Pancreatic Cancer Analysis Report

Youxin (Anthony) Tan

2024-12-25

Contents

1 Introduction

Pancreatic cancer is one of the deadliest malignancies worldwide, with a five-year survival rate of less than 10% due to its aggressive nature(Chen, 2023), limited treatment options and late diagnosis. Among pancreatic cancers, pancreatic adenocarcinoma (PAAD) accounts for the majority of cases and creates challenges to early detection and therapeutic intervention. Understanding the molecular mechanisms driving PAAD is important for developing novel diagnostic markers and therapeutic targets to improve patient care(Wang, 2021).

This study aims to identify genes that are differentially expressed between normal pancreatic tissue and pancreatic adenocarcinoma tissue, leveraging genetic data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) project. By integrating advanced bioinformatics approaches, we seek to uncover key biological processes, pathways, and molecular functions associated with PAAD development and progression.

Our research addresses the following key questions:

What genes are differentially expressed between normal pancreatic tissue and pancreatic adenocarcinoma tissue?

Which biological processes and pathways are enriched in the differentially expressed genes? How do these findings contribute to our understanding of the molecular mechanisms underlying pancreatic adenocarcinoma?

Through data preprocessing, differential gene expression analysis(DEGs), pathway enrichment analysis(GSEA), and network inference(PANDA), this project will provide insights into the biology of pancreatic adenocarcinoma. By stratifying data based on sex, diabetes status,and alcohol consumption, we further aim to investigate molecular features that may refine our understanding of PAAD heterogeneity and its clinical implications.

2 Methods & Result: TCGA VS GTEx

2.0.1 Data Sources

- **TCGA (PAAD):** Pancreatic adenocarcinoma samples.
- **GTEx:** Normal tissue samples.

2.0.2 Tools and Packages

- Differential expression analysis was conducted using **limma** and **voom**.
- Gene Ontology (GO) and KEGG pathway enrichment analyses were performed.

```
library(Biobase)
library(limma)
library(ggfortify)
library(biomaRt)
library(gplots)
library(SummarizedExperiment)
library('ggplot2')      # For plotting
library('reshape2')     # For data processing
library('visNetwork')
library('fgsea')
```

2.0.3 Exploratory Data Analysis (EDA)

- **Sample Details:**
  - TCGA (179 cancer samples) with gene information for 34,032 genes.
  - GTEx (349 normal samples) with gene information for 31,530 genes.
  - Includes metadata on sex, age, diabetes, and alcohol history.

```
## Load Data and Preprocess

# downloading data
gtex <- readRDS("gtex_pancreas.rds")
tcga <- readRDS("tcga_paad.rds")

# structure of both datasets
gtex
tcga
```

- **Processing Steps:**
    1. Batch effect detection.

```
# checking sequencing platform
unique(colData(tcga)$tcga.gdc_platform)
names(colData(tcga))[1:5]
```

2. Calculated TPM (transcripts per million) of genes.

```
raw_counts <- assay(tcga, "raw_counts")
gene_lengths <- rowData(tcga)$bp_length/1000
rpk <- raw_counts/gene_lengths
scaling_factors <- colSums(rpk)
tpm <- sweep(rpk, 2, scaling_factors, FUN = "/") * 10^6
tpm[1, 0:5]
```

3. Focused analysis on early stages (1–2).

```
# filtering TCGA data to only include first and second grade tumors
keep_stages <- tcga@colData@listData$tcga.gdc_cases.diagnoses.tumor_stage %in%
  c("stage i", "stage ia", "stage ib", "stage iia", "stage iib")
dim(tcga)
tcga <- tcga[, keep_stages]
dim(tcga)

# defining a column that outlines the source: GTEX or TCGA
colData(gtex)$source <- "GTEX"
colData(tcga)$source <- "TCGA"
```

4. Merge normal/Tumour genes, filtered genes with low counts.

```
## Analysis for Normal vs Tumour

# intersection of columns (features of sample data)
common_cols <- intersect(names(colData(gtex)), names(colData(tcga)))

# identification of common genes for both GTEX and TCGA
common_genes <- intersect(rownames(assay(gtex, "logtpm")),
                          rownames(assay(tcga, "logtpm")))

# updating datasets by applying common genes and common columns filters
gtex_upd <- gtex[common_genes, ]
tcga_upd <- tcga[common_genes, ]
colData(gtex_upd) <- colData(gtex_upd)[, common_cols, drop = FALSE]
colData(tcga_upd) <- colData(tcga_upd)[, common_cols, drop = FALSE]

# merging
merged_logtpm <- cbind(assay(gtex_upd, "logtpm"), assay(tcga_upd, "logtpm"))
dim(merged_logtpm)
merged_colData <- as.data.frame(rbind(colData(gtex_upd), colData(tcga_upd)))
dim(merged_colData)
```

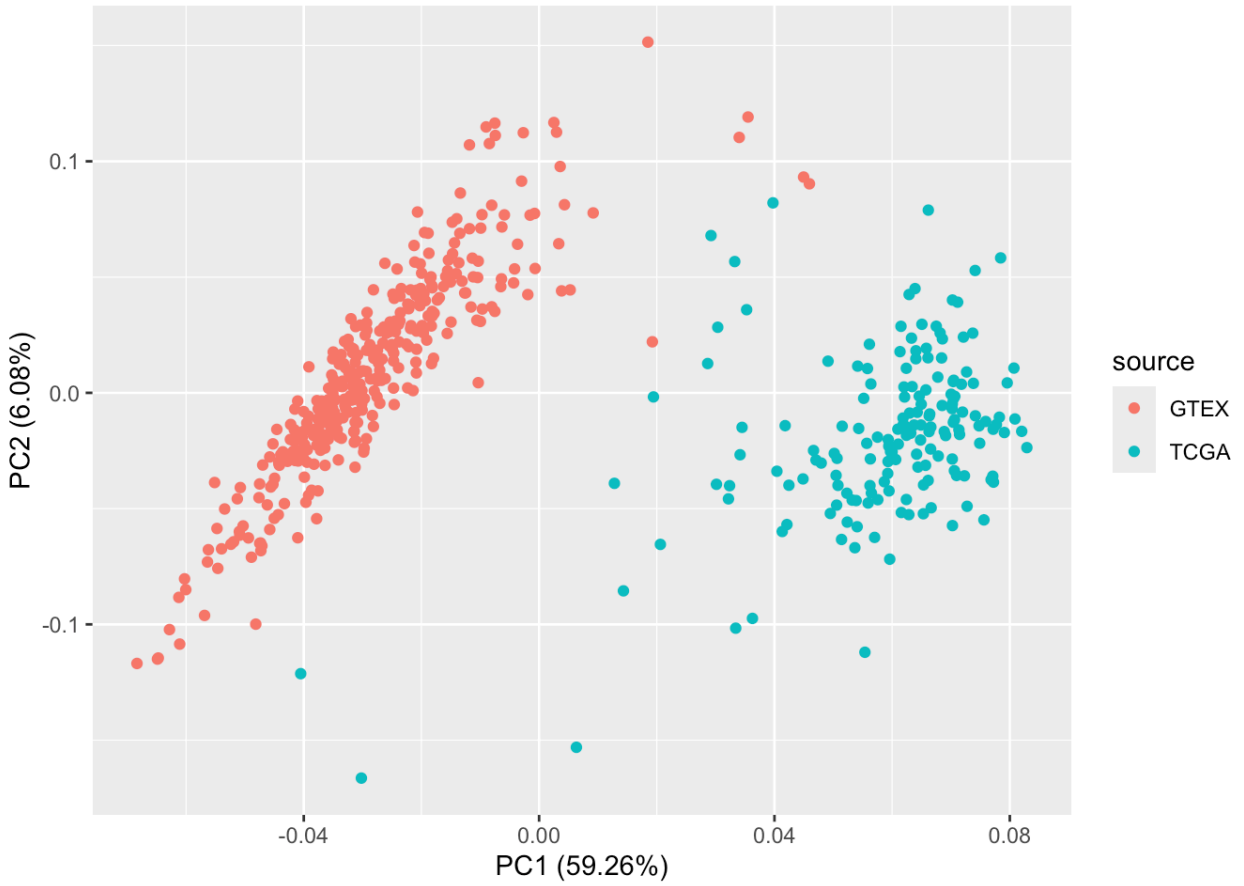
```
# genes will be filtered if they have very low counts (tpm < 1) in 25% of the samples
merged_exprs <- cbind(assay(gtex_upd, "raw_counts"), assay(tcga_upd, "raw_counts"))
merged_gene_lengths <- rowData(gtex_upd)$bp_length/1000
merged_rpk <- merged_exprs/merged_gene_lengths
merged_scaling_factors <- colSums(merged_rpk)
merged_tpm <- sweep(merged_rpk, 2, merged_scaling_factors, FUN = "/") * 10^6
keep <- rowSums(merged_tpm > 1) >= ncol(merged_tpm)/4
gtex_upd <- gtex_upd[keep, ]
tcga_upd <- tcga_upd[keep, ]
saveRDS(gtex_upd, file("gtex.rds"))
merged_exprs <- merged_exprs[keep, ]
sum(keep)
```

5. Conducted principal component analysis (PCA).

```
PCA <- prcomp(t(assay(tcga_filtered, "logtpm")))
p <- autoplot(PCA, data = as.data.frame(colData(tcga)),
              colour="tcga.gdc_cases.demographic.gender")
p
PCA <- prcomp(t(merged_logtpm[keep,]))
p <- autoplot(PCA, data = as.data.frame(rbind(colData(gtex_upd[keep,]), colData(tcga_upd[keep,]))), col
p
```

```
PCA <- prcomp(t(merged_logtpm))
p <- autoplot(PCA, data = merged_colData, colour="source")
p
```

```
knitr::include_graphics("./PCA.png")
```

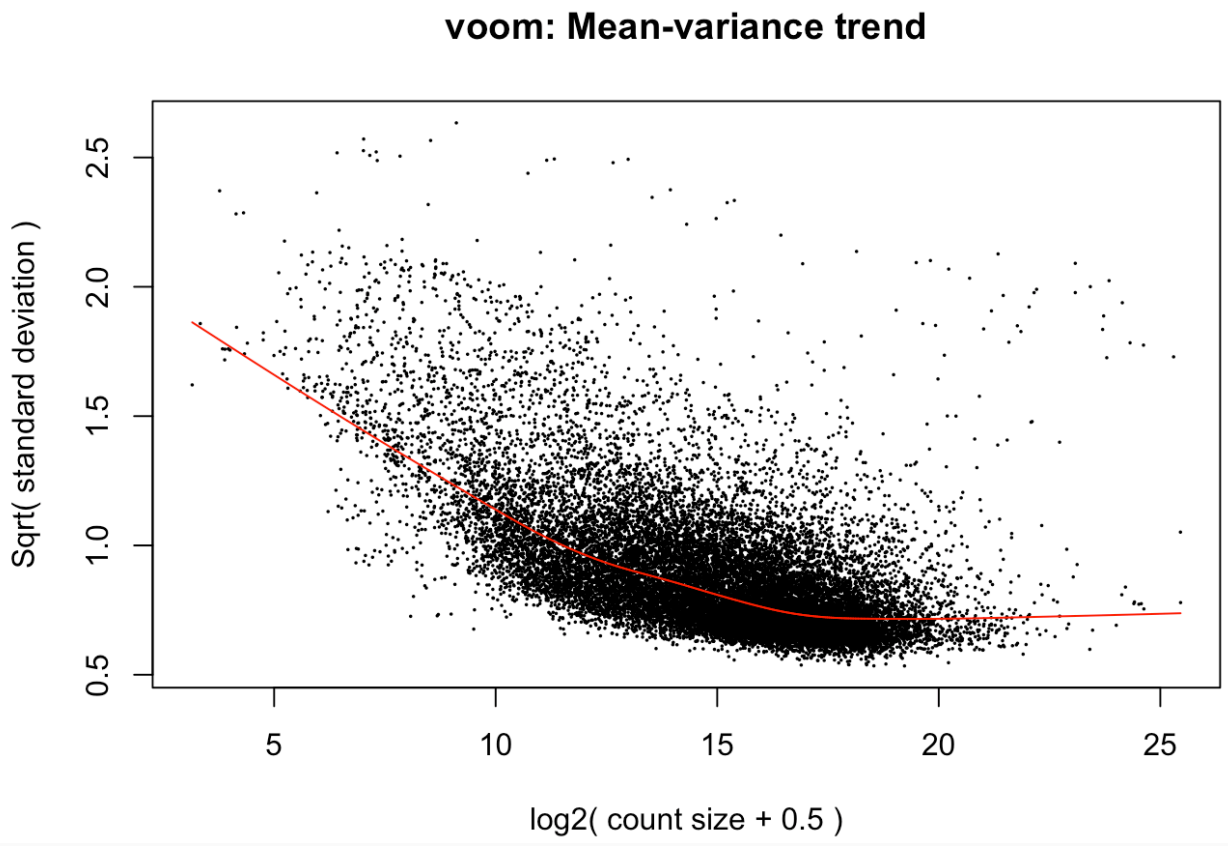


2.0.3.1 PCA Interpretation We can see the GTEx and TCGA are clearly separated into two clusters, especially in the PC1 Axis. It indicates there is major difference between the GTEx and TCGA dataset, which is a good sign for further differential analysis.

2.1 Differential Expression Analysis

2.1.1 Voom Mean-variance trend

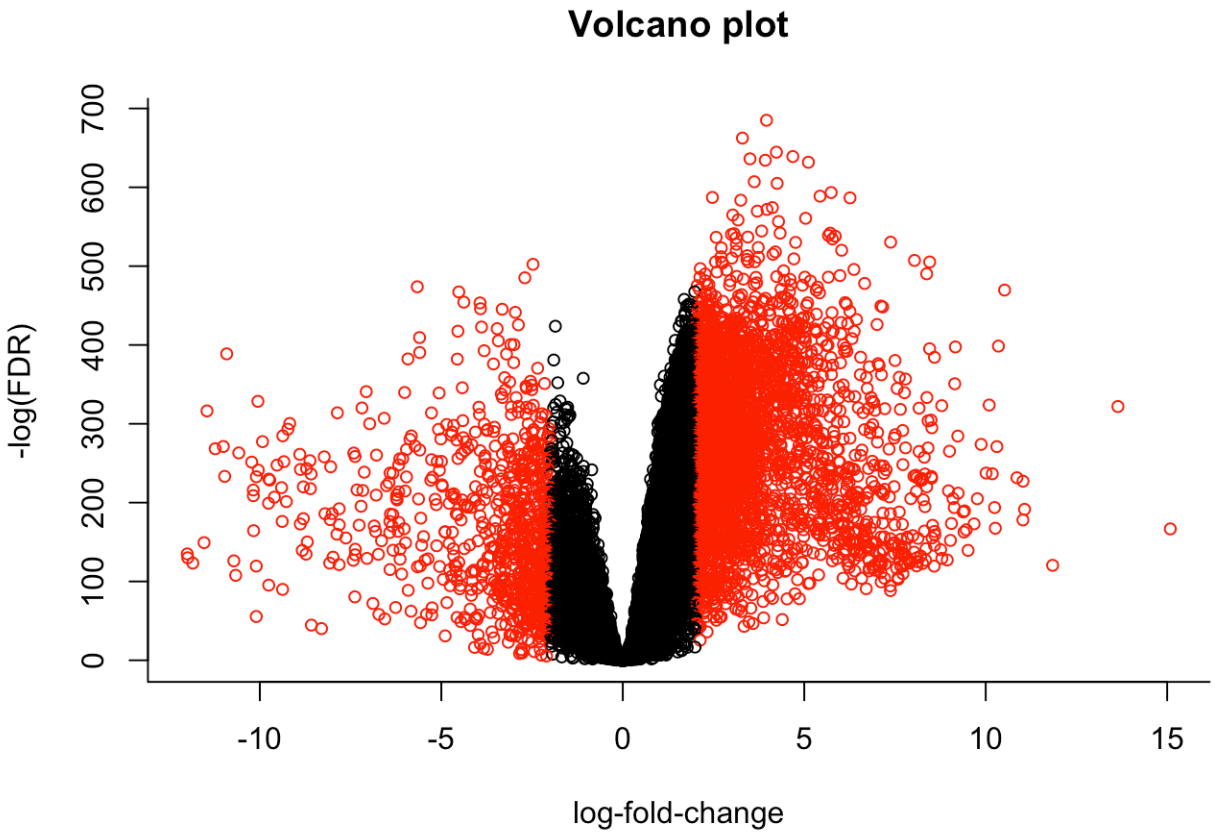
```
knitr::include_graphics("./Voom.png")
```



2.1.1.1 Voom Interpretation: The graph is as expected by statistical theory of variance, where variance decreases as count size increases.(UC Davis Bioinformatics, 2018)

2.1.2 Volcano Plot

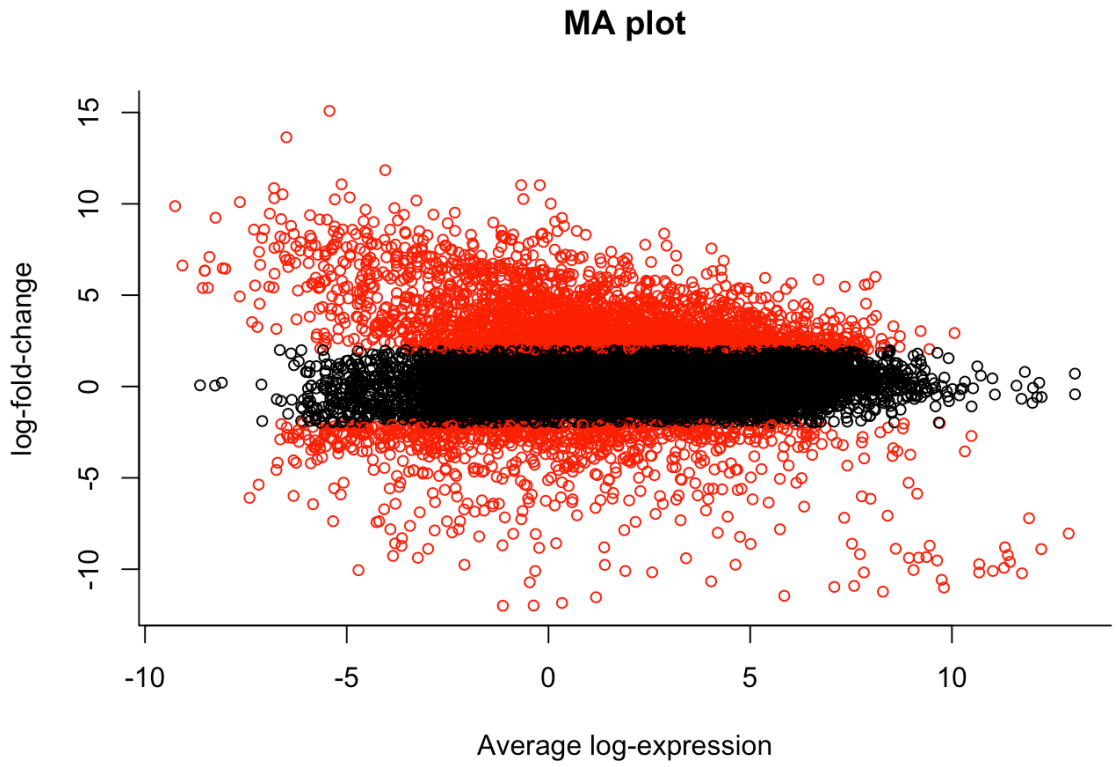
```
knitr::include_graphics("./Volcano.png")
```



2.1.2.1 Volcano Interpretation: The graph does not show any concerning pattern and follow the lecture example. Plot shows significant number of differentially expressed genes between the TGCA and GTEX.

2.1.3 MA Plot

```
knitr::include_graphics("./MA.png")
```



2.1.3.1 MA plot Interpretation: The graph does not show any concerning pattern and follow the lecture example. With steady logFC throughout all average log-expression.

2.2 GO and KEGG Enrichment Analysis

2.2.1 Gene Ontology (GO)

```
# Perform over-representation analyses for Gene Ontology terms using the limma package
go_tcga <- goana(up_tcga_ID, species="Hs")
# This is the list of top GO terms enriched for genes highly expressed in tcga
topGO(go_tcga, ontology = "BP")
```

```
knitr::include_graphics("./over-tcga.png")
```

GO:0007155	cell
GO:0002376	immune syste
GO:0009605	response to external
GO:0002682	regulation of immune syste
GO:0030155	regulation of cell
GO:0050896	response to
GO:0001775	cell a
GO:0051239	regulation of multicellular organisma
GO:0032501	multicellular organisma
GO:0006952	defense
GO:0048518	positive regulation of biologica
GO:0048522	positive regulation of cellula
GO:0048583	regulation of response to
GO:0016477	cell
GO:0045321	leukocyte a
GO:0006955	immune
GO:0098609	cell–cell
GO:0045785	positive regulation of cell
GO:0048584	positive regulation of response to
GO:0002684	positive regulation of immune syste

2.2.1.1 Over Expressed GO terms in TCGA

The GO terms can be grouped into key functional categories related to pancreatic cancer biology.

Immune System and Defense – Overexpressed terms include **immune system process (GO:0002376)**, **regulation of immune system process (GO:0002682)**, **immune response (GO:0006955)**, **defense response (GO:0006952)**, **leukocyte activation (GO:0045321)**, **positive regulation of immune system process (GO:0002684)**, and **cell activation (GO:0001775)**. Tumors manipulate immune responses to promote growth and evade detection(Wikipedia, n.d.), contributing to an inflammatory microenvironment, a hallmark of pancreatic cancer.

Cell Adhesion and Migration – Key terms include **cell adhesion (GO:0007155)**, **regulation of cell adhesion (GO:0030155)**, **cell-cell adhesion (GO:0098609)**, **positive regulation of cell adhesion (GO:0045785)**, and **cell migration (GO:0016477)**. Tumor progression and metastasis involve epithelial-to-mesenchymal transition (EMT), disrupting adhesion and enhancing cell motility and invasiveness.

Stimulus Response and Regulation – This category includes **response to external stimulus (GO:0009605)**, **response to stimulus (GO:0050896)**, **regulation of response to stimulus (GO:0048583)**, **positive regulation of response to stimulus (GO:0048584)**, **positive regulation of biological process (GO:0048518)**, and **positive regulation of cellular process (GO:0048522)**. Overexpression of these genes in pancreatic adenocarcinoma (PAAD) aids tumor survival by promoting immune evasion, adaptation to hypoxia, and resistance to therapy, while driving metastasis through EMT.

Multicellular Organism Processes – Terms include **multicellular organismal process (GO:0032501)** and **regulation of multicellular organismal process (GO:0051239)**. These pathways regulate tumor growth, tissue remodeling, and interactions between cancer cells and the microenvironment. Overexpression supports proliferation, immune modulation, and communication with stromal cells, facilitating tumor survival and metastasis.

This clustering highlights processes driving tumor progression, immune evasion, and metastatic behavior in pancreatic cancer.

```
go_gtex <- goana(up_gtex_ID, species="Hs")
# This is the list of top GO terms enriched for genes highly expressed in gtex
topGO(go_gtex, ontology = "BP")
```

```
knitr::include_graphics("./over-gtex.png")
```

```
GO:0007586
GO:0003013                              circulatory s
GO:1901606                       alpha-amino acid cata
GO:0046395                        carboxylic acid cata
GO:0016054                           organic acid cata
GO:0170040               proteinogenic amino acid cata
GO:0009063                               amino acid cata
GO:0008015                                     bloo
GO:0170035                           L-amino acid cata
GO:0044282                         small molecule cata
GO:0098739                            import across pl
GO:1901605                       alpha-amino acid meta
GO:0006575          cellular modified amino acid meta
GO:0048878                                   chemica
GO:0010749 regulation of nitric oxide mediated signal
GO:0006520                               amino acid meta
GO:0043436                                  oxoacid meta
GO:0006082                           organic acid meta
GO:0030001                                     metal
GO:0042219          cellular modified amino acid cata
```

2.2.1.2 Over Expressed GO terms in GTEx

The GO terms can be grouped into key functional categories related to normal pancreatic tissue biology.

Metabolism and Catabolism – Overexpressed terms include alpha-amino acid catabolic process (GO:1901606), carboxylic acid catabolic process (GO:0046395), organic acid catabolic process (GO:0016054), proteinogenic amino acid catabolic process (GO:0170040), amino acid catabolic process (GO:0009063), L-amino acid catabolic process (GO:0170035), small molecule catabolic process (GO:0044282), alpha-amino acid metabolic process (GO:1901605), amino acid metabolic process (GO:0006520), oxoacid metabolic process (GO:0043436), and organic acid metabolic process (GO:0006082). These pathways reflect the high metabolic activity required to maintain homeostasis and nutrient processing in normal pancreatic tissues, supporting energy production and normal cellular function. In contrast, cancer cells reduce these catabolic and metabolic processes as they shift towards anabolic pathways to fuel rapid proliferation and tumor growth.

Digestion and Circulation – Key terms include digestion (GO:0007586), circulatory system process (GO:0003013), and blood circulation (GO:0008015). These processes are critical for maintaining normal pancreatic

functions, such as enzyme secretion(like insulin) and nutrient absorption. In normal tissues, digestion and circulation ensure homeostasis and energy supply. However, cancer cells downregulate these functions as they prioritize growth and survival over normal physiological activities.(Wikipedia, n.d.)

Homeostasis and Transport – This category includes chemical homeostasis (GO:0048878), metal ion transport (GO:0030001), and import across plasma membrane (GO:0098739). Normal cells regulate ion transport and maintain chemical balance to support stable internal environments and proper enzyme activity in the pancreas. Cancer cells disrupt these pathways to promote proliferation and invasion, resulting in lower expression of these homeostatic functions.

Amino Acid and Nitric Oxide Regulation – Terms include cellular modified amino acid metabolic process (GO:0006575), cellular modified amino acid catabolic process (GO:0042219), and regulation of nitric oxide mediated signal transduction (GO:0010749). In normal cells, these processes contribute to metabolic balance and cellular communication. Pancreatic cancer cells downregulate these pathways as disrupted signaling and metabolic reprogramming favor tumor progression, immune evasion, and survival.

This clustering highlights the metabolic, circulatory, and homeostatic processes essential for normal pancreatic function, contrasting with the metabolic rewiring and disrupted physiological pathways characteristic of pancreatic cancer cells.

2.2.2 KEGG Pathways

```
# Perform over-representation analyses for KEGG pathways
kegg_tcga <- kegga(up_tcga_ID, species="Hs")
# This is the list of top KEGG pathwayss enriched for genes highly expressed in tcga
topKEGG(kegg_tcga)
```

```
knitr::include_graphics("./kegg-tcga.png")
```

```
                                                                                  P
hsa04145                                                                       Pha
hsa04514                                                          Cell adhesion mo
hsa04820                                                       Cytoskeleton in muscl
hsa04061 Viral protein interaction with cytokine and cytokine re
hsa04060                                 Cytokine-cytokine receptor inte
hsa04640                                                    Hematopoietic cell
hsa05416                                                             Viral myoca
hsa05323                                                          Rheumatoid ar
hsa04512                                                        ECM-receptor inte
hsa05150                                              Staphylococcus aureus in
hsa05140                                                               Leishma
hsa04659                                                     Th17 cell differen
hsa05332                                                     Graft-versus-host
hsa05330                                                           Allograft re
hsa05200                                                          Pathways in
hsa05166                                 Human T-cell leukemia virus 1 in
hsa04940                                                      Type I diabetes m
hsa04610                                         Complement and coagulation ca
hsa04510                                                               Focal a
hsa05152                                                                Tuber
```

2.2.2.1 Over expressed TCGA

Interpretation:

The top overexpressed KEGG pathways in pancreatic cancer shows key processes driving tumor growth, immune evasion, and metastasis. **Phagosome, cytokine-cytokine receptor interaction, and cell adhesion molecules** pathways shows an inflammatory microenvironment and enhanced immune signaling that promote tumor survival and immune suppression. **ECM-receptor interaction and focal adhesion**(Dafrazi, 2023) pathways indicate increased cell migration and invasion, for metastasis. These pathways indicate how pancreatic tumors manipulate immune responses and cellular interactions to sustain growth and spread.

```
# Perform over-representation analyses for KEGG pathways
kegg_gtex <- kegga(up_gtex_ID, species="Hs")
# This is the list of top KEGG pathwayss enriched for genes highly expressed in gtex
topKEGG(kegg_gtex)
```

```
knitr::include_graphics("./kegg-gtex.png")
```

	Pathway	N	DE	
hsa04972	Pancreatic secretion	106	26	7
hsa04974	Protein digestion and absorption	105	20	9
hsa00260	Glycine, serine and threonine metabolism	41	8	1
hsa00670	One carbon pool by folate	39	7	9
hsa04975	Fat digestion and absorption	43	7	1
hsa04080	Neuroactive ligand-receptor interaction	370	20	3
hsa00500	Starch and sucrose metabolism	40	6	1
hsa04950	Maturity onset diabetes of the young	26	5	1
hsa04973	Carbohydrate digestion and absorption	52	6	5
hsa04970	Salivary secretion	97	8	6
hsa01230	Biosynthesis of amino acids	75	7	6
hsa01100	Metabolic pathways	1564	48	7
hsa00330	Arginine and proline metabolism	50	5	2
hsa00270	Cysteine and methionine metabolism	52	5	3
hsa04971	Gastric acid secretion	76	6	3
hsa00750	Vitamin B6 metabolism	6	2	5
hsa04978	Mineral absorption	61	5	6
hsa00561	Glycerolipid metabolism	65	5	8
hsa04964	Proximal tubule bicarbonate reclamation	23	3	9
hsa04020	Calcium signaling pathway	254	11	1

2.2.2.2 Over Expressed GTEx

Interpretation:

The top overexpressed KEGG pathways in normal pancreatic tissue reflect essential functions for digestion, secretion, and metabolic regulation. **Pancreatic secretion, protein digestion and absorption, and fat digestion** pathways shows the pancreas's role in enzyme production and nutrient processing, which are reduced in tumor cells that prioritize growth over normal function. **Glycine, serine, threonine metabolism, and one-carbon pool by folate** pathways indicate active amino acid and folate metabolism, supporting cellular maintenance and DNA synthesis. **Carbohydrate and starch metabolism** reflects energy utilization, which is often disrupted in tumor cells that shift toward anabolic pathways. These pathways demonstrate the pancreas's role in maintaining metabolic homeostasis, which is lost as cancer progresses. It corroborates with our findings in the GO analysis.

2.3 Network Analysis

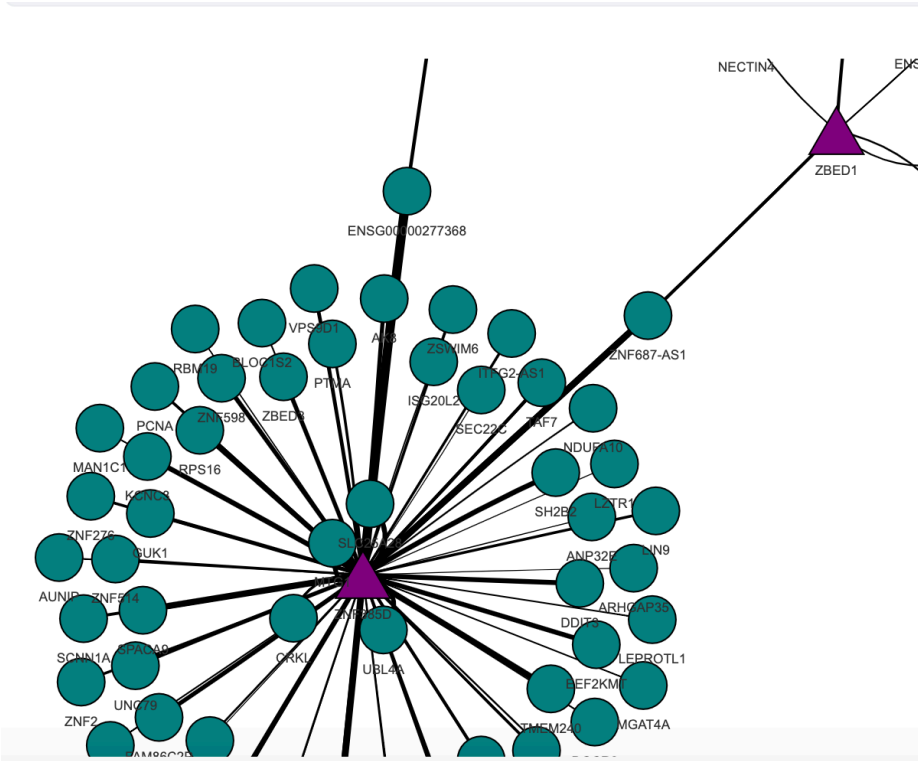
2.3.1 PANDA Networks

```
male_network_full <- read.table("output_males.txt", header=TRUE, stringsAsFactors=FALSE)
female_network_full <- read.table("output_females.txt", header=TRUE, stringsAsFactors=FALSE)
```

```
## This is the TCGA network
edges <- tcga_network
```

```
colnames(edges) <- c("from", "to", "value")
nodes       = data.frame(id = unique(as.vector(as.matrix(edges[,c(1,2)]))),
                         label=unique(as.vector(as.matrix(edges[,c(1,2)]))))
nodes$group = ifelse(nodes$id %in% edges$from, "TF", "gene")
net <- visNetwork(nodes, edges, width = "100%")
net <- visGroups(net, groupname = "TF", shape = "triangle",
                 color = list(background = "purple", border="black"))
net <- visGroups(net, groupname = "gene", shape = "dot",
                 color = list(background = "teal", border="black"))
visLegend(net, main="Legend", position="right", ncol=1)
```

```
knitr::include_graphics("./panda-tcga-network.png")
```

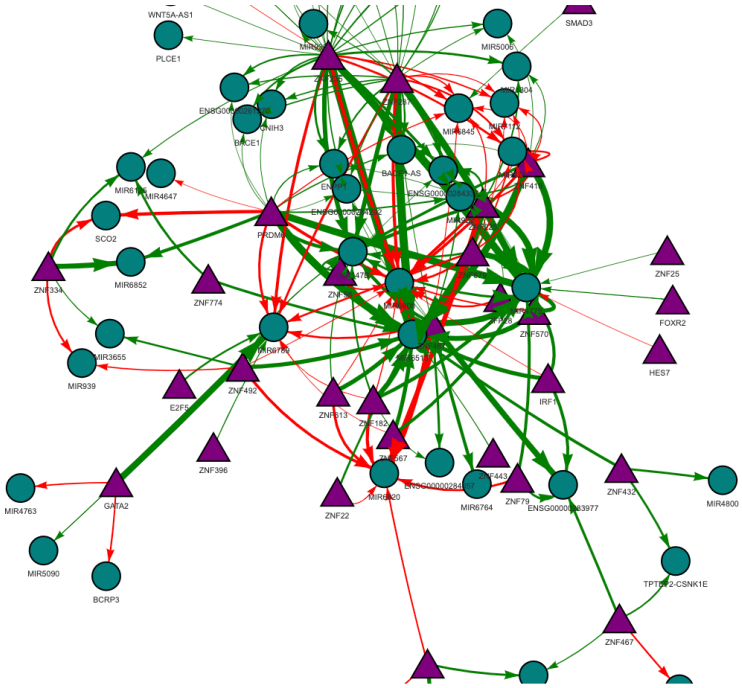


2.3.1.1 PANDA-TCGA NETWORK

```
## This plots the edges that change the most in cancer vs normal
edges <- diffnet
colnames(edges) <- c("from", "to", "value")
edges$arrows = "to"
edges$color  = ifelse(edges$value > 0, "green", "red")
edges$value  = abs(edges$value)
edges <- edges[order(edges$value, decreasing = TRUE), ]
edges <- edges[1:200,]
nodes       = data.frame(id = unique(as.vector(as.matrix(edges[,c(1,2)]))),
                         label=unique(as.vector(as.matrix(edges[,c(1,2)]))))
```

```
nodes$group = ifelse(nodes$id %in% edges$from, "TF", "gene")
net <- visNetwork(nodes, edges, width = "100%")
net <- visGroups(net, groupname = "TF", shape = "triangle",
                 color = list(background = "purple", border="black"))
net <- visGroups(net, groupname = "gene", shape = "dot",
                 color = list(background = "teal", border="black"))
visLegend(net, main="Legend", position="right", ncol=1)
```

```
knitr::include_graphics("./panda-cancer-vs-normal.png")
```



2.3.1.2 PANDA CANCER VS NORMAL

Plot interpretation:

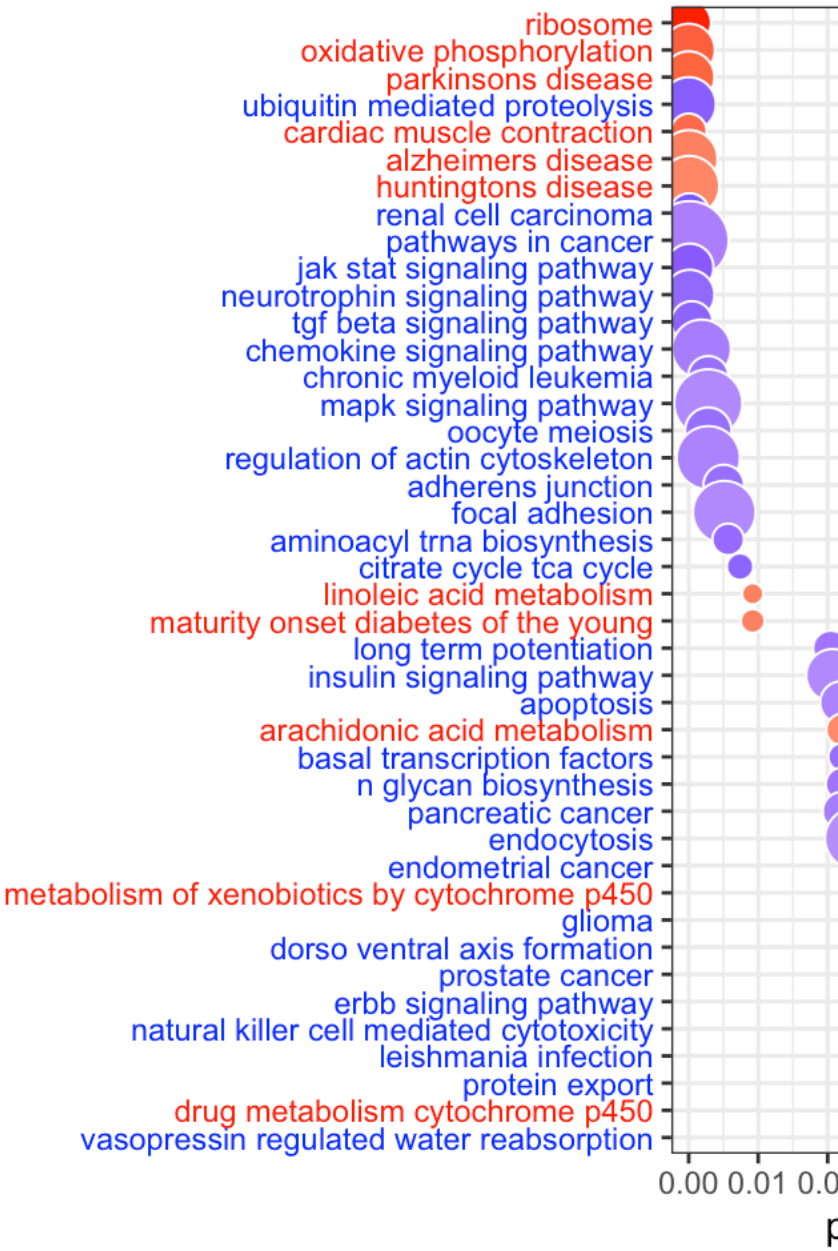
In the plot above, green arrows mean UP regulated in cancer, red arrows mean DOWN regulated in cancer and we can see a few genes are significantly UP regulated in cancer indicates that pancreatic cancer might replies significantly on certain genes' function. As we mentioned above, certain functions in GO analysis are specifically over-expressed in cancer such as Immune system and defense, cell adhesion etc, we suspect the genes with thick green arrows are related to some functions that cancer heavily relies on.

```
## This plots a bubble plot for the differentially expressed pathways


dat <- data.frame(fgseaRes)
# Settings
fdrcut <- 0.05 # FDR cut-off to use as output for significant signatures
dencol_neg <- "blue" # bubble plot color for negative ES
```

```
dencol_pos <- "red" # bubble plot color for positive ES
signnamelength <- 4 # set to remove prefix from signature names (2 for "GO", 4 for "KEGG", 8 for "REACT
asp <- 3 # aspect ratio of bubble plot
charcut <- 100 # cut signature name in heatmap to this nr of characters
a <- as.character(dat$pathway) # 'a' is a great variable name to substitute row names with something mo
for (j in 1:length(a)){
  a[j] <- substr(a[j], signnamelength+2, nchar(a[j]))
}
a <- tolower(a) # convert to lower case (you may want to comment this out, it really depends on what si
for (j in 1:length(a)){
  if(nchar(a[j])>charcut) { a[j] <- paste(substr(a[j], 1, charcut), "...", sep=" ")}
} # cut signature names that have more characters than charcut, and add "..."
a <- gsub("_", " ", a)
dat$NAME <- a
dat2 <- dat[dat[,"padj"]<fdrcut,]
dat2 <- dat2[order(dat2[,"padj"]),]
dat2$signature <- factor(dat2$NAME, rev(as.character(dat2$NAME)))
sign_neg <- which(dat2[,"NES"]<0) #Up regulated in cancer
sign_pos <- which(dat2[,"NES"]>0)
signcol <- rep(NA, length(dat2$signature))
signcol[sign_neg] <- dencol_neg # text color of negative signatures
signcol[sign_pos] <- dencol_pos # text color of positive signatures
signcol <- rev(signcol) # need to revert vector of colors, because ggplot starts plotting these from be
g<-ggplot(dat2, aes(x=padj,y=signature,size=size))
g+geom_point(aes(fill=NES), shape=21, colour="white")+
  theme_bw()+ # white background, needs to be placed before the "signcol" line
  xlim(0,fdrcut)+
  scale_size_area(max_size=10,guide="none")+
  scale_fill_gradient2(low=dencol_neg, high=dencol_pos)+
  theme(axis.text.y = element_text(colour=signcol))+
  theme(aspect.ratio=asp, axis.title.y=element_blank()) # test aspect.ratio
```

```
knitr::include_graphics("./panda-bubble.png")
```



2.3.1.3 Differentially Expressed Pathways

Bubble chart Network Analysis Discussion:

Running PANDA on TCGA and GTEx expression data, followed by GSEA, revealed the top targeted pathways in pancreatic cancer. The analysis identified aberrant signaling in pathways such as **JAK/STAT signaling(Wang, 2021), TGF- signaling, and the MAPK pathway**, which have been previously linked to various cancer types. Additionally, changes in the activity of pathways like **regulation of actin cytoskeleton, TCA cycle, and oxidative phosphorylation** can serve as indicators of transformations commonly observed in different cancers. This network analysis highlights key pathways that may contribute to pancreatic cancer development and progression.

2.4 Single Gene Analysis

Upon conducting differential expression analysis comparing TCGA, that is, cancer samples, to the GTEx normal samples, using limma and voom, the plot of the final model mean-variance trend appeared not to display any concerning patterns, and the volcano plot indicated a large number of genes that are differentially expressed. A few of these differentially expressed genes will be discussed below.

Using the thresholds of a 0.05 adjusted p-value and a log fold change of 2, we found 3199 genes up-regulated in the Cancerous TCGA samples. Genes indicated as being upregulated included ANXA8 (logFC 6.08), FAM83A (logFC 8.52), KRT6A (logFC 3.27), MET (logFC 2.87), NT5E (logFC 4.48), and SLC2A1 (logFC 5.09). In a study(Posta et al., 2023) attempting to identify prognostic factors for pancreatic cancer, high expression of all of these genes were found to be linked with shorter survival. In addition, all of these genes were also found to affect this poor prognosis independently of each other. An overview of the function of these genes is provided in the table below, and the MET gene will be further discussed:

```
knitr::include_graphics("./single-gene-1.png")
```

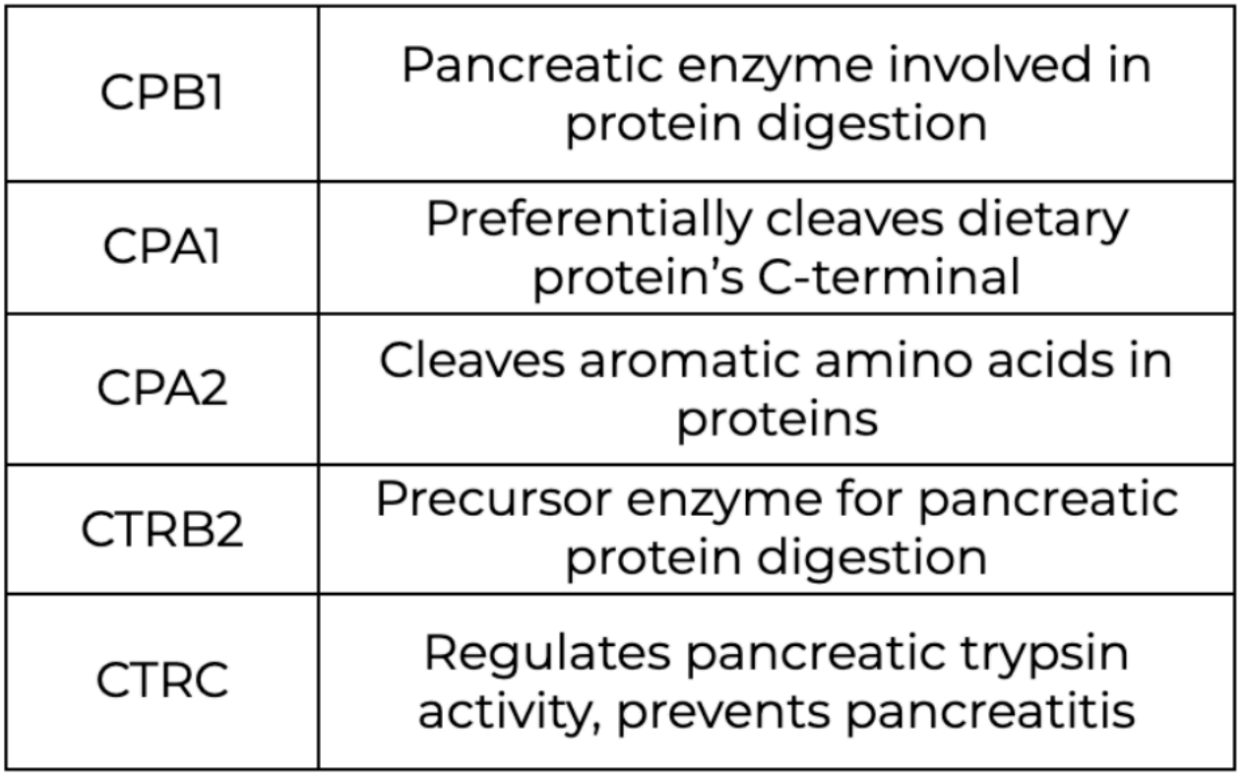
ANXA8	late endosome localization
FAM83A	biomarker in many cancers
NT5E	enzyme serves to convert AMP to adenosine
SLC2A1	facilitates glucose transport across plasma membranes
KRT6A	type II cytoskeletal keratin involved in epithelial wound healing.
MET	Proto-oncogene associated with cancers

The MET gene(NCBI, 2024) is located on the q arm of chromosome 7, and encodes a receptor for the Hepatocyte Growth Factor, upon binding to which the MET receptor activates pathways that regulate cell growth, survival, migration, and invasion. MET is a proto-oncogene that has been linked to multiple cancers, and specifically in pancreatic cancer, papers(Qin et al., 2022 & Lin et al., 2011 & Kim et al., 2024) suggest that the expression of MET is fundamental to the growth of the cancer cells, and increases the migration ability of cancer cells, leading to metastasis. MET has also been associated with cancer-related pain, via perineural invasion, and has been implicated in chemoresistance(Qin et al., 2022).

Using the thresholds of a 0.05 adjusted p-value and a log fold change of -2, we found 843 genes down-regulated in the Cancerous TCGA samples. Genes indicated as being downregulated included CPB1 (logFC -8.80),

CPA1 (logFC -10.22), CPA2 (logFC -9.53), CTRB2 (logFC -9.60), and CTRC (-8.71). In a study(Xiao et al, 2022) attempting to identify genes relevant to drug development for the treatment of pancreatic cancer, these genes were found to be highly downregulated hub genes in protein-protein interaction networks, implicating these genes as critical in the tumor microenvironment. An overview of the function of these genes is provided in the table below, and the CTRB2 gene will be further discussed:

```
knitr::include_graphics("./single-gene-2.png")
```



The CTRB2 gene(NCBI, 2024) is located on the q arm of chromosome 16, and encodes a principal precursor involved in the activation of pancreatic enzymes, thus dysregulation of this gene results in imbalances in pancreatic enzyme activity, and inversions in this gene are associated with chronic inflammation of the pancreas. Specifically in relation to pancreatic cancer, low expression(Maturana et al., 2021) of CTRB2, via gene variations(Jermusyk et al., 2021) or otherwise, confers risk of cancer. The relationship between CTRB2 and type 2 diabetes(Maturana et al., 2021), as well as chronic pancreatitis, risk factors for pancreatic cancer, may further explain the association of this gene with pancreatic cancer.

3 Methods & Result: Alcoholic Consumption

4 Methods & Result: Diabetic VS Non-Diabetic

- Differential network analysis highlighted:
  - **Diabetes vs Non-Diabetes:** Over-regulation of PINX1.
  - **Sex Differences:** DEGs mostly on X/Y chromosomes.

5 Methods & Result: Male vs Female

6 References

Chen, D., Zhang, S., & Fang, W. (2023). A single-cell atlas of the tumor, stromal and immune microenvironments in pancreatic cancer liver metastasis (GSE197177) [Data set]. NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197177

Dafrazi, A. A., Mehrabi, T., & Malekinejad, F. (2023). A bioinformatics study for recognition of hub genes and pathways in pancreatic ductal adenocarcinoma. *arXiv preprint arXiv:2303.14440*. https://doi.org/10.48550/arXiv.2303.14440

Jermusyk, A., Zhong, J., Connelly, K. E., Petersen, G. M., Westlake, C. J., & Amundadottir, L. T. (2021). A 584 bp deletion in CTRB2 inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer. *The American Journal of Human Genetics, 108*(10), 1885–1898. https://doi.org/10.1016/j.ajhg.2021.09.012

Kim, J., Lee, T. S., Lee, M. H., Cho, I. R., Ryu, J. K., Kim, Y. T., Lee, S. H., & Paik, W. H. (2024). Pancreatic cancer treatment targeting the HGF/c-MET pathway: The MEK inhibitor trametinib. *Cancers, 16*(5), 1056. https://doi.org/10.3390/cancers16051056

Li, C., Wu, J.-J., Hynes, M., Welling, T. H., Pasca di Magliano, M., & Simeone, D. M. (2011). c-Met is a marker of pancreatic cancer stem cells and therapeutic target. *Gastroenterology, 141*(6), 2218–2227. https://doi.org/10.1053/j.gastro.2011.07.050

López de Maturana, E., Rodríguez, J. A., Alonso, L., Lao, O., Molina-Montes, E., Martín-Antoniano, I. A., Gómez-Rubio, P., Lawlor, R., Carrato, A., Hidalgo, M., Iglesias, M., Molero, X., Löhr, M., Michalski, C., Perea, J., O'Rorke, M., Barberà, V. M., Tardón, A., Farré, A., … Malats, N. (2021). A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. *Genome Medicine, 13*(15). https://doi.org/10.1186/s13073-020-00816-4

National Center for Biotechnology Information (NCBI). (2024). *CTRB2 chymotrypsinogen B2 [Homo sapiens (human)]*. National Library of Medicine (NLM). Retrieved from https://www.ncbi.nlm.nih.gov/gene/440387

National Center for Biotechnology Information (NCBI). (2024). *MET proto-oncogene, receptor tyrosine kinase [Homo sapiens (human)]*. National Library of Medicine (NLM). Retrieved from https://www.ncbi.nlm.nih.gov/gene/4233

Posta, M., & Győrffy, B. (2023). Analysis of a large cohort of pancreatic cancer transcriptomic profiles to reveal the strongest prognostic factors. *Clinical and Translational Science, 16*(8), 1479–1491. https://doi.org/10.1111/cts.13563

Qin, T., Xiao, Y., Qian, W., Wang, X., Gong, M., Wang, Q., An, R., Han, L., Duan, W., Ma, Q., & Wang, Z. (2022). HGF/c-Met pathway facilitates the perineural invasion of pancreatic cancer by activating the mTOR/NGF axis. *Cell Death & Disease, 13*(4). https://doi.org/10.1038/s41419-022-04799-5

UC Davis Bioinformatics Core. (2018, June). Differential expression with limma-voom. Retrieved from https://ucdavis-bioinformatics-training.github.io/2018-June-RNA-Seq-Workshop/thursday/DE.html

Wang, S., Zheng, Y., Yang, F., Zhu, L., Zhu, X.-Q., Wang, Z.-F., Wu, X.-L., Zhou, C.-H., Yan, J.-Y., Hu, B.-Y., Kong, B., Fu, D.-L., Bruns, C., Zhao, Y., Qin, L.-X., & Dong, Q.-Z. (2021). The molecular biology of pancreatic adenocarcinoma: Translational challenges and clinical perspectives. *Signal Transduction and Targeted Therapy, 6*(1), 249. https://doi.org/10.1038/s41392-021-00659-4

Wikipedia contributors. (n.d.). Immune system. Wikipedia, The Free Encyclopedia. Retrieved December 25, 2024, from https://en.wikipedia.org/wiki/Immune_system

Wikipedia contributors. (n.d.). Tumor metabolome. Wikipedia, The Free Encyclopedia. Retrieved October 29, 2024, from https://en.wikipedia.org/wiki/Tumor_metabolome

Xiao, X., Wan, Z., Liu, X., Chen, H., Zhao, X., Ding, R., Cao, Y., Zhou, F., Qiu, E., Liang, W., Ou, J., Chen, Y., Chen, X., & Zhang, H. (2023). Screening of therapeutic targets for pancreatic cancer by bioinformatics methods. *International Journal of Gastrointestinal Cancer, 55*(6), 420–425. https://doi.org/10.1055/a-2007-2715

Xiao, Y., Zhang, B., Cloyd, J. M., Xu, G., Du, S., Mao, Y., & Pawlik, T. M. (2022). Gene signature and connectivity mapping to assist with drug prediction for pancreatic ductal adenocarcinoma. *Surgical Oncology, 44,* 101849. https://doi.org/10.1016/j.suronc.2022.101849