

به نام خدا



مقدمه‌ای بر بیوانفورماتیک

Bioinformatics-based screening of hub genes for
prostate cancer bone metastasis and analysis of
immune infiltration

استاد:

دکتر شریفی

نام و نام خانوادگی اعضاء:

حمیدرضا علیپور - ۴۰۰۱۰۸۸۸۲

سپهر میزانیان - ۴۰۰۱۰۹۶۸۴

کیارش جولایی - ۴۰۰۱۰۰۹۴۹

مقدمه

مقالاتی که پژوهه ما مسئولیت کار بر روی آن را بر عهده داشت، Bioinformatics-based screening of hub genes for prostate cancer bone metastasis and analysis of immune infiltration نام دارد. این مقاله با استفاده از روش‌های بیوانفورماتیکی، ژن‌های کلیدی مرتبه با متاستاز استخوانی سرطان پروستات (PCa) و ارتباط آنها با نفوذ سلول‌های ایمنی را شناسایی کرده است. این تجزیه و تحلیل بر روی مجموعه داده‌های GSE32269 انجام شده است.

در این مقاله، 197 ژن با متاستاز استخوان سرطان پروستات متناسب‌اند. در نهایت هم ده ژن کلیدی شناسایی شده‌اند که در نمونه‌های متاستاز استخوانی بیان بسیار بالاتری داشته و با کاهش نرخ بقای بیماران هم متناسب بودند. همچنین برای بررسی نفوذ سلول‌های ایمنی، الگوریتم CIBERSORTx و پایگاه داده TIMER به کار گرفته شدند.

این مقاله، از 5 بخش تشکیل شده که بخش 1، شامل توضیح مساله، بخش 4 شامل خلاصه نتایج به دست آمده و بخش 5 شامل نتیجه‌گیری نهایی است. بخش‌های اصلی و نیاز به پیاده‌سازی مقاله، بخش‌های 2 و 3 هستند. بخش 2، خود از 9 قسمت تشکیل شده و متدولوژی کار را توضیح می‌دهد. بخش 3 نیز خود دارای 8 قسمت است که شرح نتایج به دست آمده از قسمت‌های بخش 2 است.

ما در این پژوهه، تلاش بر پیاده‌سازی قسمت‌های بخش 2 و رسیدن به نتایج متناظر آنها در بخش 3 کردیم که در ادامه به تفصیل، جزئیات مربوطه آمده است.

بخش 1

در ابتداء داده‌های خام را پیدا می‌کنیم. برای اینکار از دیتابست GSE32269 استفاده می‌کنیم و آن را از سایت gene expression omnibus دانلود می‌کنیم:

The screenshot shows the GEO Accession Display page for Series GSE32269. The page header includes the NCBI logo and the GEO Gene Expression Omnibus logo. The main content area displays the following details:

- Status:** Public on Sep 22, 2011
- Title:** Expression data for primary localized prostate cancer versus castration-resistant bone metastatic prostate cancer
- Organism:** Homo sapiens
- Experiment type:** Expression profiling by array
- Summary:** We compared 22 primary PCa (hormone-dependent) versus 29 metastatic PCa (CRPC). The expression of genes related to cell cycle, proliferation, DNA synthesis, and androgen metabolism are significantly increased in CRPC group. The expression of AR-stimulated genes were partially reactivated. TMPRSS2-ERG fusion status was determined for the samples by PCR. The expression of ERG was highly increased in fusion positive versus negative.
- Overall design:** 120 snap-frozen CT-guided bone marrow biopsies from patients with castration-resistant prostate cancer were collected as a source of material. 29 independent biopsies containing mostly tumor were identified as CRPC group through carefully microscopical examination. 4 samples containing no tumor were identified as normal bone marrow group. Primary tumor was isolated by LCM from frozen biopsies of hormone-naïve patients. 22 samples were selected as primary PCa group.
- Contributor(s):** Cai C, Stambrough M, Febbo PG, Balk SP
- Citation(s):** Cai C, Wang H, He HH, Chen S et al. ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. *J Clin Invest* 2013 Mar;123(3):1109-22. PMID: 23426182

At the bottom of the page, there is a blue button labeled "Analyze with GEO2R". Below the button, the submission date is listed as Sep 21, 2011, and the last update date is Aug 10, 2018.

این داده‌ها به شکل داده‌های خام با فرمت CEL هستند که شامل ۵۴ سمپل است. ۲۲ سمپل اولیه مربوط به بیماران PCa است که بیماری‌شان در مراحل اولیه است و ۲۹ سمپل بعدی مربوط به بیمارانی است که سرطان متاستاز کرده و به استخوان رسیده است. در نهایت هم ۴ سمپل مربوط به افراد بدون بیماری است که همانطور که در مقاله گفته شده با آن‌ها کاری نداریم و فقط از ۵۱ سمپل ابتدایی استفاده می‌کنیم.

۲ بخش

در همان سایت geo گزینه analyze with GEO2R را انتخاب می‌کنیم. تا به صفحه زیر برویم:

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance.

[Full instructions](#) [YouTube](#)

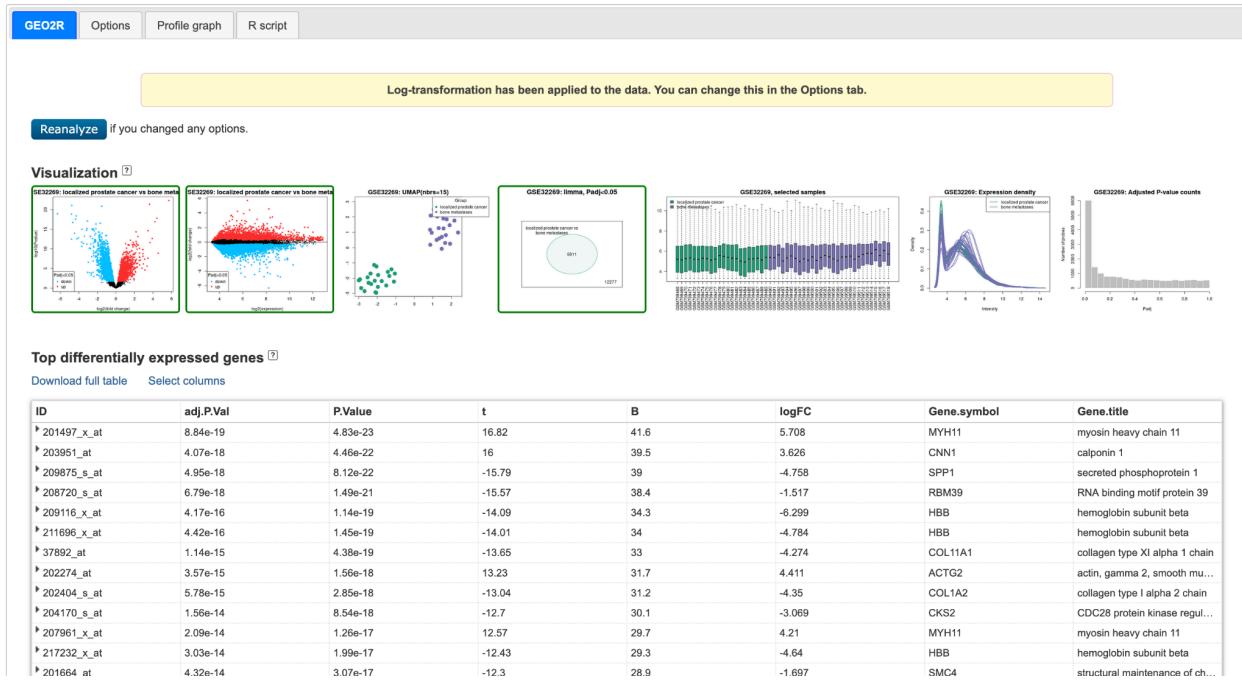
GEO accession Set Expression data for primary localized prostate cancer versus castration-resistant bone metastatic prostate cancer

▼ Samples		▼ Define groups				Selected 51 out of 55 samples	
localized prostate ...	GSM799483	Enter a group name:	List	sample 16	Primary localized prostate cancer	localized prostate cancer	no negative
localized prostate ...	GSM799484	<input checked="" type="checkbox"/> Cancel selection		sample 17	Primary localized prostate cancer	localized prostate cancer	no negative
localized prostate ...	GSM799485	<input type="checkbox"/>	localized prostate cancer (22 samples)	sample 18	Primary localized prostate cancer	localized prostate cancer	no positive
localized prostate ...	GSM799486	<input type="checkbox"/>	bone metastases (29 samples)	sample 19	Primary localized prostate cancer	localized prostate cancer	no positive
localized prostate ...	GSM799487	<input type="checkbox"/>		sample 20	Primary localized prostate cancer	localized prostate cancer	no positive
localized prostate ...	GSM799488			prostate cancer sample 21	Primary localized prostate cancer	localized prostate cancer	no positive
localized prostate ...	GSM799489			prostate cancer sample 22	Primary localized prostate cancer	localized prostate cancer	no positive
bone metastases	GSM799490			prostate cancer sample 23	Bone metastatic prostate cancer	bone metastases (CRPC)	yes positive
bone metastases	GSM799491			prostate cancer sample 24	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative
bone metastases	GSM799492			prostate cancer sample 25	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative
bone metastases	GSM799493			prostate cancer sample 26	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative
bone metastases	GSM799494			prostate cancer sample 27	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative
bone metastases	GSM799495			prostate cancer sample 28	Bone metastatic prostate cancer	bone metastases (CRPC)	yes positive
bone metastases	GSM799496			prostate cancer sample 29	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative
bone metastases	GSM799497			prostate cancer sample 30	Bone metastatic prostate cancer	bone metastases (CRPC)	yes negative

در این صفحه دو گروه bone metastases و localized prostate cancer را ایجاد می‌کنیم و ۵۱ سمپل ابتدایی را در این گروه‌ها قرار می‌دهیم. سپس با استفاده از تنظیمات زیر که در واقع تنظیمات دیفالت است داده‌ها را آنالیز می‌کنیم.

GEO2R	Options	Profile graph	R script
Apply adjustment to the P-values. More...	Apply log transformation to the data. More...	Category of Platform annotation to display on results.	Plot displays. More...
<input checked="" type="radio"/> Benjamini & Hochberg (False discovery rate)	<input type="radio"/> Auto-detect	<input type="radio"/> Submitter supplied	Significance level cut-off (enter number between 0 and 1) <input type="text" value="0.05"/>
<input type="radio"/> Benjamini & Yekutieli	<input type="radio"/> Yes	<input checked="" type="radio"/> NCBI generated	Log 2 fold change threshold <input type="text" value="0"/>
<input type="radio"/> Bonferroni	<input type="radio"/> No		
<input type="radio"/> Holm			
	Apply limma precision weights (vooma). More...		Volcano and Mean-difference plot contrasts (select up to 5) 0 selected (clear)
	<input type="radio"/> Yes	<input type="checkbox"/> localized prostate cancer vs bone metastases	
	<input checked="" type="radio"/> No		
	Force normalization. More...		
	<input type="radio"/> Yes		
	<input checked="" type="radio"/> No		
If you edit Options after performing an analysis, click Rereanalyze to apply the edits:			
Rereanalyze			

و خروجی سایت به شکل زیر است:



با استفاده از گزینه دانلود این داده‌ها را به شکل یک جدول می‌توانیم دریافت کنیم(این جدول در مستندات پروژه با نام GSE32269.top.table.tsv قرار گرفته است. حال با استفاده از کد زیر فیلتر داده‌ها را انجام می‌دهیم و ژنهایی که adjusted p_value کمتر از ۵ صدم و logFC بیشتر از ۱ دارند را نگه می‌داریم و در فایل DEGs_filtered.tsv ذخیره می‌کنیم. این کد با اسم filter_data.R در مستندات پروژه قرار دارد:

```

# Load necessary library
library(dplyr)

# Set working directory (adjust the path to your project folder)
setwd("~/Documents/university/"بیو/پروژه)

# Read the original data file from GEO2R
data <- read.table("GSE32269.top.table.tsv", header = TRUE, sep = "\t", stringsAsFactors = FALSE)

# Check column names to ensure correct filtering criteria (optional)
print(names(data))

# Filter for DEGs: adjusted P-value < 0.05 and |logFC| ≥ 1
# Adjust the column names below if needed (e.g., "adj.P.Val" and "logFC")
filtered_data <- data %>%
  filter(adj.P.Val < 0.05, abs(logFC) >= 1)

# Optionally, check how many DEGs you have
num_degs <- nrow(filtered_data)
cat("Number of differentially expressed genes:", num_degs, "\n")

# Write the filtered data to a new TSV file
write.table(filtered_data, "DEGs_filtered.tsv", sep = "\t", quote = FALSE, row.names = FALSE)

# Message when done
cat("Filtered data saved as 'DEGs_filtered.tsv'\n")

```

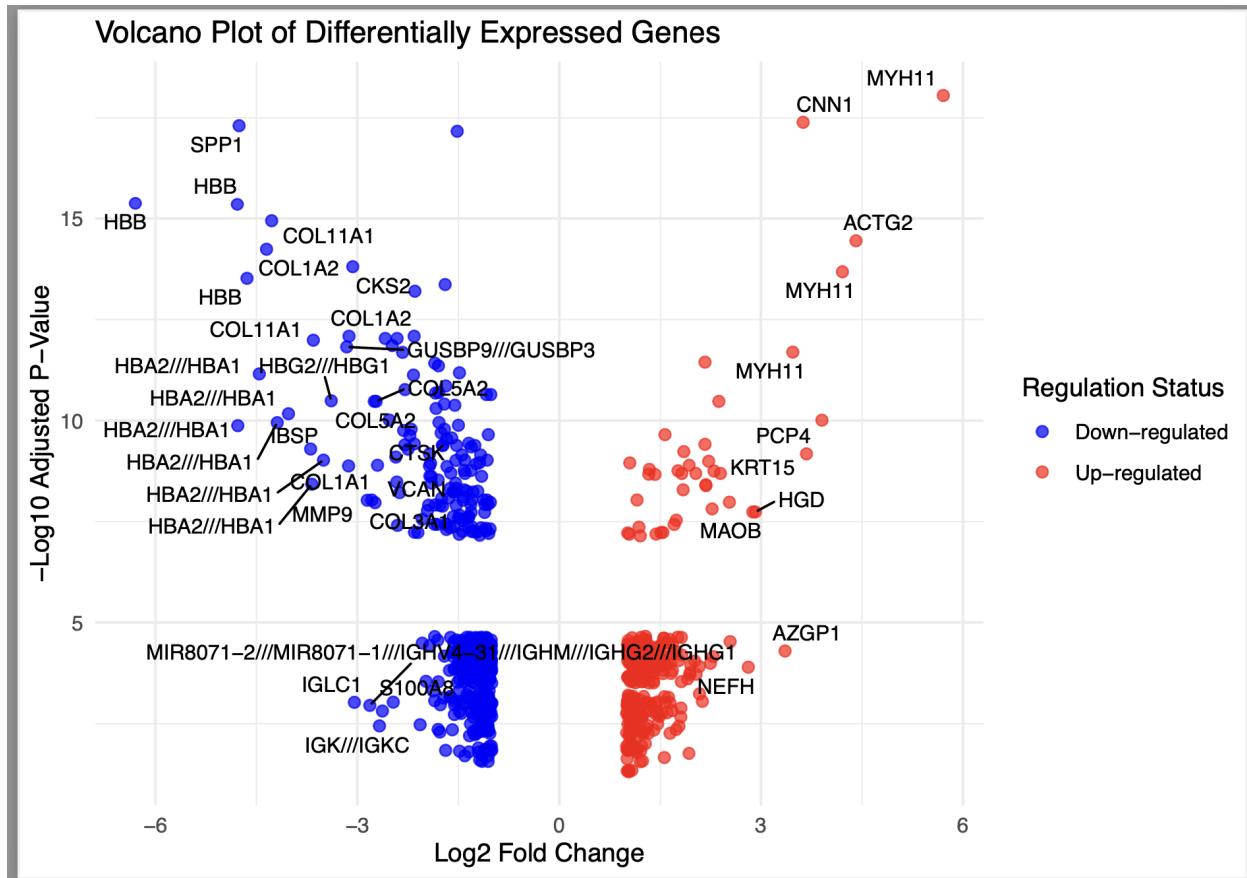
حال باید در این بخش مشابه مقاله heatmap و volcano plot برای ۴۰ ژن با بیشترین تفاوت بیان را رسم کنیم. برای اینکار کد volcano_heatmap.R را نوشتیم(خود GEO2R البته volcano plot را رسم کرد که در شکل هم مشخص است ولی heatmap را نکشید. در نتیجه این کد R نوشته شد تا هر دو کار را انجام دهد) در این کد ابتدا داده‌ها را فیلتر می‌کنیم(البته می‌توانستیم به جای اینکار مستقیم از DEGs_filtered.tsv بخوانیم) و همچنین ۴۰ ژن برتر بر حسب logFC را نیز بدست می‌آوریم. با استفاده از کتابخانه ggplot نمودار volcano را رسم کردیم:

```

# --- Volcano Plot with correct colors ---
volcano_plot <- ggplot(filtered_data, aes(x = logFC, y = -log10(adj.P.Val))) +
  geom_point(aes(color = ifelse(logFC > 1, "Up-regulated", ifelse(logFC < -1, "Down-regulated", "Not Regulated"))),
             size = 2, alpha = 0.7) +
  scale_color_manual(values = c("Up-regulated" = "red", "Down-regulated" = "blue", "Not Regulated" = "grey"))
  theme_minimal() +
  geom_text_repel(data = top_genes, aes(label = Gene.symbol), size = 3) +
  labs(title = "Volcano Plot of Differentially Expressed Genes",
       x = "Log2 Fold Change",
       y = "-Log10 Adjusted P-Value",
       color = "Regulation Status") +
  theme(legend.position = "right")
print(volcano_plot)

```

که به شکل زیر است و در مستندات هم قرار داده شده:



برای رسم heatmap نیاز به دانستن بیان هر یک از ژن‌ها در هر یک از سمپل‌ها داریم پس باید ابتدا آن را تولید کنیم. برای اینکار کد `create_expression_matrix.R` را می‌نویسیم که با استفاده از کتابخانه `affy` و خواندن داده‌های خامی که قبلاً دانلود کرده بودیم `expression_matrix` را می‌سازد و آن را ذخیره می‌کند. هر سطر این ماتریس ID یک ژن را مشخص می‌کند و هر ستون آن نیز یکی از سمپل‌ها را نشان می‌دهد و هر خانه میزان بیان آن ژن در آن سمپل را مشخص می‌کند. توجه کنید به اینکه تنها با ۵۱ سمپل ابتدایی کار داریم هم در اینجا در نظر گرفته شده:

```

if(!require(BiocManager)) install.packages("BiocManager")
if(!require(affy)) BiocManager::install("affy")
if(!require(limma)) BiocManager::install("limma")
library(affy)
library(limma)

# Set working directory to the folder containing .CEL files
setwd("~/Documents/university/بیو/پروژه/GSE32269_RAW") # Update this to your local path

# Read and normalize .CEL files
raw_data <- ReadAffy()      # reads all .CEL files from your directory
normalized_data <- rma(raw_data) # RMA normalization

# Extract expression values
expression_matrix <- exprs(normalized_data)

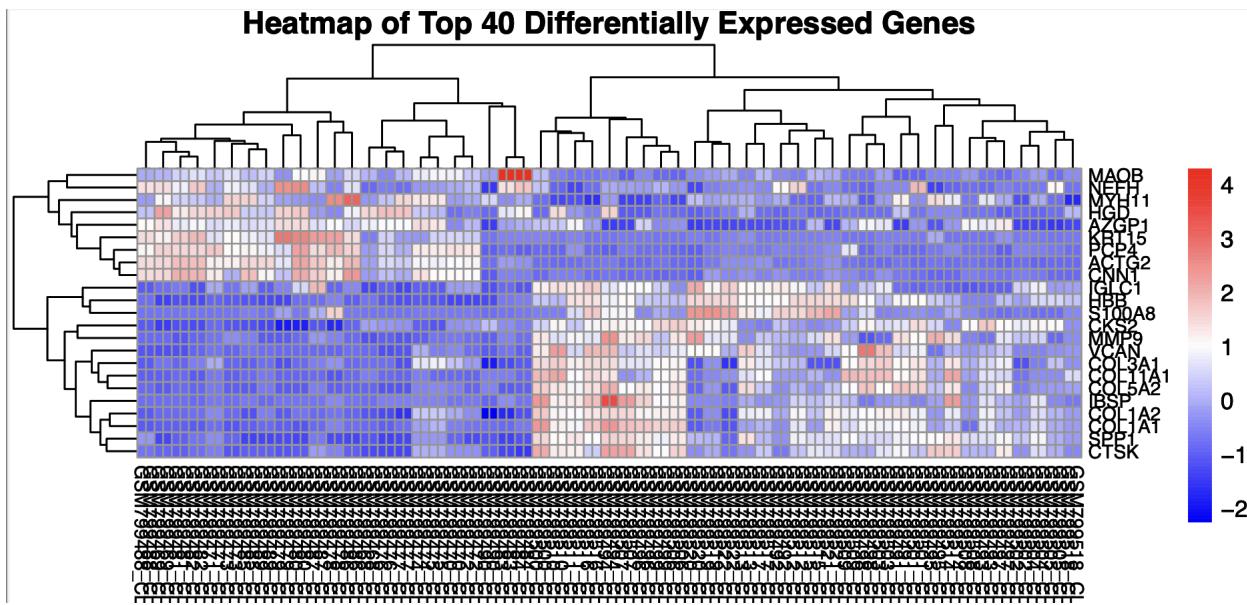
# Remove the last 4 samples (normal samples)
# Assuming that the columns are in the order described
expression_matrix_filtered <- expression_matrix[, 1:51]

# Save the filtered expression matrix as a .tsv file
write.table(expression_matrix_filtered,
            file = "expression_data.tsv",
            sep = "\t",
            quote = FALSE,
            row.names = TRUE,
            col.names = NA)

cat("Filtered expression data saved as 'expression_data.tsv'\n")

```

همچنین توجه کنید که ما به یک mapping بین symbol و ID ژن‌ها هم نیاز داریم و آن را با استفاده از annotation file مربوط به این دیتاست بدست می‌آوریم که از همان سایت geo قابل دانلود است. Annotation file این سمپل در واقع GPL96-57554.txt است که در مستندات پروژه هم قرار داده شده است. با خواندن این مقادیر heatmap را می‌کشیم که به شکل زیر است:



بخش ۳

حال به سراغ آنالیز WGCNA می‌رویم. آنطور که در مقاله ذکر شده این آنالیز با استفاده از سایت sanger box و به صورت آنلاین انجام شده و این سایت کل امکانات مربوط به این آنالیز را فراهم می‌کند. ولی با مراجعه به این سایت مشاهده شد که برای استفاده باید ثبت نام کنیم و برای ثبت نام باید شماره چینی داشته باشیم و شماره مجازی چینی هم نتوانستیم بخریم. در نتیجه مراحل این آنالیز را با استفاده از کد R و پکیج WGCNA انجام دادیم. کدهای مربوط به این بخش در دو فایل R و WGCNA_threshold_plot.R و WGCNA.R نوشته شده است.

ابتدا همانطور که در مقاله گفته شده داده‌های مربوط expression_matrix را می‌خوانیم و آنها را نرمالایز می‌کنیم و فرمت آن را درست می‌کنیم و همچنین داده‌های ناقص را حذف می‌کنیم:

```

# Install packages if not already installed
if (!require("WGCNA")) install.packages("WGCNA")
if (!require("data.table")) install.packages("data.table") # for fast file reading

# Set working directory (adjust the path)
setwd("~/Documents/university") # بیو/بروزه

# Read expression data (assuming tab-delimited with row names)
# Adjust header and separator if needed
exprData <- fread("expression_data.tsv", data.table = FALSE)
# If your gene IDs are in the first column and not already rownames:
rownames(exprData) <- exprData[,1]
exprData <- exprData[,-1]
geneVariance <- apply(exprData, 1, var, na.rm = TRUE)
varianceThreshold <- quantile(geneVariance, 0.2)
filteredExprData <- exprData[geneVariance > varianceThreshold, ]

# Convert data to numeric if not already, and transpose (WGCNA expects samples as rows)
datExpr <- as.data.frame(t(filteredExprData))
# Check for missing values
gsg <- goodSamplesGenes(datExpr, verbose = 3)
if (!gsg$allOK) {
  # Optionally, remove genes/samples with too many missing values:
  datExpr <- datExpr[gsg$goodSamples, gsg$goodGenes]
}

```

حال مشابه مقاله نمودار مربوط به soft threshold power را برای ۱ تا ۳۰ می‌کشیم. این پارامتر در واقع دارد مشخص می‌کند که وزن کوریلیشن بین ژن‌ها چقدر بزرگ باشد و در تحلیل WGCNA باید مشخص شود. هر چقدر این پارامتر بزرگ‌تر باشد یعنی تاثیر کوریلیشن‌های بزرگ‌تر بیشتر است و تاثیر کوچک‌ترها کم می‌شود. ما هم در نهایت مانند مقاله threshold را برابر ۱۴ قرار دادیم. همچنین R^2 را هم مانند مقاله برابر ۰.۸۹ قرار دادیم. کد این بخش به شکل زیر است:

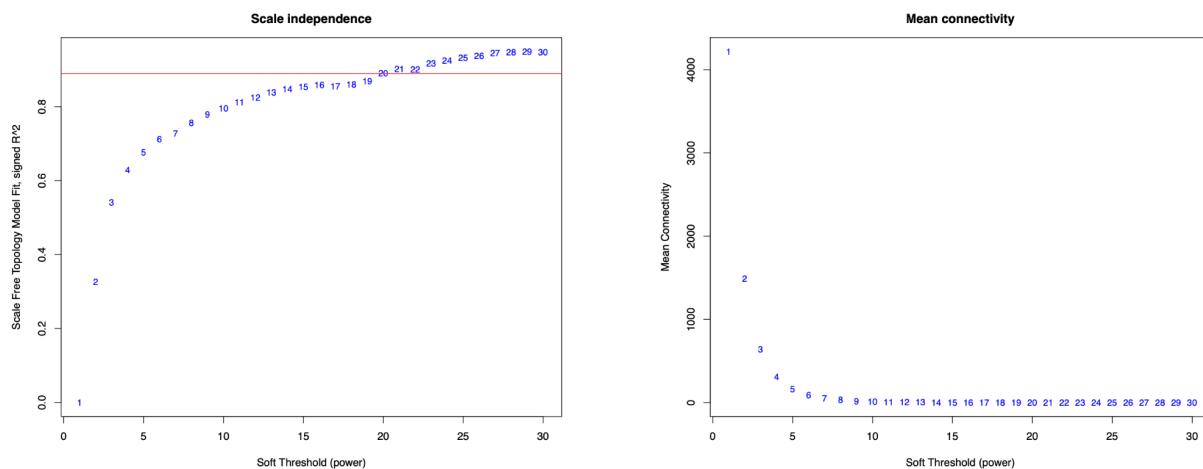
```

# Choose a set of soft-thresholding powers
powers = c(1:30)
# Call the network topology analysis function
sft <- pickSoftThreshold(datExpr, powerVector = powers, verbose = 5)
# Plot the results to help select a power
par(mfrow = c(1,2))
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
      xlab = "Soft Threshold (power)", ylab = "Scale Free Topology Model Fit, signed R^2",
      type = "n", main = "Scale independence")
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
      labels = powers, cex = 0.8, col = "blue")
abline(h = 0.89, col = "red") # you might choose power where R^2 reaches 0.89

plot(sft$fitIndices[,1], sft$fitIndices[,5],
      xlab = "Soft Threshold (power)", ylab = "Mean Connectivity", type = "n",
      main = "Mean connectivity")
text(sft$fitIndices[,1], sft$fitIndices[,5], labels = powers, cex = 0.8, col = "blue")

```

و نمودار آن به این شکل است:



و مشاهده می‌شود که threshold=۱۴ از حداقل مقدار مجاز یعنی ۰.۸۹ کمتر است پس ولید است.

حال برای ادامه کار نیاز داریم که یک trait_data.tsv بسازیم که در واقع شامل این اطلاعات است که هر یک از سمپل‌ها متعلق به کدام یک از دو دسته localized prostate cancer و bone metastases است تا با استفاده از این داده‌ها بتوانیم تاثیر تفاوت بیان این ژن در گسترش بیماری را بهتر درک کنیم و بتوانیم آن‌ها را دسته‌بندی کنیم. برای ساخت این جدول از

کد زیر استفاده کردیم و البته از این دانش که ۲۲ سمپل ابتدایی مربوط به localized prostate و ۲۹ تای بعدی مربوط به bone metastases cancer هستند.

```
# Install packages if not already installed
if (!require("WGCNA")) install.packages("WGCNA")
if (!require("data.table")) install.packages("data.table") # for fast file reading

# Set working directory (adjust the path)
setwd("~/Documents/university/"بیو/بروزه)

# Read expression data (assuming tab-delimited with row names)
# Adjust header and separator if needed
exprData <- fread("expression_data.tsv", data.table = FALSE)
rownames(exprData) <- exprData[,1]
exprData <- exprData[,-1]
# Create a trait vector for the 51 samples:
# 0 for primary samples (first 22) and 1 for metastasis samples (next 29)
trait <- c(rep(0, 22), rep(1, 29))

# Optionally, name the vector with the sample names from the filtered expression data:
sample_names <- colnames(exprData)
names(trait) <- sample_names

# Save the trait data to a file (optional)
write.table(trait, file = "trait_data.tsv", sep = "\t", quote = FALSE, row.names = TRUE, col.names = NA)
cat("Trait data saved as 'trait_data.tsv'\n")
```

حال در کد مربوط به WGCNA این داده‌ها را می‌خوانیم soft threshold power را هم برابر ۱۴ قرار می‌دهیم. حال با استفاده از آن ماتریس مجاورت و سپس topological overlap matrix را ایجاد می‌سازیم:

```
# Ensure samples match between expression and traits
trait <- trait[colnames(exprData), , drop=FALSE]

softPower <- 14
# Create adjacency matrix
adjacency <- adjacency(datExpr, power = softPower)

# Turn adjacency matrix into topological overlap matrix (TOM)
TOM <- TOMsimilarity(adjacency)
dissTOM <- 1 - TOM
```

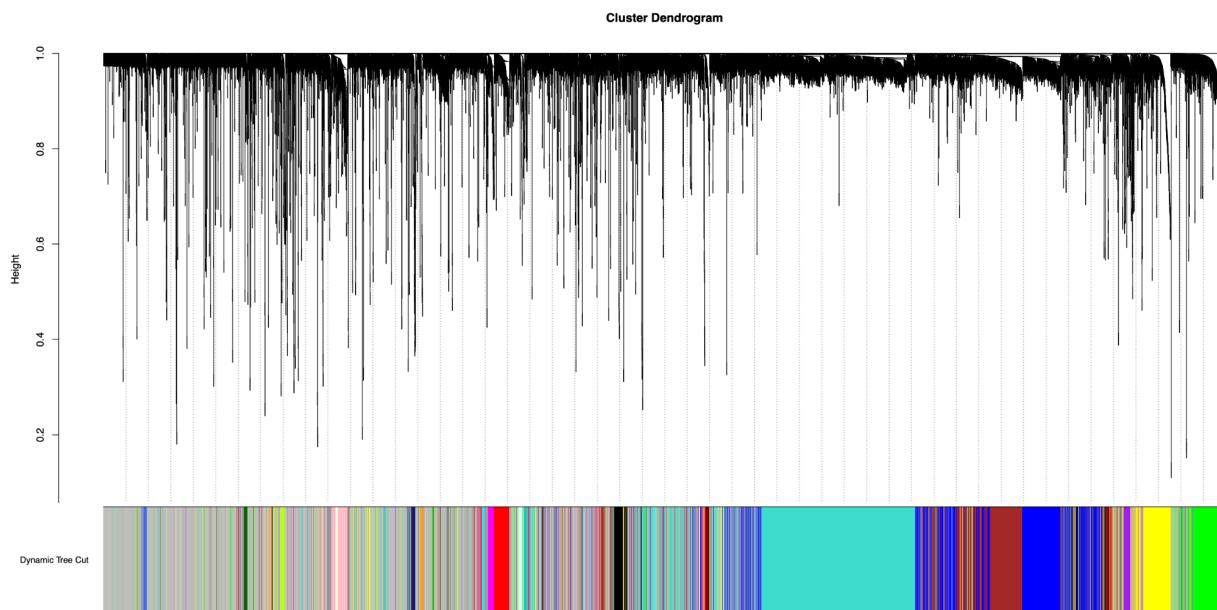
حال با استفاده از کد زیر و روش dynamic cut tree ژن‌ها را مازول بندی می‌کنیم و هر مازول را نیز با یک رنگ نشان می‌دهیم:

```
geneTree <- hclust(as.dist(dissTOM), method = "average")
plot(geneTree, main = "Gene Clustering", xlab = "", sub = "")

# Identify modules using dynamic tree cut
minModuleSize <- 30
dynamicMods <- cutreeDynamic(dendro = geneTree, distM = dissTOM,
                               deepSplit = 2, pamRespectsDendro = FALSE,
                               minClusterSize = minModuleSize)
dynamicColors <- labels2colors(dynamicMods)

# Plot initial modules
plotDendroAndColors(geneTree, dynamicColors, "Dynamic Tree Cut",
                     dendroLabels = FALSE, hang = 0.03, addGuide = TRUE, guideHang = 0.05)
```

در این کد همانند مقاله فرض کردیم که حداقل تعداد ژن موجود در هر مازول باید ۳۰ باشد.
نمودار آن هم به شکل زیر است:



حال همانند مقاله سعی کردیم که با یک merging threshold که برابر ۰.۲۵ بود بعضی از این مازول‌ها را با هم مرج کنیم:

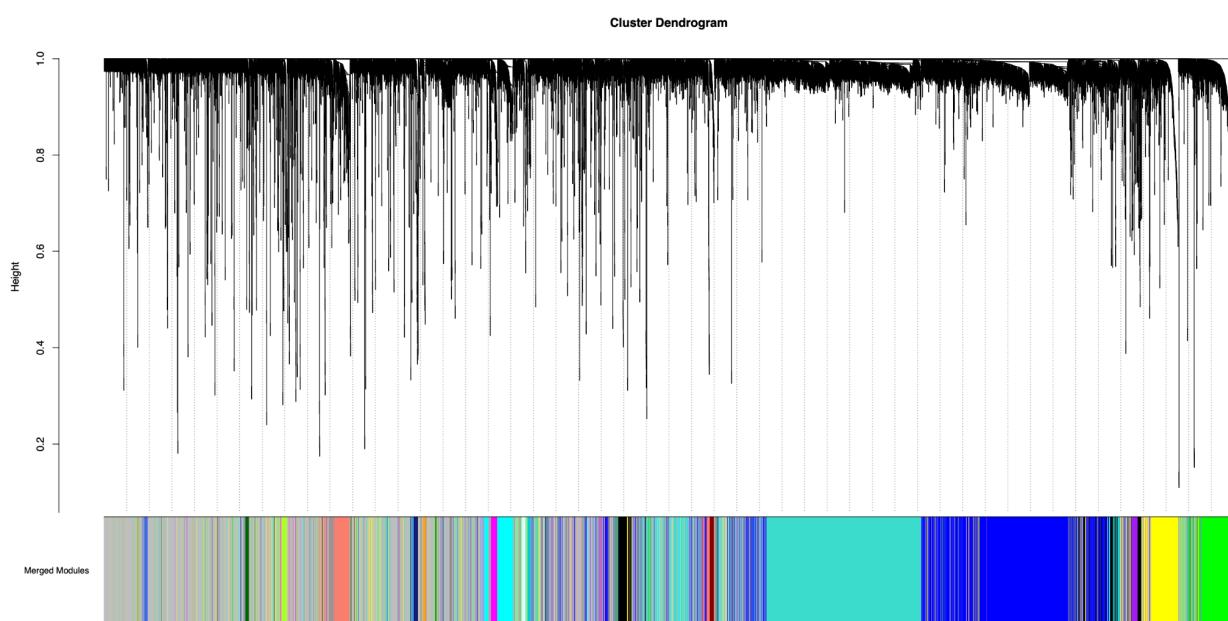
```

# Merge modules with a merging threshold of 0.25
mergedModules <- mergeCloseModules(datExpr, dynamicColors, cutHeight = 0.25, verbose = 3)
mergedColors <- mergedModules$colors
mergedMEs <- mergedModules$newMEs

# Plot dendrogram with merged modules
plotDendroAndColors(geneTree, mergedColors, "Merged Modules",
                     dendroLabels = FALSE, hang = 0.03, addGuide = TRUE, guideHang = 0.05)

```

نتیجه به شکل زیر شد:



حال نیاز داریم که هر یک از این مژول‌ها را بررسی کنیم تا در نهایت یکی از آن‌ها را انتخاب کنیم و فقط داده‌های آن را نگه داریم. برای اینکار ابتدا eigengen‌ها را حساب می‌کنیم. سپس با استفاده از آن‌ها بدست می‌آوریم که میزان کوریلیشن بیان ژن‌های هر مژول نسبت به داده‌های (اینکه بیان ژن به نسبت سمپل‌های دو گروه چگونه تغییر می‌کند) trait چقدر است. همچنین آن را هم بدست می‌آوریم که چقدر احتمال دارد این داده‌ها شانسی وجود داشته باشند. در نهایت مشابه مقاله برای مقایسه مژول‌ها این داده‌ها را پرینت می‌کنیم:

```

# Calculate module eigengenes
MEList <- moduleEigengenes(datExpr, colors = mergedColors)
MEs <- MEList$eigengenes

# Calculate correlations between module eigengenes and trait
trait <- data.frame(MetastasisStatus = trait)
moduleTraitCor <- cor(MEs, trait, use = "p")
moduleTraitPvalue <- corPvalueStudent(moduleTraitCor, nrow(datExpr))

# Define color scheme
textMatrix <- paste(signif(moduleTraitCor, 2), "\n", signif(moduleTraitPvalue, 1), "", sep = "")
par(mar = c(6, 8.5, 3, 3))

# Plot
labeledHeatmap(Matrix = moduleTraitCor,
               xLabels = colnames(trait),
               yLabels = names(MEs),
               ySymbols = names(MEs),
               colorLabels = FALSE,
               colors = blueWhiteRed(50),
               textMatrix = textMatrix,
               setStdMargins = FALSE,
               cex.text = 0.9,
               zlim = c(-1, 1))

```

خروجی به شکل زیر است:



مشاهده می‌شود که مازول با رنگ cyan کمترین p_value و یکی از بیشترین کوریلیشن‌ها را دارد. همچنین از جدول‌های قبلی می‌توانستیم ببینیم که این مازول شامل تعداد قابل توجهی

ژن نیز هست. پس همین مازول را انتخاب می‌کنیم و داده‌های مربوط به آن را در فایل Module_cyan_genes.tsv ذخیره می‌کنیم.

```
module <- "cyan"
moduleGenes <- names(datExpr)[which(mergedColors == sub("ME", "", module))]
write.table(moduleGenes, paste0("Module_", module, "_genes.tsv"), sep = "\t", quote = FALSE, row.names = TRUE)
```

بخش ۴

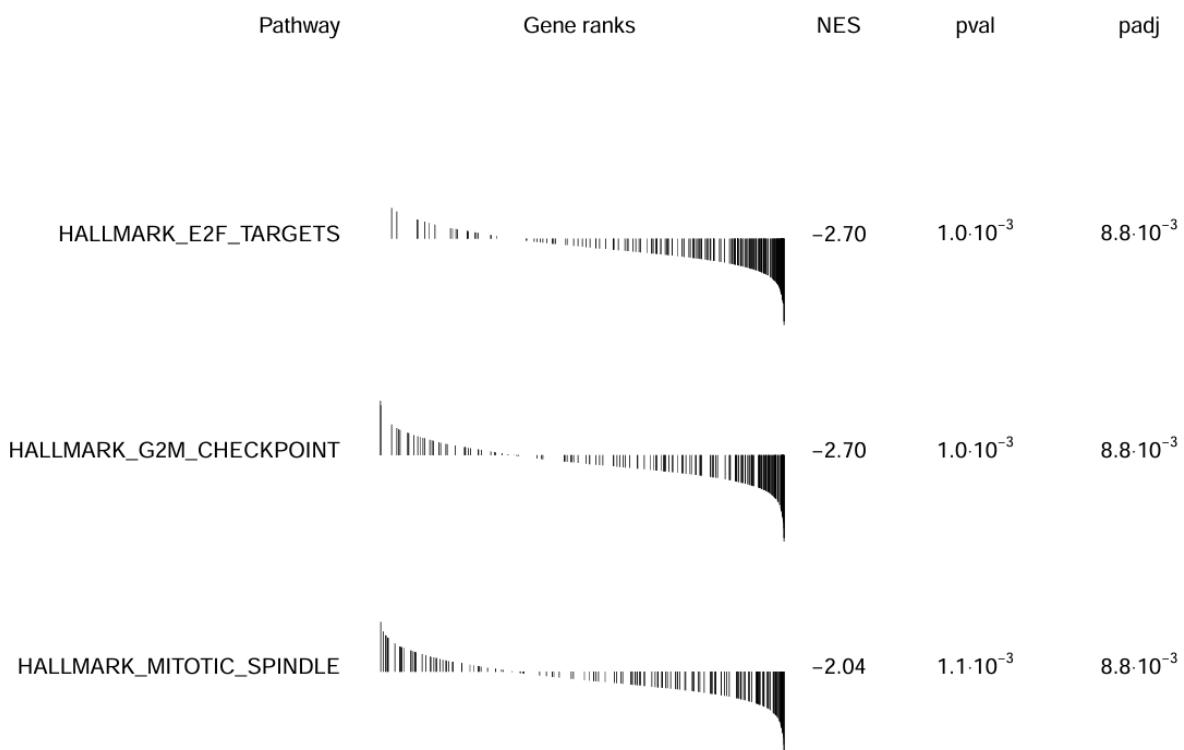
در این بخش مشابه مقاله باید با استفاده از ابزار GSEA، کار pathway analysis را انجام دهیم. با استفاده از کد، ابتدا تحلیل بیان دیفرانسیلی را انجام داده و سپس ژن‌ها را بر اساس مرتب می‌کنیم t-value.

```
25 # Perform differential expression analysis
26 design <- model.matrix(~ metadata$condition) # Adjust based on your metadata
27 fit <- lmFit(data, design)
28 fit <- eBayes(fit)
29 results <- topTable(fit, coef=2, number=Inf, sort.by="t")
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49 # Extract ranking scores (t-values)
50 ranked_genes <- results$t
51 names(ranked_genes) <- rownames(results) # Assign probe IDs first
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
```

در پایان، دیتابیس pathway‌ها به نام MSigDB را لود کرده و با استفاده از ابزار مربوط به GSEA ویژه R (کتابخانه fgsea)، می‌توانیم pathway‌های مربوط به ژن‌ها را مشاهده و تحلیل کنیم.

```
102 # Load MSigDB gene sets
103 msigdb_data <- msigdbr(species = "Homo sapiens", category = "H")
104
105 # Convert to required format
106 pathways <- split(msigdb_data$entrez_gene, msigdb_data$gs_name)
107
108 # Run fgsea
109 fgsea_results <- fgsea(
110   pathways = pathways,
111   stats = ranked_genes, # Fixed ranked genes
112   minSize = 15,
113   maxSize = 500,
114   nperm = 1000
115 )
```

خود مقاله، از روی این تحلیل بدین نتیجه رسیده است که سه مجموعه ژن E2F_TARGETS و G2M_CHECKPOINT، MITOTIC_SPINDLE بیشترین میزان غلظت را داشتند. نمودار enrichment مربوط به این سه دسته ژن هم که در شکل ۳ مقاله آمده، به صورت زیر توسط ما بازترسیم شد.



این نمودار در فایل Figure3.pdf همراه با این گزارش پیوست شده است. کد R این بخش با نام Gene_set_enrichment_analysis_(GSEA)_pathway_analysis_4_2 هم در دسترس است و پیوست شده است.

بخش ۵

در این بخش مشابه مقاله باید از داده هایی که در دو بخش دوم و سوم بدست آوردیم اشتراک بگیریم و اشتراک آنها را به عنوان ژن های با خطر بالا در فایل High_Risk_DEGs.tsv ذخیره کنیم.(این فایل هم در مستندات موجود است) برای این کار از کد زیر استفاده می کنیم:

```

library(VennDiagram)

# Read files
cyan_genes <- read.table("Module_cyan_genes.tsv", header = TRUE, sep = "\t", stringsAsFactors = FALSE)
degs <- read.table("DEGs_filtered.tsv", header = TRUE, sep = "\t", stringsAsFactors = FALSE)

# Convert to character and trim whitespace
cyan_genes$x <- trimws(as.character(cyan_genes$x))
degs$ID <- trimws(as.character(degs$ID))

# Find intersection
common_genes <- degs[degs$ID %in% cyan_genes$x, ]

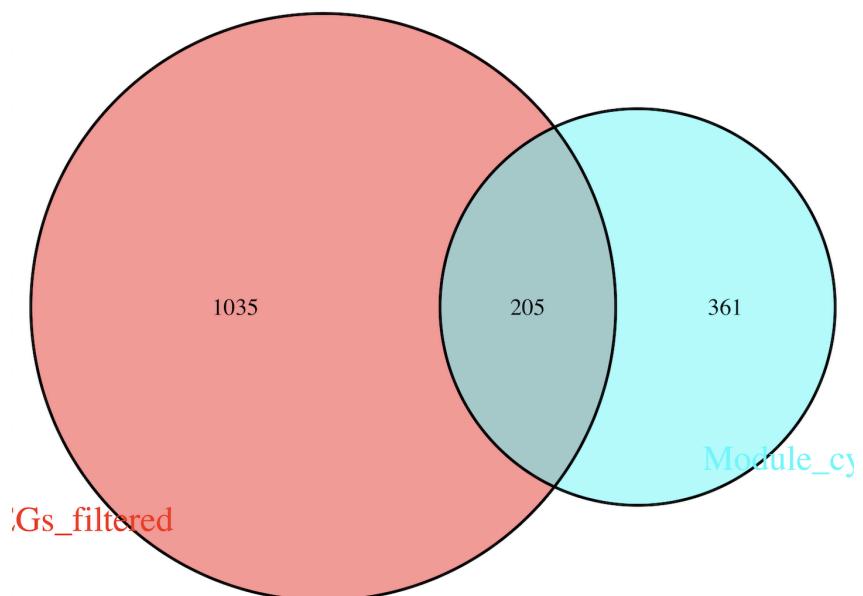
# Save output
write.table(common_genes, "High_Risk_DEGs.tsv", sep = "\t", quote = FALSE, row.names = FALSE)

venn.diagram(
  x = list(
    "Module_cyan_g" = cyan_genes$x,
    "DEGs_filtered" = degs$ID
  ),
  filename = "venn_diagram.png",
  output = TRUE,
  fill = c("cyan", "red"),
  alpha = 0.5,
  cat.col = c("cyan", "red"),
  cat.cex = 1.5
)

# Display the number of common genes
print(paste("Number of common genes:", nrow(common_genes)))

```

و با استفاده از آن نمودار ون این داده‌ها را می‌کشیم که به شکل زیر است:



و اشتراك آنها شامل ۲۰۵ ژن است. توجه كنيد که اعدادي که ما بدست آورديم با مال مقاله تفاوت هايی دارد. اين تفاوتها می تواند در بعضی از جزئيات باشد. مثلا در بخش ۲ در مقاله ذکر نشده که در سایت geo تنظيمات آناليز باید به چه شکل باشد و برای مثال روش محاسبه adjusted p_value با استفاده از کدام يك از آپشن های سایت است. و يا موارد ديگر. در مورد بخش ۳ هم دقيقا در مقاله ذکر نشده بود که روش نرماليزيشن و حذف داده های ناقص به چه صورت است. در نتيجه ماژول بندی ما با مال مقاله تفاوت دارد و در نهايit هم تعداد ژن های خروجي ما با مال مقاله کمي فرق دارد. ولی در نهايit ما به تعداد ژن نزديك تعداد ژن های درون مقاله رسيديم(ما ۲۰۵ ژن بدست آورديم در حالی که در مقاله ۱۹۷ تا بود). اين ژن هايي که الان بدست آورديم در واقع پر ريسک ترين ژن ها هستند. ژن هايي که در صورت بيان زياد آنها انتظار اين است که خطر متاستاز سرطان به استخوان بيشتر شود.

بخش ۶

در اين بخش می خواهيم با استفاده از سایت آناليز <https://davidbioinformatics.nih.gov> functional enrichment را بر روی ۲۰۵ ژنی که تا الان بدست آورديم انجام دهيم. برای اينكار در سایت گزينه start analysis را می زنیم. در صفحه بعدی لیست ۲۰۵ ژنمان را پیست می کنیم و identifier را هم برابر AFFYMETRIX_3PRIME_IVT_ID انتخاب می کنیم:

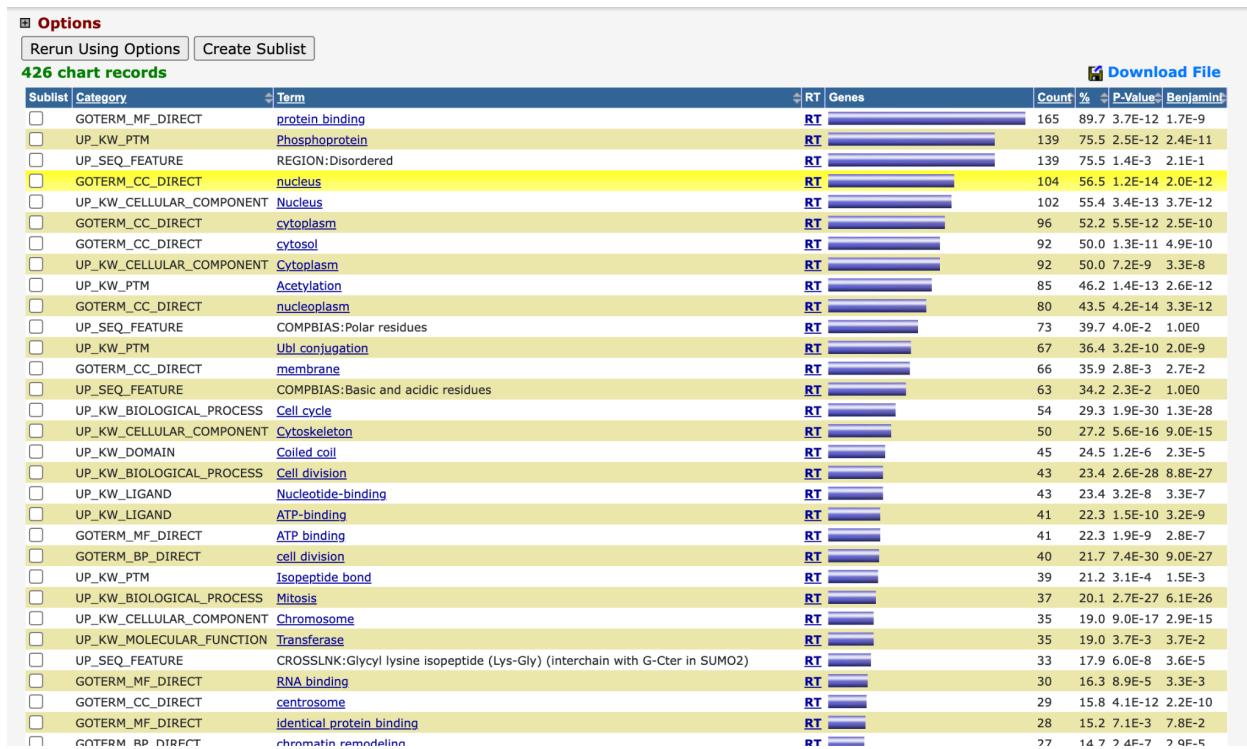
در صفحه بعدی گزینه functional annotation tool می‌کنیم:

The screenshot shows the DAVID Analysis Wizard interface. On the left, the "Gene List Manager" section is visible, showing a list of species: "Homo sapiens(205)". Below this is a "List Manager" section with "List_1" selected. On the right, the "Analysis Wizard" section displays "Step 1. Successfully submitted gene list" with "Current Gene List: List_1" and "Current Background: Homo sapiens". It also shows "Step 2. Analyze above gene list with one of DAVID tools" with a dropdown menu expanded to show "Functional Annotation Tool" (with sub-options: Functional Annotation Clustering, Functional Annotation Chart, Functional Annotation Table), "Gene Functional Classification Tool", "Ortholog Tool", "Gene ID Conversion Tool", and "Gene Name Batch Viewer". A blue arrow points from Step 1 to Step 2.

در صفحه بعدی همانند مقاله گزینه homo sapiens را انتخاب می‌کنیم و سپس annotation chart را می‌زنیم:

The screenshot shows the DAVID Annotation Summary Results page. On the left, the "Gene List Manager" section is visible, showing "Current Gene List: List_1" and "Current Background: Homo sapiens". The "Annotation Summary Results" section on the right shows a list of selected annotations: "184 DAVID IDs", "Check Defaults", "Clear All", and a list of categories with counts: Disease (2 selected), Functional_Annotations (6 selected), Gene_Ontology (3 selected), General_Annotations (0 selected), Interactions (1 selected), Literature (0 selected), Pathways (3 selected), Protein_Domains (4 selected), and Tissue_Expression (0 selected). Below this is a note: "***Red annotation categories denote DAVID defined defaults***". Under "Combined View for Selected Annotation", there are three buttons: "Functional Annotation Clustering", "Functional Annotation Chart", and "Functional Annotation Table".

خروجی به این شکل است. در این جدول ما می‌توانیم برای functionality های سلولی مختلف بدست بیاوریم که ژن‌های انتخابی ما در ارتباط با کدام یک از آن‌ها هستند. در هر سطر یک functionality نوشته شده و همچنین تعداد ژن‌های ما که با آن عمل مرتبط هستند و همچنین p_value آن ثبت شده است. در عکس زیر آن‌ها را بر حسب تعداد ژن‌های مرتبط کردم و همانطور که مشخص است protein binding بیشترین ژن مرتبط را دارد.



حال اگر آن‌ها را بر حسب p_value سورت کنیم مشاهده می‌شود که cell و cell cycle division بیشترین ارتباط را با ژن‌های ما دارند. دانلود شده این داده‌ها در فایل functional_enrichment.txt در مستندات پروژه قرار داده شده است.

حال با استفاده از کد زیر نمودار آن را مانند مقاله برای ۴ گروه از functionality ها می‌کشیم:

```

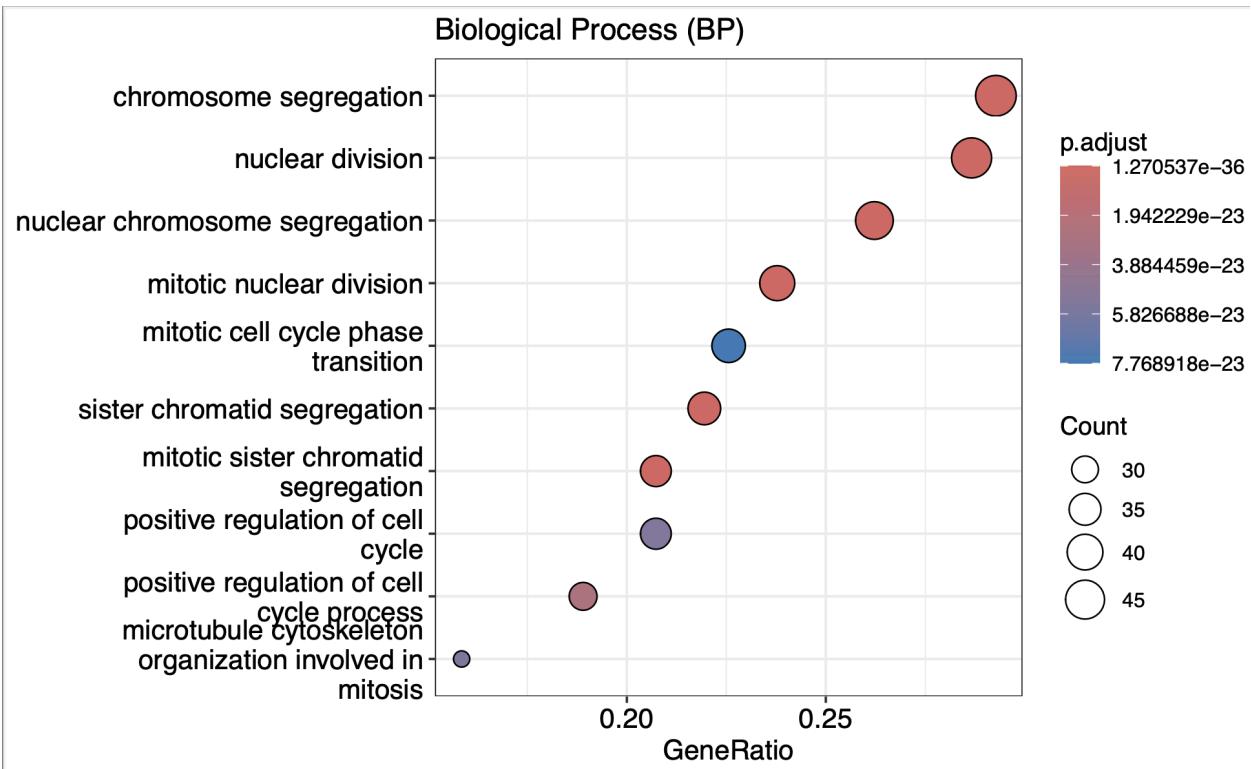
library(ggplot2)
library(readr)
library(org.Hs.eg.db)
setwd("~/Documents/university/("بيو/بروزه/")

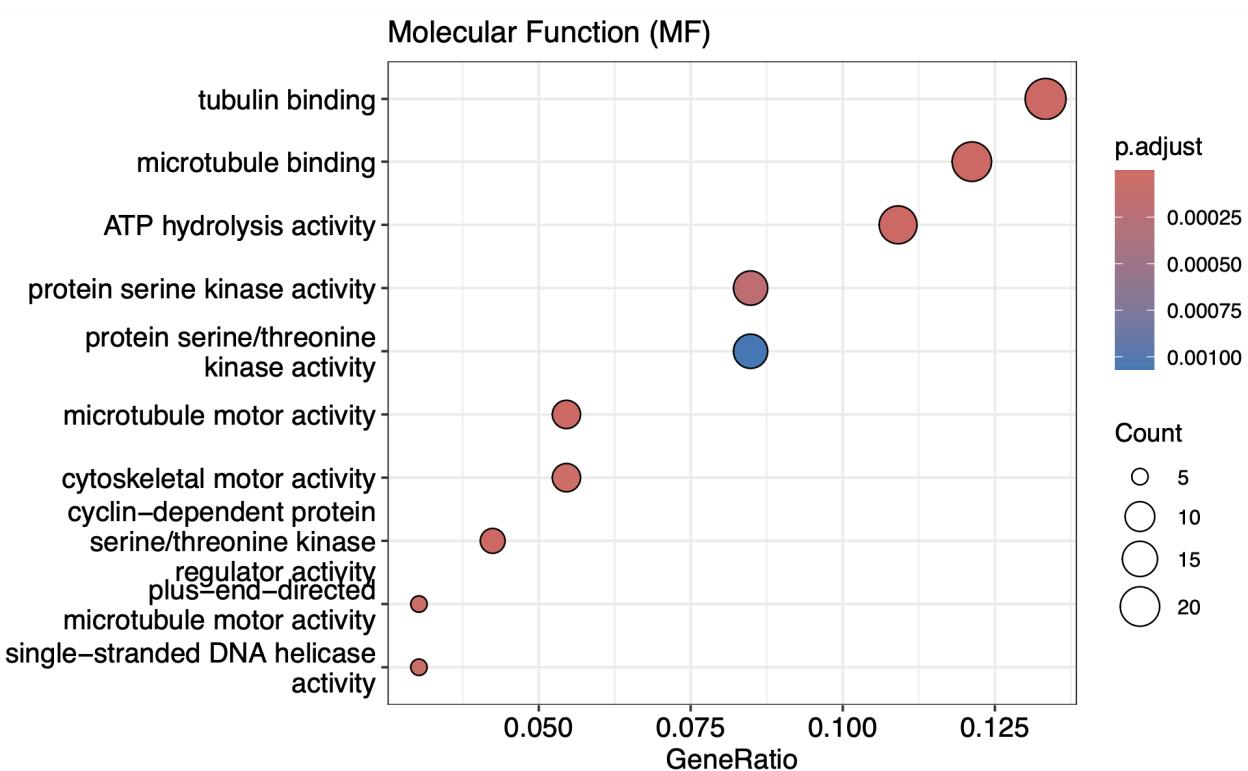
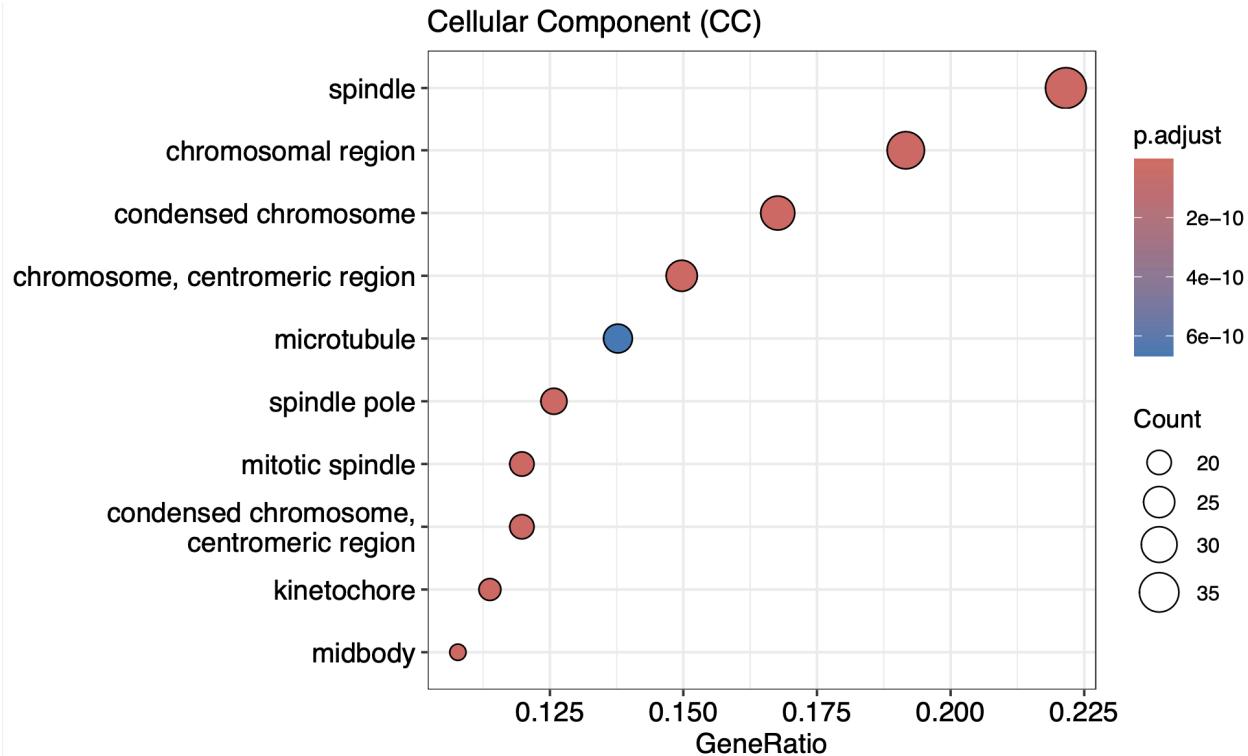
# Read the TSV file (assuming first column contains gene symbols)
genes_df <- read.table("High_Risk_DEGs.tsv", header = TRUE, sep = "\t")
# Extract gene symbols (adjust column name if needed)
genes <- genes_df$Gene.symbol # Replace with the actual column name

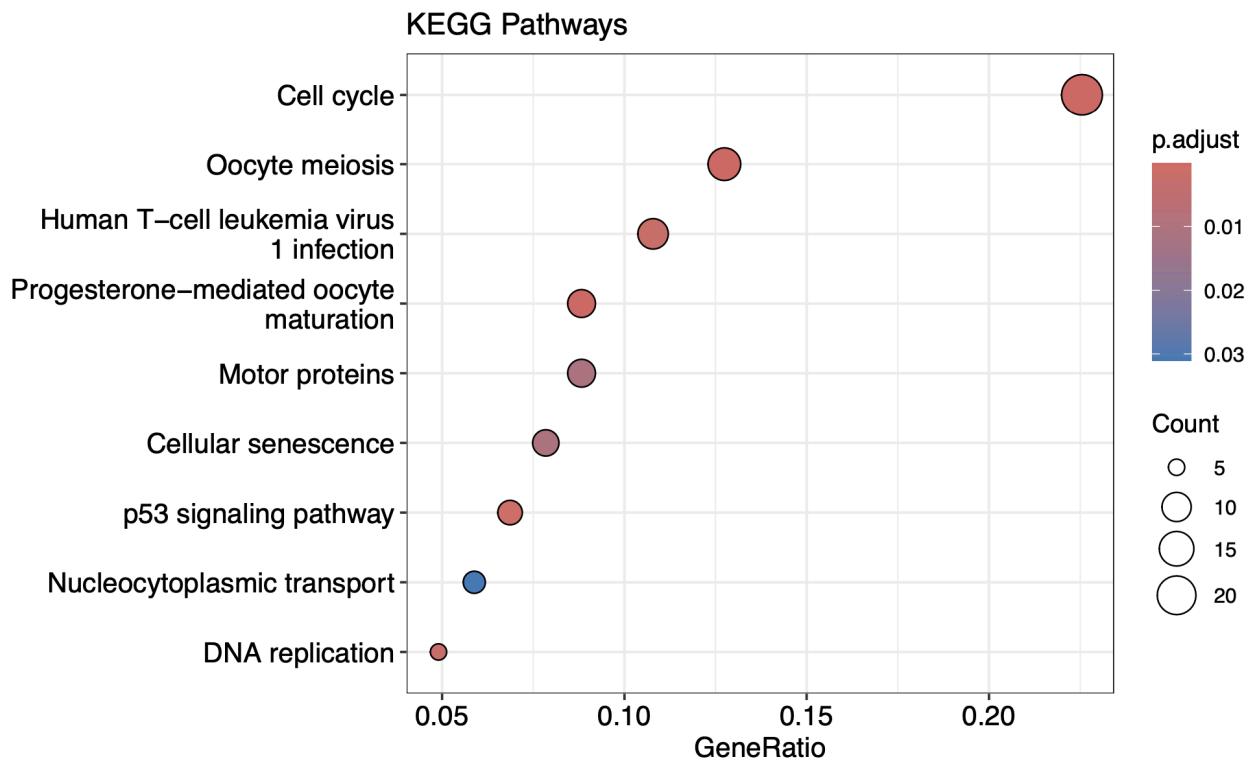
# Convert gene symbols to Entrez IDs
gene_entrez <- bitr(genes, fromType="SYMBOL", toType="ENTREZID", OrgDb=org.Hs.eg.db)
# Perform GO enrichment analysis (BP, CC, MF)
ego_bp <- enrichGO(gene = gene_entrez$ENTREZID, OrgDb = org.Hs.eg.db, keyType = "ENTREZID", ont = "BP")
ego_cc <- enrichGO(gene = gene_entrez$ENTREZID, OrgDb = org.Hs.eg.db, keyType = "ENTREZID", ont = "CC")
ego_mf <- enrichGO(gene = gene_entrez$ENTREZID, OrgDb = org.Hs.eg.db, keyType = "ENTREZID", ont = "MF")
# Perform KEGG pathway analysis
kegg <- enrichKEGG(gene = gene_entrez$ENTREZID, organism = "hsa")

# Plot the results
p <- dotplot(ego_bp, showCategory=10) + ggtitle("Biological Process (BP)")
print(p)
p <- dotplot(ego_cc, showCategory=10) + ggtitle("Cellular Component (CC)")
print(p)
p <- dotplot(ego_mf, showCategory=10) + ggtitle("Molecular Function (MF)")
print(p)
p <- dotplot(kegg, showCategory=10) + ggtitle("KEGG Pathways")
print(p)

```







بخش ۷

در این بخش از مقاله، قصد داریم تا شبکه برهم کنش‌های پروتئین پروتئین را ترسیم کنیم. برای این کار، از دیتابیس String استفاده می‌کنیم. آدرس این دیتابیس [/https://string-db.org](https://string-db.org) است. برای تحلیل با کمک این دیتابیس، ابتدا تحلیل بیان دیفرانسیلی ژن‌ها را در یک فایل csv ذخیره کرده و در دیتابیس String آپلود می‌کنیم. دیتابیس خود اقدام به تحلیل داده‌ها کرده و اطلاعات مربوط به شبکه را به ما می‌دهد.

Version: 12.0 SEPEHR MIZANIAN | LOGOUT | SURVEY

 STRING

Stored Geneset

reference:	"my gene set no. 1"
organism:	Homo sapiens
distinct input items:	5132
distinct genes mapped:	4652
enable statistics use:	yes
last updated:	2025-02-04, 16:37:47

[too large to be viewed as interaction network]

page 1 of 33 first previous next last

input rank	input term	mapped gene	annotation
200	SARAF	SARAF	Store-operated calcium entry-associated regulatory factor; Negative regulator of store-operated C...
199	PPP1CA	PPP1CA	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit; Protein phosphatase that asso...
198	PRDX6	PRDX6	Peroxiredoxin-6; Thiol-specific peroxidase that catalyzes the reduction of hydrogen peroxide and ...
197	EPRS1	EPRS1	Bifunctional glutamate/proline-tRNA ligase; Multifunctional protein which is primarily part of the ...
196	EPRS1	EPRS1	Bifunctional glutamate/proline-tRNA ligase; Multifunctional protein which is primarily part of the ...
195	KARS1	KARS1	Lysine-tRNA ligase; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a ...
194	CTSB	CTSB	Cathepsin B heavy chain; Thiol protease which is believed to participate in intracellular degradatio...
193	BCAP31	BCAP31	B-cell receptor-associated protein 31; Functions as a chaperone protein. Is one of the most abund...
192	RAP1B	RAP1B	Ras-related protein Rap-1b; GTP-binding protein that possesses intrinsic GTPase activity. Contrib...
191	CTR9	CTR9	RNA polymerase-associated protein CTR9 homolog; Component of the PAF1 complex (PAF1C) w...
190	PLOD1	PLOD1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1; Part of a complex composed of PLOD1, P3H3 ...
189	HYOU1	HYOU1	Hypoxia up-regulated protein 1; Has a pivotal role in cytoprotective cellular mechanisms triggered...
188	TPI1	TPI1	Triosephosphate isomerase; Triosephosphate isomerase is an extremely efficient metabolic enzy...
187	TPI1P1	--	--
186	PSMD8	PSMD8	26S proteasome non-ATPase regulatory subunit 8; Component of the 26S proteasome, a multipro...

تحلیل‌های این سایت مربوط به داده‌های میزان غنی بودن پروتئین در بدن و نمودارهای مربوط به gene ontology است.

Version: 12.0 SEPEHR MIZANIAN | LOGOUT | SURVEY

 STRING

Geneset Analysis

Interaction Stats at Default Settings

number of nodes:	4652	expected number of edges:	108654
number of edges:	119258	PPI enrichment p-value:	<1.0e-16
average node degree:	51.3	your network has significantly more interactions than expected (what does that mean?)	
avg. local clustering coefficient:	0.302		

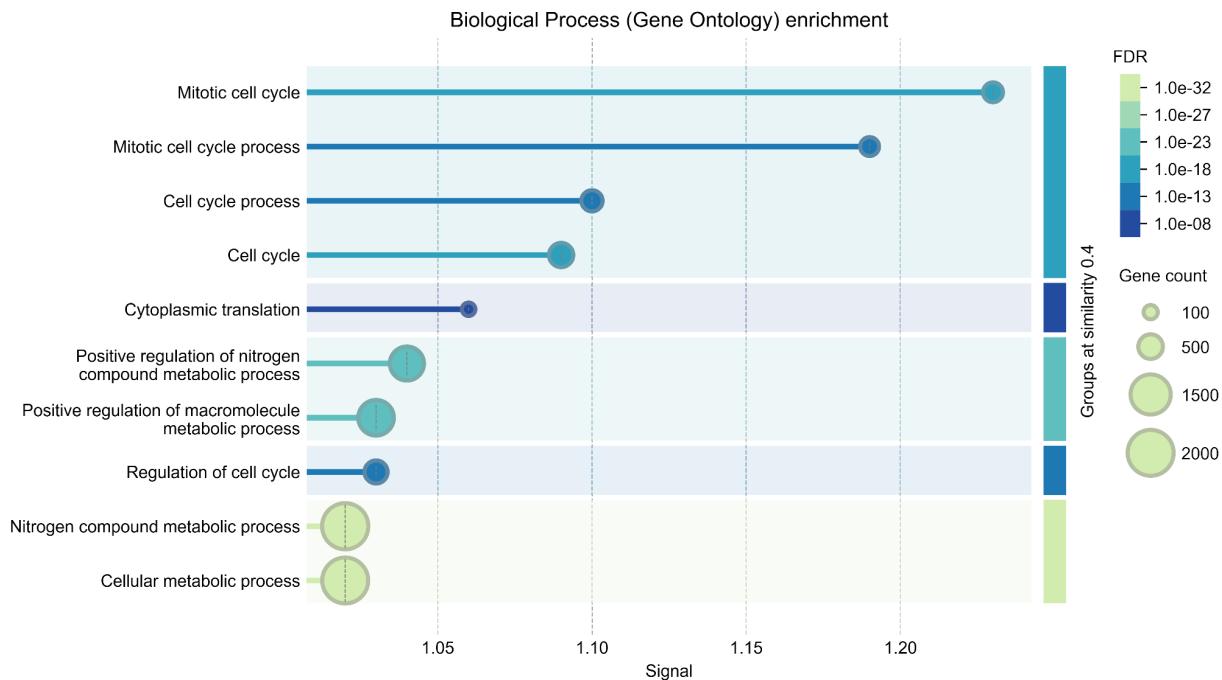
Functional enrichments in your set of proteins

explain columns

Biological Process (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0000278	Mitotic cell cycle	287 of 631	0.28	1.23	7.67e-17
GO:1903047	Mitotic cell cycle process	248 of 537	0.29	1.19	5.13e-15
GO:0022402	Cell cycle process	342 of 835	0.24	1.1	6.45e-15
GO:0007049	Cell cycle	479 of 1246	0.21	1.09	4.50e-17
GO:0002181	Cytoplasmic translation	76 of 123	0.42	1.06	3.25e-08

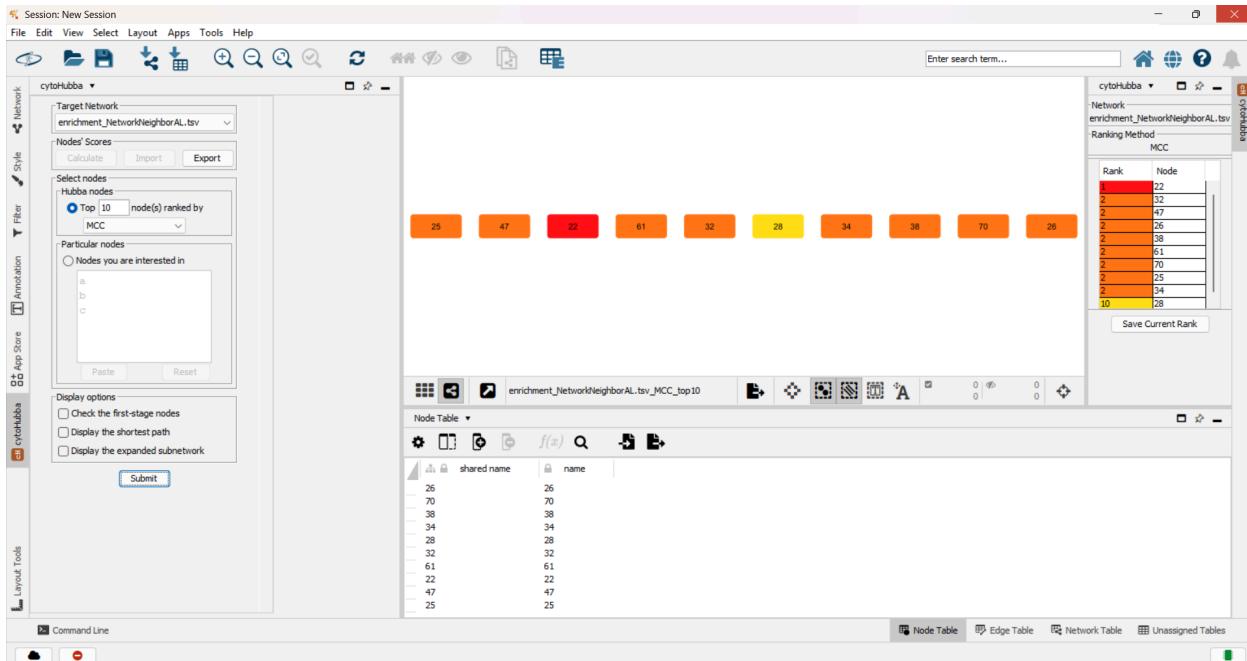
Molecular Function (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0019899	Enzyme binding	754 of 2084	0.19	1.11	2.05e-22
GO:0003723	RNA binding	613 of 1672	0.19	1.08	1.24e-18
GO:0005515	Protein binding	2188 of 7242	0.11	1.02	1.63e-34
GO:1901363	Heterocyclic compound binding	1825 of 5977	0.11	1.0	8.64e-28
GO:0097159	Organic cyclic compound binding	1838 of 6050	0.11	0.99	3.49e-27

Cellular Component (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate



در نهایت، یک فایل با فرمت tsv از این سایت خروجی گرفته می‌شود که اطلاعات مربوط به گراف pathway ژن‌هاست. حالا باید این اطلاعات را با استفاده از ابزار Cytoscape بررسی کنیم.

هنگامی که یک فایل tsv را با استفاده از Cytoscape بررسی می‌کنیم، باید مشخص کنیم که کدام ستون‌ها و کدام target score هستند. پس از انجام این کار، با استفاده از پلاگین cytoHubba، می‌توانیم ده Hub برتر این شبکه را مشخص کنیم.



در نهایت هم می‌توان با استفاده از R، گرافی مشابه شکل 6 مقاله ترسیم کرد که در فایل Figure6.pdf پیوست شده است. مطابق با نتایج، که در مقاله نیز اشاره شد، ده هاب ژنی در RACGAP1، NUSAP1، .NCAPG و ASPM، CDK1، AURKA، CCNA2، PBK، DLGAP5، KIF4A

بخش ۸

در این بخش از مقاله، قصد داریم از بین ده هاب ژنی برتر، ژن‌های با بیشترین بیان را شناسایی کرده و ببینیم آیا بیان آن‌ها ارتباط معنی‌داری با متاستاز استخوان و کاهش بقای بیمار دارد یا خیر. برای این منظور، از دیتابیس cBioPortal استفاده می‌کنیم.

The screenshot shows the cBioPortal homepage. At the top, there's a navigation bar with links for Data Sets, Web API, Tutorials/Webinars, FAQ, News, Visualize Your Data, About, cBioPortal Installations, and Donate. A 'Login' button is also present. Below the navigation is a green banner with the text 'Re-introducing the cBioPortal Newsletter! Subscribe via LinkedIn or Google Groups'. The main content area has tabs for 'Query' and 'Quick Search Beta!'. A search bar at the top right includes a 'Please cite cBioPortal' link. On the left, there's a sidebar titled 'Select Studies for Visualization & Analysis:' with a list of study categories and their counts: PanCancer Studies (11), Pediatric Cancer Studies (14), Immunogenomic Studies (8), Cell lines (3), PreCancerous Studies (1), Adrenal Gland (3), Ampulla of Vater (1), Biliary Tract (16), Bladder/Urinary Tract (22), Bone (4), and Bowel (26). A 'Quick select' dropdown shows 'TCGA PanCancer Atlas Studies' and 'Curated set of non-redundant studies'. To the right of the study list is a table of sample counts for various cancer types, such as 25040 samples for MSK-CHORD and 25775 samples for MSK MetTropism. Below the study list are buttons for 'Query By Gene' and 'Explore Selected Studies'. On the far right, a 'What's New' section lists recent datasets added to the portal, including CHORD, cDNA Sequencing Cohort, Prostate Cancer, Hepatocellular Carcinoma, Pancreatic Cancer, and Metastatic Pancreatic Neuroendocrine Tumor. There's also a 'Read the latest cBioPortal Newsletter!' link and social media sharing buttons for LinkedIn and Google Groups.

با استفاده از تحلیل بیان تفاضلی بین این ده هاب ژنی در بین بیماران با سرطان پروستات اولیه و متاستاز استخوان، متوجه می‌شویم که ژن‌های CCNA2، NUSAP1 و PBK دارای بیشترین اثر هستند.

در نهایت، با استفاده از Kaplan–Meier survival analyses، میزان اثر این ژن‌ها روی بقا بررسی می‌شود. این یک روش آماری غیرپارامتری جهت بررسی این است که چند درصد از افراد پس از یک واقعه مانند سرطان زنده می‌مانند.

```

59 # Kaplan-Meier Survival Analysis
60 surv_data <- metadata[, c("survival_time", "survival_status")]
61 surv_data$survival_status <- as.numeric(surv_data$survival_status)
62
63 # Loop through each hub gene
64 for (gene in hub_genes) {
65   high_expr <- expr_data[gene, ] > median(expr_data[gene, ])
66   surv_data$group <- ifelse(high_expr, paste("High", gene, "Group"), paste("Low", gene, "Group"))
67
68 # Fit Kaplan-Meier Model
69 km_fit <- survfit(Surv(survival_time, survival_status) ~ group, data = surv_data)
70
71 # Plot Kaplan-Meier Curve
72 ggsurvplot(km_fit, data = surv_data, pval = TRUE, conf.int = TRUE,
73             risk.table = TRUE, ggtheme = theme_minimal(),
74             title = paste("Kaplan-Meier Survival Curve for", gene))
75 }
```

برای این تحلیل در R از دو پکیج survival و survminer استفاده می‌شود و در نهایت نمودار برای ژن‌ها رسم می‌شود که بیانگر احتمال بقای تجمعی بر حسب زمان است و Kaplan–Meier

به مرور زمان افت پیدا می‌کند. اگر تفاوت میزان بقای تجمعی برای دو گروه در افتها بیشتر از 0.05 باشد یعنی اثر این ژن معنی‌دار است.

بخش ۹

در این بخش ابتدا به سایت CIBERSORTx Impute Fractions رفته و گزینه CIBERSORTx را انتخاب می‌کنیم. هدف ما بررسی میزان نسبت سلول‌های ایمنی در سمپل‌های ژنی متاستاز استخوان و غیر متاستاز است. بدین منظور فایل Mixture سمپل‌ها که همان فایل GSE32269 است را در سایت آپلود می‌کنیم.

User data generated before 8/15/23 currently not available.

The screenshot shows the CIBERSORTx website. At the top, there is a red banner with the text "User data generated before 8/15/23 currently not available.". Below this, the CIBERSORTx logo is displayed. A green box contains the text: "CIBERSORTx is an analytical tool from the Alizadeh Lab and Newman Lab to impute gene expression profiles and provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data." Another green box below it says: "EcoTyper, a machine learning framework for the identification of cell states and ecosystems from bulk, single-cell, and spatially-resolved expression data, is now available. EcoTyper extends CIBERSORTx for large-scale profiling of cellular ecosystems." A yellow box at the bottom left contains the text: "The CIBERSORT legacy website has been incorporated into the CIBERSORTx website under the menu item CS Archives." A green box at the bottom right contains the text: "Notice: user login is back up. We apologize for the inconvenience and appreciate your patience."

پس از آپلود فایل Signature Matrix مربوط به 22 سلول ایمنی مانند مقاله انتخاب می‌کنیم. همینطور تعداد جایگشت‌ها را برابر 500 قرار می‌دهیم.

Select Analysis Mode:

[Example](#) [Custom](#)

Configure Custom Input:

Signature matrix file* [LM22 \(22 immune cell types\)](#) [More Information...](#)

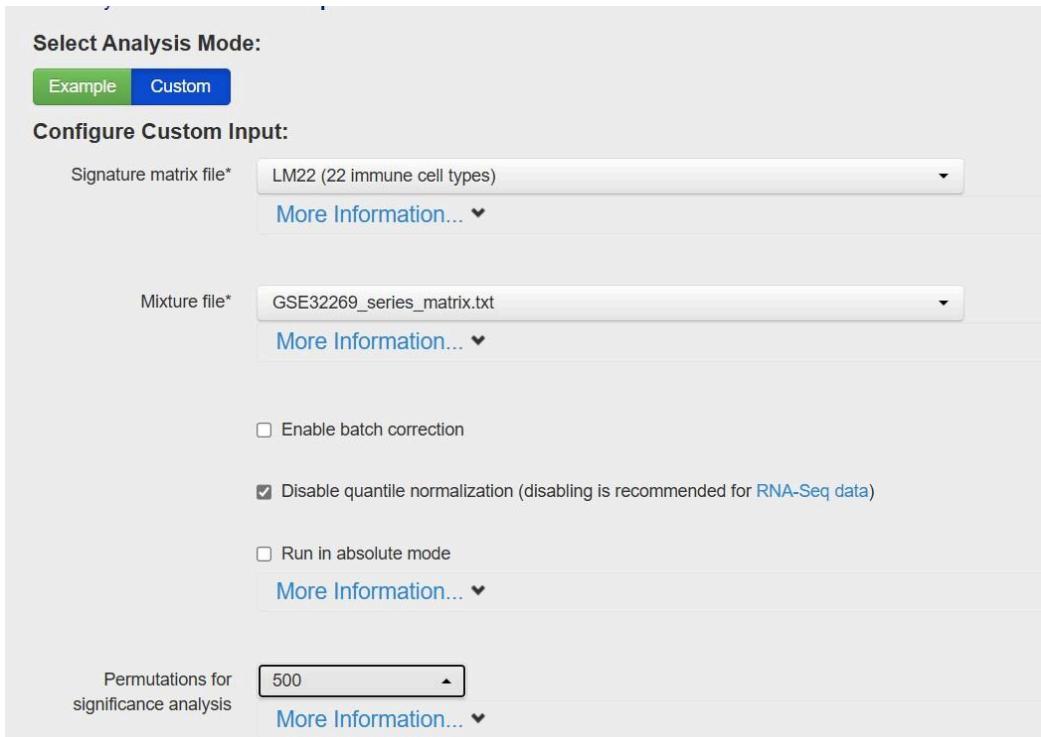
Mixture file* [GSE32269_series_matrix.txt](#) [More Information...](#)

Enable batch correction

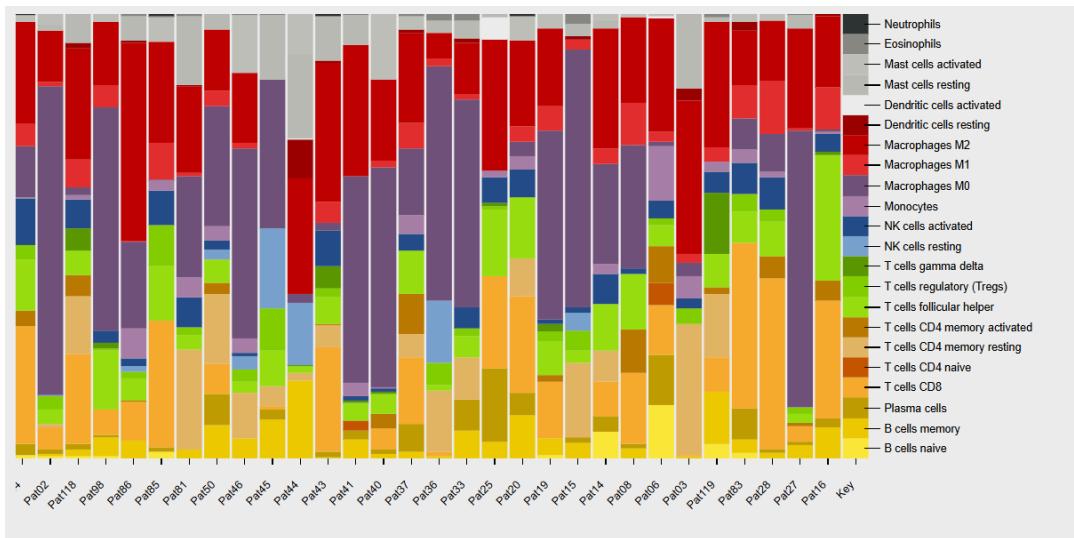
Disable quantile normalization (disabling is recommended for RNA-Seq data)

Run in absolute mode
[More Information...](#)

Permutations for significance analysis [500](#) [More Information...](#)



با اجرای Run الگوریتم Reverse Convolution اجرا شده و خروجی به صورت زیر خواهد بود.



Input Sample	B cells naive	B cells memory	Plasma cells	T cells CD8 naive	T cells CD4 memory resting	T cells CD4 memory activated	T cells CD4 follicular helper	T cells regulatory (Tregs)	T cells gamma delta	NK cells resting	NK cells activated	Monocytes	Macrophages M0	Macrophages M1	Macrophages M2	Dendritic cells resting	Dendritic cells activated	Mast cells resting	Mast cells activated	Eosinophils	P-value	Correlation			
Pat26	0.016	0.03	0.106	0	0.052	0.038	0.167	0	0	0.04	0.03	0.054	0.076	0.02	0	0	0.11	0.000	0.396	0.018					
Pat23	0.044	0	0.055	0.221	0	0	0.099	0.166	0	0	0.07	0.002	0	0.039	0.040	0.014	0	0	0.001	0.074	0.144				
Pat91	0.001	0.026	0.027	0.15	0	0	0	0.118	0	0	0.055	0.03	0.175	0.072	0.167	0.023	0	0	0.136	0.019	0.000	0.144			
Pat50	0.084	0.129	0.073	0	0.048	0.019	0.113	0	0.039	0.009	0	0.147	0.084	0.248	0	0	0.015	0	0	0.000	0.010	0.299			
Pat86	0.000	0.019	0.006	0	0	0	0.027	0.115	0	0	0.111	0.032	0.176	0.085	0.061	0	0	0	0.039	0.000	0.214	0.058			
Pat80	0.007	0.012	0.001	0.146	0	0	0.02	0.048	0.029	0	0	0	0	0.046	0.11	0	0	0	0.041	0	0.000	0.140	0.087		
Pat79	0.072	0	0.049	0.25	0	0	0.123	0.135	0.01	0	0.012	0.02	0.051	0.054	0.206	0.005	0	0.012	0	0	0.000	0.072	0.147		
Pat49	0.011	0.048	0.076	0.204	0	0	0.017	0.112	0.02	0	0	0.094	0.025	0.03	0.099	0.075	0.013	0	0	0.02	0.009	0.000	0.032	0.203	
Pat47	0.072	0	0.057	0.055	0	0	0.051	0.098	0	0	0.001	0	0	0.017	0.244	0	0.03	0	0.067	0	0.000	0.070	0.150		
Pat39	0	0.053	0	0.086	0	0.217	0	0.007	0.005	0.018	0	0.018	0	0.033	0	0.009	0.249	0	0	0.022	0.009	0	0.000	0.425	
Pat38	0.014	0	0.005	0.263	0	0	0.031	0.081	0	0	0.092	0.006	0.238	0.073	0.184	0	0	0	0.013	0	0.000	0.034	0.198		
Pat29	0.122	0.198	0.1	0.009	0.016	0.102	0.027	0.058	0.022	0	0.006	0.017	0	0.171	0.027	0.124	0	0	0.001	0	0	0.000	0.002	0.500	
Pat04	0.008	0.025	0.044	0	0	0.034	0.115	0.032	0	0	0.106	0.004	0.117	0.05	0.229	0.004	0	0	0.014	0	0.000	0.198	0.062		
Pat02	0.006	0.003	0.01	0.05	0	0.007	0.004	0.031	0.031	0	0.003	0	0	0.011	0.115	0	0	0	0.014	0.024	0	0.000	0.000	0.774	
Pat118	0.004	0.016	0.012	0.204	0	0.13	0.047	0.054	0	0.052	0	0.064	0.009	0.018	0.062	0.251	0.013	0	0.064	0	0	0.000	0.162	0.075	
Pat98	0.005	0.043	0.004	0.06	0	0	0.131	0.006	0.012	0	0.028	0	0	0.05	0.143	0	0	0	0.017	0	0.000	0.054	0.168		
Pat86	0	0.041	0	0.087	0	0	0.002	0.051	0.015	0	0.013	0.017	0.068	0.194	0.001	0.004	0	0.055	0	0.006	0.000	0.176	0.069		
Pat55	0.015	0	0.009	0.0743	0	0	0	0.123	0.092	0	0	0.079	0.022	0.001	0.082	0.228	0	0	0	0.056	0.000	0.005	0.126	0.096	
Pat81	0	0.02	0.001	0	0	0.225	0	0.032	0.018	0	0	0.067	0.045	0.228	0.009	0.195	0.003	0	0.156	0	0.005	0.000	0.432	0.014	
Pat50	0	0.074	0.072	0.067	0	0.156	0.026	0.053	0	0	0.023	0.019	0.033	0.271	0.033	0.139	0	0	0.035	0	0	0.000	0.034	0.197	
Pat46	0	0.044	0	0	0	0.103	0	0.027	0.028	0	0.028	0.009	0.033	0	0.012	0.157	0	0	0	0.132	0.002	0.000	0.032	0.203	
Pat45	0	0.088	0.023	0.006	0	0.045	0	0.081	0.094	0	0.182	0	0	0.017	0	0	0.020	0	0.144	0	0.009	0.000	0.378	0.022	
Pat44	0	0.174	0	0	0.001	0.017	0	0.014	0	0.004	0.14	0	0	0.019	0	0.24	0.089	0.001	0.189	0.091	0	0.000	0.440	0.012	
Pat43	0	0.063	0.013	0.235	0	0.05	0.003	0.06	0.02	0.05	0	0.081	0	0.017	0.049	0.31	0	0.006	0	0.099	0	0	0.004	0.238	0.048
Pat41	0	0.042	0.02	0	0.024	0	0	0.036	0.003	0.004	0	0.011	0.03	0	0	0	0	0	0.066	0	0.003	0.000	0.212	0.058	
Pat40	0	0.011	0.001	0.048	0	0	0.033	0.045	0	0.005	0	0.007	0.005	0.047	0	0.016	0.182	0	0	0	0.147	0	0.000	0.006	0.382
Pat37	0	0.015	0.062	0.151	0	0.052	0.09	0.096	0	0	0.038	0.044	0.227	0	0.020	0.009	0	0	0.031	0.004	0.000	0.040	0.186		
Pat36	0	0.065	0	0.009	0	0.138	0	0.013	0.049	0	0.142	0	0	0.016	0.059	0	0	0	0.027	0	0.015	0.000	0.092	0.116	
Pat33	0	0.063	0.07	0	0	0.096	0	0.047	0.017	0	0.048	0	0	0.017	0.114	0.011	0	0.038	0	0.016	0.000	0.186	0.066		
Pat25	0	0.038	0.166	0.207	0	0	0	0.149	0.007	0.008	0	0.058	0.014	0	0	0.047	0	0.051	0.007	0	0	0.000	0.010	0.293	
Pat20	0	0.098	0.049	0.219	0	0.085	0	0.137	0	0	0	0.062	0.031	0.032	0.035	0.191	0	0	0	0.055	0.000	0.004	0.034	0.197	
Pat19	0.008	0.037	0	0.128	0	0	0.014	0.075	0.022	0.019	0	0.001	0	0.021	0	0.055	0.175	0	0	0	0.033	0	0.000	0.078	0.139
Pat15	0	0.034	0.014	0	0	0.168	0	0.025	0.045	0	0.041	0.013	0	0	0.024	0	0.006	0	0.027	0	0.024	0.000	0.078	0.139	
Pat14	0.06	0	0.034	0.077	0	0.071	0	0.105	0	0	0	0.068	0.02	0.027	0.034	0.271	0	0	0.021	0.012	0	0.000	0.074	0.144	
Pat08	0	0.023	0.008	0.161	0	0	0.098	0.124	0	0	0.014	0	0.275	0	0.096	0.194	0	0	0	0.006	0	0.000	0.154	0.079	
Pat06	0.12	0	0.113	0.112	0.051	0	0.081	0.049	0.014	0	0.04	0.123	0.01	0.023	0.255	0	0.006	0	0.005	0	0.000	0.008	0.334		
Pat02	0	0.005	0	0.008	0	0	0.001	0.001	0	0.003	0	0.023	0.049	0.031	0.018	0.028	0	0.028	0	0.168	0	0	0.000	0.196	0.064
Pat119	0.032	0.118	0	0.077	0	0.144	0.019	0.075	0	0.136	0	0.05	0.021	0	0.033	0.249	0	0	0.01	0	0.007	0	0.000	0.402	0.017
Pat83	0.014	0.029	0.07	0.173	0	0	0.071	0.04	0	0	0	0.07	0.03	0.069	0.076	0.122	0.02	0	0	0.019	0	0.000	0.048	0.175	
Pat28	0	0.012	0.007	0.169	0	0	0.051	0.076	0.028	0	0	0.074	0.011	0.086	0.121	0.133	0	0	0	0.016	0	0.000	0.032	0.202	
Pat27	0	0.031	0.006	0.035	0	0.007	0.009	0.021	0.013	0	0	0	0	0	0.001	0.226	0	0	0	0.029	0	0.003	0.004	0.447	
Pat16	0	0.07	0.02	0.003	0	0	0.045	0.002	0.008	0	0.039	0.006	0.006	0	0.095	0.16	0.004	0	0	0	0	0.000	0.080	0.137	

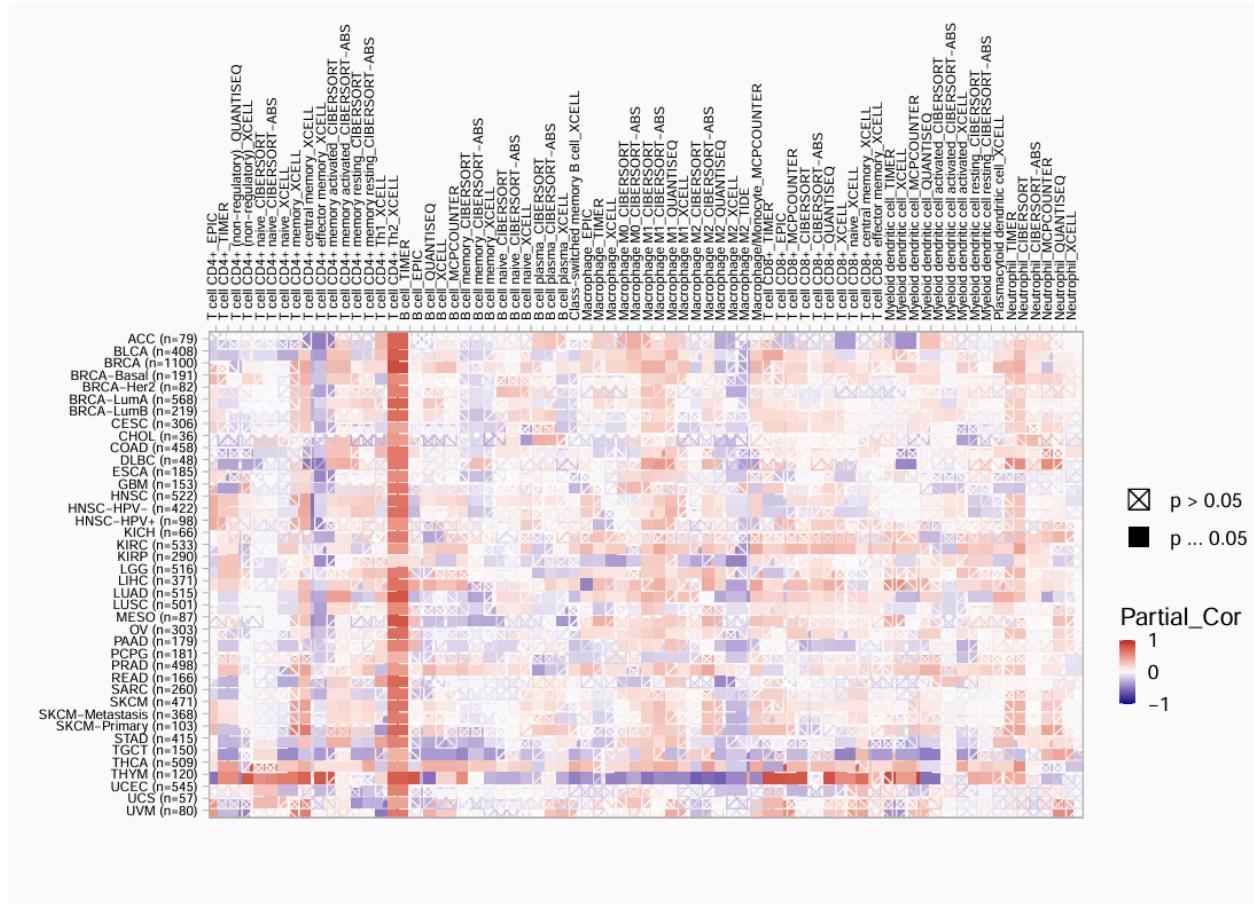
می توان مشاهده کرد که اشخاصی که مبتلاست از سرطان پروستات داشتند ، دارای درصد بالاتری از ماکروفاز M0 و سلول های Treg بوده اند.

در ادامه به بررسی سلول های ایمنی در سه ژن هاب به دست آمده از بخش های قبلی یعنی NUSAP1 و CCNA2 ، PBK می پردازیم. بدین وسیله از دیتابیس TIMER استفاده می کنیم.

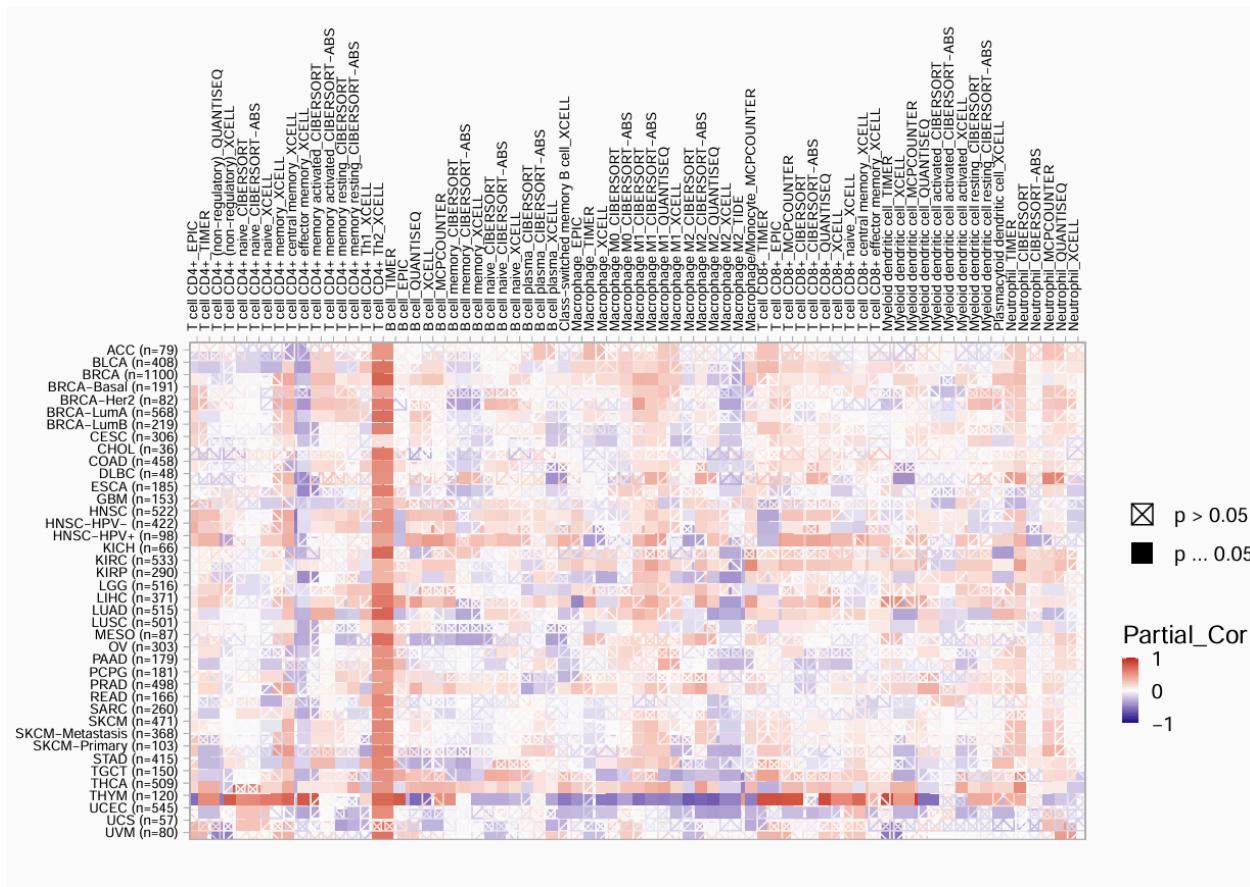
به صورت زیر مانند مقاله 6 سلول ایمنی اصلی را انتخاب کرده و ارتباط آنها را با هر کدام از ژن هاب ها می سنجیم.

The screenshot shows the TIMER2.0 web application interface. At the top, there is a navigation bar with tabs: Home, Immune, Exploration, and Estimation. The 'Immune' tab is currently selected. Below the navigation bar, there is a sub-navigation bar with tabs: Gene (selected), Mutation, sCNA, and Outcome. The main content area has two sections: 'Gene Expression:' and 'Immune Infiltrates:'. Under 'Gene Expression:', a dropdown menu is open, showing 'CCNA2' as the selected value. Under 'Immune Infiltrates:', several buttons are displayed, each representing a type of immune cell infiltration: T cell CD8+, B cell, Macrophage, T cell CD4+, Neutrophil, and DC. A blue 'Submit' button is located at the bottom left of the main content area.

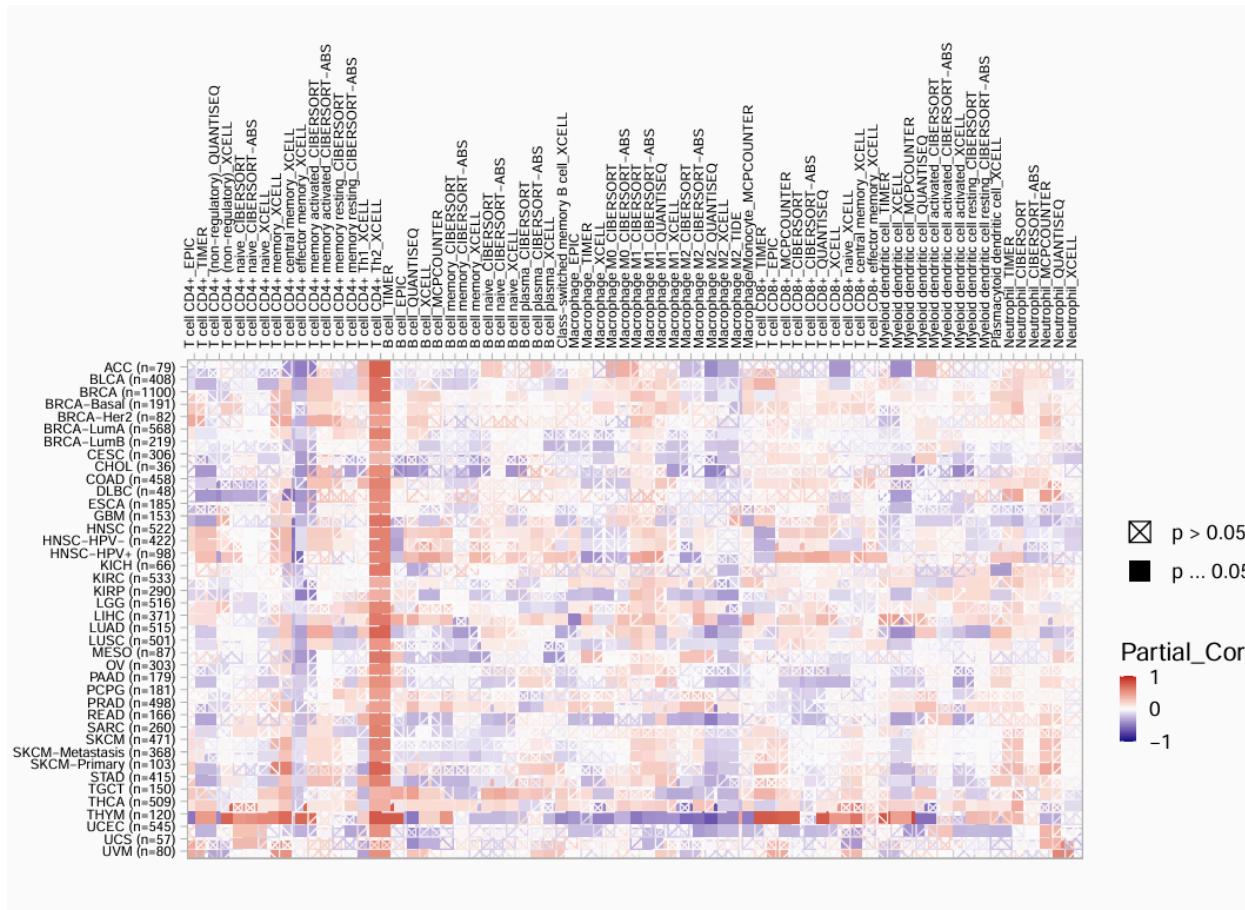
خروجی برای هر کدام از هاب ها به صورت زیر می باشد.



CCNA2



NUSAP1



PBK

مشابه نتایج مقاله ، سلول های PBK + با CD4T کورلیشن منفی دارند و باقی سلول ها کورلیشن مثبت دارند.