Iranian Journal of Breast Diseases. 2022;14(1):49-63.

Original Article

Identification of Prognostic Genes in HER2-enriched Breast Cancer by Gene Co-expression Network Analysis

Darzi M[1], Gorgin S[1], Majidzadeh-A K[2], Esmaeili R[2*]

[1]Department of Electrical Engineering and Information Technology, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran
[2]Genetics Department, Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran

Receive: 31/1/2021
Accepted: 27/3/2021

[*]**Corresponding Author:**
esmaeili.rezvan@gmail.com

Ethics Approval:
Not Applicable

Abstract

Introduction: HER2-enriched subtype of breast cancer has a worse prognosis than luminal subtypes. Recently, the discovery of targeted therapies in other groups of breast cancer has increased patient survival. The aim of this study was to identify genes that affect the overall survival of this group of patients based on a systems biology approach.

Methods: Gene expression data and clinical information on 58 patients with HER2-enriched cancer were downloaded from The Cancer Genome Atlas (TCGA). Co-expression modules were identified using the weighted gene co-expression network analysis (WGCNA). The Cox regression was used to determine the modules that had a significant relationship with the overall survival (OS) endpoint. Single-gene survival analysis was performed within the selected module. Finally, functional annotation to explore the significance of genes was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID).

Results: Of the six identified co-expression modules, two had significantly poor prognoses. Single-gene survival analysis showed that 39% of genes in the selected modules were identified as significant. The genes were mainly related to the biological pathways such as Ubiquitin-mediated proteolysis and RNA degradation. *CHAMP1*, *PPP1R26*, *PRRC2B*, *KANSL3*, and *ANAPC2* were identified as the 5 most important genes associated with reduced OS, in order of significance.

Conclusion: The systems biology approach can provide appropriate results relate to patient survival analysis. In this study, some genes were identified to be used as prognostic biomarkers in experimental studies related to the OS in the HER2-enriched subgroup. These genes can be considered potential candidates for therapeutic targets in this group of patients.

Keywords: HER2-enriched Breast Cancer, WGCNA, Co-expression Network, TCGA

[Downloaded from ijbd.ir on 2028-01-05]

[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

فصلنامه بیماری های پستان ایران
۱۴۰۰؛ ۱۴(۱): ۴۹-۶۳

مقاله پژوهشی

شناسایی ژن‌های مرتبط با پیش‌آگهی در سرطان پستان Her2-enriched با استفاده از تجزیه و تحلیل شبکه هم‌بیانی ژنی

محمد درزی[۱]، سعید گرگین[۱]، کیوان مجیدزاده[۲]، رضوان اسمعیلی[۲*]

[۱] پژوهشکده برق و فناوری اطلاعات، سازمان پژوهش‌های علمی و صنعتی ایران، تهران، ایران
[۲] دپارتمان ژنتیک، مرکز تحقیقات سرطان پستان، پژوهشکده سرطان معتمد، جهاد دانشگاهی، تهران، ایران

تاریخ ارسال: ۹۹/۱۱/۱۱
تاریخ پذیرش: ۱۴۰۰/۱/۷

* نویسنده مسئول:
esmaeili.rezvan@gmail.com

چکیده

مقدمه: زیرگروه Her2-enriched در سرطان پستان پیش‌آگهی بدتری نسبت به زیرگروه‌های لومینال دارد. و اخیرا کشف درمان‌های هدفمند در سایر گروه‌های سرطان پستان منجر به افزایش بقاء بیماران شده است. هدف این مطالعه شناخت ژن‌های موثر در بقای کلی این دسته از بیماران مبتنی بر زیست‌شناسی سامانه‌ای است.

روش بررسی: مجموعه داده‌های بیان ژن و اطلاعات بالینی ۵۸ بیمار مبتلا به زیرگروه Her2-enriched از اطلس ژنوم سرطان دانلود شد. با استفاده از شبکه هم‌بیان ژنی وزن‌دار، ماژول‌های هم‌بیان شناسایی شدند. به‌منظور شناسایی ماژول‌های موثر بر بقای کلی، از رگرسیون کاکس استفاده شد. آنالیز بقاء تک‌ژنی در درون ماژول منتخب انجام و در پایان ژن‌های معنی‌دار با استفاده از ابزار DAVID مورد تفسیر عملکردی قرار گرفتند.

یافته‌ها: از میان شش ماژول هم‌بیان شناسایی‌شده، دو ماژول به طور معنی‌دار با پیش‌آگهی بدتر در بیماران مرتبط بودند. بر مبنای آنالیز بقای تک‌ژنی، ۳۹٪ از ژن‌های ماژول منتخب، معنادار شناخته شد که رابطه معناداری با مسیرهای زیستی "Ubiquitin mediated proteolysis" و " RNA degradation" داشتند. ژن‌های *CHAMP1*، *PPP1R26*، *PRRC2B*، *KANSL3* و *ANAPC2* به‌ترتیب به عنوان ۵ ژن مهم در کاهش بقای کلی شناسایی شدند. همچنین این ژن‌ها در مسیرهای مرتبط با تقسیم سلولی نقش دارند که تاییدکننده نقش آن‌ها در سرطان است.

نتیجه‌گیری: رویکرد زیست‌شناسی سامانه‌ای می‌تواند نتایج مناسبی در تحلیل بقاء بیماران داشته باشد. در این مطالعه ژن‌هایی شناسایی شدند که می‌توانند در مطالعات آزمایشگاهی به عنوان بیومارکرهای پیش‌آگهی‌دهنده بقای کلی بیماران زیرگروه Her2-enriched و همچنین کاندیداهای بالقوه برای درمان هدفمند این دسته از بیماران مورد استفاده قرار گیرند.

واژه‌های کلیدی: سرطان، زیست‌شناسی سامانه‌ای، شبکه‌های وزندار، شبکه‌های هم‌بیانی، Her2-enriched، WGCNA

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

مقدمه

رشد روزافزون سرطان پستان در چند دهه اخیر و اثرات سرطان پستان شایع‌ترین سرطان در بین زنان و دومین عامل مرگ‌ومیر ناشی از سرطان‌ها بعد از سرطان ریه در بین این قشر به شمار می‌رود. اگرچه شیوع سرطان پستان افزایش داشته ولی به‌دلیل تغییر تدابیر درمانی مرگ‌ومیر ناشی از آن کاهش یافته است. با این وجود هنوز حدود ۲۴٪/۵ از کل موارد سرطان و ۱۵/۵٪ مرگ و میر ناشی از آن مربوط به سرطان پستان است (۱).

این سرطان در صورت تشخیص در مراحل اولیه از شانس بالایی برای درمان برخوردار است. عوامل متداول تعیین‌کننده پیش‌آگهی[1] در سرطان پستان عبارتند از وضعیت غده‌های لنفاوی زیربغلی، اندازه و درجه تومور، هیستولوژی، وضعیت گیرنده‌های استروژن[2]، پروژسترون[3]، و فاکتور رشد اپیدرمی ۲[4]. علاوه بر این با توجه به ظهور فناوری‌های وابسته به بررسی بیان ژن‌ها با بازدهی بالا مانند میکروآرایه[5] و نسل جدید توالی‌یابی[6]، اخیرا تقسیم‌بندی مولکولی نیز در سرطان پستان انجام شده است و در حال حاضر پیش‌بینی بیماری از این طریق انجام می‌شود. بر این اساس سرطان پستان به ۵ زیرگروه دسته‌بندی می‌شود. لومینال A[7] با فراگیری ۵۰٪ شایع‌ترین سرطان‌های مهاجم پستان را تشکیل می‌دهد. این گروه از سرطان از نظر میزان بیان ER/PR مثبت بوده و از نظر بیان HER2 منفی هستند. ژن‌های مرتبط با این زیرگروه عموماً دارای پیش‌آگهی خوب هستند. لومینال B[8] دسته دیگری از سرطان پستان است که حدود ۲۰٪ از سرطان‌های مهاجم پستان را شامل می‌شود. میزان بیان ER/PR در آن‌ها مثبت و بیان HER2/neu در آن‌ها به صورت متغیر (مثبت یا منفی) است. این تنوع در بیان ژن *HER2* باعث بروز پاسخ‌های متفاوت نسبت به هورمون‌درمانی[9] و شیمی‌درمانی[10] در این گروه شده است. به طور کلی دارای پیش‌آگهی ضعیف‌تری نسبت به مبتلایان گروه لومینال A هستند (۲). سرطان پستان سه گانه منفی[11] یا بازال[12] شامل ۲۰-۱۰٪ از تمام موارد سرطان پستان است که متاسفانه در زنان جوان به ویژه در قبل از دوره یائسگی[13] بیشتر دیده می‌شود. این بدخیمی، فنوتیپ خاصی از سرطان پستان است که فاقد بیان بیش از حد گیرنده استروژن، گیرنده پروژسترون و پروتئین گیرنده رشد اپیدرمی فاکتور ۲ بوده و در نتیجه در حال حاضر شیمی‌درمانی تنها درمان استاندارد برای مبتلایان این گروه به شمار می‌آید. شواهد حاصل از مطالعات آزمایشگاهی نشان می‌دهد که سرطان پستان سه‌گانه منفی دارای میزان نسبتا بالایی از عود و متاستاز دوره‌ای با بقای کلی ضعیف است (۳- ۵).

HER2-enriched یکی دیگر از زیرگروه‌های سرطان پستان است که در آن گیرنده هورمونی منفی است (گیرنده استروژن و گیرنده پروژسترون منفی) و HER2 مثبت است. سرطان HER2-enriched سریع‌تر از سرطان‌های لومینال رشد می‌کند و در گذشته پیش‌آگهی بدتری داشته‌اند اما آن‌ها اغلب با درمان‌های هدف‌مند با هدف پروتئین HER2 مانند Enhertu و Herceptin با موفقیت درمان می‌شوند (۶، ۷). Normal-like آخرین زیرگروه سرطان پستان است که شامل مبتلایانی است که گیرنده هورمون مثبت (گیرنده استروژن و یا گیرنده پروژسترون مثبت) و HER2 منفی دارند. در این دسته از مبتلایان به سرطان پستان، سطح پروتئین Ki-67 است که به کنترل سرعت رشد سلول‌های سرطانی کمک می‌کند. شرایط این گروه شبیه گروه لومینال A است. بیماران این گروه پیش‌آگهی خوبی دارند.

ناهمگنی[14] در میان بیماران با بیان Her2 بالاست و دسته وسیعی از بیماران از لومینال A و B، Her2-enriched و Normal-like را شامل می‌شود. همچنین با وجودی که بعضی از بیماران بیان بالای ژن Her2 را دارند اما میزان بیان پروتئین Her2 که از طریق روش ایمنوهیستوشیمی[15] اندازه‌گیری می‌شود در برخی از آن‌ها پایین است. بیماران Her2-enriched با توجه به بیان بالای ژن‌های دیگری نظیر ژن‌های مسیر گیرنده فاکتور

[1] Prognosis
[2] Estrogen Receptor (ER)
[3] Progesterone Receptor (PR)
[4] Human Epidermal growth factor Receptor 2 (HER-2)
[5] Microarray
[6] Next Generation Sequencing (NGS)
[7] Luminal A
[8] Luminal B
[9] Endocrine Therapy
[10] Chemotherapy
[11] Triple Negative Breast Cancers (TNBC)
[12] Basal-Like
[13] Pre-menopausal
[14] Heterogeneity
[15] Immunohistochemistry (IHC)

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

رشد اپیدرمال[16] و تایروزین کیناز[17] کاندیدای پاسخ به درمان‌های جدید هستند.

یکی از راهبردهایی که در سال‌های اخیر با گسترش داده‌های با بازدهی بالا[18] بیش از پیش مورد توجه قرار گرفته است استفاده از رویکرد زیست‌شناسی سامانه‌ای[19] برای تحلیل برهمکنش ژن‌ها بر روی یکدیگر و تاثیر این برهمکنش بر روی ویژگی‌های بالینی بیماران بوده است. با استفاده از این رویکرد می‌توان ژن‌های موثر در بقای بیماران را شناسایی و کاندیداهای درمانی مناسب را پیشنهاد نمود. مبنای این رویکرد استفاده از دیدگاه شبکه[20] برای بررسی روابط بین ژن‌ها و تاثیر مجموعه‌ای از آن‌ها بر روی ویژگی‌ها و صفات بیرونی[21] و بالینی است.

تحلیل شبکه هم‌بیان ژنی وزن‌دار (WGCNA[22]) به عنوان یکی از چارچوب‌های مطرح در پیاده‌سازی رویکرد زیست‌شناسی سامانه‌ای محسوب می‌شود. هدف اصلی از طراحی این چارچوب تحلیل رفتار هم‌بیانی مجموعه‌ای از ژن‌ها به جای ژن‌های انفرادی بوده است (۸). در این روش به جای ارتباط ده‌ها هزار ژن با یک متغیر بالینی می‌توان همبستگی چند گروه (ماژول) ژنی را با آن متغیر خاص بررسی نمود. در سال‌های اخیر WGCNA برای ساخت و تحلیل شبکه هم‌بیانی سرطان‌های مختلف همچون (۹-۱۳) و بیماری‌های متعددی مانند (۱۴-۱۷) مورد استفاده قرار گرفته است. شناسایی الگوهای هم‌بیانی می‌تواند به شناسایی عملکرد ژن‌های ناشناخته نیز منجر شود که این امکان با عنوان guilt-by-association (GBA) شناخته می‌شود.

بررسی محققان این پژوهش نشان می‌دهد Her2-enriched یکی از زیرگروه‌های سرطان پستان است که شبکه برهمکنش ژن‌های آن مورد بررسی قرار نگرفته است. در این پژوهش از WGCNA به عنوان الگوریتمی برای شناسایی ماژول‌های ژنی هم‌بیان بیماران زیرگروه Her2-enriched استفاده شده است. در ادامه، همبستگی ماژول‌های هم‌بیانی شناسایی شده با ویژگی‌های بالینی همچون بقای کلی بیماران مورد بررسی قرار می‌گیرد. و در نهایت ژن‌های موجود در ماژول‌های موثر بر بقای کلی مورد بررسی تفسیر کارکردی[23] و تحلیل غنی‌سازی[24] قرار می‌گیرند.

مواد و روش‌ها

طراحی این پژوهش و مراحل انجام آن در نمودار گردش کار در شکل ۱ ارایه شده است. بر این اساس مراحل جمع‌آوری[25]، پیش‌پردازش[26]، و پالایش داده‌ها[27] به عنوان گام‌های پیش از ساخت شبکه هم‌بیانی تعریف شدند.



شکل ۱: نمودار گردش کار مراحل اجرای پژوهش

سه گام نخست مربوط به آماده‌سازی داده‌هاست. پس از آماده‌سازی با ساخت شبکه، ماژول‌های هم‌بیان شناسایی می‌شوند. سپس آنالیز بقاء به ازاء ماژول‌ها انجام می‌شود و پس از این مرحله با شناسایی ماژول-های موثر، ژن‌های درون ماژول‌های منتخب تفسیر کارکردی می‌شوند.

شبکه هم‌بیانی ژنی بر اساس داده‌های حاصل از اجرای سه مرحله قبلی ساخته و ماژول‌های هم‌بیان شناسایی شدند. پس از ساخت شبکه و شناسایی ماژول‌ها، آنالیز بقاء به

[16] Epidermal growth factor Receptor
[17] Tyrosine kinase
[18] High-throughput
[19] System Biology
[20] Network Approach
[21] External Trait
[22] Weighted gene Co-expression Network Analysis
[23] Functional Annotation Analysis
[24] Enrichment Analysis
[25] Data Collection
[26] Data Preprocessing
[27] Data Filtering

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

ازای یکایک ماژول‌های شناسایی شده انجام شد. در این راستا شناسایی ماژول‌های مرتبط با پیش‌آگهی[۲۸] به عنوان گام بعدی این مطالعه بود. نهایتا در گام پایانی، تحلیل آنتولوژی ژن و مسیرهای زیستی متناسب با ماژول‌های مرتبط با پیش‌آگهی انجام شد. در این پژوهش برای اجرای تمامی مراحل از نسخه ۳,۶,۱ پلتفرم R استفاده گردید.

مجموعه داده و پیش‌پردازش داده‌ها

در این مطالعه از داده‌های بیان ژن و داده‌های بالینی در قالب دو مجموعه داده استفاده شد. داده‌های بیان-ژن شامل ۵۸ نمونه سرطان پستان از زیرگروه Her2-enriched و داده‌های بالینی از اطلس ژنوم سرطان[۲۹] به‌دست آمد (۱۸، ۱۹).

فایل‌های شمارش خوانش خام که خروجی ابزار HTSeq-Count محسوب می‌شود برای نمونه‌های سرطان پستان با استفاده از پکیج "TCGAbiolinks" (۲۰) دانلود شد. این داده‌ها با استفاده از روش TPM[۳۰] نرمال‌سازی شدند. بررسی اثر دسته‌ای با استفاده از پکیج "MBatch" در تارنمای https://bioinformatics.mdanderson.org انجام شد. بر اساس نتایج حاصله از آنالیز آنلاین این تارنما، مجموعه داده سرطان پستان دارای اثر دسته‌ای قوی[۳۱] نبود. در ادامه با استفاده از دسته‌بندی زیرگروه‌های سرطان پستان ارائه شده در پیوست تکمیلی مقاله‌ی مطالعه گسترده بر روی داده‌های سرطان پستان اطلس ژنوم سرطان (۱۸) نمونه‌های Her2-enriched انتخاب شد. سپس، بیان نمونه‌ها در مبنای لگاریتم دو محاسبه و پس از این مرحله به‌منظور شناسایی داده‌های پرت[۳۲] خوشه‌بندی سلسله مراتبی بر روی نمونه‌ها اجرا شد (شکل ۲).

پکیج "dynamicTreeCut" برای هرس نمودن انطباقی شاخه‌های[۳۳] دندروگرام تولیدشده از خوشه‌بندی سلسله مراتبی استفاده گردید. در پی اجرای این مرحله دو نمونه شامل TCGA-A2-A0CY و TCGA-A8-A08X به عنوان داده پرت شناخته و حذف شدند. در این مطالعه ۵۰۰۰ ژن تولیدکننده‌ی پروتئین جهت بررسی در شبکه هم‌بیانی انتخاب شد. این ژن‌ها امکان بیشترین ارتباط[۳۴] به سایر ژن‌ها را داشتند. کیفیت نمونه‌ها و ژن‌ها با استفاده از تابع goodSamplesGenes در پکیج WGCNA مورد بررسی قرار گرفت به‌طوری که ژن‌های با واریانس صفر در تمامی نمونه‌ها حذف شدند.

ساخت شبکه هم‌بیان ژنی وزن‌دار

شبکه هم‌بیان ژنی زیرگروه Her2-enriched با استفاده از پکیج WGCNA ساخته شد. همبستگی بین هر جفت ژن با استفاده از تابع biweight midcorrelation (bicor) محاسبه شد. پایداری[۳۵] در برابر نویز علت استفاده از این روش در مقایسه با ضریب همبستگی پیرسون محسوب می‌شود (۲۱). در این مطالعه از رویکرد «هیبریدی علامت‌دار[۳۶]» برای ساخت شبکه هم‌بیانی استفاده شد. در این نوع از شبکه از آستانه نرم و سخت[۳۷] هم‌زمان استفاده می‌شود. ماتریس همبستگی[۳۸] از طریق تابع مجاورت در پکیج WGCNA و با توان بتا[۳۹] به ماتریس همجواری[۴۰] تبدیل می‌شود. به‌منظور دستیابی به خاصیت مقیاس آزاد[۴۱] می‌بایست مقدار بتا طوری تعیین شود که این ویژگی در شبکه حاصل شود. در شبکه هم‌بیانی Her2-enriched مجموعه‌ای بین ۱ تا ۲۰ برای بتا، مورد بررسی قرار می‌گیرد و در نهایت با در نظر گرفتن شاخص‌هایی مقدار بتا تعیین و بر اساس آن ماتریس مجاورت ساخته می‌شود.

در چارچوب WGCNA به منظور شناسایی خوشه‌های ژنی با بیشترین هم‌بیانی از نسخه تعمیم‌یافته‌معیار همپوشانی توپولوژیکی[۴۲] استفاده می‌شود. با استفاده از معیار همپوشانی توپولوژیکی علاوه بر محاسبه شباهت دو ژن، هم‌زمان شباهت ژن‌های متصل به دو ژن مذکور نیز محاسبه می‌شود. در این راستا با استفاده از تابع TOMsimilarity نهاده‌های ماتریس مجاورت به مقادیر جدیدی در ماتریس همپوشانی توپولوژیکی تبدیل می‌شوند. تمامی نهاده‌های این ماتریس مثبت بوده و خود ماتریس نیز متقارن است. از آنجا که عدم شباهت

[28] Prognostic Modules
[29] The Cancer Genome Atlas (TCGA)
[30] Transcripts Per Million
[31] Strong Batch Effect
[32] Outlier
[33] Adaptive Branch Pruning

[34] Most Connected Genes
[35] Robustness
[36] Signed Hybrid Network
[37] Hard and Soft Thresholding
[38] Correlation Matrix
[39] Power β
[40] Adjacency Matrix
[41] Scale-free Network
[42] Generalized Version of Topological Overlap Measure (TOM)

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

همپوشانی توپولوژیکی[43] باعث تفکیک بهتر خوشه‌ها از یکدیگر می‌شود، در چارچوب WGCNA به جای ماتریس TOM از ماتریس 1-TOM استفاده می‌شود. در گام بعدی خوشه‌بندی سلسله‌مراتبی داده‌های بیان ژن‌های Her2-enriched بر مبنای الگوریتم «پیوند میانگین[44]» و با استفاده از تابع hClust انجام می‌شود. بر این اساس خوشه‌های (ماژول‌های) ژنی استخراج می‌شوند. بر حسب تعریف، ماژول در شبکه عبارت است از زیرمجموعه‌ای از نودها که زیرشبکه‌هایی را در ذیل شبکه اصلی تشکیل می‌دهند. در اینجا ماژول‌های ژنی عبارت از گروه‌هایی از ژن‌هاست که دارای بیان مشابه هستند. در این بخش نیز هرس شاخه‌های دندروگرام با استفاده از تابع cutreeDynamic انجام می‌شود. در تابع مذکور و در این مطالعه آرگومان‌های "deepSplit" و " minimum cluster size" به ترتیب برابر با ۲ و ۵۰ در نظر گرفته شد. مطابق با پیش‌فرض چارچوب WGCNA ماژول‌های مختلف شناسایی‌شده هر کدام با یک رنگ شناخته می‌شوند. و بر اساس مطالعات صورت گرفته هر ماژول به مجموعه‌ای از فرایندهای زیستی نگاشت می‌شود (۲۲). در چارچوب WGCNA یکی از متغیرهای مهمی که تولید می‌شود Module Eigengene (ME) است که معادل با مقادیر[45] ویژه در تحلیل مولفه اصلی[46] است. آنچه در اینجا به عنوان ME در هر ماژول مورد استفاده قرار می‌گیرد نخستین مولفه اصلی (PC1) است که بیشترین درصد واریانس مقادیر بیان تمامی ژن‌ها را در درون یک ماژول بیان می‌نماید. با استفاده از ME به جای مقادیر یکایک بیان ژن‌های درون هر ماژول یک نماینده وجود دارد که می‌توان از آن در محاسبات آتی استفاده نمود. ME برای هر ماژول با استفاده از تابع moduleEigengenes محاسبه شد. از آن جا که میزان ME برخی ماژول‌ها بسیار به هم نزدیک است می‌توان بر اساس تعیین حد آستانه و با استفاده از تابع mergeCloseModules آن‌ها را ادغام نمود.

شناسایی ماژول‌های مطلوب

پس از شناسایی ماژول‌ها آنچه مهم و ضروری است تعیین ارتباط بین ماژول‌ها و ویژگی‌های بالینی است که در چارچوب WGCNA با استفاده از ME هر ماژول انجام می‌شود. بدین صورت که با استفاده از تابع cor که بیانگر ضریب همبستگی پیرسون است، میزان همبستگی بین ME متعلق به هر ماژول با صفات و ویژگی‌های بالینی محاسبه می‌شود. در این مطالعه ویژگی‌های بالینی عبارتند از سایز تومور[47]، تعداد لنف نودهای درگیر با تومور[48]، مرحله تومور بیمار[49]، سن تشخیص تومور[50]، نوع بافت[51]، وضعیت یائسگی[52]. همچنین وضعیت بقای کلی[53]، زمان بقای کلی[54]، وضعیت بازه‌ی عاری از پیشرفت تومور[55] زمان مربوط به آن[56] از دیگر صفات بالینی مورد بررسی در این مطالعه بوده است.

تحلیل بقاء

چندین پکیج از جمله Survival (23)، Survminer (24) و RegParallel (25) برای بررسی ارتباط ماژول‌های هم‌بیانی شناسایی‌شده با عامل بقاء مورد استفاده قرار گرفت. در این قسمت نیز از ME به عنوان نماینده هر ماژول برای محاسبه ارتباط بین ماژول‌ها و بقای کلی مورد استفاده قرار گرفت. مقادیر حاصل از ME برای ماژول‌ها به دو دسته مثبت و منفی تقسیم شدند (۱۲). در این راستا رگرسیون کاکس تک‌متغیره و نمودار k-M برای هر ماژول با نمایندگی ME آن ماژول اجرا و در ادامه ماژول‌هایی که پی-مقدار[57] آن‌ها در تحلیل کاکس کمتر از ۰/۰۵ بود به عنوان ماژول‌های معنادار مرتبط با عامل بقای کلی انتخاب شدند. در گام پایانی نیز تحلیل بقای تک‌ژنی[58] برای ژن‌های داخل ماژول‌های مرتبط با پیش‌آگهی انجام شد.

آنتولوژنی ژن و تحلیل مسیرهای زیستی

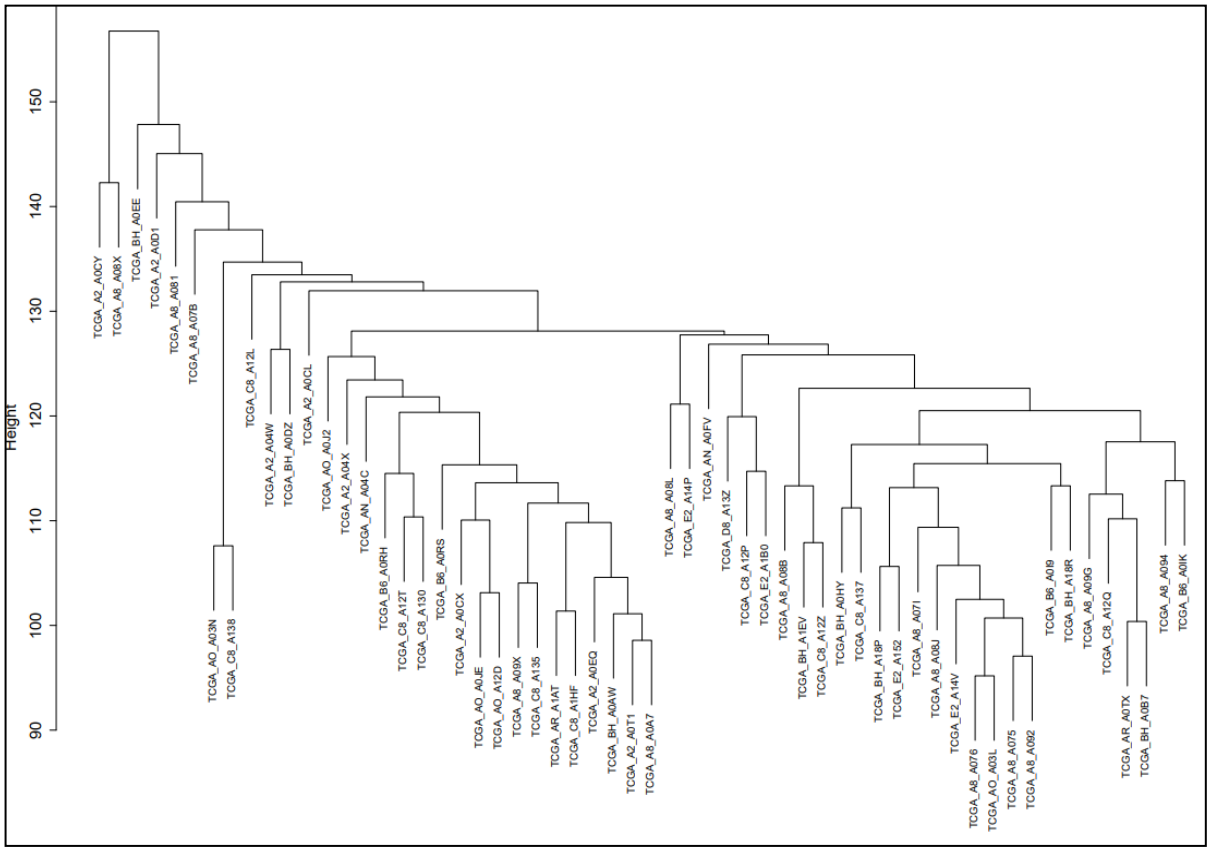
تحلیل غنی‌سازی ژنی[59] بر روی ژن‌های درون ماژول‌های منتخب با استفاده از ابزار DAVID[60] (26) انجام شد. بر مبنای خروجی ابزار DAVID آنتولوژی ژن و تحلیل

[43] TOM-based dissimilarity
[44] Average Linkage
[45] Eigenvalue
[46] Principal Component Analysis (PCA)

[47] Tumor Size
[48] Node
[49] AJCC Stage
[50] Age of Diagnosis
[51] Histological Type
[52] Menopause Status
[53] Overall Survival (OS) Status
[54] Overall Survival (OS) Time
[55] Progression Free Interval (PFI) Status
[56] Progression Free Interval (PFI) Time
[57] P-value
[58] Single-gene Survival Analysis
[59] Gene Enrichment Analysis
[60] Database for Annotation, Visualization, and Integrated Discovery (DAVID)

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

مسیرهای زیستی برای ماژول‌های منتخب مورد بررسی قرار گرفت. ما در تحلیل مسیرهای زیستی[61] دانشنامه ژن‌ها و ژنوم‌های کیوتو[62] و پایگاه داده ری‌اکتوم[63] را مورد بررسی قرار دادیم. پی- مقدار کمتر از ۰/۰۵ و نرخ کشف اشتباه[64] کمتر از ۰/۰۵ برای تشخیص معناداری فرایندها و مسیرهای زیستی به عنوان حد آستانه تعریف شد.



شکل ۲: خوشه‌بندی سلسله مراتبی ۵۸ نمونه Her2-enriched برای تعیین نمونه‌های پَرت

یافته‌ها

شش ماژول در شبکه هم‌بیانی Her2-enriched

همان‌طور که پیش از این بیان شد برای ساخت شبکه هم‌بیانی داده‌های بیماران Her2-enriched از چارچوب WGCNA استفاده شد. در این راستا و با استفاده از آستانه‌ی نرم، شبکه هم‌بیانی Her2-enriched ساخته و شش ماژول در آن شناسایی شد. در این میان پارامتر بتا نقش مهمی در ایجاد خاصیت مقیاس آزاد برای شبکه هم‌بیانی ایفا می‌نماید. بر این اساس ابتدا مجموعه‌ای از مقادیر ممکن برای بتا تولید شد (جدول ۱). در انتخاب مقدار بتا باید موارد ذیل را مدنظر قرار داد:

۱- $R^2 > ۰/۸$

۲- بیشینه و میانگین درجه در شبکه

۳- شیب خط برازش برای نمودار $\log_{10}(k)$ در برابر $\log_{10}(P(k))$ باید تقریباً برابر ۱- باشد.

جدول ۱ پارامترهای مورد بررسی برای انتخاب توان «بتا» مناسب به‌منظور ساخت شبکه هم‌بیان ژنی وزندار. ستون Power بیانگر مقادیر مختلف بتا است. همچنین k بیانگر ارتباط است به‌طوری که mean.k. بیانگر مقدار میانگین ارتباط و max.k. بیانگر بیشنیه ارتباط است.

بر این اساس و به منظور دستیابی به جواب‌های قابل قبول، مقدار بتا برابر با ۸ در نظر گرفته شد. با انتخاب بتا برابر با ۸ مقدار R^2 برابر با ۰/۹۳ به‌دست آمد. در شکل شماره ۳ و در نمودار (الف) ارتباط بین R^2 و مقادیر مختلف بتا و در نمودار (ب) ارتباط بین مقادیر مختلف بتا با میانگین درجه در شبکه Her2-enriched ارائه شده

[61] Pathway Analysis
[62] Kyoto Encyclopedia of Genes and Genomes (KEGG)
[63] Reactome Database
[64] False Discovery Rate (FDR

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

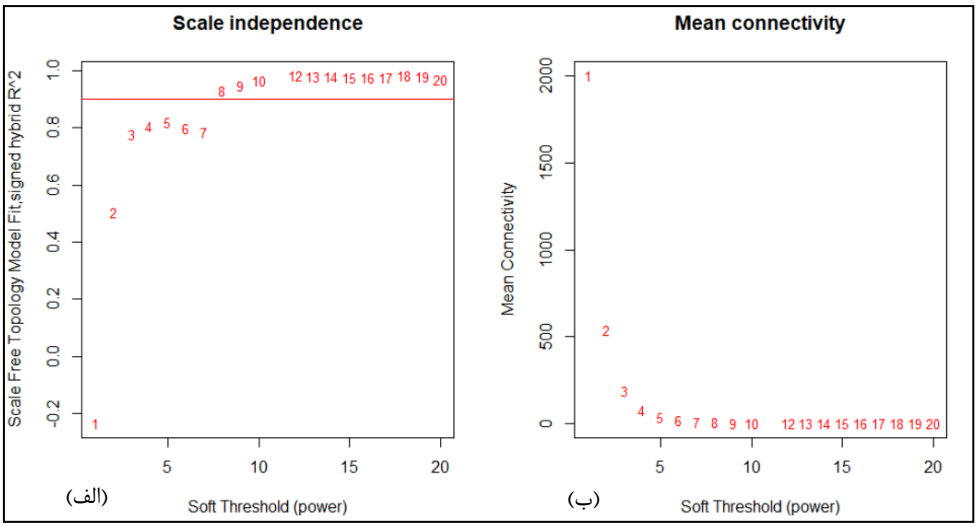
است. در ادامه با استفاده از ماتریس بیان ژن و مقدار بتای انتخاب شده ماتریس مجاورت ساخته و سپس با استفاده از ماتریس همپوشانی توپولوژیکی، خوشه‌بندی سلسله‌مراتبی انجام شد. با مطالعه ماژول‌ها مشخص شد برخی از MEهای ماژول‌ها بسیار به هم نزدیک بوده که نشان از شباهت آن‌ها با هم دارد. به همین دلیل ماژول‌هایی که تنها ۲۰٪ با هم تفاوت داشتند در هم ادغام شدند. در بالای شکل شماره ۴ دندروگرام حاصل از خوشه‌بندی Her2-enriched و در پایین آن دو سطح ماژول‌های شکل گرفته در قبل و بعد از فرایند ادغام نمایش داده شده است.

رنگ‌ها در شکل شماره ۴ بیانگر ماژول‌های ژنی شناخته شده هستند. همان‌طور که در این شکل ملاحظه می‌شود ماژول‌های شناسایی شده در گام نخست ۷ ماژول بود که پس از فرایند ادغام به ۶ ماژول کاهش پیدا نمود. مجموعه‌ای از ژن‌ها که اساسا مدل نتوانسته آن‌ها را شناسایی نماید و در هیچ یک از ماژول‌ها دسته‌بندی نشده‌اند در ماژول خاکستری قرار می‌گیرند.

جدول ۱: پارامترهای مورد بررسی برای انتخاب توان «بتا» مناسب به منظور ساخت شبکه هم‌بیان ژنی وزندار

Power	R.square	Slope	mean.k.	median.k.	max.k.
1	0.24	1.01	2001.55	1858.18	4410.73
2	0.50	-1.49	541.87	443.57	1927.70
3	0.78	-2.02	188.37	129.54	997.66
4	0.81	-2.18	77.81	43.24	585.72
5	0.82	-2.18	36.58	16.35	368.06
6	0.80	-2.15	19.04	6.82	242.64
7	0.78	-2.04	10.76	3.11	165.82
8	0.93	-1.73	6.51	1.51	121.47
9	0.95	-1.72	4.18	0.77	102.79
10	0.96	-1.67	2.82	0.41	88.05
12	0.98	-1.58	1.45	0.13	66.40
13	0.98	-1.54	1.09	0.08	58.69
14	0.98	-1.51	0.85	0.04	52.33
15	0.98	-1.47	0.67	0.03	46.90

ستون **Power** بیانگر مقادیر مختلف بتا است. همچنین **k** بیانگر ارتباط است به طوری که **mean.k.** بیانگر مقدار میانگین ارتباط و **max.k.** بیانگر بیشنیه ارتباط است.

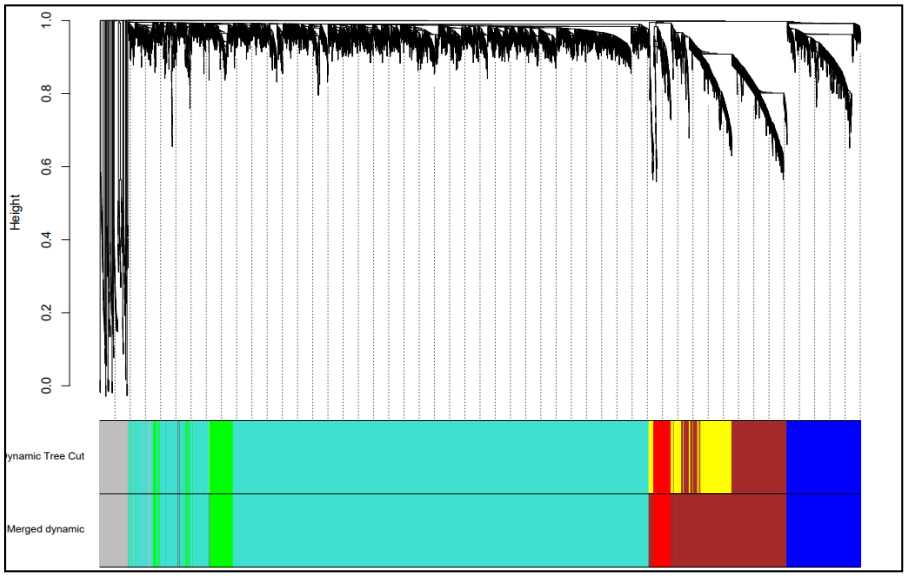


شکل ۲: نتایج بررسی و تحلیل توپولوژی شبکه Her2-enriched بر اساس مقادیر مختلف بتا

الف: در این نمودار محور افقی مقادیر مختلف بتائی که مورد بررسی قرار گرفت را نشان می‌دهد و در محور عمودی شاخص برازش R نشان داده شده است. حد آستانه ۰/۹ برای انتخاب بتا در نظرگرفته شده که با خط افقی قرمز مشخص است.

ب: این نمودار میانگین ارتباطات در محور عمودی و مقادیر بتا در محور افقی را نشان می‌دهد. در انتخاب بتای مناسب، مقدار میانگین ارتباطات یکی از شاخص‌هائی است که برای ایجاد شبکه می‌بایست مورد توجه قرار گیرد.

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]



شکل ۳: دندروگرام ژنی Her2-eriched

ماژول‌ها مطابق با شاخه‌های دندروگرام از یکدیگر تفکیک می‌شوند. رنگ‌های مختلف بیانگر خوشه‌های شناسایی‌شده در فرایند خوشه‌بندی است. در ذیل دندروگرام دو سطح رنگی به همراه نمایش ماژول‌ها در قبل و بعد از فرایند ادغام نمایش داده شده است. سطح رنگی بالا (Dynamic Tree Cut) وضعیت ماژول‌ها قبل از فرایند ادغام و سطح رنگی پایین (dynamic Merged) وضعیت ماژول‌ها را بعد از ادغام نشان می‌دهد که تعداد ماژول‌ها بر اساس حد آستانه شباهت ۲۰٪ از هفت به شش ماژول کاهش پیدا نمود.

شناسایی ماژول‌های موثر بر پیش‌آگهی

ارتباط بین ماژول‌ها و ویژگی‌های بالینی: امکان بررسی ارتباط بین پروفایل بیان ژن بیماران و صفات بالینی ایشان از مزیت‌های WGCNA محسوب می‌شود. بدین منظور همبستگی ME هر ماژول با اطلاعات بالینی بیماران مورد مطالعه و بررسی قرار می‌گیرد. در این پژوهش مجموعا ۱۰ ویژگی بالینی و پاتولوژیکی از بیماران مبتلا به Her2-enriched اطلس ژنوم سرطان استخراج و همبستگی آن‌ها با ME هر ماژول با استفاده از ضریب پیرسون محاسبه شد. نتیجه محاسبه‌ی این همبستگی در شکل شماره ۵ ارائه شده است. در این راستا، ماژول سبز با «سن بیماران در زمان تشخیص تومور» رابطه‌ی مثبت (۰/۲۹) و با پی–مقداری برابر با ۰/۰۳ رابطه معنی‌دار داشت. ماژول قهوه‌ای با متغیرهای «تعداد لنف نودهای درگیر با تومور»، «مرحله تومور بیمار»، و «وضعیت بازه عاری از پیشرفت تومور» به ترتیب با پی–مقدارهای ۰/۰۳، ۰/۰۵ و ۰/۰۳ رابطه معنادار داشت. همچنین ماژول قرمز با متغیر «تعداد لنف نودهای درگیر با تومور» رابطه مثبت (۰/۴۲) و پی-مقدار ۰/۰۰۱ رابطه م عناداری داشت.

تحلیل بقاء: از مدل رگرسیون کاکس برای تعیین تاثیر هر ماژول به عنوان متغیر مستقل بر زمان و رویداد بقای کلی به عنوان متغیرهای وابسته در بیماران مبتلا به سرطان استفاده شد. بر این اساس، از ME هر ماژول به عنوان نماینده در رگرسیون کاکس استفاده کردیم که از میان ماژول‌ها دو ماژول فیروزه‌ای و سبز با پی- مقدارهای به ترتیب برابر با ۰/۰۰۸ و ۰/۰۳ در ارتباط با بقای کلی معنادار شناخته شدند (P<=0.05).

در جدول شماره ۲ نتایج آنالیز کاکس ارائه شده است. بر مبنای آنچه در جدول ۲ ارائه شده است مقدار نسبت مخاطره[۶۵] برای ماژول‌های فیروزه‌ای و سبز به ترتیب برابر با ۲/۵۹ و ۲/۰۸ است. در نتیجه افزایش بیان ژن‌های موجود در این دو ماژول باعث پیش‌آگهی بد[۶۶] نسبت به بقای کلی در بیماران گروه Her2-enriched می‌شود. به عبارت بهتر افزایش بیان ژن‌های ماژول فیروزه‌ای و

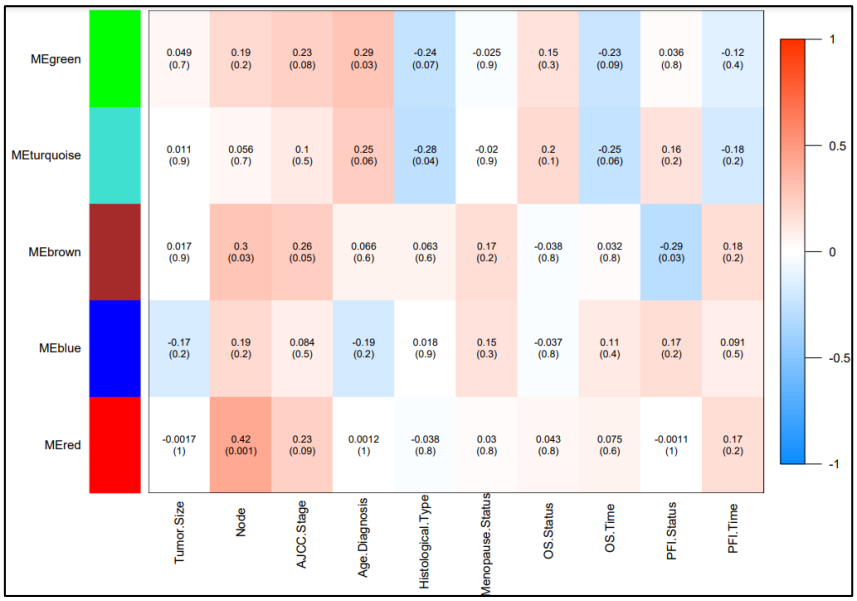
[65] Hazard Ratio (HR)
[66] Poor Prognosis

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

سبز باعث کاهش بقای بیماران این زیرگروه از سرطان پستان خواهد شد. در ادامه بر اساس تحلیل تک‌متغیره بر روی ژن‌های موجود در دو ماژول مذکور مشخص شد که به‌ترتیب ۳۰٪ و ۳۹٪ از ژن‌های ماژول سبز و فیروزه‌ای در ارتباط با بقای کلی با آستانه ۰/۰۵ معنادار هستند (پیوست تکمیلی ۱).

تحلیل غنی‌سازی عملکردی ژن‌های ماژول پیش‌آگهی

در این مطالعه از بین دو ماژول مذکور به دلیل «نسبت مخاطره» بالاتر، ماژول فیروزه‌ای برای تجزیه و تحلیل غنی‌سازی ژنی انتخاب شد. این مرحله از آنالیز که با استفاده از ابزار DAVID انجام شد فرایندها و مسیرهای زیستی معنادار و مرتبطی با ژن‌های موجود در این ماژول که پی-مقدار آن‌ها در آنالیز کمتر از ۰/۰۱ بود، بررسی شد. همان‌طور که در جدول شماره ۳ مشاهده می‌شود مجموعه ژن‌های درون این ماژول به‌طور معناداری با فرایندهای زیستی همچون "transcription, DNA-templated"، "protein sumoylation"، "mRNA export from nucleus" و "intracellular protein transport" غنی‌سازی شده‌اند. همچنین بر اساس پایگاه KEGG مسیرهای زیستی "Ubiquitin mediated proteolysis" و "RNA degradation" در ارتباط با ژن‌های موجود در ماژول فیروزه‌ای معنادار شدند.



شکل ۴: ارتباط بین ماژول‌ها و متغیرهای بالینی

در هر سلول از نمودار دو عدد مشاهده می‌شود که عدد بالایی بیانگر میزان همبستگی و عدد داخل پرانتز پی-مقدار ناشی از محاسبه هبستگی بین ME و صفات بالینی است. مقادیر مختلف همبستگی با طیف رنگی مابین قرمز (همبستگی مثبت) و آبی (همبستگی منفی) مدرج شده است. به طور نمونه ماژول قرمز با «تعداد لنف نودهای درگیر با تومور» همبستگی معادل با ۰/۴۲ دارد که پی-مقدار ۰/۰۰۱ این ارتباط را معنادار نموده است.

جدول ۲: نتایج اجرای کاکس بر روی ME ماژول‌های شناسایی شده

Module Name	No.Genes	Overall Survival		
		HR	P-value	CI
Brown	796	1.05	0.9	0.59-1.87
Turquoise	3178	2.59	0.008	1.23-5.47
Blue	485	0.89	0.7	0.49-1.60
Green	213	2.08	0.03	1.04-4.14
Red	117	1.14	0.7	0. 63-2.05
Grey	211	-	-	-

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

جدول ۳: عناوین تفسیر کارکردی در ماژول فیروزه‌ای

Functional Annotation Term	Count	FDR	Functional Annotation Term	Count	FDR
GO-Biological Process (BP)			**GO-Molecular Function (MF)**		
protein sumoylation	19	2.39E-06	protein binding	315	2.30E-11
nuclear-transcribed mRNA poly(A) tail shortening	8	0.006	poly(A) RNA binding	69	7.12E-08
mRNA export from nucleus	13	0.006	DNA binding	74	0.003
DNA duplex unwinding	9	0.006	nucleotide binding	25	0.004
transcription, DNA-templated	82	0.006	ATP binding	67	0.004
intracellular transport of virus	9	0.012	helicase activity	11	0.007
intracellular protein transport	19	0.012	ATP-dependent helicase activity	7	0.007
tRNA export from nucleus	7	0.032	ubiquitin protein ligase binding	21	0.007
regulation of glucose transport	7	0.034	chromatin binding	25	0.010
mRNA splicing, via spliceosome	17	0.042	ATP-dependent DNA helicase activity	7	0.010
protein transport	24	0.047	nucleic acid binding	46	0.016
GO-Cellular Component			DNA helicase activity	6	0.016
nucleoplasm	168	1.65E-26	ligase activity	19	0.016
cytoplasm	200	4.02E-09	RNA binding	30	0.016
nucleus	203	1.29E-08	nuclear localization sequence binding	6	0.027
nucleolus	50	1.01E-05	transcription coactivator activity	17	0.036
nuclear pore	12	1.56E-04	ATP-dependent RNA helicase activity	8	0.051
membrane	91	1.89E-04	zinc ion binding	49	0.055
cytosol	118	0.005	**KEGG Pathway**		
nuclear membrane	17	0.014	Ubiquitin mediated proteolysis	13	0.020
chromosome	11	0.016	RNA degradation	11	0.005

بحث

در این مطالعه از WGCNA برای تحلیل داده‌های بیان mRNA به‌منظور شناسایی ماژول‌های هم‌بیانی زیرگروه Her2-enriched سرطان پستان و ارتباط آن‌ها با صفات و ویژگی‌های بالینی استفاده شد. نمونه‌های سرطان پستان و برخی از زیرگروه‌های آن با استفاده از رویکرد زیست‌شناسی سامانه‌ای مورد بررسی قرار گرفته‌اند. زیرگروه Her2-enriched از جمله زیرگروه‌های سرطان پستان بود که با رویکرد شبکه‌ای مورد تحلیل قرار نگرفته بود. بدین منظور الگوریتم WGCNA به‌عنوان یکی از چارچوب‌هایی که مبتنی بر روش‌های بدون نظارت اجرا می‌شود در این مطالعه مورد استفاده قرار گرفت. امکان بررسی همبستگی میان پروفایل بیان ژن نمونه‌ها با ویژگی‌های بالینی آن‌ها از مزایای WGCNA محسوب می‌شود. در این مطالعه ۶ ماژول از بین ۵۰۰۰ ژن با بیشترین درجه ارتباط با سایر ژن‌ها شناسایی شد که یکی از آن‌ها خاکستری و غیرقابل استفاده بود. در تحلیل انجام شده با استفاده از رگرسیون کاکس، دو ماژول فیروزه‌ای و سبز ارتباط معناداری با بقای کلی داشته و به‌عنوان ماژول‌های موثر بر پیش‌آگهی بقای کلی شناسایی شدند. بر اساس «نسبت مخاطره» به‌دست آمده برای این دو ماژول مشخص شد افزایش بیان ژن‌های این دو ماژول باعث کاهش بقای کلی بیماران زیرگروه Her2-enriched می‌شود. با توجه به مقدار بالاتر «نسبت مخاطره» در ماژول فیروزه‌ای، ژن‌های درون این ماژول در ارتباط با فرایندها و مسیرهای زیستی مورد تحلیل و بررسی قرار گرفتند. بر اساس آنالیز کاکس برای ژن‌های درون ماژول فیروزه‌ای، به‌ترتیب ژن‌های *CHAMP1*، *PPP1R26*، *PRRC2B*، *KANSL3* و *ANAPC2* به عنوان ۵ ژن مهم شناسایی شدند.

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

بر مبنای تحلیل آنتولوژی ژنی بیشتر ژن‌های ماژول فیروزه‌ای با فرایندهای زیستی همچون transcription, DNA-templated (GO:0006355)، protein transport (GO:0015031)، intracellular protein transport (GO:0006886)، protein sumoylation (GO:0016925)، mRNA export from nucleus (GO:0006406) ارتباط معنادار داشته‌اند. همچنین ژن‌های این ماژول ارتباط معناداری با مسیرهای زیستی همچون Ubiquitin mediated proteolysis (hsa04120) و RNA degradation (hsa03018) داشته است.

در ادامه به بررسی ۵ ژن مهم در این آنالیز می‌پردازیم. *CHAMP1* در ارتباط میکروتوبول‌ها[۶۷] به کینتوکور[۶۸] نقش دارد و مطالعات نشان داده است که در ایجاد برخی از ناهنجاری‌های ژنتیکی نقش اساسی دارد و تاکنون نقش این پروتئین در ایجاد سرطان بررسی شده است. اما با توجه به نقش مهم آن در تقسیم سلول می‌تواند کاندیدای مطالعات آتی در زمینه سرطان باشد.

PPP1r26 یک زیرواحد تنظیم‌کننده‌ی پروتئین فسفاتاز است. مطالعات قبلی نشان داده است که این ژن در سرطان پستان افزایش بیان دارد اما عملکرد آن نامشخص است (۲۷). کاهش در این پروتئین از متاستاز سرطان پستان به ریه جلوگیری می‌کند. در این مطالعه نیز بیان این پروتئین باعث کاهش بقاء بیماران شده است. در نتیجه این پروتئین می‌تواند به عنوان یک هدف دارویی در نظر گرفته شود.

موتاسیون‌های ژن *PRRC2B* در سرطان لنفوم سلول‌های T[۶۹] نشان داده شده است (۲۸). همچنین ارتباط این ژن به ژن *NUP214* در لوسمی حاد میلوئیدی[۷۰] مشاهده شده است.

KANSL3 نیز یک پروتئین تنظیمی مرتبط با هیستون‌ها است. این ژن در بیماران مبتلا به سرطان پستان که در ادامه سلول‌های سرطانی ایشان به مغز متاستاز داده‌اند تفاوت بیان جدی داشته است (۲۹). مطالعات محدودی که بیانگر حضور این ژن‌ها در سرطان‌های مختلف است، تاییدیه‌ای بر اهمیت مطالعه این ژن‌ها در سرطان پستان است.

اما به طور کلی مطالعات کمی در مورد این ژن‌ها در سرطان انجام شده است. لذا مطالعه حاضر تاییدی بر پتانسیل بالقوه این ژن‌ها در تعیین پیش‌آگهی زیرگروه Her2-enriched سرطان پستان و همچنین معرفی این ژن‌ها به عنوان گزینه مناسب برای درمان هدفمند در سرطان پستان است.

بعضی از مطالعات که در زمینه استفاده از شبکه‌های هم بیان ژنی و سرطان پستان انجام شده است به شرح زیر است. همان‌طور که ملاحظه می‌شود تاکنون مطالعه در مورد این شبکه در گروه Her2-enriched انجام نشده است. لیگانگ بائو و همکاران با استفاده از چارچوب WGCNA ژن‌های موثر بر پیش‌آگهی بیماران سرطان پستان سه‌گانه منفی را بررسی و معرفی نمودند. بر این اساس بیان بالای *GIPC1*، *HES6*، *KIAA1543*، *MYLK2*، و *PPAN* منجر به بقای کلی بد[۷۱] در این دسته از بیماران سرطان پستان شد (۳۰). کالین کِلرک و همکاران در یک مطالعه گسترده ارتباط ماژول‌های مختلف زیرگروه‌های luminal A، luminal B، HER2+، و basal-like را با صفات بالینی از جمله بقای کلی مورد بررسی قرار دادند. این مطالعه نشان داد افزایش بیان ژن *PDLIM4* بیانگر پیش‌آگهی خوب در زیرگروه‌های luminal A و luminal B است در صورتی که این ژن نقش متضادی در پیش‌آگهی زیرگروه basal-like دارد. *FOXM1* یکی از ژن‌هایی است که در این مطالعه نشان داده شد پیش‌آگهی بد بر روی بیماران زیرگروه luminal B دارد. ژن *CD53* از دیگر یافته‌های مطالعه مذکور محسوب می‌شود که رابطه معناداری با پیش‌آگهی خوب هم در زیرگروه HER2+ و هم در زیرگروه basal-like داشت (۳۱). هوئیجی شی و همکاران نیز در مطالعه‌ای با استفاده از WGCNA داده‌های سرطان پستان را فارغ از زیرگروه‌های آن بررسی نمودند. از میان ماژول‌هایی که در این مطالعه ساخته شد تنها یک ماژول، ارتباط معنی‌داری با متغیر «بقاء بدون عود[۷۲]» داشت. ایشان ۱۰ ژن کلیدی شامل *CCNB2*، *UBE2C*، *PRC1*، *CDC20*، *ASPM*، *FOXM1*، *KIF4A*، *NUSAP1*، *PTTG1* و *CEP55* را در ماژول منتخب انتخاب و آن‌ها را با رویکرد بقای تک‌ژنی تحلیل نمودند.

[67] Microtubule
[68] Kinetochore
[69] T Cell Lymphoma
[70] Acute myeloid leukemia (AML)
[71] Poor Overall Survival
[72] Relapse Free Survival

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

نتایج این تحلیل نشان داد تمامی این ژن‌ها با «بقاء بدون عود» همبستگی معناداری داشتند (۳۲).

نتیجه‌گیری

در این مطالعه داده‌های بیان ژن زیرگروه Her2-enriched از بیماران مبتلا به سرطان پستان با استفاده از چارچوب WGCNA مورد بررسی قرار گرفت. در این تحلیل همبستگی ۵ ماژول شناسایی شده با متغیرهای بالینی و پاتولوژیکی مورد بررسی قرار گرفت. دو ماژول فیروزه‌ای و سبز با بقای کلی ارتباط موثر داشتند که برای ادامه تحلیل ماژول فیروزه‌ای انتخاب شد. مطالعه ما نشان داد افزایش بیان ژن‌های ماژول فیروزه‌ای باعث پیش‌آگهی بد و کاهش بقای کلی بیمارن این زیرگروه از سرطان پستان می‌شود. در ماژول فیروزه‌ای به ترتیب ژن‌های *CHAMP1*، *PPP1R26*، *PRRC2B*، *KANSL3* و *ANAPC2* به عنوان ۵ ژن مهم در کاهش بقای کلی شناسایی شدند. در این مطالعه فهرستی از ژن‌ها استخراج شد که می‌تواند در آنالیزهای آزمایشگاهی مورد استفاده قرار گیرد. بر این اساس پژوهش بر روی مکانیسم عملکرد ژن‌های شناسایی شده و همچنین مطالعه‌ی یکپارچه‌ی داده‌های اُمیک Her2-enriched شامل اپی‌ژنومیک، ترنسکریپتومیک، پروتئومیک، و متابولیک از گام‌های بعدی این مطالعه محسوب می‌شود.

تشکر و قدردانی

این مقاله مستخرج از رساله دکتری مهندسی کامپیوتر با عنوان «مقایسه گروه سرطان سارکوم با کارسینوما به منظور استخراج الگوهای ژنوم مشابه با استفاده از رویکرد زیست‌شناسی سامانه‌ای» مصوب پژوهشکده برق و فناوری اطلاعات سازمان پژوهش‌های علمی و صنعتی ایران است. این رساله به صورت مشترک در پژوهشکده سرطان معتمد جهاد دانشگاهی اجرا شد. از تمامی افرادی که در انجام این پژوهش ما را یاری نموده‌اند، تشکر می‌شود.

تعارض منافع

نویسندگان اعلام می‌دارند که هیچ تعارض منافعی در پژوهش حاضر وجود ندارد.

References

1. Global cancer statistics 2020 [Internet]. International Agency for Research on Cancer. 2020. Available from: https://gco.iarc.fr/today.
2. Makki J. Diversity of breast carcinoma: histological subtypes and clinical relevance. Clinical Medicine Insights:Pathology. 2015; 8:CPath-S31563.
3. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer. 2007; 109(1):25-32.
4. Aysola K, Desai A, Welch C, Xu J, Qin Y, Reddy V, Matthews R, Owens C, Okoli J, Beech DJ, Piyathilake CJ. Triple negative breast cancer– an overview. Hereditary genetics: current research. 2013; 2013(Suppl 2).
5. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA. Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical cancer research. 2007; 13(15):4429-34.
6. Plevritis SK, Munoz D, Kurian AW, Stout NK, Alagoz O, Near AM, Lee SJ, Van Den Broek JJ, Huang X, Schechter CB, Sprague BL. Association of screening and treatment with breast cancer mortality by molecular subtype in US women, 2000-2012. Jama. 2018; 319(2):154-64.
7. Vaz-Luis I ,Winer E, Lin N. Human epidermal growth factor receptor-2-positive breast cancer: does estrogen receptor status define two distinct subtypes? Annals of Oncology. 2013; 24(2): 283-91.
8. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008; 9(1): 1-3

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

9. Zhang H, Zhao X, Wang M, Ji W. Key modules and hub genes identified by coexpression network analysis for revealing novel biomarkers for larynx squamous cell carcinoma. Journal of Cellular Biochemistry. 2019; 120(12): 19832-40.
10. Tian H, Guan D, Li J. Identifying osteosarcoma metastasis associated genes by weighted gene co-expression network analysis (WGCNA). Medicine. 2018; 97(24).
11. Qiu J, Du Z, Wang Y, Zhou Y, Zhang Y, Xie Y, Lv Q. Weighted gene co-expression network analysis reveals modules and hub genes associated with the development of breast cancer. Medicine. 2019; 98(6).
12. Zhang C, Sun Q. Weighted gene co-expression network analysis of gene modules for the prognosis of esophageal cancer. Journal of Huazhong University of Science and Technology [Medical Sciences]. 2017; 37(3):319-25.
13. Zhang X, Feng H, Li Z, Li D, Liu S, Huang H, Li M. Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis. OncoTargets and therapy. 2018; 11:6001.
14. Wang T, Wu B, Zhang X, Zhang M, Zhang S, Huang W, Liu T, Yu W, Li J, Yu X. Identification of gene coexpression modules, hub genes, and pathways related to spinal cord injury using integrated bioinformatics methods. Journal of cellular biochemistry. 2019; 120(5): 6988-97.
15. Zhou X, Zhen X, Liu Y, Cui Z, Yue Z, Xu A, Han J. Identification of Key Modules, Hub Genes, and Noncoding RNAs in Chronic Rhinosinusitis with Nasal Polyps by Weighted Gene Coexpression Network Analysis. BioMed Research International. 2020 Jan 23; 2020.
16. Guo SM, Wang JX, Li J, Xu FY, Wei Q, Wang HM, Huang HQ, Zheng SL, Xie YJ, Zhang C. Identification of gene expression profiles and key genes in subchondral bone of osteoarthritis using weighted gene coexpression network analysis. Journal of cellular biochemistry. 2018; 19(9): 7687-95.
17. Li W, Wang L, Wu Y, Yuan Z, Zhou J. Weighted gene co-expression network analysis to identify key modules and hub genes associated with atrial fibrillation. International journal of molecular medicine. 2020; 45(2):401-16.
18. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490(7418):61.
19. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell. 2018; 173(2):400-16.
20. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, Noushmehr H, Colaprico A, Papaleo E. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS computational biology. 2019; 15(3): e1006701.
21. Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC bioinformatics. 2012; 13(1): 1-21.
22. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005; 4(1).
23. Therneau TM, Grambsch PM. Statistics for biology and health. Modeling survival data: extending the Cox model: Springer-Verlag. 2000; 87-152.
24. Kassambara A, Kosinski M, Biecek P, Fabian S. Drawing Survival Curves using "ggplot2". R package version 0.2. 2018;2.
25. Blighe K. RegParallel: Standard regression functions in R enabled for parallel processing over large data-frames. bioconductor. 2019.
26. Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44.
27. Zheng T, Lu M, Wang T, Zhang C, Du X. NRBE3 promotes metastasis of breast cancer by down-regulating E-cadherin expression. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2018; 1865(12): 1869-77.
28. Donner I, Katainen R, Kaasinen E, Aavikko M, Sipilä LJ, Pukkala E, Aaltonen LA. Candidate susceptibility variants in angioimmunoblastic T-cell lymphoma. Familial cancer. 2019; 18(1): 113-9.
29. Mamoor S. The KAT8 regulatory NSL subunit complex subunit 2 (KANSL2) is over-expressed in brain metastatic breast cancer. 2020.
30. Bao L, Guo T, Wang J, Zhang K, Bao M. Prognostic genes of triple-negative breast cancer identified by weighted gene co-expression network analysis. Oncology letters. 2020; 19(1):127-38.

[Downloaded from ijbd.ir on 2028-01-05]
[DOR: 20.1001.1.17359406.1400.14.1.7.4]
[DOI: 10.30699/ijbd.14.1.49]

31. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis. 2013; 34(10):2300-8.

32. Shi H, Zhang L, Qu Y, Hou L, Wang L, Zheng M. Prognostic genes of breast cancer revealed by gene co-expression network analysis. Oncology letters. 2017; 14(4): 4535-42.

[Downloaded from ijbd.ir on 2028-01-05]

[DOR: 20.1001.1.17359406.1400.14.1.7.4]

[DOI: 10.30699/ijbd.14.1.49]