

5-2022

Identification of Synonymous Genes and Pathways Implicated in Irritable Bowel Disease and Pancreatic Duct Adenocarcinoma

Lavanya Uppala
uppala821@gmail.com

Follow this and additional works at: https://digitalcommons.unomaha.edu/university_honors_program

 Part of the [Bioinformatics Commons](#), and the [Biology Commons](#)

Recommended Citation

Uppala, Lavanya, "Identification of Synonymous Genes and Pathways Implicated in Irritable Bowel Disease and Pancreatic Duct Adenocarcinoma" (2022). *Theses/Capstones/Creative Projects*. 183.
https://digitalcommons.unomaha.edu/university_honors_program/183

This Dissertation/Thesis is brought to you for free and open access by the University Honors Program at DigitalCommons@UNO. It has been accepted for inclusion in Theses/Capstones/Creative Projects by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

Identification of Synonymous Genes and Pathways Implicated in Irritable Bowel Disease and Pancreatic Duct Adenocarcinoma

Lavanya Uppala

Abstract—Better understanding and genetic characterization of the gut microbiome will allow for the identification of clinically distinct gastrointestinal diseases. Facilitated by high throughput technologies, intestinal flora analyses have elucidated a broad spectrum of neuropsychiatric, immunological, and allergic disorders linked with this organ system. Microbiome research especially has shed light on underlying factors of intestinal disorders. This interplay of environmental bacteria versus host tissue gene expression may have implications for disease pathogenicity and etiological determination. For instance, pancreatic disorders are common symptoms of irritable bowel disease (IBD), which is thought to affect approximately 7% to 21% of the population [1]. However, IBD is often difficult to diagnose and treat as it presents with a constellation of symptoms, many of which are similar to other common gastrointestinal diseases. Thus, clinical diagnosis of IBD is typically only made when biomarkers for other common gastrointestinal diseases are absent. Similarly, pancreatic ductal adenocarcinoma (PDAC) accounts for over 90% of known malignancies of the pancreas [2]. PDAC also presents with a similar clinical prodrome to other common gastrointestinal illnesses, leading to late diagnosis. These potentially pleiotropic genes involved in such diseases can be identified as biomarkers through gene expression analyses to use in the form of a diagnostic tool. Such a mechanism may inform clinical decision-making, in turn leading to better patient outcomes.

Index Terms—irritable bowel disorder (IBD), irritable bowel syndrome (IBS), pancreatic duct adenocarcinoma (PDAC), gastrointestinal illness, gut microbiome, host tissue expression, gene expression, pathway analysis.

I. INTRODUCTION

THE gastrointestinal system is a complex amalgamation of organs that helps an individual digest food, process nutrients, and regulate various hormone levels in the body. Due to its function, this organ system is subject to influence from both the environment and intrinsic physiological factors of the patient. Current research identifies that the biggest of these factors are the gut microbiome and more recently the immune system which may cause the symptom of chronic inflammation, a often seen prodrome, or early clinical marker, of various gastrointestinal illnesses [3].

However, this interplay between the host and the gut microbiome complicates medical diagnoses, and makes it difficult to identify clinical biomarkers of disease with their presentations. For instance, irritable bowel disease (IBD), also known as irritable bowel syndrome (IBS), is marked by symptoms

including cramping, abdominal pain, bloating, gas, diarrhea, or constipation which are akin to many other common gastrointestinal disorders as severe as celiac disease, pancreatitis, ulcers, Crohn's disease, or as simple as the common stomach ache or appendicitis, and more [1]. In turn, individuals who have some food sensitivities or have a modicum of intestinal dysbiosis due to uncontrollable factors, present with similar symptoms [4]. This, IBD's constellation of symptoms and similarities to other common gastrointestinal illnesses, is what makes it difficult to diagnose and treat, with professional diagnosis often being difficult to achieve and undertreatment being rampant.

Historically, there have been many different approaches to the diagnosis of gastrointestinal disorders, including blood tests, endoscopies, colonoscopies, stool tests, barium swallowing radiography, CT or MRI scans, sonograms, biopsies, and many more [5]. However, such methods typically must be used in conjunction with each other, which may pose an unnecessarily invasive and expensive burden on the patient. Instead, using gene expression profile and pathway analysis of tissue samples from routine clinical procedures or from stool samples may provide an more easily obtainable research specimen and are often more insightful regarding the diseases' genetic and symptomatic heterogeneity. That is, gene expression panels, as diagnostic tools, have proven utility in the clinical sphere. One such panel was published by Plevy *et al* [6]. This group found that the A4-Fla2 and FlaX serological markers differentiated non-IBD, Crohn's disease, and ulcerative colitis (UC) patients with 87% accuracy (95% CI, p-value<0.001), while Crohn's vs. UC differentiation was as high as 93% (95% CI, p-value<0.001) [6]. More specifically, signaling of specific genes, such as IFN- γ in *Helicobacter pylori* infections, have been shown to differentially exacerbate disease severity [6], [7]. Thus, better characterization of the gene expression profiles of common gastrointestinal diseases serves the dual purpose of creating a pseudo-standardized scale of disease severity on which patients can be rated, as well as preventing potentially costly overtreatment [8].

Many gastrointestinal disorders also have root in the dysfunction of the gut microbiome, which impacts or regulates such disease pathways. Yet, given the incredible diversity and abundance of microbes on the planet; especially those that are implicated in human health, it is nearly impossible to connect physiological states with microbial community composition [9]. This effect is further confounded when one considers the fact that environmental factors, such as diet,

L. Uppala is with the College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE, 68182 USA.

Manuscript received May 6, 2022; revised May 6, 2022.

stress, exercise, medication use, are also significant regulators of the gut microbiome along with more intrinsic factors such as genetics and immune systems. Moreover, as the gut microbiome undergoes periods of growth and variation, gastrointestinal disorders vary in their severity, at times even having periods of remission without the necessity of medical intervention [10]. A study by Halfvarson *et al.* even found that the microbiomes of patients with IBD fluctuated more than those of healthy patients [11]. Further supporting this theory, Morgan *et al.* found that although microbial composition was altered in disease sample data sets compared to controls, microbial function was more often altered, with major changes occurring with nutrient transport and uptake [12]. Accordingly, this provides initial evidence that altered gut microbiome conditions, and thus altered bacterial mechanistic functions and pathways, have large scale involvement in gastrointestinal disease pathogenesis.

However, many studies have also intimated that there is an interplay between immunological response to food-borne stimuli involved in disorders under the IBD umbrella. In other words, some groups believe that IBD-related illnesses may simply be an immune reaction to diet or environment. While this juxtaposition of causes, between the gut microbiome and the immune system, is well known for this group of illnesses, it is not well characterized. As aforementioned, due to the environmental and symptomatic variation of patients with gastrointestinal disease, cross-sectional studies are difficult to design, likely contributing to the lack of data regarding these interactions. However, bioinformatics methodologies are beginning to allow for solutions to these difficulties with the ability to integrate “big data”, which can account for widespread public health concerns. More specifically, RNA sequencing analyses show great utility in this realm. RNA, or more specifically tRNA, are core components of all living cells and can be used to identify the taxonomic grouping of different organisms [9]. RNA also supports the downstream proteomic or multi-omic analyses which can help identify functional elements within an analysis [13]. Such methodologies of RNA-sequencing based gene expression profiles have even helped some labs identify molecular subtypes and biomarkers of colorectal cancer [14], [15], one of the most common types of cancer in the world [16]. In other words, RNA provides transcriptomic understanding of intestinal and immune system illnesses, due to their ability to characterize organism composition as well as functional relationships.

Other studies have shown how gene expression analyses may also provide insight into environmental factors [17], and the role of the gut microbiota in the development of immune responses [18]. Many groups, like Siegel *et al.* and Tarallo *et al.*, have found that altered gene expression profiles in gastrointestinal illnesses can also be correlated with transcriptomic changes in the gut microbiome [10], [16]–[18]. These examples show that metatranscriptomic analyses of the gut microbiome can elucidate differences between common gastrointestinal diseases. With better knowledge of environmental and immunological factors implicated in such clinically indistinct diseases, identification of genetic components contributing to such an illness’s etiology becomes easier.

Identifying and utilizing these biomarkers for the creation of diagnostic tools can better inform treatment of this umbrella of illnesses, in turn leading to better clinical outcomes for patients [19].

Accordingly, the research directive focuses on how gene expression profiling and pathway analysis of these illnesses (IBD and PDAC) can identify unique pathways, genes, and proteins implicated in each disorder. Due to the varied nature of IBD and PDAC, identifying the genetic intersection of tissue and microbiome gene expression will reveal genes which are confirmed to be involved in the illnesses with a low false discovery rate. These distinct biomarkers may then be connected to clinical presentations of the disorders, to better characterize and diagnose the disorder, allowing patients to be better treated. The goal of this study is to identify gene expression profiles and associated pathways in common intestinal diseases. A special focus on the interplay between immunology and environmental or gut microbiome factors involved in IBD will be taken. Thus, identifying the role of physiology versus lifestyle will provide insight into treatment options, as IBD and PDAC may be underdiagnosed and undertreated based on disease etiology or prodromic conditions.

II. METHODS

A. Data Retrieval and Preprocessing

The bioinformatics workflow used for the experiment is seen in Figure 1. First, gene expression datasets for irritable bowel disease (IBD) and pancreatic duct adenocarcinoma (PDAC), for both tissue cells and the microbiome of the respective organ, were collected from the NCBI Gene Expression Omnibus (GEO) [20] database. All data was generated using the Affymetrix Human Transcriptome Array 2.0 platform and were from human patients. Effectively, given that the data is coming from a variety of different sources, the data was also be processed or cleaned to ensure that all information is standardized. Data pre-processing, including comparisons, were accomplished by original programs written in Python. Currently, the main dataset that utilized for IBD was supplied by Chang and Mayer, which contains 25 IBD patients and 21 healthy controls (GSE63379) [21]. The Chang and Mayer dataset contains gene expression profiling of peripheral blood mononuclear cells (PBMCs). PBMCs are known to be implicated in selective immune system responses, as well as being the major cells involved in immunity overall (i.e. lymphocytes, monocytes, macrophages, etc.) [22]. Thus, given the putative highly immunologic nature of gastrointestinal disease, these cells act as the bridge between the physiological and environmental aspects of IBD, and thus are a critical host of information. Validation of these results is conducted using the expression data from 4 healthy volunteers and 5 IBD patients dataset published by Martinez and Santos which elucidates histopathological factors which implicate the immune system in gastrointestinal disease etiology (GSE14841) [23]. Essentially, the Martinez and Santos dataset acts as a measure of the IBD gut microbiome between healthy and ill patients using the jejunum to which the Chang and Mayer

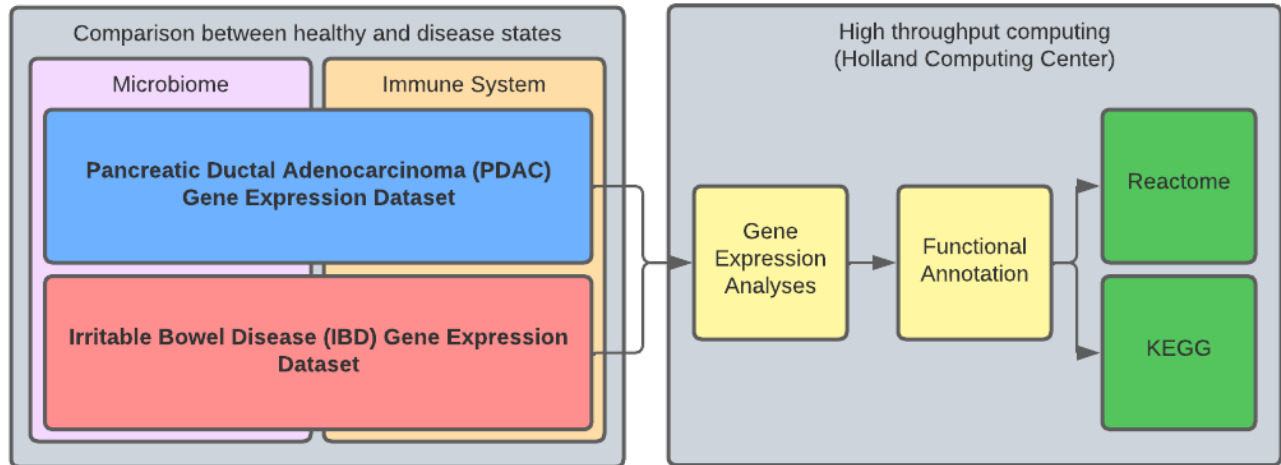


Fig. 1. Bioinformatics workflow used to compare and analyze gene expression and pathway differences between PDAC and IBD, health and ill, patient datasets. Functional annotation was achieved using Bioconductor goFunction package, and Kyoto Encyclopedia of Genes and Genomes (KEGG).

dataset acts as a measure of the host tissue cells. The jejunum is the middle part of the small intestine between the duodenum and ileum which assists in further digestion of food, as well as the absorption of nutrients.

Similarly, Schmittigen *et al.* have published data of gene expression profiling using 8 healthy and 14 PDAC tissue samples (GSE71989) [24], while Hirahoka *et al.* use data which represents the pancreas microbiome (GSE19650) [25]. The latter contains 7 healthy samples and 15 samples with PDAC. That is, these pancreatic tissue and pancreatic microbiome datasets present the intersection of the host tissue and gut microbiome for PDAC-affected patients. For PDAC, the immune aspect is represented by intraductal papillary-mucinous neoplasm tissue (IPMN) gene expression, while pancreatic biome expression is taken from the tissues of the organ itself. The benefit of using neoplastic tissues is that it accounts for the carcinogenesis of PDACs, and the transition from the benign papillary-mucinous adenoma variant to the malignant carcinoma.

Volcano plots of gene expression were generated in R to qualitatively visualize the relative numbers of significantly differentially expressed genes for both illnesses.

B. Gene Expression Analyses

From these datasets, probes were first mapped to genes using a Python script and the Affymetrix Human Transcriptome Array 2.0 probe set provided by the company. Differentially expressed genes between the healthy and ill patients were then identified using the R limma [26], affy [27], and hgu-133plus2.db [28] libraries, with further filtration performed using another novel Python script. In turn, this gene set was again compared between the microbiome and immune system to find genes that are confirmed to be implicated in the illness. Identification of significantly ($p < 0.05$) and differentially expressed genes and comparison between datasets was conducted

using a slurm script on the Holland Computing Center (HCC). Given that the Benjamini and Hochberg adjusted p-value yielded no significantly differentially expressed genes for the IBD tissue sample dataset, the unaltered p-value was preferred in this scenario. Computing was performed on the HCC due to the large data set and subsequently great processing power required for adequate efficiency.

C. Pathway Analysis and Functional Annotation

Using the gene expression data collected from NCBI GEO [20], for each of the diseases, pathway analysis in R was conducted using various Bioconductor libraries, such as GOFUNCTION [29], org.Hs.eg.db [30], pathview [31], and ReactomePA [32]. This code enabled the visualization of differentially expressed genes to their respective pathways in the intestine and pancreas using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [33]–[35], REACTOME [36], and gene ontologies (GO). To accomplish this, the obtained gene symbols were first converted to Entrez ids, removing any components which could not be mapped. These pathway charts were then compared, and a literature review conducted to better understand the roles of specific genes in the context of gastrointestinal disease. A focus was placed on identifying any correlations between pathways and clinical symptoms.

III. RESULTS

Prior to full gene expression analyses, differential gene expression for each illness was characterized using volcano plots (Figures 2 and 3). However, we see here that there are seemingly next to no significantly differentially expressed genes when using the IBD PBMC or immune system dataset when using the Benjamini and Hochberg adjusted p-value (Figure 2). Accordingly, to mitigate this issue, genes that were significantly expressed based on an unaltered p-value were utilized. While this may slightly increase the false discovery

rate, the overlap of the datasets is able to minimize this through the amount of data used for analysis. That is, later discovery shows that the overlap between datasets is fairly minimal despite cross checking thousands of genes. This same calculation was not an issue for the PDAC datasets, as is seen through the volcano plots in Figures 3. Still, for maximum possible standardization, a non-adjusted p-value was utilized. After processing the raw data, it was discovered that the PDAC dataset had thousands of differentially expressed genes, with hundreds for the IBD dataset.

Overall, initial results revealed that there is much smaller overlap between the expression of the gut microbiome and tissue samples involved in IBD versus the overlap seen between pancreatic cells and the gene expression seen in the pancreatic duct, despite having relatively similar numbers of expressed genes (Table 1). Comparing the gene lists, 64 commonly differentially expressed genes were identified between the IBD gut microbiome and stomach cells, while 19099 commonly differentially expressed genes were identified between the pancreatic duct biome and pancreatic cells. There are 27 commonly differentially expressed genes between both illnesses with duplicates and known homologs removed (Table 1).

TABLE I
DIFFERENTIALLY EXPRESSED GENES ASSOCIATED WITH IBD AND PDAC DISEASE STATES

Gene Symbol	IBD Microbiome logFC	PDAC Microbiome logFC
SLC19A1	0.890537	2.253729
TAPBP	0.597237	1.378669
SQSTM1	0.591454	-0.88035
GADD45B	0.545506	2.524668
CSAD	0.426277	2.132834
CARS2	0.382721	-0.68654
LTBR	0.379235	1.409514
MED20	0.332045	-1.52301
DAGLB	0.306691	0.712591
SMCR8	0.264237	-0.56907
PNRC1	0.261289	1.569592
HLA-C	0.259588	-1.24977
TMUB2	-0.18629	-0.68216
BRD7	-0.21192	0.514129
SMG8	-0.23621	0.570852
MTX1	-0.24454	-0.82138
MZF1	-0.28754	-1.14308
PTAFR	-0.32534	1.579793
TTYH2	-0.34464	3.633956
CTRL	-0.35785	8.672169
PGLS	-0.37849	0.949866
HK2	-0.43142	-1.89536
IL12RB1	-0.44687	2.033092
HECTD2	-0.48684	0.846792
NFIC	-0.5669	2.322305
ELOVL6	-0.68895	-2.47517

Top significantly (p<0.05) differentially expressed genes associated with both IBD and PDAC disease states, as intersected between gut microbiome and tissue gene expression datasets. There are 27 total genes common to both illnesses. Genes are ordered according to fold change between healthy and affected sample groups based on microbiome. LogFC refers to the log fold change of gene expression.

However, looking at a more low level view of the data, there were seemingly no similarities between the top differentially expressed genes between the IBD and PDAC datasets before merging or comparing. More interestingly, the fold change for the top up and downregulated samples for the PDAC samples were much greater than those for the IBD samples, at nearly 6-

TABLE II
TOP DIFFERENTIALLY EXPRESSED GENES ASSOCIATED WITH IRRITABLE BOWEL SYNDROME

Irritable Bowel Syndrome			
Downregulated	Fold Change	Upregulated	Fold Change
C1ORF21	-0.76838	GSTT1	1.528697
ELOVL6	-0.68895	SLC19A1	0.890537
SLC16A5	-0.65941	DUSP3	0.601571
KIAA1683	-0.58717	TAPBP	0.597237
NFIC	-0.5669	SQSTM1	0.591454

Top significantly (p<0.05) differentially expressed genes associated with both IBD gut microbiome and tissue samples. Genes are ordered according to fold change between healthy and affected sample groups.

TABLE III
TOP DIFFERENTIALLY EXPRESSED GENES ASSOCIATED WITH PANCREATIC DUCT ADENOCARCINOMA

Pancreatic Duct Adenocarcinoma			
Downregulated	Fold Change	Upregulated	Fold Change
S100P	-8.34346	CPA1	13.16132
CEACAM5	-6.99256	CEL	13.08126
MUC5AC	-6.64225	PNLIP	12.16057
CLDN18	-6.61624	CPB1	11.95104
SLC6A14	-5.77941	CELA2B	11.83255

Top significantly (p<0.05) differentially expressed genes associated with both PDAC pancreatic microbiome and tissue samples. Genes are ordered according to fold change between healthy and affected sample groups.

20 times their counterparts, as shown by Tables 1 and 2. All of the top differentially expressed genes are poorly characterized and have only generalized function such as chromosome binding or being termed an “intracellular protein”, except for *CPA1* and *S100P* which showed a high fold change in PDAC samples (Table 2) [37]–[40]. *CPA1* is produced in the pancreas and preferentially cleaves C-terminal branched-chain and aromatic amino acids from dietary proteins, thus showing function in digestive processes [37]. Similarly, *S100p* is known to be involved in cellular proliferation of pancreatic cancer as well as survival and invasion of the tumor [38].

In turn, analyzing these genes using REACTOME and KEGG also yielded curious results. Processing the IBD dataset revealed that there were no significantly enriched pathways associated with the illness, which may be in part to the limited size of the dataset post-processing. In contrast, with the PDAC sampleset, the CDC42 GTPase cycle, extracellular matrix organization, RAC1 GTPase cycle, RHO GTPase cycle, diseases of signal transduction by growth factor receptors and second messengers, EPH-Ephrin signaling, synthesis of PC, RHOB GTPase cycle, and EPHB-mediated forward signaling, as well as crosslinking of collagen fibrils pathways were all enriched (Figure 4). All enriched pathways associated with the PDAC dataset were highly significant, and over half had a pathway count of over 100. The RHO GTPase cycle was the most enriched pathway associated with PDAC at a count of over 225 and p-value of less than 0.004, while the crosslinking of collagen fibrils pathway was the least enriched with a count of less than 25 and p-value of approximately 0.016 (Figure 4).

This validates alternative functional annotation of PDAC and IBD associated pathways via and KEGG diagram mappings. The KEGG diagram for irritable bowel disease (hsa05321) only showed two factors that were upregulated: IL-

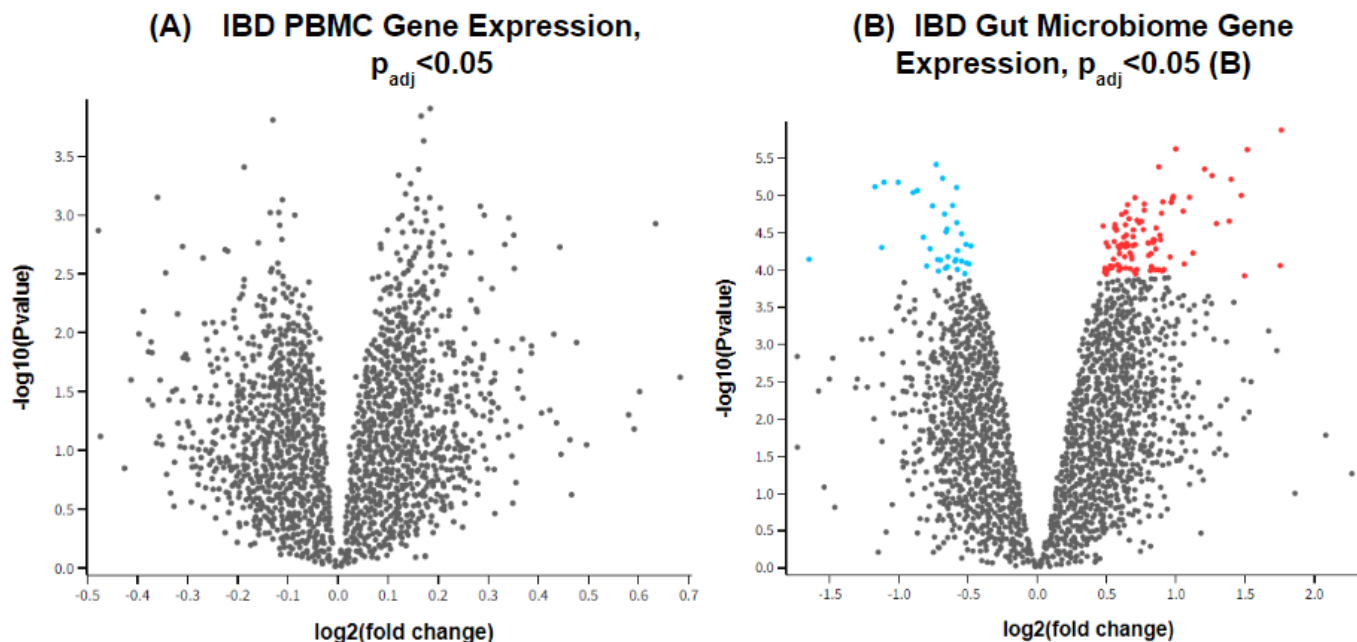


Fig. 2. Volcano plot of gene expression, of the most differentially expressed genes ($p_{adj} < 0.05$) between IBD-affected and healthy patient samples. Blue represents significantly downregulated, while red represents significantly upregulated genes. A) Comparison using peripheral blood mononuclear cells [21]. (PBMC). B) Comparison using gut microbiome expression data [23].

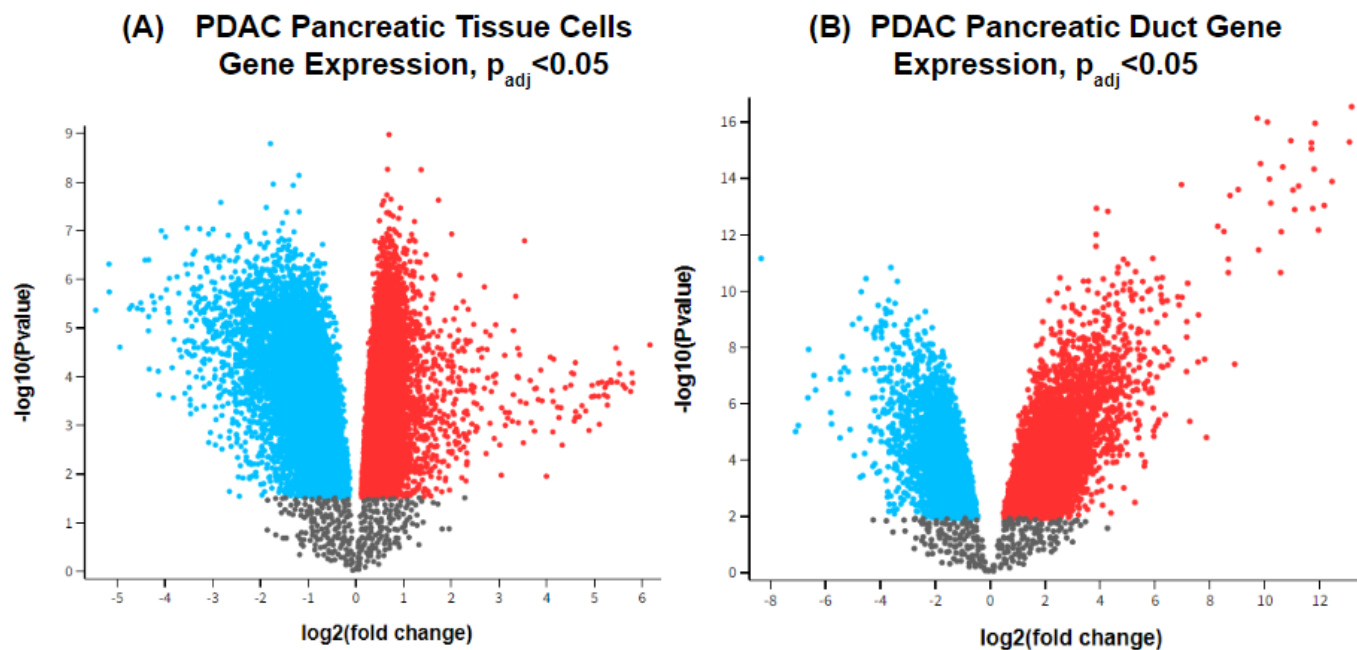


Fig. 3. Volcano plot of gene expression, of the most differentially expressed genes ($p_{adj} < 0.05$) between healthy patients and patients with pancreatic duct adenocarcinoma. Blue represents significantly downregulated, while red represents significantly upregulated genes. A) Comparison using healthy and tumorous tissues [24]. B) Comparison using pancreatic duct biome gene expression [25].

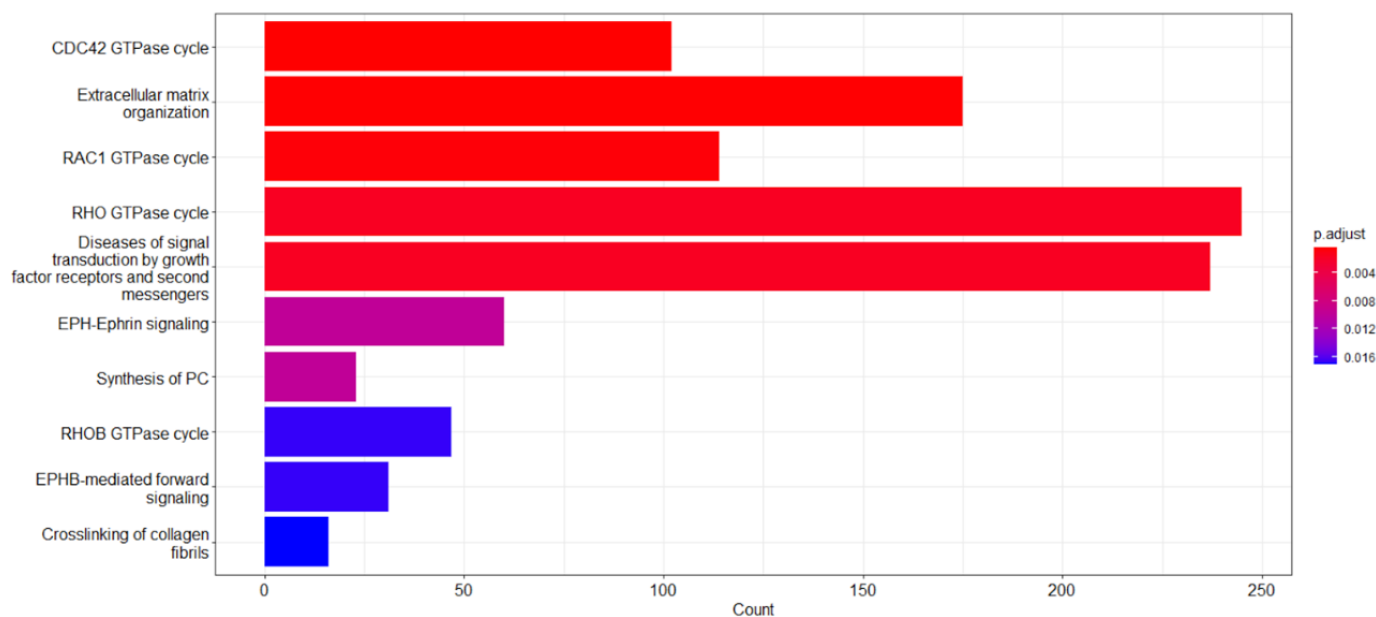


Fig. 4. REACTOME chart of enriched pathways associated with PDAC: CDC42 GTPase cycle, extracellular matrix organization, RAC1 GTPase cycle, RHO GTPase cycle, diseases of signal transduction by growth factor receptors and second messengers, EPH-Ephrin signaling, synthesis of PC, RHOB GTPase cycle, and EPHB-mediated forward signaling, and crosslinking of collagen fibrils pathways. The RHO GTPase cycle is the most enriched pathway at a count of over 225 and p-value of less than 0.004. Crosslinking of collagen fibrils is one of the least enriched pathways with a count of less than 25 and p-value of approximately 0.016.

17 and IL-12R β 1, a component of IL-12 (Figure 5). Accordingly, KEGG indicates that these genes are most involved with the Th17 cell differentiation pathway as well as with cytokine-cytokine receptor interaction. The opposite is seen when mapping the KEGG pathways for pancreatic cancer, which was considered synonymous with the KEGG pathway depicting pancreatic malignancy for the purposes of this experiment (hsa05212). However, again, we see that more components are upregulated than downregulated with PDAC. Implicated genes include *RacGEF*, *Rac*, *NF κ B*, *PKB/Akt*, *CASP9*, *P13K*, *Raf*, *K-Ras*, *RAI*, *RAIBP1*, *PLD1*, *Cdb-12/Rac*, *JNK*, *ERK*, *EGF*, *EGFR*, *Jak1*, *STAT3*, *STAT1*, *mTOR*, *S6K*, *p16*, *CDK4/6*, *E2F*, *p21*, *GADD45*, *Bax*, *POLK*, *BRCA2*, *TGF β* , *TGF β RI*, *SMAD2/3*, and *SMAD4* (Figure 6). Many of these genes, such as *BRCA2*, *EGF*, *Ras*, *TGF β* , and *Bax* are well known genes associated with cellular proliferation, and thus may be considered putative protooncogenes or associated with pancreatic malignancies. Generally, KEGG indicates genes are most affiliated with the P13K-Akt signaling pathway, ErbB signaling pathway, Jak-STAT signaling pathway, TGF- β signaling pathway, cell cycle, and end products of the p53 signaling pathway (Figure 6). For both illnesses, KEGG mapping of genes with a negative fold change indicated that all genes displayed in both the IBD and PDAC diagrams were minimally downregulated.

IV. DISCUSSION

The results elucidate gene and pathway components which are common to both irritable bowel disorder (IBD) and pancreatic duct adenocarcinoma (PDAC). Despite implicating different genes, the illnesses involve similar pathways. For

instance, *NFIC* and *ELOVL6* are common significantly expressed genes between the two illnesses' microbiomes (Table 1), with *ELOVL6* expressed at a fold change of -2.47 and *NFIC* a fold change of 2.32 in IBD samples.

ELOVL6 (Elongation Of Long Chain Fatty Acids, member 6) was also the most downregulated common gene between both illnesses. This gene is known for having an integral function in catalyzing the first and rate-limiting reaction of the long-chain fatty acid elongation cycle, as well as a more minor role in other biological processes such as building membrane lipids and lipid mediators [41]. Other sources indicate that *ELOV6* is involved in energy metabolism, a mechanism of the gut, and insulin sensitivity, a responsibility of the pancreas [42]. Dysfunction of this single gene showed correlation with insulin resistance. This provides reserved validation of the initial results that pancreatic and gut dysfunction relating to *ELOVL6* has far reaching implications into the cause of common gastrointestinal disorders.

In turn, Nuclear Factor 1 C, or *NFIC*, is a dimeric DNA-binding protein which acts as a transcription or replication factor for adenovirus [43]. Adenoviruses are a common family of viruses which tend to cause a variety of symptoms, collectively known as the common cold, although they may also cause gastrointestinal illnesses [44]. However, to be more specific, *NFIC* binds to a palindromic sequence present in viral and cellular promoters for the replication of these microbes, so it also has potential function in the replication of specific bacteria or host cells that display this same primer region. To this effect, this primer region is found in many animal and bacterial cells. *NFIC* also has well characterized function in gastric tumorigenesis due to its role as a transcriptional factor [45]. As such, this may allude to *NFIC* being a genetic

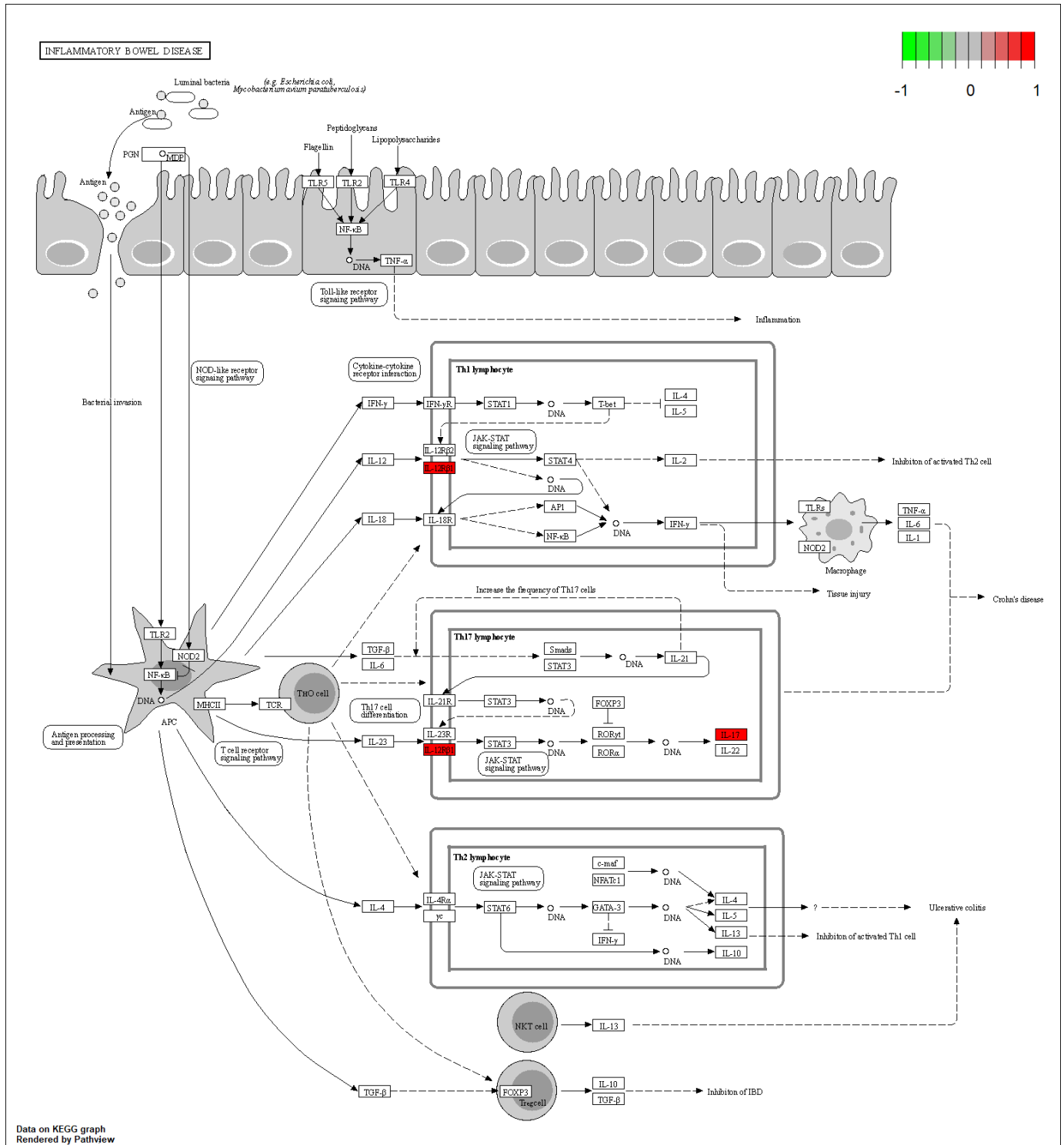


Fig. 5. KEGG annotation of differentially upregulated genes involved in IBD. KEGG pathway hsa05321 used, referring to illnesses which are implicated in chronic inflammation of the gastrointestinal tract. These genes are most affiliated with the Th17 cell differentiation pathway as well as with cytokine-cytokine receptor interaction.

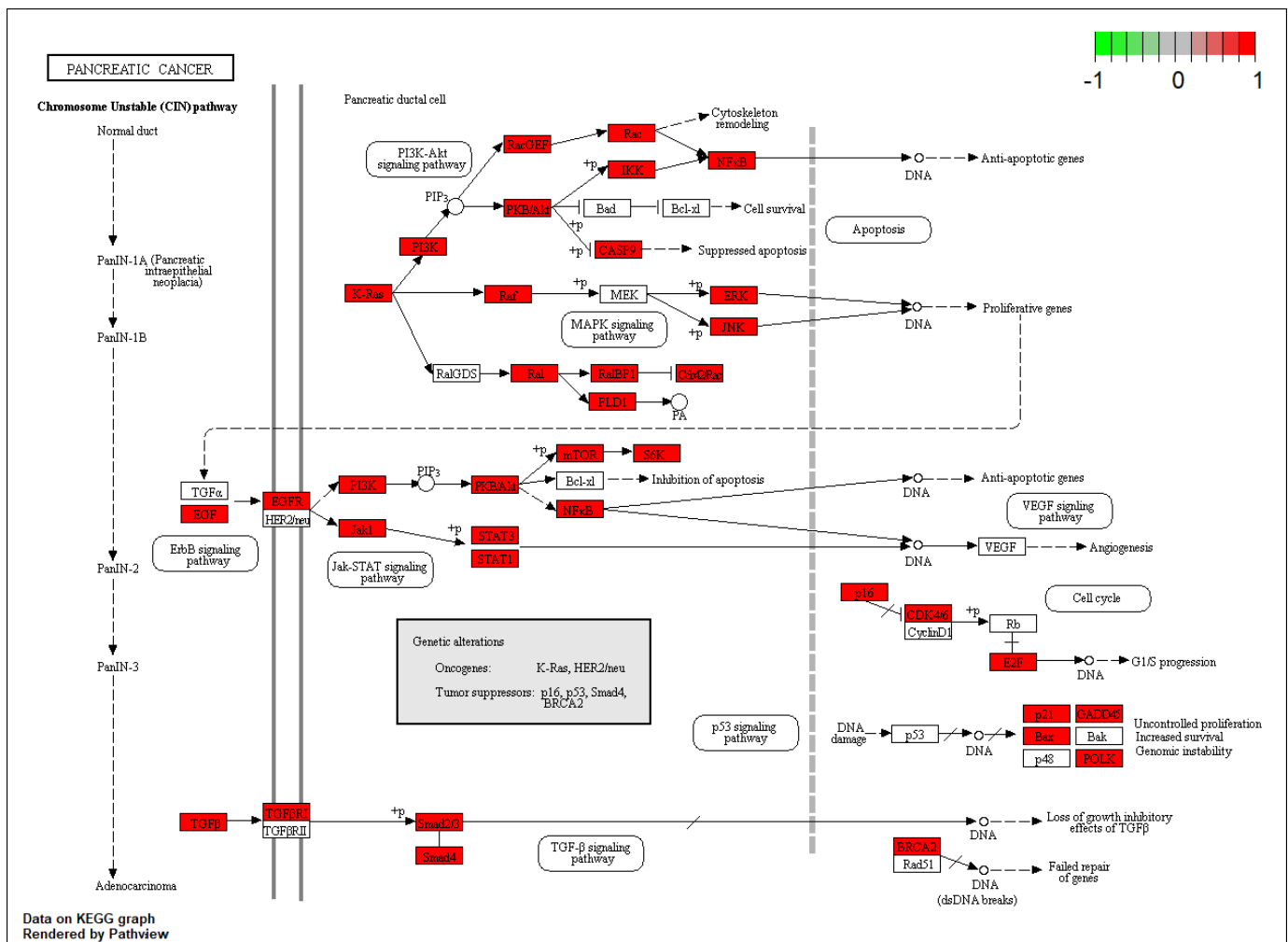


Fig. 6. KEGG pathway of differentially expressed genes involved in PDAC. KEGG pathway hsa05212 used, referring to infiltrating duct adenocarcinoma, the most common variety of PDAC. These genes are most affiliated with the PI3K-Akt signaling pathway, ErbB signaling pathway, Jak-STAT signaling pathway, TGF-β signaling pathway, cell cycle, and end products of the p53 signaling pathway.

component which acts as part of the intersection between the immune system and microbiome through its pleiotropic effects.

SLC19A1 (Solute Carrier Family 19 [Folate Transporter], member 1), another common gene, was one of the most upregulated between both diseases (Table 1). Moreover, pathway analysis revealed that enriched pathways associated with PDAC primarily work in the GTPase cycle, an important signal transduction intermediary, or have other signaling functions (Figure 4). *SLC19A1* is also involved in immunostimulatory signaling and intracellular transport of folate, a vitamin B complex that is necessary for healthy cell growth [46]. Thus, mutations in this gene are known to cause a variety of different cancers, particularly metastatic varieties, including colorectal cancer [46] and deficiencies in nutrient uptake [47]. As such, upregulation of this gene may indicate a corresponding increase in immune signaling, or that IBD and PDAC have common immune reactions. The solute carrier family also has critical function in the influx transport of bile acids, steroidal hormones, various drugs, and several other substrates, a role which connects this gene group to glucose

and energy homeostasis [48]. Such roles are also managed by the digestive system, and thus malignancies in SLC-family genes may strongly support disease pathogenesis.

CTRL was the most upregulated gene in the discovered gene set, with a fold change of 8.67 when looking at the PDAC pancreatic duct microbiome. However, while this can be considered more of a background gene as it has known enzymatic function in pancreas-associated digestion, it was interestingly also significantly expressed in irritable bowel syndrome [49]. Consequently, this may validate the aforementioned fact that the vast majority of IBD cases also present with pancreatic malignancies [50]. More specifically, mutations in this gene are known to present with life threatening immune disorders such as Netherton syndrome which in mild cases presents with food allergies and inflammatory skin disorders, and in more severe cases can cause the patient to be more susceptible to disease by lacking skin protection and sepsis [51]. That is, *CTRL* is also implicated in immunodeficiency and thus also potentially links the gastrointestinal system with the immune system.

Another gene of interest, *CPA1* and *S100P*, were both

one of most differentially expressed genes common to both illnesses, but also the only genes to not simply be termed an intracellular protein with fairly uncharacterized function. That is, mutations in *CPAI* are known to cause chronic pancreatitis, while elevated levels of the protein are known to be associated with pancreatic malignancies including cancer [37]. In turn, *S100P* is known to promote pancreatic cancer, as well as to stimulate cellular proliferation and survival in many other types of cancer [38]. Thus, these preliminary results validate that the workflow is correctly identifying the results that we desire. However, these differences in fold change may also allude to the adjusted p-value concerns, as a more severe fold change would likely be more significant than a fold change that is more minimal.

Pathway analysis also yielded interesting results, as many of the enriched pathways were involved in GTPase cycling. Specifically, GTPases are molecular switches or timers in many fundamental cellular processes, such as signal transduction in response to activation of cell surface receptors [52]. Aberrant signaling of GTPases are known to be connected with various disorders, such as in the formation of cytoskeletal barriers which cause inflammation characteristic of IBD [53]. Similarly, from the REACTOME diagram we saw that the RHO GTPase cycle was the most enriched with an extremely low p-value of less than 0.004 and count close to 250 (Figure 4). Again, this confirms the findings that we found with the REACTOME chart for PDAC, that the most enriched pathways are those associated with cell signaling and cytoskeletal rearrangement, which are hallmarks of cellular proliferation. Comparing this to the IBD results, we see while the gene functions involve tissue-based carcinogenesis, there is little data to implicate the immune system in PDAC pathogenesis.

Conversely, functional annotation of the IBD sample set yielded minimal results, with only IL-12 and IL-17 being indicated as the most upregulated components involved in the illness. Mutations in IL-12 are known to impair the development of interleukin-17-producing T lymphocytes and result in increased susceptibility to mycobacterial and *Salmonella* infections [54], [55]. In turn, IL-17 is a well known proinflammatory cytokine, indicating that upregulation of this gene may potentially increase gut inflammation, and thus validating the immune system's contribution to disease pathogenesis. The phenomenon of IL-17 overproduction has been shown to contribute to immunosuppression and nutritional impairment in patients with gastrointestinal cancers [56].

V. CONCLUSION

In conclusion, the results provide preliminary evidence that these common gastrointestinal disorders, irritable bowel disorder (IBD) and pancreatic duct adenocarcinoma (PDAC), have common etiologies. There are a total of 27 genes common to both illnesses, expressed by both the bacterial microbiome and tissue of the patient's pancreas and gut. Moreover, the illnesses involve comparable pathways through which malignancies may putatively lead to similar clinical symptoms. For instance *NFIC* and *ELOVL6* are common significantly expressed genes between the two illnesses' microbiomes (Table 1). *ELOVL6* is

expressed at a fold change of -2.47 and *NFIC* a fold change of 2.32 in PDAC samples. *ELOVL6* was the most down-regulated common gene between both illnesses. *SLC19A1*, another common gene, was one of the most upregulated between both. Each of these genes, *ELOVL6*, *SLC19A1*, and *NFIC* all have dual functions of being involved in metabolism as well as immunostimulatory signaling. Moreover, pathway analysis reveals that enriched pathways associated with PDAC primarily work in the GTPase cycle, an important signal transduction intermediary, or have other signaling functions. Overall, these findings provide substantial evidence which may support the creation of a diagnostic gene panel for better clinical distinguishment between IBD and PDAC which are often underdiagnosed.

ACKNOWLEDGMENT

The author would like to thank Dr. Dhundy Bastola and Ishwor Thapa for their support and guidance of this study. The author also thanks the University of Nebraska at Omaha, College of Information Science and Technology, School of Interdisciplinary Informatics, for their funding of the project.

REFERENCES

- [1] William D Chey, Jacob Kurlander, and Shanti Eswaran. Irritable bowel syndrome: a clinical review. *Jama*, 313(9):949–958, 2015.
- [2] Michael Orth, Philipp Metzger, Sabine Gerum, Julia Mayerle, Günter Schneider, Claus Belka, Maximilian Schnurr, and Kirsten Lauber. Pancreatic ductal adenocarcinoma: Biological hallmarks, current status, and future perspectives of combined modality treatment approaches. *Radiation Oncology*, 14(1):1–20, 2019.
- [3] Davide Giuseppe Ribaldone, Rinaldo Pellicano, and Giovanni Clemente Actis. Inflammation in gastrointestinal disorders: Prevalent socioeconomic factors. *Clinical and Experimental Gastroenterology*, 12:321, 2019.
- [4] George B Saffouri, Robin R Shields-Cutler, Jun Chen, Yi Yang, Heather R Lekatz, Vanessa L Hale, Janice M Cho, Eric J Battaglioli, Yogesh Bhattarai, Kevin J Thompson, et al. Small intestinal microbial dysbiosis underlies symptoms associated with functional gastrointestinal disorders. *Nature communications*, 10(1):1–11, 2019.
- [5] Mirpouya Mirmozaffari. Presenting an expert system for early diagnosis of gastrointestinal diseases. *International Journal of Gastroenterology Sciences*, 1(1):21–27, 2020.
- [6] Scott Plevy, Mark S Silverberg, Steve Lockton, Tom Stockfish, Lisa Croner, Jordan Stachelski, Michelle Brown, Cheryl Triggs, Emil Chuang, Fred Princen, et al. Combined serological, genetic, and inflammatory markers differentiate non-ibd, crohn's disease, and ulcerative colitis patients. *Inflammatory bowel diseases*, 19(6):1139–1148, 2013.
- [7] Cody C Allison, Jonathan Ferrand, Louise McLeod, Mohammad Hassan, Maria Kaparakis-Liaskos, Alexandra Grubman, Prithi S Bhatthal, Anouk Dev, William Sievert, Brendan J Jenkins, et al. Nucleotide oligomerization domain 1 enhances ifn- γ signaling in gastric epithelial cells during helicobacter pylori infection and exacerbates disease severity. *The Journal of Immunology*, 190(7):3706–3715, 2013.
- [8] L Bauer, A Hapfelmeier, S Blank, M Reiche, J Slotta-Huspenina, M Jesinghaus, A Novotny, T Schmidt, M Grosser, M Kohlruss, et al. A novel pretherapeutic gene expression-based risk score for treatment guidance in gastric cancer. *Annals of Oncology*, 29(1):127–132, 2018.
- [9] Michael H Schwartz, Haipeng Wang, Jessica N Pan, Wesley C Clark, Steven Cui, Matthew J Eckwahl, David W Pan, Marc Parisien, Sarah M Owens, Brian L Cheng, et al. Microbiome characterization by high-throughput transfer rna sequencing and modification analysis. *Nature communications*, 9(1):1–13, 2018.
- [10] Ruben AT Mars, Yi Yang, Tonya Ward, Mo Houtti, Sambhawa Priya, Heather R Lekatz, Xiaojia Tang, Zhifu Sun, Krishna R Kalari, Tal Korem, et al. Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell*, 182(6):1460–1473, 2020.

- [11] Jonas Halfvarson, Colin J Brislawn, Regina Lamendella, Yoshiki Vázquez-Baeza, William A Walters, Lisa M Bramer, Mauro D'amato, Ferdinando Bonfiglio, Daniel McDonald, Antonio Gonzalez, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature microbiology*, 2(5):1–7, 2017.
- [12] Xochitl C Morgan, Timothy L Tickle, Harry Sokol, Dirk Gevers, Kathryn L Devaney, Doyle V Ward, Joshua A Reyes, Samir A Shah, Neal LeLeiko, Scott B Snapper, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology*, 13(9):1–18, 2012.
- [13] Tae Woong Whon, Na-Ri Shin, Joon Yong Kim, and Seong Woon Roh. Omics in gut microbiome analysis. *Journal of Microbiology*, 59(3):292–297, 2021.
- [14] Rachel V Purcell, Martina Visnovska, Patrick J Biggs, Sebastian Schmeier, and Frank A Frizelle. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Scientific reports*, 7(1):1–12, 2017.
- [15] Atsushi Yamada, Pingjian Yu, Wei Lin, Yoshinaga Okugawa, C Richard Boland, and Ajay Goel. A rna-sequencing approach for the identification of novel long non-coding rna biomarkers in colorectal cancer. *Scientific reports*, 8(1):1–10, 2018.
- [16] Rebecca L Siegel, Kimberly D Miller, Ann Goding Sauer, Stacey A Fedewa, Lynn F Butterly, Joseph C Anderson, Andrea Cercek, Robert A Smith, and Ahmedin Jemal. Colorectal cancer statistics, 2020. *CA: a cancer journal for clinicians*, 70(3):145–164, 2020.
- [17] Sonia Tarallo, Giulio Ferrero, Gaetano Gallo, Antonio Francavilla, Giuseppe Clerico, Alberto Realis Luc, Paolo Manghi, Andrew Maltez Thomas, Paolo Vineis, Nicola Segata, et al. Altered fecal small rna profiles in colorectal cancer reflect gut microbiome composition in stool samples. *Msystems*, 4(5):e00289–19, 2019.
- [18] Jason M Knight, Laurie A Davidson, Damir Herman, Camilia R Martin, Jennifer S Goldsby, Ivan V Ivanov, Sharon M Donovan, and Robert S Chapkin. Non-invasive analysis of intestinal development in preterm and term infants using rna-sequencing. *Scientific reports*, 4(1):1–9, 2014.
- [19] Michael Camilleri, Paula Carlson, Andres Acosta, Irene Busciglio, Asha A Nair, Simon J Gibbons, Gianrico Farrugia, and Eric W Klee. Rna sequencing shows transcriptomic changes in rectosigmoid mucosa in patients with irritable bowel syndrome-diarrhea: a pilot case-control study. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 306(12):G1089–G1098, 2014.
- [20] Emily Clough and Tanya Barrett. The gene expression omnibus database. In *Statistical genomics*, pages 93–110. Springer, 2016.
- [21] Swapna Mahurkar, Christos Polytrachou, Dimitrios Iliopoulos, Charalabos Pothoulakis, Emeran A Mayer, and Lin Chang. Genome-wide dna methylation profiling of peripheral blood mononuclear cells in irritable bowel syndrome. *Neurogastroenterology & Motility*, 28(3):410–422, 2016.
- [22] Jalal Pourahmad and Ahmad Salimi. Isolated human peripheral blood mononuclear cell (pbmc), a cost effective tool for predicting immunosuppressive effects of drugs and xenobiotics. *Iranian journal of pharmaceutical research: IJPR*, 14(4):979, 2015.
- [23] Bruno K Rodiño-Janeiro, Cristina Martínez, Marina Fortea, Beatriz Lobo, Marc Pigrau, Adoración Nieto, Ana María González-Castro, Eloísa Salvo-Romero, Danila Guagnozzi, Cristina Pardo-Camacho, et al. Decreased tesk1-mediated cofilin 1 phosphorylation in the jejunum of ibs-d patients may explain increased female predisposition to epithelial dysfunction. *Scientific reports*, 8(1):1–11, 2018.
- [24] Jinmai Jiang, Ana Clara P Azevedo-Pouly, Roxana S Redis, Eun Joo Lee, Yuriy Gusev, David Allard, Dhruvitkumar S Sutaria, Mohamed Badawi, Ola A Elgamal, Megan R Lerner, et al. Globally increased ultraconserved noncoding rna expression in pancreatic adenocarcinoma. *Oncotarget*, 7(33):53165, 2016.
- [25] Nobuyoshi Hiraoka, Rie Yamazaki-Itoh, Yoshinori Ino, Yasunori Mizuguchi, Tesshi Yamada, Setsuo Hirohashi, and Yae Kanai. Cxcl17 and icam2 are associated with a potential anti-tumor immune response in early intraepithelial stages of human pancreatic carcinogenesis. *Gastroenterology*, 140(1):310–321, 2011.
- [26] Gordon K Smyth. Limma: linear models for microarray data. *Springer*, pages 397–420, 2005.
- [27] Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184, 2009.
- [28] Jonatan Taminau, David Steenhoff, Alain Coletta, Stijn Meganck, Cosmin Lazar, Virginie de Schaezen, Robin Duque, Colin Molter, Hugues Bersini, Ann Nowé, et al. insilicodb: an r/bioconductor package for accessing human affymetrix expert-curated datasets from geo. *Bioinformatics*, 27(22):3204–3205, 2011.
- [29] Jing Wang, Xianxiao Zhou, Jing Zhu, Yunyan Gu, Wenyuan Zhao, Jinfeng Zou, and Zheng Guo. Go-function: deriving biologically relevant functions from statistically significant functions. *Briefings in bioinformatics*, 13(2):216–227, 2012.
- [30] M Carlson. org. hs. eg. db: Genome wide annotation for human. r package version 3.8. 2. 2019.
- [31] Weijun Luo and Cory Brouwer. Pathview: an r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14):1830–1831, 2013.
- [32] Guangchuang Yu and Qing-Yu He. Reactomepa: an r/bioconductor package for reactome pathway analysis and visualization. *Molecular BioSystems*, 12(2):477–479, 2016.
- [33] Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.
- [34] Minoru Kanehisa. Toward understanding the origin and evolution of cellular organisms. *Protein Science*, 28(11):1947–1951, 2019.
- [35] Minoru Kanehisa, Miho Furumichi, Yoko Sato, Mari Ishiguro-Watanabe, and Mao Tanabe. Kegg: integrating viruses and cellular organisms. *Nucleic acids research*, 49(D1):D545–D551, 2021.
- [36] G Joshi-Tope, Marc Gillespie, Imre Vastrik, Peter D'Eustachio, Esther Schmidt, Bernard de Bono, Bijay Jassal, GR Gopinath, GR Wu, Lisa Matthews, et al. Reactome: a knowledgebase of biological pathways. *Nucleic acids research*, 33(suppl_1):D428–D432, 2005.
- [37] Eszter Hegyi and Miklós Sahin-Tóth. Human cpal mutation causes digestive enzyme misfolding and chronic pancreatitis in mice. *Gut*, 68(2):301–312, 2019.
- [38] Thiruvengadam Arumugam, Diane M Simeone, Kenneth Van Golen, and Craig D Logsdon. S100p promotes pancreatic cancer growth, survival, and invasion. *Clinical Cancer Research*, 11(15):5356–5364, 2005.
- [39] Adam Stevens, Leonibus Chiara De, Pierre Chatelain, and Peter Clayton. Gene expression networks associated with changes in serum markers of metabolism and growth in gh-treated children with gh deficiency. *ESPE Abstracts*, 82, 2014.
- [40] Sally Pemble, Klaus R Schroeder, Sharon R Spencer, David J Meyer, Ernst Hallier, Hermann M Bolt, Brian Ketterer, and John B Taylor. Human glutathione s-transferase theta (gstt1): cdna cloning and the characterization of a genetic polymorphism. *Biochemical Journal*, 300(1):271–276, 1994.
- [41] Takashi Matsuzaka and Hitoshi Shimano. Elovl6: a new player in fatty acid metabolism and insulin sensitivity. *Journal of molecular medicine*, 87(4):379–384, 2009.
- [42] Sonsoles Morcillo, Gracia María Martín-Núñez, Gemma Rojo-Martínez, María Cruz Almaraz, Eva García-Escobar, María Luisa Mansego, Griselda De Marco, Felipe J Chaves, and Federico Sorriquer. Elovl6 genetic variation is related to insulin sensitivity: a new candidate gene in energy metabolism. *PLoS one*, 6(6):e21198, 2011.
- [43] Amy Baldwin. *Modulation of human papillomavirus type 16 early gene expression by transforming growth factor-beta*. University of South Carolina, 2003.
- [44] Jason G Smith, Christopher M Wiethoff, Phoebe L Stewart, and Glen R Nemerow. Adenovirus. *Cell Entry by Non-Enveloped Viruses*, pages 195–224, 2010.
- [45] Amanda K Riffel, Eric Schuenemann, and Carrie A Vyhlidal. Regulation of the cyp3a4 and cyp3a7 promoters by members of the nuclear factor i transcription factor family. *Molecular pharmacology*, 76(5):1104–1114, 2009.
- [46] Liu Huang, Tao Zhang, Conghua Xie, Xin Liao, Qianqian Yu, Jueping Feng, Hong Ma, Jing Dai, Min Li, Jigui Chen, et al. Slc10b1 and slc19a1 gene variants and irinotecan-induced rapid response and survival: a prospective multicenter pharmacogenetics study of metastatic colorectal cancer. *PLoS One*, 8(10):e77223, 2013.
- [47] Michelle R O'Byrne, Kit Sing Au, Alanna C Morrison, Jone-Ing Lin, Jack M Fletcher, Kathryn K Ostermaier, Gayle H Tyerman, Sabine Doebel, and Hope Northrup. Association of folate receptor (folr1, folr2, folr3) and reduced folate carrier (slc19a1) genes with meningomyelocele. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 88(8):689–694, 2010.
- [48] Tatiana Claro Da Silva, James E Polli, and Peter W Swaan. The solute carrier family 10 (slc10): beyond bile acid transport. *Molecular aspects of medicine*, 34(2-3):252–269, 2013.
- [49] Jiajia Wu, Zedong Li, Kai Zeng, Kangjian Wu, Dong Xu, Jun Zhou, and Lijian Xu. Key genes associated with pancreatic cancer and their association with outcomes: A bioinformatics analysis. *Molecular medicine reports*, 20(2):1343–1352, 2019.
- [50] Henrik Laurell, Michèle Bouisson, Philippe Berthelémy, Philippe Rochaix, Sébastien Déjean, Philippe Besse, Christiane Susini, Lucien

- Pradayrol, Nicole Vaysse, and Louis Buscail. Identification of biomarkers of human pancreatic adenocarcinomas by expression profiling and validation with gene expression analysis in endoscopic ultrasound-guided fine needle aspiration samples. *World Journal of Gastroenterology: WJG*, 12(21):3344, 2006.
- [51] H Wiegmann, J Reunert, D Metze, T Marquardt, T Engel, V Kunde, S Ehl, D Foell, I van den Heuvel, V Oji, et al. Refining the dermatological spectrum in primary immunodeficiency: mucosa-associated lymphoid tissue lymphoma translocation protein 1 deficiency mimicking netherton/omenn syndromes. *British Journal of Dermatology*, 182(1):202–207, 2020.
- [52] Sandrine Etienne-Manneville and Alan Hall. Rho gtpases in cell biology. *Nature*, 420(6916):629–635, 2002.
- [53] Rocío López-Posadas, Michael Stürzl, Imke Atreya, Markus F Neurath, and Nathalie Britzen-Laurent. Interplay of gtpases and cytoskeleton in cellular barrier defects during gut inflammation. *Frontiers in immunology*, 8:1240, 2017.
- [54] A O’Garra, N Hosken, S Macatonia, CA Wenner, and K Murphy. The role of macrophage-and dendritic cell-derived il12 in th1 phenotype development. *Research in immunology*, 146(7-8):466–472, 1995.
- [55] Noé Ramirez-Alejo and Leopoldo Santos-Argumedo. Innate defects of the il-12/IFN- γ axis in susceptibility to infections by mycobacteria and salmonella. *Journal of interferon & cytokine research*, 34(5):307–317, 2014.
- [56] Takashi Yazawa, Masahiko Shibata, Kenji Gonda, Takeshi Machida, Satoshi Suzuki, Akira Kenjo, Izumi Nakamura, Takao Tsuchiya, Yoshihisa Koyama, Kenichi Sakurai, et al. Increased il-17 production correlates with immunosuppression involving myeloid-derived suppressor cells and nutritional impairment in patients with various gastrointestinal cancers. *Molecular and clinical oncology*, 1(4):675–679, 2013.