



دانشگاه تهران

دانشکده ریاضی، آمار و علوم کامپیوتر

تعیین نقش دومین پروتئین‌ها در سرطان‌ها بعد از بهینه کردن خصیصه‌ها

نگارش

سیرانا هاشمی رنجبر

اساتید راهنما

دکتر عباس نوذری دالینی

دکتر زهرا رزاقی مقدم کاشانی

استاد مشاور

دکتر علی محمد بنائی مقدم

رساله برای دریافت درجه کارشناسی ارشد

در رشته علوم کامپیوتر

۱۳۹۵ تیر

تَعْدِيْمُهُمْ كَسَانِي كَه
در سختی ها و ناملایات زندگی پشتیبان و مایه دل کرمی ام بودند.

چکیده

یکی از بزرگ‌ترین چالش‌های علوم زیستی حال حاضر، شناختن عوامل ایجاد سرطان و راه‌های پشگیری و درمان آن است. سرطان به دلایل مختلفی رخ می‌دهد که مهمترین آن به دلیل تغییرات ایجاد شده بر روی ژن‌های انسان است. وقتی ژنی معیوب شود پروتئینی که از روی آن کدنویسی می‌شود، هم می‌تواند دچار نقص کارکرد شود. پروتئین‌ها از حیاتی‌ترین درشت‌مولکول‌ها در طبیعت هستند که در ساخت و ساز سلول‌های بدن انسان نقش یسار مهمی دارند. هر قطعه از پروتئین که می‌تواند به صورت غیر وابسته به یک ساختار فشرده و پایدار چین بخورد دومین پروتئین نام داد که می‌تواند به تنها‌یی یا در ارتباط با سایر دومین‌های پروتئین فعالیت مشترکی را انجام دهد. از این رو شناخت اثر تغییرات ایجاد شده در ژنوم بر روی دومین‌های پروتئینی می‌تواند به شناخت عوامل ایجاد یا توسعه سرطان کمک کند. از آنجایی که تعداد ژن‌ها و پروتئین‌های انسانی زیاد است و رابطه پیچیده‌ای در تراکنش میان پروتئین‌ها و دومین‌های آن‌ها وجود دارد و یافتن عوامل اصلی ایجاد سرطان به روش‌های آزمایشگاهی بسیار مشکل است، مسئله یافتن دومین‌ها، پروتئین‌ها و ژن‌های دخیل در سرطان توسط نرم افزار و به کارگیری دانش رایانه و آمار مطرح می‌شود. در این پایان نامه مسئله یافتن دومین‌های دخیل در ایجاد و توسعه سرطان بررسی شده است. به این صورت که ابتدا میزان احتمال رخدادن جهش‌ها با روش پیشنهادی بر روی ژنوم محاسبه شده و با تعیین یک حد آستانه به مشخص کردن دومین‌های کاندید هر بیماری پرداخته شده است. به علاوه، بر اساس دومین‌های کاندید، به معرفی ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده کاندید هر سرطان نیز پرداخته شده است.

كلمات کلیدی: دومین پروتئین، سرطان، کلاس‌بندی، کت، پی‌فم، میتوکندری، سلول بنیادی، سلول ترمیم کننده، مطالعه یکپارچه سرطان، مدل آماری.

پیش‌گفتار

سرطان‌ها علاوه بر عامل اصلی مرگ و میر در دنیا، جز بیماری‌هایی هستند که درمان آن‌ها پروسه‌ای طولانی مدت و پرهزینه‌ای دارد^[۱] و این بیماری در اکثر موارد به نقص عضو فرد بیمار انجامیده و سال‌ها زندگی او را تحت تاثیر قرار می‌دهد. بنابراین سال‌هاست محققان سعی در یافتن راههایی برای شناخت زود هنگام سرطان‌ها و یافتن درمان‌های پیشگیرانه یا سرکوب کننده سرطان‌ها هستند^[۲]، تا تلفات جانی و مادی را به حداقل برسانند. با وجود تلاهای زیادی که انجام شده و موفقیت‌هایی که کسب شده هنوز هم، هر روزه افراد بسیاری به این بیماری مهلك دچار شده و جان خود را از دست می‌دهند. در چندین سال اخیر با کدگشایی از ژنوم مرجع، برخی موسسات شروع به جمع آوری داده‌های مربوط به ژنوم بیماران سلطانی کردند تا با مطالعه و مقایسه آن‌ها با ژنوم مرجع به شناسایی عوامل ایجاد و توسعه سرطان‌ها بپردازند^[۳]. یکی از روش‌هایی که برای این منظور به کار می‌رود شناسایی ژن‌های دخیل در سرطان است. اما بسیاری معتقدند مطالعه ژن‌ها برخی عوامل ایجاد سرطان‌ها را نادیده می‌گیرد و آن دومین‌های پروتئینی در سلول‌هاست که جهش‌های موجود بر روی آن‌ها می‌تواند مشکلاتی در سلول ایجاد کند و حتی ممکن است باعث ایجاد بیماری شود^[۴]. بنابراین برخی از مطالعات به این سمت میل کرده است. در مطالعه نقش دومین‌های پروتئینی در سرطان، جهش‌های موجود بر روی مناطق کدکننده ژن‌های کدکننده پروتئین به پروتئینی که آن ژن‌ها می‌سازند نگاشت داده می‌شود. جهش‌هایی که بعد از نگاشت بر روی دومین واقع می‌شوند برای مطالعه انتخاب می‌شوند. در این پایان نامه هم از این روش برای شناسایی عوامل ایجاد و توسعه سرطان‌ها استفاده شده است. برای این کار جهش‌های روی ژنوم بیماران به دومین‌های پروتئینی نگاشت می‌شود و سپس از روши برای تعیین جهش‌های راننده و تمایز آن‌ها از جهش‌های مسافر مشخص می‌گردد. در اینجا از توزیع دو جمله‌ای برای محاسبه احتمال رخدادن تصادفی جهش‌ها بر روی دومین‌ها استفاده شده است. سپس با توزیع نرمال و محاسبه خطای جمعی به معرفی دومین‌های دخیل در سرطان پرداخته می‌شود. از آن جایی که تعداد دومین‌های ساختار محور که در آزمایشگاه شناسایی می‌شوند، کم است، از دومین‌های توالی محور که توسط نرم افزار تهیه می‌شوند استفاده می‌شود. در اینجا از دومین‌های پی‌فم به عنوان دومین‌های توالی محور استفاده شده و از دومینهای کت برای محک زدن میزان درستی دومین‌های پی‌فم و قابلیت اطمینان نتایج مطالعه

بر روی دومین‌های پی‌فم استفاده شده است. سه عاملی که به نظر می‌رسد در ایجاد سرطان‌ها دخیل باشند در کنار مطالعه دومین‌ها، مورد ارزیابی قرار گرفته‌اند. اولین عامل دی‌ان‌آ میتوکندری است [۵]. با توجه به ساختار خاص میتوکندری و وظایفی که در سلول دارد، توجه بسیاری به نقش آن در بیماری‌های مختلف از جمله سرطان می‌شود [۶]. در اینجا نیز، همه جهش‌های موجود بر روی دی‌ان‌آ میتوکندری بیماران مربوط به همه ۳۷ ژن آن مورد مطالعه قرار گرفت و نشان داده شد در ایجاد برخی از سرطان‌ها میتوکندری دخالت دارد اما در برخی دیگر هیچ تاثیری ندارد. عامل دوم ژن‌های سلول‌های بنیادی هستند یعنی ژن‌هایی که بیان آن‌ها باعث می‌شود سلول به مانند سلول بنیادی رفتار کند. مجموعه‌های مختلفی با نام ژن‌های سلول‌های بنیادی ارائه شده‌اند. در اینجا با انتخاب یکی از مجموعه‌های موجود به مطالعه نقش این ژن‌ها در ایجاد و توسعه سرطان پرداخته شده است. با بررسی انجام شده مشاهده شد که تعداد کمی ژن از این مجموعه در ایجاد سرطان‌ها دخیل‌اند. عامل سوم ژن‌های ترمیم کننده هستند. با این ژن‌ها هم درست مانند ژن‌های سلول‌های بنیادی رفتار شده است. در مورد این مجموعه هم نتایج مانند ژن‌های سلول‌های بنیادی است. تعداد کمی ژن از این مجموعه در ایجاد سرطان‌ها دخیل‌اند. مطالعه دیگری هم در اینجا انجام شده و آن مطالعه یکپارچه سرطان‌هاست. یعنی تمامی سرطان‌های مورد مطالعه در این پایان نامه همانند یک نوع سرطان یکسان و واحد در نظر گرفته شده و دومین‌های کاندید دخیل در ایجاد و توسعه آن پیدا شده و هر سه عامل ذکر شده در بالا، برای مطالعه یکپارچه هم مورد ارزیابی قرار گرفته است.

این پایان نامه در چهار فصل آماده شده است. در فصل اول به بیان مفاهیم زیستی مورد نیاز مسئله مانند سلول، دی‌ان‌آ، پروتئین، ژن، جهش، مفهوم میتوکندری، سلول بنیادی و ژن ترمیم کننده پرداخته و مقدمات آماری مورد نیاز شرح داده شده است. در فصل دوم، به معرفی سرطان، مطالعات قبلی انجام شده در مورد سرطان‌ها و یافتن ژن‌ها و دومین‌های دخیل در سرطان، مطالعات مربوط به ارتباط جهش‌های میتوکندریابی و سرطان پرداخته اسامی ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده که از مطالعات مختلف تهیه شده توضیح داده شده است. همچنین به معرفی مطالعاتی که به مقایسه دومین‌های ساختار محور و توالی محور پرداخته‌اند اشاره شده است. سابقه کمی هم از مطالعات یکپارچه سرطان‌ها ذکر شده است. در فصل سه، به معرفی یکی از روش‌های مطالعه دومین محور سرطان [۷] پرداخته شده است. در آخر و در فصل چهار، به معرفی روش پیشنهادی برای شناسایی دومین‌های دخیل در سرطان پرداخته شده و به ذکر نتایج مطالعه سه عامل ذکر شده در بالا و ارزیابی آن‌ها پرداخته می‌شود. با مقایسه دومین‌های کت و پی‌فم میزان قابلیت اطمینان روش پی‌فم ارزیابی شده و در بخش آخر به معرفی وبسایتی می‌پردازد که برای بارگیری نتایج این آزمون و مشاهده گرافیکی نتایج آن ساخته شده است.

با مشکر از دوست بسیار عزیزم

آدین

که با نظرات و راهنمایی‌های سازنده اش مرایاری داد.

با مشکر از دوست کرامی ام

فرشاد جعفری

بایت‌یاری اش در طراحی و پیاده‌سازی وب‌سایت پایان‌نامه

و با مشکر از

مرکز محاسبات ابری دانشگاه صنعتی امیرکبیر

فهرست مطالب

۱	مفاهیم زیستی مقدمات و تعاریف اولیه
۱	۱.۱ سلول
۳	۲.۱ هسته سلول
۴	۳.۱ دی.ان.آ
۷	۴.۱ چرخه سلولی
۹	۵.۱ آر.ان.آ
۱۰	۶.۱ پروتئین
۱۲	۷.۱ تبدیل دی.ان.آ به پروتئین
۱۶	۸.۱ ایزوفرم‌های پروتئینی
۱۸	۹.۱ میتوکندری
۱۸	۹.۱.۱ ژنوم میتوکندری
۲۱	۹.۱.۱.۱ نقش زیستی میتوکندری
۲۲	۱۰.۱ انواع تغییرات توالی دی.ان.آ
۲۲	۱۰.۱.۱ چند ریختی در یک نوکلئوتید واحد
۲۳	۱۰.۱.۱.۱ جهش
۲۵	۱۱.۱ تغییرات ژنتیکی و غیر ژنتیکی سلول
۲۶	۱۲.۱ سلول‌های بنیادی
۲۷	۱۲.۱.۱ تعریف سلول‌های بنیادی
۲۸	۱۲.۱.۱.۱ سلول‌های بنیادی سرطان
۲۸	۱۳.۱ ترمیم دی.ان.آ
۲۹	۱۳.۱.۱ آسیب دی.ان.آ و جهش

۳۰	مکانیسم‌های بازسازی دی.ان.آ.	۲.۱۳.۱
۳۰	دومین‌های پروتئینی	۱۴.۱
۳۱	آمار و کاربرد آن در بیوانفورماتیک	۱۵.۱
۳۲	توزیع احتمال دوجمله‌ای	۱.۱۵.۱
۳۳	توزیع نرمال	۲.۱۵.۱
۳۳	آزمون فرض	۳.۱۵.۱
۳۴	آزمون فرضیه‌های متعدد	۴.۱۵.۱
۳۵	جمع‌بندی	۱۶.۱
۳۶	۲ تعریف مسئله و پیشینه تحقیق	
۳۶	سرطان	۱.۲
۳۹	میتوکندری و سرطان	۲.۲
۴۷	ژن‌های سلول‌های بنیادی و سرطان	۳.۲
۵۰	بازسازی دی.ان.آ و سرطان	۴.۲
۵۰	مطالعه یکپارچه سرطان‌ها	۵.۲
۵۱	مقایسه دومین‌های توالی محور و ساختار محور	۶.۲
۵۴	دومین‌های پروتئینی و سرطان	۷.۲
۵۶	بانک‌های اطلاعاتی مرتبط	۸.۲
۵۶	بانک اطلاعاتی اطلس ژنوم سرطانی	۱.۸.۲
۵۸	بانک اطلاعاتی ژنوم مرجع	۲.۸.۲
۶۰	بانک اطلاعاتی کت	۳.۸.۲
۶۱	بانک اطلاعاتی اسکاپ	۴.۸.۲
۶۱	بانک اطلاعاتی پی‌فم	۵.۸.۲
۶۲	بانک اطلاعاتی پروتئین	۶.۸.۲
۶۲	بانک اطلاعاتی یونیپرات	۷.۸.۲
۶۳	بانک اطلاعاتی یوسیاس‌سی	۸.۸.۲
۶۳	بانک اطلاعاتی هوگو	۹.۸.۲
۶۴	بانک اطلاعاتی رفسک	۱۰.۸.۲

۶۴	بانک اطلاعاتی منبع ادغام ساختارها و توالی‌ها	۱۱.۸.۲
۶۴	تعریف مسئله	۹.۲
۶۵	جمع‌بندی	۱۰.۲
۳ نگاشت دومین‌ها و نقدی بر روش شناسایی		
۶۶	نگاشت دومین‌های پروتئینی جهش‌های سوماتیک بیماران سرطانی	۱.۳
۶۷	معرفی روش	۱.۱.۳
۷۱	نقشه جهش ژنی و دومینی	۲.۱.۳
۷۳	مقایسه نقشه سرطانی، سرطان‌های پستان و روده بزرگ	۳.۱.۳
۷۷	روش‌ها و ابزارهای تهیه نقشه جهش ژنی و دومین	۴.۱.۳
۷۹	نتیجه‌گیری	۲.۳
۴ روش بیشنهادی و نتایج		
۸۰	جمع‌آوری داده‌ها و آماده کردن آن‌ها برای مطالعه	۱.۴
۸۲	روش امتیازدهی به جهش‌ها	۲.۴
۸۶	ژن‌های میتوکندری	۳.۴
۹۳	ژن‌های سلول‌های بنیادی	۴.۴
۹۷	ژن‌های ترمیم کننده	۵.۴
۱۰۱	نقش دومین‌ها در سرطان	۶.۴
۱۰۱	دومین‌های کت	۱۶.۴
۱۰۲	دومین‌های پی‌فیم	۲۶.۴
۱۰۴	نتیجه‌گیری و مقایسه دومین‌های ساختاری و دنباله‌ای	۷.۴
۱۰۷	معرفی وب سایت	۸.۴
۱۰۸	بخش اول، جدول‌های مورد نیاز مطالعه	۱۸.۴
۱۰۹	بخش دوم، نمایش گرافیکی اطلاعات	۲۸.۴
۱۰۹	بخش سوم، جدول‌های تهیه شده از تحلیل نتایج	۳۸.۴
۱۱۰	بخش دوم، نمایش گرافیکی نتایج	۴۸.۴
۱۱۰	جمع‌بندی	۹.۴

۱۲۰

آ جدول‌ها

۱۴۸

واژه‌نامه فارسی به انگلیسی

۱۵۶

واژه‌نامه انگلیسی به فارسی

فهرست جداول

۱۷	جدول علائم نام‌گذاری پروتئین‌ها ^[۸]	۱.۱
۲۰	اسامی ژن‌های میتوکندری و اطلاعات مربوط به آن‌ها	۲.۱
۴۵	نتایج مطالعات مختلف بر روی جهش‌های میتوکندری در انواع سرطان‌ها ^[۹]	۱.۲
۴۶	اسامی رونوشت‌های نامتعارف میتوکندری	۲.۲
۵۲	ژن‌های انتخاب شده برای مطالعه ^[۷]	۳.۲
۶۸	تعداد جهش‌های سرطان روده ^[۷]	۱.۳
۷۰	اسامی دومین‌های معنادار سرطان‌های روده بزرگ و پستان ^[۷]	۲.۳
۷۲	ژن‌های انتخاب شده با تعداد جهش‌های بالا در سرطان روده بزرگ ^[۷]	۳.۳
۷۳	دومین‌های پروتئینی انتخاب شده با تعداد جهش بالا ^[۷]	۴.۳
۸۳	تعداد بیماران هر سرطان با حداقل یک جهش بر روی اگزوم و پروتئین دومین‌ها	۱.۴
۸۷	تعداد بیماران دارای جهش‌های میتوکندریایی و تعداد کل جهش‌های میتوکندریایی در هر بیماری	۲.۴
۱۱۱	ژن‌ها و مجموعه‌های کاندید سرطان‌ها	۳.۴
۱۱۲	تعداد جهش‌های مربوط به رونوشت‌های نامتعارف میتوکندری	۴.۴
۱۱۳	ژن‌های سلول‌های بنیادی کاندید سرطان‌ها	۵.۴
۱۱۴	ژن‌های ترمیم کننده کاندید سرطان‌ها	۶.۴
۱۱۵	ژن‌های سلول‌های بنیادی با حداقل یک دومین کاندید کت در هر سرطان	۷.۴
۱۱۶	ژن‌های ترمیم کننده با حداقل یک دومین کاندید کت در هر سرطان	۸.۴
۱۱۷	ژن‌های سلول‌های بنیادی با حداقل یک دومین کاندید پی‌فم در هر سرطان	۹.۴
۱۱۸	ژن‌های ترمیم کننده با حداقل یک دومین کاندید پی‌فم در هر سرطان	۱۰.۴

۱۲۰	اسامی فایل‌های دانلود شده برای بیماری‌های مختلف	۱.آ
۱۲۲	اسامی کامل و مخفف انواع سرطان‌ها	۲.آ
۱۲۲	لینک دانلود داده‌های مربوط به ژنوم ، پروتئین و دومین‌های پروتئینی	۳.آ
۱۲۳	اسامی ژن‌های سلول‌های بنیادی منتخب [۱۰]	۴.آ
۱۲۴	اسامی ژن‌های ترمیم کننده منتخب [۱۱، ۱۲، ۱۳]	۵.آ
۱۲۵	اسامی دومین‌های کاندید کت سرطان‌ها	۶.آ
۱۲۸	اسامی دومین‌های کاندید پی‌فم سرطان‌ها	۷.آ

فهرست تصاویر

۳	نمونه یک سلول و اجزای آن [۱]	۱.۱
۴	نمونه ای از یک هسته سلول [۱]	۲.۱
۵	فرمول شیمیایی بازهای آلی [۱]	۳.۱
۶	رشته دی‌ان‌آ [۱۴]	۴.۱
۹	مراحل تقسیم سلولی میوز و میتوز [۵]	۵.۱
۱۱	تصویر سه بعدی ساختارهای پروتئین [۱]	۶.۱
۱۳	شکل مراحل تولید پروتئین از روی دی‌ان‌آ [۱]	۷.۱
۱۴	مراحل رونویسی آر‌ان‌آ از روی دی‌ان‌آ [۱]	۸.۱
۱۵	ترجمه آر‌ان‌آ به پروتئین [۱]	۹.۱
۱۸	نحوه ایجاد ایزوفرم‌های مختلف یک پروتئین [۱۵]	۱۰.۱
۲۳	اسنیپ‌ها بر روی ژنوم [۱۶]	۱۱.۱
۲۴	دسته‌بندی انواع جهش‌ها [۱۷]	۱۲.۱
۲۹	شكلی از سلول‌های بنیادی سرطانی و درمان‌های خاص [۱۸]	۱۳.۱
۳۷	نشانه‌های سلول‌های سرطانی [۱۹]	۱.۲
۵۳	نحوه همپوشانی دومین‌ها بر روی پروتئین [۲۰]	۲.۲
۵۴	نگاشت دومین‌های پی‌فم به دومین‌های اسکاپ [۲۱]	۳.۲
۵۴	نگاشت جهش‌ها به پروتئین دومین‌ها [۲۲]	۴.۲
۵۵	نقاط جهش‌یافته خاص انواع سرطان [۲۲]	۵.۲
۷۱	نقشه جهش‌های دومینی و ژنی برای سرطان‌های روده بزرگ و پستان [۷]	۱.۳
۷۴	قله‌های دومینی باقی‌مانده بعد از حذف جهش‌های مربوط به ژن‌های قله [۷]	۲.۳

۷۵	قله‌های ژنی و دومینی مشترک بین سلطان‌های روده بزرگ و پستان [۷]	۳.۳
۱۰۵	نمودار درصد بیماران دارای حداقل یک دومین کاندید کت و پی‌فم در سلطان‌ها	۱.۴
۱۰۶	حالت‌های ممکن میزان همپوشانی دو دومین کت و پی‌فم	۲.۴
۱۰۷	نمایش دومین‌های کت و پی‌فم یک پروتئین	۳.۴

فصل ۱

مفاهیم زیستی مقدمات و تعاریف اولیه

رکن اصلی و شالوده علم بیوانفورماتیک مفاهیم زیستی است، در واقع در این رشته، علوم رایانه برای کمک به حل مسائل زیستی آمده است. برای پیدا کردن جواب‌های درست و قابل اطمینان در این زمینه باید هرچه بیشتر و بهتر با مفاهیم زیستی و آماری مورد نیاز آشنا شد. در این فصل مفاهیم زیستی مرتبط با پایان نامه شامل مفاهیم سلول، دی.ان.آ، آر.ان.آ، پروتئین، تقسیم سلولی، سلول‌های بنیادی، ژن‌های ترمیم کننده، جهش و آمار و احتمال مقدماتی مرور خواهد شد.

۱.۱ سلول

علوم زیستی برای جانداران تعریف می‌شود. آنچه که جانداران مانند گل‌ها، حیوانات و انسان‌ها را از غیر جانداران مانند سنگ و آهن جدا می‌کند این است که همه جانداران از واحدهای منظم تشکیل شده‌اند که سلول^۱ نام دارند. جانداران تولید مثل می‌کنند. از یک شکل ساده شروع به رشد می‌کنند. انرژی و ماده را از محیط گرفته و آن را تبدیل به مواد دیگر می‌کنند. به محرک‌ها پاسخ داده و خود را با محیط سازگار می‌کنند؛ کارهایی که غیر جانداران در انجام آن‌ها ناتوانند. سلول یا یاخته، واحد زیستی، ساختاری و عملگری در همه جانداران شناخته شده است. سلول‌ها کوچکترین واحد اصلی زنده هستند که می‌توانند به طور مستقل جایگزین شوند. ساده‌ترین جانداران، نک سلولی‌ها هستند. انسان‌ها هم مانند بسیاری از موجودات زنده دیگر مانند جانوران، گیاهان و قارچ‌ها از اجتماع سلولی‌های مختلف به وجود آمده‌اند که از رشد و تقسیم یک سلول منفرد اولیه تشکیل شده‌اند که هر کدام از سلول‌ها نقش اختصاصی را به عهده گرفته و در هماهنگی و همکاری با سایر سلول‌ها به موجودات زنده معنا می‌دهند^[۱]. با مطالعه سلول‌ها به عنوان واحدهای حیاتی، زیست شناسان به دنبال یافتن بسیاری از سوالات از جمله یافتن سرمنشأ حیات، دلیل تنوع زیستی جانداران و گیاهان، علت پیری و بیماری‌ها هستند تا با پاسخ به آن‌ها زندگی

^۱Cell

سالمتر و به دور از بیماری را برای انسان‌ها به ارمغان آورده‌اند و از نابودی طبیعت تا جای ممکن جلوگیری کنند. سلول‌ها در عین داشتن شباهت‌های زیاد، کارکرد و تفاوت‌های زیادی نیز دارند. پس باید دانست به عنوان مثال، سلول‌های یک پروانه و یک باکتری چه شباهت‌هایی دارند و یا تفاوت سلول‌های مغز و ریه آدمی در چیست. تفاوت‌هایی که برای سلول‌ها شناسوایی شده است شامل تفاوت در اندازه، شکل و ظاهر، کارکرد، مواد مورد نیاز تغذیه‌ای، توانایی در حرکت، تمایل به همکاری با سایر سلول‌ها و برخی اجزای ساختاری آن‌هاست. برای توضیح بیشتر باید گفت مثلاً بعضی از سلول‌ها دارای غشای نازکی هستند و برخی دیگر در سطح بیرونی غشا یک لایه لزج‌مانند دارند و یا برخی دیگر وظیفه تولید فاکتورهایی مانند مواد چربی یا نشاسته‌ای، هورمون‌ها و یا رنگدانه را بر عهده دارند [۱].

اما در کنار این همه تفاوت، مطالعات زیست‌شناسان و زیست‌شیمی‌شناسان نشان داده است سلول‌ها به طور حیرت‌آوری در اجزاء به یکدیگر شباهت دارند. سلول‌ها از اجزای مختلفی به نام اندامک‌ها^۱ تشکیل شده‌اند، که وظایف مختلفی بر عهده دارند (ممکن است برخی از اندامک‌ها در برخی گونه‌های زیستی موجود نباشند). بخش‌های مختلف سلول عبارت است از: غشاء^۲، هسته^۳، سیتوزول^۴، میتوکندری^۵، کلروپلاست^۶، لیزوژوم‌ها^۷، دستگاه گلزی^۸، پراکسی‌زوم‌ها^۹، وزیکول^{۱۰}، ریبوزوم^{۱۱}، شبکه آندوپلاسمی^{۱۲} و اسکلت سلولی^{۱۳}. در شکل ۱.۱ شمای یک سلول و اجزای آن نشان داده شده است [۱].

اسکلت سلولی، کارکردی مانند داربست یا استخوان‌بندی برای سلول دارد. غشاء سلولی، اطراف همه سلول‌ها را غشائی دربر گرفته است تا سلول و اجزای آن را از محیط اطراف جدا کرده، از آسیب حفظ نموده و شکل واحدی به آن دهد. سیتوپلاسم، داخل فضای سلولی و اطراف اندامک‌های سلول را پوشش می‌دهد. میتوکندری به عنوان تولیدکننده انرژی شیمیایی سلول فعالیت دارد. هسته سلول، بزرگ‌ترین اندامک سلول است که دو وظیفه عمده بر عهده دارد؛ یکی کنترل واکنش‌های شیمیایی داخل سیتوپلاسم و دیگری نگهداری اطلاعات لازم برای

^۱Organelles

^۲Plasma membranes

^۳Cell Nucleus

^۴Cytosol

^۵Cytoplasm

^۶Mitochondrion

^۷Chloroplast

^۸Lysosome

^۹Golgi apparatus

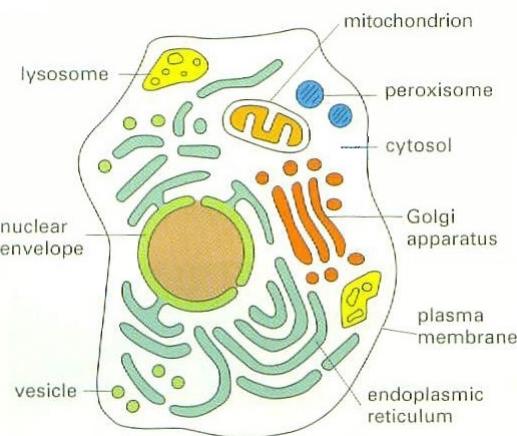
^{۱۰}Peroxisome

^{۱۱}Vesicle

^{۱۲}Ribosome

^{۱۳}Endoplasmic reticulum

^{۱۴}Cytoskeleton



شکل ۱.۱: نمونه یک سلول و اجزای آن [۱]

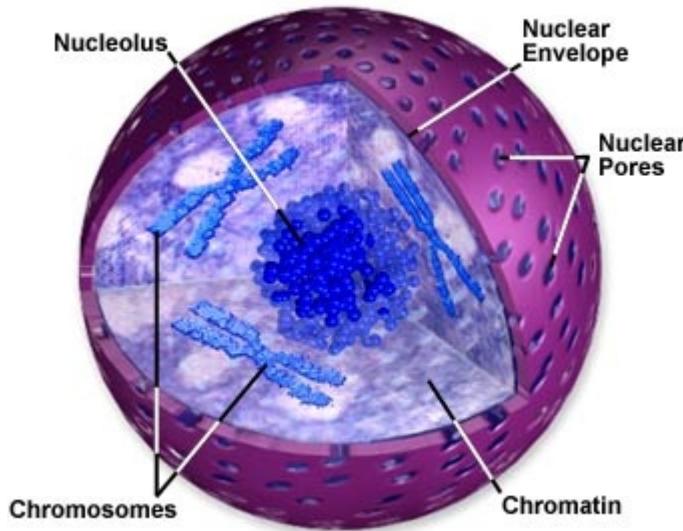
تقسیم سلولی. ریبوزوم به پروتئین‌سازی کمک می‌کند. اگر سلولی در سطح خود ریبوزوم نداشته باشد، وظایفی مانند متابولیسم چربی‌ها، خنثی‌سازی سموم و ذخیره کلسیم را شبکه آندوپلاسمی به عهده می‌گیرد و در صورت وجود ریبوزوم، این شبکه در پروتئین‌سازی دخالت می‌کند. در دستگاه گلتری، بر روی پروتئین‌های ساخته شده در شبکه آندوپلاسمی تغییراتی اعمال می‌شود، مانند افزودن مواد مختلف و یا بریدن قطعات اضافی. لیزوژوم‌ها دستگاه گوارش سلول محسوب شده و قادر به هضم اندامک‌های فرسوده و مواد خارجی وارد شده به سلول هستند. سیتوزول، محل انجام واکنش‌های شیمیایی سلول است. وزیکول‌ها، محلی بر روی غشاء، هستند که تبادلات بین سلولی را انجام می‌دهند و پراکسیزوم‌ها دارای فضایی برای انجام واکنش‌های شیمیایی هستند [۱]. در ادامه به بررسی جزئی‌تر هسته پرداخته می‌شود.

۲.۱ هسته سلول

با توجه به این که هسته وظیفه تکثیر سلول را به عهده دارد، یکی از اساسی‌ترین اجزای سلول است؛ و از آن جایی که برخی سلول‌های برخی جانداران فاقد هسته هستند، این ویژگی یعنی داشتن یا نداشتن هسته در سلول، به عنوان معیاری برای دسته‌بندی جانداران به کار می‌رود. مثلاً سلول باکتری فاقد هسته است. به دسته جانداران که سلول‌هایشان هسته ندارد پروکاریوت^۱ می‌گویند و در مقابل به جاندارانی که سلول‌هایشان هسته دارد مانند انسان، یوکاریوت^۲ می‌گویند. همان‌طور که گفته شد، هسته یکی از اجزای اصلی سلول است که نقش تکثیر سلولی را به عهده دارد. بیشتر پروکاروت‌ها به صورت موجودات تک سلولی زندگی می‌کنند، اما بعضی از آن‌ها با هم در یک زنجیره، خوشی یا سایر ساختارهای چندسلولی سازماندهی شده و زندگی می‌کنند. از نظر شکل و ساختار،

^۱prokaryote

^۲Eukaryote



شکل ۲.۱: نمونه‌ای از یک هسته سلول [۱]

پروکاریوت‌ها ممکن است محدود به نظر برسند اما از لحاظ شیمیایی جزسلول‌های بسیار متنوع و مبتکر به شمار می‌آیند. بیشتر جانداران پیچیده، شامل گیاهان، جانوران و قارچ‌ها از سلول‌های یوکاریوتی ساخته شده‌اند [۱]. هسته با دو غشاء از سیتوپلاسم جدا می‌شود که به آن پوشش هسته^۱ گفته می‌شود. در میان این دو غشاء، منافذی^۲ وجود دارد که تبادل دوطرفه‌ی مواد را میان سیتوپلاسم و هسته امکان‌پذیر می‌کند. درون هسته یک یا دو هستک^۳ وجود دارد که وظیفه اصلی آن تولید پروتئین است. اطراف هستک‌ها را نوکلئوپلاسم^۴ احاطه کرده است. نوکلئوپلاسم مایعی ژلاتینی است که مواد بسیاری از جمله پروتئین را در خود حل می‌کند. اطلاعات ژنتیکی نیز در هسته وجود دارد. در شکل ۲.۱ نمایی از یک هسته سلول نشان داده شده است [۲۳].

۳.۱ دی.ان.آ

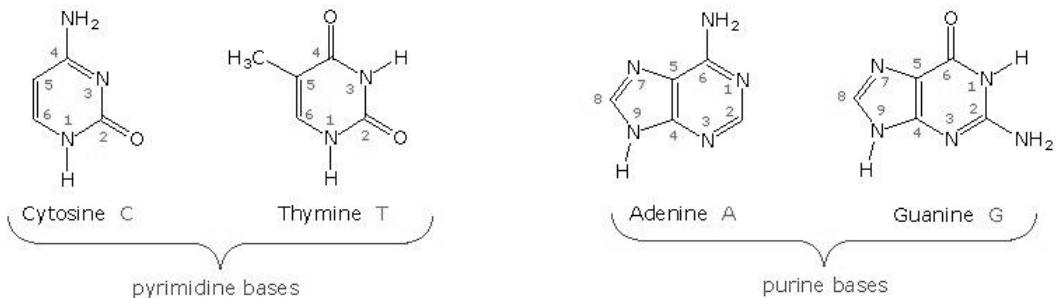
حدوداً ۹۲ عنصر در طبیعت وجود دارد که تنها چهار عنصر کربن، هیدروژن، نیتروژن و اکسیژن ۹۶.۵٪ وزن موجودات زنده را تشکیل می‌دهند. بین این چهار عنصر نقش کربن خاص‌تر است؛ زیرا کربن، اتم کوچکی است که چهار الکترون و چهار جای خالی دارد و می‌تواند پیوند محکم کوالانسی ایجاد کرده و به ساخت درشت‌مولکول‌ها بیانجامد. زنجیره یا حلقه‌های کربنی تولید شده را مولکول‌های آلی می‌نامند. از ترکیبات مختلف مولکول‌های آلی

^۱Nuclear Envelope

^۲Nuclear Pores

^۳Nucleus

^۴Nucleoplasm



شکل ۳.۱: فرمول شیمیایی بازهای آلی [۱]

در سلول‌ها چهار خانواده از درشت‌مولکول‌ها شکل می‌گیرد. پلی‌ساقاریدها^۱ که از قند تشکیل می‌شوند، چربی‌ها^۲ از اسیدهای چرب، پروتئین‌ها از اسید آمینه‌ها^۳، و اسیدهای نوکلئیک^۴ از ترکیب نوکلئوتیدها^۵ به جود می‌آیند. اسید آمینه در شیمی به هر ملکولی که شامل گروه‌های کاربردی آمینه و کربوکسیلیک اسید است گفته می‌شود و نوکلئوتیدها از ترکیب، حلقه‌های حاوی نیتروژن متصل به حلقه قند پنج کربنی که یک یا چند گروه فسفات دارد، ساخته شده‌اند. دی‌ان‌آ^۶ نوعی اسید نوکلئیک است. حلقه‌های حاوی نیتروژن را باز آلی می‌نامند؛ چهار نوع باز در ساختار دی‌ان‌آ به کار رفته که سیتوزین^۷، تیمین^۸، گوانین^۹ و آدنین^{۱۰} نامگذاری شده‌اند و به اختصار به ترتیب C، T، G و A خوانده می‌شوند. شباهت خانوادگی زیادی بین بازهای نوکلئوتیدی وجود دارد. به دسته سیتوزین و تیمین، پیریمیدین^{۱۱} و به دسته گوانین و آدنین، هم پورین^{۱۲} می‌گویند. شکل ۳.۱ فرمول بازهای آلی به همراه ترکیب شیمیایی آن‌ها را نشان می‌دهد [۱].

بازهای سیتوزین و تیمین تمایل به ایجاد پیوند هیدروژنی بین هم دارند و همین‌طور بازهای گوانین و آدنین نیز تمایل دارند پیوند هیدروژنی بین خود برقرار کنند. این تمایل باعث شده است تا دی‌ان‌آ به صورت ساختار مارپیچ دورشتهای^{۱۳} منظمی از ترکیب درشت‌مولکول‌های نوکلئوتید با پیوند هیدروژنی بین بازهای هم خانواده در

^۱polysaccharide

^۲Lipid

^۳Amino acid

^۴Nucleic acid

^۵Nucleotide

^۶Deoxyribonucleic Acid(DNA)

^۷Cytosine

^۸Thymine

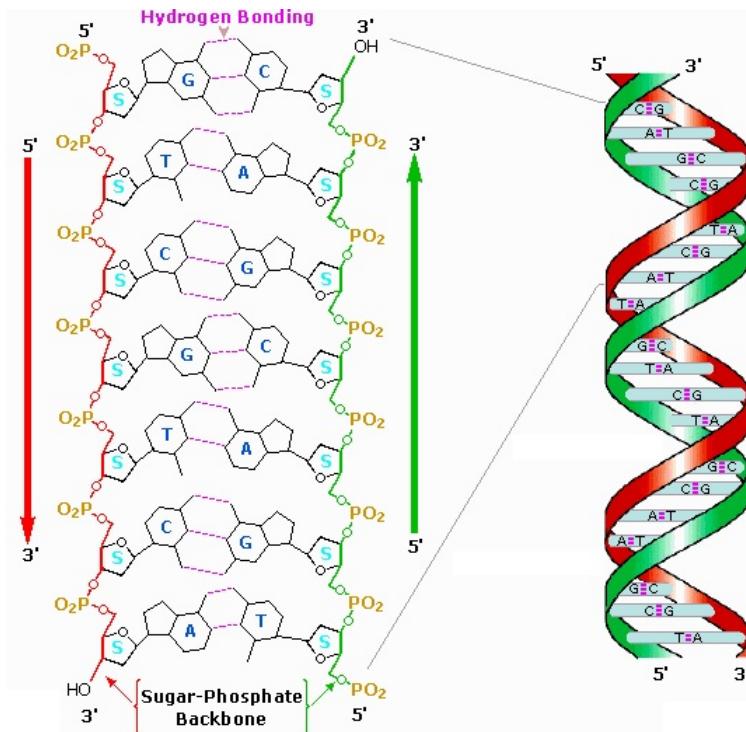
^۹Guanine

^{۱۰}Adenine

^{۱۱}Pyrimidine

^{۱۲}Purine

^{۱۳}Double helix

شکل ۱.۴: رشته دی.ان.آ.^[۱۴]

دو رشته مقابله هم به وجود آید. در شرکت مولکول های دی.ان.آ از دو رشته تشکیل شده اند که هر دو رشته^۱ اطلاعات ژنتیکی یکسانی را در بر دارند [۲۴، ۲]. اسکلت دی.ان.آ از قند پنج کربنی تشکیل شده است که با گروه فسفات کربن سوم خود به گروه فسفات کربن پنجم قد پنج کربنی بعدی می چسبد. پس به جز قند اول و آخر، بقیه قندها از هر دو گروه فسفات در جایگاه سوم و پنجم با قند مجاور پیوند دارند. به قندی که جایگاه پنجم آن بدون پیوند است پنج پریم^۲، و به قندی که جایگاه سوم آن بدون پیوند است سه پریم^۳ می گویند. واضح است که اگر در ابتدای رشته ای جایگاه آزاد، قند شماره پنج باشد در انتهای رشته جایگاه آزاد، قند شماره سه خواهد بود، و جهت آن را پنج پریم به سه پریم می خوانند و بر عکس. با توجه به این خصوصیت، رشته های دی.ان.آ به صورت عکس مقابل هم قرار دارند. شکل ۱.۴ تصویر دی.ان.آ و نحوه قرار گیری بازه های آلی در دو رشته آن را نشان می دهد [۱۴].

تعداد زیادی دی.ان.آ برای نگهداری همه اطلاعات لازم برای تکامل یک باکتری تک سلولی نیاز است و دی.ان.آ بیشتری نیاز است تا دستورات لازم برای تکامل موجودات پرسلولی مانند انسان را نگهداری کند. این اطلاعات بر روی دی.ان.آ نگهداری می شود. هر سلول انسانی حاوی حدود دو متر دی.ان.آ است، در حالی که قطر هسته فقط پنج تا هشت میکرومتر است. پس برای اینکه این حجم زیاد دی.ان.آ داخل هسته قرار گیرد باید بسیار فشرده

^۱Strand^۲5'^۳3'

شود^[۱]. برای این کار رشته‌های دی.ان.آ دور درشت‌مولکول‌های پپتیدی به نام هیستون‌ها^۱ پیچیده می‌شوند. به مجموعه دی.ان.آ و هیستون‌ها کروماتین^۲ می‌گویند. دی.ان.آ در سلول به صورت رشته‌های بلند سازماندهی شده است که کروموزوم^۳ نام دارند. در طول تقسیم سلولی برای تکثیر شدن، این کروموزوم‌ها در راستای کپی شدن دی.ان.آ رونویسی می‌شوند تا هر سلول یک سری کامل از کروموزوم‌ها را داشته باشد. پس همه سلول‌های یک جاندار با وجود تفاوت در ساختار و کارکرد دی.ان.آ یکسانی دارند^[۲].

دی.ان.آ انسان در هسته، درون ۲۴ کروموزوم توزیع شده است که حدود $\frac{2}{3}$ میلیارد نوکلئوتید طول کل آن می‌باشد. به استثناء سلول‌های زایا (اسپرم و تخمک) و سلول‌های بسیار تخصص یافته مانند گلبول‌های قرمز، هر سلول انسانی شامل دو کپی از هر کروموزوم می‌باشد که یکی را از مادر و دیگری را از پدر به ارث برده است که به آنها کروموزوم‌های همولوگ می‌گویند. به هر کدام از کپی‌های کروموزوم یک کروماتید^۴ می‌گویند. کروموزوم‌های جنسی در مرد تنها جفت کروموزوم‌هایی هستند که غیر همولوگ می‌باشند، چون کروموزوم Y را از پدر و کروموزوم X را از مادر به ارث می‌برد. اما علاوه بر هسته سلول، میتوکندری (کروموزوم M) و در گیاهان کلروپلاست هم مقدار کمی دی.ان.آ دارند^[۳].

۴.۱ چرخه سلولی

چرخه سلولی^۵ یک مجموعه از فعالیت‌های متوالی است که باعث تقسیم شدن سلول به دو سلول مجزا و یکسان می‌شود به طوری که سلول جدید ایجاد شده قادر است همه فعالیت‌های سلول اول را انجام دهد و مجدداً سلول جدیدی تولید کند. چرخه سلولی یک سلول یوکاریوت، شامل دو مرحله اینترفاز^۶ و تقسیم است. مرحله‌ای از چرخه سلولی را که در آن سلول تقسیم نمی‌شود اینترفاز می‌نامند. در مرحله اینترفاز، سلول با سنتز دی.ان.آ و انجام تدارکات لازم جهت تفکیک اطلاعات ژنتیکی، همانندسازی اندامک‌ها و متراکم شدن کروموزوم‌ها را انجام می‌دهد. در مرحله تقسیم یا میتوز^۷، کروموزوم‌های خواهری از هم جدا می‌شوند. این مرحله، مرحله اساسی در جدا شدن دو سلول دختر است. و در نهایت سلول به دو سلول دختر تقسیم می‌شود. هر سلول دختر تقریباً نیمی از محتويات سیتوپلاسمی، از جمله یکی از دو هسته همانندسازی شده را به عنوان میراث دریافت می‌کند. با پایان این مرحله یک چرخه کامل می‌شود^[۴].

^۱Histones

^۲Chromatin

^۳Chromosome

^۴Chromatid

^۵Cell cycle

^۶Interphase

^۷Mitosis

علاوه بر چرخه سلولی که در همه سلول‌ها رخ می‌دهد در اندام‌های جنسی نوع دیگری از تقسیم هم صورت می‌گیرد به این نوع تقسیم شدن، میوز^۱ می‌گویند. تقسیم میوز شامل دو بخش میوز اول و میوز دوم است. در اثر تقسیم میوز، گامت‌ها به وجود می‌آیند. این تقسیم عموماً قبل از تشکیل گامت‌ها یا همزمان با تولید آن‌ها صورت می‌گیرد. این فرایند سبب می‌شود که در موقع تشکیل تخم، تعداد کروموزوم‌ها مضاعف نشود. تقسیم میوز در اندام تولید مثلی نر و ماده صورت می‌گیرد. این سلول‌ها دو تقسیم متواالی را طی می‌کنند، اما کروموزوم‌ها فقط یک بار مضاعف می‌شوند. از این تقسیم چهار سلول حاصل می‌آید که تعداد کروموزوم‌های هر یک نصف تعداد اولیه است. در مرحله اول کروموزوم‌های همساخت به ترتیب ویژه‌ای جفت می‌شوند. کروموزوم‌ها کوتاهتر و ضخیخت‌تر می‌شوند. حال هر کدام یک تتراد^۲ هستند که از دو کروموزوم همساخت یا چهار کروماتید تشکیل شده‌اند. هر کروماتید از یک تتراد، به دور کروماتید خواهر خود می‌پیچد و کوتاهتر و ضخیم‌تر می‌شود. دو کروماتید خواهر از هر کروموزوم همساخت، قطعاتی را بین خود مبادله می‌کنند. تبادل قطعات بین دو کروماتید از دو کروموزوم همساخت را تقاطع کروموزومی^۳ گویند. کروموزوم‌ها، جدا شدن از یکدیگر را آغاز می‌کنند، اما چون در بعضی نقاط تبادل صورت گرفته است، لذا در این نقاط متصل به یکدیگر باقی می‌مانند. این مرحله از نظر ژنتیکی دارای اهمیت فراوانی است، زیرا تبادل بین کروماتیدهای ناخواهری در این مرحله صورت می‌گیرد که سبب تشکیل کروماتیدهای نوترکیب می‌شود. در مرحله دوم میوز، کروماتیدهای ناخواهری از هم جدا می‌گردند. این کروماتیدها، با کروموزوم پدری و مادری خود تفاوت دارند. تقسیم میوزی کامل می‌شود و چهار سلول به وجود می‌آید. در بسیاری از جانداران ماده، سیتوپلاسم سلول‌ها در میوز به طور نامساوی تقسیم می‌شود و فقط یک سلول به جای چهار سلول حاصل می‌آید که سیتوپلاسم فراوان دارد و به تخمک تبدیل می‌شود. سه سلول کوچک باقیمانده معمولاً می‌میرند. در بعضی از جانداران نر، چهار سلول حاصل به اسپرم تبدیل می‌شوند. شکل ۵.۱ مراحل تقسیم میوز و میتوز را نمایش می‌دهد^[۵].

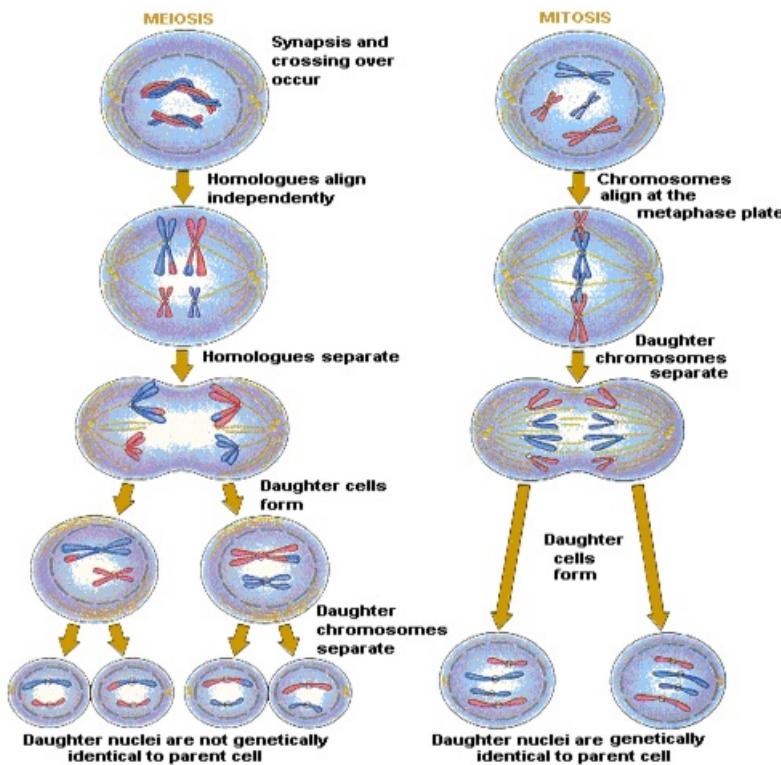
به طور خلاصه تفاوت‌های این دو تقسیم عبارتند از:

- تعداد کروموزوم‌ها در میوز نصف می‌شود ولی در میتوز ثابت می‌ماند.
- میوز فقط در اندام‌های جنسی صورت می‌گیرد اما میتوز در همه اندام‌های در حال رشد صورت می‌گیرد.
- حاصل هر بار میتوز دو سلول دختری است اما حاصل هر بار میوز چهار سلول دختری است.
- در میوز امکان نوترکیبی وجود دارد اما در میتوز خیر^[۶].

^۱Meiosis

^۲Tetrad

^۳Crossing-over



شکل ۱: مراحل تقسیم سلولی میوز و میتوز [۵]

۵.۱ آر.ان.آ

آر.ان.آ^۱ از لحاظ ساختاری همانند دی.ان.آ نوعی اسیدنوكلئیک است و تنها فرقش در عدم داشتن یکی از بازهای آلی است. آر.ان.آ به جای باز تیمین باز دیگری به نام یوراسیل^۲ دارد و به جای ساختار مارپیچ دورشته‌ای، ساختار تک رشته‌ای دارد که بازهای یوراسیل و آدنین آن با هم و بازهای سیمین و گوانین هم با یکدیگر در روی همان رشته پیوند هیدروژنی برقرار می‌کنند. آر.ان.آ انواع مختلفی دارد که هر یک وظایف مختلفی در سلول بر عهده دارند. برای مثال، آر.ان.آ پیامرسان^۳ وظیفه اصلی پروتئین سازی را به عهده دارد و حاوی کدهای ژنتیکی برای ساخت پروتئین می‌باشد. آر.ان.آ ریبوزومی^۴ اصلی‌ترین جز تشکیل دهنده ریبوزوم‌ها است و آر.ان.آ ناقل^۵ وظیفه انتقال اسیدآمینه‌ها به داخل جایگاه خاص ریبوزیم را دارد؛ در واقع عمل اصلی ترجمه در پروتئین‌سازی به عهده آر.ان.آ ناقل است [۲].

^۱Ribonucleic Acid (RNA)

^۲Uracil

^۳Messenger RNA(mRNA)

^۴Ribosomal RNA(rRNA)

^۵Transfer RNA(tRNA)

۶.۱ پروتئین

هنگام مشاهده یک سلول در زیر میکروسکوپ یا بررسی فعالیت‌های الکتریکی یا بیوشیمی آن، در حقیقت پروتئین‌ها^۱ را می‌بینیم و بررسی می‌کنیم. پروتئین‌ها واحدهای ساختمانی تشکیل دهنده سلول هستند و بیشترین وزن سلول (بدون در نظر گرفتن آب) را تشکیل می‌دهند. پروتئین‌ها علاوه بر نقشی که در شکل و ساختار سلولی دارند، بسیاری از عملکردهای سلول را نیز انجام می‌دهند. برای مثال، آنزیم‌ها با ایجاد سطح مولکولی پیچیده که به وسیله برآمدگی‌ها و فرورفتگی‌ها ایجاد می‌شوند، سطح مکملی برای مولکول‌های خاص ایجاد کرده و به این ترتیب واکنش شیمیایی درون سلولی را انجام می‌دهند. پروتئین‌های درون غشای پلاسمایی، کانال‌ها و پمپ‌هایی را ایجاد می‌کنند که انتقال مواد غذایی و سایر مولکول‌های کوچک به داخل و خارج سلول را کنترل می‌کند. پروتئین‌های دیگر پیام‌ها را از یک سلول به سلول دیگر انتقال می‌دهند و یا به عنوان جمع‌کننده پیام عمل کرده و اطلاعات را از غشای پلاسمایی به هسته سلول‌های خاص حمل می‌نمایند. برخی دیگر به عنوان ماشین‌های مولکولی کوچک با بخش‌های متحرک به کار می‌روند، مانند کینزین که به حرکت اندامک‌ها در سرتاسر سیتوپلاسم کمک می‌کند. برخی دیگر نظیر هلیکاز^۲، مولکول دی‌ان‌آ دو رشته‌ای را باز می‌نماید. همچنین پروتئین‌های تخصص یافته می‌توانند به عنوان آنتی‌بادی، سم، هورمون، مولکول‌های ضدیخ، رشته‌های الاستیک یا ژنراتورهای نورافشان عمل کنند. برای دانستن این که ژن‌ها چگونه کار می‌کنند، ماهیچه‌ها چگونه منقبض می‌شوند، چگونه اعصاب پیام الکتریکی را منتقل می‌کنند، جنین چگونه تکوین می‌باید و یا چگونه بدن کار می‌کند باید ابتدا پروتئین‌ها را شناخت [۱]. پروتئین چهار ساختار دارد. ساختار اول پروتئین همان توالی اسید‌آمینه آن است. ساختار دوم آن، یک ساختار سه بعدی است که از ایجاد پیوند در اسیدهای آمینه نزدیک به هم در زنجیره خطی پروتئین به وجود می‌آید. ساختار سه بعدی هر پروتئین منحصر به فرد است، و اغلب دو الگوی تاخوردگی منظم در این ساختار وجود دارد. یکی از الگوها مارپیچ آلفا^۳ نام دارد و دیگری به صفحه بتا^۴ معروف است. این الگوها شکل فضایی منظم و تکرارپذیری به خود می‌گیرند که به آن‌ها موتیف‌های ساختاری^۵ می‌گویند. یک مارپیچ آلفا زمانی ایجاد می‌شود که زنجیره درشت‌مولکولی پروتئینی به دور خود بچرخد و استوانه‌ای محکم ایجاد نماید. صفحات بتا زمانی تشکیل می‌شوند که بین قطعاتی از زنجیره درشت‌مولکولی پروتئینی پیوندهای هیدروژنی ایجاد شده و آن‌ها را در کنار هم قرار می‌دهد. اگر به ساختار دوم، قوس یا چین دیگری اضافه شود، ساختار سوم می‌گویند. ساختار سوم پروتئین، ساختاری سه بعدی و کروی شکل است که در آن زنجیره مارپیچ بصورت یک الگوی هندسی معینی تا

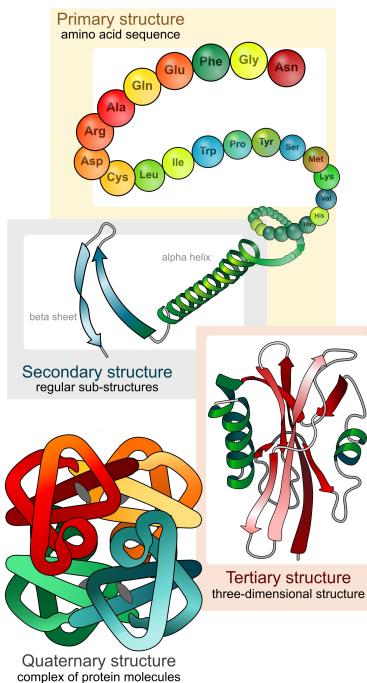
^۱Protein

^۲Helicase

^۳ α helix

^۴ β sheet

^۵Structural motif



شکل ۶.۱: تصویر سه بعدی ساختارهای پروتئین [۱]

شده و بهم تابیده شده است. بسیاری از آنزیم‌ها پروتئین‌های کروی دارند. ساختارهای نوع سوم با انواع متفاوتی از پیوندهای شیمیایی بهم متصل می‌شوند. و در نهایت، اگر یک مولکول پروتئین خاص به صورت کمپلکسی از بیش از یک زنجیره درشت‌مولکولی پروتئینی ساخته شود به آن ساختار کامل، ساختار چهارم می‌گویند. آنچه در مطالعه ساختار چهارم اهمیت دارد واحدهایی به نام حوزه‌های پروتئینی^۱ است. شکل ۶.۱ تصویر سه بعدی ساختارهای پروتئین را نمایش می‌دهد [۱].

حوزه یا دومین پروتئینی، هر قطعه از زنجیره درشت‌مولکولی پروتئینی است که بتواند به صورت غیروابسته، به یک ساختار فشرده و پایدار چین بخورد [۱]. هر دومین می‌تواند به تنها یی و یا در ارتباط با سایر دومین‌های پروتئین فعالیت مشترکی را انجام دهد [۲۵]. یک دومین معمولاً بین ۱۰۰ تا ۲۵۰ اسید‌آمینه دارد. دومین‌های مختلف یک پروتئین اغلب با عملکردهای مختلفی در ارتباط هستند. مولکول‌های پروتئین‌های کوچک، همانند پروتئین عضلانی حامل اکسیژن میوگلوبین، معمولاً تنها دارای یک دومین هستند. پروتئین‌های بزرگتر می‌توانند دارای چندین دومین باشند، که معمولاً توسط قطعه‌ای از زنجیره درشت‌مولکولی پروتئینی بدون ساختار، به یکدیگر متصل می‌شوند. با توجه به اینکه ۲۰ نوع اسید‌آمینه در ساخت پروتئین‌ها دخالت دارند، برای هر پروتئین به طول n اسید‌آمینه، 20^n زنجیره مختلف می‌توان ساخت. با این حال، تنها نسبت ناچیزی از این تعداد غیرقابل تصور، می‌توانند به حالت پایدار سه بعدی معین چین بخورند. اکثر مولکول‌های پروتئینی می‌توانند دارای حالت‌های نهایی

^۱Protein domains

پایدار بسیار متفاوتی باشند، که هر کدام از حالت‌ها دارای ویژگی‌های شیمیابی متفاوتی هستند. اما همه این حالت‌ها از لحاظ بیولوژیکی برای موجود زنده مفید نیستند و در انتخاب طبیعی از بین می‌روند^[۱]. زمانی که پروتئین ساختار پایدار نهایی خود را با ویژگی‌های مفید تکامل داد، ساختار آن می‌تواند طی زمان تغییر کند تا پروتئین را قادر به انجام عملکردهای جدید نماید. این عمل طی تکامل انجام شده است، زیرا بسیاری از پروتئین‌های کنونی را می‌توان به خانواده‌های پروتئینی گروه‌بندی کرد که در آن هر عضو خانواده دارای توالی اسید‌آمینه‌ای و حالت پایدار نهایی شبیه به سایر اعضای خانواده را دارد^[۱]. دومین‌ها هم مانند پروتئین‌ها طی تکامل دچار تغییر شده‌اند و همین‌ها باعث شده است تا دومین‌ها هم قابل دسته‌بندی به خانواده‌های مختلفی باشند که از لحاظ ساختار و کارکرد شباهت‌های زیادی به یکدیگر دارند. می‌توان گفت پروتئین‌های با دومین‌های زیاد از تکرار و ترکیب تک دومین‌ها ساخته شده‌اند^[۲۵].

۷.۱ تبدیل دی.ان.آ به پروتئین

به مجموع کل دی.ان.آ که حاوی کل اطلاعات وراثتی است ژنوم^۱ می‌گویند. گفته شد که سلول‌ها با هم از لحاظ مختلف تفاوت دارند، همین‌طور گفته شد که سلول‌ها اطلاعات ژنتیکی یکسانی را هم حمل می‌کنند، پس چه چیزی باعث به وجود آمدن تفاوت بین عملکرد سلول‌ها می‌شود؟ آنچه که باعث ایجاد تفاوت می‌شود به پروتئین‌های متفاوتی که سلول‌های مختلف دارند مربوط می‌شود. یک ژن^۲ بخشی از دی.ان.آ است که دستورات لازم برای تولید یک پروتئین مشخص یا گروهی از پروتئین‌های مشخص را شامل می‌شود. اگرچه این تعریف برای اکثر ژن‌ها صادق است، اما برخی ژن‌ها به جای پروتئین، یک مولکول آر.ان.آ به عنوان محصول نهایی تولید می‌کنند که نیازی به ترجمه شدن به پروتئین را ندارد و مشابه پروتئین‌ها اعمال ساختاری و عملکردی مختلفی را در سلول انجام می‌دهند. ژنوم انسان حاوی حدوداً ۲۵۰۰۰ ژن است. کروموزوم‌ها در بیشتر یوکاریوت‌ها از جمله انسان، علاوه بر ژن‌ها دارای دی.ان.آ اضافی^۳ نیز هستند که اطلاعات مهمی ندارند^[۱]. اما ژن‌ها بر روی کروموزوم‌ها عیناً مثل هم نیستند و می‌توانند چینش‌های مختلفی داشته باشند که به این چینش‌های مختلف دگرچینی یا آلل^۴ می‌گویند. این دگرچینی‌ها در طول نسل‌های مختلف به وجود می‌آیند و اگر باعث سازگاری کمتر جاندار در زیستگاهش نشوند و یا بقاء آن را به خطر نیندازد، ماندگار شده و به نسل‌های بعد منتقل می‌شوند. رنگ‌های متفاوت رنگینه چشم در افراد متفاوت نمونه‌ای از دگرچینی است^[۲۶].

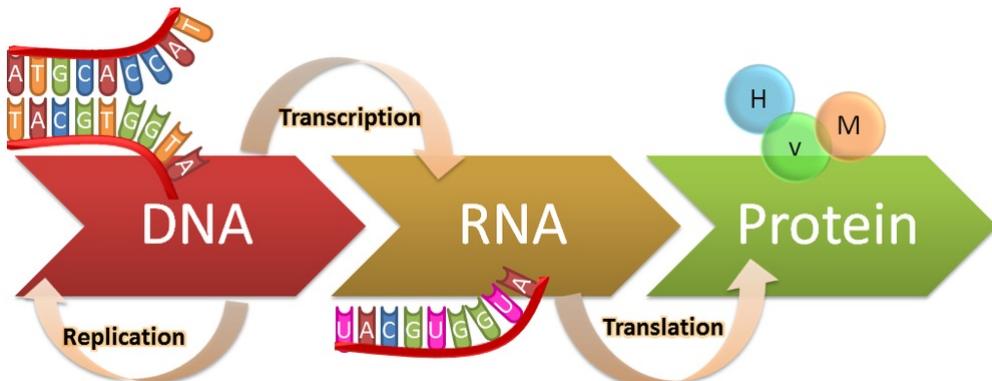
همان‌طور که گفته شد، پروتئین، درشت‌مولکولی است که از ترکیب اسیدهای آمینه به وجود آمده است.

^۱Genome

^۲Gene

^۳Junk DNA

^۴Allele

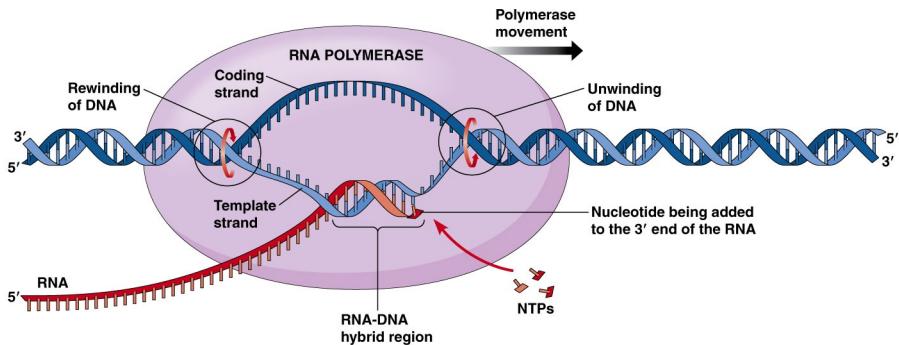
شکل ۷.۱: شکل مراحل تولید پروتئین از روی دی.ان.آ.^[۱]

پروتئین‌ها در حقیقت با استفاده از دی.ان.آ در یک فرآیند دو مرحله‌ای به وجود می‌آیند. در اولین مرحله که رونویسی نام دارد، آر.ان.آ ساخته می‌شود و با استفاده از یکی از رشته‌های دی.ان.آ، به عنوان الگو، و رعایت روابط مکملی بین نوکلئوتیدها، ساخته می‌شود. انتقال اطلاعات از دی.ان.آ به آر.ان.آ در هسته سلول و انتقال از آر.ان.آ به پروتئین در سیتوپلاسم صورت می‌گیرد. این دو مرحله اصلی عبارتند از رونویسی^۱، که اطلاعات ژنتیکی را از دی.ان.آ به مولکول آر.ان.آ انتقال می‌دهد و ترجمه^۲، که اطلاعات موجود در آر.ان.آ را به صورت پروتئین بیان می‌کند. شکل ۷.۱ مراحل تولید پروتئین از روی دی.ان.آ را نمایش می‌دهد^[۱].

رونویسی در هسته سلول رخ می‌دهد. در این فرآیند، یک مولکول آر.ان.آ از روی یک رشته دی.ان.آ الگو، رونویسی می‌شود؛ مشابه اتفاقی است برای یک رشته دی.ان.آ در هنگام همانندسازی رخ می‌دهد. در همانندسازی دو رشته دی.ان.آ باید در محلی که این فرآیند شروع می‌شود از یکدیگر جدا شوند. در رونویسی فقط یکی از رشته‌های دی.ان.آ به عنوان الگو برای تشکیل مولکول جدید به کار می‌رود. نوکلئوتیدهایی که مولکول آر.ان.آ جدید را می‌سازند، هر کدام توسط پیوند هیدروژنی بین بازهای مکمل، به نوکلئوتیدهای دی.ان.آ الگو وصل می‌شوند. نوکلئوتیدهای آر.ان.آ توسط آنزیم رونویسی، که آر.ان.آ پلی‌مراز^۳ نام دارد به هم متصل می‌شوند^[۱].

آر.ان.آ پلی‌مراز باید محلهای شروع و پایان فرآیند رونویسی ژن را بشناسد؛ شروع رونویسی، یک توالی نوکلئوتیدی است که راهانداز^۴ نامیده می‌شود. اولین مرحله رونویسی این است که آر.ان.آ به راهانداز اتصال یابد و شروع به رونویسی آر.ان.آ کند. در طی مرحله رونویسی، همچنان که سنتز آر.ان.آ ادامه می‌یابد، رشته آر.ان.آ از دی.ان.آ الگوی خود جدا می‌شود تا دو رشته‌ی دی.ان.آ باز شده، در ناحیه‌ای که رونویسی انجام شده است، به طرف

^۱Transcription^۲Translation^۳RNA polymerase^۴Promoter



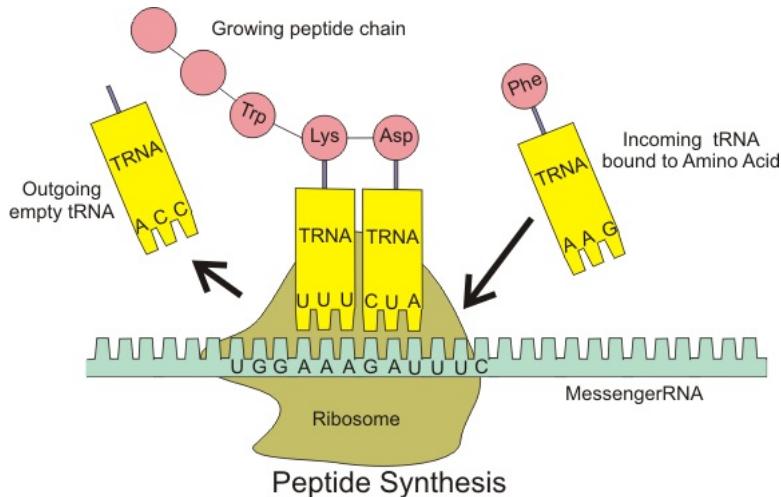
شکل ۱.۸: مراحل رونویسی آر.ان.آ از روی دی.ان.آ [۱]

یکدیگر برگردند. سرانجام در سومین مرحله یا مرحله پایان، آر.ان.آ پلیمراز به یک توالی ویژه از بازها در دی.ان.آ الگو می‌رسد که به آن توالی پایان‌دهنده^۱ می‌گویند. این توالی نشان‌دهنده انتهای ژن است. در این نقطه، آر.ان.آ پلیمراز از مولکول ژن جدا می‌شود. به این نوع آر.ان.آ ایجاد شده که اطلاعات دی.ان.آ الگو را کد می‌کند، آر.ان.آ پیام‌سان گفته می‌شود؛ زیرا این مولکول اطلاعات ژنتیکی را از داخل هسته سلول به ماشین ترجمه سلول که در سیتوپلاسم قرار دارد، انتقال می‌دهد. طی مرحله رونویسی علاوه بر آر.ان.آ پیام‌سان، دو نوع دیگر از آر.ان.آها (آر.ان.آ ریبوزومی و آر.ان.آ ناقل) نیز تولید می‌شوند که در سنتز پروتئین‌ها نقش دارند. شکل ۱.۸ مراحل رونویسی آر.ان.آ از روی دی.ان.آ را به خوبی نمایش می‌دهد [۱].

ترجمه پیام ژنتیکی حمل شده در آر.ان.آ پیام‌سان به زبان اسید‌آمینه پروتئین، نیاز به یک مترجم خاص دارد. در سلول، این عمل توسط آر.ان.آ ناقل انجام می‌شود. درواقع آر.ان.آ ناقل اطلاعاتی که در آر.ان.آ پیام‌سان به صورت نوکلئوتید وجود دارد را به واحدهای سازنده پروتئین تبدیل می‌کند. یک مولکول آر.ان.آ ناقل از یک رشته منفرد آر.ان.آ که یک زنجیره درشت‌مولکولی نوکلئوتیدی شامل تقریباً ۸۰ نوکلئوتید است ساخته می‌شود. البته یک مولکول آر.ان.آ ناقل به تنها یک نمی‌تواند عمل ترجمه را انجام دهد و برای این‌که یک اسید‌آمینه مناسب به آر.ان.آ ناقل متصل شود، به نوعی آنزیم نیاز دارد. بنابراین، اطلاعات به شکل مولکول‌های آر.ان.آ پیام‌سان، برای تطبیق اطلاعات مولکول آر.ان.آ ناقل و آنزیم‌هایی برای متصل کردن اسیدهای آمینه به آر.ان.آ ناقل آماده‌اند که درشت‌مولکول‌های پپتیدی^۲ مناسب را در ریبوزوم بسازند. ریبوزوم تعدادی آر.ان.آ مخصوص خود به نام آر.ان.آ ریبوزومی و تعدادی هم پروتئین دارد که می‌تواند درشت‌مولکول‌های پپتیدی را تولید کند. به طور معمول، یک مولکول آر.ان.آ پیام‌سان به طور همزمان توسط چندین ریبوزوم ترجمه می‌شود. وقتی که توالی نوکلئوتیدی را انداز از اولین ریبوزوم خارج می‌شود، ریبوزوم اسیدی می‌تواند به آن اتصال یابد، پس ممکن است چندین ریبوزوم به دنبال هم روی یک آر.ان.آ پیام‌سان حرکت کنند. پس از این که درشت مولکول پپتیدی ساخته شد، پیچ و تاب

^۱Terminator

^۲Poly peptide



شکل ۹.۱: ترجمه آر.ان.آ به پروتئین [۱]

می خورد و یک شکل سه بعدی به خود می گیرد. چندین درشتمولکول پپتیدی گرد هم می آیند و یک پروتئین می سازند. در شکل ۹.۱ مراحل و نحوه ترجمه آر.ان.آ به پروتئین نشان داده شده است. [۱]

از آنجایی که دی.ان.آ دور شته ای است، در راه انداز دو نوع آر.ان.آ مختلف می تواند ساخته شود؛ یکی از چپ به راست و دیگری از راست به چپ. اما راه انداز نامتقارن بوده و فقط در یک جهت متصل می شود، بنابراین بعد از قرار گیری آر.ان.آ در راه انداز، این آنزیم رونویسی را در جهت پنج پریم به سه پریم انجام می دهد. اما در کل جهت رونویسی ژن ها متفاوت است. گفته شد دی.ان.آ علاوه بر ژن ها حاوی اطلاعات اضافی نیز است. در بیشتر ژن های یوکاریوتی توالی های کد کننده توسط توالی غیر کد کننده به نام اینtron^۱ گسترش می شود. قطعات کد شونده که در میان اینtron ها پخش شده اند را اگزون^۲ می نامند، که معمولاً کوتاه تر از اینtron ها هستند. طول اینtron ها می تواند از یک تا ده هزار نوکلئوتید باشد. نسبت بین طول اینtron ها و اگزون ها در هر ژن نسبت به ژن دیگر متفاوت است. برای تولید یک آر.ان.آ پیامبر در سلول های یوکاریوتی کل ژن شامل همه اینtron ها و اگزون ها به شکل آر.ان.آ رونویسی می شود و سپس با ادامه کار آر.ان.آ پلی مراز، فرآیند پردازش آر.ان.آ^۳ شروع شده و طی آن توالی اینtron ها حذف شده و اگزون ها به طور متواالی به هم می چسبند [۱]. پروتئین ها متنوع ترین درشتمولکول ها از لحاظ ساختاری در سلول هستند. اگرچه اندازه آن ها از ۳۰ تا ۱۰،۰۰۰ اسید آمینه است، اما اکثر آن ها بین ۵۰ تا ۲۰۰۰ اسید آمینه طول دارند. در طبیعت حدود ۱۰۰ نوع اسید آمینه شناخته شده وجود دارد که فقط ۲۰ عدد از آن ها در ساختار پروتئین ها شرکت دارند [۱]. برای نام گذاری اسید های آمینه از کدهای سه حرفی استفاده می شود، که از سه اسید نوکلئوتید متواالی تشکیل شده است. به هر گروه سه تایی از اسید های نوکلئوتید متواالی

^۱Intron

^۲Exon

^۳RNA splicing

در آر.ان.آ یک کدن^۱ گفته می‌شود. در جدول ۱.۱ اسامی اسیدهای آمینه و توالی سه تایی آن‌ها آمده است. توالی آر.ان.آ می‌تواند در هر یک از سه قالب خواندن^۲ قرائت شود و بستگی دارد که فرآیند رمزگشایی از کجا آغاز شود. اما فقط یکی از سه قالب خواندن می‌تواند پروتئین صحیح را تولید کند[۲۷]. در پروتئین‌ها هم مانند دی.ان.آ برای مشخص شدن سر و ته آن‌ها از اسامی N-terminus، برای مشخص شدن سمتی که اسیدآمینه آغازی آن با COOH- پایان می‌یابد و از C-terminus برای مشخص شدن سمتی که اسیدآمینه پایانی آن با NH2- پایان می‌یابد، استفاده می‌کنند. در حالت کلی، پروتئینی که از آر.ان.آ پیامبر ترجمه می‌شود از سمت N-terminus به سمت C-terminus ادامه می‌یابد[۸]. هر پروتئین معمولاً به یک شکل پایدار نهایی منحصر به فرد چین می‌خورد، اما گاهی اوقات زمانی که پروتئین با مولکول‌های دیگر در سلول میانکنش می‌دهد، این ساختار به طور جزئی دچار تغییر می‌شود. اما اگر پروتئین‌ها به طور نادرست چین بخورند، ممکن است تجمعاتی را ایجاد کنند که به سلول آسیب برساند و یا به کلی آن را از بین ببرد. پروتئین‌های تجمع‌یافته، زمینه تعدادی از اختلالات عصبی مانند آلزایمر^۳، یا بیماری‌هایی مانند نوعی عفونت مغزی هستند که در نهایت مرگ بیمار را درپی دارند. از آنجایی که پروتئین‌ها ساختار پیچیده‌ای دارند و نگهداری کامل اطلاعات مربوط به این ساختار دشوار است، برای پروتئین‌ها یک ساختار سلسله مراتبی در چهار سطح تعریف می‌شود[۱]. در ساختار اول فقط توالی اسیدآمینه پروتئین‌ها مدنظر قرار می‌گیرد. ساختار اول از سه لحاظ اهمیت دارد:

- تغییر اسیدآمینه در هر جایگاه در زنجیره پروتئین موجب تغییر در عملکرد و فعالیت آن می‌شود مانند عملکرد کاتالیزوری یک آنزیم.
- توالی اسیدآمینه، سطوح دیگر ساختار پروتئین را تعیین می‌کند و تحلیل ارتباطات بین توالی‌های اسیدآمینه، ساختار سه‌بعدی پروتئین و قوانین حاکم بر تاخورده‌گی زنجیره درشت مولکولی پروتئینی را آشکار می‌سازد.
- توالی پروتئین بیانگر اطلاعات زیادی درباره تاریخچه تکامل آن است. بنابراین تغییرات مولکولی ناشی از تکامل می‌تواند از توالی اسیدهای آمینه دنبال شود[۲۸].

۸.۱ ایزوفرم‌های پروتئینی

ایزوفرم‌های^۴ مختلف یک پروتئین شکل‌های متفاوت از یک پروتئین واحد هستند. شکل‌های متفاوت یک پروتئین می‌تواند به وسیله ژن‌های کپی شده خیلی نزدیک به هم ساخته شود. این ژن‌ها در اثر کپی شدن یک ژن اجدادی

^۱Codon

^۲Reading frame

^۳Alzheimer's disease

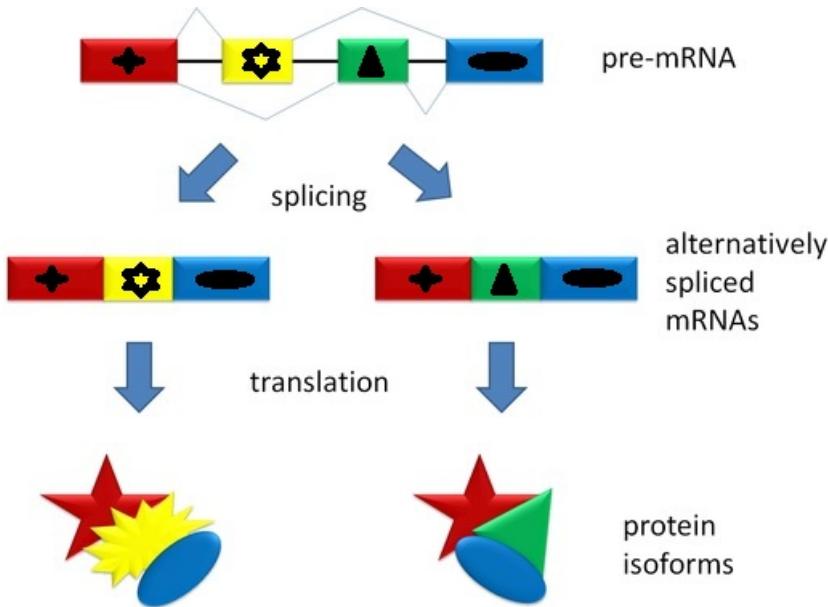
^۴Isoforms

جدول ۱.۱: جدول علائم نامگذاری پروتئین‌ها [۸]

		Second Letter									
		U		C		A		G			
First Letter	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U	Third Letter
		UUC		UCC		UAC		UGU		C	
		UUA	Leu	UCA		UAA	stop	UGA	stop	A	
		UUC		UCG		UAG		UGG	Trp	G	
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	Gln	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U	
		AUC		ACC		AAC		AGC		C	
		AUA		ACA		AAA	Lys	AGA	Arg	A	
		AUG	Met	ACG		AAG		AGG		G	
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	Glu	GGA		A	
		GUG		GCG		GAG		GGG		G	

به دو ژن و ایجاد تغییرات ساختاری در طی تکامل به وجود آمده‌اند. اما همچنان پروتئین‌هایی که تولید می‌کنند شکل‌های متفاوت از یک پروتئین واحد هستند. ممکن است ایزوفرم‌های مختلف یک پروتئین توسط یک ژن واحد ساخته شود و این عمل با حذف یا اضافه شدن اگزون‌هایی همراه است [۲۹]. فضای ساختاری پروتئوم^۱ انسان به دلیل حضور انواع ایزوفرم‌های پروتئین‌های مختلف، بزرگ و متنوع است. این تنوع، در اثر تغییرات پس از ترجمه^۲، انواع باهم متصل شدن‌ها^۳، تغییرات ژنتیکی^۴ و نوترکیبی سوماتیک^۵ به وجود می‌آید. به عنوان مثال، دهها میلیون مولکول IgG متفاوت در بدن انسان وجود دارد که در یک زمان معین موجب پردازش استادانه نوترکیبی سوماتیک و تغییرات هدف‌دار می‌شود. بخش بزرگی از ژن‌های کد کننده پروتئین (حدود ۸۰٪) دارای انواع سرهم کردن اگزون‌ها هستند، که محصولات پروتئین با اندازه‌های مختلف تولید می‌کنند [۱۵]. شکل ۱۰.۱ نحوه ایجاد ایزوفرم‌های مختلف یک پروتئین در یک ژن خاص را نشان می‌دهد.

¹Proteome²posttranslational modifications³splice variants⁴genetic variations⁵somatic recombination



شکل ۱۰.۱: نحوه ایجاد ایزوفرم‌های مختلف یک پروتئین [۱۵]

۹.۱ میتوکندری

میتوکندری در سلول، اندامکی است که وظیفه آن تنفس سلولی و همین‌طور انتقال انرژی است. به علاوه، میتوکندری وظایف دیگر درون سلولی نیز دارد که شامل سیگنال دهنده، تمایز سلولی، مرگ سلولی، کنترل رشد و حفظ سلول می‌باشد. شاخصه‌های متعدد و متفاوت، میتوکندری را منحصر به فرد کرده است. تعداد میتوکندری‌ها در سلول بسته به نوع ارگانیسم‌ها، بافت‌ها و نوع سلول‌ها متفاوت می‌باشد. برای مثال سلول قرمز خون، فاقد میتوکندری می‌باشد پس انرژی‌اش را از مسیر دیگری به دست می‌آورد، در حالی که سلول سازنده کبد بیش از ۲۰۰۰ میتوکندری در ساختار خود دارد. این اندامک شامل چندین توابع مختلف است که هر کدام عملکرد منحصر به فردی دارند [۳۰].

۱.۹.۱ ژنوم میتوکندری

همانند سازی آر.ان.آ، دی.ان.آ سازی و پروتئین سازی در میتوکندری هم صورت می‌گیرد. این فرایند توسط آنزیم‌ها و ملکول‌های خاص خود اندامک صورت می‌گیرد. دی.ان.آ میتوکندری اغلب موجودات، حلقوی است. جایگاه دی.ان.آ در ماده زمینه میتوکندری و بعضی مواقع چسبیده به غشای داخلی میتوکندری است. محصولاتی که توسط دی.ان.آ میتوکندری رمز می‌شوند شامل آر.ان.آهای ریبوزومی میتوکندری، آر.ان.آهای انتقال دهنده و برخی از پروتئین‌های مسیر تنفس می‌باشند. بعضی از پروتئین‌های میتوکندری در هسته رمز می‌شوند و پس از ساخته شدن در سیتوزول وارد اندامک می‌شوند. در انسان میتوکندری کوتاهترین کروموزوم با ۱۶۵۷۱ اسیدنوکلئوتید

طول و ۳۷ ژن است. از این تعداد ۱۳ عدد در پروتئین‌سازی دخالت دارند. ۲۲ عدد در ساختن آر.ان.آهای انتقال دهنده میتوکندریالی و دو عدد نیز در ساختن آر.ان.آهای ریبوزومی دخالت دارند. البته میتوکندری می‌تواند شامل دو تا ۱۰ نسخه از هر دی.ان.آ خود باشد. در بسیاری از گونه‌ها از جمله انسان دی.ان.آ میتوکندری فقط از مادر به ارث می‌رسد. میتوکندری انسان حلقوی است. موضوع قابل توجه دیگر در مورد پروتئین میتوکندری‌ها مربوط به کدن آن است که برخی کدن‌ها معنای متفاوتی با آنچه در پروتئین‌سازی هسته انجام می‌شود، دارند. این ۳۷ ژن میتوکندری با توجه به نوع فعالیتشان به شش مجموعه تقسیم‌بندی می‌شوند. این تقسیم‌بندی در برخی مطالعات دارای اهمیت است، زیرا نگاه ویژه‌ای به ژن‌ها می‌بخشد. جدول ۲.۱ اسامی ژن‌های میتوکندری و مجموعه‌های آن را نشان می‌دهد [۳۰].

جدول ۲.۱: اسامی ژن‌های میتوکندری و اطلاعات مربوط به آن‌ها

Group	Approved Symbols	Gene Type	Length
rRNA	MT-RNR1	non-coding RNA	953
	MT-RNR2		1558
tRNA	MT-TA		68
	MT-TC		65
	MT-TD		67
	MT-TE		68
	MT-TF		70
	MT-TG		67
	MT-TH		68
	MT-TI		68
	MT-TK		69
	MT-TL1		74
	MT-TL2		70
	MT-TM		67
	MT-TN		72
	MT-TP		67
	MT-TQ		71
	MT-TR		64
	MT-TS1		68
	MT-TS2		58
	MT-TT		65
	MT-TV		68
	MT-TW		67
	MT-TY		65
ATP synthase	MT-ATP6	protein-coding gene	681
	MT-ATP8		207
COMPLEX IV	MT-CO1		1542
	MT-CO2		684
	MT-CO3		784
COMPLEX III	MT-CYB		1141
COMPLEX I	MT-ND1		956
	MT-ND2		1042
	MT-ND3		346
	MT-ND4		1378
	MT-ND4L		297
	MT-ND5		1812
	MT-ND6		525

۲.۹.۱ نقش زیستی میتوکندری

میتوکندری در سلول‌های متفاوت و در زمان‌های متفاوت نقش‌های متفاوتی را بازی می‌کند. می‌توان خصیصه‌های متفاوتی را به آن نسبت داد.

- تنفس هوایی سلول‌ها

تمام مواد انرژی‌زا، ضمن تغییرات متابولیکی درون سیتوپلاسمی، به بستره میتوکندری می‌رسند و بعد از اعمال تغییرات بر روی آن‌ها در درون میتوکندری موجب تولید انرژی می‌شوند.

- سنتز اسیدهای چرب

یکی از راه‌های تولید اسید چرب، سیستم میتوکندری‌ای است که عکس اکسیداسیون یا تجزیه آن‌ها می‌باشد.

- دخالت میتوکندری در گوارش چربی‌ها

در هنگام گرسنگی، میتوکندری‌ها به طرف ذرات چربی حرکت کرده و روی ذرات چرب خم شده و آنزیم‌های میتوکندری‌ای شروع به هضم چربی و آزادسازی انرژی می‌کنند.

- ذخیره و تجمع مواد در میتوکندری‌ها

میتوکندری‌ها می‌توانند در اتاق داخلی خود مواد مختلف را انباسته کنند که این مواد عبارت‌اند از: ترکیبات آهن‌دار، چربی‌ها، پروتئین‌ها، کاتیون‌ها و آب.

- محل میتوکندری‌ها در سلول

میتوکندری‌ها اغلب در اطراف هسته دیده می‌شوند. البته میتوکندری‌ها در داخل سلول‌ها جایه‌جا شده و خود را به جایی که نیاز به انرژی بیشتر است می‌رسانند.

- تعداد میتوکندری‌ها در سلول

تشخیص تعداد میتوکندری‌های یک سلول دشوار است. اما اغلب بر حسب نوع سلول و مرحله عمل سلول متفاوت می‌باشد. یک سلول معمولی کبد بیشترین تعداد و در حدود ۱۰۰۰ تا ۱۶۰۰ میتوکندری وجود دارد که در اثر تحلیل رفتن سلول و نیز سرطانی شدن آن کاهش می‌یابد. در مقابل، تعداد میتوکندری در بافت لنفی، خیلی کمتر است. در سلول‌های گیاهی هم کمتر از جانوری می‌باشد چون بسیاری از اعمال میتوکندری‌ها، به وسیله کلروپلاست انجام می‌شود [۳۰].

۱۰.۱ انواع تغییرات توالی دی.ان.آ

در این بخش، انواع تغییرات مربوط به توالی دی.ان.آ شرح داده خواهد شد. شناسایی انواع تغییرات و تمایز آن‌ها از هم، دانستن تفاوت‌ها و شباهت‌های آن‌ها به درک بهتر مسئله پیش رو کمک می‌کند.

۱.۱۰.۱ چند ریختی در یک نوکلئوتید واحد

چند ریختی در یک نوکلئوتید واحد یا اسنیپ^۱، رایج‌ترین نوع تنوع ژنتیکی در میان مردم است؛ یعنی در بیشتر از یک درصد مردم جامعه دیده می‌شوند. هر اسنیپ نشان‌دهنده تفاوت در یک بلوک ساختمانی دی.ان.آ، به نام نوکلئوتید است. اسنیپ‌ها به طور معمول در طول دی.ان.آ یک فرد رخ می‌دهند. به طور متوسط یک اسنیپ در هر ۳۰۰ نوکلئوتید رخ می‌دهد. یعنی ۱۰۰۰۰۰۰ نقطه در ژنوم انسان هست که می‌تواند اسنیپ داشته باشد [۳۱]. اما هر فرد به طور متوسط سه تا پنج میلیون اسنیپ به ارث می‌برد و حدود ۳۰ اسنیپ جدید در ارث به دست می‌آورد. دو نوع اسنیپ وجود دارد. اسنیپ‌های غیرکدکننده^۲، که در پنج منطقه ترجمه نشده نخست UTR^۳ و سه منطقه ترجمه نشده نخست ۳'UTR اینترون‌ها و شبه‌زن‌ها^۴ رخ می‌دهند، یا نقش تنظیمی در پیرایش^۵ مقررات رونویسی (پرومотор)، محل‌های اتصال عامل ترجمه^۶ و مقررات ترجمه (شروع یا فسخ) دارند. دسته دوم، اسنیپ‌های کدکننده^۷ هستند، که شامل اسنیپ‌های متراff^۸ و اسنیپ‌های غیرمتراff^۹ است [۳۲]. اسنیپ‌های متراff تغییری در توالی اسید آمینه ایجاد نمی‌کنند اما باعث تغییر ایستایی آر.ان.آ پیامبر می‌شوند [۳۳].

اسنیپ‌های غیر متراff باعث تغییر توالی اسید آمینه می‌شوند اما ممکن است این تغییر قابل قبول باشد، یا نباشد و باعث بروز بیماری شود [۳۲]. اغلب اسنیپ‌ها بین ژن‌ها رخ می‌دهند. به همین خاطر آن‌ها نشانگرهای خوبی برای کار بر روی ژنوم هستند. به بیان دیگر، هنگامی که اسنیپ‌ها در روی ژن و یا نواحی نزدیک نظارتی ژن رخ می‌دهند ممکن است نقش مستقیم در ایجاد بیماری‌های مؤثر بر عملکرد ژن را داشته باشند. بیشتر اسنیپ‌ها هیچ تأثیری بر سلامت و رشد ندارند اما باعث برخی تفاوت‌های ژنتیکی می‌شوند. با این حال، در مطالعه سلامت انسان بسیار مهم هستند، زیرا، تغییرات اسنیپ‌ها ممکن است باعث نشان دادن حساسیت به برخی داروها شود و یا استعداد ابتلاء به بیماری‌های خاص را افزایش دهد. از اسنیپ‌ها برای پیگیری توارث ژن بیماری در خانواده هم

^۱SNP Single Nucleotide Polymorphism

^۲Noncoding SNPs

^۳Pseudogenes

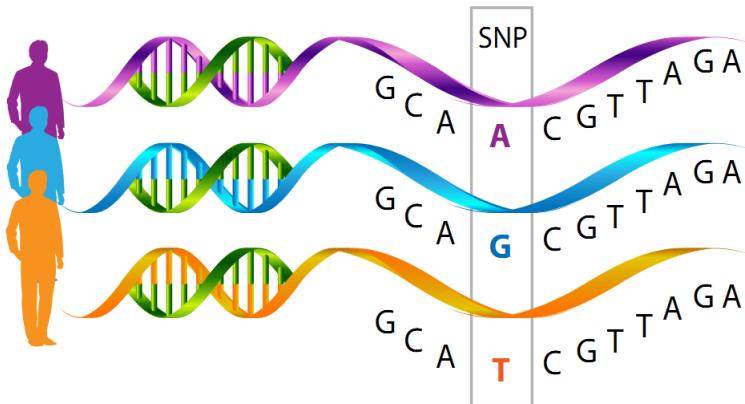
^۴Splicing

^۵TF binding site

^۶Coding SNPs

^۷Synonymous SNPs

^۸Nonsynonymous SNPs



شکل ۱۱.۱: اسنیپ‌ها بر روی ژنوم [۱۶]

استفاده می‌کنند. صمطالعه بر روی اسنیپ‌ها ادامه دارد تا ارتباط آن‌ها با بیماری‌های پیچیده مانند بیماری‌های قلبی، دیابت و سرطان مشخص شود. شکل ۱۱.۱ نمایش اسنیپ بر روی ژنوم را نشان می‌دهد [۱۶].

۲.۱۰.۱ جهش

جهش یک تغییر دائمی در توالی نوکلئوتیدی در ژنوم یک موجود زنده یا دی‌ان‌آی‌های سیتوپلاسمی یا سایر عناصر ژنتیکی است. جهش‌ها نتیجه آسیب دیدن همیشگی دی‌ان‌آی هستند که بر اثر تأثیرات فیزیکی، شیمیایی یا پرتوئی به وجود می‌آیند؛ خطاها یی که باعث حذف، اضافه یا جایگزینی بخشی از آن می‌شوند. این تغییرات در کمتر از یک درصد افراد جامعه رخ می‌دهند و معمولاً عواقب شدیدی دارند. این دو خصیصه، تفاوت اصلی بین اسنیپ و جهش است [۳۴، ۳۵، ۳۶]. جهش‌ها می‌توانند تغییرات ظاهری قابل مشاهده در جاندار ایجاد کنند یا نکنند. جهش‌ها در تکامل، ایجاد سرطان و توسعه سیستم ایمنی بدن نقش دارند. جهش‌ها در ژنوم می‌توانند در محصولی که ژن تولید می‌کند اثر داشته باشند یا نداشته باشند، یا مانع از عملگری ژن به طور جزئی یا کامل شوند [۱۷]. جهش‌های روی ژنوم با نوع تأثیر بر روی توالی پروتئینی به پنج دسته تقسیم می‌شوند. شکل ۱۲.۱ دسته‌بندی انواع جهش‌ها را نشان می‌دهد.

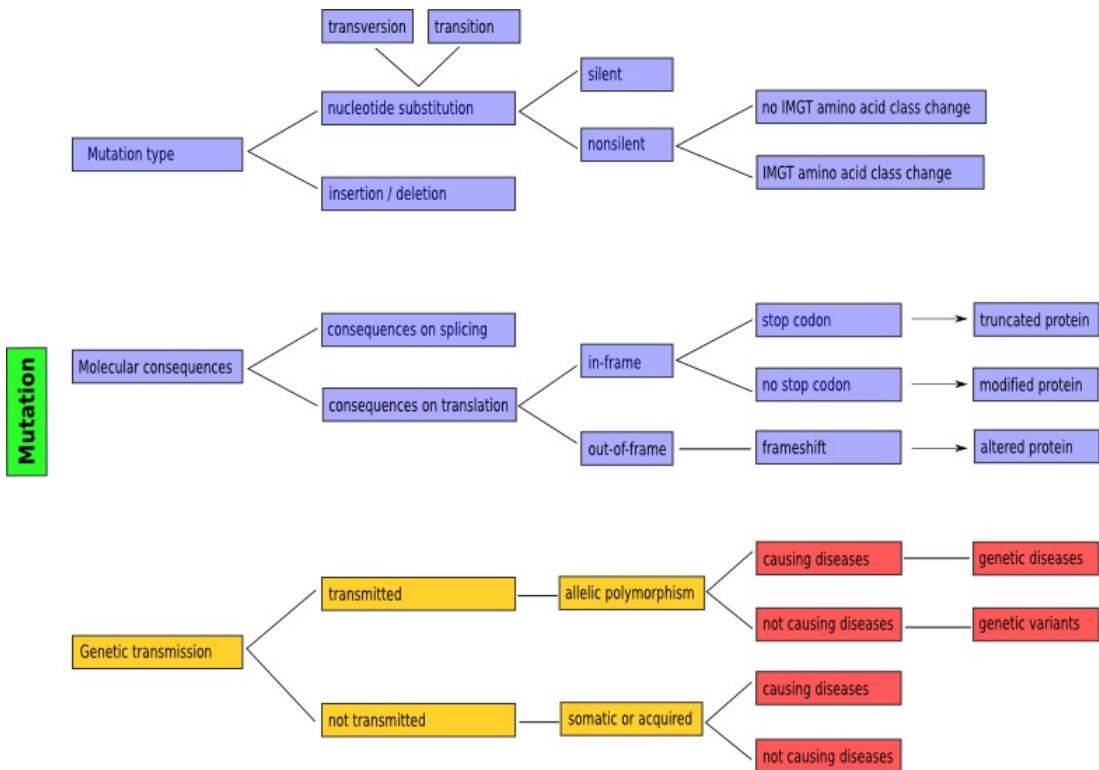
- جهش‌های جابه‌جایی قالب^۱

اضافه شدن به، یا کاسته شدن از موقعیتی در دی‌ان‌آ که مضربی از عدد سه نباشد، باعث جابه‌جایی قالب می‌شود. این افزایش و کاهش می‌تواند باعث ترجمه کاملاً متفاوتی از آر ان‌آ شود. در مقابل، جهش‌های با مضرب سه وجود دارند که به نام جهش‌های داخل قالب شناخته می‌شوند [۳۶].

- جهش‌های بی‌معنی^۲

^۱Frameshift mutation

^۲Nonsense mutations



شکل ۱۲.۱: دسته‌بندی انواع جهش‌ها [۱۷]

جهش‌های نقطه‌ای هستند که باعث جایگزینی یک اسیدآmine در یک موقعیت خاص می‌شوند. این جهش‌ها باعث ایجاد یک کدن توقف زودهنگام یا کدن بی‌معنی در آر.ان.آ پیامبر شده و باعث تولید پروتئین‌های کوتاه‌تر، غیر کامل و اغلب بدون عملگر می‌شوند [۳۷].

• جهش‌های بدمعنی^۱ یا جهش‌های نامتراوف^۲

این نوع جهش‌ها، جهش‌های نقطه‌ای هستند که باعث جایگزینی یک نوکلئوتید با یک نوکلئوتید دیگر شده و باعث می‌شوند پروتئین تولید شده توانایی خود را از دست بدهد و قادر به انجام وظایف خود نباشد. این نوع جهش‌ها باعث به وجود آمدن برخی بیماری‌ها از قبیل اپیدرمولیز بولوسا^۳ و بیماری داسی شکل سلول^۴ می‌شوند [۳۷، ۳۸].

• جهش‌های خنثی^۵

جهش‌های تک نوکلئوتیدی هستند که باعث می‌شوند از لحاظ شیمیایی اسیدآmine مشابه با اسیدآmine

^۱ missense mutations

^۲ nonsynonymous mutations

^۳ Epidermolysis bullosa

^۴ sicklecell disease

^۵ Neutral mutations

اصلی به وجود آید. این جهش باعث می‌شود یا تغییری در پروتئین ایجاد شده دیده نشود یا تغییرات خیلی کم باشد. برای مثال، تغییر اسیدآمینه آرگینین به لیسین که با تغییر در نوکلئوتید وسط از A به G به وجود می‌آید. این دو اسیدآمینه از لحاظ شیمیایی به هم شبیه می‌باشند [۳۹].

• جهش‌های خاموش^۱

جهش‌های خاموش، جهش‌هایی هستند که معمولاً در توالی غیر کدکننده پروتئین رخ می‌دهند، اگر در ناحیه کدکننده پروتئین رخ دهن، تغییری در تولید پروتئین ایجاد نمی‌کنند. گاهاً به این جهش‌ها، جهش‌های متراوف^۲ هم می‌گویند که البته این دو نام کاملاً معادل نیستند زیرا جهش‌های متراوف جهش‌های خاموشی هستند که فقط در نواحی اگزون رخ می‌دهند [۴۰].

• جهش‌های محل شکاف^۳

این جهش‌ها، جهش‌های ژنتیکی شامل اضافه، حذف یا تغییر تعدادی از نوکلئوتیدها در محل‌های مخصوصی هستند [۳۹]. این جهش ممکن است باعث شود بخشی از اینtron، جزئی از اگزون تلقی شود و از آران. آی ایجاد شده، پروتئین غیر طبیعی تولید شود. یا ممکن است بخش یا کل یک اگزون حذف شود و در نهایت پروتئینی بی عمل تولید شود [۴۰]. این جهش‌ها در ایجاد بیماری‌هایی مثل صرع^۴ و یا کمبود پاراتیروئید^۵ نقش دارند [۴۱].

۱۱.۱ تغییرات ژنتیکی و غیر ژنتیکی سلول

دو نوع تغییرات ممکن است در سلول‌های یوکاریوتی رخ دهد. یکی در مرحله تقسیم در چرخه سلولی و دیگر در مرحله تقسیم در اندام‌های جنسی. این تغییرات بسیار مهم هستند، زیرا بر روی بقاء خود موجود و حتی نسل‌های بعدی آن موجود تأثیرگذارند. جهش‌های دودمان زایشی^۶ در اندام‌های جنسی و هنگام تولید سلول‌های جنسی روی می‌دهد. این تغییرات بر نسل آینده موجود اعمال می‌شوند بنابراین می‌توانند گسترش یابند [۱۶]. این نوع تغییرات را می‌توان به صورت عمدی به شکل زیر دسته‌بندی کرد:

• جهش درون ژنی: یک یا چندین نوکلئوتید تغییر کرده، حذف با اضافه شوند.

^۱Silent mutations

^۲synonymous mutation

^۳Splitsite mutations

^۴Dementia

^۵Parathyroid Deficiency

^۶Germ-line mutation

- جهش در دی.ان.آ تنظیمی یک ژن: زمان و جایی که بیان یک ژن می‌تواند تحت تأثیر جهش‌ها در توالی‌های دی.ان.آ که فعالیت آن را تنظیم می‌کند، قرار بگیرد.
- مضاعف شدن ژن‌ها: یک ژن، بخش زیادی از دی.ان.آ و یا حتی کل ژنوم، می‌تواند مضاعف شده و تعدادی از ژن‌های کاملاً یکسان در سلول را به وجود آورد.

- جابه‌جا شدن اگزون‌ها: دو یا تعداد بیشتری ژن می‌توانند شکسته شده و مجددأ به یکدیگر متصل شوند و ژن هیبریدی حاوی قطعات دی.ان.آ که اساساً متعلق به ژن‌های مختلفی هستند تشکیل شود.^[۱]

نوع دیگری از تغییرات، جهش‌های سوماتیک^۱ یا اکتسابی هستند که در مرحله تقسیم چرخه سلولی رخ می‌دهند. این نوع تغییرات در طول نسل باقی نمی‌مانند و فقط تغییراتی را در بدن همان موجود زنده به وجود می‌آورند. این تغییرات می‌توانند گاه‌آ به صورت بیماری یا سرطان بروز نماید. پس می‌توان جهش‌های سوماتیک را به دو دسته مسافر^۲ و راننده^۳ تقسیم کرد. جهش‌ها یا تغییرات مسافر تمایلی به گسترش از بافتی به بافت دیگر را ندارند و فقط یک بافت را درگیر می‌کنند و سلامت بافت را به خطر نمی‌اندازند، اما جهش‌های راننده تمایل دارند به بافت‌های دیگر هم گسترش یابند و سلامت بافت را تهدید کرده و نهایتاً آن را از بین می‌برند. بنابراین می‌توان گفت سرطان در اثر جهش‌های راننده در بدن موجود زنده به وجود می‌آید^[۱۶]. سلول در صدد است تا این تغییرات را به حداقل برساند و آسیب‌های وارد شده به دی.ان.آ را سریعاً و تا جای ممکن ترمیم کند، چون ممکن است این تغییرات بقاء نسل و یا بقاء خود موجود را به خطر اندازد. به همین دلیل یک سیستم پشتیبان به نام ترمیم جفت ناجور^۴ برای تصحیح خطاهای دارد. ماشین هماندسازی به میزان یک در ده ملیون نوکلئوتید اشتباه می‌کند، و سیستم ترمیم جفت ناجور ۹۹٪ آن‌ها را تصحیح کرده و در نهایت میزان اشتباهات به یک در یک میلیارد کاهش می‌یابد؛ یعنی حدوداً سه تغییر نوکلئوتیدی در کل ژنوم. سیستم ترمیم جفت ناجور نقش مهمی در جلوگیری از سرطان دارد^[۱].

۱۲.۱ سلول‌های بنیادی

اهمیت مطالعه سلول‌های بنیادی در رفتاری است که دارند. این سلول‌ها می‌توانند تقسیم شود و سلول‌های جدیدی بسازند. امری که سلول‌های سرطانی آن را به طور نامحدود انجام می‌دهند. در این بخش به تعریف سلول بنیادی و انواع آن پرداخته می‌شود.

^۱Somatic Mutation

^۲Passenger

^۳Driver

^۴Poor DNA repair

۱.۱۲.۱ تعریف سلول‌های بنیادی

سلول‌های بنیادی^۱، دسته‌ای از سلول‌های تغییر نیافته‌ای هستند که می‌توانند طی تقسیم میتوуз سلول‌های بنیادی جدیدی را تولید کنند و یا تغییر یافته و به شکل سلول‌های خاصی درآیند. در جانداران چند سلولی، دو دسته مختلف سلول‌های بنیادی وجود دارد. سلول‌های رویانی^۲ و سلول‌های بنیادی بالغ^۳. در موجودات بالغ، سلول‌های بنیادی به عنوان یک سیستم تعمیر برای تجدید بافت‌های بالغ بدن عمل می‌کنند. در جنین در حال رشد، سلول‌های بنیادی می‌توانند به تمام سلول‌های تخصصی تمایز یابند. در بدن انسان سه منبع برای دستیابی به سلول‌های بنیادی بالغ وجود دارد، مغز استخوان^۴ (به خصوص استخوان ران و لگن خاصره)، بافت‌های چربی^۵ و خون. همچنین به جز این سه مورد، از خون بند ناف بلافارسله بعد از تولد نوزاد هم می‌توان سلول‌های بنیادی استخراج کرد. سلول‌های بنیادی توسط توانایی‌های بالقوه خود در تمایز به انواع دیگر سلول به پنج دسته تقسیم می‌شوند. طبقه بندی کامل سلول‌های بنیادی به شرح زیر است [۱۸].

بس‌توان^۶ که توانایی تمایز به تمام انواع سلول را دارد. به عنوان نمونه: تخم تشکیل شده در لقاح و اولین چند سلول که از تقسیم تخم شکل می‌گیرد. پرتوان^۷ که توانایی تمایز به تقریباً همه انواع سلول‌ها را دارد. برای مثال، سلول‌های بنیادی جنینی و سلول‌های لایه‌های مختلف، که در مرحله آغاز جنینی در اثر تمایز سلول‌های بنیادی تشکیل شده‌اند. چندتوان^۸ که توانایی افتراق به یک خانواده نزدیک و مرتبط از سلول را دارند. برای مثال، سلول‌های بنیادی خونساز بزرگسالان، که می‌توانند به سلول‌های قرمز خون، سلول‌های سفید خون و پلاکت‌ها تبدیل شوند. کم‌توان^۹ که توانایی تمایز به چند سلول را دارد. برای مثال، سلول‌های لنفوئیدی یا سلول‌های بنیادی مغز استخوانی در بزرگسالان. تک‌توان^{۱۰}، که تنها توانایی تولید سلول از نوع خود را دارد. برای مثال، سلول‌های بنیادی بزرگسالان [۱۸].

گاه‌ها سلول‌ها دچار نوعی نقص می‌شوند که نمی‌توانند همنوع خود را تولید کنند. این نقص به مرور زمان باعث ایجاد مشکل در کارکرد سلول‌ها می‌شوند و یا باعث حمله سیستم ایمنی به آن‌ها و نهایتاً باعث از بین رفتن آن سلول‌ها می‌شود. از بیماری‌هایی که به این شکل در بدن انسان شکل می‌گیرند می‌توان به دیابت نوع

^۱Stem Cells

^۲Embryonic Cells

^۳Adult Stem Cell

^۴Bone Marrow

^۵Adipose Tissue

^۶Totipotent

^۷Pluripotent

^۸Multipotent

^۹Oligopotent

^{۱۰} Unipotent

یک، سوختگی، آرتروز و روماتیسم اشاره کرد. دانشمندان و محققان به چند دلیل به سلول‌های بنیادی علاقه‌مند می‌باشند. اگر چه سلول‌های بنیادی هیچ یک عملی انجام نمی‌دهند، اما ظرفیت زیادی برای انجام عملی بعد از تبدیل شدن به سلول‌های خاص را دارند. هر سلول در بدن، برای مثال، از سلول‌های بنیادی اولیه کمی، در مراحل اولیه رشد جنینی مشتق شده است. بنابراین، از سلول‌های بنیادی استخراج شده از جنین می‌توان برای استخراج هر نوع سلول مورد نظری استفاده کرد. این ویژگی باعث قدرت دادن کافی به سلول‌های بنیادی برای بازسازی بافت آسیب دیده تحت شرایط مناسب، می‌شود. سلول‌های بنیادی بالغ اغلب در درمان‌های پزشکی استفاده می‌شود، به عنوان مثال در پیوند مغز استخوان [۱۸].

۲.۱۲.۱ سلول‌های بنیادی سرطان

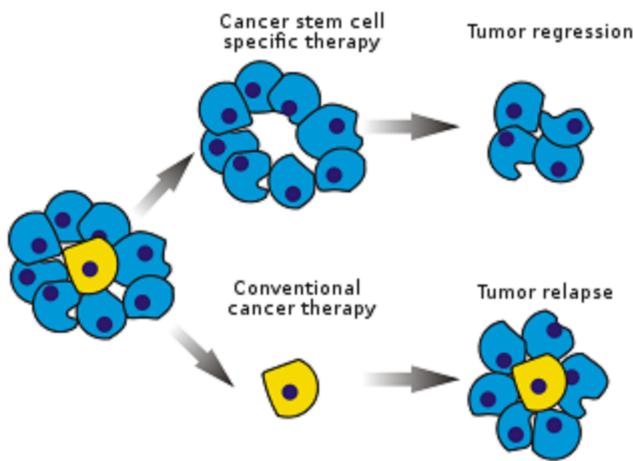
سلول‌های بنیادی سرطانی^۱، نوعی سلول‌های سرطانی هستند که ویژگی‌های سلول‌های بنیادی طبیعی با توانایی تبدیل به تمام انواع سلول‌ها در یک نمونه سرطان خاص را دارند [۱۸]. اما همه سلول‌های سرطانی این خاصیت را ندارد و سلول بنیادی سرطانی نیستند. در واقع علت توسعه سرطان و عود آن به خاطر وجود سلول‌های بنیادی سرطانی است که آن‌ها را قادر به تکثیرشدن می‌کند. بنابراین دستیابی به نحوه کنترل و حذف سلول‌های بنیادی یکی از بزرگ‌ترین امیدها در درمان سرطان است. روش‌هایی مانند شیمی‌درمانی در کشتن سلول‌هایی موفق هستند که قادر به ساختن سلول‌های جدید نیستند. درواقع بزرگ‌ترین ایراد این روش درمانی در عدم توانایی آن در از بین بردن سلول‌های بنیادی سرطانی است و همین دلیل باعث عود سرطان در عده‌ای از بیماران تحت درمان شیمی‌درمانی می‌شود. شکل ۱۳.۱ سلول‌های بعد از درمان‌های مختلف را نشان می‌دهد. مسئله وجود سلول‌های بنیادی سرطانی حدود ۲۰ سال پیش مطرح شد و هنوز هم به عنوان بزرگ‌ترین چالش در درمان سرطان معرفی می‌شود [۱۸].

۱۳.۱ ترمیم دی.ان.آ

بازسازی دی.ان.آ.^۲ به صورت مجموعه‌ای از فرایندهایی است که در آن آسیب واردہ به یک سلول شناسایی شده و اصلاح می‌گردد. در سلول‌های انسان، هم فعالیت‌های متابولیکی طبیعی و هم فاکتورهای زیست محیطی مانند نور و تشعشع فرابنفش می‌تواند باعث آسیب به دی.ان.آ شود. بیشتر این آسیب‌ها مربوط به آسیب ساختاری مولکول دی.ان.آ است و می‌تواند توانایی سلول برای رونویسی ژن در کدگذاری دی.ان.آ را تحت تاثیر قرار دهد. آسیب‌های دیگری که به صورت بالقوه منجر به ایجاد تغییرات مضر در ژنوم سلولی می‌شوند، بر روی زنده ماندن سلول‌های

^۱Cancer Stem Cells

^۲DNA Repair



شکل ۱۳.۱: شکلی از سلول‌های بنیادی سرطانی و درمان‌های خاص [۱۸]

دختر بعد از تقسیم سلولی اثر دارند. فرایند بازسازی دی.ان.آ به صورت ثابتی فعال است تا آسیب وارد شده بر ساختار دی.ان.آ را تصحیح کند. زمانی که فرایند بازسازی عادی دچار مشکل می‌گردد یا زمانی که خزان یاخته‌ای رخ می‌دهد، ممکن است توانایی بازسازی از بین برود. توانایی بازسازی دی.ان.آ در یک سلول برای یکپارچگی ژنوم و عملکرد طبیعی ارگانیسم نیز حیاتی است [۴۲]. میزان بازسازی دی.ان.آ بستگی به عوامل زیادی دارد که شامل نوع سلول، سن سلول و محیط درونی آن است. آسیب‌های بازسازی نشده در ژن‌های حیاتی مانند ژن سرکوبگر غده، می‌تواند مانع توانایی سلول در اجرای عملکردش و افزایش ارزیابی احتمال شکل‌گیری تومور شود. آسیب دی.ان.آ می‌تواند به دو نوع اصلی تقسیم گردد.

- آسیب‌های درون‌زاد یا خود به خودی که در اثر رونویسی دی.ان.آ به وجود می‌آیند. هم‌چنین آسیب‌هایی که در دی.ان.آ ایجاد شده اما تصحیح نشده‌اند و در هر تقسیم سلولی به سلول جدید منتقل می‌شوند.
- آسیب‌های برون‌زا که توسط عوامل خارجی به وجود می‌آیند، عبارتند از: اشعه ماوراء بنفس خورشید، دیگر فرکانس‌های تشعشعی شامل پرتو ایکس و پرتو گام، آبکافت یا تخریب حرارتی، زهراوهای گیاهی خاص، مواد شیمیایی ساخته دست بشر که جهش‌زا هستند به خصوص ترکیبات آروماتیک که به عنوان عوامل افزایش‌زا دی.ان.آ نیز عمل می‌کنند و ویروس‌ها [۴۲].

۱.۱۳.۱ آسیب دی.ان.آ و جهش

تکرار دی.ان.آ آسیب دیده قبل از تقسیم سلولی می‌تواند منجر به درآمیختن غلط بازهای مخالف که آسیب‌زا هستند، شود. سلول‌های دختری که از این بازهای غلط به ارث رسیده‌اند، جهش‌هایی از توالی اصلی دی.ان.آ را انتقال می‌دهند که غیرقابل جبران است (به جز در مورد نادر). تشخیص میان دی.ان.آ آسیب دیده و جهش یافته

مهم است، این دو حالت دو نوع اصلی از خطا در دی.ان.آ. هستند. آسیب‌های دی.ان.آ و جهش به صورت اساسی با هم متفاوت هستند. آسیب‌ها به صورت وضعیت‌های فیزیکی غیر طبیعی در دی.ان.آ. هستند مانند شکستن رشته دوتایی یا مجزا. آسیب‌های دی.ان.آ می‌توانند به وسیله آنزیم‌ها شناسایی گردند و بنابراین آن‌ها در صورتی که اطلاعات اضافی مانند توالی غیر آسیب دیده و سالم در رشته مکمل دی.ان.آ یا در کروموزوم همولوگ برای کپی کردن در دسترس باشد، می‌توانند به درستی بازسازی گردند. اگر دی.ان.آ آسیب دیده از بین نرود، می‌تواند مانع از نسخه‌نویسی ژن می‌شود و بنابراین ترجمه دی.ان.آ به پروتئین مسدود می‌گردد و ممکن است سلول بمیرد [۴۲].
جهش، تغییر در توالی بازی دی.ان.آ است. جهش نمی‌تواند با کمک آنزیم‌ها تشخیص داده شود و بنابراین جهش قابل بازسازی شدن نیست. در سطح سلولی، جهش‌ها می‌توانند باعث تغییر در عملکرد پروتئین و نظم آن‌ها شوند. جهش‌ها زمانی تکرار می‌شوند که سلول تکثیر می‌گردد. دی.ان.آ آسیب دیده و جهش یافته به هم مرتبط هستند زیرا دی.ان.آهای آسیب دیده اغلب منجر به ایجاد خطاهایی در سنتز دی.ان.آ در طول فراند تکثیر یا بازسازی می‌شوند، این خطاها منبع اصلی جهش هستند [۴۲].

۲.۱۳.۱ مکانیسم‌های بازسازی دی.ان.آ

سلول‌ها در صورتی که دی.ان.آ آسیب دیده منجر به از دست دادن یکپارچگی و عدم دسترسی به اطلاعات ضروری ژنوم گردد نمی‌توانند فعالیت نمایند، اما بقیه سلول‌ها در زمانی که ژن‌های غیر اساسی از بین می‌روند و یا آسیب می‌بینند تا حدی به فعالیت می‌پردازند. بسته به نوع آسیب وارد شده، روش‌های بازسازی مختلفی برای حفظ و یابی اطلاعات به کار بردۀ می‌شود. در ابتدا آسیب دی.ان.آ به صورت موضعی تعیین می‌گردد، مولکول‌های خاص برای بازسازی دی.ان.آ به آن و یا در نزدیکی مکان آسیب اتصال می‌یابند که منجر به القای اتصال مولکول‌های دیگر و ایجاد ساختار پیچیده‌ای می‌شود که منجر به ایجاد بازسازی حقیقی و واقعی می‌گردد. بعد از وارد شدن آسیب به دی.ان.آ، نقاط وارسی چرخه سلولی فعال می‌گردد. فعال شدن نقاط کنترل منجر به توقف چرخه سلولی می‌شود و به سلول زمانی را برای بازسازی آسیب قبل از ادامه فعالیت تقسیم می‌دهد [۴۳].

۱۴.۱ دومین‌های پروتئینی

تجزیه و تحلیل ساختارهای پروتئینی معمولاً با تجزیه پروتئین به دومین‌ها آغاز می‌شود. هدف اساسی این تجزیه کاهش یک ساختار پروتئین پیچیده به مجموعه‌های ساده و در عین حال به لحاظ ساختاری، واحد معنادار است، که هر کدام می‌توانند به طور مستقل مورد تجزیه و تحلیل قرار گیرند. دومین‌ها می‌توانند به عنوان واحدهای اساسی از ساختارهای پروتئینی شناخته شوند که خصیصه‌های اصلی در کل پروتئین را دارا هستند. برخی از ویژگی‌ها شامل : ثبات، یکپارچگی، حضور در هسته آبگریز، و توانایی در پیچیده شدن به طور مجزا است. این ویژگی‌ها

باعث شده است تا پیشنهاد شود که برهمنش بین اتمی داخل یک دومین خیلی گسترده‌تر از برهمنش بین دو دومین است. از این رو، دومین‌ها را می‌توان گروههایی با حداکثر ارتباط اتمی داخل گروهی و حداقل ارتباط بین گروهی در نظر گرفت. دومین‌ها بلوک‌های ساختمانی پروتئین هستند که ترکیب‌های مختلف آن‌ها به ساختارهای پروتئینی مختلف منجر می‌شود. دو روش برای شناسایی دومین‌ها وجود دارد. اگر دومین‌ها به طور دستی و نیمه اتوماتیک در آزمایشگاهها و با توجه به فشردگی و حفاظت از پروتئین شناسایی شوند، آن‌ها را پروتئین دومین‌های ساختار محور می‌نامند. اما اگر شناسایی دومین‌ها بر اساس توالی آن‌ها انجام شود، آن‌ها را پروتئین دومین‌های توالی محور می‌نامند. روش‌های توالی محور نرم‌افزاری هستند. برای دومین‌های ساختار محور می‌توان از دو روش معروف کت و اسکاپ نام برد. روش‌های توالی محور بسیاری وجود دارد که شناخته‌شده‌ترین آن‌ها پی‌فم است. دومین‌ها در اسکاپ براساس تاخوردگی و در کت براساس هندسه کلاس‌بندی می‌شوند. انواع زیادی از ساختارهای پروتئینی وجود دارد که از ترکیب دومین‌های مختلف با ساختار واحد به دست آمده‌اند. غالباً، اطلاعات توالی پروتئین برای مشخص کردن ساختار آن کافی نیست زیرا ساختارهای یکسانی را می‌توان با توالی های متقاوی به دست آورد. بنابراین، آشنایی با ساختار پروتئین در دست‌یابی به دومین‌ها مفید خواهد بود. روش‌های ساختاری به روش‌های محاسباتی الیت دارند. زیرا در روش‌های دستی هزاران الگوریتم براساس دانش اولیه، زیست‌شناسی و حساسیت‌های ساختاری به کار می‌رود و در نهایت ساختاری به دست می‌آید که ممکن است مغایر با حالت‌های قبلی بوده باشد. روش‌های محاسباتی، بر خلاف روش‌های دستی قادر نیستند ناسازگاری‌های درونی که ناشی از طبیعت و پیچیدگی زیستی است را شناسایی کنند. برای همین روش‌های محاسباتی دارای مشکلاتی هستند اما بسیار سریع عمل می‌کنند و با توجه به گسترده‌گی و وسعت مطالعات می‌توانند کارآیی مناسبی در تحقیقات داشته باشند [۴۴].

۱۵.۱ آمار و کاربرد آن در بیوانفورماتیک

بیوانفورماتیک به معنای ادغام شاخه‌هایی از علوم ریاضی، آمار و احتمالات و علوم کامپیوتر به منظور تجزیه و تحلیل داده‌های خیلی حجمی بیولوژیکی است. مکانیزم‌های تصادفی زیادی در ایجاد داده‌های بیوانفورماتیکی نقش دارند. به این دلیل که در تکامل ژنتیکی بسیاری از فرآیندها تصادفی اتفاق می‌افتد. در حال حاضر الگوریتم‌های آماری زیادی در بیوانفورماتیک استفاده می‌شود. مانند بلست^۱ که ابزاری برای جستجوی توالی‌یابی محلی ابتدائی است. این ابزار از اولین ابزارهای علم آمار است که به کمک زیست‌شناسان رفته است. زنجیره‌های مارکوف^۲ و مدل مخفی

^۱BLAST

^۲Markov chains

مارکوف^۱ هم از دیگر ابزارهایی هستند که به وفور در این علم برای تجزیه و تحلیل نظریه تکامل استفاده می‌شوند. اما استفاده از آمار و احتمالات در رشتۀ بیوانفورماتیک فقط به این موارد ختم نمی‌شود بلکه شاخه بسیار وسیعی از این علم را در بر می‌گیرد. استفاده از هر کدام از این شاخه‌ها بستگی به نوع سؤال طرح شده در بیوانفورماتیک دارد. در اینجا نیز با توجه به نوع مسئله نیاز است تا برای تجزیه و تحلیل داده‌ها از علم آمار و احتمالات بهره گرفته شده است.

۱.۱۵.۱ توزیع احتمال دوجمله‌ای

توزیع دوجمله‌ای نوعی توزیع پرکاربرد در آمار، اقتصاد، و علوم تجربی است. در نظریه آمار و احتمال توزیع دوجمله‌ای، توزیعی گسسته از تعداد موفقیت‌ها در دنباله‌ای شامل n آزمایش مستقل برنولی با احتمال موفقیت p است. در واقع متغیر تصادفی X (تعداد موفقیت‌ها) را متغیر دوجمله‌ای با پارامترهای n و p می‌گویند. یک آزمایش دوجمله‌ای بایستی دارای ویژگی‌های زیر باشد.

- آزمایش دارای n تعداد آزمون یکسان و عیناً مشابه باشد.
- نتیجه هر آزمون فقط به یکی از این دو صورت باشد: موفق یا ناموفق.
- احتمال موفقیت آزمونی را اگر با p نشان می‌دهند، از آزمون دیگر یکسان باشد و تغییر نکند.
- احتمال ناموفقیت را با q نشان می‌دهند که برابر است با $1 - p$.
- آزمون‌ها مستقل باشند.

در حالت کلی اگر X یک متغیر تصادفی دوجمله‌ای با پارامترهای p و q باشد، آن را بصورت $X \sim B(n, p)$ نمایش می‌دهند. تعداد راههای ممکن در انجام n آزمایش برنولی که به k موفقیت منتهی می‌شود برابر است با تعداد دنباله‌های مختلف به طول n از حروف a, b با k حرف a (موفقیت) و $n - k$ حرف b (شکست). اما تعداد این دنباله‌ها برابر است با $\binom{n}{k}$ ، زیرا تعداد جایگشت‌های متمایز n حرف از دو نوع مختلف، که k تا همانند از نوع اول و $n - k$ تا همانند از نوع دوم وجود دارد برابر است با $\frac{n!}{k!(n-k)!}$. با توجه به استقلال آزمون‌ها، چون احتمال هر یک از این دنباله پیشامدها برابر $p^k(1-p)^{n-k}$ است، احتمال به دست آوردن k موفقیت با تابع جرم احتمال زیر مشخص می‌شود:

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n \quad (1.1)$$

^۱Hidden Markov Model

دلیل اینکه به این توزیع دو جمله‌ای می‌گویند این است که قضیه بسط دو جمله‌ای که در فرمول ۱.۲ آمده است، تضمین می‌کند که $Pr(X = k)$ یک تابع جرم احتمال باشد. به عبارت دیگر:

$$\sum_{k=0}^n Pr(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1 \quad (2.1)$$

یعنی مجموع تمام احتمال‌های حساب شده برای تمامی حالت‌های ممکن k برابر یک خواهد بود.

۲.۱۵.۱ توزیع نرمال

متغیر تصادفی X را که منحنی توزیع آن زنگی شکل باشد متغیر تصادفی نرمال می‌نامند و با نماد $n(x; \mu, \sigma)$ نشان می‌دهند. تابع چگالی احتمال این توزیع به صورت زیر است.

$$n(x; \mu, \sigma) = f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R} \quad (3.1)$$

متغیر تصادفی نرمال با میانگین μ و واریانس σ^2 را متغیر تصادفی نرمال استاندارد می‌نامند و آن را با Z نشان می‌دهند. اگر X یک متغیر تصادفی دو جمله‌ای با میانگین np و واریانس $(1-p)n$ باشد، شکل حدی توزیع متغیر تصادفی $Z = \frac{X-np}{\sqrt{npq}}$ هنگامی که n به سمت بینهایت می‌کند، دارای توزیع نرمال استاندارد است. تقریب دو جمله‌ای به وسیله توزیع نرمال حتی برای مقادیر کوچک n هم خوب است. پس می‌توان از فرمول زیر در موقع لزوم استفاده کرد.

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) \quad (4.1)$$

۳.۱۵.۱ آزمون فرض

آزمون فرض یکی از مهم‌ترین شاخه‌های نظریه تصمیم‌گیری است. فرض یا بیان یا حدسی درباره توزیع جامعه یا پارامترهای جامعه را فرض آماری می‌نامند و درست بودن فرض را باید از نتایج به دست آمده از نمونه‌گیری بررسی کرد. واضح است که درستی یا نادرستی یک فرض وقتی به طور دقیق معلوم می‌شود که تمام جامعه مورد مطالعه قرار گیرد که این کار در خیلی از موارد عملاً امکان‌پذیر نیست. لذا، به نمونه‌گیری از جامعه پرداخته می‌شود و تشخیص درست یا نادرست بودن فرض بر اساس نمونه، مورد بررسی قرار می‌گیرد. این عمل را آزمون فرض آماری که برای رد شدن تنظیم می‌شود را فرض صفر (خنثی) می‌نامند و آن را با H_0 نشان می‌دهند. فرض آماری که در مقابل فرض صفر قرار می‌گیرد را فرض مقابل نامیده و آن را با H_1 نشان می‌دهند. فرض صفر را همیشه شرایط موجود در نظر می‌گیرند و فرض مقابل، همواره طوری انتخاب می‌شود که می‌خواهد

شرایط موجود را تغییر دهد. رد یا قبول یک فرض آماری به کمک نمونه انجام می‌شود، لذا وقوع خطا امکان‌ناپذیر است. امکان دو نوع خطا اجتناب ناپذیر است، یکی آن که H_0 رد شود در صورتی که درست باشد (خطای نوع اول) و دیگری آن که فرض H_0 پذیرفته شود در صورتی که درست نباشد (خطای نوع دوم).

$$\alpha = P(\text{rejecting } H_0 | H_0 \text{ is true}) \quad (5.1)$$

$$\beta = P(\text{accepting } H_0 | H_1 \text{ is true}) \quad (5.1)$$

α یا همان خطای نوع اول را میزان معنی‌دار بودن یا سطح تشخیص هم می‌نامند. ناحیه‌ای را که باعث رد فرض صفر می‌شود ناحیه بحرانی یا ناحیه‌ای را که باعث قبول فرض صفر می‌شود ناحیه قبولی و مرز بین این دو ناحیه را نقطه بحرانی می‌نامند. می‌توان گفت یک آزمون فرض عبارت است از افزار فضای نمونه به دو ناحیه که ناحیه بحرانی و ناحیه قبول نامیده می‌شود. یک تصمیم گیری خوب وقتی انجام می‌شود که هر دو نوع خطا کم باشد. به عبارت دیگر نه فرضی را بی‌مورد رد و نه فرضی را بی‌مورد قبول کرده باشد. اگر در توزیع دوجمله‌ای n نسبتاً بزرگ باشد، تقریباً دارای توزیع نرمال با میانگین $np = \mu$ و $\sigma = \sqrt{np(1-p)}$ است و در نتیجه تصمیم گیری براساس آماره زیر خواهد بود.

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \quad (7.1)$$

۴.۱۵.۱ آزمون فرضیه‌های متعدد

آزمون فرضیه آماری برپایه رد فرض صفر استوار است اگر احتمال داده‌های مشاهده شده تحت فرضیه صفر کم باشد، اگر مقایسه چندگانه انجام شود و یا فرضیه‌های متعدد مورد آزمایش قرار گیرد، شанс رویدادهای نادر افزایش می‌باید در نتیجه، احتمال اشتباه رد فرضیه صفر افزایش می‌باید. تصحیح بونفرونی^۱ برای کم کردن چنین خطاهایی استفاده می‌شود. اساس این ایده به این صورت است که اگر یک آزمایش گر m آزمایش با سطح معناداری α انجام دهد، می‌بایست از نرخ خطای جمعی^۲ برای کم کردن احتمال اشتباه رد فرضیه صفر استفاده کند. یعنی سطح معناداری هر آزمون را به جای حداقل α/m در نظر بگیرد. به عبارتی اگر سطح معناداری مورد نظر برای تمام آزمایشات α باشد، پس از تصحیح بونفرونی سطح معناداری هر فرضیه برابر α/m می‌شود. اگر

^۱Bonferroni correction

^۲familywise (FWER)

یک مجموعه از فرضیه‌ها باشند و p_1, \dots, p_m اندازه پی-مقدارهای^۱ (به احتمال رد فرض صفر، پی-مقدار نیز گفته می‌شود) مربوط به آن‌ها باشند. نرخ خطای جمعی برابر با احتمال رد حداقل یک فرض درست است. تصحیح بونفرونی ادعا می‌کند که شرط رد فرض صفر $\sum_{i=1}^m p_i \leq \frac{\alpha}{m}$ برای تک تک آزمایش‌ها، خطای جمعی H_i را کنترل می‌کند که به روش زیر حساب می‌شود.

$$FWER = P\left\{ \bigcup_{i=1}^m \left(p_i \leq \frac{\alpha}{m} \right) \right\} \leq \sum_{i=1}^m P(p_i \leq \frac{\alpha}{m}) \leq m \cdot \frac{\alpha}{m} \leq m \frac{\alpha}{m} = \alpha \quad (A.1)$$

این کنترل هیچ نیازی به فرض در مورد وابستگی پی-مقدارها ندارد [۴۵].

۱۶.۱ جمع‌بندی

پروتئین‌ها مهم‌ترین ترکیبات در موجودات زنده هستند. آن‌ها به عنوان هورمون، گیرنده، ذخیره‌کننده، مدافع، آنزیم و به عنوان انتقال‌دهنده ذرات در بدن عمل می‌کنند. پروتئین‌ها به طور کلی واحد عملیاتی سلول‌های یوکاریوتی هستند که نقش سلول‌ها را مشخص می‌کنند. دومین‌های پروتئینی واحدهای مجرزا و مستقل هستند که اختلال در کار آن‌ها می‌تواند به ایجاد اختلال در کار پروتئین بیانجامد. بنابراین، این ویژگی باعث سوق مطالعه به سمت تعیین نقش دومین‌های پروتئینی در بیماری‌ها از جمله سرطان شده است. با توجه به وسعت ژنوم انسان و تعداد زیاد ژن‌ها، پروتئین‌ها و دومین‌ها دانش رایانه‌ای برای یافتن جواب سوالات بیشماری که در این زمینه وجود دارد به کمک علوم زیستی آمده است.

^۱P-value

۲ فصل

تعریف مسئله و پیشینه تحقیق

مطلوب مربوط به زن، پروتئین، دومین پروتئینی و جهش در فصل پیش شرح داده شد. در این فصل به شرح بیماری سرطان و عوامل مختلف ایجاد آن می‌برداریم. سپس به تعریف مسئله پرداخته و مطالعات پیشین انجام شده در این رابطه به طور مختصر شرح داده می‌شود و در نهایت ابزارهای آماری استفاده شده در این پایان‌نامه شرح داده می‌شود.

۱.۲ سرطان

سرطان در اثر تغییرات ژنتیکی یا غیرژنتیکی که بر سلول‌ها اعمال می‌شود به وجود می‌آید. اما مطالعه سرطان‌ها و یافتن درمان و پیشگیری از آن از اهمیت زیادی دارد زیرا امروزه عامل اصلی مرگ و میر در دنیا انواع سرطان‌ها هستند. میزان ابتلاء به سرطان بین سال‌های ۱۹۷۵ تا ۲۰۰۰ میلادی دوبرابر شده است، و همچنان در حال افزایش است و پیش‌بینی می‌شود تا سال ۲۰۲۰ مجدداً دو برابر شده و تا سال ۲۰۳۰ سه برابر شود^[۴۶]. بالغ بر یک صد نوع سرطان شناسایی شده است. با توجه به بافت و اندامی که سرطان در آن رخ می‌دهد، سرطان‌ها شناسه‌های مختلفی به خود می‌گیرند. ۸۵٪ سرطان‌ها در سلول‌های مخاطی^۱ رخ می‌دهند و جز دسته کارسینوما^۲ هستند. سرطان‌هایی که در سلول‌های لایه میانی^۳، مانند استخوان و ماهیچه رخ می‌دهند، جز دسته سارکوما^۴ و سرطان‌هایی که در بافت‌های غده‌ای^۵ مانند پستان رخ می‌دهند، جزء دسته آدنوکارسینوما^۶ هستند. سرطان‌ها مربوط به بافت‌های مختلف خصیصه‌های مختلفی دارند. برای مثال سرطان پوست دارای خصیصه‌هایی است که

^۱Epithelial

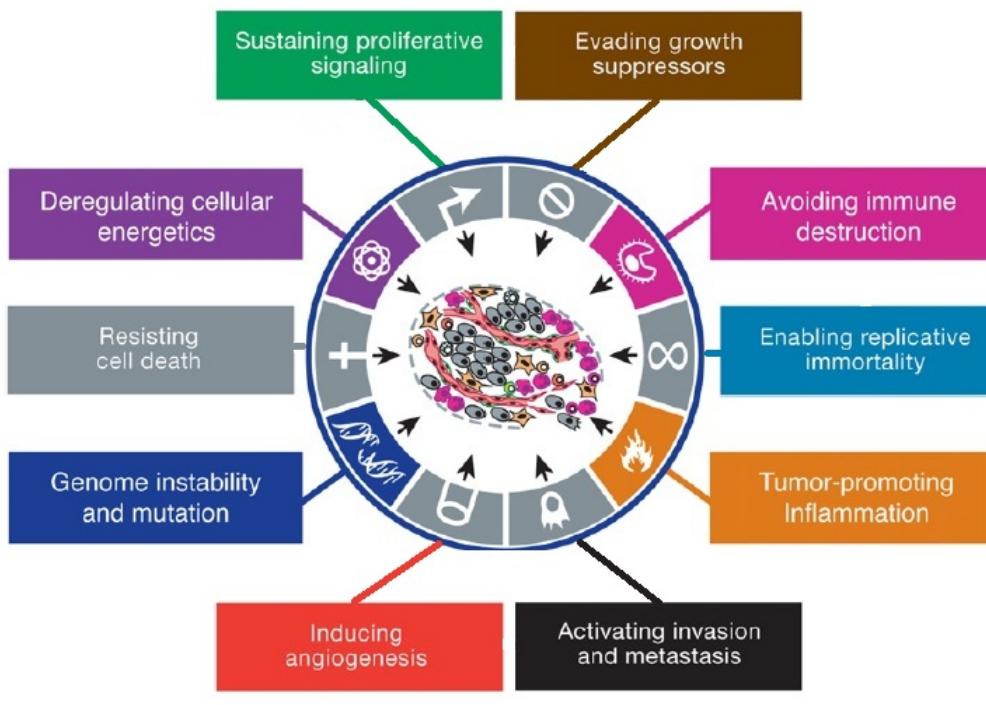
^۲Carcinoma

^۳Mesoderm

^۴Sarcoma

^۵Glandular

^۶Adenocarcinoma



[۱۹]

شکل ۱.۲: نشانه‌های سلول‌های سرطانی [۱۹]

آن را از سرطان ریه متمایز می‌کند. عوامل مختلفی در ایجاد سرطان‌ها دخیل هستند. برای مثال، اشعه فرابنفش خورشیدی به راحتی می‌تواند بافت پوست را مورد حمله قرار دهد، و دود سیگار و مواد دخانی می‌توانند بافت ریه را تخریب کنند [۸]. با وجود تفاوت‌ها بین سرطان‌ها، آن‌ها دارای شباهت‌هایی هم هستند که می‌توان آن‌ها را به صورتی که در شکل ۱.۲ نشان داده است به ۱۰ گروه دسته‌بندی کرد [۱۹]. این ۱۰ گروه عبارتند از:

- داشتن علامت رشد مستقل: سلول‌های طبیعی برای تقسیم شدن نیاز به دریافت علامت رشد دارند که از خارج به آن‌ها می‌رسد، اما سلول‌های سرطانی تحت تأثیر این عامل نبوده و سرعت رشدشان زیاد است.
- طفره رفتن از علامت منع رشد: رشد سلول‌ها با سیگنالی به نام علامت منع رشد کنترل می‌شود. سلول‌ها با دریافت این سیگنال از رشد باز می‌مانند. سلول‌های سرطانی این سیگنال را نادیده می‌گیرند.
- طفره رفتن از مرگ: سلول‌های طبیعی با دریافت علامتی از بین می‌روند، اما سلول‌های سرطانی از دریافت این عامل هم طفره می‌روند.
- داشتن نیروی تکرارپذیری نامحدود: سلول‌های طبیعی قدرت تقسیم‌پذیری محدودی دارند، اما برای سلول‌های سرطانی تکرارپذیری نامحدود است.
- داشتن قدرت رگ‌زایی: سلول‌های طبیعی مواد غذایی و اکسیژن مورد نیاز خود را از طریق رگ‌ها دریافت

می‌کنند که میزان رگ‌ها در بدن انسان‌های سالم تقریباً ثابت است، اما سلول‌های سرطانی قدرت رگ‌زایی دارند و در صورت نیاز خود به مواد غذایی و اکسیژن بیشتر، بدن را وادار به رگ‌زایی می‌کنند.

- تهاجم به سایر بافت‌ها: سلول‌های طبیعی موقعیت خود در سلول را حفظ کرده و به طور کلی مهاجرت نمی‌کنند، اما سلول‌های سرطانی به سایر بافت‌ها مهاجرت کرده و بیشترین عامل مرگ و میر سرطان‌ها هستند.

- برنامه‌ریزی مجدد انرژی متابولیسم: سلول‌های طبیعی میزان مشخصی انرژی برای بقاء و تکثیر دریافت می‌کنند، اما سلول‌های سرطانی این میزان را افزایش می‌دهند تا بقاء و تکثیر بیشتر خود را تضمین کنند.

- طفره رفتن از تخریب ایمنی: سلول‌های طبیعی مرتبأً توسط سیستم ایمنی بدن بازدید می‌شوند تا مانع رشد و یا باعث ریشه‌کنی غده‌ها شوند، اما سلول‌های سرطانی سعی می‌کنند نظارت سیستم ایمنی بر روی خود را به حداقل برسانند.

- داشتن جهش در ژنوم: سلول‌های سالم همه انسان‌ها دارای ژنوم تقریباً یکسانی هستند. اما سلول‌های سرطانی دارای جهش‌های خاصی هستند که باعث شده آن‌ها از حالت عادی خارج شده و به سلول‌های مخرب بافت تبدیل شوند.

- داشتن رفتار مشابه سلول‌های ترمیم زخم: در مناطقی از بدن که دچار زخم می‌شود. سلول‌ها نیاز دارند انرژی و غذای زیادی جذب کنند، قدرت رگ‌زایی داشته باشند و سلول‌های جدیدی تولید کنند. سلول‌های سرطانی هم مانند این سلول‌ها رفتار می‌کنند.

دو دسته اصلی برای کلاس‌بندی ژن‌های سرطانی وجود دارد: آنکوژن‌ها^۱ و ژن‌های سرکوب‌گر غده^۲. آنکوژن‌ها یا ژن‌های تومورزا ژن‌های تغییر یافته‌ای هستند که در حالت عادی پروتئین‌هایی، که در کنترل رشد و تکثیر سلول‌ها نقش دارند را بیان می‌کنند. آنکوژن‌ها باعث بروز سرطان می‌شوند. جهش‌های موجود بر روی این ژن‌ها اغلب باعث بیان بیش از حد فاکتورهای کنترلی، افزایش تعداد ژن‌های کد کننده آن‌ها و یا تغییر فاکتورهای کنترلی بصورتی که فعالیت فاکتورهای، افزایش یابد و یا نیمه عمر آن‌ها در سلول زیاد شود، می‌گردد. ژن سرکوب‌گر غده، ژنی است که سلول را از رسیدن به راه سرطان باز می‌دارد. اگر در این ژن جهشی روی دهد که به از دست دادن یا کاهش کارایی بینجامد، سرطان گسترش خواهد یافت. این روند بطور معمول با دگرگونی‌های ژنتیکی دیگری نیز

¹Oncogenes

²Tumor suppressor genes

همراه می‌باشد [۸]. عوامل محیطی که باعث ایجاد سرطان می‌شود را می‌توان در چهار دسته جای داد. این عوامل بر بافت‌ها اثر کرده و باعث تغییرات غیرژنتیکی و ایجاد سرطان می‌شوند:

- تابش اشعه آفتاب، یا سایر اشعه‌های مضر: که در برخی محیط‌های کاری وجود دارد.
- تغذیه و رژیم غذایی نامناسب: مثلاً خوردن مواد غذایی که در آن‌ها از مواد نگهدارنده استفاده شده است. یا خوردن غذاهای داغ که به ایجاد سرطان مری کمک می‌کند.
- استعمال دخانیات: که باعث ایجاد حداقل ۸۱ نوع سرطان در اندام‌های ریه، پانکراس، مثانه، کلیه، دهان و شکم می‌شود. ۴۰٪ مرگ و میر سرطان‌ها به خاطر کشیدن سیگار است.
- سن زایمان در زنان: مثلاً بالا بودن سن زنان در زمان اولین زایمان خطر ابتلاء به سرطان پستان را افزایش می‌دهد [۸].

در دسته‌بندی دیگری می‌توان علل پیدایش سرطان را به دو دسته ژنتیکی^۱ و وراژنتیک^۲ تقسیم کرد. علل ژنتیکی که باعث بروز تغییرات در دی‌ان‌آ می‌شوند، شامل جهش‌های نقطه‌ای، اسنیپ، تغییرات تعداد کپی^۳ است. عوامل وراژنتیک، عواملی هستند که بدون ایجاد تغییر در توالی دی‌ان‌آ باعث تغییر در بیان ژن‌ها می‌شوند، مانند تغییرات میزان متیلاسیون^۴، تغییرات پس‌ترجمه^۵، پروتئین‌های هیستونی و تغییرات ریز آر‌ان‌آ^۶.

۲.۲ میتوکندری و سرطان

میتوکندری می‌تواند در چندین نوع بیماری تأثیر داشته باشد مثلاً می‌توان به اختلال عملکرد قلبی اشاره کرد. میتوکندری همچنین می‌تواند در پروسه پیری نیز مؤثر باشد. تحقیقات و پژوهش‌های اخیر نشان می‌دهند که اوتیسم به ویژه اوتیسم حاد با نقص در سیستم میتوکندری در ارتباط است. دیگر بیماری‌های با علت مربوط به اختلال در میتوکندری شامل اسکیزوفرنی، اختلال دوقطبی، زوال عقل، آلزایمر، بیماری پارکینسون، صرع، سکته مغزی، بیماری‌های قلبی عروقی، سندروم خستگی مزمن، و دیابت می‌باشد [۳۰]. در اکثر این موارد عامل توارث در ابتلاء به این بیماری نقش زیادی را ایفا می‌کند [۳۰]. اتو والبرگ در سال ۱۹۳۷ میلادی نشان داد [۴۷] که سلول‌های سرطانی، لاکتوز را در حضور اکسیژن هم تولید می‌کنند. وی فرض کرد که یک نقص در تنفس میتوکندری در

^۱Genetic

^۲Epigenetic

^۳Copy Number Variation(CNV)

^۴Methylation

^۵posttranslational modifications

^۶microRNA

سلول‌های سرطانی می‌تواند دلیلی بر این مشاهده باشد. این اولین تلاش برای نشان دادن نقش میتوکندری در سلول‌های سرطانی بود. در طول دهه‌های گذشته مطالعاتی بر روی بیولوژی میتوکندری انجام شده است و نقش آن در مرگ کنترل شده سلول‌ها و یا تأثیر آن در ایجاد برخی بیماری‌ها مشخص شده است. اما به طور کلی مطالعات کمی بر روی جهش‌های بالقوه‌ای که می‌تواند فعالیت میتوکندری در سرطان را تحت تأثیر قرار دهد و یا نقش ضدسرطانی و درمانی میتوکندری که می‌تواند ایفا کند، انجام شده است. در حالی که جهش‌های مختلفی بر روی میتوکندری اتفاق می‌افتد و گاه‌آماً به حذف بخشی از آن می‌انجامد اما بعيد است که به حذف کلی آن بیانجامد. به این ترتیب، سلول‌هایی با میتوکندری‌های جهش یافته مختلف می‌تواند مدل مناسبی برای تحقیق بر روی تأثیر این جهش‌ها بر کارکرد سلول و عوامل ضد سرطان باشد. حدود ۸۳ عدد پروتئین برای انجام عمل اکسیده شدن فسفر^۱ لازم است که ۱۳ عدد از این پروتئین‌ها در میتوکندری قرار دارند. اگرچه میتوکندری مکانیسم لازم برای بیان ژن‌هایش را دارد اما بسیاری از اجزای پروتئینی آن به وسیله دی‌ان‌آ هسته‌ای کدگذاری می‌شود. با توجه به این که میتوکندری اینترون خیلی کمی دارد بنابراین بسیاری از جهش‌ها بر روی مناطق کدکننده پروتئین رخ می‌دهند و تأثیرات بیولوژیکی دارند. اگرچه مکانیسم‌های بین سلولی برای تقلیل دادن اثر سوء جهش‌ها در میتوکندری وجود دارد، این پدیده در مخمرها و گیاهان به خوبی مشهود است اما در مورد پستانداران هنوز مشاهده نشده است. شاید به خاطر تک وارث بودن میتوکندری در این گونه از جانداران باشد. برخی معتقدند دلیل آن می‌تواند به خاطر جلوگیری از ایجاد اختلال در کار میتوکندری با افزایش سن باشد. برخی نیز معتقدند این روند میتوکندری است حفظ عملکرد میتوکندری حتی با وجود جهش‌ها باشد. زیرا هر سلول شامل چندین کپی از میتوکندری است و ممکن است نسخه اصلی و جهش یافته با هم وجود داشته باشند. هنگام تقسیم سلولی، میتوکندری به طور تصادفی به دو سلول دختر تقسیم می‌شود و می‌تواند کاملاً بدون تغییر باقی بماند و یا رفته جهش خورده و در نهایت کاملاً تغییر کند. به این ترتیب، اثر بیولوژیکی یک جهش می‌تواند بسته به نسبت آن در داخل میتوکندری متفاوت باشد. این اثر به ایجاد فنوتیپ‌های مختلف می‌انجامد که جهش‌های بیماری‌زای مشابهی دارند. جهش‌هایی که در میتوکندری رخ می‌دهد در دو دسته جای می‌گیرند. دسته اول جهش‌هایی هستند که به تکثیر سلول‌های سرطانی کمک می‌کنند و دسته دوم باعث سازگاری سلول‌های سرطانی با محیط جدید شده و سراحت بیماری به سایر اجزا بدن را موجب می‌شوند [۴۷].

مطالعاتی بر روی رابطه جهش‌های میتوکندری و برخی از انواع سرطان‌ها انجام شده است که در ادامه به اختصار ذکر می‌شوند.

● سرطان سینه

^۱oxidative phosphorylation

در مورد سرطان سینه به دو مطالعه می‌توان اشاره کرد. در مطالعه اول [۹] زن سالم و جهش خورده ۱۹ بیمار با هم مقایسه شده و مشخص شده است که در ۱۴ بیمار، حدود ۸۰٪ جهش‌ها در مناطق خاصی مرسوم به دی-لوپ^۱ قرار داشت. همچنین در این مطالعه سه زن 6 16S rRNA, ND2, ATPase به عنوان زن‌های کاندید مرتبط با سرطان سینه معرفی شده‌اند که جهش‌های زیادی روی خود داشتند. در مطالعه دوم [۹] که بر روی ۱۸ بیمار انجام شده است، ۶۱٪ جهش‌ها در مناطق دی-لوپ و بقیه جهش‌ها بر روی زن‌های b ND1, ND4, ND5, Cytochrome

• سرطان تخمدان

در تحقیقی بر روی ۱۰ بیمار [۹]، مشخص شده است که عمدۀ جهش‌های دی.ان.آ. میتوکندری از نوع تغییر اسید‌آمینه تیمین و گوانین به آدنین بوده است. همین‌طور اکثر جهش‌های سوماتیک به مناطق دی-لوپ دو زن 12S rRNA, 16S rRNA محدود می‌شوند.

• سرطان روده بزرگ

در تحقیقی که بر روی ۱۰ بیمار انجام شده است [۹]، مشخص شد که هفت بیمار جهش‌هایی بر روی دی.ان.آ. میتوکندری داشتند و بیشتر این جهش‌ها در تبدیل آدنین به تیمین و گوانین به سیتوزین بوده است.

• سرطان معده

در مطالعه بر روی ۳۲ بیمار دارای سرطان معده [۹]، مشخص شد که حذف بخش مشخصی از د.ان.آ. میتوکندری در ۵۴٪ بیمارها به طور یکسان رخ داده است و بخش عمدۀ جهش‌ها بر روی سه زن ND1, ND5, COXI میتوکندری قرار داشته است.

• سرطان کبد

در تحقیقاتی که انجام شده [۹]، بیان شده است که ارتباط مستقیمی بین سرطان کبد و جهش‌های موجود بر دی.ان.آ. میتوکندری وجود ندارد. به نظر می‌رسد به تحقیقات بیشتری در این زمینه نیاز است تا این عدم وابستگی اثبات و یا رد شود.

• سرطان مری

در تحقیقی [۹] ادعا شده است که در ۴۰٪ بیمارها جهش‌ها بر روی دی-لوپ مشاهده شده‌اند. در تحقیق

^۱D-Loop

دیگری این جهش‌ها بسیار نادر گزارش شده‌اند. بنابراین در مورد سرطان مری نیز به تحقیقات بیشتری نیاز است تا درستی و نادرستی ادعاهای قبلی ثابت یا رد شود.

• سرطان پانکراس

در مطالعه‌ای [۹] که بر روی ۱۵ بیمار صورت گرفته است مشخص شده است که تقریباً تمامی بیماران جهش‌هایی بر روی دی.ان.آ. میتوکندری خود دارند. البته این جهش‌ها به ژن‌های خاصی محدود نمی‌شوند و تقریباً تمامی ژن‌های میتوکندری را در بر می‌گیرند.

• سرطان کلیه

در تحقیقی [۹] به وجود رابطه بین حذف باز بر روی ژن ND1 میتوکندری در سرطان کلیه اشاره شده است. در تحقیقی دیگر [۹] به اهمیت جهش‌های موجود بر روی ژن ND3 میتوکندری در سرطان کلیه اشاره شده است. در مطالعه دیگری [۹] که بر روی ۲۵ بیمار دارای سرطان کلیه انجام شده است مشخص شده است که آنزیم میتوکندریایی و دی.ان.آ. بافت سالم و سرطانی افراد تفاوت‌های زیادی دارند، اما کاهش محتوای زیر واحد زنجیره تنفسی بیشتر از تغییرات دی.ان.آ. میتوکندری با تومور سرطانی در ارتباط است.

• سرطان پروستات

برای سرطان پروستات، تحقیقاتی انجام شده است [۹]. هرچند گمان می‌رود که ارتباطی بین جهش‌های موجود بر روی برخی ژن‌های میتوکندری و این سرطان وجود داشته باشد اما این احتمالات بسیار ضعیف هستند و نیاز به مطالعات بیشتری برای پاسخ دادن به این گمانه زنی‌ها می‌باشد.

• سرطان مغز

در مطالعات مختلف انجام شده [۹]، مشخص شده است که بین تغییرات میتوکندری و این نوع سرطان رابطه نزدیکی وجود دارد. اما این وابستگی مربوط به تغییرات تعداد کپی می‌شود نه جهش‌های موجود بر روی دی.ان.آ. میتوکندری.

• سرطان تیروئید

بیشتر از دو دهه از مشخص شدن این که سلول‌های سرطانی تیروئید شامل تعداد بالایی تغییرات غیرطبیعی میتوکندریایی هستند، می‌گذرد. در مطالعات جداگانه‌ای [۹] که بعدها صورت گرفت این ایده تقویت شد. بیشترین رابطه مربوط به وجود تغییرات در ژن‌های کدکننده مجموعه پروتئینی یک^۱ و مجموعه پروتئینی

^۱Protein Complex I

چهار^۱ بودند.

• سرطان خون

مطالعات مختلفی هم بر روی رابطه تغییرات ایجاد شده بر روی میتوکندری و سرطان خون انجام شده است[۹]. هر کدام از آن‌ها به نحوی وجود ارتباط بین این بیماری و تغییرات مختلف را گزارش کرده‌اند اما اتفاق نظری بر روی این تغییرات نیست و تقریباً تمامی ژن‌های موجود دی.ان.آ میتوکندری را شامل می‌شود.

• سایر سرطان‌ها

در مورد سرطان‌های سر و گردن، ریه و مثانه تحقیقات کمی انجام شده است [۹]. هیچ یک از این مطالعات به وجود رابطه بین جهش‌های موجود در دی.ان.آ میتوکندری و سرطان‌های مذکور پی نبرده‌اند. بنابراین به مطالعه و تحقیق بیشتری در این زمینه نیاز است تا وجود این رابطه به اثبات برسد و یا به طور کلی نقض شود.

در کل به نظر می‌رسد دو خصیصه در تغییرات میتوکندری صرف نظر از نوع تومور در سرطان‌ها مشاهده می‌شود. اول، حجم زیاد جهش‌ها از تیمین به سیتوزین و از گوانین به آدنین است. دوم، تنوع ژن‌هایی است که دچار جهش می‌شوند. اما همچنان تجزیه و تحلیل دقیق‌تر به منظور تعیین اهمیت تغییرات دی.ان.آ میتوکندری و نقش آن‌ها در ایجاد و توسعه سرطان مورد نیاز است. ممکن است برخی تغییرات دی.ان.آ میتوکندری بعد از شیمی درمانی رخ داده باشد. تحقیقات آینده باید بر روی جنبه‌های عملکردی جهش‌های میتوکندری و پاسخ‌های درمانی انجام شود و به احتمال زیاد پیامدهای بالینی قابل توجهی در پیشگیری و درمان سرطان خواهد داشت. سوالات مهمی می‌توان مطرح کرد که نیاز به پاسخ‌های دقیق‌تری دارند. مثلاً نقش دقیق دی.ان.آ میتوکندری در ایجاد و توسعه سرطان چیست؟ چگونه جهش‌ها در میتوکندری به وجود می‌آیند و تکامل یافته و همه نسخه‌های وحشی به حالت جهش یافته تبدیل شده و سرطانی می‌شوند؟ کدام جهش‌ها بیشتر منجر به ایجاد سرطان‌ها و توسعه به دیگر بافت‌ها می‌شوند؟ آیا ارتباطی مکانیکی بین جهش‌های دی.ان.آ میتوکندری، تنفس سلولی و تغییر سرعت مرگ و میر سلولی وجود دارد؟ چگونه می‌توان به طور مؤثری سلول‌های سرطانی که دی.ان.آ میتوکندری جهش یافته دارند را از بین برد و به عامل ضد سرطان پایدار تبدیل کرد؟ پاسخ به این سوالات بی‌شك می‌تواند به دانستن نقش میتوکندری در سرطان کمک زیادی بکند و در فراهم آوردن مبنایی برای طراحی داروهای جدید و مؤثر در پیشگیری و درمان سرطان بسیار کارا باشد. جدول شماره ۱.۲ نمایانگر اسامی ژن‌هایی از میتوکندری است که جهش‌های موجود بر روی آنها در ارتباط با سرطان‌های مختلف در مطالعات مختلف ذکر شده در بالا به

^۱Protein Complex IV

عنوان عوامل مهم در ایجاد و یا پیشرفت این بیماری‌ها نام برده شده‌اند [۹]. علاوه بر ۳۷ زن یاد شده، پنج رونوشت نامتعارف هم روی دی.ان.آ. میتوکندری وجود دارد که رونوشتی از برخی نقاط دی.ان.آ. میتوکندری هستند که به طور غیر طبیعی رخ می‌دهند و ممکن است باعث ایجاد برخی سرطان‌ها شوند [۴۸]. اسمی این اجزا به همراه طول و مکان رونوشت برداری آن‌ها از روی دی.ان.آ. در جدول ۲.۲ آمده است.

جدول ۱.۲: نتایج مطالعات مختلف بر روی جهش‌های میتوکندری در انواع سرطان‌ها^[۹]

Type of Cancer	Mutated Gene/Region	Respiratory Complex Affected
Breast	Dloop	None
	MT-RNR2	rRNA
	MT-ND1	Complex I
	MT-ND2	Complex I
	MT-ND4	Complex I
	MT-ND5	Complex I
	MT-CYB	Complex III
	MT-ATP6	Complex V
Ovarian	Dloop	None
	MT-RNR1	rRNA
	MT-RNR2	rRNA
	MT-CYB	Complex III
Colorectal	MT-RNR1	rRNA
	MT-RNR2	rRNA
	MT-ND1	Complex I
	MT-ND4L	Complex I
	MT-ND5	Complex I
	MT-CYB	Complex III
	MT-CO1	Complex IV
	MT-CO2	Complex IV
	MT-CO3	Complex IV
Gastric	Dloop	None
	MT-ND1	Complex I
	MT-ND5	Complex I
	MT-CO1	Complex IV
Hepatic	Dloop	None
Esophageal	Dloop	None
Pancreatic	Dloop	None
	MT-RNR1	rRNA
	MT-RNR2	rRNA
	MT-ND1	Complex I
	MT-ND2	Complex I
	MT-ND4	Complex I
	MT-ND4L	Complex I
	MT-ND5	Complex I
	MT-ND6	Complex I
	MT-CYB	Complex III
	MT-CO1	Complex IV
	MT-CO2	Complex IV
	MT-CO3	Complex IV
	MT-ATP6	Complex V
	Dloop	None
	MT-RNR2	rRNA

Prostate	MT-ND1	Complex I
	MT-ND4	Complex I
Thyroid	Dloop	None
	MT-RNR1	rRNA
	MT-RNR2	rRNA
	MT-TN	tRNA
	MT-TR	tRNA
	MT-TL2	tRNA
	MT-ND1	Complex I
	MT-ND2	Complex I
	MT-ND3	Complex I
	MT-ND4	Complex I
	MT-ND4L	Complex I
	MT-ND5	Complex I
	MT-ND6	Complex I
	MT-CYB	Complex III
Bladder, Head, Neck, Lung	MT-CO1	Complex IV
	MT-CO2	Complex IV
	MT-CO3	Complex IV
	MT-ATP6	Complex V
	Dloop	None
MDS	MT-CYB	Complex III
	MT-CO1	Complex IV
	MT-CO2	Complex IV
	MT-ATP8	Complex V
ALL	MT-TT	tRNA
	MT-ND5	Complex I
	MT-CO2	Complex IV
	MT-ATP6	Complex V

جدول ۲.۲: اسامی رونوشت‌های نامتعارف میتوکندری

Patent Name	Start Position	End Position	Length
BC018860	7380	7515	135
J01415.25	5512	5579	67
JA040725	7587-15504	7982-15888	779
JA429504	654	674	20
JA429830	5544	5566	22

۳.۲ ژن‌های سلول‌های بنیادی و سرطان

سلول‌های سلطانی همه یکسان نیستند. در یک تومور بدخیم و یا در میان سلول‌های سلطانی می‌تواند انواع مختلف سلول وجود داشته باشد. نظریه سلول‌های بنیادی سرطان، پیشنهاد می‌کند که در میان تمام سلول‌های سلطانی، سلول‌های کمی به عنوان سلول‌های بنیادی خود را تکثیر کرده و به حفظ سرطان کمک می‌کنند، این سلول‌ها بسیار شبیه به سلول‌های بنیادی طبیعی هستند به طور معمول تجدید شده و به حفظ اندامها و بافت‌ها کمک می‌کنند. در این دیدگاه، سلول‌های سلطانی که سلول‌های بنیادی نیستند، می‌توانند باعث ایجاد مشکلاتی شوند، اما آن‌ها نمی‌توانند در دراز مدت بدن را تهدید کنند^[۱۸]. این ایده که سرطان در درجه اول از یک جمعیت کوچک‌تر از سلول‌های بنیادی مشتق شده است، دارای پیامدهای مهمی است. به عنوان مثال، بسیاری از درمان‌های ضد سرطان جدید را بر اساس توانایی برای کوچک کردن توده ارزیابی می‌کنند، اما اگر درمان‌ها، سلول‌های بنیادی سلطانی را نکشند، تومور به زودی رشد خواهد کرد. می‌شود یک مقایسه با روش وجین کردن علف هرز انجام داد بر پایه این‌که، ریز ریز کردن علف هرز مهم نیست، اگر ریشه علف هرز را وجین نکنیم، بعد از چند هفته دوباره علف‌های هرز رشد خواهند کرد. مفهوم مهم دیگر این است که سلول بنیادی سلطانی که منجر به متاستاز^۱ می‌شود (متاستاز یا فراگستربه گسترش و مهاجرت سلول‌های سلطانی از یک بافت به بافت‌های دیگر، گفته می‌شود)، می‌تواند به عنوان یک مخزن سلول‌های سلطانی عمل کند و موجب عود پس از جراحی، پرتو درمانی یا شیمی درمانی شود؛ حتی وقتی همه نشانه‌های قابل مشاهده سرطان حذف شده است. یکی از مولفه‌های نظریه سلول‌های بنیادی سلطان مربوط چگونگی ایجاد سرطان است. برای این‌که یک سلول به سلول سلطانی تبدیل شود، باید تعداد قابل توجهی از تغییرات اساسی در توالی دی‌ان‌آ سلول رخ دهد. نظریه سرطان متعارف این است که هر سلول در بدن می‌تواند تغییر کند و به سلول مهاجم سلطانی تبدیل شود. سلول‌های بنیادی طبیعی، تنها سلول‌هایی هستند که خود را تکثیر می‌کنند و در نتیجه قادر به جمع آوری تغییرات لازم برای تولید سرطان هستند. بنابراین، نظریه این است که سلول‌های بنیادی سلطانی از سلول‌های بنیادی سلطانی به وجود می‌آیند. پس، مفهوم مهم دیگر نظریه سلول‌های بنیادی سلطانی این است که سلول‌های بنیادی سلطانی بسیار شبیه به سلول‌های بنیادی طبیعی بوده و بسیاری از رفتارها و ویژگی‌های سلول‌های بنیادی طبیعی را دارا هستند. سلول‌های سلطانی دیگر تولید شده توسط سلول‌های بنیادی سلطانی باید بسیاری از قوانین سلول‌های دختر در بافت‌های طبیعی را دنبال کنند. برخی از محققان می‌گویند که سلول‌های سلطانی مانند یک کاریکاتور از سلول‌های طبیعی هستند: آن‌ها نمایش بسیاری از ویژگی‌های بافت‌های سالم را نمایش می‌دهند، اما در یک مسیر انحرافی. اگر این درست باشد، پس می‌توان از اطلاعات سلول‌های بنیادی طبیعی برای شناسایی سلول‌های

^۱Metastasis

بنیادی سرطانی و نابودی آن‌ها استفاده کرد [۱۸].

اگر سلول‌های بنیادی سرطانی محرك اصلی رشد سرطان و متاستاز باشند، پس درمان مؤثر سرطان باید حمله به سلول‌های بنیادی سرطانی باشد. اگر سلول‌های بنیادی سرطان حذف نشوند، کاهش حجم تومور سرطانی یا کاهش سرعت رشد سلول‌های سرطانی ممکن است تسکین موقت ارائه دهد، اما یک درمان طولانی مدت نیست. اگر فقط سلول‌های بنیادی سرطان حذف شوند، بقیه سلول‌های سرطانی توسط سیستم ایمنی بدن مورد حمله قرار گرفته و یا به طور طبیعی خواهند مرد؛ زیرا سلول‌های سرطانی که تعریف سلول بنیادی ندارند، توانایی تولید مثل ندارد. بنابراین، درمان مؤثر سرطان این است که سلول‌های بنیادی سرطان به طور خاص مورد هدف قرار گیرند. انواع دیگر این گونه درمان‌ها به احتمال زیاد عوارض جانبی کمی دارند زیرا آن‌ها انواع دیگر سلول‌ها را دست نخورده باقی می‌گذارند. پیدا کردن چنین درمان‌هایی نیاز به دانستن بیشتر در مورد کیفیت و رفتار سلول‌های بنیادی سرطان دارد. شکل ۶.۲ سلول‌های بنیادی سرطانی و درمان‌های خاص را نشان می‌دهد [۴۹].

منظور از مجموعه ژن‌های سلول بنیادی، مجموعه‌ای از ژن‌های است که مشخصات بیان آن‌ها، یک هسته حفاظت شده محکمی از برنامه رونویسی سلول‌های بنیادی را نشان می‌دهد. گروه‌های مطالعاتی مختلفی لیست مجموعه ژن‌های سلول بنیادی را تهیه کرده‌اند. لیستی که در اینجا معرفی می‌شود لیستی است که توسط پالمر و همکاران [۱۰] در سال ۲۰۱۲ میلادی تهیه شده است. این مجموعه از یک بانک اطلاعاتی داده با کیفیت بالا به نام کونکور迪ا مشتق شده است. کونکور迪ا ترکیبی از تجزیه متنی خودکار، گزینش انسان و روش‌های نرمال کردن است [۵۰]. پیش‌تر، اسپرگر و همکاران [۵۱]، اسکوتیم و همکاران [۵۲] و آلمستراپ و همکاران [۵۳]، با روش‌های تحلیلی مختلف مجموعه ژن‌های سلول بنیادی متفاوتی را ارائه کرده‌اند. تعداد ژن‌های این مجموعه‌ها و اسامی آن‌ها در این مطالعات بسیار متفاوت بوده است. به طوری که مجموعه ژن‌های اعلام شده از سوی آلمستراپ و همکاران [۵۳]، حاوی فقط ۴۴ ژن است، در حالی که لیست اسپرگر و همکاران [۵۱] شامل ۱۰۸۵ ژن است. این چهار لیست در مجموع هیج ژن مشترکی ارائه نداده‌اند. لیستی که در اینجا شرح داده می‌شود، شامل ۱۸۹ ژن است و به چهار دسته تقسیم می‌شوند.

- ژن‌های سلول بنیادی در تکثیر دی‌ان‌آ / مازول چرخه سلولی^۱ (۵۵ ژن)
- ژن‌های سلول بنیادی در رونویسی آر.ان‌آ / مازول سنتز پروتئین^۲ (۵۳ ژن)
- ژن‌های سلول بنیادی در سوخت و ساز بدن / سیگنانلینگ هورمون / مازول سنتز پروتئین^۳ (۳۳ ژن)

^۱SCGS genes in the DNA replication / cell cycle module

^۲SCGS genes in the RNA transcription / protein synthesis module

^۳SCGS genes in the metabolism / hormone signaling / protein synthesis module

• ژن‌های سلول بنیادی در سیگنالینگ چند سلولی / سیگنالینگ ایمنی / مازول هویت سلول^۱ (۴۸ ژن)

یکی از مقالاتی که اخیراً چاپ شده است و مورد توجه زیاد محققان قرار گرفته است مطالعه‌ای است که توomasی و همکارش [۵۴] انجام داده‌اند. این مقاله با عنوان "میزان خطر سرطان بین بافت‌های مختلف می‌تواند به وسیله تعداد تقسیمات سلول‌های بنیادی توجیه شود" در سال ۲۰۱۵ چاپ شد. برخی از بافت‌های بدن میلیون‌ها بار بیشتر از برخی بافت‌های دیگر سرطانی می‌شوند. این مطالعه می‌گوید خطر طول عمر^۲ (خطر طول عمر، میزان شانس ابتلا به یک بیماری توسط یک فرد در طول زندگی را نشان می‌دهد) سرطان در بسیاری از بافت‌های بدن به میزان خیلی بالایی (۸۱٪) با تعداد کل تقسیمات سلول‌های خود تجدید پذیر عادی در ارتباط است. این تحقیق نشان می‌دهد که حدود فقط یک سوم از خطرات سرطان در بافت‌های مختلف مربوط به عوامل محیطی یا ارثی است و بخش عمده آن با "شانس بد" در ارتباط است و آن مربوط به جهش‌های تصادفی است که در جایگزینی دی‌ان‌آ^۳ در سلول‌های بنیادی غیرسرطانی رخ می‌دهد. خطر طول عمر برای سرطان‌های مختلف مشخص شده است برای مثال سرطان ریه ۶.۹٪ خطر طول عمر دارد. این خطر برای سرطان تیروئید ۱۰.۸٪ و برای سرطان مغز ۶.۰٪ است. برخی از تفاوت‌ها در میزان ابتلا مربوط به عوامل شناخته شده‌ای مانند استعمال دخانیات، الکل و تابش اشعه ماوراء بنفسج یا ویروس پاپیلومای انسانی می‌باشد. عامل دیگر در ابتلا به سرطان عوامل ارثی است. فقط پنج تا ۱۰٪ سرطان‌ها عامل ارثی دارند. این مطالعه علاوه بر این دو عامل، عامل سومی را هم معرفی می‌کند و آن طول عمر تعداد تقسیم سلولی در هر بافت است. در علم سرطان شناسی برای هر عامل غیر ارثی، لفظ "محیطی" به کار می‌رود. در این مطالعه نشان داده شده است نه تنها تأثیرات اتفاقی از جایگزینی دی‌ان‌آ به طور عدی قابل تخمین زدن است و مجزا از عوامل محیطی است بلکه بسیار پر اهمیت‌تر از دو عامل ارثی و محیطی است. در این مطالعه، ۳۱ سرطان مختلف مورد بررسی قرار گرفت و با استفاده از روش‌های یادگیری ماشین، این بیماری‌ها به دو دسته تقسیم شده‌اند. نه عدد از بیماری‌ها بیشتر تحت تأثیر عوامل محیطی و ارثی قرار داشتند و ۲۲ بیماری باقی مانده بیشتر تحت تأثیر جهش‌های تصادفی در جایگزینی دی‌ان‌آ بودند. این دسته بندی بدون داشتن اطلاعاتی در مورد عوامل محیطی و ارثی انجام شده است و تنها به طول عمر خطر بیماری و تعداد تقسیمات سلول‌های بنیادی بسته است. برای بیماری‌های دسته اول که کمتر تحت تأثیر عوامل ارثی هستند می‌توان به درمان‌های پیشگیری در عدم ابتلا به این بیماری‌ها امیدوار بود. اما برای دسته دوم بیماری‌ها که بیشتر تحت تأثیر عامل سوم یعنی جهش‌های تصادفی در جایگزینی دی‌ان‌آ هستند، باید بیشتر به درمان سرطان بعد از ابتلا امیدوار بود تا پیشگیری از آن [۵۴].

^۱SCGS genes in the multicellular signaling / immune signaling / cell identity module

^۲Lifetime Risk

^۳DNA Replication

۴.۲ بازسازی دی.ان.آ و سرطان

به دلیل محدودیت‌های وراثتی در زمینه بازسازی دی.ان.آ، اگر انسان‌ها به اندازه کافی عمر نمایند در این صورت آن‌ها ممکن است گاهی اوقات مبتلا به سرطان شوند حداقل ۳۴ جهش در ژن بازسازی دی.ان.آ انسان به صورت ارثی وجود دارد که خطر ابتلا به سرطان را افزایش می‌دهد. بیشتر این جهش‌ها منجر به بازسازی دی.ان.آ به صورت کمتر موثر در مقایسه با حالت عادی می‌شوند. به صورت خاصی، سرطان کولورکتال غیر پولیپوزیس ارثی^۱ به مقدار زیادی در ارتباط با جهش‌های خاصی در روش بازسازی عدم انطباق دی.ان.آ است. ژن‌های BRCA1 و BRCA2، دو ژن معروف هستند که جهش‌های صورت گرفته در آن‌ها به مقدار زیادی امکان ابتلا به سرطان را افزایش می‌دهد. زمانی که ۱۱۳ توالی مربوط به افراد دارای سرطان روده بزرگ مورد آزمون قرار گرفت تنها چهار نفر فاقد جهش در ژن بازسازی دی.ان.آ با نام MGMT بودند، در حالی که اکثر آن‌ها دارای بیان کاهش یافته‌ای از MGMT بودند که ناشی از متیلی شدن ناحیه شناساگر MGMT بود[۵۵]. در پنج مطالعه[۵۵] مختلف نشان داده شده که ۹۰٪ تا ۴۰٪ از افراد مبتلا به سرطان روده بزرگ دارای بیان کاهش یافته‌ای از MGMT بودند که این امر ناشی از متیل شدن ناحیه شناساگر MGMT بود. در مطالعه[۵۵] مشابهی به غیر از ۱۱۹ بیمار مبتلا به سرطان روده بزرگ دارای نقص بازسازی غیر انطباقی که ناشی از فاقد ژن بازسازی دی.ان.آ با نام PMS2 است، بودند. شش مورد نقص در ژن PMS2 داشتند. در مورد ۱۰ بیمار دیگر، از دست دادن بیان شدن PMS2 به احتمال زیاد ناشی از بیان شدن بیش از حد miR-155 microRNA بود که به مقدار کمتری به تنظیم ژن BRCA1 می‌پرداخت. مشکلات اپی‌ژنتیک به میزان ۱۳٪ تا ۱۰۰٪ برای ژن‌های بازسازی دی.ان.آ، MLH1 و WRN، FANCB، FANCF، MGMT، MLH1، MSH2، MSH4، ERCC1، XPF، NEIL1 و ATM مشاهده شدند. این مشکلات و نقص اپی‌ژنتیکی در سرطان‌های مختلفی رخ می‌دهد (برای مثال سرطان پستان، تخم‌دان، روده بزرگ و سر و گردن). دو یا سه نقص در بیان ژن ERCC1، XPF یا PMS2 به صورت همزمان در اکثر ۴۹ مورد مبتلا به سرطان روده بزرگ رخ داده است که به وسیله فاسیتا و همکاران[۵۵] مورد ارزیابی قرار گرفته است.

۵.۲ مطالعه یکپارچه سرطان‌ها

تجزیه و تحلیل یکپارچه سرطان‌ها^۲ با هدف یافتن تفاوت‌ها و شباهت‌های تغییرات ژنومی و سلوی در میان انواع تومورهاست. در مطالعه یکپارچه سرطان‌ها مانند مطالعه تکی سرطان‌ها از داده‌های مختلف در دسترس مانند

^۱HNPPCC

^۲Pan-Cancer

جهش‌های ژنوم، میزان متیلاسیون، اطلاعات بالینی، توالی آر.ان.آ و غیره استفاده می‌شود. اهمیت مطالعه یکپارچه سلطان‌ها به دلیل رفتار یکسان آن‌ها در سرعت رشد بالا، داشتن قدرت رگ‌زایی، نامیرایی و تهاجم به سایر بافت‌هاست. اما با وجود این شباهت‌ها، دارای تفاوت‌هایی هم هستند که آن‌ها را از هم متمایز می‌کند. مطالعات گوناگونی در زمینه مطالعه یکپارچه سلطان‌ها صورت گرفته است که در ادامه به شرح خلاصه‌ای از نتایج یکی از این مطالعات پرداخته خواهد شد.

در این مطالعه^[۵۶] که بر مبنای تجزیه شبکه‌ای یکپارچه سلطان‌ها^۱ بر روی جهش‌های ژنوم ۳۲۸۱ بیمار از ۱۲ گونه مختلف سلطان انجام شده است، نشان می‌دهد علاوه بر برخی ژن‌های شناخته شده، ژن‌هایی نیز وجود دارند که در مطالعه مجزای سلطان‌ها، جهش‌های بسیار کمی دارند و گفته می‌شود نقش کمرنگی در ایجاد سلطان‌ها داشته باشند، اما برخی از همین ژن‌ها در مطالعه یکپارچه سلطان‌ها جهش‌های زیادی را در بر می‌گیرند و از این دیدگاه به نظر می‌رسد که باید نقش پررنگی در ایجاد سلطان‌ها داشته باشند. این تحقیق به دنبال مطالعه یکپارچه سلطان، به معنی ۱۴۷ ژن می‌پردازد که متعلق به ۱۶ زیرشبکه مختلف هستند و خاصیت گفته شده را دارند. یعنی در مطالعه تکی سلطان‌ها جهش‌های زیادی ندارند ولی در مطالعه یکپارچه جهش‌های زیادی را به خود اختصاص می‌دهند. جدول ۳.۲ اسامی این ۱۴۷ را نشان می‌دهد. برای انجام این مطالعه از توالی ژنوم، تعداد کمی آرایه^۲ و توالی آر.ان.آ استفاده شده است. بعد از غربال جهش‌ها و انتخاب تغییرات تک نوکلئوتیدی، حذف و اضافه‌های کوچک در ۱۹۴۲۴ ژن، حذف نمونه‌هایی با تعداد جهش‌های بالا و ژن‌های با بیان بسیار کم در همه سلطان‌ها، ۱۱۵۶۵ ژن جهش یافته در ۳۱۱۰ بیمار باقی ماند که مطالعات اصلی بر روی آن‌ها صورت گرفته است. بسیاری از ژن‌های معروف و مشخص دخیل در سلطان مانند RTK, TP53, PI3K, NOTCH در نتایج این مطالعه نیز گزارش شده‌اند. اما از ۱۴۷ ژن معنی شده، ۹۲ ژن در هیچ یک از پنج مرجع مطالعه مورد مقایسه این مطالعه، گزارش نشده بود. ۱۳ ژن تنها در یکی از این پنج مرجع گزارش شده‌اند^[۵۶].

۶.۲ مقایسه دومین‌های توالی محور و ساختار محور

مطالعاتی برای مقایسه دومین‌های توالی محور و ساختار محور انجام شده است^[۲۱، ۲۰]، تا به شناسایی شباهت‌ها و تفاوت‌های موجود در این دو نوع روش مختلف بپردازند. در این بخش به بیان خلاصه‌ای از دو مطالعه می‌پردازیم که به مقایسه نگاشت دومین‌های پروتئینی ساختارمحور و توالی محور پرداخته‌اند. مطالعه اول^[۲۰] به مقایسه دیتاهای دو بانک اطلاعاتی اسکاپ و پی‌فم پرداخته که از روش زیر برای مقایسه این دو بانک اطلاعاتی استفاده کرده است. برای نگاشت دادن و مقایسه دومین‌ها بر روی یک پروتئین، دو مسئله را باید در نظر گرفت. مسئله اول

^۱Pan-Cancer Network Analysis

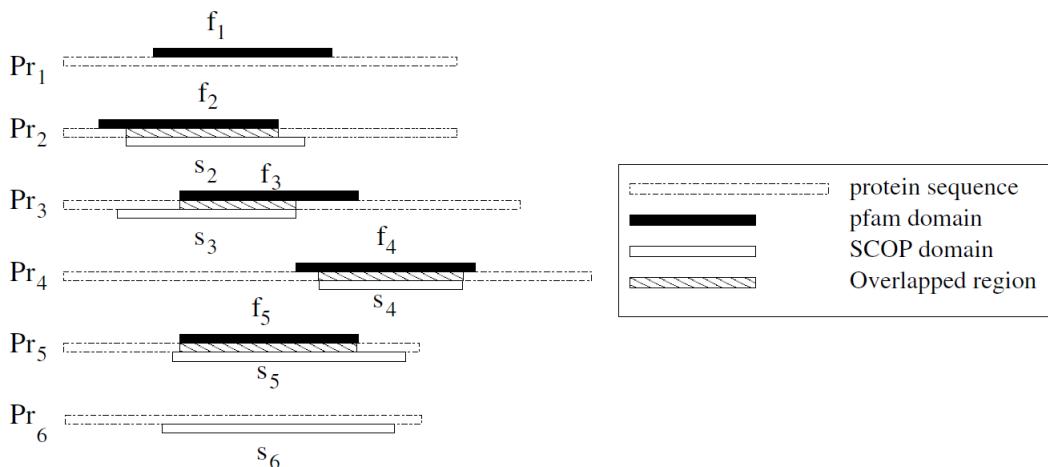
^۲Array Copy Number

جدول ۳.۲: ژن‌های انتخاب شده برای مطالعه [۷]

147 Selected Genes by Gene Landscape Method						
ADNP	CCND1	EPHA3	LPP	NRAS	SHPRH	UACA
AGBL2	CCNE1	ERBB2	LRIG1	OSMR	SMARCA4	VARS2
ALS2CR8	CD1D	ERBB4	MAGI2	P4HTM	SMARCB1	WAC
AMOTL1	CDKN2A	FAM101A	MAML1	PBRM1	SMC1A	WAPAL
ANKRD12	CDKN2AIP	FAM118B	MAST3	PDS5A	SMC2	ZFP91
ANKRD17	CDON	FBXW7	MDM4	PDS5B	SMC4	ZMIZ1
AREG	CELSR3	FOXK1	MLL2	PIK3CA	SMG1	ZNF227
ARHGAP4	CHD6	FOXK2	MLL3	PLCE1	SMG5	ZNF668
ARID1A	CHD8	FOXM1	MLL5	PLEKHA8	SMG7	ATM
ARID1B	CLASP1	HECW2	MMS22L	PRDM2	SNRK	CABLES1
ARID2	CLASP2	HLA-A	MYD88	PRKRIR	SPOP	CCDC88A
ASXL1	CLIP2	HLA-B	N4BP2	PROSER1	SPTBN2	HRAS
ASXL2	CUL9	IFT140	NCAPD2	PTEN	STAG1	MAP2K4
B2M	DCLRE1C	IWS1	NCAPD3	PTMA	STAG2	MAP3K1
BAP1	DLL1	JAG1	NCAPG2	RAD21	STK11IP	NOTCH2
BOC	DTX4	KDM1B	NCAPH2	RASGRF2	STRADA	PIK3R1
BRAF	E2F3	KDM6A	NFE2L2	RASGRP2	SWAP70	RASA1
BRPF1	EGFR	KEAP1	NOTCH1	RGL1	TEP1	RNF20
CACHD1	EIF2AK1	KLF5	NOTCH3	RUNX1	TP53	STK11
CARNS1	ELF3	KRAS	NOTCH4	SCAPER	TRIP12	TP53INP1
CBFB	ELF4	LFNG	NPM1	SETD2	TTLL5	WT1

نوع همپوشانی دو دومین از دو بانک اطلاعاتی مختلف است که سه حالت برای آن وجود دارد: حالت اول: یکی از بانک‌های اطلاعاتی عاری از دومین برای آن پروتئین است. حالت دوم هر دو، دومین‌هایی را مشخص کرده‌اند اما هیچ همپوشانی بین آن‌ها نیست و حالت سوم این است که دو دومین مشخص شده توسط دو روش، همپوشانی دارند. شکل ۲.۲ نحوه همپوشانی دومین‌ها به دو روش مختلف را نشان می‌دهد. مسئله دوم این است که ممکن است برای یک پروتئین، یک دومین از یکی از بانک‌های اطلاعاتی با دو دومین یا بیشتر از بانک اطلاعاتی دیگر همپوشانی داشته باشد. با در نظر گرفتن این دو مسئله از فرمولی برای محاسبه میزان همپوشانی هر دو دومین از دو بانک اطلاعاتی مختلف استفاده شده است. به فرض اگر s_i بیان‌گر i امین دومین اسکاپ باشد و f_j نیز بیان‌گر j امین دومین پی‌فم باشد، برای محاسبه میزان همپوشانی دومین‌ها از امتیاز $M(s_i, f_j)$ استفاده شده است که به صورت زیر محاسبه می‌شود:

$$M(s_i, f_j) = \frac{2}{freq(s_i) + freq(f_j)} \sum_{p_i \in P} \frac{overlap(s_i^k, f_j^k)}{\min(\text{length}(s_i^k), \text{length}(f_j^k))} \quad (1.2)$$

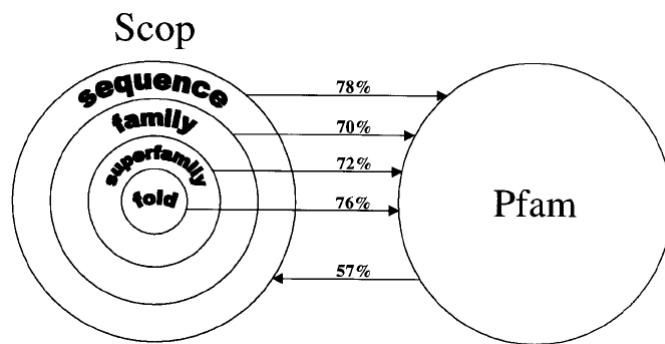


شکل ۲.۲: نحوه همپوشانی دومین‌ها بر روی پروتئین [۲۰]

در اینجا مجموعه پروتئین‌ها با P نمایش داده می‌شود و p_k نماد k امین پروتئین مجموعه است. $overlap(s_i^k, f_j^k)$ میزان همپوشانی دو دومین را نشان می‌دهد. و $length(f_j^k)$ هم به ترتیب نشان‌گر طول دومین‌های i و j ام اسکاپ و پی‌فم هستند. $freq(f_j)$ هم میزان فرکانس دومین i ام و j ام اسکاپ و پی‌فم را نشان می‌دهند. این روش بر روی ۸۲۵۹ پروتئین، مورد مطالعه قرار گرفت و مشخص شد، ۲۰۰۸ دومین از بین ۲۵۱۲ دومین اسکاپ با حداقل یک دومین پی‌فم همپوشانی دارند و از بین ۲۰۸۱ دومین پی‌فم ۲۰۷۵ عدد آن‌ها با حداقل یک دومین اسکاپ همپوشانی دارند. حدود یک پنجم دومین‌های اسکاپ هیچ همپوشانی با دومین‌های پی‌فم ندارند، و تنها شش دومین پی‌فم با هیچ دومین اسکاپی همپوشانی ندارند. همچنین در این مطالعه مشخص شد هرچه طول دومین‌ها کمتر باشد همپوشانی بین آن‌ها هم بیشتر است.

مطالعه دوم [۲۱] که در اینجا به طور خلاصه شرح داده می‌شود مجدداً مربوط به مقایسه دو بانک اطلاعاتی اسکاپ و پی‌فم است. در این مقایسه مشخص شد ۷۰٪ دومین‌های اسکاپ در بانک اطلاعاتی پی‌فم وجود دارد و ۵۷٪ دومین‌های پی‌فم هم نیز جز زیر مجموعه بانک اطلاعاتی اسکاپ هستند. در این مطالعه دومین‌های پی‌فم به چهار مرحله آخر طبقه‌بندی دومین‌های اسکاپ نگاشت داده شده‌اند. شکل ۳.۲ میزان همپوشانی دومین‌های پی‌فم و اسکاپ را در چهار مرحله مختلف نشان می‌دهد.

از دیگر نتایج این مطالعه، یافتن این موضوع بود که دومین‌های اسکاپ به طور متوسط ۲۰٪ از دومین‌های پی‌فم کوتاه‌تر بودند. همین طور، مشخص شد بیشتر خانواده‌های پی‌فم و اسکاپ با هم برابر بودند. از ۸۰۲ خانواده پی‌فم ۷۱۲ خانواده فقط با یک خانواده از اسکاپ همپوشانی داشت. و از ۶۴۲ خانواده مورد مطالعه اسکاپ خانواده فقط با یک خانواده از پی‌فم همپوشانی داشت.

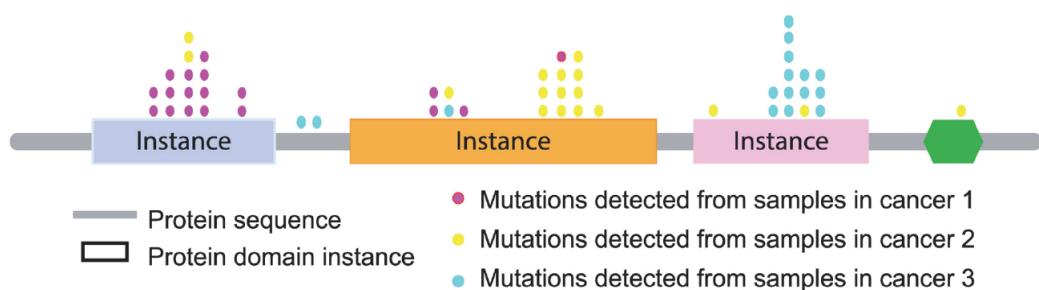


شکل ۳.۲: نگاشت دومین‌های پروتئین به دومین‌های اسکاپ [۲۱]

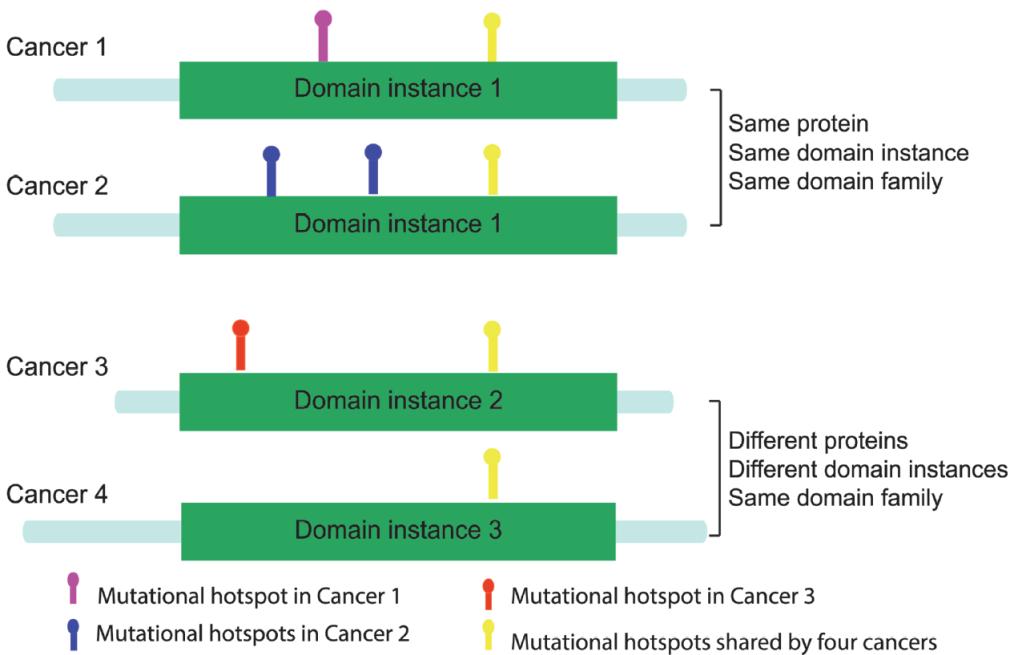
۷.۲ دومین‌های پروتئینی و سرطان

تحقیقات زیادی بر روی ژن‌های افراد مبتلا به سرطان، میزان متیلاسیون، تغییرات تعداد کپی، توالی آر.ان.آ و غیره انجام شده، اما هنوز پاسخ بسیاری از سوالات مطرح در این زمینه نامشخص است. یکی از روش‌هایی که اخیراً برای شناسایی عوامل ابتلا به سرطان به کار گرفته می‌شود، مطالعه بر روی دومین‌های پروتئینی است. با توجه به ساختار، کارکرد، اهمیت و نقش دومین‌های پروتئینی، محققان سعی دارند تا به ارتباط و شباهت الگوهای جهش بیماران دست یابند.

یکی از تحقیقات انجام شده، مطالعه بر روی جهش‌های سوماتیک بیماران سرطانی در ۲۹ بیماری مختلف است، که توسط یانگ و همکاران [۲۲] در سال ۲۰۱۵ میلادی انجام شده است. این تحقیق سه بخش مهم دارد، در بخش اول جهش‌های بی‌معنی سوماتیک بر روی دومین‌های پروتئینی بیماران مورد ارزیابی و سپس مورد مقایسه با مطالعات قبلی مشابه قرار گرفته است. دوم جهش‌های موجود بر روی پروتئین‌هایی مورد مطالعه قرار گرفته که حداقل دو دومین داشته و در حداقل یک نوع سرطان دارای جهش باشند، مشخص شد از بین ۹۲ پروتئین مورد نظر، ۵۴ پروتئین دارای رفتار تقریباً مشابهی بودند، یعنی تجمع جهش در سرطان‌های مختلف بر روی دومین‌های خاصی از پروتئین‌ها بود. شکل ۴.۲ نگاشت بخش‌های سرطان‌های مختلف به دومین‌های یک پروتئین را نشان



شکل ۴.۲: نگاشت جهش‌ها به پروتئین دومین‌ها [۲۲]



شکل ۵.۲: نقاط جهش‌یافته خاص انواع سرطان [۲۲]

می‌دهد.

سوم اینکه، آن‌ها دریافتند، نه تنها رفتار انواع سرطان‌ها در دومین‌های یکسان از خانواده دومینی یکسان بر روی پروتئین یکسان مشابه است. بلکه همین رفتار یعنی داشتن محل تجمعی جهشی خاص و متفاوت، بر روی دومین‌های مختلف از خانواده‌های دومینی یکسان در پروتئین‌های مختلف هم دیده می‌شود. شکل ۵.۲ مثالی از رفتار سرطان‌های مختلف بر روی خانواده‌های دومین‌های مختلف را نشان می‌دهد.

مطالعه دوم توسط نرت و همکاران [۷] انجام شده، و سعی در یافتن دومین‌های دخیل در سرطان روده بزرگ را دارد. آن‌ها از سرطان پستان به عنوان نوعی محب برای ارزیابی نتایج تولید شده از مطالعه سرطان پستان استفاده کرده‌اند. علاوه بر مشخص کردن دومین‌های مهم هر بیماری، دومین‌های مشترک هر دو بیماری و ارتباط این دومین‌ها با ژن‌های مهمی که از مطالعات پیشین به دست آمده‌اند را مطالعه کرده است. در اینجا نیز از داده‌های سوماتیک بیماران استفاده شده است. آن‌ها در یافتن علاوه بر جهش‌های مهمی که بر روی دومین‌های برخی ژن‌های خاص مانند APC, TP53 و KRAS قرار دارد، فرکانس جهش بالایی بر دومین‌های خاصی وجود دارد که ژن‌های شامل این دومین‌ها به تنها بی‌تعداد بسیار محدودی جهش دارند. اما مجموع جهش‌هایی برخی دومین‌های خاص بر روی این ژن‌ها معنادار است. در ادامه و فصل سه به شرح کامل این تحقیق خواهیم پرداخت.

۸.۲ بانک‌های اطلاعاتی مرتبط

در این بخش به معرفی بانک‌های اطلاعاتی مختلفی پرداخته می‌شود که برای مطالعه به اطلاعات آن‌ها نیاز است. این بانک‌های اطلاعاتی عمدهاً مربوط به مؤسسه‌های معروف بیوانفورماتیک آمریکا و اروپا و یا مؤسسه سرطان آمریکا هستند.

۱.۸.۲ بانک اطلاعاتی اطلس ژنوم سرطانی

مؤسسه اطلس ژنوم سرطانی^۱، که به بانک اطلاعاتی تی‌سی‌جی‌ای معروف است، در سال ۲۰۰۶ میلادی بعد از سه سال کار آزمایشی توسط مؤسسه ملی سرطان^۲ و مؤسسه ملی تحقیقات ژنوم انسانی^۳ و سرمایه‌گذاری اولیه یکصد میلیون دلاری این مؤسسات راهاندازی شد. مؤسسه اطلس ژنوم سرطانی از هفت مرکز تشکیل شده که هر کدام بخشی از کارهای مربوط به تولید داده از روی نمونه‌ها را انجام می‌دهند. در بخش مرکز اول، داده‌ها از لحاظ کیفیت بررسی شده و اطلاعات کاملی از بیماران جمع‌آوری شده و نمونه‌ها ذخیره می‌شوند. در مرکز دوم، تجزیه و تحلیل تغییرات ژنومی با استفاده از تکنولوژی‌های مختلف انجام می‌شود. در مرکز سوم، تغییرات دی‌ان‌آ که با سرطان خاصی در ارتباط هستند را مشخص می‌کنند. مرکز چهارم، کل داده‌های پروتئینی را تجزیه و تحلیل می‌کند. مرکز پنجم، وظیفه جمع‌آوری و مرتب کردن داده‌های تولید شده از چهار مرکز اول را به عهده دارد. داده‌ها به چهار روش مختلف و در چهار مرحله مختلف آماده شده‌اند تا برای مطالعه‌ها با اهداف مختلف مناسب باشند. در مرکز ششم، توالی ژنوم سرطانی ذخیره سازی می‌شود. و در نهایت مرکز هفتم، جدیدترین ابزارهای انفورماتیکی را برای کار بر روی نمونه‌ها و جمع‌آوری داده‌ها فراهم می‌آورد تا داده‌های با کیفیت بیشتری در اختیار محققانی که از این داده‌ها استفاده می‌کنند، قرار بگیرد. در هر مرحله ممکن است تعدادی از نمونه‌ها به خاطر نداشتن استاندارهای لازم حذف شوند. در نهایت نمونه‌های باقی مانده کیفیت بالایی خواهند داشت. اطلاعات تکیمیلی‌تر در وب سایت مؤسسه^۴ موجود می‌باشد. پژوهشگران این مؤسسه بر روی بیش از بیست نوع سرطان مختلف کار می‌کنند. این سرطان‌ها به دو دلیل انتخاب شده‌اند: یک، تشخیص ضعیف و تأثیر کلی بهداشت عمومی. دو، وجود غده‌های سرطانی و نمونه بافت‌های سالم آن‌ها که از بیماران با رضایت آن‌ها گرفته می‌شود و کیفیت و کمیت نمونه‌ها که باید طبق استانداردهای مؤسسه باشد^[۳۱]. این نمونه‌ها مربوط به بافت‌های زیر هستند:

• پستان

^۱The Cancer Genome Atlas(TCGA)

^۲National Cancer Institute(NCI)

^۳National human Genome Research Institute(NHGRI)

^۴<http://cancergenome.nih.gov/>

• سیستم اعصاب مرکزی

• داخل جمجمه

• شکمی

• سر و گردن

• خون

• پوست

• بافت‌های نرم

• قفسه سینه

• دستگاه تناسلی و ادراری

مزیت استفاده از داده‌های مؤسسه اطلس ژنوم سرطانی به این خاطر است که داده‌ها از کیفیت بالایی برخوردارند و همه داده‌ها به روز هستند. یعنی همچنان نمونه‌هایی در حال اضافه شدن به داده‌های قبلی است. برای هر نوع سرطان سعی می‌شود بیش از ۵۰۰ نمونه جمع‌آوری شود. نمونه‌های یکسانی در بخش‌های مختلف مؤسسه مورد مطالعه قرار می‌گیرد. هر نمونه با چندین پلتفرم^۱ تجزیه و تحلیل شده است. نمونه‌های سالم و نمونه‌های سرطانی هر شخص وجود دارد، و تقریباً همه داده‌ها به صورت رایگان در اختیار همه قرار می‌گیرد. برای تولید داده‌های با کیفیت بالا اقدامات مختلفی انجام می‌شود. برای داشتن داده باکیفیت، حداقل ۶۰٪ نمونه جدا شده از بافت شخص بیمار باید بافت سرطانی باشد. هیچ درمان اولیه‌ای نظیر رادیودرمانی یا شیمی‌درمانی بر روی بیماران قبل از برداشت نمونه انجام نشده باشد، زیرا این اقدامات ممکن است ژن‌های سرطانی را تغییر دهد. نمونه مربوط به غده اولیه حتماً باید موجود باشد. یعنی اگر در فردی نمونه از بافت سرطانی که در اثر گسترش سرطان به سایر بافت‌ها به دست آمده، نمونه بافت سرطانی اولیه هم موجود است. به جز نمونه‌گیری از بافت سرطانی، نمونه دی‌ان‌آ مردی به دودمان زایشی آن فرد هم برای مقایسه نمونه سرطانی با آن موجود است. این نمونه اکثراً از خون فرد گرفته شده است که برای مقایسه ژنوم سالم با ژنوم بافت سرطانی بیمار از آن استفاده می‌شود. این داده‌ها برای مطالعات تکمیلی‌تر و ایجاد بانک سلول‌های سرطانی به صورت یخزده نگهداری می‌شوند.^[۳۱]

^۱Platform

برای توالی دی.ان.آ تعیین توالی کل ژنوم^۱ و تعیین توالی کلی اگزون^۲ موجود است. برای جهش‌های سرطانی از داده‌های تعیین توالی کل اگزون استفاده شده است [۳۱]. تعیین توالی کل ژنوم پردازشی است که طی آن تعیین توالی کامل دی.ان.آ فرد به طور واحد انجام می‌شود. امروزه هزینه زیادی باید صرف این نوع تعیین توالی کرد، اما احتمالاً در آینده نزدیک این میزان بسیار کاهش یابد [۵۷]. تعیین توالی کل اگزون، تکنیکی برای تعیین توالی همه ژن‌های کدکننده پروتئین (اگزون‌ها) است. حدود ۱۸۰۰۰ اگزون در ژنوم وجود دارد که مجموعاً حدوداً ۳۰ میلیون نوکلئوتید طول دارند. یعنی تقریباً یک درصد طول کل ژنوم را شامل می‌شوند. جهش‌هایی که در ناحیه اگزون‌ها رخ می‌دهد بسیار عاقب شدیدتری نسبت به جهش‌های موجود در ۹۹٪ باقی‌مانده ژنوم دارند. هدف از تعیین توالی کل اگزون، مشخص کردن تغییرات ژنتیکی است که مسئول بسیاری از بیماری‌ها هستند، بدون اینکه هزینه بسیار زیادی را صرف تعیین توالی کل ژنوم کرده باشند. در کل، تعیین توالی ژنوم دقیق‌تری از تعیین توالی کل اگزون دارد. در مورد تغییرات تعداد کپی، تعیین توالی کل ژنوم بهتر از تعیین توالی کل اگزون عمل می‌کند، زیرا اگزون‌ها پیوسته نیستند و تغییرات تعداد کپی گسترده‌تر از مناطق تحت پوشش در روش تعیین توالی کل اگزون است. پس در روش تعیین توالی کل اگزون برخی از اطلاعات از بین می‌رود. به علاوه همه جهش‌هایی که در ناحیه خارج ژن‌ها رخ می‌دهند هم در این روش شناسای نمی‌شوند [۵۸].

۲.۸.۲ بانک اطلاعاتی ژنوم مرجع

ژنوم همه انسان‌ها باید شبیه به هم باشد مگر در جاهایی که اسنیپ وجود دارد. بنابراین یک ژنوم به عنوان ژنوم مرجع معرفی شده است که توالی همه کروموزوم‌های انسانی را نشان می‌دهد. با مقایسه این ژنوم با ژنوم افراد بیمار می‌توان نواحی تغییر یافته در ژنوم بیماران را شناسایی کرد. برای پیدا کردن جهش‌ها بر روی ژنوم انسانی، از ژنوم مرجع واحدی استفاده می‌شود تا بتوان از داده‌ها، نتایج آماری بهتری به دست آورد. در ژنوم‌های مرجع مختلف محل ژن‌های شناخته شده با هم کمی متفاوت است و این به خاطر پیشرفتی است که رفته رفته بر ژنوم مرجع اعمال می‌شود تا خطاهای موجود در ژنوم‌های مرجع قبلی تا جای ممکن تصحیح شود. برخی از نقاط ژنوم که قبلاً رمزگشایی نشده بود، در ژنوم مرجع جدید شناسایی می‌شود. هنگامی که پروژه ژنوم انسانی در سال ۲۰۰۳ کامل شد، بسیاری از افراد تصور می‌کردند که ژنوم انسان به طور کامل بازسازی شده است، اما در واقع برخلاف این تصور، بخش‌هایی از ژنوم ناشناخته باقی ماند که شامل مناطق بدون توالی مشخص در ژنوم، اطلاعات از دست رفته در مراحل مختلف پروژه، مناطقی که انطباق اشتباه^۳ داشتند و مناطقی که بد نمایش داده شده بودند، بود. برای تصحیح این خطاهای کنسرسیو ژنوم انسانی با مشارکت چهار مؤسسه معتبر در اروپا ایجاد شد تا در راستای بهبود

^۱Whole Genome Sequencing(WGS)

^۲Whole Exon Sequencing(WES)

^۳Misaligned

بخشیدن به ژنوم انسانی تلاش کند. تاکنون ژنوم‌های مرجع مختلفی در سال‌های مختلف ارائه شده و از طریق وب سایت^۱ در دسترس عموم قرار گرفته است. داشتن ژنوم انسانی کامل‌تر و با خطای کمتر از اهمیت زیادی برخوردار است زیرا محققان زیادی برای یافتن علل بیماری‌ها از جمله سرطان از ژنوم مرجع استفاده می‌کنند. پس وجود خطا در آن ممکن است موجب تفسیر نادرست از تغییرات موجود در ژنوم شود و نتایج به دست آمده درست نباشد.^[۵۹]

سختی‌های یافتن ژنوم مرجع شامل: وجود مناطق با توالی‌های تکراری یا توالی‌های متغیر است؛ مناطقی که به طور نامناسب گردآوری شده‌اند و همچنین، مناطقی که هیچ توالی برای آن‌ها پیدا نشده است. با توجه به همه این سختی‌ها ژنوم مرجع‌های مختلفی در سال‌های ۲۰۰۳، ۲۰۰۷ و ۲۰۱۳ ارائه شده است و قرار است به زودی نسخه جدیدی نیز ارائه شود. البته در فاصله معرفی ژنوم مرجع جدید، برای ژنوم مرجع به روز، ویرایش‌های^۲ جدیدی ارائه می‌شود تا ایرادات قبلی تاحدی رفع شود. داده‌هایی که از بانک اطلاعاتی موسسه اطلس ژنوم انسانی برای مطالعه در این پژوهه استفاده شده‌اند، بر پایه ژنوم مرجع ارائه شده در سال ۲۰۰۷ تفسیر شده‌اند. البته به جز ژنوم انسان، ژنوم مرجع موش و ماهی گورخری هم در این کنسرسیوم مورد مطالعه قرار می‌گیرد.^[۶۰] برای تهیه ژنوم مرجع از تعدادی فرد داوطلب استفاده می‌شود. ژن‌ها متعلق به فرد خاصی نیستند و از ترکیب نمونه‌های مختلف به دست می‌آیند. برای مثال ژنوم مرجع^۳ که در سال ۲۰۰۷ ارائه شده است، از ۱۳ داوطلب از منطقه‌ای در آمریکا جمع‌آوری شده است. این نسخه جدید ارائه شده نسبت به نسخه قبلی که در سال ۲۰۰۳ ارائه شده بود پیشرفت‌های چشمگیری داشت به‌طوری که، ژنوم مرجعی که در سال ۲۰۰۳ ارائه شد، شامل حدوداً ۱۵۰۰۰۰ منطقه بدون توالی بود که این میزان به ۲۵۰ منطقه در ژنوم مرجع سال ۲۰۰۷ کاهش یافت.^[۶۱، ۶۲]

در تعیین توالی ژنوم دو اصطلاح به چشم می‌خورد: کانتیگ^۴ و داربست^۵. در واقع ژنوم از سرهمندی کانتیگ‌ها و داربست‌ها به وجود می‌آید. کانتیگ‌ها توالی‌های اجتماعی ممتد هستند که از مجموعه‌های از خوانش‌های مشترک ساخته شده‌اند. در حالی که داربست‌ها مجموعه‌هایی مرتب و جهت‌دار هستند که به وسیله جفت‌های مکمل^۶، کانتیگ‌ها را به هم متصل می‌کنند. دلیل وجود کانتیگ و داربست به این خاطر است که برای تعیین توالی کل ژنوم انسانی از روش پایین به بالای توالی یابی دی.ان.آ. استفاده می‌شود. در این روش، دی.ان.آ. به قطعات کوچکی شکسته می‌شود. این قطعات کوچک تعیین توالی شده و مجدداً به هم متصل می‌شوند و کانتیگ‌ها و در نهایت ژنوم را می‌سازند. تکنولوژی‌های کنونی اجازه توالی یابی بین ۳۰۰ تا ۱۰۰۰ جفت باز را می‌دهند، پس به ناچار

^۱<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc>

^۲Patches

^۳Contig

^۴Scaffold

^۵Mate pair

مجبور به خرد کردن دی.ان.آ هستند. میزان موفقیت در سرهم کردن کانتیگ‌ها به میزان خوانش‌های مشترک بین آن‌ها بستگی دارد و فاصله بین آن‌ها هم با داربست‌ها پر می‌شود. از آنجایی که برش‌ها بر روی توالی دی.ان.آ به طور تصادفی و بر روی کپی‌های مختلف دی.ان.آ انجام می‌شوند، پس هر بخش از ژنوم باید بارها دیده شود و خوانش‌های مشترک بین آن‌ها دیده شود^[۳۷]. اخیراً از روش دیگری به نام توالی یابی جفت پایانی^۱ استفاده می‌شود که در این روش هر جز از دو طرف تعیین توالی می‌شود. در این روش اطلاعات اضافی از جهت‌گیری کانتیگ‌ها به دست می‌آید که در یافتن ژنوم با خطای کمتر مؤثر است^[۶۳]. در نهایت برای یافتن توالی‌های داربست‌ها از روش‌های پی‌سی‌آر^۲ (برای توالی‌های با طول کم) و از شبیه‌سازی بی‌ای‌سی^۳ (برای توالی‌های با طول زیاد) استفاده می‌شود^[۶۴].

بانک اطلاعاتی انسembل^۴ در راستای اجرای پروژه ژنوم انسانی راهاندازی شده است و یک منبع متمرکز از ژنوم گونه‌های مختلف از جمله انسان، برخی مهره‌داران است. این بانک اطلاعاتی یکی از منابع ژنوم شناخته شده است که اطلاعات ژنومی مناسبی در اختیار محققین قرار می‌دهد^[۶۴].

۳.۸.۲ بانک اطلاعاتی کت

بانک اطلاعاتی کلاس‌بندی ساختار پروتئین کت^۵ شامل اطلاعات مربوط به یک روش شناسایی ساختار محور دومین‌های پروتئینی است. این روش نیمه اتوماتیک است. از روش‌های آزمایشگاهی برای تعیین ساختار دومین‌های کت استفاده می‌شود. در این کلاس‌بندی دومین‌های پروتئینی، از چهار مرحله مختلف استفاده می‌شود. مرحله اول کلاس نام دارد و شامل ساختار دوم دومین پروتئینی است. مرحله دوم، معماری است و شامل شباهت ساختاری است. مرحله سوم توپولوژی دومین است و شباهت‌های ساختاری خاص را شامل می‌شود. ممکن است دو دومین معماری یکسان اما ساختار توپولوژی متفاوتی داشته باشند و در نهایت مرحله چهارم، رابطه تکاملی اثبات شده را نشان می‌دهد که بر اساس شباهت ساختار و عملکرد دومین‌ها تعیین می‌شود^[۶۵، ۶۶]. برای مطالعه پروتئین‌ها و یافتن دومین‌های آن‌ها در آزمایشگاه، محدودیت وجود دارد. این محدودیت‌ها گاهاً به خاطر طول زیاد پروتئین‌ها و یا حل نبودن رشته پروتئینی در محلول‌های مورد نظر و یا وجود بخش‌های بسیار منعطف در رشته‌ها می‌باشد که مطالعه را بسیار دشوار یا غیر ممکن می‌سازد. بنابراین محققان مجبورند بخش‌هایی از رشته پروتئینی را جدا کنند، یا توالی مشخصی را به ابتدای انتهای آن‌ها اضافه کرده و به رشته مناسب جهت مطالعه دست یابند. می‌توان

^۱pairedend sequencing

^۲Polymerase chain reaction(PCR)

^۳Bacterial artificial chromosome(BAC)

^۴Ensembl

^۵CATH Protein Structure Classification

گفت اگر کل رشته پروتئینی و بدون هیچ کم و کاستی مورد مطالعه قرار بگیرد شاید دومین‌های بیشتر و یا حتی متفاوت‌تری نسبت به این مطالعات به دست آید اما فعلاً و با توجه به محدودیت‌های آزمایشگاهی‌باید به این یافته‌ها بسند کرد [۶۷، ۶۸]. برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۱ مراجعه کرد.

۴.۸.۲ بانک اطلاعاتی اسکاپ

بانک اطلاعاتی اسکاپ^۲ هم مانند بانک اطلاعاتی کت یک نوع کلاس‌بندی دومین‌هایی پروتئینی است که به طور نیمه اتوماتیک انجام می‌شود. این بانک اطلاعاتی از سال ۱۹۹۴ میلادی راهاندازی شده است. دومین‌ها در اسکاپ براساس تاخورده‌گی طبقه‌بندی می‌شوند. هفت طبقه مختلف در اسکاپ برای دسته‌بندی دومین‌ها در نظر گرفته شده است. این هفت طبقه به این صورت است: کلاس (شامل نوع تاخورده‌گی)، تاخورده‌گی (شامل شکل‌های مختلف دومین‌های داخل یک کلاس)، فوق خانواده (شامل دومین‌های داخل یک کلاس که حداقل یک جد دور یکسان داشته باشند)، خانواده (دومین‌های داخل یک فوق خانواده که جدهای نزدیک‌تر مشابه داشته باشند)، پروتئین (دومین‌های یک خانواده که شامل پروتئین‌های ضرورتاً یکسان هستند)، گونه (شامل دومین‌های پروتئینی که به گونه مشابه تعلق دارند) و دومین (شامل بخش یا کل یک پروتئین است). برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۳ می‌توان مراجعه کرد.

۵.۸.۲ بانک اطلاعاتی پی‌فم

بانک اطلاعاتی خانواده پروتئینی پی‌فم^۴ با همترازی و مدل مخفی مارکوف بر اساس دومین‌های کت ایجاد شده است [۷۰]. این بانک اطلاعاتی بر اساس توالی پروتئین به تعیین ساختار دومین‌ها می‌پردازد. برای هر دومین تعیین شده توسط روش پی‌فم دو امتیاز در نظر گرفته شده است. از امتیازها برای بیان رابطه زیستی یافته‌ها استفاده می‌شود. اولین امتیازی که استفاده می‌شود ارزش مورد انتظار^۵ است. این امتیاز تعداد همترازی‌های متفاوتی را نشان می‌دهد که احتمال رخ دادن آن باید از امتیاز خاصی کمتر باشد. ارزش مورد انتظار اگر برابر یک باشد یعنی نتیجه کاملاً اتفاقی به دست آمده است. هرچه مقدار ارزش مورد انتظار کم باشد یعنی نتیجه وابستگی کمتری با شانس دارد. این ارزش با اندازه بانک اطلاعاتی مورد مطالعه وابستگی دارد به همین خاطر از امتیاز دوم به نام امتیاز بیت^۶ استفاده می‌شود. اگر نتیجه شباهت کاملاً اتفاقی باشد امتیاز بیت برابر صفر خواهد بود. اگر امتیاز یک

^۱<http://www.cathdb.info/>

^۲SCOP

^۳<http://scop2.mrc-lmb.cam.ac.uk/front.html>

^۴Pfam

^۵E-Value

^۶Bit Score

باشد یعنی احتمال تولید نتیجه توسط مدل دوباره احتمال تولید نتیجه به طور شناسی خواهد بود. در حالت کلی اگر امتیاز بیت عدد n باشد یعنی احتمال تولید نتیجه توسط مدل 2^n بار بیشتر از احتمال تولید نتیجه به طور شناس خواهد بود [۷۱]. برای هرخانواده در پی فم می‌توان کارهای زیر را بر روی توالی‌ها انجام داد [۷۲].

- توالی یابی چندگانه برای توالی‌های مورد نظر
- دیدن ساختار دومین پروتئین
- بررسی توزیع گونه
- موجود بودن لینک ارتباطی به سایر سایتها برای گرفتن اطلاعات بیشتر
- مشاهده ساختار پروتئین

در پی فم اطلاعاتی درمورد دومین‌های پروتئین و خانواده‌های پروتئینی وجود دارد. برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۱ مراجعه کرد.

۶.۸.۲ بانک اطلاعاتی پروتئین

بانک اطلاعاتی پروتئین^۲، یک بانک اطلاعاتی از ساختارهای سه بعدی از مولکول‌های زیستی بزرگ مانند پروتئین و اسیدنوکلئوتیدهاست. بسیاری از ساختارهای پروتئین موجود در این سایت به وسیله بانک‌های اطلاعاتی دیگری همچون کت مورد استفاده قرار می‌گیرد. حدود ۱۰۷۰۰۰ پروتئین در این بانک وجود دارد. علاوه بر ساختارهای سه بعدی اطلاعاتی نظیر برخی اسامی ژن‌ها به همراه محل اگزون‌ها و اینترون‌های ژن‌های کدکننده پروتئین و غیره نیز از این سایت قابل دسترسی است [۷۳]. برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۳ مراجعه کرد.

۷.۸.۲ بانک اطلاعاتی یونیپرات

یونیپرات^۴ یک بانک اطلاعاتی با کیفیت بالا و دسترسی رایگان از توالی‌های پروتئین و اطلاعات عملگری است که از پروژه‌های توالی ژnom مشتق شده است. حدود ۸۰٪ توالی‌های پروتئین موجود در پی فم دارای یک شبیه در بانک اطلاعاتی یونیپرات هم هستند. این سایت معتبر تحت نظر دو مؤسسه بیوانفورماتیک اروپا و مؤسسه بیوانفورماتیک

^۱<http://pfam.xfam.org/>

^۲Protein Data Bank(PDB)

^۳<http://www.rcsb.org/pdb/home/home.do>

^۴Uniprot

سوئیس فعالیت می‌کند و شامل بانک‌های اطلاعاتی مختلفی است [۷۴]. برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۱ مراجعه کرد.

۸.۸.۲ بانک اطلاعاتی یوسیاسی

بانک اطلاعاتی یوسیاسی^۲ زیر نظر مرکز زیست مولکولی دانشگاه کالیفرنیا، سانت کروز^۳ می‌باشد و شامل اطلاعات زیادی در مورد ژن‌ها، اسمی مختلف آن‌ها، تعداد اگزون‌ها، محل آن‌ها، نقاط شروع و پایان مناطق کدکننده پروتئین‌ها است. این بانک اطلاعاتی سعی کرده اطلاعات جامعی در مورد ژن‌ها از بانک‌های اطلاعاتی معتبر جمع کرده و به طور یکجا و در قالب جدول‌های متعددی در اختیار عموم قرار دهد [۷۵]. برای دسترسی به سایت می‌توان به آدرس اینترنتی سایت^۴ مراجعه کرد.

۹.۸.۲ بانک اطلاعاتی هوگو

کمیته نام‌گذاری ژن‌ها، هوگو^۵ با هدف نام‌گذاری استاندارد برای ژن‌ها ایجاد شده است این بانک اطلاعاتی ژن‌ها را با توجه به محصولی که تولید می‌کنند دسته‌بندی کرده و برای بارگیری در اختیار عموم قرار می‌دهد. ژن‌ها به سه دسته تقسیم می‌شوند.

- ژن‌های کدکننده پروتئین^۶ شامل ژن‌هایی که محصول نهایی آن‌ها پروتئین است. این ژن‌ها می‌توانند گاهًا ایزوفرم‌های مختلف یک پروتئین را تولید کنند.
- ژن‌های غیرکدکننده آر.ان.آ.^۷، این دسته از ژن‌ها پروتئین کد نمی‌کنند، اما آر.ان.آهایی را می‌سازند که این آر.ان.آها فعالیت‌های مختلفی را در سلول انجام می‌دهند.
- شبه ژن‌ها^۸، این دسته از ژن‌ها توالی‌هایی از دی.ان.آ هستند که شبیه به ژن‌های کدکننده پروتئین هستند اما هیچ پروتئین عملگری را تولید نمی‌کنند [۷۶].

برای دسترسی به سایت می‌توان به آدرس اینترنتی سایت^۹ مراجعه کرد.

^۱<http://www.uniprot.org/>

^۲UCSC

^۳Center for Biomolecular Science & Engineering ,University of California, Santa Cruz

^۴<https://genome.ucsc.edu/>

^۵HUGO Gene Nomenclature Committee

^۶Protein Coding Genes

^۷non-Coding RNA

^۸pseudogenes

^۹<http://www.genenames.org/>

۱۰.۸.۲ بانک اطلاعاتی رفسک

بانک اطلاعاتی رفسک^۱ یک بانک اطلاعاتی باز است، توضیحات کاملی درباره توالی دی.ان.آ و آر.ان.آ و محصولات پروتئینی تولید شده ژن‌ها در اختیار عموم قرار می‌دهد. علاوه بر ژنوم انسان اطلاعاتی درباره برخی باکتری‌ها و ویروس‌ها هم در این بانک اطلاعاتی یافت می‌شود[۷۷].

۱۱.۸.۲ بانک اطلاعاتی منبع ادغام ساختارها و توالی‌ها

این بانک اطلاعاتی که به سیفتز معروف است، برای نگاشت دادن اطلاعات مربوط به توالی‌های پروتئینی استفاده شده در بانک اطلاعاتی کت و پی‌فم، اسکاپ و یونیپرات به پروتئین‌های انسانی به کار می‌رود. برای شناسایی دومین‌های پروتئینی در آزمایشگاه، معمولاً بخشی از رشته پروتئین انتخاب می‌شود نه همه آن یا گاهی یک توالی خاص به رشته‌ها اضافه و یا کم می‌شود. این رشته‌ها را رشته‌های پی‌دی‌بی می‌نامند. از آنجایی که ساخت رشته پی‌دی‌بی از قانون خاصی پیروی نمی‌کند، پس برای تک تک پروتئین‌ها، نگاشت محل دومین از رشته پی‌دی‌بی به پروتئین اصلی باید به صورت دستی انجام شود. سیفتز^۲[۷۸] یک منبع به روز از نگاشت رشته‌های پی‌دی‌بی به پروتئین‌های است. این منبع هر هفته یکبار به روز می‌شود تا اگر رشته جدیدی مورد مطالعه قرار گرفته است، نتایج آن در دسترس محققین قرار بگیرد. برای اطلاعات بیشتر می‌توان به آدرس اینترنتی سایت^۳ می‌توان مراجعه کرد.

۹.۲ تعریف مسئله

با توجه به تعاریف و مقدمات گفته شده در فصل یک و این فصل مشخص شد که مسئله تشخیص عامل‌های ایجاد یا توسعه سرطان، یک مسئله زیستی است که مقدار زیادی داده دارد. برای کار بر روی این داده‌ها باید از علم آمار و علوم کامپیوتر کمک گرفت. با توجه به شیوع بالای بیماری سرطان و تنوع این بیماری در بافت‌های مختلف بدن و میزان مرگ و میر بالای آن و هزینه‌های مالی زیادی که روند درمان آن به دولتها و اشخاص تحمل می‌کند. زیست‌شناسان و محققان زمان زیادی صرف یافتن علل ایجاد یا توسعه سرطان می‌کنند. یکی از نتایج این تحقیقات این است که جهش‌های رخ داده در ژنوم از عوامل اصلی و کلیدی ایجاد و توسعه سرطان‌ها هستند. با توجه به نقش دومین‌های پروتئینی در سلول، تمرکز زیست‌شناسان بر روی یافتن دومین‌هایی است که جهش‌های رخ داده بر روی این دومین‌ها باعث ایجاد سرطان می‌شود. همین طور گفته شد که حدود ۱۹۰۰۰ ژن در بدن انسان مسئول کدکردن پروتئین‌ها هستند و بر روی هر پروتئین ممکن است یک یا چند دومین وجود داشته باشد.

¹RefSeq

²SIFTS(Structure integration with function, taxonomy and sequence)

³<http://www.ebi.ac.uk/pdbe/docs/sifts/>

با توجه به تعداد زیاد ژن‌ها و دومین‌ها و تنوع الگوی‌های جهشی بر روی ژنوم بیماران و بودن جهش‌های مسافر در ژنوم، و بسیار پرهزینه بودن تحقیقات آزمایشگاهی در این زمینه، از علم آمار و احتمال و علوم کامپیوتر برای یافتن دومین‌های دخیل پیشنهادی در سلطان استفاده می‌شود تا با محدود کردن مطالعات آزمایشگاهی به دومین‌های پیشنهادی، با صرف هزینه بسیار کمتر و در زمان بسیار کمتر به شناسایی دومین‌های واقعی دخیل در ایجاد یا توسعه سلطان‌ها پرداخت و به دنبال درمان‌های موثرتر بود. در این پایان نامه نیز تمرکز بر روی همین مسئله است و سعی می‌شود با نگاشت جهش‌های دخیل در سلطان بر روی دی.ان.آ و دومین‌های پروتئینی، دومین‌های دخیل در ایجاد یا توسعه سلطان‌ها را شناسایی کرد.

۱۰.۲ جمع‌بندی

در این بخش در مورد سلطان، دلایل ایجاد سلطان در بدن انسان پرداخته شد. همین انواع بانک‌های اطلاعاتی مورد استفاده شرح داده شد و در ادامه به معرفی مطالعات انجام شده در زمینه سلطان و ارتباط آن با میتوکندری، سلول‌های بنیادی، ژن‌ها ترمیم کننده و مطالعه یکپارچه سلطان پرداخته شد. در فصل سوم، به بیان مفصل مطالعه انجام شده در زمینه نقش دومین‌های پروتئینی در سلطان و بحث بر روی مطالعه انجام شده در مورد اعتبار مطالعات مختلف انجام شده بر روی سلطان پرداخته خواهد شد.

فصل ۳

نگاشت دومین‌ها و نقدی بر روش شناسایی

این فصل به شرح کامل مطالعه نرت و همکاران [۷] پرداخته می‌شود که به نگاشت جهش‌های سوماتیک به دومین‌های پروتئینی در سرطان روده بزرگ پرداخته است.

۱.۳ نگاشت دومین‌های پروتئینی جهش‌های سوماتیک بیماران سرطانی

ظهور تعیین توالی کل ژنوم، زمینه‌ای برای مقایسه ژنوم صدھا نمونه سالم و بیمار در یک مطالعه واحد را به وجود آورد. این مطالعه به دنبال شناسایی جهش‌های ژنتیکی که عامل شروع و رشد تومورهای سرطانی هستند، می‌گردد. شناسایی و تمایز جهش‌ها به دو دسته راننده و مسافر، مسئله اصلی در یافتن مکانیسم ایجاد سرطان و درمان آن به شمار می‌رود [۱]. بسیاری از مطالعات از روش‌های ژن محور برای مطالعه در این زمینه استفاده می‌کنند. به عبارت دیگر هدف اصلی آن‌ها مشخص کردن جهش‌های راننده‌ای است که در تعداد زیادی از بیماران مشاهده می‌شود. متاسفانه این گونه روش‌های مطالعه به شناسایی زیرمجموعه کوچکی از ژن‌ها محدود می‌شوند که تعداد محدودی از بیماران را پوشش می‌دهند. به علاوه، مشخص کردن جهش‌هایی که در روی ژن‌های نادر هستند، یعنی در بیماران زیادی دیده نمی‌شوند اما ممکن است رابطه قوی با سرطان‌ها دارند در در روش ژن محور بسیار مشکل است. در واقع، امروزه این موضوع، یک مسئله حیاتی در راه مبارزه با سرطان است. همین‌طور، رویه‌های ژن محور، برای مشخص کردن ژن‌های راننده هیچ تمایزی برای جهش‌ها بر روی نقاط مختلف ژن قائل نیستند که این، بی‌توجهی به اطلاعات مهمی است که مکان جهش‌ها دارند. یک مطالعه که اخیراً میزان بدتفسیری جهش‌ها در روش‌های مطالعه ژن محور را نشان داده است، ادعا کرده است تغییراتی که موجب حذف کامل یک پروتئین می‌شوند اغلب متمایز از جهش‌هایی هستند که در مناطق خاصی از پروتئین رخ می‌دهند. این تغییرات باعث از بین بردن همه تعاملات پروتئین می‌شوند. به طور خلاصه، محققان با نشان دادن مثال‌های مختلف سعی در نشان دادن اهمیت جایگاه جهش بر روی پروتئین هستند و این که مطالعات ژنی تصویر ناواضح و ناکاملی از

اهمیت عملکرد جهش ارائه می‌دهند. برای مثال وجود جهش بر روی جایگاه‌های مختلف یک دومین باعث بروز فنوتیپ‌های متفاوت یک بیماری خاص می‌شود. این مطالعات باعث به وجود آمدن روش دومین محور شده است. در رویکرد ژن محور، جهش‌های راننده، که به طور برجسته در تعداد زیادی از نمونه‌های مورد مطالعه وجود داشته باشند، شناسایی می‌گردد. اما این رویکرد به موقعیت جهش‌ها بر روی ژن‌ها و زمینه کاربردی برخی مکان‌های خاص ژن‌ها نمی‌پردازد. در اینجا روش جدید دومین محور که توسط نرت و همکاران^[۷] برای نگاشت کردن جهش‌ها بر روی دومین‌های پروتئینی ارائه شده است، معرفی می‌شود که می‌تواند به طور عملی‌تر عوامل مؤثر در ایجاد سرطان را معرفی کرده و با جمع آوری جهش از تمام ژن‌های حاوی یک دومین پروتئین خاص، به شناسایی جهش‌هایی که در سطح ژن کم هستند اما در دومین خاصی به طور فراوان دیده می‌شوند، بپردازد و در نهایت، دومین‌های پروتئینی با تعداد جهش‌های بالا را به عنوان عامل ایجاد و توسعه سرطان معرفی کند.

۱.۱.۳ معرفی روش

در مطالعه نرت و همکاران^[۷]، جهش‌های سوماتیک کل ژنوم ۱۰۰ بیمار که همگی سرطان روده^۱ داشتند، به ژن‌ها و دومین‌های پروتئینی نگاشت داده شده و یک نقشه توپوگرافی از آن‌ها تهیه شده است. ژن‌های شناخته شده دخیل در سرطان مانند ATP, TP53, KRAS در این مطالعه نیز شامل جهش‌های بالایی هستند. به علاوه، دومین‌هایی با تعداد جهش‌های زیادی شناسایی شدند که بر روی پروتئین‌های کدشده از این ژن‌ها قرار دارند، در حالی که بقیه دومین‌های این ژن‌ها تعداد خیلی کمی جهش را به خود اختصاص می‌دهند. در یک مطالعه که توسط اسکوبلوم و همکاران^[۷] بر روی سرطان روده بزرگ انجام شده است، دو دومین MH1,MH2 به عنوان دومین‌های موثر در این بیماری نام برده شده‌اند. اگرچه بعداً در مطالعه دیگری، چیتندن و همکاران^[۷] ذکر کردند، ۷۰٪ جهش‌های پروتئین‌های حاوی دومین MH2 بر روی دومین MH1 قرار دارند، اما در مرور دومین MH1 خلاف این مطلب صادق است، یعنی در پروتئین‌های حاوی دومین MH1 هیچ کدام از جهش‌ها بر روی این دومین قرار ندارند. در مطالعات مشابه دیگری^[۷]، دومین‌های دیگری نیز شناسایی شده‌اند که به طور معناداری با سرطان‌ها رابطه دارند، مانند دومین‌های پروتئین‌های کیناز، دومین‌های دخیل در نگهداری و ترمیم دی.ان.آ و پروتئین‌های تنظیم ترجمه. باید توجه داشت گاه‌آ، برخی مطالعات با هدف مطالعه دومین‌های خاص برخی ژن‌های خاص انجام می‌شوند که این امر باعث می‌شود نتایج حاصل از این مطالعات تحت تاثیر این نوع انتخاب باشند و باعث معرفی نادرست برخی دومین‌ها به عنوان دومین‌های مهم در ایجاد بیماری‌ها شوند.

در مطالعه نرت و همکاران^[۷]، از جهش‌های اگزومی، شامل جهش‌های تک نوکلئوتیدی، اضافه و حذف شدن‌های کوتاه ۱۰۰ بیمار مبتلا به سرطان روده بزرگ که از پروژه اطلس ژنوم سرطان^[۳۱] به دست آمده‌اند،

^۱colon adenocarcinoma

جدول ۱.۳: تعداد جهش‌های سرطان روده^[۷]

Mutation counts for colon cancer	
Total patients	100
Total mutations	21572
Total nonsynonymous SNVs	17,174 (79.6%)
Total frameshift insertions	2,527 (11.7%)
Total nonframeshift insertions	239 (1.1%)
Total frameshift deletions	5 (0.0%)
Total nonframeshift deletions	0 (0.0%)
Total stop-loss SNVs	33 (0.2%)
Total stop-gain SNVs	1,594 (7.4%)
Mutations in domain regions	10,647 (49.4%)
Average mutations per patient	216 (+ 552)
Number of mutations per patient	214880

استفاده و روش جدیدی برای تفسیر جهش‌های سوماتیک در روش مطالعه دومین محور ارائه شده است. به طور کلی انتظار می‌رود جهش در ژن‌های کدکننده پروتئین مانند جهش تک نوکلئوتیدی که باعث تغییرات اسید آمینه^۱، اضافه یا حذف شدن‌های کوتاه که باعث تغییر در قاب خواندن و جهش به دست آوردن یا از دست دادن یک کدون توقف، منجر به توسعه سرطان شوند. بنابراین، تمام انواع جهش‌های سوماتیک ذکر شده در بالا به علاوه جهش حذف و اضافه خارج از قالب، در ژن‌های کدکننده پروتئین نمونه‌های سرطانی برای مطالعه انتخاب شده اند. در مجموع ۲۱۵۷۲ جهش در ۱۰۰ نمونه وجود دارد. به طور متوسط ۲۱۵ جهش برای هر بیمار وجود دارد. حدود ۸۰٪ این جهش‌ها از نوع جهش‌های غیر خاموش، ۱۲٪ جهش تغییر قاب، ۷٪ جهش افزایش توقف، ۰٪ جهش از دست دادن توقف و یک درصد جهش اضافه شدن اسیدنوکلئوتید هستند. جدول ۱.۳ تعداد جهش‌های هر نوع جهش را نشان می‌دهد. علاوه بر این، تقریباً نیمی (۴۹٪/۴۹٪) از همه جهش‌ها داخل مناطق دومین پروتئین رخ داده است.

این روش یک نقشه جهش ژنی تولید می‌کند. "نقشه جهش ژنی"^۲ یک نقشه توپوگرافی است که در آن به هر ژن، کوه گفته می‌شود و تعداد جهش‌های هر ژن هم ارتفاع کوه را نشان می‌دهد. به کوههای با ارتفاع بیشتر قله و کوههای با ارتفاع کم نیز تپه گفته می‌شود. این روش برای نشان دادن فرکانس جهش ژن‌های مختلف کل ژنوم استفاده می‌شود. "قله‌های ژنی"^۳ این نقشه توپوگرافی، ژن‌هایی با فرکانس جهش بالا در یک بیماری خاص و یا مجموعه بیماری را نشان می‌دهند. در این مطالعه، علاوه بر نقشه جهش ژنی، نقشه جهش دومینی سرطان روده بزرگ هم به همین ترتیب رسم شده است با این تفاوت که از دومین به جای ژن استفاده شده است. با مقایسه این دو نقشه نشان داده شده است خصوصیت‌های مهمی توسط رسم نقشه دومینی فاش می‌شود که در روش نقشه

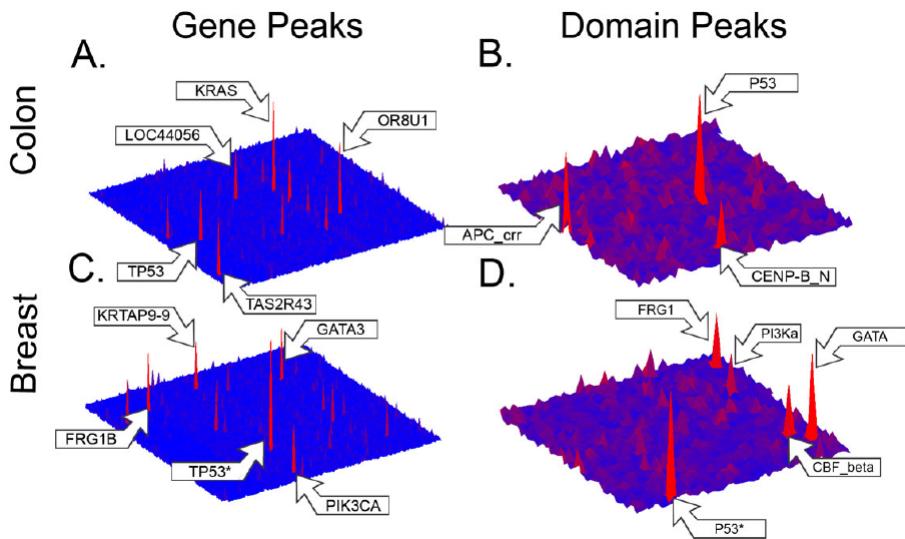
^۱SNVs nonsynonymous (nsSNVs)^۲"Domain Mutational Landscape"^۳Gene Peaks

ژنی قابل مشاهده نیست. زیرا دومین‌های با تعداد جهش‌های بالا در روش ژنی قابل شناسایی نیستند. به علاوه، نشان داده شده است که چطور تمرکز بر روی جهش‌های ژن باعث بد شناسایی ژن‌های دخیل در سرطان می‌شود در حالی که، ژن‌هایی که دومین‌های با جهش‌های بالا دارند بیشتر با سرطان ارتباط دارند. بنابراین، کشیدن نقشه جهش دومینی باعث شناسایی جهش‌های سوماتیک راننده می‌شود. این روش، ابزار جدیدی برای تجزیه و تحلیل مطالعات مقیاس بزرگ سرطان معرفی می‌کند.

برای مقایسه با انواع دیگر سرطان، نقشه جهش دومین و ژنی ۵۲۲ بیماران سرطان پستان مهاجم نیز رسم شده است. با وجود تعداد کم جهش‌ها (۲۵۸۰۷ جهش)، بیماران مبتلا به سرطان پستان به طور متوسط تعداد کمتری جهش (۴۹.۷ جهش) به ازای هر نفر داشتند، اما توزیع انواع جهش و تعداد جهش داخل مناطق دومین، تقریباً شبیه به مجموعه سرطان روده بزرگ بود. در این مطالعه، در دو نوع سرطان پستان و روده بزرگ، با استفاده از علم آمار و روش نرمال کردن داده‌ها سعی شد ژن‌ها و دومین‌های معنادار معرفی شود. ۱۵۴ ژن معنادار برای سرطان روده بزرگ، ۱۵۱ ژن معنادار برای سرطان پستان، ۴۵ دومین معنادار برای سرطان روده بزرگ و ۴۱ دومین معنادار هم برای سرطان پستان معرفی کردند و بیان کردند که تعداد کمی ژن مشترک با فرکانس جهش بالا، به اصطلاح "قله" مشترک در نقشه ژنومی دو سرطان مانند APC, KRAS, TP53 وجود دارد. اسامی دومین‌های معنادار این دو سرطان در جدول ۲.۳ آمده است. علاوه بر مقایسه نتایج سرطان روده بزرگ و پستان، نرت و همکاران [۷]، نتایج خود را با نتایج دو مطالعه دیگر [۷] که مطالعه کل ژنوم بودند نیز مقایسه کردند. یکی از مطالعات ۸۹ ژن کاندید و دیگری ۱۹۱ ژن کاندید معرفی کردند. تنها شش ژن مشترک بین ژن‌های کاندید این مطالعات و ژن‌های کاندید مطالعه نرت و همکاران [۷] وجود داشت.

جدول ۲.۳: اسمى دومين‌های معنادار سرطان‌های روده بزرگ و پستان [۷]

Top domains highly mutated in breast cancer tumor genomes	Top domains highly mutated in colon cancer tumor genomes
P53	P53
APC_crr	GATA
CENP-B_N	FRG1
Tep10_C	CBF_beta
TAS2R	PI3Ka
WAP	BID
Rad52_Rad22	PI3K_p85B
DUF3591	Fork_head
MH2	S-antigen
zf-DHHC	Asp-B-Hydro_N
Dpy-30	SRP-alpha_N
bZIP_2	Tis11B_N
PI3K_p85B	zf-CCCH
RunxI	LSM
zf-C2H2_jaz	Acetyltransf_1
CK_II_beta	ELM2
Mesd	bZIP_2
Fapy_DNA_glyco	RNA_pol_L
Tmemb_55A	S_100
Fox-1_C	LIF_OSM
MutS_IV	IL8
S_100	Prion
BRICHOS	Lyase_1
CTP_transf_1	Cyt-b5
14-3-3	INSIG
FAA_hydrolase	Pou
IL8	Ets
5-nucleotidase	HD
bZIP_1	CBFD_NFYB_HMF
Ribosomal_L7Ae	bZIP_1
Tctex-1	HNF-1B_C
LSM	PTEN_C2
OGFr_N	RPEL
APC_basic	T-box
UNC45-central	Ski_Sno
Casc1	Complex1_49kDa
ARS2	UPF0041
PIH1	Elongin_A
PKD	Keratin_B2
CTP_transf_3	RNase_P_Rpp14
Armet	Med15
Acyl_CoA_thio	
TNFR_c6	
Miro	
KOW	



شکل ۱.۳: نقشه جهش‌های دومینی و ژنی برای سرطان‌های روده بزرگ و پستان [۷]

۲.۱.۳ نقشه جهش ژنی و دومینی

برای تهیه نقشه جهش ژنی، فرکانس جهش سرطان روده بزرگ و پستان، برای تک تک ژن‌ها بر روی یک نقشه دو بعدی که در آن هر ژن به عنوان یک مربع با اندازه دلخواه در یک شبکه با مختصات دکارتی نگاشت شده است. نقشه جهش دومین هم به همین شکل تهیه شده است. شکل ۱.۳ نقشه جهش ژنی و دومینی دو سرطان پستان و سرطان روده بزرگ را نشان می‌دهد. شکل A1.۳ و شکل B1.۳ نقشه جهش ژنی دو سرطان روده بزرگ و پستان را نشان می‌دهند. نقشه‌های ژنی و دومینی موجود بر روی هر کدام با ذکر اسامی مشخص شده است.

ارتفاع قله (محور Z) روی نقشه ژنی متناسب با فراوانی جهش سوماتیک هر ژن تقسیم بر اندازه طول پروتئین کدشده آن ژن (به عنوان مثال طولانی‌ترین ایزوفرم پروتئین) است. این نقشه چشم انداز کلی از ژن‌های جهش یافته در بیماران مبتلا به سرطان روده بزرگ را نشان می‌دهد. در مقایسه با مطالعه قبلی، این روش، شامل تعداد زیادی قله از ژن‌های جهش یافته است؛ از جمله دو ژن KRAS و TP53. فرکانس جهش قابل توجهی هم در ژن APC پیدا شده است. جدول ۳.۳ اسامی ژن‌های منتخب سرطان روده بزرگ با تعداد بخش‌های بالا در این مطالعه را نشان می‌دهد. لازم به ذکر است که در نقشه جهش ژنی، به طور کلی تعداد بسیار زیادی ژن با فرکانس جهش پایین دیده می‌شود. در این مطالعه از نرخ کشف کاذب محلی^۱ برای شناسایی ژن‌ها و دومین‌ها با فرکانس جهش قابل توجه، استفاده شده است. انتظار می‌رود با فرض این‌که جهش‌های مسافر با توزیع یکنواخت در سراسر ژنوم توزیع شده‌اند، جهش‌های راننده با این روش قابل شناسایی باشند. تعداد جهش هر دومین نسبت به طول دامنه تجمعی، نرمال شده است. این باعث می‌شود که نقشه دومین به سمت دومین‌هایی که جهش‌های زیادی در

¹Local False Discovery

جدول ۳.۳: ژن‌های انتخاب شده با تعداد جهش‌های بالا در سرطان روده بزرگ [۷]

Gene	Protein Accession	Mutations	Protein Length
APC	NP_001120982	76	2843
BRAF	NP_004324	14	766
FBXW7	NP_361014	21	745
KRAS	NP_203524	30	189
LOC440563	NP_001130033	33	293
NRAS	NP_002515	7	189
OR8U1	NP_001005204	48	309
TAS2R19	NP_795369	22	299
TAS2R30	NP_001091112	28	319
TAS2R31	NP_795366	31	309
TAS2R43	NP_795365	36	309
TP53	NP_000537	31	393

آن‌ها اتفاق افتاده است، سوق پیدا نکند. با استفاده از یک حد آستانه ^۱/۰، برای نرخ کشف کاذب محلی، ۱۵۴ ژن، برای سرطان روده بزرگ و ۱۵۱ ژن برای سرطان پستان به عنوان ژن‌های کاندید اعلام شده‌اند. ژن‌های مهمی که از روش نرمال کردن برای سرطان روده بزرگ در قله قرار گرفته‌اند، شامل ژن‌های KRAS, OR8U1 و TAS2R43 است. علاوه بر این، فرکانس جهش قابل توجهی در شش ژن یافته شده است که پیش‌تر از این نیز در مطالعات قبلی شناسایی شده بودند: APC, KRAS, TP53, FBXW7, SMAD4 و GRID1. از شش ژن با بیشترین فرکانس جهش در میان ژن‌های منتخب، چهار عدد در بین ۲۰ ژن با فرکانس جهش بالا در نتایج مطالعه قبلی هم وجود داشتند. ژن PIK3CA تنها ژنی است که اهمیت فرکانس جهش معناداری را در این مطالعه کسب نکرده است. با این حال، فرکانس جهش دومین PI3K_p85B واقع در ژن PIK3CA معنادار است. دو ژن منتخب دیگر SMAD2 و SMAD3 فرکانس جهش بالای ندرند، اما فراوانی جهش قابل توجهی در دومین MH2 موجود در هر دو ژن مشاهده شده است.

شكل ۳.۱ و شکل D1.۳ به نقشه جهش دومین دو سرطان یاد شده اشاره دارند. در اینجا تعداد جهش برای هر دومین نسبت به طول تجمعی آن دومین در ژن‌های جهش یافته شناسایی شده در ژنوم، نرمال شده است و با استفاده از یک حد آستانه ^۱/۰ منجر به شناسایی ۴۵ دومین با فرکانس جهش قابل توجه در مجموعه سرطان روده بزرگ و ۴۱ دومین در مجموعه سرطان پستان شده است. در نقشه جهش دومینی برای سرطان روده بزرگ تنها چند قله مربوط به دومین‌های CENP-B_N P53, CRR, APC وجود دارد که در مقابل آن تعداد زیادی تپه دومین وجود دارد.

ساختن نقشه جهش ژنی و دومینی امکان شناسایی تعداد زیادی از قله‌های دومین که به طور قابل توجهی

^۱threshold

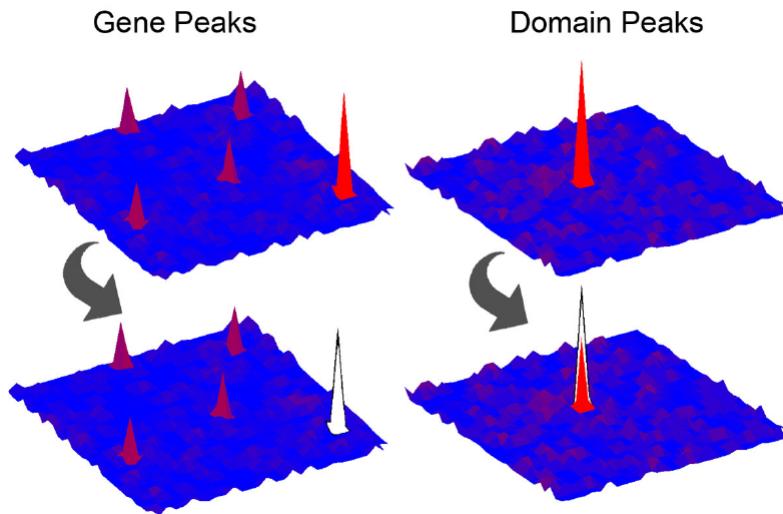
جدول ۴.۳: دومین‌های پروتئینی انتخاب شده با تعداد جهش بالا [۷]

Name	Genes (Number of Mutations)
APC_basic	APC(5)
APC_crr	APC(8)
CortBP2	CTTNBP2NL (1), CTTNB2P (2), FILIP1 (5), FILIP1L (1)
MH2	SMAD4 (9), SMAD9 (2), SMAD2 (1), SMAD3 (1), GARS (1)
Miro	KRAS (28), NRAS (7), RAB27B (2), RAB11B (2), RABL3 (2) + 46 genes with 36 additional mutations
MutS_IV	MSH4 (2), MSH6 (2), MSH5 (1), MSH2 (1)
P13K_P85B	PIK3CA (4)
P53	TP53 (27), TP63(1)
TAS2R	TAS2R43 (36), TAS2R31(31), TAS2R30 (25), TAS2R19 (21) + 19 genes with 27 additional mutations
WAP	WFDC8 (3), SLP1 (1), WFDC5 (1), KAL1 (1)

جهش یافته‌اند را می‌دهد. برخی از قله‌های دومین، بیشتر یا همه جهش خود را بر روی ژن‌هایی دارند که در روش ژن محور نیز بیشترین فرکانس‌های جهش را داشتنند. جدول ۴.۳ دومین‌های پروتئینی انتخاب شده با مقدار جهش بالا را نشان می‌دهد. به عنوان مثال، در مجموعه سرطان روده بزرگ دو دومین، P53 و APC_crr تقریباً تمام جهش‌های خود را از ژن‌های TP53 و APC، که هر دو به عنوان قله ژن در نقشه ژنی ظاهر شده‌اند، دریافت کرده‌اند. با این حال، برخی از قله‌های دومین که فرکانس جهش بالایی دارند بر روی ژن‌هایی قرار دارند که تک تک ژن‌ها به صورت جداگانه حاوی تعداد قابل توجهی جهش نیستند. برای مثال، دومین CENP-B_N شان داده شده در نقشه جهش دومینی، دارای جهش‌های جمع آوری شده از ژن‌هایی است که هیچ کدام از ژن‌های TIGD7 و JRKL فرکانس جهش قابل توجهی در نقشه ژن محور ندارند. در مقایسه نقشه ژنی و دومینی باید گفت که امکان شناسایی تعدادی از دومین‌هایی که در آن جهش‌های با فرکانس بالا حتی بعد از حذف ژن‌های با فرکانس جهش بالای شامل آن دومین‌ها، هم وجود دارد. یعنی دومین‌هایی هستند که جهش‌های زیادی دارند که بخشی از این جهش‌ها مربوط به ژن‌هایی است که این ژن‌ها در مطالعه ژنی هم جهش‌های معنادار زیادی دارند. اما حتی با حذف چنین ژن‌هایی دومین‌های یادشده همچنان فرکانس جهش بالایی دارند. در شکل ۲.۳ مثالی از چنین دومینی نشان داده شده است.

۳.۱.۳ مقایسه نقشه سرطانی، سرطان‌های پستان و روده بزرگ

در این مطالعه ژن‌ها و دومین‌های مشترکی بین دو سرطان روده بزرگ و سرطان پستان پیدا شده است که در هر دو سرطان، فرکانس جهش بالایی دارند. این ژن‌ها، TP53, KRAS, CELA1, SERTAD3, HIST1H1C, BCL2L11 و DCAF4L2 هستند که در هر دو سرطان در نقشه جهش ژنی جز مرتفع‌ترین قلل هستند. به

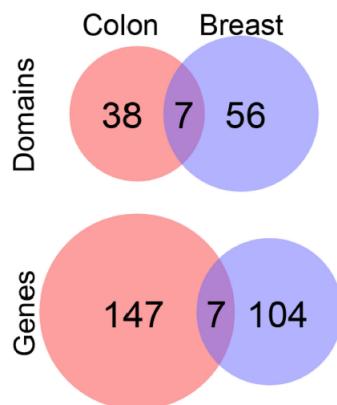


شکل ۲.۳: قلهای دومینی باقیمانده بعد از حذف جهش‌های مربوط به ژن‌های قله [۷]

علاوه دومین‌های S_100, P53, PI3K_p85B, bZIP_1, bZIP_2, IL8, LSM علاوه دومین‌های جز مرتفع‌ترین قلل هستند. نمودار ونی رسم شده در شکل ۳.۳ تعداد قلل ژنی و دومینی برای هر دو نوع سرطان و مشترکات آن‌ها را نشان می‌دهد. تعداد دومین مشترک که فرکانس جهش بالایی در هر دو بیماری دارند بیشتر از تعداد ژن‌های مشترک با فرکانس جهش بالا است. به عنوان مثال، در حالی که ۴/۵٪ (۷ نفر از ۱۵۴) از ژن‌های با قلل مرتفع سرطان روده بزرگ در افراد مبتلا به سرطان پستان نیز مشاهده شده است، حدود ۱۵٪ (۷ نفر از ۴۵) از دومین‌های با قلل مرتفع دومین سرطان روده بزرگ در افراد مبتلا به سرطان پستان نیز مشاهده شده است. به علاوه، مشخص شده است که از بین دومین‌های با قلل مرتفع در نقشه جهش دومینی فقط یک دومین P53 به ژن TP53 متعلق است.

تجزیه و تحلیل نتایج حاصل از نقشه ژنی باعث بازشناسایی مجدد ژن‌های دخیل در سرطان شده است، این نشان می‌دهد که روش این مطالعه در شناسایی ژن‌های مهم در سرطان خوب عمل می‌کند. هدف اصلی این مطالعه، تهیه نقشه جهش‌های دومینی سرطان روده بزرگ و مقایسه آن با نقشه جهش‌های ژنی آن است. در کل، ۴۵ دومین با تعداد جهش بالا در نمونه‌های سرطان روده بزرگ شناسایی شده است، که سه دومین, P53, CENP-B_N DNA-binding APC_crr, CENP-B_N APC_crr در صدر لیست قرار دارند. دومین N که یک دومین است، شامل جهش‌هایی است که متعلق به ژن TIGD7 و JRKL است، که نقش شناخته شده‌ای در سرطان ندارند. دومین‌های P53 و APC_crr که جز دومین‌های پرجهش هستند، مربوط به ژن‌های TP53 و APC می‌باشند که جز ژن‌های شناخته شده در سرطان هستند. با این حال، نگاشت دادن جهش‌ها به دومین‌ها، روش با ارزشی برای شناخت جهش‌های مهم در سرطان را به ارمغان می‌آورد.

تهیه و مطالعه نقشه دومینی اهمیت توجه به پروتئین‌ها را در ایجاد و توسعه سرطان‌ها مشخص کرده است.



شکل ۳.۳: قله‌های ژنی و دومینی مشترک بین سرطان‌های روده بزرگ و پستان [۷]

اما سوال این است که آیا تمام نقش رشد تومورها به عهده پروتئین‌های است یا عوامل دیگری نیز دخیل هستند؟ برای مثال دومین P53 یک دومین DNA-binding است که بیشتر از ۹۰٪ جهش‌های ژن TP53 را شامل می‌شود. این در حالی است که این دومین فقط نصف طول پروتئین را دربر می‌گیرد (۱۹۳ اسید‌آمینه از ۳۹۳ اسید‌آمینه کل پروتئین). ۲۷ جهش از ۳۱ جهش پروتئین TP53 بر روی دومین P53 قرار دارد. وجود جهش بر روی این دومین می‌تواند باعث کاهش کارآیی DNA binding، بد تا خوردن پروتئین، عدم ثبات پروتئین و عوامل دیگر شود. ژن APC شامل هفت تکرار از دومین APC_crr است به علاوه ۱۳ دومین متمایز دیگر. جهش‌های کوتاه‌کننده بر روی این ژن، به طور عمده در مناطقی از پروتئین که دومین یا سومین تکرار از دومین APC_crr هستند رخ می‌دهد. علی‌رغم نرسیدن به اطلاعات مهمی در سطح ژن در مجموعه سرطان روده بزرگ، ژن PIK3CA در مجموعه سرطان پستان، در صدر پنج ژن با بیشترین جهش‌ها بعد از نرمال کردن نسبت به طول ژن قرار گرفته است. ژن PI3KCA واسطه ارسال سیگنال برای پردازش‌هایی مانند رشد و بقای سلول است و در چندین نوع سرطان مختلف با فرکانس جهش بالا دیده می‌شود. این ژن شامل پنج دومین است، از مطالعه نقشه جهش دومینی این ژن مشخص شده است، دومین PI3K_p85B که مسئول اتصال PI3K_p85 که برای شکل‌دهی هترودایتمر^۱ است، جهش‌های معناداری در هر دو سرطان دارد. میزان پی-مقدار آن در سرطان روده بزرگ بالا (آستانه ۴٪) بود. برای دو دومین PI3K_C2 و PI3K_rdb فرکانس جهش بالا و یا میزان پی-مقدار بالا مشاهده نشده است و برای دو دومین آخر، PI3_Ka و PI4 که شامل جهش‌های بی‌معنی در سرطان‌های مختلفی از جمله سرطان روده بزرگ و پستان هستند، دومین PI3_Ka فرکانس جهش و میزان پی-مقدار بالایی در سرطان پستان دارد اما در مورد سرطان روده بزرگ این چنین نیست. همین‌طور جهش‌های کمی بر روی دومین PI3_PI4 وجود دارد.

روی هم رفته، این مثال‌ها مزايا و معایب روش مطالعه دومین را نشان می‌دهند. در حالی که روش سنتی،

^۱Heterodimer

مطالعه ژنی محل جهش‌ها بر روی ژن را آشکار نمی‌کند، در روش دومین محور اهمیت نقش برخی دومین‌ها در ایجاد و توسعه سرطان‌ها مشخص می‌شود. همین‌طور دومین‌هایی که فرکانس جهش بالایی دارند در روش ژن محور قابل تشخیص نیستند اما در روش دومین محور شناسایی می‌شوند؛ مانند دومین PI3K_p85B. در روش مطالعه دومین محور تجمع دومین‌ها از ژن‌های مختلف مورد بررسی قرار می‌گیرد، و جهش‌هایی که خارج از دومین رخ می‌دهند حذف می‌شوند، بنابراین، هنوز هم به مطالعه بیشتر برای بسط دادن این روش برای سایر قسمت‌های ژن وجود دارد.

با مقایسه نقشه ژنی و دومینی مشخص شده است که تعداد کمی از دومین‌هایی که فرکانس جهش بالایی دارند (در کل هفت دومین) حتی بعد از حذف ژن‌هایی که فرکانس جهش بالایی دارند، هنوز هم فرکانس جهش بالایی KAL1 و WFDC5، ALPI و WFDC1 را به خود اختصاص می‌دهند. فرکانس جهش‌های دومین WAP بر روی سه ژن WFDC8، ALPI و KAL1 بعد از حذف جهش‌های این دومین بر روی ژن WFDC8، که این ژن فرکانس جهش بالایی داشت هم معنادار بود. ژن WFDC8 ژن شناخته شده مرتبط با سرطان نیست، اما ژن WFDC5 نقش شناخته شده‌ای دارد. و ژن SLPI نیز در توسعه سرطان ریه نقش دارد. به علاوه، جهش‌های ژن KAL1 مسئول سنتروم کالمن^۱ است. به خاطر مشخص بودن ارتباط جهش‌های دومین WAP با بیماری‌ها و سرطان‌ها، به خاطر وجود آن بر روی ژن‌هایی که با سرطان ارتباط دارند، شناسایی نقش ژن WFDC8 در سرطان روده بزرگ نیاز به مطالعات بیشتر دارد. مثال‌هایی که در بالا مطرح شد، در آن‌ها قلل دومین با ارزش به حداقل یک ژن با ارزش تعلق داشت، فقط ۱۴ دومین از ۴۵ دومین از مجموعه دومین‌های با فرکانس جهش بالا را شامل می‌شد. بقیه ۳۱ دومین مربوط به ژن‌هایی بودند که جهش‌های با ارزشی در نقشه جهش ژنی نداشتند به همین دلیل این دومین‌ها از طریق روش مطالعه بر ژن محور سنتی قابل شناسایی نبودند. اما ممکن است همین نقاط عامل کلیدی در بافت سرطانی باشند. یکی از این دومین‌ها CortBP2 است که در چهار ژن CTTNBP2NL (یک جهش)، ژن CTTNBP2 (دو جهش)، ژن FILIP1L (پنج جهش)، و ژن FILIP1 (یک جهش) دارد. پروتئین FILIP1L، یک پروتئین ضد رگزا است و به عنوان سرکوب‌گر تومور عمل می‌کند و از دست دادن آن در ایجاد سرطان تخدمان، سرطان سر و گردن سلول سنگفرشی و سرطان مغز درجه پایین^۲ نقش دارد. در حالی که فرکانس جهش ژن FILIP1L در این مطالعه ما با ارزش نیست، اما دومین CortBP2 رتبه ۷۵ دومین‌های با ارزش را دارد، که نقش جدیدی را برای ژن FILIP1L و سایر ژن‌های شامل دومین CortBP2 سرطان روده پیشنهاد می‌دهد.

برای اثبات یا رد درستی نتایج روش‌های غیرآزمایشگاهی مانند روش‌های پردازش رایانه‌ای، در نهایت باید از روش‌های آزمایشگاهی و تجربی استفاده کرد. روش‌های جدیدی برای مشخص کردن تمایز بین جهش‌های مسافر

^۱Kallmann Syndrome

^۲oligodendrogiomas

و راننده از بین میلیون‌ها تنوع جهشی در پروژه‌های مطالعه توالی تومورها در مقیاس بزرگ نیاز است. روش نرت و همکاران [۷] با تهیه نقشه جهش دومینی برای شناخت جهش‌های راننده مزایای زیادی نسبت به روش سنتی ژن محور دارد، که این امر به خاطر دخیل کردن نقش دومین‌ها در مطالعه، برای شناخت بهتر نقش جهش‌ها در ایجاد و توسعه سرطان‌ها است. مقایسه نقشه جهشی برای دو سرطان پستان و روده بزرگ، برای شناسایی شباهت‌های عمل‌کردی دومین‌ها در سرطان‌های مختلف است. مشخص کردن این‌که چه دومین‌هایی در ایجاد سرطان‌ها دخیل هستند می‌تواند مجموعه جدیدی از ژن‌ها و دومین‌ها را برای مطالعات دارویی و ساخت داروهای جدیدتر و مؤثر فراهم آورد.

۴.۱.۳ روش‌ها و ابزارهای تهیه نقشه جهش ژنی و دومین

داده‌های مربوط به جهش‌های کل اگزون‌های ۱۰۰ بیمار سرطان روده بزرگ و ۵۲۲ بیمار سرطان پستان از بانک اطلاعاتی تی‌سی‌جی‌ای [۳۱] بارگیری شده است. برای بیمارها جهش‌های سوماتیک و دودمان زایشی و تغییرات تک نوکلئوتیدی و حذف و اضافه‌های کوتاه به همراه نمونه بافت سالم برای همه این جهش‌ها موجود است. در این جا جهش‌های دودمان زایشی و جهش‌هایی که کیفیت لازم مورد نظر را نداشته‌اند، حذف شده‌اند. یک مجموعه از جهش‌های سوماتیک تک نوکلئوتیدی و حذف و اضافه شدن برای هر دو مجموعه سرطان مورد مطالعه گردآوری شده است، به طوری که هر جهش حداقل در یک فرد بیمار مشاهده شده باشد. از برنامه آنوار^۱ برای فیلتر کردن جهش‌های یک ریختی، استفاده شده است. به نظر می‌رسد این جهش‌ها نمی‌توانند در ایجاد سرطان دخیل باشند. همچنین از آنوار برای تفسیر جهش در مناطق کدکننده پروتئین با اثرات مرتبط با آن‌ها استفاده شده است. تغییرات تک نوکلئوتیدی به چند دسته تقسیم بندی شده‌اند؛ براساس تغییر دادن یا ندادن اسید‌آمینه. بر اساس ایجاد کردن یک کدن ایست و یا حذف آن. حذف و اضافه شدن‌ها هم بر اساس این‌که باعث جابه‌جا‌یی قاب خواندن اسیدهای آمینه می‌شوند و یا آن را به هم می‌ریزند دسته‌بندی می‌شوند. جهش‌های از جنس تغییرات تک نوکلئوتیدی که اسید‌آمینه را تغییر نمی‌دادند حذف شدند، زیرا به نظر می‌رسد باعث ایجاد سرطان نمی‌شوند. برای هر ژن و دومین در همه بیماران، تعداد جهش‌ها بعد از اعمال تمامی فیلترهای ذکر شده انجام شده (مانند حذف اسنیپ‌ها)، محاسبه شده است. برای دومین‌های مورد مطالعه، یک مجموعه از ۳۹۶۳ پروتئین انسانی از بانک اطلاعاتی رفسک^۲ مورد استفاده قرار گرفته است. برای همترازسازی چند توالی، تمامی دومین‌های پی‌فم، از "بانک اطلاعاتی دومین‌های نگهداری شده"^۳ استفاده شده است.

از میان ۱۱۹۱۲ دومین پی‌فم ۴۲۶۵ دومین به حداقل یک پروتئین نگاشت شده است. همه جهش‌های

^۱ANNOVAR

^۲RefSeq

^۳Conserved Domain Database (CDD)

سوماتیک بر روی مناطق کدکننده پروتئین با استفاده از نرم افزار آنوار به پروتئین‌ها و دومین‌ها نگاشت داده شده‌اند. "مجموعه ژن‌های سرطانی"^۱ به عنوان مجموعه‌ای از ژن‌ها معرفی شده که بر روی هر کدام از این ژن‌ها حداقل یک جهش وجود دارد، این مجموعه بعد از فیلتر کردن چند ریختی‌ها^۲ و جهش‌های تک نوکلئوتیدی منطقی^۳ تهیه شده است. برای مطالعه بر روی دومین‌های پروتئینی، طولانی‌ترین ایزوفرم هر پروتئین از مجموعه رفسک انتخاب شده و جهش‌ها فقط به این ایزوفرم نگاشت داده شده‌اند. برای مشخص کردن فرکانس جهش دومین‌ها و ژن‌های مجموعه بیمار، ابتدا تعداد جهش‌های سوماتیک همه بیماران که داخل منطقه کدکننده پروتئین و دومین هستند، شمرده شده‌اند. برای این‌که انتظار می‌رود مناطق با طول بیشتر تعداد جهش‌های بیشتری نسبت به مناطق با تعداد جهش کمتر داشته باشند، تعداد جهش‌های هر پروتئین نسبت به طول آن پروتئین نرمال شده است. برای دومین‌ها هم تعداد جهش‌های هر دومین نسبت به مجموع طول آن دومین در کل "مجموعه دومین سرطانی" موجود، نرمال شد. برای ارزیابی ژن و دومین‌های معنادار، از ارزیابی شرط هستی شناسی ژن^۴ تمامی ژن و دومین‌های مورد مطالعه در دو سلطان پستان و روده بزرگ استفاده شده است. بسیاری از دومین‌ها هنوز ناشناخته هستند و یا نقش ناشناخته‌ای دارند. بنابراین، این روش فقط قادر به دست‌یابی به حدود ۴۰٪ دومین‌ها بر حسب شرط جی-او که حاوی حداقل یک جهش باشند، است.

به طور خلاصه می‌توان گفت با نگاشت کردن جهش‌های سوماتیک بر روی دومین‌های پروتئینی اطلاعاتی به دست می‌آید که در روش ژن محور قابل دستیابی نیست. مشخص شد برخی دومین‌های معنادار با تعداد جهش‌های بالا وجود دارد که ژن‌های کدکننده پروتئین آن دومین‌ها در مطالعه ژن محور جهش‌های اندکی دارند که معنادار نیست. همچنین برخی دومین‌ها وجود دارند که برخی ژن‌های کدکننده آن‌ها در مطالعه ژن محور جهش‌های معنادار دارند. برای بعضی از چنین دومین‌هایی حتی بعد از حذف جهش‌های مربوط به ژن‌های با جهش‌های معنادار، آن دومین باز هم جهش‌های معنادار دارد. به طور خلاصه مراحل کار این مطالعه را می‌توان به صورت این صورت بیان کرد: جهش‌های سوماتیک تعدادی بیمار سرطان روده بزرگ غربال شده تا جهش‌های خاموش و اسنیپ‌ها حذف شوند. جهش‌های باقی مانده به ژن‌ها نگاشت داده شده و نقشه جهش ژنی تهیه شده است که در آن هر ژن معادل یک کوه است و ارتفاع آن عدد حاصل از تقسیم تعداد جهش‌های هر ژن بر طول همان ژن است. باز دیگر جهش‌ها به دومین‌های پروتئین‌های پی‌فم نگاشت داده شده و نقشه جهش دومینی بر اساس دومین‌های پی‌فم رسم شده است. این نقشه شبیه نقشه جهش ژنی است و برای محاسبه ارتفاع هر کوه دومینی تعداد جهش‌های هر دومین بر طول تجمعی آن دومین تقسیم شده است. از جهش‌های سوماتیک سرطان

^۱Cancer Gene Set

^۲Polymorphisms

^۳Synonymous

^۴Gene Ontology term(GO term)

پستان هم عیناً مانند سرطان روده بزرگ دو نقشه جهش ژنی و دومینی تهیه شده است. با انجام روش آماری، به معرفی مجموعه کاندید از ژنهای دومین‌ها در دو سرطان روده بزرگ و پستان پرداخته شده است. و در نهایت، نتایج حاصل از دو سرطان با هم مقایسه شده‌اند. نتایج حاصل از نقشه جهش ژنی سرطان روده بزرگ با نتایج دو مطالعه دیگر نیز مقایسه شده است.

۲.۳ نتیجه‌گیری

در این فصل و با بیان نتایج مطالعه مشخص شد که مطالعه دومین‌های پروتئینی و تعیین نقش آن‌ها در سرطان می‌تواند مفید باشد. زیرا در روش مطالعه دومین‌ها اطلاعاتی به دست می‌آید که در روش‌های ژن محور قابل دستیابی نیستند.

فصل ۴

روش پیشنهادی و نتایج

در سه فصل اول مقدمات زیستی و ابزارهای لازم به همراه مطالعات انجام شده برای شناسایی نقش ژن‌ها و دومین‌های دخیل در سرطان ذکر شد. در این فصل ابتدا به نحوه جمع آوری داده‌های مورد مطالعه و آماده کردن آن‌ها برای مطالعه پرداخته شده، راه حل جدید پیدا کردن ژن‌ها و دومین‌های دخیل در سرطان شرح داده شده، سپس با استفاده از همین روش ژن‌های میتوکندری، ژن‌های سلول‌های بنیادی، ژن‌های ترمیم کننده و دومین‌های دخیل در سرطان معرفی شده، میزان شباهت دومین‌های پی‌فم و کت پروتئین‌های انسانی حساب شده تا میزان قابلیت اطمینان دومین‌های پی‌فم ارزیابی شود و در نهایت نتایج به دست آمده، با نتایج مطالعات قبلی مقایسه خواهد شد.

۱.۴ جمع آوری داده‌ها و آماده کردن آن‌ها برای مطالعه

داده‌هایی که برای مطالعه انتخاب شده است از وبسایت‌های مختلفی جمع آوری شده‌اند. اسامی و معرفی این وبسایت‌ها در فصل دو آمده است. در اینجا به ذکر اطلاعاتی پرداخته می‌شود که از این وبسایت‌ها بارگیری شده است. ابتدا داده‌های مربوط به جهش‌های سوماتیک کل اگزوم بر روی همه کروموزوم‌های بیماران سرطانی ۲۹ سرطان مختلف از سایت تی‌سی‌جی‌ای بارگیری شد. لیست تمامی فایل‌های دانلودی در جدول ?? پیوست موجود می‌باشد. برای هر بیماری بین یک تا شش فایل مختلف وجود دارد. هر کدام از فایل‌ها مربوط به یکی از روش‌های آزمایشگاهی است که داده‌های اگزومی بیماران براساس آن تهیه شده‌اند. برای توالی یابی اگزوم بیماران گاه‌آغاز از بیش از یک روش آزمایشگاهی استفاده می‌شود که این روش‌ها دقت‌های مختلفی در تشخیص توالی دارند. این تفاوت در روش‌ها باعث می‌شود گاه‌آغاز توالی یابی اگزوم متفاوتی ارائه شود. یعنی در یک روش ممکن است یک نوکلتوتید A خوانده شده اما در روش دیگر G خوانده شود. و نتایج این روش‌های آزمایشگاهی در فایل‌های مختلف قرار داده شده است. در این‌جا، نتایج همه روش‌های مختلف توالی یابی یک بیماری با هم ادغام شده‌اند و

اگر نوکلئوتیدی در حداقل یک روش، جهش یافته گزارش شده باشد، به عنوان جهش پذیرفته شده است. بیمارانی که قبیل از این و یا همزممان با این سرطان، سرطان دیگری داشته‌اند حذف شده‌اند و بیماران مورد مطالعه سابقه هیچ سرطان دیگری نداشته و قبیل از نمونه‌گیری تحت هیچ نوع درمان رادیوتراپی یا شیمی‌درمانی قرار نگرفته‌اند. به دلیل طولانی بودن اسمی سرطان‌ها، در کلیه جدول‌ها از نام اختصاری آن‌ها استفاده شده است. جدول ۲.۰۱ اسمی کامل و مخفف همه سرطان‌ها را نشان می‌دهد.

در فصل یک، به تعریف انواع جهش‌های موجود بر روی ژنوم پرداخته شد. گفته شد، جهش‌های خاموش، جهش‌هایی هستند که در توالی اسید‌آمینه پروتئین تغییری ایجاد نمی‌کنند. بنابراین، می‌توان گفت جهش‌های خاموش، جهش‌هایی هستند که به احتمال خیلی زیادی در ایجاد سرطان دخیل نیستند. پس جهش‌های خاموش هم از مطالعه حذف شدند. پس، بعد از این و در ادامه فصل چهار، هرجایی که لفظ جهش به کار رود منظور جهش‌های غیر خاموش خواهد بود و نه همه انواع جهش‌ها.

بعد از بارگیری داده‌های مربوط به جهش‌های اگزوم بیماران، نیاز بود تا اسمی ژن‌های مختلف به یک شکل استاندارد و واحدی تبدیل شوند تا از داشتن ژن‌هایی با نام‌های متفاوت جلوگیری شود. اسمی که برای ژن‌ها در بانک اطلاعاتی تی‌سی‌جی‌ای به کار گرفته شده است، از استاندارد خاصی تعییت نمی‌کند. بنابراین، برای یافتن و یکسان کردن اسمی ژن‌ها از اسمی مختلف و معادلی که دو بانک اطلاعاتی یوسی‌اس‌سی و هوگو در اختیار گذاشته بودند استفاده شد و اسمی تمامی ژن‌ها به اسم استاندار ژن در هوگو که روش استانداری برای نام‌گذاری ژن‌هاست، تغییر کرد. حال با داشتن اسمی ژن‌های افراد بیمار، باید تشخیص داد کدام ژن‌ها، ژن‌های کدکننده پروتئین هستند. برای شناسایی ژن‌های کدکننده پروتئین هم از استاندار هوگو استفاده شد.

بعد از شناخت ژن‌های کدکننده پروتئین باید دانست موقعیت هر ژن بر روی ژنوم چیست و تعداد اگزون‌ها چقدر است و موقعیتشان بر روی هر ژن چیست، چه پروتئینی تولید می‌کنند و بر روی کدام رشته دی‌ان‌آ قرار دارند. برای یافتن محل ژن‌ها و اگزون‌های آن‌ها و محل شروع و پایان کدکردن پروتئین از روی ژن، از صفحات وبسایت پی‌دی‌بی و بانک اطلاعاتی یوسی‌اس‌سی (جدول‌های مربوط به یوسی‌اس‌سی، رفسک و انسمبل) استفاده شد. هیچ کدام از این جدول‌ها به تنها‌ی شامل اطلاعات لازم در مورد اگزون‌هایی که در پروتئین سازی دخالت دارند نبودند، زیرا پروتئین‌ها ایزوفرم‌های مختلفی دارند و همه این جدول‌ها اطلاعات مربوط به همه ایزوفرم‌های یک پروتئین را نداشتند.

بعد از شناخت پروتئین‌های هر ژن، باید دومین‌های همه پروتئین‌ها در صورت وجود مشخص می‌شد. برای دومین‌های پروتئین از دو بانک اطلاعاتی کت و پی‌فم استفاده شد تا علاوه بر مطالعه دومین‌های پروتئینی، از بانک اطلاعاتی کت که نوعی بانک اطلاعاتی ساختار محور است برای محک زدن اعتبار دومین‌های پی‌فم استفاده شود.

همان طور که گفته شد، برای پیدا کردن دومین‌های پی‌دی‌بی استفاده می‌شود که رشته‌ها بخش‌هایی از پروتئین اصلی هستند که ممکن است توالی‌هایی هم به آن‌ها اضافه شده، یا بخش‌هایی حذف شده باشد. بنابراین از بانک اطلاعاتی سیفتز برای نگاشت دادن دومین گزارش شده بر روی رشته‌های پی‌دی‌بی به پروتئین اصلی استفاده شد. برای دومین‌های کت، از دومین‌هایی استفاده شده است که به طور ۱۰۰٪ بر روی پروتئین اصلی قرار دارند و هیچ اسیدآمینه‌ای به دومین اضافه نشده و یا از آن کاسته نشده است. یعنی دومین کت موجود بر روی رشته پی‌دی‌بی، توالی کاملاً مشابهی در رشته پروتئین اصلی هم داشته باشد و هیچ حذف و اضافه‌ای مربوط به رشته پی‌دی‌بی، درون دومین وجود نداشته باشد. همه دومین‌های انسانی پی‌فم بدون در نظر گرفتن مقدار امتیاز بیت و ارزش مورد انتظار انتخاب شده‌اند. جدول ۳.۳ پیوست شامل آدرس صفحات وب و فایل‌های مختلفی است که بارگیری شده‌اند.

برای کار بر روی داده‌های بیماران سرطانی باید همه اطلاعات بارگیری شده با هم ادغام می‌شوند. در اغلب مطالعات، به جای این کار از نرم افزارهای تهیه شده برای سرعت بخشیدن به مطالعه استفاده می‌شود. این نرم افزارها ممکن است به روز نباشند و یا دارای خطاهای در تولید نتیجه باشند. با توجه به گستردگی این شاخه از علم، اطلاعات متفاوتی برای برخی ژن‌ها یا دومین‌ها وجود دارد. مثلاً یک ژن ممکن در یک بانک اطلاعاتی پروتئینی تولید نکند اما در بانک اطلاعاتی دیگری دارای چند محصول پروتئینی باشد. در این مطالعه با بررسی و مطالعه چندین بانک اطلاعاتی از اجتماع اطلاعات موجود آن‌ها استفاده شده است تا حداکثر همخوانی اطلاعات به دست آید و داده‌های با خطای کمتر تولید شود. نتیجه این مطالعه، تولید جدولی یکپارچه است که اسامی مختلف ژن‌های کدکننده پروتئین، محل آن‌ها بر روی هر کروموزوم، تعداد اگزون‌ها و محل آن‌ها بر روی ژن، اسامی پروتئین کدشونده توسط ژن‌ها و اسامی و محل آغاز و پایان همه دومین‌های پروتئین موجود همه ژن‌های کدکننده پروتئین را نشان می‌دهد. به دلیل حجم بودن، این جدول فقط از طریق وب سایتی که برای این کار تهیه شده است و در ادامه این بخش شرح داده خواهد شد، قابل دسترسی است. بعد از آماده کرده جدول مربوط به اطلاعات ژن‌ها و جدول‌های اطلاعات مربوط به بیماران سرطانی، نگاشت جهش‌ها بر روی ژن‌ها، اگزوم و دومین‌های پروتئینی انجام شده است. جدول ۱.۴ اطلاعات مربوط به تعداد بیماران هر سرطان با حداقل یک جهش بر روی مناطق مختلف ژنوم، مناطق کدکننده پروتئین و دومین‌های پروتئینی را نشان می‌دهد.

۲.۴ روش امتیازدهی به جهش‌ها

همان طور که گفته شد در این مرحله جهش‌های موجود بر روی ژن‌ها مشخص شدن و جهش‌هایی منتظر با آن که روی دومین پروتئین تولید شده توسط آن ژن رخ می‌دهد، تعیین شدن. برای راحتی در نوشتن از این

جدول ۱.۴: تعداد بیماران هر سرطان با حداقل یک جهش بر روی اگزوم و پروتئین دومین‌ها

Cancer Type	Genome	Protein Coding Region	Pfam Domain	CATH Domain
BRCA	983	983	980	914
COAD	269	269	268	265
GBM	275	275	275	270
KIRC	451	451	451	438
KIRP	169	169	169	167
LGG	460	460	460	452
LUAD	545	545	544	541
OV	229	229	228	220
UCEC	248	248	248	248
ACC	92	92	92	87
BLCA	412	412	412	408
CHOL	36	36	36	36
ESCA	182	182	182	181
HNSC	524	524	524	520
KICH	66	66	66	64
LIHC	202	202	202	199
LUSC	178	178	177	176
PAAD	171	170	169	165
PCPG	175	175	175	151
PRAD	425	425	425	375
READ	116	116	116	114
SARC	254	254	254	251
SKCM	75	75	75	74
STAD	387	387	387	383
TGCT	144	144	144	141
THCA	356	356	355	320
THYM	123	123	123	123
UCS	57	57	57	57
UVM	80	80	80	78
Pan Cancer	7684	7683	7674	7418

به بعد از کلمه جهش روی دومین برای نشان دادن جهش‌های روی دومین که حاصل جهش روی ژن کدکننده پروتئین حاوی این دومین استفاده می‌شود. بعد از نگاشت دادن جهش‌ها به دومین‌های پروتئینی، باید جهش‌های راننده پیدا شوند. هنگام تقسیم سلولی جهش‌ها بر روی ژنوم رخ می‌دهند، برخی از آن‌ها تصحیح می‌شوند اما برخی دیگر تصحیح نشده و به سلول جدید منتقل می‌شوند. این جهش‌ها می‌توانند هیچ تغییری در عملکرد سلول ایجاد نکنند یا بر عکس باعث بروز برخی بیماری‌ها در فرد شوند. جهش‌ها، به طور تصادفی و مستقل از هم رخ می‌دهند. پس می‌توان احتمال رخ دادن هر جهش بر روی ژنوم را با توزیع دوچمله‌ای حساب کرد. از آنجایی که جهش‌های موجود بر روی ژنوم به پروتئین‌های سازنده هر کدام از ژن‌های کدکننده پروتئین نگاشت داده شده است، و جهش‌های موجود بر روی دومین‌های پروتئینی مشخص شده است و به علاوه، چون نمونه‌های مورد مطالعه مربوط به بیماران است، پس انتظار می‌رود تعداد جهش‌های بیشتری بر روی دومین‌های خاصی که در ایجاد سرطان دخیل هستند مشاهده شود. یعنی تصادف در رخداد آن‌ها دخالت کمتری دارد. به تعبیری دیگر، اگر جهش‌های موجود بر روی دومین خاصی با احتمال اتفاق افتاده باشند، پس با اطمینان بیشتری می‌توان

گفت که جهش‌های موجود بر روی این دومین نقش بیشتری در ایجاد بیماری دارند.

همان‌طور که گفته شد، توزیع جهش‌ها بر روی ژنوم از نوع توزیع دوچمله‌ای است. اگر n برابر تعداد کل جهش‌های رخداده بر روی همه دومین‌ها باشد، p احتمال رخدادن یک جهش بر روی یک دومین خاص است که برابر طول آن دومین در کل مجموعه پروتئین‌ها تقسیم بر مجموع طول همه دومین‌ها در پروتئین‌های موجود در ژنوم است، و k تعداد جهش‌های موجود بر روی دومین خاص است. هدف این مسئله یافتن محاسبه احتمال رخداد k جهش بر روی دومین خاص است که آن را با $Pr(X = k)$ نمایش می‌دهند و به صورت زیر حساب می‌شود.

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)} \quad (1.4)$$

با توجه به طول زیاد مجموعه دومین‌های پروتئینی انسانی و تعداد زیاد جهش‌ها، برای کاهش خطای محاسباتی از لگاریتم احتمال دوچمله‌ای حساب شده برای مقایسه استفاده می‌شود. پس فرمول زیر به دست می‌آید:

$$\log(Pr(X = k)) = \log \binom{n}{k} + k \log p + (n - k) \log (1 - p) \quad (2.4)$$

اما حتی محاسبه مقدار $\log \binom{n}{k}$ هم به دلیل نوع مسئله دچار خطای محاسباتی بالایی است و باعث سر ریز می‌شود. پس برای حل این مشکل از تقریب استرلینگ^۱ برای محاسبه $\log \binom{n}{k}$ استفاده شده است که به این صورت است [۷۹].

$$\log_2 \binom{n}{k} \sim nH\left(\frac{k}{n}\right) \quad (3.4)$$

که در اینجا،

$$H(\epsilon) = -\epsilon \log_2(\epsilon) - (1 - \epsilon) \log_2(1 - \epsilon) \quad (4.4)$$

آنتروپی دوتایی^۲ برای متغیر ϵ است. یعنی در نهایت از فرمول

^۱Stirling's approximation

^۲binary entropy

(۵.۴)

$$\log_{\gamma} Pr(X = k) = n \left(\frac{k}{n} \log_{\gamma} \left(\frac{k}{n} \right) - \left(1 - \frac{k}{n} \right) \log_{\gamma} \left(1 - \frac{k}{n} \right) + k \log_{\gamma} p + (n - k) \log_{\gamma} (1 - p) \right)$$

برای محاسبه میزان احتمال رخداد تعداد جهش‌های هر دومین استفاده شده است. در فصل دو گفته شد که توزیع دو جمله‌ای را می‌توان با توزیع نرمال استاندارد تقریب زد. پس می‌توان برای این مسئله از آزمون فرض به روش زیر استفاده کرد.

فرض صفر = جهش‌ها بر روی دومین به طور تصادفی رخ داده‌اند.

فرض مقابل = جهش‌ها بر روی دومین به طور تصادفی رخ نداده‌اند.

آزمون صفر وقتی رد می‌شود که احتمال فرض صفر از سطح معنی‌داری تعیین شده کمتر نباشد. پس وقتی فرض صفر رد می‌شود که رابطه زیر برقرار باشد.

$$Pr(X > x) \simeq P(Z > \frac{x - np}{\sqrt{np(1-p)}}) = 1 - P(Z < \frac{x - np}{\sqrt{np(1-p)}}) < \alpha = .05 \quad (6.4)$$

همین‌طور در فصل دو گفته شد از آزمون فرضیه‌های متعدد برای بهبود نتیجه آزمون فرض استفاده می‌شود. در اینجا برای هر دسته بیماری هر آزمون فرض به تعداد دومین‌ها اجرا می‌شود پس بنابر آزمون فرضیه‌های متعدد و تصحیح بونفرونی، اگر تعداد دومین‌ها در کل مطالعه را m فرض کنیم سطح معناداری هر فرضیه باید حداقل $\frac{1}{m}\alpha$ باشد. پس در نهایت برای رد فرض صفر (یافتن دومین دخیل در ایجاد یا توسعه سرطان) نیاز داریم تا نامساوی زیر برقرار باشد.

$$1 - P(Z < \frac{x - np}{\sqrt{np(1-p)}}) < \frac{1}{m}\alpha \Rightarrow P(Z < \frac{x - np}{\sqrt{np(1-p)}}) > 1 - (\frac{1}{m}\alpha) \quad (7.4)$$

در نتیجه ابتدا احتمال رخداد جهش‌ها بر روی دومین‌ها حساب شده و دومین‌هایی که احتمال رخدادن جهش روی آن‌ها کمتر از پنج درصد باشد، به عنوان دومین‌های با شанс بالا برای انتخاب شدن به عنوان دومین‌های کاندید که جهش‌های موجود بر روی این دومین‌ها، در ایجاد و توسعه سرطان نقش دارند اعلام می‌شود. در مرحله بعد، از تصحیح بونفرونی برای پیدا کردن دومین‌های کاندید از بین دومین‌های با شанс بالا استفاده می‌شود. به این صورت که اگر تصحیح خطای جمعی هر دومین با شанс بالا، کمتر از پنج درصد باشد به عنوان دومین کاندید

معرفی می‌شود. پس بعد از این و در ادامه فصل چهار، دومین کاندید، به دومینی اطلاق می‌شود که هم احتمال رخداد کل جهش‌های آن کمتر از ۵ درصد باشد و هم تصحیح خطای جمعی آن کمتر از پنج درصد باشد. و جهش معنادار، جهشی است که بر روی یک دومین کاندید قرار داشته باشد. برای محاسبه تصحیح خطای جمعی تقسیم پنج درصد بر (30×30) تعداد کل دومین‌ها) استفاده شده است. عدد ۳۰ برابر تعداد انواع سرطان‌های مورد مطالعه + ۱ می‌باشد. زیرا به جز ۲۹ سرطان مورد مطالعه، یک مطالعه یک‌پارچه سرطان هم داریم پس هر فرضیه به تعداد دومین‌های هر مجموعه \times (تعداد انواع سرطان + ۱) بار اجرا می‌شود. یعنی برای هر دومین پی‌فم از سطح معناداری $\frac{1}{30 \times 30} < \alpha$ و برای هر دومین کت هم از سطح معناداری $\frac{1}{30 \times 759} < \alpha$ استفاده شده است.

۳.۴ ژن‌های میتوکندری

در فصل یک و دو به نقش میتوکندری در بدن انسان و اهمیت آن در سلول‌ها و نقش آن در پیشگیری، بروز و یا درمان سرطان پرداخته شد. گفته شد که ۳۷۴ ژن بر روی دی‌ان‌آ میتوکندری وجود دارد که ۲۲ ژن آن پروتئین تولید می‌کنند. در این بخش، همه ۳۷ ژن میتوکندری بدون در نظر گرفتن محصول تولیدی مورد مطالعه قرار گرفته است. ۳۷ ژن میتوکندری مجموعاً ۱۵۳۹۲ اسیدونوکلئوتید طول دارند و ۹۳ درصد طول میتوکندری را می‌پوشانند. از کل ۳۳۸۳۷۵۶ جهش بر روی ژن‌های ۱۱۲۱ بیمار مبتلا به انواع سرطان‌های مورد مطالعه، ۲۰۴۰ جهش بر روی میتوکندری قرار دارد. ژن‌های میتوکندری با توجه به نوع عملکردشان به شش مجموعه مختلف هم تقسیم می‌شوند. در اینجا یک بار با مطالعه ژن‌ها و جهش‌های آن‌ها به معرفی ژن‌های کاندید پرداخته و بار دیگر با مطالعه مجموعه‌ها به معرفی مجموعه‌های کاندید پرداخته شده است. برای مطالعه مجموعه‌ها، مجموع جهش‌های موجود بر روی همه ژن‌های آن مجموعه به عنوان جهش‌های آن مجموعه و مجموع طول ژن‌های آن مجموعه به عنوان طول مجموعه در نظر گرفته شده است. برای محاسبه معناداری هر ژن یا مجموعه در میتوکندری از روشی که در بخش قبل گفته شد، استفاده شده است. با این تفاوت که اگر n برابر تعداد کل جهش‌های رخداده بر روی همه ژن‌های میتوکندری باشد، p احتمال رخدادن یک جهش تصادفی بر روی یک ژن خاص آن است که برابر طول آن ژن تقسیم بر مجموع طول همه ژن‌های میتوکندری است. برای هر ژن میتوکندری هم از سطح معناداری $\frac{1}{30 \times 37} < \alpha$ استفاده شده است. برای مجموعه‌ها اگر n برابر تعداد کل جهش‌های رخداده بر روی همه مجموعه‌های میتوکندری باشد، p احتمال رخدادن یک جهش بر روی یک مجموعه خاص آن است که برابر طول آن مجموعه تقسیم بر مجموع طول همه مجموعه‌های دی‌ان‌آ میتوکندری است، و درنهایت k تعداد جهش‌های موجود بر روی یک مجموعه در میتوکندری است. برای هر مجموعه میتوکندری هم از سطح معناداری $\frac{1}{30 \times k} < \alpha$ استفاده شده است. با اعمال این روش

جدول ۲.۴: تعداد بیماران دارای جهش‌های میتوکندریایی و تعداد کل جهش‌های میتوکندریایی در هر بیماری

Cancer Type	Total patients	Number of Mutations on M Chromosome	Patients with Mutations on M Chromosome
BRCA	983	305	184
COAD	269	0	0
GBM	275	0	0
KIRC	451	76	60
KIRP	169	23	16
LGG	460	15	11
LUAD	545	106	62
OV	230	0	0
UCEC	248	86	119
ACC	92	80	40
BLCA	412	17	11
CHOL	36	87	28
ESCA	182	239	98
HNSC	524	24	14
KICH	66	128	41
LIHC	202	128	86
LUSC	178	0	2
PAAD	171	8	7
PCPG	175	0	0
PRAD	425	26	19
READ	116	0	0
SARC	254	136	97
SKCM	75	9	7
STAD	387	51	13
TGCT	144	104	49
THCA	356	45	33
THYM	123	180	61
UCS	57	74	39
UVM	80	0	0
Pan Cancer	7685	2040	1121

نتایج زیر حاصل شده است. همان‌طور که قبلاً هم گفته شده از ژنوم مرتع برای تعیین ژن‌هایی که در ایجاد یا توسعه سرطان نقش دارند استفاده شده است. یعنی با مقایسه ژنوم افراد بیمار با ژنوم مرتع، نقاطی از ژنوم افراد بیمار که با ژنوم مرتع فرق داشت به عنوان جهش مورد مطالعه قرار گرفت. جدول ۲.۴ تعداد بیمارانی که دارای جهش‌هایی بر روی میتوکندری هستند و تعداد کل جهش‌های میتوکندریایی هر بیماری را نشان می‌دهد.

• سرطان مهاجم پستان

از بین ۹۸۳ بیمار مبتلا به سرطان مهاجم سینه، ۱۸۵ بیمار جهش‌هایی بر روی دی.ان.آ میتوکندری داشتند. در مطالعات قبلی [۹] ژن‌های MT-RNR2, MT-ND1, MT-ND2, MT-ND4, MT-ND5, MT-CYB, MT-ATP6 میتوکندری به عنوان ژن‌های دخیل در این سرطان معرفی شده بودند. اما در اینجا، فقط سه ژن MT-RNR1, MT-RNR2, MT-TT و مجموعه rRNA کاندید هستند.

• سرطان روده بزرگ

برای ۲۶۹ بیمار دارای سرطان روده بزرگ، هیچ جهش میتوکندریایی مشاهده نشد و میتوان گفت که جهش‌های میتوکندریایی تاثیری در ایجاد این بیماری ندارد. اما در مطالعات قبلی [۹] جهش‌هایی بر روی ژن‌های میتوکندریایی مشاهده شده بود.

• سرطان تومور مغز

برای این سرطان نیز مانند سرطان روده هیچ جهش میتوکندریایی مشاهده نشد. ۲۷۵ بیمار دارای سرطان تومور مغز مورد آزمایش قرار گرفتند. به نظر می‌رسد به نمونه‌های بیشتری برای مطالعه اهمیت این نوع جهش‌ها در این سرطان نیاز است. در مطالعات قبلی [۹] جهش‌هایی دی‌ان‌آ میتوکندری مشاهده نشده است.

• سرطان سلول روشن کلیه

۶۰ نفر از ۴۵۱ نفر دارای جهش‌هایی بر روی میتوکندری هستند. ژن‌های زیادی دارای جهش هستند. اما هیچ ژن یا مجموعه‌ای جهش‌های معنادار ندارد. در مطالعات قبلی [۹] تمایزی بین انواع سرطان‌های کلیه MT-ND1, MT-ND3 قائل نشده و همه را به طور یک‌جا مطالعه کرده و جهش‌هایی بر روی دو ژن گزارش کرده‌اند.

• سرطان سلول پاپیلاری کلیه

برای این سرطان ۱۶۹ نفر از ۱۶ نفر دارای جهش‌های میتوکندریایی هستند. پراکندگی این جهش‌ها بسیار زیاد است. بنابراین وجود دخالت جهش‌های میتوکندریایی در ایجاد این سرطان کمزنگ به نظر می‌رسد. نتیجه آزمایش معناداری که در اینجا استفاده شده است هم این موضوع را تائید می‌کند. زیرا هیچ جهش معناداری بر روی ژن یا مجموعه وجود ندارد. دو جهش بر روی رونوشت‌های نامتعارف BC018860, JA040725 قرار داشت.

• تومور درجه پایین مغز

رفتار جهش‌های میتوکندریایی برای تومور درجه پایین مغز هم مانند سرطان سلول پاپیلاری کلیه است یعنی جهش‌های کم و پراکنده بر روی دی‌ان‌آ میتوکندری ۴۶۰ بیمار سرطانی تومور درجه پایین مغز دیده می‌شود. اما هیچ کدام از ژن‌ها و مجموعه‌ها دارای اهمیت خاصی نیستند. یک جهش بر روی رونوشت نامتعارف BC018860 قرار داشت.

• سرطان ریه

در این سرطان، برای ۶۲ نفر از ۵۴۵ نفر جهش‌های مورد نظر مشاهده شد. ژن‌های MT-TL1, MT-

TY, MT-CYB, MT-ND1, MT-ND4, MT-CO2, MT-ND5, MT-CO1, MT-CO3, MT-RNR2, MT-ND4L, MT-ATP6 دارای جهش بودند. با این که هیچ ژن کاندیدی وجود ندارد اما مجموعه COMPLEX III دارای جهش‌های معنادار است که فقط شامل ژن MT-CYB است. شش جهش بر روی رونوشت نامتعارف BC018860 و یک جهش هم بر روی JA040725 قرار داشت.

• سرطان سیروز تخمداهن

از بین ۲۳۰ نفر بیمار مورد مطالعه هیچ شخصی دارای جهش‌های میتوکندریایی نبود. در مطالعات قبلی [۹] برخی بیماران جهش‌هایی بر روی ژن‌های RNR1,RNR2, MT-CYB داشتند.

• سرطان مخاط رحم

در این سرطان، حدود ۵۰٪ بیماران دارای جهش‌های میتوکندریایی هستند. ژن‌های MT-TV, MT-TC, MT-RNR1, MT-TI, MT-TL1, MT-RNR2, MT-TW, MT-TM, MT-TG, MT-TS1, MT-TH, MT-ND1, MT-CYB دارای rRNA دارای RNR1,RNR2,MT-TV جهش‌های معنادار بود. از بین مجموعه‌ها هم مجموعه JA040725، سه جهش بر روی مخاط رحم بود. بیشترین تعداد جهش‌های رونوشت‌های نامتعارف میتوکندری مریبوط به سرطان جهش‌های معنادار بود. ۱۶ جهش بر روی JA429504 و چهار جهش هم بر روی BC018860 قرار داشت.

• سرطان غدد فوق کلیه

حدود ۴۵ درصد بیماران یعنی، ۴۰ نفر از ۹۲ نفر، جهش‌های میتوکندریایی داشتند. ۱۲ ژن زیر از مجموعه کل ژن‌های میتوکندری جهش داشتند:

MT-TL1, MT-TY, MT-CYB, MT-ND5, MT-ND4, MT-CO1, MT-RNR2, MT-ND6, MT-CO2, MT-CO3, MT-ND4L, MT-ND1 تنها ژن MT-TL1 دارای جهش‌های معنادار بود. هیچ مجموعه‌ای جهش‌های معنادار نداشت.

• سرطان مثانه

از بین ۴۱۲ بیمار تنها ۱۱ نفر جهش میتوکندریایی داشتند و این جهش‌ها روی ژن‌ها کاملاً پراکنده بودند. هیچ جهش معناداری وجود ندارد. در مطالعات قبلی جهش‌ها فقط بر روی مناطق دی-لوپ قرار داشتند.

• سرطان مجرای صفوراوی

بیشتر جهش‌های میتوکندریایی مریبوط به سرطان مجرای صفوراوی است. از بین ۳۶ نفر با جهش‌های

مورد نظر مشاهده شد. ژن MT-RNR2 و مجموعه COMPLEX III دارای جهش‌های معنادار بودند.

• سرطان مری

بیمارهایی با جهش‌های میتوکندریایی ۵۵٪ بیمارها را تشکیل می‌دادند؛ یعنی ۱۰۰ بیمار از کل ۱۸۲ بیمار مورد مطالعه. ژن MT-CYB و مجموعه COMPLEX III دارای جهش‌های معنادار بودند. یک جهش هم بر روی رونوشت نامتعارف J01415.25 داشت. در مطالعات قبلی [۹] دی-لوپ دارای جهش بود و جهش‌های نادری بر روی چند ژن قرار داشته است.

• سرطان سروگردان

تنها سه درصد بیماران از بین ۵۲۴ نفر بیمار مورد مطالعه، جهش‌هایی بر روی میتوکندری داشتند و این جهش‌ها بر روی ۱۱ ژن پراکنده بود؛ جهش‌های معنادار با ۱۰ جهش به ژن MT-ND5 تعلق داشت. هیچ مجموعه‌ای جهش‌های معنادار نداشت. در مطالعات قبلی [۹] جهش‌ها بر روی مناطق دی-لوپ مشاهده شده بود.

• سرطان کروموفوب کلیه

۴۱ بیمار از ۶۶ بیمار دارای جهش‌های میتوکندریایی بر روی ۲۰ ژن بودند. جهش‌های معنادار به ژن MT-ND1 و مجموعه COMPLEX III تعلق داشت. یک جهش بر روی BC0188604 قرار داشت.

• سرطان هپاتوسلولار کبد

۴۳٪ بیماران (۸۶ بیمار) جهش‌های میتوکندریایی داشتند. از بین ۱۳ ژن جهش یافته، ژن MT-ND5 و دو مجموعه COMPLEX I، COMPLEX III جهش‌های معنادار داشتند. یک جهش هم بر روی رونوشت نامتعارف J01415.25 موجود بود. در مطالعات قبلی [۹] ارتباطی بین جهش‌های میتوکندری و سرطان کبد مشخص نشده است.

• سرطان سلول سنگفرشی ریه

تنها دو نفر از ۱۷۸ نفر بر روی JA040725 دارای جهش بودند. در مطالعات قبلی [۹] نیز جهش‌ها فقط بر روی مناطق دی-لوپ مشاهده شده بودند.

• سرطان پانکراس

۷ نفر از ۱۷۱ نفر بر روی شش ژن جهش داشتند. هیچ ژن یا مجموعه‌ای جهش‌های معنادار نداشت. همین پراکندگی جهش‌ها در مطالعات قبلی [۹] نیز مشاهده شده است.

• فتوکروموموستوم و پاراگانگلیوم

هیچ جهشی بر روی دی.ان.آ میتوکندری ۲۱ نفر بیمار موجود یافت نشد.

• سرطان پروستات

از بین ۴۲۵ مرد سرطانی، ۲۲ نفر جهش میتوکندریایی داشتند. تنها ژن کاندید، ژن MT-RNR2 بود.

هیچ مجموعه‌ای جهش‌های معنادار نداشت. در مشاهدات قبلی [۹] سه ژن، MT-RNR2, MT-ND1, MT-ND4 دارای جهش بودند.

• سرطان رکتوم

هیچ جهشی بر روی دی.ان.آ میتوکندری ۱۱۶ بیمار مورد مطالعه یافت نشد.

• سرطان سارکوم

۱۰۷ بیمار از ۲۵۴ بیمار جهش‌های میتوکندریایی داشتند. این جهش‌ها بر روی ۲۰ ژن پراکنده بودند. از این بین، ژن‌های COMPLEX IV, rRNA و مجموعه‌های MT-RNR2, MT-CO2 دارای جهش‌های معنادار بودند. یک جهش هم به رونوشت نامتعارف J01415.25 تعلق داشت.

• سرطان ملانومای پوست

از بین ۷۵ بیمار مورد مطالعه هفت بیمار بر روی هفت ژن میتوکندریایی دارای جهش بودند. هیچ جهش معنادار یافت نشد.

• سرطان معده

۳۸۷ بیمار مورد مطالعه قرار گرفتند و ۱۳ بیمار با جهش‌های مورد نظر شناسایی شدند. این جهش‌ها بر روی ۱۵ ژن قرار داشتند. نکته جالب در مورد جهش‌های این سرطان تعداد زیاد جهش‌ها برای هم بیمار بود به طور متوسط هر بیمار ۲.۶ جهش بر روی دی.ان.آ میتوکندری داشت، که این مقدار دو برابر میانگین جهش هر بیمار در تمامی سرطان‌های مورد مطالعه است. اما هیچ کدام معنادار نیستند. دو جهش هم بر روی BC018860 گزارش شده است. در تحقیقات قبلی [۹] سه ژن، MT-ND1, MT-ND5, MT-CO1 و دی-لوپ، مناطقی بودند که در بیماران مورد مطالعه دارای جهش بودند.

• سرطان سلول‌های زایشی بیضه

۳۴ درصد٪ بیماران (۴۹ نفر) جهش‌های میتوکندریایی داشتند، که بر روی ۱۷ ژن پراکنده بود. مجموعه COMPLEX III دارای جهش‌های معنادار بود.

• سرطان تیروئید

۳۳ نفر از ۳۵۶ نفر جهش میتوکندریایی داشتند. این جهش‌ها بر روی ۱۲ ژن پراکنده بودند. در مطالعات قبلی [۹] دو مجموعه COMPLEX I, COMPLEX IV دارای جهش بودند. اما در اینجا فقط ژن MT-ND4 و مجموعه COMPLEX I دارای جهش‌های معنادار هستند.

• سرطان تیموس

۶۱ نفر از ۱۲۳ نفر جهش‌های میتوکندریایی داشتند، و ژن rRNA و مجموعه MT-RNR2 دارای جهش‌های معنادار بودند.

• سرطان رحم

۴۰ بیمار از مجموع ۵۷ بیمار مورد مطالعه جهش‌های میتوکندریایی داشتند. که این جهش‌ها بر روی ۱۶ ژن قرار داشت. اما جهش معناداری وجود ندارد. سه جهش بر روی BC018860 و یک جهش هم بر روی JA040725 قرار دارد.

• سرطان چشم

هیچ جهشی در بیماران این دسته (۸۰ نفر) وجود نداشت.

• مطالعه یکپارچه سرطان

در مطالعه همه سرطان‌ها به طور یک‌جا و واحد، دو ژن MT-CYB, MT-RNR2 از بین ۳۷ ژن rRNA جهش‌های معنادار دارند. همچنین در مطالعه یکپارچه، در بین شش مجموعه تنها دو مجموعه COMPLEX III جهش معنادار ندارد.

با توجه به مطالعه تک‌تک بیماری‌ها به طور جداگانه می‌توان گفت، نقش جهش‌های موجود بر روی دی‌ان‌آ میتوکندری برای بیماری‌های مختلف متفاوت است. برای برخی سرطان‌ها، جهش‌های موجود در ایجاد آن‌ها بی‌تأثیرند اما برای برخی دیگر میزان مشارکت جهش‌های میتوکندریایی در ایجاد یا توسعه سرطان زیاد است. جدول ۴.۴ ژن‌ها و مجموعه‌های معنادار برای هرکدام از بیماری‌ها را در صورت وجود نشان می‌دهد. علاوه بر نام ژن‌ها، تعداد بیمارانی که بر روی هر کدام از این ژن‌ها حداقل یک جهش دارند هم مشخص شده است. اما برای جهش‌های موجود بر روی رونوشت‌های نامتعارف میتوکندری، فقط به ذکر تعداد جهش‌ها و تعداد بیماران دارای این رونوشت‌ها پرداخته شده است؛ جدول ۴.۴ این اطلاعات را نشان می‌دهد.

با این‌که میتوکندری فقط از مادر به ارث می‌رسد بنابراین شاید به نظر بررسد جهش‌های موجود بر روی آن در ایجاد سرطان‌های مخصوص زنان پرنگ‌تر از سایر سرطان‌ها باشد، اما در این مشاهدات با این‌که ۵۰ درصد بیماران

سرطان مخاط رحم جهش‌های میتوکندریایی داشتند اما هیچ یک از ۲۳۰ بیمار سرطان سیروز تخدمان جهش میتوکندریایی نداشتند. پس نمی‌توان گفت جهش‌های میتوکندریایی در سرطان‌های مخصوص زنان دخالت دارند.

۴.۴ ژن‌های سلول‌های بنیادی

در فصل یک و دو درمورد اهمت ژن‌های سلول‌های بنیادی و نقشی که این سلول‌ها در ایجاد و توسعه سرطان‌ها دارند صحبت شد. در این قسمت به طور مجزا به مطالعه نقش ژن‌هایی که در مطالعه پالمر و همکاران [۵۰] به عنوان ژن‌های سلول‌های بنیادی معرفی شده‌اند، پرداخته می‌شود. اسمای این ۱۸۲ ژن در جدول آ.^۴ پیوست آمده است. برای مطالعه بر روی ژن‌های سلول‌های بنیادی جهش‌های موجود بر روی مناطق کدکننده پروتئین آن‌ها مورد مطالعه قرار گرفت. برای این منظور، اگر n برابر تعداد کل جهش‌های رخداده بر روی همه ژن‌های سلول‌های بنیادی باشد، p احتمال رخدادن یک جهش بر روی یک ژن خاص این مجموعه است که برابر طول آن ژن تقسیم بر مجموع طول همه ژن‌های سلول‌های بنیادی است و k تعداد جهش‌های موجود بر روی یک ژن سلول بنیادی است. برای هر ژن سلول بنیادی از سطح معناداری $\frac{1}{\alpha \times 182} < \alpha$ استفاده شده است. در نهایت با این تحلیل نتایج زیر به دست آمده است:

- سرطان مهاجم پستان

ژن‌های کاندید برای سرطان مهاجم پستان سه ژن KMT2C, PILRB, HLA-DRB5 از مجموعه ژن‌های سلول‌های بنیادی هستند.

- سرطان روده بزرگ

در این مطالعه ژن HLA-DPA1 کاندید برای سرطان روده بزرگ انتخاب شد.

- سرطان تومور مغز

سرطان تومور مغزی جهش‌های معناداری بر روی ژن‌های CHEK2, TSHZ2 داشت.

- سرطان سلول روشن کلیه

نه ژن سلول بنیادی DNMT3B, CHEK2, RRAS2, NREP, FYB, SMARCC2, RCSD1, HLA-DRB5, TNFSF10 در این سرطان جهش معنادار دارند.

- سرطان سلول پاپیلاری کلیه

برخلاف سرطان سلول روشن کلیه که نه ژن کاندید دارد، در سلول پاپیلاری کلیه فقط دو ژن, CHEK2

DPH3 کاندید هستند و فقط ژن کاندید CHEK2 با سرطان سلول روشن کلیه مشترک است.

• تومور درجه پایین مغز

سه ژن ZBTB20 در مطالعه انجام شده، به عنوان ژن‌های کاندید این سرطان انتخاب شدند.

• سرطان ریه

بیشترین سهم معناداری ژن‌های کاندید سلول‌های بنیادی در سرطان‌ها مربوط به سرطان ریه با ۱۰ ژن کاندید زیر است.

SPDL1, CHEK2, TRPC4, CDH6, KMT2C, PILRB, TSHZ2, NPR3, GIMAP1, FYB

٪ ۸۰ بیماران در این سرطان دارای جهش‌هایی بر روی یکی از ژن‌های سلول‌های بنیادی بودند.

• سرطان سیروز تخمداهن

دو ژن BOD1, HAS2 به عنوان ژن‌های کاندید انتخاب شدند. تنها ۳۰٪ افراد (۷۵ نفر از ۲۳۰ نفر) جهش‌هایی بر روی این دسته از ژن‌ها داشتند.

• سرطان مخاط رحم

فقط ژن ATP11C دارای جهش‌های معنادار است.

• سرطان غدد فوق کلیه

این نوع سرطان کلیه چهار ژن کاندید دارد. این ژن‌ها HDAC2, ERCC2, GARS, PRR34 هستند. و هیچ اشتراکی با ژن‌های کاندید دو نوع سرطان سلول روشن و سلول پاپیلاری کلیه ندارند.

• سرطان مثانه

سه ژن کاندید برای سرطان مثانه وجود دارد. این ژن‌ها CHEK2, ERCC2, KMT2C هستند.

• سرطان مجرای صفراءوی

برای مجاری صفراءوی سه ژن GIMAP8 کاندید شناخته شدند.

• سرطان مری

۸۰٪ بیماران جهش‌هایی بر روی ژن‌های سلول‌های بنیادی داشتند اما تنها دو ژن NREP, BRINP1 کاندید ایجاد‌کننده یا توسعه این سرطان انتخاب شده‌اند.

- سرطان سروگردن

پنج ژن CHEK2, LIN28B, BRINP1, KMT2C, NPR3 در این سرطان کандید شناخته شدند.
- سرطان کروموفوب کلیه

ژن‌هایی که در این سرطان جهش‌های معنادار داشتند، ژن‌های DIMT1, KMT2C, HLA-DRB5, HLA-DQA1 بودند.
- سرطان هپاتوسولولار کبد

ژن HTR7, ژن کандید این بیماری شناخته شد.
- سرطان سلول سنگفرشی ریه

با اینکه ۹۰٪ بیماران جهش‌هایی بر روی ژن‌های مورد مطالعه دارند. اما هیچ یک از این جهش‌ها معنادار نیستند.
- سرطان پانکراس

۸۰٪ بیماران جهش‌هایی بر روی ژن‌های میتوکنندی دارند. و تعداد ژن‌های کандید این سرطان به نسبت زیاد است. اسامی این ژن‌ها به شرح زیر است.

CHEK2, BBS9, GARS, SLC24A1, KMT2C, SMARCC2, NPR3, AFTPH
- فوکروموسیتوم و پاراگانگلیوم

ژن‌های کандید این سرطان CHEK2, NUSAP1, KMT2C, HLA-DRB5 هستند. با این‌که فقط ۴۵ نفر از ۱۷۵ نفر بر روی ژن‌های سلول‌های بنیادی جهش دارند اما تعداد ژن‌های کандید نسبتاً زیاد است.
- سرطان پروستات

دو ژن CHEK2, KMT2C دارای جهش‌های معنادار هستند.
- سرطان رکتوم

هیچ ژن کандیدی وجود ندارد. ۵۰٪ بیماران ۲۸۰ جهش بر روی ۱۱۶ ژن مختلف داشتند.
- سرطان سارکوم

دو ژن ZNF788, BRINP1 ژن‌هایی هستند که در مطالعه ژن‌های سلول‌های بنیادی این سرطان، جهش‌های معنادار داشتند.

• سرطان ملانومای پوست

سرطان پوست جز سرطان‌هایی است که ژن‌های با جهش‌های معنادار زیادی نسبت به بقیه دارد. هشت ژن GDF3, CCDC90B, CDH6, KMT2C, GIMAP5, GIMAP7, GIMAP1, GIMAP8 برای این بیماری کاندید هستند. معناداری چهار ژن هم خانواده GIMAP5, GIMAP7, GIMAP1، GIMAP8 در این سرطان می‌تواند برای مطالعات تکمیلی‌تر این سرطان حائز اهمیت باشد.

• سرطان معده

این سرطان هم جهش‌های معناداری بر روی ژن‌های CHEK2, SOHLH2, BRINP1, KMT2C، TSHZ2, ZBTB20 داشت.

• سرطان سلول‌های زایشی بیضه

این سرطان سه ژن کاندید به نام‌های C10orf128, HLA-DRB5, HLA-DQA1 دارد.

• سرطان تیروئید

سرطان تیروئید ژن‌های کاندید متعددی دارد. این ژن‌ها GDF3, RIOK2, HLA-DRB5 هستند.

• سرطان تیموس

تنها ژنی که جهش‌های معناداری در این سرطان داشت، ژن CHEK2 بود.

• سرطان رحم

دو ژن کاندید این سرطان، ژن‌های CHEK2, KMT2C هستند.

• سرطان چشم

سه ژن کاندید در سرطان چشم هستند. CHEK2, NUSAP1, HLA-DRB5.

• مطالعه یکپارچه سرطان

از بین تمامی ژن‌ها، ۱۳ ژن زیر در مطالعه یکپارچه دارای اهمیت بودند. تمامی این ژن‌ها در حداقل یک نوع سرطان خاص هم دارای جهش‌های معنادار بودند.

CHEK2, SOHLH2, BRINP1, TRPC4, CDH6, KMT2C, PILRB, HLA-DRB5, TSHZ2, NPR3, GIMAP1, GIMAP8, ZBTB20

از بین ۱۸۲ ژن مورد مطالعه ۵۷ ژن در حداقل یک نوع بیماری جهش‌های معنادار داشتند. دو ژن, CHK2 به ترتیب در ۱۸ و ۲۰ بیماری جهش‌های معنادار داشتند. پس نمی‌توان اهمیت این ژن‌ها در سرطان را نادیده گرفت. اما بقیه ۵۵ ژن در حداکثر هفت بیماری جهش‌های معنادار داشتند. مجموع مناطق کدکننده پروتئینی ۱۸۲ ژن سلول‌های بنیادی مورد مطالعه ۳۲۶۲۵۳ اسیدنوکلئوتید است. به جز کروموزوم‌های ۱۶، ۷ و M بقیه کروموزوم‌ها شامل یک یا چند ژن سلول بنیادی هستند. ۱۴۹۰۲ جهش بر روی این ژن‌ها قرار دارد. به جز دو سرطان سلول سنگفرشی ریه و سرطان رکتوم بقیه سرطان‌ها بین یک تا ۱۰ ژن کاندید دارند. ۱۳۷ ژن از ۱۸۲ ژن هیچ جهش معناداری ندارند. پر تکرارترین ژن بین تمامی سرطان‌ها، ژن CHEK2 است که در ۱۶ سرطان جهش‌های معنادار دارد. جدول ۵.۴ ژن‌های کاندید هر سرطان را صورت وجود نمایش می‌دهد. به طور متوسط ۶۰٪ بیماران جهش‌هایی بر روی ژن‌های سلول‌های بنیادی دارند. این میزان بین ۲۰٪ تا ۹۰٪ برای سرطان‌های مختلف متفاوت است. در جدول ۵.۴ علاوه بر اسمی ژن‌ها، میزان بیمارانی که دارای ژن‌های سلول‌های بنیادی با جهش‌های معنادار هستند، نشان داده شده است. در بخش مربوط به مطالعه دومین‌ها به نقش دومین‌های ژن‌های سلول‌های بنیادی در سرطان‌ها پرداخته خواهد شد.

۵.۴ ژن‌های ترمیم کننده

در فصل یک و دو در مورد ژن‌های ترمیم و اهمیت مطالعه آن‌ها در سرطان‌ها توضیح داده شد. روش انتخاب ژن‌های ترمیم کننده کاندید برای هر سرطان، درست مانند ژن‌های سلول‌های بنیادی است. تعداد این ژن‌ها ۱۷۴ عدد است که در جدول آ.۵ پیوست، اسمی این ژن‌ها را آمده است. در ادامه به تفکیک به ارائه اسمی ژن‌های کاندید هر سرطان به تفکیک بیماری پرداخته می‌شود. در این بخش از مطالعه از شرط‌های زیر برای محاسبه احتمالات استفاده شده است. اگر n برابر تعداد کل جهش‌های رخداده بر روی همه ژن‌های ترمیم کننده باشد، p احتمال رخدادن یک جهش بر روی یک ژن خاص این مجموعه است که برابر طول آن ژن تقسیم بر مجموع طول همه ژن‌های ترمیم کننده است و k تعداد جهش‌های موجود بر روی یک ژن ترمیم خاص است. برای هر ژن ترمیم کننده از سطح معناداری $\frac{1}{3.0 \times 174} < \alpha$ استفاده شده است. با اعمال این روش و این پارامترها، نتایج ذیل حاصل شده است.

• سرطان مهاجم پستان

تنها ژن ترمیم کننده کاندید برای سرطان مهاجم پستان، ژن معروف TP53 است.

• سرطان روده بزرگ

در این سرطان فقط ژن TP53 کاندید است.

• سرطان تومور مغز

ژن ترمیم کننده کاندید در اغلب سرطان‌ها یعنی TP53، در این سرطان نیز حضور دارد.

• سرطان سلول روشن کلیه

این سرطان جهش‌های معنادار بر روی ژن‌های FANCE, DDB1, RPA1, TP53, CHEK2 داشت.

• سرطان سلول پاپیلاری کلیه

ژن‌های کاندید این بیماری OGG1, MSH3, TDG, TP53, CHEK2 هستند. اشتراک ژن‌های کاندید این سرطان و سرطان سلول روشن کلیه تنها در دو ژن CHEK2 و TP53 است.

• تومور درجه پایین مغز

نتیجه مطالعه این سرطان شامل دو ژن کاندید CHEK2 و TP53 است.

• سرطان ریه

۸۵٪ بیماران یعنی ۴۶۳ بیمار از بین ۵۴۵ بیمار جهش‌هایی بر روی ۱۶۲ ژن ترمیم کننده دارند، که فقط دو ژن ERCC5, TP53, CHEK2 دارای جهش‌های معنادار هستند.

• سرطان سیروز تخدمدان

فقط ژن TP53 دارای جهش‌های معنادار است.

• سرطان مخاط رحم

دو ژن MSH2 و CHEK2 ژن‌های کاندید این بیماری هستند.

• سرطان غدد فوق کلیه

این سرطان سه ژن ترمیم کننده کاندید دارد. این ژن‌ها MSH3, TP53 و ERCC2 هستند.

• سرطان مثانه

ژن‌های ATM, TP53, ERCC2, CHEK2 ژن‌های کاندید سرطان مثانه هستند.

• سرطان مجرای صفراء

فقط دو ژن TP53, CHEK2 دارای جهش‌های معنادار هستند.

- سرطان مری
ژن TP53 جهش‌های معنادار دارد.
- سرطان سروگردن
برای سرطان سروگردن ژن‌های TP53, CHEK2 جهش‌های معنادار داشتند.
- سرطان کروموفوب کلیه
فقط ژن TP53 جهش‌های معنادار دارد.
- سرطان هپاتوسلولار کبد
همانند سرطان کروموفوب کلیه، فقط ژن TP53 جهش‌های معنادار دارد.
- سرطان سلول سنگفرشی ریه
ژن TP53 کاندید انتخاب شد.
- سرطان پانکراس
این سرطان ژن‌های کاندید زیادی دارد که شامل ژن‌های ERCC3, XPC, WRN, TDG, FAN1, EME2, TP53, CHEK2 است.
- فتوکروموموستیوم و پاراگانگلیوم
ژن‌های کاندید این سرطان FANCD2, ERCC8, TDG, CHEK2 هستند.
- سرطان پروستات
سرطان پروستات چهار ژن ترمیم کننده کاندید به اسمی ATM, TP53, POLI, CHEK2 دارد.
- سرطان رکتوم
این سرطان یک ژن کاندید به نام TP53 دارد.
- سرطان سارکوم
جهش‌های معنادار برای این سرطان مربوط به ژن TP53 بود.
- سرطان ملانومای پوست
سرطان پوست دارای ژن‌های کاندید BLM, MPG, TP53, CHEK2 است.

• سرطان معده

ژن‌های کاندید این سرطان چهار ژن UVSSA, SLX4, TP53, CHEK2 هستند.

• سرطان سلول‌های زایشی بیضه

این سرطان تنها سرطانی است که هیچ جهش معناداری بر روی هیچ یک از ژن‌های ترمیم کننده ندارد.

• سرطان تیروئید

سرطان تیروئید چهار ژن کاندید به نام‌های SMUG1, TDG, TP53, CHEK2 دارد.

• سرطان تیموس

تنها ژن کاندید، ژن CHEK2 است.

• سرطان رحم

فقط ژن TP53 که پرتفکرارترین ژن کاندید است، در این سرطان، جهش‌های معنادار دارد.

• سرطان چشم

در سرطان چشم ۳۰ بیمار از ۸۰ بیمار از ۲۹ ژن جهش ۲۴ ژن مختلف ترمیم کننده دارند. از این بین، چهار ژن کاندید به نام‌های FANCD2, CCNH, TDG, CHEK2 وجود دارد.

• مطالعه یک پارچه سرطان

چهار ژن ATM, TDG, TP53, CHEK2 در مطالعه یک پارچه سرطان، ژن‌های کاندید شناخته شدند.

از بین ۱۷۴ ژن شناخته شده به عنوان ژن‌های ترمیم کننده، ۲۷ ژن در حداقل یک بیماری جهش‌های معنادار دارند. بیشترین سهم مشارکت را ژن TP53 دارد. این ژن در ۲۵ سرطان از ۲۹ سرطان جهش‌های معنادار دارد و هر سرطانی حداقل یک ژن کاندید ترمیم کننده دارد. به طور متوسط ۶۹ درصد بیماران در حداقل یک ژن از ۱۷۴ ژن ترمیم کننده دارای جهش‌های معنادار هستند. سرطان تیروئید با ۲۱ درصد و سرطان مری با ۹۶ درصد کمترین و بیشترین جهش را بر روی مجموعه مورد مطالعه دارند. اما بیشترین ژن‌های ترمیم کننده کاندید را سرطان پانکراس با هشت ژن دارد. و سرطان سلول‌های زایشی بیضه هیچ جهش معناداری در این مجموعه ژن ندارد. جدول ۶.۴ اسامی ژن‌های کاندید هر بیماری و تعداد بیماران دارای این ژن‌ها را به تفکیک نشان می‌دهد. همانند دو بخش قبل در اینجا هم برای ژن‌های کاندید تعداد بیماران با حداقل یک جهش بر روی هر کدام از این

ژن‌ها مشخص شده است. در بخش مربوط به مطالعه دومین‌ها، به نقش دومین‌های ژن‌های سلول‌های بنیادی در سرطان‌ها پرداخته خواهد شد.

۶.۴ نقش دومین‌ها در سرطان

بخش اصلی پایان نامه، یافتن و معرفی دومین‌های پروتئینی دخیل در هر سرطان است. این مطالعه در دو بخش کت و پی‌فم انجام شده است. در هر بخش به معرفی دومین‌های کاندید پرداخته می‌شود. علاوه بر این، به معرفی ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده‌ای پرداخته می‌شود که حداقل یک دومین معنادار دارند. نتایج این مطالعه با نتایج به دست آمده از سایر مطالعات [۷، ۲۲] مقایسه شده است. برای مطالعه اهمیت نقش دومین‌ها در سرطان‌ها، نگاشت جهش‌های سوماتیک دی‌ان‌آ به دومین‌ها به این صورت انجام می‌شود: تنها جهش‌های سوماتیک موجود بر روی بخش‌هایی از دی‌ان‌آ که بعد از کدشدن به پروتئین، درون نواحی دومین آن پروتئین واقع می‌شوند، انتخاب می‌شوند.

۱.۶.۴ دومین‌های کت

بعد از محاسبه احتمال رخداد جهش‌های موجود بر روی دومین‌های کت و محاسبه پی مقدار برای هر کدام از دومین‌ها، دومین‌های کاندید انتخاب شدن. این روش در ابتدای فصل شرح داده شد و گفته شد که اگر n برابر تعداد کل جهش‌های رخداده بر روی همه دومین‌های کت باشد، p احتمال رخدادن یک جهش بر روی یک دومین خاص آن است که برابر طول آن دومین تقسیم بر مجموع طول همه دومین‌های کت است، و درنهایت k تعداد جهش‌های موجود بر روی یک دومین است. برای هر دومین کت هم از سطح معناداری $\frac{1}{30 \times 759} < \alpha$ استفاده شده است.

تعداد دومین‌های کاندید کت برای بیماران بین یک تا ۱۹ دومین، برای سرطان‌های مختلف متغیر است. در مطالعه یک پارچه سرطان ۹۳ دومین از ۷۵۹ دومین کت، جهش‌های معنادار دارند؛ یعنی حدود ۱۲/۳٪ دومین‌های کت شناخته شده انسانی. جدول آ.۶ پیوست اسامی دومین‌های کاندید همه سرطان‌ها را نشان می‌دهد. برای همه دومین‌های کاندید هر بیماری میزان پوشش افراد بیمار هم محاسبه شده است. سهم پوششی هر دومین بین ۰/۲٪ تا ۹۲٪ بیماران هر دسته است. در کل دومین‌های کاندید هر نوع سرطانی بین ۲۸٪ تا ۹۸٪ بیماران را پوشش می‌دهند. این اطلاعات برای تمامی دومین‌های کاندید در جدول آ.۶ پیوست قابل مشاهده است. برای مطالعه بیشتر بر روی ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده، لیست دومین‌های آن‌ها تهیه شد و سپس مشاهده شد که چند درصد از دومین‌های آن‌ها جز دومین‌های کاندید هستند. آیا ژنی وجود دارد که جز ژن‌های کاندید نباشد اما دومین کاندید داشته باشد و بر عکس یانه. برای هر بیماری هم ژن‌هایی وجود دارد که کاندید اهستند

اما یا اصلاً دومین ندارد یا دومین دارند ولی کاندید نیست و هم ژن‌های غیر کاندیدی وجود دارد که دومین‌های کاندید دارند. بیشتر ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده دومین‌های کت ندارند. از بین ۱۸۲ ژن سلول‌های بنیادی ۴۵ ژن دارای دومین هستند و از این تعداد ژن، ۲۴ ژن کاندید بیماری هستند ولی در همان بیماری دومین کاندید ندارد. به علاوه ۷۱ ژن سلول بنیادی اصلاً دومین کت ندارد. ۱۰۵ دومین کاندید وجود دارد که متعلق به هیچ ژن کاندیدی نیستند و ۱۶ ژن کاندید وجود دارد که حداقل یک دومین کاندید دارند. از بین ۱۷۴ ژن ترمیم کننده ۷۲ ژن دارای دومین هستند و از این تعداد ژن، ۳۵ ژن کاندید بیماری هستند ولی در همان بیماری دومین کاندید ندارد. ۱۱۸ دومین کاندید وجود دارد که متعلق به هیچ ژن کاندیدی نیستند و در نهایت ۴۱ ژن کاندید وجود دارد که حداقل یک دومین کاندید دارند. جدول ۷.۴ لیست ژن‌های سلول‌ها بنیادی را نشان می‌دهد که حداقل یک دومین کاندید در مجموعه کت دارند. جدول ۸.۴ لیست ژن‌های ترمیم کننده را نشان می‌دهد که حداقل یک دومین کاندید در مجموعه کت دارند.

۲.۶.۴ دومین‌های پی‌فم

برای پیدا کردن دومین‌های کاندید پی‌فم هم مانند مجموعه دومین کت عمل شده است. اگر n برابر تعداد کل جهش‌های رخداده بر روی همه دومین‌های پی‌فم باشد، p احتمال رخدادن یک جهش بر روی یک دومین خاص آن است که برابر طول آن دومین تقسیم بر مجموع طول همه دومین‌های پی‌فم است، و در نهایت k تعداد جهش‌های موجود بر روی یک دومین است. برای هر دومین پی‌فم هم از سطح معناداری $\alpha < \frac{1}{3.666}$ استفاده شده است. نقش دومین‌ها در سلطان‌های مختلف متفاوت است برخی تعداد دومین‌های کاندید کم و برخی دیگر تعداد دومین‌های زیادی دارند. به طوری که بیماری سلطان سیروز تخدمان فقط سه دومین کاندید دارد در حالی که سلطان هپاتوسلوار کبد ۹۳ دومین کاندید دارد. در مطالعه یک پارچه سلطان نیز ۲۲۲ دومین کاندید پی‌فمی وجود دارد؛ یعنی حدود ۳/۴٪ دومین‌های پی‌فمی شناخته شده انسانی. جدول آ.۷ پیوست اسامی دومین‌های کاندید همه سلطان‌ها را نشان می‌دهد. مقدار پوشش دومین‌های کاندید هر بیماری به طور جداگانه محاسبه شده است و در جدول آ.۷ پیوست در کنار نام هر دومین ذکر شده است. سهم پوششی هر دومین بین ۰٪ تا ۹۸٪ است. بیماران هر دسته است. در کل دومین‌های کاندید هر نوع سلطانی بین ۷۴٪ تا ۱۰۰٪ بیماران را پوشش می‌دهند. مطالعه بر روی ژن‌های سلول‌های بنیادی و ترمیم کننده و دومین‌های پی‌فم آن‌ها هم انجام شده است. از بین ۱۸۲ ژن سلول‌های بنیادی ۱۷۵ ژن دارای دومین و از این تعداد، ۷۷ ژن کاندید بیماری هستند ولی در همان بیماری دومین کاندید ندارد. ۱۲۵ دومین کاندید وجود دارد که متعلق به هیچ ژن کاندیدی نیستند. ۳۴ ژن وجود دارد که حداقل یک دومین کاندید دارند. از بین ۱۷۴ ژن ترمیم کننده ۱۶۷ ژن دارای دومین هستند. از این تعداد، ۴۱ ژن کاندید بیماری هستند ولی در همان بیماری دومین کاندید ندارد. ۴۵ دومین کاندید وجود دارد.

که متعلق به هیچ ژن کاندیدی نیستند و در آخر ۳۵ ژن وجود دارد که حداقل یک دومین کاندید دارند. جدول ۹.۴ لیست ژن‌های سلول‌ها بنیادی را نشان می‌دهد که حداقل یک دومین کاندید در مجموعه پی‌فم دارند. جدول ۱۰.۴ لیست ژن‌های ترمیم کننده را نشان می‌دهد که حداقل یک دومین کاندید در مجموعه پی‌فم دارند.

نکته دارای اهمیت در مورد ژن CHEK2 که در بین سلول‌های بنیادی و ترمیم کننده مشترک است، این است که این ژن در ۲۴ بیماری دارای جهش‌های معنادار بر روی دومین یا ژن و یا هر روی آن‌هاست. پنج بیماری که در هیچ جهش معنادار بر روی ژن یا دومین CHEK2 ندارد، سرطان سیروز تخمدا، سرطان غدد فوق کلیه، سرطان کروموفوب کلیه، سرطان سلول سنگفرشی ریه و سرطان سلول‌های زایشی بیضه هستند. در مورد ژن TP53 هم باید گفت به جز چهار سرطان فئوکروموسیتوم و پاراگانگلیوم، سلول‌های زایشی بیضه، تیموس و چشم در بقیه سرطان‌ها دارای جهش‌های معنادار بر روی ژن یا دومین‌ها بود. این نتایج حاکی از دخالت قوی این دو ژن در ایجاد و توسعه سرطان‌ها دارد. ژن CHEK2 نقش تقسیم سلولی دارد و پروتئین تولیدی آن با پروتئین‌های زیادی میانکنش دارد مانند پروتئین TP53. تشخیص قوی این ژن‌ها در این مطالعه می‌تواند تاییدی بر درستی روش به کار گرفته شده در شناسایی دومین‌های پروتئینی دخیل در سرطان باشد.

در تحقیق نرت و همکاران [۷]، ۱۰ دومین پی‌فم به عنوان مهم‌ترین دومین‌ها برای سرطان روده بزرگ معرفی شده‌اند. لیست این دومین‌ها در جدول ۲.۳ فصل سه آمده است. در مطالعه‌ای که در اینجا انجام شده است، ۲۳ دومین کاندید به نام‌های 7tm_1, 7tm_4, APC_crr, APC_u5, APC_u9, Androgen_recep, Arm(APC_u3, Cadherin, DUF1388, EPL1, Ion_trans, Keratin_2_tail, MH2, P53, Keratin_B2_2, MHC_I_NKAP, OGFr_III, P53, P53_tetramer, PI3K_p85B, PI3Ka, Ras, TMEM247 برای سرطان روده بزرگ وجود دارد. در بین این دومین‌ها چهار دومین مشترک وجود دارد. این دومین‌ها APC_crr, MH2, PI3K_p85B هستند.

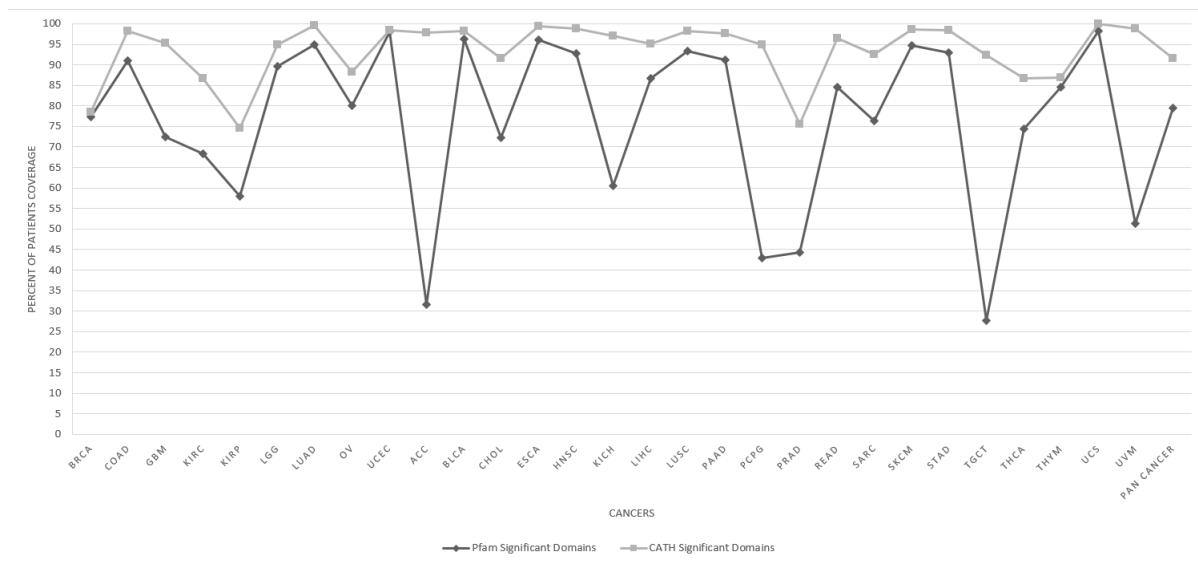
روشی که نرت و همکاران [۷] برای یافتن دومین‌های دخیل در سرطان به کار برده‌اند، ایراداتی دارد. اولین ایراد این است که برای نرمال کردن تعداد جهش‌های هر دومین، تعداد جهش‌های آن دومین را نسبت مجموع طول جمعی آن دومین در ژنوم نرمال گرداند. یعنی اگر دومینی بر روی سه پروتئین مختلف وجود داشته باشد اما در مطالعه آن‌ها در دو دومین از این سه دومین جهش داشته باشد، تعداد جهش‌ها تقسیم بر مجموع طول دو دومین جهش خورده شده است. به علاوه نرت و همکاران [۷] از ژن‌هایی که ژن منتخب مطالعات قبلی بودند برای یافتن دومین‌ها استفاده کرده‌اند. بنابر این، اگر دومینی روی این ژن‌ها نبوده ولی در مطالعه کل ژنوم بر روی سایر ژنها جهش‌های معنادار داشته باشد، تعداد جهش‌ها از مطالعه آن‌ها حذف شده است. به علاوه نرت و همکاران [۷] پروتئین‌ها و دومین‌هایی که کمتر از ۱۵۰ اسید‌آمینه طول داشتند را از لیست دومین مطالعه حذف کرده‌اند. آن‌ها از

نرخ کشف کاذب ۰.۰ استفاده کرده اندکه این نرخ سطح معناداری بسیار کمی دارد و در مطالعات آماری معمولاً توصیه می شود از حداقل سطح معناداری ۰/۰۵ استفاده شود. در مقایسه نتایج این مطالعه با نتایج مطالعه یانگ و همکاران [۲۲] باید گفت دومین P53 در سرطان ریه، پستان، کبد، تخمدان، دو دومین Pkinase-Tyr، Ras در سرطان ریه، دومین Ras در سرطان پانکراس و دومین MH2 جز دومین های معنادار بودند که در مشاهدات این مطالعه نیز جز دومین های معنادار این سرطانها هستند. دومین P53-tetramer که در مطالعه یانگ و همکاران جز دومین های معنادار سرطان کبد شناخته شده بود، در این مطالعه معنادار نبود.

۷.۴ نتیجه‌گیری و مقایسه دومین‌های ساختاری و دنباله‌ای

مطالعات انجام شده اختصاصی بر روی ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده نشان داد که علی‌رغم تاکید زیادی که بر نقش این ژن‌ها در ایجاد و توسعه سرطان می شود بیشتر این ژن‌ها هیچ جهش معناداری نداشتند، حتی دومین‌های پروتئین‌های (کت و پی‌فم) متناظر کدشده از روی بیشتر این ژن‌ها هم جهش‌های معنادار نداشتند. به طوری که از ۱۸۲ ژن سلول‌های بنیادی ۱۲۴ ژن هیچ جهش معناداری در مطالعه ژن‌ها و دومین‌ها (کت و پی‌فم) نداشتند. همین‌طور ۱۱۳ ژن ترمیم کننده نیز هیچ جهش معناداری بر روی ژن یا دومین (کت و پی‌فم) نداشتند. از بین ۷۶۸۵ بیمار مورد مطالعه در مجموع به جز دو بیمار که هیچ جهش غیرخاموش بر روی مناطق کدکننده پروتئین نداشتند، بقیه بیماران حداقل یک جهش غیر خاموش بر روی این نقاط داشتند. از این بین ۷۶۷۴ نفر دارای جهش بر روی دومین‌های پی‌فم و ۷۴۱۸ نفر دارای جهش بر روی دومین‌های کت بودند. دومین‌های کاندیدی که برای مجموعه دومین‌های کت معرفی شده‌اند ۶۱۰۸ نفر را پوشش می‌دهند. یعنی این افراد حداقل یک جهش بر روی یکی از دومین‌های کاندید کت دارند. برای دومین‌های پی‌فم این میزان ۷۰۴۱ نفر است یعنی ۹۳٪ بیماران مورد مطالعه. شکل ۱.۴ میزان پوشش بیماران توسط دومین‌های معنادار برای دو مجموعه توسط کت و پی‌فم را نشان می‌دهد.

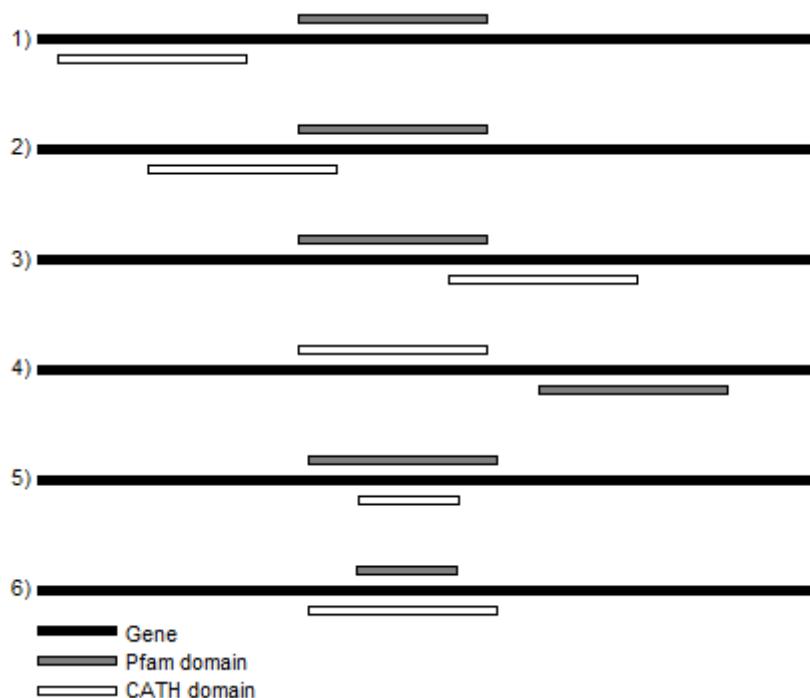
با توجه به مطالعه انجام شده و نتایج حاصل از آن باید گفت که در این روش برای مشخص کردن دومین‌های دخیل در سرطان از کل ژن‌های کدکننده پروتئین استفاده شده است، و نتیجه کلی پوشش ۹۲٪ از کل بیماران مورد مطالعه است. این درصد پوشش بیماران، توانایی این روش در شناسایی دومین‌های کاندید در سرطان‌ها را نشان می‌دهد. تنها کمتر از هشت درصد بیماران مورد مطالعه شناسایی نشدند که می‌توان آن را به عوامل دیگری مانند تغییرات کپی، تغییرات متیلاسیون دی‌ان‌آ یا نقص کارکرد آر-ان‌آ دانست که نیاز به مطالعه بیشتری دارد. معرفی تعداد دومین‌های کاندید مناسب یعنی حدود ۱۲٪ در مقابل پوشش ۹۲٪ بیماران دست آورد خوبی است و نشان از موفق بودن روش استفاده شده دارد.



شکل ۱.۴: نمودار درصد بیماران دارای حداقل یک دومین کاندید کت و پی‌فم در سرطان‌ها

شکل ۱.۴ درصد بیمارانی که توسط دومین‌های کاندید کت و پی‌فم پوشش داده می‌شود را نشان می‌دهد. در همه بیماری‌ها به جز بیماری سرطان رحم که هر دو مجموعه دومین تعداد یکسانی بیمار را پوشش می‌دهند. مجموعه پی‌فم تعداد افراد بیشتری را تحت پوشش دارد. در مورد بیماری سلول‌های زابشی بیضه مجموعه دومین کت بسیار ضعیف بوده که شاید به خاطر دومین شناخته شده ساختاری کت برای پروتئین‌هایی باشد که در این سرطان دخیل‌اند و یا به خاطر وجود عوامل موثر دیگر در ایجاد این سرطان مانند تغییرات تعداد کپی باشد. در مطالعات مشابه انجام شده پیشین، میزان شباهت کل دومین‌های اسکاپ و پی‌فم با هم مقایسه شده بود. اما در اینجا تنها دومین‌های انسانی دو بانک اطلاعاتی کت (ساختار محور) و پی‌فم (توالی محور) با هم مقایسه شده است. از آنجایی که بخش اصلی مطالعه بر پایه نقش دومین‌ها در ایجاد و توسعه سرطان‌هاست، و چون تعداد دومین‌های ساختاری برای ژن‌های انسانی اندک است، باید از دومین‌های پیش‌گویی شده توسط نرمافزارهای مختلف استفاده کرد. در اینجا از مجموعه دومین‌پی‌فم که مجموعه دومین توالی محور است، برای مطالعه اصلی استفاده شده است. مجموعه دومین کت که مجموعه دومین ساختار محور است برای ارزیابی دومین‌های پی‌فم انتخاب شده است.

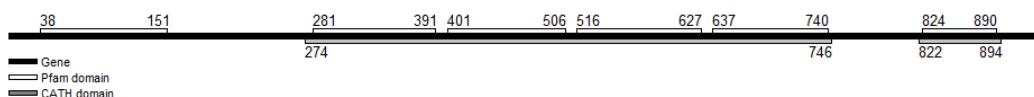
۲۹۹۳ زن، دومین‌های کت و ۱۷۷۲۲ زن، هم دومین‌های پی‌فم دارند. دومین‌های کت روی پروتئوم انسان، شامل ۷۵۹ خانواده هستند و جمماً ۱۰۴۲۷۳۹ اسیدآمینه طول دارند. خانواده‌های دومینی کت بین حداقل یک و حداقل ۲۱۷ تکرار در روی پروتئوم هستند یعنی هر دومینی می‌تواند بر روی پروتئین‌های مختلف وجود داشته باشد. حتی ممکن است یک پروتئین چندین بار در روی یک پروتئین ظاهر شده باشد که میزان این تکرارها برای دومین‌های مختلف کت بین یک تا ۲۱۷ متفاوت است. دومین‌های پی‌فم هم شامل ۶۰۰۹ خانواده با مجموع طول



شکل ۲.۴: حالت‌های ممکن میزان همپوشانی دو دومین کت و پی‌فم

۵۰۳۴۸۵۱ اسیدآمینه و حداقل یک و حداقل تکرار بر روی پروتئوم هستند. برای مقایسه میزان شباهت بین این دو دسته، فقط ژن‌هایی که در هر دو دسته دارای دومین بودند، انتخاب شده‌اند. زیرا تعداد ژن‌های با دومین کت بسیار کمتر از ژن‌های دارای دومین پی‌فم بود. میزان همپوشانی دومین‌های کت و پی‌فم، سه حالت مختلف دارد. حالت اول، دو دومین ممکن است با هم همپوشانی نداشته باشند. حالت دوم، بخشی از دو دومین با هم همپوشانی داشته باشند و حالت سوم، یکی از آن‌ها زیر مجموعه دیگری باشد که در شکل ۲.۴ همه حالت‌های مختلف نشان داده شده است.

برای حساب مقدار همپوشانی دو مجموعه دومین، محوریت بر روی وجود اشتراک بین دو مجموعه دومین بود. یعنی، دومین‌های کت موجود بر روی یک ژن با هم ادغام شده و به عنوان یک دومین کت معروفی شده‌اند. همین طور دومین‌های مختلف پی‌فم موجود بر روی یک ژن هم با هم ادغام شده و به صورت یک دومین از پی‌فم معروفی شده‌اند. برای مثال، برای ژن ACTN2، دو دومین کت دارد که دومین ۱۲۰.۵۸.۶۶ در منطقه (۷۴۶-۲۷۴) و دومین ۱.۱۰.۲۳۸.۱ در منطقه (۸۹۴-۸۲۲) واقع شده است. و پی‌فم هم سه دومین بر روی این ژن دارد که دومین ۶۲۷-۵۱۶، ۵۰۶-۴۰۱، ۳۹۱-۲۷۹ در منطقه (۱۵۱-۳۸)، دومین Spectrin در منطقه (۳۹۱-۲۸۱) و (۱۵۱-۳۸) و دومین EFhand_Ca_insen در منطقه (۸۹۰-۸۲۴) واقع شده است. و دومین سوم هم به نام ۷۴۰-۶۳۷ واقع شده است. برای محاسبه میزان شباهت دومین‌های پی‌فم به دومین‌های کت، فرض شده است که این ژن یک دومین کت در منطقه (۳۹۱-۲۸۱)، دومین پی‌فم در هفت منطقه (۱۵۱-۳۸)، (۸۹۴-۸۲۲) و (۷۴۶-۲۷۴) و یک دومین پی‌فم در منطقه (۳۹۱-۲۸۱)، (۱۵۱-۳۸) واقع شده است. برای محاسبه میزان شباهت دومین‌های پی‌فم به دومین‌های کت، فرض شده است که این ژن یک دومین کت در منطقه (۳۹۱-۲۸۱)، دومین پی‌فم در هفت منطقه (۱۵۱-۳۸)، (۸۹۴-۸۲۲) و (۷۴۶-۲۷۴) و یک دومین پی‌فم در منطقه (۳۹۱-۲۸۱)، (۱۵۱-۳۸) واقع شده است.



شکل ۳.۴: نمایش دومین‌های کت و پی‌فم یک پروتئین

۵۴۶، (۳۹۱-۲۷۹)، (۵۰۶-۴۰۱)، (۶۲۷-۵۱۶)، (۸۹۰-۸۲۴) و (۷۴۰-۶۳۷) دارد. دومین کت در مجموع اسیدآمینه و دومین پی‌فم ۶۱۴ اسیدآمینه طول دارد. همان طور که در شکل ۳.۴ مشخص است دومین کت در شش منطقه (۷۴۶-۲۷۴)، (۳۹۱-۲۸۱)، (۵۰۶-۴۰۱)، (۶۲۷-۵۱۶)، (۷۴۰-۶۳۷) پی‌فم و در منطقه (۸۹۴-۸۲۲) هم دومین پی‌فم در منطقه (۸۹۰-۸۲۴) را پوشش می‌دهد. بنابراین میزان همپوشانی دو دومین کت و پی‌فم در روی این ژن ۵۰۰ اسیدآمینه است. یعنی این محدوده مشترک ۹۲٪ دومین‌های کت و پی‌فم در روی این ژن را پوشش می‌دهد.

محاسبات بالا برای تمامی ژن‌های شامل هر دو نوع دومین کت و پی‌فم انجام شده است. از ۲۹۷۴ پروتئین که شامل هر نوع دومین بودند، ۵۶ پروتئین هیچ همپوشانی بین دومین‌های کت و پی‌فم نداشتند. فقط چهار پروتئین دومین کت و پی‌فم عیناً مشابه داشتند. در ۱۵۷۳ پروتئین، دومین پیشنهادی پی‌فم زیرمجموعه دومین کت بودند و در ۷۷ مورد هم دومین کت زیر مجموعه دومین پی‌فم بود. در ۱۲۷۱ مورد دو دومین همپوشانی داشتند. اما هیچ یک به طور کامل دیگری را تحت پوشش قرار نمی‌داد. میانگین بخش همپوشان دو مجموعه، ۷۹٪ دومین‌های کت و ۷۵٪ دومین‌های پی‌فم را شامل می‌شود. علت استفاده از این روش برای محاسبه میزان همپوشانی دومین‌های کت و پی‌فم به این خاطراست که در مطالعه پیش رو، تنها قرار داشتن اسیدآمینه بر روی دومین دارای اهمیت است. اما مهم نیست این اسیدآمینه در کجا دومین قرار گرفته باشد. با توجه به این میزان همپوشانی دومین‌های کت و پی‌فم، به نظر می‌رسد دومین‌های پی‌فم با هر پی-مقدار و بیت ارزش تولید شده توانایی خوبی در پیدا کردن محل دومین‌های پروتئینی دارند.

۸.۴ معرفی وب سایت

یکی از کارهایی برای بهتر جلوه دادن نتایج این مطالعه انجام شد، نمایش نتایج در قالب یک وب‌سایت بود. آدرس این وب‌سایت <http://mostafa.ac.ir/Canserouspdomains> می‌باشد. این سایت شامل چهار بخش است. یک بخش مربوط به تمامی جداول تولید شده مورد نیاز مطالعه بر روی ژنوم بیماران سرطانی است. بخش دوم، نمایش گرافیکی اطلاعات مربوط جدول‌های بخش یک همراه با امکان جستجو برای ژن، پروتئین است. بخش سوم، جدول‌ها و نمودارهای مربوط به نتایج مطالعه انجام شده است. بخش چهارم هم نمایش گرافیکی نتایج مطالعه است.

۱.۸.۴ بخش اول، جدول‌های مورد نیاز مطالعه

اطلاعات درون جدول‌های موجود در این بخش از وب سایت‌های معتبر بسیاری که قبلاً اسامی همه آن‌ها آمده، تهیه شده است. به دلیل تعداد زیاد ژن‌های کدکننده پروتئین و اطلاعات مربوط به آن‌ها، امکان پیوست این جدول به پایان نامه نبود و فقط از طریق وب‌سایت پایان نامه قابل دسترسی و بارگیری است. همه جداول موجود در این بخش به فرمت .xlxs. قابل بارگیری هستند. اسامی جدول‌های تهیه شده در زیر آمده است.

- ژن‌ها، اسامی و پروتئین‌ها

این جدول اسامی مختلف ژن‌های کدکننده پروتئین انسانی که از یانک‌های اطلاعاتی مختلف جمع‌آوری شده‌اند را نمایش می‌دهد. به علاوه مشخص می‌کند کدام ژن بر روی چه کروموزومی و کدام رشته دی‌ان‌آ قرار دارد و اسم پروتئینی که این ژن می‌سازد چیست. تعداد اگزون‌های هر ژن، محل اگزون‌ها، طول پروتئین آن ژن، اسامی دومین‌های کت و پی‌فم و محل آن‌ها بر روی پروتئین چیست و کجاست. اهمیت این جدول از این نظر است که برای ژن‌ها اسم واحدی وجود ندارد و در مطالعات مختلف اسامی مختلفی برای ژن‌ها بیان می‌شود که گاه‌آین امر باعث گمراهی می‌شود. البته برای برخی مطالعات دانستن این که ژن از روی رشته پیشرو ساخته می‌شود یا پسرو مهم است. زیرا برای ساختن پروتئین از روی ژن‌ها دانستن این امر ضروری است.

- جدول لیست دومین‌های کت

این جدول شامل تمام دومین‌های کت انسانی است. اطلاعاتی که این جدول به نمایش می‌گذارد: اسامی دومین‌های کت پروتئین انسانی، تعداد تکرار هر کدام روی پروتئین‌ها، طول هر تکرار روی پروتئین‌ها و مجموع کل طول دومین‌ها.

- جدول پروتئین‌ها و دومین‌های کت آن‌ها

این جدول جایگاه همه دومین‌های کت پروتئین‌ها برای ژن‌های کدکننده انسانی را به طور مجزا نشان می‌دهد.

- جدول لیست دومین‌های پی‌فم

این جدول شامل تمام دومین‌های پی‌فم انسانی است. اطلاعاتی که این جدول به نمایش می‌گذارد: اسامی دومین‌های پی‌فم پروتئین انسانی، تعداد تکرار هر کدام روی پروتئین‌ها، طول هر تکرار روی پروتئین‌ها، مجموع کل طول دومین‌ها.

• جدول پروتئین‌ها و دومین‌های پی‌فم آن‌ها

این جدول حایگاه همه دومین‌های پی‌فم پروتئین‌ها برای ژن‌های کدکننده انسانی را به طور مجزا نشان می‌دهد.

• ژنوم و مناطق اگزون و اینترون بر روی آن

این جدول نیز برای نمایش تعداد اگزون‌های فعال در پروتئین‌سازی هر ژن، به همراه طول پروتئین، طول اسیدنوكلئوتیدها، ابتدا و انتهای هر ژن و ابتدا و انتهای منطقه کدشونده است. اطلاعات این جدول برای ساختن پروتئین از روی ژن‌ها به کار می‌رود.

• علاوه بر این جدول‌ها، جدول مربوط به طول، اگزون‌ها و مناطق کدکننده پروتئین آن‌ها در چهار جدول مختلف مربوط به پی‌دی‌بی، یوسی‌اس‌سی، رفسک، انسمبل و جدول ادغامی مت Shankل از این چهار جدول که در این مطالعه استفاده شده، اسامی ژن‌های ترمیم کننده و ژن‌های سلول‌های بنیادی استفاده شده در این پایان نامه نیز در این بخش وجود دارد.

لازم به ذکر است که همه این جدول‌ها شامل یک فایل متنی راهنمای جهت استفاده راحت‌تر می‌باشند.

۲.۸.۴ بخش دوم، نمایش گرافیکی اطلاعات

در این بخش امکان نمایش گرافیکی هر ژن، پروتئین و دومین وجود دارد. به این صورت که با وارد کردن نام هر ژن کدکننده پروتئین یا نام پروتئین، نمایش گرافیکی آن ژن، حایگاه اگزون‌ها و اینترون‌ها بر روی آن به همراه نمایش گرافیکی پروتئین آن ژن در صورت وجود، با مشخص بودن حایگاه دومین‌های کت و پی‌فم آن به نمایش در می‌آید.

۳.۸.۴ بخش سوم، جدول‌های تهیه شده از تحلیل نتایج

این بخش شامل جدول‌هایی است که از مطالعه بر روی بیماری‌ها به دست آمده‌اند. این جدول‌ها شامل، جدول مربوط به تعداد جهش‌های مربوط به انواع بیماری‌ها، تعداد بیماران و ژن‌های هر دسته، جدول مربوط به تعداد جهش‌های موجود بر روی دی‌ان‌آ میتوکندری به طور مستقل، جدول مربوط به تعداد جهش‌های مربوط به ژن‌های ترمیم کننده و ژن‌های سلول‌های بنیادی، جدول‌های مربوط به تعداد جهش‌های موجود بر روی هر دومین کت و پی‌فم در همه سرطان‌ها، جدول مربوط به تعداد بیماران تحت پوشش دومین‌های کاندید کت و پی‌فم و جدول‌های معرفی دومین‌های کاندید است. همه جدول‌های موجود در این بخش به فرمت .xlsx قابل بارگیری هستند.

۴.۸.۴ بخش دوم، نمایش گرافیکی نتایج

نمایش گرافیکی این بخش، به نمایش نتایج مطالعه جهش‌های سوماتیک بیماران سرطانی می‌پردازد. به طور عمده از گراف برای نمایش نتایج استفاده شده است. برای مثال در بخش مربوط به نمایش گرافیکی دومین‌های کاندید پی‌فم، از گره‌های گراف برای نمایش سرطان‌ها و دومین‌ها و از یال‌های گراف برای نمایش ارتباط بین سرطان‌ها و دومین‌های کاندید استفاده شده است.

۹.۴ جمع‌بندی

بعد از مطالعه ژن‌های موجود بر روی دی‌ان‌آ میتوکندری مشخص شد، برخی ژن‌های میتوکندری در ایجاد برخی سرطان‌ها نقش دارند اما بدر برخی دیگر از سرطان‌ها جهش‌های موجود بر روی دی‌ان‌آ میتوکندری هیچ تاثیری در ایجاد یا توسعه آن سرطان‌ها ندارند. با مطالعه ژن‌های سلول‌های بنیادی و ژن‌های ترمیم کننده مشخص شد بسیاری از این ژن‌ها نقشی در ایجاد سرطان‌ها ندارند. اما برخی دیگر نقش بسیار پررنگی دارند. بعد از مطالعه جهش‌های موجود بر روی دومین‌های پروتئینی مشخص شد دومین‌هایی وجود دارند که دارای جهش‌های معنادار هستند که در ایجاد سرطان‌ها دخیل‌اند. با مقایسه دومین‌های پروتئینی کت و پی‌فم مشخص شد دومین‌های پی‌فم از قابلیت اطمینان زیادی برخوردارند و تا حد زیادی می‌توان به دومین‌های پی‌فم اعتماد کرد و نتایج مطالعات بر پایه دومین‌های پی‌فم را معتبر شمرد.

و در آخر باید خاطر نشان شد، تمامی مراحل این پایان نامه توسط نرم افزارهای نوشته شده توسط خود شخص انجام شده و از هیچ نرم افزار دیگری برای یافتن ژن‌ها، توالی پروتئین‌ها و دومین‌ها استفاده نشده است. در بسیاری از مطالعات از نرم افزارهای نوشته شده توسط شخص ثالث برای این منظور استفاده می‌شود که به منظور کم کردن زمان مطالعه است. اما غالباً این روش‌ها عاری از خطا نیستند و سایر مطالعات را هم تحت تاثیر قرار می‌دهند. در این مطالعه تنها از اطلاعات وب سایت‌های مختلف که داده‌های کلی را در اختیار قرار می‌دادند استفاده شده و تمامی مراحل جستجو و غربال داده‌ها شخصاً انجام شده است. از طرفی این مطالعه یک مطالعه جامع است که در سطح کل ژنوم انجام شده و تمامی ژن‌های کدکننده پروتئین و دومین‌های شناخته شده تا زمان مطالعه را شامل می‌شود.

جدول ۳.۴: زن‌ها و مجموعه‌های کاندید سرطان‌ها

Cancer	Genes	Complexes
BRCA	MT-RNR1(0.036)	MT-RNR2(0.056) MT-TT(0.008) rRNA(0.084)
LUAD		COMPLEX III(0.011)
UCEC	MT-RNR1(0.117)	MT-TV(0.12) rRNA(0.417)
ACC	MT-TL1(0.326)	
CHOL	MT-RNR2(0.333)	rRNA
ESCA	MT-CYB(0.137)	COMPLEX III(0.137)
HNSC	MT-ND5(0.010)	
KICH	MT-ND1(0.136)	
LIHC	MT-ND5(0.129)	COMPLEX I(0.183)
PRAD	MT-RNR2(0.019)	
SARC	MT-RNR2(0.126)	rRNA(0.142) COMPLEX IV(0.157)
TGCT		COMPLEX III(0.063) rRNA(0.0276)
THYM	MT-RNR2(0.276)	
Pan Cancer	MT-RNR2	rRNA COMPLEX III

جدول ۴.۴: تعداد جهش‌های مربوط به رونوشت‌های نامتعارف میتوکندری

Cancers \ Patents	BC018860	J01415.25	JA040725	JA429504	JA429830
KIRP	1	0	1	0	0
LGG	1	0	0	0	0
LUAD	6	0	1	0	0
UCEC	16	0	94	4	3
ESCA	0	1	0	0	0
KICH	1	0	0	0	0
LIHC	0	1	0	0	0
LUSC	0	0	2	0	0
SARC	0	1	0	0	0
SKCM	0	0	0	0	0
STAD	2	0	0	0	0
UCS	3	0	1	0	0
Pan Cancer	30	3	99	4	3

جدول ۵.۴: ژن‌های سلول‌های بنیادی کاندید سرطان‌ها

Cancer Type	Genes			
BRCA(0.082)	KMT2C(0.069)	PILRB(0.008)	HLA_DRB5(0.007)	
COAD(0.015)	HLA_DPA1(0.015)			
GBM(0.044)	CHEK2(0.018)	TSHZ2(0.025)		
KIRC(0.111)	DNMT3B(0.031)	CHEK2(0.022)	RRAS2(0.018)	NREP(0.007)
	FYB(0.029)	SMARCC2(0.031)	RCSD1(0.020)	HLA_DRB5(0.018)
	TNFSF10(0.013)			
KIRP(0.071)	CHEK2(0.059)	DPH3(0.012)		
LGG(0.093)	CHEK2(0.039)	HDAC2(0.017)	ZBTB20(0.046)	
LUAD(0.420)	SPDL1(0.018)	CHEK2(0.072)	TRPC4(0.072)	CDH6(0.072)
	KMT2C(0.178)	PILRB(0.022)	TSHZ2(0.068)	NPR3(0.046)
	GIMAP1(0.022)	FYB(0.055)		
OV(0.026)	BOD1(0.009)	HAS2(0.017)		
UCEC(0.097)	ATP11C(0.097)			
ACC(0.565)	HDAC2(0.054)	ERCC2(0.207)	GARS(0.380)	PRR34(0.087)
BLCA(0.303)	CHEK2(0.061)	ERCC2(0.097)	KMT2C(0.209)	
CHOL(0.333)	CHEK2(0.083)	KMT2C(0.250)	GIMAP8(0.028)	
ESCA(0.093)	NREP(0.022)	BRINP1(0.077)		
HNSC(0.204)	CHEK2(0.038)	LIN28B(0.015)	BRINP1(0.031)	KMT2C(0.120)
	NPR3(0.027)			
KICH(0.212)	DIMT1(0.015)	KMT2C(0.136)	HLA_DRB5(0.076)	HLA_DQA1(0.030)
LIHC(0.050)	HTR7(0.050)			
PAAD(0.573)	CHEK2(0.170)	BBS9(0.094)	GARS(0.058)	SLC24A1(0.094)
	KMT2C(0.170)	SMARCC2(0.135)	NPR3(0.088)	AFTP(0.135)
PCPG(0.143)	CHEK2(0.051)	NUSAP1(0.040)	KMT2C(0.051)	HLA_DRB5(0.011)
PRAD(0.089)	CHEK2(0.035)	KMT2C(0.054)		
SARC(0.043)	ZNF788(0.028)	BRINP1(0.020)		
SKCM(0.373)	GDF3(0.080)	CCDC90B(0.040)	CDH6(0.107)	KMT2C(0.160)
	GIMAP5(0.067)	GIMAP7(0.067)	GIMAP1(0.067)	GIMAP8(0.120)
STAD(0.323)	CHEK2(0.054)	SOHLH2(0.044)	BRINP1(0.059)	KMT2C(0.165)
	TSHZ2(0.072)	ZBTB20(0.090)		
TGCT(0.028)	C10orf128(0.021)	HLA_DRB5(0.021)	HLA_DQA1(0.014)	
THCA(0.048)	CHEK2(0.014)	GDF3(0.008)	RIOK2(0.008)	HLA_DRB5(0.017)
THYM(0.057)	CHEK2(0.057)			
UCS(0.158)	CHEK2(0.070)	KMT2C(0.105)		
UVM(0.138)	CHEK2(0.075)	NUSAP1(0.050)	HLA_DRB5(0.050)	
PanCancer(0.243)	CHEK2(0.041)	SOHLH2(0.010)	BRINP1(0.022)	TRPC4(0.024)
	CDH6(0.021)	KMT2C(0.106)	PILRB(0.008)	HLA_DRB5(0.011)
	TSHZ2(0.026)	NPR3(0.017)	GIMAP1(0.009)	GIMAP8(0.017)
	ZBTB20(0.021)			

جدول ۴.۶: ژن‌های ترمیم کننده کاندید سرطان‌ها

Cancer	Genes			
BRCA(0.334)	TP53(0.334)			
COAD(0.520)	TP53(0.520)			
GBM(0.287)	TP53(0.287)			
KIRC(0.109)	FANCE(0.044)	DDB1(0.049)	RPA1(0.022)	TP53(0.042)
	CHEK2(0.022)			
KIRP(0.172)	OGG1(0.024)	MSH3(0.041)	TDG(0.036)	TP53(0.041)
	CHEK2(0.059)			
LGG(0.504)	TP53(0.480)	CHEK2(0.039)		
LUAD(0.578)	ERCC5(0.033)	TP53(0.547)	CHEK2(0.072)	
OV(0.848)	TP53(0.848)			
UCEC(0.347)	MSH4(0.073)	TP53(0.290)		
ACC(0.370)	MSH3(0.065)	TP53(0.196)	ERCC2(0.207)	
BLCA(0.636)	ATM(0.136)	TP53(0.498)	ERCC2(0.097)	CHEK2(0.061)
CHOL(0.222)	TP53(0.139)	CHEK2(0.083)		
ESCA(0.879)	TP53(0.879)			
HNSC(0.716)	TP53(0.712)	CHEK2(0.038)		
KICH(0.333)	TP53(0.333)			
LIHC(0.322)	TP53(0.322)			
LUSC(0.792)	TP53(0.792)			
PAAD(0.842)	ERCC3(0.088)	XPC(0.099)	WRN(0.146)	TDG(0.099)
	FAN1(0.099)	EME2(0.111)	TP53(0.673)	CHEK2(0.170)
PCPG(0.177)	FANCD2(0.051)	ERCC8(0.011)	TDG(0.074)	CHEK2(0.051)
PRAD(0.193)	ATM(0.045)	TP53(0.108)	POLI(0.014)	CHEK2(0.035)
READ(0.672)	TP53(0.672)			
SARC(0.370)	TP53(0.370)			
SKCM(0.227)	BLM(0.067)	MPG(0.040)	TP53(0.107)	CHEK2(0.093)
STAD(0.579)	UVSSA(0.044)	SLX4(0.067)	TP53(0.499)	CHEK2(0.054)
THCA(0.045)	SMUG1(0.008)	TDG(0.022)	TP53(0.008)	CHEK2(0.014)
THYM(0.057)	CHEK2(0.057)			
UCS(0.912)	TP53(0.912)			
UVM(0.200)	FANCD2(0.063)	CCNH(0.025)	TDG(0.037)	CHEK2(0.075)
Pan Cancer(0.452)	ATM(0.055)	TDG(0.017)	TP53(0.391)	CHEK2(0.040)

جدول ۷.۴: ژن‌های سلول‌های بنیادی با حداقل یک دومین کاندید کت در هر سرطان

Cancers	Genes					
BRCA	CHEK2	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	
COAD	CHEK2	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	
GBM	CCNB1	CCNA2				
KIRC	SKP2	TRIM24				
KIRP	CHEK2	DPH3				
LUAD	CHEK2	KDELC1	IL10RA	PTPRC	HLA-DRB5	HLA-DPA1
UCEC	CHEK2	CCNB1	CCNA2	SKP2	KDELC1	TRNT1
	HLA-DRB5	HLA-DPA1	IL10RA			
BLCA	CHEK2	CCNB1	RRAS2	CCNA2	TRIM24	RAB3B
	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	TARDBP	
CHOL	BMPR1A	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	
ESCA	CHEK2					
HNSC	CHEK2	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	
LIHC	CHEK2	CCNB1	CCNA2	KDELC1	IL10RA	HLA-DRB5
	HLA-DPA1					
LUSC	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1		
PAAD	CHEK2	TARDBP				
PCPG	RRAS2	RAB3B	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1
PRAD	CHEK2					
READ	CHEK2					
SARC	CHEK2	CCNB1	CCNA2	KDELC1	IL10RA	HLA-DRB5
	HLA-DPA1					
SKCM	CHEK2	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1	
STAD	CHEK2	SKP2	TRIM24	KDELC1	IL10RA	PTPRC
	HLA-DPA1	HLA-DRB5				
TGCT	KDELC1	IL10RA	HLA-DRB5	HLA-DPA1		
THCA	CHEK2	RRAS2	RAB3B			
THYM	CHEK2	CTSS				
UVM	LIN28B					
Pan Cancer	DNMT3B	CHEK2	CCNB1	RRAS2	CCNA2	BMPR1A
	KDELC1	IL10RA	PTPRC	HLA-DRB5	HLA-DPA1	PDE4C
	TRIM24	RAB3B	FGF2	TARDBP	SKP2	

جدول ۸.۴: ژن‌های ترمیم کننده با حداقل یک دومین کاندید کت در هر سرطان

Cancers	Genes						
BRCA	CLK2	CDK7	CHEK1	TP53	CHEK2		
COAD	CLK2	NHEJ1	CDK7	CHEK1	TP53	CHEK2	
GBM	CCNH	TP53					
KIRC	SETMAR	TP53					
KIRP	CLK2	MUTYH	OGG1	CDK7	CHEK1	TP53	CHEK2
LGG	TP53						
LUAD	CLK2	CDK7	CHEK1	TP53	CHEK2		
OV	TP53						
UCEC	CLK2	CCNH	CDK7	CHEK1	TP53	CHEK2	
ACC	TP53						
BLCA	CLK2	RPA2	CCNH	CDK7	WRN	MGMT	CHEK1
	RAD51	TP53	CHEK2	DMC1			
CHOL	TP53						
ESCA	CLK2	CDK7	CHEK1	TP53	CHEK2		
HNSC	CLK2	CDK7	CHEK1	TP53	CHEK2		
KICH	TP53						
LIHC	CLK2	CCNH	CDK7	CHEK1	TP53	CHEK2	
LUSC	TP53						
PAAD	CLK2	REV1	TOPBP1	CDK7	MDC1	POLM	POLL
	CHEK1	LIG4	TP53BP1	BRCA1	LIG3	TP53	XRCC1
	CHEK2						
PCPG	TDG	UNG	RAD51	LIG1	DMC1		
PRAD	CLK2	MAD2L2	CDK7	CHEK1	TP53	CHEK2	
READ	CLK2	NHEJ1	CDK7	CHEK1	TP53	CHEK2	
SARC	CLK2	CCNH	CDK7	CHEK1	TP53	CHEK2	
SKCM	CLK2	CDK7	CHEK1	TP53	CHEK2		
STAD	CLK2	CDK7	CHEK1	TP53	CHEK2		
THCA	CLK2	CDK7	CHEK1	TDG	UNG	RAD51	CHEK2
	DMC1						
THYM	CLK2	CDK7	CHEK1	TP53	CHEK2		
UCS	TP53						
UVM	RPA2	RPA3	TDG	UNG	RPA1	LIG1	
Pan Cancer	CLK2	RPA2	NHEJ1	REV1	SETMAR	TOPBP1	CCNH
	UNG	LIG4	RAD51	TP53BP1	BRCA1	LIG3	TP53
	CDK7	MDC1	POLM	WRN	MGMT	POLL	CHEK1
	XRCC1	CHEK2	DMC1	TDG			

جدول ۹.۴: ژن‌های سلول‌های بنیادی با حداقل یک دومین کاندید پی‌فم در هر سرطان

Cancers	Genes					
BRCA	CDH6					
COAD	RRAS2	RAB3B	GPR176	HTR7	ERCC2	TRPC4
	CDH6	CCR1				
GBM	GPR176	HTR7	CCR1			
KIRC	TRIM24	ORC1	NREP	ADH5		
LUAD	RRAS2	BMPR1A	ROR1	RAB3B	BRINP1	GPR176
HTR7	TRPC4	CDH6	PILRB	GIMAP5	GIMAP6	
CCR1	NPR3	GIMAP1	PTPRC	GIMAP7	GIMAP8	
UCEC	ZNF788	TRPC4	CDH6	ZBTB20		
ACC	LIN28B	FBXO22	ERCC2			
BLCA	RRAS2	ZNF788	RAB3B	ERCC2	CDH6	KMT2C
	ZBTB20					
CHOL	HLA-DPA1	HLA-DQA1				
HNSC	BRINP1	CDH6	KMT2C	NPR3		
KICH	IGF2BP1	TARDBP	HLA-DPA1	HLA-DQA1		
LUSC	ROR1	GPR176	HTR7	CDH6	PTPRC	CCR1
	NPR3					
PAAD	TET1	IGF2BP1	RRAS2	ROR1	RAB3B	TARDBP
	BBS9	KMT2C				
PCPG	RRAS2	RAB3B	CCDC90B	ERCC2		
PRAD	CDH6					
READ	RRAS2	RAB3B	GPR176	HTR7	ERCC2	CCR1
SARC	GPX8					
SKCM	BMPR1A	ROR1	BRINP1	TRPC4	CDH6	PILRB
	GIMAP5	GIMAP6	PTPRC	GIMAP7	NPR3	GIMAP1
	GIMAP8					
STAD	MSH2	MSH6	HESX1	ROR1	FGF2	GPR176
	HTR7	HDX	ALX1	TRPC4	CDH6	PTPRC
	CCR1	NPR3				
TGCT	RRAS2	RAB3B	GPX8	C10orf128	HLA-DPA1	HLA-DQA1
THCA	RRAS2	BMPR1A	ROR1	RAB3B		
UCS	IPO5	KMT2C				
UVM	DDX1	SNX8	KMT2C	HLA-DMB	HLA-DRB5	
Pan Cancer	TET1	RRAS2	BMPR1A	ZNF788	HESX1	ROR1
	GIMAP6	HLA-DMB	PTPRC	HLA-DRB5	HLA-DPA1	GIMAP7
	RAB3B	BRINP1	GPR176	HTR7	HDX	ALX1
	HLA-DQA1	CCR1	NPR3	GIMAP1	GIMAP8	ZBTB20
	TRPC4	HAS2	CDH6	KMT2C	PILRB	GIMAP5

جدول ۱۰.۴: ژن‌های ترمیم کننده با حداقل یک دومین کاندید پی‌فم در هر سرطان

Cancers	Genes					
BRCA	H2AFX	TP53				
COAD	LIG4	TP53	ERCC2			
GBM	TP53					
KIRC	NEIL3	FANCE	RPA1	TP53		
KIRP	PRPF19					
LGG	PMS1	TP53				
LUAD	H2AFX	SERPINA2	TP53			
OV	TP53					
UCEC	TP53					
ACC	UVSSA	TP53	ERCC2			
BLCA	H2AFX	BRIP1	TP53	ERCC2		
CHOL	TP53					
ESCA	TP53					
HNSC	H2AFX	TP53				
KICH	RDM1	TP53				
LIHC	TP53					
LUSC	H2AFX	TP53				
PAAD	PARP1	PARP3	MUS81	FANCM	PARP2	EME2
	ERCC4	EME1	RDM1	TP53		
PCPG	FANCD2	H2AFX	SMUG1	TDG	UNG	BRIP1
	ERCC2					
PRAD	H2AFX	TP53				
READ	TP53	ERCC2				
SARC	TP53					
SKCM	SERPINA2	TP53				
STAD	MSH4	MSH2	MSH6	MSH3	NEIL1	TP53
TGCT	FANCA					
THCA	SMUG1	TDG	UNG			
UCS	SPRTN	TP53				
UVM	FANCD2	DDB1	SMUG1	TDG	UNG	BRCA2
Pan Cancer	SETMAR	H2AFX	SERPINA2	TP53		

پیوست‌ها

پیوست آ

جدول‌ها

جدول آ.۱: اسامی فایل‌های دانلود شده برای بیماری‌های مختلف

Cancer Type	File Name
ACC	broad.mit.edu__IlluminaGA_curated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA_automated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA_curated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__Mixed\curated\DNA\sequencing_level2.maf
BLCA	bcgsc.ca__IlluminaHiSeq\automated\DNA\sequencing_level2.maf
	broad.mit.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	broad.mit.edu__IlluminaGA\curated\DNA\sequencing_level2.maf
BRCA	genome.wustl.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	genome.wustl.edu__IlluminaGA\curated\DNA\sequencing_level2.maf
CHOL	bcgsc.ca__IlluminaHiSeq\automated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__Mixed\curated\DNA\sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	ucsc.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
COAD	hgsc.bcm.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	hgsc.bcm.edu__ABI\SOLiD\DNA\Sequencing_level2.maf
ESCA	bcgsc.ca__IlluminaHiSeq\automated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	ucsc.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
GBM	broad.mit.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
HNSC	bcgsc.ca__IlluminaHiSeq\automated\DNA\sequencing_level2.maf
	broad.mit.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	broad.mit.edu__IlluminaGA\curated\DNA\sequencing_level2.maf
KICH	bcgsc.ca__IlluminaHiSeq\automated\DNA\sequencing_level2.maf
	broad.mit.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	hgsc.bcm.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	hgsc.bcm.edu__Mixed\curated\DNA\sequencing_level2.maf
KIRC	hgsc.bcm.edu__Mixed\DNA\Sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA\curated\DNA\sequencing_level2.maf
	hgsc.bcm.edu__Mixed\curated\DNA\sequencing_level2.maf
	broad.mit.edu__Illumina\Genome\Analyzer\DNA\Sequencing_level2.maf
	broad.mit.edu__IlluminaGA\automated\DNA\sequencing_level2.maf
	broad.mit.edu__IlluminaGA\curated\DNA\sequencing_level2.maf

		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
LGG		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_curated_DNA_sequencing_level2.maf
LIHC		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__Mixed_curated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
LUAD		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_curated_DNA_sequencing_level2.maf
LUSC		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
OV		hgsc.bcm.edu__SOLiD_curated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		genome.wustl.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
PAAD		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_curated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
PCPG		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
PRAD		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_curated_DNA_sequencing_level2.maf
READ		hgsc.bcm.edu__ABI_SOLiD_DNA_Sequencing_level2.maf
		hgsc.bcm.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
SARC		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		genome.wustl.edu__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
SKCM		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
STAD		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_curated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
TGCT		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__Mixed_curated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
THCA		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
THYM		bcgsc.ca__IlluminaHiSeq_automated_DNA_sequencing_level2.maf
		hgsc.bcm.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		ucsc.edu__IlluminaGA_automated_DNA_sequencing_level2.maf
		broad.mit.edu__Illumina_Genome_Analyzer_DNA_Sequencing_level2.maf

UCEC	genome.wustl.edu__Illumina__Genome__Analyzer__DNA__Sequencing_level2.maf
UCSC	bcgsc.ca__IlluminaHiSeq__automated__DNA__sequencing_level2.maf
	broad.mit.edu__IlluminaGA__automated__DNA__sequencing_level2.maf
	broad.mit.edu__IlluminaGA__curated__DNA__sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA__automated__DNA__sequencing_level2.maf
UVM	bcgsc.ca__IlluminaHiSeq__automated__DNA__sequencing_level2.maf
	broad.mit.edu__IlluminaGA__automated__DNA__sequencing_level2.maf
	hgsc.bcm.edu__IlluminaGA__automated__DNA__sequencing_level2.maf
	ucsc.edu__IlluminaGA__automated__DNA__sequencing_level2.maf

جدول آ.۲: اسامی کامل و مخفف انواع سرطان‌ها

Persian Name	abbreviation	Latin Name
سرطان غدد فوق کلیه	ACC	Adrenocortical carcinoma
سرطان مثانه	BLCA	Bladder Urothelial Carcinoma
سرطان مهاجم پستان	BRCA	Breast invasive carcinoma
سرطان مجرای صفراوی	CHOL	Cholangiocarcinoma
سرطان روده بزرگ	COAD	Colon adenocarcinoma
سرطان مری	ESCA	Esophageal carcinoma
تومور مغز	GBM	Glioblastoma multiforme
سرطان سلول سینگرفشی سر و گردن	HNSC	Head and Neck squamous cell carcinoma
سرطان کروموفوب کلیه	KICH	Kidney Chromophobe
سرطان سلول روشن کلیه	KIRC	Kidney renal clear cell carcinoma
سرطان سلول پاپیلاری کلیه	KIRP	Kidney renal papillary cell carcinoma
تومور درجه پایین مغز	LGG	Brain Lower Grade Glioma
سرطان هپاتوسلولار کبد	LIHC	Liver hepatocellular carcinoma
سرطان ریه	LUAD	Lung adenocarcinoma
سرطان سلول سینگرفشی ریه	LUSC	Lung squamous cell carcinoma
سرطان سیروز تخدمان	OV	Ovarian serous cystadenocarcinoma
سرطان پانکراس	PAAD	Pancreatic adenocarcinoma
فتوکروموموستوم و پاراگانگلیوم	PCPG	Pheochromocytoma and Paraganglioma
سرطان پروستات	PRAD	Prostate adenocarcinoma
سرطان رکتوم	READ	Rectum adenocarcinoma
تومور بد خیم بافت پیوندی	SARC	Sarcoma
سرطان ملانومای پوست	SKCM	Skin Cutaneous Melanoma
سرطان معده	STAD	Stomach adenocarcinoma
سرطان سلول‌های زایشی بیضه	TGCT	Testicular Germ Cell Tumors
سرطان تیروئید	THCA	Thyroid carcinoma
سرطان تیموس	THYM	Thymoma
سرطان مخاط رحم	UCEC	Uterine Corpus Endometrial Carcinoma
سرطان رحم	UCS	Uterine Carcinosarcoma
سرطان چشم	UVM	Uveal Melanoma

جدول آ.۳: لینک دانلود داده‌های مربوط به ژنوم ، پروتئین و دومین‌های پروتئینی

download data links	
UCSC	https://genome.ucsc.edu/cgi-bin/hgTables
PDB	http://www.rcsb.org/pdb/gene/+UNIPROT ID*
Pfam	http://pfam.xfam.org/protein/+UNIPROT ID*
CATH	ftp://ftp.biochem.ucl.ac.uk/pub/cath/v4_0_0/CathDomall.v4.0.0
SIFTS	ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/pdb_chain_uniprot.csv.gz
	ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/pdb_chain_pfam.csv.gz
	ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/pdb_chain_cath_uniprot.csv.gz
	ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/uniprot_pdb.csv.gz

جدول آ.۴: اسامی ژن‌های سلول‌های بنیادی منتخب [۱۰]

SCGS genes in the multicellular signaling / immune signaling / cell identity module.			
DNMT3B	TBCE	MTHFD1L	NA
MCM6	RIOK2	ARMC9	MLL3
CDC25A	BCKDHB	XPOT	MXI1
PFAS	RAD1	IARS	FKSG49
MCM4	NREP	HDX	FAM185BP
XRCC5	ADH5	ACTRT3	ARRB2
HAUS6	PLRG1	ERCC2	SMARCC2
TET1	ROR1	TBC1D16	WASH3P
IGF2BP1	RAB3B	GARS	PILRB
PLAA	LOC285431	KIF7	CTSH
DEPDC1B	DBC1	UBE2K	SAT1
TEX10	KIF23	SLC25A3	JUNB
CCDC99	DIAPH3	ICMT	CD53
MSH2	GNL2	UGGT2	PECAM1
BUB1B	FGF2	ATP11C	IL10RA
MSH6	TARDBP	SLC24A1	RCSD1
DLGAP5	NMNAT2	EIF2AK4	ARHGDIIB
SKIV2L2	ZNF167	GPX8	GIMAP5
CENPE	KIF20A	ALX1	GIMAP6
CHEK2	CENPI	OSTC	HLA-DMB
SOHLH2	DDX1	TRPC4	PTPRC
CCNB1	XXYLT1	HAS2	C10orf128
RRAS2	GPR176	FZD2	CMBL
PRIM1	FBXO22	TRNT1	HLA-DRB5
PAICS	BBS9	MMADHC	HLA-DPA1
CCNA2	C14orf166	SNX8	ABCG1
CPSF3	BOD1	CDH6	GIMAP7
NUSAP1	CDC123	HAT1	HLA-DQA1
LIN28B	SNRPD3	SEC11A	TSHZ2
IPO5	FAM118B	DIMT1	RGCC
KIF11	DPH3	TM2D2	CCR1
BMPR1A	EIF2B3	FST	NPR3
NDC80	KDELC1	GBE1	RSAD2
BCAT1	RPF2		GIMAP1
CCNG1	APLP1		TNFSF10
ZNF788	DACT1		AFTP
ASCC3	PDHB		NA
FANCB	C14orf119		MALAT1
MCM10	DTD1		UBXN2A
HMGA2	SAMM50		PDE4C
SKP2	CCL26		GIMAP8
TRIM24	C4orf52		FYB
ORC1	CCDC90B		MS4A7
HDAC2	MED20		C5orf56
HESX1	UTP6		LOC400931
C1orf135	RARS2		MLLT6
INHBE	KIAA0020		CTSS
MIS18A	ARMCX2		ZBTB20
DCUN1D5	RARS		
POLE2	MTHFD2		
MRPL3	DHX15		
CENPH	HTR7		
MYCN	HIST1H4C		
HAUS1			
GDF3			

SCGS genes in the RNA transcription / protein synthesis module.

SCGS genes in the DNA replication / cell cycle module.

جدول آ.5: اسامی ژن‌های ترمیم کننده منتخب [۱۳، ۱۲، ۱۱]

Gene ID					
CLK2	RNF168	TDP2	MMS19	SERPINA2	PER1
DCLRE1B	SETMAR	HUS1	POLL	FANCM	RAD51C
EXO1	TOPBP1	NUDT1	ALKBH3	MLH3	RAD51D
C1ORF86	TREX1	PMS2	ATM	MNAT1	RDM1
MAD2L2	XPC	POLM	CHEK1	PARP2	RECQL5
MSH4	HELQ	RPA3	DDB1	RAD51B	RPA1
MUTYH	NEIL3	SHFM1	DDB2	TDP1	TP53
PARP1	POLN	MPLKIP	FANCF	XRCC3	POLI
RAD54L	RNF4	XRCC2	FEN1	BLM	RBBP8
RPA2	UVSSA	CNOT7	GTF2H1	DUT	ERCC1
SPRTN	CCNH	NBN	H2AFX	FAN1	ERCC2
APLF	CDK7	NEIL2	MRE11A	FANCI	C19ORF40
ERCC3	ERCC8	POLB	MUS81	NEIL1	LIG1
FANCL	GTF2H2	PRKDC	PRPF19	POLG	PNKP
GEN1	MSH3	RAD54B	RAD9A	RAD51	POLD1
MSH2	POLK	RECQL4	ALKBH2	TP53BP1	RAD23A
MSH6	RAD17	RRM2B	GTF2H3	SLX4	XAB2
NHEJ1	RAD50	UBE2V2	NABP2	EME2	XRCC1
PMS1	UBE2B	WRN	POLE	ERCC4	PCNA
REV1	XRCC4	APTX	RAD52	FANCA	SPO11
XRCC5	FANCE	FANCC	RECQL	SLX1B	CHEK2
ATRIP	GTF2H4	FANCG	SMUG1	MPG	DMC1
FANCD2	GTF2H5	RAD23B	TDG	NTHL1	XRCC6
HLTF	MDC1	INSL6	UBE2N	PALB2	APEX2
MBD4	MSH5	XPA	UNG	BRCA1	CETN2
MLH1	POLH	DCLRE1A	BRCA2	BRIP1	FANCB
OGG1	REV3L	DCLRE1C	ERCC5	EME1	RPA4
PARP3	RNF8	ERCC6	LIG4	ENDOV	TREX2
RAD18	SHPRH	MGMT	APEX1	LIG3	UBE2A

جدول آ.۶: اسامی دومین‌های کاندید کت سرطان‌ها

Cancers	Genes			
BRCA(0.773)	1.10.1070.11(0.339)	1.10.220.60(0.003)	1.10.437.10(0.017)	1.10.510.10(0.258)
	2.170.260.10(0.007)	2.40.250.10(0.020)	2.60.200.10(0.018)	2.60.40.10(0.202)
	2.60.40.1110(0.043)	2.60.40.60(0.045)	2.60.40.720(0.333)	4.10.365.10(0.007)
COAD(0.911)	1.10.10.440(0.026)	1.10.1070.11(0.268)	1.10.287.450(0.100)	1.10.30.10(0.059)
	1.10.510.10(0.502)	1.10.565.10(0.182)	2.130.10.10(0.175)	2.60.200.10(0.152)
	2.60.40.10(0.405)	2.60.40.1110(0.063)	2.60.40.720(0.524)	3.10.20.230(0.037)
GBM(0.724)	3.10.50.10(0.056)	3.40.190.10(0.089)	3.40.50.1370(0.037)	3.40.50.2300(0.045)
	1.10.1070.11(0.116)	1.10.150.210(0.015)	1.10.472.10(0.062)	1.20.5.50(0.029)
	2.10.220.10(0.247)	2.60.40.1110(0.287)	2.60.40.720(0.276)	3.30.505.10(0.105)
KIRC(0.683)	3.40.718.10(0.051)			
	1.10.1070.11(0.051)	1.10.750.10(0.519)	1.20.1280.50(0.033)	1.20.245.10(0.035)
	1.20.5.900(0.009)	1.20.920.10(0.200)	2.10.220.10(0.058)	2.10.310.10(0.018)
	2.60.40.1120(0.029)	2.60.40.200(0.024)	2.60.40.720(0.033)	3.30.300.30(0.024)
KIRP(0.580)	2.130.10.130(0.040)	2.170.270.10(0.062)	2.40.50.120(0.033)	2.60.40.1110(0.047)
	3.30.70.340(0.013)	3.40.1500.10(0.013)	4.10.75.10(0.038)	
	1.10.1670.10(0.024)	1.10.510.10(0.491)	2.60.40.1110(0.030)	2.60.40.1230(0.024)
	2.60.40.720(0.041)	3.10.660.10(0.012)	3.20.20.330(0.036)	3.30.1250.10(0.018)
LGG(0.896)	3.40.980.10(0.018)	3.90.280.10(0.018)	3.90.940.10(0.012)	
	1.10.1070.11(0.089)	1.10.950.10(0.002)	2.10.220.10(0.076)	2.60.40.1110(0.050)
LUAD(0.949)	2.60.40.720(0.476)	3.30.1370.10(0.033)	3.40.718.10(0.774)	
	1.10.101.10(0.040)	1.10.1070.11(0.103)	1.10.167.10(0.090)	1.10.490.10(0.061)
	1.10.510.10(0.783)	1.20.1250.10(0.136)	1.20.58.60(0.051)	1.20.930.40(0.105)
	2.140.10.30(0.083)	2.170.8.10(0.040)	2.30.39.10(0.244)	2.40.20.10(0.061)
	2.60.60.20(0.048)	2.70.98.20(0.042)	3.10.100.10(0.149)	3.10.200.10(0.072)
	3.40.190.10(0.077)	3.40.50.1820(0.233)	3.90.190.10(0.259)	4.10.75.10(0.024)
	2.10.25.10(0.143)	2.10.70.10(0.127)	2.110.10.10(0.051)	2.120.10.80(0.092)
	2.40.50.40(0.092)	2.60.40.10(0.556)	2.60.40.1120(0.048)	2.60.40.720(0.495)
	3.15.10.10(0.029)	3.20.20.100(0.108)	3.30.300.30(0.061)	3.30.450.50(0.017)
OV(0.800)	1.10.287.650(0.009)	2.60.40.720(0.800)	3.30.450.40(0.009)	
UCEC(0.980)	1.10.1070.11(0.556)	1.10.150.60(0.073)	1.10.1520.10(0.032)	1.10.1540.10(0.056)
	1.20.1280.50(0.056)	1.20.5.110(0.028)	1.25.10.10(0.266)	2.10.310.10(0.028)
	2.60.40.10(0.516)	2.60.40.1110(0.677)	2.60.40.720(0.294)	3.10.20.230(0.044)
	1.10.3090.10(0.048)	1.10.472.10(0.153)	1.10.510.10(0.577)	1.10.630.10(0.206)
	2.130.10.10(0.258)	2.140.10.30(0.121)	2.30.39.10(0.202)	2.60.210.10(0.137)
	3.30.1330.10(0.044)	3.30.505.10(0.395)	3.30.890.10(0.036)	3.40.50.10140(0.069)
ACC(0.315)	4.10.230.10(0.016)	4.10.280.10(0.073)		
	1.10.1200.10(0.022)	1.10.220.60(0.011)	1.10.405.10(0.022)	1.20.120.20(0.087)
	2.60.20.10(0.054)	2.60.40.720(0.174)	3.30.429.10(0.033)	
BLCA(0.964)	1.10.10.10(0.192)	1.10.1070.11(0.269)	1.10.150.210(0.029)	1.10.150.50(0.080)
	1.10.565.10(0.260)	1.20.1260.10(0.027)	1.20.920.10(0.129)	1.25.40.20(0.148)
	2.60.40.1110(0.032)	2.60.40.720(0.488)	3.30.450.50(0.029)	3.30.70.330(0.257)
	1.10.150.60(0.041)	1.10.1820.10(0.010)	1.10.472.10(0.153)	1.10.510.10(0.660)
	2.10.220.10(0.197)	2.130.10.10(0.129)	2.30.30.190(0.061)	2.60.40.10(0.561)
	3.40.50.10300(0.029)	3.40.50.300(0.517)		
CHOL(0.722)	1.10.1170.10(0.111)	1.10.246.10(0.083)	1.20.1260.10(0.083)	2.10.60.10(0.056)
	2.60.40.10(0.500)	2.60.40.1110(0.028)	2.60.40.720(0.139)	3.40.718.10(0.139)
	4.10.720.10(0.028)			
ESCA(0.962)	1.10.1070.11(0.126)	1.10.150.210(0.022)	1.10.3630.10(0.022)	1.10.510.10(0.747)
	1.20.930.40(0.099)	1.25.10.10(0.220)	1.25.40.20(0.165)	2.10.220.10(0.132)
	2.10.310.10(0.016)	2.60.200.10(0.093)	2.60.40.1110(0.044)	2.60.40.720(0.835)
	3.30.70.940(0.016)			
	1.10.1070.11(0.219)	1.10.246.10(0.036)	1.10.506.10(0.019)	1.10.510.10(0.519)
	1.25.40.20(0.248)	2.10.25.10(0.107)	2.130.10.10(0.115)	2.140.10.30(0.061)

HNSC(0.927)	2.60.40.10(0.469)	2.60.40.1110(0.027)	2.60.40.720(0.664)	2.60.60.20(0.031)
	3.10.100.10(0.090)	3.30.450.50(0.015)		
KICH(0.606)	1.10.630.10(0.136)	1.20.58.120(0.030)	2.60.40.1110(0.091)	2.60.40.1150(0.030)
	2.60.40.720(0.303)	3.10.490.10(0.015)	3.30.500.10(0.212)	3.40.50.1580(0.045)
	3.90.470.20(0.015)	4.10.40.20(0.015)		
LIHC(0.866)	1.10.246.10(0.114)	1.10.472.10(0.104)	1.10.510.10(0.609)	1.10.640.10(0.030)
	2.60.40.1110(0.050)	2.60.40.1260(0.030)	2.60.40.720(0.292)	3.30.9.10(0.035)
	1.25.10.10(0.282)	2.10.310.10(0.015)	2.40.10.230(0.020)	2.60.40.10(0.545)
	3.40.50.2300(0.045)	3.90.1230.10(0.059)		
LUSC(0.933)	1.10.1070.11(0.230)	1.20.930.40(0.118)	1.25.40.20(0.197)	2.120.10.80(0.067)
	2.140.10.30(0.124)	2.60.40.10(0.669)	2.60.40.1110(0.079)	2.60.40.720(0.730)
	3.10.100.10(0.185)	3.30.300.30(0.045)	3.30.450.40(0.022)	
PAAD(0.912)	1.10.10.440(0.047)	1.10.150.210(0.023)	1.10.418.10(0.170)	1.10.437.10(0.158)
	1.20.120.560(0.076)	1.20.5.510(0.006)	1.20.5.900(0.006)	1.25.40.20(0.333)
	1.10.510.10(0.643)	1.10.630.10(0.205)	1.10.8.60(0.088)	1.20.120.310(0.023)
	2.60.20.10(0.041)	2.60.200.10(0.181)	2.60.40.720(0.626)	2.70.98.20(0.058)
	3.30.70.330(0.257)	3.40.220.10(0.064)	3.40.50.10190(0.099)	3.90.1170.10(0.023)
PCPG(0.429)	1.10.30.10(0.017)	1.10.3260.10(0.017)	1.10.750.10(0.017)	1.25.40.210(0.011)
	2.30.90.10(0.006)	2.60.40.10(0.177)	3.30.500.10(0.069)	3.40.470.10(0.057)
	3.40.50.300(0.171)			
PRAD(0.442)	1.10.1070.11(0.033)	1.10.1450.10(0.005)	1.10.510.10(0.216)	2.120.10.80(0.014)
	2.20.160.10(0.005)	2.40.320.10(0.007)	2.60.200.10(0.021)	2.60.210.10(0.118)
	2.60.40.1110(0.033)	2.60.40.720(0.099)	3.30.900.10(0.007)	3.40.718.10(0.012)
	4.10.365.10(0.005)			
READ(0.845)	1.10.1070.11(0.103)	1.10.287.450(0.103)	1.10.510.10(0.336)	2.130.10.10(0.164)
	2.60.200.10(0.138)	2.60.40.1110(0.052)	2.60.40.720(0.647)	
SARC(0.764)	1.10.472.10(0.102)	1.10.510.10(0.417)	2.40.50.120(0.043)	2.60.40.10(0.339)
	2.60.40.1110(0.028)	2.60.40.720(0.343)	3.10.200.10(0.067)	3.40.800.10(0.035)
SKCM(0.947)	1.10.510.10(0.920)	1.10.630.10(0.320)	1.10.880.10(0.027)	1.20.930.40(0.120)
	2.10.220.10(0.173)	2.10.70.10(0.200)	2.10.90.10(0.200)	2.30.39.10(0.240)
	2.60.40.770(0.040)	3.30.360.50(0.013)	3.40.190.10(0.107)	3.40.718.10(0.067)
	2.60.20.10(0.080)	2.60.40.10(0.573)	2.60.40.1110(0.107)	2.60.40.720(0.147)
	3.90.1290.10(0.080)			
STAD(0.930)	1.10.10.440(0.044)	1.10.10.60(0.106)	1.10.1070.11(0.235)	1.10.510.10(0.584)
	1.10.565.10(0.227)	1.10.630.10(0.181)	1.20.1270.60(0.065)	1.20.1280.50(0.028)
	2.130.10.10(0.145)	2.30.42.10(0.196)	2.40.20.10(0.054)	2.60.200.10(0.114)
	2.70.98.20(0.067)	3.10.20.230(0.034)	3.30.1120.30(0.031)	3.30.9.10(0.031)
	1.20.58.60(0.088)	1.20.920.10(0.129)	1.25.40.20(0.163)	2.10.220.10(0.160)
	2.60.40.10(0.519)	2.60.40.1110(0.080)	2.60.40.60(0.101)	2.60.40.720(0.473)
	3.40.50.2300(0.047)	3.90.190.10(0.274)		
TGCT(0.278)	1.20.930.40(0.069)	2.60.40.10(0.181)	3.10.25.10(0.028)	3.10.390.10(0.021)
	3.10.50.10(0.042)	3.10.620.10(0.014)	3.30.500.10(0.042)	3.40.1230.10(0.021)
	3.40.50.1370(0.021)			
THCA(0.744)	1.10.510.10(0.629)	1.50.10.20(0.017)	2.60.40.1110(0.006)	3.40.470.10(0.011)
	3.40.50.300(0.160)			
THYM(0.846)	1.10.510.10(0.732)	1.10.630.10(0.244)	1.20.5.290(0.008)	2.60.40.720(0.057)
	3.90.70.10(0.179)			
UCS(0.982)	1.10.1070.11(0.368)	1.10.1520.10(0.035)	1.10.1820.10(0.018)	1.10.3430.10(0.053)
	1.25.10.10(0.281)	2.130.10.10(0.439)	2.130.10.110(0.035)	2.60.210.10(0.088)
	2.60.40.1110(0.175)	2.60.40.720(0.807)	2.60.40.820(0.035)	
UVM(0.512)	1.10.150.210(0.025)	1.25.10.10(0.113)	2.40.50.140(0.125)	3.30.1520.10(0.087)
	3.30.500.10(0.263)	3.40.470.10(0.037)		
	1.10.10.10(0.079)	1.10.10.440(0.009)	1.10.10.60(0.042)	1.10.101.10(0.016)
	1.10.150.50(0.030)	1.10.150.60(0.013)	1.10.1520.10(0.005)	1.10.1540.10(0.015)
	1.10.472.10(0.047)	1.10.490.10(0.025)	1.10.506.10(0.008)	1.10.510.10(0.449)
	1.10.720.50(0.006)	1.10.750.10(0.033)	1.10.800.10(0.016)	1.20.1050.10(0.050)
	1.20.245.10(0.009)	1.20.5.100(0.012)	1.20.5.110(0.005)	1.20.5.50(0.019)

Pan Cancer(0.915)	1.25.10.10(0.086)	1.25.40.20(0.083)	2.10.220.10(0.075)	2.10.25.10(0.067)
	2.130.10.10(0.074)	2.130.10.130(0.031)	2.140.10.30(0.040)	2.170.270.10(0.042)
	2.40.128.20(0.042)	2.40.20.10(0.019)	2.40.250.10(0.004)	2.40.50.40(0.044)
	2.60.210.10(0.028)	2.60.40.10(0.339)	2.60.40.1110(0.067)	2.60.40.1120(0.016)
	2.80.10.50(0.052)	3.10.100.10(0.059)	3.10.20.230(0.009)	3.10.200.10(0.036)
	3.30.1370.10(0.023)	3.30.1490.20(0.021)	3.30.300.30(0.017)	3.30.450.40(0.005)
	3.30.800.10(0.020)	3.30.9.10(0.008)	3.40.190.10(0.037)	3.40.470.10(0.011)
	3.40.50.300(0.257)	3.40.718.10(0.058)	3.90.1170.10(0.004)	3.90.1230.10(0.016)
	1.10.1070.11(0.155)	1.10.1300.10(0.059)	1.10.1380.10(0.023)	1.10.150.210(0.008)
	1.10.167.10(0.037)	1.10.246.10(0.022)	1.10.287.450(0.009)	1.10.437.10(0.023)
	1.10.555.10(0.039)	1.10.565.10(0.107)	1.10.630.10(0.110)	1.10.640.10(0.010)
	1.20.1250.10(0.054)	1.20.1260.10(0.013)	1.20.1280.50(0.012)	1.20.1340.10(0.016)
	1.20.58.60(0.023)	1.20.82.10(0.010)	1.20.920.10(0.066)	1.20.930.40(0.042)
	2.10.310.10(0.005)	2.10.60.10(0.018)	2.10.70.10(0.063)	2.120.10.80(0.019)
	2.170.8.10(0.012)	2.30.30.190(0.012)	2.30.39.10(0.088)	2.30.42.10(0.089)
	2.60.120.200(0.044)	2.60.120.260(0.046)	2.60.20.10(0.025)	2.60.200.10(0.040)
	2.60.40.60(0.023)	2.60.40.720(0.366)	2.60.60.20(0.016)	2.70.98.20(0.024)
	3.10.50.10(0.026)	3.10.620.10(0.004)	3.20.20.100(0.046)	3.20.20.140(0.045)
	3.30.450.50(0.009)	3.30.70.1230(0.009)	3.30.70.1470(0.009)	3.30.70.330(0.120)
	3.40.50.10140(0.015)	3.40.50.10190(0.042)	3.40.50.1370(0.008)	3.40.50.2300(0.015)
	3.90.190.10(0.133)	4.10.280.10(0.011)	4.10.365.10(0.003)	4.10.75.10(0.007)

جدول آ.7: اسامی دومین‌های کاندید پی‌فم سرطان‌ها

Cancers	Domains		
BRCA(0.785)	7tm_4(0.415)	ATP-synt_A(0.008)	Atrophin-1(0.028)
	Cadherin(0.239)	Cytochrom_B_N_2(0.011)	DUF4647(0.013)
	G_path_suppress(0.011)	H-K_ATPase_N(0.003)	Histone(0.073)
	Oxidored_q5_N(0.008)	P53(0.285)	P53_tetramer(0.020)
	PTEN_C2(0.020)	Proton_antipo_M(0.036)	Runt(0.024)
	CBF_beta(0.022)	COX1(0.020)	COX3(0.011)
	FAM219A(0.005)	FRG1(0.013)	GATA(0.040)
	NADH5_C(0.009)	NADHdh(0.016)	Oxidored_q4(0.007)
	PI3K_C2(0.025)	PI3K_P85_iSH2(0.023)	PI3K_p85B(0.010)
	T-box(0.042)	TMEM247(0.011)	Tis11B_N(0.010)
COAD(0.981)	7tm_1(0.651)	7tm_4(0.729)	APC_crr(0.074)
	Arm(APC_u3(0.141)	Cadherin(0.457)	DNA_ligase_IV(0.011)
	DUF504(0.015)	Drf_FH1(0.041)	EPL1(0.175)
	Ion_trans(0.565)	Keratin_2_tail(0.071)	Keratin_B2_2(0.323)
	OGFr_III(0.033)	P53(0.472)	P53_tetramer(0.033)
	APC_u5(0.052)	APC_u9(0.149)	Androgen_recep(0.093)
	DUF1388(0.171)	DUF4483(0.026)	DUF4594(0.045)
	Folliculin_C(0.037)	GLTP(0.059)	HBB(0.033)
	MH2(0.119)	MHC_I(0.119)	NKAP(0.041)
	PI3K_p85B(0.056)	PI3Ka(0.141)	Quaking_NLS(0.011)
	RHIM(0.026)	RRN7(0.011)	Ras(0.576)
	TMEM247(0.030)	Xylo_C(0.045)	
	2-oxogl_dehyd_N(0.011)	7TM_GPCR_Srx(0.011)	7tm_1(0.469)
	Ephrin_lbd(0.047)	Filaggrin(0.044)	Furin-like(0.156)
GBM(0.953)	Metallothi_Euk2(0.007)	P53(0.265)	PARG_cat(0.015)
	7tm_4(0.753)	DSPc(0.160)	DUF1943(0.015)
	GF_recep_IV(0.084)	Iso_dh(0.051)	Keratin_B2_2(0.091)
	PI3K_P85_iSH2(0.055)	PI3K_p85B(0.033)	PI3Ka(0.047)
	PTEN_C2(0.145)	RB_A(0.033)	RB_B(0.015)
	7tm_4(0.499)	ADH_N(0.031)	AMP-binding(0.146)
	BAH(0.071)	Bromodomain(0.142)	C2-set(0.044)
	Connexin43(0.004)	Cystatin(0.035)	Cytochrom_B_N_2(0.013)
	DUF2052(0.018)	DUF2181(0.047)	DUF2363(0.004)
	Dynamitin(0.024)	ERM(0.029)	ETS_PEA3_N(0.027)
HCC(0.953)	FAST_2(0.022)	FA_FANCE(0.042)	GRIM-19(0.004)
	HNH_3(0.009)	HRCT1(0.007)	KRTAP7(0.007)
	MARCKS(0.020)	MCM3AP_GANP(0.027)	MRP-L51(0.011)
	Myosin_N(0.020)	NADHdh(0.018)	NYAP_N(0.031)
	Oxysterol_BP(0.053)	P33MONOX(0.029)	P53(0.031)
	Peptidase_C12(0.049)	Peptidase_M49(0.031)	PhyH(0.022)
	Rap1_C(0.009)	Rcd1(0.013)	Ribosomal_L50(0.018)
	SRI(0.009)	SSXRD(0.031)	STAT_bind(0.058)
	TMEM37(0.011)	TRH(0.018)	TSKS(0.044)
	Recep_L_domain(0.051)	Sp38(0.022)	UDPGT(0.084)
	Aldolase_II(0.035)	Alveol-reg_P311(0.004)	Ant_C(0.029)
	Cation_ATPase_N(0.042)	Chorein_N(0.027)	Clat_adaptor_s(0.040)
	DRY_EERY(0.018)	DUF167(0.007)	DUF2045(0.007)
	DUF3371(0.022)	DUF4594(0.031)	DUF4757(0.018)
	Ependymin(0.009)	FAM104(0.033)	FAM219A(0.018)
	Glyco_hydro_18(0.064)	Got1(0.020)	GpcrRhopsn4(0.020)
	Keratin_2_tail(0.020)	Keratin_B2(0.029)	LAG1-DNAbind(0.013)
	Metallothio(0.040)	Methyltransf_FA(0.020)	Mpv17_PMP22(0.027)
	Neurensin(0.044)	Neuro_bHLH(0.033)	Oxidored_q3(0.018)
	PACT_coil_coil(0.022)	PLDc(0.009)	PTPlike_phytase(0.016)
	Propeptide_C1(0.004)	RBM39linker(0.013)	REPA_OB_2(0.020)

	SHR-BD(0.027)	SOUL(0.047)	SRA1(0.013)
	Scramblase(0.031)	Skp1_PPOZ(0.007)	Sod_Cu(0.024)
	TruB_N(0.018)	U1snRNP70_N(0.011)	UCR_TM(0.007)
	UTP25(0.020)	VHL(0.512)	WAP(0.044)
	dNK(0.044)	zf-RanBP(0.038)	zf-Sec23_Sec24(0.044)
KIRP(0.746)	ANAPC5(0.012)	Acyl-CoA_dh_N(0.053)	Agouti(0.012)
	Coatomer_g_Cpla(0.006)	Cornifin(0.041)	DUF1388(0.059)
	FLYWCH_N(0.012)	GPHH(0.036)	GTPase_binding(0.006)
	Keratin_2_tail(0.030)	Keratin_B2(0.101)	Keratin_B2_2(0.183)
	Prp19(0.012)	RIBIOP_C(0.083)	SNF5(0.024)
	Tau95(0.018)	Telomerase_RBD(0.018)	Thioredoxin_7(0.018)
	Androgen_recep(0.053)	BAGE(0.006)	CLU(0.018)
	DUF4801(0.012)	DUF846(0.219)	DcpS(0.018)
	HNF-1A_C(0.012)	HRCT1(0.030)	IGF2_C(0.018)
	MAPKK1_Int(0.018)	NAC(0.018)	Oxidored_q6(0.018)
	SPATA6(0.030)	SSXT(0.018)	Synapsin_N(0.024)
	Tubulin(0.154)	Ubiq-Cytc-red_N(0.018)	VPS11_C(0.006)
	7tm_4(0.502)	BAGE(0.011)	CCER1(0.007)
	FAM47(0.048)	Furin-like(0.050)	H-K_ATPase_N(0.002)
LGG(0.950)	Keratin_B2(0.041)	Keratin_B2_2(0.085)	Med9(0.007)
	PI3K_P85_iSH2(0.030)	PTEN_C2(0.030)	SAA(0.009)
	CD99L2(0.009)	COX7B(0.009)	DUF3736(0.009)
	HMG_box(0.109)	Iso_dh(0.815)	Keratin_2_tail(0.009)
	P53(0.457)	P53_tetramer(0.017)	PARG_cat(0.013)
	SPAN-X(0.020)	Synapsin_N(0.017)	TMEM252(0.007)
	TRM13(0.007)		
	7tm_1(0.730)	7tm_3(0.191)	7tm_4(0.916)
	CABS1(0.029)	CCER1(0.020)	CLCA(0.127)
	Cadherin_C(0.231)	Cadherin_C_2(0.017)	Cation_ATPase_N(0.066)
F5_F8_type_C(0.167)	Cystatin(0.095)	DUF2668(0.024)	DUF4527(0.018)
	DUF4688(0.040)	DUF676(0.040)	Ephrin_lbd(0.116)
	FAM75(0.092)	FRG1(0.024)	Filaggrin(0.103)
	HIN(0.061)	HNOB(0.033)	Histone(0.207)
	Ig_3(0.334)	Integrin_alpha2(0.246)	Ion_trans(0.626)
	Kringle(0.101)	LCE(0.143)	LRR_5(0.083)
	Lectin_C(0.310)	Lig_chan(0.187)	Lipase(0.147)
	MIEAP(0.022)	Myosin_tail_1(0.394)	NACHT(0.178)
	Neurexophilin(0.064)	Neuromodulin(0.015)	NtCtMGAM_N(0.048)
	PARG_cat(0.031)	PC-Esterase(0.042)	PMG(0.158)
	PaaSYMP(0.013)	Peptidase_M10(0.161)	Peptidase_S41_N(0.037)
	Proton_antipo_M(0.050)	Rap1_C(0.020)	Ras(0.571)
	SOUL(0.033)	SPATIAL(0.029)	SPRR2(0.037)
	Sushi(0.530)	Synapsin_N(0.018)	T4_deiodinase(0.048)
	TSC21(0.017)	TSP_1(0.354)	Ten_N(0.042)
	V-set(0.413)	WTX(0.064)	Xin(0.044)
	AIG1(0.095)	ANF_receptor(0.372)	BAGE(0.022)
	COesterase(0.261)	CUB(0.450)	Cadherin(0.644)
	Collagen(0.473)	Cornifin(0.048)	Crisp(0.018)
	DUF4584(0.022)	DUF4599(0.035)	DUF4636(0.017)
	F5_F8_type_C(0.167)	FAM104(0.026)	FAM47(0.217)
	Forkhead(0.193)	Glyco_hydro_1(0.161)	Glyco_hydro_31(0.163)
	I-set(0.705)	IRK(0.182)	Ig_2(0.391)
	KRTAP(0.105)	Keratin_B2(0.073)	Keratin_B2_2(0.259)
	Laminin_G_2(0.294)	Laminin_G_3(0.088)	Ldl_recept_a(0.253)
	MACPF(0.116)	MAGE(0.229)	MHC_I_3(0.083)
	NMDAR2_C(0.084)	Neur_chan_LBD(0.345)	Neur_chan_memb(0.297)
	OLF(0.150)	P53(0.420)	P53_tetramer(0.055)

	POM121(0.055)	PROL5-SMR(0.029)	PTPlike_phytase(0.017)
	Pkinase_Tyr(0.607)	Pro-rich(0.083)	Protocadherin(0.081)
	Reprolysin(0.255)	Ribosomal_L50(0.020)	RyR(0.075)
	Sarcoglycan_1(0.064)	Serpin(0.409)	Sortilin-Vps10(0.117)
	TB(0.101)	TLD(0.040)	TRP_2(0.033)
	Tsc35(0.017)	UDPGT(0.264)	UNC80(0.028)
	Xlink(0.110)	fn3(0.664)	zf-C2H2_assoc(0.020)
	zf-Sec23_Sec24(0.028)		
OV(0.883)	7tm_4(0.496)	DUF2462(0.004)	DUF4552(0.013)
	MRP-S32(0.009)	NtCtMGAM_N(0.022)	ODAM(0.013)
	P53(0.726)	P53_tetramer(0.048)	PTCRA(0.009)
	Sam68-YY(0.013)	UPF0054(0.009)	
UCEC(0.984)	7tm_4(0.806)	BEX(0.133)	CBFB_NFYA(0.020)
	DUF1992(0.020)	DUF3518(0.113)	DUF3591(0.121)
	NADHdh(0.048)	Oxidored_q3(0.040)	P53(0.278)
	CUB(0.319)	Cadherin(0.532)	DSPc(0.492)
	Ion_trans(0.536)	LRR_5(0.077)	MATH(0.077)
	PI3K_P85_iSH2(0.306)	PI3K_p85B(0.133)	PI3Ka(0.226)
	PTEN_C2(0.383)	Pro-rich(0.157)	Proton_antipo_M(0.113)
ACC(0.978)	7tm_4(0.630)	ATP_bind_3(0.054)	ATP_synt_H(0.043)
	C8(0.228)	COX1(0.130)	Clathrin_lg_ch(0.033)
	DLEU7(0.065)	DUF1388(0.098)	DUF2043(0.022)
	DUF846(0.087)	EBP(0.054)	ESSS(0.011)
	FRG1(0.120)	GGN(0.065)	GLTP(0.207)
	HBB(0.207)	Hydantoinase_B(0.076)	IGFBP(0.098)
	Keratin_B2_2(0.467)	LYRIC(0.043)	MANEC(0.098)
	Nop53(0.087)	OGFr_III(0.109)	P53(0.109)
	Proton_antipo_M(0.163)	RGM_N(0.098)	SPATA9(0.033)
	zf-C2H2(0.472)	zf-C2H2_6(0.319)	
	An_peroxidase(0.261)	Apolipoprotein(0.087)	B-block_TFIIC(0.011)
	Cornifin(0.043)	Cytochrom_B_C(0.022)	Cytochrom_B_N_2(0.065)
	DUF4639(0.152)	DUF4644(0.033)	DUF504(0.022)
	FAM47(0.109)	FAM86(0.065)	FIST_C(0.022)
	GRIN_C(0.120)	Galactosyl_T(0.174)	Glyco_hydro_39(0.163)
	Kelch_3(0.065)	Keratin_2_head(0.152)	Keratin_2_tail(0.033)
	MHC_I(0.130)	Menin(0.054)	NADH5_C(0.043)
	P53_tetramer(0.043)	PMG(0.076)	PRY(0.109)
	SRCR(0.228)	Syntaxin_2(0.152)	TMEM247(0.217)
	TSP_1(0.337)	Ubiq-Cytc-red_N(0.130)	zf-CCHC(0.022)
	4F5(0.017)	7tm_4(0.854)	ASXH(0.041)
	Cadherin(0.600)	DEAD_2(0.063)	DUF3518(0.044)
	Ets(0.102)	FRG1(0.036)	FYRN(0.022)
	HAT_KAT11(0.078)	Hamartin(0.053)	Histone(0.284)
	MT-A70(0.039)	MTP18(0.007)	P53(0.427)
	RB_B(0.034)	Ras(0.481)	Rb_C(0.049)
	SIP1(0.041)	SMC_Nse1(0.019)	SPT6_acidic(0.007)
BLCA(0.983)	THRAP3_BCLAF1(0.085)	TPR_17(0.017)	Tis11B_N(0.022)
	zf-C2H2(0.585)	zf-C2H2_6(0.376)	zf-CCCH(0.063)
	CDI(0.029)	CK_II_beta(0.010)	CMS1(0.019)
	DUF3643(0.015)	DUF4547(0.015)	Epiglycanin_TR(0.044)
	Filaggrin(0.070)	Furin-like(0.109)	GPHH(0.036)
	Hydrolase(0.112)	K167R(0.075)	Keratin_B2(0.068)
	P53_tetramer(0.036)	PI3Ka(0.180)	RB_A(0.053)
	Ribosomal_S13(0.024)	RsfS(0.012)	SAM_PNT(0.049)
	STAG(0.029)	Slu7(0.022)	Synapsin_N(0.024)
	UMPH-1(0.027)	WTX(0.051)	Wyo sine_form(0.017)
	6PGD(0.056)	ANAPC3(0.056)	ARA70(0.056)
	COX3(0.056)	Cys_rich_FGFR(0.111)	Cytochrom_B_C(0.056)

CHOL(0.917)	FRG1(0.111)	GRIM-19(0.028)	G_path_suppress(0.028)
	MHC_II_alpha(0.083)	NDUFB10(0.028)	Oxidored_q2(0.056)
	Phostensin(0.056)	Proton_antipo_M(0.111)	RIBIOP_C(0.083)
	SLBB(0.028)	SMG1(0.056)	SRP54_N(0.056)
	UPF0697(0.028)	ecTbetaR2(0.028)	zinc_ribbon_2(0.028)
	COX1(0.139)	COX2(0.028)	COX2_TM(0.083)
	DUF4578(0.056)	EloA-BP1(0.056)	Epiglycanin_TR(0.083)
	Involucrin(0.111)	Iso_dh(0.194)	Isy1(0.028)
	Oxidored_q3(0.056)	P53(0.139)	PDDEXK_1(0.028)
	Rav1p_C(0.139)	Ribonuclease_3(0.056)	Ribosomal_L44(0.028)
	Thioredox_DsbH(0.056)	UFD1(0.028)	UMPH-1(0.056)
	7tm_4(0.923)	BAGE(0.016)	COX1(0.077)
	Cytochrom_B_C(0.038)	Cytochrom_B_N_2(0.077)	DUF4499(0.016)
ESCA(0.995)	GatB_Yqey(0.022)	Involucrin(0.088)	MHC_I_3(0.082)
	P53(0.775)	P53_tetramer(0.033)	Pro-rich(0.115)
	COX2(0.038)	COX2_TM(0.022)	COX3(0.038)
	ERM(0.082)	FAM153(0.060)	FRG1(0.082)
	Myb_DNA-bind_7(0.016)	NADH5_C(0.027)	NADHdh(0.055)
	Proton_antipo_M(0.071)	RtcB(0.027)	ecTbetaR2(0.022)
	zf-C2H2_7(0.022)		
	2-oxogl_dehyd_N(0.010)	7tm_4(0.830)	ANF_receptor(0.277)
	Cadherin_C(0.103)	CwfJ_C_1(0.013)	DCR(0.011)
	Ependymin(0.017)	Epiglycanin_TR(0.032)	FAM47(0.097)
HNSC(0.989)	Glyco_hydro_31(0.107)	HAT_KAT11(0.065)	Histone(0.233)
	Lipase(0.080)	MACPF(0.055)	MHC_I(0.090)
	PARG_cat(0.032)	PI3K_P85_iSH2(0.025)	PI3Ka(0.116)
	Synapsin_N(0.011)	TMEM252(0.013)	TSP_1(0.239)
	Ank_5(0.183)	ArgoMid(0.013)	Cadherin(0.517)
	DED(0.044)	DUF2369(0.019)	EGF(0.170)
	FRG1(0.048)	FYRN(0.013)	Filaggrin(0.071)
	Keratin_B2(0.069)	Keratin_B2_2(0.162)	LRR_5(0.050)
	NYD-SP12_N(0.031)	P53(0.601)	P53_tetramer(0.044)
	PRIMA1(0.010)	Pro-rich(0.111)	Sushi(0.324)
	zf-C2H2_7(0.011)	zf-C2H2_assoc(0.011)	zf-HC5HC2H(0.031)
	7tm_4(0.652)	ATP-synt_A(0.061)	Aha1_N(0.015)
	COX3(0.045)	Connexin43(0.015)	Cytochrom_B_C(0.015)
	DUF4482(0.045)	DUF974(0.015)	EBP(0.030)
KICH(0.970)	FAD_binding_2(0.061)	FRG1(0.167)	FRG2(0.061)
	HNF-1_N(0.030)	Hamartin(0.030)	Keratin_B2_2(0.227)
	MIP(0.091)	Med8(0.015)	Methyltransf_28(0.015)
	P53(0.273)	P53_TAD(0.015)	P53_tetramer(0.030)
	Pro-rich(0.152)	Proton_antipo_M(0.197)	Proton_antipo_N(0.030)
	Ribosom_S12_S23(0.015)	Ribosomal_L5_C(0.015)	Ripply(0.030)
	TRC8_N(0.076)	Swi5(0.015)	TAS2R(0.242)
	CCDC74_C(0.030)	COX1(0.076)	COX2(0.015)
	Cytochrom_B_N_2(0.061)	DDDD(0.015)	DUF4456(0.015)
	ER_lumen_recept(0.030)	Eaf7(0.015)	Epiglycanin_TR(0.091)
	Forkhead(0.197)	Gal-3-0_sulfotrl(0.076)	HGAL(0.015)
	MCD(0.015)	MHC_I(0.212)	MHC_II_alpha(0.030)
	NADHdh(0.136)	NAD_Gly3P_dh_N(0.015)	Oxidored_q5_N(0.045)
	PK_C(0.015)	PRAS(0.015)	PTEN_C2(0.091)
	RDD(0.015)	RRM_1(0.318)	Rb_C(0.015)
	Snakin_Pallidin(0.015)		
	7tm_4(0.827)	AKAP2_C(0.025)	ALS2CR8(0.025)
	Calpain_u2(0.010)	Connexin43(0.010)	Cytochrom_B_N_2(0.040)
	FRG1(0.089)	Keratin_B2_2(0.272)	LCE(0.119)
	Oxidored_q4(0.030)	P53(0.277)	Proton_antipo_M(0.139)

LIHC(0.950)	Androgen_recep(0.035)	COX1(0.045)	COX2_TM(0.020)
	DUF2363(0.020)	DUF846(0.069)	Epiglycanin_TR(0.069)
	NADH5_C(0.040)	Orn_DAP_Arg_deC(0.074)	Oxidored_q3(0.040)
	RB_A(0.045)	SAA(0.030)	Serum_albumin(0.109)
	TAS2R(0.188)		
LUSC(0.983)	7tm_1(0.843)	7tm_4(0.978)	ANF_receptor(0.404)
	Cadherin(0.697)	Cadherin_C(0.180)	CitMHS(0.056)
	Dppa2_A(0.039)	Histone(0.309)	I-set(0.803)
	Laminin_G_2(0.331)	Lipase(0.185)	Neur_chan_LBD(0.281)
	Sam68-YY(0.028)	Sarcoglycan_1(0.090)	Sushi(0.539)
	APOBEC_C(0.028)	CPSase_L_D3(0.034)	CUB(0.478)
	Collagen(0.528)	DPPIV_N(0.157)	DUF5050(0.062)
	INPP5B_PH(0.028)	Ig_2(0.388)	LRR_5(0.084)
	P53(0.663)	P53_tetramer(0.034)	PTEN_C2(0.084)
	TEX13(0.045)	TSP_1(0.399)	UDPGT(0.247)
	fn3(0.742)		
	2-oxogl_dehyd_N(0.076)	7tm_4(0.889)	AF1Q(0.164)
	Ago_hook(0.094)	Androgen_recep(0.058)	Ank_5(0.199)
	CD52(0.012)	CD99L2(0.117)	CR6_interact(0.035)
K562(0.988)	Cu_amine_oxidN2(0.047)	DDDD(0.053)	DIL(0.082)
	DUF3437(0.035)	DUF3498(0.129)	DUF3508(0.064)
	DUF4607(0.047)	DUF4609(0.006)	DUF4647(0.076)
	Daxx(0.018)	Dppa2_A(0.041)	Drf_FH1(0.088)
	Epiglycanin_TR(0.135)	Exostosin(0.094)	FAM104(0.041)
	FRG2(0.070)	FYRN(0.018)	Fumble(0.105)
	HJURP_C(0.070)	HMMR_C(0.047)	HNH_3(0.023)
	IL17(0.152)	Ig_4(0.076)	Interfer-bind(0.111)
	Keratin_B2_2(0.503)	Kringle(0.175)	LCE(0.345)
	Med12-LCEWAV(0.094)	Med15(0.240)	Med9(0.117)
	NAP(0.275)	NGF(0.088)	NKAP(0.064)
	OGFr_III(0.029)	Ofd1_CTDD(0.041)	Orexin(0.029)
	P68HR(0.023)	PARG_cat(0.058)	PHO4(0.111)
	Phostensin_N(0.070)	Pro-rich(0.164)	Prominin(0.146)
	RGS12_us2(0.035)	RIBIOP_C(0.070)	RNA_polII_A34(0.082)
	RRM_1(0.415)	RRM_6(0.170)	RYDR_ITPR(0.146)
	Ribosomal_L44(0.018)	SNARE_assoc(0.082)	SPAN-X(0.129)
	Synaphin(0.111)	Synapsin_N(0.053)	Synaptobrevin(0.199)
	TMEM132D_C(0.064)	TPP1(0.041)	TPR_1(0.158)
	Tektin(0.123)	Tmemb_185A(0.088)	Tmemb_cc2(0.146)
	WAC_Acf1_DNA_bd(0.041)	WGR(0.053)	bZIP_2(0.088)
	AJAP1_PANP_C(0.105)	ANAPC3(0.070)	Abhydrolase_3(0.187)
	Atg14(0.105)	B12-binding_2(0.152)	BAGE(0.029)
	Cadherin_C_2(0.129)	Clat_adaptor_s(0.269)	Coillin_N(0.053)
	DNA_pol3_delta2(0.035)	DOR(0.146)	DUF1771(0.023)
	DUF3591(0.099)	DUF4464(0.041)	DUF846(0.058)
	DUF4665(0.058)	DUF4688(0.105)	ERCC4(0.111)
	EBP(0.053)	EMP70(0.140)	FRG1(0.322)
	FAM47(0.152)	FANCM-MHF_bd(0.058)	H-K_ATPase_N(0.082)
	Gal-3-0_sulfotrl(0.187)	Granin(0.175)	IFT46_B_C(0.082)
	HRCT1(0.199)	Helicase_C_4(0.070)	Keratin_B2(0.287)
	K_channel_TID(0.058)	Keratin_2_head(0.123)	MH2(0.140)
	LIAS_N(0.117)	MDF1(0.152)	NAD_binding_6(0.082)
	Myosin_tail_1(0.409)	NAD_binding_4(0.088)	Nramp(0.088)
	Ndc1_Nup(0.088)	Nitroreductase(0.035)	P53(0.579)
	Otopetrin(0.135)	P5-ATPase(0.105)	Peptidase_M48(0.053)
	PHTB1_N(0.082)	PRCC(0.058)	RGM_N(0.117)
	RDD(0.053)	RED_N(0.099)	RPAP3_C(0.047)
	RNA_pol_3_Rpc31(0.064)	RNA_pol_Rpb2_1(0.181)	Retinal(0.129)

	Rabaptin(0.105)	Ras(0.819)	Succ_DH_flav_C(0.041)
	SPT6_acidic(0.099)	SP_C-Propep(0.094)	TMCCDC2(0.105)
	TAF1D(0.035)	TIP_N(0.035)	Tcell_CD4_C(0.006)
	TSP_3(0.175)	Taxilin(0.117)	VEFS-Box(0.023)
	UDPG_MGDP_dh_N(0.047)	V-ATPase_C(0.117)	zf-CXXC(0.170)
	bcl-2I13(0.129)	dsrm(0.146)	
PCPG(0.949)	AChE_tetra(0.011)	ANAPC3(0.057)	Arm_3(0.023)
	DEAD_2(0.029)	DIL(0.023)	DND1_DSRM(0.017)
	DUF846(0.074)	Epiglycanin_TR(0.034)	FAM104(0.017)
	FancD2(0.029)	Fumble(0.029)	GFO_IDH_MocA(0.023)
	HJURP_C(0.011)	Histone_H2A_C(0.046)	IRK_N(0.011)
	MHC_I(0.074)	MIP(0.046)	Methyltransf_16(0.040)
	PARG_cat(0.023)	PBD(0.034)	PTEN_C2(0.029)
	RNA_pol_Rpb2_1(0.029)	RPOL_N(0.040)	Ras(0.160)
	SPC12(0.057)	SRA1(0.011)	Shisa(0.029)
	TAP42(0.011)	TAS2R(0.080)	THAP(0.046)
	TRAPP(0.063)	TRC8_N(0.086)	TSGA13(0.006)
	UDG(0.057)	V-ATPase_G(0.017)	VHL(0.017)
	BRO1(0.057)	C8(0.074)	Citrate_synt(0.040)
	DUF1308(0.017)	DUF1640(0.006)	DUF504(0.011)
	FAM86(0.034)	FRG1(0.080)	FRG2(0.023)
	G_glu_transpept(0.040)	Glyco_trans_1_4(0.023)	HIF-1(0.046)
	Keratin_2_head(0.080)	Keratin_B2(0.091)	Keratin_B2_2(0.314)
	Nop53(0.040)	ODR4-like(0.051)	OGFr_III(0.029)
	Pro-rich(0.086)	RAI16-like(0.034)	RIBIOP_C(0.051)
	Ribosomal_S13(0.017)	Ribosomal_S3Ae(0.023)	SH2(0.080)
	Staufen_C(0.017)	Striatin(0.040)	Syntaxin_2(0.017)
	TILa(0.063)	TMEM223(0.011)	TPR_1(0.086)
	Tektit(0.057)	Telethonin(0.006)	Trypsin(0.166)
	Wyosine_form(0.017)		
PRAD(0.755)	7tm_4(0.459)	AChE_tetra(0.005)	CDI(0.007)
	Creb_binding(0.009)	DUF4209(0.002)	DUF4597(0.009)
	FAM47(0.035)	FISNA(0.007)	FRG1(0.014)
	HRCT1(0.005)	Histone(0.066)	Keratin_B2(0.042)
	MATH(0.115)	MH2(0.021)	Med15(0.016)
	PTEN_C2(0.021)	SLIDE(0.012)	Synapsin_N(0.012)
	Cadherin(0.212)	Coilin_N(0.009)	Cornifin(0.012)
	DUF4722(0.007)	DUF4725(0.005)	DUF776(0.005)
	Flt3_lig(0.007)	Furin-like_2(0.009)	HNF_C(0.007)
	Keratin_B2_2(0.092)	Keratin_matx(0.009)	LCE(0.040)
	Mustang(0.002)	P53(0.092)	PARG_cat(0.014)
	UvrD_C_2(0.005)		
READ(0.966)	7tm_1(0.595)	APC_15aa(0.034)	APC_crr(0.147)
	Androgen_recep(0.060)	Arm(APC_u3(0.103))	Augurin(0.017)
	EPL1(0.267)	HBB(0.034)	IL10(0.060)
	MH2(0.103)	Maelstrom(0.026)	Neural_ProG_Cyt(0.026)
	S60S1(0.034)	SPT2(0.017)	Syndecan(0.052)
	APC_u5(0.086)	APC_u9(0.267)	AlaDh_PNT_N(0.034)
	DUF1388(0.164)	DUF4606(0.017)	DUF4721(0.026)
	KRTAP(0.095)	Keratin_2_tail(0.069)	Keratin_B2_2(0.233)
	OGFr_III(0.026)	P53(0.629)	Ras(0.500)
	TEX15(0.034)	Tom7(0.017)	

	FRG1(0.067)	GSHPx(0.063)	HIRA_B(0.008)
	Keratin_2_head(0.126)	LEP503(0.012)	Loricrin(0.035)
	Pro-rich(0.067)	Proton_antipo_M(0.055)	RB_A(0.051)
	RIBIOP_C(0.043)		
SKCM(0.987)	7tm_4(0.880)	AIG1(0.213)	ANF_receptor(0.427)
	Cadherin(0.667)	Cadherin_C(0.187)	Collagen(0.560)
	DUF4681(0.053)	DUF4698(0.120)	Glyco_hydro_1(0.200)
	Ig_3(0.533)	Ion_trans(0.693)	LELP1(0.053)
	NEXCaM_BD(0.080)	NMDAR2_C(0.173)	NYD-SP12_N(0.133)
	Pro-rich(0.213)	RIC3(0.040)	RNA_pol_Rpb2_5(0.013)
	UDPGT(0.440)	V-set(0.453)	fn3(0.773)
	Astacin(0.147)	CO_deh_flav_C(0.040)	CUB(0.480)
	Connexin50(0.040)	DCR(0.067)	DUF4392(0.027)
	HGAL(0.040)	I-set(0.733)	Ig_2(0.533)
	MACPF(0.213)	Myosin_tail_1(0.440)	NCD3G(0.107)
	P53(0.133)	PGK(0.147)	Pkinase_Tyr(0.840)
	Serpin(0.413)	Sushi(0.640)	Ten_N(0.107)
	p450(0.600)		
	7tm_1(0.654)	7tm_4(0.855)	AKAP7_RIRII_bdg(0.041)
	CLU(0.021)	Cadherin(0.651)	Cadherin_C(0.168)
	DUF3518(0.052)	DUF3591(0.070)	DUF3736(0.062)
STAD(0.984)	Filaggrin(0.096)	Forkhead(0.160)	Furin-like(0.090)
	I-set(0.685)	IRK(0.183)	Ion_trans(0.568)
	MAP2_projctn(0.075)	MH2(0.124)	MutS_II(0.065)
	Neil1-DNA_bind(0.013)	P53(0.432)	P53_tetramer(0.018)
	PUFD(0.049)	RGS12_us2(0.018)	SRP-alpha_N(0.057)
	TSP_1(0.295)	WBP-1(0.023)	Xylo_C(0.090)
	ANF_receptor(0.339)	An_peroxidase(0.212)	BTB_2(0.183)
	CitMHS(0.047)	DIE2_ALG10(0.070)	DUF2962(0.026)
	FAM209(0.026)	FAM70(0.049)	FGF(0.140)
	G-gamma(0.080)	HRCT1(0.021)	Homeobox(0.318)
	LCE(0.109)	LMWPc(0.023)	La(0.070)
	Myb_DNA-bind_7(0.028)	NESP55(0.034)	NYAP_N(0.067)
	PGM_PMM_I(0.098)	PTEN_C2(0.096)	PTN_MK_N(0.023)
	Sushi(0.460)	TORC_C(0.031)	TRP_2(0.041)
	fn3(0.602)		
KIRP(0.984)	7tm_4(0.375)	AChE_tetra(0.014)	AKNA(0.021)
	C1-set_C(0.028)	CD52(0.014)	CD99L2(0.021)
	Cor1(0.042)	Cornifin(0.111)	Cytochrom_B_C(0.021)
	DUF4514(0.014)	DUF4542(0.021)	DUF498(0.021)
	EBP(0.028)	EPL1(0.146)	Epiglycanin_TR(0.028)
	Fanconi_A_N(0.021)	GSHPx(0.063)	Glyco_hydro_39(0.021)
	HR1(0.042)	Helicase_RecD(0.021)	K167R(0.028)
	Keratin_B2_2(0.278)	LCE(0.042)	MHC_I(0.056)
	Myelin-PO_C(0.014)	NADH5_C(0.014)	NADHdh(0.014)
	Oscp1(0.021)	Oxidored_q3(0.014)	PHD20L1_u1(0.014)
	Ras(0.306)	STG(0.014)	SUIM_assoc(0.035)
	APC_u14(0.014)	APOC4(0.021)	Androgen_recep(0.056)
	COX1(0.035)	COX3(0.021)	Complex1_LYR_1(0.014)
	DHFR_1(0.035)	DUF1388(0.042)	DUF1754(0.014)
	DUF525(0.014)	DUF846(0.194)	Dimerisation2(0.021)
	FAM104(0.049)	FANCA_interact(0.014)	FRG1(0.049)
	Glycophorin_A(0.028)	Glyoxalase_4(0.014)	HJURP_C(0.028)
	Kelch_3(0.056)	Keratin_2_head(0.076)	Keratin_2_tail(0.056)
	MHC_I_C(0.007)	MRP-L51(0.014)	Methyltransf_2(0.021)
	NYD-SP12_N(0.021)	Nnf1(0.014)	Nt_Gln_amidase(0.014)
	Pro-rich(0.076)	Proton_antipo_M(0.063)	RPA_interact_M(0.014)

	Spin-Ssty(0.042)	TAS2R(0.076)	TPP_enzyme_N(0.035)
THCA(0.868)	7tm_4(0.334)	AXIN1_TNKS_BD(0.008)	BAGE(0.008)
	Ebp2(0.008)	Epiglycanin_TR(0.028)	FRG1(0.028)
	K167R(0.017)	Keratin_2_tail(0.006)	Keratin_B2_2(0.067)
	Ndufs5(0.003)	Nop16(0.011)	Oxidored_q5_N(0.006)
	Prokineticin(0.008)	Proton_antipo_M(0.034)	RIC3(0.008)
	SIP1(0.006)	SPATIAL(0.011)	SRA1(0.006)
	VitD-bind_III(0.006)	WASH_WAHD(0.008)	UDG(0.017)
	COX1(0.020)	DUF4523(0.006)	DUF4721(0.006)
	FRG2(0.014)	Filaggrin(0.025)	HCR(0.014)
	LNP1(0.006)	MaoC_dehydratas(0.008)	NADHdh(0.017)
	PAD_M(0.014)	Pkinase_Tyr(0.607)	Pro-rich(0.025)
	Ras(0.160)	Ribosomal_S13(0.008)	SEEK1(0.003)
	SURF6(0.008)	TEP1_N(0.008)	
	COX1(0.081)	COX14(0.008)	COX2(0.033)
	FRG1(0.130)	GTF2I(0.520)	Gb3_synth(0.041)
	Oxidored_q5_N(0.016)	Phospholamban(0.008)	Pro-rich(0.130)
THYM(0.870)	COX2_TM(0.049)	COX3(0.033)	Cytochrom_B_N_2(0.049)
	Keratin_B2_2(0.228)	NADHdh(0.073)	Oxidored_q3(0.073)
	Proton_antipo_M(0.065)	Staufen_C(0.024)	TMIE(0.041)
	ATG13(0.018)	BAGE(0.053)	COX1(0.140)
	Cytochrom_B_C(0.035)	Cytochrom_B_N_2(0.053)	DUF3518(0.105)
UCS(1.000)	FYRN(0.018)	HEAT(0.140)	HIG_1_N(0.053)
	Leu_zip(0.035)	MATH(0.070)	Maelstrom(0.035)
	Musclin(0.035)	NADH_dehy_S2_C(0.035)	NADHdh(0.088)
	PI3K_p85B(0.053)	PI3Ka(0.140)	PMAIP1(0.018)
	Proton_antipo_N(0.035)	SBF(0.070)	SRA1(0.035)
	WBP1(0.035)	tRNA_bind(0.070)	V1R(0.053)
	COX2_TM(0.035)	COX3(0.088)	Coatomer_b_Cpla(0.035)
	DUF3736(0.035)	DUF4503(0.035)	FRG1(0.070)
	HPIP(0.035)	LAMTOR5(0.018)	Las1(0.018)
	Membralin(0.053)	Methyltransf_FA(0.018)	Mucin-like(0.053)
	Oxidored_q3(0.053)	P53(0.789)	PI3K_P85_iSH2(0.123)
	PTEN_C2(0.070)	Pex14_N(0.018)	Proton_antipo_M(0.263)
	SprT-like(0.035)	Telomere_Sde2_2(0.018)	
	A2M_N_2(0.050)	ANAPC3(0.025)	Arm_3(0.037)
	Clat_adaptor_s(0.050)	Cytochrom_B561(0.037)	DIL(0.050)
	DUF846(0.050)	Dak1(0.025)	FAM86(0.025)
UVM(0.988)	G-alpha(0.512)	GPHH(0.037)	Glyco_trans_1_4(0.037)
	Keratin_2_head(0.212)	Keratin_B2(0.113)	Keratin_B2_2(0.325)
	MMS1_N(0.063)	Methyltransf_16(0.063)	Mucin-like(0.025)
	PAXIP1_C(0.013)	PBD(0.037)	PGM_PMM_I(0.037)
	Pkinase_C(0.063)	Pro-rich(0.125)	RIBIOP_C(0.025)
	SIP1(0.025)	SPRY(0.188)	SUI1(0.013)
	TILA(0.050)	TMEM132(0.063)	TRC8_N(0.150)
	COX16(0.013)	Catalase-rel(0.013)	Citrate_synt(0.037)
	DND1_DSRM(0.025)	DUF1771(0.013)	DUF504(0.050)
	FANCA_interact(0.013)	FRG1(0.100)	FancD2(0.050)
	HJURP_C(0.037)	Ig_3(0.163)	Katanin_con80(0.025)
	MBD_C(0.037)	MHC_I(0.263)	MHC_II_beta(0.025)
	Nitroreductase(0.025)	Nop10p(0.013)	PAG(0.013)
	PTEN_C2(0.050)	PX(0.087)	Peptidase_C12(0.100)
	Ribosomal_L5(0.013)	Ribosomal_S13(0.025)	SF3b10(0.025)
	Staufen_C(0.037)	Striatin(0.050)	TAP42(0.025)
	Tcp11(0.063)	Tektin(0.063)	Tower(0.013)
	Transmemb_17(0.025)	UDG(0.037)	zf-HC5HC2H(0.050)
	2-oxogl_dehyd_N(0.005)	7tm_1(0.472)	7tm_3(0.079)

AIG1(0.037)	AJAP1_PANP_C(0.009)	AKAP7_RIRII_bdg(0.005)
APC_u5(0.006)	APC_u9(0.011)	ASXH(0.009)
Cadherin_2(0.019)	Cadherin_C(0.074)	Cadherin_C_2(0.007)
NMDAR2_C(0.032)	NYAP_N(0.024)	Neur_chan_LBD(0.141)
HJURP_C(0.009)	HNH_3(0.002)	HRCT1(0.011)
Sulfotransfer_1(0.107)	Sushi(0.240)	Synapsin_N(0.008)
Ependymin(0.005)	Ephrin_lbd(0.048)	Epiglycanin_TR(0.022)
PYRIN(0.033)	Peptidase_M10(0.064)	Piwi(0.035)
Laminin_G_2(0.133)	Ldl_recept_a(0.106)	Lig_chan(0.085)
p450(0.253)	zf-C2H2(0.330)	zf-C2H2_6(0.211)
CABS1(0.010)	CCER1(0.006)	CD99L2(0.014)
MHC_I(0.059)	MHC_I_3(0.030)	MIP(0.053)
Filaggrin(0.041)	Forkhead(0.081)	Furin-like(0.036)
Rb_C(0.008)	Ribosomal_L16(0.012)	Ribosomal_L50(0.004)
DUF3591(0.024)	DUF3736(0.008)	DUF4584(0.012)
PARG_cat(0.013)	PAX(0.024)	PC-Esterase(0.019)
KRTAP(0.030)	K_channel_TID(0.004)	Keratin_2_head(0.060)
Tektin(0.043)	Ten_N(0.020)	Tis11B_N(0.007)
COX2(0.005)	COX2_TM(0.005)	COX3(0.011)
Methyltransf_FA(0.004)	Myosin_tail_1(0.195)	NACHT(0.069)
Glyco_transf_29(0.078)	Glycos_transf_2(0.073)	Granin(0.034)
SPAN-X(0.026)	SPT6_acidic(0.005)	SP_C-Propep(0.004)
DUF676(0.010)	DUF846(0.019)	Dppa2_A(0.006)
PI3Ka(0.074)	PMG(0.046)	POM121(0.022)
Kringle(0.046)	LBP_BPI_CETP(0.037)	LCE(0.056)
V-set(0.212)	VHL(0.034)	Vinculin(0.072)
B12-binding_2(0.005)	BAGE(0.007)	BSP_II(0.009)
Lig_chan-Glu_bd(0.045)	Lipase(0.056)	MACPF(0.045)
FAM47(0.073)	FAM70(0.012)	FISNA(0.005)
Protocadherin(0.025)	Proton_antipo_M(0.033)	RB_A(0.016)
Cytochrom_B_N_2(0.009)	DDDD(0.003)	DPPIV_N(0.043)
OLF(0.056)	Otopetrin(0.027)	Oxidored_q3(0.008)
Ig_2(0.171)	Ig_3(0.145)	Ion_trans(0.368)
TORC_C(0.007)	TRC8_N(0.022)	TRP_2(0.010)
Cornifin(0.015)	NtCtMGAM_N(0.015)	I-set(0.434)
Neuro_bHLH(0.013)	Cu_amine_oxidN2(0.007)	THRAP3_BCLAF1(0.040)
7tm_4(0.649)	AF-4(0.080)	AF1Q(0.008)
ANAPC3(0.009)	ANF_receptor(0.174)	APC_crr(0.008)
An_peroxidase(0.084)	Androgen_recep(0.018)	ArgoMid(0.009)
Cation_ATPase_N(0.025)	CitMHS(0.012)	Connexin43(0.002)
Neur_chan_memb(0.124)	Neurensin(0.012)	Neurexophilin(0.027)
Histone(0.127)	Histone_H2A_C(0.020)	Homeobox(0.161)
T-box(0.054)	TAFA(0.013)	TAS2R(0.107)
FAM104(0.011)	FAM209(0.007)	Cystatin(0.036)
Pkinase_Tyr(0.333)	Pro-rich(0.057)	OATP(0.101)
zf-HC5HC2H(0.023)	zf-Sec23_Sec24(0.008)	IRK(0.081)
zf-C2H2_assoc(0.004)	zf-CXXC(0.015)	TMEM132(0.035)
CDI(0.005)	CMS1(0.005)	COX1(0.018)
MYT1(0.017)	Med15(0.020)	Med9(0.004)
G-gamma(0.022)	Glyco_hydro_1(0.062)	Glyco_hydro_31(0.064)
RyR(0.023)	S6OS1(0.013)	SET(0.092)
DUF4589(0.012)	DUF4636(0.006)	DUF4688(0.012)
PI3K_C2(0.024)	PI3K_P85_iSH2(0.027)	PI3K_p85B(0.015)
Keratin_2_tail(0.011)	Keratin_B2(0.046)	Keratin_B2_2(0.160)
Trefoil(0.011)	Tyrosinase(0.019)	UDPGT(0.094)
COesterase(0.107)	CUB(0.197)	Cadherin(0.368)
NADH5_C(0.006)	NADHdh(0.015)	NKAP(0.008)
H-K_ATPase_N(0.003)	HAT_KAT11(0.020)	HIN(0.019)

SRP-alpha_N(0.008)	SUIM_assoc(0.005)	Serpin(0.157)
Drf_FH1(0.015)	E1_dh(0.034)	EPL1(0.032)
PTEN_C2(0.054)	PTPlike_phytase(0.006)	PUFD(0.007)
LIAS_N(0.004)	LLC1(0.005)	LRR_5(0.029)
WTX(0.026)	Xylo_C(0.013)	fn3(0.373)
BTB_2(0.072)	C1-set(0.069)	C2-set(0.010)
MAGE_N(0.036)	MATH(0.016)	MH2(0.033)
FMO-like(0.044)	FRG1(0.038)	FYRN(0.006)
RIBIOP_C(0.018)	Rap1_C(0.004)	Ras(0.273)
DSPc(0.097)	DUF1388(0.017)	DUF3518(0.017)
P53(0.330)	P53_tetramer(0.022)	PACT_coil_coil(0.006)
Iso_dh(0.075)	K167R(0.034)	KCNQ_channel(0.021)
TSC21(0.004)	TSKS(0.016)	TSP_1(0.172)
Cytochrom_B_C(0.004)	IRK_N(0.005)	IL17(0.013)
OGFr_III(0.006)	TMEM247(0.007)	TMCCDC2(0.007)

منابع

- [1] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell (5th edit.) Garland Science.* 2007.
- [2] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. *Biochemistry. 5th.* New York: WH Freeman, 2002.
- [3] Peter Russell. *iGenetics: A Mendelian approach.* Pearson/Benjamin Cummings San Francisco/London, 2006.
- [4] Anthea Maton, David Lahart, Jean Hopkins, Maryanna Quon Warner, Susan Johnson, and Jill D Wright. *Cells: Building blocks of life.* Pearson Prentice Hall, 1997.
- [5] Inna I Kruman. *Cell Cycle and DNA Damage Response in Postmitotic Neurons.* INTECH Open Access Publisher, 2011.
- [6] Conly L Rieder. *Mitosis and meiosis.* Academic Press, 1999.
- [7] Nathan L Nehrt, Thomas A Peterson, DoHwan Park, and Maricel G Kann. Domain landscapes of somatic mutations in cancer. *BMC genomics,* 13(Suppl 4):S9, 2012.
- [8] Lauren Pecorino. *Molecular biology of cancer: mechanisms, targets, and therapeutics.* Oxford university press, 2012.
- [9] Jennifer S Carew and Peng Huang. Mitochondrial defects in cancer. *Molecular cancer,* 1(1):1, 2002.

- [10] Nathan P Palmer, Patrick R Schmid, Bonnie Berger, and Isaac S Kohane. A gene expression profile of stem cell pluripotentiality and differentiation is conserved across diverse solid and hematopoietic cancers. *Genome Biol*, 13(8):R71, 2012.
- [11] Richard D Wood, Michael Mitchell, and Tomas Lindahl. Human dna repair genes, 2005. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 577(1):275–283, 2005.
- [12] Sabine S Lange, Kei-ichi Takata, and Richard D Wood. Dna polymerases and cancer. *Nature Reviews Cancer*, 11(2):96–110, 2011.
- [13] Errol C Friedberg, Graham C Walker, Wolfram Siede, and Richard D Wood. *DNA repair and mutagenesis*. American Society for Microbiology Press, 2005.
- [14] Anirban Ghosh and Manju Bansal. A glossary of dna structures from a to z. *Acta Crystallographica Section D: Biological Crystallography*, 59(4):620–626, 2003.
- [15] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, et al. Tissue-based map of the human proteome. *Science*, 347(6220):1260419, 2015.
- [16] Gary S Stein, Maria Borowski, Mai X Luong, Meng-Jiao Shi, Kelly P Smith, and Priscilla Vazquez. *Human Stem Cell Technology and Biology: A Research Guide and Laboratory Manual*. John Wiley & Sons, 2011.
- [17] Vincent Burrus and Matthew K Waldor. Shaping bacterial genomes with integrative and conjugative elements. *Research in microbiology*, 155(5):376–386, 2004.
- [18] Thomas Dittmar and Kurt S Zänker. *Role of cancer stem cells in cancer biology and therapy*. CRC Press, 2013.

- [19] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.
- [20] Ya Zhang, John-Marc Chandonia, Chris Ding, and Stephen R Holbrook. Comparative mapping of sequence-based and structure-based protein domains. *BMC bioinformatics*, 6(1):1, 2005.
- [21] Arne Elofsson and EL Sonnhammer. A comparison of sequence and structure protein domain families as a basis for structural genomics. *Bioinformatics*, 15(6):480–500, 1999.
- [22] Fan Yang, Evangelia Petsalaki, Thomas Rolland, David E Hill, Marc Vidal, and Frederick P Roth. Protein domain-level landscape of cancer-type-specific somatic mutations. *PLOS Comput Biol*, 11(3):e1004147, 2015.
- [23] Daniele Hernandez-Verdun. Nucleolus: from structure to dynamics. *Histochemistry and cell biology*, 125(1-2):127–137, 2006.
- [24] JD Watson and FHC Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *JAMA*, 269(15):1966–1967, 1993.
- [25] Christine Vogel, Matthew Bashton, Nicola D Kerrison, Cyrus Chothia, and Sarah A Teichmann. Structure, function and evolution of multidomain proteins. *Current opinion in structural biology*, 14(2):208–216, 2004.
- [26] W Gregory Feero, Alan E Guttmacher, W Gregory Feero, Alan E Guttmacher, and Francis S Collins. Genomic medicine—an updated primer. *New England Journal of Medicine*, 362(21):2001–2011, 2010.
- [27] DE Avner, WE Harmon, P Niaudet, and N Yoshikawa. *Pediatric Nephrology*. 2009. Springer.

- [28] PP Vaidyanathan. Genomics and proteomics: a signal processor's tour. *Circuits and Systems Magazine, IEEE*, 4(4):6–29, 2004.
- [29] David Brett, Heike Pospisil, Juan Valcárcel, Jens Reich, and Peer Bork. Alternative splicing and genome complexity. *Nature genetics*, 30(1):29–30, 2002.
- [30] Justin C St John. *Mitochondrial DNA, Mitochondria, Disease and Stem Cells*. Springer Science & Business Media, 2012.
- [31] Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz. The cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemporary oncology*, 19(1A):A68, 2015.
- [32] Alain Vignal, Denis Milan, Magali SanCristobal, André Eggen, et al. A review on snp and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3):275–306, 2002.
- [33] Jubao Duan, Mark S Wainwright, Josep M Comeron, Naruya Saitou, Alan R Sanders, Joel Gelernter, and Pablo V Gejman. Synonymous mutations in the human dopamine receptor d2 (drd2) affect mrna stability and synthesis of the receptor. *Human molecular genetics*, 12(3):205–216, 2003.
- [34] John S Bertram. The molecular biology of cancer. *Molecular aspects of medicine*, 21(6):167–223, 2000.
- [35] Yael T Aminetzach, J Michael Macpherson, and Dmitri A Petrov. Pesticide resistance via transposition-mediated adaptive gene truncation in drosophila. *Science*, 309(5735):764–767, 2005.
- [36] John S Bertram. The molecular biology of cancer. *Molecular aspects of medicine*, 21(6):167–223, 2000.

- [37] Frank Eisenhaber. *Discovering biomolecular mechanisms with computational biology*. Springer, 2006.
- [38] David P Minde, Zeinab Anvarian, Stefan GD Rüdiger, and Madelon M Maurice. Messing up disorder: how do missense mutations in the tumor suppressor protein apc lead to cancer? *Molecular cancer*, 10(1):1, 2011.
- [39] Motoo Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, 1984.
- [40] JV Chamary, Joanna L Parmley, and Laurence D Hurst. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, 7(2):98–108, 2006.
- [41] David B Parkinson and Rajesh V Thakker. A donor splice site mutation in the parathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *Nature genetics*, 1(2):149–152, 1992.
- [42] Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, James Darnell, et al. *Molecular cell biology*, volume 4. WH Freeman New York, 2000.
- [43] Lawrence C Panasci and Moulay A Alaoui-Jamali. *DNA repair in cancer therapy*. Springer Science & Business Media, 2004.
- [44] Stella Veretnik, Jenny Gu, and Shoshana Wodak. Identifying structural domains in proteins. *Structural Bioinformatics*, 2:485–513, 2009.
- [45] Jelle J Goeman and Aldo Solari. Multiple hypothesis testing in genomics. *Statistics in medicine*, 33(11):1946–1978, 2014.
- [46] SE Pories. *Cancer. Biographies of Disease Series*. Portsmouth, Greenwood Press, 2009.

- [47] Guido Kroemer. Mitochondria in cancer. *Oncogene*, 25(34):4630–4632, 2006.
- [48] K Robinson, J Creed, B Reguly, C Powell, R Wittock, D Klein, A Maggrah, L Klotz, RL Parr, and GD Dakubo. Accurate prediction of repeat prostate biopsy outcomes by a mitochondrial dna deletion assay. *Prostate cancer and prostatic diseases*, 13(2):126–131, 2010.
- [49] Brenton Thomas Tan, Christopher Yongchul Park, Laurie Elizabeth Ailles, and Irving L Weissman. The cancer stem cell hypothesis: a work in progress. *Laboratory Investigation*, 86(12):1203–1207, 2006.
- [50] Tanya Barrett, Dennis B Troup, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, Katherine H Phillippy, Patti M Sherman, et al. Ncbi geo: archive for functional genomics data sets—10 years on. *Nucleic acids research*, 39(suppl 1):D1005–D1010, 2011.
- [51] Jamie M Sperger, Xin Chen, Jonathan S Draper, Jessica E Antosiewicz, Chris H Chon, Sunita B Jones, James D Brooks, Peter W Andrews, Patrick O Brown, and James A Thomson. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proceedings of the National Academy of Sciences*, 100(23):13350–13355, 2003.
- [52] Rolf I Skotheim, Guro E Lind, Outi Monni, Jahn M Nesland, Vera M Abeler, Sophie D Fosså, Nur Duale, Gunnar Brunborg, Olli Kallioniemi, Peter W Andrews, et al. Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. *Cancer research*, 65(13):5588–5598, 2005.
- [53] Kristian Almstrup, Christina E Hoei-Hansen, Ute Wirkner, Jonathon Blake, Christian Schwager, Wilhelm Ansorge, John E Nielsen, Niels E Skakkebæk, Ewa Rajpert-De Meyts, and Henrik Leffers. Embryonic stem cell-like features of

- testicular carcinoma in situ revealed by genome-wide gene expression profiling. *Cancer Research*, 64(14):4736–4743, 2004.
- [54] Cristian Tomasetti and Bert Vogelstein. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 347(6217):78–81, 2015.
- [55] Felix Dietlein, Lisa Thelen, and H Christian Reinhardt. Cancer-specific defects in dna repair pathways as targets for personalized therapeutic approaches. *Trends in Genetics*, 30(8):326–339, 2014.
- [56] Mark DM Leiserson, Fabio Vandin, Hsin-Ta Wu, Jason R Dobson, Jonathan V Eldridge, Jacob L Thomas, Alexandra Papoutsaki, Younhun Kim, Beifang Niu, Michael McLellan, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature genetics*, 47(2):106–114, 2015.
- [57] Christian Gilissen, Jayne Y Hehir-Kwa, Djie Tjwan Thung, Maartje van de Vorst, Bregje WM van Bon, Marjolein H Willemsen, Michael Kwint, Irene M Janssen, Alexander Hoischen, Annette Schenck, et al. *Genome sequencing identifies major causes of severe intellectual disability*. Nature Publishing Group, 2014.
- [58] Sarah B Ng, Emily H Turner, Peggy D Robertson, Steven D Flygare, Abigail W Bigham, Choli Lee, Tristan Shaffer, Michelle Wong, Arindam Bhattacharjee, Evan E Eichler, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261):272–276, 2009.
- [59] David A Wheeler, Maithreyan Srinivasan, Michael Egholm, Yufeng Shen, Lei Chen, Amy McGuire, Wen He, Yi-Ju Chen, Vinod Makhijani, G Thomas Roth,

- et al. The complete genome of an individual by massively parallel dna sequencing. *nature*, 452(7189):872–876, 2008.
- [60] Stewart Scherer. *short guide to the human genome*. Cold Spring Harbor Laboratory Press, 2008.
- [61] Rusk N. E pluribus unum. *Nature Methods*, page 331, 2010.
- [62] Nicholas Wade. Genome of dna pioneer is deciphered. *New York Times*, pages 10–18, 2007.
- [63] Melissa J Fullwood, Chia-Lin Wei, Edison T Liu, and Yijun Ruan. Next-generation dna sequencing of paired-end tags (pet) for transcriptome and genome analyses. *Genome research*, 19(4):521–532, 2009.
- [64] G Gibson, S Muse, and John Boyle. Book reviews-a primer of genome science. *Biochemistry and Molecular Biology Education*, 33(4):313, 2005.
- [65] Ian Sillitoe, Tony E Lewis, Alison Cuff, Sayoni Das, Paul Ashford, Natalie L Dawson, Nicholas Furnham, Roman A Laskowski, David Lee, Jonathan G Lees, et al. Cath: comprehensive structural and functional annotations for genome sequences. *Nucleic acids research*, page gku947, 2014.
- [66] Christine A Orengo, AD Michie, S Jones, David T Jones, MB Swindells, and Janet M Thornton. Cath—a hierarchic classification of protein domain structures. *Structure*, 5(8):1093–1109, 1997.
- [67] Ryan Day, David AC Beck, Roger S Armen, and Valerie Daggett. A consensus view of fold space: combining scop, cath, and the dali domain dictionary. *Protein Science*, 12(10):2150–2160, 2003.

- [68] Christine A Orengo, AD Michie, S Jones, David T Jones, MB Swindells, and Janet M Thornton. Cath—a hierarchic classification of protein domain structures. *Structure*, 5(8):1093–1109, 1997.
- [69] Alexey G Murzin, Steven E Brenner, Tim Hubbard, and Cyrus Chothia. Scop: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*, 247(4):536–540, 1995.
- [70] Alex Bateman, Lachlan Coin, Richard Durbin, Robert D Finn, Volker Hollich, Sam Griffiths-Jones, Ajay Khanna, Mhairi Marshall, Simon Moxon, Erik LL Sonnhammer, et al. The pfam protein families database. *Nucleic acids research*, 32(suppl 1):D138–D141, 2004.
- [71] Robert D Finn, Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y Eberhardt, Sean R Eddy, Andreas Heger, Kirstie Hetherington, Liisa Holm, Jaina Mistry, et al. Pfam: the protein families database. *Nucleic acids research*, page gkt1223, 2013.
- [72] Robert D Finn, Penelope Coggill, Ruth Y Eberhardt, Sean R Eddy, Jaina Mistry, Alex L Mitchell, Simon C Potter, Marco Punta, Matloob Qureshi, Amaia Sangrador-Vegas, et al. The pfam protein families database: towards a more sustainable future. *Nucleic acids research*, page gkv1344, 2015.
- [73] Frances C Bernstein, Thomas F Koetzle, Graeme JB Williams, Edgar F Meyer, Michael D Brice, John R Rodgers, Olga Kennard, Takehiko Shimanouchi, and Mitsuo Tasumi. The protein data bank. *European Journal of Biochemistry*, 80(2):319–324, 1977.
- [74] UniProt Consortium et al. Uniprot: a hub for protein information. *Nucleic acids research*, page gku989, 2014.

- [75] Donna Karolchik, Angela S Hinrichs, Terrence S Furey, Krishna M Roskin, Charles W Sugnet, David Haussler, and W James Kent. The ucsc table browser data retrieval tool. *Nucleic acids research*, 32(suppl 1):D493–D496, 2004.
- [76] Kristian A Gray, Bethan Yates, Ruth L Seal, Mathew W Wright, and Elspeth A Bruford. Genenames.org: the hgnc resources in 2015. *Nucleic acids research*, page gku1071, 2014.
- [77] Kim D Pruitt, Tatiana Tatusova, and Donna R Maglott. Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 33(suppl 1):D501–D504, 2005.
- [78] Sameer Velankar, José M Dana, Julius Jacobsen, Glen van Ginkel, Paul J Gane, Jie Luo, Thomas J Oldfield, Claire O’Donovan, Maria-Jesus Martin, and Gerard J Kleywegt. Sifts: structure integration with function, taxonomy and sequences resource. *Nucleic acids research*, page gks1258, 2012.
- [79] B Robert. Ash. information theory, 1990.

واژه‌نامه فارسی به انگلیسی

Epidermolysis bullosa	اپیدرمولیز بولوسا
E-value	ارزش مورد انتظار
SCOP	اسکاپ
SCOP	اسکاپ
Cytoskeleton	اسکلت سلولی
Noncoding SNPs	اسنیپ‌های غیرکدکننده
Nonsynonymous SNPs	اسنیپ‌های غیرمتراff
Coding SNPs	اسنیپ‌های کدکننده
Synonymous SNPs	اسنیپ‌های متراff
Amino acid	اسیدهای آمینه
Nucleic acid	اسیدهای نوکلئیک
Exon	اگرون
Bit-Score	امتیاز بیت
Organelles	اندامک ها
Ensembl	انسمبل
Misaligned	انطباق اشتباه
Isoform	ایزوفرم
Interphase	اینترفاز
Intron	اینtron
Adenocarcinoma	آدنوکارسینوما
Adenine	آدنین

Ribonucleic Acid(RNA)	آر ان آ
Messenger RNA(mRNA)	آر ان آ پیام‌رسان
Ribosomal RNA(rRNA)	آر ان آ ریبوزومی
Transfer RNA(tRNA)	آر ان آ ناقل
RNA polymerase	آر ان آ پلی مراز
non-Coding RNA	آر ان آ غیر کد کننده
threshold	آستانه
Allele	آلل
binary entropy	آنتروپی دوتایی
Oncogenes	آنکوژن ها
ANNOVAR	آنوار
DNA Repair	بازسازی دی‌ان‌آ
Glandular	بافت‌های غده ای
Adipose Tissue	بافت‌های چربی
RefSeq	بانک اطلاعاتی رف-سک
Protein Data Bank(PDB)	بانک اطلاعاتی پروتئین
Conserved Domain Database (CDD)	بانک اطلاعاتی دومین‌های نگهداری شده
UCSC	بانک اطلاعاتی یو-اس-سی
splice variants	باهم متصل کردن
Totipotent	بس توان
BLAST	بیلست
Alzheimer's disease	بیماری آلزایمر
sicklecell disease	بیماری داسی شکل سلول
Terminator	پایان دهنده
Peroxisome	پرکسی زوم ها
Pluripotent	پرتوان
RNA splicing	پردازش آر ان آ

Proteome	پروتئوم
Protein	پروتئین
prokaryote	پروکاریوت
Platforms	پلت فرم ها ..
polysaccharide	پلی ساکارید ..
Purine	پورین ..
Nuclear Envelope	پوشش هسته ..
P-value	پی ارزش ..
Pfam	پی فم ..
Splicing	پیرایش ..
Pyrimidine	پیریمیدین ..
Tetrad	تتراد ..
Pan-Cancer Network Analysis	تجزیه شبکه‌ای یکپارچه سرطان‌ها ..
Local False Discovery	تجزیه و تحلیل نرخ کشف کاذب محلی ..
Translation	ترجمه ..
Poor DNA repair	ترمیم جفت ناجور ..
Bonferroni correction	تصحیح بونفرونی ..
Array Copy Number	تعداد کپی آرایه ..
SNVs nonsynonymous (nsSNVs)	تغییرات اسید آمینه ..
posttranslational modifications	تغییرات پس‌ترجمه ..
Copy Number Variation(CNV)	تغییرات تعداد کپی ..
microRNA	تغییرات ریز آر. ان. آ ..
genetic variations	تغییرات ژنتیکی ..
Crossing-over	تقاطع کروموزومی ..
Stirling's Approximation	تقریب استرلینگ ..
Unipotent	تک توان ..
Whole Genome Sequencing(WGS)	توالی کل ژنوم ..

Whole Exon Sequencing(WES)	توالی کلی اگزون
pairedend sequencing	توالی یابی جفت پایانی
Polymerase chain reaction(PCR)	توالی‌های داربست‌ها از روش‌های پی. آر
Thymine	تیمین
In frameshift mutation	جابه‌جایی درون قالب
DNA Replication	جایگزینی دی. ان. آ
Mate pair	جفت‌های همسر
missense mutations	جهش‌های بدمعنی
Somatic Mutation	جهش‌های بدنی
Germ-line mutation	جهش‌های دودمان زایشی
nonsynonymous mutations	جهش‌های نامتراالف
Nonsense mutations	جهش‌های بی معنی
Frameshift mutation	جهش‌های جابه‌جایی قالب
Silent mutations	جهش‌های خاموش
Neutral mutations	جهش‌های طبیعی
synonymous mutation	جهش‌های مترادف
Splitsite mutations	جهش‌های محل شکاف
Lipid	چربی
Cell cycle	چرخه سلولی
Multipotent	چند توان
SNPSingle Nucleotide Polymorphism	چند شکلی در یک نوکلئوتید واحد یا اسنیپ
Polymorphisms	چندریختی‌ها
Protein domains	حوزه‌های پروتئینی
Lifetime Risk	خطر طول عمر
Blood	خون
Scaffold	داربست
Poly peptide	درشت‌مولکول‌های پپتیدی

Golgi apparatus	دستگاه گلزی
D-Loop	دی-لوپ
Deoxyribonucleic Acid(DNA)	دی ان آ
Junk DNA	دی ان آ اضافی
Driver	راننده
Promoter	راه انداز
Strand	رشته
RefSeq	رفسک
Transcription	رونویسی
Ribosome	ریبوزوم
Markov chains	زنجیره‌های مارکوف
Gene	ژن
Tumor suppressor genes	ژن‌های متوقف کننده غده
Genetic	ژنتیک
SCGS genes in the DNA replication	ژن‌های سلول بنیادی در تکثیر دی.ان.آ
SCGS genes in the RNA transcription	ژن‌های سلول بنیادی در رونویسی آر.ان.آ
SCGS genes in the metabolism	ژن‌های سلول بنیادی در سوخت و ساز بدن
SCGS genes in the multicellular signaling	ژن‌های سلول بنیادی در سیگنالینگ چند سلولی
Protein Coding Genes	ژن‌های کدکننده پروتئین
Genome	ژنوم
Reference Genome	ژنوم مرجع
Sarcoma	سارکوما
Cancer Gene Census	سرشماری ژن سرطان
colon adenocarcinoma	سرطان روده
oligodendroglomas	سرطان مغز درجه پایین
cell	سلول
Mesoderm	سلول‌های لایه میانی

Epithelial	سلول های مخاطی
Stem Cells	سلول های بنیادی
Adult Stem Cell	سلول های بنیادی بالغ
Cancer Stem Cells	سلول های بنیادی سرطان
Embryonic Cells	سلول های رویانی
Kallman Syndrome	سندرم کالمن
Cytoplasm	سیتوپلاسم
Cytosol	سیتوزول
Cytosine	سیتوزین
SIFTS(Structual integration with function, (banک اطلاعاتی منبع ادغام ساختارها و توالی‌ها)	taxonomy and sequence)
immune signaling	سیگنالینگ ایمنی
hormone signaling	سیگنالینگ هورمون
Endoplasmic reticulum	شبکه آندوپلاسمی
Pseudogenes	شبه‌زن‌ها
Bacterial artificial chromosome(BAC)	شبيه‌سازی بي. اي. سي
Gene Ontology terms (GO terms)	شرایط هستی شناسی زن
Dementia	صرع
sheet	صفحه بتا
oxidative phosphorylation	عمل اکسیده شدن فسفر
Plasma membranes	غشاء
Reading frame	قالب خواندن
Gene Peaks	قله‌های ژنی
Carcinoma	کارسینوها
Contig	کانتیگ
Codon	کدن
Chromatid	کروماتید

Chromatin	کروماتین
Chromosome	کروموزوم
CATH Protein Structure Classification	کلاس‌بندی ساختار پروتئین کت
Chloroplast	کلروپلاست
Oligopotent	کم توان
Parathyroid Deficiency	کمبود پاراتیروئید
Genome Reference Consortium(GRC)	کنسرسیو姆 ژنوم مرجع
Guanine	گوانین
Lysosome	لیزozوم ها
helix	مارپیچ آلفا
Double helix	مارپیچ دورشته ای
cell cycle module	ماژول چرخه سلولی
protein synthesis module	ماژول سنتز پروتئین
cell identity module	ماژول هویت سلول
Metastasis	متاستاز
Methylation	متیلاسیون
Protein Complex IV	مجموعه پروتئینی چهار
Protein Complex I	مجموعه پروتئینی یک
Cancer Gene Set	مجموعه ژن‌های سرطانی
TF binding site	محل‌های اتصال عامل ترجمه
Hidden Markov Model	مدل مخفی مارکوف
Passenger	مسافر
Bone Marrow	مغز استخوان
Large-scale	مقیاس بزرگ
Nuclear Pores	منافذ
Synonymous	منطقی
Structural motif	موتیف‌های ساختاری

The Cancer Genome Atlas(TCGA)	مؤسسه اطلس ژنوم سرطانی
National human Genome Research Institute(NHGRI)	مؤسسه ملی تحقیقات ژنوم انسانی
National Cancer Institute(NCI)	مؤسسه ملی سرطان
Mitosis	میتوز
Mitochondrion	میتوکندری
Meiosis	میوز
familywise(FWER)	نرخ خطای جمعی
Domain Mutational Landscape	نقشه جهش دومینی
somatic recombination	نوترکیبی سوماتیک
Nucleoplasm	نوکلئوپلاسم
Nucleotide	نوکلئوتیدها
Hetrodimer	هترودایمر
Nucleus	هستک‌ها
Cell Nucleus	هسته
Helicase	هلیکاز
HUGO Gene Nomenclature Committee	هوگو، کمیته نام‌گذاری ژن‌ها
Histones	هیستون‌ها
Epigenetic	وراثنیک
Vesicle	وزیکول
Patches	ویرایش‌ها
Uracil	یوراسیل
Eukaryote	یوکاریوت
Uniprot	یونی پرات

واژه‌نامه انگلیسی به فارسی

Adenine	آدنین
Adenocarcinoma	آدنوکارسينوما
Adipose Tissue	بافت‌های چربی
Adult Stem Cell	سلول‌های بنیادی بالغ
Allele	آلل
Alzheimer's disease	بیماری آلزایمر
Amino acid	اسیدهای آمینه
ANNOVAR	آنوار
Array Copy Number	تعداد کپی آرایه
Bacterial artificial chromosome(BAC)	شبیه‌سازی بی. ای. سی
binary entropy	آنتروپی دوتایی
Bit-Score	امتیاز بیت
BLAST	بیلست
Blood	خون
Bone Marrow	مغز استخوان
Bonferroni correction	تصحیح بونفرونی
Cancer Gene Census	سرشماری ژن سرطان
Cancer Gene Set	مجموعه ژن‌های سرطانی

Cancer Stem Cells	سلول‌های بنیادی سرطان
Carcinoma	کارسینوها
CATH Protein Structure Classification	کلاس‌بندی ساختار پروتئین کت
cell	سلول
Cell cycle	چرخه سلولی
cell cycle module	ماژول چرخه سلولی
cell identity module	ماژول هویت سلول
Cell Nucleus	هسته
Chloroplast	کلروپلاست
Chromatid	کروماتید
Chromatin	کروماتین
Chromosome	کروموزوم
Coding SNPs	اسنیپ‌های کدکننده
Codon	کدن
colon adenocarcinoma	سرطان روده
Conserved Domain Database (CDD)	بانک اطلاعاتی دومین‌های نگهداری شده
Contig	کانتیگ
Copy Number Variation(CNV)	تغییرات تعداد کپی
Crossing-over	تقاطع کروموزومی
Cytoplasm	سیتوپلاسم
Cytosine	سیتوزین
Cytoskeleton	اسکلت سلولی
Cytosol	سیتوزول

Dementia	صرع
Deoxyribonucleic Acid(DNA)	دی ان آ
D-Loop	دی-لوپ
DNA Repair	بازسازی دی.ان.آ.
DNA Replication	جایگزینی دی.ان.آ.
Domain Mutational Landscape	نقشه جهش دومینی
Double helix	مارپیچ دورشته ای
Driver	راننده
Embryonic Cells	سلول‌های رویانی
Endoplasmic reticulum	شبکه آندوپلاسمی
Ensembl	انSEMBL
Epidermolysis bullosa	اپیدرمولیز بولوسا
Epigenetic	وراثناتیک
Epithelial	سلول‌های مخاطی
Eukaryote	یوکاریوت
E-value	ارزش مورد انتظار
Exon	اگزون
familywise(FWER)	نرخ خطای جمعی
Frameshift mutation	جهش‌های جابه‌جاوی قالب
Gene	ژن
Gene Ontology terms (GO terms)	شرایط هستی شناسی ژن
Gene Peaks	قلمهای ژنی
Genetic	ژنتیک

genetic variations	تغییرات ژنتیکی
Genome	ژنوم
Genome Reference Consortium(GRC)	کنسرسیو姆 ژنوم مرجع
Germ-line mutation	جهش های دودمان زایشی
Glandular	بافت های غده ای
Golgi apparatus	دستگاه گلتری
Guanine	گوانین
Helicase	هلیکاز
Hetrodimer	هترودایمر
Hidden Markov Model	مدل مخفی مارکوف
Histones	هیستون ها
hormone signaling	سیگنالینگ هورمون
HUGO Gene Nomenclature Committee	هوگو، کمیته نام‌گذاری ژن‌ها
immune signaling	سیگنالینگ ایمنی
In frameshift mutation	جابه‌جایی درون قالب
Interphase	اینترفاز
Intron	اینtron
Isoform	ایزوفرم
Junk DNA	دی ان آ اضافی
Kallman Syndrome	سندرم کالمان
Large-scale	مقیاس-بزرگ
Lifetime Risk	خطر طول عمر
Lipid	چربی

Local False Discovery	تجزیه و تحلیل نرخ کشف کاذب محلی
Lysosome	لیزوزوم ها
Markov chains	زنجیره‌های مارکوف
Mate pair	جفت های همسر
Meiosis	میوز
Mesoderm	سلول های لایه میانی
Messenger RNA(mRNA)	آر ان آ پیام‌رسان
Metastasis	متاستاز
Methylation	متیلاسیون
microRNA	تغییرات ریز آر ان آ
Misaligned	انطباق اشتباه
missense mutations	جهش های بدمعنی
Mitochondrion	میتوکندری
Mitosis	میتوز
Multipotent	چند توان
National Cancer Institute(NCI)	مؤسسه ملی سرطان
National human Genome Research Institute(NHGRI)	مؤسسه ملی تحقیقات ژنوم انسانی
Neutral mutations	جهش‌های طبیعی
non-Coding RNA	آر ان آ غیرکد کننده
Noncoding SNPs	اسنیپ‌های غیرکد کننده
Nonsense mutations	جهش‌های بی معنی
nonsynonymous mutations	جهش های نامترادف
Nonsynonymous SNPs	اسنیپ‌های غیرمترادف

Nuclear Envelope	پوشش هسته
Nuclear Pores	منافذ
Nucleic acid	اسیدهای نوکلئیک
Nucleoplasm	نوکلئوپلاسم
Nucleotide	نوکلئوتیدها
Nucleus	هستک‌ها
oligodendrogiomas	سرطان مغز درجه پایین
Oligopotent	کم توان
Oncogenes	آنکوژن‌ها
Organelles	اندامک‌ها
oxidative phosphorylation	عمل اکسیده شدن فسفر
pairedend sequencing	توالی یابی جفت پایانی
Pan-Cancer Network Analysis	تجزیه شبکه‌ای یکپارچه سرطان‌ها
Parathyroid Deficiency	کمبود پاراتیروئید
Passenger	مسافر
Patches	ویرایش‌ها
Peroxisome	پراکسی زوم‌ها
Pfam	پی‌فم
Plasma membranes	غشاء
Platforms	پلت‌فرم‌ها
Pluripotent	پرتوان
Poly peptide	درشت‌مولکول‌های پپتیدی
Polymerase chain reaction(PCR)	توالی‌های داربست‌ها از روش‌های پی. آر. سی.

Polymorphisms	چندریختی‌ها
polysaccharide	پلی ساکارید
Poor DNA repair	ترمیم جفت ناجور
posttranslational modifications	تغییرات پس‌ترجمه
prokaryote	پروکاریوت
Promoter	راه انداز
Protein	پروتئین
Protein Coding Genes	ژن‌های کدکننده پروتئین
Protein Complex I	مجموعه پروتئینی یک
Protein Complex IV	مجموعه پروتئینی چهار
Protein Data Bank(PDB)	بانک اطلاعاتی پروتئین
Protein domains	حوزه‌های پروتئینی
protein synthesis module	ماژول سنتز پروتئین
Proteome	پروتئوم
Pseudogenes	شبه‌ژن‌ها
Purine	پورین
P-value	پی-ارزش
Pyrimidine	پیریمیدین
Reading frame	قالب خواندن
Reference Genome	ژنوم مرجع
RefSeq	بانک اطلاعاتی رف-سک
RefSeq	رف-سک
Ribonucleic Acid(RNA)	آر ان آ

Ribosomal RNA(rRNA)	آر ان آ ریبوزومی
Ribosome	ریبوزوم
RNA polymerase	آر ان آ پلی مراز
RNA splicing	پردازش آر ان آ
Sarcoma	سارکوما
Scaffold	داربست
SCGS genes in the DNA replication	ژن‌های سلول بنیادی در تکثیر دی‌ان‌آ
SCGS genes in the metabolism	ژن‌های سلول بنیادی در سوخت و ساز بدن
SCGS genes in the multicellular signaling	ژن‌های سلول بنیادی در سیگنالینگ چند سلولی
SCGS genes in the RNA transcription	ژن‌های سلول بنیادی در رونویسی آر ان آ
SCOP	اسکاپ
SCOP	اسکاپ
sicklecell disease	بیماری داسی شکل سلول
SIFTS(Structual integration with function, سیفتز(بانک اطلاعاتی منبع ادغام ساختارها و توالی‌ها)	taxonomy and sequence)
Silent mutations	جهش‌های خاموش
SNPSingle Nucleotide Polymorphism	چند شکلی در یک نوکلئوتید واحد یا اسنیپ
SNVs nonsynonymous (nsSNVs)	تغییرات اسید آمینه
Somatic Mutation	جهش‌های بدنی
somatic recombination	نوترکیبی سوماتیک
splice variants	باهم متصل کردن
Splicing	پیرایش
Splitsite mutations	جهش‌های محل شکاف

Stem Cells	سلول‌های بنیادی
Stirling's Approximation	تقریب استرلینگ
Strand	رشته
Structural motif	موتیف‌های ساختاری
Synonymous	منطقی
synonymous mutation	جهش‌های متراծ
Synonymous SNPs	اسنیپ‌های متراծ
Terminator	پایان دهنده
Tetrad	تتراد
TF binding site	محل‌های اتصال عامل ترجمه
The Cancer Genome Atlas(TCGA)	مؤسسه اطلس ژنوم سرطانی
threshold	آستانه
Thymine	تیمین
بس توان	
Transcription	رونویسی
Transfer RNA(tRNA)	آر ان آ ناقل
Translation	ترجمه
Tumor suppressor genes	ژن‌های متوقف کننده غده
UCSC	بانک اطلاعاتی یو-سی-اس-سی
Unipotent	تک توان
Uniprot	یونی پرات
Uracil	یوراسیل
Vesicle	وزیکول

Whole Exon Sequencing(WES)	توالی کلی اگزون
Whole Genome Sequencing(WGS)	توالی کل ژنوم
helix	مارپیچ آلفا
sheet	صفحه بتا

Abstract

A protein domain is a conserved part of a given protein sequence that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-dimensional structure and often can be independently stable and folded. Study the effects of changes in the genome on protein domain might also help to identify the factors cause of cancer. The number of human genes and proteins is high and relation between proteins and their domains is complex and recognizing the major causes of cancer by laboratory methods is also very difficult. All of these make it necessary to use computer science and statistics in the study of cancer. In this thesis the problem of finding protien domains involved in the formation and development of cancer has been studied. In this approach initially the probability of mutation on the genome has been calculated by a proposed method. In the next step by setting a threshold, candidate domains involved in each cancer has been extracted and based on these candidate domains, candidate stem cells and repair genes for each type of cancer has been discussed.

Keywords: *Protein Domain, Classification, Pfam, CATH, Cancer, Stem Cell, Repair Gene, Mitochondria, Pan Cancer, Computational models.*



University of Tehran
School of Mathematics, Statistics and Computer Science

Incorporating Protein Domain for Sparse Classifiers in Cancer Associations

By
Seirana Hashemi Ranjbar

Supervisors
Dr. Abbas Nowzari Dalini
Dr. Zahra Razaghi Moghadam Kashani

Advisor
Dr. Ali Mohamad Banaei Moghadam

Master of Science
Computer Science

July 2016