation of TP53 is elusive in literature. In this regard, it is more plausible that present somatic mutation of TP53 in LUAD patients caused this dysfunctional state rather than low AR activity. Thereby, alterations of the proposed upstream regulatory pathway of TP53 might not depict actual physiological mechanisms. Unfortunately, it is not known if the patients underly a somatic mutation of TP53. As a result, reprogrammed AA uptake in LUAD might be attributed to TP53 mutation. This mutation is observed in 50 % of all cases (Herbst *et al.*, 2018). COSMOS does not incorporate mutations, since the tool only depicts the molecular phenotype of the cancer, whereby deregulated events are explained in a mechanistic manner. This points to a limitation of COSMOS, which is that interactions within the PKN might be utilized that do not necessarily reflect the actual disease physiology (Dugourd *et al.*, 2021).

The COSMOS network also showed that dysregulation of TP53 plays a role in down-regulated Caveolin-1 (CAV1) expression. CAV1 is a scaffolding protein involved in many processes, such as invagination of the plasma membrane for endocytosis. Strongly decreased expression of CAV1 is commonly observed in early stages of LUAD, whereby re-expression is highly correlated to advancement into late stages of

the tumor, concerning invasiveness and metastasis. CAV1 exhibits huge potential for early diagnosis of the malignancy and, due to its re-expression behavior, for the prediction of significantly shorter survival. However, this has not yet been transferred to clinical trials (Yan *et al.*, 2019; Zhan *et al.*, 2012). It is coherent with the results of this study in the respect that samples cover early stages of LUAD and exhibit strongly decreased CAV1 expression. Further investigation regarding the linkage of TP53 dysfunction to deregulation of CAV1 expression could be interesting for LUAD, as this mechanism could explain one possible cause. However, the re-emergence of CAV1 expression, associated with late cancer progression of LUAD, must be due to another regulatory event. A further report by Yan *et al.* also highlighted the predictive value of CAV1. They also suggested DCN expression as a novel biomarker in predicting patient survival (Yan *et al.*, 2019). Hereby, DCN has also been identified as a highly deregulated actor among the most enriched pathways, as shown in figure 3B.

4.3. Over-active MYC offers many therapeutic and diagnostic interventions

MYC's over-active state (figure 5) gives further mechanistic insight into LUAD pathology. Overexpression of MYC has been attributed to promoting tumor growth by inducing angiogenesis (Knies-Bamforth *et al.*, 2004). Furthermore, MYC expression is associated with causing metastasis in NSCLC. This respective study also showed that in order for MYC to drive tumor progression in NSCLC mice models, additional regulatory events are required, such as C-RAF or KRAS oncogenic activation (Rapp *et al.*, 2009). MYC functions as a TF and its constitutive expression induces stress on the cell caused by the simultaneous over-expression of several proteins in LUAD. COSMOS states that this activation of MYC is achieved by increased activity of upstream kinases, such as MAPK3, CSNK2A1 and TFs, like TCF2L and HNF4A. Contrary to the observed enhancing role of MAPK3 for MYC expression in LUAD, the differentially abundance analysis states its down-regulation. Thus, MAPK3 should not have physiological value in this respective mechanism. Another discrepancy occurs in MYC's activity status between the "forward" and "backward" run. However, the over-active state, as shown in figure 5, seems much more plausible. The reason is that various downstream targets exhibit strongly elevated expression, which can be linked to MYC. Whereas the stated inactive state of MYC in the "forward" run (appendix figure S1) seems mechanistically insignificant. Furthermore, the positive regulation of MYC

is also supported by the “MYC targets V1” set being the most enriched pathway of the proteomic layer (figure 2).

Besides that, a valuable insight was generated by CARNIVAL in connecting the alpha subunit of casein kinase II (CSNK2A1) as an upstream activator of MYC. In this interplay, CSNK2A1 increases the activity of downstream TFs, like TCF7L2 and YY1, via phosphorylation which then contributes to MYC overexpression. Especially interesting is that TCF, as part of the signaling pathway Wnt, can act in a positive feedback loop with MYC expression, which was shown in breast cancer. The respective study by Cowling and Cole recommends the MYC-Wnt pathway as an attractive therapeutical target approach (Cowling and Cole, 2007). Furthermore, CSNK2A1 might even phosphorylate MYC to directly increase its TF activity, which has been linked to tumor proliferation in lymphoma (Channavajhala and Seldin, 2002). Therefore, it seems highly likely that the casein kinase accounts for an attractive therapeutic target in LUAD. For this, Huang *et al.* proposed an allosteric inhibitor specific to the alpha subunit of CSNK2 named hematein. The therapeutical agent showed efficient inhibition of cell growth and induction of apoptosis in lung cancer cells of a murine xenograft model. Hematein interferes with the Wnt/TCF pathway (Hung *et al.*, 2013). Other allosteric inhibitors of protein kinase CSNK2, such as azonaphthalene derivates, have also been shown to exhibit anti-tumor activity (Moucadel *et al.*, 2011). Given the mechanism shown in the subnetwork of figure 5, it seems plausible that hematein would also contribute to the inhibition of MYC expression. Due to the fact, that CSNK2A and TCF2L are upstream regulators of MYC in LUAD. This could be investigated experimentally. Besides the potential usage of hematein as a therapeutical agent, there have been no clinical trials for lung cancer so far. The biomedical efficacy of hematein in NSCLC and its occurring toxicity would be very interesting to know.

Besides the aforementioned mechanistic reasons for hyperactive MYC function, genetic mechanisms also cause aberrant MYC expression. Alterations of MYC include gene amplification, which is observed in 31 % of cases in LUAD, or translocation of the gene downstream of highly active promoters (Kalkat *et al.*, 2017). So far, treating MYC as a direct therapeutical target has been avoided since the TF is involved in numerous significant cellular regulations and interfering would be highly toxic. These immense side effects emerge due to poor bioavailability of MYC inhibitors. Contrary to this, a

recent study by Beaulieu *et al.* emphasizes the promising impact of a cancer cell penetrating mini-protein drug named Omomyc (OMO-103), with high specificity towards MYC and tolerable side effects. For NSCLC, it is currently in clinical trials phase I/II and intended as an anti-metastatic drug. If this succeeds, other cancer types, such as breast cancer, might also benefit from Omomyc in the future (Beaulieu *et al.*, 2019).

Another approach to utilize overexpression of MYC is for clinical diagnosis of LUAD, whereby downstream targets can be examined. High expression of MYC targets commonly contributes to cell growth and proliferation. For instance, overexpression of CAD enhances pyrimidine synthesis, while elevated levels of the translation initiation factor EIF4A3 contribute to increased RNA metabolism. Such aberrant changes in signaling and metabolism enable the cell to go through cell cycles and proliferate (Grandori and Eisenman, 1997). Thereby, these two mentioned actors, CAD and EIF4A3, have been shown to predict patient survival and response to treatment (Qiu *et al.*, 2022; Wu *et al.*, 2020).

4.4. The PKN of COSMOS determines the quality of gained insights

Another potential disease mechanism of LUAD presented by COSMOS concerns the role of vitamin D deficiency in regulating signaling, as shown in the subnetwork of figure 6. Interestingly, COSMOS links the depletion of vitamin D to low abundance of linoleic acid in LUAD. Hereby, the PKN of COSMOS comprises a hypothetical interaction of linoleic acid acting as a positive regulator of the integral membrane protein CD36, which in turn is responsible for vitamin D (calcitriol) uptake. In this case, due to lacking linoleic acid in LUAD, the positive regulation of CD36 is absent and LUAD cells become vitamin D deficient. There is experimental evidence for both modes of regulation. Firstly, fatty acids have been shown to regulate CD36 (Smith *et al.*, 2008). Secondly, knock-out of the CD36 gene in mice resulted in vitamin D deficiency (Kiourtzidis *et al.*, 2020). However, it can only be speculated if the interplay between uptake of vitamin D by CD36 and its regulation by fatty acids depicts a true mechanism. If this is the case, it would remain hard to tell whether this process has any significant value in describing vitamin D depletion in cancer. Still, further investigation might elucidate this respective interplay.

Discussion

Calcitriol, the active form of vitamin D, is crucial for preventing carcinogenesis of many cancer types due to its involvement in processes, such as the resolution of chronic inflammation, inhibiting cell proliferation and pro-apoptotic effects for cancer cells. Furthermore, vitamin D deficiency correlates with a higher risk of developing cancer (Díaz *et al.*, 2015). Regarding the anti-cancer effect of vitamin D and its regulatory impact on caspase expression, a completely different mechanism than depicted in the subnetwork of figure 6 becomes evident. Studies are confident that higher vitamin D levels increase the activation of caspases, whereby caspase signaling may result in the programmed death of cancer cells (McGuire *et al.*, 2001). However, COSMOS suggests that lacking vitamin D (calcitriol) levels are unable to repress the expression of caspase-3 (CASP3). The induction of the signaling cascade by low fatty acid abundance thereby displays a physiological incorrect mechanism since the interaction should indicate activation rather than inhibition. One way to correct such an error in COSMOS is to simply remove the given interaction from the PKN, as it does not support biological evidence. Hence, upon regenerating the “backward” run, this interaction will not be constructed by CARNIVAL and the network topology would be impacted slightly. However, this effect was neglected here, since the other extracted mechanistic insight of the “backward” run concerning the over-active state of MYC provides a highly plausible disease hypothesis in the context of LUAD.

In addition, this observed result demonstrates the main determining factor of the performance of COSMOS, which regards its prior knowledge set. The performance of COSMOS in deciphering disease mechanisms is mainly dependent on the quality and completeness of the PKN, for which a few limitations arise. The PKN of COSMOS was built by implementing multiple databases. Unfortunately, a considerable proportion of interactions between biomolecules remains unknown in these resources. CARNIVAL will thereby not consider these potential regulations. Furthermore, MS only captures the abundance of a limited number of proteins and metabolites. Thereby, input data is incomplete too, and significant interactions between nodes might not be covered (Dugourd *et al.*, 2021).

Fortunately, the presented limitations offer room for improvement. COSMOS can continually be refined in its quality by maintaining the PKN. This involves the resources Recon3D, STITCHdb and OmniPath, which are constantly curated so that interactions reflect physiological evidence. Moreover, newly identified biomolecular interactions

can be provided by further implementing high-quality literature resources to the PKN of COSMOS. Additionally, advancements in omics technologies will make the generated input data more complete (Dugourd *et al.*, 2021).

5. Conclusion

In conclusion, this study captured two strongly altered regulatory events that provide potential targets for therapeutical intervention and applicable biomarkers. Firstly, a crosstalk initiated by inactive TP53 resulting in elevated metabolism concerning AA uptake was observed. This is due to lacking inhibition of SLC38A2 protein expression by p53. This dysfunctional state is most likely caused by somatic mutation of TP53 rather than observed dysregulation of upstream sources. The therapeutical agent MeAIB targets SLC38A2 as a selective substrate analog and its anti-tumor activity in LUAD should be investigated more carefully. On the other hand, [^{11}C]-MeAIB is already an established tracer for oncologic imaging via PET and shows great diagnostic accuracy. Furthermore, COSMOS suggested that CAV1 down-regulation is also due to the dysfunctional state of the tumor suppressor. CAV1 can be applied to predict survival in LUAD patients. Secondly, a further mechanistic insight of LUAD revealed the over-active state of MYC in a COSMOS subnetwork. Therefore, the interaction of CSNK2 in activating MYC suggests the possible usage of casein kinase inhibitors on LUAD. Furthermore, the novel therapeutic agent Omomyc might even directly target MYC for NSCLC, for which the outcome of the clinical trials needs to be awaited.

Thereby, this study demonstrates that the method COSMOS offers valuable insights into disease mechanisms. Nevertheless, obtained results remain hypothetical and require further validation in experimental setups. The application of COSMOS to multi-omics LUAD data coincided with available studies and known therapeutic targets. On the other hand, the subnetwork regarding the impact of lacking calcitriol not repressing CASP3 expression could not be linked to physiological evidence. This finding illustrates the impact prior knowledge has in generating mechanistic insights. In this regard, curation of resources used in the PKN will complete missing regulatory interactions and adjust existing ones. Refinement of this will further enhance the performance of COSMOS in the future.

6. Appendix

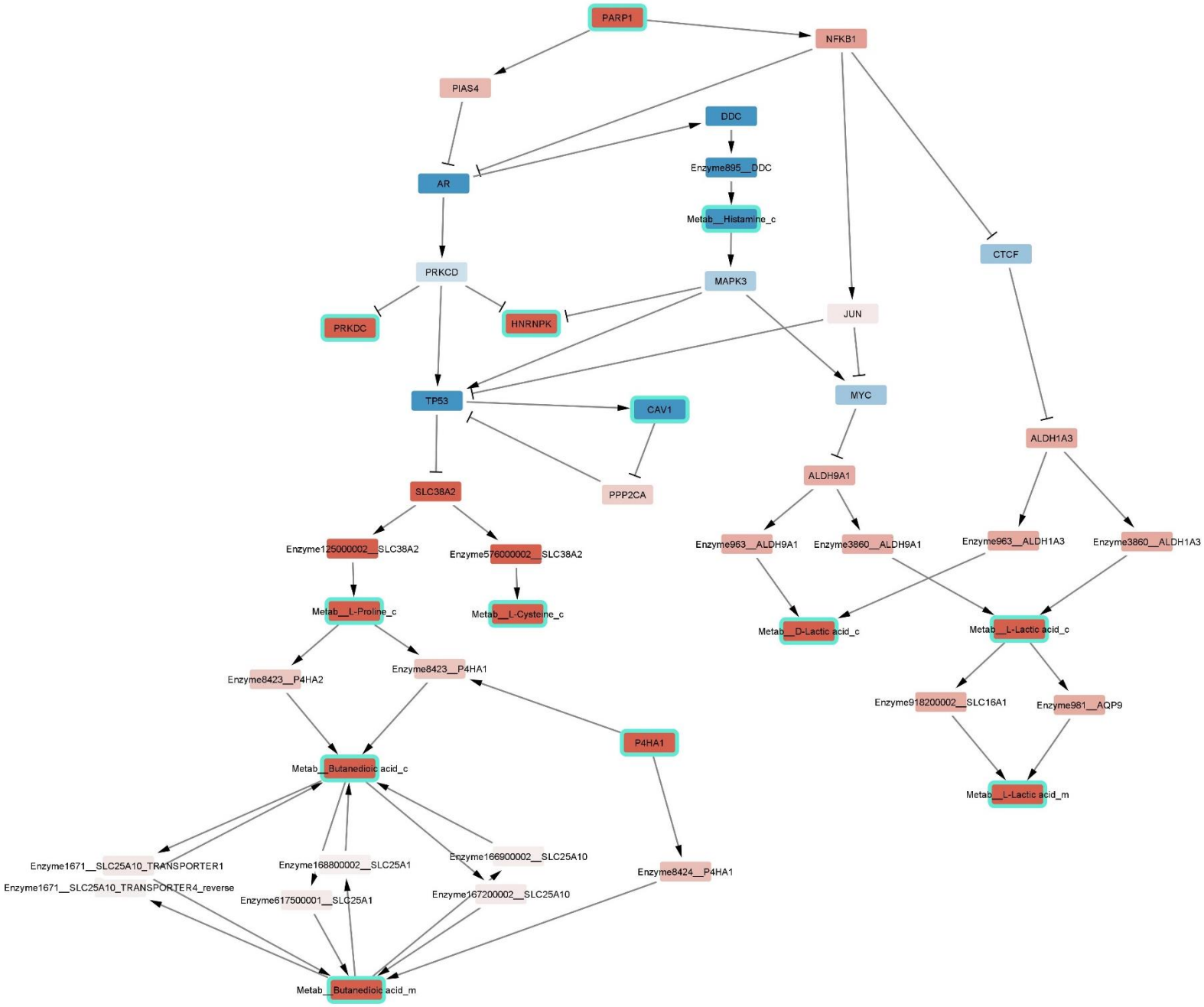


Figure S1: Network generated by COSMOS in the "forward" run, connecting signaling to metabolite nodes. The here displayed network for the "forward" run was used to extract the first presented network in figure 4. Here, the role of inactive TP53 in mediating metabolomic up-regulation is also clearly observable.

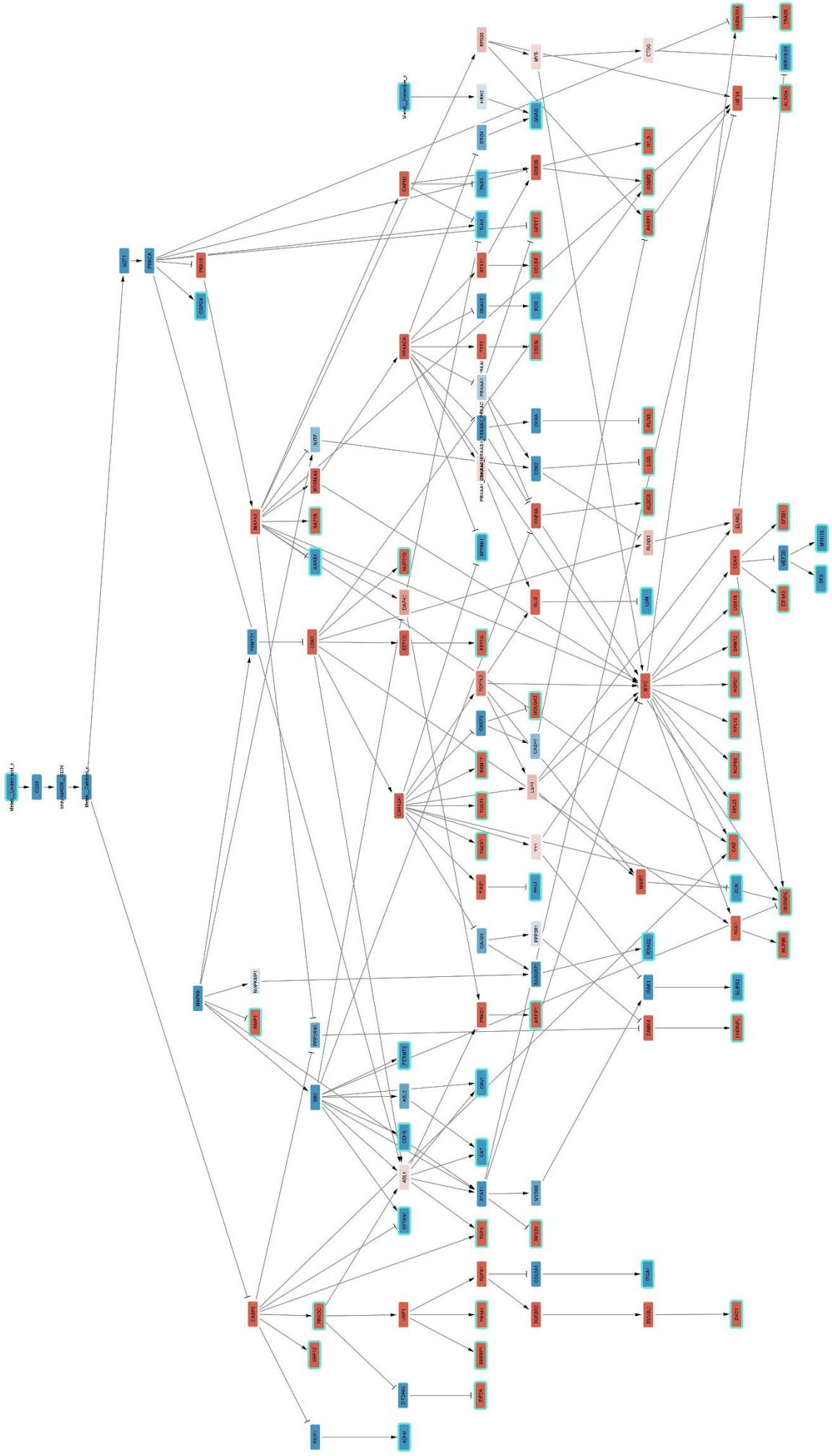


Figure S2: Network generated by COSMOS in the "backward" run, connecting metabolite to signaling nodes. The network is flipped sidewards, as it is very wide. The subnetwork involving targets and upstream regulators of MYC was extracted from this respective "backward" run. Furthermore, the mechanism focusing on linoleic acid as the initiator of a signaling cascade was extracted from this network too. Higher quality of this figure is available at the GitHub repository mentioned in the methods section.

7. References

Ahrendt, S.A., Hu, Y., Buta, M., McDermott, M.P., Benoit, N., Yang, S.C., Wu, L., and Sidransky, D. (2003). p53 Mutations and Survival in Stage I Non-Small-Cell Lung Cancer: Results of a Prospective Study. JNCI: Journal of the National Cancer Institute *95*, 961-970.

Badia-I-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O., and Saez-Rodriguez, J. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances *2*.

Beaulieu, M.E., Jauset, T., Massó-Vallés, D., Martínez-Martín, S., Rahl, P., Maltais, L., Zacarias-Fluck, M.F., Casacuberta-Serra, S., Serrano Del Pozo, E., Fiore, C., Foradada, L., Cano, V.C., Sánchez-Hervás, M., Guenther, M., Romero Sanz, E., Oteo, M., Tremblay, C., Martín, G., Letourneau, D., Montagne, M., Morcillo Alonso, M., Whitfield, J.R., Lavigne, P., and Soucek, L. (2019). Intrinsic cell-penetrating activity propels Omomyc from proof of concept to viable anti-MYC therapy. Sci Transl Med *11*.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289-300.

Bhutia, Y.D., and Ganapathy, V. (2016). Glutamine transporters in mammalian cells and their functions in physiology and cancer. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1863*, 2531-2539.

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

Channavajhala, P., and Seldin, D.C. (2002). Functional interaction of protein kinase CK2 and c-Myc in lymphomagenesis. Oncogene *21*, 5280-5288.

Cowling, V.H., and Cole, M.D. (2007). Turning the tables: Myc activates Wnt in breast cancer. Cell Cycle *6*, 2625-2627.

Díaz, L., Díaz-Muñoz, M., García-Gaytán, A.C., and Méndez, I. (2015). Mechanistic Effects of Calcitriol in Cancer Biology. Nutrients *7*, 5020-5050.

Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K.B., Vieira, V., Bekker-Jensen, D.B., Kranz, J., Bindels, E.M.J., Costa, A.S.H., Sousa, A., Beltrao, P., Rocha, M., Olsen, J.V., Frezza, C., Kramann, R., and Saez-Rodriguez, J. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. Molecular Systems Biology *17*, e9730.

Dugourd, A., and Saez-Rodriguez, J. (2019). Footprint-based functional analysis of multiomic data. Current Opinion in Systems Biology *15*, 82-90.

Gegner, H., Müller, T., Naake, T., Dugourd, A., Schilling, D., Kliewer, G., Kunze-Rohrbach, N., Saez-Rodriguez, J., Schneider, M.A., Huber, W., Poschet, G., Hell, R., and Krijgsveld, J. (2022). Evaluation of a single-sample workflow for the joint analysis of metabolomics and proteomics. Unpublished Work.

Grandori, C., and Eisenman, R.N. (1997). Myc target genes. Trends in Biochemical Sciences *22*, 177-181.

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. Genome Biology *18*.

Herbst, R.S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. Nature *553*, 446-454.

Huang, S., Chaudhary, K., and Garmire, L.X. (2017). More Is Better: Recent Progress in Multi-Omics Data Integration Methods. Frontiers in genetics *8*, 84-84.

Hung, M.S., Xu, Z., Chen, Y., Smith, E., Mao, J.H., Hsieh, D., Lin, Y.C., Yang, C.T., Jablons, D.M., and You, L. (2013). Hematein, a casein kinase II inhibitor, inhibits lung cancer tumor growth in a murine xenograft model. Int J Oncol *43*, 1517-1522.

Imai, H., Kaira, K., Oriuchi, N., Shimizu, K., Tominaga, H., Yanagitani, N., Sunaga, N., Ishizuka, T., Nagamori, S., Promchan, K., Nakajima, T., Yamamoto, N., Mori, M., and Kanai, Y. (2010). Inhibition of L-type amino acid transporter 1 has antitumor activity in non-small cell lung cancer. Anticancer Res *30*, 4819-4828.

Jewell, J.L., Russell, R.C., and Guan, K.-L. (2013). Amino acid signalling upstream of mTOR. Nature Reviews Molecular Cell Biology *14*, 133-139.

Kalkat, M., De Melo, J., Hickman, K.A., Lourenco, C., Redel, C., Resetca, D., Tamachi, A., Tu, W.B., and Penn, L.Z. (2017). MYC Deregulation in Primary Human Cancers. Genes (Basel) *8*.

Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. PLoS Computational Biology *8*, e1002375.

Kiourtzidis, M., Kühn, J., Brandsch, C., and Stangl, G.I. (2020). Vitamin D Status of Mice Deficient in Scavenger Receptor Class B Type 1, Cluster Determinant 36 and ATP-Binding Cassette Proteins G5/G8. Nutrients *12*.

Knies-Bamforth, U.E., Fox, S.B., Poulsom, R., Evan, G.I., and Harris, A.L. (2004). c-Myc interacts with hypoxia to induce angiogenesis in vivo by a vascular endothelial growth factor-dependent mechanism. Cancer Res *64*, 6563-6570.

Krassowski, M., Das, V., Sahu, S.K., and Misra, B.B. (2020). State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. Frontiers in Genetics *11*.

Liberti, M.V., and Locasale, J.W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem Sci *41*, 211-218.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Jill, and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems *1*, 417-425.

Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J., and Saez-Rodriguez, J. (2019). From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. NPJ Syst Biol Appl *5*, 40.

Madden, E., Logue, S.E., Healy, S.J., Manie, S., and Samali, A. (2019). The role of the unfolded protein response in cancer progression: From oncogenesis to chemoresistance. Biol Cell *111*, 1-17.

Maleki, F., Ovens, K., Hogan, D.J., and Kusalik, A.J. (2020). Gene Set Analysis: Challenges, Opportunities, and Future Research. Frontiers in Genetics *11*.

Marshall, A.D., van Geldermalsen, M., Otte, N.J., Anderson, L.A., Lum, T., Vellozzi, M.A., Zhang, B.K., Thoeng, A., Wang, Q., Rasko, J.E., and Holst, J. (2016). LAT1

is a putative therapeutic target in endometrioid endometrial carcinoma. Int J Cancer *139*, 2529-2539.

McGuire, T.F., Trump, D.L., and Johnson, C.S. (2001). Vitamin D(3)-induced apoptosis of murine squamous cell carcinoma cells. Selective induction of caspase-dependent MEK cleavage and up-regulation of MEKK-1. J Biol Chem *276*, 26365-26373.

Morotti, M., Zois, C.E., El-Ansari, R., Craze, M.L., Rakha, E.A., Fan, S.J., Valli, A., Haider, S., Goberdhan, D.C.I., Green, A.R., and Harris, A.L. (2021). Increased expression of glutamine transporter SNAT2/SLC38A2 promotes glutamine dependence and oxidative stress resistance, and is associated with worse prognosis in triple-negative breast cancer. Br J Cancer *124*, 494-505.

Moucadel, V., Prudent, R., Sautel, C.F., Teillet, F., Barette, C., Lafanechere, L., Receveur-Brechot, V., and Cochet, C. (2011). Antitumoral activity of allosteric inhibitors of protein kinase CK2. Oncotarget *2*, 997-1010.

Müller, T., Kalxdorf, M., Longuespée, R., Kazdal, D.N., Stenzinger, A., and Krijgsveld, J. (2020). Automated sample preparation with SP3 for low-input clinical proteomics. Mol Syst Biol *16*, e9111.

Myers, D.J., and Wallen, J.M. (2022). Lung Adenocarcinoma. In StatPearls, (StatPearls Publishing).

Naake, T., and Huber, W. (2021). MatrixQCvis: shiny-based interactive data quality exploration for omics data. Bioinformatics *38*, 1181-1182.

Nishii, R., Higashi, T., Kagawa, S., Kishibe, Y., Takahashi, M., Yamauchi, H., Motoyama, H., Kawakami, K., Nakaoku, T., Nohara, J., Okamura, M., Watanabe, T., Nakatani, K., Nagamachi, S., Tamura, S., Kawai, K., and Kobayashi, M. (2013). Diagnostic usefulness of an amino acid tracer, α-[N-methyl-11C]-methylaminoisobutyric acid (11C-MeAIB), in the PET diagnosis of chest malignancies. Annals of Nuclear Medicine *27*, 808-821.

Qiu, M., Chen, M., Lan, Z., Liu, B., Xie, J., and Li, X. (2022). Plasmacytoma variant translocation 1 stabilized by EIF4A3 promoted malignant biological behaviors of lung adenocarcinoma by generating circular RNA LMNB2. Bioengineered *13*, 10123-10140.

Rapp, U.R., Korn, C., Ceteci, F., Karreman, C., Luetkenhaus, K., Serafin, V., Zanucco, E., Castro, I., and Potapenko, T. (2009). MYC is a metastasis gene for non-small-cell lung cancer. PLoS One *4*, e6029.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research *43*, e47.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research *13*, 2498-2504.

Smith, J., Su, X., El-Maghrabi, R., Stahl, P.D., and Abumrad, N.A. (2008). Opposite regulation of CD36 ubiquitination by fatty acids and insulin: effects on fatty acid uptake. J Biol Chem *283*, 13578-13585.

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L.J., Bork, P., and Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res *44*, D380-384.

Travis, W.D. (2011). Pathology of Lung Cancer. Clinics in Chest Medicine *32*, 669-692.

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nature Methods *13*, 966-967.

Wu, G., Zhao, Z., Yan, Y., Zhou, Y., Wei, J., Chen, X., Lin, W., Ou, C., Li, J., Wang, X., Xiong, K., Zhou, J., and Xu, Z. (2020). CPS1 expression and its prognostic significance in lung adenocarcinoma. Ann Transl Med *8*, 341.

Yan, J., Risacher, S.L., Shen, L., and Saykin, A.J. (2018). Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. Briefings in Bioinformatics *19*, 1370-1381.

Yan, Y., Xu, Z., Qian, L., Zeng, S., Zhou, Y., Chen, X., Wei, J., and Gong, Z. (2019). Identification of CAV1 and DCN as potential predictive biomarkers for lung adenocarcinoma. Am J Physiol Lung Cell Mol Physiol *316*, L630-l643.

Zhan, P., Shen, X.K., Qian, Q., Wang, Q., Zhu, J.P., Zhang, Y., Xie, H.Y., Xu, C.H., Hao, K.K., Hu, W., Xia, N., Lu, G.J., and Yu, L.K. (2012). Expression of caveolin-1

is correlated with disease stage and survival in lung adenocarcinomas. Oncol Rep *27*, 1072-1078.

Zhang, L., Zhang, Z., and Yu, Z. (2019). Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. Journal of Translational Medicine *17*, 423.