مجله علمی ـ پژوهشی زیست‌شناسی جانوری تجربی
سال دهم، شماره سوم، پیاپی سی و نهم، زمستان ۱۴۰۰ (۷۴-۶۳)

DOI: 10.30473/eab.2021.58111.1823

مقاله پژوهشی:

NGS در خدمت ارتقای پزشکی شخص‌محور: گزارش موردی یک بیمار مبتلا به سرطان پستان تهاجمی

لیلا میرصادقی[۱]، رضا حاجی‌حسینی[۲]، علی محمد بنایی مقدم[۳]، کاوه کاوسی[۴*]

۱. دکتری، گروه بیولوژی، دانشگاه پیام نور، تهران، ایران
۲. استاد، گروه بیولوژی، دانشگاه پیام نور، تهران، ایران
۳. استادیار، آزمایشگاه ژنومیکس و اپی‌ژنومیکس (LGE)، گروه بیوشیمی، مرکز تحقیقات بیوشیمی و بیوفیزیک (IBB)، دانشگاه تهران، تهران، ایران
٤. 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: ۱٤۰۰/۷/۲۵)

چکیده

ظهور پزشکی شخص‌محور با تکیه‌بر 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 می دهد. 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 بایگانی‌شده از بافت تثبیت شده با فرمالین و جاسازی شده در پارافین مربوط به یک بیمار زن ایرانی مبتلا به سرطان پستانِ 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 شخص‌محور، ترادف‌یابی نسل بعدی، داده‌های جهش، کارسینومای پستان تهاجمی‌شده.

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

Leila Mirsadeghi[1], Reza Haji Hosseini[2], Ali Mohammad Banaei-Moghaddam[3], Kaveh Kavousi[4*]

1. Ph. D., Department of Biology, Payame Noor University, Tehran, Iran
2. Professor, Department of Biology, 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)

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.

* نویسنده مسئول: کاوه کاوسی Email: kkavousi@ut.ac.ir

مقدمه

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 این‌حال 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 پستانِ 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). متاسفانه 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 زیست نشانگرهای مولکولی در هر سرطان جهت پیش‌آگهی و درمان زودهنگام، همواره مورد توجه سرطان شناسان بوده‌است. متخصصان آنکولوژی برای شناسایی این نشانگرها به داده‌های فراوان نیاز دارند (McKenzie *et al*., 2019). امروزه، تمرکز بر تکنیک مولکولی 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 زودهنگام و در مراحل اولیه غربالگری شوند. این تغییر چشم‌انداز که در آن توجه متخصصین به تغییرات خاص مولکولی تومور معطوف است، دریچه‌ای برای درمانِ 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

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 آنکوژنز 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 جهش‌های محرک و کمی‌سازی اطلاعات، اساس یک رویکرد جدید در مقابله با سرطان را ایجاد کرده‌است (Morganti *et al*., 2019). طی دو دهه‌ی اخیر، پیشرفت‌های بسیاری در تکنیک‌های مولکولی مرتبط با عرصه تشخیص‌های بالینی انجام شده است. پزشکی شخص‌محور با تکیه بر به خدمت‌گرفتن چنین تکنیک‌هایی و همچنین بر پایه وجود تنوع ژنتیکی بین افراد و نژادهای مختلف توانسته است روند و عملکردِ درمان بیماران را بهبود بخشد. پیش‌بینی شده است که موجی از درمان‌های آینده مربوط به سرطان، به‌کمک پزشکی شخص‌محور امکانپذیر خواهد شد (Sheikine *et al*., 2017). امروزه مطابق با پروتکل‌های شورای تحقیقات ملی آمریکا (*NRC*) که زیر مجموعه‌ی اینستیتو ملی سلامت (*NIH*) است بهتر است از عبارت پزشکی دقیق (Precision medicine) به جای پزشکی شخص‌محور استفاده شود اگرچه که همچنان در جوامع پزشکی و علمی از این دو اصطلاح به جای هم استفاده می‌شود و همپوشانی بسیاری بین مفهوم این دو عبارت وجود دارد. در حیطه سرطان، پزشکی دقیق سرطان شناسی دقیق (Precision oncology) خوانده می‌شود و هدف آن یافتن ژن‌های کلیدی هدفمند و نشانگر برای هر سرطان است تا به جای ترادف یابی کل ژنوم، ترادف‌یابی پانل ژنی هدفمند (Targeted gene panel sequencing) انجام شود تا هزینه‌ها کاهش و سرعت تشخیص و تحلیل نتایج برای هر بیمار بهبود یابد (Sheikine *et al*., 2017). مطالعات مختلف شواهدی از خدمات پزشکی شخص‌محور و دقیق در پیشگیری و درمان بیماران را ارائه داده‌اند. در یک پژوهش انجام شده روی زنانِ دارای جهش در ژن‌های *BRCA1* و *BRCA2* که خود به‌عنوان ژن‌های سرکوبگر تومور هستند، نرخ ابتلا به سرطان پستان سه برابر گزارش شده است (Chai *et al*., 2015). بنابراین با اقدام زودهنگام جهت ترادف‌یابی ژنومِ افراد، به‌ویژه کسانی‌که دارای زمینه‌های خانوادگی در بروز سرطان پستان می‌باشند می‌توان افراد در ریسک را پیش از موعد در میان انبوه جمعیت غربالگری نمود تا تحت درمان‌های پیشگیرانه مؤثر قرار گیرند. در یک مطالعه دیگر مبتنی بر *NGS* بر روی ۱۷۸۱ شرکت‌کننده، فراوانی جهش در ۲۵ ژن بررسی شد. یافته‌ها نشان می‌دهد که تقریباً ۶۶ درصد جهش‌های ژنی این افراد مربوط به ژن‌های *BRCA1* و *BRCA2* است. شواهد بالینی وجود دارد که در سرطان‌هایی که نقص در این دو ژن فراوانی بیشتری داشته‌اند، مبتلایان، به درمان با مهارکننده‌های پلاتین و پلی*ADP*–ریبوز پلیمراز–۱ پاسخ خوبی داده‌اند. این مهارکننده‌ها می‌توانند به‌عنوان داروی هدف برای این بیماران استفاده شوند (Tung *et al*., 2015). در سال ۲۰۱۹ و مبتنی بر اطلاعات به‌دست‌آمده از *NGS*، ۷۹ ژن در شناسایی بیماران مبتلا به سرطان پستان اولیه و سرطان پستان ثانویه ارزیابی و تأیید شدند. این ژن‌ها می‌توانند 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 نمونه‌های بایگانی شده بودند، انتخاب و به‌طور موردی بررسی شد. بدین ترتیب که پس از استخراج و تخلیص *DNA* از بافت پستانِ تثبیت‌شده در فرمالین و جاسازی‌شده در پارافین (*FFPE*)، ترادف‌یابی کل اگزوم *DNA* انجام شد و جهش‌های رخ‌داده در آن مورد تجزیه و تحلیل بیوانفورماتیکی قرار گرفت.

مواد و روش‌ها

از میان ۴۰ نمونه *FFPE* به‌دست‌آمده از بافت پستان

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 ۱×۱/۵×۱/۸ 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.

استخراج و خالص‌سازی *DNA*

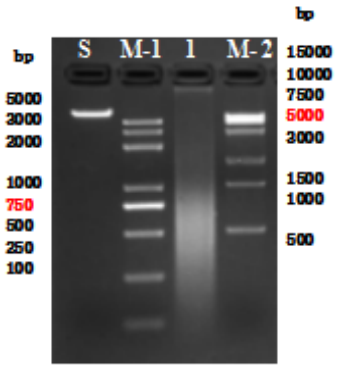
در مطالعه حاضر، فرایند استخراج و خالص‌سازی DNA ی ژنومیک از بافت جاسازی شده در پارافین به‌کمک *QIAamp DNA FFPE Tissue Kit* انجام شد.

کنترل کیفی *DNA* استخراج شده و خلوص آن

در ادامه خلوص و یکپارچگی *DNA*ی تخلیص‌شده با آزمون کنترل کیفی (*QC*) و به‌کمک تکنیک الکتروفورز ژل آگارز ارزیابی و پایش شد. به‌طور جداگانه، الکتروفورز برای *DNA*ی ژنومیک و محصول *PCR* به‌ترتیب بر روی ژل آگارز ۱ و ۲ درصد انجام شد. در فرایند استخراج، ۴/۲۸۰ میکروگرم *DNA* با قطعاتی بین ۵۰۰ تا ۱۰۰۰جفت باز به‌دست آمده است. در مجموع، نتیجه آزمایش براساس معیارهای *QC*، توضیح می‌دهد که نمونه حاضر می‌تواند ملزومات ساخت کتابخانه برای ترادف‌یابی کل اگزوم را برآورده سازد (شکل ۱).

آماده‌سازی کتابخانه برای ترادف‌یابی

برای تهیه کتابخانه *DNA*، از کل ۴/۲۸۰ 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 ۲۸۰-۱۸۰ جفت باز ایجاد شدند. در ادامه، خوشه‌هایی از قطعات *DNA* تولید و میلیون‌ها نسخه از *DNA*ی 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 سایز قطعات ارزیابی شود.



S: نمونه استاندارد
M-1: Trans 2k plus DNA ladder
1: نمونه مورد مطالعه*
M-2: Trans 15k plus DNA ladder
*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.

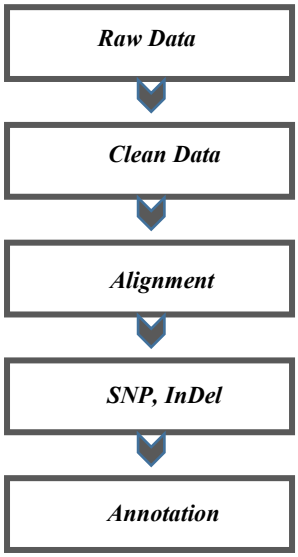
شکل ۱. نتاج الکتروفورز مشاهده‌شده روی ژل آگارز. شرح علائم اختصاری در کادر راهنما آمده است.

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 *HiSeq PE150*، ترادف‌یابی شد. روند خوانش تا جایی ادامه می‌یابد که کل اگزوم ترادف‌یابی شود. در مطالعه حاضر، عمق ترادف‌یابی ۲۰۰*X* 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 شده‌است (شکل ۲).



شکل ۲. روندکاری مراحل تجزیه و تحلیل بیوانفورماتیکی

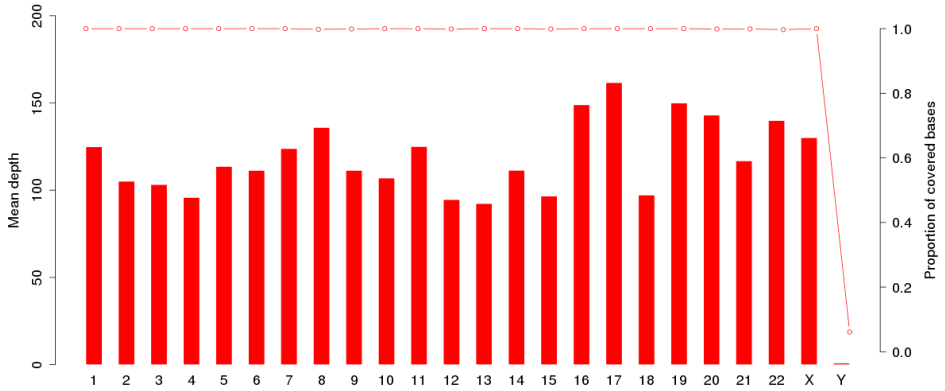
در مرحله تهیه فایل حاشیه‌نویسی برای واریانت‌ها، از روش ANNOVAR استفاده شده است (Wang & Hakonarson, 2010)، تا جنبه‌های مختلفی از داده‌های جهش، از جمله تغییرات در کدگذاری پروتئین‌ها آشکار شوند.

ارزیابی

در این گام برای یافتن مناطق ژنومی تحت‌تأثیر که تغییرات پروتئینی نشان می‌دهند، از پایگاه‌های داده *RefSeq* و *Gencode* استفاده شده‌است. برای پیش‌بینی مخرب بودن جهش‌ها، ابزارهای *SIFT*، *PolyPhen*، *MutationAssessor*، *LRT* و *CADD* به‌کار گرفته شدند. همچنین پایگاه‌های داده *dbSNP*، *COSMIC*، *OMIM*، *GWAS Catalog* و *HGMD* استفاده شدند تا بتوانیم اطلاعات موجود گزارش شده مرتبط با گستره ژنوم و برخی از انجمن‌های سرطان را با داده‌های جهش به‌دست‌آمده از اگزوم بیمارِ مورد مطالعه مقایسه کنیم. در فایل حاشیه‌نویسی، برای پیداکردن مسیرهای بیوشیمیایی مرتبط با هر ژن از داده‌های موجود در پایگاه‌های داده *Gene Ontology* (Consortium, 2004)، *KEGG* (Kanehisa & Goto, 2000)، *Reactome* (Fabregat *et al*., 2017)، *Biocarta* (Nishimura, 2001) و PID (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، نگاشت ۷۳۸۷۲۴۶۸ قطعه به ژنوم مرجع انجام شد. بر این اساس، داده‌های متعدد مربوط به ترادف‌یابی کل اگزوم به‌دست آمد که ما به اختصار نتایج را در سطح چندشکلی‌های تک نوکلئوتیدی (*SNPs*) و اضافه و حذف‌ها (*InDels*) ی رخداده در ژن‌ها بررسی و نیز اطلاعاتی را برمبنای تحلیل‌های بیوانفورماتیکی به‌دست‌آمده 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

*SNP*های تشخیص داده‌شده

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.

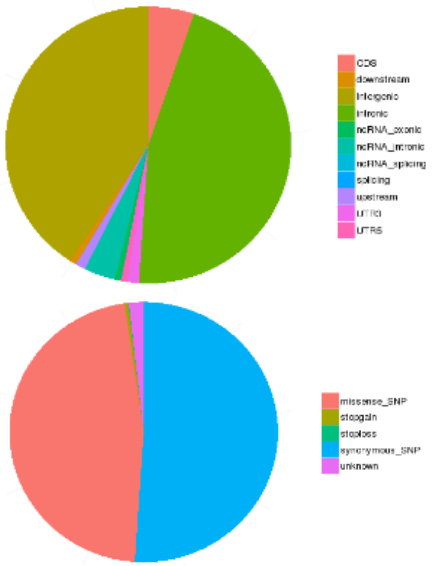
جدول ۱. فراوانی *SNP*های مشاهده‌شده

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

از این تعداد، ۲۲۰۵۷ پلی‌مورفیسم مربوط به نواحی اگزونیک است. علاوه بر این، ۶۱/۰۶ درصد *SNP*ها از پیش در پایگاه‌های داده مرتبط با *SNP* گزارش شده‌اند و بقیه جدید هستند. باید توجه داشت اگرچه ترادف‌یابی 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 می دهند (Truty *et al.*, 2019). فراوانی *SNP*ها در مناطق مختلف ژنوم و مناطق کدکننده نمونه مورد مطالعه در نمودار دایره‌ای مقایسه شده‌اند (شکل ۴).

*InDel*های 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 شناسایی‌شده در نمونه بیمار تحت مطالعه در جدول ۲ لیست شده است.



شکل ٤. 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

	انواع *InDel*	تعداد
مناطق کدکننده	frameshift_deletion	۱۱۹
	frameshift_insertion	۱۰۱
	nonframeshift_deletion	۲۱۳
	nonframeshift_insertion	۱۷۹
	stopgain	۸
مناطق غیر کدکننده	stoploss	۱
	ناشناخته	۹۰
	intronic	۲۸۶۹۳
	UTR3	۹۳۰
	UTR5	۵۳۷
	splicing	۸۶
	ncRNA_exonic	۳۶۳
	ncRNA_intronic	۲۳۲۴
	ncRNA_splicing	۴
	upstream	۸۵۲
	downstream	۵۰۱
	Intergenic	۲۲۹۷۱
		۵۷۹۷۲

از مجموع ۵۷۹۷۲ جهش *InDel*، ۳۹۷۷۹ نوع یعنی ۶۸/۵۶ درصد قبلاً در پایگاه‌های داده مرتبط گزارش شده اند و ۱۸۲۴۰ نوع از این تغییرات جدید می‌باشند و تاکنون در هیچ پایگاه داده عمومی مربوط به جهش *InDel* گزارش نشده‌اند. علاوه‌براین تعداد *InDel*ها در مناطق مختلف ژنوم و نیز به‌طور منحصربه‌فرد در مناطق کدکننده در یک نمودار دایره‌ای ترسیم و مقایسه شده‌اند (شکل ۵).

حاشیه‌نویسی واریانت‌ها

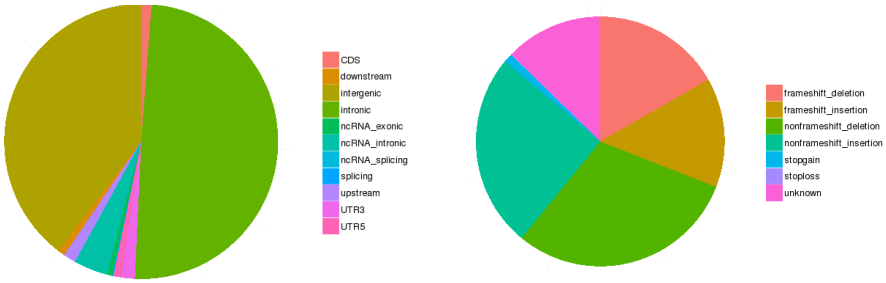
ارزیابی فایل‌های حاشیه‌نویسی مربوط به داده‌های *SNP* و *InDel* بخشی دیگر از تجزیه و تحلیل بیوانفورماتیکی این مطالعه است.

ارزیابی داده‌های *SNP*

در این مطالعه، از بین ژن‌هایی که در ناحیه اگزونیک، پلی‌مورفیسم تک نوکئوتیدی نشان دادند، پس از حذف *synonymous_SNP*ها که در آن هر دوی آلل‌ها ترادف یکسانی از پلی‌پپتید را تولید می‌کنند، ده ژن *AHNAK2*، *HLA-DRB1*، *MUC16*، *HLA-A*، *HLA-C*، *HLA-DQA1*، *PDE4DIP*، *HLA-B*، *HLA-DRB5*، و *HLA-DQB1* بیشترین پلی‌مورفیسم تک نوکلئوتیدی را نشان دادند. این نتیجه در مورد *HLA*ها با توجه به این‌که بیشترین تنوع را در گستره‌ی ژنوم دارند بدیهی است (Gourraud *et al*., 2014). در نواحی غیراگزونیک، بیشترین *SNP*ها در *RBFOX1*، *CSMD1*، *HLA-DRB1*، *RYR2*، *HLA-DQB1*، *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-DRB1*، *HLA-B*، *ZNF595*، *SARM1*، *HLA-DRB5*، *HLA-DQA1* و *PABPC3* بیشتر است.



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 تأییدشده در فهرست ژن‌های شناخته‌شده سرطان پستان

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

* 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.

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

نتایج نشان می دهد که در نواحی غیراگزونیکِ 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 *HLA-DRB1*، *DMD*، *RBFOX1*، *CSMD1*، *RBFOX3*، *HLA-DQB1*، *HLA-DQA1*، 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 شناخته‌شده هستند در جدول ۴ آمده‌اند.

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

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

* 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 *HCMDB*، تأیید می‌شوند.

افزون‌بر این، ۲۴۴ ژنِ 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

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، به‌طوری‌که که نژاد و شیوه زندگی همواره بر شیوع و میزان پراکنش هر سرطان در نقاط مختلف جهان تأثیرگذار بوده است. بنابراین نیاز است تا این اطلاعات برای هر نژاد، منطقه و هر کشور بومی‌سازی شوند. علاوه‌براین فراوانی داده‌های موجود برای سرطان‌های تهاجمی و متاستازیافته 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

پژوهش محسوب می‌شود. با این وجود، *NGS* امکان ترادف‌یابی همزمان ژن‌ها را فراهم می کند 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 به کارگرفته شده، دستیابی به فایل حاشیه‌نویسی و تحلیل جهش‌ها غیرممکن بود. از سوی دیگر، شواهد به‌دست‌آمده از تجزیه و تحلیل فایل حاشیه‌نویسی اگزوم 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 آللی 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 اگزوم بیمارِ 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 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 اگزوم متمرکز شده‌ایم، در حالی‌که مطالعه و یکپارچه‌سازی داده‌های اومیکس در سطوح مختلف شامل اپی‌ژنوم، ترنسکریپتوم، 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/اگزوم در مطالعات بعدی حذف شوند و نقشه راه بهینه‌ای در اختیار متخصصینِ 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 جمع‌آوری‌شدن داده‌های بیشتر برای بیماران ایرانی نیز قابل انجام است.

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.