



بسمه تعالی  
طرح تحقیق پایان نامه کارشناسی ارشد

کد ملی : 1292119306

نام و نام خانوادگی دانشجو	شماره دانشجویی	دانشکده	بخش/گروه	امضا دانشجو
بهار مهدوی	40157571337	علوم ریاضی	علوم کامپیوتر	

مشخصات اساتید راهنما و مشاور	نام و نام خانوادگی	رتبه دانشگاهی	محل خدمت	امضا و تاریخ
استاد راهنمای اصلی	دکتر منصور رزقی آهق	استادیار	دانشگاه تربیت مدرس	
استاد راهنمای دوم (در صورت نیاز)				
درصد سهم استاد راهنمای دوم:				
استاد مشاور (در صورت نیاز)	دکتر خسرو رجبی زاده			

<b>عنوان:</b> شناسایی محاسباتی امضاهای جهشی در تومورهای سرطانی فردی شناسایی امضاهای جهشی در تومورهای سرطانی فردی با استفاده از رویکردهای مبتنی بر یادگیری
<b>Title:</b> Computational identification of mutational signatures in individual cancer tumours Identifying mutational signatures in individual cancer tumours using learning-based approaches

### تعریف مساله، اهداف، و سوالات تحقیق:

پیشرفت‌های اخیر در فن‌آوری‌های توالی‌یابی DNA با کارایی بالا<sup>۱</sup>، مطالعاتی را امکان‌پذیر کرده است که هزاران ژنوم یا اگزوم سرطان کامل<sup>۲</sup> را بررسی می‌کند. توالی‌یابی کل ژنوم<sup>۳</sup>، جامعه ژنومیک سرطان را وارد قلمرو جدیدی کرده است. سرطان یک بیماری ژنومی است که در آن تکثیر کلونال<sup>۴</sup> کنترل نشده توسط تغییرات ژنومی در سلول‌های پیکری<sup>۵</sup> شروع و تقویت می‌شود (1). علیرغم این واقعیت که یک ژنوم سرطان ممکن است بین ده‌ها تا میلیون‌ها جهش پیکری را حمل کند (2، 3)، تنها زیرمجموعه کوچکی از این

<sup>1</sup> high-throughput DNA sequencing technologies

<sup>2</sup> whole cancer genomes or exomes

<sup>3</sup> whole-genome sequencing

<sup>4</sup> clonal proliferation

<sup>5</sup> somatic cells

جهش‌ها که جهش‌های "محرك"<sup>6</sup> نامیده می‌شوند، باعث گسترش نئوپلاستیک<sup>7</sup> می‌شوند (1, 4). عموماً اعتقاد بر این است که مابقی جهش‌ها تحت عنوان جهش‌های "مسافری"<sup>8</sup>، مزیت انتخابی در فرآیندهای دخیل در جهش‌زایی ایجاد نمی‌کنند (5, 6).

تجمع جهش‌های محرك در ژنوم سلول پیکری نتیجه یک یا چند فرآیند جهش‌زایی است که به طور مداوم یا متناوب در طول عمر ارگانیسم عمل می‌کند (7). چنین فرآیندهای جهش‌زا، شامل آسیب DNA<sup>9</sup> توسط ژنوتوکسین عوامل برون زا<sup>10</sup> یا درون زا<sup>11</sup>، همانندسازی معیوب DNA، مسیرهای ترمیم DNA معیوب، و ویرایش آنزیمی DNA است (8). بسیاری از این فرآیندها الگوی مشخصی از جهش‌ها را در ژنوم نشان می‌دهند که به عنوان "امضای جهشی"<sup>12</sup> شناخته می‌شود (2, 9). مفهوم امضاهای جهشی در سال 2012 به دنبال اثبات این موضوع معرفی شد که تجزیه و تحلیل همه جهش‌های جایگزینی<sup>13</sup> در مجموعه‌ای از توالی کامل ژنومی سرطان سینه می‌تواند الگوهای ثابت جهش‌زایی در سراسر تومورها که در طول تومورزایی<sup>14</sup> به وجود می‌آیند را نشان دهد (10). اگرچه امضاهای جهشی یک مفهوم نسبتاً جدید در بیولوژی سرطان هستند، اولین توصیفات انحرافات ژنومی<sup>15</sup> ناشی از یک فرآیند خاص به اوایل قرن بیستم باز می‌گردد، زمانی که اشعه ایکس برای ایجاد شکستگی کروموزوم در سلول‌های تحت تابش یافت شد (11-13). سایر پیوندهای علی بین عوامل جهش‌زا و الگوهای تغییرات پیکری نیز ایجاد شده است، مانند جابجایی‌های نوکلئوتیدی که توسط مواد سرطان‌زا<sup>16</sup> موجود در دود تنباکو ایجاد می‌شوند (14, 15). علاوه بر این، برخی از عوامل شیمی درمانی<sup>17</sup> نیز جهش‌زا هستند و ممکن است امضای جهشی خود را در ژنوم سرطان بیماران مبتلا به بدخیمی‌های ثانویه<sup>18</sup> نقش کنند (16, 17). این مثال‌ها اهمیت مطالعه الگوهای جهش پیکری را برای درک ما از مکانیسم‌های مولکولی نئوپلازی<sup>19</sup> نشان می‌دهند، که به طور بالقوه امکان کشف جهش‌زاهای جدید را فراهم می‌کند (2, 7, 8, 18).

الگوهای جهش‌های متعدد در ژنوم‌های سرطانی معمولاً روی یکدیگر قرار می‌گیرند و داده‌ها را غیرقابل درک می‌کنند. در سال 2012 L. Alexandrov، راهی برای حل ریاضی این مسئله ارائه کرد و نشان داد که الگوهای جهش از جهش‌زاهای فردی یافت شده در یک تومور را می‌توان با استفاده از یک رویکرد ریاضی به نام جداسازی منبع کور<sup>20</sup> از یکدیگر متمایز کرد (10). در سال 2013، تیم وی اولین چارچوب محاسباتی را برای رمزگشایی امضاهای جهشی از داده‌های ژنومیک سرطان منتشر کردند (19). متعاقباً، آن‌ها این چارچوب را برای بیش از هفت هزار ژنوم سرطانی به کار بردند و اولین نقشه جامع از امضاهای جهشی در سرطان انسان را ایجاد کردند (2). در حال حاضر، بیش از صد امضای جهش یافته در فهرست سرطان انسان شناسایی شده است (20-24).

امضای جهشی را می‌توان از نظر ریاضی به عنوان رابطه‌ای بین یک فرآیند جهش‌زا (معلوم یا ناشناخته) و مجموعه‌ای از انواع جهش پیکری تعریف کرد. بسیاری از دسته‌های تغییرات ژنومی (شکل 1) می‌توانند به‌عنوان ویژگی‌های<sup>21</sup> یک امضای جهشی عمل کنند، از

<sup>6</sup> driver

<sup>7</sup> neoplastic expansion

<sup>8</sup> passenger

<sup>9</sup> DNA damage

<sup>10</sup> exogenous

<sup>11</sup> endogenous

<sup>12</sup> mutational signature

<sup>13</sup> substitution mutations

<sup>14</sup> tumorigenesis

<sup>15</sup> genomic aberrations

<sup>16</sup> carcinogens

<sup>17</sup> chemotherapeutic

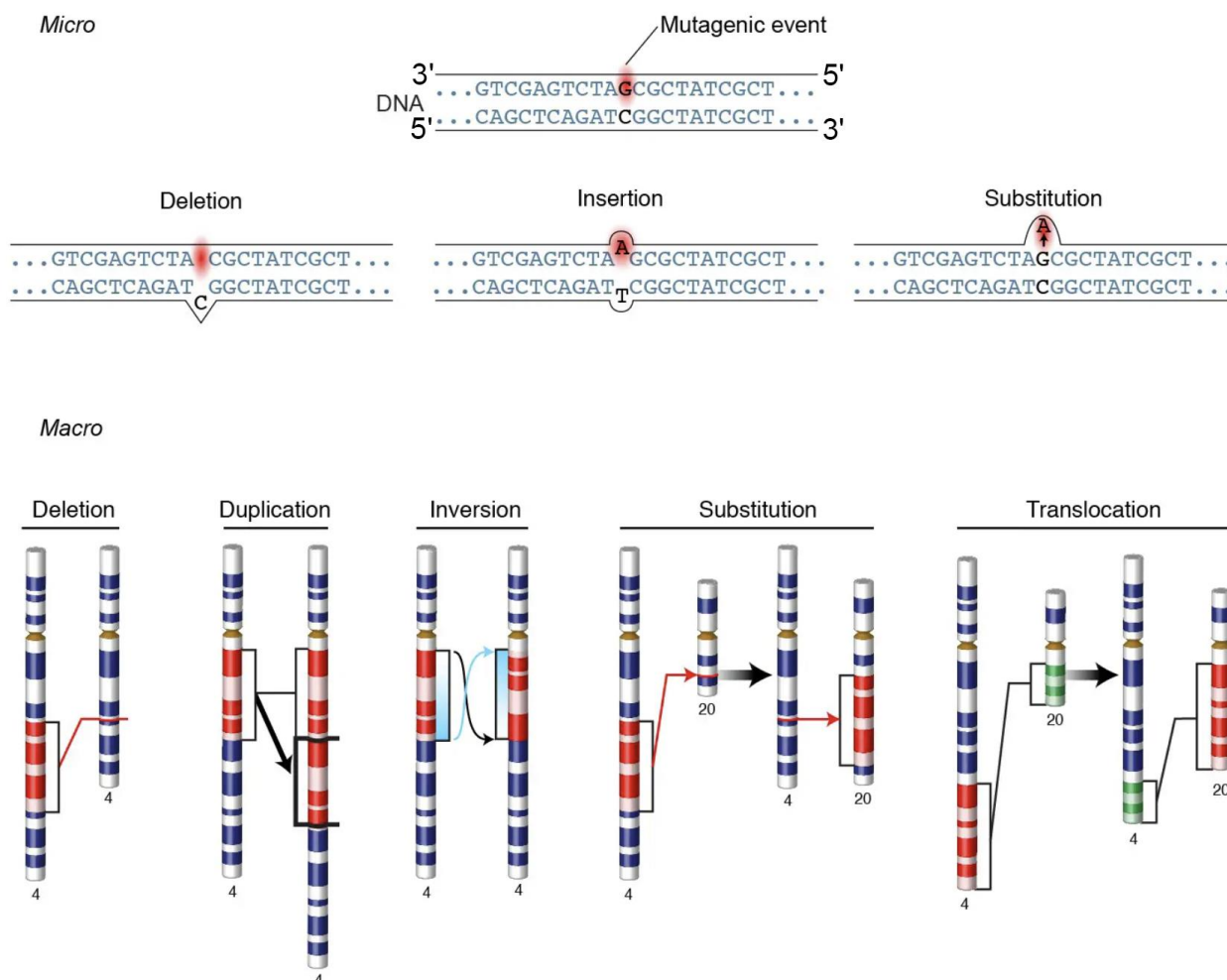
<sup>18</sup> secondary malignancies

<sup>19</sup> neoplasia

<sup>20</sup> blind source separation

<sup>21</sup> features

جمله جایگزین‌های تک بازی<sup>۲۲</sup> (SBSs) یا دوگانه (DBS)<sup>۲۳</sup>، درج‌ها و حذف‌های کوچک (این‌دِل‌ها)<sup>۲۴</sup>، تغییرات تعداد کپی<sup>۲۵</sup>، بازآرایی‌های ساختاری<sup>۲۶</sup>، رویدادهای ادغام عناصر قابل انتقال<sup>۲۷</sup>، هایپرجهش موضعی (کاتائگیس)<sup>۲۸</sup> و تغییرات اپی ژنتیکی<sup>۲۹</sup>. با توجه به اینکه بیشتر مطالعات تا به امروز بر روی جایگزینی‌های تک بازی متمرکز شده است، در عمل تنها تعداد محدودی از ویژگی‌ها را می‌توان در انتزاع ریاضی<sup>۳۰</sup> یک امضای جهشی گنجانده. با این حال، امضاهای مبتنی بر این‌دِل (25, 26) یا انواع ساختاری<sup>۳۱</sup> (25-27) و تغییرات تعداد کپی (28) نیز توصیف شده است. اگرچه ما هنوز در مراحل اولیه درک نحوه طبقه‌بندی این‌دِل‌ها، بازآرایی‌ها و تعداد کپی‌ها هستیم و مدل‌سازی دقیق آن‌ها چالش برانگیزتر است. بنابراین آن‌ها هنوز به طور گسترده مورد استفاده قرار نگرفته‌اند.



شکل 1. دسته‌های کلی تغییرات ژنومی و انواع مختلف جهش‌ها.

<sup>22</sup> single-base substitutions

<sup>23</sup> single-nucleotide or dinucleotide substitutions

<sup>24</sup> small insertions and deletions (indels)

<sup>25</sup> copy number changes

<sup>26</sup> structural rearrangements

<sup>27</sup> transposable element integration events

<sup>28</sup> localized hypermutation (kataegis)

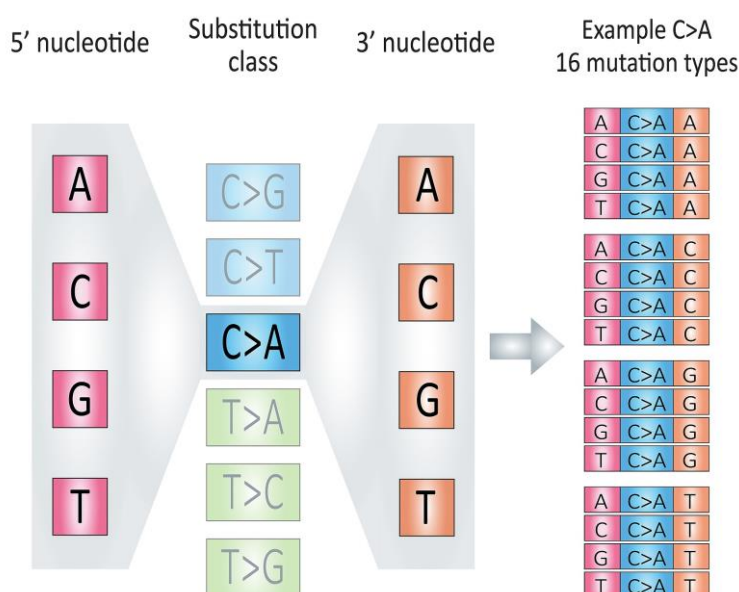
<sup>29</sup> epigenetic changes

<sup>30</sup> mathematical abstraction

<sup>31</sup> structural variants

یکی از معروفترین روش‌های مدلسازی جایگزینی‌های تک بازی به این صورت است که، شش کلاس جایگزینی پایه وجود دارد:  $C>A$ ,  $C>G$ ,  $C>T$ ,  $T>A$ ,  $T>C$ ,  $T>G$ . جایگزینی  $G>T$  معادل جایگزینی  $C>A$  در نظر گرفته می‌شود زیرا تشخیص اینکه در ابتدا جایگزینی در کدام رشته DNA (رو به جلو<sup>۳۲</sup> یا معکوس<sup>۳۳</sup>) رخ داده است، امکان‌پذیر نیست. بنابراین هر دو جایگزین  $C>A$  و  $G>T$  به عنوان بخشی از کلاس " $C>A$ " محاسبه می‌شوند. به دلیل مشابه جهش‌های  $G>C$ ,  $G>A$ ,  $A>T$ ,  $A>G$  و  $A>C$  به ترتیب به عنوان بخشی از کلاس‌های " $C>G$ ", " $C>T$ ", " $T>A$ ", " $T>C$ " و " $T>G$ " محاسبه می‌شوند.

گرفتن اطلاعات از جفت بازهای مجاور 5' و 3' منجر به ایجاد 96 نوع جهش ممکن می‌شود (مانند  $A[C>A]T$ ,  $A[C>A]A$  و غیره) (شکل 2). کاتالوگ جهش یک تومور با دسته‌بندی هر جهش نوکلئوتیدی منفرد<sup>۳۴</sup> (SNV) در یکی از 96 نوع جهش و شمارش تعداد کل جایگزینی‌ها برای هر یک از این 96 نوع جهش ایجاد می‌شود.



شکل 2. با در نظر گرفتن باز مجاور 5' (A, C, G, T)، 6 کلاس جایگزینی ( $C>A$ ,  $C>G$ ,  $C>T$ ,  $T>A$ ,  $T>C$ ,  $T>G$ ) و باز مجاور 3' (A, C, G, T) 96 نوع جهش ایجاد می‌گردد ( $4 \times 6 \times 4 = 96$ ). 16 نوع جهش کلاس جایگزینی  $C>A$  به عنوان مثال نشان داده شده است.

هنگامی که کاتالوگ یا ماتریس جهشی (به عنوان مثال تعداد 96 نوع جهش برای جایگزینی‌های تک بازی) یک تومور به دست آید، دو رویکرد برای رمزگشایی مشارکت امضاهای جهشی مختلف در چشم انداز ژنومی تومور دنبال می‌شود. ابتدا کاتالوگ جهشی تومور با کاتالوگ جهشی مرجع یا مجموعه داده مرجع امضاهای جهشی مقایسه می‌گردد. سپس مدل‌سازی امضاهای جهشی می‌تواند با استفاده از روش‌های آماری مانند فاکتورسازی ماتریس غیرمنفی<sup>۳۵</sup> (NMF) برای شناسایی فرایندهای جهش جدید بالقوه صورت گیرد (2). شناسایی سهم امضاهای جهشی متنوع در سرطان‌زایی بینشی در مورد بیولوژی تومور فراهم می‌کند و می‌تواند فرصت‌هایی را برای درمان هدفمند ارائه دهد.

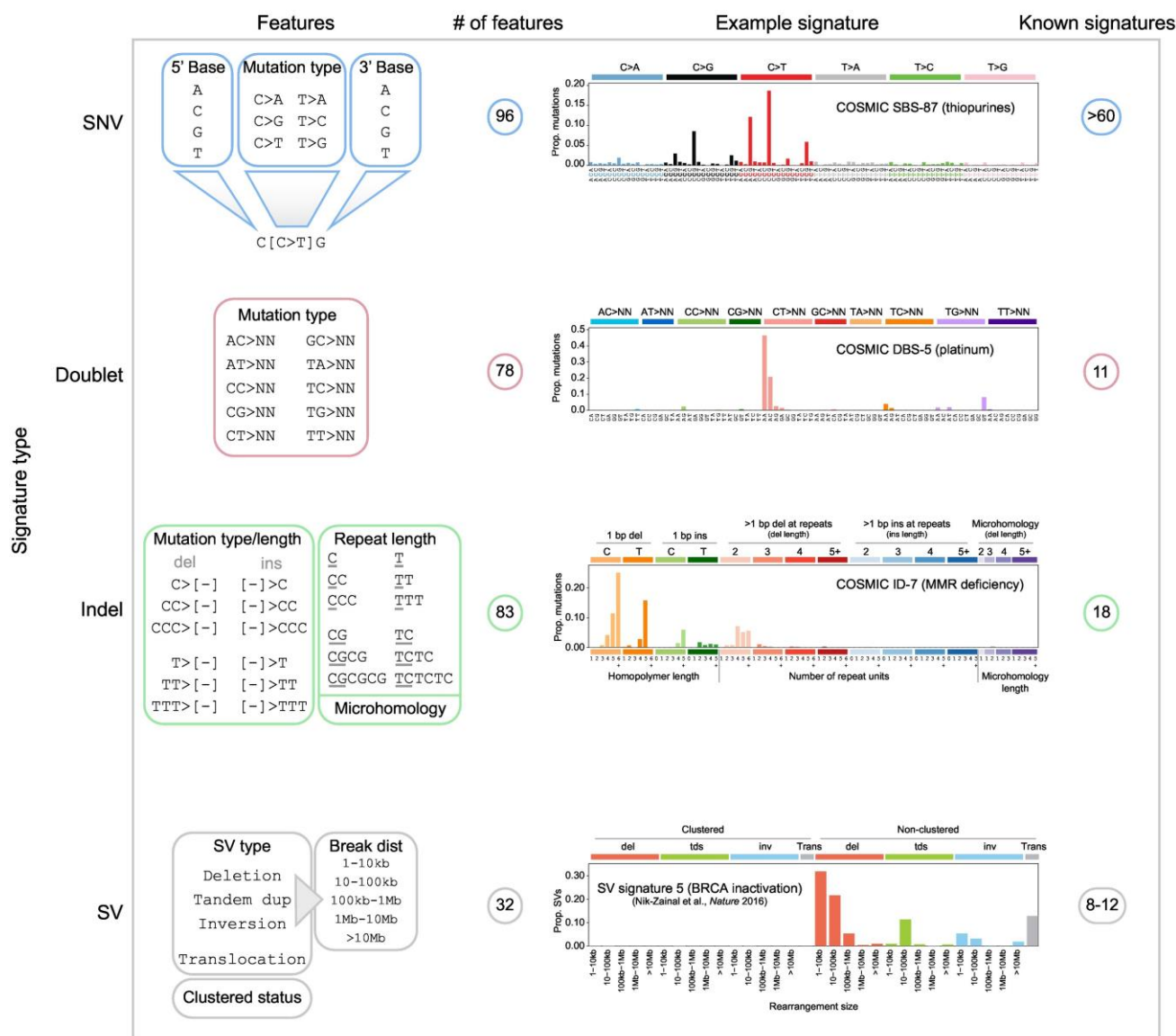
<sup>32</sup> forward

<sup>33</sup> reverse

<sup>34</sup> single nucleotide variant (SNV)

<sup>35</sup> non-negative matrix factorization (NMF)

تا سال 2021 و در نسخه 3.2 دیتابیس COSMIC<sup>۳۶</sup>، بیش از 60 امضای جهشی SBS در سراسر سرطان‌ها بر اساس 96 نوع ممکن که زمینه سه نوکلئوتیدی SNV را در نظر می‌گیرند، شناسایی و گنجانده شده است (20). DBS 11 و 18 امضای ایندل نیز شناسایی شده‌اند (20). در نهایت، 8-12 امضاهای SV بر اساس الگوهای شماره کپی مجاور<sup>۳۷</sup>، جهت‌گیری نقطه شکست<sup>۳۸</sup>، وجود SV های خوشه‌ای، و زمینه توالی نزدیک<sup>۳۹</sup> شناسایی شده‌اند (25، 29-31) (شکل 3). که البته در این چند سال اخیر این تعداد به طور منظم به روز شده است.



شکل 3. ویژگی‌های انواع تک نوکلئوتیدی (SNV یا SBS)، دوگانه (DBS)، درج-حذف (indel)، و تغییرات ساختاری (SV). ستون اول ویژگی‌های ژنومی هر نوع امضا را نشان می‌دهد. برای SNV ها، زمینه سه نوکلئوتیدی، شامل بازهای 5' و 3' محل جهش یافته، همراه با نوع واریانت (به عنوان مثال، C>T) در نظر گرفته می‌شود. برای دوگانه‌ها، فقط آل‌های مرجع و تغییر یافته در نظر گرفته می‌شوند (N نشان دهنده هر بازی است). برای ایندل‌ها، نوع و طول جهش در نظر گرفته می‌شود، جایی که «[-]» دنباله‌های حذف شده (del) یا درج شده (in) را نشان می‌دهد. طول تکرارهای موضعی نیز در نظر گرفته می‌شود (به عنوان مثال، اگر GCGG به GCCG تبدیل شود، 3 برای C خواهیم داشت). و همچنین اینکه آیا میکروهومولوژی<sup>۴۰</sup> در مجاورت ناحیه حذف شده وجود

<sup>36</sup> Catalogue of Somatic Mutations in Cancer (COSMIC) database

<sup>37</sup> adjacent copy number patterns

<sup>38</sup> breakpoint orientation

<sup>39</sup> nearby sequence context

<sup>40</sup> microhomology

دارد یا خیر. SV ها به چهار نوع نشان داده شده، کلاسه‌بندی می‌شوند و SV های داخل کروموزومی (تکثیرات پشت سر هم<sup>۴۱</sup>، حذف‌ها، و وارونگی‌ها<sup>۴۲</sup>) بر اساس فاصله بین دو نقطه شکست (فاصله شکست<sup>۴۳</sup>) ساب کلاسه‌بندی می‌شوند. اینکه آیا SV های مشابه در یک منطقه ژنومی خاص خوشه‌بندی شده‌اند نیز در نظر گرفته می‌شود. ستون دوم تعداد ویژگی‌های مورد استفاده برای طبقه‌بندی هر امضا را نشان می‌دهد. به عنوان مثال، امضاهای SNV دارای چهار باز احتمالی 5'، شش نوع جهش احتمالی، و چهار باز احتمالی 3' هستند که 96 ویژگی را ایجاد می‌کند. ستون سوم نمونه‌هایی از امضاها را نشان می‌دهد، با محور Y که نسبت جهش‌ها در هر ویژگی را نشان می‌دهد. شناسه امضاهای شناسایی شده COSMIC (در صورت وجود) و علت‌شناسی<sup>۴۴</sup> امضا را نشان می‌دهد. ستون چهارم تعداد امضاهای گزارش شده در مطالعات منتشر شده یا COSMIC را نشان می‌دهد. مخفف: MMR، بازسازی نابرابر<sup>۴۵</sup>.

آنالیز امضاهای جهشی یک رویکرد قدرتمند برای درک فرآیندهای جهش‌زایی است که تکامل ژنوم سرطان را شکل داده است. برای ارزیابی امضاهای جهشی فعال در یک ژنوم سرطانی، ابتدا باید فعالیت‌های آن‌ها را با تخمین تعداد جهش‌های حک شده<sup>۴۶</sup> توسط هر امضا، کمی‌سازی کرد. مطالعه این امضاهای جهشی دارای پتانسیل مهمی برای درک بیشتر ما از علل سرطان‌های فردی است و می‌تواند بینش‌های جدیدی در رابطه با پیشگیری و درمان سرطان ارائه دهد. امروزه، آنالیزهای امضای جهشی به یک جزء استاندارد مطالعات ژنومی تبدیل شده‌اند، زیرا می‌توانند منابع محیطی<sup>۴۷</sup> و درون‌زای جهش‌زایی را در هر تومور نشان دهند. در واقع، این رشته نوپا در جهت‌های محاسباتی، تجربی و بالینی در حال توسعه و گسترش است و به سمت استفاده از روش بالینی معنادار، آگاهی‌رسانی، تلاش‌هایی جهت پیشگیری سرطان، شناسایی پتانسیل‌های سرطانی ناشناخته (32، 33)، طبقه‌بندی بیماری‌زایی جهش‌های رده‌زایی<sup>۴۸</sup> (34)، هدایت روش‌های تشخیصی، پیش‌بینی پاسخ به درمان برای مدیریت بالینی بیماران سرطانی (29)، شناسایی حساسیت به درمان‌های ضد سرطان (35) و مداخلات شخصی‌سازی شده سرطان<sup>۴۹</sup> پیش می‌رود (16، 27، 35-40).

برای پزشکان، استفاده از امضاهای جهشی قابل اعتماد برای طبقه‌بندی بالینی بسیار مهم است. در حالی که توسعه روش‌هایی برای کشف امضاهای جهشی به موفقیت قابل توجهی دست یافته است، این هنوز یک زمینه نوظهور است که ناشی از پیشرفت‌های تحلیلی و تکنولوژیکی اخیر است و علیرغم پیشرفت‌ها در این زمینه، منشأ بسیاری از امضاهای جهشی همچنان مبهم است. علاوه بر این، مطالعات و تجزیه و تحلیل‌های جدیدتر، نسخه‌های متعدد و متفاوتی از این امضاها را نسبت به موارد قبلی گزارش کردند (20)، که جامعه را به این سوال سوق می‌دهد که این یافته‌ها که حاصل نتایج ریاضی انتزاعی‌اند تا چه میزان دقیق بوده و بدون تأیید تجربی قابل اعتماد و استناد هستند (41) و آیا می‌توان با استفاده از بهبود الگوریتم‌های موجود، الگوهای امضاهای جهشی را با دقت بالاتری شناسایی کرد؟

## پیش فرض ها و فرضیه ها:

### مفروضات:

1. هر تومور سرطانی، حاوی الگوهای ثابت و منحصر به فردی از تغییرات ژنومی، تحت عنوان امضای جهشی است که توسط فرآیندهای جهش‌زایی خاص درون‌زا یا عوامل خارجی مانند اشعه ایکس و مواد سرطان‌زا در طول تومورزایی ایجاد می‌شود.

<sup>41</sup> tandem duplications

<sup>42</sup> inversions

<sup>43</sup> break dist

<sup>44</sup> etiology

<sup>45</sup> mismatch repair

<sup>46</sup> imprinted mutations

<sup>47</sup> environmental

<sup>48</sup> germline variants

<sup>49</sup> personalized cancer interventions

2. با توجه به پیشرفت‌های فناوری توالی‌یابی و افزایش داده‌های در دسترس، اطلاعات حاصل از تحلیل امضاهای جهشی می‌تواند به دقت بیشتری در تعیین علت‌ها و مکانیسم‌های مولکولی نئوپلازی در سرطان‌ها کمک کند و به توسعه راهکارهای نوین برای پیشگیری و درمان سرطان منجر شود.

3. دیتاهای مورد استفاده در این مطالعه، داده‌های از قبل شبیه‌سازی شده و در دسترس از مطالعات پیشین است (21, 28).

#### فرضیه‌ها:

1. تاکنون، منشأ بسیاری از امضاهای جهشی همچنان ناشناخته است و اطلاعات بسیاری از امضاهای جهشی شناسایی شده به وسیله روش‌های محاسباتی هنوز تأیید تجربی نشده‌اند، بنابراین نیاز به تحقیقات بیشتر در این زمینه حائز اهمیت است.

2. مدل‌های طراحی شده، الگوریتم‌های و راهکارهای نوظهوری که قابلیت شناسایی الگوهای امضاهای جهشی در داده‌های ژنوم تومورها را دارا هستند می‌توانند با استفاده از تکنیک‌های ریاضی مدرن بهبود بیشتری یافته تا به دقت و صحت بالاتری در شناسایی و تحلیل‌های الگوهای جهشی منجر شود.

#### مواد و روش انجام تحقیق:

مجموعه داده مورد استفاده در این مطالعه، شامل الگوهای SBS از 2700 ژنوم سرطان قبلاً شبیه‌سازی شده از پروژه PCAWG<sup>50</sup>، مربوط به 300 تومور از 9 نوع سرطان مختلف، از جمله: کارسینوم سلول انتقالی مثانه<sup>51</sup>، آدنوکارسینوم مری<sup>52</sup>، آدنوکارسینوم سینه<sup>53</sup>، کارسینوم سلول سنگفرشی ریه<sup>54</sup>، کارسینوم سلول کلیه<sup>55</sup>، آدنوکارسینوم تخمدان<sup>56</sup>، استئوسارکوم<sup>57</sup>، آدنوکارسینوم دهانه رحم<sup>58</sup> و آدنوکارسینوم معده<sup>59</sup> است که ژنوم سرطان این نمونه‌ها با استفاده از 21 امضای مرجع COSMIC مختلف شبیه‌سازی شده‌اند (21, 28).

حداقل دو رویکرد مجزا برای آنالیز امضاهای جهشی وجود دارد. استخراج *de novo* یک رویکرد یادگیری ماشینی بدون نظارت<sup>60</sup> است که امکان شناسایی الگوهای امضاهای جهشی‌یافته شناخته شده و ناشناخته قبلی را فراهم می‌کند (21). این نوع آنالیز از آنجایی که به گروه‌های بزرگی شامل معمولاً بیش از 100 نمونه نیاز دارد، عمده‌تاً برای استخراج امضاهای مرجع استفاده می‌شود. در مقابل، بازسازی امضاهای جهشی یک رویکرد بهینه‌سازی عددی است که با تعیین تعداد جهش‌های منتسب به هر عامل امضا در آن نمونه، امکان تخصیص امضاهای شناخته شده (در بیشتر موارد، مرجع) را به یک نمونه جداگانه فراهم می‌کند. در حالی که بازسازی نمی‌تواند

<sup>50</sup> Pan-Cancer Analysis of Whole Genomes

<sup>51</sup> bladder transitional cell carcinoma

<sup>52</sup> esophageal adenocarcinoma

<sup>53</sup> breast adenocarcinoma

<sup>54</sup> lung squamous cell carcinoma

<sup>55</sup> renal cell carcinoma

<sup>56</sup> ovarian adenocarcinoma

<sup>57</sup> osteosarcoma

<sup>58</sup> cervical adenocarcinoma

<sup>59</sup> stomach adenocarcinoma

<sup>60</sup> unsupervised

فعالیت‌های امضاهای جهشی ناشناخته قبلی را شناسایی یا کمی کند، این رویکرد به طور گسترده در گروه‌های کوچک و برای نمونه‌های بالینی که ارزیابی‌ها تقریباً به طور انحصاری برای یک بیمار سرطانی انجام می‌شود، به کار می‌رود (42).

## 1. استخراج de novo امضاهای جهشی:

استخراج de novo امضاهای جهشی (19) یک رویکرد یادگیری ماشینی بدون نظارت است که در آن یک ماتریس،  $M$ ، که مربوط به جهش‌های پیکری در مجموعه‌ای از نمونه‌های سرطانی تحت یک طبقه‌بندی جهشی (43) است، با حاصلضرب دو ماتریس با رتبه پایین  $S$  و  $A$  تقریب زده می‌شود. ماتریس  $S$  مجموعه‌ای از امضاهای جهشی را منعکس می‌کند، در حالی که ماتریس  $A$  شامل فعالیت‌های امضاها می‌شود. یک فعالیت مربوط به تعداد جهش‌های ایجاد شده توسط یک امضا در نمونه سرطان است. برای به حداکثر رساندن تأثیر تجزیه و تحلیل امضای جهشی، درک دقیق ریاضی و الگوریتم‌های قوی برای افزایش دقت و قابلیت تفسیر آن مورد نیاز است. این زمینه در دهه گذشته شاهد توسعه سریع تکنیک‌های محاسباتی بوده است. ابزارهای محبوبی مانند MuSiCal (44)، SigProfilerExtractor (21)، SignatureAnalyzer (20, 45, 46)، signature.tools.lib (22, 24) و سایرین به موفقیت قابل توجهی دست یافته‌اند (47). چارچوب‌های محاسباتی برای استخراج de novo امضاهای جهشی توسعه یافته تا به امروز در **جدول 1** نشان داده شده است.

Tool name	Input	Platform	Factorization method	Factorization engine	GPU	Manual selection	Automatic selection	Automatic algorithm	Mutational catalog support	Plotting support	COSMIC comparison
Emu <sup>(49)</sup>	matrix	C++	EM	original implementation <sup>(49)</sup>	no	yes	yes*	BIC <sup>(48)</sup>	SBS-96	no	no
Maftools <sup>(51)</sup>	matrix, MAF	R- Bioconductor	NMF	NMF R package <sup>(50)</sup>	no	yes	no	–	SBS-96	SBS-96	1 to 1
MutationalPatterns <sup>(52)</sup>	matrix, VCF	R- Bioconductor	NMF	NMF R package <sup>(50)</sup>	no	yes	no	–	SBS-96, SBS-192	SBS-96, SBS-192	1 to 1
★ MuSiCal <sup>(44)</sup>	matrix	Python	mvNMF	original implementation	yes	yes	yes*	sparse NNLS	new mutational catalog <sup>(44)</sup>	new catalog <sup>(44)</sup>	1 to many
MutSignatures <sup>(54)</sup>	matrix, VCF, MAF	R	NMF	Brunet et al. <sup>(53)</sup>	no	no	no	–	SBS-96	SBS-96	1 to 1
MutSpec <sup>(55)</sup>	matrix, VCF, custom	Galaxy, Perl, R	NMF	NMF R package <sup>(50)</sup>	no	yes	no	–	SBS-96, SBS-192	SBS-96, SBS-192	1 to 1
SigFit <sup>(58)</sup>	matrix	R	Bayesian inference	Stan R package <sup>(57)</sup>	no	yes	yes*	Elbow method <sup>(56)</sup>	SBS-96	SBS-96, SBS-192	1 to 1
SigMiner <sup>(60)</sup>	matrix, MAF	R	(automatic) Bayesian NMF, (manual) NMF	(automatic) SignatureAnalyzer implementation, <sup>(46)</sup> (manual) NMF R package <sup>(50)</sup>	no	yes*	yes	ARD <sup>(59)</sup>	SBS-96, DBS-78, ID-83	generic	1 to 1
SignatureAnalyzer <sup>(45, 46)</sup>	matrix, MAF	R (CPU), <sup>(61)</sup> Python (GPU) <sup>(62)</sup>	Bayesian NMF	original implementation <sup>(45, 46)</sup>	yes	no	yes	ARD <sup>(59)</sup>	SBS-96, DBS-78, ID-83	SBS-96, DBS-78, ID-83	1 to 1
SignatureToolsLib <sup>(22)</sup>	matrix, VCF, custom	R	NMF	NMF R package <sup>(50)</sup>	no	yes	no	–	SBS-96, DBS-78, ID-83, SV-32	SBS-96, SV-32, generic	1 to 2
SigneR <sup>(63)</sup>	matrix, VCF	R- Bioconductor, C++	Bayesian NMF	original implementation <sup>(63)</sup>	no	yes	yes*	BIC <sup>(48)</sup>	SBS-96	SBS-96	no
★ SigProfilerExtractor <sup>(21)</sup>	matrix, VCF, MAF, custom	Python, R wrapper	NMF	original implementation	yes	yes	yes*	NMFk <sup>(64)</sup>	SBS-96, DBS-78, ID-83, CN-48, others, <sup>(65)</sup> any	SBS-96, DBS-78, ID-83, CN-48, SV-32, others, <sup>(43)</sup> generic	1 to many
SigProfiler_PCAWG <sup>(20)</sup>	matrix, VCF, MAF, custom	Python, MATLAB	NMF	Brunet et al. <sup>(53)</sup>	no	yes	no	–	SBS-96, DBS-78, ID-83, others, <sup>(43)</sup> any	SBS-96, DBS-78, ID-83	no
SomaticSignatures <sup>(66)</sup>	matrix, VCF	R- Bioconductor	NMF, PCA	NMF R package <sup>(30)</sup> pcaMethods R package <sup>(65)</sup>	no	yes	no	–	SBS-96	SBS-96	no
TensorSignatures <sup>(68)</sup>	VCF	Python	NTF	TensorFlow <sup>(67)</sup>	yes	yes	yes*	BIC <sup>(48)</sup>	tensor	SBS-96 with strand bias	no

**جدول 1. مروری بر ابزارهای بیوانفورماتیک توسعه داده شده جهت استخراج de novo امضاهای جهشی.** ابزارها بر اساس حروف الفبا مرتب شده‌اند. 1 تا 1 به یک امضای de novo اشاره دارد که دقیقاً با یک امضای COSMIC مطابقت دارد. 1 تا 2 به یک امضای de novo اشاره دارد که با ترکیبی از حداکثر دو امضای COSMIC مطابقت دارد. 1 to many به یک امضای de novo اشاره دارد که با ترکیبی از یک یا چند امضای COSMIC مطابقت دارد.

<sup>61</sup> low-rank matrices



MAF، فرمت حاشیه‌نویسی جهش<sup>۶۲</sup>؛ VCF، فرمت فراخوانی جهش<sup>۶۳</sup>؛ EM، الگوریتم به حداکثر رساندن انتظار<sup>۶۴</sup>؛ NMF، فاکتورسازی ماتریس غیر منفی. PCA، تجزیه و تحلیل اجزای اصلی<sup>۶۵</sup>؛ NTF، فاکتورسازی تانسور غیر منفی<sup>۶۶</sup>؛ ARD، تعیین ارتباط خودکار<sup>۶۷</sup>؛ BIC، معیار اطلاعات بیزی<sup>۶۸</sup>؛ COSMIC، کاتالوگ جهش‌های جسمی در سرطان. SBS، جایگزین‌های تک بازی. DBS، جایگزین‌های بازی دوگانه؛ ID، درج‌ها و حذف‌های کوچک؛ CN، شماره کپی؛ SV، جهش‌های ساختاری.

\* رویکرد پیش‌فرض برای انتخاب تعداد کل امضاها وقتی ابزاری از انتخاب دستی و خودکار پشتیبانی می‌کند.

یکی از چالش‌ها در مرحله کشف de novo امضاهای جهشی، این است که در آن امضاهای جهشی مستقیماً از یک گروه استخراج می‌شوند. ابزارهای متعددی توسعه داده شده است که بیشتر آن‌ها از نظر الگوریتمی، بر پایه فاکتورسازی ماتریس غیرمنفی (NMF) (69) یا روی رویکردهای ریاضی مشابه NMF (61، 62، 70) متکی هستند. مزیت اصلی NMF نسبت به سایر رویکردهای فاکتورسازی، توانایی آن در به دست آوردن عوامل غیرمنفی است که بخشی از داده‌های اصلی هستند. بنابراین امکان تفسیر بیولوژیکی عوامل غیرمنفی شناسایی شده را فراهم می‌کند (69). با این حال آن‌ها حتی برای یک مجموعه داده منجر به ایجاد نتایج بسیار متغیری می‌شوند (20، 47). بنابراین مقایسه امضاهای کشف شده در طول مطالعات دشوار است، به ویژه تعیین اینکه آیا یک امضا، جدید است یا صرفاً تغییری از امضای شناخته شده قبلی به دلیل بایاس الگوریتمی است. این موضوع ممکن است زیربنای تعداد زیادی از امضاهایی باشد که بیش از حد مشابه هستند یا می‌توانند به عنوان ترکیبات خطی یکدیگر در فهرست پرکاربرد امضاهای شناخته شده از کاتالوگ جهش‌های پیکری در سرطان (COSMIC) بیان شوند. همچنین ممکن است ویژگی بافت مشاهده شده برای برخی از امضاها را مخدوش کند (22).

در این میان روش MuSiCal که توسط تیم Peter J. Park از دانشگاه هاروارد توسعه داده و در فوریه 2024 منتشر شد (44)، یک چارچوب جامع را ارائه می‌کند که تخصیص امضای دقیق و همچنین کشف امضای قوی و حساس را امکان‌پذیر می‌کند. MuSiCal برای حل چالش عنوان شده، از چندین روش جدید استفاده می‌کند، از جمله NMF حداقل حجم<sup>۶۹</sup> (mvNMF) (71-74)، حداقل مربعات غیرمنفی پراکنده مبتنی بر احتمال<sup>۷۰</sup> (NNLS) و یک رویکرد مبتنی بر داده برای بهینه‌سازی سیستماتیک پارامترها و اعتبارسنجی مناسب. این رویکرد اخیر در قیاس با روش بسیار معروف SigProfilerExtractor که توسط تیم L. Alexandrov از دانشگاه سن دیگو در 2022 معرفی شده بود نتایج بهتری را بدست آورد.

## 2. بازسازی و تخصیص امضاهای جهشی:

تخصیص امضاهای جهشی به نمونه‌های سرطان فردی و جهش‌های پیکری فردی فرصتی را برای شناسایی فرآیندهای مسئول جهش‌های سوماتیکی به صورت نمونه به نمونه<sup>۷۱</sup> فراهم می‌کند و ما را قادر می‌سازد تا فرآیندهای جهشی را در ژنوم سرطان مشخص کنیم.

<sup>62</sup> mutation annotation format

<sup>63</sup> variant call format

<sup>64</sup> expectation maximization algorithm

<sup>65</sup> principal component analysis

<sup>66</sup> nonnegative tensor factorization

<sup>67</sup> automatic relevance determination

<sup>68</sup> Bayesian information criterion

<sup>69</sup> minimum-volume NMF (mvNMF)

<sup>70</sup> , likelihood-based sparse nonnegative least squares (NNLS)

<sup>71</sup> sample-by-sample

در دهه گذشته، ابزارهای متعددی برای بازسازی مجدد امضاهای شناخته شده، از جمله deconstructSigs (42)، Mutational Patterns (75، 76)، sigLASSO (77) و SignatureTool (22، 24) توسعه یافته‌اند (جدول 2). اکثر این ابزارها تقریباً به طور انحصاری از امضاهای SBS پشتیبانی می‌کنند و فاقد یک رابط آنلاین هستند. اگرچه چند ابزار وب، از جمله MuSiCa (78) و Mutalisk (79) نیز وجود دارد.

Tool name	Input data (mutations)	Platform	Optimization Method	Algorithm	Computational engine	Penalties	Post hoc filter (TMB threshold)
DeconstructSigs (2016) <sup>(42)</sup>	matrix, custom	R	Multiple linear regression with a nonnegative cutoff on activities	Golden-section search algorithm <sup>(80)</sup>	Original implementation	Addition penalty (SSE; default: 0.001)	Yes (6%)
MutationalPatterns (2018) <sup>(75)</sup>	matrix, VCF	R	NNLS	Levenberg Marquardt algorithm <sup>(82)</sup>	Pracma R package <sup>(81)</sup>	No penalties	No
MutationalPatterns (strict) (2022) <sup>(76)</sup>	matrix, VCF	R	NNLS	Levenberg Marquardt algorithm <sup>(82)</sup>	Original implementation (penalty framework) and Pracma R package <sup>(81)</sup>	Removal penalty similarity; (cosine default: 0.004)	No
sigLASSO (2020) <sup>(77)</sup>	matrix, VCF, MAF, custom	R	Non-negative linear LASSO regression	Alternative convex search algorithm <sup>(84)</sup>	Original implementation (framework) and glmnet R package (Lasso regression) <sup>(83)</sup>	Optimized penalty (L1 norm). Priors. Lambda hyperparameter.	No
SignatureToolsLib (2020, 2022) <sup>(22, 24)</sup>	matrix, VCF, BEDPE, custom	R Web app	Non-negative linear regression (KL Divergence objective function)	Lee's multiplicative algorithm <sup>(69)</sup>	NNLM R package <sup>(85)</sup>	No penalties	Yes (5%)
★ SigProfilerAssignment (2023) <sup>(28)</sup>	matrix, VCF, MAF, custom segmentation,	Python R Web app	NNLS	Lawson Hanson algorithm <sup>(87)</sup>	Original implementation (penalty framework) and Scipy python package (NNLS) <sup>(86)</sup>	Initial removal, addition, and removal penalties (L2 norm; default: 0.05, 0.05 and 0.01)	No

**جدول 2. مروری بر ابزارهای بیوانفورماتیک توسعه داده شده جهت تخصیص امضاهای جهشی.** ابزارها بر اساس حروف الفبا مرتب شده‌اند. ستون‌های جدول موارد زیر را نشان می‌دهند: نام چهارچوب محک زدن ابزار، انواع داده‌های ورودی پشتیبانی‌شده، پلت‌فرم‌های عملیاتی سازگار، روش بهینه‌سازی استفاده‌شده، الگوریتم برازش اولیه <sup>۷۲</sup>، موتور محاسباتی استفاده‌شده، جریمه‌های اضافی اعمال‌شده، و آستانه درصد بار جهش تومور برای جلوگیری از بیش‌برازش امضاها. MAF، فرمت حاشیه‌نویسی جهش؛ matrix: ماتریس جهشی <sup>۷۳</sup>؛ NNLM؛ مدل‌های خطی غیر منفی <sup>۷۴</sup>؛ NNLS، کمترین مربعات غیر منفی <sup>۷۵</sup>؛ SSE، مجموع مربعات خطاها <sup>۷۶</sup>؛ TMB، بار جهش تومور <sup>۷۷</sup>؛ VCF، فرمت فراخوانی جهش.

در این میان SigProfilerAssignment، یک دسکتاپ و یک چارچوب محاسباتی آنلاین برای تخصیص تمام انواع امضاهای جهشی، از جمله مجموعه‌های COSMIC از امضاهای مرجع SBS، DBS، ID و CN به نمونه‌های مجزا را ارائه می‌کند که توسط تیم L. Alexandrov در دسامبر 2023 معرفی شد (28).

اکنون یک رابط آنلاین کاربرپسند از ابزار SigProfilerAssignment جهت استفاده، به عنوان بخشی از وب‌سایت امضاهای جهشی COSMIC (88) در <https://cancer.sanger.ac.uk/signatures/assignment/> در دسترس است. SigProfilerAssignment اولین ابزاری است که امکان آنالیز امضاهای تعداد کمی و تخصیص احتمالی امضاها را به جهش‌های

<sup>72</sup> primary fitting algorithm

<sup>73</sup> mutational matrix

<sup>74</sup> non-negative linear models

<sup>75</sup> non-negative least squares

<sup>76</sup> sum of squared errors

<sup>77</sup> tumor mutational burden

پیکری فردی فراهم می‌کند که پیش‌بینی کننده خوبی برای بقای بالینی هستند (89, 90). علاوه بر این، SigProfilerAssignment از تخصیص امضاهای جهشی استخراج شده de novo و مجموعه‌ای از امضاهای سفارشی ارائه شده توسط کاربر پشتیبانی می‌کند. این ابزار به عنوان موتور محاسباتی خود، از پیاده‌سازی سفارشی الگوریتم مرحله‌ای پیش‌رونده<sup>۷۸</sup> (91) برای رگرسیون پراکنده<sup>۷۹</sup> و کمترین مربعات غیرمنفی (NNLS) بر اساس روش لاوسون-هانسون<sup>۸۰</sup> (87) برای بهینه‌سازی عددی استفاده می‌کند و نتایج مقایسه روش‌های مختلف در این زمینه نشان می‌دهد که SigProfilerAssignment از سایر ابزارهای رایج بهتر عمل می‌کند. در ادامه الگوریتم استفاده شده در این ابزار شرح داده می‌شود:

از نظر ریاضی، یک طرح جهش را می‌توان به عنوان یک الفبای محدود<sup>۸۱</sup>  $\Xi$  از انواع جهش نشان داد که در مجموع شامل حروف  $\xi$  است. در اینجا، یک امضای جهشی به عنوان یک تابع جرم احتمال<sup>۸۲</sup> با دامنه الفبای  $\Xi$  تعریف شده است. در نماد برداری<sup>۸۳</sup>، یک امضای جهشی را می‌توان به عنوان  $\vec{s} = [s_1, s_2, \dots, s_\xi]^T$  نشان داد، که در آن  $s_k$ ،  $1 \leq k \leq \xi$ ، احتمالی است برای امضای جهش،  $\vec{s}$  باعث ایجاد جهش‌هایی از نوع متناظر با حرف  $k$  ام الفبای  $\Xi$  شود. از آنجایی که یک امضای جهشی یک تابع جرم احتمالی است،  $\sum_{k=1}^{\xi} s_k = 1$  و  $0 \leq s_k \leq 1$  به این ترتیب، مجموعه‌ای از  $n$  امضای جهشی شناخته شده را می‌توان به عنوان یک ماتریس امضا،  $S \in \mathbb{R}_+^{\xi \times n}$ ، که در آن  $S = [\vec{s}^1, \vec{s}^2, \dots, \vec{s}^n]$  بیان کرد. علاوه بر این، مجموعه‌ای از جهش‌ها در ژنوم سرطان را می‌توان به عنوان  $v: \Xi \rightarrow \mathbb{N}_+^\xi$  تعریف کرد. در نماد برداری، مجموعه‌ای از جهش‌ها در ژنوم سرطان  $\vec{v} = [v_1, v_2, \dots, v_\xi]^T$ ، که در آن  $v_k$ ،  $1 \leq k \leq \xi$ ، تعداد جهش‌های آن ژنوم سرطان از نوع جهش مربوط به حرف  $k$  ام الفبای  $\Xi$  را منعکس می‌کند.

SigProfilerAssignment یک ماتریس امضا،  $S$  و مجموعه‌ای از جهش‌ها،  $\vec{v}$ ، را به عنوان ورودی می‌گیرد تا بردار ستونی از فعالیت‌های  $\vec{a} = [a_1, a_2, \dots, a_n]^T$ ، که در آن  $a_t \in \mathbb{N}_0^n$ ،  $1 \leq t \leq n$ ، متناظر با تعداد جهش‌های سوماتیک منتسب به  $t$  امین امضای جهش است را به عنوان خروجی تولید کند. فرض اساسی تخصیص امضاهای جهشی این است که جهش‌های درون یک نمونه را می‌توان به عنوان برهم‌نهی<sup>۸۴</sup> امضاهای جهشی شناخته شده و فعالیت‌های آن‌ها تقریب زد:

$$\vec{v} \approx S\vec{a} \quad (1)$$

بنابراین، با توجه به  $\vec{a} \geq 0$ ، باید بردار  $\vec{a}$  را استخراج کرد که به بهترین وجه با داده‌های ورودی ارائه شده مطابقت دارد. برای حل این مشکل بهینه‌سازی، SigProfilerAssignment از پیاده‌سازی سفارشی الگوریتم مرحله‌ای پیش‌رونده (91) استفاده می‌کند و حداقل مربعات غیرمنفی (NNLS) (92) را بر اساس روش لاوسون-هانسون (92) اعمال می‌کند:

$$\min_{\vec{a} \geq 0} \|\vec{v} - S\vec{a}\|_2^2 \quad (2)$$

<sup>78</sup> forward stagewise algorithm

<sup>79</sup> sparse regression

<sup>80</sup> Lawson-Hanson method

<sup>81</sup> finite alphabet

<sup>82</sup> probability mass function

<sup>83</sup> vector notation

<sup>84</sup> superposition

الگوریتم ابتدا با محاسبه حداقل خطای نسبی<sup>۸۵</sup>،  $\epsilon_{min} \frac{\|\vec{v}-S\vec{a}\|_2^2}{\|\vec{v}\|_2^2}$ ، با استخراج بردار  $\vec{a}$  غیرمنفی بهینه برای مجموعه کامل همه امضاهای مرجع،  $S$ ، با استفاده از معادله (2) شروع می‌شود. این حداقل خطا بهترین توضیح ممکن را در مورد داده‌ها ارائه می‌دهد، اما همچنین منجر به بیش برآزش<sup>۸۶</sup> می‌شود زیرا از همه امضاهای موجود استفاده می‌شود.

در مرحله بعد، این ابزار به ترتیب از مراحل حذف و اضافه کردن امضا بر اساس الگوریتم‌های مرحله‌ای پیش‌رونده و پس‌رونده<sup>۸۷</sup> استفاده می‌کند (91). ابتدا، امضاها با استفاده از یک الگوریتم مرحله‌ای پس‌رونده حذف می‌شوند (الگوریتم 1). به طور خاص، هر امضا از مجموعه امضای مرجع،  $S$ ، به طور مکرر حذف می‌شود و مجموعه امضای باقیمانده،  $\hat{S}$ ، با اعمال رابطه (2) به نمونه  $\vec{v}$  نسبت داده می‌شود. افزایش خطای نسبی،  $\epsilon_j \frac{\|\vec{v}-\hat{S}\vec{a}\|_2^2}{\|\vec{v}\|_2^2} - \epsilon_{min}$ ، به دلیل حذف یک امضا با اضافه کردن امضای  $j$  ام از  $S$  محاسبه می‌شود. امضایی با کمترین افزایش نسبی در میزان خطا از مجموعه امضا،  $S$ ، حذف می‌شود، مشروط بر اینکه افزایش، کمتر از یک آستانه خاص (مقدار پیش فرض 0.01) باشد. پس از حذف نهایی امضا با حداقل افزایش نرخ خطای نسبی، حداقل خطای نسبی،  $\epsilon_{min}$ ، و مجموعه‌ی امضاها،  $S$ ، به روز می‌شوند تا این حذف را منعکس کنند. مرحله حذف تکرار می‌شود تا زمانی که همه امضاهایی که شرایط را برآورده می‌کنند از  $S$  حذف شوند. مراحل حذف با مراحل جمع بر اساس الگوریتم مرحله‌ای پیش‌رونده (الگوریتم 1) دنبال می‌شود. به طور خاص، هر یک از امضاهای مرجع حذف شده قبلی به طور مکرر به  $S$  مجدد اضافه می‌شود و مجموعه امضای جدید،  $\hat{S}$ ، با اعمال معادله (2) برای نمونه  $\vec{v}$  مناسب است. بنابراین، کاهش خطای نسبی،  $\epsilon_l = \epsilon_{min} - \frac{\|\vec{v}-\hat{S}\vec{a}\|_2^2}{\|\vec{v}\|_2^2}$ ، به دلیل اضافه کردن یک امضا، با اضافه کردن امضای  $l$  ام به  $S$  محاسبه می‌شود. امضا با حداکثر کاهش نسبی میزان خطا، به مجموعه امضا،  $S$ ، برمی‌گردد، مشروط بر اینکه افزایش، بیش از یک آستانه خاص (مقدار پیش فرض 0.05) باشد. پس از افزودن نهایی امضا با بیشترین کاهش نرخ نسبی، حداقل خطای نسبی،  $\epsilon_{min}$ ، و مجموعه امضاها،  $S$ ، به روز می‌شوند تا این اضافه را منعکس کنند. مرحله جمع تکرار می‌شود تا زمانی که همه امضاهایی که شرایط را برآورده می‌کنند به  $S$  مجدد اضافه شوند. در نهایت، مراحل جمع و حذف تا زمان همگرایی تکرار می‌شود، جایی که هیچ امضایی اضافه یا از لیست امضاها حذف نمی‌شود (الگوریتم 1).

SigProfilerAssignment علاوه بر تعیین کمیت فعالیت هر امضای جهشی، امضاهای شناخته شده را نیز بر اساس زمینه جهش خاصی به تک تک جهش‌ها اختصاص می‌دهد.

$$p_k^t = \frac{s_k^t a_t}{[S\vec{a}]_k} \quad (3)$$

که در آن،  $p_k^t$  نشان دهنده احتمال جهش مربوط به حرف  $k$  ام الفبای  $\Sigma$  است که توسط امضای  $t$  ام در نمونه ایجاد می‌شود؛  $s_k^t$  احتمال امضای  $t$  ام است که منجر به ایجاد جهش مربوط به حرف  $k$  ام الفبای  $\Sigma$  است؛  $a_t$  تعداد جهش‌هایی است که به امضای جهش  $t$  ام نسبت داده می‌شود؛ و  $[S\vec{a}]_k$  مقدار  $k$  امین عنصر برداری است که از ضرب ماتریسی ماتریس امضا،  $S$ ، و فعالیت‌های امضای مشتق شده،  $\vec{a}$ ، به دست می‌آید.

<sup>85</sup> minimum relative error

<sup>86</sup> overfitting

<sup>87</sup> backward and forward stepwise algorithms

```

Input:  $\vec{v} \in \mathbb{N}_+^{\xi \times 1}$  (a vector corresponding to a set of mutations in a sample) and
 $S \in \mathbb{R}_+^{\xi \times n}$  (a matrix corresponding to a set of  $n$  known mutational signatures)
Output:  $\vec{a} \in \mathbb{N}_+^{n \times 1}$  (the vector reflecting the activities of the  $n$  known signatures in sample  $\vec{v}$ )

1:  $\epsilon_{\min}, \vec{a} = \text{calcNNLS}(\vec{v}, S)$ 
    $S^{\text{all}} = S$ 

2: While FLAG = True:
3:    $\epsilon_{\min}, S = \text{removeSignatures}(\vec{v}, S, \epsilon_{\min})$ 
4:    $\epsilon_{\min}, S = \text{addSignatures}(\vec{v}, S^{\text{all}}, S, \epsilon_{\min})$ 
5:   Set FLAG = False if  $S$  remains constant and there is no addition or removal of signatures
   END While

6:  $\epsilon_{\min}, \vec{a} = \text{calcNNLS}(\vec{v}, S)$ 
7: Return  $\vec{a}$ 

8: FUNCTION removeSignatures ( $\vec{v}, S, \epsilon_{\min}$ )
9:   While FLAG = True:
10:    For  $j$  in 1 to size( $S, 2$ ) do                                // loop from 1 to the total number of signatures in  $S$ 
11:       $\hat{S} = S[:, -j]$                                               // remove the  $j^{\text{th}}$  signature from  $S$ 
12:       $\epsilon[j], \vec{a}_j = \text{calcNNLS}(\vec{v}, \hat{S})$ 
    END For
13:    minIndex, minValue = min( $\epsilon$ )                                // find the signature set with least relative error
14:    If (minValue -  $\epsilon_{\min} \leq 0.01$ )
15:       $S = S[:, -\text{minIndex}]$ 
    else
16:      Return minValue,  $S$ 
    END If
  END While
END removeSignatures

17: FUNCTION addSignatures ( $\vec{v}, S^{\text{all}}, S, \epsilon_{\min}$ )
18:   While FLAG = True:
19:    For  $p$  in 1 to size( $S^{\text{all}}, 2$ ) do                                // loop from 1 to the total number of signatures in  $S^{\text{all}}$ 
20:       $\hat{S} = [S; S^{\text{all}}[:, p]]$                                     // add the  $p^{\text{th}}$  signature from  $S^{\text{all}}$ 
21:       $\epsilon[j], \vec{a}_j = \text{calcNNLS}(\vec{v}, \hat{S})$ 
    END For
22:    minIndex, minValue = min( $\epsilon$ )                                // find the signature set with least relative error
23:    If ( $\epsilon_{\min} - \text{minValue} \geq 0.05$ )
24:       $S = [S; S^{\text{all}}[:, \text{minIndex}]]$ 
    else
25:      Return minValue,  $S$ 
    END If
  END While
END addSignatures

26: FUNCTION calcNNLS( $\vec{v}, S$ )
27:    $\vec{a} = \text{nnls}(S, \vec{v})$                                           // Calculating NNLS with the Lawson-Hanson method
28:    $\epsilon = \|\vec{v} - S\vec{a}\|_2^2 / \|\vec{v}\|_2^2$                           // Computing relative error
29:   Return  $\epsilon, \vec{a}$ 
END calcNNLS

```

در نهایت، جهت ارزیابی مدل نهایی، پس از تخصیص امضاها، انتساب هر امضا به هر نمونه به عنوان نتیجه مثبت واقعی<sup>88</sup> (TP)، مثبت کاذب<sup>89</sup> (FP)، یا منفی کاذب<sup>90</sup> (FN) طبقه‌بندی می‌گردد. اگر حداقل یک جهش به امضا اختصاص داده شود و فعالیت‌های یافته‌های عینی<sup>91</sup> امضا بزرگتر از صفر باشد، یک امضای شناخته شده TP در نظر گرفته می‌شود. در مقابل، زمانی که امضا توسط یک ابزار تخصیص داده شد، اما فعالیت یافته‌های عینی صفر بود، به عنوان FP طبقه‌بندی می‌گردد. در نهایت، نتایج FN امضاهایی با فعالیت‌های یافته‌های عینی بالای صفر است که هیچ جهش پیکری به آن‌ها اختصاص داده نشده است. این معیارهای استاندارد امکان محاسبه دقت<sup>92</sup>، حساسیت<sup>93</sup> و امتیاز F1<sup>94</sup> مدل را در هر نمونه فراهم می‌کند که به صورت زیر تعریف می‌شود:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$F_1 \text{ score} = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \quad (6)$$

علاوه بر این، زمان اجرا و استفاده از حافظه مدل نیز در قیاس سایر مدل‌ها بررسی می‌گردد.

در این مطالعه، تلاش بر این است که با بهره‌گیری از تغییرات در مدل‌های ریاضی و تکنیک‌های محاسباتی مرتبط با شناسایی و آنالیز امضاهایی جهشی، دقت را بهبود بخشیده و در زمینه شناسایی امضاهای جهشی در تومورهای سرطانی پیشرفت بیشتری ایجاد کنیم. مقالاتی که مبنای اصلی کار این طرح تحقیقاتی قرار دارد شامل 2 مقاله پایه‌ای در حیطه امضاهایی جهشی شامل مقاله چاپ شده توسط L. B. Alexandrov, S. Nik-Zainal و همکارانشان در سال 2012 و 2013 (2, 10)، همچنین 4 مقاله از جدیدترین و پیشرفته‌ترین رویکردهای بکار برده شده در این زمینه در سال‌های اخیر است (21, 28, 44, 93).

## جنبه جدید بودن و نوآوری:

زیبایی چارچوب امضای جهشی فقط در الگوریتم‌های ریاضی آن نهفته نیست زیرا این الگوریتم‌ها اغلب به سادگی رویکردهای تجزیه ماتریسی هستند، بلکه به نحوه طبقه‌بندی جهش‌ها قبل از فاکتورسازی<sup>95</sup> هم مربوط است. جایگزین‌های تک باز، با ترکیب زمینه‌های توالی طرفین هر جایگزین احتمالی طبقه‌بندی می‌شدند، که منجر به یک الگوی 96 کانالی برای SBSs شد (10, 94). امضاهای جایگزین دو بازی<sup>96</sup> (DBS) توسط 78 کانال تعریف می‌شوند (20, 95) و امضاهای بازآرایی دارای الگوی 32 کانالی است (25). تعداد بیش از حد کانال‌های فاقد اطلاعات مهم قدرت تشخیص امضا را کاهش می‌دهد. برعکس، کانال‌هایی که خیلی کم هستند ممکن است

<sup>88</sup> true positive (TP)

<sup>89</sup> false positive (FP)

<sup>90</sup> false negative (FN)

<sup>91</sup> ground truth

<sup>92</sup> precision

<sup>93</sup> sensitivity

<sup>94</sup> F1 score

<sup>95</sup> decomposition

<sup>96</sup> Double-base substitution signatures (DBSs)

احتمال تشخیص بیولوژیکی جدید را کاهش دهند. از طرفی کانال‌هایی که بیش از حد پیچیده هستند، قابلیت استفاده را کاهش می‌دهند و احتمالاً به امضاهای مختلط<sup>97</sup> منجر می‌شوند و تفسیر را به چالشی غیرضروری تبدیل می‌کنند.

در این طرح تحقیقاتی، سعی بر این است که با کمک ارائه راهکارهای نوین در طبقه‌بندی جهش‌ها یا مدل‌های ریاضی و تکنیک‌های محاسباتی که اساس آنالیزهای امضای جهشی را تشکیل می‌دهند (21, 28, 96)، جهت ارتقاء دقت تحلیل‌ها و پیشبرد در زمینه شناسایی امضاهای جهشی در تومورهای سرطانی گامی برداشته شود. امید است تا این گام به دستاوردهای معتبری در زمینه تشخیص و درمان سرطان منجر گردد و به بهبود سلامت و کیفیت زندگی بیماران سرطانی کمک کند.

---

<sup>97</sup> mixed signatures

- .1 Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-24.
- .2 Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-21.
- .3 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. Cancer genome landscapes. *science*. 2013;339(6127):1546-58.
- .4 Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, et al. Genetic progression and the waiting time to cancer. *PLoS computational biology*. 2007;3(11):e225.
- .5 Attolini CSO, Michor F. Evolutionary theory of cancer. *Annals of the New York Academy of Sciences*. 2009;1168(1):23-51.
- .6 Yates LR, Campbell PJ. Evolution of the cancer genome. *Nature Reviews Genetics*. 2012;13(11):795-806.
- .7 Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current opinion in genetics & development*. 2014;24:52-60.
- .8 Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. *Nature Reviews Cancer*. 2014;14(12):786-800.
- .9 Pfeifer GP. Environmental exposures and mutational patterns of cancer genomes. *Genome medicine*. 2010;2:1-4.
- .10 Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979-93.
- .11 MULLER H, editor Further studies on the nature and causes of gene mutations. *Proceedings of the 6th International Congress of Genetics, 1932; 1932*.
- .12 Bauer H, Demerec M, Kaufmann BP. X-ray induced chromosomal alterations in *Drosophila melanogaster*. *Genetics*. 1938;23(6):610.
- .13 Sax K. Chromosome aberrations induced by X-rays. *Genetics*. 1938;23(5):494.
- .14 Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*. 2012;150(6):1121-34.
- .15 Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*. 2002;21(48):7435-51.
- .16 Harris RS. Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. *Genome medicine*. 2013;5:1-3.
- .17 Hunter C, Smith R, Cahill DP, Stephens P, Stevens C, Teague J, et al. A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer research*. 2006;66(8):3987-91.
- .18 Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nature reviews genetics*. 2014;15(9):585-98.
- .19 Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*. 2015;9(1):3-3.
- .20 Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101.
- .21 Islam SA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics*. 2022;2(11).
- .22 Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nature cancer*. 2020;1(2):249-63.
- .23 Ledford H. Trove of tumour genomes offers clues to cancer origins. *Nature*. 2022;604(7907):609.-
- .24 Degasperi A, Zou X, Dias Amarante T, Martinez-Martinez A, Koh GCC, Dias JM, et al. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science*. 2022;376(6591):ab19283.
- .25 Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47-54.
- .26 Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, et al. The topography of mutational processes in breast cancer genomes. *Nature communications*. 2016;7(1):11.383
- .27 Secrier M, Li X, De Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nature genetics*. 2016;48(10):1131-41.



- .28 Díaz-Gay M ,Vangara R, Barnes M, Wang X, Islam SA, Vermes I, et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics*. 2023;39(12):btad756.
- .29 Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nature medicine*. 2017;23(4):517-25.
- .30 Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nature genetics*. 2014;46(2):116-25.
- .31 Hoang PH, Cornish AJ, Dobbins SE, Kaiser M, Houlston RS. Mutational processes contributing to the development of multiple myeloma. *Blood cancer journal*. 2019;9(8):60.
- .32 Georgeson P, Pope BJ, Rosty C, Clendenning M, Mahmood K, Joo JE, et al. Evaluating the utility of tumour mutational signatures for identifying hereditary colorectal cancer and polyposis syndrome carriers. *Gut* . 49-2138:(11)70;2021
- .33 Grolleman JE, De Voer RM, Elsayed FA, Nielsen M, Weren RD, Palles C, et al. Mutational signature analysis reveals NTHL1 deficiency to cause a multi-tumor phenotype. *Cancer cell*. 2019;35(2):256-66. e5.
- .34 Georgeson P, Harrison TA ,Pope BJ, Zaidi SH, Qu C, Steinfeldt RS, et al. Identifying colorectal cancer caused by biallelic MUTYH pathogenic variants using tumor mutational signatures. *Nature communications*. 2022;13(1):3254.
- .35 Levatić J, Salvadores M, Fuster-Tormo F, Supek F .Mutational signatures are markers of drug sensitivity of cancer cells. *Nature communications*. 2022;13(1):2926.
- .36 Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. *Nature communications*. 2015;6(1):8683.
- .37 Fox EJ, Salk JJ, Loeb LA. Exploring the implications of distinct mutational signatures and mutation rates in aging and cancer. *Genome medicine*. 2016;8(1):1-3.
- .38 Li X, Wu WK, Xing R, Wong SH, Liu Y, Fang X, et al. Distinct subtypes of gastric cancer defined by molecular characterization include novel mutational signatures with prognostic capability. *Cancer research*. 2016;76(7):1724-32.
- .39 Poon S, Huang M, Choo Y, McPherson J, Yu W, Heng H, et al. Mutation signatures implicate aristolochic acid in bladder cancer development. *Genome Med* 7: 38. 2015.
- .40 Poon SL, McPherson JR, Tan P, Teh BT, Rozen SG. Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. *Genome medicine*. 2014;6:1-14.
- .41 Koh G, Degasperis A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nature reviews cancer*. 2021;21(10):619-37.
- .42 Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome biology*. 2016;17(1):1-11.
- .43 Bergstrom EN, Huang MN, Mahto U, Barnes M, Stratton MR, Rozen SG, et al .SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC genomics*. 2019;20(1):1-12.
- .44 Jin H, Gulhan DC, Geiger B, Ben-Isvy D, Geng D, Ljungström V, et al. Accurate and sensitive mutational signature analysis with MuSiCal. *Nature Genetics*. 2024:1-12.
- .45 Taylor-Weiner A, Aguet F, Haradhvala NJ, Gosai S, Anand S, Kim J, et al. Scaling computational genomics to millions of individuals with GPUs. *Genome biology*. 2019;20(1):1-5.
- .46 Kasar S, Kim J, Improgo R ,Tiao G, Polak P, Haradhvala N, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nature communications*. 2015;6(1):8866.
- .47 Omichessan H, Severi G, Perduca V .Computational tools to detect signatures of mutational processes in DNA from tumours: a review and empirical comparison of performance. *PloS one*. 2019;14(9):e0221235.
- .48 Schwarz G. Estimating the dimension of a model. *The annals of statistics*. 1978:461-4.
- .49 Fischer A, Illingworth CJ, Campbell PJ, Mustonen V. EMu: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome biology*. 2013;14(4):1-10.
- .50 Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC bioinformatics*. 2010;11(1):1-9.
- .51 Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. 2018;28(11):1747-56.
- .52 Blokzijl F, Janssen R ,Van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome medicine*. 2018;10:1-11.

- .53 Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization .Proceedings of the national academy of sciences. 2004;101(12):4164-9.
- .54 Fantini D, Vidimar V, Yu Y, Condello S, Meeks JJ. MutSignatures: an R package for extraction and analysis of cancer mutational signatures. Scientific Reports. 2020;10(1):18217.
- .55 Ardin M, Cahais V, Castells X, Bouaoun L, Byrnes G, Herceg Z, et al. MutSpec: a Galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. BMC bioinformatics. 2016;17:1-10.
- .56 Thorndike RL. Who belongs in the family? Psychometrika. 1953;18(4):267-76.
- .57 Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. Journal of statistical software. 2017;76.
- .58 Gori K, Baez-Ortega A. sigfit: flexible Bayesian inference of mutational signatures. bioRxiv. 2018:372896.
- .59 Tan VY, Févotte C. Automatic relevance determination in nonnegative matrix factorization with the/spl beta/-divergence. IEEE transactions on pattern analysis and machine intelligence. 2012.605-1592:(7)35;
- .60 Wang S, Li H, Song M, Tao Z, Wu T, He Z, et al. Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. Plos Genetics. 2021;17(5):e1009557.
- .61 Dempster AP ,Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society: series B (methodological). 1977;39(1):1-22.
- .62 Suri P, Roy NR, editors. Comparison between LDA & NMF for event-detection from large text stream data. 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT); 2017: IEEE.
- .63 Rosales RA, Drummond RD, Valieris R, Dias-Neto E, Da Silva IT. signeR: an empirical Bayesian approach to mutational signature discovery. Bioinformatics. 2017;33(1):8-16.
- .64 Nebgen BT, Vangara R, Hombrados-Herrera MA, Kuksova S, Alexandrov BS. A neural network for determination of latent dimensionality in non-negative matrix factorization. Machine Learning: Science and Technology. 2021;2(2):025012.
- .65 Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics. 2007;23(9):1164-7.
- .66 Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics. 2015;31(22):3673-5.
- .67 Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:160304467. 2016.
- .68 Vöhringer H, Hoeck AV, Cuppen E, Gerstung M. Learning mutational signatures and their multidimensional genomic properties with TensorSignatures. Nature communications. 2021;12(1):3628.
- .69 Lee DD, Seung HS .Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401(6755):788-91.
- .70 Févotte C, Cemgil AT, editors. Nonnegative matrix factorizations as probabilistic inference in composite models. 2009 17th European Signal Processing Conference; 2009: IEEE.
- .71 Craig MD. Minimum-volume transforms for remotely sensed data. IEEE Transactions on Geoscience and Remote Sensing. 1994;32(3):542-52.
- .72 Miao L, Qi H. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. IEEE Transactions on Geoscience and Remote Sensing. 2007;45(3):765-77.
- .73 Ang AMS, Gillis N. Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2019;12(12):4843-53.
- .74 Leplat V, Gillis N, Ang AM. Blind audio source separation with minimum-volume beta-divergence NMF. IEEE Transactions on Signal Processing. 2020;68:3400-10.
- .75 Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. BioMed Central; 2018.
- .76 Manders F, Brandsma AM, de Kanter J, Verheul M, Oka R, van Roosmalen MJ, et al. MutationalPatterns: the one stop shop for the analysis of mutational processes. BMC genomics. 2022;23(1):134.
- .77 Li S, Crawford FW, Gerstein MB. Using sigLASSO to optimize cancer mutation signatures jointly with sampling likelihood. Nature communications. 2020.3575:(1)11;
- .78 Díaz-Gay M, Vila-Casadesús M, Franch-Expósito S, Hernández-Illán E, Lozano JJ, Castellví-Bel S. Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples. BMC bioinformatics.6-1:(1)19;2018 .
- .79 Lee J, Lee AJ, Lee J-K, Park J, Kwon Y, Park S, et al. Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. Nucleic acids research. 2018;46(W1):W102-W8.

- .80 Kiefer J. Sequential minimax search for a maximum. *Proceedings of the American mathematical society*. 1953;4(3):502-6.
- .81 Borchers HW. Pracma: Practical numerical math functions (2.4. 2) 2022 [Available from: <https://CRAN.R-project.org/package=pracma> last accessed.
- .82 Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*. 1944;2(2):164-8.
- .83 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*. 2010;33(1):1.
- .84 Gorski J, Pfeuffer F, Klamroth K. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical methods of operations research*. 2007;66:373-407.
- .85 Lin X, Boutros PC. Optimization and expansion of non-negative matrix factorization. *BMC bioinformatics*. 2020;21(1):1-10.
- .86 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*. 2020;17(3):261-72.
- .87 Ling RF. Solving least squares problems. JSTOR; 1977.
- .88 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic acids research*. 2019;47(D1):D941-D7.
- .89 Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, et al. A pan-cancer compendium of chromosomal instability. *Nature*. 2022;606(7916):976-83.
- .90 Steele CD, Abbasi A, Islam SA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022;606(7916):984-91.
- .91 Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*: Springer; 2009.
- .92 Lawson CL, Hanson RJ. *Solving least squares problems*: SIAM; 1995.
- .93 Khandekar A, Vangara R, Barnes M, Diaz-Gay M, Abbasi A, Bergstrom EN, et al. Visualizing and exploring patterns of large mutational events with SigProfilerMatrixGenerator. *bioRxiv*. 2023:2023.02. 03.527015.
- .94 Koh G, Zou X, Nik-Zainal S. Mutational signatures: experimental design and analytical framework. *Genome biology*. 2020;21:1-13.
- .95 Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A compendium of mutational signatures of environmental agents. *Cell*. 2019;177 .36-821:(4)7e16.
- .96 Jin H, Gulhan DC, Geiger B, Ben-Isvy D, Geng D, Ljungstrom V, et al. Accurate and sensitive mutational signature analysis with MuSiCal. *bioRxiv*. 2022:2022.04. 21.489082.

عنوان مصوب:

Approved title:

کمیته تخصصی گروه

نام و نام خانوادگی	عنوان	رتبه علمی	رای	امضا

توضیحات

--

امضای مدیر گروه  
تاریخ

تکمیل این قسمت اجباری می باشد موضوع پایان نامه طبق لیست پیوست جزء بخش دوم گروه  
..... بند ..... اولویتهای تحقیقاتی طبق بخشنامه شماره 30/51682 مورخ 89/7/6  
و 30/51185 مورخ 89/7/5 می باشد:  
عنوان اولویت:

تاریخ تصویب در شورای پژوهشی  
دانشکده/گروه تخصصی:

تکمیل این قسمت اجباری می باشد.موضوع پایان نامه طبق لیست پیوست جزء بخش دوم گروه  
..... بند ..... اولویتهای تحقیقاتی طبق بخشنامه شماره 30/51682 مورخ  
89/7/6 و 30/51185 مورخ 89/7/5 می باشد:  
عنوان اولویت:

امضاء رئیس شورا/مدیر گروه