

Giải trình tự DNA thế hệ mới và các ứng dụng trong y tế, sức khỏe

TS Nguyễn Ngọc Kim Vy

Đại học California, San Diego, Hoa Kỳ

Hiện nay, công nghệ giải trình tự gen đang được ứng dụng nhiều ở lĩnh vực chẩn đoán, tầm soát và chữa trị bệnh. Các thai phụ có thể làm những chẩn đoán trước sinh nhanh và chính xác mà không cần phải chịu những rủi ro và nguy hiểm do chọc ối. Các bệnh di truyền có thể được phát hiện sớm hơn. Trong điều trị, trình tự DNA của bệnh nhân có thể giúp các bác sĩ điều chỉnh phác đồ điều trị phù hợp và hiệu quả hơn. Tất cả những điều này có được là nhờ vào những tiến bộ của công nghệ giải trình tự gen thế hệ mới với thời gian và chi phí ngày càng giảm.

Tên gọi giải trình tự gen thế hệ mới (next generation sequencing - NGS) là để phân biệt với phương pháp giải trình tự thế hệ đầu tiên do Frederick Sanger phát triển vào năm 1977. Vào thời điểm đó, phương pháp Sanger không phải là duy nhất, nhưng hiệu quả và dễ sử dụng hơn cả, nên rất được ưa chuộng. Trong suốt 25 năm sau đó, phương pháp này liên tiếp được cải tiến và tự động hoá để chế tạo thành máy giải trình tự (sequencer) thế hệ đầu tiên và đóng vai trò chủ đạo trong dự án giải trình tự bộ gen người đã được công bố vào năm 2001.

Vào đầu những năm 2000, một loạt công nghệ mới được phát triển nhằm tăng năng suất và hiệu quả trong việc giải trình tự DNA. Trong số đó, phương pháp của Công ty Solexa tỏ ra hiệu quả hơn cả nên đã được Công ty Illumina mua lại và phát triển thành một thế hệ máy giải trình tự mới. Hiện nay các dòng máy giải trình tự của Illumina như Miseq, HiSeq, NextSeq... là phổ biến nhất trên thị trường, được sử dụng rộng rãi trong nghiên cứu cơ bản lẫn chẩn đoán và điều trị.



Hình 1. Các sợi DNA.

Để hình dung sự khác nhau cơ bản trong nguyên tắc của hai phương pháp giải trình tự gen thế hệ mới và thế hệ đầu tiên, chúng ta hãy nhớ lại một chút về hình dạng của phân tử DNA. Nếu được phóng đại lên, sợi DNA có hình dạng giống sợi mì (hình 1), nhưng thay vì được cấu thành từ các đơn phân là glucose như sợi mì, thì sợi DNA này được kết nối bởi hàng triệu phân tử nucleotide gọi tắt là A, T, C, G (mỗi chữ cái tương ứng với một trình tự). Giải trình tự gen là đi tìm thứ tự đúng của từng phân tử A, T, C, G trên một gen nhỏ hoặc trong toàn bộ

gen của sinh vật.

Mỗi tế bào người (thường) có 46 “sợi mì” DNA khổng lồ (tức là các nhiễm sắc thể). Lấy ví dụ một sợi mì nhỏ nhất của người là nhiễm sắc thể thứ 21, có khoảng 50 triệu trình tự. Trong khi đó, các phương pháp hiện nay chỉ giải được cho một cọng mì nhỏ xíu có chiều dài khoảng 500-2.000 trình tự mà thôi. Như vậy, người ta phải cắt nhỏ sợi mì thứ 21 ra thành 20.000-100.000 cọng mì nhỏ để giải (con số có tính chất ước lượng, trên thực tế có thể nhiều hơn vì các cọng mì

bị cắt ra có thể chồng lên nhau và phải được giải nhiều lần để tăng độ chính xác), rồi ráp nối kết quả lại với nhau. Mỗi lần giải với phương pháp Sanger thường chỉ giải 1 cọng mì nhỏ, nên chi phí khá cao và hiệu quả thấp. Ngược lại, phương pháp NGS cho phép xử lý hàng triệu cọng mì nhỏ cùng một lúc, kết quả được ghi chép và lưu trữ dưới dạng dữ liệu tin học. Vì vậy, hai điểm mạnh của NGS là rút ngắn thời gian và giảm chi phí. Nếu lấy trên đơn vị là 1 triệu trình tự, thì phương pháp Sanger sẽ tốn tới 2.400 USD, trong khi phương pháp NGS chỉ tốn 0,05-0,15 USD.

Với giá thành thấp như vậy, người ta có thể giải với số lượng nhiều và thu được lượng dữ liệu rất lớn. Nhưng từ kho dữ liệu này đến kết quả đáng tin cậy để gửi trả cho khách hàng là một đoạn đường dài, đòi hỏi các phương pháp tính toán từ sinh - tin học và xác suất thống kê hợp lý. Vì vậy, ngoài máy móc và công nghệ, thì quan trọng là người phân tích cần có hiểu biết trong cả 3 lĩnh vực: sinh học, tin học và tiến hoá phân tử.

Với cách thiết kế đa dạng và linh động, phương pháp giải trình tự thế hệ mới có thể được biến đổi để phù hợp cho nhiều mục đích khác nhau. Trong ứng dụng, có 3 quy trình phổ biến nhất:

Giải trình tự cả bộ gen: là xác định trình tự cho cả bộ gen của sinh vật, trong đó có con người. Mặc dù công nghệ hiện nay tiến bộ như vậy, nhưng vẫn chưa có bộ gen của động vật có vú nào, kể cả con người được giải mã hoàn toàn. Mặc dù vẫn còn nhiều hạn chế và thách thức khi ứng dụng rộng rãi, nhưng quy trình này đã tìm ra được các vị trí trong bộ gen liên quan đến nhiều loại bệnh như tự kỷ, xơ cứng teo một bên,



Hình 2. Máy NextSeq đang hoạt động tại Viện Di truyền y học, TP Hồ Chí Minh.

rối loạn trương lực (dystonia)...

Giải toàn bộ exon: exon chỉ chiếm 1% của bộ gen nhưng chứa thông tin trực tiếp mã hoá cho các protein thực hiện chức năng trong cơ thể. Vì vậy, những thay đổi trình tự của exon có thể liên quan trực tiếp đến bệnh tật hay tình trạng sức khỏe của con người. Quy trình này đã được thương mại hoá nên dễ thực hiện và khá phổ biến.

Giải một/vài gen cụ thể: là xác định trình tự của một/vài gen có liên quan đến một loại bệnh nào đó, ví dụ nhóm các gen liên quan đến hội chứng ung thư di truyền. Vì số lượng ít và mục tiêu rõ ràng, nên kết quả và độ chính xác của việc giải trình tự này sẽ cao hơn, còn chi phí sẽ thấp hơn.

Hiện nay, ngoài Mỹ thì Trung Quốc cũng phát triển rất mạnh và nắm giữ nhiều công nghệ giải trình tự mới. Họ cũng đã tự sản xuất được máy giải trình tự. Điển hình là Viện Di truyền học Bắc Kinh đã tiến hành giải trình tự bộ gen rất nhiều loài động, thực vật và tất nhiên họ có nguồn dữ liệu khổng lồ về bộ gen con người. Với tiềm lực về công nghệ, tài chính và nhân lực như vậy, các công ty Trung Quốc có thể dễ dàng xâm

nhập vào thị trường Việt Nam và các nước châu Á.

Ngoài kinh doanh và lợi nhuận, có một vấn đề quan trọng cần được quan tâm nếu chúng ta sử dụng công nghệ và dịch vụ của nước ngoài. Đó là họ sẽ có quyền sở hữu và khai thác dữ liệu về bộ gen của người Việt Nam. Bộ gen chứa vô vàn thông tin về đặc điểm di truyền, sức khỏe... của người Việt. Đó là một nguồn dữ liệu vô cùng quý giá mà chúng ta dù chưa khai thác được hết nhưng không thể để mất. Và nếu không muốn viễn cảnh đó xảy ra, thì không cách gì khác là người Việt phải nắm bắt được công nghệ này và đưa vào sử dụng để phục vụ nhu cầu đang ngày một tăng lên trong ngành y tế và sức khỏe. Và hiển nhiên, sự chọn lựa và hiểu biết của khách hàng khi sử dụng dịch vụ cũng là một yếu tố quan trọng. Nhiều máy giải trình tự của Illumina hiện đã được sử dụng tại Việt Nam. Hy vọng một ngày không xa, công nghệ này và các ứng dụng của nó sẽ được thực hiện rộng rãi hơn, tổ chức chặt chẽ và bài bản hơn (kết hợp với tư vấn di truyền) và giá thành thấp hơn, phù hợp với khả năng tài chính của người Việt.

TÀI LIỆU THAM KHẢO

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4391380/>.
2. <https://www.statnews.com/2017/06/20/human-genome-not-fully-sequenced/>.
3. https://en.wikipedia.org/wiki/DNA_sequencing.
4. https://en.wikipedia.org/wiki/Whole_genome_sequencing.