

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده علوم و فنون نوین
گروه شبکه



استنتاج درخت فیلوژنی تومور سرطانی با استفاده از داده‌های تک‌سلولی و تغییرات تعداد تکرار

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی فناوری اطلاعات
گرایش سامانه‌های شبکه‌ای

افشین بزرگ‌پور

اساتید راهنما

دکتر سامان هراتی‌زاده و دکتر ابوالفضل مطهری

مرداد ۱۴۰۰

چکیده

این راهنما، نمونه‌ای از قالب پروژه، پایان‌نامه و رساله دانشگاه تهران می‌باشد که با استفاده از کلاس -tehran thesis و بسته‌ی پرشین در \LaTeX تهیه شده است. این قالب به گونه‌ای طراحی شده است که مطابق با دستورالعمل نگارش و تدوین پایان‌نامه کارشناسی ارشد و دکتری، مورخ ۹۳/۰۶/۰۳ پردیس دانشکده‌های فنی دانشگاه تهران باشد و حروف چینی بسیاری از قسمت‌های آن، مطابق با استاندارد قالب‌های فارسی پایان‌نامه در لاتک، به طور خودکار انجام می‌شود.

چکیده بخشی از پایان‌نامه است که خواننده را به مطالعه آن علاقمند می‌کند و یا از آن می‌گریزند. چکیده باید ترجیحاً در یک صفحه باشد. در نگارش چکیده نکات زیر باید رعایت شود. متن چکیده باید مزین به کلمه‌ها و عبارات سلیس، آشنا، بامعنی و روشن باشد. بگونه‌ای که با حدود ۳۰۰ تا ۵۰۰ کلمه بتواند خواننده را به خواندن پایان‌نامه راغب نماید. چکیده، جدای از پایان‌نامه باید به تنهایی گویا و مستقل باشد. در چکیده باید از ذکر منابع، اشاره به جداول و نمودارها اجتناب شود. تمیز بودن مطلب، نداشتن غلط‌های املائی یا دستور زبانی و رعایت دقت و تسلسل روند نگارش چکیده از نکات مهم دیگری است که باید در نظر گرفته شود. در چکیده پایان‌نامه باید از درج مشخصات مربوط به پایان‌نامه خودداری شود. چکیده باید منعکس‌کننده اصل موضوع باشد. در چکیده باید اهداف تحقیق مورد توجه قرار گیرد. تأکید روی اطلاعات تازه (یافته‌ها) و اصطلاحات جدید یا نظریه‌ها، فرضیه‌ها، نتایج و پیشنهادها متمرکز شود. اگر در پایان‌نامه روش نوینی برای اولین بار ارائه می‌شود و تا به حال معمول نبوده است، با جزئیات بیشتری ذکر شود. شایان ذکر است چکیده فارسی و انگلیسی باید حتماً به تأیید استاد راهنما رسیده باشد.

کلمات کلیدی در انتهای چکیده فارسی و انگلیسی آورده می‌شود. محتوای چکیده‌ها بر اساس موضوع و گرایش تحقیق طبقه‌بندی می‌شود و به همین جهت وجود کلمات شاخص و کلیدی، مراکز اطلاعاتی را در طبقه‌بندی دقیق و سریع پایان‌نامه یاری می‌دهد. کلمات کلیدی، راهنمای نکات مهم موجود در پایان‌نامه هستند. بنابراین باید در حد امکان کلمه‌ها یا عباراتی انتخاب شود که ماهیت، محتوا و گرایش کار را به وضوح روشن نماید.

واژگان کلیدی حداکثر ۵ کلمه یا عبارت، متناسب با عنوان، قالب پایان‌نامه، لاتک

فهرست مطالب

پ	فهرست تصاویر
ت	فهرست جداول
ث	فهرست الگوریتم‌ها
ج	فهرست برنامه‌ها
۱	فصل ۱: مقدمه
۵	فصل ۲: مبانی تحقیق
۵	۱.۲ تنوع ژنتیکی
۸	۲.۲ تکامل تومور ^۱
۹	۳.۲ تکنولوژی‌های توالی‌یابی و فراوانی تغییرات آلل ^۲
۱۰	۴.۲ ناهمگنی ژنومی تومور
۱۳	۵.۲ بازسازی زیر کلونال
۱۵	۶.۲ تغییرات تعداد کپی
۱۷	۷.۲ جهش‌های ساده بدنی
۱۸	۸.۲ ترک آللی ^۳
۱۹	۹.۲ مقدمه‌ای بر مدل‌سازی احتمالی

¹Tumor Evolution

²Variant allele frequency

³Allele dropout

۱.۹.۲	زنجیره مارکوف مونت کارلو ^۴	۲۰
فصل ۳:	روش‌های پیشین	۲۳
فصل ۴:	روش پیشنهادی	۲۴
۱.۴	مقدمه	۲۴
۲.۴	معرفی دادگان ورودی	۲۴
۳.۴	روش پیشنهادی اول (درخت‌بازی)	۲۵
۱.۳.۴	پیش‌پردازش	۲۵
۱.۱.۳.۴	تصادفی	۲۵
فصل ۵:	نتایج تجربی	۲۷
فصل ۶:	بحث و نتیجه‌گیری	۲۸
مراجع		۲۹

⁴Markov Chain Monte Carlo (MCMC)

فهرست کارهای باقیمانده

فصل ۱

مقدمه

تومور^۱ از رشد غیر طبیعی سلول با احتمال حمله یا گسترش به سایر قسمت‌های بدن تشکیل می‌شود. تومورهای بدخیم^۲ معمولاً سرطان^۳ نامیده می‌شوند. سرطان علل مختلفی از جمله تغییرات ژنتیکی، آلودگی محیط زیست یا انتخاب‌های نادرست در سبک زندگی دارد. یک تومور ممکن است از زیرجمعیت‌های سلولی با تغییرات ژنومی مشخص تشکیل شده باشد، این پدیده ناهمگنی تومور^۴ نامیده می‌شود. ناهمگنی تومور احتمالاً برای درمان سرطان و کشف نشانگر زیستی، به ویژه در روش‌های درمانی هدفمند، تأثیراتی خواهد داشت [۱۹]. درمان‌های فعلی، سرطان را به عنوان یک بیماری همگن درمان می‌کنند [۳۸].

داروهای هدفمند در برابر زیرجمعیت‌های تک یا چند سلولی با انکوژن^۵ جهش‌یافته که آن‌ها را هدف قرار می‌دهند، تولید شده‌اند، در حالی که آن دسته از زیرجمعیت‌های سلولی که هیچ گونه تأثیری از داروهای به واسطه جهش خود، نمی‌گیرند بدون درمان باقی مانده و ممکن است منجر به عود مجدد تومور یا عدم درمان تومور می‌شوند [۱۹]. این زیرجمعیت‌های سلولی بدون درمان ممکن است منجر به پیشرفت تومور پس از درمان دارویی شوند [۱۹]. به عنوان مثال، رشد مجدد سلول‌های تومورزا در سرطان روده بزرگ^۶ سرطان پستان و گلیوبلاستوم^۷ پس از تابش یا درمان سیکلوفسفامید مشاهده شده است [۳۸]. بنابراین، مطالعه روند رشد تومور و ناهمگنی آن

¹Tumor

²Malignant tumor

³Cancer

⁴Tumor heterogeneity

⁵Oncogene

⁶Colorectal carcinoma

⁷Glioblastomas

تأثیرات زیادی بر تشخیص و درمان سرطان دارد.

تومورها می‌توانند خوش خیم، بدخیم و دارای رفتاری نامشخص یا ناشناخته باشند [۲]. تومورهای خوش خیم شامل فیبروئیدهای رحمی^۸ و خال‌های ملانوسیتیک^۹ است. آن‌ها محدود و محلی^{۱۰} هستند و به سرطان تبدیل نمی‌شوند [۴]. تومورهای بالقوه بدخیم^{۱۱} شامل سرطان در محل^{۱۲} هستند. آن‌ها به سایر بافت‌ها حمله نکرده و از بین نمی‌روند اما ممکن است به سرطان تبدیل شوند [۳]. تومورهای بدخیم را معمولاً سرطان می‌نامند. آن‌ها به بافت اطراف حمله کرده و از بین می‌روند، ممکن است متاستاز^{۱۳} ایجاد کنند و اگر درمان نشوند یا به درمان پاسخ ندهند، کشنده خواهد بود [۳].

ناهمگنی تومور توضیح می‌دهد که تومور بیش از یک نوع سلول شامل می‌شود. انواع مختلف سلول‌های داخل تومور دارای ویژگی‌های مورفولوژیکی و فیزیولوژیکی متمایزی مانند گیرنده‌های سطح سلول، تکثیر^{۱۴} و رگ‌زایی^{۱۵} هستند. ناهمگنی تومور می‌تواند بین تومورها (ناهمگنی بین توموری) و یا درون تومورها (ناهمگنی درون توموری) رخ دهد. به طور گسترده‌ای پذیرفته شده است که توسعه تومور یک روند تکاملی است [۹]، و پیشرونده^{۱۶} معمولاً از یک سلول منشأ می‌گیرند و گروهی از سلول‌ها را تشکیل می‌شوند که در نهایت یک توده را شکل می‌دهند.

دو مدل برای ناهمگنی تومور وجود دارد (شکل ۱.۱). یک مدل تشکیل سرطان از طریق سلول‌های بنیادی بوده که قابلیت ارث‌بری ندارند و مدل دیگر تشکیل سرطان از طریق تکامل کلونی^{۱۷} بوده که قابلیت ارث‌بری دارد. [۹]. مفهوم سلول‌های بنیادی سرطانی بیان می‌کند که رشد و پیشرفت بسیاری از تومورها توسط کسری کمی از سلول‌ها کنترل می‌شود و اکثر سلول‌های موجود در تومور محصولات تمایز غیر طبیعی سلول‌های بنیادی سرطانی هستند [۹]. بنابراین، برای توصیف و از بین بردن سلول‌های بدخیم در تومورها، لازم است که بر بخش کوچکی از سلول‌های تومور تمرکز کنیم [۲۷]. مفهوم تکامل کلونی بیان می‌کند که تومور از یک سلول طبیعی ژنتیکی بوجود می‌آید که به تعداد زیادی سلول تبدیل می‌شود. در این تکامل، جهش‌های تصادفی به طور مداوم تولید می‌شوند و

⁸Uterine fibroid

⁹Melanocytic nevi

¹⁰Local

¹¹Potentially malignant tumor

¹²Carcinoma In Situ

¹³Metastases

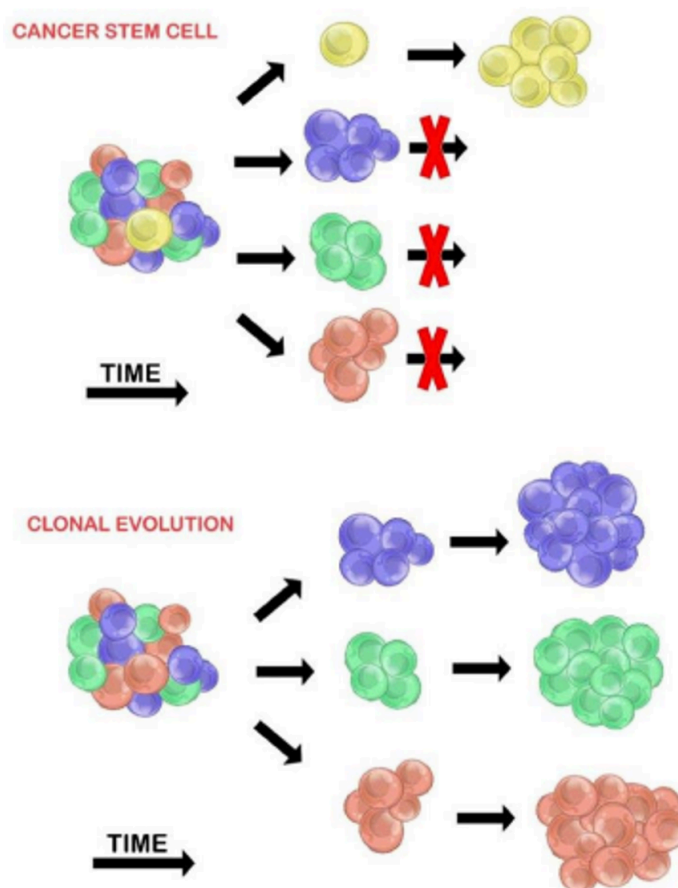
¹⁴Proliferative

¹⁵Angiogenic

¹⁶Spontaneous

¹⁷Clonal

در نهایت تومور حاصل میلیاردها سلول بدخیم است که حاصل از تجمع تعداد زیادی جهش است [۲۴]. تکامل تومور به عنوان توالی پیدرپی گسترش کلونی توصیف می‌شود، که در آن در هر حالت جدید یک رویداد جهش اضافی ایجاد می‌شود [۹].



شکل ۱.۱: دو مدل برای ناهمگونی تومور

یکی از توالی‌های پی در پی گسترش کلونی، یک مدل خطی از جانشینی کلونی است، جایی که جهش‌های متوالی پیدرپی باعث ایجاد توالی خطی از مجموعه‌های گسترش کلون می‌شوند و منجر به رشد کلون می‌شوند [۹]. مورد دیگر یک مدل چند کلونی از پیشرفت تومور است، که در آن یک سلول منفرد از طریق مکانیزم تقسیم به چندین زیرکلون گسترش می‌یابد [۳۰]. این مدل بیش از مدل خطی با ناهمگنی تومور مرتبط است. جهش‌های اکتسابی منجر به افزایش بی‌ثباتی ژنومی با هر نسل متوالی می‌شود [۱۳].

تومورهای ناهمگن^{۱۸} که متشکل از چندین کلون هستند، می‌توانند حساسیت‌های مختلفی را نسبت به داروهای سمیت سلولی^{۱۹} در نشان دهند. علاوه بر این، می‌زان ناهمگنی تومور می‌تواند خود به عنوان نشانگر زیستی^{۲۰} مورد استفاده قرار گیرد زیرا هر چقدر می‌زان ناهمگنی تومور بیشتر باشد، احتمال حضور کلون‌های مقاوم در برابر درمان بیشتر است [۴۰]. دلایل حساسیت‌های مختلف می‌تواند تعاملات بین کلون‌ها باشد که ممکن است اثر درمانی را مهار یا تغییر دهد [۹]. تومورهایی با ناهمگنی زیاد، با احتمال بیشتری از کلون‌های گوناگون تشکیل شده است که به درمان مقاوم هستند و ممکن است منجر به عدم موفقیت در درمان شوند. روش‌های نوین درمان تومورها با هدف شخصی‌سازی برنامه‌های درمانی از طریق هدف قرار دادن جمعیت‌های سلولی توموری موجود در یک بیمار، توسعه می‌یابند [۱۸]. ناهمگنی‌های توموری یکی از عوامل اصلی مقاومت در برابر دارو است و بنابراین، یک عامل بالقوه در شکست درمان محسوب می‌شود. [۱۸]. تومورها می‌توانند از راه‌های مختلف به طور همزمان به مقاومت دارویی دست یابند، بنابراین هدف قرار دادن فقط یک مکانیسم مقاومت برای غلبه بر نارسایی درمانی، می‌تواند مزیت درمان‌های هدفمند را محدود کند [۱۱]. بنابراین، ناهمگنی تومور می‌تواند برای درک توسعه تومور، پیچیدگی ایجاد کند و توسعه روش‌های موفقیت آمیز را با چالش روبرو کند [۱۸]. مطالعه ناهمگنی تومور می‌تواند منجر به پیشرفت و توسعه روش‌های درمانی شخصی سازی شده شوند و درک ما را از روابط عملکردی بین کلون‌ها در طول درمان افزایش دهند [۱۱]. برای مطالعه ناهمگنی تومور، بسیاری از ابزارهای محاسباتی موثر برای تجزیه و تحلیل اطلاعات کلونی تومور و تاریخچه تکامل آن تولید شده است. این ابزارها با استفاده از داده‌های تغییرپذیری ژنتیکی، تولید شده توسط فناوری‌های توالی‌یابی نسبتاً دقیق، قادر هستند تا ترکیب‌های کلونی تومور و رابطه اجداد بین کلون‌ها نتیجه دهند. این اطلاعات برای درک پیشرفت تومور و کمک به پیشرفت‌های درمانی کارآمد مهم است.

در ادامه مفاهیم حوزه تحقیق مثل مدل‌های ناهمگنی توموری، روش‌های مختلف توالی‌یابی، روش‌های مختلف ساخت درخت فیلوژنی تومور، مباحث مرتبط به یادگیری عمیق و یادگیری تقویتی به اختصار توضیح داده شد. در فصل سوم تحقیق پیشرو، به بررسی الگوریتم‌هایی که با استفاده از داده‌های توالی‌یابی تکسلولی، درخت فیلوژنی تومور را استنباط کرده‌اند پرداخته شد. هر یک از این روش‌ها برای ساخت درخت فیلوژنی به همراه دادگان مورد استفاده، مورد ارزیابی قرار گرفت و در انت‌های فصل سوم مقایسه‌های بین روش‌های مختلف صورت گرفت. در فصل چهارم روش پیشنهادی استنباط درخت فیلوژنی بر مبنای یادگیری تقویتی و داده‌های

¹⁸Heterogenetic¹⁹Cytotoxic²⁰Biomarker

توالی‌یابی تک‌سلولی به تفصیل بیان شده و در فصل پایانی نتایج بدست آمده و مقایسه آن با نتایج پیشین، گزارش شده است. در پایان موضوعات پیشنهادی که در کارهای آتی در راستای ادامه این پژوهش می‌تواند مورد بررسی قرار گیرند، توضیح داده شد.

فصل ۲

مبانی تحقیق

در این فصل ابتدا مفاهیم مورد نیاز جهت تعریف مسئله مانند مدل-های ناهمگنی تومور، روش-های یافتن درخت تکاملی تومور، روش-های توالی-یابی داده مورد بررسی قرار می-گیرند. در ادامه مدل-های مورد استفاده برای استنباط درخت تکاملی تومور معرفی می-شوند. در پایان مفاهیم مرتبط با یادگیری ماشینی، یادگیری عمیق و یادگیری تقویتی به منظور استنباط درخت تکاملی تومور با رویکرد مبتنی بر داده^۱ توضیح داده می-شوند.

۱.۲ تنوع ژنتیکی

دی‌ان‌ای^۲ یک مولکول بیولوژیکی است که توسط نوکلئوتیدها^۳ پلیمری شده است. در دی‌ان‌ای چهار نوع نوکلئوتید وجود دارد: آدنین^۴ (A)، تیمین^۵ (T)، سیتوزین^۶ (C) و گوانین^۷ (G). دی‌ان‌ای اساس توالی اسیدهای آمینه است که پروتئین را تشکیل می‌دهد. یک مولکول دی‌ان‌ای از دو رشته تشکیل شده است. که در موازات^۸ هم و در جهت-های مخالف قرار دارند و ساختاری از مارپیچ دوتایی ایجاد می‌کنند. هر نوع نوکلئوتید روی یک رشته

¹Data driven

²DNA

³Nucleotid

⁴Adenine

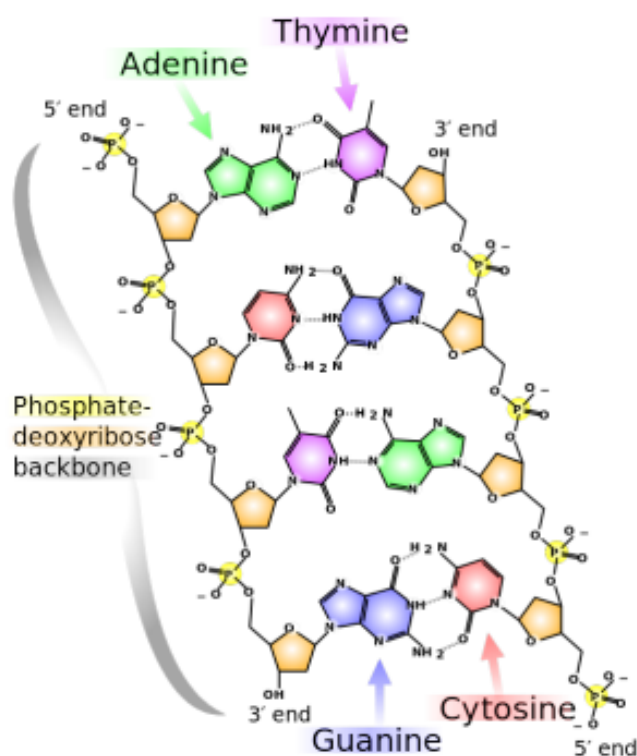
⁵Thymine

⁶Cytosine

⁷Guanine

⁸Antiparallel

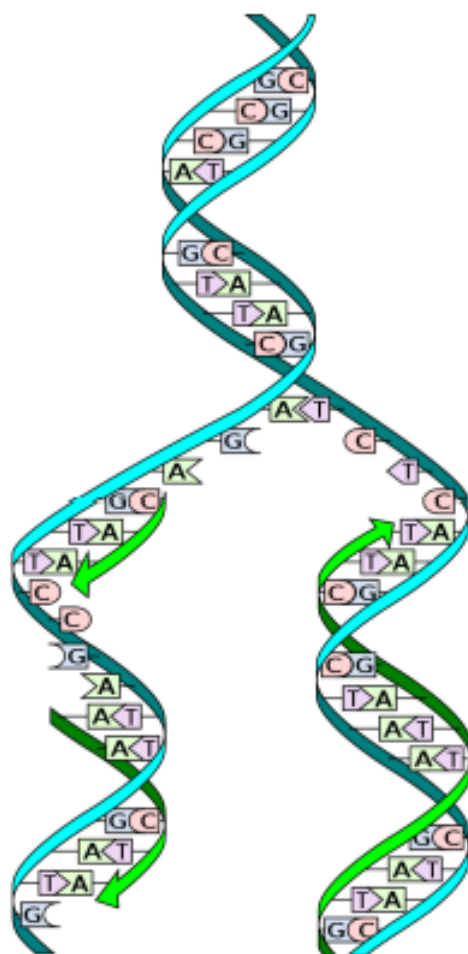
با نوع دیگری از نوکلئوتید در رشته دیگر مرتبط است: A با T ؛ C با G (شکل ۱.۲) [۵]. این به عنوان قانون پایه جفت شدن نوکلئوتیدها در هر رشته از دی-ان-ای شناخته می‌شود.



شکل ۱.۲: مارپیچ دوگانه دی‌ان‌ای

همانند سازی دی‌ان‌ای فرآیند تولید دو مولکول دی‌ان‌ای یکسان از مولکول دی‌ان‌ای اصلی است. وقتی تکثیر شروع می‌شود، دو رشته یک مولکول دی‌ان‌ای از یکدیگر جدا می‌شوند و هر رشته به عنوان الگویی برای ساخت نمونه مشابه خود عمل می‌کند. نوکلئوتیدها در هر موقعیت از یک رشته با نوع دیگری از نوکلئوتید مبتنی بر قانون پایه جفت شدن، به منظور سنتز همتای این رشته، متصل می‌شود. پس از همانند سازی، مولکول دی‌ان‌ای اصلی به دو مولکول یکسان تبدیل می‌شود (شکل ۲.۲) [۵].

ژن ناحیه‌ای از دی‌ان‌ای است و به عنوان مولکول واحد وراثت شناخته می‌شود. ژن‌های متعددی در ساختار دی‌ان‌ای با عملکردهای متفاوت وجود دارد. جهش به تغییر دائمی توالی هسته‌ای ژنوم اطلاق می‌شود. جهش‌ها می‌توانند در حین فرآیند تکثیر دی‌ان‌ای و با جفت‌گیری اشتباه در قسمت‌های



شکل ۲.۲: همانندسازی دی‌ان‌ای

مختلف دی-ان-ای ایجاد می-شود. انواع مختلفی از جهش-ها مانند جهش تک نوکلئوتیدی^۹ (جهش نقطه‌ای^{۱۰}) (شکل ۳.۲) و تغییرات ساختاری^{۱۱} شامل درج^{۱۲}، حذف^{۱۳} و برگشت^{۱۴} (شکل ۴.۲) وجود دارد. جهش-های سلولی می-توانند به بنا بر دلایلی چون مواد شیمیایی، سمیت یا ویروس ایجاد شوند. جهش در یک ژن می‌تواند محصولات آن را تغییر دهد (مانند ایجاد پروتئین متفاوت) یا از عملکرد صحیح ژن جلوگیری کند [۵].

^۹Single nucleotide mutation

^{۱۰}Point mutation

^{۱۱}Single variant

^{۱۲}Insertion

^{۱۳}Deletion

^{۱۴}reversion

original sequence:

ACTTGGTCAGAAATTCCCAGGTGTCA

point mutation:

ACTTGGTCATAATTCCCAGGTGTCA

شکل ۳.۲: جهش تک‌نوکلئوتیدی

insertion:

ACTTGGTCAGAAATTCCCAGGTGTCA



ACTTGGTCAGATAGGCAATTCCCAGGTGTCA

deletion:

ACTTGGTCAGAAATTCCCAGGTGTCA



ACTTGGTCACCCAGGTGTCA

reversion:

ACTTGGTCAGAAATTCCCAGGTGTCA



ACTTGGTCCTAAGACCCAGGTGTCA

شکل ۴.۲: تغییرات ساختاری

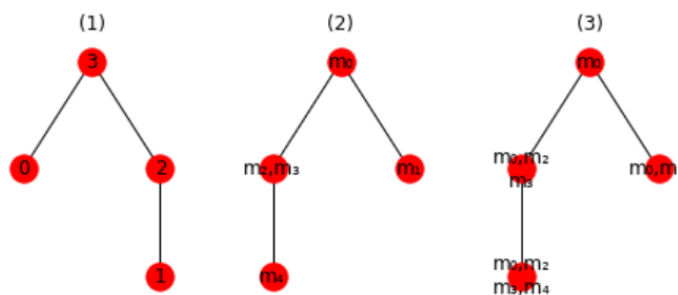
۲.۲ تکامل تومور

جهشی که در هر سلول از بدن اتفاق می‌افتد، به استثنای سلول‌های جنسی (اسپرم و تخمک)، جهش جسمی^{۱۵} نامیده می‌شود [۱]. تجمع جهش بدنی در طول زندگی یک فرد می‌تواند منجر به رشد کنترل نشده مجموعه‌ای از سلول (تومور) شود [۳۲] و می‌تواند باعث شکل‌گیری سرطان یا بیماری‌های دیگر شود [۱]. بدلیل تجمع سلول‌های گوناگون، بیش از یک نوع سلول در تومور وجود خواهد داشت. به گروه‌های سلول با مجموعه‌ای از جهش مشخص، کلون یا جمعیت سلولی تومور گفته می‌شود. کلون‌های موجود در تومور از نظر فیلوژنتیک با

¹⁵somatic

هم مرتبط هستند و رابطه آنها را می‌توان با یک درخت فیلوژنتیک نشان داد [۹]. درخت فیلوژنتیک رابطه تکاملی بین کلون و ترتیب وقوع هر جهش را نشان می‌دهد. به عنوان مثال، شکل ۵.۲:

- یک درخت فیلوژنتیک از یک تومور با چهار کلون با برچسب ۰ تا ۳ را نشان می‌دهد.
 - جهش جدیدی را نشان می‌دهد که در هر کلون در طول تکامل این تومور رخ داده است.
- همچنین هر کلون جهشی را در مسیر از کلون بالایی به سمت خود به ارث می‌برد. به عنوان مثال، کلون ۰ جهش‌های m^0 ، m^1 دارد. کلون ۱ دارای جهش m^0 ، m^2 ، m^3 ، m^4 است.



شکل ۵.۲: درخت فیلوژنتیک تومور

۳.۲ تکنولوژی‌های توالی‌یابی و فراوانی تغییرات آل

تعیین توالی دی‌ان‌ای روشی برای تشخیص ترتیب دقیق نوکلئوتیدها در یک رشته دی‌ان‌ای است. روش توالی‌یابی نسل بعدی^{۱۶} از تعدادی فناوری مدرن توالی تشکیل شده است که امکان تعیین هزینه و زمان توالی‌یابی را به طور موثر فراهم می‌کند. با استفاده از نمونه بیولوژیکی به عنوان ورودی این تکنولوژی‌ها، توالی‌های کوتاه نوکلئوتیدی تولید می‌شود (که به آن خوانش^{۱۷} گفته می‌شود). سپس خوانش با استفاده از الگوریتم هم‌ترازی^{۱۸} متنوعی مانند الگوریتم تبدیل Burrows-Wheeler با ژنوم مرجع تراز می‌شوند. پس از ترازبندی، می‌توان با جمع‌آوری خوانش‌های همپوشانی^{۱۹}، توالی اجماعی^{۲۰} ایجاد کرد (شکل ۶.۲). در موقعیتی از توالی اجماع به

¹⁶Next generation sequencing

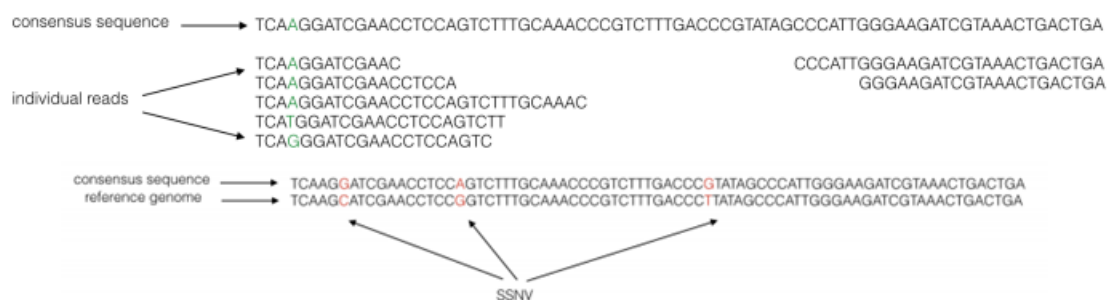
¹⁷Read

¹⁸Alignment

¹⁹Overlapping read

²⁰Consensus

دلیل همپوشانی خوانش‌ها، ممکن است بیش از یک نوع خوانش از نوکلئوتید تراز شده وجود داشته باشد (تعداد کل قرائت مرتبط با یک نوع جهش، را پوشش خوانش^{۲۱} نامیده می‌شود). نوکلئوتید موجود در این موقعیت به عنوان رایج-ترین نوکلئوتید تراز شده، مشخص می‌شود. به عنوان مثال، در شکل ۶.۲، سه آدنین (A)، یک گوانین (G) و یک تیمین (T) در موقعیت سوم توالی اجماع تراز می‌شوند، سپس نوکلئوتید در آن موقعیت به عنوان آدنین (A) تعیین می‌شود. پس از ایجاد توالی اجماع، نوکلئوتیدهای موجود در آن توالی، که متفاوت از ژنوم مرجع هستند، شناسایی شده و به عنوان تغییرات بدنی تک نوکلئوتیدی^{۲۲} شناخته می‌شود. با استفاده از نمونه‌های متعدد استخراج شده از یک نمونه تومور، ما می‌توانیم تغییرات بدنی تک نوکلئوتیدی را در هر نمونه با فناوری تعیین توالی-یابی تشخیص دهیم. نسبت تعداد سلول‌های موجود در یک نمونه حاوی تغییرات بدنی تک نوکلئوتیدی به کل سلول‌ها، فراوانی تغییرات آلل یک تغییر بدنی تک نوکلئوتیدی در این نمونه نامیده می‌شود. مقادیر فراوانی تغییرات آلل برای هر تغییر بدنی تک نوکلئوتیدی در هر نمونه تومور قابل محاسبه است. ابزارهای زیادی برای بازسازی درخت فیلوژنتیک تومور از مقادیر فراوانی تغییرات آلل تومور به عنوان ورودی الگوریتم استفاده می‌کنند.



شکل ۶.۲: تشخیص تغییر بدنی تک نوکلئوتیدی از طریق خوانش هم‌ترازی

²¹Read coverage

²²Somatic single nucleotide variation

۴.۲ ناهمگنی ژنومی تومور

سرطان بیماری-ای است که بدلیل ایجاد ناهنجاری-های اساسی در فرآیند-های بنیادی سلول مانند تکثیر^{۲۳}، تمایز^{۲۴} و مرگ^{۲۵} سلول ایجاد می-شود [۲۶]. این ناهنجاری منجر به رشد کنترل نشده تومور و به-کارگیری بافت غیرسرطانی برای حمایت از این رشد می شود. علت اصلی این تغییرات جهش است. جهش یک اصطلاح گسترده است که چندین دسته از تغییرات ژنتیکی را پوشش می-دهد. هنگام حاملگی، یک جنین دارای یک ژنوم خاص و منحصر به فرد است. این ژنوم که به ژنوم جوانه‌زنی^{۲۶} معروف است، می-تواند با ژنوم انسانی مرجع مقایسه شود. ژنوم انسانی مرجع یک نمونه از ژنوم انسان است و از دی‌ان‌ای چند نفر تشکیل شده است. تفاوت بین ژنوم جوانه‌زنی و ژنوم مرجع به عنوان جهش ژنوم جوانه‌زنی شناخته می-شود. جهش-های جوانه‌زنی می-توانند مسئول افزایش خطر ابتلا به سرطان باشند [۳۶]، اما بندرت خود مسئول مستقیم توسعه تومور هستند. معمولاً تومورها در اثر جهش-های اکتساب شده پس از لقاح، که معروف به جهش-های بدنی هستند، ایجاد می شوند. جهش-های بدنی نتیجه اشتباهات در تکثیر دی‌ان‌ای [۸]، قرار گرفتن در معرض جهش-های با منشأ داخلی یا خارجی یا وارد-شدن توالی-های دی‌ان‌ای با منشأ بیرونی بدلیل قرار گرفتن در معرض ویروس است [۳۹]. غالباً در سرطان، جهش‌های بدنی باعث ایجاد اختلال در روند تکثیر دی‌ان‌ای یا ترمیم آن می شوند و حتی جهش‌های بدنی بیشتری ایجاد می کنند [۳۷]. نظریه کلونی بودن سرطان [۳۲] سرطان را به عنوان یک تک سلولی با منشأ غیرجنسی در نظر می-گیرد که در اثر تولید مثل فراوان، یک توده متشکل از کلون-های سلولی گوناگون را ایجاد می-کند. در این مدل سلولهای توموری با یکدیگر در رقابت هستند و جهش-های بدنی که مزیت رشد را ایجاد می کنند در جمعیت سلول-های توموری از نسبت بیشتری برخوردار خواهند بود. جهش-های بدنی که باعث رشد تومور شده و از سلولی به سلولی دیگر منتقل می-شوند به عنوان جهش‌های راننده^{۲۷} شناخته می-شوند. اولین سلولی که دارای جهش راننده بوده و آن را به جهش-های بعدی منتقل می-کند به عنوان سلول بنیانگذار شناخته می شود. همه فرزندان این سلول بنیانگذار، جهش راننده و هر جهش دیگری را که سلول بنیانگذار قبل از به دست آوردن جهش راننده بدست آورده است، دارند. این جهش-های دیگر، که مزیتی

²³Replication

²⁴Differentiation

²⁵Death

²⁶Germline genome

²⁷Driver mutation

برای رشد و گسترش تنوع توموری ندارند، به عنوان جهش‌های مسافر^{۲۸} شناخته می‌شوند. شایان ذکر است که تعریف جهش راننده و مسافر به زمینه ژنتیکی و محیطی بستگی دارد. به عنوان مثال، شیمی درمانی داروهای سمیت سلولی (سیتوتوکسیک) می‌تواند باعث تغییر جهش از مسافر به جهش راننده شود و عامل اصلی مقاومت در برابر درمان باشد. همچنین جهش‌ها را می‌توان بر اساس نوع تغییری که در دی‌ان‌ای ایجاد می‌شود، به طبقات متمایز تقسیم کرد. حذف و تغییر تک‌نوکلئوتیدها^{۲۹} جهش‌هایی هستند که یک پایه در ژنوم را به پایه دیگری تغییر می‌دهند. ایندل^{۳۰} درج یا حذف یک بخش دی‌ان‌ای است که می‌تواند کوتاه یا طولانی باشد. از ایندل کوتاه و تغییرات تک‌نوکلئوتیدی در مجموع به عنوان جهش‌های ساده بدنی^{۳۱} یاد می‌شود. در همه قسمت‌های یک ژنوم، از جمله کل کروموزوم‌ها، قابلیت حذف یا کپی شدن قسمتی از ژنوم وجود دارد. تغییرات شماره کپی به جهشی اتلاق می‌شود که منجر به حذف یا کپی شدن قسمتی از ژنوم می‌شود. تغییرات شماره کپی^{۳۲} نوعی تغییر ساختاری هستند که شامل وارونگی (وقتی قسمت بزرگی از ژنوم معکوس شده باشد) و انتقال متعادل (جایی که دو بخش ژنومی مکان‌های خود را با یکدیگر تعویض می‌کنند) می‌باشند [۳۷]. این گونه‌های مختلف جهش مستقل از یکدیگر نیستند و می‌توانند در رابطه با یکدیگر اتفاق بیفتند (به عنوان مثال یک جهش می‌تواند منجر به تقویت یک وارونگی شود).

تکنیک توالی‌یابی نسل بعدی این امکان را فراهم کرده است تا با صرف هزینه بسیار کم و با استفاده از یک نمونه توموری، توالی‌یابی از دی‌ان‌ای صورت پذیرد و همین امر منجر به تحول گسترده‌ای در زمینه مطالعه تکامل تومور شده زیر امکان نمونه‌برداری در تعداد بسیار بالا را از تومور فراهم می‌کند. نمونه‌گیری در حجم بالا این امکان را فراهم آورده است تا ناهمگنی تومور از نقطه منظر ژنتیکی مورد بررسی قرار گیرد و پاسخ به درمان بیماران سرطانی با جزییات بیشتری مورد ارزیابی قرار گیرد.

تقریباً همه نمونه‌های استخراج شده از تومور ترکیبی از سلول‌ها با ژنوتیپ‌های مختلف را شامل می‌شود. یک نمونه توموری به ندرت فقط شامل بافت سرطانی است زیرا شامل سلول‌های غیر سرطانی از استرومای اطراف^{۳۳} یا سلول‌های ایمنی نفوذی^{۳۴} است. مطالعات ژنومیک نشان داده است که حتی در میان سلولهای سرطانی، غالباً زیرجمعیت‌های متعدد سرطانی نیز وجود دارد. به عنوان مثال، در یک مطالعه مهم در سال ۲۰۱۲

²⁸Passenger mutation²⁹Single nucleotide variants (SNV)³⁰Indel³¹Single Somatic Mutation³²Copy number alteration³³Surrounding stroma³⁴Infiltrating immune cell

، گرلینگر و همکارانش [۲۲] توالی-یابی ژنوم و تغییرات شماره کپی را از طریق نمونه-های مکانی مجزا استخراج شده از سرطان کلیه اولیه و نقاط متاستاز ثانویه بدست آورده-اند. با بررسی این نمونه-های متعدد، مشخص شد که یک ناهمگنی ژنتیکی قابل توجهی در تومور وجود دارد. تعداد بسیار زیادی از جهش-های شناسایی شده در همه سلول-های توموری مشاهده نشدند و این بدان معناست که این جهش-ها بیش از آن-که یک ناحیه کلونی باشند، به صورت یک ناحیه زیر کلونی بوده-اند. با استفاده از روش-های پردازش غیراتوماتیک، تغییرات تک نوکلئوتیدی ها و تغییرات شماره کپی بر اساس نمونه-هایی که از آن استخراج شده-اند، به خوشه-های مجزا دسته-بندی شده و یک درخت فیلوژنی به آن-ها نسبت داده شد. بازسازی درخت فیلوژنیک تومور این امکان را فراهم آورد تا سیر تکاملی تومور با استفاده از شاخه-های مختلف درخت فیلوژنی شامل جهش-هایی با عملکرد یکسان از سه ژن متفاوت مورد بررسی قرار گیرد.

در همان سال، یک مطالعه مهم دیگر، "تاریخچه زندگی ۲۱ سرطان پستان" [۳۱]، حضور ITH را نیز نشان داد. در این مطالعه آنها توالی-یابی کامل ژنوم را در عمق متوسط 188X بر روی تومور پستان PD4120a انجام دادند. این عمق اجازه می‌دهد تا جمعیت‌های شیوع تا ۵٪ کم باشد. آنها مشاهده کردند که تغییرات تک نوکلئوتیدی‌ها در تعداد کمی از خوشه‌های مجزا مشاهده می‌شوند که با توجه به کسر نوع آلل (VAF) آنها مشاهده می‌شود، نسبت خواندن‌ها در یک مکان متفاوت شامل آلل نوع. علاوه بر این، آنها توانستند نشان دهند که برخی از این خوشه‌های مجزا را نمی‌توان با جهش‌های موجود در تمام جمعیت‌های سرطانی توضیح داد، که این نشان دهنده حضور تغییرات تک نوکلئوتیدی‌های تحت کلونال است. در همان زمان، آنها دریافتند که بسیاری از جهش‌ها در تمام سلول‌های سرطانی موجود در نمونه وجود دارد، که نشان می‌دهد جد مشترک اخیر نسبتاً دیر در زمان تکامل رشد کرده است. مشاهده اینکه جهش‌های زیر کلونال به جای توزیع یکنواخت یا مطابق قانون قدرت در خوشه‌های متمایز پیدا شده است، شواهدی را نشان می‌دهد که این جهش‌های زیرکلونالی بیش از آنکه ناشی از تکامل خنثی یا مصنوعات فنی باشد، در زیر مجموعه‌های متمایز ناشی از فشارهای انتخابی یافت می‌شود. نویسندگان همچنین با تأیید اینکه جهش‌های زیر کلونال محدود به تغییرات تک نوکلئوتیدی نیستند، توانستند حضور تغییرات شماره کپی‌های کلونال و زیرکلونال را تأیید کنند. نویسندگان یک الگوریتم خوشه‌بندی غیر پارامتریک (یک مدل مخلوط فرآیند دیریشله (DPMM)) را با استدلال قابل توجه دستی برای استنباط فیلوژنی شاخه‌ای از چهار زیر جمعیت سرطانی در آن نمونه منفرد تومور ترکیب کردند. درک معماری ژنتیکی این زیرجمعیت‌ها می‌تواند به مطالعه زیست شناسی سرطان کمک کند و نشان داده شده است که در پیش‌بینی بقا در بسیاری از انواع سرطان مفید است [۷]. به عنوان مثال، زیرجمعیت‌های مختلف، که توسط مجموعه جهش‌های جسمی حمل شده تعریف

می‌شوند، توانایی‌های مختلفی در مقاومت در برابر درمان و متاستاز دارند. برای انجام این کار، باید از یک یا تعداد کمی از نمونه‌های تومور فله، ژنوتیپ‌های موجود در نمونه را شناسایی کرد. این مسئله، تحت عنوان بازسازی ساب کلونال، موضوع اصلی این پایان‌نامه است. مطالعات پیشگام که نشان داد ITH برای انجام این بازسازی به استدلال دستی قابل توجهی نیاز دارد. استدلال دستی کند، مستعد خطا است و به تخصص قابل توجهی نیاز دارد. مزایای بازسازی کاملاً خودکار بدیهی است. این بخش پیش زمینه مشکل بازسازی زیر کلونال، چگونگی پرداختن به آن برای انواع مختلف جهش، خصوصیات اصلی الگوریتم‌های بازسازی زیر کلونال و خلاصه‌ای از کارهای موجود در این زمینه را توصیف می‌کند.

۵.۲ بازسازی زیر کلونال

بازسازی ساب کلونال سعی دارد ژنوتیپ‌های موجود در تومور را از تعداد کمی از نمونه‌های توالی دی‌ان‌ای از آن تومور استنباط کند. تعداد ژنوتیپ‌های موجود در تومور از قبل مشخص نیست. این ژنوتیپ‌های زیر کلونال به طور معمول با جهش‌هایی که در مقایسه با ژنوم خط جوانه‌ای دارند، توصیف می‌شوند. ژنوم جوانه‌زنی علاوه بر نمونه(های) تومور، با تعیین توالی یک نمونه غیرسرطانی تعیین می‌شود. در حال حاضر در هنگام تعریف این جمعیت از دو نوع جهش به طور معمول استفاده می‌شود: جهش‌های ساده بدنی‌های متشکل از تعویض‌ها و درج / حذف کوچک (ایندل) و CNA حاصل از تغییرات ساختاری بزرگتر. مشاهده انواع جهش‌های دیگر، مانند مجموعه گسترده‌ای از SVها که شامل بازآرایی هستند، مشاهده آنها دشوارتر است و روش‌های شناسایی آنها در مراحل اولیه رشد است.

به طور متوسط، حتی در شرایط ایده آل، هر سلول در هر بخش یک جهش پیدا می‌کند [۸]، به همین ترتیب، بیشتر سلول‌های تومور ژنوتیپ منحصر به فردی خواهند داشت. بنابراین، به طور دقیق، اکثر سلول‌های تومور می‌توانند به طور بالقوه نمایانگر زیرجمعیت منحصر به فرد خود باشند. با این حال، به طور عملی، جهش‌هایی که مختص سلول‌های منفرد است یا فقط تعداد کمی از سلول‌ها آنها را به اشتراک می‌گذارد، در حین فراخوانی نوع شناسایی نمی‌شوند. تماس متغیر در بخش ۳.۵.۲ بیشتر مورد بحث قرار گرفته است. بعلاوه، سلول‌هایی که بخش عمده‌ای از جهش‌های خود را به اشتراک می‌گذارند، خصوصاً جهش‌های راننده، صفات مشابهی دارند. به همین ترتیب، من قرارداد گسترده‌ای را اتخاذ کرده و یک زیر جمعیت را به عنوان تمام سلول‌هایی که دارای زیر

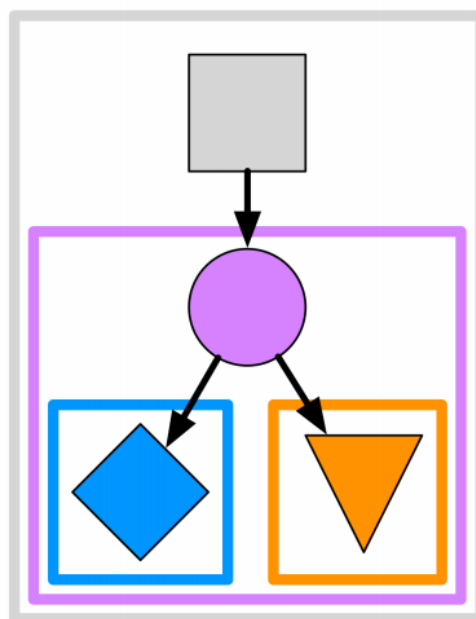
مجموعه یکسان جهش‌های بدنی در هنگام فراخوانی نوع هستند، تعریف می‌کنم.

یک گام مهم در بازسازی ساب کلونال محاسبه شیوع سلولی تبارهای زیر کلونال و سپس، در نهایت، زیر جمعیت‌های سرطانی است. شیوع سلولی یک زیر جمعیت، نسبت سلول‌های نمونه توالی شده متعلق به آن است. غالباً، شیوع سلولی با تقسیم بر خلوص نمونه، یعنی نسبت سلولهای سرطانی در نمونه، به بخش سلولهای سرطانی، نسبت سلولهای سرطانی، تبدیل می‌شود. هر سلول دقیقاً به یک زیر مجموعه تعلق دارد، بنابراین این شیوع باید در یک جمع باشد. به طور کلی، سلول‌های غیر سرطانی در یک زیر مجموعه واحد قرار می‌گیرند. با این حال، از آنجا که جهش‌ها اغلب در زیر جمعیت‌های متعدد وجود دارند، شیوع سلولی بسیاری از زیر جمعیت‌ها را نمی‌توان مستقیماً از جهش‌های آن استنباط کرد. برای پرداختن به این موضوع، ما یک نسب زیر کلونال برای یک جهش به عنوان مجموعه زیر جمعیت‌هایی که در آن وجود دارد، تعریف می‌کنیم. به طور رسمی، دودمان‌های زیر کلونال از زیر جمعیت بنیانگذار تشکیل می‌شود (جایی که جهش برای اولین بار ظاهر می‌شود) و همه زیر جمعیت‌های بعدی آن (که وراثت جهش) علاوه بر جهش‌های خاص خود، این زیر مجموعه‌های فرزندی حاوی تمام جهش‌های موجود در نژاد تعریف کننده زیر جمعیت هستند (به جز در صورت حذف محل منبع جهش، برای جزئیات بیشتر به فصل ۳ مراجعه کنید). نسب مربوط به یک زیر درخت (یا کلاد) از درخت کلون تومور است. شیوع سلولی یک تبار مجموع شیوع سلولی زیر جمعیت‌هایی است که متعلق به آن تبار هستند. از آنجا که سلول‌ها می‌توانند در چندین نژاد زیرکلونال وجود داشته باشند، شیوع نسب در یک جمع نیست.

شکل ۷.۲ تصویری از یک درخت کلون نمونه را ارائه می‌دهد. گره‌های موجود در درخت، همانطور که در بالا تعریف شد، نشان دهنده زیر جمعیت است. فلش‌ها از جمعیت والدین به سمت فرزندان نشان هدایت می‌شوند. دودمانهای زیر کلونال به صورت مستطیل نشان داده می‌شوند و با توجه به زیر مجموعه بنیادی آنها که در ریشه تیغه یافت می‌شوند، رنگی هستند.

۶.۲ تغییرات تعداد کپی

بیشتر ژنوم انسان دیپلوئید است، به این معنی که دو نسخه از توالی دی‌ان‌ای ما در سلول‌های ما وجود دارد، یکی از پدر و دیگری از مادر. تغییرات شماره کپی این تغییر را می‌دهند، یا با تغییر در تعداد نسخه‌ها (مثلاً از طریق تکثیر کل ژنوم)، نسبت کپی‌های مادر به پدر (مثلاً از دست دادن خنثی هتروزیگوزیته در تعداد کپی‌ها،



شکل ۷.۲: درخت کلون تومور

جایی که برای همان منطقه یک ژنوم والدین تکثیر می‌شود و دیگری حذف شده است) یا هر دو (به عنوان مثال کپی کروموزوم مادر). بیشتر این تغییرات (به استثنای تکثیر کل ژنوم) دامنه محدودی از ژنوم را تحت تأثیر قرار می‌دهد، اما می‌تواند از تأثیر یک ژن تا یک کروموزوم کامل باشد. این بخش از ژنوم تغییر یافته به عنوان یک بخش شناخته می‌شود.

تغییرات شماره کپی می‌توانند تعداد کپی کل یک بخش و / یا تعداد نسبی والدین را تغییر دهند. هر یک از این تغییرات توسط توالی-یابی ژنومی هسته قابل تشخیص است. تغییر در تعداد کپی کل یک بخش را می‌توان تشخیص داد زیرا نسبت خواندن آن نقشه به آن بخش بین خط خوانه زنی و نمونه تومور متفاوت خواهد بود. بخش از یک قطعه نسبت ورود خوانده شده است که به یک قطعه در یک نمونه غیر سرطانی ترسیم شده است به نسبت خوانده شده که به یک بخش در یک نمونه سرطانی ترسیم شده است. از نسبت نسبت‌ها برای محاسبه این واقعیت استفاده می‌شود که تعداد کل قرائت‌ها اغلب بین توالی یابی سرطانی و غیرسرطانی متفاوت است، در مناطق مختلف ژنوم عمق خواندن بیشتر یا پایین تر ناشی از محتوای GC یا نقشه برداری وجود دارد و تردستی یک تومور با بافت طبیعی متفاوت است. تکرار یک ژنوم، میانگین تعداد کپی از هر کروموزوم است که برای طول کروموزوم نرمال می‌شود.

با تغییر در کسر آلل می‌توان عدم تعادل در تعداد نسخه‌های مادری و پدری این بخش را تشخیص داد. در مناطق دیپلوئید ژنوم‌ها، اگر یک بازه بین کپی‌های مادر و پدر متفاوت باشد، موقعیت هتروزیگوت نامیده می‌شود. جهش‌های تک پایه، خط جوانه زنی همچنین به عنوان چند شکلی تک هسته‌ای نامیده می‌شوند. وقتی یک ژنوم توالی یابی شود، حدود نیمی از قرائت آن مکان هتروزیگوت حاوی هر یک از بازها خواهد بود، در نتیجه کسر آلل ۵۰ است. این امر تا زمانی که نسبتی برابر با نسخه‌های مادرانه و پدری وجود داشته باشد، صادق خواهد بود. اگر این نسبت تغییر کند، کسر آلل تمام پولیمورفیسم تک هسته‌ای در بخش آسیب دیده تغییر می‌کند. پولیمورفیسم تک هسته‌ای هتروزیگوت به طور متوسط هر ۱۵۰۰ باز [۱۲] رخ می‌دهد و بنابراین برای بخش‌های طولانی بسیاری از پولیمورفیسم تک هسته‌ای هتروزیگوت تحت تأثیر قرار می‌گیرند. توزیع کسر آلل S تمام پولیمورفیسم تک هسته‌ای در بخش، حالت دوگانه‌ای پیدا می‌کند که هر حالت نشان دهنده نسبت نسخه‌های آن بخش از هر والد است.

فراخوانی CNA چالش برانگیز است زیرا با مشاهده مستقل هر بخش، مسئله هنوز مشخص نشده است. حتی با فرض اینکه هر بخش فقط توسط یک CNA تحت تأثیر قرار گیرد، CNA موسوم به سه پارامتر (نسبت سلولهای حاوی CNA، تعداد کپی‌های مادر و تعداد کپی‌های پدری) وجود دارد و فقط دو مشاهده برای توضیح وجود دارد (و کسر آلل)

همه روش‌ها با فرض اینکه تعداد کمی از نژادهای زیرکلونال مسئول بیشتر یا تمام تغییرات شماره کپی هستند، این ابهام را برطرف می‌کنند. روشی که توسط الگوریتم باتبرگ [۳۱] به کار رفته است، به بیشتر تغییرات شماره کپی وابسته به یک نژاد زیر کلونال منفرد و شایع به نام تبار کلونال متکی است. تحت این روش، شیوع این تبار، همراه با تعداد کپی اصلی و جزئی در تمام تغییرات تعداد کپی کلونال، می‌تواند با یک فرآیند دو مرحله‌ای تخمین زده شود. در گام اول، این روش با فرض شیوع نژاد کلون f_c آغاز می‌شود. شیوع تبار کلونال در بیشتر موارد با خلوص نمونه تومور برابر است. با توجه به شیوع کلونال، هر بخش پس از آن فقط دو متغیر برای توضیح دارد (تعداد کپی بزرگ و جزئی). از آنجا که هر بخش دارای دو مشاهدات است، اکنون مسئله هنوز به درستی تعیین نشده است و بهترین کپی اصلی و مینور متناسب است. سپس، ترکیب کلی مقدار Φ_c فرض شده با ترکیب مناسب در تمام بخشها تعیین می‌شود. الگوریتم با بهینه سازی این تناسب بهترین مقدار Φ_c را انتخاب می‌کند. سپس برای هر بخش، شماره کپی اصلی و جزئی با بهینه سازی متناسب بودن قطعه با بهترین مقدار Φ_c انتخاب می‌شود. این روش فرض می‌کند که تمام تغییرات شماره کپی به نژاد کلونال تعلق دارند، که همیشه درست نیست. در مرحله بعدی، بخشهایی که حاوی تغییرات تعداد کپی تحت کلونال هستند با جستجوی بخشهایی با اطلاعات مناسب

ضعیف با استفاده از Φ_c استنباط شده مشخص می‌شوند. در این بخش‌ها، روش به طور همزمان و مستقل از هر بخش دیگر، عدد Φ_i و عدد کپی بزرگ و جزئی را استنباط می‌کند.

از آنجا که سه متغیر وجود دارد و تنها دو مشاهده وجود دارد، راه حل‌های بسیاری با تناسب داده برابر وجود دارد که از نظر زیست‌شناختی برای این تغییرات تعداد کپی زیر کلونال قابل قبول است. این ابهام با انتخاب راه حلی که نزدیکترین شماره به شماره نسخه طبیعی است برطرف می‌شود، اما تعدادی از موارد متداول وجود دارد که این ابتکار عمل ناموفق است. سپس این روش‌ها انتساب تغییرات تعداد کپی زیرکلونال به دودمان و تمام استنباط‌های فیلوژنتیک را برای روش‌های پایین دست رها می‌کنند.

رویکرد عمده دیگر این است که فرض کنیم همه تغییرات شماره کپی از تعداد کمی تبار ساب کلونال به وجود می‌آیند. الگوریتم‌هایی که از این روش استفاده می‌کنند به طور مشترک شیوع این نژادها و تعداد کپی بزرگ و جزئی را برای هر بخش استنتاج می‌کنند (به عنوان مثال THetA [۴۳، ۴۱] و TITAN). تعداد دودمانهای زیر کلونال معمولاً با استفاده از احتمال جریمه شده‌ای مانند معیار اطلاعات بیزی (BIC) یا انواع BIC تعیین می‌شود (به عنوان مثال THetA از BIC اصلاح شده با پارامتر مقیاس گذاری استفاده می‌کند [۴۳]). بنابراین این روش‌ها هم تغییرات شماره کپی را خوانی می‌کنند و هم آنها را به دودمان‌های زیرکلونال اختصاص می‌دهند. هیچ روش موجود این دودمان‌ها را در یک درخت فیلوژنتیک قرار نمی‌دهد

۷.۲ جهش‌های ساده بدنی

جهش‌های ساده بدنی جهش‌های کوچکی هستند که می‌توانند مستقیماً از طریق توالی‌یابی و نسبت کروموزوم‌های موجود در نمونه حاوی آنها از تعداد قرائت‌های حاوی جهش و تعداد کل خوانده‌ها در آن مکان، مشاهده شوند. نسبت قرائت حاوی جهش به کل قرائت به عنوان VAF جهش شناخته می‌شود. جهش‌های ساده بدنی‌ها معمولاً با بررسی مشترک ترازها و یک نمونه غیرسرطانی خوانده می‌شوند. این استنباط مشترک برای جداسازی انواع بدنی و ژرمینال مورد نیاز است.

این فرایند به دلیل انواع مختلف خطاها و تعصبات که در داده‌های NGS وجود دارد، دشوار می‌شود [۲۰]. یک مشکل اساسی در تشخیص جهش‌های ساده بدنی این است که به نظر می‌رسد خطاهای توالی جهش‌های ساده بدنی شیوع کمی دارند. به طور خاص، در Illumina Hiseq2000 که به طور گسترده استفاده می‌شود، از

هر ۱۰۰۰ پایه یکی از آنها دارای یک خطا است (به طور معمول یک تعویض) [۳۳]. به همین ترتیب، در طول سه میلیارد پایه ژنوم انسانی، یک احتمال غیر قابل اغماض وجود دارد که در بعضی موقعیت‌ها، چندین بار خواندن دقیقاً شامل خطای توالی دقیقاً در همان موقعیت‌ها است. به نظر می‌رسد این خطاها شیوع کم جهش‌های ساده بدنی دارند. تمایز بین این خطاها و شیوع کم واقعی جهش‌های ساده بدنی‌ها شامل یک معامله بین حساسیت و ویژگی و در حالت ایده آل، یک مدل نویز بسیار دقیق است. حل این مشکل امتداد طبیعی کار گسترده‌ای است که در زمینه فراخوانی جهش‌های جوانه‌زنی انجام شده است و الگوریتم‌های زیادی برای انجام این کار وجود دارد (به عنوان مثال [۲۰، ۱۵])

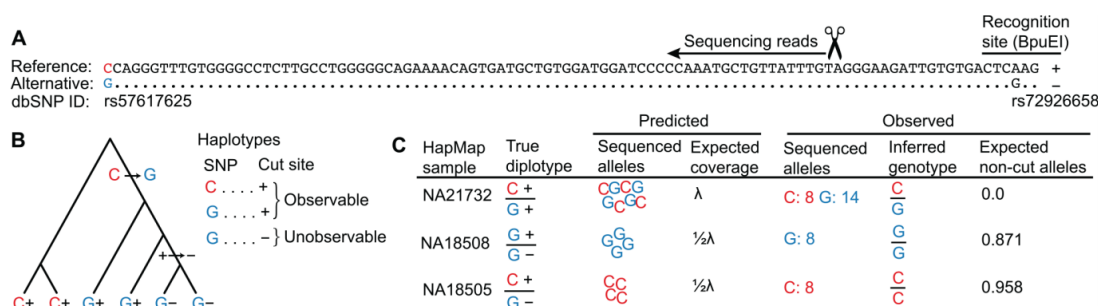
۸.۲ ترک آلی

اگرچه روش‌های تعیین توالی با بازدهی بالا [۲۹] ارزان هستند، اما تحت تاثیر مقدار بایاس هستند و مارکرهای ژنتیکی-ای تولید می‌کنند که تقریباً به طور تصادفی در کل ژنوم تقسیم می‌شوند. این روش‌ها با موفقیت در نگاشت^{۳۵} صفات [۲۳، ۳۲]، ساخت مپ پیوندی [۱۷، ۳۴]، اسکن انتخاب [۱۶، ۴۲]، و برآورد تنوع ژنتیکی [۱۴] استفاده شده است. یکی از این روش‌ها، تعیین ژنوتیپ براساس توالی [۶] (GBS) است. در GBS، هدف توالی یابی فقط با اتصال آداپتورهای توالی به محل‌های برش آنزیم محدود کننده، به کمتر از ۵٪ از ژنوم کاهش می‌یابد (شکل زیر). قرائت GBS همچنین می‌تواند به صورت کانکت‌های کوتاه مونتاژ شود، که بدون نیاز به توالی ژنوم فراخوانی یک نوع تغییر تک هسته‌ای (تغییرات تک نوکلئوتیدی) را امکان پذیر می‌کند [۲۵]. از این رو، GBS یک روش محبوب در سیستم‌های غیر مدلی است که به طور معمول فاقد منابعی مانند مجموعه ژنوم و ریزآرایه‌ها است.

بر خلاف توالی یابی کل ژنوم (WGS)، GBS مستعد ابتلا به خطاهای مختلف تماس به دلیل محدودیت چندشکلی‌های سایت است (کاهش آللیک). کاهش آللیک در GBS می‌تواند برنامه‌هایی را که به فراخوانی دقیق تغییرات نادر، از جمله تخمین طیف فرکانس سایت در ژنتیک جمعیت متکی هستند، را دچار اختلال کند. یک رویکرد آماری سیستماتیک برای تشخیص کاهش آللیک در داده‌های توالی GBS، اجرا شده و در بسته نرم افزاری منبع باز GBStools وجود دارد. این روش مبتنی بر این واقعیت است که کاهش آللیک متناسب با تعداد آللهای سایت محدود کننده بدون برش که در آنجا حمل می‌کند، میزان خوانش نمونه را در یک سایت خاص کاهش

³⁵Mapping

می‌دهد. بنابراین GBStools پوشش هر نمونه را در یک سایت خاص به عنوان یک متغیر تصادفی پواسون مورد استفاده قرار می‌دهد که از توزیع با میانگین λ (آلیک های بدون برش صفر)، توزیع با میانگین $\lambda/2$ (یک آلیک بدون برش)، یا با میانگین صفر (دو آلیک بدون برش). GBStools حداکثر احتمال پارامتر λ را با استفاده از تعداد واقعی آلیک های بدون برش در هر نمونه که به عنوان متغیرهای نهفته (مشاهده نشده) در نظر گرفته می‌شود و از طریق حداکثر رساندن مقدار چشم انتظاری (EM)، محاسبه می‌کند. از مقادیر مورد انتظار این متغیرهای نهفته می‌توان برای تخمین اینکه کدام نمونه‌ها یک آلیک بدون برش دارند استفاده کرد. به طور همزمان، GBStools فرکانس سایت آلل های SNP مرجع قابل مشاهده و جایگزین، φ_1 و φ_2 ، و آلیک بدون برش، φ_3 ، که در آن $\varphi_1 + \varphi_2 + \varphi_3 = 1$ برآورد می‌کند و در نهایت، آزمون نسبت احتمال با مقایسه فرضیه صفر $\varphi_3 = 0$ با فرضیه $\varphi_3 > 0$ جایگزین می‌کند. GBStools در اجرای فعلی خود نمی‌تواند ژنوتیپ های واقعی پنهان شده توسط کاهش آلیک را استنباط کند، اما می‌توان با فیلتر کردن سایت هایی که نسبت احتمال آنها زیاد است خطاها را حذف کند.



شکل ۸.۲: نمایی از تطابق ژنتیکی

در شکل بالا، آلل بدون برش ناشی از SNP rs72926658 با برچسب “-” و آلل برش با “+” برچسب گذاری شده است. آلل “-” در هاپلوتیپ با آلل G مشتق شده بوجود آمده و باعث شده تا برخی از آلل های G توسط GBS قابل مشاهده نباشند. نمونه های نشان داده شده دارای سه دیپلوتیپ هتروزیگوت است. نتایج توالی با پیش بینی ها مطابقت داشت و نمونه NA18505 به اشتباه هموزیگوت نامیده می‌شد، اما انتظار می‌رود تعداد آلل های کاهش یافته محاسبه شده توسط GBStools (0.958) با تعداد واقعی (۱) مطابقت داشته باشد، و آن را به عنوان یک تماس اشتباه احتمالی مشخص کند.

۹.۲ مقدمه‌ای بر مدل سازی احتمالی

وظیفه اصلی یادگیری ماشین، یادگیری از داده‌ها است، کاری که به عنوان استنباط شناخته می‌شود. برای یادگیری از داده‌ها، باید فرضیاتی را مطرح کرد. توصیف رسمی فرضیات صورت گرفته به عنوان یک مدل ذکر می‌شود. یک مدل احتمالی مفروضات ارائه شده را تعریف می‌کند که اطلاعات آموخته شده را با استفاده از متغیرهای تصادفی و توزیع‌های احتمال به داده‌های مشاهده شده پیوند می‌دهد. توزیع‌های احتمال توابع ریاضی هستند که یک رویداد را ورودی می‌کنند و احتمال آن واقعه را بیرون می‌آورند. توزیع احتمال می‌تواند تابعی بیش از واقعه باشد و این متغیرهای اضافی به عنوان پارامترهای توزیع شناخته می‌شوند [۲۵]. رویکرد بیزی در یادگیری ماشین شامل استنباط احتمالی مقادیر پارامترهای منوط به مشاهدات است [۲۶]. چهار مولفه دارد:

- احتمال: احتمال مشاهده داده‌ها است، مشروط به تنظیم پارامتر $P(\text{data} | \text{parameters})$

- پارامترهای احتمال

- پارامترهای قبلی

- داده‌های مشاهده شده

پارامترها خود مجموعه‌ای از متغیرهای تصادفی هستند که از توزیع قبلی $P(\text{پارامترها})$ گرفته شده‌اند، که باورهای ما را در مورد احتمال حالت‌های مختلف پارامتر در غیاب مشاهده مشاهده می‌کند. این اصطلاحات با استفاده از قانون بیز با هم ترکیب می‌شوند:

- $$P(\text{parameters}|\text{data}) = P(\text{data}|\text{parameters}) * P(\text{parameters}) / P(\text{data})$$

- $$\text{Posterior} \propto \text{likelihood} * \text{prior}$$

پس زمینه توزیع پارامترهای مشروط به مشاهده داده‌ها است و خروجی اصلی استنتاج بیزی است. از توزیع پسین می‌توان برای انجام کارهایی مانند پیش‌بینی مشاهدات آینده استفاده کرد.

۱.۹.۲ زنجیره مارکوف مونت کارلو

برای انجام استنتاج بیزی^{۳۶}، ما اغلب می‌خواهیم در توزیع پسین ادغام شده، پیش‌بینی کنیم یا خلاصه‌هایی پیدا کنیم، به عنوان مثال میانگین پارامتر پسین. به طور کلی، انجام چنین ادغامی (جمع بندی در مورد متغیرهای گسسته) از نظر تحلیلی غیرقابل حل است. با این حال، می‌توان چنین ادغام‌هایی را با استفاده از نمونه‌هایی که از قسمت پسین ترسیم شده‌اند تقریبی داد:

$$E[f] = \int f(x)p(x)dx \approx 1/N \sum_{1..N} f(x_i) \quad (1.2)$$

که در آن x_i نمونه i از $p(x)$ و $p(x)$ و $f(x)$ به ترتیب توزیع و عملکرد مورد نظر ما است. به ندرت می‌توان مستقیماً از توزیع پسین نمونه برداری کرد. برای تولید موثر نمونه‌ها از توزیع، حتی در ابعاد بالا، می‌توان از تکنیک زنجیره مارکوف مونت کارلو استفاده کرد. زنجیره مارکوف مونت کارلو یک زنجیره مارکوف می‌سازد که در آن توزیع تعادل توزیع پسین است. سپس مقادیر زنجیره می‌تواند به عنوان نمونه از پسین با توجه به همگرایی کافی به توزیع تعادل مورد استفاده قرار گیرد. برای انجام زنجیره مارکوف مونت کارلو، تاز زمانی که بتوان $p \propto p(x)$ را محاسبه کرد، نیازی به محاسبه $p(x)$ نیست. این زنجیره مارکوف مونت کارلو را قادر می‌سازد تا از محاسبه ثابت‌های نرمال سازی، که اغلب غیرقابل حل هستند، خودداری کند. یک زنجیره مارکوف به عنوان یک سری متغیرهای تصادفی تعریف می‌شود که دارای ویژگی استقلال شرطی زیر هستند:

$$p(z^{N+1} | z^1..z^N) = p(z^{N+1} | z^N) \quad (2.2)$$

نمونه‌ای از الگوریتم زنجیره مارکوف مونت کارلو الگوریتم Metropolis-Hastings (MH) است [۲۸]. الگوریتم MH از حالت دلخواه Z^t شروع می‌شود. سپس یک حالت پیشنهادی z از توزیع پروپوزال $q(z|z^t)$ ترسیم می‌شود. این حالت پیشنهادی z با احتمال زیر پذیرفته می‌شود:

$$\min(1, \hat{p}(z^*) q(z^t | z^*) / \hat{p}(z^t) q(z^* | z^t)) \quad (3.2)$$

³⁶Bayesian

می‌توان نشان داد که الگوریتم MH تعادل دقیق را برآورده می‌کند و از این رو، $p(x)$ توزیع تعادل است [۱۰]. در حالی که توازن دقیق برای اثبات اینکه در محدوده نمونه‌های بی‌نهایت زنجیره به توزیع مورد نظر همگراست کافی است، اما در عمل فقط تعداد محدودی از نمونه‌ها را می‌توان ترسیم کرد. واضح است که نمونه‌های ابتدای زنجیره، که از یک مکان دلخواه در فضای حالت شروع می‌شوند، بعید است از توزیع تعادل باشد. این نمونه‌ها به عنوان نمونه‌های سوختنی کنار گذاشته می‌شوند. هرچه همگرایی زنجیره مارکوف سریعتر باشد، نمونه‌های کمتری باید کنار گذاشته شوند و می‌توان از تعداد بیشتری برای محاسبه انتظارات استفاده کرد. با بررسی اثری از مقادیر مهم پارامتر یا احتمال همگرایی می‌توان نظارت کرد، اما این امر ممکن است چند حالت را از دست بدهد. متأسفانه دانستن اینکه آیا همگرایی حاصل شده است غیرممکن است، فقط گاهی اوقات می‌توان همگرایی را رد کرد [۲۱]. گذشته از همگرایی، یکی دیگر از خصوصیات اصلی یک زنجیره مارکوف میزان اختلاط زنجیره است. با توجه به n نمونه مستقل از توزیع، واریانس میانگین پارامتر برآورد σ_n است که σ انحراف استاندارد توزیع خلفی پارامتر است. نمونه‌های گرفته شده از زنجیره مارکوف مستقل نیستند، زیرا به وضعیت فعلی زنجیره بستگی دارند (یعنی فقط از نظر شرطی مستقل هستند). برای تخمین اندازه نمونه موثر یک زنجیره مارکوف، یعنی تعداد نمونه‌های مستقل با همان خطای استاندارد همان زنجیره، می‌توان از معادله زیر استفاده کرد:

$$ESS = \frac{n}{1 + 2 \sum_{j=1}^{\infty} \rho_j} \quad (۴.۲)$$

حاصل جمع بی‌نهایت محاسبه ESS را می‌توان با استفاده از برآوردگر پریدوگرام کوتاه تطبیقی Sokal [۳۵] تخمین زد.

فصل ۳

روش‌های پیشین

فصل ۴

روش پیشنهادی

۱.۴ مقدمه

پس از آشنایی با روش‌های پیشین که برای حل مسئله مشابه مورد استفاده قرار گرفته‌اند، حال می‌توانیم به معرفی و تشریح روش‌های پیشنهادی خود برای حل مسئله پیش رو پردازیم. در این فصل ابتدا داده‌های ورودی مسئله را همراه با فرضیات در نظر گرفته شده بیان می‌کنیم و پس از آن دو روش پیشنهادی متفاوت را بیان خواهیم نمود. در روش اول که به رویکردهای پیشین نزدیک‌تر است با تغییری از جنس روش‌های نوین در مراحل میانی به یک روش جدید می‌رسیم که به علت افزایش سرعت همگرایی می‌توان فرض و داده‌های جدیدی را از طریق CNV به آن افزود و پاسخ گرفت. اما روش دوم کاملاً متفاوت بوده و با رویکردی جدید در حوزه یادگیری ماشین همراه است که به کمک یادگیری تقویتی به حل مسئله مورد نظر می‌پردازد.

۲.۴ معرفی دادگان ورودی

قبل از وارد شدن به بخش روش‌های پیشنهادی نیاز است تا دادگان ورودی را مشخص و معرفی نماییم تا در قسمت‌های بعدی بتوانیم از نمادهای معرفی شده در این بخش استفاده نماییم. دادگان مورد استفاده

۳.۴ روش پیشنهادی اول (درخت‌بازی)

۱.۳.۴ پیش‌پردازش

قبل از شروع باید بر روی داده‌ها یک پیش‌پردازش اعمال کنیم که وابسته به سیاست در نظر گرفته شده می‌تواند باعث تغییر در پاسخ نهایی نیز شود. به این منظور داده‌هایی که miss شده‌اند با روش‌های زیر می‌تواند برای ورود به مرحله بعد تخمین زده شود.

۱.۱.۳.۴ تصادفی

پر کردن کاملاً تصادفی میس‌ها

جدول ۱.۴: اندیس‌های به کار رفته در مدل ریاضی

I, J	بیماران
k	مرحله زمان‌بندی (بستری، اتاق عمل، ریکاوری)
L_k	ماشین (تخت یا اتاق عمل) در مرحله k
n	جراح

جدول ۲.۴: پارامترهای مدل ریاضی

t_{ik}	زمان خدمت‌دهی به بیمار در مرحله k ام
\tilde{t}_{ik}	زمان فاری خدمت‌دهی به بیمار در مرحله k ام
t_{ik}^p	مقدار بدبینانه (حداکثر) برای زمان خدمت‌دهی به بیمار در مرحله k ام
t_{ik}^m	محتمل‌ترین مقدار برای زمان خدمت‌دهی به بیمار در مرحله k ام
t_{ik}^o	مقدار خوشبینانه (حداقل) برای زمان خدمت‌دهی به بیمار در مرحله k ام

جدول ۳.۴: متغیرهای مدل ریاضی

X_{ild_k}	متغیر صفر-یک تخصیص بیمار به تخت/اتاق عمل
S_{ild_k}	زمان شروع خدمت‌دهی به بیمار
Y_{ijkl_k}	متغیر صفر-یک توالی بیماران
V_{ni}	متغیر صفر-یک تخصیص جراح به بیمار

فصل ۵

نتایج تجربی

فصل ۶

بحث و نتیجه گیری

مراجع

- [1] Nci dictionary of cancer terms: somatic mutation definition. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/somatic-mutation?redirect=true>. 10
- [2] Ii neoplasms. 19 June 2014. 3
- [3] Cancer - activity 1 - glossary. page page 4 of 5, 2008. 3
- [4] Abrams, Gerald. Neoplasia i. 23 January 2012. 3
- [5] Alberts, Bruce, Johnson, Alexander, Lewis, Julian, Raff, Martin, Roberts, Keith, and Walter, Peter. Molecular biology of the cell 4th edition. New York: Garland Science, 1463, 2002. 8, 9
- [6] Anderson, Kristina, Lutz, Christoph, Van Delft, Frederik W, Bateman, Caroline M, Guo, Yan-ping, Colman, Susan M, Kempinski, Helena, Moorman, Anthony V, Titley, Ian, Swansbury, John, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, 469(7330):356–361, 2011. 21
- [7] Andor, Noemi, Graham, Trevor A, Jansen, Marnix, Xia, Li C, Aktipis, C Athena, Petritsch, Claudia, Ji, Hanlee P, and Maley, Carlo C. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature medicine*, 22(1):105–113, 2016. 15
- [8] Behjati, Sam, Huch, Meritxell, van Boxtel, Ruben, Karthaus, Wouter, Wedge, David C, Tamuri, Asif U, Martincorena, Iñigo, Petljak, Mia, Alexandrov, Ludmil B, Gundem, Gunes, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature*, 513(7518):422–425, 2014. 13, 16
- [9] Birbrair, Alexander, Zhang, Tan, Wang, Zhong-Min, Messi, Maria Laura, Olson, John D, Mintz, Akiva, and Delbono, Osvaldo. Type-2 pericytes participate in normal and tumoral angiogenesis. *American Journal of Physiology-Cell Physiology*, 307(1):C25–C38, 2014. 3, 4, 5, 11

- [10] Bishop, Christopher M. Pattern recognition. *Machine learning*, 128(9), 2006. [25](#)
- [11] Burrell, Rebecca A and Swanton, Charles. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular oncology*, 8(6):1095–1111, 2014. [5](#)
- [12] Chen, Rui, Mias, George I, Li-Pook-Than, Jennifer, Jiang, Lihua, Lam, Hugo YK, Chen, Rong, Miriami, Elana, Karczewski, Konrad J, Hariharan, Manoj, Dewey, Frederick E, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 148(6):1293–1307, 2012. [19](#)
- [13] Cooper, Geoffrey M. *Elements of human cancer*. Jones & Bartlett Learning, 1992. [4](#)
- [14] de Visser, J Arjan GM and Rozen, Daniel E. Clonal interference and the periodic selection of new beneficial mutations in escherichia coli. *Genetics*, 172(4):2093–2100, 2006. [21](#)
- [15] Demicheli, R, Retsky, MW, Hrushesky, WJM, Baum, M, and Gukas, ID. The effects of surgery on tumor growth: a century of investigations. *Annals of oncology*, 19(11):1821–1828, 2008. [21](#)
- [16] Dentro, Stefan C, Leshchiner, Ignaty, Haase, Kerstin, Tarabichi, Maxime, Wintersinger, Jeff, Deshwar, Amit G, Yu, Kaixian, Rubanova, Yulia, Macintyre, Geoff, Vázquez-García, Ignacio, et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *BioRxiv*, page 312041, 2018. [21](#)
- [17] Fearon, Eric R and Vogelstein, Bert. A genetic model for colorectal tumorigenesis. *cell*, 61(5):759–767, 1990. [21](#)
- [18] Fedele, Clare, Tothill, Richard W, and McArthur, Grant A. Navigating the challenge of tumor heterogeneity in cancer therapy. *Cancer discovery*, 4(2):146–148, 2014. [5](#)
- [19] Fisher, Rosie, Pusztai, Lazos, and Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *British journal of cancer*, 108(3):479–485, 2013. [2](#)
- [20] Friedl, Peter and Wolf, Katarina. Plasticity of cell migration: a multiscale tuning model. *Journal of Cell Biology*, 188(1):11–19, 2010. [20](#), [21](#)
- [21] Gelman, Andrew, Shirley, Kenneth, et al. Inference from simulations and monitoring convergence. *Handbook of markov chain monte carlo*, 6:163–174, 2011. [25](#)
- [22] Gerlinger, Marco, Rowan, Andrew J, Horswell, Stuart, Larkin, James, Endesfelder, David, Gronroos, Eva, Martinez, Pierre, Matthews, Nicholas, Stewart, Aengus, Tarpey, Patrick, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl j Med*, 366:883–892, 2012. [14](#)

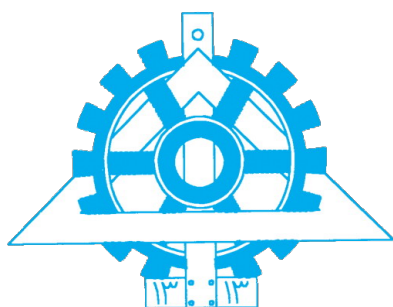
- [23] Greaves, Mel and Maley, Carlo C. Clonal evolution in cancer. *Nature*, 481(7381):306–313, 2012. [21](#)
- [24] Halford, S, Rowan, A, Sawyer, E, Talbot, I, and Tomlinson, Ian. O6-methylguanine methyltransferase in colorectal cancers: detection of mutations, loss of expression, and weak association with g: C> a: T transitions. *Gut*, 54(6):797–802, 2005. [4](#)
- [25] Hanahan, Douglas and Weinberg, Robert A. The hallmarks of cancer. *cell*, 100(1):57–70, 2000. [21](#), [23](#)
- [26] Hanahan, Douglas and Weinberg, Robert A. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011. [12](#), [23](#)
- [27] Handa, Osamu, Naito, Yuji, and Yoshikawa, Toshikazu. Redox biology and gastric carcinogenesis: the role of helicobacter pylori. *Redox Report*, 16(1):1–7, 2011. [3](#)
- [28] Hastings, W Keith. Monte carlo sampling methods using markov chains and their applications. 1970. [24](#)
- [29] Hugo, Honor, Ackland, M Leigh, Blick, Tony, Lawrence, Mitchell G, Clements, Judith A, Williams, Elizabeth D, and Thompson, Erik W. Epithelial—mesenchymal and mesenchymal—epithelial transitions in carcinoma progression. *Journal of cellular physiology*, 213(2):374–383, 2007. [21](#)
- [30] Lee, Kyung-Hwa, Lee, Ji-Shin, Nam, Jong-Hee, Choi, Chan, Lee, Min-Cheol, Park, Chang-Soo, Juhng, Sang-Woo, and Lee, Jae-Hyuk. Promoter methylation status of hmlh1, hmsh2, and mgmt genes in colorectal cancer associated with adenoma–carcinoma sequence. *Langenbeck’s archives of surgery*, 396(7):1017–1026, 2011. [4](#)
- [31] Nik-Zainal, Serena, Van Loo, Peter, Wedge, David C, Alexandrov, Ludmil B, Greenman, Christopher D, Lau, King Wai, Raine, Keiran, Jones, David, Marshall, John, Ramakrishna, Manasa, et al. The life history of 21 breast cancers. *Cell*, 149(5):994–1007, 2012. [15](#), [19](#)
- [32] Nowell, Peter C. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976. [10](#), [13](#), [21](#)
- [33] Sabeh, Farideh, Shimizu-Hirota, Ryoko, and Weiss, Stephen J. Protease-dependent versus-independent cancer cell invasion programs: three-dimensional amoeboid movement revisited. *Journal of Cell Biology*, 185(1):11–19, 2009. [20](#)
- [34] Sakr, WA, Haas, GP, Cassin, BF, Pontes, JE, and Crissman, JD. The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *The Journal of urology*, 150(2):379–385, 1993. [21](#)

- [35] Sokal, Alan. Monte carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration*, pages 131–192. Springer, 1997. [25](#)
- [36] Stewart, BWKP and Wild, CP. World cancer report 2014. health, 2017. [13](#)
- [37] Stratton, Michael R, Campbell, Peter J, and Futreal, P Andrew. The cancer genome. *Nature*, 458(7239):719–724, 2009. [13](#), [14](#)
- [38] Sun, Xiao-xiao and Yu, Qiang. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacologica Sinica*, 36(10):1219–1227, 2015. [2](#)
- [39] Talbot, Simon J and Crawford, Dorothy H. Viruses and tumours—an update. *European Journal of Cancer*, 40(13):1998–2005, 2004. [13](#)
- [40] Truninger, Kaspar, Menigatti, Mirco, Luz, Judith, Russell, Anna, Haider, Ritva, Gebbers, Jan-Olaf, Bannwart, Fridolin, Yurtsever, Hueseyin, Neuweiler, Joerg, Riehle, Hans-Martin, et al. Immunohistochemical analysis reveals high frequency of pms2 defects in colorectal cancer. *Gastroenterology*, 128(5):1160–1171, 2005. [5](#)
- [41] Vander Heiden, Matthew G, Cantley, Lewis C, and Thompson, Craig B. Understanding the warburg effect: the metabolic requirements of cell proliferation. *science*, 324(5930):1029–1033, 2009. [20](#)
- [42] Waclaw, Bartlomiej, Bozic, Ivana, Pittman, Meredith E, Hruban, Ralph H, Vogelstein, Bert, and Nowak, Martin A. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264, 2015. [21](#)
- [43] Zhu, Aizhi, Lee, Daniel, and Shim, Hyunsuk. Metabolic positron emission tomography imaging in cancer detection and therapy response. In *Seminars in oncology*, volume 38, pages 55–69. Elsevier, 2011. [20](#)

Abstract

This thesis studies on writing projects, theses and dissertations using tehran-thesis class. It ...

Keywords SNV, CNV, Phylogenetic, Tree, Q-learning, Deep learning



University of Tehran
College of Engineering
Faculty of New Science and
Technology
Network



Inference of Phylogenetic Tree for Inter Tumor using Single Cell Mutations and CNV

A Thesis submitted to the Graduate Studies Office
In partial fulfillment of the requirements for
The degree of Master of Science
in Information Technology - Network Science

By:

Afshin Bozorgpour

Supervisors:

Dr. Saman Haratizadeh and Dr. Abolfazl Motahari

Jul 2021