مدل گرافیکی احتمالی برای کشف جامع امضاهای جهش‌

مینا شایگان, سید علی کتان‌فروش

4تیر, 1399

چکیده

شناخت فرایند‌های جهش‌زایی که در طی گسترش سرطان عمل می‌کنند، موضوع اصلی بیولوژی سرطان است. جهش‌های جسمی ناشی از این فرایند‌ها با تکثیر دی‌ان‌ای یا قرار گرفتن در معرض عوامل محیطی و روش زندگی مرتبط است. دانستن فعالیت فرایند‌های جهش‌‌زا در یک تومور می‌تواند در روش‌های درمانی شخصی‌سازی شده، تشخیص زودهنگام، و تحقیقات تومورشناسی مورد استفاده قرارگیرد. روش‌های محاسباتی 30 امضای معتبر از فرایند‌های جهش‌زا فعال در سرطان‌های انسانی شناسایی کرده‌اند. هر امضا یک الگو از انواع جهش است.‌ هرچند که مشخص است که امضاهای مختلف از بافت‌های مختلفی بدست آمده‌اند که ممکن است ضایعه اولیه یکسان داشته باشند. برای تعیین یک امضای جهش‌ متناسب با یک فرایند جهش‌زا لازم است امضاهای بدست آمده از همه ضایعات اولیه‌ای که فرایند جهش‌زا در آن عمل کرده است را با یکدیگر ادغام کنیم. نمونه‌های هر ضایعه اولیه نباید جداگانه بررسی کرد بلکه باید همه نمونه‌ها با هم در نظر بگیریم. مدل‌های قبلی انواع ضایعات اولیه را جداگانه بررسی کردند، که برای تحلیل مقایسه‌ای نمونه‌ها ناکارا است. ما زمینه‌ای را فراهم می‌کنیم که می‌توان مجموعه نمونه‌ها را به زیرگروه‌هایی از نوع ضایعه اولیه تقسیم کرد. امضاهای جهش بین زیرگروه‌ها مشترک هستند، اما نسبت امضاها برای هر نوع ضایعه اولیه متناسب با آن نوع ضایعه اولیه است. ما مدل تخصیص دیریکله پنهان ترکیبی همراه با استنباط بیزی وردشی جهت استخراج امضاهای جهش‌ استفاده می‌شود تا بتوانیم ویژگی‌های هر ضایعه اولیه بدست آوریم. از کران‌های پایین وردشی جهت پیدا کردن تعداد الگوهای جهش ممکن استفاده می‌شود.

کلمات کلیدی: فرایند جهش‌زا، امضا جهش، سرطان، مدل گرافیکی احتمال

1. مقدمه

فرایند‌های جهش‌زا متعددی از جمله خطاهای تکثیر دی‌ان‌ای و قرار گرفتن در معرض عوامل جهش‌زایی مانند مواد شیمیایی، رادیواکتیو، دخانیات سرطان زا [1]، نور فرابنفش [2]، و واکنشهای التهابی باعث ایجاد جهش‌ها هستند. جهش‌های محرک1 که در فرایند تبدیل شدن سلول سالم به سرطانی دخیل هستند، قابلیت رشد را به سلول‌های سرطانی می‌دهند و در ریزمحیط بافتی که سرطان ایجاد می‌شود به طور مثبت انتخاب می‌شوند (نسبت جهش‌هایی که یک آمینواسید را تغییر می‌دهند به نسبت جهش‌هایی که آمینواسیدی را تغییر نمی‌دهند, بیش‌تر از انتظار است). برای حفظ نهایی سرطان نیازی به یک جهش محرک نیست (اگرچه همچنان حضور دارند) اما باید در بعضی از مراحل پیشرفت سرطان انتخاب شوند. در مقابل جهش‌های گذرگر2 به طور مثبت انتخاب نمی‌شوند و قابلیت رشد تجمعی را ایجاد نمی‌کند و در نتیجه در پیشرفت سرطان دخالت ندارد. جهش‌های گذرگر در ژنوم سرطانی یافت می‌شود زیرا جهش‌های جسمی بدون عواقب عملکردی اغلب در طول تقسیم سلولی رخ می‌دهند. بنابراین یک سلولی که یک جهش محرک بدست می‌آورد، از قبل جهش‌های جسمی بی‌اثر زیستی در ژنوم خود دارد. این جهش‌ها به دنبال توسعه تجمعی همراه می‌شوند و در نتیجه در تمام سلول‌های سرطان نهایی حضور دارند [3]. مطالعات ژنوم سرطانی معمولا بر روی شناسایی جهش‌‌های محرک تمرکز دارند. این جهش‌ها در درک مکانیسم توسعه سرطان کمک می‌کنند. با این حال جهش‌های گذرگر نیز اطلاعات مهمی در اختیار ما قرار می‌دهند، زیرا که اغلب آنها الگوهایی (امضاهای جهش) را نشان می‌دهند که باعث ایجاد جهش‌های جسمی می‌شوند. مطالعات کلاسیک با توجه به محدودیت توالی‌یابی برای درک فرایند‌های جهش جسمی بر روی ژن‌های سرطانی محدودی تمرکز می‌کردند که فراوانی جهش در آنها زیاد است. این جهش‌ها در افراد مختلف با یک نوع سرطان جهت به دست آوردن جهش‌های کافی برای تجزیه و تحلیل جمع‌آوری می‌شوند. سپس پروفایل‌های الگو جهش با انواع مختلف سرطان با یکدیگر مقایسته می‌شوند. با این حال، از آنجایی که

1 driver mutation

2 passenger mutation

بسیاری از جهش‌ها در ژن‌های سرطانی جهش‌های محرک ناشی از تکثیر سلولی هستند، پروفایل جهش حاصل یک نمایش سودار1 از فرایند جهش متناظر است. علاوه‌ بر این کمبود داده‌های جهش امکان مقایسته الگوهای جهش افراد وجود ندارد. به دلیل توسعه توالی‌یابی نسل جدید2، داده‌هایی با مقیاس بزرگ از ژنوم سرطانی به سرعت در دسترس همگان قرار گرفته است و فرصت‌های جدیدی را برای بررسی امضاهای جهش به صورت فردی با یک نمایش ناسودار با استفاده از داده‌های جهش جسمی گسترده ژنوم فراهم می‌کند[4].

هر فرآیند جهش‌زا منجر به امضای جهش‌ خاصی می‌شود و انباشت جهش‌ها از بدو تولد تا به امروز نتیجه ترکیبی از برخی از امضاهای جهش‌ در نظر گرفته می‌شود [5]. از این رو ، انتظار می‌رود که تفکیک‌سازی امضاهای جهش مکانیسم‌های سرطان‌زا را آشکار کند و ممکن است به عنوان نشانگرهای زیستی برای تشخیص زودهنگام عمل کنند [6،7،8]. اخیرا با استفاده از آزمایش‌های مبتنی بر ویرایش ژن CRISPR-Cas9 در یک سیستم سلول مهندسی‌شده ژنتیکی انسان، مجدداً امضاهای جهش سرطانی در شرایط آزمایشگاهی بازتولید شدند. این نتیجه از وجود امضاهای جهش‌ پشتیبانی می‌کند [9].

امضاها با شناسایی الگوهای مشترک در بسیاری از تومورها براساس تعداد جهش‌ها و نوع آنها کشف می‌شوند. روش اصلی کشف امضا براساس فاکتورگیری نامنفی ماتریس3 بود [10] و 30 امضاء جهش‌ در انواع مختلف سرطان آشکار کرد [14،11،12،13]4. رویکرد "signeR" نیز از روش فاکتورگیری نامنفی ماتریس در ترکیب با رویکرد بیزی تجربی برای یافتن تعداد صحیح امضا‌ها استفاده می‌کند [15]. روش دیگر از مدل‌سازی موضوعی استفاده می‌کند که براساس مدل‌ تخصیص پنهان دیریکله5 [16] انجام می‌شود[4، 17].

در مقایسته با روش فاکتورگیری نامنفی ماتریس، مدل‌سازی موضوعی ساختار‌های احتمالی از جهش‌ها در

1 biased

2 next-generation sequencers

3 Non-Negative Matrix Factorization

4 <http://cancer.sanger.ac.uk/cosmic>

5 latent dirichlet allocation

نمونه‌ها فرض می‌کند. انتظار می رود این رویکرد عملکرد کلی سازی را بهبود بخشد [18].

اگرچه این روش‌ها در برآورد امضاهای جهش‌ مؤثر هستند ، اما آنها دید جامع و کاملی از امضاهای جهش‌ ندارند. برای تعیین یک امضا جهش متناظر با یک فرایند جهش‌زا لازم است یک امضا به عنوان نماینده انتخاب شود یا امضاهای بدست آمده از همه ضایعات اولیه بوجود آمده از آن فرایند جهش‌زا را با یکدیگر ادغام کنیم. اگرچه توزیع امضاهای بدست آمده ممکن است کمی متفاوت باشند، نمونه‌های هر ضایعه اولیه نباید از دیگر ضایعات اولیه جدا باشند، بلکه نمونه‌ها را به صورت جامع تجزیه و تحلیل می‌کنیم تا بتوانیم این مشکل را حل کنیم. با این وجود ، در این حالت ، پیچیدگی مجموعه داده افزایش می‌یابد و شناسایی امضاها دشوار می‌شود. علاوه بر این ، هنگام جستجوی امضای جدید ، انتخاب مدل مبنای مهمی برای تصمیم‌گیری در مورد جدید بودن امضا است. یک روش قابل توجه برای تعیین تعداد امضاهای جهش ، استفاده از روش ""Emu [20] است که در آن معیار اطلاع بیزی-شوارتز1 برای انتخاب مدل استفاده می‌شود. با این حال، معیار اطلاع بیزی-شوارتز فقط برای مدل‌های احتمالی که ماتریس اطلاعات فیشر آنها به طور منظم باشد از نظر ریاضیاتی مناسب است زیرا توزیع پسین پارامترها می‌تواند تقریب شده لاپلاسی باشد که در مدل منظم آماری قضیه حد مرکزی در آن برقرار باشد. بنابراین ، معیار اطلاع بیزی-شوارتز نباید در مدل‌های ترکیبی شامل متغیرهای پنهان مانند تخصیص پنهان دیریکله استفاده شود. توجه داشته باشید که "signeR" که در پاراگراف قبلی توضیح داده شده است، از معیار اطلاع بیزی-شوارتز نیز برای انتخاب مدل استفاده کرده است. برای پرداختن به این موضوعات، ما یک روش جدید برای پیش‌بینی امضاهای جهش پیشنهاد می‌کنیم و یک مدل ترکیبی احتمالی جدید برای تخصیص پنهان دیریکله با استنباط بیزی وردشی2 انتخاب می‌کنیم. پارامترهای توزیع پیشین در تخصیص پنهان دیریکله ویژگی‌های هر ضایعه اولیه را می‌گیرد. با استفاده از چنین مدلی ، با به اشتراک‌گذاری توزیع‌های امضاهای جهش بین ضایعات اولیه، می‌توان نمونه‌ها را به صورت جمعی تجزیه و تحلیل کرد. این مدل مجموعه ای از امضاها را تولید کند که

1 bayesian information criterion

2 variational bayes inference

شامل چندین امضا از همان فرآیند جهش زا نیست و به هر فرایند جهش‌زا یک امضا جهش اختصاص می‌دهد. از استنباط بیزی وردشی برای انتخاب مدل و تعداد امضاهای جهش استفاده می‌کنیم. ما تأیید کردیم که روش ما در یک شبیه‌سازی عددی به اندازه کافی دقیق است. @@@

1. مواد و روش ها
   1. جهش‌ها در ژنوم‌های سرطان

با توجه به پیشرفت‌های اخیر توالی‌یابی نسل جدید، تعداد ژنوم‌های شناخته شده سرطان به سرعت در حال رشد است. با یک ژنوم خاص سرطان، مجموعه‌ای از جهش‌ها موجود در ژنوم را بدست می‌آوریم. جهش‌های جسمی معمولاً به چهار کلاس گروه‌بندی می‌شوند: جایگزینی1، جایگیری و حذف2، جابجایی3 و مضاعف شدگی4. از 1865 ژنوم کامل شامل بیشتر انواع سرطان، 79,793,266 جایگزینی تکی، و 4,122,233 جایگیری و حذف به دست آوردند که برای امضاهای جهش مورد بررسی قرار می‌گیرد [21].

یک جایگزینی تکی از A به C به صورت [A > C] نمایش داده می‌شود. هر نوکلئوتید می‌تواند با 3 نوع نوکلئوتید دیگر جایگزین شود. اگر یک نوکلئوتید در یک رشته دی‌ان‌ای جایگزین شود متناسب با نوع جایگزینی در رشته مکمل نیز جایگزینی رخ می‌دهد. مثلا اگر جایگزینی [C > T] رخ دهد، جایگزینی [G > A] نیز در رشته مکمل آن رخ می‌دهد [13]. در نتیجه 6 نوع جایگزینی تکی داریم:

1 substitution

2 indel

3 rearrangement

4 copy number change

* 1. تخصیص دیریکله پنهان برای مدل‌سازی امضاهای جهش

مراجع

1. Pfeifer, GP. et al. (2002) DNA damage and p53 mutations in smoking-associated cancers. Oncogene, 21 (48):7435–7451.
2. Pfeifer, GP. et al. (2005) Mutations induced by ultraviolet light. Mutat Res, 571(1- 2):19–31.
3. Stratton, M.R. et al. (2009) The cancer genome. Nature, 458, 719–724.
4. Shiraishi, Y. et al. (2015) A simple model-based approach to inferring and visualizing cancer mutation signatures. PLoS Genet., 11, e1005657.
5. Stratton, M.R. (2011) Exploring the genomes of cancer cells: progress and promise. Science, 331, 1553–1558.
6. Harris, R.S. (2013) Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. Genome Med., 5, 87.
7. Temko, D. et al. (2018) The effects of mutational processes and selection on driver mutations across cancer types. Nat. Commun., 9, 1857.
8. Wagener, R. et al. (2015) Analysis of mutational signatures in exomes from B-cell lymphoma cell lines suggest APOBEC3 family members to be involved in the pathogenesis of primary effusion lymphoma. Leukemia, 29, 1612–1615.
9. Zou, X. et al. (2018) Validating the concept of mutational signatures with isogenic cell models. Nat. Commun., 9, 1744.
10. Lee, D.D. and Seung, H.S. (2001) Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, pp. 556–562.
11. Nik-Zainal, S. et al. (2012) Mutational processes molding the genomes of 21 breast cancers. Cell, 149, 979–993.
12. Nik-Zainal, S. et al. (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature, 534, 47
13. Alexandrov, L.B. et al. (2013) Deciphering signatures of mutational processes operative in human cancer. Cell Rep., 3, 246–259.
14. Alexandrov, L.B. et al. (2015) Clock-like mutational processes in human somatic cells. Nat. Genet., 47, 1402.
15. Rosales, R.A. et al. (2017) Signer: an empirical bayesian approach to mutational signature discovery. Bioinformatics, 33, 8–16.
16. Blei, D.M. et al. (2003) Latent dirichlet allocation. J. Mach. Learn. Res., 3, 993–1022.
17. Taro Matsutani. et al. (2019) Discovering novel mutation signatures by latent Dirichlet allocation with variational Bayes inference, Bioinformatics, Volume 35, Issue 22, Pages 4543–4552.
18. Hofmann, T. (1999) Probabilistic latent semantic indexing. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 50–57.
19. Alexandrov, L.B. et al. (2013) Signatures of mutational processes in human cancer. Nature, 500, 415–421.
20. Fischer,A. et al. (2013) Emu: probabilistic inference of mutational processes and their localization in the cancer genome. Genome Biol., 14, R39.
21. Alexandrov, L.B. et al. (2020) The repertoire of mutational signatures in human cancer. Nature, 578, 94–101.