



# مقدمه‌ای بر بیوانفورماتیک

پروژه

دانشکده مهندسی کامپیوتر

دانشگاه صنعتی شریف

نیم سال اول ۹۹-۰۰

اساتید:

جناب آقای دکتر شریفی و سرکار خانم دکتر کوهی

نام و نام خانوادگی:

امیرمهدی نامجو



## ۱ مقدمه

در این پروژه، قصد داریم به تحلیل داده‌های ریزآرایه برای لوکمی حاد مغزاستخوان<sup>۱</sup> بپردازیم. این بیماری نوعی سرطان خون و مغزا استخوان است که در آن سلول‌های Myeloid مغزا استخوان، به تعداد بیشتر می‌تکثیر می‌شوند. این موضوع منجر به تولید بیش از اندازه پیش‌سازهای گلبول‌های سفید<sup>۲</sup> می‌شود اما این پیش‌سازها توانایی مقابله با عفونت‌ها را ندارند و در نتیجه بدن نسبت به عفونت ضعیف می‌شود. در این بیماری شاهد نوعی از کم‌خونی و کاهش تعداد گلبول‌های قرمز اصلی بدن می‌شود و در عوض سلول‌های خونی غیرعادی تولید می‌شود که منجر به نارسایی‌های مختلف در فرد بیمار می‌شود.

در این پروژه به کمک داده‌های ریزآرایه قصد داریم ارتباط بین تغییرات بیان‌ژن در افراد مبتلا به این بیماری را مشخص کنیم و سپس براساسی ژن‌هایی که تغییر بیان را نشان می‌دهند، در مورد های مرتبط با این Gene ها و عملکرد آنان، اطلاعاتی را ارائه کنیم.

## ۲ آماده سازی اولیه

در ابتدای کد تمامی کتابخانه‌های مورد نیاز را اضافه می‌کنیم:

```
1 library(GEOquery)
2 library(limma)
3 library(Bioconductor)
4 library(pheatmap)
5 library(reshape2)
6 library(dplyr)
7 library(ggplot2)
8 library(stringr)
```

بعد از این، باید داده‌ها را به کمک کتابخانه GEOquery بخوانیم. با توجه به لینک داده شده در مستند تعریف پروژه، متوجه می‌شویم که نام دیتاست مورد استفاده GSE48558 و بر پایه پلتفرم GPL6244 است. این دو عبارت را به صورت یک متغیر ذخیره می‌کنیم تا اگر بعداً خواستیم تحلیل را روی دیتاست مشابهی انجام دهیم، این قسمت کد بدون تغییر قابل استفاده باشد.

```
1 gset <- getGEO(seriesName, GSEMatrix =TRUE, AnnotGPL=
2 TRUE , destdir = "Data")
3 if (length(gset) > 1){
4   idx <- grep(platformName, attr(gset, "names"))
5 } else {
6   idx <- 1
7 }
8 gset <- gset[[idx]]
```

<sup>1</sup>Acute myeloid leukemia (AML)

<sup>2</sup>Precursor Blood Cells



عبارت if گذاشته شده برای این است که اگر دیتاست روی چند پلتفرم مختلف موجود باشد، دیتای یکی از آن‌ها انتخاب کنیم. در اینجا البته این طور نیست و در اصل همان حالت else اتفاق می‌افتد.

حال نوبت به انتخاب و اسم گذاری روی گروه‌ها می‌رسد. برای این کار مواردی که مربوط به بیماری بودند، همگی Test نام‌گذاری شده‌اند. برای مواردی که Normal بودند، نوع Name Source هم به انتهای نام گروه آن‌ها اضافه شده است چون در ادامه کار به آن نیاز داریم. ضمن این که برای CD34، کلمه اضافی HSPC هم وجود داشت که از نام گروه خارج شده است.

```
1 gset<- gset[,which(gset$source_name_ch1 == "AML Patient"  
| gset$`phenotype:ch1` == "Normal")]  
2 func <- function(x) {  
3   if (gset$source_name_ch1[x] == "AML Patient") {  
4     return("Test")  
5   } else {  
6     sp1l <- strsplit2(gset$source_name_ch1[x] , "\\\\+")[1, 1]  
7     return(paste0("Normal_" , sp1l))  
8   }  
9 }  
10  
11  
12 gr <- sapply(1:length(gset$`phenotype:ch1`), func)  
13
```

### ۳ تحلیل داده‌ها و توضیح کدهای مربوطه

ابتدا ماتریس بیان ژن را تشکیل می‌دهیم و ماکسیمم و مینیمم آن را چک می‌کنیم.

```
1 expr <- exprs(gset)  
2 print(paste0("Max Expr: " , max(expr)))  
3 print(paste0("Min Expr: " , min(expr)))  
4
```

خروجی این دو پرینت قسمت عبارات زیر است:

```
"Max Expr: 13.76153622"  
"Min Expr: 1.611473179"
```

این نشان دهنده این است که داده‌ها در مقیاس لگاریتمی هستند و نیازی به تغییر مقیاس نداریم.

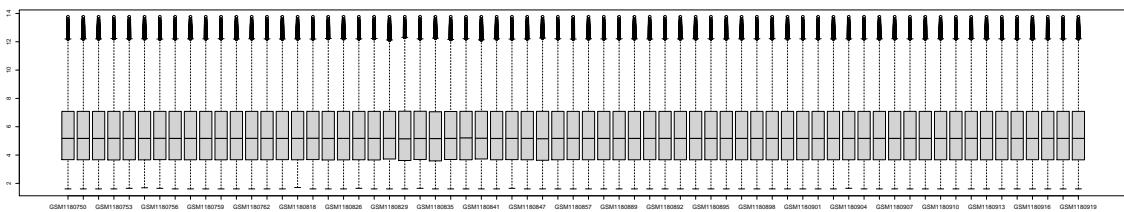
کد تبدیل کننده به مقیاس لگاریتمی هم در فایل اصلی به صورت کامنت شده موجود است.



## ۱.۳ بررسی نرمال بودن داده‌ها

```
1 boxplot(expr)  
2
```

خروجی این boxplot به صورت زیر است (تصویر قابل زوم کردن):



شکل ۱: نمودار Boxplot داده‌های ماتریس بیان ژن

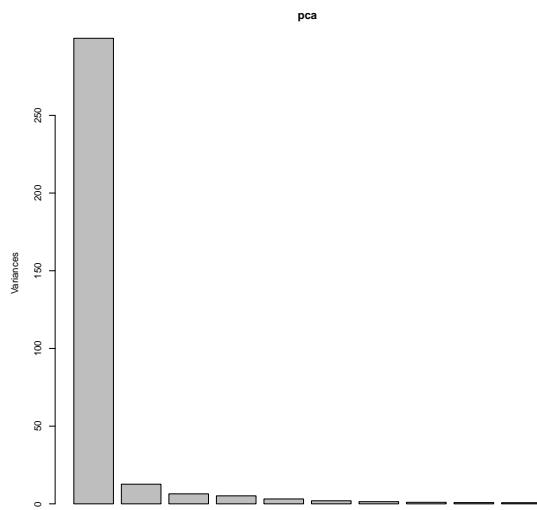
با توجه شکل ۱، میانه داده‌ها و همچنین چارک‌های اول و سوم آن‌ها بسیار نزدیک به یکدیگر است. در نتیجه نیازی به نرمالایز کردن داده‌ها نداریم. با این وجود به صورت کامنت شده کد نرمالایز کردن قرار داده شده است.

## ۲.۳ کاهش ابعاد و PCA

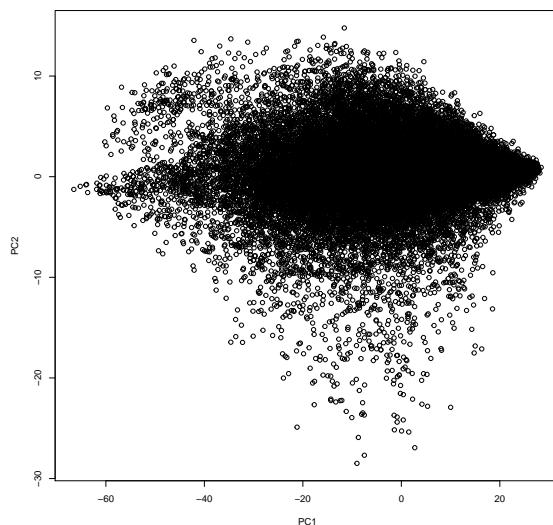
برای کاهش ابعاد از روش تحلیل مولفه‌های اصلی<sup>۳</sup> استفاده شده است. ابتدا به شکل ساده از آن استفاده می‌کنیم و سپس با بررسی مشکلی که دارد، با تغییراتی آن را بهتر می‌کنیم.

```
1 pca <- prcomp(expr)  
2 plot(pca)  
3 plot(pca$x[, 1:2])  
4
```

نتیجه این دو بخش در شکل‌های صفحه بعد دیده می‌شود. همان طور که در شکل ۲ دیده می‌شود بخش عمده تغییرات تنها در یک مولفه به تصویر کشیده شده است. در شکل ۳ هم مشخص است که عمده تغییرات فقط در راستای محور افقی نمود پیدا کرده‌اند و این موضوع نکته خوبی نیست. در اصل چیزی که الان می‌بینیم، این است که یکسری ژن‌ها بیان خیلی بالایی داشته‌اند و یکسری ژن‌ها هم اصلاً بیان نشده‌اند. این کمکی به ما در تمایز میان ژن‌های مربوط به AML نمی‌کند. زیرا علاوه‌نحوی که بیان خیلی بالایی داشتند، عموماً ژن‌های Housekeeping دارند که برای فعالیت‌های حیاتی سلول لازم هستند و علاوه در هر سلولی بیان بالایی دارند و ژن‌های با بیان کم هم ژن‌هایی هستند که در تمامی سلول‌های این نوع بیان کمی دارند. از این رو، باید از طریق Scale کردن و صفر کردن میانگین بیان هر ژن در تمامی نمونه‌ها، کاری کنیم که اختلاف میان AML و سایرین آشکار شود.



شکل ۲: نمودار ستونی میزان تغییرات بیان شده توسط هر یک از مولفه‌های PCA



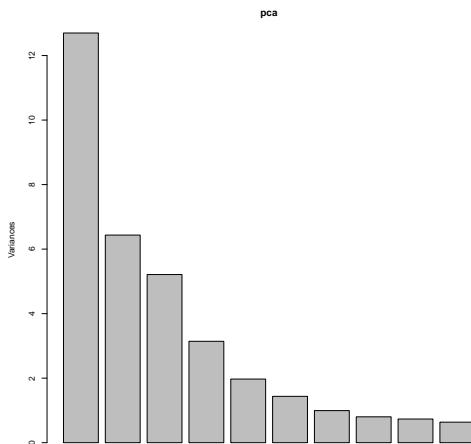
شکل ۳: نمودار نقطه‌ای (scatterplot) داده‌ها براساس دو مولفه با بیشترین اهمیت

برای Scale کردن از کد زیر استفاده می‌کنیم:

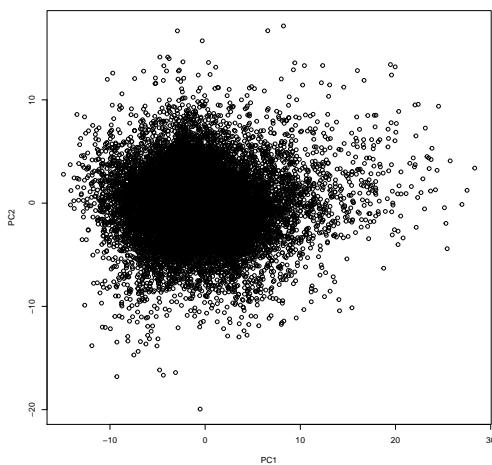
```
1 expr.scaled <- t(scale(t(expr) , scale = FALSE))
2 pca <- prcomp(expr.scaled)
3 plot(pca)
4 plot(pca$x[,1:2])
5
```



نتیجه کار به صورت دو شکل ۴ و ۵ می‌شود که از نظر توزیع اهمیت بین مولفه‌های مختلف، بسیار بهتر از قبل است:



شکل ۴: نمودار ستونی میزان تغییرات بیان شده توسط هر یک از مولفه‌های PCA پس از Scale شدن



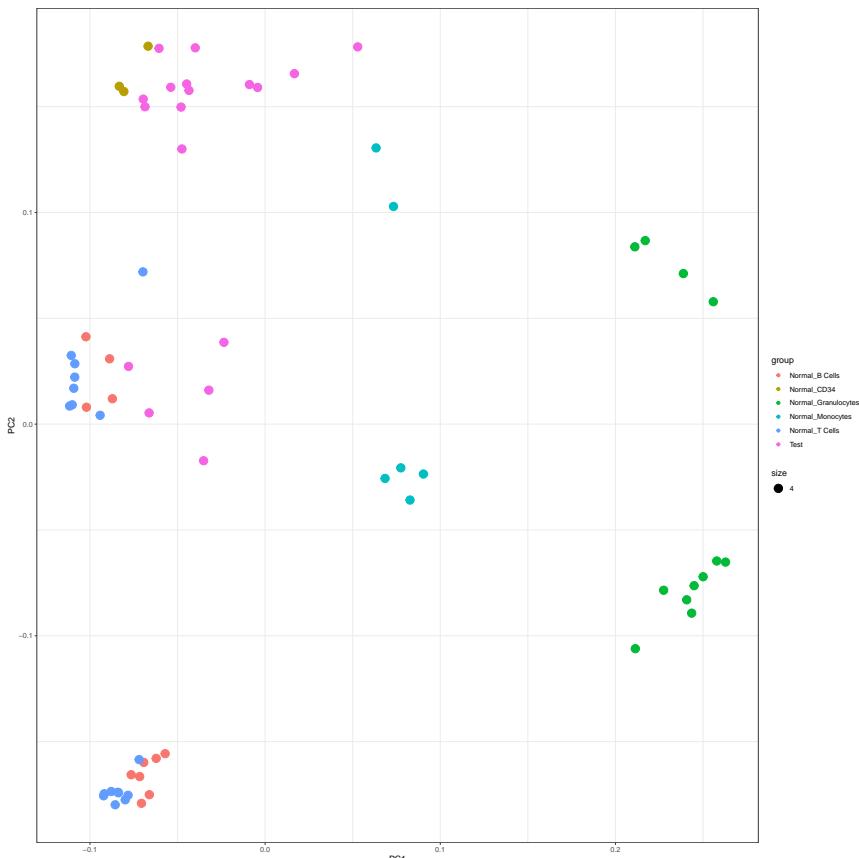
شکل ۵: نمودار نقطه‌ای (scatterplot) داده‌ها براساس دو مولفه با بیشترین اهمیت پس از Scale شدن



حال با کمک مولفه rotation در PCA، نمودار مربوط به نمونه‌های مختلف را با رنگ‌های مختلف رسم می‌کنیم.

```
1 pcar <- data.frame(pca$rotation[,1:3] , group = gr)
2 ggplot(pcar , aes(PC1 , PC2 , color = group , size = 4))
+ geom_point() + theme_bw()
3
```

شکل نمودار در زیر آورده شده است:



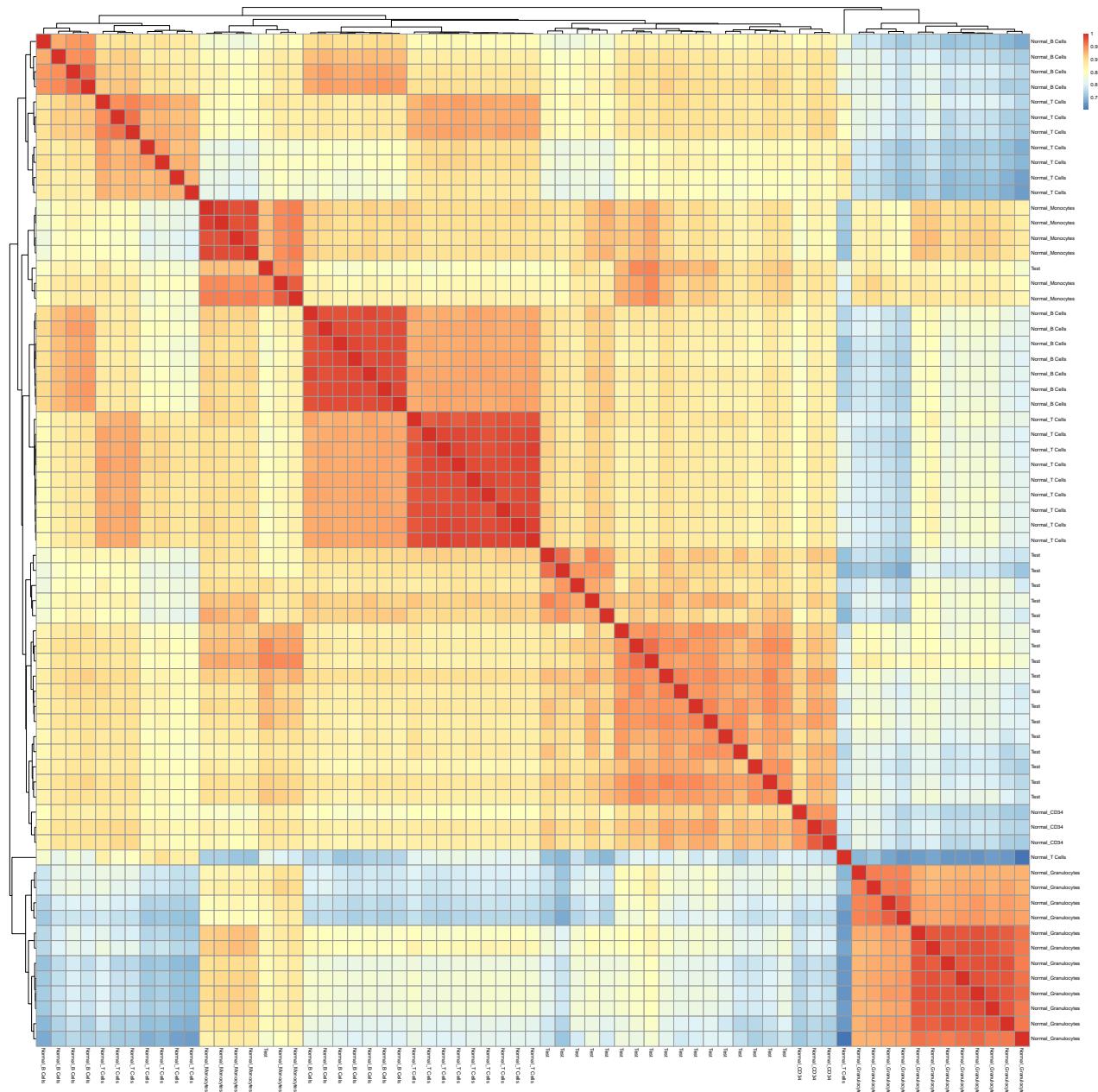
شکل ۶: نمودار Scatterplot نمونه‌ها براساس دو مولفه اصلی در PCA و رنگ شده براساس نوع مولفه

### ۳.۳ نمودار همبستگی Heatmap

برای رسم نمودار همبستگی و Heatmap از کد زیر استفاده می‌کنیم:

```
1 pheatmap(cor(expr) , labels_row = gr , labels_col = gr
)
2
```

نتیجه بدین صورت خواهد بود:



شکل ۷: نمودار Heatmap همبستگی بین نمونه‌های مختلف

### ۴.۳ پیدا کردن تفاوت‌های بیان ژن در سلول بیماران و سلول‌های سالم

با توجه به شکل‌های ۶ و ۹ می‌بینیم که سلول‌های تست به سلول‌های CD34 نزدیک‌تر هستند. برای همین ابتدا آن دسته از سلول‌های تست که به CD34 نزدیک‌تر هستند را با نماد دیگری در گروه‌ها مشخص می‌کنیم. با توجه به شکل ۶ می‌بینیم که برای این دسته مولفه PC2 بیش از ۰.۱۱ و آن‌ها کمتر از -۰.۰۲۵ است. آن‌ها را با نماد "AML\_Near\_CD34" مشخص می‌کنیم. کد این کار در



زیر آورده شده است:

```
1 gr2 <- gr
2 gr2[which(pcar$PC2 > 0.11 & pcar$group == "Test")] <-
3 "AML_Near_CD34"
4 gr2 <- factor(gr2)
5 gset$description <- gr2
```

دلیل استفاده از تابع factor این است که عبارات درون gr2 صرفا به عنوان یکسری رشته معمولی نباشند و به عنوان سطوح مختلف از یک متغیر مشخص در نظر گرفته شوند.  
پس از آن باید یک ماتریس بسازیم که برای هر نمونه، مشخص کند که در کدام دسته قرار دارد.  
این ماتریس را برای این که بتوانیم پس از آن از قابلیت fit linear model کتابخانه Limma استفاده کنیم، باید به صورت One-Hot بسازیم. یعنی برای هر نمونه، ستون تمامی دسته‌ها وجود داشته باشد و دسته مربوط مقدار 1 و مابقی مقدار 0 داشته باشند. ضمناً در نام یکسری از دسته‌ها اسپیس وجود دارد که برای این که در پکیج‌های بعدی به مشکل تخریب، این اسپیس با \_ جایگزین شده است. در کد پایین توجه کنید که تابع str\_replace به کتابخانه stringr مربوط به کتابخانه است.

```
1 onehot <- model.matrix(~ description + 0, gset)
2 colnames(onehot) <- levels(gr2)
3
4 colnames(onehot) <-
5 c(sapply(colnames(onehot), function(x)
6 str_replace(x, " ", "_")))
```

بعد از این ابتدا یک Fit می‌کنیم. سپس تفاوت بین بیان‌زن‌های دسته سلول‌های بیمار و دسته سلول‌های CD34 را بدست می‌آوریم. پس از آن از طریق یک مدل بیزین، مدل جدیدی را Fit می‌کنیم. در این مدل بیزین، دانش پیشین ما بر این فرض است که حدود 1 درصد ژن‌ها باید تفاوت بیان نشان بدهند. این مدل براساس این دانش Posterior، توزیع نهایی را بدست اورده و در نهایت جدول نهایی که شامل مواردی نظیر Adjusted P Value و همچنین log Fold Change برای ژن‌های مختلف است را بدست می‌آورد. در این جدول اطلاعات دیگری هم موجود است ولی ما تنها نام و نشانه ژن‌ها و همچنین adj.P.Val و logFC را برای مرحله بعد نیاز داریم.

```
1 fit <- lmFit(gset, onehot)
2
3 contrast <-
4 makeContrasts(AML_Near_CD34 - Normal_CD34, levels =
onehot)
5
6 fit2 <- contrasts.fit(fit, contrast)
7 fit2 <- eBayes(fit2, 0.01)
8 tT <- topTable(fit2,
9 adjust = "fdr",
10 sort.by = "B", number = Inf)
```



```
11 tT2 <- subset(tT , select = c("Gene.symbol" , "Gene.
12 title" , "adj.P.Val" , "logFC"))
```

با توجه به این که logFC بزرگتر از 1 نشان دهنده تفاوت بیش از دوبرابری در بیان بیشتریک ژن و کمتر از 1- نشان دهنده تفاوت بیش از دوبرابری در بیان کمتریک ژن است، از این طریق و با توجه به adj.P.Val کمتر از 0.05 ژن هایی که تفاوت بیان معنی دار داشتند را بدست می آوریم. از آنجایی که بعضی از ژن ها چند نماد دارند که با /// جدا شده اند، از همه این نمادها در جواب استفاده می کنیم.

```
1 aml.up <- subset(tT2, logFC > 1 & adj.P.Val < 0.05)
2 aml.up.genes <- unique(as.character(
3 strsplit2( (aml.up$Gene.symbol), "/\\/")))
4 aml.down <- subset(tT2, logFC < -1& adj.P.Val < 0.05)
5 aml.down.genes <- unique(as.character(
6 strsplit2( (aml.down$Gene.symbol), "/\\/")))
7
```

## ۴ تحلیل کمک Pathways و Gene Ontology

### ۱.۴ تحلیل ژن ها با بیان بالاتر در سلول های بیمار

#### ۱.۱.۴ تحلیل Pathway

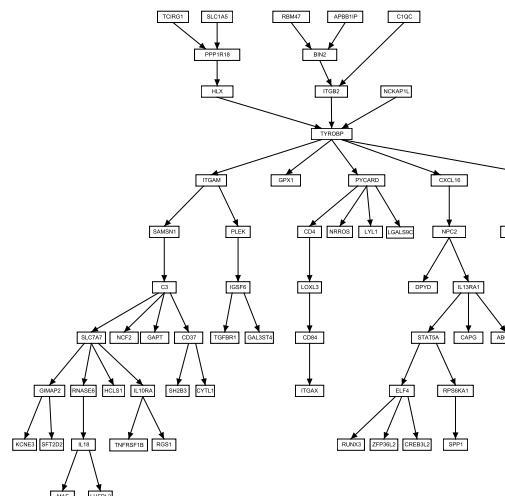
برای تحلیل از Enrichr استفاده می کنیم [۲، ۱]. با تحلیل ژن هایی که در بیماران بیان بالاتری داشته اند، در میان Pathway ها براساس مجموعه WikiPathways 2019 Human موارد زیر قابل توجه بودند:

- TYROBP Causal Network :  $APV = 2.264 \times 10^{-8}$
- Microglia Pathogen Phagocytosis Pathway:  $APV = 1.143 \times 10^{-9}$
- Platelet-mediated interactions with vascular and circulating cells :  $APV = 0.0002054$

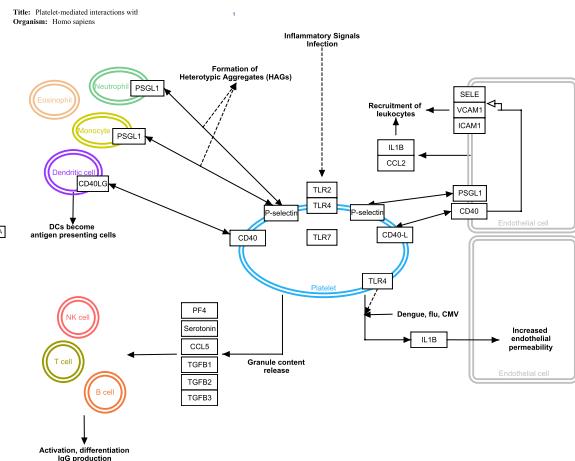
البته موارد دیگری هم هستند ولی این موارد، از نظر موضوعی هم در مقالات مرتبط با AML به خوبی بررسی شده اند. (جدول کامل در کنار این فایل قرار دارد) در صفحه بعد تصویر این Pathway ها قرار گرفته است.



Title: TYROBP Causal Network  
Organism: Homo sapiens

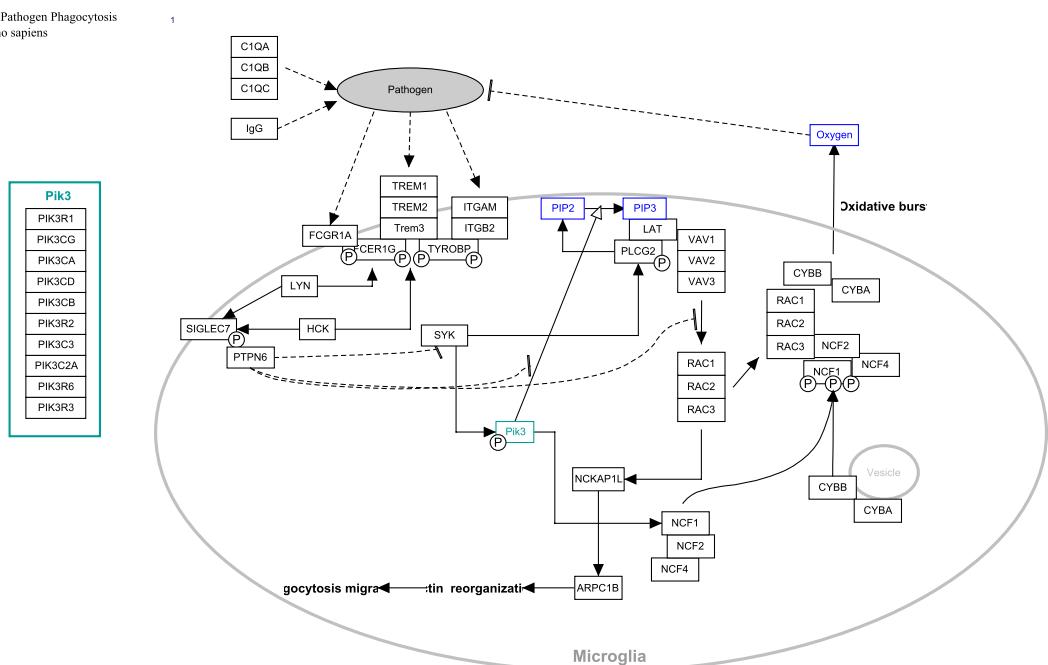


(ب)



(ا)

Title: Microglia Pathogen Phagocytosis  
Organism: Homo sapiens



(ج)

شکل ۸: قسمت (آ) مربوط به مربوط به  
Microglia Pathogen Phagocytosis Pathway  
قسمت (ب) مربوط به TYROBP Causal Network  
و قسمت (ج) مربوط به است.



## ۲.۱.۴ تحلیل Gene Ontology

در زمینه Molecular Function، موارد زیر قابل توجه هستند:

- Toll-like receptor binding (GO:0035325) :  $APV = 0.003139$
- kinase binding (GO:0019900):  $APV = 1.846 \times 10^{-7}$
- aryl sulfotransferase activity (GO:0004062) :  $APV = 0.005231$

در زمینه Biological Process موارد زیر قابل توجه هستند:

- neutrophil activation involved in immune response (GO:0002283) :  $APV = 4.830 \times 10^{-37}$
- neutrophil degranulation (GO:0043312):  $APV = 6.631 \times 10^{-37}$
- neutrophil mediated immunity (GO:0002446) :  $APV = 2.158 \times 10^{-37}$

در زمینه Cellular Component موارد زیر قابل توجه هستند:

- specific granule membrane (GO:0035579) :  $APV = 7.816 \times 10^{-14}$
- specific granule (GO:0042581):  $APV = 2.843 \times 10^{-14}$
- tertiary granule (GO:0070820) :  $APV = 3.434 \times 10^{-14}$

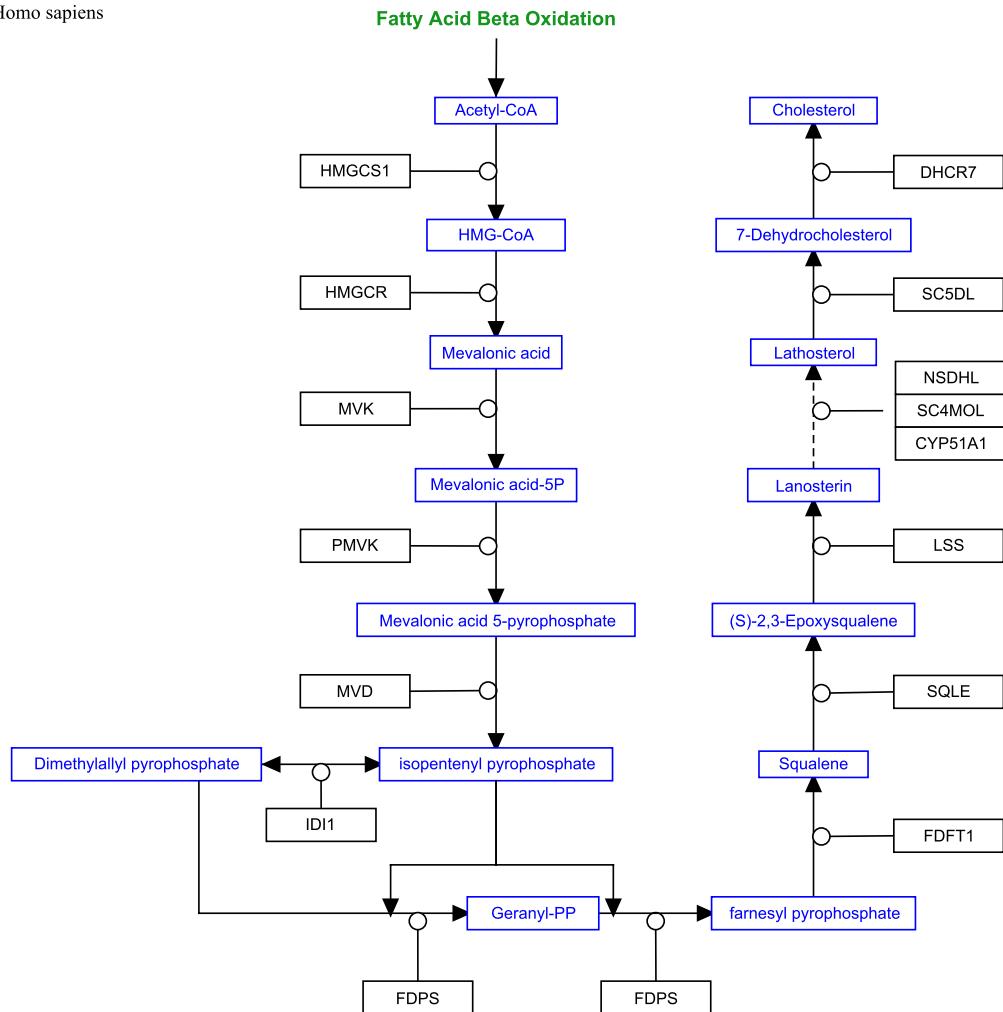
## ۲.۴ تحلیل ژن‌ها با بیان پایین‌تر در سلول‌های بیمار

### ۱.۲.۴ Pathway تحلیل

در زمینه Pathway های این قسمت، بعضی P-Value بالایی می‌دهند که چندان مناسب نیست. اما مثلا در بخش ARCHS4 Kinase، شاهد این هستیم که PRKG2 human kinase ARCHS4 coex-pression با  $APV = 1.269 \times 10^{-7}$  نشان از تفاوت معنی دار در این زمینه دارد. همچنین در بخش Cholesterol biosynthesis، یک Pathway با عنوان Bioplanet2019 دارد، اما با جست و جو به مقاله‌ای مرتبط با این AML Pathway و می‌رسیم که به نظر نشان دهنده وجود ارتباط معنادار هستند. تصویر Cholesterol biosynthesis در صفحه بعد قرار دارد. [۳]



Title: Cholesterol Biosynthesis Pathway  
Last modified: 2/22/2013  
Organism: Homo sapiens



شکل ۹: Cholesterol Biosynthesis Pathway

#### ۲.۲.۴ تحلیل Gene Ontology

در مورد Gene Ontology ، اصلی‌ترین موردی که با APV پایین دیده می‌شود مربوط به دو Cellular Component با نام‌های (GO:0032389) MutLalpha complex و (GO:0071556) component of luminal side of endoplasmic reticulum membrane با APV = 0.005508 و 0.03364 است و در هر دو مورد مقاله‌هایی مرتبط با موضوع یافت می‌شود. [۵، ۴]



## ۵ توضیحات مختلف در مورد AML

### ۱.۵ آشنایی با AML

#### ۱.۱.۵ AML چیست؟

برای بررسی AML ابتدا باید ببینیم که لوکمی چیست.  
لوکمی<sup>۴</sup> به سرطان‌هایی می‌گویند که در سلول‌هایی ایجاد می‌شود که در حالت معمول، قرار است به سلول‌های مختلف خونی تبدیل شوند. در رابطه تربین شکل، لوکمی در حالات اولیه سلول‌های سفید خون ایجاد می‌شود اما در بعضی موارد، در سلول‌های دیگری هم این مسئله ایجاد می‌شود. لوکمی بسته به سرعت رشد به دو دسته کلی حاد<sup>۵</sup> یا مزمن<sup>۶</sup> تقسیم می‌شوند. همچنین بعضی لوکمی‌ها در سلول‌های مربوط به مغزاستخوان<sup>۷</sup> شروع شده و برخی دیگر در سلول‌های لنفاوی<sup>۸</sup>.  
لوکمی حاد مغزاستخوان، در بافت مغز استخوان - که قسمت نرم درونی برخی استخوان‌هاست و سلول‌های خونی در آن جا ساخته می‌شود - آغاز می‌شود. این نوع لوکمی در اغلب اوقات به سرعت به خون هم منتقل می‌شود. در حالت‌های حادتر، این لوکمی می‌تواند به بخش‌های دیگر بدن نظیر گره‌های لنفاوی، کبد، سیستم عصبی مرکزی (مغز و نخاع) و طحال هم متاستاز<sup>۹</sup> بدهد و به آن‌ها گسترش یابد.

#### ۲.۱.۵ مغز استخوان

برای مطالعه بهتر این موضوع، باید بررسی کنیم که مغزاستخوان دقیقاً چیست و چه وظیفه‌ای دارد.  
مغز استخوان قسمت نرم درونی بعضی از استخوان‌های بدن است که در سلول‌های تولید کننده خون، سلول‌های چربی و بافت همبند تشکیل شده است. بخش اندکی از سلول‌های تولید کننده خون، سلول‌های بنیادی خونی<sup>۱۰</sup> نامیده می‌شوند.

درون مغز استخوان، این سلول‌های بنیادی خونی، به سلول‌های خونی مختلف تبدیل می‌شوند. در طی این فرایند سلول‌های بنیادی خونی، یا تبدیل به لفونوسیت‌ها - که نوعی سلول سفید خونی است - تبدیل شده و یا به نوع دیگری از سلول‌های تولید کننده خون به نام میلوبئید تبدیل می‌شوند. سلول‌های میلوبئید می‌توانند تبدیل به گلبول‌های قرمز، گلبول‌های سفید (متفاوت با لفونوسیت) و پلاکت بشونند. سلول‌هایی که در بیماری AML رفتار غیرنرمال دارند، همین سلول‌های میلوبئید هستند.

#### ۳.۱.۵ انواع سلول‌های خونی

برای اطلاع بیشتر لازم است کمی در مورد سلول‌های خونی هم آگاهی پیدا کنیم.  
به طور کلی سه نوع سلول اصلی خونی وجود دارد:

- سلول‌های خونی قرمز: این سلول‌های وظیفه دارند که اکسیژن را از شش‌ها به همه بافت‌های مختلف بدن انتقال داده و در کربن دی‌اکسید تولیدی آن‌ها را به شش‌ها بازگردانند.

<sup>4</sup>Leukemia

<sup>5</sup>Acute

<sup>6</sup>Chronic

<sup>7</sup>myeloid

<sup>8</sup>lymphoid

<sup>9</sup>metastasis

<sup>10</sup>Blood Stem Cells



• پلاکتها: پلاکتها در اصل قطعات سلولی هستند که توسط نوع خاصی از سلول‌های مغز استخوان موسوم به مگاکاریوسیت<sup>۱۱</sup> ساخته می‌شوند. یکی از اصلی‌ترین وظایف پلاکتها جلوگیری از خونریزی است. آن‌ها کمک می‌کنند سوراخ‌هایی که روی رگ‌ها در اثر ضربه یا خراش ایجاد شده است، پوشیده شوند.

• سلول‌های سفید خونی: وظایف این سلول‌ها مبارزه با عفونت‌ها در بدن است.

سلول‌های سفید، خود دسته‌های مختلفی دارند. از جمله مهم‌ترین آن‌ها سه دسته زیر است:

• Granulocytes: گرانولوسیت‌ها سلول‌های سفید بالغی هستند که از میلوبلاست‌ها<sup>۱۲</sup> که خود یک نوع سلول تولید کننده خون در مغز استخوان است، ایجاد می‌شود. گرانولوسیت‌ها به صورت سلول‌هایی که یکسری نقطه روی خود دارند زیر میکروسکوپ دیده می‌شوند. این نقاط شامل آنزیم‌ها و موادی هستند که می‌توانند میکروب‌ها را نابود کنند. سه نوع گرانولوسیت به نام‌های Eosinophil، Neutrophil و Basophil وجود دارند.

• Monocytes: مونوسیت‌ها از مونوبلاست‌ها<sup>۱۳</sup> در مغز استخوان ایجاد می‌شوند. این سلول‌ها بعد از چرخیدن در سراسر بدن در طول حدود یک روز، وارد بافت‌ها شده و تبدیل به ماکروفاز<sup>۱۴</sup> می‌شود. ماکروفازها با احاطه کردن میکروب‌ها و هضم کردن آنان به اینمنی بدن کمک کنند. این سلول‌ها به لنفوسيت‌ها در شناسایی میکروب‌ها و ترشح آنتی بادی مناسب هم کمک می‌کنند.

• Lymphocytes: لنفوسيت‌ها، از لنفوبلاست<sup>۱۵</sup> در مغز استخوان ایجاد می‌شوند. این سلول‌ها جزء اصلی سازنده بافت لنفاوی هستند که بخش بزرگی از سیستم ایمنی بدن است. بافت لنفاوی در گره‌های لنفاوی، تیموس، طحال و بخش‌های مختلفی از بدن یافت می‌شود. دو نوع اصلی لنفوسيت‌ها سلول‌های T و B نامیده می‌شوند.

#### ۴.۱.۵ ریسک فاکتورها

عوامل مختلفی وجود دارند که می‌توانند باعث افزایش احتمال ابتلا به AML بشوند. در زیر به طور خلاصه بعضی از این عوامل را ذکر کرده‌ایم:

- سن - احتمال ابتلا با افزایش سن رابطه مستقیم دارد
- جنسیت - AML در مردان شایع‌تر از زنان است.
- سیگار کشیدن - همانند بسیاری از بیماری‌های دیگر، سیگار کشیدن احتمال ابتلا به AML را هم افزایش می‌دهد. دلیل این موضوع این است که مواد سرطان‌زاوی در تنباکو وجود دارد که در شش جذب شده و به سراسر بدن منتقل می‌شوند و می‌توانند باعث ایجاد سرطان‌های مختلف بشوند.

<sup>11</sup>megakaryocyte

<sup>12</sup>Myeloblasts

<sup>13</sup>monoblasts

<sup>14</sup>Macrophage

<sup>15</sup>Lymphoblast



- قرار گرفتن در معرض بعضی مواد شیمیایی - به عنوان مثال برخورد طولانی مدت با بنزن و فرمالدھید (مثلا در محل کار) ریسک ابتلا به AML را افزایش می‌دهند.
- قرار گرفتن در معرض مواد رادیواکتیو - همانند اکثر سرطان‌ها، قرار گرفتن در معرض پرتوهای ناشی از مواد رادیواکتیو احتمال ابتدا به AML را افزایش می‌دهد.
- وجود یکسری ناهنجاری‌های خونی - بعضی از ناهنجاری‌ای خونی نظیر myelodysplastic syndrome احتمال ابتلا به AML را بالا می‌برند.
- وجود مشکلات ژنتیکی - بعضی جهش‌های ژنتیکی نظیر Bloom Fancone Anemia یا سندروم یا مشکلات ژنتیکی خیلی شدیدتر نظیر مشکلات کروموزومی نظیر سندروم داون امکان ابتلا به AML را افزایش می‌دهد.
- سابقه خانوادگی - افرادی که در خانواده خود افرادی با AML را داشته‌اند، احتمال ابتلا به این بیماری در آن‌ها بیش‌تر است.

### 5.1.5 علائم AML

برای شناسایی زودهنگام این بیماری باید علائم آن را شناخت. علائم این بیماری را می‌توان به چند دسته تقسیم کرد.

#### علائم ناشی از کمبود سلول‌های قرمز خونی:

این علائم به دلیل کمبود سلول‌های قرمز خون پیش می‌آیند. کمبود سلول‌های قرمز (کم خونی) می‌تواند به دلایل دیگری هم پیش بیاید ولی یکی از علل آن ممکن است AML باشد. این علائم شامل موارد زیر است:

- خستگی
- ضعف
- احساس سرما
- سرگیجه
- سردرد
- رنگ پریدگی
- تنگی نفس

#### علائم ناشی از کمبود سلول‌های سفید نرمال خونی:

سلول‌های سفید خونی مسئول مبارزه با عفونت‌ها و بیماری‌ها هستند. فرد مبتلا به AML ممکن است به طور مداوم دچار بیماری‌های عفونی بشود که به راحتی برطرف نمی‌شوند و یا با فاصله کم بعد از بهبودی از یک بیماری، درگیری بعدی بشود. با وجود این که در بیماران AML ممکن است تعداد سلول‌های سفید خون زیاد باشد، اما این افزایش تعداد به دلیل سلول‌های لوکمی هستند که از حالت رفتاری نرمال خارج شده و در مقابل بیماری‌های عفونی، از بدن محافظت نمی‌کنند.

#### علائم ناشی از کمبود پلاکت

این علائم شامل موارد زیر هستند:



- کبودی‌های مختلف روی پوست
- خونریزی طولانی که بند نیاید
- خونریزی‌های شدید و متداول بینی
- خونریزی لثه

#### علائم ناشی از افزایش سلول‌های لوکمی

سلول‌های سرطانی AML معمولاً بزرگ‌تر از سلول‌های معمول سفید خونی هستند و به آسانی نمی‌توانند در رگ‌ها جا به جا شوند. با این وجود اگر تعداد آن‌ها خیلی زیاد بشود، امکان ایجاد لخته‌های خونی درون رگ‌ها وجود دارد و در نتیجه آن سلول‌های قرمز هم امکان جا به جایی نخواهند داشت. به این پدیده Leukostasis گفته می‌شود. این پدیده نادر است ولی در صورت وقوع، باید به طور اورژانسی به آن رسیدگی شود. از علائم آن موارد زیر هستند:

- سردرد
- ضعف در یک سمت بدن
- تکلم گنگ و نامفهوم
- گیجی و پریشانی
- خواب آلودگی

#### علائم عمومی:

علائم زیر علائم عمومی هستند که هر چند در AML هم مشاهده می‌شوند اما در بسیاری از بیماری‌های دیگر هم وجود دارند و وجود آن‌ها به تنها‌بی احتمالاً علت دیگری دارد.

- کاهش وزن
- خستگی
- تب
- عرق شبانه
- کاهش اشتها

#### علائم مربوط به پخش شدن بیماری به سایر اندام‌ها:

در صورت انتشار لوکمی به سایر اندام‌های بدن و به خصوص انتقال به مغز و نخاع، علائم زیر ممکن است رخدادهند:

- سردرد
- ضعف



- حمله صرع
- استفراغ
- مشکل در تعادل
- بی حسی صورت
- تاری دید

#### ۶.۱.۵ علت ابتلا به AML

پیدا کردن دلیل اصلی ابتدا به این بیماری در اکثر افراد کار دشواری است. با این وجود به طور کلی می‌توان در مورد علل ابتلا به این نوع سرطان توضیحاتی ارائه داد.

ژن‌های خاصی در سلول‌ها وجود دارند که مسئول رشد و تکثیر و مرگ سلول‌ها هستند. ژن‌هایی که مسئول رشد و تکثیر و زندگاندن سلول هستند Oncogene و سلول‌هایی که مسئول مرگ سلول در زمان مناسب هستند، ژن‌های سرکوب‌گر تومور<sup>۱۶</sup> نامیده می‌شوند.

از آن جایی که فرآیند تقسیم سلولی، فرآیند کاملاً دقیقی نیست و امکان خطأ در آن وجود دارد، گاهی به دلیل جهش‌هایی خاص، Oncogene‌ها بیش از حد انتظار فعال شده و یا ژن‌های سرکوب‌گر تومور، غیرفعال شوند. در چنین شرایطی سلول به طور بی‌رویه به رشد و تکثیر ادامه داده و در تومور و در موارد شدیدتر سرطان ایجاد می‌شود. به طور مثلاً در AML، دیده شده که ژن‌هایی نظری FLT3 یا RAS در این زمینه نقش مهمی دارند.

تغییرات مختلفی ممکن است در کروموزوم‌های یک سلول AML اتفاق افتد. بعضی از آن‌ها به شرح زیر است:

- جا به جایی کروموزمی<sup>۱۷</sup>: این موضوع، شایع‌ترین مورد در تغییرات کروموزومی است. در این حالت بخشی از یک کروموزوم شکسته شده و به کروموزوم دیگر متصل می‌شود. در صورتی که این تغییر منجر به فعال شدن Oncogene‌ها و یا غیرفعال شدن ژن‌های سرکوب‌کننده تومور نظری RUNX1 بشود، امکان ایجاد تومور وجود دارد.
- حذف کروموزومی<sup>۱۸</sup>: در این حالت قسمتی از کروموزوم حذف می‌شود. این حذف ممکن است منجر به از دست رفتن ژن‌های سرکوب‌گر تومور بشود و در نتیجه رشد آن سلول از کنترل خارج بشود.
- معکوس شدن<sup>۱۹</sup>: در این حالت، کروموزم معکوس شده و ترتیب آن بر عکس می‌شود. در این حالت تعداد زیادی از ژن‌ها ممکن است از دست بروند. عملاً این حالت مانند خواندن یک کتاب بر عکس است. در نتیجه تعدادی زیادی از ژن‌های مربوط به آن کروموزوم وظیفه خود را به درستی انجام نخواهند داد.

<sup>16</sup>Tumor Suppressor Gene

<sup>17</sup>Translocation

<sup>18</sup>Deletion

<sup>19</sup>Inversion



- تکرار<sup>۲۰</sup>: در این حالت، بخشی از کروموزوم تکرار می‌شود. این موضوع می‌تواند منجر به ایجاد کپی‌های زیادی از یک ژن درون سلول بشود و در نتیجه اگر این ژن‌ها جزو Oncogene ها باشند، امکان تکثیر بی‌رویه سلول وجود دارد.

#### ۷.۱.۵ پیشگیری از AML

با توجه به این که بسیاری از ریسک فاکتورهای AML موارد غیرقابل تغییر هستند، در اکثر موارد امکان جلوگیری از وقوع AML در یک فرد وجود ندارد. با این حال، چندین عامل مهم وجود دارند که کنترل آن‌ها می‌تواند به پیشگیری از این سرطان منجر شود.

اولین مورد که اصلی ترین عامل قابل کنترل است، جلوگیری از سیگار کشیدن است. سیگار و دخانیات رابطه مستقیم با ابتلا به انواع سرطان‌ها دارند و کنترل این رفتارها نیز کاملاً بر عهده خود افراد است. در نتیجه جلوگیری از مصرف دخانیات نقش موثری در کاهش احتمال ابتلا به AML دارد. شیمی درمانی و قرار گرفتن در معرض تابش مواد رادیواکتیو، هر چند برای درمان یک سرطان دیگر استفاده می‌شود اما ممکن است باعث ایجاد نوع دیگر سرطان نظیر AML بشود. در حال حاضر تحقیقات مختلفی در دنیا در حال انجام است تا روش‌هایی برای این کار پیدا شود که باعث افزایش احتمال ابتلا به یک سرطان جانبی نشوند. با این حال، همچنان درمان سرطان‌های مهلك به کمک شیمی درمانی و پرتودرمانی، اهمیت بالایی دارد. زیرا هر چند ممکن است این درمان‌ها در آینده منجر به سرطان دیگری بشوند، اما حداقل باعث می‌شوند که سرطان فعلی فرد تا حد خوبی کنترل بشود و در نهایت منجر به افزایش طول عمر می‌شوند.

در صورتی که فرد با مواد شیمی سرطان‌زا نظری بنزن کار می‌کند، احتمال ابتلای او به سرطان‌های مختلف نظیر AML افزایش می‌یابد. عوض کردن شغل و یا برخورد کاملاً محافظت شده با این مواد می‌تواند به کاهش احتمال ابتلا به AML موثر باشد. مطالب این مقدمه برگرفته از [۶] بودند.

### ۲.۵ بررسی ارتباطات نتایج تحلیل دادگان ریزآرایه و AML

در قسمت Microglia Pathogen Phagocytosis Pathway یکی از مواردی که به آن اشاره کردیم<sup>۲۰</sup> بود. یکی از مقالاتی که به این موضوع پرداخته، [۷] است. Microglia یا میکروگلیا، در اصل سلول‌هایی کوچک و غیرعصبی هستند که ساختار حمایتی دستگاه عصبی مرکزی را تشکیل می‌دهند. این سلول‌ها از جمله اولین سلول‌هایی هستند که نسبت به عفونت‌های مغزی نظیر انسفالوپاتی واکنش نشان می‌دهند. یکی از ژن‌های که خیلی به AML مرتبط است، ژن RUNX1 است که با نام AML1 هم شناخته می‌شود. این ژن در مراحلی از ایجاد میکروژلیا در دوران جنینی تاثیر دارد و باعث تغییرات ساختاری و مورفوЛОژیکی در آن می‌شود. افزایش آن در AML باعث ایجاد اختلال‌هایی در کارهای بدن می‌شود.

موضوع دیگر مربوط به TYROBP Causal Network<sup>۲۱</sup> می‌شود. TYROBP نوعی پروتئین تیروزین کیناز است. این پروتئین در تشکیل سلول‌های ایمنی (ایمونوگلوبین) بدن نقش دارد و همان طور که می‌دانیم، در جریان AML سلول‌های سفید خونی که فعالیت نرمалی ندارند، افزایش می‌یابند. در نتیجه رابطه بین این مورد با AML هم مشخص می‌شود. این مورد در [۸] و [۹] بررسی شده است.

<sup>20</sup>Addition or Duplication

<sup>21</sup>Tyrosine Kinase



از طرف دیگر در بخش Biological Process شاهد حضور مواردی مرتبط با Neutrophil ها هستیم. همان طور که در بخش آنشایی با AML گفته شد، Neutrophil ها یکی از انواع گرانولوسیت‌ها هستند و گرانولوسیت‌ها خود جزو گلبول‌های سفید به حساب می‌آمدن. افزایش بیان ژن‌هایی که منجر به فعال‌سازی این سلول‌ها می‌شود با توجه به ذات AML که منجر به تکثیر بی‌رویه سلول‌های خونی می‌شود، منطقی به نظر می‌رسد. در مقاله‌های [۱۰] و [۱۱] هم به افزایش فعالیت‌های مربوط به Neutrophil ها در AML پرداخته شده است.

در قسمت Molecular Function مشاهده می‌کنیم که وجود دارد Kinase Binding است. کیناز یک آنزیم است که بر روی فسفات‌های پرانرژی نظری ATP اثر گذار است. در این بیماری شاهد این هستیم که ژن‌هایی مربوط به فرآیندهای Kinase بیان بیشتری داشته‌اند. با تحقیق در مقالات، می‌بینیم که این مسئله به شکل خیلی جدی بررسی شده است. یکی از درمان‌هایی که برای AML پیشنهاد می‌شود، مبتنی بر استفاده از بازدارنده‌های تیروزین کیناز<sup>۲۲</sup> و بازدارنده‌های تئوینین/سرین کیناز<sup>۲۳</sup> است. مکانیزم اثر این داروها به این شکل است که در برای پیوند دادن با کیناز، با ATP وارد رفابت شده و سعی می‌کنند در مواردی باعث شوند که ATP نتواند با کیناز وارد پیوند کاتالیزی بشود. در نتیجه بدین شکل جلوی فعالیت بیش از حد کیناز را می‌گیرند. از این داروها در سرطان‌های دیگر هم استفاده می‌شود. مقاله [۱۲] به این داروها پرداخته و مقاله [۱۳] هم به بررسی تاثیر این داروها در AML و پیشرفت‌ها و چالش‌های پیش‌رو در استفاده از آنان پرداخته است. در مورد Gene Ontology های مربوط به کاهش بیان ژن، به MutLAlpha اشاره شد. در اصل MutL ها جزو سیستم اصلاح اشتباہات DNA<sup>۲۴</sup> هستند. این که مشاهده می‌کنیم در هنگام وقوع این سرطان، بیان ژن‌های مربوط به آن کاهش یافته است، امری منطقی است. زیرا عملاً این سرطان و ایجاد سلول‌هایی که رشد آن‌ها به صورت بی‌رویه است، همان طور که در بخش‌های قبل هم گفتیم در اثر ایرادات و اشتباہاتی که در تقسیم سلولی روی می‌دهد اتفاق می‌افتد که ژن‌های نامناسبی فعال می‌شوند. در این مورد سیستم اصلاح اشتباہات دچار مشکل شده و در نتیجه آن این سلول‌های نامعمول لوکمی ایجاد شده‌اند. این موضوع در [۴] بررسی شده است.

## ۶ فایل‌های تحویلی

در پوشه src دو فایل قرار دارد. یک فایل main.r که کد اصلی برنامه است و در کنار آن یک فایل Project.ipynb هم قرار دارد که یک فایل Jupyter است و همان کدهای main.r را دارد ولی در کنار آن، نحوه کارکرد قسمت‌های مختلف کد را هم توضیح داده‌ام.

در پوشه Results نتایج بدست آمده نظری نمودارها و همچنین فایل‌های txt شامل اسامی ژن‌هایی که تفاوت بیان داشته‌اند و همچنین جدولی که بعد از فیت کردن مدل نهایی بدست آمده، قرار گرفته است.

در نهایت این فایل گزارش که در حال مطالعه آن هستید هم در کنار این فایل‌ها قرار گرفته است.

## ۷ نتیجه‌گیری نهایی

در این پروژه، ما به تحلیل داده‌های ریزآرایه برای بیماری لوکمی حاد مغزاستخوان پرداختیم. ابتدا به تحلیل مواردی نظری همبستگی نمونه‌های مختلف پرداخته و پس از آن با توجه به شباهت بین

<sup>22</sup>Tyrosine Kinase Inhibitor

<sup>23</sup>Serine/Threonine Kinase inhibitors

<sup>24</sup>DNA Mismatch Repair



دسته‌ای از نمونه‌های بیمار با سلول‌های CD34، اقدام به تحلیل تفاوت بیان ژن در آن‌ها کردیم. در نهایت براساس این تفاوت‌ها و به کمک Enrichr، در مورد Pathway ها و Gene Ontology های مربوط به آن اطلاعاتی کسب کردیم. در بخش‌هایی نهایی نیز توضیحاتی در مورد خود بیماری AML داده و همچنین به بررسی نتایجی که در مقالات دیگر مرتبط با این بیماری و Pathway و Gene Ontology های مربوطه وجود دارد، پرداختیم.

## مراجع

- [1] Edward Chen, Christopher Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Meirelles, Neil Clark, and Avi Ma'ayan. Enrichr: Interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128, 04 2013.
- [2] Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–W97, 05 2016.
- [3] Henry Y. Li, Frederick R. Appelbaum, Cheryl L. Willman, Richard A. Zager, and Deborah E. Bunker. Cholesterol-modulating agents kill acute myeloid leukemia cells and sensitize them to therapeutics by blocking adaptive cholesterol responses. *Blood*, 101(9):3628–3634, 2003.
- [4] Elizabeth C. Matheson and Andrew G. Hall. Assessment of mismatch repair function in leukaemic cell lines and blasts from children with acute lymphoblastic leukaemia. *Carcinogenesis*, 24(1):31–38, 01 2003.
- [5] Linjun Hu, Yuling Gao, Zhan Shi, Yang Liu, Junjun Zhao, Zunqiang Xiao, Jiayin Lou, Qiuran Xu, and Xiangmin Tong. Dna methylation-based prognostic biomarkers of acute myeloid leukemia patients. *Annals of Translational Medicine*, 7:737–737, 12 2019.
- [6] Acute myeloid leukemia (aml) in adults, <https://www.cancer.org/cancer/acute-myeloid-leukemia>.
- [7] Debasis Nayak, Theodore L. Roth, and Dorian B. McGavern. Microglia development and function. *Annual Review of Immunology*, 32(1):367–402, 2014. PMID: 24471431.
- [8] Tyrobp, <https://en.wikipedia.org/wiki/Tyrobp>, Dec 2020.
- [9] Daniel Trombly, Troy Whitfield, Srivatsan Padmanabhan, Jonathan Gordon, Jane Lian, Andre Wijnen, Sayyed Zaidi, Janet Stein, and Gary Stein. Genome-wide co-occupancy of aml1-eto and n-cor defines the t(8;21) aml signature in leukemic cells. *BMC genomics*, 16:309, 04 2015.



- [10] Adam Lamble and Evan Lind. Targeting the immune microenvironment in acute myeloid leukemia: A focus on t cell immunity. *Frontiers in Oncology*, 8:213, 06 2018.
- [11] Daniel Rosenblum and S Petzold. Neutrophil alkaline phosphatase: Comparison of enzymes from normal subjects and patients with polycythemia vera and chronic myelogenous leukemia. *Blood*, 45:335–43, 04 1975.
- [12] Sanjay B. Hari, Ethan A. Merritt, and Dustin J. Maly. Conformation-selective atp-competitive inhibitors control regulatory interactions and noncatalytic functions of mitogen-activated protein kinases. *Chemistry & Biology*, 21(5):628–635, 2014.
- [13] Yuan Ling, Zikang Zhang, Hua Zhang, and Zunnan Huang. Protein kinase inhibitors as therapeutic drugs in aml: Advances and challenges. *Current pharmaceutical design*, 23, 07 2017.