Article: Deciphering Signatures of Mutational Processes Operative in Human Cancer

Journal: Cell

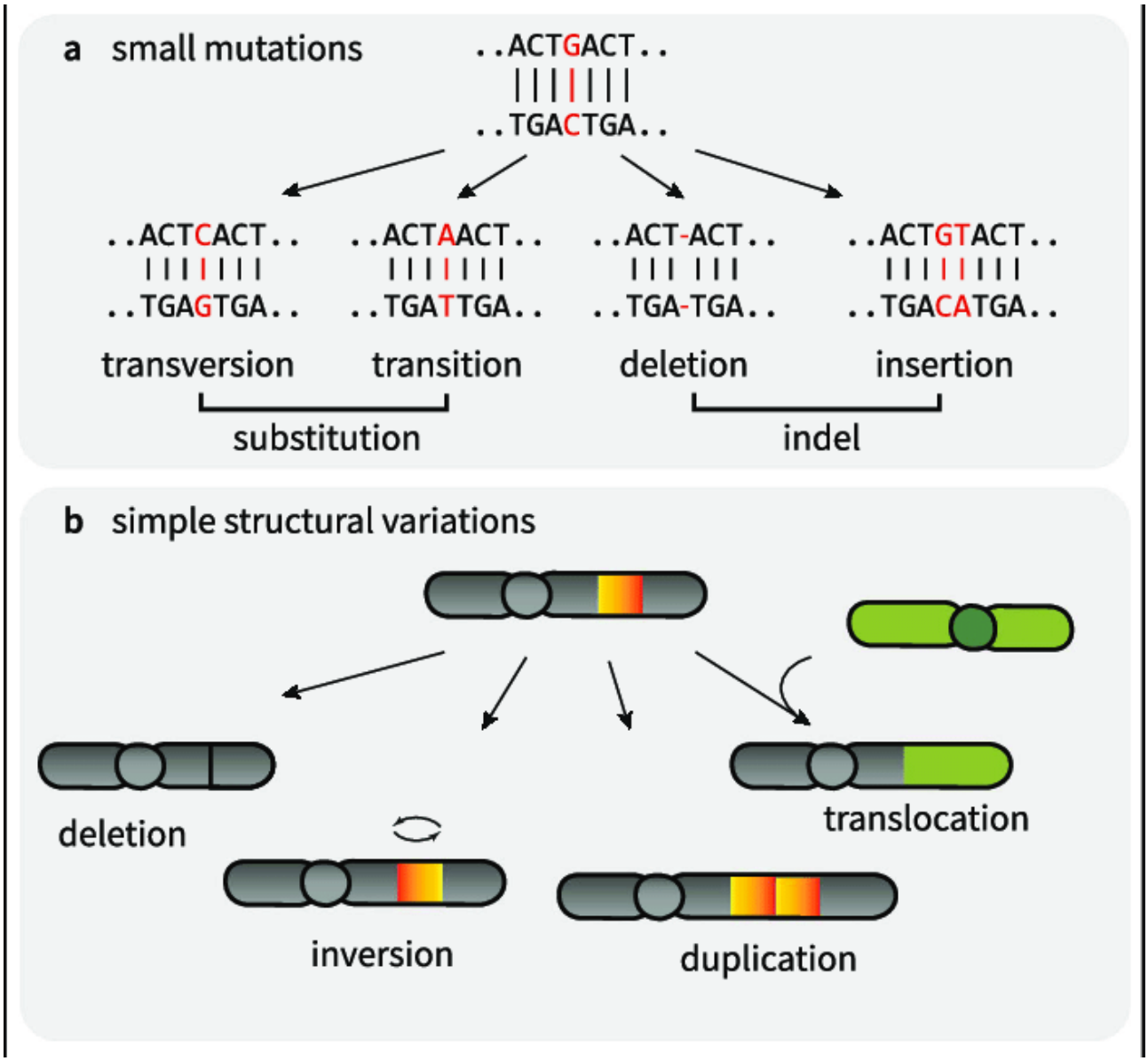
Authors: Alexandrov, L.B. et al.

Year: 2013

Citiations: 592

link: https://www.cell.com/cell-reports/fulltext/S2211-1247(12)00433-0

جهش های جسمی معمولاً به چهار کلاس گروه بندی می شوند.تعویض نوکلئوتید ، ایندل های کوچک ، بازآرایی مجدد و تغییر تعداد کپی ها.




امضای یک فرآیند جهش به عنوان یک تابع چگالی احتمال گسسته با دامنه ای از ویژگی های جهش از پیش انتخاب شده نشان داده می‌شود.

به طور ریاضی، می‌توان ویژگی‌های جهش را به صورت مجموعه الفبای متناهی Ξ با K حرف (هر حرف متناظر با یک ویژگی جهش) بیان کرد و طبق تعریف، فرایند جهش $P_1$ یک K-tuple با ترتیب lexicographical است؛ $P_1 = [p^1_1, p^2_1, p^K_1]^T$، که $p^i_1$ احتمال فرایند $p_1$ که ویژگی جهش متناظر با i ام حرف از الفبای Ξ باعث می‌شود را نشان می‌دهد و چون $p^i$ احتمال هستند:

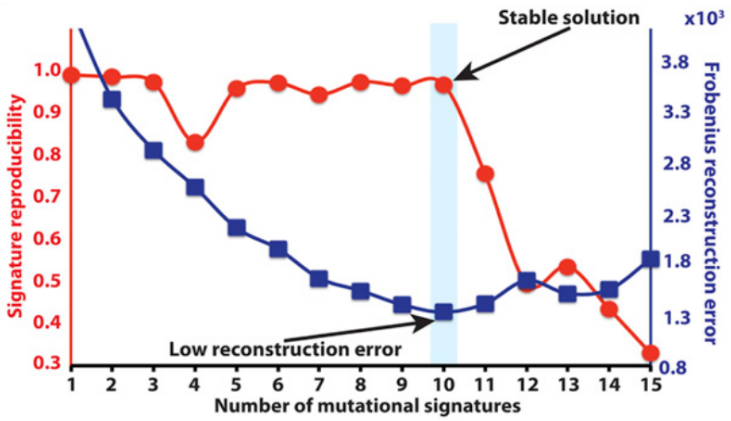
$$P_1 = [p^1_1, p^2_1, \ldots, p^K_1]^T \quad \sum_{k=1}^{K} p^k_1 = 1$$

ژنوم‌های سرطانی متفاوت می‌توانند در معرض فرایند جهش‌زا با شدت‌های متفاوت قرار گیرند. به عنوان مثال ، یک فرایند جهش‌زا می تواند باعث ایجاد ۱۰۰۰ جهش در یک ژنوم سرطان شود در حالی که باعث ایجاد ۲۰،۰۰۰ جهش در دیگری می شود. بنابراین یک فرایند جهش‌زا با امضای $P_1$ دارای شدت (تعداد جهش‌هایی که سبب می‌شود)، $e^1_g$ ، در ژنوم سرطانی g است. توجه داشته باشید که اندیس پایین امضا $P_1$ با اندیس بالا $e^1_g$ مطابقت دارد. بنابراین شدت $e^1_g$ با امضا $P_1$ مطابقت دارد.

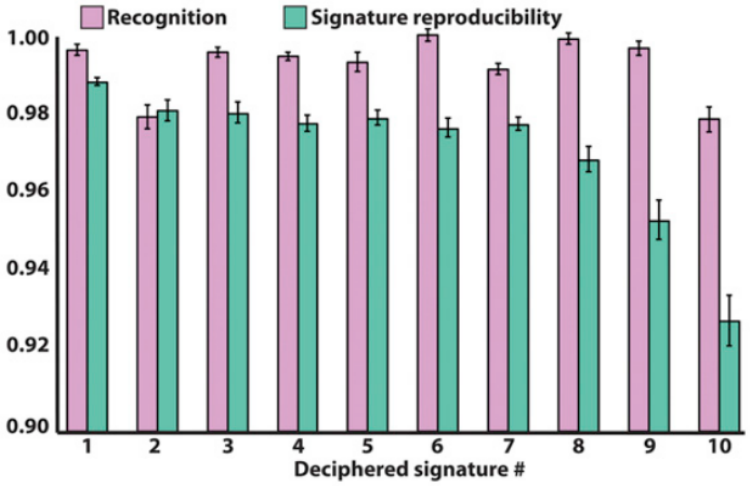
****

روش NMF بر روی ۱۰۰ فهرست جهش‌زا ژنوم سرطانی شبیه‌سازی شده پیاده‌سازی شده است. مشابه بسیاری از ژنوم‌های سرطانی انسان، هر ژنوم سرطانی شامل ۵۰۰ تا ۵۰۰۰۰ جهش جایگزینی است. جهش های شبیه‌سازی شده با استفاده از ۱۰ فرایند جهش‌زا با امضا مجزا هر کدام با ۹۶ نوع جهش (۶ نوع جهش جایگزین و دو همسایه مجاور) تولید شده‌اند.

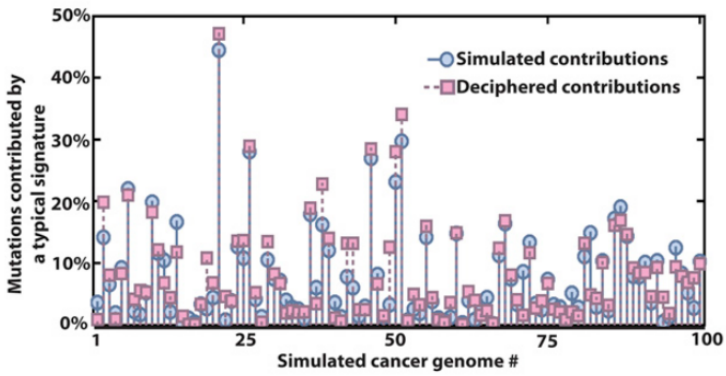
شناسایی تعداد، N، فرایندهای جهش‌زا در مجموعه ژنوم‌های سرطانی به یک دانش پیشین را برای کشف امضاهایشان نیاز دارد. در روش NMF مقدار N را با استفاده از مقادیر مختلف شناسایی می‌کنیم. برای هر N، شباهت بین فرایندهای به دست آمده در هر تکرار (قابلیت بازتولید فرایند) را ارزیابی می‌کنیم. علاوه بر این، برای هر N، متوسط خطای بازسازی (reconstruction error) فروبنیوس از متوسط امضاهای کشف شده $\bar{P}$ و شدت آن‌ها $\bar{E}$ ( $\|M - \bar{P} \times \bar{E}\|_F^2$ ) تعیین می‌شود. خطای کم بازسازی نشانگر توصیف دقیق از فهرست‌های ژنوم سرطان اصلی است. مقدار N ای را انتخاب می‌کنیم که فرایندهای به دست آمده قابلیت بازتولید و خطای بازسازی کمی داشته باشند. برای ۱۰۰ ژنوم سرطانی شبیه‌سازی شده، می‌توانیم جواب‌های قابل بازتولید را برای N با مقادیر بین ۲ و ۱۰، شناسایی کنیم (میزان شباهت بین تکرارهای یک فرایند به عبارت دیگر احتمال اینکه فرایندی که در هر تکرار تولید می‌شود، در یک خوشه قرار گیرد. ) . افزایش تعداد امضاها از ۲ به ۱۰ به طور قبل ملاحظه‌ای خطای بازسازی را کاهش می‌دهد، افزایش بیش از ۱۰ امضا بیش از این خطا را کاهش نمی‌دهد. این نشان می دهد که رویکرد ما می تواند "به صورت بهینه" امضاهای ده فرایند جهش را تشخیص دهد ، دقیقاً عددی که در ابتدا برای شبیه سازی فهرست های جهش یافته ۱۰۰ ژنوم سرطان استفاده شده است.



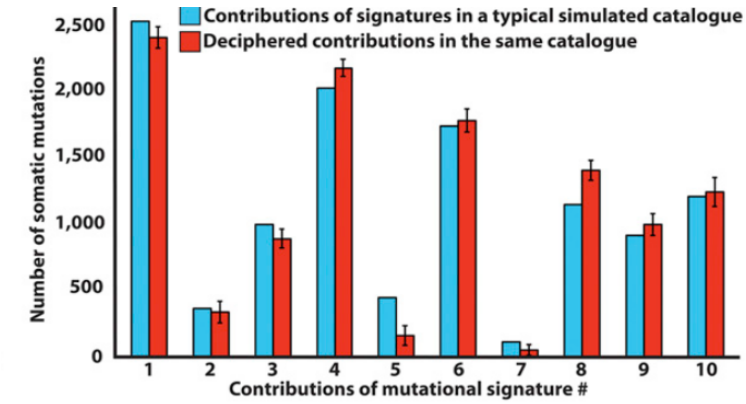

۱۰ امضای کشف شده قابلیت بازتولیدی بالایی دارند (شباهت امضاهایی تولید شده در هر تکرار در یک خوشه قرار می‌گیرند ۰.۹۶ <silhouette width) و همچنین شباهت بالایی به امضاهایی دارند که برای تولید ۱۰۰ فهرست جهش‌زا مورد استفاده قرار گرفتند (متوسط شباهت کسینوسی ۰.۹۸<).



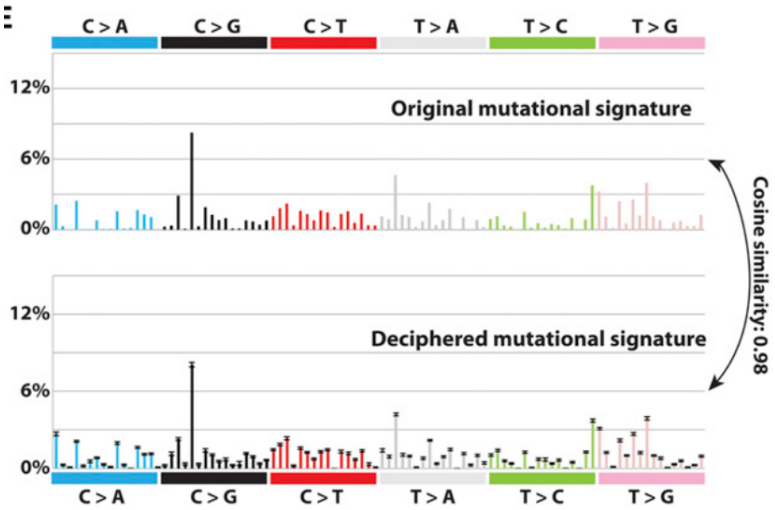

علاوه بر این ، روش NMF قادر به شناسایی دقیق تعداد جهش های انجام شده توسط هر یک از ده فرآیند در هر یک از ژنوم ها است. مقایسه بین سهم تعداد جهش‌ها در یکی از امضاهای اصلی و امضا کشف شده‌اش در کلیه ژنومها در شکل زیر نشان داده شده است.




مقایسه سهم تعداد جهش هر ۱۰ امضا در یک ژنوم واحد در شکل زیر نشان داده شده است.




مقایسه بین یک امضاء اصلی و کشف شده در شکل زیر نشان داده شده است.




مقایسه بین یک فهرست جهش‌زا اصلی و بازسازی شده یک ژنوم در شکل زیر نشان داده شده است.