

آنالیز شبکه هم بیانی وزن دار هیپاتوسلولار کارسینوما پروتئین BDH2 را به عنوان یک بیومارکر تشخیصی و تارگت دارویی

معرفی می کند

آی سان نقابی

ابوالحسن بهاری

امیر مهدی محمودی

چکیده:

کارسینومای هیپاتوسلولار کبد (LIHC) یکی از انواع بدخیم سرطان است که میزان بروز بالایی داشته و پیش‌آگهی نامطلوبی دارد. این مطالعه با هدف بررسی ارتباط ژن‌های کلیدی (hub genes) با پیشرفت LIHC و اهداف درمانی بالقوه انجام شد. برای این منظور، از داده‌های پایگاه The Cancer Genome Atlas مربوطه (TCGS-LIHC) و GTex و پروفایل‌های بیان ژنی GSE54236 از پایگاه Gene Expression Omnibus (GEO) استفاده شد تا ژن‌های با بیان افتراقی بین بافت‌های سرطانی و نرمال شناسایی شوند.

ژن‌های با بیان افتراقی با استفاده از روش‌های تحلیل افتراقی بیان ژن (Differential Gene Expression Analysis) استخراج شدند.

در مجموع، ۶۸ ژن با بیان افتراقی شناسایی شدند که به‌طور عمده در فعال‌سازی کمپلمان (فرایند زیستی)، تریمرهای کلاژنی (بخش سلولی)، اتصال به کربوهیدرات و فعالیت لیگاند‌های گیرنده (عملکرد مولکولی) و مسیر برهم‌کنش سایتوکاین-گیرنده سایتوکاین نقش داشتند.

مقدمه:

کارسینومای هیپاتوسلولار (LIHC) یکی از شایع‌ترین سرطان‌ها در سراسر جهان است و از نظر بروز در رتبه ششم و از نظر مرگومیر مرتبط با سرطان در رتبه سوم قرار دارد. بر اساس آمار جهانی سرطان، سرطان اولیه کبد در سال ۲۰۲۰ ششمین سرطان شایع و سومین علت مرگومیر ناشی از سرطان بوده است، به‌طوری‌که LIHC حدود ۷۵ تا ۸۵ درصد از کل موارد سرطان کبد را شامل می‌شود. عوامل خطر کلیدی در ایجاد این بیماری شامل عفونت‌های مزمن با ویروس هپاتیت B یا C، مصرف مواد غذایی آلوده به آفلاتوکسین، مصرف بیش از حد الکل، چاقی، دیابت نوع ۲ و سیگار کشیدن هستند. روش‌های درمانی رایج برای LIHC شامل جراحی برداشت تومور، پیوند کبد و روش‌های تخریبی است. با وجود پیشرفت در روش‌های تشخیصی و درمانی، پیش‌آگهی این بیماری همچنان ضعیف است که عمدتاً به دلیل نرخ بالای عود و متاستاز می‌باشد. یکی از چالش‌های اصلی در درمان این بیماری، نبود یک سیستم دقیق برای پیش‌بینی پیش‌آگهی و روش مؤثر برای تشخیص زودهنگام است. از این رو، ارزیابی نشانگرهای زیستی بالقوه برای تشخیص و درمان LIHC از اهمیت بالایی برخوردار است.

در این مطالعه، ما به‌طور نظام‌مند ژن‌های مرکزی کلیدی درگیر در LIHC را شناسایی کرده‌ایم و چارچوبی جامع برای تحلیل ژن‌های با بیان هم‌زمان افتراقی (differentially co-expressed genes) ارائه داده‌ایم. این رویکرد، درک ما را از علل زیربنایی و مکانیسم‌های مولکولی پیشرفت LIHC بهبود می‌بخشد.

داده‌ها

برای شناسایی ژن‌های با بیان افتراقی بین LIHC و بافت‌های نرمال، داده‌های پروفایل بیان ژن از سه منبع جمع‌آوری شد: پایگاه TCGA شامل ۳۶۹ نمونه بافت توموری اولیه (portal.gdc.cancer.gov/projects/TCGA-LIHC) همراه با متادیتای مربوطه و ماتریکس کلینیکال

پایگاه GEO مجموعه داده GSE54236 شامل ۷۷ نمونه مجاور غیرتوموری و ۷۸ نمونه LIHC

((www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54236))

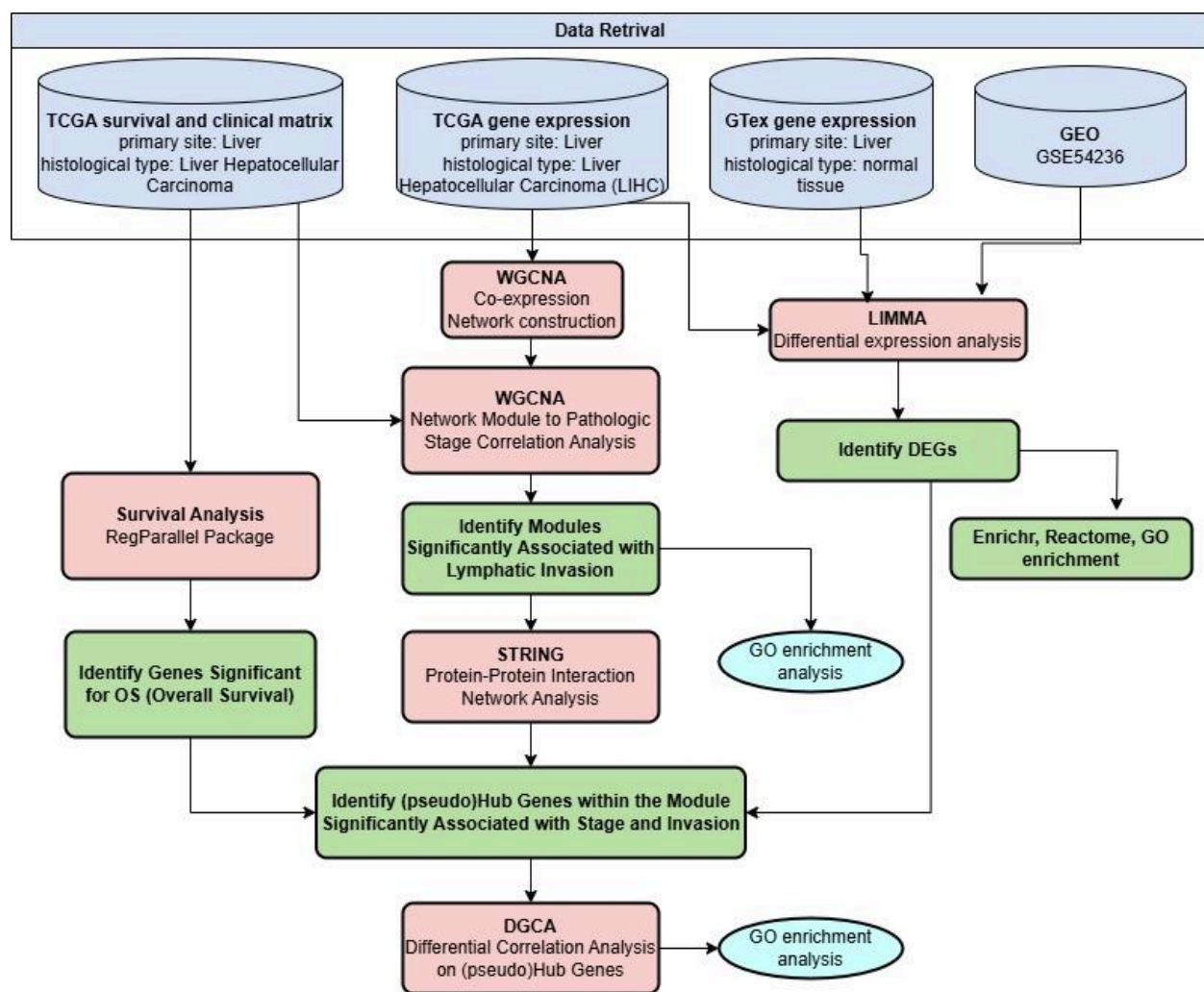
پایگاه GTex شامل ۱۱۰ نمونه بافت کبد افراد سالم

داده های TCGA و GTex با استفاده از پکیج UCSCXenaTools از دیتابیس Xena Hubs دانلود شد.

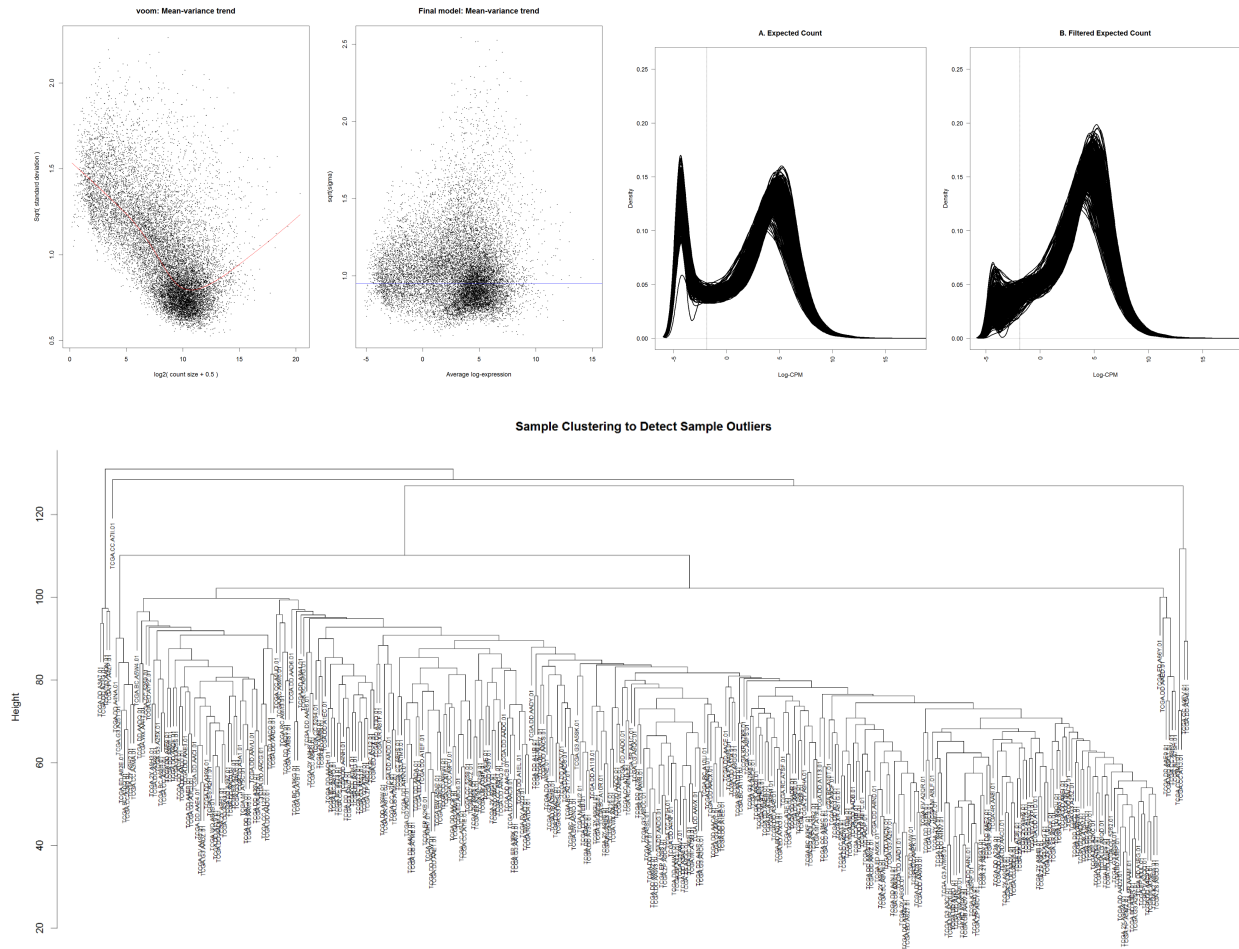
روش کار

شناسایی ژن‌های با بیان افتراقی (DEGs):

برای شناسایی ژن‌های با بیان افتراقی (DEGs) در مجموعه داده‌های TCGA-LIHC و GSE54236، از پکیج limma در R استفاده شد. معیارهای انتخاب ژن‌های افتراقی به‌صورت زیر تعیین شدند: ژن‌های دارای بیان متفاوت در افراد سالم و بیمار در دیتاست GSE54236 براساس $|\log FC| > 1$ و مقدار $P\text{-value} < 0.05$ فیلتر شدند. برای ژن‌های متفاوت بیان شده در دیتای TCGA نسبت به GTex، $\log FC \neq 0$ و کات آف $P\text{-value}$ برابر با ۰.۰۵ در نظر گرفته شد. برای نمایش بصری ژن‌های افتراقی، از پکیج ggplot2 در R استفاده شده و نمودار آتشفشانی (Volcano Plot) ترسیم شد.

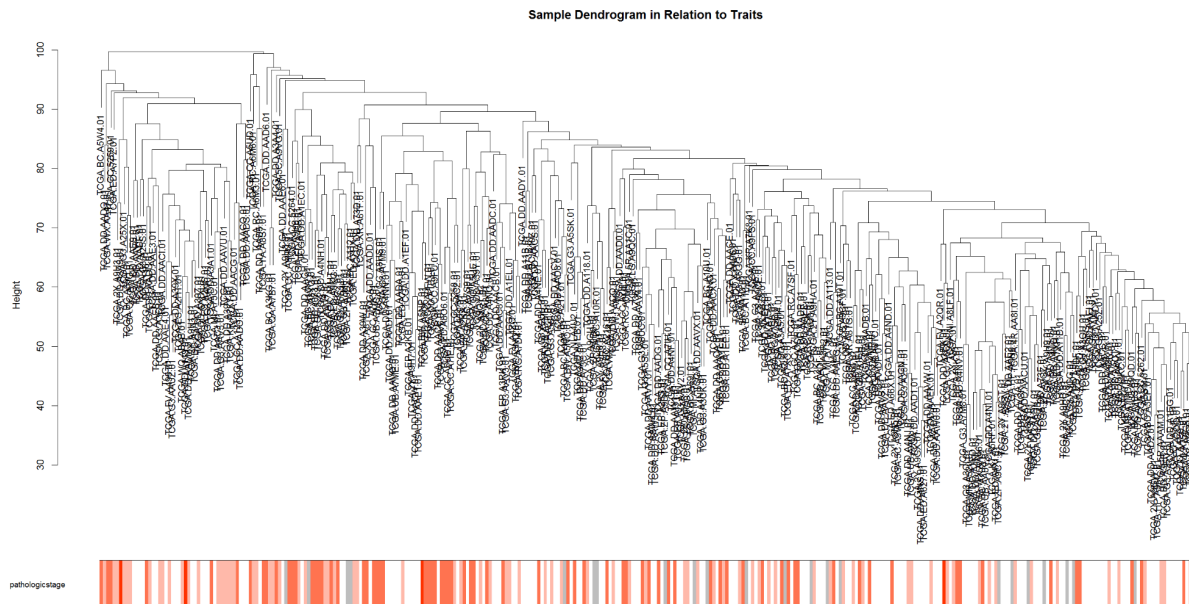


آنالیز داده‌های TCGA و GTex و تحلیل شبکه هم‌بیانی وزن دار ژن‌ها (WGCNA)
 با توجه به این که stage نشان دهنده‌ی میزان تهاجمی بودن و متاستاز سرطان است، ویژگی pathologicstage از ماتریکس کلینیکال دیتای TCGA جهت بررسی ارتباط آن با هم‌بیانی ژن‌ها انتخاب شد. طی پیش‌پردازش ماتریکس کلینیکال، استیج‌های I، II، III و IV به ترتیب به اعداد ۱، ۲، ۳ و ۴ تبدیل شد. ساخت و تحلیل شبکه هم‌بیانی وزن دار ژن‌ها با استفاده از پکیج WGCNA و dplyr انجام شد. نمونه‌ها براساس فاصله‌ی اقلیدسی دسته‌بندی شدند و نمونه‌های Outlier با ارتفاع شاخه بیش از ۱۰۰ حذف شدند.

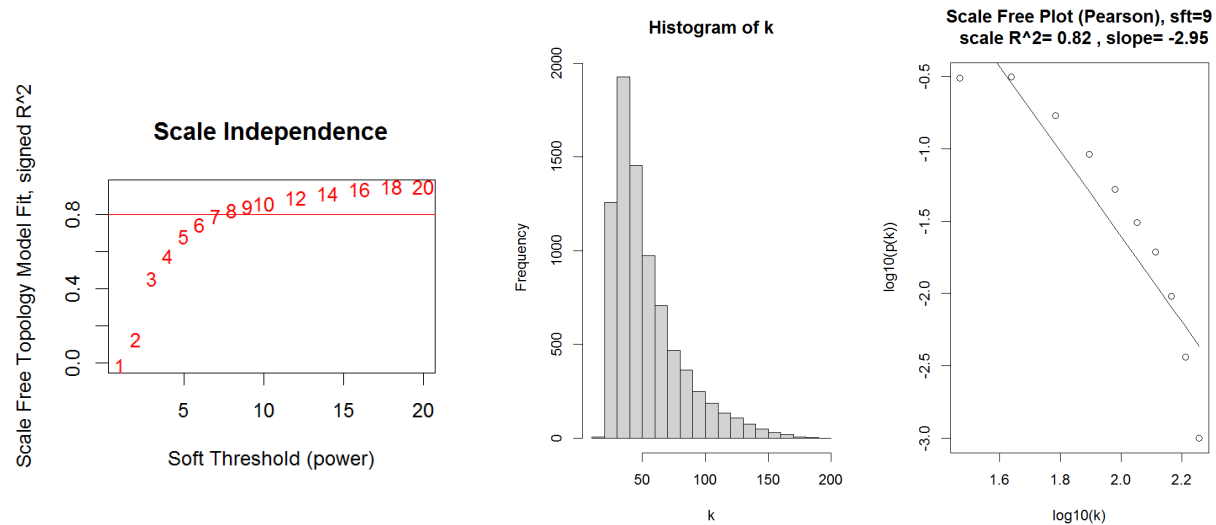


نمودار بالا تمامی سمپل ها را نشان می دهد. سمپل هایی که ارتفاع بالاتر از ۱۰۰ داشتند، حذف شدند. نمودار پایین، علاوه بر دندروگرام سمپل های مناسب، مقادیر stage سرطان را در ارتباط با سمپل ها نشان می دهد. رنگ قرمز به معنی ارتباط زیاد و

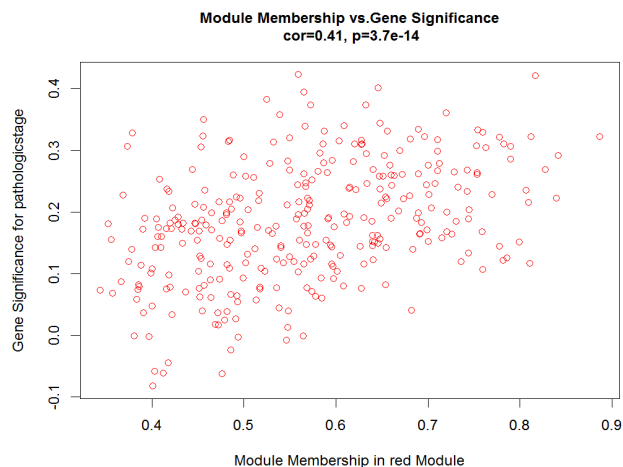
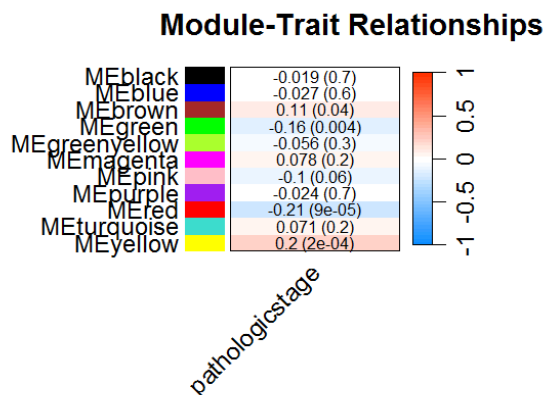
سفید به معنای ارتباط کم است. خاکستری به معنای عدم وجود اطلاعات می باشد.



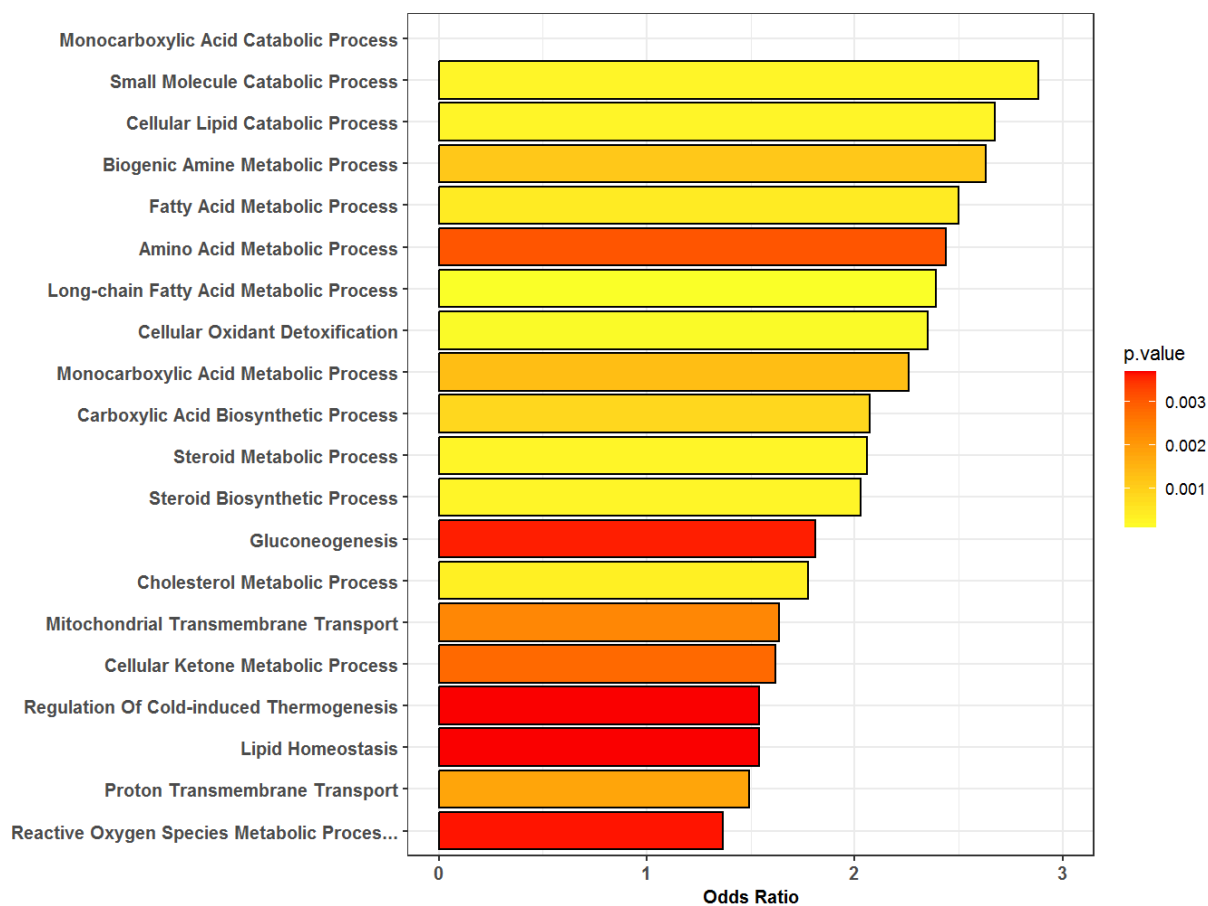
scale تعیین شد و حداقل مقداری که scale-free topology fit براساس soft-thresholding power ضریب بتا یا همان free topology fit را بیش از ۰.۸ کند به عنوان ضریب بتا در نظر گرفته شد



ماژول های ژنی براساس ضریب بتا شناسایی شدند و ارتباط آنها با stage سرطان کبد و همچنین معنی دار بودن این ارتباط بررسی شد. ماژول قرمز به دلیل داشتن correlation منفی کاملاً معنی دار انتخاب شد. نمودار زیر ارتباط بین اهمیت ژن ها در استیج بیماری و عضویت آنها در ماژول قرمز را نشان می دهد. همان طور که می بینید بین عضویت در ماژول قرمز و اهمیت ژن در استیج بیماری ارتباط مشخصی وجود دارد. آنالیز Gene Ontology Enrichment برای ژن های ماژول قرمز انجام شد.



Enriched GO Biological Processes



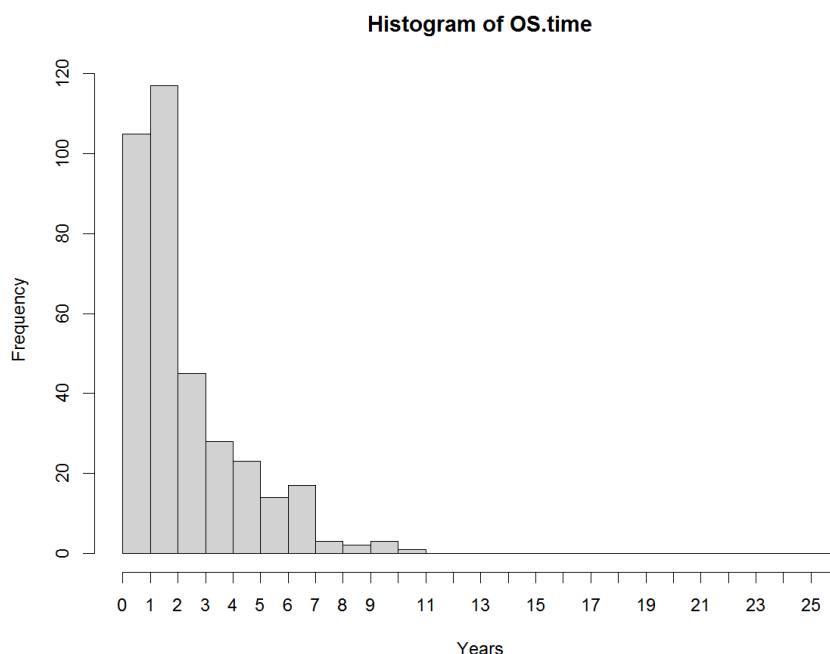
آنالیز بقا

با استفاده از پکیج RegParallel وجود دیتای بقا برای بازه های زمانی مختلف در سمل ها بررسی شد. مدل بقا برای همه ی ژن های ماژول قرمز به روش رگرسیون cox ایجاد شد. سپس مهمترین ژن های ماژول قرمز براساس سه حیطه تعیین شدند:

۱- مرتبط بودن با ویژگی بالینی مورد بررسی (stage)

۲- تفاوت بیان معنی دار

۳- اثر معنی دار بر بقا
بر این اساس، ۳۸ ژن در صورت کاهش بیان در بیماران، اثر منفی معنی داری بر بقا داشتند.



مصورسازی شبکه برهمکنش پروتئین ها (PPI) و غربالگری برای یافتن hub gene ها
شبکه ساخته شده را وارد نرم افزار cytoscape کردیم و پس از دانلود دیتای STRING برای پروتئین ها، شبکه را stringify کردیم. در شبکه ی ساخته شده، hub genes را براساس معیارهای زیر مشخص کردیم. ابتدا در Excel، ژن ها را براساس degree از بیشترین به کمترین مرتب کردیم. سپس ژن های دارای degree > 10 که adjusted p-value ی مناسبی در آنالیز بقا داشتند و LogFC آن ها مخالف صفر و بزرگتر از ۰.۵+ یا کوچکتر از -۰.۵ بود انتخاب شدند.

شناسایی تغییرات در همبستگی ژن ها به روش Differential Gene Correlation Analysis
در این مرحله با استفاده از پکیج DGCA، تغییرات همبستگی بیان ژن ها سنجیده شد. همچنین، برای ژن BDH2 که یکی از Hub node های اصلی بود، تغییرات همبستگی بیان ژن ها و تقویت مسیرهای زیستی بررسی شد.

نتایج:

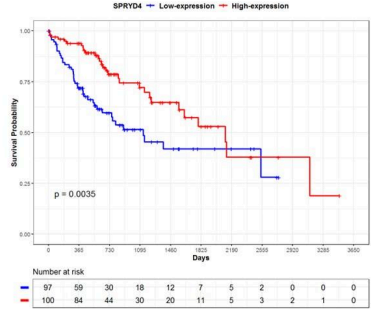
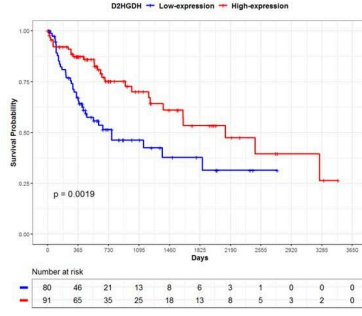
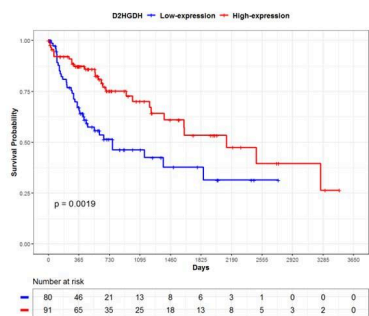
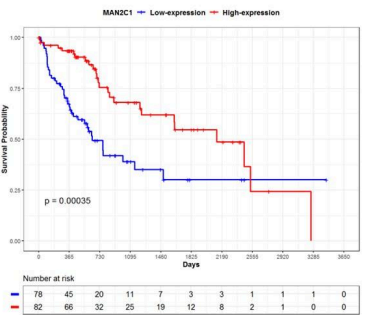
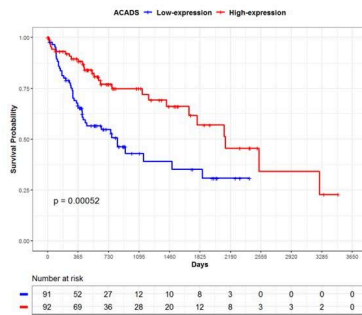
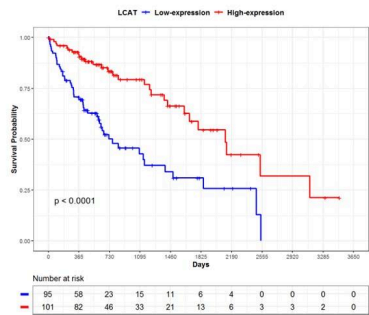
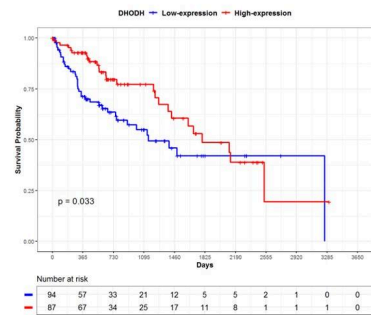
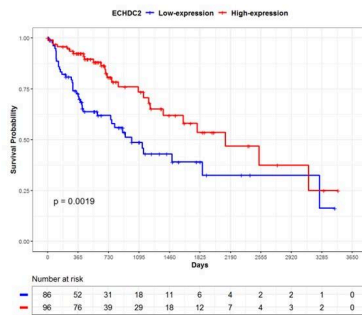
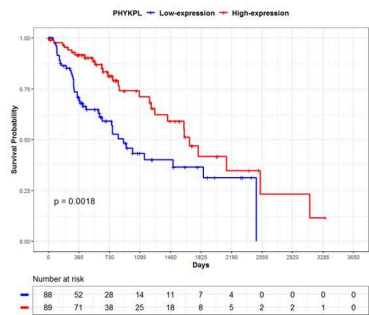
مهمترین ژن های down-regulated موثر در بقای بیمار و تهاجم سرطان در ماژول بررسی شده عبارتند از:

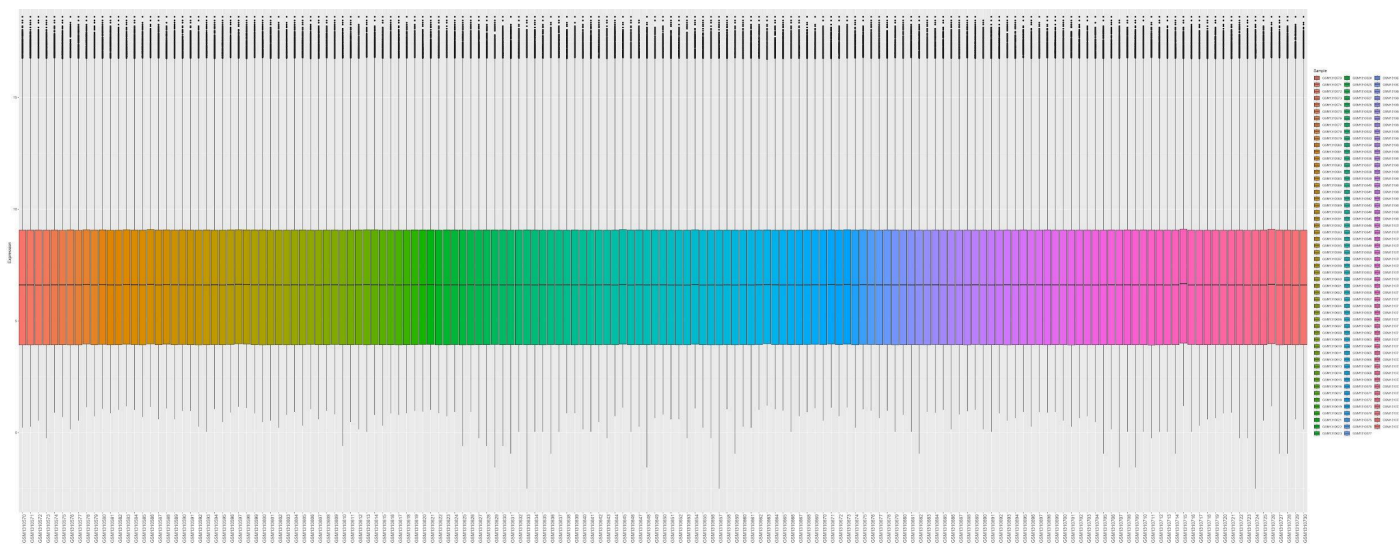
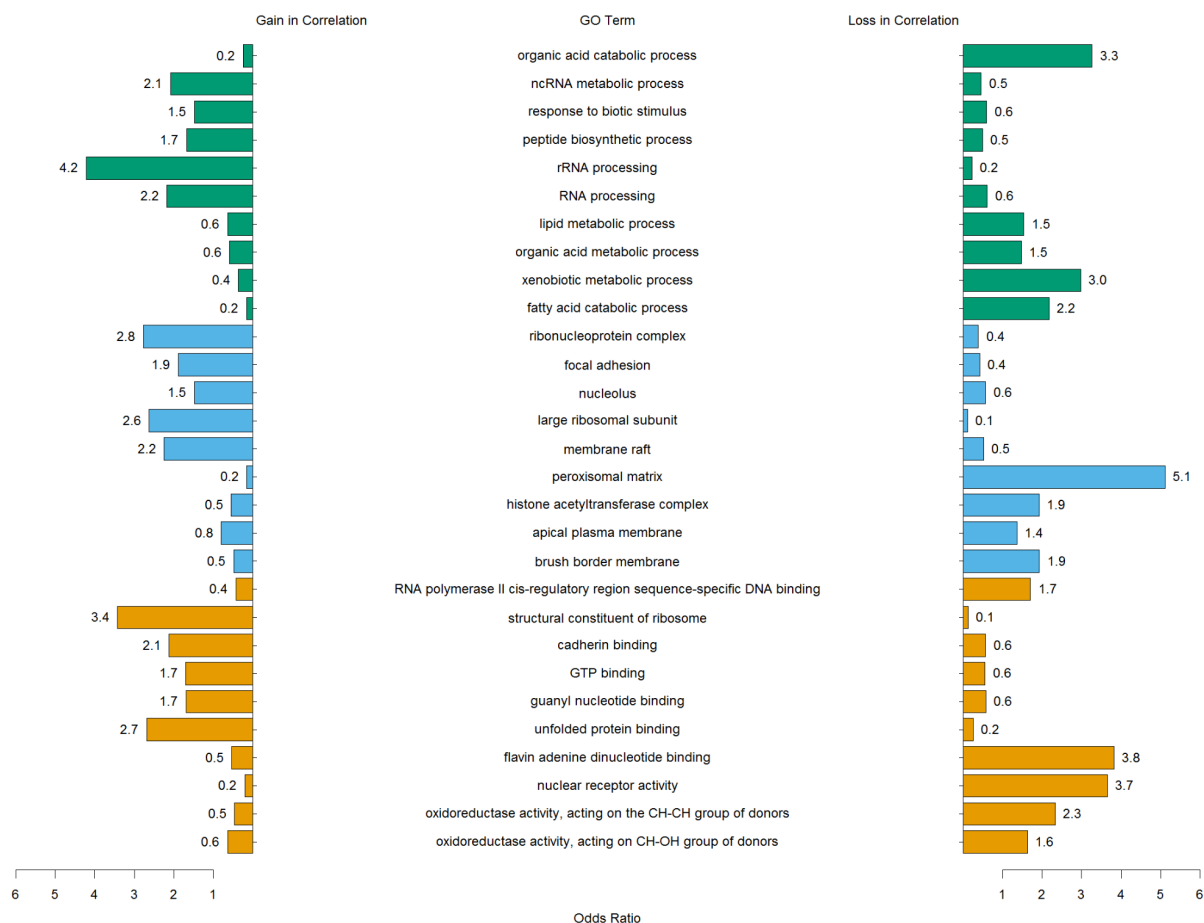
PHYKPL
ECHDC2
DHODH
LCAT
MST1
BDH2
ACADS
FAHD2A
CBR1
EPHX1

براساس غربالگری شبکه و بررسی هاب نود های بدست آمده، بهترین ژن ها عبارتند از:

ACAA2
ACAA1
ACADS
PC
ALDH4A1
ALDH7A1
IVD
UQCRC1
APOE
ECHDC2
SHMT1
DECR2
FDXR
ALDH1A1
FBP1
DHODH
BDH2
MLYCD
ITIH4

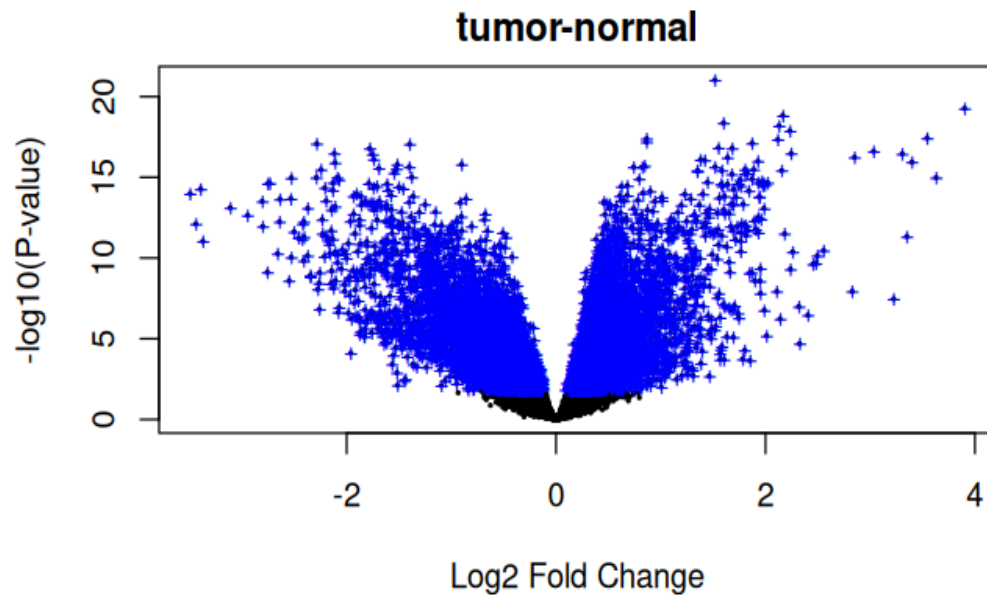
پروتئین BDH2 به عنوان یک Tumor suppressor عمل می کند. مطالعات قبلی نشان داده است که این پروتئین، اتوفازی و ROS-induced cell death را تحریک می کند. همچنین پژوهش های اخیر نشان داده است که غیر فعال شدن این پروتئین باعث تکثیر و متاستاز نازوفانگال کارسینوما می شود. از این رو این پروتئین می تواند مورد بررسی بیشتر از جهت هدف قرار گرفتن به عنوان بیومارکر تشخیصی و یا تارگت دارویی قرار گیرد. همچنین سایر نتایج این پژوهش نیز می تواند درک ما از مکانیسم مولکولی سرطان کبد را افزایش دهد و زمینه را برای درمان های هدفمند و موثرتر فراهم کند.





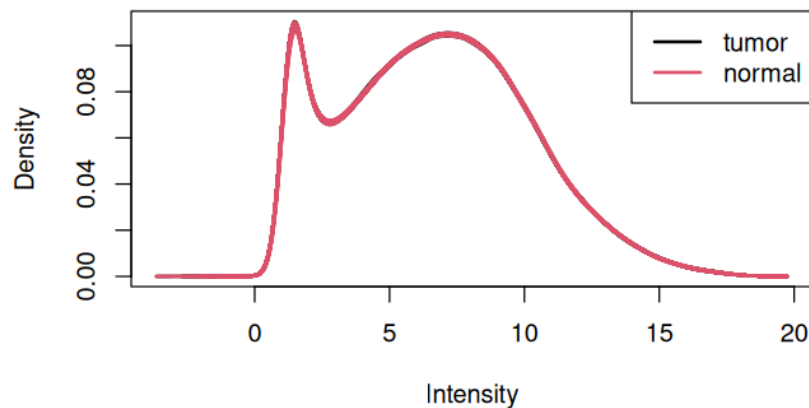
توزیع نمونه‌های GSE54236

داده‌ها به کمک تکنولوژی میکروآرای و با استفاده از پلتفرم GPL6480 جمع‌آوری شده است. نمونه‌ها از بافت کارسینومای کبدی و همچنین بافت غیرتوموری از کبد همان بیماران گرفته شده‌اند. همانطور که نشان داده شده است، چارکهای داده‌های مسئله با هم برابر بوده و داده‌ها اصلاح شده و نرمال‌سازی شده هستند.

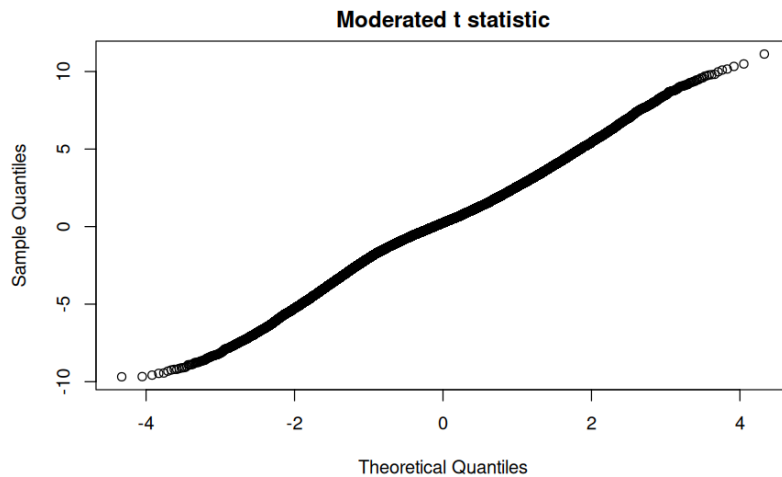


نمودار آتش فشان برای نمایش منهای لگاریتم پی-مقدار در مبنای ۱۰ نسبت به مقدار تغییرات لگاریتم بیان ژن‌ها، استفاده می‌کنیم. در این نمودار ژن‌هایی مهم هستند، که پی-مقدار در آنها کم و تغییرات ژن‌ها در دو دسته زیاد باشد، بنابراین هرچه ژن به سمت بالا-راست یا بالا-چپ نمودار نزدیکتر باشد، اهمیت بیشتری می‌یابد.

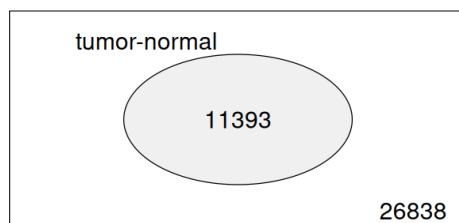
GSE54236/GPL6480 value distribution



این نمودار نشان‌دهنده توزیع شدت سیگنال‌های بیان ژن در نمونه‌های توموری و نرمال است. منحنی‌های قرمز و مشکی، توزیع مقادیر بیان ژن در این دو گروه را نمایش می‌دهند. یکنواخت بودن توزیع‌ها و همپوشانی آن‌ها نشان می‌دهد که داده‌ها به درستی نرمال‌سازی شده‌اند.

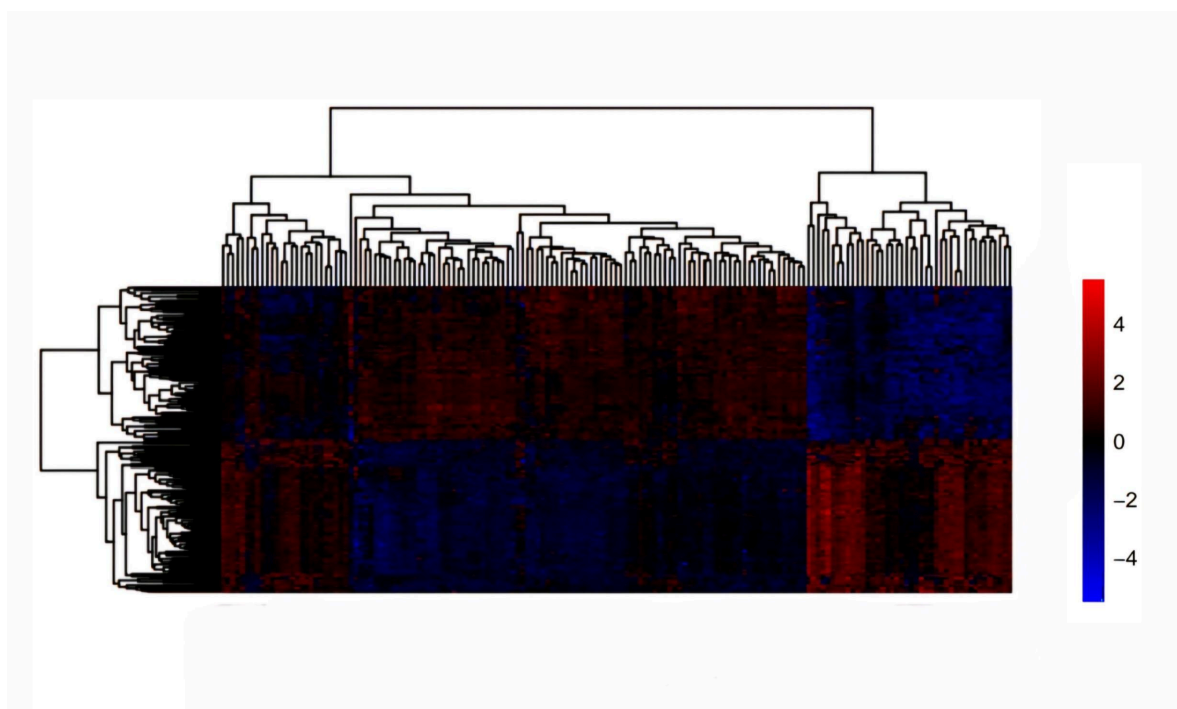


این نمودار Q-Q plot است که توزیع آماره t تعدیل شده را در برابر کوانتایل های توزیع نظری نشان می دهد. اگر داده ها به درستی پردازش و نرمال سازی شده باشند، نقاط باید تقریباً روی خط قطری قرار گیرند.



این نمودار نشان می دهد که از مجموع ۲۶,۸۳۸ ژن، تعداد ۱۱,۳۹۳ ژن به طور مشترک بین بافت های توموری و نرمال بیان شده اند. این مقایسه به شناسایی ژن هایی که در هر دو گروه مشترک هستند یا به طور اختصاصی در یکی از آنها بیان می شوند کمک می کند.

تحلیل ژن های استخراج شده:



این نقشه حرارتی الگوی بیان ژن‌های افتراقی را در نمونه‌های توموری و نرمال نشان می‌دهد. رنگ قرمز بیان بالاتر ژن را نشان می‌دهد. رنگ آبی بیان پایین‌تر ژن را نمایش می‌دهد. همچنین، شاخه‌های درختی در بالا و سمت چپ نمودار، گروه‌بندی نمونه‌ها و ژن‌ها را بر اساس شباهت بیان آن‌ها مشخص می‌کنند. این نمودار کمک می‌کند تا تفاوت‌های بیان ژنی بین نمونه‌های LIHC و نرمال بهتر درک شود.

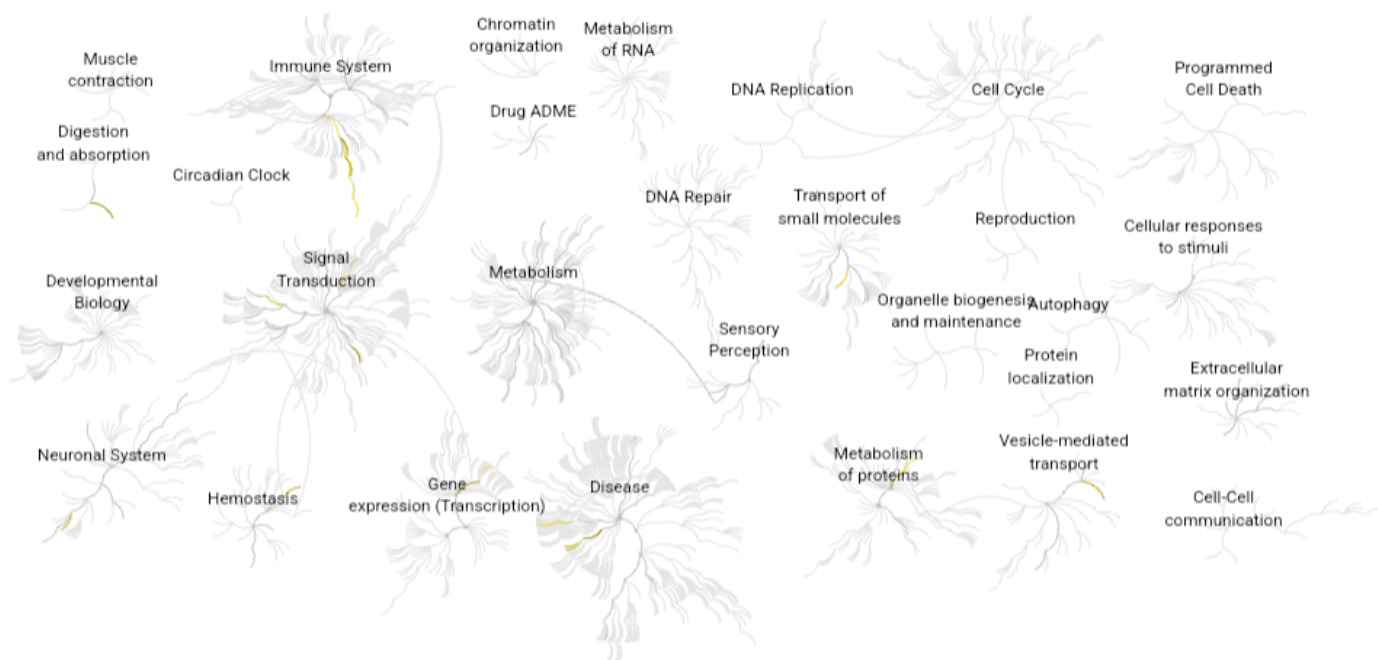
| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|--|-------------|------------------|------------|----------------|
| 1 | Lectin Pathway of Complement Activation | 4.394e-11 | 8.349e-9 | 454.77 | 10845.49 |
| 2 | Ficolins Bind to Repetitive Carbohydrate Structures on the Target Cell Surface | 1.067e-9 | 1.013e-7 | 1076.81 | 22245.45 |
| 3 | Initial Triggering of Complement | 0.000003242 | 0.0002053 | 16.86 | 213.06 |
| 4 | Complement Cascade | 0.00001722 | 0.0008181 | 12.40 | 136.01 |
| 5 | Creation of C4 and C2 Activators | 0.00003671 | 0.001395 | 14.93 | 152.43 |
| 6 | Defective B3GALT1 Causes PpS | 0.0004363 | 0.01279 | 22.73 | 175.85 |
| 7 | O-glycosylation of TSR Domain-Containing Proteins | 0.0004713 | 0.01279 | 22.10 | 169.25 |
| 8 | Scavenging by Class A Receptors | 0.002460 | 0.05691 | 30.81 | 185.12 |
| 9 | Diseases Associated With O-glycosylation of Proteins | 0.002696 | 0.05691 | 11.68 | 69.09 |
| 10 | Class B 2 (Secretin Family Receptors) | 0.005930 | 0.1127 | 8.72 | 44.70 |

نتیجه فرایندهای زیستی حاصل از ژن‌های استخراج شده از بافت توموری نسبت به بافت مجاور غیرتوموری.

کارسینوم هپاتوسلولار کبد (LIHC) ممکن است با تأثیرگذاری بر مسیر لکتین فعال‌سازی کمپلمان، در پیشرفت بیماری نقش داشته باشد. در شرایط طبیعی، این مسیر با شناسایی الگوهای قندی خاص روی سطح پاتوژن‌ها، سیستم ایمنی را فعال می‌کند. در LIHC، تغییرات در بیان ژن‌های مرتبط با این مسیر، مانند کاهش تولید لکتین متصل‌شونده به مانوز (MBL)، می‌تواند منجر به کاهش فعالیت کمپلمان و در نتیجه، کاهش حذف سلول‌های سرطانی شود. این اختلال ممکن است به رشد و گسترش تومور کمک کند.

مطالعات نشان داده‌اند که تغییرات در مسیر لکتین فعال‌سازی کمپلمان می‌تواند با پیشرفت سرطان کبد مرتبط باشد:

<https://doi.org/10.3389/fimmu.2022.1007382>



این نقشه از Reactome نشان می‌دهد که LIHC مسیرهای کلیدی زیستی را تحت تأثیر قرار می‌دهد، از جمله: مسیرهای ایمنی و کمپلمان (مثل Lectin Pathway of Complement Activation) که ممکن است منجر به کاهش پاسخ ایمنی ضدتوموری شوند. مسیرهای گلیکوزیلاسیون و سیگنالینگ سلولی که می‌توانند باعث رشد و متاستاز سلول‌های سرطانی شوند. مسیرهای متابولیکی که در تنظیم انرژی و بیوسنتز نقش دارند و ممکن است برای تأمین نیازهای تومور تغییر کنند.

نتیجه‌گیری:

نتایج این مطالعه نشان می‌دهد که LIHC مسیرهای زیستی مختلفی را تحت تأثیر قرار می‌دهد که می‌توانند بر پیشرفت سرطان، تنظیم سیستم ایمنی، و تغییرات متابولیکی نقش داشته باشند. فعال‌سازی مسیر لکتین کمپلمان یکی از فرآیندهای مهمی است که در LIHC مختل می‌شود، و کاهش بیان ژن‌های مرتبط با آن ممکن است کاهش پاسخ ایمنی و افزایش رشد و متاستاز سلول‌های توموری را تسهیل کند. علاوه بر این، مسیرهای O-glycosylation و متابولیسم سلولی به‌طور معناداری تغییر یافته‌اند که می‌تواند اهداف بالقوه‌ای برای درمان هدفمند ارائه دهد. در مجموع، بررسی این مسیرهای زیستی می‌تواند به بهبود تشخیص، شناسایی نشانگرهای زیستی، و توسعه روش‌های درمانی جدید برای LIHC کمک کند.

مقالات:

**Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular
Carcinoma. Gastroenterology. 2019;156(2):477-91.e1**

**Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al.
Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality
Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians.
2021;71(3):209-49**