

NGS Serves to Promote Personalized Medicine: A Case- Report of a Patient with Invasive Breast Carcinoma

Leila Mirsadeghi¹, Reza Haji Hosseini²,
Ali Mohammad Banaei-Moghaddam³,
Kaveh Kavousi^{4*}

1. Ph. D., Department of Biology, Faculty of Science, Payame
Noor University, Tehran, Iran

2. Professor, Department of Biology, Faculty of Science,
Payame Noor University, Tehran, Iran

3. Assistant Professor, Laboratory of Genomics and
Epigenomics (LGE), Department of Biochemistry, Institute of
Biochemistry and Biophysics (IBB), University of Tehran,
Tehran, Iran

4. Assistant Professor, Laboratory of Complex Biological
Systems and Bioinformatics (CBB), Department of
Bioinformatics, Institute of Biochemistry and Biophysics
(IBB), University of Tehran, Tehran, Iran

(Received: Mar. 13, 2021 - Accepted: Oct. 17, 2021)

NGS در خدمت ارتقای پزشکی شخص محور: گزارش موردی یک بیمار متلا به سرطان پستان تهاجمی

لیلا میرصادقی^۱، رضا حاجی حسینی^۲،

علی محمد بنایی مقدم^۳، کاوه کاوسی^{۴*}

۱. دکتری، گروه بیولوژی، دانشکده علوم، دانشگاه پیام نور، تهران، ایران

۲. استاد، گروه بیولوژی، دانشکده علوم، دانشگاه پیام نور، تهران، ایران

۳. استادیار، آزمایشگاه ژنومیکس و اپیژنومیکس (LGE)، گروه بیوشیمی،

مرکز تحقیقات بیوشیمی و بیوفیزیک (IBB)، دانشگاه تهران، تهران، ایران

۴. استادیار، آزمایشگاه سیستم‌های زیستی پیچیده و بیوانفورماتیک (CBB)،

گروه بیوانفورماتیک، مرکز تحقیقات بیوشیمی و بیوفیزیک (IBB)، دانشگاه

تهران، تهران، ایران

(تاریخ دریافت: ۱۳۹۹/۱۲/۲۳ - تاریخ پذیرش: ۱۴۰۰/۷/۲۵)

Abstract

The emergence of personalized medicine based on molecular techniques, such as next-generation sequencing, has increased our understanding of drivers of complex diseases, including cancers. In many cases due to the complexity of cancer, it is difficult for human physicians and biologists to make decisions on the basis solely of clinical practice or laboratory evidence. Thus, the personalized medicine approach comes into play and provides large volumes and valuable data for experts. Further, data analysis with bioinformatic tools has opened a new horizon in the process of prognosis and screening of in risk individuals. It has caused significant recent advances in diagnostic technology and improved targeted treatments. In the present study, archived formalin-fixed paraffin-embedded tissue from an Iranian female patient with invasive breast carcinoma was investigated. In this way, after DNA extraction and purification, the whole exome was sequenced and the mutation data were analyzed. Obtained information could help to the enrichment of the Iranian genome databases. In the light of this research and by studying other Iranian samples, we can provide an optimized roadmap for precision oncologists to increase the life expectancy of breast cancer patients.

Keywords: Bioinformatics, Invasive breast carcinoma, Mutation data, Next-generation sequencing, Personalized medicine.

چکیده

ظهور پزشکی شخص محور با تکیه بر تکنیک‌های مولکولی، از جمله ترادف‌یابی نسل بعدی، درک ما را از محرک‌های مؤثر بر بیماری‌های پیچیده، از جمله سرطان افزایش داده است. در بسیاری از موارد به دلیل پیچیدگی سرطان، تصمیم‌گیری فقط بر اساس اقدامات بالینی یا شواهد آزمایشگاهی برای پزشکان و زیست‌شناسان امری دشوار است. بنابراین، رویکرد پزشکی شخصی وارد عمل می‌شود و داده‌های انبوه و ارزشمندی را در اختیار متخصصان قرار می‌دهد. علاوه بر این، تجزیه و تحلیل داده‌ها با ابزارهای بیوانفورماتیک، افق جدیدی را در روند پیش‌آگهی و غربالگری افراد در معرض خطر گشوده است. این امر باعث پیشرفت‌های چشمگیر اخیر در فناوری تشخیص شده و درمان‌های هدفمند را بهبود بخشیده است. در مطالعه حاضر، نمونه بایگانی شده از بافت تثبیت شده با فرمالین و جاسازی شده در پارافین مربوط به یک بیمار زن ایرانی مبتلا به سرطان پستان تهاجمی به‌طور موردی بررسی شد. به این ترتیب که پس از استخراج و خالص‌سازی DNA، ترادف‌یابی کل اگزوم انجام شد و داده‌های جهش مورد تجزیه و تحلیل قرار گرفت. اطلاعات به دست آمده می‌تواند به غنی سازی پایگاه‌های داده ژنوم ایرانی کمک کند. در پرتو این پژوهش و با مطالعه نمونه‌های ایرانی بیشتر، ما می‌توانیم یک نقشه راه بهینه برای زیست‌شناسان پژوهشگر در حیطه سرطان و آنکولوژیست‌های دقیق ارائه دهیم تا امید به زندگی بیماران مبتلا به سرطان پستان افزایش یابد.

واژه‌های کلیدی: بیوانفورماتیک، پزشکی شخص محور، ترادف‌یابی نسل بعدی، داده‌های جهش، کارسینومای پستان تهاجمی شده.

مقدمه

ترادف‌یابی نسل بعدی (NGS)، از جمله ترادف‌یابی کل ژنوم (WGS) یا ترادف‌یابی کل اگزوم (WES) می‌تواند حجم زیادی از داده‌های مرتبط با ژن‌های جهش یافته تولید کند که داده‌های جهش نامیده می‌شوند (Sheikine *et al.*, 2017). در حال حاضر، لیستی از اهداف ژنی برای درمان‌های مورد تأیید FDA و نیز به‌عنوان معیارهایی برای آزمایشات بالینی ارائه شده‌اند و مصوبات اخیر انجام آزمایش NGS بر روی تمام بیماران سرطانی در مرحله متاستاتیک را ضروری دانسته‌اند (McKenzie *et al.*, 2019). اما مشکل استفاده از داده‌های جهش، انبوه بودن آنهاست، به نحویکه تجزیه و تحلیل این داده‌های گسترده نیاز به استفاده از رویکردهای محاسباتی قوی برای بهره‌برداری کارآمدتر اطلاعات دارد. در این راه، ظهور ابزارهای بیوانفورماتیک به موازات توسعه تکنیک‌های مولکولی می‌تواند اطلاعات ارزشمند نهفته در این داده‌های انبوه را استخراج کند و در اختیار پزشکان و متخصصین قرار دهد. این ابزارها تشخیص تغییرات مولکولی پیش آمده و پیش‌آگهی بیماری را بهبود بخشیده‌اند و کمک می‌کنند تا افراد در ریسک زود هنگام و در مراحل اولیه غربالگری شوند. این تغییر چشم‌انداز که در آن توجه متخصصین به تغییرات خاص مولکولی تومور معطوف است، دریچه‌ای برای درمان شخصی باز کرده است تا درمان هدفمند برای بیماران با توجه به داده‌های به‌دست‌آمده از هر شخص انتخاب شود (Gambardella *et al.*, 2020). چنین رویکردی دانش پزشکی شخص‌محور (Personalized medicine) نامیده می‌شود و مدیون پیشرفت‌های اخیر قابل توجه در تکنولوژی‌های مولکولی به ویژه تکنیک NGS و نیز استفاده از دانش بیوانفورماتیک است. آن یک مدل پزشکی است که پروسه بهداشت و درمان را برای هر شخص سفارشی‌سازی می‌نماید و این اطمینان را حاصل می‌کند که انتخاب دارو و نوع درمان برای فرد مبتلا به‌طور ویژه طراحی شود و تصمیمات پزشکی متناسب و بهینه، بر پایه زمینه ژنتیکی هر بیمار انتخاب شود (Sheikine *et al.*, 2017). این تکنیک از سال

جهش‌ها تغییرات کوچکی در ژن‌ها ایجاد می‌کنند. اگر ژن‌ها آسیب ببینند و ترمیم نشوند، منجر به بروز ناهنجاری‌های چند عاملی می‌شوند که بیماری‌های پیچیده نامیده می‌شوند. تعداد ژن‌های جهش‌یافته که ترادف‌یابی می‌شوند به سرعت در حال افزایش است. با این حال بیشتر این ژن‌ها از نوع مسافر (Passenger) هستند که قابلیت برای بیماری‌زایی ندارند (Greenman *et al.*, 2007; Watson *et al.*, 2013). ژن‌ها با جهش‌های مهم را می‌توان به قله‌ی کوه تشبیه کرد، درحالی‌که اکثریت قریب به اتفاق ژن‌ها به تپه‌هایی شباهت دارند که به ندرت جهش‌های مهم در آنها اتفاق می‌افتد (Wood *et al.*, 2007). به ژن‌های اصلی در بروز هر بیماری، ژن‌های محرک (Driver) گفته می‌شود. سرطان‌ها، یکی از بیماری‌های پیچیده هستند که توسط ژن‌های محرک معیوب القا می‌شوند. در این میان، سرطان پستان اولیه به‌عنوان یک بیماری پیچیده، شایع‌ترین سرطان تشخیص داده شده در زنان سراسر جهان است که در صورت پیشرفت به مرحله ثانویه و متاستاتیک، کشنده خواهد بود. به‌طور کلی، آن پنجمین علت شایع مرگ و میر ناشی از سرطان است. در میان زنان، شایع‌ترین علت مرگ ناشی از سرطان در مناطق در حال توسعه است و در زنان مناطق توسعه یافته پس از سرطان ریه دومین علت شایع مرگ‌ومیر سرطانی محسوب می‌شود (Kumar & Singla, 2019). متأسفانه سن بروز این سرطان در زنان کشور ما دست‌کم یک دهه زودتر از زنان در کشورهای توسعه‌یافته است به نحویکه بسیاری از مبتلایان در محدوده سنی ۴۰ تا ۴۹ سال قرار دارند (Ebrahimi *et al.*, 2015)، که بر اهمیت موضوع می‌افزاید. از طرفی شناخت زیست‌شناسی نشانه‌های مولکولی در هر سرطان جهت پیش‌آگهی و درمان زود هنگام، همواره مورد توجه سرطان‌شناسان بوده‌است. متخصصان آنکولوژی برای شناسایی این نشانه‌ها به داده‌های فراوان نیاز دارند (McKenzie *et al.*, 2019). امروزه، تمرکز بر تکنیک مولکولی

BRCA2 که خود به عنوان ژن های سرکوبگر تومور هستند، نرخ ابتلا به سرطان پستان سه برابر گزارش شده است (Chai *et al.*, 2015). بنابراین با اقدام زودهنگام جهت ترادف یابی ژنوم افراد، به ویژه کسانی که دارای زمینه های خانوادگی در بروز سرطان پستان می باشند می توان افراد در ریسک را پیش از موعد در میان انبوه جمعیت غربالگری نمود تا تحت درمان های پیشگیرانه مؤثر قرار گیرند. در یک مطالعه دیگر مبتنی بر *NGS* بر روی ۱۷۸۱ شرکت کننده، فراوانی جهش در ۲۵ ژن بررسی شد. یافته ها نشان می دهد که تقریباً ۶۶ درصد جهش های ژنی این افراد مربوط به ژن های *BRCA1* و *BRCA2* است. شواهد بالینی وجود دارد که در سرطان هایی که نقص در این دو ژن فراوانی بیشتری داشته اند، مبتلایان، به درمان با مهارکننده های پلاتین و پلی ADP-ریبوز پلیمراز-۱ پاسخ خوبی داده اند. این مهارکننده ها می توانند به عنوان داروی هدف برای این بیماران استفاده شوند (Tung *et al.*, 2015). در سال ۲۰۱۹ و مبتنی بر اطلاعات به دست آمده از *NGS*، ۷۹ ژن در شناسایی بیماران مبتلا به سرطان پستان اولیه و سرطان پستان ثانویه ارزیابی و تأیید شدند. این ژن ها می توانند در اهداف بالینی و درمانی جدید مورد تحقیق بیشتر قرار گیرند و در جهت درمان هدفمند مفید واقع شوند (Smith *et al.*, 2019). در همین راستا و در پژوهش حاضر، نمونه منشأ گرفته از بافت پستان یک بیمار زن ایرانی مبتلا به سرطان پستان تهاجمی که از قبل پارافینه شده بود، از بین ۴۰ مورد با هویت ناشناس که همگی از نمونه های بایگانی شده بودند، انتخاب و به طور موردی بررسی شد. بدین ترتیب که پس از استخراج و تخلیص *DNA* از بافت پستان تثبیت شده در فرمالین و جاسازی شده در پارافین (*FFPE*)، ترادف یابی کل اگزوم *DNA* انجام شد و جهش های رخ داده در آن مورد تجزیه و تحلیل بیوانفورماتیکی قرار گرفت.

مواد و روش ها

از میان ۴۰ نمونه *FFPE* به دست آمده از بافت پستان

۲۰۰۶ به صورت تجاری در دسترس است. به این ترتیب که با اتمام پروژه ژنوم انسان در سال ۲۰۰۳، تحقیقات در زمینه آنکولوژی با هدف درک بهتر مبنای ژنتیکی آنکوژن و شناسایی تغییرات ایجاد شده بر ترادف ژنوم متمرکز شد و منجر به ظهور "عصر ژنومیک" در عرصه تحقیقات و درمان سرطان گردید. *NGS* با امکان شناسایی جهش های محرک و کمی سازی اطلاعات، اساس یک رویکرد جدید در مقابله با سرطان را ایجاد کرده است (Morganti *et al.*, 2019). طی دو دهه ی اخیر، پیشرفت های بسیاری در تکنیک های مولکولی مرتبط با عرصه تشخیص های بالینی انجام شده است. پزشکی شخص محور با تکیه بر به خدمت گرفتن چنین تکنیک هایی و همچنین بر پایه وجود تنوع ژنتیکی بین افراد و نژادهای مختلف توانسته است روند و عملکرد درمان بیماران را بهبود بخشد. پیش بینی شده است که موجی از درمان های آینده مربوط به سرطان، به کمک پزشکی شخص محور امکان پذیر خواهد شد (Sheikine *et al.*, 2017). امروزه مطابق با پروتکل های شورای تحقیقات ملی آمریکا (*NRC*) که زیر مجموعه ی اینستیتو ملی سلامت (*NIH*) است بهتر است از عبارت پزشکی دقیق (*Precision medicine*) به جای پزشکی شخص محور استفاده شود اگرچه که همچنان در جوامع پزشکی و علمی از این دو اصطلاح به جای هم استفاده می شود و همپوشانی بسیاری بین مفهوم این دو عبارت وجود دارد. در حیطه سرطان، پزشکی دقیق سرطان شناسی دقیق (*Precision oncology*) خوانده می شود و هدف آن یافتن ژن های کلیدی هدفمند و نشانگر برای هر سرطان است تا به جای ترادف یابی کل ژنوم، ترادف یابی پانل ژنی هدفمند (*Targeted gene panel sequencing*) انجام شود تا هزینه ها کاهش و سرعت تشخیص و تحلیل نتایج برای هر بیمار بهبود یابد (Sheikine *et al.*, 2017). مطالعات مختلف شواهدی از خدمات پزشکی شخص محور و دقیق در پیشگیری و درمان بیماران را ارائه داده اند. در یک پژوهش انجام شده روی زنان دارای جهش در ژن های *BRCA1* و

مربوط به زنان مبتلا به کارسینومای تهاجمی پستان که از نمونه‌های موجود در بایگانی بخش پاتولوژی یک بیمارستان تخصصی در شهر تهران، بدون دخالت در هرگونه کارآزمایی بالینی جمع‌آوری شد، نمونه مربوط به یک زن ۴۴ ساله مبتلا به کارسینومای داکتال تهاجمی پستان (*Invasive ductal carcinoma*) با هویت ناشناس (*anonymous*) جهت انجام استخراج *DNA* و انجام *WES* انتخاب شد. بیمار در گرید سوم بیماری و تومور مربوطه با ابعاد $1/5 \times 1/8 \times 1/1$ و در مرحله *T1cN1aMx* تشخیص داده شده بود. در ادامه، نمونه جهت انجام تست‌های مولکولی به نمایندگی شرکت *Novogene* در آسیا ارسال شد. بررسی اسلاید پاتولوژی موجود در بایگانی نشان داد که بیشتر بافت برداشته شده از نوع توموری بوده و درصد سلول‌های نرمال در نمونه ناچیز می‌باشد.

آماده‌سازی کتابخانه برای ترادف‌یابی

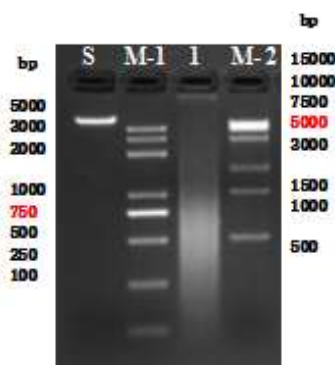
برای تهیه کتابخانه *DNA*، از کل $4/280$ میکروگرم از *DNA* ژنومیک استخراج شده، مقدار یک میکروگرم به‌عنوان ماده ورودی استفاده شده‌است. کتابخانه با استفاده از کیت *Agilent SureSelect Human All Exon* (*Agilent Technologies, CA, USA*) در شرکت *Novogene* تولید شد. قطعات $280-180$ جفت باز ایجاد شدند. در ادامه، خوشه‌هایی از قطعات *DNA* تولید و میلیون‌ها نسخه از *DNA* تک رشته‌ای تولید شده‌اند. در پایان این مرحله، محصولات به‌دست‌آمده با استفاده از سیستم *AMPure XP* (*Beckman Coulter, Beverly, USA*) خالص‌سازی و با روش *Agilent high sensitivity DNA* روی سیستم *Agilent Bioanalyzer 2100* کمی‌سازی شدند تا میانگین سائز قطعات ارزیابی شود.

استخراج و خالص‌سازی *DNA*

در مطالعه حاضر، فرایند استخراج و خالص‌سازی *DNA* ژنومیک از بافت جاسازی شده در پارافین به‌کمک *QIAamp DNA FFPE Tissue Kit* انجام شد.

کنترل کیفی *DNA* استخراج شده و خلوص آن

در ادامه خلوص و یکپارچگی *DNA* تخلیص شده با آزمون کنترل کیفی (*QC*) و به‌کمک تکنیک



S: نمونه استاندارد

Trans 2k plus DNA ladder :M-1

1: نمونه مورد مطالعه*

Trans 15k plus DNA ladder :M-2

*ظاهر نشدن باند تک بر روی ژل در استخراج *DNA* از نمونه‌های بافت پارافینه، به دلیل اجتناب‌ناپذیری تخریب بخشی از *DNA* در طول زمان جاسازی شدن در پارافین نتیجه‌ای نرمال است.

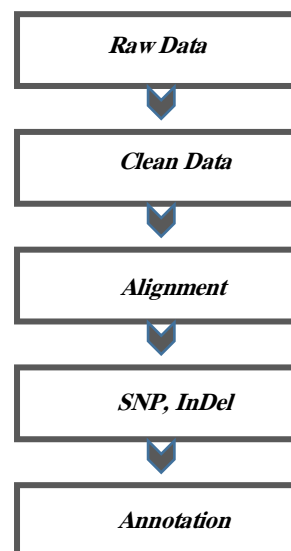
شکل ۱. نتایج الکتروفورز مشاهده‌شده روی ژل آگارز. شرح علائم اختصاری در کادر راهنما آمده است.

ترادفیابی

پس از آنکه واجد شرایط بودن کتابخانه در شرکت *Novogene* کنترل و تأیید شد، آن روی یک بستر *Illumina* قرار گرفت و با توجه به غلظت مؤثر و حجم داده مورد نیاز، کل اگزوم در دستگاه *sequencer* بر پایه تکنولوژی *PE150 HiSeq* ترادفیابی شد. روند خوانش تا جایی ادامه می‌یابد که کل اگزوم ترادفیابی شود. در مطالعه حاضر، عمق ترادفیابی $200\times$ محاسبه شد. معیار عمق، به صورت میانگین تعداد دفعاتی که هر باز آلی در ژنوم اندازه‌گیری می‌شود تعریف می‌گردد (Jiang et al., 2019).

تجزیه و تحلیل بیوانفورماتیکی داده‌های تولیدشده

تجزیه و تحلیل بیوانفورماتیکی داده‌های خروجی با توجه به روند کاری شرکت *Novogene* طی چندین مرحله انجام شد که پایپ‌لاین مربوطه با استفاده از یک نمودار جریان نشان داده شده است (شکل ۲).



شکل ۲. روند کاری مراحل تجزیه و تحلیل بیوانفورماتیکی

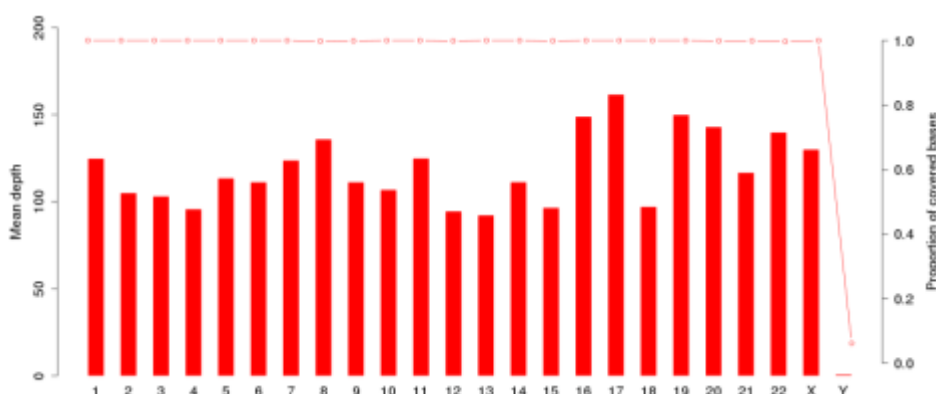
در مرحله تهیه فایل حاشیه‌نویسی برای واریانت‌ها، از روش ANNOVAR استفاده شده است (Wang & Hakonarson, 2010)، تا جنبه‌های مختلفی از داده‌های جهش، از جمله تغییرات در کدگذاری پروتئین‌ها آشکار شوند.

ارزیابی

در این گام برای یافتن مناطق ژنومی تحت تأثیر که تغییرات پروتئینی نشان می‌دهند، از پایگاه‌های داده *RefSeq* و *Gencode* استفاده شده است. برای پیش‌بینی مخرب بودن جهش‌ها، ابزارهای *SIFT*، *PolyPhen*، *MutationAssessor*، *LRT* و *CADD* به کار گرفته شدند. همچنین پایگاه‌های داده *GWAS Catalog*، *OMIM*، *COSMIC*، *dbSNP* و *HGMD* استفاده شدند تا بتوانیم اطلاعات موجود گزارش شده مرتبط با گستره ژنوم و برخی از انجمن‌های سرطان را با داده‌های جهش به دست آمده از اگزوم بیمار مورد مطالعه مقایسه کنیم. در فایل حاشیه‌نویسی، برای پیدا کردن مسیرهای بیوشیمیایی مرتبط با هر ژن از داده‌های موجود در پایگاه‌های داده *Gene Ontology* (Consortium, 2004)، *KEGG*، *Reactome* (Kanehisa & Goto, 2000)، *Biocarta* (Fabregat et al., 2017)، *PID* (Nishimura, 2001) و (Schaefer et al., 2009) استفاده شده است.

نتایج

در این مطالعه، ترادفیابی در سطح اگزوم یک بیمار مبتلا به سرطان تهاجمی پستان انجام شد و ۷۳۹۷۱۷۱۰ قطعه نوکلئوتید خوانش شدند که از این تعداد، نگاهت ۷۳۸۷۲۴۶۸ قطعه به ژنوم مرجع انجام شد. بر این اساس، داده‌های متعدد مربوط به ترادفیابی کل اگزوم به دست آمد که ما به اختصار نتایج را در سطح چندشکلی‌های تک نوکلئوتیدی (*SNPs*) و اضافه و حذف‌ها (*InDels*)ی رخ داده در ژن‌ها بررسی و نیز اطلاعاتی را بر مبنای تحلیل‌های بیوانفورماتیکی به دست آمده از فایل حاشیه‌نویسی واریانت‌ها گزارش می‌نماییم. اطلاعات کلی مربوط به ترادفیابی انجام شده، از جمله متوسط عمق خوانش و همچنین میزان *coverage* بازهای نیتروژندار مربوط به هر کروموزوم به کمک نمودار نمایش داده شده است (شکل ۳).



شکل ۳. نمایش عمق متوسط خوانش (نمودار میله‌ای) و Coverage هر باز (نمودار خط-نقطه) در هر کروموزوم

پیش در پایگاه‌های داده مرتبط با *SNP* گزارش شده‌اند و بقیه جدید هستند. باید توجه داشت اگرچه ترادف یابی در نواحی اگزونی انجام شده است اما در جدول، *SNP*‌های موجود در برخی نواحی دیگر ژنوم نیز گزارش شده‌است. علت آن است که گاهی در مرحله ساخت کتابخانه، به دام‌انداختن اگزوم از مرز این نواحی فراتر می‌رود و به اینترون‌ها و سایر نواحی می‌رسد. این نکته مهم است که تقریباً ۱۵ درصد انواع واریانت‌های بیماری‌زا مربوط به نواحی اینترونیک هستند و مرتبط با جهش‌هایی هستند که در مرز اگزون-اینترون رخ می‌دهند (Truty et al., 2019). فراوانی *SNP*‌ها در مناطق مختلف ژنوم و مناطق کدکننده نمونه مورد مطالعه در نمودار دایره‌ای مقایسه شده‌اند (شکل ۴).

InDel‌های تشخیص داده شده

اضافه و حذف (*InDel*)‌های کوچکی که طول آنها کمتر از ۵۰ جفت باز است، دسته دیگری از انواع تغییرات در ژنوم انسان را تشکیل می‌دهد. اگر تعداد نوکلئوتیدهای وارد شده یا حذف شده سه یا مضربی از عدد سه نباشد، کل چارچوب خوانش تغییر خواهد کرد. چنین *InDel*‌هایی که در منطقه کدکننده یا جایگاه‌های *splicing* رخ می‌دهند می‌توانند باعث تغییر در رونوشت‌ها و پروتئین‌ها شوند. داده‌های آماری مربوط به *InDel*‌های شناسایی شده در نمونه بیمار تحت مطالعه در جدول ۲ لیست شده است.

SNP‌های تشخیص داده شده

چندشکلی‌های تک نوکلئوتیدی (*SNPs*) که با عنوان تنوع تک نوکلئوتیدی (*SNVs*) نیز شناخته می‌شوند، بزرگترین دسته از تغییرات هستند که باعث ایجاد تنوع در ژنوم می‌شوند. کل ژنوم معمولی انسان حدود ۳/۶ میلیون *SNP* دارد. در نمونه مورد مطالعه، تعداد ۴۲۴۰۴۷ چند شکلی نوکلئوتیدی مربوط به ۱۹۳۲۷ ژن مشاهده می‌شود که آمار جزئیات آن در جدول ۱ نشان داده شده است.

جدول ۱. فراوانی *SNP*‌های مشاهده شده

تعداد	انواع <i>SNP</i>	مناطق کدکننده
۱۱۲۶۲	synonymous	
۱۰۲۹۵	missense	
۹۲	stopgain	
۱۵	stoploss	
۳۹۳	ناشناخته	
۱۹۴۳۳۴	intronic	مناطق غیر کدکننده
۵۷۰۷	UTR3	
۲۹۲۵	UTR5	
۶۵	splicing	
۲۸۸۲	ncRNA_exonic	
۱۵۰۹۹	ncRNA_intronic	
۱۳	ncRNA_splicing	
۴۸۲۱	upstream	
۳۳۷۳	downstream	
۱۷۲۷۷۱	intergenic	
۴۲۴۰۴۷	مجموع	

از این تعداد، ۲۲۰۵۷ پلی مورفیسم مربوط به نواحی اگزونیک است. علاوه بر این، ۶۱/۰۶ درصد *SNP*‌ها از

منحصربه‌فرد در مناطق کدکننده در یک نمودار دایره‌ای ترسیم و مقایسه شده‌اند (شکل ۵).

حاشیه‌نویسی واریانت‌ها

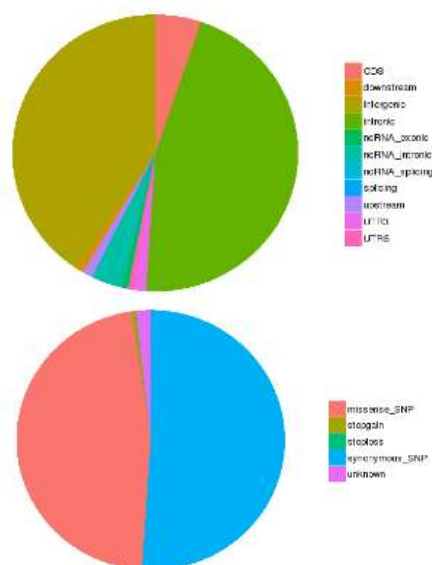
ارزیابی فایل‌های حاشیه‌نویسی مربوط به داده‌های SNP و InDel بخشی دیگر از تجزیه و تحلیل بیوانفورماتیکی این مطالعه است.

ارزیابی داده‌های SNP

در این مطالعه، از بین ژن‌هایی که در ناحیه اگزونیک، پلی‌مورفیسم تک نوکلئوتیدی نشان دادند، پس از حذف *synonymous_SNP*‌ها که در آن هر دوی آلل‌ها مترادف یکسانی از پلی‌پپتید را تولید می‌کنند، ده ژن *HLA-A*، *MUC16*، *HLA-DRB1*، *AHNAK2*، *HLA-B*، *PDE4DIP*، *HLA-DQA1*، *HLA-C*، *HLA-DRB5* و *HLA-DQB1* بیشترین پلی‌مورفیسم تک نوکلئوتیدی را نشان دادند. این نتیجه در مورد *HLA*‌ها با توجه به این که بیشترین تنوع را در گستره‌ی ژنوم دارند بدیهی است (Gourraud et al., 2014). در نواحی غیراگزونیک، بیشترین SNP‌ها در *RBFOX1*، *HLA-DQB1*، *RXR2*، *HLA-DRB1*، *CSMD1* و *DMD*، *BCAS3*، *HLA-DQA1*، *RBFOX3* و *CUX1* دیده می‌شوند. همچنین، غنی‌سازی داده‌های SNP در فهرست ژن‌های شناخته شده سرطان پستان موجود در چندین پایگاه داده عمومی مرتبط با سرطان انجام شده که نتایج تأییدشده در جدول ۳ آمده است.

ارزیابی داده‌های InDel

در این مطالعه ۱۱۷۲۸ ژن با تغییرات از نوع InDel یافت شدند. در نواحی اگزونیک پس از حذف *nonframeshift_deletion*‌ها و *nonframeshift_insertion*‌ها که تغییر در چارچوب خوانش ایجاد نمی‌کنند، InDel در ۲۴۱ ژن دیده شد که فراوانی آن در هفت ژن *HLA-B*، *HLA-DRB1*، *HLA-DQA1*، *HLA-DRB5*، *SARM1*، *ZNF595* و *PABPC3* بیشتر است.

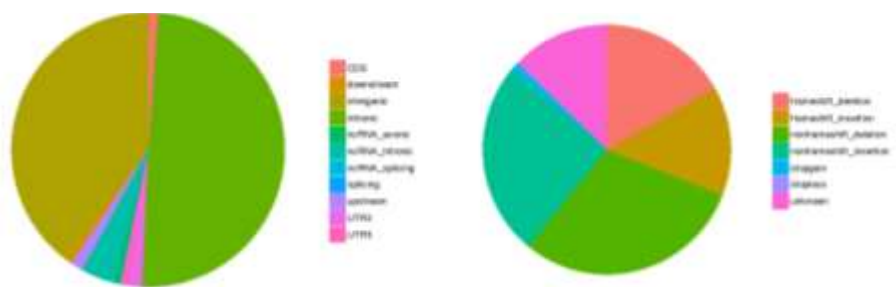


شکل ۴. فراوانی SNP‌ها در مناطق مختلف ژنوم (طرح سمت چپ)؛ فراوانی انواع مختلف SNP در منطقه کدکننده (طرح سمت راست). CDS، اشاره به تعداد SNP‌ها در منطقه کدکننده دارد.

جدول ۲. فراوانی InDel‌های مشاهده شده

تعداد	انواع InDel	
۱۱۹	frameshift_deletion	مناطق کدکننده
۱۰۱	frameshift_insertion	
۲۱۳	nonframeshift_deletion	
۱۷۹	nonframeshift_insertion	
۸	stopgain	
۱	stoploss	مناطق غیر کدکننده
۹۰	ناشناخته	
۲۸۶۹۳	intronic	
۹۳۰	UTR3	
۵۳۷	UTR5	
۸۶	splicing	
۳۶۳	ncRNA_exonic	
۲۳۲۴	ncRNA_intronic	
۴	ncRNA_splicing	
۸۵۲	upstream	
۵۰۱	downstream	
۲۳۹۷۱	Intergenic	
۵۷۹۷۲		

از مجموع ۵۷۹۷۲ جهش InDel، ۳۹۷۷۹ نوع یعنی ۶۸/۵۶ درصد قبلاً در پایگاه‌های داده مرتبط گزارش شده‌اند و ۱۸۲۴۰ نوع از این تغییرات جدید می‌باشند و تاکنون در هیچ پایگاه داده عمومی مربوط به جهش InDel گزارش نشده‌اند. علاوه بر این تعداد InDel‌ها در مناطق مختلف ژنوم و نیز به‌طور



شکل ۵. مقایسه فراوانی *InDel* ها در مناطق مختلف ژنوم (چپ)؛ فراوانی انواع مختلف *InDel* در منطقه کدکننده (راست). *CDS*، اشاره به تعداد *InDel* ها در منطقه کدکننده دارد

جدول ۳. ژن‌های *SNP* یافته تأیید شده در فهرست ژن‌های شناخته شده سرطان پستان

نام پایگاه‌های داده در ارتباط با سرطان				
<i>HGMD</i>	<i>HCMDB</i> **	<i>OMIM</i>	<i>gwas_Catalog</i>	<i>Cosmic</i>
<i>MIIP</i>	<i>ERBB2</i> *	<i>PTPRF</i>	<i>SLC45A1</i>	<i>MSH2</i>
<i>XRCC3</i> *	<i>TP53</i> *	<i>BRCA2</i> *	<i>MDM4</i>	<i>UBXN4</i>
<i>ZNF350</i>	<i>KISS1</i>	<i>AKT1</i>	<i>TERT</i> *	<i>GATA2</i>
<i>RRP1B</i>	<i>RHOA</i>	<i>NQO1</i>	<i>CCDC170</i>	<i>HGFAC</i>
<i>CCDC22</i>	<i>TNFRSF11B</i>	<i>TP53</i> *	<i>LSP1</i>	<i>HLA-B</i>
<i>EPCAM</i>	<i>MMP9</i>	<i>RAD54L</i>	<i>PLXNC1</i>	<i>EGFR</i> *
<i>CASP8</i> *	<i>PTGS2</i>	<i>CASP8</i> *	<i>RAD51B</i>	<i>KMT2C</i>
<i>BARD1</i> *	<i>VEGFC</i>	<i>BARD1</i> *	<i>COL1A1</i>	<i>IFNA4</i>
<i>NEK10</i>	<i>VEGFA</i>	<i>PIK3CA</i>		<i>UBAP2</i>
<i>ATXN7</i>	<i>TGFB1</i>	<i>CDS1</i>		<i>FAM205A</i>
<i>BMPR1B</i>	<i>ESR1</i> *	<i>PRLR</i>		<i>CTSL</i>
<i>TERT</i> *	<i>MUC1</i>	<i>HMMR</i>		<i>C9orf129</i>
<i>DROSHA</i>	<i>BRMS1</i>	<i>NQO2</i>		<i>OR13C5</i>
<i>ATG10</i>	<i>EGFR</i> *	<i>ESR1</i> *		<i>ORM1</i>
<i>LOX</i>	<i>MMP1</i>	<i>RB1CC1</i>		<i>TNFSF15</i>
<i>EHMT2</i>	<i>TNFSF11</i>	<i>TSG101</i>		<i>RAPGEF1</i>
<i>DMBT1</i>	<i>HIF1A</i>	<i>ATM</i>		<i>DBH</i>
<i>BRCA2</i> *	<i>CDH1</i> *	<i>KRAS</i>		<i>C9orf62</i>
<i>FANCA</i>	<i>SPP1</i>	<i>XRCC3</i> *		<i>TUBBP5</i>
<i>BRIP1</i> *	<i>CD44</i>	<i>KLC1</i> *		<i>FGFR2</i>
<i>XRCC1</i>	<i>MMP2</i>	<i>PALB2</i>		<i>HNF1A</i>
<i>MIR499A</i>	<i>ITGAV</i>	<i>AR</i>		<i>ABHD12B</i>
<i>MIR499B</i>	<i>CTNNB1</i>	<i>CDH1</i> *		<i>PIGQ</i>
<i>KLC1</i> *	<i>SYK</i>	<i>BRCA1</i> *		<i>CREBBP</i>
<i>TMPRSS3</i>	<i>CXCL12</i>	<i>PHB</i>		<i>SHCBP1</i>
	<i>MKI67</i>	<i>RAD51C</i>		<i>CYLD</i>
	<i>PEBP1</i>	<i>PPM1D</i>		<i>TP53</i> *
	<i>MMP14</i>	<i>BRIP1</i> *		<i>ERBB2</i> *
	<i>BRCA1</i> *	<i>BACH1</i>		<i>ANKRD62</i>
	<i>NME1</i>	<i>CHEK2</i>		<i>PPP2R3B</i>
	<i>TNFRSF11A</i>			
	<i>SRC</i>			

* ژن‌های *SNP* یافته که در بیش از یک پایگاه داده تأیید می‌شوند.

** ژن‌های *SNP* یافته که در بیشتر از چهار مطالعه گزارش شده در پایگاه داده *HCMDB* تأیید می‌شوند.

علاوه بر این غنی‌سازی ژن‌های دارای *SNP* به مسیرهای بیوشیمیایی مرتبط با سرطان در پایگاه داده *KEGG* انجام شد. نتایج نشان می‌دهد که ۳۳۵ ژن از مجموع ۱۹۳۲۷ ژن در مسیرهای مرتبط با سرطان نقش دارند.

نتایج نشان می‌دهد که در نواحی غیراگزونیک *DQAI* و *SYNE1* مشاهده می‌شود. همچنین فایل حاشیه‌نویسی حاوی اطلاعات غنی‌سازی داده‌های *InDel* در پایگاه‌های داده عمومی مرتبط با سرطان نیز می‌باشد که ژن‌های تأیید شده در این پایگاه‌های داده که در سرطان پستان شناخته شده هستند در جدول ۴ آمده‌اند.

نتایج نشان می‌دهد که در نواحی غیراگزونیک نزدیک به مرز، *InDel* های بیشتری رخ داده‌اند و در مقایسه با نواحی اگزونی، فراوانی جهش در هر ژن نیز بسیار بیشتر است. از بین آنها بیشترین فراوانی در ژن‌های *RBFOX1*، *DMD*، *HLA-DRB1*، *CSMD1*، *HLA-DQB1*، *RBFOX3*، *HLA-*

جدول ۴. ژن‌های *InDel* یافته تأییدشده در فهرست ژن‌های شناخته‌شده سرطان پستان

نام پایگاه‌های داده در ارتباط با سرطان					
<i>HGMD</i>	<i>HCMDDB**</i>	<i>OMIM</i>	<i>Cosmic</i>		
<i>LHCGR</i>	<i>ERBB2</i>	<i>PTPRF</i>	<i>HNF1A</i>	<i>FAM71E2</i>	<i>MST1L</i>
	<i>MMP9</i>	<i>CASP8</i>	<i>ORAI1</i>	<i>ZNF772</i>	<i>LOR</i>
	<i>PTGS2</i>	<i>BARD1</i>	<i>WDR66</i>	<i>ZNF814</i>	<i>KISS1*</i>
	<i>VEGFC</i>	<i>PIK3CA</i>	<i>NCOR2</i>	<i>SIRPA</i>	<i>ADRA2B</i>
	<i>TGFB1</i>	<i>CDS1</i>	<i>EP400</i>	<i>RALY</i>	<i>FMNL2</i>
	<i>ESR1*</i>	<i>NQO2</i>	<i>OR4L1</i>	<i>CTSA</i>	<i>SCRN3</i>
	<i>BRMS1</i>	<i>ESR1*</i>	<i>RBM23</i>	<i>MED15</i>	<i>HOXD9</i>
	<i>EGFR</i>	<i>RB1CC1</i>	<i>HOMEZ</i>	<i>BAIAP2L2</i>	<i>PRKRA</i>
	<i>CDH1*</i>	<i>TSG101</i>	<i>NOP9</i>	<i>SYNGR1</i>	<i>FAM171B</i>
	<i>TP53*</i>	<i>ATM</i>	<i>DDHD1</i>	<i>SHROOM4</i>	<i>CCR5</i>
	<i>MMP2</i>	<i>BRCA2</i>	<i>GOLGA8R</i>	<i>NUDT11</i>	<i>GPX1</i>
	<i>KISS1*</i>	<i>XRCC3</i>	<i>LRRC49</i>	<i>CACNA1B</i>	<i>KIAA1211</i>
	<i>ITGAV</i>	<i>AKT1</i>	<i>WDR73</i>	<i>KAT6B</i>	<i>MAML3</i>
	<i>RHOA</i>	<i>CDH1*</i>	<i>FAM174B</i>	<i>PHRF1</i>	<i>MSH3</i>
	<i>SYK</i>	<i>NQO1</i>	<i>MEF2A</i>	<i>DNHD1</i>	<i>KCNN2</i>
	<i>CXCL12</i>	<i>TP53*</i>	<i>IL32</i>	<i>FNBP4</i>	<i>PCDHB10</i>
	<i>MMP14</i>	<i>PHB</i>	<i>RBBP6</i>	<i>TMPRSS13</i>	<i>GPRIN1</i>
	<i>TNFRSF11A</i>	<i>RAD51C</i>	<i>MAZ</i>	<i>ZNF384</i>	<i>HLA-A</i>
	<i>SRC</i>	<i>PPM1D</i>	<i>ASPHD1</i>	<i>FAM90A1</i>	<i>CDSN</i>
		<i>BRIP1</i>	<i>THAP11</i>	<i>A2ML1</i>	<i>HLA-B</i>
		<i>BACH1</i>	<i>PDXDC2P</i>	<i>DDX12P</i>	<i>HLA-DRB1</i>
		<i>CHEK2</i>	<i>PKD1L2</i>	<i>PHLDA1</i>	<i>AK9</i>
		<i>AR</i>	<i>PIEZO1</i>	<i>KANK3</i>	<i>MAP3K4</i>
			<i>BCL6B</i>	<i>KMT2B</i>	<i>TBP</i>
			<i>RAI1</i>	<i>NR1H2</i>	<i>TMEM184A</i>
			<i>USP36</i>	<i>LENG9</i>	<i>EPDR1</i>
			<i>SLC38A10</i>	<i>MAFA</i>	<i>STAG3L2</i>
			<i>TSPAN10</i>	<i>LURAP1L</i>	<i>ZAN</i>
			<i>HDGFRP2</i>	<i>USP20</i>	<i>TMEM229A</i>
					<i>IRF5</i>

* ژن‌های *InDel* یافته که در بیش از یک پایگاه داده تأیید می‌شوند.

** ژن‌های *InDel* یافته که در بیشتر از چهار مطالعه گزارش شده در پایگاه داده *HCMDDB* تأیید می‌شوند.

افزون بر این، ۲۴۴ ژن دارای جهش از نوع *InDel* به مسیرهای بیوشیمیایی مرتبط با سرطان در پایگاه داده *KEGG* غنی‌سازی می‌شوند و بدین ترتیب دخالت آنها در مسیر سرطان تأیید می‌شود.

بحث و نتیجه‌گیری

نیاز است تا این اطلاعات برای هر نژاد، منطقه و هر کشور بومی‌سازی شوند. علاوه بر این فراوانی داده‌های موجود برای سرطان‌های تهاجمی و متاستاز یافته کمتر است و همین امر شناخت ما را نسبت به این مرحله از بیماری محدودتر کرده‌است. با توجه به این شواهد و در یک رویکرد کمینه‌گرا، ما ۴۰ نمونه بافت پارافینه بایگانی‌شده را از بیماران مبتلا به سرطان پستان تهاجمی با هویت ناشناس جمع‌آوری کردیم. در گام نخست با توجه به هزینه‌های بالای استخراج *DNA* و ترادفیابی کل اگزوم، تنها نمونه یک بیمار زن ۴۴ ساله مبتلا به سرطان پستان تهاجمی شده ترادفیابی و ارزیابی شد که این موضوع از محدودیت‌های این

امروزه داده‌های بسیاری در سطح ژنوم در پایگاه‌های داده از جمله (<https://portal.gdc.cancer.gov>) *TCGA* و (<http://cbioportal.org>) *cBio* *Cancer Genomics Portal* در مورد بیماران مبتلا به سرطان وجود دارد که کمک بسیاری به روند درمان و انتخاب داروهای هدف و مناسب برای بیماران کرده‌است. با این وجود، سرطان بیماری پیچیده‌ای است که در نژادهای مختلف، الگوی ژنتیکی بروز آن تا حدودی متفاوت است، به طوری که نژاد و شیوه زندگی همواره بر شیوع و میزان پراکنش هر سرطان در نقاط مختلف جهان تأثیرگذار بوده است. بنابراین

شناخته‌شده محرک سرطان تأیید می‌شوند. با توجه به ایرانی‌بودن مورد تحت بررسی، اطلاعات به‌دست‌آمده می‌تواند به غنی‌سازی داده‌های موجود در پایگاه‌های داده ژنتیک بومی کمک نماید. در صورتی‌که نمونه‌های جمع‌آوری شده از ۳۹ بیمار دیگر که همگی مبتلا به سرطان تهاجمی پستان بوده‌اند نیز ترادف‌یابی شوند و تحت مطالعه و تجزیه و تحلیل دقیق قرار گیرند به درمان هدفمند بیماران بومی کمک شایانی خواهد کرد و داده‌های مربوط به این سرطان بومی‌سازی خواهند شد. در پرتو چنین مطالعاتی است که جنبه‌هایی از این بیماری که در حال حاضر مشخص نیست آشکار می‌شود و امید به زندگی بیماران مبتلا افزایش خواهد یافت. این نکته مهم است که ما در این مطالعه تنها بر مطالعه در سطح آگروم متمرکز شده‌ایم، در حالی‌که مطالعه و یکپارچه‌سازی داده‌های اومیکس در سطوح مختلف شامل اپی‌ژنوم، ترنسکرپتوم، پروتئوم، متابولوم و میکروبیوم نیاز است (Griffith & Gray, 2011) تا ابعاد این بیماری از زوایای مختلف مطالعه شود و این کاستی از معایب و محدودیت‌های دیگر این مطالعه است. ما امیدوار هستیم با ترادف‌یابی نمونه‌های بیشتر و به‌کارگیری ابزارهای کارآمد در حیطه دانش بیوانفورماتیک و زیست‌شناسی محاسباتی مثل یادگیری ماشین و همچنین یکپارچه‌سازی داده‌ها در سطوح مختلف اومیکس بتوانیم یک پانل چند ژنی هدفمند در رابطه با بیماران ایرانی مبتلا به سرطان پستان تهاجمی شده پیشنهاد نماییم تا هزینه‌های ترادف‌یابی کل ژنوم/آگروم در مطالعات بعدی حذف شوند و نقشه راه بهینه‌ای در اختیار متخصصین حیطه‌ی آنکولوژی دقیق قرار گیرد. این مطالعه با به‌کارگیری رویکرد یادگیری ماشینی گروهی بر روی داده‌های صدها بیمار مشابه که اطلاعات آنها در *TCGA* و *cBioPortal* موجود بودند انجام شده است (Mirsadeghi et al., 2021) و در صورت جمع‌آوری شدن داده‌های بیشتر برای بیماران ایرانی نیز قابل انجام است.

پژوهش محسوب می‌شود. با این وجود، *NGS* امکان ترادف‌یابی همزمان ژن‌ها را فراهم می‌کند و در مقایسه می‌تواند برای درمان هدفمند بیمار کارآمدتر باشد و در کل هزینه و آسیب کمتری نسبت به سایر روش‌های تشخیصی که در آنها از آزمایشات پی‌درپی استفاده می‌شود خواهد داشت (Tung et al., 2015) که از مزایای این روش است. از طرف دیگر، این تکنیک داده‌های فراوانی را در اختیار قرار می‌دهد. در این مطالعه، استفاده از دانش بیوانفورماتیک کمک شایانی در تجزیه و تحلیل داده‌های انبوه به‌دست‌آمده از *NGS* نمود که بدون استفاده از ابزارهای به‌کارگرفته شده، دستیابی به فایل حاشیه‌نویسی و تحلیل جهش‌ها غیرممکن بود. از سوی دیگر، شواهد به‌دست‌آمده از تجزیه و تحلیل فایل حاشیه‌نویسی آگروم نمونه تحت بررسی نشان می‌دهد که ژن‌های خانواده *HLA* از تنوع ژنتیکی بسیار زیادی برخوردار هستند که با توجه به گزارش‌های موجود، در جمعیت انسانی طبیعی است. به نظر می‌رسد از بین ژن‌های این خانواده فراوانی تنوع آلی در *HLA-DRB1* بیشتر است که در مقایسه با مطالعات دیگر، شواهد مشابهی به‌دست آمد (Aureli et al., 2020). علاوه‌براین، یافت‌شدن *SNP* ها و *InDel* های جدید در آگروم بیمار تحت بررسی تامل‌برانگیز است و در صورتیکه در سایر نمونه‌های مربوط به جمعیت ایرانی تکرار شود نیاز به مطالعه و بررسی بالینی دقیق ایجاد می‌شود. امروزه استفاده از رویکردهایی مثل ابزارهای محاسباتی توسعه‌یافته در حیطه سرطان و یادگیری ماشین قادرند تا تعداد زیادی از نمونه‌های *WES* را همزمان ارزیابی نمایند (Liu et al., 2015). مقایسه یافته‌های این رویکردها با اطلاعات موجود در پایگاه‌های داده عمومی مرتبط با سرطان و مطالعات دیگری که داده‌های جهش را ارزیابی کرده‌اند معنی‌دارتر است. با این حال شواهد نشان می‌دهد بسیاری از ژن‌هایی که در مطالعه حاضر فراوانی جهش بالایی نشان داده‌اند در فهرست ژن‌های

REFERENCES

- Aureli, A.; Canossi, A.; Del Beato, T.; Buonomo, O.; Rossi, P.; Roselli, M.; ... Sconocchia, G. (2020). Breast Cancer Is Associated with Increased HLA-DRB1* 11: 01 and HLA-DRB1* 10: 01 Allele Frequency in a Population of Patients from Central Italy. *Immunological Investigations*; 49(5), 489-497.
- Chai, X.; Domchek, S.; Kauff, N.; Rebbeck, T.; Chen, J. (2015). RE: Breast Cancer Risk After Salpingo-Oophorectomy in Healthy BRCA1/2 Mutation Carriers: Revisiting the Evidence for Risk Reduction. *Journal of the National Cancer Institute*; 107(9), djv217.
- Consortium, G.O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*; 32(suppl_1), D258-D261.
- Ebrahimi, M.; Vahdaninia, M.; Montazeri, A. (2015). Risk factors for breast cancer in Iran: a case-control study. *Breast Cancer Research*; 4: 4.
- Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; ... May, B. (2017). The reactome pathway knowledgebase. *Nucleic Acids Research*; 46(D1): D649-D655.
- Gambardella, V.; Tarazona, N.; Cejalvo, J.M.; Lombardi, P.; Huerta, M.; Roselló, S.; ... Cervantes, A. (2020). Personalized medicine: recent progress in cancer therapy. *Cancers*; 12(4): 1009.
- Gourraud, P.-A.; Khankhanian, P.; Cereb, N.; Yang, S.Y.; Feolo, M.; Maiers, M.; ... Oksenberg, J. (2014). HLA diversity in the 1000 genomes dataset. *PloS One*; 9(7): e97282.
- Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G.L.; Hunter, C.; Bignell, G.; ... Stevens, C. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*; 446(7132): 153-158.
- Griffith, O.L.; Gray, J.W. (2011). 'Omic approaches to preventing or managing metastatic breast cancer. *Breast Cancer Research*; 13(6): 230.
- Jiang, Y.; Jiang, Y.; Wang, S.; Zhang, Q.; Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinformatics*; 20(1): 1-12.
- Kanehisa, M.; Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*; 28(1): 27-30.
- Kumar, A.; Singla, A. (2019). Epidemiology of Breast Cancer: Current Figures and Trends. In *Preventive Oncology for the Gynecologist* (pp. 335-339). Springer.
- Liu, Y.; Tian, F.; Hu, Z.; DeLisi, C. (2015). Evaluation and integration of cancer gene classifiers: identification and ranking of plausible drivers. *Scientific Reports*, 5.
- McKenzie, A.J.H.; Dilks, H.; Jones, S.F.; Burris III, H. (2019). Should next-generation sequencing tests be performed on all cancer patients? *Expert Review of Molecular Diagnostics*; 19(2): 89-93.
- Mirsadeghi, L.; Hosseini, R.H.; Banaei-Moghaddam, A.M.; Kavousi, K. (2021). EARN: an ensemble machine learning algorithm to predict driver genes in metastatic breast cancer. *BMC Medical Genomics*; 14(1), 1-19.
- Morganti, S.; Tarantino, P.; Ferraro, E.; D'Amico, P.; Duso, B.A.; Curigliano, G. (2019). cancer. *Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics*; 9-30.
- Nishimura, D. (2001). BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient*; 2(3): 117-120.
- Schaefer, C.F.; Anthony, K.; Krupa, S.; Buchoff, J.; Day, M.; Hannay, T.; Buetow, K.H. (2009). PID: the pathway interaction database. *Nucleic Acids Research*; 37(suppl_1), D674-D679.

- Sheikine, Y.; Kuo, F.C.; Lindeman, N.I. (2017). Clinical and technical aspects of genomic diagnostics for precision oncology. *Journal of Clinical Oncology*; 35(9): 929-933.
- Smith, N.G.; Gyanchandani, R.; Shah, O.S.; Gurda, G.T.; Lucas, P.C.; Hartmaier, R.J.; ... Kota, K. (2019). Targeted mutation detection in breast cancer using MammaSeq™. *Breast Cancer Research*; 21(1): 22.
- Truty, R.; Paul, J.; Kennemer, M.; Lincoln, S.E.; Olivares, E.; Nussbaum, R.L.; Aradhya, S. (2019). Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. *Genetics in Medicine*; 21(1): 114-123.
- Tung, N.; Battelli, C.; Allen, B.; Kaldate, R.; Bhatnagar, S.; Bowles, K.; ... Ellisen, L. (2015). Frequency of mutations in individuals with breast cancer referred for BRCA 1 and BRCA 2 testing using next-generation sequencing with a 25-gene panel. *Cancer*; 121(1): 25-33.
- Wang, K.; Li, M.; Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*; 38(16): e164-e164.
- Watson, I.R.; Takahashi, K.; Futreal, P.A.; Chin, L. (2013). Emerging patterns of somatic mutations in cancer. *Nature Reviews Genetics*; 14(10): 703-718.
- Wood, L.D.; Parsons, D.W.; Jones, S.; Lin, J.; Sjöblom, T.; Leary, R.J.; ... Ptak, J. (2007). The genomic landscapes of human breast and colorectal cancers. *Science*; 318(5853): 1108-1113.