

شناسایی ژن‌های کانونی مرتبط با پیشرفت کارسینومای غیر سلول کوچک ریه با تحلیل داده‌های ریزآرایه

ناعمه جعفرپور^۱، خدیجه صادق^۱، امیرحسین احمدی^{۱*}

(۱) گروه علوم و فناوری‌های زیستی، دانشکده‌ی علوم و فناوری‌های نانو و زیستی، دانشگاه خلیج فارس، بوشهر، ایران

چکیده:

کارسینومای سلول‌های غیر کوچک (NSCLC) شایع‌ترین نوع سرطان ریه است که سالانه منجر به مرگ حدود یک میلیون نفر در جهان می‌شود. اگرچه مبتلایان به NSCLC در مراحل ابتدایی بهتر به داروها پاسخ می‌دهند، میزان بقای کلی پنج ساله برای بیمارانی که در مراحل پیشرفته این سرطان قرار دارند ضعیف باقی مانده است. از این رو درک مکانیسم‌های مولکولی از جمله تغییرات بیان ژن‌ها که باعث پیشرفت NSCLC می‌شود برای شناخت اهداف درمانی جدید و افزایش بقای بیماران حائز اهمیت است. ابزارهای بیوانفورماتیک در کنار مطالعات ریزآرایه و توالی‌یابی نسل جدید نقش مهمی در شناسایی ژن‌های کانونی داشته‌اند که بیشترین تأثیر را در شبکه ژن‌های تغییر بیان یافته دارند. به این ترتیب هدف از مطالعه‌ی حاضر شناسایی ژن‌های کانونی مرتبط با پیشرفت NSCLC و بررسی ارتباط آن‌ها با میزان بقای این بیماران است.

در این مطالعه دو پروفایل بیان ژنی GSE21933 و GSE21933 از مطالعات ریزآرایه از پایگاه GEO استفاده شد. به منظور شناسایی ژن‌هایی که بیان آن‌ها بین نمونه‌های توموری و کنترل تغییر بیان یافته بود از نرم افزار R استفاده شد. تجزیه و تحلیل هستی‌شناسی و غنی شدن ژن‌ها در مسیرهای KEGG با استفاده از پایگاه GSEA انجام شد. برای یافتن ژن‌های کانونی از نرم افزار Cytoscape و افزونه Cytohubba استفاده گردید. ارتباط تغییر بیان ژن‌ها بر میزان بقای بیماران از پایگاه داده Kaplan Meier plotter استخراج شد.

نتایج مطالعه نشان داد که ژن‌های کانونی در شبکه ژن‌های تغییر بیان یافته، مختص به مراحل پیشرفته NSCLC شامل EGFR، CCT2، CCT5، PDIA6، AR، WDR12، PNPLA2، PENK، NUP37 و SNAI2 می‌باشند که به غیر از SNAI2 و PNPLA2 همگی با میزان بقای کلی بیماران ارتباط معناداری دارند.

ژن‌های کانونی مرتبط با مرحله پیشرفته NSCLC که در این مطالعه معرفی شده‌اند می‌توانند به عنوان اهدافی در جلوگیری از پیشرفت بیشتر و افزایش بقای مبتلایان به NSCLC مورد توجه قرار گیرند.

واژگان کلیدی: سرطان غیر سلول کوچک ریه، پروفایل بیان ژن، ژن‌های کانونی، بقای کل

* نویسنده مسئول:

دکتر امیرحسین احمدی، گروه علوم و فناوری‌های زیستی، دانشکده‌ی علوم و فناوری‌های نانو و زیستی، دانشگاه خلیج فارس، بوشهر، ایران. پست الکترونیک: ahahmadi@pgu.ac.ir

مقدمه:

سرطان ریه یکی از سرطان‌های رایج است که هر ساله منجر به مرگ بیش از یک میلیون نفر در سراسر دنیا می‌شود [۱]. از نظر میزان مرگ و میرهای ناشی از سرطان در ایران، سرطان ریه به ترتیب رتبه دوم و سوم را در آقایان و خانم‌ها به خود اختصاص داده است [۲]. سرطان ریه به دو نوع عمده، سرطان سلول‌های غیر کوچک ریه^۱ (NSCLC) و سرطان سلول‌های کوچک ریه^۲ (SCLC) تقسیم می‌شود. NSCLC شایع‌ترین نوع سرطان ریه است به طوری که ۸۵ درصد از موارد سرطان ریه را به خود اختصاص داده است. سلول‌های سرطانی در NSCLC نسبت به SCLC آهسته‌تر تقسیم می‌شوند و دیرتر گسترش می‌یابند [۳]. از نظر بافت‌شناسی آدنوکارسینوما شایع‌ترین نوع NSCLC به شمار می‌رود و در حدود ۵۰ درصد از کل سرطان‌های ریه را شامل می‌شود. همچنین کارسینومای سلول‌های سنگفرشی^۳ (SCC) نوع دیگری از NSCLC به شمار می‌رود [۴]. درمان‌های کلاسیک سرطان ریه شامل جراحی، شیمی درمانی و پرتودرمانی است که هرکدام متناسب با نیازها و شرایط بیمار انجام می‌شود [۵]. با وجود سال‌ها تحقیق و پیشرفت‌های چشمگیر در روش‌های تشخیص زودهنگام و درمان NSCLC، میزان نسبی بقای کلی^۴ (OS) ۵ ساله، ضعیف باقی مانده است به طوری که این میزان از ۶۸ درصد در مبتلایان در مرحله IB تا کمتر از ۱۰ درصد در مبتلایان در مرحله IV است [۶، ۷]. این آمار به ویژه از این نظر حائز اهمیت است که در حدود ۷۵ درصد از بیماران در مراحل میانی پیشرفت NSCLC شناسایی می‌شوند. به این ترتیب اگرچه شناسایی NSCLC در مراحل اولیه و درمان آن با موفقیت‌هایی همراه بوده است، همچنان نیاز به توسعه روش‌های جدید برای درمان مبتلایان به مراحل پیشرفته این سرطان وجود دارد. بنابراین شناخت مکانیسم‌های دقیق مولکولی که منجر به پیشرفت این بیماری می‌شود، برای یافتن اهداف بالقوه به منظور بکارگیری در راهبردهای درمانی موثر بر مبتلایان به مراحل پیشرفته،

بسیار مهم است [۸]. در سال‌های اخیر، با توسعه تراشه پروفایل بیان ژن و تکنولوژی‌های توالی‌یابی نسل جدید، حجم داده‌های بیان ژن‌ها در سرطان ریه بسیار گسترش یافته است که زمینه را برای مطالعه جامع ژن‌های درگیر در سرطان و یافتن ژن‌های کانونی^۵ به عنوان اهداف درمانی و یا نشانگرهای زیستی پیش‌آگهی‌دهنده فراهم می‌کند [۹]. ژن‌های کانونی ژن‌هایی هستند که همبستگی زیادی با دیگر ژن‌های موجود در شبکه ژنی دارند و درجه (تعداد اتصال آن‌ها با دیگر ژن‌های شبکه) بیشتر از ۱۰ است و بنابراین تغییر در بیان آن‌ها بیشترین اثر را می‌تواند بر روی سایر ژن‌ها به دنبال داشته باشد. تاکنون مطالعات بیوانفورماتیکی متعددی برای شناسایی ژن‌های کانونی که مرتبط با مراحل اولیه NSCLC هستند انجام شده است. با این حال شناسایی ژن‌های مرتبط با مراحل پیشرفته NSCLC تاکنون مورد بررسی قرار نگرفته است. در این مطالعه از دو پروفایل بیان ژن موجود در پایگاه GEO که دارای نمونه‌های بیماران در مراحل اولیه و پیشرفته بودند برای مقایسه و شناسایی ژن‌های مرتبط با مراحل پیشرفته NSCLC استفاده شد. همچنین شبکه برهمکنش پروتئین-پروتئین ژن‌ها^۶ (PPI) و ژن‌های کانونی و همبستگی آن‌ها بر میزان بقای بیماران با ابزار بیوانفورماتیکی ارزیابی شده است.

مواد و روش‌ها:**داده‌های ریزآرایه^۷**

برای تعیین ژن‌های مرتبط با مراحل پیشرفته NSCLC دو پروفایل بیان ژنی GSE21933 با پلت فرم Phalanx Human OneArray و GSE33532 با پلت فرم Affymetrix Human Genome U133 Plus 2.0 Array از پایگاه داده GEO به نشانی <http://www.ncbi.nlm.nih.gov/geo/> دانلود شد. نمونه‌های داده GSE21933 شامل ۱۱ بیمار مبتلا به آدنوکارسینومای ریه و ۱۰ بیمار مبتلا به سرطان سلول‌های سنگفرشی ریه به همراه نمونه‌هایی از بافت‌های سالم اطراف تومور از این بیماران بود که در دو گروه

¹ Non-small-cell lung cancer² Small-cell lung cancer³ Squamous cell carcinoma⁴ Overall Survival⁵ Hub genes⁶ Protein-Protein Interaction Network⁷ Microarray data

سطح اطمینان امتیاز ترکیبی بیش از ۰/۴ وارد نرم‌افزار Cytoscape (نسخه 3.8.0) به نشانی <http://www.cytoscape.org/> شد [۱۳] تا PPI آن رسم شود. همچنین ژن‌های مرکزی در شبکه PPI با استفاده از افزونه cytoHubba و براساس درجه شناسایی شدند [۱۴].

تجزیه و تحلیل میزان بقا ژن‌های کانونی

برای ارزیابی اهمیت پیش‌آگهی ژن‌های کانونی و بررسی تاثیر تغییر بیان ژن‌ها بر میزان بقای بیماران از پایگاه داده Kaplan Meier plotter به نشانی <http://kmplot.com/analysis/> استفاده شد [۱۵].

این پایگاه شامل داده‌های بالینی و داده‌های بیان ژن سرطان‌های مختلف از جمله NSCLC است که برای ارزیابی اهمیت پیش‌آگهی هر ژن کانونی در NSCLC مورد استفاده قرار می‌گیرد. به منظور ارزیابی بهتر در مورد ارزش پیش‌آگهی ژن‌های کانونی، نمونه‌های موجود در پایگاه با توجه به بیان میانه ژن، به دو گروه با بیان بالا (بیش از ۵۰٪) و با بیان پایین (کمتر از ۵۰٪) تقسیم شد. میزان بقای کلی (OS) بیماران مبتلا به سرطان ریه با استفاده از نمودار Kaplan Meier مورد بررسی قرار گرفت. مقادیر P-value با آزمون Log rank test و نسبت خطر^۸ (HR) با فاصله اطمینان^۹ (CI) ۹۵٪ نیز محاسبه شد.

یافته‌ها:

شناسایی ژن‌ها با بیان متمایز (DEGs)

از نمودار ون^{۱۰} برای شناسایی ژن‌های مشترک مرتبط با مراحل اولیه و پیشرفته بیماری (دو گروه مربوط به مراحل اولیه از دو مطالعه و یک گروه مربوط به مرحله پیشرفته بیماری از یک مطالعه) و همچنین ژن‌هایی که به طور اختصاصی در مرحله پیشرفته بیماری دچار تغییر بیان شده بودند استفاده شد. تغییر بیان ۳۳۸ ژن در حالت اشتراک بین مراحل اولیه و پیشرفته (۱۵۵ ژن افزایش بیان یافته و ۲۴۳ ژن کاهش بیان یافته) و ۲۷۸ ژن به طور اختصاصی در مرحله پیشرفته بیماری (شامل

مرحله اولیه^۱ (ES) بیماری (شامل مراحل اول و دوم) و مرحله پیشرفته^۲ (LS) بیماری (شامل مراحل سوم و چهارم) دسته بندی شده بودند. به این ترتیب از پروفایل بیان ژن مربوط به ۲۱ بیمار مبتلا به NSCLC (۱۰ فرد در مرحله یک و دو و ۱۱ فرد در مرحله ۳ و ۴) برای مطالعه حاضر انتخاب شد. همچنین نمونه‌های مربوط به داده GSE33532 همگی در مرحله اولیه بیماری (یک و دو) بودند و شامل ۲۰ نمونه بافت سالم اطراف تومور و ۸۰ نمونه از قسمت‌های مختلف تومور (چهار نمونه توموری از هر فرد) از ۲۰ بیمار بود.

شناسایی ژن‌ها با بیان متمایز^۳ (DEGs)

به منظور شناسایی ژن‌هایی که بیان آن‌ها بین نمونه‌های بیمار و کنترل تغییر یافته بود از نرم‌افزار R (نسخه 4.0.2) به نشانی <https://www.r-project.org> استفاده شد. داده‌های خام به کمک پکیج‌های Biobase، Limma و GEOquery مورد آنالیز قرار گرفتند. ژن‌های بدست آمده براساس $P\text{-value} < 0.05$ و $|\log_2\text{-fold change}| \geq 1.5$ جداسازی و انتخاب شدند.

تجزیه و تحلیل عملکرد و مسیر غنی‌سازی

تجزیه و تحلیل هستی‌شناسی^۴ (GO) [۱۰، ۱۱] و غنی شدن ژن‌ها^۵ در مسیرهای KEGG^۶ [۱۲]، از طریق پایگاه داده GSEA^۷ به نشانی اینترنتی <https://www.gsea-msigdb.org/gsea/index.jsp> انجام شد. ژن‌های افزایش و کاهش بیان یافته مشترک بین حالت مراحل اولیه و پیشرفته NSCLC و ژن‌های اختصاصی مربوط به حالت پیشرفته سرطان به صورت جداگانه مورد بررسی قرار گرفتند.

تجزیه و تحلیل شبکه برهمکنش پروتئین-پروتئین

(PPI) و شناسایی ژن‌های کانونی

برای درک برهمکنش بین پروتئین‌ها شبکه PPI با استفاده از پایگاه داده STRING به نشانی <https://string-db.org/> رسم شد. برهمکنش‌هایی با

¹ Early stage

² Late stage

³ Differentially expressed genes

⁴ Gene ontology

⁵ Gene enrichment analysis

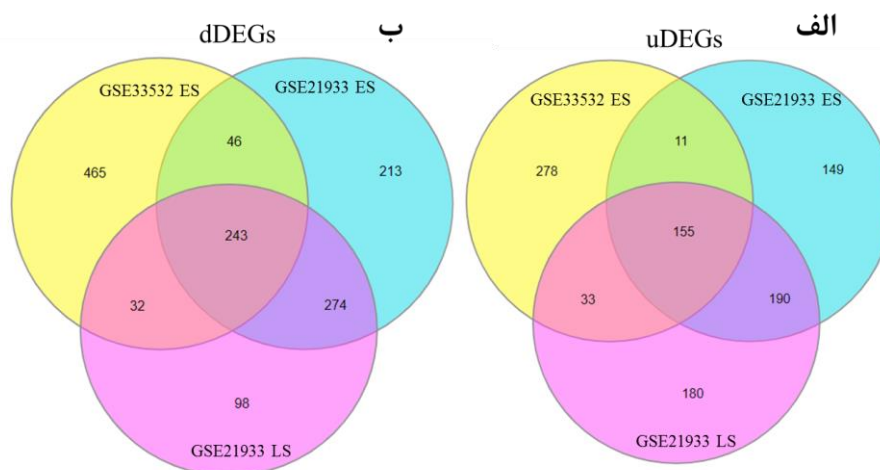
⁶ Kyoto Encyclopedia of Genes and Genomes

⁷ Gene Set Enrichment Analysis

⁸ Hazard ratio

⁹ Confidence intervals

¹⁰ Venn diagram



شکل ۱ - نمودار ون برای یافتن ژن‌های تغییر بیان یافته مشترک و اختصاصی برای مراحل اولیه و پیشرفته سرطان غیر سلول کوچک ریه. الف) برای ژن‌های افزایش بیان یافته (uDEGs) مشترک بین مراحل اولیه و پیشرفته ۱۵۵ ژن و ژن‌های اختصاصی مراحل پیشرفته ۱۸۰ ژن شناسایی شد. ب) برای ژن‌های کاهش بیان یافته (dDEGs) مشترک بین مراحل اولیه و پیشرفته ۲۴۳ ژن و ژن‌های اختصاصی مراحل پیشرفته ۹۸ ژن شناسایی شد. (uDEGs: upregulated differentially expressed genes, dDEGs: downregulated differentially expressed genes)

سلولی، پیام‌رسانی P53 و تعامل گیرنده‌های خارج سلولی بود (شکل ۲-د).

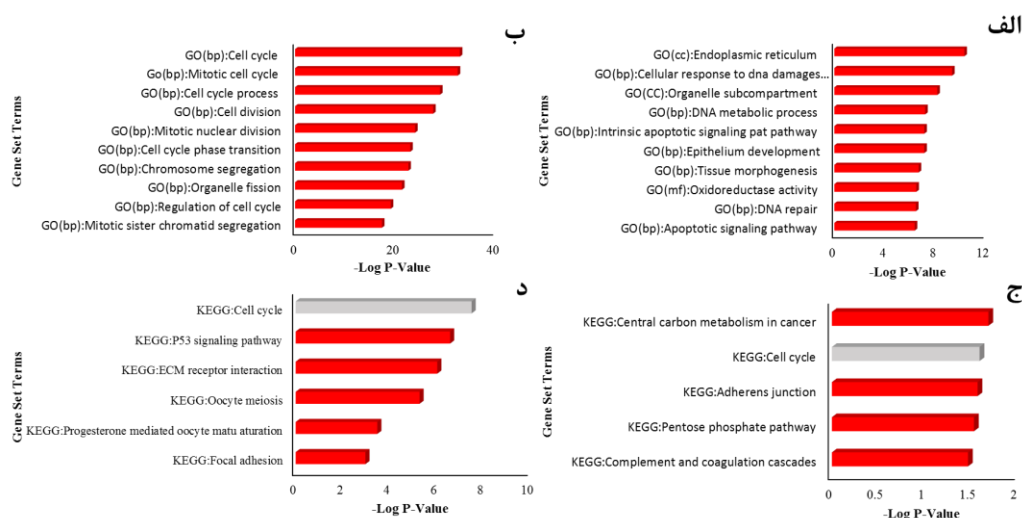
ژن‌های کاهش بیان یافته مختص به مرحله پیشرفته بیماری از نظر هستی‌شناسی بیشتر در فعالیت‌های مربوط به پیام‌رسانی سلولی، فرایندهای مبتنی بر اکتین و فیلامنت‌ها و تنظیم تمایز سلولی بودند (شکل ۳-الف). ژن‌های کاهش بیان یافته‌ی مشترک بین مراحل اولیه و پیشرفته بیشتر مربوط به چسبندگی‌های سلولی، توسعه رگ و ساختار درگیر در مورفوژن بودند (شکل ۳-ب). مسیر تکثیر جمعیت سلولی در بین هر دو حالت مشترک بود. در بررسی مسیرهای KEGG برای ژن‌های کاهش بیان یافته مختص به مرحله پیشرفته بیشتر ژن‌ها در مسیر تولید استروئید دخالت داشتند (شکل ۳-ج). در مقابل در بررسی مسیرهای KEGG در حالت کاهش به طور عمده مسیر مربوط به واکنش‌های انعقادی و آبشاری، ABC ترانسپورتر^۱ و مولکول چسبندگی سلولی سلولی ثبت شد (شکل ۳-د).

شناسایی ژن‌هایی کانونی از طریق بررسی شبکه برهمکنش پروتئین-پروتئین

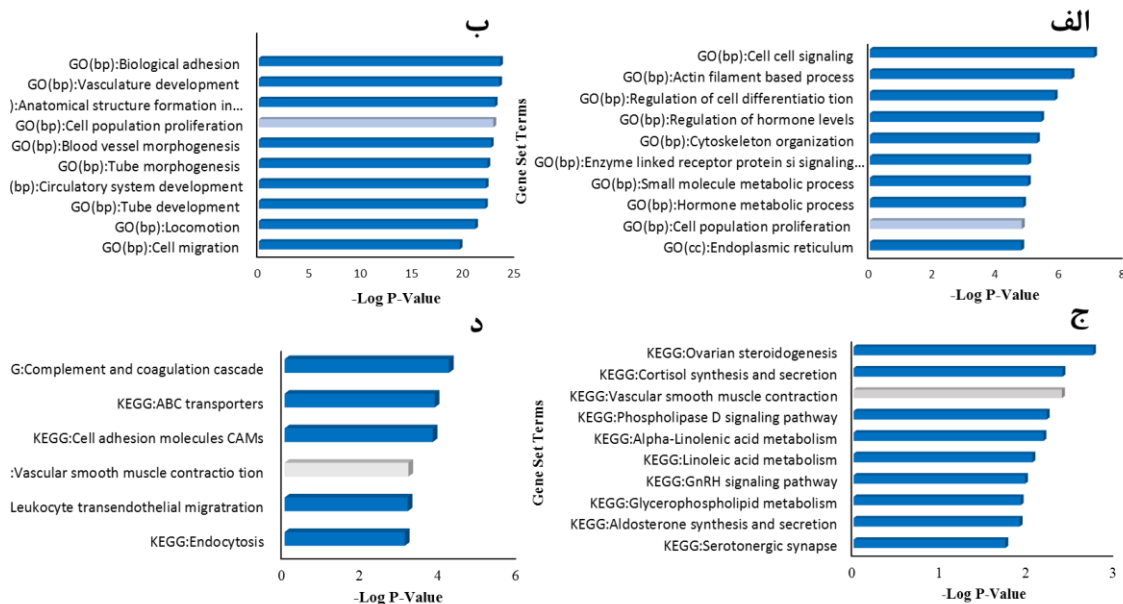
۱۸۰ ژن افزایش بیان یافته و ۹۸ ژن کاهش بیان یافته) ثبت شد (شکل ۱).

تجزیه و تحلیل عملکرد و مسیر غنی‌سازی ژن‌ها
مطابق نتایج به دست آمده از پایگاه داده GSEA ژن‌های افزایش و کاهش بیان یافته در ده مسیر عمده و در حالت اشتراک و اختصاصی با یکدیگر مقایسه شدند. ژن‌های افزایش بیان یافته مختص به مرحله پیشرفته بیماری از نظر هستی‌شناسی بیشتر در فعالیت‌های مربوط به شبکه آندوپلاسمی، پاسخ سلولی به آسیب DNA تحت تاثیر محرک‌ها و تشکیل اندامک‌ها نقش داشتند (شکل ۲-الف). ژن‌های افزایش بیان یافته مشترک بین نمونه‌های مراحل اولیه و پیشرفته بیشتر مربوط به چرخه سلولی و فرایندهای چرخه سلولی بودند (شکل ۲-ب). بررسی غنی شدن ژن‌ها در مسیرهای KEGG برای ژن‌های افزایش یافته مختص به مراحل پیشرفته نشان داد که این ژن‌ها بیشتر در مسیر متابولیسم کربن تمرکز دارند (شکل ۲-ج). در طرف دیگر بررسی غنی شدن ژن‌ها در مسیرهای KEGG برای ژن‌های افزایش یافته‌ی مشترک نشان دهنده نقش آن‌ها در مسیر چرخه

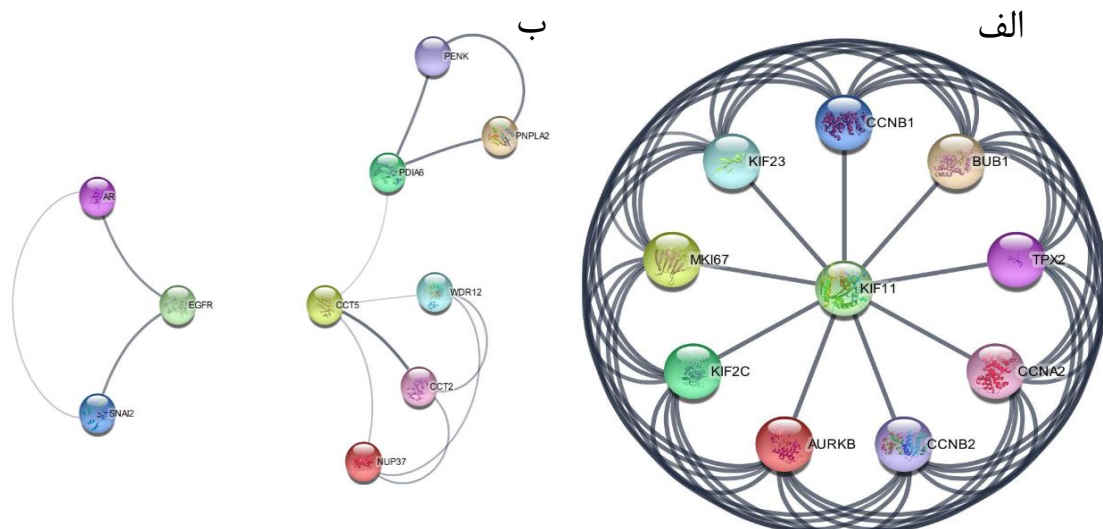
¹ ATP-binding cassette transporter



شکل ۲- بررسی GO و مسیر KEGG برای ژن‌های افزایش بیان یافته. (الف) GO مربوط به ژن‌های افزایش بیان یافته در حالت اشتراک بین مراحل اولیه و پیشرفته بیماری. (ب) GO مربوط به ژن‌های افزایش بیان یافته در حالت اشتراک بین مراحل اولیه و پیشرفته بیماری. (ج) مسیر KEGG مربوط به ژن‌های افزایش بیان یافته مختص به مرحله پیشرفته بیماری. (د) مسیر KEGG مربوط به ژن‌های افزایش بیان یافته در حالت اشتراک بین مرحله اولیه و پیشرفته بیماری. رنگ خاکستری مسیر KEGG مشابه برای ژن‌های مختص به مرحله پیشرفته و ژن‌های مشترک بین مراحل اولیه و پیشرفته را نشان می‌دهد.



شکل ۳- بررسی GO و مسیر KEGG برای ژن‌های کاهش بیان یافته. (الف) GO مربوط به ژن‌های کاهش بیان یافته در حالت اشتراک بین مراحل اولیه و پیشرفته بیماری. (ب) GO مربوط به ژن‌های کاهش بیان یافته در حالت اشتراک بین مراحل اولیه و پیشرفته بیماری. (ج) مسیرهای KEGG مربوط به ژن‌های کاهش بیان یافته مختص به مرحله پیشرفته بیماری. (د) مسیرهای KEGG مربوط به ژن‌های کاهش بیان یافته در حالت اشتراک بین مرحله اولیه و پیشرفته بیماری. رنگ آبی کم رنگ هستی‌شناسی مشابه برای ژن‌های مختص به مرحله پیشرفته و ژن‌های مشترک بین مراحل اولیه و پیشرفته را نشان می‌دهد. رنگ خاکستری مسیر KEGG مشابه برای ژن‌های مختص به مرحله پیشرفته و ژن‌های مشترک بین مراحل اولیه و پیشرفته را نشان می‌دهد.



شکل ۴ - شبکه PPI برای ژن‌ها بایان متمایز. (الف) شبکه PPI برای ۱۰ ژن کانونی در حالت اشتراک بین مراحل اولیه و پیشرفته بیماری. (ب) شبکه PPI برای ۱۰ ژن کانونی در حالت اختصاصی برای مرحله پیشرفته بیماری.

طوری که خطر مرگ را دو برابر افزایش داده است. از بین ۱۰ ژن کانونی مختص به مرحله پیشرفته بیماری ۷ ژن *WDR12*, *PDIA6*, *CCT5*, *CCT2*, *EGFR*, *AR* و *NUP37* دچار افزایش بیان و ۳ ژن *PNPLA2* و *PENK* دچار کاهش بیان بودند. از این تعداد به غیر از *SNAI2* و *PNPLA2* همگی ارتباط معناداری با میزان بقای بیماران داشتند (شکل ۶). به عنوان مثال افزایش بیان بیش از ۵۰ درصد *CCT5* با کاهش ۱/۶ برابری میزان بقا همراه است در حالی که کاهش بیان بیش از ۵۰ درصد *AR* با ۵۰ درصد بقای کمتر همراه می‌باشد.

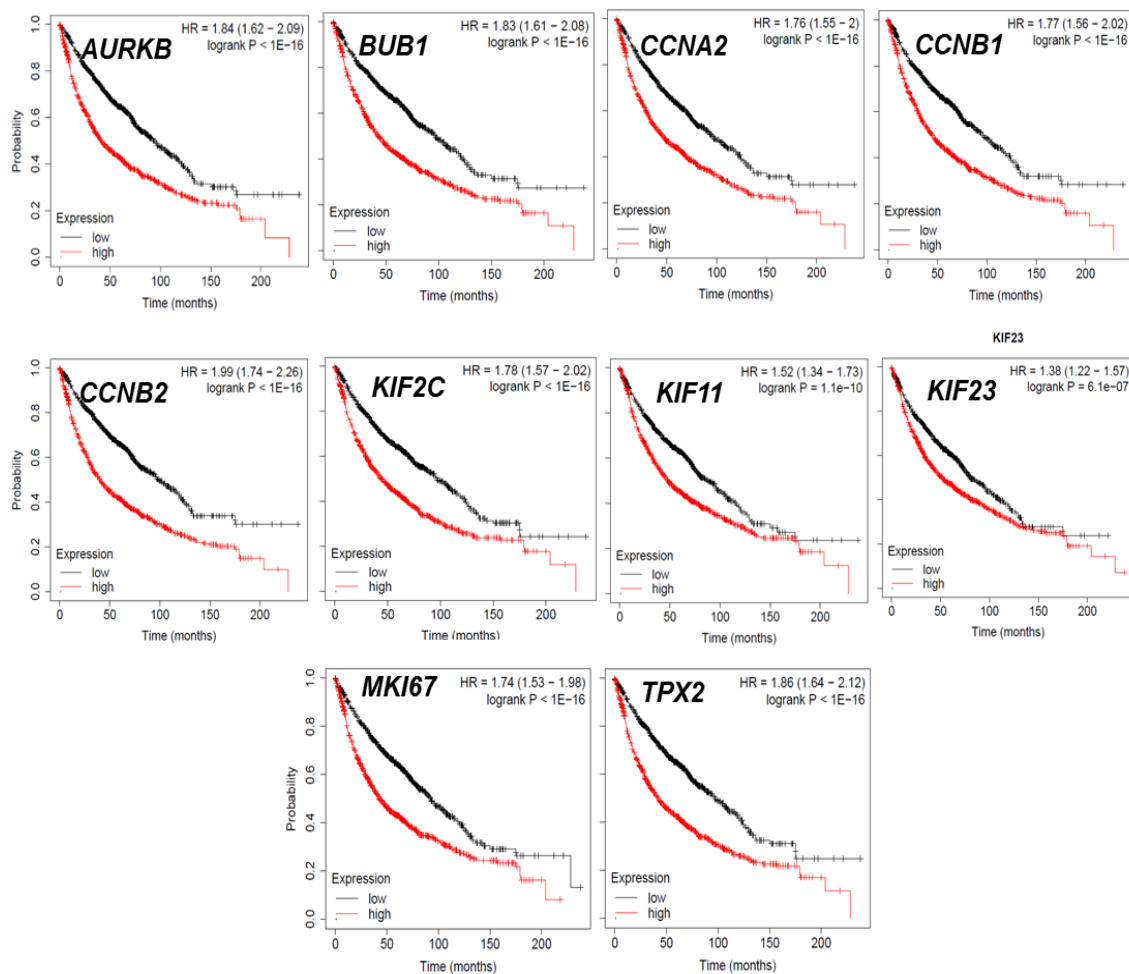
بحث و نتیجه‌گیری:

تومورزایی، پیشرفت و متاستاز سرطان ریه، مانند هر سرطان دیگر، فرایندی پیچیده و شامل انحراف چندین ژن و مسیر سلولی است. ژن‌های کانونی که بیشترین ارتباط را با سایر ژن‌ها دارند احتمالاً ژن‌های اصلی در پیشرفت سرطان هستند. برای بهبود تشخیص و درمان NSCLC، یافتن ژن‌های کانونی (که بیان متمایزی در بافت سرطانی نسبت به بافت سالم دارند) و درک نقش آن‌ها، از اهمیت حیاتی برخوردار است. توسعه فناوری‌های ریزآرایه و توالی‌یابی نسل جدید و توسعه ابزارهای بیوانفورماتیک به شناسایی ژن‌های کانونی در سرطان و

پیشرفته که بیان متمایزی دارند شامل ۳۹۸ گره و ۲۴۲۵ اتصال و برای ژن‌های اختصاصی مرحله پیشرفته بیماری شامل ۲۷۳ گره و ۳۴۰ اتصال است. به کمک افزونه cytoHubba ده ژن کانونی در شبکه ژن‌های تغییر بیان یافته مشترک بین مراحل اولیه و پیشرفته و شبکه ژن‌های تغییر بیان یافته مختص به مراحل پیشرفته مشخص گردیدند. در حالت اول بین ۱۰ ژن کانونی، ۴۵ اتصال مشخص شدند که ژن‌ها به ترتیب براساس بیشترین اتصالات شامل *KIF11*, *MKI67*, *CCNA2*, *KIF23*, *KIF2C*, *TPX2*, *BUB1*, *CCNB1*, *CCNB2* و *AURKB* بودند (شکل ۴-الف). در حالت دوم بین ۱۰ ژن کانونی، ۱۳ اتصال مشاهده شدند که براساس تعداد اتصالات به ترتیب شامل *EGFR*, *CCT2*, *CCT5*, *PDIA6*, *AR*, *WDR12*, *PNPLA2*, *PENK* و *SNAI2* بودند (شکل ۴-ب).

پیش‌آگهی ژن‌های کانونی

همانطور که در شکل ۵ نشان داده شده است، افزایش بیان ۱۰ ژن کانونی مشترک بین مراحل اولیه و پیشرفته همگی به طور معناداری با کاهش بقای بیماران همراه بودند (شکل ۵). از این بین بیان بالای ژن *CCNB2* در گروه بیمارانی که افزایش بالای ۵۰ درصد داشته‌اند، بیشترین تاثیر را بر کاهش بقای بیماران نشان داد به



شکل ۵ - ارتباط بقا کلی با ۱۰ ژن کانونی در بین ژن‌های مشترک بین مراحل اولیه و پیشرفته NSCLC، براساس داده‌های سایت Kaplan-Meier-plotter

پیش‌بینی اهداف بالقوه درمانی کمک شایانی کرده است [۸]. مطالعه حاضر اولین بار با تلفیق اطلاعات دو مطالعه ریزآرایه GSE33532 و GSE21933 به یافتن ژن‌های مرتبط با مراحل پیشرفته NSCLC پرداخته است و برخی ژن‌های کانونی را براساس مرحله بیماری برای نخستین بار گزارش می‌نماید. نتایج مطالعه حاضر نشان می‌دهد که ژن‌های افزایش و کاهش بیان یافته‌ای که مختص به مرحله پیشرفت بیماری هستند، به ترتیب بیشتر در مسیرهای متابولیسم کربن در سرطان و سنتز استروئیدها نقش ایفا می‌کنند. در طرف مقابل ژن‌های افزایش و کاهش بیان یافته مشترک بین مراحل اولیه و

پیش‌بینی اهداف بالقوه درمانی کمک شایانی کرده است [۸].

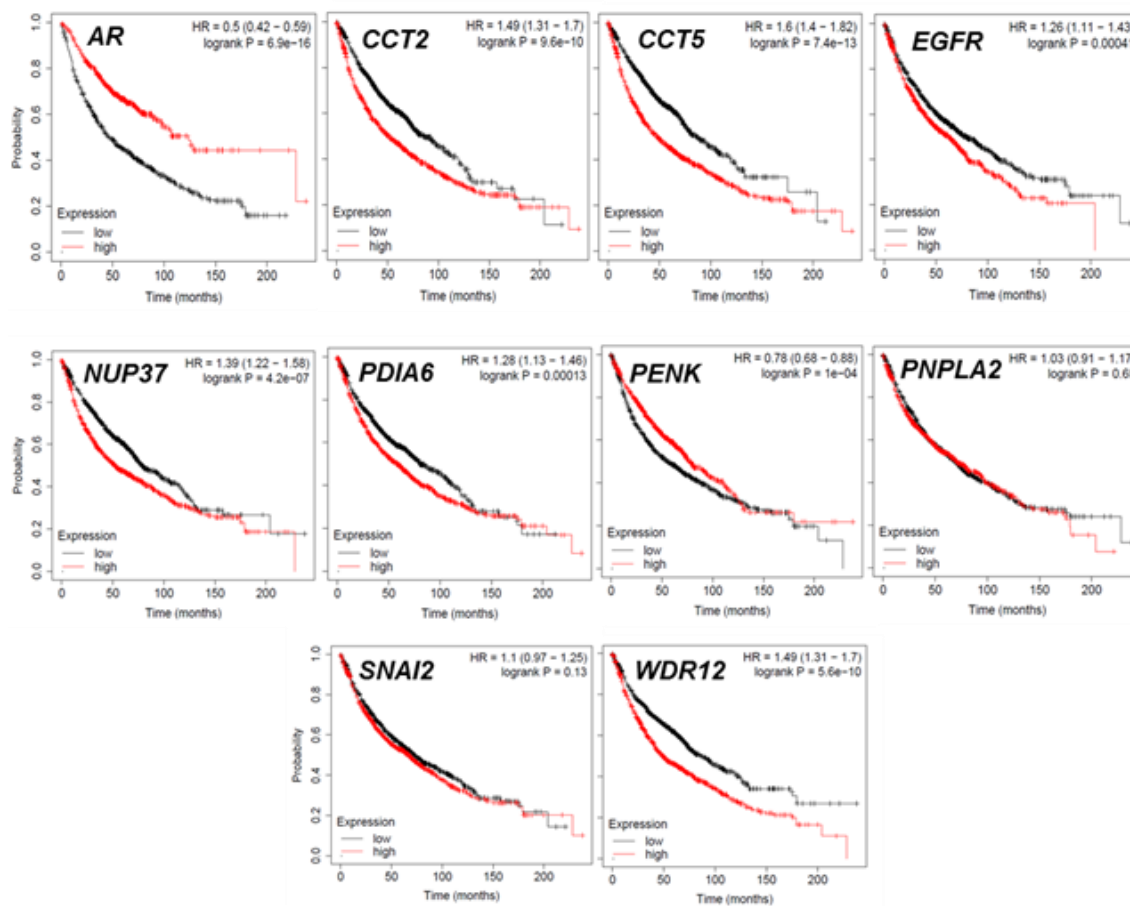
پیشرفته به ترتیب در چرخه سلولی^۱ و آبشار کمپلمان^۲ و کوآگولاسیون^۳ تمرکز دارند. به این ترتیب هدف قرار دادن مسیر متابولیسم کربن در NSCLC می‌تواند راهکاری برای درمان NSCLC در مراحل پیشرفته آن باشد. این یافته با گزارش سایر مطالعات هماهنگ است. به عنوان مثال دو و همکارانش گزارش داده‌اند که وجود پلی‌مورفیسم‌های تک نوکلئوتیدی^۴ (SNP) در ژن‌های مسیر متابولیسم کربن می‌تواند به عنوان فاکتورهای پیش‌آگهی دهنده از میزان بقای بیماران مبتلا به NSCLC باشد [۱۶].

¹ Cell cycle

² Complement cascade

³ Coagulation

⁴ Single nucleotide polymorphism



شکل ۶ - ارتباط بقا کلی با ۱۰ ژن کانونی در بین ژن‌هایی که بر اساس داده‌های سایت Kaplan Meier-plotter به طور اختصاصی بیان متمایزی در مرحله پیشرفته NSCLC دارند.

پیش‌آگهی در پیش بینی بقا یا عود زودرس باشد [۸]. *CCNB2* به عنوان یک واسطه مهم در طول انتقال G2/M از چرخه سلولی عمل می‌کند و یک زیر واحد تنظیم کننده در فاکتور پیش‌برنده میتوز^۲ (MPF) می‌باشد. گزارش شده است که القای کاهش در بیان *CCNB2* در سلول‌های سرطان مثانه تا حد زیادی از توانایی‌های متاستاتیک آن‌ها جلوگیری می‌نماید. همچنین مطالعات مسدود سازی بیان *CCNB2* می‌تواند منجر به افزایش بقا شود [۱۹]. *CCNB1* یا همان سایکلین B1 یک پروتئین تنظیم کننده است که نقش مهمی در میتوز دارد. بیان بیش از حد *CCNB1* در NSCLC تشخیص داده شده است؛ از این رو می‌تواند به عنوان یک نشانگر^۳ در تشخیص NSCLC بکار رود [۲۰]. یوان و همکاران گزارش داده‌اند که کاهش بیان

در مطالعه حاضر براساس تحلیل شبکه PPI، ۱۰ ژن کانونی در شبکه ژن‌های مشترک بین مراحل اولیه و پیشرفته سرطان شناسایی شدند که شامل *CCNA2*، *BUB1*، *CCNB1*، *KIF23*، *KIF11*، *MKI67*، *TPX2*، *KIF2C*، *CCNB2* و *AURKB* بودند. در این میان، پیشتر ژن‌های *CCNB1*، *CCNB2* و *BUB1* به عنوان ژن کانونی در مراحل اولیه NSCLC گزارش شده بودند [۱۷، ۱۸]. چهار ژن از این ده ژن یعنی *CCNA2*، *CCNB2*، *CCNB1* و *BUB1* در مسیر چرخه سلولی KEGG تمرکز دارند. ژن *CCNA2* یک سایکلین^۱ است که در انتقال از G1/S و G2/M نقش دارد. بیان پروتئین این ژن در انواع تومورها، از جمله سرطان پستان، کبد، پروستات و ریه افزایش می‌یابد و به نظر می‌رسد این ژن نشانگر

^۲ Mitose promoting factor

^۳ Marker

^۱ Cyclin-A2

سرطان‌های پیشرفته ریه ارزشمند باشد [۲۷]. ژن *AURKB* نقش اساسی در برهمکنش بین کروموزوم‌ها و میکروتوبول‌ها در حین میتوز دارد [۲۸]. مطالعات قبلی فعالیت‌های ضد توموری و حساسیت پرتویی داورینول^۷ را در سلول‌های سرطانی ریه انسان از طریق مهار *AURKB* تایید کرده است [۳].

در مطالعه حاضر همچنین ۱۰ ژن کانونی مختص به مراحل پیشرفته بیماری گزارش شد که شامل ژن‌های *WDR12 AR PDIA6 CCT5 CCT2 EGFR* *PNPLA2* *PENK* *NUP37* و *SNAI2* بودند. از بین این ژن‌ها *EGFR* و *SNAI2* در تنظیم اتصالات سلولی نقش دارند [۲۹، ۳۰]. بر طبق نتایج مطالعه حاضر ژن *EGFR* کانونی‌ترین ژن در مراحل پیشرفته NSCLC است. *EGFR* گیرنده فاکتور رشد اپیدرمال^۸ اپیدرمال^۸ است. مهارکننده‌های تیروزین کیناز^۹ (TKI) مانند جفیتینیب^{۱۰} و ارلوتینیب^{۱۱} اولین درمان هدفمند برای NSCLC در افراد دارای جهش *EGFR* می‌باشد. همچنین *EGFR*-TKI در مقایسه با شیمی‌درمانی مبتنی بر پلاتین که جزء درمان‌های رایج NSCLC است، بقای بیماران را به طور قابل توجهی طولانی کرده است. با این حال، فقط ۲۰ تا ۳۰ درصد از کل بیماران NSCLC با جهش‌های حساس به *EGFR* می‌توانند از *EGFR*-TKI بهره‌مند شوند [۳۱].

افزایش بیان *CCT2* دیگر ژن کانونی مرتبط با مراحل پیشرفته NSCLC، در دیگر سرطان‌ها مانند سرطان کبد، پروستات و پستان نیز گزارش شده است و با شدت سرطان و پیش‌آگهی نامطلوب ارتباط دارد [۳۲]. افزایش بیان *CCT5* در NSCLC توسط گائو و همکاران گزارش شده است [۳۳]. *PDIA6* عضوی از خانواده دی‌سولفید ایزومراز^{۱۲} است که در شبکه اندوپلاسمی قرار دارد. بیان این ژن در NSCLC افزایش می‌یابد و با پیش‌آگهی ضعیف بیماران NSCLC همراه است [۳۴]. ژن *AR* دیگر ژن کانونی مختص به مرحله پیشرفته است

CCNB1 باعث کاهش تکثیر سلولی و القا آپوپتوز^۱ در سلول‌های توموری می‌شود [۲۱]. همچنین سوریا و همکاران بیان کرده‌اند که تغییر بیان *CCNB1* در کسر مشخصی از مبتلایان به NSCLC رخ می‌دهد. آن‌ها همچنین نشان دادند انواع NSCLC علاوه بر تفاوت‌های زیستی در میزان بیان این ژن نیز متفاوت هستند. بیشترین تغییر بیانی برای این ژن مربوط به سرطان سلول‌های غیر کوچک سرطان ریه از نوع سنگفرشی است [۲۲]. *BUB1* دیگر ژن کانونی است که در چندین مسیر تومورزایی از جمله میوز تخمک، چرخه سلولی و مسیر سیگنالینگ P53^۲ نقش دارد. *BUB1* با کینازهای مرتبط با تومور^۳، میکروRNAها^۴ و عوامل رونویسی^۵ در ارتباط است [۲۳].

همچنین سایر ژن‌های کانونی نیز به نوعی با تقسیم سلولی مرتبط هستند. به عنوان مثال ژن *KIF2C* پروتئینی را رمزگذاری می‌کند که بخشی از خانواده پروتئین‌های شبه کیناز است. این پروتئین در تنظیم حرکت میکروتوبول در سلول‌ها نقش دارد و برای جداسازی کروموزوم در طول آنافاز مهم است. بیان غیرطبیعی این ژن بر میتوز تاثیر می‌گذارد و می‌تواند منجر به سرطان شود [۲۴]. دیگر ژن کانونی به نام *MKI67* یک آنتی‌ژن هسته‌ای را در طی مراحل S، G1 و G2/M در سلول‌های در حال تکثیر کد می‌کند [۲۵]. *KIF11* عضوی از خانواده اکتین است که جهت ساخت دوک دو قطبی برای تقسیم سلولی ضروری است و با داشتن این عملکرد می‌تواند به عنوان یک هدف درمانی باشد [۲۶]. *KIF23* عضوی از خانواده شبه کاینزین^۶ است. پروتئین ساخته شده توسط این ژن در انتقال اندامک‌ها در درون سلول و حرکت کروموزوم‌ها حین تقسیم نقش دارد [۲۷]. کاتو و همکاران بیان کردند که افزایش بیان این ژن در اکثر بافت‌های سرطانی متاستاتیک ریه در مقایسه با نمونه کنترل مشاهده شده است. از این رو برای رشد و بقای سلول‌های سرطانی ریه بسیار مهم است و می‌تواند برای اهداف درمانی جدید در

⁷ Daurinol⁸ Epidermal growth factor receptor⁹ Tyrosine kinase inhibitors¹⁰ Gefitinib¹¹ Erlotinib¹² Disulfide-isomerase¹ Apoptosis² P53 signaling³ Tumor-associated kinase⁴ microRNAs⁵ Transcription factors⁶ Kinesins

- International Journal of Oncology. 2012;41(6):1943-1949.
- 2) Khazaei S, Mansori K, Soheylizad M, et al. Epidemiology of lung cancer in Iran: sex difference and geographical distribution. Middle East Journal of Cancer. 2017;8(4):223-228.
 - 3) Tang Q, Zhang H, Kong M, Mao X, Cao X. Hub genes and key pathways of non-small lung cancer identified using bioinformatics. Oncology Letters. 2018;16(2):2344-2354.
 - 4) Wang S, Xu X, Kong W. Identification of hub genes associated with lung adenocarcinoma based on bioinformatics analysis. Computational and Mathematical Methods in Medicine. 2021;2021. <https://doi.org/10.1155/2021/5550407>
 - 5) Itani R, Al Faraj A. siRNA conjugated nanoparticles—a next generation strategy to treat lung cancer. International Journal of Molecular Sciences. 2019;20(23):6088.
 - 6) Ni M, Liu X, Wu J, et al. Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. Frontiers in Genetics. 2018;9:469.
 - 7) Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clinic Proceedings; 2019;94(8):1623-1640.
 - 8) Xiao Y, Feng M, Ran H, Han X, Li X. Identification of key differentially expressed genes associated with non-small cell lung cancer by bioinformatics analyses. Molecular Medicine Reports. 2018;17(5):6379-6386.
 - 9) Huang H, Huang Q, Tang T, et al. Differentially expressed gene screening, biological function enrichment, and correlation with prognosis in non-small cell lung cancer. Medical Science Monitor. 2019;25:4333-4341.
 - 10) Kuznetsova I, Lugmayr A, Siira SJ, Rackham O, Filipovska A. CirGO: an alternative circular way of visualising gene ontology terms. BMC Bioinformatics. 2019;20(1):1-7.
 - 11) Pomaznoy M, Ha B, Peters B. GOnet: a tool for interactive Gene Ontology analysis. BMC Bioinformatics. 2018;19(1):1-8.
 - 12) Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000;28(1):27-30.
 - 13) Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. Journal of Proteome Research. 2018;18(2):623-632.
 - 14) Chin C-H, Chen S-H, Wu H-H, et al. CytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Systems Biology. 2014;8(4):1-7.

که ارتباط این ژن با انواع بدخیمی ازجمله سرطان ریه توسط هوانگ و همکاران گزارش شده است [۳۵].

تجزیه و تحلیل بقا نشان داده که افزایش بیان *WDR12* با کاهش بقا کلی در NSCLC همراه است [۳۶].

PNPLA2 آنزیمی است که سرکوب آن منجر به تجمع چربی در سلول‌های اپیتلیال ریه انسان می‌شود [۳۷]. ژن *PENK* در بلوغ عصبی و سیستم غدد درون‌ریز اعصاب بیان می‌شود و با تاثیر بر مسیر opioid در تنظیم مرگ و بقا سلول نقش دارد [۳۸]. بیان دو ژن *PNPLA2* و *PENK* نیز در مرحله پیشرفته بیماری کاهش می‌یابد.

NUP37 جزئی از کمپلکس منفذ هسته‌ای^۱ (NPC) است [۳۹]. یانگ و همکاران افزایش بیان این ژن را در مراحل پیشرفته NSCLC گزارش کرده‌اند. همچنین بیان شده است که این پروتئین در تنظیم چرخه سلولی نقش دارد و تغییر در بیان آن با میزان بقا در مبتلایان به NSCLC ارتباط دارد از این رو می‌تواند به عنوان یک نشانگر کلیدی در درمان سرطان ریه نقش ایفا کند [۳۹].

در این مطالعه داده‌ها فقط بر اساس مرحله‌ای که نمونه‌های بیمار در آن قرار داشتند مورد تجزیه و تحلیل قرار گرفته‌اند. از این رو چنانچه از همین داده‌ها برای مقایسه نوع سلول‌های درگیر در بیماران نیز استفاده شود می‌توان به ژن‌های ارزشمندتری دست یافت و مشخص کرد که در هر نوع از NSCLC و در هر مرحله از بیماری بیان چه ژن‌هایی تغییر می‌یابد.

این مطالعه با استفاده از تجزیه و تحلیل بیوانفورماتیکی ۳۹۸ ژن تغییر یافته مشترک و ۲۷۸ ژن تغییر یافته مختص به مرحله پیشرفته انجام گرفت. GO و KEGG ژن‌ها مشخص و ۲۰ ژن کانونی هم در مراحل مختلف بیماری شناسایی شدند که ژن‌های مورد اعتمادی برای پیش‌آگهی سرطان ریه و مراحل پیشرفته آن بودند. ژن‌های کانونی معرفی شده همچنین می‌توانند به عنوان اهدافی برای راهبردهای درمانی مختلف از جمله RNAهای کوچک مداخله‌گر باشند.

منابع:

- 1) Kloskowski T, Gurtowska N, Olkowska J, et al. Ciprofloxacin is a potential topoisomerase II inhibitor for the treatment of NSCLC.

¹ Nuclear pore complex

- 26) Li T, Gao X, Han L, Yu J, Li H. Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis. *World Journal of Surgical Oncology*. 2018;16(1):1-12.
- 27) Kato T, Wada H, Patel P, et al. Overexpression of KIF23 predicts clinical outcome in primary lung cancer patients. *Lung Cancer*. 2016;92:53-61.
- 28) Luciano AM, Lodde V, Franciosi F, Ceciliani F, Peluso JJ. Progesterone receptor membrane component 1 expression and putative function in bovine oocyte maturation, fertilization, and early embryonic development. *Reproduction*. 2010;140(5):663.
- 29) Katafiasz D, Smith LM, Wahl JK 3rd. Slug (SNAIL2) expression in oral SCC cells results in altered cell-cell adhesion and increased motility. *Cell Adhesion & Migration*. 2011;5(4):315-322.
- 30) Rao TC, Ma VP-Y, Blanchard A, et al. EGFR activation attenuates the mechanical threshold for integrin tension and focal adhesion formation. *Journal of Cell Science*. 2020;133(13):jcs238840.
- 31) Wang Y, Zheng C, Lu W, et al. Bioinformatics-based identification of HDAC inhibitors as potential drugs to target EGFR wild-type non-small-cell lung cancer. *Frontiers in Oncology*. 2021;11:339.
- 32) Klimczak M, Biecek P, Zylicz A, Zylicz M. Heat shock proteins create a signature to predict the clinical outcome in breast cancer. *Scientific Reports*. 2019;9(1):1-15.
- 33) Gao H, Zheng M, Sun S, et al. Chaperonin containing TCP1 subunit 5 is a tumor associated antigen of non-small cell lung cancer. *Oncotarget*. 2017;8(38):64170.
- 34) Bai Y, Liu X, Qi X, et al. PDIA6 modulates apoptosis and autophagy of non-small cell lung cancer cells via the MAP4K1/JNK signaling pathway. *EBioMedicine*. 2019;42:311-325.
- 35) Huang Q, Tan Q, Mao K, et al. The role of adrenergic receptors in lung cancer. *American Journal of Cancer Research*. 2018;8(11):2227.
- 36) Li W, Li N, Gao L, You C. Integrated analysis of the roles and prognostic value of RNA binding proteins in lung adenocarcinoma. *PeerJ*. 2020;8:e8509.
- 37) Tang M, Hu Z, Rao C, et al. Burkholderia pseudomallei interferes with host lipid metabolism via NR1D2-mediated PNPLA2/ATGL suppression to block autophagy-dependent inhibition of infection. *Autophagy*. 2020:1-16.
- 38) Roperch J-P, Incitti R, Forbin S, et al. Aberrant methylation of NPY, PENK, and WIF1 as a promising marker for blood-based
- 15) Hou G-X, Liu P, Yang J, Wen S. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan–Meier plotter. *PloS One*. 2017;12(3):e0174515.
- 16) Do SK, Choi SH, Lee SY, et al. Genetic variants in one-carbon metabolism pathway predict survival outcomes of early-stage non-small cell lung cancer. *Oncology*. 2020;98(12):897-904.
- 17) Xu X, Xu L, Huang H, et al. Identification of hub genes as biomarkers correlated with the proliferation and prognosis in lung cancer: a weighted gene co-expression network analysis. *BioMed Research International*. 2020; 2020:3416807. <https://doi.org/10.1155/2020/3416807>
- 18) Zeng Y, Li N, Chen R, et al. Screening of hub genes associated with prognosis in non-small cell lung cancer by integrated bioinformatics analysis. *Translational Cancer Research*. 2020;9(11):7149-7164.
- 19) Tu H, Wu M, Huang W, Wang L. Screening of potential biomarkers and their predictive value in early stage non-small cell lung cancer: a bioinformatics analysis. *Translational Lung Cancer Research*. 2019;8(6):797.
- 20) Li G, Liu X, Zhang D, Liu D, Li Z. The expression and significance of cyclin B1 and survivin in human non-small cell lung cancer. *The Chinese-German Journal of Clinical Oncology*. 2011;10(4):192-197.
- 21) Yuan J, Yan R, Krämer A, et al. Cyclin B1 depletion inhibits proliferation and induces apoptosis in human tumor cells. *Oncogene*. 2004;23(34):5843-5852.
- 22) Soria J-C, Jang SJ, Khuri FR, et al. Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. *Cancer Research*. 2000;60(15):4000-4004.
- 23) Wang L, Yang X, An N, Liu J. Bioinformatics analysis of BUB1 expression and gene regulation network in lung adenocarcinoma. *Translational Cancer Research*. 2020;9(8):4820-4833.
- 24) Liao Y, Yin G, Wang X, et al. Identification of candidate genes associated with the pathogenesis of small cell lung cancer via integrated bioinformatics analysis. *Oncology Letters*. 2019;18(4):3723-3733.
- 25) Yang C, Su H, Liao X, et al. Marker of proliferation Ki-67 expression is associated with transforming growth factor beta 1 and can predict the prognosis of patients with hepatic B virus-related hepatocellular carcinoma. *Cancer Management and Research*. 2018;10:679.

- diagnosis of colorectal cancer. BMC Cancer. 2013;13(1):1-10.
- 39) Huang L, Wang T, Wang F, et al. NUP37 silencing induces inhibition of cell proliferation, G1 phase cell cycle arrest and apoptosis in non-small cell lung cancer cells. Pathology-Research and Practice. 2020;216(3):152836.

Identification of Hub Genes Associated with the Progression of Non-small Cell Lung Carcinoma

Naemeh Jafarpour¹, Khadijeh Sadegh¹, Amirhossein Ahmadi^{1*}

1) Department of Biological science and technology, Faculty of Nano and Bio Science and Technologies, Persian Gulf University, Bushehr, Iran

Abstract:

Non-small cell lung carcinoma is the most prevalent type of lung cancer which causes about one million death worldwide. Although patients with the early stage of NSCLC have better responses to drugs, the overall five-year survival remains poor for the patients with the late stages of this cancer. Therefore, understanding the molecular mechanisms underlying the progression of NSCLC, like alterations in expression profile, is essential to find new therapeutic targets and increase the patient's survival. Bioinformatic tools and microarray together with next-generation sequencing studies have important roles in finding hub genes in the network of differentially expressed genes. Thus, this study aims to find hub genes which are associated with NSCLC and study their correlations with the patients' survival.

In this study, two microarray studies of GSE21933 and GSE21933 from GEO database were selected. To find differentially expressed genes in tumor samples versus controls, the R software was used. Gene ontology and KEGG gene enrichment analysis were performed by GSEA website. Cytoscape software and Cytohubba plugin were used to identify the hub genes. The correlation between each canonical gene and patients' survival was extracted from the Kaplan Meier plotter website.

The results showed that the hub genes in the network of late stage-specific differentially expressed genes are EGF, CCT2, CCT5, PDIA6, AR, WDR12, PNPLA2, PENK, NUP37 and SNAI2. All genes except PNPLA2 and SNAI2, showed a significant correlation with the patients' overall survival.

This study introduced hub genes which are associated with the late-stage NSCLC. Possibly, they could be targeted to inhibit NSCLC progression and increase the survival of patients with NSCLC.

Keywords: Non-small cell lung carcinoma, Gene expression profile, Hub genes, Overall survival

* Corresponding Author:

Amirhossein Ahmadi, PhD. Department of Biological Science and Technology, Faculty of Nano and Bio Science and Technologies, Persian Gulf University, Bushehr, 75169, Iran. Email: ahahmadi@pgu.ac.ir