



مقاله پژوهشی

ارائه یک الگوریتم فراابتکاری جهت کشف توالی تنظیمی در ژنوم بیماران مبتلا به سرطان سینه

محمد جواد حسین پور^{۱،*}، حمید پروین^{۲،*}، صمد نجاتیان^{۳،۴}، وحیده رضایی^{۵،۶}، کرم الله باقری فرد^{۷،۸}

- ۱ - دانشکده مهندسی کامپیوتر، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
- ۲ - دانشکده مهندسی کامپیوتر، واحد نورآباد ممسنی، دانشگاه آزاد اسلامی، نورآباد ممسنی، ایران
- ۳ - دانشکده مهندسی برق، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
- ۴ - دانشکده ریاضیات، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
- ۵ - باشگاه پژوهشگران جوان و نخبگان، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
- ۶ - باشگاه پژوهشگران جوان و نخبگان، واحد نورآباد ممسنی، دانشگاه آزاد اسلامی، نورآباد ممسنی، ایران
- ۷ - باشگاه پژوهشگران جوان و نخبگان، واحد استهبان، دانشگاه آزاد اسلامی، استهبان، ایران

تاریخ پذیرش مقاله: ۱۳۹۸/۱۱/۱۲

تاریخ دریافت مقاله: ۱۳۹۸/۰۵/۲۵

چکیده

زمینه و هدف: امروزه در دانش پزشکی، روش‌های تشخیص علائم و کشف بیماری‌ها، بسیار وسیع و گسترده شده است. در نتیجه، تحلیل کلیه عوامل دخیل در یک بیماری، اغلب چالش برانگیز است؛ بنابراین به یک سیستم مکانیزه برای کمک به کشف قوانین، شناسایی الگوهای موجود و پیش‌بینی رخ داده‌های آینده، کاملاً احساس می‌شود. در این تحقیق قصد داریم با استفاده از یک الگوریتم فراابتکاری چندهدفه، روشی را ارائه کنیم که بتواند توالی‌های تنظیمی (Regulatory Elements) با طول متغیر از ژنوم را کشف و استخراج کرده و برهم‌کنش‌های بین آن‌ها را نیز شمارش کند. درواقع این توالی تنظیمی می‌تواند نقش قابل توجهی در بروز و تشدید بیماری سرطان ایفا کنند.

مواد و روش‌ها: در این پژوهش، از یک متد پیشنهادی جهت کشف توالی تنظیمی در ژنوم بیمار مبتلا به سرطان سینه استفاده شده است. متد پیشنهادی در نرم‌افزار Matlab پیاده‌سازی شده است. همچنین، جهت سنجش عملکرد و کارایی روش پیشنهادی، الگوریتم موردنظر بر روی مجموعه داده‌های Hi-C (Hi-C Dataset) مربوط به بیماران مبتلا به سرطان سینه در دو رده سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ معرفی شده توسط میفسود و همکاران، اجرا گردید.

نتایج: اجرای روش پیشنهادی با روش HiCUP مورد مقایسه قرار گرفت. نتایج نشان می‌دهند که روش MSARE در کشف توالی‌های تنظیمی، عملکرد بهتری را نسبت به روش HiCUP داشته است.

نتیجه‌گیری: در این پژوهش، بررسی‌های تجربی نشان می‌دهند که دو پروموتور BLC۶ و HOTTIP کشف شده توسط روش پیشنهادی در هر دو سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ در بروز و تشدید بیماری سرطان سینه تأثیر معنی‌داری داشته‌اند.

کلمات کلیدی: توالی تنظیمی، الگوریتم فراابتکاری، پروموتور BLC۶، پروموتور HOTTIP

مقدمه

تنظیم ژن (gene regulation) فعالیتی است که منجر به بیان ژن می‌شود و یا برعکس می‌تواند بیان ژن را متوقف کند. عواملی مانند (پروموتور، تشدیدکننده) می‌توانند بر روی عملیات تنظیم ژن تأثیر داشته باشند، به این عوامل توالی‌های تنظیمی گفته می‌شود. در صورتی که این توالی‌های تنظیمی شناسایی شوند، آنگاه عملاً می‌توان تنظیم بیان ژن را مشخص و تعیین نمود. امروزه ابزارها و تکنولوژی‌هایی به وجود آمده‌اند که می‌توانند توالی‌های تنظیمی را شناسایی کنند. همچنین این ابزارها توانایی آن را دارند که برهم‌کنش‌های مختلف بین این توالی‌های تنظیمی را نیز شناسایی کنند (۱). Hi-C یکی از این تکنولوژی‌ها است. این تکنولوژی، به جهت کشف برهم‌کنش‌های بیشتر در توالی‌های

تنظیم ژن (gene regulation) فعالیتی است که منجر به بیان ژن می‌شود و یا برعکس می‌تواند بیان ژن را متوقف کند. عواملی مانند (پروموتور، تشدیدکننده) می‌توانند بر روی عملیات تنظیم ژن تأثیر داشته باشند، به این عوامل توالی‌های تنظیمی گفته می‌شود. در صورتی که این توالی‌های تنظیمی شناسایی شوند، آنگاه عملاً می‌توان تنظیم بیان ژن را مشخص و تعیین نمود. امروزه ابزارها و تکنولوژی‌هایی به وجود آمده‌اند که می‌توانند توالی‌های تنظیمی را شناسایی کنند. همچنین این ابزارها توانایی آن را دارند که برهم‌کنش‌های مختلف بین این توالی‌های تنظیمی را نیز شناسایی کنند (۱). Hi-C یکی از این تکنولوژی‌ها است. این تکنولوژی، به جهت کشف برهم‌کنش‌های بیشتر در توالی‌های

*نویسنده مسئول: حمید پروین، دانشکده مهندسی کامپیوتر، واحد نورآباد ممسنی، دانشگاه آزاد اسلامی، نورآباد ممسنی، ایران
Email: parvin@iust.ac.ir
https://orcid.org/0000-0002-4877-182X

HiC-bench همانند HiC-Pro می‌توانست به‌صورت موازی بر روی چند سیستم اجرا شود. ابزار پیشنهادی آن‌ها، قادر بود، توالی‌های تنظیمی و برهم‌کنش‌های بین آن‌ها را در مدت‌زمان محدودی کشف و استخراج کند (۱۲).

مطابق تحقیقات انجام‌شده در سال‌های گذشته، تکنولوژی Hi-C و ابزارهای آن می‌توانند داده‌های Hi-C را به‌عنوان ورودی گرفته و تعداد برهم‌کنش‌های بین توالی‌های تنظیمی موجود در هر کروموزوم را شناسایی کند (۱۳)؛ اما این تکنولوژی و ابزارهای مورد‌استفاده در آن، تنها می‌توانند با قطعات (Fragments) با طول ثابت (Fix-bin size) از ژنوم (که تشکیل توالی‌های تنظیمی را می‌دهند) کشف و استخراج کرده و برهم‌کنش‌های بین آن‌ها را شمارش کنند. در این صورت، خروجی این ابزارها که تنها قادر هستند، با فرگمنت‌های به طول ثابت کار کنند، نمی‌تواند بهینه باشد. درواقع، دلیل بهینه نبودن این ابزارها ثابت نبودن طول توالی‌های تنظیمی و متغیر بودن طول آن‌ها است (۱۴، ۱۵). حال، در این تحقیق قصد داریم با ارائه یک الگوریتم چندهدفه مبتنی بر شبیه‌سازی تبرید روشی را بیان کنیم که بتواند فرگمنت‌های با طول متغیر از ژنوم که تشکیل توالی‌های تنظیمی با طول متغیر را می‌دهند کشف و استخراج کرده و برهم‌کنش‌های بین آن‌ها را شمارش کنیم. درواقع این توالی‌های تنظیمی می‌توانند پروموتور/تشدیدکننده‌های بالقوه‌ای باشند که نقش قابل‌توجهی در بروز و تشدید بیماری سرطان را ایفا کنند.

مواد و روش‌ها

در این مقاله، قصد داریم با استفاده از الگوریتمی چندهدفه مبتنی بر شبیه‌سازی تبرید توالی‌های تنظیمی با طول متغیر را در ژنوم بیماران سرطانی کشف و استخراج کنیم. توالی‌های تنظیمی کشف‌شده می‌توانند پروموتور/تشدیدکننده‌های بالقوه‌ای باشند که در بیماری سرطان مؤثر بوده‌اند. جهت سنجش عملکرد و کارایی الگوریتم پیشنهادی، الگوریتم موردنظر بر روی مجموعه داده‌های Hi-C معرفی‌شده توسط میفسود و همکاران مربوط به سلول‌های خونی GM12878 و CD34⁺ در چهار طول فرگمنت (Fragment size) ۱۰Kbp، ۱۰۰Kbp، ۵۰۰Kbp و ۱Mbp، اجرا شده است. در ادامه ابتدا تابع تناسب مورد‌استفاده در الگوریتم پیشنهادی توضیح داده می‌شود. سپس نحوه اجرای الگوریتم پیشنهادی و نتایج اجرای آن بر روی مجموعه داده Hi-C مربوط به این دو نوع سلول خونی توضیح داده خواهد شد.

تنظیمی به وجود آمد (۲). این تکنولوژی برای اولین بار توسط آیدین و همکاران در سال ۲۰۰۹ جهت شبیه‌سازی همه برهم‌کنش‌های کروموزومی در یک آزمایش خاص مطرح گردید (۳). در حقیقت، تکنولوژی Hi-C به همراه ابزارهای آن، برهم‌کنش‌های بین نقاط مختلف ژنوم را در یک ساختار سه‌بعدی ارائه می‌دهد. این نقاط، می‌توانند توالی‌های تنظیمی باشند. بدین معنی که از نظر ساختاری کدام توالی از ژنوم با کدام توالی دیگر از ژنوم برهم‌کنش دارد (۴). این اطلاعات می‌توانند دانش زیادی را در زمینه ساختار برهم‌کنش ژنوم ارائه کند و بسیار گران‌قیمت نیز می‌باشند. امروزه از تکنولوژی Hi-C، برای شناسایی توالی‌های منظم، خصوصاً شناسایی پروموتور/تشدیدکننده‌ها استفاده شده است (۵). مقالات (۶-۸) با استفاده از این تکنولوژی توانسته‌اند پروموتور/تشدیدکننده‌های جدید را کشف کنند، که با یکدیگر برهم‌کنش داشته و بیان ژن را تنظیم می‌کنند.

در سال‌های گذشته، پژوهش‌های زیادی در این زمینه صورت گرفته است که در ادامه به چند مورد از آن‌ها اشاره می‌شود. کابروس و همکاران در سال ۲۰۱۵، الگوریتمی جهت تشخیص اجماعی بر روی داده‌های Hi-C در ساختار DNA موش و انسان بکار بردند. الگوریتم پیشنهادی آن‌ها قادر به کشف اجماع‌ها در DNA به‌صورت پویا بود (۹). میفسود و همکاران در سال ۲۰۱۵، در مقاله‌ای به کشف و استخراج پروموتورها در سلول‌های انسانی با استفاده از تکنولوژی HiCUP در رزولوشن بالا پرداختند. آن‌ها در این پژوهش، از ضبط Hi-C (Chi-C) جهت کشف برهم‌کنش‌های موجود بین توالی ژنوم در ۲ نوع از سلول‌های خونی انسان استفاده کردند (۱۰). سروانت و همکاران در سال ۲۰۱۵، در مقاله‌ای ابزار HiC-Pro را جهت کشف توالی تنظیمی و برهم‌کنش‌های بین آن‌ها در کروموزوم‌های انسانی ارائه دادند. ابزار پیشنهادی آن‌ها می‌توانست برهم‌کنش‌های یک توالی خاص از ژنوم را با دیگر توالی از ژنوم به دست آورد. همچنین HiC-Pro می‌توانست به‌صورت موازی اجرا شود؛ بنابراین سرعت تحلیل و آنالیز مجموعه داده‌های Hi-C در آن بالا بود (۱۱). چارلامبوس و همکاران در سال ۲۰۱۷، در مقاله‌ای، ابزاری بنام HiC-bench را جهت تجزیه و تحلیل مجموعه داده‌های Hi-C ارائه کردند. این ابزار، علاوه بر آنکه قادر به کشف توالی تنظیمی شامل پروموتور/تشدیدکننده بود، می‌توانست، برهم‌کنش بین آن‌ها را نیز کشف و ضبط کند. کارایی عملکرد این ابزار در برخورد با مجموعه داده‌های Hi-C، همانند HiCUP و HiC-Pro است. همچنین

تابع تناسب

تابع تناسب برای دو توالی کشف شده به عنوان پروموتر و تشدیدکننده بالقوه در یک کروموزوم، مطابق با فرمول ۱، برابر با جمع تمامی برهم کنش های موجود بین این دو توالی تقسیم بر مجموع طول هر دو توالی پروموتر و تشدیدکننده است. مقدار تابع تناسب عددی در بازه [۰،۱] است؛ و هرچقدر این مقدار بیشتر باشد نشان می دهد که این دو توالی با یکدیگر برهم کنش بالاتری داشته و مقدار تابع تناسب آن ها بهتر است.

$$\text{مجموع برهم کنش های بین دو توالی کشف شده} \\ \text{مجموع طول هر دو توالی} = \text{تابع تناسب}$$

الگوریتم موردنظر یک الگوریتم چند هدفه مبتنی بر SA به نام MSARE (Multi-objective Simulation Annealing) جهت کشف و استخراج توالی های منظم، در مجموعه داده های Hi-C است. الگوریتم MSARE علاوه بر کشف و استخراج این عناصر قادر به شمارش تمامی برهم کنش های بین آن ها نیز است. ورودی های الگوریتم مجموعه داده Hi-C شامل اطلاعات برهم کنش های بین ژنومی در هر کروموزوم، بیشینه تکرار (Maxitr) و بیشینه زیر تکرار (Itr) است. در مرحله پیش پردازش عملیات توالی سازی (Sequencing) و حذف برهم کنش های برون کروموزومی بر روی مجموعه داده ورودی انجام می گیرد. سپس به اندازه بیشینه تکرار، جمعیت اولیه به صورت تصادفی جهت کشف توالی های تنظیمی ایجاد می شود. جمعیت ورودی موردنظر، بر اساس مشخص کردن ابتدا و انتهای پروموتر و ابتدا و انتهای تشدیدکننده تنظیم می شود. در مرحله بعد با استفاده از تابع تناسب هر یک از جمعیت ورودی مورد ارزیابی قرار می گیرد. پس از محاسبه تابع تناسب برای هر عضو، جمعیت ورودی موردنظر جهت انجام پردازش بعدی وارد فاز الگوریتم شبه سازی تبرید می شود. در این فاز، به اندازه بیشینه زیر تکرار، ۲۰ همسایه پیرامون بهترین عضو از جمعیت ورودی مرحله قبل ایجاد می شود. سپس هر یک از همسایه های ایجاد شده، با استفاده از فرمول ۱ مورد ارزیابی قرار می گیرند. سپس، بهترین پاسخ یافته شده در هر تکرار که دارای بهترین مقدار تابع ارزیابی است، ذخیره می شود. بهترین پاسخ در هر تکرار، نشان دهنده توالی های تنظیمی بهینه کشف شده در مجموعه داده ورودی است. در مرحله آخر تمامی توالی های تنظیمی کشف شده در مجموعه داده به عنوان خروجی نمایش داده می شود.

پارامترها

در الگوریتم پیشنهادی، بیشینه تکرار برابر با ۱۰۰۰، تعداد زیر تکرار برابر با ۲۰، تعداد جمعیت اولیه ۲۵ و تعداد همسایگان قابل کشف در نزدیکی جواب بهینه ۲۰ در نظر گرفته شده است. همچنین در الگوریتم شبیه سازی تبرید نیز دمای اولیه برابر با ۱۰ و میزان کاهش دما ۰/۹۹ تنظیم شده است. برای هر عضو از جمعیت ورودی با استفاده از فرمول ۱، مقدار تابع تناسب آن محاسبه می شود. سپس از بین جمعیت ورودی، عضوی که نشان دهنده پروموتر/تشدیدکننده بالقوه با بهترین مقدار تابع تناسب است به عنوان بهترین پاسخ فعلی در نظر گرفته می شود. حال به اندازه بیشینه تکرار، مراحل زیر تکرار می گردد. در نزدیکی پاسخ بهینه موجود به اندازه ۲۰ همسایه، جمعیت جدید تولید می شود، سپس مقدار تابع تناسب هر عضو از آن جمعیتی محاسبه می گردد. در مرحله بعد، مقدار تابع تناسب هر عضو از جمعیت همسایه با بهترین پاسخ حاصله مقایسه می شود. حال اگر پاسخ جدید از بهترین پاسخ یافته شده بهتر بود، پاسخ جدید را به عنوان بهترین پاسخ در نظر گرفته می شود، اما اگر بهتر نبود، آن را با درصد احتمالی به صورت پاسخ مشروط در نظر می گیرد. سپس الگوریتم موردنظر، بهترین پاسخ یافته شده را به روزرسانی کرده و کاهش دما نیز به عنوان آخرین مرحله انجام می گیرد. گردش کار موردنظر، به اندازه تعداد تکرار ادامه می کند و در هر تکرار پاسخ بهینه (شامل بهترین توالی پروموتر/تشدیدکننده بالقوه کشف شده)، ذخیره می شود.

نتایج

الگوریتم پیشنهادی در محیط Matlab 2017a پیاده سازی شده است. اجرای الگوریتم پیشنهادی جهت حصول نتایج، بر روی سرور با مشخصات (3 CPU Cor i7, 16 G RAM, 2 T H.D.D, Winodws Server 2016) انجام گردیده است. الگوریتم پیشنهادی از نظر عملکرد بر اساس مقدار تابع تناسب، همچنین کشف توالی های پروموتر/انهسر بالقوه و شمارش برهم کنش های بین آن ها با روش استفاده شده در میفسود (HiCUP) مقایسه شده است. در ادامه، تجزیه و تحلیل هر یک از نتایج حاصله در قالب جداول و نمودارها توضیح داده می شود. در جدول ۱، میانگین مقدار تابع تناسب مربوط به هر کروموزوم در مجموعه داده های Hi-C ورودی با فرگمنت های به طول

جدول ۱- میانگین مقدار تابع تناسب در هر کروموزوم در فرگمنت‌های مختلف مربوط به روش پیشنهادی و روش HiCUP

شماره کروموزوم	Kbp۱۰		Kbp۱۰۰		Kbp۵۰۰		Mbp۱	
	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP
۱	۰۶E-۲۰/۳	۰۷E-۷۲/۱	۰۷E-۳۷/۹	۰۹E-۴۵/۴	۰۷E-۷۴/۲	۰۹E-۴۵/۴	۰۸E-۴۸/۱	۰۹E-۳۳/۴
۲	۰۶E-۳۳/۲	۰۷E-۷۳/۱	۰۶E-۰۶/۱	۰۹E-۲۱/۴	۰۷E-۷۷/۲	۰۹E-۲۱/۴	۰۸E-۵۵/۱	۰۹E-۱۲/۴
۳	۰۶E-۰۰/۳	۰۷E-۸۵/۱	۰۶E-۰۷/۱	۰۹E-۱۵/۵	۰۷E-۶۶/۲	۰۹E-۱۵/۵	۰۸E-۴۹/۱	۰۹E-۰۸/۵
۴	۰۶E-۲۵/۳	۰۷E-۹۳/۱	۰۶E-۰۴/۱	۰۹E-۴۰/۵	۰۷E-۹۰/۲	۰۹E-۴۰/۵	۰۸E-۶۳/۱	۰۹E-۲۷/۵
۵	۰۶E-۶۵/۳	۰۷E-۱۶/۲	۰۶E-۰۴/۱	۰۹E-۶۹/۵	۰۷E-۷۳/۲	۰۹E-۶۹/۵	۰۸E-۷۷/۱	۰۹E-۵۶/۵
۶	۰۶E-۶۱/۳	۰۷E-۴۰/۲	۰۶E-۱۳/۱	۰۹E-۹۶/۵	۰۷E-۹۰/۲	۰۹E-۹۶/۵	۰۸E-۷۵/۱	۰۹E-۹۰/۵
۷	۰۶E-۵۴/۳	۰۷E-۷۸/۲	۰۶E-۱۷/۱	۰۹E-۴۵/۶	۰۷E-۷۴/۲	۰۹E-۴۵/۶	۰۸E-۴۸/۱	۰۹E-۳۱/۶
۸	۰۶E-۰۹/۴	۰۷E-۸۴/۲	۰۶E-۱۸/۱	۰۹E-۹۸/۶	۰۷E-۹۱/۲	۰۹E-۹۸/۶	۰۸E-۴۳/۱	۰۹E-۸۵/۶
۹	۰۶E-۵۴/۴	۰۷E-۳۸/۳	۰۶E-۰۱/۱	۰۹E-۴۳/۸	۰۷E-۸۲/۲	۰۹E-۴۳/۸	۰۸E-۸۹/۱	۰۹E-۰۰/۸
۱۰	۰۶E-۵۸/۴	۰۷E-۶۰/۳	۰۶E-۱۲/۱	۰۹E-۵۴/۷	۰۷E-۸۵/۲	۰۹E-۵۴/۷	۰۸E-۴۸/۱	۰۹E-۴۳/۷
۱۱	۰۶E-۸۲/۴	۰۷E-۷۶/۳	۰۶E-۱۶/۱	۰۹E-۵۵/۷	۰۷E-۶۴/۲	۰۹E-۵۵/۷	۰۸E-۴۲/۱	۰۹E-۴۶/۷
۱۲	۰۶E-۴۷/۴	۰۷E-۱۳/۴	۰۶E-۰۱/۱	۰۹E-۶۱/۷	۰۷E-۰۱/۳	۰۹E-۶۱/۷	۰۸E-۵۹/۱	۰۹E-۵۳/۷
۱۳	۰۶E-۴۰/۴	۰۷E-۳۴/۴	۰۶E-۱۷/۱	۰۸E-۰۳/۱	۰۷E-۸۶/۲	۰۸E-۰۳/۱	۰۸E-۵۷/۱	۰۸E-۰۲/۱
۱۴	۰۶E-۹۴/۴	۰۷E-۱۸/۵	۰۶E-۰۲/۱	۰۸E-۱۳/۱	۰۷E-۹۸/۲	۰۸E-۱۳/۱	۰۸E-۵۳/۱	۰۸E-۱۱/۱
۱۵	۰۶E-۴۸/۵	۰۷E-۳۹/۶	۰۶E-۰۷/۱	۰۸E-۲۱/۱	۰۷E-۸۹/۲	۰۸E-۲۱/۱	۰۸E-۴۴/۱	۰۸E-۱۹/۱
۱۶	۰۶E-۴۷/۵	۰۷E-۱۴/۷	۰۶E-۱۰/۱	۰۸E-۲۵/۱	۰۷E-۶۹/۲	۰۸E-۲۵/۱	۰۸E-۶۲/۱	۰۸E-۲۲/۱
۱۷	۰۶E-۷۴/۵	۰۷E-۸۸/۸	۰۶E-۱۵/۱	۰۸E-۲۶/۱	۰۷E-۸۷/۲	۰۸E-۲۶/۱	۰۸E-۸۳/۱	۰۸E-۲۳/۱
۱۸	۰۶E-۹۹/۵	۰۷E-۸۲/۶	۰۷E-۹۳/۹	۰۸E-۳۳/۱	۰۷E-۶۸/۲	۰۸E-۳۳/۱	۰۸E-۴۱/۱	۰۸E-۲۹/۱
۱۹	۰۶E-۱۸/۶	۰۶E-۱۵/۱	۰۶E-۱۰/۱	۰۸E-۷۴/۱	۰۷E-۹۴/۲	۰۸E-۷۴/۱	۰۸E-۰۴/۲	۰۸E-۶۹/۱
۲۰	۰۶E-۸۱/۶	۰۷E-۷۳/۹	۰۶E-۰۸/۱	۰۸E-۶۴/۱	۰۷E-۰۳/۳	۰۸E-۶۴/۱	۰۸E-۶۷/۱	۰۸E-۶۱/۱
۲۱	۰۶E-۱۸/۷	۰۷E-۳۳/۹	۰۶E-۳۰/۱	۰۸E-۶۹/۲	۰۷E-۲۲/۳	۰۸E-۶۹/۲	۰۸E-۸۱/۱	۰۸E-۵۶/۲
۲۲	۰۶E-۱۶/۷	۰۶E-۷۳/۱	۰۶E-۱۹/۱	۰۸E-۷۹/۲	۰۷E-۰۰/۳	۰۸E-۷۹/۲	۰۸E-۴۳/۱	۰۸E-۷۰/۲
X	۰۶E-۳۸/۷	۰۷E-۸۵/۸	۰۶E-۰۵/۱	۰۹E-۷۷/۶	۰۷E-۴۹/۲	۰۹E-۷۷/۶	۰۸E-۲۷/۱	۰۹E-۶۶/۶
Y	۰۵E-۶۷/۱	۰۵E-۹۹/۲	۰۶E-۴۳/۳	۰۷E-۱۱/۶	۰۷E-۸۱/۴	۰۷E-۱۱/۶	۰۸E-۱۰/۲	۰۷E-۶۸/۲
M	۰۵E-۸۹/۴	۰۵E-۳۳/۳	۰۶E-۰۰/۵	۰۶E-۰۰/۱	۰۶E-۰۰/۱	۰۶E-۰۰/۱	۰۷E-۰۰/۵	۰۷E-۰۰/۵

تابع تناسب الگوریتم پیشنهادی در هر کروموزوم نسبت به روش HiCUP بهتر است؛ بنابراین میانگین مقدار تابع تناسب در روش پیشنهادی نسبت به روش HiCUP که پروموتور/تشدیدکننده‌های بالقوه با طول ثابت را استخراج می‌کند، بهتر بوده است. در حقیقت

۱۰Kbp، ۱۰۰Kbp، ۵۰۰Kbp و ۱Mbp به تصویر کشیده شده است. همان‌طور که مشاهده می‌کنید، به دلیل اینکه، روش پیشنهادی قادر به کشف پروموتور/تشدیدکننده‌های بالقوه با طول‌های متغیر است؛ بنابراین، مطابق با جدول زیر، میانگین مقدار



جدول ۲- میانگین تعداد برهم کنش های شمارش شده بین توالی پروموتور/تشدیدکننده بالقوه کشف شده در هر کروموزوم در فرگمنت های مختلف مربوط به روش پیشنهادی و روش HiCUP

شماره کروموزوم	Kbp۱۰	Kbp۱۰۰		Kbp۵۰۰		Mbp۱		
	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP	روش پیشنهادی	HiCUP
۱	۱۸	۳۴۸۹۹۲۷۲۱/۱	۵۶۵۲۱۷۴/۷۸۶	۰۳۵۴۵۵۸۶۵/۵	۲۶۱۹۰۵/۲۷۹۶	۳۶۳۶۶۷۳۷/۵۴	۸۸۲۳۵/۵۳۳۷	۰۳۸۷۴۳۹/۲۱۰
۲	۴۲/۱۴	۳۲۰۰۶۱۶۳۴/۱	۰۴۱۶۶۶۷/۶۳۹	۳۰۰۸۰۳۴۳۹/۴	۲۸۹۴۷۴/۲۸۱۳	۷۷۳۶۲۴۸۱/۴۳	۶۵/۵۱۴۱	۹۶۹۲۲۹۲/۱۶۹
۳	۷۱/۱۷	۳۴۸۰۳۵۹۳۸/۱	۳۶/۷۶۹	۵۵۰۶۴۳۰۳۹/۴	۳۷۹۳۱/۲۶۲۱	۹۸۹۵۷۳۳۲/۵۳	۲۲۲۲۲۲/۵۲۸۱	۴۲۰۶۹۸۲/۲۱۲
۴	۹۲/۱۵	۳۱۷۴۰۲۲۲۲/۱	۳۴۴	۲۵۰۴۱۵۱۹۳/۴	۷۱۴۲۸۶/۲۴۵۴	۷۲۱۵۳۴/۴۷	۰۶۶۶۶۷/۵۰۰۷	۴۰۵۰۷۳۸/۱۸۵
۵	۲۵/۲۲	۳۵۱۷۳۹۶۵۳/۱	۴/۵۹۱	۹۱۳۰۸۰۷۸۷/۴	۲۷۵۸۶۲/۲۶۰۷	۷۰۸۵۰۴۱/۵۶	۹۴۷۳۶۸/۵۵۳۴	۹۷۵۷۰۶۷/۲۲۱
۶	۴۶/۲۲	۳۶۵۷۴۰۰۹۱/۱	۲۸۵۷۱۴/۶۳۵	۹۹۳۱۲۶۷۵/۴	۱۷۱۴۲۹/۲۵۱۵	۷۲۵۰۱۷۲۱/۶۱	۱۱۱۱۱۱/۵۵۵۶	۹۹۳۷۴۷۸/۲۴۲
۷	۳۱۳۰۴۳۴۸/۲۳	۳۷۵۶۵۹۵۲۵/۱	۹۳۳۳۳۳۳/۴۷۱	۹۰۷۵۴۳۷۵۹/۴	۲/۲۴۰۵	۶۱۸۰۰۸۴۷/۵۷	۹۰۹۰۹۱/۵۲۳۲	۸۷۵۲۸۹۴/۲۲۳
۸	۸۲۶۹۲۳۰۸/۲۰	۳۶۰۶۱۸۰۱۴/۱	۵/۴۷۱	۰۷۵۸۲۲۱۴۲/۵	۱۷۸۵۷۱/۲۳۳۵	۲۵۱۰۳/۶۶	۱۵۳۸۴۶/۳۸۸۱	۳۰۵۷۷۳۴/۲۵۸
۹	۸/۲۶	۴۱۹۷۹۳۷۵۲/۱	۳۵/۵۶۸	۹۴۵۸۸۱۷۳۹/۵	۱۹۳۵۴۸/۲۴۴۱	۹۰۳۸۸۱۴۵/۶۸	۲۶۳۱۵۸/۶۶۴۴	۵۷۶۱۸۴۴/۲۵۹
۱۰	۶۵۵۵۵۵۵۶/۲۸	۴۰۳۹۶۱۴۴۵/۱	۰۶۲۵/۵۱۶	۴۳۷۱۰۸۱۵۲/۵	۱۱۵۳۸۵/۲۴۶۷	۰۲۲۸۸۸۲۷/۷۰	۷۵/۳۸۴۹	۹۷۷۳۹۱۱/۲۷۴
۱۱	۶۶۹۳۵۴۸۴/۲۳	۴۰۷۲۷۷۸۵۲/۱	۶۶۶۶۶۶۷/۷۱۳	۷۴۴۷۰۵۳۴/۶	۶۸۱۸۱۸/۳۱۵۷	۶۵۷۴۴۵۶۶/۹۱	۳/۴۰۳۲	۰۴۴۵۵۱۷/۳۶۱
۱۲	۳۴۹۲۰۶۳۵/۱۲	۴۱۸۴۱۵۵۶۳/۱	۷۵۱	۱۰۹۴۶۱۶۴۱/۶	۷۶۶۶۶۷/۲۳۸۸	۶۶۷۲۳۵۸۹/۸۴	۹۳۷۵/۵۱۲۵	۵۵۲۲۲۳۵/۳۳۳
۱۳	۷۳۹۱۳۰۴۳/۱۰	۳۸۴۶۱۹۲۱۶/۱	۲۲۲۲۲۲۲/۴۳۰	۳۵۸۹۹۴۸۵۵/۵	۵/۱۷۰۵	۵۴۰۳۴۸۴۴/۷۹	۲۵/۴۱۳۶	۱۴۰۹۶۳۶/۳۱۳
۱۴	۸۰۸۲۱۹۱۸/۱۱	۴۴۹۲۵۲۲۲۳/۱	۶/۵۸۲	۳۲۳۰۹۴۰۱۸/۷	۸۵۷۱۴۳/۲۹۵۹	۳۵۷۸۶۳۲/۱۱۷	۳۸۴۶۱۵/۵۸۴۶	۲۳۱۵۱۸۵۰۴۵۹
۱۵	۶۵۹۳۴۰۶۶/۱۰	۵۰۷۹۲۵۸۰۱/۱	۴۲۱۰۵۲۶/۶۲۱	۴۴۲۰۴۵۳۷۱/۷	۱۹۴۴۴۴/۲۴۶۳	۴۳۱۱۴۰۴/۱۱۰	۷۲۷۲۷۳/۴۱۰۷	۳۶۱۴۴۵۸/۴۳۳
۱۶	۹۵۹۵۹۵۹۶/۹	۴۵۲۴۱۲۹۸۶/۱	۶۵/۵۴۳	۲۵۴۳۹۹۲۴۶/۷	۶۸/۲۰۶۱	۲۷۱۹۵۸۷/۱۰۱	۴/۴۴۷۸	۳۵۶۸۲/۳۸۹
۱۷	۷۸۰۴۸۷۸۰۵/۹	۵۴۹۷۵۹۱۳۳/۱	۴۵/۷۶۸	۸۳۸۶۵۷۶۷۲/۸	۳۲۵/۲۸۴۶	۳۰۹۸۲۴۳/۱۳۵	۲۱۴۲۸۶/۵۲۹۱	۸۶۳۸۸۸۹/۵۲۲
۱۸	۳۴۲۶۵۷۳۴/۱۱	۴۳۹۶۶۰۱۷۸/۱	۳۰۶	۹۲۵۰۴۲۸۹۴/۵	۷۰۸۳۳۳/۱۴۴۶	۸۲۱۱۰۲۵۳/۹۸	۰۹۰۹۰۹/۲۸۵۸	۴۱۰۰۸۴/۳۷۸
۱۹	۵۸۳۳۳۳۳۳۳/۸	۴۶۴۳۹۳۶۱۶/۱	۲۲۲۲۲۲۲/۶۷۰	۶۴۵۷۶۸۱۱/۱۰	۴۳۴۷۸۳/۳۴۰۰	۷۹۰۴۹۵۱/۱۷۹	۳۳۳۳۳۳/۷۸۶۵	۶۳۸۲۲۳۳/۶۸۷
۲۰	۷۶۰۶۳۸۳/۱۰	۵۱۱۳۱۶۱۸۳/۱	۵۳۸۴۶۱۵/۵۶۱	۶۶۵۳۴۰۵۹۶/۸	۵/۲۲۱۵	۴۸۵۷۳۷۶/۱۵۱	۷۵/۴۹۰۷	۷۶۰۹۷۳/۵۸۹



۲۱	۵۴۱۶۶۶۶۶۷/۹	۴۸۱۲۶۷۷۳۷/۱	۴۶۶۶۶۶۷/۵۵۱	۵۷۶۱۱۴۱۳/۱۲	۳۵۲۹۴۱/۲۵۴۷	۹۶۶۱۳۲۶/۲۱۷	۸۱۸۱۸۲/۶۴۱۳	۷۳۹۵۴۱۲/۸۰۷
۲۲	۷۹۰۱۲۳۴۵۷/۶	۵۸۰۲۰۱۵۴۵/۱	۳۵۲۹۴۱۲/۴۶۶	۲۲۸۳۹۵۰۶/۱۲	۵/۱۸۹۹	۸۱۳۲۸۶۲/۲۱۸	۲۵/۳۲۳۹	۸۲۷۳۲۷۳/۸۳۵
X	۰۸۱۸۷۱۳۵/۱۰	۳۰۲۹۴۷۵۴۴/۱	۱۱۱۱۱۱۱/۲۷۵	۴۰۶۴۳۰۶۳۱/۴	۵۳۸۴۶۲/۱۵۲۱	۵۰۹۸۱۸۸/۵۶	۶۶۶۶۷/۳۱۸۹	۴۶۵۸۵۳۷/۲۲۰
Y	۳	۶۳۸۸۸۸۸۸۹/۲	۲۴	۹۸۶۱۱۱۱۱/۴	۵	۴۸۱۴۸۱۴۸/۶	۷۵/۳	۷۵۶۰۹۷۵۶۱/۸
M	۳۵۸	۳۳۳۳۳۳۳/۱۴۸	۵۸۸۳	۵۸۸۳	۵۸۸۳	۵۸۸۳	۵۸۸۳	۵۸۸۳

جدول ۳- تعداد توالی‌های پروموتور/تشدیدکننده بالقوه کشف‌شده توسط روش پیشنهادی به تفکیک هر کروموزوم در فرگمنت‌های مختلف

شماره کروموزوم	۱۰Kbp	۱۰۰Kbp	۵۰۰Kbp	۱Mbp
۱	۰	۱۸	۳	۰
۲	۰	۱۰	۵	۰
۳	۰	۱۲	۵	۰
۴	۰	۶	۶	۳
۵	۱	۱۰	۲	۱
۶	۱	۸	۸	۱
۷	۰	۳	۲	۰
۸	۴	۸	۴	۲
۹	۱	۸	۲	۱
۱۰	۳	۹	۶	۰
۱۱	۶	۷	۳	۱
۱۲	۳	۱۲	۳	۲
۱۳	۵	۴	۳	۲
۱۴	۷	۸	۲	۲
۱۵	۱۴	۱۰	۳	۱
۱۶	۲۱	۱۳	۲	۰
۱۷	۲۳	۱۲	۳	۱
۱۸	۲۴	۷	۵	۱
۱۹	۱۷	۹	۳	۱
۲۰	۴۳	۶	۲	۱
۲۱	۲۶	۸	۱	۱
۲۲	۲۷	۹	۴	۱
X	۷۹	۲	۲	۱
Y	۰	۰	۰	۰
M	۰	۰	۰	۰

معنی است که تعداد برهم‌کنش‌های بین پروموتور/تشدیدکننده بالقوه استخراج‌شده با طول متغیر بیشتر از حالت با طول ثابت باشد. همچنین مطابق با جدول ۱، میانگین مقدار تابع تناسب

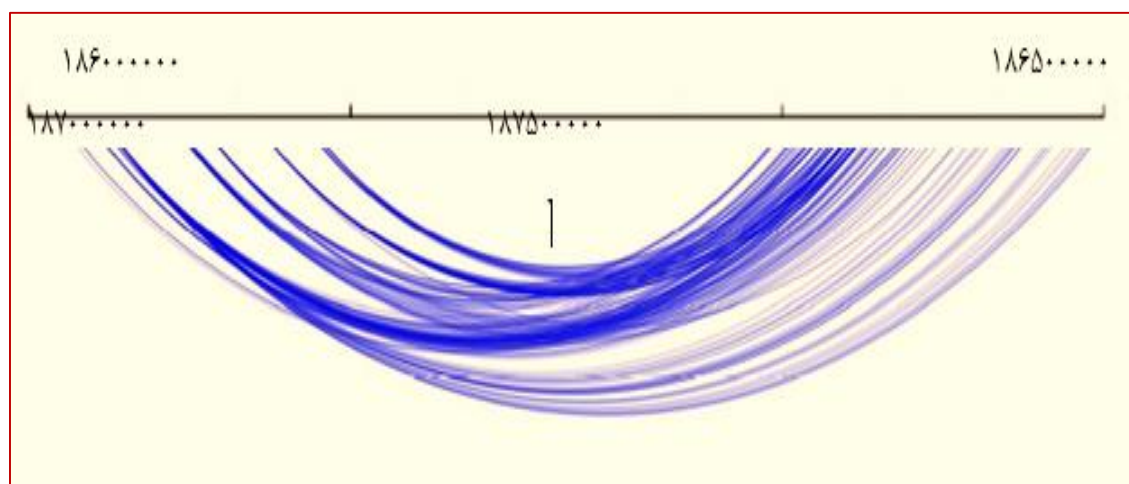
الگوریتم پیشنهادی، توالی‌های پروموتور/تشدیدکننده بالقوه با طول‌های متغیر را که تابع تناسب آن‌ها نسبت به حالت با طول ثابت بهتر باشد، استخراج می‌کند. بهتر بودن تابع تناسب بدین

به ترتیب شامل فرگمنت‌های به طول ۱۰ Kbp، ۱۰۰ Kbp، ۵۰۰ Kbp و ۱ Mbp که روش HiCUP قادر به کشف آن‌ها نبوده است را به تفکیک هر کروموزوم نشان می‌دهد. همان‌طور که مشاهده می‌شود، تعداد ۶۰۶ توالی‌های پروموتور/تشدیدکننده بالقوه در فرگمنت‌های مختلف توسط روش پیشنهادی استخراج گردیده است که روش میفسود قادر به کشف و استخراج آن نبوده است. مطابق با این جدول، در بین این ۶۰۶ توالی کشف‌شده، تعداد ۳۰۵ توالی به تفکیک هر کروموزوم در مجموعه داده با فرگمنت ۱۰ Kbp، تعداد ۱۹۹ توالی مربوط به مجموعه داده با فرگمنت ۱۰۰ Kbp، تعداد ۷۹ توالی در مجموعه داده با فرگمنت ۵۰۰ Kbp و در آخر تعداد ۲۳ توالی در مجموعه داده با فرگمنت ۱ Mbp است. همچنین این نتایج نشان می‌دهند که در مجموعه داده‌های با طول فرگمنت پایین روش پیشنهادی نسبت به روش میفسود در کشف برهم‌کنش‌ها عملکرد بهتری را دارا است.

در ادامه دو پروموتور کشف‌شده توسط روش پیشنهادی که در بروز و تشدید بیماری سرطان سینه تأثیر داشته‌اند را بررسی می‌کنیم. طبق تحقیقات انجام‌شده توسط میفسود و همکاران در (۱۰) دو پروموتور BLC۶ و HOTTIP در هر دو سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ تأثیر زیادی در بروز و تشکیل بیماری سرطان سینه داشته‌اند. مطابق با شکل ۱، پروموتور BLC۶ کشف‌شده توسط روش پیشنهادی، در هر دو سلول خونی GM۱۲۸۷۸

روش پیشنهادی در فرگمنت‌های با طول‌های ۱۰ Kbp، ۱۰۰ Kbp و ۵۰۰ Kbp از روش HiCUP بیشتر است. در مقابل، در مجموعه داده با فرگمنت ۱ Mbp مقدار میانگین این تابع در روش پیشنهادی تقریباً در همه کروموزوم‌ها نزدیک به روش HiCUP است. این نشان‌دهنده می‌دهد که روش پیشنهادی در فرگمنت‌های با طول کوچک‌تر نسبت به روش HiCUP عملکرد بهتری را در کشف پروموتور/تشدیدکننده‌های بالقوه دارد. همچنین به دلیل این که میانگین مقدار تابع تناسب مربوط به توالی‌های تنظیمی کشف‌شده توسط روش پیشنهادی در هر فرگمنت بهتر بوده است. در نتیجه، تعداد برهم‌کنش‌های شمارش‌شده بین آن‌ها نیز در روش پیشنهادی نسبت به روش مورد مقایسه بهتر است.

جدول ۲، میانگین تعداد برهم‌کنش‌های شمارش‌شده بین توالی پروموتور/تشدیدکننده بالقوه کشف‌شده مربوط به هر کروموزوم مجموعه داده‌های Hi-C ورودی، به ترتیب شامل فرگمنت‌های به طول ۱۰ Kbp، ۱۰۰ Kbp، ۵۰۰ Kbp و ۱ Mbp را نشان می‌دهد. همان‌طور که مشاهده می‌کنید، میانگین تعداد برهم‌کنش‌های شمارش‌شده در روش پیشنهادی بر اساس هر کروموزوم در همه فرگمنت‌ها بهتر از روش HiCUP است. در نتیجه، توالی‌های تنظیمی کشف‌شده توسط روش پیشنهادی می‌تواند قابل اطمینان‌تر^۱ از توالی‌های تنظیمی کشف‌شده توسط روش میفسود باشند.



شکل ۱- پروموتور BLC۶ کشف‌شده توسط روش پیشنهادی، در هر دو سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ و برهم‌کنش‌های آن

و CD۳۴⁺ در توالی ۱۸۶۰۰۰۰۰ تا ۱۸۶۵۰۰۰۰ در کروموزوم ۳ است که با توالی ۱۸۷۰۰۰۰۰ تا ۱۸۷۵۰۰۰۰ از همان

جدول ۳، تعداد توالی‌های پروموتور/تشدیدکننده بالقوه کشف‌شده توسط روش پیشنهادی در مجموعه داده‌های Hi-C ورودی،

^۱ robustness

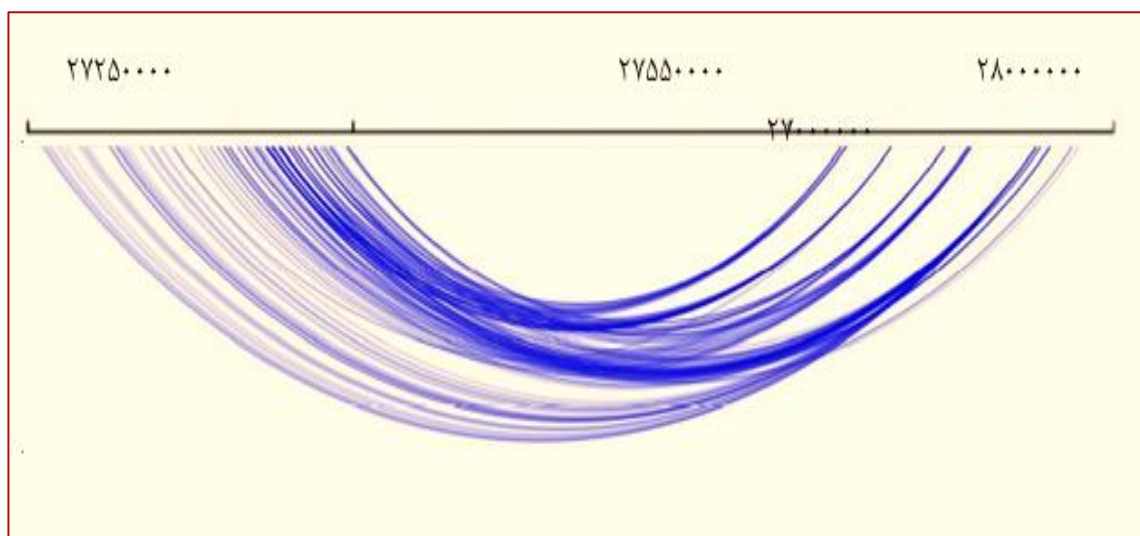
مقابل میفسود و همکاران با به کارگیری HiCUP، توالی تنظیمی را که این پروموتور موردنظر در آن قرار دارد توالی ۲۵۶۰۰۰۰۰ تا ۲۹۲۰۰۰۰۰ از کروموزوم ۷ کشف و استخراج کرده‌اند. این نشان می‌دهد که روش پیشنهادی در کشف این پروموتور و برهم‌کنش‌های آن نیز همانند پروموتور قبل عملکرد بهتری نسبت به روش میفسود داشته است.

بحث و نتیجه‌گیری

امروزه در علم ژنتیک به اثبات رسیده است که توالی‌های تنظیمی در کروموزوم‌های یک انسان بیمار که با یکدیگر برهم‌کنش دارند، دارای طول ثابت (Fix-bin size) نمی‌باشند بلکه طول آن‌ها می‌تواند متغیر باشد (۱۶-۱۴). همچنین تاکنون، در مطالعات انجام گرفته در (۸-۶) روش‌هایی پیشنهاد شده‌اند که توالی ثابت ژنوم که با یکدیگر برهم‌کنش دارند را کشف و استخراج

کروموزوم برهم‌کنش‌های زیادی را داشته است. توالی‌ای که پروموتور مذکور با آن برهم‌کنش دارد تقریباً در توالی‌ای به فاصله ۱Mbp نسبت به توالی پروموتور قرار گرفته است. روش پیشنهادی، توالی تنظیمی^۲ را که پروموتور BLC6 در آن قرار دارد توالی ۱۸۶۰۰۰۰۰ تا ۱۸۷۵۰۰۰۰۰ از کروموزوم ۳ کشف و استخراج کرده است؛ اما در مقابل میفسود و همکاران با به کارگیری HiCUP، توالی تنظیمی را که این پروموتور موردنظر در آن قرار دارد توالی ۱۸۵۸۰۰۰۰۰ تا ۱۸۹۴۰۰۰۰۰ از کروموزوم ۳ کشف و استخراج کرده‌اند. این نشان می‌دهد که روش پیشنهادی به صورت بهتری نسبت به روش میفسود قادر به کشف این پروموتور و برهم‌کنش‌های مؤثر آن بوده است.

پروموتور HOTTIP، پروموتور کشف شده توسط روش پیشنهادی است که در بروز و تشدید سرطان سینه مؤثر بوده است. مطابق با شکل ۲، پروموتور HOTTIP کشف شده توسط روش پیشنهادی،



شکل ۲- پروموتور HOTTIP کشف شده توسط روش پیشنهادی، در هر دو سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ و برهم‌کنش‌های آن

می‌کنند. در این صورت ممکن است توالی‌های تنظیمی کشف شده (که این توالی‌ها می‌توانند پروموتور/انهرهای بالقوه باشند)، بهینه نبوده و شمارش برهم‌کنش‌های بین آن‌ها دقیق نباشد. در این پژوهش، روشی مبتنی بر شبیه‌سازی تبرید جهت حل مسئله استخراج توالی‌های تنظیمی با طول‌های متغیر ارائه گردیده است. روش پیشنهادی می‌تواند توالی‌های تنظیمی با طول متغیر را کشف و استخراج کند. همچنین این روش قادر به شمارش برهم‌کنش‌های بین این توالی‌ها، نیز است. از جمله مزایای روش

در هر دو سلول خونی GM۱۲۸۷۸ و CD۳۴⁺ در توالی ۲۶۰۰۰۰۰۰ تا ۲۶۵۰۰۰۰۰ از کروموزوم ۷ است که با توالی ۲۷۵۰۰۰۰۰ تا ۲۸۰۰۰۰۰۰۰ از همان کروموزوم برهم‌کنش‌های زیادی را داشته است. توالی‌ای که پروموتور مذکور با آن برهم‌کنش دارد تقریباً در توالی‌ای به فاصله ۱/۱Mbp نسبت به توالی پروموتور در کروموزوم ۷ قرار گرفته است. روش پیشنهادی توالی تنظیمی را که پروموتور HOTTIP در آن قرار دارد توالی ۲۷۰۰۰۰۰۰ تا ۲۸۰۰۰۰۰۰۰ از کروموزوم ۷ کشف و استخراج کرده است؛ اما در

² Regulatory regions

موردبررسی قرار گرفت. نتایج نشان می‌دهند که دو پروموتور BLC6 و HOTTIP کشف‌شده توسط روش پیشنهادی در هر دو سلول خونی GM12878 و CD34⁺ در بروز و تشدید بیماری سرطان سینه تأثیر معنی‌داری داشته‌اند. مطابق مطالعات صورت گرفته توسط میفسود و همکاران در (۱۰) این دو پروموتور در هر دو سلول خونی تأثیرگذار در بروز و تشکیل بیماری سرطان سینه بوده‌اند. در نتیجه، روش پیشنهادی عملکرد قابل قبولی در کشف توالی تنظیمی دخیل در بروز سرطان را داشته است. مطابق نتایج و بررسی‌های صورت گرفته در بخش قبلی، توالی کشف‌شده پروموتور BLC6 در کروموزوم ۳ و همچنین توالی کشف‌شده پروموتور HOTTIP در کروموزوم ۷ توسط روش پیشنهادی دقیق-تر از روش میفسود بوده است. در همین راستا، شمارش برهم-کنش‌های هر دو پروموتور با توالی‌های دیگر از کروموزوم ۳ و ۷ نیز توسط روش پیشنهادی دقیق‌تر از روش میفسود است. این نشان می‌دهد که روش پیشنهادی به‌صورت بهتری نسبت به روش میفسود قادر به کشف این پروموتورها و برهم‌کنش‌های مؤثر آنها بوده است؛ بنابراین روش پیشنهادی که عملکرد بهتری در کشف و استخراج این توالی‌ها را دارد، می‌تواند به‌عنوان ابزاری جهت بهبود تشخیص بیماری سرطان به کار گرفته شود.

تشکر و قدردانی

مقاله موردنظر مستخرج از رساله دکتری با کد ۱۲۰۴۱۰۰۶۹۵۱۰۰۳ انجام‌شده توسط محمدجواد حسین پور به راهنمایی دکتر حمید پروین و دکتر صمد نجاتیان در دانشگاه آزاد اسلامی واحد یاسوج است. لذا نویسندگان از کلیه افرادی که در این مقاله همکاری نموده‌اند کمال تشکر را دارند.

تعارض منافع

نویسندگان هیچ‌گونه تعارض منافی را اعلام ننموده‌اند.

پیشنهادی می‌توان به همین دو مورد (اول، کشف و استخراج توالی‌های تنظیمی با طول متغیر و دوم، شمارش برهم‌کنش‌های بین آنها) اشاره کرد؛ که می‌تواند دقت در تشخیص این نواحی را بالا ببرد. همچنین سرعت‌پایین اجرای روش پیشنهادی نیز از جمله عیبی است که در این روش وجود دارد. نتایج و مقایسه-های انجام‌گرفته، حاکی از عملکرد بالا و بهینه روش پیشنهادی در کشف توالی‌های تنظیمی با طول متغیر نسبت به روش بکار رفته در میفسود (HiCUP) است. درواقع روش HiCUP به‌عنوان ابزاری قدرتمند در آنالیز داده‌های Hi-C بشمار می‌رود؛ اما این روش توانایی کشف توالی‌های تنظیمی با طول متغیر را ندارد.

به دلیل اینکه، روش پیشنهادی قادر به کشف توالی‌های تنظیمی با طول‌های متغیر است؛ بنابراین، در همه طول فرگمنت‌های مورد آزمایش، میانگین مقدار تابع تناسب الگوریتم پیشنهادی در هر کروموزوم نسبت به روش HiCUP نیز بهتر است. در حقیقت الگوریتم پیشنهادی، توالی‌های تنظیمی با طول-های متغیر را که تابع تناسب آنها نسبت به حالت با طول ثابت بهتر باشد، استخراج می‌کند. بهتر بودن تابع تناسب بدین معنی است که تعداد برهم‌کنش‌های بین توالی‌های تنظیمی استخراج‌شده با طول متغیر بیشتر از حالت با طول ثابت باشد. نتایج مربوط به میانگین مقدار تابع تناسب نشان می‌دهد که روش پیشنهادی در مجموعه داده با فرگمنت‌های با طول کوچک‌تر نسبت به روش HiCUP عملکرد بهتری را در کشف توالی‌های تنظیمی با طول متغیر دارد. همچنین تعداد برهم‌کنش‌های شمارش‌شده بین این توالی‌ها در روش پیشنهادی نسبت به روش مورد مقایسه نیز بهتر است.

در پایان، عملکرد روش پیشنهادی در تشخیص و استخراج توالی‌های تنظیمی دخیل در بروز و تشدید بیماری سرطان سینه

References

1. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326(1):289–93.
2. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58(1):268–76



3. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep.* 2015; 10(1):1297–309.
4. Hwang YC, Zheng Q, Gregory BD, Wang LS. High-throughput Identification of Long-range Regulatory elements and their Target Promoters in the Human Genome. *Nucleic Acids Res.* 2013; 41(9):4835–4846.
5. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010; 38(4):576-589.
6. Sauria ME, Phillips-Cremens JE, Corces VG, Taylor J. HiFive: a tool suite for easy and efficient HiC and 5C data analysis. *Genome Biol.* 2015; 16(237):1-10.
7. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature letter* 2013; 503(7475):290-294.
8. Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics.* 2014; 30(19):2808–10.
9. Cabrerios I, Abbe E, Tsirig A. Detecting Community Structures in Hi-C Genomic Data. Cornell Library University. 2015; 1509(1):1-17.
10. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics* 2015; 47(6):598-606.
11. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an Optimized and Flexible pipeline for Hi-C Data Processing. *Genome Biol.* 2015; 16(259): 1-11.
12. Charalampos L, Stephen KBP, Panagiotis N, Iannis A, Aristotelis T, HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarKbping. *BMC Genomics.* 2017; 18(22):1-16.
13. Riccardo, C, Qiuyang, W, Jihong, G, and Sheng, Z, GITAR: an open source tool for analysis and visualization of Hi-C data. *Genomics Proteomics Bioinf.* 2018;16(5):365-372.
14. Akdemir KC, Chin L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* 2015; 16(198):1-8.
15. Liu F, Li H, Ren C, Bo X and Shu W. PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Nature.* 2016; 6(28517):1-14.
16. Hosseinpour M, Parvin H, Nejatian S, Rezaei V. Detection and Extraction of Potential Promoter/Enhancer Interactions in Genome of Cancer Patients using an Evolutionary Multi-Objective Algorithm. *Journal of Health and Biomedical Informatics.* 2018; 5 (2):304-313.



Original Article

Presenting a Meta-Heuristic Algorithm to Detect Regulatory Elements in the Genome of Breast Cancer Patients

Hosseinpoor MJ^{1,7}, Parvin H^{2,6*}, Nejatian S^{3,5}, Rezaee V^{4,5}, Bagherifard K^{1,5}

1. Department of Computer Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

2. Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

3. Department of Electrical Engineering, Yasooj Branch, Islamic Azad University, Yasooj, Iran

4. Department of Mathematics, Yasooj Branch, Islamic Azad University, Yasooj, Iran

5. Young Researchers and Elite Club, Yasooj Branch, Islamic Azad University, Yasooj, Iran

6. Young Researchers and Elite Club, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

7. Young Researchers and Elite Club, Estahban Branch, Islamic Azad University, Estahban, Iran

Received: 16 Aug 2019

Accepted: 01 Feb 2020

Abstract

Background & Objectives: Nowadays, in medical sciences, the amount of data on symptoms of people affected with various illnesses on one hand, and finding assistive techniques for the diagnosis of those diseases on the other, has been widespread. Consequently, the analysis and consideration of all factors involved in a disease are often challenging. Thus, a mechanized system to help discover the rules, identify patterns, and predict future events is absolutely needed. In this research, we intend to use a multi-objective algorithm to provide a method capable of detecting, extract sequences of variable-length from the genome, and count the interactions among them. In fact, these regulatory elements could play a significant role in the incidence and exacerbation of cancer.

Material & Methods: In this research, a proposed method for the detection of regulatory elements in the genome of a breast cancer patient has been used. The proposed method is implemented in MATLAB software. Also, to measure the performance and effectiveness of the suggested method, the proposed algorithm is implemented on HiC dataset, regarding patients with breast cancer in two blood cells GM12878 and CD34+ introduced by Mifsud et al.

Results: The results of implementing the proposed method are compared with the HiCUP method. The results show that the MSARE method has a better performance in detecting regulatory elements compared to the HiCUP method.

Conclusion: Experimental studies have shown that the two promoters BLC6 and HOTTIP discovered by the proposed method have had a significant effect on the incidence and severity of breast cancer in both blood cells GM12878 and CD34+.

Keywords: Regulatory Elements, Meta-Heuristic Algorithm, BLC6 Promoter, HOTTIP Promote

***Corresponding Author:** Hamid Parvin, Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

Email: parvin@iust.ac.ir

<https://orcid.org/0000-0002-4877-182X>