

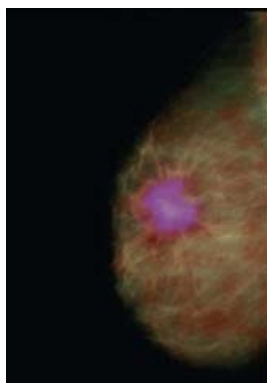
پیدا کردن ژنی که مسؤول سرطان است

نویسنده: برنار پرن^۱

مترجم: حوری سپهری

ویراستاران: فرج‌الله محمودی، ارسلان شادمان

پیشرفت‌های بیولوژی مدرن و به‌ویژه ژنتیک ملکولی نیاز به ابزار جدید ریاضی دارند. مثال آن آمار و نقش آن در جستجوی ژن مسؤول سرطان سینه است.

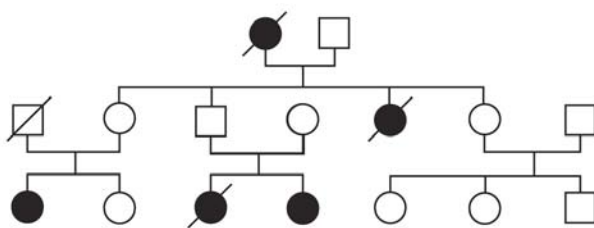


در این ماموگرافی (عکسبرداری از پستان) با رنگ غیرواقعی، یک غده سرطانی به رنگ صورتی به وضوح مشاهده می‌شود. قسمت مهمی از پژوهش‌های سرطان سینه معطوف به منظره ژنتیکی آن است. در این پژوهش‌ها تئوری آماری نقش اصلی را بازی می‌کند.

^۱ Prum, Bernard: *Trouver un gène responsable de cancer*:
in: *L'explosion des mathématiques*, SMF et SMAI, Paris, 2002, p. 28-31

بسیاری از بیماری‌ها ریشهٔ وراثتی دارند: احتمال بروز بیماری در افرادی که کم و بیش حامل ژن بیماری هستند بیشتر است. بدین جهت علم ژنتیک جدید در جستجوی آگاهی بر نقش ژن‌های مختلف و به‌ویژه عمل آن در پیدایش بیماری است، به این امید که روزی بتوان آن را معالجه کرد. مثلاً اشخاص مبتلا به سرطان سینه را در نظر بگیریم که در فرانسه از هر ۸ زن حدوداً یک نفر به آن مبتلاست. در کنار عوامل متعدد خطر سرطان (تغذیه، سیگار، قرار گرفتن در مقابل اشعه و غیره) چند سالی است ژنی را شناسایی کرده‌اند که موتاسیون‌های آن در درصد بالایی از زنان مبتلا به این نوع سرطان دخیل است. این ژن BRCA1 (سرطان سینه ۱)^۱ نامگذاری شده است. این نتیجهٔ بیومدیkal جز با نتایج آنالیزهای آماری نمی‌توانست بدست آید و به طوریکه خواهیم دید این نتایج اجازه داد که با دقت روزافزون محل ژن را به‌طور دقیق معین کنند.

علم ژنتیک مدت‌ها طبیعت مادی ژن را نمی‌شناخت. بیش از ۲۰ سال نیست که به سکانس‌های متعدد DNA دسترسی پیدا کرده‌اند. DNA زنجیره‌ای ملکولی می‌باشد که اطلاعات ژنتیکی را از والدین به فرزندان منتقل می‌کند. در عین حال، ندانستن ترکیب شیمیایی ژن‌ها مانع از گرفتن نتایج وراثتی دقیق دربارهٔ برخی صفات نبوده است.



شکل ۱. خانواده‌ای را نشان می‌دهد که تعدادی از افراد آن دارای سرطان سینه هستند. مربع‌ها برای مردها، دایره‌ها برای زن‌ها، دایرهٔ سیاه مربوط به فرد مبتلا و دایرهٔ سیاه خط‌دار مربوط به فرد مبتلایی است که مرده است. ملاحظه می‌کنیم که مادربزرگ، یکی از دختران و سه نوهٔ دختری او دارای سرطان سینه بوده‌اند. مطمئناً در سایر افراد خانواده بیماری می‌تواند حضور داشته باشد. به وسیلهٔ نتایج این شجره‌نامه، ژنتیک‌دانان به فرض وجود یک عامل ژنتیکی در این بیماری هدایت شده‌اند.

در مقابل یک بیماری نظیر سرطان سینه، نخستین سوال مطرح این است که «آیا این بیماری وراثتی است؟» و آیا ژن‌هایی مسؤوّل آن هستند؟ جواب برای سرطان مدت‌ها

مشکوک بود. برای جواب مثبت به این سؤال باید منتظر بود تا تمرکز تعداد افراد مبتلا در فامیل تعیین شود. مثلاً معلوم شود آیا دختر یا خواهرزنی مبتلا نسبت به سایرین بیشتر در معرض ابتلا است. مدت‌ها متخصصین آمار- ژنتیک نتایج را به صورت شجره‌نامه^۱ شکل شماره یک نمایش می‌دادند. باین شجره‌نامه چه باید کرد؟ تقریباً از مندل^۱ به بعد می‌دانیم که یک صفت وراثتی اغلب به وسیله ژنی مشخص می‌شود که دارای فرم‌های مختلف بنام آلل^۲ است. هر فرد یک آلل از پدر و یک آلل از مادر به ارث می‌برد، این فرد یکی از آن دو را به طور اتفاقی به فرزند خود منتقل می‌کند. ژنتیک‌دان‌ها برای انتقال یک بیماری مورد مطالعه، مدلی پیشنهاد می‌کنند که دخالت چند ژن و آلل را فرض می‌گیرد. ارزش این مدل با آزمون‌های آماری ویژه تعیین می‌شود، مثلاً باعث حذف فرضیه ساده‌ای از قبیل «بیماری مطالعه شده هیچ عامل ژنتیکی ندارد» می‌شوند.

در بیماری‌هایی که دارای عوامل پیچیده هستند «نظیر سرطان سینه» و عوامل متعدد محیطی و سن در آن دخالت دارند، باید داده‌ها را برحسب وابستگی به زمان بررسی کرد. در این حال باید از آمار فرایندها استفاده کرد که شاخه‌ای پیشرفته از ریاضی است و قسمت بزرگی از این پیشرفت مدیون نتایج حاصله از مکتب احتمالات فرانسه در سال‌های ۱۹۸۰ (پ.آ. میر، ژ. ژاکو)^۳ و مکتب آماری اسکاندیناوی می‌باشد.

آمارهایی برای مشخص کردن کروموزوم حامل ژن

پس از آن که آنالیز شجره‌نامه‌ها وجود ژن حساسیت به سرطان پستان را نشان داد، مرحله بعد تعیین محل ژن ولو به گونه‌ای اجمالی روی یک کروموزوم از ۲۳ کروموزوم انسانی است. برای این کار از سال‌های ۱۹۸۰ به بعد از نشانه‌گذارهایی استفاده می‌کنند، این نشانه‌گذارها زنجیرهای کوچک DNA کاملاً شناخته شده‌ای می‌باشند که می‌توان آنها را به آسانی و با کمترین هزینه مثلاً توسط تجزیه شیمیایی سریع تشخیص داد. علامت‌گذاری تعیین محل آن آسان است. علامت‌گذارها اجازه می‌دهند که شباهت بین مناطق کروموزوم‌های مورد امتحان در افراد بیمار و وابسته ارزیابی شوند. هر قدر شباهت یک منطقه کروموزومی در خویشاوندان مبتلا بیشتر باشد، به همان اندازه احتمال اینکه این

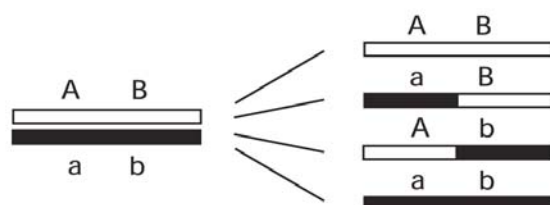
^۱ Mendel

^۲ allele

^۳ (P. A. Meyer - J. Jacod)

منطقه دارای یک ژن دخیل در بیماری باشد بیشتر است.

چنین آنالیز آماری مطمئناً پیچیده است زیرا هر والدی کروموزوم‌هایی را که از والدین خود گرفته است عیناً به فرزندان خود منتقل نمی‌کند بلکه به صورت ترکیب مجددی نظیر شکل (۲) آن‌ها را منتقل می‌نماید. اگر دو ژن را روی یک کروموزوم اولیه ملاحظه کنیم، آن‌ها را پس از ترکیب مجدد می‌توان روی دو کروموزوم متفاوت مشاهده کرد. به همان میزان که دو ژن مورد مطالعه از هم دورتر باشند احتمال پدیده فوق بیشتر است پس باید نرخ مشابهت در طول یک کروموزوم را تحلیل کرد، یعنی به بررسی یک فرایند تصادفی پرداخت. بنابراین با بررسی آماری فرایندها می‌توان فاصله‌ای را که در آن ژن مورد نظر قرار دارد مشخص کرد. به کار بردن علامت‌گذاری‌ها به اکیپ آمریکایی (جف م. هال)^۱ از دانشگاه برکلی اجازه داد که در سال ۱۹۹۰ ژن BRCA1 را روی کروموزوم ۱۷ تعیین نمایند.



شکل ۲. برای هر زوج از کروموزوم‌های یک فرد، یک کروموزوم از سوی پدر به ارث می‌رسد (با رنگ سیاه) و یک کروموزوم از سوی مادر (با رنگ سفید). به هر فرزند از سوی هر یک از والدین فقط یک کروموزوم منتقل می‌شود. اما پیش از انتقال، کروموزوم‌های هر زوج می‌توانند قطعاتی را بین خود به طور تصادفی مبادله کنند. این فرآیند که تجدید ترکیب نامیده می‌شود موجب آن است که از سوی والدین کروموزومی با ترکیب جدید منتقل شود که به شکل یکی از چهار صورت درج شده در شکل است (در این شکل فرض شده است که کروموزوم‌ها دو ناحیه را با هم مبادله کرده‌اند).

خواندن ملکول DNA برای مشخص کردن کامل ژن و ناهنجاری‌های آن

کار بعدی مشخص کردن دقیق محل ژن و شناختن ساختار آن است. می‌دانیم که ماده ژنتیکی DNA زنجیر ملکولی کشیده‌ای است که با الفبای مرکب از چهار حروف t, g, c, a نوشته می‌شود (این چهار حرف، حرفهای اول چهار نوع ملکول تشکیل دهنده زنجیره DNA می‌باشند). بانک‌های داده‌های ژنتیکی چندین میلیارد از این حروف را در اختیار

^۱ Jeff. M. Hall

دارند (حدود ۲۵ میلیون در روز به آن اضافه می‌شود).

با دقتی که در روش علامت‌گذارها وجود دارد می‌توان یک ژن را روی یک سکانس DNA که دارای ۴ میلیون حرف است معین کرد. برای دانستن این‌که کدام آلل یا کدام موتاسیون مسؤول مثلاً بیماری سرطان سینه است، باید این سکانس‌ها را در افراد سالم و بیمار مقایسه کرد. این عمل مانند آن‌است که یک غلط تایپی را در مطلبی با ۴ میلیون کاراکتر یعنی کتابی حدود ۲۰۰۰ صفحه پیدا کرد، یا بهتر بگوییم باید به تعداد افراد مورد مطالعه کتاب ۲۰۰۰ صفحه‌ای در نظر بگیریم و در آنها یک غلط تایپی پیدا کنیم. حتی با استفاده از امکانات وسیع کامپیوتری این کار بسیار سنگین است. اما در انسان ژن‌ها بیش از ۳٪ کروموزوم‌ها را تشکیل نمی‌دهند و بقیه ماده کروموزومی را بین ژنی نامند. بنابراین اگر پژوهش محدود به غلط تایپی ژن شود تعداد صفحات به ۳۰ صفحه می‌رسد که برای همه کامپیوترها قابل محاسبه می‌باشد.

اما چگونه ژن‌ها را از بقیه کروموزوم تشخیص می‌دهند. مسلم شده است که «سبک» نوشتن ژن‌ها با روش مواد بین ژنی متفاوت است: فرکانس توالی حروف یکسان نیست و از این تفاوت می‌توان برای تشخیص سکانس‌های بین ژنی استفاده کرد. می‌توان این اختلاف سبک را مورد بهره‌برداری قرار داد و به کمک آن در سکانس مورد نظر ژن‌ها و بخش‌های بین ژنی را از هم تفکیک کرد. این چالش مشکل است. باید از مدل‌های آماری که زنجیره مارکف^۱ مخفی خوانده می‌شود و در سال‌های ۱۹۸۰ در ارتباط با مسائل بازشناسی خودبخود گفتار بسط یافته بود کمک گرفت. استفاده از زنجیره‌های مارکف مخفی در ژنومیک هم زمان با استفاده از الگوریتم‌هایی صورت گرفت که هم قادر بودند سبک‌های مختلف را مشخص کنند و هم به هر وضعیت روی کروموزوم سبک مناسب آن را نسبت دهند.

به این ترتیب می‌توان محل BRCA1 را به صورت دقیقی مشخص کرد و از این پس به سهولت در هر بیمار آن را خواند. ژن حساسیت به سرطان سینه ۵۵۹۲ حرف و بیشتر از ۸۰ آلل دارد. کار جدید آمارگران ارتباط بین آلل‌های متفاوت و اهمیت آن‌ها در بیماری سرطان است.

بیولوژی زمینه جدید فعالیت را به ریاضی تقدیم می‌کند

مثال ژن BRCA1 مؤید این است که احتمالاً بیولوژی در مقابل ریاضی همان نقشی را

^۱ Markov..

بازی خواهد کرد که در قسمت مهمی از قرن بیستم فیزیک در مقابل ریاضی داشته است: دادن زمینه کاربردی به ابزارهای تئوری جدید و تدارک ابزارهای جدید (در اینجا به ابزارهای آماری متوسل می‌شویم اما می‌توانستیم از یکی دیگر از زمینه‌های ریاضی نظیر سیستم دینامیک، بهینه‌سازی تا هندسه کمک بگیریم. می‌دانیم که تشکل فضایی ملکول‌ها در عمل آنها نقش اساسی دارد). امروزه چالشی جدید برای متخصصین آمار بوجود آمده است: در حال حاضر می‌توان چندین میلیون معرف را روی یک سانتیمتر مربع شیشه جا داد و فهمید کدامیک از ژن‌ها در چه بافت‌هایی و در چه شرایط آزمایشگاهی یا... در کدام سلول سرطانی عمل می‌نمایند. درصدها مورد با شرایط مختلف اندازه‌گیری‌هایی در آزمایشگاه‌ها انجام شده و در نتیجه تعداد زیادی داده عددی فراهم شده است که بیان عبارتهای هزاران ژن را عرضه می‌کنند. در حال حاضر تنها آنالیزهای آماری میتواند ادعا کند قادر است آنها را بررسی کند و به‌طور دقیق ارتباط بین ژن‌ها و بیماری‌ها را روشن سازد.

برنار پرن

آزمایشگاه آمار و ژنوم

(UMR CNRS 8071)

ژنوپول، دانشگاه اوری

چند مرجع

- B. Prum, "Statistique et génétique" dans *Development of Mathematics 1950-2000* (sous la dir. de J.-P. Pier, Birkhäuser, 2000).
- C. Bonaïti-Pellié, F. Doyon et M. G. Lé, "Où en est l'épidémiologie du cancer en l'an 2001", *Médecine-Science*, 17, pp. 586-595 (2001).
- F. Muri-Majoube et B. Prum, "Une approche statistique de l'analyse des génomes", *Gazette des mathématiciens*, n° 89, pp. 63-98 (juillet 2001).
- B. Prum, "La recherche automatique des gènes", *La Recherche*, n° 346, pp. 84-87 (2001).

- M. S. Waterman, *Introduction to computational biology* (Chapman & Hall, 1995).

Bernard Prum
Laboratoire Statistique et Génome
(UMR CNRS 8071),
La Génopole, Université d'Évry