



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẢNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0041317

(51)⁸ C12Q 1/68; G06F 19/24; G06F 19/22; (13) B
G06F 19/00; G06F 19/18

(21) 1-2018-03888

(22) 20/12/2016

(86) PCT/US2016/067886 20/12/2016

(87) WO 2017/136059 A1 10/08/2017

(30) 62/290,891 03/02/2016 US; 15/382,508 16/12/2016 US

(45) 25/10/2024 439

(43) 26/11/2018 368A

(73) VERINATA HEALTH, INC. (US)

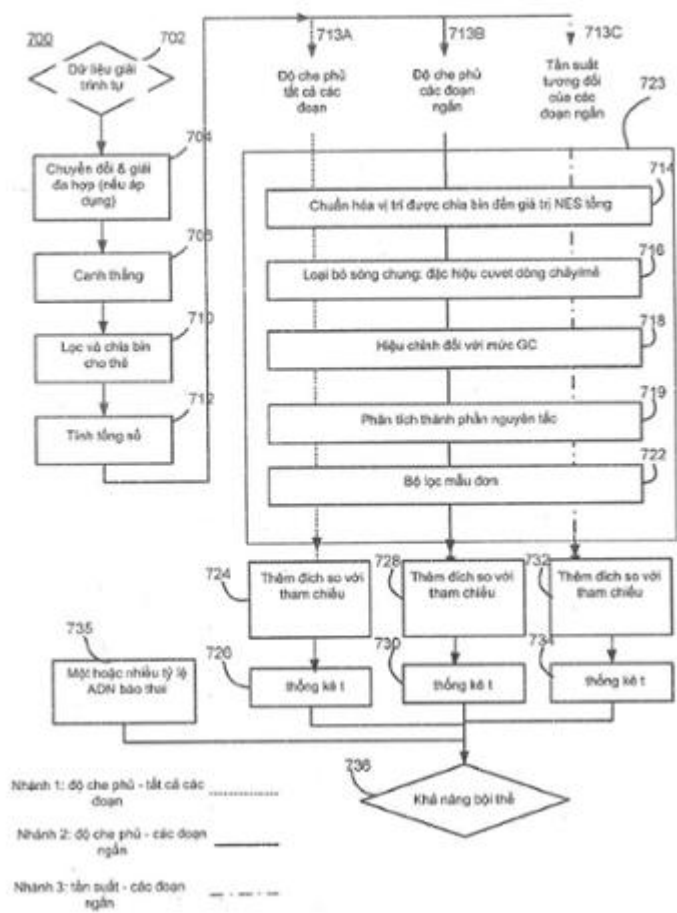
5200 Illumina Way, San Diego, California 92122, United States of America

(72) DUENWALD, Sven (US); COMSTOCK, David A. (US); BARBACIORU, Catalin (US); CHUDOVA, Darya I. (US); RAVA, Richard P. (US); JONES, Keith W. (US); CHEN, Gengxin (CN); SKVORTSOV, Dimitri (DE).

(74) Công ty TNHH Quốc tế D & N (D&N INTERNATIONAL CO.,LTD.)

(54) PHƯƠNG PHÁP VÀ HỆ THỐNG XÁC ĐỊNH BIẾN THỂ SỐ LƯỢNG BẢN SAO CỦA TRÌNH TỰ AXIT NUCLEIC

(57) Sáng chế đề xuất các phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) được biết hoặc nghi là có liên quan đến nhiều tình trạng bệnh lý. Theo một số phương án, sáng chế đề xuất các phương pháp xác định biến thể số lượng bản sao của bào thai nhờ sử dụng các mẫu từ người mẹ chứa ADN ngoài tế bào của người mẹ và bào thai. Sáng chế còn đề xuất hệ thống ước lượng số lượng bản sao của trình tự axit nucleic.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề xuất các phương pháp và hệ thống xác định biến thể số lượng bản sao (copy number variation - CNV) được biết hoặc nghi là có liên quan đến nhiều tình trạng bệnh lý.

Tình trạng kỹ thuật của sáng chế

Một trong các nỗ lực có tính quyết định trong nghiên cứu y học ở người là việc phát hiện ra các bất thường di truyền mà gây ra các hậu quả xấu đối với sức khỏe. Trong nhiều trường hợp, các gen đặc hiệu và/hoặc gen đánh dấu chẩn đoán quan trọng đã được nhận diện ở các phần của bộ gen mà có mặt ở các số lượng bản sao bất thường. Chẳng hạn, trong chẩn đoán tiền sản, việc thừa hoặc thiếu bản sao của toàn bộ nhiễm sắc thể là các thương tổn di truyền thường xảy ra. Ở bệnh ung thư, sự khuyết đoạn hoặc nhân lên của các bản sao của toàn bộ nhiễm sắc thể hoặc các đoạn nhiễm sắc thể, và sự khuếch đại mức độ cao của các vùng gen đặc hiệu, là các biểu hiện phổ biến.

Hầu hết thông tin về biến thể số lượng bản sao (CNV) đã được đưa ra bằng cách giải di truyền học tế bào mà cho phép sự nhận biết các bất thường cấu trúc. Các quy trình thông thường để sàng lọc di truyền và định lượng sinh học đã sử dụng các thủ thuật xâm lấn, ví dụ, chọc ối, lấy máu bào thai, hoặc sinh thiết nhau thai (chorionic villus sampling - CVS), để thu được các tế bào để phân tích các kiểu hình nhân. Nhận thấy nhu cầu về các phương pháp thử nghiệm nhanh hơn mà không đòi hỏi việc nuôi cấy tế bào, kỹ thuật lai huỳnh quang tại chỗ (fluorescence in situ hybridization - FISH), PCR huỳnh quang định lượng (quantitative fluorescence PCR - QF-PCR) và lai so sánh bộ gen dựa trên mảng (array-comparative Genomic Hybridization - array-CGH) đã được phát triển làm các phương pháp di truyền học phân tử-tế bào để phân tích các biến thể số lượng bản sao.

Một trong các nỗ lực có tính quyết định trong nghiên cứu y học ở người là việc phát hiện ra các bất thường di truyền mà gây ra các hậu quả xấu đối với sức khỏe. Trong nhiều trường hợp, các gen đặc hiệu và/hoặc gen đánh dấu chẩn đoán quan trọng đã được nhận diện ở các phần của bộ gen mà có mặt ở các số lượng bản sao bất thường. Chẳng hạn, trong chẩn đoán tiền sản, việc thừa hoặc thiếu bản sao của toàn bộ nhiễm sắc thể là các thương tổn di

truyền thường xảy ra. Ở bệnh ung thư, sự khuyết đoạn hoặc nhân lên của các bản sao của toàn bộ nhiễm sắc thể hoặc các đoạn nhiễm sắc thể, và sự khuếch đại mức độ cao của các vùng gen đặc hiệu, là các biểu hiện phổ biến. Hầu hết thông tin về biến thể số lượng bản sao (CNV) đã được đưa ra bằng cách giải di truyền học tế bào mà cho phép sự nhận biết các bất thường cấu trúc. Các quy trình thông thường để sàng lọc di truyền và định lượng sinh học đã sử dụng các thủ thuật xâm lấn, ví dụ, chọc ối, lấy máu bào thai, hoặc sinh thiết nhau thai (chorionic villus sampling - CVS), để thu được các tế bào để phân tích các kiểu hình nhân. Nhận thấy nhu cầu về các phương pháp thử nghiệm nhanh hơn mà không đòi hỏi việc nuôi cấy tế bào, kỹ thuật lai huỳnh quang tại chỗ (fluorescence in situ hybridization - FISH), PCR huỳnh quang định lượng (quantitative fluorescence PCR - QF-PCR) và lai so sánh bộ gen dựa trên mảng (array-comparative Genomic Hybridization - array-CGH) đã được phát triển làm các phương pháp di truyền học phân tử-tế bào để phân tích các biến thể số lượng bản sao.

Tiến bộ công nghệ mà cho phép giải trình tự toàn bộ bộ gen trong thời gian tương đối ngắn, và phát hiện về ADN lưu thông ngoài tế bào (cell-free ADN - cfADN) đã tạo ra cơ hội so sánh vật liệu di truyền có nguồn gốc từ một nhiễm sắc thể để so sánh với vật liệu di truyền của nhiễm sắc thể khác mà không gặp phải các nguy cơ do các phương pháp lấy mẫu xâm lấn, điều này tạo ra công cụ để chẩn đoán các loại các biến thể số lượng bản sao khác nhau của các trình tự gen quan tâm.

Các hạn chế của các phương pháp hiện có trong chẩn đoán tiền sản không xâm lấn, mà bao gồm độ nhạy thấp xuất phát từ các mức hạn chế của cfADN, và sai số giải trình tự của công nghệ xuất phát từ bản chất vốn có của thông tin di truyền, là cơ sở cho nhu cầu tiếp tục về các phương pháp không xâm lấn mà sẽ mang lại bất kỳ hoặc tất cả trong số tính đặc hiệu, độ chính xác và tính ứng dụng, để chẩn đoán một cách đáng tin cậy sự thay đổi số lượng bản sao ở nhiều bối cảnh lâm sàng. Đã cho thấy rằng độ dài trung bình của các đoạn cfADN của bào thai là ngắn hơn so với các đoạn cfADN người mẹ trong huyết tương của phụ nữ mang thai. Khác biệt giữa cfADN bào thai và người mẹ này được khai thác trong phương án thực hiện sáng chế ở đây để xác định CNV và/hoặc tỷ lệ ADN bào thai. Các phương án được bộc lộ ở đây đáp ứng một số trong số các nhu cầu trên. Một số phương án có thể được thực hiện với việc tạo thư viện không dùng PCR kết hợp với giải trình tự ADN đầu bắt cặp (paired end).

Một số phương án đưa đến độ nhạy và tính đặc hiệu phân tích cao cho các chẩn đoán tiền sản không xâm lấn và các chẩn đoán cho nhiều loại bệnh.

Bản chất kỹ thuật của sáng chế

Theo một số phương án, sáng chế đề xuất các phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của thể lệch bội bào thai bất kỳ, và các CNV được biết hoặc nghi là có liên quan đến nhiều tình trạng y học. Các CNV mà có thể được xác định theo phương pháp của sáng chế bao gồm các thể ba nhiễm và các thể đơn nhiễm của bất kỳ trong số một hoặc nhiều nhiễm sắc thể 1-22, X và Y, các thể đa bội nhiễm sắc thể khác, và các đột biến mất đoạn và/hoặc sao đôi của các đoạn của một hoặc nhiều trong số các nhiễm sắc thể bất kỳ. Theo một số phương án, phương pháp bao gồm việc nhận dạng các CNV của trình tự axit nucleic quan tâm, ví dụ, trình tự liên quan về mặt lâm sàng, trong mẫu thử nghiệm. Phương pháp này đánh giá biến thể số lượng bản sao của trình tự quan tâm cụ thể.

Theo một số phương án, phương pháp được thực hiện ở hệ thống máy tính mà bao gồm một hoặc nhiều bộ xử lý và bộ nhớ hệ thống để ước lượng số lượng bản sao của trình tự axit nucleic quan tâm trong mẫu thử nghiệm chứa các axit nucleic của một hoặc nhiều bộ gen.

Một khía cạnh của sáng chế đề xuất phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của trình tự axit nucleic quan tâm trong mẫu thử nghiệm bao gồm các đoạn axit nucleic ngoài tế bào có nguồn gốc từ hai hoặc nhiều bộ gen. Phương pháp bao gồm các bước: (a) nhận các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm; (b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu bao gồm trình tự quan tâm, nhờ đó tạo ra các thể trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin; (c) xác định các kích thước đoạn của ít nhất một số trong số các đoạn axit nucleic ngoài tế bào có mặt trong mẫu thử nghiệm; (d) tính độ che phủ của các thể trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin: (i) xác định số lượng các thể trình tự sắp thẳng với bin, và (ii) chuẩn hóa số lượng các thể trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; (e) xác định tham số thống kê cho trình tự quan tâm nhờ sử dụng độ che phủ của các bin ở trình tự quan tâm và độ che phủ của các bin trong vùng tham chiếu đối với trình tự quan tâm; và (f) xác định biến thể số lượng bản sao ở trình tự quan tâm

nhờ sử dụng tỷ số khả dĩ tính được từ tham số thống kê t và thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.

Theo một số phương án thực hiện, phương pháp bao gồm việc thực hiện (d) và (e) hai lần, một lần cho các đoạn ở domain kích thước thứ nhất và một lần nữa cho các đoạn ở domain kích thước thứ hai. Theo một số phương án thực hiện, domain kích thước thứ nhất bao gồm các đoạn axit nucleic ngoài tế bào có gần như tất cả các kích thước trong mẫu, và domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn kích thước xác định. Theo một số phương án thực hiện, domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn khoảng 150 bp. Theo một số phương án thực hiện, tỷ số khả dĩ được tính từ tham số thống kê t thứ nhất cho trình tự quan tâm nhờ sử dụng các thẻ trình tự cho các đoạn trong khoảng kích thước thứ nhất, và tham số thống kê t thứ hai cho trình tự quan tâm nhờ sử dụng các thẻ trình tự cho các đoạn trong khoảng kích thước thứ hai.

Theo một số phương án thực hiện, tỷ số khả dĩ được tính là khả năng thứ nhất mà mẫu thử nghiệm là mẫu lệch bội so với khả năng thứ hai mà mẫu thử nghiệm là mẫu nguyên bội.

Theo một số phương án thực hiện, tỷ số khả dĩ được tính từ một hoặc nhiều giá trị của tỷ lệ ADN bào thai ngoài tham số thống kê t và thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.

Theo một số phương án thực hiện, một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin về kích thước của các đoạn axit nucleic ngoài tế bào. Theo một số phương án thực hiện, giá trị của tỷ lệ ADN bào thai được tính bằng cách: thu được phân bố tần suất của các kích thước đoạn; và áp dụng phân bố tần suất này cho mô hình liên hệ tỷ lệ ADN bào thai với tần suất kích thước đoạn để thu được giá trị tỷ lệ ADN bào thai. Theo một số phương án thực hiện, mô hình liên hệ tỷ lệ ADN bào thai với tần suất của kích thước đoạn bao gồm mô hình tuyến tính chung có nhiều điều kiện và hệ số cho nhiều kích thước đoạn.

Theo một số phương án thực hiện, một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin độ che phủ đối với các bin của bộ gen tham chiếu; Theo một số phương án thực hiện, giá trị của tỷ lệ ADN bào thai được tính bằng cách áp dụng các giá trị độ che phủ của nhiều bin cho mô hình liên hệ tỷ lệ ADN bào thai với độ che phủ của bin để thu được giá trị tỷ lệ ADN bào thai. Theo một số phương

án thực hiện, mô hình liên hệ tỷ lệ ADN bào thai với độ che phủ của bin bao gồm mô hình tuyến tính chung có nhiều điều kiện và hệ số cho nhiều bin. Theo một số phương án thực hiện, nhiều bin có sự tương quan cao giữa tỷ lệ ADN bào thai và độ che phủ ở các mẫu huấn luyện.

Theo một số phương án thực hiện, một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng các tần suất của nhiều 8-mer tìm thấy trong các trình tự đọc. Theo một số phương án thực hiện, giá trị của tỷ lệ ADN bào thai được tính bằng cách áp dụng các tần suất nhiều 8-mer cho mô hình liên hệ tỷ lệ ADN bào thai với tần suất 8-mer để thu được giá trị tỷ lệ ADN bào thai. Theo một số phương án thực hiện, mô hình liên hệ tỷ lệ ADN bào thai với tần suất 8-mer bao gồm mô hình tuyến tính chung có nhiều điều kiện và hệ số cho nhiều 8-mer. Theo một số phương án thực hiện, nhiều 8-mer có sự tương quan cao giữa tỷ lệ ADN bào thai và tần suất 8-mer.

Theo một số phương án thực hiện, một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin độ che phủ đối với các bin thuộc nhiễm sắc thể giới tính.

Theo một số phương án thực hiện, tỷ số khả dĩ được tính từ tỷ lệ ADN bào thai, tham số thống kê t của các đoạn ngắn, và tham số thống kê t của tất cả các đoạn, trong đó các đoạn ngắn là các đoạn axit nucleic ngoài tế bào trong khoảng kích thước thứ nhất nhỏ hơn kích thước tiêu chuẩn, và tất cả các đoạn này là các đoạn axit nucleic ngoài tế bào bao gồm các đoạn ngắn và các đoạn dài hơn kích thước tiêu chuẩn. Theo một số phương án thực hiện, tỷ số khả dĩ được tính:

$$LR = \frac{\sum_{ff_{total}} q(ff_{total}) * p_1(T_{short}, T_{all} | ff_{est})}{p_0(T_{short}, T_{all})}$$

trong đó p_1 tượng trưng cho khả năng mà dữ liệu đến từ phân bố chuẩn nhiều biến biểu diễn mô hình 3 bản sao hoặc 1 bản sao, p_0 tượng trưng cho khả năng mà dữ liệu đến từ phân bố chuẩn nhiều biến biểu diễn mô hình 2 bản sao, T_{short} , T_{all} là các điểm số T tính được từ độ che phủ nhiễm sắc thể được tạo ra từ các đoạn ngắn và tất cả các đoạn, và $q(ff_{total})$ là phân bố mật độ của tỷ lệ ADN bào thai.

Theo một số phương án thực hiện, tỷ số khả dĩ được tính từ một hoặc nhiều giá trị của tỷ lệ ADN bào thai ngoài tham số thống kê t và thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.

Theo một số phương án thực hiện, tỷ số khả dĩ được tính cho thẻ đơn nhiệm X, thẻ ba nhiệm X, thẻ ba nhiệm 13, thẻ ba nhiệm 18, hoặc thẻ ba nhiệm 21.

Theo một số phương án thực hiện, việc chuẩn hóa số lượng các thẻ trình tự bao gồm: chuẩn hóa đối với nội dung GC của mẫu, chuẩn hóa đối với biên dạng sóng toàn phần của biến thiên của tập huấn luyện, và/hoặc chuẩn hóa đối với một hoặc nhiều thành phần thu được từ phân tích thành phần nguyên tắc.

Theo một số phương án thực hiện, trình tự quan tâm là nhiễm sắc thể người được chọn từ nhóm gồm có nhiễm sắc thể 13, nhiễm sắc thể 18, nhiễm sắc thể 21, nhiễm sắc thể X, và nhiễm sắc thể Y.

Theo một số phương án thực hiện, vùng tham chiếu là tất cả các nhiễm sắc thể khỏe mạnh, các nhiễm sắc thể khỏe mạnh không bao gồm trình tự quan tâm, ít nhất là nhiễm sắc thể bên ngoài trình tự quan tâm, và/hoặc tập con nhiễm sắc thể được chọn từ các nhiễm sắc thể khỏe mạnh. Theo một số phương án thực hiện, vùng tham chiếu bao gồm các nhiễm sắc thể khỏe mạnh mà đã được xác định là tạo ra khả năng dò tín hiệu tốt nhất cho tập mẫu huấn luyện.

Theo một số phương án thực hiện, phương pháp này còn bao gồm việc tính các giá trị của tham số kích thước đối với các bin bằng cách, đối với mỗi bin: (i) xác định giá trị của tham số kích thước từ các kích thước của các đoạn axit nucleic ngoài tế bào trong bin, và (ii) chuẩn hóa giá trị của tham số kích thước bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao. Phương pháp này còn bao gồm việc xác định tham số thống kê t dựa trên kích thước cho trình tự quan tâm nhờ sử dụng các giá trị của tham số kích thước của các bin ở trình tự quan tâm và các giá trị của tham số kích thước của các bin trong vùng tham chiếu cho trình tự quan tâm. Theo một số phương án thực hiện, tỷ số khả dĩ của (f) được tính từ tham số thống kê t và tham số thống kê t dựa trên kích thước. Theo một số phương án thực hiện, tỷ số khả dĩ của (f) được tính từ tham số thống kê t dựa trên kích thước và tỷ lệ ADN bào thai.

Theo một số phương án thực hiện, phương pháp này còn bao gồm việc so sánh tỷ số khả dĩ với tiêu chuẩn xác định bất thường bản sao để xác định biến thể số lượng bản sao ở trình tự quan tâm. Theo một số phương án thực hiện, tỷ số khả dĩ được chuyển đổi thành tỷ số khả dĩ loga trước khi so sánh với tiêu chuẩn xác định bất thường bản sao. Theo một số

phương án thực hiện, tiêu chuẩn xác định bất thường bản sao thu được bằng cách áp dụng các tiêu chuẩn khác nhau cho tập huấn luyện của mẫu huấn luyện, và chọn tiêu chuẩn mà tạo ra độ nhảy xác định và độ nhảy xác định.

Theo một số phương án thực hiện, phương pháp này còn bao gồm việc thu nhiều tỷ số khả dĩ và áp dụng nhiều tỷ số khả dĩ này vào cây quyết định để xác định trường hợp bội thể cho mẫu.

Theo một số phương án thực hiện, phương pháp còn bao gồm thu được nhiều tỷ số khả dĩ và một hoặc nhiều giá trị độ che phủ của trình tự quan tâm, và áp dụng nhiều tỷ số khả dĩ này và một hoặc nhiều giá trị độ che phủ của trình tự quan tâm vào cây quyết định để xác định trường hợp bội thể cho mẫu.

Một khía cạnh khác của sáng chế đề xuất phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của trình tự axit nucleic quan tâm trong mẫu thử nghiệm bao gồm các đoạn axit nucleic ngoài tế bào có nguồn gốc từ hai hoặc nhiều bộ gen. Phương pháp bao gồm các bước: (a) nhận các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm; (b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu bao gồm trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin; (c) tính độ che phủ của các thẻ trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin: (i) xác định số lượng các thẻ trình tự sắp thẳng với bin, và (ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; Phương pháp còn bao gồm các bước: (d) xác định tham số thống kê t cho trình tự quan tâm nhờ sử dụng độ che phủ của các bin ở trình tự quan tâm và độ che phủ của các bin trong vùng tham chiếu đối với trình tự quan tâm; (e) ước tính một hoặc nhiều trị số của tỷ lệ ADN bào thai của các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm; và (f) xác định biến thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng tham số thống kê t và một hoặc nhiều trị số của tỷ lệ ADN bào thai.

Theo một số phương án thực hiện, (f) bao gồm việc tính tỷ số khả dĩ từ tham số thống kê t và một hoặc nhiều trị số của tỷ lệ ADN bào thai. Theo một số phương án thực hiện, tỷ số

khả dĩ được tính cho thẻ đơn nhiễm X, thẻ ba nhiễm X, thẻ ba nhiễm 13, thẻ ba nhiễm 18, hoặc thẻ ba nhiễm 21.

Theo một số phương án thực hiện, việc chuẩn hóa số lượng các thẻ trình tự bao gồm: chuẩn hóa đối với nội dung GC của mẫu, chuẩn hóa đối với biên dạng sóng toàn phần của biến thiên của tập huấn luyện, và/hoặc chuẩn hóa đối với một hoặc nhiều thành phần thu được từ phân tích thành phần nguyên tắc.

Theo một số phương án thực hiện, trình tự quan tâm là nhiễm sắc thể người được chọn từ nhóm gồm có nhiễm sắc thể 13, nhiễm sắc thể 18, nhiễm sắc thể 21, nhiễm sắc thể X, và nhiễm sắc thể Y.

Một khía cạnh nữa của sáng chế đề xuất phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của trình tự axit nucleic quan tâm trong mẫu thử nghiệm bao gồm các đoạn axit nucleic ngoài tế bào có nguồn gốc từ hai hoặc nhiều bộ gen. Phương pháp bao gồm các bước: (a) nhận các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm; (b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu bao gồm trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin; (c) xác định các kích thước đoạn của các đoạn axit nucleic ngoài tế bào tồn tại trong mẫu thử nghiệm; (d) tính độ che phủ của các thẻ trình tự đối với các bin của bộ gen tham chiếu nhờ sử dụng các thẻ trình tự cho các đoạn axit nucleic ngoài tế bào có các kích thước ở domain kích thước thứ nhất; (e) tính độ che phủ của các thẻ trình tự đối với các bin của bộ gen tham chiếu nhờ sử dụng các thẻ trình tự cho các đoạn axit nucleic ngoài tế bào có các kích thước ở domain kích thước thứ hai, trong đó domain kích thước thứ hai khác với domain kích thước thứ nhất; (f) tính các đặc trưng kích thước đối với các bin của bộ gen tham chiếu nhờ sử dụng các kích thước đoạn xác định được ở (c); và (g) xác định biến thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (d) và (e) và các đặc trưng kích thước tính được ở (f).

Theo một số phương án thực hiện, domain kích thước thứ nhất bao gồm các đoạn axit nucleic ngoài tế bào có gần như tất cả các kích thước trong mẫu, và domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn kích thước xác định. Theo một số

phương án thực hiện, domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn khoảng 150 bp.

Theo một số phương án thực hiện, trình tự quan tâm là nhiễm sắc thể người được chọn từ nhóm gồm có nhiễm sắc thể 13, nhiễm sắc thể 18, nhiễm sắc thể 21, nhiễm sắc thể X, và nhiễm sắc thể Y.

Theo một số phương án thực hiện, (g) bao gồm việc tính tham số thống kê t cho trình tự quan tâm nhờ sử dụng độ che phủ của các bin ở trình tự quan tâm tính được ở (d) và/hoặc (e). Theo một số phương án thực hiện, trong đó việc tính tham số thống kê t cho trình tự quan tâm bao gồm việc sử dụng độ che phủ của các bin ở trình tự quan tâm và độ che phủ của các bin trong vùng tham chiếu đối với trình tự quan tâm.

Theo một số phương án thực hiện, (g) bao gồm việc tính tham số thống kê t cho trình tự quan tâm nhờ sử dụng các đặc trưng kích thước của các bin ở trình tự quan tâm tính được ở (f). Theo một số phương án thực hiện, việc tính tham số thống kê t cho trình tự quan tâm bao gồm việc sử dụng các đặc trưng kích thước của các bin ở trình tự quan tâm và các đặc trưng kích thước của các bin trong vùng tham chiếu đối với trình tự quan tâm.

Theo một số phương án thực hiện, đặc trưng kích thước cho bin bao gồm tỷ số của các đoạn của size nhỏ hơn giá trị xác định với tổng các đoạn trong bin.

Theo một số phương án thực hiện, (g) bao gồm việc tính tỷ số khả dĩ từ tham số thống kê t .

Theo một số phương án thực hiện, (g) bao gồm việc tính tỷ số khả dĩ từ tham số thống kê t thứ nhất cho trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (d), và tham số thống kê t thứ hai cho trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (e).

Theo một số phương án thực hiện, (g) bao gồm việc tính tỷ số khả dĩ từ tham số thống kê t thứ nhất cho trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (d), tham số thống kê t thứ hai cho trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (e), và tham số thống kê t thứ ba cho trình tự quan tâm nhờ sử dụng các đặc trưng kích thước tính được ở (f).

Theo một số phương án thực hiện, tỷ số khả dĩ được tính từ một hoặc nhiều giá trị của tỷ lệ ADN bào thai ngoài ít nhất là tham số thống kê t thứ nhất và thứ hai. Theo một số phương án thực hiện, phương pháp này còn bao gồm việc tính một hoặc nhiều giá trị của tỷ lệ ADN bào thai nhờ sử dụng thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.

Theo một số phương án thực hiện, phương pháp này còn bao gồm việc tính một hoặc nhiều giá trị của tỷ lệ ADN bào thai nhờ sử dụng thông tin độ che phủ đối với các bin của bộ gen tham chiếu. Theo một số phương án thực hiện, một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin độ che phủ đối với các bin thuộc nhiễm sắc thể giới tính. Theo một số phương án thực hiện, tỷ số khả dĩ được tính cho thể đơn nhiễm X, thể ba nhiễm X, thể ba nhiễm 13, thể ba nhiễm 18, hoặc thể ba nhiễm 21.

Theo một số phương án thực hiện, (d) và/hoặc (e) bao gồm: (i) xác định số lượng các thẻ trình tự sắp thẳng với bin, và (ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; Theo một số phương án thực hiện, việc chuẩn hóa số lượng các thẻ trình tự bao gồm: chuẩn hóa đối với nội dung GC của mẫu, chuẩn hóa đối với biên dạng sóng toàn phần của biến thiên của tập huấn luyện, và/hoặc chuẩn hóa đối với một hoặc nhiều thành phần thu được từ phân tích thành phần nguyên tắc.

Theo một số phương án thực hiện, (f) bao gồm việc tính các giá trị của tham số kích thước đối với các bin bằng cách, đối với mỗi bin: (i) xác định giá trị của tham số kích thước từ các kích thước của các đoạn axit nucleic ngoài tế bào trong bin, và (ii) chuẩn hóa giá trị của tham số kích thước bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao.

Một khía cạnh khác của sáng chế đề cập đến hệ thống để ước lượng số lượng bản sao của trình tự axit nucleic quan tâm trong mẫu thử nghiệm, hệ thống bao gồm: bộ giải trình tự để nhận các đoạn axit nucleic từ mẫu thử nghiệm và đưa ra các thông tin trình tự axit nucleic của của mẫu thử nghiệm; bộ xử lý; và một hoặc nhiều vật ghi đọc được bằng máy tính có có lưu trữ các lệnh để thực thi trên bộ xử lý này. Các lệnh này bao gồm lệnh để: (a) nhận các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm; (b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu bao gồm trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin; (c) xác định các kích thước đoạn của ít nhất một số trong số các đoạn axit nucleic ngoài tế bào có mặt trong mẫu thử nghiệm; (d) tính độ che phủ của các thẻ trình tự đối với các bin

của bộ gen tham chiếu bằng cách, đối với mỗi bin: (i) xác định số lượng các thẻ trình tự sắp thẳng với bin, và (ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; Phương pháp còn bao gồm các bước: (e) xác định tham số thống kê t cho trình tự quan tâm nhờ sử dụng độ che phủ của các bin ở trình tự quan tâm và độ che phủ của các bin trong vùng tham chiếu đối với trình tự quan tâm; và (f) xác định biến thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng tỷ số khả dĩ tính được từ tham số thống kê t và thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.

Theo một số phương án thực hiện, hệ thống được tạo cấu hình để thực hiện phương pháp bất kỳ trong số các phương pháp nêu trên.

Một khía cạnh nữa của sáng chế đề cập đến sản phẩm chương trình máy tính bao gồm một hoặc nhiều vật ghi bất biến đọc được bằng máy tính có lưu trữ trong đó các lệnh thực thi bởi máy tính mà, khi được thực thi bởi một hoặc nhiều bộ xử lý của hệ thống máy tính, khiến cho hệ thống máy tính thực hiện phương pháp bất kỳ trong số các phương pháp nêu trên.

Mặc dù các ví dụ ở đây liên quan đến người và ngôn ngữ chủ yếu hướng đến mối quan tâm là con người, nhưng các khái niệm được mô tả ở đây có thể áp dụng được cho các hệ gen từ thực vật hoặc động vật bất kỳ. Dưới đây, các đối tượng và dấu hiệu này và khác của sáng chế sẽ trở nên rõ ràng hơn từ phần mô tả sau và các yêu cầu bảo hộ đính kèm, hoặc sẽ hiểu được từ việc thực hiện sáng chế đã nêu.

Kết hợp bằng cách viện dẫn

Tất cả các sáng chế, đơn yêu cầu cấp bằng sáng chế, và các công bố khác, bao gồm tất cả các trình tự được bộc lộ trong các tài liệu này, được nói đến ở đây được đưa vào đây bằng cách viện dẫn, với cùng phạm vi như thể từng công bố, bằng sáng chế hoặc đơn sáng chế được chỉ ra là được đưa vào bằng cách viện dẫn. Tất cả các tài liệu được viện dẫn, ở phần thích hợp, được đưa vào đây bằng cách viện dẫn toàn bộ dùng cho mục đích được chỉ ra bởi ngữ cảnh viện dẫn tài liệu đó trong bản mô tả này. Tuy nhiên, việc viện dẫn tài liệu bất kỳ không có nghĩa là thừa nhận rằng tài liệu đó là tình trạng kỹ thuật đã biết đối với sáng chế.

Mô tả vắn tắt các hình vẽ

Fig. 1 là lưu đồ của phương pháp 100 để xác định sự có mặt hoặc vắng mặt của biến thể số lượng bản sao trong mẫu thử nghiệm chứa hỗn hợp các axit nucleic.

Fig.2A thể hiện một cách có chủ đề về việc sử dụng phương pháp giải trình tự đầu bắt cặp như thế nào để xác định cả kích thước đoạn và độ che phủ trình tự.

Fig.2B thể hiện lưu đồ của quy trình sử dụng độ che phủ dựa trên kích thước để xác định biến thể số lượng bản sao của trình tự axit nucleic quan tâm trong mẫu thử nghiệm.

Fig.2C thể hiện lưu đồ của quy trình xác định tham số kích thước đoạn cho trình tự axit nucleic quan tâm được sử dụng để ước lượng số lượng bản sao.

Fig.2D thể hiện lưu đồ của hai nhánh chồng lấn của tiến trình.

Fig.2E thể hiện lưu đồ về quy trình ba nhánh để ước lượng số lượng bản sao.

Fig.2F thể hiện các phương án thực hiện mà áp dụng tham số thống kê t cho phân tích số lượng bản sao để cải thiện độ chính xác của phân tích.

Fig.2G thể hiện quy trình ví dụ để xác định tỷ lệ ADN bào thai từ thông tin độ che phủ theo một số phương án thực hiện của sáng chế.

Fig.2H thể hiện quy trình xác định tỷ lệ ADN bào thai từ thông tin phân bố kích thước theo một số phương án thực hiện.

Fig.2I thể hiện quy trình ví dụ để xác định tỷ lệ ADN bào thai từ thông tin tần suất 8-mer theo một số phương án thực hiện của sáng chế.

Fig.2J thể hiện tiến trình xử lý các trình tự đọc mà thông tin của chúng có thể được sử dụng để thu được ước tính tỷ lệ ADN bào thai.

Fig.3A thể hiện lưu đồ về ví dụ về quy trình để giảm nhiễu trong dữ liệu trình tự từ mẫu thử nghiệm.

Các hình vẽ từ Fig.3B đến Fig.3K thể hiện các phân tích dữ liệu thu được ở các giai đoạn khác nhau của quy trình được minh họa trên Fig.3A.

Fig.4A thể hiện lưu đồ của quy trình tạo mặt nạ trình tự để làm giảm nhiễu trong dữ liệu trình tự.

Fig.4B thể hiện rằng điểm số MapQ có mối tương quan đơn điệu mạnh với CV có lượng che phủ được chuẩn hóa.

Fig.5 là sơ đồ khối của hệ thống phân tán để xử lý mẫu thử nghiệm và sau cùng là thực hiện phép chẩn đoán.

Fig.6 là sơ đồ minh họa cách mà các công đoạn khác nhau trong xử lý các mẫu thử nghiệm có thể được nhóm lại để được xử lý bởi các bộ phận khác nhau của hệ thống.

Các hình vẽ từ Fig.7A và 7B thể hiện điện di đồ của thư viện giải trình tự cfADN được chuẩn bị theo phương thức tóm tắt mô tả trong Ví dụ 1a (Fig.7A), và phương thức được mô tả trong Ví dụ 1b (Fig.7B).

Fig.8 thể hiện tiến trình tổng quát và đồng thời gian cho phiên bản mới của NIPT so với tiến trình công việc trong phòng thí nghiệm chuẩn.

Fig.9 thể hiện hiệu suất thư viện giải trình tự dưới dạng hàm của cfADN tách chiết đầu vào, chỉ rõ mối tương quan tuyến tính mạnh với sự tập trung của thư viện với sự tập trung đầu vào với hiệu suất chuyển hóa cao.

Fig.10 thể hiện sự phân bố kích thước đoạn cfADN đo được từ 324 mẫu từ các phụ nữ mang thai bé trai.

Fig.11 thể hiện tỷ lệ ADN bào thai liên quan từ tổng số đếm của các trình tự đọc có đầu bắt cặp được ánh xạ so với số đếm từ các trình tự đọc có đầu bắt cặp mà nhỏ hơn 150 bp.

Fig.12 thể hiện điểm số lệch bội thống kê t kết hợp cho phát hiện các mẫu thể ba nhiễm 21 cho (A) tổng số đếm của tất cả các đoạn; (B) tổng số đếm của chỉ các đoạn ngắn (<150bp); (C) phần của các đoạn ngắn (tổng số đếm giữa 80 và 150 bp/tổng số đếm <250bp); (D) tham số thống kê t kết hợp từ (B) và (C); và (E) kết quả cho cùng mẫu thu được nhờ sử dụng quy trình phòng thí nghiệm Illumina Redwood City CLIA với trung bình là 16 M số đếm/mẫu.

Fig.13 thể hiện các tỷ lệ ADN bào thai ước tính được từ các bin được chọn so với các phần đo được bằng các giá trị nhiễm sắc thể chuẩn hóa (normalized chromosome value - REF) cho X-nhiễm sắc thể. Tập 1 được dùng để hiệu chuẩn giá trị tỷ lệ ADN bào thai và tập 2 độc lập để kiểm định sự tương quan.

Mô tả chi tiết sáng chế

Các định nghĩa

Trừ khi có quy định khác, việc thực hiện phương pháp và hệ thống được bộc lộ trong bản mô tả bao gồm các kỹ thuật thông thường và các thiết bị thường được sử dụng trong lĩnh vực sinh học phân tử, vi sinh vật học, tinh chế protein, thiết kế protein, việc giải trình tự protein và ADN và lĩnh vực ADN tái tổ hợp, đều nằm trong hiểu biết chuyên môn trong lĩnh vực này. Các kỹ thuật và thiết bị này là đã biết đối với người có hiểu biết trung bình trong

lĩnh vực kỹ thuật này và được mô tả trong nhiều văn bản và tài liệu tham khảo (xem, ví dụ, Sambrook et al., "Molecular Cloning: A Laboratory Manual," tái bản lần 3 (Cold Spring Harbor), [2001]); và Ausubel et al., "Current Protocols in Molecular Biology" [1987]).

Khoảng số là bao gồm cả các số định rõ khoảng này. Được dự tính là mọi giới hạn số lớn nhất được đưa ra trong toàn bộ bản mô tả này bao gồm mọi giới hạn số thấp hơn, như thể các giới hạn số thấp hơn này được viết rõ ràng trong bản mô tả. Mọi giới hạn số nhỏ nhất được đưa ra trong toàn bộ bản mô tả này sẽ bao gồm mọi giới hạn số cao hơn, như thể là các giới hạn số cao hơn này được viết rõ ràng trong bản mô tả. Mọi khoảng số được đưa ra trong toàn bộ bản mô tả này sẽ bao gồm mọi khoảng số hẹp hơn mà nằm trong khoảng số rộng hơn này, như thể các khoảng số hẹp hơn này được viết rõ ràng toàn bộ trong bản mô tả.

Các tiêu đề được đưa ra trong bản mô tả không nhằm để giới hạn phần mô tả.

Trừ khi được quy định khác, tất cả các thuật ngữ kỹ thuật và khoa học được sử dụng trong bản mô tả này có nghĩa giống như được hiểu thông thường bởi người có hiểu biết trung bình trong lĩnh vực này. Các từ điển khoa học khác nhau mà chứa các thuật ngữ được bao gồm trong bản mô tả này là được biết rõ và khả dụng đối với người có hiểu biết trung bình trong lĩnh vực. Mặc dù phương pháp và vật liệu bất kỳ tương tự hoặc tương đương với phương pháp và vật liệu được mô tả trong bản mô tả này được sử dụng trong việc thực hiện hoặc thử nghiệm các phương án được bộc lộ trong bản mô tả, một số phương pháp và vật liệu được mô tả dưới đây.

Các thuật ngữ được định nghĩa ngay dưới đây được mô tả đầy đủ hơn bằng cách tham chiếu đến toàn bộ bản mô tả. Cần phải hiểu rằng bản mô tả này không bị giới hạn bởi phương pháp luận cụ thể, các quy trình và chất phản ứng được mô tả, vì chúng có thể khác nhau, tùy thuộc vào ngữ cảnh chúng được sử dụng bởi những người có hiểu biết trung bình trong lĩnh vực. Như được sử dụng ở đây, các thuật ngữ dạng số ít "một", "một cái", và "cái này" bao gồm cả dạng số nhiều trừ khi ngữ cảnh quy định ngược lại một cách rõ ràng.

Trừ khi có quy định khác, axit nucleic được viết lần lượt từ trái sang phải theo chiều 5' đến 3' và trình tự axit amin được viết từ trái sang phải theo chiều từ amino đến carboxy.

Thuật ngữ "tham số" được sử dụng ở đây biểu thị dấu hiệu vật lý mà trị số hoặc đặc tính khác của nó có tác động đến tình trạng bệnh lý liên quan chẳng hạn sự biến thể số lượng bản sao. Trong một số trường hợp, thuật ngữ tham số được sử dụng với việc tham chiếu đến một biến số mà có ảnh hưởng đến kết quả của một hệ thức hoặc mô hình toán học, biến số này có thể là biến số độc lập (tức là, đầu vào cho mô hình) hoặc biến trung gian dựa vào một hoặc nhiều biến độc lập. Tùy thuộc vào phạm vi của mô hình, đầu ra của một mô hình có thể trở thành đầu vào của mô hình khác, nhờ đó trở thành tham số của mô hình khác.

Thuật ngữ "tham số kích thước đoạn" chỉ tham số mà có liên quan đến kích thước hoặc chiều dài của một đoạn hoặc một tập hợp gồm nhiều đoạn như các đoạn axit nucleic, ví dụ, các đoạn cfADN thu được từ dịch cơ thể. Như được sử dụng ở đây, một tham số "lệch về phía kích thước hoặc khoảng kích thước của đoạn" nếu: 1) tham số này được gán trọng số thuận đối với kích thước hoặc khoảng kích thước của đoạn, ví dụ, tổng số được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này so với kích thước hoặc khoảng kích thước khác; hoặc 2) tham số thu được từ giá trị mà được gán trọng số thuận đối với kích thước hoặc khoảng kích thước này của đoạn, ví dụ, tỷ số thu được từ số đếm được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này. Kích thước hoặc khoảng kích thước của đoạn có thể là một đặc trưng của bộ gen hoặc một phần của nó khi bộ gen tạo ra các đoạn axit nucleic được làm giàu hoặc có sự tập trung kích thước hoặc khoảng kích thước cao hơn so với đoạn axit nucleic từ bộ gen khác hoặc phần khác của cùng một bộ gen.

Thuật ngữ "gán trọng số" chỉ việc thay đổi ở lượng như tham số hoặc biến số nhờ sử dụng một hoặc nhiều giá trị hoặc hàm số, được xem là "trọng số." Theo các phương án nhất định, tham số hoặc biến số được nhân với trọng số. Theo các phương án khác, tham số hoặc biến số thay đổi theo hàm số mũ. Theo một số phương án, hàm số có thể là hàm tuyến tính hoặc hàm không tuyến tính. Ví dụ về các hàm không tuyến tính có thể áp dụng được bao gồm, nhưng không chỉ giới hạn ở hàm bước Heaviside, hàm hộp xe, hàm bậc thang, hoặc hàm sigma. Việc gán trọng số một tham số hoặc biến số gốc có thể làm tăng hoặc giảm một cách có hệ thống giá trị của biến số được gán trọng số. Theo các phương án khác nhau, việc gán trọng số có thể dẫn đến các giá trị dương, không âm hoặc âm.

Thuật ngữ "biến thể số lượng bản sao" ở đây chỉ sự biến đổi về số lượng bản sao của trình tự axit nucleic có mặt trong mẫu thử nghiệm so với số lượng bản sao của trình tự axit nucleic có mặt trong mẫu đối chứng. Theo các phương án nhất định, trình tự axit nucleic là 1 kb hoặc lớn hơn. Trong một số trường hợp, trình tự axit nucleic là toàn bộ nhiễm sắc thể hoặc một phần đáng kể của nó. "Biến thể số lượng bản sao" chỉ trình tự axit nucleic mà trong đó tìm thấy sự khác nhau về số lượng bản sao so với trình tự axit nucleic đích trong mẫu thử nghiệm có mức mong muốn trình tự axit nucleic. Ví dụ, mức trình tự axit nucleic đích trong mẫu thử nghiệm so với mức có mặt trong mẫu đạt yêu cầu. Biến thể/biến đổi số lượng bản sao bao gồm mất đoạn, bao gồm mất đoạn nhỏ, chèn đoạn, bao gồm chèn đoạn nhỏ, lặp đoạn, lặp nhiều đoạn và chuyển đoạn. CNV bao gồm các thể lệch bội nhiễm sắc thể và thể lệch bội một phần.

Thuật ngữ "thể lệch bội" ở đây chỉ sự mất cân bằng của vật liệu di truyền gây ra bởi sự mất hoặc tăng thêm toàn bộ nhiễm sắc thể hoặc một phần của nhiễm sắc thể.

Thuật ngữ "thể lệch bội nhiễm sắc thể" và "thể lệch bội nhiễm sắc thể toàn phần" ở đây chỉ sự mất cân bằng của vật liệu di truyền gây ra bởi sự mất hoặc tăng thêm toàn bộ nhiễm sắc thể, và bao gồm thể lệch bội dòng mầm và thể lệch bội khảm.

Thuật ngữ "thể lệch bội một phần" và "thể lệch bội nhiễm sắc thể một phần" ở đây chỉ sự mất cân bằng của vật liệu di truyền gây ra bởi sự mất hoặc tăng thêm một phần của nhiễm sắc thể, ví dụ, thể một nhiễm sắc thể một phần và thể ba nhiễm sắc thể một phần, và bao gồm sự mất cân bằng do chuyển đoạn, mất đoạn và chèn đoạn.

Thuật ngữ "nhiều" chỉ nhiều hơn một yếu tố. Ví dụ, thuật ngữ được sử dụng ở đây khi đề cập đến số lượng phân tử axit nucleic hoặc thể trình tự mà đủ để nhận dạng sự sai khác đáng kể về biến thể số lượng bản sao trong mẫu thử nghiệm và mẫu đạt yêu cầu bằng cách sử dụng các phương pháp được bộc lộ ở đây. Theo một số phương án, thu được ít nhất khoảng 3×10^6 thể trình tự nằm trong từ khoảng 20 đến 40bp cho mỗi mẫu thử nghiệm. Theo một số phương án, mỗi mẫu thử nghiệm đưa ra dữ liệu cho ít nhất khoảng 5×10^6 , 8×10^6 , 10×10^6 , 15×10^6 , 20×10^6 , 30×10^6 , 40×10^6 hoặc 50×10^6 thể trình tự, mỗi thể trình tự bao gồm từ khoảng 20 đến 40bp.

Thuật ngữ "trình tự đọc đầu bắt cặp" chỉ các trình tự đọc từ việc giải trình tự đầu bắt cặp mà thu được một trình tự đọc từ mỗi đầu cuối của đoạn axit nucleic. Giải trình tự đầu bắt cặp có thể ba nhiễmo gồm việc phân đoạn các sợi polynucleotit thành các trình tự ngắn được gọi là các đoạn chèn. Việc phân đoạn là tùy ý và không cần thiết đối với polynucleotit tương đối ngắn như các phân tử ADN ngoài tế bào.

Thuật ngữ "polynucleotit," "axit nucleic" và "phân tử axit nucleic" được sử dụng thay thế cho nhau và chỉ trình tự nucleotit được liên kết cộng hóa trị (tức là, ribonucleotit đối với ARN và deoxyribonucleotit đối với ADN) trong đó vị trí 3' của đường pentoza của một nucleotit được nối bằng nhóm phosphodiester với vị trí 5' của đường pentoza của nucleotit kế tiếp. Nucleotit bao gồm các trình tự axit nucleic ở dạng bất kỳ, bao gồm, nhưng không chỉ giới hạn ở các phân tử ARN và ADN như phân tử cfADN. Thuật ngữ "polynucleotit" bao gồm, không giới hạn, polynucleotit sợi đơn và sợi đôi.

Thuật ngữ "mẫu thử nghiệm" ở đây chỉ mẫu, thường có nguồn gốc từ dịch sinh học, tế bào, mô, cơ quan hoặc cơ thể, chứa axit nucleic hoặc hỗn hợp của axit nucleic gồm ít nhất một trình tự axit nucleic cần được sàng lọc về biến thể số lượng bản sao. Theo các phương án nhất định, mẫu chứa ít nhất một trình tự axit nucleic mà số lượng bản sao của nó bị nghi là đã bị biến đổi. Các mẫu như vậy bao gồm, nhưng không chỉ giới hạn ở đờm/dịch khoang miệng, dịch màng ối, máu, phân đoạn máu, hoặc mẫu sinh thiết kim nhỏ (ví dụ, sinh thiết giải phẫu, sinh thiết kim nhỏ, v.v.), nước tiểu, dịch màng bụng, dịch màng phổi, và mẫu tương tự. Mặc dù mẫu thường được lấy từ đối tượng người (ví dụ, người bệnh), các thử nghiệm có thể được sử dụng đối với biến thể số lượng bản sao (CNV) ở mẫu từ động vật có vú bất kỳ, bao gồm, nhưng không chỉ giới hạn ở chó, mèo, ngựa, dê, cừu, gia súc, lợn, v.v.. Mẫu có thể được sử dụng trực tiếp dưới dạng thu được từ nguồn sinh học hoặc sau khi xử lý sơ bộ để làm thay đổi đặc tính của mẫu. Ví dụ, quá trình xử lý sơ bộ này có thể bao gồm tách huyết tương từ máu, pha loãng dịch nhầy và v.v.. Phương pháp xử lý sơ bộ có thể còn bao gồm, nhưng không chỉ giới hạn ở, lọc, kết tủa, pha loãng, chưng cất, trộn, ly tâm, đông lạnh, khô lạnh, cô, khuếch đại, phân đoạn axit nucleic, làm bất hoạt các thành phần gây nhiễu, bổ sung các chất phản ứng, ly giải, v.v.. Nếu các phương pháp xử lý sơ bộ này được dùng cho mẫu, các phương pháp xử lý sơ bộ này thường sao cho axit nucleic đích vẫn còn trong mẫu thử nghiệm, đôi khi ở

nồng độ tỷ lệ với nồng độ trong mẫu thử nghiệm không được xử lý (ví dụ, cụ thể là, mẫu không được đưa vào bất kỳ phương pháp xử lý sơ bộ như vậy). Các mẫu "đã được xử lý" hoặc "đã được chế biến" vẫn được xem là các mẫu "thử nghiệm" sinh học đối với các phương pháp được mô tả ở đây.

Thuật ngữ "mẫu đạt yêu cầu" hoặc "mẫu không bị bệnh" ở đây chỉ mẫu chứa hỗn hợp axit nucleic có mặt với số lượng bản sao đã biết mà axit nucleic trong mẫu thử nghiệm được so sánh với nó, và đó là mẫu bình thường, tức là không phải thể lệch bội, đối với trình tự axit nucleic đích. Theo một số phương án, mẫu đạt yêu cầu được dùng làm mẫu huấn luyện không bị bệnh của tập huấn luyện để suy ra mặt nạ trình tự hoặc biên dạng trình tự. Theo các phương án nhất định, mẫu đạt yêu cầu được dùng để nhận dạng một hoặc nhiều nhiễm sắc thể chuẩn hóa hoặc các đoạn đối với nhiễm sắc thể cần xem xét. Ví dụ, mẫu đạt yêu cầu có thể được dùng để nhận dạng hoặc chuẩn hóa nhiễm sắc thể đối với nhiễm sắc thể 21. Trong trường hợp này, mẫu đạt yêu cầu là mẫu không phải mẫu thể ba nhiễm sắc thể 21. Ví dụ khác bao gồm sử dụng chỉ nữ giới làm mẫu đạt yêu cầu đối với nhiễm sắc thể X. Mẫu đạt yêu cầu có thể còn được dùng cho các mục đích khác như xác định ngưỡng để nhận dạng mẫu bị bệnh, xác định ngưỡng để định rõ vùng che chắn trên trình tự tham chiếu, xác định lượng che phủ mong muốn đối với các vùng khác nhau của bộ gen, và tương tự.

Thuật ngữ "tập huấn luyện" ở đây chỉ tập hợp các mẫu huấn luyện mà có thể ba nhiễm gồm các mẫu bị bệnh và/hoặc các mẫu không bị bệnh và được dùng để phát triển mô hình phân tích mẫu thử nghiệm. Theo một số phương án, tập huấn luyện bao gồm các mẫu không bị bệnh. Theo các phương án này, ngưỡng để xác định CNV được thiết lập bằng cách sử dụng các tập hợp mẫu huấn luyện mà không bị bệnh đối với biến thể số lượng bản sao đích. Các mẫu không bị bệnh trong tập huấn luyện có thể được dùng làm mẫu đạt yêu cầu để nhận dạng các trình tự chuẩn hóa, ví dụ, các nhiễm sắc thể chuẩn hóa, và định lượng nhiễm sắc thể của các mẫu không bị bệnh được dùng để thiết lập ngưỡng đối với mỗi trình tự, ví dụ, các nhiễm sắc thể quan tâm. Theo một số phương án, tập huấn luyện bao gồm các mẫu bị bệnh. Các mẫu bị bệnh trong tập huấn luyện có thể được sử dụng để kiểm chứng mẫu thử nghiệm bị bệnh có thể được phân biệt một cách dễ dàng với các mẫu không bị bệnh.

Tập huấn luyện còn là mẫu thống kê trong quần thể quan tâm, mẫu thống kê này không được lẫn với mẫu sinh học. Mẫu thống kê thường bao gồm nhiều cá thể, dữ liệu về các cá thể này được dùng để xác định một hoặc nhiều giá trị định lượng đích có thể khái quát hóa cho quần thể. Mẫu thống kê là tập hợp con gồm các cá thể trong quần thể quan tâm. Các cá thể có thể là người, động vật, mô, tế bào, các mẫu sinh học khác (tức là, mẫu thống kê có thể bao gồm nhiều mẫu sinh học), và các thực thể riêng lẻ khác cung cấp các điểm dữ liệu để phân tích thống kê.

Thông thường, tập huấn luyện được dùng kết hợp với tập hợp xác thực. Thuật ngữ "tập hợp xác thực được dùng để chỉ tập hợp gồm các cá thể trong mẫu thống kê, dữ liệu của các cá thể này được dùng để xác thực hoặc đánh giá các giá trị định lượng đích được xác định nhờ sử dụng tập huấn luyện. Theo một số phương án, chẳng hạn, tập huấn luyện cung cấp dữ liệu để tính toán mặt nạ cho trình tự tham chiếu, trong khi tập hợp xác thực cung cấp dữ liệu để đánh giá tính xác thực hoặc tính hiệu quả của mặt nạ.

"Ước lượng số lượng bản sao" được sử dụng ở đây khi đề cập đến ước lượng thống kê tình trạng của trình tự gen có liên quan đến số lượng bản sao của trình tự này. Ví dụ, theo một số phương án, việc đánh giá bao gồm việc xác định sự có mặt hoặc không có mặt của một trình tự gen. Theo một số phương án, việc đánh giá bao gồm việc xác định thể lệch bội một phần hay toàn phần của trình tự gen. Theo các phương án khác, việc đánh giá bao gồm việc phân biệt giữa hai hoặc nhiều mẫu dựa vào số lượng bản sao của trình tự gen. Theo một số phương án, việc đánh giá bao gồm các phân tích thống kê, ví dụ, chuẩn hóa và so sánh, dựa vào số lượng bản sao của trình tự gen.

Thuật ngữ "axit nucleic định tính" được dùng thay thế cho nhau với "trình tự định tính," là trình tự mà số lượng trình tự hoặc axit nucleic đích được so sánh với nó. Trình tự định tính là trình tự có mặt trong mẫu sinh học, tốt hơn là với mức biểu thị đã biết, tức là, lượng trình tự định tính là đã biết. Thông thường, trình tự định tính là trình tự có mặt trong "mẫu đạt yêu cầu." "Trình tự định tính quan tâm" là trình tự định tính mà lượng trình tự này đã biết trong mẫu đạt yêu cầu, và là trình tự có liên quan đến sự sai khác của trình tự quan tâm giữa đối tượng đối chứng và cá thể mắc tình trạng bệnh lý.

Thuật ngữ "trình tự quan tâm" hoặc "trình tự axit nucleic quan tâm" ở đây chỉ trình tự axit nucleic mà có liên quan đến sự sai khác về sự biểu hiện trình tự giữa cá thể khỏe mạnh và cá thể bị bệnh. Trình tự quan tâm có thể là trình tự trên nhiễm sắc thể mà được biểu hiện sai, tức là, được biểu hiện quá mức hoặc dưới mức, trong tình trạng di truyền hoặc bệnh lý. Trình tự quan tâm có thể là một phần của nhiễm sắc thể, tức là, đoạn nhiễm sắc thể, hoặc toàn bộ nhiễm sắc thể. Ví dụ, trình tự quan tâm có thể là nhiễm sắc thể mà được biểu hiện quá mức ở tình trạng thể lệch bội, hoặc gen mã hóa chất chống u mà được biểu hiện dưới mức ở bệnh ung thư. Các trình tự quan tâm bao gồm các trình tự mà được biểu hiện quá mức hoặc dưới mức trong tổng quần thể, hoặc quần thể con các tế bào của đối tượng. "Trình tự định tính đích" là trình tự quan tâm trong mẫu đạt yêu cầu. "Trình tự quan tâm thử nghiệm" là trình tự quan tâm trong mẫu thử nghiệm.

Thuật ngữ "trình tự chuẩn hóa" ở đây chỉ trình tự mà được dùng để chuẩn hóa số lượng thể trình tự được ánh xạ đến trình tự quan tâm kết hợp với trình tự chuẩn hóa. Theo một số phương án, trình tự chuẩn hóa gồm nhiễm sắc thể khỏe mạnh. "Nhiễm sắc thể khỏe mạnh" là nhiễm sắc thể không bị lệch bội lẻ. Trong một số trường hợp bao gồm nhiễm sắc thể của người, nhiễm sắc thể khỏe mạnh là nhiễm sắc thể bất kỳ không phải nhiễm sắc thể X, nhiễm sắc thể Y, nhiễm sắc thể 13, nhiễm sắc thể 18, và nhiễm sắc thể 21. Theo một số phương án, trình tự chuẩn hóa biểu thị độ biến thiên về số lượng thể trình tự mà được ánh xạ đến nó giữa các mẫu và các lần chạy giải trình tự mà gần đúng với độ biến thiên của trình tự quan tâm mà nó được dùng làm tham số chuẩn hóa. Trình tự chuẩn hóa có thể phân biệt mẫu bị bệnh với một hoặc nhiều mẫu không bị bệnh. Theo một số cách thực hiện, trình tự chuẩn hóa phân biệt tốt nhất hoặc có hiệu quả, khi so sánh với các trình tự chuẩn hóa tiềm năng khác như nhiễm sắc thể khác, mẫu bị bệnh với một hoặc nhiều mẫu không bị bệnh. Theo một số phương án, độ biến thiên của trình tự chuẩn hóa được tính là độ biến thiên về định lượng nhiễm sắc thể đối với trình tự quan tâm giữa các mẫu và các lần chạy giải trình tự. Theo một số phương án, trình tự chuẩn hóa được xác định trong tập hợp mẫu không bị bệnh.

"Nhiễm sắc thể chuẩn hóa," "nhiễm sắc thể mẫu thức chuẩn hóa," hoặc "trình tự nhiễm sắc thể chuẩn hóa" là ví dụ về "trình tự chuẩn hóa." "Trình tự nhiễm sắc thể chuẩn hóa" có thể gồm có một nhiễm sắc thể hoặc gồm một nhóm các nhiễm sắc thể. Theo một số phương

án, trình tự chuẩn hóa gồm hai hoặc nhiều nhiễm sắc thể khỏe mạnh. Theo các phương án nhất định, các nhiễm sắc thể khỏe mạnh đều là các nhiễm sắc thể thường không phải các nhiễm sắc thể, X, Y, 13, 18, và 21. "Đoạn chuẩn hóa" là một ví dụ khác về "trình tự chuẩn hóa." "Trình tự đoạn chuẩn hóa" có thể gồm một đoạn của nhiễm sắc thể hoặc có thể gồm hai hoặc nhiều đoạn của cùng nhiễm sắc thể hoặc của các nhiễm sắc thể khác nhau. Theo các phương án nhất định, trình tự chuẩn hóa được dự tính để chuẩn hóa độ biến thiên như độ biến thiên liên quan đến quá trình xử lý giữa các nhiễm sắc thể (trong một lần chạy), và độ biến thiên giữa các lần giải trình tự (giữa các lần chạy).

Thuật ngữ "khả năng phân biệt" ở đây chỉ đặc tính của nhiễm sắc thể chuẩn hóa mà cho phép người có hiểu biết trung bình trong lĩnh vực phân biệt một hoặc nhiều mẫu không bị bệnh, tức là, bình thường với một hoặc nhiều mẫu bị bệnh, tức là, thể lệch bội. Nhiễm sắc thể chuẩn hóa biểu hiện "khả năng phân biệt" lớn nhất là nhiễm sắc thể hoặc nhóm gồm các nhiễm sắc thể mà đưa ra sự khác biệt thống kê lớn nhất giữa sự phân bố của định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm trong tập hợp mẫu đạt yêu cầu và định lượng nhiễm sắc thể đối với cùng nhiễm sắc thể quan tâm này trong nhiễm sắc thể tương ứng trong một hoặc nhiều mẫu bị bệnh.

Thuật ngữ "độ biến thiên" ở đây chỉ một đặc tính khác của nhiễm sắc thể chuẩn hóa mà cho phép người có hiểu biết trung bình trong lĩnh vực phân biệt một hoặc nhiều mẫu không bị bệnh, tức là, mẫu bình thường với một hoặc nhiều mẫu bị bệnh, tức là, thể lệch bội. Độ biến thiên của nhiễm sắc thể chuẩn hóa, được đo trong tập hợp mẫu đạt yêu cầu, chỉ độ biến thiên về số lượng thể trình tự được ánh xạ đến nó mà gần đúng với độ biến thiên về số lượng thể trình tự được ánh xạ đến nhiễm sắc thể quan tâm mà nó dùng làm tham số chuẩn hóa.

Thuật ngữ "mật độ thể trình tự" ở đây chỉ số lượng trình tự đọc mà được ánh xạ đến trình tự bộ gen tham chiếu, ví dụ, mật độ thể trình tự đối với nhiễm sắc thể 21 là số lượng trình tự đọc được tạo ra bằng phương pháp giải trình tự mà được ánh xạ đến nhiễm sắc thể 21 của bộ gen tham chiếu.

Thuật ngữ "tỷ lệ mật độ thể trình tự" ở đây chỉ tỷ lệ giữa số lượng thể trình tự mà được ánh xạ đến nhiễm sắc thể của bộ gen tham chiếu, ví dụ, nhiễm sắc thể 21, đến chiều dài của nhiễm sắc thể của bộ gen tham chiếu.

Thuật ngữ "định lượng trình tự" ở đây chỉ tham số mà có liên quan đến số lượng thể trình tự hoặc một tham số khác được xác định đối với trình tự quan tâm và số lượng thể trình tự hoặc tham số khác được xác định đối với trình tự chuẩn hóa. Trong một số trường hợp, định lượng trình tự là tỷ lệ giữa độ che phủ của thể trình tự hoặc tham số khác đối với trình tự quan tâm với độ che phủ của thể trình tự hoặc tham số khác đối với trình tự chuẩn hóa. Trong một số trường hợp, định lượng trình tự chỉ tham số có liên quan đến mật độ thể trình tự của trình tự quan tâm với mật độ thể trình tự của trình tự chuẩn hóa. "Định lượng trình tự thử nghiệm" là tham số mà có liên quan đến mật độ thể trình tự hoặc tham số khác của trình tự quan tâm, ví dụ, nhiễm sắc thể 21, mà trình tự chuẩn hóa, ví dụ, nhiễm sắc thể 9, được xác định trong mẫu thử nghiệm. Tương tự, "định lượng trình tự định tính" là tham số mà có liên quan đến mật độ thể trình tự hoặc tham số khác của trình tự quan tâm mà trình tự chuẩn hóa được xác định trong mẫu đạt yêu cầu.

Thuật ngữ "độ che phủ" chỉ độ nhiều của thể trình tự được ánh xạ đến một trình tự xác định được. Độ che phủ có thể được chỉ báo định lượng bởi mật độ thể trình tự (hoặc số lượng thể trình tự), tỷ lệ mật độ thể trình tự, lượng che phủ được chuẩn hóa, giá trị độ che phủ được điều chỉnh, v.v..

Thuật ngữ "lượng che phủ" chỉ sự biến đổi của độ che phủ thô và thường biểu thị số lượng tương đối của thể trình tự (đôi khi được gọi là số đếm) trong vùng của bộ gen như ngăn. Lượng che phủ có thể thu được bằng cách chuẩn hóa, điều chỉnh và/hoặc hiệu chỉnh độ che phủ thô hoặc số đếm đối với vùng của bộ gen. Ví dụ, lượng che phủ được chuẩn hóa đối với vùng có thể thu được bằng cách lấy số lượng thể trình tự được ánh xạ đến vùng này chia cho tổng số lượng thể trình tự được ánh xạ đến toàn bộ bộ gen. Lượng che phủ được chuẩn hóa cho phép so sánh độ che phủ của một bin giữa các mẫu khác nhau, có thể có độ sâu giải trình tự khác nhau. Nó khác với định lượng trình tự ở chỗ định lượng trình tự thường thu được bằng cách chia cho số lượng thể được ánh xạ đến tập hợp con của toàn bộ bộ gen. Tập hợp con là một hoặc nhiều đoạn hoặc nhiễm sắc thể chuẩn hóa. Lượng che phủ, cho dù có được chuẩn hóa hay không, có thể được hiệu chỉnh cho sự biến thiên biên dạng chung từ vùng này đến vùng kia trên bộ gen, các biến thiên tỷ lệ G-C, khoảng lệch ở các nhiễm sắc thể khỏe mạnh, v.v..

Thuật ngữ "giải trình tự thế hệ mới (Next Generation Sequencing – NGS)" ở đây chỉ phương pháp giải trình tự mà cho phép giải trình tự số lượng lớn đồng thời các phân tử được khuếch đại bằng cách tách dòng và các phân tử axit nucleic đơn. Ví dụ không giới hạn về NGS bao gồm giải trình tự bằng phương pháp tổng hợp nhờ sử dụng các đầu cuối nhuộm thuận nghịch, và giải trình tự bằng cách nối.

Thuật ngữ "tham số" ở đây chỉ giá trị số mà định rõ đặc điểm đặc trưng của hệ thống. Thông thường, tham số định rõ đặc điểm bằng số cho tập hợp dữ liệu định lượng và/hoặc mối tương quan số giữa các tập hợp dữ liệu định lượng. Ví dụ, tỷ lệ (hoặc hàm của tỷ lệ) giữa số lượng thể trình tự được ánh xạ đến nhiễm sắc thể và chiều dài của nhiễm sắc thể mà thể được ánh xạ đến, là tham số.

Các thuật ngữ "giá trị ngưỡng" và "giá trị ngưỡng định tính" ở đây chỉ số bất kỳ được dùng làm điểm ngắt để định rõ đặc điểm cho mẫu như mẫu thử nghiệm chứa axit nucleic từ sinh vật bị nghi là mắc tình trạng bệnh lý. Ngưỡng có thể được so sánh với giá trị tham số để xác định liệu mẫu đưa ra giá trị tham số này có đề xuất rằng sinh vật này có mắc tình trạng bệnh lý hay không. Theo các phương án nhất định, giá trị ngưỡng định tính được tính nhờ sử dụng tập hợp dữ liệu định tính và dùng làm giới hạn chẩn đoán biến thể số lượng bản sao, ví dụ, thể lệch bội, ở sinh vật. Nếu kết quả thu được từ các phương pháp được bộc lộ ở đây vượt quá ngưỡng, đối tượng có thể được chẩn đoán là có biến thể số lượng bản sao, ví dụ, thể ba nhiễm 21. Các giá trị ngưỡng thích hợp đối với phương pháp được bộc lộ ở đây có thể được xác định bằng cách phân tích các giá trị được chuẩn hóa (ví dụ định lượng nhiễm sắc thể, NCV hoặc NSV) được tính đối với tập hợp mẫu huấn luyện. Giá trị ngưỡng có thể được xác định bằng cách sử dụng mẫu đạt yêu cầu (tức là, không bị bệnh) trong tập huấn luyện mà bao gồm cả các mẫu đạt yêu cầu (tức là, không bị bệnh) và các mẫu bị bệnh. Các mẫu trong tập huấn luyện đã biết là có thể lệch bội nhiễm sắc thể (tức là, các mẫu bị bệnh) có thể được sử dụng để xác nhận rằng các ngưỡng được chọn có tác dụng trong việc phân biệt các mẫu bị bệnh với các mẫu không bị bệnh trong tập hợp thử nghiệm (xem các ví dụ ở đây). Việc lựa chọn ngưỡng phụ thuộc vào mức độ tin cậy mà người dùng mong muốn có để thực hiện phân loại. Theo một số phương án, tập huấn luyện được dùng để xác định giá trị ngưỡng thích hợp gồm ít nhất 10, ít nhất 20, ít nhất 30, ít nhất 40, ít nhất 50, ít nhất 60, ít nhất 70, ít nhất 80, ít

nhất 90, ít nhất 100, ít nhất 200, ít nhất 300, ít nhất 400, ít nhất 500, ít nhất 600, ít nhất 700, ít nhất 800, ít nhất 900, ít nhất 1000, ít nhất 2000, ít nhất 3000, ít nhất 4000, hoặc nhiều mẫu đạt yêu cầu hơn. Có thể là có lợi nếu sử dụng tập hợp mẫu đạt yêu cầu lớn hơn để làm tăng tính hữu dụng trong chẩn đoán của giá trị ngưỡng.

Thuật ngữ "bin" chỉ một đoạn của trình tự hoặc một đoạn của bộ gen. Theo một số phương án, các bin tiếp giáp nhau trong bộ gen hoặc nhiễm sắc thể. Mỗi bin có thể xác định một trình tự nucleotit trong bộ gen tham chiếu. Kích thước của bin có thể là 1 kb, 100 kb, 1Mb, v.v., tùy thuộc vào phép phân tích được yêu cầu bởi các ứng dụng cụ thể và mật độ thể trình tự. Ngoài các vị trí của nó trong trình tự tham chiếu, các bin có thể có các đặc tính khác như các đặc tính về độ che phủ mẫu và cấu trúc trình tự như phân đoạn G-C.

Thuật ngữ "ngưỡng che chắn" được dùng ở đây chỉ số lượng mà giá trị dựa vào số lượng thể trình tự trong bin trình tự được so sánh với nó, trong đó bin có giá trị vượt quá ngưỡng che chắn được che chắn. Theo một số phương án, ngưỡng che chắn có thể là bậc phần trăm (percentile rank), số đếm tuyệt đối, điểm chất lượng ánh xạ, hoặc các giá trị thích hợp khác. Theo một số phương án, ngưỡng che chắn có thể được xác định là bậc phần trăm của hệ số biến đổi giữa nhiều mẫu không bị bệnh. Theo các phương án khác, ngưỡng che có thể được xác định là điểm chất lượng ánh xạ, ví dụ, điểm MapQ, mà có liên quan đến độ tin cậy của việc sắp thẳng đọc với bộ gen tham chiếu. Lưu ý rằng giá trị ngưỡng che khác với giá trị ngưỡng về biến thể số lượng bản sao (CNV), giá trị ngưỡng về biến thể số lượng bản sao là điểm ngắt để định rõ đặc tính mẫu chứa axit nucleic từ cơ thể bị nghi mắc tình trạng bệnh lý liên quan đến CNV. Theo một số phương án, giá trị ngưỡng CNV được xác định liên quan đến giá trị của nhiễm sắc thể được chuẩn hóa (NCV) hoặc giá trị của đoạn được chuẩn hóa (NSV) được mô tả ở nơi khác trong bản mô tả.

Thuật ngữ "giá trị được chuẩn hóa" ở đây chỉ giá trị số mà có liên quan đến số lượng thể trình tự được xác định đối với trình tự (ví dụ nhiễm sắc thể hoặc đoạn nhiễm sắc thể) đích với số lượng thể trình tự được xác định đối với trình tự chuẩn hóa (ví dụ nhiễm sắc thể chuẩn hóa hoặc đoạn nhiễm sắc thể chuẩn hóa). Ví dụ, "giá trị được chuẩn hóa" có thể là định lượng nhiễm sắc thể như được mô tả ở nơi khác trong bản mô tả, hoặc nó có thể là NCV, hoặc có thể là NSV như được mô tả ở nơi khác trong bản mô tả.

Thuật ngữ "trình tự đọc" chỉ trình tự thu được từ một phần của mẫu axit nucleic. Thông thường, mặc dù không nhất thiết, trình tự đọc biểu thị một trình tự ngắn gồm các cặp bazơ liên tiếp trong mẫu. Trình tự đọc có thể được biểu thị bằng ký tự bởi trình tự cặp bazơ (A, T, C, hoặc G) của một phần mẫu. Nó có thể được lưu giữ trong thiết bị nhớ và được xử lý thích hợp để xác định liệu nó có đúng với trình tự tham chiếu hoặc đáp ứng các tiêu chuẩn khác. Trình tự đọc có thể thu được một cách trực tiếp từ thiết bị giải trình tự hoặc gián tiếp từ thông tin trình tự liên quan đến mẫu. Trong một số trường hợp, trình tự đọc là trình tự ADN có chiều dài đủ (ví dụ, ít nhất khoảng 25 bp) mà có thể được sử dụng để nhận dạng trình tự hoặc vùng lớn hơn, ví dụ, mà có thể được sắp thẳng và chỉ định cụ thể cho nhiễm sắc thể hoặc vùng bộ gen hoặc gen.

Thuật ngữ "trình tự đọc bộ gen" được dùng khi đề cập đến trình tự đọc thuộc đoạn bất kỳ trong toàn bộ bộ gen của cá thể.

Thuật ngữ "thẻ trình tự" được dùng ở đây thay đổi với thuật ngữ "thẻ trình tự được ánh xạ" chỉ trình tự đọc mà được chỉ định cụ thể, tức là, được ánh xạ, đến trình tự lớn hơn, ví dụ, bộ gen tham chiếu, bằng cách sắp thẳng. Thẻ trình tự được ánh xạ duy nhất đến bộ gen tham chiếu, tức là, chúng được gán cho một vị trí duy nhất đến bộ gen tham chiếu. Trừ khi có quy định khác, các thẻ mà ánh xạ đến cùng trình tự trên trình tự tham chiếu được đếm một lần. Các thẻ có thể được tạo ra dưới dạng cấu trúc dữ liệu hoặc sự lắp ghép dữ liệu khác. Theo các phương án nhất định, thẻ chứa trình tự đọc và thông tin liên quan đến trình tự đọc đó như vị trí của trình tự trong bộ gen, ví dụ, vị trí trên nhiễm sắc thể. Theo các phương án nhất định, vị trí này được chỉ rõ cho hướng sợi dương. Thẻ có thể được xác định để cho phép số lượng giới hạn lỗi bất cặp trong khi sắp thẳng với bộ gen tham chiếu. Theo một số phương án, thẻ có thể được ánh xạ đến nhiều hơn một vị trí trên bộ gen tham chiếu, tức là, thẻ mà không ánh xạ duy nhất, có thể không được bao gồm trong phân tích này.

Thuật ngữ "thẻ trình tự không dư" chỉ thẻ trình tự không ánh xạ đến cùng vị trí, được đếm để xác định giá trị nhiễm sắc thể được chuẩn hóa (NCV) theo một số phương án. Đôi khi, nhiều trình tự đọc được sắp thẳng đến cùng các vị trí trên bộ gen tham chiếu, tạo ra thẻ trình tự thừa hoặc lặp lại. Theo một số phương án, thẻ trình tự lặp lại không ánh xạ đến cùng một vị trí được bỏ qua hoặc được đếm là một "thẻ trình tự không dư" để xác định NCV. Theo

một số phương án, thẻ trình tự không dư được sắp thẳng đến các điểm không bị loại trừ được đếm tạo ra "số đếm điểm không loại trừ" (số đếm NES) để xác định NCV.

Thuật ngữ "điểm" chỉ vị trí duy nhất (tức là ID nhiễm sắc thể, vị trí và hướng nhiễm sắc thể) trên bộ gen tham chiếu. Theo một số phương án, điểm có thể đưa ra vị trí cho một gốc, thẻ trình tự, hoặc đoạn trên trình tự.

"Điểm được loại trừ" là các điểm được tìm thấy trong các vùng của bộ gen tham chiếu mà được loại trừ đối với mục đích đếm thẻ trình tự. Theo một số phương án, các điểm được loại trừ được tìm thấy trong các vùng của các nhiễm sắc thể mà chứa các trình tự lặp lại, ví dụ, tâm động và điểm cuối, và các vùng của nhiễm sắc thể mà chung cho nhiều hơn một nhiễm sắc thể, ví dụ, các vùng có mặt trên nhiễm sắc thể Y mà cũng có mặt trên nhiễm sắc thể X.

"Các điểm không loại trừ" (NES) là các điểm mà không bị loại trừ trong bộ gen tham chiếu cho mục đích đếm thẻ trình tự.

"Số đếm điểm không loại trừ" (số đếm NES) là số lượng thẻ trình tự mà được ánh xạ đến NES trên bộ gen tham chiếu. Theo một số phương án, số đếm NES là số lượng thẻ trình tự không dư được ánh xạ đến NES. Theo một số phương án, độ che phủ và các tham số liên quan như lượng che phủ được chuẩn hóa, biên dạng chung đã loại đi lượng che phủ, và định lượng nhiễm sắc thể được dựa vào số đếm NES. Theo một ví dụ, định lượng nhiễm sắc thể được tính là tỷ lệ của số đếm NES đối với nhiễm sắc thể quan tâm so với số đếm đối với nhiễm sắc thể chuẩn hóa.

Giá trị nhiễm sắc thể được chuẩn hóa (NCV) liên kết độ che phủ của mẫu thử nghiệm với độ che phủ của tập hợp mẫu huấn luyện/định tính. Theo một số phương án, NCV là dựa vào định lượng nhiễm sắc thể. Theo một số phương án, NCV liên quan đến sự chênh lệch giữa định lượng nhiễm sắc thể của nhiễm sắc thể quan tâm trong mẫu thử nghiệm và giá trị trung bình của định lượng nhiễm sắc thể tương ứng trong tập hợp mẫu đạt yêu cầu, và có thể được tính như sau:

$$NCV_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j}$$

trong đó $\hat{\mu}_j$ và $\hat{\sigma}_j$ lần lượt là giá trị trung bình ước tính và độ lệch chuẩn, đối với định lượng nhiễm sắc thể thứ j trong tập hợp mẫu đạt yêu cầu, và x_{ij} là tỷ số nhiễm sắc thể thứ j quan sát được (định lượng) đối với mẫu thử nghiệm i .

Theo một số phương án, NCV có thể được tính "trên chương trình chạy" bằng cách liên kết định lượng nhiễm sắc thể của nhiễm sắc thể quan tâm trong mẫu thử nghiệm với số trung bình của định lượng nhiễm sắc thể tương ứng trong mẫu ghép được giải trình tự trên cùng dòng tế bào là:

$$NCV_{ij} = \frac{x_{ij} - M_j}{\hat{\sigma}_j}$$

trong đó M_j là số trung bình ước tính cho định lượng nhiễm sắc thể thứ j trong tập hợp mẫu ghép được giải trình tự trên cùng cuvet dòng chảy; $\hat{\sigma}_j$ là độ lệch chuẩn đối với định lượng nhiễm sắc thể thứ j trong một hoặc nhiều tập hợp mẫu ghép được giải trình tự trên một hoặc nhiều cuvet dòng chảy, và x_{ij} là định lượng nhiễm sắc thể thứ j quan sát được đối với mẫu thử nghiệm i . Theo một phương án, mẫu thử nghiệm i là một trong các mẫu ghép được giải trình tự trên cùng cuvet dòng chảy mà từ đây M_j được xác định.

Ví dụ, đối với nhiễm sắc thể quan tâm 21 trong mẫu thử nghiệm A, được giải trình tự là một trong số 64 mẫu ghép trên một cuvet dòng chảy, NCV đối với nhiễm sắc thể 21 trong mẫu thử nghiệm A được tính là định lượng của nhiễm sắc thể 21 trong mẫu A trừ đi số trung bình của định lượng đối với nhiễm sắc thể 21 được xác định trong 64 mẫu ghép, chia cho độ lệch chuẩn của định lượng đối với nhiễm sắc thể 21 được xác định đối với 64 mẫu ghép trên cuvet dòng chảy 1, hoặc các cuvet dòng chảy bổ sung.

Như được sử dụng ở đây, thuật ngữ "được sắp thẳng," "sự sắp thẳng," hoặc "sắp thẳng" chỉ quá trình căn chỉnh trình tự đọc hoặc thể với trình tự tham chiếu và qua đó xác định liệu trình tự tham chiếu có chứa trình tự đọc hay không. Nếu trình tự tham chiếu chứa trình tự đọc, trình tự đọc có thể được ánh xạ đến trình tự tham chiếu hoặc, theo các phương án nhất định, đến vị trí cụ thể trong trình tự tham chiếu. Trong một số trường hợp, việc sắp thẳng đơn giản cho biết có hay không trình tự đọc là một số của một trình tự tham chiếu cụ thể (tức là, liệu trình tự đọc này có mặt hay không có mặt trong trình tự tham chiếu). Ví dụ, việc sắp thẳng đọc với trình tự tham chiếu đối với nhiễm sắc thể 13 của người sẽ cho biết liệu trình tự đọc

này có mặt trong trình tự tham chiếu cho nhiễm sắc thể 13 hay không. Công cụ cung cấp thông tin này có thể được gọi là bộ kiểm tra thành viên của tập hợp (set membership tester). Trong một số trường hợp, việc sắp thẳng còn chỉ ra vị trí trong trình tự tham chiếu trong đó trình tự đọc hoặc thể ánh xạ đến. Ví dụ, nếu trình tự tham chiếu là trình tự bộ gen hoàn chỉnh của người, việc sắp thẳng có thể chỉ ra rằng trình tự đọc có mặt trên nhiễm sắc thể 13, và có thể còn chỉ ra rằng trình tự đọc này là trên một sợi và/hoặc điểm cụ thể của nhiễm sắc thể 13.

Các trình tự đọc hoặc thể được sắp thẳng là một hoặc nhiều trình tự được xác định là phù hợp về trình tự của các phân tử axit nucleic của nó với trình tự đã biết từ bộ gen tham chiếu. Việc sắp thẳng có thể được thực hiện bằng tay, mặc dù nó thường được thực hiện bằng thuật toán máy tính, bởi vì sẽ không thể sắp thẳng các trình tự đọc trong khoảng thời gian hợp lý để thực hiện các phương pháp được bộc lộ ở đây. Một ví dụ về thuật toán từ việc sắp thẳng các trình tự là chương trình máy tính sắp thẳng dữ liệu nucleotit nội vùng hiệu quả (Efficient Local Alignment of nucleotit Data - ELAND) được phân phối dưới dạng một phần của đường tin Illumina Genomics Analysis pipeline. Theo cách khác, bộ lọc Bloom hoặc bộ kiểm tra thành viên của tập hợp tương tự có thể được dùng để sắp thẳng các trình tự đọc với bộ gen tham chiếu. Xem đơn sáng chế Mỹ số 61/552,374 nộp ngày 27 tháng 10 năm 2011 tài liệu này được kết hợp vào bản mô tả bằng cách viện dẫn toàn bộ. Việc so khớp trình tự đọc trong sắp thẳng có thể là khớp 100% trình tự hoặc ít hơn 100% (không khớp hoàn toàn).

Thuật ngữ "ánh xạ" được dùng ở đây chỉ việc gán cụ thể một trình tự đọc cho trình tự lớn hơn, ví dụ, bộ gen tham chiếu, bằng cách sắp thẳng.

Như được sử dụng ở đây, thuật ngữ "bộ gen tham chiếu" hoặc "trình tự tham chiếu" chỉ trình tự bộ gen đã biết cụ thể bất kỳ, dù là một phần hay toàn bộ, của sinh vật hoặc virus bất kỳ mà có thể được dùng để tham chiếu các trình tự được nhận dạng từ đối tượng. Ví dụ, bộ gen tham chiếu được dùng cho đối tượng là người cũng như nhiều sinh vật khác được tìm thấy tại trung tâm thông tin công nghệ sinh học quốc gia (National Center for Biotechnology Information) tại địa chỉ ncbi.nlm.nih.gov. "Bộ gen" chỉ toàn bộ thông tin di truyền của sinh vật hoặc virus, được biểu hiện trong trình tự axit nucleic.

Theo các phương án khác nhau, trình tự tham chiếu lớn hơn đáng kể so với trình tự đọc mà được sắp thẳng với nó. Ví dụ, có thể là lớn hơn ít nhất khoảng 100 lần, hoặc lớn hơn

ít nhất khoảng 1000 lần, hoặc lớn hơn ít nhất khoảng 10.000 lần, hoặc lớn hơn ít nhất khoảng 105 lần, hoặc lớn hơn ít nhất khoảng 106 lần, hoặc lớn hơn ít nhất khoảng 107 lần.

Trong một ví dụ, trình tự tham chiếu là trình tự của bộ gen có chiều dài đủ của người. Các trình tự này có thể được gọi là các trình tự bộ gen tham chiếu. Trong một ví dụ khác, trình tự tham chiếu bị giới hạn ở nhiễm sắc thể cụ thể của người như nhiễm sắc thể 13. Theo một số phương án, nhiễm sắc thể Y tham chiếu là trình tự nhiễm sắc thể Y từ bộ gen của người phiên bản hg19. Các trình tự này có thể được gọi là trình tự nhiễm sắc thể tham chiếu. Ví dụ khác về trình tự tham chiếu bao gồm bộ gen của các loài khác, cũng như các nhiễm sắc thể, vùng dưới nhiễm sắc thể (như sợi), v.v., của loài bất kỳ.

Theo các phương án khác nhau, trình tự tham chiếu là trình tự liên tiếp hoặc tổ hợp khác có nguồn gốc từ nhiều cá thể. Tuy nhiên, trong một số ứng dụng, trình tự tham chiếu có thể được lấy từ một cá thể cụ thể.

Thuật ngữ "trình tự liên quan về mặt lâm sàng" ở đây chỉ trình tự axit nucleic đã biết hoặc nghi là đi kèm hoặc có liên quan với trình trạng bệnh lý hoặc di truyền. Việc xác định không có mặt hoặc có mặt của trình tự liên quan về mặt lâm sàng có thể có tác dụng trong việc xác định việc chẩn đoán hoặc khẳng định chẩn đoán tình trạng bệnh lý, hoặc đưa ra tiên lượng về sự tiến triển của bệnh.

Thuật ngữ "có nguồn gốc" được dùng trong ngữ cảnh về axit nucleic hoặc hỗn hợp các axit nucleic, ở đây chỉ cách thức mà nhờ đó thu được axit nucleic từ nguồn mà từ đó chúng bắt nguồn. Ví dụ, theo một phương án, hỗn hợp các axit nucleic mà có nguồn gốc từ hai bộ gen khác nhau có nghĩa là các axit nucleic này, ví dụ, cfADN, được giải phóng trong tự nhiên bởi các tế bào thông qua các quá trình có trong tự nhiên như sự chết hoại hoặc sự chết tế bào theo chương trình. Theo một phương án khác, hỗn hợp axit nucleic mà có nguồn gốc từ hai bộ gen khác nhau có nghĩa là các axit nucleic này được tách chiết từ hai loại tế bào khác nhau từ đối tượng.

Thuật ngữ "dựa vào" khi được dùng trong ngữ cảnh về việc thu được giá trị định lượng cụ thể, ở đây chỉ việc dùng một đại lượng khác làm đầu vào để tính giá trị định lượng cụ thể dưới dạng đầu ra.

Thuật ngữ "mẫu bệnh" ở đây chỉ mẫu sinh học thu được từ người bệnh, tức là, người nhận sự quan tâm, chăm sóc hoặc điều trị y tế. Mẫu bệnh có thể là mẫu bất kỳ được mô tả ở đây. Theo các phương án nhất định, mẫu bệnh thu được bằng các quy trình không xâm lấn, ví dụ, mẫu máu ngoại vi hoặc mẫu phân. Các phương pháp được mô tả ở đây không cần giới hạn chỉ ở người. Do đó, các ứng dụng thú y khác nhau được dự tính trong đó trường hợp mẫu bệnh có thể là mẫu từ động vật có vú không phải người (ví dụ, mèo, lợn, ngựa, bò và tương tự).

Thuật ngữ "mẫu hỗn hợp" ở đây chỉ mẫu chứa hỗn hợp axit nucleic, có nguồn gốc từ các bộ gen khác nhau.

Thuật ngữ "mẫu từ người mẹ" ở đây chỉ mẫu sinh học thu được từ đối tượng mang thai, ví dụ, người phụ nữ.

Thuật ngữ "dịch sinh học" ở đây chỉ dịch lỏng được lấy từ nguồn sinh học và bao gồm, ví dụ, máu, huyết thanh, huyết tương, đờm, dịch rửa, dịch não tủy, nước tiểu, tinh dịch, mồ hôi, nước mắt, nước bọt, và tương tự. Như được sử dụng ở đây, thuật ngữ "máu," "huyết tương" và "huyết thanh" tuyệt đối bao gồm các phân đoạn hoặc các phần đã được xử lý của nó. Tương tự, trong đó mẫu được lấy từ sinh thiết, băng gạc, kính phết, v.v., "mẫu" tuyệt đối bao gồm phân đoạn hoặc phần đã được xử lý có nguồn gốc từ sinh thiết, băng gạc, kính phết, v.v..

Các thuật ngữ "axit nucleic người mẹ" và "axit nucleic bào thai" ở đây lần lượt chỉ axit nucleic của đối tượng nữ giới có thai và axit nucleic của bào thai được mang bởi phụ nữ có thai này.

Như được sử dụng ở đây, thuật ngữ "tương ứng với" đôi khi chỉ trình tự axit nucleic, ví dụ, gen hoặc nhiễm sắc thể, có mặt trong bộ gen của các đối tượng khác nhau, và không nhất thiết phải có cùng trình tự trong tất cả bộ gen, nhưng dùng để đưa ra độ tương đồng chứ không phải thông tin di truyền của trình tự quan tâm, ví dụ, gen hoặc nhiễm sắc thể.

Như được sử dụng ở đây, thuật ngữ "phân đoạn bào thai" chỉ phân đoạn axit nucleic của bào thai có mặt trong mẫu chứa axit nucleic của bào thai và mẹ. Phân đoạn bào thai thường được dùng để định rõ đặc tính của cfADN trong máu người mẹ.

Như được sử dụng ở đây thuật ngữ "nhiễm sắc thể" chỉ vật mang gen mang tính di truyền của tế bào sống, có nguồn gốc từ sợi nhiễm sắc chứa các thành phần ADN và protein (đặc biệt là histon). Hệ thống đánh số nhiễm sắc thể bộ gen của người thông thường được thừa nhận quốc tế được sử dụng trong bản mô tả.

Như được sử dụng ở đây, thuật ngữ "chiều dài polynucleotit" chỉ số lượng nucleotit tuyệt đối trong trình tự hoặc trong vùng của bộ gen tham chiếu. Thuật ngữ "chiều dài nhiễm sắc thể" chỉ chiều dài đã biết của nhiễm sắc thể cho trước theo cặp bazơ, ví dụ, được đưa ra trong tổ hợp NCBI36/hg18 của nhiễm sắc thể người được tìm thấy tại địa chỉ lgenomel.lucsl.ledu/cgi-bin/hgTracks?hgsid=167155613&chromInfoPage= trên mạng toàn cầu.

Thuật ngữ "đối tượng" ở đây chỉ đối tượng người cũng như đối tượng không phải là người như động vật có vú, động vật không xương sống, động vật có xương sống, nấm, vi khuẩn và virus. Mặc dù các ví dụ ở đây liên quan đến người và ngôn ngữ chủ yếu hướng đến mối quan tâm là con người, nhưng các khái niệm được bộc lộ ở đây có thể áp dụng được cho bộ gen từ thực vật hoặc động vật bất kỳ, và hữu dụng trong lĩnh vực y học thú y, khoa học động vật và các phòng thí nghiệm nghiên cứu và tương tự.

Thuật ngữ "tình trạng bệnh lý" ở đây chỉ "tình trạng y tế" là thuật ngữ rộng bao gồm tất cả các bệnh và rối loạn, mà còn có thể bao gồm các tổn thương và các tình huống sức khỏe bình thường, như mang thai, mà có thể ảnh hưởng đến sức khỏe của người, có lợi từ các trợ giúp y tế, hoặc có các liên quan để điều trị y tế.

Thuật ngữ "toàn phần" khi được dùng trong việc đề cập đến thể lệch bội nhiễm sắc thể ở đây chỉ sự tăng thêm hoặc mất hoàn toàn một nhiễm sắc thể.

Thuật ngữ "một phần" khi được dùng trong việc đề cập đến thể lệch bội nhiễm sắc thể ở đây chỉ sự tăng thêm hoặc mất một phần, tức là, một đoạn của nhiễm sắc thể.

Thuật ngữ "khảm" ở đây chỉ biểu thị sự có mặt của hai quần thể tế bào với hai kiểu nhân khác nhau trong một cá thể mà được phát triển từ một trứng được thụ tinh duy nhất. Hiện tượng khảm có thể là do đột biến trong quá trình phát triển mà được truyền chỉ cho nhóm nhỏ tế bào trưởng thành.

Thuật ngữ "không khảm" ở đây chỉ sinh vật, ví dụ, bào thai người, được cấu thành gồm các tế bào của một kiểu nhân.

Thuật ngữ "độ nhạy" như được sử dụng ở đây chỉ xác suất mà kết quả thử nghiệm sẽ là dương tính nếu có mặt tình trạng bệnh lý đích. Có thể được tính là số lượng dương tính thật chia cho tổng số lượng dương tính thật và âm tính giả.

Thuật ngữ "tính đặc hiệu" như được sử dụng ở đây chỉ xác suất mà kết quả thử nghiệm sẽ là âm tính nếu không có mặt tình trạng bệnh lý đích. Có thể được tính là số lượng âm tính thật chia cho tổng số lượng âm tính thật và dương tính giả.

Thuật ngữ "làm giàu" ở đây chỉ quá trình khuếch đại axit nucleic đích đa hình có chứa trong một phần của mẫu từ người mẹ, và kết hợp sản phẩm được khuếch đại với phần còn lại của mẫu từ người mẹ mà từ đó phần này được lấy ra. Ví dụ, phần còn lại của mẫu từ người mẹ có thể là mẫu từ người mẹ ban đầu.

Thuật ngữ "mẫu từ người mẹ ban đầu" ở đây chỉ mẫu sinh học không được làm giàu thu được từ đối tượng mang thai, ví dụ, người phụ nữ, người dùng làm nguồn mà từ người này một phần được lấy ra để khuếch đại axit nucleic đích đa hình.

"Mẫu ban đầu" có thể là mẫu bất kỳ thu được từ đối tượng mang thai, và các phân đoạn đã qua xử lý của nó, ví dụ, mẫu cfADN tinh chế được tách chiết từ mẫu huyết tương của người mẹ.

Thuật ngữ "đoạn môi," như được sử dụng ở đây chỉ oligonucleotit phân lập mà có thể hoạt động như điểm khởi đầu quá trình tổng hợp khi được đặt trong các điều kiện cảm ứng cho quá trình tổng hợp sản phẩm kéo dài (ví dụ, các điều kiện bao gồm nucleotit, tác nhân cảm ứng như ADN polymeraza, và nhiệt độ và pH phù hợp). Đoạn môi tốt hơn là sợi đơn để cho hiệu quả khuếch đại tối đa, nhưng có thể theo cách khác là sợi đôi. Nếu là sợi đôi, đoạn môi trước tiên được xử lý để tách các sợi của nó ra trước khi được dùng để tạo ra sản phẩm kéo dài. Tốt hơn, đoạn môi là oligodeoxyribonucleotit. Đoạn môi phải có độ dài đủ để khơi mào quá trình tổng hợp sản phẩm kéo dài khi có mặt của tác nhân cảm ứng. Chiều dài chính xác của đoạn môi sẽ phụ thuộc vào nhiều yếu tố, bao gồm nhiệt độ, nguồn cung cấp môi, và sử dụng phương pháp, và các tham số được dùng để thiết kế môi.

Giới thiệu và bối cảnh

CNV trong bộ gen của người có ảnh hưởng đáng kể đến tính đa dạng và bản chất bệnh tật ở người (Redan et al., *Nature* 23:444-454, Shaikh et al. *Genome Res* 19:1682-1690). Các bệnh này bao gồm, nhưng không chỉ giới hạn ở bệnh ung thư, bệnh lây nhiễm và bệnh tự miễn, bệnh thuộc hệ thần kinh, bệnh về trao đổi chất và/hoặc tim mạch, và tương tự.

CNV đã biết là góp phần gây ra bệnh di truyền thông qua các cơ chế khác nhau, dẫn đến sự mất cân bằng về định lượng gen hoặc sự đứt gãy gen trong hầu hết các trường hợp. Ngoài mối tương quan trực tiếp của chúng với các bệnh di truyền, CNV được biết là trung gian gây ra các thay đổi về kiểu hình mà có thể là nguy hại. Gần đây, một số nghiên cứu đã báo cáo có sự tăng gánh nặng của CNV hiếm gặp hoặc CNV khởi tạo (*de novo* CNV) trong các rối loạn phức tạp như tự kỷ, ADHD, và tâm thần phân liệt so với các đối chứng bình thường, làm nổi bật tính gây bệnh tiềm tàng của các CNV hiếm gặp hoặc duy nhất (Sebat et al., 316:445 - 449; Walsh et al., *Science* 320:539- 543). CNV xuất hiện do sự sắp xếp lại bộ gen, chủ yếu là do các biến cố mất đoạn, lặp đoạn, chèn đoạn và chuyển đoạn mất cân bằng.

Thấy rằng các đoạn cfADN từ nguồn gốc bào thai trung bình là ngắn hơn các đoạn từ nguồn gốc của mẹ. Thử nghiệm không xâm lấn trên phụ nữ mang thai (Non-invasive prenatal testing - NIPT) dựa trên dữ liệu NGS đã được thực hiện thành công. Các phương pháp hiện tại bao gồm giải trình tự mẫu từ người mẹ sử dụng các trình tự đọc ngắn (25bp-36bp), sắp thẳng với bộ gen, tính toán và chuẩn hóa vùng che phủ dưới nhiễm sắc thể, và cuối cùng đánh giá sự biểu hiện quá mức của các nhiễm sắc thể quan tâm (13/18/21/X/Y) so với vùng che phủ được chuẩn hóa dự tính đi kèm với bộ gen lưỡng bội bình thường. Do đó, thử nghiệm và phân tích NIPT truyền thống dựa vào số lượng hoặc độ che phủ để đánh giá sự có thể đúng của thể lệch bội của bào thai.

Do các mẫu huyết tương của người mẹ có mặt hỗn hợp cfADN của người mẹ và bào thai, sự thành công của phương pháp NIPT bất kỳ đưa ra phụ thuộc vào độ nhạy của nó để phát hiện sự thay đổi về số lượng bản sao trong các mẫu phân đoạn bào thai nhỏ. Đối với các phương pháp dựa vào việc đếm, độ nhạy của chúng được xác định bằng cách (a) độ sâu giải trình tự và (b) khả năng chuẩn hóa dữ liệu để làm giảm sai lệch do kỹ thuật. Bản mô tả này đề xuất phương pháp phân tích đối với NIPT và các ứng dụng khác bằng cách rút ra các thông

tin về kích thước đoạn từ, ví dụ, các trình tự đọc đầu bắt cặp, và sử dụng thông tin này trong đường dẫn phân tích. Độ nhạy phân tích được cải thiện đưa ra khả năng ứng dụng phương pháp NIPT ở độ che phủ giảm (ví dụ, độ sâu giải trình tự giảm) cho phép việc sử dụng kỹ thuật này để xét nghiệm cho các phụ nữ mang thai có nguy cơ trung bình với chi phí thấp hơn.

Phương pháp, thiết bị và hệ thống được bộc lộ ở đây để xác định số lượng bản sao và biến thể số lượng bản sao (CNV) của các trình tự quan tâm khác nhau trong mẫu thử nghiệm chứa hỗn hợp axit nucleic có nguồn gốc từ hai hoặc nhiều bộ gen khác nhau, và đã biết hoặc nghi ngờ khác về số lượng của một hoặc nhiều trình tự quan tâm. Biến thể số lượng bản sao được xác định bằng phương pháp và thiết bị được bộc lộ ở đây bao gồm tăng thêm hoặc mất đi toàn bộ nhiễm sắc thể, các thay đổi bao gồm các đoạn nhiễm sắc thể rất lớn mà có thể nhìn được bằng kính hiển vi, và sự dư thừa biến thể số lượng bản sao dưới kính hiển vi của các đoạn ADN có kích thước khoảng từ một nucleotit đến kilo bazơ (kb), đến mega bazơ (Mb).

Theo một số phương án, sáng chế đề xuất các phương pháp xác định biến thể số lượng bản sao (CNV) của bào thai nhờ sử dụng các mẫu từ người mẹ chứa ADN ngoài tế bào của người mẹ và bào thai. Một số phương pháp thực hiện sử dụng chiều dài đoạn (hoặc kích thước đoạn) của cfADN để cải thiện độ nhạy và tính đặc hiệu để phát hiện thể lệch bội ở bào thai từ cfADN trong huyết tương của người mẹ. Một số phương án được thực hiện với việc tạo thư viện không dùng PCR cùng với giải trình tự ADN đầu bắt cặp. Theo một số phương án, cả kích thước đoạn và độ che phủ được dùng để tăng cường sự phát hiện thể lệch bội ở bào thai. Theo một số phương án, phương pháp bao gồm kết hợp việc đếm độc lập các đoạn ngắn hơn với phân đoạn liên quan của các đoạn ngắn hơn trong các bin trong bộ gen.

Một số phương án được bộc lộ ở đây đề xuất các phương pháp để cải thiện độ nhạy và/hoặc độ đặc hiệu của phân tích dữ liệu trình tự bằng cách loại bỏ độ lệch hàm lượng GC trong mẫu. Theo một số phương án, việc loại bỏ độ lệch hàm lượng GC trong mẫu là dựa vào dữ liệu trình tự được hiệu chỉnh cho biến thiên hệ thống thường thấy trong các mẫu huấn luyện không bị bệnh.

Một số phương án được bộc lộ đề xuất phương pháp thu được các tham số có tín hiệu cao để gây nhiễu tỷ lệ từ các đoạn axit nucleic ngoài tế bào, để xác định các tình trạng bệnh lý di truyền khác nhau liên quan đến số lượng bản sao và CNV, với độ nhạy, tính chọn lọc,

và/hoặc tính hiệu quả được cải thiện so với các phương pháp thông thường. Các tham số bao gồm, nhưng không chỉ giới hạn ở, độ che phủ, độ che phủ được gán trọng số kích thước đoạn, phân đoạn hoặc tỷ lệ của các đoạn trong khoảng xác định, mức độ metyl hóa của các đoạn, thống kê thu được từ độ che phủ, ước tính tỷ lệ ADN bào thai thu được từ thông tin về độ che phủ, v.v.. Quy trình được mô tả đã được phát hiện là đặc biệt có hiệu quả ở chỗ cải thiện tín hiệu trong mẫu có phân đoạn ADN tương đối thấp từ bộ gen xem xét (ví dụ, bộ gen của bào thai). Ví dụ về mẫu này là mẫu máu của người mẹ từ một người mang thai sinh đôi, sinh ba, v.v., trong đó quá trình này đánh giá biến thể số lượng bản sao trong bộ gen của một trong số bào thai.

Theo một số phương án, có thể đạt được độ nhạy và độ đặc hiệu phân tích cao với việc tạo thư viện đơn giản nhờ sử dụng đầu vào cfADN rất thấp mà không cần khuếch đại bằng PCR. Phương pháp không dùng PCR đơn giản hóa tiến trình công việc, cải thiện thời gian quay vòng và loại bỏ được các sai số hệ thống vốn có khi dùng phương pháp PCR. Theo một số phương án, việc phát hiện thể lệch bội của bào thai từ huyết tương của người mẹ có thể được thực hiện chắc chắn và hiệu quả hơn các phương pháp thông thường, cần ít đoạn cfADN khác lạ hơn. Kết hợp lại, đạt được độ nhạy và độ đặc hiệu phân tích được cải thiện với thời gian quay vòng rất nhanh với số lượng đoạn cfADN thấp hơn đáng kể. Điều này cho phép xét nghiệm NIPT được thực hiện với chi phí thấp hơn đáng kể, tạo thuận lợi cho việc áp dụng trong quần thể sản khoa nói chung.

Trong các cách thực hiện khác nhau, việc tạo thư viện không dùng PCR là có thể bằng phương pháp được bộc lộ. Một số các thực hiện loại bỏ sai số hệ thống vốn có của phương pháp PCR, làm giảm độ phức tạp của thử nghiệm, giảm độ sâu giải trình tự yêu cầu (thấp hơn 2,5 lần), tạo ra thời gian quay vòng nhanh hơn, ví dụ, quay vòng trong một ngày, cho phép đo phân đoạn bào thai (FF) trong khi xử lý, tạo thuận lợi cho sự phân biệt giữa cfADN của người mẹ và của bào thai/nhau thai nhờ sử dụng thông tin kích thước đoạn.

Đánh giá CNV

Phương pháp xác định CNV

Bằng cách sử dụng giá trị độ che phủ trình tự, các tham số kích thước đoạn, và/hoặc mức độ metyl hóa được đưa ra bởi các phương pháp được bộc lộ ở đây, người có hiểu biết

trung bình trong lĩnh vực có thể xác định các tình trạng bệnh lý di truyền khác nhau có liên quan đến số lượng bản sao và CNV của trình tự, các nhiễm sắc thể, hoặc đoạn nhiễm sắc thể có độ nhạy, tính chọn lọc và/hoặc tính hiệu quả được cải thiện so với sử dụng giá trị độ che phủ trình tự thu được bằng các phương pháp thông thường. Ví dụ, theo một số phương án, trình tự tham chiếu bị che chắn được dùng để xác định sự có mặt hoặc không có mặt của hai hoặc nhiều thể lệch bội nhiễm sắc thể toàn phần ở bào thai bất kỳ trong mẫu thử nghiệm từ người mẹ chứa các phân tử axit nucleic của bào thai và người mẹ. Các phương pháp ví dụ được đưa ra dưới đây sắp thẳng các trình tự đọc với các trình tự tham chiếu (bao gồm bộ gen tham chiếu). Việc sắp thẳng được thực hiện trên các trình tự tham chiếu không bị che chắn hoặc bị che chắn, nhờ đó tạo ra các thể trình tự được ánh xạ đến trình tự tham chiếu. Theo một số phương án, chỉ các thể trình tự nằm trên các đoạn không bị che chắn của trình tự tham chiếu được đếm để xác định biến thể số lượng bản sao.

Theo một số phương án, việc đánh giá mẫu axit nucleic về CNV bao gồm việc định rõ đặc tính tình trạng của thể lệch bội nhiễm sắc thể hoặc đoạn bằng một trong ba kiểu xác định: "bình thường" hoặc "không bị bệnh," "bị bệnh," và "không xác định." Ngưỡng để xác định bình thường và bị bệnh là tập hợp điển hình. Tham số liên quan đến thể lệch bội hoặc biến thể số lượng bản sao khác được đo trong mẫu và giá trị đo được được so sánh với ngưỡng. Đối với thể lệch bội dạng lặp, việc xác định là bị bệnh nếu định lượng nhiễm sắc thể hoặc đoạn (hoặc lượng trình tự có giá trị đo được khác) cao hơn tập hợp ngưỡng xác định đối với các mẫu bị bệnh. Đối với các thể lệch bội này, việc xác định là bình thường nếu định lượng nhiễm sắc thể hoặc đoạn thấp hơn tập hợp ngưỡng đối với các mẫu bình thường. Ngược lại, đối với các thể lệch bội dạng mất, việc xác định là bị bệnh nếu định lượng nhiễm sắc thể hoặc đoạn thấp hơn ngưỡng xác định đối với các mẫu bị bệnh, và việc xác định là bình thường nếu định lượng nhiễm sắc thể hoặc đoạn cao hơn tập hợp ngưỡng đối với các mẫu bình thường. Ví dụ, khi có mặt của thể ba nhiễm, xác định là "bình thường" bằng giá trị của tham số, ví dụ, định lượng nhiễm sắc thể xét nghiệm mà thấp hơn ngưỡng tin cậy được xác định bởi người dùng, và xác định là "bị bệnh" bằng tham số, ví dụ, định lượng nhiễm sắc thể xét nghiệm, mà cao hơn ngưỡng tin cậy được xác định bởi người dùng. Xác định kết quả là "không xác định" bằng tham số, ví dụ, định lượng nhiễm sắc thể xét nghiệm mà nằm giữa các ngưỡng để xác

định là "bình thường" hoặc "bị bệnh". Thuật ngữ "không xác định" được dùng thay đổi với "không xếp loại".

Các tham số mà có thể được dùng để xác định CNV bao gồm, nhưng không chỉ giới hạn ở, độ che phủ, độ che phủ chệch/được gán trọng số bởi kích thước đoạn, phần hoặc tỷ lệ của các đoạn trong khoảng kích thước xác định, và mức độ metyl hóa của các đoạn. Như được bàn luận ở đây, độ che phủ thu được từ việc đếm các trình tự đọc được sắp thẳng với một vùng của bộ gen tham chiếu và tùy ý được chuẩn hóa để tạo ra số đếm thể trình tự. Theo một số phương án, số đếm thể trình tự có thể được gán trọng số bởi kích thước đoạn.

Theo một số phương án, tham số kích thước đoạn bị chệch theo đặc trưng kích thước đoạn của một trong các bộ gen. Tham số kích thước đoạn là tham số mà có liên quan đến kích thước của đoạn. Tham số mà được làm lệch theo kích thước đoạn nếu: 1) tham số này được gán trọng số thuận đối với kích thước, ví dụ, tổng số được gán trọng số có sức nặng hơn đối với kích thước này so với các kích thước khác; hoặc 2) tham số này thu được từ giá trị mà được gán trọng số thuận đối với kích thước đoạn này, ví dụ, tỷ số thu được từ tổng số được gán trọng số có sức nặng hơn đối với kích thước này. Kích thước là một đặc trưng của bộ gen khi bộ gen có nồng độ axit nucleic có kích thước này được làm giàu hoặc cao hơn so với bộ gen khác hoặc phần khác của cùng một bộ gen.

Theo một số phương án, phương pháp xác định sự có mặt hoặc không có mặt của thể lệch bội nhiễm sắc thể toàn phần bất kỳ của bào thai trong mẫu thử nghiệm của người mẹ bao gồm (a) thu nhận thông tin trình tự về axit nucleic của bào thai và người mẹ trong mẫu thử nghiệm của người mẹ; (b) sử dụng thông tin trình tự này và phương pháp được bộc lộ trên đây để xác định số lượng thể trình tự, lượng che phủ trình tự, tham số kích thước đoạn, hoặc tham số khác đối với mỗi trong số các nhiễm sắc thể quan tâm được chọn từ các nhiễm sắc thể 1-22, X và Y và để xác định số lượng thể trình tự hoặc tham số khác đối với một hoặc nhiều trình tự nhiễm sắc thể chuẩn hóa; (c) sử dụng số lượng thể trình tự hoặc tham số khác xác định được đối với mỗi trong số các nhiễm sắc thể quan tâm và số lượng thể trình tự hoặc tham số khác xác định được đối với mỗi trong số các nhiễm sắc thể chuẩn hóa để tính mỗi định lượng nhiễm sắc thể đối với mỗi trong số các nhiễm sắc thể quan tâm; và (d) so sánh mỗi định lượng nhiễm sắc thể này với giá trị ngưỡng, và qua đó xác định sự có mặt hoặc không

có mặt của thể lệch bội nhiễm sắc thể toàn phần bất kỳ của bào thai trong mẫu thử nghiệm của người mẹ.

Theo một số phương án, bước (a) được mô tả trên đây có thể bao gồm giải trình tự ít nhất một phần của phân tử axit nucleic của mẫu thử nghiệm để thu nhận thông tin trình tự nói trên về phân tử axit nucleic của bào thai và người mẹ trong mẫu thử nghiệm. Theo một số phương án, bước (c) bao gồm việc tính toán mỗi định lượng nhiễm sắc thể đối với mỗi trong số các nhiễm sắc thể quan tâm dưới dạng tỷ lệ của số lượng thể trình tự hoặc tham số khác xác định được đối với mỗi trong số các nhiễm sắc thể quan tâm và số lượng thể trình tự hoặc tham số khác xác định được đối với (các) trình tự nhiễm sắc thể chuẩn hóa. Theo các phương án khác, định lượng nhiễm sắc thể là dựa vào lượng che phủ trình tự được xử lý thu được từ số lượng thể trình tự hoặc tham số khác. Theo một số phương án, chỉ các thể trình tự không dư, duy nhất được dùng để tính lượng che phủ trình tự được xử lý hoặc tham số khác. Theo một số phương án, lượng che phủ trình tự được xử lý là tỷ lệ mật độ thể trình tự, là số lượng thể trình tự được tiêu chuẩn hóa theo chiều dài trình tự. Theo một số phương án, lượng che phủ trình tự được xử lý hoặc tham số khác là thể trình tự được chuẩn hóa hoặc tham số được chuẩn hóa khác, là số lượng thể trình tự hoặc tham số khác của trình tự quan tâm chia cho số lượng thể trình tự hoặc tham số khác của tất cả hoặc một phần đáng kể của bộ gen. Theo một số phương án, lượng che phủ trình tự được xử lý hoặc tham số khác như tham số kích thước đoạn được điều chỉnh theo biên dạng chung của trình tự quan tâm. Theo một số phương án, lượng che phủ trình tự được xử lý hoặc tham số khác được điều chỉnh theo mối tương quan trong phạm vi mẫu giữa hàm lượng GC và độ che phủ trình tự đối với mẫu được xét nghiệm. Theo một số phương án, lượng che phủ trình tự được xử lý hoặc tham số khác là kết quả của sự kết hợp các quá trình này, sẽ được mô tả kỹ hơn ở nơi khác trong bản mô tả.

Theo một số phương án, định lượng nhiễm sắc thể được tính là tỷ lệ của độ che phủ trình tự được xử lý hoặc tham số khác đối với mỗi trong số các nhiễm sắc thể quan tâm và độ che phủ trình tự được xử lý hoặc tham số khác đối với (các) trình tự nhiễm sắc thể chuẩn hóa.

Theo phương án bất kỳ trong số các phương án nêu trên, thể lệch bội nhiễm sắc thể toàn phần được chọn từ thể ba nhiễm sắc thể toàn phần, thể đơn nhiễm sắc thể toàn phần và thể đa bội nhiễm sắc thể toàn phần. Thể lệch bội nhiễm sắc thể toàn phần được chọn từ thể

lệch bội toàn phần của một nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể 1-22, X, và Y. Ví dụ, các thể lệch bội nhiễm sắc thể toàn phần khác nhau của bào thai nói trên được chọn từ thể ba nhiễm 2, thể ba nhiễm 8, thể ba nhiễm 9, thể ba nhiễm 20, thể ba nhiễm 21, thể ba nhiễm 13, thể ba nhiễm 16, thể ba nhiễm 18, thể ba nhiễm 22, , 47,XXX, 47,XYY, và thể đơn nhiễm X.

Theo phương án bất kỳ trong số các phương án nêu trên, các bước (a)-(d) được lặp lại đối với mẫu thử nghiệm từ các đối tượng người mẹ khác nhau, và phương pháp bao gồm việc xác định sự có mặt hoặc không có mặt của hai hoặc nhiều thể lệch bội nhiễm sắc thể toàn phần khác nhau của bào thai bất kỳ trong mỗi mẫu thử nghiệm.

Theo phương án bất kỳ trong số các phương án nêu trên, phương pháp có thể còn bao gồm việc tính giá trị nhiễm sắc thể được chuẩn hóa (NCV), trong đó NCV liên kết định lượng nhiễm sắc thể với giá trị trung bình của định lượng nhiễm sắc thể tương ứng trong tập hợp mẫu đạt yêu cầu là:

$$NCV_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j}$$

trong đó $\hat{\mu}_j$ và $\hat{\sigma}_j$ lần lượt là giá trị trung bình ước tính và độ lệch chuẩn, đối với định lượng nhiễm sắc thể thứ j trong tập hợp mẫu đạt yêu cầu, và x_{ij} là định lượng nhiễm sắc thể thứ j quan sát được đối với mẫu thử nghiệm i.

Theo một số phương án, NCV có thể được tính "trên chương trình chạy" bằng cách liên kết định lượng nhiễm sắc thể của nhiễm sắc thể quan tâm trong mẫu thử nghiệm với số trung bình của định lượng nhiễm sắc thể tương ứng trong mẫu ghép được giải trình tự trên cùng dòng tế bào là:

$$NCV_{ij} = \frac{x_{ij} - M_j}{\hat{\sigma}_j}$$

trong đó M_j là số trung bình ước tính cho định lượng nhiễm sắc thể thứ j trong tập hợp mẫu ghép được giải trình tự trên cùng cuvet dòng chảy; $\hat{\sigma}_j$ là độ lệch chuẩn đối với định lượng nhiễm sắc thể thứ j trong một hoặc nhiều tập hợp mẫu ghép được giải trình tự trên một hoặc nhiều cuvet dòng chảy, và x_i là định lượng nhiễm sắc thể thứ j quan sát được đối với mẫu thử nghiệm i.

Theo một phương án, mẫu thử nghiệm i là một trong các mẫu ghép được giải trình tự trên cùng cuvet dòng chảy mà từ đây M_j được xác định.

Theo một số phương án, sáng chế đề xuất phương pháp xác định sự có mặt hoặc không có mặt của các thể lệch bội nhiễm sắc thể một phần khác nhau của bào thai trong mẫu thử nghiệm của người mẹ chứa axit nucleic của bào thai và người mẹ. Phương pháp này bao gồm các quy trình tương tự như phương pháp để phát hiện thể lệch bội toàn phần như nêu trên. Tuy nhiên, thay vì phân tích toàn bộ nhiễm sắc thể, thì chỉ phân tích một đoạn của nhiễm sắc thể. Xem đơn sáng chế Mỹ số 2013/0029852, được kết hợp bằng cách viện dẫn.

Fig.1 thể hiện phương pháp xác định sự có mặt của biến thể số lượng bản sao theo một số phương án. Quá trình 100 được thể hiện trên Fig.1 sử dụng độ che phủ của thể trình tự dựa trên số lượng thể trình tự (tức là, số đếm thể trình tự) để xác định CNV. Tuy nhiên, tương tự như phần mô tả trên đây đối với việc tính NCV, các biến số hoặc tham số khác, như kích thước, tỷ lệ kích thước, và mức độ metyl hóa, có thể được dùng thay cho độ che phủ. Theo một số cách thực hiện, hai hoặc nhiều biến số được kết hợp để xác định CNV. Hơn nữa, độ che phủ và các tham số khác có thể được gán trọng số dựa vào kích của các đoạn mà từ đó các thể được rút ra. Để dễ hiểu, chỉ độ che phủ được đề cập trong quy trình 100 được minh họa trên Fig.1, nhưng cần lưu ý rằng các tham số khác, như kích thước, tỷ lệ kích thước, và mức độ metyl hóa, tổng số được gán trọng số theo kích thước, v.v. có thể được dùng thay cho độ che phủ.

Trong các quá trình thao tác 130 và 135, độ che phủ của thể trình tự định tính (hoặc trị số của tham số khác) và độ che phủ của thể trình tự thử nghiệm (hoặc giá trị của tham số khác) được xác định. Bản mô tả này đề xuất các quy trình xác định lượng che phủ mà đưa đến độ nhạy và tính chọn lọc được cải thiện so với các phương pháp thông thường. Quá trình thao tác 130 và 135 được đánh dấu bằng dấu sao và được làm nổi bật bằng các hộp đường nét đậm để chỉ rõ các thao tác này góp phần làm cải thiện các giải pháp kỹ thuật trước đó. Theo một số phương án, lượng che phủ của thể trình tự được chuẩn hóa, điều chỉnh, sắp xếp, và được xử lý theo cách khác để cải thiện độ nhạy và tính chọn lọc của phân tích. Các quá trình này được mô tả chi tiết hơn ở nơi khác trong bản mô tả.

Từ góc độ khái quát, phương pháp thực hiện việc sử dụng các trình tự chuẩn hóa của mẫu huấn luyện định tính trong việc xác định CNV của mẫu thử nghiệm. Theo một số phương án, mẫu huấn luyện định tính không bị bệnh và có số lượng bản sao bình thường. Các trình tự chuẩn hóa đưa đến cơ chế để chuẩn hóa các phép đo đối với các độ biến đổi trong và giữa các lần chạy. Các trình tự chuẩn hóa được xác định nhờ sử dụng thông tin trình tự từ tập hợp mẫu đạt yêu cầu thu được từ đối tượng đã biết là chứa tế bào có số lượng bản sao bình thường đối với trình tự quan tâm bất kỳ, ví dụ, nhiễm sắc thể hoặc đoạn của nó xác định các trình tự chuẩn hóa được nêu trong các bước 110, 120, 130, 145 và 146 của phương án theo phương pháp được mô tả trên Fig.1. Theo một số phương án, các trình tự chuẩn hóa được dùng để tính định lượng trình tự đối với các trình tự thử nghiệm. Xem bước 150. Theo một số phương án, các trình tự chuẩn hóa còn được dùng để tính ngưỡng mà định lượng trình tự của trình tự quan tâm được so sánh với ngưỡng này. Xem bước 150. Thông tin trình tự thu được từ trình tự chuẩn hóa và trình tự thử nghiệm được dùng để xác định nhận dạng có ý nghĩa thống kê của thể lệch bội nhiễm sắc thể trong mẫu thử nghiệm (bước 160).

1 Trở lại các chi tiết của phương pháp xác định sự có mặt của biến thể số lượng bản sao theo một số phương án, Fig.1 đưa ra lưu đồ 100 của một phương án để xác định CNV của trình tự quan tâm, ví dụ, nhiễm sắc thể hoặc đoạn của nó trong mẫu sinh học. Theo một số phương án, mẫu sinh học thu được từ đối tượng và chứa hỗn hợp axit nucleic được đóng góp từ các bộ gen khác nhau. Các bộ gen khác nhau này có thể được đóng góp vào mẫu bởi hai cá thể, ví dụ, các bộ gen khác nhau được đóng góp bởi bào thai và người mẹ mang bào thai này. Ngoài ra, các bộ gen khác nhau này có thể được đóng góp vào mẫu bởi ba hoặc nhiều hơn ba cá thể, ví dụ, các bộ gen khác nhau được đóng góp bởi hai hoặc nhiều hơn hai bào thai và người mẹ mang các bào thai này. Theo cách khác, các bộ gen được đóng góp vào mẫu bởi tế bào ung thư thể lệch bội và tế bào thể bội chính bình thường từ cùng một đối tượng, ví dụ, mẫu huyết tương của người bệnh ung thư.

Ngoài việc phân tích mẫu thử nghiệm của người bệnh, một hoặc nhiều nhiễm sắc thể chuẩn hóa hoặc một hoặc nhiều đoạn nhiễm sắc thể chuẩn hóa được chọn cho mỗi nhiễm sắc thể quan tâm có thể. Các nhiễm sắc thể hoặc đoạn chuẩn hóa được nhận dạng không đồng thời từ việc xét nghiệm bình thường các mẫu bệnh, có thể xảy ra trong thực hành lâm sàng.

Nói cách khác, các nhiễm sắc thể hoặc đoạn chuẩn hóa được nhận dạng trước khi xét nghiệm các mẫu bệnh. Mỗi quan hệ giữa các nhiễm sắc thể hoặc đoạn chuẩn hóa và các nhiễm sắc thể hoặc đoạn đích được lưu giữ để sử dụng trong quá trình xét nghiệm. Như được giải thích dưới đây, mỗi quan hệ này được duy trì điển hình trong các khoảng thời gian trải qua xét nghiệm nhiều mẫu. Phần bản luận sau đây liên quan đến các phương án để chọn lựa các nhiễm sắc thể hoặc đoạn nhiễm sắc thể chuẩn hóa cho các nhiễm sắc thể hoặc đoạn đích của cá thể.

Tập hợp các mẫu đạt yêu cầu thu được để nhận dạng các trình tự chuẩn hóa định tính và đưa ra giá trị phương sai để sử dụng trong việc xác định nhận dạng có ý nghĩa thống kê CNV trong mẫu thử nghiệm. Trong bước 110, nhiều mẫu sinh học định tính thu được từ nhiều đối tượng đã biết là chứa tế bào có số lượng bản sao bình thường đối với một trình tự quan tâm bất kỳ. Theo một phương án, mẫu đạt yêu cầu thu được từ những người mẹ mang thai mà đã được xác nhận nhờ sử dụng các phương pháp di truyền tế bào là có số lượng bản sao bình thường của các nhiễm sắc thể. Mẫu sinh học định tính có thể là dịch sinh học, ví dụ, huyết tương, hoặc mẫu thích hợp bất kỳ như được mô tả dưới đây. Theo một số phương án, mẫu đạt yêu cầu chứa hỗn hợp các phân tử axit nucleic, ví dụ, các phân tử cfADN. Theo một số phương án, mẫu đạt yêu cầu là mẫu huyết tương của người mẹ chứa hỗn hợp các phân tử cfADN của người mẹ và bào thai. Thông tin trình tự để chuẩn hóa các nhiễm sắc thể và/hoặc các đoạn của nó thu được bằng cách giải trình tự ít nhất một phần của axit nucleic, ví dụ, axit nucleic của bào thai và người mẹ, sử dụng phương pháp giải trình tự bất kỳ đã biết. Tốt hơn là, phương pháp bất kỳ trong số các phương pháp giải trình tự thế hệ mới (NGS) được mô tả ở nơi khác trong bản mô tả được dùng để giải trình tự axit nucleic của bào thai và người mẹ dưới dạng một phân tử hoặc các phân tử được khuếch đại bằng cách nhân dòng. Theo các phương án khác nhau, mẫu đạt yêu cầu được xử lý như được bộc lộ dưới đây trước và trong quá trình giải trình tự. Chúng có thể được xử lý bằng cách sử dụng thiết bị, hệ thống, và kit như được bộc lộ ở đây.

Trong bước 120, ít nhất một phần của mỗi trong số tất cả các axit nucleic định tính chứa trong mẫu đạt yêu cầu được giải trình tự để tạo ra hàng triệu trình tự đọc, ví dụ, các trình tự đọc 36bp, được sắp thẳng với bộ gen tham chiếu, ví dụ, hg18. Theo một số phương án, các trình tự đọc gồm khoảng 20bp, khoảng 25bp, khoảng 30bp, khoảng 35bp, khoảng 40bp,

khoảng 45bp, khoảng 50bp, khoảng 55bp, khoảng 60bp, khoảng 65bp, khoảng 70bp, khoảng 75bp, khoảng 80bp, khoảng 85bp, khoảng 90bp, khoảng 95bp, khoảng 100bp, khoảng 110bp, khoảng 120bp, khoảng 130, khoảng 140bp, khoảng 150bp, khoảng 200bp, khoảng 250bp, khoảng 300bp, khoảng 350bp, khoảng 400bp, khoảng 450bp, hoặc khoảng 500bp. Mong muốn rằng các tiến bộ của công nghệ này sẽ cho phép các trình tự đọc đầu cuối đơn lớn hơn 500bp cho phép các trình tự đọc lớn hơn khoảng 1000bp nếu trình tự đọc đầu bắt cặp được tạo ra. Theo một phương án, các trình tự đọc được ánh xạ gồm 36bp. Theo một phương án khác, các trình tự đọc được ánh xạ gồm 25bp.

Các trình tự đọc được sắp thẳng với bộ gen tham chiếu, và các trình tự đọc mà được ánh xạ duy nhất đến bộ gen tham chiếu được biết là các thẻ trình tự. Các thẻ trình tự nằm trong đoạn được che chắn của trình tự tham chiếu được che chắn không được đếm để phân tích CNV.

Theo một phương án, ít nhất khoảng 3×10^6 thẻ trình tự định tính, ít nhất khoảng 5×10^6 thẻ trình tự định tính, ít nhất khoảng 8×10^6 thẻ trình tự định tính, ít nhất khoảng 10×10^6 thẻ trình tự định tính, ít nhất khoảng 15×10^6 thẻ trình tự định tính, ít nhất khoảng 20×10^6 thẻ trình tự định tính, ít nhất khoảng 30×10^6 thẻ trình tự định tính, ít nhất khoảng 40×10^6 thẻ trình tự định tính, hoặc ít nhất khoảng 50×10^6 thẻ trình tự định tính gồm các trình tự đọc 20 đến 40bp thu được từ các trình tự đọc mà ánh xạ duy nhất đến bộ gen tham chiếu.

Trong bước 130, tất cả các thẻ thu được từ việc giải trình tự axit nucleic trong các mẫu đạt yêu cầu được đếm để thu được độ che phủ của thẻ trình tự định tính. Tương tự, trong thao tác 135, tất cả các thẻ thu được từ mẫu thử nghiệm được đếm để thu được độ che phủ của thẻ trình tự thử nghiệm. Bản mô tả này đề xuất các quy trình xác định lượng che phủ mà đưa đến độ nhạy và tính chọn lọc được cải thiện so với các phương pháp thông thường. Quá trình thao tác 130 và 135 được đánh dấu bằng dấu sao và được làm nổi bật bằng các hộp đường nét đậm để chỉ rõ các thao tác này góp phần làm cải thiện các giải pháp kỹ thuật trước đó. Theo một số phương án, lượng che phủ của thẻ trình tự được chuẩn hóa, điều chỉnh, sắp xếp, và được xử lý theo cách khác để cải thiện độ nhạy và tính chọn lọc của phân tích. Các quá trình này được mô tả chi tiết hơn ở nơi khác trong bản mô tả.

Do tất cả các thẻ trình tự định tính được ánh xạ và được đếm trong mỗi trong số các mẫu đạt yêu cầu, độ che phủ của thẻ trình tự đối với trình tự quan tâm, ví dụ, trình tự liên quan về mặt lâm sàng, trong mẫu đạt yêu cầu được xác định, như là độ che phủ của thẻ trình tự đối với các trình tự bổ sung mà từ đó các trình tự chuẩn hóa được nhận dạng về sau.

Theo một số phương án, trình tự quan tâm là nhiễm sắc thể mà có liên quan đến thẻ lệch bội nhiễm sắc thể toàn phần, ví dụ, nhiễm sắc thể 21, và trình tự chuẩn hóa định tính là một nhiễm sắc thể hoàn chỉnh mà không đi kèm với thẻ lệch bội nhiễm sắc thể và độ biến thiên của nó về độ che phủ của thẻ trình tự gần đúng với độ biến thiên của trình tự (tức là, nhiễm sắc thể) quan tâm, ví dụ, nhiễm sắc thể 21. Các nhiễm sắc thể chuẩn hóa được chọn có thể là một hoặc một nhóm mà gần đúng nhất với độ biến thiên của trình tự quan tâm. Một hoặc nhiều nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể 1-22, X, và Y có thể là trình tự quan tâm, và một hoặc nhiều nhiễm sắc thể có thể được nhận dạng là trình tự chuẩn hóa cho mỗi trong số nhiễm sắc thể bất kỳ trong các nhiễm sắc thể 1-22, X và Y trong mẫu đạt yêu cầu. Nhiễm sắc thể chuẩn hóa có thể là một nhiễm sắc thể hoặc nó có thể là một nhóm các nhiễm sắc thể như được mô tả ở nơi khác trong bản mô tả.

Theo một phương án khác, trình tự quan tâm là một đoạn của nhiễm sắc thể có liên quan đến thẻ lệch bội một phần, ví dụ, mất đoạn hoặc chèn đoạn trên nhiễm sắc thể, hoặc sự chuyển đoạn mất cân bằng trên nhiễm sắc thể, và trình tự chuẩn hóa là một đoạn nhiễm sắc thể (hoặc nhóm gồm nhiều đoạn) mà không đi kèm với thẻ lệch bội một phần và độ biến thiên của nó về độ che phủ của thẻ trình tự gần đúng với độ biến thiên của đoạn nhiễm sắc thể có liên quan đến thẻ lệch bội một phần. (Các) đoạn nhiễm sắc thể chuẩn hóa được chọn có thể là một hoặc nhiều đoạn mà gần đúng nhất với độ biến thiên của trình tự quan tâm. Một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể 1-22, X, và Y có thể là trình tự quan tâm.

Theo các phương án khác, trình tự quan tâm là một đoạn của nhiễm sắc thể đi kèm với thẻ lệch bội nhiễm sắc thể và trình tự chuẩn hóa là toàn bộ nhiễm sắc thể hoặc các nhiễm sắc thể. Theo các phương án khác nữa, trình tự quan tâm là toàn bộ nhiễm sắc thể đi kèm với thẻ lệch bội và trình tự chuẩn hóa là đoạn hoặc các đoạn nhiễm sắc thể mà không đi kèm với thẻ lệch bội.

Liệu một trình tự hoặc nhóm các trình tự được xác định trong mẫu đạt yêu cầu là (các) trình tự chuẩn hóa cho một hoặc nhiều trình tự quan tâm, trình tự chuẩn hóa định tính có thể được chọn để có độ biến thiên về độ che phủ của thẻ trình tự hoặc tham số kích thước đoạn mà gần đúng nhất hoặc có hiệu quả với độ che phủ của thẻ trình tự của trình tự quan tâm như được xác định trong mẫu đạt yêu cầu. Ví dụ, các trình tự chuẩn hóa định tính là trình tự mà tạo ra độ biến thiên nhỏ nhất trong các mẫu đạt yêu cầu khi được dùng để chuẩn hóa trình tự quan tâm, tức là, độ biến thiên của trình tự chuẩn hóa là gần nhất với độ biến thiên của trình tự quan tâm được xác định trong mẫu đạt yêu cầu. Nói cách khác, trình tự chuẩn hóa định tính là trình tự được chọn để tạo ra độ biến thiên ít nhất về định lượng trình tự (đối với trình tự quan tâm) giữa các mẫu đạt yêu cầu. Do đó, quá trình này chọn lọc một trình tự mà khi được sử dụng làm nhiễm sắc thể chuẩn hóa được mong muốn là tạo ra độ biến thiên nhỏ nhất về định lượng nhiễm sắc thể giữa các lần chạy đối với trình tự quan tâm.

Trình tự chuẩn hóa được nhận dạng trong mẫu đạt yêu cầu đối với một hoặc nhiều trình tự quan tâm bất kỳ vẫn là trình tự chuẩn hóa được lựa chọn để xác định sự có mặt hoặc không có mặt của thẻ lệch bội trong mẫu thử nghiệm qua các ngày, tuần, tháng, và có thể là năm, với điều kiện là các quy trình cần phải tạo các thư viện giải trình tự, và giải trình tự các mẫu nhất thiết không được thay đổi theo thời gian. Như được mô tả trên đây, các trình tự chuẩn hóa để xác định sự có mặt của thẻ lệch bội được chọn vì (có thể là trong số các lý do khác nữa) độ biến thiên về số lượng thẻ trình tự hoặc giá trị của tham số kích thước đoạn mà được ánh xạ đến nó trong số các mẫu, ví dụ, các mẫu khác nhau, và các lần chạy giải trình tự, ví dụ, các lần chạy giải trình tự diễn ra trong cùng một ngày và/hoặc các ngày khác nhau, mà gần đúng nhất với độ biến thiên của trình tự quan tâm mà nó được dùng làm tham số chuẩn hóa. Các thay đổi quan trọng trong các quy trình này sẽ ảnh hưởng đến số lượng thẻ mà được ánh xạ đến tất cả các trình tự, từ đó sẽ xác định trình tự hoặc nhóm trình tự nào sẽ có độ biến thiên trong số các mẫu trong cùng một lần chạy giải trình tự và/hoặc trong các lần chạy giải trình tự khác nhau, trong cùng một ngày hoặc trong các ngày khác nhau mà gần đúng nhất với độ biến thiên của (các) trình tự quan tâm, sẽ đòi hỏi rằng tập hợp các trình tự chuẩn hóa cần được xác định lại. Các thay đổi quan trọng trong quy trình bao gồm các thay đổi về quy trình chuẩn trong phòng thí nghiệm được sử dụng để tạo thư viện giải trình tự, bao gồm các thay

đổi liên quan đến việc chuẩn bị mẫu để giải trình tự đa hợp thay cho giải trình tự đơn, và các thay đổi về nền giải trình tự, bao gồm các thay đổi về hóa học được dùng trong giải trình tự.

Theo một số phương án, trình tự chuẩn hóa được chọn để chuẩn hóa trình tự quan tâm cụ thể là trình tự mà phân biệt tốt nhất một hoặc nhiều mẫu đạt yêu cầu từ một hoặc nhiều mẫu bị bệnh, ngụ ý rằng trình tự chuẩn hóa là trình tự mà có khả năng phân biệt lớn nhất, tức là, khả năng phân biệt của trình tự chuẩn hóa là sao cho nó đưa ra sự phân biệt tối ưu với trình tự quan tâm trong mẫu thử nghiệm bị bệnh để dễ dàng phân biệt mẫu thử nghiệm bị bệnh với các mẫu không bị bệnh khác. Theo các phương án khác, trình tự chuẩn hóa là trình tự có sự kết hợp của độ biến thiên nhỏ nhất và khả năng phân biệt lớn nhất.

Mức độ của khả năng phân biệt có thể được xác định dưới dạng độ lệch thống kê giữa các định lượng trình tự, ví dụ, định lượng nhiễm sắc thể hoặc định lượng đoạn, trong quần thể mẫu đạt yêu cầu và định lượng nhiễm sắc thể trong một hoặc nhiều mẫu thử nghiệm như được mô tả dưới đây và được thể hiện trong các ví dụ. Ví dụ, khả năng phân biệt có thể được biểu diễn bằng số dưới dạng giá trị trong thử nghiệm, thể hiện độ chênh lệch thống kê giữa định lượng nhiễm sắc thể trong tập hợp mẫu đạt yêu cầu và định lượng nhiễm sắc thể trong một hoặc nhiều mẫu thử nghiệm. Tương tự, khả năng phân biệt có thể dựa trên định lượng đoạn thay vì định lượng nhiễm sắc thể. Theo cách khác, khả năng phân biệt có thể được biểu diễn bằng số dưới dạng giá trị nhiễm sắc thể được chuẩn hóa (Normalized Chromosome Value - NCV), là điểm số z đối với định lượng nhiễm sắc thể miễn là sự phân bố đối với NCV là bình thường. Tương tự, trong trường hợp mà các đoạn nhiễm sắc thể là các trình tự quan tâm, khả năng phân biệt của định lượng đoạn có thể được biểu diễn bằng số dưới dạng giá trị đoạn được chuẩn hóa (Normalized Segment Value-NSV), là điểm số z đối với định lượng đoạn nhiễm sắc thể miễn là sự phân bố đối với NSV là bình thường. Trong xác định điểm số z , độ lệch trung bình và độ lệch chuẩn của định lượng nhiễm sắc thể hoặc đoạn trong tập hợp mẫu đạt yêu cầu có thể được sử dụng. Theo cách khác, độ lệch trung bình và độ lệch chuẩn của định lượng nhiễm sắc thể hoặc đoạn trong tập huấn luyện bao gồm mẫu đạt yêu cầu và mẫu bị bệnh có thể được sử dụng. Theo các phương án khác, trình tự chuẩn hóa là trình tự mà có độ biến thiên nhỏ nhất và khả năng phân biệt lớn nhất hoặc sự kết hợp tối ưu của độ biến thiên nhỏ và khả năng phân biệt lớn.

Phương pháp nhận dạng các trình tự mà vốn có các đặc tính tương tự và thiên về các biến thiên tương tự trong số các mẫu và chạy giải trình tự, và hữu dụng để xác định định lượng trình tự trong mẫu thử nghiệm.

Xác định định lượng trình tự

Theo một số phương án, định lượng nhiễm sắc thể hoặc đoạn đối với một hoặc nhiều nhiễm sắc thể hoặc đoạn đích được xác định trong tất cả các mẫu đạt yêu cầu như được mô tả trong bước 146 được thể hiện trên Fig.1, và trình tự nhiễm sắc thể hoặc đoạn chuẩn hóa được nhận dạng trong bước 145. Một số trình tự chuẩn hóa được tạo ra trước khi định lượng trình tự được tính toán. Sau đó, một hoặc nhiều trình tự chuẩn hóa được nhận dạng theo các tiêu chuẩn khác nhau như được mô tả hơn nữa dưới đây, xem bước 145. Theo một số phương án, ví dụ, trình tự chuẩn hóa nhận dạng được dẫn đến độ biến thiên nhỏ nhất về định lượng trình tự đối với trình tự quan tâm qua tất cả các mẫu đạt yêu cầu.

Trong bước 146, dựa vào mật độ thể định tính tính được, định lượng trình tự định tính, tức là, định lượng nhiễm sắc thể hoặc định lượng đoạn, đối với trình tự quan tâm được xác định là tỷ lệ của độ che phủ của thể trình tự đối với trình tự quan tâm và độ che phủ của thể trình tự định tính đối với các trình tự bổ sung mà từ đó các trình tự chuẩn hóa được nhận dạng tiếp trong bước 145. Trình tự chuẩn hóa được nhận dạng được dùng để nhận dạng tiếp các định lượng trình tự trong mẫu thử nghiệm.

Theo một phương án, định lượng trình tự trong mẫu đạt yêu cầu là định lượng nhiễm sắc thể được tính là tỷ lệ của số lượng thể trình tự hoặc tham số kích thước đoạn đối với nhiễm sắc thể quan tâm và số lượng thể trình tự đối với trình tự nhiễm sắc thể chuẩn hóa trong mẫu đạt yêu cầu. Trình tự nhiễm sắc thể chuẩn hóa có thể là một nhiễm sắc thể đơn, nhóm các nhiễm sắc thể, một đoạn của một nhiễm sắc thể, hoặc một nhóm các đoạn từ các nhiễm sắc thể khác nhau. Do đó, định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm được xác định trong mẫu đạt yêu cầu là tỷ lệ giữa số lượng thể đối với nhiễm sắc thể quan tâm và số lượng thể đối với (i) trình tự nhiễm sắc thể chuẩn hóa bao gồm nhiễm sắc thể đơn, (ii) trình tự nhiễm sắc thể chuẩn hóa bao gồm hai hoặc nhiều nhiễm sắc thể, (iii) trình tự đoạn chuẩn hóa bao gồm một đoạn của nhiễm sắc thể, (iv) trình tự đoạn chuẩn hóa bao gồm hai hoặc nhiều đoạn tạo ra một nhiễm sắc thể, hoặc (v) trình tự đoạn chuẩn hóa bao gồm hai hoặc nhiều

đoạn của hai hoặc nhiều nhiễm sắc thể. Ví dụ để xác định định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm 21 theo (i)-(v) là như sau: định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm, ví dụ, nhiễm sắc thể 21, được xác định là tỷ lệ của độ che phủ của thể trình tự của nhiễm sắc thể 21 và một trong các độ che phủ của thể trình tự sau đây: (i) mỗi trong số tất cả các nhiễm sắc thể còn lại, tức là, các nhiễm sắc thể 1-20, nhiễm sắc thể 22, nhiễm sắc thể X, và nhiễm sắc thể Y; (ii) tất cả các tổ hợp có thể có của hai hoặc nhiều nhiễm sắc thể còn lại; (iii) một đoạn của nhiễm sắc thể khác, ví dụ, nhiễm sắc thể 9; (iv) hai đoạn của một nhiễm sắc thể khác, ví dụ, hai đoạn của nhiễm sắc thể 9; (v) hai đoạn của hai nhiễm sắc thể khác, ví dụ, một đoạn của nhiễm sắc thể 9 và một đoạn của nhiễm sắc thể 14.

Theo một phương án khác, định lượng trình tự trong mẫu đạt yêu cầu là định lượng đoạn thay vì định lượng nhiễm sắc thể, định lượng đoạn này được tính dưới dạng tỷ lệ giữa số lượng thể trình tự đối với đoạn đích, mà không phải toàn bộ nhiễm sắc thể, và số lượng thể trình tự đối với trình tự đoạn chuẩn hóa trong mẫu đạt yêu cầu. Trình tự đoạn chuẩn hóa có thể là trình tự bất kỳ trong số các trình tự nhiễm sắc thể hoặc đoạn chuẩn hóa được bàn luận trên đây.

Nhận diện các trình tự chuẩn hóa

Ở bước 145, một trình tự chuẩn hóa được nhận diện cho một trình tự quan tâm. Theo một số phương án, trình tự chuẩn hóa là trình tự dựa trên các định lượng trình tự được tính toán, ví dụ, mà dẫn đến độ biến đổi nhỏ nhất về định lượng trình tự đối với trình tự quan tâm trong số tất cả các mẫu huấn luyện đủ tiêu chuẩn. Phương pháp nhận diện các trình tự mà vốn đã có các đặc tính giống nhau và dễ có các biến thiên tương tự giữa các mẫu và các lần chạy giải trình tự và là hữu ích để xác định định lượng trình tự trong các mẫu thử nghiệm.

Trình tự chuẩn hóa cho một hoặc nhiều trình tự quan tâm có thể được nhận diện trong tập mẫu đạt yêu cầu, và các trình tự này có thể được nhận diện trong các mẫu đạt yêu cầu sau đó được sử dụng để tính toán định lượng trình tự cho một hoặc nhiều trình tự quan tâm trong mỗi mẫu trong số các mẫu thử nghiệm (bước 150) để xác định sự có mặt hoặc không có mặt của thể lệnh bội lẻ trong mỗi mẫu trong số các mẫu thử nghiệm. Trình tự chuẩn hóa được nhận diện cho các nhiễm sắc thể hoặc các đoạn nhiễm sắc thể quan tâm có thể khác nhau khi các kỹ thuật giải trình tự khác nhau được sử dụng và/hoặc khi có sự khác nhau trong quá trình

trình sạch axit nucleic mà được giải trình tự và/hoặc sự khác nhau ở việc tạo thư viện giải trình tự. Việc sử dụng các trình tự chuẩn hóa theo phương pháp được mô tả trong bản mô tả này cung cấp phép đo đặc hiệu và nhạy đối với biến thể về số lượng bản sao của một nhiễm sắc thể hoặc đoạn của nó bất mầu nào và/hoặc kỹ thuật giải trình tự nào được sử dụng.

Theo một số phương án, nhiều hơn một trình tự chuẩn hóa được nhận diện, tức là, các trình tự chuẩn hóa khác nhau có thể được xác định cho một trình tự quan tâm và nhiều định lượng trình tự có thể được xác định cho một trình tự quan tâm. Ví dụ, độ biến thiên, ví dụ, hệ số biến thiên ($CV = \text{độ lệch chuẩn} / \text{trung bình}$), về định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm 21 là nhỏ nhất khi độ che phủ trình tự của nhiễm sắc thể 14 được sử dụng. Tuy nhiên, hai, ba, bốn, năm, sáu, bảy, tám hoặc nhiều trình tự chuẩn hóa hơn có thể được nhận diện để sử dụng trong việc xác định lượng trình tự cho trình tự quan tâm trong mẫu thử nghiệm. Ví dụ, định lượng thứ hai đối với nhiễm sắc thể 21 trong một mẫu thử nghiệm bất kỳ có thể được sử dụng nhờ sử dụng nhiễm sắc thể 7, nhiễm sắc thể 9, nhiễm sắc thể 11 hoặc nhiễm sắc thể 12 làm trình tự nhiễm sắc thể chuẩn hóa vì các nhiễm sắc thể này có CV gần với CV của nhiễm sắc thể 14.

Theo một số phương án, khi một nhiễm sắc thể đơn được chọn làm trình tự nhiễm sắc thể chuẩn hóa cho nhiễm sắc thể quan tâm, trình tự nhiễm sắc thể chuẩn hóa này là nhiễm sắc thể mà đem lại định lượng nhiễm sắc thể cho nhiễm sắc thể quan tâm có độ biến thiên nhỏ nhất ở tất cả các mẫu thử nghiệm, ví dụ, các mẫu đạt yêu cầu. Trong một số trường hợp, nhiễm sắc thể chuẩn hóa tốt nhất có thể không có độ biến thiên nhỏ nhất, nhưng có thể có phân bố định lượng đạt yêu cầu mà phân biệt tốt nhất mẫu hoặc các mẫu thử nghiệm với các mẫu đạt yêu cầu, tức là nhiễm sắc thể chuẩn hóa tốt nhất có thể không có độ biến thiên nhỏ nhất nhưng có khả năng phân biệt lớn nhất.

Theo một số phương án, trình tự chuẩn hóa gồm một hoặc nhiều trình tự hoặc nhiễm sắc thể thường khỏe mạnh hoặc đoạn của chúng. Theo một số phương án, nhiễm sắc thể thường khỏe mạnh gồm tất cả các nhiễm sắc thể thường ngoại trừ (các) nhiễm sắc thể quan tâm. Theo một số phương án, nhiễm sắc thể thường khỏe mạnh gồm tất cả các nhiễm sắc thể thường ngoại trừ các nhiễm sắc thể X, Y, 13, 18, và 21. Theo một số phương án, nhiễm sắc thể thường khỏe mạnh gồm tất cả nhiễm sắc thể thường ngoại trừ nhiễm sắc thể được xác định

từ mẫu là không thuộc trạng thái lưỡng bội thông thường, có thể là hữu ích để xác định bộ gen ung thư có số lượng bản sao bất thường so với bộ gen lưỡng bội thông thường.

Xác định các thể lệnh bội lẻ ở các mẫu thử nghiệm

Dựa vào sự nhận diện (các) trình tự chuẩn hóa ở các mẫu đạt yêu cầu, định lượng trình tự được xác định cho trình tự quan tâm ở mẫu thử nghiệm gồm hỗn hợp các axit nucleic nhận được từ các bộ gen khác nhau ở một hoặc nhiều trình tự quan tâm.

Ở bước 115, mẫu thử nghiệm thu được từ đối tượng bị nghi ngờ hoặc biết là mang CNV liên quan về mặt lâm sàng của trình tự quan tâm. Mẫu thử nghiệm có thể là chất dịch sinh học, ví dụ, huyết tương hoặc mẫu thích hợp bất kỳ được mô tả dưới đây. Như đã nêu, mẫu này có thể thu được bằng cách sử dụng quy trình không xâm lấn như lấy máu đơn giản. Theo một số phương án, mẫu thử nghiệm chứa hỗn hợp các phân tử axit nucleic, ví dụ phân tử cfADN. Theo một số phương án, mẫu thử nghiệm là mẫu huyết tương của người mẹ chứa hỗn hợp các phân tử cfADN bào thai và người mẹ.

Ở bước 125, ít nhất một phần các axit nucleic thử nghiệm trong mẫu thử nghiệm được giải trình tự như được mô tả đối với các mẫu đạt yêu cầu để tạo ra hàng triệu trình tự đọc, ví dụ trình tự đọc 36bp. Theo các phương án khác nhau, các trình tự đọc có đầu bắt cặp 2x36 bp được sử dụng để giải trình tự có đầu bắt cặp. Như ở bước 120, các đoạn đầu được tạo ra từ việc giải trình tự axit nucleic trong mẫu thử nghiệm được ánh xạ hoặc sắp thẳng duy nhất với bộ gen đối chứng để tạo ra các thể. Như được mô tả ở bước 120, ít nhất khoảng 3×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 5×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 8×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 10×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 15×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 20×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 30×10^6 thể trình tự đạt yêu cầu, ít nhất khoảng 40×10^6 thể trình tự đạt yêu cầu, hoặc ít nhất khoảng 50×10^6 thể trình tự đạt yêu cầu gồm các trình tự đọc từ 20 đến 40bp thu được từ các trình tự đọc ánh xạ duy nhất lên bộ gen đối chứng. Theo một số phương án, các trình tự đọc được tạo ra bởi thiết bị giải trình tự được cung cấp ở định dạng điện tử. Việc sắp thẳng được thực hiện bằng cách sử dụng thiết bị tính toán như được mô tả dưới đây. Các trình tự đọc riêng biệt được so sánh với bộ gen đối chứng, mà thường lớn (hàng triệu cặp bazơ) để nhận diện các vị trí mà các trình tự đọc duy nhất tương ứng với bộ gen đối chứng. Theo một số phương án, thu

tục sắp thẳng cho phép có sự bắt cặp sai giới hạn giữa các trình tự đọc và bộ gen đối chứng. Trong một số trường hợp, 1, 2, hoặc 3 cặp bazơ trong một trình tự đọc được phép bắt cặp sai với cặp bazơ tương ứng trong bộ gen đối chứng và việc ánh xạ vẫn được thực hiện.

Ở bước 135, tất cả hoặc phần lớn các thẻ thu được từ việc giải trình tự các axit nucleic ở các mẫu thử nghiệm được đếm để xác định độ che phủ thẻ trình tự thử nghiệm nhờ sử dụng thiết bị tính toán như được mô tả dưới đây. Theo một số phương án, mỗi trình tự đọc được sắp thẳng với một vùng cụ thể của bộ gen đối chứng (trong phần lớn các trường hợp là nhiễm sắc thể hoặc đoạn nhiễm sắc thể) và trình tự đọc được chuyển thành thẻ bằng cách gắn thông tin vị trí vào trình tự đọc. Vì quy trình này mở, nên thiết bị tính toán có thể vẫn số đếm liên tục số lượng thẻ/trình tự đọc ánh xạ đến mỗi vùng của bộ gen đối chứng (trong phần lớn các trường hợp là nhiễm sắc thể hoặc đoạn nhiễm sắc thể). Các số đếm được lưu trữ đối với mỗi nhiễm sắc thể hoặc đoạn nhiễm sắc thể quan tâm và mỗi nhiễm sắc thể hoặc đoạn nhiễm sắc thể chuẩn hóa tương ứng.

Theo một số phương án, bộ gen đối chứng có một hoặc nhiều vùng loại trừ là một phần của bộ gen sinh học thật mà không có trong bộ gen đối chứng. Các trình tự đọc có khả năng sắp thẳng với các vùng loại trừ này không được đếm. Ví dụ về vùng loại trừ gồm các vùng có trình tự lặp lại dài, các vùng có độ tương đồng giữa các nhiễm sắc thể X và Y, v.v.. Nhờ sử dụng trình tự đối chứng được che thu được nhờ kỹ thuật che nêu trên, chỉ các thẻ trên các đoạn không bị che của trình tự đối chứng được đếm để phân tích CNV.

Theo một số phương án, phương pháp này xác định xem có đếm thẻ nhiều hơn một lần khi các trình tự đọc sắp thẳng với cùng vị trí trên bộ gen hoặc trình tự đối chứng hay không. Có thể có trường hợp khi có hai thẻ có cùng trình tự và do vậy sắp thẳng với vị trí giống nhau trên trình tự đối chứng. Trong một số trường hợp, phương pháp sử dụng để đếm các thẻ có thể loại bỏ việc đếm các thẻ giống nhau có nguồn gốc từ cùng mẫu được giải trình tự. Nếu có số lượng không cân đối của các thẻ giống nhau trong mẫu xác định, gợi ý rằng có một sự lệch lớn hoặc thiết sót khác trong quy trình này. Do vậy, theo một số phương án, phương pháp đếm này không đếm các thẻ từ mẫu xác định mà giống với các thẻ từ mẫu này mà đã được đếm trước đó.

Các tiêu chí khác nhau có thể được thiết lập để chọn khi loại bỏ thẻ giống nhau ra khỏi một mẫu đơn. Theo một số phương án, tỷ lệ phần trăm xác định của các thẻ được đếm phải là duy nhất. Nếu nhiều thẻ hơn ngưỡng này không phải là duy nhất, chúng được loại bỏ. Ví dụ, tỷ lệ phần trăm xác định yêu cầu ít nhất 50% là duy nhất, các thẻ giống nhau không được đếm cho đến khi tỷ lệ của các thẻ duy nhất lớn hơn 50% đối với mẫu. Theo các phương án khác, số lượng thẻ duy nhất ngưỡng là ít nhất khoảng 60%. Theo các phương án khác, tỷ lệ phần trăm ngưỡng của các thẻ duy nhất là ít nhất khoảng 75%, hoặc ít nhất khoảng 90%, hoặc ít nhất khoảng 95%, hoặc ít nhất khoảng 98%, hoặc ít nhất khoảng 99%. Ngưỡng có thể được thiết lập là 90% đối với nhiễm sắc thể 21. Nếu 30 triệu thẻ được sắp thẳng với nhiễm sắc thể 21, thì ít nhất 27 triệu thẻ phải là duy nhất. Nếu 3 triệu thẻ được đếm không là duy nhất và 30 triệu và thẻ đầu tiên không phải là duy nhất, thì thẻ này không được đếm. Việc chọn ngưỡng cụ thể hoặc các tiêu chí khác được sử dụng để xác định khi nào không đếm các thẻ giống nhau khác có thể được chọn bằng cách sử dụng phép phân tích thống kê thích hợp. Một yếu tố ảnh hưởng đến ngưỡng này hoặc tiêu chí khác là lượng tương đối của mẫu được giải trình tự so với kích thước của bộ gen mà các thẻ có thể được sắp thẳng. Các yếu tố khác gồm kích thước của trình tự đọc và các yếu tố tương tự.

Theo một phương án, số lượng thẻ trình tự thử nghiệm được ánh xạ đến trình tự quan tâm được chuẩn hóa đến độ dài đã biết của trình tự quan tâm mà chúng được ánh xạ đến để tạo ra tỷ lệ mật độ thẻ trình tự thử nghiệm. Như được mô tả đối với mẫu đạt yêu cầu, việc chuẩn hóa đến độ dài đã biết của trình tự quan tâm là không cần thiết và có thể được đưa vào dưới dạng một bước để giảm số lượng chữ số trong một số đề đơn giản hóa số này để diễn dịch ở người. Vì tất cả các thẻ trình tự thử nghiệm được ánh xạ được đếm trong mẫu thử nghiệm, độ che phủ thẻ trình tự cho trình tự quan tâm, ví dụ, trình tự liên quan về mặt lâm sàng, trong các mẫu thử nghiệm được xác định, như là độ phủ thẻ trình tự đối với các trình tự khác tương ứng với ít nhất một trình tự chuẩn hóa được nhận diện trong các mẫu đạt yêu cầu.

Ở bước 150, dựa vào sự giống nhau của ít nhất một trình tự chuẩn hóa trong mẫu đạt yêu cầu, định lượng trình tự thử nghiệm được xác định đối với trình tự quan tâm trong mẫu thử nghiệm. Theo các phương án khác nhau, định lượng trình tự thử nghiệm được xác định bằng cách tính toán sử dụng các độ che phủ thẻ trình tự của trình tự quan tâm và trình tự chuẩn

hóa tương ứng như được mô tả trong bản mô tả này. Thiết bị tính toán chịu trách nhiệm cho việc này sẽ truy cập điện tử vào mối liên quan giữa trình tự quan tâm và trình tự chuẩn hóa liên quan của nó, có thể được lưu ở cơ sở dữ liệu, bảng, biểu đồ hoặc ở dạng mã trong các lệnh chương trình.

Như được mô tả ở phần khác của bản mô tả này, ít nhất một trình tự chuẩn hóa có thể là một trình tự hoặc nhóm các trình tự. Định lượng trình tự đối với trình tự quan tâm trong mẫu thử nghiệm là tỷ lệ của độ che phủ thể trình tự được xác định đối với trình tự quan tâm trong mẫu thử nghiệm và độ che phủ thể trình tự của ít nhất một trình tự chuẩn hóa được xác định trong mẫu thử nghiệm, trong đó trình tự chuẩn hóa trong mẫu thử nghiệm tương ứng với trình tự chuẩn hóa được nhận diện trong mẫu đạt yêu cầu đối với một trình tự quan tâm cụ thể. Ví dụ, nếu trình tự chuẩn hóa được nhận diện cho nhiễm sắc thể 21 trong mẫu đạt yêu cầu được xác định là nhiễm sắc thể, ví dụ, nhiễm sắc thể 14, thì định lượng trình tự thử nghiệm đối với nhiễm sắc thể 21 (trình tự quan tâm) được xác định là tỷ lệ của độ che phủ thể trình tự đối với nhiễm sắc thể 21 và độ che phủ thể trình tự đối với nhiễm sắc thể 14, mỗi độ che phủ này được xác định trong mẫu thử nghiệm. Tương tự, định lượng nhiễm sắc thể cho các nhiễm sắc thể 13, 18, X, Y, và các nhiễm sắc thể khác liên quan đến các thể lệch bội nhiễm sắc thể được xác định. Trình tự chuẩn hóa đối với nhiễm sắc thể quan tâm có thể là một hoặc một nhóm nhiễm sắc thể hoặc một hoặc một nhóm đoạn nhiễm sắc thể. Như nêu trên, trình tự quan tâm có thể là một phần của nhiễm sắc thể, ví dụ một đoạn nhiễm sắc thể. Do đó, định lượng trình tự đối với đoạn nhiễm sắc thể được xác định là tỷ lệ của độ che phủ thể trình tự được xác định đối với đoạn này trong mẫu thử nghiệm và độ che phủ thể trình tự đối với đoạn nhiễm sắc thể chuẩn hóa trong mẫu thử nghiệm, trong đó đoạn nhiễm sắc thể chuẩn hóa trong mẫu thử nghiệm tương ứng với đoạn nhiễm sắc thể chuẩn hóa (một đoạn hoặc nhóm các đoạn nhiễm sắc thể) được nhận diện trong mẫu đạt yêu cầu đối với một đoạn quan tâm cụ thể. Đoạn nhiễm sắc thể có thể có kích thước nằm trong khoảng từ vài kilobazơ (kb) đến megabazơ (Mb) (ví dụ, khoảng từ 1kb đến 10 kb, hoặc khoảng từ 10 kb đến 100 kb, hoặc từ 100kb đến 1Mb).

Ở bước 155, các giá trị ngưỡng được suy ra từ các giá trị độ lệch chuẩn được thiết lập đối với các định lượng trình tự đạt yêu cầu được xác định trong các mẫu đạt yêu cầu và định

lượng trình tự được xác định đối với các mẫu đã biết là thể lệnh bội lẻ đối với trình tự quan tâm. Lưu ý rằng phép toán này thường được thực hiện không đồng bộ với bước phân tích các mẫu thử nghiệm của người bệnh. Thao tác này có thể được thực hiện, ví dụ, đồng thời với bước chọn các trình tự chuẩn hóa từ mẫu đạt yêu cầu. Sự phân loại chính xác phụ thuộc vào các sai phân giữa các phân bố xác suất đối với các nhóm khác nhau, tức là loại thể lệnh bội lẻ. Trong một số ví dụ, các ngưỡng được chọn từ phân bố kinh nghiệm đối với mỗi loại thể lệnh bội lẻ, ví dụ thể ba nhiễm 21. Các trị số ngưỡng có thể có được thiết lập để phân loại các thể ba nhiễm 13, thể ba nhiễm 18, thể ba nhiễm 21, và thể lệch bội đơn nhiễm sắc thể X như được mô tả trong các ví dụ, mô tả việc sử dụng phương pháp xác định các thể lệnh bội lẻ nhiễm sắc thể bằng cách giải trình tự cfADN tách chiết từ mẫu của người mẹ gồm hỗn hợp các axit nucleic của bào thai và người mẹ. Trị số ngưỡng mà được xác định để phân biệt các mẫu có thể lệnh bội lẻ về nhiễm sắc thể có thể giống hoặc khác với ngưỡng đối với thể lệnh bội lẻ khác. Như được thể hiện trong các ví dụ, trị số ngưỡng đối với mỗi nhiễm sắc thể quan tâm được xác định từ độ biến thiên định lượng của nhiễm sắc thể quan tâm giữa các mẫu và các lần chạy giải trình tự. Định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm bất kỳ biến thiên càng ít, thì độ trải rộng về định lượng đối với nhiễm sắc thể quan tâm qua tất các mẫu không bị bệnh càng hẹp, mà các mẫu này được sử dụng để thiết lập ngưỡng để xác định các thể lệnh bội lẻ khác nhau.

Quay trở lại dòng quy trình liên quan đến việc phân loại mẫu thử nghiệm của người bệnh, ở bước 160, biến thể số lượng bản sao của trình tự quan tâm được xác định trong mẫu thử nghiệm bằng cách so sánh định lượng trình tự thử nghiệm đối với trình tự quan tâm với ít nhất một ngưỡng được thiết lập từ các định lượng trình tự đạt yêu cầu. Phép toán này có thể được thực hiện bởi cùng một thiết bị tính toán được sử dụng để xác định độ che phủ thể trình tự và/hoặc tính toán các định lượng đoạn.

Ở bước 160, định lượng tính toán đối với trình tự thử nghiệm quan tâm được so sánh với định lượng được thiết lập là các trị số ngưỡng được chọn theo "ngưỡng độ tin cậy" được xác định theo người dùng để phân loại mẫu là "bình thường", "bị bệnh" hoặc "không xác định." Các mẫu "không xác định" là các mẫu mà không thể thực hiện sự chẩn đoán cuối cùng có độ tin cậy đối với các mẫu này. Mỗi loại mẫu bị bệnh (ví dụ, thể ba nhiễm 21, thể ba nhiễm

một phần 21, thể đơn nhiễm sắc thể X) có các ngưỡng của nó, một cho các mẫu bình thường (không bị bệnh) và một cho các mẫu bị bệnh (mặc dù trong một số trường hợp hai ngưỡng này trùng nhau). Như được mô tả ở phần khác của bản mô tả, trong một số trường hợp, mẫu không xác định được chuyển thành mẫu xác định (bị bệnh hoặc bình thường) nếu tỷ lệ axit nucleic của bào thai trong mẫu thử nghiệm là đủ lớn. Việc phân loại trình tự thử nghiệm có thể được báo cáo bởi thiết bị tính toán được sử dụng trong các phép toán khác của quy trình này. Trong một số trường hợp, việc phân loại được báo cáo theo định dạng điện tử và có thể được hiển thị, gửi thư điện tử, được đánh thành văn bản, v.v. cho người quan tâm.

Theo một số phương án, việc xác định CNV gồm bước tính toán NCV hoặc NSV liên hệ định lượng nhiễm sắc thể hoặc đoạn nhiễm sắc thể với giá trị trung bình của định lượng nhiễm sắc thể hoặc đoạn nhiễm sắc thể tương ứng trong tập các mẫu đạt yêu cầu như nêu trên. Sau đó CNV có thể được xác định bằng cách so sánh NCV/NSV với trị số ngưỡng ước lượng số lượng bản sao đã xác định trước.

Ngưỡng ước lượng số lượng bản sao có thể được chọn để tối ưu tỷ lệ dương tính giả và âm tính giả. Ngưỡng ước lượng số lượng bản sao càng cao sự xuất hiện dương tính giả càng ít. Tương tự, ngưỡng này càng thấp, sự xuất hiện âm tính giả càng ít. Do vậy, có một sự thỏa hiệp giữa ngưỡng lý tưởng thứ nhất mà ở trên ngưỡng này chỉ có các dương tính thật được phân loại và ngưỡng lý tưởng thứ hai mà nằm dưới ngưỡng này chỉ có các âm tính thật được phân loại.

Các ngưỡng được thiết lập phần lớn tùy thuộc vào độ biến thiên về định lượng nhiễm sắc thể đối với một nhiễm sắc thể quan tâm cụ thể như được xác định trong tập hợp mẫu không bị bệnh. Độ biến thiên này tùy thuộc vào một số yếu tố, bao gồm tỷ lệ của cADN bào thai có mặt trong mẫu. Độ biến thiên (CV) được xác định bởi giá trị trung bình và độ lệch chuẩn đối với các định lượng nhiễm sắc thể ở quần thể các mẫu không bị bệnh. Do vậy, (các) ngưỡng để phân loại thể lệnh bộ lễ sử dụng các NCV, theo:

$$NCV_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j}$$

(trong đó $\hat{\mu}_j$ và $\hat{\sigma}_j$ lần lượt là giá trị trung bình và độ lệch chuẩn ước tính, đối với định lượng nhiễm sắc thể thứ j trong tập hợp mẫu đạt yêu cầu, và x_{ij} là định lượng nhiễm sắc thể thứ j quan sát được đối với mẫu thử nghiệm i).

với tỷ lệ ADN bào thai liên quan là:

$$FF_{ij} = 2 \times \left| \frac{NCV_{ij} \times \hat{\sigma}_j}{\hat{\mu}_j} \right| = 2 \times NCV \times CV$$

Do đó, đối với mỗi NCV của nhiễm sắc thể quan tâm, tỷ lệ ADN bào thai kỳ vọng liên quan đến một trị số NCV cho trước có thể được tính toán từ CV dựa trên giá trị trung bình và độ lệch chuẩn của tỷ lệ nhiễm sắc thể đối với nhiễm sắc thể quan tâm trên quần thể mẫu không bị bệnh.

Sau đó, dựa trên mối quan hệ giữa tỷ lệ ADN bào thai và các trị số NCV, biên quyết định có thể được chọn mà ở trên biên này các mẫu được xác định là dương tính (bị bệnh) dựa trên các điểm phân vị phân bố chuẩn. Như nêu trên, theo một số phương án, ngưỡng được thiết lập cho sự cân bằng tối ưu giữa việc phát hiện các dương tính thật và tỷ lệ các kết quả âm tính giả. Tức là, ngưỡng được chọn để tối đa hóa tổng các dương tính thật và âm tính thật hoặc tối thiểu hóa tổng các dương tính giả và âm tính giả.

Một số phương án đề xuất phương pháp chẩn đoán trước sinh về thể lệnh bội lẻ nhiễm sắc thể ở bào thai trong mẫu sinh học gồm các phân tử axit nucleic của bào thai và người mẹ. Sự chẩn đoán được thực hiện dựa trên việc thu nhận các thông tin giải trình tự từ ít nhất một phần của hỗn hợp các phân tử axit nucleic của bào thai và người mẹ thu được từ mẫu thử nghiệm sinh học, ví dụ, mẫu huyết tương của người mẹ, tính toán từ dữ liệu giải trình tự định lượng nhiễm sắc thể chuẩn hóa đối với một hoặc nhiều nhiễm sắc thể quan tâm, và/hoặc định lượng đoạn nhiễm sắc thể chuẩn hóa cho một hoặc nhiều đoạn nhiễm sắc thể quan tâm, và xác định sai phân có ý nghĩa về mặt thống kê giữa định lượng nhiễm sắc thể đối với nhiễm sắc thể quan tâm và/hoặc định lượng đoạn nhiễm sắc thể đối với đoạn quan tâm, tương ứng trong mẫu thử nghiệm và trị số ngưỡng được thiết lập ở các mẫu đạt yêu cầu (bình thường) và cung cấp phương pháp chẩn đoán trước sinh dựa trên sai phân thống kê này. Như được mô tả ở bước 160 của phương pháp này, sự chẩn đoán là bình thường hay mắc bệnh được thực

hiện. "Không xác định" được đưa ra trong trường hợp mà không thể thực hiện sự chẩn đoán là bình thường hay bị bệnh với độ tin cậy.

Theo một số phương án, hai ngưỡng có thể được chọn. Ngưỡng thứ nhất được chọn để tối thiểu hóa tỷ lệ dương tính giả, trên ngưỡng này các mẫu sẽ được phân loại là "bị bệnh", và ngưỡng thứ hai được chọn để tối thiểu hóa tỷ lệ âm tính giả, dưới ngưỡng này các mẫu sẽ được phân loại là "không bị bệnh". Các mẫu có NCV trên ngưỡng thứ hai nhưng dưới ngưỡng thứ nhất có thể được phân loại là mẫu "Bị nghi ngờ có thể lệnh bội lẻ" hoặc "không xác định", đối với các mẫu này sự có mặt hay không có mặt của thể lệnh bội lẻ có thể được xác nhận bằng các phương pháp độc lập. Vùng nằm giữa ngưỡng thứ nhất và ngưỡng thứ hai có thể được gọi là vùng "không xác định".

Theo một số phương án, các ngưỡng nghi ngờ và không xác định được thể hiện trong bảng 1. Như có thể thấy, các ngưỡng NCV thay đổi đối với các nhiễm sắc thể khác nhau. Theo một số phương án, các ngưỡng này thay đổi theo FF đối với mẫu như được mô tả trên đây. Các kỹ thuật ngưỡng được áp dụng trong bản mô tả này đem lại độ nhạy và độ chọn lọc được cải thiện theo một số phương án.

Bảng 1 Ngưỡng NCV nghi ngờ và mắc bệnh khoảng không xác định được đặt trong ngoặc

	Nghi ngờ	Mắc bệnh
Chr 13	3,5	4,0
Chr 18	3,5	4,5
Chr 21	3,5	4,0
Chr X (XO, XXX)	4,0	4,0
Chr Y (XX so với XY)	6,0	6,0

Phân tích kích thước đoạn và độ che phủ trình tự

Như nêu trên, các tham số kích thước đoạn cũng như độ che phủ có thể được sử dụng để đánh giá CNV. Kích thước của đoạn axit nucleic ngoài tế bào, ví dụ, đoạn cfADN có thể thu được bằng phương pháp giải trình tự đầu bắt cặp, điện di (ví dụ, điện di mao quản dựa trên microchip) và các phương pháp đã biết trong lĩnh vực này. Fig.2A thể hiện một cách có

chủ đề về việc giải trình tự đầu bắt cặp có thể được sử dụng như thế nào để xác định cả kích thước đoạn và độ che phủ trình tự.

Nửa trên của Fig.2A thể hiện sơ đồ của một đoạn ADN tự do của bào thai và một đoạn ADN tự do của người mẹ tạo ra một mẫu cho quy trình giải trình tự đầu bắt cặp. Thông thường, trình tự axit nucleic dài được chia đoạn thành các trình tự ngắn hơn để đọc được theo quy trình giải trình tự có đầu bắt cặp. Các đoạn này còn được gọi là các đoạn chèn. Việc chia đoạn là không cần thiết đối với ADN tự do do chúng đã tồn tại ở các đoạn phần lớn ngắn hơn 300 cặp bazơ. Được thể hiện rằng các đoạn ADN tự do của bào thai trong huyết tương của người mẹ dài hơn các đoạn ADN tự do của người mẹ. Như được thể hiện ở phần trên cùng của Fig.2, ADN tự do có nguồn gốc bào thai có độ dài trung bình khoảng 167 cặp bazơ, còn ADN tự do có nguồn gốc từ người mẹ có độ dài trung bình khoảng 175 cặp bazơ. Theo phương pháp giải trình tự đầu bắt cặp trên một số nền tảng, như giải trình tự Illumina bởi nền tảng tổng hợp như được mô tả dưới đây, trình tự trình tự nối, trình tự chỉ số và/hoặc trình tự đoạn mỗi được nối vào hai đầu của một đoạn (không được thể hiện trên Fig.2A). Đoạn là trình tự đọc thứ nhất theo một hướng, tạo ra trình tự đọc 1 từ một đầu của đoạn. Sau đó, trình tự đọc thứ hai bắt đầu từ đầu ngược lại của đoạn này, tạo ra trình tự đọc 2. Sự tương ứng giữa trình tự đọc 1 và trình tự đọc 2 có thể được nhận diện bởi các tọa độ của chúng trên cuvet dòng chảy. Sau đó trình tự đọc 1 và trình tự đọc 2 được ánh xạ đến trình tự đối chứng ở dạng một cặp thể mà gần nhau như được thể hiện ở nửa dưới của Fig.2. Theo một số phương án, nếu các trình tự đọc đủ dài, hai trình tự đọc có thể ở phần giữa của đoạn chèn. Sau khi cặp được sắp thẳng với trình tự đối chứng, khoảng cách tương đối giữa hai trình tự đọc và độ dài của đoạn này có thể được xác định từ các vị trí của hai trình tự đọc. Do các trình tự đọc có đầu được ghép cặp tạo ra số lượng cặp bazơ gấp hai lần các trình tự đọc một đầu có cùng độ dài trình tự đọc, chúng giúp cải thiện chất lượng sắp thẳng, đặc biệt đối với các trình tự có nhiều sự lặp lại hoặc trình tự không duy nhất. Theo nhiều phương án, trình tự đối chứng được chia thành các bin, như các bin 100K cặp bazơ. Sau khi các trình tự đọc đầu bắt cặp được sắp thẳng với trình tự đối chứng, số lượng các trình tự đọc được sắp thẳng với một bin có thể được xác định. Số lượng cũng như độ dài của đoạn chèn (ví dụ, các đoạn cfADN) cũng có thể được xác định cho một bin. Theo một số phương án, nếu một đoạn chèn lưỡng lự giữa hai bin, một nửa đoạn chèn có thể được phân bổ cho mỗi bin.

Fig.2B thể hiện một phương án đề xuất quy trình 220 sử dụng độ che phủ dựa trên kích thước để xác định biến thể số lượng bản sao của trình tự axit nucleic quan tâm trong mẫu thử nghiệm gồm các đoạn axit nucleic tự do có nguồn gốc từ hai hoặc nhiều bộ gen. Như được sử dụng ở đây, một tham số "lệch về phía kích thước hoặc khoảng kích thước của đoạn" khi: 1) tham số này được gán trọng số thuận đối với kích thước hoặc khoảng kích thước của đoạn, ví dụ, số đếm được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này hơn đối với kích thước hoặc khoảng kích thước khác; hoặc 2) tham số thu được từ giá trị mà được gán trọng số thuận đối với kích thước hoặc khoảng kích thước này của đoạn, ví dụ, tỷ số thu được từ số đếm được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này. Kích thước hoặc khoảng kích thước của đoạn có thể là một đặc tính của bộ gen hoặc một phần của nó khi bộ gen tạo ra các đoạn axit nucleic được làm giàu hoặc có sự tập trung kích thước hoặc khoảng kích thước cao hơn so với đoạn axit nucleic từ bộ gen khác hoặc phần khác của cùng một bộ gen.

Quy trình 220 bắt đầu bằng việc nhận các trình tự đọc trình tự thu được bằng cách giải trình tự đoạn axit nucleic tự do trong mẫu thử nghiệm. Xem khối 222. Hai hoặc nhiều bộ gen trong mẫu thử nghiệm có thể là bộ gen của người mẹ mang thai và bộ gen của bào thai mà người mẹ mang thai mang. Theo một số ứng dụng, mẫu thử nghiệm gồm ADN tự do từ các tế bào khối u và các tế bào không bị bệnh. Theo một số phương án, do tỷ lệ tín hiệu trên nhiễu cao được tạo ra bởi độ che phủ được làm lệch về kích thước, việc giải trình tự các đoạn axit nucleic tự do được thực hiện mà không cần khuếch đại các đoạn axit nucleic nhờ sử dụng PCR. Quy trình 200 còn bao gồm bước sắp thẳng các trình tự đọc của các đoạn axit nucleic tự do với bộ gen đối chứng gồm trình tự quan tâm và được chia thành các bin. Việc sắp thẳng thành công tạo ra các thẻ trình tự thử nghiệm, mà gồm trình tự và vị trí của nó trên trình tự đối chứng. Xem bước 224. Sau đó quy trình 220 tiếp tục bằng cách xác định kích thước của các đoạn axit nucleic tự do có trong mẫu thử nghiệm. Một số phương án áp dụng phương pháp giải trình tự đầu bắt cặp để cung cấp độ dài của đoạn chèn gắn với thẻ trình tự. Xem bước 226. Các thuật ngữ "kích thước" và "độ dài" được sử dụng hoán đổi cho nhau khi chúng được sử dụng liên quan đến các trình tự hoặc các đoạn axit nucleic. Theo phương án thể hiện trong bản mô tả này, quy trình 220 còn bao gồm bước gán trọng số các thẻ trình tự thử nghiệm dựa trên kích thước của các đoạn axit nucleic tự do mà từ đó thu được các thẻ Xem bước 228. Như

được sử dụng trong bản mô tả này, "gán trọng số" chỉ việc biến đổi một đại lượng nhờ sử dụng một hoặc nhiều biến số hoặc hàm. Một hoặc nhiều biến số hoặc hàm được gọi là "trọng số." Theo nhiều phương án, biến số được nhân với trọng số. Theo các phương án khác, biến số có thể được biến đổi lũy thừa hoặc theo cách khác. Theo một số phương án, bước gán trọng số cho các thẻ trình tự thử nghiệm được thực hiện bằng cách dịch chuyển độ che phủ về phía thẻ trình tự thử nghiệm thu được từ các đoạn axit nucleic tự do có kích thước hoặc đặc tính khoảng kích thước của một bộ gen trong mẫu thử nghiệm. Như nêu trên, kích thước là một đặc trưng của bộ gen khi bộ gen này có nồng độ axit nucleic có kích thước này được làm giàu hoặc cao so với bộ gen khác hoặc phần khác của cùng một bộ gen.

Theo một số phương án, hàm gán trọng số là hàm tuyến tính hoặc hàm không tuyến tính. Ví dụ về các hàm không tuyến tính có thể áp dụng được bao gồm, nhưng không chỉ giới hạn ở hàm bước Heaviside, hàm hình hộp (box-car), hàm bậc thang, hoặc hàm sigma. Theo một số phương án, hàm Heaviside hoặc hàm hình hộp được sử dụng, để một thẻ trong khoảng kích thước cụ thể nhân với trọng số là 1 và các thẻ ngoài khoảng này nhân với trọng số là 0. Theo một số phương án, các đoạn có kích thước trong khoảng từ 80 đến 150 cặp bazơ được gán trọng số là 1, còn các đoạn nằm ngoài khoảng này được gán trọng số là 0. Theo các ví dụ này, việc gán trọng số là thậm trọng, là 0 hay 1 tùy thuộc vào xem tham số của tất cả các trị số nằm trong hay ngoài khoảng cụ thể. Theo cách khác, các trọng số được tính toán ở dạng hàm liên tục của kích thước đoạn hoặc khía cạnh khác của trị số tham số liên quan.

Theo một số phương án, các trọng số đối với các đoạn trong một khoảng kích thước là dương và trong một khoảng kích thước khác là âm. Điều này có thể được sử dụng để giúp tăng cường tín hiệu khi các hướng của sai phân giữa hai bộ gen có dấu ngược nhau. Chẳng hạn, số đếm trình tự đọc được gán trọng số là 1 đối với đoạn chèn có kích thước từ 80 đến 150 cặp bazơ và trọng số là -1 đoạn chèn có kích thước từ 160 đến 200 cặp bazơ.

Các trọng số có thể được đưa ra cho các số đếm cũng như các tham số khác. Chẳng hạn, việc gán trọng số cũng có thể được áp dụng đối với các tham số phân số hoặc tỷ lệ mà sử dụng kích thước đoạn. Ví dụ, tỷ lệ có thể đưa cho các đoạn trong một số khoảng con nhất định trọng số lớn hơn các đoạn và các bin có kích thước khác.

Sau đó độ che phủ được tính toán đối với các bin trên cơ sở các thể trình tự thử nghiệm được gán trọng số. Xem khối 230. Các độ che phủ này được xem là được làm lệch về kích thước. Như được giải thích trên đây, trị số được làm lệch về phía kích thước đoạn hoặc khoảng kích thước đoạn khi tham số này được gán trọng số thuận đối với kích thước đoạn khoảng kích thước đoạn. Quy trình 200 còn bao gồm bước nhận diện biến thể số lượng bản sao ở trình tự quan tâm từ các độ che phủ tính toán được. Xem bước 232. Theo một số phương án, như được giải thích dưới đây liên quan đến các Fig.2C, Fig.3A-Fig.3K, và Fig.4, các độ che phủ có thể được điều chỉnh hoặc hiệu chỉnh để loại bỏ nhiễu trong dữ liệu, nhờ đó tăng tỷ lệ tín hiệu trên nhiễu. Trong một số ứng dụng, độ che phủ dựa trên các thể được gán trọng số thu được trong quy trình 220 đem đến độ nhạy cao và/hoặc độ chọn lọc cao so với các độ che phủ không được gán trọng số trong việc xác định biến thể số lượng bản sao. Trong một số ứng dụng, quy trình ví dụ được đưa ra dưới đây còn có thể cải thiện độ nhạy và độ chọn lọc đối với phân tích CNV.

Quy trình ví dụ để phân tích kích thước đoạn và/hoặc độ che phủ trình tự

Một số phương án bộc lộ đề xuất phương pháp xác định lượng che phủ trình tự có nhiều thấp và/hoặc tín hiệu cao, cung cấp dữ liệu để xác định các tình trạng về gen khác nhau liên quan đến số lượng bản sao và CNV với độ nhạy, độ chọn lọc và/hoặc hiệu quả được cải thiện so với lượng che phủ trình tự thu được bằng phương pháp thông thường. Theo một số phương án, các trình tự của mẫu thử nghiệm được xử lý để thu được các lượng che phủ trình tự.

Quy trình này có thể sử dụng một số thông tin sẵn có từ các nguồn khác. Theo một số phương án, tất cả các thông tin này thu được tập các mẫu huấn luyện đã biết là không bị bệnh (ví dụ, không có thể lệnh bội lẻ). Theo các phương án khác, một số hoặc tất cả thông tin thu được từ các mẫu thử nghiệm khác, mà có thể được cung cấp "trên chương trình chạy" nhiều mẫu được phân tích trong cùng một quy trình.

Theo một số phương án, các mặt nạ trình tự (sequence mask) được sử dụng để giảm nhiễu dữ liệu. Theo một số phương án, cả trình tự quan tâm và trình tự chuẩn hóa của nó đều được che. Theo một số phương án, có thể sử dụng các mặt nạ khác nhau khi các nhiễm sắc thể hoặc các đoạn nhiễm sắc thể quan tâm khác nhau được xem xét. Ví dụ, một mặt nạ (hoặc nhóm mặt nạ) có thể được sử dụng khi nhiễm sắc thể 13 là nhiễm sắc thể quan tâm và mặt nạ

(hoặc nhóm mặt nạ) khác có thể được sử dụng khi nhiễm sắc thể 21 là nhiễm sắc thể quan tâm. Theo một số phương án, mặt nạ được xác định ở độ phân giải của bin. Do vậy, trong một ví dụ, độ phân giải của mặt nạ là 100 kb. Theo một số phương án, một mặt nạ riêng biệt có thể được sử dụng cho nhiễm sắc thể Y. Các vùng loại trừ được che đối với nhiễm sắc thể Y có thể được tạo ra ở độ phân giải nhỏ (1kb) hơn so với các nhiễm sắc thể quan tâm khác, như được mô tả trong đơn Mỹ tạm thời số 61/836,057, nộp ngày 17/06/2013 [số hồ sơ ARTEP008P]. Các mặt nạ được tạo ra ở dạng các tập tin nhận diện các vùng gen được loại trừ.

Theo một số phương án, quy trình này sử dụng trị số kỳ vọng của độ che phủ chuẩn hóa để loại bỏ sự biến thiên bin với bin trong lược sử của trình tự quan tâm, sự biến thiên này không cung cấp thông tin để xác định CNV đối với mẫu thử nghiệm. Quy trình này điều chỉnh các đại lượng che phủ chuẩn hóa theo trị số kỳ vọng của độ che phủ chuẩn hóa đối với mỗi bin trên toàn bộ bộ gen hoặc ít nhất các bin của các nhiễm sắc thể khỏe mạnh ở bộ gen đối chứng (sử dụng trong phép toán 317 dưới đây). Các tham số ngoài độ che phủ có thể được cải thiện bởi quy trình này. Trị số kỳ vọng có thể được xác định từ tập huấn luyện gồm các mẫu không bị bệnh. Ví dụ, trị số kỳ vọng có thể là trị số trung bình của các tập các mẫu huấn luyện. Trị số độ che phủ được kỳ vọng của các mẫu có thể được xác định là số lượng các thể không dư duy nhất được căn chỉnh với một bin chia cho tổng số lượng các thể dư duy nhất được căn chỉnh với tất cả các bin ở các nhiễm sắc thể khỏe mạnh của bộ gen đối chứng.

Fig.2C thể hiện lưu đồ về quy trình 200 để xác định tham số kích thước đoạn đối với trình tự quan tâm, tham số được sử dụng để ước lượng số lượng bản sao của trình tự quan tâm trong mẫu thử nghiệm ở khối 214. Quy trình này loại bỏ biến thiên hệ thống thông thường ở các mẫu huấn luyện không bị bệnh, biến thiên làm tăng nhiễu trong phân tích để đánh giá CNV. Quy trình này cũng loại bỏ độ lệch GC riêng đối với mẫu thử nghiệm, nhờ đó tăng tỷ lệ tín hiệu trên nhiễu trong phân tích dữ liệu. Lưu ý rằng quy trình 200 cũng có thể được áp dụng cho độ che phủ, bất kể độ che phủ này có được làm lệch về kích thước hay không. Tương tự, các quy trình trên Fig.2D, Fig.3, và Fig.4 được áp dụng ngang bằng nhau đối với độ che phủ, độ che phủ được gán trọng số kích thước đoạn, phần hoặc tỷ lệ của các đoạn trong khoảng kích thước xác định, mức methyl hóa của các đoạn, v.v..

Quy trình 200 bắt đầu bằng việc tạo ra các trình tự đọc trình tự của mẫu thử nghiệm như được thể hiện ở bước 202. Theo một số phương án, các trình tự đọc trình tự thu được bằng cách giải trình tự các đoạn ADN thu được từ máu của phụ nữ mang thai gồm cfADN của người mẹ và bào thai. Quy trình thực hiện việc so sánh các trình tự đọc trình tự với bộ gen đối chứng gồm trình tự quan tâm, tạo ra các thẻ trình tự thử nghiệm. Bước 204. Theo một số phương án, các trình tự đọc được sắp thẳng với nhiều hơn một vị trí được loại trừ. Theo một số phương án, nhiều trình tự đọc được sắp thẳng với cùng một vị trí được loại trừ và giảm xuống thành một lần đếm trình tự đọc. Theo một số phương án, các trình tự đọc được sắp thẳng với các vị trí loại trừ cũng bị loại trừ. Do vậy, theo một số phương án, chỉ có các thẻ không dư, được sắp thẳng duy nhất được sắp thẳng với các vị trí không loại trừ được đếm để tạo ra số đếm vị trí không loại trừ (số đếm NES) để xác định độ che phủ hoặc các tham số khác của mỗi bin.

Sau đó, quy trình 200 đưa ra các kích thước của các đoạn axit nucleic tự do có trong mẫu thử nghiệm. Theo một số phương án sử dụng phương pháp giải trình tự đầu bắt cặp, kích thước/độ dài đoạn chèn có thể thu được từ các vị trí của một cặp trình tự đọc ở các đầu của đoạn chèn. Các kỹ thuật khác có thể được sử dụng để xác định kích thước đoạn. Xem bước 205. Sau đó, ở các bin của bộ gen đối chứng, gồm các bin ở trình tự quan tâm, quy trình 200 xác định các trị số của tham số kích thước đoạn lệch về kích thước đoạn đặc trưng của một trong số các bộ gen. Thuật ngữ "tham số kích thước đoạn" chỉ tham số mà có liên quan đến kích thước hoặc chiều dài của một đoạn hoặc một tập hợp gồm nhiều đoạn như các đoạn axit nucleic, ví dụ, các đoạn cfADN thu được từ dịch thể. Như được sử dụng ở đây, một tham số "lệch về phía kích thước hoặc khoảng kích thước của đoạn" nêu: 1) tham số này được gán trọng số thuận đối với kích thước hoặc khoảng kích thước của đoạn, ví dụ, số đếm được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này hơn đối với kích thước hoặc khoảng kích thước khác; hoặc 2) tham số thu được từ giá trị mà được gán trọng số thuận đối với kích thước hoặc khoảng kích thước này của đoạn, ví dụ, tỷ số thu được từ số đếm được gán trọng số có sức nặng hơn khi đi kèm với các đoạn có kích thước hoặc khoảng kích thước này. Kích thước hoặc khoảng kích thước của đoạn có thể là một đặc trưng của bộ gen hoặc một phần của nó khi bộ gen tạo ra các đoạn axit nucleic được

làm giàu hoặc có sự tập trung kích thước hoặc khoảng kích thước cao hơn so với đoạn axit nucleic từ bộ gen khác hoặc phần khác của cùng một bộ gen.

Theo một số phương án, tham số kích thước đoạn là số đếm được gán trọng số kích thước. Theo một số phương án đoạn được gán trọng số 1 trong một khoảng và 0 ngoài khoảng này. Theo các phương án khác, tham số kích thước đoạn là tỷ phần hoặc tỷ lệ của các đoạn trong một khoảng kích thước. Xem bước 206. Theo một số phương án, trị số của tham số kích thước đoạn (hoặc độ che phủ như nêu trên) của mỗi bin được chia cho trị số của tham số của trình tự chuẩn hóa trong cùng một mẫu thử nghiệm, tạo ra tham số chuẩn hóa.

Sau đó quy trình 200 đưa ra lược tả chung của trình tự quan tâm. Lược tả chung gồm trị số tham số kỳ vọng ở mỗi bin thu được từ tập huấn luyện gồm các mẫu huấn luyện không bị bệnh. Bước 208. Quy trình 200 loại bỏ biến thiên chung trong mẫu huấn luyện bằng cách điều chỉnh các trị số tham số chuẩn hóa của các thẻ trình tự thử nghiệm theo các trị số tham số kỳ vọng để thu được các trị số được hiệu chỉnh theo lược tả chung của tham số đối với trình tự quan tâm. Bước 210. Theo một số phương án, trị số kỳ vọng của tham số thu được từ tập huấn luyện được tạo ra ở bước 208 là giá trị trung bình của các mẫu huấn luyện. Theo một số phương án, phép toán 210 điều chỉnh trị số chuẩn hóa của tham số bằng cách lấy trị số chuẩn hóa của tham số trừ đi trị số kỳ vọng của tham số này. Theo các phương án khác, phép toán 210 chia trị số chuẩn hóa của tham số cho trị số kỳ vọng của tham số này của mỗi bin để tạo ra trị số hiệu chỉnh theo lược tả chung của tham số.

Ngoài hoặc thay vì hiệu chỉnh lược tả chung, quy trình 200 loại bỏ các độ lệch GC riêng đối với mẫu thử nghiệm bằng cách điều chỉnh trị số tham số. Như được thể hiện ở bước 212, quy trình này điều chỉnh trị số tham số hiệu chỉnh theo lược tả chung dựa trên mối quan hệ giữa mức hàm lượng GC và độ che phủ được hiệu chỉnh theo lược tả chung trong mẫu thử nghiệm, nhờ đó thu được trị số hiệu chỉnh mẫu-GC của tham số kích thước đoạn. Sau khi điều chỉnh biến thiên hệ thống chung ở các mẫu huấn luyện không bị bệnh và ở các độ lệch GC, quy trình này cung cấp trị số kích thước đoạn được hiệu chỉnh đối với lược tả chung và/hoặc phương sai GC, trị số này được sử dụng để đánh giá CNV của mẫu có độ nhạy và độ đặc hiệu. Theo một số phương án, trị số kích thước đoạn có thể được điều chỉnh bằng cách sử dụng phương pháp phân tích thành phần chính loại bỏ các thành phần của phương sai không liên

quan đến biến thể số lượng bản sao của trình tự quan tâm như còn được mô tả với sự tham chiếu đến bước 719 trên Fig.2F. Theo một số phương án, trị số kích thước đoạn có thể chính xác hóa bằng cách loại bỏ các bin ngoại lai trong một mẫu như được mô tả liên quan đến bước 321 trên Fig.3A.

Quy trình nhiều nhánh để xác định số lượng bản sao sử dụng nhiều tham số

Như được nhấn mạnh ở trên, quy trình được mô tả thích hợp để xác định CNV sử dụng nhiều tham số, nhưng không giới hạn ở độ che phủ, độ che phủ được gán trọng số kích thước đoạn, kích thước đoạn, phần hoặc tỷ lệ của các đoạn trong khoảng kích thước xác định, mức methyl hóa của đoạn, v.v.. Mỗi trong số các tham số này có thể được xử lý riêng để góp phần riêng biệt vào việc xác định CNV cuối cùng.

Theo một số phương án, các quy trình giống nhau có thể được áp dụng cho phân tích độ che phủ được gán trọng số kích thước và phân tích kích thước đoạn, cả hai là các tham số kích thước đoạn. Fig.2D thể hiện lưu đồ về hai nhánh chồng lấn của dòng quy trình 600, nhánh 1 cho độ che phủ có gán trọng số kích thước và nhánh 2 cho phân tích kích thước đoạn. Theo phương án khác không được thể hiện trong bản mô tả này, mức methyl hóa có thể được xử lý theo một nhánh khác. Hai nhánh này có thể gồm các phép toán so sánh được để thu được thông tin độ che phủ được điều chỉnh mà việc xác định CNV dựa trên thông tin này.

Phần nhánh đơn ban đầu của quy trình bắt đầu bằng việc thu dữ liệu trình tự, xem bước 602 và tiếp tục tính toán số đếm như nêu trên, xem bước 612. Tại thời điểm này, quy trình thể hiện chia thành hai nhánh như nêu trên. Quay trở lại phần ban đầu của quy trình, chuỗi phép toán biến đổi dữ liệu trình tự thành các trình tự đọc. Khi dữ liệu trình tự thu được từ việc giải trình tự ghép kênh, các trình tự đọc cũng được giải ghép kênh để nhận diện nguồn dữ liệu. Xem khối 604. Các trình tự đọc sau đó được sắp thẳng với trình tự đối chứng, trong đó các trình tự đọc đã sắp thẳng được cung cấp ở dạng thẻ trình tự. Xem bước 606. Sau đó các thẻ trình tự được lọc để thu được các vị trí không loại trừ, mà các vị trí này là các thẻ trình tự được ánh xạ rõ ràng, không được nhân đôi. Các thẻ trình tự được sắp xếp thành các bin có độ dài trình tự cụ thể như 1 kb, 100 kb hoặc 1Mb. Xem bước 610. Theo một số phương án liên quan đến việc phân tích của vùng riêng cho hội chứng, các bin có độ dài 100 kb. Theo một số phương án, các bin có độ biến thiên lớn có thể được che nhờ sử dụng mặt nạ trình tự thu được

từ các mẫu không bị bệnh theo cách thức như được mô tả trên Fig.3A, khối 313. Sau đó, các thể ở các NES được đếm để tạo ra độ che phủ được chuẩn hóa và điều chỉnh để phân tích CNV. Xem bước 612.

Theo phương án được thể hiện, các phép toán 604, 606, 610, và 612 được thực hiện một lần và phần lớn các phép toán còn lại được thực hiện hai lần, một lần để phân tích độ che phủ được gán trọng số kích thước (nhánh 1) và một lần để phân tích kích thước đoạn (nhánh 2). Theo các phương án khác, một hoặc nhiều phép toán được thể hiện là được thực hiện theo hai nhánh chỉ được thực hiện một lần và kết quả được dùng chung cho cả hai quá trình. Ví dụ về các phép toán dùng chung này gồm các phép toán 614, 616, và 618.

Theo các phương án được thể hiện, độ che phủ thu được (số đếm được gán trọng số kích thước) hoặc tham số kích thước đoạn (phần hoặc tỷ lệ kích thước) của NES được chuẩn hóa bằng cách, ví dụ, chia trị số NES của một bin cho tổng NES của bộ gen hoặc tập các nhiễm sắc thể chuẩn hóa. Theo một số phương án, chỉ có độ che phủ được chuẩn hóa, còn tham số kích thước đoạn không cần phải chuẩn hóa, do nó không chịu ảnh hưởng bởi độ sâu của việc giải trình tự giống như độ che phủ. Xem khối 614. Sau đó, theo một số phương án, phương sai chung cho tập huấn luyện gồm các mẫu không bị bệnh được loại bỏ, phương sai này không liên quan đến CNV quan tâm. Theo phương án được thể hiện, phương sai chung được biểu diễn ở dạng lược tả sóng chung thu được từ các mẫu không bị bệnh theo cách thức giống như lược tả sóng chung nêu trên. Theo một số phương án như được thể hiện trên Fig.6, các mẫu không bị bệnh được sử dụng để thu được lược tả sóng chung gồm các mẫu lấy từ cùng một cuvet dòng chảy hoặc mẻ xử lý. Xem khối 616. Việc tính toán cuvet dòng chảy riêng cho sóng chung còn được giải thích dưới đây. Theo phương pháp được thể hiện, sau khi lược tả sóng chung được loại bỏ, các độ che phủ được hiệu chỉnh đối với mức GC trên cơ sở riêng cho mẫu. Xem khối 616. Một số thuật toán để hiệu chỉnh GC được mô tả chi tiết dưới đây liên quan đến Fig.3A, khối 319.

Theo phương án được thể hiện, ở cả hai nhánh 1 phân tích độ che phủ được gán trọng số và nhánh 2 phân tích kích thước đoạn, dữ liệu có thể còn được lọc nhiễu riêng đối với một mẫu cụ thể, ví dụ, dữ liệu của các bin ngoại lệ mà có độ che phủ rất khác với các bin khác có thể được loại bỏ ra khỏi phân tích, sự khác nhau này không được cho là biến thể số lượng bản

sao quan tâm. Xem khối 622. Phép toán lọc trong mẫu này có thể tương ứng với khối 321 trên Fig.3A.

Theo một số phương án, sau khi lọc một mẫu, các trị số độ che phủ được gán trọng số của nhánh 1 và tham số kích thước đoạn của nhánh 2 được làm giàu tín hiệu đích hơn đối chứng. Xem các bước 624 và 628. Sau đó, mỗi độ che phủ và tham số kích thước đoạn đối với nhiễm sắc thể được sử dụng để tính toán định lượng nhiễm sắc thể và trị số nhiễm sắc thể chuẩn hóa (normalized chromosome value-NCV) như nêu trên. Sau đó NCV có thể được so sánh với tiêu chí để xác định điểm thể hiện xác suất của CNV. Xem các bước 626 và 630. Sau đó điểm từ hai nhánh có thể được kết hợp tạo ra điểm ghép cuối cùng, mà điểm này xác định có nên xác định là có thể lệnh bội lẻ hay không. Theo một số phương án, điểm của bước 626 và 630 là các số thống kê thử nghiệm t hoặc các trị số Z. Theo một số phương án, điểm cuối cùng là trị số chi bình phương. Theo các phương án khác, điểm cuối cùng là bình phương trung bình của hai trị số t hoặc điểm z. Các phương pháp khác để kết hợp hai điểm từ hai nhánh có thể được sử dụng để cải thiện độ nhạy và độ chọn lọc trong phát hiện CNV. Theo cách khác, có thể kết hợp hai điểm từ hai nhánh bằng các phép toán logic, ví dụ, phép toán VÀ hoặc phép toán HOẶC. Chẳng hạn, khi độ nhạy cao được ưu tiên để đảm bảo âm tính giả thấp, việc xác định CNV có thể được thực hiện khi điểm của nhánh 1 hoặc nhánh 2 đáp ứng tiêu chí xác định. Mặt khác, nếu độ nhạy cao là mong muốn để đảm bảo dương tính giả thấp, việc xác định CNV có thể được thực hiện chỉ khi điểm của nhánh 1 VÀ nhánh 2 đáp ứng tiêu chí xác định.

Lưu ý rằng có sự thỏa hiệp giữa độ nhạy và độ chọn lọc nhờ sử dụng phép toán logic nêu trên. Theo một số phương án, phương pháp giải trình tự hai bước được áp dụng để giải quyết sự thỏa hiệp này như được mô tả thêm dưới đây. Một cách vắn tắt, việc cho điểm ban đầu của một mẫu được so sánh với ngưỡng thứ nhất tương đối thấp được thiết kế để tăng độ nhạy và nếu điểm của mẫu cao hơn ngưỡng thứ nhất, mẫu này phải trải qua chu kỳ giải trình tự thứ hai, sâu hơn chu kỳ thứ nhất. Mẫu này sau đó được xử lý và phân tích lại theo quy trình giống với quy trình nêu trên. Sau đó điểm thu được được so sánh với ngưỡng thứ hai tương đối cao được thiết kế để cải thiện độ nhạy. Theo một số phương án, các mẫu trải qua chu kỳ

giải trình tự thứ hai có số điểm tương đối thấp trong số các mẫu mà có số điểm lớn hơn ngưỡng thứ nhất, bằng cách này giảm số lượng mẫu cần phải giải trình tự lại.

Theo một số phương án, nhánh thứ ba sử dụng tham số thứ ba có thể được sử dụng. Một ví dụ về nhánh thứ ba này là metyl hóa. Mức metyl hóa có thể được xác định trực tiếp bằng cách xác định mức metyl hóa của axit nucleic của mẫu hoặc gián tiếp ở dạng tham số tương quan với kích thước đoạn của axit nucleic tự do.

Theo một số phương án, tham số thứ ba này là độ che phủ hoặc tham số trên cơ sở số đếm thứ hai, trong đó số đếm là dựa trên kích thước đoạn ngoài kích thước đoạn chính được sử dụng trong tham số trên cơ sở số đếm thứ nhất. Khi các đoạn có kích thước từ 80 đến 150 cặp bazơ được sử dụng để tạo ra tham số số đếm hoặc độ che phủ, chúng loại khoảng 70% kết quả ra khỏi quá trình giải trình tự. Ở mức độ mà các trình tự đọc loại trừ vẫn có tín hiệu hữu ích tiềm năng nào đó, thì các trình tự đọc này có thể được sử dụng trong tham số thứ ba mà gồm các trình tự đọc loại trừ hoặc trình tự đọc ở phần được làm lệch về kích thước nằm ngoài hoặc chồng lấn phần phần được làm lệch về kích thước được sử dụng trong tham số thứ nhất. Về mặt này, các trình tự đọc và các trị số độ che phủ liên quan thu được từ các đoạn loại trừ có thể được gán trọng số thấp hơn. Nói cách khác, tham số biến thể số lượng bản sao được tính toán nhờ sử dụng các trình tự đọc này có thể được gán cho là ít quan trọng trong việc xác định biến thể số lượng bản sao cuối cùng. Theo cách khác, như nêu trên, các thẻ nằm ngoài khoảng kích thước trong tham số thứ nhất có thể có trị số âm khi hai bộ gen có đặc tính trái ngược nhau trong hai khoảng kích thước.

Theo các phương án khác nhau, các độ che phủ trong các quy trình 200, 220, và 600 được làm lệch về phía các thẻ của các đoạn ở giới hạn ngắn hơn của phổ kích thước đoạn. Theo một số phương án, độ che phủ được làm lệch về phía các thẻ của các đoạn có kích thước ngắn hơn trị số tính toán. Theo một số phương án, độ che phủ được làm lệch về phía thẻ của các đoạn trong khoảng kích thước đoạn và giới hạn trên của khoảng này là khoảng 150 cặp bazơ hoặc ít hơn.

Theo các phương án khác nhau của quy trình 200, 220, và 600, các trình tự đọc trình tự thu được bằng cách giải trình tự các đoạn axit nucleic tự do mà không đầu tiên sử dụng PCR để khuếch đại axit nucleic của các đoạn axit nucleic tự do. Theo các phương án khác nhau,

các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic tự do đến độ sâu không lớn hơn khoảng 6 triệu đoạn trên mỗi mẫu. Theo một số phương án, độ sâu giải trình tự không lớn hơn 1 triệu đoạn trên mỗi mẫu. Theo một số phương án, các trình tự đọc thu được bằng cách giải trình tự đa thành phần và số lượng các mẫu được ghép ít nhất là khoảng 24.

Theo các phương án khác nhau của quy trình 200, 220, và 600, mẫu thử nghiệm gồm huyết tương của một cá thể. Theo một số phương án, quy trình này còn bao gồm bước thu axit nucleic tự do từ mẫu thử nghiệm. Theo một số phương án, quy trình này còn bao gồm bước giải trình tự đoạn axit nucleic tự do có nguồn gốc từ hai hoặc nhiều bộ gen.

Theo các phương án khác nhau về quy trình 200, 220, và 600, hai hoặc nhiều bộ gen bao gồm các bộ gen của người mẹ và bào thai. Theo một số phương án, biến thể số lượng bản sao ở trình tự quan tâm có thể lệnh bội lẻ ở bộ gen của bào thai.

Theo các phương án khác nhau về quy trình 200, 220, và 600, hai hoặc nhiều bộ gen bao gồm các bộ gen từ tế bào ung thư và soma. Theo một số phương án, quy trình này bao gồm bước sử dụng biến thể số lượng bản sao trong bộ gen ung thư để chẩn đoán bệnh ung thư, giám sát sự tiến triển của bệnh ung thư và/hoặc xác định điều trị cho bệnh ung thư. Theo một số phương án, biến thể số lượng bản sao gây ra bất thường về gen.

Theo một số phương án về quy trình 200, 220, và 600, độ che phủ lệch về phía các thẻ của các đoạn ở giới hạn dài của của phổ kích thước đoạn. Theo một số phương án, độ che phủ lệch về phía các thẻ của các đoạn có kích thước dài hơn trị số quy định. Theo một số phương án, độ che phủ lệch về phía thẻ của các đoạn trong khoảng kích thước đoạn và trong đó giới hạn dưới của khoảng này là khoảng 150 cặp bazơ hoặc lớn hơn.

Theo một số phương án về quy trình 200, 220, và 600, quy trình này còn gồm: bước xác định ở các bin của bộ gen đối chứng, gồm trình tự quan tâm, mức metyl hóa của các đoạn axit nucleic tự do trong bin này và sử dụng mức metyl hóa, ngoài hoặc thay cho độ che phủ tính toán được hoặc các trị số tham số kích thước đoạn để nhận diện biến thể số lượng bản sao. Theo một số phương án, sử dụng mức metyl hóa để nhận diện biến thể số lượng bản sao bao gồm bước tạo ra lược sử metyl hóa chung đối với các bin của trình tự quan tâm. Lược sử metyl hóa chung gồm các mức metyl hóa kỳ vọng ở ít nhất một bin của trình tự quan tâm. Theo một số phương án, các mức metyl hóa kỳ vọng thu được từ độ dài của các đoạn axit

nucleic tự do trong tập huấn luyện gồm các mẫu huấn luyện không bị bệnh gồm các axit nucleic được giải trình tự và căn chỉnh theo cách thức về cơ bản giống với các đoạn axit nucleic của mẫu thử nghiệm, mức methyl hóa kỳ vọng có sự khác nhau giữa các bin. Theo một số phương án, quy trình này bao gồm bước điều chỉnh trị số mức methyl hóa nhờ sử dụng các mức methyl hóa kỳ vọng ở các bin của ít nhất một trình tự quan tâm, nhờ đó thu được các trị số mức methyl hóa hiệu chỉnh theo lược tả chung cho trình tự quan tâm, quy trình này còn bao gồm bước nhận diện biến thể số lượng bản sao bằng cách sử dụng độ che phủ được hiệu chỉnh theo lược tả chung và mức methyl hóa được hiệu chỉnh theo lược tả chung. Theo một số phương án, bước nhận diện biến thể số lượng bản sao sử dụng độ che phủ được hiệu chỉnh theo lược tả chung và mức methyl hóa được hiệu chỉnh theo lược tả chung còn bao gồm bước: điều chỉnh độ che phủ được hiệu chỉnh theo lược tả chung và mức methyl hóa được hiệu chỉnh theo lược tả chung dựa trên mức hàm lượng GC, bằng cách này thu được độ che phủ được hiệu chỉnh theo GC và các trị số mức methyl hóa được hiệu chỉnh theo GC cho trình tự quan tâm; và nhận diện biến thể số lượng bản sao sử dụng độ che phủ được hiệu chỉnh theo GC và mức methyl hóa được hiệu chỉnh theo GC.

Theo một số phương án về quy trình 200, 220, và 600, tham số kích thước đoạn bao gồm phân số hoặc tỷ số gồm phần các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm có các kích thước đoạn ngắn hơn hoặc dài hơn giá trị ngưỡng. theo một số phương án, tham số kích thước đoạn gồm phân số gồm (i) số đoạn trong mẫu thử nghiệm nằm trong khoảng kích thước thứ nhất gồm 110 cặp bazơ, và (ii) số đoạn trong mẫu thử nghiệm nằm trong khoảng kích thước thứ hai gồm khoảng kích thước thứ nhất và các kích thước ngoài khoảng kích thước thứ nhất.

Xác định số lượng bản sao sử dụng quy trình ba nhánh, tỷ số hợp lệ, thống kê t và/hoặc tỷ lệ ADN bào thai

Fig.2E thể hiện lưu đồ về quy trình ba nhánh để ước lượng số lượng bản sao. Lưu đồ này gồm ba nhánh quy trình chồng lẫn nhau 700, gồm nhánh 1 (hoặc 713A) phân tích độ che phủ của các trình tự đọc liên quan đến các đoạn có tất cả các kích thước, nhánh 2 (hoặc 713B) phân tích độ che phủ của các trình tự đọc liên quan đến các đoạn có kích thước ngắn hơn và nhánh 3 (hoặc 713C) phân tích về tần suất tương đối của các trình tự đọc ngắn so với tất cả các trình tự đọc.

Quy trình 700 tương tự quy trình 600 về cách bố trí chung. Các phép toán được thể hiện bởi các bước 702, 704, 706, 710, 712 có thể được thực hiện theo cách thức giống hoặc tương tự với các phép toán được thể hiện bởi các bước 602, 604, 606, và 610, và 612. Sau khi thu được số đếm trình tự đọc, độ che phủ được tính toán nhờ sử dụng các trình tự đọc của các đoạn có tất cả các kích thước theo nhánh 713A. Độ che phủ được xác định nhờ sử dụng các trình tự đọc của các đoạn ngắn theo nhánh 713B. Tần suất của các trình tự đọc của các đoạn ngắn so với toàn bộ các trình tự đọc được xác định ở nhánh 713C. Tần suất tương đối cũng được gọi là tỷ lệ kích thước hoặc phân số kích thước. Nó là một ví dụ về đặc trưng kích thước đoạn. Theo một số phương án, các đoạn ngắn là các đoạn có kích thước ngắn hơn khoảng 150 cặp bazơ. Theo các phương án khác nhau, các đoạn ngắn có thể có kích thước nằm trong khoảng 50-150, 80-150, hoặc 110-150 cặp bazơ. Theo một số phương án, nhánh thứ ba hoặc nhánh 713C, là tùy chọn.

Dữ liệu của ba nhánh 713A, 713B, và 713C đều trải qua các phép toán chuẩn hóa 714, 716, 718, 719, và 722 để loại bỏ phương sai không liên quan đến số lượng bản sao của trình tự quan tâm. Các phép toán chuẩn hóa này nằm trong các ô của bước 723. Phép toán 714 bao gồm chuẩn hóa đại lượng phân tích của trình tự quan tâm bằng cách chia đại lượng phân tích cho tổng trị số của đại lượng của trình tự đối chứng. Bước chuẩn hóa này có thể sử dụng các trị số thu được từ mẫu thử nghiệm. Tương tự, các phép toán 718 và 722 chuẩn hóa đại lượng phân tích sử dụng các trị số thu được từ mẫu thử nghiệm. Các phép toán 716 và 719 sử dụng các trị số thu được từ tập huấn luyện gồm các mẫu không bị bệnh.

Phép toán 716 loại bỏ phương sai của sóng chung thu được từ tập huấn luyện gồm các mẫu không bị bệnh, sử dụng phương pháp giống hoặc tương tự như được mô tả liên quan đến bước 616. Phép toán 718 loại bỏ phương sai của phương sai GC riêng đối với cá thể sử dụng phương pháp giống hoặc tương tự như được mô tả liên quan đến bước 618.

Phép toán 719 còn loại bỏ phương sai bằng cách sử dụng phương pháp phân tích thành phần chính (principal component analysis-PCA). Phương sai được loại bỏ bằng phương pháp PCA là do các yếu tố không liên quan đến số lượng bản sao của trình tự quan tâm. Đại lượng phân tích ở mỗi bin (độ che phủ, tỷ lệ kích thước đoạn, v.v) cung cấp biến số độc lập cho PCA, và các mẫu của tập huấn luyện không bị bệnh cung cấp các trị số cho các biến độc lập

này. Tất cả các mẫu của tập huấn luyện này gồm các mẫu có số lượng bản sao giống nhau của trình tự quan tâm, ví dụ, hai bản sao của nhiễm sắc thể sôma, một bản sao của nhiễm sắc thể X (khi các mẫu nam giới được sử dụng làm các mẫu không bị bệnh), hoặc hai bản sao của nhiễm sắc thể X (khi các mẫu nữ giới được sử dụng làm các mẫu không bị bệnh). Do vậy, phương sai ở các mẫu không do thể lệnh bội lẻ hoặc sai phân về số lượng bản sao. PCA của tập huấn luyện tạo ra các thành phần chính mà không liên quan đến số lượng bản sao của trình tự quan tâm. Các thành phần chính này sau đó có thể được sử dụng để loại bỏ biến số ở mẫu thử nghiệm không liên quan đến số lượng bản sao của trình tự quan tâm.

Theo một số phương án, phương sai của một hoặc nhiều thành phần chính được loại bỏ ra khỏi dữ liệu của mẫu thử nghiệm bằng cách sử dụng các hệ số được ước tính từ dữ liệu của mẫu không bị bệnh trong vùng nằm ngoài trình tự quan tâm. Theo một số phương án, vùng này là tất cả các nhiễm sắc thể khỏe mạnh. Chẳng hạn, PCA được thực hiện trên dữ liệu độ che phủ bin đã chuẩn hóa của các mẫu chuẩn huấn luyện, do đó tạo ra các thành phần chính tương ứng với kích thước trong đó phần lớn phương sai ở dữ liệu này có thể thu nạp. Phương sai thu nạp được không liên quan đến biến thể số lượng bản sao ở trình tự quan tâm. Sau khi các thành phần chính thu được từ các mẫu chuẩn huấn luyện, các thành phần này được áp dụng cho dữ liệu thử nghiệm. Mô hình hồi quy tuyến tính với mẫu thử nghiệm là biến số đáp lại và các thành phần chính là các biến số phụ thuộc được tạo ra ở các bin từ vùng nằm ngoài trình tự quan tâm. Các hệ số hồi quy thu được được sử dụng để chuẩn hóa độ che phủ bin của vùng quan tâm bằng cách trừ đi tổ hợp tuyến tính của các thành phần chính được xác định bởi các hệ số hồi quy ước tính. Việc này loại bỏ phương sai không liên quan đến CNV ra khỏi trình tự quan tâm. Xem khối 719. Dữ liệu dư được sử dụng cho phân tích xuôi. Ngoài ra, phép toán 722 loại bỏ các điểm dữ liệu ngoại lệ sử dụng các phương pháp được mô tả liên quan đến bước 622.

Sau khi thực hiện các phép toán chuẩn hóa ở bước 723, các trị số độ che phủ của tất cả các bin đã được "chuẩn hóa" để loại bỏ các nguồn biến thể ngoài thể lệnh bội lẻ hoặc biến thể số lượng bản sao khác. Về một ý nghĩa nào đó, các bin của trình tự quan tâm được làm giàu hoặc biến đổi tương đối so với các bin khác nhằm mục đích phát hiện biến thể số lượng bản sao. Xem bước 724, không phải là một phép toán nhưng thể hiện các trị số độ che phủ thu

được. Các phép toán chuẩn hóa ở bước lớn 723 có thể tăng tín hiệu và/hoặc giảm nhiễu của đại lượng phân tích. Tương tự, trị số độ che phủ của các đoạn ngắn đối với các bin được chuẩn hóa để loại bỏ nguồn biến thể ngoài thể lệnh bội lẻ hoặc biến thể số lượng bản sao khác như được thể hiện ở bước 728, và tần số tương đối của các đoạn ngắn (hoặc tỷ lệ kích thước) đối với các bin được chuẩn hóa theo cách tương tự để loại bỏ các nguồn biến thể ngoài thể lệnh bội lẻ hoặc biến thể số lượng bản sao khác như được thể hiện ở bước 732. Như đối với bước 724, các bước 728 và 732 không phải là các phép toán nhưng thể hiện độ che phủ và các trị số tần số tương đối sau bước lớn xử lý 723. Cần hiểu rằng các phép toán ở bước lớn 723 có thể được biến đổi, sắp xếp lại hoặc loại bỏ. Ví dụ, theo một số phương án, phép toán PCA 719 không được thực hiện. Theo các phương án khác, bước hiệu chỉnh cho phép toán GC 718 không được thực hiện. Theo các phương án khác, thứ tự của các phép toán có thể thay đổi; ví dụ, phép toán PCA 719 được thực hiện trước khi hiệu chỉnh cho phép toán GC 718.

Độ che phủ của tất cả các đoạn sau khi chuẩn hóa và loại bỏ phương sai như được thể hiện trong bước 724 được sử dụng để thu thống kê t ở bước 726. Tương tự, độ che phủ của các đoạn ngắn sau khi chuẩn hóa và loại bỏ phương sai ở bước 728 được sử dụng để thu được thống kê t ở bước 730, và tần số tương đối của các đoạn ngắn sau khi chuẩn hóa và loại bỏ phương sai được thể hiện ở bước 732 được sử dụng để thu được thống kê t ở bước 734.

Fig.2F chứng minh tại sao cần áp dụng thống kê t để phân tích số lượng bản sao có thể giúp cải thiện độ chính xác của phân tích. Fig.2F thể hiện, ở mỗi bảng, phân bố tần số của độ che phủ bin được chuẩn hóa của trình tự quan tâm và trình tự đối chứng, với phân bố của trình tự quan tâm chồng lấn và che phân bố của trình tự đối chứng. Ở bảng ở trên, độ che phủ bin đối với mẫu có độ che phủ cao được thể hiện, có trên 6 triệu trình tự đọc; ở bảng dưới, độ che phủ bin đối với mẫu có độ che phủ thấp được thể hiện, có ít hơn 2 triệu trình tự đọc. Trục nằm ngang thể hiện độ che phủ chuẩn hóa so với độ che phủ trung bình của trình tự đối chứng. Trục dọc thể hiện mật độ xác suất tương đối liên quan đến số lượng bin có trị số độ che phủ trung bình. Fig.2F là một loại thống kê đồ. Phân bố đối với trình tự quan tâm được thể hiện ở phía trước và phân bố của trình tự đối chứng được thể hiện phía sau. Trung bình đối với phân bố của trình tự quan tâm thấp hơn trung bình của trình tự đối chứng, thể hiện số lượng bản sao giảm ở mẫu. Độ lệch trung bình giữa trình tự quan tâm và trình tự đối chứng là tương tự

đối với mẫu có độ che phủ cao ở bảng trên và mẫu có độ che phủ thấp ở bảng dưới. Do vậy, độ lệch trung bình, theo một số phương án, có thể được sử dụng để nhận diện biến thể số lượng bản sao ở trình tự quan tâm. Lưu ý rằng các phân bố của mẫu có độ che phủ cao có các phương sai nhỏ hơn các phân bố của mẫu có độ che phủ thấp. Việc chỉ sử dụng trung bình để phân biệt hai phân bố không thu được sai phân giữa hai phân bố như việc sử dụng cả giá trị trung bình và phương sai. Thống kê t có thể phản ánh cả giá trị trung bình và phương sai của phân bố.

Theo một số phương án, phép toán 726 tính thống kê t như sau:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

trong đó $\bar{x}_1$ là độ che phủ bin của trình tự quan tâm, $\bar{x}_2$ là độ che phủ bin của vùng/trình tự đối chứng, s_1 là độ lệch chuẩn của độ che phủ của trình tự quan tâm, s_2 là độ lệch chuẩn của độ che phủ của vùng tham chiếu, n_1 là số bin của trình tự quan tâm; và n_2 là số bin cùng vùng tham chiếu.

Theo một số phương án, vùng tham chiếu gồm tất cả các nhiễm sắc thể khỏe mạnh (ví dụ, nhiễm sắc thể ngoài nhiễm sắc thể có nhiều khả năng có thể lệnh bội lẻ). Theo một số phương án, vùng tham chiếu gồm ít nhất một nhiễm sắc thể ngoài trình tự quan tâm. Theo một số phương án, vùng tham chiếu gồm các nhiễm sắc thể khỏe mạnh không gồm trình tự quan tâm. Theo các phương án khác, vùng tham chiếu gồm một tập nhiễm sắc thể (ví dụ, tập con các nhiễm sắc thể được chọn từ các nhiễm sắc thể khỏe mạnh) mà được xác định có khả năng dò tín hiệu tốt nhất cho tập các mẫu huấn luyện. Theo một số phương án, khả năng dò tín hiệu dựa trên khả năng của vùng tham chiếu phân biệt các bin chứa biến thể số lượng bản sao với các bin không chứa biến thể số lượng bản sao. Theo một số phương án, vùng tham chiếu được nhận diện theo cách thức tương tự như được sử dụng để xác định "nhiễm sắc thể chuẩn hóa" như được mô tả trong phần có tiêu đề "Nhận diện trình tự chuẩn hóa."

Quay trở lại Fig.2E, một hoặc nhiều ước tính tỷ số ADN bào thai (bước 735) có thể được kết hợp với thống kê bất kỳ trong số các thống kê t ở bước 726, 730 và 734 để thu được ước tính thích hợp cho mức bội thể. Xem khối 736. Theo một số phương án, một hoặc nhiều

tỷ lệ ADN bào thai của bước 740 thu được bởi quy trình bất kỳ trong số quy trình 800 trên Fig.2G, quy trình 900 trên Fig.2H, hoặc quy trình 1000 trên Fig.2I. Các quy trình này có thể được thực hiện song song nhờ sử dụng một quy trình như quy trình 1100 trên Fig.2J.

Fig.2G thể hiện quy trình ví dụ để xác định tỷ lệ ADN bào thai từ thông tin độ che phủ theo một số phương án thực hiện của sáng chế. Quy trình 800 bắt đầu bằng việc thu thông tin độ che phủ (ví dụ, các trị số định lượng trình tự) của các mẫu huấn luyện từ tập huấn luyện. Xem khối 802. Mỗi mẫu của tập huấn luyện thu được từ phụ nữ mang thai được biết là mang bào thai trai. Tức là, mẫu chứa cfADN của bào thai trai. Theo một số phương án, phép toán 802 có thể thu được độ che phủ trình tự được chuẩn hóa theo các cách khác với định lượng trình tự như được mô tả hoặc có thể thu được các trị số độ che phủ khác.

Sau đó quy trình 800 bao gồm bước tính các tỷ lệ ADN bào thai của mẫu huấn luyện. Theo một số phương án, tỷ lệ ADN bào thai có thể được tính từ trị số định lượng trình tự:

$$FF_j = -2 \times \frac{Rx_j - \text{median}(Rx_i)}{\text{median}(Rx_i)}$$

trong đó Rx_j là định lượng trình tự cho mẫu bé trai, trung bình (Rx_i) là trung bình của các định lượng trình tự đối với các mẫu bé gái. Theo các phương án khác, trung bình hoặc các số đo có xu hướng trung tâm khác có thể được sử dụng. Theo một số phương án, FF có thể thu được bằng các phương pháp khác, như tần số tương đối của nhiễm sắc thể X và Y. Xem khối 804.

Quy trình 800 còn bao gồm bước chia trình tự đối chứng thành nhiều bin trình tự con. Theo một số phương án, trình tự đối chứng là bộ gen hoàn chỉnh. Theo một số phương án, các bin là bin 100 kb. Theo một số phương án, bộ gen được chia thành 25.000 bin. Sau đó quy trình này tạo ra độ che phủ của các bin. Xem bước 806. Theo một số phương án, độ che phủ được sử dụng ở bước 806 thu được sau khi thực hiện các phép toán chuẩn hóa được thể hiện ở bước 1123 trên Fig.2J. Theo các phương án khác, độ che phủ của khoảng kích thước khác có thể được sử dụng.

Mỗi bin gắn với các độ che phủ của các mẫu trong tập huấn luyện. Do vậy, đối với mỗi bin có thể thu được một tương quan giữa độ che phủ của các mẫu và tỷ lệ ADN bào thai của các mẫu. Quy trình 800 gồm bước thu nhận các tương quan giữa tỷ lệ ADN bào thai và độ

che phủ đối với tất cả các bin. Xem khối 808. Sau đó quy trình này chọn các bin có trị số tương quan lớn hơn ngưỡng. Xem khối 810. Theo một số phương án, các bin có 6000 trị số tương quan cao nhất được chọn. Mục đích của việc nhận diện các bin là thể hiện sự tương quan cao giữa độ che phủ và tỷ lệ ADN bào thai trong các mẫu thử nghiệm. Sau đó các bin này có thể được sử dụng để dự đoán tỷ lệ ADN bào thai trong mẫu thử nghiệm. Mặc dù các mẫu huấn luyện là các mẫu bào thai trai, nhưng tương quan giữa tỷ lệ ADN bào thai và độ che phủ có thể được khái quát hóa cho các mẫu thử nghiệm trai và gái.

Bằng cách sử dụng các bin đã chọn có các trị số tương quan cao, quy trình thu được mô hình tuyến tính liên hệ tỷ lệ ADN bào thai với độ che phủ. Xem bước 812. Mỗi bin đã chọn cung cấp một biến số độc lập cho mô hình tuyến tính. Do đó, mô hình tuyến tính thu được cũng gồm tham số hoặc trọng số cho mỗi bin. Các trọng số của các bin được điều chỉnh để cho mô hình khớp với dữ liệu. Sau khi thu được mô hình tuyến tính, quy trình 800 bao gồm bước áp dụng dữ liệu độ che phủ trình tự đọc của mẫu thử nghiệm vào mô hình để xác định tỷ lệ ADN bào thai cho mẫu thử nghiệm. Xem khối 814. Dữ liệu độ che phủ được áp dụng của mẫu thử nghiệm là đối với các bin có tương quan cao giữa tỷ lệ ADN bào thai và độ che phủ.

Fig.2J thể hiện tiến trình 1100 để xử lý các trình tự đọc được mà thông tin của chúng có thể được sử dụng để thu được ước tính tỷ lệ ADN bào thai. Quy trình 1100 có các bước xử lý tương tự như quy trình 600 trên Fig.2D. Các bước 1102, 1104, 1106, 1110, 1112, 1123, 1114, 1116, 1118, và 1122 lần lượt tương ứng với các bước 602, 604, 606, 610, 612, 623, 614, 616, 618, và 622. Theo một số phương án, một hoặc nhiều phép toán chuẩn hóa ở bước 123 là tùy chọn. Nhánh 1 cung cấp thông tin độ che phủ, mà có thể được sử dụng ở bước 806 của quy trình 800 như được thể hiện trên Fig.2G. Sau đó quy trình 800 có thể tạo ra ước tính tỷ lệ ADN bào thai 1150 trên Fig.2J.

Theo một số phương án, các ước tính tỷ lệ ADN bào thai (ví dụ, 1150 và 1152 trên Fig.2J) có thể được kết hợp để tạo ra ước tính tỷ lệ ADN bào thai kết hợp (ví dụ, 1154). Các phương pháp khác nhau có thể được sử dụng để thu được các ước tính tỷ lệ ADN bào thai. Chẳng hạn, tỷ lệ ADN bào thai có thể thu được từ thông tin độ che phủ. Xem bước 1150 trên Fig.2J và quy trình 800 trên Fig.2G. Theo một số phương án, tỷ lệ ADN bào thai cũng có thể

được ước tính từ phân bố kích thước của các đoạn. Xem bước 1152 trên Fig.2J và quy trình 900 trên Fig.2H. Theo một số phương án, tỷ lệ ADN bào thai cũng có thể được ước tính từ phân bố tần số 8-mer. Xem bước 1152 trên Fig.2J và quy trình 1000 trên Fig.2I.

Ở mẫu thử nghiệm có cfADN của bào thai trai, tỷ lệ ADN bào thai cũng có thể được ước tính từ độ che phủ của nhiễm sắc thể Y và/hoặc nhiễm sắc thể X. Theo một số phương án, ước tính kết hợp của tỷ lệ ADN bào thai (xem, ví dụ, bước 1155) đối với bào thai trai giả định thu được nhờ sử dụng thông tin chọn từ nhóm gồm: tỷ lệ ADN bào thai thu được từ thông tin độ che phủ của bin, tỷ lệ ADN bào thai thu được từ thông tin kích thước đoạn, tỷ lệ ADN bào thai thu được từ độ che phủ của nhiễm sắc thể Y, tỷ lệ ADN bào thai thu được từ nhiễm sắc thể X, và tổ hợp bất kỳ của chúng. Theo một số phương án, giới tính giả định của bào thai thu được nhờ sử dụng độ che phủ của nhiễm sắc thể Y. Hai hoặc nhiều tỷ lệ ADN bào thai (ví dụ, 1150 và 1152) có thể kết hợp theo các cách khác nhau để cung cấp ước tính kết hợp của tỷ lệ ADN bào thai (ví dụ, 1155). Chẳng hạn, phương pháp trung bình hoặc trung bình được gán trọng số có thể được sử dụng theo một số phương án, trong đó việc gán trọng số có thể dựa trên độ tin cậy thống kê của ước tính tỷ lệ ADN bào thai.

Theo một số phương án, ước tính kết hợp về tỷ lệ ADN bào thai đối với bào thai gái giả định thu được nhờ sử dụng thông tin được chọn từ nhóm gồm: tỷ lệ ADN bào thai thu được từ thông tin độ che phủ của các bin, tỷ lệ ADN bào thai thu được từ thông tin kích thước đoạn và tổ hợp của chúng.

Fig.2H thể hiện quy trình xác định tỷ lệ ADN bào thai từ thông tin phân bố kích thước theo một số phương án. Quy trình 900 bắt đầu bằng việc thu thông tin độ che phủ (ví dụ, các trị số định lượng trình tự) của các mẫu huấn luyện trai từ tập huấn luyện. Xem khối 902. Quy trình 900 sau đó bao gồm bước tính tỷ lệ ADN bào thai của các mẫu huấn luyện sử dụng các phương pháp nêu trên liên quan đến bước 804. Xem khối 904.

Quy trình 900 thực hiện chia khoảng kích thước thành các bin tạo các bin dựa trên kích thước đoạn và xác định tần số của các trình tự đọc đối với các bin dựa trên kích thước đoạn. Xem khối 906. Theo một số phương án, tần số của các bin dựa trên kích thước đoạn thu được mà không cần chuẩn hóa đối với các yếu tố như được thể hiện ở bước 1123. Xem nhánh 1124 trên Fig.2J. Theo một số phương án, tần số của các bin dựa trên kích thước đoạn thu được sau

khi tùy ý thực hiện các phép toán chuẩn hóa được thể hiện ở bước 1123 trên Fig.2J. Theo một số phương án, khoảng kích thước IS được chia thành 40 bin. Theo một số phương án, bin ở giới hạn thấp gồm các đoạn có kích thước nhỏ hơn khoảng 55 cặp bazơ. Theo một số phương án, bin ở giới hạn thấp gồm các đoạn có kích thước nằm trong khoảng từ 50 đến 55 cặp bazơ, mà loại trừ thông tin đối với các trình tự đọc ngắn hơn 50 bp. Theo một số phương án, bin ở giới hạn cao gồm các đoạn có kích thước lớn hơn khoảng 245 cặp bazơ. Theo một số phương án, bin ở giới hạn cao gồm các đoạn có kích thước nằm trong khoảng từ 245 đến 250 cặp bazơ, mà loại trừ thông tin đối với các trình tự đọc dài hơn 250 bp.

Quy trình 900 tiếp tục bằng cách thu nhận mô hình tuyến tính liên hệ tỷ lệ ADN bảo thai với tần số trình tự đọc đối với các bin trên cơ sở kích thước đoạn, sử dụng dữ liệu của các mẫu huấn luyện. Xem khối 908. Mô hình tuyến tính thu được bao gồm các biến số độc lập cho tần số trình tự đọc của các bin trên cơ sở kích thước. Mô hình này cũng gồm tham số hoặc trọng số cho mỗi bin trên cơ sở kích thước. Các trọng số của các bin được điều chỉnh để cho mô hình khớp với dữ liệu. Sau khi thu được mô hình tuyến tính, quy trình 900 bao gồm bước áp dụng dữ liệu tần số trình tự đọc của mẫu thử nghiệm lên mô hình xác định tỷ lệ ADN bảo thai cho mẫu thử nghiệm. Xem khối 910.

Theo một số phương án, tần số 8-mer có thể được sử dụng để tính tỷ lệ ADN bảo thai. Fig.2I thể hiện quy trình ví dụ 1000 để xác định tỷ lệ ADN bảo thai từ thông tin tần số 8-mer theo một số phương án thực hiện của sáng chế. Quy trình 1000 bắt đầu bằng việc thu thông tin độ che phủ (ví dụ, các trị số định lượng trình tự) của các mẫu huấn luyện trải từ tập huấn luyện. Xem khối 1002. Quy trình 1000 sau đó bao gồm bước tính tỷ lệ ADN bảo thai của các mẫu huấn luyện sử dụng các phương pháp bất kỳ được mô tả cho bước 804. Xem khối 1004.

Quy trình 1000 còn gồm bước thu nhận các tần số 8-mer (ví dụ, tất cả các hoán vị có thể có của 4 nucleotit ở 8 vị trí) từ các trình tự đọc của mỗi mẫu huấn luyện. Xem khối 1006. Theo một số phương án, tối đa 65.536 hoặc gần số đó nhiều 8-mer và tần số của chúng thu được. Theo một số phương án, các tần số 8-mer thu được mà không cần chuẩn hóa đối với các yếu tố như được thể hiện ở bước 1123. Xem nhánh 1124 trên Fig.2J. Theo một số phương án, tần số 8-mer thu được sau khi tùy ý thực hiện các phép toán chuẩn hóa được thể hiện ở bước 1123 trên Fig.2J.

Mỗi 8-mer gắn với các tần số của các mẫu trong tập huấn luyện. Do vậy, đối với 8-mer có thể thu được một tương quan giữa tần số 8-mer của các mẫu và tỷ lệ ADN bào thai của các mẫu. Quy trình 1000 gồm bước thu nhận các tương quan giữa tỷ lệ ADN bào thai và tần số 8-mer đối với tất cả các 8-mer. Xem khối 1008. Sau đó quy trình này chọn các 8-mer có trị số tương quan lớn hơn ngưỡng. Xem khối 1010. Mục đích của việc nhận diện 8-mer là thể hiện sự tương quan cao giữa tần số 8-mer và tỷ lệ ADN bào thai trong các mẫu thử nghiệm. Sau đó các bin này có thể được sử dụng để dự đoán tỷ lệ ADN bào thai trong mẫu thử nghiệm. Mặc dù các mẫu huấn luyện là các mẫu bào thai trai, nhưng tương quan giữa tỷ lệ ADN bào thai và tần số 8-mer có thể được khái quát hóa cho các mẫu thử nghiệm trai và gái.

Bằng cách sử dụng các 8-mer đã chọn có các trị số tương quan cao, quy trình thu được mô hình tuyến tính liên hệ tỷ lệ ADN bào thai với tần số 8-mer. Xem khối 1012. Mỗi bin đã chọn cung cấp một biến số độc lập cho mô hình tuyến tính. Do đó, mô hình tuyến tính thu được cũng gồm tham số hoặc trọng số cho mỗi bin. Sau khi thu được mô hình tuyến tính, quy trình 1000 bao gồm bước áp dụng dữ liệu tần số 8-mer của mẫu thử nghiệm lên mô hình xác định tỷ lệ ADN bào thai cho mẫu thử nghiệm. Xem khối 1014.

Trở lại Fig.2E, theo một số phương án, quy trình 700 gồm bước thu nhận mức bội thể khả dĩ cuối cùng trong phép toán 736 nhờ sử dụng thống kê t dựa trên độ che phủ của tất cả các đoạn được cung cấp bởi phép toán 726, ước tính tỷ lệ ADN bào thai được đưa ra bởi phép toán 726, và thống kê t dựa trên độ che phủ của các đoạn ngắn được đưa ra bởi phép toán 730. Các phương án án kết hợp các kết quả từ nhánh 1 và nhánh 2 sử dụng mô hình chuẩn nhiều chiều. Theo một số phương án để đánh giá CNV, khả năng bội thể là khả năng thể lệnh bội lẻ, là một khả năng của mô hình giả định thể lệnh bội lẻ (ví dụ, trisomy hoặc monosomy) trừ đi khả năng của mô hình có giả định nguyên bội trong đó mô hình này sử dụng thống kê t dựa trên độ che phủ của tất cả các đoạn, ước tính tỷ lệ ADN bào thai và thống kê t dựa trên độ che phủ của các đoạn ngắn làm đầu vào và đưa ra khả năng làm đầu ra.

Theo một số phương án, khả năng của mức bội thể được thể hiện là tỷ số khả năng. Theo một số phương án, tỷ số khả năng được lập mô hình là:

$$LR = \frac{\sum_{ff_{total}} q(ff_{total}) * p_1(T_{short}, T_{all} | ff_{est})}{p_0(T_{short}, T_{all})}$$

trong đó p_1 là khả năng mà dữ liệu từ phân bố chuẩn nhiều chiều thể hiện mô hình 3 bản sao hoặc 1 bản sao, p_0 là khả năng mà dữ liệu đến từ phân bố chuẩn nhiều chiều thể hiện mô hình 2 bản sao, T_{short} , T_{all} là các điểm số T tính được từ độ che phủ nhiễm sắc thể được tạo ra từ các đoạn ngắn và tất cả các đoạn, và còn $q(ff_{total})$ là phân bố mật độ của tỷ lệ ADN bào thai (được ước tính từ dữ liệu huấn luyện) xem xét lỗi liên quan đến ước tính tỷ lệ ADN bào thai. Mô hình này kết hợp độ che phủ được tạo ra từ các đoạn ngắn với độ che phủ được tạo ra bởi tất cả các đoạn, điều này giúp cải thiện việc phân tách điểm của độ che phủ của các mẫu không bị bệnh và mắc bệnh. Theo phương án được thể hiện, mô hình này cũng sử dụng tỷ lệ ADN bào thai, nhờ đó cải thiện thêm khả năng phân biệt mẫu bị bệnh và mẫu không bị bệnh. Do vậy, tỷ số hợp lệ được tính toán nhờ sử dụng thống kê t dựa trên độ che phủ của tất cả các đoạn (726), thống kê t dựa trên độ che phủ của các đoạn ngắn (730), và ước tính tỷ lệ ADN bào thai do quy trình 800 (hoặc bước 726), 900, hoặc 1000 cung cấp như được mô tả ở trên. Theo một số phương án, tỷ số hợp lệ này được sử dụng để phân tích các nhiễm sắc thể 13, 18, và 21.

Theo một số phương án, khả năng mức bội thể thu được bởi phép toán 736 chỉ sử dụng thống kê t thu được dựa trên tần số tương đối của các đoạn ngắn thu được bởi phép toán 734 của nhánh 3 và ước tính tỷ lệ ADN bào thai được cung cấp bởi phép toán 726, quy trình 800, 900, hoặc 1000. Tỷ số hợp lệ có thể được tính toán theo phương trình sau:.

$$LR = \frac{\sum_{ff_{total}} q(ff_{total}) * p_1(T_{short_freq} | ff_{est})}{p_0(T_{short_freq})}$$

trong đó p_1 là khả năng mà dữ liệu từ phân bố chuẩn nhiều chiều thể hiện mô hình 3 bản sao hoặc 1 bản sao, p_0 là khả năng mà dữ liệu đến từ phân bố chuẩn nhiều chiều thể hiện mô hình 2 bản sao, $T_{short-freq}$ là điểm số T tính được từ tần số tương đối của các đoạn ngắn và $q(ff_{total})$ là phân bố mật độ của tỷ lệ ADN bào thai (được ước tính từ dữ liệu huấn luyện) xem xét lỗi liên quan đến ước tính tỷ lệ ADN bào thai. Tỷ số hợp lệ được tính bằng cách sử dụng thống kê t trên cơ sở tần số tương đối của các đoạn ngắn (734), và ước tính tỷ lệ ADN bào thai do quy trình 800 cung cấp (hoặc bước 726), 900, hoặc 1000 nêu trên. Theo một số phương án, tỷ số hợp lệ này được sử dụng để phân tích nhiễm sắc thể X.

Theo một số phương án, tỷ số hợp lẽ được tính toán nhờ sử dụng thống kê t dựa trên độ che phủ của tất cả các đoạn (726), thống kê t dựa trên độ che phủ của các đoạn ngắn (730), và tần số tương đối của các đoạn ngắn (734). Ngoài ra, tỷ lệ ADN bào thai thu được nêu trên có thể kết hợp với các thống kê t để tính tỷ số hợp lẽ. Bằng cách kết hợp thông tin từ nhánh bất kỳ trong số ba nhánh 713A, 713B, và 713C, khả năng phân biệt đánh giá mức bội thể có thể được cải thiện. Xem, ví dụ 2 và Fig.12. Theo một số phương án, các tổ hợp khác nhau có thể được sử dụng để thu được tỷ số hợp lẽ cho một nhiễm sắc thể, ví dụ thống kê t của tất cả ba nhánh, thống kê t của nhánh thứ nhất và thứ hai, tỷ lệ ADN bào thai và ba thống kê t, tỷ lệ ADN bào thai và một thống kê t. v.v.. Sau đó, tổ hợp tối ưu có thể được chọn dựa trên hiệu suất của các mô hình.

Theo một số phương án để đánh giá nhiễm sắc thể thường, tỷ số hợp lẽ được lập mô hình là khả năng của dữ liệu được lập mô hình thu được từ mẫu thể ba nhiễm hoặc thể đơn nhiễm so với khả năng của dữ liệu được lập mô hình thu được mẫu thể lưỡng bội. Tỷ số hợp lẽ này có thể được sử dụng để xác định thể ba nhiễm hoặc thể đơn của các nhiễm sắc thể thường theo một số phương án.

Theo một số phương án để đánh giá nhiễm sắc thể giới tính, tỷ số hợp lẽ đối với nhiễm sắc thể X và tỷ số hợp lẽ đối với thể ba nhiễm X được đánh giá. Hơn nữa, phép đo độ che phủ nhiễm sắc thể (ví dụ, CNV hoặc điểm số của độ che phủ z đối với nhiễm sắc thể X và điểm số đối với nhiễm sắc thể Y cũng có thể được đánh giá. Theo một số phương án, bốn trị số này được đánh giá nhờ sử dụng cây xác định để xác định số lượng bản sao của nhiễm sắc thể giới tính. Theo một số phương án, cây xác định cho phép xác định mức bội thể là XX, XY, X, XXY, XXX, hoặc XYY.

Theo một số phương án, tỷ số hợp lẽ được biến đổi thành tỷ số hợp lẽ logarit và tiêu chuẩn hoặc ngưỡng xác định thể lệnh bội lẻ hoặc biến thể số lượng bản sao có thể được thiết lập theo kinh nghiệm để đạt được độ nhạy và độ chọn lọc cụ thể. Chẳng hạn, tỷ số hợp lẽ logarit là 1,5 có thể được thiết lập để xác định thể ba nhiễm 13 hoặc thể ba nhiễm 18 dựa trên độ nhạy và độ chọn lọc của mô hình khi áp dụng lên mẫu huấn luyện. Ngoài ra, chẳng hạn, trị số tiêu chuẩn xác định đột biến là 3 có thể được thiết lập cho thể ba nhiễm của nhiễm sắc thể 21 trong một số ứng dụng.

Mô tả chi tiết về quá trình xác định độ che phủ trình tự làm ví dụ

Fig.3A thể hiện ví dụ về quy trình 301 để giảm nhiễu trong dữ liệu chuỗi từ mẫu thử nghiệm. Các hình vẽ từ Fig.3B đến Fig.3J thể hiện các phân tích dữ liệu ở các giai đoạn khác nhau của quy trình. Điều này đưa ra một ví dụ về dòng quy trình có thể được sử dụng trong quy trình nhiều bước như được mô tả trên Fig.2D.

Quy trình 301 được thể hiện trên Fig.3A sử dụng độ che phủ trình tự dựa vào số lượng các thể xác định trình tự để ước lượng số lượng bản sao. Tuy nhiên, tương tự như phần mô tả trên đây về quy trình 100 để xác định CNV dựa theo Fig.1, các biến số hoặc tham số khác, chẳng hạn như kích thước, tỷ lệ kích thước, và mức metyl hóa, có thể được sử dụng thay cho độ che phủ cho quy trình 400. Theo một số phương án thực hiện, hai hoặc nhiều biến số có thể không trải qua cùng một quy trình để thu được hai điểm số chỉ ra xác suất của CNV, như được thể hiện trên đây dựa theo Fig.2D. Khi đó hai điểm số này có thể có thể được kết hợp để xác định CNV. Hơn nữa, độ che phủ và các tham số khác có thể được gán trọng số dựa vào kích thước của các đoạn mà từ đó các thể được rút ra. Để dễ hiểu, chỉ có độ che phủ được đề cập trong quy trình 300, nhưng cần lưu ý rằng các tham số khác như kích thước, tỷ lệ kích thước, và mức metyl hóa, số đếm được gán trọng số theo kích thước, v.v. có thể được sử dụng thay cho độ che phủ.

Như được thể hiện trên Fig.3A, quy trình được mô tả bắt đầu với việc chiết cfADN từ một hoặc nhiều mẫu. Xem khối 303. Các quy trình và thiết bị chiết phù hợp được mô tả ở phần khác của bản mô tả. Trong một số phương án, quy trình chiết cfADN được mô tả trong đơn sáng chế Mỹ số 61/801,126 nộp ngày 15/03/2013 (đưa viện dẫn toàn bộ vào bản mô tả này). Theo một số phương án thực hiện, thiết bị xử lý cfADN từ nhiều mẫu đồng thời cung cấp các thư viện dồn kênh và dữ liệu trình tự. Xem các khối 305 và 307 trên Fig.3A. Theo một số phương án, thiết bị xử lý cfADN song song từ tám hoặc nhiều mẫu thử nghiệm. Như được mô tả ở phần khác của bản mô tả, hệ thống giải trình tự có thể xử lý cfADN đã chiết để tạo thư viện gồm các đoạn cfADN đã được tạo mã (ví dụ, tạo mã thanh). Bộ giải trình tự sẽ sắp xếp thư viện cfADN theo trình tự để tạo ra lượng đọc trình tự rất lớn. Mỗi lần tạo mã mẫu cho phép giải dồn kênh các đoạn đọc trong các mẫu đã được dồn kênh. Mỗi mẫu trong số tám hoặc nhiều mẫu có thể có hàng nghìn hoặc hàng triệu đoạn đọc. Quy trình có thể lọc các đoạn đọc trước các phép toán bổ sung trên Fig.3A. Theo một số phương án, lọc đoạn đọc là quy

trình lọc chất lượng vận hành bởi các chương trình phần mềm được thực hiện trong bộ giải trình tự để lọc ra các đoạn đọc có lỗi và chất lượng kém. Ví dụ, các chương trình phần mềm kiểm soát trình tự (Sequencing Control Software - SCS) Illumina và đánh giá chung về trình tự và biến thể lọc ra các đoạn đọc có lỗi và chất lượng kém bằng cách chuyển đổi dữ liệu hình ảnh gốc do các phản ứng trình tự tạo ra thành điểm cường độ, cuộc gọi cơ sở, sắp xếp điểm chất lượng, và định dạng bổ sung để cung cấp thông tin liên quan đến sinh học để phân tích xuôi dòng.

Sau khi bộ giải trình tự hoặc thiết bị khác tạo ra các đoạn đọc cho mẫu, một bộ phận của hệ thống sẽ sắp xếp bằng máy tính các đoạn đọc thành một hệ gen chuẩn. Xem khối 309. Việc sắp xếp được mô tả ở phần khác của bản mô tả. Việc sắp xếp này tạo ra các thẻ chứa các trình tự đọc với thông tin về vị trí được chú thích chỉ rõ các vị trí đặc biệt trên hệ gen chuẩn. Theo các phương án thực hiện cụ thể, hệ thống thực hiện sắp xếp đường thứ nhất mà không tính các đoạn đọc trùng nhau, tức là hai hoặc nhiều đoạn đọc có các trình tự giống nhau, và sau đó loại bỏ các đoạn đọc trùng nhau hoặc tính các đoạn đọc trùng nhau thành một đoạn đọc để tạo ra các thẻ xác định trình tự không trùng lặp. Theo các phương án thực hiện khác, hệ thống không loại bỏ các đoạn đọc trùng nhau. Theo một số phương án, quy trình không tính đến các đoạn đọc được điều sắp xếp đến nhiều vị trí trên hệ gen để tạo ra các thẻ sắp xếp duy nhất. Trong một số phương án, các thẻ xác định trình tự không dư được sắp xếp duy nhất mà được ánh xạ thành các vị trí không bị loại trừ (nonexcluded site - NES) là cơ sở để tạo ra số đếm vị trí không bị loại trừ (NES), cung cấp dữ liệu để ước tính độ che phủ.

Như đã được giải thích, các vị trí loại trừ là các vị trí trong các vùng của hệ gen chuẩn đã được loại trừ không đếm các thẻ xác định trình tự. Theo một số phương án, vị trí loại trừ được tìm thấy nằm trong các vùng của các nhiễm sắc thể chứa các trình tự lặp lại, ví dụ, đoạn trung tâm và đoạn cuối, và các vùng chung giữa các nhiễm sắc thể này với nhiều hơn một nhiễm sắc thể khác, ví dụ, các vùng có trên nhiễm sắc thể Y và cũng có trên nhiễm sắc thể X. Các vị trí không loại trừ (NES) là các vị trí không bị loại trừ trong một hệ gen chuẩn để đếm các thẻ xác định trình tự.

Tiếp theo, hệ thống chia các thẻ đã sắp xếp thành các bin trên một hệ gen chuẩn. Xem khối 311. Các bin được đặt dọc theo chiều dài của hệ gen chuẩn. Theo một số phương án, toàn bộ hệ gen chuẩn được chia thành các bin liên tiếp, các bin này có thể có kích thước xác

định bằng nhau (ví dụ, 100kb). Tùy chọn, các bin có thể có độ dài được xác định động, có thể dựa trên cơ sở từng mẫu. Độ sâu trình tự tác động đến sự lựa chọn kích thước bin tối ưu. Các bin có kích thước động có thể có kích thước được xác định bằng kích thước theo thư viện. Ví dụ, kích thước bin có thể được xác định là bằng độ dài trình tự yêu cầu để chứa trung bình 1000 thẻ.

Mỗi bin có một số thẻ từ mẫu cần xem xét. Số thẻ này phản ánh “độ che phủ” của trình tự được điều chỉnh, làm việc như một điểm bắt đầu lọc hoặc làm sạch mẫu dữ liệu để xác định chính xác biến thể số lượng bản sao trong mẫu. Fig.3A thể hiện các phép toán dọn dẹp trong các khối từ 313 đến 321.

Theo phương án được mô tả trên Fig.3A, quy trình áp dụng mặt nạ cho các bin của hệ gen chuẩn. Xem khối 313. Hệ thống có thể loại trừ không xem xét đến độ che phủ trong các bin đã được che đi trong một số hoặc tất cả các phép toán theo quy trình sau đây. Trong nhiều trường hợp, các giá trị độ che phủ từ các bin đã được che đi không được xét đến bất kỳ phép toán còn lại nào trên Fig.3A.

Theo các phương án thực hiện khác nhau, một hoặc nhiều mặt nạ được áp dụng để loại bỏ các bin cho các vùng của hệ gen được tìm ra là biểu hiện có độ biến thể cao giữa mẫu này và mẫu khác. Các mặt nạ này được cung cấp cho cả các nhiễm sắc thể quan tâm (ví dụ, chr13, 18, và 21) và các nhiễm sắc thể khác. Như đã được giải thích, nhiễm sắc thể quan tâm là nhiễm sắc thể cần xem xét do có thể chứa biến thể số lượng bản sao hoặc sai lệch khác.

Trong một số phương án thực hiện, các mặt nạ được nhận biết từ tập huấn luyện của các mẫu đạt tiêu chuẩn bằng cách sử dụng phương pháp sau đây. Đầu tiên, mỗi tập huấn luyện mẫu được xử lý và lọc theo các bước 315 đến 319 trên Fig.3A. Sau đó, đại lượng che phủ được chuẩn hóa và được hiệu chỉnh sẽ được dành cho mỗi bin và các số liệu thống kê chẳng hạn như độ lệch chuẩn, độ lệch tuyệt đối trung bình, và/hoặc hệ số biến đổi được tính cho mỗi bin. Các tổ hợp lọc khác nhau có thể được ước lượng cho mỗi nhiễm sắc thể quan tâm. Các tổ hợp lọc cung cấp một lớp lọc cho các bin của nhiễm sắc thể quan tâm và lớp lọc khác cho các bin của tất cả các nhiễm sắc thể còn lại.

Trong một số phương án thực hiện, việc lựa chọn nhiễm sắc thể chuẩn hóa (hoặc nhóm nhiễm sắc thể) được xem xét lại sau khi thu được các mặt nạ (ví dụ, bằng cách chọn các ngưỡng cắt nhiễm sắc thể quan tâm như được mô tả trên đây). Sau khi áp dụng mặt nạ trình

tự, quy trình chọn (các) nhiễm sắc thể chuẩn hóa có thể được thực hiện như được mô tả ở phần khác của bản mô tả này. Ví dụ, tất cả các tổ hợp nhiễm sắc thể có thể có được đánh giá dưới dạng các nhiễm sắc thể chuẩn hóa và xếp hạng theo khả năng phân biệt các mẫu bị bệnh và không bị bệnh. Quy trình này có thể (hoặc có thể không) tìm ra nhiễm sắc thể hoặc hoặc nhóm nhiễm sắc thể chuẩn hóa khác biệt tối ưu. Theo các phương án khác, nhiễm sắc thể chuẩn hóa là các nhiễm sắc thể tạo ra tính biến thiên nhỏ nhất trong định lượng trình tự đối với trình tự quan tâm qua tất cả các mẫu đạt tiêu chuẩn. Nếu một nhiễm sắc thể hoặc nhóm nhiễm sắc thể chuẩn hóa khác được nhận diện, quy trình tùy ý thực hiện việc nhận diện các bin để lọc được mô tả trên đây. Có thể (các) nhiễm sắc thể chuẩn hóa mới sẽ tạo ra các ngưỡng cắt mới.

Theo một số phương án, mặt nạ khác được áp dụng cho nhiễm sắc thể Y. Ví dụ về mặt nạ nhiễm sắc thể Y phù hợp được mô tả trong đơn sáng chế tạm thời của Mỹ số 61/836,057, nộp ngày 17/06/2013 [số hồ sơ ARTEP008P], được đề cập đến trong bản mô tả sáng chế này với mục đích tham khảo.

Sau khi hệ thống che theo cách tính toán các bin, nó sẽ chuẩn hóa theo cách tính toán các giá trị độ che phủ ở các bin không bị loại trừ bởi các mặt nạ. Xem khối 315. Theo một số phương án, hệ thống chuẩn hóa các giá trị độ che phủ mẫu thử nghiệm trong mỗi bin (ví dụ, số đếm NES trên mỗi bin) đối với phần lớn hoặc tất cả độ che phủ trong hệ gen chuẩn hoặc một phần của nó (ví dụ, độ che phủ trong các nhiễm sắc thể khỏe mạnh của hệ gen chuẩn). Trong một số trường hợp, hệ thống chuẩn hóa các giá trị độ che phủ mẫu thử nghiệm (trên mỗi bin) bằng cách lấy số đếm cho bin cần xem xét chia cho tổng số tất cả các vị trí không bị loại trừ canh thẳng theo tất cả các nhiễm sắc thể khỏe mạnh trong hệ gen chuẩn. Theo một số phương án, hệ thống chuẩn hóa các giá trị độ che phủ mẫu thử nghiệm (trên mỗi bin) bằng cách thực hiện hồi quy tuyến tính. Ví dụ, đầu tiên là hệ thống tính toán độ che phủ cho tập bin con trong các nhiễm sắc thể khỏe mạnh theo công thức $y_a = \text{intercept} + \text{slope} * \text{gwp}_a$, trong đó y_a là độ che phủ cho bin a, và gwp_a là biên dạng chung cho cùng bin đó. Sau đó hệ thống sẽ tính toán các độ che phủ được chuẩn hóa z_b theo công thức: $z_b = y_b / (\text{intercept} + \text{slope} * \text{gwp}_b) - 1$.

Như đã được giải thích ở trên, nhiễm sắc thể khỏe mạnh là nhiễm sắc thể không có khả năng bị lệch bội. Theo một số phương án, các nhiễm sắc thể khỏe mạnh là tất cả các nhiễm

sắc thể thường ngoài các nhiễm sắc thể 13, 18, và 21. Theo một số phương án, các nhiễm sắc thể khỏe mạnh là tất cả các nhiễm sắc thể thường ngoài các nhiễm sắc thể được xác định là lệch khỏi hệ gen lưỡng bội bình thường.

Giá trị đếm biến đổi của bin hoặc độ che phủ được gọi là "đại lượng che phủ được chuẩn hóa" cho việc xử lý sau. Việc chuẩn hóa được thực hiện bằng cách sử dụng thông tin riêng cho mỗi mẫu. Thông thường, không có thông tin nào từ tập huấn luyện được sử dụng. Việc chuẩn hóa cho phép các đại lượng che phủ từ các mẫu có các kích thước thư viện khác nhau (và do đó có số đoạn đọc và thể khác nhau) cần được xử lý trên cơ sở bình đẳng. Một số phép toán xử lý theo sau sử dụng các đại lượng che phủ có được từ các mẫu huấn luyện mà có thể được giải trình tự từ các thư viện lớn hoặc nhỏ hơn các thư viện đã dùng cho mẫu thử nghiệm cần xem xét. Không có sự chuẩn hóa dựa vào số lượng đoạn đọc được canh thẳng theo toàn bộ hệ gen chuẩn (hoặc ít nhất là các nhiễm sắc thể khỏe mạnh), phương pháp xử lý sử dụng các tham số thu được từ tập huấn luyện có thể không tin cậy hoặc không tổng quát hóa theo một số phương án thực hiện.

Fig.3B thể hiện độ che phủ qua các nhiễm sắc thể 21, 13, và 18 đối với nhiều mẫu. Một số mẫu được xử lý theo cách khác nhau. Kết quả là, có thể thấy sự biến đổi lớn giữa các mẫu khác tại bất kỳ vị trí gen cho trước. Chuẩn hóa giúp loại bỏ một số biến đổi giữa các mẫu. Phần bên trái của Fig.3C thể hiện các đại lượng che phủ được chuẩn hóa qua toàn bộ hệ gen.

Theo phương án trên Fig.3A, hệ thống loại bỏ hoặc làm giảm "biên dạng chung" ra khỏi các đại lượng che phủ được chuẩn hóa sinh ra từ phép toán 315. Xem khối 317. Phép toán này loại bỏ độ lệch hệ thống trong các đại lượng che phủ được chuẩn hóa sinh ra từ cấu trúc của hệ gen, quy trình tạo thư viện, và quy trình giải trình tự. Ngoài ra, phép toán này được thiết kế để hiệu chỉnh độ lệch tuyến tính hệ thống bất kỳ khác với biên dạng kỳ vọng trong mẫu cho trước bất kỳ.

Trong một số phương án thực hiện, việc loại bỏ biên dạng chung bao gồm bước lấy đại lượng che phủ được chuẩn hóa của mỗi bin chia cho giá trị kỳ vọng tương ứng của mỗi bin. Theo các phương án khác, việc loại bỏ biên dạng chung bao gồm bước lấy đại lượng che phủ được chuẩn hóa của mỗi bin trừ đi giá trị kỳ vọng của mỗi bin. Giá trị kỳ vọng có thể thu được từ tập huấn luyện gồm các mẫu không bị bệnh (hoặc các mẫu cái không bị bệnh đối với nhiễm sắc thể X). Các mẫu không bị bệnh là các mẫu từ các đối tượng riêng lẻ đã biết là không có

đột biến lệch bội đối với nhiễm sắc thể quan tâm. Theo một số phương án thực hiện, việc loại bỏ biên dạng chung bao gồm bước lấy đại lượng che phủ được chuẩn hóa của mỗi bin trừ đi giá trị kỳ vọng của mỗi bin (thu được từ tập huấn luyện). Theo một số phương án, quy trình sử dụng giá trị trung bình của các đại lượng che phủ được chuẩn hóa cho mỗi bin như được xác định bằng cách sử dụng tập huấn luyện. Nói cách khác, các giá trị trung bình là các giá trị kỳ vọng.

Trong một số phương án, việc loại bỏ biên dạng chung được thực hiện bằng cách sử dụng kỹ thuật hiệu chỉnh tuyến tính đối với sự phụ thuộc của mẫu độ che phủ trên biên dạng chung. Như đã được thể hiện, biên dạng chung là giá trị kỳ vọng đối với mỗi bin như được xác định từ tập huấn luyện (ví dụ giá trị trung bình đối với mỗi bin). Các phương án này có thể sử dụng mô hình tuyến tính mạnh thu được bằng cách làm cho các đại lượng che phủ được chuẩn hóa của mẫu thử nghiệm khớp với biên dạng chung trung bình thu được cho mỗi bin. Trong một số phương án, mô hình tuyến tính thu được bằng cách hồi quy các đại lượng che phủ được chuẩn hóa quan sát được của mẫu dựa vào biên dạng chung trung bình (hoặc giá trị kỳ vọng khác).

Mô hình tuyến tính dựa vào giả định rằng các đại lượng che phủ mẫu có quan hệ tuyến tính với các giá trị biên dạng chung, quan hệ tuyến tính này cần duy trì đối với cả các nhiễm sắc thể/vùng khỏe mạnh và trình tự quan tâm. Xem Fig.3D. Trong trường hợp này, việc hồi quy các đại lượng che phủ mẫu được chuẩn hóa trên các đại lượng che phủ kỳ vọng của biên dạng chung sẽ tạo ra một đường có độ dốc và hệ số chặn. Theo một số phương án, độ dốc và hệ số chặn của đường này được dùng để tính toán đại lượng che phủ "được dự đoán" từ giá trị biên dạng chung cho một bin. Trong một số phương án thực hiện, việc hiệu chỉnh theo biên dạng chung bao gồm bước tạo mô hình cho đại lượng che phủ được chuẩn hóa của mỗi bin bằng các đại lượng che phủ được dự đoán đối với bin. Theo một số phương án thực hiện, độ che phủ của các thẻ trình tự thử nghiệm được điều chỉnh bằng cách: (i) thu được mối quan hệ toán học giữa độ che phủ của các thẻ trình tự thử nghiệm so với độ che phủ kỳ vọng trong nhiều bin trong một hoặc nhiều nhiễm sắc thể hoặc vùng khỏe mạnh, và (ii) áp dụng mối quan hệ toán học này vào các bin ở trình tự quan tâm. Theo một số phương án thực hiện, các độ che phủ trong mẫu thử nghiệm được hiệu chỉnh biến đổi bằng cách sử dụng quan hệ tuyến tính giữa các giá trị độ che phủ kỳ vọng từ các mẫu huấn luyện không bị bệnh và các giá trị

độ che phủ cho mẫu thử nghiệm trong các nhiễm sắc thể khỏe mạnh hoặc các vùng khỏe mạnh khác của hệ gen. Sự điều chỉnh này tạo ra các độ che phủ được hiệu chỉnh theo biên dạng chung. Trong một số trường hợp, việc điều chỉnh này bao gồm bước thu được các độ che phủ cho mẫu thử nghiệm đối với tập còn gồm các bin trong các nhiễm sắc thể hoặc vùng khỏe mạnh như sau:

$$y_a = \text{intercept} + \text{slope} * \text{gwp}_a$$

trong đó y_a là độ che phủ của bin a cho mẫu thử nghiệm trong một hoặc nhiều nhiễm sắc thể hoặc vùng khỏe mạnh, và gwp_a là biên dạng chung cho bin a đối với các mẫu huấn luyện không bị bệnh. Khi đó quy trình sẽ tính toán độ che phủ được hiệu chỉnh theo biên dạng chung z_b cho trình tự hoặc vùng quan tâm theo công thức:

$$z_b = y_b / (\text{intercept} + \text{slope} * \text{gwp}_b) - 1$$

trong đó y_b là độ che phủ quan sát được của bin b cho mẫu thử nghiệm ở trình tự quan tâm (có thể ở bên ngoài nhiễm sắc thể hoặc vùng khỏe mạnh), và gwp_b là biên dạng chung cho bin b đối với các mẫu huấn luyện không bị bệnh. Mẫu số $(\text{intercept} + \text{slope} * \text{gwp}_b)$ là độ che phủ cho bin b được dự đoán là cần phải quan sát trong các mẫu thử nghiệm không bị bệnh dựa vào mối quan hệ được ước tính từ các vùng khỏe mạnh của hệ gen. Trong trường hợp trình tự quan tâm có biến thể số lượng bản sao, độ che phủ quan sát được và do đó giá trị độ che phủ được hiệu chỉnh theo biên dạng chung cho bin b sẽ lệch tương đối lớn so với độ che phủ của mẫu không bị bệnh. Ví dụ, độ che phủ được hiệu chỉnh z_b sẽ cân xứng với tỷ lệ ADN bào thai trong trường hợp của mẫu có thể ba nhiễm đối với các bin trên nhiễm sắc thể bị bệnh. Quy trình chuẩn hóa này trong mẫu bằng cách tính toán hệ số chặn và độ dốc trên các nhiễm sắc thể khỏe mạnh, và sau đó đánh giá xem vùng gen quan tâm lệch như thế nào với mối liên hệ (như được mô tả bằng độ dốc và hệ số chặn) mà có giá trị đối với các nhiễm sắc thể khỏe mạnh trong cùng một mẫu.

Độ dốc và hệ số chặn thu được từ đường như được thể hiện trên Fig.3D. Ví dụ về việc loại bỏ biên dạng chung được mô tả trên Fig.3C. Bảng bên trái thể hiện sự biến đổi cao giữa các bin ở các đại lượng che phủ được chuẩn hóa qua nhiều mẫu. Bảng bên phải cũng thể hiện các đại lượng che phủ được chuẩn hóa đó sau khi loại bỏ biên dạng chung như được mô tả trên đây.

Sau khi hệ thống loại bỏ hoặc làm giảm các biến đổi biên dạng chung tại khối 317, hệ thống hiệu chỉnh cho các biến đổi nội dung trong mẫu GC (guanin-xitozin). Xem khối 319. Mỗi bin có sự phân chia tỷ lệ của chính nó từ GC. Tỷ lệ được xác định bằng cách chia số lượng nucleotit G và C trong bin b cho tổng số nucleotit trong bin (ví dụ, 100.000). Một số bin sẽ có tỷ lệ GC lớn hơn các bin khác. Như được thể hiện trên Fig.3E và Fig.3F, các mẫu khác nhau thể hiện độ lệch GC khác nhau. Sự khác nhau này và việc hiệu chỉnh chúng sẽ được mô tả kỹ hơn dưới đây. Các hình vẽ từ Fig.3E đến Fig.3G thể hiện đại lượng che phủ được chuẩn hóa có biên dạng chung được hiệu chỉnh (trên mỗi bin) dưới dạng hàm của tỷ lệ GC (trên mỗi bin). Điều ngạc nhiên là các mẫu khác nhau thể hiện tính phụ thuộc GC khác nhau. Một số mẫu thể hiện tính phụ thuộc giảm đơn điệu (như trên Fig.3E), trong khi một số khác thể hiện tính phụ thuộc có hình dấu phẩy (như trên Fig.3F và Fig.3G). Vì các biên dạng này có thể là duy nhất đối với mỗi mẫu, việc hiệu chỉnh được mô tả ở bước này được thực hiện riêng biệt và duy nhất đối với mỗi mẫu.

Trong một số phương án, hệ thống tính toán sắp xếp các bin trên cơ sở tỷ lệ GC như được thể hiện trên các hình vẽ từ Fig.3E đến Fig.3G. Sau đó hệ thống sẽ hiệu chỉnh biên dạng chung đã hiệu chỉnh, đại lượng che phủ được chuẩn hóa của bin bằng cách sử dụng thông tin từ các bin khác có cùng nội dung GC. Việc hiệu chỉnh này được áp dụng cho mỗi bin không được che.

Trong một số quy trình, mỗi bin được hiệu chỉnh về nội dung GC theo cách sau. Hệ thống tính toán lựa chọn các bin có tỷ lệ GC giống với bin cần xem xét và sau đó xác định tham số hiệu chỉnh từ thông tin trong các bin đã chọn. Theo một số phương án, các bin mà có cùng tỷ lệ GC sẽ được lựa chọn bằng cách sử dụng giá trị loại bỏ tương đồng được xác định tùy ý. Trong một ví dụ, 2% trong số tất cả các bin được chọn. Các bin này là các bin nằm trong số 2% có nội dung GC tương đồng nhất với bin cần xem xét. Ví dụ, 1% các bin có nhiều nội dung GC hơn một chút và 1% có ít nội dung GC hơn một chút được chọn.

Sử dụng các bin được chọn, hệ thống tính toán xác định tham số hiệu chỉnh. Trong một ví dụ, tham số hiệu chỉnh là giá trị tiêu biểu của các đại lượng che phủ được chuẩn hóa (sau khi loại bỏ biên dạng chung) trong các bin được chọn. Các ví dụ về giá trị tiêu biểu này bao gồm số trung bình của các đại lượng che phủ được chuẩn hóa trong các bin được chọn. Hệ thống áp dụng tham số hiệu chỉnh đã tính cho bin cần xem xét cho đại lượng che phủ được

chuẩn hóa (sau khi loại bỏ biên dạng chung) cho bin cần xem xét. Theo một số phương án thực hiện, giá trị tiêu biểu (ví dụ, giá trị trung bình) được trừ đi từ đại lượng che phủ được chuẩn hóa của bin cần xem xét. Theo một số phương án, giá trị trung bình (hoặc giá trị tiêu biểu khác) của các đại lượng che phủ được chuẩn hóa được lựa chọn bằng cách chỉ sử dụng các đại lượng che phủ cho các nhiễm sắc thể thường khỏe mạnh (tất cả các nhiễm sắc thể thường ngoại trừ các nhiễm sắc thể 13, 18, và 21).

Trong một ví dụ sử dụng, ví dụ, các bin 100kb, mỗi bin sẽ có giá trị tỷ lệ GC duy nhất, và các bin được chia thành nhóm dựa vào nội dung tỷ lệ GC của nó. Ví dụ, các bin được chia thành 50 nhóm, trong đó các biên nhóm tương ứng với (0, 2, 4, 6, ..., và 100) các điểm phân vị của phân bố %GC. Đại lượng che phủ được chuẩn hóa trung bình được tính cho mỗi nhóm bin từ các nhiễm sắc thể khỏe mạnh thường ánh xạ đến cùng nhóm GC (trong mẫu), và khi đó giá trị trung bình được trừ đi từ các đại lượng che phủ được chuẩn hóa (đối với tất cả các bin trong toàn bộ hệ gen trong cùng nhóm GC). Điều này áp dụng hiệu chỉnh GC được ước tính từ các nhiễm sắc thể khỏe mạnh nằm trong mẫu cho trước bất kỳ cho các nhiễm sắc thể có thể bị bệnh trong cùng một mẫu. Ví dụ, tất cả các bin trên các nhiễm sắc thể khỏe mạnh có nội dung GC nằm giữa 0,338660 và 0,344720 được nhóm lại với nhau, tính số trung bình cho nhóm này và lấy độ che phủ được chuẩn hóa của các bin nằm trong dãy GC này trừ đi số trung bình đó, các bin có thể được tìm thấy ở mọi nơi trên hệ gen (ngoại trừ các nhiễm sắc thể 13, 18, 21, và X). Theo một số phương án, nhiễm sắc thể Y được loại trừ ra khỏi quy trình hiệu chỉnh GC này.

Fig.3G thể hiện việc áp dụng hiệu chỉnh GC sử dụng các đại lượng che phủ được chuẩn hóa trung bình làm tham số hiệu chỉnh như vừa được mô tả. Bảng bên trái thể hiện các đại lượng che phủ không được hiệu chỉnh so với biên dạng tỷ lệ GC. Như được thể hiện, biên dạng có dạng phi tuyến tính. Bảng bên phải thể hiện các đại lượng che phủ được hiệu chỉnh. Fig.3H thể hiện các độ che phủ được chuẩn hóa cho nhiều mẫu trước hiệu chỉnh tỷ lệ GC (bảng trái) và sau hiệu chỉnh tỷ lệ GC (bảng phải). Fig.3I thể hiện hệ số biến động (coefficient of variation - CV) của các độ che phủ được chuẩn hóa cho nhiều mẫu thử nghiệm trước hiệu chỉnh tỷ lệ GC (màu đỏ) và sau hiệu chỉnh tỷ lệ GC (màu xanh), trong đó hiệu chỉnh GC dẫn đến sự biến động tương đối nhỏ ở các độ che phủ được chuẩn hóa.

Quy trình trên đây là phương án thực hiện tương đối đơn giản của hiệu chỉnh GC. Các phương pháp thay thế để hiệu chỉnh độ lệch GC thực hiện hàm nối trục hoặc kỹ thuật khớp phi tuyến tính khác, có thể được áp dụng trong khoảng GC liên tục và không yêu cầu gộp các đại lượng che phủ bằng nội dung GC. Ví dụ về các kỹ thuật phù hợp bao gồm hiệu chỉnh Loess liên tục và hiệu chỉnh hàm nối trục nhẵn. Hàm điều chỉnh (fitting function) có thể rút ra từ đại lượng che phủ được chuẩn hóa từng bin so với nội dung GC cho mẫu cần xem xét. Việc hiệu chỉnh cho mỗi bin được tính toán bằng cách áp dụng nội dung GC cho bin cần xem xét cho hàm điều chỉnh. Ví dụ, đại lượng che phủ được chuẩn hóa có thể được điều chỉnh bằng cách trừ đi giá trị độ che phủ kỳ vọng của hàm nối trục tại nội dung GC của bin cần xem xét. Tùy ý, có thể đạt được điều chỉnh này bằng cách chia giá trị độ che phủ kỳ vọng theo hàm nối trục.

Sau khi hiệu chỉnh độ phụ thuộc GC trong thao tác 319, hệ thống loại bỏ theo tính toán các bin ngoại lệ trong mẫu cần xem xét - Xem khối 321. Phép toán này có thể được gọi là lọc hoặc cắt mẫu đơn. Fig.3J thể hiện rằng ngay cả sau khi hiệu chỉnh GC, độ che phủ vẫn có độ biến động theo từng mẫu cụ thể ở các vùng nhỏ. Xem ví dụ về độ che phủ vị trí 1.1 e8 trên nhiễm sắc thể 12, ở đó có độ lệch cao không mong muốn so với các kết quả có giá trị kỳ vọng. Có thể độ lệch này là kết quả của biến thể số lượng bản sao nhỏ trong hệ gen vật chất. Tùy ý, điều này có thể là do các nguyên nhân kỹ thuật trong khi giải trình tự không liên quan đến biến thể số lượng bản sao. Thông thường, phép toán này chỉ áp dụng cho các nhiễm sắc thể khỏe mạnh.

Theo một ví dụ, các hệ thống lọc theo tính toán các bin bất kỳ có đại lượng che phủ được chuẩn hóa được hiệu chỉnh GC của nhiều hơn 3 độ lệch tuyệt đối trung bình so với trung bình của đại lượng che phủ được chuẩn hóa được hiệu chỉnh GC qua tất cả các bin trong nhiễm sắc thể có bin cần xem xét lọc. Trong một ví dụ, giá trị loại bỏ được xác định là 3 độ lệch tuyệt đối trung bình được điều chỉnh phù hợp với độ lệch chuẩn, do đó thực tế là giá trị loại bỏ là $1,4826 \times \text{độ lệch tuyệt đối trung bình từ giá trị trung bình}$. Theo một số phương án, phép toán này được áp dụng cho tất cả các nhiễm sắc thể trong mẫu, bao gồm cả các nhiễm sắc thể khỏe mạnh và các nhiễm sắc thể nghi ngờ có đột biến lệch bội.

Theo các phương án thực hiện cụ thể, phép toán bổ sung mà có thể được định rõ là kiểm soát chất lượng được thực hiện. Xem khối 323. Theo một số phương án, việc đo kiểm soát

chất lượng bao gồm bước dò xem có bất kỳ nhiễm sắc thể mẫu thức tiềm năng nào, tức là "các nhiễm sắc thể chuẩn hóa" hoặc "các nhiễm sắc thể khỏe mạnh" bị đột biến lệch bội hoặc nếu không thì không phù hợp cho việc xác định liệu mẫu thử nghiệm có biến thể số lượng bản sao ở trình tự quan tâm hay không. Khi quy trình xác định rằng nhiễm sắc thể khỏe mạnh là không phù hợp, quy trình có thể bỏ qua mẫu thử nghiệm và không thực hiện xác định đột biến. Theo cách khác, kết quả đo QC không đạt có thể kích hoạt việc sử dụng tập nhiễm sắc thể chuẩn hóa thay thế để thực hiện xác định đột biến. Trong một ví dụ, phương pháp kiểm soát chất lượng so sánh các giá trị độ che phủ được chuẩn hóa chuẩn dành cho các nhiễm sắc thể khỏe mạnh so với các giá trị kỳ vọng dành cho các nhiễm sắc thể thường khỏe mạnh. Các giá trị kỳ vọng có thể thu được bằng cách khớp mô hình chuẩn nhiều biến theo các biên dạng chuẩn hóa của các mẫu huấn luyện không bị bệnh, lựa chọn cấu trúc mô hình tốt nhất theo tính hợp lệ của dữ liệu hoặc tiêu chí Bayesian (ví dụ, mô hình được chọn bằng cách sử dụng tiêu chí thông tin Akaike hoặc có thể là tiêu chí thông tin Bayesian), và ấn định mô hình tốt nhất để dùng trong QC. Các mô hình chuẩn của các nhiễm sắc thể khỏe mạnh có thể được thu bằng cách, ví dụ, sử dụng kỹ thuật gộp nhóm mà nhận dạng hàm xác suất có độ lệch trung bình và độ lệch chuẩn cho các độ che phủ nhiễm sắc thể trong các mẫu chuẩn. Tất nhiên là cũng có thể sử dụng các dạng mô hình khác. Quy trình đánh giá tính hợp lệ của độ che phủ được chuẩn hóa quan sát được trong bất kỳ mẫu thử nghiệm sắp tới có các tham số mô hình cố định. Có thể thực hiện việc này bằng cách tính điểm mỗi mẫu thử nghiệm sắp tới với mô hình để thu được tính hợp lệ và qua đó nhận dạng các giá trị ngoại lệ liên quan đến tập mẫu thông thường. Độ lệch trong tính hợp lệ của mẫu thử nghiệm so với độ lệch của các mẫu huấn luyện có thể cho thấy sự bất thường ở các nhiễm sắc thể chuẩn hóa hoặc mẫu có mang thành phần lạ / mẫu thử nghiệm xử lý thành phần lạ mà có thể dẫn đến phân loại mẫu sai. Kết quả đo QC này có thể được sử dụng để làm giảm lỗi khi phân loại đi kèm với các thành phần lạ trong mẫu này. Fig.3K, bảng bên phải, thể hiện số lượng nhiễm sắc thể trên trục x và trục y thể hiện độ che phủ nhiễm sắc thể chuẩn hóa dựa vào kết quả so sánh với mô hình QC thu được như được mô tả trên đây. Đồ thị này thể hiện một mẫu có độ che phủ thừa cho nhiễm sắc thể 2 và mẫu khác có độ che phủ thừa cho nhiễm sắc thể 20. Các mẫu này sẽ bị loại bỏ bằng cách sử dụng kết quả đo QC được mô tả ở đây hoặc đổi sang sử dụng tập nhiễm sắc thể chuẩn hóa thay thế. Bảng bên trái trên Fig.3K thể hiện NCV so với tính hợp lệ cho một nhiễm sắc thể.

Trình tự được mô tả trên Fig.3A có thể được sử dụng cho tất cả các bin của tất cả các nhiễm sắc thể trong hệ gen. Theo một số phương án, quy trình khác được áp dụng cho nhiễm sắc thể Y. Để tính định lượng nhiễm sắc thể hoặc phân đoạn, NCV, và/hoặc NSV, các đại lượng che phủ được chuẩn hóa được hiệu chỉnh (như được xác định trên Fig.3A) từ các bin trong các nhiễm sắc thể hoặc phân đoạn dùng khi thể hiện định lượng, NCV, và/hoặc NSV được sử dụng. Xem khối 325. Theo một số phương án, đại lượng che phủ được chuẩn hóa trung bình được tính từ tất cả các bin trong nhiễm sắc thể quan tâm, nhiễm sắc thể chuẩn hóa, phân đoạn quan tâm, và/hoặc phân đoạn chuẩn hóa được dùng để tính định lượng trình tự, NCV, và/hoặc NSV như được mô tả ở phần khác của bản mô tả.

Theo một số phương án, nhiễm sắc thể Y được xử lý theo cách khác. Nó có thể được lọc bằng cách che tập các bin duy nhất đối với nhiễm sắc thể Y. Theo một số phương án, việc lọc nhiễm sắc thể Y được xác định theo quy trình trong đơn sáng chế tạm thời Mỹ số 61/836,057, đã được đưa vào ở trên bằng cách viện dẫn. Theo một số phương án, bộ lọc che đi các bin nhỏ hơn các bin trong bộ lọc của các nhiễm sắc thể còn lại. Ví dụ, mặt nạ nhiễm sắc thể Y có thể lọc ở mức 1kb, trong khi các mặt nạ nhiễm sắc thể khác có thể lọc ở mức 100kb. Tuy nhiên, nhiễm sắc thể Y có thể được chuẩn hóa ở cùng kích thước bin như các nhiễm sắc thể khác (ví dụ, 100kb).

Theo một số phương án, nhiễm sắc thể Y đã được lọc được chuẩn hóa như được mô tả trên đây trong phép toán 315 trên Fig.3A. Tuy nhiên, ngược lại thì nhiễm sắc thể Y sẽ không được hiệu chỉnh thêm nữa. Do đó, các bin của nhiễm sắc thể Y không phải trải qua quá trình loại bỏ biên dạng chung. Tương tự, các bin của nhiễm sắc thể Y không trải qua hiệu chỉnh GC hoặc các bước lọc khác được thực hiện sau đây. Điều này là bởi vì khi mẫu được xử lý, quy trình không biết đây là mẫu đực hay mẫu cái. Mẫu cái cần phải không có đoạn đọc nào đồng chỉnh với nhiễm sắc thể Y tham chiếu.

Tạo mặt nạ trình tự

Một số phương án bộc lộ ở đây sử dụng chiến lược lọc ra (hoặc che) các đoạn đọc trình tự không phân biệt trên trình tự quan tâm bằng cách sử dụng các mặt nạ trình tự, điều này tạo ra tín hiệu cao hơn và nhiễu thấp hơn, tương đối so với các giá trị được tính bằng các phương pháp thông thường, ở các giá trị độ che phủ được dùng để đánh giá CNV. Các mặt nạ này có thể được nhận dạng bằng nhiều kỹ thuật khác nhau. Theo một phương án, mặt nạ được nhận

dạng bằng cách sử dụng kỹ thuật được thể hiện trên các hình vẽ Fig.4A-4B như được giải thích chi tiết dưới đây.

Theo một số phương án thực hiện, mặt nạ được nhận dạng bằng cách sử dụng tập huấn luyện gồm các mẫu đại diện được biết là có số lượng bản sao bình thường của trình tự quan tâm. Các mặt nạ có thể được nhận dạng bằng cách sử dụng kỹ thuật mà trước tiên là chuẩn hóa tập huấn luyện các mẫu, sau đó hiệu chỉnh về sự biến thiên hệ thống trong dãy trình tự (ví dụ, biên dạng), và sau đó hiệu chỉnh chúng về độ biến thiên GC như được mô tả dưới đây. Việc chuẩn hóa và hiệu chỉnh được thực hiện trên các mẫu từ tập huấn luyện, chứ không phải các mẫu thử nghiệm. Mặt nạ được nhận dạng một lần và sau đó được áp dụng cho nhiều mẫu thử nghiệm.

Fig.4A thể hiện lưu đồ cho quy trình 400 để tạo ra mặt nạ trình tự, mặt nạ này có thể được áp dụng cho một hoặc nhiều mẫu thử nghiệm để loại bỏ các bin trên trình tự để không xem xét trong ước lượng số lượng bản sao. Quy trình 400 được thể hiện trên Fig.4 sử dụng độ che phủ để xác định trình tự dựa vào số lượng các thẻ xác định trình tự để thu được mặt nạ trình tự. Tuy nhiên, tương tự như phần mô tả trên đây về quy trình 100 để xác định CNV dựa theo Fig.1, các biến số hoặc tham số khác, chẳng hạn như kích thước, tỷ lệ kích thước, và mức methyl hóa, có thể được sử dụng để bổ sung hoặc thay cho độ che phủ cho quy trình 400. Theo một số phương án thực hiện, một mặt nạ được tạo cho mỗi tham số trong số hai hoặc nhiều tham số. Hơn nữa, độ che phủ và các tham số khác có thể được gán trọng số dựa vào kích thước của các đoạn mà từ đó các thẻ được rút ra. Để dễ hiểu, chỉ có độ che phủ được đề cập trong quy trình 400, nhưng cần lưu ý rằng các tham số khác như kích thước, tỷ lệ kích thước, và mức methyl hóa, số đếm được gán trọng số theo kích thước, v.v. có thể được sử dụng thay cho độ che phủ.

Quy trình 400 bắt đầu bằng cách tạo tập huấn luyện bao gồm các đoạn đọc trình tự từ nhiều mẫu huấn luyện không bị bệnh. Khối 402. Khi đó quy trình căn chỉnh các đoạn đọc trình tự của tập huấn luyện với hệ gen chuẩn chứa trình tự quan tâm, qua đó tạo ra các thẻ trình tự huấn luyện cho các mẫu huấn luyện. Khối 404. Theo một số phương án, chỉ có các thẻ không dư được căn chỉnh duy nhất được ánh xạ đến các vị trí không loại trừ là được dùng cho phân tích sau. Quy trình bao gồm bước chia hệ gen chuẩn thành nhiều bin và xác định cho mỗi mẫu huấn luyện không bị bệnh một độ che phủ của các thẻ trình tự huấn luyện trong

mỗi bin cho mỗi mẫu huấn luyện. Khối 406. Quy trình còn xác định cho mỗi bin độ che phủ kỳ vọng của các thẻ trình tự huấn luyện trong tất cả các mẫu huấn luyện. Khối 408. Theo một số phương án, độ che phủ kỳ vọng của mỗi bin là giá trị trung bình của các mẫu huấn luyện. Các độ che phủ kỳ vọng tạo thành biên dạng chung. Khi đó quy trình điều chỉnh độ che phủ của các thẻ trình tự huấn luyện trong mỗi bin cho mỗi mẫu huấn luyện bằng cách loại bỏ sự biến động ở biên dạng chung, qua đó thu được các độ che phủ được hiệu chỉnh theo biên dạng chung của các thẻ trình tự huấn luyện trong các bin cho mỗi mẫu huấn luyện. Sau đó quy trình tạo ra mặt nạ trình tự bao gồm các bin không được che và được che trong hệ gen chuẩn. Mỗi bin đã che có đặc điểm phân bố vượt ngưỡng che. Đặc điểm phân bố được tạo cho các độ che phủ được điều chỉnh của các thẻ trình tự huấn luyện trong bin qua các mẫu huấn luyện. Theo một số phương án thực hiện, ngưỡng che có thể liên quan đến độ biến thể quan sát được trong độ che phủ được chuẩn hóa trong bin qua các mẫu huấn luyện. Các bin có hệ số biến thể cao hoặc độ lệch tuyệt đối trung bình của độ che phủ được chuẩn hóa qua các mẫu có thể được nhận dạng dựa vào sự phân bố thực nghiệm của các số đo tương ứng. Theo một số phương án thay thế, ngưỡng che có thể liên quan đến độ biến thể quan sát được trong độ che phủ được chuẩn hóa trong bin qua các mẫu huấn luyện. Các bin có hệ số biến thể cao hoặc độ lệch tuyệt đối trung bình của độ che phủ được chuẩn hóa qua các mẫu có thể được che đi dựa vào sự phân bố thực nghiệm của các số đo tương ứng.

Trong một số phương án thực hiện, các giá trị loại bỏ riêng để nhận dạng các bin được che, tức là, các ngưỡng che, được xác định cho nhiễm sắc thể quan tâm và cho tất cả các nhiễm sắc thể khác. Hơn nữa, các ngưỡng che riêng có thể được xác định cho mỗi nhiễm sắc thể quan tâm một cách riêng biệt, và một ngưỡng che cho tập gồm tất cả các nhiễm sắc thể không bị bệnh. Ví dụ, mặt nạ dựa vào ngưỡng che nhất định được xác định cho nhiễm sắc thể 13 và ngưỡng che khác được dùng để xác định mặt nạ cho các nhiễm sắc thể khác. Các nhiễm sắc thể không bị bệnh có thể còn có các ngưỡng mặt nạ được xác định cho từng nhiễm sắc thể.

Các tổ hợp ngưỡng che khác nhau có thể được ước lượng cho mỗi nhiễm sắc thể quan tâm. Các tổ hợp ngưỡng che tạo ra một mặt nạ cho các bin của nhiễm sắc thể quan tâm và mặt nạ khác cho các bin của tất cả các nhiễm sắc thể còn lại.

Theo một phương pháp, khoảng giá trị cho hệ số biến thiên (CV) hoặc số đo giá trị loại bỏ phân bố mẫu được xác định dưới dạng các số phần trăm (ví dụ, 95, 96, 97, 98, 99) của phân bố thực nghiệm của các giá trị CV bin và các giá trị loại bỏ này được áp dụng cho tất cả các nhiễm sắc thể thường ngoại trừ các nhiễm sắc thể quan tâm. Hơn nữa, khoảng giá trị loại bỏ phân vị cho CV được xác định cho phân bố CV thực nghiệm và các giá trị loại bỏ này được áp dụng cho nhiễm sắc thể quan tâm (ví dụ, nhiễm sắc thể 21). Theo một số phương án, các nhiễm sắc thể quan tâm là nhiễm sắc thể X và các nhiễm sắc thể 13, 18, và 21. Tất nhiên, có thể cân nhắc đến các phương pháp khác; ví dụ, bước tối ưu hóa riêng biệt có thể được thực hiện cho mỗi nhiễm sắc thể. Các khoảng giá trị cần được đồng thời tối ưu hóa song song (ví dụ, một khoảng cho nhiễm sắc thể quan tâm cần xem xét và một khoảng khác cho tất cả các nhiễm sắc thể khác) xác định lưới tổ hợp loại bỏ CV. Xem Fig.4B. Hoạt động của hệ thống trên tập huấn luyện được đánh giá qua hai giá trị loại bỏ (một cho các nhiễm sắc thể chuẩn hóa (hoặc các nhiễm sắc thể thường khác nhiễm sắc thể quan tâm) và một dành cho nhiễm sắc thể quan tâm), và tổ hợp hoạt động tốt nhất sẽ được chọn cho cấu hình cuối cùng. Tổ hợp này có thể khác nhau đối với mỗi nhiễm sắc thể quan tâm. Theo một số phương án, hiệu quả được đánh giá trên tập hợp lệ thay cho tập huấn luyện, tức là tính hợp lệ chéo được dùng để đánh giá hiệu quả.

Theo một số phương án, hiệu quả được tối ưu hóa để xác định khoảng giá trị loại bỏ là hệ số biến đổi của các định lượng nhiễm sắc thể (dựa vào việc lựa chọn thử các nhiễm sắc thể chuẩn hóa). Quy trình chọn tổ hợp các giá trị loại bỏ làm giảm thiểu CV của định lượng nhiễm sắc thể (ví dụ, tỷ lệ) của nhiễm sắc thể quan tâm bằng cách sử dụng (các) nhiễm sắc thể chuẩn hóa hiện đang được lựa chọn. Theo một phương pháp, quy trình thử hiệu quả của mỗi tổ hợp giá trị loại bỏ trong lưới theo cách sau: (1) áp dụng tổ hợp giá trị loại bỏ để xác định mặt nạ cho tất cả các nhiễm sắc thể và áp dụng các mặt nạ đó để lọc các thẻ của tập huấn luyện; (2) tính các độ che phủ được chuẩn hóa qua tập huấn luyện của các mẫu không bị bệnh bằng cách áp dụng quy trình trên Fig.3A cho các thẻ đã lọc; (3) xác định độ che phủ được chuẩn hóa làm đại diện trên mỗi nhiễm sắc thể by, ví dụ, cộng tổng các độ che phủ được chuẩn hóa của bin cho nhiễm sắc thể cần xem xét; (4) tính các định lượng nhiễm sắc thể bằng cách sử dụng các nhiễm sắc thể chuẩn hóa hiện tại, và (5) xác định CV của các định lượng nhiễm sắc thể. Quy trình có thể đánh giá biểu hiện của các bộ lọc được chọn bằng cách áp dụng chúng cho tập

mẫu thử nghiệm được tách từ phần ban đầu của tập huấn luyện. Tức là, quy trình tách tập huấn luyện ban đầu thành các tập con huấn luyện và thử nghiệm. Tập con huấn luyện được dùng để xác định giá trị loại bỏ mặt nạ như được mô tả trên đây.

Theo các phương án thay thế, thay vì xác định các mặt nạ dựa vào CV của các độ che phủ, các mặt nạ có thể được xác định bằng việc phân bố các điểm chất lượng ánh xạ từ kết quả sắp xếp qua các mẫu huấn luyện trong các bin. Điểm chất lượng ánh xạ phản ánh tính duy nhất mà đoạn đọc được ánh xạ đến hệ gen chuẩn. Nói cách khác, các điểm chất lượng ánh xạ xác định số xác suất mà đoạn đọc được sắp xếp sai. Điểm chất lượng ánh xạ thấp kéo theo tính duy nhất thấp (xác suất sắp thẳng lệch lớn). Tính duy nhất là đại lượng chịu trách nhiệm cho một hoặc nhiều lỗi trong trình tự đoạn đọc (như được tạo ra từ bộ giải trình tự). Mô tả chi tiết về các điểm chất lượng ánh xạ được viết trong tài liệu của Li H, Ruan J, Durbin R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores.. *Genome Research* **18**:1851-8, được đưa vào đây bằng cách viện dẫn toàn bộ. Theo một số phương án thực hiện, điểm chất lượng ánh xạ ở đây được gọi là điểm MapQ. Fig.4B thể hiện rằng điểm MapQ có tính tương quan đơn điệu lớn với CV của các độ che phủ được xử lý. Ví dụ, các bin có CV cao hơn 0,4 gần như hoàn toàn gộp vào phía bên trái của sơ đồ trên Fig.4B, có điểm MapQ thấp hơn khoảng 4. Do đó, việc che đi các bin có MapQ nhỏ có thể tạo ra mặt nạ khá giống với mặt nạ được xác định bằng cách che đi các bin có CV cao.

Mẫu và xử lý mẫu

Mẫu

Các mẫu được dùng để xác định CNV, ví dụ, các đột biến lệch bội của nhiễm sắc thể, đột biến lệch bội một phần, và tương tự, có thể bao gồm các mẫu được lấy từ tế bào, mô, hoặc bộ phận bất kỳ trong đó các biến thể số lượng bản sao đối với một hoặc nhiều trình tự quan tâm sẽ được xác định. Như kỳ vọng thì các mẫu chứa các axit nucleic có trong tế bào và/hoặc axit nucleic "ngoài tế bào" (ví dụ, cfADN).

Theo một số phương án sẽ có lợi khi thu các axit nucleic ngoài tế bào, ví dụ, ADN ngoài tế bào (cfADN). Các axit nucleic ngoài tế bào, bao gồm ADN ngoài tế bào, có thể thu được bằng nhiều phương pháp đã biết trong lĩnh vực từ các mẫu sinh học bao gồm nhưng không giới hạn ở huyết tương, huyết thanh và nước tiểu (ví dụ xem tài liệu của Fan et al., *Proc Natl Acad Sci* 105:16266-16271 [2008]; Koide et al., *Prenatal Diagnosis* 25:604-607 [2005]; Chen

et al., *Nature Med.* 2: 1033-1035 [1996]; Lo et al., *Lancet* 350: 485-487 [1997]; Botezatu et al., *Clin Chem.* 46: 1078-1084, 2000; và Su et al., *J Mol. Diagn.* 6: 101-107 [2004]). Để tách ADN ngoài tế bào từ các tế bào trong mẫu, các phương pháp bao gồm, nhưng không giới hạn ở cắt phân đoạn, ly tâm (ví dụ, ly tâm chênh lệch tỷ trọng), kết tủa ADN đặc hiệu, hoặc phân loại tế bào lưu lượng cao và/hoặc các phương pháp tách khác có thể được dùng. Hiện có các kit bán sẵn trên thị trường để tách cfADN tự động hoặc bằng tay (Roche Diagnostics, Indianapolis, IN, Qiagen, Valencia, CA, Macherey-Nagel, Duren, DE). Các mẫu sinh học bao gồm cfADN đã được dùng trong các thử nghiệm để xác định sự có mặt hoặc không của các bất thường nhiễm sắc thể, ví dụ, thể ba nhiễm 21, bằng các thử nghiệm giải trình tự có thể phát hiện các đột biến lệch bội của nhiễm sắc thể và/hoặc các hiện tượng đa hình khác.

Theo các phương án khác, cfADN có trong mẫu có thể được làm giàu đặc hiệu hoặc không đặc hiệu trước khi sử dụng (ví dụ, trước khi tạo thư viện giải trình tự). Việc làm giàu ADN mẫu không đặc hiệu chỉ toàn bộ khuếch đại hệ gen của các đoạn ADN bộ gen của mẫu có thể được dùng để làm tăng mức của ADN mẫu trước khi tạo thư viện giải trình tự cfADN. Việc làm giàu ADN mẫu không đặc hiệu có thể là sự làm giàu có lựa chọn của một trong hai hệ gen có trong mẫu bao gồm nhiều hơn một hệ gen. Ví dụ, sự làm giàu không đặc hiệu có thể có tính lựa chọn của hệ gen bào thai trong mẫu mẹ, điều này có thể thu được bằng các phương pháp đã biết để tăng tỷ lệ tương ứng giữa ADN bào thai và ADN của mẹ trong mẫu. Tuy ý, sự làm giàu không đặc hiệu có thể là sự khuếch đại không chọn lọc của cả hai hệ gen có trong mẫu. Ví dụ, sự khuếch đại không đặc hiệu có thể là thuộc về ADN mẹ và con trong mẫu bao gồm hỗn hợp ADN từ các hệ gen của mẹ và bào thai. Các phương pháp khuếch đại toàn bộ hệ gen đều đã được biết đến trong lĩnh vực. PCR sử dụng môi là oligonucleotit thoái biến (degenerate oligonucleotide-primed - DOP), kỹ thuật PCR kéo dài môi (primer extension PCR - PEP) và khuếch đại đa dịch chuyển (multiple displacement amplification - MDA) là ví dụ về các phương pháp khuếch đại toàn bộ hệ gen. Theo một số phương án, mẫu bao gồm hỗn hợp của cfADN từ các hệ gen khác nhau không được làm giàu cho cfADN của các hệ gen có trong hỗn hợp. Theo phương án khác, mẫu bao gồm hỗn hợp của cfADN từ các hệ gen khác nhau được làm giàu không đặc hiệu cho bất kỳ hệ gen nào có mặt trong mẫu.

Mẫu bao gồm (các) axit nucleic mà các phương pháp được mô tả ở đây được áp dụng cho thì thường bao gồm mẫu sinh học ("mẫu thử nghiệm"), ví dụ, như được mô tả trên đây.

Theo một số phương án, (các) axit nucleic cần lọc cho một hoặc nhiều CNV được tinh sạch hoặc phân lập bằng các phương pháp đã biết bất kỳ.

Theo đó, theo một số phương án mẫu bao gồm hoặc gồm có polynucleotit được tinh sạch hoặc phân lập, hoặc có thể bao gồm các mẫu chẳng hạn như mẫu mô, mẫu dịch lỏng sinh học, mẫu tế bào, và tương tự. Các mẫu dịch lỏng sinh học thích hợp bao gồm, nhưng không giới hạn ở máu, huyết tương, huyết thanh, mô hôi, nước mắt, đờm, nước tiểu, đờm, dịch tai, bạch huyết, nước bọt, dịch não tủy, ravages, dịch huyền phù tủy xương, dịch âm đạo, dịch rửa qua cổ, dịch não, dịch màng bụng, sữa, các chất tiết của hệ hô hấp, đường ruột và đường sinh dục niệu, dịch ối, sữa, và các mẫu gạn bạch cầu. Theo một số phương án, mẫu là mẫu có thể dễ dàng thu được bằng các quy trình không xâm lấn, ví dụ, máu, huyết tương, huyết thanh, mô hôi, nước mắt, đờm, nước tiểu, đờm, dịch tai, nước bọt hoặc phân. Theo một số phương án mẫu là mẫu máu ngoại vi, hoặc các phần huyết tương và/hoặc huyết thanh của mẫu máu ngoại vi. Theo các phương án khác, mẫu sinh học là miếng gạc hoặc kính phết, mẫu sinh thiết, hoặc nuôi cấy tế bào. Theo phương án khác, mẫu là hỗn hợp của hai hoặc nhiều mẫu sinh học, ví dụ, mẫu sinh học có thể bao gồm hai hoặc nhiều mẫu dịch lỏng sinh học, mẫu mô, và mẫu nuôi cấy tế bào. Như được sử dụng ở đây, thuật ngữ "máu," "huyết tương" và "huyết thanh" rõ ràng bao gồm các phân đoạn hoặc các phần đã được xử lý của nó. Tương tự, khi mẫu được lấy từ sinh thiết, băng gạc, kính phết, v.v., "mẫu" rõ ràng bao gồm phân đoạn hoặc phần đã được xử lý có nguồn gốc từ sinh thiết, băng gạc, kính phết, v.v..

Theo một số phương án, các mẫu có thể thu được từ các nguồn, bao gồm, nhưng không giới hạn ở, các mẫu từ những cá thể khác nhau, các mẫu từ các giai đoạn khác nhau của cùng một cá thể hoặc từ những cá thể khác nhau, các mẫu từ các cá thể mang bệnh khác nhau (ví dụ, các cá thể mắc ung thư hoặc nghi ngờ có rối loạn di truyền), các cá thể bình thường, các mẫu thu được ở các giai đoạn bệnh khác nhau trong một cá thể, các mẫu thu được từ một cá thể trải qua các quá trình trị liệu bệnh khác nhau, các mẫu từ các cá thể trải qua các yếu tố môi trường khác nhau, các mẫu từ các cá thể có bệnh lý bẩm sinh, các mẫu các cá thể phơi nhiễm tác nhân gây bệnh lây truyền (ví dụ, HIV), và tương tự.

Theo phương án minh họa khác nhưng không làm giới hạn sáng chế, mẫu là mẫu mẹ thu được từ cá thể cái mang thai, ví dụ phụ nữ mang thai. Trong trường hợp này, mẫu có thể được phân tích bằng cách sử dụng các phương pháp được mô tả ở đây để cung cấp các chuẩn

đoán trước khi sinh về nguy cơ hình thành các dị tật nhiễm sắc thể ở bào thai. Mẫu mẹ có thể là mẫu mô, mẫu dịch lỏng sinh học, hoặc mẫu tế bào. Dịch lỏng sinh học bao gồm, đây là các ví dụ không giới hạn, máu, huyết tương, huyết thanh, mồ hôi, nước mắt, đờm, nước tiểu, đờm, dịch tai, bạch huyết, nước bọt, dịch não tủy, ravages, dịch huyền phù tủy xương, dịch âm đạo, dịch rửa qua cổ, dịch não, dịch màng bụng, sữa, các chất tiết của hệ hô hấp, đường ruột và đường sinh dục niệu, và các mẫu gạn bạch cầu.

Theo phương án minh họa khác nhưng không làm giới hạn sáng chế, mẫu mẹ là hỗn hợp của hai hoặc nhiều mẫu sinh học, ví dụ, mẫu sinh học có thể bao gồm hai hoặc nhiều mẫu dịch lỏng sinh học, mẫu mô, và mẫu nuôi cấy tế bào. Theo một số phương án, mẫu là mẫu có thể dễ dàng thu được bằng các quy trình không xâm lấn, ví dụ, máu, huyết tương, huyết thanh, mồ hôi, nước mắt, đờm, nước tiểu, sữa, đờm, dịch tai, nước bọt và phân. Theo một số phương án, mẫu sinh học là mẫu máu ngoại vi, và/hoặc các phần huyết tương và huyết thanh của mẫu máu ngoại vi. Theo các phương án khác, mẫu sinh học là miếng gạc hoặc kính phết, mẫu sinh thiết, hoặc mẫu nuôi cấy tế bào. Như đã bộc lộ trên đây, thuật ngữ "máu," "huyết tương" và "huyết thanh" rõ ràng bao gồm các phân đoạn hoặc các phần đã được xử lý của nó. Tương tự, trong đó mẫu được lấy từ sinh thiết, băng gạc, kính phết, v.v., "mẫu" rõ ràng bao gồm phân đoạn hoặc phần đã được xử lý có nguồn gốc từ sinh thiết, băng gạc, kính phết, v.v..

Theo một số phương án, các mẫu có thể còn thu được từ các tế bào, mô nuôi cấy trong ống nghiệm, hoặc các nguồn chứa polynucleotit khác. Các mẫu nuôi cấy có thể được lấy từ các nguồn bao gồm, nhưng không giới hạn ở, các mẹ nuôi cấy (ví dụ, mô hoặc các tế bào) được duy trì trong các môi trường và điều kiện khác nhau (ví dụ, pH, áp suất, hoặc nhiệt độ), các mẹ nuôi cấy (ví dụ, mô hoặc các tế bào) được duy trì trong khoảng thời gian khác nhau, các mẹ nuôi cấy (ví dụ, mô hoặc các tế bào) được xử lý bằng các yếu tố hoặc thuốc thử khác nhau (ví dụ, thuốc ứng viên, hoặc chất điều biến), hoặc các mẹ nuôi cấy thuộc loại mô và/hoặc các tế bào khác nhau.

Các phương pháp phân lập các axit nucleic từ các nguồn sinh học đã được biết rõ và sẽ khác nhau tùy theo bản chất của nguồn. Người có hiểu biết trung bình trong lĩnh vực hoàn toàn có thể phân lập (các) axit nucleic từ nguồn khi cần cho phương pháp được mô tả ở đây. Trong một số trường hợp, có thể có lợi khi phân đoạn các phân tử axit nucleic trong mẫu axit nucleic. Việc phân đoạn có thể ngẫu nhiên, hoặc có thể đặc hiệu đạt được, ví dụ, bằng cách

sử dụng sự phân cắt bằng enzym cắt giới hạn. Phương pháp phân đoạn ngẫu nhiên được biết rõ trong lĩnh vực, và bao gồm, ví dụ, phân cắt bằng ADNaza giới hạn, xử lý kiềm và cắt vật lý. Theo một phương án, các axit nucleic mẫu thu được từ cfADN không trải qua bước phân đoạn.

Tạo thư viện giải trình tự

Theo một phương án, các phương pháp được mô tả ở đây có thể sử dụng các công nghệ giải trình tự thế hệ tiếp theo (NGS), cho phép nhiều mẫu được giải trình tự riêng dưới dạng các phân tử thuộc hệ gen (tức là, giải trình tự đơn phức) hoặc dưới dạng các mẫu chung bao gồm các phân tử thuộc hệ gen được tạo chỉ số (ví dụ, giải trình tự phức) trên một chu trình giải trình tự. Các phương pháp này có thể tạo ra hàng trăm triệu đoạn đọc của các trình tự ADN. Theo các phương án khác nhau, các trình tự axit nucleic bộ gen, và/hoặc trình tự axit nucleic bộ gen được tạo chỉ số có thể được xác định bằng cách sử dụng, ví dụ, các công nghệ giải trình tự thế hệ tiếp theo (NGS) được mô tả ở đây. Theo các phương án khác nhau, việc phân tích lượng dữ liệu trình tự không lồ thu được bằng cách sử dụng NGS có thể được thực hiện bằng cách sử dụng một hoặc nhiều bộ xử lý như được mô tả ở đây.

Theo các phương án khác nhau, việc sử dụng các công nghệ giải trình tự này không yêu cầu việc tạo thư viện giải trình tự.

Tuy nhiên, theo một số phương án, các phương pháp giải trình tự được dự tính ở đây yêu cầu việc tạo thư viện giải trình tự. Theo một phương án minh họa, việc tạo thư viện giải trình tự bao gồm bước tạo tập hợp ngẫu nhiên gồm các đoạn ADN (ví dụ, các polynucleotit) được biến đổi phù hợp sẵn sàng được giải trình tự. Các thư viện giải trình tự gồm các polynucleotit có thể được tạo ra từ ADN hoặc ARN, bao gồm các dạng tương đương, dạng tương tự của ADN hoặc cADN, ví dụ, ADN hoặc cADN là ADN bổ sung hoặc bản sao được tạo ra từ khuôn ARN, nhờ hoạt động của enzym phiên mã ngược. Các polynucleotit có thể bắt nguồn ở dạng sợi đôi (ví dụ, dsADN chẳng hạn như các đoạn ADN thuộc hệ gen, cADN, các sản phẩm khuếch đại PCR, và tương tự) hoặc, theo một số phương án, các polynucleotit có thể bắt nguồn từ dạng sợi đơn (ví dụ, ssADN, ARN, v.v.) và chuyển đổi thành dạng dsADN. Ví dụ, theo một số phương án, các phân tử mARN sợi đơn có thể được sao chép thành các cADN sợi đôi phù hợp dùng trong việc tạo thư viện giải trình tự. Trình tự chính xác của các phân tử polynucleotit sơ cấp thường không quan trọng đối với phương pháp tạo thư viện, và

có thể đã được biết đến hoặc chưa. Theo một phương án, các phân tử polynucleotit là các phân tử ADN. Cụ thể hơn, theo một số phương án, các phân tử polynucleotit thể hiện toàn bộ phần bổ sung di truyền của một sinh vật hoặc gần như toàn bộ phần bổ sung di truyền của một sinh vật, và là các phân tử ADN bộ gen (ví dụ, ADN tế bào, ADN ngoài tế bào (cfADN), v.v.), thường bao gồm cả trình tự intron và trình tự exon (trình tự mã hóa), cũng như các trình tự điều hòa không mã hóa chẳng hạn như các trình tự gen khởi đầu và gen tăng cường. Theo một số phương án, các phân tử polynucleotit sơ cấp bao gồm các phân tử ADN của bộ gen người, ví dụ, các phân tử cfADN có trong máu ngoại vi của người mang thai.

Việc tạo các thư viện giải trình tự cho một số nền tảng giải trình tự NGS được tạo điều kiện thuận lợi hơn nhờ dùng các polynucleotit có khoảng kích thước đoạn đặc hiệu. Việc điều chế các thư viện này thường bao gồm việc phân đoạn các polynucleotit lớn (ví dụ ADN bộ gen của tế bào) để thu được các polynucleotit nằm trong khoảng kích thước kỳ vọng.

Việc phân đoạn có thể đạt được bằng các phương pháp bất kỳ mà người có hiểu biết trung bình trong lĩnh vực đều biết. Ví dụ, việc phân đoạn có thể đạt được bằng các phương tiện cơ học bao gồm, nhưng không giới hạn ở khí dung hóa, sóng siêu âm và cắt thủy lực. Tuy nhiên việc phân đoạn cơ học này thường chỉ khung chính ADN tại các liên kết C-O, P-O và C-C tạo ra hỗn hợp không đồng nhất của các đầu bằng và các đầu nhô ở đầu 3' và 5' với liên kết C-O, P-O và/ C-C bị bẻ gãy (xem, ví dụ, Alnemri and Liwack, J Biol. Chem 265:17323-17333 [1990]; Richards and Boyer, J Mol Biol 11:327-240 [1965]) có thể cần được sửa chữa do chúng có thể thiếu 5'-phosphat cần thiết cho các phản ứng enzym tiếp theo, ví dụ, bước nối các trình tự nối (adaptor) giải trình tự mà cần thiết để tạo ADN để giải trình tự.

Ngược lại, cfADN, thường tồn tại dưới dạng các phân đoạn có ít hơn 300 cặp bazơ và do đó, việc phân đoạn thường là không cần thiết để tạo thư viện giải trình tự bằng cách sử dụng các mẫu cfADN.

Thông thường, dù các polynucleotit bị buộc phải phân đoạn (ví dụ, được phân đoạn trong ống nghiệm), hay tồn tại tự nhiên dưới dạng các đoạn, chúng cũng được chuyển đổi thành ADN đầu bằng có 5'-phosphat và 3'-hydroxyl. Các giao thức chuẩn, ví dụ, giao thức giải trình tự bằng cách sử dụng, ví dụ, nền tảng Illumina như được mô tả ở phần khác của bản mô tả, hướng dẫn người dùng sửa đầu ADN mẫu, tinh sạch các sản phẩm được sửa đầu trước

khi tạo đuôi dA, và tinh sạch các sản phẩm tạo đuôi dA trước các bước nối trình tự nối của quá trình tạo thư viện.

Các phương án khác nhau của các phương pháp tạo thư viện giải trình tự được mô tả ở đây loại bỏ yêu cầu thực hiện một hoặc nhiều bước thường bị bắt buộc bởi các giao thức chuẩn để thu được chế phẩm ADN có thể được giải trình tự bởi NGS. Phương pháp rút gọn (phương pháp ABB), phương pháp 1 bước, và phương pháp 2 bước là các ví dụ về phương pháp tạo thư viện giải trình tự, có thể được tìm thấy ở đơn sáng chế số 13/555,037 nộp ngày 20/07/2012, đưa viện dẫn toàn bộ vào bản mô tả này.

Các axit nucleic đánh dấu để theo dõi và xác nhận tính nguyên vẹn của mẫu

Theo các phương án khác nhau, việc xác minh tính nguyên vẹn của các mẫu và theo dõi mẫu có thể được hoàn thành bằng cách giải trình tự cho các hỗn hợp của các mẫu axit nucleic bộ gen, ví dụ, cfADN, và các axit nucleic đánh dấu đi kèm mà đã được đưa vào các mẫu, ví dụ, trước khi xử lý.

Các axit nucleic đánh dấu có thể được kết hợp với mẫu thử nghiệm (ví dụ, mẫu nguồn sinh học) và trải qua các quá trình bao gồm, ví dụ, một hoặc nhiều bước trong số phân đoạn mẫu nguồn sinh học, ví dụ, thu phần huyết tương về cơ bản là ngoài tế bào từ toàn bộ mẫu máu, tinh sạch các axit nucleic từ huyết tương đã phân đoạn, hoặc mẫu nguồn sinh học không được phân đoạn, ví dụ như mẫu mô, và giải trình tự. Theo một số phương án, việc giải trình tự bao gồm bước tạo thư viện giải trình tự. Trình tự hoặc tổ hợp các trình tự của các phân tử đánh dấu được kết hợp với mẫu nguồn được chọn làm mẫu nguồn duy nhất. Theo một số phương án, tất cả các phân tử đánh dấu duy nhất trong mẫu đều có cùng trình tự. Theo các phương án khác, các phân tử đánh dấu duy nhất trong mẫu là nhiều trình tự, ví dụ, tổ hợp của hai, ba, bốn, năm, sáu, bảy, tám, chín, 10, 15, 20, hoặc nhiều trình tự khác nhau.

Theo một phương án, tính nguyên vẹn của mẫu có thể được xác nhận bằng cách sử dụng nhiều phân tử axit nucleic đánh dấu có các trình tự giống nhau. Tùy ý, tính đồng nhất của mẫu có thể được xác nhận bằng cách sử dụng nhiều phân tử axit nucleic đánh dấu có ít nhất hai, ít nhất ba, ít nhất bốn, ít nhất năm, ít nhất sáu, ít nhất bảy, ít nhất tám, ít nhất chín, ít nhất mười, ít nhất 11, ít nhất 12, ít nhất 13, ít nhất 14, ít nhất 15, ít nhất 16, ít nhất 17, ít nhất 18, ít nhất 19, ít nhất 20, ít nhất 25, ít nhất 30, ít nhất 35, ít nhất 40, ít nhất 50, hoặc nhiều trình tự khác nhau. Việc xác minh tính nguyên vẹn của nhiều mẫu sinh học, tức là, hai hoặc nhiều mẫu sinh

học, đòi hỏi mỗi mẫu trong số hai hoặc nhiều mẫu được đánh dấu bởi các axit nucleic đánh dấu có trình tự duy nhất đối với mỗi mẫu thử nghiệm trong số nhiều đang được đánh dấu. Ví dụ, mẫu thứ nhất có thể được đánh dấu bởi các axit nucleic đánh dấu có trình tự A, và mẫu thứ hai có thể được đánh dấu bởi các axit nucleic đánh dấu có trình tự B. Tùy ý, mẫu thứ nhất có thể được đánh dấu bởi tất cả các phân tử axit nucleic đánh dấu có trình tự A, và mẫu thứ hai có thể được đánh dấu bởi hỗn hợp của các trình tự B và C, trong đó các trình tự A, B và C là các phân tử đánh dấu có trình tự khác nhau.

(Các) axit nucleic đánh dấu có thể được thêm vào mẫu ở giai đoạn tạo mẫu bất kỳ mà xảy ra trước bước tạo thư viện (nếu cần tạo thư viện) và giải trình tự. Theo một phương án, các phân tử đánh dấu có thể được kết hợp với mẫu nguồn chưa được xử lý. Ví dụ, các axit nucleic đánh dấu có thể được đưa vào ống thu gom dùng để lấy mẫu máu. Tùy ý, các axit nucleic đánh dấu có thể được thêm vào mẫu máu sau khi lấy máu. Theo một phương án, các axit nucleic đánh dấu được thêm vào bình dùng để đựng mẫu dịch lỏng sinh học, ví dụ, (các) axit nucleic đánh dấu được thêm vào ống lấy máu dùng để lấy mẫu máu. Theo phương án khác, (các) axit nucleic đánh dấu được thêm vào một phần của mẫu dịch lỏng sinh học. Ví dụ, các axit nucleic đánh dấu được thêm vào phần huyết tương và/hoặc huyết thanh của mẫu máu, ví dụ, mẫu huyết tương của người mẹ. Theo phương án khác nữa, các phân tử đánh dấu được thêm vào mẫu được tinh sạch, ví dụ, mẫu axit nucleic đã được tinh sạch từ mẫu sinh học. Ví dụ, các axit nucleic đánh dấu được thêm vào mẫu cfADN được tinh sạch của mẹ và bào thai. Tương tự, các axit nucleic đánh dấu có thể được thêm vào mẫu sinh thiết trước khi xử lý mẫu thử nghiệm. Theo một số phương án, các axit nucleic đánh dấu có thể được kết hợp với chất mang vận chuyển các phân tử đánh dấu vào các tế bào của mẫu sinh học. Các chất mang vận chuyển tế bào bao gồm các hạt mỡ cation và nhạy với pH.

Theo các phương án khác nhau, các phân tử đánh dấu có các trình tự bổ sung với bộ gen, đây là các trình tự không có trong hệ gen của mẫu nguồn sinh học. Theo một phương án mẫu, các phân tử đánh dấu được dùng để xác minh tính nguyên vẹn của mẫu nguồn sinh học ở người có các trình tự không xuất hiện ở hệ gen người. Theo một phương án thay thế, các phân tử đánh dấu có các trình tự không xuất hiện ở mẫu nguồn và ở bất kỳ một hoặc nhiều hệ gen đã biết khác. Ví dụ, các phân tử đánh dấu mà được dùng để xác minh tính nguyên vẹn của mẫu nguồn sinh học ở người có các trình tự không xuất hiện ở hệ gen người và hệ gen

chuột. Phương án này cho phép xác nhận tính nguyên vẹn của mẫu thử nghiệm gồm hai hoặc nhiều hệ gen. Ví dụ, tính nguyên vẹn của ADN mẫu ngoài tế bào ở người thu được từ đối tượng bị bệnh bởi mầm bệnh, ví dụ, vi khuẩn, có thể được xác minh bằng cách sử dụng các phân tử đánh dấu có các trình tự không xuất hiện ở cả hệ gen người và hệ gen của vi khuẩn gây ảnh hưởng. Các trình tự của các hệ gen của nhiều mầm bệnh, ví dụ, vi khuẩn, virus, nấm men, nấm, động vật nguyên sinh v.v., được công bố công khai ở địa chỉ web ncbi.nlm.nih.gov/genomes. Theo phương án khác, các phân tử đánh dấu là các axit nucleic có các trình tự không xuất hiện ở bất kỳ hệ gen đã biết. Trình tự của các phân tử đánh dấu có thể được tạo ngẫu nhiên bằng thuật toán.

Theo các phương án khác nhau, các phân tử đánh dấu có thể là các axit deoxyribonucleic (ADN), axit ribonucleic có trong tự nhiên hoặc các thể tương tự axit nucleic nhân tạo (mô phỏng axit nucleic) bao gồm axit nucleic peptit (PNA), axit nucleic morpholino, axit nucleic dạng khóa, axit nucleic glycol, và axit nucleic threose, các loại này khác với ADN hoặc ARN có trong tự nhiên ở các thay đổi về khung chính của phân tử hoặc thể mô phỏng ADN không có khung chính phosphodiester. Các axit deoxyribonucleic có thể là từ các hệ gen có trong tự nhiên hoặc có thể được tạo ra trong phòng thí nghiệm nhờ sử dụng các enzym hoặc tổng hợp hóa học pha rắn. Các phương pháp hóa học có thể còn được dùng để tạo các thể mô phỏng ADN không có trong tự nhiên. Các dẫn xuất ADN có sẵn trong đó liên kết phosphodiester đã được thay thế nhưng deoxyriboza được giữ lại bao gồm nhưng không giới hạn ở thể mô phỏng ADN có các khung chính được tạo bởi liên kết thioformaxetal hoặc carboxamid, đây đều là các thể mô phỏng ADN có cấu trúc tốt. Các thể mô phỏng ADN khác bao gồm dẫn xuất morpholino và axit nucleic peptit (PNA), chứa khung chính giả peptit gốc N-(2-aminoethyl)glyxin (Ann Rev Biophys Biomol Struct 24:167-183 [1995]). PNA là một thể mô phỏng có cấu trúc cực tốt của ADN (hoặc của axit ribonucleic [ARN]), và các oligome PNA có thể tạo ra các cấu trúc đôi rất bền vững với các oligome ADN và ARN (hoặc PNA) bổ sung Watson-Crick, và chúng còn có thể liên kết với các đích trong ADN đôi bằng phương pháp xâm nhập sợi xoắn ốc (Mol Biotechnol 26:233-248 [2004]). Thể mô phỏng/tương tự có cấu trúc tốt khác của thể tương tự ADN mà có thể được dùng làm các phân tử đánh dấu là ADN phosphorothioat trong đó một trong số các oxy không cầu được thay thế bằng lưu huỳnh. Sự thay đổi này làm giảm hoạt động của endo- và exonucleaza 2 bao gồm exonucleaza ADN

POL 1 5' đến 3' và 3' đến 5', các enzym nucleaza SI và PI, RNaza, nucleaza huyết thanh và phosphodiesteraza nọc rắn.

Độ dài của các phân tử đánh dấu có thể khác hoặc không khác với độ dài của các axit nucleic mẫu, tức là, độ dài của các phân tử đánh dấu có thể giống với độ dài của các phân tử trong gen mẫu, hoặc có thể lớn hơn hoặc nhỏ hơn độ dài của các phân tử trong gen mẫu. Độ dài của các phân tử đánh dấu được đo bởi số lượng các cặp bazơ nucleotit hoặc tương tự nucleotit mà cấu thành các phân tử đánh dấu. Các phân tử đánh dấu có độ dài khác với độ dài của các phân tử trong gen mẫu có thể được phân biệt với các axit nucleic nguồn nhờ sử dụng các phương pháp tách đã biết trong lĩnh vực. Ví dụ, sự chênh lệch độ dài của các phân tử axit nucleic mẫu và các phân tử axit nucleic đánh dấu có thể được xác định bằng cách tách điện di, ví dụ, điện di mao quản. Khác biệt kích thước có thể có lợi khi định lượng và đánh giá chất lượng của các axit nucleic mẫu và các axit nucleic đánh dấu. Tốt hơn là, các axit nucleic đánh dấu ngắn hơn các axit nucleic bộ gen, và có độ dài đủ để loại trừ chúng không bị ảnh xạ đến hệ gen của mẫu. Ví dụ, trình tự 30 cặp bazơ ở người là cần thiết để ánh xạ duy nhất nó đến bộ gen người. Theo đó theo một số phương án, các phân tử đánh dấu được dùng trong các thử nghiệm sinh học giải trình tự ở các mẫu của người cần phải có độ dài ít nhất là 30 bp.

Việc chọn độ dài của các phân tử đánh dấu được xác định sơ bộ bằng công nghệ giải trình tự được dùng để xác minh tính nguyên vẹn của mẫu nguồn. Độ dài của các axit nucleic bộ gen mẫu đang được giải trình tự có thể cũng được xem xét. Ví dụ, một số công nghệ giải trình tự sử dụng kỹ thuật khuếch đại vô tính các polynucleotit, việc này có thể yêu cầu rằng các polynucleotit bộ gen mà sẽ được khuếch đại vô tính phải có độ dài nhỏ nhất. Ví dụ, việc giải trình tự sử dụng bộ phân tích trình tự Illumina GAII bao gồm khuếch đại vô tính trong ống nghiệm bằng phương pháp PCR bắc cầu (còn gọi là khuếch đại nhóm) của các polynucleotit có độ dài tối thiểu là 110bp, trình tự nối được nối để tạo ra axit nucleic có ít nhất 200 bp và ít hơn 600 bp có thể được khuếch đại vô tính và được giải trình tự. Theo một số phương án, độ dài của các phân tử đánh dấu được nối trình tự nối nằm giữa khoảng 200bp và khoảng 600bp, giữa khoảng 250bp và 550bp, giữa khoảng 300bp và 500bp, hoặc giữa khoảng 350 và 450. Theo các phương án khác, độ dài của các phân tử đánh dấu được nối trình tự nối là khoảng 200bp. Ví dụ, khi giải trình tự cfADN của bào thai có trong mẫu mẹ, độ dài của các phân tử đánh dấu có thể được chọn giống với độ dài của các phân tử cfADN của bào

thai. Do đó, theo một phương án, độ dài của các phân tử đánh dấu được dùng trong thí nghiệm bao gồm bước giải trình tự hoàn toàn song song của cfADN trong mẫu mẹ để xác định sự có hay không có đột biến lệch bội nhiễm sắc thể ở bào thai, có thể là khoảng 150 bp, khoảng 160bp, 170 bp, khoảng 1180bp, khoảng 190bp hoặc khoảng 200bp; tốt hơn là, các phân tử đánh dấu là khoảng 170 pp. Các cách giải trình tự khác, ví dụ, giải trình tự SOLiD, giải trình tự Polony và giải trình tự 454 sử dụng PCR nhũ tương để khuếch đại vô tính các phân tử ADN để giải trình tự, và mỗi công nghệ quy định độ dài tối thiểu và tối đa của các phân tử cần khuếch đại. Độ dài của các phân tử đánh dấu cần giải trình tự như các axit nucleic được khuếch đại vô tính có thể lên đến khoảng 600bp. Theo một số phương án, độ dài của các phân tử đánh dấu cần giải trình tự có thể lớn hơn 600bp.

Các công nghệ giải trình tự phân tử đơn mà không sử dụng kỹ thuật khuếch đại vô tính của các phân tử, và có thể giải trình tự các axit nucleic trên khoảng độ dài khuôn rất rộng, trong hầu hết các trường hợp không yêu cầu rằng các phân tử cần giải trình tự phải có độ dài cụ thể. Tuy nhiên, số lượng trình tự trên một khối lượng đơn vị tùy thuộc vào số lượng các nhóm hydroxyl đầu 3', và do đó các khuôn tương đối ngắn để giải trình tự sẽ hiệu quả hơn khuôn dài. Nếu bắt đầu với các axit nucleic dài hơn 1000 nt, thường là có lợi khi cắt các axit nucleic thành độ dài trung bình từ 100 đến 200 nt sao cho có nhiều thông tin trình tự có thể được tạo từ cùng một lượng axit nucleic. Do đó, độ dài của các phân tử đánh dấu có thể nằm trong khoảng từ mười bazơ đến hàng nghìn bazơ. Độ dài của các phân tử đánh dấu được dùng để giải trình tự phân tử đơn có thể lên đến khoảng 25bp, lên đến khoảng 50bp, lên đến khoảng 75bp, lên đến khoảng 100bp, lên đến khoảng 200bp, lên đến khoảng 300bp, lên đến khoảng 400bp, lên đến khoảng 500bp, lên đến khoảng 600bp, lên đến khoảng 700bp, lên đến khoảng 800 bp, lên đến khoảng 900bp, lên đến khoảng 1000bp, hoặc dài hơn.

Độ dài được chọn cho phân tử đánh dấu còn được xác định bằng độ dài của axit nucleic bộ gen đang được giải trình tự. Ví dụ, cfADN lưu thông trong dòng máu của người dưới dạng các đoạn gen gồm ADN bộ gen tế bào. Các phân tử cfADN bào thai tìm thấy trong huyết tương của sản phụ thường ngắn hơn các phân tử cfADN của người mẹ (Chan et al., Clin Chem 50:8892 [2004]). Việc cắt phân đoạn kích thước của ADN bào thai đang lưu thông đã khẳng định rằng độ dài trung bình của các đoạn ADN của bào thai đang lưu thông là <300 bp, trong khi ADN của người mẹ được ước tính là giữa khoảng 0,5 và 1 Kb (Li et al., Clin Chem, 50:

1002-1011 [2004]). Số liệu này khớp với số liệu trong tài liệu của Fan và các cộng sự, bằng cách sử dụng kỹ thuật NGS, họ đã xác định rằng cfADN ở bào thai hiếm khi >340bp (Fan et al., Clin Chem 56:1279-1286 [2010]). ADN lấy từ nước tiểu bằng phương pháp chuẩn dựa trên silic oxit gồm có hai đoạn, ADN trọng lượng phân tử cao, hình thành từ các tế bào được cắt và đoạn có trọng lượng phân tử thấp (150-250 cặp bazơ) của ADN transrenal (Tr-ADN) (Botezatu et al., Clin Chem. 46: 1078-1084, 2000; and Su et al., J Mol. Diagn. 6: 101-107, 2004). Việc ứng dụng kỹ thuật mới được phát triển để tách các axit nucleic ngoài tế bào từ dịch cơ thể nhằm tách các axit nucleic transrenal cho thấy trong nước tiểu có các đoạn ADN và ARN ngắn hơn nhiều so với 150 cặp bazơ (công bố đơn sáng chế Mỹ số 20080139801). Theo các phương án, trong đó cfADN là axit nucleic bộ gen được giải trình tự, các phân tử đánh dấu được chọn có thể lên đến khoảng độ dài của cfADN. Ví dụ, độ dài của các phân tử đánh dấu được dùng trong các mẫu cfADN ở người mẹ cần giải trình tự dưới dạng các phân tử axit nucleic đơn hoặc các axit nucleic được khuếch đại vô tính có thể nằm trong khoảng giữa 100 bp và 600. Theo các phương án khác, các axit nucleic bộ gen mẫu là các đoạn gồm các phân tử lớn hơn. Ví dụ, axit nucleic bộ gen mẫu được giải trình tự là ADN tế bào được phân đoạn. Theo các phương án, khi ADN tế bào được phân đoạn được giải trình tự, độ dài của các phân tử đánh dấu có thể lên đến độ dài của các đoạn ADN. Theo một số phương án, độ dài của các phân tử đánh dấu bằng ít nhất là độ dài tối thiểu yêu cầu để ánh xạ đoạn đọc trình tự duy nhất đến bộ gen chuẩn phù hợp. Theo các phương án khác, độ dài của các phân tử đánh dấu bằng độ dài tối thiểu yêu cầu để loại trừ các phân tử đánh dấu để khỏi bị ánh xạ đến bộ gen chuẩn làm mẫu.

Ngoài ra, các phân tử đánh dấu có thể được dùng để xác nhận các mẫu không được thử nghiệm bằng quá trình giải trình tự axit nucleic, và có thể được xác nhận bằng kỹ thuật sinh học thông thường ngoài kỹ thuật giải trình tự, ví dụ, PCR thời gian thực.

Mẫu đối chứng (ví dụ, các mẫu chứng dương trong quá trình để giải trình tự và/hoặc phân tích).

Theo các phương án khác nhau, các trình tự đánh dấu được đưa vào các mẫu, ví dụ, như được mô tả trên đây, có thể đóng vai trò làm các đối chứng dương để xác nhận độ chính xác và tính hiệu quả của việc giải trình tự và việc xử lý và phân tích sau đó.

Theo đó, sáng chế đề xuất các chế phẩm và phương pháp cung cấp mẫu chứng dương trong quá trình (in-process positive control - IPC) để giải trình tự ADN trong mẫu. Theo một số phương án, sáng chế đề xuất các mẫu chứng dương để giải trình tự cfADN trong mẫu bao gồm hỗn hợp của các bộ gen. IPC có thể được dùng để liên kết các biến đổi đường cơ sở về thông tin trình tự thu được từ các tập mẫu khác nhau, ví dụ, các mẫu được giải trình tự ở các thời điểm khác nhau trên các chu trình giải trình tự khác nhau. Do đó, ví dụ, IPC có thể liên kết thông tin trình tự thu được cho mẫu thử nghiệm của mẹ với thông tin trình tự thu được từ tập mẫu đạt tiêu chuẩn đã được giải trình tự ở thời điểm khác.

Tương tự, trong trường hợp phân tích phân đoạn, IPC có thể liên kết thông tin trình tự thu được từ đối tượng đối với (các) phân đoạn cụ thể với trình tự thu được từ tập mẫu đạt tiêu chuẩn (của các trình tự tương tự) đã được giải trình tự ở thời điểm khác. Theo một số phương án, IPC có thể liên kết thông tin trình tự thu được từ đối tượng về các locut cụ thể liên quan đến bệnh ung thư với thông tin trình tự thu được từ tập mẫu đạt tiêu chuẩn (ví dụ, từ quá trình khuếch đại/xóa đoạn đã biết, và tương tự).

Ngoài ra, các IPC có thể được dùng làm mẫu đánh dấu để theo vết (các) mẫu theo dõi trong quy trình giải trình tự. Các IPC còn có thể cung cấp giá trị định lượng trình tự dương định tính, ví dụ, NCV, cho một hoặc nhiều đột biến lệch bội của các nhiễm sắc thể quan tâm, ví dụ, thể ba nhiễm 21, thể ba nhiễm 13, thể ba nhiễm 18 để cung cấp sự giải thích thích hợp, và để đảm bảo độ tin cậy và chính xác của dữ liệu. Theo một số phương án các IPC có thể được tạo bao gồm các axit nucleic từ các bộ gen của đực và cái để cung cấp các định lượng cho các nhiễm sắc thể X và Y trong mẫu mẹ nhằm xác định bào thai có phải là nam hay không.

Loại và số lượng các mẫu chứng trong quy trình tùy thuộc vào loại hoặc bản chất của thử nghiệm cần làm. Ví dụ, đối với thử nghiệm yêu cầu giải trình tự ADN từ mẫu bao gồm hỗn hợp của các bộ gen để xác định có tồn tại đột biến lệch bội nhiễm sắc thể hay không, mẫu chứng trong quy trình có thể bao gồm ADN thu được từ mẫu đã biết là chứa đột biến lệch bội nhiễm sắc thể giống với loại đang được kiểm tra. Theo một số phương án, IPC bao gồm ADN từ mẫu đã biết là có đột biến lệch bội của nhiễm sắc thể quan tâm. Ví dụ, IPC để kiểm tra xác định bào thai có thể ba nhiễm hay không, ví dụ, thể ba nhiễm 21, trong mẫu mẹ có ADN thu được từ một cá thể có thể ba nhiễm 21. Theo một số phương án, IPC bao gồm hỗn hợp ADN thu được từ hai hoặc nhiều cá thể có các đột biến lệch bội khác nhau. Ví dụ, để kiểm tra xác

định xem có hay không có thể ba nhiễm 13, thể ba nhiễm 18, thể ba nhiễm 21, và thể đơn nhiễm X, IPC bao gồm tổ hợp các mẫu ADN thu được từ các sản phụ mang thai có một trong số các thể ba nhiễm đang được kiểm tra. Ngoài ra để hoàn thiện các đột biến lệch bội của nhiễm sắc thể, các IPC có thể được tạo để cung cấp mẫu chứng dương cho các thử nghiệm xác định có hay không có đột biến lệch bội một phần.

IPC làm mẫu chứng cho việc phát hiện đột biến lệch bội đơn có thể được tạo ra bằng cách sử dụng hỗn hợp của ADN bộ gen tế bào thu được từ hai đối tượng trong đó một đối tượng là có bộ gen lệch bội. Ví dụ, IPC được tạo làm mẫu chứng cho thử nghiệm xác định thể ba nhiễm ở bào thai, ví dụ, thể ba nhiễm 21, có thể được tạo ra bằng cách kết hợp ADN bộ gen từ đối tượng đực hoặc cái mang thể ba nhiễm với ADN bộ gen của đối tượng cái đã biết là không mang thể ba nhiễm. ADN bộ gen có thể được tách từ các tế bào của cả hai đối tượng, và được cắt để tạo các đoạn có khoảng 100 - 400 bp, giữa khoảng 150 - 350 bp, hoặc giữa khoảng 200-300 bp để mô phỏng các đoạn cfADN lưu thông trong các mẫu mẹ. Tỷ lệ của ADN được phân đoạn từ đối tượng có đột biến lệch bội, ví dụ, thể ba nhiễm 21, được chọn để mô phỏng tỷ lệ của cfADN lưu thông ở bào thai tìm thấy trong các mẫu mẹ để tạo ra IPC bao gồm hỗn hợp của ADN được phân đoạn bao gồm khoảng 5%, khoảng 10%, khoảng 15%, khoảng 20%, khoảng 25%, khoảng 30%, trong số ADN từ đối tượng mang đột biến lệch bội. IPC có thể bao gồm ADN từ các đối tượng khác nhau mang đột biến lệch bội khác nhau. Ví dụ, IPC có thể bao gồm khoảng 80% số ADN cái không bị bệnh, và 20% còn lại có thể là ADN từ ba đối tượng khác nhau mang thể ba nhiễm 21, thể ba nhiễm 13, và thể ba nhiễm 18. Hỗn hợp của ADN được phân đoạn được điều chế để giải trình tự. Việc xử lý hỗn hợp của ADN được phân đoạn có thể bao gồm bước tạo thư viện giải trình tự, có thể được giải trình tự bằng cách sử dụng các phương pháp song song ở bất kỳ theo kiểu đơn công hoặc đa công. Các dung dịch gốc của IPC bộ gen có thể được lưu trữ và được dùng trong nhiều thử nghiệm chuẩn đoán.

Tùy ý, IPC có thể được tạo ra bằng cách sử dụng cfADN thu được từ người mẹ đã biết là có thai mắc một thể đột biến lệch bội nhiễm sắc thể đã biết. Ví dụ, cfADN có thể thu được từ phụ nữ mang thai mắc thể ba nhiễm 21. cfADN được tách từ mẫu mẹ, và dòng hóa vào vector vi khuẩn và nuôi trong vi khuẩn để cung cấp nguồn IPC liên tục. ADN có thể được tách ra từ vector vi khuẩn bằng cách sử dụng các enzym cắt giới hạn. Theo cách khác, cfADN được

dòng hóa có thể được khuếch đại bằng, ví dụ, PCR. IPC ADN có thể được xử lý để giải trình tự trong các chu trình giống nhau do cfADN từ các mẫu thử nghiệm cần được phân tích về sự có mặt hoặc không có mặt của các đột biến lệch bội của nhiễm sắc thể.

Mặc dù việc tạo các IPC được mô tả trên đây đối với các thể ba nhiễm, cần phải hiểu rằng các IPC có thể được tạo ra để phản ánh thể đột biến lệch bội một phần khác bao gồm, ví dụ, khuếch đại đoạn khác nhau và/hoặc mất đoạn. Do đó, ví dụ, trong khi nhiều bệnh ung thư đã biết là đi kèm với các thể khuếch đại cụ thể (ví dụ, ung thư vú đi kèm với 20Q13), các IPC có thể được tạo ra kết hợp với các loại khuếch đại đã biết này.

Phương pháp giải trình tự

Như được chỉ ra trên đây, các mẫu đã được điều chế, (ví dụ, Thư Viện Giải Trình Tự) được giải trình tự như là một phần của quy trình xác định (các) biến thể số lượng bản sao. Có thể sử dụng kỹ thuật bất kỳ trong số các kỹ thuật giải trình tự.

Một số kỹ thuật giải trình tự là có sẵn trên thị trường, như hệ thống giải trình tự bằng lai ghép (sequencing-by-hybridization platform) của Affymetrix Inc. (Sunnyvale, CA) và hệ thống giải trình tự bằng tổng hợp (sequencing-by-synthesis platforms) của hãng 454 Life Sciences (Bradford, CT), Illumina/Solexa (Hayward, CA) và Helicos Biosciences (Cambridge, MA), và hệ thống giải trình tự bằng cách nối (sequencing-by-ligation platform) của hãng Applied Biosystems (Foster City, CA), như được mô tả dưới đây. Ngoài việc giải trình tự phân tử đơn được thực hiện bằng cách sử dụng kỹ thuật giải trình tự bằng tổng hợp của hãng Helicos Biosciences, các kỹ thuật giải trình tự phân tử đơn khác bao gồm, nhưng không chỉ giới hạn ở, kỹ thuật SMRTTM của hãng Pacific Biosciences, kỹ thuật ION TORRENTTM, và kỹ thuật giải trình tự nanopore được phát triển bởi, ví dụ, hãng Oxford Nanopore Technologies.

Mặc dù phương pháp Sanger tự động được cho là công nghệ “thế hệ đầu tiên”, kỹ thuật giải trình tự Sanger bao gồm giải trình tự Sanger tự động, cũng có thể được sử dụng trong các phương pháp được mô tả ở đây. Các phương pháp giải trình tự thích hợp khác bao gồm, nhưng không chỉ giới hạn ở các kỹ thuật chụp ảnh axit nucleic, ví dụ, kính hiển vi lực nguyên tử (atomic force microscopy-AFM) hoặc kính hiển vi điện tử truyền qua (transmission electron microscopy-TEM). Các kỹ thuật giải trình tự minh họa được mô tả chi tiết hơn dưới đây.

Theo một phương án minh họa, nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm các bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu lấy từ người mẹ, cfADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, sử dụng kỹ thuật giải trình tự bằng tổng hợp của Illumina và kỹ thuật hóa học giải trình tự dựa trên ký tự kết thúc khả nghịch (ví dụ như được mô tả trong tài liệu Bentley et al., *Nature* 6:53-59). ADN mẫu có thể là ADN bộ gen, ví dụ, ADN tế bào hoặc cfADN. Theo một số phương án, ADN bộ gen từ các tế bào được phân lập được dùng làm mẫu, và nó được phân đoạn thành các độ dài gồm một vài trăm cặp bazơ. Theo các phương án khác, cfADN được dùng làm mẫu, và sự phân đoạn là không cần thiết vì cfADN tồn tại dưới dạng các đoạn ngắn. Ví dụ cfADN dạng bào phôi di chuyển trong dòng máu dưới dạng các phân đoạn có độ dài xấp xỉ 170 cặp bazơ (bp) (Fan et al., *Clin Chem* 56:1279-1286), và không cần phân đoạn ADN trước khi giải trình tự. Kỹ thuật giải trình tự của Illumina dựa vào sự gắn ADN bộ gen đã phân đoạn vào bề mặt phẳng, trong suốt về mặt quang học mà trên đó các mỏ neo (anchor) oligonucleotit được gắn kết. ADN khuôn được sửa chữa ở đầu để tạo ra các đầu bằng 5' được phosphorylat hóa, và hoạt tính polymeraza của đoạn Klenow được sử dụng để thêm một bazơ A vào đầu 3' của các đoạn ADN phosphorylat hóa đầu bằng. Sự bổ sung này tạo ra các đoạn ADN để nối vào các trình tự nối (adapter) oligonucleotit, mà có phần nhô (overhang) của bazơ T đơn ở đầu 3' của nó để gia tăng hiệu quả nối. Các oligonucleotit của trình tự nối là sợi bổ sung với các oligo mỏ neo trong cuvet dòng chảy (không nhầm lẫn với các trình tự đọc mỏ neo/có mỏ neo trong phân tích lập-mở rộng). Trong các điều kiện pha loãng giới hạn, ADN khuôn sợi đơn, đã được biến đổi trình tự nối được đưa vào cuvet dòng chảy và làm bất động bằng cách lai với các oligo mỏ neo. Các đoạn ADN đã được gắn được kéo dài và khuếch đại bậc cầu để tạo ra cuvet dòng chảy giải trình tự mật độ siêu cao với hàng trăm triệu cụm (cluster), mỗi cụm chứa khoảng 1.000 bản sao của cùng một khuôn. Theo một phương án, ADN bộ gen đã phân đoạn ngẫu nhiên được khuếch đại bằng cách sử dụng PCR trước khi nó được đưa vào khuếch đại cụm. Theo cách khác, ngân hàng gen không khuếch đại (ví dụ, không PCR) được sử dụng, và bộ gen ADN đã phân đoạn ngẫu nhiên được làm giàu bằng cách chỉ sử dụng khuếch đại cụm (Kozarewa et al., *Nature Methods* 6:291-295 [2009]). Các khuôn được giải trình tự bằng cách sử dụng kỹ thuật giải trình tự bằng tổng hợp ADN bốn màu mạnh (robust four-color DNA sequencing -by-synthesis technology) mà sử

dụng các đầu khả nghịch với các thuốc nhuộm huỳnh quang có thể loại bỏ được. Việc dò huỳnh quang độ nhạy cao được thực hiện bằng cách sử dụng chiếu xạ laze và quang phản xạ toàn phần. Các trình tự đọc ngắn khoảng 10 đến vài trăm cặp bazơ được canh thẳng với bộ gen tham chiếu và ánh xạ duy nhất của các trình tự đọc ngắn đến bộ gen tham chiếu được nhận diện bằng cách sử dụng phần mềm đường dẫn phân tích dữ liệu được phát triển riêng. Sau khi hoàn thành trình tự đọc đầu tiên, các khuôn có thể được tái sinh tại chỗ để cho phép trình tự đọc thứ hai từ đầu đối diện của các đoạn này. Do đó, kỹ thuật giải trình tự một đầu hoặc đầu bắt cặp của các đoạn ADN có thể được sử dụng.

Các phương án khác nhau của sáng chế có thể sử dụng kỹ thuật giải trình tự bằng tổng hợp mà cho phép giải trình tự đầu bắt cặp. Theo một số phương án, hệ thống giải trình tự bằng tổng hợp của Illumina bao gồm các đoạn tạo cụm. Việc tạo cụm là quá trình trong đó mỗi đoạn phân tử được khuếch đại đẳng nhiệt. Theo một số phương án, như ví dụ được mô tả ở đây, đoạn này có hai trình tự nối khác nhau được gắn vào hai đầu của đoạn, các trình tự nối này cho phép đoạn lai với hai oligo khác nhau trên bề mặt của làn cuvet dòng chảy. Đoạn này còn bao gồm hoặc được liên kết với hai trình tự chỉ thị ở hai đầu của đoạn, các trình tự chỉ thị này tạo ra các nhãn để nhận diện các mẫu khác nhau trong kỹ thuật giải trình tự đa thành phần. Trong một số hệ thống giải trình tự, đoạn cần được giải trình tự còn được gọi là đoạn chèn.

Theo một số phương án, cuvet dòng chảy để tạo cụm trong hệ thống Illumina là phiên kính có các làn. Mỗi làn là một kênh kính được phủ lớp gồm hai loại oligo. Quá trình lai được kích hoạt bởi loại thứ nhất trong số hai loại oligo trên bề mặt. Oligo này là bổ sung cho trình tự nối thứ nhất trên một đầu của đoạn. Polymeraza tạo ra sợi bổ sung của đoạn được lai. Phân tử sợi kép được làm biến tính, và sợi khuôn ban đầu được rửa sạch. Sợi còn lại, cùng với nhiều sợi còn lại khác, được khuếch đại đơn dòng bằng khuếch đại bắc cầu.

Trong khuếch đại bắc cầu, sợi gấp lại, và vùng trình tự nối thứ hai trên đầu thứ hai của sợi này lai với loại oligo thứ hai trên bề mặt cuvet dòng chảy. Polymeraza tạo ra sợi bổ sung, tạo thành phân tử cầu nối sợi kép. Phân tử sợi kép này được làm biến tính tạo ra hai phân tử sợi đơn được nối vào cuvet dòng chảy thông qua hai oligo khác nhau. Sau đó, quá trình này được lặp đi lặp lại, và xảy ra đồng thời đối với hàng triệu cụm tạo ra sự khuếch đại đơn dòng của

tất cả các đoạn. Sau quá trình khuếch đại bậc cầu, các sợi ngược được cắt và rửa sạch, chỉ còn lại các sợi xuôi. Các đầu 3' được phong bế để ngăn chặn việc gắn nối không mong muốn.

Sau khi tạo cụm, việc giải trình tự bắt đầu với việc kéo dài đoạn mỗi giải trình tự thứ nhất để tạo ra trình tự đọc thứ nhất. Với mỗi chu trình, các nucleotit được đánh dấu huỳnh quang cạnh tranh để bổ sung vào chuỗi đang phát triển. Chỉ một nucleotit được đưa vào trên cơ sở trình tự mẫu. Sau khi bổ sung từng nucleotit, cụm này được kích thích bởi nguồn sáng, và tín hiệu huỳnh quang đặc trưng được phát. Số lượng chu trình xác định chiều dài trình tự đọc. Bước sóng phát xạ và cường độ tín hiệu xác định vị trí gắn bazơ. Đối với một cụm cho trước tất cả các sợi giống nhau đều được đọc một cách đồng thời. Hàng trăm triệu cụm được giải trình tự theo cùng một cách gần như đồng thời. Khi kết thúc trình tự đọc thứ nhất, sản phẩm đọc được rửa sạch.

Trong bước tiếp theo của quay trình bao gồm hai đoạn mỗi chỉ thị, đoạn mỗi chỉ thị 1 Được đưa vào và lai với vùng chỉ thị 1 trên mẫu. Các vùng chỉ thị tạo ra sự đồng nhất của các đoạn, mà là hữu ích cho các mẫu tách kênh trong quy trình giải trình tự đa kênh. Trình tự đọc chỉ thị 1 được tạo ra tương tự với trình tự đọc thứ nhất. Sau khi kết thúc trình tự đọc chỉ thị 1, sản phẩm đọc được rửa sạch và đầu 3' của sợi này được loại bỏ nhóm bảo vệ. Sau đó sợi khuôn gấp lại và liên kết với oligo thứ hai trên cuvet dòng chảy. Trình tự chỉ thị 2 được đọc theo cùng cách như trình tự chỉ thị 1. Sau đó sản phẩm đọc chỉ thị 2 được rửa sạch sau khi hoàn thành bước này.

Sau khi đọc hai trình tự chỉ thị, trình tự đọc 2 bắt đầu bằng cách sử dụng polymeraza để kéo dài oligo cuvet dòng chảy thứ hai, tạo thành cầu nối sợi kép. ADN sợi kép này được làm biến tính, và đầu 3' được phong bế. Sợi xuôi ban đầu được cắt và rửa sạch, còn lại sợi ngược. Trình tự đọc 2 bắt đầu bằng việc đưa vào đoạn mỗi giải trình tự cho trình tự đọc 2. Giống như trình tự đọc 1, các bước giải trình tự được lặp lại cho đến khi đạt được chiều dài mong muốn. Sản phẩm trình tự đọc 2 được rửa sạch. Toàn bộ quá trình này tạo ra hàng triệu trình tự đọc, thể hiện tất cả các đoạn. Các trình tự từ các thư viện mẫu tập hợp lại được tách ra dựa trên các chỉ thị duy nhất được đưa vào trong khi chuẩn bị mẫu. Đối với mỗi mẫu, các trình tự đọc của các đoạn kéo dài tương tự của các vị trí gắn bazơ được tạo cụm cục bộ. Các trình tự đọc xuôi và

ngược được bắt cặp tạo thành các trình tự tiếp giáp. Các trình tự tiếp giáp này được canh thẳng với bộ gen tham chiếu để nhận diện biến thể.

Ví dụ về giải trình tự bằng tổng hợp mô tả trên đây liên quan đến các trình tự đọc có đầu bắt cặp, được sử dụng theo nhiều trong số các phương án của các phương pháp được bộc lộ. Giải trình tự có đầu bắt cặp bao gồm 2 trình tự đọc từ hai đầu của đoạn. Khi cặp trình tự đọc được ánh xạ đến các trình tự đối chứng, thì khoảng cách cặp bazơ giữa hai trình tự đọc có thể được xác định, khoảng cách này sau đó có thể được sử dụng để xác định độ dài của các đoạn mà từ đó các trình tự đọc thu được. Trong một số trường hợp, đoạn phân làm hai bin sẽ có một đoạn đầu của trình tự đọc đầu cặp được căn chỉnh với một bin và trình tự đọc còn lại cho bin liền kề. Điều này trở nên hiếm hơn khi các bin dài hơn hoặc các trình tự đọc ngắn hơn. Các phương pháp khác nhau có thể được sử dụng để giải thích về tư cách thành viên bin của các đoạn này. Ví dụ, chúng có thể được lược bỏ trong việc xác định tần số kích thước đoạn của bin; chúng có thể được tính cho cả hai bin liền kề; chúng có thể được gán cho bin mà chứa đựng số lượng lớn hơn các cặp bazơ của hai bin; hoặc chúng có thể được gán cho cả hai bin với trọng lượng tương ứng với phần của các cặp bazơ ở mỗi bin.

Các trình tự đọc có đầu bắt cặp có thể sử dụng đoạn chèn có độ dài khác nhau (tức là, kích thước đoạn khác nhau cần được giải trình tự). Như ý nghĩa được mặc định trong bản mô tả này, các trình tự đọc có đầu bắt cặp được sử dụng để chỉ các trình tự đọc thu được từ các đoạn chèn khác nhau. Trong một số trường hợp, để phân biệt các trình tự đọc có đầu bắt cặp có đoạn chèn ngắn với các trình tự đọc có đầu bắt cặp có đoạn chèn dài, trình tự đọc có đầu bắt cặp có đoạn chèn dài còn được gọi là các trình tự đọc cặp liên kết. Theo một số phương án liên quan đến các trình tự đọc cặp liên kết, hai trình tự nối tiếp hợp biotin đầu tiên được gắn kết vào hai đầu của đoạn chèn tương đối dài (ví dụ, một vài kb). Các trình tự nối tiếp hợp biotin sau đó liên kết hai đầu của đoạn chèn để tạo ra phân tử hình vòng tròn. Đoạn còn chứa đựng các trình tự nối tiếp hợp biotin có thể sau đó được thu được bằng cách phân đoạn tiếp phân tử hình vòng tròn. Đoạn còn có hai đầu của đoạn gốc theo thứ tự trình tự ngược nhau có thể sau đó được giải trình tự bằng quy trình giống như đối với giải trình tự có đầu bắt cặp có đoạn chèn ngắn mô tả trên đây. Các chi tiết khác về giải trình tự bắt cặp bằng cách sử dụng

hệ thống Illumina được thể hiện trong công bố trực tuyến tại địa chỉ URL dưới đây, mà được đưa vào bằng cách viện dẫn trong tính toàn vẹn của nó:

res|.illumina|.com/documents/products/technotes/technote_nextera_matepair_data_processing. Các thông tin bổ sung về giải trình tự có đầu bắt cặp có thể được tìm thấy trong patent Mỹ số 7601499 và công bố patent Mỹ số 2012/0,053,063, mà được đưa vào bằng cách viện dẫn liên quan tới các tài liệu về các phương pháp và thiết bị giải trình tự có đầu bắt cặp.

Sau khi giải trình tự các đoạn ADN, các trình tự đọc có độ dài định trước, ví dụ, 100 bp, được ánh xạ hoặc canh thẳng theo bộ gen tham chiếu đã biết. Các trình tự đọc được ánh xạ hoặc canh thẳng và các vị trí tương ứng của chúng về các trình tự đối chứng còn được gọi là thẻ. Theo một phương án, giải trình tự bộ gen tham chiếu là giải trình tự NCB136/hg18, mà có sẵn trên địa chỉ [www ở genome chấm ucsc chấm edu/cgi-bin/hgGateway?org=Human&db=hg18&hgid=166260105](http://www.genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgid=166260105)). Theo cách khác, giải trình tự bộ gen tham chiếu là GRCh37/hg19, mà có sẵn trên địa chỉ [www ở genome chấm ucsc chấm edu/cgi-bin/hgGateway](http://www.genome.ucsc.edu/cgi-bin/hgGateway). Các nguồn thông tin trình tự công khai khác bao gồm GenBank, dbEST, dbSTS, EMBL (Phòng thí nghiệm sinh học phân tử của Châu Âu - European Molecular Biology Laboratory), và DDBJ (Ngân Hàng dữ liệu ADN Nhật Bản - ADN Databank of Japan). Các thuật toán máy tính có sẵn cho việc canh thẳng các chuỗi, bao gồm nhưng không giới hạn ở BLAST (Altschul et al., 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), FASTA (Person & Lipman, 1988), BOWTIE (Langmead et al., Bộ gen Biology 10:R25,1-R25,10), hoặc ELAND (Illumina, Inc., San Diego, CA, MỸ). Theo một phương án, một đầu của các bản sao khuếch đại đơn dòng của các phân tử cfADN huyết tương được giải trình tự và xử lý bởi phép phân tích canh thẳng tin sinh học cho thiết bị phân tích bộ gen Illumina, mà sử dụng phần mềm canh thẳng cơ sở dữ liệu nucleotit quy mô lớn hiệu quả (Efficient Large-Scale Alignment of Nucleotit Databases -ELAND).

Theo một phương án minh họa nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu lấy từ người mẹ, cfADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, sử dụng kỹ thuật giải trình tự đơn phân tử của kỹ thuật giải trình tự đơn phân tử chính xác Helicos (Helicos True Single Molecule Sequencing - tSMS) (ví dụ như

được mô tả trong tài liệu Harris T.D. et al., Science 320:106-109). Trong kỹ thuật tSMS, mẫu ADN được phân cắt thành các sợi gồm xấp xỉ 100 đến 200 nucleotit, và trình tự polyA được thêm vào đầu 3' của mỗi sợi ADN. Mỗi sợi được đánh dấu bằng cách bổ sung vào adenosin nucleotit được đánh dấu huỳnh quang. Các sợi ADN sau đó được lai vào cuvet dòng chảy, chứa hàng triệu vị trí bắt giữ oligo T được cố định vào bề mặt cuvet dòng chảy. Theo một số phương án nhất định, các mẫu có thể ở mật độ vào khoảng 100 triệu mẫu/cm². Cuvet dòng chảy sau đó được nạp vào dụng cụ, ví dụ, thiết bị giải trình tự HeliScope™, và thiết bị laser chiếu sáng bề mặt của cuvet dòng chảy, bộc lộ vị trí của mỗi mẫu. Camera CCD có thể ánh xạ vị trí của các khuôn trên bề mặt cuvet dòng chảy. Nhãn huỳnh quang của khuôn sau đó được phân cắt và rửa sạch. Phản ứng giải trình tự bắt đầu bằng cách đưa vào ADN polymeraza và nucleotit đánh dấu huỳnh quang. Axit oligo T nucleic dùng làm đoạn mồi. Polymeraza kết hợp các nucleotit đã đánh dấu vào đoạn mồi theo cách thức định hướng khuôn. Polymeraza và nucleotit không kết hợp được loại bỏ. Các khuôn có sự kết hợp có định hướng của nucleotit đánh dấu huỳnh quang được nhận biết bằng cách chụp ảnh bề mặt cuvet dòng chảy. Sau khi chụp ảnh, bước phân cắt loại bỏ nhãn huỳnh quang, và quy trình này được lặp lại với các nucleotit được gắn nhãn huỳnh quang khác cho tới khi đạt được độ dài trình tự đọc mong muốn. Thông tin trình tự được thu gom với mỗi bước bổ sung nucleotit. Việc giải trình tự toàn bộ bộ gen bằng các công nghệ giải trình tự đơn phân tử loại trừ hoặc thường xóa bỏ sự khuếch đại dựa trên PCR trong quá trình tạo thư viện giải trình tự, và các phương pháp cho phép đo trực tiếp mẫu chứ không phải đo bản sao của mẫu đó.

Theo một phương án minh họa khác nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu thử nghiệm lấy từ người mẹ, cfADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, bằng cách sử dụng giải trình tự 454 (Roche) (ví dụ như được mô tả trong tài liệu Margulies, M. et al. Nature 437:376-380). Giải trình tự 454 thường bao gồm hai bước. Ở bước thứ nhất, ADN được cắt xén thành nhiều đoạn xấp xỉ 300 đến 800 cặp bazơ, và các đoạn được tạo đầu bằng. Các trình tự nối oligonucleotit sau đó được nối ở các đầu của các đoạn. Các trình tự nối dùng làm các đoạn mồi để khuếch đại và giải trình tự cho các đoạn này. Các đoạn có thể được gắn vào các hạt bắt giữ ADN, ví dụ, các hạt được bọc bởi streptavidin nhờ sử dụng, ví dụ, trình tự nối B, chứa thẻ 5'-biotin. Các đoạn

gắn vào các hạt được khuếch đại PCR trong các giọt nhỏ của nhũ tương dầu-nước. Kết quả là có nhiều bản sao của các đoạn ADN được khuếch đại dòng vô tính trên mỗi hạt. Ở bước thứ hai, các hạt được bắt giữ trong các lỗ (ví dụ, các lỗ có kích thước giọt mực). Giải trình tự bằng tổng hợp (Pyrosequencing) được thực hiện song song trên mỗi đoạn ADN. Sự bổ sung vào một hoặc nhiều nucleotit tạo ra tín hiệu ánh sáng mà được ghi bởi camera CCD ở thiết bị giải trình tự. Cường độ tín hiệu là tỷ lệ với số lượng nucleotit được kết hợp. Giải trình tự bằng tổng hợp sử dụng pyrophosphat (pyrophosphate - PPi) mà được giải phóng ngay khi bổ sung nucleotit. PPi được biến đổi thành ATP bởi ATP sulfurylaza khi có mặt adenosin 5' phosphosulfat. Luciferaza sử dụng ATP để biến đổi luxiferin thành oxyluciferin, và phản ứng này tạo ra ánh sáng mà được đo và phân tích.

Theo một phương án minh họa khác nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, ctADN trong mẫu thử nghiệm lấy từ người mẹ, ctADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, bằng cách sử dụng công nghệ SOLiD™ (các hệ thống sin học ứng dụng - Applied Biosystems). Trong công nghệ giải trình tự bằng phương pháp nối SOLiD™, bộ gen ADN được cắt xén thành các đoạn, và các trình tự nối được gắn vào các đầu 5' và 3' của các đoạn để tạo thư viện đoạn. Theo cách khác, các trình tự nối bên trong có thể được đưa vào bằng cách nối các trình tự nối ở các đầu 5' và 3' của các đoạn, tạo vòng cho các đoạn, cắt đoạn được tạo vòng để tạo ra trình tự nối bên trong, và gắn các trình tự nối vào các đầu 5' và 3' của các đoạn thu được để tạo thư viện bắt cặp. Tiếp đó, các tập hợp hạt dòng vô tính được chuẩn bị trong các lò vi phản ứng chứa các hạt, các đoạn mồi, khuôn, và các thành phần PCR. Tiếp theo PCR, các khuôn được làm biến tính và các hạt được làm giàu để tách các hạt với có các khuôn được kéo dài. Các khuôn ở các hạt đã chọn được trải qua sự biến đổi 3' mà cho phép liên kết với phiên kính thủy tinh. Trình tự có thể được xác định bằng cách lai và nối theo trình tự của các oligonucleotit ngẫu nhiên một phần với bazơ được xác định ở trung tâm (hoặc cặp bazơ) được nhận biết bởi huỳnh quang đặc hiệu. Sau khi màu sắc được ghi, oligonucleotit đã nối được phân cắt và loại bỏ và quy trình này sau đó được lặp lại.

Theo một phương án minh họa khác nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, ctADN trong mẫu thử nghiệm lấy từ người mẹ, ctADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, sử dụng kỹ thuật giải trình tự đơn phân tử theo thời gian thực (the single molecule, real-time sequencing technology - SMRT™) của Pacific Biosciences. Trong công nghệ giải trình tự SMRT, sự kết hợp liên tục các nucleotit được gắn nhãn bằng thuốc nhuộm được chụp ảnh trong suốt quá trình tổng hợp ADN. Các đơn phân tử ADN polymeraza được gắn vào bề mặt đáy của các bộ dò bước sóng ở chế độ zero (các bộ dò ZMW) riêng biệt mà thu nhận các thông tin trình tự trong khi các nucleotit được liên kết phospho đang được kết hợp vào sợi mỗi đang phát triển. Bộ dò ZMW bao gồm cấu trúc giới hạn mà cho phép quan sát thấy sự kết hợp của nucleotit đơn lẻ bởi ADN polymeraza vào nền gồm các nucleotit huỳnh quang mà nhanh chóng khuếch tán ở phần ngoài của ZMW (ví dụ, tính bằng micrô giây). Nó thường mất vài mili giây để kết hợp nucleotit vào sợi đang phát triển. Trong suốt thời gian này, nhãn huỳnh quang được kích thích và tạo ra tín hiệu huỳnh quang, và thẻ huỳnh quang được tách ra. Phép đo huỳnh quang tương ứng của thuốc nhuộm cho biết bazơ nào được kết hợp. Quy trình này được lặp lại để tạo ra trình tự.

Theo một phương án minh họa khác nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu thử nghiệm lấy từ người mẹ, cfADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, bằng cách sử dụng kỹ thuật giải trình tự có lỗ cỡ nano (ví dụ như được mô tả trong tài liệu Soni GV và Meller A. Clin Chem 53: 1996-2001). Các kỹ thuật phân tích giải trình tự ADN có lỗ cỡ nano được phát triển bởi nhiều công ty, bao gồm, ví dụ, Oxford Nanopore Technologies (Oxford, Vương quốc Anh), Sequenom, NABsys, và tương tự. Giải trình tự lỗ cỡ nano là công nghệ giải trình tự đơn phân tử nhờ đó đơn phân tử của ADN được giải trình tự một cách trực tiếp khi nó đi qua lỗ cỡ nano. Lỗ cỡ nano là một lỗ nhỏ, thường có đường kính vào khoảng 1 nanomet. Ngâm lỗ cỡ nano vào trong chất lỏng dẫn điện và đặt điện thế (điện áp) trên khắp lỗ cơ nano tạo ra một dòng điện nhẹ do sự truyền dẫn các ion qua lỗ cỡ nano. Lượng dòng điện truyền qua này là nhạy với kích thước và hình dạng của lỗ cỡ nano. Vì phân tử ADN đi qua lỗ cỡ nano, nên mỗi nucleotit trên phân tử ADN làm tắc nghẽn lỗ cỡ nano ở một mức độ khác, làm thay đổi cường độ của dòng điện

qua lỗ cỡ nano ở các mức độ khác nhau. Vì vậy, sự thay đổi về dòng điện khi phân tử ADN đi qua lỗ cỡ nano tạo ra trình tự đọc của trình tự ADN.

Theo một phương án minh họa khác nhưng không giới hạn sáng chế, các phương pháp được mô tả ở đây bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu thử nghiệm lấy từ người mẹ, cfADN hoặc ADN tế bào ở đối tượng được sàng lọc ung thư, và tương tự, bằng cách sử dụng mảng điện trở có tách dụng trường nhạy hóa học (chemical-sensitive field effect transistor - chemFET) (ví dụ, như được mô tả trong công bố đơn sáng chế Mỹ 2009/0026082). Trong một ví dụ về kỹ thuật này, các phân tử ADN có thể được đưa vào các khoang phản ứng, và các phân tử khuôn có thể được lai với đoạn môi giải trình tự được gắn kết với polymeraza. Sự kết hợp của một hoặc nhiều triphosphat vào một sợi axit nucleic mới ở đầu 3' của đoạn môi giải trình tự có thể được nhận biết dưới dạng sự thay đổi về dòng điện bởi chemFET. Mảng có thể có nhiều cảm biến chemFET. Trong một ví dụ khác, các axit nucleic đơn lẻ có thể được gắn vào các hạt, và các axit nucleic có thể được khuếch đại trên hạt, và các hạt riêng lẻ có thể được chuyển vào các khoang phản ứng riêng lẻ trên mảng chemFET, với mỗi khoang có cảm biến chemFET, và các axit nucleic có thể được giải trình tự.

Theo một phương án khác, phương pháp theo sáng chế bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu thử nghiệm lấy từ người mẹ, sử dụng kính hiển vi điện tử truyền qua (transmission electron microscopy - TEM). Phương pháp, gọi là phương pháp truyền nano nhanh sắp đặt phân tử riêng lẻ (Individual Molecule Placement Rapid Nano Transfer - IMPRNT), bao gồm bước sử dụng hình ảnh kính hiển vi điện tử truyền qua độ phân giải đơn nguyên tử về ADN trọng lượng phân tử cao (150kb hoặc lớn hơn) được đánh dấu chọn lọc với các đánh dấu nguyên tử nặng và bố trí các phân tử này trên các màng siêu mỏng ở các mảng song song siêu đậm đặc (3nm sợi - sợi) với khoảng cách bazơ - bazơ thống nhất. Kính hiển vi điện tử được sử dụng để chụp ảnh cho các phân tử trên các màng để xác định vị trí của chất đánh dấu nguyên tử nặng và để trích xuất thông tin trình tự bazơ từ ADN. Phương pháp này còn được mô tả trong công bố sáng chế quốc tế WO 2009/046445. Phương pháp này cho phép giải trình tự các bộ gen người hoàn chỉnh trong ít hơn mười phút.

Theo một phương án khác, công nghệ giải trình tự ADN là giải trình tự đơn phân tử Ion Torrent, mà kết hợp với công nghệ bán dẫn với hóa học giải trình tự đơn giản để trực tiếp dịch các thông tin mã hóa hóa học (A, C, G, T) thành các thông tin dạng số (0, 1) trên chip bán dẫn. Về bản chất, khi nucleotit được kết hợp vào sợi ADN bằng polymeraza, thì ion hydro được giải phóng dưới dạng sản phẩm phụ. Ion Torrent sử dụng mảng mật độ cao gồm các lỗ được tạo cỡ micro bằng máy để thực hiện quy trình sinh hóa này theo cách song song ô ạt. Mỗi lỗ chứa một phân tử ADN khác. Bên dưới các lỗ là lớp nhạy ion và bên dưới lớp đó là cảm biến ion. Khi nucleotit, ví dụ C, được thêm vào khuôn ADN và sau đó được kết hợp vào sợi ADN, ion hydro sẽ được giải phóng. Điện tích từ ion đó sẽ làm thay đổi độ pH của dung dịch, điều này có thể được phát hiện thấy bởi cảm biến ion của Ion Torrent. Thiết bị giải trình tự - về cơ bản là dụng cụ đo độ pH trạng thái rắn nhỏ nhất trên thế giới - xác định bazơ, chuyển trực tiếp từ các thông tin hóa học thành các thông tin dạng số. Bộ giải trình tự gen người trên máy ion (Ion Personal Genome Machine - PGM™) sau đó nạp đầy vào chip lần lượt các nucleotit. Nếu nucleotit tiếp theo mà nạp vào chip không phải là nucleotit khớp nhau, thì sẽ không ghi được sự thay đổi điện áp và sẽ không xác định bazơ. Nếu có hai bazơ trùng nhau trên sợi ADN, thì điện áp sẽ gấp đôi, và chip sẽ ghi hai bazơ trùng nhau được xác định. Sự phát hiện trực tiếp cho phép ghi nhận sự kết hợp nucleotit tính bằng giây.

Theo một phương án khác, phương pháp theo sáng chế bao gồm bước thu nhận thông tin trình tự cho các axit nucleic trong mẫu thử nghiệm, ví dụ, cfADN trong mẫu thử nghiệm lấy từ người mẹ, sử dụng giải trình tự bằng quá trình lai. Giải trình tự bằng phép lai bao gồm cho các trình tự polynucleotit tiếp xúc với các mẫu dò polynucleotit, trong đó mỗi trong số các mẫu dò polynucleotit có thể tùy ý được gắn vào nền. Nền có thể là bề mặt phẳng bao gồm mảng gồm nhiều trình tự nucleotit đã biết. Mẫu lai vào mảng có thể được sử dụng để xác định các chuỗi polynucleotit có mặt trong mẫu. Theo các phương án khác, mỗi mẫu dò được gắn vào hạt, ví dụ, hạt từ tính hoặc tương tự. Sự lai vào các hạt có thể được xác định và được sử dụng để nhận dạng các trình tự polynucleotit trong mẫu.

Theo một số phương án trong số các các phương pháp được mô tả ở đây, các thẻ trình tự được ánh xạ bao gồm các trình tự đọc khoảng 20bp, khoảng 25bp, khoảng 30bp, khoảng 35bp, khoảng 40bp, khoảng 45bp, khoảng 50bp, khoảng 55bp, khoảng 60bp, khoảng 65bp, khoảng

70bp, khoảng 75bp, khoảng 80bp, khoảng 85bp, khoảng 90bp, khoảng 95bp, khoảng 100bp, khoảng 110bp, khoảng 120bp, khoảng 130, vào khoảng 140bp, khoảng 150bp, khoảng 200bp, khoảng 250bp, khoảng 300bp, khoảng 350bp, khoảng 400bp, khoảng 450bp, hoặc khoảng 500bp. Mong đợi rằng các tiến bộ công nghệ sẽ cho phép các trình tự đọc một đầu lớn hơn 500bp cho phép các trình tự đọc lớn hơn khoảng 1000bp khi các trình tự đọc có đầu bắt cặp được tạo ra. Theo một phương án, các thẻ trình tự được ánh xạ bao gồm các trình tự đọc mà là 36bp. Ánh xạ các thẻ trình tự được đạt được bằng cách so sánh trình tự của thẻ với trình tự của đối chứng để xác định gốc nhiễm sắc thể của phân tử axit nucleic được giải trình tự (ví dụ cfADN), và thông tin trình tự di truyền riêng biệt là không cần thiết. Mức độ thấp về sự không khớp (0-2 sự không khớp trên mỗi thẻ trình tự) có thể được cho phép để giải thích về các hiện tượng đa hình nhỏ mà có thể tồn tại giữa bộ gen đối chứng và các bộ gen trong mẫu hỗn hợp.

Thường thu được nhiều thẻ trình tự trên mỗi mẫu. Theo một số phương án, ít nhất vào khoảng 3×10^6 thẻ trình tự, ít nhất vào khoảng 5×10^6 thẻ trình tự, ít nhất vào khoảng 8×10^6 thẻ trình tự, ít nhất vào khoảng 10×10^6 thẻ trình tự, ít nhất vào khoảng 15×10^6 thẻ trình tự, ít nhất vào khoảng 20×10^6 thẻ trình tự, ít nhất vào khoảng 30×10^6 thẻ trình tự, ít nhất vào khoảng 40×10^6 thẻ trình tự, hoặc ít nhất vào khoảng 50×10^6 thẻ trình tự chứa từ 20 đến 40bp trình tự đọc, ví dụ, 36bp, được thu được từ việc ánh xạ các trình tự đọc đến bộ gen đối chứng trên mỗi mẫu. Theo một phương án, tất cả các trình tự đọc được ánh xạ đến tất cả các vùng của bộ gen đối chứng. Theo một phương án, các thẻ mà đã được ánh xạ đến tất cả các vùng, ví dụ, tất cả các nhiễm sắc thể, của bộ gen đối chứng đều được đếm, và CNV, tức là, sự biểu diễn quá mức hoặc dưới mức của trình tự được quan tâm, ví dụ, nhiễm sắc thể hoặc một phần của nó, trong mẫu ADN hỗn hợp được xác định. Phương pháp này không yêu cầu sự phân biệt giữa hai bộ gen.

Độ chính xác cần thiết để xác định đúng xem có hay không có CNV, ví dụ, hiện tượng lệch bội, trong mẫu, được dựa vào sự biến đổi về số lượng thẻ trình tự mà ánh xạ đến bộ gen đối chứng trong số các mẫu trong quá trình giải trình tự (khả năng biến đổi giữa các nhiễm sắc thể), và sự biến đổi về số lượng các thẻ trình tự mà ánh xạ đến bộ gen đối chứng trong các quá trình giải trình tự khác nhau (khả năng biến đổi giữa quá trình giải trình tự). Ví dụ, các

biến đổi này có thể được thông báo cụ thể cho các thể mà ảnh xạ đến các trình tự đối chứng giàu GC hoặc nghèo GC. Các biến đổi khác có thể thu được từ việc sử dụng các giao thức khác nhau để tách chiết và làm sạch các axit nucleic, tạo thư viện giải trình tự, và sử dụng các nền tảng giải trình tự khác nhau. Phương pháp theo sáng chế sử dụng các định lượng trình tự (các định lượng nhiễm sắc thể, hoặc các liệu lượng đoạn) dựa trên sự hiểu biết về các trình tự chuẩn hóa (các trình tự nhiễm sắc thể chuẩn hóa hoặc các trình tự đoạn chuẩn hóa), để giải thích về bản chất về khả năng biến đổi tích lũy bắt nguồn từ khả năng biến đổi giữa các nhiễm sắc thể (trong một lần chạy), và giữa các lần giải trình tự (giữa các lần chạy) và phụ thuộc vào nền tảng. Các định lượng nhiễm sắc thể là dựa trên sự hiểu biết về trình tự nhiễm sắc thể chuẩn hóa, mà có thể bao gồm một nhiễm sắc thể, hoặc hai hoặc nhiều nhiễm sắc thể được chọn lọc từ các nhiễm sắc thể 1- 22, X, và Y. Theo cách khác, các trình tự nhiễm sắc thể chuẩn hóa có thể bao gồm một đoạn nhiễm sắc thể, hoặc hai hoặc nhiều đoạn trong số một nhiễm sắc thể hoặc hai hoặc nhiều nhiễm sắc thể. Các liệu lượng đoạn là dựa trên sự hiểu biết về trình tự đoạn chuẩn hóa, mà có thể bao gồm một đoạn của một nhiễm sắc thể bất kỳ, hoặc hai hoặc nhiều đoạn trong số hai hoặc nhiều nhiễm sắc thể trong số các nhiễm sắc thể 1- 22, X, và Y.

CNV và các chẩn đoán trước khi sinh

ADN và ARN bào thai tự do ngoài tế bào lưu thông trong máu người mẹ có thể được sử dụng cho việc chẩn đoán sớm trước sinh không xâm lấn (non-invasive prenatal diagnosis - NIPD) về số lượng tăng của các tình trạng di truyền, cả cho việc quản lý thai kỳ và để hỗ trợ việc đưa ra quyết định sinh sản. Sự có mặt của ADN tự do ngoài tế bào lưu thông trong dòng máu đã được biết đến trên 50 năm. Mới gần đây, phát hiện thấy có các lượng nhỏ ADN bào thai lưu thông trong dòng máu người mẹ trong suốt quá trình thai kỳ (Lo et al., Lancet 350:485-487). Cho rằng bắt nguồn từ các tế bào nhau thai đang chết, ADN bào thai tự do ngoài tế bào (ctADN) đã được chứng minh bao gồm các đoạn ngắn thường có độ dài ít hơn 200 bp (Chan et al., Clin Chern 50:88-92), mà có thể được nhận biết sớm khi thai kỳ được 4 tuần (Illanes et al., Early Human Dev 83:563-566), và được biết là được tách ra khỏi sự lưu hành máu người mẹ trong nhiều giờ sinh đẻ (Lo et al., Am J Hum Genet 64:218-224). Ngoài ctADN, các đoạn của ARN bào thai tự do ngoài tế bào (cell-free fetal RNA - cfRNA) cũng có thể được nhận biết trong dòng máu người mẹ, bắt nguồn từ các gen mà được sao chép ở bào thai hoặc nhau

thai. Sự tách chiết và phân tích tiếp theo về các thành phần di truyền trong bào thai từ mẫu máu người mẹ tạo ra nhiều cơ hội mới cho NIPD.

Phương pháp theo sáng chế là phương pháp độc lập với hiện tượng đa hình, phương pháp này là để dùng trong NIPD và không đòi hỏi phân biệt ctADN bào thai với ctADN trong người mẹ để cho phép xác định hiện tượng lệch bội ở bào thai. Theo một số phương án, hiện tượng lệch bội là thể ba nhiễm hoặc thể đơn nhiễm sắc thể toàn phần, hoặc thể ba nhiễm hoặc thể đơn nhiễm một phần. Các hiện tượng lệch bội một phần được gây ra bởi sự mất hoặc tăng một phần của nhiễm sắc thể, và chứa đựng các mất cân bằng nhiễm sắc thể do có các chuyển đoạn không cân bằng, chuyển đoạn không cân bằng, mất đoạn và chèn đoạn. Hơn nữa, hiện tượng lệch bội được biết phổ biến nhất tương thích với tuổi thọ là thể ba nhiễm 21, tức là, Hội chứng Down (Down Syndrome - DS), mà gây ra bởi sự có mặt của một phần hoặc toàn bộ nhiễm sắc thể 21. Hiếm khi, DS có thể bị gây ra bởi dị tật di truyền hoặc ngẫu nhiên nhờ đó bản sao bổ sung của tất cả hoặc một phần của nhiễm sắc thể 21 trở nên được gắn vào một nhiễm sắc thể khác (thông thường nhiễm sắc thể 14) để tạo ra một nhiễm sắc thể khác thường. DS có liên quan đến sự suy giảm trí tuệ, các khó khăn nghiêm trọng trong học tập và tỷ lệ tử vong quá cao bị gây ra bởi các vấn đề sức khỏe kéo dài như bệnh tim. Các hiện tượng lệch bội khác với ý nghĩa lâm sàng đã biết bao gồm hội chứng Edward (thể ba nhiễm 18) và hội chứng Patau (thể ba nhiễm 13), mà thường tử vong trong vòng một vài tháng đầu đời. Các hiện tượng bất thường liên quan tới số lượng nhiễm sắc thể giới tính cũng được biết và bao gồm thể đơn nhiễm X, ví dụ, hội chứng Turner (XO), và hội chứng ba X (XXX) ở các trường hợp sinh nữ giới và hội chứng Klinefelter (XXY) và hội chứng XYY ở các trường hợp sinh nam giới, tất cả các hội chứng này đều có liên quan tới các kiểu hình khác nhau bao gồm mất khả năng sinh đẻ và giảm sút về kỹ năng trí tuệ. Thể đơn nhiễm X [45, X] là nguyên nhân phổ biến về sảy thai sớm chiếm khoảng 7% trong số các trường hợp sảy thai tự phát. Dựa vào tần suất trẻ sinh ra còn sống 45,X (còn được gọi là hội chứng Turner) 1-2/10,000, ước tính ít hơn 1% trong số các ca thụ tinh 45,X sẽ sống sót được đến lúc sinh. Khoảng 30% trong số các bệnh nhân mắc hội chứng Turners là khảm với cả dòng tế bào 45,X và dòng tế bào 46,XX hoặc một dòng tế bào chứa nhiễm sắc thể X được sắp xếp lại (Hook and Warburton 1983). Kiểu hình ở trẻ sơ sinh sinh ra còn sống là tương đối nhẹ khi xem xét khả năng gây chết phôi cao và có thể tất cả các bé gái sinh ra còn sống mắc hội chứng Turner được giả thiết là mang

dòng tế bào chứa hai nhiễm sắc thể giới tính. Thể đơn nhiễm X có thể xảy ra ở các bé gái dưới dạng 45,X hoặc dưới dạng 45,X/46XX, và ở các bé trai dưới dạng 45,X/46XY. Các thể đơn nhiễm sắc thể thường ở người thường được giả định là không tương thích với sự sống; tuy nhiên, có khá nhiều báo cáo di truyền học tế bào mô tả thể đơn nhiễm đầy đủ của một nhiễm sắc thể 21 ở các trẻ sinh ra còn sống (Vosranova et al., *Molecular Cytogen.* 1:13; Joosten et al., *Prenatal Diagn.* 17:271-5. Phương pháp mô tả ở đây có thể được sử dụng để chẩn đoán các hiện tượng bất thường nhiễm sắc thể này và hiện tượng bất thường nhiễm sắc thể khác trước khi sinh.

Theo một số phương án, các phương pháp bậc lộ ở đây có thể xác định sự có mặt hoặc vắng mặt của các thể ba nhiễm của nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể 1-22, X và Y. Các ví dụ về các thể ba nhiễm mà có thể được phát hiện thấy theo phương pháp theo sáng chế bao gồm nhưng không giới hạn ở thể ba nhiễm 21 (T21; Hội chứng Down), thể ba nhiễm 18 (T18; Hội chứng Edward), thể ba nhiễm 16 (T16), thể ba nhiễm 20 (T20), thể ba nhiễm 22 (T22; Hội chứng Mất Mèo), thể ba nhiễm 15 (T15; hội chứng Prader Willi), thể ba nhiễm 13 (T13; hội chứng Patau), thể ba nhiễm 8 (T8; hội chứng Warkany), thể ba nhiễm 9, và XXY (Hội chứng Klinefelter), XYY, hoặc các thể ba XXX. Các thể ba nhiễm toàn phần của các thể nhiễm sắc điển hình khác tồn tại ở trạng thái không khảm là gây chết người, nhưng có thể tương thích với sự sống khi có mặt ở trạng thái khảm. Cần hiểu rằng các thể ba toàn phần khác nhau, cho dù tồn tại ở trạng thái khảm hoặc trạng thái không khảm, và các thể ba một phần có thể được xác định trong cfADN dạng phôi theo các bậc lộ được nêu ra ở đây.

Các ví dụ không giới hạn về các thể ba nhiễm một phần mà có thể được xác định bởi phương pháp theo sáng chế bao gồm nhưng không bị giới hạn ở, thể ba nhiễm một phần 1q32-44, thể ba nhiễm 9p, khảm nhiễm thể ba nhiễm 4, thể ba nhiễm 17p, thể ba nhiễm một phần 4q26-pter, thể ba nhiễm một phần 2p, thể ba nhiễm một phần 1q, và/hoặc thể ba nhiễm một phần 6p/thể đơn nhiễm 6q.

Các phương pháp bậc lộ ở đây cũng có thể được sử dụng để xác định thể đơn nhiễm sắc thể X, thể đơn nhiễm sắc thể 21, và các thể đơn nhiễm một phần như, thể đơn nhiễm 13, thể đơn nhiễm 15, thể đơn nhiễm 16, thể đơn nhiễm 21, và thể đơn nhiễm 22, mà được biết là có liên quan đến sảy thai trong thai kỳ. Thể đơn nhiễm một phần của các nhiễm sắc thể thường liên

quan đến hiện tượng lệch bội toàn phần cũng có thể được xác định bằng phương pháp được mô tả ở đây. Các ví dụ không giới hạn về các hội chứng mất đoạn mà có thể được xác định theo phương pháp theo sáng chế bao gồm các hội chứng do các mất đoạn một phần của các nhiễm sắc thể gây ra. Các ví dụ về các mất đoạn một phần mà có thể được xác định theo các phương pháp được mô tả ở đây bao gồm nhưng không giới hạn ở các mất đoạn một phần của các nhiễm sắc thể 1, 4, 5, 7, 11, 18, 15, 13, 17, 22 và 10, mà được mô tả dưới đây.

Hội chứng mất đoạn 1q21.1 hoặc vi mất đoạn 1q21.1 (hội quy) là sự sai hình hiếm thấy của nhiễm sắc thể 1. Tiếp theo hội chứng mất đoạn, còn có hội chứng lặp đoạn 1q21.1. Mặc dù có một phần của ADN mất theo hội chứng mất đoạn trên một điểm cụ thể, nhưng có hai hoặc ba bản sao của phần tương tự của ADN trên cùng điểm này có hội chứng lặp đoạn. Tài liệu chuyên ngành đề cập đến cả mất đoạn và lặp đoạn dưới dạng các biến thể số lượng bản sao 1q21.1 (copy-number variation - CNV). Mất đoạn 1q21.1 có thể được kết hợp với hội chứng TAR (giảm tiểu cầu với thiếu xương quay).

Hội chứng Wolf-Hirschhorn (WHS) (OMIN #194190) là hội chứng mất đoạn gen liên kết kết hợp với mất đoạn bán hợp tử của nhiễm sắc thể 4p16.3. Hội chứng Wolf-Hirschhorn là hội chứng dị dạng bẩm sinh được đặc trưng bởi thiếu năng phát triển trước và sau khi sinh, khuyết tật phát triển có mức độ thay đổi, các dấu hiệu sọ mặt đặc trưng (về ngoài kiểu “mặt nạ chiến binh Hy Lạp” như mũi, trán cao, gốc mũi nhô cao, hiện tượng cách xa nhau của hai bộ phận, lông mày cong cao, mắt lồi, các nếp rãnh quạt, nhân trung ngắn, miệng khác biệt với các khóm miệng hướng xuống dưới, và tật hàm nhỏ), và rối loạn co giật.

Mất đoạn một phần của nhiễm sắc thể 5, còn được biết đến là trù 5p- hoặc 5p, và được gọi là hội chứng Cris du Chat (OMIN#123450), bị gây ra bởi sự mất đoạn của nhánh ngắn (nhánh p) của nhiễm sắc thể 5 (5p15.3-p15.2). Những trẻ sơ sinh mắc tình trạng bệnh này thường xuyên khóc có âm vực cao mà nghe giống như tiếng kêu của mèo. Rối loạn này được đặc trưng bởi khuyết tật trí tuệ và phát triển chậm, kích thước đầu nhỏ (hiện tượng nhỏ đầu), trọng lượng khi sinh nhẹ, và giọng nói cường độ yếu (giảm trương lực) ở giai đoạn trẻ sơ sinh, các đặc điểm khuôn mặt khác biệt và có thể có các dị tật tim.

Hội chứng Williams-Beuren còn được biết đến là hội chứng mất đoạn nhiễm sắc thể 7q11.23 (OMIN 194050) là hội chứng mất đoạn gen liên kết dẫn đến rối loạn đa hệ gây ra bởi mất đoạn bán hợp tử nằm trong khoảng từ 1,5 đến 1,8 JVIb trên nhiễm sắc thể 7q11.23, mà chứa xấp xỉ 28 gen.

Hội chứng Jacobsen, còn được biết đến là rối loạn mất đoạn 11q, là rối loạn bẩm sinh hiếm gặp do mất đoạn của vùng đầu cuối của nhiễm sắc thể 11 chứa dải 11q24.1. Nó có thể gây ra các khuyết tật trí tuệ, hình dạng bề ngoài khuôn mặt khác biệt, và các vấn đề thể chất bao gồm cả các dị tật tim và rối loạn đông máu.

Thẻ đơn nhiễm một phần của nhiễm sắc thể 18, được biết đến là thẻ đơn nhiễm 18p là rối loạn nhiễm sắc thể hiếm gặp trong đó tất cả hoặc một phần của nhánh ngắn (p) của nhiễm sắc thể 18 bị mất đoạn (thẻ đơn nhiễm). Rối loạn này thông thường được đặc trưng bởi vóc người nhỏ bé, các mức độ biến thiên của tâm lý chậm phát triển, sự chậm nói, các dị dạng của vùng sọ và mặt (sọ mặt), và/hoặc các bất thường thể chất khác. Các dị tật sọ mặt liên quan có thể thay đổi mạnh về phạm vi và mức độ nghiêm trọng qua từng trường hợp.

Các tình trạng bệnh gây ra bởi các thay đổi về cấu trúc hoặc số lượng bản sao của nhiễm sắc thể 15 bao gồm Hội chứng Angelman và Hội chứng Prader-Willi, các hội chứng này liên quan tới mất hoạt tính gen ở cùng phần của nhiễm sắc thể 15, vùng 15q11-q13. Cần hiểu rằng một số hoán vị và vi mất đoạn có thể là không triệu chứng ở bố mẹ người mang bệnh, nhưng lại có thể gây ra bệnh di truyền chính ở con đẻ. Ví dụ, người mẹ khỏe mạnh mang vi mất đoạn 15q11-q13 có thể sinh ra đứa trẻ mắc hội chứng Angelman, rối loạn suy nhược thần kinh nghiêm trọng. Vì vậy, các phương pháp, thiết bị và hệ thống mô tả ở đây có thể được sử dụng để nhận dạng mất đoạn một phần như vậy và các mất đoạn khác ở bào thai.

Thẻ đơn nhiễm một phần 13q là rối loạn nhiễm sắc thể hiếm gặp mà xảy ra khi mất một phần của nhánh dài (q) của nhiễm sắc thể 13 (thẻ đơn nhiễm). Các trẻ sơ sinh sinh ra với thẻ đơn nhiễm một phần 13q có thể có trọng lượng khi sinh nhẹ, các dị dạng của đầu và mặt (vùng sọ mặt), các bất thường về bộ xương (đặc biệt là tay và chân), và các bất thường thể chất khác. Tâm lý chậm phát triển là đặc trưng cho tình trạng bệnh này. Tỷ lệ tử vong trong suốt giai đoạn trẻ sơ sinh là cao trong số những cá nhân được sinh ra mắc rối loạn này. Hầu hết tất cả

trường hợp của thể đơn nhiễm một phần 13q đều xảy ra ngẫu nhiên mà không có lý do rõ ràng (ngẫu nhiên).

Hội chứng Smith-Magenis (SMS - OMIM #182290) bị gây ra bởi mất đoạn, hoặc mất vật liệu di truyền, trên một bản sao của nhiễm sắc thể 17. Hội chứng được biết rộng rãi này có liên quan tới chậm phát triển, tâm lý chậm phát triển, các dị thường bẩm sinh như các dị tật tim và thận, và các bất thường về hành vi thần kinh như các rối loạn giấc ngủ nghiêm trọng và hành vi tự gây tổn thương. Hội chứng Smith-Magenis (SMS) được gây ra ở hầu hết các trường hợp (90%) bởi sự mất đoạn giữa 3. 7-Mb trong nhiễm sắc thể 17p11.2.

Hội chứng mất đoạn 22q11.2, còn được biết đến là hội chứng DiGeorge, là hội chứng gây ra do mất đoạn một phần của nhiễm sắc thể 22. Mất đoạn (22 q11.2) xảy ra ở gần giữa của nhiễm sắc thể trên nhánh dài của một trong số nhiễm sắc thể. Các dấu hiệu của hội chứng này thay đổi nhiều, thậm chí giữa các thành viên của cùng một gia đình, và ảnh hưởng đến nhiều bộ phận của cơ thể. Các dấu hiệu và triệu chứng đặc trưng có thể bao gồm các dị tật khi sinh như bệnh tim bẩm sinh, các dị tật ở vòm miệng, liên quan phổ biến nhất tới các vấn đề thần kinh cơ với sự đóng cơ (thiếu năng vòm miệng), các khuyết tật học tập, các khác biệt nhỏ ở các dấu hiệu trên khuôn mặt, và các nhiễm khuẩn tái diễn. Các vi mất đoạn ở vùng nhiễm sắc thể 22q11.2 có liên quan đến nguy cơ mắc bệnh tâm thần phân liệt tăng gấp 20 đến 30 lần.

Các mất đoạn trên nhánh ngắn của nhiễm sắc thể 10 có liên quan tới hội chứng DiGeorge giống như kiểu hình. Thể đơn nhiễm một phần của nhiễm sắc thể 10p là hiếm gặp nhưng đã được quan sát thấy ở một phần trong số các người bệnh biểu hiện các dấu hiệu của hội chứng DiGeorge.

Theo một phương án, các phương pháp, thiết bị, và hệ thống mô tả ở đây được sử dụng để xác định các thể đơn nhiễm một phần bao gồm nhưng không bị giới hạn ở thể đơn nhiễm một phần của các nhiễm sắc thể 1, 4, 5, 7, 11, 18, 15, 13, 17, 22 và 10, ví dụ, thể đơn nhiễm một phần 1q21.11, thể đơn nhiễm một phần 4p16.3, thể đơn nhiễm một phần 5p15.3-p15.2, thể đơn nhiễm một phần 7q11.23, thể đơn nhiễm một phần 11q24.1, thể đơn nhiễm một phần 18p, thể đơn nhiễm một phần của nhiễm sắc thể 15 (15q11-q13), thể đơn nhiễm một phần 13q, thể

đơn nhiễm một phần 17p11.2, thể đơn nhiễm một phần của nhiễm sắc thể 22 (22q11.2), và thể đơn nhiễm một phần 10p cũng có thể được xác định nhờ sử dụng phương pháp này.

Các thể đơn nhiễm một phần khác mà có thể được xác định theo các phương pháp được mô tả ở đây bao gồm chuyển đoạn không cân bằng t(8;11)(p23.2;p15.5); vi mất đoạn 11q23; mất đoạn 17p11.2; mất đoạn 22q13.3; vi mất đoạn Xp22.3; mất đoạn 10p14; vi mất đoạn 20p, [del(22)(q11.2q11.23)], các mất đoạn 7q11.23 và 7q36; mất đoạn 1p36; vi mất đoạn 2p; u sợi thần kinh loại 1 (vi mất đoạn 17q 11.2), mất đoạn Yq; vi mất đoạn 4p16,3; vi mất đoạn 1p36.2; mất đoạn 11q14; vi mất đoạn 19q13.2; Rubinstein-Taybi (vi mất đoạn 16 p13.3); vi mất đoạn 7p21; Hội chứng Miller-Dieker (17p13.3); và vi mất đoạn 2q37. Các mất đoạn một phần có thể là các mất đoạn nhỏ của phần của nhiễm sắc thể, hoặc chúng có thể là các vi mất đoạn của nhiễm sắc thể ở đó mất đoạn của một gen có thể xảy ra.

Một số hội chứng lặp đoạn bị gây ra do lặp đoạn một phần của các nhánh nhiễm sắc thể đã được nhận dạng (xem OMIN [Di truyền thuyết Mendel trực tuyến ở Người được xem trực tuyến tại ncbi.nlm.nih.gov/omim]). Theo một phương án, phương pháp theo sáng chế có thể được sử dụng để xác định có hoặc không có các lặp đoạn và/hoặc các phép nhân của các đoạn của nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể 1-22, X và Y. Các ví dụ không giới hạn về các hội chứng lặp đoạn mà có thể được xác định theo phương pháp theo sáng chế bao gồm các lặp đoạn một phần của các nhiễm sắc thể 8, 15, 12, và 17, mà được mô tả dưới đây.

Hội chứng lặp đoạn 8p23.1 là rối loạn di truyền hiếm gặp gây ra do lặp đoạn của một vùng từ nhiễm sắc thể người 8. Hội chứng lặp đoạn này có mức độ thường xảy ra ước tính ở 1 trong 64.000 ca sinh và là nghịch đảo của hội chứng mất đoạn 8p23.1. Lặp đoạn 8p23.1 có liên quan tới kiểu hình biến đổi bao gồm một hoặc nhiều trong các chậm nói, chậm phát triển, dị hình biến dạng nhẹ, vớ trán cao và lông mày uốn vòng cung, và bệnh tim bẩm sinh (CHD).

Hội chứng lặp đoạn nhiễm sắc thể 15q (Dup15q) là hội chứng có thể nhận dạng về lâm sàng mà tạo thành từ các lặp đoạn của nhiễm sắc thể 15q11-13.1. Các trẻ mắc Dup15q thường bị giảm trương lực (giọng nói cường độ yếu), chậm phát triển; chúng có thể được sinh ra với dị tật sứt môi và/hoặc khe hở hàm ếch hoặc các dị dạng của tim, thận hoặc các cơ quan khác;

chúng thể hiện một số mức độ về chậm/khuyết tật nhận thức (tâm lý chậm phát triển), chậm nói và chậm ngôn ngữ, và các rối loạn xử lý giác quan.

Hội chứng Pallister Killian là kết quả của vật liệu nhiễm sắc thể bổ sung #12. Thường có sự pha trộn của các tế bào (hiện tượng khảm nhiễm thể), một số có vật liệu bổ sung #12, và một số mà bình thường (46 nhiễm sắc thể mà không có vật liệu bổ sung #12). Các trẻ mắc hội chứng này có nhiều vấn đề bao gồm tâm lý chậm phát triển nghiêm trọng, giọng nói cường độ yếu, các dấu hiệu "dữ tợn" trên khuôn mặt, và trán cao. Chúng xu hướng có môi trên rất mỏng với môi dưới dày hơn và mũi ngắn. Các vấn đề sức khỏe khác bao gồm co giật, ăn uống kém, các khớp nối cứng, các bệnh đục thể mắt ở tuổi trưởng thành, mất thính giác, và các di tật tim. Những người mắc hội chứng Pallister Killian có tuổi thọ ngắn.

Các cá nhân mắc tình trạng bệnh di truyền được gọi là dup(17)(p11.2p11.2) hoặc dup 17p mang thông tin di truyền bổ sung (được biết đến là lặp đoạn) trên nhánh ngắn của nhiễm sắc thể 17. Lặp đoạn của nhiễm sắc thể 17p11.2 nằm trong Hội chứng Potocki-Lupski (PTLS), là tình trạng bệnh di truyền mới được thừa nhận với chỉ một vài chục trường hợp được nêu trong tài liệu chuyên ngành y khoa. Các người bệnh mà có lặp đoạn này thường có giọng nói cường độ thấp, ăn uống kém, và không phát triển trong suốt giai đoạn trẻ sơ sinh, và cũng thể hiện với sự phát triển trễ các dấu mốc quan trọng như vận động và biết nói. Nhiều cá nhân mà mắc PTLS gặp khó khăn với sự phát âm rõ ràng và xử lý ngôn ngữ. Ngoài ra, các người bệnh có thể có các đặc điểm về hành vi tương tự với các đặc điểm nhìn thấy ở những người mắc tự kỷ hoặc rối loạn tự kỷ. Các cá nhân mắc hội chứng PTLS có thể có các di tật tim và ngưng thở khi ngủ. Đã biết lặp đoạn của vùng lớn trong nhiễm sắc thể 17p12 mà bao gồm gen PMP22 gây ra bệnh sâu răng Charcot-Marie.

CNV có liên quan tới các trường hợp chết non. Tuy nhiên, do các giới hạn vốn có của di truyền tế bào thông thường, nên sự đóng góp của CNV đối với sự chết non được cho là bị miêu tả không đúng mức (Harris et al., Prenatal Diagn 31:932-944). Như được thể hiện trong các ví dụ và mô tả chỗ khác trong bản mô tả này, phương pháp theo sáng chế có khả năng xác định sự có mặt của các hiện tượng lệch bội một phần, ví dụ, các mất đoạn và lặp đoạn của các đoạn nhiễm sắc thể, và có thể được sử dụng để nhận dạng và xác định có hoặc không có CNV mà có liên quan tới các trường hợp chết non.

Xác định CNV của các rối loạn lâm sàng

Ngoài việc xác định sớm các dị tật khi sinh, các phương pháp được mô tả ở đây có thể được áp dụng cho việc xác định bất thường trong biểu diễn các trình tự di truyền trong bộ gen. Các bất thường trong biểu diễn các trình tự di truyền trong bộ gen có liên quan tới nhiều bệnh lý khác nhau. Các bệnh lý như vậy bao gồm nhưng không bị giới hạn ở ung thư, các bệnh nhiễm khuẩn và tự miễn dịch, các bệnh của hệ thần kinh, các bệnh chuyển hóa và/hoặc tim mạch, và tương tự.

Do vậy, theo các phương án khác nhau, sử dụng các phương pháp được mô tả ở đây trong việc chẩn đoán, và/hoặc theo dõi, và hoặc điều trị các bệnh lý như vậy được đề xuất. Ví dụ, các phương pháp có thể được áp dụng vào việc xác định có hoặc không có bệnh, vào việc theo sự tiến triển của bệnh và/hoặc hiệu quả của chế độ điều trị, vào việc xác định có hoặc không có các axit nucleic của mầm bệnh ví dụ virus; vào việc xác định các bất thường nhiễm sắc thể liên quan tới bệnh đoạn ghép chống lại vật chủ (graft versus host disease - GVHD), và vào việc xác định sự đóng góp của các cá nhân trong các phép phân tích pháp y.

Các CNV trong Ung thư

ADN huyết tương và huyết thanh của các người bệnh ung thư chứa số lượng đo được của ADN khối u, mà có thể được phục hồi và được dùng làm nguồn thay thế của ADN khối u, và các khối u được đặc trưng bởi hiện tượng lệch bội, hoặc các số lượng không phù hợp của các trình tự gen hoặc thậm chí toàn bộ các nhiễm sắc thể. Vì vậy, xác định sự khác biệt về lượng trình tự cho trước tức là trình tự được quan tâm, trong mẫu từ cá thể có thể được áp dụng trong việc dự báo hoặc chẩn đoán tình trạng bệnh y khoa. Theo một số phương án, phương pháp theo sáng chế có thể được sử dụng để xác định có hoặc không có hiện tượng lệch bội nhiễm sắc thể ở người bệnh bị nghi ngờ hoặc được biết là bị ung thư.

Một số phương án thực hiện ở đây đề xuất các phương pháp phát hiện ung thư, theo dõi đáp ứng điều trị và bệnh tồn dư tối thiểu dựa trên các mẫu ctADN lưu thông trong máu bằng cách sử dụng giải trình tự nông đối với các mẫu bằng phương pháp đầu bắt cặp và sử dụng các thông tin kích thước đoạn có sẵn từ các trình tự đọc đầu bắt cặp để xác định sự có mặt của ADN tế bào chết theo chương trình được metyl hóa khác biệt từ các tế bào ung thư trong môi

trường của các tế bào bình thường. cfADN có nguồn gốc khối u đã được chứng minh là ngắn hơn cfADN không có nguồn gốc khối u ở một số bệnh ung thư. Do đó, phương pháp dựa trên kích thước mô tả ở đây có thể được sử dụng để xác định CNV bao gồm các hiện tượng lệch bội có liên quan tới các bệnh ung thư này, cho phép (a) phát hiện khối u có mặt trong quá trình thiết lập sàng lọc hoặc chẩn đoán; (b) theo dõi đáp ứng điều trị; (c) theo dõi bệnh tồn dư tối thiểu.

Theo một số phương án nhất định, hiện tượng lệch bội là đặc trưng về bộ gen của đối tượng và dẫn đến sự gia tăng nói chung về khuynh hướng dễ mắc ung thư. Theo một số phương án nhất định, hiện tượng lệch bội là đặc trưng về các tế bào cụ thể (ví dụ, các tế bào khối u, các tế bào tiền khối u, v.v.) mà là hoặc có khuynh hướng gia tăng mắc khối đa u. Các hiện tượng lệch bội cụ thể có liên quan tới các bệnh ung thư cụ thể hoặc các khuynh hướng dễ mắc các bệnh ung thư cụ thể như mô tả dưới đây. Theo một số phương án, phương pháp giải trình tự đầu bắt cặp rất nông có thể được sử dụng để phát hiện / theo dõi sự có mặt của bệnh ung thư theo cách hiệu quả chi phí.

Do vậy, các phương pháp theo các phương án khác nhau được mô tả ở đây thực hiện việc xác định biến thể số lượng bản sao của (các) trình tự đang được quan tâm ví dụ (các) trình tự liên quan tới lâm sàng, trong mẫu thử nghiệm từ đối tượng ở đó một số biến đổi nhất định về số lượng bản sao tạo ra chỉ báo về sự có mặt và/hoặc khuynh hướng dễ mắc bệnh ung thư. Theo một số phương án nhất định, mẫu bao gồm hỗn hợp các axit nucleic có nguồn gốc từ hai hoặc nhiều loại tế bào. Theo một phương án, hỗn hợp các axit nucleic có nguồn gốc từ các tế bào bình thường và ung thư có nguồn gốc từ đối tượng mắc phải tình trạng bệnh y khoa ví dụ ung thư.

Sự phát triển của bệnh ung thư thông thường đi kèm với sự thay đổi về số lượng của toàn bộ nhiễm sắc thể tức là hiện tượng lệch bội nhiễm sắc thể toàn phần, và/hoặc thay đổi về số lượng các đoạn của các nhiễm sắc thể tức là hiện tượng lệch bội một phần, gây ra bởi quy trình được biết đến là bất ổn định nhiễm sắc thể (chromosome instability - CIN) (Thoma et al., Swiss Med Weekly 2011;141:w13170). Tin rằng nhiều khối u rắn, như ung thư vú, sự tiến triển từ khởi mào đến di căn thông qua sự tích lũy của một số sai hình di truyền. [Sato et al., Cancer Res., 50: 7184-7189; Jongsma et al., J Clin Pathol: Mol Path 55:305-309]]. Các sai hình di

truyền như vậy, khi chúng tích lũy, có thể mang lại những lợi thế tăng sinh, bất ổn định di truyền và khả năng đi kèm là phát triển sự kháng thuốc nhanh chóng, và tăng cường sự tạo mạch, sự phân hủy protein và di căn. Các sai hình di truyền có thể ảnh hưởng đến hoặc "gen ức chế ung thư" lặn hoặc các gen gây ung thư hoạt động mạnh. Tin rằng các mất đoạn và tái tổ hợp dẫn đến sự mất tính dị hợp tử (loss of heterozygosity - LOH) đóng vai trò chính trong sự phát triển khối u bằng cách mở ra các gen tương ứng ức chế khối u đột biến.

cfADN đã được phát hiện thấy trong quá trình di chuyển các người bệnh được chẩn đoán mắc ung thư ác tính bao gồm nhưng không bị giới hạn ở ung thư phổi (Pathak et al. Clin Chern 52:1833-1842), ung thư tuyến tiền liệt (Schwartzbach et al. Clin Cancer Res 15:1032-8), và ung thư vú (Schwartzbach et al. available online at breast-cancer-research.com/content/1115/R71). Nhận dạng các bất ổn định bộ gen liên quan tới các ung thư mà có thể được xác định trong cfADN di chuyển trong máu ở các người bệnh ung thư là công cụ chẩn đoán và dự đoán tiềm tàng. Theo một phương án, các phương pháp mô tả ở đây được sử dụng để xác định CNV của một hoặc nhiều (các) trình tự đang được quan tâm trong mẫu, ví dụ, mẫu chứa hỗn hợp các axit nucleic có nguồn gốc từ đối tượng mà bị nghi ngờ hoặc được biết là mắc ung thư ví dụ ung thư biểu mô, ung thư mô liên kết, u lymphô, bệnh bạch cầu, các u tế bào mầm và u nguyên bào. Theo một phương án, mẫu này là mẫu huyết tương có nguồn gốc (được xử lý) từ máu ngoại vi mà có thể chứa hỗn hợp cfADN có nguồn gốc từ các tế bào bình thường và ung thư. Theo một phương án khác, mẫu sinh học cần thiết để xác định xem có CNV hay không là có nguồn gốc từ tế bào mà, nếu có ung thư, chứa hỗn hợp các tế bào ung thư và không ung thư từ các mô sinh học khác bao gồm nhưng không bị giới hạn ở các chất lỏng sinh học như huyết thanh, mô hôi, nước mắt, nước bọt, nước tiểu, nước bọt, dịch tai, bạch huyết, nước bọt, dịch não tủy, các dịch tả, dịch huyền phù tủy xương, dịch âm đạo, rửa bàng quang, dịch não, dịch tích tụ ở bụng, sữa, các chất bài tiết qua các đường hô hấp, ruột và tiết niệu, và các mẫu gan tách bạch cầu, hoặc ở các sinh tiết mô, các miếng gạc, hoặc các kính phết. Theo các phương án khác, mẫu sinh học là mẫu chất thải rắn từ người (phân).

Các phương pháp được mô tả ở đây không bị giới hạn ở phép phân tích cfADN.

Cần hiểu rằng các phép phân tích tương tự có thể được thực hiện trên các mẫu tế bào ADN.

Theo các phương án khác nhau, (các) trình tự đang được quan tâm bao gồm (các) trình tự axit nucleic được biết hoặc bị nghi ngờ đóng một vai trò trong sự phát triển và/hoặc tiến triển của bệnh ung thư. Các ví dụ về trình tự đang được quan tâm bao gồm các trình tự axit nucleic ví dụ các nhiễm sắc thể toàn phần và/hoặc các đoạn của các nhiễm sắc thể, mà được khuếch đại hoặc mất đoạn ở các tế bào ung thư như mô tả dưới đây.

Tổng số lượng CNV và nguy cơ mắc bệnh ung thư

Mỗi SNP ung thư phổ biến và CNV ung thư phổ biến tương tự đều có thể chỉ tạo ra sự gia tăng nhẹ về nguy cơ mắc bệnh. Tuy nhiên, nói chung chúng có thể gây ra nguy cơ mắc ung thư tăng đáng kể. Về việc này, lưu ý rằng tăng hoặc mất dòng giao tử của các đoạn ADN lớn đã được báo cáo dưới dạng các nhân tố khiến cho các cá nhân có khuynh hướng dễ mắc u nguyên bào sợi thần kinh, tuyến tiền liệt và ung thư đại trực tràng, ung thư vú, vòm họng buồng trứng liên quan tới BRCA1 (xem, ví dụ, Krepischi et al. Breast Cancer Res., 14: R24; Diskin et al. Nature 2009, 459:987-991; Liu et al. Cancer Res 2009, 69: 2176-2179; Lucito et al. Cancer Biol Ther 2007, 6:1592-1599; Thean et al. Genes Chromosomes Cancer 2010, 49:99-106; Venkatachalam et al. Int J Cancer 2011, 129:1635-1642; and Yoshihara et al. Genes Chromosomes Cancer 2011, 50:167-177). Lưu ý rằng các CNV thường xuyên được tìm thấy trong nhóm quần thể khỏe mạnh (các CNV phổ biến) được cho là đóng một vai trò trong căn bệnh ung thư (xem, ví dụ, Shlien and Malkin (2009) Genome Medicine, 1(6): 62). Trong một nghiên cứu thử nghiệm giả thiết rằng các CNV phổ biến có liên quan tới ung thư ác tính (Shlien et al. Proc Natl Acad Sci USA 2008, 105:11264-11269) một sơ đồ về mọi CNV đã biết có quỹ tích trùng với quỹ tích của các gen liên quan tới ung thư lành tính (như được giới thiệu bởi Higgins et al. Nucleic Acids Res 2007, 35:D721-726) được tạo ra. Chúng được gọi là "các CNV ung thư". Trong phép phân tích ban đầu (Shlien et al. Proc Natl Acad Sci USA 2008, 105:11264-11269), 770 bộ gen khỏe mạnh được đánh giá bằng cách sử dụng tập hợp mảng Affymetrix 500K, có khoảng cách trung bình giữa các mẫu dò là 5,8kb. Vì các CNV thông thường được cho là bị yếu đi ở các vùng gen (Redon et al. (2006) Nature 2006, 444:444-454), nên bất ngờ khi phát hiện ra 49 gen ung thư mà được bao gồm hoặc chồng lấn trực tiếp bởi CNV ở nhiều hơn một người trong nhóm quần thể đối chứng lớn. Trong mười gen đầu tiên, các CNV ung thư có thể được tìm thấy trong bốn người hoặc nhiều hơn.

Vì vậy tin rằng tần suất CNV có thể được dùng làm thước đo nguy cơ mắc bệnh ung thư (xem, ví dụ, công bố sáng chế Mỹ số 2010/0261183 A1). Tần suất CNV có thể được xác định đơn giản bằng bộ gen cấu thành của sinh vật hoặc nó có thể biểu diễn một phần có nguồn gốc từ một hoặc nhiều khối u (các tế bào khối u) nếu có các khối u như vậy.

Theo một số phương án nhất định, các CNV trong mẫu thử nghiệm (ví dụ, mẫu bao gồm axit nucleic cấu tạo (dòng giao tử)) hoặc hỗn hợp các axit nucleic (ví dụ, axit nucleic dòng giao tử và (các) axit nucleic có nguồn gốc từ các tế bào khối u) được xác định nhờ sử dụng các phương pháp được mô tả ở đây cho các biến thể số lượng bản sao. Việc nhận dạng số lượng gia tăng của các CNV trong mẫu thử nghiệm, ví dụ, so với giá trị đối chứng chỉ ra nguy cơ hoặc khuynh hướng dễ mắc ung thư ở đối tượng. Cần hiểu rằng giá trị đối chứng có thể thay đổi theo nhóm quần thể xác định. Cũng cần hiểu rằng giá trị tuyệt đối của sự gia tăng về tần suất CNV sẽ thay đổi tùy theo độ chính xác của phương pháp dùng để xác định tần suất CNV và các tham số khác. Thông thường, sự gia tăng về tần suất CNV ít nhất vào khoảng 1,2 lần giá trị đối chứng được xác định để chỉ báo nguy cơ mắc ung thư (xem, ví dụ, Công bố patent Mỹ 2010/0261183 A1), ví dụ sự gia tăng về tần suất CNV ít nhất hoặc vào khoảng 1,5 lần giá trị đối chứng hoặc nhiều hơn, như 2 đến 4 lần giá trị đối chứng là chỉ báo về nguy cơ gia tăng mắc bệnh ung thư (ví dụ, so với nhóm quần thể đối chứng khỏe mạnh bình thường).

Xác định sự biến đổi cấu trúc trong bộ gen của động vật có vú so với giá trị đối chứng cũng được cho là chỉ báo nguy cơ mắc ung thư. Trong ngữ cảnh này, theo một phương án, thuật ngữ "biến đổi cấu trúc" có thể được định nghĩa là tần suất CNV ở động vật có vú được nhân với kích thước CNV trung bình (tính theo bp) ở động vật có vú. Vì vậy, các số điểm biến đổi cấu trúc cao sẽ thu được do tần suất CNV gia tăng và/hoặc do xảy ra các mất đoạn hoặc lặp đoạn axit nucleic thuộc bộ gen lớn. Theo đó, theo một số phương án nhất định các CNV trong mẫu thử nghiệm (ví dụ, mẫu bao gồm axit nucleic cấu tạo (dòng giao tử)) được xác định bằng cách sử dụng các phương pháp được mô tả ở đây để xác định kích thước và số lượng của các biến thể số lượng bản sao. Theo một số phương án nhất định, tổng điểm số biến đổi cấu trúc trong ADN bộ gen lớn hơn khoảng 1 triệu bazơ, hoặc lớn hơn khoảng 1,1 triệu bazơ, hoặc lớn hơn khoảng 1,2 triệu bazơ, hoặc lớn hơn khoảng 1,3 triệu bazơ, hoặc lớn hơn

khoảng 1,4 triệu bazơ, hoặc lớn hơn khoảng 1,5 triệu bazơ, hoặc lớn hơn khoảng 1,8 triệu bazơ, hoặc lớn hơn khoảng 2 triệu bazơ của ADN chỉ báo nguy cơ mắc bệnh ung thư.

Tin rằng các phương pháp này cung cấp phép đo về nguy cơ mắc ung thư bao gồm nhưng không bị giới hạn ở, các bệnh bạch cầu cấp tính và mãn tính, các u lymphô, rất nhiều khối u rắn của trung mô phôi hoặc biểu mô phôi, ung thư não, vú, gan, dạ dày, ruột kết, u lymphô tế bào B, ung thư phổi, ung thư phế quản, ung thư đại trực tràng, ung thư tuyến tiền liệt, ung thư vú, ung thư tuyến tụy, ung thư dạ dày, ung thư buồng trứng, ung thư bong da, ung thư hệ thần kinh não hoặc trung ương, ung thư hệ thần kinh ngoại vi, ung thư thực quản, ung thư cổ, u melanin, ung thư tử cung hoặc nội mạc tử cung, ung thư khoang miệng hoặc họng, ung thư gan, ung thư thận, ung thư đường mật, ung thư ruột non hoặc ruột thừa, ung thư tuyến nước bọt, ung thư tuyến giáp, ung thư tuyến thượng thận, ung thư xương, ung thư mô liên kết sụn, u mỡ ác tính, ung thư tinh hoàn, và u mô bào sợi ác tính, và các ung thư khác.

Các hiện tượng lệch bội nhiễm sắc thể toàn phần

Như nêu trên đây, có tồn tại tần suất cao của hiện tượng lệch bội ở bệnh ung thư. Trong một số nghiên cứu kiểm tra mức độ thường xảy ra của các biến thể số lượng bản sao soma (somatic copy number alterations - SCNAs) ở bệnh ung thư, phát hiện thấy rằng một phần tư của bộ gen của tế bào ung thư thông thường bị bệnh hoặc bởi các SCNA toàn bộ nhánh hoặc bởi các SCNA toàn bộ nhiễm sắc thể của hiện tượng lệch bội (xem, ví dụ, Beroukhi et al. Nature 463: 899-905). Các biến đổi toàn bộ nhiễm sắc thể được quan sát thấy hồi quy ở một số loại ung thư. Ví dụ, nhìn thấy sự gia tăng của nhiễm sắc thể 8 ở 10 đến 20% trong số các trường hợp mắc bệnh bạch cầu tủy bào cấp tính (acute myeloid leukaemia - AML), cũng như một số khối u rắn, bao gồm ung thư mô liên kết Ewing và các khối u xơ cứng (xem, ví dụ, Barnard et al. Leukemia 10: 5-12; Maurici et al. Cancer Genet. Cytogenet. 100: 106-110; Qi et al. Cancer Genet. Cytogenet. 92: 147-149; Barnard, D. R. et al. Blood 100: 427-434; và tương tự. Danh sách các tăng và giảm nhiễm sắc thể minh họa nhưng không giới hạn sáng chế ở các bệnh ung thư người được thể hiện trong Bảng 2.

BẢNG 2. Các tăng và giảm nhiễm sắc thể hồi quy đặc biệt minh họa ở bệnh ung thư người (xem, ví dụ, Gordon et al. (2012) Nature Rev. Genetics, 13: 189-203).

Nhiễm sắc thể	Tăng Loại ung thư	Giảm Loại ung thư
1	Đa u tủy xương Ung thư biểu mô tuyến (vú)	Ung thư biểu mô tuyến (thận)
2	U nguyên bào gan Ung thư mô liên kết của Ewing	
3	Đa u tủy xương U lymphô tế bào B lớn khuếch tán	U melanin Ung thư biểu mô tuyến (thận)
4	Bệnh bạch cầu nguyên bào cấp tính	Ung thư biểu mô tuyến (thận)
5	Đa u tủy xương Ung thư biểu mô tuyến (thận)	
6	Bệnh bạch cầu nguyên bào cấp tính Khối u Wilm	Ung thư biểu mô tuyến (thận)
7	Ung thư biểu mô tuyến (thận) Ung thư biểu mô tuyến (ruột)	Bệnh bạch cầu tủy bào cấp tính Bệnh bạch cầu dòng tủy và mono bào vị thành niên
8	Bệnh bạch cầu tủy bào cấp tính Bệnh bạch cầu tủy bào mạn tính Ung thư mô liên kết Ewing	Ung thư biểu mô tuyến (thận)
9	Đa u tủy xương Bệnh tăng hồng cầu vô căn	
10	Bệnh bạch cầu nguyên bào cấp tính Ung thư biểu mô tuyến (tử cung)	U bào hình sao Đa u tủy xương
11	Đa u tủy xương	
12	Bệnh bạch cầu lymphô bào mạn tính Khối u Wilms	Đa u tủy xương
13	Bệnh bạch cầu tủy bào cấp tính Khối u Wilms	Đa u tủy xương
14	Bệnh bạch cầu nguyên bào cấp tính	Ung thư biểu mô tuyến (thận) U màng não
15	Đa u tủy xương	
16	Ung thư biểu mô tuyến (thận)	Đa u tủy xương
17	Ung thư biểu mô tuyến (thận) Bệnh bạch cầu nguyên bào cấp tính	
18	Bệnh bạch cầu nguyên bào cấp tính Khối u Wilm	Ung thư biểu mô tuyến (thận)
19	Đa u tủy xương Bệnh bạch cầu tủy bào mạn tính	Ung thư biểu mô tuyến (Vú) U màng não
20	U nguyên bào gan Ung thư biểu mô tuyến (thận)	

21	Bệnh bạch cầu nguyên bào cấp tính Bệnh bạch cầu nguyên mẫu tiêu cầu cấp tính	
22	Bệnh bạch cầu nguyên bào cấp tính	U màng não
X	Bệnh bạch cầu nguyên bào cấp tính U lymphô có nang	
Y		

Theo các phương án khác nhau, các phương pháp được mô tả ở đây có thể được sử dụng để phát hiện và/hoặc định lượng các hiện tượng lệch bội nhiễm sắc thể toàn bộ mà có liên quan tới ung thư nói chung, và/hoặc có liên quan tới các bệnh ung thư cụ thể. Vì vậy, ví dụ, theo một số phương án nhất định, sự phát hiện thấy và/hoặc định lượng các hiện tượng lệch bội nhiễm sắc thể toàn bộ đặc trưng bởi sự tăng hoặc giảm được thể hiện trong Bảng 2 được xem xét.

Các biến thể số lượng bản sao đoạn nhiễm sắc thể cấp độ nhánh.

Nhiều nghiên cứu đã báo cáo các kiểu mẫu của các biến thể số lượng bản sao cấp độ nhánh trên toàn bộ số lượng lớn các mẫu ung thư (Lin et al. Cancer Res 68, 664- 673 (2008); George et al. PLoS ONE 2, e255 (2007); Demichelis et al. Genes Chromosomes Cancer 48: 366-380 (2009); Beroukheim et al. Nature. 463(7283): 899-905). Ngoài ra, quan sát thấy rằng tần suất của các biến thể số lượng bản sao cấp độ nhánh giảm theo độ dài của các nhánh nhiễm sắc thể. Điều chỉnh theo xu hướng này, phần lớn các nhánh nhiễm sắc thể đều có bằng chứng chắc chắn về sự tăng hoặc giảm ưu tiên, nhưng hiếm gặp cả hai, ở nhiều dòng ung thư (xem, ví dụ, Beroukheim et al. Nature. 463(7283): 899-905).

Do vậy, theo một phương án, các phương pháp mô tả ở đây được sử dụng để xác định các CNV cấp độ nhánh (các CNV bao gồm một nhánh nhiễm sắc thể hoặc gần như một nhánh nhiễm sắc thể) trong mẫu. Các CNV có thể được xác định trong các CNV trong mẫu thử nghiệm bao gồm axit nucleic cấu tạo (dòng giao tử) và các CNV cấp độ nhánh có thể được nhận dạng ở các axit nucleic cấu tạo này. Theo một số phương án nhất định, các CNV cấp độ nhánh được nhận dạng (nếu có mặt) trong mẫu bao gồm hỗn hợp các axit nucleic (ví dụ, các axit nucleic có nguồn gốc từ các axit bình thường và các axit nucleic có nguồn gốc từ các tế bào khối u). Theo một số phương án nhất định, mẫu này có nguồn gốc từ đối tượng mà bị

nghi ngờ hoặc được biết mắc bệnh ung thư ví dụ ung thư biểu mô, ung thư mô liên kết, u lymphô, bệnh bạch cầu, các u tế bào mầm, u nguyên bào, và tương tự. Theo một phương án, mẫu này là mẫu huyết tương có nguồn gốc (được xử lý) từ máu ngoại vi mà có thể bao gồm hỗn hợp của cfADN có nguồn gốc từ các tế bào bình thường và ung thư. Theo một phương án khác, mẫu sinh học mà được sử dụng để xác định xem có CNV hay không có nguồn gốc từ tế bào mà, nếu có ung thư, bao gồm hỗn hợp của các tế bào ung thư và không ung thư từ các mô sinh học khác bao gồm nhưng không bị giới hạn ở các chất lỏng sinh học như huyết thanh, mô hôi, nước mắt, nước bọt, nước tiểu, dịch tai, bạch huyết, nước bọt, dịch não tủy, dịch huyệt, dịch huyền phù tủy xương, dịch âm đạo, dịch rửa qua cổ, dịch não, dịch tích tụ ở bụng, sữa, các chất bài tiết của các đường hô hấp, ruột và tiết niệu, và các mẫu gạn tách bạch cầu, hoặc trong các sinh tiết mô, các miếng gạc, hoặc các kính phết. Theo các phương án khác, mẫu sinh học là mẫu chất thải rắn từ người (phân).

Theo các phương án khác nhau, các CNV được nhận dạng là chỉ báo về sự có mặt của ung thư hoặc nguy cơ gia tăng về mắc ung thư bao gồm nhưng không bị giới hạn ở các CNV cấp độ nhánh được liệt kê trong Bảng 3. Như được minh họa trong Bảng 3, một số CNV nhất định mà bao gồm tăng cấp độ nhánh quan trọng là chỉ ra sự có mặt của ung thư hoặc an nguy cơ gia tăng về mắc một số bệnh ung thư. Vì vậy, ví dụ, tăng lên ở 1q là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc bệnh bạch cầu tăng lymphô bào cấp tính (acute lymphoblastic leukemia - ALL), ung thư vú, GIST, HCC, NSC phổi, u nguyên bào tủy, u melanin, MPD, ung thư buồng trứng, và/hoặc ung thư tuyến tiền liệt. Tăng lên ở 3q là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư vảy thực quản, SC phổi, và/hoặc MPD. Tăng lên ở 7q là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư đại trực tràng, u thần kinh đệm, HCC, NSC phổi, u nguyên bào tủy, u melanin, ung thư tuyến tiền liệt, và/hoặc ung thư thận. Tăng lên ở 7p là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư vú, ung thư đại trực tràng, ung thư biểu mô tuyến thực quản, u thần kinh đệm, HCC, NSC phổi, u nguyên bào tủy, u melanin, và/hoặc ung thư thận. Tăng lên ở 20q là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư vú, ung thư đại trực tràng, u mô ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, ung thư u thần kinh đệm, HCC, NSC phổi, u melanin, ung thư buồng trứng, và/hoặc ung thư thận, và v.v..

Tương tự như được minh họa trong Bảng 3 một số CNV nhất định mà bao gồm mất cấp độ nhánh đáng kể là chỉ ra sự có mặt của và/hoặc nguy cơ gia tăng mắc một số bệnh ung thư. Vì vậy, ví dụ, giảm ở 1p là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc u mô đệm đường tiêu hóa. Giảm ở 4q là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư đại trực tràng, ung thư biểu mô tuyến thực quản, SC phổi, u melanin, ung thư buồng trứng, và/hoặc ung thư thận. Giảm ở 17p là chỉ báo về sự có mặt hoặc nguy cơ gia tăng mắc ung thư vú, ung thư đại trực tràng, ung thư biểu mô tuyến thực quản, HCC, NSC phổi, SC phổi, và/hoặc ung thư buồng trứng, và tương tự.

BẢNG 3. Các biến thể số lượng bản sao đoạn nhiễm sắc thể cấp độ nhánh quan trọng ở mỗi trong số 16 phân nhóm ung thư (vú, trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, ung thư vảy thực quản, GIST (u mô đệm đường tiêu hóa), u thần kinh đệm, HCC (ung thư biểu mô tế bào gan), NSC phổi, SC phổi, u nguyên bào tủy, u melanin, MPD (bệnh tăng sinh tủy), buồng trứng, tuyến tiền liệt, bệnh bạch cầu tăng lymphô bào cấp tính (ALL), và thận) (xem ví dụ. Beroukhi et al. Nature (2010) 463(7283): 899-905).

Nhánh	Các loại ung thư tăng lên đáng kể ở	Các loại ung thư giảm xuống đáng kể ở	Gen gây ung thư/Gen ức chế khối u đã biết
1p	----	GIST	
1q	ALL, vú, GIST, HCC, NSC phổi, u nguyên bào tủy, u melanin, MPD, buồng trứng, tuyến tiền liệt	----	
3p	----	Vảy thực quản, NSC phổi, SC phổi, thận	<i>VHL</i>
3q	Vảy thực quản, SC phổi, MPD	----	
4p	ALL	Vú, ung thư biểu mô tuyến thực quản, thận	
4q	ALL	Trực kết tràng, ung thư biểu mô tuyến thực quản, SC phổi, u melanin, buồng trứng, thận	
5p	Ung thư vảy thực quản, HCC, NSC phổi, SC phổi, ung thư thận	----	<i>TERT</i>
5q	HCC, thận	Ung thư biểu mô tuyến thực quản, NSC phổi	<i>APC</i>
6p	ALL, HCC, NSC phổi, u melanin	----	
6q	ALL	U melanin, thận	

7p	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, u thần kinh đệm, HCC, NSC phổi, u nguyên bào tủy, u melanin, thận	----	<i>EGFR</i>
7q	Trực kết tràng, u thần kinh đệm, HCC, NSC phổi, u nguyên bào tủy, u melanin, tuyến tiền liệt, thận	----	<i>BRAF, MET</i>
8p	ALL, M PD	Ung thư vú, HCC, NSC phổi, u nguyên bào tủy, tuyến tiền liệt, thận	
8q	ALL, vú, trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, HCC, NSC phổi, M P D, buồng trứng, tuyến tiền liệt	U nguyên bào tủy	<i>MYC</i>
9p	MPD	ALL, vú, ung thư biểu mô tuyến, NSC phổi, u melanin, buồng trứng, thận	<i>CDKN2A/B</i>
9q	ALL, MPD	NSC phổi, u melanin, ung thư buồng trứng, ung thư thận	
10p	ALL	U thần kinh đệm, SC phổi, u melanin	
10q	ALL	U thần kinh đệm, SC phổi, u nguyên bào tủy, u melanin	<i>PTEN</i>
11p	----	U nguyên bào tủy	<i>WT1</i>
11q	----	U mỡ ác tính phân biệt hóa, u nguyên bào tủy, U melanin	<i>ATM</i>
12p	Trực kết tràng, thận	----	<i>KRAS</i>
12q	Thận	----	
13q	Trực kết tràng	Vú, u mỡ ác tính phân biệt hóa, u thần kinh đệm, NSC phổi, buồng trứng	<i>RB/IBRCA2</i>
14q	ALL, NSC phổi, SC phổi, tuyến tiền liệt	GIST, u melanin, thận	
15q	----	GIST, NSC phổi, SC phổi, buồng trứng	
16p	Vú	----	
16q	----	Ung thư vú, HCC, u nguyên bào tủy, buồng trứng, tuyến tiền liệt	
17p	ALL	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, HCC, NSC phổi, SC phổi, buồng trứng	<i>TP53</i>
17q	ALL, HCC, NSC phổi, u nguyên bào tủy	Vú, buồng trứng	<i>ERBB2, NF1/BRCA1</i>
18p	ALL, u nguyên bào tủy	Trực kết tràng, NSC phổi	

18q	ALL, u nguyên bào tủy	Ung thư trực kết tràng, ung thư biểu mô tuyến thực quản, NSC phổi	<i>SMAD2, SMAD4</i>
19p	U thần kinh đệm	Ung thư biểu mô tuyến thực quản, NSC phổi, u melanin, buồng trứng	
19q	U thần kinh đệm, SC phổi	Ung thư biểu mô tuyến thực quản, NSC phổi	
20p	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, GIST, u thần kinh đệm, HCC, NSC phổi, u melanin, thận	----	
20q	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, u thần kinh đệm, HCC, NSC phổi, u melanin, buồng trứng, thận	----	
21q	ALL, GIST, MPD	----	
22q	U melanin	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, GIST, NSC phổi, SC phổi, buồng trứng, tuyến tiền liệt	<i>NF2</i>

Các ví dụ về các kết hợp giữa các biến thể số lượng bản sao cấp độ nhánh được dự định có tính chất minh họa chứ không giới hạn sáng chế. Các biến thể số lượng bản sao cấp độ nhánh khác và các kết hợp ung thư của chúng được biết đối với người có trình độ trong lĩnh vực kỹ thuật này.

Các biến thể số lượng bản sao nhỏ hơn, ví dụ tiêu điểm

Như nêu trên đây, theo một số phương án nhất định, các phương pháp được mô tả ở đây có thể được sử dụng để xác định có hoặc không có khuếch đại nhiễm sắc thể. Theo một số phương án, khuếch đại nhiễm sắc thể là sự tăng lên của một hoặc nhiều toàn bộ nhiễm sắc thể. Theo các phương án khác, khuếch đại nhiễm sắc thể là sự tăng lên của một hoặc nhiều đoạn của nhiễm sắc thể. Theo các phương án khác, khuếch đại nhiễm sắc thể là sự tăng lên của hai hoặc nhiều đoạn trong số hai hoặc nhiều nhiễm sắc thể. Theo các phương án khác nhau, khuếch đại nhiễm sắc thể có thể liên quan tới sự tăng lên của một hoặc nhiều gen gây ung thư.

Các gen hoạt động mạnh liên quan tới các khối u rắn ở người thông thường gây ra tác động bởi sự biểu thị quá mức hoặc biểu thị đã biến đổi. Khuếch đại gen là cơ chế phổ biến dẫn

đến tăng điều hòa mức biểu hiện gen. Các bằng chứng từ các nghiên cứu tế bào di truyền chỉ ra rằng khuếch đại đáng kể xảy ra ở trên 50% các bệnh ung thư vú ở người. Đáng lưu ý nhất là, sự khuếch đại của thụ thể yếu tố tăng trưởng biểu bì người tiền gen ung thư loại 2 (HER2) nằm trên nhiễm sắc thể 17 (17(17q21- q22)), gây ra sự biểu hiện quá mức của các thụ thể HER2 trên bề mặt tế bào dẫn đến sự báo hiệu quá mức và mất điều hòa ở bệnh ung thư vú và các ung thư ác tính khác (Park et al., *Clinical Breast* 8:392-401). Các gen gây ung thư đã được phát hiện thấy là được khuếch đại ở các ung thư ác tính khác ở người. Các ví dụ về sự khuếch đại của các gen gây ung thư tế bào ở các khối u người bao gồm các khuếch đại của: c-myc ở dòng tế bào của bệnh bạch cầu tiền tủy bào HL60, và ở dòng tế bào của ung thư biểu mô phổi tế bào nhỏ, N-myc ở các u nguyên bào sợi thần kinh sơ cấp (các giai đoạn III và IV), các dòng tế bào của u nguyên bào sợi thần kinh, dòng tế bào của u nguyên bào võng mạc và các khối u sơ cấp, và các dòng và khối u của ung thư biểu mô tế bào nhỏ, L-myc ở dòng tế bào và khối u của ung thư biểu mô phổi tế bào nhỏ, c-myb ở bệnh bạch cầu tủy xương cấp tính và ở dòng tế bào của ung thư biểu mô ruột kết, c-erbB ở tế bào ung thư biểu mô dạng u biểu bì, và các u thần kinh đệm sơ cấp, c-K-ras-2 ở các ung thư biểu mô sơ cấp của phổi, ruột kết, bàng quang, và ruột thẳng, N-ras ở dòng tế bào của ung thư biểu mô ở động vật có vú (Varmus H., *Ann Rev Genetics* 18: 553-612 (1984) [trích dẫn trong Watson et al., *Molecular Biology of the Gene* (4th ed.; Benjamin/Cummings Publishing Co. 1987)]).

Các lặp đoạn của các gen gây ung thư là nguyên nhân phổ biến của nhiều loại ung thư, như trường hợp với khuếch đại P70-S6 Kinaza 1 và ung thư vú. Trong các trường hợp như vậy, lặp đoạn di truyền xảy ra ở tế bào xôma và chỉ ảnh hưởng đến bộ gen của chính các tế bào ung thư, chứ không phải toàn bộ cơ thể, ít ảnh hưởng hơn nhiều đến con cháu sau này. Các ví dụ khác về các gen gây ung thư mà được khuếch đại ở các bệnh ung thư người bao gồm MYC, ERBB2 (EGFR), CCND1 (Cyclin D1), FGFR1 và FGFR2 ở ung thư vú, MYC và ERBB2 ở ung thư cổ, HRAS, KRAS, và MYB ở ung thư đại trực tràng, MYC, CCND1 và MDM2 ở ung thư thực quản, CCNE, KRAS và MET ở ung thư dạ dày, ERBB1, và CDK4 ở u nguyên bào đệm, CCND1, ERBB1, và MYC ở ung thư đầu và cổ, CCND1 ở ung thư tế bào gan, MYCB ở u nguyên bào sợi thần kinh, MYC, ERBB2 và AKT2 ở ung thư buồng trứng, MDM2 và CDK4 ở ung thư mô liên kết, và MYC ở ung thư phổi tế bào nhỏ. Theo một phương án, phương pháp theo sáng chế có thể được sử dụng để xác định có hoặc không có sự khuếch

đại của gen gây ung thư liên quan tới ung thư. Theo một số phương án, khuếch đại gen gây ung thư có liên quan tới ung thư vú, ung thư cổ, ung thư đại trực tràng, ung thư thực quản, ung thư dạ dày, u nguyên bào đệm, ung thư đầu và cổ, ung thư tế bào gan, u nguyên bào sợi thần kinh, ung thư buồng trứng, ung thư mô liên kết, và ung thư phổi tế bào nhỏ.

Theo một phương án, phương pháp theo sáng chế có thể được sử dụng để xác định có hoặc không có sự mất đoạn nhiễm sắc thể. Theo một số phương án, mất đoạn nhiễm sắc thể là sự mất của một hoặc nhiều toàn bộ nhiễm sắc thể. Theo các phương án khác, mất đoạn nhiễm sắc thể là sự mất của một hoặc nhiều đoạn của nhiễm sắc thể. Theo các phương án khác, mất đoạn nhiễm sắc thể là sự mất của hai hoặc nhiều đoạn trong số hai hoặc nhiều nhiễm sắc thể. Mất đoạn nhiễm sắc thể có thể liên quan tới mất một hoặc nhiều gen ức chế ung thư.

Các mất đoạn nhiễm sắc thể liên quan tới gen ức chế ung thư được cho là đóng một vai trò quan trọng trong sự phát triển và tiến triển của các khối u rắn. Gen ức chế khối u nguyên bào võng mạc (Rb-1), đặt trong nhiễm sắc thể 13q14, là gen ức chế khối u được mô tả khái quát nhất. Sản phẩm gen Rb-1, phosphoprotein nhân 105 kDa, rõ ràng đóng một vai trò quan trọng trong việc điều hòa chu kỳ tế bào (Howe et al., Proc Natl Acad Sci (Mỹ) 87:5883-5887). Sự biểu hiện biến đổi hoặc bị mất của protein Rb được gây ra bởi sự khử hoạt tính của cả các alen gen hoặc thông qua đột biến điểm hoặc mất đoạn nhiễm sắc thể. Các biến đổi gen Rb-1 đã được phát hiện thấy có mặt không chỉ ở các u nguyên bào võng mạc mà còn ở các ung thư ác tính khác như các ung thư xương, ung thư phổi tế bào nhỏ (Rygaard et al., Cancer Res 50: 5312-5317 [1990]) và ung thư vú. Các nghiên cứu đa hình chiều dài đoạn cắt giới hạn (Restriction fragment length polymorphism - RFLP) đã chỉ ra rằng các loại khối u như vậy thường xuyên mất tính dị hợp tử ở 13q gợi ý rằng một trong số các alen Rb-1 bị mất do sự mất đoạn toàn bộ nhiễm sắc thể (Bowcock et al., Am J Hum Genet, 46: 12 [1990]). Các bất thường ở nhiễm sắc thể số 1 bao gồm các lặp đoạn, các mất đoạn và chuyển đoạn không cân bằng liên quan đến nhiễm sắc thể số 6 và các nhiễm sắc thể cặp khác chỉ ra rằng các vùng của nhiễm sắc thể số 1, cụ thể là 1q21-1q32 và 1p11-13, có thể chứa các gen gây ung thư hoặc gen ức chế ung thư mà là liên quan về mầm bệnh đến cả hai giai đoạn mạn tính và giai đoạn cuối của bệnh u tăng sinh tủy (Caramazza et al., Eur J Hematol 84:191-200 [2010]). U tăng sinh tủy còn liên quan đến các mất đoạn của nhiễm sắc thể số 5. Các mất đoạn toàn phần hoặc mất

đoạn giữa của nhiễm sắc thể số 5 là bất thường kiểu nhân thường gặp nhất ở các hội chứng loạn sản tủy (MDS). Những người bệnh del(5q)/5q- MDS đã được phân lập có tiên lượng tốt hơn so với các người bệnh có khiếm khuyết kiểu nhân bổ sung, là những người có xu hướng phát triển u tăng sinh tủy (MPN) và bệnh bạch cầu tủy xương cấp tính. Tần xuất các mất đoạn nhiễm sắc thể số 5 không cân bằng dẫn đến ý kiến rằng 5q chứa một hoặc nhiều gen ức chế khối u mà có vai trò to lớn trong việc kiểm soát sự phát triển của các tế bào mầm/gốc tạo huyết (hematopoietic stem/progenitor cells - SCs/HPCs). Bản đồ di truyền tế bào của các vùng mất đoạn thông thường (commonly deleted regions - CDRs) tập trung ở 5q31 và 5q32 được nhận dạng là các gen ức chế khối u ứng viên, bao gồm tiểu đơn vị ribosome RPS14, yếu tố phiên mã Egr1/Krox20 và protein tổ chức lại khung tế bào, alpha-catenin (Eisenmann et al., *Oncogen* 28:3429-3441). Các nghiên cứu di truyền tế bào và kiểu alen của các khối u mới và các dòng tế bào khối u đã chỉ ra rằng sự mất alen từ các vùng phân biệt trên nhiễm sắc thể 3p, bao gồm 3p25, 3p21-22, 3p21.3, 3p12-13 và 3p14, là các bất thường hệ gen sớm nhất và thường xuyên nhất đối với phổ rộng các bệnh ung thư biểu mô chủ yếu ở phổi, vú, thận, đầu và cổ, buồng trứng, cổ tử cung, ruột kết, tuyến tụy, thực quản, bàng quang và các cơ quan khác. Một số gen ức chế ung thư đã được ánh xạ đến vùng nhiễm sắc thể 3p, và được cho rằng quá trình tăng methyl hóa các mất đoạn giữa hoặc gen khởi đầu xảy ra trước khi mất 3p hoặc toàn bộ nhiễm sắc thể số 3 trong quá trình phát triển ung thư biểu mô (Angeloni D., *Briefings Functional Genomics* 6:19-39).

Trẻ sơ sinh và trẻ nhỏ mắc hội chứng Down (DS) thường mắc phải bệnh bạch cầu thoáng qua bẩm sinh và có nguy cơ mắc bệnh ung thư bạch cầu dạng tủy cấp tính và ung thư bạch cầu tăng lymphô bào cấp tính. Nhiễm sắc thể số 21, có khoảng 300 gen, có thể liên quan đến nhiều dạng sai hình cấu trúc, ví dụ, các hoán vị, các mất đoạn, và các khuếch đại, ở bệnh bạch cầu, u lymphô, và khối u rắn. Ngoài ra, các gen nằm trên nhiễm sắc thể số 21 đã được xác định rằng đóng vai trò quan trọng trong việc hình thành khối u. Các dạng sai hình về số xôma cũng như cấu trúc của nhiễm sắc thể số 21 có liên quan tới bệnh bạch cầu, và các gen đặc hiệu bao gồm RUNX1, TMPRSS2, và TFF, mà nằm trên 21q, đóng vai trò trong việc hình thành khối u (Fonatsch C *Gene Chromosomes Cancer* 49:497-508).

Từ các vấn đề trên, theo các phương án khác nhau, các phương pháp được mô tả ở đây có thể được sử dụng để xác định các CNV đoạn mà đã biết là bao gồm một hoặc nhiều gen gây ung thư hoặc gen ức chế ung thư, và/hoặc đã biết là có liên quan tới ung thư hoặc nguy cơ gia tăng mắc ung thư. Theo một số phương án nhất định, các CNV có thể được xác định trong mẫu thử nghiệm bao gồm axit nucleic cấu tạo (dòng giao tử) và đoạn này có thể được nhận dạng trong các axit nucleic cấu tạo này. Theo một số phương án nhất định các CNV đoạn được nhận dạng (nếu có mặt) trong mẫu bao gồm hỗn hợp các axit nucleic (ví dụ, các axit nucleic có nguồn gốc từ các axit bình thường và các axit nucleic có nguồn gốc từ các tế bào khối u). Theo một số phương án nhất định mẫu này có nguồn gốc từ đối tượng mà bị nghi ngờ hoặc được biết mắc bệnh ung thư ví dụ ung thư biểu mô, ung thư mô liên kết, u lymphô, bệnh bạch cầu, các u tế bào mầm, u nguyên bào, và tương tự. Theo một phương án, mẫu này là mẫu huyết tương có nguồn gốc (được xử lý) từ máu ngoại vi mà có thể bao gồm hỗn hợp của cfADN có nguồn gốc từ tế bào bình thường và tế bào ung thư. Theo một phương án khác, mẫu sinh học mà được sử dụng để xác định xem có CNV hay không có nguồn gốc từ tế bào mà, nếu có ung thư, bao gồm hỗn hợp của các tế bào ung thư và không ung thư từ các mô sinh học khác bao gồm nhưng không chỉ giới hạn ở các chất lỏng sinh học như huyết thanh, mô hôi, nước mắt, nước bọt, nước tiểu, nước bọt, dịch tai, bạch huyết, nước bọt, dịch não tủy, dịch huyệt, dịch huyền phù tủy xương, dịch âm đạo, dịch rửa qua cổ, dịch não, các dịch tích tụ ở bụng, sữa, các chất bài tiết của các đường hô hấp, ruột và tiết niệu, và các mẫu gạn tách bạch cầu, hoặc trong các sinh tiết mô, các miếng gạc, hoặc các kính phết. Theo các phương án khác, mẫu sinh học là mẫu chất thải rắn từ người (phân).

Các CNV được sử dụng để xác định sự có mặt của bệnh và/hoặc nguy cơ mắc ung thư gia tăng có thể bao gồm sự khuếch đại hoặc các mất đoạn.

Theo các phương án khác nhau, các CNV được xác định là biểu thị sự có mặt của ung thư hoặc nguy cơ gia tăng mắc ung thư bao gồm một hoặc nhiều trong số các khuếch đại được thể hiện ở Bảng 4.

BẢNG 4: Các đoạn nhiễm sắc thể minh họa nhưng không giới hạn sáng chế được đặc trưng bởi các khuếch đại có liên quan tới ung thư. Các loại ung thư được liệt

kê là các ung thư được nhận dạng trong tài liệu Beroukhim *et al.* Nature 18: 463: 899-905.

Vùng định	Độ dài (Mb)	Các loại ung thư được nhận dạng trong phép phân tích này chứ không phải trong các ấn phẩm công bố trước đó
chr1:119996566-120303234	0,228	Vú, SC phổi, u melanin
chr1:148661965-149063439	0,35	Vú, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, tế bào gan, SC phổi, u melanin, buồng trứng, tuyến tiền liệt, thận
chr1:1-5160566	4,416	Ung thư biểu mô tuyến thực quản, buồng trứng
chr1:158317017-159953843	1,627	U mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, tuyến tiền liệt, thận
chr1:169549478-170484405	0,889	Ung thư trực kết tràng, u mỡ ác tính phân biệt hóa, tuyến tiền liệt, thận
chr1:201678483-203358272	1,471	Tuyến tiền liệt
chr1:241364021-247249719	5,678	NSC phổi, u melanin, buồng trứng
chr1:39907605-40263248	0,319	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, NSC phổi, SC phổi
chr1:58658784-60221344	1,544	Vú, u mỡ ác tính phân biệt hóa, SC phổi
chr3:170024984-173604597	3,496	Vú, ung thư biểu mô tuyến thực quản, U thần kinh đệm
chr3:178149984-199501827	21,123	Vảy thực quản, NSC phổi
chr3:86250885-95164178	8,795	SC phổi, u melanin
chr4:54471680-55980061	1,449	NSC phổi
chr5:1212750-1378766	0,115	U mỡ ác tính phân biệt hóa
chr5:174477192-180857866	6,124	Vú, NSC phổi
chr5:45312870-49697231	4,206	SC phổi
chr6:1-23628840	23,516	Ung thư biểu mô tuyến thực quản
chr6:135561194-135665525	0,092	Vú, ung thư biểu mô tuyến thực quản
chr6:43556800-44361368	0,72	Ung thư biểu mô tuyến thực quản, tế bào gan, buồng trứng

chr6:63255006-65243766	1,988	Ung thư biểu mô tuyến thực quản, NSC phổi
chr7:115981465-116676953	0,69	Ung thư biểu mô tuyến thực quản, NSC phổi, u melanin, buồng trứng
chr7:54899301-55275419	0,363	Ung thư biểu mô tuyến thực quản, vảy thực quản
chr7:89924533-98997268	9,068	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản, buồng trứng
chr8:101163387-103693879	2,516	NSC phổi, u melanin, buồng trứng
chr8:116186189-120600761	4,4	Vú, tế bào gan, NSC phổi, Buồng trứng
chr8:128774432-128849112	0,009	Ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, SC phổi, u nguyên bào tủy, rối loạn tăng sinh tủy, buồng trứng
chr8:140458177-146274826	5,784	NSC phổi, u nguyên bào tủy, u melanin, Buồng trứng
chr8:38252951-38460772	0,167	Trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản
chr8:42006632-42404492	0,257	Ung thư biểu mô tuyến thực quản, NSC phổi, SC phổi, buồng trứng, tuyến tiền liệt
chr8:81242335-81979194	0,717	Vú, u melanin
chr9:137859478-140273252	2,29	Ung thư trực kết tràng, u mỡ ác tính phân biệt hóa
chr10:74560456-82020637	7,455	Vú, buồng trứng, tuyến tiền liệt
chr11:101433436-102134907	0,683	NSC phổi, SC phổi
chr11:32027116-37799354	5,744	Vú, u mỡ ác tính phân biệt hóa, NSC phổi, SC phổi
chr11:69098089-69278404	0,161	U mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, tế bào gan, SC phổi, buồng trứng
chr11:76699529-78005085	1,286	U mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, SC phổi, buồng trứng
chr12:1-1311104	1,271	NSC phổi
chr12:25189655-25352305	0,112	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản, vảy thực quản, buồng trứng
chr12:30999223-32594050	1,577	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, NSC phổi, SÂNG CHẾ phổi
chr12:38788913-42596599	3,779	Vú, ung thư trực kết tràng, u mỡ ác tính phân biệt hóa, vảy thực quản, NSC phổi, SC phổi
chr12:56419524-56488685	0,021	U mỡ ác tính phân biệt hóa, u melanin, thận

chr12:64461446-64607139	0,041	U mỡ ác tính phản biệt hóa, thận
chr12:66458200-66543552	0,058	U mỡ ác tính phản biệt hóa, vảy da thực quản, thận
chr12:67440273-67566002	0,067	Vú, u mỡ ác tính phản biệt hóa, vảy thực quản, u melanin, thận
chr12:68249634-68327233	0,06	Vú, u mỡ ác tính phản biệt hóa, vảy thực quản, thận
chr12:70849987-70966467	0,036	U mỡ ác tính phản biệt hóa, thận
chr12:72596017-73080626	0,23	Thận
chr12:76852527-77064746	0,158	U mỡ ác tính phản biệt hóa
chr12:85072329-85674601	0,272	U mỡ ác tính phản biệt hóa
chr12:95089777-95350380	0,161	U mỡ ác tính phản biệt hóa
chr13:108477140-110084607	1,6	Vú, ung thư biểu mô tuyến thực quản, NSC phổi, SC phổi
chr13:1-40829685	22,732	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản
chr13:89500014-93206506	3,597	Vú, ung thư biểu mô tuyến thực quản, u nguyên bào tủy
chr14:106074644-106368585	0,203	Vảy thực quản
chr14:1-23145193	3,635	Bệnh bạch cầu tăng lymphô bào cấp tính, vảy thực quản, tế bào gan, SC phổi
chr14:35708407-36097605	0,383	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, tuyến tiền liệt
chr15:96891354-97698742	0,778	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, NSC phổi, u nguyên bào tủy, u melanin
chr17:18837023-19933105	0,815	Vú, tế bào gan
chr17:22479313-22877776	0,382	Vú, NSC phổi
chr17:24112056-24310787	0,114	Vú, NSC phổi
chr17:35067383-35272328	0,149	Trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản
chr17:44673157-45060263	0,351	U melanin
chr17:55144989-55540417	0,31	NSC phổi, u nguyên bào tủy, u melanin, buồng trứng
chr17:62318152-63890591	1,519	Vú, NSC phổi, u melanin, buồng trứng

chr17:70767943-71305641	0,537	Vú, NSC phổi, u melanin, buồng trứng
chr18:17749667-22797232	5,029	Trực kết tràng, ung thư biểu mô tuyến thực quản, buồng trứng
chr19:34975531-35098303	0,096	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản
chr19:43177306-45393020	2,17	NSC phổi, buồng trứng
chr19:59066340-59471027	0,321	Vú, NSC phổi, buồng trứng
chr2:15977811-16073001	0,056	SC phổi
chr20:29526118-29834552	0,246	Buồng trứng
chr20:51603033-51989829	0,371	Tế bào gan, NSC phổi, buồng trứng
chr20:61329497-62435964	0,935	Tế bào gan, NSC phổi
chr22:19172385-19746441	0,487	Trực kết tràng, u melanin, buồng trứng
chrX:152729030-154913754	1,748	Vú, NSC phổi, thận
chrX:66436234-67090514	0,267	Buồng trứng, tuyến tiền liệt

Theo một số phương án nhất định kết hợp với các khuếch đại mô tả trên đây (ở đây), hoặc riêng biệt, các CNV được nhận dạng là chỉ ra sự có mặt của ung thư hoặc nguy cơ gia tăng về mắc ung thư bao gồm một hoặc nhiều trong số các mất đoạn được thể hiện trong Bảng 5.

BẢNG 5. Các đoạn nhiễm sắc thể minh họa nhưng không giới hạn sáng chế được đặc trưng bởi các mất đoạn có liên quan tới ung thư. Các loại ung thư được liệt kê là các loại ung thư được nhận dạng trong Beroukhim *et al.* Nature 18: 463: 899-905.

Vùng đỉnh	Độ dài (Mb)	Các loại ung thư được nhận dạng ở phép phân tích này chứ không phải trong các ấn phẩm công bố trước đó
chr1:110339388-119426489	1p13.2	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản, NSC phổi, SC phổi, u melanin, buồng trứng, tuyến tiền liệt
chr1:223876038-247249719	1q43	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, SC phổi, u melanin, tuyến tiền liệt

chr1:26377344-27532551	1p36.11	Vú, ung thư biểu mô tuyến thực quản, Vảy thực quản, NSC phổi, SÁNG CHẾ phổi, u nguyên bào tủy, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt
chr1:3756302-6867390	1p36.31	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, vảy thực quản, tế bào gan, NSC phổi, SC phổi, u nguyên bào tủy, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt, thận
chr1:71284749-74440273	1p31.1	Vú, ung thư biểu mô tuyến thực quản, u thần kinh đệm, tế bào gan, NSC phổi, SC phổi, u melanin, buồng trứng, thận
chr2:1-15244284	2p25.3	NSC phổi, buồng trứng
chr2:138479322-143365272	2q22.1	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, buồng trứng, tuyến tiền liệt, thận
chr2:204533830-206266883	2q33.2	Ung thư biểu mô tuyến thực quản, tế bào gan, NSC phổi, u nguyên bào tủy, thận
chr2:241477619-242951149	2q37.3	Vú, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, SC phổi, u nguyên bào tủy, u melanin, buồng trứng, thận
chr3:116900556-120107320	3q13.31	U mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, tế bào gan, NSC phổi, u melanin, rối loạn tăng sinh tủy, tuyến tiền liệt
chr3:1-2121282	3p26.3	Trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, NSC phổi, u melanin, rối loạn tăng sinh tủy
chr3:175446835-178263192	3q26.31	Bệnh bạch cầu tăng lymphô bào cấp tính, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, NSC phổi, u melanin, rối loạn tăng sinh tủy, tuyến tiền liệt
chr3:58626894-61524607	3p14.2	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, SC phổi, u nguyên bào tủy, u melanin, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt, thận
chr4:1-435793	4p16.3	Rối loạn tăng sinh tủy
chr4:186684565-191273063	4q35.2	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản, NSC phổi, u nguyên bào tủy, u melanin, tuyến tiền liệt, thận

chr4:91089383-93486891	4q22.1	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản, tế bào gan, NSC phổi, thận
chr5:177541057-180857866	5q35.3	Vú, NSC phổi, rối loạn tăng sinh tủy, buồng trứng
chr5:57754754-59053198	5q11.2	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, SC phổi, u melanin, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt
chr5:85837489-133480433	5q21.1	Trực kết tràng, u mỡ ác tính phân biệt hóa, NSC phổi, SC phổi, rối loạn tăng sinh tủy, buồng trứng
chr6:101000242-121511318	6q22.1	Trực kết tràng, NSC phổi, SC phổi
chr6:1543157-2570302	6p25.3	Trực kết tràng, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, NSC phổi, SC phổi, buồng trứng, tuyến tiền liệt
chr6:161612277-163134099	6q26	Trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, NSC phổi, SÂNG CHÈ phổi, buồng trứng, tuyến tiền liệt
chr6:76630464-105342994	6q16.1	Trực kết tràng, tế bào gan, NSC phổi
chr7:141592807-142264966	7q34	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, buồng trứng, tuyến tiền liệt, thận
chr7:144118814-148066271	7q35	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản, NSC phổi, u melanin, rối loạn tăng sinh tủy, buồng trứng
chr7:156893473-158821424	7q36.3	Vú, ung thư biểu mô tuyến thực quản, vảy thực quản, NSC phổi, u melanin, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt
chr7:3046420-4279470	7p22.2	U melanin, rối loạn tăng sinh tủy, buồng trứng
chr7:65877239-79629882	7q21.11	Ung thư vú, u nguyên bào tủy, u melanin, rối loạn tăng sinh tủy, buồng trứng
chr8:1-392555	8p23.3	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, rối loạn tăng sinh tủy
chr8:2053441-6259545	8p23.2	Bệnh bạch cầu tăng lymphô bào cấp tính, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, rối loạn tăng sinh tủy

chr8:22125332-30139123	8p21.2	Bệnh bạch cầu tăng lymphô bào cấp tính, u mỡ ác tính phân biệt hóa, tế bào gan, rối loạn tăng sinh tủy, buồng trứng, thận
chr8:39008109-41238710	8p11.22	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, u mỡ ác tính phân biệt hóa, vảy thực quản, tế bào gan, NSC phổi, rối loạn tăng sinh tủy, thận
chr8:42971602-72924037	8q11.22	Vú, u mỡ ác tính phân biệt hóa, vảy thực quản, tế bào gan, NSC phổi, rối loạn tăng sinh tủy, thận
chr9:1-708871	9p24.3	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, NSC phổi, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt
chr9:21489625-22474701	9p21.3	Ung thư trực kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, rối loạn tăng sinh tủy, buồng trứng
chr9:36365710-37139941	9p13.2	Rối loạn tăng sinh tủy
chr9:7161607-12713130	9p24.1	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, ung thư trực kết tràng, ung thư biểu mô tuyến thực quản, tế bào gan, SC phổi, u nguyên bào tủy, u melanin, rối loạn tăng sinh tủy, buồng trứng, tuyến tiền liệt, thận
chr10:1-1042949	10p15.3	Trực kết tràng, NSC phổi, SC phổi, buồng trứng, tuyến tiền liệt, thận
chr10:129812260-135374737	10q26.3	Vú, trực kết tràng, u thần kinh đệm, NSC phổi, SC phổi, u melanin, buồng trứng, thận
chr10:52313829-53768264	10q11.23	Trực kết tràng, NSC phổi, SC phổi, buồng trứng, thận
chr10:89467202-90419015	10q23.31	Vú, SC phổi, buồng trứng, thận
chr11:107086196-116175885	11q23.1	Ung thư biểu mô tuyến thực quản, u nguyên bào tủy, thận
chr11:1-1391954	11p15.5	Vú, u mỡ ác tính phân biệt hóa, ung thư biểu mô tuyến thực quản, NSC phổi, u nguyên bào tủy, buồng trứng
chr11:130280899-134452384	11q25	Ung thư biểu mô tuyến thực quản, vảy thực quản, tế bào gan, NSC phổi, u nguyên bào tủy, thận
chr11:82612034-85091467	11q14.1	U melanin, thận
chr12:11410696-12118386	12p13.2	Vú, tế bào gan, rối loạn tăng sinh tủy, tuyến tiền liệt
chr12:131913408-132349534	12q24.33	U mỡ ác tính phân biệt hóa, NSC phổi, Rối loạn tăng sinh tủy
chr12:97551177-99047626	12q23.1	Vú, trực kết tràng, ung thư vảy thực quản, NSC phổi, rối loạn tăng sinh tủy

chr13:111767404-114142980	13q34	Vú, tế bào gan, NSC phổi
chr13:1-23902184	13q12.11	Vú, SC phổi, buồng trứng
chr13:46362859-48209064	13q14.2	Tế bào gan, SC phổi, rối loạn tăng sinh tủy, tuyến tiền liệt
chr13:92308911-94031607	13q31.3	Vú, tế bào gan, NSC phổi, thận
chr14:1-29140968	14q11.2	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản, rối loạn tăng sinh tủy
chr14:65275722-67085224	14q23.3	U mỡ ác tính phân biệt hóa, rối loạn tăng sinh tủy
chr14:80741860-106368585	14q32.12	Bệnh bạch cầu tăng lymphô bào cấp tính, u mỡ ác tính phân biệt hóa, u melanin, rối loạn tăng sinh tủy
chr15:1-24740084	15q11.2	Bệnh bạch cầu tăng lymphô bào cấp tính, vú, ung thư biểu mô tuyến thực quản, NSC phổi, rối loạn tăng sinh tủy, buồng trứng
chr15:35140533-43473382	15q15.1	Ung thư biểu mô tuyến thực quản, NSC phổi, rối loạn tăng sinh tủy
chr16:1-359092	16p13.3	Ung thư biểu mô tuyến thực quản, tế bào gan, NSC phổi, thận
chr16:31854743-53525739	16q11.2	Vú, tế bào gan, NSC phổi, u melanin, thận
chr16:5062786-7709383	16p13.3	Tế bào gan, NSC phổi, u nguyên bào tủy, u melanin, rối loạn tăng sinh tủy, buồng trứng, thận
chr16:76685816-78205652	16q23.1	Vú, trực kết tràng, ung thư biểu mô tuyến thực quản, tế bào gan, NSC phổi, SC phổi, u nguyên bào tủy, thận
chr16:80759878-82408573	16q23.3	Ung thư trực kết tràng, tế bào gan, thận
chr16:88436931-88827254	16q24.3	Ung thư trực kết tràng, tế bào gan, NSC phổi, tuyến tiền liệt, thận
chr17:10675416-12635879	17p12	NSC phổi, SC phổi, rối loạn tăng sinh tủy
chr17:26185485-27216066	17q11.2	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, NSC phổi, SC phổi, u melanin, rối loạn tăng sinh tủy, buồng trứng
chr17:37319013-37988602	17q21.2	Vú, trực kết tràng, u mỡ ác tính phân biệt hóa, SC phổi, u melanin, rối loạn tăng sinh tủy, buồng trứng
chr17:7471230-7717938	17p13.1	SC phổi, rối loạn tăng sinh tủy
chr17:78087533-78774742	17q25.3	Trực kết tràng, rối loạn tăng sinh tủy
chr18:1-587750	18p11.32	Rối loạn tăng sinh tủy

chr18:46172638-49935241	18q21.2	Ung thư biểu mô tuyến thực quản, NSC phổi
chr18:75796373-76117153	18q23	Trục kết tràng, ung thư biểu mô tuyến thực quản, vảy thực quản, buồng trứng, tuyến tiền liệt
chr19:1-526082	19p13.3	Tế bào gan, NSC phổi, thận
chr19:21788507-34401877	19p12	Tế bào gan, NSC phổi, thận
chr19:52031294-53331283	19q13.32	Vú, tế bào gan, NSC phổi, u nguyên bào tủy, buồng trứng, thận
chr19:63402921-63811651	19q13.43	Vú, trục kết tràng, u mỡ ác tính phân biệt hóa, tế bào gan, NSC phổi, u nguyên bào tủy, buồng trứng, thận
chr20:1-325978	20p13	Vú, u mỡ ác tính phân biệt hóa, NSC phổi
chr20:14210829-15988895	20p12.1	Ung thư biểu mô tuyến thực quản, NSC phổi, u nguyên bào tủy, u melanin, rối loạn tăng sinh tủy, tuyến tiền liệt, thận
chr21:38584860-42033506	21q22.2	Vú
chr22:20517661-21169423	22q11.22	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản
chr22:45488286-49691432	22q13.33	Vú, tế bào gan, NSC phổi, SC phổi
chrX:1-3243111	Xp22.33	Ung thư biểu mô tuyến thực quản, NSC phổi, SC phổi
chrX:31041721-34564697	Xp21.2	Bệnh bạch cầu tăng lymphô bào cấp tính, ung thư biểu mô tuyến thực quản, u thần kinh đệm

Các hiện tượng lệch bội lẻ được nhận dạng là đặc trưng về các bệnh ung thư khác nhau (ví dụ, các hiện tượng lệch bội lẻ được nhận dạng trong các Bảng 4 và 5) có thể chứa các gen được biết là có liên quan tới các căn bệnh ung thư (ví dụ, các chất ức chế khối u, các gen gây ung thư, v.v.). Các hiện tượng lệch bội này cũng có thể được thăm dò để nhận dạng các gen liên quan nhưng chưa biết trước đó.

Ví dụ, Beroukhir et al. *supra*, đánh giá các gen gây ung thư tiềm năng trong các biến thể số lượng bản sao nhờ sử dụng các mối quan hệ gen giữa các Loci20 liên quan (GRAIL - Gene Relationships Among Implicated Loci20), một thuật toán mà tìm kiếm các các mối quan hệ chức giữa các vùng thuộc bộ gen. GRAIL tính điểm cho mỗi gen trong tập hợp các vùng thuộc bộ gen về 'tính liên quan' của nó tới các gen ở các vùng khác dựa trên sự tương tự nguyên bản giữa các đoạn tóm tắt công bố của tất cả các tờ báo viện dẫn đến gen, với ý định là một số gen

đích sẽ hoạt động theo các quá trình phổ biến. Các phương pháp này cho phép nhận dạng /mô tả các gen trước đó không liên quan tới các bệnh ung thư cụ thể đang đề cập. Bảng 6 minh họa các gen đích được biết là nằm trong đoạn khuếch đại được nhận dạng và các gen được dự báo, và Bảng 7 minh họa các gen đích được biết là nằm trong đoạn mất được nhận dạng và các gen được dự báo.

BẢNG 6. Các đoạn và gen nhiễm sắc thể minh họa nhưng không giới hạn sáng chế được biết hoặc dự báo là có trong các vùng có đặc điểm là khuếch đại ở nhiều ung thư khác nhau (xem ví dụ, Beroukheim *et al. supra.*).

Nhiễm sắc thể và băng nhiễm sắc thể	Vùng đỉnh	# các gen	Đích đã biết	Đích hàng đầu GRAIL
8q24.21	chr8:128774432-128849112	1	<i>MYC</i>	<i>MYC</i>
11q13.2	chr11:69098089-69278404	3	<i>CCND1</i>	<i>ORAOV1</i>
17q12	chr17:35067383-35272328	6	<i>ERBB2</i>	<i>ERBB2</i> , <i>C17orf37</i>
12q14.1	chr12:56419524-56488685	7	<i>CDK4</i>	<i>TSPAN31</i>
14q13.3	chr14:35708407-36097605	3	<i>NKX2-1</i>	<i>NKX2-1</i>
12q15	chr12:67440273-67566002	1	<i>MDM2</i>	<i>MDM2</i>
7p11.2	chr7:54899301-55275419	1	<i>EGFR</i>	<i>EGFR</i>
1q21.2	chr1:148661965-149063439	9	<i>MCL1</i>	<i>MCL1</i>
8p12	chr8:38252951-38460772	3	<i>FGFR1</i>	<i>FGFR1</i>
12p12.1	chr12:25189655-25352305	2	<i>KRAS</i>	<i>KRAS</i>
19q12	chr19:34975531-35098303	1	<i>CCNE1</i>	<i>CCNE1</i>

22q11.21	chr22:19172385-19746441	11	<i>CRKL</i>	<i>CRKL</i>
12q15	chr12:68249634-68327233	2		<i>LRRC10</i>
12q14.3	chr12:64461446-64607139	1	<i>HMGA2</i>	<i>HMGA2</i>
Xq28	chrX:152729030-154913754	53		<i>SPRY3</i>
5p15.33	chr5:1212750-1378766	3	<i>TERT</i>	<i>TERT</i>
3q26.2	chr3:170024984-173604597	22	<i>PRKCI</i>	<i>PRKCI</i>
15q26.3	chr15:96891354-97698742	4	<i>IGF1R</i>	<i>IGF1R</i>
20q13.2	chr20:51603033-51989829	1		<i>ZNF217</i>
8p11.21	chr8:42006632-42404492	6		<i>PLAT</i>
1p34.2	chr1:39907605-40263248	7	<i>MYCL1</i>	<i>MYCL1</i>
17q21.33	chr17:44673157-45060263	4		<i>NGFR, PHB</i>
2p24.3	chr2:15977811-16073001	1	<i>MYCN</i>	<i>MYCN</i>
7q21.3	chr7:89924533-98997268	62	<i>CDK6</i>	<i>CDK6</i>
13q34	chr13:108477140-110084607	4		<i>IRS2</i>
11q14.1	chr11:76699529-78005085	14		<i>GAB2</i>
20q13.33	chr20:61329497-62435964	38		<i>BIRC7</i>
17q23.1	chr17:55144989-55540417	5		<i>RPS6KB1</i>

1p12	chr1:119996566-120303234	5		<i>REG4</i>
8q21.13	chr8:81242335-81979194	3		<i>ZNF704</i> , <i>ZBTB10</i>
6p21.1	chr6:43556800-44361368	18		<i>VEGFA</i>
5p11	chr5:45312870-49697231	0		
20q11.21	chr20:29526118-29834552	5	<i>BCL2L1</i>	<i>BCL2L1</i> , <i>ID1</i>
6q23.3	chr6:135561194-135665525	1	<i>MYB</i>	<i>hsa-mir-548a-2</i>
1q44	chr1:241364021-247249719	71		<i>AKT3</i>
5q35.3	chr5:174477192-180857866	92		<i>FLT4</i>
7q31.2	chr7:115981465-116676953	3	<i>MET</i>	<i>MET</i>
18q11.2	chr18:17749667-22797232	21		<i>CABLES1</i>
17q25.1	chr17:70767943-71305641	13		<i>GRB2</i> , <i>ITGB4</i>
1p32.1	chr1:58658784-60221344	7	<i>JUN</i>	<i>JUN</i>
17q11.2	chr17:24112056-24310787	5		<i>DHRS13</i> , <i>FLOT2</i> , <i>ERAL1</i> , <i>PHF12</i>
17p11.2	chr17:18837023-19933105	12		<i>MAPK7</i>
8q24.11	chr8:116186189-120600761	13		<i>NOV</i>
12q15	chr12:66458200-66543552	0		

19q13.2	chr19:43177306-45393020	60		<i>LGALS7, DYRK1B</i>
11q22.2	chr11:101433436-102134907	8	<i>BIRC2, YAP1</i>	<i>BIRC2</i>
4q12	chr4:54471680-55980061	7	<i>PDGFRA, KIT</i>	<i>KDR, KIT</i>
12p11.21	chr12:30999223-32594050	9		<i>DDX11, FAM60A</i>
3q28	chr3:178149984-199501827	143	<i>PIK3CA</i>	<i>PIK3CA</i>
1p36.33	chr1:1-5160566	77		<i>TP73</i>
17q24.2	chr17:62318152-63890591	12		<i>BPTF</i>
1q23.3	chr1:158317017-159953843	52		<i>PEA15</i>
1q24.3	chr1:169549478-170484405	6		<i>BAT2D1, MYOC</i>
8q22.3	chr8:101163387-103693879	14		<i>RRM2B</i>
13q31.3	chr13:89500014-93206506	3		<i>GPC5</i>
12q21.1	chr12:70849987-70966467	0		
12p13.33	chr12:1-1311104	10		<i>WNK1</i>
12q21.2	chr12:76852527-77064746	0		
1q32.1	chr1:201678483-203358272	21	<i>MDM4</i>	<i>MDM4</i>
19q13.42	chr19:59066340-59471027	19		<i>PRKCG, TSEN34</i>
12q12	chr12:38788913-42596599	12		<i>ADAMTS20</i>
12q23.1	chr12:95089777-95350380	2		<i>ELK3</i>

12q21.32	chr12:85072329-85674601	0		
10q22.3	chr10:74560456-82020637	46		<i>SFTPA1B</i>
3p11.1	chr3:86250885-95164178	8		<i>POU1F1</i>
17q11.1	chr17:22479313-22877776	1		<i>WSB1</i>
8q24.3	chr8:140458177-146274826	97		<i>PTP4A3, MAFA, PARP10</i>
Xq12	chrX:66436234-67090514	1	<i>AR</i>	<i>AR</i>
6q12	chr6:63255006-65243766	3		<i>PTP4A1</i>
14q11.2	chr14:1-23145193	95		<i>BCL2L2</i>
9q34.3	chr9:137859478-140273252	76		<i>NRARP, MRPL41, TRAF2, LHX3</i>
6p24.1	chr6:1-23628840	95		<i>E2F3</i>
13q12.2	chr13:1-40829685	110		<i>FOXO1</i>
12q21.1	chr12:72596017-73080626	0		
14q32.33	chr14:106074644-106368585	0		
11p13	chr11:32027116-37799354	35		<i>WT1</i>

BẢNG 7. Các đoạn và gen nhiễm sắc thể minh họa nhưng không giới hạn sáng chế được biết hoặc dự báo là có trong các vùng có đặc điểm là khuếch đại ở nhiều ung thư khác nhau (xem ví dụ, Beroukhir et al. supra.).

Nhiễm sắc thể và băng nhiễm sắc thể	Vùng đỉnh	# các gen	Đích đã biết	Đích hàng đầu của GRAIL
9p21.3	chr9:21489625- 22474701	5	<i>CDKN2A/B</i>	<i>CDKN2A</i>
3p14.2	chr3:58626894- 61524607	2	<i>FHIT</i>	<i>FHIT</i>
16q23.1	chr16:76685816- 78205652	2	<i>WWOX</i>	<i>WWOX</i>
9p24.1	chr9:7161607- 12713130	3	<i>PTPRD</i>	<i>PTPRD</i>
20p12.1	chr20:14210829- 15988895	2	<i>MACROD2</i>	<i>FLRT3</i>
6q26	chr6:161612277- 163134099	1	<i>PARK2</i>	<i>PARK2</i>
13q14.2	chr13:46362859- 48209064	8	<i>RBI</i>	<i>RBI</i>
2q22.1	chr2:138479322- 143365272	3	<i>LRP1B</i>	<i>LRP1B</i>
4q35.2	chr4:186684565- 191273063	15		<i>FRG2, TUBB4Q</i>
5q11.2	chr5:57754754- 59053198	5	<i>PDE4D</i>	<i>PLK2, PDE4D</i>
16p13.3	chr16:5062786- 7709383	2	<i>A2BP1</i>	<i>A2BP1</i>
7q34	chr7:141592807- 142264966	3	<i>TRB</i>	<i>PRSSI</i>
2q37.3	chr2:241477619- 242951149	19		<i>TMEM16G, ING5</i>
19p13.3	chr19:1-526082	10		<i>GZMM, THEG, PPAP2C, C19orf20</i>
10q23.31	chr10:89467202- 90419015	4	<i>PTEN</i>	<i>PTEN</i>

8p23.2	chr8:2053441-6259545	1	<i>CSMD1</i>	<i>CSMD1</i>
1p36.31	chr1:3756302-6867390	23		<i>DFFB</i> , <i>ZBTB48</i> , <i>AJAP1</i>
4q22.1	chr4:91089383-93486891	2		<i>MGC48628</i>
18q23	chr18:75796373-76117153	4		<i>PARD6G</i>
6p25.3	chr6:1543157-2570302	2		<i>FOXC1</i>
19q13.43	chr19:63402921-63811651	17		<i>ZNF324</i>
Xp21.2	chrX:31041721-34564697	2	<i>DMD</i>	<i>DMD</i>
11q25	chr11:130280899-134452384	12	<i>OPCML</i> , <i>HNT</i>	<i>HNT</i>
13q12.11	chr13:1-23902184	29		<i>LATS2</i>
22q13.33	chr22:45488286-49691432	38		<i>TUBGCP6</i>
15q11.2	chr15:1-24740084	20		<i>A26B1</i>
22q11.22	chr22:20517661-21169423	3		<i>VPREB1</i>
10q26.3	chr10:129812260-135374737	35		<i>MGMT</i> , <i>SYCE1</i>
12p13.2	chr12:11410696-12118386	2	<i>ETV6</i>	<i>ETV6</i>
8p23.3	chr8:1-392555	2		<i>ZNF596</i>
1p36.11	chr1:26377344-27532551	24		<i>SFN</i>
11p15.5	chr11:1-1391954	49		<i>RASSF7</i>
17q11.2	chr17:26185485-27216066	10	<i>NF1</i>	<i>NF1</i>

11q23.1	chr11:107086196-116175885	61	<i>ATM</i>	<i>CADMI</i>
9p24.3	chr9:1-708871	5		<i>FOXD4</i>
10q11.23	chr10:52313829-53768264	4	<i>PRKG1</i>	<i>DKK1, PRKG1</i>
15q15.1	chr15:35140533-43473382	109		<i>TUBGCP4</i>
1p13.2	chr1:110339388-119426489	81		<i>MAGI3</i>
Xp22.33	chrX:1-3243111	21		<i>SHOX</i>
3p26.3	chr3:1-2121282	2		<i>CHL1</i>
9p13.2	chr9:36365710-37139941	2	<i>PAX5</i>	<i>MELK</i>
17p13.1	chr17:7471230-7717938	10	<i>TP53</i>	<i>ATP1B2</i>
12q24.33	chr12:131913408-132349534	7		<i>CHFR</i>
7q36.3	chr7:156893473-158821424	7	<i>PTPRN2</i>	<i>NCAPG2</i>
6q16.1	chr6:76630464-105342994	76		<i>FUT9, C6orf165, C6orf162, GJA10</i>
5q21.1	chr5:85837489-133480433	142	<i>APC</i>	<i>APC</i>
8p11.22	chr8:39008109-41238710	7		<i>C8orf4, ZMAT4</i>
19q13.32	chr19:52031294-53331283	25		<i>BBC3</i>
10p15.3	chr10:1-1042949	4		<i>TUBB8</i>
1p31.1	chr1:71284749-74440273	4	<i>NEGR1</i>	<i>NEGR1</i>
13q31.3	chr13:92308911-94031607	2	<i>GPC6</i>	<i>GPC6, DCT</i>

16q11.2	chr16:31854743-53525739	37		<i>RBL2</i>
20p13	chr20:1-325978	10		<i>SOX12</i>
5q35.3	chr5:177541057-180857866	43		<i>SCGB3A1</i>
1q43	chr1:223876038-247249719	173	<i>RYR2</i>	<i>FH, ZNF678</i>
16p13.3	chr16:1-359092	16		<i>HBZ</i>
17q21.2	chr17:37319013-37988602	22		<i>CNP</i>
2p25.3	chr2:1-15244284	51		<i>MYT1L</i>
3q13.31	chr3:116900556-120107320	1		<i>LSAMP</i>
7q21.11	chr7:65877239-79629882	73	<i>MAGI2</i>	<i>CLDN4</i>
7q35	chr7:144118814-148066271	3	<i>CNTNAP2</i>	<i>CNTNAP2</i>
14q32.12	chr14:80741860-106368585	154		<i>PRIMA1</i>
16q24.3	chr16:88436931-88827254	9		<i>C16orf3</i>
3q26.31	chr3:175446835-178263192	1	<i>NAALADL2</i>	<i>NAALADL2</i>
17q25.3	chr17:78087533-78774742	8		<i>ZNF750</i>
19p12	chr19:21788507-34401877	12		<i>ZNF492, ZNF99</i>
12q23.1	chr12:97551177-99047626	3	<i>ANKS1B</i>	<i>ANKS1B</i>
4p16.3	chr4:1-435793	4		<i>ZNF141</i>
18p11.32	chr18:1-587750	4		<i>COLEC12</i>
2q33.2	chr2:204533830-206266883	1	<i>PARD3B</i>	<i>PARD3B</i>

8p21.2	chr8:22125332-30139123	63		<i>DPYSL2, STMN4</i>
8q11.22	chr8:42971602-72924037	86	<i>SNTG1</i>	<i>FLJ23356, ST18, RB1CC1</i>
16q23.3	chr16:80759878-82408573	2	<i>CDH13</i>	<i>CDH13</i>
11q14.1	chr11:82612034-85091467	6	<i>DLG2</i>	<i>CCDC89, CCDC90B, TMEM126A</i>
14q23.3	chr14:65275722-67085224	7		<i>GPHN, MPP5</i>
7p22.2	chr7:3046420-4279470	1	<i>SDK1</i>	<i>SDK1</i>
13q34	chr13:111767404-114142980	25		<i>TUBGCP3</i>
17p12	chr17:10675416-12635879	5	<i>MAP2K4</i>	<i>MAP2K4, ZNF18</i>
21q22.2	chr21:38584860-42033506	19	<i>DSCAM, TMPRSS2/ERG</i>	<i>DSCAM</i>
18q21.2	chr18:46172638-49935241	7	<i>SMAD4, DCC</i>	<i>DCC</i>
6q22.1	chr6:101000242-121511318	87		<i>GTF3C6, TUBE1, ROS1</i>
14q11.2	chr14:1-29140968	140		<i>ZNF219, NDRG2</i>

Theo các phương án khác nhau, dự định sử dụng các phương pháp nêu ở đây để xác định CNV của các đoạn bao gồm các gen hoặc vùng đã khuếch đại được xác định trong Bảng 6 và/hoặc sử dụng các phương pháp nêu ở đây để xác định CNV của các đoạn bao gồm các gen hoặc vùng mất đoạn được xác định trong Bảng 7.

Theo một phương án, các phương pháp được mô tả ở đây cung cấp phương tiện đánh giá mối liên quan giữa khuếch đại gen và mức độ phát triển khối u. Tương quan giữa sự

Hiện tại, và liên quan đến các gen được liệt kê ở đây, không có bằng chứng về mối liên

khuếch đại và/hoặc mất đoạn và giai đoạn hoặc cấp độ của bệnh ung thư có thể có ý nghĩa tiên lượng quan trọng bởi vì thông tin này có thể góp phần vào việc xác định cấp độ khối u dựa vào di truyền sẽ dự đoán tốt hơn diễn biến trong tương lai của bệnh với các khối u tiến triển mạnh hơn có tiên lượng xấu nhất. Ngoài ra, thông tin về các sự kiện khuếch đại và/hoặc mất đoạn gen sớm có thể hữu ích trong việc liên hệ các sự kiện này là yếu tố dự đoán cho tiến triển bệnh sau này.

Khuếch đại và mất đoạn gen như được xác định bởi phương pháp này có thể được kết hợp với các tham số đã biết khác như cấp độ khối u, mô học, chỉ số gắn nhãn Brd/Urd, tình trạng hormon, sự liên quan về nút, kích thước khối u, thời gian sống và các đặc điểm khối u khác có sẵn từ các nghiên cứu dịch tễ học và thống kê sinh học. Ví dụ, ADN khối u cần thử nghiệm bằng phương pháp này có thể bao gồm tăng sản không điển hình, ung thư biểu mô ống dẫn sữa tại chỗ, ung thư giai đoạn I-III và các hạch bạch huyết di căn để cho phép xác định mối liên quan giữa sự khuếch đại, mất đoạn và giai đoạn. Mối liên quan được tạo ra có thể giúp can thiệp trị liệu hiệu quả. Ví dụ, các vùng được khuếch đại ổn định có thể chứa một gen biểu hiện quá mức, sản phẩm của gen này có thể được tấn công trị liệu (ví dụ, thụ thể yếu tố tăng trưởng tyrosin kinaza, p185^{HER2}).

Theo các phương án khác nhau, các phương pháp được mô tả ở đây có thể được sử dụng để xác định các sự kiện khuếch đại và/hoặc mất đoạn liên quan đến hiện tượng kháng thuốc bằng cách xác định biến thể số lượng bản sao của trình tự axit nucleic từ ung thư nguyên phát đến biến thể số lượng bản sao của các tế bào đã di căn đến vị trí khác. Nếu khuếch đại và/hoặc mất đoạn gen là biểu hiện của tính không ổn định nhiễm sắc thể cho phép phát triển nhanh chóng hiện tượng kháng thuốc, mong đợi là có sự khuếch đại và/hoặc mất đoạn trong các khối u nguyên phát ở người bệnh kháng thuốc hóa trị nhiều hơn là trong các khối u ở người bệnh nhạy cảm với hóa trị. Ví dụ, sự khuếch đại của các gen cụ thể chịu trách nhiệm về sự phát triển của kháng thuốc, vùng quanh các gen này được mong đợi là sẽ được khuếch đại ổn định trong các tế bào khối u từ sự tràn dịch màng phổi ở người bệnh kháng thuốc hóa trị nhưng không có trong các khối u nguyên phát. Việc phát hiện ra mối liên quan giữa khuếch đại và/hoặc mất đoạn gen và sự phát triển của hiện tượng kháng thuốc có thể cho phép xác định người bệnh sẽ hoặc sẽ không hưởng lợi từ liệu pháp bổ trợ.

Theo cách tương tự với cách được mô tả để xác định sự có mặt hay không có mặt của đột biến lệch bội nhiễm sắc thể bào thai toàn bộ và/hoặc một phần trong mẫu người mẹ, các phương pháp, thiết bị, và hệ thống được mô tả ở đây có thể được sử dụng để xác định sự có mặt hay không có mặt của đột biến lệch bội nhiễm sắc thể toàn bộ và/hoặc một phần trong mẫu người bệnh bất kỳ chứa các axit nucleic, ví dụ ADN hoặc cfADN (bao gồm các mẫu người bệnh không phải mẫu người mẹ). Mẫu người bệnh có thể là loại mẫu sinh học được mô tả ở phần khác trong bản mô tả này. Tốt hơn là, mẫu này thu được bằng các quy trình không xâm lấn. Ví dụ, mẫu này có thể là mẫu máu, hoặc phân đoạn huyết thanh và huyết tương của mẫu máu. Theo cách khác, mẫu này có thể là mẫu nước tiểu hoặc mẫu phân. Theo các phương án khác nữa, mẫu này là mẫu sinh thiết mô. Trong mọi trường hợp, mẫu đều chứa các axit nucleic ví dụ cfADN hoặc ADN gen, được lọc, và xác định trình tự bằng cách sử dụng phương pháp bất kỳ trong số các phương pháp xác định trình tự NGS mô tả trước đó.

Cả đột biến lệch bội nhiễm sắc thể toàn bộ và một phần liên quan đến sự hình thành, tiến triển bệnh ung thư có thể được xác định theo phương pháp này.

Theo các phương án khác nhau, khi sử dụng các phương pháp mô tả ở đây để xác định sự có mặt và/hoặc tăng nguy cơ ung thư, sự chuẩn hóa dữ liệu có thể được thực hiện đối với (các) nhiễm sắc thể mà CNV được xác định. Theo các phương án nhất định, sự chuẩn hóa dữ liệu có thể được thực hiện đối với (các) nhánh nhiễm sắc thể mà CNV được xác định. Theo các phương án nhất định, sự chuẩn hóa dữ liệu có thể được thực hiện đối với (các) đoạn cụ thể mà CNV được xác định.

Ngoài vai trò của CNV trong bệnh ung thư, các CNV có liên quan đến số lượng các bệnh phức tạp thông thường đang gia tăng, bao gồm virus gây suy giảm miễn dịch ở người (human immunodeficiency virus - HIV), các bệnh tự miễn và các phổ bệnh rối loạn tâm thần kinh.

CNV ở bệnh tự miễn và bệnh lây nhiễm

Đến nay, một số nghiên cứu đã báo cáo mối liên quan giữa CNV trong gen liên quan đến viêm và đáp ứng miễn dịch và HIV, bệnh hen suyễn, bệnh Crohn và các rối loạn tự miễn dịch khác (Fanciulli *et al.*, Clin Genet 77:201-213 [2010]). Ví dụ, CNV trong *CCL3L1*, được cho là liên quan đến tính nhạy cảm với HIV/AIDS (*CCL3L1*, mất đoạn 17q11.2), viêm khớp dạng thấp (*CCL3L1*, mất đoạn 17q11.2), và bệnh Kawasaki (*CCL3L1*, lặp đoạn 17q11.2);

CNV trong *HBD-2*, đã được báo cáo là dẫn đến bệnh Crohn đại tràng (*HBD-2*, mất đoạn 8p23.1) và bệnh vẩy nến (*HBD-2*, mất đoạn 8p23.1); CNV trong *FCGR3B*, cho thấy là dẫn đến bệnh viêm cầu thận trong bệnh lupus ban đỏ hệ thống (*FCGR3B*, mất đoạn 1q23, lặp đoạn 1q23), viêm huyết quản liên quan đến kháng thể bào tương kháng bạch cầu trung tính (anti-neutrophil cytoplasmic antibody - ANCA) (*FCGR3B*, mất đoạn 1q23), và làm tăng nguy cơ phát triển viêm khớp dạng thấp. Có ít nhất hai bệnh tự miễn hoặc bệnh viêm đã cho thấy là có liên quan đến CNV ở các locus gen khác nhau. Ví dụ, bệnh Crohn liên quan đến số lượng bản sao thấp ở *HBD-2*, mà còn liên quan đến hiện tượng đa hình mất đoạn thông thường ở thượng nguồn của gen *IGRM* mã hóa thành viên trong họ *GTPaza* liên quan đến tính miễn dịch p47. Ngoài sự liên quan đến số lượng bản sao *FCGR3B*, tính nhạy với SLE cũng đã được báo cáo là tăng lên đáng kể ở những đối tượng có số lượng bản sao của thành phần bổ sung C4 thấp hơn.

Sự liên quan giữa mất đoạn gen ở các locus *GSTM1* (*GSTM1*, mất đoạn 1q23) và *GSTT1* (*GSTT1*, mất đoạn 22q11.2) và tăng nguy cơ bị hen dị ứng đã được báo cáo trong một số nghiên cứu độc lập. Theo một số phương án, các phương pháp được mô tả ở đây có thể được sử dụng để xác định sự có mặt hoặc không có mặt của CNV liên quan đến các bệnh viêm và/hoặc bệnh tự miễn. Ví dụ, các phương pháp này có thể được sử dụng để xác định sự có mặt của CNV ở người bệnh nghi ngờ nhiễm HIV, bệnh hen suyễn, hoặc bệnh Crohn. Các ví dụ về CNV liên quan đến các bệnh như vậy bao gồm nhưng không giới hạn mất đoạn ở 17q11.2, 8p23.1, 1q23, và 22q11.2, và lặp đoạn tại 17q11.2, và 1q23. Theo một số phương án, phương pháp này có thể được sử dụng để xác định sự có mặt của CNV trong các gen bao gồm nhưng không giới hạn ở *CCL3L1*, *HBD-2*, *FCGR3B*, *GSTM*, *GSTT1*, *C4*, và *IRGM*.

Các bệnh CNV của hệ thần kinh

Sự liên quan giữa CNV mới và CNV kế thừa với một số bệnh tâm thần và thần kinh thông thường đã được báo cáo ở bệnh tự kỷ, tâm thần phân liệt và động kinh, và một số trường hợp của bệnh thoái hóa thần kinh như bệnh Parkinson, bệnh xơ cứng cột bên teo cơ (amyotrophic lateral sclerosis - ALS) và bệnh Alzheimer trội trên nhiễm sắc thể thường (Fanciulli *et al.*, Clin Genet 77:201-213 [2010]). Bất thường trong di truyền tế bào đã được quan sát thấy ở các người bệnh bị tự kỷ và rối loạn phổ tự kỷ (autism spectrum disorder - ASD) có lặp đoạn ở 15q11-q13. Theo Công đoàn Dự án Gen Tự kỷ, 154 CNV bao gồm một

số CNV lặp, trên nhiễm sắc thể 15q11-q13 hoặc tại các vị trí gen mới chứa nhiễm sắc thể 2p16, 1q21 và tại 17p12 trong vùng liên quan đến hội chứng Smith-Magenis chồng lấn với ASD. Vi mất đoạn hoặc vi lặp đoạn hồi quy trên nhiễm sắc thể 16p11.2 đã nhấn mạnh quan sát rằng các CNV mới được phát hiện ở các locut với các gen như *SHANK3* (mất đoạn 22q13.3), neurexin 1 (*NRXN1*, mất đoạn 2p16.3) và các tế bào thần kinh đệm (*NLGN4*, mất đoạn Xp22.33) được biết đến là dùng để điều hòa sự biệt hóa tiếp hợp và điều hòa sự giải phóng chất truyền dẫn thần kinh hệ kích thích. Bệnh tâm thần phân liệt cũng được cho là có liên quan đến nhiều CNV mới. Vi mất đoạn và vi lặp đoạn liên quan đến bệnh tâm thần phân liệt chứa quá nhiều gen thuộc quá trình phát triển thần kinh và kích thích, đề xuất rằng nhiều CNV ảnh hưởng đến các gen này có thể góp phần trực tiếp vào sự phát sinh của bệnh tâm thần phân liệt ví dụ *ERBB4*, mất đoạn 2q34, *SLC1A3*, mất đoạn 5p13.3; *RAPGEF4*, mất đoạn 2q31.1; *CIT*, mất đoạn 12.24; và nhiều gen với CNV mới. CNV cũng được cho là liên quan đến các rối loạn thần kinh khác bao gồm bệnh động kinh (*CHRNA7*, mất đoạn 15q13.3), bệnh Parkinson (*SNCA* lặp đoạn 4q22) và ALS (*SMN1*, mất đoạn 5q12.2.-q13.3; và mất đoạn *SMN2*). Theo một số phương án, các phương pháp được mô tả ở đây có thể được sử dụng để xác định sự có mặt hoặc không có mặt của CNV liên quan đến các bệnh về hệ thần kinh. Ví dụ, các phương pháp này có thể được sử dụng để xác định sự có mặt của CNV ở người bệnh nghi ngờ mắc bệnh tự kỷ, bệnh tâm thần phân liệt, bệnh động kinh, các bệnh thoái hóa thần kinh như bệnh Parkinson, bệnh xơ cứng cột bên teo cơ (ALS) hoặc bệnh Alzheimer trội trên nhiễm sắc thể thường. Các phương pháp này có thể được sử dụng để xác định CNV của các gen liên quan đến các bệnh về hệ thần kinh bao gồm nhưng không giới hạn ở bệnh bất kỳ trong số các bệnh: rối loạn phổ tự kỷ (ASD), bệnh tâm thần phân liệt, và bệnh động kinh, và CNV của các gen liên quan đến các chứng rối loạn thoái hóa thần kinh như bệnh Parkinson. Ví dụ về CNV liên quan đến các bệnh này bao gồm nhưng không giới hạn lặp đoạn tại 15q11-q13, 2p16, 1q21, 17p12, 16p11.2, và 4q22, và mất đoạn tại 22q13.3, 2p16.3, Xp22.33, 2q34, 5p13.3, 2q31.1, 12.24, 15q13.3, và 5q12.2. Theo một số phương án, các phương pháp này có thể được sử dụng để xác định sự có mặt của CNV trong các gen bao gồm nhưng không giới hạn ở *SHANK3*, *NLGN4*, *NRXN1*, *ERBB4*, *SLC1A3*, *RAPGEF4*, *CIT*, *CHRNA7*, *SNCA*, *SMN1*, và *SMN2*.

CNV và các bệnh chuyển hóa hoặc tim mạch

Mối liên quan giữa thể bệnh chuyển hóa và tim mạch, như bệnh tăng cholesterol máu di truyền (familial hypercholesterolemia - FH), xơ vữa động mạch và bệnh động mạch vành, và CNV đã được báo cáo trong một số nghiên cứu (Fanciulli *et al.*, Clin Genet 77:201-213 [2010]). Ví dụ, sự tái sắp xếp dòng phôi, chủ yếu là mất đoạn, đã được quan sát thấy trong gen *LDLR* (*LDLR*, mất đoạn/ lặp đoạn 19p13.2) ở một số người bệnh FH không mang các đột biến *LDLR* khác. Một ví dụ khác là gen *LPA* mã hóa apolipoprotein(a) (apo(a)) mà nồng độ huyết tương của protein này liên quan đến nguy cơ mắc bệnh động mạch vành, nhồi máu cơ tim (myocardial infarction - MI) và đột quy. Nồng độ huyết tương của apo(a) chứa lipoprotein Lp(a) khác nhau trên 1000 lần giữa các cá thể và 90% sự biến thiên này được xác định theo di truyền ở *LPA*, với nồng độ huyết tương và kích thước isoform Lp(a) tỷ lệ với số lượng biến thiên cao của trình tự lặp 'kringle 4' (trong khoảng 5-50). Dữ liệu này cho thấy rằng CNV trong ít nhất hai gen có thể có liên quan đến nguy cơ tim mạch. Các phương pháp được mô tả ở đây có thể được sử dụng trong những nghiên cứu lớn để tìm kiếm cụ thể mối liên quan của CNV với các rối loạn tim mạch. Theo một số phương án, phương pháp này có thể được sử dụng để xác định sự có mặt hoặc không có mặt của CNV liên quan đến bệnh chuyển hóa hoặc tim mạch. Ví dụ, phương pháp này có thể được sử dụng để xác định sự có mặt của CNV ở người bệnh nghi ngờ bị bệnh tăng cholesterol máu di truyền. Các phương pháp được mô tả ở đây có thể được sử dụng để xác định CNV của các gen liên quan đến bệnh chuyển hóa hoặc tim mạch, ví dụ bệnh tăng cholesterol máu. Ví dụ về CNV liên quan đến các bệnh này bao gồm nhưng không giới hạn mất đoạn/ lặp đoạn 19p13.2 của gen *LDLR*, và sự nhân đoạn ở gen *LPA*.

Thiết bị và hệ thống xác định CNV

Phân tích dữ liệu giải trình tự và chẩn đoán suy ra từ đó thường được thực hiện bằng cách sử dụng các thuật toán và chương trình thực thi bằng máy tính khác nhau. Do đó, các phương án nhất định sử dụng các quy trình liên quan đến dữ liệu được lưu trữ trong hoặc được truyền qua một hoặc nhiều hệ thống máy tính hoặc các hệ thống xử lý khác. Các phương án được bộc lộ ở đây cũng liên quan đến thiết bị thực hiện các hoạt động này. Thiết bị này có thể được thiết lập đặc biệt cho mục đích yêu cầu, hoặc có thể là máy tính (hoặc nhóm máy tính) đa dụng được kích hoạt chọn lọc hoặc được cấu hình lại bởi chương trình máy tính và/hoặc

cấu trúc dữ liệu lưu trữ trong máy tính. Theo một số phương án, nhóm các bộ xử lý thực hiện một vài hoặc tất cả các hoạt động phân tích đã nêu theo cách hợp tác (ví dụ, qua mạng hoặc điện toán đám mây) và/hoặc song song. Bộ xử lý hoặc nhóm các bộ xử lý để thực hiện các phương pháp được mô tả ở đây có thể thuộc nhiều loại khác nhau bao gồm bộ vi điều khiển hoặc bộ vi xử lý như các thiết bị lập trình được (ví dụ, CPLD và FPGA) và các thiết bị không lập trình được như ASIC mảng công hoặc các bộ vi xử lý đa dụng.

Ngoài ra, các phương án nhất định đề cập đến sản phẩm chương trình máy tính hoặc phương tiện đọc được bằng máy tính hữu hình và/hoặc bất biến bao gồm các lệnh chương trình và/hoặc dữ liệu (bao gồm các cấu trúc dữ liệu) để thực hiện các thao tác thực thi bởi máy tính khác nhau. Ví dụ về vật ghi đọc được bằng máy tính bao gồm nhưng không giới hạn ở các thiết bị nhớ bán dẫn, phương tiện từ tính như ổ đĩa, băng từ, phương tiện quang như đĩa CD, phương tiện từ - quang, và các thiết bị phần cứng được tạo cấu hình đặc biệt để lưu trữ và thực hiện các lệnh chương trình, như bộ nhớ chỉ đọc (read-only memory - ROM) và bộ nhớ truy cập ngẫu nhiên (random access memory - RAM). Phương tiện đọc được bằng máy tính có thể được điều khiển trực tiếp bởi người dùng cuối hoặc phương tiện có thể được điều khiển trực tiếp bởi người dùng cuối. Ví dụ về phương tiện được điều khiển trực tiếp bao gồm phương tiện đặt tại cơ sở người dùng và/hoặc phương tiện không dùng chung với các thực thể khác. Ví dụ về phương tiện được điều khiển gián tiếp bao gồm phương tiện truy cập được gián tiếp với người dùng qua mạng bên ngoài và/hoặc qua dịch vụ cung cấp tài nguyên dùng chung như "đám mây". Ví dụ về các lệnh chương trình bao gồm cả mã máy, như mã máy được tạo ra bởi bộ biên dịch, và các tập tin chứa mã mức cao hơn có thể được thực thi bởi máy tính sử dụng bộ diễn dịch.

Theo các phương án khác nhau, dữ liệu hoặc thông tin được sử dụng trong các phương pháp và thiết bị được cung cấp ở định dạng điện tử. Dữ liệu hoặc thông tin này có thể bao gồm các trình tự đọc và thể có nguồn gốc từ mẫu axit nucleic, số lượng hoặc mật độ của các thể này mà căn chỉnh với các vùng cụ thể của trình tự tham chiếu (ví dụ, so sánh với nhiễm sắc thể hoặc đoạn nhiễm sắc thể), các trình tự tham chiếu (bao gồm trình tự tham chiếu cung cấp các dạng đa hình duy nhất hoặc chủ yếu), định lượng nhiễm sắc thể và đoạn nhiễm sắc thể, kết quả xác định như kết quả xác định đột biến lệch bội, các giá trị nhiễm sắc thể và đoạn nhiễm sắc thể chuẩn hóa, các cặp nhiễm sắc thể và đoạn nhiễm sắc thể và các nhiễm sắc thể

và đoạn nhiễm sắc thể chuẩn hóa tương ứng, khuyến nghị tư vấn, chẩn đoán, và thông tin tương tự. Như được sử dụng ở đây, dữ liệu hoặc thông tin khác được cung cấp ở định dạng điện tử có sẵn để lưu trữ trên máy hoặc để truyền giữa các máy. Thông thường, dữ liệu ở định dạng điện tử được cung cấp dạng số hóa và có thể được lưu trữ dưới dạng các bit và/hoặc byte trong các cấu trúc dữ liệu, danh mục, cơ sở dữ liệu, v.v. khác nhau. Dữ liệu có thể được biểu diễn ở dạng điện tử, quang học, v.v..

Một phương án đề xuất sản phẩm chương trình máy tính để tạo ra kết quả cho biết sự có mặt hoặc không có mặt của đột biến lệch bội, ví dụ, đột biến lệch bội ở bào thai hoặc ung thư, trong mẫu thử. Sản phẩm chương trình máy tính có thể chứa các lệnh để thực hiện một hoặc nhiều phương pháp bất kỳ trong số các phương pháp mô tả trên đây để xác định sự bất thường nhiễm sắc thể. Như đã giải thích, sản phẩm chương trình máy tính có thể bao gồm phương tiện đọc được bằng máy tính bất biến và/hoặc hữu hình có logic (ví dụ, các lệnh) thực thi được hoặc biên dịch được bằng máy tính được ghi trên đó để khiến cho bộ xử lý xác định được định lượng nhiễm sắc thể và, trong một số trường hợp, xác định có hay không có đột biến lệch bội ở bào thai. Trong một ví dụ, sản phẩm chương trình máy tính bao gồm phương tiện đọc được bằng máy tính có logic (ví dụ, các lệnh) thực thi được hoặc biên dịch được bằng máy tính được ghi trên đó để khiến cho bộ xử lý chẩn đoán đột biến lệch bội ở bào thai bao gồm: quy trình nhận để nhận dữ liệu trình tự từ ít nhất một phần của phân tử axit nucleic từ mẫu sinh học của người mẹ, trong đó dữ liệu trình tự bao gồm định lượng nhiễm sắc thể và/hoặc đoạn nhiễm sắc thể; logic hỗ trợ bởi máy tính để phân tích đột biến lệch bội ở bào thai từ dữ liệu nhận được nói trên; và quy trình xuất ra để tạo ra kết quả biểu thị sự có mặt, không có mặt hoặc loại đột biến lệch bội ở bào thai nói trên.

Thông tin trình tự từ mẫu đang xem xét có thể được ánh xạ đến các trình tự tham chiếu nhiễm sắc thể để xác định số thể trình tự cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ và để xác định số thể trình tự cho trình tự đoạn chuẩn hóa cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ nói trên. Theo các phương án khác nhau, các trình tự tham chiếu được lưu trữ trong cơ sở dữ liệu như cơ sở dữ liệu quan hệ hoặc đối tượng chẳng hạn.

Cần hiểu rằng sẽ là không thực tế, hoặc thậm chí là không thể trong hầu hết các trường hợp, khi để cho người không được trợ giúp thực hiện các thao tác tính toán theo các phương

pháp được bộc lộ ở đây. Ví dụ, việc ánh xạ một trình tự đọc 30 bp từ mẫu đến một nhiễm sắc thể bất kỳ trong số các nhiễm sắc thể người có thể đòi hỏi nhiều năm nỗ lực mà không có sự trợ giúp của thiết bị tính toán. Tất nhiên, vấn đề này là phức tạp bởi vì việc gọi ra đột biến lệch bội đáng tin cậy thường đòi hỏi ánh xạ hàng nghìn (ví dụ, ít nhất khoảng 10.000) hoặc thậm chí hàng triệu trình tự đọc đến một hoặc nhiều nhiễm sắc thể.

Các phương pháp được bộc lộ ở đây có thể được thực hiện bằng cách sử dụng hệ thống để ước lượng số lượng bản sao của trình tự gen quan tâm trong mẫu thử. Hệ thống bao gồm: (a) bộ giải trình tự để nhận các axit nucleic từ mẫu thử nghiệm cung cấp thông tin trình tự của axit nucleic từ mẫu; (b) bộ xử lý; và (c) một hoặc nhiều vật ghi đọc được bằng máy tính lưu trữ trên đó các lệnh để thực thi trên bộ xử lý để thực hiện phương pháp xác định CNV bất kỳ, ví dụ, các đột biến lệch bội nhiễm sắc thể hoặc đột biến lệch bội một phần.

Theo một số phương án, các phương pháp được lệnh bởi phương tiện đọc được bằng máy tính lưu trữ trên đó các lệnh đọc được bằng máy tính để thực hiện phương pháp xác định CNV bất kỳ, ví dụ, các đột biến lệch bội nhiễm sắc thể hoặc đột biến lệch bội một phần. Do đó một phương án đề xuất sản phẩm chương trình máy tính bao gồm một hoặc nhiều vật ghi bất biến đọc được bằng máy tính lưu trữ trên đó các lệnh thực thi được bằng máy tính mà, khi được thực thi bởi một hoặc nhiều bộ xử lý của hệ thống máy tính, khiến cho hệ thống máy tính thực hiện phương pháp ước lượng số lượng bản sao của trình tự quan tâm trong mẫu thử bao gồm các axit nucleic ngoài tế bào của người mẹ và bào thai. Phương pháp bao gồm các bước: (a) nhận các trình tự đọc thu được bằng cách giải trình tự các đoạn axit nucleic ngoài tế bào trong mẫu thử; (b) so sánh các trình tự đọc của các đoạn axit nucleic ngoài tế bào với bộ gen tham chiếu bao gồm trình tự quan tâm, do đó cung cấp các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin; (c) xác định kích thước của các đoạn axit nucleic ngoài tế bào tồn tại trong mẫu thử; (d) gán trọng số cho các thẻ trình tự thử nghiệm dựa vào kích thước của các đoạn axit nucleic ngoài tế bào mà từ đó các thẻ này thu được; (e) tính toán độ che phủ đối với các bin dựa vào các thẻ được gán trọng số ở bước (d); và (f) nhận diện biến thể số lượng bản sao ở trình tự quan tâm từ độ che phủ đã tính toán. Trong một số phương án thực hiện, việc gán trọng số các thẻ trình tự thử nghiệm bao gồm dịch chuyển độ che phủ về phía các thẻ trình tự thử nghiệm thu được từ các đoạn axit nucleic ngoài tế bào có đặc trưng kích thước hoặc khoảng kích thước của một bộ gen trong mẫu thử. Trong một số

phương án thực hiện, việc gán trọng số cho các thẻ trình tự thử nghiệm bao gồm gán giá trị 1 cho các thẻ thu được từ các đoạn axit nucleic ngoài tế bào có kích thước hoặc khoảng kích thước này, và gán giá trị 0 cho các thẻ khác. Trong một số phương án thực hiện, phương pháp này còn bao gồm bước xác định, trong các bin của bộ gen tham chiếu, bao gồm trình tự quan tâm, các giá trị của tham số kích thước đoạn bao gồm số lượng các đoạn axit nucleic ngoài tế bào trong mẫu thử có các kích thước đoạn ngắn hơn hoặc dài hơn giá trị ngưỡng. Ở đây, việc xác định sự biến thể số lượng bản sao ở trình tự quan tâm bao gồm việc sử dụng các giá trị của tham số kích thước đoạn cũng như độ che phủ đã tính trong (e). Trong một số phương án thực hiện, hệ thống này được tạo cấu hình để ước lượng số lượng bản sao trong mẫu thử bằng cách sử dụng các phương pháp và quy trình nêu trên.

Theo một số phương án, các lệnh còn có thể bao gồm thông tin ghi tự động thích hợp với phương pháp như định lượng nhiễm sắc thể và sự có mặt hoặc không có mặt của đột biến lệch bội nhiễm sắc thể ở bào thai trong hồ sơ bệnh án cho đối tượng người cung cấp mẫu thử của người mẹ. Hồ sơ bệnh án có thể được bảo quản bởi, ví dụ, thư viện, phòng khám, bệnh viện, tổ chức duy trì sức khỏe, công ty bảo hiểm, hoặc trang web hồ sơ bệnh án cá nhân. Hơn nữa, dựa vào các kết quả của phân tích được thực hiện bởi bộ xử lý, phương pháp này có thể còn bao gồm bước kê đơn, bắt đầu, và/hoặc thay đổi quy trình điều trị đối tượng người mà mẫu thử của người mẹ được lấy từ người đó. Điều này có thể bao gồm thực hiện một hoặc nhiều thử nghiệm hoặc phân tích bổ sung trên các mẫu bổ sung được lấy từ đối tượng.

Các phương pháp được bộc lộ cũng có thể được thực hiện bằng cách sử dụng hệ thống xử lý máy tính được làm thích ứng hoặc được tạo cấu hình để thực hiện phương pháp xác định CNV bất kỳ, ví dụ, các đột biến lệch bội nhiễm sắc thể hoặc đột biến lệch bội một phần. Một phương án cung cấp hệ thống xử lý máy tính được làm thích ứng hoặc được tạo cấu hình để thực hiện phương pháp như được mô tả ở đây. Theo một phương án, thiết bị này bao gồm thiết bị giải trình tự được làm thích ứng hoặc được tạo cấu hình để giải trình tự ít nhất một phần của phân tử axit nucleic trong mẫu để thu được loại thông tin trình tự được mô tả ở phần khác trong bản mô tả này. Thiết bị này cũng có thể bao gồm các thành phần để xử lý mẫu. Các thành phần này được mô tả ở phần khác trong bản mô tả.

Trình tự hoặc dữ liệu khác, có thể được nhập vào máy tính hoặc được lưu trữ trên vật ghi đọc được bằng máy tính trực tiếp hoặc gián tiếp. Theo một phương án, hệ thống máy tính được nối trực tiếp với thiết bị giải trình tự để đọc và/hoặc phân tích các trình tự axit nucleic từ các mẫu. Trình tự hoặc thông tin khác từ công cụ này được cung cấp qua giao diện trong hệ thống máy tính. Theo cách khác, các trình tự xử lý bởi hệ thống được cung cấp từ nguồn lưu trữ trình tự như cơ sở dữ liệu kho lưu trữ khác. Khi khả dụng với thiết bị xử lý, thiết bị nhớ hoặc thiết bị lưu trữ khối đệm hoặc lưu trữ, ít nhất là tạm thời, các chuỗi axit nucleic. Ngoài ra, thiết bị nhớ có thể lưu trữ số đếm thẻ cho các nhiễm sắc thể hoặc bộ gen, v.v. khác nhau. Bộ nhớ cũng có thể lưu trữ các thủ tục và/hoặc chương trình khác nhau để phân tích việc trình diễn trình tự hoặc dữ liệu được ánh xạ. Các chương trình/ thủ tục này có thể bao gồm các chương trình để thực hiện phân tích thống kê, v.v..

Trong một ví dụ, người dùng cung cấp mẫu vào thiết bị giải trình tự. Dữ liệu được thu và/hoặc được phân tích bởi thiết bị giải trình tự kết nối với máy tính. Phần mềm trên máy tính cho phép thu thập và/hoặc phân tích dữ liệu. Dữ liệu có thể được lưu trữ, hiển thị (qua màn hình hoặc thiết bị tương tự khác), và/hoặc truyền đến vị trí khác. Máy tính có thể được kết nối internet dùng để truyền dữ liệu đến thiết bị cầm tay sử dụng bởi người dùng từ xa (ví dụ, bác sĩ, nhà khoa học hoặc nhà phân tích). Cần hiểu rằng dữ liệu có thể được lưu trữ và/hoặc được phân tích trước khi truyền. Theo một số phương án, dữ liệu thô được thu thập và được truyền đến người dùng hoặc thiết bị từ xa để phân tích và/hoặc lưu trữ dữ liệu. Cuộc truyền có thể diễn ra qua mạng internet, nhưng cũng có thể diễn ra qua kết nối vệ tinh hoặc kết nối khác. Theo cách khác, dữ liệu có thể được lưu trữ trên phương tiện đọc được bằng máy tính và phương tiện này có thể được chuyển đến người dùng cuối (ví dụ, qua thư). Người dùng từ xa có thể ở cùng hoặc ở khác vị trí địa lý, bao gồm nhưng không giới hạn ở tòa nhà, thành phố, bang, quốc gia hoặc lục địa.

Theo một số phương án, các phương pháp này cũng bao gồm thu thập dữ liệu liên quan đến các trình tự polynucleotit (ví dụ, các trình tự đọc, các thẻ và/hoặc các trình tự nhiễm sắc thể tham chiếu) và truyền dữ liệu đến máy tính hoặc hệ thống tính toán khác. Ví dụ, máy tính có thể được nối với thiết bị phòng thí nghiệm, ví dụ, thiết bị thu thập mẫu, thiết bị khuếch đại nucleotit, thiết bị giải trình tự nucleotit, hoặc thiết bị lai hóa. Sau đó máy tính có thể thu thập dữ liệu phù hợp thu được bởi thiết bị phòng thí nghiệm. Dữ liệu có thể được lưu trữ trên máy

tính ở bước bất kỳ, ví dụ, trong khi được thu thập theo thời gian thực, trước khi truyền, trong khi hoặc kết hợp với bước truyền, hoặc sau khi truyền. Dữ liệu có thể được lưu trữ trên phương tiện đọc được bằng máy tính mà có thể được trích xuất ra từ máy tính. Dữ liệu được thu thập hoặc lưu trữ có thể được truyền từ máy tính đến vị trí ở xa, ví dụ, qua mạng cục bộ hoặc mạng diện rộng như internet. Tại vị trí ở xa, các thao tác khác nhau có thể được thực hiện trên dữ liệu đã truyền như được mô tả dưới đây.

Các loại dữ liệu định dạng điện tử có thể được lưu trữ, truyền, phân tích, và/hoặc sử dụng trong các hệ thống, thiết bị, và các phương pháp được bộc lộ ở đây bao gồm các loại dữ liệu sau:

Trình tự đọc thu được bởi việc giải trình từ các axit nucleic trong mẫu thử

Thẻ thu được bởi việc căn chỉnh trình tự đọc với bộ gen tham chiếu hoặc trình tự hoặc các trình tự tham chiếu khác

Bộ gen hoặc trình tự tham chiếu

Mật độ thẻ trình tự - Số đếm hoặc số lượng thẻ cho mỗi trong số hai hoặc nhiều vùng (thông thường là các nhiễm sắc thể hoặc đoạn nhiễm sắc thể) của bộ gen tham chiếu các trình tự tham chiếu khác

Thông tin nhận dạng của các nhiễm sắc thể hoặc đoạn nhiễm sắc thể chuẩn hóa cho các nhiễm sắc thể hoặc đoạn nhiễm sắc thể quan tâm cụ thể

Định lượng cho các nhiễm sắc thể hoặc đoạn nhiễm sắc thể (hoặc các vùng khác) thu được từ các nhiễm sắc thể hoặc đoạn nhiễm sắc thể quan tâm và các nhiễm sắc thể hoặc đoạn nhiễm sắc thể chuẩn hóa tương ứng

Ngưỡng để xác định định lượng nhiễm sắc thể là bị bệnh, không bị bệnh, hoặc không xác định

Các kết quả xác định thực tế của các định lượng nhiễm sắc thể

Chẩn đoán (tình trạng lâm sàng đi kèm với các kết quả xác định)

Khuyến nghị thử nghiệm thêm suy ra từ các kết quả xác định và/hoặc chẩn đoán

Kế hoạch điều trị và/hoặc theo dõi suy ra từ các kết quả xác định và/hoặc chẩn đoán

Các loại dữ liệu khác nhau có thể được thu, lưu trữ, truyền, phân tích, và/hoặc sử dụng ở một hoặc nhiều vị trí bằng cách sử dụng thiết bị riêng biệt. Các tùy chọn xử lý trải ra trên phổ rộng. Ở một đầu của phổ, tất cả hoặc phần nhiều thông tin này được lưu trữ và sử dụng tại vị trí mà mẫu thử được xử lý, ví dụ, phòng khám của bác sỹ hoặc cơ sở khám bệnh khác. Ở đầu cực còn lại, mẫu được thu tại một vị trí, mẫu được xử lý và được giải trình tự tùy ý tại một vị trí khác, trình tự đọc được so sánh và xác định đột biến được thực hiện ở một hoặc nhiều vị trí khác, và các chẩn đoán, khuyến nghị, và/hoặc kế hoạch được chuẩn bị tại vị trí khác nữa (đây có thể là vị trí mà thu được mẫu tại đó).

Theo các phương án khác nhau, trình tự đọc được tảo ra bằng thiết bị giải trình tự và sau đó được truyền đến vị trí ở xa mà chúng được xử lý để tạo ra xác định đột biến lệch bội. Tại vị trí ở xa này, ví dụ, trình tự đọc được so sánh với trình tự tham chiếu để tạo ra các thể, chúng được đếm và được gán cho các nhiễm sắc thể hoặc đoạn nhiễm sắc thể quan tâm. Cũng tại vị trí ở xa này, số đếm này được biến đổi thành định lượng bằng cách sử dụng các nhiễm sắc thể hoặc đoạn nhiễm sắc thể chuẩn hóa liên quan. Hơn nữa, tại vị trí ở xa, định lượng có thể được sử dụng để tạo ra xác định lệch bội.

Một số thao tác xử lý có thể được sử dụng tại các vị trí riêng biệt bao gồm như sau:

- Thu thập mẫu
- Xử lý mẫu sơ bộ để giải trình tự
- Giải trình tự
- Phân tích dữ liệu trình tự và suy ra các kết quả xác định đột biến lệch bội
- Chẩn đoán
- Báo cáo chẩn đoán và/hoặc kết quả xác định đột biến cho người bệnh hoặc nhân viên y tế
- Phát triển kế hoạch để điều trị, thử nghiệm và/hoặc theo dõi thêm
- Thực hiện kế hoạch
- Tư vấn

Một hoặc nhiều thao tác bất kỳ trong số các thao tác này được tự động hóa như được mô tả ở phần khác trong bản mô tả. Thông thường, việc giải trình tự và phân tích dữ liệu trình tự và suy ra kết quả xác định đột biến lệch bội sẽ được thực hiện bằng cách tính toán. Các thao tác khác có thể được thực hiện thủ công hoặc tự động.

Ví dụ về các vị trí mà tại đó việc thu thập mẫu có thể được thực hiện bao gồm phòng khám của bác sỹ, phòng khám, nhà người bệnh (tại đó công cụ hoặc bộ công cụ thu thập mẫu được cung cấp), và các phương tiện chăm sóc sức khỏe di động. Ví dụ về các vị trí mà tại đó việc xử lý mẫu trước khi giải trình tự có thể được thực hiện bao gồm phòng khám của bác sỹ, phòng khám, nhà người bệnh (tại đó thiết bị hoặc bộ thiết bị xử lý mẫu được cung cấp), các phương tiện chăm sóc sức khỏe di động, và các cơ sở của bên cung cấp phân tích đột biến lệch bội. Ví dụ về các vị trí mà tại đó việc giải trình tự có thể được thực hiện bao gồm phòng khám của bác sỹ, phòng khám, nhà người bệnh (tại đó thiết bị hoặc bộ thiết bị giải trình tự mẫu được cung cấp), các phương tiện chăm sóc sức khỏe di động, và các cơ sở của bên cung cấp phân tích đột biến lệch bội. Vị trí mà tại đó việc giải trình tự diễn ra có thể được cung cấp với kết nối mạng dành riêng để truyền dữ liệu trình tự (thông thường là trình tự đọc) ở định dạng điện tử. Kết nối này có thể là có dây hoặc không dây và có thể được tạo cấu hình để truyền dữ liệu đến vị trí mà tại đó dữ liệu có thể được xử lý và/hoặc tổng hợp lại trước khi truyền đến vị trí xử lý. Bộ tổng hợp dữ liệu có thể được duy trì bởi các tổ chức y tế như các Tổ chức duy trì sức khỏe (Health Maintenance Organization - HMO).

Các thao tác phân tích và/hoặc suy ra có thể được thực hiện tại vị trí bất kỳ trong số các vị trí nói trên hoặc theo cách khác tại vị trí ở xa khác nữa dành riêng cho việc tính toán và/hoặc dịch vụ phân tích dữ liệu trình tự axit nucleic. Các vị trí này bao gồm, ví dụ, các cụm như cụm máy chủ đa dụng, các cơ sở của doanh nghiệp dịch vụ phân tích đột biến lệch bội, và tương tự. Theo một số phương án, thiết bị tính toán dùng để thực hiện phân tích được thuê hoặc cho thuê. Tài nguyên tính toán có thể là một phần của việc thu thập truy cập được bằng internet của các bộ xử lý như các tài nguyên xử lý được gọi theo cách thông thường là đám mây. Trong một số trường hợp, việc tính toán được thực hiện bởi nhóm các bộ xử lý song song hoặc song song khối lượng lớn liên kết hoặc không liên kết với nhau. Việc xử lý có thể được thực hiện bằng cách sử dụng quy trình xử lý phân tán như điện toán cụm, điện toán lưới, và tương tự. Trong các phương án này, cụm hoặc lưới tài nguyên tính toán cùng nhau tạo thành một siêu máy tính ảo bao gồm nhiều bộ xử lý hoặc máy tính cùng đóng vai trò thực hiện phân tích và/hoặc suy ra được mô tả ở đây. Các công nghệ này cũng như các siêu máy tính thông thường khác có thể được sử dụng để xử lý dữ liệu trình tự như được mô tả ở đây. Mỗi trong số chúng là một dạng điện toán song song phụ thuộc vào các bộ xử lý hoặc máy tính. Trong trường hợp

điện toán lưới, các bộ xử lý này (thường là toàn bộ máy tính) được kết nối với mạng (cá nhân, công cộng, hoặc Internet) bởi giao thức mạng thông thường như Ethernet. Ngược lại, siêu máy tính có nhiều bộ xử lý kết nối bởi các buýt máy tính tốc độ cao cục bộ.

Theo các phương án nhất định, chẩn đoán (ví dụ, bào thai mắc hội chứng Downs hoặc người bệnh mắc một loại ung thư cụ thể) được tạo ra tại cùng vị trí với thao tác phân tích. Trong các phương án khác, chẩn đoán được thực hiện tại vị trí khác. Trong một số ví dụ, việc báo cáo chẩn đoán được thực hiện tại vị trí mà mẫu được lấy, mặc dù không cần thiết phải như vậy. Ví dụ về các vị trí mà tại đó chẩn đoán có thể được tạo ra hoặc báo cáo và/hoặc tại đó việc phát triển kế hoạch được thực hiện bao gồm phòng khám của bác sỹ, phòng khám, các vị trí trên internet truy cập được bằng máy tính, và các thiết bị cầm tay như điện thoại di động, máy tính bảng, điện thoại thông minh, v.v. có kết nối không dây hoặc có dây với mạng. Ví dụ về các vị trí mà tại đó việc tư vấn được thực hiện bao gồm phòng khám của bác sỹ, phòng khám, các vị trí trên internet truy cập được bằng máy tính, các thiết bị cầm tay, v.v..

Theo một số phương án, các thao tác thu thập mẫu, xử lý mẫu, và giải trình tự được thực hiện tại vị trí thứ nhất và thao tác phân tích và suy ra được thực hiện tại vị trí thứ hai. Tuy nhiên, trong một số trường hợp, việc thu thập mẫu được tiến hành tại một vị trí (ví dụ, phòng khám của bác sỹ hoặc phòng khám) và việc xử lý mẫu và giải trình tự được thực hiện tại vị trí khác mà có thể là tùy ý cùng vị trí với nơi diễn ra việc phân tích và suy ra.

Theo các phương án khác nhau, trình tự của các thao tác nêu trên có thể được khởi động bởi người dùng hoặc thực thể bắt đầu thu thập mẫu, xử lý mẫu và/hoặc giải trình tự. Sau khi một hoặc nhiều thao tác đã bắt đầu thực thi, các thao tác khác có thể sẽ tự nhiên diễn ra sau đó. Ví dụ, thao tác giải trình tự có thể khiến cho trình tự đọc được thu thập tự động và được truyền đến thiết bị xử lý mà sau đó thực hiện, thường là tự động và có thể không cần đến sự can thiệp thêm của người dùng, thao tác phân tích và suy ra đột biến lệch bội. Theo một số phương án thực hiện, kết quả của thao tác xử lý này sau đó được tự động chuyển giao, có thể là với việc định dạng lại dưới dạng chẩn đoán, đến bộ phận hệ thống hoặc thực thể xử lý các báo cáo thông tin đến chuyên gia y tế và/hoặc người bệnh. Như đã giải thích, thông tin này cũng có thể được xử lý tự động để tạo ra kế hoạch điều trị, thử nghiệm, và/hoặc theo dõi, có thể là cùng với thông tin tư vấn. Do đó, việc bắt đầu thao tác ở giai đoạn sớm có thể khởi động trình tự đầu cuối trong đó chuyên gia y tế, người bệnh hoặc bên liên quan khác được cung cấp

chẩn đoán, kế hoạch, thông tin tư vấn và/hoặc thông tin khác hữu ích để hoạt động trong một điều kiện vật lý. Điều này có thể đạt được mặt dù các phần của toàn bộ hệ thống được tách nhau về mặt vật lý và có thể ở xa vị trí của, ví dụ, mẫu và thiết bị trình tự.

Fig.5 thể hiện một phương án thực hiện của hệ thống phân tán để tạo ra xác định đột biến hoặc chẩn đoán từ mẫu thử. Vị trí thu thập mẫu 01 được sử dụng để thu được mẫu thử từ người bệnh như phụ nữ mang thai hoặc người bệnh được cho là mắc ung thư. Sau đó các mẫu này được cung cấp đến vị trí xử lý và giải trình tự 03 tại đó mẫu thử có thể được xử lý và giải trình tự như được mô tả trên đây. Vị trí 03 bao gồm thiết bị xử lý mẫu cũng như thiết bị giải trình tự mẫu đã xử lý. Kết quả của việc giải trình tự, như được mô tả ở phần khác trong bản mô tả, là sự thu thập trình tự đọc thường được cung cấp ở định dạng điện tử và cung cấp đến mạng như mạng Internet được biểu thị bằng số tham chiếu 05 trên Fig.5.

Dữ liệu trình tự được cung cấp đến vị trí ở xa 07 tại đó việc phân tích và tạo kết quả xác định đột biến được thực hiện. Vị trí này có thể bao gồm một hoặc nhiều thiết bị tính toán mạnh mẽ như máy tính hoặc bộ xử lý. Sau khi tài nguyên tính toán tại vị trí 07 đã hoàn thành việc tính toán và tạo ra kết quả xác định đột biến từ thông tin trình tự nhận được, xác định đột biến này được chuyển tiếp trở lại mạng 05. Theo một số phương án thực hiện, không chỉ xác định đột biến được tạo ra ở vị trí 07 mà chẩn đoán liên quan cũng được tạo ra. Kết quả xác định đột biến và/hoặc chẩn đoán sau đó được truyền qua mạng và trở lại vị trí thu thập mẫu 01 như được minh họa trên Fig.5. Như đã giải thích, đây chỉ đơn giản là một trong nhiều biến đổi về cách các thao tác khác nhau liên quan đến việc tạo ra xác định đột biến hoặc chẩn đoán có thể được chia cho các vị trí khác nhau. Một biến thể thông thường bao gồm cung cấp thu thập và xử lý mẫu và giải trình tự tại một vị trí. Biến thể khác bao gồm cung cấp xử lý và giải trình tại cùng vị trí với phân tích và tạo xác định đột biến.

Fig.6 mô tả chi tiết về các tùy chọn để thực hiện các thao tác khác nhau tại các vị trí riêng biệt. Theo cách chi tiết nhất được mô tả trên Fig.6, mỗi trong số các thao tác sau đây được thực hiện ở vị trí riêng: thu thập mẫu, xử lý mẫu, giải trình tự, căn chỉnh trình tự đọc, xác định đột biến, phân tích, và báo cáo và/hoặc phát triển kế hoạch.

Theo một phương án tổng hợp một số thao tác này, xử lý mẫu và giải trình tự được thực hiện tại một vị trí và căn chỉnh trình tự đọc, xác định đột biến và chẩn đoán được thực hiện tại một vị trí riêng. Xem phần trên Fig.6 được nhận dạng bởi ký tự tham chiếu A. Theo phương

án thực hiện khác, được nhận dạng bởi ký tự B trên Fig.6, tất cả thu thập mẫu, xử lý mẫu, và giải trình tự đều được thực hiện tại cùng vị trí. Theo phương án thực hiện này, căn chỉnh trình tự đọc và xác định đột biến được thực hiện tại vị trí thứ hai. Cuối cùng, chẩn đoán và báo cáo và/hoặc phát triển kế hoạch được thực hiện tại vị trí thứ ba. Theo phương án thực hiện được mô tả bởi ký tự C trên Fig.6, thu thập mẫu được thực hiện tại vị trí thứ nhất, tất cả xử lý mẫu, giải trình tự, căn chỉnh trình tự đọc, xác định đột biến, và chẩn đoán được thực hiện cùng nhau tại vị trí thứ hai, và báo cáo và/hoặc phát triển kế hoạch được thực hiện tại vị trí thứ ba. Cuối cùng, theo phương án thực hiện được ký hiệu D trên Fig.6, thu thập mẫu được thực hiện tại vị trí thứ nhất, tất cả xử lý mẫu, giải trình tự, căn chỉnh trình tự đọc, và xác định đột biến được thực hiện tại vị trí thứ hai, và chẩn đoán và báo cáo và/hoặc quản lý kế hoạch được thực hiện tại vị trí thứ ba.

Một phương án đề xuất hệ thống dùng để xác định sự có mặt hoặc không có mặt của một hoặc nhiều đột biến lệch bội nhiễm sắc thể toàn phần ở bào thai khác nhau bất kỳ trong mẫu thử của người mẹ bao gồm các axit nucleic ở người mẹ và bào thai, hệ thống này bao gồm bộ giải trình tự để nhận mẫu axit nucleic và cung cấp thông tin trình tự axit nucleic ở người mẹ và bào thai từ mẫu; bộ xử lý; và vật ghi đọc được bằng máy chứa các lệnh để thực thi trên bộ xử lý, các lệnh này bao gồm:

- (a) mã để thu được thông tin trình tự các axit nucleic ở người mẹ và bào thai trong mẫu;
- (b) mã để sử dụng thông tin trình tự này để xác định bằng cách tính toán số lượng thể trình tự từ các axit nucleic ở người mẹ và bào thai cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ được chọn từ các nhiễm sắc thể 1-22, X, và Y và xác định số lượng thể trình tự cho ít nhất một trình tự nhiễm sắc thể chuẩn hóa hoặc trình tự đoạn nhiễm sắc thể chuẩn hóa cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ;
- (c) mã để sử dụng số lượng thể trình tự được xác định cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ và số lượng thể trình tự được xác định cho mỗi trình tự nhiễm sắc thể chuẩn hóa hoặc trình tự đoạn nhiễm sắc thể chuẩn hóa để tính toán một định lượng nhiễm sắc thể cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ; và
- (d) mã để so sánh mỗi trong số từng định lượng nhiễm sắc thể cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ với giá trị ngưỡng tương ứng cho mỗi trong số một

hoặc nhiều nhiễm sắc thể quan tâm, và do đó xác định sự có mặt hoặc không có mặt của một hoặc nhiều đột biến lệch bội nhiễm sắc thể toàn phần ở bào thai khác nhau bất kỳ trong mẫu.

Theo một số phương án, mã để tính toán một định lượng nhiễm sắc thể cho mỗi trong số một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ bao gồm mã để tính toán định lượng nhiễm sắc thể cho một nhiễm sắc thể được chọn trong số các nhiễm sắc thể quan tâm dưới dạng tỷ lệ của số lượng thể trình tự được xác định cho nhiễm sắc thể quan tâm được chọn và số lượng thể trình tự được xác định cho ít nhất một trình tự nhiễm sắc thể chuẩn hóa hoặc trình tự đoạn nhiễm sắc thể chuẩn hóa cho nhiễm sắc thể quan tâm được chọn.

Theo một số phương án, hệ thống này còn bao gồm mã để lặp lại việc tính toán định lượng nhiễm sắc thể cho mỗi trong số các đoạn nhiễm sắc thể còn lại bất kỳ của một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ.

Theo một số phương án, một hoặc nhiều nhiễm sắc thể quan tâm được chọn từ các nhiễm sắc thể 1-22, X, và Y bao gồm ít nhất hai mươi nhiễm sắc thể được chọn từ các nhiễm sắc thể 1-22, X, và Y, và trong đó các lệnh này bao gồm các lệnh để xác định sự có mặt hoặc không có mặt của ít nhất hai mươi đột biến lệch bội nhiễm sắc thể toàn phần ở bào thai khác nhau được xác định.

Theo một số phương án, ít nhất một trình tự nhiễm sắc thể chuẩn hóa là nhóm gồm các nhiễm sắc thể được chọn từ các nhiễm sắc thể 1-22, X, và Y. Theo các phương án khác, ít nhất một trình tự nhiễm sắc thể chuẩn hóa là một nhiễm sắc thể được chọn từ các nhiễm sắc thể 1-22, X, và Y.

Phương án khác đề xuất hệ thống dùng để xác định sự có mặt hoặc không có mặt của một hoặc nhiều đột biến lệch bội nhiễm sắc thể ở bào thai khác nhau một phần bất kỳ trong mẫu thử của người mẹ bao gồm các axit nucleic ở người mẹ và bào thai, hệ thống này bao gồm: bộ giải trình tự để nhận mẫu axit nucleic sample và cung cấp thông tin trình tự axit nucleic ở người mẹ và bào thai từ mẫu; bộ xử lý; và vật ghi đọc được bằng máy chứa các lệnh để thực thi trên bộ xử lý, các lệnh này bao gồm:

- (a) mã để thu được thông tin trình tự các axit nucleic ở người mẹ và bào thai trong mẫu;
- (b) mã để sử dụng thông tin trình tự này để xác định bằng cách tính toán số lượng thể trình tự từ các axit nucleic ở người mẹ và bào thai cho mỗi trong số một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ được chọn từ các nhiễm sắc thể 1-22,

X, và Y và xác định số lượng thẻ trình tự cho ít nhất một trình tự đoạn chuẩn hóa cho mỗi trong số một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ;

(c) mã hóa bằng cách sử dụng số lượng thẻ trình tự được xác định cho mỗi trong số một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ và số lượng thẻ trình tự được xác định cho trình tự đoạn chuẩn hóa để tính toán một định lượng đoạn nhiễm sắc thể cho mỗi trong số một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ; và

(d) mã để so sánh mỗi trong số từng định lượng đoạn nhiễm sắc thể cho mỗi trong số một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ với giá trị ngưỡng tương ứng cho mỗi trong số một hoặc nhiều đoạn nhiễm sắc thể bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ, và do đó xác định sự có mặt hoặc không có mặt của một hoặc nhiều đột biến lệch bội nhiễm sắc thể một phần ở bào thai khác nhau trong mẫu này.

Theo một số phương án, mã để tính toán một định lượng đoạn nhiễm sắc thể bao gồm mã để tính toán định lượng đoạn nhiễm sắc thể cho một đoạn được chọn trong số các đoạn nhiễm sắc thể dưới dạng tỷ lệ của số lượng thẻ trình tự được xác định cho đoạn nhiễm sắc thể được chọn và số lượng thẻ trình tự được xác định cho trình tự đoạn chuẩn hóa tương ứng cho đoạn nhiễm sắc thể được chọn.

Theo một số phương án, hệ thống này còn bao gồm mã để lặp lại việc tính toán định lượng đoạn nhiễm sắc thể cho mỗi trong số các đoạn nhiễm sắc thể còn lại bất kỳ của một hoặc nhiều đoạn bất kỳ của một hoặc nhiều nhiễm sắc thể quan tâm bất kỳ.

Theo một số phương án, hệ thống này còn bao gồm (i) mã để lặp các bước từ (a)-(d) cho các mẫu thử từ các đối tượng người mẹ khác nhau, và (ii) mã để xác định sự có mặt hoặc không có mặt của một hoặc nhiều đột biến lệch bội nhiễm sắc thể một phần ở bào thai khác nhau bất kỳ trong mỗi mẫu.

Theo các phương án khác của hệ thống bất kỳ trong số các hệ thống được đề xuất trong đây, mã này còn bao gồm mã để tự động ghi sự có mặt hoặc không có mặt của đột biến lệch bội nhiễm sắc thể ở bào thai như được xác định trong bước (d) trong hồ sơ bệnh án cho đối tượng người cung cấp mẫu thử của người mẹ, trong đó việc ghi này được thực hiện bằng cách sử dụng bộ xử lý.

Theo một số phương án của hệ thống bất kỳ trong số các hệ thống được đề xuất trong đây, bộ giải trình tự được tạo cấu hình để thực hiện giải trình tự thế hệ mới (next generation giải trình tự - NGS). Theo một số phương án, bộ giải trình tự được tạo cấu hình để thực hiện giải trình tự song song khối lượng lớn bằng cách sử dụng kỹ thuật giải trình tự bằng phương pháp tổng hợp với các đầu cuối nhuộm thuận nghịch. Theo các phương án khác, bộ giải trình tự được tạo cấu hình để thực hiện giải trình tự bằng phương pháp nối. Theo các phương án khác nữa, bộ giải trình tự được tạo cấu hình để thực hiện giải trình tự một phân tử.

Ví dụ thực hiện sáng chế

Ví dụ 1

Tạo và giải trình tự các thư viện giải trình tự cơ bản và được làm giàu

a. Tạo thư viện giải trình tự - giao thức rút gọn (ABB)

Tất cả các thư viện giải trình tự, tức là, các thư viện cơ bản và được làm giàu, được điều chế từ xấp xỉ 2 ng cfADN tinh sạch được tách chiết từ huyết tương của người mẹ. Việc tạo thư viện được thực hiện bằng cách sử dụng chất phản ứng của tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1 (Phần số E6000L; New England Biolabs, Ipswich, MA), cho Illumina® như sau. Bởi vì ADN huyết tương ngoài tế bào được phân đoạn tự nhiên, nên không cần phân đoạn thêm bằng khí dung hóa hoặc siêu âm phá tế bào thực hiện trên các ADN mẫu huyết tương. Các đầu nhô của xấp xỉ 2 ng đoạn cfADN tinh chế chứa trong 40 µl được biến đổi thành các đầu bằng phosphoryl hóa theo Modulun khôi phục đầu NEBNext® bằng cách nuôi cấy trong ống ly tâm 1,5 ml cfADN với 5 µl dung dịch đệm phosphoryl hóa 10X, 2 µl hỗn hợp dung dịch deoxynucleotit (10 mM mỗi dNTP), 1 µl dung dịch pha loãng 1:5 ADN Polymeraza I, 1 µl T4 ADN Polymeraza và 1 µl T4 Polynucleotit Kinaza được cung cấp trong tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1 trong 15 phút ở 20°C. Các enzym sau đó được khử hoạt tính bằng nhiệt bằng cách nuôi cấy hỗn hợp phản ứng ở 75°C trong 5 phút. Hỗn hợp được làm lạnh xuống 4°C, và việc gắn đuôi dA của ADN đầu bằng được thực hiện bằng cách sử dụng 10 µl hỗn hợp gốc để gắn đuôi dA chứa đoạn Klenow (từ 3' đến 5' exo minus) (tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1), và ủ trong 15 phút ở 37°C. Sau đó, đoạn Klenow được khử hoạt tính bằng nhiệt bằng cách ủ hỗn hợp phản ứng ở nhiệt độ 75°C trong 5 phút. Sau khi khử hoạt tính của đoạn Klenow, 1 µl dung dịch pha loãng 1:5 chứa Hỗn hợp oligo trình tự nối thuộc bộ gen Illumina (Phần số

1000521; Illumina Inc., Hayward, CA) được sử dụng để nối các trình tự nối Illumina (trình tự nối Y không chỉ số) vào các ADN có đuôi dA bằng cách sử dụng 4 µl T4 ADN ligaza cung cấp trong tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1, bằng cách ủ hỗn hợp phản ứng trong 15 phút ở 25°C. Hỗn hợp này được làm lạnh xuống 4°C, và cfADN được nối trình tự nối được tinh sạch khỏi các trình tự nối không được nối, các dạng nhị hợp của trình tự nối, và các chất phản ứng khác bằng cách sử dụng các hạt từ tính được cung cấp trong hệ thống tinh chế Agencourt AMPure XP PCR (Phần số A63881; Beckman Coulter Genomics, Danvers, MA). Mười tám chu kỳ PCR được thực hiện để làm giàu có chọn lọc cfADN được nối trình tự nối (25 µl) bằng cách sử dụng Hỗn hợp gốc có độ chính xác cao Phusion® (25µl; Finnzymes, Woburn, MA) và môi PCR của Illumina (0,5 µM cho mỗi môi) bổ sung với trình tự nối (Phần số 1000537 và 1000538). ADN được nối trình tự nối được đưa vào phản ứng PCR (98°C trong 30 giây; 18 chu kỳ 98°C trong 10 giây, 65°C trong 30 giây, và 72°C trong 30 giây; giãn nở lần cuối ở 72°C trong 5 phút, và giữ ở 4°C) bằng cách sử dụng môi PCR bộ gen Illumina (các phần số 1000537 và 1000538) và Hỗn hợp gốc Phusion HF PCR được cung cấp trong tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1, theo hướng dẫn của nhà sản xuất. Sản phẩm đã khuếch đại được tinh chế bằng cách sử dụng hệ thống tinh chế Agencourt AMPure XP PCR (Agencourt Bioscience Corporation, Beverly, MA) theo hướng dẫn của nhà sản xuất truy cập được từ trang www.beckmangenomics.com/products/AMPureXPProtocol_000387v001.pdf. Sản phẩm khuếch đại tinh chế được rửa giải trong 40 µl dung dịch đệm Qiagen EB, và nồng độ và sự phân bố kích thước của thư viện khuếch đại được phân tích bằng cách sử dụng kit Agilent ADN 1000 cho thiết bị 2100 Bioanalyzer (Agilent technologies Inc., Santa Clara, CA).

b. Tạo thư viện giải trình tự - giao thức độ dài đầy đủ

Giao thức độ dài đầy đủ được mô tả ở đây về bản chất là giao thức tiêu chuẩn cung cấp bởi Illumina, và chỉ khác với giao thức Illumina ở việc tinh chế thư viện khuếch đại. Giao thức Illumina lệnh cho thư viện khuếch đại được tinh chế bằng cách sử dụng điện di trên gel, trong khi đó giao thức được mô tả ở đây sử dụng các hạt từ tính cho cùng bước tinh chế. Xấp xỉ 2 ng cfADN tinh sạch được tách chiết từ huyết tương của người mẹ được sử dụng để tạo thư viện giải trình tự cơ bản bằng cách sử dụng tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1 (Phần số E6000L; New England Biolabs, Ipswich, MA) cho Illumina®

về cơ bản là theo hướng dẫn của nhà sản xuất. Tất cả các bước trừ bước tinh sạch cuối cùng của sản phẩm được nối trình tự nối, được thực hiện bằng cách sử dụng các hạt từ tính Agencourt và chất phản ứng thay vì cột tinh chế, được thực hiện theo giao thức này cùng với các chất phản ứng để điều chế mẫu NEBNext™ cho thư viện ADN bộ gen mà được giải trình tự bằng cách sử dụng Illumina® GAI. Giao thức NEBNext™ về cơ bản là tuân thủ theo giao thức cung cấp bởi Illumina, truy cập được từ trang grcf.jhmi.edu/hts/protocols/11257047_ChIP_Sample_Prep.pdf.

Phần nhô ra của xấp xỉ 2 ng đoạn cfADN tinh sạch chứa trong 40 µl được biến đổi thành các đầu bằng phosphoryl hóa theo Môđun khôi phục đầu NEBNext® bằng cách nuôi cấy 40µl cfADN với 5µl dung dịch đệm phosphoryl hóa 10X, 2 µl hỗn hợp dung dịch deoxynucleotit (10 mM mỗi dNTP), 1 µl dung dịch pha loãng 1:5 ADN Polymeraza I, 1 µl T4 ADN Polymeraza và 1 µl T4 Polynucleotit Kinaza được cung cấp trong tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1 trong ống ly tâm 200 µl trong máy chu trình nhiệt trong 30 phút ở 20°C. Mẫu được làm lạnh xuống 4°C, và được tinh chế bằng cách sử dụng cột QIAQuick được cung cấp trong Bộ tinh chế QIAQuick PCR (QIAGEN Inc., Valencia, CA) như sau. 50 µl chất phản ứng được chuyển đến ống ly tâm 1,5 ml, và 250 µl dung dịch đệm Qiagen PB được thêm vào. 300 µl thu được được chuyển đến cột QIAquick, được ly tâm với tốc độ 13.000 vòng/phút trong một phút trong ống ly tâm. Cột này được rửa bằng 750 µl dung dịch đệm Qiagen PE, và được ly tâm lại. Etanol dư được loại bỏ ra bằng quy trình ly tâm bổ sung trong 5 phút với tốc độ 13.000 vòng/phút. ADN được rửa giải trong 39 µl dung dịch đệm Qiagen EB bằng ly tâm. Việc nối đuôi dA của 34 µl ADN đầu bằng thu được bằng cách sử dụng 16 µl hỗn hợp chính có đuôi dA chứa đoạn Klenow (từ 3' đến 5' exo minus) (tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1), và nuôi cấy trong 30 phút ở 37°C theo Môđun nối đuôi dA NEBNext® của nhà sản xuất. Mẫu được làm lạnh xuống 4°C, và được tinh chế bằng cách sử dụng cột được cung cấp trong Bộ tinh chế MinElute PCR (QIAGEN Inc., Valencia, CA) như sau. 50 µl chất phản ứng được chuyển đến ống ly tâm 1,5 ml, và 250 µl dung dịch đệm Qiagen PB được thêm vào. 300 µl này được chuyển đến cột MinElute, được ly tâm với tốc độ 13.000 vòng/phút trong một phút trong ống ly tâm. Cột này được rửa bằng 750 µl dung dịch đệm Qiagen PE, và được ly tâm lại. Etanol dư được loại bỏ ra bằng quy trình ly tâm bổ sung trong 5 phút với tốc độ 13.000 vòng/phút. ADN được rửa

giải trong 15 µl dung dịch đệm Qiagen EB bằng ly tâm. Mười micrôlít nước rửa giải ADN được ủ với 1 µl dung dịch pha loãng 1:5 Hỗn hợp oligo trình tự nối thuộc bộ gen Illumina (Phần số 1000521), 15 µl dung dịch đệm phản ứng nối nhanh 2X, và 4 µl Quick T4 ADN ligaza, trong 15 phút ở 25°C theo Modul nối nhanh NEBNext®. Mẫu được làm lạnh xuống 4°C, và được tinh chế bằng cách sử dụng cột MinElute như sau. Một trăm năm mươi micrôlít dung dịch đệm Qiagen PE được thêm vào 30 µl phản ứng, và toàn bộ thể tích được chuyển đến cột MinElute, được ly tâm với tốc độ 13.000 vòng/phút trong một phút trong ống ly tâm. Cột này được rửa bằng 750 µl dung dịch đệm Qiagen PE, và được ly tâm lại. Etanol dư được loại bỏ ra bằng quy trình ly tâm bổ sung trong 5 phút với tốc độ 13.000 vòng/phút. ADN được rửa giải trong 28 µl dung dịch đệm Qiagen EB bằng ly tâm. Hai mươi ba micrôlít nước rửa giải ADN được nối trình tự nối được đưa vào 18 chu kỳ phản ứng PCR (98°C trong 30 giây; 18 chu kỳ 98°C trong 10 giây, 65°C trong 30 giây, và 72°C trong 30 giây; giãn nở lần cuối ở 72°C trong 5 phút, và giữ ở 4°C) bằng cách sử dụng mỗi PCR bộ gen Illumina (các phần số 100537 và 1000538) và Hỗn hợp chính Phusion HF PCR được cung cấp trong tập hợp chất phản ứng ADN điều chế ADN mẫu NEBNext™ 1, theo hướng dẫn của nhà sản xuất. Sản phẩm đã khuếch đại được tinh chế bằng cách sử dụng hệ thống tinh chế Agencourt AMPure XP PCR (Agencourt Bioscience Corporation, Beverly, MA) theo hướng dẫn của nhà sản xuất truy cập được từ trang www.beckmangenomics.com/products/AMPureXPgiao_thức_000387v001.pdf. Hệ thống tinh chế Agencourt AMPure XP PCR loại bỏ các dNTP không kết hợp, các đoạn môi, dạng nhị hợp của môi, muối và các tạp chất khác, và khôi phục các đơn vị siêu sao chép lớn hơn 100 bp. Sản phẩm đã khuếch đại được rửa giải từ các hạt Agencourt trong 40 µl dung dịch đệm Qiagen EB và thành phần cỡ hạt của các thư viện được phân tích bằng cách sử dụng Bộ Agilent ADN 1000 cho thiết bị 2100 Bioanalyzer (Agilent technologies Inc., Santa Clara, CA).

c. Phân tích các thư viện giải trình tự được tạo theo các giao thức rút gọn (a) và giao thức độ dài đầy đủ (b)

Các điện di đồ được tạo ra bởi thiết bị Bioanalyzer được thể hiện trên các hình vẽ Fig.7A và Fig.7B. Fig.7A thể hiện điện di đồ của ADN thư viện được điều chế từ cfADN tinh chế từ mẫu huyết tương M24228 bằng cách sử dụng giao thức độ dài đầy đủ được mô tả trong (a), và Fig.7B thể hiện điện di đồ của ADN thư viện được điều chế từ cfADN tinh chế từ mẫu

huyết tương M24228 bằng cách sử dụng giao thức độ dài đầy đủ được mô tả trong (b). Trên cả hai hình vẽ, các đỉnh 1 và 4 biểu diễn lần lượt dấu chuẩn dưới 15 bp, và dấu chuẩn trên 1.500; các số ở trên đỉnh biểu thị số lần di chuyển với các đoạn thư viện; và các đường ngang biểu thị ngưỡng thiết lập để tích hợp. Điện di đồ trên Fig.7A thể hiện đỉnh phụ của các đoạn là 187 bp và đỉnh chính của các đoạn là 263 bp, trong khi đó điện di đồ trên Fig.7B chỉ thể hiện một đỉnh ở 265 bp. Sự tích hợp các vùng đỉnh tạo ra nồng độ tính toán là 0,40 ng/ μ l cho ADN của đỉnh 187 bp trên Fig.7A, nồng độ 7,34 ng/ μ l cho ADN của đỉnh 263bp trên Fig.7A, và nồng độ 14,72 ng/ μ l cho ADN của đỉnh 265 bp trên Fig.7B. Các trình tự nối Illumina được nối vào cfADN được biết là 92 bp, được trừ từ 265 bp, biểu thị rằng kích thước đỉnh của cfADN là 173 bp. Đỉnh phụ tại 187 bp biểu diễn các đoạn của hai môi được nối đầu - cuối. Các đoạn hai môi tuyến tính được loại bỏ ra khỏi sản phẩm thư viện cuối cùng khi giao thức rút gọn được sử dụng. Giao thức rút gọn cũng loại bỏ các đoạn khác nhỏ hơn 187 bp. Trong ví dụ này, nồng độ cfADN được nối trình tự nối tính chế gấp đôi nồng độ cfADN được nối trình tự nối được tạo ra bằng cách sử dụng giao thức độ dài đầy đủ. Cần lưu ý rằng nồng độ của cfADN được nối trình tự nối của các đoạn cfADN được nối trình tự nối luôn lớn hơn nồng độ thu được bằng cách sử dụng giao thức độ dài đầy đủ (dữ liệu không được thể hiện).

Do đó ưu điểm của việc tạo thư viện giải trình tự bằng cách sử dụng giao thức rút gọn đó là thư viện thu được một cách phù hợp chỉ bao gồm một đỉnh chính trong khoảng 262-267 bp trong khi đó chất lượng của thư viện được điều chế bằng cách sử dụng giao thức độ dài đầy đủ biến đổi như được phản ánh bởi số lượng và tính di động của các đỉnh không phải đỉnh biểu diễn cfADN. Các sản phẩm không cfADN sẽ chiếm không gian trên cuvet dòng chảy và làm giảm chất lượng của khuếch đại cụm và việc chụp ảnh sau đó của các phản ứng giải trình tự, điều này làm cơ sở cho việc quy hoàn toàn cho tình trạng đột biến lệch bội. Giao thức rút gọn được thể hiện không ảnh hưởng đến việc giải trình tự thư viện.

Ưu điểm khác của việc tạo thư viện giải trình tự bằng cách sử dụng giao thức rút gọn đó là ba bước enzym gồm tạo đầu bằng, nối đuôi d-A, và nối trình tự nối, tốn ít hơn một giờ để hoàn thành, hỗ trợ cho việc phê duyệt và thực hiện dịch vụ chẩn đoán đột biến lệch bội nhanh.

Ưu điểm khác đó là ba bước enzym gồm tạo đầu bằng, nối đuôi d-A, và nối trình tự nối, được thực hiện trong cùng ống phản ứng, do đó tránh nhiều lần chuyển mẫu mà có thể dẫn đến mất vật liệu, và quan trọng hơn là có thể dẫn đến làm lẫn lộn mẫu và nhiễm bản mẫu.

Ví dụ 2

Xét nghiệm trước sinh không xâm lấn bằng cách sử dụng kích thước đoạn

Giới thiệu

Kể từ lần giới thiệu thương mại vào cuối năm 2011 và đầu năm 2012, xét nghiệm trước sinh không xâm lấn (non-invasive prenatal testing - NIPT) đối với ADN ngoài tế bào (cell free ADN - cfADN) trong huyết tương của người mẹ đã nhanh chóng trở thành phương pháp được lựa chọn để sàng lọc phụ nữ mang thai có nguy cơ bị đột biến lệch bội bào thai cao. Các phương pháp cơ bản dựa vào việc phân lập và giải trình tự cfADN trong huyết tương của phụ nữ mang thai, và đếm số lượng đoạn cfADN để so sánh với các vùng cụ thể của bộ gen người tham chiếu (tài liệu tham khảo: Fan et al., Lo et al.). Các phương pháp đếm phân tử và giải trình tự ADN cho phép xác định với độ chính xác cao các số lượng bản sao liên quan cho mỗi trong số các nhiễm sắc thể trên bộ gen. Đã đạt được độ nhạy và độ đặc hiệu cao để phát hiện các thể ba 21, 18 và 13 theo cách lặp lại trong nhiều nghiên cứu lâm sàng (các tài liệu tham khảo, trích dẫn phân tích tổng hợp Gil/Nicolaides).

Gần đây hơn, các nghiên cứu lâm sàng bổ sung đã cho thấy rằng phương pháp này có thể được mở rộng cho tập hợp sản khoa chung. Không có sự khác biệt nhận thấy được trong các tỷ lệ ADN bào thai giữa tập hợp có nguy cơ cao và tập hợp nguy cơ trung bình (các tài liệu tham khảo). Các kết quả nghiên cứu lâm sàng cho thấy rằng NIPT sử dụng đếm phân tử bằng giải trình tự cfADN thực hiện tương đương nhau ở cả hai tập hợp. Cải thiện đáng kể về mặt thống kê trong giá trị tiên đoán dương (positive predictive value - PPV) trên sàng lọc huyết thanh tiêu chuẩn đã được thể hiện (các tài liệu tham khảo). Các kết quả thử nghiệm dương tính giả thấp hơn, so với kết quả đo sinh hóa huyết thanh và sàng lọc độ mờ da gáy, đã làm giảm rõ rệt nhu cầu cần thực hiện các quy trình chẩn đoán xâm lấn (xem các tài liệu tham khảo Larion et al. trên nhóm của Abuhamad).

Với kết quả NIPT tốt trong tập hợp sản khoa chung, tính đơn giản trong tiến trình công việc và chi phí giờ đây đã trở thành yếu tố xem xét chính để thực hiện giải trình tự cfADN cho toàn bộ việc phát hiện đột biến lệch bội nhiễm sắc thể trong tập hợp sản khoa chung (tài

liệu tham khảo: ISPD Debate 1, Brisbane). Hầu hết các phương pháp NIPT trong phòng thí nghiệm sử dụng bước khuếch đại phản ứng chuỗi Polymeraza (Polymeraza chain reaction - PCR) sau khi tạo thư viện và giải trình tự một đầu dò hồi 10-20 triệu đoạn cfADN đơn nhất để đạt được độ nhạy hợp lý để phát hiện đột biến lệch bội. Độ phức tạp của tiến trình công việc dựa trên PCR và các yêu cầu giải trình tự sâu hơn đã giới hạn tiềm năng của xét nghiệm NIPT và dẫn đến chi phí cao hơn.

Tại đây chứng minh được rằng độ nhạy và độ đặc hiệu phân tích cao có thể đạt được với việc tạo thư viện đơn giản nhờ sử dụng đầu vào cfADN rất thấp không yêu cầu khuếch đại PCR. Phương pháp không PCR đơn giản hóa tiến trình công việc, cải thiện thời gian quay vòng và loại bỏ được các độ lệch vốn có ở phương pháp PCR. Tiến trình công việc không khuếch đại có thể được kết hợp với giải trình tự đầu bắt cặp để cho phép xác định độ dài đoạn cho mỗi thể và tổng tỷ lệ ADN bào thai trong mỗi mẫu. Bởi vì các đoạn cfADN thai ngắn hơn các đoạn của người mẹ [tài liệu tham khảo Quake 2010, cũng nên trích dẫn bài viết Lo's Science Clin Translation], nên việc phát hiện đột biến lệch bội ở bào thai từ huyết tương của người mẹ có thể được thực hiện hiệu quả và mạnh mẽ hơn nhiều, đòi hỏi ít đoạn cfADN duy nhất hơn. Kết hợp lại, đạt được độ nhạy và độ đặc hiệu phân tích được cải thiện với thời gian quay vòng rất nhanh với số lượng đoạn cfADN thấp hơn đáng kể. Điều này cho phép xét nghiệm NIPT được thực hiện với chi phí thấp hơn đáng kể, tạo thuận lợi cho việc áp dụng trong quần thể sản khoa nói chung.

Phương pháp

Các mẫu máu ngoại biên được đưa vào các ống BCT (Streck, Omaha, NE, Mỹ) và được chuyển đến phòng thí nghiệm Illumina CLIA ở thành phố Redwood để dùng cho xét nghiệm NIPT thương mại. Các mẫu thỏa thuận người bệnh đã ký cho phép các phân phân ước huyết tương thứ hai được ẩn danh dữ liệu và sử dụng cho nghiên cứu lâm sàng, ngoại trừ các mẫu người bệnh được gửi từ Bang New York. Các mẫu huyết tương cho công việc này được chọn để bao gồm cả bào thai bị đột biến lệch bội và không bị mắc bệnh trong khoảng của nồng độ cfADN và tỷ lệ ADN bào thai.

Đơn giản hóa xử lý thư viện

cfADN được tách chiết từ 900µL huyết tương của người mẹ bằng cách sử dụng kit tinh chế máu NucleoSpin 96 lỗ (Macherey-Nagel, Düren, Đức) với một số điều chỉnh nhỏ để chứa

được đầu vào dịch ly giải lớn hơn. cfADN phân lập được đặt trực tiếp vào quy trình thư viện giải trình tự mà không có sự chuẩn hóa bất kỳ đầu vào cfADN. Các thư viện giải trình tự được tạo bằng kit thư viện ADN không dùng PCR TruSeq (Illumina, San Diego, CA, Mỹ) với các chỉ số kép để tạo mã vạch các đoạn cfADN để nhận dạng mẫu. Những điều chỉnh sau đây cho giao thức thư viện được sử dụng để cải thiện sự tương thích của quá trình tạo thư viện với nồng độ cfADN đầu vào thấp. Thể tích đầu vào của khuôn được tăng lên, trong khi đó hỗn hợp chính để sửa chữa đầu, gắn đuôi A và nối và nồng độ trình tự nối được giảm xuống. Ngoài ra, sau khi sửa chữa đầu, bước dùng nhiệt được đưa vào để khử hoạt tính các enzym, bước tinh chế hạt SPRI (nhà cung cấp) sau khi khôi phục đầu được loại bỏ, và việc rửa giải trong bước tinh chế hạt SPRI sau khi nối sử dụng dung dịch đệm HT1 (Illumina).

Một bộ điều chỉnh chất lỏng MICROLAB® STAR (Hamilton, Reno, NV, Mỹ), được tạo cấu hình với đầu 96 kênh và 8 kênh hút nhỏ giọt 1-mL, được sử dụng để xử lý từng đợt 96 mẫu huyết tương cùng lúc. Bộ điều chỉnh chất lỏng này xử lý từng mẫu huyết tương riêng lẻ thông qua tách chiết ADN, điều chế và lượng tử hóa thư viện giải trình tự. Các thư viện mẫu riêng lẻ được lượng tử hóa với AccuClear (Biotium, Hayward, CA, Mỹ) và nhóm 48 mẫu được điều chế với các đầu vào chuẩn hóa tạo ra nồng độ cuối cùng 32 pM để giải trình tự.

Giải trình tự đầu bắt cặp

Giải trình tự ADN được thực hiện với dụng cụ Illumina NextSeq 500 sử dụng giải trình tự đầu bắt cặp 2x36 bp, cộng với 16 chu kỳ bổ sung để giải trình tự các mã vạch mẫu. Tổng số 364 mẫu được chạy trên đợt giải trình tự độc lập.

Các trình tự ADN ghép cặp được giải ghép bằng cách sử dụng bcl2fastq (Illumina) và được ánh xạ đến bộ gen người tham chiếu (hg19) bằng cách sử dụng thuật toán căn chỉnh trình tự bowtie2 [tham khảo Landmead]. Trình tự đọc ghép cặp phải khớp với các sợi xuôi và ngược mới được đếm. Tất cả các cặp ánh xạ được đếm vượt số điểm chất lượng ánh xạ 10 (Ruan et al.) với trình tự đọc thứ nhất duy nhất trên toàn bộ được gán cho các bin bộ gen độ rộng cố định liên tiếp không chồng lấn có kích thước 100 kb. Xấp xỉ 2% bộ gen thể hiện độ che phủ biến thiên cao trên tập hợp độc lập gồm các mẫu NIPT và được loại ra khỏi phân tích sau này.

Bằng việc sử dụng thông tin vị trí bộ gen và kích thước đoạn có sẵn từ các vị trí được ánh xạ của mỗi trong số hai đầu của đoạn cfADN được giải trình tự, hai biến được suy ra cho mỗi cửa sổ 100kb: (a) tổng số đếm đoạn ngắn dài dưới 150 cặp bazơ, và (b) một phần của các đoạn trong khoảng 80 và 150 cặp bazơ trong tập hợp gồm tất cả các đoạn dưới 250 cặp bazơ. Việc giới hạn kích thước các đoạn xuống dưới 150 cặp bazơ làm tăng nồng độ các đoạn bắt nguồn từ nhau thai, đây chính là đại diện cho ADN bào thai. Một phần của các đoạn ngắn biểu thị lượng cfADN bào thai liên quan trong hỗn hợp huyết tương. CfADN từ bào thai có thể ba nhiễm được mong đợi là có tỷ lệ trình tự đọc ngắn cao hơn ánh xạ đến nhiễm sắc thể ba so với bào thai thể nguyên bội mà lưỡng thể với nhiễm sắc thể đó.

Số đếm và tỷ lệ của các đoạn ngắn được chuẩn hóa độc lập để loại bỏ các độ lệch thử nghiệm mang tính hệ thống và các biến thiên riêng từng mẫu do hàm lượng guanin cytosin (GC) thuộc bộ gen bằng cách sử dụng quy trình thể hiện trên Fig.2D. Các giá trị đã chuẩn hóa được xét bớt bằng cách loại bỏ các bin lệch ra khỏi đường trung bình toàn bộ nhiễm sắc thể ở mức nhiều hơn 3 số đo độ lệch chuẩn mạnh. Cuối cùng, với mỗi trong số hai biến, các giá trị chuẩn hóa được xén bớt liên quan đến nhiễm sắc thể đích được so sánh với các giá trị trên nhiễm sắc thể tham chiếu chuẩn hóa để tạo ra thống kê t.

Dữ liệu từ mỗi lần chạy giải trình tự đầu bắt cặp tuân thủ theo bốn bước phân tích: 1) chuyển đổi trình tự đọc, 2) phân bin đặc điểm ở độ phân giải 100kb, 3) chuẩn hóa mỗi đặc điểm (số đếm và tỷ lệ) ở độ phân giải 100kb và 4) kết hợp các đặc điểm và chấm điểm phát hiện đột biến lệch bội. Ở bước 1, dữ liệu mẫu được giải ghép từ các mã vạch riêng lẻ, căn chỉnh với bộ gen và được lọc đối với chất lượng trình tự. Bước 2: tổng số đếm đoạn ngắn dài dưới 150 cặp bazơ, và tỷ lệ của các đoạn trong khoảng 80 và 150 cặp bazơ trong tập hợp gồm tất cả các đoạn dưới 250 cặp bazơ được xác định cho mỗi bin. Chênh lệch thử nghiệm và các biến thiên đặc hiệu mẫu được loại bỏ trong bước 3. Cuối cùng, nồng độ cao so với mẫu tham chiếu được xác định và được chấm điểm bằng cách sử dụng thử nghiệm t cho mỗi trong số số đếm và tỷ lệ, và được kết hợp để có điểm số cuối cùng để phát hiện đột biến lệch bội.

Phát hiện đột biến lệch bội nhiễm sắc thể

Các tác giả sáng chế đã thử nghiệm xem dữ liệu số đếm và tỷ lệ có thể được kết hợp để tăng cường khả năng phát hiện thể ba nhiễm 21 ở bào thai hay không. Mười sáu mẫu huyết tương từ phụ nữ mang thai với thể ba nhiễm 21 được xác nhận kiểu hình nhân và 294 mẫu từ

phụ nữ mang thai không bị bệnh được phân bố ngẫu nhiên trên các đợt xử lý, tạo ra chín cuvet dòng chảy để giải trình tự. Mỗi bước thuật toán được kiểm tra riêng rẽ để xác định khả năng của mỗi bước và kết hợp các bước để phát hiện đột biến lệch bội. Điểm số cuối cùng để phát hiện đột biến lệch bội ở bào thai trong trường hợp kết hợp được xác định là căn bậc hai của tổng bình phương của hai thống kê riêng lẻ, và một ngưỡng được áp dụng để tạo ra kết quả xác định là "phát hiện đột biến lệch bội" so với "không phát hiện đột biến lệch bội".

Tính toán tỷ lệ ADN bào thai

Với mỗi mẫu, tỷ lệ ADN bào thai được ước tính bằng cách sử dụng tỷ lệ giữa tổng số đoạn có kích thước [111,136 bp] với tổng số đoạn có kích thước [165,175 bp] trong tập con gồm các bin bộ gen 100 kb. Bằng cách sử dụng các mẫu từ phụ nữ đang mang thai bé trai, 10% đầu tiên của bin bộ gen có tương quan cao nhất với tỷ lệ ADN bào thai được suy ra từ số lượng bản sao của nhiễm sắc thể X [tài liệu tham khảo Rava] được xác định. Tương quan giữa ước tính tỷ lệ ADN bào thai dựa trên kích thước đoạn và ước tính suy ra từ nhiễm sắc thể X ở bào thai bé trai đã biết được tính bằng cách sử dụng phân tích kiểm chứng chéo bỏ ra một [REF] bao gồm cả chọn bin và ước tính tham số mô hình hồi quy. Tỷ lệ ADN bào thai đã ước tính sau đó được suy ra bằng cách sử dụng mô hình hồi quy tuyến tính từ tỷ lệ kích thước đoạn.

Kết quả:

Đơn giản hóa xử lý thư viện

Fig.8 thể hiện tiến trình tổng quát và dòng thời gian cho phiên bản mới này của NIPT so với tiến trình công việc trong phòng thí nghiệm chuẩn. Toàn bộ tiến trình công việc điều chế 96 mẫu để phân lập huyết tương, tách chiết cfADN, thiết lập thư viện, lượng tử hóa và hợp nhất có thể xử lý các mẫu trong ít hơn 6 giờ tổng thời gian điều chế trên một Hamilton STAR. Điều này so với 9 giờ và hai Hamilton STARs với các phương pháp dựa trên PCR sử dụng trong phòng thí nghiệm CLIA. Lượng cfADN tách chiết trên mỗi mẫu được lấy trung bình 60 pg/ μ L, và kết quả của đầu ra thư viện giải trình tự được tương quan tuyến tính ($R^2=0,94$) với đầu vào cfADN như được thể hiện trên Fig.9. Mức thu hồi trung bình lớn hơn 70% (thêm khoảng), biểu thị mức thu hồi hiệu quả cao của cfADN sau khi tinh chế hạt SPRI. Mỗi chuỗi giải trình tự sử dụng lượng chuẩn hóa gồm 48 mẫu được dồn kênh và tốn xấp xỉ

14 giờ để hoàn thành. Số lượng trung bình của các trình tự đọc ghép gắp được ánh xạ đơn nhất là XXX M với 95% mẫu trên YYY.

Giải trình tự đầu bắt cặp

Tổng thời gian giải trình tự mỗi đợt 48 mẫu là ít hơn 48 giờ trên NextSeq 500. Điều này so với 40 giờ (1 cuvet dòng chảy, 96 mẫu) hoặc 50 giờ (2 cuvet dòng chảy, 192 mẫu) cho quy trình phòng thí nghiệm trên HiSeq 2500. Các vị trí bộ gen được ánh xạ của cả hai đầu đoạn cfADN cung cấp thông tin kích thước đoạn cfADN. Fig.10 thể hiện sự phân bố kích thước đoạn cfADN đo được từ 324 mẫu từ các phụ nữ mang thai bé trai. Kích thước của các đoạn được ánh xạ đến các nhiễm sắc thể thường được biết đến là nguyên bội và cơ bản là đại diện cho các nhiễm sắc thể của người mẹ được biểu diễn bằng đường cong mảnh. Kích thước trung bình của đoạn chèn vào là 175 bp với XX% đoạn có kích thước đo trong khoảng 100bp và 200bp. Đường cong đậm biểu diễn kích thước đoạn chỉ phát sinh từ nhiễm sắc thể Y chỉ đại diện cho các đoạn cfADN bào thai. Sự phân bố kích thước từ các trình tự riêng cho nhiễm sắc thể Y nhỏ hơn, trung bình là 167 bp với chu kỳ 10 bazơ tại các kích thước đoạn ngắn hơn.

Bởi vì các đoạn ngắn hơn của cfADN được làm giàu với ADN bào thai, phân tích chọn lọc bằng cách chỉ sử dụng các đoạn ngắn hơn chỉ mong đợi là tăng khả năng đại diện bào thai liên quan do chọn lọc ưu tiên các trình tự đọc bào thai. Fig.11 thể hiện tỷ lệ ADN bào thai liên quan từ tổng số đếm của các trình tự đọc có đầu bắt cặp được chỉ ra so với số đếm từ các trình tự đọc có đầu bắt cặp mà nhỏ hơn 150 bp. Cuối cùng, tỷ lệ ADN bào thai trung bình tăng với hệ số 2 so với tổng số đếm mặc dù có tăng một chút trong phương sai. Điểm ngắt kích thước ở 150 bp cho thấy tạo ra sự cân bằng tối ưu cho số đếm với việc tăng khả năng đại diện bào thai so với phương sai trong số đếm.

Phát hiện đột biến lệch bội nhiễm sắc thể

Mỗi trong số các số đo, tổng số đếm, số đếm nhỏ hơn 150 bp, tỷ lệ số đếm được làm giàu với cfADN bào thai (số đếm trong khoảng 80 và 150 bp/số đếm <250bp) và kết hợp số đếm đoạn ngắn hơn với tỷ lệ, được kiểm tra về khả năng phân biệt các mẫu thể ba nhiễm 21 với thể nguyên bội ở nhiễm sắc thể 21. Fig.12 thể hiện kết quả cho mỗi trong số các số đo này. Tổng số đếm có giá trị trung bình XX số đếm trong khi số đếm nhỏ hơn 150 bp có giá trị trung bình YY số đếm. Tuy nhiên, như có thể thấy trên Fig 4A và 4B, số đếm nhỏ hơn cho thấy sự tách biệt tốt hơn giữa thể ba và nguyên bội cơ bản là vì số đo này được làm giàu cho

cfADN bào thai. Chỉ riêng phần cũng gần như hiệu quả bằng tổng số đếm để phân biệt đột biến lệch bội (Fig.4C), nhưng khi được sử dụng kết hợp với số đếm đoạn ngắn (Fig. 4D) cung cấp khả năng phân biệt cải thiện hơn so với chỉ riêng số đếm đoạn ngắn. Điều này cho thấy rằng tỷ lệ này cung cấp thông tin độc lập tăng cường khả năng phát hiện thể ba nhiễm 21. Khi so với tiến trình công việc phòng thí nghiệm CLIA hiện thời bằng cách sử dụng quá trình tạo thư viện với khuếch đại PCR và trung bình của 16M số đếm/mẫu, tiến trình công việc giải trình tự đầu bắt cặp, không dùng PCR thể hiện hiệu suất tương đương với số đếm/mẫu ít hơn đáng kể (ví dụ, 6 M số đếm/mẫu hoặc ít hơn) và tiến trình công việc chuẩn bị mẫu đơn giản hơn, ngắn hơn.

Tính toán tỷ lệ ADN bào thai

Bằng cách sử dụng các kết quả nhiễm sắc thể X từ phụ nữ mang thai bé trai, các giá trị nhiễm sắc thể chuẩn hóa có thể được sử dụng để xác định các tỷ lệ ADN bào thai cho số đếm (tham khảo ClinChem) và được so sánh với các kích thước đoạn cfADN khác nhau. Các tỷ lệ ADN bào thai suy ra từ nhiễm sắc thể X được sử dụng để hiệu chỉnh tỷ lệ cho tập hợp gồm 140 mẫu và ước tính hiệu suất bằng cách sử dụng kiểm chứng chéo bỏ ra một. Fig.13 thể hiện các kết quả của dự báo tỷ lệ ADN bào thai được kiểm chứng chéo và chứng minh tương quan giữa hai tập hợp dữ liệu, biểu thị rằng các ước tính tỷ lệ ADN bào thai có thể thu được từ các mẫu bất kỳ, bao gồm các kết quả từ phụ nữ mang thai bé gái khi tập hợp hiệu chỉnh được đo.

Thảo luận

Chứng minh được rằng độ nhạy và độ đặc hiệu phân tích cao để phát hiện đột biến lệch bội ở bào thai từ cfADN trong huyết tương của người mẹ có thể đạt được với quy trình tạo thư viện không dùng PCR kết hợp với giải trình tự ADN đầu bắt cặp. Phương pháp này đơn giản hóa tiến trình công việc, cải thiện thời gian quay vòng (Fig.8) và loại bỏ một số độ lệch vốn có khi dùng phương pháp PCR. Giải trình tự đầu bắt cặp cho phép xác định kích thước độ dài đoạn và tỷ lệ ADN bào thai có thể được sử dụng sau này để tăng cường phát hiện đột biến lệch bội ở các số đếm thể thấp hơn đáng kể so với các phương pháp thương mại hiện đang được thực hiện. Hiệu suất của phương án thực hiện đầu bắt cặp không dùng PCR có vẻ tương tự với các phương pháp giải trình tự một đầu sử dụng lên đến ba lần số thể.

Đơn giản hóa xử lý thư viện

Tiến trình công việc không dùng PCR có một số ưu điểm với phòng thí nghiệm lâm sàng. Bởi vì hiệu suất cao và đặc điểm tuyến tính của quá trình tạo thư viện, nhóm các mẫu chuẩn hóa để giải trình tự có thể được thực hiện trực tiếp từ các nồng độ thư viện mẫu riêng lẻ. Độ lệch vốn có trong khuếch đại PCR của quy trình tạo thư viện do đó được loại bỏ. Ngoài ra, không cần phải phân lập các bộ điều chỉnh chất lỏng riêng cho các hoạt động trước và sau PCR, điều này làm giảm gánh nặng về vốn cho phòng thí nghiệm. Tiến trình công việc được đơn giản hóa này cho phép các đợt mẫu được điều chế trong một ca của phòng thí nghiệm lâm sàng, và sau đó được giải trình tự và phân tích qua đêm. Cuối cùng, chi phí vốn giảm bớt, thời gian "thao tác tại chỗ" giảm bớt và quay vòng nhanh cho phép có thể giảm đáng kể chi phí và tính mạnh mẽ tổng thể của NIPT.

Giải trình tự đầu bắt cặp

Việc giải trình tự đầu bắt cặp trên hệ thống NextSeq 500 có một số ưu điểm để đếm các đoạn cfADN. Trước tiên, với các mã vạch chỉ số kép, các mẫu có thể được dồn kênh ở mức cao cho phép chuẩn hóa và hiệu chỉnh biến đổi chuỗi-chuỗi với độ tin cậy thống kê cao. Ngoài ra, bởi vì 48 các mẫu được dồn kênh trên mỗi chuỗi, và lượng cần thiết trên cuvet dòng chảy để nhóm cụm bị giới hạn, yêu cầu đầu vào trên mỗi mẫu được giảm đi đáng kể, cho phép tiến trình công việc tạo thư viện không dùng PCR được sử dụng. Với kết quả cfADN điển hình xấp xỉ 5 ng mỗi mẫu, các nhà nghiên cứu có thể nhận được 2-3 chuỗi giải trình tự trên mỗi mẫu mặc dù không có khuếch đại PCR. Điều này trái ngược với các phương pháp khác đòi hỏi lượng đầu vào huyết tương đáng kể từ nhiều ống máu để tạo ra đủ cfADN để xác định đột biến lệch bội (REF). Cuối cùng, quy trình giải trình tự đầu bắt cặp cho phép xác định kích thước đoạn cfADN và làm giàu phân tích cho cfADN bào thai.

Phát hiện đột biến lệch bội nhiễm sắc thể

Kết quả của chúng tôi đã cho thấy rằng số đếm các đoạn cfADN dưới 150 bp có thể phân biệt tốt hơn đột biến lệch bội từ các nhiễm sắc thể nguyên bội hơn là tổng số đếm. Kết quả quan sát này trái ngược với các kết quả của Fan et al. đề xuất rằng độ chính xác của thống kê đếm có thể được giảm bớt bằng cách sử dụng các đoạn ngắn hơn (Fan et al.) do giảm số lượng số đếm có sẵn. Phần của các đoạn rút gọn cũng cung cấp một số biệt hóa để phát hiện thể ba 21 như đề cập bởi Yu et al., mặc dù có khoảng động nhỏ hơn số đếm. Tuy nhiên, kết

hợp các kết quả số đếm và phần trong việc tách các mẫu thể ba 21 từ nguyên bội, và cho thấy rằng hai số đo này là các phương pháp bổ sung nhau để biểu diễn nhiễm sắc thể. Các số đo sinh học khác, ví dụ metyl hóa, cũng có thể cung cấp thông tin trực giao có thể tăng cường tỷ lệ tín hiệu trên nhiễu để phát hiện đột biến lệch bội.

Tính toán tỷ lệ ADN bào thai

Các phương pháp được trình bày ở đây ước tính tỷ lệ ADN bào thai trong mỗi mẫu mà không tạo ra thêm công việc trong phòng thí nghiệm. Với nhiều mẫu trên mỗi cuvet dòng chảy, xấp xỉ một nửa số đó là bé trai, ước tính tỷ lệ ADN bào thai chính xác có thể thu được cho tất cả các mẫu bằng cách hiệu chuẩn phép đo tỷ lệ ADN bào thai từ thông tin kích thước đoạn với thông tin được xác định từ các mẫu bé trai. Trong bối cảnh thương mại, thử nghiệm lâm sàng của các nhà nghiên cứu đã cho thấy rằng các phương pháp đếm chuẩn sử dụng số lượng lớn các thể một đầu đã dẫn đến tỷ lệ âm giả rất thấp ngay cả khi không có các số đo tỷ lệ ADN bào thai cụ thể (REF). Với giới hạn tương tự của phát hiện được quan sát thấy ở đây, mong đợi đạt được hiệu suất xét nghiệm tương đương.

Kết luận

Chứng minh được rằng độ nhạy và độ đặc hiệu phân tích cao để phát hiện đột biến lệch bội ở bào thai từ cfADN trong huyết tương của người mẹ có thể đạt được với việc tạo thư viện không dùng PCR kết hợp với giải trình tự ADN đầu bắt cặp. Tiến trình công việc đã đơn giản hóa này có thời gian quay vòng rất nhanh, có khả năng cho phép NIPT được thực hiện với chi phí thấp hơn đáng kể để sử dụng trong tập hợp sản khoa chung. Ngoài ra, các kỹ thuật giải trình tự đầu bắt cặp có khả năng đo lường các hiện tượng sinh học khác, cũng như cung cấp các ứng dụng lâm sàng khác. Ví dụ, thông tin kích thước từ các vùng metyl hóa cụ thể của bộ gen hoặc các đảo CpG có thể cung cấp số đo trực giao khác để tăng cường việc phát hiện các biến thể số lượng bản sao trên bộ gen.

Sáng chế có thể được bộc lộ ở dạng cụ thể khác mà không nằm ngoài tinh thần hoặc các đặc tính của nó. Các phương án được mô tả phải được xem xét theo tất cả các khía cạnh chỉ mang tính minh họa và không giới hạn. Phạm vi của sáng chế, do đó, được bộc lộ bằng các điểm yêu cầu bảo hộ kèm theo chứ không phải là phần mô tả trên đây. Tất cả những thay đổi nằm trong ý nghĩa và khoảng tương đương của các điểm yêu cầu bảo hộ được coi là nằm trong phạm vi của chúng.

YÊU CẦU BẢO HỘ

1. Phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của trình tự axit nucleic quan tâm trong mẫu thử nghiệm chứa các đoạn axit nucleic ngoài tế bào có nguồn gốc từ hai hoặc nhiều bộ gen, phương pháp được thực hiện nhờ sử dụng hệ thống máy tính bao gồm một hoặc nhiều bộ xử lý và bộ nhớ hệ thống, bao gồm các bước:

(a) nhận, bằng hệ thống máy tính, các trình tự đọc thu được bằng cách giải trình tự đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm;

(b) sắp thẳng, bởi một hoặc nhiều bộ xử lý, các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu chứa trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin;

(c) xác định các kích thước đoạn của ít nhất một số trong số các đoạn axit nucleic ngoài tế bào có mặt trong mẫu thử nghiệm;

(d) đối với các đoạn axit nucleic ngoài tế bào xác định được là nằm trong domain kích thước thứ nhất, tính, bởi một hoặc nhiều bộ xử lý, các độ che phủ thứ nhất của các thẻ trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin:

(i) xác định số lượng thẻ trình tự sắp thẳng với bin, và

(ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao;

(e) đối với các đoạn axit nucleic ngoài tế bào xác định được là nằm trong domain kích thước thứ hai, tính, bởi một hoặc nhiều bộ xử lý, các độ che phủ thứ hai của các thẻ trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin:

(i) xác định số lượng thẻ trình tự sắp thẳng với bin, và

(ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao;

và

(f) xác định biến thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng tỷ số khả dĩ tính được từ các độ che phủ thứ nhất và các độ che phủ thứ hai.

2. Phương pháp theo điểm 1, trong đó tỷ số khả dĩ được tính từ tham số thống kê t của các độ che phủ thứ nhất và tham số thống kê t của các độ che phủ thứ hai, trong đó tham số thống kê t được tính nhờ sử dụng các độ che phủ của các bin ở trình tự quan tâm và các độ che phủ của các bin trong trình tự tham chiếu đối với trình tự quan tâm.

3. Phương pháp theo điểm 1, trong đó domain kích thước thứ nhất bao gồm các đoạn axit nucleic ngoài tế bào có gần như tất cả các kích thước trong mẫu thử nghiệm, và domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn kích thước xác định.
4. Phương pháp theo điểm 1, trong đó domain kích thước thứ hai chỉ bao gồm các đoạn axit nucleic ngoài tế bào nhỏ hơn khoảng 150 bp.
5. Phương pháp theo điểm 1, trong đó tỷ số khả dĩ được tính là khả năng thứ nhất mà mẫu thử nghiệm là mẫu lệch bội so với khả năng thứ hai mà mẫu thử nghiệm là mẫu nguyên bội.
6. Phương pháp theo điểm 1, trong đó tỷ số khả dĩ được tính từ một hoặc nhiều giá trị của tỷ lệ ADN bào thai (fetal fraction) cùng với các độ che phủ thứ nhất và các độ che phủ thứ hai.
7. Phương pháp theo điểm 6, trong đó một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin về kích thước của các đoạn axit nucleic ngoài tế bào.
8. Phương pháp theo điểm 7, trong đó giá trị của tỷ lệ ADN bào thai được tính bằng cách: thu phân bố tần suất của kích thước của các đoạn axit nucleic ngoài tế bào; và áp dụng phân bố tần suất này cho mô hình liên hệ tỷ lệ ADN bào thai với tần suất kích thước đoạn để thu được giá trị tỷ lệ ADN bào thai.
9. Phương pháp theo điểm 6, trong đó một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị của tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin độ che phủ đối với các bin của bộ gen tham chiếu.
10. Phương pháp theo điểm 9, trong đó giá trị của tỷ lệ ADN bào thai được tính bằng cách: áp dụng các giá trị độ che phủ của nhiều bin cho mô hình liên hệ tỷ lệ ADN bào thai với độ che phủ của bin để thu được giá trị tỷ lệ ADN bào thai.
11. Phương pháp theo điểm 6, trong đó một hoặc nhiều giá trị của tỷ lệ ADN bào thai bao gồm giá trị tỷ lệ ADN bào thai tính được nhờ sử dụng thông tin độ che phủ đối với các bin

của nhiễm sắc thể giới tính.

12. Phương pháp theo điểm 6, trong đó tỷ số khả dĩ được tính từ tỷ lệ ADN bào thai, tham số thống kê t của các đoạn ngắn, và tham số thống kê t của tất cả các đoạn, trong đó các đoạn ngắn là các đoạn axit nucleic ngoài tế bào trong khoảng kích thước thứ nhất nhỏ hơn kích thước tiêu chuẩn, và tất cả các đoạn này là các đoạn axit nucleic ngoài tế bào bao gồm các đoạn ngắn và các đoạn dài hơn kích thước tiêu chuẩn.

13. Phương pháp theo điểm 12, trong đó tỷ số khả dĩ được tính:

$$LR = \frac{\sum_{ff_{total}} q(ff_{total}) * p_1(T_{short}, T_{all} | ff_{est})}{p_0(T_{short}, T_{all})}$$

trong đó p_1 tượng trưng cho khả năng mà dữ liệu đến từ phân bố chuẩn nhiều biến biểu diễn mô hình 3 bản sao hoặc 1 bản sao, p_0 tượng trưng cho khả năng mà dữ liệu đến từ phân bố chuẩn nhiều biến biểu diễn mô hình 2 bản sao, T_{short} , T_{all} là các điểm số T tính được từ độ che phủ nhiễm sắc thể được tạo ra từ các đoạn ngắn và tất cả các đoạn, và $q(ff_{total})$ là phân bố mật độ của tỷ lệ ADN bào thai.

14. Phương pháp theo điểm 1, trong đó tỷ số khả dĩ được tính cho thể đơn nhiễm X, thể ba nhiễm X, thể ba nhiễm 13, thể ba nhiễm 18, hoặc thể ba nhiễm 21.

15. Phương pháp theo điểm 1, trong đó chuẩn hóa số lượng các thể trình tự bao gồm: chuẩn hóa đối với nội dung GC của mẫu thử nghiệm, chuẩn hóa đối với biên dạng sóng toàn phần của biến thiên của tập huấn luyện, và/hoặc chuẩn hóa đối với một hoặc nhiều thành phần thu được từ phân tích thành phần nguyên tắc.

16. Phương pháp theo điểm 2, trong đó vùng tham chiếu được chọn từ nhóm gồm có: tất cả các nhiễm sắc thể khỏe mạnh, các nhiễm sắc thể khỏe mạnh không bao gồm trình tự quan tâm, ít nhất nhiễm sắc thể bên ngoài trình tự quan tâm, và tập con nhiễm sắc thể được chọn từ nhiễm sắc thể khỏe mạnh, trong đó nhiễm sắc thể khỏe mạnh là các nhiễm sắc thể thường không phải các nhiễm sắc thể 13, 18, và 21.

17. Phương pháp theo điểm 16, trong đó vùng tham chiếu bao gồm các nhiễm sắc thể khỏe mạnh mà đã được xác định là mang lại khả năng dò tín hiệu tốt nhất cho tập mẫu huấn luyện.

18. Phương pháp theo điểm 2, phương pháp này còn bao gồm bước tính các giá trị của tham số kích thước cho các bin bằng cách, đối với mỗi bin:
- (i) xác định giá trị của tham số kích thước từ các kích thước của các đoạn axit nucleic ngoài tế bào trong bin, và
 - (ii) chuẩn hóa giá trị của tham số kích thước bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; và
- xác định tham số thống kê dựa trên kích thước cho trình tự quan tâm nhờ sử dụng các giá trị của tham số kích thước của các bin ở trình tự quan tâm và các giá trị của tham số kích thước của các bin trong vùng tham chiếu cho trình tự quan tâm.
19. Phương pháp theo điểm 18, trong đó tỷ số khả dĩ của (f) được tính từ tham số thống kê thứ nhất, tham số thống kê thứ hai, và tham số thống kê dựa trên kích thước.
20. Phương pháp theo điểm 18, trong đó tỷ số khả dĩ của (f) được tính từ tham số thống kê dựa trên kích thước và tỷ lệ ADN bào thai.
21. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm việc so sánh tỷ số khả dĩ với tiêu chuẩn xác định bất thường bản sao để xác định biến thể số lượng bản sao ở trình tự quan tâm.
22. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm việc thu nhiều tỷ số khả dĩ và áp dụng nhiều tỷ số khả dĩ này vào cây quyết định để xác định trường hợp bội thể cho mẫu thử nghiệm.
23. Hệ thống ước lượng số lượng bản sao của trình tự axit nucleic quan tâm trong mẫu thử nghiệm, hệ thống này bao gồm:
- bộ giải trình tự để nhận các đoạn axit nucleic ngoài tế bào từ mẫu thử nghiệm và đưa ra các thông tin trình tự axit nucleic của mẫu thử nghiệm;
 - bộ xử lý; và
 - một hoặc nhiều vật ghi đọc được bằng máy tính có lưu trữ các lệnh để thực thi trên bộ xử lý này để:
- (a) nhận các trình tự đọc thu được bằng cách giải trình tự đoạn axit

nucleic ngoài tế bào trong mẫu thử nghiệm;

(b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu chứa trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin;

(c) xác định các kích thước đoạn của ít nhất một số trong số các đoạn axit nucleic ngoài tế bào có mặt trong mẫu thử nghiệm;

(d) đối với các đoạn axit nucleic ngoài tế bào xác định được là nằm trong domain kích thước thứ nhất, tính các độ che phủ thứ nhất của các thẻ trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin:

(i) xác định số lượng thẻ trình tự sắp thẳng với bin, và

(ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao;

(e) đối với các đoạn axit nucleic ngoài tế bào xác định được là nằm trong domain kích thước thứ hai, tính các độ che phủ thứ hai của các thẻ trình tự đối với các bin của bộ gen tham chiếu bằng cách, đối với mỗi bin:

(i) xác định số lượng thẻ trình tự sắp thẳng với bin, và

(ii) chuẩn hóa số lượng các thẻ trình tự sắp thẳng với bin này bằng cách tính đến các biến thiên giữa các bin do các yếu tố khác ngoài biến thể số lượng bản sao; và

(f) xác định biến thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng tỷ số khả dĩ tính được từ các độ che phủ thứ nhất và các độ che phủ thứ hai.

24. Phương pháp xác định biến thể số lượng bản sao (copy number variation - CNV) của trình tự axit nucleic quan tâm trong mẫu thử nghiệm chứa các đoạn axit nucleic ngoài tế bào có nguồn gốc từ hai hoặc nhiều bộ gen, phương pháp bao gồm các bước:

(a) nhận các trình tự đọc thu được bằng cách giải trình tự đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm;

(b) sắp thẳng các trình tự đọc của các đoạn axit nucleic ngoài tế bào hoặc sắp thẳng các đoạn chứa các trình tự đọc với các bin của bộ gen tham chiếu chứa trình tự quan tâm, nhờ đó tạo ra các thẻ trình tự thử nghiệm, trong đó bộ gen tham chiếu được chia thành nhiều bin;

(c) xác định các kích thước đoạn của các đoạn axit nucleic ngoài tế bào có trong mẫu thử nghiệm;

(d) tính độ che phủ của các thể trình tự đối với các bin của bộ gen tham chiếu nhờ sử dụng các thể trình tự cho các đoạn axit nucleic ngoài tế bào có các kích thước ở domain kích thước thứ nhất;

(e) tính độ che phủ của các thể trình tự đối với các bin của bộ gen tham chiếu nhờ sử dụng các thể trình tự cho các đoạn axit nucleic ngoài tế bào có các kích thước ở domain kích thước thứ hai, trong đó domain kích thước thứ hai khác với domain kích thước thứ nhất;

(f) tính các đặc trưng kích thước đối với các bin của bộ gen tham chiếu nhờ sử dụng các kích thước đoạn xác định được ở (c); và

(g) xác định biên thể số lượng bản sao ở trình tự quan tâm nhờ sử dụng độ che phủ tính được ở (d) và (e) và các đặc trưng kích thước tính được ở (f).

25. Phương pháp theo điểm 24, trong đó (g) bao gồm việc tính tham số thống kê t cho trình tự quan tâm nhờ sử dụng các đặc trưng kích thước của các bin ở trình tự quan tâm tính được ở (f).

26. Phương pháp theo điểm 2, trong đó tham số thống kê t được tính như sau:

$$t = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

trong đó

x_1 là độ che phủ bin của trình tự quan tâm,

x_2 là độ che phủ bin của vùng tham chiếu,

s_1 là độ lệch chuẩn của các độ che phủ của các bin ở trình tự quan tâm,

s_2 là độ lệch chuẩn của các độ che phủ của các bin trong vùng tham chiếu,

n_1 là số lượng bin ở trình tự quan tâm, và

n_2 là số lượng bin trong vùng tham chiếu.

27. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm, trước bước (a), bước:

tách chiết các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm từ mẫu huyết tương của phụ nữ mang thai, trong đó các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm chứa axit nucleic có nguồn gốc từ bào thai và axit nucleic có nguồn gốc từ người mẹ; và

giải trình tự các đoạn axit nucleic ngoài tế bào này để thu được các đoạn đọc trình tự.

28. Phương pháp theo điểm 27, trong đó phương pháp này còn bao gồm bước: xác định rằng bào thai bị bệnh bởi bất thường di truyền liên quan đến biến thể số lượng bản sao ở trình tự quan tâm.

29. Phương pháp theo điểm 28, trong đó phương pháp này còn bao gồm bước kê đơn, bắt đầu, và/hoặc thay đổi phát đồ điều trị, trong đó phát đồ điều trị được thiết kế để điều trị bất thường di truyền ảnh hưởng đến bào thai.

30. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm, trước bước (a), bước:

tách chiết các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm từ cá thể, trong đó các đoạn axit nucleic ngoài tế bào này chứa axit nucleic có nguồn gốc từ các tế bào ung thư; và

giải trình tự các đoạn axit nucleic ngoài tế bào này để thu được các đoạn đọc trình tự.

31. Phương pháp theo điểm 30, trong đó phương pháp này còn bao gồm bước: xác định rằng cá thể này bị bệnh bởi bệnh ung thư liên quan đến biến thể số lượng bản sao ở trình tự quan tâm.

32. Phương pháp theo điểm 31, phương pháp này còn bao gồm bước kê đơn, bắt đầu, và/hoặc thay đổi phát đồ điều trị, trong đó phát đồ điều trị được thiết kế để điều trị bệnh ung thư ảnh hưởng đến cá thể này.

33. Phương pháp theo điểm 30, trong đó các đoạn axit nucleic ngoài tế bào trong mẫu thử nghiệm này được tách chiết từ mẫu huyết tương của cá thể này.

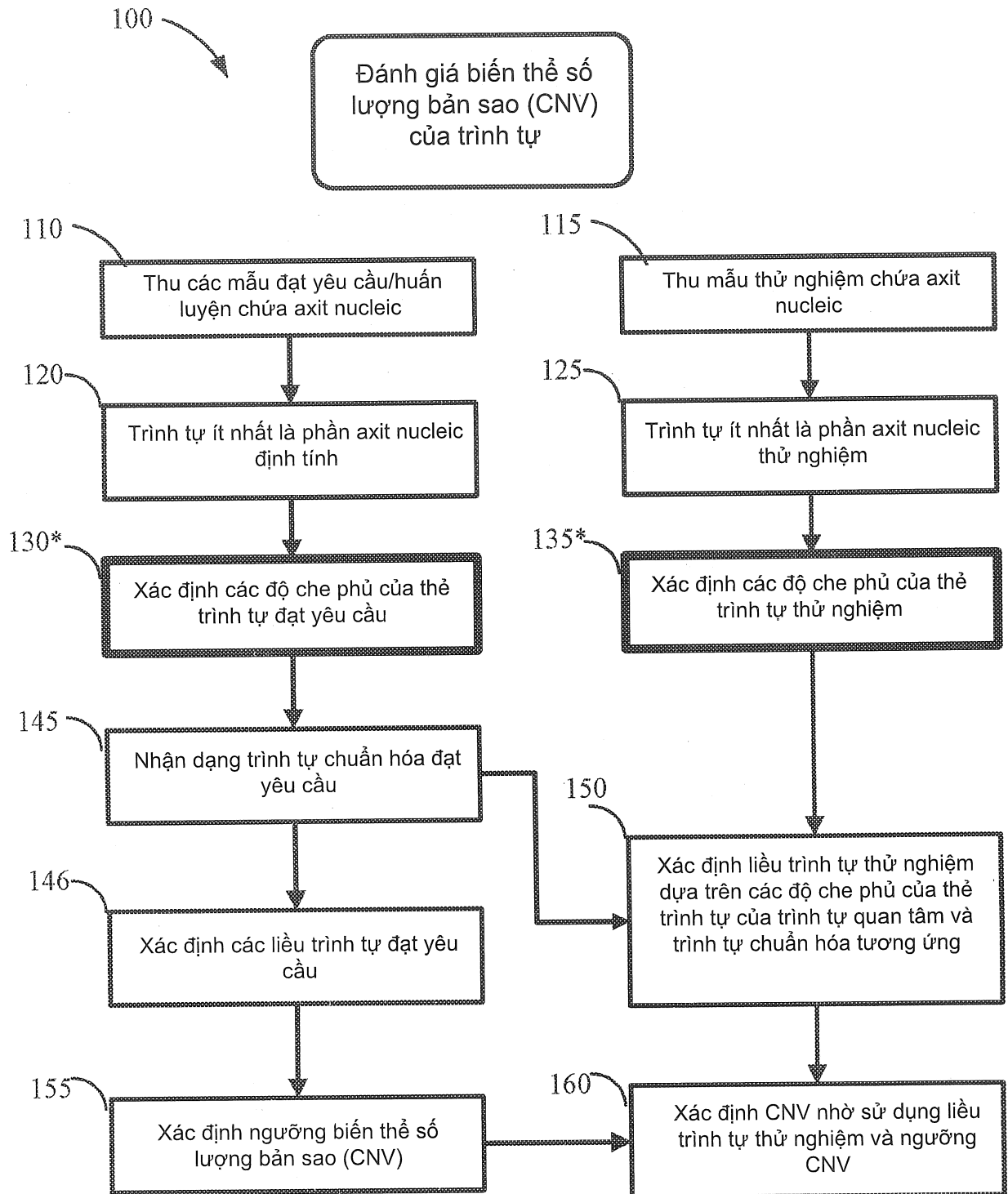


Fig.1

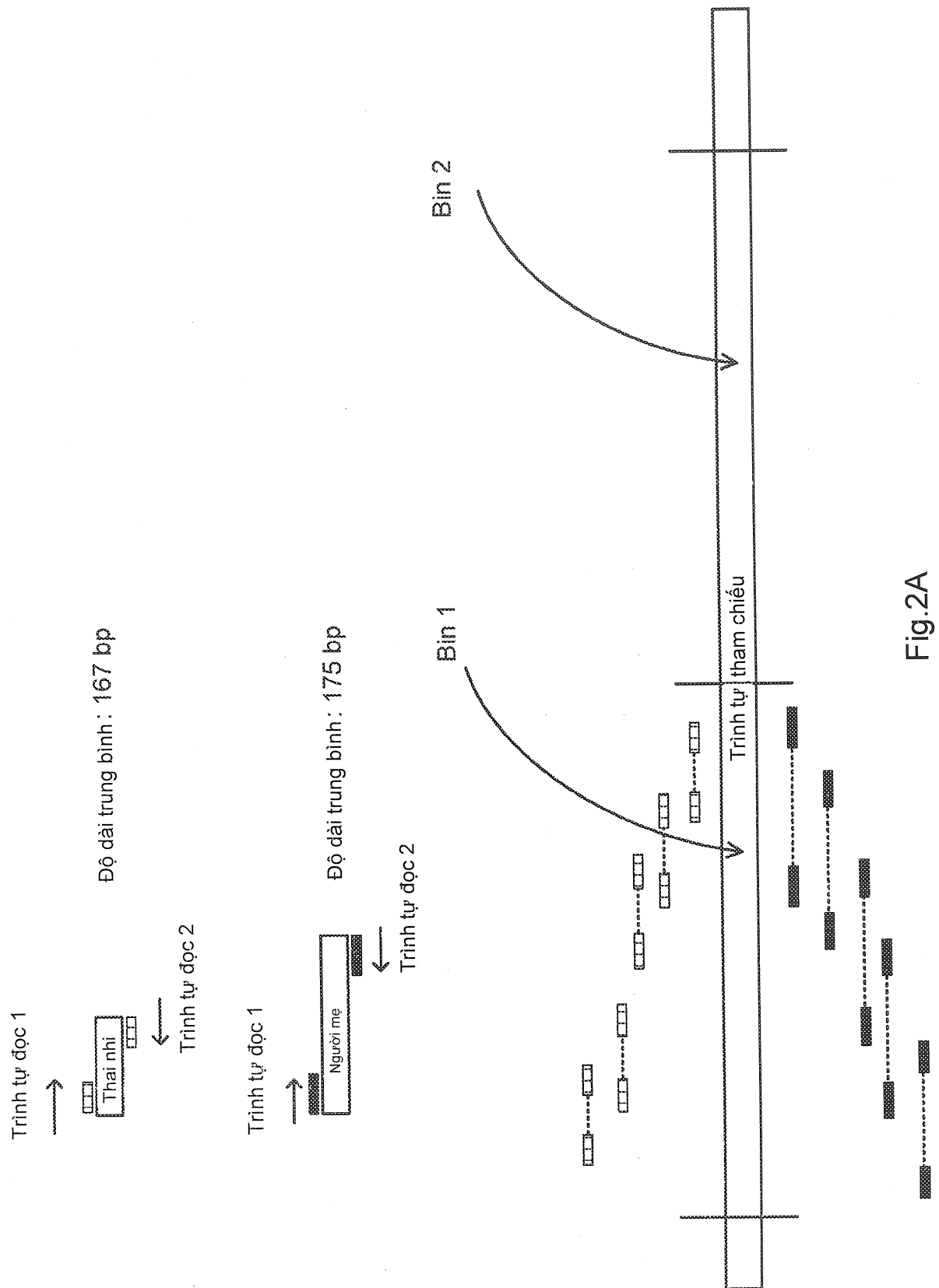


Fig.2A

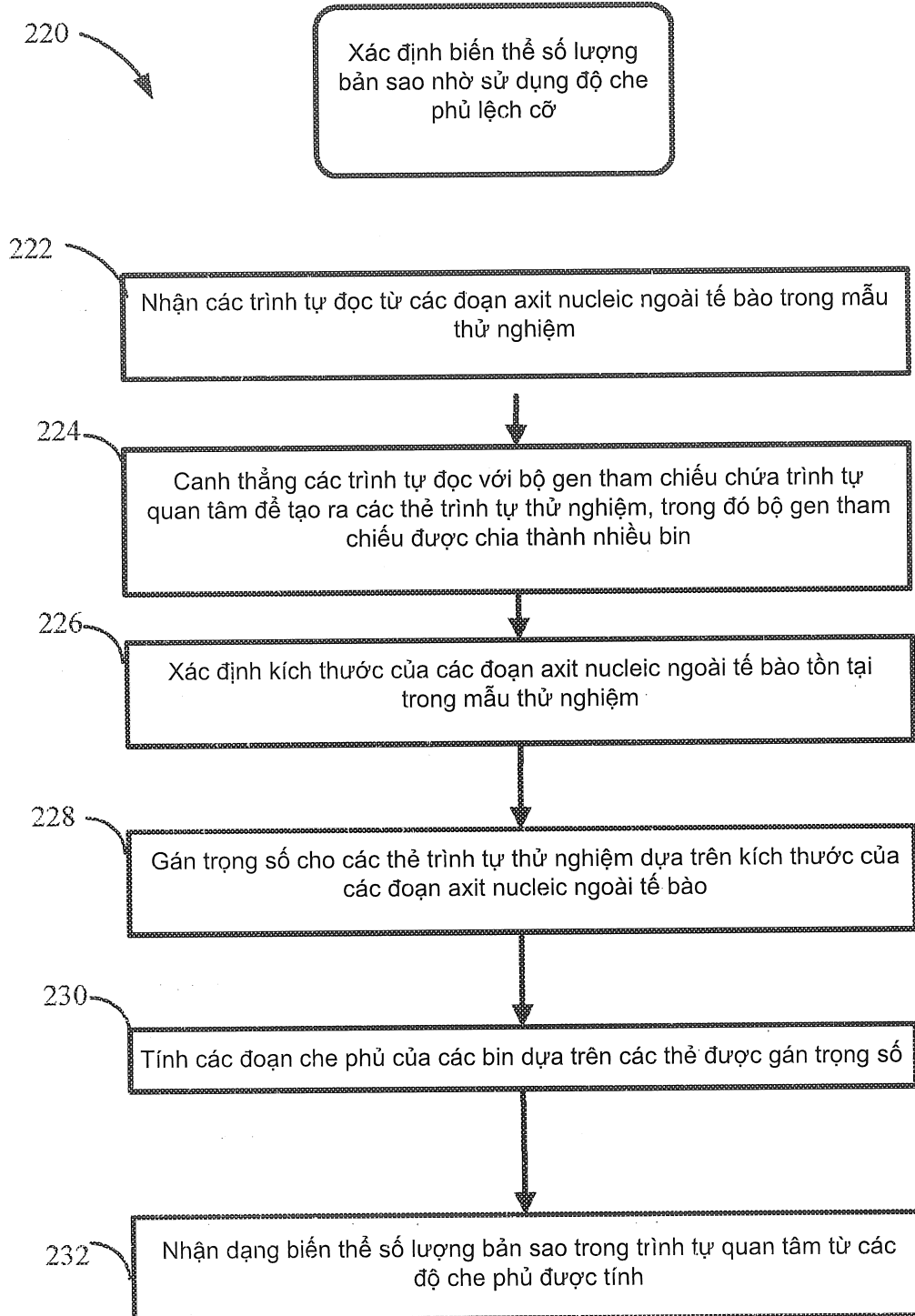


Fig.2B

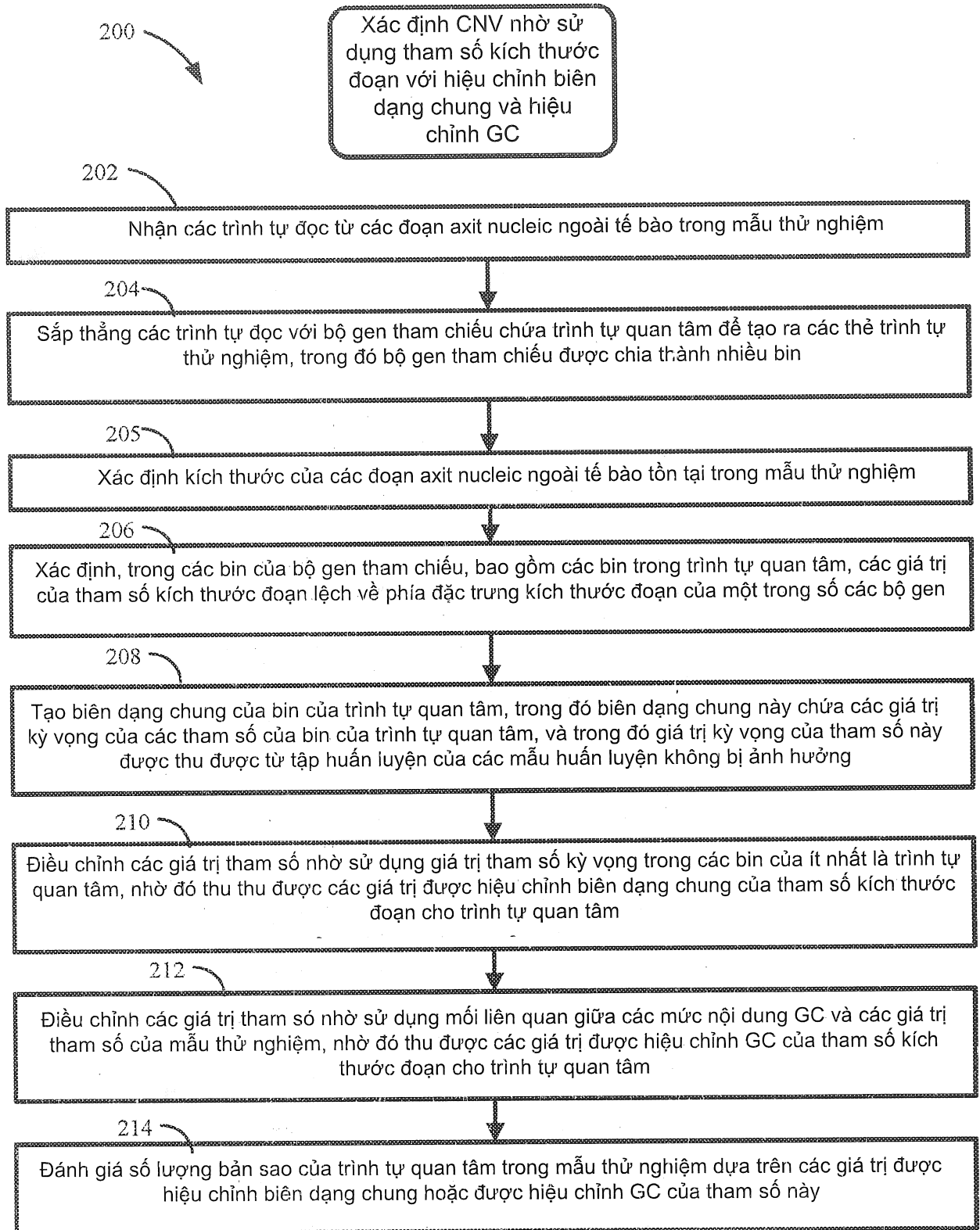
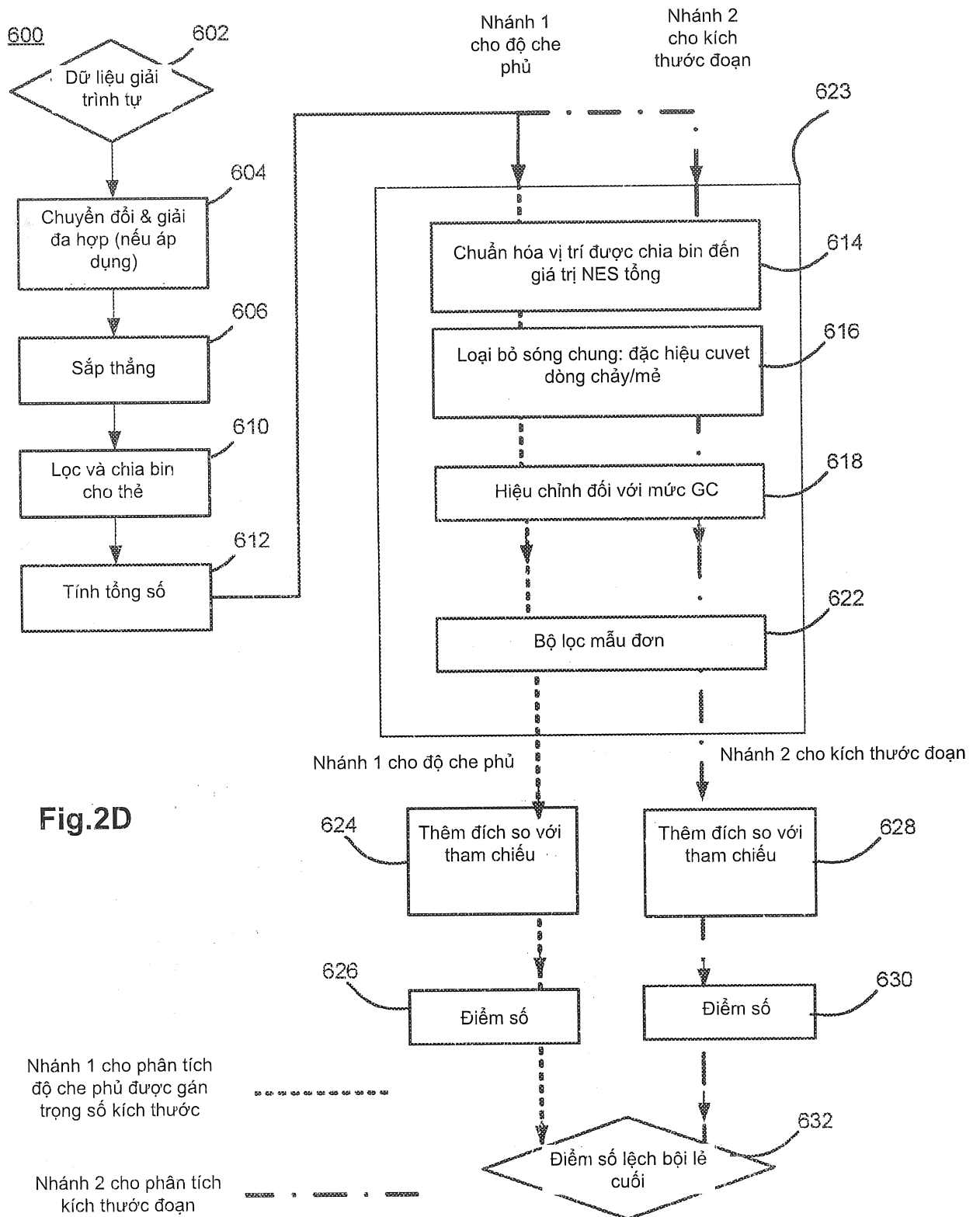


Fig.2C



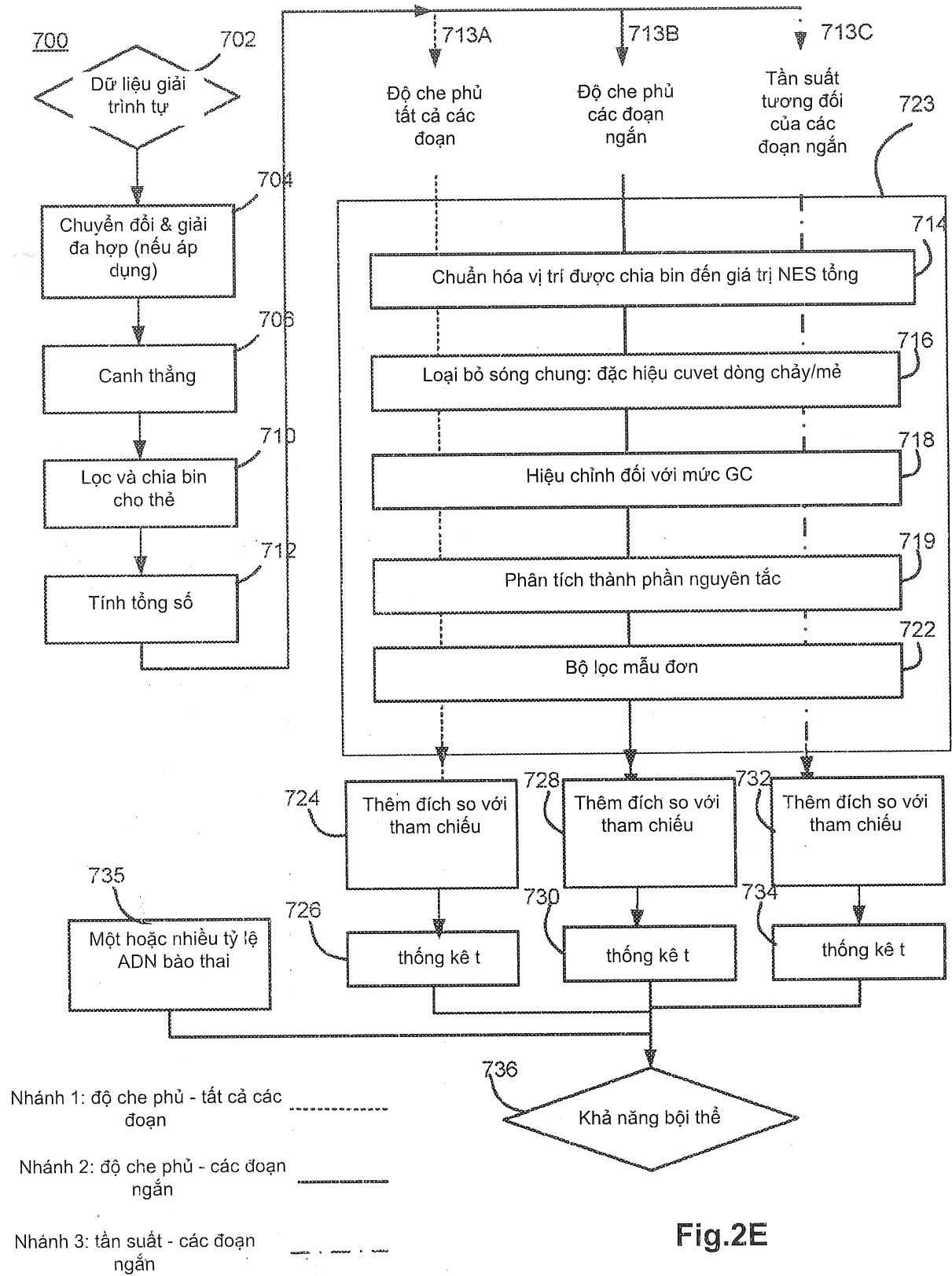
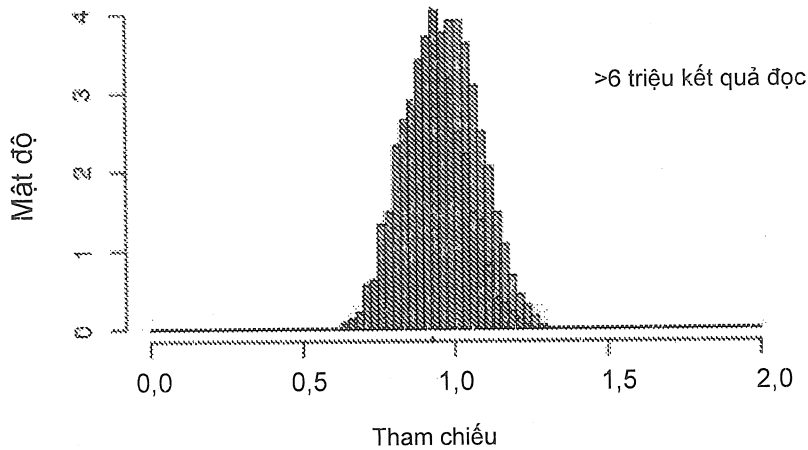
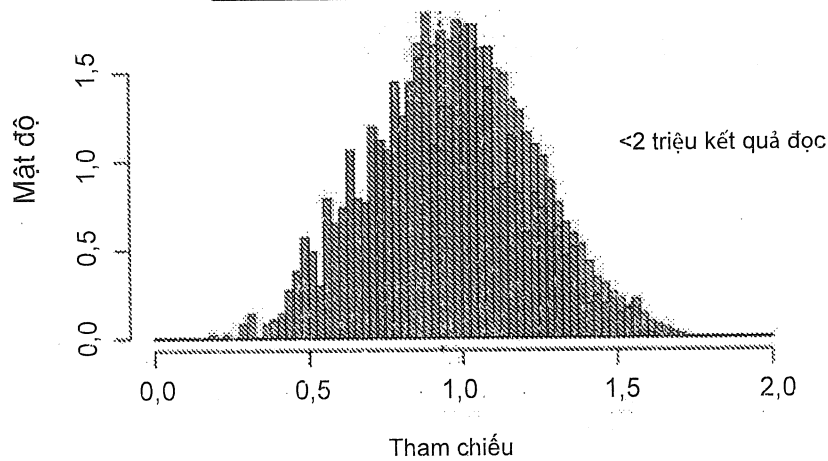


Fig.2E

Phân bố độ che phủ bin đối với mẫu độ che phủ cao



Phân bố độ che phủ bin đối với mẫu độ che phủ thấp



$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

X_1 = độ che phủ bin của nhiễm sắc thể quan tâm

X_2 = độ che phủ bin của vùng tham chiếu

Fig.2F

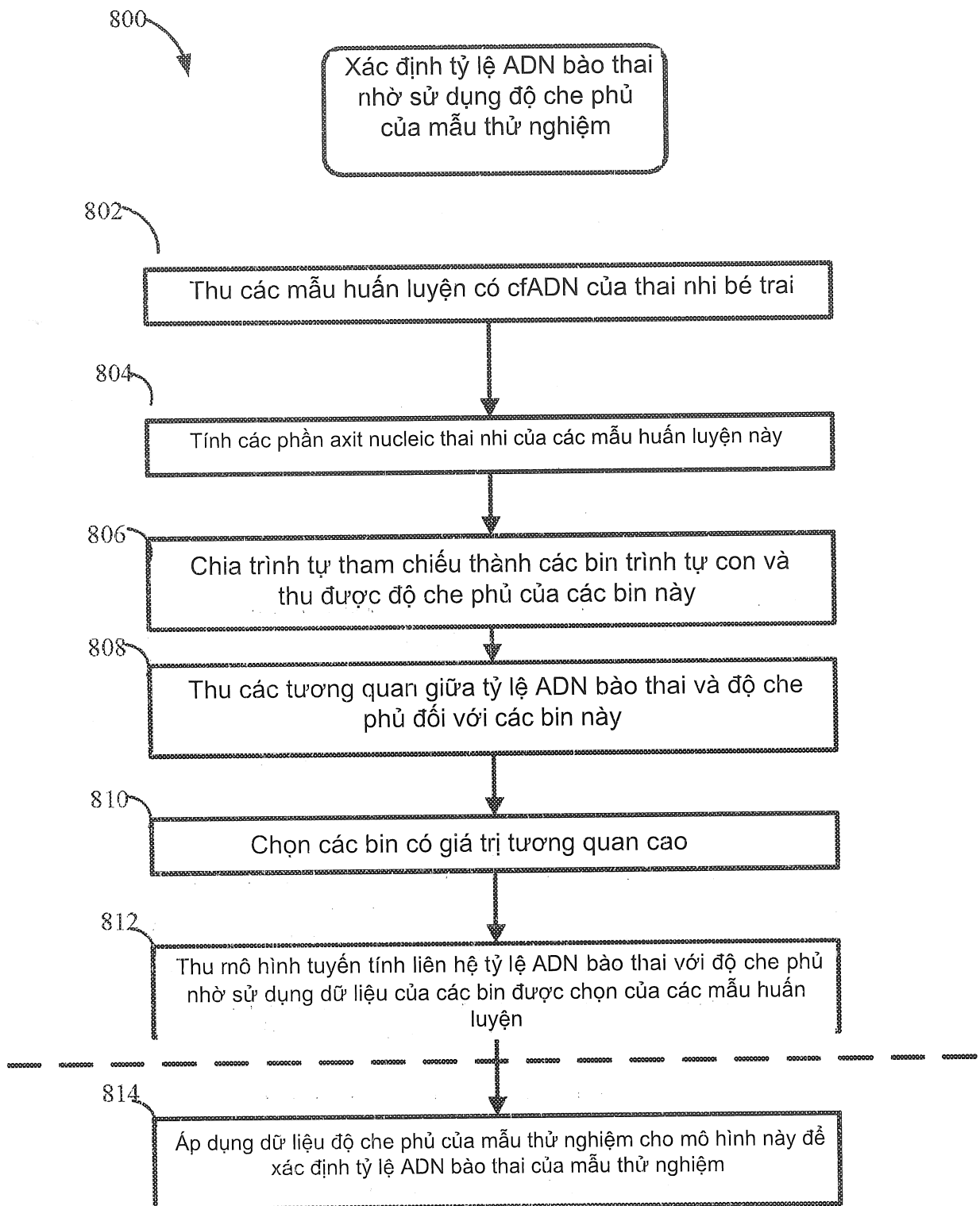


Fig.2G

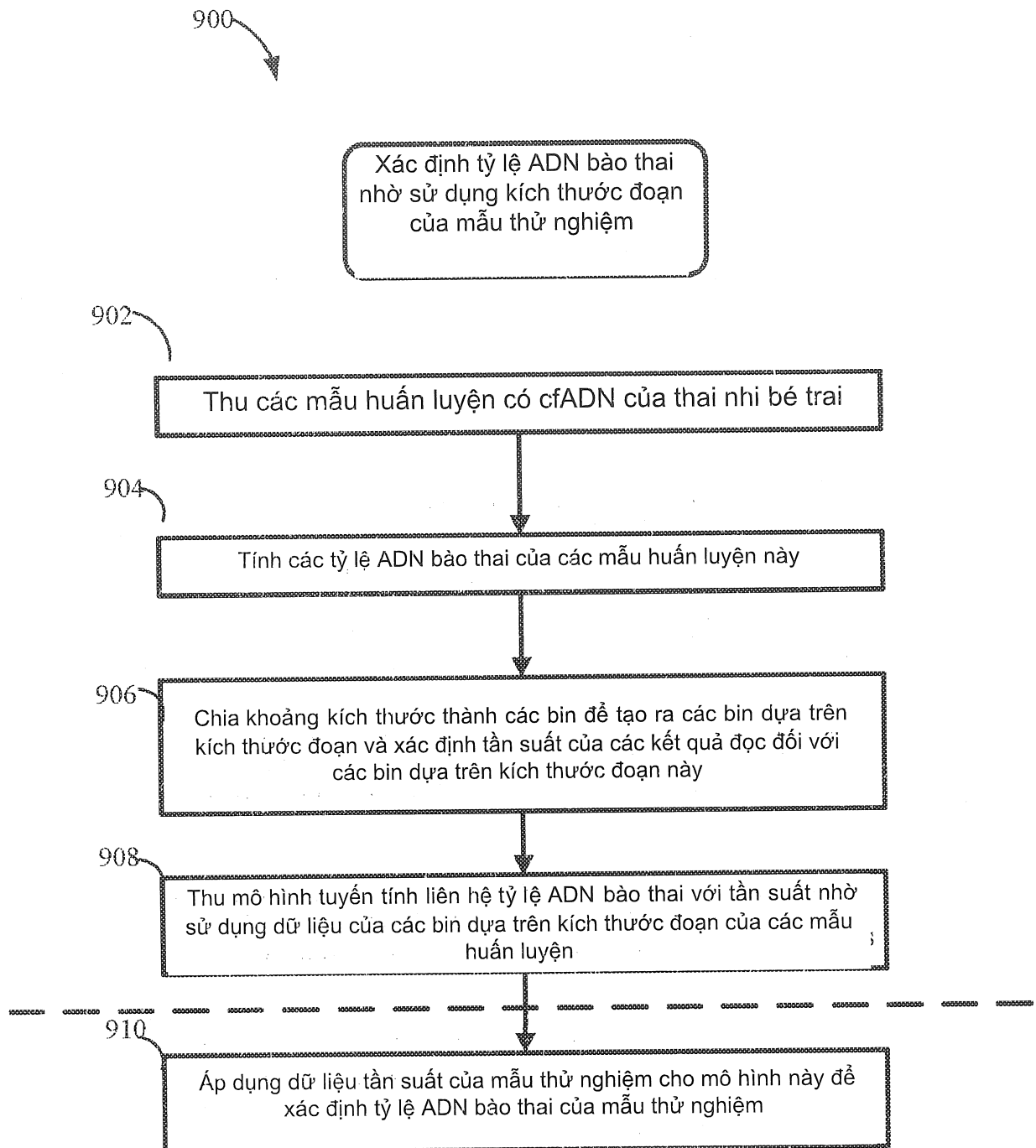


Fig.2H

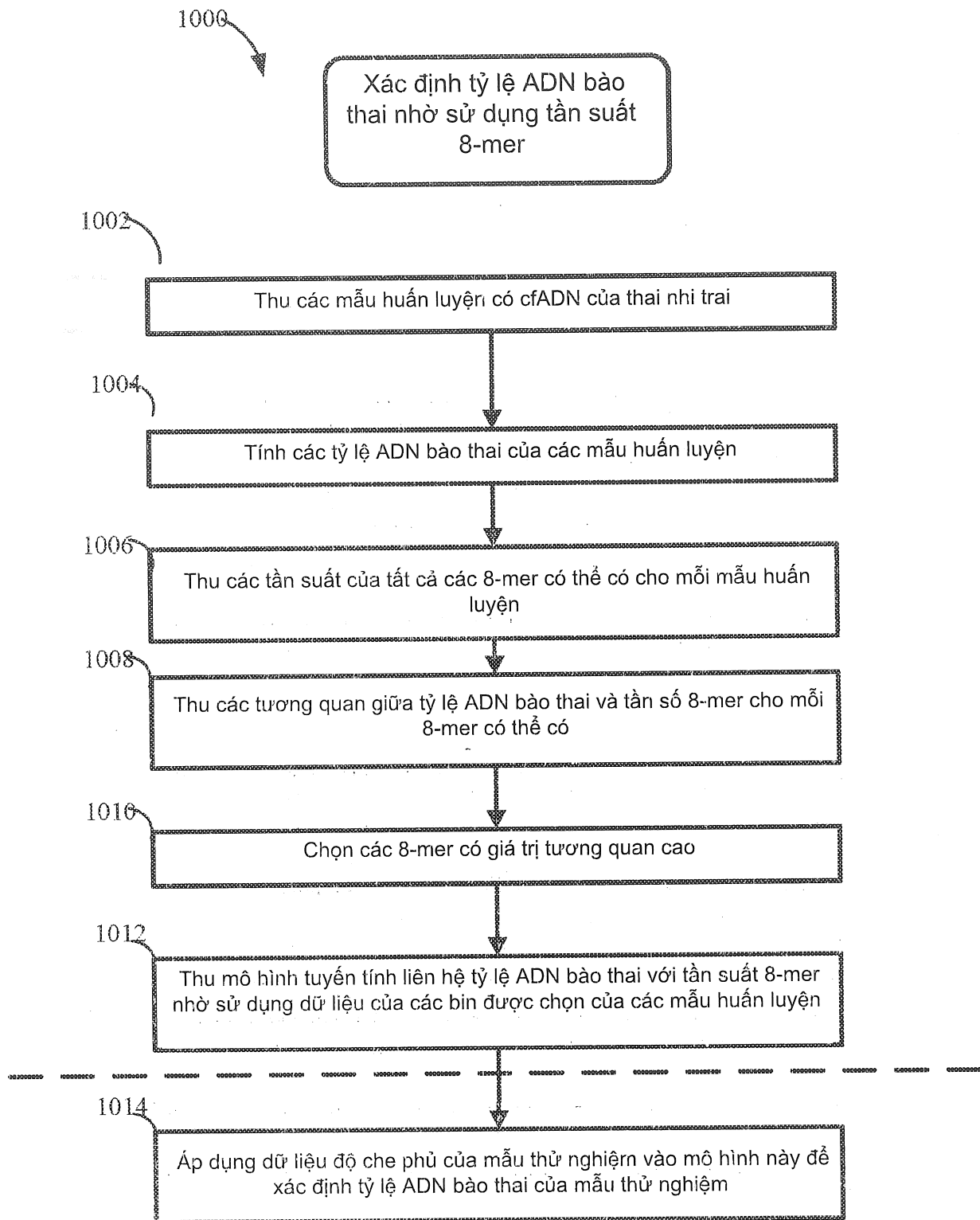


Fig.2I

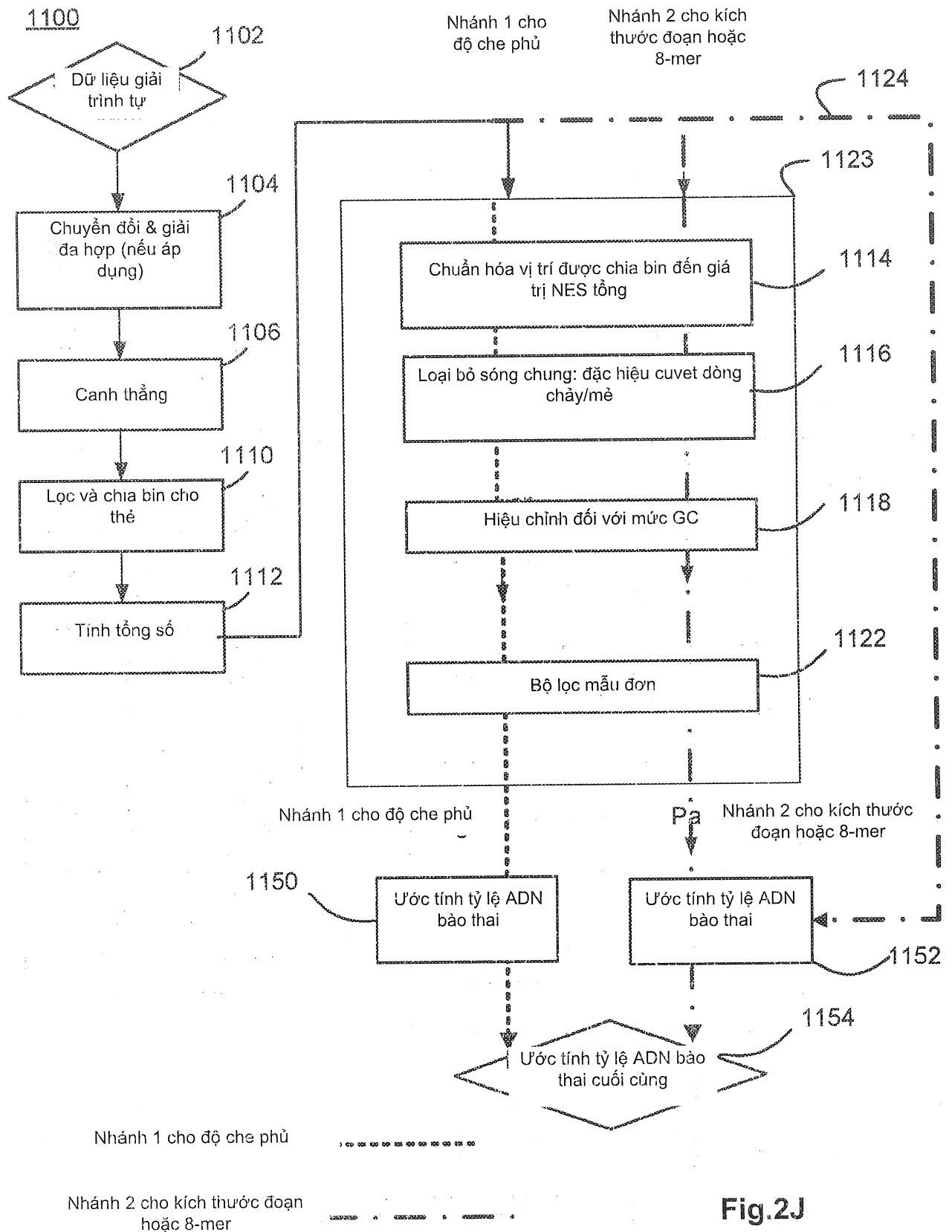


Fig.2J

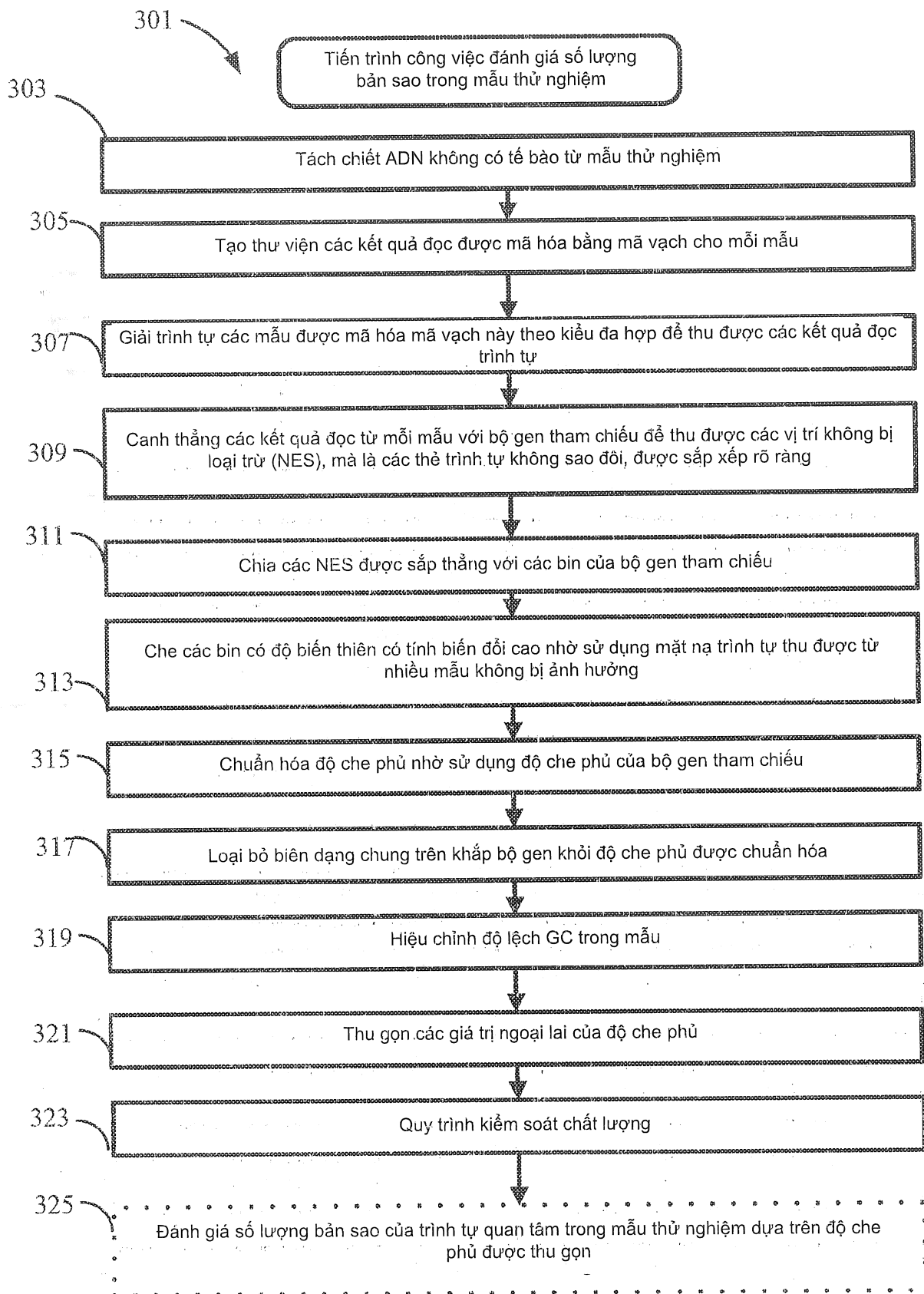


Fig.3A

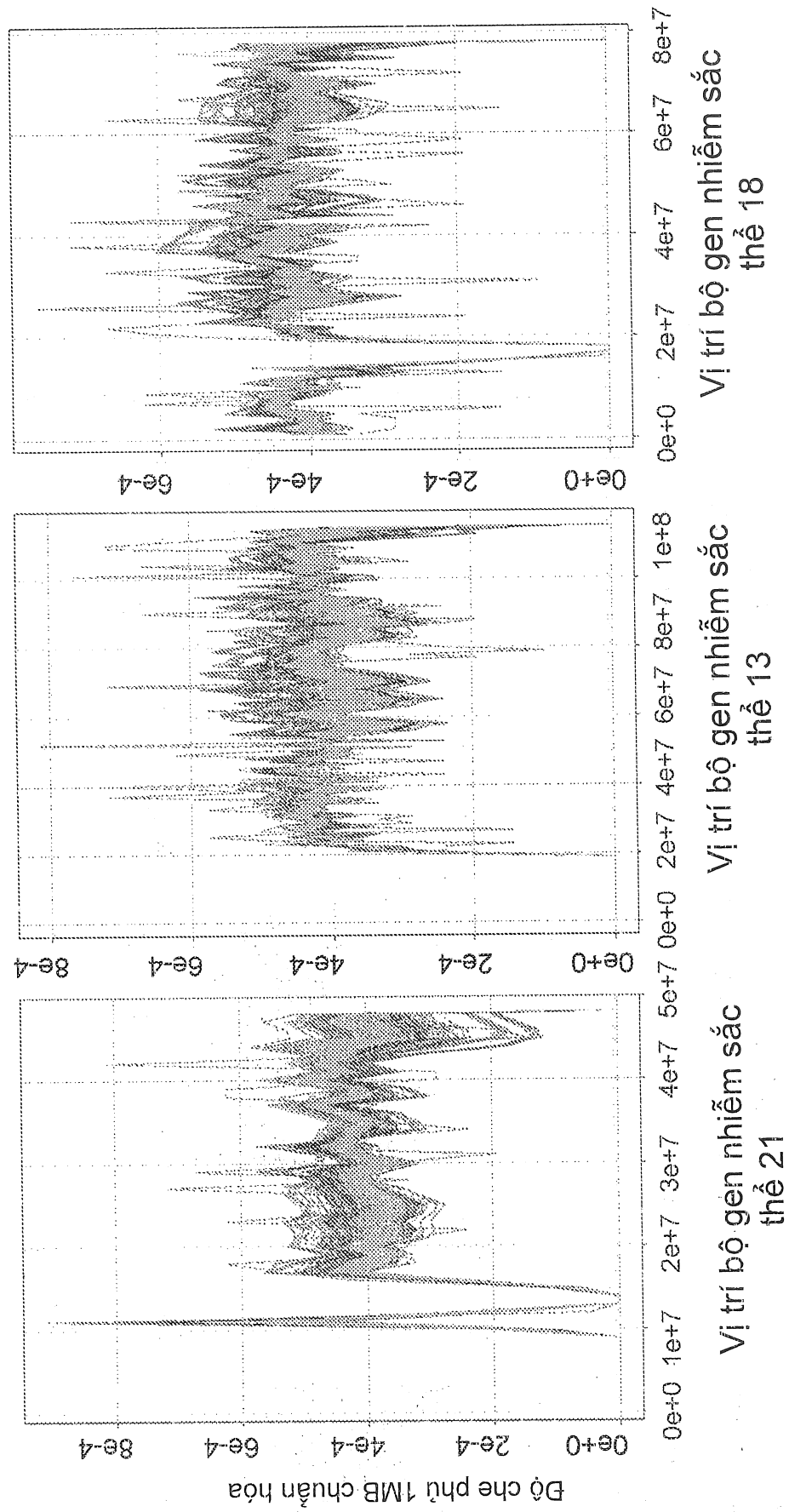


Fig.3B

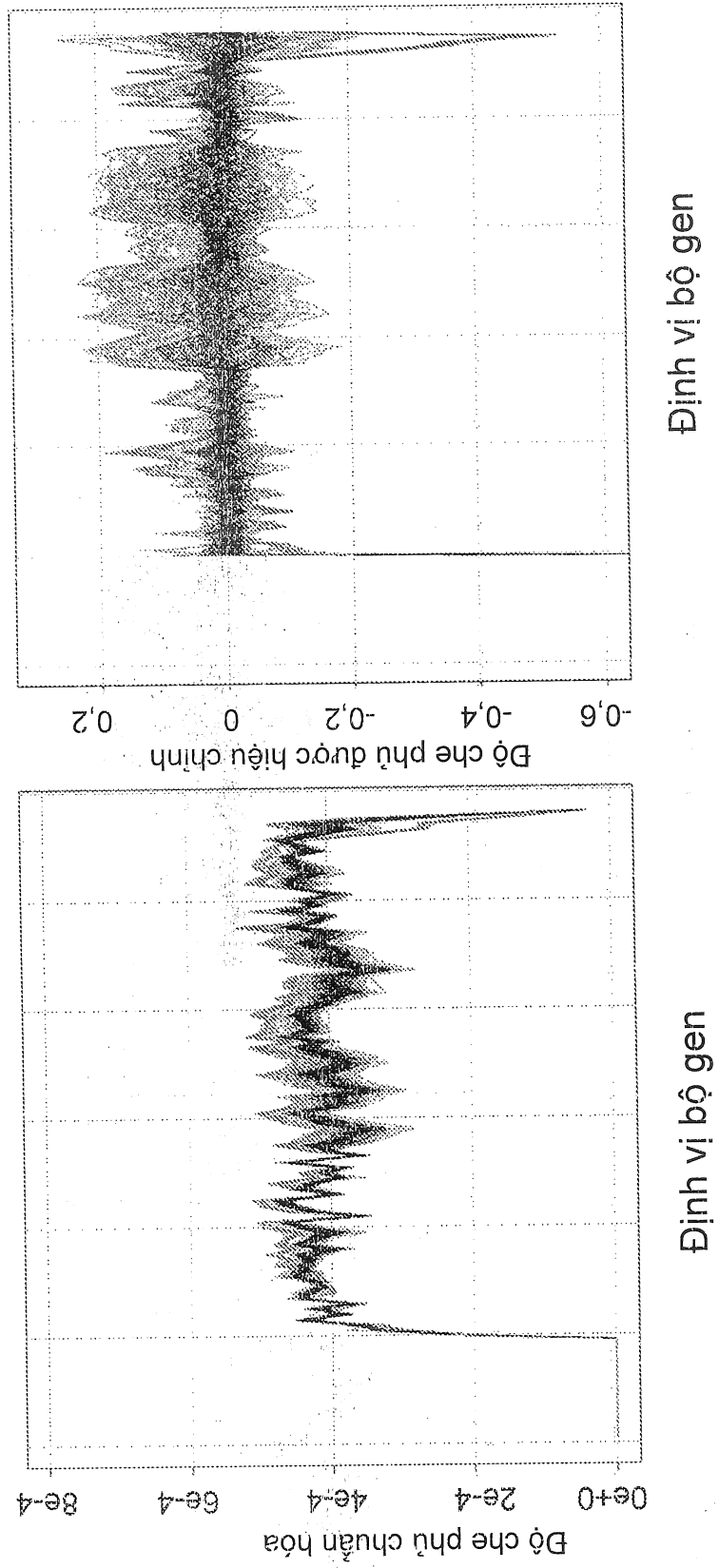


Fig.3C

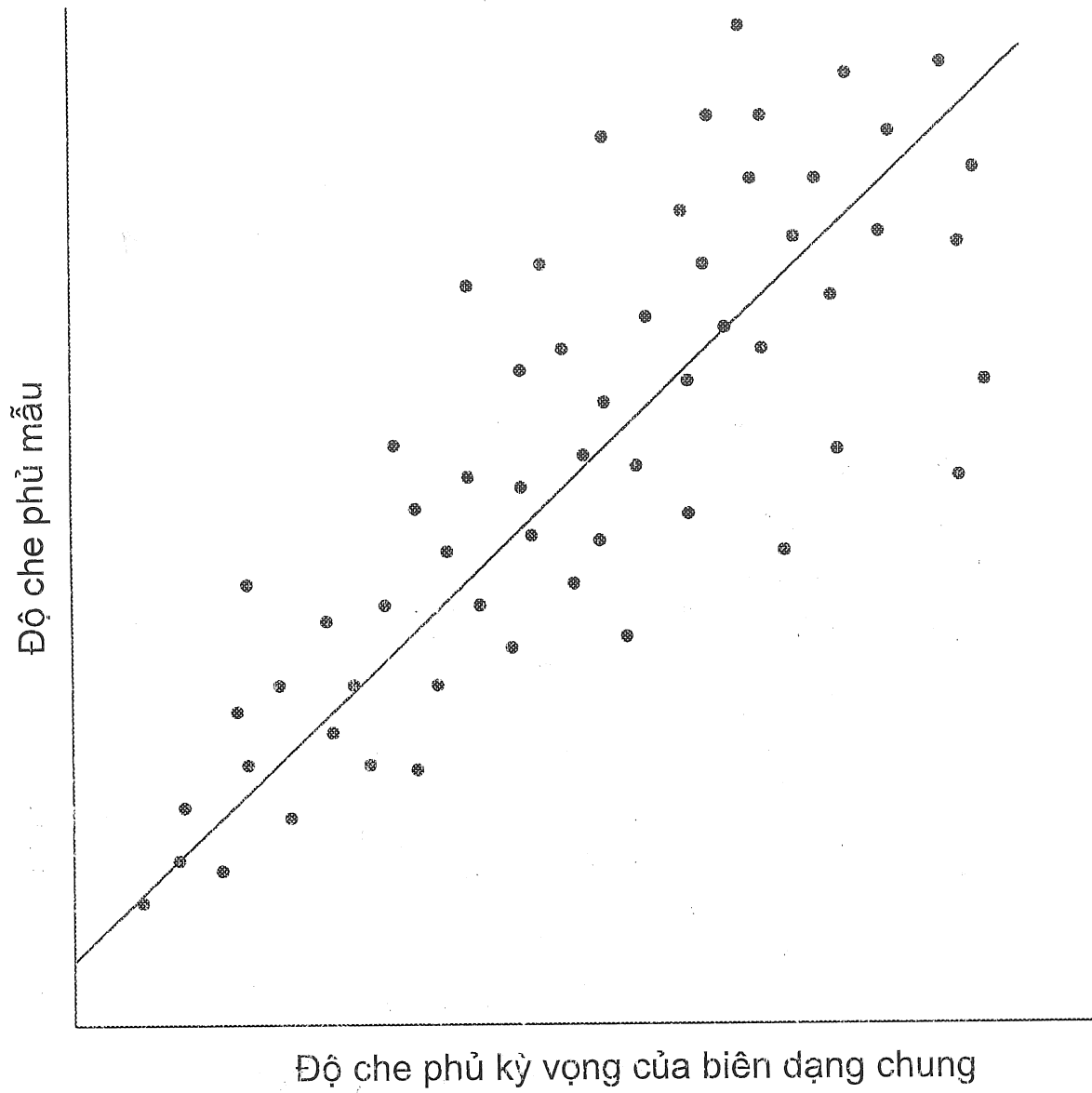


Fig.3D

HiSeqTrFC | C204CACXX | CLIA233

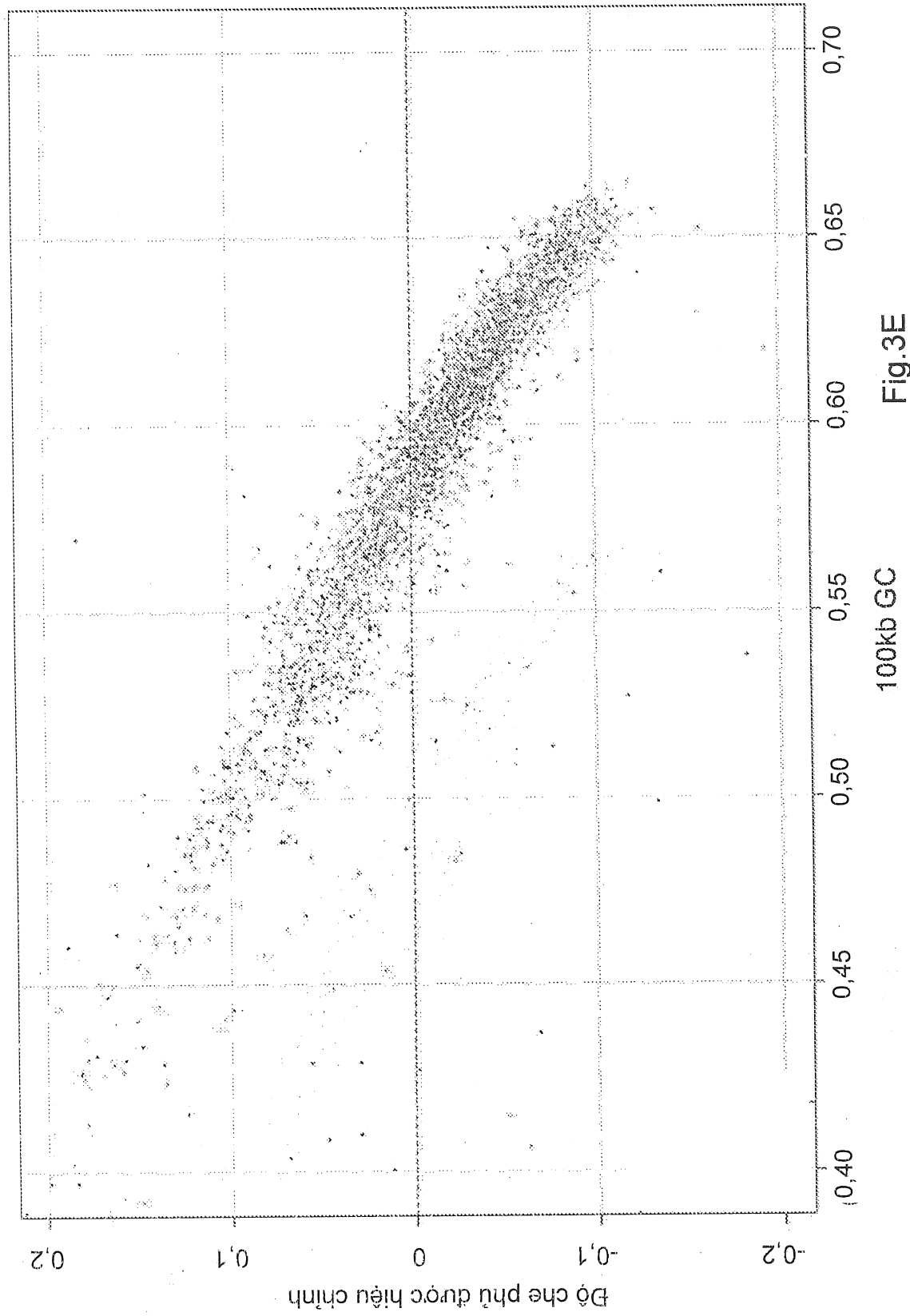
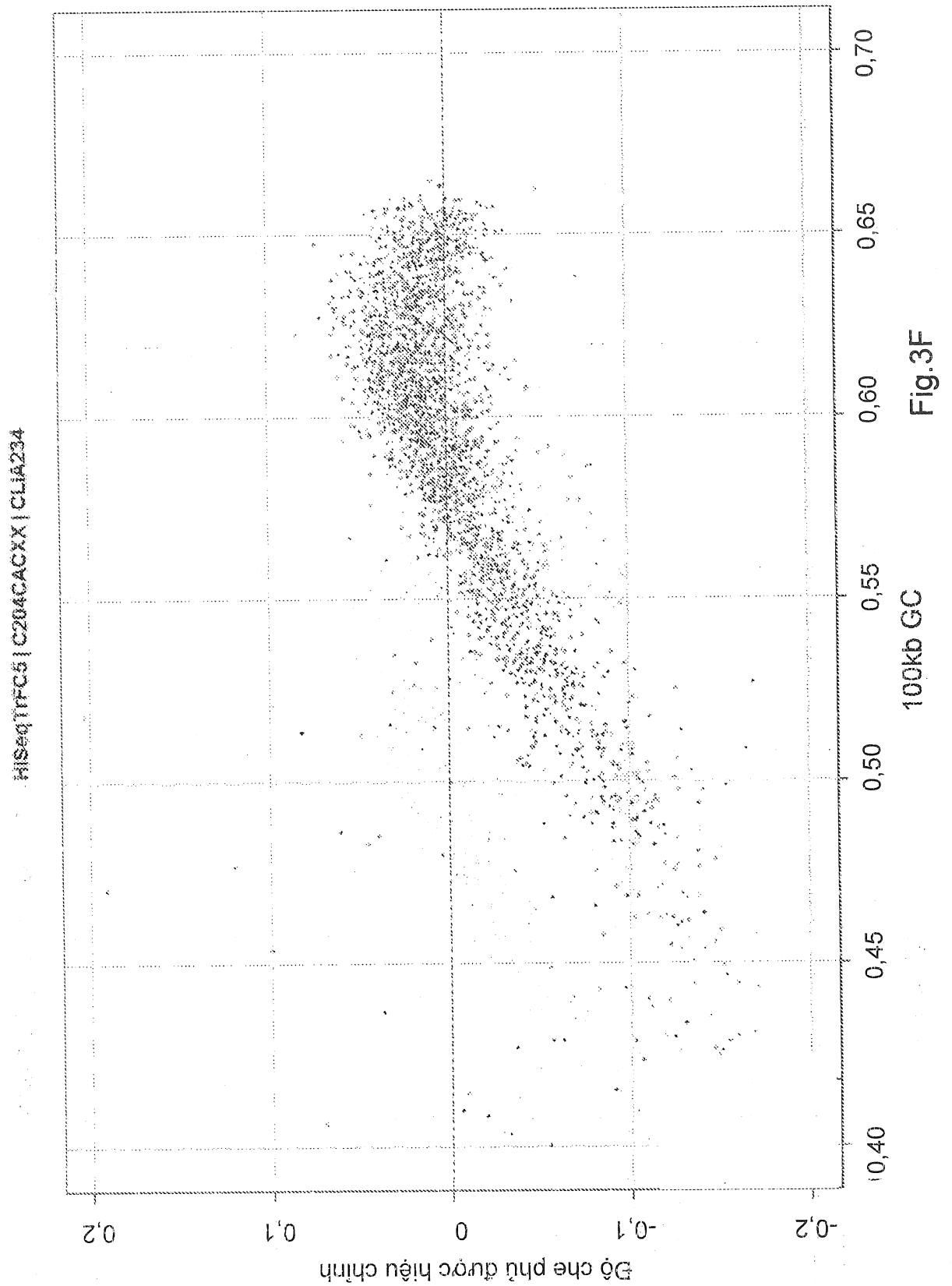
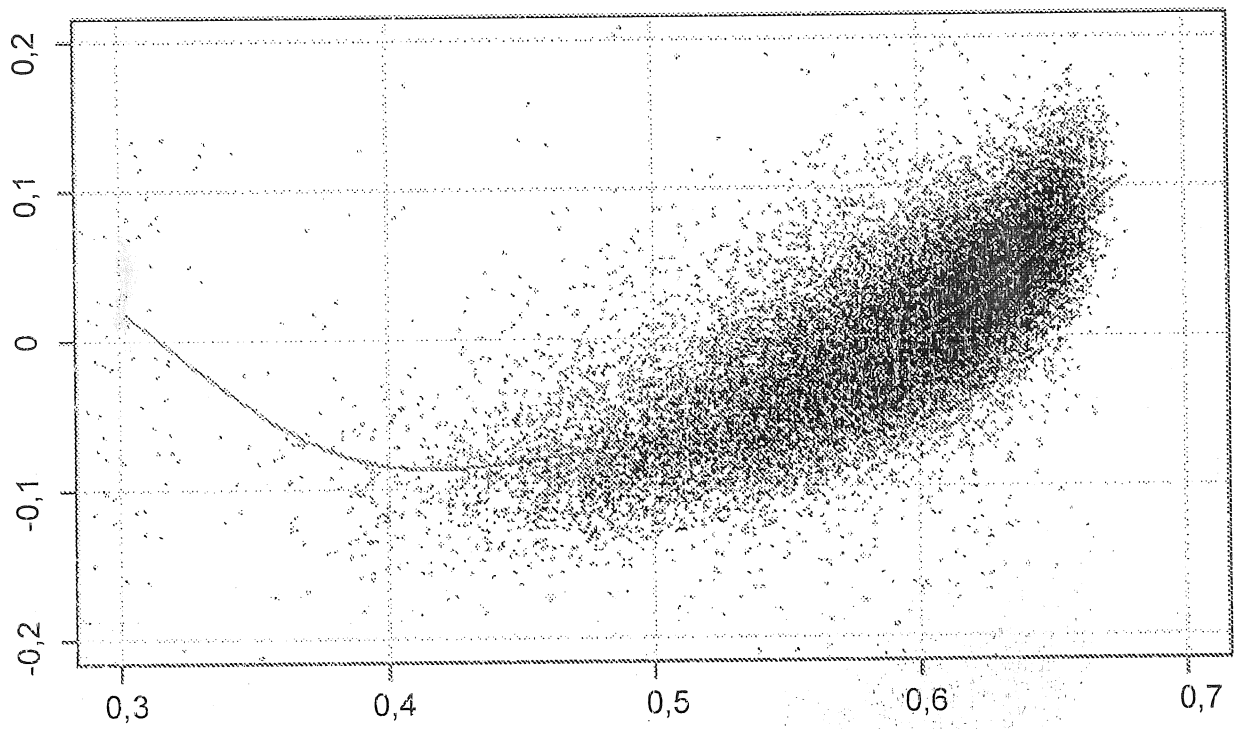


Fig.3E



Phân bố so với gc% trước hiệu chỉnh



Phân bố so với gc% sau hiệu chỉnh

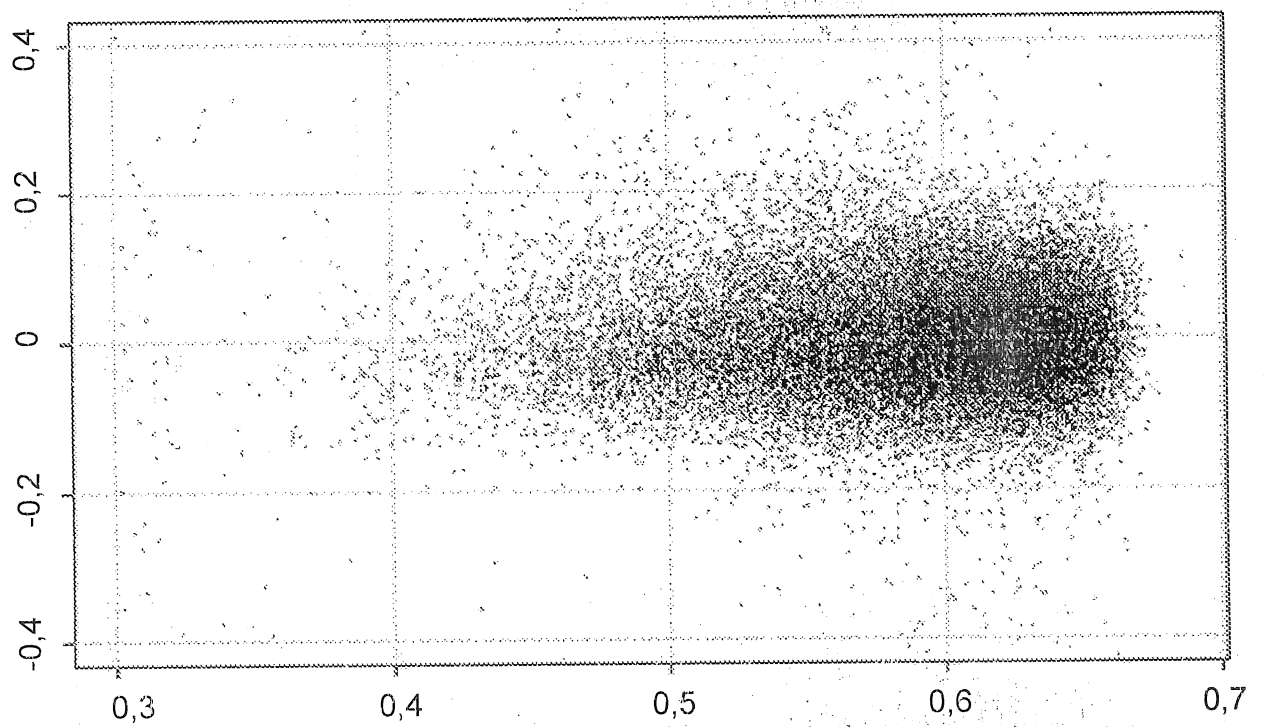


Fig.3G

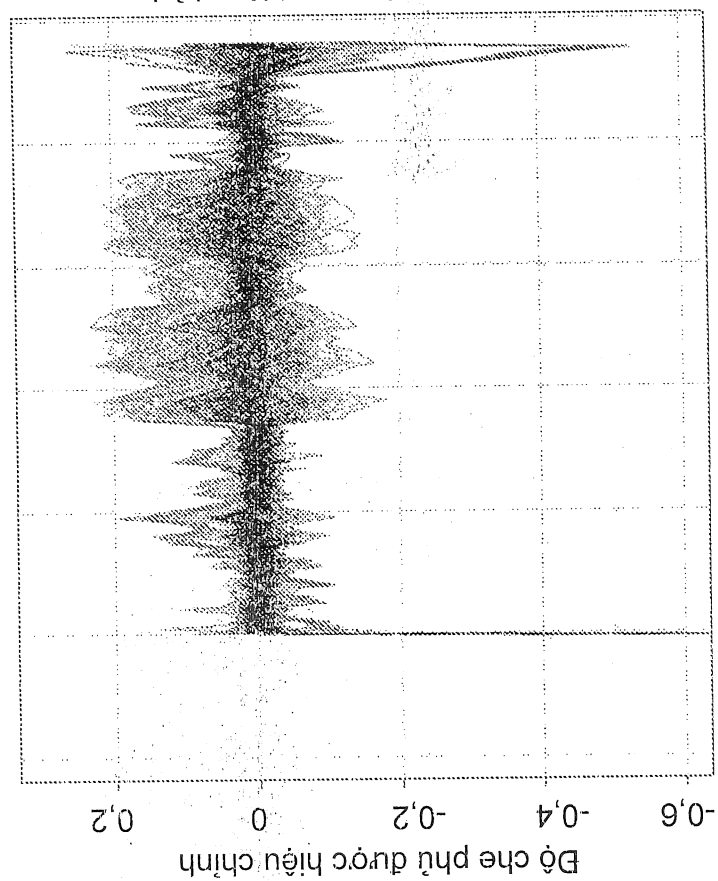
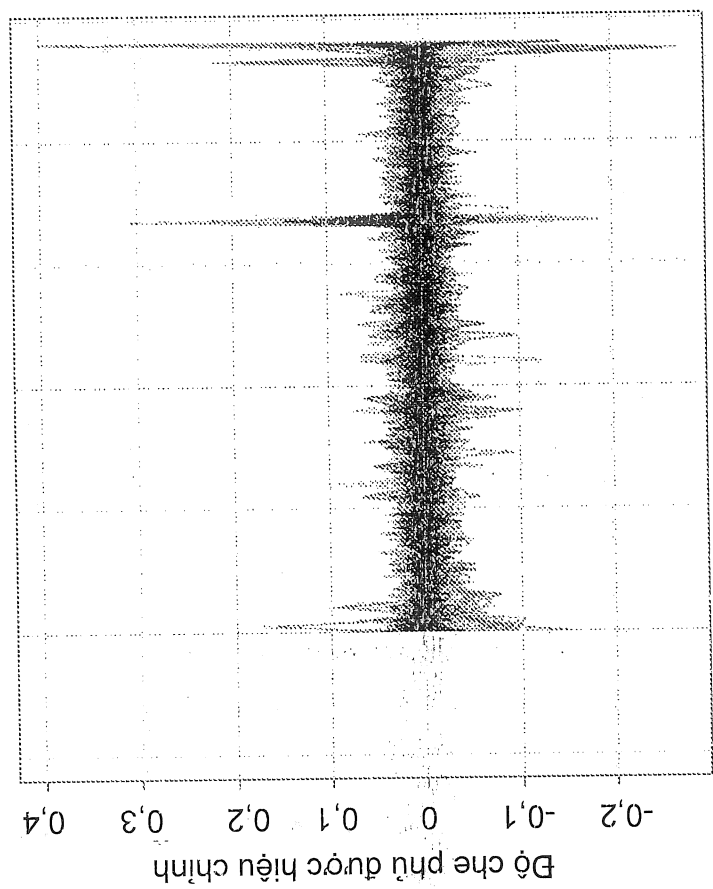


Fig.3H

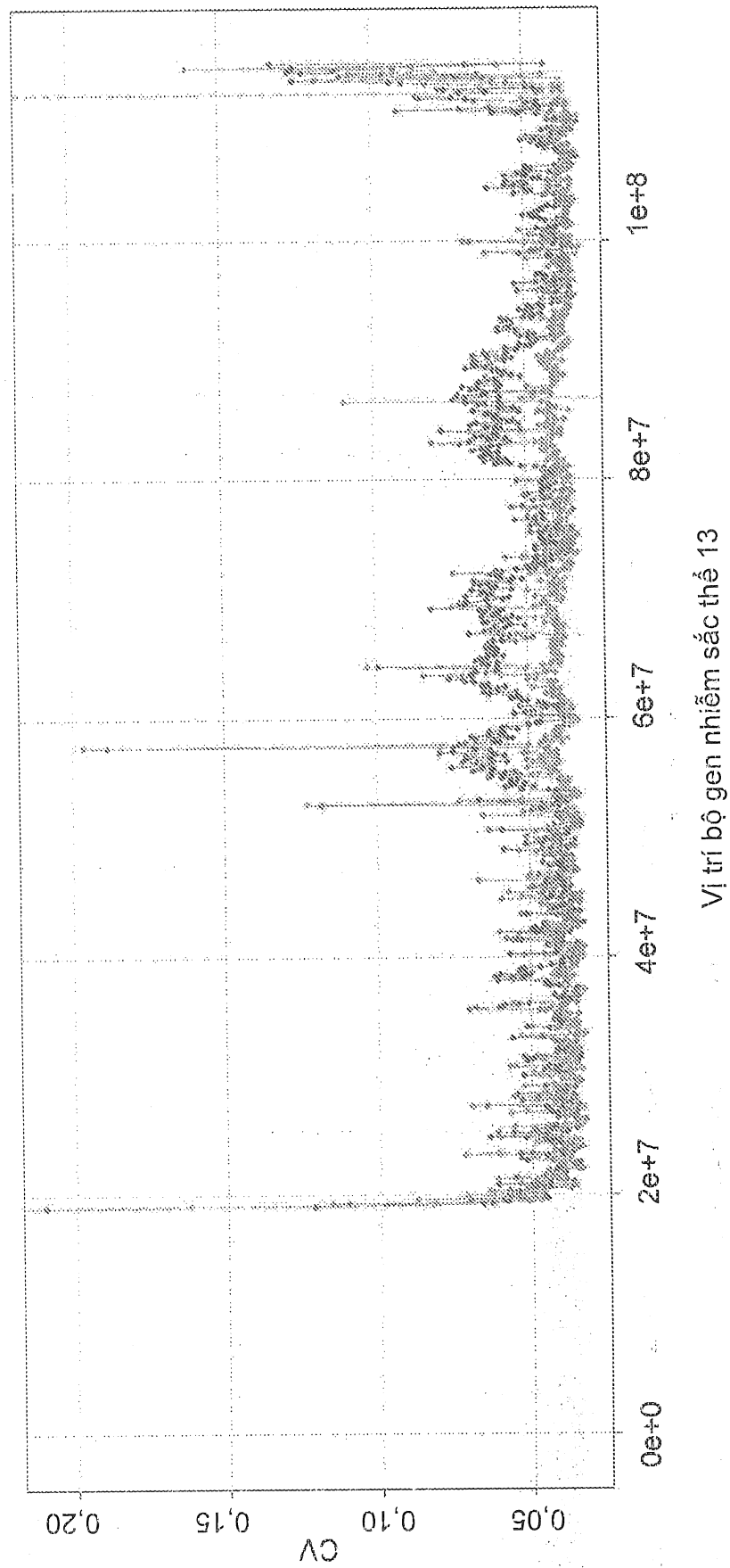


Fig.3I

D2F7KACXX | 320573 | chr12

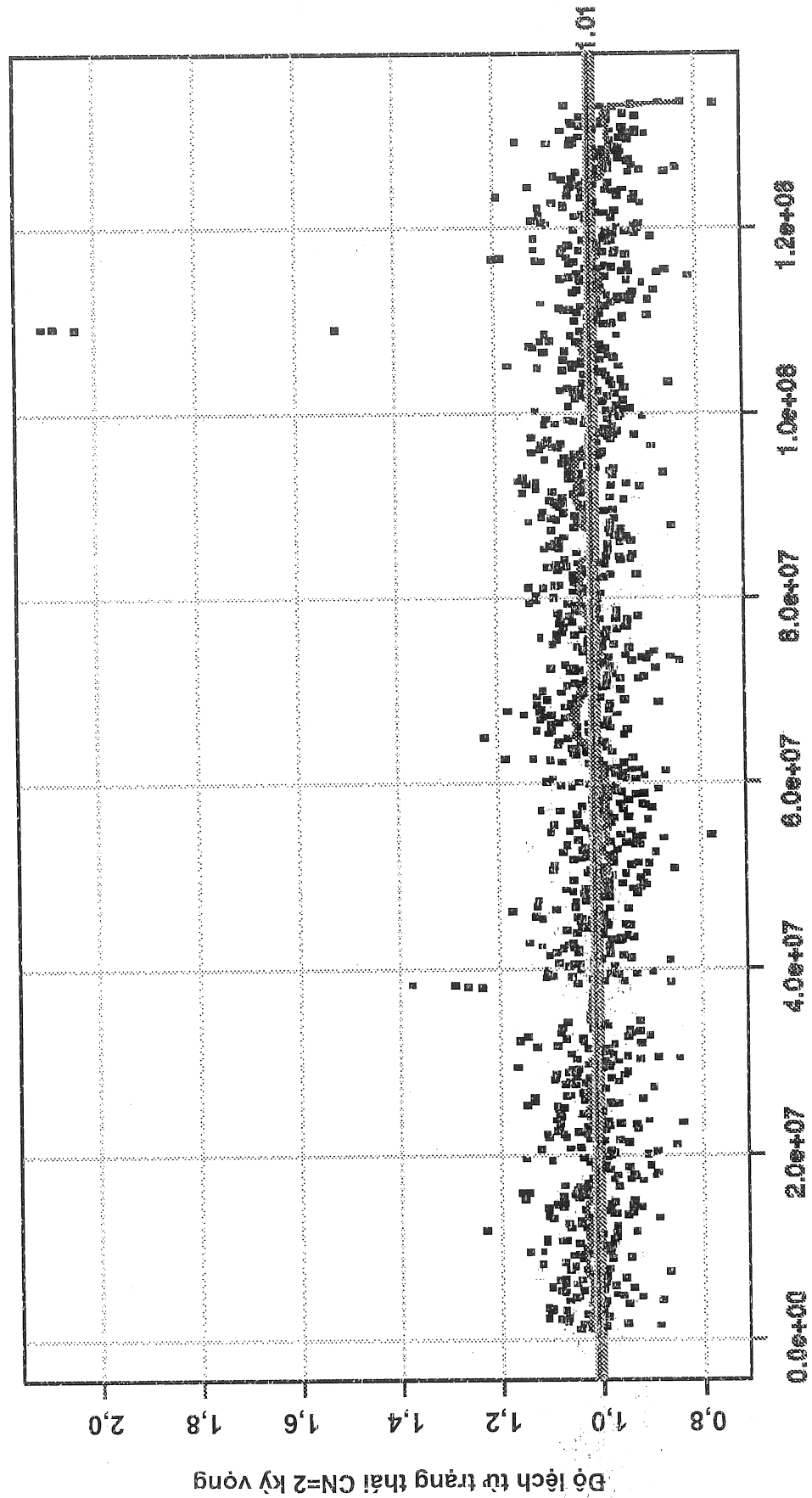


Fig.3J

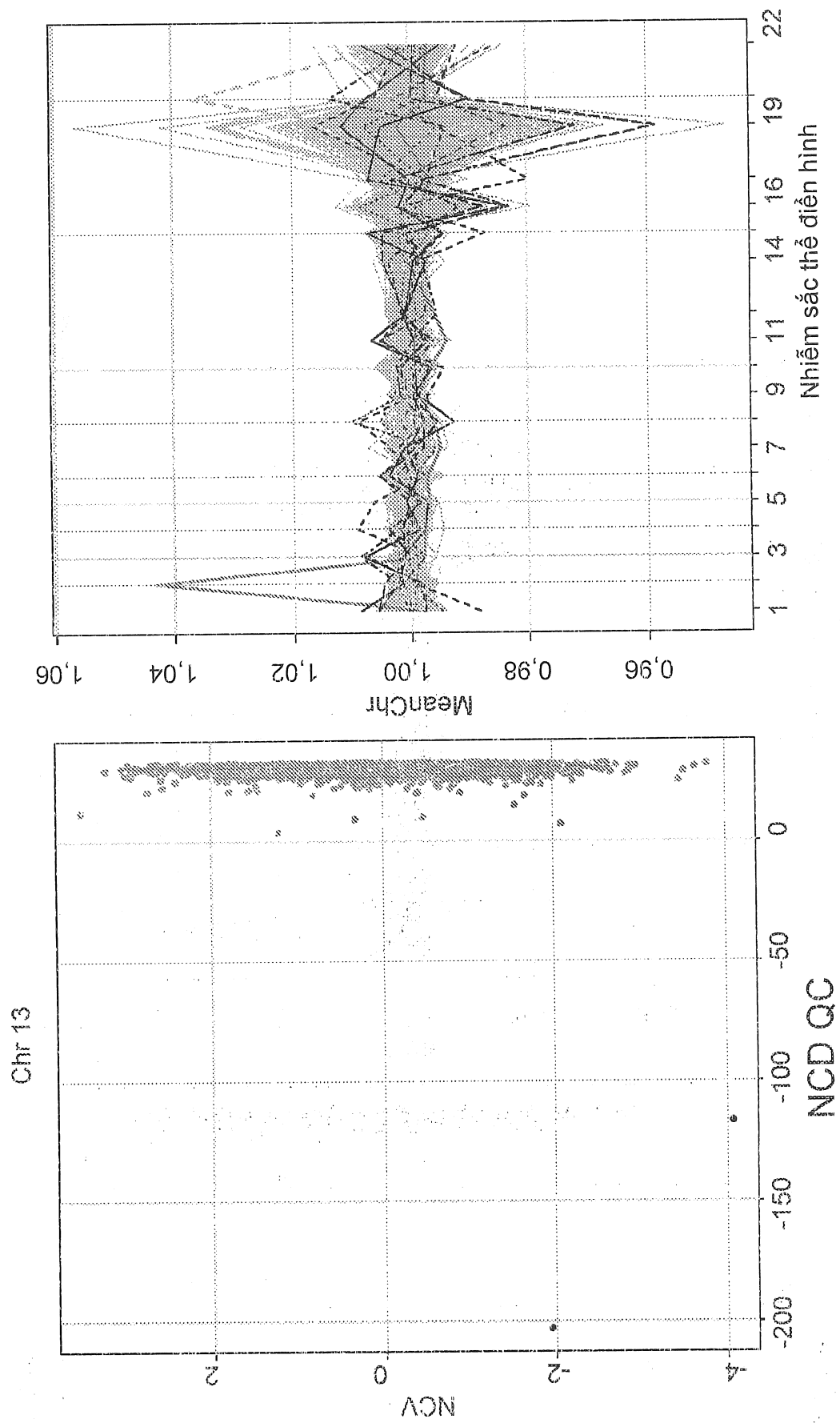


Fig.3K

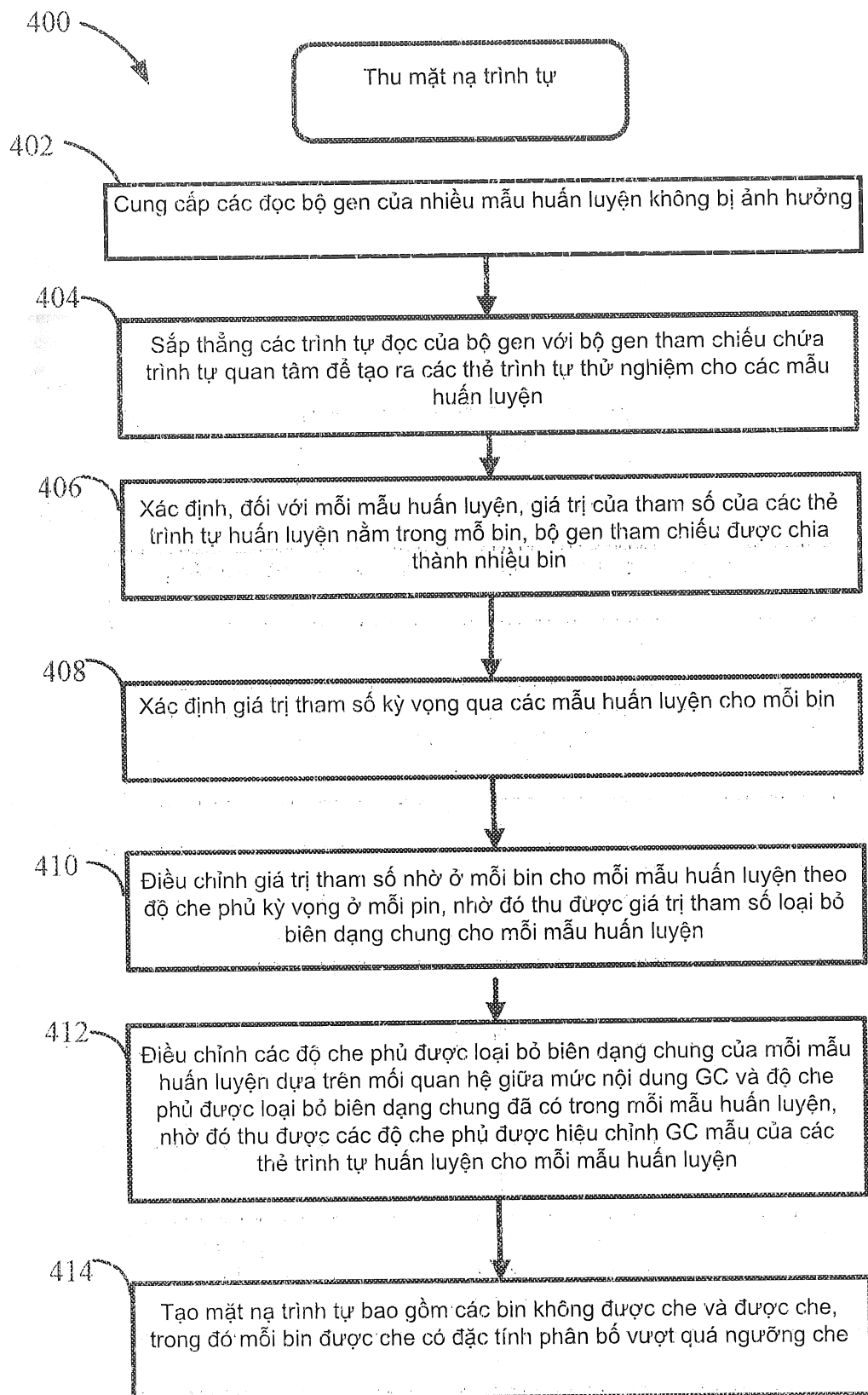


Fig.4A

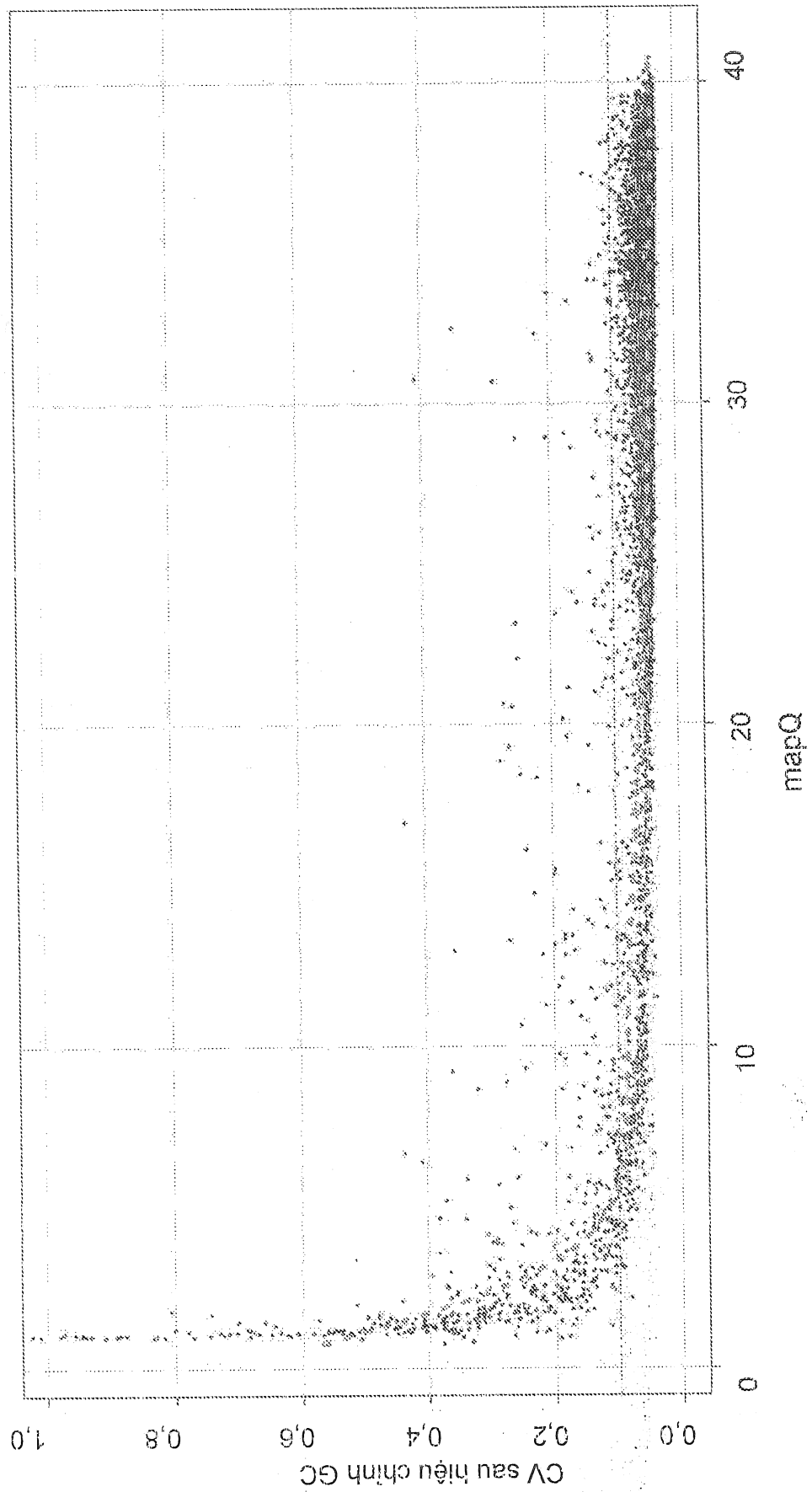


Fig.4B

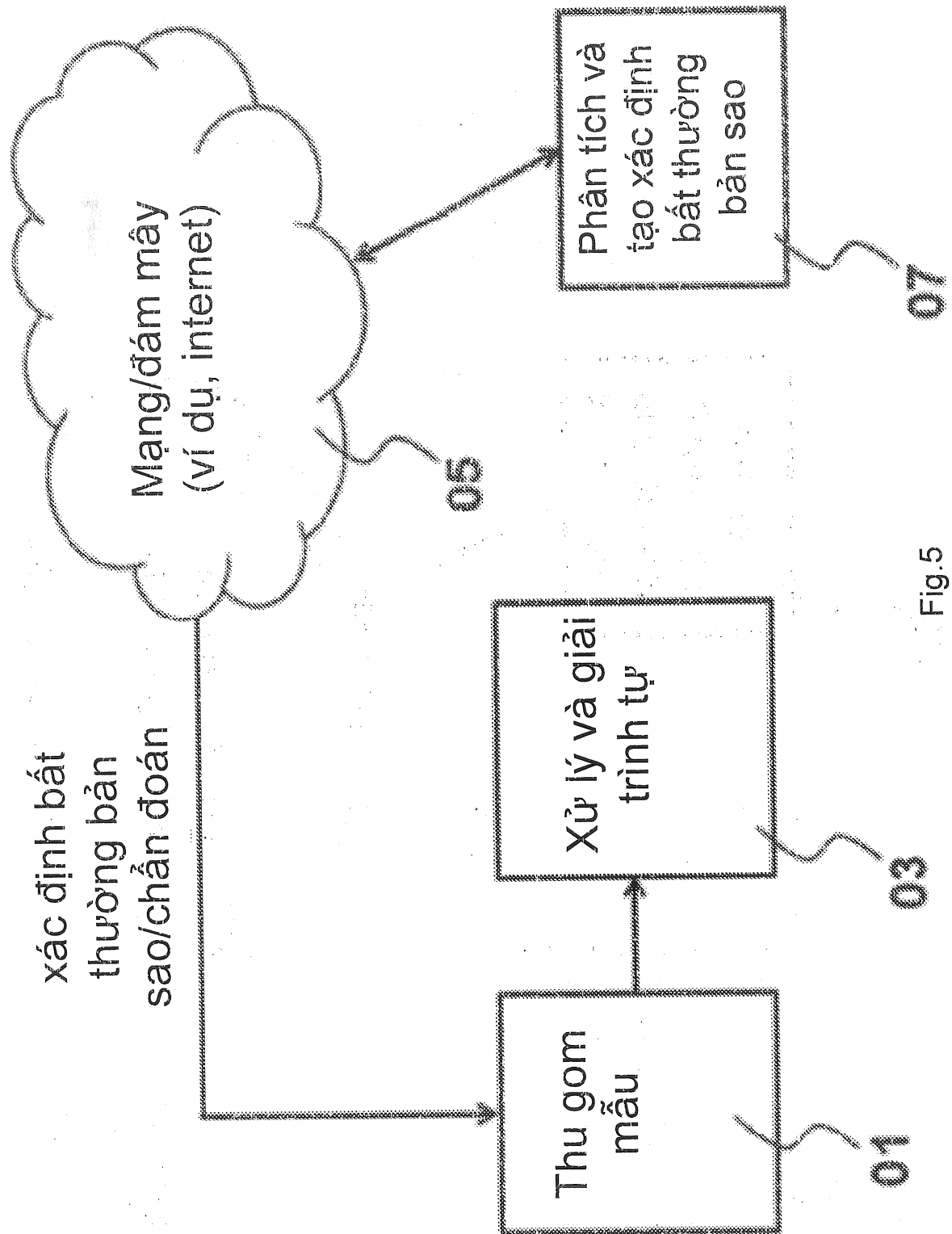


Fig.5

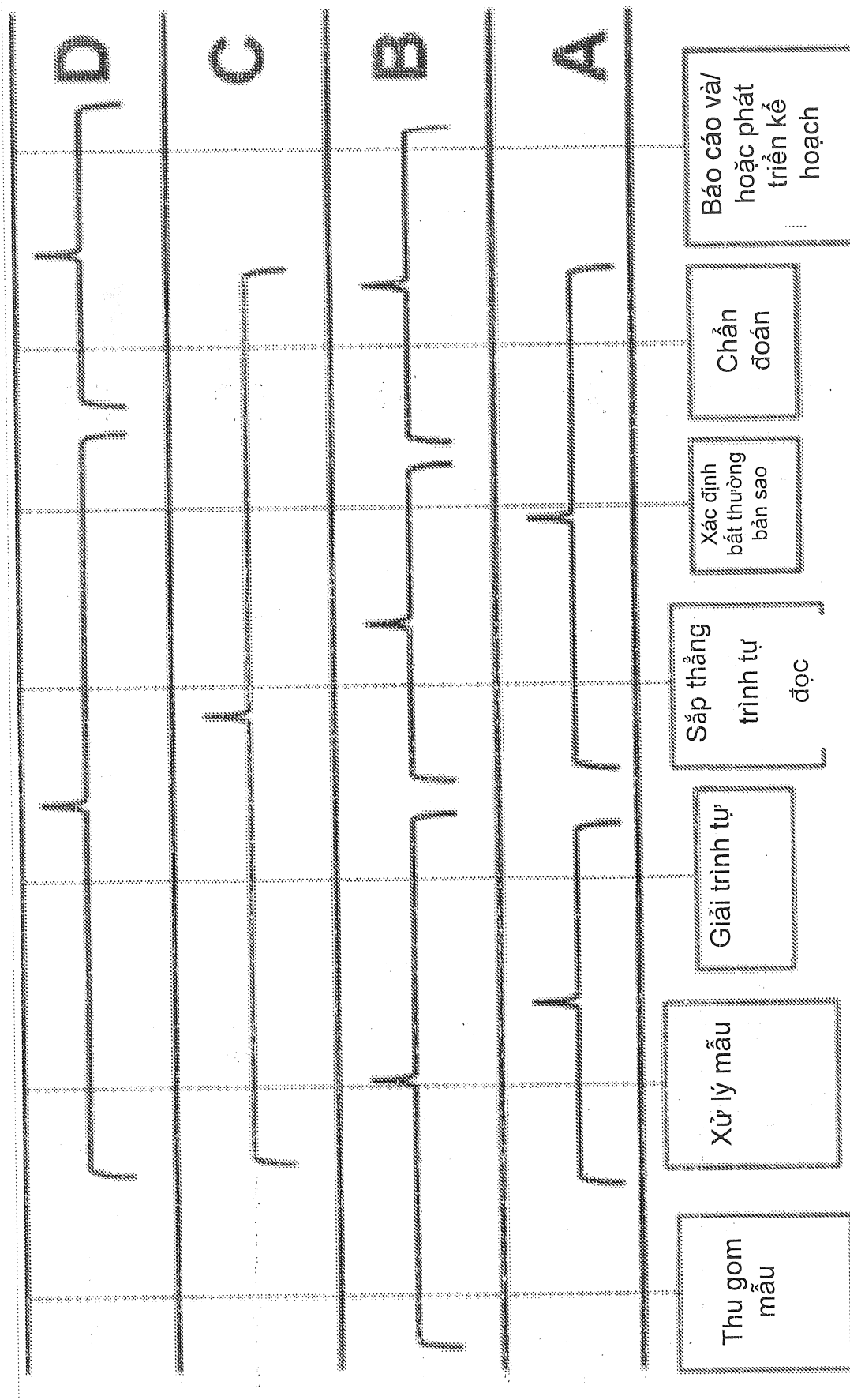


Fig.6

41317

27/33

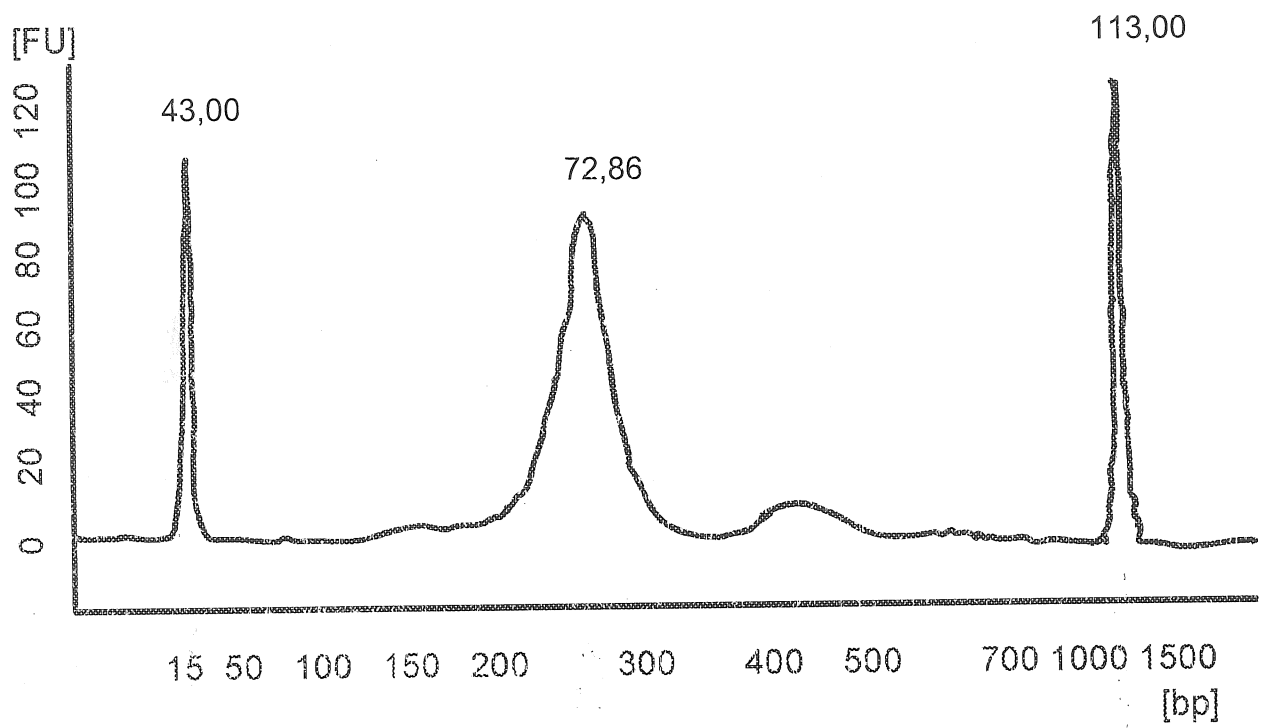


Fig.7A

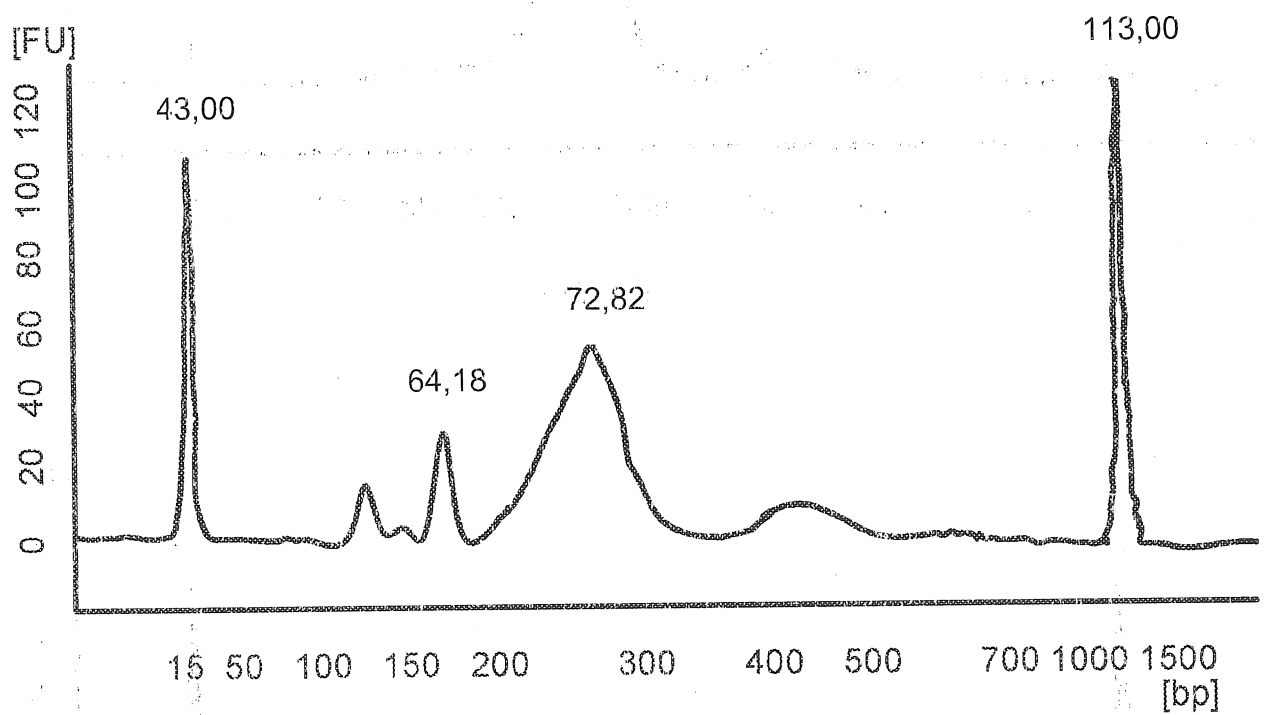


Fig.7B

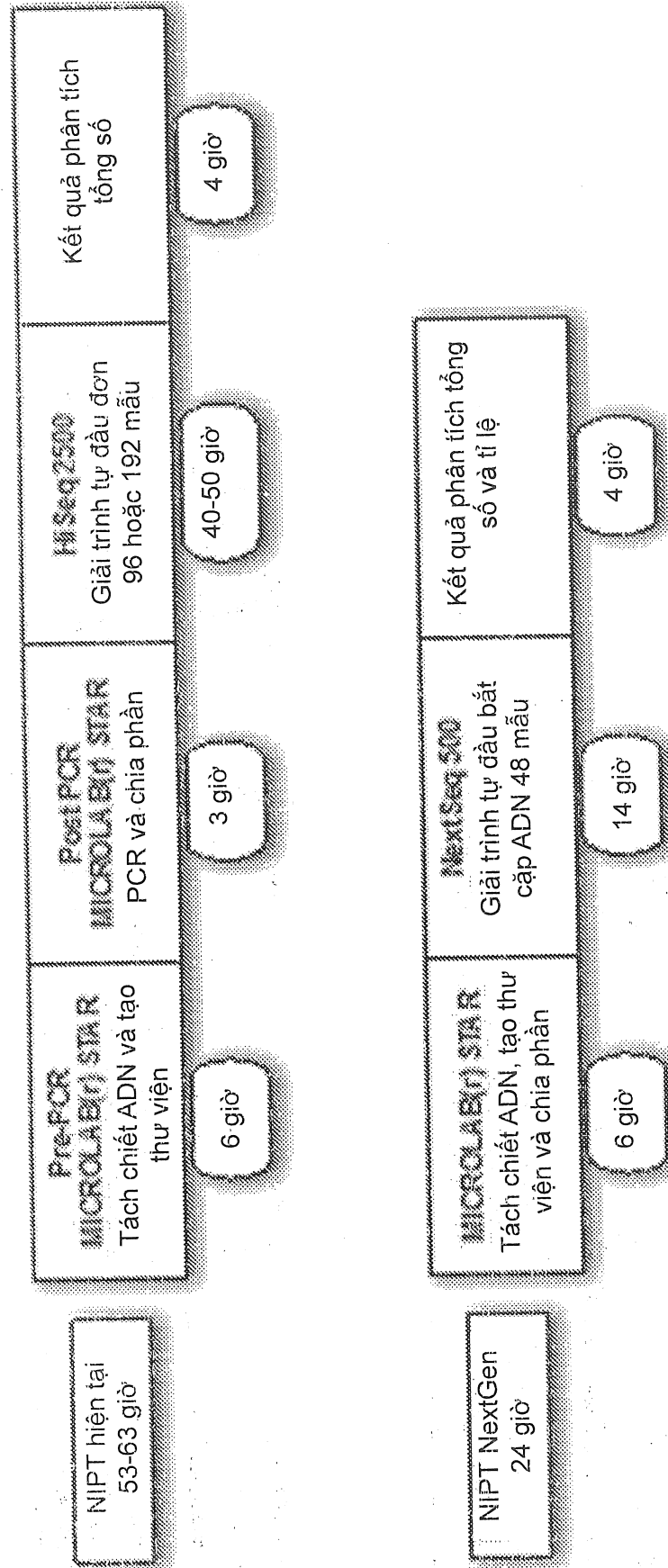


Fig.8

Hiệu suất thư viện

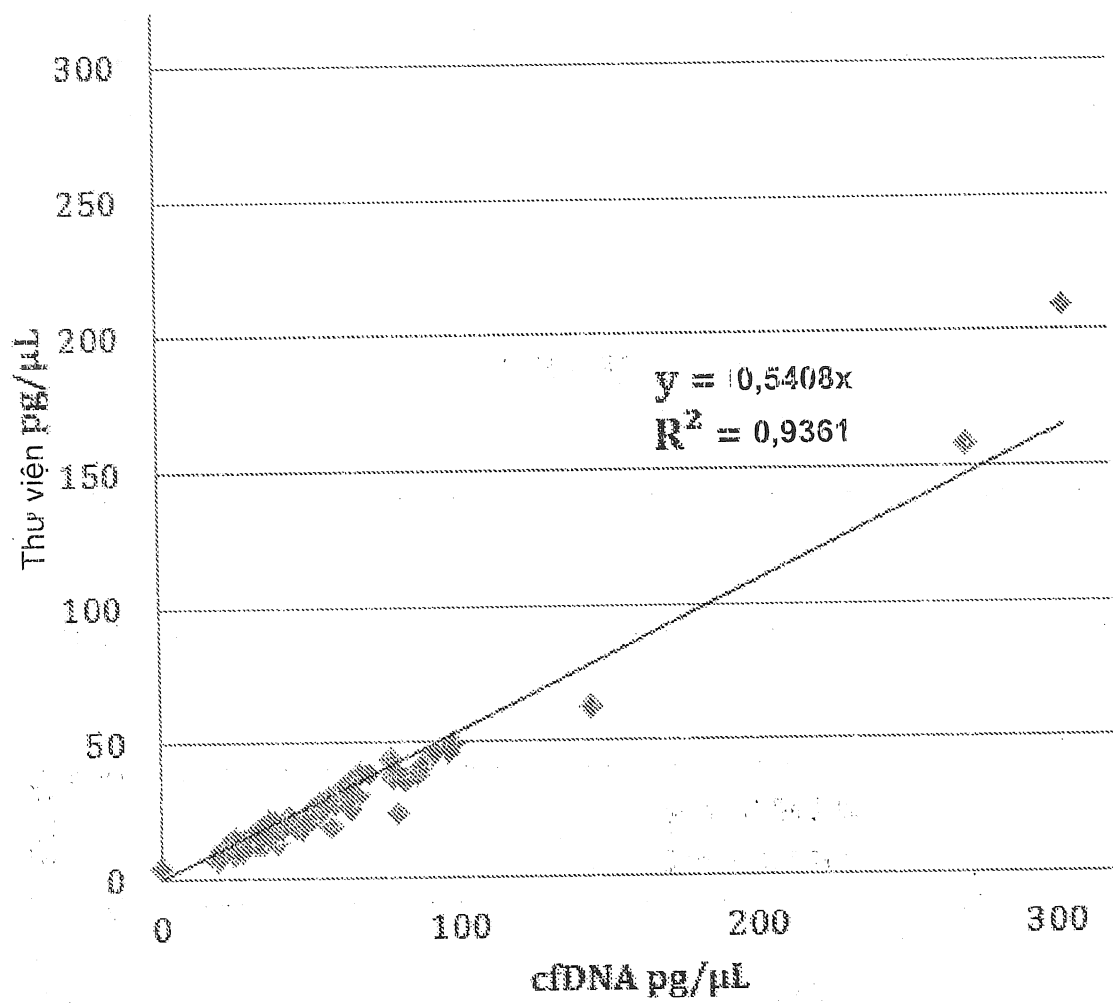


Fig.9

So sánh phân bố kích thước của người mẹ và thai nhi

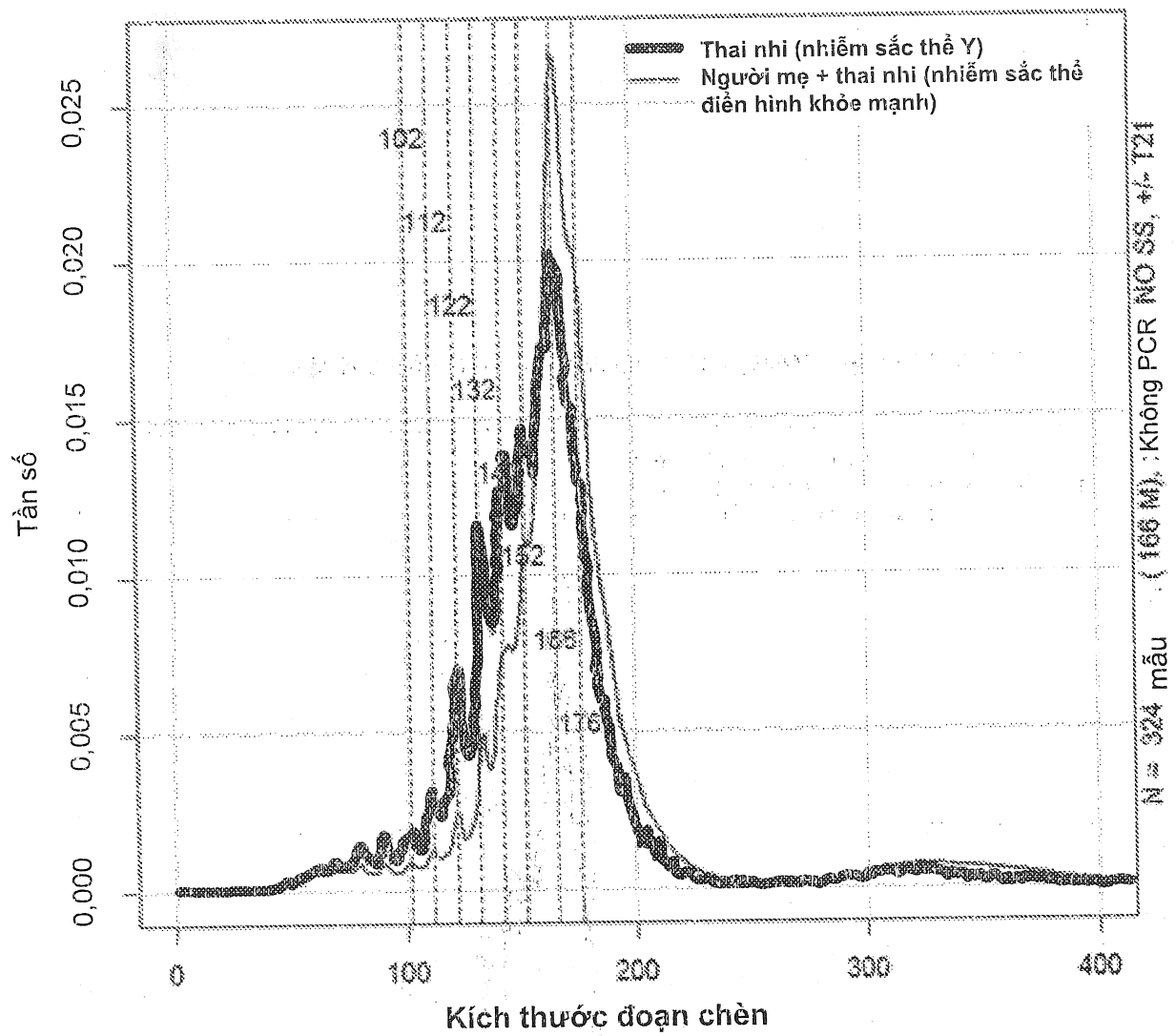


Fig.10

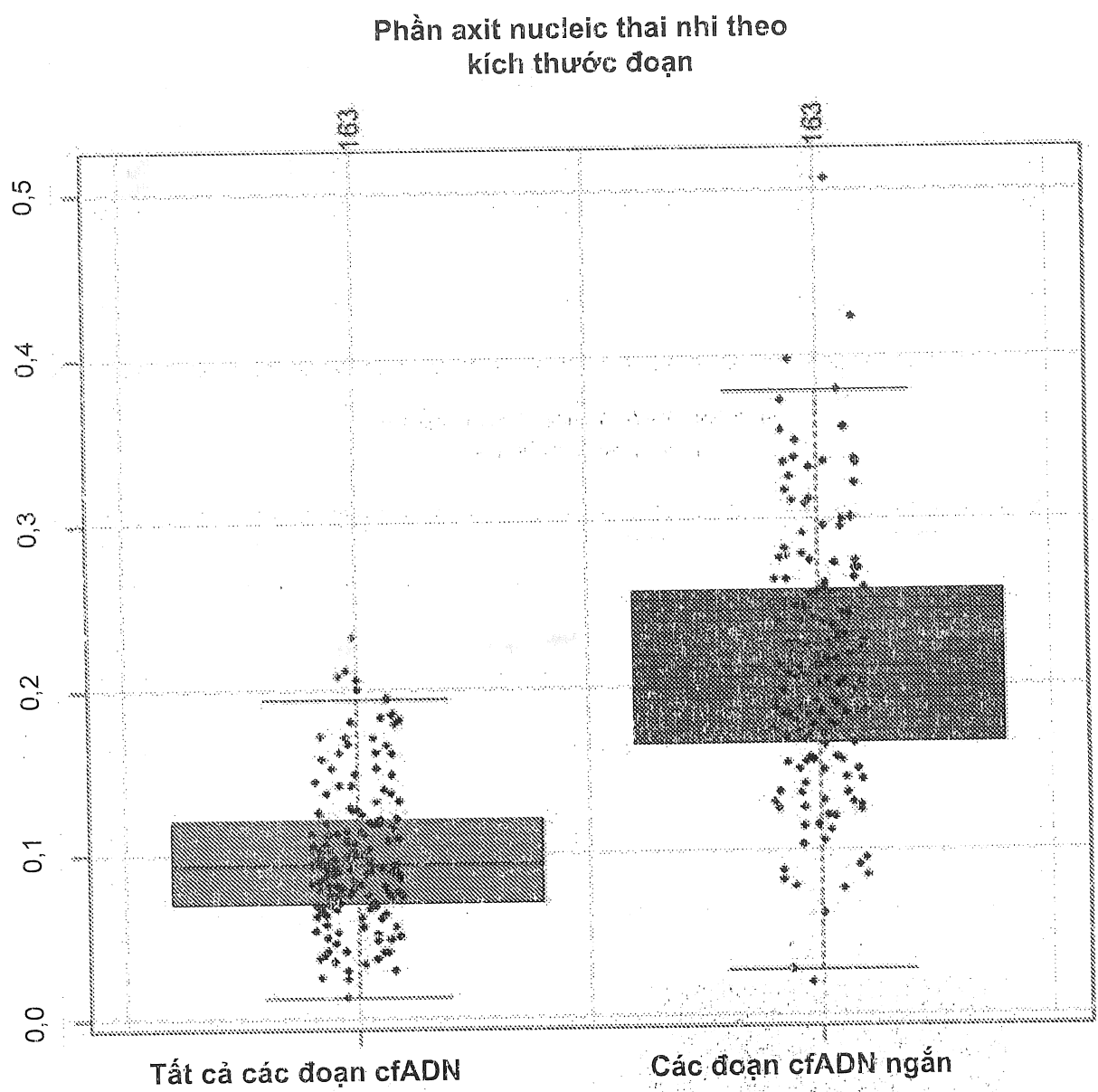


Fig.11

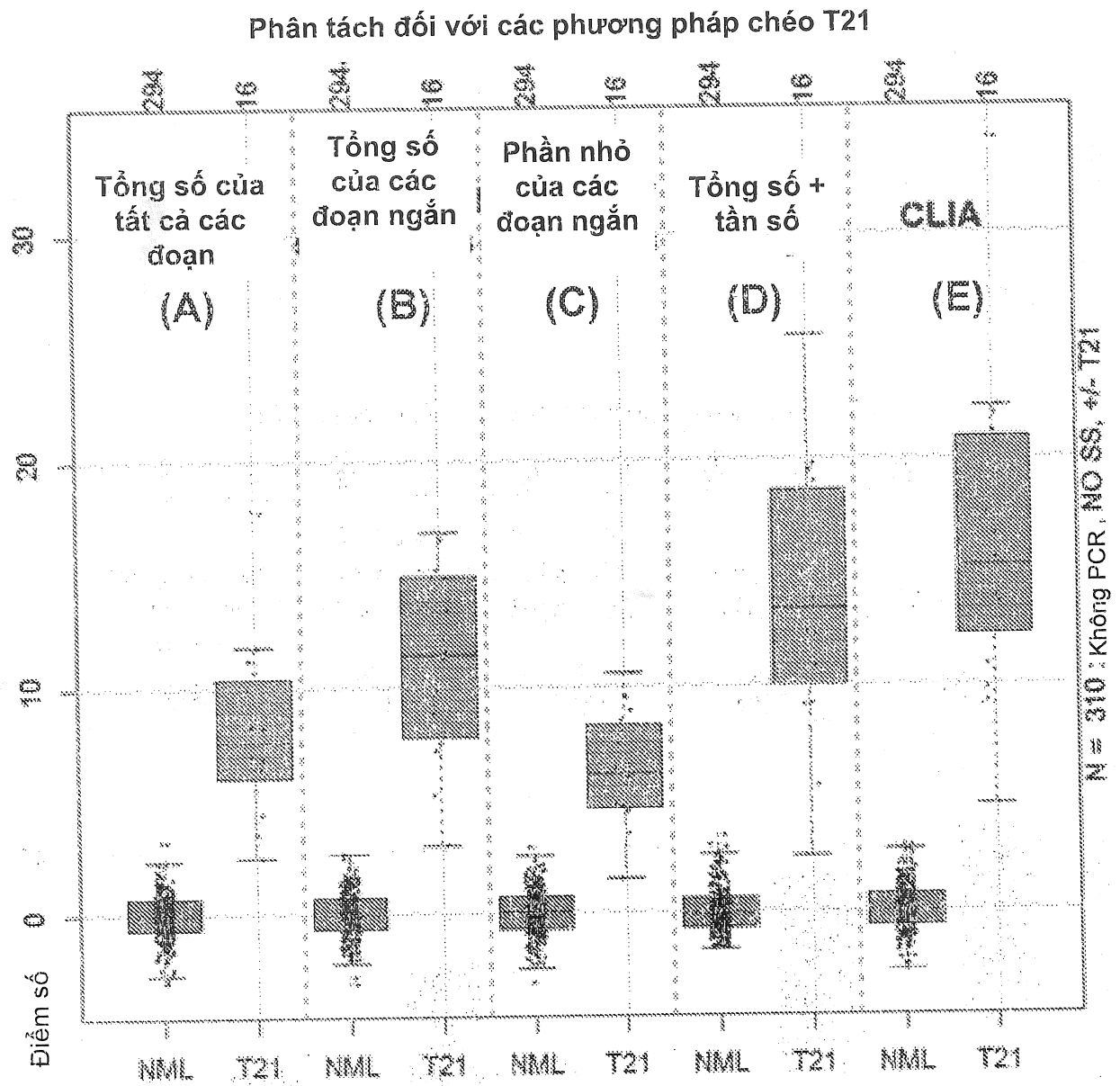


Fig.12

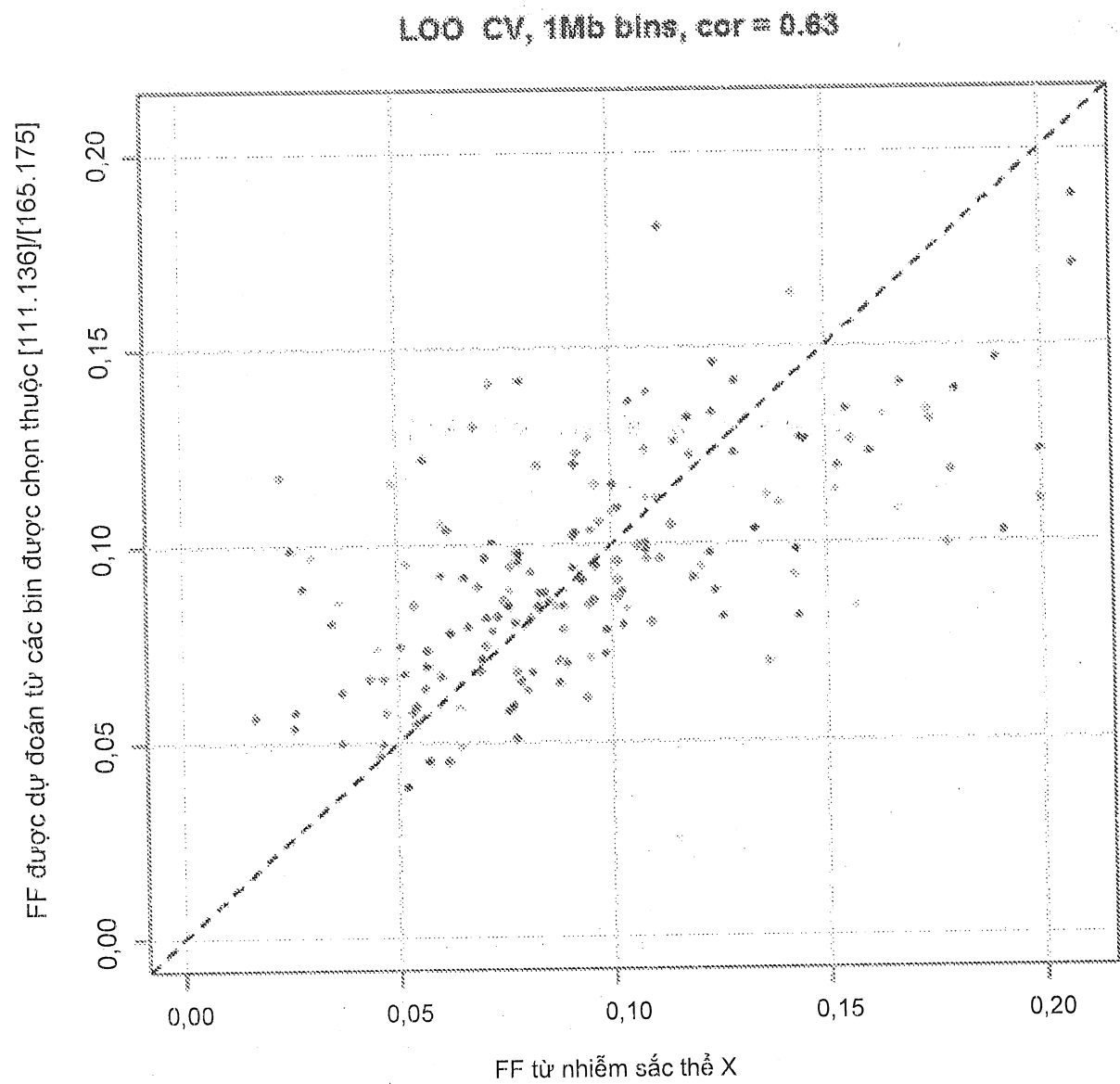


Fig.13