**سهند نوعی 9923087**

**تمرین سوم**

**1. بخش اول: تحلیل بیان ژن تفاضلی با استفاده از ابزارهای GEO2R**

**وارد پایگاه داده GEO شوید و دیتاست GSE15852 را پیدا کنید. این دیتاست شامل ۴۳ نمونه سرطان سینه و بافت های سالم متناظر آن است. با استفاده از ابزار GEO2R نمونه های سرطانی را با سالم مقایسه کنید.**

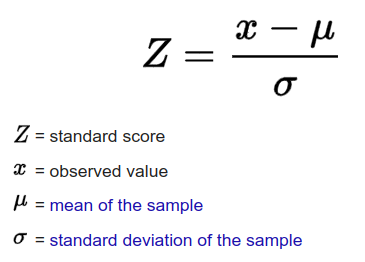
1. **توضیح دهید در بیان ژن تفاضلی مشکل استفاده از p-value چیست؟ آزمایشی طراحی کنید که نشان دهد احتمال رخ دادن خطای نوع اول با افزایش تعداد آزمون‌ها چگونه تغییر می‌کند. درباره روش‌های اصلاح p-value تحقیق کنید و دو روش Benjamini-Hochberg(BH) و Bonferroni را توضیح دهید و تفاوت آنها را بیان کنید.**

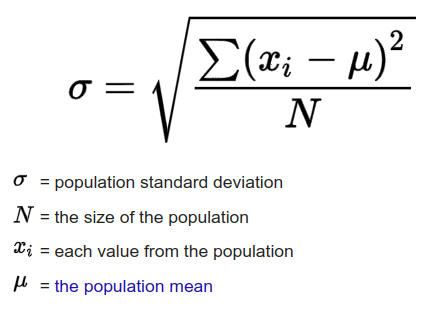
برای به دست آوردن p-value از روش های مختلفی میتوان استفاده کرد. مهم ترین آنها

Student's t-test, Standard score(z-score), chi-square test

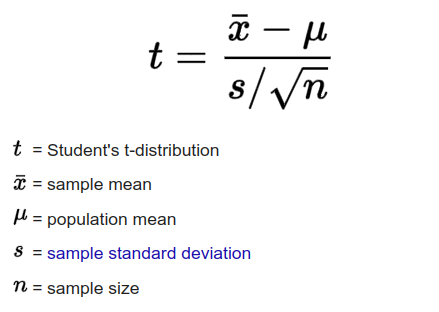
است که با توجه به فرمول به دست آوردن هر یک از آنها(که در ادامه آورده شده است) مقادیر به دست آمده از آنها به اندازه جامعه آماری و تعداد نمونه ها بستگی دارد.

فرمول z-score:

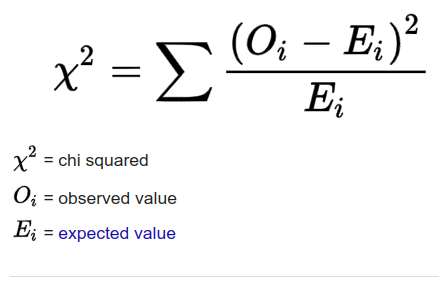


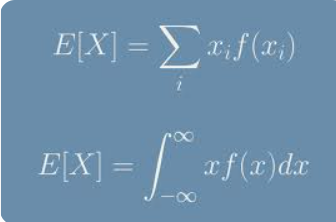


فرمول t student:



فرمول chi-square:





با توجه به مقادیر به دست آمده از آنها به جدول مربوطه رجوع کرده و مقدار p-value را به دست می آوریم. بنابراین p-value به طور مستقیم به اندازه نمونه بستگی دارد و این مشکلات متعددی را ایجاد میکند.

برای بررسی مشکلات فرض میکنیم α = 0.05 است:

1. افزایش اندازه نمونه میتواند منجر به ایجاد خطای نوع اول شود یعنی فرض صفر میتواند به اشتباه رد شود. مثلا اگر ۱۰ نمونه داشته باشیم و هیچ کدام از این ده نمونه تفاوت معناداری نداشته باشند هیچ کدام داده significant محسوب نمیشوند اما اگر از هر کدام از ۱۰ نمونه ۱۰۰ نمونه کپی کنیم و دوباره آزمایش را انجام بدهیم، انتظار میرود حدود ۵۰ داده significant گزارش شود در حالی که در عمل هیچ کدام significant نیستند.

**آزمایش**: اگر دو گروه داشته باشیم که تفاوت significantی در آنها وجود نداشته باشند و 10 داده با توزیع نرمال از آنها تولید کنیم داده significantی با توجه به مقدار α احتمالا وجود نخواهد داشت. حال اگر تعدا همین داده های مشابه را به ۲۰، ۱۰۰، ۱۰۰۰ و بیشتر افزایش دهیم، خواهیم دید که تعداد داده های significant بیشتر میشود چون در z score، مقدار n افزایش یافته، مقدار σ کاهش پیدا میکند، مقدار z افزایش می یابد و در نهایت مقدار p-value کاهش می یابد و با توجه به ثابت ماندن مقدار α تعداد داده های significant افزایش می یابد در حالی که در عمل هیچ داده significantی وجود نداشته است.

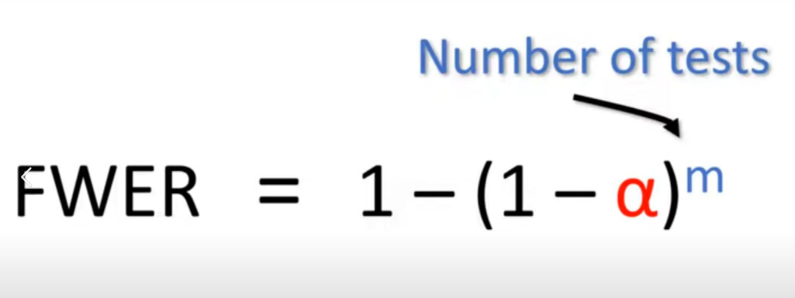
1. برعکس این مورد هم صادق است. اگر تعداد داده ها کم باشد، تفاوت بین داده ها significant باشد و این را از قبل بدانیم اما مقدار α برای نشان دادن این تفاوت کافی نباشد، داده ای significant تشخیص داده نمیشود.
2. در واقع انتخاب مقدار α به صورت سلیقه ای یا در حالت ایده آل بر اساس تجربه است و این میتواند برای بعضی آزمایش ها جواب بدهد اما برای بعضی نه. انتخاب نادرست مقدار p-value میتواند منجر به ایجاد خطای اول(false positive - بخصوص در تعداد داده های زیاد) یا خطای دوم(false negative - بخصوص در داده های کم) بشود.

روش های مختلفی برای اصلاح مشکل p-value وجود دارد. از جمله معروف ترین آنها میتوان به موارد زیر اشاره کرد که در سایت NCBI هم ذکر شده است:

* Benjamini & Hochberg (False discovery rate)
* Benjamini & Yekutieli
* Bonferroni
* Holm

روش Bonferroni:

Family-wise error rate معیاری برای بررسی احتمال رخ دادن خطا با انجام m تست است و فرمول آن به شرح زیر است:



اگر نمودار آن را بررسی کنیم خواهیم دید که اگر یک تست داشته باشیم با مقدار α حدود 0.05 احتمال وجود false positive برابر 0.05 است که مطلوب ما هست اما اگر 60 تست همزمان انجام شوند احتمال وجود false positive حدود 95 درصد است. برای رفع این مشکل از روش bonferroni استفاده میشود که در احتمال را در همان حدود 5 درصد ثابت نگه میدارد. در این روش مقدار α تقسیم بر تعداد تست میشود مثلا اگر 60 تست داشته باشیم 0.05/60 میشود و این مقدار α جدید خواهد بود و مقدار p-value داده ها برای تشخیص diffferential significant بودن داده ها با این مقدار مقایسه میشود. همچنین میتوان به جای تغییر α، مقدار p-value را مطابق سوال تغییر داد. برای این کار مقادیر p-value به دست آمده اولیه در تعداد تست ها ضرب میشود و این p-valueها با α قبلی مقایسه میشود.

روش Benjamin-Hochberg:

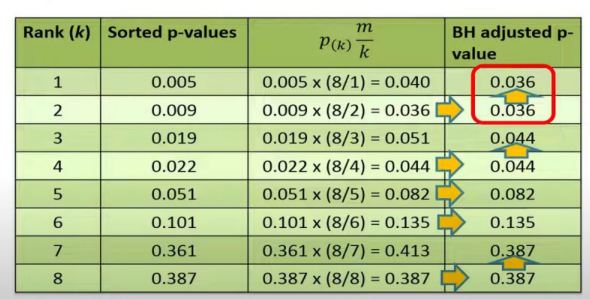
در این روش هم مجدد هم میتوان α را تغییر داد هم میتوان p-value را تغییر داد. در اینجا من فقط روش تغییر p-value را توضیح میدهم. در این روش ابتدا تست ها را به ترتیب p-value بدست آمده به صورت صعودی مرتب میکنیم و با شروع از 1 به هر کدام از این تست ها rank میدهیم. حال p-value جدید را از فرمول زیر بدست می آوریم:



که در این رابطه p-value هر کدام از تست ها در تعداد تست ها صرب میشود و تقسیم بر رنک آن تست میشود. سپس از آخرین رنک به سمت اولین رنک به ترتیب زیر عمل میکنیم:

* در رنک m مقدار به دست آمده را p-value جدید در نظر میگیریم.
* در رنک های دیگر در صورتی که p-value به دست آمده در k - 1 کوچکتر مساوی p-value رنک k بود، p-value به دست آمده را به عنوان p-value جدید در نظر میگیریم
* در صورتی که p-value رنک k - 1 بزرگتر از k بود، p-value به دست آمده برای رنک k را برای k - 1 در نظر میگیریم.

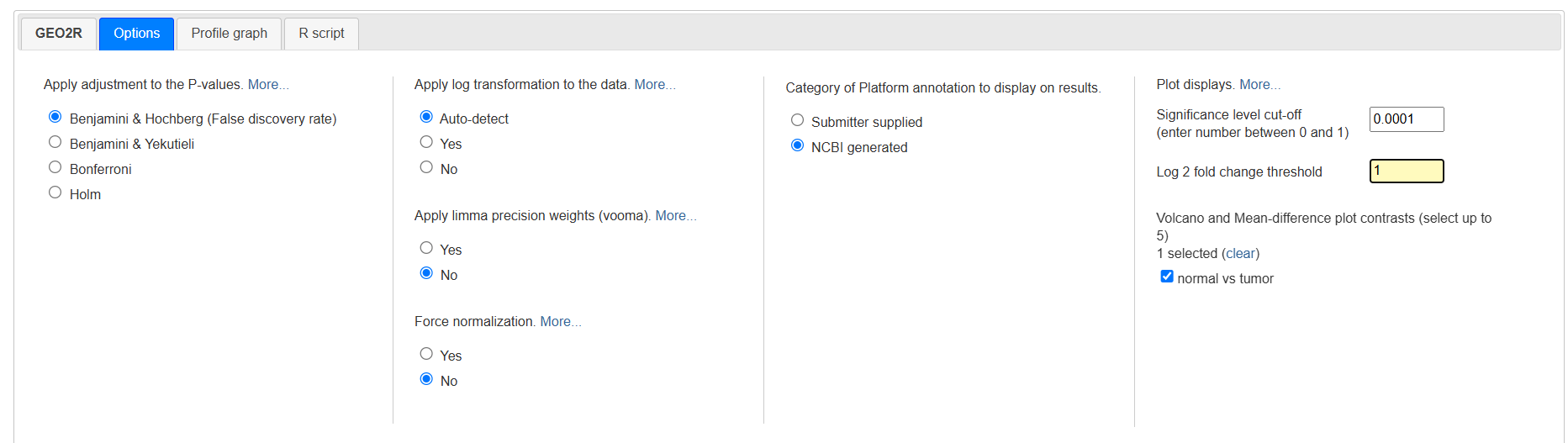
همانند شکل زیر:



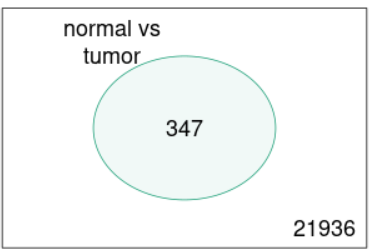
برای انتخاب تست هایی که null hypophysis را رد میکنند از رنک آخر شروع میکنیم و به محض رسیدن به بزرگترین رنکی که مقدار p-value آن از مقدار α کمتر بود از آن رنک به قبل تست null hypophysis را رد میکنند و به عنوان داده differential significant محسوب میشوند.

تفاوت این دو روش آن است که روش bonferroni سخت گیرانه تر عمل میکند و تعداد تست های false positive در آن قطعا کوچکتر مساوی تعداد در BH است. بنابراین در مواردی که false positive نداشتن بسیار اهمیت دارد روش Bonferroni مناسب تر است. در مواردی هم که false negative نداشتن بسیار مهم است روش BH بهتر است.

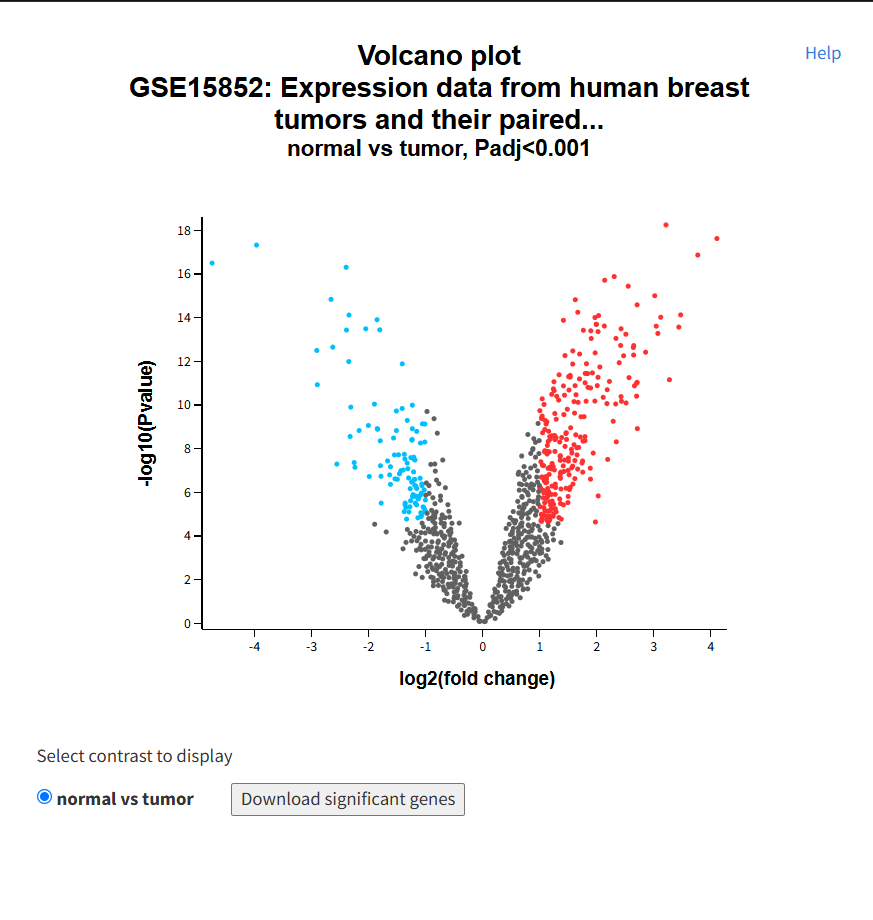
1. **ژن‌هایی را معنادار در نظر بگیرید که adj p-value < 0.001 و 1 < logFC باشد. سپس نمودار‌های Volcano plot و Mean Difference(MD) plot را رسم کنید و هرکدام را توضیح دهید. ژن‌های معنا‌دار را ذخیره کنید. از آنها برای بخش دوم تمرین استفاده خواهیم کرد.**



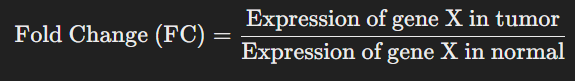
با توجه به این تنظیمات 347 ژن معنادار پیدا شدند که فایل csv آنها ضمیمه شده است. به دو گروه تقسیم شدند. گروه tumor و گروه normal



تحلیل نمودار Volcano plot:



با توجه به نمودار نقاط قرمز و آبی ژن های significant هستند. هر چه به سمت بالای محور yها میرویم نشان دهنده بیشتر significant بودن ژن ها است. چون هرچه -log(Pvalue) بیشتر میشود یعنی مقدار p-value به دست آمده آن کمتر است و در واقع احتمال این که واقعا موثر بوده است بیشتر میشود و احتمال false positive بودن آن کمتر است. هر چه به سمت دو سر نمودار میرویم اندازه تفاوت بین دو گروه بیشتر است. رابطه Fold Change به صورت به صورت زیر است:



اگر لگاریتم این رابطه در مبنای 2 را بگیریم log2FC به دست می آید و هر چه این عدد از صفر فاصله بگیرد یعنی میزان تفاوت بیان ژن در دو گروه بیشتر است. حال هر چه بزرگ تر از صفر باشد یعنی میزان بیان ژن X در داده های داده های تومور و ناسالم بیشتر است(upregulated) که با قرمز نشان داده شده است و برعکس هر چه منفی تر باشد(downregulated) یعنی در داده های سالم با میزان بیشتری تفاوت بیان بیشتری دارد که با آبی نشان داده شده است. بنابراین اگر بخواهیم ژن هایی که نسبت به بقیه با احتمال بیشتری true positive هستند را در بیاوریم میتوان به صورت زیر اشاره کرد:

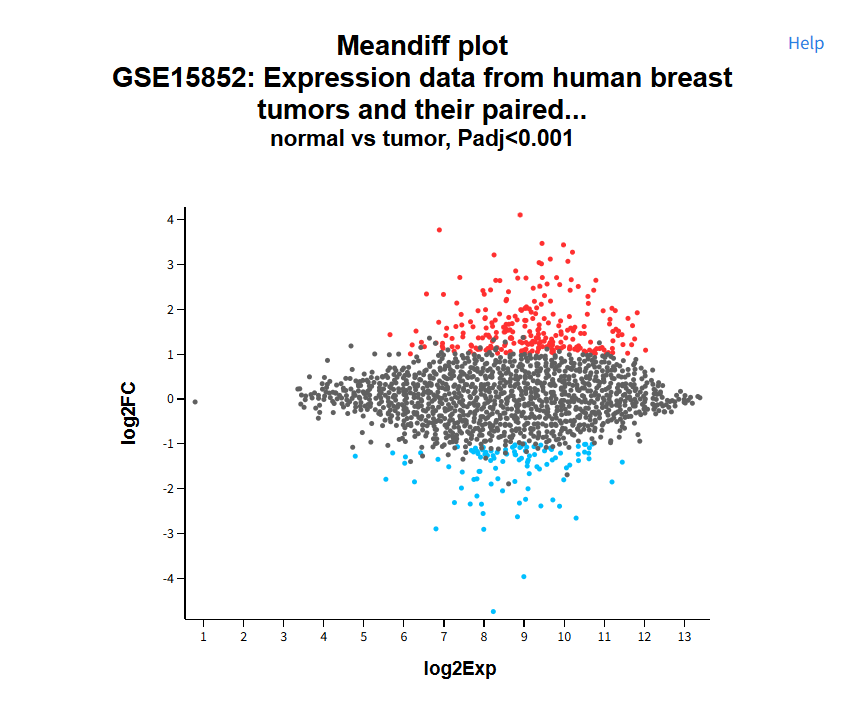
RBP4, PDE3B, PPP1R1A

Significantترین با بیشترین تفاوت بیان ژن که در بیماران بیشترین بیان را دارند

KRT19, CD24, EPCAM, TACSTD2

Significantترین با بیشترین تفاوت بیان ژن که در افراد سالم بیشترین بیان را دارند.

تحلیل نمودار Mean Difference plot:



محور افقی لگاریتم میانگین بیان هر ژن در دو گروه را نشان میدهد و در واقع شدت سیگنال فلورسنت حاصل است. میبینیم که اکثر ژن ها اقلا شدت بیان ژن 8 را دارند. اکثرا ژن هایی که بیشترین و کمترین بیان ژن را داشتند جز ژن های significant نیستند. همچنین احتمالا دستگاه اندازه گیری هم سالم بوده است چون نسبت به شدت بیان خاصی بایاس نداشته است.

با توجه به نتایج بدست آمده:

ژن های RBP4 و PDE3B و SAA2 و CFD به میزان خوبی بیان شده اند و fold change زیادی دارند و در سلول های سرطانی احتمالا موثرترند. ژن FHL1 ژن موثر در سرطان است که به بیشترین بیان را داشته است.

همچنین ژن های KRT19 و CD24 و TACSTD2 بیشترین تفاوت بیان در سلول های سالم را دارند. ژن CD24 بیشترین میزان بیان در سلول های سالم را در میان ژن های significant داشته است.

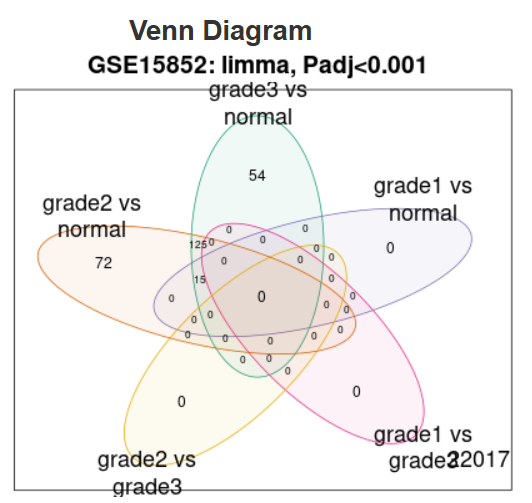
1. **در انتها، نمونه‌ها را براساس Grade به چهار گروه تقسیم کنید. سپس مجدد آنالیز را انجام دهید. نمودار‌های UMAP، Venn diagram، Boxplot را رسم کنید و آنها را تحلیل کنید.**



تحلیل نمودار UMAP:

UMAP یک روش برای dimensionality reduction است که در اینجا به دو بعد کاهش یافته است. با توجه به به وجود نیامدن clusterهای مجزا، اطلاعات خاصی از این نمودار به دست نمی‌آید. این نشان دهنده آن است که دو بعد برای دریافت اطلاعات significant کافی نیست یا تعداد neighborها به درستی انتخاب نشده است.

تحلیل Venn diagram:



با توجه به نمودار 22017 ژن بی تاثیر هستند.

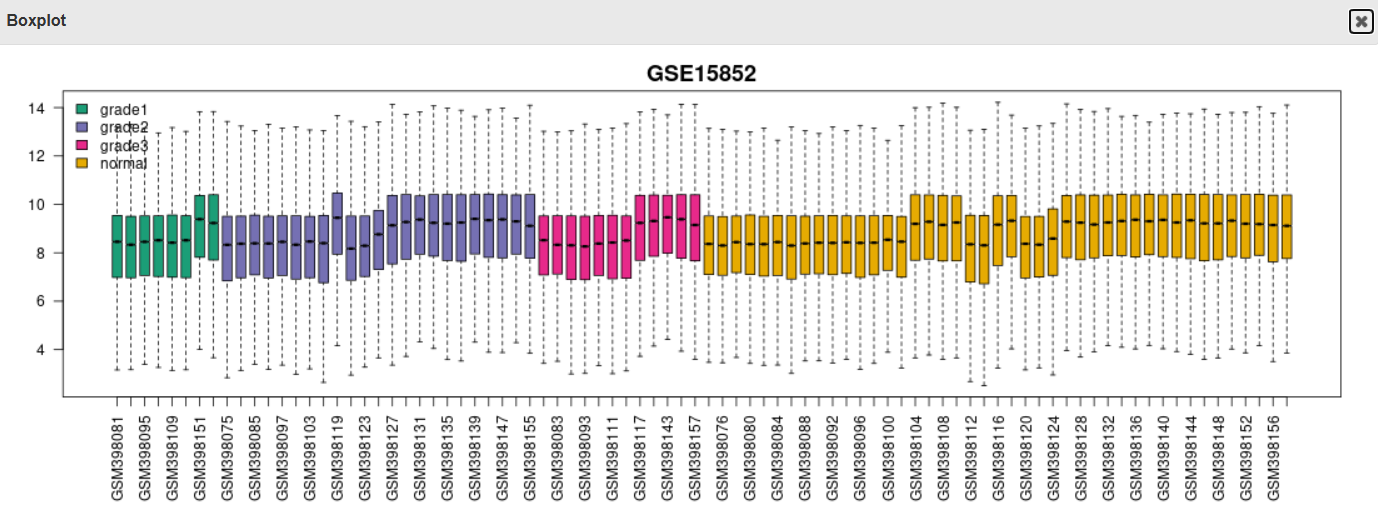
72 ژن significant در مقایسه grade2 و normal پیدا شده است

54 ژن significant در مقایسه grade3 و normal پیدا شده است

15 ژن significant هم در مقایسه grade2 و normal و هم در مقایسه grade3 و normal و هم در مقایسه grade1 و normal پیدا شده است.

125 ژن significant هم در مقایسه grade2 و normal و هم در مقایسه grade3 و normal پیدا شده است.

تحلیل Box plot:



خط داخل جعبه نشان دهنده میانه هست

دو سر جعبه محدوده چارک اول و سوم را نشان می دهد

دو سر هر نمودار جعبه ای محدوده داده ها را مشخص میکند

در نهایت با توجه به نمودار داده outlier در این بررسی احتمالا وجود نداشته و داده ها دارای کیفیت خوبی بوده اند و قبل از آزمایش داده های پرت حذف شده اند. بنابراین میتوان به نتایج آزمایش استناد کرد.

بخش دوم: غنی سازی مسیرهای زیستی

**در این بخش، ژنهای معنادار شناسایی شده در مرحله قبل را با استفاده از ابزار Enrichr از نظر مسیرها ی زیستی و عملکردها ی ژنی تحلیل کنید. برای این منظور از پایگاه داده های KEGG و GO استفاده کنید.**

1. **تحلیل مسیرهای زیستی در KEGG :**

* **لیست ژنهای معنادار را در ابزار Enrichr وارد کنید و مسیرها ی زیستی معنادار در پایگاه داده KEGG را شناسایی کنید. توضیح دهید که این مسیرهای زیستی چه نقشی دارند و چگونه با پیشرفت یا شکل گیری سرطان مرتبط هستند. نتایج را با جداول و نمودارهای مناسب ارائه دهید و نکات برجسته هر مسیر زیستی را توضیح دهید.**

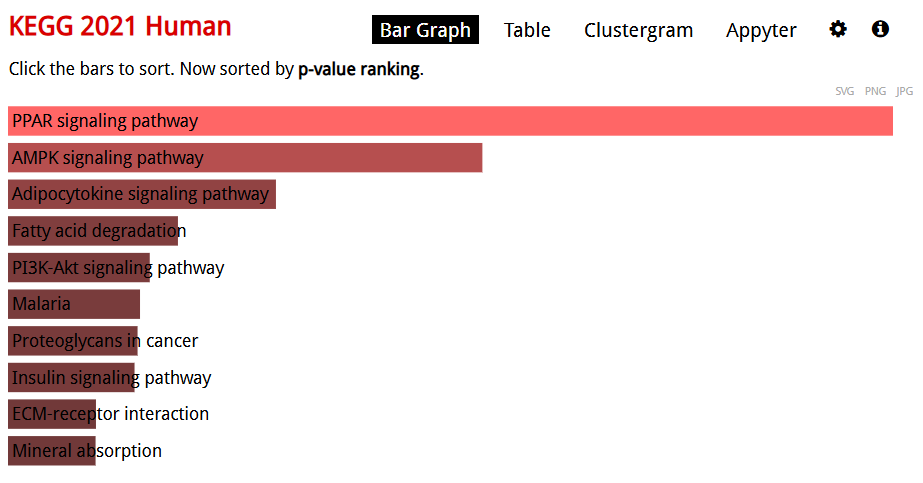
ابتدا سمبل ژن های significant را وارد میکنیم که 347 عدد هستند.

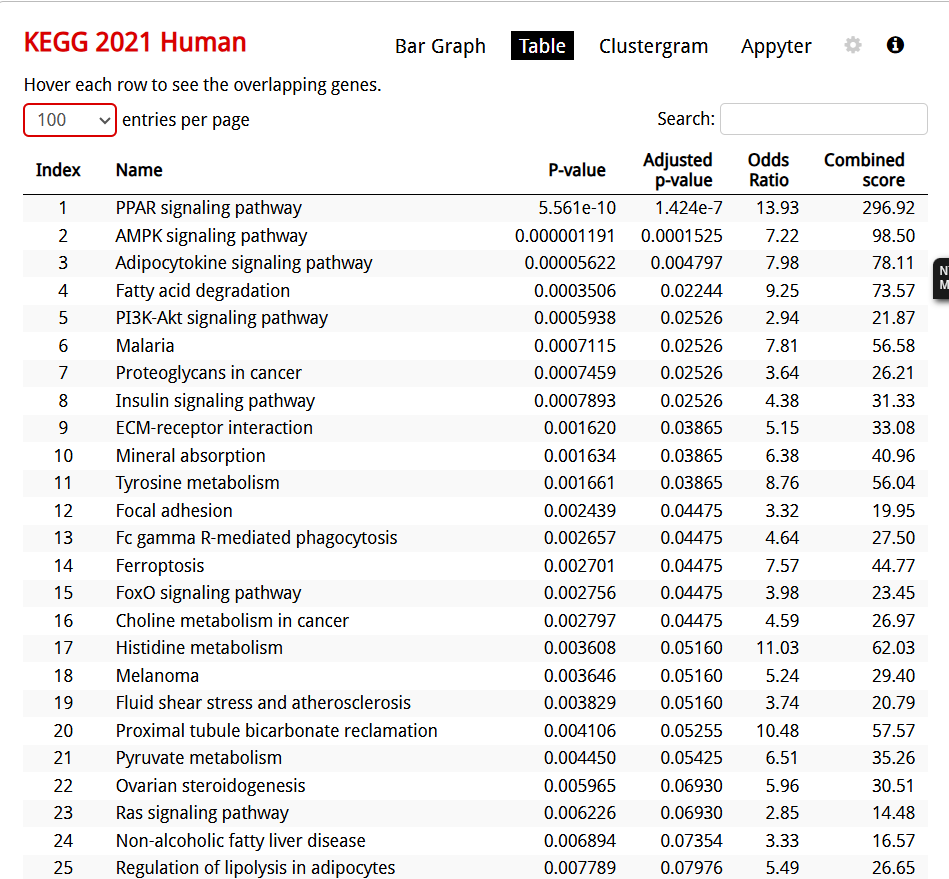
ژن هایی که در خروجی geo2r تکراری بودن مثل HBA2///HBA1 در برخی موارد مورد اول را نگه داشتم و در برخی موارد مورد دوم را. به نظرم این کار منطقی تر از نگه داری فقط یکی از آنها باشد هرچند در نتیجه کار ممکن است تاثیری نداشته باشد.

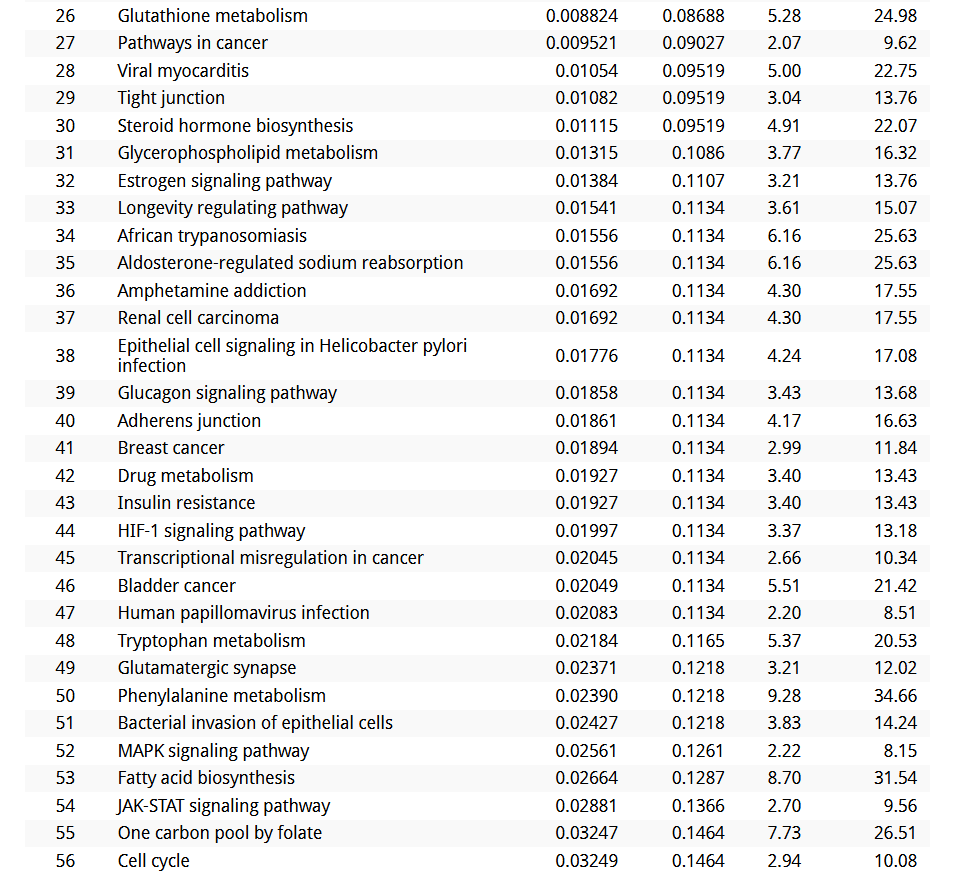
نرم افزار بعد از حذف کردن تکراری ها 285 ژن را در نظر میگیرد.

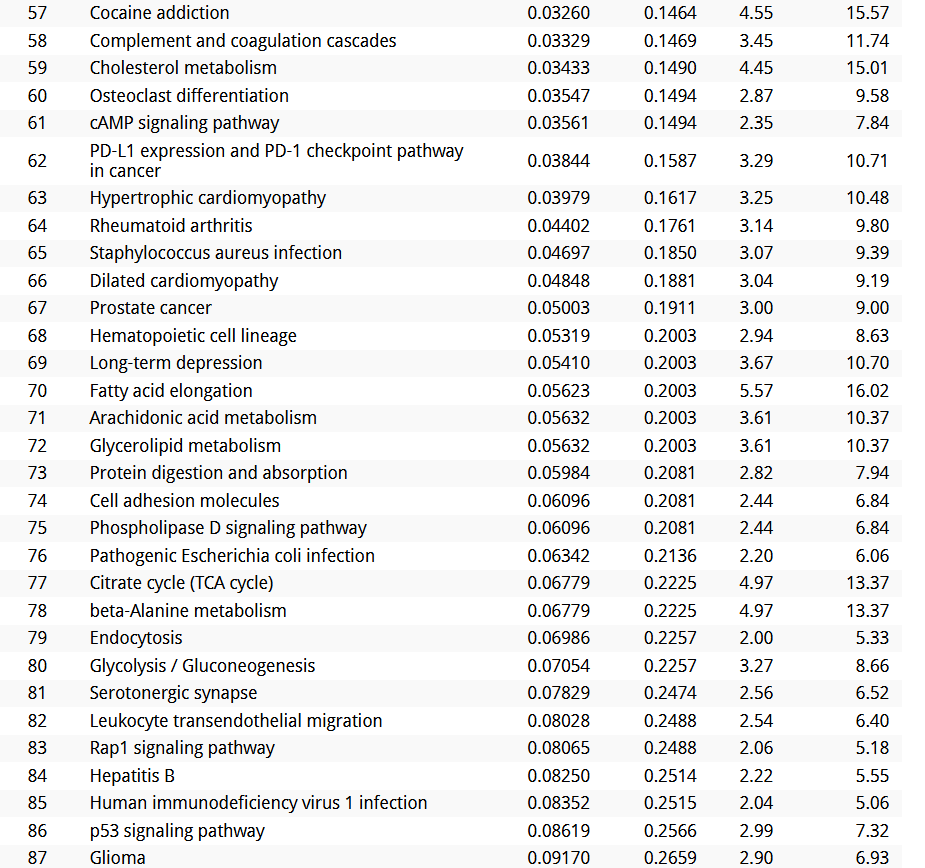


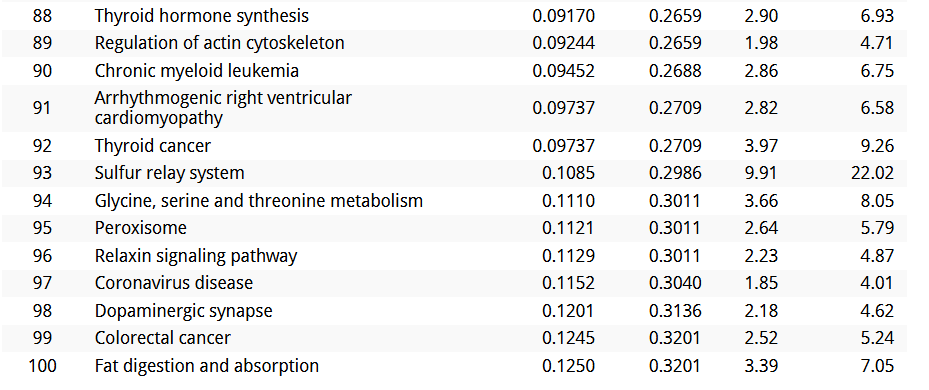
KEGG:







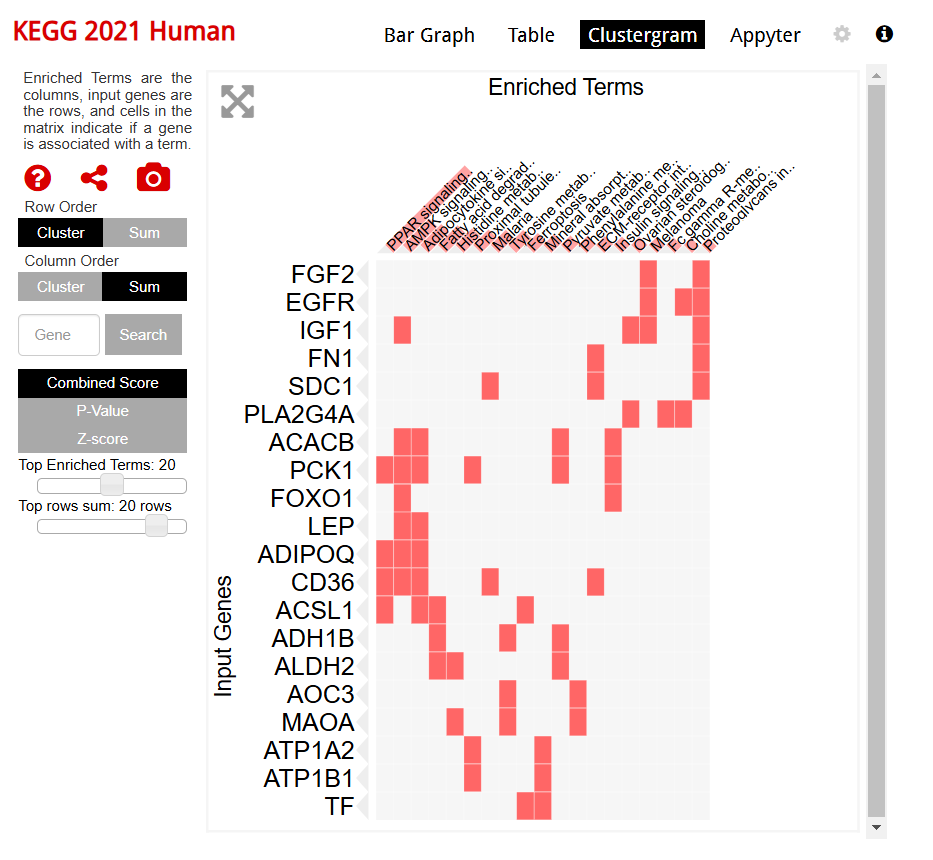




اگر در اینجا هم p-value های کوچکتر از 0.001 را در نظر بگیریم 8 مسیر زیستی معنادار به دست می آید

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | PPAR signaling pathway | 5.561e-10 | 1.424e-7 | 13.93 | 296.92 |
| 2 | AMPK signaling pathway | 0.000001191 | 0.0001525 | 7.22 | 98.50 |
| 3 | Adipocytokine signaling pathway | 0.00005622 | 0.004797 | 7.98 | 78.11 |
| 4 | Fatty acid degradation | 0.0003506 | 0.02244 | 9.25 | 73.57 |
| 5 | PI3K-Akt signaling pathway | 0.0005938 | 0.02526 | 2.94 | 21.87 |
| 6 | Malaria | 0.0007115 | 0.02526 | 7.81 | 56.58 |
| 7 | Proteoglycans in cancer | 0.0007459 | 0.02526 | 3.64 | 26.21 |
| 8 | Insulin signaling pathway | 0.0007893 | 0.02526 | 4.38 | 31.33 |

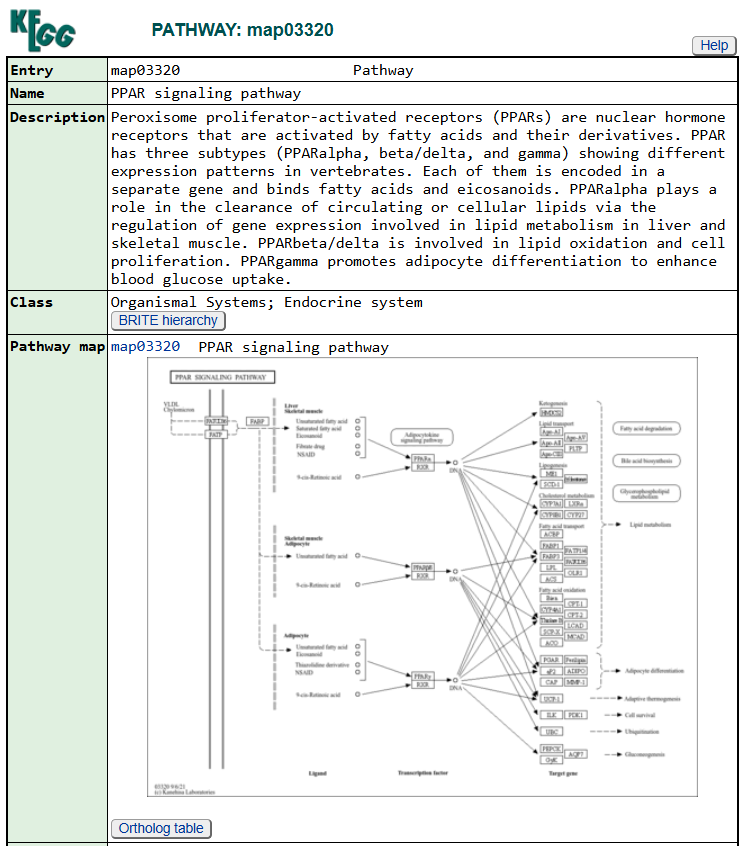
و ژن های مرتبط با این مسیر های زیستی نشان داده شده اند.

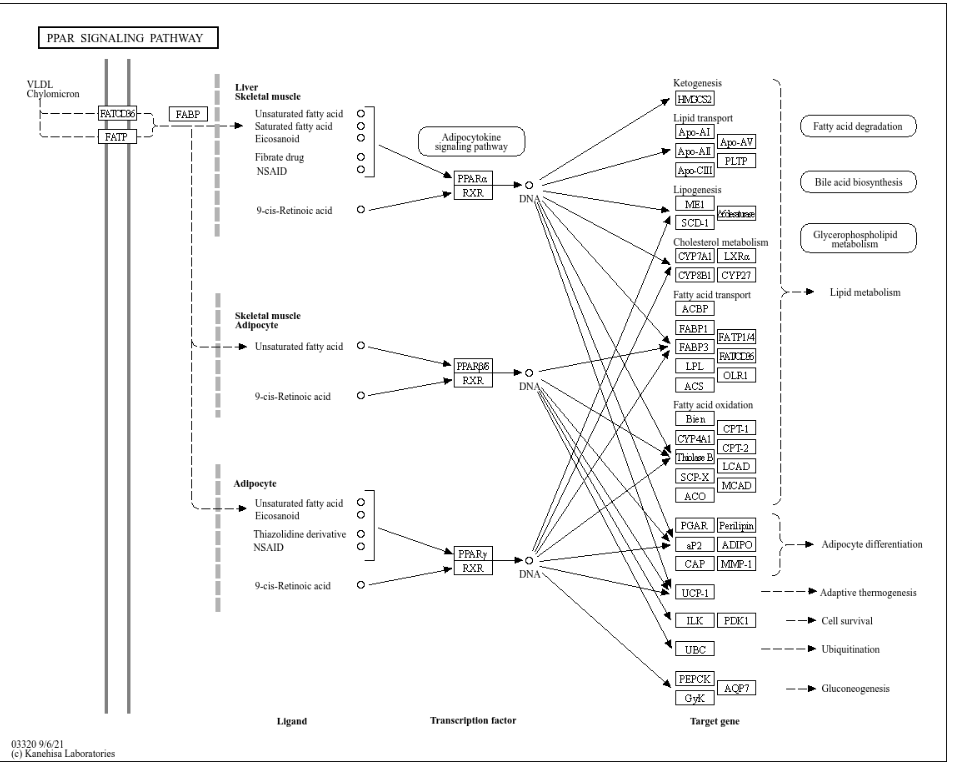


### PPAR Signaling Pathway

نقش: این مسیر در تنظیم متابولیسم اسیدهای چرب و هومئوستاز گلوکز نقش دارد و همچنین باعث مهار التهاب و استرس اکسیداتیو می‌شود.

ارتباط با سرطان سینه: اختلال در عملکرد این مسیر می‌تواند باعث تغییر در متابولیسم انرژی و ایجاد شرایطی مانند التهاب مزمن شود که با رشد و بقای سلول‌های سرطانی در سرطان سینه مرتبط است.

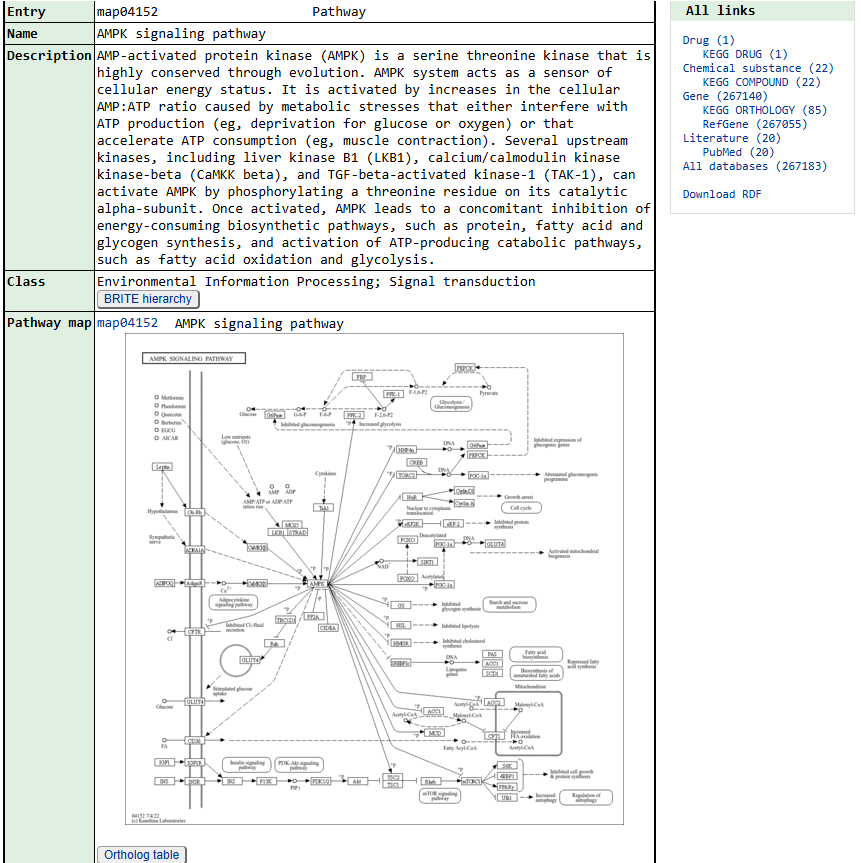


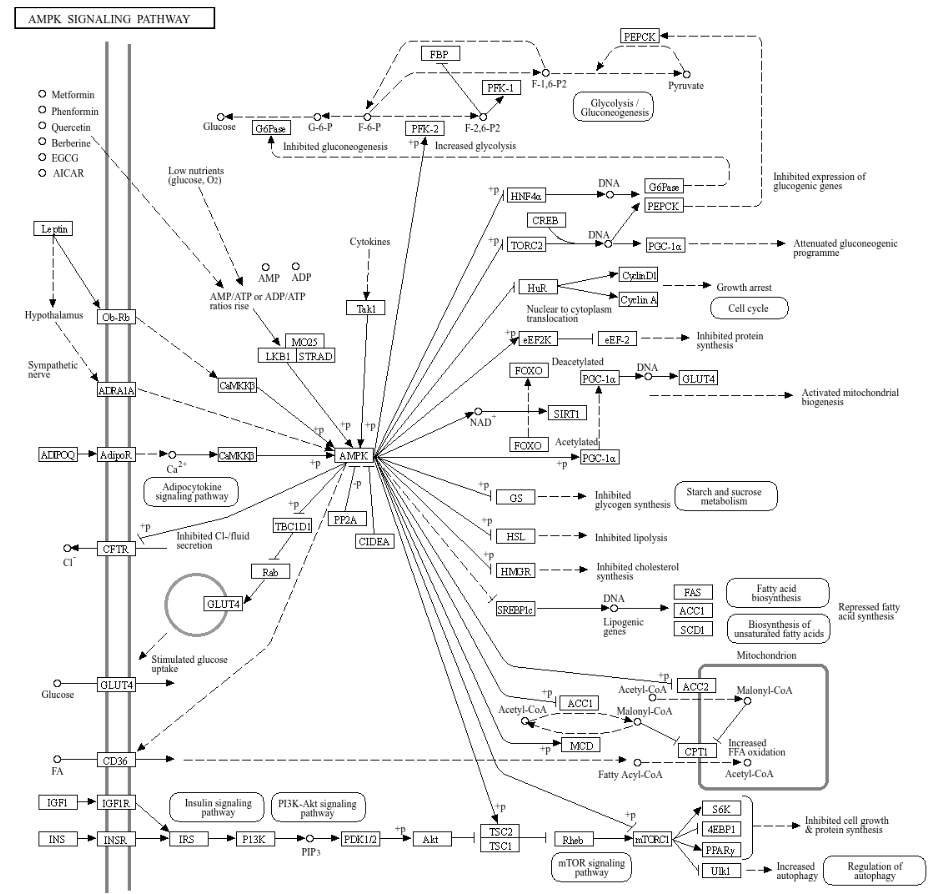


### AMPK Signaling Pathway

نقش: AMPK به عنوان حسگر انرژی سلولی، تعادل انرژی را تنظیم کرده و مسیرهای مرتبط با رشد سلولی را کنترل می‌کند.

ارتباط با سرطان سینه: سرکوب یا فعال‌سازی غیرطبیعی AMPK می‌تواند منجر به تنظیم غیرطبیعی رشد و تکثیر سلولی در سرطان سینه شود. AMPK معمولاً به عنوان یک مسیر محافظ در برابر رشد سلول‌های سرطانی عمل می‌کند.

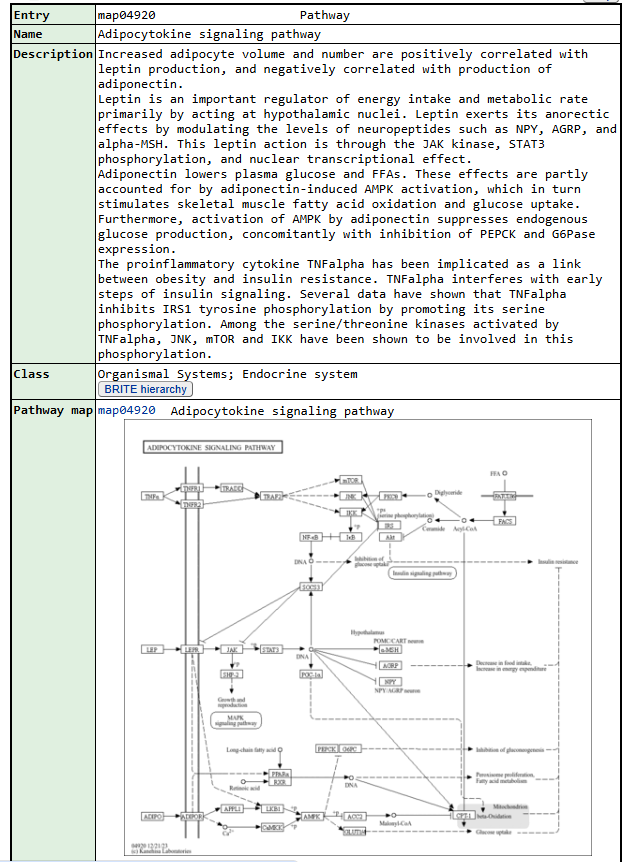


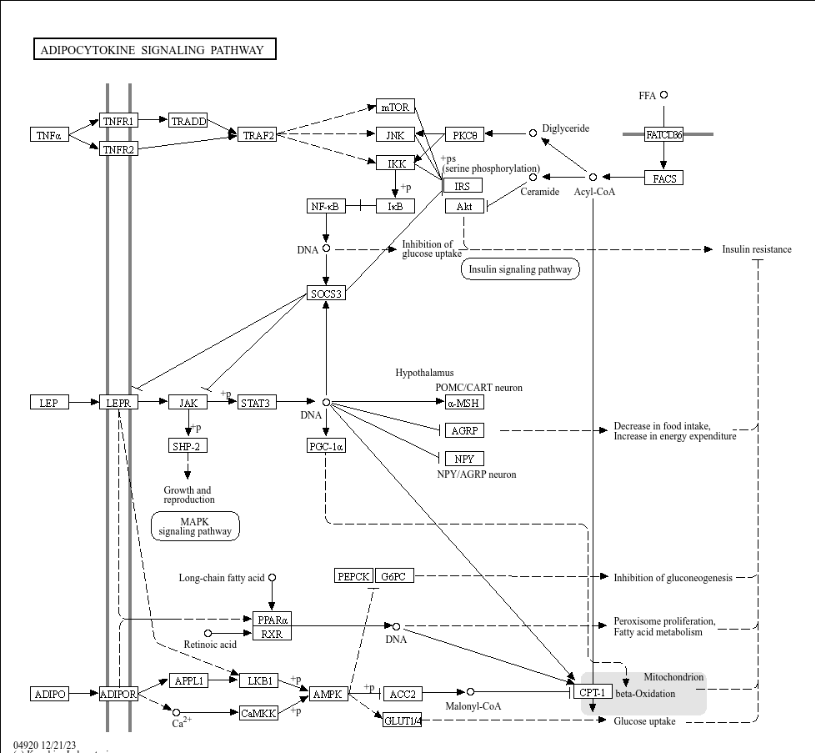


### Adipocytokine Signaling Pathway

نقش: این مسیر شامل سیگنال‌دهی آدیپوکاین‌ها (مانند لپتین و آدیپونکتین) است که متابولیسم انرژی، التهاب و حساسیت به انسولین را تنظیم می‌کند.

ارتباط با سرطان سینه: لپتین که در این مسیر نقش دارد، می‌تواند رشد و مهاجرت سلول‌های سرطانی را تحریک کند. سطوح پایین آدیپونکتین نیز با افزایش خطر سرطان سینه همراه است.

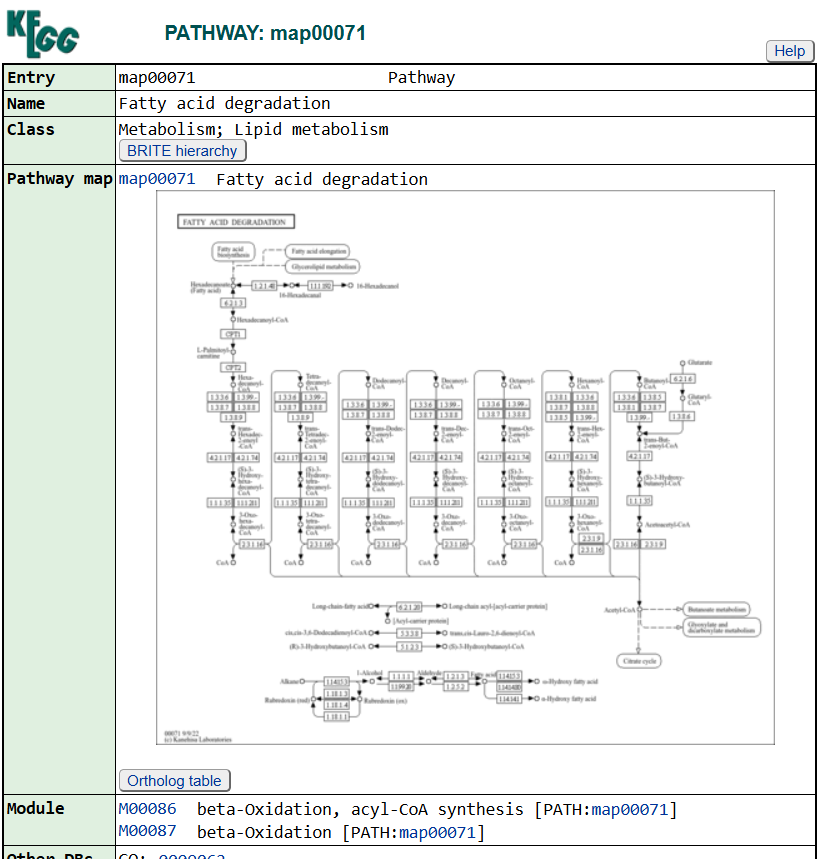


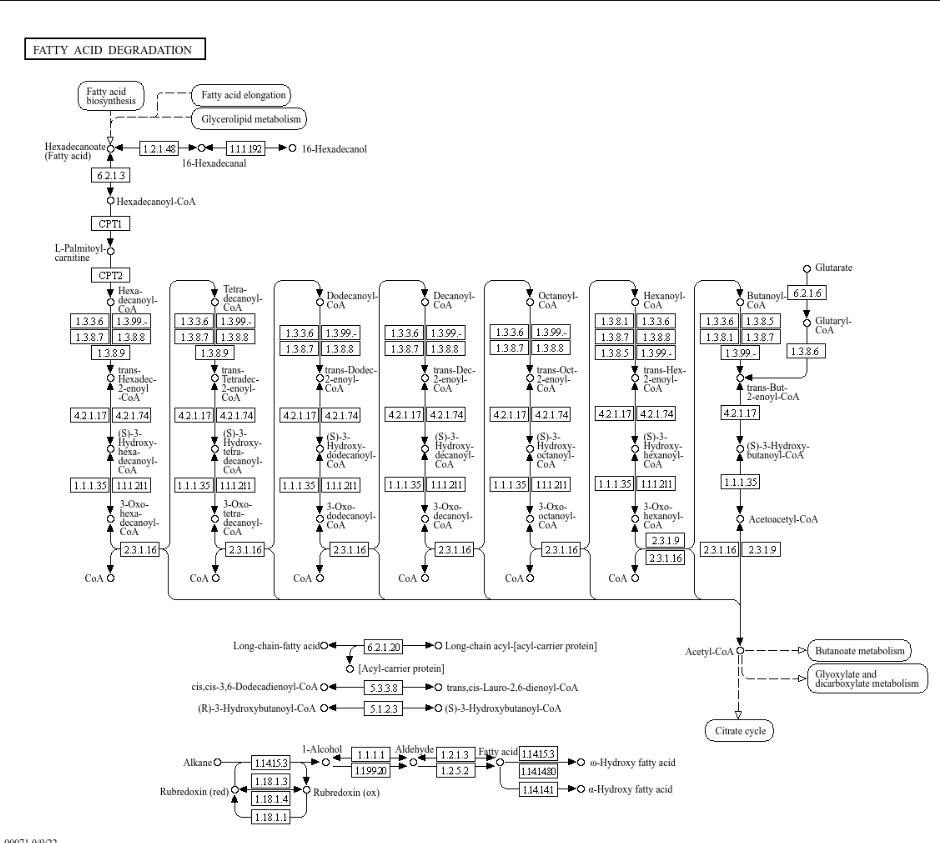


### Fatty Acid Degradation

نقش: این مسیر مسئول تجزیه اسیدهای چرب و تولید انرژی است.

ارتباط با سرطان سینه: سلول‌های سرطانی می‌توانند برای تأمین انرژی خود به این مسیر وابسته باشند. تغییر در متابولیسم اسیدهای چرب معمولاً در سلول‌های سرطانی مشاهده می‌شود.

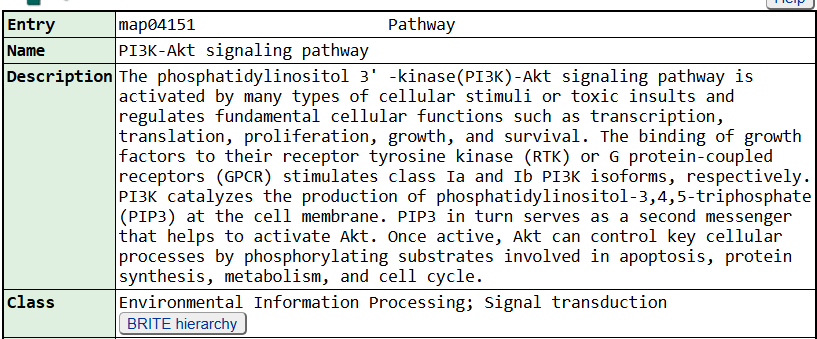


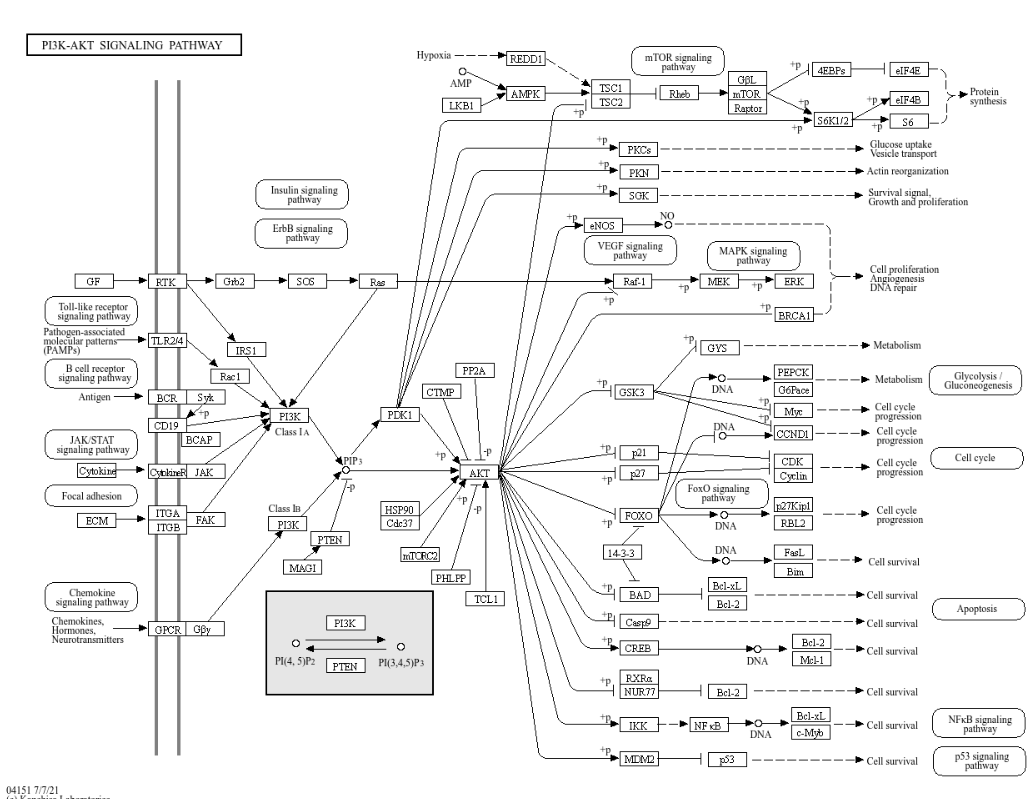


### PI3K-Akt Signaling Pathway

نقش: این مسیر در تنظیم رشد، بقا و متابولیسم سلولی نقش دارد.

ارتباط با سرطان سینه: PI3K-Akt یک مسیر شناخته‌شده در سرطان سینه است. جهش‌ها یا فعال‌سازی بیش‌ازحد این مسیر می‌تواند باعث رشد غیرطبیعی سلولی و مقاومت به درمان شود.



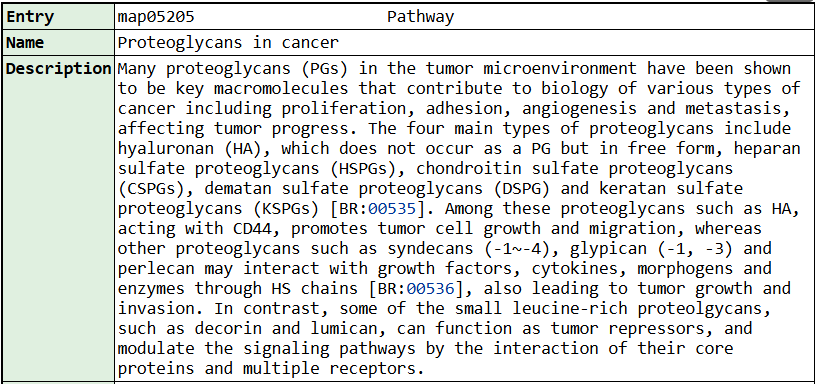


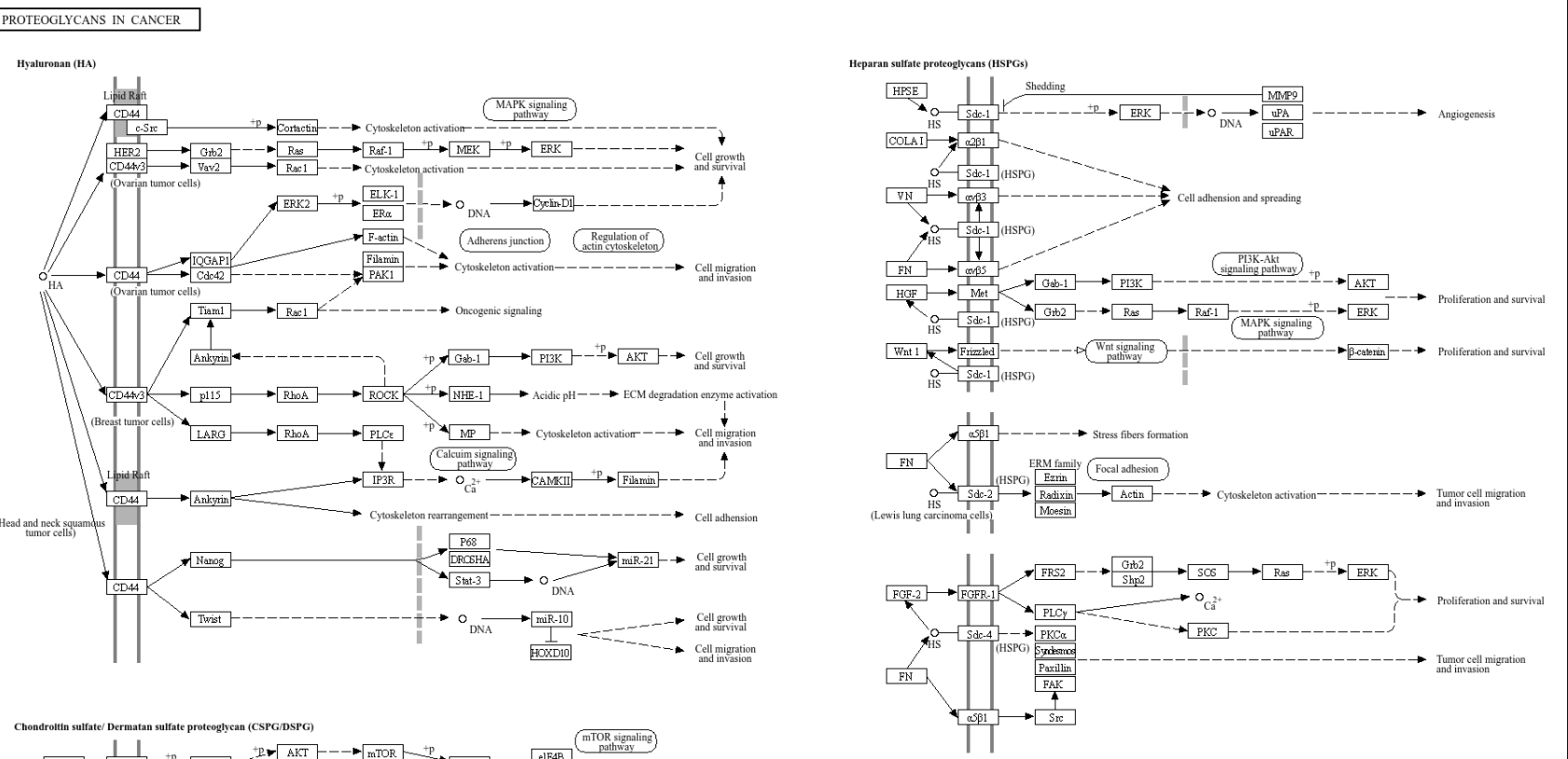
### 

### Proteoglycans in Cancer

نقش: پروتئوگلیکان‌ها در ماتریکس خارج‌سلولی قرار دارند و در فرآیندهای چسبندگی سلولی و انتقال سیگنال نقش دارند.

ارتباط با سرطان سینه: پروتئوگلیکان‌ها می‌توانند از طریق تعامل با گیرنده‌های سطح سلولی، مسیرهای مرتبط با رشد و مهاجرت سلولی را فعال کنند و متاستاز سرطان سینه را تسهیل نمایند.

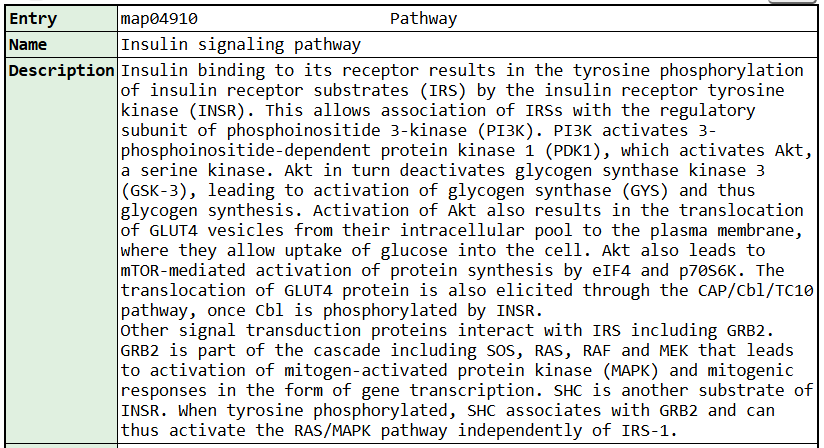


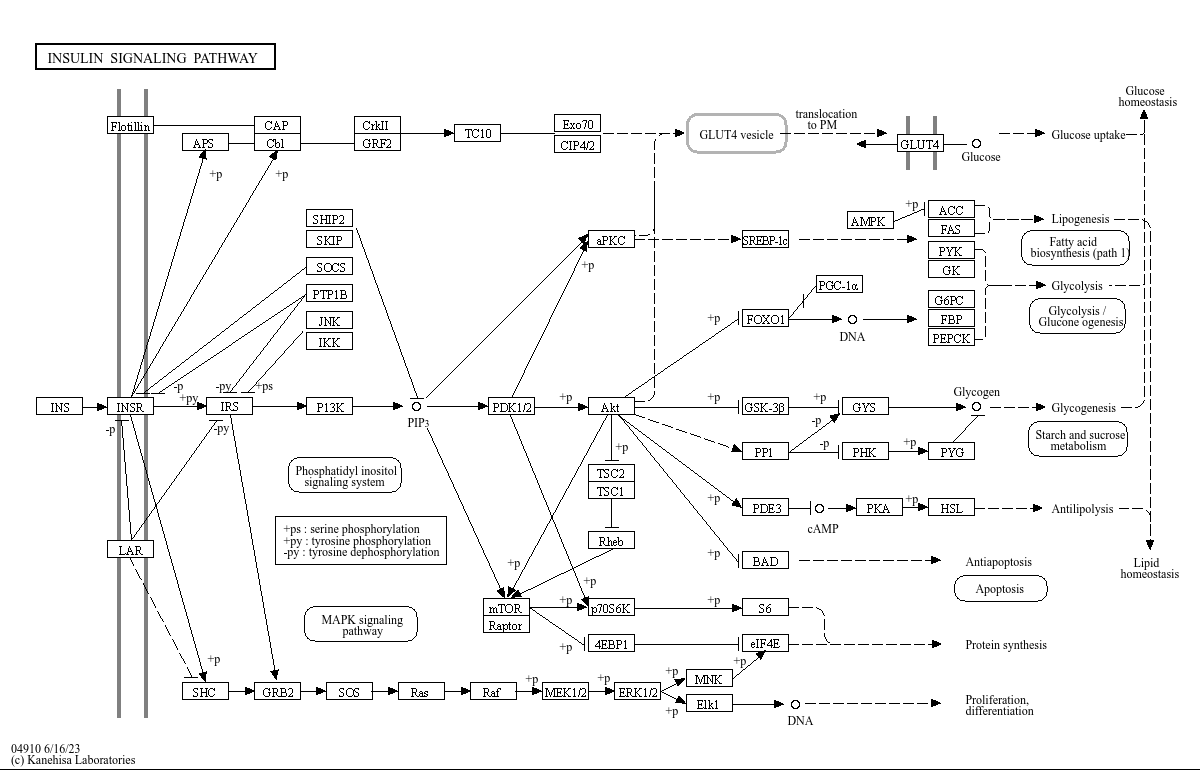


### Insulin Signaling Pathway

نقش: این مسیر متابولیسم گلوکز و لیپیدها را تنظیم می‌کند و نقش مهمی در رشد سلولی دارد.

ارتباط با سرطان سینه: مقاومت به انسولین و افزایش سطح انسولین می‌تواند مسیرهای رشد سلولی مرتبط با سرطان سینه را فعال کند. این موضوع به‌ویژه در افراد دارای اضافه‌وزن یا چاقی اهمیت دارد.





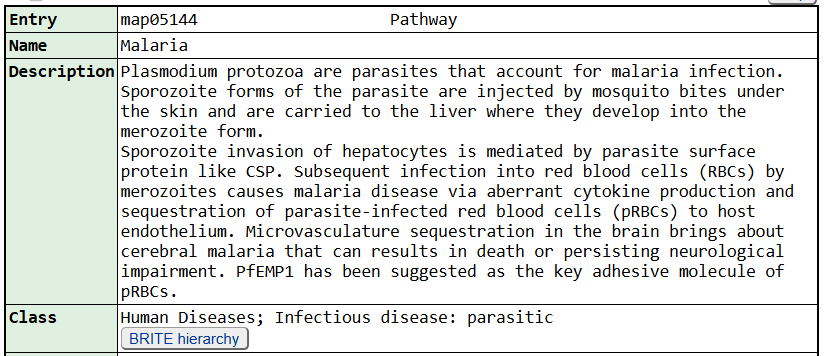
### Malaria Pathway

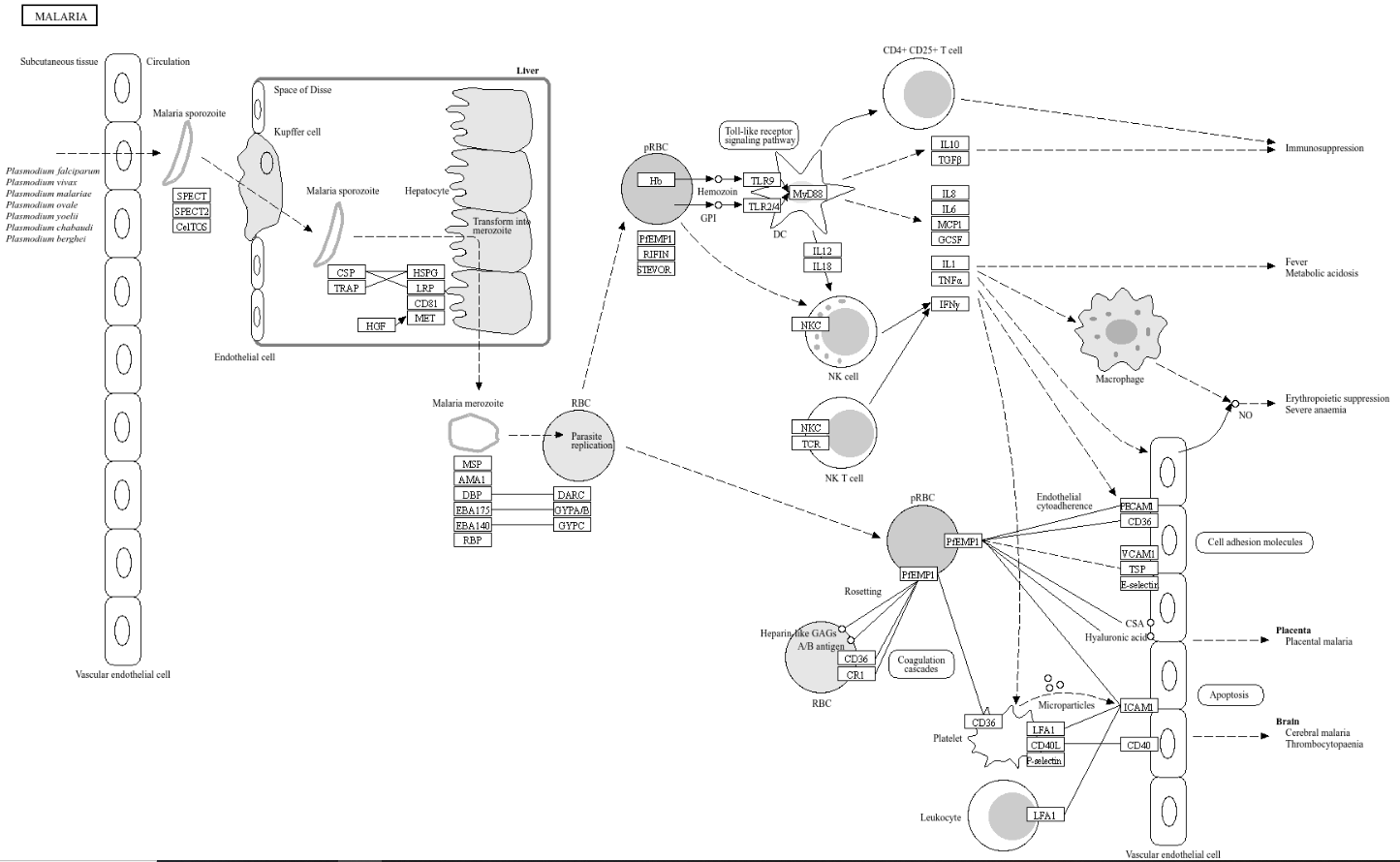
نقش: این مسیر به طور خاص به واکنش‌های ایمنی بدن در برابر انگل Plasmodium اشاره دارد که عامل بیماری مالاریا است. این مسیر شامل تنظیمات ایمنی، تولید سیتوکین‌ها و عوامل التهابی است.

التهاب مزمن: مالاریا با ایجاد التهاب سیستمیک، ریسک شکل‌گیری سرطان، از جمله سرطان سینه، را افزایش می‌دهد.

سرکوب ایمنی: ضعف سیستم ایمنی ناشی از عفونت‌های طولانی‌مدت مالاریا به سلول‌های سرطانی کمک می‌کند از شناسایی فرار کنند.

استرس اکسیداتیو: افزایش استرس اکسیداتیو در مالاریا باعث آسیب به DNA و جهش‌های مرتبط با سرطان می‌شود.

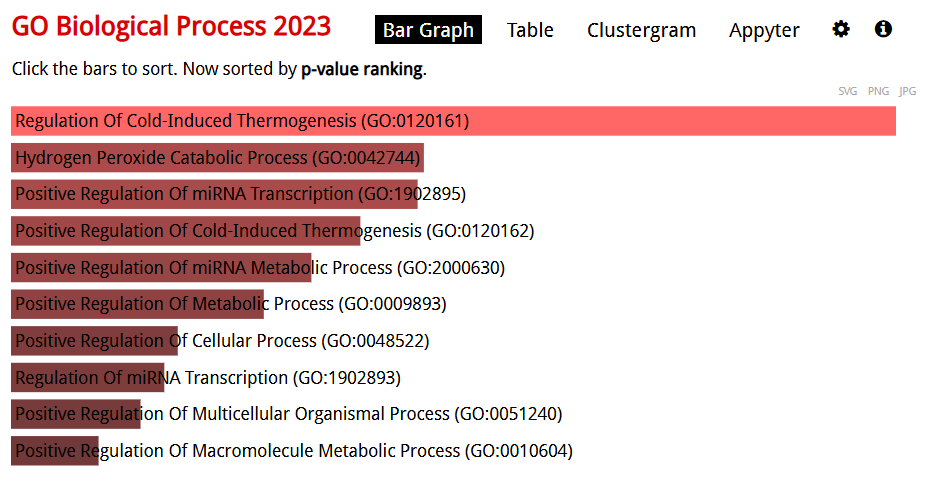
ر

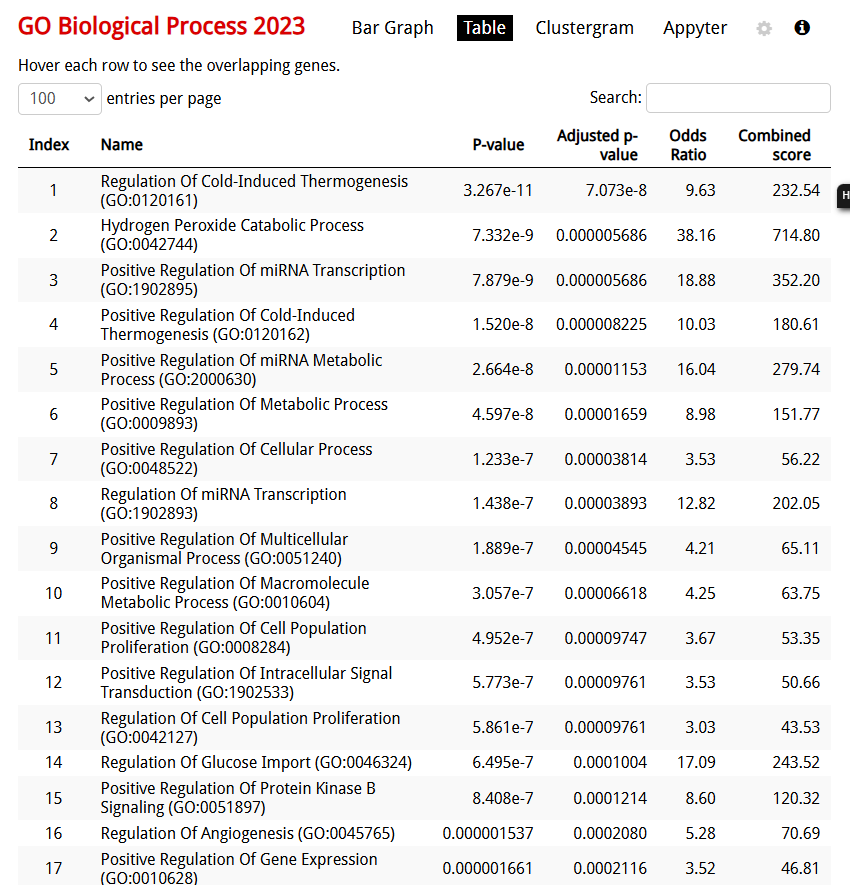


1. **تحلیل عملکرد ژنها در GO :**

**دیتاست GO را بررسی کنید و ژنهای معنادار را در سه دسته زیر تحل یل کنید. نتایج را با استفاده از نمودار و جداول مناسب نشان دهید.**

* **فرایندها ی زیستی)Process Biological): بررسی کنید کدام فرایندهای زیستی معنادار هستند و نقش آنها در سرطان را توضیح دهید.**





در اینجا 6 فرایند اول را به عنوان significant در نظر میگیریم

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | Regulation Of Cold-Induced Thermogenesis (GO:0120161) | 3.267e-11 | 7.073e-8 | 9.63 | 232.54 |
| 2 | Hydrogen Peroxide Catabolic Process (GO:0042744) | 7.332e-9 | 0.000005686 | 38.16 | 714.80 |
| 3 | Positive Regulation Of miRNA Transcription (GO:1902895) | 7.879e-9 | 0.000005686 | 18.88 | 352.20 |
| 4 | Positive Regulation Of Cold-Induced Thermogenesis (GO:0120162) | 1.520e-8 | 0.000008225 | 10.03 | 180.61 |
| 5 | Positive Regulation Of miRNA Metabolic Process (GO:2000630) | 2.664e-8 | 0.00001153 | 16.04 | 279.74 |
| 6 | Positive Regulation Of Metabolic Process (GO:0009893) | 4.597e-8 | 0.00001659 | 8.98 | 151.77 |

### Regulation Of Cold-Induced Thermogenesis

این فرآیند مربوط به تولید گرما در واکنش به سرما از طریق فعالیت میتوکندری‌ها است. در سرطان سینه، تغییرات متابولیسم انرژی و فعالیت میتوکندریایی ممکن است بقای سلول‌های سرطانی را در شرایط استرس‌زا، مانند کمبود اکسیژن یا مواد مغذی، تسهیل کند.

### Hydrogen Peroxide Catabolic Process

این فرآیند شامل تجزیه هیدروژن پراکسید (H₂O₂)، یکی از گونه‌های فعال اکسیژن (ROS)، به ترکیبات بی‌ضرر است. ROS می‌تواند نقش تنظیمی داشته باشد، اما تولید بیش‌ازحد آن باعث آسیب DNA و جهش‌های سرطانی می‌شود. در سرطان سینه، سلول‌های سرطانی این مسیر را فعال نگه می‌دارند تا تعادل اکسیداتیو خود را حفظ کرده و از مرگ سلولی جلوگیری کنند.

### Positive Regulation Of miRNA Transcription

این فرآیند به افزایش بیان miRNA‌ها می‌پردازد که نقش‌های تنظیمی مهمی در بیان ژن‌ها دارند. در سرطان سینه، تغییرات در تنظیم miRNA‌ها می‌تواند به عدم تعادل در بیان ژن‌های مرتبط با رشد سلولی، متاستاز و بقا منجر شود و نقش کلیدی در پیشرفت بیماری داشته باشد.

### Positive Regulation Of Cold-Induced Thermogenesis

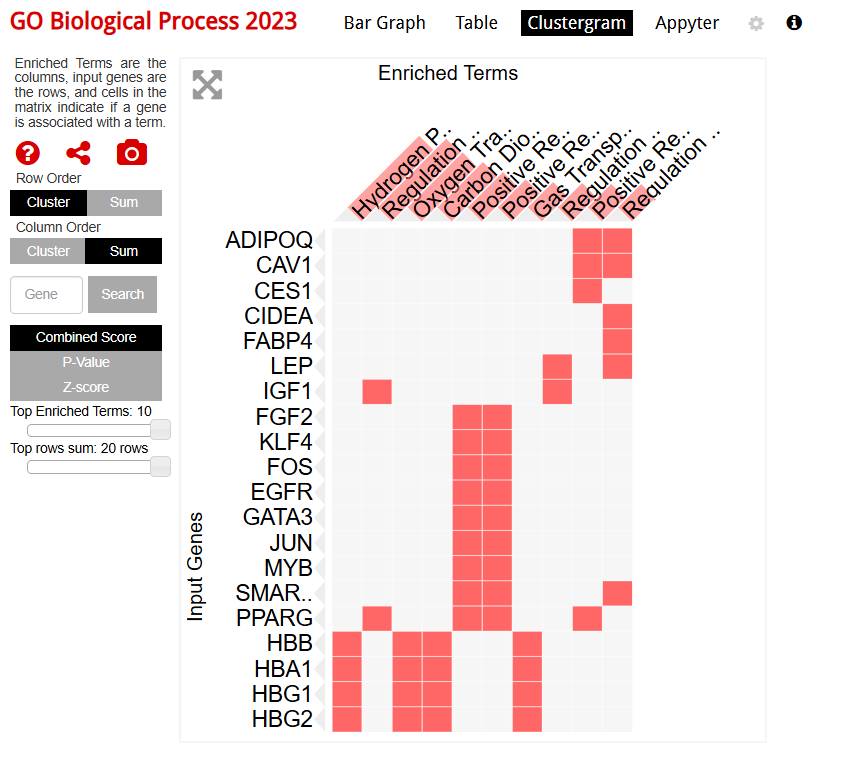
این مسیر مشابه مورد اول است و به تحریک ترموژنز ناشی از سرما مرتبط است. در سرطان سینه، تنظیم انرژی و متابولیسم سلولی که از ویژگی‌های ترموژنز است، می‌تواند به تطابق سلول‌های سرطانی با محیط‌های کم‌انرژی کمک کرده و بقا و رشد آنها را تسریع کند.

### Positive Regulation Of miRNA Metabolic Process

این فرآیند به افزایش متابولیسم miRNA‌ها، شامل پردازش و فعالیت آنها، می‌پردازد. در سرطان سینه، تغییرات در متابولیسم miRNA‌ها می‌تواند بر بیان ژن‌های دخیل در تکثیر، مهاجرت و متاستاز سلولی اثر گذاشته و به پیشرفت سرطان کمک کند.

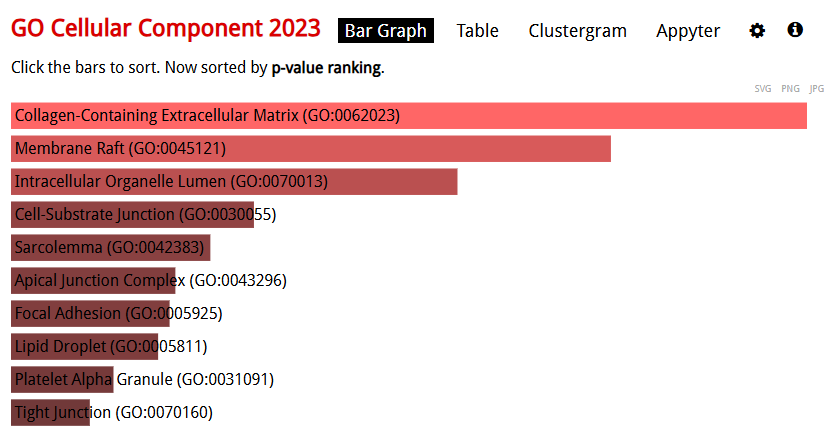
### Positive Regulation Of Metabolic Process

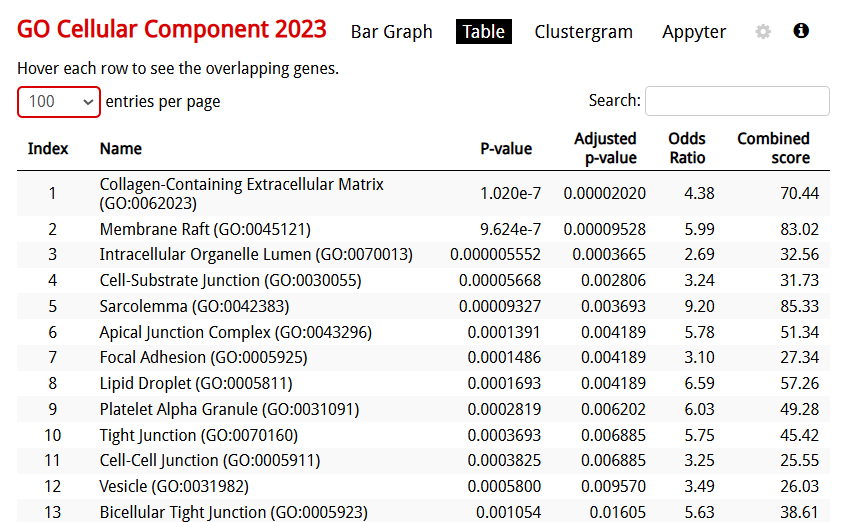
این فرآیند شامل تحریک متابولیسم سلولی است که در سرطان‌ها به‌طور قابل‌توجهی بازبرنامه‌ریزی می‌شود. در سرطان سینه، سلول‌های سرطانی از متابولیسم گلیکولیتیک (اثر واربورگ) برای تولید سریع انرژی و زیست‌مولکول‌های مورد نیاز برای تکثیر و رشد استفاده می‌کنند.



ژن های مربوط به هر کدام از این فرایند های زیستی هم در clustergram بالا مشخص است.

* **اجزای سلولی )Component Cellular): مشخص کنید ژنهای معنادار در کدام بخش های سلول فعال هستند و اهمیت این بخش ها را توضیح دهید.**





در اینجا هم در صورتی که p-value های کمتر از 0.001 را در نظر بگیریم 12 بخش سلولی فعال سلولی پیدا میشود.

به عنوان نمونه سه مورد از این بخش های سلولی را توضیح میدهیم:

تحلیل **اجزای سلولی** (Cellular Component) به شناسایی بخش‌هایی از سلول می‌پردازد که ژن‌های موردنظر در آن فعال هستند. در تصویر ارائه‌شده، داده‌ها از طریق آنالیز GO (Gene Ontology) مشخص کرده‌اند که برخی از اجزای سلولی به طور معناداری با ژن‌های شما مرتبط هستند. در ادامه برخی از مهم‌ترین بخش‌ها و اهمیت آن‌ها آورده شده است:

### Collagen-Containing Extracellular Matrix

### ماتریکس خارج‌سلولی حاوی کلاژن یکی از اجزای مهم محیط خارج سلولی است که نقش اصلی آن فراهم کردن حمایت ساختاری برای سلول‌ها و تنظیم سیگنال‌دهی سلولی است. این بخش در فرایندهای حیاتی مانند بازسازی بافت، حفظ پایداری سلول‌ها، و مهاجرت سلولی نقش اساسی دارد. تغییرات در این ساختار معمولاً به بیماری‌هایی نظیر سرطان و فیبروز مرتبط است و مطالعه آن می‌تواند به درک عمیق‌تر بیماری‌های بافتی کمک کند.

### Membrane Raft

### رفت‌های غشایی، میکرو دامنه‌هایی در غشای سلولی هستند که غنی از لیپیدها و پروتئین‌های خاص می‌باشند. این نواحی به دلیل نقششان در انتقال سیگنال‌های سلولی، سازمان‌دهی پروتئین‌های غشایی، و تقسیم سلولی بسیار مهم هستند. رفت‌های غشایی در بسیاری از فرایندهای فیزیولوژیکی و بیماری‌های مرتبط با اختلالات سیگنال‌دهی، از جمله سرطان و بیماری‌های ایمنی، نقشی کلیدی ایفا می‌کنند.

### Intracellular Organelle Lumen

### حفره اندامک‌های داخل سلولی شامل فضای داخلی اندامک‌هایی مانند میتوکندری، هسته و شبکه اندوپلاسمی است که محل انجام فرایندهای ضروری سلولی نظیر ترجمه، تاخوردگی پروتئین، و متابولیسم انرژی می‌باشد. این فضاها به دلیل نقشی که در عملکرد صحیح سلولی و تنظیم پاسخ‌های استرسی دارند، از اهمیت بالایی برخوردارند. اختلال در عملکرد آن‌ها ممکن است منجر به بیماری‌های متابولیکی و ژنتیکی شود.

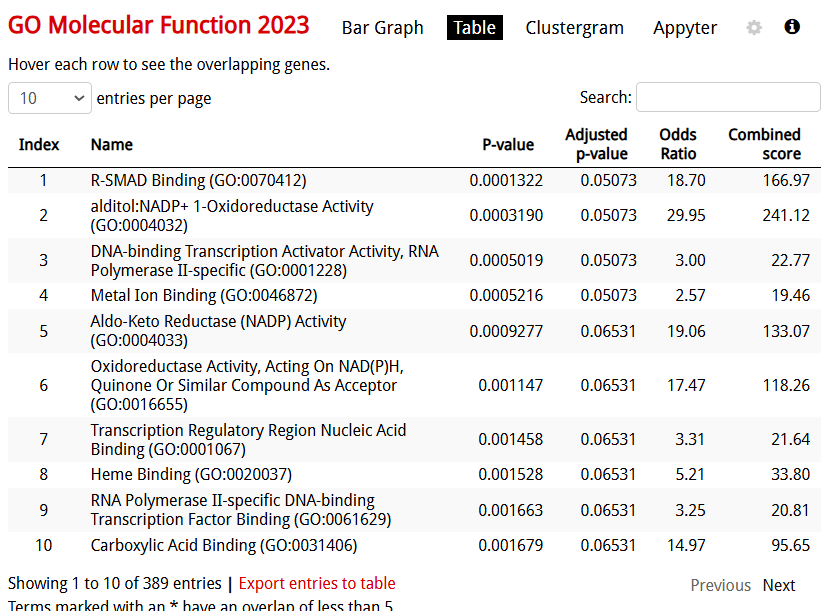


مجددا ژن های مرتبط هم در نمودار بالا مشخص شده اند.

### 

* **عملکردهای مولکولی )Function Molecular): عملکردهای مولکولی معنادار را بررسی کرده و ارتباط آن ها با فرایندها ی سرطان زا را توضیح دهید.**





اگر p-value های کمتر از 0.001 را در نظر بگیریم 5 molecular function معنادار پیدا میشود.

در این جدول، عملکردهای مولکولی معنادار (GO Molecular Functions) شناسایی شده‌اند و ارتباط هر عملکرد با فرایندهای بیولوژیکی، از جمله سرطان‌زایی، قابل بررسی است. در ادامه هر عملکرد به همراه ارتباط آن با سرطان توضیح داده شده است:

### R-SMAD Binding

این عملکرد شامل اتصال به پروتئین‌های SMAD است که بخشی از مسیر سیگنال‌دهی TGF-β محسوب می‌شوند. مسیر TGF-β در تنظیم رشد سلول و تمایز نقش دارد و تغییرات آن به رشد کنترل‌نشده سلول‌ها در سرطان منجر می‌شود. افزایش یا کاهش اتصال R-SMAD می‌تواند در فرایندهای متاستاز و سرکوب ایمنی در سرطان تأثیرگذار باشد.

### alditol:NADP+ 1-Oxidoreductase Activity

این عملکرد مرتبط با مسیرهای متابولیکی است که در تبدیل قندها نقش دارند. تغییرات متابولیکی یکی از ویژگی‌های اصلی سلول‌های سرطانی است که انرژی موردنیاز برای رشد سریع آن‌ها را تأمین می‌کند. فعالیت غیرطبیعی این آنزیم‌ها می‌تواند در مقاومت سلولی نسبت به درمان‌های متابولیکی دخیل باشد.

### DNA-binding Transcription Activator Activity, RNA Polymerase II-specific

این عملکرد شامل تنظیم بیان ژن‌ها از طریق RNA Polymerase II است. در سرطان، این فرایند ممکن است منجر به افزایش بیان ژن‌های مرتبط با بقا، تکثیر، و مقاومت به دارو شود. تنظیم نابجای عوامل رونویسی نقش مهمی در سرطان‌زایی دارد.

### Metal Ion Binding

اتصال به یون‌های فلزی در تنظیم فعالیت‌های آنزیمی و انتقال الکترون‌ها ضروری است. برخی یون‌ها مانند آهن و مس در سرطان می‌توانند نقش دوگانه‌ای داشته باشند: تسهیل استرس اکسیداتیو که موجب تخریب سلول‌های سالم می‌شود، یا حمایت از رشد سلول‌های سرطانی.

### 

### Aldo-Keto Reductase (NADP) Activity

این آنزیم‌ها در متابولیسم لیپیدها و کربوهیدرات‌ها نقش دارند. در سرطان، آن‌ها می‌توانند با تنظیم مقاومت سلول به استرس و حمایت از رشد تومور در محیط‌های پرتنش نقش داشته باشند. افزایش فعالیت این آنزیم‌ها ممکن است مقاومت به شیمی‌درمانی را تقویت کند.



در انتها نیز ژن های مرتبط با این عملکردهای مولکولی مشخص شده اند.