NETWORK MEDICINE-BASED DISCOVERY OF BIOMARKERS FOR PANCREATIC CANCER

by

Şafak Kılıç

B.S., Columbia University in the City of New York, 2017

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2021

ACKNOWLEDGEMENTS

I would like to extend my sincerest thanks to my thesis supervisor, Prof. Dr. Kutlu Ülgen, without whose wise guidance and endless patience this work would have been impossible. I am forever grateful for her support of, and belief in, my work. I would also like to thank Prof. Dr. Ahmet Kerim Avcı and the entire Boğaziçi University Faculty of Chemical Engineering for their support.

Furthermore, I would like to acknowledge the friendship and encouragement of my peers and co-workers Elif Esvap, Begüm Yağcı, Müge Kasım, Özlem Özbek, Ekin Yurdakul, Feyza Kevser Öner, Mustafa Sertbaş, Elif Gençtürk, Atakan Yüksel and Ali İsmail Duman, who have provided me with mentorship, kindness, and camaraderie throughout this process.

Lastly, I would like to thank my parents, who taught me to be curious and value learning, gave me their love and encouragement, and made it possible for me to write this thesis.

ABSTRACT

NETWORK MEDICINE-BASED DISCOVERY OF BIOMARKERS FOR PANCREATIC CANCER

Pancreatic ductal adenocarcinoma (PDAC) is the most common type of pancreatic cancer. Due to its high lethality and difficult diagnosis, PDAC is a prime candidate for biomarker discovery using novel approaches. This study uses high-throughput gene expression and protein-protein interaction data to apply differential gene expression analysis, constraint-based reconstruction and analysis (COBRA) of the metabolic network, and the construction of protein-protein interaction (PPI) and gene coexpression networks, comparing cancerous and healthy pancreatic tissue. Clustering of PPI and coexpression networks is used to place differential expression and COBRA results in biological context and find gene clusters which describe cancer-related processes. Differential expression analysis on TCGA and GTEx databases found 826 differentially expressed genes (DEGs). For COBRA methods, gene expression data was used to filter the 10,600 metabolic reactions in the Recon3D generic human metabolic model into cancer and healthy tissue-specific models, containing 5,879 and 5,812 reactions, respectively. The metabolic flux profiles of these models were discovered via flux sampling, revealing 1,960 significantly different internal reactions and 338 significantly different exchange reactions. These were narrowed down to 250 internal reactions and 14 exchange reactions whose mean flux difference between two models was greater than two standard deviations of the flux distributions. 55 genes were found to both be DEGs and exhibit significant flux differences in their associated reactions. These genes and the top 5 PPI or coexpression clusters with the highest proportion of DEGs or genes with significant metabolic differences were annotated with Gene Ontology (GO) and Disease Ontology (DO) terms, as well as with known biological pathways. The annotation revealed associations between cancer genes and cell-cell adhesion, immune response, and reactive oxygen species (ROS) stress processes, among others.

ÖZET

PANKREAS KANSERİ İÇİN AĞ TIBBI TABANLI BİYOBELİRTEÇ KEŞFİ

Pankreas duktal adenokarsinomu (PDAC), pankreas kanserinin en sık görülen türüdür. Yüksek ölüm oranı ve teşhisinin güçlüğü nedeniyle PDAC yeni yaklaşımlarla biyobelirteç keşfi için öncelikli bir adaydır. Bu çalışmada ayrımsal gen ekspresyon analizi, kısıtlama bazlı metabolik ağ inşası ve analizi (COBRA), ve protein-protein etkileşim (PPI) ve gen eş ekspresyon ağları kullanılarak kanser ve sağlıklı pankreas dokusunun kıyaslanmasında yüksek hacimli veri kullanılmıştır. PPI ve eş ekspresyon ağlarının kümelenmesi, ayrımsal ekspresyon ve COBRA analizlerinin sonucunu biyolojik bağlam içerisine yerleştirmek için kullanılmıştır. Ayrımsal ekspresyon analizi, TCGA ve GTEx veritabanlarından alınan gen ekspresyon verisi içerisinde 826 farklı ekspresyona sahip gen (DEG) keşfetmiştir. COBRA yöntemlerinin kullanılabilmesi için gen ekspresyon verisi Recon3D isimli 10,600 reaksiyon içeren jenerik insan metabolik modelini sırasıyla 5,879 ve 5,812 reaksiyon içeren kanser ve sağlıklı dokuya özgü modellere çevirmekte kullanılmıştır. Bu iki modelin metabolik akı profilleri akı örnekleme yöntemiyle keşfedilmiş ve reaksiyon bazında birbiriyle kıyaslanarak anlamlı fark gösteren 1,960 içsel reaksiyon, 338 de değişim reaksiyonu ortaya çıkarılmıştır. Bu reaksiyonların kapsamı, sadece iki model arasındaki akı ortalaması farkı, akı profillerinin iki standart sapmasından büyük olan 250 içsel reaksiyonu ve 14 değişim reaksiyonunu kapsayacak şekilde daraltılmıştır. 55 genin aynı anda DEG oldukları ve anlamlı akı farklılıkları gösteren reaksiyonlarla ilintili oldukları bulunmuştur. Bu genlerin, ve bunlara ek olarak PPI ve eş ekspresyon ağlarındaki en çok DEG ya da metabolik farklılık gösteren geni içeren 5 kümedeki genlerin Gen Ontolojisi (GO) terimleri, Hastalık Ontolojisi (DO) terimleri ve bilinen biyolojik yolaklar ile ilişkilendirilmesi yapılmıştır. Bu ilişkilendirme sonucunda kanser genleri ve hücre-hücre adhezyonu, bağışıklık tepkisi ve reaktif oksijen türleri (ROS) baskısı süreçleri ve diğer süreçler arasındaki ilişkiler ortaya konmuştur.

GRAPHICAL ABSTRACT

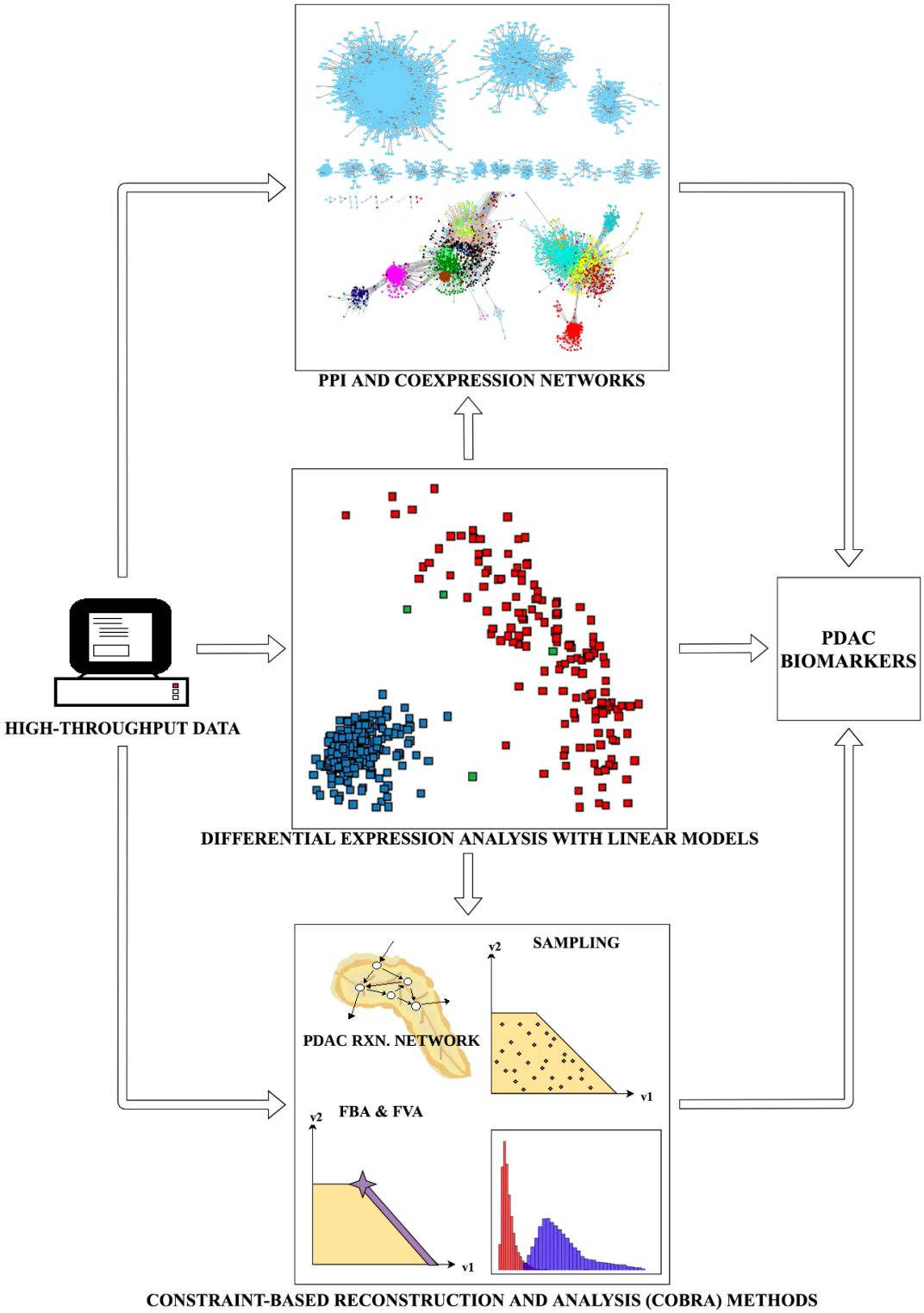

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a_{ij}^{norm}	Normalized graph adjacency between nodes i and j
$cytC$	Cytochrome C
d_{ij}^{norm}	Normalized graph distance between nodes i and j
$DHAP$	Dihydroxyacetone phosphate
f_{q_i}	The value of the q'th arbitrary feature for a node i
$Fe(II)$	Iron (Reduced)
$Fe(III)$	Iron (Oxidized)
g_i	Reaction inclusion penalty for low-expression reaction i (GIMME Algorithm)
H^+	Hydrogen cation
H_2O	Water
k_i	Degree of node i
L	Mean RNA-Seq library size
l_i	Lower flux bound of reaction i
l_{ij}	Product of the edge weights nodes i and j share among their neighbors
M	Median RNA-Seq library size
NAD^+	Nicotinamide adenosine diphosphate (Oxidized)
$NADH$	Nicotinamide adenosine diphosphate (Reduced)
O_2	Oxygen
O_2^-	Superoxide anion
P_k	Probability of randomly finding a node in the graph with degree k
Ri	Chemical reaction i in a metabolic system
RE_i	Gene expression level mapped to reaction i (INIT Algorithm)
$\underline{\underline{S}}$	Stoichiometric matrix containing the coefficients of the mass balance
u_i	Upper flux bound of reaction i

v_i	Reaction rate ("flux") of reaction i
w_i	Reaction "weight" (INIT Algorithm)
x_i	Gene expression level mapped to reaction i (GIMME Algorithm) *or* Index specifying whether metabolite i is included in the model (INIT Algorithm)
y_i	Index specifying whether reaction i is included in the model (INIT Algorithm)
Z	Objective function
Z_0	Optimal FBA objective function value
λ_i	i'th decision variable for an arbitrary LP problem
γ	Proportion of optimal FBA objective function value achieved by another LP approach
μ	Mean
σ	Standard deviation
ω_{ij}	Topological overlap between nodes i and j

LIST OF ACRONYMS/ABBREVIATIONS

ACHR	Artificially Centered Hit-and-Run
AR	Androgen Receptor
BH4	Tetrahydrobiopterin
BioGRID	Biological General Repository for Interaction Datasets
CA19-9	Carbohydrate Antigen 19-9
COBRA	Constraint-Based Reconstruction and Analysis
COX	Cyclooxygenase
CPM	Counts per Million
DEG	Differentially Expressed Gene
DNA	Deoxyribonucleic Acid
DO	Disease Ontology
DSD	Difference-to-Standard-Deviation Ratio
EMT	Epithelial-Mesenchymal Transition
ER	Estrogen Receptor
FBA	Flux Balance Analysis
FC	Fold Change
FDA	Food and Drug Administration
FDR	False Discovery Rate
FPKM	Fragments per Kilobase Million
FVA	Flux Variability Analysis
GIMME	Gene Inactivity Moderated by Metabolism and Expression
GAG	Glycosaminoglycan
GO	Gene Ontology
GSE	Gene Set Enrichment
GSMM/GEM	Genome-scale Metabolic Model
GTEx	The Genotype-Tissue Expression Project

HGNC	Human Genome Organization Gene Nomenclature Comittee
HPRD	Human Proteome Reference Database
INIT	Integrative Network Inference for Tissues
IPMN	Intraductal Papillary Mucinous Neoplasm
KEGG	Kyoto Encyclopedia of Genes and Genomes
LOWESS	Locally Weighted Scatterplot Smoothing
LOX	Lipoxygenase
LP	Linear Programming
MCL	Markov Clustering Algorithm
MCN	Mucinous Cystic Neoplasm
MDS	Multidimensional Scaling
MILP	Mixed Integer Linear Programming
NCBI	National Center for Biotechnology Information
NIH	National Institutes of Health
NO	Nitric Oxide
OMIM	Online Mendelian Inheritance in Man
PAAD	Pancreatic Adenocarcinoma
PanIN	Pancreatic Intraepithelial Neoplasm
PCC	Pearson Correlation Coefficient
PDAC	Pancreatic Ductal Adenocarcinoma
PNET	Pancreatic Neuroendocrine Tumor
PPI	Protein-Protein Interaction
RMF	Required Metabolic Functionalities
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
ROS	Reactive Oxygen Species
siRNA	Small Interacting RNA
TCGA	The Cancer Genome Atlas
TOM	Topological Overlap Matrix
TPM	Transcripts per Million

TTD	Therapeutic Target Database
WGCNA	Weighted Gene Coexpression Network Analysis

1. INTRODUCTION

1.1. The Pancreas & Pancreatic Cancer

The pancreas in humans is an organ with two fundamental areas of function: the first are “exocrine” functions, encompassing the secretion of various digestive enzymes into the small intestine, and the second are the “endocrine” functions, involving hormone secretion into the bloodstream. Most of the exocrine pancreatic tissue (greater than 90%) is made up of acinar cells, which produce digestive juices. The second major component of exocrine tissue are the ductal cells, which line the pancreatic ducts and control the flow of enzymes into the intestine. The endocrine cells are mostly localized in discrete clusters called the “Islets of Langerhans,” secreting hormones directly into blood vessels. Altogether, the cells of the endocrine pancreas make up only around 1-2% of the pancreatic tissue [1].

The term “pancreatic cancer” can be used to refer to any cancer originating in the pancreas, but the tumor causing the disease can initially originate in any kind of pancreatic tissue. About 90% of all cases of pancreatic cancer are found to have started in ductal cells, a comparatively small subset of cells in exocrine tissue. This type of condition is termed a pancreatic ductal adenocarcinoma (PDAC). Along with the much rarer condition of pancreatic acinar adenocarcinoma (which starts in acinar cells and comprises around 2% of pancreatic cancer cases) and several other subtypes of cancer, exocrine tissue cancers make up the vast majority of all cases of pancreatic cancer. Cancers of the endocrine tissue, termed pancreatic neuroendocrine tumors (PNET), make up only around 1% of pancreatic cancers [2]. This study will therefore focus on ductal adenocarcinomas, since they are known to be not only much more common, but also more lethal [3].

Pancreatic cancer, including all subtypes, is a disease which has been responsible for nearly 330,000 deaths in 2012. In the same year, the incidence rate for all pancreatic cancer types was reported to be 338,000. With a very high mortality-to-incidence ratio of 0.98, pancreatic cancer is a very lethal disease. While it is estimated to be the 12th most common type of cancer in men and 11th in women by incidence, pancreatic cancer is the 7th most

common type of cancer by mortality due to this reason [2]. The majority of these deaths are accounted for by ductal adenocarcinomas, the most common subtype. Without surgery, ductal adenocarcinoma patients have overall 5-year-survival rates close to 5%, whereas for a typical PNET patient this rate is nearly 15%. While the removal of most pancreatic tissue along with the duodenum (Whipple Procedure), is a surgical procedure which has been shown to improve 5-year-survival rates to near 10% for ductal adenocarcinoma patients, this is somewhat stymied by the fact that few pancreatic tumors at the time of discovery are deemed operable (around 15-20%) [3]. It is assumed that pancreatic ductal adenocarcinomas are so deadly specifically because they tend to be asymptomatic in early stages, and even in later stages the symptoms of cancer are easy to confuse with those of other pathologies, by which time the cancer may have spread to other tissues [2, 3]. This is exemplified by the fact that the mean patient age at discovery of pancreatic cancers is over 65, which also contributes to the difficulty of operating on pancreatic cancers [2].

1.2. Types of Biomarkers

The term "biomarker" can mean different things depending on context. The US' FDA-NIH Biomarker Working Group [4] defines biomarkers as "A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention." In this context, a disease biomarker can indicate whether a patient has developed or will develop the disease or whether the progression of the disease will be better or worse compared to others. These definitions correspond to subtypes of biomarkers, termed "diagnostic" and "prognostic" biomarkers, respectively, and are not exhaustive of the existing biomarker subtypes [5]. A very desirable quality of a diagnostic biomarker is detection in blood or other bodily fluids, without having to sample the tissue that is suspected to be cancerous. Since a prognostic biomarker is only used after cancer has been diagnosed, such a property is no longer as essential. On the other hand, prognostic biomarkers need to predict disease progression, and not just presence, necessitating them to show level-dependent effects, predicting better or worse progression of cancer at higher or lower biomarker levels. Within this context, where possible, this study will highlight genes related to secretion processes and cell cycle-related processes as potential diagnostic and prognostic biomarkers, respectively.

Drug targets, while not technically biomarkers of either kind themselves, are similarly highly sought after, especially for diseases like pancreatic cancer. However, they have different specifications than either kind of biomarker in that they need to be “druggable” (i.e., it must be possible to chemically affect them inside the body) and non-essential to the normal functioning of healthy tissues and organs.

1.3. Network Medicine

Such a combination of properties as high lethality, difficulty of treatment post-diagnosis, and difficulty of physiological diagnosis makes PDAC a prime target for novel diagnostic or treatment approaches. Network medicine is such a paradigm which encompasses various methods of analyzing the totality of genes, gene transcripts, proteins or metabolic reactions within a cell or tissue of interest as an interconnected network (the genome, transcriptome, proteome, and metabolome, respectively). The basis for this paradigm is the fact that most pathological conditions are caused not only by singular perturbations in the genome or metabolism of the cell, but by a combination thereof. Additionally, such perturbations may manifest themselves in effects “downstream,” in the genes or proteins affected or regulated by the perturbed genes or proteins [6]. Therefore, analyzing such networks may offer more insight into the cells or tissues of interest than analyzing the components in isolation when searching for biomarkers or drug targets. The network medicine paradigm has therefore been gaining much traction in recent cancer research [6].

Three common types of biological networks which are frequently used in research are presented below. Table 1.1 contains descriptions of the networks which this study makes use of to find genes or proteins of interest to diagnosing or treating pancreatic cancer.

Once such networks are constructed, they are amenable to topological analysis, which is the analysis of a network based on its pure graph theoretical properties, stripped of any biological considerations. Subnetworks (termed “modules” or “clusters”) based on such properties can be found within these networks that can offer a more focused view of a smaller set of genes or proteins [7, 8].

Table 1.1. Descriptions of biological network types.

Network	Description
Protein-Protein Interaction (PPI)	Contain a mapping of physical interaction or complex formation between proteins [7].
Coexpression	Contain a mapping of which genes get expressed together relatively often, and which genes are rarely expressed together [9].
Metabolic	Contain a mapping of metabolic reactions for a given organism, and the substrate and product metabolites for each reaction, as well as the protein catalyzing this reaction [7].

To reconcile the graph theoretical significance of genes or proteins with biological significance, these networks can be then annotated with biological significance markers, such as shown below in Table 1.2. Whether these annotations offer a significant theme or meaning for a set of genes or proteins of interest can be determined by an approach called enrichment analysis.

Table 1.2. Descriptions of various annotation types.

Annotation	Description
Known Disease Gene	Annotation with diseases known to be significantly associated with genes or gene clusters.
Known Drug Target	Annotation with drugs known to target this gene or protein.
Gene Ontology	Annotation with a three-part scheme which seeks to classify each gene by the molecular function, cellular component of localization, and metabolic process of its associated protein [10, 11].
Survival	Annotation with whether the expression of each gene determines survival outcomes in cancer patients.

1.4. Metabolic Networks and Constraint-Based Modeling

For metabolic networks, an alternative (non-graph-theoretical) avenue of analysis is "Constraint-Based Reconstruction and Analysis" (COBRA) [12]. This approach entails representing each of the metabolites in the cell or tissue of interest as a steady-state mass balance, with metabolic reactions taking on positive or negative quantities depending on whether they produce or consume this metabolite. These models are often called genome scale metabolic models (GSMMs/GEMs). The approach depends on Flux Balance Analysis (FBA), the simultaneous solution of the mass balances within such a model subject to constraints of reaction reversibility, empirical values of metabolite uptakes, and a metabolic objective that is to be optimized. The solution gives information both on the rates of individual metabolic reactions, and on the general metabolic profile of a cell or tissue. FBA and a suite of related methods thus allow the exploration of metabolic networks not topologically, but as a system of equations [12].

The aim of this study is to find differences between the healthy pancreas and PDAC via these network medicine methods and to discuss their suitability for use in diagnostic or treatment approaches. Firstly, differentially expressed genes (DEGs) between the healthy and cancerous pancreas are found using public domain gene expression data. The DEGs themselves are considered for their diagnostic or therapeutic usefulness before being used alongside other data sources to generate metabolic, PPI and coexpression networks for both the healthy and the cancerous pancreas. The metabolic networks are then compared through a COBRA approach to find diagnostically or therapeutically relevant metabolic differences. Simultaneously, the PPI and coexpression networks are analyzed based on their topological properties. Any topological differences between the cancer and healthy networks are noted. Furthermore, the topological properties of DEGs and any groups of metabolically different proteins identified through the COBRA approach are investigated and compared to the topological properties of the networks at large. Any groups of genes exhibiting significant differences between the cancerous and healthy pancreas are then annotated with terms indicating their biological significance for interpretation. All groups of interrelated genes found by either the PPI or coexpression networks ("clusters" or "modules") are also subject to such annotation, regardless of whether they exhibit differences between the healthy and

cancerous pancreas, to reveal which of these groups contain a statistically significant proportion of genes found previously to be cancer-related or pancreas-related. The value of any genomic, metabolomic, proteomic or transcriptomic difference found is discussed with a focus on relevance as a potential biomarker or drug target. The “Theoretical Background” section will provide information on network medicine concepts and the known biological markings of pancreatic cancer. The “Methods” section will explain how these networks are prepared and what data are collected, in addition to explaining the key mathematical ideas behind some of the analysis methods used. The “Results & Discussion” section includes the differences identified between the cancerous and healthy pancreas, the topological properties of the networks these differences are found in, and the annotation enrichment analyses of these differences. The results are discussed by comparison to established literature. In the “Conclusion” section the study’s strengths and weaknesses are evaluated, its key findings are underlined, and forward-facing recommendations are made.

2. THEORETICAL BACKGROUND

2.1. Pancreatic Cancer Pathways

The development of pancreatic ductal adenocarcinoma (PDAC) has been associated with genetic alterations to several genes that make up the pancreatic ductal cell genome. The development of cancer shows a progression that can take the shape of several precursor lesions, known as pancreatic intraepithelial neoplasms (PanINs), intraductal papillary mucinous neoplasms (IPMNs) or mucinous cystic neoplasms (MCNs) [13–16], but there are some commonalities in each case, such as the dense stroma, or scar tissue, pancreatic lesions tend to cause. This scarring is one of the defining features of pancreatic cancer development, and is thought to be one of the reasons for its drug-resistant properties [17, 18].

In terms of genetic mutations, in over 50% of PDAC cases, mutations were discovered in the oncogene KRAS, and tumor suppressor genes p16/CDKN2A, TP53 and SMAD4 [2].The KRAS and p16/CDKN2A mutations have been typically found in earlier stages of PDAC development than other mutations and are therefore considered initiating events in PDAC development [16, 17].

KRAS has a variety of functions, and the mutation causes its product, the *Kras2* protein, to be constitutively produced [17], upregulating the phosphoinositide-3-kinase (PI3K) pathway [19–21]. The activation of this pathway has downstream effects including the suppression of cell death and changing the shape of the cells [17, 19–21]. KRAS mutations also upregulate the mitogen-activated protein kinase (MAPK) pathway leading to further cell proliferation [19–21]. KRAS mutations are found in almost all cases of ductal adenocarcinoma and activate many other pathways related to cancer development [16, 17]. Downstream from KRAS activation of the MAPK pathway, a tissue developing pancreatic cancer begins to overexpress growth factors and their receptors, such as the epidermal growth factor receptor (EGFR) and the vascular endothelial growth factor (VEGF), promoting further proliferation of the cancerous tissue [16, 19–21].

While the KRAS oncogene signals for the dysregulation of downstream pathways (including that of p16/CDKN2A), the mutations in the tumor suppressor genes remove the cell's natural pathways of inhibiting their own growth and repairing DNA mutations. p16/CDKN2A and TP53 express proteins that control the cell cycle – the growth and division behavior of a cell, and the inactivation of the SMAD4 gene, which occurs in nearly 55% of cases, leads to a loss of cellular control over gene transcription via the transforming growth factor β (TGF-β) pathway [17].

The TGF-β pathway is implicated in another phenomenon termed epithelial-to-mesenchymal transition (EMT). This phenomenon is one of the ways by which cells in healthy tissue, which are differentiated and shaped to perform their metabolic functions wherever they are located, lose their shape and cellular bonds to the rest of the tissue and migrate in the body. In addition to TGF-β pathway activation, EMT is characterized by the loss or underexpression of proteins that make up and regulate epithelial cellular junctions, such as E-cadherin. While this phenomenon is not exclusive to cancer, it is heavily implicated in cancer metastasis, and pancreatic cancer is no exception [16, 17].

2.2. Known Biomarkers in Pancreatic Cancer

Currently, carbohydrate antigen 19-9 (CA19-9) in blood serum is the only diagnostic and prognostic biomarker of pancreatic cancer approved by the USA's FDA. Even so, its median sensitivity was found to be 79%, and its median specificity was 82%. CA19-9's performance is further confounded by the fact that some individuals (found to be around 10% of the US' Caucasian population) naturally do not express the genes that lead to CA19-9 production [18, 22].

Experimental biomarker research is taking many different routes, including detection of the expression levels of cancer-related genes like KRAS or VEGF. Other gene expression-based methods include the detection of GPC1 or PD-L1 expression, which are implicated in EMT and immune T-cell infiltration of cancerous tissue, respectively [18].

Metabolic biomarkers like CA19-9 are also sought after, with a broad spread of different classes of molecules under consideration. One such class of metabolites under consideration are phospholipids, such as phosphatidylcholine, lysophosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine [23–26]. A very long chained lipid molecule made up of these smaller phospholipids named PC-594 is also under close scrutiny [26–28]. Another class of biomarker candidate molecules in pancreatic cancer is sphingolipids [24, 26, 29]. Sphingolipids are also of interest since their varieties have been associated with both increased and decreased cancer proliferation and survival [30–33]. Lastly, smaller, simpler molecules like amino acids, creatine, choline and taurine have been considered as tentative cancer biomarkers as well in works by [23–25, 28, 29, 34, 35].

2.3. Graph Theory & Clustering

2.3.1. Graph Theory

Graph Theory is a branch of mathematics which examines structures made up of nodes connected to each other pairwise by edges, making up what is termed a graph. Such graph structures have topological properties that are independent of the real-world context of their constituents. Some salient properties of graphs must be discussed to contextualize the work done in this study. Two major properties of any graph are the “weightedness” and “directedness” of its edges. For a “weighted graph,” each edge has a certain “weight,” a value denoting the magnitude of the connection or association between two nodes. For an “unweighted graph,” each edge is either existent or nonexistent. A “directed graph” has edges which imply a directional relationship between one node to another, whereas an “undirected graph” has edges that imply a symmetrical relationship between any two nodes [36]. Protein-protein interaction networks are generally conceptualized as unweighted, undirected networks; coexpression networks are weighted, undirected networks; and metabolic networks are weighted, directed networks [37]. The properties of each node as a function of its relationship to the rest of the graph are termed the topological properties of a node. Some topological properties of nodes relevant to this study are defined below in Table 2.1:

Table 2.1. Definitions of node topological properties [36, 38].

Property	**Unweighted Undirected Graph Definition**	**Weighted Undirected Graph Definition**
Degree	Number of edges the node has; number of neighbors.	The sum of the weights of all edges of the node.
Average Shortest Path Length	The average minimum number of edges linking this node to any other node.	The average value of the lowest sum of edge weights along edges linking this node to any other node.
Clustering Coefficient	The proportion of the number of edges this node's immediate neighbors have between each other, to the number of edges they possibly could have.	The proportion of the sum of edge weights between this node's immediate neighbors, to the maximum value of the sum, weighted by the "immediacy" of the neighbors themselves (i.e., edge weights).
Closeness Centrality	The reciprocal of the average shortest path length, average measure of how "close" this node is to every other node.	
Betweenness Centrality	The proportion of shortest paths between any two nodes which pass through this node to the number of shortest paths between any two nodes in the whole graph.	

2.3.2. Clustering

Outside of graph theory, clustering among any distinguishable set of data objects can be used to find similar and dissimilar subsets. Within graph theory, clustering approaches serve to identify subgraphs or "clusters" based on a graph's topology alone. Many different classes of clustering algorithms can extract subgraphs with higher interconnectivity given a larger graph, but the choice of clustering algorithm is non-trivial for several research considerations. The major factor in deciding on a clustering approach is whether nodes are

defined by a set of continuous or discrete features, or by a set of weighted or unweighted edges [39].

If edges are not explicitly defined, and each node is defined by a set of continuous or discrete features, then each node can be considered a point in n-space, where n is the number of features. These nodes' distance to each other in this space can then be calculated using these features and normalized by dividing by the highest distance between any two nodes. Subtracting these normalized distances from 1 yields an "adjacency" (also variously called "closeness" or "similarity") value between any two nodes. The normalized Euclidean distance (henceforth just "distance") and adjacency, d_{ij}^{norm} and a_{ij}^{norm} between nodes i and j is defined as in [39], by writing

$$d_{ij}^{norm} = d_{ji}^{norm} = \frac{\sqrt{\sum_q^n \left(f_{q_i} - f_{q_j}\right)^2}}{\max\left(\sqrt{\sum_q^n \left(f_{q_i} - f_{q_j}\right)^2}\right)} \tag{2.1}$$

$$a_{ij}^{norm} = 1 - d_{ij}^{norm} \tag{2.2}$$

where f_{q_i} is defined as the value of the q'th feature out of n for node i. This process resulsts conceptually in a "fully connected" graph, where each node is connected to each other by an edge with weight equal to the adjacency between two nodes. The distance or adjacency of each node to each other node can be expressed as a two-dimensional matrix. A graph constructed this way is amenable to similarity-based clustering methods that work on the adjacency matrix directly, such as hierarchical, k-means or k-medoids clustering. These methods are noise-sensitive but can be very resource and time-efficient. These methods are also usable in clustering data objects that are not nodes in a graph [39].

The coexpression networks and clusters constructed in this study with the WGCNA algorithm are an example of such a process, using a method called the Topological Overlap Matrix (TOM) to calculate similarity scores and hierarchical clustering to form clusters [40, 41]. In this approach, each gene is a node, with its expression in each tissue sample being a

feature. The topological overlap ω_{ij} between nodes i and j is related to the adjacency a_{ij}^{norm} and is calculated as in [41] by first defining

$$k_i = \sum_{u \neq i} a_{iu}^{norm} \tag{2.3}$$

$$l_{ij} = \sum_{u \neq i,j} a_{iu}^{norm} a_{uj}^{norm} \tag{2.4}$$

where k_i is the degree or connectivity of each node i in the sense of adjacency, and l_{ij} is the product of the edge weights nodes i and j share among their neighbors. l_{ij} can be therefore considered a measure of the extent to which nodes i and j share neighbors. ω_{ij}, the topological overlap between nodes i and j, is then calculated as

$$\omega_{ij} = \frac{l_{ij} + a_{ij}^{norm}}{\min\{k_i, k_j\} + 1 - a_{ij}^{norm}} \tag{2.5}$$

and gives a similarity measure that captures more information about the closeness of nodes in the neighborhood of i and j compared to the adjacency a_{ij}^{norm} [41]. All aforementioned clustering approaches that work on the adjacency matrix are also applicable to the TOM.

If nodes are defined only by the edges they have between themselves and other nodes, these nodes' relationship to others can only be understood in the context of these edges, not as distances in space. Therefore, flow-based clustering approaches such as the Markov Clustering Algorithm (MCL) that simulate "flow" along edges in the graph are used to find groups that are better connected. These clustering methods are robust but are quite time and resource intensive [39, 42, 43].

If nodes are defined both by a set of features and by edges, either approach is possible. Furthermore, it is possible to convert either definition of nodes into the other if one desires to use a specific clustering algorithm. However, this process usually requires a sufficiently high amount of time and processing power to discourage such efforts [39].

Heuristic-based clustering algorithms are another alternative to those outlined above, using node-wise topological properties of neighboring nodes to assign nodes with similar topological properties in a local region to the same cluster. These have the disadvantage of potentially not searching every node in the graph [44–46].

2.4. Network Medicine & Biological Networks

The advent of network medicine as a paradigm owes much to the graph theoretical work done to interpret biological networks as a generalized concept. In a review by Barabasi et al., some of the key features of general biological networks are summarized. One of these features of general biological networks is their "scale-free topology". A "scale-free" network, as opposed to a "random" network, is not made up of nodes randomly connected to each other, but has very few well-connected nodes, with a majority of sparsely connected nodes surrounding them. The degree distribution of such networks can be modeled by a power law as in

$$P_k = ak^b \tag{2.6}$$

where the probability of randomly finding a node in the graph with degree k is P_k, and a and b are arbitrary constants [6, 37].

It is theorized that this property affords robustness to protein-protein interaction networks, where the removal of a few nodes from the network, or even the removal of one well-connected node (a "hub"), is not sufficient to disconnect major portions of the network as long as most of the hubs remain within the network [7, 37]. This property of being "scale-free" is termed thus because the general shape, degree distribution, and other properties of such a network are independent of the number of nodes in the network. Additionally, thanks to the presence of such hubs, biological networks exhibit something called the "small world effect", which is simply the fact that any pair of nodes are likely to be separated by a short path through the network. This is why a perturbation of one node (e.g. a point mutation in one gene) has the potential to affect the whole network. Lastly, biological networks tend to show a high degree of clustering, containing subregions where nodes are densely connected

to each other. This implies that such regions may have some common reason for association, such as to form specific protein complexes in a PPI network, or to encode a set of proteins for the same biological pathway in a gene coexpression network [6, 7, 37]. The identification of such related modules is an important procedure to ascribe physical meaning to network phenomena.

In cancer research, genes found to be mutated or deleted in disease are found to be on average much better connected than the rest of the genes in the network and therefore much more likely to be hubs in either coexpression or PPI networks [6]. Goh et al.'s research indicates that this is not a broadly applicable assumption for disease genes [47]. Using disease gene data from the Online Mendelian Inheritance in Man (OMIM) database, Goh et al. show that the apparent high degree of disease genes is due to a small subset (22%) of disease genes overlapping with essential genes having very high degree. The non-essential disease genes, by contrast, have a small number of neighbors. They make the caveat, however, for somatic gene mutations as in the case of cancers, the disease-associated genes are still likely to have high degree [47].

Beyond individual topological properties of genes or proteins, the clustering approach in network medicine attempts to contextualize individual gene or protein properties as a phenomenon arising from the interrelated action of several related genes or proteins. Van Dam et al. explain the use of clustering to discover how these individual differences between phenotypes are regulated or driven by overarching phenomena. They also outline the guilt-by-association approach in coexpression networks, which implicates genes clustered together are all likely to be associated with a phenotype (like cancer) if many nodes in the cluster are known to be associated [48]. Gillis and Pavlidis caution against the use of this method in PPI networks specifically, stating that the majority of phenotypic change comes not from the interaction of all proteins in a cluster, but by a few specific interactions that nevertheless have great effect [49].

Especially for PPI networks, these kinds of qualifications and objections are more common, such as in research by Mora and Donaldson: The researchers state that while previous studies had identified drug targets as being high-degree nodes of PPI networks,

they have found node degree to be a poor predictor of a node being a drug target. Additionally, they state that the merging of PPI networks tends to increase the overrepresentation drug targets among high-degree nodes. They also concede that cancer drug targets seem to be well-predicted by higher degree nodes [50]. One additional study on the inadequacies of node degree as a topological property for predicting drug targets by Peng & Schork notes that the highest-degree nodes in a PPI network tend to be too central to attack with chemotherapeutic drugs, potentially being highly toxic to patients. These authors suggest "eigenvector centrality", which for any given node is proportional to the sum of the degrees of that node's neighboring nodes, as a predictive property for drug targets. This kind of topological property is likely to find nodes in the network that connect hubs, instead of the hubs themselves. They furthermore suggest using gene expression level data to prune PPI networks for tissue-specificity. By pruning nodes whose associated genes are weakly expressed in the tissue of interest, highly unlikely protein-protein interactions can be ruled out, affecting the topological analysis results significantly [51]. Işık et al. in a recent study have proposed a novel topological property termed "local radiality" for nodes in PPI networks, requiring additional gene expression data, which has been shown to perform especially well in predicting drug target nodes with low degree. Local radiality of a node is proportional to the number of shortest paths this node is on, given that these shortest paths also contain perturbed nodes (which are determined from the gene expression) [52].

2.5. Linear Programming and Related Methods in Constraint-Based Analyses

Linear Programming (LP) is an approach to solving optimization problems with the following specifications as discussed in [12, 53]:

- The optimization objective is a linear function of constants c_i variables λ_i.
- The constraints on the optimization problems are all linear functions of variables λ_i describing equalities or inequalities.
- There exist n decision variables λ_i, with m equality constraints, where $m < n$.

Therefore, these problems can be formulated as

$$\begin{aligned} &maximize\ Z = c_1\lambda_1 + c_2\lambda_2 + \cdots + c_n\lambda_n = \underline{c}^T\underline{\lambda} \\ &such\ that\ \ a_{1,1}\lambda_1 + a_{1,2}\lambda_2 + \cdots + a_{1,n}\lambda_n = b_1 \\ &\qquad\qquad a_{2,1}\lambda_1 + a_{2,2}\lambda_2 + \cdots + a_{2,n}\lambda_n = b_2 \\ &\qquad\qquad\qquad\qquad \vdots \\ &\qquad\qquad a_{m,1}\lambda_1 + a_{m,2}\lambda_2 + \cdots + a_{m,n}\lambda_n = b_m \\ &and \qquad \underline{\lambda} \geq 0 \\ &where \quad a_{j,i}, b_j, c_i\ are\ constants. \end{aligned} \tag{2.7}$$

The linear equality constraints in Equation (2.7) can be expressed as $\underline{\underline{A}}\,\underline{\lambda} = \underline{b}$, with $\underline{\underline{A}}$ being an m-by-n matrix of arbitrary coefficients $a_{j,i}$. It can be shown that the constraints in Equation (2.7) constrain the n-dimensional solution space for all λ_i into a convex region. Due to the region's convexity, a global maximum solution to the linear programming problem can be guaranteed. Furthermore, for any such non-empty convex region (i.e. there is at least one point which satisfies all constraints), this optimum solution is guaranteed to occur on the vertices of the region for an optimization objective that is also a linear function of λ_i. Dantzig's simplex method, which checks the vertices of this region [53], and Karmarkar's interior point method, which traverses the interior of this region [54], are the two algorithms upon which modern algorithms for the solution of such problems are based.

2.5.1. Flux Balance Analysis (FBA)

Flux Balance Analysis is essentially the application of linear programming methods to a system of metabolic reactions to estimate the rate of each reaction simultaneously, subject to an optimization objective and linear constraints [12, 55–57]. Firstly, the entirety of metabolic reactions in a system are defined, as a set of reactions like in this toy example:

$$
\begin{aligned}
&R1\colon A \leftrightarrow B + C \\
&R2\colon B + 2C \rightarrow D \\
&R3\colon A \leftrightarrow \\
&R4\colon D \rightarrow \\
&\vdots
\end{aligned}
\tag{2.8}
$$

Here, the one-sided arrow implies an irreversible reaction, whereas the two-sided arrow implies a reversible reaction. The reactions $R3$ and $R4$ are termed "exchange" reactions. These tend to signify the uptake or secretion of a metabolite into the defined system – generally a cell or tissue. Whereas $R3$ shows that A can be taken up or secreted, $R4$ shows that D can only be secreted. Special cases of exchange reactions termed "demand" or "sink" reactions exist. While demand reactions represent intracellular irreversible accumulation of a metabolite, sink reactions are reversible and indicate metabolites that take part in metabolism, but cannot be produced by the existing metabolic model, either because they are produced by non-metabolic processes or because the model is incomplete [12, 55–57]. Assuming each reaction Ri occurs at rate v_i, and that the biological system is at steady-state, these reactions can be translated into the set of mass balance equations

$$
\begin{aligned}
&-v_1 - v_3 = \frac{d[A]}{dt} = 0 \\
&v_1 - v_2 = \frac{d[B]}{dt} = 0 \\
&v_1 - 2v_2 = \frac{d[C]}{dt} = 0 \\
&v_2 - v_4 = \frac{d[D]}{dt} = 0 \\
&\vdots
\end{aligned}
\tag{2.9}
$$

These mass balance equations can be expressed as $\underline{\underline{S}}\,\underline{v} = \underline{b}$, where $\underline{v}$ is a vector of reaction rates, termed "fluxes," $\underline{b}$ is a vector of all zeros, corresponding in length to the number of metabolites, and $\underline{\underline{S}}$ is the stoichiometric matrix

$$\underline{\underline{S}} = \begin{bmatrix} -1 & 0 & -1 & 0 & \cdots \\ 1 & -1 & 0 & 0 & \\ 1 & -2 & 0 & 0 & \cdots \\ 0 & 1 & 0 & -1 & \\ \vdots & & \vdots & & \ddots \\ & & & & \end{bmatrix} \tag{2.10}$$

which contains the coefficients of the mass balances.

These reactions all have upper and lower bounds on their fluxes that can be set based on empirical observations of some fluxes. In the absence of such information, the upper and lower bounds are generally set to an arbitrarily large number (typically 1000 and -1000) as in [12, 55, 56, 57], representing near-infinite flux. If a reaction is irreversible, its lower bound is set to zero flux. These upper and lower bounds can be represented as

$$l_i \leq v_i \leq u_i \; for \; all \; v_i \tag{2.11}$$

or equivalently in vector form, as

$$\underline{l} \leq \underline{v} \leq \underline{u}. \tag{2.12}$$

An objective function is defined as a linear combination of reaction fluxes or lumped into the flux of a single reaction that needs to be maximized. This reaction is frequently ATP production, biomass production or biomass maintenance. The biomass reactions are generally formulated as demand reactions that are the linear combination of several other reaction fluxes, and their coefficients are determined empirically. In illustrative terms, this indicates the disappearance of metabolites at a certain rate to keep the cell growing or maintaining its core functions. Thus, instead of defining the objective function as

$$Z = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = \underline{c}^T \underline{v} \tag{2.13}$$

and assigning c_i for each reaction, a reaction $Rbiom$ can be defined, updating $\underline{v}$, $\underline{u}$, $\underline{l}$, $\underline{b}$, and $\underline{\underline{S}}$ accordingly:

$$\begin{gathered} Rbiom: p_A A + p_B B + p_C C + p_D D \ldots \rightarrow \\ Z = c_{biom} v_{biom}. \end{gathered} \tag{2.14}$$

The final form of the FBA problem formulation is as follows:

$$\begin{aligned} &maximize\ Z = \underline{c}^T \underline{v} \\ &such\ that\ \ \underline{\underline{S}}\,\underline{v} = 0 \\ &and \qquad \underline{l} \leq \underline{v} \leq \underline{u}. \end{aligned} \tag{2.15}$$

This problem can be reduced to the linear programming problem and solved with any of the associated algorithms, finding an optimal flux distribution for the cell or tissue.

2.5.2. Flux Variability Analysis (FVA)

The linear programming solution algorithms discussed here guarantee the optimality of their solutions but cannot guarantee their uniqueness [12, 53, 58]. There may be other possible flux distributions that give the same optimal value of the objective function. To overcome this difficulty, an approach called Flux Variability Analysis (FVA) is employed [12, 55, 56, 58]. Given an optimal FBA solution Z_0 has been found for a system, the FVA problem is formulated as follows for each reaction:

$$\begin{aligned} & maximize/minimize\ v_i \\ & such\ that \quad \underline{\underline{S}}\,\underline{v} = 0 \\ & and \quad \underline{l} \le \underline{v} \le \underline{u} \\ & and \quad \underline{c}^T \underline{v} \ge \gamma Z_0 \end{aligned} \tag{2.16}$$

Here, γ is a parameter between 0 and 1 indicating the proportion of the optimal solution the FVA solutions need to achieve. When γ is 1, the solutions of this problem all achieve the FBA optimum. It is generally assumed that reactions whose range of possible fluxes is narrow are more critical to the functioning of the metabolic network. The FVA problems can also be solved using LP algorithms as mentioned earlier in this section.

2.5.3. Minimization of Total Flux

Biological models tend to contain loops for which the linear programming algorithms cannot guarantee realistic results. Without further corrective steps, the FBA flux profile will fail to find realistic fluxes for reactions that are in such loops. Furthermore, some loops result in the production of thermodynamically infeasible products. As biological models get larger, the likelihood that such loops exist grows significantly. While approaches such as Schellenberger et al.'s loop law deal with this problem explicitly, they require the solution of a mixed-integer linear programming (MILP) problem. These problems are essentially LP problems where some of the decision variables can only take on discrete integer values [59]. MILP problems tend to scale poorly with large models, so an alternative approach can be taken to minimize the absolute value sum of all reaction fluxes in $\underline{v}$ (also termed the "1-norm" or the "taxicab norm" of $\underline{v}$).

Given an optimal FBA solution Z_0 has been found for a system, this minimization can be formulated as in [56] by solving

$$\begin{aligned} &minimize \sum_{i}^{n} |v_i| \\ &such\ that\ \underline{\underline{S}}\ \underline{v} = 0 \\ &and \qquad \underline{l} \leq \underline{v} \leq \underline{u} \\ &and \qquad \underline{c}^T \underline{v} = Z_0. \end{aligned} \tag{2.17}$$

This formulation can also be converted into an LP problem and solved using dedicated LP algorithms.

2.5.4. Finding Tissue-Specific Models from Genome-Scale Metabolic Models (GSMMs)

2.5.4.1. Mapping. The approaches mentioned here both use LP principles to extract a tissue-or-context-specific submodel from a general GSMM, using empirical gene expression or protein abundance data to provide tissue specificity. If gene expression data are used, the expression of each gene must be mapped onto the reactions in the model, accounting for reactions that depend on multiple genes or genes that transcribe enzymes for multiple reactions. To achieve this, the general metabolic model needs to contain gene-reaction association rules, annotating each metabolic reaction with a string of gene IDs, separated by AND or OR statements. An AND statement implies that the reaction's expression level is equal to the minimum level among genes or gene groups separated by AND. One possible parsing of an OR statement is that it implies that the reaction's expression level is equal to the sum of expression levels among genes or gene groups separated by OR. Another possible parsing is that an OR statement implies that the reaction's expression level is equal to the maximum level among genes or gene groups separated by OR. This second method assumes that many reactions' activity is driven by a single gene and saves a lot of computation time. If data are not available on a gene, its expression level is set at -1. If parsing OR statements

such that the expression level of a reaction is the sum of genes with OR statements, an expression level of -1 is considered to be 0 [56].

2.5.4.2. The GIMME Algorithm. Gene Inactivity Moderated by Metabolism and Expression (GIMME) was first described in Becker & Palsson's [60] work as an LP approach to finding a set of reactions in a metabolic model that were consistent with gene expression data while still achieving Required Metabolic Functionalities (RMF).

The algorithm first uses FBA to maximize RMF – expressed as a linear objective function – in the non-tissue-specific model. Subsequently, given an optimal FBA solution Z_0 has been found, the GIMME LP problem is formulated as

$$\begin{aligned} &minimize \sum_i^n g_i|v_i|\text{s} \\ &such\ that\ \underline{\underline{S}}\,\underline{v} = 0 \\ &and \quad \underline{l} \le \underline{v} \le \underline{u} \\ &and \quad \underline{c}^T\underline{v} \ge \gamma Z_0 \\ &where \quad g_i = \{x_{cutoff} - x_i, if\ x_{cutoff} > x_i \\ &\qquad\qquad 0, if\ x_{cutoff} \le x_i\} \end{aligned} \tag{2.18}$$

Here, γ is a parameter between 0 and 1 indicating the proportion of the optimal solution the FVA solutions need to achieve. When γ is 1, the solutions of this problem all achieve the FBA optimum. The vector $\underline{x}$ contains the expression levels mapped to each reaction, and x_{cutoff} is an adjustable parameter that is usually selected as some quantile of $\underline{x}$. Selection of x_{cutoff} determines how heavily a reaction is penalized for having a low expression level.

2.5.4.3. The INIT Algorithm. INIT, or Integrative Network Inference for Tissues, is developed by Mardinoglu et al. [61] as a method of tissue-specific network construction method capable of simulating human tissues in that it can accumulate or secrete a broad

range of byproducts and work in the absence of a biomass function given the low proliferative rates of human cells compared to microbes. INIT poses the MILP problem

$$\begin{aligned} &maximize\ \sum_i^n w_i y_i + \sum_j^m x_j \\ &where \quad y_i = \{0, reaction\ excluded \\ &\qquad\qquad 1, reaction\ included\} \\ &and \quad\ x_j = \{0, metabolite\ excluded \\ &\qquad\qquad 1, metabolite\ included\} \\ &and \quad\ w_i = 5 \log\left(\frac{RE_i}{RE_{i,baseline}}\right) \end{aligned} \tag{2.19}$$

where w is a vector of weights for each reaction, where RE_i indicates the reaction expression of reaction i in the tissue of interest, as explained in subsection 2.5.4.1. $RE_{i,baseline}$ indicates the reaction expression of reaction i across all tissues. Please note that this MILP does not necessitate the solution of any FBA or FBA-like problem, only maximizing the metabolites and the sum of weights of reactions included.

3. METHODS

3.1. Data Import and Preparation

3.1.1. Gene/Protein Nomenclature Harmonization

Genes or gene products in the various networks constructed in this study were linked together, across databases and analysis tools, by a unique name or identifier. The names which connect genes across various networks were downloaded from Human Genome Organization Gene Nomenclature Committee (HGNC) [62], which is maintained by the European Bioinformatics Institute. The HGNC database links nearly 42,000 genes and proteins across many common nomenclature standards, including Ensembl Gene IDs, UniProt IDs, RefSeq IDs, NCBI Gene IDs, and HGNC gene symbols. The size and scope of this database determined the maximal set of genes and gene products this study can contain, since cross-database and cross-network analysis is impossible for genes and proteins whose names or symbols are not covered in the HGNC database.

3.1.2. Gene Expression Data

RNA sequencing (RNA-Seq) data regarding gene expression were downloaded from two sources to compare the cancerous transcriptome to the healthy transcriptome. Pancreatic RNA-Seq data from The Cancer Genome Atlas (TCGA), maintained by the United States' NIH was filtered for the solid pancreatic tissue samples of the TCGA-PAAD project. This dataset comprised 175 samples expressing some 60,843 genes, out of which 23 samples were cut for likely not meeting study protocol. Out of the remaining 152 samples, 4 samples were healthy cancer adjacent tissues. This led to 148 TCGA-cancer samples and 4 TCGA-healthy samples [63, 64]. RNA-Seq data from 249 healthy pancreas samples, covering 56,202 genes, were downloaded from The Genotype-Tissue Expression Project (GTEx), maintained by the Broad Institute at MIT and Harvard [65].

Both sources offer data in both raw read count format, and in a per-million-reads format. The per-million-reads format TCGA offers is "Fragments Per Kilobase Million (FPKM)", whereas GTEx offers "Transcripts Per Million (TPM)." The FPKM-formatted TCGA data were converted into TPM for consistency. For any gene i, its TPM expression level TPM_i can be found as

$$TPM_i = \frac{FPKM_i}{\sum_j FPKM_j} \times 10^6, \tag{3.1}$$

dividing its FPKM expression level $FPKM_i$ by the sum of the FPKM expression levels of all genes j, and multiplying by one million.

For both the raw read count and TPM data, the gene entries common to both GTEx and TCGA databases comprised 53,196 genes. Both TCGA and GTEx offer clinical data for the patients these samples were harvested from. The age and sex data of each patient were also downloaded to help remove confounding effects during the identification of differentially expressed genes.

3.1.3. PPI Data

Protein-protein interaction data are notoriously incomplete due to existing databases being maintained primarily by manual curation, interaction detection protocols being generally unrepresentative of biological conditions, and many proteins simply being not “popular” (e.g. because these proteins are not in any disease pathways) enough to be studied [6, 7].

To have a complete yet reliable reconstructed network, three databases were merged and then filtered based on confidence. These databases are the Human Proteome Reference Database (HPRD) [66–68], Biological General Repository for Interaction Datasets (BioGRID) [69], and European Bioinformatics Institute’s IntAct [70]. These databases hold data as a table of binary interacting pairs of proteins, along with additional information regarding each interaction, depending on database. This type of information is automatically

amenable to being turned into a network via software such as Cytoscape [71], as was done in this study.

HPRD catalogues around 39,240 human protein-protein interactions (containing two different interacting proteins) from nearly 9,617 unique human proteins. The confidence measure this database uses is the number of corroborating publications. Interactions corroborated by more than one publication, however, make up less than 12% of the dataset. HPRD being the smallest database used, it was included in the final PPI network without any reliability filtering to not further reduce its size [66–68].

BioGRID contains 395,219 protein-protein interactions comprising 17,257 human proteins whose physical interaction was experimentally observed. Additionally, it contains 1,787 “genetic interactions”, which are interactions of one protein with the gene encoding another protein [69]. These were removed from the database prior to any further processing. As a confidence measure, BioGRID reports the unscaled confidence score given in the publication where the interaction was first reported. Thus, confidence scores are not consistent across publications, and some interactions are not scored at all. To ensure relatively high-confidence interactions are maintained, the BioGRID database were treated as follows: For interactions with confidence scores, where publications report confidence in a standardized 0-to-1 scale, only interactions above a 0.5 level of confidence were kept in the data. Where publications use a non-standard confidence scale, only interactions with an above-average confidence score were kept. These scored interactions make up nearly 30% of the database, and such a pruning procedure removed about 5% of these, amounting to 1.5% of the whole database being pruned this way. The unscored interactions which make up a large majority of the database were fully included in the final PPI network, assuming that the proportion of low-confidence interactions among the unscored interactions is the same as that among the scored interactions, such that the less reliable interactions admitted into the PPI network make up less than 3.5% of the BioGRID dataset. This reduced the number of interactions admitted into the PPI network to around 368,982 and the number of proteins to 17,097.

IntAct contains 351,497 human protein-protein interactions comprising 23,458 proteins. IntAct assigns a confidence score called MIscore to each interaction [70]. MIscore is an aggregate scoring scheme which considers three factors on a scale from 0 to 1 [72]:

- The method of observation of the interaction: The more definitive the method, the higher score. For instance, “X-ray crystallography” is scored higher than “experimental interaction evidence”.
- The type of interaction: The closer the interaction has been recorded, the higher score. For instance, “direct interaction” is scored higher than “co-localization”.
- The number of publications corroborating the interaction.

The database was filtered to keep only interactions with MIscore over 0.5, reducing the size of the database by 65%. The remaining 122,952 interactions corresponding to 10,251 proteins were admitted to the PPI network.

The final human PPI network data pool therefore comprises nearly 18,000 proteins, making over 380,000 interactions with other proteins, before further analysis. This process has so far indicated that BioGRID is the largest contributor to the PPI network, containing over 90% of all proteins.

3.1.4. Healthy Pancreas and Pancreatic Cancer “Seed Genes”

Gene IDs in the form of HGNC gene symbols for “seed genes” were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG). These “seed genes” are a set of known genes whose protein products are known to be significantly enriched in either the healthy pancreas compared to other healthy tissues (“healthy seed genes”), or in the cancerous pancreas compared to the healthy pancreas (“cancer seed genes”). The 74 cancer seed genes were downloaded from KEGG’s pancreatic cancer dataset (“hsa05212”) and the 96 healthy seed genes were downloaded from KEGG’s pancreatic secretion dataset (“hsa04972”). The secretion dataset was chosen because it highlights the proteins most involved in the process of transporting exocrine pancreatic secretions from the acinar cells

through ductal cells and into the lumen of the pancreas [19–21]. A table of these seed genes is included in the appendix, Table A.1 and Table A.2.

3.1.5. Drug Interaction Data

The Therapeutic Target Database (TTD), a database matching 5,059 approved drugs to 215 approved human drug target genes or proteins was downloaded and pruned to include only the 159 anticancer agents [73]. These 159 drugs were found to target 83 unique human proteins. These drugs, as well as their current indications and targets are shown in Table A.3 of the appendix.

3.1.6. Recon3D

The most recent distribution of the Recon3D Human Metabolic Model, which is a version of the Recon3D Human Metabolic Reconstruction modified for amenability to flux balance analysis was downloaded from the Virtual Metabolic Human website, vmh.life [74].

3.1.7. Metabolic Uptake Data

For a constraint-based analysis of the metabolic models in the study, the 140 metabolic uptake and secretion constraints were obtained from the work of Jain et al. [75]. These same constraints were chosen due to being used in Opdam et al.'s [76] work in reconstructing tissue-specific networks. The reactions corresponding to the uptake and secretion of these metabolites in Recon3D were found to be incomplete, lacking reactions for the secretion of phosphocholine, phosphoethanolamine, asymmetric diethylarginine (ADMA), hippurate and salicylurate, and the reaction for maleate uptake. These reactions were not added to the model but were not found to cause a metabolic deficiency. Jain et al.'s data are reported in the appendix Table A.4.

Since the data came from analysis of non-pancreatic cancer cell lines, the assumption was made that approximate uptake and secretion rates could be inferred from this data, at

best. To this end, the uptake and secretion rates across the various cancer cell lines were clustered using hierarchical clustering, eliminating cell lines that were clearly outliers. To do this, R's default hclust function was used with a visually determined clustering tree cut height of 800.

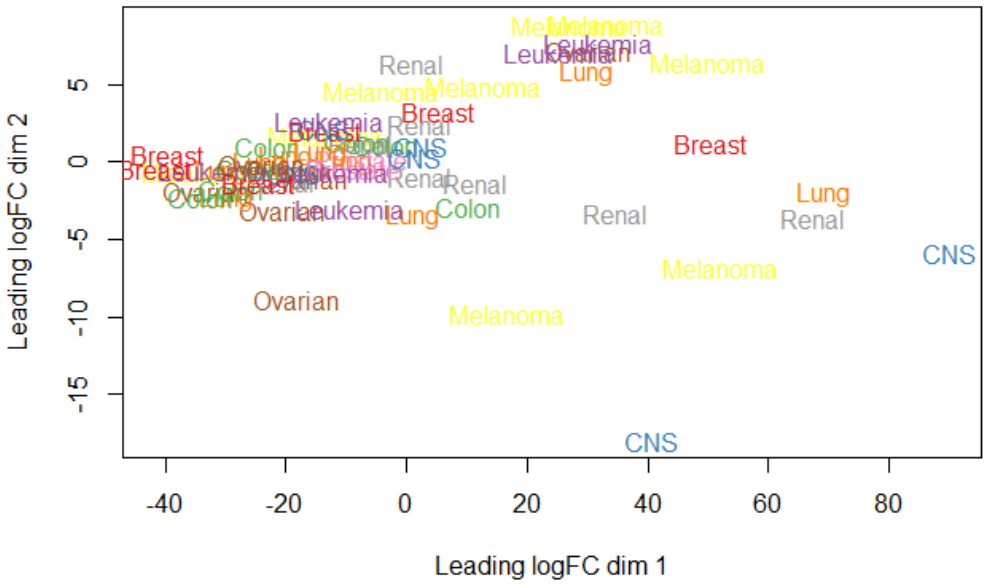

Figure 3.1. Uptake/secretion data from each cell line, MDS plot.

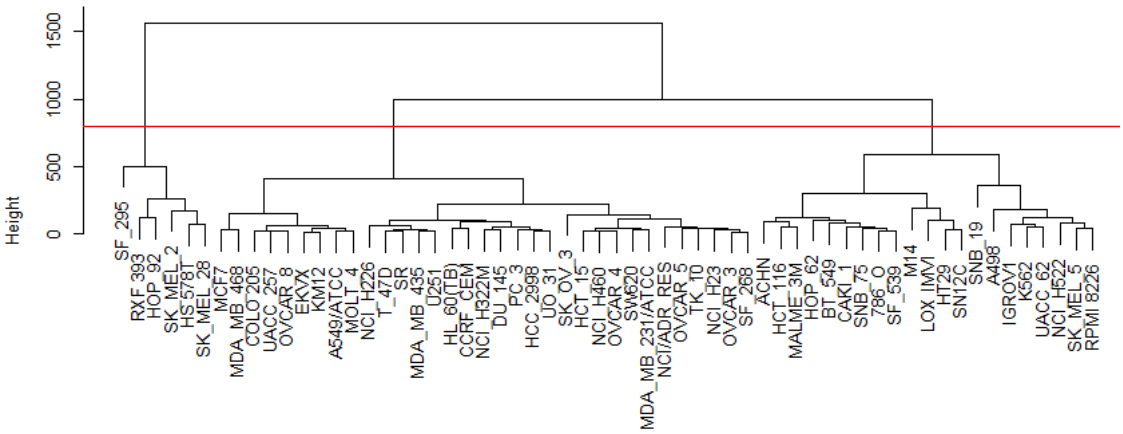

Figure 3.2. Clustering dendrogram of uptake/secretion data from each cell line, outliers' cutoff shown in red.

An MDS plot of the unmodified data, as well as the associated hierarchical clustering dendrogram are shown in Figure 3.1 and Figure 3.2, respectively.

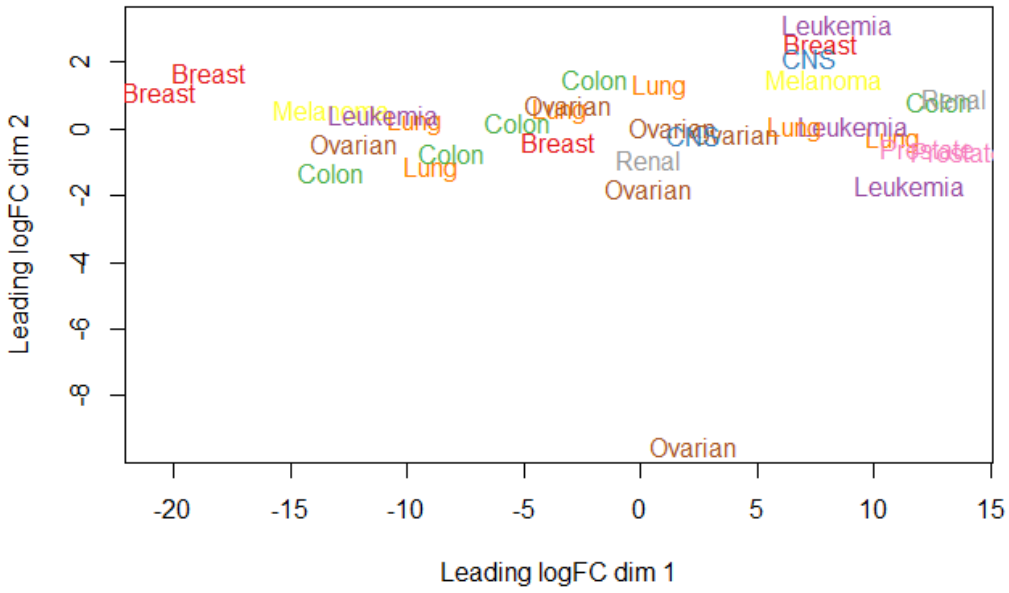

Figure 3.3. Uptake/secretion data from each cell line, MDS plot with outliers removed.

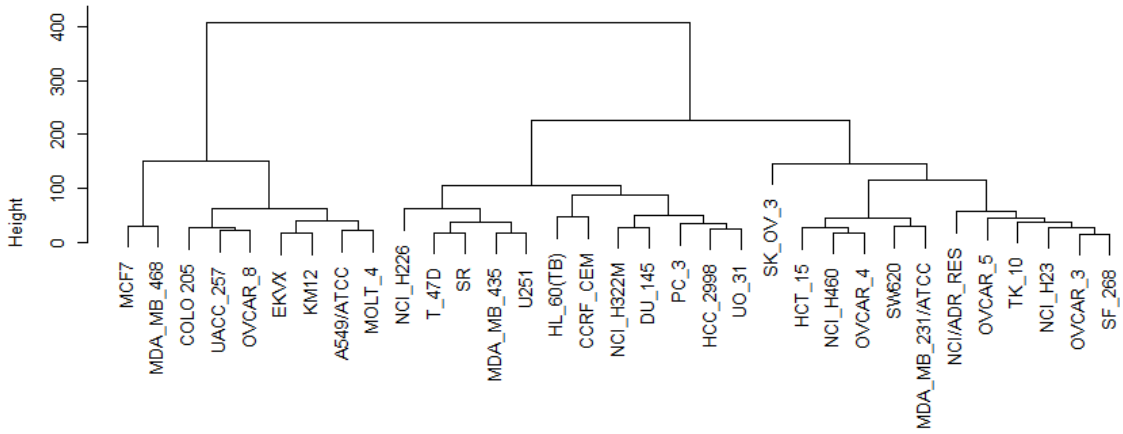

Figure 3.4. Clustering dendrogram of uptake/secretion data from each cell line, outliers removed.

Figure 3.3 and Figure 3.4 represent the MDS plot and clustering dendrogram after some cell lines were removed as outliers. The rest of the data were averaged together and scaled to the nearest power of 10, to indicate the rough order-of-magnitude of uptake and secretion. To avoid over-constraining the model, uptakes and secretion rates with orders-of-magnitude below 0.001 fmol/cell/h were set to 0.001 fmol/cell/h.

It was further found that Jain et al.'s constraints did not contain human-essential lipids, linoleic and linolenic acid, so these were added to the data with uptake rates equal to the minimum order-of-magnitude uptake rate value among amino acids, namely 0.1 fmol/cell/h [77, 78]. The table of uptake/secretion flux constraints used in this study are available in Table A.4 in the appendix.

3.2. Differential Gene Expression

The R software packages edgeR and limma was used to find the most significantly differentially expressed genes (DEGs) between the healthy and cancer groups [79–81]. To remove the confounding effects from this analysis, a design matrix was generated which classified each sample by the relevant experimental identifier (cancer/healthy) in addition to all the confounding identifiers. In this case, the confounding identifiers were dataset (the different processing of TCGA and GTEx, irrespective of cancer), patient age in bins of 10 years, and sex. The gene expression data in terms of raw read counts obtained from GTEx (healthy) and TCGA (mostly cancer, 4 healthy samples) were preprocessed by the filterByExpr function in the edgeR package, with the given design matrix. This function removes genes whose log2-count-per-million (logCPM) expression levels are very low and therefore tend to exhibit a proportionally large variance. Subsequently, the data were converted to units of logCPM and normalized using the "Trimmed Mean of m-values" (TMM) method [82]. The heteroskedasticity of the normalized data was reduced using a pipeline of the calcNormFactors function in the edgeR package and the voom function in the limma package.

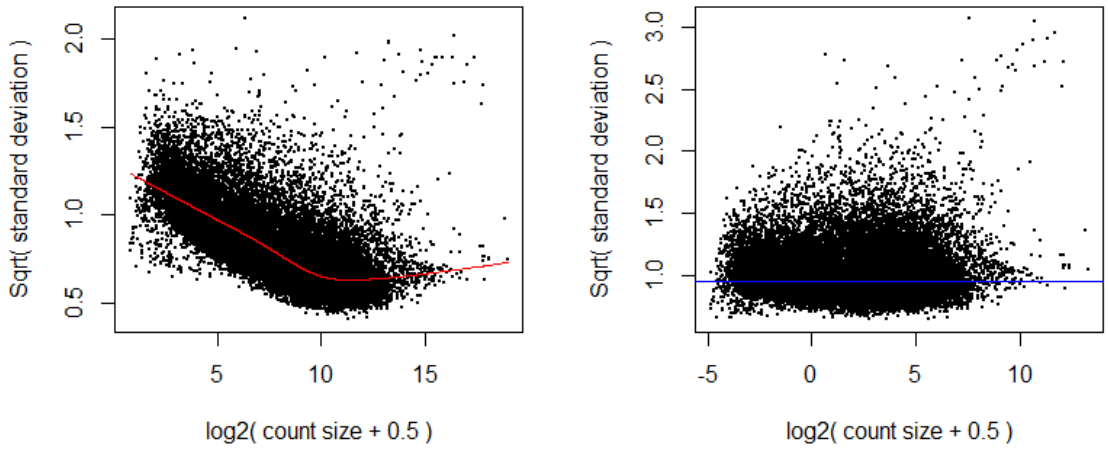

Figure 3.5. Expression mean-variance scatterplots; red and blue lines are the LOWESS fits to the data without and with variance stabilization, respectively.

The voom function [79] estimates gene-wise variances by fitting a Locally Weighted Scatterplot Smoothing (LOWESS) curve to the mean-variance trend and assigning weights to expression data equal to the inverse squared standard deviation estimated by the LOWESS curve. These weighted expression values are then used in the linear model. The mean variance trend of the data before weighting, as well as the effective mean-variance trend of the weighted data are shown in Figure 3.5.

The normalized, variance-stabilized expression data were analyzed via an empirical Bayes procedure consisting of the limma lmFit and eBayes functions, accounting for data heteroskedasticity. The topTable function in the limma package was used to report the results of the empirical Bayes procedure using the Benjamini-Hochberg correction [80] for multiple testing at a false discovery rate threshold (FDR) of 1%.

3.3. Metabolic Networks and Constraint-Based Modelling and Analysis (COBRA)

The MATLAB software package COBRA Toolbox [56, 83] was used in the steps outlined here. The solver & optimizer interface Gurobi [84] was used for calculations.

3.3.1. Objective and Constraints Setup

The Recon3D Metabolic Model's "biomass_maintenance" objective represents the consumption of metabolites in maintaining the routine activities of a typical human cell. This was set as the objective function for the remainder of analyses here, representing the general metabolic activity and survival of the tissue, using COBRA Toolbox's changeObjective function. The averaged order-of-magnitude estimates of uptake and secretion adapted from Jain et al. [75], including those of linoleic and linolenic acid, were supplied to the model as maxima and minima of uptake and secretion to the tissue. The default unit of the uptake and secretion data, fmol/cell/h, define the units of all following analyses. The uptake and secretion of inorganic salts and gases were fully allowed by setting their exchange reactions' upper and lower bounds to 1000 and -1000 fmol/cell/h, respectively. The reactions in the Recon3D subsystems "Drug metabolism" and "Xenobiotics metabolism" had their upper and lower bounds set to zero, to constrain the behavior of the tissue-specific models to what they would likely exhibit under their typical growth conditions.

3.3.2. Generation of Tissue Specific Models

The Recon3D Metabolic Model by default comes with a mapping of the reactions it contains onto genes, represented by NCBI Gene IDs. This allows the use of gene expression data taken from GTEx and TCGA to exclude the reactions from the models based on the abundances of their genes in the relevant tissues. First, the gene expression data in TPM format from these sources were mapped onto reactions in the Recon3D model using the mapExpressionsToReactions default function. Recon3D comes with a pre-defined set of gene-reaction relations, indicating for each reaction the necessary and sufficient set of genes being expressed. For each condition, cancer and healthy, this resulted in an array of "reaction expressions," which would be used to indicate how priority would be assigned to each reaction in constructing tissue-specific networks. Reactions for which the gene expression data did not exist or was ambiguous were given a reaction expression value of -2.

COBRA Toolbox's default function for this kind of model generation, createTissueSpecificModel, can use one of several published algorithms, including GIMME

[60] and INIT [61]. Preliminary work in this study has confirmed the findings by Opdam et al. that while a "GIMME-like" algorithm guarantees the resulting metabolic network to be solvable via a flux balance analysis, an "INIT-like" algorithm offers much higher fidelity to the reference tissue [76]. This study used a hybrid two-step approach to harness the advantages of both algorithms, as explained below.

3.3.2.1. INIT-like Step. The INIT setting of the createTissueSpecificModel function requires the user to supply an array of reaction weights, whereby the algorithm attempts to maximize flux through reactions with high weight and to minimize flux through reactions with low weight. Weights were assigned based on the transformation to the reaction expression suggested by prior work [61, 76], with reactions with uncertain reaction expression maintaining a weight of -2. The weights of the reactions "biomass_maintenance," which represents the tissue expending energy and nutrients to maintain its livelihood [74]; and "ATPS4mi," which represents the activity of mitochondrial ATP synthetase, were set to the maximum value among all weights to ensure that the algorithm would prioritize including these reactions. The createTissueSpecificModel function was then allowed to run for an hour each using the INIT algorithm and such weights, for each condition: cancer and healthy. The INIT algorithm was generally found to generate a tissue specific model for which flux balance analysis was mathematically infeasible. In this case, the GIMME-like process was applied, as explained below.

3.3.2.2. GIMME-like Step. The GIMME setting of the createTissueSpecificModel function requires the user to supply the reaction expression data, and a cutoff value which separates this data into two groups with "high expression" reactions that have expression values above the cutoff, and "low expression" reactions that have values below. Opdam et al. suggest that for the most accurate results, this cutoff must be chosen high (between the 75^{th} and 90^{th} percentile of the reaction expression data) [76].

With the aim of keeping as many reactions obtained from the INIT algorithm in the final model as possible, the expression levels of the reactions that were included by the INIT algorithm were set to the maximum value among all reactions. The original unmodified model was then subjected to the GIMME process, repeated for a series of expression cutoffs

corresponding to percentiles between the 90th and 50th percentile in 10% decrements. The highest percentile cutoff leading to a model amenable to FBA analysis and reaching at least 75% of the biomass maintenance objective of the initial general model being chosen as the final model for both the healthy and cancer conditions.

Finally, exchange reactions and reactions from the central carbon metabolism, corresponding to the Recon3D subsystems "Glycolysis/gluconeogenesis," "Citric acid cycle" and "Oxidative phosphorylation" were re-added to the tissue specific model to avoid overconstraining it. The models were checked for leaks and an erroneous ability to produce energy from inorganic inputs, as detailed in Fritzemeier et al.'s [85] work.

3.3.3. Flux Balance Analysis & Flux Variability Analysis

COBRA Toolbox's optimizeCbModel function was used to carry out FBA on both models. The flux balance analysis was done with an additional objective of minimizing the absolute value sum ("taxicab norm") of all the reaction fluxes in the network to minimize flux through loops with no physical significance.

Since FBA optimal solutions are not necessarily unique, the COBRA Toolbox fluxVariability function was used to find, for each condition – cancer and healthy – the maximum and minimum flux value for each reaction at which it is still possible to obtain the optimal FBA solution.

3.3.4. Sampling

The COBRA Toolbox function gpSampler uses the Artificially Centered Hit-and-Run (ACHR) algorithm [86] to uniformly sample the possible solution space of the metabolic model, constrained by the directionality of the reactions and the limitations on the uptake rates. The upper and lower bounds of each reaction were further constrained to the upper and lower bounds given by the fluxVariability function, such that each sampled flux distribution corresponded to a state at the optimal value of the FBA objective. The function

was used with the default parameters to generate over 10,000 possible flux profiles for the cancer and healthy phenotypes. Reactions absent in either model were assigned flux upper and lower bounds of 0 for purposes of comparison.

Each reaction contained in both metabolic networks were subject to a 2-sample t-test with unequal variances comparing whether the sampling flux profiles indicated statistically significant differences between the two conditions. A Benjamini-Hochberg correction was applied to correct for multiple testing with an acceptable false discovery rate of 0.01. The difference-to-standard-deviation ratio (DSD) was calculated for each reaction such that:

$$DSD = \frac{|\mu_c - \mu_h|}{\sigma_c + \sigma_h}, if\ \sigma_c, \sigma_h > 0$$
$$DSD = \frac{|\mu_{c/h}|}{2\sigma_{c/h}}, if\ \sigma_c\ or\ \sigma_h = 0 \quad (3.1)$$

where μ_c and μ_h stand for the flux sampling means of a reaction in the cancer and healthy metabolic models, respectively. $\mu_{c/h}$ represents the flux sampling mean of a reaction in either model in case the reaction does not exist in the other model. Similarly, σ_c and σ_h stand for the flux sampling standard deviations of a reaction in the cancer and metabolic models, with $\sigma_{c/h}$ representing the standard deviation of a reaction in either model in case the reaction does not exist in the other model. DSD represents a further metric of significance when comparing the means of two distributions. To further narrow down the list of significantly different reactions, reactions with DSD values over 1 were extracted as a subset from the results of the t-testing. This corresponds to reactions whose flux sampling means differed at least by two average standard deviations between the healthy and cancer conditions. Average sampling results for each subsystem were used to classify subsystems into related clusters by hierarchical clustering using the R hclust function.

3.3.5. Synthetic Lethality

The fastSL function by Pratapa et al. [87] was imported into MATLAB to carry out synthetic lethality analysis on both the healthy and cancer metabolic models. This function first finds “single lethals” by removing reactions from the model one at a time, and then finds “double lethals” by removing pairs of reactions. If the removal of a reaction or a pair of reactions causes a greater than 99% reduction in the objective flux, the reaction or the pair of reactions are termed lethal to the tissue simulated by the metabolic model. The cancer single lethal reactions and double lethal reaction pairs were compared to the healthy single lethal reactions and healthy double lethal reaction pairs. Reactions and reaction pairs which were lethal to both simulated tissues were removed to find the set of lethal reactions and reaction pairs that could be uniquely lethal to the cancer cell.

3.4. Protein-Protein Interaction (PPI) Networks

After differential gene expression and constraint-based modelling approached identified genes or proteins differentiating cancer and healthy phenotypes, PPI networks were constructed with the aim of finding the interactions of these proteins with known disease genes, known drug targets, and each other. This process was used to discover interaction partners of significant genes that could be easier to target than the genes themselves or interacting groups of significant genes that could be co-targeted.

3.4.1. Network Construction & Topological Analysis

The integrated data from the HPRD, BioGRID and IntAct comprise a list of pairs of interacting protein IDs, which is sufficient to form and analyze a network using software such as Cytoscape. However, additional processing was done to ensure the relevance of the networks constructed in this study to healthy and cancerous pancreas tissue. To that end, genes in the list of interacting pairs which did not have at least one gene ontology (GO) term in common with their condition’s corresponding seed genes (“healthy” or “cancer”) in each of the three gene ontology categories – molecular function, cellular compartment and

biological process – were pruned from the list [10, 11, 88]. The R package GO.db [89] was used to access the GO terms of each gene algorithmically, indexed by HGNC gene symbols. Only GO terms which had been assigned to genes as a result of *in vivo* or *in vitro* experimentation were considered, and the terms which were assigned through *in silico* studies were discarded in the interest of not accumulating errors due to repeated computational studies. These lists of interacting protein pairs were pruned further to include only proteins which had non-negligible expression levels in the pancreas, using the gene expression data, as carried out by Peng & Schork [51].

Two PPI networks were then constructed using the Cytoscape software, one filtered according to cancer seed genes and gene expression, and the other filtered according to healthy seed genes and gene expression. The topological analysis of the networks was carried out using Cytoscape's NetworkAnalyzer plugin [71].

3.4.2. Clustering

The Markov Clustering (MCL) algorithm [42] which is made available on the Cytoscape platform through a plugin called Clustermaker2 [90] was used to construct clusters from the PPI data. This algorithm was recommended for PPI cluster discovery in a study by Brohee and van Helden [46], where the authors suggest a value of 1.8 for the inflation parameter, the only adjustable parameter for the algorithm that determines the granularity of the clustering. The effect of the inflation parameter was first tested on the healthy PPI network, generating a series of PPI clusterings with inflation parameters in a range between 1.8 and 2.5. The best inflation parameter was chosen semi-quantitatively, attempting to minimize the number of unclustered nodes while ensuring that the largest clusters remained at tractable and similar sizes (ideally around 10 to 500 nodes in each cluster) to make subsequent analysis more consistent. The "ideal" inflation parameter was chosen as 2.2 and subsequent clustering on both the healthy and cancer networks was carried out at this level of the inflation parameter. Since the healthy and cancer PPI networks had similar characteristics when unclustered, this was found to yield a similar size distribution of clusters.

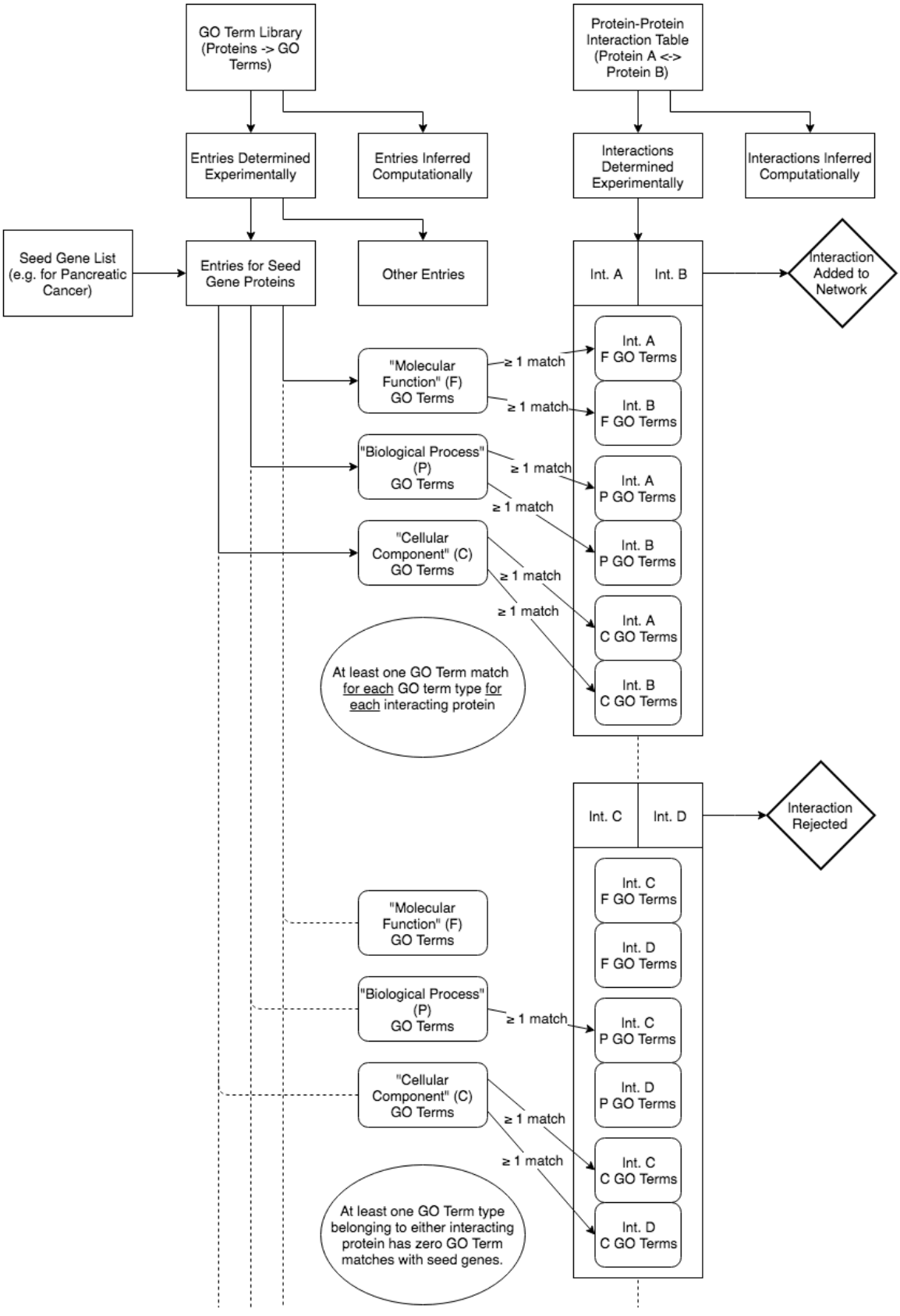

Figure 3.6. Filtering PPI data by GO terms.

3.5. Coexpression Networks

Similar to PPI networks, coexpression networks were constructed with the aim of finding coexpression partners or clusters for the genes discovered to be different between the cancer and healthy phenotypes. It was surmised that these coexpression partners, while themselves not found to be significantly different in expression, could allow for alternative ways of targeting genes for therapeutic, diagnostic, or prognostic purposes. Furthermore, genes discovered to be within coexpression clusters of known disease genes not discovered in this study could still be genes of interest, through a guilt-by-association approach.

An algorithmic approach termed "Weighted Gene Co-expression Network Analysis (WGCNA)" [40] was utilized through the similarly named R package WGCNA [91] to find coexpression networks and modules in both the healthy and cancer conditions. The raw count data from the TCGA and GTEx databases were imported. To ensure that the TCGA data constituted a true cancer coexpression network, the 4 healthy samples were removed from the data. To limit the outsized effect of very low-expressed-high-variance genes on the networks, genes that did not at least have 10 counts in over more than half of the samples were removed from each dataset. The voom function in the limma R package [79] was used to transform raw counts into logCPM and normalize these values using the quantile normalization method in one step. Samples in both datasets were clustered using hierarchical clustering to visually identify outliers – samples with highly atypical expression profiles, which were removed from the datasets. The R programming language's default function hclust was used for this purpose.

The WGCNA process assumes that the networks created are scale-free. To ensure this, the WGCNA package's pickSoftThreshold function was used to calculate Pearson correlation coefficients (PCC) between each pair of genes in each dataset, and raise these PCCs to various powers, finding the lowest power ("soft threshold") for which the resulting weighted network showed at least an 80% similarity to a scale-free network. This was defined as the weighted network having at least an R^2 value of 0.8 when a power law of best fit is fit to its degree distribution. These powers were found to be 16 for the cancer network, and 14 for the healthy network.

After such preprocessing, WGCNA's blockwiseModules function was used to calculate a topological overlap matrix (TOM) using the PCCs and soft thresholds previously estimated. The package's default settings were used with the exception of the mergeCutHeight parameter, which indicates the inverse of the level of correlation between clusters at which the clusters will be merged. This parameter was set to 0.1, indicating merging clusters with over 90% correlation. It should be noted that this study used a "signed" network and a "signed" TOM. Practically this implies that a strong anticorrelation between the expression level of two genes will be distinct from a strong correlation, and topological overlap will take on a range of 0 to 1, 1 implying maximum correlation and 0 implying maximum anticorrelation. Anticorrelated genes, that nonetheless may have a significant effect on each other's expression, are not likely to be clustered together.

After this step, the clusters were exported to Cytoscape for visualization. The Cytoscape NetworkAnalyzer plugin [71] was used for their topological analysis.

3.6. Annotation & Enrichment Analyses

Annotation with known disease genes, drug targets and GO terms was done through the use of clusterProfiler, an R package based on Yu et al.'s [92, 93] work, using algorithmic improvements from Korotkevich et al.'s package fgsea [94]. clusterProfiler offers two classes of functions which check whether a certain group of genes or proteins is "enriched" in a certain biological concept (such as "pancreatic cancer") using databases of controlled biological vocabularies. The "enricher" class of functions checks enrichment using a hypergeometric test, comparing only the NCBI Gene IDs of the genes or proteins in a group against the relevant database. The "GSEA" class of functions uses the Gene Set Enrichment Analysis (GSEA) paradigm [95] and requires the fold changes of the genes or proteins as well. The GSEA approach weights annotation terms by their associated genes' fold changes over a background, which in this case was obtained from the GTEx data for healthy pancreatic tissue. [65] Both functions were used to seek enrichment with terms from the following corpuses: DOSE [93] for annotations with known disease genes, the TTD [73] for annotations with known cancer drug targets, the Bioconductor databases GO.db and

org.Hs.eg.db [89, 96] for annotations with GO terms, and KEGG [19–21] for annotations with pathways reported in the KEGG database.

Any group of genes subject to analysis with clusterProfiler was first matched with the results of the limma DEG pipeline to discover the fold changes of each of their constituents. Genes or proteins whose fold change could not be discovered through the limma pipeline due to having very low expression or other reasons were still subject to enrichment analysis using the hypergeometric test approach, but were considered not to be part of the group for the GSEA approach. For each group, the genes were sorted by their associated log2-fold-changes and then put through the clusterProfiler functions seeking enrichment with terms from all the aforementioned corpuses. The p-value cutoff parameter for both classes of functions was set to 0.05, and the Benjamini-Hochberg correction for multiple testing was utilized at the default false discovery rate threshold of 0.2. The parameters minGSSize and maxGSSize, dictating the range of gene set sizes eligible for enrichment analysis were set to 2 and 2000, respectively. While each annotation dataset contains gene sets of varying sizes, these settings were found to exclude overly general annotation terms, and terms associated with only one gene, which could result in uninformative associations.

The groups subject to enrichment analysis via the hypergeometric test and GSEA methods in this study were:

- DEGs identified with the limma pipeline (1 group)
- DEGs that are also associated with metabolic reactions with significantly different flux (1 group)
- Each coexpression network cluster identified by WGCNA, for both healthy pancreas and pancreatic cancer networks (46 cancer groups, 98 healthy groups)
- Each PPI network cluster identified using MCL clustering, for both healthy pancreas and pancreatic cancer networks (1330 cancer groups, 1325 healthy groups)

In addition to the enrichment analysis carried out via clusterProfiler, each of these groups were manually annotated with the 83 known drug targets of anticancer agents obtained from TTD. The outputs of the PPI and coexpression networks were further

annotated with the 74 cancer and 96 healthy seed genes obtained from KEGG. A graphical summary of the procedure follows this section.

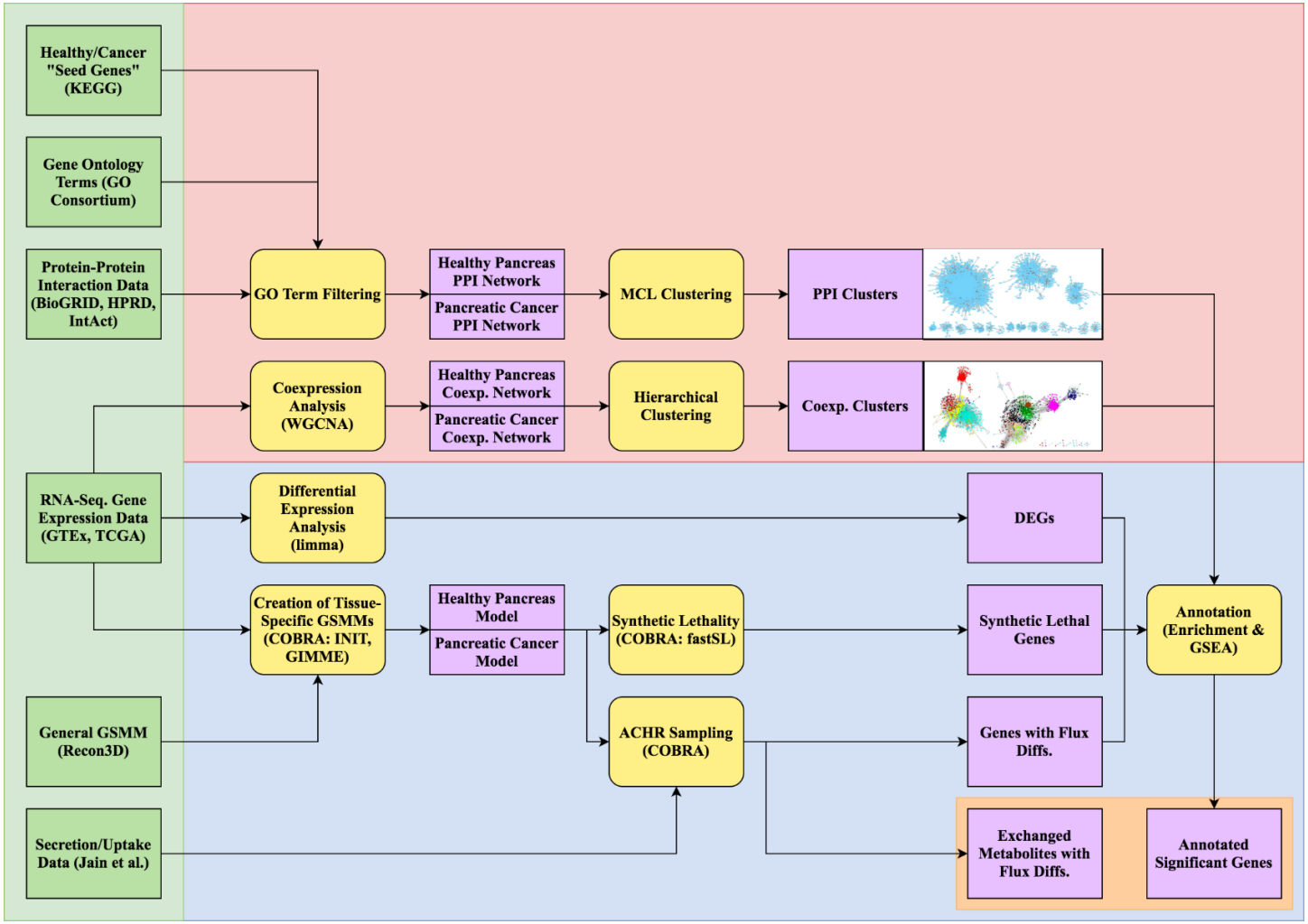

Figure 3.7. The biomarker discovery pipeline, showing all raw data (green), analyses (yellow) and results (purple); with the red section containing methods using network topology, the blue section containing those which are not, and the orange section containing the results.

4. RESULTS & DISCUSSION

4.1. Differentially Expressed Genes (DEGs)

4.1.1. Preprocessing and Estimation of Confounding Effects in Data

The 53,196 genes common to the GTEx and TCGA datasets were subjected to the edgeR/limma analysis pipeline to determine DEGs in cancer using a linear model. This pipeline requires each sample to be classified by each covariate that is being examined, where the covariate of interest is whether the sample comes from cancerous tissue. The other covariates are age, sex, and database (whether the data come from the TCGA/GTEx dataset). The object of the linear model is to identify significant gene expression differences between healthy and cancer conditions given the variation in gene expression due to age, sex, or database. A breakdown of all 400 samples in this study by each of these covariates is given below in Table 4.1. It should be noted that the columns in Table 4.1 are independent of each other.

Table 4.1. Linear model sample breakdown by covariates.

Cancer/Healthy		**Database**		**Sex**		**Age**	
Level	Freq.	Level	Freq.	Level	Freq.	Level	Freq.
Cancer	148	TCGA	152	Male	230	20-29	20
						30-39	22
						40-49	68
						50-59	127
Healthy	252	GTEx	248	Female	170	60-69	102
						70-79	49
						80-89	12

Since the linear model needs to correct for the effect of these other covariates, it is desirable to estimate the variance introduced by each covariate prior to using the linear model. Between the two datasets, the edgeR function filterByExpr discovered that 32,556 genes were, on average, expressed below a logCPM cutoff value of -2.1. This cutoff value is calculated as

$$Cutoff = \log_2\left(\frac{10 \times 10^6}{M} + \frac{2 \times 10^6}{L}\right) \tag{4.1}$$

and was taken from Law et al.'s [97] tutorial for processing RNA-Seq data. M is the median value among library sizes (sums of all expression read counts in each sample) and L is the mean value among library sizes.

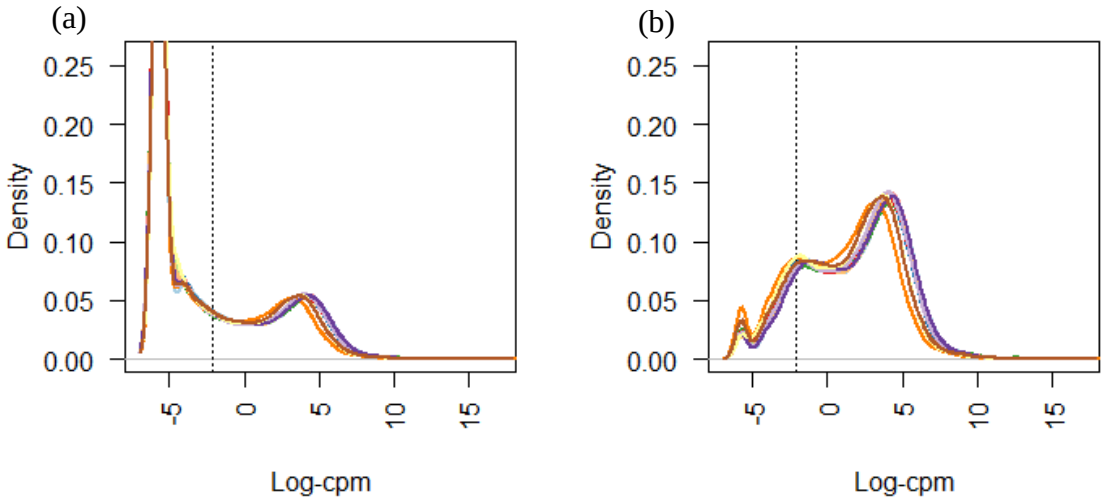

Figure 4.1. Gene expression data (a) before and (b) after filtering to remove low-count genes.

The filterByExpr function does not remove all genes with mean expression below this cutoff but removes genes which exhibit expression below this cutoff in a "sufficiently high number of samples." The "sufficiently high number of samples" is picked to be the size of the smallest group among all groups included as factors in the linear model [98], which in

this case was 12 samples corresponding to the 80-89 year old age group. The filtering removed 32,556 genes, or approximately 61% of the dataset. The logCPM expression profiles are shown in Figure 4.1. before and after filtering. Each colored plot represents a single sample.

The removal of genes with very low expression values is a necessary preprocessing step for both the estimation of confounding covariate effects, and for the use of the linear model itself, since the variance behavior of such low-expression genes is different to the rest of the data. After the removal of low-expression genes, and subsequent normalization and variance-stabilization, the data are seen to be very significantly confounded by the dataset from which the data were taken.

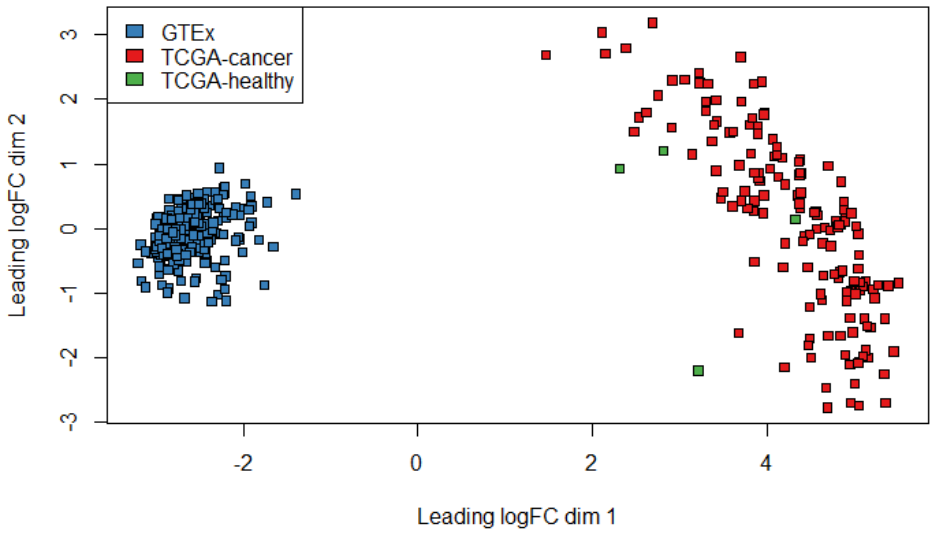

Figure 4.2. An MDS plot illustrating the amount of variance explained by dataset choice.

Figure 4.2. shows a Multidimensional Scaling (MDS) plot indicating a majority of the differences between samples comes from the choice of dataset, not the cancer/healthy condition. The confounding effects of the other clinical variables were examined.

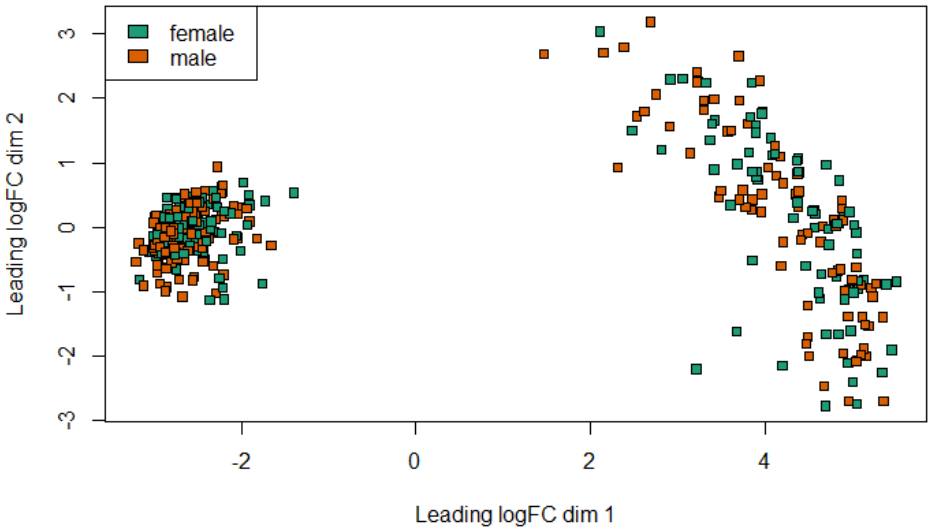

Figure 4.3. An MDS plot illustrating the amount of variance explained by sex.

The sex characteristic of the samples was found to be mostly balanced between the cancer and healthy conditions, and no significant variance due to sex can be seen in the MDS plot in Figure 4.3.

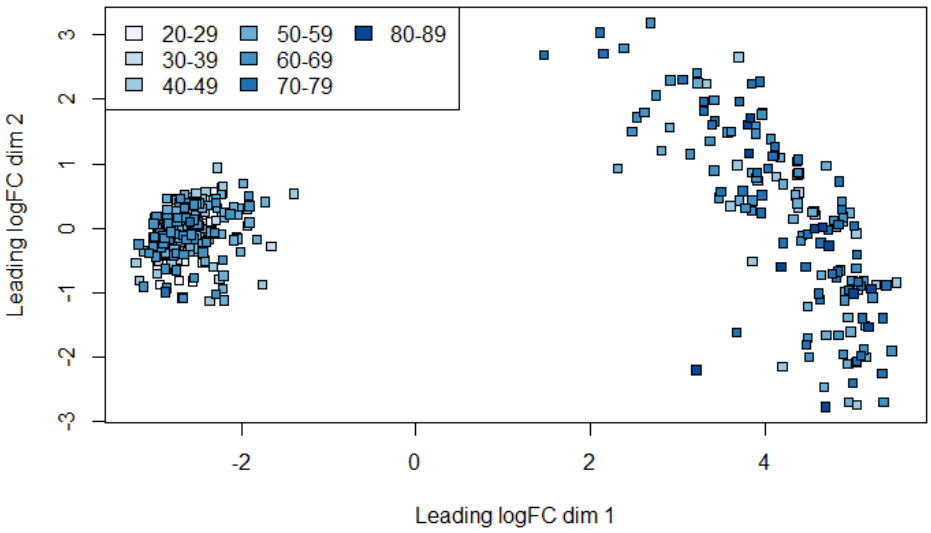

Figure 4.4. An MDS plot illustrating the amount of variance explained by age.

Age groups showed differences between the cancer and healthy conditions, which were illustrated in the MDS plot in Figure 4.4. The cancer group in the study is decidedly older.

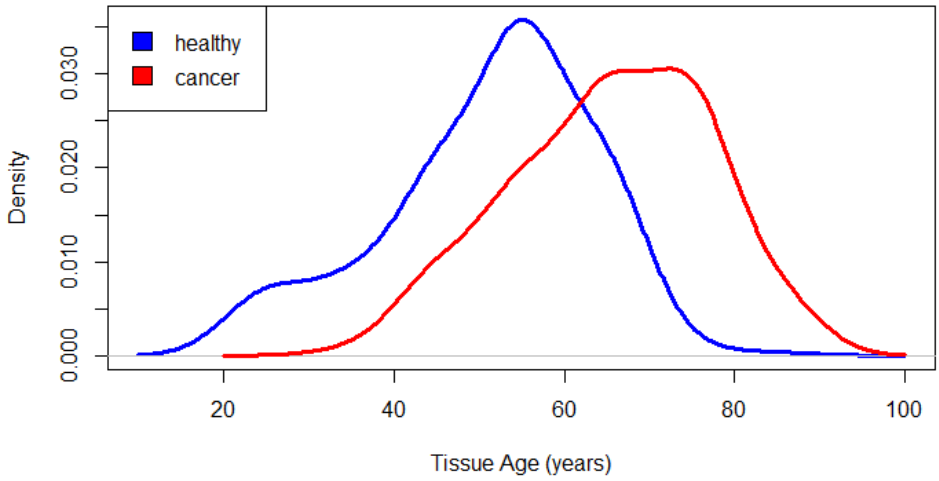

Figure 4.5. A density plot illustrating age variance between the cancer and healthy conditions.

Figure 4.5., a density plot showing the age distributions of the healthy and cancer sample groups was also included, since the many different age levels may impede interpretation of the MDS plot. The plots show that age could indeed be a source of variance between datasets. These plots justify the use of a linear model taking data source and age into account when discovering DEGs. Furthermore, since the pipelines used further along in this study (such as WGCNA) assume identical processing of data [91], the significant confounding factors indicate that the data cannot be analyzed together in those pipelines. Therefore, in subsequent analysis, the cancer and healthy data were analyzed separately and compared after topological analysis and annotation. An option to overcome this issue and analyze the data together by using a batch-effect-removal tool was rejected since these tools rely on groups to be balanced within batches, which is not the case in this study [99, 100]. Due to the large confounding effects, especially due to dataset size, the linear model is expected to therefore make very conservative estimates of which genes are significantly different between the cancer and healthy conditions.

4.1.2. Significant DEGs

After necessary preprocessing, 826 DEGs were discovered between the healthy and cancer conditions using a Benjamini-Hochberg false discovery rate threshold of 0.01. The full table of all 826 DEGs with adjusted p-values below 0.01 can be found in the appendix, in Table A.5. The most significantly differentially expressed genes with adjusted p-values below 10^{-7} were illustrated in Table 4.2. A brief description of the top 5 most significant genes found in Table 4.2, as well as a survey of existing literature relating them to pancreatic cancer, follows.

Table 4.2. DEGs with Adjusted p-Values Below 10^{-7}.

HGNC Gene Symbol	**Average Expression (logCPM)**	**log2 Fold Change**	**Adj.P.Val (Benjamini-Hochberg)**
EPCAM	7.246	2.763	$< 10^{-12}$
GPRC5A	5.128	4.239	$< 10^{-10}$
MAL2	6.179	2.522	$< 10^{-10}$
RAB25	5.571	2.781	$< 10^{-9}$
CGN	6.583	2.644	$< 10^{-9}$
KRT19	6.761	3.853	$< 10^{-9}$
FXYD3	5.254	4.541	$< 10^{-9}$
NQO1	4.081	3.223	$< 10^{-9}$
TSPAN1	3.870	3.981	$< 10^{-9}$
PERP	6.766	2.161	$< 10^{-9}$
LAD1	6.720	2.227	$< 10^{-9}$
ERBB3	6.996	2.596	$< 10^{-9}$
MYH14	6.383	2.484	$< 10^{-8}$

Table 4.2. DEGs with Adjusted p-Values Below 10^{-7}. (cont.)

HGNC Gene Symbol	**Average Expression (logCPM)**	**log2 Fold Change**	**Adj.P.Val (Benjamini-Hochberg)**
KLF5	5.052	2.792	$< 10^{-8}$
ESRP1	6.075	2.558	$< 10^{-7}$
MISP	2.289	4.580	$< 10^{-7}$
PPL	5.647	2.517	$< 10^{-7}$
AGR2	5.052	5.674	$< 10^{-7}$
B3GNT3	3.424	4.277	$< 10^{-7}$
CEACAM6	4.883	4.990	$< 10^{-7}$
PKP3	5.680	2.790	$< 10^{-7}$

The EPCAM gene, which is by far the gene found to show the most significant difference, encodes the epithelial cellular adhesion molecule (also named EpCAM). This molecule is a cell membrane protein that regulates cell-cell interactions and has been implicated in cancer research showing increased expression in many different types of cancer. While most research on EPCAM is on its properties as a pan-cancer indicator, these studies [101–104] also note that EPCAM overexpression is a much stronger indicator of breast, colon and pancreatic cancers compared to other cancer types. A study by Spizzo et al [105] found that EPCAM overexpression was positively correlated with most gastrointestinal primary tumors and metastatic tumors, patients with hepatocellular carcinomas, clear cell renal cell cancer, urothelial cancer and squamous cell cancers did not show such behavior. The study further found that EPCAM overexpression correlated with some breast cancer subtypes and not others. Akita et al. [101, 106] indicate that for most cancers, EPCAM overexpression is a biomarker for poor prognosis, whereas it indicates a relatively good prognosis for pancreatic cancer. Considering CA19-9's non-specificity limitations, such a nuanced expression behavior could be a boon of investigating EPCAM

as a pancreatic cancer biomarker. A further point of interest about EPCAM raised in studies like [101, 103, 104, 106] is its association with not-yet-cancerous, cancer precursor cells termed tumor-initiating cells (TICs) or cancer stem cells (CSCs). Studies using bispecific EpCAM/CD3 antibodies targeting pancreatic cell lines in *in vitro* and animal models were shown to be effective in inhibiting cancer progression and eliminating CSCs [107, 108]. Last but not least, the EpCAM protein was found to be amenable to immunocapture in whole blood in circulating pancreatic cells (CPCs) and pancreatic circulating tumor cells (CTCs) [109]. This illustrates the possibility of EPCAM being a valid biomarker for diagnostic purposes, when it is unclear whether the patient has cancer or which organ is the cancer primary site.

GPRC5A, also known as RAI3, is a gene that codes for a G-protein-coupled receptor whose expression was found to be dysregulated in a variety of human cancers [110]. While GPRC5A expression is not consistently increased or decreased in all cancers [110], it's been found to be consistently upregulated in pancreatic cancer [110–112]. Liu et al. [112] found that knocking out GPRC5A was associated with reduced cancer proliferation and reduced resistance to chemotherapeutic drugs. Similarly, Zhou et al. [110] report that treatment of cancer cells with the chemotherapeutic drug gemcitabine was found to induce GPRC5A expression *in vitro*, with GPRC5A knockout sensitizing the cells to gemcitabine treatment. Jahny et al. have recently reported being able to use GPRC5A profiling in harvested pancreatic ductal cells to identify pancreatic cancer, healthy tissue, chronic pancreatitis and intraductal papillary mucinous neoplasms (IPMNs) with a sensitivity of 89.7% and a specificity of 94.3%. Jahny's study has used the fact that apical ductal tissue, a morphological variant of ductal tissue in the pancreas, has a different GPRC5A expression profile compared to bulk ductal tissue [111]. The results relating to IPMNs are significant due to IPMNs being a pancreatic ductal tumor precursor which are benign by themselves, but show a propensity for transforming into malignant pancreatic ductal adenocarcinoma [2, 3, 17].

The MAL2 gene encodes a transmembrane protein found to be overexpressed in breast and ovarian carcinoma, in addition to pancreatic ductal adenocarcinoma [113–115]. It was especially found by Bhandari et al. and Eguchi et al.'s [114, 115] work, respectively, that

MAL2 increases invasiveness and metastatic potential by mediating the epithelial-to-mesenchymal transition in cancer cells.

RAB25 encodes the similarly named Rab25 signaling protein, whose normal function in cells is the recycling of cell-surface proteins through vesicle formation and transport [116–118]. To date, its function in pancreatic cancer has not been well studied, but some studies have shown it to act as an oncogene in renal, breast and ovarian cancers, as well as some types of lung cancer [118, 119]. Conversely, it has been found to have tumor suppressor properties in esophageal and colorectal cancers [117, 118], as well as in triple-negative lung cancer in a study by Jeong et al. [119].

The CGN gene encodes a protein named cingulin, which takes part in the organization of tight junctions between cells in epithelial tissues. An early study by Citi et al. [120] showed an upregulation of cingulin in patients presenting cancers of epithelial tissues. However, cingulin's impact on cancer was found to be very complex, and especially dependent on its effect on the expression of other tight junction proteins, most significantly claudins [121, 122]. Specifically, claudin-4 was found to be overexpressed in pancreatic cancer and likely involved in the epithelial-to-mesenchymal transition [123, 124].

Overall, it must be noted that this list of top DEGs is dominated by genes whose function is the production or regulation of proteins which form the extracellular matrix and the intercellular junctions in pancreatic tissue. KRT19 and CEACAM6 are such genes which have already been placed under scrutiny in the case of pancreatic ductal adenocarcinoma by Yao et al. and Pandey et al., respectively in [125, 126]. Meanwhile, PKP3 [127], PPL [128], LAD1 [129] and ESRP1 [130] are genes with similar function that have been implicated in other cancer types.

4.1.3. Annotation & Enrichment Analyses of DEGs

The 826 DEGs were found to overlap with 7 out of the 83 possible anticancer drug targets. These results can be seen in Table 4.3. Table 4.3 shows very few DEGs are directly

amenable to being targeted by existing drugs. This is further compounded by the fact that several DEGs that are drug targets for existing drugs are significantly downregulated in pancreatic cancer, so their disruption could have an exacerbating effect on cancer progression. MET, SRD5A3, and ERBB2 are those genes that are upregulated in pancreatic cancer that have been successfully targeted with anticancer agents in other contexts.

The MET gene encodes the mesenchymal-epithelial transition factor (c-Met), which is a receptor for hepatocellular growth factor HGF. HGF was shown by Di Renzo et al. and Ebert et al. [131, 132] to stimulate the proliferation of pancreatic cancer cell lines *in vitro,* and there is ongoing research about the viability of the MET-HGF pathway as a prognostic biomarker or drug target in PDAC [133, 134]. This pathway was also indicated to be active in colorectal cancer and other tumors *in vitro* [135, 136]. One of the drugs targeting it, Cabozantinib, also targets VEGFR2 [73], a gene that encodes the vascular endothelial growth factor receptor, which is implicated in several cancers due to its promotion of angiogenesis – the proliferation of blood vessels in cancerous tissue [119, 137].

The SRD5A3 gene encodes for an enzyme named polyprenol reductase (Steroid 5-alpha-reductase 3). One of this gene's functions is the conversion of testosterone into another androgen, 5-alpha-dihydrotestosterone (5-alpha-DHT). As such, Uemura et al. [138] have investigated its effect on prostate cancer progression, finding that the gene's knockout reduced cancer viability in prostate cancer cells *in vitro.*

Table 4.3. DEGs which are known drug targets [73], genes marked *D* are downregulated in cancer.

HGNC Gene Symbol	**Average Expression (logCPM)**	**log2 Fold Change**	**Adj.P.Val (B-H)**	**Drugs**	**Indications**
MET	5.275	1.851	< 0.001	Cabozantinib*, Crizotinib**	Thyroid cancer*, renal cell carcinoma*, hepatocellular carcinoma*, non-small cell lung cancer**
CSF3R^{D}	2.685	-1.974	< 0.006	Pegfilgrastim	Malignant solid tumor
FLT3^{D}	1.738	-1.839	< 0.007	Intedanib	Colorectal cancer
TLR7^{D}	-0.463	-1.966	< 0.01	Imiquimod	Skin cancer
BCL2^{D}	2.811	-1.460	< 0.002	GDC-0199, Taxol, Paclitaxel	Breast cancer
ADRA1D^{D}	-1.751	-2.282	< 0.002	Armodafinil	Pediatric cancer
SRD5A3	4.486	1.333	< 0.002	Dutasteride, Polyestradiol Phosphate, Finasteride	Prostate cancer
ERBB2	6.860	1.087	< 0.004	BIBW 2992, Masoprocol, Dacomitinib, Trastuzumab, Neratinib Maleate, Tyverb, Lapatinib, Pertuzumab	Breast cancer, non-small-cell lung cancer, prostate cancer

The ERBB2 gene is also known as HER-2 and encodes a kind of epidermal growth factor receptor. It has been widely studied [139–141], being overexpressed in many different cancer phenotypes. Special attention has been given to ERBB2 as a drug target in breast cancer. However, attempts to use Trastuzumab, which is the anticancer agent used in breast cancer treatment to treat pancreatic cancer *in vitro* have not shown strong results [142, 143].

The DEGs have been annotated with the R clusterProfiler package, drawing upon the disease ontology and gene ontology databases, using a p-value threshold of 0.05, followed by a Benjamini-Hochberg false discovery rate threshold of 0.2 to determine significance. All results of the annotation process that have met this significance threshold are represented below.

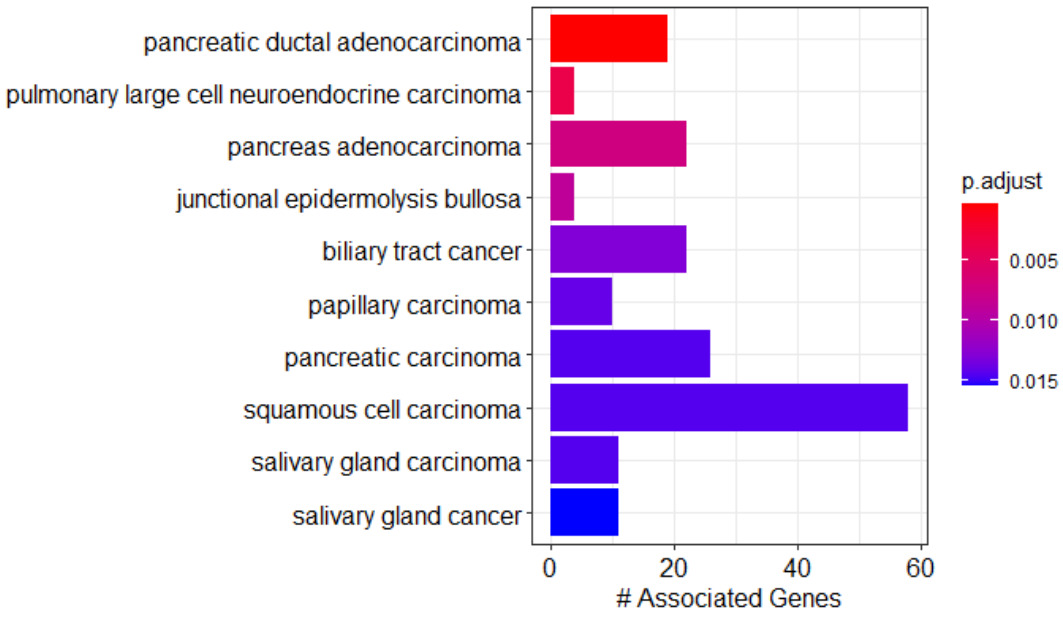

Figure 4.6. Disease ontology annotation for DEGs.

Figure 4.6. represents the 10 annotation terms most significantly associated with DEGs in the disease ontology corpus. While pancreatic ductal adenocarcinoma and other pancreatic cancers are represented at highly significant levels, other cancer types have also shown strong associations with this set of DEGs, primarily squamous cell carcinoma, which is a

term covering cancers in several different tissues, originating in squamous cells that are flat in shape. These cells make up the epidermis, the outermost layer of human skin, but are also found in the linings of the gastrointestinal tract, lungs and some glands [3].

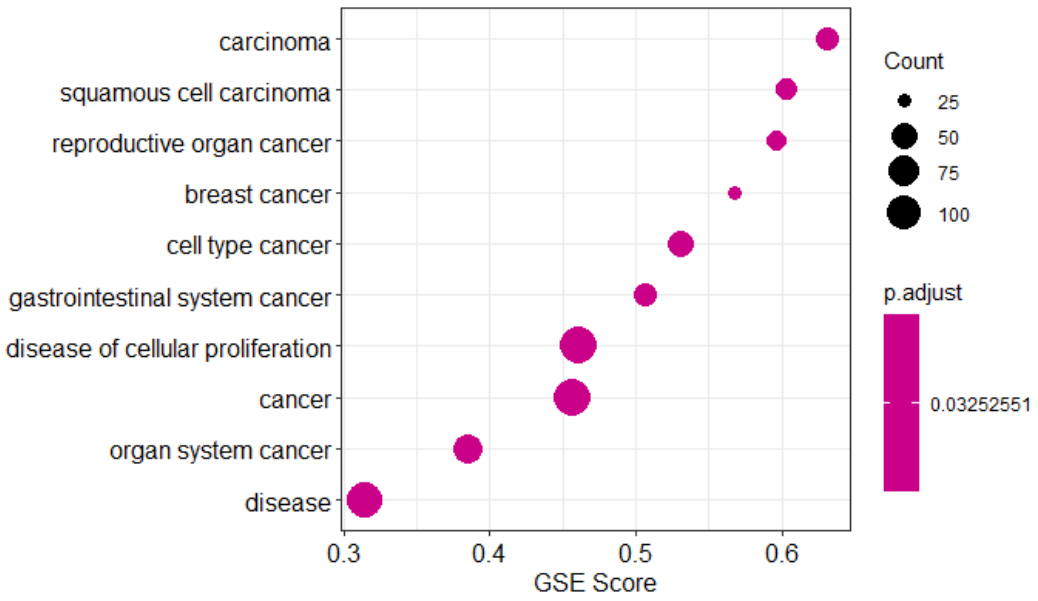

Figure 4.7. Disease ontology annotation for DEGs, GSE.

Figure 4.7. shows the same disease ontology corpus annotated to the DEGs using the gene set enrichment analysis (GSEA) method. GSEA weights the annotation terms by the amount of their associated genes' expression differences from a background level. In this case, while "carcinoma" has been found to represent this group of genes most significantly, other significantly associated terms do not include pancreatic cancer.

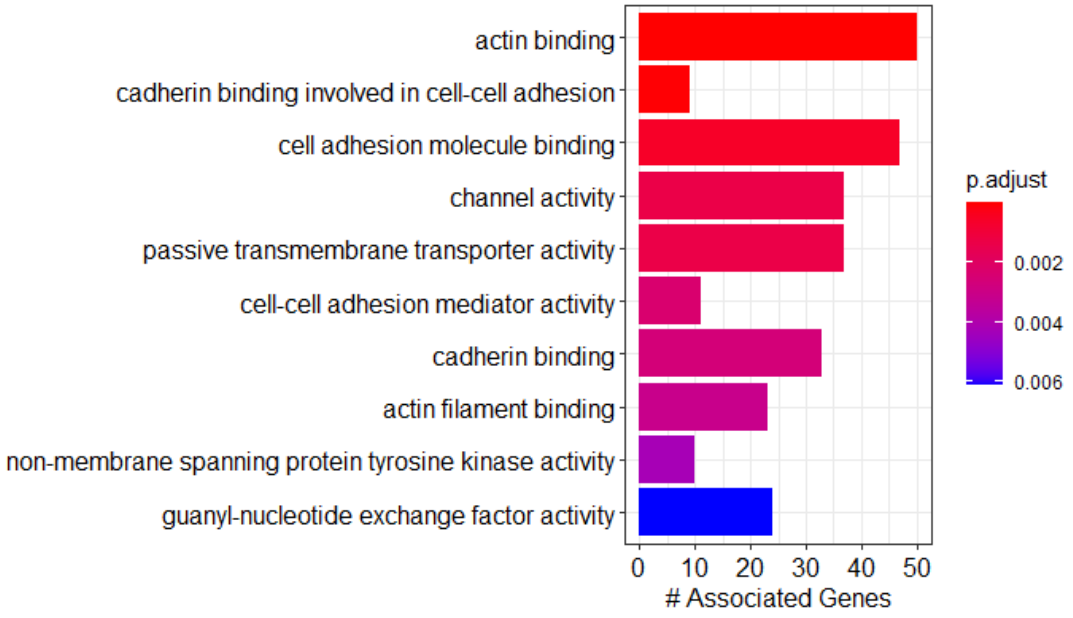

Figure 4.8. Gene Ontology: Molecular function annotation for DEGs.

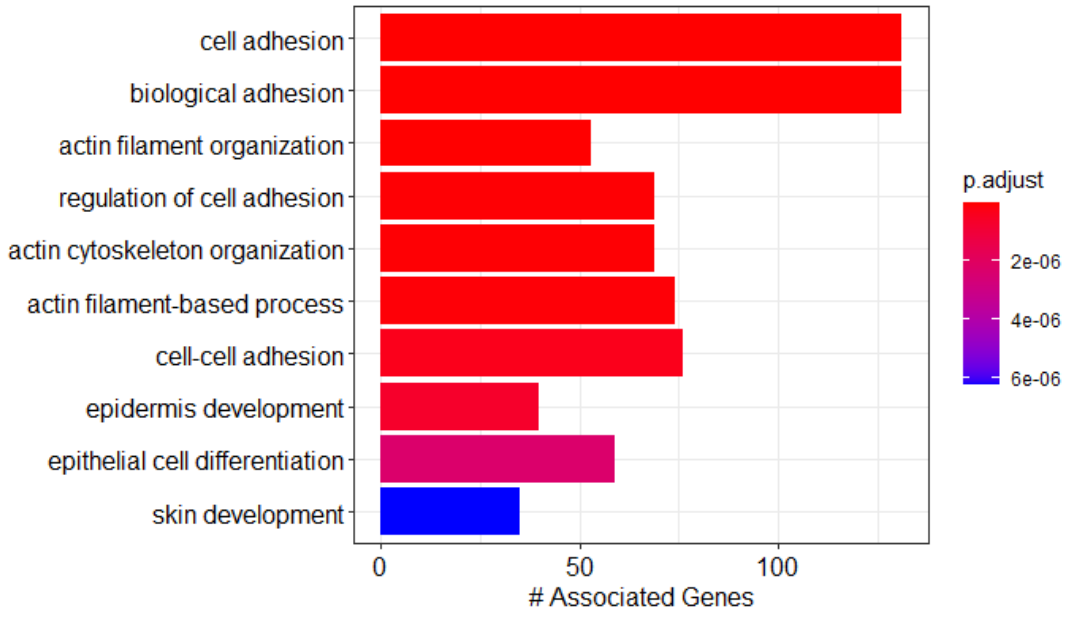

Figure 4.9. Gene Ontology: Biological process annotation for DEGs.

Gene ontology annotations reveal that the DEGs that have been discovered are significantly associated with cell-cell adhesion phenomena and the binding of cells through actin and cadherin proteins, which mediate cell-cell junctions. Figure 4.8. and Figure 4.9. contain the “molecular function” and “biological process” gene ontology associations of the DEGs, respectively.

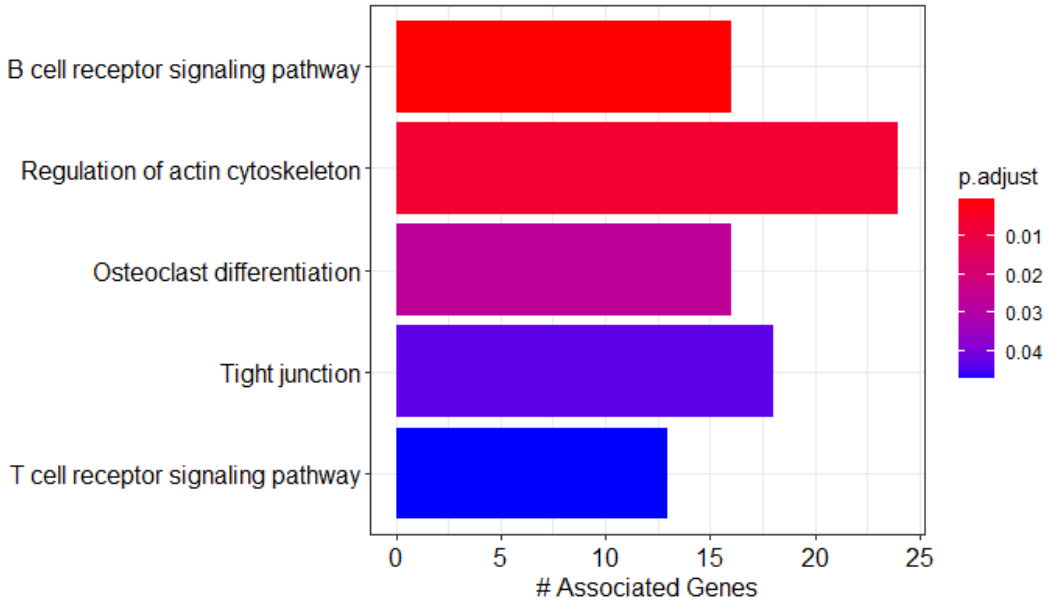

Figure 4.10. KEGG pathway annotation for DEGs.

Figure 4.10. contains the most significant annotations of the DEGs with KEGG pathways. This is in line with the top DEGs containing many genes involved in cell adhesion. DEGs are significantly enriched in the genes on KEGG’s immune cell response pathways (T and B cell receptor signaling pathways), in addition to the cell-cell-junction-related pathways previously discussed in this section. The inclusion of T and B cell receptor signaling pathways could indicate an immune response to cancer. Immune cell infiltration in cancerous tissue was studied by Li et al. in pancreatic cancer, finding that T-cell infiltration did not inhibit cancer growth, but reduced metastatic potential and improved responses to immunotherapy [144]. Li et al. further suggest that programmed cell death protein (PDCD1) and chemokine receptor 3 (CXCR3) need to be expressed in the cancer cells for such infiltration to occur, and that chemokine ligand 1 (CXCL1) expression

prevents infiltration [144]. In the present study, the expression of all three genes was found to be reduced, although not at significant levels. ($p > 0.05$) In Table 4.2, the gene B3GNT3 likely represents the immune-response-related effects of pancreatic cancer, since its expression was found to inhibit T cell infiltration in pancreatic cancer in a study by Zhuang et al. [145].

4.2. Metabolic Models

4.2.1. General Properties of Generated Tissue-Specific Models

To generate tissue specific models, the gene expression data from the TCGA and GTEx databases were mapped onto the Recon3D general model using Recon3D's pre-defined [74] gene-reaction associations. It was found that out of the 10,600 reactions in the Recon3D model, 4,688 were either associated with none of the genes in the databases, or were associated with at least one gene not in either database whose expression was necessary for the reaction to occur. This did not preclude these reactions being allowed in the tissue-specific models but indicates the high proportion (nearly 45%) of reactions for which the createTissueSpecificModel algorithm did not have expression data to work with. It was found that of these 4,688 reactions, 2,187 (nearly 47%) were exchange reactions or in the subsystems "drug metabolism" or "xenobiotics metabolism." Therefore, exchange reactions were left in the model regardless of the createTissueSpecificModel function's output, their activities being limited by the uptake/secretion rate constraints adapted from Jain et al.'s work [75]. The adapted constraints and Jain et al.'s original values are reported in appendix Table A.4. The reactions belonging to the subsystems "drug metabolism" or "xenobiotics metabolism" were set to zero flux. This process used the hybrid GIMME/INIT process as detailed in the Methods section.

The models generated were found to both be leak free and unable to generate energy without carbon input. The "biomass_maintenance" reaction, which was the optimization objective for constraint-based analysis took on an optimal flux value at 0.0079 fmol/cell/h in both models. "Biomass_maintenance" is a synthetic reaction empirically specified in Recon3D, signifying the consumption of several types of metabolites to maintain the

livelihood of an arbitrary human cell. Table 4.4 shows the stoichiometric coefficients of each metabolite consumed in the "biomass_maintenance" reaction, as well as the effective consumption rate given the "biomass_maintenance" flux of 0.0079 fmol/cell/h.

Table 4.4. Biomass maintenance precursor metabolites' stoichiometric coefficients and consumption rates; metabolites have been color coded: water – blue; sugar – light red; nucleotide triphosphates – green; amino acids – yellow; lipids – dark red.

Metabolite	**Stoichiometric Coefficient**	**Maintenance Consumption (fmol/cell/h)**
Water	20.6508	0.1631
Glucose-6-phosphate	0.2752	0.0022
ATP	20.7045	0.1636
CTP	0.0390	3.08×10^{-4}
GTP	0.0361	2.85×10^{-4}
UTP	0.0534	4.22×10^{-4}
L-Glutamate	0.3859	3.05×10^{-3}
L-Aspartate	0.3526	2.79×10^{-3}
L-Alanine	0.5056	3.99×10^{-3}
L-Asparagine	0.2794	2.21×10^{-3}
L-Cysteine	0.0466	3.68×10^{-4}
L-Glutamine	0.3260	2.58×10^{-3}
Glycine	0.5389	4.26×10^{-3}
L-Serine	0.3925	3.10×10^{-3}
L-Threonine	0.3127	2.47×10^{-3}
L-Lysine	0.5921	4.68×10^{-3}
L-Arginine	0.3593	2.84×10^{-3}
L-Methionine	0.1530	1.21×10^{-3}
L-Histidine	0.1264	9.99×10^{-4}
L-Tyrosine	0.1597	1.26×10^{-3}
L-Isoleucine	0.2861	2.26×10^{-3}
L-Leucine	0.5455	4.31×10^{-3}

Table 4.4. Biomass maintenance precursor metabolites' stoichiometric coefficients and consumption rates; metabolites have been color coded: water – blue; sugar – light red; nucleotide triphosphates – green; amino acids – yellow; lipids – dark red. (cont.)

Metabolite	**Stoichiometric Coefficient**	**Maintenance Consumption (fmol/cell/h)**
L-Tryptophan	0.0133	1.05×10^{-4}
L-Phenylalanine	0.2595	2.05×10^{-3}
L-Proline	0.4125	3.26×10^{-3}
L-Valine	0.3526	2.79×10^{-3}
Cardiolipin	0.0117	9.21×10^{-5}
Cholesterol	0.0204	1.61×10^{-4}
Phosphatidylcholine	0.1545	1.22×10^{-3}
Phosphatidylethanolamine	0.0554	4.37×10^{-4}
Phosphatidylglycerol	0.0029	2.30×10^{-5}
Phosphatidylinositol	0.0233	1.84×10^{-4}
Phosphatidylserine	0.0058	4.60×10^{-5}
Sphingomyelin	0.0175	1.38×10^{-4}

After tissue specific models were generated, of the 10,600 reactions in the Recon3D model, the healthy pancreas model had 7,599 reactions, while the pancreatic cancer model had 7,666 reactions. Since this process forced the inclusion of exchange reactions and the removal of the "drug metabolism" or "xenobiotics metabolism" subsystems, the effective number of reactions that were subject to tissue specific modeling were 8,238 reactions in the Recon3D model, and the effective number of tissue-specific reactions that the algorithm assigned to the healthy and cancer models were 5,812 and 5,879, respectively. This corresponds to an approximate reduction in size of 29% for both networks.

4.2.2. Subsystem-Scale Comparison and Evaluation of Models

The 88 subsystems in the cancer model and 89 subsystems in the cancer model were subjected to comparison based on their sizes and their inclusion of reactions with

significantly different flux between the two models. In addition, the set of significantly different reactions in each subsystem underwent enrichment analyses separately to pinpoint which subsystems were more likely to contain pathways relevant to cancer development.

4.2.2.1. Subsystem Size Differences. The sizes and reaction breakdowns-by-subsystem of the resulting models are given in the appendix in Table A.6. for each subsystem with a size difference. The "exchange/demand," "drug metabolism" and "xenobiotics metabolism" were removed from Table A.6. since they are completely identical across models. The average reaction number difference between the cancer and healthy models of all subsystems was 2. An abbreviated version of Table A.6. is reproduced in Table 4.5, which contains only subsytems whose size difference is over 3 reactions between the cancer and healthy models, and no transport subsystems, which are harder to interpret. In Table 4.5, entries marked L mean that the Recon3D subsystem contains 100 reactions or more, whereas entries marked C mean over 75% of the Recon3D subsystem has been conserved in both tissue specific models, cancer and healthy.

Table 4.5. Subsystems whose size differs between the healthy and cancer models by more than 3 reactions.

Subsystem	# Reactions (Recon3D)	# Reactions (Healthy)	# Reactions (Cancer)	Abs. Diff. (Cancer-Healthy)	Abs. Diff. as % of Subsystem Size
ROS detoxification	7	1	6	5	71%
Keratan sulfate synthesis	22	11	22	11	50%
Thiamine metabolism	6	3	6	3	50%
Androgen and estrogen synthesis and	26	8	19	11	42%
Biotin metabolism	12	6	10	4	33%
Blood group synthesis	47	45	31	14	30%
Tryptophan metabolism	45	25	35	10	22%
Tetrahydrobiopterin metabolism[C]	18	18	15	3	17%
NAD metabolism	27	21	24	3	11%
Inositol phosphate metabolism	65	46	40	6	9%
Steroid metabolism	90	28	36	8	9%
Tyrosine metabolism	93	63	70	7	8%
Arachidonic acid metabolism	41	24	27	3	7%
Keratan sulfate degradation	76	15	20	5	7%
Methionine and cysteine metabolism[C]	46	41	38	3	7%
Glycine, serine, alanine, and threonine	47	29	32	3	6%
Sphingolipid metabolism[L]	133	93	85	8	6%
Bile acid synthesis[L]	185	93	104	11	6%
Glycosphingolipid metabolism	71	40	44	4	6%
Fatty acid synthesis[L]	239	169	160	9	4%
Cholesterol metabolism[L]	242	100	97	3	1%
Fatty acid oxidation[L,C]	961	747	738	9	1%

It can be observed in Table 4.5 that the cancer model contains more reactions in the "ROS (reactive oxygen species) detoxification" subsystem. Cancer's footprint is well-known to include oxidative stress due to the buildup of ROS inside the cell [146]. The expression of the KRAS oncogene in pancreatic cancer triggers several pathways leading to an increase in the level of ROS in the cell, which in turn promotes proliferation. However, an excess of ROS is damaging to the cell, so the cell's antioxidant metabolism needs to compensate for this effect [147, 148]. This finding is therefore promising for the predictive capabilities of the cancer metabolic model.

The cancer model was found to have more reactions in subsystems related to keratan sulfate metabolism. Keratan sulfates are a subtype of the glycosaminoglycan (GAG) class of molecules found primarily in the cornea, but also is prevalent in cartilage and bone. GAGs are generally found bound to surfaces of proteins, and take part in hydration and organization of the extracellular environment in tissues [149]. Keratan sulfates also take part in cell signaling by mediating interactions between proteins. The sulfation patterns of these molecules has been found to regulate the nature of such interactions [149, 150]. Research on the activity of this keratan sulfates and heparan sulfates, which are an adjacent class of GAGs, have found that they were instrumental in stimulating cell proliferation and growth in healthy cells, but were shown to trigger apoptosis in cancer cells. In the case of inhibiting cancer growth, these chemicals were found to bind to and inhibit the TGF-β pathway by binding to the epidermal growth factor receptor, EGFR [150–152]. In general, heparan sulfates are a better studied class of GAGs and some heparan sulfates like the glypican class of molecules are already considered putative biomarkers of pancreatic cancer [18, 150, 153]. Radwanska et al. [154] have found that lumican, a proteoglycan whose surface is modified by the attachment of keratan and heparan sulfates, inhibited the migration of A375 melanoma cells *in vitro*. In another study, Leiphrakpam et al. [155] have found increased levels of keratan sulfates in primary tumors of PDAC cases compared to healthy pancreatic tissue. They observed that keratan sulfate levels were further enhanced in lung-metastasized tumors with pancreatic origin, compared to the pancreatic primary tumor. However, the same effect was not observed with liver-metastasized tumors with pancreatic origin. This finding of keratan sulfate metabolism being more populated in the cancer model than in the healthy model might be coupled to the fact that the "methionine and cysteine metabolism"

subsystem is less populated in the cancer model. It could be surmised that the expression of the genes in the keratan sulfate metabolism have drawn metabolic flux away from the metabolism of sulfur-containing amino acids. This subsystem size difference could therefore point at sulfur stress, or as previously discussed, a dysregulation of the extracellular matrix and cell-cell adhesion phenomena in PDAC.

"Tetrahydrobiopterin metabolism", shown in Table 4.5 to have 3 fewer reactions in the cancer model compared to the healthy model, comprises the reactions of the cofactor tetrahydrobiopterin, also known as BH4. BH4 is an important cofactor in tryptophan and phenylalanine metabolisms, the production of nitric oxide (NO), and the immune pain response [156]. Since NO is known to promote angiogenesis, the proliferation of blood vessels, BH4 has been studied for its relevance to angiogenesis in cancer [157, 158]. However, a study by Cronin et al. has investigated the effects of BH4 in modulating T-cell activation through an immune response, to have an antitumor effect. They report that immune response to tumors is suppressed due to the action of a tryptophan-related molecule, kynurenine, and increased amounts of BH4 lead to T-cell antitumor activity. They furthermore claim that the immune response promoted by BH4 occurs long before NO production increases [159]. Such findings indicate that tetrahydrobiopterin and tryptophan metabolisms may be interlinked in their relevance to pancreatic cancer.

Another area where the two models differ are the subsystems "androgen and estrogen synthesis of metabolism" and "steroid metabolism." Androgens and estrogen being hormones that are steroids themselves, these subsystems can be interpreted together. Breast and prostate cancers have especially been investigated for the involvement of such steroid hormones in their growth and proliferation. These cancers were found to express estrogen receptor (ER) and androgen receptor (AR), which upon being stimulated by tissue estrogen or androgens, initiate a signaling cascade leading to proliferation [160–162]. In pancreatic cancer, Fernandez-del Castillo et al. [163] have found pancreatic cancer to be defined by lowered serum testosterone. TSPAN1, which in Table 4.2 was shown to be significantly differentially expressed, has its expression regulated by androgen metabolism and was found to be related to pancreatic cancer metastasis by Munkley et al. [164]. SRD5A3's status as a DEG and a drug target were discussed previously. These steroids which make up the human

sex hormones may be an important avenue of exploration indeed. In terms of non-sex-hormone steroids, a study by Liu et al. [165] has found that dexamethasone, a steroid molecule commonly used as a drug, had the effect of inducing cancer proliferation and enhancing TGF-β expression in grafts of human pancreatic cancer. Yet another study by Chen et al. [166] showed that total serum cholesterol was correlated with lowered pancreatic cancer risk. These results, more than anything, point to the complications posed by the many different types of steroid molecules that coexist and coregulate the metabolism.

"Arachidonic acid metabolism" and "bile acid metabolism" subsystems are two more subsystems exhibiting size differences between the phenotypes whose involvement in pancreatic cancer may be interlinked. Arachidonic acid is a precursor metabolite that is converted by cyclooxygenase (COX) and lipoxygenase (LOX) enzymes into prostaglandins and leukotrienes, respectively. These signaling metabolites are often produced at sites of trauma and inflammation, and are found to be in abundance in tumors, tumor-adjacent tissues, and tumor precursor lesions [167, 168]. Studies have shown COX-2 upregulation in several gastrointestinal cancer types, as well as in pancreatic cancer [169–172]. While the COX and LOX enzymes can be understood to be results of metabolic dysregulation, research has focused on ways in which an increase in the activity of these enzymes could be a cause of disease, particularly pancreatic cancer [168]. Covey et al. [173] have suggested that the activity of these enzymes leads to the inhibition of the PTEN tumor suppressor protein, allowing oncogenic signaling through the PI3K/AKT pathway to continue. Similarly, bile acids are known to be involved in the development of pancreatic inflammation, like due to bile reflux into the pancreatic ducts [174, 175]. Tucker et al.'s [176] study found that bile acid introduction caused dose-dependent increases in COX-2 and prostaglandin concentrations in pancreatic cancer cell lines. Outside of the pancreas, bile acids are implicated especially in inflammation and cancer development in the colon. An increase in secondary bile acid levels in the colon is known to indicate higher risk for colon inflammation and cancer [177–179].

The two models were found to show marked differences in the "biotin metabolism" and "thiamine metabolism" subsystems. Both these chemicals are part of the vitamin B complex. Biotin, known to regulate fatty acid metabolism, has been considered as a cancer

biomarker in some exploratory work, going as far back as 1949 [180, 181]. However, the main focus on biotin in cancer research has been as a cofactor of drug molecules, aiding in the targeting and complexation of cancer cell proteins with the anticancer agents [182, 183]. While thiamine deficiency has been associated with diabetes and alcoholism, conditions with known deleterious effects in the pancreas [184–186], no pancreatic cancer-specific literature is available. It is known that thiamine can stimulate cell growth, and mice experiments by Comin-Anduix et al. [187] have shown that cancer cells exhibit thiamine deficiency and increased growth when given thiamine. A vitamin B complex molecule not mentioned in Table 4.5 is folate. Folate metabolism was found to differ only by a single reaction between the two models. However, folate, which is central to pathways of producing and methylating DNA is already a molecule under much scrutiny [188–190]. There already exist anticancer drugs such as Methotrexate targeting dihydrofolate reductase (DHFR) [191] but Koseki et al. [192] have suggested that there exist other potential enzymes on this pathway that are more attractive targets. Most current research in the folate metabolism's role in pancreatic cancer focuses on the methylene tetrahydrofolate reductase protein, encoded by the MTHFR gene [193–195]. Studies by Nie et al. and Liu et al. have shown mutations in the MTHFR gene to increase pancreatic cancer risk [194, 195]. A study by Chittiboyina et al. on the mechanism of the folate metabolism in pancreatic cancer reports that MTHFR (among others) exhibits mutations, whereas DHFR does not. They further suggest that the dysregulation in the folate metabolism can be linked to dysregulation in the methionine and choline metabolisms, since methylene tetrahydrofolate and betaine, itself a choline derivative, compete as alternative methyl group donors in the production of methionine from homocysteine [193]. This may indicate that the dysregulation of the vitamin B metabolism may be related to dysregulation in the keratan sulfate metabolism, pointing to the interconversion of sulfur-containing amino acids and sulfur stress as salient aspects in pancreatic cancer.

The "inositol phosphate metabolism" subsystem is another difference that bears examination. Inositol and its various phosphates are a class of molecules with regulatory activity in cellular signaling, both on the insulin pathway and on the PI3K pathway, which is relevant to cellular growth, proliferation, and apoptosis [196]. Assmann et al. have proposed that insulin itself, in pancreatic endocrine cells, has a regulatory effect on PI3K

pathways [197]. Diminished levels of inositol phosphates have been reported in pancreatic cancer by Battini et al. [34] and the cancer inhibitory effect of supplemental inositol hexaphosphate is currently the subject of much study across various cancers [198–200], including pancreatic cancer [201].

The fatty acid, sphingolipid and glycosphingolipid metabolisms showing differences between the models is also an indication that the dysregulation of the lipid metabolism is a significant factor in distinguishing the two phenotypes. This phenomenon is already subject to study in the context of many cancers [35, 202], especially for sphingolipids and their derivatives. Several authors such as Akita et al. and Mayerle et al. [24, 26] have already identified serum sphingolipid levels to be potential biomarkers. Sphingolipids and glycosphingolipids' activity in promoting or suppressing tumor proliferation and growth is a complex area of research, due to the opposing effects the different types of these molecules have on cell metabolism, and the interconvertibility of these molecules. Specifically, the apoptotic effect of ceramide versus the proliferative effect of sphingosine-1-phosphate is oft-discussed [30–33].

4.2.2.2. Average Reaction Flux Differences. The differences between the flux sampling distributions for each reaction were quantified by a two-sample t-test between the healthy and cancer metabolic models. The two models were found to be different across 4,003 reactions, corresponding to 1,345 genes. Of these reactions, 2,043 were found to be transport or exchange reactions. These reactions were removed prior to interpretation, leaving 1,960 metabolic reactions corresponding to 1,095 genes. To further narrow down the list of reactions to be interpreted individually, a subset of these metabolic reactions whose difference-to-standard-deviation ratio (DSD) were above 1 was extracted. This subset captured those reactions which showed differences between the cancer and healthy models exceeding twice their average standard deviation, and included 250 metabolic reactions, corresponding to 206 genes. Additionally, only 14 exchange reactions were found to have DSD ≥ 1. This process and how it has affected the subsystem-wise distribution of significantly different reactions can be seen below.

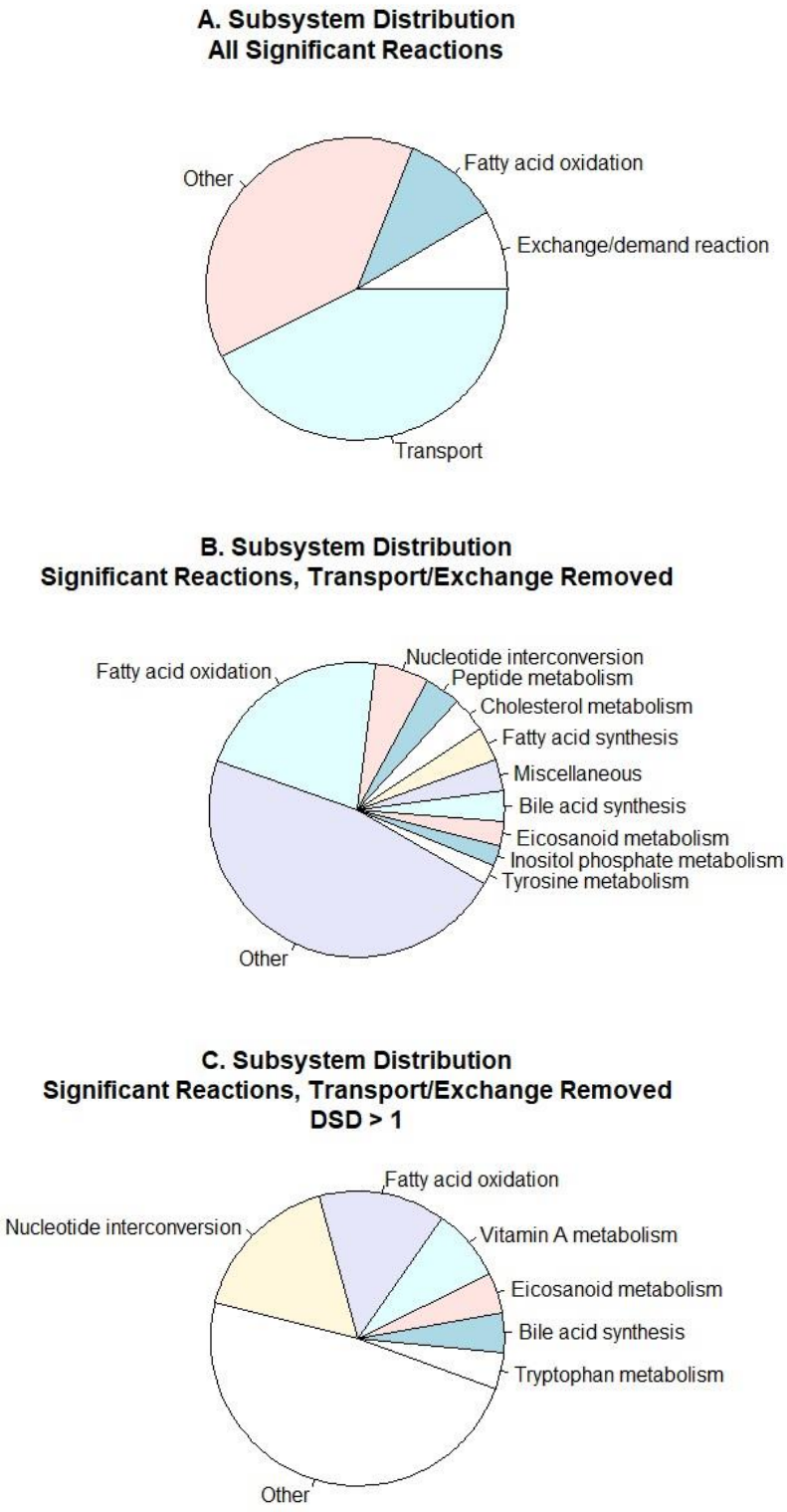

Figure 4.11. Subsystem distribution of significantly different reactions between healthy and cancer metabolic models.

The same information is tabulated in abbreviated form in Table 4.6. Table 4.6 additionally includes the proportion of significantly different reactions in each subsystem to the total number of reactions in that subsystem in the unmodified Recon3D model. For illustrative purposes, Table 4.6 contains only those subsystems which contain more than 5 reactions with significant flux differences and whose proportion of significantly different reactions is over 20%.

Table 4.6. Distribution of significantly different reactions across subsystems as number of reactions and proportion of subsystem size, subsystems where 5 or more significantly different reactions are found, and 20% of total reactions are significantly different.

Subsystem Name	**Number of Sig. Diff. Reactions**	**Total Number of Reactions**	**Proportion of Sig. Diff. Reactions**
Citric acid cycle	20	20	100%
Pyruvate metabolism	29	32	91%
Glutamate metabolism	14	16	88%
Taurine and hypotaurine metabolism	7	8	88%
ROS detoxification	6	7	86%
Oxidative phosphorylation	8	10	80%
Histidine metabolism	11	14	79%
Tryptophan metabolism	35	45	78%
Valine, leucine, and isoleucine metabolism	35	45	78%
Glycolysis/gluconeogenesis	32	42	76%
NAD metabolism	20	27	74%
Methionine and cysteine metabolism	34	46	74%
Propanoate metabolism	11	15	73%
Nucleotide interconversion	117	164	71%
Urea cycle	28	40	70%
Beta-Alanine metabolism	7	10	70%
Heme synthesis	13	19	68%

Table 4.6. Distribution of significantly different reactions across subsystems as number of reactions and proportion of subsystem size, subsystems where 5 or more significantly different reactions are found, and 20% of total reactions are significantly different. (cont.)

Subsystem Name	Number of Sig. Diff. Reactions	Total Number of Reactions	Proportion of Sig. Diff. Reactions
Inositol phosphate metabolism	44	65	68%
Fructose and mannose metabolism	15	23	65%
Chondroitin sulfate degradation	28	44	64%
Androgen and estrogen synthesis and metabolism	16	26	62%
Lysine metabolism	19	31	61%
Tetrahydrobiopterin metabolism	11	18	61%
Vitamin C metabolism	6	10	60%
Vitamin A metabolism	28	47	60%
Pentose phosphate pathway	24	41	59%
Biotin metabolism	7	12	58%
Pyrimidine synthesis	12	21	57%
Eicosanoid metabolism	53	93	57%
Glutathione metabolism	9	16	56%
Phenylalanine metabolism	20	36	56%
Vitamin B2 metabolism	5	9	56%
Glycine, serine, alanine, and threonine metabolism	26	47	55%
Blood group synthesis	25	47	53%
Purine catabolism	18	34	53%
Keratan sulfate synthesis	11	22	50%
Pyrimidine catabolism	16	34	47%
Transport	1705	3744	46%
Miscellaneous	69	152	45%
Tyrosine metabolism	42	93	45%

Table 4.6. Distribution of significantly different reactions across subsystems as number of reactions and proportion of subsystem size, subsystems where 5 or more significantly different reactions are found, and 20% of total reactions are significantly different. (cont.)

Subsystem Name	Number of Sig. Diff. Reactions	Total Number of Reactions	Proportion of Sig. Diff. Reactions
Fatty acid oxidation	423	961	44%
Aminosugar metabolism	14	32	44%
Arginine and proline metabolism	17	40	43%
Folate metabolism	25	59	42%
Arachidonic acid metabolism	17	41	41%
Steroid metabolism	37	90	41%
CoA synthesis	5	14	36%
Bile acid synthesis	64	185	35%
Glycosphingolipid metabolism	24	71	34%
Chondroitin synthesis	15	45	33%
Ubiquinone synthesis	6	18	33%
Peptide metabolism	77	242	32%
Cholesterol metabolism	75	242	31%
Fatty acid synthesis	73	239	31%
Sphingolipid metabolism	35	133	26%
Phosphatidylinositol phosphate metabolism	9	40	23%
Glycerophospholipid metabolism	37	165	22%
Keratan sulfate degradation	17	76	22%

Table 4.6 shows that subsystems with large numbers of significantly different reactions are not necessarily those that with a high proportion of significantly different reactions. However, many of the subsystems with significant size differences between the cancer and healthy models (as seen in Table 4.5) reappear as subsystems with many reactions with significantly different flux. However, Table 4.6 also shows that subsystems with many

significantly different reactions have a low or middling amount of significantly different reactions in proportion to their sizes. Furthermore, little overlap was found between subsystems with many significantly different reactions and subsystems with a high level of average flux difference. A scatter plot of subsystems according to their average flux differences, DSD, and number of significantly different reactions is shown in Figure 4.12. All subsystems were then clustered via hierarchical clustering with respect to these variables, and the result of this clustering is shown in Figure 4.13.

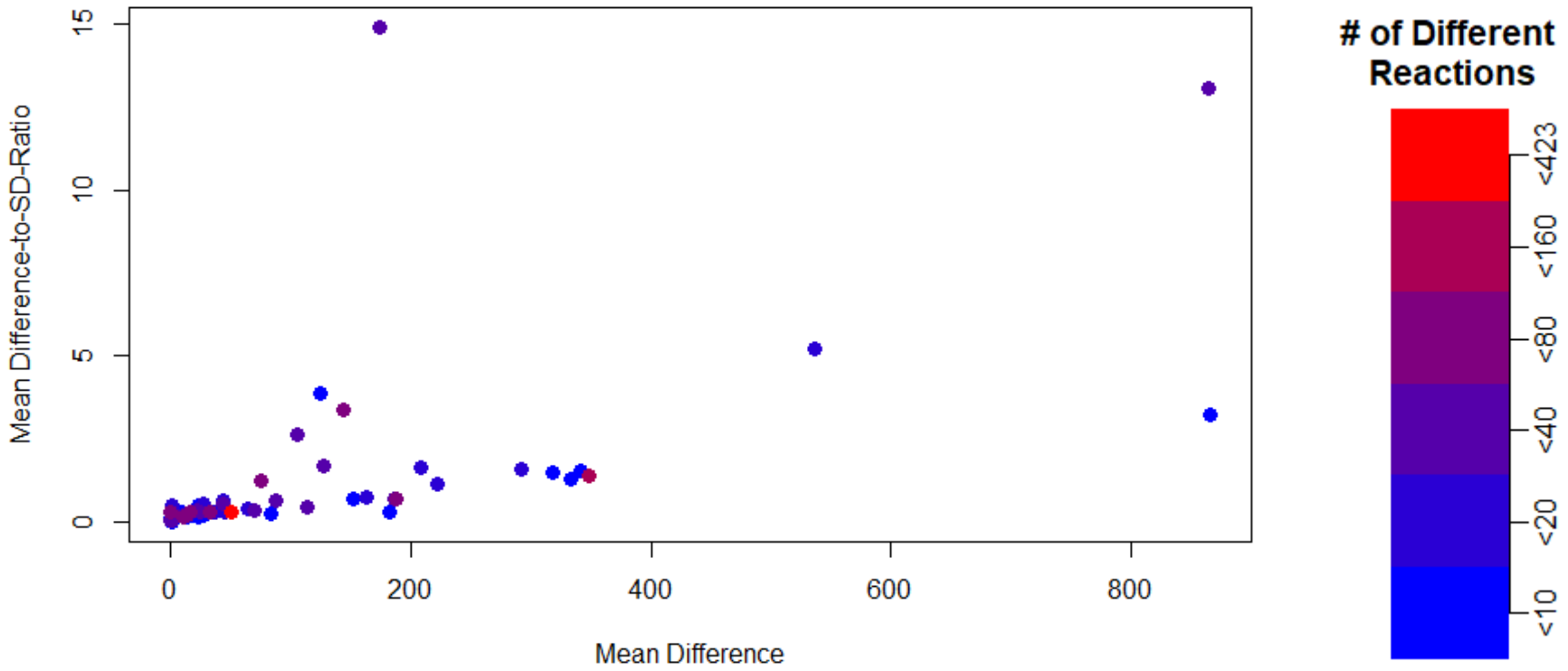

Figure 4.12. Scatter plot of subsystems according to significant differences between healthy and cancer metabolic models.

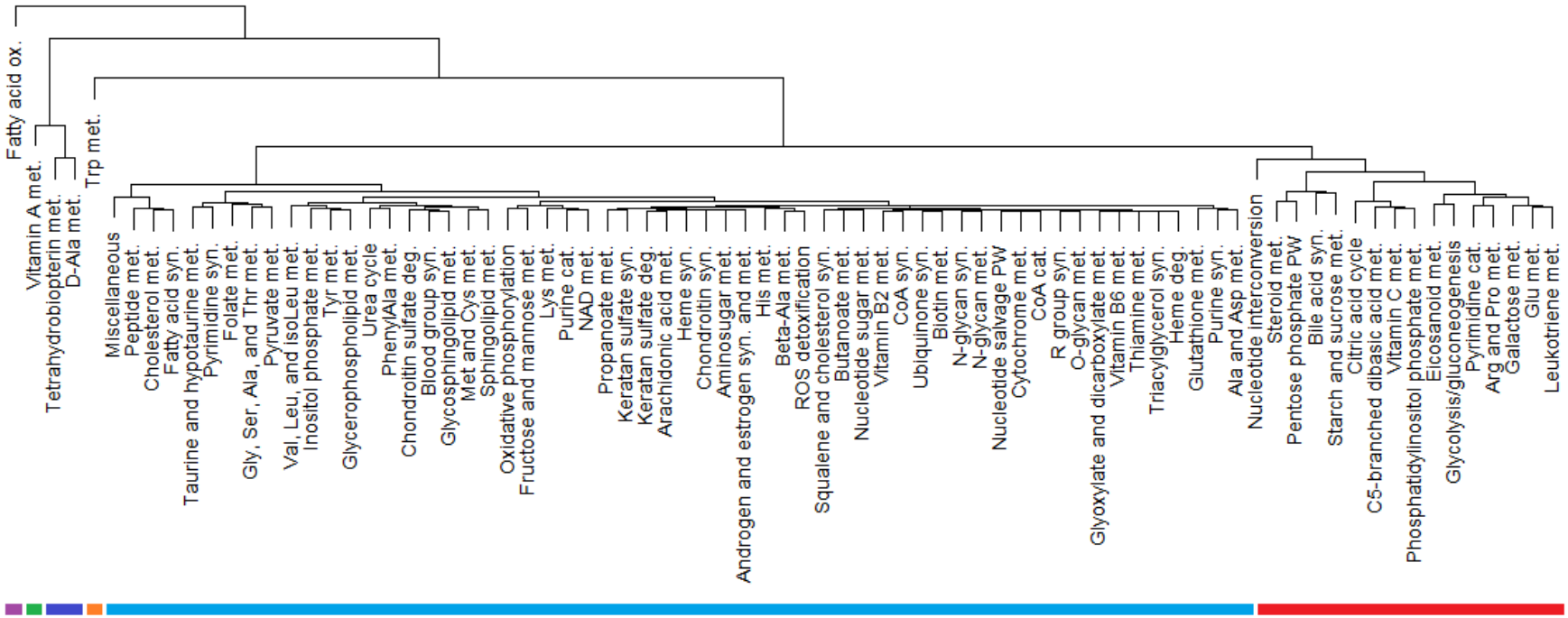

Figure 4.13. Subsystems clustered via hierarchical clustering.

4.2.3. Individual Reaction Flux Differences Between Models

Here the individual significant flux differences of each reaction between the healthy and cancer models, found via sampling of the optimal flux ranges obtained via FVA are reported. With a Benjamini-Hochberg false-discovery rate cutoff of 0.01, this corresponds to 1,960 metabolic reactions (neither transport nor exchange reactions) and 338 exchange reactions. A preliminary look at the distribution of reaction flux differences revealed that, for most reactions found to be significantly different, a flux difference over 10^{-6} fmol/cell/h was observed. However, some reactions with flux differences markedly below this value (in the 10^{-10} – 10^{-8} fmol/cell/h range) were observed especially in the "Histidine metabolism," "Peptide metabolism," and "Exchange/demand reactions" subsystems. These low-difference reactions composed 81%, 51%, and 17%, respectively, of all significantly different reactions in these three subsystems.

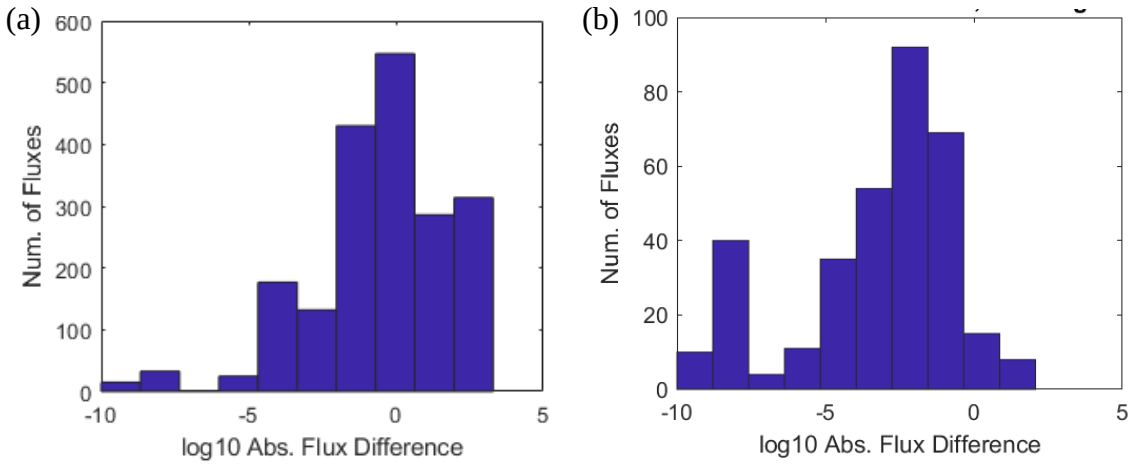

Figure 4.14. Flux difference distributions among significantly different reactions for (a) metabolic and (b) exchange reactions.

The flux difference distributions of the significantly different reactions among metabolic and exchange reactions are shown in Figure 4.14. These reactions were removed from the pool of significantly different reactions before further analysis, leaving 282 exchange and 1,861 metabolic reactions for consideration. Excluding these low-difference reactions still leaves many reactions for analysis with very low FDR-adjusted p-values.

What follows is the discussion of these significantly different reaction fluxes among exchange reactions and the central carbon metabolism. Attention has been focused on these subsystems since the central carbon metabolism is the best studied one for many organisms, and exchange reactions elucidate the metabolism's uptake or secretion behavior, potentially leading to the discovery of biomarkers that can be discovered outside of the tissue of interest.

4.2.3.1. Differences Among Exchange Reactions. The 282 exchange reactions showing significant differences between the healthy and cancer metabolic networks and whose fluxes differ by over 10^{-6} fmol/cell/h between the models are reported here. Green fluxes (negative) indicate uptake rates, yellow fluxes (positive) indicate secretion rates. Table 4.7 contains those exchange reactions whose metabolites are taken up both in the cancer and the healthy models.

Table 4.7. Exchange reaction flux differences for uptake reactions: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$.

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
Nicotinamide*	$<10^{-300}$	$(-84.8 \pm 0.7) \times 10^{-4}$	$(-1.1 \pm 0.1) \times 10^{-4}$	-8.4×10^{-3}	76
Taurocholate*	$<10^{-300}$	$(-81.2 \pm 0.1) \times 10^{-5}$	$(-21.7 \pm 0.2) \times 10^{-5}$	-6.0×10^{-4}	3.8
Sulfate	$<10^{-40}$	-0.37 ± 0.02	-1.01 ± 0.04	0.64	0.37
Water	$<10^{-300}$	-85 ± 1	-194 ± 1	109	0.44
D-Carnitine*	$<10^{-300}$	$(-64.4 \pm 0.1) \times 10^{-4}$	$(-33.3 \pm 0.1) \times 10^{-4}$	-3.1×10^{-3}	1.9
L-Methionine*	$<10^{-300}$	-0.770 ± 0.001	-0.471 ± 0.001	-0.30	1.6
Iron (II)	$<10^{-300}$	-351 ± 1	-233 ± 1	-118	1.5
Phosphate	$<10^{-10}$	-2.3 ± 0.1	-3.4 ± 0.1	1.1	0.69

Table 4.7. Exchange reaction flux differences for uptake reactions: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
L-Tyrosine*	$<10^{-300}$	-0.722 ± 0.001	-0.936 ± 0.001	0.21	0.77
L-Tryptophan	$<10^{-300}$	-0.522 ± 0.001	-0.415 ± 0.001	-0.11	1.3
Choline	$<10^{-100}$	-0.816 ± 0.005	-0.950 ± 0.003	0.13	0.86
L-Homoserine	$<10^{-300}$	-0.568 ± 0.001	-0.654 ± 0.001	0.086	0.87
L-Aspartate	$<10^{-4}$	-0.86 ± 0.02	-0.95 ± 0.01	0.090	0.91
L-Asparagine	$<10^{-2}$	-0.95 ± 0.01	-0.87 ± 0.02	-0.073	1.1
L-Phenylalanine	$<10^{-300}$	-0.941 ± 0.001	-0.870 ± 0.001	-0.071	1.1
L-Lysine	$<10^{-300}$	-9.567 ± 0.005	-9.979 ± 0.001	0.41	0.96
Glycocholate	$<10^{-300}$	(-994.5 ± 0.2) x 10^{-6}	(-961.2 ± 0.5) x 10^{-6}	-3.3 x 10^{-5}	1.03
L-Serine	$<10^{-50}$	-9.71 ± 0.01	-9.95 ± 0.01	0.25	0.98
L-Isoleucine	$<10^{-300}$	-9.986 ± 0.001	-9.829 ± 0.003	-0.16	1.02
Oxygen	$<10^{-50}$	-879.7 ± 0.6	-892.3 ± 0.6	13	0.99
L-Leucine	$<10^{-200}$	-9.989 ± 0.001	-9.899 ± 0.002	-0.090	1.01
Glyco-deoxycholate	$<10^{-9}$	(-996.4 ± 0.3) x 10^{-6}	(-993.4 ± 0.3) x 10^{-6}	-3.0 x 10^{-6}	1.003

Table 4.7. Exchange reaction flux differences for uptake reactions: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
Linoleate	$<10^{-30}$	$(-996.4 \pm 0.1) \times 10^{-4}$	$(-993.7 \pm 0.2) \times 10^{-4}$	-2.7×10^{-4}	1.003
Linolenate	$<10^{-5}$	$(-994.5 \pm 0.2) \times 10^{-4}$	$(-993.2 \pm 0.2) \times 10^{-4}$	-1.3×10^{-4}	1.001
L-Valine	$<10^{-2}$	$(-9997.8 \pm 0.3) \times 10^{-3}$	$(-9999.2 \pm 0.2) \times 10^{-3}$	1.4×10^{-3}	0.9999

Table 4.7 shows that while many significantly different reactions with low p-values were discovered, actual fold changes for many uptake reactions are very low. The models differ much more significantly in the metabolites they secrete compared to those they take up. Of the 156 metabolites with significantly different flux secreted by both models, only those with over 5-fold secretion increase or decrease (corresponding to FC greater than 5 or less than 0.2) were reported in Table 4.8 in the interest of brevity.

Table 4.8. Exchange reaction flux differences for secretion reactions with over 5-fold increase or decrease: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$.

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
Glycourso-deoxycholate	$<10^{-10}$	0.011 ± 0.01	(4.1 ± 0.5) x 10^{-7}	0.011	2600
Taurourso-deoxycholate	$<10^{-10}$	(4.5 ± 0.6) x 10^{-4}	(7.8 ± 0.7) x 10^{-7}	4.5 x 10^{-4}	580
Sulfite	$<10^{-20}$	(6.5 ± 0.6) x 10^{-4}	0.31 ± 0.03	-0.31	2.1 x 10^{-3}
L-Sulfolactate	$<10^{-20}$	(4.8 ± 0.5) x 10^{-4}	0.20 ± 0.02	-0.20	2.5 x 10^{-3}
L-Cysteate	$<10^{-10}$	(3.4 ± 0.4) x 10^{-4}	0.10 ± 0.01	-0.098	3.4 x 10^{-3}
L-Phe-L-Tyr	$<10^{-300}$	(4.3 ± 0.5) x 10^{-4}	(79.1 ± 0.7) x 10^{-3}	-0.079	5.4 x 10^{-3}
2-Oxoadipate*	$<10^{-300}$	0.036 ± 0.001	6.57 ± 0.01	-6.5	5.5 x 10^{-3}
Glutaryl-carnitine*	$<10^{-300}$	(47.4 ± 0.1) x 10^{-4}	(4.3 ± 0.2) x 10^{-5}	4.7 x 10^{-3}	110
1-Methyl-nicotinamide	$<10^{-300}$	(9.5 ± 0.1) x 10^{-3}	(11.2 ± 0.8) x 10^{-6}	9.4 x 10^{-3}	85
Homovanillate	$<10^{-300}$	(1.1 ± 0.1) x 10^{-3}	(85.5 ± 0.7) x 10^{-3}	-0.084	0.014
Indole-3-acetate	$<10^{-300}$	(12.4 ± 0.8) x 10^{-4}	(79.6 ± 0.7) x 10^{-3}	-0.078	0.016
L-Val-L-Trp-L-Val	$<10^{-300}$	(5.8 ± 0.6) x 10^{-4}	(29.8 ± 0.4) x 10^{-3}	-0.029	0.019

Table 4.8. Exchange reaction flux differences for secretion reactions with over 5-fold increase or decrease: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
L-Ile-L-Trp-L-Tyr	$<10^{-300}$	$(5.1 \pm 0.6) \times 10^{-4}$	$(23.9 \pm 0.4) \times 10^{-3}$	-0.023	0.021
3-Hydroxy-glutarate	$<10^{-300}$	0.278 ± 0.003	$(6.7 \pm 0.6) \times 10^{-3}$	0.27	41
L-Trp-L-Phe	$<10^{-300}$	$(7.4 \pm 0.7) \times 10^{-4}$	$(29.1 \pm 0.4) \times 10^{-3}$	-0.028	0.025
3-Hydroxy-isovaleryl-carnitine	$<10^{-300}$	$(7.8 \pm 0.6) \times 10^{-6}$	$(30.6 \pm 0.4) \times 10^{-5}$	-3.0×10^{-4}	0.025
Taurine	$<10^{-20}$	$(20.6 \pm 0.9) \times 10^{-4}$	0.08 ± 0.01	-0.073	0.027
Cholesterol	$<10^{-100}$	$(3.0 \pm 0.7) \times 10^{-3}$	0.103 ± 0.004	-0.10	0.029
L-Lys-L-Phe-L-Ile	$<10^{-300}$	$(5.5 \pm 0.5) \times 10^{-4}$	$(18.6 \pm 0.3) \times 10^{-3}$	-0.018	0.030
L-Tyr-L-Trp-L-Phe	$<10^{-300}$	$(4.2 \pm 0.5) \times 10^{-4}$	$(14.1 \pm 0.3) \times 10^{-3}$	-0.014	0.030
2-Hydroxy-isovalerate	$<10^{-300}$	0.147 ± 0.003	$7.0 \pm 0.6) \times 10^{-3}$	0.14	21
L-Asp-L-Met-L-Asp	$<10^{-300}$	$(6.3 \pm 0.6) \times 10^{-4}$	$(12.7 \pm 0.3) \times 10^{-3}$	-0.012	0.050
2-Hydroxyl-3-methylvalerate	$<10^{-200}$	0.067 ± 0.002	$(3.8 \pm 0.5) \times 10^{-3}$	0.063	18

Table 4.8. Exchange reaction flux differences for secretion reactions with over 5-fold increase or decrease: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
L-Tyr-L-Phe-L-Tyr	$<10^{-300}$	$(26.3 \pm 0.9) \times 10^{-4}$	$(45.8 \pm 0.4) \times 10^{-3}$	-0.043	0.057
Taurocheno-deoxycholate	$<10^{-10}$	$(4.3 \pm 0.4) \times 10^{-4}$	0.007 ± 0.001	-6.1×10^{-3}	0.065
L-Phe-L-Leu	$<10^{-300}$	$(25.2 \pm 0.4) \times 10^{-3}$	$(1.8 \pm 0.1) \times 10^{-3}$	0.023	14
L-Met-L-Met-L-Ile	$<10^{-300}$	$(32.0 \pm 0.9) \times 10^{-4}$	$(43.3 \pm 0.3) \times 10^{-3}$	-0.040	0.074
L-Trp-L-Ile-L-Trp	$<10^{-300}$	$(14.4 \pm 0.6) \times 10^{-4}$	$(19.2 \pm 0.2) \times 10^{-3}$	-0.018	0.075
L-Trp-L-Tyr-L-Tyr*	$<10^{-300}$	$(9.8 \pm 0.2) \times 10^{-3}$	$(124.2 \pm 0.7) \times 10^{-3}$	-0.11	0.079
L-Lactate	$<10^{-200}$	0.156 ± 0.004	0.013 ± 0.003	0.14	12
Methylamine	$<10^{-40}$	$(2.9 \pm 0.6) \times 10^{-4}$	$(3.3 \pm 0.2) \times$ 10-3	-3.1×10^{-3}	0.087
15-HPETE	$<10^{-20}$	$(1.8 \pm 0.1) \times 10^{-4}$	$(1.9 \pm 0.4) \times 10^{-5}$	1.6×10^{-4}	9.9
L-Trp-L-Lys	$<10^{-100}$	$(4.2 \pm 0.1) \times 10^{-3}$	$(4.5 \pm 0.5) \times 10^{-4}$	3.7×10^{-3}	9.3
L-Trp-Gly-L-Leu	$<10^{-100}$	$(7.0 \pm 0.6) \times 10^{-4}$	$(6.2 \pm 0.2) \times 10^{-3}$	-5.5×10^{-3}	0.11
L-Ile-L-Arg-L-Ile	$<10^{-200}$	$(2.8 \pm 0.3) \times 10^{-3}$	$(23.1 \pm 0.5) \times 10^{-3}$	-0.020	0.12

Table 4.8. Exchange reaction flux differences for secretion reactions with over 5-fold increase or decrease: Sorted by absolute-log-fold-change (cancer/healthy), (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**FC (Cancer/ Healthy)**
L-Tyr-L-Val-L-Met	$<10^{-100}$	$(5.0 \pm 0.5) \times 10^{-4}$	$(4.1 \pm 0.1) \times 10^{-3}$	-3.6×10^{-3}	0.12
L-Val-L-Val	$<10^{-300}$	$(29.3 \pm 0.9) \times 10^{-3}$	0.218 ± 0.002	-0.19	0.14
Tryptamine	$<10^{-90}$	$(4.9 \pm 0.5) \times 10^{-4}$	$(3.5 \pm 0.1) \times 10^{-3}$	-3.0×10^{-3}	0.14
L-Met-L-Trp-L-Phe	$<10^{-200}$	$(9.2 \pm 0.7) \times 10^{-4}$	$(6.4 \pm 0.2) \times 10^{-3}$	-5.5×10^{-3}	0.14
Acetone	$<10^{-300}$	1.80 ± 0.03	0.27 ± 0.01	1.5	6.8
Urate*	$<10^{-300}$	$(47.7 \pm 0.5) \times 10^{-5}$	$(31.8 \pm 0.1) \times 10^{-4}$	-2.7×10^{-3}	0.15
L-Leu-L-Val	$<10^{-70}$	0.030 ± 0.001	$(4.9 \pm 0.5) \times 10^{-3}$	0.025	6.1
Mevalonate	$<10^{-100}$	0.39 ± 0.01	0.07 ± 0.01	0.32	5.8
L-Leu-L-Trp	$<10^{-300}$	$(3.4 \pm 0.1) \times 10^{-3}$	$(17.4 \pm 0.3) \times 10^{-3}$	-0.014	0.19
Cholesterol-C16 Lipid Ester	$<10^{-80}$	$(9.7 \pm 0.4) \times 10^{-4}$	$(1.9 \pm 0.2) \times 10^{-4}$	7.9×10^{-4}	5.2
W-Hydroxy-leukotriene B4	$<10^{-30}$	$(3.7 \pm 0.2) \times 10^{-4}$	$(7.2 \pm 0.9) \times 10^{-5}$	3.0×10^{-4}	5.1

The two models had between them 101 exchange reactions whose fluxes were found to be “significantly different” in one of the following two ways: either only one model

included the reaction (the reaction in the other model was treated as having zero flux), or the reaction was reversible and showed flux in different directions between models (i.e., one model producing the "products" from the "reactants," the other producing "reactants" from "products" using the same enzymes). For such reactions, the concept of fold change is meaningless. Here in Table 4.9, the 19 such exchange reactions with DSD values over 0.05 are shown, corresponding to those reactions among this subset which have shown the greatest mean difference fluxes in proportion to their standard deviation.

Table 4.9. Exchange reaction flux differences for reactions that exist in only one model or switch directions between models, with $DSD \geq 0.05$: Sorted by DSD, (*) indicates reactions with $DSD \geq 1$.

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**DSD**
Anthranilate*	$<10^{-300}$	0.457 ± 0.001	0	0.46	1.3
Succinate	$<10^{-300}$	(26.8 ± 0.1) x 10^{-3}	0	0.027	0.97
3-Hydroxy-anthranilate	$<10^{-300}$	(81.6 ± 0.7) x 10^{-6}	0	8.2 x 10^{-5}	0.50
4-Methylpentanal	$<10^{-300}$	0.516 ± 0.009	0	0.52	0.24
R-Pantothenate	$<10^{-100}$	(17.7 ± 0.7) x 10^{-5}	(-138.0 ± 0.9) x 10^{-8}	0.00018	0.20
3-Hydroxy-L-Kynurenine	$<10^{-300}$	(9.2 ± 0.2) x 10^{-6}	0	9.2 x 10^{-6}	0.17
Kynurenate	$<10^{-300}$	(7.9 ± 0.2) x 10^{-6}	0	7.9 x 10^{-6}	0.15
Picolinate	$<10^{-200}$	(7.6 ± 0.2) x 10^{-3}	0	0.0076	0.15
Testosterone	$<10^{-200}$	0.216 ± 0.006	0	0.22	0.15

Table 4.9. Exchange reaction flux differences for reactions that exist in only one model or switch directions between models, with $DSD \geq 0.05$: Sorted by DSD, (*) indicates reactions with $DSD \geq 1$. (cont.)

Exchange Reaction of Metabolite	**Adj. P. Val. (B-H)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**DSD**
Lithocholate	$<10^{-100}$	$(-2.9 \pm 0.1) \times 10^{-6}$	0	-2.9×10^{-6}	0.088
3,4-Dihydroxy-L-phenylalanine	$<10^{-70}$	$(3.7 \pm 0.2) \times 10^{-3}$	0	0.0037	0.075
Superoxide	$<10^{-70}$	$(15.2 \pm 0.8) \times 10^{-3}$	0	0.015	0.073
Serotonin	$<10^{-60}$	0	$(1.8 \pm 0.1) \times 10^{-5}$	-1.80×10^{-5}	0.071
Estradiol	$<10^{-60}$	0.034 ± 0.002	0	0.034	0.070
7-alpha-Hydroxy-cholest-4-en-3-one	$<10^{-60}$	0.043 ± 0.002	0	0.043	0.069
Docosane-dioate	$<10^{-50}$	0.059 ± 0.004	0	0.059	0.065
7-Ketolitho-cholate	$<10^{-50}$	0.032 ± 0.002	0	0.032	0.063
Lanosterol	$<10^{-40}$	0	0.032 ± 0.002	-0.032	0.059
16-alpha-Hydroxy-estrone	$<10^{-40}$	0.022 ± 0.002	0	0.022	0.057

The most different reaction in Table 4.7 is the exchange of nicotinamide, a water-soluble form of vitamin B3 that is a precursor for NADH/NADPH, two ubiquitous redox cofactors. The cancer model was found to take up many times the nicotinamide the healthy model took up from the available pool. Nicotinamide, as a vitamin, is comparatively widely available as a supplement, and studies have focused on its preventative effect on skin cancer [203]. Table 4.8 also shows the cancer model secreting a significantly larger amount of 1-methylnicotinamide compared to the healthy model, which in Recon3D is a byproduct of the metabolic conversion of methionine to homocysteine. One of the most active pathways by which nicotinamide is converted into cofactors and enters the cellular metabolism is through the action of the nicotinamide phosphoribosyl-transferase (Nampt) enzyme, encoded by the NAMPT gene. Nampt exists both in intracellular and extracellular form, acting both as an enzyme and as a signaling protein. Sun et al. and Guo et al. have found both that NAMPT was overexpressed in cancerous tissue and that Nampt's inhibition led to cancer cell mortality and reduced proliferation, in prostate cancer and glioblastoma, respectively [204, 205]. While Guo et al.'s work makes no distinction between intracellular and extracellular Nampt, Sun et al. have specifically discovered increased levels of extracellular Nampt, and their therapeutic intervention has targeted the inactivation of extracellular Nampt [204]. A study by Chini et al. [206] have found that inhibition of Nampt inhibits cell growth and proliferation in pancreatic cancer cell lines. NAMPT expression was not found to be significantly different between pancreatic cancer cell lines and healthy pancreatic cells in either Chini et al.'s study or this present study. Paradoxically, the Nampt enzyme (represented by a reaction labeled "NMNS" in Recon3D) was found to exist only in the healthy metabolic model in this study, with a mean sampling flux of 0.250 ± 0.003 fmol/cell/h.

Another group of metabolites represented well among these significant exchange reactions are related to the metabolism of the amino acid tryptophan. Anthranilate and indole-3-acetate are precursors for tryptophan production, a process that also yields glyceraldehyde-3-phosphate (G3P) which is a glycolysis intermediate. Kynurenine and related molecules are intermediates in the production of nicotinamides and picolinate, and their production is related to the cellular response to oxidative stress [19–21, 207]. Many of these intermediate metabolites have immunosuppressive effects on T-cells [207, 208].

Kynurenine and anthranilate can then be re-utilized by the central carbon metabolism by entering glycolysis, during which process 2-oxoadipate is an important intermediate. Finally, tryptophan is a precursor for the synthesis of the neurotransmitter metabolite serotonin, and tryptamine is a metabolic intermediate during serotonin synthesis [19–21, 207]. Considering its many varied functions, the tryptophan pathway likely reflects a lot of the disparate-seeming exchange reactions seen in Table 4.7. to Table 4.9. Cancer research on this pathway has most significantly focused on the kynurenine portion of the pathway, especially regarding the activity of the indoleamine 2,3-dioxygenase enzymes (encoded by IDO1 and IDO2 genes) in preventing successful treatment. This present study has found IDO1 to show significant metabolic difference ($p < 0.01$) between the healthy and cancer phenotypes and $DSD \geq 1$, even though it did not show any significant differential expression. Nevler et al. have published a study that relates these proteins to treatment resistance in pancreatic ductal adenocarcinoma [209], and other researchers have echoed these findings for other subtypes of cancer [210, 211]. Beyond attempting to overcome treatment resistance, Huang et al. have suggested that serum 3-hydroxyanthranilic acid and 3-hydroxykynurenine levels could be used as a diagnostic biomarker of pancreatic cancer [212]. Meanwhile, Nam et al. have reported in [213] that in the case of breast cancer, increased indoleacetates, increased homovanillate, and decreased urea in urine could be used as a diagnostic biomarker panel. These findings mirror this present study's exchange fluxes.

Taurocholate, along with glycocholate and glycodeoxycholate, are three conjugated bile acids that have shown differences in uptake rate between the cancer and healthy models. In each case, the cancer model showed higher uptake of the metabolite. Of these, taurocholate is the most different between models, taken up at a rate in the cancer model that is over 500 times that of the healthy model. The cancer model was also found to secrete tauroursodeoxycholate, glycoursodeoxycholate and taurochenodeoxycholate, all conjugated bile acids, more than the healthy model. Interestingly, a study by Goldman et al. [214] found that glycoursodeoxycholate had a protective effect against DNA damage and oxidative stress in esophageal cancer. Conjugated bile acids are the form of bile acids that have been secreted into the intestine before modification by gut bacteria into becoming secondary bile acids. While secondary bile acids are implicated generally in promoting colon cancer [177–179], it is proposed by Ridlon et al. and Wolf et al. in [177, 178] that taurocholate metabolism by

certain gut bacteria causes hydrogen sulfide generation, exacerbating colon cancer risk in a diet-dependent fashion. While the pancreas has no “gut bacteria” to speak of, Table 4.9 shows that the cancer model secretes 7-ketolithocholate, which is a secondary bile acid. Such a finding could be related with the fact that the healthy model was found to take up significantly more sulfate ions from the medium and secrete significantly more sulfite ions, sulfolactate, and sulfur-containing amino acid byproducts like cysteate and tripeptides containing methionine, implicating differences in sulfur stress setting the models apart. Furthermore, the secretion of leukotriene and 15-HPETE were also found to be increased in the cancer model. These are metabolites in the arachidonic acid metabolism, related to trauma and inflammation. Subsection 4.2.2.1 discussed how the bile acid metabolism and arachidonic acid metabolisms are interrelated.

The succinate secretion seen in the cancer metabolic model is seemingly in line with findings by Rahn et al. in [215] that upregulation of succinate dehydrogenase (SDH) is associated with a decrease in succinate accumulation and cancer cell proliferation in pancreatic ductal adenocarcinoma. A study on mouse and human lung cancer cell lines by Wu et al. reveals that succinate has a dual signaling role in enabling cancer proliferation: intracellularly, accumulated succinate promotes tumorigenesis and EMT, and extracellularly, secreted succinate recruits macrophages to protect tumor tissue from an immune response [216].

Other notable differences between the cancer and healthy metabolic models’ exchange flux differences are briefly discussed here. Several exchange reactions were also occupied by the various steroids that were secreted by the cancer and healthy metabolic models. The cancer model showed a decreased secretion of cholesterol and lanosterol, and an increased secretion of the steroid sex hormones, testosterone and estradiol. A significant uptake difference was found for D-carnitine as well. Differences in carnitine metabolism are contextualized in a review by Melone et al. [217], which states that a divergent carnitine metabolism may be a significant enabler of how cancer cells use fatty acid biosynthesis and oxidation to supplement their aberrant glycolysis-based energy metabolism (the Warburg effect). Last but not least, the cancer metabolic model showed a marked decrease in its water uptake compared to the healthy metabolic model, and AQP5, which encodes a subunit of the

aquaporin protein involved in forming cellular water channels, was found to be a DEG overexpressed more than 32-fold.

4.2.3.2. Differences in the Central Carbon Metabolism. The "Glycolysis/gluconeogenesis," "Citric acid cycle," and "Oxidative phosphorylation" subsystems make up the central carbon metabolism and contain between them 61 reactions found to be significantly different between models. In these subsystems, the most salient difference was found in the cancer model's use of a variant cytochrome C oxidase reaction which produced the reactive oxygen species superoxide anion (O_2^-) as a byproduct. The cytochrome C oxidase enzyme exists on the internal membrane of the mitochondrion and is oxidized by oxygen to pump protons from the lumen of the mitochondrion to the intermembrane space. These protons subsequently generate the concentration gradient which allows ATP synthase to produce ATP from ADP. The healthy reaction, labeled "CYOOm2i" in Recon3D, and the variant reaction, labeled "CYOOm3i" in Recon3D, can be presented as

$$4\,Fe(II) - cytC + O_2 + 8\,H^+(m) \rightarrow 4\,Fe(III) - cytC + 2\,H_2O + 4\,H^+(i)\ \textit{(CYOOm2i)} \tag{4.2}$$

and

$$4\,Fe(II) - cytC + O_2 + 7.92\,H^+(m) \rightarrow 4\,Fe(III) - cytC + 1.96\,H_2O + 4\,H^+(i) + 0.02\,O_2^-\ \textit{(CYOOm3i)} \tag{4.3}$$

respectively.

In Equation (4.2) and Equation (4.3), H^+(m) refers to protons within the mitochondrial lumen, and H^+(i) refers to protons in the intermembrane space. Table 4.10 shows the flux differences between the healthy and cancer models of these two reactions.

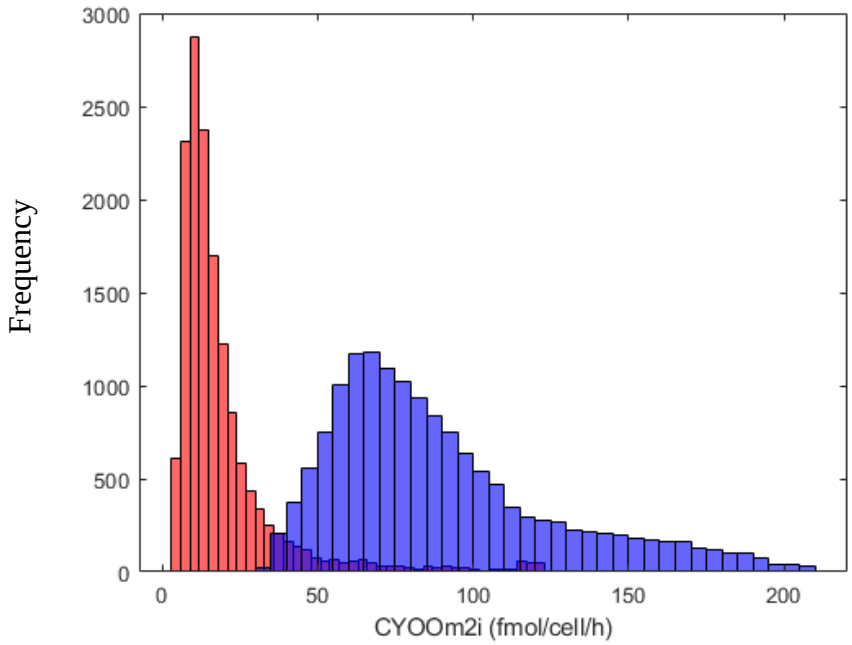

Figure 4.15. Flux sampling results for cytochrome C oxidase, blue: healthy, red: cancer.

Figure 4.15 illustrates the sampling results of CYOOm2i, showing the cancer model utilizing this reaction at a lower rate, instead using CYOOm3i. Since cytochrome C oxidase is an enzyme with multiple subunits, it is encoded by a variety of genes acting in conjunction. These genes are MT-CO1 through MT-CO3, and COX4 through COX8, with the MT-CO genes being named so to prevent confusion with the unrelated cyclooxygenase-encoding COX genes. Even though the flux differences between the models for this reaction have been found to be significant, none of the genes encoding the subunits of this enzyme have been found as significantly differentially expressed genes in cancer. The most significant differential expression was found in COX6B2 ($0.05 < p < 0.06$). This finding corroborates the possibility of using reactive oxygen species as potential biomarkers for pancreatic cancer.

Table 4.10. Flux differences for CYOOm3i and CYOOm2i: Sorted alphabetically, (*) indicates reactions with $DSD \geq 1$.

Reaction Symbol	**Adj. P. Val. (Benjamini-Hochberg)**	**Cancer Model Mean Flux (fmol/cell/h)**	**Healthy Model Mean Flux (fmol/cell/h)**	**Flux Diff. (Cancer – Healthy) (fmol/cell/h)**	**DSD**
CYOOm2i*	$< 10^{-300}$	20.6 ± 0.2	90.3 ± 0.3	-69.7	1.2
CYOOm3i*	$< 10^{-300}$	71.9 ± 0.3	0	71.9	1.8

The second salient finding was the massive overproduction of lactate in the cancer model compared to the healthy model. The lactate dehydrogenase (LDH) enzyme, combining its cytoplasmic (labeled "LDH_L" in the Recon3D model) and peroxisomal activity (labeled "r0173" in the Recon3D model), was found to produce lactate from pyruvate in both models. However, the mean production rate of lactate from pyruvate was found to be 251 ± 2 fmol/cell/h for the cancer model, compared to 32 ± 2 fmol/cell/h for the healthy model, corresponding to a nearly 8-fold amplification. The reversible reaction LDH catalyzes is

$$NAD^+ + lactate \leftrightarrow NADH + H^+ + pyruvate. \tag{4.4}$$

This overactivity of the glycolytic pathway, leading to more lactate fermentation, is a well-studied cancer behavior termed the Warburg effect, as discussed previously in [218–221]. The cancer model does indeed produce more ATP and other nucleosyl triphosphates (like GTP) overall from the glycolytic pathway compared to the healthy model (1953 ± 9 fmol/cell/h compared to 1901 ± 9 fmol/cell/h), but this is merely a 3% (or 1.03-fold) change.

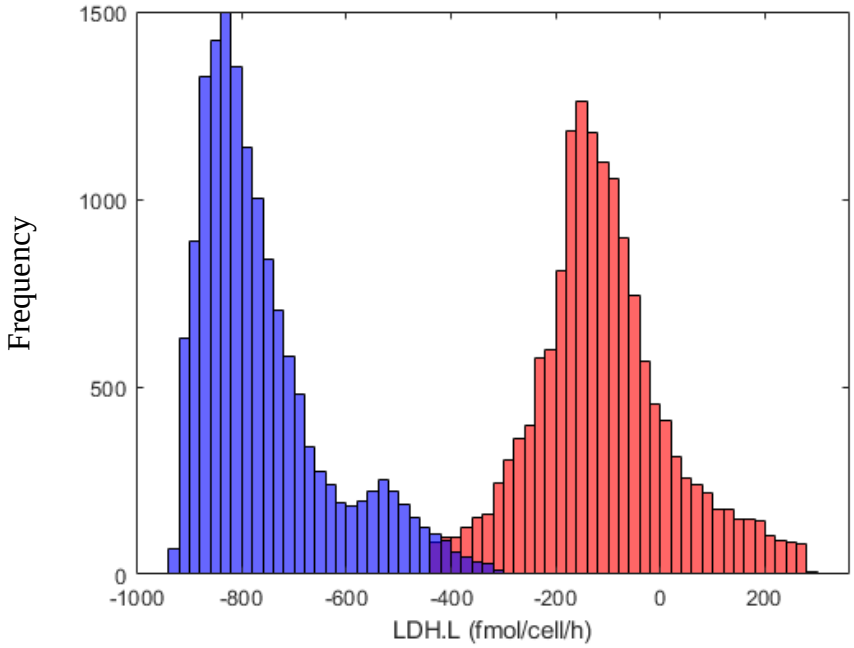

Figure 4.16. Flux sampling results for lactate dehydrogenase (cytoplasmic), blue: healthy, red: cancer.

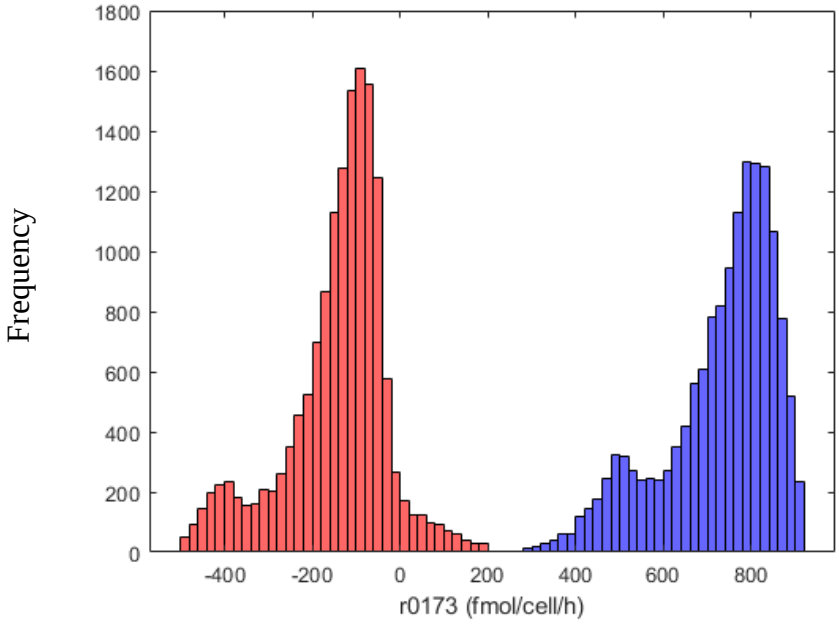

Figure 4.17. Flux sampling results for lactate dehydrogenase (peroxisomal), blue: healthy, red: cancer.

Figure 4.16 and Figure 4.17 shows the sampling results for these enzymes. Once again, no genes encoding for LDH subunits or variants were found to be significantly differentially expressed.

A third finding concerns the metabolite dihydroxyacetone phosphate (DHAP), which is ordinarily produced and consumed as a part of healthy glycolysis. DHAP is also the bridge between glycolysis and fatty acid metabolism, being the entry point of glycerol into the energy metabolism. The enzyme that converts phosphorylated glycerol into DHAP, glycerol-3-phosphate dehydrogenase (labeled "r0202" in the Recon3D model, encoded by gene GPD1), can act reversibly to produce glycerol from DHAP as well. The reaction it catalyzes is

$$NAD^+ + glycerol - 3 - phosphate \leftrightarrow NADH + H^+ + DHAP. \quad (4.5)$$

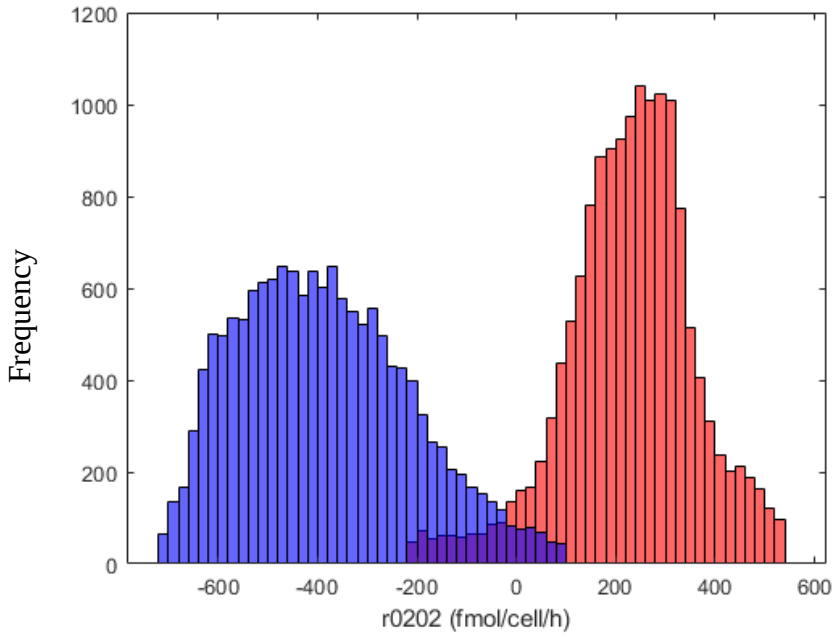

Figure 4.18. Flux sampling results for glycerol-3-phosphate dehydrogenase, blue: healthy, red: cancer.

In this study, it was found that the cancer model uses glycerol-3-phosphate dehydrogenase to *produce* DHAP from glycerol-3-phosphate at a rate of 225 ± 1 fmol/cell/h, while the healthy model *consumed* DHAP to produce more glycerol-3-phosphate at a rate of 385 ± 1 fmol/cell/h. This difference can be seen in Figure 4.18.

A study by Marin-Hernandez et al. [222] proposes DHAP accumulation to be part of the mechanism of action of some anticancer agents, since DHAP accumulation is known to be toxic through the inhibition of glucose uptake. A different approach to DHAP's relevance to cancer has been suggested by studies indicating that DHAP production acts as a metabolic activation signal to mTORC1, a signaling protein that promotes growth and proliferation [223, 224]. The mTOR family of genes and their associated proteins are known to play a role in pancreatic cancer through the action of the PI3K signaling pathway, as well as in the regulation of lipid metabolism [19–21]. Research has shown mTOR genes to be a potential avenue for therapeutic interventions in pancreatic cancer [225–227], and existing drugs such as Temsirolimus, Sirolimus and Zotarolimus already target the activity of mTOR proteins [73]. It is therefore possible that the investigation of DHAP metabolism can lead to advancements in the areas of both biomarker and drug discovery. Although considered "drug targets" and "cancer seed genes" in this study, no genes of the mTOR family have been found to be significantly differentially expressed. It is also likely that this variance in DHAP metabolism is related to the kynurenine pathway and how it affects G3P, as discussed in the previous subsection.

4.2.4. Overlap Between DEGs and Reactions with Significantly Different Flux

The genes whose reactions were found to have significantly different fluxes between the metabolic models and genes found to be significant DEGs contain 55 overlapping members. These genes are of interest because their differential expression and differential enzyme activity are both relatively high, making the detection of these genes even likelier in a real-life scenario for the diagnosis or prognosis of pancreatic cancer. Table 4.11 presents these genes, along with the number of reactions they are associated with which have significant reaction flux differences. Of these 55 "overlap genes," a selection of those with

adjusted p-values under 10^{-7}, $DSD \geq 1$, or log2-fold-changes over 5 (or under -5, corresponding to a 32-fold amplification or inhibition) are discussed in detail.

Table 4.11. Genes which are both DEGs and associated with significantly different flux, (*) indicates genes with at least one $DSD \geq 1$ reaction, red fold changes indicate increased cancer expression, blue fold changes indicate decreased cancer expression.

HGNC Gene Symbol	Average Expression (logCPM)	log2 Fold Change	Adj.P.Val (Benjamini-Hochberg)	Associated Sig. Rxns.
NQO1	4.081	3.223	$< 10^{-9}$	1
B3GNT3	3.424	4.277	$< 10^{-7}$	7
ALAS2	-2.800	-3.257	$< 10^{-4}$	2
PLCD3	3.882	1.999	$< 10^{-4}$	1
FADS3*	4.573	-1.349	$< 10^{-4}$	3
NAGK	5.364	-0.806	$< 10^{-4}$	1
PYGB	6.578	1.411	$< 10^{-4}$	2
ACSM5	-0.785	-2.191	< 0.001	2
SLC47A1	-0.592	-1.881	< 0.001	2
PLPP2	5.162	1.570	< 0.001	1
SLC2A1	4.642	2.172	< 0.001	4
CD36	4.057	-2.356	< 0.001	11
HMOX1	3.049	-1.805	< 0.001	1
MBOAT2	3.369	-1.235	< 0.001	1
LIPA*	5.330	1.605	< 0.001	7
TSTA3	6.251	1.333	< 0.001	2
FUT3	3.697	2.408	< 0.001	3
CERS4	5.132	-1.304	< 0.001	2
PIK3CD	3.405	-1.345	< 0.002	1
HCST	0.518	-1.578	< 0.002	1
GGTLC1*	-3.195	-2.763	< 0.002	1
PHOSPHO1	-1.743	-1.939	< 0.002	2

Table 4.11. Genes which are both DEGs and associated with significantly different flux, (*) indicates genes with at least one $DSD \geq 1$ reaction, red fold changes indicate increased cancer expression, blue fold changes indicate decreased cancer expression. (cont.)

HGNC Gene Symbol	Average Expression (logCPM)	log2 Fold Change	Adj.P.Val (Benjamini-Hochberg)	Associated Sig. Rxns.
SRD5A3	4.486	1.333	< 0.002	1
FIG4	3.626	-0.580	< 0.002	1
ALDH3B1	4.401	1.475	< 0.002	19
XDH	1.522	3.048	< 0.002	4
SLC35A2	4.787	0.845	< 0.002	2
UGT1A10	-2.805	7.777	< 0.002	6
SDR16C5*	0.302	5.097	< 0.002	4
SLC26A11	3.601	-0.856	< 0.002	1
CKMT1A	0.398	2.952	< 0.003	1
DPEP2	0.175	-1.685	< 0.004	1
RENBP	1.508	-1.357	< 0.005	1
B4GALT4	5.015	0.856	< 0.005	2
CKMT1B	1.108	2.667	< 0.006	1
GDA	2.772	2.543	< 0.006	1
GFPT1	6.699	1.005	< 0.006	1
PIK3R5	2.250	-1.640	< 0.006	1
AK4	2.134	2.225	< 0.007	2
KMO	-1.507	-1.789	< 0.007	1
PYCR1	5.553	1.199	< 0.007	4
SLC2A6	2.345	-1.217	< 0.007	1
GPX2	4.413	2.736	< 0.007	2
GM2A	5.018	-0.885	< 0.008	3
ME2	4.420	-0.729	< 0.008	2
GALNT3	5.443	1.072	< 0.008	1
PSAP	9.557	-0.723	< 0.008	2

Table 4.11. Genes which are both DEGs and associated with significantly different flux, (*) indicates genes with at least one $DSD \geq 1$ reaction, red fold changes indicate increased cancer expression, blue fold changes indicate decreased cancer expression. (cont.)

HGNC Gene Symbol	Average Expression (logCPM)	log2 Fold Change	Adj.P.Val (Benjamini-Hochberg)	Associated Sig. Rxns.
SLC9A5	0.628	-1.003	< 0.008	1
SLC52A3	1.871	1.600	< 0.008	1
ITPKB	4.384	-0.741	< 0.009	1
AQP5*	1.700	4.830	< 0.009	1
CNDP2	6.840	-0.679	< 0.009	2
GALNT5	-0.021	3.752	< 0.01	1
INPP5B*	4.152	-0.596	< 0.01	2
ACSL5	4.263	1.336	< 0.01	31

NQO1 encodes a quinone oxidoreductase enzyme whose main function is the relief of oxidative stress on ubiquinone (also known as coenzyme Q10) and other quinone molecules. Ubiquinone being a key coenzyme in oxidative phosphorylation, this genetic difference could be one of the main drivers of the CYOOm2i/CYOOm3i metabolic difference observed between healthy and cancerous metabolic models, discussed previously. It is found to be overexpressed in many cancer types [228], including pancreatic cancer [229]. Lewis et al. [230] have suggested inactivating NQO1 expression as a therapeutic approach to pancreatic cancer, due to the induction of greater oxidative stress in the cancer tissue.

B3GNT3 encodes a cell membrane signaling protein which is overexpressed in pancreatic cancer and lends it resilience against immune T cell infiltration [145]. These findings by Zhuang et al. were previously discussed in subsection 4.1.3. within the general context of differentially expressed cancer genes being annotated with immune response-related terms.

FADS3, encoding a type of the fatty acid desaturase enzyme, is involved in lipid interconversion, especially the production of eicosanoids like arachidonic acid, involved in inflammation [231]. It is unclear whether its apparent downregulation actually indicates more activity in this pathway through other proteins.

LIPA encodes the digestive enzyme lipase, known to be often elevated in pancreatic cancer. Lipase levels are neither specific nor selective enough to be considered a pancreatic cancer biomarker instead of CA19-9. A study by Stotz et al. [232] suggests that the ratio of blood lipase to blood amylase may be a good biomarker for pancreatic cancer, although the gene associated with amylase production (AMY1A) was not found to be significantly different in this present study.

UGT1A10 is the highest fold-change gene found in Table 4.11, encoding a variant of uridine diphosphate glucuronosyltransferase. These uridine diphosphate glucuronosyltransferases (UGTs) have a function of modifying hydrophobic drugs and xenobiotics for their subsequent metabolic processing [233]. Since these drugs include chemotherapy drugs, it is likely that this gene's overexpression shown in Table 4.11 is related to the medication used by the patients the samples were taken from. This may still be clinically relevant as polymorphisms in the protein of a related gene, UGT1A1, was found to be related to resistance to chemotherapeutic drug irinotecan in refractory pancreatic cancer patients [234].

SDR16C5 encodes Epidermal Retinol Dehydrogenase 2, which is involved in the oxidation of retinol into retinaldehyde (two forms of vitamin A). It was found to be significantly overexpressed in pancreatic cancer. Another study by Qi et al. [235] reported SDR16C5 to be underexpressed in triple negative breast cancer. Outside of SDR16C5, Bleul et al. [236] discovered a reduction in retinoic acid concentrations and retinoid receptors in pancreatic cancer patients.

AQP5 encodes a subunit of the protein aquaporin, which allows water transport in and out of cells. Its status as a DEG was mentioned in the previous subsection. AQP5 was found to be overexpressed in pancreatic ductal adenocarcinoma in a study by Direito et al. [237].

Another study by Woo et al [238] suggests, in murine cancer cell lines, that AQP5 activates the Ras pathway. This indicates that AQP5 may not only be an indicator or biomarker for pancreatic cancer, but a viable drug target as well.

INPP5B is a gene that encodes an inositol phosphate phosphatase. Inositol phosphates are generally related to cancer due to their relation to the phosphatidylinositol-3-kinase (PI3K) pathway that is an important signaling pathway implicated in many cancer types [196]. Battini et al.'s finding of lowered phosphatidylinositol phosphates in pancreatic cancer was previously discussed [34], although it is difficult to point out whether the downregulation of INPP5B is consistent with this finding. This study found that according to the reactions defined in Recon3D, INPP5B is more involved in interconversion of phosphatidylinositol phosphates than their production. A review by Ijuin contains a survey of inositol phosphate phosphatases, stating that these genes were found to be upregulated in some cancers, downregulated in some others, and mutated in others, due to having a complex mechanism of interacting with the cancer metabolic and signaling pathways [239]. While Ijuin's review does not make explicit mention of INPP5B, its closest relations in the review are INPP4A, which was found to be mutated in pancreatic ductal adenocarcinoma by Basturk et al. [240], and INPP5D, which was found to be upregulated in intraductal papillary mucinous neoplasms (IPMNs) but not in pancreatic cancer in a study by Hallas et al. [241].

4.2.5. Annotation & Enrichment Analyses of Significant Flux Differences

Using annotation and enrichment analyses on the total set of 1,960 metabolic reactions found to be significantly different between the two models, corresponding to 1,095 genes is uninformative, since the results are flooded with "low-level" (overly general) annotations such as "metabolic reaction" or the subsystems these reactions belong to, which are already known. Therefore, annotation and enrichment analyses were carried out on the smaller set of 55 genes where significance in terms of differential expression and in terms of metabolic flux differences overlapped. However, it was found that of these 55 genes, only SRD5A3 was a known drug target. Therefore, annotation with known drug targets was carried out on the entire set of 1,095 significant genes found through metabolic modelling, which were

found to contain between them 11 of the 83 drug targets from the TTD database. Table 4.12 summarizes these findings.

Table 4.12. Genes whose reactions show different flux between models, which are also drug targets: Genes marked *D* are downregulated in cancer, (*) indicates genes with at least one reaction having $DSD \geq 1$.

HGNC Gene Symbol	**Average Expression (logCPM)**	**log2 Fold Change**	**Adj.P.Val (Benjamini-Hochberg)**	**Drugs**	**Indications**
ABCB1[D]	2.215	-0.837	> 0.01	Biricodar	Ovarian cancer
AOX1[D]	5.923	-0.477	> 0.01	Isovanillin	Solid tumor
CYP17A1	-0.593	0.340	> 0.01	Abiraterone	Prostate cancer
CYP19A1[D]	-1.073	-0.272	> 0.01	Letrozole, Exemestane, Anastrozole, Fadrozole, Amino-glutehimide, Testolactone	Breast cancer
DHFR*	1.456	0.527	> 0.01	Leucovorin, Methotrexate, Trimethoprim, Amino-salicylic acid	Solid tumor
DPYD[D, *]	4.131	-0.130	> 0.01	Fluorouracil	Solid tumor
IMPDH1[D]	5.434	-0.136	> 0.01	Tyverb, Tiazofurin	Breast cancer
RRM2*	1.490	0.859	> 0.01	Gemcitabine	Solid tumor

Table 4.12. Genes whose reactions show different flux between models, which are also drug targets: Genes marked *D* are downregulated in cancer, (*) indicates genes with at least one reaction having $DSD \geq 1$. (cont.)

HGNC Gene Symbol	Average Expression (logCPM)	log2 Fold Change	Adj.P.Val (Benjamini-Hochberg)	Drugs	Indications
SRD5A3	4.486	1.333	< 0.002	Dutasteride, Finasteride, Polyestradiol phosphate	Prostate cancer
SRD5A1	3.295	0.638	> 0.01	Dutasteride, Finasteride, Polyestradiol phosphate	Prostate cancer
TXNRD1	5.358	0.288	> 0.01	Fotemustine	Solid tumor

Out of the potential drug targets shown in Table 4.12, only SRD5A3's upregulation has been corroborated by differential gene expression analysis. In addition to SRD5A3, RRM2, and DHFR can be considered to have potential as drug targets since they show clearly differing metabolic behavior between the two models, even though it is not certain they are upregulated in pancreatic cancer. The likely downregulation in cancer of the DPYD gene makes it a worse avenue of exploration as a drug target. SRD5A3, being a significant DEG identified in this study and a known drug target for prostate cancer, has been discussed previously in Section 4.1.3.

RRM2 encodes one of the subunits of ribonucleotide reductase, ordinarily involved in DNA repair in many tissues. Its overexpression was found to promote epithelial-to-mesenchymal transition and metastasis in prostate and breast cancers [242, 243]. Similarly, a study by Duxbury et al. [244] found that inhibition of RRM2 expression through small interacting RNA (siRNA) fragments resulted in lower metastatic behavior. Most research on RRM2's relationship with pancreatic cancer has been focused on pancreatic cancer cells'

resistance to the ubiquitous antitumor agent gemcitabine and RRM2's role in this resistance. Duxbury et al. [244] further discovered that siRNA inactivation of RRM2 sensitized pancreatic cancer cells to gemcitabine treatment. A study by Xia et al. [245] found the same relationship where RRM2 inactivation sensitizes pancreatic cancer cells to gemcitabine treatment, using the plant extract gambogic acid, usually found in herbal supplements. Studies by Tu et al. and Lewis et al. [246, 247] have investigated the mechanism by which RRM2 is activated in pancreatic cancer to confer gemcitabine resistance, finding the angiogenesis promoter vasohibin 2 (VASH2) and the regulatory enzyme sphingosine kinase-2 (SPHK2), respectively. The present study has found that SPHK2 shows no differential gene expression between the healthy and cancer conditions, however the Recon3D "SPHK21c" reaction, which corresponds to sphingosine kinase-2 activity shows significantly increased flux in cancer ($p < 0.01$) with a 1.2-fold change. Surprisingly, the present study finds VASH2 to be a significantly differentially expressed gene between the cancer and healthy phenotypes ($p < 0.01$), albeit exhibiting lower expression levels in cancer, being diminished nearly 3-fold.

DHFR, which encodes dihydrofolate reductase, is a target of the well-studied drug methotrexate [73, 191, 192, 248]. It was previously mentioned in subsection 4.2.2.1 as part of a discussion on the significance of folate metabolism in comparison with MTHFR. While it was not found in the drug target database, it must be mentioned that the Recon3D "MTHFD" reaction that corresponds to the MTHFR gene was also found to be significantly ($p < 0.01$) more active in cancer, showing a flux of 720 ± 2 fmol/cell/h in the cancer model against 396 ± 2 fmol/cell/h in the healthy model.

The set of 55 genes with differential expression and metabolic flux difference overlap (henceforth "overlap genes") have been annotated with the R clusterProfiler package, drawing upon the disease ontology and gene ontology databases, using a p-value threshold of 0.05, followed by a Benjamini-Hochberg false discovery rate threshold of 0.2 to determine significance.

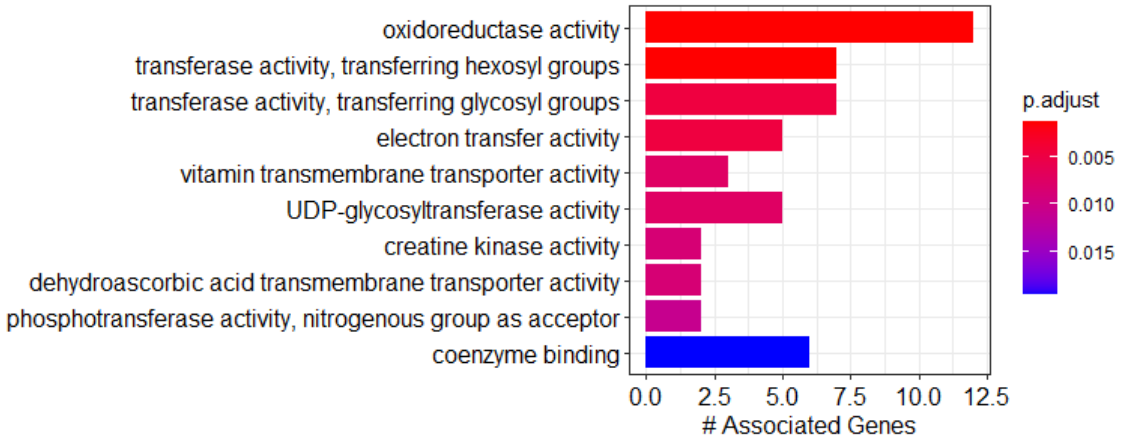

Figure 4.19. Gene Ontology: Molecular function annotation for overlap genes.

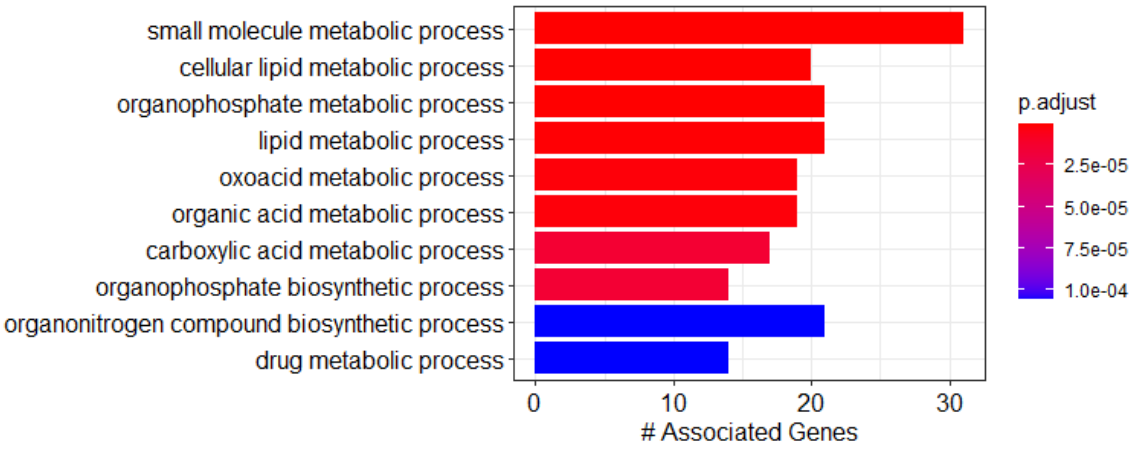

Figure 4.20. Gene Ontology: Biological process annotation for overlap genes.

All results of the annotation process that have met this significance threshold are represented below. Figure 4.19 and Figure 4.20 represent the 10 annotation terms most significantly associated with these “overlap genes” in the gene ontology molecular function and biological process corpuses, respectively. While many annotation terms revealed about the overlap genes by Figure 4.19 and Figure 4.20 are very broad, they still offer insights that uridine diphosphate (UDP) metabolism, as well as various vitamin metabolisms and lipid

metabolisms are highly represented among these overlap genes. Figure 4.21 contains the most significant annotations of the overlap genes with KEGG pathways.

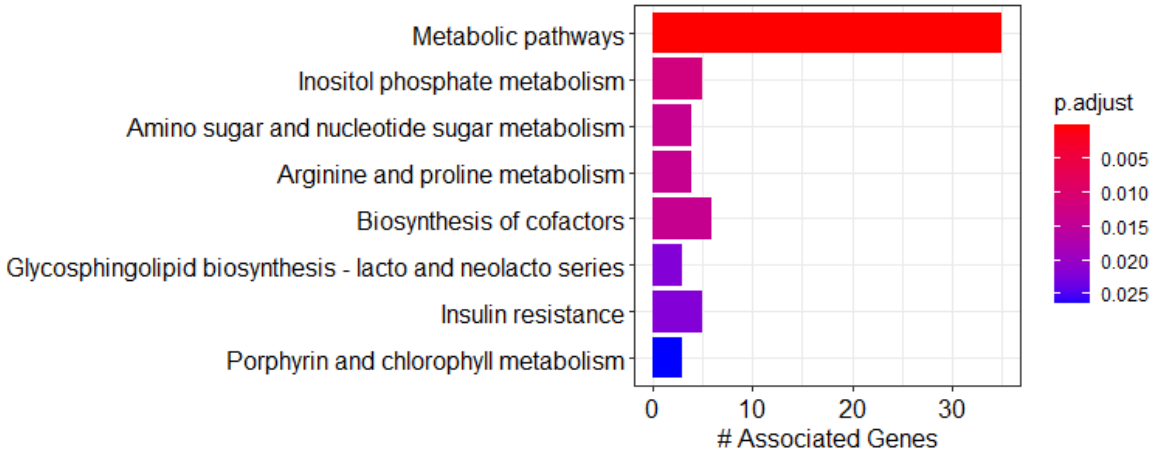

Figure 4.21. KEGG pathway annotation for overlap genes.

Figure 4.21 shows inositol phosphate and sphingolipid metabolisms are well-represented in the group of overlap genes, which is unsurprising due to these pathways being subject of much research [24, 35, 106, 196, 202], in addition to amino sugar, nucleotide sugar, arginine and proline metabolisms. Insulin resistance pathways are also apparently enhanced among this group of genes, which is promising since insulin is a measurable pancreatic secretion. Insulin secretion's regulation by inositol phosphates may be a contributing factor to this finding [196].

4.2.6. Synthetic Lethality

Pratapa et al.'s synthetic lethality approach [87] was applied to the healthy and cancer metabolic models created to discover single reactions or pairs of reactions whose removal would cause each model to "die" by having its biomass production rate drop below 1% of the optimal rate. Reactions whose removal by themselves was lethal to the cancer model without being lethal to the healthy model are presented in Table 4.13.

Table 4.13. Cancer single-lethal reactions, their associated subsystems and associated genes, (*) indicates the gene is a significant DEG, (**) indicates the gene has reactions with significant flux differences.

Reaction Recon3D Label	**Reaction Name**	**Reaction Subsystem**	**Associated Genes**
CDSm	Phosphatidate Cytidylyltransferase	Glycerophospholipid metabolism	CDS1* or CDS2
HMR_0586	CDP-Diacylglycerol:Phosphatidylglycerol 3-Phosphatidyltransferase	Glycerophospholipid metabolism	CRLS1**
CYTK1m	Cytidylate Kinase (CMP), Mitochondrial	Nucleotide interconversion	CMPK2
NDPK3m	Nucleoside-Diphosphate Kinase (ATP:CDP), Mitochondrial	Nucleotide interconversion	NME4** or NME6**
HMR_0587	Cardiolipin transport	Transport, mitochondrial	N/A

These lethal reactions focus on three areas: preventing the cancer metabolism from producing necessary glycerophospholipids and moving cardiolipin to the mitochondrion. In fact, excluding the transport reaction, all reactions of Table 4.13 can be summed together to indicate a cardiolipin-generating cycle that consumes ATP. Since transport reactions are not necessarily associated with genes in Recon3D, we can only guess why this transportation is necessary. Since all other reactions in Table 4.13 are mitochondrial reactions, a likely guess is that the gene expression profile of the cancer metabolic model prevents it from carrying out a necessary reaction in the cytoplasm, due to its tissue-specificity constraints. Then, shuttling metabolites between organelles can become necessary for the cancer model's survival. Pairs of reactions whose concurrent removal was lethal to the cancer model without being lethal to the healthy model are presented in Table 4.14.

Table 4.14. Cancer double-lethal reactions, colored in pairs, with their associated subsystems and associated genes, (*) indicates the gene is a significant DEG, (**) indicates the gene has reactions with significant flux differences.

Reaction Recon3D Label	**Reaction Name**	**Reaction Subsystem**	**Associated Genes**
r0276	NADH:Guanosine-5-Phosphate Oxidoreductase (Deaminating)	Purine catabolism	GMPR** or GMPR2**
r1384	Guanosine Aminohydrolase	Nucleotide interconversion	N/A
r0276	NADH:Guanosine-5-Phosphate Oxidoreductase (Deaminating)	Purine catabolism	GMPR** or GMPR2**
PUNP3	Purine-Nucleoside Phosphorylase (Guanosine)	Purine catabolism	PNP**
PA_HStm	Transport of Phosphatidate, Mitochondrial	Transport, mitochondrial	N/A
PCHOLPm_hs	Choline Phosphatase	Glycerophospholipid metabolism	PLD1**
PA_HStm	Transport of Phosphatidate, Mitochondrial	Transport, mitochondrial	N/A
PETOHMm_hs	Phosphatidylethanolamine N-Methyltransferase	Glycerophospholipid metabolism	PEMT
PPItm	Diphopshate Transporter, Mitochondrial	Transport, mitochondrial	N/A
PPAm	Inorganic Diphosphatase	Oxidative phosphorylation	PPA2

Table 4.14. Cancer double-lethal reactions, colored in pairs, with their associated subsystems and associated genes, (*) indicates the gene is a significant DEG, (**) indicates the gene has reactions with significant flux differences. (cont.)

Reaction Recon3D Label	**Reaction Name**	**Reaction Subsystem**	**Associated Genes**
PGLYCtm	Transport of Phosphatidylglycerol, Endoplasmatic Reticulum	Transport, mitochondrial	N/A
PGPP_hsc	Transport of Phosphatidyl Glycerol Phosphate, Mitochondrial	Glycerophospholipid metabolism	N/A

Tables 4.13 and 4.14 reveal most significantly that gene overexpression is a poor indicator of synthetic lethality potential, only CDS1 being found to be a significant DEG ($p < 0.01$). Table 4.13 further reaffirms the importance of the glycerophospholipid metabolism and use of alternate nucleotide triphosphates as energy sources to the cancer metabolism. Table 4.14, similarly to prior findings reported in this study, additionally shows that glycerophospholipid metabolism, especially at the point of phosphatidylinositol production, may be a relevant target.

4.3. Protein-protein Interaction (PPI) Networks

Discoveries through differential expression analysis and metabolic modeling give some insight into where differences exist between the cancerous and healthy phenotypes. However, it is likely that genes or proteins whose expression or activity are different are inaccessible to desirable diagnostic or therapeutic methods. Creating PPI networks allows the discovery of genes or proteins that are directly or indirectly able to affect the genes or proteins discovered earlier.

4.3.1. Network Topological Properties

The topological properties of healthy and cancer PPI networks were analyzed with Cytoscape's NetworkAnalyzer plugin. Table 4.15 contains a breakdown of overall network topological properties for the two networks.

Table 4.15. Overall PPI network topological properties.

Network Property	Healthy Network	Cancer Network
Nodes	11,195	11,261
Edges	208,438	209,708
Avg. Degree	37.24	37.25
Avg. Shortest Path Length	2.852	2.852
Avg. Clustering Coefficient	0.131	0.131
Avg. Betweenness Centrality	0.000165	0.000165

The cancer and healthy networks, differing only by the seed genes used in their construction, show remarkably little difference in terms of the degree distribution.

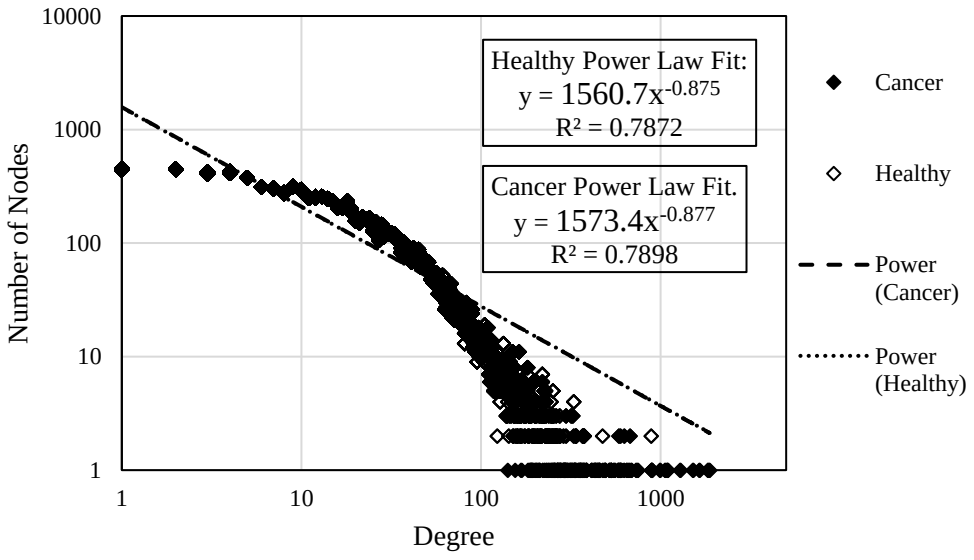

Figure 4.22. Power law fits to degree distributions in PPI networks.

It is known that PPI networks tend to be scale-free, and therefore their logarithmic degree distributions can be expressed as a power law. The present study has also found this to be true as in Figure 4.22, with very similar power law best-fit lines. Plotting each gene's topological properties against each other further reinforces that the two networks are nearly identical.

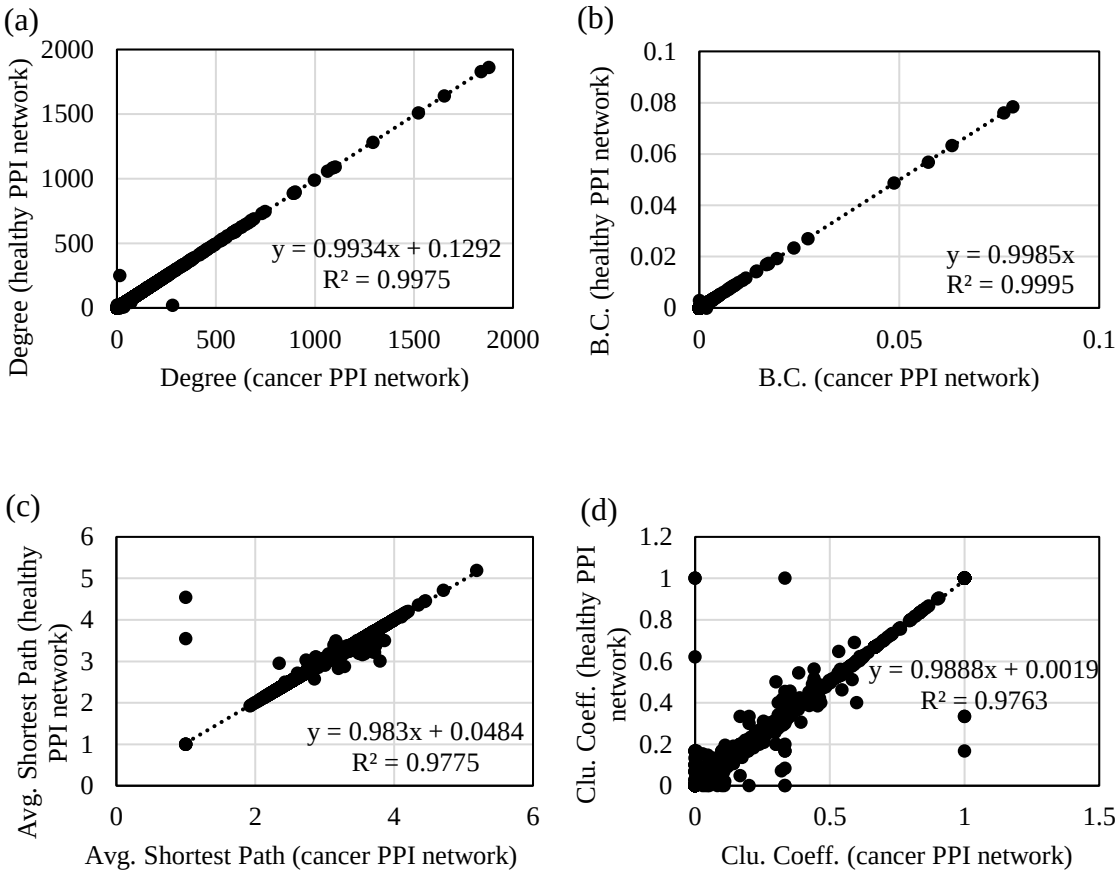

Figure 4.23. Topological properties of cancer and healthy PPI networks, by (a) degree, (b) betweenness centrality, (c) average shortest path length, and (d) clustering coefficient.

Figure 4.23 reveals most significantly that the cancer and healthy PPI networks show similarities beyond their average properties, and even individual nodes show remarkable similarity between their topological properties between networks. This is an indication that the network generation method that differentiates cancer and healthy PPI networks, algorithmic matching of GO terms to a set of seed genes (illustrated in Figure 3.6), does not

lead to a strong dependence of the network on the initial set of seed genes. Therefore, further analysis regarding PPI networks will only take into account the PPI cancer network, which is slightly larger. While the two networks differ very little in terms of key average topological properties, average topological properties belonging to subsets of nodes in the PPI networks show some small differences. These values are provided below, in Table 4.16.

Table 4.16. PPI network topological properties for subgroups of interest.

Network Property	**Significant DEGs in Network**	**Genes with Sig. Diff. Metabolic Reactions in Network**	**Cancer Seed Genes in Network**	**Healthy Seed Genes in Network**
Avg. Degree	36.40	21.10	37.69	37.28
Avg. Shortest Path Length	2.86	2.93	2.85	2.85
Avg. Clustering Coefficient	0.129	0.106	0.132	0.131
Avg. Betweenness Centrality	0.000159	0.000058	0.000168	0.000166

The meanings of these network topological properties are detailed in Table 2.1. For all intents and purposes, these properties can be treated as “centrality measures” for any given node, which indicate how well-connected a given node is to the rest of the graph. In the case of each topological property, except for shortest path length, a higher value indicates a better-connected node. Table 4.16 shows that in terms of all centrality measures, DEGs exhibit less central behavior compared to the complete network, and genes belonging to significantly different metabolic reactions exhibit much less central behavior compared to DEGs or the complete network. In every metric measured in this study, these disease-related genes have been found to be less well-connected and having fewer interaction partners. Previous approaches by researchers such as Xu et al. [249] have claimed that disease genes tend to be more central, and Barabasi et al. have echoed that this is indeed true for known disease genes. [6] However, recent research is more in line with this present work’s finding that while

known disease genes tend to be central, centrality of genes in terms of PPI is not a good predictor of disease [47, 50]. This can be seen in the fact that seed gene subgroups have, on average, higher or equivalent centrality to the total network, regardless of whether the seed genes are cancer genes or pancreatic secretion genes. It can be assumed that, as previously discussed, there is an inherent bias in the discovery of both "seed genes" (i.e., genes broadly accepted to be related to a certain phenotype) and of PPI relations. This bias skews towards finding more connections and relations belonging to already well-studied genes.

4.3.2. Topological Properties of Individual Genes

Prior to analysis of how genes fall into functional clusters within the PPI network, it may be desirable to look at the topological properties of genes as individual nodes within such a network. Here, the topological properties of genes found to indicate significant differences between the cancer and healthy phenotypes previously are compared to the top 10 outliers in terms of each topological property, which are termed "hubs," due to their superlative centrality within these networks.

4.3.2.1. Hub Genes. Sets of top 10 hub genes were determined according to each centrality measure indicated in Table 4.15 and Table 4.16. Since biological networks tend to be scale-free, there was found to be great overlap between genes with the highest degree and those with the highest betweenness centrality. In other words, any shortest path between two arbitrary nodes was found to likely to pass through one of the massively central, highest-degree nodes. To find genes that had more of a bridging role between densely connected areas of the network without being themselves closely connected to many unrelated genes, an additional centrality measure, the ratio of betweenness centrality to degree was calculated for each node. The same problem is identified, and an alternative solution is proposed, in a work on psychiatric disease networks by Johnson et al. [250]. These genes are henceforth called "bridge genes" or "bridge nodes" in this study. Additionally, when discovering the top 10 hub genes in terms of clustering coefficient, those genes with clustering coefficient exactly equal to 1 were discarded. A clustering coefficient of 1 indicates a node whose every neighbor are also neighbors with each other. Theoretically, this can indicate tightly connected subnetworks, but it was found that most nodes with a clustering coefficient of 1

existed in very small subnetworks (with less than 5 nodes) that were all interconnected, but poorly connected to the rest of the network. Table 4.17 contains the topological properties of the 20 genes with highest degree, betweenness centrality, clustering coefficient, and betweenness centrality/degree ratio, as well as those with lowest average shortest path length. These genes have been sorted into groups and color coded in Table 4.17: high centrality in all except for clustering coefficient – green; high centrality in all except for clustering coefficient and B.C./degree ratio – yellow; high degree, low average shortest path, high B.C., or a combination – grey; only high clustering coefficient – orange; only high B.C./degree ratio – blue. Furthermore, genes marked *SG* are cancer seed genes, genes marked *DEG* were discovered to be significant DEGs, genes marked *MET* were discovered to have significant metabolic flux differences. Please note that the list contains far fewer genes than 50, since there is much overlap between these hub genes.

Table 4.17. Topological properties for the top 10 PPI hub genes considering each centrality measure shown.

HGNC Gene Symbol	**Degree**	**Avg. Shortest Path**	**Clustering Coefficient**	**Betweenness Centrality**	**B.C./Degree Ratio**
TRIM25	1878	1.93	0.010	0.076	4.1×10^{-5}
APP	1653	1.97	0.007	0.078	4.7×10^{-5}
ELAVL1	1523	1.99	0.009	0.063	4.2×10^{-5}
HNRNPL	1293	2.01	0.008	0.057	4.4×10^{-5}
NTRK1	1840	1.94	0.016	0.049	2.7×10^{-5}
XPO1	1102	2.08	0.017	0.024	2.2×10^{-5}
EGFR[SG]	1090	2.04	0.019	0.027	2.5×10^{-5}
CUL3	1063	2.05	0.032	0.017	1.6×10^{-5}
NXF1[DEG]	998	2.12	0.012	0.019	2.0×10^{-5}
UBC	900	2.07	0.030	0.017	1.9×10^{-5}
MOV10	898	2.15	0.012	0.017	1.9×10^{-5}
TP53[SG]	892	2.07	0.039	0.014	1.6×10^{-5}

Table 4.17. Topological properties for the top 10 PPI hub genes considering each centrality measure shown. (cont.)

HGNC Gene Symbol	Degree	Avg. Shortest Path	Clustering Coefficient	Betweenness Centrality	B.C./Degree Ratio
CHD5	14	3.07	0.857	1.7×10^{-7}	1.2×10^{-8}
COMMD5	13	3.13	0.846	3.7×10^{-7}	2.9×10^{-8}
POLR3K	10	3.26	0.844	1.6×10^{-7}	1.6×10^{-8}
KCNK15	7	3.00	0.905	5.3×10^{-9}	7.6×10^{-10}
CDH3	6	3.00	0.867	1.3×10^{-8}	2.2×10^{-9}
ZNF79	5	3.09	0.900	7.5×10^{-8}	1.5×10^{-8}
GSAP	4	2.92	0.833	3.7×10^{-6}	9.2×10^{-7}
$PTGIS^{MET}$	4	2.92	0.833	3.2×10^{-6}	7.9×10^{-7}
PEX12	4	3.29	0.833	3.2×10^{-7}	7.9×10^{-8}
MFNG	4	3.25	0.833	1.5×10^{-8}	3.8×10^{-9}
EFCAB6	4	3.11	0.833	1.2×10^{-8}	3.0×10^{-9}
TYROBP	4	3.45	0.000	1.8×10^{-4}	4.5×10^{-5}
EIF4E3	4	2.97	0.000	1.8×10^{-4}	4.4×10^{-5}
SHLD2	3	2.88	0.333	1.8×10^{-4}	5.9×10^{-5}
$IL10RA^{DEG}$	3	3.19	0.000	3.6×10^{-4}	1.2×10^{-4}
ERCC6L2	2	3.71	0.000	1.8×10^{-4}	8.9×10^{-5}
IL10RB	2	4.19	0.000	1.8×10^{-4}	8.9×10^{-5}

Table 4.17 shows significant overlap between degree, average shortest path length and betweenness centrality, and there is only partial overlap between these other centrality measures and betweenness-centrality-to-degree ratio. Considering this finding, further analysis of gene centrality in PPI networks will only take into account betweenness centrality. This reduces the number of parameters required to characterize each gene, and therefore the dimensionality of the data. This excludes genes with high betweenness centrality-to-degree-ratio only partially, and genes with high clustering coefficient almost completely. This was deemed an acceptable trade-off since these parameters do not find

globally central genes as well as betweenness centrality alone. Another discovery in Table 4.17 is that there seems to be almost no overlap between clustering coefficient and the other centrality measures. DEGs and genes with significant reaction flux differences are represented only sporadically among the top hub genes, implying that PPI topological properties are by themselves not a good indicator of pancreatic cancer significance.

4.3.2.2. DEGs and Genes whose Reactions Show Flux Differences. In this section, genes already discovered to be significant either as DEGs or as genes with reaction flux differences are ranked based on their betweenness centrality and clustering coefficient, reporting the 10 genes with the highest betweenness centrality among each group. These most central DEGs are reported in Table 4.18.

Table 4.18. Topological properties for the top 10 PPI hub genes among DEGs, genes marked *SG* are cancer seed genes, genes marked *MET* were discovered to have significant metabolic flux differences.

HGNC Gene Symbol	**Average Expression (logCPM)**	**log2 Fold Change**	**Adj.P.Val (Benjamini-Hochberg)**	**Betweenness Centrality**
NXF1	6.538	-0.585	< 0.007	0.019453
LMNA	7.664	0.919	< 0.006	0.014371
CDH1	7.975	1.029	< 0.006	0.008818
GRB2	5.565	-0.517	< 0.01	0.007254
ARRB2	4.424	-1.113	< 0.002	0.002603
LGALS3	6.536	1.497	< 0.001	0.002352
ERBB2[SG]	6.860	1.087	< 0.004	0.001307
TRAF1	3.767	-1.265	< 0.001	0.001248
FOXA1	0.607	3.095	< 0.007	0.001152
MAPK3[SG]	5.284	0.785	< 0.002	0.001103

Table 4.19 shows the top 10 highest betweenness centrality genes in the PPI network that also exhibit significant flux differences.

Table 4.19. Topological properties for the top 10 PPI hub genes among genes with significant metabolic differences, genes marked *SG* are cancer seed genes, (*) indicates genes with at least one reaction having $DSD \geq 1$.

HGNC Gene Symbol	Average Expression (logCPM)	log2 Fold Change	Adj.P.Val (Benjamini-Hochberg)	Betweenness Centrality
SHMT2*	6.179	0.107	> 0.01	0.003657
PTEN*	6.347	-0.287	> 0.01	0.001452
ENO1	8.304	0.100	> 0.01	0.001363
PIK3R1^{SG}	5.076	-0.366	> 0.01	0.001267
TK1	3.007	0.913	> 0.01	0.001229
CTBP1	7.014	-0.180	> 0.01	0.001167
CFTR	7.796	-0.018	> 0.01	0.001111
BSG*	8.040	0.468	> 0.01	0.001017
WWOX	2.706	0.058	> 0.01	0.000988
GAPDH	8.749	0.842	> 0.01	0.000859

4.3.3. PPI Network Clustering and Annotation

Each gene in the cancer and healthy PPI networks were assigned into clusters using the flow-based MCL algorithm in the clusterMaker2 package for Cytoscape. With an inflation parameter of 2.2, the cluster size distribution shows power law-like behavior, with the first 3 clusters having over a hundred genes each, whereas the majority of clusters have 2 genes.

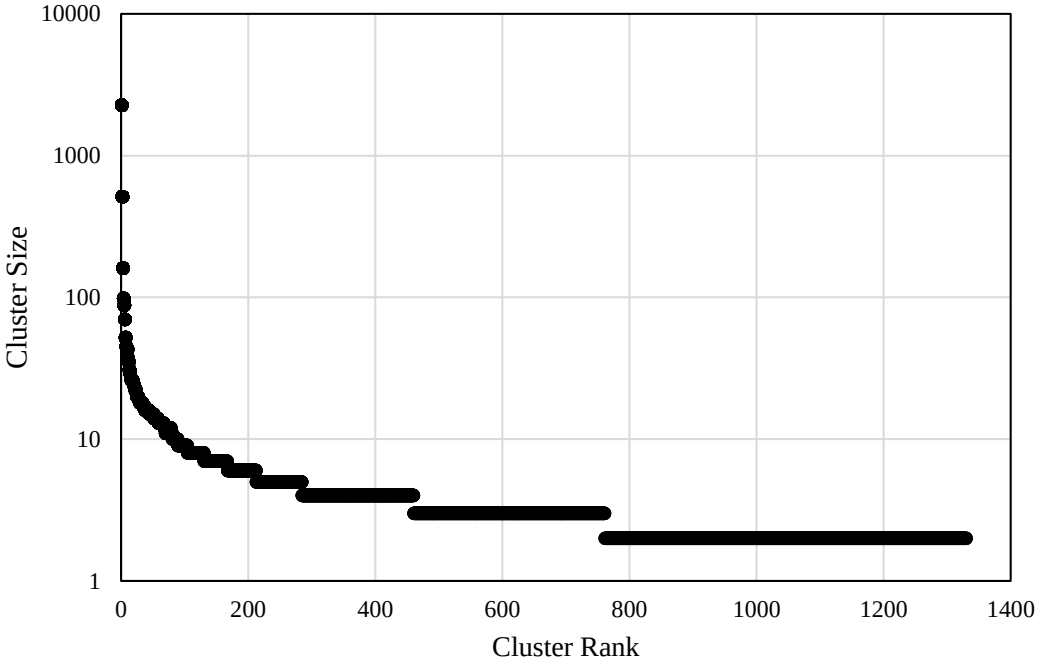

Figure 4.24. Cluster size distribution of the PPI network.

How genes previously discovered to be significant are distributed across PPI clusters can also be of interest. It was observed that that DEGs or genes with significant reaction flux differences discovered in sections 4.1. and 4.2. are most concentrated in smaller clusters.

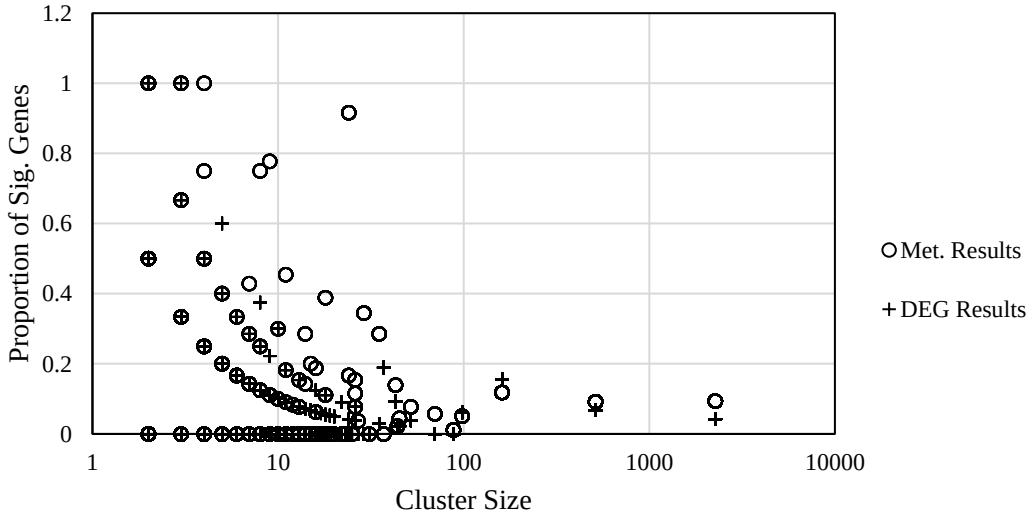

Figure 4.25. Distribution of significant genes across PPI clusters.

Figure 4.25 primarily shows that it is much more likely to find a high concentration of significant genes in very small PPI clusters. However, these clusters are unlikely to be very informative. For the range of clusters with 10-100 genes, where relevant PPI information can be obtained, DEGs seem to be much less concentrated than genes with significant metabolic differences. The 5 PPI clusters found to contain the greatest proportion of DEGs and genes with metabolic differences in this study are shown in Table 4.20. Clusters are named by the rank of their size among all clusters. The results of such PPI clustering analyses reveal, importantly, how certain groups of genes are "closely connected" to each other in terms of protein interaction, which must be kept in mind as results are interpreted.

Table 4.20. PPI clusters with the greatest proportion of DEGs or genes with metabolic differences, color coded: proportionally enriched in metabolically different genes – orange; proportionally enriched in DEGs – green.

PPI Cluster Rank/Name	Cluster Size	# of DEGs in Cluster	# of Met. Diff. Genes in Cluster	% of DEGs in Cluster	% of Met. Diff. Genes in Cluster
14	29	0	10	0	34%
20	24	0	22	0	92%
31	18	1	7	6%	39%
70	11	0	5	0	45%
85	10	0	3	0	30%
3	161	25	19	16%	12%
11	37	7	0	19%	0
73	11	2	0	19%	0
84	10	3	0	30%	0
89	10	3	0	30%	0

These clusters were subsequently annotated using the R clusterProfiler package, drawing upon the disease ontology and gene ontology databases, using a p-value threshold of 0.05, followed by a Benjamini-Hochberg false discovery rate threshold of 0.2 to determine

significance. All results of the annotation process that have met this significance threshold are represented in the rest of this section.

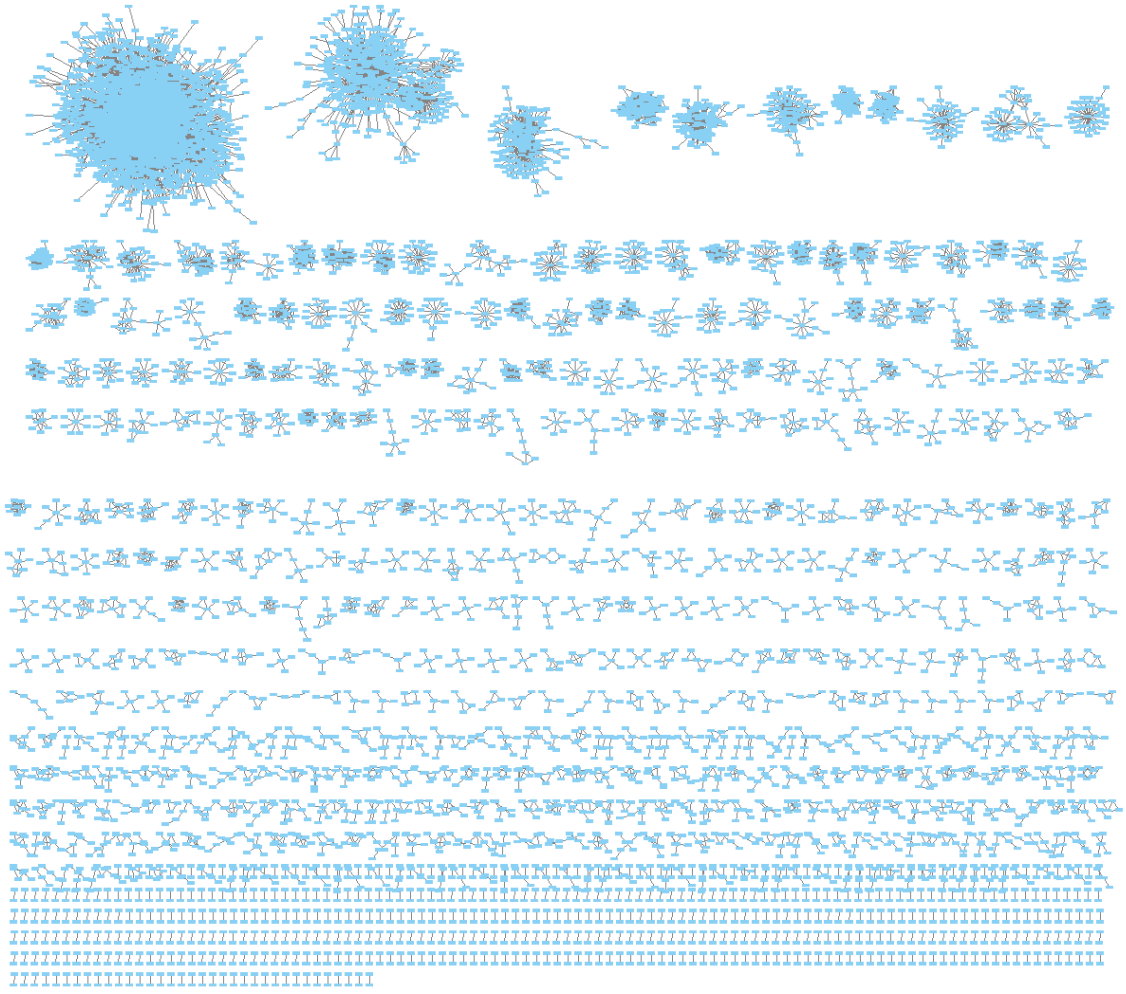

Figure 4.26. Pictorial representation of the 1330 clusters belonging to the cancer PPI network.

4.3.3.1. Enrichment of PPI Clusters with a High Proportion of Metabolically Significant Genes. This subsection contains the illustrations of the enrichment analyses on PPI networks, starting with a relatively large cluster, cluster 14.

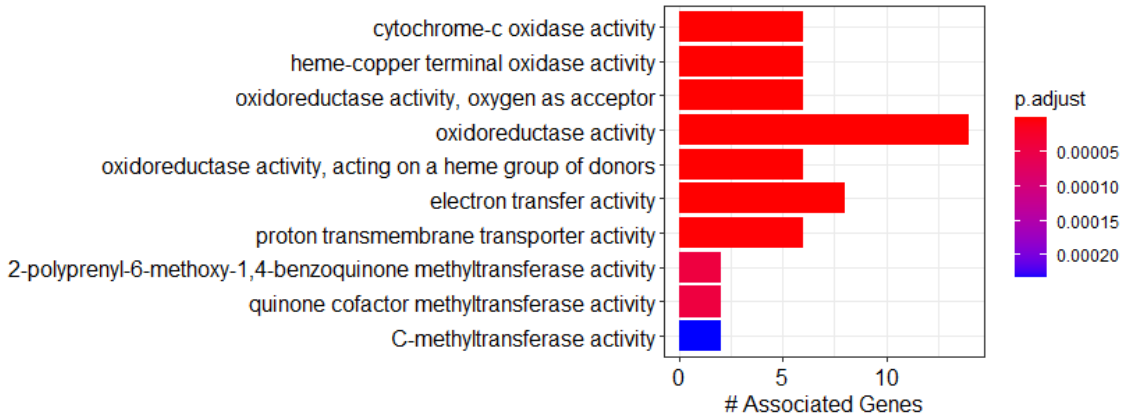

Figure 4.27. Gene Ontology: Molecular function annotation for PPI cluster 14.

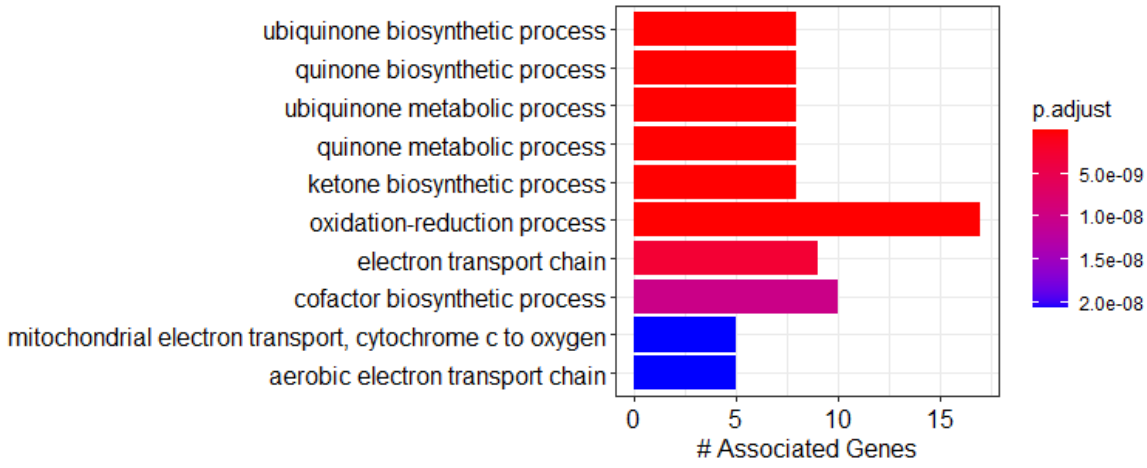

Figure 4.28. Gene Ontology: Biological process annotation for PPI cluster 14.

Figure 4.27 and Figure 4.28 illustrate the significant GO molecular function and biological process terms, respectively, associated with PPI cluster 14. This cluster is shown as being relevant to the oxidative phosphorylation pathways used in aerobic respiration.

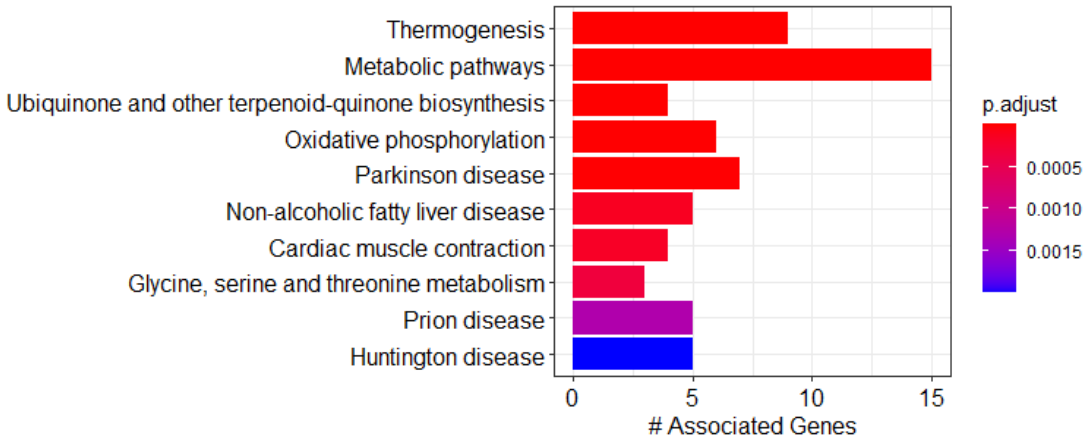

Figure 4.29. KEGG pathway annotation for PPI cluster 14.

The KEGG annotation in Figure 4.29 seems to corroborate the previous figures' results. In addition to these annotation results, cluster 14 also contains the cytochrome C oxidase subunit-encoding genes COX6A1, COX6B1, COX6C, and COX7C.

Cluster 20 is similarly mostly associated with oxidative phosphorylation and is annotated using the DOSE disease ontology database to be significantly associated with "mitochondrial metabolism disease" ($p < 0.05$). Over 90% of the genes that are members of cluster 20 are significant due to metabolic flux differences.

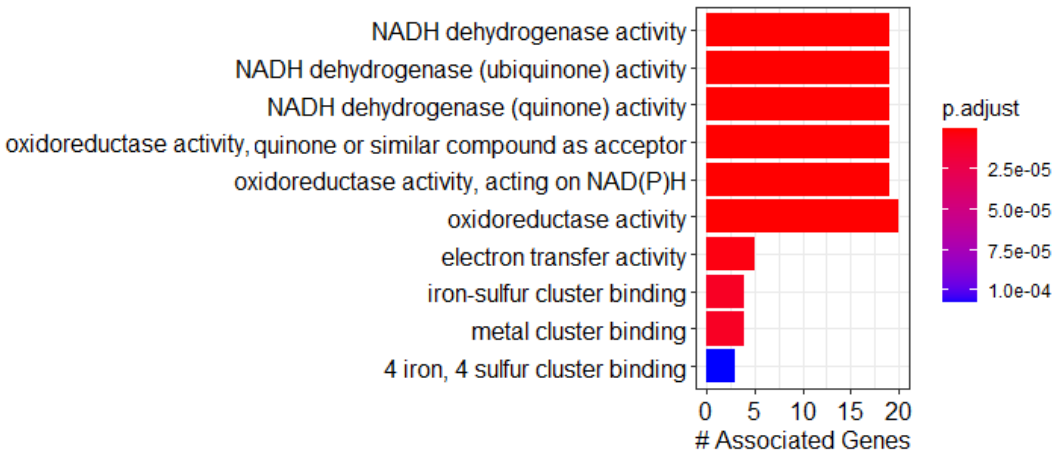

Figure 4.30. Gene Ontology: Molecular function annotation for PPI cluster 20.

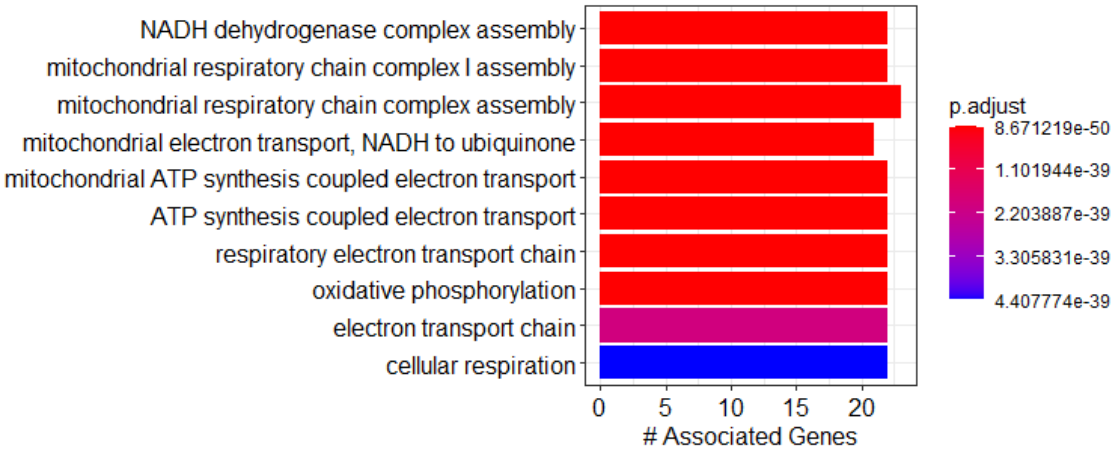

Figure 4.31. Gene Ontology: Biological process annotation for PPI cluster 20.

Figures 4.30 and 4.31 contain the GO molecular function and biological process annotations, respectively, for PPI cluster 20. Figure 4.30 and Figure 4.31 present cluster 20 to be highly associated with oxidative phosphorylation, with Figure 4.30 especially shedding light on the findings in subsection 4.2.3.1., regarding the varied uptake rates of various sulfur-containing ions and iron. It is likely upon viewing cluster 20's enrichment results, that these metabolites also indicated a dysregulation of the cancer model's oxidative phosphorylation subsystem

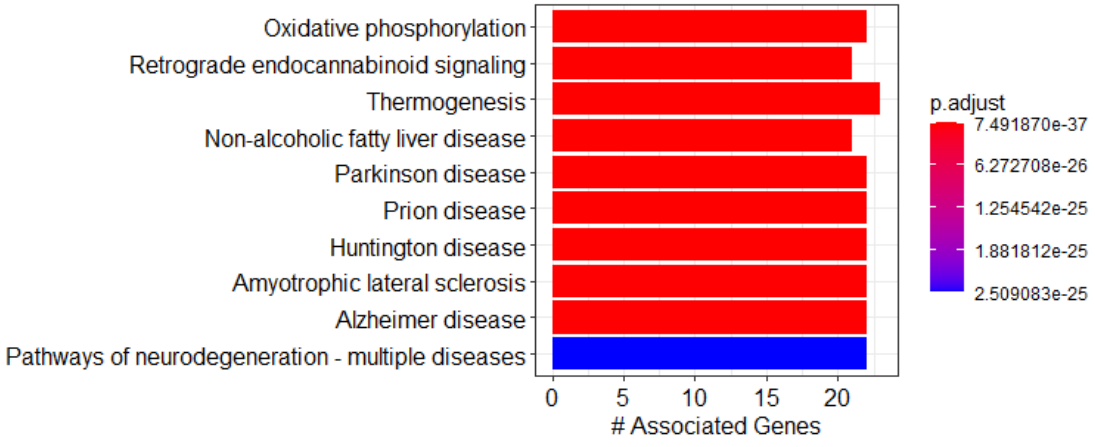

Figure 4.32. KEGG pathway annotation for PPI cluster 20.

PPI cluster 20 also includes UQCR10, encoding ubiquinol-cytochrome C reductase. This gene was found to contribute to reactions with metabolic flux differences with $DSD \geq 1$ in the oxidative phosphorylation subsystem of the metabolic models.

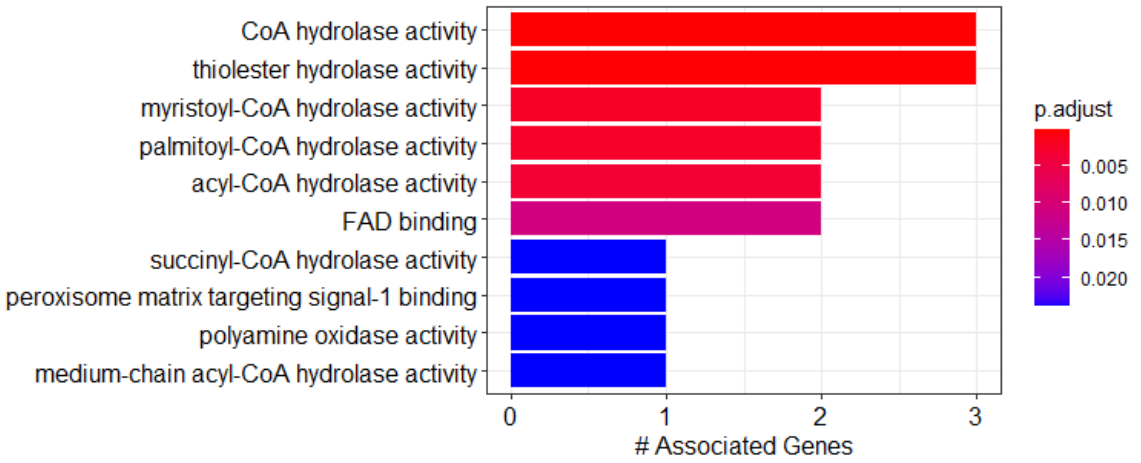

Figure 4.33. Gene Ontology: Molecular function annotation for PPI cluster 31.

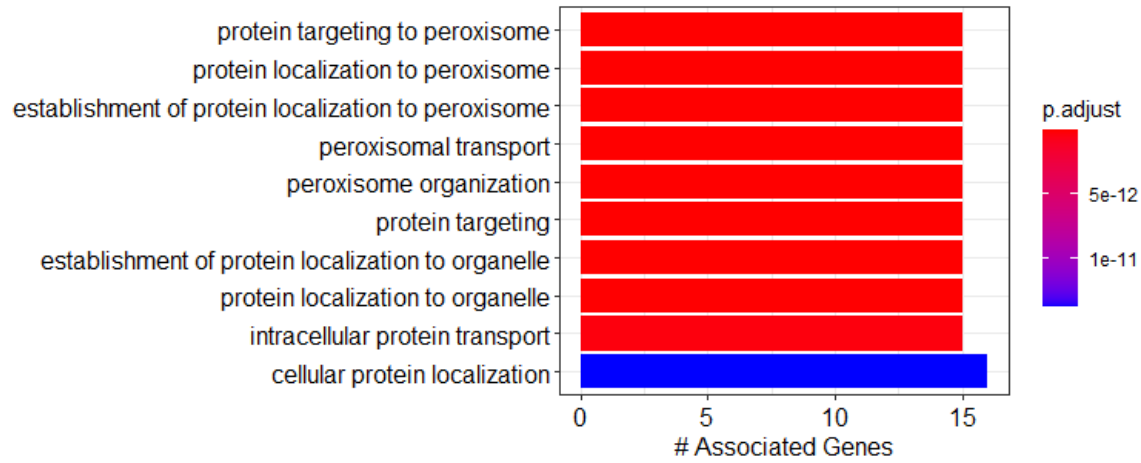

Figure 4.34. Gene Ontology: Biological process annotation for PPI cluster 31.

Figure 4.33 and Figure 4.34 contain the GO molecular function and biological process annotations, respectively, for PPI cluster 31. These Gene Ontology annotations place peroxisomal dysregulation, likely due to bile acid metabolism, in the spotlight due to their

strong association with genes whose reactions have been perturbed in the cancer metabolic model.

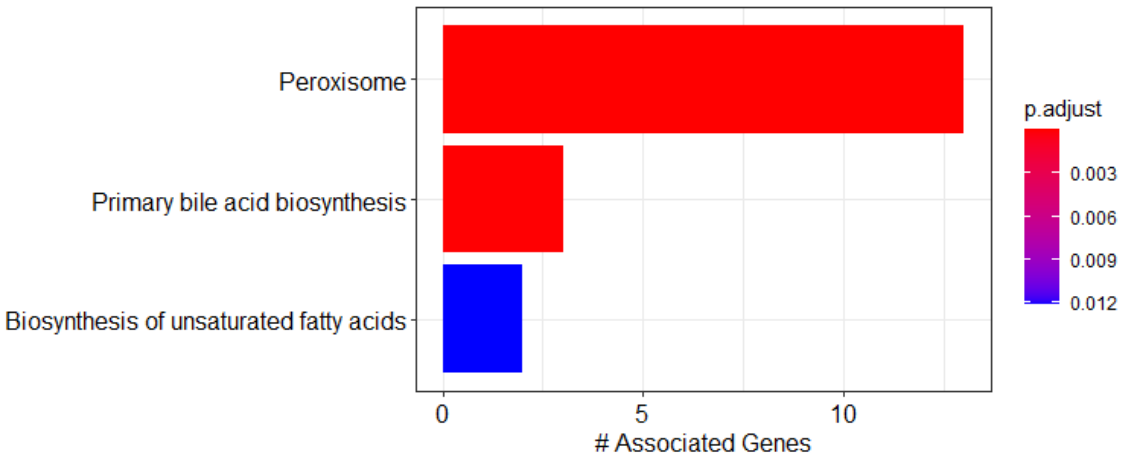

Figure 4.35. KEGG pathway annotation for PPI cluster 31.

PPI cluster 31 also includes BAAT, encoding bile acid acyltransferase and CRAT, encoding carnitine acyltransferase. These genes were found to contribute to reactions with metabolic flux differences with $DSD \geq 1$. This finding links the dysregulation in the cancer model's carnitine and bile acid metabolisms together.

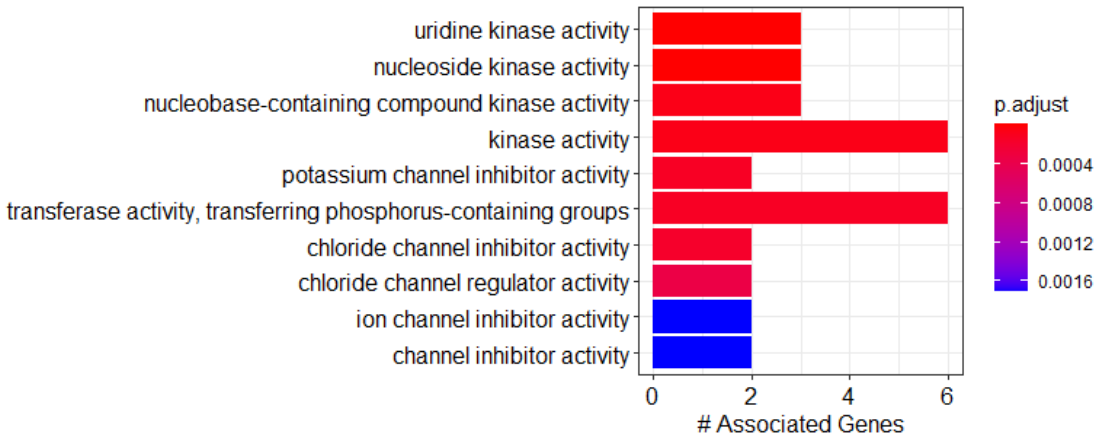

Figure 4.36. Gene Ontology: Molecular function annotation for PPI cluster 70.

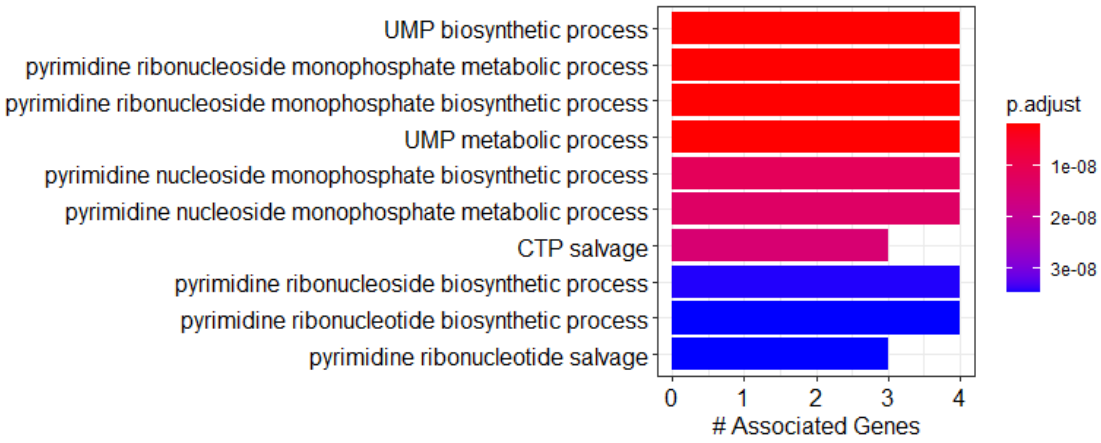

Figure 4.37. Gene Ontology: Biological process annotation for PPI cluster 70.

PPI cluster 70 is enriched in uridine kinase metabolism among molecular function and biological process GO terms, as Figure 4.36 and Figure 4.37 reveal. It does not contain any significant DEGs, but contains UCK1, UCK2, and UCK1L as genes whose associated reactions have high reaction flux differences with $DSD \geq 1$. These genes encode uridine cytidine kinases, which phosphorylate pyrimidine nucleosides uridine and cytidine into pyrimidine nucleotides UMP and CMP. The continuous phosphorylation of these nucleosides is an important recycle loop that allows cells to keep producing DNA and RNA necessary for replication. If this process is interrupted due to nucleotide insufficiency, the cell is driven to programmed cell death, or apoptosis, usually by the activation of the tumor suppressor gene TP53. Malami and Abdul's review explain all of the above, subsequently linking an overexpression of UCK1L and UCK2 to the promotion of cancerous growth, due to lack of apoptosis [251]. Such upregulation is found to be a putative cancer biomarker for lung, breast and hepatocellular cancer in several studies [252–254]. An in silico study by El Hassouni et al. [255] found the same UCK2 overexpression in pancreatic cancer. Although this study showed that these genes were associated with high levels of reaction flux differences between cancer and healthy metabolic models, a significant gene expression difference was not observed for any of the aforementioned genes ($p > 0.01$). It is worth mentioning that UCK2 was found to have a Benjamini-Hochberg adjusted p-value below 0.03 in terms of its quality as a DEG in this study as well and would have been considered significantly overexpressed with a more relaxed experimental methodology. The KEGG

pathway enrichment results for cluster 70 confirm the strong uridine kinase associations of this cluster.

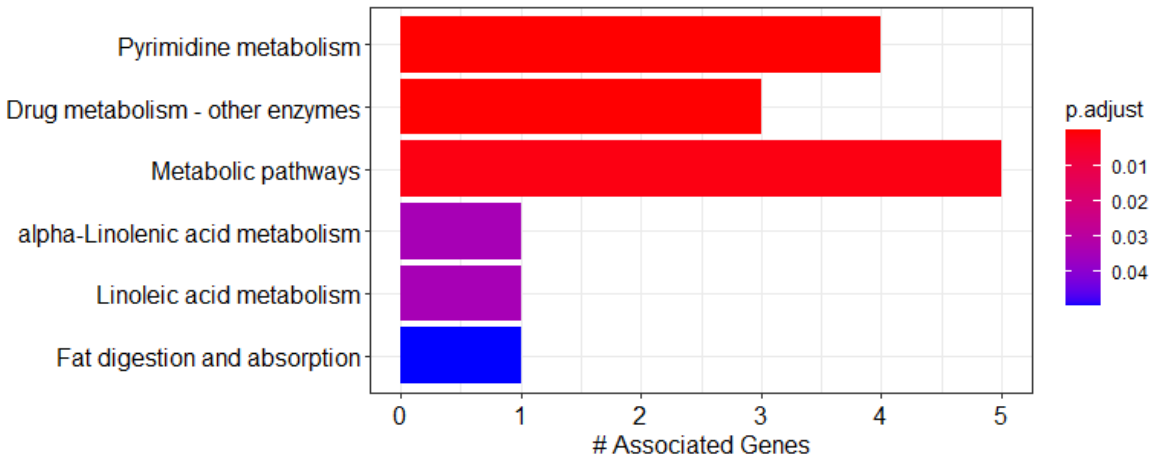

Figure 4.38. KEGG pathway annotation for PPI cluster 70.

PPI cluster 85 also revealed significant results in the enrichment analysis for GO molecular function terms, once again implicating uridine kinase metabolism.

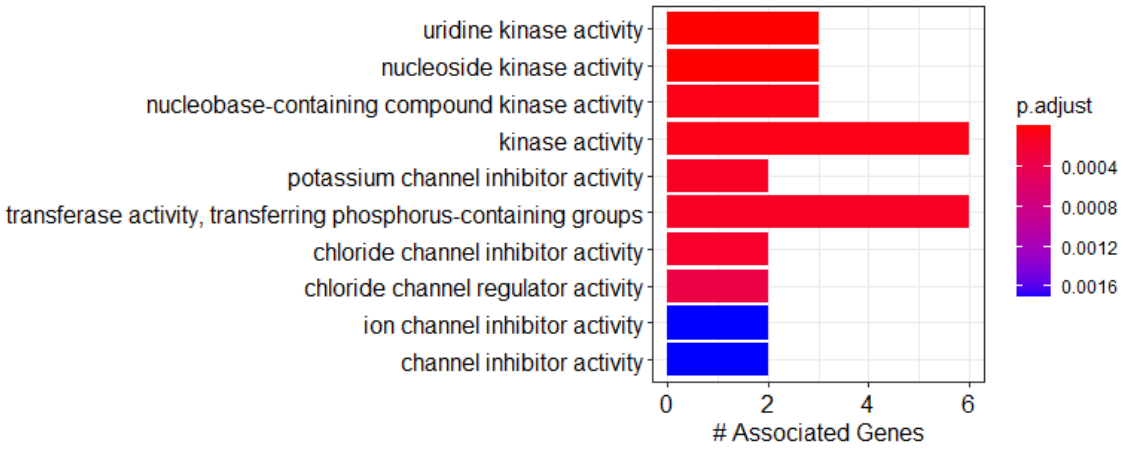

Figure 4.39. Gene Ontology: Molecular function annotation for PPI cluster 85.

This result can be seen in Figure 4.39. Cluster 85 also contains the IMPDH1 gene, a breast cancer drug target shown earlier in Table 4.12. Due to IMPDH1 being found to be underexpressed in pancreatic cancer in this study, IMPDH1 was discounted as a viable drug target in pancreatic cancer. However, this PPI clustering result implies that a therapeutic approach inhibiting IMPDH1 activity can be considered viable for pancreatic cancer treatment if the intervention can sufficiently inactivate pancreatic-cancer-related proteins that are neighbors of IMPDH1 without damaging the healthy metabolism too severely.

4.3.3.2. Enrichment of PPI Clusters with a High Proportion of Significant DEGs. PPI cluster 3 contains several cancer drug targets: ERBB2, SRC, EGFR, ABL1, KIT, EPOR, of which only ERBB2 was found to be a DEG in this study. Being such a large PPI cluster, this result may have been expected, yet cluster 3 is markedly more enriched in DEGs compared to its neighboring clusters 2 and 4. Additionally, this cluster also contains several high metabolic activity genes like MT-CO2, involved in oxidative phosphorylation, and SLC5A1, which is involved in glucose transport. These genes both have reactions with DSD ≥ 1, suggesting that the central carbon metabolism may be well represented in this cluster.

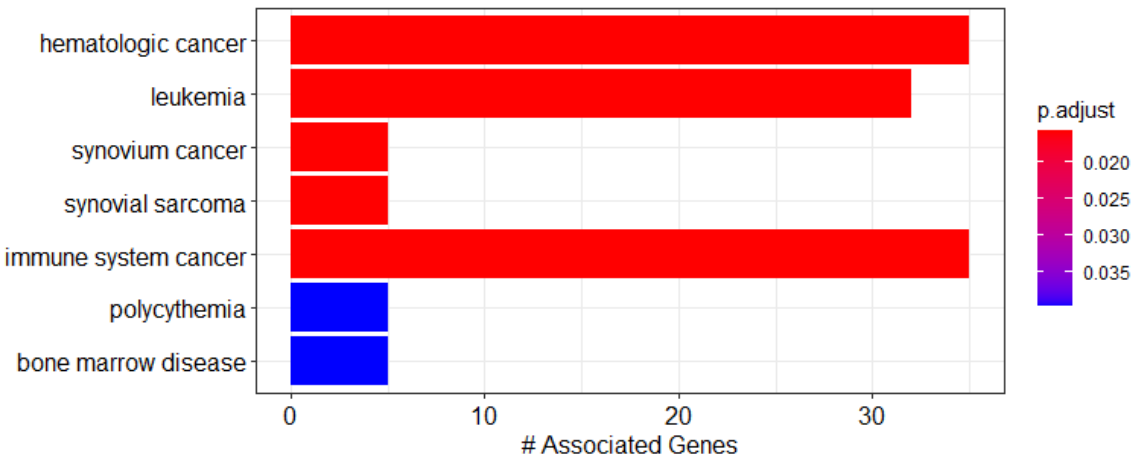

Figure 4.40. Disease ontology annotation for PPI cluster 3.

Figure 4.40 represents the disease ontology enrichment analysis results done on this cluster, showing cancer enrichment. The results of GO molecular function and biological

process enrichment analyses on cluster 3 do not, unfortunately, provide information beyond confirming that cluster 3 is enriched in ERBB2/EGFR signaling and therefore related to pro-growth signaling pathways.

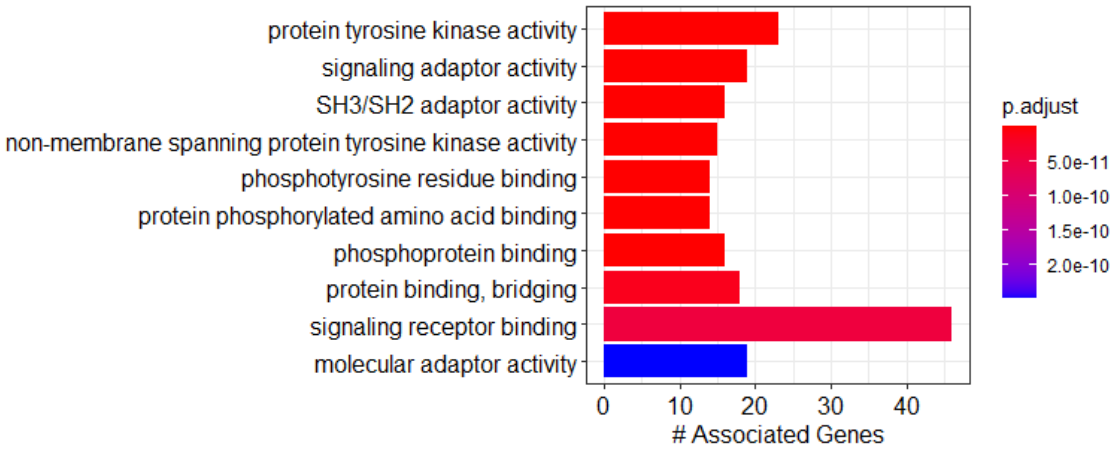

Figure 4.41. Gene Ontology: Molecular function annotation for PPI cluster 3.

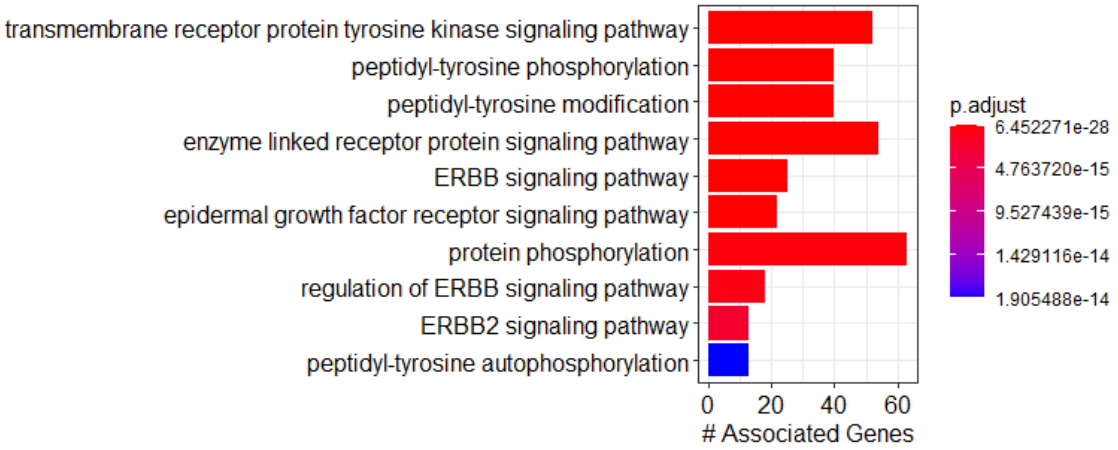

Figure 4.42. Gene Ontology: Biological process annotation for PPI cluster 3.

KEGG pathway enrichment results reveal that cluster 3 is enriched in insulin signaling as well as ERBB2/EGFR signaling. Although pancreatic ductal adenocarcinoma is not a pancreatic neuroendocrine tumor (PNET) and is therefore not ordinarily associated with

insulin-secreting tissue, this may indicate that perturbations to this cluster may cause the disruption of insulin secretion. Once again, insulin secretion's coregulation with the PI3K signaling pathway may be a promising avenue of research [196, 197].

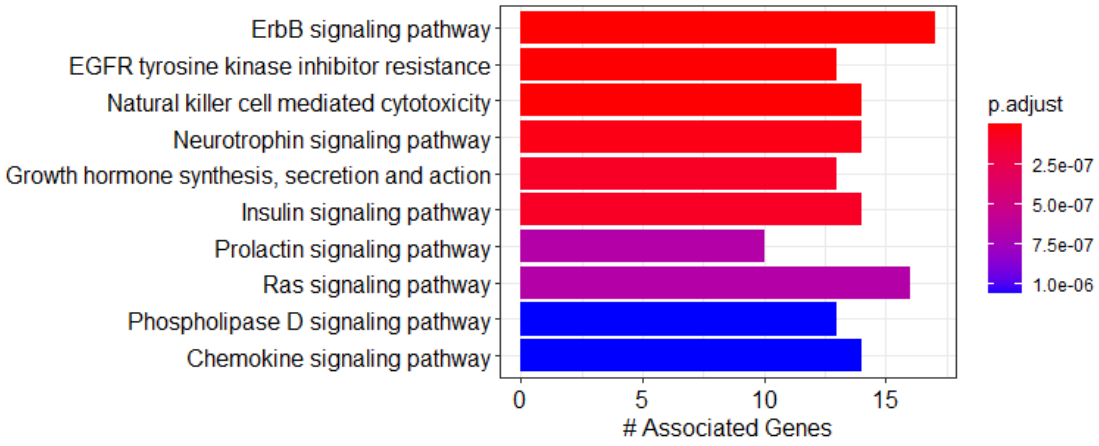

Figure 4.43. KEGG pathway annotation for PPI cluster 3.

The PPI cluster 11 shows enrichment in terms related to cell adhesion and cell-cell junctions, which are phenomena that have characterized many findings of this present work. The fact that many DEGs related to cell adhesion co-occur in a PPI cluster is an indication that therapeutic approaches that can target one such protein can have outsized effects on neighboring proteins.

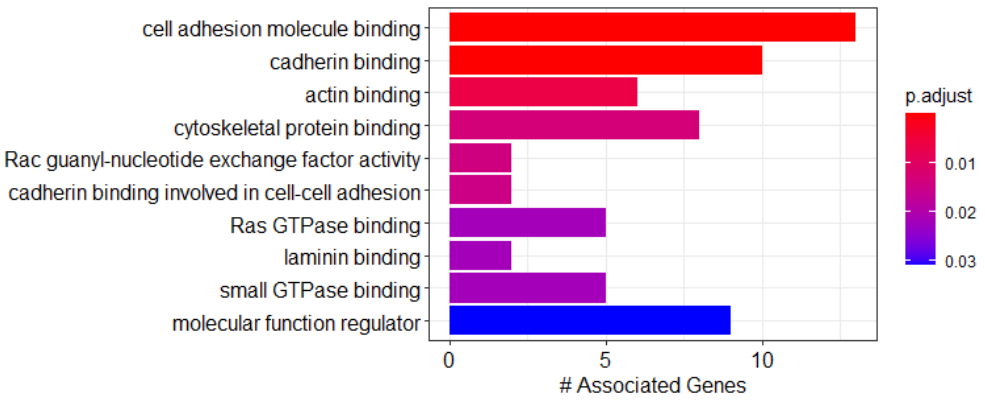

Figure 4.44. Gene Ontology: Molecular function annotation for PPI cluster 11.

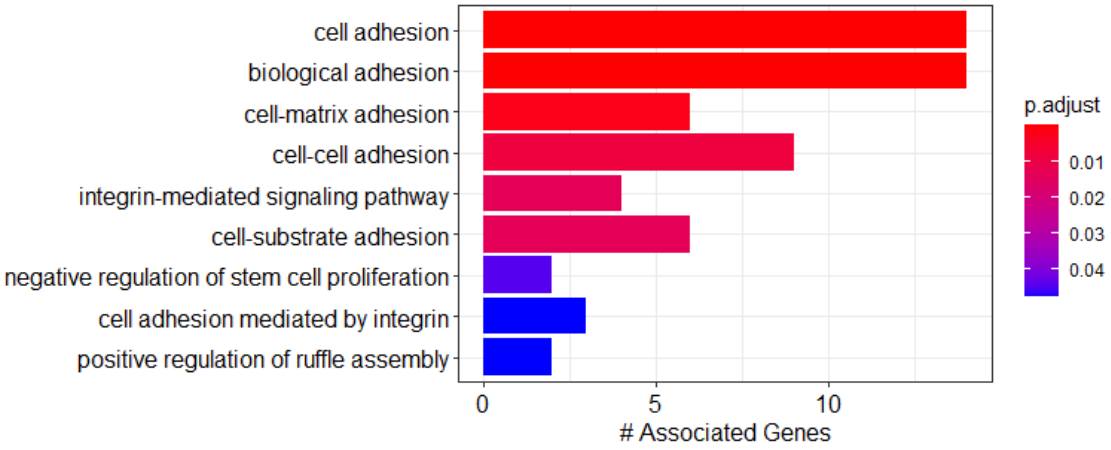

Figure 4.45. Gene Ontology: Biological process annotation for PPI cluster 11.

Figure 4.44 and Figure 4.45 show that cluster 11 contains these adhesion-related genes which interact with signaling cascades such as the Ras GTPase pathway, which seed gene KRAS is a major constituent of.

PPI cluster 84's enrichment analyses show an eclectic assortment of related terms, indicating this cluster has activity in hormone binding, cell cycle regulation, extracellular signal transduction, and protein catabolism through ubiquitination among GO term enrichment results.

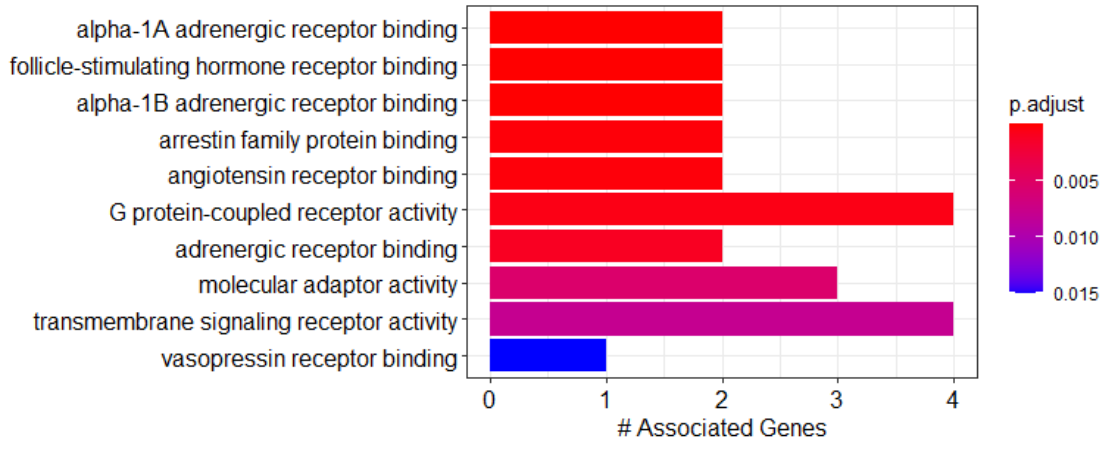

Figure 4.46. Gene Ontology: Molecular function annotation for PPI cluster 84.

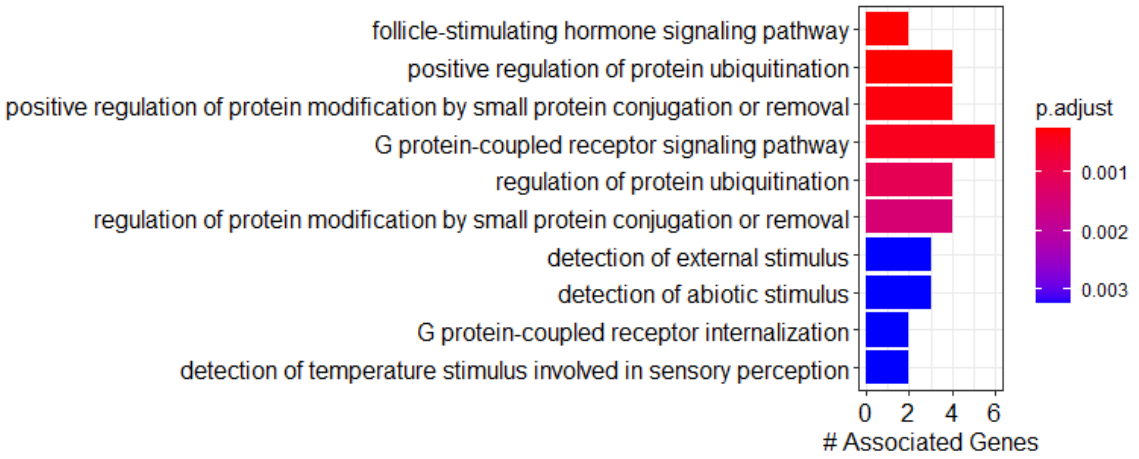

Figure 4.47. Gene Ontology: Biological process annotation for PPI cluster 84.

Cluster 84's KEGG pathway annotations, show a similar eclectic mixture, but clarify that this cluster may be most significantly concerned with intercellular signaling through intermediate metabolites such as neurotransmitters or chemokines. This makes cluster 84 a likely candidate for diagnostic approaches that measure these signaling metabolites.

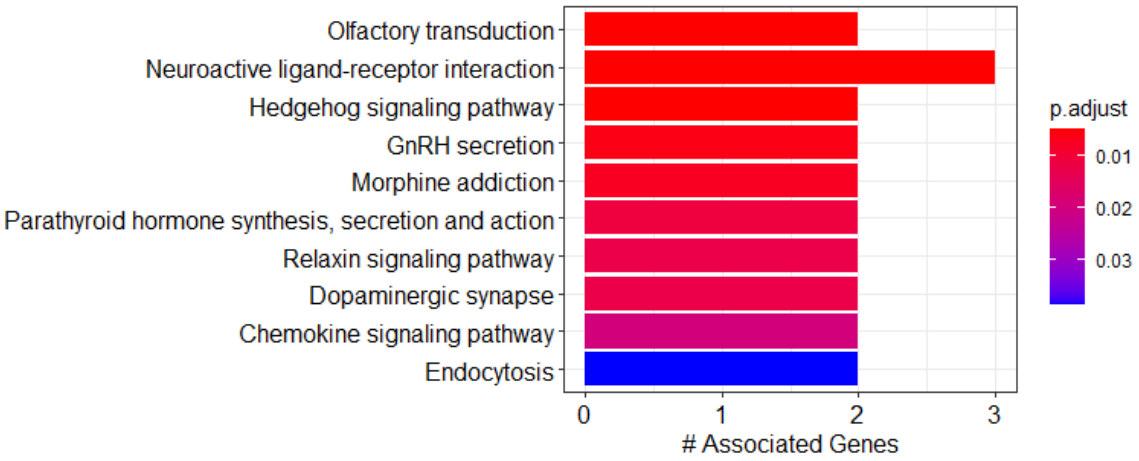

Figure 4.48. KEGG pathway annotation for PPI cluster 84.

Additionally, cluster 84 contains several genes that are by themselves interesting. ARRB1 and ARRB2 are genes encoding arrestin-beta variants. Both of these genes encode

signaling proteins that dampen the cellular response to external stimuli, and ARRB2 was found to be significantly downregulated in pancreatic cancer in this study ($p < 0.01$). There is an interesting parallel between the discovery of ARRB2's differential expression and insulin metabolism in pancreatic endocrine cells. It has already been discussed in this study that the cell cycle is coregulated with insulin secretion through the action of inositol phosphate signaling and the PI3K pathway [196, 197]. A study by Ravier et al. [256] found that ARRB2 deletion in mice caused a decrease in the mass of pancreatic endocrine beta cells. Cluster 84 could therefore inform how the endocrine behavior of the pancreas could be useful in understanding whether the exocrine pancreas is likely to have cancer. In addition to ARRB1 and ARRB2, cluster 84 was found to contain the OXER1 gene, found to be significantly underexpressed in pancreatic cancer ($p < 0.01$). OXER1 is an eicosanoid receptor mediating inflammation response. This is an unexpected finding due to existing research having found OXER1 to be overexpressed in breast and prostate cancer, and involved in promoting a survival response in cancer cells [257, 258]. Lastly, cluster 84 contains the gene NTSR1, which encodes neurotensin receptor. While it was not found to be a significant DEG in this study, there is growing evidence that neurotensin receptors are pancreatic ductal adenocarcinoma biomarkers and several studies have attempted to use NTSR1 expression levels to distinguish the healthy pancreas, pancreatitis, pancreatic cancer precursor lesions like PanINs, and pancreatic cancer from each other, finding an overexpression of NTSR1 in pancreatic cancer compared to all other conditions tested [259–261].

Cluster 89 can also be seen to be highly enriched in junction protein-encoding genes. Its disease ontology annotations. can be read as a survey of the diseases that are possibly related to any dysregulation in tight junction formation, since cluster 89, small as it is, contains 5 types of claudins and 3 types of TJP genes, encoding tight junction signaling proteins.

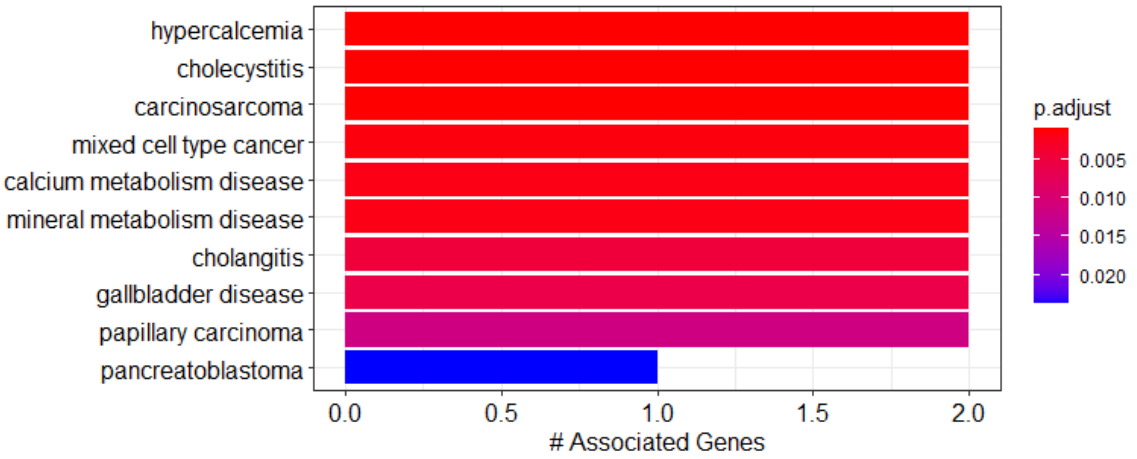

Figure 4.49. Disease ontology annotation for PPI cluster 89.

The GO molecular function and biological process enrichment analysis results for PPI cluster 89 also show the enrichment of tight junction and gap junction terms.

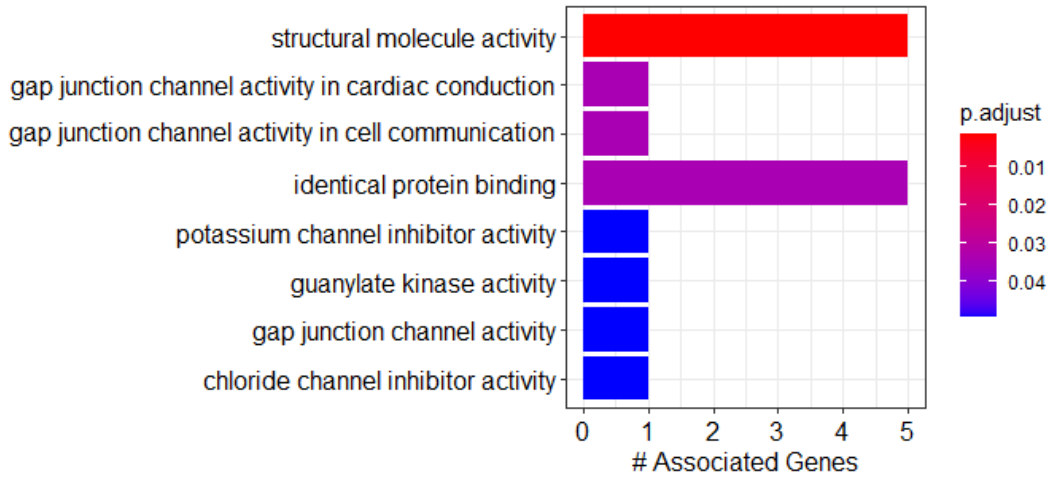

Figure 4.50. Gene Ontology: Molecular function annotation for PPI cluster 89.

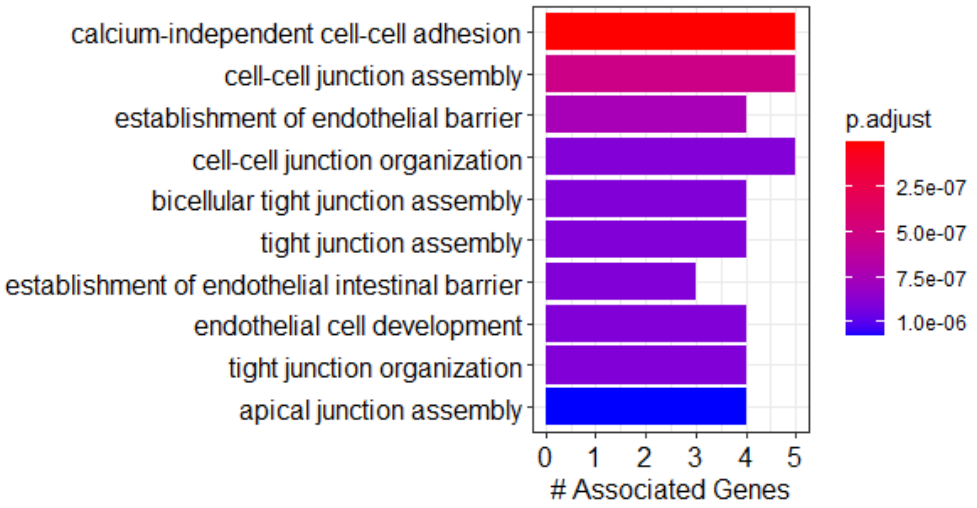

Figure 4.51. Gene Ontology: Biological process annotation for PPI cluster 89.

PPI cluster 89's annotation with KEGG pathways reveals, interestingly, the enrichment of immune response and infection related terms in this cluster, relating the prior findings of immune response dysregulation and cell-cell junction dysregulation to each other.

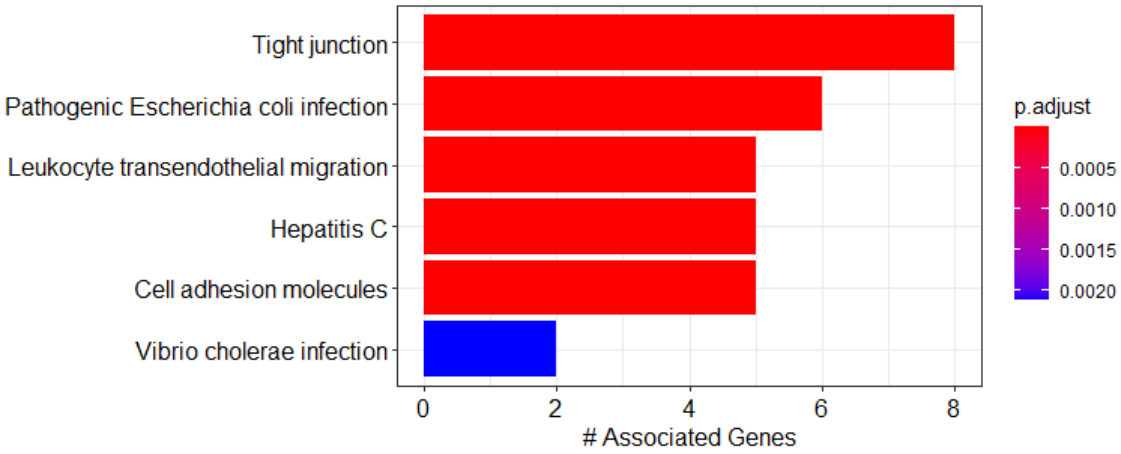

Figure 4.52. KEGG pathway annotation for PPI cluster 89.

4.3.3.3. Enrichment of Other PPI Clusters with Cancer-Related Enrichment Results. In addition to the clusters shown in Table 4.20, two other PPI clusters, clusters 12 and 17, were subjected to annotation analysis. Cluster 12 has a total of 35 genes, 1 of which (CERS4, encoding ceramide synthase 4) was found to be both a significant DEG and a gene with significant reaction flux differences. Additionally, this cluster contains SLC19A1, SLC38A1, SLC6A6, SLCO4A1, all genes found to have DSD ≥ 1 among their reactions, encoding solute carrier proteins transporting folate, thiamine, glutamine, taurine, beta-alanine, and conjugated bile acids, among others. Furthermore, cluster 12 has 5 more genes whose reactions show significant metabolic fluxes with DSD < 1. The aggregation of these properties makes cluster 12 worth investigating.

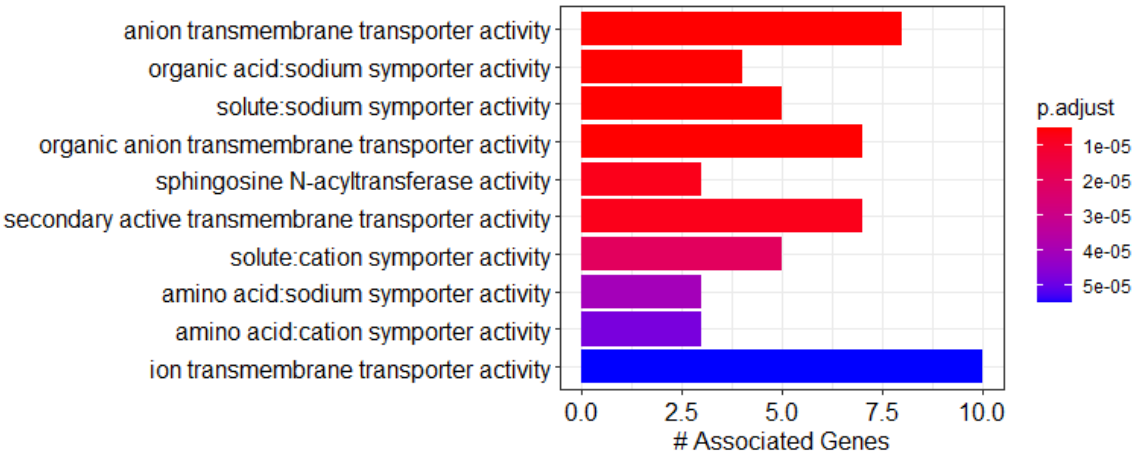

Figure 4.53. Gene Ontology: Molecular function annotation for PPI cluster 12.

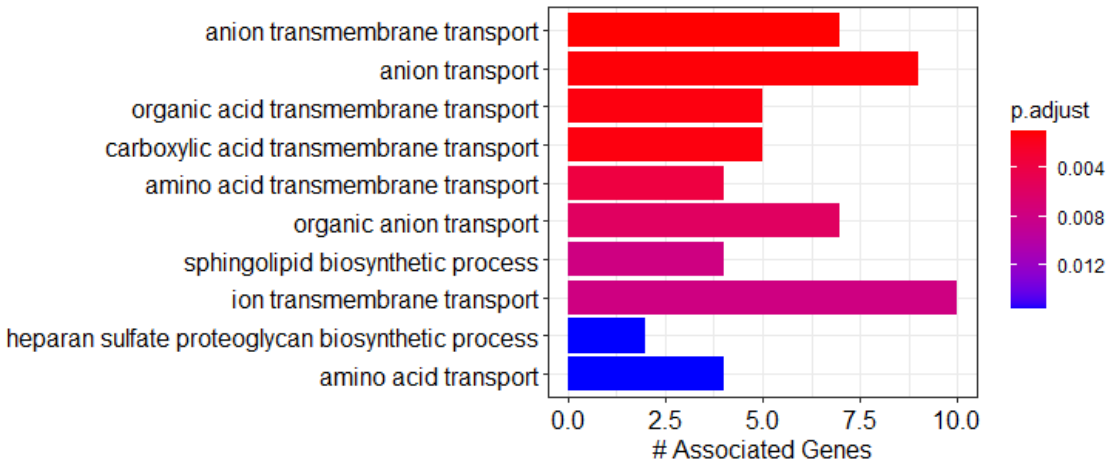

Figure 4.54. Gene Ontology: Biological process annotation for PPI cluster 12.

The very strong associations PPI cluster 12 has with secretion processes make it noteworthy in finding diagnostic biomarkers, since secreted metabolites can be discovered relatively non-invasively. Additionally, PPI cluster 12 shows enrichment in terms pointing towards sphingolipid metabolism, which is, as previously discussed, a popular area of cancer research especially based on the sphingolipids' erratic behavior in different cancer types [24, 26, 29]. Sphingolipids are known to play a part in the regulation of many other subsystems or pathways so far discovered in this study to be likely relevant to pancreatic cancer, such as the arachidonic acid pathway. In a review by Brachtendorf et al., existing research on ceramide synthase 4 is summarized to show that generally, an increase in CERS4 expression is correlated with less cancer proliferation and more cancer apoptosis [262].

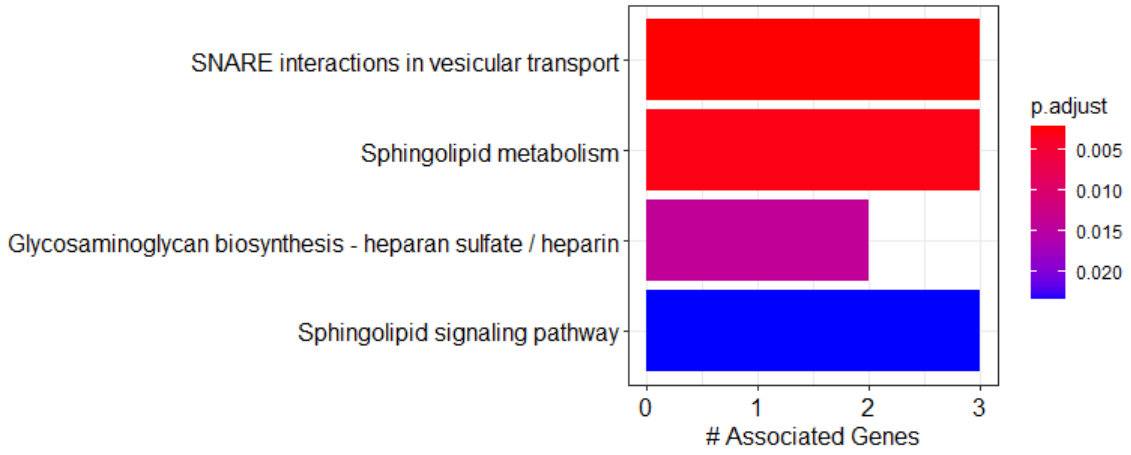

Figure 4.55. KEGG pathway annotation for PPI cluster 12.

Both Figure 4.54 and Figure 4.55, which show biological process GO term annotations and KEGG pathway annotations, respectively, for cluster 12, show the enrichment of sphingolipid metabolism within this cluster, in addition to enrichment in heparan sulfate metabolism. Figure 4.53 shows that the proteins in this cluster are involved in transmembrane transport.

Cluster 17 has 26 genes, of which only 3 are DEGs or genes with significant reaction flux differences. However, this cluster was chosen for investigation since it contains NQO1, which is a gene found to be very significant in this study both as a DEG and in its associated reactions' flux differences. Additionally, cluster 17 contains TP53, which is one of the most central hub genes in the entire PPI network and a cancer seed gene, encoding tumor suppressor protein p53.

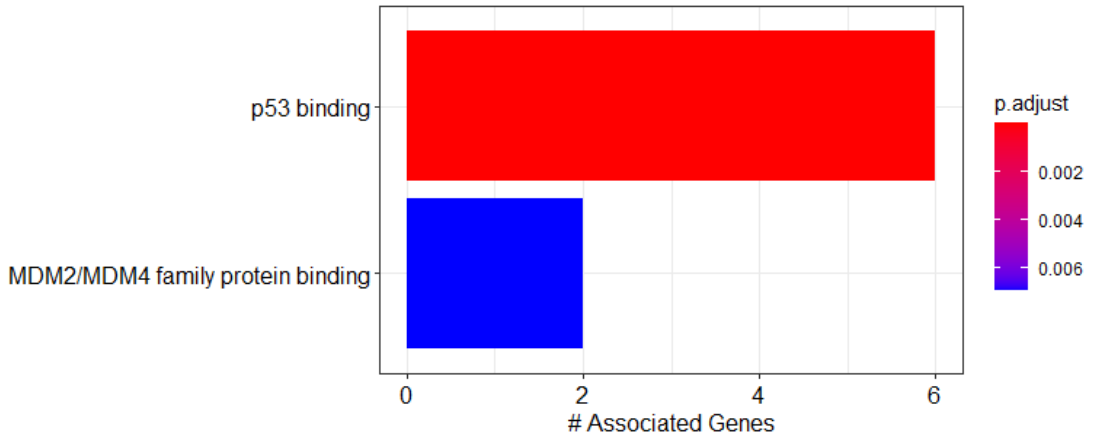

Figure 4.56. Gene Ontology: Molecular function annotation for PPI cluster 17.

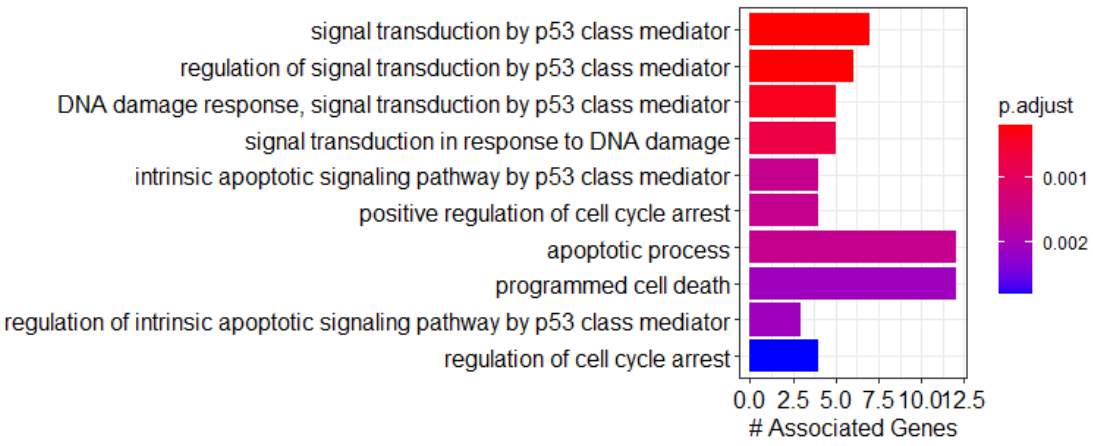

Figure 4.57. Gene Ontology: Biological process annotation for PPI cluster 17.

4.4. Coexpression Networks

Coexpression networks were separately generated using the R package WGCNA for the healthy and cancer phenotypes. 148 cancer samples and 252 healthy samples were considered, spanning 53,196 genes. The 32,556 genes, found by the limma R package's filterByExpr package to be expressed at a very low level were filtered before WGCNA analysis as well, leaving 20,640 genes for both networks. Please refer to the results from

section 4.1.1 for information on the filtering process and the various ancillary characteristics of these datasets.

4.4.1. Network Construction

The WGCNA algorithm itself starts with a filtering step, whereby outlier samples are filtered. This process was carried out visually, by inspection of hierarchical clustering dendrogram plots of the expression profiles of each sample. The cancer network had one outlying sample removed, and the healthy network had no removal of samples.

Subsequently, WGCNA determines a closeness or distance measure between each gene using a topological overlap matrix (TOM), which necessitates a power, termed the "soft threshold," to which the Pearson correlation score between each gene will be raised. WGCNA prescribes algorithmically determining this power by fitting a scale-free topology model to the network at each possible value of the soft threshold and picking the lowest such power that satisfies some arbitrarily high "fit," defined by the R^2 value. This threshold R^2 value was chosen to be 0.8 for the present study, finding soft threshold powers in both networks to have a value of 12.

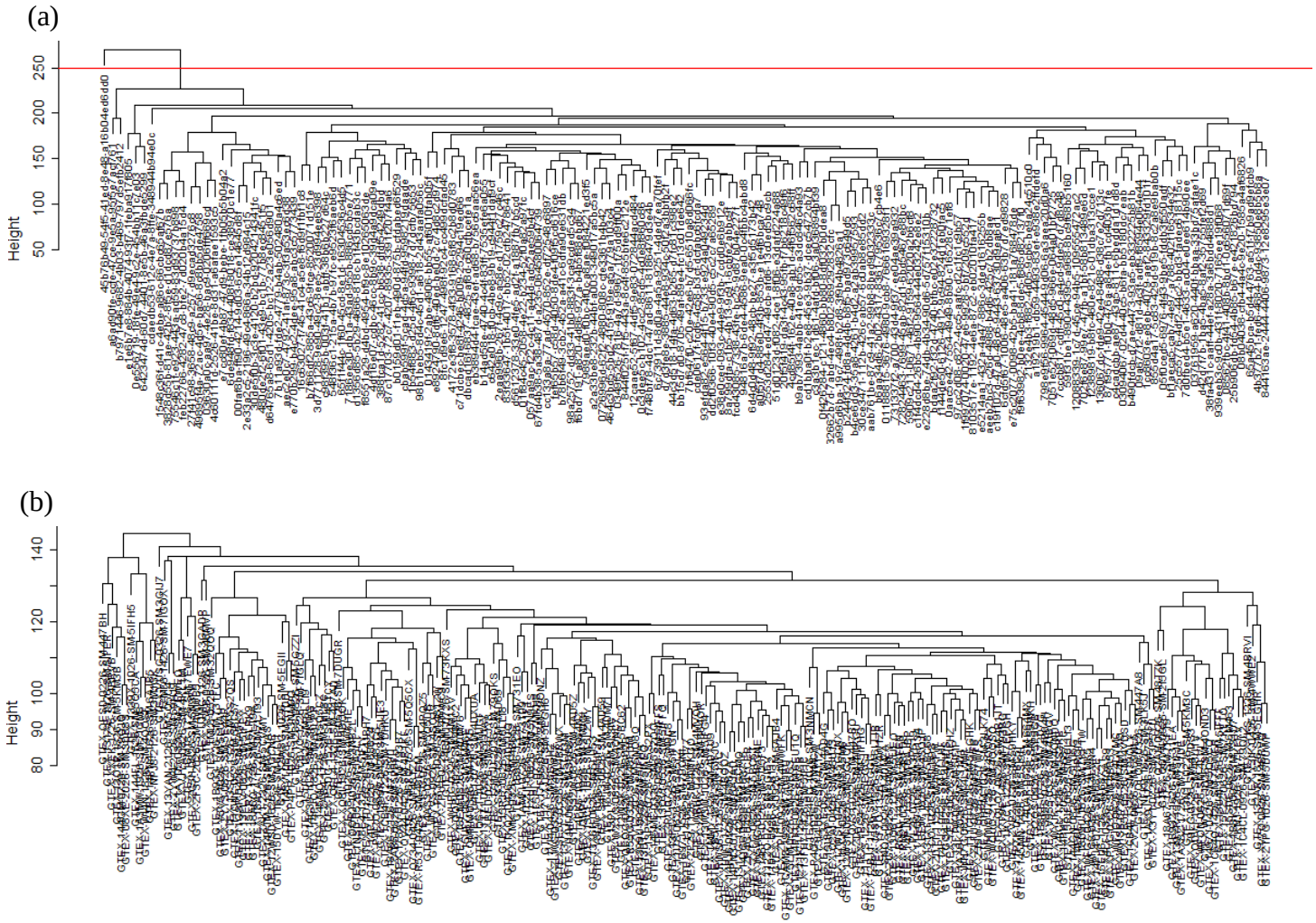

Figure 4.58. Hierarchical clustering to remove outliers in (a) the cancer dataset, and (b) the healthy dataset, the horizontal red line in (a) indicates the cut height, determined visually to indicate outliers.

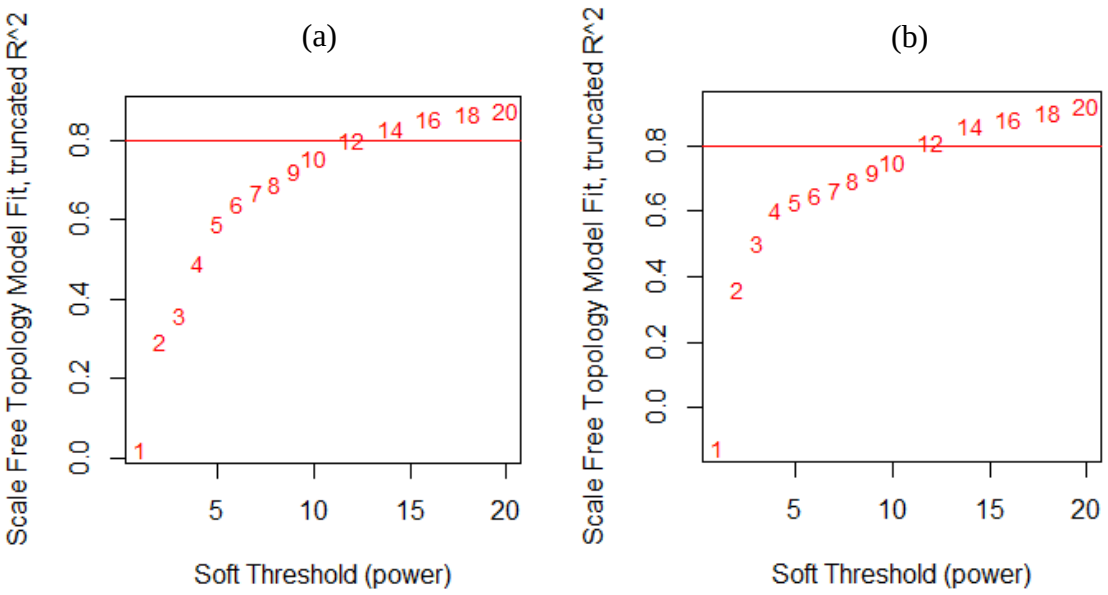

Figure 4.59. Soft threshold discovery via fit to scale free topology in the (a) cancer and (b) healthy WGCNA coexpression networks.

The Figure 4.58 is an illustration of the hierarchical clustering tree-cutting process and Figure 4.59 illustrates the range of soft thresholding powers tried to fit the TOM to a scale free topology, as well as the soft thresholds chosen.

After the TOM is constructed with the found soft thresholds, WGCNA automatically assigns all genes to clusters using a hierarchical clustering approach, and then merging clusters whose Pearson correlation exceeds a certain arbitrarily high value. In this study, this value was 90%. The resulting clusters are presented as hierarchical clustering dendrograms in below.

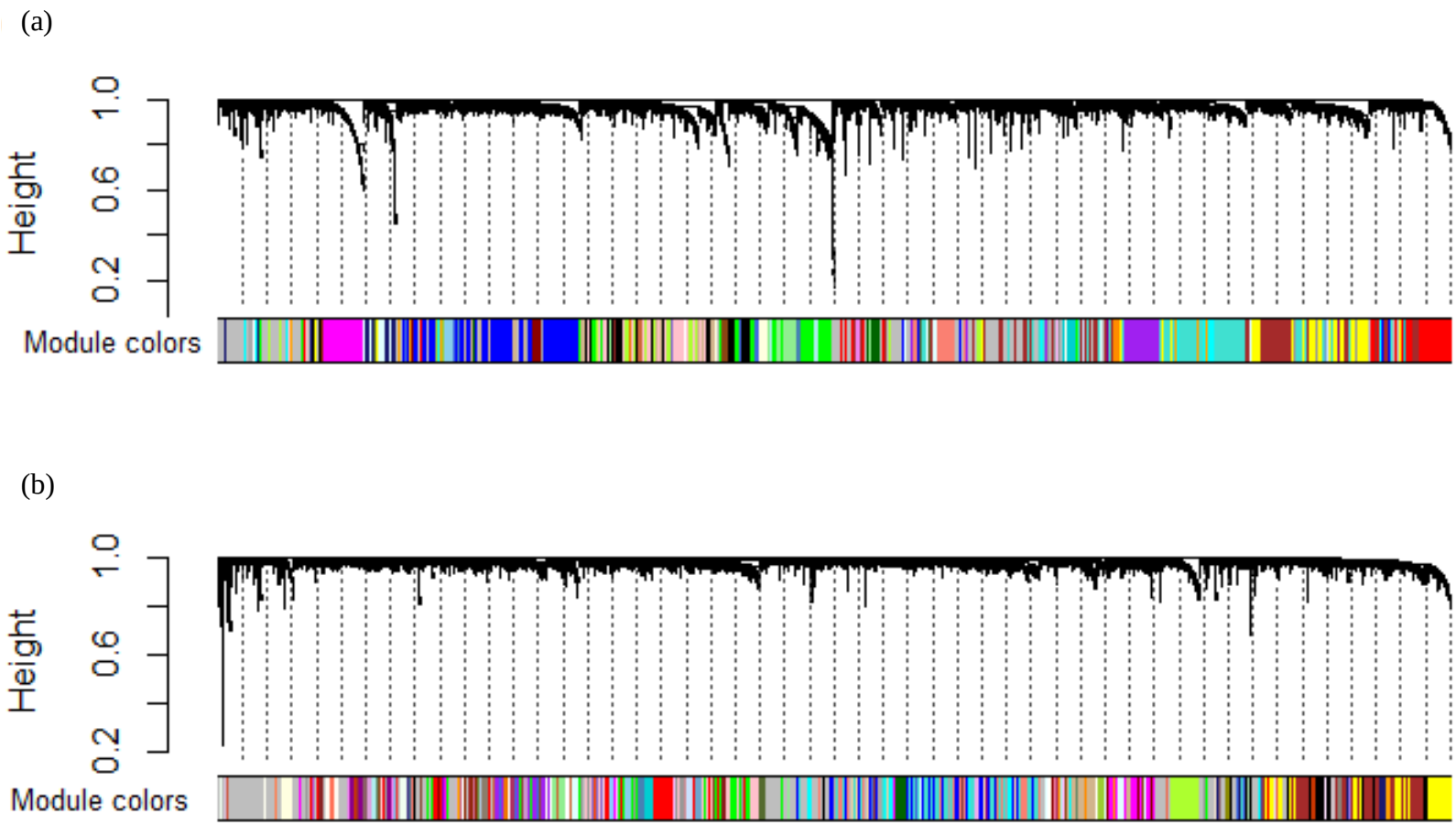

Figure 4.60. Coexpression clustering dendrograms for the (a) cancer and (b) healthy WGCNA coexpression networks.

4.4.2. Network Topological Properties

The WGCNA algorithm does not automatically calculate topological properties, so the fully connected, weighted network produced by WGCNA was imported into the Cytoscape platform. To cut down on computation time and make visualization easy, any edges with a topological overlap (or edge weight) less than 0.05 were pruned, leaving the network not fully connected. The networks' overall topological properties are in Table 4.21.

Table 4.21. Overall coexpression network topological properties.

Network Property	Healthy Network	Cancer Network
Nodes	2,910	7,321
Edges	52,139	355,306
Avg. Degree	35.83	97.06
Avg. Shortest Path Length	2.425	5.855
Avg. Clustering Coefficient	0.718	0.794
Avg. Betweenness Centrality	0.007857	0.003010

Whereas PPI networks showed broad similarity in their average topological properties, coexpression networks are very different, and especially show surprising topological behavior such as betweenness centrality being uncorrelated with degree between networks.

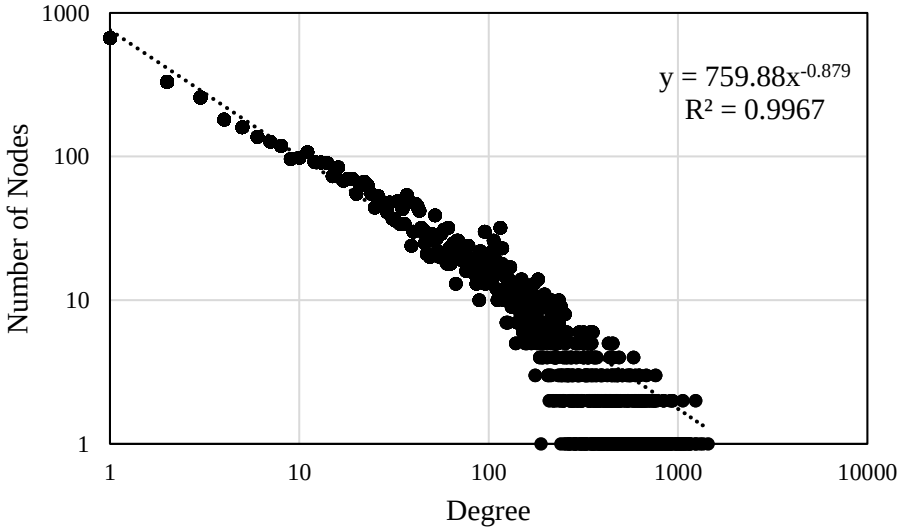

Figure 4.61. Power law fit to degree distribution in the cancer coexpression network.

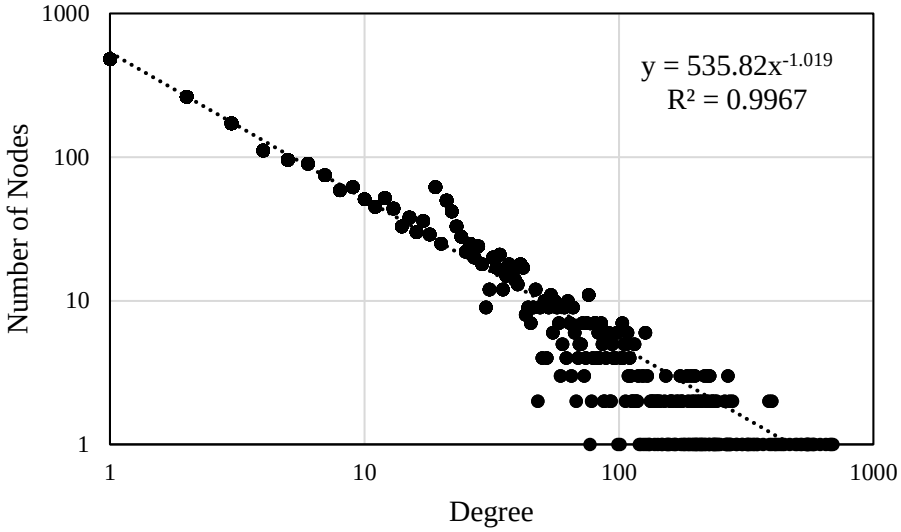

Figure 4.62. Power law fit to degree distribution in the healthy coexpression network.

Coexpression networks are also scale-free by design, which is illustrated, along with the fitted power law, in Figure 4.61 and Figure 4.62.

Table 4.22 is an illustration of how average network topological properties vary in subnetworks chosen from certain sets of significant genes.

Table 4.22. Coexpression network topological properties for subgroups of interest, color coded: cancer network properties – orange; healthy network properties – blue.

Network Property	Significant DEGs in Network	Genes with Sig. Diff. Metabolic Reactions in Network	Cancer Seed Genes in Network	Healthy Seed Genes in Network
Avg. Degree	106.60	98.88	99.81	98.45
Avg. Shortest Path Length	5.81	5.86	5.86	5.88
Avg. Clustering Coefficient	0.790	0.799	0.796	0.799
Avg. Betweenness Centrality	0.003058	0.003106	0.000664	0.000261
Avg. Degree	38.22	35.98	36.35	33.64
Avg. Shortest Path Length	2.43	2.43	2.42	2.44
Avg. Clustering Coefficient	0.721	0.718	0.718	0.717
Avg. Betweenness Centrality	0.003106	0.007725	0.007806	0.008122

The topological properties of the coexpression networks exhibit many differences to those of PPI networks. Most strikingly, the major difference between any two networks' or subnetworks' topological properties comes from whether they are generated using healthy or cancer gene expression data, whereas the choice of significant subnetworks matters comparatively little. Furthermore, DEGs were noticeably higher degree than the average gene in both coexpression networks. It is likely that this is a necessary part of the WGCNA algorithm, where the cancer network contains many genes that have been co-perturbed by a single driving condition (cancer), leading to necessarily highly interdependent expression levels.

4.4.3. Coexpression Network Clustering and Annotation

By default, WGCNA marks groups of coexpressed genes as clusters, but returns a fully connected network with weighted edges. This means that even if some low-weight edges were pruned, clusters could be connected to each other. Therefore, the following network visualizations do not depict clusters as separate entities, but as different colors. The cancer network was broken down into 46 clusters, whereas the healthy networks was broken down into 97. The results of these cluster enrichment analyses reveal, importantly, how certain groups of genes are coexpressed, which must be kept in mind as results are interpreted. Any group of genes in a certain cluster are not likely to show any concurrent or concordant behavior except for coexpression, in contrast to the PPI networks analyzed in the previous section.

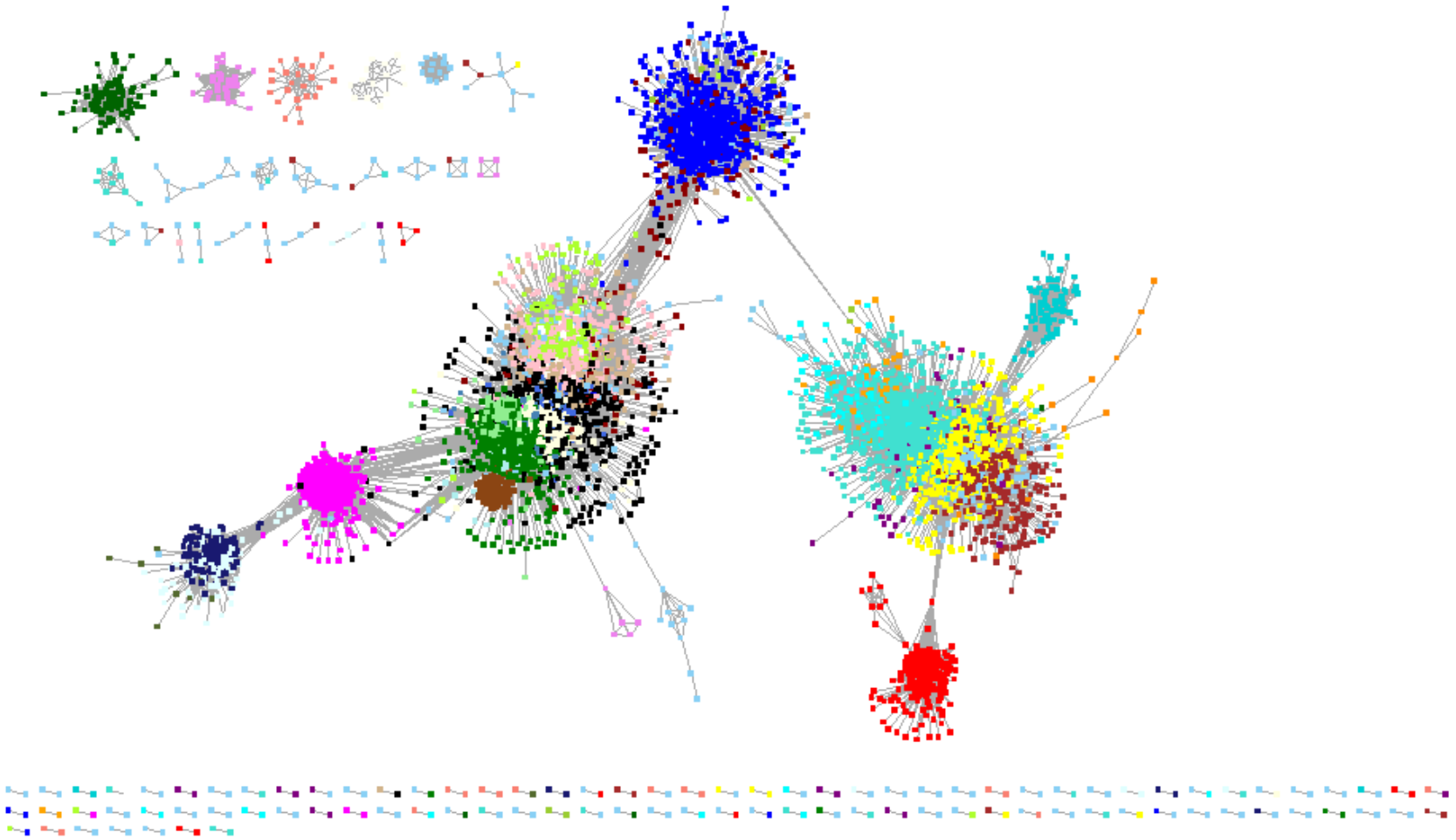

Figure 4.63. The cancer coexpression network, each cluster is given a distinct color, pale blue indicates unclustered nodes.

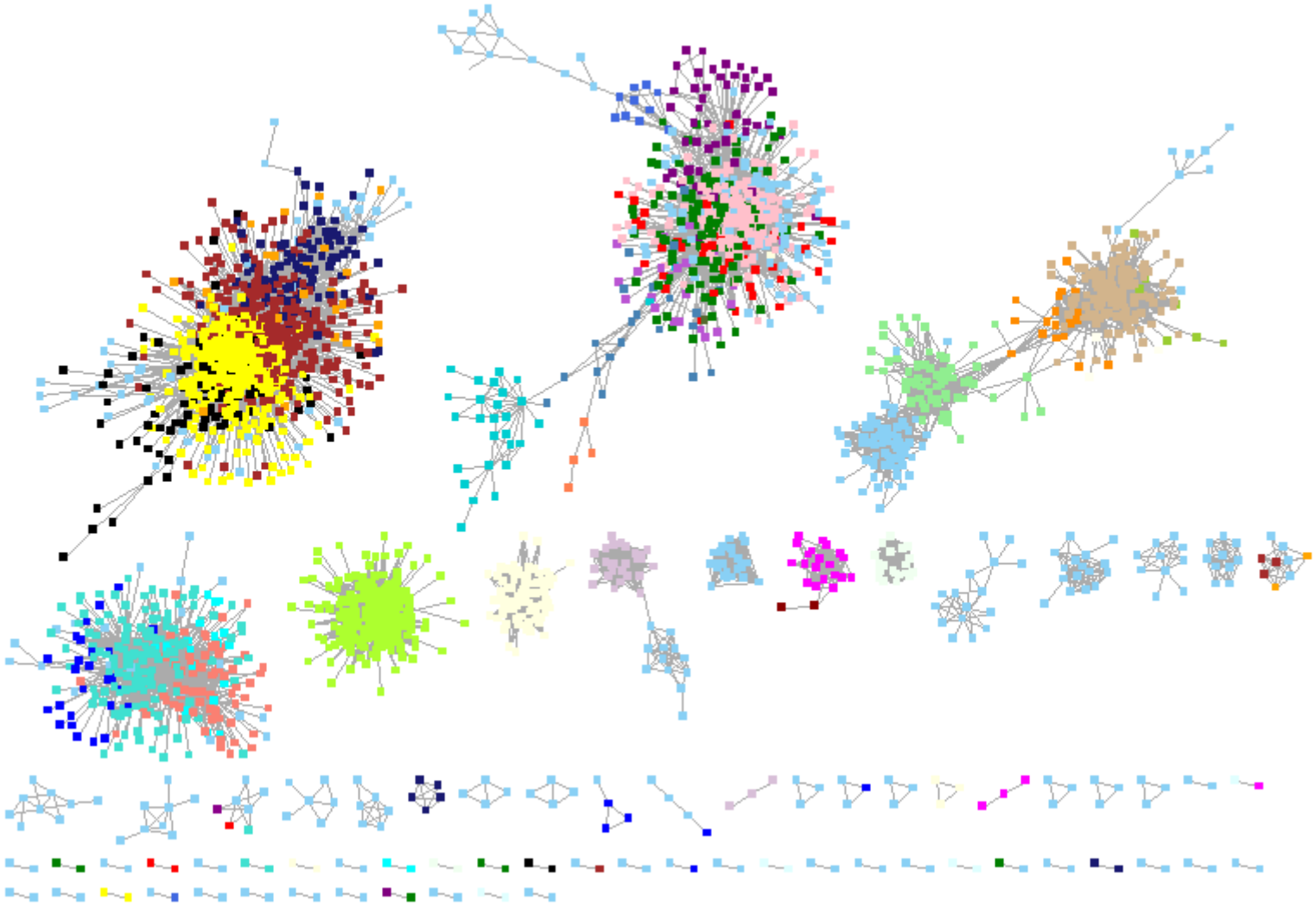

Figure 4.64. The healthy coexpression network, each cluster is given a distinct color, pale blue indicates unclustered nodes.

The WGCNA coexpression networks are different from the PPI networks in their clustering behavior as well, especially showing a much flatter cluster size distribution.

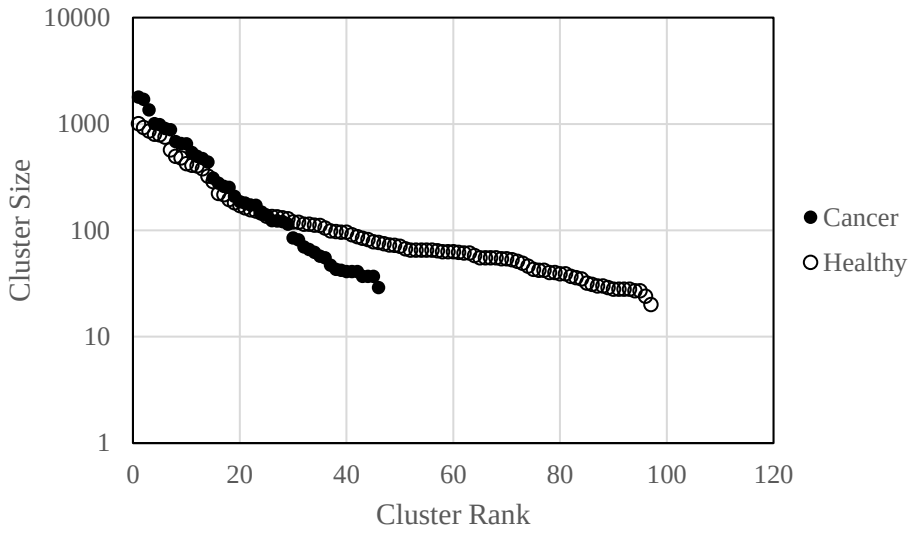

Figure 4.65. Cluster size distribution of the coexpression networks.

Furthermore, the distribution of significant genes (either DEGs or genes whose reactions exhibit metabolic flux differences) across clusters is not as skewed towards smaller clusters, and that DEGs are more represented at each cluster size range than genes with reactions having flux differences.

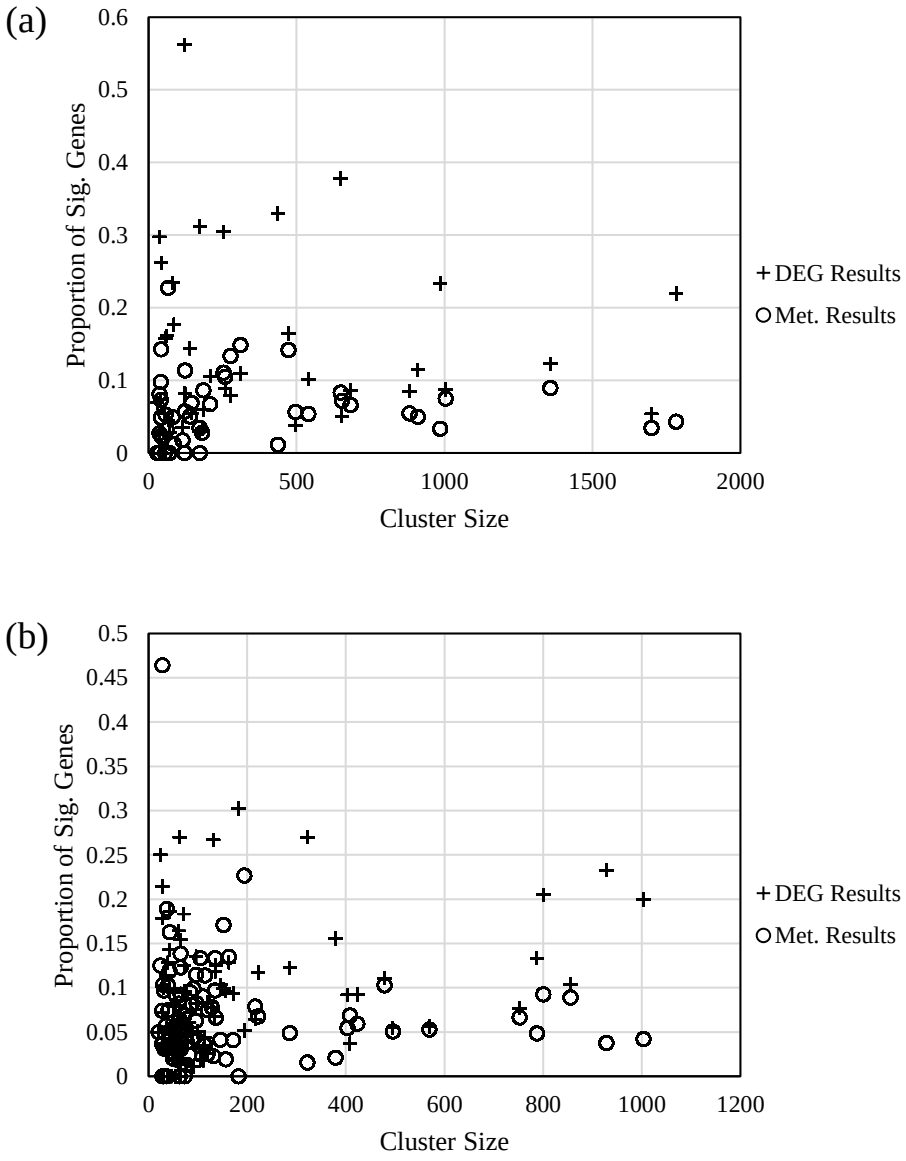

Figure 4.66. Distribution of significant genes across (a) cancer and (b) healthy coexpression network clusters.

Table 4.23 reports the 5 clusters with the highest proportion of DEGs and genes with metabolic flux differences in the cancer coexpression network. Since in this study the focus was on the cancerous transcriptome, clusters of interest were identified only for the cancer coexpression network. Table 4.23 is color coded: proportionally enriched in metabolically different genes – orange; proportionally enriched in DEGs – green.

Table 4.23. Cancer coexpression network clusters with the greatest proportion of DEGs or genes with metabolic differences.

Coexpression Cluster Name	Cluster Size	# of DEGs in Cluster	# of Met. Diff. Genes in Cluster	% of DEGs in Cluster	% of Met. Diff. Genes in Cluster
Salmon	473	78	67	17%	14%
Midnightblue	310	34	46	11%	15%
Lightcyan	277	22	37	8%	13%
Darkolivegreen	66	3	15	5%	23%
Orangered4	42	11	6	26%	14%
Purple	649	245	54	38%	8%
Cyan	437	144	5	33%	1%
Lightgreen	253	77	28	30%	11%
Darkgreen	173	54	0	31%	0
Skyblue	121	68	0	56%	0

These clusters were subsequently annotated using the R clusterProfiler package, drawing upon the disease ontology and gene ontology databases, using a p-value threshold of 0.05, followed by a Benjamini-Hochberg false discovery rate threshold of 0.2 to determine significance. All results of the annotation process that have met this significance threshold are represented below. Since coexpression clusters show interrelated expression behavior among their constituent genes, whether each cluster is upregulated or downregulated in cancer is also reported.

4.4.3.1. Enrichment of Coexpression Clusters with a High Proportion of Metabolically Significant Genes. The “salmon” coexpression cluster was annotated with molecular function and biological process GO terms as it was found to be enriched in both metabolically significant genes and DEGs, alongside being a sizable cluster.

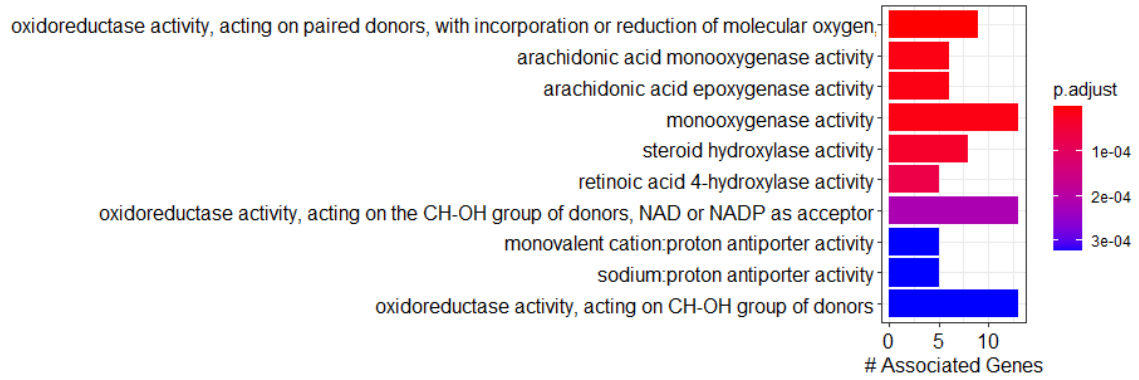

Figure 4.67. Gene Ontology: Molecular function annotation for cancer coexpression cluster “salmon”.

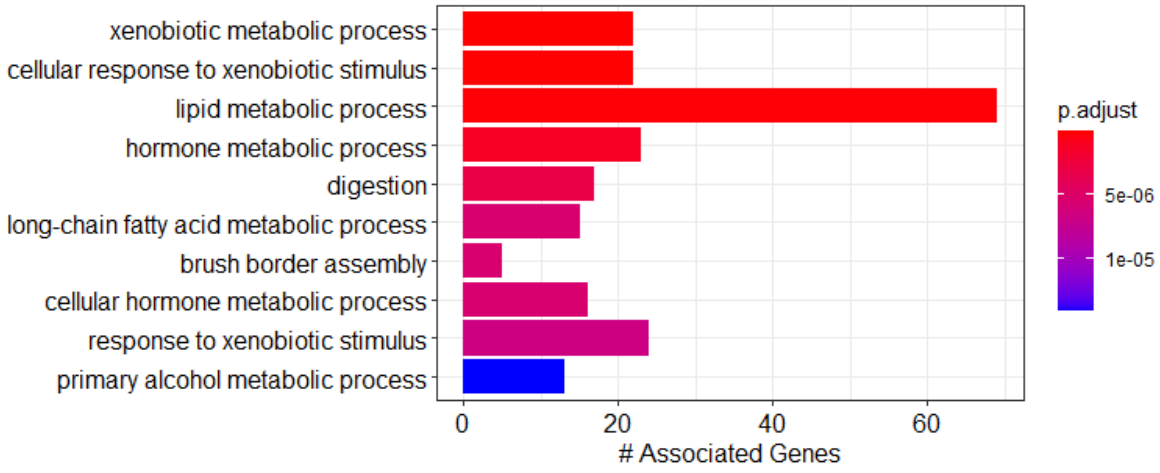

Figure 4.68. Gene Ontology: Biological process annotation for cancer coexpression cluster “salmon”.

Figure 4.67 and Figure 4.68 reveal the salmon cluster’s involvement in several pathways, such as the inflammation/arachidonic acid pathway, steroid metabolism, and retinol (vitamin A) metabolism. The cluster showed differential expression in a positive direction, that is, its constituent genes were mostly upregulated in cancer. Results of enrichment analysis with KEGG pathways paint a similar eclectic picture, this time also revealing enrichment in the “chemical carcinogenesis” pathway.

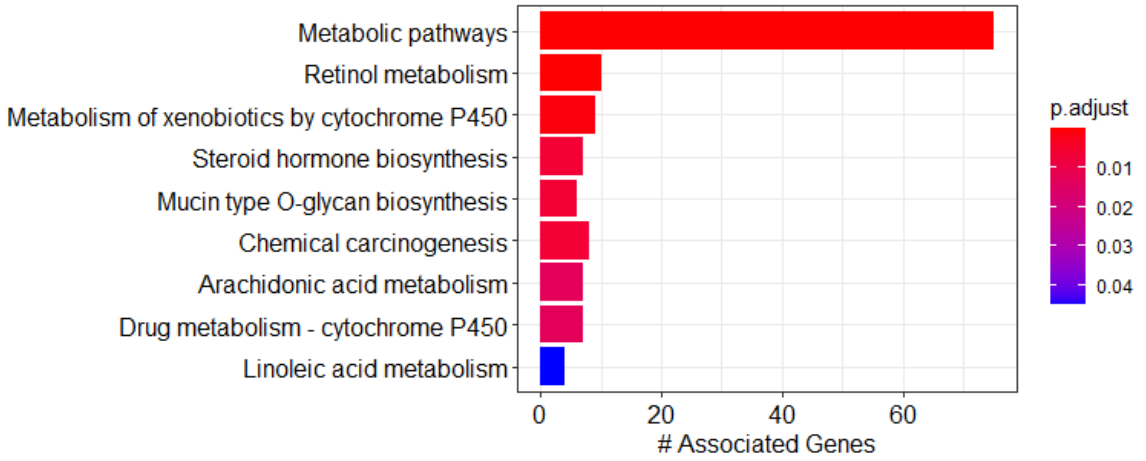

Figure 4.69. KEGG pathway annotation for cancer coexpression cluster “salmon”.

The “midnightblue” coexpression cluster was found to be downregulated in cancer, even though it contained the double lethal gene GMPR, discussed in an earlier subsection. This raises doubt on a therapeutic approach aimed at inactivating GMPR, since it is likely that such an approach will damage healthy tissues more. This cluster was found to be mostly related to pancreatic secretion and digestion.

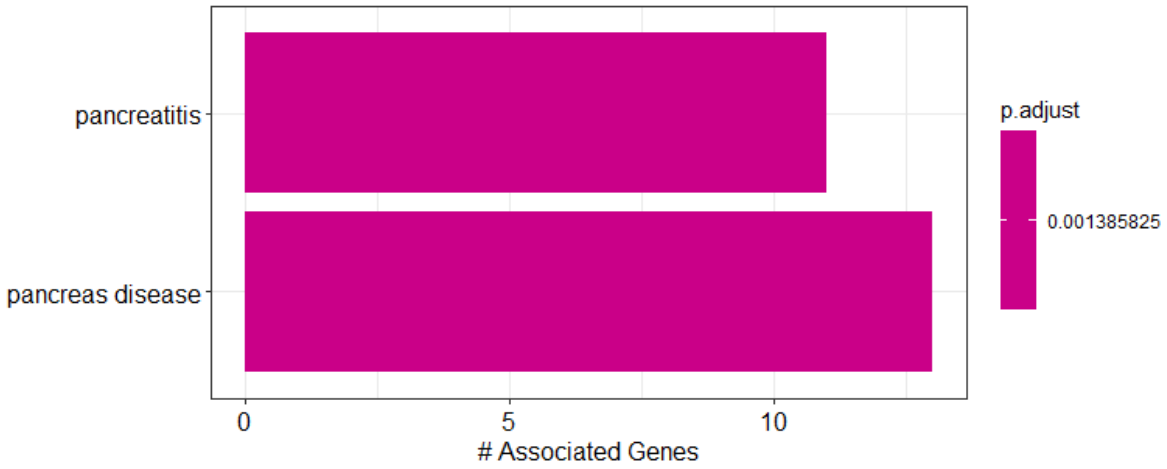

Figure 4.70. Disease ontology annotation for cancer coexpression cluster “midnightblue”.

Figure 4.70. shows the disease ontology enrichment analysis result showing that genes in this cluster were implicated in pancreatitis, but not in pancreatic cancer. Still, this cluster contains both DEGs and genes with significant metabolic flux variations at a proportion over 10%, and it may be relevant in diagnostic biomarker discovery since digestion and secretion processes can be indirectly interfered with, leading to less invasive tests.

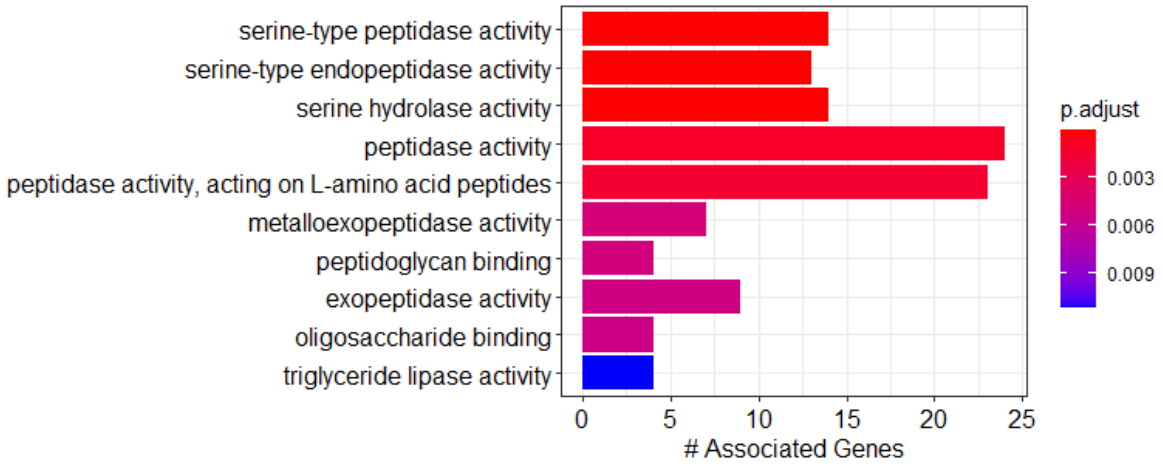

Figure 4.71. Gene Ontology: Molecular function annotation for cancer coexpression cluster “midnightblue”.

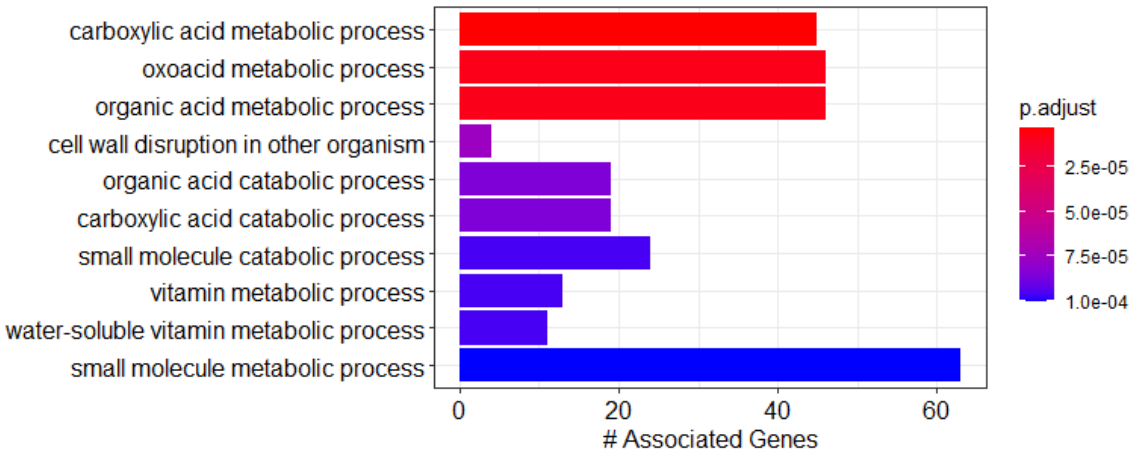

Figure 4.72. Gene Ontology: Biological process annotation for cancer coexpression cluster “midnightblue”.

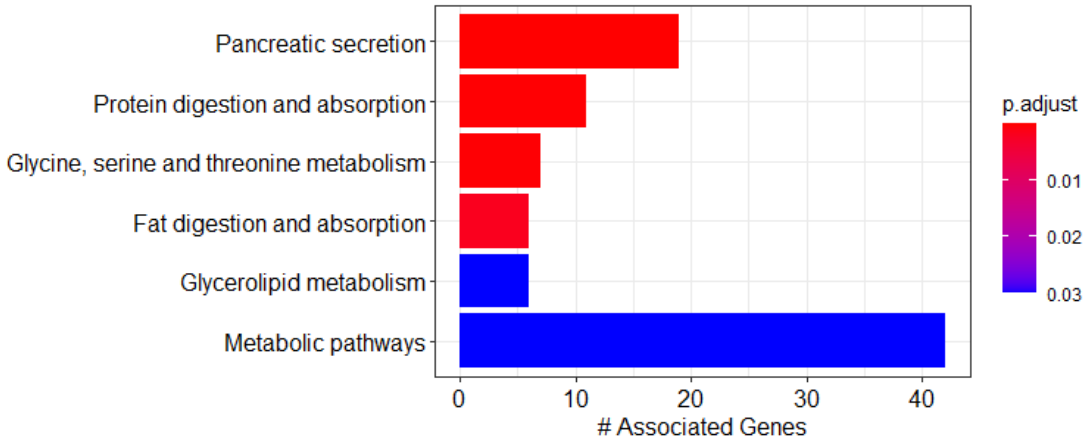

Figure 4.73. KEGG pathway annotation for cancer coexpression cluster “midnightblue”.

Figure 4.71, Figure 4.72 and Figure 4.73, depicting GO molecular function, GO biological process and KEGG pathway annotation results all corroborate the “midnightblue” cluster’s digestive functions.

The “lightcyan” coexpression cluster is also underexpressed in cancer. The most significant DEG it contains is CDH4, encoding the cell-cell junction protein cadherin 4, already previously discussed.

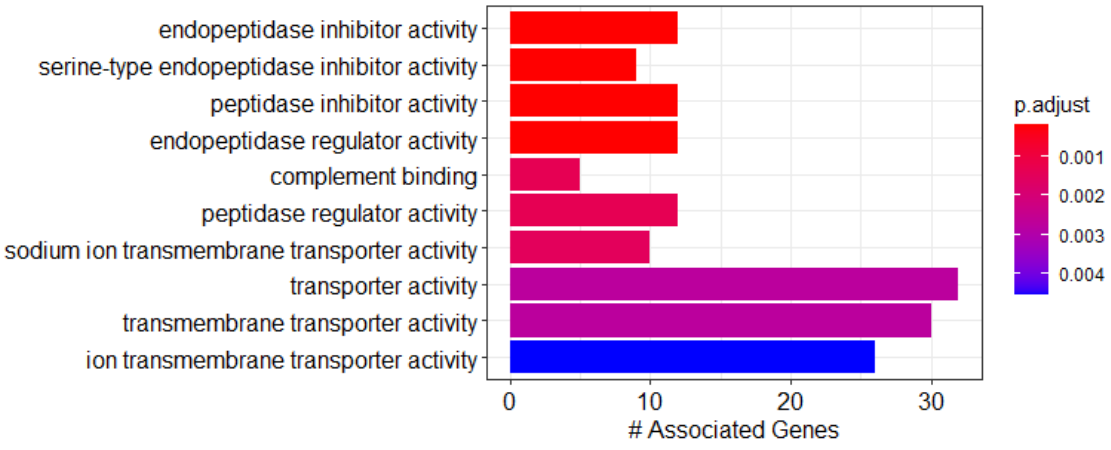

Figure 4.74. Gene Ontology: Molecular function annotation for cancer coexpression cluster “lightcyan”.

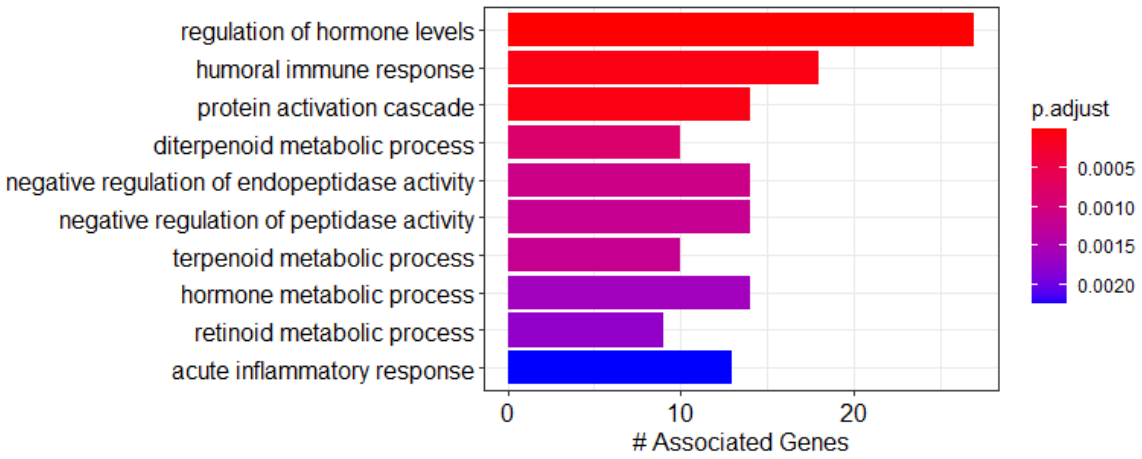

Figure 4.75. Gene Ontology: Biological process annotation for cancer coexpression cluster "lightcyan".

Figure 4.75 shows that the "lightcyan" cluster is enriched in genes related to many concepts that have been covered in this study already, including hormone metabolism, immune response and retinol (vitamin A) metabolism.

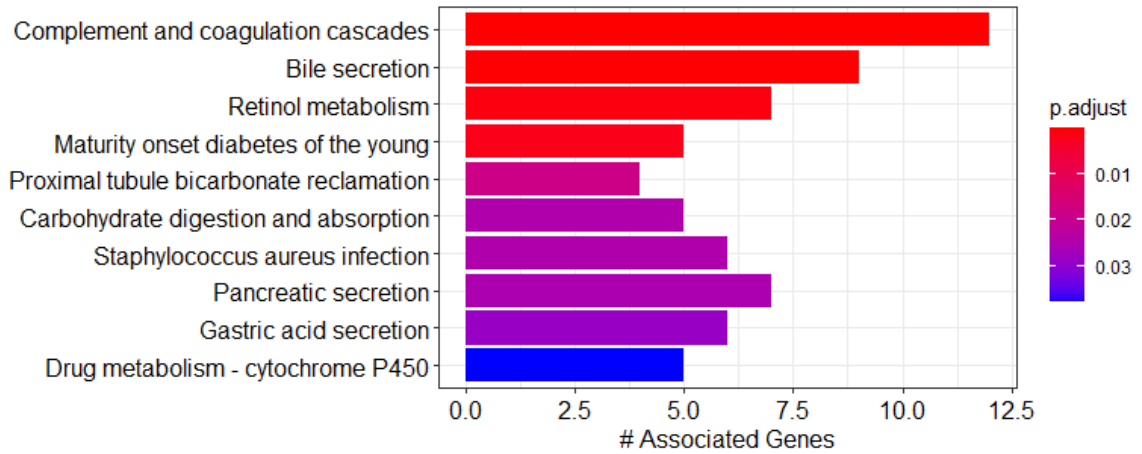

Figure 4.76. KEGG pathway annotation for cancer coexpression cluster "lightcyan".

The “darkolivegreen” coexpression cluster shows enrichment in some of the same terms as “lightcyan,” like bile secretion and retinol metabolism, but is an overexpressed cluster.

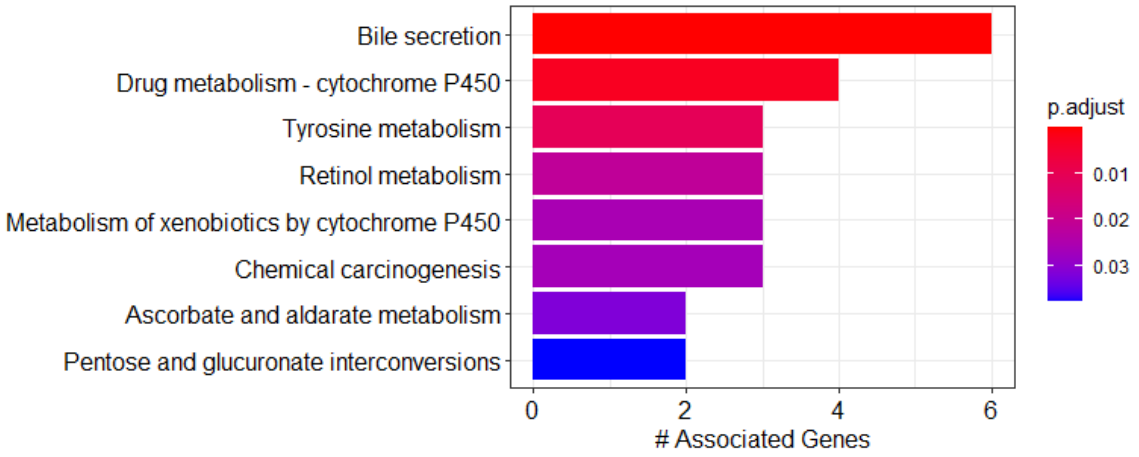

Figure 4.77. KEGG pathway annotation for cancer coexpression cluster “darkolivegreen”.

In Figure 4.77, it can also be seen to show enrichment in drug and xenobiotics metabolism, although it contains no known drug targets.

4.4.3.2. Enrichment of Coexpression Clusters with a High Proportion of Significant DEGs.

The “purple” coexpression cluster is one of the largest ones in the cancer coexpression network but it contains a large proportion of DEGs, including the most significant DEG, EPCAM. It also contains the previously discussed single lethal gene CDS1 and drug target SRD5A3.

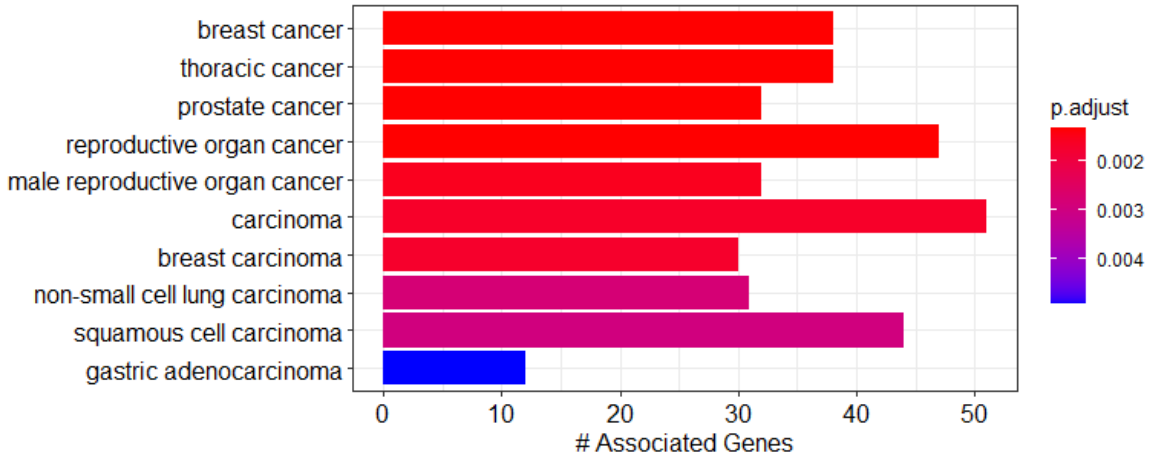

Figure 4.78. Disease ontology annotation for cancer coexpression cluster “purple”.

“Purple” is an upregulated cluster, and Figure 4.78 represents its enrichment in cancer-related genes, although pancreatic cancer enrichment is not found to be significant. The “purple” cluster also contains the KEGG seed genes, MAPK3, SMAD3. The MAPK pathway was previously discussed as promoting growth in subsection 2.1. More interesting is the upregulation of SMAD3, since SMAD4 was discussed, also in subsection 2.1., to be likely downregulated or mutated in cancer.

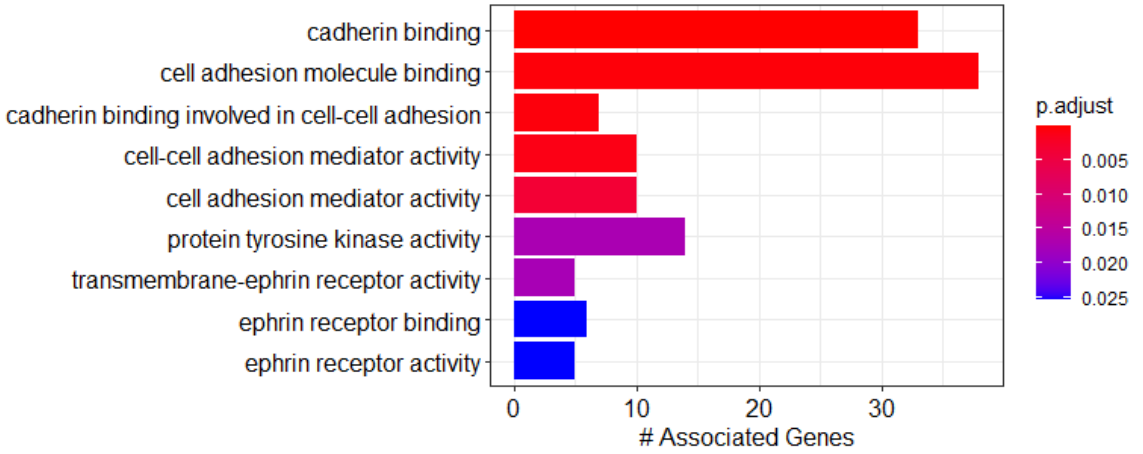

Figure 4.79. Gene Ontology: Molecular function annotation for cancer coexpression cluster “purple”.

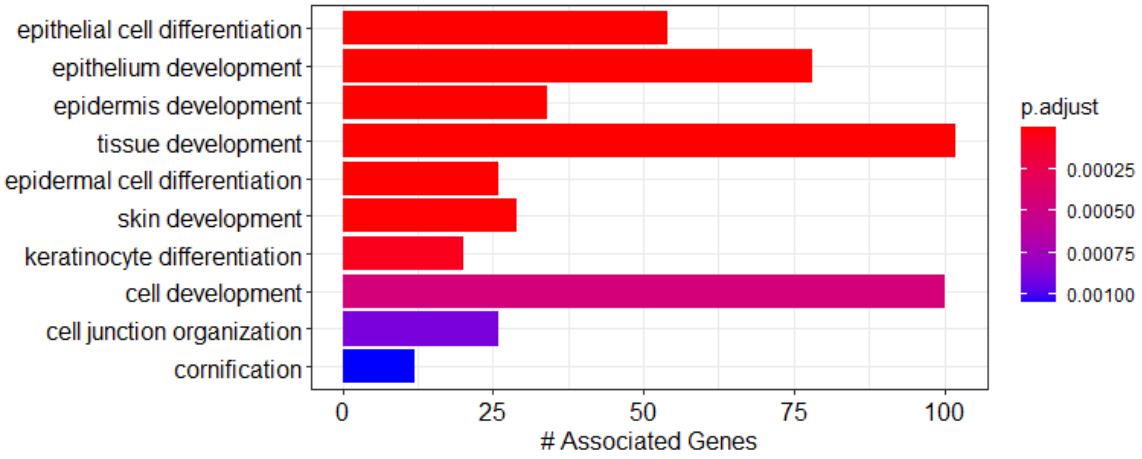

Figure 4.80. Gene Ontology: Biological process annotation for cancer coexpression cluster "purple".

Figure 4.79 and Figure 4.80 contain the GO molecular function and biological process enrichment results, respectively, and heavily represent enrichment in genes related to cell-cell junctions and cell adhesion.

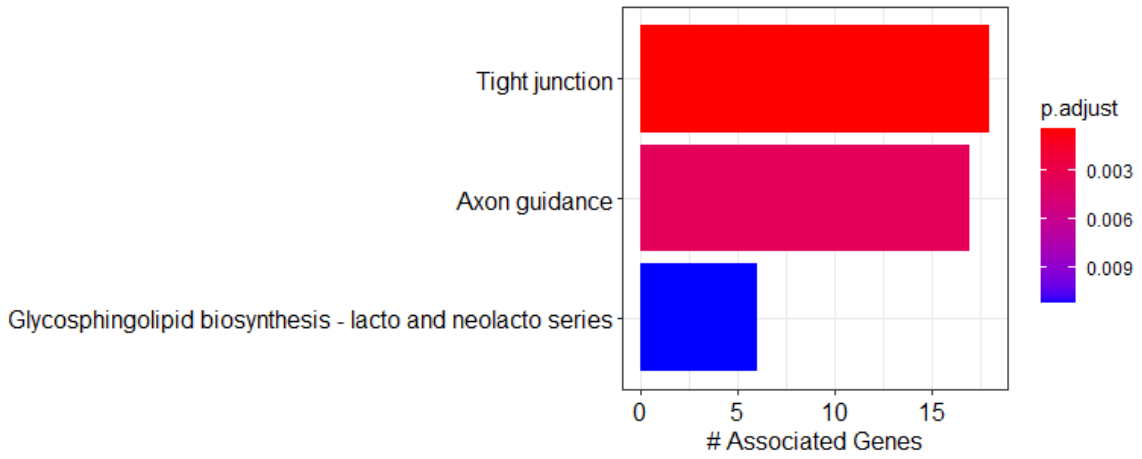

Figure 4.81. KEGG pathway annotation for cancer coexpression cluster "purple".

The KEGG pathway enrichment analysis results in Figure 4.81 show that the cluster is also somewhat enriched in glycosphingolipids.

The “lightgreen” cluster is a downregulated cluster. Disease ontology enrichment results for the cluster shows many associations with immune response.

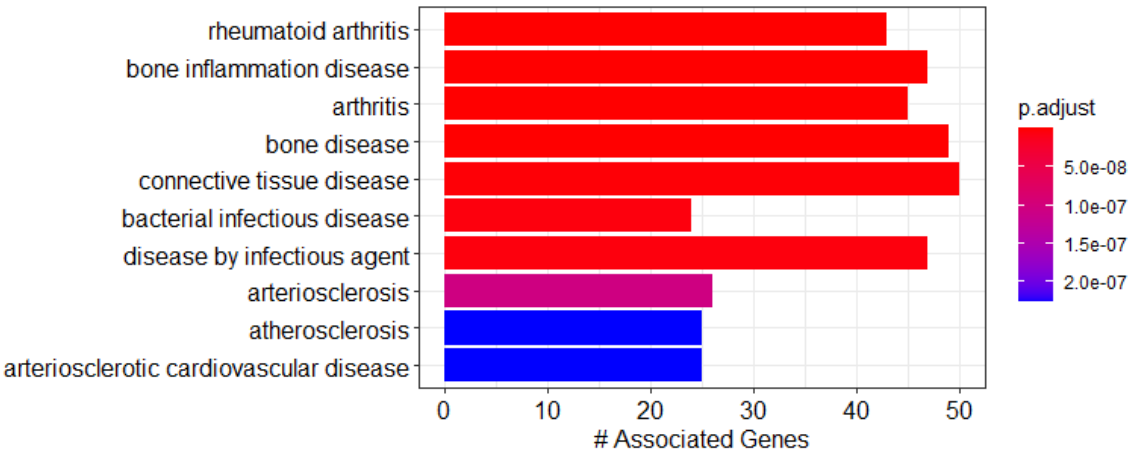

Figure 4.82. Disease ontology annotation for cancer coexpression cluster “lightgreen”.

Not pictured in Figure 4.82, but found to be a significant result, is the disease ontology annotations the “lightgreen” cluster was found to be enriched in titled “hematologic cancer” and “immune cancer,” with adjusted p-values between 10^{-3} and 10^{-4}, respectively.

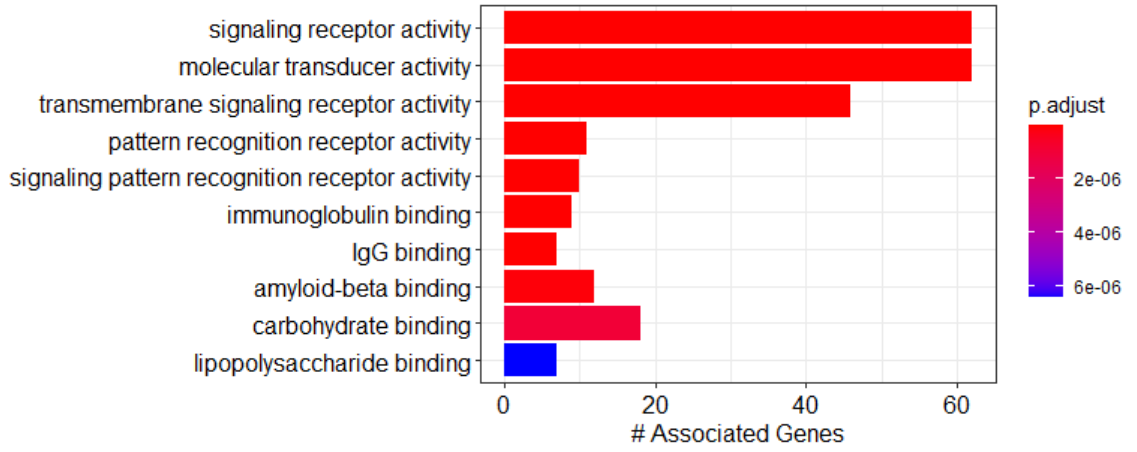

Figure 4.83. Gene Ontology: Molecular function annotation for cancer coexpression cluster “lightgreen”.

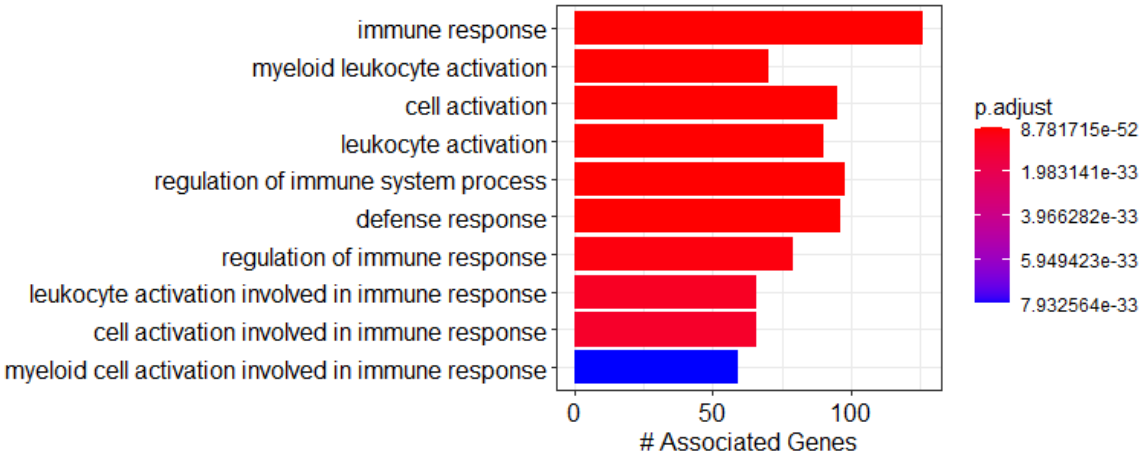

Figure 4.84. Gene Ontology: Biological process annotation for cancer coexpression cluster "lightgreen".

Figure 4.83 and Figure 4.84 contain the "lightgreen" cluster's GO term annotations for molecular function and biological process categories, respectively, and contain similarly immune response-related terms to Figure 4.82. Not pictured in Figure 4.84 but found to be significant nonetheless ($p < 10^{-14}$) were the GO biological process enrichment result that the "lightgreen" cluster was enriched in tumor necrosis factor production. The "lightgreen" cluster's KEGG pathway annotations also showed enrichment in immune response related genes.

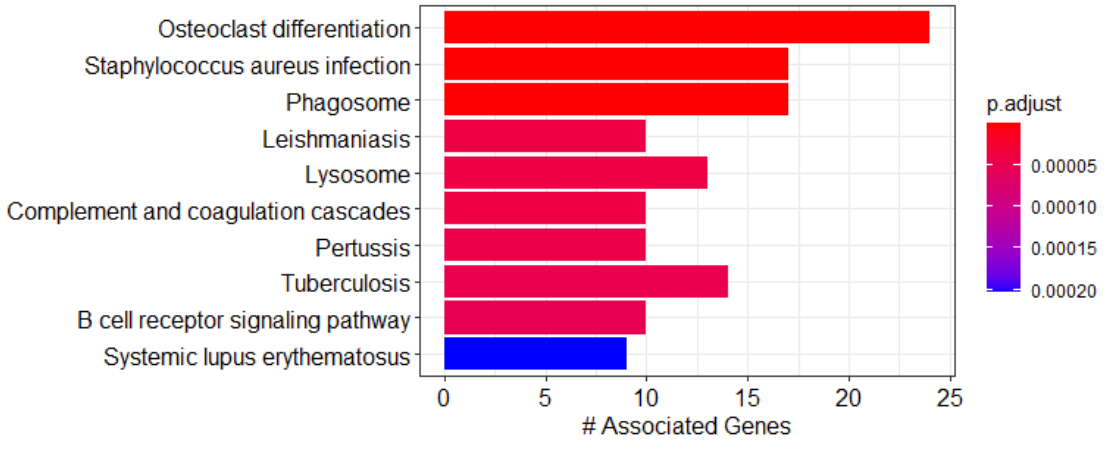

Figure 4.85. KEGG pathway annotation for cancer coexpression cluster "lightgreen".

4.4.3.3. Enrichment of Other Coexpression Clusters with Cancer-Related Enrichment Results. Several coexpression clusters where the proportion of significant genes was too low to be included in Table 4.23, nonetheless have shown clear enrichment analysis results pointing at relevant pathways. This section details the "green," "red," and "greenyellow" coexpression clusters which fit this description.

The "green" cluster is a downregulated cluster containing many genes on the inositol phosphate/PI3K pathway such as INPP5B and PIK3CD, as well as the breast cancer drug target BCL2, which was discarded as a viable drug target for pancreatic cancer due to its downregulation. This cluster is very strongly associated with immune response.

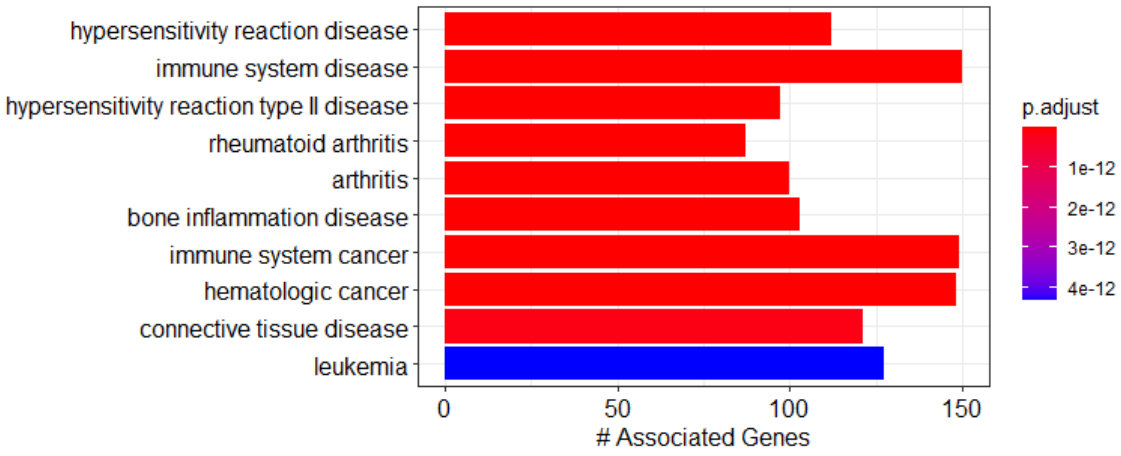

Figure 4.86. Disease ontology annotation for cancer coexpression cluster "green".

The "green" cluster is implicated in the immune response by every enrichment analysis in this study. Its enrichment in immune response disease ontology terms is apparent in Figure 4.86, but it is also enriched in immune response terms in GO molecular function, GO biological process and KEGG pathway annotation & enrichment analyses, pictured below.

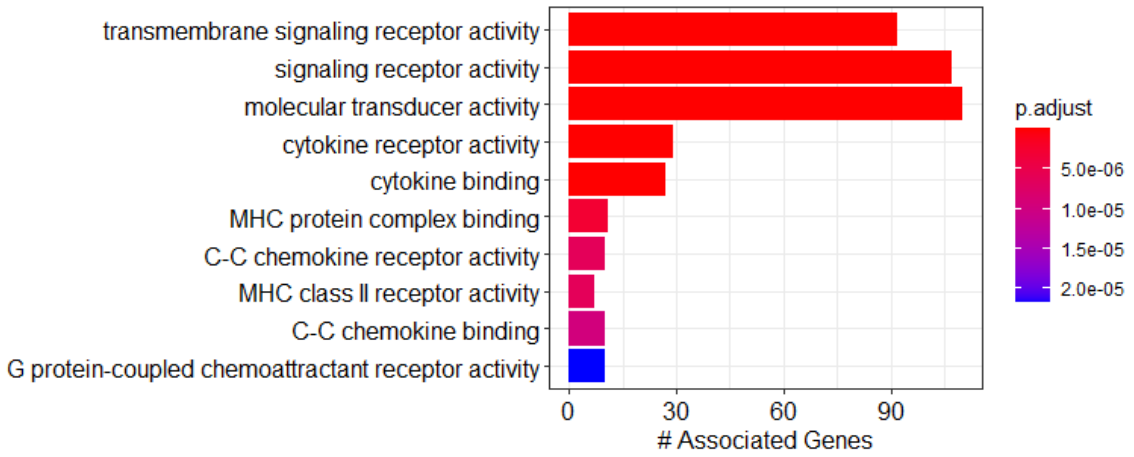

Figure 4.87. Gene Ontology: Molecular function annotation for cancer coexpression cluster “green”.

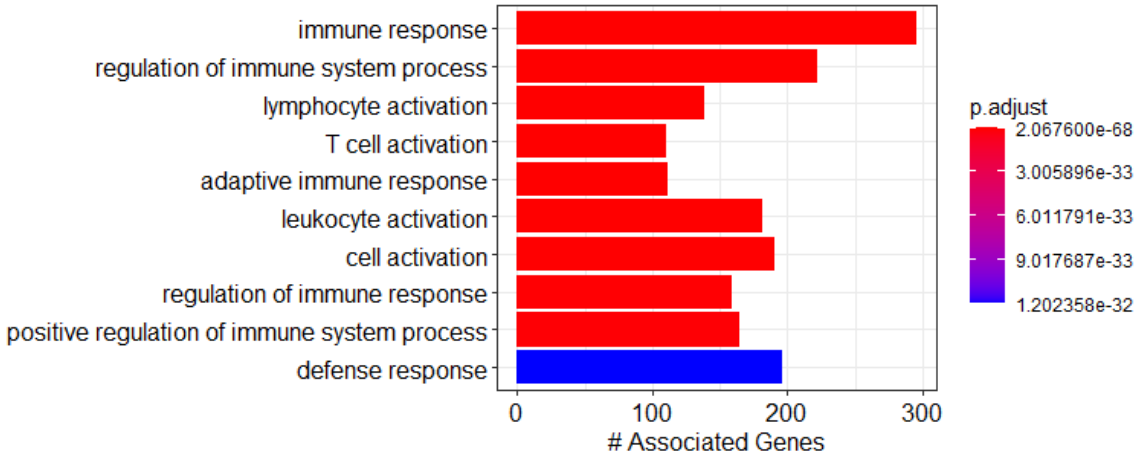

Figure 4.88. Gene Ontology: Biological process annotation for cancer coexpression cluster “green”.

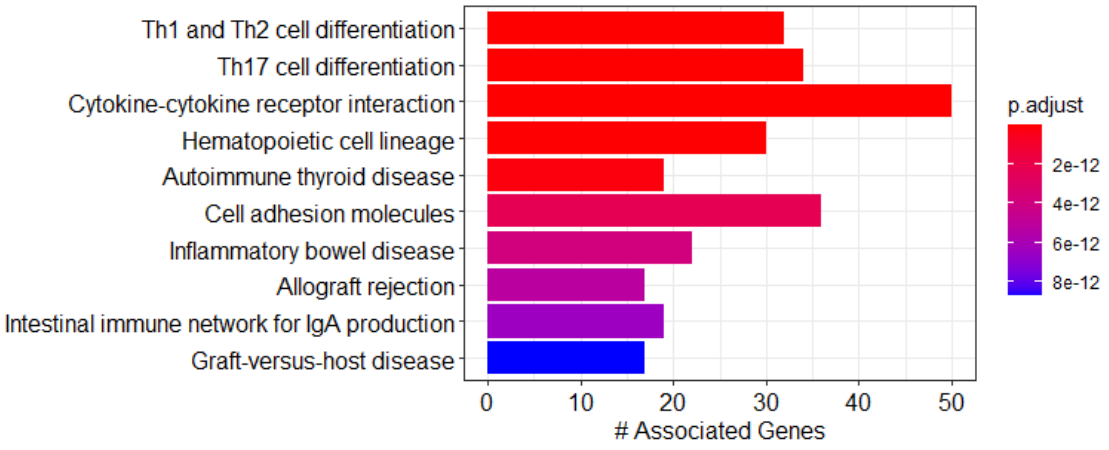

Figure 4.89. KEGG pathway annotation for cancer coexpression cluster “green”.

The “red” coexpression cluster is an overexpressed cluster that is large and contains many cell cycle related genes, as well as the previously discussed putative drug target genes DHFR and RRM2, in addition to NME6, the single lethal gene. Paradoxically, the most significant DEG in this cluster is PERP, which encodes the p53 apoptosis effector related to PMP22 and was found to be upregulated, but ordinarily has an apoptotic effect [263]. This could be due to an insufficient anti-cancer signaling response found in cancerous tissue.

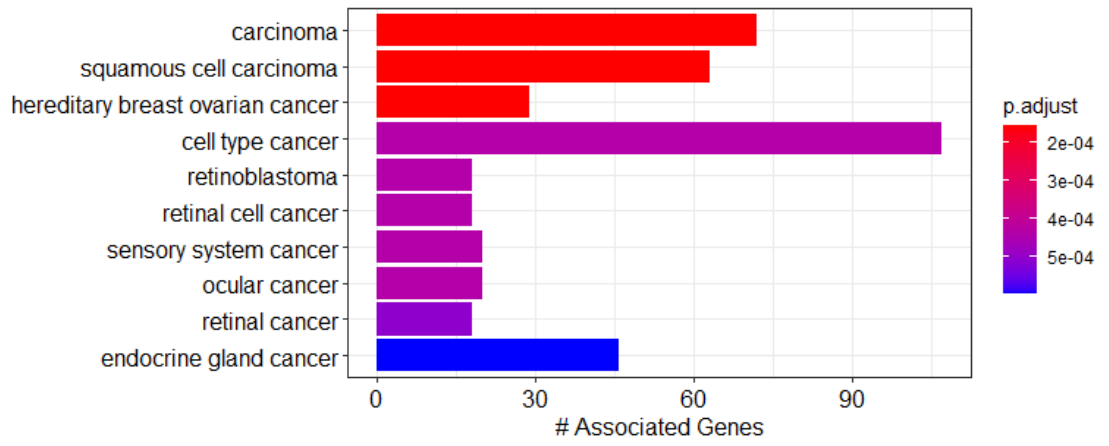

Figure 4.90. Disease ontology annotation for cancer coexpression cluster “red”.

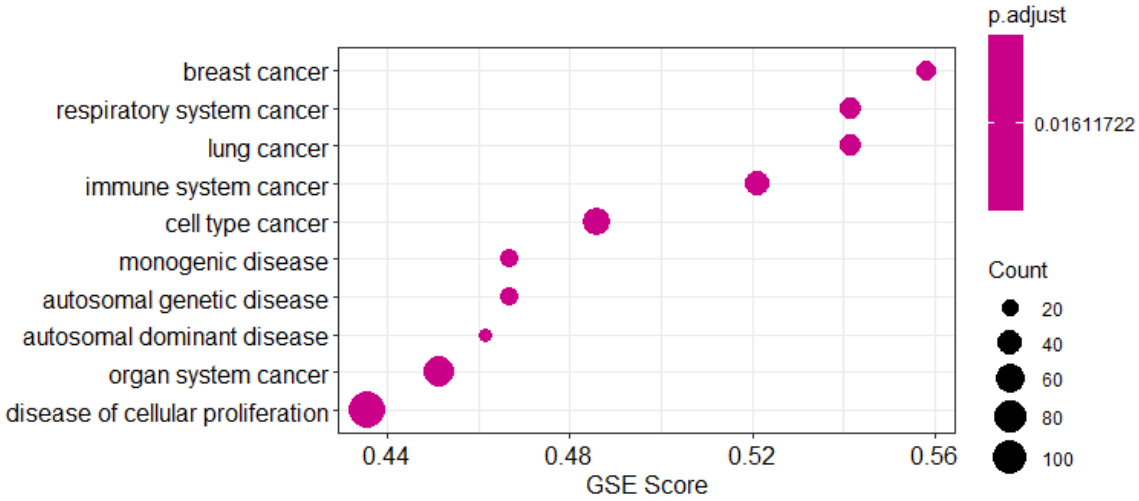

Figure 4.91. Disease ontology annotation for cancer coexpression cluster “red”, GSE.

Figure 4.90 shows that this cluster is indeed very associated with various cancers. Figure 4.91 shows the gene set enrichment analysis of the same terms in the disease ontology corpus, revealing a cluster with a much stronger breast-or-lung-cancer association than any other disease.

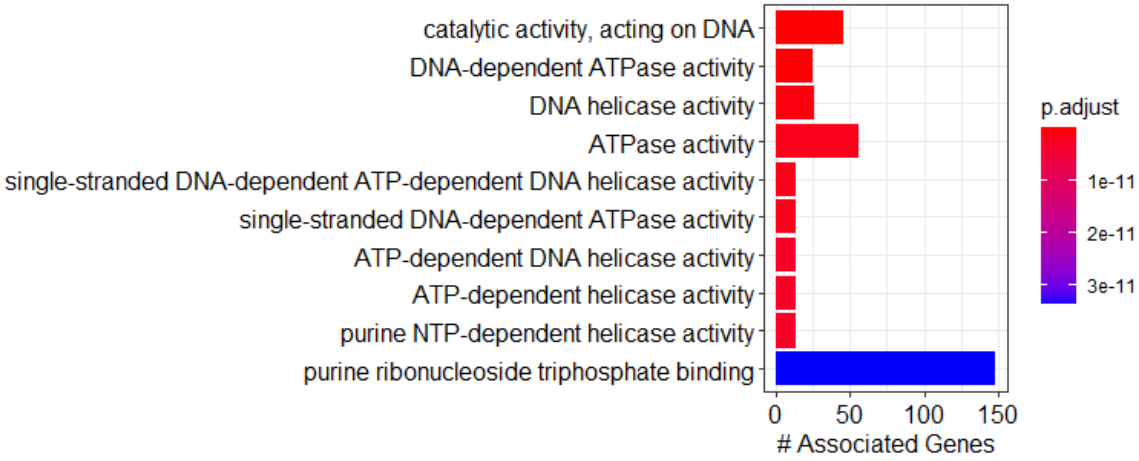

Figure 4.92. Gene Ontology: Molecular function annotation for cancer coexpression cluster “red”.

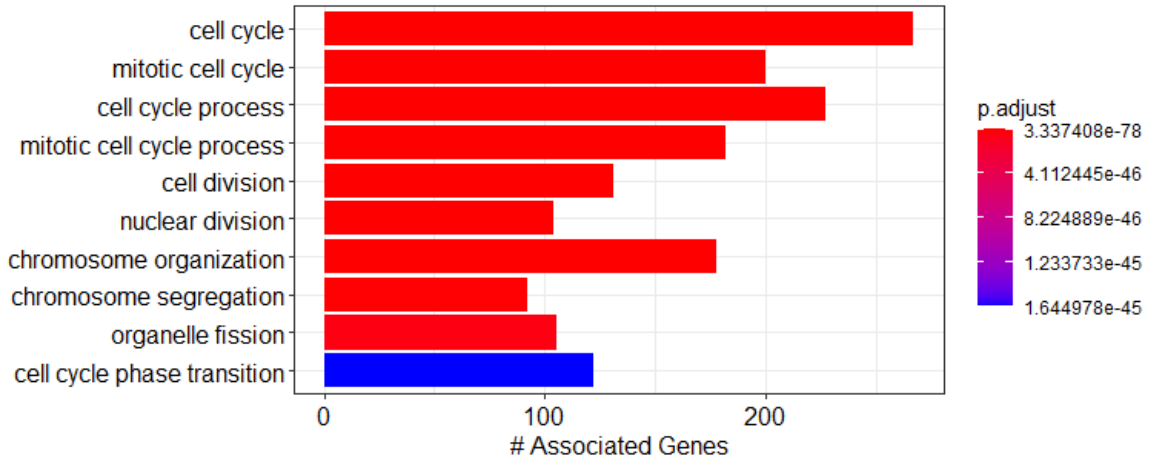

Figure 4.93. Gene Ontology: Biological process annotation for cancer coexpression cluster "red".

Figure 4.92. and Figure 4.93. show the results of GO molecular function and biological process enrichment analyses on the "red" cluster, respectively, showing that genes in this cluster have a lot of direct activity on DNA replication, especially helicase activity. The KEGG pathway enrichment analysis results for the "red" cluster also finds it strongly enriched in cell cycle-related genes.

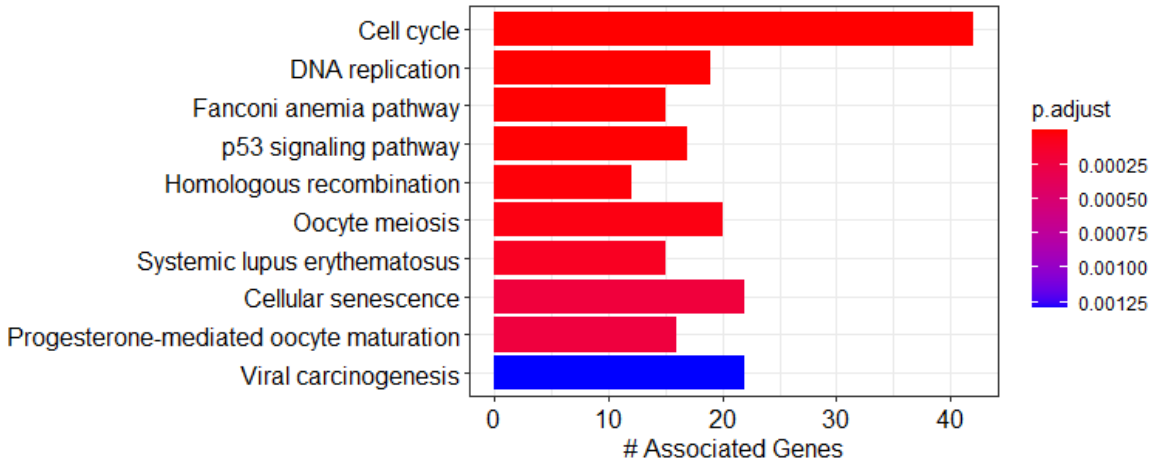

Figure 4.94. KEGG pathway annotation for cancer coexpression cluster "red".

The “greenyellow” cluster is a weakly overexpressed cluster that is mostly significant due to its strong association to cell adhesion-related genes, and its containing of the SLC6A6 taurine transporter gene discussed previously. Its disease ontology enrichment and GSE analyses reveal a host of diseases that relate the cluster to connective tissue dysfunction, cancer being among them.

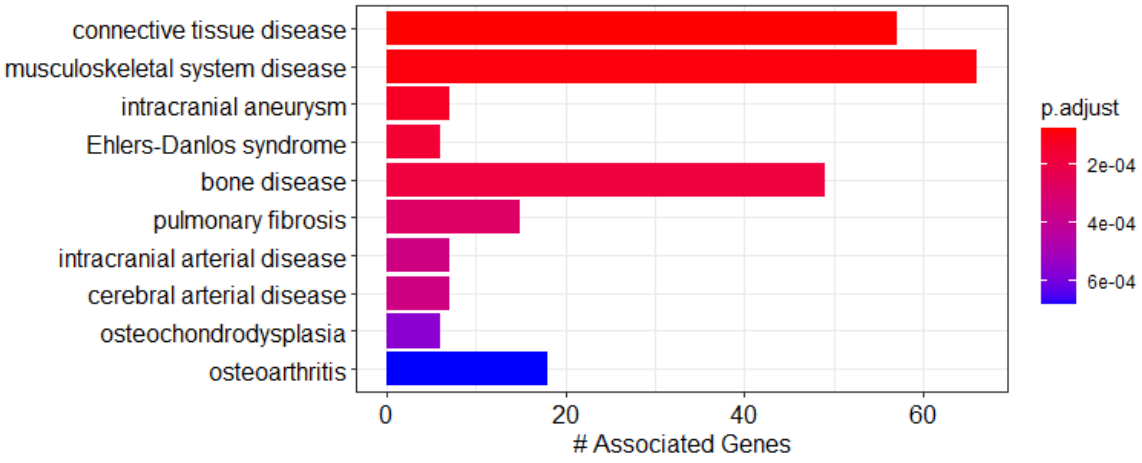

Figure 4.95. Disease ontology annotation for cancer coexpression cluster “greenyellow”.

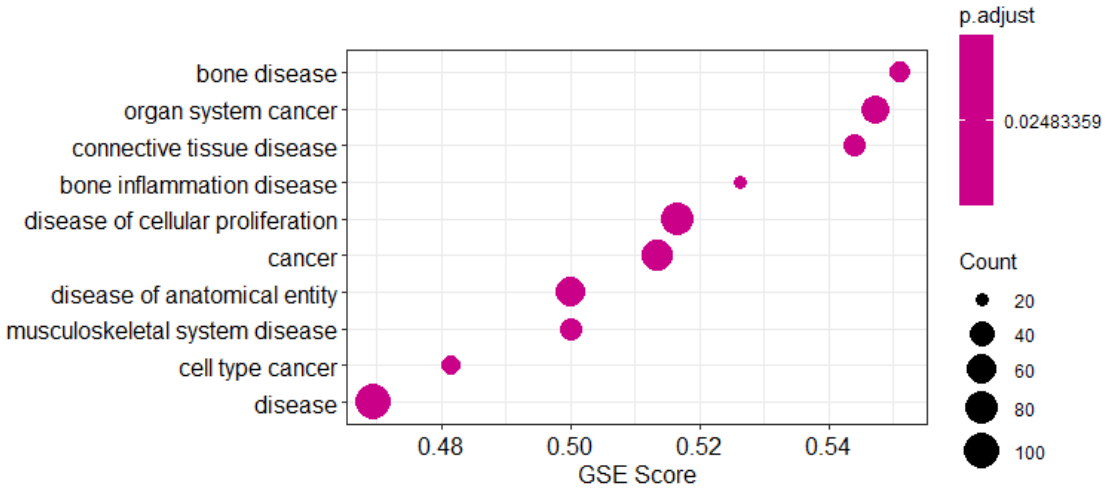

Figure 4.96. Disease ontology annotation for cancer coexpression cluster “greenyellow”, GSE.

GO molecular function and biological process enrichment analyses on the "greenyellow" cluster show enrichment in terms related to many different proteins which help anchor cells to the extracellular matrix or connect them to each other.

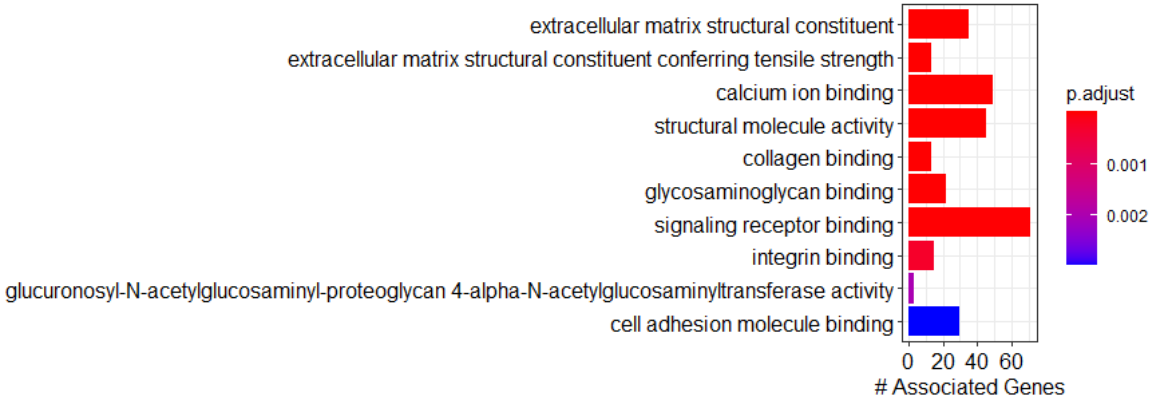

Figure 4.97. Gene Ontology: Molecular function annotation for cancer coexpression cluster "greenyellow".

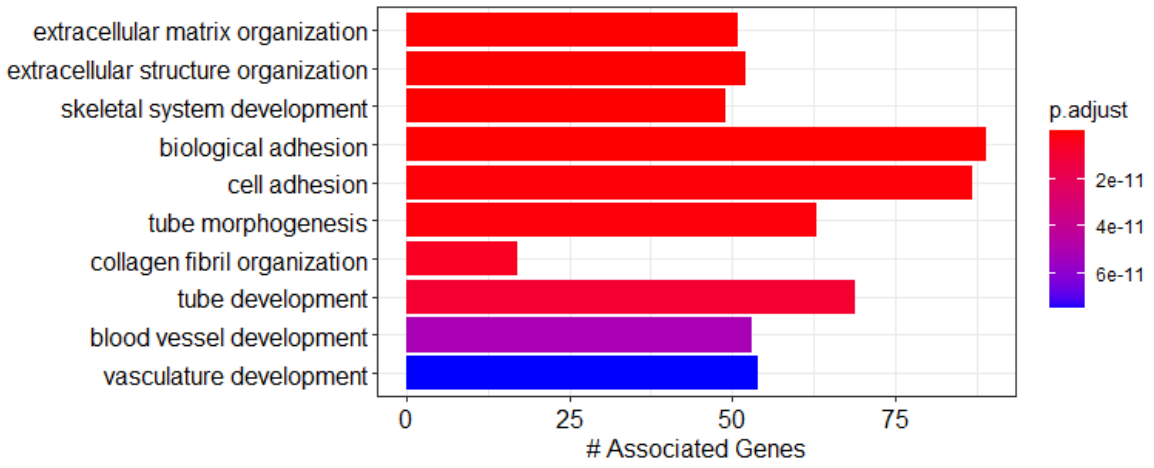

Figure 4.98. Gene Ontology: Biological process annotation for cancer coexpression cluster "greenyellow".

The KEGG pathway annotation terms for the "greenyellow" cluster are related to cellular communication with the extracellular matrix through receptor action.

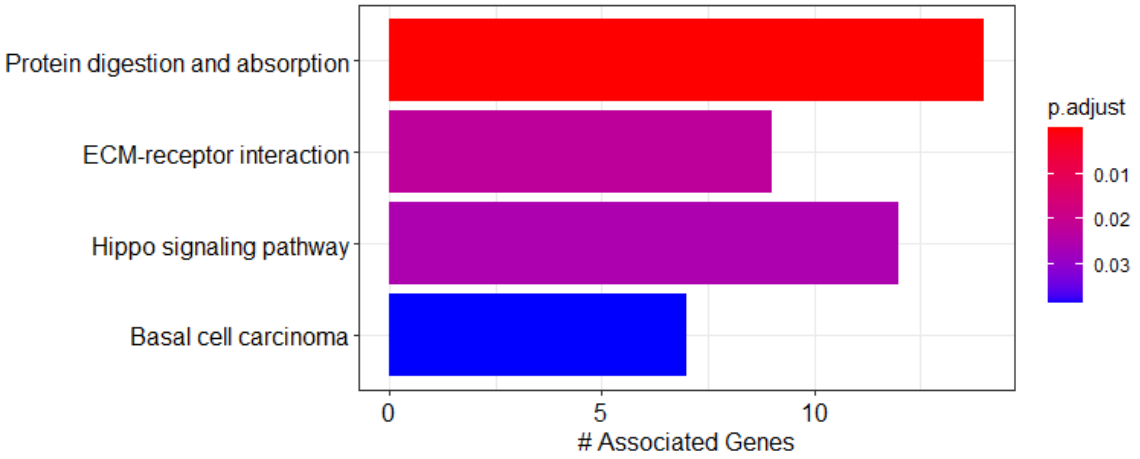

Figure 4.99. KEGG pathway annotation for cancer coexpression cluster “greenyellow”.

5. CONCLUSION & RECOMMENDATIONS

This study attempted to apply a multi-omics, network medicine approach to the discovery of biomarkers and drug targets in pancreatic cancer. Such a multi-perspective approach was found necessary, not only due to the lethality and poor diagnostic and prognostic prospects of pancreatic cancer, but also due to single-method *in silico* approaches presenting various dead ends in data interpretation. Thanks to the proliferation of high throughput data and the development of more powerful computational methods, a multifaceted approach is becoming a possibility for more and more researchers.

The first approach applied to the pancreatic cancer and healthy pancreas gene expression datasets was differential expression analysis through the limma package, using the statistically very robust Bayesian linear modelling approach. This approach was successfully used to discover 826 DEGs, remaining robust to large confounding batch effects. However, these batch effects were so large that subsequent analyses, using network medicine approaches, needed to construct separate healthy and cancer networks, resulting in more computation time, and reducing the interpretability of each network. A consensus network approach would have been more powerful but was simply infeasible given the difference between the datasets used. Additionally, the conservative p-values the Bayesian model generated in response to such batch effects likely excluded some genes that should have been discovered by this line of analysis. Therefore, an ancillary conclusion of this study is that as they exist, the high throughput data need to be harmonized to use a multi-omics approach to its full potential.

Constraint-based metabolic modeling was the second approach utilized by this study. While it shares some features of other network medicine approaches like PPI and coexpression networks, the results this approach provides depend very little on network topology. This approach was used successfully to pare down the number of reactions in the overly general Recon3D human model by 29% for both the healthy and the cancer network models, and simulate flux landscapes in each that could be compared to each other statistically, giving an estimate of how different each reaction would be from each other in

real tissue. During modeling, it was found that the primary determinants of how tissue-specific models are created were the uptake/secretion profiles and biomass functions of each network. Determining these is most likely a herculean effort but is necessary for any constraint-based metabolic modeling approach. Furthermore, the demand for specificity in uptake/secretion profiles and biomass functions increases exponentially as the *in silico* experiments begin to ask more and more specific questions, whereby comparing the pancreas to a general human cell, comparing pancreatic ductal tissue to the whole pancreas, and comparing pancreatic ductal adenocarcinoma to healthy ductal tissue each require multiplicative amounts of preparation. Even the INIT algorithm that theoretically can eschew a biomass function requires a decision to be made by the individual researcher on how to define a baseline to assign weights to reactions, or else requires an external database, such as the HPA, to define these weights *a priori.* This study is a proof-of-concept of an approach whereby high throughput data from any source can be used by any researcher to find flux differences between two metabolic models, while retaining INIT's level of human-specificity and GIMME's amenability to FBA-based methods. While this approach was used to find over 4,000 statistically significant differences between the healthy and cancer models' reactions, it suffered from the same problem as differential expression analysis. The sheer number of significant results made it near-impossible to put these results in biological context.

This necessitated the use of PPI and coexpression networks, especially through clustering and annotation, to provide this needed context to prior non-network based (or at least non-topological) results. In terms of PPI networks, the first discovery was that high centrality was a poor proxy for cancer-related significance, either as a DEG or as a gene whose reactions show differential flux. Furthermore, the GO term-based filtering approach cannot create sufficiently distinct healthy and cancer PPI networks. In the presence of high throughput gene expression data, it would be far quicker and more informative to generate a network as well-connected as possible, and then prune genes with low expression levels. In contrast to PPI networks, coexpression network generation using WGCNA creates much more distinct networks, where seemingly the healthy and cancer networks have marked differences in centrality. However, the presence of cancer as a strong driving factor of gene expression in one network but not the other would call such a conclusion into question. As

mentioned before, a consensus network would have allowed the independent comparison of these topological properties. Despite such shortcomings, the PPI and coexpression networks were seen to possess an accurate scale-free topology and were successfully clustered using the MCL and WGCNA algorithms, respectively. After clustering and annotation using enrichment analysis, these networks became a powerful tool in assigning meaning to the non-network-based results from earlier. As discussed in subsection 1.2., secreted or secretion-related genes or metabolites were highlighted as more likely diagnostic, and cell-cycle related genes were highlighted as more likely prognostic. These associations required contextualization of large amounts of differential expression and COBRA reaction flux data, which the network clustering approaches provided. Therefore, it can be concluded that such a multi-omics, multi-method approach combines the statistical power of the differential expression and constraint-based modeling approaches with the power of networks to contextualize and relate information.

The genes found to indicate the most significant differences, and the terms they were found to be most enriched in between the cancer and healthy models were as follows: 826 DEGs were found, of which 21 had superlatively low adjusted p-values ($< 10^{-7}$). These DEGs were mostly enriched in genes related to cellular adhesion, cell-cell junctions, and immune response. 3 of these genes were existing drug targets for cancers of other tissues. The constraint-based metabolic models revealed 282 exchange and 1,861 metabolic reactions with significant flux differences. These 1,861 metabolic reactions were encoded by 1,095 unique genes. Those reactions with very high flux differences compared to their flux variance ($DSD \geq 1$) comprised 250 metabolic reactions, corresponding to 206 unique genes. 3 of these genes were existing drug targets for cancers of other tissues. The genes with significant flux differences that were also DEGs made up 55 genes, 2 of which were very significant DEGs ($p < 10^{-7}$) and 6 of which had reactions with very high proportional flux differences ($DSD \geq 1$). These genes were found to be enriched in genes related to inositol phosphate, creatine, and sphingolipid metabolism, as well as showing significant differences in their central carbon metabolism and response to reactive oxygen species (ROS). Considering only those very highly significant DEGs and genes which are both DEGs and show significant reaction flux differences, this study has found 74 genes with potential to be pancreatic cancer biomarkers. Looking at the PPI clustering results, NQO1 and GPX2 in PPI

cluster 17, or CERS4 in PPI cluster 12 can be implicated as most likely biomarkers of pancreatic cancer among those 74 genes. The former cluster is related to cancer's response to ROS stress, whereas the second cluster is related to sphingolipid synthesis. It is more likely that ROS stress genes are related to the secretion of lactic acid and other oxidative stress markers, making NQO1 and GPX2 likely to be diagnostic biomarkers. Sphingolipids, on the other hand, are involved in proliferation regulation, making CERS4 more likely to be a prognostic biomarker. While coexpression clusters are harder to interpret directly, they can offer a diverse range of potential targets coexpressed with an overarching cellular process. For instance, a panel of genes from the "red" coexpression cluster among these 74 genes will give a panel of SLC2A1, MBOAT2, and PERP, which are genes that encode a glucose transporter, lysophosphatidylcholine-producing enzyme, and an auxiliary tumor suppressor protein, respectively, while also being coexpressed with cancer cell cycle activity. This panel of genes are as likely to be diagnostic as prognostic, due to being secretion-or-uptake-related and cell cycle-related at the same time. In addition to this, 14 small molecule metabolites with exchange reactions having $DSD \geq 1$ were discovered, where a diminished 2-oxoadipate secretion in cancer compared to the healthy pancreas can be suggested as the highest DSD diagnostic biomarker metabolite.

This study has shown that further work in the field would be best characterized by such multi-method approaches as well, but with a narrower focus on the desired type of biomarker or drug target. A survival analysis approach can add an extra dimension of predictive power especially when searching for prognostic biomarkers, both in a constraint-based modeling approach [140] and in a coexpression network topological analysis approach [264, 265]. Drug-protein interaction networks, or other approaches seeking to map drug targeting on PPI networks can be utilized for greater focus on drug target discovery [266–268]. Once again, keeping in mind the specifications for drug targets, such studies focusing on drug target discovery instead of biomarker discovery should use multi-tissue models or data, to ensure found drug targets will be selective towards cancer cells [269, 270].

REFERENCES

1. Longnecker, D. S., "Anatomy and Histology of the Pancreas", *Pancreapedia: The Exocrine Pancreas Knowledge Base,* 2014, https://pancreapedia.org/reviews/anatomy-and-histology-of-pancreas, accessed in September 2020.

2. Hruban, R. H., *Pancreatic Cancer – World Cancer Report 2014*, Tech. Rep., World Health Organization, Geneva, 2014.

3. Edge, S. B., D. R. Byrd, C. C. Compton, A. G. Fritz and F. L. Greene, *AJCC Cancer Staging Manual*, Tech. Rep., American Joint Committee on Cancer, Chicago, IL, 2010.

4. FDA-NIH Biomarker Working Group, *BEST (Biomarkers, EndpointS, and other Tools) Resource*, Food and Drug Administration (US), Silver Spring, MD, 2016, http://www.ncbi.nlm.nih.gov/books/NBK326791, accessed in September 2020.

5. Califf, R. M., "Biomarker Definitions and Their Applications", *Experimental Biology and Medicine (Maywood)*, Vol. 243, No. 3, pp. 213–221, 2018.

6. Barabási, A.-L., N. Gulbahce and J. Loscalzo, "Network Medicine: A Network-Based Approach to Human Disease", *Nature Reviews Genetics*, Vol. 12, No. 1, pp. 56-68, 2011.

7. Millan, P. P., *Network Analysis of Protein Interaction Data: An Introduction,* https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction, 2020, accessed in September 2020.

8. Pavlopoulos, G.A., M. Secrier, C. N. Moschopoulos, T. G. Soldatos, S. Kossida, J. Aerts, R. Schneider and P. G. Bagos, "Using Graph Theory to Analyze Biological Networks", *BioData Mining*, Vol. 4, No. 1, p. 10, 2011.

9. Horvath, S. and J. Dong, “Geometric Interpretation of Gene Coexpression Network Analysis”, *PLoS Computational Biology*, Vol. 4, No. 8, 2008.

10. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, “Gene Ontology: Tool for the Unification of Biology”, *Nature Genetics*, Vol. 25, No. 1, pp. 25–29, 2000.

11. The Gene Ontology Consortium, “The Gene Ontology Resource: 20 Years and Still GOing Strong”, *Nucleic Acids Research*, Vol. 47, No. D1, pp. D330–D338, 2019.

12. Palsson, B. *Ø*., *Systems Biology: Constraint-Based Reconstruction and Analysis*, Cambridge University Press, Cambridge, 2015.

13. Distler, M., D. Aust, J. Weitz, C. Pilarsky and R. Grützmann, “Precursor Lesions for Sporadic Pancreatic Cancer: PanIN, IPMN, and MCN”, *Biomed Research International*, Vol. 2014, 2014.

14. Matthaei, H., R. D. Schulick, R. H. Hruban and A. Maitra, “Cystic Precursors to Invasive Pancreatic Cancer”, *Nature Reviews Gastroenterology & Hepatology*, Vol. 8, No. 3, pp. 141-150, 2011.

15. Brugge, W. R., G. Y. Lauwers, D. Sahani, C. Fernandez-del Castillo and A. L. Warshaw, “Cystic Neoplasms of the Pancreas”, *New England Journal of Medicine*, Vol. 351, No. 12, pp. 1218–1226, Sep. 2004.

16. Polireddy, K. and Q. Chen, “Cancer of the Pancreas: Molecular Pathways and Current Advancement in Treatment”, *Journal of Cancer*, Vol. 7, No. 11, pp. 1497–1514, 2016.

17. Iacobuzio-Donahue, C. A., “Genetic Evolution of Pancreatic Cancer: Lessons Learnt from the Pancreatic Cancer Genome Sequencing Project”, *Gut*, Vol. 61, No. 7, pp. 1085–94, 2012.

18. Hasan, S., R. Jacob, U. Manne and R. Paluri. “Advances in Pancreatic Cancer Biomarkers”, *Oncology Reviews,* Vol. 13, No. 1, 2019.

19. Kanehisa, M., “Toward Understanding the Origin and Evolution of Cellular Organisms”, *Protein Science,* Vol. 28, No. 11, pp. 1947–1951, 2019.

20. Kanehisa, M., Y. Sato, M. Furumichi, K. Morishima and M. Tanabe, “New Approach for Understanding Genome Variations in KEGG”, *Nucleic Acids Research,* Vol. 47, No. D1, pp. D590–D595, 2019.

21. Kanehisa, M. and S. Goto, “KEGG: Kyoto Encyclopedia of Genes and Genomes”, *Nucleic Acids Research,* Vol. 28, No. 1, pp. 27–30, 2000.

22. Goonetilleke, K. S. and A. K. Siriwardena, “Systematic Review of Carbohydrate Antigen (CA 19-9) as a Biochemical Marker in the Diagnosis of Pancreatic Cancer”, *European Journal of Surgical Oncology,* Vol. 33, No. 3, pp. 266–270, 2007.

23. Mehta, K. Y., H.-J. Wu, S. S. Menon, Y. Fallah, X. Zhong, N. Rizk, K. Unger, M. Mapstone, M. S. Fiandaca, H. J. Federoff and A.K. Cheema, “Metabolomic Biomarkers of Pancreatic Cancer: A Meta-Analysis Study”, *Oncotarget,* Vol. 8, No. 40, pp. 68899–68915, 2017.

24. Mayerle, J., H. Kalthoff, R. Reszka, B. Kamlage, E. Peter, B. Schniewind, S. González Maldonado, C. Pilarsky, C. D. Heidecke, P. Schatz, M. Distler, J. A. Scheiber, U. M. Mahajan, F. U. Weiss, R. Grützmann and M. M. Lerch, “Metabolic Biomarker Signature to Differentiate Pancreatic Ductal Adenocarcinoma from Chronic Pancreatitis”, *Gut,* Vol. 67, No. 1, pp. 128–137, 2018.

25. Kobayashi, T., S. Nishiumi, A. Ikeda, T. Yoshie, A. Sakai, A. Matsubara, Y. Izumi, H. Tsumura, M. Tsuda, H. Nishisaki, N. Hayashi, S. Kawano, Y. Fujiwara, H. Minami, T. Takenawa, T. Azuma and M. Yoshida, “A Novel Serum Metabolomics-Based Diagnostic Approach to Pancreatic Cancer”, *Cancer Epidemiology, Biomarkers & Prevention,* Vol. 22, No. 4, pp. 571–579, 2013.

26. Akita, H., S. A. Ritchie, I. Takemasa, H. Eguchi, E. Pastural, W. Jin, Y. Yamazaki, D. B. Goodenowe, H. Nagano, M. Monden, M. Mori and Y. Doki, "Serum Metabolite Profiling for the Detection of Pancreatic Cancer: Results of a Large Independent Validation Study", *Pancreas*, Vol. 45, No. 10, pp. 1418–1423, 2016.

27. Ritchie, S. A., B. Chitou, Q. Zheng, D. Jayashinghe, W. Jin, A. Mochizuki and D. B. Goodenowe, "Pancreatic Cancer Serum Biomarker PC-594: Diagnostic Performance and Comparison to CA19-9", *World Journal of Gastroenterology*, Vol. 21, No. 21, pp. 6604–6612, 2015.

28. Martín-Blázquez, A., C. Jimenez-Luna, C. Diaz, J. Martinez-Galán, J. Prados, F. Vicente, C. Melguizo, O. Genilloud, J. Perez del Palacio and O. Caba, "Discovery of Pancreatic Adenocarcinoma Biomarkers by Untargeted Metabolomics", *Cancers (Basel)*, Vol. 12, No. 4, 2020.

29. Itoi, T., M. Sugimoto, J. Umeda, A. Sofuni, T. Tsuchiya, S. Tsuji, R. Tanaka, R. Tonozuka, M. Honjo, F. Moriyasu, K. Kasuya, Y. Nagakawa, Y. Abe, K. Takano, S. Kawachi, M. Shimazu, T. Soga, M. Tomita and M. Sunamura, "Serum Metabolomic Profiles for Human Pancreatic Cancer Discrimination", *International Journal of Molecular Science*, Vol. 18, No. 4, 2017.

30. Ogretmen, B., "Sphingolipid Metabolism in Cancer Signalling and Therapy", *Nature Reviews Cancer*, Vol. 18, No. 1, 2018.

31. Furuya, H., Y. Shimizu and T. Kawamori, "Sphingolipids in Cancer", *Cancer and Metastasis Reviews*, Vol. 30, No. 3, pp. 567–576, 2011.

32. Zhuo, D., X. Li and F. Guan, "Biological Roles of Aberrantly Expressed Glycosphingolipids and Related Enzymes in Human Cancer Development and Progression", *Frontiers in Physiology*, Vol. 9, 2018.

33. Ponnusamy, S., M. Meyers-Needham, C. E. Senkal, S. A. Saddoughi, D. Sentelle, S. P. Selvam, A. Salas and B. Ogretmen, "Sphingolipids and Cancer: Ceramide and

Sphingosine-1-Phosphate in the Regulation of Cell Death and Drug Resistance", *Future Oncology*, Vol. 6, No. 10, pp. 1603–1624, 2010.

34. Battini, S., F. Faitot, A. Imperiale, A. E. Cicek, C. Heimburger, G. Averous, P. Bachellier and I. J. Namer, "Metabolomics Approaches in Pancreatic Adenocarcinoma: Tumor Metabolism Profiling Predicts Clinical Outcome of Patients", *BMC Medicine*, Vol. 15, 2017.

35. Fest, J., S. Vijfhuizen, J. J. Goeman, O. Veth, A. Joensuu, M. Perola, S. Männistö, E. Ness-Jensen, K. Hveem, T. Haller, N. Tonisson, K. Mikkel, A. Metspalu, C. M. van Duijn, A. Ikram, B. H. Stricker, R. Ruiter, C. H. J. van Eijck, G.-J. B. van Ommen and P. A. C. 't Hoen, "Search for Early Pancreatic Cancer Blood Biomarkers in Five European Prospective Population Biobanks Using Metabolomics", *Endocrinology*, Vol. 160, No. 7, pp. 1731–1742, 2019.

36. Diestel, R., *Graph Theory*, Springer-Verlag, Berlin Heidelberg, 5th Ed., 2017.

37. Barabási A.-L. and Z. N. Oltvai, "Network Biology: Understanding the Cell's Functional Organization", *Nature Reviews Genetics*, Vol. 5, No. 2, 2004

38. Newman, M. E. J., "A Measure of Betweenness Centrality Based on Random Walks", *Social Networks*, Vol. 27, No. 1, pp. 39–54, 2005.

39. Rokach, L. and O. Maimon, "Clustering Methods", *Data Mining and Knowledge Discovery Handbook*, pp. 321-352, Springer US, Boston, MA, 2005.

40. Zhang, B. and S. Horvath, "A General Framework for Weighted Gene Co-Expression - Network Analysis", *Statistical Applications in Genetics and Molecular Biology*, Vol. 4, 2005.

41. Yip, A. M. and S. Horvath, "The Generalized Topological Overlap Matrix for Detecting Modules in Gene Networks", *Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology*, Las Vegas, NV, 2006.

42. Van Dongen, S., *Graph Clustering by Flow Simulation*, PhD Thesis, University of Utrecht, 2000.

43. Van Dongen, S., *Performance Criteria for Graph Clustering and Markov Cluster Experiments*, Tech. Rep., Centrum voor Wiskunde en Informatica, Utrecht, 2000.

44. Bader, G. D. and C. W. Hogue, "An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks", *BMC Bioinformatics*, Vol. 4, No. 1, p. 2, 2003.

45. Li, C., Y. Zhang and X. Chen, "Heuristic Clustering Based on Centroid Learning and Cognitive Feature Capturing", *Mathematical Problems in Engineering*, 2019.

46. Brohée, S. and J. van Helden, "Evaluation of Clustering Algorithms for Protein-Protein Interaction networks", *BMC Bioinformatics*, Vol. 7, p. 488, 2006.

47. Goh, K. I., M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabási, "The Human Disease Network", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104, No. 21, pp. 8685–8690, 2007.

48. Van Dam, S., U. Võsa, A. van der Graaf, L. Franke and J. P. de Magalhães, "Gene Co-Expression Analysis for Functional Classification and Gene-Disease Predictions", *Briefings in Bioinformatics*, Vol. 19, No. 4, pp. 575–592, 2018.

49. Gillis, J. and P. Pavlidis, "'Guilt by Association' Is the Exception Rather Than the Rule in Gene Networks", *PLOS Computational Biology*, Vol. 8, No. 3, 2012.

50. Mora, A. and I. M. Donaldson, "Effects of Protein Interaction Data Integration, Representation and Reliability on the Use of Network Properties for Drug Target Prediction", *BMC Bioinformatics*, Vol. 13, No. 1, 2012.

51. Peng, Q. and N. J. Schork, "Utility of Network Integrity Methods in Therapeutic Target Identification", *Frontiers in Genetics*, Vol. 5, No. FEB, pp. 1–17, 2014.

52. Isik, Z., C. Baldow, C. V. Cannistraci and M. Schroeder, “Drug Target Prioritization by Perturbed Gene Expression and Network Information”, *Scientific Reports*, Vol. 5, pp. 1–13, 2015.

53. Dantzig, G. B. and R. Cottle, *The Basic George B. Dantzig*, Stanford University Press, Redwood City, CA, 2003.

54. Karmarkar, N., “A New Polynomial-Time Algorithm for Linear Programming”, *Combinatorica*, Vol. 4, No. 4, pp. 373–395, 1984.

55. Orth, J. D., I. Thiele and B. Ø. Palsson, “What is Flux Balance Analysis?”, *Nature Biotechnology*, Vol. 28, No. 3, pp. 245–248, 2010.

56. Heirendt, L., S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdottir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. J. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. El Assal, L. V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. B. Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. P. Oliveira, P. T. Vuong, L. P. El Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. Aragón Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Ø. Palsson, I. Thiele & R. M. T. Fleming, “Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0”, *Nature Protocols*, Vol. 14, No. 3, pp. 639–702, 2019.

57. Schilling, C. H., D. Letscher and B. Ø. Palsson, “Theory for the Systemic Definition of Metabolic Pathways and their Use in Interpreting Metabolic Function from a Pathway-Oriented Perspective”, *Journal of Theoretical Biology*, Vol. 203, No. 3, pp. 229–248, 2000.

58. Mahadevan, R. and C. H. Schilling, “The Effects of Alternate Optimal Solutions in Constraint-Based Genome-Scale Metabolic Models”, *Metabolic Engineering*, Vol. 5, No. 4, pp. 264–276, 2003.

59. Schellenberger, J., N. E. Lewis and B. Ø. Palsson, “Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models”, *Biophysical Journal*, Vol. 100, No. 3, pp. 544–553, 2011.

60. Becker, S. A. and B. O. Palsson, “Context-Specific Metabolic Networks Are Consistent with Experiments”, *PLoS Computational Biology*, Vol. 4, No. 5, p. 1000082, 2008.

61. Agren, R., S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew and J. Nielsen, “Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT”, *PLoS Computational Biology*, Vol. 8, No. 5, p. e1002518, 2012.

62. Braschi, B., P. Denny, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates and E. Bruford, “Genenames.org: The HGNC and VGNC Resources in 2019”, *Nucleic Acids Res.*, Vol. 47, No. D1, pp. D786–D792, 2019.

63. Grossman, R. L., A. P. Heath, V. Ferretti, H. E. Varmus, D. R. Lowry, W. A. Kibbe and L. M. Standt, “Toward a Shared Vision for Cancer Genomic Data”, *N. Engl. J. Med.*, Vol. 375, No. 12, pp. 1109–1112, 2016.

64. National Cancer Institute, *TCGA: The Cancer Genome Atlas*, National Institutes of Health (US), https://www.cancer.gov/tcga, accessed in February 2021.

65. Carithers, L. J., K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. DeLuca, J. Peter-Demchok, E. T. Gelfand, P. Guan, G. E. Korzeniewski, N. C. Lockhart, C. A. Rabiner, A. K. Rao, K. L. Robinson, N. V. Roche, S. J. Sawyer, A. V. Segre, C. E. Shive, A. M. Smith, L. H. Sobin, A. H. Undale, K. M. Valentino, J. Vaught, T. R. Young, H. M. Moore and the GTEx Consortium, “A Novel Approach to

High-Quality Postmortem Tissue Procurement: The GTEx Project", *Biopreservation and Biobanking*, Vol. 13, No. 5, pp. 311–319, 2015.

66. Mishra, G., M. Suresh, K. Kumaran, N. Kannabiran, S. Suresh, P. Bala, K. Shivakumar, N. Anuradha, R. Reddy, T. Madhan Raghavan, S. Menon, G. Hanumanthu, M. Gupta, S. Upendran, S. Gupta, M. Mahesh, B. Jacob, P. Mathew, P. Chatterjee, K. S. Arun, S. Sharma, K. N. Chandrika, N. Deshpande, K. Palvankar, R. Raghavnath, R. Krishnakanth, H. Karathia, B. Rekha, R. Nayak, G. Vishnupriya, H. G. M. Kumar, M. Nagini, G. S. S. Kumar, R. Jose, P. Deepthi, S. S. Mohan, T. K. B. Gandhi, H. C. Harsha, K. S. Deshpande, M. Sarker, T. S. K. Prasad and A. Pandey, "Human Protein Reference Database - 2006 Update", *Nucleic Acids Research*, Vol. 34, pp. D411–D414, 2006.

67. Peri, S., J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T. K. B. Gandhi, M. Gronborg, N. Ibarrola, N Deshpande, K. Shanker, H. N. Shivashankar, B. P. Rashmi, M. A. Ramya, Z. Zhao, K. N. Chandrika, N. Padma, H. C. Harsha, A. J. Yatish, M. P. Kavitha, M. Menezes, D. R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S. K. Anand, V. Madavan, A. Joseph, G. W. Wong, W. P. Schiemann, S. N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G. C. Blobe, C. V. Dang, J. G. N. Garcia, J. Pevsner, O. N. Jensen, P. Roepstorff, K. S. Deshpande, A. M. Chinnaiyan, A. Hamosh, A. Chakravarti and A. Pandey, "Development of Human Protein Reference Database as An Initial Platform for Approaching Systems Biology in Humans", *Genome Research*, Vol. 13, pp. 2363–2371, 2003.

68. Prasad, T. S. K., R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. Abdul Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, "Human Protein Reference Database - 2009 Update", *Nucleic Acids Research*, Vol. 37, pp. D767–D772, 2009.

69. Stark, C., B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, "BioGRID: A General Repository for Interaction Datasets.", *Nucleic acids research*, Vol. 34, pp. D535-9, 2006.

70. Orchard, S., M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni and H. Hermjakob, "The MIntAct Project—IntAct as a Common Curation Platform for 11 Molecular Interaction Databases", *Nucleic Acids Research*, Vol. 42, No. D1, pp. D358–D363, 2014.

71. Shannon, P., A. Markiel, O. Ozier, S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks", *Genome Res*, Vol. 13, No. 11, pp. 2498–2504, 2003.

72. Villaveces, J.M., R. C. Jimenez, P. Porras, N. del-Toro, M. Duesbury, M. Dumousseau, S. Orchard, H. Choi, P. Ping, N. C. Zong, M. Askenazi, B. H. Haberman and H. Hermjakob, "Merging And Scoring Molecular Interactions Utilising Existing Community Standards: Tools, Use-Cases and A Case Study.", *Database : The Journal of Biological Databases and Curation*, Vol. 2015, 2015.

73. Wang, Y., S. Zhang, F. Li, Y. Zhou, Y. Zhang, Z. Wang, R. Zhang, J. Zhu, Y. Ren, Y. Tan, C. Qin, Y. Li, X. Li, Y. Chen and F. Zhu, "Therapeutic Target Database 2020: Enriched Resource for Facilitating Research and Early Development of Targeted Therapeutics", *Nucleic Acids Research*, Vol. 48, No. D1, pp. D1031–D1041, 2020.

74. Brunk, E., S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Draeger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich, A. Prlic, A. Sastry, A. D. Danielsdottir, A. Heinken, A. Noronha, P. W. Rose, S. K. Burley, R. M. T. Fleming, J. Nielsen, I. Thiele

and B. O. Palsson, “Recon3D Enables a Three-Dimensional View of Gene Variation in Human Metabolism”, *Nature Biotechnology*, Vol. 36, No. 3, 2018.

75. Jain, M., R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A. L. Souza, R. Kafri, M. W. Kirschner, C. B. Clish and V. K. Mootha, “Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation”, *Science*, Vol. 336, No. 6084, pp. 1040–1044, 2012.

76. Opdam, S., A. Richelle, B. Kellman, S. Li, D. C. Zielinski and N. E. Lewis, “A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models”, *Cell Systems*, Vol. 4, No. 3, 2017.

77. Whitney, E. and S. R. Rolfes, *Understanding Nutrition*. Cengage Learning, Boston, MA, 2010.

78. Whitford, W. and J. Manwaring, *Lipids in Cell Culture Media*, Tech. Rep., Fisher Scientific, Waltham, MA, 2020.

79. Law, C. W., Y. Chen, W. Shi and G. K. Smyth, “voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts”, *Genome Biology*, Vol. 15, No. 2, p. R29, 2014.

80. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth, “limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies”, *Nucleic Acids Research*, Vol. 43, No. 7, pp. e47–e47, 2015.

81. Robinson, M. D., D. J. McCarthy and G. K. Smyth, “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data”, *Bioinformatics*, Vol. 26, No. 1, pp. 139–140, 2010.

82. Robinson, M. D. and A. Oshlack, “A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data”, *Genome Biology*, Vol. 11, No. 3, p. R25, 2010.

83. Vlassis, N., M. P. Pacheco and T. Sauter, “Fast Reconstruction of Compact Context-Specific Metabolic Network Models”, *PLoS Computational Biology*, Vol. 10, No. 1, p. 1003424, 2014.

84. Gurobi Optimization, Inc., *Gurobi Optimizer Version 3.0.*, Houston, TX, 2010.

85. Fritzemeier, C. J., D. Hartleb, B. Szappanos, B. Papp and M. J. Lercher, “Erroneous Energy-Generating Cycles in Published Genome Scale Metabolic Networks: Identification and Removal”, *PLoS Computational Biology*, Vol. 13, No. 4, p. e1005494, 2017.

86. Kaufman, D. E. and R. L. Smith, “Direction Choice for Accelerated Convergence in Hit-and-Run Sampling”, *Operations Research*, Vol. 46, No. 1, pp. 84–95, 1998.

87. Pratapa, A., S. Balachandran and K. Raman, “Fast-SL: An Efficient Algorithm to Identify Synthetic Lethal Sets in Metabolic Networks”, *Bioinformatics*, Vol. 31, No. 20, pp. 3299–3305, 2015.

88. Maraziotis, I. A., K. Dimitrakopoulou and A. Bezerianos, “Growing Functional Modules from a Seed Protein via Integration of Protein Interaction and Gene Expression Data.”, *BMC Bioinformatics*, Vol. 8, p. 408, 2007.

89. Carlson, M., *GO.db: A Set of Annotation Maps Describing the Entire Gene Ontology*, https://bioconductor.org/packages/release/data/annotation/html/GO.db.html, 2019, accessed in September 2020.

90. Morris J. H., L. Apeltsin, A. M. Newman, J. Baumbach, T. Wittkop, G. Su, G. D. Baker and T. E. Ferrin, “clusterMaker: A Multi-Algorithm Clustering Plugin for Cytoscape”, *BMC Bioinformatics*, Vol. 12, No. 1, p. 436, 2011.

91. Langfelder, P. and S. Horvath, “WGCNA: An R Package for Weighted Correlation Network Analysis”, *BMC Bioinformatics*, Vol. 9, No. 1, p. 559, 2008.

92. Yu, G., L.-G. Wang, Y. Han and Q.-Y. He, “clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters”, *OMICS: A Journal of Integrative Biology*, Vol. 16, No. 5, pp. 284–287, 2012.

93. Yu, G., L.-G. Wang, G.-R. Yan and Q.-Y. He, “DOSE: An R/Bioconductor Package for Disease Ontology Semantic and Enrichment Analysis”, *Bioinformatics*, Vol. 31, No. 4, pp. 608–609, 2015.

94. Korotkevich, G., V. Sukhov and A. Sergushichev, “Fast Gene Set Enrichment Analysis”, *bioRxiv*, 2019.

95. Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, “Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles”, *Proceedings of the National Academy of Sciences*, Vol. 102, No. 43, pp. 15545–15550, 2005.

96. Carlson, M., *org.Hs.eg.db: Genome Wide Annotation for Human*, https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html, 2019, accessed in September 2020.

97. Law, C., M. Alhamdoosh, S. Su, X. Dong, L. Tian, G. K. Smyth and M. E. Ritchie, *RNA-Seq Analysis is Easy as 1-2-3 with limma, Glimma and edgeR*, https://www.bioconductor.org/packages/devel/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html, 2018, accessed in September 2020.

98. Chen, Y., A. T. L. Lun and G. K. Smyth, “From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the edgeR Quasi-Likelihood Pipeline”, *F1000Research*, Vol. 5, 2016.

99. Johnson, W. E., C. Li and A. Rabinovic, “Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods”, *Biostatistics*, Vol. 8, No. 1, pp. 118–127, 2007.

100. Nygaard, V., E. A. Rødland and E. Hovig, “Methods that Remove Batch Effects While Retaining Group Differences May Lead to Exaggerated Confidence in Downstream Analyses”, *Biostatistics*, Vol. 17, No. 1, pp. 29–39, 2016.

101. Imrich, S., M. Hachmeister and O. Gires, “EpCAM and its Potential Role in Tumor-Initiating Cells”, *Cell Adhesion & Migration*, Vol. 6, No. 1, pp. 30–38, 2012.

102. Went, P. T. H., A. Lugli, S. Meier, M. Bundi, M. Mirlacher, G. Sauter and S. Dirnhofer, “Frequent EpCam Protein Expression in Human Carcinomas”, *Human Pathology*, Vol. 35, No. 1, pp. 122–128, 2004.

103. Gires, O., C. A. Klein and P. A. Baeuerle, “On the Abundance of EpCAM on Cancer Stem Cells”, *Nature Reviews Cancer*, Vol. 9, No. 2, 2009.

104. Bauerle, P. A. and O. Gires, “EpCAM (CD326) Finding its Role in Cancer”, *British Journal of Cancer*, Vol. 96, No. 3, 2007.

105. Spizzo, G., D. Fong, M. Wurm, C. Ensinger, P. Obrist, C. Hofer, G. Mazzoleni, G. Gastl and P. Went, “EpCAM Expression in Primary Tumour Tissues and Metastases: An Immunohistochemical Analysis”, *Journal of Clinical Pathology*, Vol. 64, No. 5, pp. 415–420, 2011.

106. Akita, H., H. Nagano, Y. Takeda, H. Eguchi, H. Wada, S. Kobayashi, S. Marubashi, M. Tanemura, H. Takahashi, H. Ohigashi, Y. Tomita, O. Ishikawa, M. Mori and Y. Doki, “Ep-CAM is a Significant Prognostic Factor in Pancreatic Cancer Patients by Suppressing Cell Activity”, *Oncogene*, Vol. 30, No. 31, 2011.

107. Cioffi, M., J. Dorado, P. A. Baeuerle and C. Heeschen, “EpCAM/CD3-Bispecific T-cell Engaging Antibody MT110 Eliminates Primary Human Pancreatic Cancer Stem Cells”, *Clinical Cancer Research*, Vol. 18, No. 2, pp. 465–474, 2012.

108. Salnikov, A. V., A. Groth, A. Apel, G. Kallifatidis, B. M. Beckermann, A. Khamidjanov, E. Ryschich, M. W. Büchler, I. Herr and G. Moldenhauer “Targeting of Cancer Stem Cell

Marker EpCAM by Bispecific Antibody EpCAMxCD3 Inhibits Pancreatic Carcinoma", *Journal of Cellular and Molecular Medicine,* Vol. 13, No. 9b, pp. 4023–4033, 2009.

109. Thege, F. I., T. B. Lannin, T. N. Saha, S. Tsai, M. L. Kochman, M. A. Hollingsworth, A. D. Rhim and B. J. Kirby, "Microfluidic Immunocapture of Circulating Pancreatic Cells Using Parallel EpCAM and MUC1 Capture: Characterization, Optimization and Downstream Analysis", *Lab Chip,* Vol. 14, No. 10, pp. 1775–1784, 2014.

110. Zhou, H., A. G. Telonis, Y. Jing, N. L. Xia, L. Biederman, M. Jimbo, F. Blanco, E. Londin, J. R. Brody and I. Rigoutsos, "GPRC5A is a Potential Oncogene in Pancreatic Ductal Adenocarcinoma Cells that is Upregulated by Gemcitabine with Help from HuR", *Cell Death & Disease,* Vol. 7, No. 7, 2016.

111. Jahny, E., H. Yang, B. Liu, B. Jahnke, F. Lademann, T. Knösel, P. Rümmele, R. Grützmann, D. E. Aust, C. Pilarsky and A. Denz, "The G Protein-Coupled Receptor RAI3 Is an Independent Prognostic Factor for Pancreatic Cancer Survival and Regulates Proliferation via STAT3 Phosphorylation", *PLoS ONE,* Vol. 12, No. 1, p. e0170390, 2017.

112. Liu, B., H. Yang, C. Pilarsky and G. F. Weber, "The Effect of GPRC5a on the Proliferation, Migration Ability, Chemotherapy Resistance and Phosphorylation of GSK-3β in Pancreatic Cancer", *International Journal of Molecular Sciences,* Vol. 19, No. 7, 2018.

113. Byrne, J. A., S. Maleki, J. R. Hardy, B. S. Gloss, R. Murali, J. P. Scurry, S. Fanayan, C. Emmanuel, N. F. Hacker, R. L. Sutherland, A. deFazio and P. M. O'Brien, "MAL2 and Tumor Protein D52 (TPD52) are Frequently Overexpressed in Ovarian Carcinoma, but Differentially Associated with Histological Subtype and Patient Outcome", *BMC Cancer,* Vol. 10, p. 497, 2010.

114. Bhandari, A., Y. Shen, N. Sindan, E. Xia, B. Gautam, S. Lv and X. Zhang, "MAL2 Promotes Proliferation, Migration and Invasion through Regulating Epithelial-

Mesenchymal Transition in Breast Cancer Cell Lines", *Biochemical and Biophysical Research Communications*, Vol. 504, No. 2, pp. 434–439, 2018.

115. Eguchi, D., K. Oluchida, S. Kozono, N. Ikenaga, K. Shindo, L. Cui, K. Fujiwara, S. Akagawa, T. Ohtsuka, S. Takahata, S. Tokunaga, K. Mizumoto and M. Tanaka, "MAL2 Expression Predicts Distant Metastasis and Short Survival in Pancreatic Cancer", *Surgery*, Vol. 154, No. 3, pp. 573–582, 2013.

116. Cheng, J.-M., L. Volk, D. K. M. Janaki, S. Vyakaranam, S. Ran and K. A. Rao, "Tumor Suppressor Function of Rab25 in Triple-Negative Breast Cancer", *International Journal of Cancer*, Vol. 126, No. 12, pp. 2799–2812, 2010.

117. Tong, M., K. W. Chan, J. Y. J. Bao, K. Y. Wong, J.-N. Chen, P. S. Kwan, K. H. Tang, L. Fu, Y.-R. Qin, S. Lok, X.-Y. Guan and S. Ma, "Rab25 is a Tumor Suppressor Gene with Antiangiogenic and Anti-Invasive Activities in Esophageal Squamous Cell Carcinoma", *Cancer Research*, Vol. 72, No. 22, pp. 6024–6035, 2012.

118. Wang, S., C. Hu, F. Wu and S. He, "Rab25 GTPase: Functional Roles in Cancer", *Oncotarget*, Vol. 8, No. 38, pp. 64591–64599, 2017.

119. Jeong, B. Y., K. H. Cho, K. J. Jeong, Y.-Y. Park, J. M. Kim, S. Y. Rha, C. G. Park, G. B. Mills, J.-H. Cheong and H. Y. Lee, "Rab25 Augments Cancer Cell Invasiveness through a β1 Integrin/EGFR/VEGF-A/Snail Signaling Axis and Expression of Fascin", *Experimental & Molecular Medicine*, Vol. 50, No. 1, 2018.

120. Citi, S., A. Amorosi, F. Franconi, A. Giotti and G. Zampi, "Cingulin, a Specific Protein Component of Tight Junctions, is Expressed in Normal and Neoplastic Human Epithelial Tissues.", *The American Journal of Pathology*, Vol. 138, No. 4, pp. 781–789, 1991.

121. Guillemot, L., Y. Schneider, P. Brun, I. Castagliuolo, D. Pizzuti, D. Martines, L. Jond, M. Bongiovanni and S. Citi, "Cingulin is Dispensable for Epithelial Barrier Function and Tight Junction Structure, and Plays a Role in the Control of Claudin-2 Expression and

Response to Duodenal Mucosa Injury", *Journal of Cell Science*, Vol. 125, No. 21, pp. 5005–5014, 2012.

122. Guillemot, L., E. Hammar, C. Kaister, J. Ritz, D. Caille, L. Jond, C. Bauer, P. Meda and S. Citi, "Disruption of the Cingulin Gene Does Not Prevent Tight Junction Formation but Alters Gene Expression", *Journal of Cell Science*, Vol. 117, No. 22, pp. 5245–5256, 2004.

123. Kojima, T., D. Kyuno and N. Sawada, "Targeting Claudin-4 in Human Pancreatic Cancer", *Expert Opinion on Therapeutic Targets*, Vol. 16, No. 9, pp. 881–887, 2012.

124. Neesse, A., H. Griesmann, T. M. Gress and P. Michl, "Claudin-4 as Therapeutic Target in Cancer", *Archives of Biochemistry and Biophysics*, Vol. 524, No. 1, pp. 64–70, 2012.

125. Yao, H., Z. Yang, Z. Liu, X. Miao, L. Yang, D. Li, Q. Zou, Y. Yuan, "Glypican-3 and KRT19 are Markers Associating with Metastasis and Poor Prognosis of Pancreatic Ductal Adenocarcinoma", *Cancer Biomarkers*, Vol. 17, No. 4, pp. 397–404, 2016.

126. Pandey, R., M. Zhou, S. Islam, B. Chen, N. K. Barker, P. Langlais, A. Srivastava, M. Luo, L. S. Cooke, E. Weterings and D. Mahadevan, "Carcinoembryonic Antigen Cell Adhesion Molecule 6 (CEACAM6) in Pancreatic Ductal Adenocarcinoma (PDA): An Integrative Analysis of a Novel Therapeutic Target", *Scientific Reports*, Vol. 9, No. 1, 2019.

127. Qian, H., D. Yuan, J. Bao, F. Liu, W. Zhang, X. Yang, G. Han, J. Huang, H. Sheng and H. Yu, "Increased Expression of Plakophilin 3 Is Associated with Poor Prognosis in Ovarian Cancer", *Medicine (Baltimore)*, Vol. 98, No. 10, 2019.

128. Matsumoto, K., M. Ikeda, Y. Sato, H. Kuruma, Y. Kamata, T. Nishimori, T. Tomonaga, F. Nomura, S. Egawa and M. Iwamura, "Loss of Periplakin Expression Is Associated with Pathological Stage and Cancer-Specific Survival in Patients with Urothelial Carcinoma of the Urinary Bladder", *Biomedical Research*, Vol. 35, No. 3, pp. 201–206, 2014.

129. Roth, L., S. Srivastava, M. Lindzen, A. Sas-Chen, M. Sheffer, M. Lauriola, Y. Enuka, A. Noronha, M. Mancini, S. Lavi, G. Tarcic, G. Pines, N. Nevo, O. Heyman, T. Ziv, O. M. Rueda, D. Gnocchi, E. Pikarsky, A. Admon, C. Caldas and Y. Yarden, "SILAC identifies LAD1 as a filamin-binding regulator of actin dynamics in response to EGF and a marker of aggressive breast tumors", *Science Signaling*, Vol. 11, No. 515, 2018.

130. Jeong, H. M., J. Han, S. H. Lee, H.-J. Park, H. J. Lee, J.-S. Choi, Y. M. Lee, Y.-L. Choi, Y. K. Shin and M. J. Kwon, "ESRP1 Is Overexpressed in Ovarian Cancer and Promotes Switching from Mesenchymal to Epithelial Phenotype in Ovarian Cancer", *Oncogenesis*, Vol. 6, No. 10, 2017.

131. Di Renzo, M. F., R. Poulsom, M. Olivero, P. M. Comoglio and N. R. Lemoine, "Expression of the Met/Hepatocyte Growth Factor Receptor in Human Pancreatic Cancer", *Cancer Research*, Vol. 55, No. 5, pp. 1129–1138, 1995.

132. Ebert, M., M. Yokoyama, H. Friess, M. W. Büchler and M. Korc, "Coexpression of the c-Met Proto-Oncogene and Hepatocyte Growth Factor in Human Pancreatic Cancer", *Cancer Research*, Vol. 54, No. 22, pp. 5775–5778, 1994.

133. Kim, J. H., H. S. Kim, B. J. Kim, J. Lee and H. J. Jang, "Prognostic Value of c-Met Overexpression in Pancreatic Adenocarcinoma: A Meta-Analysis", *Oncotarget*, Vol. 8, No. 42, pp. 73098–73104, 2017.

134. Modica, C., D. Tortarolo, P. M. Comoglio, C. Basilico and E. Vigna, "MET/HGF Co-Targeting in Pancreatic Cancer: A Tool to Provide Insight into the Tumor/Stroma Crosstalk", *International Journal of Molecular Science*, Vol. 19, No. 12, 2018.

135. Di Renzo, M. F., M. Olivero, A. Giacomini, H. Porte, E. Chastre, L. Mirossay, B. Nordlinger, S. Bretti, S. Bottardi and S. Giordano, "Expression of the Met/HGF Receptor in Normal and Neoplastic Human Tissues", *Oncogene*, Vol. 6, No. 11, pp. 1997–2003, 1991.

136. Di Renzo, M. F., M. Olivero, A. Giacomini, H. Porte, E. Chastre, L. Mirossay, B. Nordlinger, S. Bretti, S. Bottardi and S. Giordano, “Overexpression and Amplification of the Met/HGF Receptor Gene During the Progression of Colorectal Cancer”, *Clinical Cancer Research*, Vol. 1, No. 2, pp. 147–154, 1995.

137. Costache, M. I., M. Ioana, S. Iordache, D. Ene, C. A. Costache and A. Săftoiu, “VEGF expression in Pancreatic Cancer and Other Malignancies: A Review of the Literature”, *Romanian Journal of Internal Medicine*, Vol. 53, No. 3, pp. 199–208, 2015.

138. Uemura, M., K. Tamura, S. Chung, S. Honma, A. Okuyama, Y. Nakamura and H. Nakagawa, “Novel 5α-Steroid Reductase (SRD5A3, Type-3) is Overexpressed in Hormone-Refractory Prostate Cancer”, *Cancer Science*, Vol. 99, No. 1, pp. 81–86, 2008.

139. Yu, D. and M.-C. Hung, “Overexpression of ErbB2 in Cancer and ErbB2-Targeting Strategies”, *Oncogene*, Vol. 19, No. 53, 2000.

140. Leoncikas, V., H. Wu, L. T. Ward, A. M. Kierzek and N. J. Plant, “Generation of 2,000 Breast Cancer Metabolic Landscapes Reveals a Poor Prognosis Group with Active Serotonin Production”, *Scientific Reports*, Vol. 6, No. 1, p. 19771, 2016.

141. Dancer, J., H. Takei, J. Y. Ro and M. Lowery-Nordberg, “Coexpression of EGFR and HER-2 in Pancreatic Ductal Adenocarcinoma: A Comparative Study Using Immunohistochemistry Correlated with Gene Amplification by Fluorescencent *in situ* Hybridization”, *Oncology Reports*, Vol. 18, No. 1, pp. 151–155, 2007.

142. Harder, J., G. Ihorst, V. Heinemann, R. Hofheinz, M. Moehler, P. Buechler, G. Kloeppel, C. Röcken, M. Bitzer, S. Boeck, E. Endlicher, A. Reinacher-Schick, C. Schmoor and M. Geissler, “Multicentre Phase II Trial of Trastuzumab and Capecitabine in Patients with HER2 Overexpressing Metastatic Pancreatic Cancer”, *British Journal of Cancer*, Vol. 106, No. 6, 2012.

143. Kimura, K., T: Sawada, M. Komatsu, M. Inoue, K. Muguruma, T. Nishihara, Y. Yamashita, N. Yamada, M. Ohira and K. Hirakawa, “Antitumor Effect of Trastuzumab

for Pancreatic Cancer with High HER-2 Expression and Enhancement of Effect by Combined Therapy with Gemcitabine", *Clinical Cancer Research*, 2006.

144. Li, J., K. T. Byrne, F. Yan, T. Yamazoe, Z. Chen, T. Baslan, L. P. Richman, J. H. Lin, Y. H. Sun, A. J. Rech, D. Balli, C. A. Hay, Y. Sela, A. J. Merrell, S. M. Liudahl, N. Gordon, R. J. Norgard, S. Yuan, S. Yu, T. Chao, S. Ye, T. S. K. Eisinger-Mathason, R. B. Faryabi, J. W. Tobias, S. W. Lowe, L. M. Coussens, E. J. Wherry, R. H. Vonderheide and B. Z. Stanger, "Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy", *Immunity*, Vol. 49, No. 1, pp. 178-193.e7, 2018.

145. Zhuang, H., Z. Zhou, Z. Zhang, X. Chen, Z. Ma, S. Huang, Y. Gong, C. Zhang and B. Hou, "B3GNT3 Overexpression Promotes Tumor Progression And Inhibits Infiltration of CD8+ T Cells in Pancreatic Cancer", *Aging (Albany NY)*, Vol. 12, 2020.

146. Storz, P., "Reactive Oxygen Species in Tumor Progression", *Frontiers in Bioscience*, Vol. 10, pp. 1881–1896, 2005.

147. Storz, P., "KRas, ROS and the Initiation of Pancreatic Cancer", *Small GTPases*, Vol. 8, No. 1, pp. 38–42, 2016.

148. Zhang, L., J. Li, L. Zong, X. Chen, K. Chen, Z. Jiang, L. Nan, X. Li, W. Li, T. Shan, Q. Ma and Z. Ma, "Reactive Oxygen Species and Targeted Therapy for Pancreatic Cancer", *Oxidative Medicine and Cellular Longevity*, Vol. 2016, 2016.

149. Caterson, B. and J. Melrose, "Keratan Sulfate, a Complex Glycosaminoglycan with Unique Functional Capability", *Glycobiology*, Vol. 28, No. 4, pp. 182–206, 2018.

150. Afratis, N., C. Gialeli, D. Nikitovic, T. Tsegenidis, E. Karousou, A. D. Theocharis, M. S. Pavão, G. N. Tzanakakis and N. K. Karamanos, "Glycosaminoglycans: Key Players in Cancer Cell Biology and Treatment", *The FEBS Journal*, Vol. 279, No. 7, pp. 1177–1197, 2012.

151. Hildebrand, A., M. Romarís, L. M. Rasmussen, D. Heinegård, D. R. Twardzik, W. A. Border and E. Ruoslahti, “Interaction of the Small Interstitial Proteoglycans Biglycan, Decorin and Fibromodulin with Transforming Growth Factor Beta”, *Biochemical. Journal*, Vol. 302, pp. 527–534, 1994.

152. Iozzo, R. V. and L. Schaefer, “Proteoglycans in Health and Disease: Novel Regulatory Signaling Mechanisms Evoked by the Small Leucine-Rich Proteoglycans”, *The FEBS Journal*, Vol. 277, No. 19, pp. 3864–3875, 2010.

153. Melo, S. A., L. B. Luecke, C. Kahlert, A. F. Fernandez, S. T. Gammon, J. Kaye, V. S. LeBleu, E. A. Mittendorf, J. Weitz, N. Rahbari, C. Reissfelder, C. Pilarsky, M. F. Fraga, D. Piwnica-Worms and R. Kalluri, “Glypican1 Identifies Cancer Exosomes and Facilitates Early Detection of Cancer”, *Nature*, Vol. 523, No. 7559, pp. 177–182, 2015.

154. Radwanska, A., D. Baczynska, D. Nowak, S. Brézillon, A. Popow, F.-X. Maquart, Y. Wegrowski and M. Malicka-Blaszkiewicz, “Lumican Affects Actin Cytoskeletal Organization in Human Melanoma A375 Cells”, *Life Sciences*, Vol. 83, No. 19–20, pp. 651–660, 2008.

155. Leiphrakpam, P. D., P. P. Patil, N. Remmers, B. Swanson, P. M. Grandgenett, F. Qiu, F. Yu and P. Radhakrishnan, “Role of Keratan Sulfate Expression in Human Pancreatic Cancer Malignancy”, *Scientific Reports*, Vol. 9, No. 1, 2019.

156. Werner, E. R., N. Blau and B. Thöny, “Tetrahydrobiopterin: Biochemistry and Pathophysiology”, *Biochemical Journal*, Vol. 438, No. 3, pp. 397–414, 2011.

157. Chen, L., X. Zeng, J. Wang, S. S. Briggs, E. O'Neill, J. Li, R. Leek, D. J. Kerr, A. L. Harris and S. Cai, “Roles of Tetrahydrobiopterin in Promoting Tumor Angiogenesis”, *The American Journal of Pathology*, Vol. 177, No. 5, pp. 2671–2680, 2010.

158. Dai, Y., J. Cui, P. Gan and W. Li, “Downregulation of Tetrahydrobiopterin Inhibits Tumor Angiogenesis in BALB/C-Nu Mice with Hepatocellular Carcinoma”, *Oncology Reports*, Vol. 36, No. 2, pp. 669–675, 2016.

159. Cronin, S. J. F., C. Seehus, A. Weidinger, S. Talbot, S. Reissig, M. Seifert, Y. Pierson, E. McNeill, M. S. Longhi, B. L. Turnes, T. Kreslavsky, M. Kogler, D. Hoffmann, M. Ticevic, D. L. Scheffer, L. Tortola, D. Cikes, A. Jais, M. Rangachari, S. Rao, M. Paolino, M. Novatchkova, M. Aichinger, L. Barrett, A. Latremoliere, G. Wirnsberger, G. Lametschwandtner, M. Busslinger, S. Zicha, A. Latini, S. C. Robson, A. Waisman, N. Andrews, M. Costigan, K. M. Channon, G. Weiss, A. V. Kozlov, M. Tebbe, K. Johnsson, C. J. Woolf and J. M. Penninger, "The Metabolite BH4 Controls T Cell Proliferation in Autoimmunity and Cancer", *Nature*, Vol. 563, No. 7732, 2018.

160. Africander, D. and K.-H. Storbeck, "Steroid Metabolism in Breast Cancer: Where are We and What are We Missing?", *Molecular and Cellular Endocrinology*, Vol. 466, pp. 86–97, 2018.

161. Capper, C. P., J. M. Rae and R. J. Auchus, "The Metabolism, Analysis, and Targeting of Steroid Hormones in Breast and Prostate Cancer", *Hormones and Cancer*, Vol. 7, No. 3, pp. 149–164, 2016.

162. Miller, W. R. and J. S. O'Neill, "The Significance of Steroid Metabolism in Human Cancer", *The Journal of Steroid Biochemistry and Molecular Biology*, Vol. 37, No. 3, pp. 317–325, 1990.

163. Fernández-del Castillo, C., G. Robles-Díaz, V. Díaz-Sánchez and A. Altamirano, "Pancreatic Cancer and Androgen Metabolism: High Androstenedione and Low Testosterone Serum Levels", *Pancreas*, Vol. 5, No. 5, pp. 515–518, 1990.

164. Munkley, J., U. L. McClurg, K. E. Livermore, I. Ehrmann, B. Knight, P. Mccullagh, J. Mcgrath, M. Crundwell, L. W. Harries, H. Y. Leung, I. G. Mills, C. N. Robson, P. Rajan and D. J. Elliott, "The Cancer-Associated Cell Migration Protein TSPAN1 Is Under Control of Androgens and Its Upregulation Increases Prostate Cancer Cell Migration", *Scientific Reports*, Vol. 7, No. 1, 2017.

165. Liu, L., E. Aleksandrowicz, F. Schönsiegel, D. Gröner, N. Bauer, C. C. Nwaeburu, Z. Zhao, J. G., T. Hoppe-Tichy, E. Yefenof, T. Hackert, O. Strobel and I. Herr, "Dexamethasone Mediates Pancreatic Cancer Progression by Glucocorticoid Receptor, TGFβ and JNK/AP-1", *Cell Death & Disease*, Vol. 8, No. 10, p. e3064, 2017.

166. Chen, W. C.-Y., B. Boursi, R. Mamtani and Y.-X. Yang, "Total Serum Cholesterol and Pancreatic Cancer: A Nested Case–Control Study", *Cancer Epidemiology, Biomarkers & Prevention*, Vol. 28, No. 2, pp. 363–369, 2019.

167. Higgs, G. A., S. Moncada and J. R. Vane, "Eicosanoids in Inflammation", *Annals of Clinical Research*, Vol. 16, No. 5–6, pp. 287–299, 1984.

168. Knab, L. M., P. J. Grippo and D. J. Bentrem, "Involvement of Eicosanoids in the Pathogenesis of Pancreatic Cancer: The Roles of Cyclooxygenase-2 and 5-Lipoxygenase", *World Journal of Gastroenterology*, Vol. 20, No. 31, pp. 10729–10739, 2014.

169. Tucker, O. N., A. J. Dannenberg, E. K. Yang, F. Zhang, L. Teng, J. M. Daly, R. A. Soslow, J. L. Masferrer, B. M. Woerner, A. T. Koki and T. J. Fahey III, "Cyclooxygenase-2 Expression is Up-Regulated in Human Pancreatic Cancer", *Cancer Research*, Vol. 59, No. 5, pp. 987–990, 1999.

170. Thiel, A., J. Mrena and A. Ristimäki, "Cyclooxygenase-2 and Gastric Cancer", *Cancer and Metastasis Reviews*, Vol. 30, No. 3, pp. 387–395, 2011.

171. Wilson, K. T., S. Fu, K. S. Ramanujam and S. J. Meltzer, "Increased Expression of Inducible Nitric Oxide Synthase and Cyclooxygenase-2 in Barrett's Esophagus and Associated Adenocarcinomas", *Cancer Research*, Vol. 58, No. 14, pp. 2929–2934, 1998.

172. Eberhart, C. E., R. J. Coffey, A. Radhika, F. M. Giardiello, S. Ferrenbach and R. N. DuBois, "Up-Regulation of Cyclooxygenase 2 Gene Expression in Human Colorectal Adenomas and Adenocarcinomas", *Gastroenterology*, Vol. 107, No. 4, pp. 1183–1188, 1994.

173. Covey, T. M., K. Edes and F. A. Fitzpatrick, “Akt activation by arachidonic acid metabolism occurs via oxidation and inactivation of PTEN tumor suppressor”, *Oncogene*, Vol. 26, No. 39, pp. 5784–5792, 2007.

174. Muili, K. A., D. Wang, A. I. Orabi, S. Sarwar, Y. Luo, T. A. Javed, J. F. Eisses, S. M. Mahmood, S. Jin, V. P. Singh, M. Ananthanaravanan, G. Perides, J. A. Williams, J. D. Molkentin and S. Z. Husain, “Bile Acids Induce Pancreatic Acinar Cell Injury and Pancreatitis by Activating Calcineurin”, *Journal of Biological Chemistry*, Vol. 288, No. 1, pp. 570–580, 2013.

175. Hirota, M. and T. Shimosegawa, “Bile Acids and Pancreatic Disease”, *Bile Acids in Gastroenterology: Basic and Clinical*, pp. 169-176, Springer Japan, Tokyo, 2017.

176. Tucker, O. N., A. J. Dannenberg, E. K. Yang and T. J. Fahey III, “Bile Acids Induce Cyclooxygenase-2 Expression in Human Pancreatic Cancer Cell Lines”, *Carcinogenesis*, Vol. 25, No. 3, pp. 419–423, 2004.

177. Ridlon, J. M., P. G. Wolf and H. R. Gaskins, “Taurocholic Acid Metabolism by Gut Microbes and Colon Cancer”, *Gut Microbes*, Vol. 7, No. 3, pp. 201–215, 2016.

178. Wolf, P. G., H. R. Gaskins, J. M. Ridlon, S. Freels, A. Hamm, S. Goldberg, P. Petrilli, T. Schering, S. Vergis, S. Gomez-Perez, C. Yazici, C. Braunschweig, E. Mutlu and L. Tussing-Humphreys, “Effects of Taurocholic Acid Metabolism by Gut Bacteria: A Controlled Feeding Trial in Adult African American Subjects at Elevated Risk for Colorectal Cancer”, *Contemporary Clinical Trials Communications*, Vol. 19, p. 100611, 2020.

179. Zeng, H., S. Umar, B. Rust, D. Lazarova and M. Bordonaro, “Secondary Bile Acids and Short Chain Fatty Acids in the Colon: A Focus on Colonic Microbiome, Cell Proliferation, Inflammation, and Cancer”, *International Journal of Molecular Sciences*, Vol. 20, No. 5, 2019.

180. Cherbonnel-Lasserre, C. L., G. Linares-Cruz, J. P. Rigaut, L. Sabatier and B. Dutrillaux, "Strong Decrease in Biotin Content May Correlate with Metabolic Alterations in Colorectal Adenocarcinoma", *The International Journal of Cancer*, Vol. 72, No. 5, pp. 768–775, 1997.

181. Ballantyne, R. M. and E. W. McHenry, "Vitamin B6 and Biotin in Human Cancer Tissue", *Cancer Research*, Vol. 9, No. 11, pp. 689–691, 1949.

182. Yang, W., Y. Cheng, T. Xu, X. Wang and L. Wen, "Targeting Cancer Cells with Biotin–Dendrimer Conjugates", *European Journal of Medicinal Chemistry*, Vol. 44, No. 2, pp. 862–868, 2009.

183. Muhammad, N., N. Sadia, C. Zhu, C. Luo, Z. Guo and X. Wang, "Biotin-Tagged Platinum(IV) Complexes as Targeted Cytostatic Agents against Breast Cancer Cells", *Chemical Communications.*, Vol. 53, No. 72, pp. 9971–9974, 2017.

184. Page, G. L. J., D. Laight and M. H. Cummings, "Thiamine Deficiency in Diabetes Mellitus and the Impact of Thiamine Replacement on Glucose Metabolism and Vascular Disease", *International Journal of Clinical Practice*, Vol. 65, No. 6, pp. 684–690, 2011.

185. Isenberg-Grzeda, E., B. Chabon and S. E. Nicolson, "Prescribing Thiamine to Inpatients with Alcohol Use Disorders: How Well Are We Doing?", *Journal of Addiction Medicine*, Vol. 8, No. 1, pp. 1–5, 2014.

186. Subramanya, S. B., V. S. Subramanian, V. T. Sekar and H. M. Said, "Thiamine Uptake by Pancreatic Acinar Cells: Effect of Chronic Alcohol Feeding/Exposure", *The American Journal of Physiology-Gastrointestinal and Liver Physiology*, Vol. 301, No. 5, pp. G896–G904, 2011.

187. Comin-Anduix, B., J. Boren, S. Martinez, C. Moro, J. J. Centelles, R. Trebukhina, N. Petushok, W. N. Lee, L. G. Boros and M. Cascante, "The Effect of Thiamine Supplementation on Tumour Proliferation", *European Journal of Biochemistry*, Vol. 268, No. 15, pp. 4177–4182, 2001.

188. Bailey, L. B. and J. F. Gregory, “Folate Metabolism and Requirements”, *The Journal of Nutrition*, Vol. 129, No. 4, pp. 779–782, 1999.

189. Rosenzweig, A., J. Blenis and A. P. Gomes, “Beyond the Warburg Effect: How Do Cancer Cells Regulate One-Carbon Metabolism?”, *Frontiers in Cell and Developmental Biology*, Vol. 6, 2018.

190. Newman, A. C. and O. D. K. Maddocks, “One-Carbon Metabolism in Cancer”, *British Journal of Cancer*, Vol. 116, No. 12, pp. 1499–1504, 2017.

191. Schweitzer, B. I., A. P. Dicker and J. R. Bertino, “Dihydrofolate Reductase as a Therapeutic Target”, *The FASEB Journal*, Vol. 4, No. 8, pp. 2441–2452, 1990.

192. Koseki, J., M. Konno, A. Asai, H. Colvin, K. Kawamoto, N. Nishida, D. Sakai, T. Kudo, T. Satoh, Y. Doki, M. Mori and H. Ishii, “Enzymes of the One-Carbon Folate Metabolism as Anticancer Targets Predicted by Survival Rate Analysis”, *Scientific Reports*, Vol. 8, 2018.

193. Chittiboyina, S., Z. Chen, E. G. Chiorean, L. M. Kamendulis and B. A. Hocevar, “The Role of the Folate Pathway in Pancreatic Cancer Risk”, *PLoS One*, Vol. 13, No. 2, 2018.

194. Nie, F., M. Yu, K. Zhang, L. Yang, Q. Zhang, S. Liu, M. Liu, M. Shang, F. Zeng and W. Liu, “Association of MTHFR Gene Polymorphisms with Pancreatic Cancer: Meta-Analysis of 17 Case–Control Studies”, *International Journal of Clinical Oncology*, Vol. 25, No. 2, pp. 312–321, 2020.

195. Liu, X.-M., F.-H. Liu, Y. Tang and Q. Li, “MTHFR C677T Polymorphism and Pancreatic Cancer Risk: A Meta-Analysis”, *Asian Pacific Journal of Cancer Prevention*, Vol. 13, No. 8, pp. 3763–3766, 2012.

196. Tan, J., C.-Y. Yu, Z.-H. Wang, H.-Y. Chen, J. Guan, Y.-X. Chen and J.-Y. Fang, “Genetic Variants in the Inositol Phosphate Metabolism Pathway and Risk of Different Types of Cancer”, *Scientific Reports*, Vol. 5, No. 1, 2015.

197. Assmann, A., C. Hinault and R. N. Kulkarni, “Growth Factor Control of Pancreatic Islet Regeneration and Function”, *Pediatric Diabetes*, Vol. 10, No. 1, pp. 14–32, 2009.

198. Vucenik, I., “Anticancer Properties of Inositol Hexaphosphate and Inositol: An Overview”, *Journal of Nutritional Science and Vitaminology.*, Vol. 65, No. Supplement, pp. S18–S22, 2019.

199. Vucenik, I. and A. M. Shamsuddin, “Cancer Inhibition by Inositol Hexaphosphate (IP6) and Inositol: From Laboratory to Clinic”, *The Journal of Nutrition*, Vol. 133, No. 11, pp. 3778S-3784S, 2003.

200. Rizvi, I., D. R. Riggs, B. J. Jackson, A. Ng, C. Cunningham and D. W. Mcfadden, “Inositol Hexaphosphate (IP6) Inhibits Cellular Proliferation in Melanoma”, *Journal of Surgical Research*, Vol. 133, No. 1, pp. 3–6, 2006.

201. Somasundar, P., D. R. Riggs, B. J. Jackson, C. Cunningham, L. Vona-Davis and D. W. McFadden, “Inositol Hexaphosphate (IP6): A Novel Treatment for Pancreatic Cancer”, *Journal of Surgical Research*, Vol. 126, No. 2, pp. 199–203, 2005.

202. Koundouros, N. and G. Poulogiannis, “Reprogramming of Fatty Acid Metabolism in Cancer”, *British Journal of Cancer*, Vol. 122, No. 1, 2020.

203. Nikas, I. P., S. A. Paschou and H. S. Ryu, “The Role of Nicotinamide in Cancer Chemoprevention and Therapy”, *Biomolecules*, Vol. 10, No. 3, 2020.

204. Sun, B. L., X. Sun, N. Casanova, A. N. Garcia, R. Oita, A. M. Algotar, S. M. Camp, V. R. Hernon, T. Gregory, A. E. Cress and J. G. N. Garcia, “Role of Secreted Extracellular Nicotinamide Phosphoribosyltransferase (eNAMPT) in Prostate Cancer Progression: Novel Biomarker and Therapeutic Target”, *EBioMedicine*, Vol. 61, 2020.

205. Guo, Q., N. Han, L. Shi, L. Yang, X. Zhang, Y. Zhou, S. Yu and M. Zhang, “NAMPT: A Potential Prognostic and Therapeutic Biomarker in Patients with Glioblastoma”, *Oncology Reports*, Vol. 42, No. 3, pp. 963–972, 2019.

206. Chini, C. C. S., A. M. Gonzalez Guerrico, V. Nin, J. Camacho-Pereira, C. Escande, M. T. Barbosa and E. N. Chini, “Targeting of NAD Metabolism in Pancreatic Cancer Cells: Potential Novel Therapy for Pancreatic Tumors”, *Clinical Cancer Research*, Vol. 20, No. 1, pp. 120–130, 2014.

207. Mowat, C. G., “Role of Kynurenine Pathway in Cancer Biology”, *Targeting the Broadly Pathogenic Kynurenine Pathway*, pp. 273–286, Springer International Publishing, Cham, 2015.

208. Prodinger, J., L. J. Loacker, R. L. J. Schmidt, F. Ratzinger, G. Greiner, N. Witzeneder, G. Hoermann, S. Jutz, W. F. Pickl, P. Steinberger, R. Marculescu and K. G. Schmetterer, “The Tryptophan Metabolite Picolinic Acid Suppresses Proliferation and Metabolic Activity of CD4+ T Cells and Inhibits c-Myc Activation”, *Journal of Leukocyte Biology*, Vol. 99, No. 4, pp. 583–594, 2016.

209. Nevler, A., A. J. Muller, E. Sutanto-Ward, J. B. DuHadaway, K. Nagatomo, E. Londin, K. O'Hayer, J. A. Cozzitorto, H. Lavu, T. P. Yeo, M. Curtis, T. Villatoro, B. E. Leiby, L. Mandik-Nayak, J. M. Winter, C. J. Yeo, G. C. Prendergast and J. R. Brody, “Host IDO2 Gene Status Influences Tumor Progression and Radiotherapy Response in KRAS-Driven Sporadic Pancreatic Cancers”, *Clinical Cancer Research*, Vol. 25, No. 2, pp. 724–734, 2019.

210. Heng, B., A. A. Bilgin, D. B. Lovejoy, V. X. Tan, H. H. Milioli, L. Gluch, S. Bustamante, T. Sabaretnam, P. Moscato, C. K. Lim and G. J. Guillemin, “Differential Kynurenine Pathway Metabolism in Highly Metastatic Aggressive Breast Cancer Subtypes: Beyond IDO1-Induced Immunosuppression”, *Breast Cancer Research*, Vol. 22, No. 1, p. 113, 2020.

211. Hornigold, N., K. R. Dunn, R. A. Craven, A. Zougman, S. Trainor, R. Shreeve, J. Brown, H. Sewell, M. Shires, M. Knowles, T. Fukuwatari, E. R. Maher, J. Burns, S. Bhattarai, M. Menon, A. Brazma, G. Scelo, L. Feulner, Y. Riazalhosseini, M. Lathrop, A. Harris, P. J. Selby, R. E. Banks and N. S. Vasudev, “Dysregulation at Multiple Points of the

Kynurenine Pathway Is a Ubiquitous Feature of Renal Cancer: Implications for Tumour Immune Evasion", *British Journal of Cancer*, Vol. 123, No. 1, 2020.

212. Huang, J. Y., L. M. Butler, Ø. Midttun, A. Ulvik, R. Wang, A. Jin, Y.-T. Gaou, P. M. Ueland, W.-P. Koh and J.-M. Yuan, "A Prospective Evaluation of Serum Kynurenine Metabolites and Risk of Pancreatic Cancer", *PLoS One*, Vol. 13, No. 5, p. e0196465, 2018.

213. Nam, H., B. C. Chung, Y. Kim, K. Lee and D. Lee, "Combining Tissue Transcriptomics and Urine Metabolomics for Breast Cancer Biomarker Identification", *Bioinformatics*, Vol. 25, No. 23, pp. 3151–3157, 2009.

214. Goldman, A., A. Condon, E. Adler, M. Minnella, C. Bernstein, H. Bernstein and K. Dvorak, "Protective Effects of Glycoursodeoxycholic Acid in Barrett's Esophagus Cells", *Diseases of the Esophagus*, Vol. 23, No. 2, pp. 83–93, 2010.

215. Rahn, S., P. D. Barbosa, J. L. Möller, N. Ammar, T. Demetrowitsch, O. Helm, D. Wesch, B. Sipos, C. Röcken, K. Schwarz, H. Schaefer and S. Sebens, "Inflammation Associated Pancreatic Tumorigenesis: Upregulation of Succinate Dehydrogenase (Subunit B) Reduces Cell Growth of Pancreatic Ductal Epithelial Cells", *Cancers*, Vol. 12, No. 1, 2020.

216. Wu, J.-Y., T.-W. Huang, Y.-T. Hsieh, Y.-F. Wang, C.-C. Yen, G.-L. Lee, C.-C. Yeh, Y.-J. Peng, Y.-Y. Kuo, H.-T. Wen, H.-C. Lin, C.-W. Hsiao, K. K. Wu, H.-J. Kung, Y.-J. Hsu and C.-C. Kuo, "Cancer-Derived Succinate Promotes Macrophage Polarization and Cancer Metastasis via Succinate Receptor", *Molecular Cell*, Vol. 77, No. 2, pp. 213-227, 2020.

217. Melone, M. A. B., A. Valentino, S. Margarucci, U. Galderisi, A. Giordano and G. Peluso, "The Carnitine System and Cancer Metabolic Plasticity", *Cell Death & Disease*, Vol. 9, No. 2, 2018.

218. Shlomi, T., T. Benyamini, E. Gottlieb, R. Sharan and E. Ruppin, “Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect”, *PLoS Computational Biology*, Vol. 7, No. 3, p. 1002018, 2011.

219. Warburg, O., “The Metabolism of Carcinoma Cells”, *The Journal of Cancer Research*, Vol. 9, No. 1, pp. 148–163, 1925.

220. Fantin, V. R., J. St-Pierre and P. Leder, “Attenuation of LDH-A Expression Uncovers a Link between Glycolysis, Mitochondrial Physiology, and Tumor Maintenance”, *Cancer Cell*, Vol. 9, No. 6, pp. 425–434, 2006.

221. Liberti, M. V. and J. W. Locasale, “The Warburg Effect: How Does it Benefit Cancer Cells?”, *Trends in Biochemical Sciences*, Vol. 41, No. 3, pp. 211–218, 2016.

222. Marín-Hernández, Á., J. S. Rodríguez-Zavala, I. Del Mazo-Monsalvo, S. Rodríguez-Enríquez, R. Moreno-Sánchez and E. Saavedra, “Inhibition of Non-flux-Controlling Enzymes Deters Cancer Glycolysis by Accumulation of Regulatory Metabolites of Controlling Steps”, *Frontiers in Physiology*, Vol. 7, 2016.

223. Orozco, J. M., P. A. Krawczyk, S. M. Scaria, A. L. Cangelosi, S. H. Chan, T. Kunchok, C. A. Lewis and D. M. Sabatini, “Dihydroxyacetone Phosphate Signals Glucose Availability to mTORC1”, *Nature Metabolism*, Vol. 2, No. 9, 2020.

224. Saxton, R. A. and D. M. Sabatini, “mTOR Signaling in Growth, Metabolism, and Disease”, *Cell*, Vol. 168, No. 6, pp. 960–976, 2017.

225. Hassan, Z., C. Schneeweis, M. Wirth, C. Veltkamp, Z. Dantes, B. Feuerecker, G. O. Ceyhan, S. K. Knauer, W. Weichert, R. M. Schmid, R. Stauber, A. Arlt, O. H. Krämer, R. Rad, M. Reichert, D. Saur and G. Schneider, “MTOR Inhibitor-Based Combination Therapies for Pancreatic Cancer”, *British Journal of Cancer*, Vol. 118, No. 3, 2018.

226. Hua, H., Q. Kong, H. Zhang, J. Wang, T. Luo and Y. Jiang, “Targeting mTOR for Cancer Therapy”, *Journal of Hematology & Oncology*, Vol. 12, No. 1, p. 71, 2019.

227. Morran, D. C., J. Wu, N. B. Jamieson, A. Mrowinska, G. Kalna, S. A. Karim, A. Y. M. Au, C. J. Scarlett, D. K. Chang, M. Z. Pajak, Australian Pancreatic Cancer Genome Initiative (APGI), K. A. Oien, C. J. McKay, C. R. Carter, G. Gillen, S. Champion, S. L. Pimlott, K. I. Anderson, T. R. J. Evans, S. M. Grimmond, A. V. Biankin, O. J. Sansom and J. P. Morton, "Targeting mTOR Dependency in Pancreatic Cancer", *Gut*, Vol. 63, No. 9, pp. 1481–1489, 2014.

228. Oh, E.-T. and H. J. Park, "Implications of NQO1 in Cancer Therapy", *BMB Reports*, Vol. 48, No. 11, pp. 609–617, 2015.

229. Awadallah, N. S., D. Dehn, R. J. Shah, S. R. Nash, Y. K. Chen, D. Ross, J. S. Bentz, K. R. Shroyer, "NQO1 Expression in Pancreatic Cancer and Its Potential Use as a Biomarker", *Applied Immunohistochemistry & Molecular Morphology*, Vol. 16, No. 1, pp. 24–31, 2008.

230. Lewis, A. M., M. Ough, J. Du, M.-S. Tsao, L. W. Oberley and J. J. Cullen, "Targeting NAD(P)H:Quinone Oxidoreductase (NQO1) in Pancreatic Cancer", *Molecular Carcinogenesis*, Vol. 56, No. 7, pp. 1825–1834, 2017.

231. Berquin, I. M., I. J. Edwards, S. J. Kridel and Y. Q. Chen, "Polyunsaturated Fatty Acid Metabolism in Prostate Cancer", *Cancer and Metastasis Reviews*, Vol. 30, No. 0, 2011.

232. Stotz, M., D. A. Barth, J. M. Riedl, E. Asamer, E. V. Klocker, P. Kornprat, G. C. Hutterer, F. Prinz, K. Lackner, H. Stöger, A. Gerger and M. Pichler, "The Lipase/Amylase Ratio (LAR) in Peripheral Blood Might Represent a Novel Prognostic Marker in Patients with Surgically Resectable Pancreatic Cancer", *Cancers (Basel)*, Vol. 12, No. 7, 2020.

233. Rowland, A., J. O. Miners and P. I. Mackenzie, "The UDP-Glucuronosyltransferases: Their Role in Drug Metabolism and Detoxification", *The International Journal of Biochemistry & Cell Biology*, Vol. 45, No. 6, pp. 1121–1132, 2013.

234. Takahara, N., Y. Nakai, H. Isayama, T. Sasaki, Y. Satoh, D. Takai, T. Hamada, R. Uchino, S. Mizuno, K. Miyabayashi, D. Mohri, K. Kawakubo, H. Kogure, N. Yamamoto,

N. Sasahira, K. Hirano, H. Ijichi, M. Tada, Y. Yatomi, K. Koike, "Uridine Diphosphate Glucuronosyl Transferase 1 Family Polypeptide A1 Gene (UGT1A1) Polymorphisms Are Associated with Toxicity and Efficacy in Irinotecan Monotherapy for Refractory Pancreatic Cancer", *Cancer Chemotherapy and Pharmacology*, Vol. 71, No. 1, pp. 85–92, 2013.

235. Qi, F., W.-X. Qin and Y.-S. Zang, "Molecular Mechanism of Triple-Negative Breast Cancer-Associated BRCA1 and the Identification of Signaling Pathways", *Oncology Letters*, Vol. 17, No. 3, pp. 2905–2914, 2019.

236. Bleul, T., R. Rühl, S. Bulashevska, S. Karakhanova, J. Werner and A. V. Bazhin, "Reduced Retinoids and Retinoid Receptors' Expression in Pancreatic Cancer: A Link to Patient Survival", *Molecular Carcinogenesis*, Vol. 54, No. 9, pp. 870–879, 2015.

237. Direito, I., J. Paulino, E. Vigia, M. A. Brito and G. Soveral, "Differential Expression of Aquaporin-3 and Aquaporin-5 in Pancreatic Ductal Adenocarcinoma", *Journal of Surgical Oncology*, Vol. 115, No. 8, pp. 980–996, 2017.

238. Woo, J., J. Lee, M. S. Kim, S. J. Jang, D. Sidransky and C. Moon, "The Effect of Aquaporin 5 Overexpression on the Ras Signaling Pathway", *Biochemical and Biophysical Research Communications*, Vol. 367, No. 2, pp. 291–298, 2008.

239. Ijuin, T., "Phosphoinositide Phosphatases in Cancer Cell Dynamics—Beyond PI3K and PTEN", *Seminars in Cancer Biology*, Vol. 59, pp. 50–65, 2019.

240. Basturk, O., M. F. Berger, H. Yamaguchi, V. Adsay, G. Askan, U. K. Bhanot, A. Zehir, F. Carneiro, S.-M. Hong, G. Zamboni, E. Dikoglu, V. Jobanputra, K. O. Wrzeszczynski, S. Balci, P. Allen, N. Ikari, S. Takeuchi, H. Akagawa, A. Kanno, T: Shimosegawa, T. Morikawa, F. Motoi, M. Unno, R. Higuchi, M. Yamamoto, K. Shimizu, T. Furukawa and D. S. Klimstra, "Pancreatic Intraductal Tubulopapillary Neoplasm Is Genetically Distinct from Intraductal Papillary Mucinous Neoplasm and Ductal Adenocarcinoma", *Modern Pathology*, Vol. 30, No. 12, 2017.

241. Hallas, C., J. Phillipp, L. Domanowsky, B. Kah and K. Tiemann, “BCL9L Expression in Pancreatic Neoplasia with a Focus on SPN: A Possible Explanation for the Enigma of the Benign Neoplasia”, *BMC Cancer*, Vol. 16, No. 1, p. 648, 2016.

242. Mazzu, Y. Z., J. Armenia, G. Chakraborty, Y. Yoshikawa, S. A. Coggins, S. Nandakumar, T. A. Gerke, M. M. Pomerantz, X. Qiu, H. Zhao, M. Atiq, N. Khan, K. Komura, G.-S. M. Lee, S. W. Fine, C. Bell, E. O'Connor, H. W. Long, M. L. Freedman, B. Kim and P. W. Kantoff, “A Novel Mechanism Driving Poor-Prognosis Prostate Cancer: Overexpression of the DNA Repair Gene, Ribonucleotide Reductase Small Subunit M2 (RRM2)”, *Clinical Cancer Research*, Vol. 25, No. 14, pp. 4480–4492, 2019.

243. Zhuang, S., L. Li, Y. Zang, G. Li and F. Wang, “RRM2 Elicits the Metastatic Potential of Breast Cancer Cells by Regulating Cell Invasion, Migration and VEGF Expression via the PI3K/AKT Signaling”, *Oncology Letters*, Vol. 19, No. 4, pp. 3349–3355, 2020.

244. Duxbury, M. S., H. Ito, M. J. Zinner, S. W. Ashley and E. E. Whang, “RNA Interference Targeting the M2 Subunit of Ribonucleotide Reductase Enhances Pancreatic Adenocarcinoma Chemosensitivity to Gemcitabine”, *Oncogene*, Vol. 23, No. 8, 2004.

245. Xia, G., H. Wang, Z. Song, Q. Meng, X. Huang and X. Huang, “Gambogic Acid Sensitizes Gemcitabine Efficacy in Pancreatic Cancer by Reducing the Expression of Ribonucleotide Reductase Subunit-M2 (RRM2)”, *Journal of Experimental & Clinical Cancer Research*, Vol. 36, No. 1, p. 107, 2017.

246. Tu, M., H. Li, N. Lv, C. Xi, Z. Lu, J. Wei, J. Chen, F. Guo, K. Jiang, G. Song, W. Gao and Y. Miao, “Vasohibin 2 Reduces Chemosensitivity to Gemcitabine in Pancreatic Cancer Cells via Jun Proto-Oncogene Dependent Transactivation of Ribonucleotide Reductase Regulatory Subunit M2”, *Molecular Cancer*, Vol. 16, No. 1, p. 66, 2017.

247. Lewis, C. S., C. Voelkel-Johnson and C. D. Smith, “Suppression of c-Myc and RRM2 Expression in Pancreatic Cancer Cells by the Sphingosine Kinase-2 Inhibitor ABC294640”, *Oncotarget*, Vol. 7, No. 37, pp. 60181–60192, 2016.

248. Raimondi, M. V., O. Randazzo, M. La Franca, G. Barone, E. Vignoni, D. Rossi and S. Collina, "DHFR Inhibitors: Reading the Past for Discovering Novel Anticancer Agents", *Molecules*, Vol. 24, No. 6, 2019.

249. Xu J. and Y. Li, "Discovering Disease-Genes by Topological Features in Human Protein–Protein Interaction Network", *Bioinformatics*, Vol. 22, No. 22, pp. 2800–2805, 2006.

250. Jones, P. J., R. Ma and R. J. McNally, "Bridge Centrality: A Network Approach to Understanding Comorbidity", *Multivariate Behavioral Research*, 2019.

251. Malami, I. and A. B. Abdul, "Involvement of the Uridine Cytidine Kinase 2 Enzyme in Cancer Cell Death: A Molecular Crosstalk between the Enzyme and Cellular Apoptosis Induction", *Biomedicine & Pharmacotherapy*, Vol. 109, pp. 1506–1510, 2019.

252. Wu, Y., M. Jamal, T. Xie, J. Sun, T. Song, Q. Yin, J. Li, S. Pan, X. Zeng, S. Xie and Q. Zhang, "Uridine-Cytidine Kinase 2 (UCK2): A Potential Diagnostic and Prognostic Biomarker for Lung Cancer", *Cancer Science*, Vol. 110, No. 9, pp. 2734–2747, 2019.

253. Shen, G., P. He, Y. Mao, P. Li, F. Luh, G. Ding, X. Liu and Y. Yen, "Overexpression of Uridine-Cytidine Kinase 2 Correlates with Breast Cancer Progression and Poor Prognosis", *Journal of Breast Cancer*, Vol. 20, No. 2, pp. 132–141, 2017.

254. Yu, S., X. Li, X. Guo, H. Zhang, R. Qin and M. Wang, "UCK2 Upregulation Might Serve as an Indicator of Unfavorable Prognosis of Hepatocellular Carcinoma", *IUBMB Life*, Vol. 71, No. 1, pp. 105–112, 2019.

255. Hassouni, B. E., J. Infante, G. Mantini, C. Ricci, N. Funel, E. Giovannetti and G. J. Peters, "Uridine Cytidine Kinase 2 as a Potential Biomarker for Treatment with RX-3117 in Pancreatic Cancer", *Anticancer Research*, Vol. 39, No. 7, pp. 3609–3614, 2019.

256. Ravier, M. A., M. Leduc, J. Richard, N. Linck, A. Varrault, N. Pirot, M. M. Roussel, J. Bockaert, S. Dalle and G. Bertrand, "β-Arrestin2 Plays a Key Role in The Modulation of the Pancreatic Beta Cell Mass in Mice", *Diabetologia*, Vol. 57, No. 3, pp. 532–541. 2014.

257. Masi, M., E. Garattini, M. Bolis, D. Di Marino, L. Maraccani, E. Morelli, A. A. Grolla, F. Fagiani, E. Corsini, C. Travelli, S. Govoni, M. Racchi and E. Buoso, "OXER1 and RACK1-Associated Pathway: A Promising Drug Target for Breast Cancer Progression", *Oncogenesis*, Vol. 9, No. 12, 2020.

258. Sarveswaran, S. and J. Ghosh, "OXER1, a G Protein-Coupled Oxoeicosatetraenoid Receptor, Mediates the Survival-Promoting Effects of Arachidonate 5-Lipoxygenase in Prostate Cancer Cells", *Cancer Letters*, Vol. 336, No. 1, pp. 185–195, 2013.

259. Körner, M., B. Waser, O. Strobel, M. Büchler and J. C. Reubi, "Neurotensin Receptors in Pancreatic Ductal Carcinomas", *EJNMMI Research*, Vol. 5, No. 1, p. 17, 2015.

260. Reubi, J. C., B. Waser, H. Friess, M. Büchler and J. Laissue, "Neurotensin Receptors: A New Marker for Human Ductal Pancreatic Adenocarcinoma", *Gut*, Vol. 42, No. 4, pp. 546–550, 1998.

261. Wang, L., H. Friess, Z. Zhu, H. Graber, A. Zimmermann, M. Korc, J. C. Reubi and M. W. Büchler, "Neurotensin Receptor-1 mRNA Analysis in Normal Pancreas and Pancreatic Disease", *Clinical Cancer Research*, Vol. 6, No. 2, pp. 566–571, 2000.

262. Brachtendorf, S., K. El-Hindi and S. Grösch, "Ceramide Synthases in Cancer Therapy and Chemoresistance", *Progress in Lipid Research*, Vol. 74, pp. 160–185, 2019.

263. Attardi, L. D., E. E. Reczek, C. Cosmas, E. G. Demicco, M. E. McCurrach, S. W. Lowe and T. Jacks, "PERP, an Apoptosis-Associated Target of p53, Is a Novel Member of the PMP-22/Gas3 Family", *Genes & Development*, Vol. 14, No. 6, pp. 704–718, 2000.

264. Yuan, L., G. Qian, L. Chen, C.-L. Wu, H. C. Dan, Y. Xiao and X. Wang, "Co-Expression Network Analysis of Biomarkers for Adrenocortical Carcinoma", *Frontiers in Genetics*, Vol. 9, 2018.

265. Tang, J., D. Kong, Q. Cui, K. Wang, D. Zhang, Y. Gong and G. Wu, “Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis”, *Frontiers in Oncology*, Vol. 8, 2018.

266. Guney, E., J. Menche, M. Vidal and A.-L. Barábasi, “Network-Based *in silico* Drug Efficacy Screening”, *Nature Communications*, Vol. 7, p. 10331, 2016.

267. Fuentealba, M., H. M. Dönertaş, R. Williams, J. Labbadia, J. M. Thornton and L. Partridge, “Using the Drug-Protein Interactome to Identify Anti-Ageing Compounds for Humans”, *PLoS Computational Biology*, Vol. 15, No. 1, p. e1006639, 2019.

268. Cheng, F., R. J. Desai, D. E. Handy, R. Wang, S. Schneeweiss, A.-L. Barábasi and J. Loscalzo, “Network-Based Approach to Prediction and Population-Based Validation off *in silico* Drug Repurposing”, *Nature Communications*, Vol. 9, No. 1, 2018.

269. Robinson, J. L., A. Feizi, M. Uhlén and J. Nielsen, “A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome”, *Cell Reports*, Vol. 26, No. 10, pp. 2622-2635, 2019.

270. Pacheco, M. P., T. Bintener, D. Ternes, D. Kulms, S. Haan, E. Letellier and T. Sauter, “Identifying and Targeting Cancer-Specific Metabolism with Network-Based Drug Target Prediction”, *EBioMedicine*, Vol. 43, pp. 98–106, 2019.

APPENDIX A: SUPPLEMENTARY DATA TABLES

This appendix contains the additional data tables produced in this study, truncated from the main body in the interest of brevity.

Table A.1. KEGG Pancreatic Cancer seed genes [19–21].

Cancer Seed Genes	
KRAS	EGF
PIK3CA	EGFR
PIK3CD	ERBB2
PIK3CB	MTOR
PIK3R1	RPS6KB1
PIK3R2	RPS6KB2
PIK3R3	JAK1
ARHGEF6	STAT3
RAC1	STAT1
RAC2	VEGFA
RAC3	CDKN2A
NFKB1	CDK4
RELA	CDK6
AKT1	CCND1
AKT2	RB1
AKT3	E2F1
CHUK	E2F2
IKBKB	E2F3
IKBKG	TP53
BAD	CDKN1A
BCL2L1	GADD45A
CASP9	GADD45B
ARAF	GADD45G

Table A.1. KEGG Pancreatic Cancer seed genes [19–21]. (cont.)

Cancer Seed Genes	
BRAF	BAX
RAF1	BAK1
MAP2K1	DDB2
MAPK1	POLK
MAPK3	TGFB1
MAPK8	TGFB2
MAPK10	TGFB3
MAPK9	TGFBR1
RALGDS	TGFBR2
RALA	SMAD2
RALB	SMAD3
RALBP1	SMAD4
CDC42	BRCA2
TGFA	RAD51

Table A.2. KEGG Pancreatic Secretion seed genes [19–21].

Healthy Seed Genes			
CHRM3	CELA3A	ITPR3	ATP2B1
CCKAR	CELA3B	RYR2	ATP2B3
GNAQ	CPA1	CD38	ATP2B4
PLCB1	CPA2	BST1	ATP2B2
PLCB2	CPA3	TPCN2	SLC9A1
PLCB3	CPB1	ATP2A1	SLC4A2
PLCB4	CPB2	ATP2A3	SCTR
PRKCA	PNLIP	ATP2A2	GNAS
PRKCB	PNLIPRP1	CLCA1	ADCY1
PRKCG	PNLIPRP2	CLCA2	ADCY2

Table A.2. KEGG Pancreatic Secretion seed genes [19–21]. (cont.)

Healthy Seed Genes			
RAB3D	CEL	CLCA4	ADCY3
RAB8A	PLA2G10	TRPC1	ADCY4
RAB11A	PLA2G2D	SLC12A2	ADCY5
RAB27B	PLA2G2E	ATP1A1	ADCY6
RAP1A	PLA2G3	ATP1A2	ADCY7
RAP1B	PLA2G2F	ATP1A3	ADCY8
RHOA	PLA2G12A	ATP1A4	ADCY9
RAC1	PLA2G12B	ATP1B4	CFTR
PRSS3	PLA2G1B	ATP1B1	SLC26A3
PRSS2	PLA2G5	ATP1B2	SLC4A4
PRSS1	PLA2G2A	ATP1B3	CA2
CTRL	PLA2G2C	FXYD2	CTRB1
CELA2A	ITPR1	KCNQ1	AMY2A
CELA2B	ITPR2	KCNMA1	AMY2B

Table A.3. TTD Drug Database [73].

HGNC Gene Symbol	**Drug Name**	**Target Indication**	**Drug Indication**
ABCB1	Biricodar	Ovarian cancer	Ovarian cancer
ABL1	Bosutinib	NA	Breast cancer
ADRA1D	Armodafinil	NA	Pediatric cancer
ALK	Brigatinib	NA	Non-small-cell lung cancer
ALK	Crizotinib	NA	Non-small-cell lung cancer
ALK	Ceritinib	NA	Non-small-cell lung cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
ALK	Lorlatinib	NA	Non-small-cell lung cancer
ALK	Alectinib	NA	Lung cancer
AOX1	Isovanillin	NA	Solid tumour/cancer
AR	Prasterone	Mammary tumour	NA
AR	Testosterone	Mammary tumour	NA
AR	Fluoxymesterone	Prostate cancer	Breast cancer
AR	Fludrocortisone	Prostate cancer	NA
AR	Oxandrolone	Mammary tumour	NA
AR	Flufenamic Acid	Breast cancer	NA
AR	Testosterone	Breast cancer	NA
AR	Prasterone	Breast cancer	NA
AR	Flufenamic Acid	Prostate cancer	NA
AR	Bicalutamide	Prostate cancer	Prostate cancer
AR	Cyproterone	Breast cancer	Prostate cancer
AR	Oxandrolone	Breast cancer	NA
AR	Cyproterone	Mammary tumour	Prostate cancer
AR	ARN-509	Breast cancer	NA
AR	Ethylestrenol	Prostate cancer	NA

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
AR	Fluoxymesterone	Mammary tumour	Breast cancer
AR	Oxandrolone	Prostate cancer	NA
AR	Ethylestrenol	Breast cancer	NA
AR	Enzalutamide	Breast cancer	Prostate cancer
AR	Prasterone	Prostate cancer	NA
AR	Bicalutamide	Mammary tumour	Prostate cancer
AR	Nilutamide	Prostate cancer	Prostate cancer
AR	Cyproterone	Prostate cancer	Prostate cancer
AR	Flufenamic Acid	Mammary tumour	NA
AR	Dromostanolone	Prostate cancer	Mammary tumour
AR	ARN-509	Prostate cancer	NA
AR	Hydroxyflutamide	Breast cancer	Prostate cancer
AR	Fludrocortisone	Mammary tumour	NA
AR	ARN-509	Mammary tumour	NA
AR	Flutamide	Mammary tumour	Prostate cancer
AR	Bicalutamide	Breast cancer	Prostate cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
AR	Enzalutamide	Mammary tumour	Prostate cancer
AR	Nilutamide	Breast cancer	Prostate cancer
AR	Nandrolone	Prostate cancer	NA
AR	Nandrolone	Breast cancer	NA
AR	Hydroxyflutamide	Prostate cancer	Prostate cancer
AR	Nandrolone	Mammary tumour	NA
AR	Fluoxymesterone	Breast cancer	Breast cancer
AR	Fludrocortisone	Breast cancer	NA
AR	Nilutamide	Mammary tumour	Prostate cancer
AR	Flutamide	Breast cancer	Prostate cancer
AR	Dromostanolone	Mammary tumour	Mammary tumour
AR	Dromostanolone	Breast cancer	Mammary tumour
AR	Enzalutamide	Prostate cancer	Prostate cancer
AR	Flutamide	Prostate cancer	Prostate cancer
AR	Ethylestrenol	Mammary tumour	NA
AR	Hydroxyflutamide	Mammary tumour	Prostate cancer
BCL2	GDC-0199	Breast cancer	Solid tumour/cancer
BCL2	Taxol	Breast cancer	Breast cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
BCL2	Oral paclitaxel	Breast cancer	Solid tumour/cancer
BCL2	MCI-186	Breast cancer	NA
CACNA1G	Trimethadione	NA	Pancreatic cancer
CCKBR	Ceruletide	Pancreatic internal secretion disorder	NA
CCKBR	Pentagastrin	Pancreatic internal secretion disorder	NA
CD274	Avelumab	NA	Solid tumour/cancer
CDK4	Ribociclib Succinate	NA	Breast cancer
CDK4	Ribociclib Succinate	NA	Hormone receptor positive and HER2-negative advanced or metastatic Breast cancer
CDK4	LY2835219	NA	Breast cancer
CDK6	Ribociclib Succinate	NA	Hormone receptor positive and HER2-negative advanced or metastatic Breast cancer
CDK6	Ribociclib Succinate	NA	Breast cancer
CDK6	LY2835219	NA	Breast cancer
CEACAM3	Arcitumomab	Breast cancer	Breast cancer
CSF3R	Pegfilgrastim	NA	Malignant solid tumour

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
CTNNB1	Recombinant human endostatin	NA	Solid tumour/cancer
CYP11B1	FADROZOLE	NA	Breast cancer
CYP17A1	Abiraterone acetate	NA	Prostate cancer
CYP17A1	ABIRATERONE	NA	Prostate cancer
CYP19A1	Letrozole	Breast cancer	Hormonally-responsive Breast cancer
CYP19A1	Exemestane	Breast cancer	Hormonally-responsive Breast cancer
CYP19A1	Anastrozole	Breast cancer	Breast cancer
CYP19A1	FADROZOLE	Breast cancer	Breast cancer
CYP19A1	Amino-glutethimide	Breast cancer	NA
CYP19A1	Testolactone	Breast cancer	Breast cancer
DHFR	Leucovorin Calcium	Solid tumour/ cancer	Solid tumour/cancer
DHFR	Methotrexate Sodium	Solid tumour/ cancer	Solid tumour/cancer
DHFR	Trimethoprim	Solid tumour/ cancer	NA
DHFR	Aminosalicylic Acid	Solid tumour/ cancer	NA
DPYD	Fluorouracil	NA	Solid tumour/cancer
EEF2K	Lapatinib	NA	Breast cancer
EGFR	Tyverb/Tykerb	NA	Breast cancer
EGFR	Erlotinib	NA	Non-small-cell lung cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
EGFR	Dacomitinib	NA	Non-small-cell lung cancer
EGFR	NERATINIB MALEATE	NA	HER2/NEU overexpressing Breast cancer
EGFR	Vandetanib	NA	Solid tumour/cancer
EGFR	Cetuximab	NA	Colorectal cancer
EGFR	BIBW 2992	NA	Non-small-cell lung cancer
EGFR	Necitumumab	NA	Colorectal cancer
EGFR	Panitumumab	NA	Colorectal cancer
EGFR	Lapatinib	NA	Breast cancer
EGFR	Gefitinib	NA	Solid tumour/cancer
EPOR	RHuEPO	NA	Solid tumour/cancer
ERBB2	BIBW 2992	Prostate cancer	Non-small-cell lung cancer
ERBB2	Masoprocol	Prostate cancer	Prostate cancer
ERBB2	Dacomitinib	Prostate cancer	Non-small-cell lung cancer
ERBB2	Trastuzumab	Prostate cancer	Breast cancer
ERBB2	NERATINIB MALEATE	Prostate cancer	HER2/NEU overexpressing Breast cancer
ERBB2	Tyverb/Tykerb	Prostate cancer	Breast cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
ERBB2	Pertuzumab	Prostate cancer	Breast cancer
ERBB4	Dacomitinib	NA	Non-small-cell lung cancer
ESR1	Toremifene	NA	Breast cancer
ESR1	ARZOXIFENE	NA	Breast cancer
ESR1	Nomegestrol acetate	NA	Breast cancer
ESR1	Quinestrol	NA	Breast cancer
ESR1	Esterified estrogens	NA	Breast cancer
ESR1	Levormeloxifene non-steroidal	NA	Breast cancer
ESR1	Fosfestrol	NA	Solid tumour/cancer
ESR1	Tamoxifen	NA	Breast cancer
ESR1	Fulvestrant	NA	Breast cancer
ESR1	Gestrinone	NA	Breast cancer
ESR1	Estradiol	NA	Breast cancer
ESR2	ARZOXIFENE	NA	Breast cancer
ESRRA	Cortisone Acetate	NA	Solid tumour/cancer
ESRRB	Cortisone Acetate	NA	Solid tumour/cancer
ESRRG	Cortisone Acetate	NA	Solid tumour/cancer
FGFR1	Intedanib	NA	Colorectal cancer
FLT3	Intedanib	NA	Colorectal cancer
FOLH1	Capromab	NA	Prostate cancer
GNRH1	Goserelin	Breast cancer	Breast cancer
GNRH1	Goserelin	Prostate cancer	Breast cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
GNRH1	Leuprolide	Breast cancer	Prostate cancer
GNRH1	Leuprolide	Prostate cancer	Prostate cancer
GNRHR	Abarelix	NA	Prostate cancer
GNRHR	DEGARELIX	NA	Prostate cancer
GNRHR	Leuprorelin acetate	NA	Prostate cancer
IMPDH1	Tyverb/Tykerb	NA	Breast cancer
IMPDH1	Tiazofurin	NA	Solid tumour/cancer
KDR	Vandetanib	NA	Solid tumour/cancer
KDR	Sunitinib	NA	Gastrointestinal cancer
KDR	YN-968D1	NA	Breast cancer
KDR	Regorafenib	NA	Metastatic colorectal cancer
KDR	Cabozantinib	NA	Thyroid cancer
KIT	Regorafenib	NA	Metastatic colorectal cancer
LDLR	Porfimer Sodium	NA	Non-small-cell lung cancer
MAP1A	Demecolcine	NA	Solid tumour/cancer
MAP1B	Demecolcine	NA	Solid tumour/cancer
MET	Cabozantinib	NA	Thyroid cancer
MET	Crizotinib	NA	Non-small-cell lung cancer
MMP1	Prinomastat	NA	Lung cancer
MMP2	Prinomastat	NA	Lung cancer
MMP7	Prinomastat	NA	Lung cancer
MTOR	Zotarolimus	NA	Solid tumour/cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
MTOR	PF-04449913	NA	Solid tumour/cancer
NR3C1	Methyl-prednisolone Acetate	NA	Solid tumour/cancer
NR3C1	Prednisolone	NA	Solid tumour/cancer
NTRK1	Larotrectinib	NA	Solid tumour/cancer
NTRK2	Larotrectinib	NA	Solid tumour/cancer
NTRK3	Larotrectinib	NA	Solid tumour/cancer
PARP1	Nicotinamide	Peritoneal cancer	NA
PARP1	Niraparib Tosylate	Peritoneal cancer	Peritoneal cancer
PARP1	Nicotinamide	Ovarian cancer	NA
PARP1	Niraparib Tosylate	Ovarian cancer	Peritoneal cancer
PARP1	Nicotinamide	Fallopian tube cancer	NA
PARP1	KU-0058948	Peritoneal cancer	Ovarian cancer
PARP1	Niraparib Tosylate	Ovarian cancer	Fallopian tube cancer
PARP1	Niraparib Tosylate	Peritoneal cancer	Fallopian tube cancer
PARP1	Niraparib Tosylate	Fallopian tube cancer	Fallopian tube cancer
PARP1	Niraparib Tosylate	Fallopian tube cancer	Peritoneal cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
PARP1	KU-0058948	Ovarian cancer	Ovarian cancer
PDCD1	Nivolumab	NA	Non-small-cell lung cancer
PDF	Pralatrexate	NA	Breast cancer
PGF	Aflibercept	NA	Metastatic colorectal cancer
PGR	Dydrogesterone	Breast cancer	NA
PGR	Levonorgestrel	Solid tumour/ cancer	NA
PGR	Norgestimate	Breast cancer	NA
PGR	Medroxy-progesterone	Solid tumour/ cancer	Solid tumour/cancer
PGR	Medroxy-progesterone	Breast cancer	Solid tumour/cancer
PGR	Ulipristal	Solid tumour/ cancer	NA
PGR	Ulipristal	Breast cancer	NA
PGR	Etonogestrel	Solid tumour/ cancer	NA
PGR	Ethynodiol Diacetate	Breast cancer	NA
PGR	Etonogestrel	Breast cancer	NA
PGR	Progesterone	Solid tumour/ cancer	NA
PGR	Dydrogesterone	Solid tumour/ cancer	NA
PGR	Norethindrone	Breast cancer	Solid tumour/cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
PGR	Desogestrel	Solid tumour/ cancer	NA
PGR	Ethynodiol Diacetate	Solid tumour/ cancer	NA
PGR	Levonorgestrel	Breast cancer	NA
PGR	Progesterone	Breast cancer	NA
PGR	Segesterone acetate and ethinyl estradiol vaginal system	Breast cancer	NA
PGR	Megestrol	Solid tumour/ cancer	Breast cancer
PGR	Estradiol valerate/dienogest	Breast cancer	NA
PGR	Megestrol	Breast cancer	Breast cancer
PGR	Norethindrone	Solid tumour/ cancer	Solid tumour/cancer
PGR	Estradiol valerate/dienogest	Solid tumour/ cancer	NA
PGR	Desogestrel	Breast cancer	NA
PGR	Segesterone acetate and ethinyl estradiol vaginal system	Solid tumour/ cancer	NA
PGR	Norgestimate	Solid tumour/ cancer	NA
PRDX5	Motexafin gadolinium	NA	Brain cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
RET	Vandetanib	NA	Solid tumour/cancer
RET	Regorafenib	NA	Metastatic colorectal cancer
ROS1	Lorlatinib	NA	Non-small-cell lung cancer
ROS1	Crizotinib	NA	Non-small-cell lung cancer
RRM2	Gemcitabine	NA	Solid tumour/cancer
SLC6A2	Iobenguane I-123	NA	Neuroendocrine cancer
SMO	PF-04449913	NA	Solid tumour/cancer
SRC	Herbimycin A	NA	Solid tumour/cancer
SRC	Bosutinib	NA	Breast cancer
SRD5A1	Dutasteride	Prostate cancer	NA
SRD5A1	Finasteride	Prostate cancer	NA
SRD5A1	Azelaic Acid	Prostate cancer	NA
SRD5A1	Polyestradiol Phosphate	Prostate cancer	Prostate cancer
SRD5A2	Finasteride	Prostate cancer	NA
SRD5A2	Polyestradiol Phosphate	Prostate cancer	Prostate cancer
SRD5A2	Azelaic Acid	Prostate cancer	NA
SRD5A2	Dutasteride	Prostate cancer	NA

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
SRD5A3	Dutasteride	Prostate cancer	NA
SRD5A3	Polyestradiol Phosphate	Prostate cancer	Prostate cancer
SRD5A3	Finasteride	Prostate cancer	NA
SRD5A3	Azelaic Acid	Prostate cancer	NA
SSTR1	Pasireotide	Stomach cancer	NA
SSTR1	Lutetium Lu 177 dotatate	Stomach cancer	Neuroendocrine cancer
TERT	Doxorubicin	NA	Solid tumour/cancer
TERT	Fluorouracil	NA	Solid tumour/cancer
TLR7	Imiquimod	NA	Skin cancer
TNF	Adalimumab	Pancreatitis	NA
TNF	Certolizumab	Pancreatitis	NA
TNF	Etanercept	Pancreatitis	NA
TNF	Pentoxifylline	Pancreatitis	NA
TNF	Golimumab	Pancreatitis	NA
TNF	Thalidomide	Pancreatitis	NA
TNF	Certolizumab pegol	Pancreatitis	NA
TNF	Infliximab	Pancreatitis	NA
TNF	Nafamostat	Pancreatitis	NA
TNF	Lenalidomide	Pancreatitis	NA
TNF	Enbrel	Pancreatitis	NA

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
TOP1	Belotecan hydrocholoride	Lung cancer	Small-cell lung cancer
TOP1	Sphingosomal topotecan	Lung cancer	Solid tumour/cancer
TOP1	Topetecan	Lung cancer	Small-cell lung cancer
TOP1	Topotecan	Lung cancer	Ovarian cancer
TOP1	Irinotecan	Lung cancer	Colorectal cancer
TOP1	B-Lactams	Lung cancer	NA
TOP2A	Dexrazoxane	NA	Breast cancer
TOP2A	Lucanthone	NA	Solid tumour/cancer
TOP2A	Mitoxantrone	NA	Solid tumour/cancer
TOP2A	Etoposide	NA	Solid tumour/cancer
TOP2A	Dhaq diacetate	NA	Solid tumour/cancer
TOP2A	Valrubicin	NA	Bladder cancer
TOP2A	Doxorubicin	NA	Solid tumour/cancer
TOP2A	Epirubicin	NA	Solid tumour/cancer
TOP2B	Mitoxantrone	NA	Solid tumour/cancer
TOP2B	Lucanthone	NA	Solid tumour/cancer
TOP2B	Valrubicin	NA	Bladder cancer
TOP2B	Dhaq diacetate	NA	Solid tumour/cancer
TOP2B	Doxorubicin	NA	Solid tumour/cancer
TOP2B	Epirubicin	NA	Solid tumour/cancer
TOP2B	Dexrazoxane	NA	Breast cancer
TOP2B	Etoposide	NA	Solid tumour/cancer
TPSAB1	Pentamidine	Solid tumour/ cancer	NA
TPSAB1	Lactoferrin	Solid tumour/ cancer	Solid tumour/cancer

Table A.3. TTD Drug Database [73]. (cont.)

HGNC Gene Symbol	Drug Name	Target Indication	Drug Indication
TSHR	Thyrotropin Alfa	NA	Thyroid cancer
TUBB2A	Vinblastine	Solid tumour/ cancer	Solid tumour/cancer
TXNRD1	Fotemustine	NA	Solid tumour/cancer
TYMS	Fluorouracil	NA	Solid tumour/cancer
VEGFA	Ranibizumab	Colorectal cancer	NA
VEGFA	Aflibercept	Colorectal cancer	Metastatic colorectal cancer
VEGFB	Aflibercept	Colorectal cancer	Metastatic colorectal cancer

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75].

Metabolite	Flux (fmol/cell/h)	Adapted Flux (fmol/cell/h)
Lactate	241.365	100
Alanine	9.277	10
Glutamate	6.770	10
Alpha-glycerophosphocholine	4.221	10
Citrate	0.889	1
Proline	0.647	1
Ornithine	0.370	1
Glycerol_1	0.319	1
Phosphocholine	0.298	0.1
Malate	0.279	0.1
Acetoacetate	0.205	0.1
Glycine	0.148	0.1

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75]. (cont.)

Metabolite	**Flux (fmol/cell/h)**	**Adapted Flux (fmol/cell/h)**
Creatinine	0.084	0.1
Succinate	0.074	0.1
Fumarate	0.059	0.1
Bilirubin	0.045	0.1
Sorbitol	0.043	0.1
Kynurenine	0.035	0.1
Gaba	0.030	0.01
2'-deoxycytidine	0.027	0.01
Xanthine	0.027	0.01
Homocysteine	0.023	0.01
Betaine	0.020	0.01
Glutathione oxidized	0.019	0.01
Thiamine	0.019	0.01
Alpha-glycerophosphate	0.017	0.01
N-carbamoyl-beta-alanine	0.014	0.01
Cystathionine	0.013	0.01
Phosphoethanolamine	0.012	0.01
Uridine	0.012	0.01
OH-phenylpyruvate	9.0×10^{-3}	0.01
Uracil	8.6×10^{-3}	0.01
Dhap	7.4×10^{-3}	0.01
2-phosphoglycerate	7.1×10^{-3}	0.01
Serotonin	6.5×10^{-3}	0.01
Isocitrate	5.8×10^{-3}	0.01
3-phosphoglycerate	5.7×10^{-3}	0.01
Ump	5.5×10^{-3}	0.01
Urate	4.8×10^{-3}	0.01
Thymidine	4.6×10^{-3}	0.01

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75]. (cont.)

Metabolite	**Flux (fmol/cell/h)**	**Adapted Flux (fmol/cell/h)**
Thymine	4.5×10^{-3}	0.01
Adenosine	3.7×10^{-3}	0.01
5-hiaa	3.7×10^{-3}	0.01
Alpha-ketoglutarate	3.0×10^{-3}	0.001
2'-deoxyuridine	3.0×10^{-3}	0.001
Glucuronate	3.0×10^{-3}	0.001
Amp	2.7×10^{-3}	0.001
Dimethylglycine	2.5×10^{-3}	0.001
Kynurenate	2.3×10^{-3}	0.001
Pep	2.3×10^{-3}	0.001
Gmp	2.2×10^{-3}	0.001
Xmp	2.2×10^{-3}	0.001
Sucrose	2.0×10^{-3}	0.001
Cmp	1.8×10^{-3}	0.001
5'-adenosylhomocysteine	1.6×10^{-3}	0.001
Orotate	1.6×10^{-3}	0.001
Proprionate	1.4×10^{-3}	0.001
Lactose	1.1×10^{-3}	0.001
Adenine	1.1×10^{-3}	0.001
Folate	9.2×10^{-4}	0.001
Inosine	7.1×10^{-4}	0.001
4-hydroxybenzoate	5.3×10^{-4}	0.001
Spermidine	4.8×10^{-4}	0.001
Xanthosine	4.7×10^{-4}	0.001
Quinolinate	3.3×10^{-4}	0.001
Dcmp	1.4×10^{-4}	0.001
4-pyridoxate	1.4×10^{-4}	0.001
3-OH-kynurenate	1.2×10^{-4}	0.001

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75]. (cont.)

Metabolite	**Flux (fmol/cell/h)**	**Adapted Flux (fmol/cell/h)**
Imp	6.5×10^{-5}	0.001
3-hydroxyanthranilate	6.2×10^{-5}	0.001
Methylmalonate	5.5×10^{-5}	0.001
Citrate/isocitrate	3.9×10^{-5}	0.001
Nicotinic acid mononucleotide	3.1×10^{-5}	0.001
Pyruvate	2.6×10^{-5}	0.001
Adma	2.2×10^{-5}	0.001
UDP-glucuronate	2.0×10^{-5}	0.001
Omp	1.8×10^{-5}	0.001
Adipate	7.9×10^{-6}	0.001
Nmma	5.7×10^{-6}	0.001
Malonate	2.3×10^{-6}	0.001
Aconitate	1.6×10^{-6}	0.001
Thyroxine	1.4×10^{-6}	0.001
Cotinine	7.9×10^{-7}	0.001
Ascorbate	7.9×10^{-8}	0.001
Homocystine	7.0×10^{-8}	0.001
Triiodothyronine	5.3×10^{-8}	0.001
Glycerol_2	5.3×10^{-8}	0.001
Trimethylamine-N-oxide	4.2×10^{-8}	0.001
Hippurate	2.3×10^{-8}	0.001
Fru-1,6-DP/fru-2,6-DP/glc-1,6-DP	1.1×10^{-8}	0.001
UDP-galactose/UDP-glucose	9.7×10^{-9}	0.001
Salicylurate	4.2×10^{-9}	0.001
2'-deoxyadenosine	0	0
Hyodeoxycholate/ursodeoxycholate	-2.4×10^{-9}	-0.001
Taurolithocholate	-7.8×10^{-9}	-0.001
Chenodeoxycholate/deoxycholate	-8.8×10^{-9}	-0.001

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75]. (cont.)

Metabolite	**Flux (fmol/cell/h)**	**Adapted Flux (fmol/cell/h)**
Phenylacetylglycine	-2.0×10^{-8}	-0.001
Biotin	-4.6×10^{-8}	-0.001
Lithocholate	-6.5×10^{-8}	-0.001
Maleate	-1.2×10^{-6}	-0.001
Allantoin	-6.5×10^{-6}	-0.001
Spermine	-7.0×10^{-5}	-0.001
Glycocholate	-7.9×10^{-5}	-0.001
Taurocholate	-1.2×10^{-4}	-0.001
Niacin	-1.5×10^{-4}	-0.001
Taurodeoxycholate/taurochenodeoxycholate	-2.4×10^{-4}	-0.001
Carnosine	-3.1×10^{-4}	-0.001
Glycodeoxycholate/glycochenodeoxycholate	-7.2×10^{-4}	-0.001
Aminoisobutyrate	-1.1×10^{-3}	-0.001
2-aminodipate	-1.2×10^{-3}	-0.001
Pantothenate	-2.8×10^{-3}	-0.001
Anthranilate	-3.1×10^{-3}	-0.001
Cytidine	-4.7×10^{-3}	-0.01
Carnitine	-9.0×10^{-3}	-0.01
Oxalate	-0.021	-0.01
Citrulline	-0.044	-0.1
Hypoxanthine	-0.046	-0.1
Niacinamide	-0.115	-0.1
Taurine	-0.130	-0.1
Creatine	-0.229	-0.1
Cis-hydroxyproline/trans-hydroxyproline	-0.342	-1
Tryptophan	-0.460	-1
Choline	-0.595	-1
Homoserine	-0.887	-1

Table A.4. Uptake/Secretion Fluxes Adapted from Jain et al. [75]. (cont.)

Metabolite	Flux (fmol/cell/h)	Adapted Flux (fmol/cell/h)
Glyceraldehyde	-1.096	-1
Aspartate	-1.216	-1
Asparagine	-1.403	-1
Methionine	-1.523	-1
Phenylalanine	-1.644	-1
Tyrosine	-2.228	-1
Guanidinoacetate	-2.664	-1
Threonine	-2.934	-1
Arginine	-3.123	-1
Valine	-3.385	-10
Isoleucine	-3.899	-10
Lysine	-4.812	-10
Leucine	-4.853	-10
Serine	-7.262	-10
Glutamine	-48.710	-100
Glucose	-186.154	-100
Linoleic acid	0	-0.1
Linolenic acid	0	-0.1

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found.

HGNC Gene Symbol/ ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
EPCAM	7.246	2.763	7.03E-13	SLC26A11	3.601	-0.856	1.97E-03
GPRC5A	5.128	4.239	3.85E-11	NCKAP1	6.818	0.766	1.97E-03
MAL2	6.179	2.522	7.56E-11	GPS2	3.786	-0.950	1.97E-03
CGN	6.583	2.644	1.02E-10	LIMD2	2.985	-1.645	1.97E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
KRT19	6.761	3.853	1.02E-10	GNGT2	-0.884	-1.532	2.03E-03
RAB25	5.571	2.781	1.02E-10	EFNB2	4.314	1.129	2.08E-03
NQO1	4.081	3.223	1.21E-10	BCL2L1	6.438	0.992	2.09E-03
FXYD3	5.254	4.541	1.21E-10	CKMT1A	0.398	2.952	2.15E-03
TSPAN1	3.870	3.981	2.52E-10	FLI1	2.871	-1.335	2.15E-03
PERP	6.766	2.161	3.81E-10	ZNF682	1.382	-1.176	2.16E-03
LAD1	6.720	2.227	5.93E-10	BMP2K	2.270	-0.956	2.16E-03
ERBB3	6.996	2.596	6.14E-10	MROH6	2.404	2.419	2.17E-03
MYH14	6.383	2.484	2.41E-09	ARHGAP25	2.091	-1.436	2.18E-03
KLF5	5.052	2.792	3.16E-09	CRYBG2	0.150	3.389	2.20E-03
ESRP1	6.075	2.558	2.36E-08	PLEKHO2	4.153	-1.108	2.20E-03
MISP	2.289	4.580	4.36E-08	DOCK8	3.993	-1.523	2.20E-03
PPL	5.647	2.517	4.36E-08	RHOQ	4.963	-0.816	2.20E-03
AGR2	5.052	5.674	4.39E-08	EOMES	-0.806	-2.096	2.23E-03
B3GNT3	3.424	4.277	7.20E-08	ENSG 00000259976	4.618	-0.816	2.27E-03
CEACAM6	4.883	4.990	8.16E-08	ZFP3	3.140	-0.864	2.33E-03
PKP3	5.680	2.790	9.83E-08	KDELR3	4.119	1.463	2.39E-03
KEL	-0.741	-2.903	1.12E-07	POU2F2	2.076	-1.553	2.45E-03
TMPRSS4	0.907	5.900	1.12E-07	DENND4B	5.708	-0.700	2.47E-03
MUC1	8.058	3.411	3.05E-07	CARMIL1	3.922	1.058	2.47E-03
FERMT1	3.381	2.761	4.66E-07	CCAT1	-3.536	4.646	2.51E-03
MLPH	5.380	3.298	5.96E-07	NFATC1	3.240	-1.178	2.51E-03
FRMD4A	4.389	-1.224	8.02E-07	ATP6V1B2	5.950	-0.748	2.55E-03
LAMA3	4.946	2.796	8.02E-07	ZNF418	1.781	-1.108	2.58E-03
FAM83H	5.875	1.853	8.31E-07	BAIAP2L2	2.654	3.244	2.67E-03
SDC1	5.974	2.268	1.00E-06	CYSTM1	6.090	1.384	2.67E-03
SERPINB5	0.315	7.011	1.03E-06	PLAT	5.177	1.904	2.70E-03
AP1M2	5.557	1.826	1.32E-06	LRATD2	5.211	0.934	2.70E-03
TRIM29	3.015	5.033	1.60E-06	MNX1-AS1	0.884	2.100	2.70E-03
C1orf116	5.232	2.475	1.88E-06	MUC13	3.635	3.454	2.72E-03
LINC02041	-3.952	4.131	1.94E-06	SCEL	0.169	5.537	2.72E-03
GAB3	2.186	-1.588	1.95E-06	B4GALNT3	4.261	1.450	2.75E-03
TUBB4A	0.391	-2.820	2.23E-06	SEPTIN6	3.668	-1.092	2.77E-03
EVPL	4.173	3.178	2.96E-06	STX11	0.550	-1.393	2.86E-03
C19orf33	2.636	3.636	3.47E-06	F11R	6.887	0.950	2.86E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
ABHD17C	3.989	2.248	3.95E-06	CHI3L2	-0.731	-2.041	2.87E-03
ARHGEF16	4.842	2.284	6.21E-06	ADAM15	6.279	0.896	2.88E-03
SFN	2.534	3.500	7.26E-06	NCF4	2.001	-1.458	2.88E-03
LIPH	4.104	2.509	7.58E-06	TMEM45B	2.044	2.888	2.89E-03
KRBA1	3.260	-1.184	8.34E-06	SH3RF2	4.987	2.552	2.89E-03
C6orf132	4.764	2.070	9.42E-06	CIITA	3.458	-1.662	3.03E-03
CLDN4	7.230	2.332	9.65E-06	ENSG 00000273472	-1.225	-1.281	3.07E-03
ADAP2	2.346	-1.487	9.65E-06	ERBB2	6.860	1.087	3.21E-03
S100A16	5.634	1.702	9.65E-06	ZNF542P	2.057	-0.988	3.21E-03
ITGB4	6.381	1.827	9.65E-06	FCHSD2	4.337	-0.666	3.24E-03
ZDHHC9	5.506	1.147	9.65E-06	CD84	3.029	-1.708	3.24E-03
EPHA2	5.457	2.290	9.65E-06	CHFR	4.649	-0.701	3.31E-03
ALAS2	-2.800	-3.257	1.16E-05	KLHL36	4.174	-0.582	3.31E-03
RPS3AP5	-0.646	-2.501	1.26E-05	PIP4K2C	5.191	0.695	3.35E-03
CAMK2N1	4.571	1.752	1.26E-05	TMEM255A	-0.648	-1.774	3.36E-03
S100A6	7.321	2.158	1.26E-05	DPEP2	0.175	-1.685	3.42E-03
ANKRD13A	5.112	-0.850	1.26E-05	CYTH4	2.596	-1.555	3.42E-03
PRSS8	6.422	2.346	1.51E-05	ZIK1	1.016	-1.147	3.46E-03
LRRC1	4.089	1.210	1.62E-05	GPR65	-0.004	-1.577	3.48E-03
PLCD3	3.882	1.999	1.67E-05	ARHGAP9	2.381	-1.620	3.49E-03
ELF3	7.482	2.566	1.74E-05	LINC02600	0.096	-1.904	3.49E-03
RNF166	4.011	-1.153	1.74E-05	ENSG 00000239467	0.238	1.607	3.52E-03
PLS1	4.563	2.201	1.74E-05	PRELID3B	5.056	0.759	3.52E-03
S100A14	4.377	2.877	1.79E-05	ARFGEF3	5.505	1.377	3.52E-03
BAALC	0.819	-1.917	1.81E-05	TTYH2	3.308	-1.109	3.52E-03
FADS3	4.573	-1.349	1.84E-05	IRX5	-1.853	3.936	3.54E-03
LRRN2	2.280	-1.584	1.87E-05	TNFAIP8L2	0.229	-1.691	3.55E-03
SBK1	1.989	-1.756	1.87E-05	ECHDC3	3.528	-1.454	3.55E-03
CDCP1	5.278	1.820	2.06E-05	STYK1	-0.717	2.956	3.62E-03
FHL2	4.271	1.752	2.06E-05	IL18RAP	-0.477	-1.794	3.68E-03
KCNK1	3.573	2.355	2.10E-05	SCN11A	-1.609	-1.857	3.68E-03
LAMB3	4.862	2.425	2.12E-05	SEMA6A-AS1	0.053	-1.274	3.71E-03
CCM2L	0.795	-1.624	2.28E-05	ALPP	-3.894	6.311	3.71E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
MST1R	3.982	2.273	2.38E-05	ENSG 00000260328	-3.266	3.225	3.71E-03
TSPAN8	5.554	3.013	2.38E-05	MS4A7	3.005	-1.953	3.72E-03
ENSG 00000268555	0.828	-3.265	2.38E-05	BRINP2	-3.249	-2.387	3.74E-03
KIF17	-0.242	-1.492	2.38E-05	NAT8L	2.526	-1.783	3.77E-03
TMC4	6.244	1.739	3.94E-05	KRT15	1.782	3.575	3.77E-03
STAB2	-1.326	-2.536	4.27E-05	BICDL2	6.614	2.123	3.79E-03
BCL11A	2.261	-1.966	4.54E-05	LRRC25	1.284	-1.621	3.81E-03
INAVA	4.648	2.234	4.54E-05	VSIG2	2.019	3.395	3.81E-03
CAPN5	4.726	2.208	4.75E-05	ENSG 00000259436	-0.916	-1.914	3.81E-03
CDH4	0.671	-2.063	4.84E-05	STX3	5.188	0.749	3.82E-03
LMO7	6.051	1.567	5.13E-05	EPB42	-1.018	-1.445	3.83E-03
SMIM22	3.880	2.701	5.13E-05	FNBP1	4.487	-1.007	3.86E-03
TMEM131L	5.049	-1.714	5.13E-05	ENSG 00000267040	0.032	-1.007	3.86E-03
CDS1	3.915	1.654	5.41E-05	RASAL3	2.022	-1.540	3.87E-03
MPZL2	4.704	1.705	5.45E-05	ZNF589	3.734	-0.779	3.87E-03
SCNN1A	6.458	2.167	5.55E-05	ITGAX	3.285	-1.797	3.87E-03
ITGA3	6.582	1.879	5.83E-05	PPP1R14D	-1.490	4.285	3.91E-03
NPY5R	-1.860	-2.907	6.04E-05	BCAS1	2.093	3.239	3.92E-03
CDC42BPA	5.895	1.032	6.05E-05	SCML4	0.078	-1.675	3.97E-03
MARCHF1	3.072	-1.494	6.66E-05	CELF2	3.424	-1.640	4.00E-03
KRT8	9.626	1.826	6.91E-05	TMBIM1	8.315	0.959	4.00E-03
LPXN	3.350	-1.301	6.91E-05	ENSG 00000185332	-2.266	3.079	4.02E-03
ABI3	2.347	-1.553	7.37E-05	VDAC1	6.809	0.709	4.05E-03
NAGK	5.364	-0.806	7.37E-05	RIN1	3.360	1.431	4.06E-03
ARHGAP28	2.379	-1.675	7.37E-05	LRRC8A	5.983	1.076	4.07E-03
NPL	2.424	-1.414	7.76E-05	SOX9	6.126	1.160	4.09E-03
PLEKHA6	5.534	1.478	7.76E-05	RENBP	1.508	-1.357	4.09E-03
PYGB	6.578	1.411	7.76E-05	MYO1D	6.629	0.836	4.11E-03
PCDH1	5.639	1.850	8.29E-05	B4GALT4	5.015	0.856	4.12E-03
EMP2	5.098	1.278	8.87E-05	LIMD1	4.885	-0.654	4.13E-03
STEAP2	4.571	1.723	8.87E-05	GOLT1A	2.836	2.177	4.18E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
ITGAD	-3.630	-2.806	8.93E-05	PLEKHG6	3.182	1.679	4.20E-03
QSOX1	7.796	1.393	9.45E-05	ENSG 00000267287	-2.461	-1.964	4.24E-03
EVL	5.661	-1.503	9.99E-05	SERINC2	5.983	1.561	4.25E-03
EPB41L1	5.691	1.219	9.99E-05	AKAP7	4.506	-1.612	4.29E-03
GRHL2	4.657	1.873	9.99E-05	MAB21L4	0.266	3.828	4.32E-03
LRP5	6.259	1.163	1.03E-04	HMBOX1	3.452	-0.539	4.33E-03
S100P	1.335	6.394	1.08E-04	SH2B3	4.649	-0.907	4.33E-03
RNF144B	3.419	-1.200	1.09E-04	CCND1	6.302	1.209	4.33E-03
S100A10	6.582	1.585	1.11E-04	DDR1	7.691	0.956	4.33E-03
GRK3	4.436	-1.215	1.13E-04	P2RY13	-0.307	-2.072	4.34E-03
RASGRP3	3.125	-1.617	1.16E-04	GTF2IRD1	4.322	1.165	4.44E-03
CAPN2	7.524	0.972	1.16E-04	CARD9	0.915	-1.373	4.46E-03
COL17A1	1.844	6.002	1.21E-04	LINC02747	-4.355	3.383	4.50E-03
BMF	3.981	-1.578	1.27E-04	TMEM63B	4.802	0.741	4.51E-03
KIAA1522	7.458	1.363	1.30E-04	OLFM1	3.171	-1.405	4.58E-03
MARVELD2	4.222	1.586	1.31E-04	FGD2	2.512	-1.510	4.59E-03
FAT1	6.767	1.496	1.35E-04	SIGLEC7	-1.224	-1.883	4.62E-03
RNF128	4.743	2.353	1.38E-04	FTH1P20	-0.217	-1.087	4.64E-03
GGA2	6.949	-0.928	1.38E-04	ARHGEF6	3.320	-1.211	4.66E-03
MET	5.275	1.851	1.39E-04	PPM1M	3.803	-0.713	4.72E-03
PITX1	0.276	6.217	1.39E-04	ZNF100	2.470	-0.820	4.72E-03
LEFTY2	-3.055	-3.376	1.39E-04	OGFRL1	4.430	-0.910	4.73E-03
MIR600HG	2.312	-1.402	1.39E-04	ACOT11	2.325	1.362	4.73E-03
IRF6	5.531	1.533	1.39E-04	TBC1D10C	1.237	-1.769	4.78E-03
OVOL2	2.563	2.317	1.44E-04	HSPBAP1	3.189	-0.597	4.79E-03
TLR1	2.449	-1.378	1.48E-04	KBTBD8	0.165	-1.382	4.83E-03
ENSG 00000261123	-0.846	3.007	1.48E-04	LRRC42	3.876	0.688	4.88E-03
PHACTR1	2.257	-1.312	1.49E-04	CEP170B	6.571	0.985	4.90E-03
CD72	1.815	-1.792	1.57E-04	OTUD7A	0.420	-0.976	4.92E-03
GNG7	3.554	-1.579	1.59E-04	YPEL1	1.197	-1.112	4.94E-03
SPINT2	7.954	1.215	1.61E-04	DEPDC5	3.285	-0.499	5.02E-03
SCCPDH	5.007	1.137	1.61E-04	TRPM4	5.029	1.117	5.02E-03
DOK4	5.090	1.205	1.61E-04	PIK3AP1	5.274	-1.536	5.04E-03
MSL3	4.163	-0.658	1.72E-04	AP1S1	4.533	0.791	5.07E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
TACC2	5.160	1.378	1.72E-04	MPP1	4.452	-1.122	5.08E-03
BAIAP2L1	5.877	1.849	1.72E-04	BLZF1	3.783	0.746	5.10E-03
ACSM5	-0.785	-2.191	1.72E-04	HSPA12B	2.330	-1.071	5.13E-03
DOK3	2.432	-1.727	1.72E-04	SLC7A11	3.352	2.529	5.13E-03
ZNF208	0.351	-1.863	1.73E-04	AOAH	1.233	-1.765	5.14E-03
SPATA13	6.279	-1.066	1.84E-04	MIR155HG	-1.478	-1.699	5.17E-03
CD33	0.402	-1.833	1.92E-04	CKMT1B	1.108	2.667	5.17E-03
ESRP2	5.585	1.786	1.96E-04	CSF3R	2.685	-1.974	5.17E-03
SLC47A1	-0.592	-1.881	1.96E-04	FAM166C	-0.664	2.555	5.17E-03
TMEM30B	5.656	1.441	1.99E-04	FCRL6	-1.724	-2.116	5.17E-03
PRKCI	4.626	1.190	2.00E-04	KCNAB2	3.512	-1.255	5.17E-03
EXOC3L1	1.551	-1.104	2.12E-04	APOBR	2.020	-1.438	5.19E-03
KDM2B	4.204	-0.649	2.12E-04	TDRD10	-0.180	-1.397	5.24E-03
MATK	0.358	-1.823	2.12E-04	ANKRD44	2.267	-1.177	5.25E-03
BAIAP2	4.173	1.490	2.12E-04	DSG2	6.310	1.276	5.31E-03
MYO9B	6.009	-0.811	2.12E-04	CXCL6	1.880	-2.723	5.36E-03
PRKAR2B	3.304	-1.621	2.21E-04	SPI1	2.906	-1.649	5.36E-03
ANO1	4.097	1.870	2.24E-04	FAM189A1	2.584	-1.545	5.36E-03
TMEM170B	2.340	-1.265	2.24E-04	FSD1L	2.377	-0.909	5.36E-03
LINC00857	-0.586	2.456	2.24E-04	SIGLEC5	-1.787	-2.146	5.36E-03
LRP11	4.863	1.127	2.28E-04	ENSG 00000262222	-0.527	-1.219	5.36E-03
IQANK1	3.413	2.643	2.40E-04	ERICH2	-0.582	1.796	5.36E-03
IFFO1	4.033	-1.175	2.41E-04	PLA2G4C	2.432	-1.276	5.40E-03
NCAM1	2.613	-2.117	2.51E-04	LINP1	-3.972	3.524	5.40E-03
ALKAL2	3.264	-2.537	2.61E-04	TDRD6	0.935	-1.407	5.40E-03
PARVG	2.760	-2.050	2.68E-04	ENSG 00000261662	0.045	-1.101	5.40E-03
EPS8	5.315	1.180	2.72E-04	GDA	2.772	2.543	5.45E-03
CBLC	4.268	2.793	2.72E-04	MAN2B1	6.172	-0.694	5.45E-03
KRT7	6.715	2.639	2.80E-04	ARHGAP24	4.036	-1.033	5.46E-03
VAV3	3.678	-1.632	2.89E-04	NFAM1	1.923	-1.566	5.46E-03
LINC01857	-4.333	-2.647	2.91E-04	MYOM3	1.700	2.631	5.47E-03
PLEK2	3.377	2.292	2.91E-04	SASH3	2.194	-1.888	5.47E-03
RGL4	0.526	-1.268	2.91E-04	TM9SF3	7.582	0.725	5.52E-03
KRT18	8.515	1.586	2.91E-04	TMX3	4.943	-0.620	5.54E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
FAM49A	2.631	-1.450	2.91E-04	GFPT1	6.699	1.005	5.59E-03
PSTPIP2	2.561	-1.311	2.91E-04	CDH1	7.975	1.029	5.59E-03
CTNND2	3.973	-2.077	2.91E-04	ABCA12	-2.189	3.734	5.59E-03
STEAP1	2.694	2.097	3.20E-04	ICA1	5.793	0.885	5.60E-03
LGALS3	6.536	1.497	3.23E-04	DNAJA4	3.886	1.282	5.60E-03
CASP17P	1.602	-1.604	3.25E-04	ZIC2	-3.288	5.069	5.60E-03
NFIA-AS2	-3.693	-2.521	3.40E-04	GOLM1	6.704	1.143	5.65E-03
NGEF	3.244	2.212	3.43E-04	THEM6	4.786	1.113	5.66E-03
EEF1A1P24	-2.935	-2.159	3.43E-04	ZDHHC3	5.406	0.684	5.67E-03
NLRC4	0.557	-1.481	3.43E-04	SEMA4G	4.874	1.647	5.67E-03
PTK6	3.081	3.236	3.45E-04	PIK3R5	2.250	-1.640	5.81E-03
CTTN	7.504	0.801	3.53E-04	LMNA	7.664	0.919	5.81E-03
LSR	7.170	1.387	3.72E-04	PLEKHO1	3.572	-1.201	5.82E-03
FGR	2.791	-1.987	3.81E-04	SHROOM1	4.748	1.326	5.85E-03
MANSC1	5.015	1.284	3.81E-04	FOXL1	-0.095	3.540	5.85E-03
RIC3	3.648	-1.816	3.81E-04	PLAAT2	-2.702	4.211	5.85E-03
FOXQ1	2.885	3.114	3.86E-04	NCKAP1L	2.750	-1.865	5.85E-03
KCNN4	2.068	2.891	3.88E-04	WNT7B	1.345	2.729	5.86E-03
ENSG 00000273132	-3.447	3.918	3.88E-04	KCNT1	-1.071	-1.475	5.87E-03
ENSG 00000260877	-2.690	6.581	3.88E-04	GJB5	-1.054	4.835	5.88E-03
EPS8L2	6.735	1.388	4.12E-04	ERN2	1.766	5.012	5.88E-03
PLPP2	5.162	1.570	4.17E-04	NDFIP2	4.997	0.796	5.90E-03
EXPH5	2.660	1.543	4.21E-04	TP53INP1	6.499	-1.162	5.92E-03
C1orf210	3.795	2.505	4.65E-04	TMEM125	5.412	1.813	5.94E-03
GJB3	0.682	3.190	4.66E-04	SSH2	4.347	-0.739	5.94E-03
ARHGAP17	4.750	-0.535	4.66E-04	TXN	6.051	0.924	5.95E-03
SLC2A1	4.642	2.172	4.68E-04	FMNL3	4.312	-0.973	6.01E-03
PTPN3	5.438	1.250	4.68E-04	JARID2	3.678	-0.633	6.01E-03
CLDN7	6.264	1.648	4.76E-04	CLTB	5.968	1.093	6.01E-03
TUFT1	4.861	1.550	4.88E-04	RUBCNL	-0.393	-1.684	6.02E-03
BACH2	2.297	-1.638	4.96E-04	TMEM273	2.140	-1.169	6.05E-03
MGST1	5.543	2.055	5.03E-04	WDR91	5.340	-0.873	6.05E-03
MYEOV	2.743	5.741	5.03E-04	CLECL1	-2.846	-2.200	6.07E-03
ADCY5	4.146	-1.550	5.03E-04	LILRA4	-3.287	-2.511	6.07E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
PDE7A	4.606	-1.158	5.04E-04	KLK10	0.924	4.667	6.08E-03
TFEC	0.935	-2.033	5.42E-04	NPAS2	5.498	1.181	6.10E-03
CLDN12	5.055	1.218	5.47E-04	AK4	2.134	2.225	6.12E-03
ENSG 00000257512	-3.893	3.275	5.50E-04	PCED1B-AS1	0.273	-1.586	6.12E-03
ATP1B1	7.509	1.473	5.80E-04	ANXA2	8.597	1.131	6.12E-03
SLPI	4.217	2.634	5.80E-04	ADIPOR1	6.422	0.589	6.12E-03
TRAF1	3.767	-1.265	5.88E-04	ENSG 00000251034	0.195	-1.039	6.12E-03
FA2H	3.785	2.438	5.88E-04	TDRP	3.384	-0.892	6.20E-03
F2RL1	4.880	1.613	5.88E-04	ZNF736	3.239	-0.871	6.20E-03
MYRF	6.742	1.633	5.88E-04	SYK	4.937	-0.840	6.21E-03
CD36	4.057	-2.356	5.88E-04	FOXA1	0.607	3.095	6.25E-03
ITGB6	4.835	2.201	5.88E-04	NXF1	6.538	-0.585	6.28E-03
ENSG 00000233038	-0.588	-1.776	5.97E-04	EML2	5.197	0.771	6.28E-03
SCIMP	0.266	-2.066	6.17E-04	CDR2L	3.860	1.163	6.33E-03
CHRNA5	0.007	2.518	6.17E-04	CD37	2.977	-1.797	6.34E-03
MARVELD3	4.187	1.494	6.26E-04	BCAR3	3.236	1.223	6.35E-03
SCARF1	3.113	-0.954	6.38E-04	AJAP1	1.976	-2.282	6.38E-03
ARHGAP32	5.972	1.216	6.43E-04	WDFY4	1.566	-1.999	6.38E-03
FAM83E	1.745	3.731	6.43E-04	FLT3	1.738	-1.839	6.40E-03
CTSE	3.876	3.834	6.45E-04	MCOLN1	3.906	-0.621	6.50E-03
PDE1B	1.838	-1.348	6.45E-04	KMO	-1.507	-1.789	6.50E-03
HMOX1	3.049	-1.805	6.48E-04	PYCR1	5.553	1.199	6.55E-03
PLAAT3	5.126	1.460	6.63E-04	ZDHHC21	4.310	-0.711	6.55E-03
CYFIP2	3.961	-1.418	6.72E-04	BTC	1.858	1.278	6.55E-03
NMU	-2.373	5.838	6.76E-04	S100A5	-4.015	2.103	6.56E-03
BTN2A2	3.042	-1.030	6.79E-04	TM6SF1	0.752	-1.343	6.58E-03
ZNF671	2.212	-1.055	7.03E-04	ETV4	3.269	2.346	6.60E-03
FGF12	1.634	-1.565	7.04E-04	TRGC1	-2.886	-2.158	6.60E-03
PROM2	5.336	2.200	7.12E-04	PLXNA2	7.039	1.445	6.60E-03
SMAGP	3.605	1.529	6.47E-04	ENSG 00000261762	-1.818	2.312	6.50E-03
SHANK3	4.723	-1.267	6.48E-04	SMAP2	5.158	-1.070	6.52E-03
TMEM268	3.386	-0.819	6.48E-04	LINC00987	0.467	-1.401	6.53E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
TMPRSS3	3.711	2.005	7.64E-04	SLC2A6	2.345	-1.217	6.61E-03
LGALS4	5.867	2.737	7.64E-04	FBXO10	3.320	-0.847	6.63E-03
ASPH	6.298	1.223	7.96E-04	ZNF22	4.387	-0.549	6.77E-03
DIRAS1	2.231	-1.938	8.02E-04	GPX2	4.413	2.736	6.85E-03
BEND5	1.466	-1.157	8.16E-04	ADGRE1	-2.409	-2.273	6.85E-03
LPCAT4	4.969	1.572	8.18E-04	ID1	4.758	1.851	6.85E-03
MALT1	3.977	-0.743	8.18E-04	ZNF90	0.658	-1.158	6.85E-03
ENSG 00000267243	-0.171	-2.836	8.24E-04	ENSG 00000249042	2.130	1.012	6.85E-03
DTX1	3.313	-1.256	8.52E-04	ENSG 00000261437	0.686	1.624	6.85E-03
PRR15L	5.987	2.286	8.76E-04	PALD1	3.876	-1.350	6.85E-03
MAP3K8	3.029	-1.132	9.09E-04	ORAI2	4.284	-0.777	6.87E-03
MBOAT2	3.369	1.605	9.09E-04	HFM1	-0.304	-2.038	6.87E-03
LIPA	5.330	-1.235	9.09E-04	SPATS2L	6.270	0.734	6.87E-03
ATG16L2	4.562	-1.257	9.10E-04	SLC4A11	1.263	2.710	6.87E-03
SH3TC2	-0.343	2.629	9.15E-04	CD4	3.977	-1.422	6.87E-03
ENSG 00000259933	-1.719	2.777	9.15E-04	ZFP82	0.935	-0.985	6.90E-03
PKP4	5.719	0.684	9.15E-04	CALML4	3.561	1.366	6.90E-03
CASC8	-2.920	3.702	9.21E-04	TRAF3IP3	1.176	-1.554	6.91E-03
TOB1	6.721	1.630	9.22E-04	CTC1	4.750	-0.731	6.92E-03
TSTA3	6.251	1.333	9.43E-04	ZSCAN30	3.789	-0.603	6.95E-03
PLBD1	5.182	1.320	9.43E-04	ENSG 00000255197	-1.161	-1.736	6.95E-03
CORO2A	2.101	1.973	9.43E-04	LINC01754	-2.333	-2.267	6.96E-03
TTC22	4.148	1.276	9.43E-04	ZZEF1	6.063	-0.652	6.98E-03
TMC7	1.360	2.142	9.43E-04	FIGN	2.916	-1.032	6.98E-03
P2RX7	1.312	-1.530	9.43E-04	SCN4A	-0.824	-2.046	6.98E-03
PRR15	2.395	2.233	9.43E-04	C1GALT1	4.219	0.935	6.98E-03
FUT3	3.697	2.408	9.50E-04	TMEM92	1.005	2.389	7.02E-03
FAM78A	1.713	-1.593	9.52E-04	SPIRE2	2.854	2.131	7.04E-03
ARAP1	6.257	-0.688	9.61E-04	L3HYPDH	3.005	-0.830	7.04E-03
SLC44A4	4.910	2.297	9.61E-04	SDCBP2	4.263	1.837	7.04E-03
PTPN6	4.873	-0.968	9.61E-04	CARD8-AS1	0.716	-1.016	7.05E-03
RASSF4	5.883	-1.803	9.61E-04	ITGAL	2.155	-1.945	7.05E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
ZNF486	2.182	-1.503	9.62E-04	TACSTD2	6.918	1.875	7.05E-03
CPD	6.294	1.000	9.62E-04	GM2A	5.018	-0.885	7.05E-03
C19orf38	-0.243	-1.531	9.62E-04	MAP4K1	1.112	-1.582	7.05E-03
KCNAB1	2.093	-1.411	9.64E-04	ME2	4.420	-0.729	7.05E-03
LAMC2	5.090	2.106	9.70E-04	CST4	-4.088	3.219	7.05E-03
RIPK4	4.588	1.654	9.79E-04	DOCK10	2.562	-1.471	7.05E-03
NECTIN4	3.960	2.069	9.79E-04	ZNF737	2.321	-1.150	7.09E-03
SV2B	2.862	-2.057	9.87E-04	LINC00511	3.127	1.332	7.12E-03
ZNF154	2.254	-1.112	9.97E-04	GALNT3	5.443	1.072	7.13E-03
SLC39A4	3.774	2.343	9.97E-04	EHF	4.995	1.556	7.13E-03
CERS4	5.132	-1.304	9.97E-04	PSAP	9.557	-0.723	7.13E-03
ADAMTS10	4.091	-1.552	9.97E-04	PPME1	4.661	0.504	7.26E-03
LILRB1	1.006	-1.994	9.99E-04	NTF4	-3.364	4.105	7.28E-03
TJP3	4.379	1.684	1.03E-03	SEMA4D	4.679	-0.757	7.29E-03
SYT13	3.944	2.878	1.03E-03	GIMAP4	3.270	-1.211	7.29E-03
PNLDC1	-0.953	-2.385	1.03E-03	NEBL	4.817	1.385	7.30E-03
IL20RA	3.751	2.015	1.03E-03	FAM83B	0.994	1.917	7.30E-03
PRAM1	0.938	-1.704	1.04E-03	PAK4	5.529	0.909	7.30E-03
CARD8	4.545	-0.846	1.09E-03	HTRA4	-2.871	-2.089	7.35E-03
EPS8L1	5.014	2.014	1.10E-03	SPINT1	6.941	1.050	7.35E-03
ZBTB46	3.136	-0.755	1.10E-03	TDRD1	-0.871	-2.159	7.35E-03
SOWAHD	-1.486	-1.806	1.10E-03	ZAP70	1.085	-1.670	7.35E-03
SCNN1G	1.212	-3.144	1.10E-03	VASH1	3.491	-1.021	7.37E-03
PIK3CD	3.405	-1.345	1.10E-03	LAT2	2.493	-1.280	7.37E-03
BCL2	2.811	-1.460	1.10E-03	CLEC4A	0.414	-1.156	7.37E-03
ROBO4	3.454	-1.073	1.11E-03	GPRC5D	-1.170	2.239	7.37E-03
HCST	0.518	-1.578	1.11E-03	GJB2	2.461	1.947	7.42E-03
AKAP1	5.671	0.906	1.13E-03	LEMD1	-2.653	3.639	7.44E-03
LY86	0.561	-2.036	1.13E-03	CNNM1	1.968	2.094	7.48E-03
LILRA1	-1.192	-2.213	1.13E-03	FTLP3	0.194	-1.256	7.49E-03
NIPSNAP3B	1.352	-1.065	1.13E-03	SIM2	0.835	2.377	7.49E-03
CX3CR1	0.193	-2.350	1.13E-03	SIGLEC9	0.399	-1.601	7.49E-03
ADPRH	3.452	-0.965	1.13E-03	DNAH3	-0.658	2.478	7.51E-03
OCIAD2	4.807	1.207	1.13E-03	GDF7	-0.527	-2.031	7.56E-03
GYPC	4.260	-1.270	1.13E-03	ZMAT1	4.433	-1.200	7.56E-03
IL34	1.208	-1.419	1.13E-03	ASPHD1	3.273	1.708	7.56E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
FHOD1	4.646	-0.774	1.14E-03	ODR4	5.155	0.691	7.58E-03
MYO1F	3.227	-1.729	1.14E-03	PPM1F	4.902	-0.662	7.64E-03
SEZ6L2	5.628	1.278	1.14E-03	MCOLN2	0.470	-1.817	7.68E-03
DTNA	4.112	-1.357	1.14E-03	SLC9A5	0.628	-1.003	7.78E-03
DDHD1	3.234	-0.810	1.15E-03	CNNM4	4.485	0.897	7.80E-03
ARRB2	4.424	-1.113	1.15E-03	THEMIS2	3.330	-1.492	7.80E-03
GSTO2	4.529	1.386	1.15E-03	CEACAM21	-1.351	-1.577	7.83E-03
RSPH1	1.315	1.715	1.15E-03	CHMP4C	4.145	1.240	7.89E-03
FKBP15	5.283	-0.472	1.16E-03	LAIR1	3.001	-1.659	7.89E-03
CPAMD8	1.612	-1.283	1.17E-03	PKNOX2	2.392	-1.224	7.89E-03
GIMAP8	2.737	-1.432	1.18E-03	GMFG	2.401	-1.289	7.89E-03
SH2D3C	3.041	-1.229	1.18E-03	SLC52A3	1.871	1.600	7.91E-03
KANK3	2.130	-1.293	1.20E-03	AP1S2	3.446	-0.812	7.92E-03
WDR37	4.366	-0.629	1.20E-03	IFFO2	5.419	-0.818	7.92E-03
VASH2	-0.215	-1.423	1.22E-03	TBX21	-1.523	-1.742	7.92E-03
TNS4	0.822	5.662	1.24E-03	LINC00909	2.345	-0.686	7.92E-03
ENSG 00000182109	2.638	-1.208	1.24E-03	IL17RE	4.099	1.534	7.92E-03
RAPGEFL1	4.261	2.546	1.27E-03	NIN	4.435	-0.848	7.93E-03
JPH1	3.176	2.276	1.28E-03	CRB3	4.004	1.345	7.95E-03
AP1B1	6.472	-0.684	1.28E-03	SPTBN2	5.667	1.231	7.99E-03
PLEKHA5	4.783	1.052	1.28E-03	SETMAR	3.069	-0.563	8.02E-03
ITGAE	2.814	-0.713	1.29E-03	SH3D19	5.375	0.836	8.05E-03
MICU3	2.161	-1.165	1.29E-03	CORO7	4.094	-0.789	8.06E-03
FGD3	3.912	-1.323	1.29E-03	IFNLR1	3.237	-0.853	8.08E-03
STX1A	3.609	2.148	1.29E-03	ZNF676	-0.291	-1.591	8.14E-03
ADRA1D	-1.751	-2.282	1.29E-03	TMEM217	-1.158	-1.705	8.18E-03
GGTLC1	-3.195	-2.763	1.29E-03	TYK2	6.196	-0.563	8.20E-03
APBB1IP	2.573	-1.743	1.29E-03	TOR1B	4.440	0.604	8.26E-03
ITGA6	7.094	1.400	1.29E-03	GAPT	-0.665	-2.086	8.26E-03
SFMBT2	2.585	-1.225	1.31E-03	FTLP2	-2.742	-1.373	8.28E-03
RHOD	4.425	1.688	1.31E-03	TM9SF4	6.072	0.478	8.29E-03
PDXDC1	6.958	0.645	1.31E-03	TINAG	-4.213	4.778	8.29E-03
CAMK1D	5.238	-1.299	1.35E-03	PTPRO	0.114	-1.474	8.30E-03
ECT2	2.876	1.561	1.39E-03	HR	1.870	1.981	8.31E-03
SNX10	3.144	-1.767	1.40E-03	HLA-DMB	3.442	-1.543	8.31E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
CYB561A3	4.555	-0.692	1.43E-03	FAM53B	4.456	-0.865	8.32E-03
VILL	3.691	2.171	1.43E-03	HNF4A	5.044	2.048	8.35E-03
YIPF6	5.447	0.633	1.43E-03	ITPKB	4.384	-0.741	8.35E-03
SPATC1	-2.882	-2.104	1.43E-03	CREB3L1	6.577	1.724	8.35E-03
ENSG 00000261116	1.168	1.886	1.44E-03	DEF8	5.623	-0.478	8.35E-03
USP44	1.284	-1.553	1.45E-03	N4BP2L1	4.000	-0.829	8.35E-03
BTK	1.235	-1.862	1.45E-03	PREX1	4.258	-1.204	8.37E-03
GDPD5	5.070	-1.275	1.45E-03	AQP5	1.700	4.830	8.48E-03
PHOSPHO1	-1.743	-1.939	1.46E-03	LINC00092	-0.465	-1.542	8.48E-03
FAM110B	1.885	-1.053	1.47E-03	LCP2	2.911	-1.426	8.49E-03
ELMO3	5.608	1.268	1.47E-03	AIF1	2.462	-1.578	8.51E-03
OVOL1	1.610	3.171	1.52E-03	PRSS22	4.266	1.698	8.61E-03
PDE2A	3.089	-1.430	1.52E-03	MGLL	5.252	1.153	8.62E-03
SSH3	5.276	0.997	1.53E-03	SIGLEC11	0.356	-2.653	8.63E-03
SRD5A3	4.486	1.333	1.53E-03	TNFRSF21	5.617	1.223	8.63E-03
RNF146	4.553	-0.572	1.55E-03	ADGRE2	0.997	-1.482	8.72E-03
TLCD1	2.178	1.548	1.55E-03	PPP1R9B	5.569	-0.553	8.77E-03
MIR503HG	0.113	-1.660	1.55E-03	CNDP2	6.840	-0.679	8.77E-03
ENSG 00000223813	-0.818	2.171	1.55E-03	ENSG 00000271730	-1.842	-1.673	8.78E-03
GIMAP1	1.939	-1.388	1.55E-03	SLC44A1	6.207	0.837	8.79E-03
GPR39	1.538	1.748	1.55E-03	MEF2C	3.257	-1.150	8.85E-03
FIG4	3.626	-0.580	1.60E-03	CADM3-AS1	-1.241	-2.434	8.85E-03
ALDH3B1	4.401	1.475	1.60E-03	RGS19	2.925	-0.882	8.86E-03
CD160	-1.598	-1.511	1.60E-03	PHF10	4.751	-0.522	8.91E-03
AHNAK2	3.932	2.199	1.61E-03	LINC01480	-1.007	-1.617	8.94E-03
XDH	1.522	3.048	1.61E-03	NCF1	-0.319	-1.718	8.94E-03
ANKDD1A	3.037	-0.961	1.65E-03	PIAS2	3.639	-0.520	8.94E-03
CRHBP	-2.548	-2.291	1.66E-03	TNFAIP3	5.064	-1.289	8.98E-03
MSLN	1.552	5.366	1.66E-03	PRRG1	2.548	1.035	9.00E-03
SLC35A2	4.787	0.845	1.66E-03	CRNDE	1.406	2.278	9.00E-03
LILRB2	1.572	-1.808	1.67E-03	WDR81	4.819	-0.734	9.08E-03
CLN8	4.415	-0.727	1.67E-03	EEF1A1P9	0.256	-1.031	9.10E-03
UGT1A10	-2.805	7.777	1.68E-03	MCU	4.358	1.013	9.10E-03
NRROS	0.910	-1.426	1.70E-03	MMP2-AS1	-3.721	-2.370	9.10E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
PIP4K2A	4.270	-0.813	1.71E-03	ENSG 00000265206	-1.137	-2.092	9.10E-03
RUSC1	4.922	0.780	1.71E-03	GRB2	5.565	-0.517	9.13E-03
LINC01269	-4.454	2.542	1.71E-03	TLR7	-0.463	-1.966	9.18E-03
ZNF43	2.753	-1.166	1.71E-03	FES	3.806	-0.921	9.19E-03
ENSG 00000265743	-2.446	-1.651	1.71E-03	SLC15A3	3.234	-1.123	9.21E-03
MYOF	5.711	1.244	1.73E-03	ENSG 00000232756	-4.541	3.439	9.24E-03
SDR16C5	0.302	5.097	1.73E-03	MYO5A	4.399	-0.899	9.25E-03
STAC3	0.891	-1.141	1.73E-03	CD1D	0.811	-1.390	9.39E-03
FUT2	3.525	2.098	1.78E-03	DCBLD2	4.718	1.416	9.43E-03
C1orf162	2.908	-1.481	1.78E-03	TRIM31-AS1	-3.673	2.731	9.44E-03
TBC1D9	4.488	-1.028	1.78E-03	LRRC31	-3.144	4.946	9.44E-03
SLC25A30	3.502	-0.759	1.78E-03	GALNT5	-0.021	3.752	9.49E-03
GIMAP5	1.792	-1.546	1.78E-03	HCLS1	4.355	-1.635	9.49E-03
TMEM54	4.353	1.311	1.79E-03	CYB5R4	3.321	-0.567	9.53E-03
CD101	0.818	-1.176	1.80E-03	MAMLD1	3.064	-0.849	9.54E-03
CD300A	1.690	-1.567	1.81E-03	NEURL3	3.447	-1.774	9.57E-03
ENSG 00000260317	-0.026	-1.306	1.82E-03	NLRC3	1.463	-1.169	9.60E-03
CAPN8	2.754	4.897	1.83E-03	INPP5B	4.152	-0.596	9.66E-03
KDELR2	7.455	0.888	1.83E-03	MICAL1	4.920	-1.029	9.66E-03
HMGA1	5.866	1.327	1.84E-03	TJP2	5.977	0.817	9.66E-03
HVCN1	0.756	-1.433	1.85E-03	WASF3	3.119	-0.907	9.66E-03
PDE6B	0.765	-1.381	1.85E-03	SYNE1	5.912	-1.119	9.66E-03
ZNF135	2.122	-1.041	1.86E-03	PRELID2	2.069	1.217	9.66E-03
LINC02577	-3.764	4.665	1.88E-03	CD55	5.422	1.467	9.66E-03
MAPK3	5.284	0.785	1.88E-03	ZBTB14	3.611	-0.539	9.66E-03
ZNF257	-0.015	-1.649	1.89E-03	ENSG 00000232759	-2.493	-1.952	9.66E-03
HSPB8	3.770	-1.479	1.89E-03	PILRA	1.647	-1.338	9.66E-03
OXER1	1.794	-1.403	1.90E-03	IL10RA	3.191	-1.509	9.66E-03
ZNF710-AS1	1.665	-1.284	1.90E-03	SGSH	5.060	-0.688	9.80E-03
USH1C	3.625	2.545	1.92E-03	ANKRD33B	-1.615	-1.673	9.80E-03
CMBL	4.554	1.794	1.93E-03	BIN2	1.514	-1.433	9.83E-03

Table A.5. All DEGs with Adjusted p-Values Below 0.01, Ensembl Gene IDs have been provided if HGNC Gene Symbols were not found. (cont.)

HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val. (B.H.)	HGNC Gene Symbol/ENSG ID	Avg. Exp. (logCPM)	log2FC	Adj.P.Val (B.H.)
SPDEF	0.648	4.520	1.93E-03	ACSL5	4.263	1.336	9.86E-03
HCK	2.428	-1.880	1.96E-03	BEST1	1.976	-1.065	9.86E-03
NUP62CL	1.401	1.625	1.96E-03	ART4	1.209	-1.599	9.91E-03
LINC02313	-4.517	2.642	1.96E-03	GJB4	-1.600	3.627	9.92E-03
GIPC1	5.840	1.059	1.96E-03	ASAP2	3.493	1.430	9.92E-03
PLEK	2.013	-2.089	1.96E-03	ENSG 00000228643	-2.991	-1.970	9.92E-03
WAS	2.030	-1.702	1.96E-03	CNKSR1	4.760	1.040	9.94E-03
RASGRP1	1.986	-1.481	1.96E-03	PSTPIP1	1.123	-1.332	9.94E-03

Table A.6. Subsystem size differences between the cancer and healthy models.

Subsystem	Number of Reactions (Recon3D)	Number of Reactions (Healthy)	Number of Reactions (Cancer)	Abs. Difference (Cancer-Healthy)	Abs. Difference as Proportion of Subsystem Size
Protein formation	2	2	0	2	100.0%
ROS detoxification	7	1	6	5	71.4%
Keratan sulfate synthesis	22	11	22	11	50.0%
Thiamine metabolism	6	3	6	3	50.0%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Androgen and estrogen synthesis and metabolism	26	8	19	11	42.3%
Biotin metabolism	12	6	10	4	33.3%
Blood group synthesis	47	45	31	14	29.8%
CoA catabolism	4	2	3	1	25.0%
Heme degradation	4	3	4	1	25.0%
Tryptophan metabolism	45	25	35	10	22.2%
Vitamin B6 metabolism	11	2	0	2	18.2%
Tetrahydro-biopterin metabolism	18	18	15	3	16.7%
R group synthesis	6	4	5	1	16.7%
Glutathione metabolism	16	13	15	2	12.5%
NAD metabolism	27	21	24	3	11.1%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Vitamin B2 metabolism	9	8	9	1	11.1%
Transport, golgi apparatus	98	82	72	10	10.2%
Vitamin C metabolism	10	6	7	1	10.0%
Inositol phosphate metabolism	65	46	40	6	9.2%
Steroid metabolism	90	28	36	8	8.9%
Tyrosine metabolism	93	63	70	7	7.5%
Arachidonic acid metabolism	41	24	27	3	7.3%
Cytochrome metabolism	15	0	1	1	6.7%
Purine synthesis	15	15	14	1	6.7%
Keratan sulfate degradation	76	15	20	5	6.6%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Methionine and cysteine metabolism	46	41	38	3	6.5%
Glycine, serine, alanine, and threonine metabolism	47	29	32	3	6.4%
Pyruvate metabolism	32	28	30	2	6.3%
Glutamate metabolism	16	16	15	1	6.3%
Sphingolipid metabolism	133	93	85	8	6.0%
Bile acid synthesis	185	93	104	11	5.9%
Glyco-sphingolipid metabolism	71	40	44	4	5.6%
O-glycan metabolism	18	7	8	1	5.6%
Ubiquinone synthesis	18	13	14	1	5.6%
Heme synthesis	19	19	18	1	5.3%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Urea cycle	40	33	35	2	5.0%
Starch and sucrose metabolism	42	14	12	2	4.8%
Transport, endoplasmic reticular	211	164	174	10	4.7%
Chondroitin synthesis	45	43	45	2	4.4%
Valine, leucine, and isoleucine metabolism	45	42	44	2	4.4%
Fatty acid synthesis	239	169	160	9	3.8%
Lysine metabolism	31	23	22	1	3.2%
Purine catabolism	34	32	31	1	2.9%
Pyrimidine catabolism	34	22	21	1	2.9%
Phenylalanine metabolism	36	28	27	1	2.8%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Pentose phosphate pathway	41	35	34	1	2.4%
Chondroitin sulfate degradation	44	42	43	1	2.3%
Vitamin A metabolism	47	28	27	1	2.1%
Transport, lysosomal	105	87	89	2	1.9%
Folate metabolism	59	44	43	1	1.7%
Transport, nuclear	71	56	57	1	1.4%
Misc.	152	84	86	2	1.3%
Transport, peroxisomal	154	114	116	2	1.3%
Cholesterol metabolism	242	100	97	3	1.2%
Nucleotide interconv.	164	149	151	2	1.2%
Transport, extracellular	2652	1822	1847	25	0.9%
Fatty acid oxidation	961	747	738	9	0.9%

Table A.6. Subsystem size differences between the cancer and healthy models. (cont.)

Subsystem	**Number of Reactions (Recon3D)**	**Number of Reactions (Healthy)**	**Number of Reactions (Cancer)**	**Abs. Difference (Cancer-Healthy)**	**Abs. Difference as Proportion of Subsystem Size**
Transport, mitochondrial	453	372	368	4	0.9%
Glycero-phospholipid metabolism	165	66	67	1	0.6%