

تجزیه و تحلیل درون رایانه ای برخی تغییرات مولکولی بالقوه در تومورهای پستان در مقایسه

با بافت های طبیعی مجاور

الهه زاده حسینقلی^{۱*}، هانیه موسوی، محمدرضا صادقی دیزجی^۲، قادر مولوی^۳

وابستگی سازمانی

۱ گروه زیست شناسی، دانشکده علوم پایه، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

۲ گروه پزشکی مولکولی، دانشکده علوم نوین پزشکی، دانشگاه علوم پزشکی تبریز، تبریز، ایران

۳ بیمارستان امام حسین (ع)، دانشگاه علوم پزشکی تبریز، هشتگرد، ایران.

*نویسنده مسئول: zhosseingholi@gmail.com

چکیده

میزان ابتلا به بیماری سرطان سینه در میان جمعیت زنان در سراسر دنیا در حال افزایش است. در دهه های اخیر امکان جمعیت داده های مرتبط فراوان فناوری ریزآرایه و شناسایی دقیق تر ژن های مرتبط در ایجاد و توسعه سرطان ها فراهم شده است. در این مطالعه سه مجموعه داده (GSE109169، GSE80754 و GSE72644) شامل نمونه های تومور سرطان سینه و بافت های سالم مجاور تومور از پایگاه داده ی NCBI استخراج و ادغام شدند. با استفاده از نرم افزارهای مناسب ژن های کلیدی تغییر بیان یافته مشخص شدند و بیان همزمان آنها مورد ارزیابی قرار گرفت. هستی شناسی ژن های کلیدی (GO) تعیین شد و ریزآران ای ها و فاکتورهای رونویسی مرتبط ژن ها مورد بررسی قرار گرفت. نتایج این مطالعه نشان داد که ژن های PPARG، LEP، ADIPOQ، LPL، LIPE، FABP4، PLIN1 و CIDEA مهم ترین ژن های کاهش بیان یافته و ده ژن AURKA، DLGAP5، NCAPG، CENPA، KIF23، KIF11، BUB1B، RRM2، UBE2C و NUSAP1 مهم ترین ژن های افزایش بیان یافته هستند. دو ریزآران ای با نام های has-miR-6782-5p و has-miR-6836-5p بیشترین ارتباطات را با ژن های کلیدی کاهش بیان یافته و دو ریزآران ای با نام های has-miR-7110-5p و has-miR7106-5p بیشترین ارتباطات را با ژن های کلیدی افزایش یافته داشتند. PPARG مهم ترین فاکتور رونویسی تنظیم کننده ژن های کلیدی کاهش بیان یافته و E2F1، FOXM1، E2F4 و MYC مهم ترین فاکتورهایی بودند که ژن های کلیدی افزایش بیان یافته را کنترل می کردند. از نتایج پژوهش حاضر میتوان برای طراحی متدهای درمانی سودمند برای جلوگیری از سرطان پستان استفاده کرد.

واژگان کلیدی: سرطان سینه، بیوانفورماتیک، ریزآرایه، ژن های کلیدی.

مقدمه

سرطان پستان یک بیماری ژنتیکی و بالینی ناهمگن است. این سرطان شایعترین سرطان تشخیص داده شده در زنان و دومین علت شایع مرگ و میر ناشی از سرطان در بین زنان در جهان است. تومورهای پستان تمایل به گسترش لنفاوی و خونی دارند که منجر به متاستازهای دوردست می‌شود (۲۲). در حال حاضر، تحقیقات زیست‌پزشکی تلاش می‌کند مکانیسم‌هایی را توضیح دهد که توسط آن یک بیماری خاص ایجاد می‌شود، به همین دلیل، مطالعات بیان ژن منبع بسیار خوبی است (۱۵). به طور دقیق، اصطلاح "بیان ژن" از فعال سازی ژن تا زمانی که پروتئین بالغ در محل مربوطه خود قرار می‌گیرد تا عملکرد خود را انجام دهد و به بیان فنوتیپ سلول کمک کند، را شامل می‌شود (۲۹). سلول‌ها از طریق شبکه‌های تنظیم کننده ژن بیان ژن را کنترل می‌کنند. کشف شبکه‌های سلولی از اهداف اصلی زیست شناسی سامانه ای است، زیرا دانش ارزشمندی در مورد فرآیندهای بیولوژیکی ارائه می‌دهند که ممکن است به کشف نشانگرهای بیماری و طراحی درمان‌های جدید کمک کند. شبکه‌ها ممکن است شامل برهمکنش بین فاکتورهای رونویسی و ژن‌ها و یا برهمکنش‌های پروتئین-پروتئین باشند (۴۴). با وجود روش‌های نوین سنجش بیان ژن، روش ریزآرایه همچنان یک پلتفرم مقرون به صرفه برای تولید داده‌های کمی متنوع و اندازه‌گیری فرآیندهای مختلف است که پایگاه‌های داده آنالین مرتبط در دسترسی دارد. ادغام انواع مختلف داده‌های بیولوژیکی با افزایش میزان داده می‌تواند سودمند باشد. با این حال، یکپارچه سازی داده‌ها ساده نیست و ادغام سری‌های زمانی از پلتفرم‌های مختلف نیاز به دقت دارد (۳۲). در این پژوهش در گام نخست، از طریق تلفیق داده‌های موجود در پایگاه‌های داده، حجم داده‌های ریزآرایه‌ای مرتبط با بافت سرطانی پستان و بافت‌های مجاور آن افزایش داده شد. سپس از طریق تجزیه و تحلیل داده‌ها و مقایسه‌ی بیان ژن‌ها، به مسیرهای مولکولی جدید دخیل در سرطان پرداخته شد. برای دست‌یابی به اهداف مذکور، ابتدا داده‌های میکروآرای مرتبط به ترنسکریپتوم نمونه‌های تومور پستان و بافت‌های حاشیه‌ای آنها از پایگاه‌های داده جمع‌آوری شد. متعاقب حذف اثرات دسته‌ای و نرمال سازی داده‌ها، با رسم شبکه‌های هم‌بیانی و آنالیز ژن‌هایی که بیان متفاوت دارند در دو گروه داشتند یافت شد و عملکرد آنها مورد بررسی قرار گرفت.

مواد و روش‌ها

تجزیه و تحلیل تغییرات بیانی داده‌های ریزآرایه

جستجوی داده‌های ریز آرایه‌ای مرتبط با بررسی و مقایسه‌ی بیان ژن قسمت توموری سرطان سینه و بافت‌های به ظاهر نرمال حاشیه تومور در پایگاه NCBI قسمت GEOdataset (<http://www.ncbi.nlm.nih.gov/geo>) انجام شد. پروفایل بیانی شماره آزمایش‌های مرتبط با استفاده از پکیج GEOquery نرم افزار R (۸) و دستور getGEOSupFiles استخراج گردید. پس از نرمال سازی داده‌ها و ادغام آنها با استفاده از پکیج limma نرم افزار R (۳۳) از دستور removeBatchEffect همین پکیج برای حذف اثر دسته‌ای (تاثیر عوامل غیر زیستی در تغییر داده‌های تولید شده توسط آزمایش) استفاده شد. تحلیل مؤلفه اصلی (PCA) قبل و بعد از تجزیه و تحلیل اثر دسته‌ای برای آزمایش عملکرد تصحیح اثر دسته‌ای روی مجموعه داده‌های نرمال شده با استفاده از پکیج ggplot2 در نرم

افزار R (۴۱) و دستور autoplot استفاده شد. برای یافتن لیست ژن‌هایی که دارای $\log FC > 1$ و $\log FC < -1$ و $0.05 < \text{adj.P.Val}$ بودند از دستور GEOlimma در پکیج limma استفاده شد. نمودار آتشفشانی ژن‌های افزایش و کاهش بیان یافته نیز با استفاده از پکیج ggplot2 ترسیم شد.

رسم شبکه برهمکنش ژن‌های دارای تغییرات بیانی جهت شناسایی ژنهای کلیدی

لیست ژن‌های افزایش و کاهش بیان یافته مستخرج در مرحله ی قبل جهت شناسایی برهمکنش پروتئینی در پایگاه STRING (۳۷) به آدرس <https://string-db.org> وارد شد و نتایج خروجی دارای امتیاز اطمینان (confidence score) بیشتر از ۰/۴ برای رسم شبکه ی اندرکنش در نرم افزار رایگان ترسیم و آنالیز شبکه ی Cytoscape (۳۱) وارد شد. سپس با تجزیه و تحلیل های آماری افزونه ی CytoHubba (۴) و با استفاده از روش الگوریتم حداکثر مرکزیت دسته (MCC) ژنهای کلیدی (هاب) شناسایی شد. MCC موثرترین روش برای شناسایی ژن‌های با بیشترین تعداد ارتباط در شبکه است. ژنهای کلیدی یا هاب ژن‌هایی هستند که به شدت به ژن‌های دیگر در یک شبکه متصل هستند و تصور می‌شود برای تنظیم شبکه مهمتر باشند.

از ضرایب همبستگی (۳۰) برای سنجش میزان قوی بودن رابطه بین دو متغیر استفاده می‌شود. برای یافتن این ضریب از پایگاه داده GEPIA (۳۹) به آدرس <http://gepia.cancer-pku.cn> استفاده شد. برای بدست آوردن ارتباط بیانی بین جفت ژنهای کلیدی نیز از قسمت محاسبه ضریب همبستگی پایگاه داده GEPIA استفاده شد و نمودار مربوطه با نرم افزار Excel ترسیم شد.

تجزیه و تحلیل غنی سازی ژنهای کلیدی منتخب جهت شناسایی هستی شناسی ژنها (GO) و مسیرهای

بیولوژیکی

انتولوژی یا همان هستی شناسی ژن‌های افزایش و کاهش بیان یافته ی منتخب در سه سطح فرایندهای زیستی، عملکردهای مولکولی و موقعیت سلولی در پایگاه اطلاعاتی DAVID (۱۹) به آدرس <https://david.ncifcrf.gov> مشخص گردید. شناسایی مسیرهای زیستی مرتبط نیز با استفاده قسمت KEGG در پایگاه DAVID انجام شد و نتایج با $FDR < 0.05$ استخراج شد.

بررسی ریزآران‌های مرتبط با تنظیم بیان ژن‌های کلیدی منتخب

از پایگاه داده miRWalk 3.0 (۳۵) به آدرس <http://mirwalk.umm.uni-heidelberg.de> برای شناسایی ریزآران‌های متعامل با ام آر آن‌های ژن‌های کلیدی استفاده شد. ریزآران‌های پیش بینی شده با امتیاز < 0.95 و مناطق اتصال $UTR 3'$ ، $UTR 5'$ و CDS استخراج شد. آنالیز شبکه ی اتصال آنها برحسب درجه (degree) در نرم افزار Cytoscape انجام شد و رسم شبکه برای بررسی ریزآران‌های با درجه ی اتصال بالا با افزونه ی cytoHubba انجام شد.

بررسی فاکتورهای رونویسی مرتبط با ژن‌های کلیدی منتخب

به‌منظور شناسایی فاکتورهای رونویسی مرتبط با ژن‌های کلیدی از قسمت ENCODE and ChEA

Consensus TFs from ChIP-X و TRRUST Transcription Factors موجود سایت Enrichr (۴۲) به آدرس <https://maayanlab.cloud/Enrichr> استفاده شد. از سایت Expression2Kinases (X2K) (۶) به آدرس <http://www.maayanlab.net/X2K> نیز فاکتورهای رونویسی مرتبط استخراج شد. در تمام نتایج فاکتورهای رونویسی $\text{adjusted } p\text{-value} < 0.05$ مدنظر قرار گرفت و سپس فاکتورهای رونویسی مشترک از طریق رسم دیاگرام Venn قابل دسترسی در سایت Venny 2.0.2 (۲۷) به آدرس <https://bioinfogp.cnb.csic.es/tools/venny/index2.0.2.html> به دست آمد.

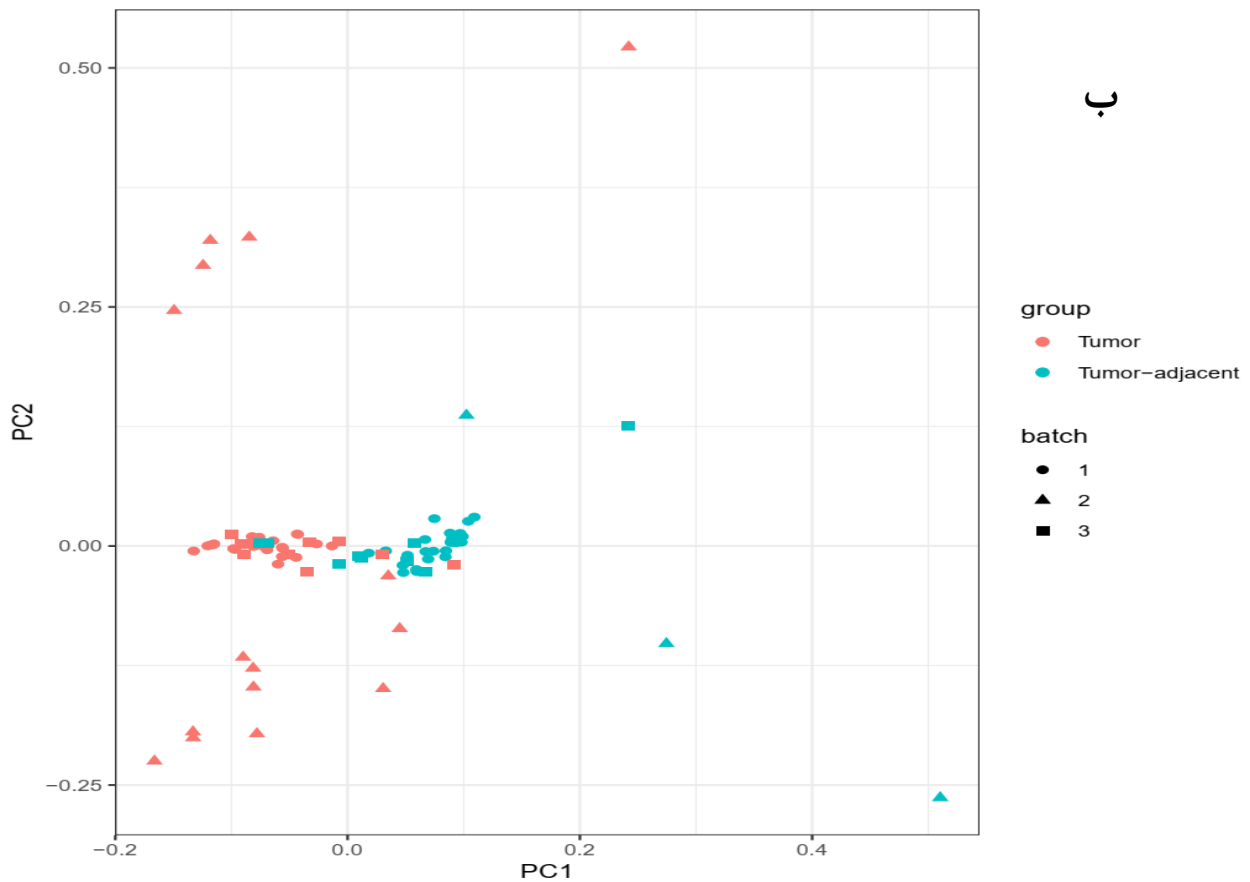
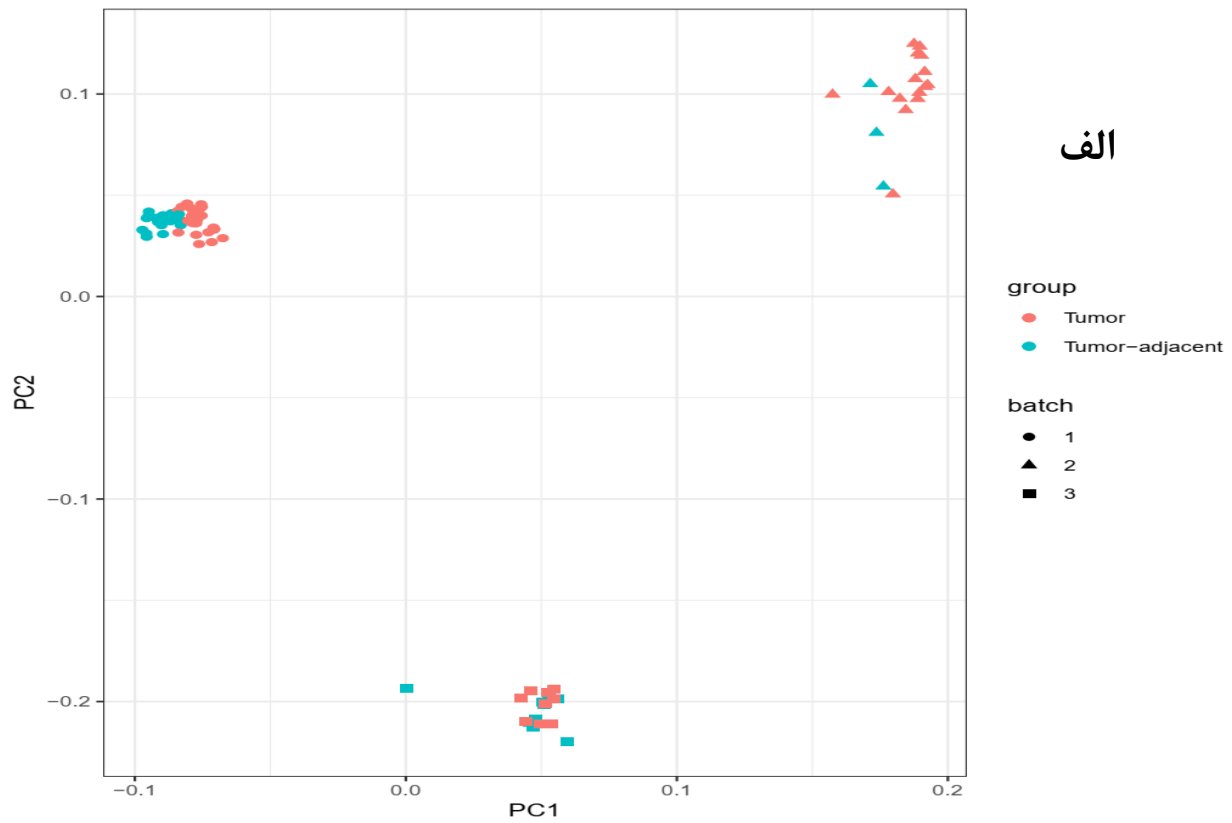
نتایج

تجزیه و تحلیل تغییرات بیانی پس از ادغام داده های زیر آرایه ها

داده‌های خام بررسی‌های ریزآرایه با شماره آزمایش‌های GSE109169، GSE80754 و GSE72644 مربوط به داده‌های بیانی نمونه‌های توموری پستان و بافت‌های نرمال مجاور آنها از پایگاه NCBI استخراج شد. توضیحات بیشتر در مورد این شماره آزمایشات در جدول ۱ آمده است. نتایج بررسی PCA دوبعدی ادغام داده ها قبل و بعد از حذف اثر دسته ای به ترتیب در شکل ۱-الف و ۱-ب آورده شده است. پس از استخراج لیست ژنی از روش ادغام داده های ریزآرایه، نمودار آتشفشانی آنها رسم شد (شکل-۲). پس از تعیین ژن‌های هاب تغییر بیان یافته، شبکه ارتباطی ژنهای کلیدی توسط Cytoscape ترسیم گردید (شکل-۳ الف و ب).

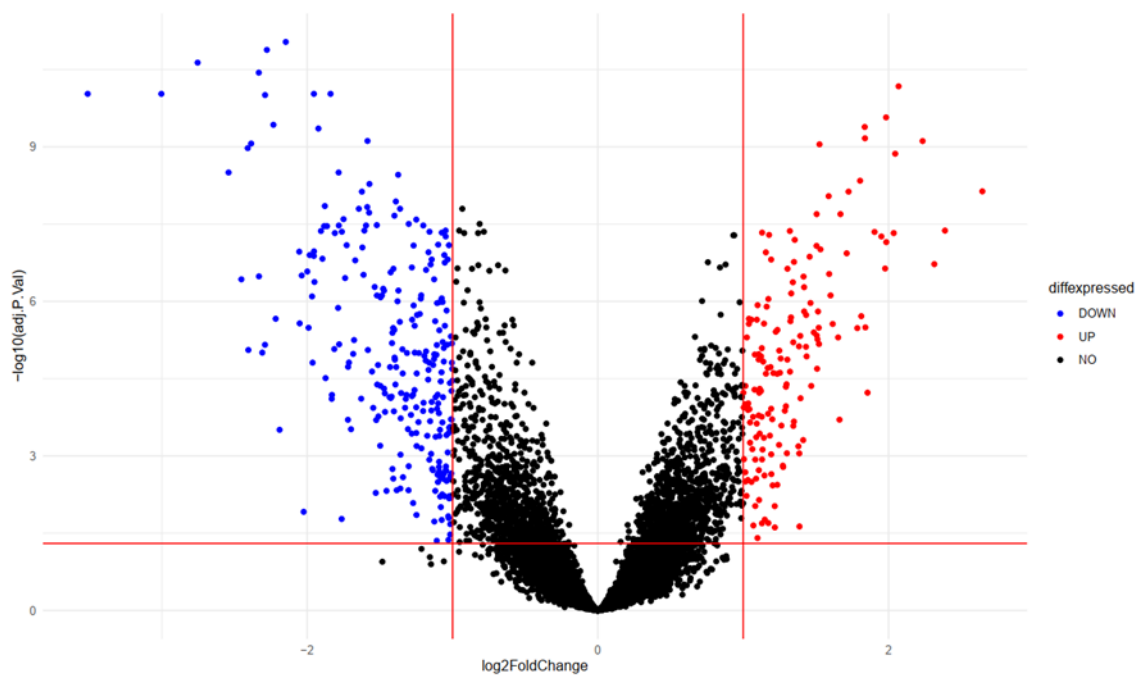
جدول ۱: مشخصات و توضیحات مرتبط با شماره آزمایشهای انتخابی

شماره آزمایش	داده های بیانی تومور	داده های بیانی بافت نرمال مجاور	توضیح آزمایش
GSE109169	GSM605340	GSM605343	مجموعه داده شامل ۲۵ جفت نمونه از تومور های سرطان پستان و اپیتلیوم طبیعی مجاور تومور است. PMID: 29706651
	GSM605341	GSM605344	
	GSM605342	GSM605345	
GSE80754	GSM2753987	GSM275399	مجموعه داده شامل ۱۵ نمونه از تومور های سرطان پستان و ۳ نمونه اپیتلیوم طبیعی مجاور تومور است. PMID: 28977854
	GSM2753988	GSM2753992	
	GSM2753989	GSM2753993	
	GSM2753990	GSM2753994	
GSE72644	GSM1144675	GSM1144678	مجموعه داده شامل ۹ بیمار مبتلا به سرطان پستان است که از آنها نمونه های اپیتلیوم طبیعی از امتداد مجرای منتهی به تومور، و خود تومور گرفته شده است. PMID: 29093438
	GSM1144676	GSM1144679	
	GSM1144677	GSM1144680	

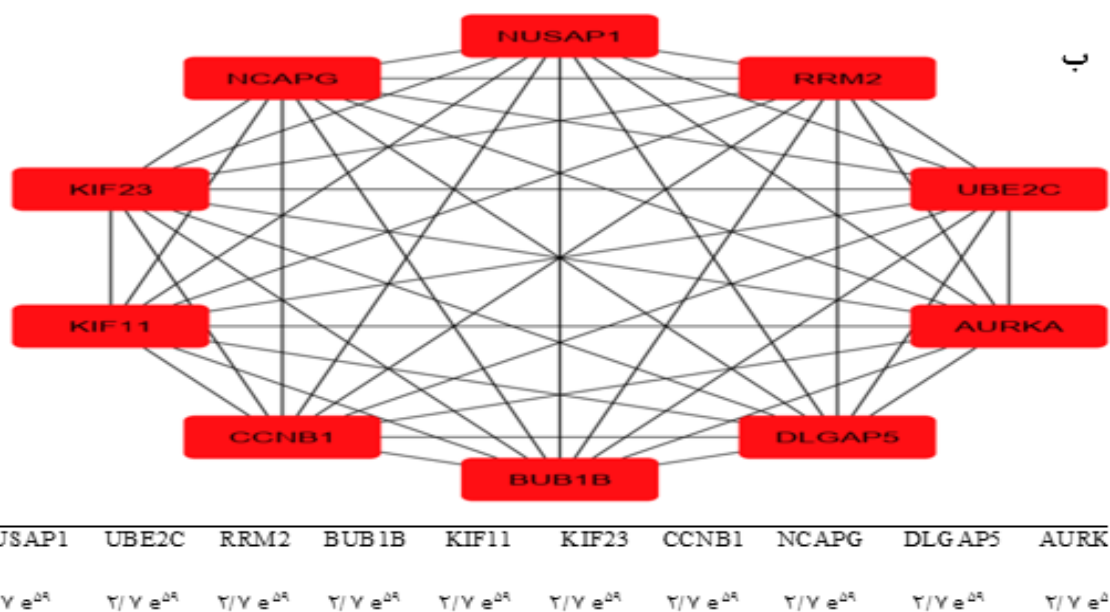
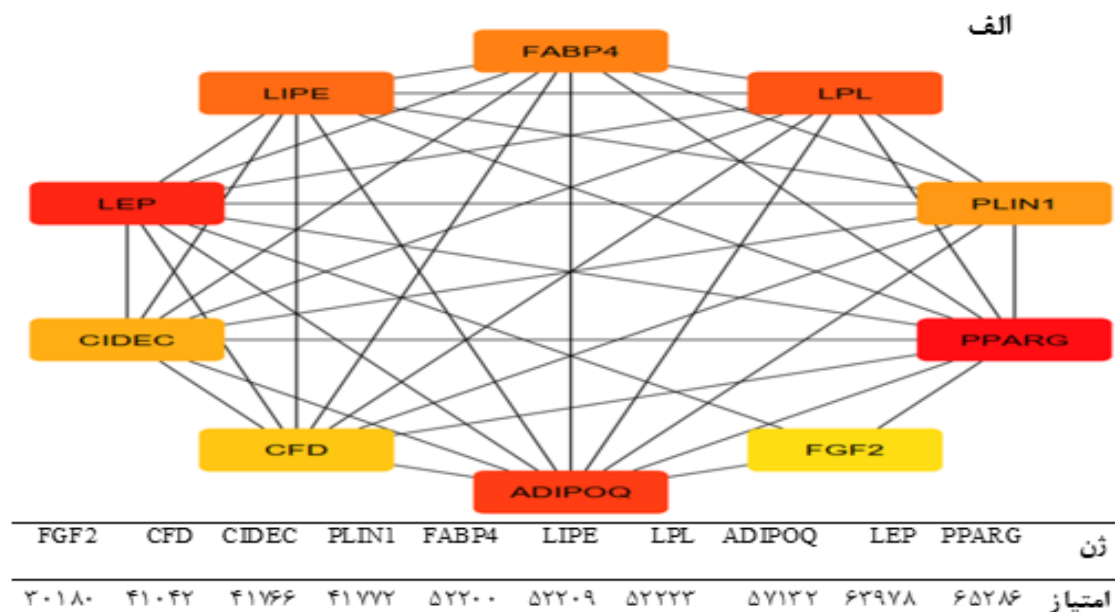


شکل-۱: پلات PCA روش ادغام داده های ریز آرایه قبل (الف) و بعد (ب) حذف اثر دسته‌ای. منظور از بچ ۱ و ۲ و ۳ به ترتیب شماره

آزمایشهای GSE109169، GSE80754 و GSE72644 است.



شکل - ۲: نمودار اشفشانی مربوط به ژن‌های مستخرج از روش ادغام داده‌های ریز آرایه. میزان ژنهای افزایش بیان یافته با رنگ قرمز در سمت راست تصویر نشان داده شده است. ژنهای بدون تغییر در مرکز تصویر با رنگ سیاه نمایش داده شده‌اند و ژنهای کاهش بیان یافته در سمت چپ تصویر به رنگ آبی نمایان هستند.



شکل - ۳: شبکه ارتباطی، اسامی و امتیاز ژنهای کلیدی کاهش بیان یافته (الف) و ژنهای کلیدی افزایش بیان یافته (ب). الگوریتم MCC در CytoHubba (مانند پروتئین‌ها) را در یک شبکه با توجه به اینکه در چند گروه کامل (به نام کلیکهای ماکسیمال) شرکت دارند، امتیاز می‌دهد. کلیک ماکسیمال گروهی از گره‌ها هستند که در آن هر گره به هر گره دیگری در آن گروه متصل است. گره‌هایی که در بیشتر این گروه‌ها یا گروه‌های بزرگ‌تر هستند، امتیاز بالاتری می‌گیرند. در تصویر امتیازها با طیف رنگی نشان داده شده‌اند.

بررسی ویژگی‌های ژن‌های کلیدی

بررسی ضریب همبستگی بیان ژن‌های هاب کاهش و افزایش‌یافته به صورت نمودار به ترتیب در تصاویر ۴ و ۵ آورده شده است. در نمودار همبستگی ژن‌های کاهش بیان یافته دو ژن CFD و FGF2 همبستگی بیانی بسیار پایینی با بقیه ی ژنها دارا هستند و از روند بقیه ی آنالیزها حذف می شوند. بیشترین همبستگی بیانی (۰/۹۶) نیز بین دو ژن PLIN1 و CIDEc وجود دارد. در نمودار همبستگی ژن‌های افزایش بیان یافته نیز بیشترین همبستگی بیانی (۰/۸۹) بین دو ژن BUB1B و NUSAP1 وجود دارد. کمترین همبستگی بیانی (۰/۴۸) نیز بین دو ژن RRM2 و UBE2C دیده می شود

PPARG	1									
LEP	0.48	1								
ADIPOQ	0.57	0.64	1							
LPL	0.43	0.54	0.69	1						
LIPE	0.55	0.79	0.87	0.62	1					
FABP4	0.53	0.68	0.75	0.57	0.76	1				
PLIN1	0.59	0.82	0.88	0.65	0.91	0.81	1			
CIDEc	0.58	0.78	0.88	0.67	0.92	0.77	0.96	1		
CFD	0.22	0.26	0.2	0.13	0.3	0.25	0.28	0.27	1	
FGF2	0.11	0.12	0.17	0.13	0.15	0.11	0.15	0.15	-0.011	1
	PPARG	LEP	ADIPOQ	LPL	LIPE	FABP4	PLIN1	CIDEc	CFD	FGF2

شکل - ۴: نمودار ضریب همبستگی ژن‌های هاب کاهش بیان یافته

AURKA	1									
DLGAP5	0.63	1								
NCAPG	0.67	0.79	1							
CCNB1	0.65	0.77	0.75	1						
KIF23	0.66	0.75	0.78	0.75	1					
KIF11	0.61	0.78	0.79	0.74	0.78	1				
BUB1B	0.62	0.76	0.78	0.75	0.81	0.81	1			
RRM2	0.57	0.67	0.64	0.58	0.61	0.64	0.64	1		
UBE2C	0.68	0.61	0.65	0.67	0.65	0.56	0.63	0.48	1	
NUSAP1	0.61	0.73	0.76	0.73	0.79	0.81	0.89	0.72	0.57	1
	AURKA	DLGAP5	NCAPG	CCNB1	KIF23	KIF11	BUB1B	RRM2	UBE2C	NUSAP1

شکل - ۵: نمودار ضریب همبستگی ژن‌های هاب افزایش بیان یافته

بررسی GO ژن‌های هاب کاهش و افزایش بیان‌یافته در روش ادغام داده های ریز آرایه به ترتیب در جداول ۲ و ۴ آورده شده است. همان‌طور که از جداول پیداست مجموع عملکرد ژن‌های هاب کاهش‌یافته، کاهش فرایندهای مرتبط با متابولیسم چربی، کاهش پاسخ به انسولین و به هم خوردن هموستاز گلوکز و کاهش پاسخ به باکتری بوده است. مجموع عملکردهای مهم ژن‌های هاب افزایش بیان‌یافته نیز افزایش تقسیم سلولی بوده است. برای ژن‌های هاب کاهش بیان‌یافته

۴ مسیر زیستی معنی‌دار شامل مسیرهای سیگنالینگ PPAR، سیگنالینگ AMPK، لیپولیز در آدیپوسیت‌ها و کاهش بیماری کبد چرب غیرالکلی نیز در بررسی‌های DAVID وجود داشت که مشخصات آن در جدول ۳ آورده شده است.

جدول - ۲: هستی شناسی (GO) ژن‌های هاب کاهش بیان یافته ی مستخرج از روش ادغام داده های زیر آرایه ها

term GO	Ontology	توضیح بیشتر در مورد GO	تعداد ژن‌های درگیر	ژن‌های درگیر	FDR
GO:0006629	P	فرایند متابولیسم لیپید	۴	LEP, LPL, PPARG, PLIN1	۰/۰۱۱۰
GO:0016042	P	فرایند متابولیسم لیپید	۴	LIPE, LPL, PLIN1	۰/۰۴۱۷
GO:0032869	P	پاسخ سلولی به محرک انسولین	۴	LEP, ADIPOQ, PPARG	۰/۰۴۱۷
GO:0042593	P	هموستاز گلوکز	۴	LEP, ADIPOQ, PPARG	۰/۰۴۱۷
GO:0009617	P	پاسخ به باکتری	۳	FABP4, ADIPOQ, LPL	۰/۰۷۸۴
GO:0005811	C	ذره لیپیدی	۴	FABP4, ADIPOQ, LPL	۹/۸۱e-۵

P: فرایند زیستی؛ C: مکان سلولی

جدول - ۳: مشخصات مسیر زیستی مرتبط با ژن‌های هاب کاهش بیان یافته

مسیر	تعداد ژن‌های درگیر	ژن‌های درگیر	FDR
hsa03320: مسیر سیگنالینگ PPAR	۵	FABP4, ADIPOQ, LPL, PPARG, PLIN1	۵e-۱/۸
hsa04152: مسیر سیگنالینگ AMPK	۴	LEP, LIPE, ADIPOQ, PPARG	۰/۰۰۳۲
hsa04923: تنظیم لیپولیز در سلول‌های چربی	۳	LIPE, FABP4, PLIN1	۰/۰۱۶۸
hsa04932: بیماری کبد چرب غیرالکلی	۳	LEP, ADIPOQ, PPARG	۰/۰۳۱۹

جدول - ۴: هستی شناسی (GO) ژن‌های هاب افزایش بیان یافته ی مستخرج از روش ادغام داده های زیر آرایه ها

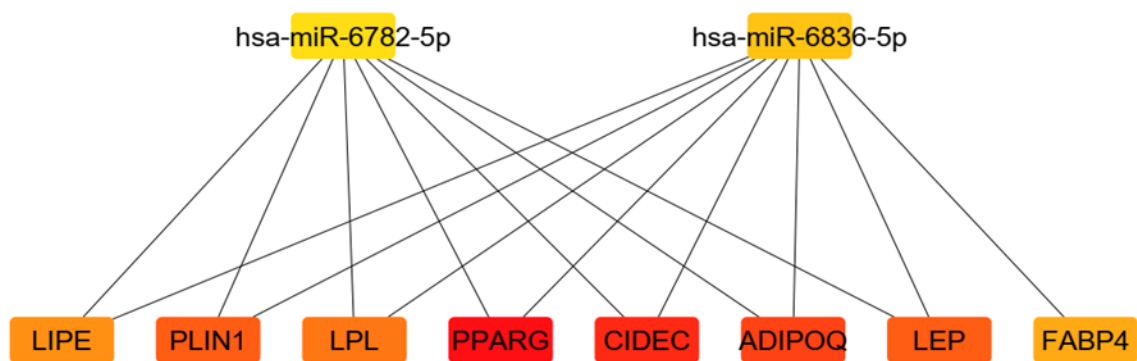
term GO	Ontology	توضیح	تعداد ژن‌های درگیر	ژن‌های درگیر	FDR
GO:0051301	P	تقسیم سلولی	۶	CCNB1, UBE2C, NCAPG, BUB1B, KIF11, AURKA	۳/۶e-۵
GO:0007052	P	سازمان‌دهی میتوزی	۴	CCNB1, KIF11, DLGAP5, AURKA	۹/۸۶e-۵
GO:0005819	C	دوک	۵	NUSAP1, BUB1B, KIF23, KIF11, AURKA	۱/۲۹e-۵
GO:0072686	C	دوک میتوزی	۴	NUSAP1, KIF23, KIF11, AURKA	۵/۱e-۴
GO:0005829	C	سیتوزول	۹	CCNB1, RRM2, UBE2C, NCAPG, BUB1B, KIF23, KIF11, DLGAP5, AURKA	۰/۰۰۲۳

GO:0005874	C	میکروتوبول	۴	NUSAP1, KIF23, KIF11, AURKA	۰/۰۰۳۲
GO:0005634	C	هسته	۹	CCNB1, RRM2, UBE2C, NUSAP1, NCAPG, KIF23, KIF11, DLGAP5, AURKA	۰/۰۰۳۲
GO:0000940	C	کینتوکور بیرونی	۲	CCNB1, BUB1B	۰/۰۰۴۲
GO:0031616	C	کینتوکور بیرونی	۲	DLGAP5, AURKA	۰/۰۰۴۶
		کروموزوم متراکم			

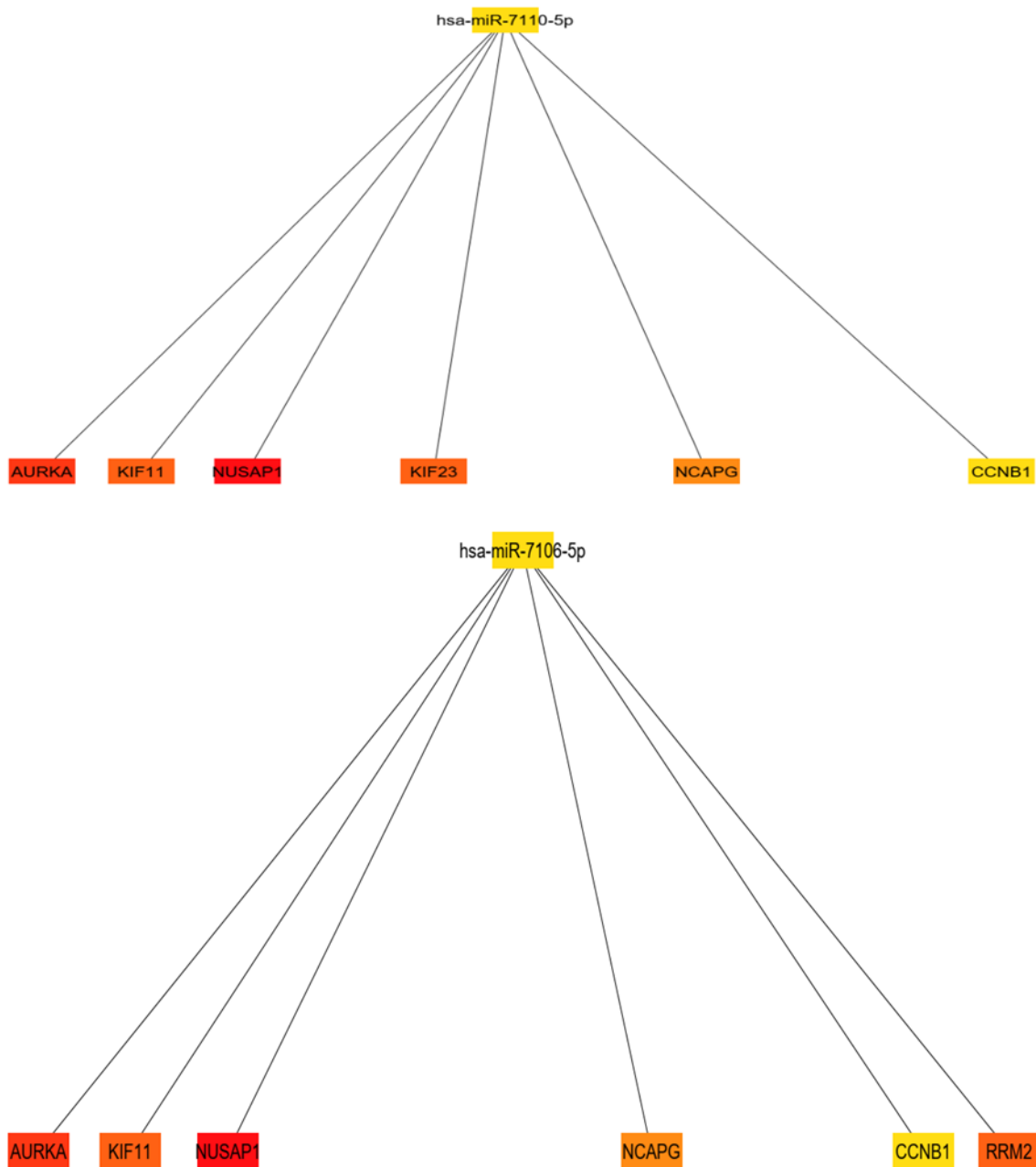
P: فرایند زیستی: C: مکان سلولی

ریزآران‌ای‌ها و فاکتورهای رونویسی مهم دخیل در تنظیم ژن‌های کلیدی

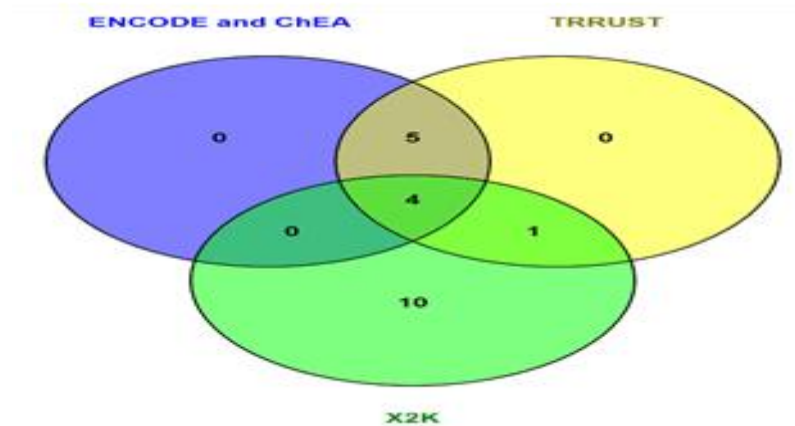
در روش استفاده شده برای مطالعه ریزآران‌ای‌های مرتبط با ژن‌های کلیدی برای ژن‌های کاهش‌یافته دو ریزآران‌ای با نام‌های has-miR-6836-5p و has-miR-6782-5p بیشترین ارتباطات را با ژن‌های هاب دارا بودند که شبکه ارتباطی آنها در شکل-۶ دیده می‌شود. در مورد ژن‌های افزایش بیان یافته نیز دو ریزآران‌ای با نام‌هایی has-miR-7110-5p و has-miR-7106-5p بیشترین ارتباطات را با ژن‌های هاب افزایش‌یافته دارا بودند که در شکل-۷ شبکه ارتباطی‌شان به صورت مجزا آورده شده است. راهکار بررسی فاکتورهای رونویسی نیز نشان داد که در مورد ژن‌های کلیدی کاهش بیان‌یافته فاکتور رونویسی PPARG به طور معنی‌داری در ارتباط با بیان ژن‌هاست. در مورد ژن‌های افزایش بیان‌یافته همان‌طور که در شکل-۸ دیده می‌شود، ۴ فاکتور رونویسی مشترک در پایگاه‌های داده وجود دارد که عبارت‌اند از E2F4, MYC, E2F1, FOXM1.



شکل-۶: شبکه ارتباطی بین دو ریزآران‌ای با درجه ارتباطی بالا با ژن‌های هاب کاهش بیان یافته. پرنسب بودن کادر اسامی نشان دهنده ی درجه ی ارتباطی بالای آنهاست.



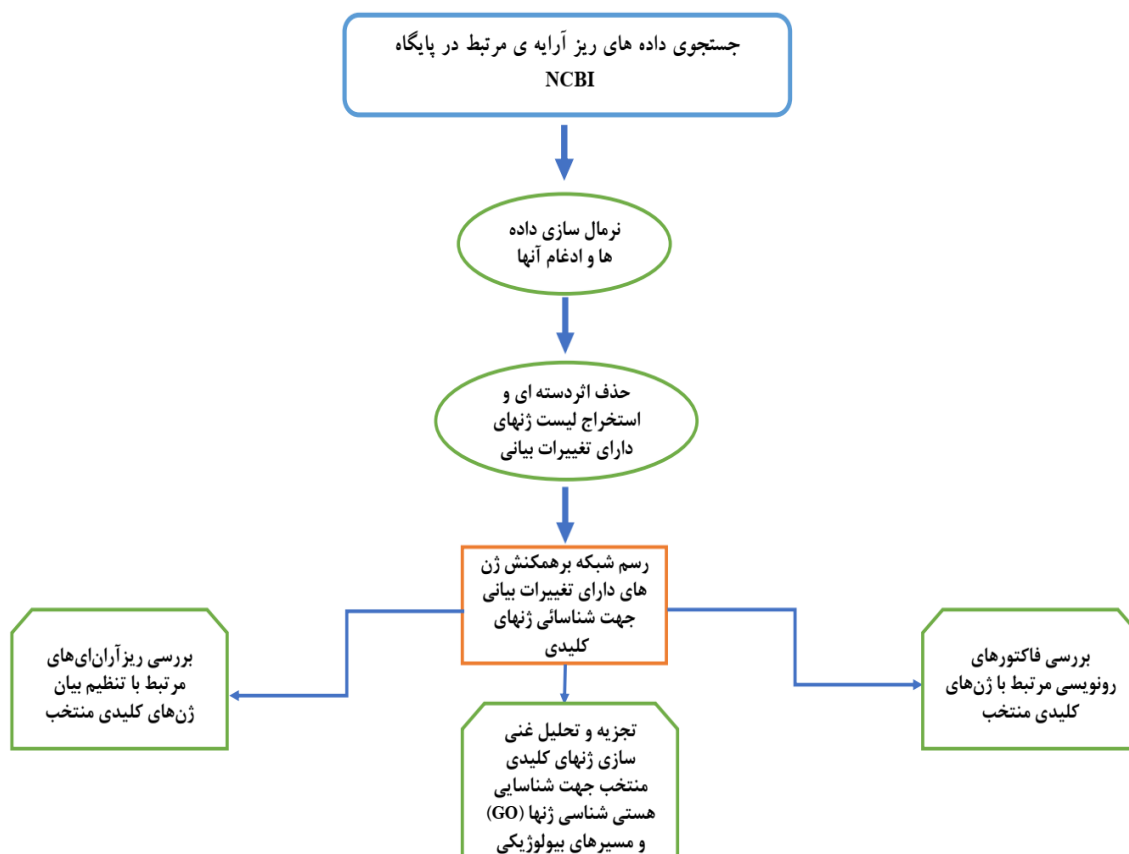
شکل - ۷: شبکه ارتباطاتی بین دو ریزآران‌ای با درجه ارتباطی بالا با ژن‌های هاب افزایش بیان یافته. پررنگ تر بودن کادر اسامی نشان دهنده ی درجه ی ارتباطی بالای آنهاست.



شکل - ۸: نمودار venny نتایج پیش بینی فاکتورهای رونویسی مرتبط با ژن‌های کلیدی افزایش یافته. ۴ فاکتور رونویسی مشترک در نتایج پیش بینی شده توسط سایت های TRRUST (دایره زرد رنگ)، سایت ENCODE AND ChEA (دایره آبی) و سایت X2K (دایره سبز) وجود دارد.

بحث

در حالی که ریزآرایه‌ها بیان هزاران ژن را به طور همزمان اندازه‌گیری می‌کنند، انتظار می‌رود که فقط زیرمجموعه کوچکی از ژن‌ها با نتیجه‌ی بیولوژیکی مورد مطالعه مرتبط باشند. لیست ژن‌های منتشر شده توسط گروه‌های مختلف مطالعه‌کننده بر روی موضوع واحد، اغلب همپوشانی کمی نشان می‌دهند. علل احتمالی این تکرارپذیری کم شامل تفاوت در روش‌های جمع‌آوری نمونه، روش‌های ساخت چیپ‌های ریزآرایه، و میزان کم نمونه‌های مورد بررسی است. ادغام مجموعه داده‌های متعدد ریزآرایه برای بهبود انتخاب ژن‌ها پیشنهاد شده است. افزایش اندازه نمونه‌ها قدرت آماری را برای بدست آوردن برآورد دقیق‌تر افزایش می‌دهد (۱). خلاصه‌ای از جریان کار مطالعه‌ی حاضر در شکل - ۹ آورده شده است. در انتخاب دیتاست‌ها دقت بر این بود که شماره آزمایش‌هایی انتخاب شوند که حاوی نمونه‌های توموری و حاشیه تومور بیماران باشند و از انتخاب دیتاست‌های حاوی رده‌های سلولی سرطان پستان اجتناب شد.



شکل - ۹: نمودار کلی جریان کار در این مطالعه

نتایج جداول GO و مسیرهای زیستی مرتبط با ژنهای هاب کاهش بیان یافته در روش ادغام داده های ریز آرایه در جدول ۲ و ۳ آورده شد که در ذیل توضیح مختصری در مورد این فرایندها و دلایل توجیهی کاهش فرایندها و مسیرهای مرتبط با این ژنها و بررسی نتایج سایر مطالعات آورده شده است. متابولیسم نامنظم چربی ها یکی از مشخصه های سلول های سرطانی و نشانه ای برای پیشرفت سرطان می باشد (۱۰). اسیدهای چرب آزاد (FFA) های تولید شده در لیپولیز در سلول های چربی به سلول های سرطانی منتقل می شوند. افزایش جذب و سنتز FFA در سلول های سرطانی می تواند منجر به افزایش بیش از حد اسیدهای چرب و سمیت چربی شود. سلول با تبدیل FFA به TAG ها و ذخیره آنها در قطرات چربی (LD) در برابر سمیت چربی مصون می ماند. سلول های چربی بیشترین نسبت را در میان سلول های تشکیل دهنده ی بافت پستان را دارند (۵). اغلب سلول های سرطانی در معرض یک محیط چالش برانگیز کمبود اکسیژن و مواد مغذی قرار می گیرند که منجر به تغییراتی در تعادل بین سنتز و جذب اسیدهای چرب می شود. (۲۴). یکی از این تغییرات کاهش لیپولیز داخل سلولی و اکسیداسیون اسید چرب می شود که نهایتاً منجر به تجمع تری گلیسیرید (TG) در قطرات چربی می شود. پس افزایش سنتز TG و کاهش لیپولیز هر دو به عنوان راهکارهایی هستند که توسط سلول های سرطانی مختلف برای اجتناب از سمیت چربی در طول هیپوکسی صورت می گیرد (۲۵). در مورد گزارش ارتباط

سرطان پستان با این بیماری تناقضاتی وجود دارد برخی از تحقیقات عدم ارتباط را گزارش کرده اند (۲۰) و برخی دیگر عدم ارتباط قطعی افزایش این بیماری را گزارش کرده اند (۲۱)

افزایش نیاز به گلوکز برای حفظ فرآیندهای اصلی سلول های سرطانی ضروری است و این سلول ها با افزایش میزان جذب گلوکز، تولید کم ATP را در مسیر گلیکولیز جبران می کند. شواهد بیولوژیکی مبنی بر نقش اختلال میزان سطح گلوکز و سایر عوامل مرتبط با متابولیسم گلوکز، مثل انسولین در ایجاد سرطان سینه وجود دارد (۲). مطالعات اپیدمیولوژیکی نشان داده است که وزیکول های خارج سلولی مشتق شده از سرطان سینه، ترشح انسولین پانکراس را سرکوب می کنند تا هموستاز گلوکز را مختل کنند (۳). مقاومت به انسولین (IR) با پیش آگهی بد سرطان سینه مرتبط است. IR باعث هایپرانسولینمی مزمن می شود که در سرطانزایی نقش دارد. همچنین هایپرانسولینمی طولانی مدت ممکن است منجر به سطوح بالاتر فاکتور رشد شبه انسولین (IGF) شود که مسیرهای سیگنالینگ تحریک کننده تومور را افزایش می دهد و به پیشرفت سرطان مقاوم در برابر درمان کمک می کند (۳۴).

تعامل پویا ثابت بین سلول ها و سیستم ایمنی برای حفظ هموستاز رخ می دهد. در حالت مزمن، سیستم ایمنی به طور مکرر در پاسخ به سیگنال های آنتی ژنی پایدار فعال می شود و ممکن است "التهاب مزمن" ایجاد شود (۷). مشابه آنچه در مورد عفونت های مزمن یا میکروارگانیزم های کومنسال بدن اتفاق می افتد، تومور ها حضور خود را با سرکوب التهاب بیش از حد و پاسخ های ایمنی ضد تومور حفظ می کند (۱۶). اسیدهای نوکلئیک و سایر محصولات که در اثر استرس یا مرگ سلولی در تومورها به دنبال پاسخ های سیتوتوکسیک آزاد می شوند، توسط حسگرهای ذاتی شناسایی می شوند که منجر به التهاب و فعال شدن سلول های ارائه دهنده آنتی ژن (APCs) می شود. بنابراین، مانند عفونت، تومورها از شروع اولیه تا پیشرفت، انتشار و در نهایت مرگ میزبان التهاب و ایمنی را تحت تأثیر قرار می دهند و تحت تأثیر قرار می گیرند (۲۸).

گیرنده های فعال شده با تکثیر پراکسی زوم (PPARs) می توانند بیان ژن های دخیل در متابولیسم لیپید، چربی زایی، حفظ هموستاز متابولیک و التهاب را تعدیل کنند. به نظر می رسد PPAR γ عمدتاً از طریق فعال سازی مسیرهای مختلف در تنظیم تومورزایی نقش دارد. بنابراین، ایجاد تغییر در مسیرهای این سیگنالینگ می تواند یک راهکار جدید بالقوه برای مهار سرطان زایی و پیشرفت تومور باشد. از آنجایی که شواهد موجود نقش سیگنال دهی PPAR را در سرطان زایی، دیابت نوع ۲ و سایر اختلالات متابولیک مانند چاقی برجسته می کند، درک مکانیسم های مولکولی بالقوه در زمینه تعامل بین متابولیسم، سیگنال دهی PPAR و سرطان ممکن است یک زمینه تحقیقاتی جالب برای توسعه و استراتژی های جدید مفید برای پیشگیری و درمان سرطان می باشد (۳۸). R.I. Glazer و همکارانش در یک مقاله مروری به نقش PPAR در شروع و ترویج تومورزایی پستانی از طریق نقش محوری آن در تنظیم متابولیسم، التهاب و تحمل ایمنی اشاره کرده اند (۱۴). در مقاله ای عملکرد PPAR را در تعدیل متابولیسم میتوکندری کبد در طول پیشرفت از بیماری کبد چرب غیرالکلی به سرطان کبد را توضیح داده شده است (۲۳). AMPK به عنوان یک کیناز، فرآیندهای بیولوژیکی متنوع را عمدتاً از طریق فسفریله کردن اهداف پایین دست خود تنظیم می کند. اختلال در فعالیت AMPK با روند پیری و بیماری های مختلف انسانی از جمله سندرم متابولیک، سکته مغزی، بیماری قلبی عروقی و سرطان مرتبط است. بنابراین،

AMPK می‌تواند به عنوان یک سرکوب کننده تومور به دلیل عملکرد آن در سرکوب سیگنالینگ انکوژن mTORC1 در نظر گرفته می‌شود (۱۸).

نتایج جداول GO با ژن‌های هاب افزایش بیان یافته پس از ادغام داده‌های زیر آرایه‌ها در جدول ۴ آورده شد که در ذیل توضیح مختصری در مورد آن آورده شده است. حذف تنظیم از چرخه سلولی یکی از مشخصه‌های سرطان است که تقسیم سلولی بی‌حد و حصر را امکان‌پذیر می‌کند. برای حمایت از این فنوتیپ بدخیم، سلول‌ها تغییرات مولکولی را به دست می‌آورند که مکانیسم‌های کنترلی را در مسیرهای سیگنالینگ و نقاط بازرسی سلولی که معمولاً برای جلوگیری از بی‌ثباتی ژنومی و تکثیر کنترل‌نشده سلولی عمل می‌کنند، لغو کنند یا دور بزنند. در نتیجه، هدف‌گیری درمانی چرخه سلولی مدت‌هاست که به عنوان یک استراتژی ضد سرطان امیدوارکننده در نظر گرفته می‌شود (۴۰).

شواهد نشان داده‌اند که ریزآران‌های بیان یک ژن را پس از رونویسی با مهار انتقال mRNA یا القای تخریب آن کنترل می‌کنند. این نشانگرهای بیولوژیکی برای تشخیص، پیش‌بینی و درمان استفاده می‌شوند (۱۷). در تصویرهای ۷ و ۸ نتایج بررسی ریزآران‌های مرتبط با ژنهای هاب کاهش و افزایش بیان یافته آورده شد. در مورد برخی از ریزآران‌ها تناقض یافته‌ای با سایر یافته‌ها وجود دارد که قطعاً نیازمند تحقیقات عملی است. در ابتدا ژنهای مرتبط با ژنهای هاب کاهش بیان یافته بیان یافته مرور می‌شود. miR-6782-5p بیان بالایی در سلول‌های سرطانی پروستات دارد و یک نشانگر زیستی برای فعالیت متاستاتیک سرطان پروستات است (۱۲). همچنین در مطالعات دیگری بیان شده است که mirR-6836-5p در سرطان لوزالمعده یک شاخص تشخیصی ارزشمندی است (۲۶). گزارشی از دو ریزآران‌ای اخیر در مورد سرطان پستان وجود ندارد. در مورد ریزآران‌های مرتبط با ژنهای افزایش بیان یافته نیز گزارش شده است که miR-7106-5p در سرطان سینه در مقایسه با بافت‌های طبیعی سینه مجاور به طور متفاوتی تنظیم می‌شود (۴۳). گزارشی از تفاوت بیانی ریزآران‌ای miR-7110-5p در سرطان پستان وجود ندارد.

بررسی فاکتورهای رونویسی در این مطالعه نشان داد که PPARG به طور معنی‌داری در ارتباط با بیان ژن‌های کلیدی کاهش بیان یافته است. در مورد ژن‌های افزایش بیان یافته نیز فاکتورهای رونویسی مؤثر عبارت بودند از FOXM1، E2F4 و E2F1 (شکل - ۸). PPARG قبلاً توضیحاتی در قسمت ژن‌های کاهش بیان یافته ارائه شد.

MYC یک فاکتور رونویسی است که در سطوح چندگانه تنظیم می‌شود و با چندین مسیر سیگنالینگ در ارتباط است. این فاکتور رونویسی اغلب در سرطان‌های پستان بیان می‌شود و در عود، متاستاز، تومور و در پاسخ شیمی درمانی نقش دارد. در سلول‌های سرطان پستان ژن‌های هدف MYC در رشد، رگ‌زایی و کنترل چرخه سلولی نقش دارند. BRAC1 از طریق تعامل با MYC با تنظیم رونویسی مرتبط است و تقویت MYC به طور قابل توجهی با فنوتیپ‌های تومور تهاجمی و نتایج بالینی ضعیف مرتبط است. مطالعات نشان داده‌اند که سیگنال‌دهی MYC میتواند سلول‌های تومور را قادر سازد تا از پاسخ ایمنی میزبان فرار کنند (۹). FOXM1 یک پروتئین‌کوژن تایید شده در سرطان سینه است و تا حدی بر حفظ سلول‌های بنیادی سینه نظارت می‌کند. تجزیه و تحلیل‌ها نشان داده‌اند که بیان بیش از حد آن در سرطان شایع است و به شدت با پیامد‌های ناگوار همراه است (۱۳). E2F1 این فاکتور رونویسی زمانی که G0 تمام و فاز S شروع می‌شود، فعال می‌شود و E2F1 فعال شده نقش مهمی در هدایت سلول‌ها به فاز S دارند. خانواده E2F1 نقش مهمی در تنظیم

چرخه سلولی به عنوان یک عامل مهم در کنترل متابولیسم نه تنها در فیزیولوژیک طبیعی بلکه در شرایط پاتولوژیک سرطان دارند. یکی از ویژگی‌های متابولیسمی بسیاری از سلول‌های سرطانی اثر واربورگ است که شامل افزایش گلیکولیز هوازی و کاهش اکسیداسیون گلوکز است. E2F1 می‌تواند این تنظیم متابولیسمی را هم با افزایش گلیکولیز و هم با سرکوب اکسیداسیون گلوکز در میتوکندری تقویت کند. این فاکتور رونویسی ممکن است یک تنظیم‌کننده اصلی در متاستاز سرطان پستان باشد، زیرا بیان ژن‌های پیش متاستاتیک متعددی را کنترل می‌کند. تغییرات در رونویسی سلول‌های سرطانی در طی مراحل مختلف متاستاز بسیار مهم است. نقش مضاعف و متناقض آن به این دلیل است که بسته به نوع بافت سلولی، E2F1 هم به عنوان انکوژن و هم به عنوان سرکوبگر عمل می‌کند. E2F4 در بسیاری از فرآیندهای سلولی و بیولوژیکی نقش دارد. به عنوان مثال تکثیر سلولی را تنظیم می‌کند و ژن‌های آپوپتوز را مهار می‌کند و یک فاکتور رونویسی پایه مهم با پتانسیل افزایش رشد تومور است. افزایش دو برابری بیان ژن E2F4 در تومورهای پستان در مقایسه با گروه کنترل مشاهده شده است (۱۱، ۳۶).

نتیجه گیری

با توجه به ناشناخته بودن جنبه‌های زیاد از وقایع مولکولی سرطان پستان به نظر می‌رسد چنین مطالعاتی سودمند باشند. اعتماد به اطلاعات حاصل از ادغام داده‌ها با توجه به افزایش تعداد داده‌ها و کاهش میزان نتایج کاذب بیشتر است. مطالعات در زمینه آنالیز داده در مورد سرطان‌ها تعداد زیادی دارند و استخراج نتایج جدید گزارش نشده بسیار کار متصور است. نتایج این مطالعه مهم‌ترین ژن‌های کاهش بیان یافته را ژن‌های LPL، ADIPOQ، LEP، PPARG، PLIN1، FABP4، LIPE، CIDEC و گزارش کرد و مهم‌ترین ژن‌های افزایش بیان یافته را ژن‌های AURKA، DLGAP5، NCAPG، CCNB1، KIF23، KIF11، BUB1B، RRM2، UBE2C و NUSAP1 معرفی کرد. همچنین پیشنهاد کرد که دو ریزآران‌ای has-miR-6782-5p و has-miR-6836-5p بیشترین ارتباطات را با ژن‌های کلیدی کاهش بیان یافته و دو ریزآران‌ای با نام‌های has-miR-7110-5p و has-miR7106-5p بیشترین ارتباطات را با ژن‌های کلیدی افزایش یافته را دارا هستند. PPARG به عنوان مهم‌ترین فاکتور رونویسی تنظیم‌کننده ژن‌های کلیدی کاهش بیان یافته و E2F1، FOXM1، E2F4 و MYC به عنوان مهم‌ترین فاکتورهای کنترل کننده ژن‌های کلیدی افزایش بیان معرفی شدند. برخی از نتایج به دست آمده در این مطالعه در هیچکدام از مقالات مرتبط با داده‌های حاصل از مطالعات سرطان پستان قبلی گزارش نشده است و از این جهت داده‌های بدست آمده ارزش بررسی‌های عملی بیشتر را دارند.

تضاد منافع

نویسندگان اعلام می‌کنند که هیچ تضاد منافی ندارند.

فهرست منابع

- 1- Acharya S, Cui L, and Pan Y (2020). Multi-view feature selection for identifying gene markers: a diversified biological data driven approach. *BMC bioinformatics*, 21: 1-31.
- 2- Alan O, Akin Telli T, Aktas B, Koca S, Ökten I N, Hasanov R, Basoglu T, Arikan R, Demircan N C, and Ercelep O (2020). Is insulin resistance a predictor for complete response in breast cancer patients who underwent neoadjuvant treatment? *World journal of surgical oncology*, 18: 1-9.
- 3- Cao M, Isaac R, Yan W, Ruan X, Jiang L, Wan Y, Wang J, Wang E, Caron C, and Neben S (2022). Cancer-cell-secreted extracellular vesicles suppress insulin secretion through miR-122 to impair systemic glucose homeostasis and contribute to tumour growth. *Nature cell biology*, 24: 954-967.
- 4- Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, and Lin C-Y (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology*, 8: 1-7.
- 5- Choi J, Cha Y J, and Koo J S (2018). Adipocyte biology in breast cancer: From silent bystander to active facilitator. *Progress in lipid research*, 69: 11-20.
- 6- Clarke D J B, Kuleshov M V, Schilder B M, Torre D, Duffy M E, Keenan A B, Lachmann A, Feldmann A S, Gundersen G W, and Silverstein M C (2018). eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic acids research*, 46: W171-W179.
- 7- Coventry B J and Henneberg M (2021). The immune system and responses to cancer: coordinated evolution. *F1000Research*, 4: 552.
- 8- Davis S and Meltzer P S (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, 23: 1846-1847.
- 9- Dhanasekaran R, Deutzmann A, Mahauad-Fernandez W D, Hansen A S, Gouw A M, and Felsher D W (2022). The MYC oncogene—the grand orchestrator of cancer growth and immune evasion. *Nature reviews Clinical oncology*, 19: 23-36.
- 10- Duncan R E, Ahmadian M, Jaworski K, Sarkadi-Nagy E, and Sul H S (2007). Regulation of lipolysis in adipocytes. *Annu. Rev. Nutr.*, 27: 79-101.
- 11- Farman F U, Haq F, Muhammad N, Ali N, Rahman H, and Saeed M (2018). Aberrant promoter methylation status is associated with upregulation of the E2F4 gene in breast cancer. *Oncology Letters*, 15: 8461-8469.
- 12- Fomicheva K, Osip'yants A, Knyazev E, Samatov T, and Shkurnikov M Y (2017). Detection of potential metastatic prostate cancer circulating biomarkers by comparison of miRNA profiles in DU145 cells and culture medium. *Bulletin of experimental biology and medicine*, 162: 792-796.
- 13- Francis R E, Myatt S S, Krol J, Hartman J, Peck B, McGovern U B, Wang J, Guest S K, Filipovic A, and Gojis O (2009). FoxM1 is a downstream target and marker of HER2 overexpression in breast cancer. *International journal of oncology*, 35: 57-68.
- 14- Glazer R I, Yuan H, Xie Z, and Yin Y (2008). PPAR γ and PPAR δ as modulators of neoplasia and cell fate. *PPAR research*, 2008: 247379.
- 15- Gobena S, Admassu B, Kinde M Z, and Gessese A T (2024). Proteomics and Its Current Application in Biomedical Area: Concise Review. *The Scientific World Journal*, 2024: 4454744.
- 16- Goldszmid R S, Dzutsev A, and Trinchieri G (2014). Host immune response to infection and cancer: unexpected commonalities. *Cell host & microbe*, 15: 295-305.
- 17- He B, Zhao Z, Cai Q, Zhang Y, Zhang P, Shi S, Xie H, Peng X, Yin W, and Tao Y (2020). miRNA-based biomarkers, therapies, and resistance in Cancer. *International journal of biological sciences*, 16: 2628.
- 18- Hsu C-C, Peng D, Cai Z, and Lin H-K, "AMPK signaling and its targeting in cancer progression and treatment," *Seminars in cancer biology*, 2022, 85:52-68.
- 19- Huang D W, Sherman B T, Tan Q, Collins J R, Alvord W G, Roayaei J, Stephens R, Baseler M W, Lane H C, and Lempicki R A (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, 8: 1-16.

- 20- Kwak M-S, Yim J Y, Yi A, Chung G-E, Yang J I, Kim D, Kim J S, and Noh D-Y (2019). Nonalcoholic fatty liver disease is associated with breast cancer in nonobese women. *Digestive and liver disease*, 51: 1030-1035.
- 21- Lee S, Jung Y, Bae Y, Yun S P, Kim S, Jo H, and Seo H-I (2017). Prevalence and risk factors of nonalcoholic fatty liver disease in breast cancer patients. *Tumori Journal*, 103: 187-192.
- 22- Lukasiewicz S, Czezelewski M, Forma A, Baj J, Sitarz R, and Stanisławek A (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. *Cancers*, 13: 4287.
- 23- Mello T, Materozzi M, and Galli A (2016). PPARs and mitochondrial metabolism: from NAFLD to HCC. *PPAR research*, 2016: 7403230.
- 24- Mukherjee A, Bilecz A J, and Lengyel E (2022). The adipocyte microenvironment and cancer. *Cancer and Metastasis Reviews*, 41: 575-587.
- 25- Munir R, Lisec J, Swinnen J V, and Zaidi N (2019). Lipid metabolism in cancer cells under metabolic stress. *British journal of cancer*, 120: 1090-1098.
- 26- O'Neill R S and Stoita A (2021). Biomarkers in the diagnosis of pancreatic cancer: Are we closer to finding the golden ticket? *World journal of gastroenterology*, 27: 4045.
- 27- Oliveros, J.C. (2024) Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/>
- 28- Quail D F and Joyce J A (2013). Microenvironmental regulation of tumor progression and metastasis. *Nature medicine*, 19: 1423-1437.
- 29- San Segundo-Val I and Sanz-Lozano C S (2016). Introduction to the gene expression analysis. *Molecular genetics of asthma*, 29-43.
- 30- Sedgwick P (2012). Pearson's correlation coefficient. *Bmj*, 345:e4483.
- 31- Shannon P, Markiel A, Ozier O, Baliga N S, Wang J T, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13: 2498-2504.
- 32- Sirbu A, Crane M, and Ruskin H J (2015). Data integration for microarrays: Enhanced inference for gene regulatory networks. *Microarrays*, 4: 255-269.
- 33- Smyth G K, Ritchie M, Thorne N, and Wettenhall J (2005). LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health.
- 34- Srinivasan M, Arzoun H, Gk L B, and Thangaraj S R (2022). A systematic review: Does insulin resistance affect the risk and survival outcome of breast cancer in women? *Cureus*, 14: e21712.
- 35- Sticht C, De La Torre C, Parveen A, and Gretz N (2018). miRWalk: an online resource for prediction of microRNA binding sites. *PloS one*, 13: e0206239.
- 36- Stubbs M, (2002) The E2F1 Transcription Factor, the ZF87 Transcription Factor, and the Control of Cell Cycle Transit, Thomas Jefferson University.
- 37- Szklarczyk D, Gable A L, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva N T, Morris J H, and Bork P (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47: D607-D613.
- 38- Tan Y, Wang M, Yang K, Chi T, Liao Z, and Wei P (2021). PPAR- α modulators as current and potential cancer treatments. *Frontiers in oncology*, 11: 599995.
- 39- Tang Z, Li C, Kang B, Gao G, Li C, and Zhang Z (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research*, 45: W98-W102.
- 40- Thu K, Soria-Bretones I, Mak T, and Cescon D (2018). Targeting the cell cycle in breast cancer: towards the next phase. *Cell Cycle*, 17: 1871-1885.
- 41- Villanueva R A M and Chen Z J, "ggplot2: elegant graphics for data analysis," ed: Taylor & Francis, 2019.
- 42- Xie Z, Bailey A, Kuleshov M V, Clarke D J, Evangelista J E, Jenkins S L, Lachmann A, Wojciechowski M L, Kropiwnicki E, and Jagodnik K M (2021). Gene set knowledge discovery with enrichr. *Current protocols*, 1: e90.
- 43- Xiong H, Chen Z, Chen W, Li Q, Lin B, and Jia Y (2020). FKBP-related ncRNA-mRNA axis in breast cancer. *Genomics*, 112: 4595-4607.

- 44- Zhou J (2024). Network Biology Reveals New Strategies for Understanding the Relationship Between Protein Function and Disease. *Computational Molecular Biology*, 14: 28-35.

In silico analysis of some potential molecular alterations in breast tumors compared to the adjacent normal tissues

Elaheh Zadeh Hosseingholi^{1*}, Haniyeh Musavi¹, Mohammadreza Sadeghi Dizaji², Ghader Molavi³

¹Department of Biology, Faculty of Basic Sciences, Azarbaijan Shahid Madani University, Tabriz, Iran

²Department of Molecular Medicine, Faculty of Advanced Medical Sciences, Tabriz, University of Medical Sciences, Tabriz, Iran

³ Emam Hossein Hospital, Tabriz University of Medical Sciences, Hashtroud, Iran.

**Corresponding Author: zhosseingholi@gmail.com*

Abstract

The incidence of breast cancer among women worldwide is on the rise. In recent decades, advancements in microarray technology have made it possible to collect and analyze extensive relevant data, allowing for a more precise identification of genes involved in the development and progression of cancers. In this study, we extracted and merged three datasets (GSE109169, GSE80754, and GSE72644) containing breast cancer tumor samples and adjacent healthy tissues from the NCBI database. Utilizing appropriate software, we identified key genes with altered expression and evaluated their co-expression. We also determined the gene ontology (GO) for these key genes and examined the associated microRNAs and transcription factors. Our findings revealed that the genes PPARG, LEP, ADIPOQ, LPL, LIPE, FABP4, PLIN1, and CIDEA exhibited significantly decreased expression. Conversely, ten genes—AURKA, DLGAP5, NCAPG, CCNB1, KIF23, KIF11, BUB1B, RRM2, UBE2C, and NUSAP1—showed increased expression. Additionally, two microRNAs, has-miR-6782-5p and has-miR-6836-5p, had the strongest associations with the down-regulated genes, while has-miR-7110-5p and has-miR-7106-5p were most closely linked to the up-regulated genes. Furthermore, PPARG was identified as the most important transcription factor regulating the key down-regulated genes, whereas E2F4, FOXM1, E2F1, and MYC were significant in controlling the up-regulated genes. The results of this study can inform the design of beneficial therapeutic methods to help prevent breast cancer.

Keywords: Breast cancer, Bioinformatics, Microarray, Hub genes