مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره سوم، شماره اول. ۲۸-۱۸، ۱۳۹۵

مقاله پژوهشی

پیش‌بینی اهداف MicroRNAهای دخیل در سرطان سینه با استفاده از روش‌های بیوانفورماتیکی

بلال صادقی[۱]*

• دریافت مقاله: ۹۴/۱۱/۲۸ • پذیرش مقاله: ۹۵/۱/۱۷

مقدمه: پیش‌بینی اهداف miRNAها از اهمیت ویژه‌ای برخوردار است. توسعه روش‌های محاسباتی و متعاقباً صرفه‌جویی در هزینه و زمان پژوهش‌های آزمایشگاهی، تأثیر به سزایی در افزایش سرعت ساخت داروهای درمانی از جمله داروهای ضد سرطانی دارد. با توجه به این که زمان کمی از شناخت miRNAها می‌گذرد، روند پژوهش‌ها در ابتدا کند بود. اما به مرور با شکل‌گیری دیتابیس‌های بیولوژیکی و درک اهمیت آن‌ها، دانشمندان بر روی این زمینه سرعت مطالعات و توجه خود را افزایش داده‌اند. تاکنون چندین روش کامپیوتری برای پیش‌بینی اهداف miRNA ها ساخته شده است ولی اکثر این روش‌ها دارای نرخ بالایی از مثبت‌های اشتباه هستند و هنوز جا برای بهبود این روش‌ها وجود دارد. از آنجا که مطالعات جدید نشان می‌دهد که miRNA ها در بافت‌های مختلف دارای اهداف متفاوتی هستند، هدف از این مقاله ارائه یک روش کامپیوتری برای پیش‌بینی اهداف، در سرطان سینه می‌باشد، تا به این طریق با اطمینان بیشتری بتوان به پیش‌بینی اهداف در این سرطان پرداخت.

روش: در این پژوهش، ابتدا انواع ویژگی‌ها استخراج شده، سپس ویژگی‌های برتر توسط دو روش CFS و ReliefF انتخاب گردیدند. انواع مدل‌های هوشمند از جمله شبکه عصبی، ماشین بردار پشتیبان با سه هسته متفاوت، الگوریتم نائیوبیز و درخت تصمیم‌گیری Random Forest بر روی داده‌ها با روش اعتبارسنجی ضربدری ۱۰ تایی تست و نتایج آن‌ها با یکدیگر مقایسه و تحلیل شد. سپس برای تأیید نتایج حاصل از داده‌های بیان ژن بهره گرفته شد.

نتایج: با استفاده از این روش ۱۲۴ اثر متقابل عملگر، شامل ۲۱ miRNA و ۳۸ mRNA، در سرطان مورد نظر پیش‌بینی گردید.

نتیجه‌گیری: این رویکرد از لحاظ بیوانفورماتیکی برای سرطان سینه مورد تأیید قرار گرفت اما برای تأیید بیشتر این پیش‌بینی‌ها، نیاز به استفاده از روش‌های آزمایشگاهی می‌باشد.

کلید واژه‌ها: پیش‌بینی اهداف miRNAها، سرطان سینه، ماشین بردار پشتیبان، بیز

•**ارجاع:** صادقی بلال. پیش‌بینی اهداف **microRNA**های دخیل در سرطان سینه با استفاده از روش‌های بیوانفورماتیکی. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۵؛ ۳(۱): ۲۸-۱۸.

۱. دکتری تخصصی ژنتیک حیوانی (ژنتیک مولکولی و بیوانفورماتیک)، استادیار گروه بهداشت مواد غذایی و بهداشت عمومی، دانشکده دامپزشکی، دانشگاه شهید باهنر کرمان، کرمان، ایران

* **نویسنده مسؤول:** کرمان، بزرگراه امام، میدان افضلی پور، دانشگاه شهید باهنر، دانشکده دامپزشکی، گروه بهداشت مواد غذایی و بهداشت عمومی.

• **شماره تماس:** ۰۳۴-۳۳۲۲۲۹۶۲

Email: sadeghi.balal@uk.ac.ir

[Downloaded from jhbmi.com on 2024-10-12]

مقدمه

سرطان سینه از رشد کنترل نشده سلول‌ها در ناحیه سینه حاصل می‌شود. بررسی‌ها نشان می‌دهد این سرطان حاصل یک فرآیند مرحله به مرحله و در نتیجه تجمع تغییرات ژنتیکی متفاوت است. تخمین‌های هر ساله در آمریکا و جوامع دیگر نشان داده است که فراوانی افرادی که از سرطان در معرض مرگ و میر قرار می‌گیرند رو به افزایش است. در حال حاضر آمار در ایالات متحده نشان می‌دهد که از هر ۴ مرگ، یک مورد در اثر ابتلاء به سرطان است [۱]. این مشکل در تمامی جمعیت‌ها و از جمله ایران به عنوان یک مشکل اساسی و رو به رشد مورد توجه است. تحقیقات متعدد پیرامون کشف روش‌های جدید در ارتباط با مکانیسم مولکولی سرطان و یافتن راه مناسب برای درمان مؤثر در حال پیشرفت است. مطالعات اخیر از طریق شناسایی نقش مؤثر miRNA ها در شروع و پیشرفت سرطان سینه، روزنه جدیدی را در جهت بهبود استراتژی‌های درمانی علیه این سرطان باز کرده است [۴-۲].

آنالیزهای اخیر رونوشت‌ها نشان داده است که ۹۰٪ ژنوم انسان به طور فعالی رونویسی می‌شود که از این میان تنها ۲٪ شامل ژن‌های رمزگذار پروتئین هستند و بقیه متعلق به خانوادهRNA های غیر رمز گذار است که هرگز به پروتئین ترجمه نمی‌گردند [۵]. یک دسته اصلی این خانواده miRNAهایی هستند که گروه بزرگی ازRNA های کوچک ۲۱-۲۲ نوکلئوتیدی را تشکیل می‌دهند و تنظیم بیان ژن‌های یوکاریوتی را بر عهده دارند و از اجزای کلیدی تنظیم بیان ژن در گیاهان و جانوران محسوب می‌شوند. این مولکول‌ها در طی تکامل حفظ شده‌اند و تنظیم بیان ژن را در مرحله بعد از رونویسی انجام می‌دهند. تنظیم و کاهش بیان ژن توسط miRNAها می‌تواند یا از طریق القای برش mRNA های هدف باشد یا از طریق سرکوب ترجمه آن‌ها انجام گیرد [۶]. در واقع با کشف نقش miRNAها در تنظیم پس از رونویسی ژن‌ها، "اصل مرکزی" زیست‌شناسی مولکولی که برای سالیان متمادی تنظیم بیان ژن‌ها را یک فرآیند خطی از DNA به RNA و سپس به پروتئین در نظر می‌گرفت نیاز به بازنگری دارد. در پستانداران miRNAها در همه کروموزوم‌ها به استثناء کروموزوم Y به صورت خوشه و یا منفرد پراکنده شده‌اند. حدود ۴۲٪ miRNA ها در نواحی بین ژنی و بقیه در نواحی درون ژنی قرار گرفته‌اند [۷].

سرطان نتیجه خروج سلول‌ها از مسیرهای درست تنظیم، تکثیر و تمایز است. غیرحساس شدن به سیگنال‌های مهارکننده رشد، خودکارآمدی در سیگنال‌های رشد، جلوگیری از مرگ سلولی برنامه‌ریزی شده، توانایی نامحدود تکثیر، حفظ رگ‌زایی و تهاجم بافتی و متاستاز منجر به بدخیم شدن سرطان می‌شوند. اولین گزارش در مورد نقش miRNAها در ایجاد سرطان در سال ۲۰۰۲ منتشر شد. طبق مطالعات صورت گرفته مشخص شد که miR-16 و miR-15 روی کروموزوم ۱۳ در ناحیه‌ای قرار گرفته‌اند که اغلب در CLL حذف می‌شود. Calin و همکاران کشف کردند که miR-16 و miR-15 در بیشتر از ۶۰٪ سلول‌های B انسان‌های دچار CLL حذف شده و یا کاهش بیان پیدا کرده است [۸]. برهمکنش miRNA ها با ژن‌های هدف، نقش آن‌ها را در رشد، مرگ برنامه‌ریزی شده، تمایز و تکثیر سلولی مشخص کرده و عملکرد مستقیم آن‌ها را در سرطان تأیید می‌کند [۱۰، ۹]. می‌توان از بیانmiRNA های غالب برای طبقه‌بندی سرطان‌ها در گروه‌هایی با ویژگی‌های متفاوت مثل نوع سلول و سبب‌شناسی بهره گرفت [۱۱].

استفاده از miRNA ها برای طبقه‌بندی تومور به مراتب مناسب‌تر از mRNA ها می‌باشد. مطالعات نشان داده است الگوی بیان miRNA ۲۱۷ بهتر از الگوی بیان ۱۶۰۰۰ mRNA، به طبقه‌بندی و تشخیص تومورها از یکدیگر قادر است که این امر به علت جفت شدگی ناقص بین miRNA با mRNA های هدف می‌باشد که در نتیجه، یک miRNA می‌تواند بیان چندین mRNA و بنابراین چندین مسیر را تنظیم کند [۱۱]. اولین گزارش در مورد عملکرد زیستی miRNAها در سرطان در سال ۲۰۰۵ منتشر شد و مشخص شد که miRNAها می‌توانند به عنوان آنکوژن یا تومور ساپرسور عمل کنند. آن‌هایی که به عنوان آنکوژن و یا القاء کننده تومور عمل می‌کنند انکومیر نامیده می‌شوند. انکومیرها در بدخیمی بافت‌های مختلف حضور دارند و اغلب در مناطقی از ژنوم که دچار حذف شدگی، مضاعف شدگی و یا موتاسیون شده‌اند، یافت می‌شوند. این دسته از miRNAها فنوتیپ آنکوژنی را از طریق کاهش بیان ژن‌های سرکوب کننده تومور و یا ژن‌های تنظیم کننده تمایز سلولی و مرگ برنامه‌ریزی شده ایجاد می‌کنند [۱۲]. دسته دیگر با هدف قرار دادن mRNAهای پروتوانکوژن و خاموش کردن آن‌ها منجر به کاهش روند سرطانی شدن می‌شوند که به این دسته miRNAها تومور ساپرسور گفته می‌شود. اغلب مطالعات صورت گرفته نشان دهنده حضور miRNAها در نواحی از ژنوم است که مستعد حذف، تکثیر و جهش هستند که نواحی ژن‌های همراه با سرطان نامیده می‌شوند. ژن‌های

[Downloaded from jhbmi.com on 2024-10-12]

miRNA قرار گرفته در داخل و یا در نزدیکی این نواحی ظاهراً با رویدادهای کروموزومی همراه می‌شوند که منجر به تومورزایی می‌شود [۱۳]. مشخص شدن نقش مهم miRNA ها در فرآیندهای متفاوت سلولی مانند همانندسازی، تمایز، آپوپتوز و دخالت آن‌ها در شروع و پیشرفت سرطان، آن‌ها را به اهداف درمانی مهمی در درمان بیماری‌ها و از جمله سرطان تبدیل کرده است. برای استفاده از miRNA ها در درمان سرطان به طور کلی می‌توان از دو استراتژی: مهار miRNA های آنکوژن و درمان براساس جایگزینی miRNA استفاده کرد [۱۴].

از مسائل مهم در زمینه miRNA ها، شناسایی اهداف آن‌ها است. مطالعات MiRNAomics، اختلال در میزان بیان miRNA ها را در طیف گسترده‌ای از بدخیمی‌های هماتولوژیک و تومورهای جامد، از جمله سرطان سینه آشکار کرده‌اند. پیش‌بینی اهداف miRNA ها در کنترل و بهبود بیماری‌ها از جمله سرطان بسیار حائز اهمیت است. با توجه به نقش کنترلی miRNA ها در خاموش کردن و کاهش سطح بیان ژن‌ها، دانشمندان می‌توانند با تحقیق بر روی این ژن‌ها و miRNA هایی که فرآورده RNA آن‌ها را مورد هدف قرار می‌دهند بر روی تکنولوژی‌هایی متمرکز شوند که با استفاده از آن‌ها بتوان با تزریق کنترل شده miRNAها، ژن‌های نامطلوب را مورد هدف قرار داد و بیماری مورد نظر را کنترل و در حالت خوشبینانه درمان کرد. تا سال ۲۰۰۳ بیشتر شناسایی miRNA ها و اهداف آنان از طریق روش‌های آزمایشگاهی و بیولوژی مولکولی که روش‌هایی زمان‌بر و بسیار پرهزینه محسوب می‌شوند، صورت می‌گرفت چرا که روش‌های کامپیوتری اندکی برای پیش‌بینی اهداف miRNAها در اختیار بود. حال با در نظر گرفتن اهمیت ژن‌های miRNA در مکانیسم‌های مختلف سلولی و در ایجاد سرطان‌های مختلف از جمله سرطان سینه که به آن‌ها اشاره گردید، مسلماً شناسایی اهداف آن‌ها می‌تواند گام بلندی در جهت کنترل و بهبود بیماری‌های مختلف از جمله سرطان سینه باشد. بنابراین استفاده از روش‌های پیش‌بینی کننده محاسباتی قابل اعتماد، الزامی است. تاکنون چندین روش کامپیوتری برای پیش‌بینی اهداف miRNAها ساخته شده است ولی اکثر این روش‌ها دارای نرخ بالایی از مثبت‌های اشتباه هستند و هنوز جا برای بهبود این روش‌ها وجود دارد. از آنجا که مطالعات جدید نشان می‌دهد که miRNAها در بافت‌های مختلف دارای اهداف متفاوتی هستند، هدف از این تحقیق ارائه یک روش کامپیوتری جدید برای پیش‌بینی اهداف، تنها در سرطان سینه می‌باشد، تا به این طریق با اطمینان بیشتری بتوان به پیش‌بینی اهداف در این سرطان پرداخت. در واقع هدف این پژوهش، شناسایی ژن‌های هدف miRNA های دخیل در سرطان سینه است که از طریق روش‌های تجربی و محاسباتی قابل انجام است. به خاطر محدودیت و هزینه زیاد روش‌های آزمایشگاهی و این که مطالعه جامع در آن‌ها بسیار پر هزینه است، راه‌های محاسباتی و بیوانفورماتیکی از اهمیت بسیار برخوردارند چرا که می‌توانند راهنمایی برای کارهای تجربی باشند.

روش

روش‌های بیوانفورماتیکی

به طور کلی، فعالیت‌های انجام شده در این پروژه شامل جمع‌آوری داده‌های مختص سرطان سینه و داده‌های منفی، استخراج ویژگی‌ها، بررسی و انتخاب ویژگی‌ها و تست انواع دسته‌بندی کننده‌های هوشمند در جواب نهایی و پیش‌بینی اهداف جدید توسط الگوریتم ارائه شده در این مطالعه است.

جمع‌آوری داده

با توجه به ماهیت راه حل ارائه شده برای پیش‌بینی ژن‌های هدف miRNA ها که در واقع استفاده از یک مدل با نظارت است، لازم بود داده‌های آموزشی، در آزمایشگاه تأیید شده باشند. به علاوه داده‌های تأیید شده آزمایشگاهی نیز باید مورد بررسی قرار می‌گرفت تا در نهایت مجموعه داده قابل اطمینانی که سایت دقیق اتصال آن‌ها مشخص شده است، به دست آید.

جمع‌آوری داده‌های مثبت

برای جمع‌آوری داده‌های مثبت به مرور منابع موجود در پایگاه داده Pubmed پرداخته شد. در این مرحله با کلید واژه‌های miRNA target و سرطان سینه، جستجو برای شناسایی مقالاتی که به شناسایی اهدافmiRNAها در سرطان سینه پرداخته‌اند صورت گرفت. از بین نتایج حاصل تنها اطلاعات مربوط به مطالعات تأیید شده با روش آزمایشگاهی گزارشگر لوسیفراز و جهش هدف‌دار به عنوان داده‌های مثبت جمع‌آوری شدند، زیرا با استفاده از این دو روش آزمایشگاهی جایگاه دقیق اتصال miRNA روی mRNA هدف مشخص می‌شود.

جمع‌آوری داده‌های منفی

در این مرحله، برخلاف اکثر الگوریتم‌های موجود که تنها از برهمکنش‌های منفی، یعنی برهمکنش‌هایی که در واقعیت وجود ندارد و تنها ساخته‌های کامپیوتری هستند استفاده کرده‌اند، روش استفاده شده در این تحقیق، از برهمکنش‌های تأیید شده برای

[Downloaded from jhbmi.com on 2024-10-12]

سرطان‌های دیگر به عنوان داده‌های منفی استفاده شد. برای این منظور از بانک داده miR2Disease برای استخراج داده‌های مربوط به سرطان ریه و کولون استفاده شد. به علاوه از داده‌های منفی ارائه شده توسط Mitra و همکاران که از کیفیت بالایی برخوردار بودند، نیز استفاده گردید تا الگوریتم قادر به شناسایی این نوع از داده‌ها و تشخیص آن‌ها از داده‌های مثبت باشد [۱۵]. از آنجا که در مقاله منتشر شده هیچ اطلاعاتی در مورد سایت اتصال miRNA ارائه نشده است ما از نرم افزار اسمیت-واترمن (Smith-Waterman) جهت بررسی همپوشانی توالی‌های miRNA بر علیه توالی‌های mRNA استفاده کردیم در اجرای این الگوریتم علاوه بر جفت شدگی واتسون-کریک، جفت شدگی کدون هرزG:U نیز مورد قبول واقع شد. امتیازهای جفت‌شدگی بینG:C ، A:U و G:U به ترتیب هفت، پنج و یک در نظر گرفته شد. امتیاز عدم جفت شدگی ۳- ایجاد فاصله ۸- و توسعه آن ۲- در نظر گرفته شد.

استخراج ویژگی‌ها

در این مرحله با استفاده از مرور منابع، ویژگی‌های لازم جهت استخراج از داده‌های جمع‌آوری شده مشخص شد. به طور کلی ویژگی‌هایی را که تاکنون استفاده شده است می‌توان در چهار گروه توالی نوکلئوتیدی، ساختاری، وابسته به موقعیت و ترمودینامیکی طبقه‌بندی کرد. در این تحقیق از ۶۰ نوع ویژگی (۳۳ ویژگی توالی نوکلئوتیدی، ۱۸ ویژگی وابسته به ساختمان، هشت ویژگی وابسته به موقعیت و یک ویژگی ترمودینامیکی) استفاده شد. لیست ویژگی‌های به کار رفته و گروه هر یک از آن‌ها به همراه توضیحات در جدول ۱ آورده شده است.

جدول ۱: لیست ویژگی‌های استفاده شده

	توضیح	ویژگی	گروه
۱	تعداد جفت شدگی‌ها در ناحیه جفت‌شدگی اصلی هشت مری	Seed	ساختاری
۲	تعداد جفت شدگی‌های پشت سرهم در انتهای ۵ miRNA	Cons_bp_mir_5p	ساختاری
۳	موقعیت ۳	Pos_3	وابسته به موقعیت
۴	موقعیت ۲	Pos_2	وابسته به موقعیت
۵	تعداد جفت شدگی‌ها به ۸ نوکلئوتید انتهای ۳ miRNA	Tail	ساختاری
۶	فراوانی جفت‌های CG در ناحیه جفت‌شدگی اصلی	CG_match	ساختاری
۷	فراوانی جفت‌های GC در ناحیه جفت‌شدگی اصلی	GC_match	ساختاری
۸	موقعیت ۶	Pos_6	وابسته به موقعیت
۹	موقعیت ۴	Pos_4	وابسته به موقعیت
۱۰	موقعیت ۵	Pos_5	وابسته به موقعیت
۱۱	موقعیت ۷	Pos_7	وابسته به موقعیت
۱۲	فراوانی جفت‌های AU در ناحیه جفت‌شدگی اصلی	AU_match	ساختاری
۱۳	فراوانی جفت‌های UA در ناحیه جفت‌شدگی اصلی	UA_match	ساختاری
۱۴	فراوانی دی‌نوکلئوتید AC در ناحیه جفت‌شدگی اصلی	AC_seed	توالی نوکلئوتیدی
۱۵	فراوانی دی‌نوکلئوتید UC در ناحیه جفت‌شدگی اصلی	UC_seed	توالی نوکلئوتیدی
۱۶	فراوانی نوکلئوتید G در ناحیه خارج ناحیه جفت‌شدگی اصلی	G_outseed	توالی نوکلئوتیدی
۱۷	موقعیت ۱	Pos_1	وابسته به موقعیت
۱۸	فراوانی G و C در رشته هدف	G+C_target	توالی نوکلئوتیدی
۱۹	فراوانی دی‌نوکلئوتید GC در ناحیه جفت‌شدگی اصلی	GC_seed	توالی نوکلئوتیدی
۲۰	فراوانی دی‌نوکلئوتید GC در ناحیه خارج ناحیه جفت‌شدگی اصلی	GC_outseed	توالی نوکلئوتیدی
۲۱	نسبت محتوای G+C در miRNA به سایت هدف	GC_ratio	وابسته به موقعیت
۲۲	فراوانی GC-CG در ناحیه جفت‌شدگی اصلی	GC-CG	ساختاری
۲۳	فراوانی دی‌نوکلئوتید AU در ناحیه جفت‌شدگی اصلی	AU_seed	توالی نوکلئوتیدی
۲۴	فراوانی UA-AU در ناحیه جفت‌شدگی اصلی	UA-AU	ساختاری
۲۵	فراوانی دی‌نوکلئوتید CC در ناحیه جفت‌شدگی اصلی	CC_seed	توالی نوکلئوتیدی
۲۶	توضیح در متن	Compactness	ساختاری
۲۷	فراوانی جفت‌های GU در ناحیه جفت‌شدگی اصلی	GU_match	ساختاری
۲۸	موقعیت ۸	Pos_8	وابسته به موقعیت
۲۹	گپ با سایز ۱ در رشته miRNA	B_miR_s1	ساختاری
۳۰	طویل‌ترین رشته از جفت بازهای پشت سرهم	Stem	ساختاری
۳۱	فراوانی دی‌نوکلئوتید UA در ناحیه جفت‌شدگی اصلی	UA_seed	توالی نوکلئوتیدی
۳۲	فراوانی دی‌نوکلئوتید GA در ناحیه خارج ناحیه جفت‌شدگی اصلی	GA_outseed	توالی نوکلئوتیدی
۳۳	اتصال عدم تقارن	Binding asymmetry	ساختاری
۳۴	گپ با سایز ۲ در رشته هدف	B_tagt_s2	ساختاری
۳۵	حداقل انرژی آزاد	MFE	ترمودینامیکی
۳۶	فراوانی دی‌نوکلئوتید AU در ناحیه خارج ناحیه جفت‌شدگی اصلی	AU_outseed	توالی نوکلئوتیدی
۳۷	فراوانی GC-GC در ناحیه جفت‌شدگی اصلی	GC-GC	ساختاری
۳۸	فراوانی جفت شدگی ها در موقعیت نوکلئوتید ۱۳-۱۷	WC_bp_po_12-17	ساختاری
۳۹	فراوانی دی‌نوکلئوتید UU در ناحیه جفت‌شدگی اصلی	UU_seed	توالی نوکلئوتیدی
۴۰	گپ با سایز ۲ در رشته miRNA	B_miR_s2	ساختاری
۴۱	تعداد کل جفت شدگی‌ها	Max_bp	ساختاری
۴۲	فراوانی GU-UA در ناحیه‌ی جفت‌شدگی اصلی	GU-UA	ساختاری
۴۳	گپ با سایز ۴ در رشته هدف	B_tagt_s4	ساختاری
۴۴	فراوانی AU-UG در ناحیه جفت‌شدگی اصلی	AU-UG	ساختاری
۴۵	فراوانی دی‌نوکلئوتید UU در ناحیه خارج ناحیه جفت‌شدگی اصلی	UU_outseed	توالی نوکلئوتیدی

[Downloaded from jhbmi.com on 2024-10-12]

جدول ۱: لیست ویژگی‌های استفاده شده (ادامه)

۴۶	گپ با سایز ۳ در رشته miRNA	B_miR_s3	ساختاری
۴۷	فراوانی GU-CG در ناحیه جفت‌شدگی اصلی	GU-CG	ساختاری
۴۸	فراوانی AU-CG در ناحیه جفت‌شدگی اصلی	AU-CG	ساختاری
۴۹	فراوانی CG-UA در ناحیه جفت‌شدگی اصلی	CG-UA	ساختاری
۵۰	فراوانی UA-CG در ناحیه جفت‌شدگی اصلی	UA-CG	ساختاری
۵۱	فراوانی UA-UG در ناحیه جفت‌شدگی اصلی	UA-UG	ساختاری
۵۲	گپ با سایز ۶ در رشته هدف	B_tagt_s6	ساختاری
۵۳	گپ با سایز ۵ در رشته هدف	B_tagt_s5	ساختاری
۵۴	گپ با سایز ۱ در رشته هدف	B_tagt_s1	ساختاری
۵۵	گپ با سایز ۷ در رشته هدف	B_tagt_s7	ساختاری
۵۶	گپ با سایز ۵ در رشته miRNA	B_miR_s5	ساختاری
۵۷	فراوانی دی‌نوکلئوتید AC در ناحیه خارج ناحیه جفت‌شدگی اصلی	AC_outseed	توالی نوکلئوتیدی
۵۸	گپ با سایز ۳ در رشته هدف	B_tagt_s3	ساختاری
۵۹	فراوانی دی‌نوکلئوتید UC در ناحیه خارج ناحیه جفت‌شدگی اصلی	UC_outseed	توالی نوکلئوتیدی
۶۰	فراوانی UA-GU در ناحیه جفت‌شدگی اصلی	UA-GU	ساختاری

جدول ضمیمه ۱: تعدادی از داده‌های استفاده شده جهت آموزش مدل

ژن هدف	miRNA	mRNA (5' to 3')	miRNA (3' to 5')	Pubmed ID
BCL2	miR-148a	GCUCCGUUUGGCAGUGCAAUGG	UGUUUCAAGACAUCACGUGACU	21455217
BCL2	miR-365	AUCUGUACAUCCUGGGGCAUUA	UAUUCCUAAAAAUCCCCGUAAU	22072615
BCL2	miR-365	AUGGAAUAACUCUGUGGCAUUA	UAUUCCUAAAAAUCCCCGUAAU	22072615
BCL6	miR-127	CUGCAGGCAGACACGGAUCUGA	UCGGUUCGAGUCUGCCUAGGCU	16766263
IGF1R	miR-139-5p	UACCAGUUUCAAUCACUGUAGA	GACCUCUGUGCACGUGACAUCU	22580051
IGF1R	miR-143	UUUAUAUGUGUAUAUAGACAAAAGAAUACAUCUCA	CUCGAUGUCACGAAGUAGAGU	25474488
IGF1R	miR-145	GAUUAUUAUUUGGGGGAACUGGAC	UCCCUAAGGACCCUUUUGACCUG	25474488
IGF1R	miR-497	ACAGAAAAAAAAAAAGCUGCUA	UGUUUGGUGUCACACGACGAC	22710713
KRAS	miR-143	AUGUUAAAAGAAGUCAUCUCA	CUCGAUGUCACGAAGUAGAGU	19137007
KRAS	miR-143	AAGUUACAGUUUGCACAAGUUCAUCUCA	CUCGAUGUCACGAAGUAGAGU	19137007

انتخاب ویژگی

به وضوح به اثبات رسیده است که انتخاب ویژگی نقش بسیار مهمی در شناسایی جایگاه‌های اتصال به عهده دارد. همواره دقت و صحت نتایج پیش‌بینی، ارتباط مستقیمی با تعداد و نوع ویژگی‌های انتخاب شده برای پیاده‌سازی الگوریتم دارد. انتخاب ویژگی فرآیندی است که زیرمجموعه‌ای از ویژگی‌های اصلی را انتخاب می‌کند و یکی از مهم‌ترین تکنیک‌های پیش‌پردازش داده است. در واقع ویژگی‌های مناسب ناشناخته‌اند، بنابراین انتخاب ویژگی باید ویژگی‌های اضافی و نامناسب را شناسایی و حذف کند. در این تحقیق برای انتخاب ویژگی‌هایی که نقش مؤثرتری در تشخیص نمونه‌های مثبت از منفی دارند از دو روش ReliefF و Correlation-based Feature Selection (CFS)استفاده شد [۱۶]. تفاوت کلیدی بین این دو روش در این است که روش ReliefF نقش هر ویژگی را به طور جداگانه مورد بررسی قرار داده و بر این اساس به هر ویژگی نمره‌ای داده می‌شود ولی روش CFS ارتباط بین ویژگی‌ها را نیز مورد بررسی قرار داده و در نهایت بهترین زیر مجموعه از ویژگی‌ها را به عنوان نتیجه نهایی انتخاب کرده و ارائه می‌کند.

برای اجرای این روش‌های انتخاب ویژگی از نرم‌افزار داده کاوی Weka [۱۷] استفاده شد.

داده‌های بیانی miRNA و mRNA گرفته شده از بیماران مبتلا به سینه سینه

پروفایل بیان ریزآرایه برای miRNA و mRNA از ۲۱۰ بیمار مبتلا به سرطان سینه از بانک داده (Gene Omnibus GEO(Expression با شماره سریال GSE22216 که توسط Buffa و همکاران انجام شده بود، گرفته شد [۱۸]. سپس به منظور شناسایی ژن‌هایی که به طور متفاوت بیان می‌شوند از تحلیل بیان افتراقی ژن‌ها استفاده شد و مشخص شد که ۱۴ miRNA افزایش بیان و ۴۳ mRNA کاهش بیان را در بافت سرطانی نشان می‌دهند (FDR<۰/۰۵). با استفاده از روش استفاده شده در این پروژه، به پیش‌بینی برهمکنش‌های احتمالی بین این miRNA ها و mRNA ها که بیان معکوسی را در بافت سرطانی سینه نشان داده‌اند، پرداخته شد.

[Downloaded from jhbmi.com on 2024-10-12]

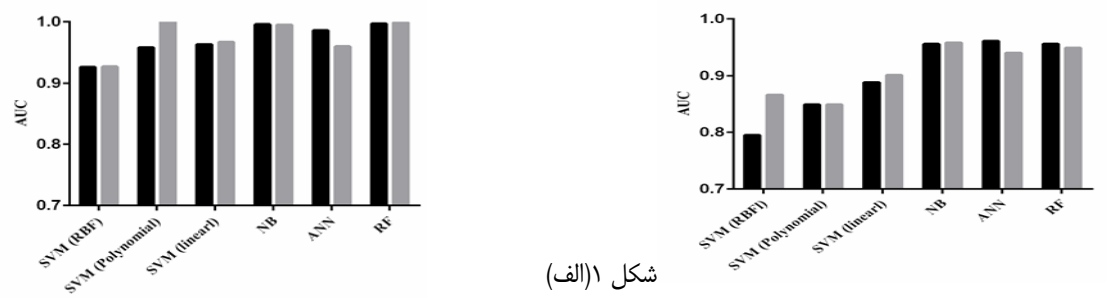
نتایج

توسط روش CFS، ۱۱ ویژگی به عنوان بهترین زیر مجموعه از ویژگی‌ها برای تشخیص برهمکنش‌های miRNA-mRNA مخصوص سرطان سینه انتخاب شد (جدول ۲). از میان ویژگی‌های انتخاب شده، شش ویژگی مربوط به ویژگی‌های ساختاری، سه ویژگی مربوط به توالی و دو عدد باقیمانده مربوط به موقعیت بود. ۱۱ ویژگی برتر انتخاب شده توسط روش ReliefF در جدول ۲ آورده شده است. همچنین در این جدول ویژگی‌های مشترک انتخاب شده توسط روش CFS و روش ReliefF به صورت ستاره مشخص گردیده است.

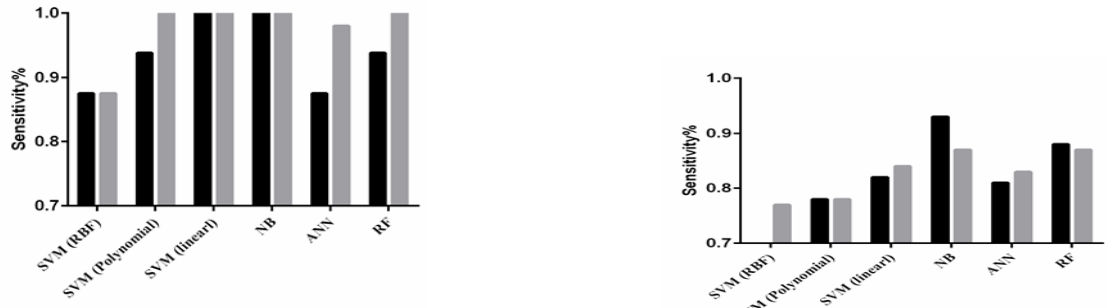
جدول ۲: لیست ۱۱ ویژگی برتر انتخاب شده توسط روش CFS

ویژگی	توضیح	اشتراک با ReliefF
G+C_target	فراوانی G و C در سایت هدف	
Seed	تعداد جفت شدگی‌ها در ناحیه جفت‌شدگی اصلی هشت مری	*
Tail	تعداد جفت شدگی‌ها به ۸ نوکلئوتید انتهای ۳ miRNA	*
CG_match	فراوانی جفت باز CG در ناحیه ناحیه جفت‌شدگی اصلی	*
GC_match	فراوانی جفت باز GC در ناحیه ناحیه جفت‌شدگی اصلی	*
AU_match	فراوانی جفت باز AU در ناحیه ناحیه جفت‌شدگی اصلی	*
Binding asymmetry	اتصال عدم تقارن	
B_tagt_s6	تعداد گپ‌ها با سایز ۶ در توالی هدف	
Pos_3	موقعیت ۳	*
UA_match	فراوانی جفت باز UA در ناحیه ناحیه جفت‌شدگی اصلی	*
Pos_7	موقعیت ۷	*

شکل ۱ (الف) نتایج حاصل از این مرحله را نشان می‌دهد. اجرای این طبقه‌بندها بر روی داده‌های آزمون مستقل نیز مورد ارزیابی قرار گرفت. همان طور که در شکل ۱ مشاهده می‌شود طبقه بند نائیوبیز تعمیم پذیری بهتری را نشان می‌دهد و در نتیجه به عنوان الگوریتم مرکزی مدل ارائه شده در این مقاله مورد استفاده قرار گرفت.



شکل ۱(الف)



شکل ۱(ب)

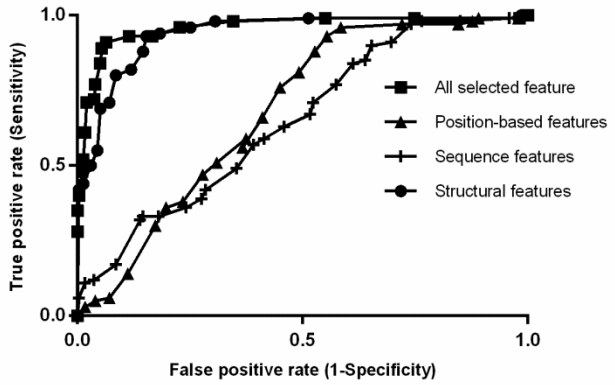
شکل ۱: نتایج حاصل از عملکرد الگوریتم‌های طبقه‌بند بر روی داده‌های آموزشی (اعتبار سنجی ضربدری ۱۰تایی) (الف). نتایج حاصل از عملکرد همان طبقه‌بندها بر روی داده‌های آزمون (ب).

[Downloaded from jhbmi.com on 2024-10-12]

محور افقی نحوه عملکرد طبقه‌بندهای متفاوت را براساس حساسیت و سطح زیر منحنی نشان می‌دهد. رنگ تیره و رنگ روشن در نمودارها به ترتیب نشان دهنده روش CFS و روش ReliefF می‌باشد.

مقایسه عملکرد طبقه‌بند نائیوبیز بر روی انواع متفاوتی از ویژگی‌ها

هدف این قسمت آزمایش شناسایی مؤثرترین نوع از ویژگی‌ها در میان ویژگی‌های انتخاب شده توسط روش CFS بر روی عملکرد مدل ارائه شده در این مقاله بود. عملکرد طبقه بند براساس روش اعتبارسنجی ضربدری ۱۰ تایی و با استفاده از مساحت زیر سطح منحنی ROC مورد بررسی قرار گرفت (شکل ۲). مقدار AUC براساس تمام ۱۱ ویژگی انتخاب شده، ۹۶٪ بود. برای ویژگی‌های ساختاری AUC به ۹۳٪ کاهش یافت، بدون این که اثر قابل توجهی در میزان حساسیت طبقه بند حاصل شود. برای ویژگی‌های توالی، مقدار AUC کاهش قابل توجهی را تا ۶۷٪ نشان داد و در نهایت با استفاده از ویژگی‌های مبتنی بر موقعیت، AUC به ۶۸٪ کاهش یافت.



شکل ۲: نمودار منحنی ROC مربوط به عملکرد ماشین بردار پشتیبان بر روی انواع متفاوتی از ویژگی‌ها

مقایسه عملکرد روش مورد استفاده در این تحقیق با الگوریتم‌های موجود

در این مرحله روش مورد استفاده با الگوریتم miTarget اولین الگوریتم ارئه شده مبتنی بر آموزش ماشینی مورد مقایسه قرار گرفت. علاوه براین با دو الگوریتم دیگر multiMitar و RFmiRTarget که تاکنون بهترین نتایج را در میان روش‌های مبتنی بر آموزش ماشینی ارائه کرده‌اند مقایسه شد. به این منظور ویژگی‌های استفاده شده توسط این الگوریتم‌ها شناسایی و استخراج شد و سپس با استفاده از داده‌های آموزشی استفاده شده در این تحقیق، دوباره آموزش داده شدند. الگوریتم مرکزی استفاده شده در مدل miTarget و multiMitar ماشین بردار پشتیبان و در مدل RFmiRTarget درخت تصمیم Ranodm Forest است. روش استفاده شده در مقایسه سایر روش‌ها از لحاظ معیارهای دقت کلی، F-measure و ضریب همبستگی متیو نشان می‌دهد، که روش بهتری می‌باشد (جدول ۳).

جدول ۳: مقایسه روش استفاده شده در این مطالعه با تعدادی از الگوریتم‌های موجود

	Sensitivity	Specificity	MCC	AUC
روش مطالعه	۰/۷۹	۰/۹۱	۰/۷۸	۰/۹۶
MultiMiTar	۰/۷۷	۰/۹۶	۰/۷۷	۰/۹۴
miTarget	۰/۶۳	۰/۹۲	۰/۵۸	۰/۸۸
RFMirTarget	۰/۶۹	۰/۹۴	۰/۶۶	۰/۹۲

[Downloaded from jhbmi.com on 2024-10-12]

پیش‌بینی برهمکنش miRNA-mRNA جدید سرطان سینه با استفاده از داده‌های بیان ژن از الگوریتم مورد استفاده شده برای پیش‌بینی جفت‌های miRNA-mRNA که پروفایل بیان معکوس در بافت سرطان سینه دارند، استفاده شد. مدل ارائه شده در این مقاله توانست ۱۲۴ جفت برهمکنش شامل ۲۱ miRNA و ۳۸ mRNA را شناسایی کند. پیش‌بینی‌های حاصل با داده‌های موجود در بانک داده miR2Disease و OncomiRDB مورد مقایسه قرار گرفت و مشخص شد که پنج مورد از آن‌ها قبلاً توسط روش‌های آزمایشگاهی تأیید شده‌اند. جدول ۴ تعدادی از اهداف پیش‌بینی شده برای هر miRNA را نشان می‌دهد.

جدول ٤: تعدادی از اهداف پیش‌بینی شده برای miRNA های افزایش بیان یافته در سرطان سینه

miRNA	mRNA
hsa-miR-25-3p	SFRP2, SNCA, ST3GAL5, STAB2, XKR4, XKR6, ZFPM2
hsa-miR-29a-3p	ARHGAP24, FOXN3, LPP, NBEA, PDE5A, SFRP1, XKR4
hsa-miR-29b-3p	ANKRD12, CAMK2D, CNR1, DOCK5, FAM46A, FOS, FOXN3,
hsa-miR-424-5p	ARHGAP24, CAMK2D, CDC14A, CHP2, CNR1, DOCK5, FOXN3,
hsa-miR-7-5p	ANKRD12, ARHGAP24, CAMK2D, CHP2, CNR1, DOCK5, FAM46A,
hsa-miR-93-5p	ANKRD12, ARHGAP24, CAMK2D, CBX7, CDC14A, CHP2, DOCK5,
hsa-miR-96-5p	ANKRD12, ARHGAP24, CAMK2D, CBX7, CNR1, CRTAM, DIRAS1,

بسیاری از گزارشات همراهی miRNA ها را با بیماری‌ها مورد بررسی قرار داده‌اند. مطالعات MiRNAomics، اختلال در میزان بیان miRNA ها را در طیف گسترده‌ای از بدخیمی‌های هماتولوژیک و تومورهای جامد، از جمله سرطان سینه آشکار کرده‌اند. شناسایی اهداف miRNA و نیز بررسی نتایج حاصل از القای miRNA ها بر روی مسیرهای سیگنالینگ و عملکرد سلول، با استفاده از روش‌های آزمایشگاهی یک فرآیند بسیار پرهزینه و وقت گیر است؛ بنابراین استفاده از روش‌های پیش‌بینی کننده محاسباتی قابل اعتماد که امکان تجزیه و تحلیل اهداف miRNA ها را در مقیاس وسیع فراهم کنند، الزامی است. تاکنون الگوریتم‌های زیادی بدین منظور طراحی شده‌اند، با این حال، هنوز هم جا برای بهبود این روش ها وجود دارد. از این رو، توسعه روش‌های محاسباتی جهت پیش‌بینی هر چه دقیق‌تر اهداف miRNA برای درک نقش miRNA ها در تنظیم ژن‌ها، مهم و اساسی است. به هر حال این روش‌ها اغلب منجر به تولید تعداد زیادی از پیش‌بینی‌های مثبت کاذب می‌شوند که ممکن است نمایانگر برهمکنش عملکردی miRNA با mRNA مورد نظر در یک بیماری خاص نباشد. در واقع، با توجه به ماهیت چند وجهی miRNA در هدف قرار دادن اهدافشان، پیش‌بینی‌های انجام شده توسط الگوریتم‌های پیش‌بینی‌کننده قبلی، قابل اعتماد برای هر شرایط پاتولوژیک نیستند [۱۹]. اگر چه تعدادی از مطالعات نشان داده‌اند که عملکرد miRNA ها مختص بافت است [۲۰] ولی تاکنون هیچ کدام از آن‌ها الگوریتمی برای پیش‌بینی اهداف miRNA برای یک بیماری خاص ارائه نکرده است. در این مطالعه ما یک رویکرد جدید برای پیش‌بینی اهداف miRNA های مختص سرطان سینه ارائه کردیم که براساس طبقه‌بند نائیوبیز است و از داده‌های آموزشی مختص سرطان استفاده کرده است. در مدل ارائه شده، استفاده از داده‌های آموزشی با کیفیت بالا که سایت‌های دقیق اتصال آن‌ها با روش‌های آزمایشگاهی تأیید شده، کارآمدی این مدل را در اجرا تضمین می‌کند. مطالعات قبل نشان داده است که استفاده یا عدم استفاده از یک گروه خاصی از ویژگی‌ها، عملکرد یک الگوریتم را به شدت تحت تأثیر قرار می‌دهد. برای مثال، Kim و همکاران گزارش کردند با حذف ویژگی‌های موقعیت از مطالعه، حساسیت الگوریتم miTarget کاهش می‌یابد [۲۱]. خوشبختانه، به دلیل وجود مقدار کافی و قابل دسترس از داده‌های مربوط به miRNA ها و اهداف آنان، در حال حاضر امکان تولید روش‌های محاسباتی که قادر به پیش‌بینی اهداف miRNA مختص بیماری باشد، وجود دارد. در این مقاله از مجموعه داده منفی، ارائه شده توسط Mitra و همکاران و Mendoza و همکاران که در چندین مطالعه دیگر نیز مورد استفاده قرار گرفته است، استفاده گردید [۲۲، ۱۵]. Bandyopadhyay و همکاران گزارش کردند که اهداف پیش‌بینی شده miRNA ها با استفاده از روش‌های محاسباتی فعلی در تمام بافت‌ها و یا بیماری‌ها دارای عملکرد نیستند. در واقع برخی از سایت‌های اتصال اهدافی که قبلاً تأیید شده‌اند، برای اتصال miRNA در بافت دیگر قابل دسترس نیستند چراکه آن‌ها در داخل ساختار ثانویه mRNA پنهان شده‌اند و یا توسط پروتئین‌های متصل شونده به RNA پوشیده شده‌اند [۲۳]. هدف از این مطالعه بررسی این موضوع بود که آیا با استفاده از داده‌های آموزشی اختصاصی سرطان سینه می‌توان به بهتر شدن

[Downloaded from jhbmi.com on 2024-10-12]

الگوریتم‌های قبلی غیر اختصاصی کمک کرد و اگر چنین است، کدام ویژگی‌ها برای سرطان سینه اختصاصی‌تر عمل می‌کنند. در بخش اول، نتایج نشان داد که در مقایسه با روش‌های قبلی، این روش با دقت بالاتری قادر به شناسایی اهداف miRNA هایی شد که قبلاً از طریق روش‌های آزمایشگاهی برهمکنش آن‌ها تأیید شده بود. با توجه به ویژگی‌های استفاده شده در این مطالعه، نتایج نشان داد که اگرچه ویژگی‌های ساختاری نقش مهمی را در حساسیت بالای مدل ایفا می‌کنند، ویژگی‌های توالی و ترکیب بازی بالاترین سهم را در ایجاد اختصاصیت مدل دارند. مطالعات قبلی نشان داده‌اند که محل‌های اتصال miRNA ها دارای ترکیبات نوکلئوتیدی خاصی هستند که به میزان قابل توجهی بین اهدافی که کاهش بیان داشته‌اند به دلیل ترانسفکشن miRNA و آن‌هایی که بیانشان پایدار باقی مانده‌اند، متفاوت است [۲۵، ۲۴]. موضوع دیگر در مورد پیش‌بینی اهداف miRNA های ویژه بافت، تأثیر ایزوفرم‌های 3'UTR جایگزین به دلیل سایت‌های پلی‌آدنیلاسیون جایگزین است. این می‌تواند منجر به از دست رفتن سایت‌های اتصال miRNA، از طریق کوتاه کردن ناحیه‌ی 3'UTR ژن‌های هدف شود [۲۶]. اگر چه، مطالعه حاضر نشان می‌دهد که miRNA های اختصاصی بافت اغلب در بیماری‌های مربوط به بافت خاص دخیل هستند، اما این که آیا ویژگی‌های اختصاصی بافت برای عملکرد miRNA وجود دارد، تا حد زیادی ناشناخته باقی مانده است و نیاز به بررسی‌های بیشتر و تأیید آزمایشگاهی دارد. به هر حال نتایج ارائه شده، نشان می‌دهد که روش طراحی شده در این مقاله توانایی پیش‌بینی اهداف miRNA های دخیل در سرطان سینه را دارد.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005; 55(2):74-108.

2. Casey MC, Sweeney KJ, Brown JA, Kerin MJ. Exploring circulating micro-RNA in the neoadjuvant treatment of breast cancer. Int J Cancer. 2016;139(1):12-22.

3. Madhavan D, Peng C, Wallwiener M, Zucknick M, Nees J, Schott S, et al. Circulating miRNAs with prognostic value in metastatic breast cancer and for early detection of metastasis. Carcinogenesis; 2016.

4. Shimomura A, Shiino S, Kawauchi J, Takizawa S, Sakamoto H, Matsuzaki J, et al. Novel combination of serum microRNA for detecting breast cancer in the early stage. Cancer Sci. 2016;107(3):326-34.

5. Costa FF. Non-coding RNAs: Meet thy masters. Bioessays. 2010; 32(7):599-608.

6. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 2004;116(2):281-97.

7. Kim VN, Nam JW. Genomics of microRNA. Trends Genet. 2006;22(3):165-73.

8. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524-9.

9. Almeida MI, Reis RM, Calin GA. MicroRNA history: discovery, recent applications, and next frontiers. Mutat Res. 2011; 717(1-2):1-8.

10. Jansson MD, Lund AH. MicroRNA and cancer. Mol Oncol. 2012; 6(6):590-610.

11. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005; 435(7043):834-8.

12. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006;6(4):259-69.

13. Zhang W, Dahlberg JE, Tam W. MicroRNAs in tumorigenesis: a primer. Am J Pathol. 2007;171(3):728-38.

14. Garzon R, Marcucci G, Croce CM. Targeting MicroRNAs in Cancer: Rationale, Strategies and Challenges. Nat Rev Drug Discov. 2010; 9(10):775-89.

15. Mitra R, Bandyopadhyay S. MultiMiTar: A Novel Multi Objective Optimization based miRNA-Target Prediction Method. PLoS ONE. 2011; 6(9):e24583.

16. Kononenko I. Estimating attributes: analysis and extensions of RELIEF. Machine Learning. 1994; 784: 171-82.

17. Frank E, Hall M, Trigg L, Holmes G, Witten IH: Data mining in bioinformatics using Weka. Bioinformatics. 2004;20(15):2479-81.

18. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, et al. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. Cancer Res. 2011;71(17):5635-45.

19. Bandyopadhyay S, Ghosh D, Mitra R, Zhao Z. MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets. Sci Rep. 2015;5:8004.

20. Li P, Hua X, Zhang Z, Li J, Wang J. Characterization of regulatory features of housekeeping and tissue-specific regulators within tissue regulatory networks. BMC Syst Biol. 2013;7:112.

21. Kim SK, Nam JW, Rhee JK, Lee WJ, Zhang BT. miTarget: microRNA target gene prediction using a support vector machine. BMC Bioinformatics. 2006; 7:1-11.

22. Mendoza MR, da Fonseca GC, Loss-Morais G, Alves R, Margis R, Bazzan AL. RFMirTarget: predicting human microRNA target genes with a random forest classifier. PLoS One. 2013;8(7):e70153.

23. Bandyopadhyay S, Ghosh D, Mitra R, Zhao Z. MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets. Sci Rep. 2015;5:8004.

[Downloaded from jhbmi.com on 2024-10-12]

24. Wang X, El Naqa IM. Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics. 2008; 24(3):325-32.
25. Xiao J, Li Y, Wang K, Wen Z, Li M, Zhang L, Guang X. In silico method for systematic analysis of feature importance in microRNA-mRNA interactions. BMC Bioinformatics. 2009;10:427
26. Nam JW, Rissland OS ,Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, et al. Global analyses of the effect of different cellular contexts on microRNA targeting. Mol Cell. 2014;53(6):1031-43.

[Downloaded from jhbmi.com on 2024-10-12]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2016;3(1): 18-28

Prediction of MicroRNAs (miRNAs) Targets in Breast Cancer Using Bioinformatics Methods

Sadeghi Balal[1]*

• **Received:** 5 Apr, 2016 • **Accepted:** 16 Feb, 2016

Introduction: Prediction of MicroRNAs (miRNAs) targets has a major importance. Development of calculating methods and also cost-effective and time-saving laboratory researches, has great effect on the production of therapeutic medicines like anti cancerous drugs. Since miRNAs have been recently identified, the researches trend was slow at first but by development of biologic databases and understanding its importance, scientists enhanced the speed of studies and paid more attention to this field. Until now, several computerized methods have been developed for prediction of microRNAs (miRNAs) targets but most of these methods have high false positive and further studies are required to improve these methods. Since, recent studies show that miRNAs have different targets in several tissues, this study aimed to provide a computerized method for prediction of miRNAs target in breast cancer.

Method: In this study, at first, all types of features were extracted, then, dominant features were selected via CFS and Relief methods. Smart models such as nervous network, Support Vector Machine with three different cores, Naïve Bayes algorithm and Random Forest decision tree using ten cross accreditation method were tested and its results compared and analyzed. In order to validate the results, gene expression profiling was used.

Results: Analyzing miRNA and gene expression profiles, the classifier predicted 124 functional interactions involving 21 miRNAs and 38 mRNAs in breast cancer.

Conclusion: In terms of bioinformatics, this approach was validated for breast cancer but for further validation, experimental methods also should be used.

Key words: MicroRNAs Targets Prediction, Breast Cancer, Support Vector Machine, Naïve Bayes

•**Citation:** Sadeghi B. Prediction of MicroRNAs (miRNAs) Targets in Breast Cancer Using Bioinformatics Methods. Journal of Health and Biomedical Informatics 2016; 3(1): 18-28.

1. Ph.D in Animal Genetics (Molecular Genetics and Bioinformatics), Assistant Professor, Food Hygiene and Public Health Dept., Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran

***Correspondence**: Food Hygiene and Public Health Dept., School of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran

• **Tel:** 034- 33222962 • **Email:** sadeghi.balal@uk.ac.ir

[Downloaded from jhbmi.com on 2024-10-12]