

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0031860
(43) 공개일자 2022년03월14일

- (51) 국제특허분류(Int. Cl.)
G16B 20/20 (2019.01) *C12Q 1/6869* (2018.01)
G16B 30/10 (2019.01) *G16B 40/10* (2019.01)
- (52) CPC특허분류
G16B 20/20 (2019.02)
C12Q 1/6869 (2018.05)
- (21) 출원번호 10-2021-7039566
 (22) 출원일자(국제) 2020년05월01일
 심사청구일자 없음
- (85) 번역문제출일자 2021년12월02일
 (86) 국제출원번호 PCT/US2020/031147
 (87) 국제공개번호 WO 2020/227137
 국제공개일자 2020년11월12일
- (30) 우선권주장
 62/842,534 2019년05월03일 미국(US)
 62/971,530 2020년02월07일 미국(US)

- (71) 출원인
 울티마 제노믹스, 인크.
 미국 94560 캘리포니아주 뉴어크 게이트웨이 블러바드 7979 스위트 101
- (72) 발명자
 에치오니, 요아브
 미국 94560 캘리포니아주 뉴어크 게이트웨이 블러바드 7979 스위트 101 울티마 제노믹스, 인크. 내
 파이글러, 심촌
 미국 94560 캘리포니아주 뉴어크 게이트웨이 블러바드 7979 스위트 101 울티마 제노믹스, 인크. 내
 (뒷면에 계속)
- (74) 대리인
 양영준, 이상남

전체 청구항 수 : 총 69 항

(54) 발명의 명칭 핵산 변이체를 검출하는 방법

(57) 요약

시험 샘플에서 짧은 유전자 변이체를 검출하는 방법이 본원에 기재되어 있다. 일부 예시적인 방법에서, 짧은 유전자 변이체는 시험 핵산 분자로부터 수득된 하나 이상의 시퀀싱 데이터 세트를 사용하여 결정된 하나 이상의 매치 스코어를 사용하여 호출되며, 여기서 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정된다. 또한, 2개 이상의 상이한 흐름-사이클 순서 및/또는 흐름 사이클 당 5개 이상의 뉴클레오타이드 흐름을 갖는 연장된 흐름-사이클 순서를 사용하여 시험 핵산 분자를 시퀀싱하는 방법이 본원에 기재되어 있다.

대표도

염기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G
0	0.0010	0.0010	0.9988	0.9988	0.0010	0.9988	0.9988	0.0001	0.0010	0.9988	0.0010	0.0010	0.0010	0.9988	0.0010	0.0010	0.9988	0.0010	0.9988	0.9988
			●○	●○		●○	●○			●○		●	●	○	●	●	○	●	○	○
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.9979	0.0010	0.9979	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.0010
	●○	●○			●○				●○		●○	○	○	●	○	○	●	○	●	●
2	0.0010	0.0010	0.0001	0.0001	0.0010	0.0001	0.0001	0.9979	0.0010	0.0001	0.0010	0.0010	0.0010	0.0001	0.0010	0.0010	0.0001	0.0010	0.0001	0.0001
								●○												
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

제1 후보 서열	T	A	-	-	T	-	-	GG	T	-	C	-	-	A	-	-	T	-	C	G
제2 후보 서열	T	A	-	-	T	-	-	GG	T	-	C	G	T	-	C	G	-	A	-	-

(52) CPC특허분류

G16B 30/10 (2019.02)

G16B 40/10 (2019.02)

C12Q 2600/156 (2013.01)

C12Q 2600/158 (2013.01)

(72) 발명자

알모기, 길라드

미국 94560 캘리포니아주 뉴어크 게이트웨이 블러
바드 7979 스위트 101 울티마 제노믹스, 인크. 내

프랫, 마크

미국 94560 캘리포니아주 뉴어크 게이트웨이 블러
바드 7979 스위트 101 울티마 제노믹스, 인크. 내

오베르스트라스, 플로리안

미국 94560 캘리포니아주 뉴어크 게이트웨이 블러
바드 7979 스위트 101 울티마 제노믹스, 인크. 내

명세서

청구범위

청구항 1

하기를 포함하는, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법:

(a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 복수의 흐름 위치를 포함하는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과인 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계;

(b) 하나 이상의 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 각각의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 각 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계;

(c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및

(d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

청구항 2

제1항에 있어서, 수득하는 단계가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 것을 포함하는 것인 방법.

청구항 3

제1항 또는 제2항에 있어서, 표적 짧은 유전자 변이체가 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택되는 것인 방법.

청구항 4

제1항 또는 제2항에 있어서, 표적 짧은 유전자 변이체가 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택되는 것인 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 포함하고, 바이오마커 패널이 표적 짧은 유전자 변이체를 포함하는 것인 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 흐름-사이클 순서를 선택하는 단계를 포함하는 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 인실리코에서 표적 서열 및 참조 서열을 시퀀싱함으로써 수득되는 것인 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 표적 시퀀싱 데이터 세트가 2개 초과인 비연속 흐름 위치에서 참조

시퀀싱 데이터와 상이한 것인 방법.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 표적 시퀀싱 데이터 세트가 2개 초과 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이한 것인 방법.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 표적 서열이 X개의 염기 위치에서 참조 서열과 상이하고, 여기서 표적 시퀀싱 데이터 세트가 (X+2)개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이한 것인 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서, 표적 시퀀싱 데이터 세트가 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이한 것인 방법.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함하는 것인 방법.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.

청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 복수의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.

청구항 15

제14항에 있어서, 단계 (c)가 하기를 포함하는 것인 방법:

해당 흐름 위치에서 표적 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 또는

해당 흐름 위치에서 참조 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계.

청구항 16

제15항에 있어서, 단계 (c)에서 결정된 매치 스코어가 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값인 방법.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 단계 (c)가 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함하는 것인 방법.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, 단계 (c)가 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함하는 것인 방법.

청구항 19

제1항 내지 제18항 중 어느 한 항에 있어서, 하나 이상의 시험 시퀀싱 데이터 세트가 복수의 시험 시퀀싱 데이터 세트를 포함하는 것인 방법.

청구항 20

제19항에 있어서, 표적 짧은 유전자 변이체의 존재 또는 부재가 하나 이상의 시험 시퀀싱 데이터 세트 각각에 대해 별도로 호출되는 것인 방법.

청구항 21

제19항 또는 제20항에 있어서, 복수의 시험 시퀀싱 데이터 세트의 적어도 일부가 상이한 시퀀싱 시작 위치를 갖는 상이한 시험 핵산 분자와 연관된 것인 방법.

청구항 22

제1항 내지 제21항 중 어느 한 항에 있어서, 흐름-사이클 순서가 동일한 순서로 반복되는 4개의 별도의 흐름을 포함하는 것인 방법.

청구항 23

제1항 내지 제22항 중 어느 한 항에 있어서, 흐름-사이클 순서가 5개 이상의 별도의 흐름을 포함하는 것인 방법.

청구항 24

제1항 내지 제23항 중 어느 한 항에 있어서, 하기를 포함하는, 컴퓨터-구현 방법인 방법:

하나 이상의 프로세서를 사용하여 표적 짧은 유전자 변이체를 선택하는 단계;

하나 이상의 프로세서에서 하나 이상의 시험 시퀀싱 데이터 세트를 수신함으로써 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계;

하나 이상의 프로세서를 사용하여 하나 이상의 매치 스코어를 결정하는 단계; 및

하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

청구항 25

하기를 포함하는, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법:

(a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계;

(b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계;

(c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및

(d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

청구항 26

제25항에 있어서, 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계, 및 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계를 포함하는 방법.

청구항 27

제25항 또는 제26항에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성, 또는 제2 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타내는 것인 방법.

청구항 28

제25항 또는 제26항에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트 둘 모두가 후보 서열과 매치할 가능성을 나타내는 것인 방법.

청구항 29

제25항 내지 제28항 중 어느 한 항에 있어서, 하나 이상의 후보 서열이 2개 이상의 상이한 후보 서열을 포함하고, 방법이 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트와 연관된 각 핵산 분자에 대해, 하기를 포함하는 것인 방법:

2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

청구항 30

제29항에 있어서, 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.

청구항 31

제29항에 있어서, 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 비연속 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.

청구항 32

제29항에 있어서, 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.

청구항 33

제29항에 있어서, 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열이 X개의 염기 위치에서 선택된 후보 서열과 상이하고, 여기서 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 (X+2)개 이상의 흐름 위치에서 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열과 상이한 것인 방법.

청구항 34

제29항에 있어서, 2개 이상의 상이한 후보 서열로부터 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 선택된 후보 서열과 상이한 것인 방법.

청구항 35

제22항 내지 제34항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함하는 것인 방법.

청구항 36

제25항 내지 제35항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능도를 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.

청구항 37

제25항 내지 제36항 중 어느 한 항에 있어서, 흐름 신호가 각 흐름 위치에서 복수의 염기 카운트에 대한 가능도를 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.

청구항 38

제37항에 있어서, 매치 스코어를 결정하는 단계가 하나 이상의 상이한 후보 서열 각각에 대해, 해당 흐름 위치에서 후보 서열의 염기 카운트에 상응하는, 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하는 것을 포함하는 것인 방법.

청구항 39

제37항에 있어서, 하나 이상의 상이한 후보 서열에 대해, 각 흐름 위치에서 후보 서열의 염기 카운트를 포함하는 후보 시퀀싱 데이터 세트를 생성하는 단계를 포함하는 방법.

청구항 40

제39항에 있어서, 후보 시퀀싱 데이터 세트가 인실리코에서 생성되는 것인 방법.

청구항 41

제37항 내지 제40항 중 어느 한 항에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값인 방법.

청구항 42

제25항 내지 제41항 중 어느 한 항에 있어서, 시험 핵산 분자의 적어도 일부가 상이한 시퀀싱 시작 위치를 갖는 것인 방법.

청구항 43

제25항 내지 제42항 중 어느 한 항에 있어서, 하기를 포함하는 방법:

표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계;

여기서 하나 이상의 후보 서열은 표적 서열 및 참조 서열을 포함함.

청구항 44

제43항에 있어서, 표적 짧은 유전자 변이체가 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택되는 것인 방법.

청구항 45

제43항에 있어서, 표적 짧은 유전자 변이체가 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택되는 것인 방법.

청구항 46

제43항 내지 제45항 중 어느 한 항에 있어서, 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 포함하고, 바이오마커 패널이 시험 샘플에 존재하는 표적 짧은 유전자 변이체를 포함하는 것인 방법.

청구항 47

제43항 내지 제46항 중 어느 한 항에 있어서, 참조 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시퀀싱되었을 경우, 참조 시퀀싱 데이터 세트가 예상 참조 시퀀싱 데이터 세트를 결정함으로써 획득되는 것인 방법.

청구항 48

제43항 내지 제46항 중 어느 한 항에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 흐름 위치에서 참조 서열과 상이한 것인 방법.

청구항 49

제43항 내지 제46항 중 어느 한 항에 있어서, 표적 서열이 제1 흐름에 따라 2개 이상의 비연속 흐름 위치에서 참조 서열과 상이한 것인 방법.

청구항 50

제43항 내지 제46항 중 어느 한 항에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 참조 서열과 상이한 것인 방법.

청구항 51

제43항 내지 제46항 중 어느 한 항에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 참조 서열과 상이한 것인 방법.

청구항 52

제43항 내지 제51항 중 어느 한 항에 있어서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서가 동일한 순서로 반복되는 4개의 별도의 흐름을 포함하는 것인 방법.

청구항 53

제43항 내지 제51항 중 어느 한 항에 있어서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서가 동일한 순서로 반복되는 5개 이상의 별도의 흐름을 포함하는 것인 방법.

청구항 54

제25항 내지 제53항 중 어느 한 항에 있어서, 하기를 포함하는 방법:

제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로의 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제 1 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 시험 핵산 분자를 시퀀싱하는 단계;

연장된 시퀀싱 프라이머를 제거하는 단계; 및

제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로의 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제 2 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 동일한 시험 핵산 분자를 시퀀싱하는 단계.

청구항 55

제25항 내지 제54항 중 어느 한 항에 있어서, 하기를 포함하는, 컴퓨터-구현 방법인 방법:

하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계;

상기 하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계;

상기 하나 이상의 프로세서를 사용하여 매치 스코어를 결정하는 단계; 및

상기 하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

청구항 56

제1항 내지 제55항 중 어느 한 항에 있어서, 별도의 흐름이 단일 염기 유형을 포함하는 것인 방법 또는 시스템.

청구항 57

제1항 내지 제56항 중 어느 한 항에 있어서, 별도의 흐름 중 적어도 하나가 2 또는 3개의 상이한 염기 유형을 포함하는 것인 방법 또는 시스템.

청구항 58

제1항 내지 제57항 중 어느 한 항에 있어서, 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 변이체 호출 파일을 생성 또는 업데이트하는 단계를 포함하는 방법 또는 시스템.

청구항 59

제1항 내지 제58항 중 어느 한 항에 있어서, 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 보고서를 생성하는 단계를 포함하는 방법 또는 시스템.

청구항 60

제59항에 있어서, 보고서가 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 텍스트, 확률, 수치 또는 그래픽 출력을 포함하는 것인 방법 또는 시스템.

청구항 61

제59항 또는 제60항에 있어서, 환자 또는 환자의 의료 대리인에게 보고서를 제공하는 단계를 포함하는 방법 또는 시스템.

청구항 62

제1항 내지 제61항 중 어느 한 항에 있어서, 짧은 유전자 변이체가 단일 뉴클레오타이드 다형성 또는 indel을 포함하는 것인 방법 또는 시스템.

청구항 63

하기를 포함하는, 핵산 분자를 시퀀싱하는 방법:

핵산 분자를 프라이머에 혼성화하여 혼성화된 주형을 형성하는 단계;

5개 이상의 별도의 뉴클레오타이드 흐름을 포함하는 반복된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 표지된 비중결 뉴클레오타이드를 사용하여 프라이머를 연장하는 단계; 및

프라이머가 뉴클레오타이드 흐름에 의해 연장됨에 따라 혼입된 표지된 뉴클레오타이드로부터의 신호 또는 신호의 부재를 검출하는 단계.

청구항 64

제63항에 있어서, 각 뉴클레오타이드 흐름 후에 신호 또는 신호의 부재를 검출하는 단계를 포함하는 방법.

청구항 65

제63항 또는 제64항에 있어서, 복수의 핵산 분자를 시퀀싱하는 단계를 포함하는 방법.

청구항 66

제65항에 있어서, 복수의 핵산 분자가 유전자좌에 대해 상이한 시퀀싱 시작 위치를 갖는 것인 방법.

청구항 67

제63항 내지 제66항 중 어느 한 항에 있어서, 흐름-사이클 순서가 무작위 시퀀싱 시작 위치의 5% 이상에서 가능한 SNP 순열의 50% 이상에 대해 2개 초과 흐름 위치에서 신호 변화를 유도하는 것인 방법.

청구항 68

제63항 내지 제67항 중 어느 한 항에 있어서, 흐름-사이클 순서가 흐름 당 0.6 이상의 염기 혼입 효율을 갖는 것인 방법.

청구항 69

하기를 포함하는 시스템:

하나 이상의 프로세서; 및

제1항 내지 제62항 중 어느 한 항의 방법을 구현하기 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체.

발명의 설명

기술 분야

- [0001] 관련 출원에 대한 상호 참조
- [0002] 본 출원은 2019년 5월 3일에 출원된 미국 특허 가출원 일련 번호 62/842,534; 및 2020년 2월 7일에 출원된 미국 특허 가출원 일련 번호 62/971,530을 우선권 주장하며; 이들 각각의 내용은 그 전문이 본원에 참조로 포함된다.
- [0003] ASCII 텍스트 파일의 서열 목록의 제출
- [0004] ASCII 텍스트 파일 상의 하기 제출의 내용은 그 전문이 본원에 참조로 포함된다: 서열 목록의 컴퓨터 판독가능 형태 (CRF) (파일명: 165272000540SEQLIST.TXT, 기록 날짜: 2020년 4월 27일, 크기: 5 KB).
- [0005] 발명의 분야
- [0006] 유전자 변이체의 검출을 포함하는 시퀀싱 데이터를 생성 및/또는 분석하는 방법을 포함하는, 폴리뉴클레오타이드를 시퀀싱하는 방법이 본원에 기재되어 있다.

배경 기술

- [0007] DNA 샘플 내의 유전자 변이체는 샘플 내의 DNA를 시퀀싱하고, 서열을 참조 서열에 정렬하고, 차이를 평가함으로써 검출될 수 있다. 시퀀싱된 DNA 및 참조 서열 사이의 높은 신뢰도 차이는 DNA 샘플이 유래된 유기체에 대한 변이체라고 불린다. 차세대 시퀀싱은 단일 샘플에서 많은 상이한 핵산 분자를 동시에 시퀀싱하여 분석할 상당한 양의 데이터를 생성하는데 필요한 도구를 연구 및 임상 실험실에 제공하였다.
- [0008] 또한, 합성에 의한 가역적-종결자 시퀀싱 (예를 들어, 가역적으로 종결된, 염료-표지된 시퀀싱 방법)은 각 염기에 대해 단일 구분된 신호를 제공하며, 따라서 단일-신호 시퀀싱 오류는 잘못된 변이체 호출을 초래할 수 있다. 일부 경우에, 이는 높은 깊이 시퀀싱에 의해 극복되어 참양성 신호로 잘못된 호출을 효과적으로 압도할 수 있지만, 이러한 높은 깊이에서의 시퀀싱은 고가이고 시간이 많이 소요된다.
- [0009] 고도로 효율적이고 정확한 염기 호출 및 변이체 호출 프로토콜에 대한 필요성은 관련 기술분야에서 여전히 요구되고 있다.

발명의 내용

- [0010] 발명의 간단한 요약
- [0011] 핵산 분자를 함유하는 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법이 본원에 기재되어 있으며, 이는 특정 실시양태에서 컴퓨터-구현 방법일 수 있다. 또한, 이러한 방법을 수행하기 위한 시스템이 본원에 기재되어

있다. 핵산 분자를 시퀀싱하는 방법이 추가로 기재되어 있다.

- [0012] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0013] 상기 방법의 일부 실시양태에서, 획득하는 단계는 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 시험 핵산 분자를 시퀀싱하는 것을 포함한다.
- [0014] 상기 방법의 일부 실시양태에서, 표적 짧은 유전자 변이체는 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택된다. 일부 실시양태에서, 표적 짧은 유전자 변이체는 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택된다. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다.
- [0015] 상기 방법의 일부 실시양태에서, 방법은 흐름-사이클 순서를 선택하는 단계를 추가로 포함한다.
- [0016] 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 예상 표적 시퀀싱 데이터 세트이거나, 참조 시퀀싱 데이터 세트는 예상 참조 시퀀싱 데이터 세트이다. 일부 실시양태에서, 예상 표적 시퀀싱 데이터 세트 및 예상 참조 시퀀싱 데이터 세트는 인실리코에서 표적 서열 및 참조 서열을 시퀀싱함으로써 획득된다.
- [0017] 상기 방법의 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 비연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 서열은 X개의 염기 위치에서 참조 서열과 상이하고, 여기서 표적 시퀀싱 데이터 세트는 (X+2)개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, (X+2) 흐름 위치 차이는 0과 실질적으로 동일한 값 및 0보다 실질적으로 큰 값 사이의 차이를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함한다.
- [0018] 상기 방법의 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트는 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타낸다. 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 복수의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트는 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타낸다.
- [0019] 상기 방법의 일부 실시양태에서, 단계 (c)는 (i) 해당 흐름 위치에서 표적 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 또는 (ii) 해당 흐름 위치에서 참조 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함한다. 일부 실시양태에서, 단계 (c)에서 결정된 매치 스코어는 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값이다. 일부 실시양태에서, 단계 (c)는 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함한다. 일부 실시양태에서, 단계 (c)는 시험 시퀀싱 데이터

세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함한다.

- [0020] 상기 방법의 일부 실시양태에서, 하나 이상의 시험 시퀀싱 데이터 세트는 복수의 시험 시퀀싱 데이터 세트를 포함한다. 일부 실시양태에서, 표적 짧은 유전자 변이체의 존재 또는 부재는 하나 이상의 시험 시퀀싱 데이터 세트 각각에 대해 별도로 호출된다. 일부 실시양태에서, 복수의 시험 시퀀싱 데이터 세트의 적어도 일부는 상이한 시퀀싱 시작 위치를 갖는 상이한 시험 핵산 분자와 연관된다.
- [0021] 상기 방법의 일부 실시양태에서, 흐름-사이클 순서는 동일한 순서로 반복되는 4개의 별도의 흐름을 포함한다. 일부 실시양태에서, 흐름-사이클 순서는 5개 이상의 별도의 흐름을 포함한다.
- [0022] 상기 방법의 일부 실시양태에서, 방법은 컴퓨터-구현 방법이다. 예를 들어, 일부 실시양태에서, 컴퓨터-구현 방법은 하나 이상의 프로세서를 사용하여 표적 짧은 유전자 변이체를 선택하는 단계; 하나 이상의 프로세서에서 하나 이상의 시험 시퀀싱 데이터 세트를 수신함으로써 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계; 하나 이상의 프로세서를 사용하여 하나 이상의 매치 스코어를 결정하는 단계; 및 하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계를 포함한다.
- [0023] 하기를 포함하는 시스템이 또한 본원에 제공된다: 하나 이상의 프로세서; 및 상기 방법을 구현하기 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체.
- [0024] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0025] 상기 방법의 일부 실시양태에서, 방법은 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계, 및 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계를 포함한다.
- [0026] 상기 방법의 일부 실시양태에서, 매치 스코어는 제1 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성, 또는 제2 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타낸다. 일부 실시양태에서, 매치 스코어는 제1 시험 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트 둘 모두가 후보 서열과 매치할 가능성을 나타낸다.
- [0027] 상기 방법의 일부 실시양태에서, 하나 이상의 후보 서열은 2개 이상의 상이한 후보 서열을 포함하고, 방법은 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트와 연관된 각 핵산 분자에 대해, 하기를 포함한다: 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 비연속 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서

들 모두에 따라 2개 이상의 비연속 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 연속 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 연속 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 3개 이상의 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 X개의 염기 위치에서 선택된 후보 서열과 상이하고, 여기서 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 (X+2)개 이상의 흐름 위치에서 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 X개의 염기 위치에서 선택된 후보 서열과 상이하고, 여기서 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 (X+2)개 이상의 흐름 위치에서 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열과 상이하다. 일부 실시양태에서, (X+2) 흐름 위치 차이는 0과 실질적으로 동일한 값 및 0보다 실질적으로 큰 값 사이의 차이를 포함한다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 선택된 후보 서열과 상이하다. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 하나 이상의 흐름-사이클에 걸쳐 선택된 후보 서열과 상이하다.

[0028] 상기 방법의 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함한다. 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트는 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타낸다. 일부 실시양태에서, 흐름 신호는 각 흐름 위치에서 복수의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트는 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타낸다. 일부 실시양태에서, 매치 스코어를 결정하는 단계는 하나 이상의 상이한 후보 서열 각각에 대해, 해당 흐름 위치에서 후보 서열의 염기 카운트에 상응하는, 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하는 것을 포함한다. 상기 방법의 일부 실시양태에서, 방법은 하나 이상의 상이한 후보 서열에 대해, 각 흐름 위치에서 후보 서열의 염기 카운트를 포함하는 후보 시퀀싱 데이터 세트를 생성하는 단계를 포함한다. 일부 실시양태에서, 후보 시퀀싱 데이터 세트는 인실리코에서 생성된다. 일부 실시양태에서, 매치 스코어는 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값이다.

[0029] 상기 방법의 일부 실시양태에서, 시험 핵산 분자의 적어도 일부는 상이한 시퀀싱 시작 위치를 갖는다.

[0030] 상기 방법의 일부 실시양태에서, 방법은 하기를 추가로 포함한다: 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; 여기서 하나 이상의 후보 서열은 표적 서열 및 참조 서열을 포함한다. 일부 실시양태에서, 표적 짧은 유전자 변이체는 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택된다. 일부 실시양태에서, 표적 짧은 유전자 변이체는 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택된다. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 시험 샘플에 존재하는 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 참조 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시퀀싱되었을 경우, 참조 시퀀싱 데이터 세트는 예상 참조 시퀀싱 데이터 세트를 결정함으로써 획득된다. 일부 실시양태에서, 참조 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 별도의 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시퀀싱되었을 경우, 참조 시퀀싱 데이터 세트는 예상 참조 시퀀싱 데이터 세트를 결정함으로써 획득된다. 일부 실시양태에서, 표적 서열은 제1 흐

름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 비연속 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 비연속 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 연속 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 연속 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 3개 이상의 흐름 위치에서 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 참조 서열과 상이하다. 일부 실시양태에서, 표적 서열은 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 하나 이상의 흐름-사이클에 걸쳐 참조 서열과 상이하다.

- [0031] 상기 기재된 방법의 일부 실시양태에서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서는 동일한 순서로 반복되는 4개의 별도의 흐름을 포함한다. 일부 실시양태에서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서는 동일한 순서로 반복되는 5개 이상의 별도의 흐름을 포함한다.
- [0032] 상기 기재된 방법의 일부 실시양태에서, 방법은 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제1 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 시험 핵산 분자를 시퀀싱하는 단계; 연장된 시퀀싱 프라이머를 제거하는 단계; 및 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제2 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 동일한 시험 핵산 분자를 시퀀싱하는 단계를 포함한다.
- [0033] 상기 기재된 방법의 일부 실시양태에서, 방법은 컴퓨터-구현 방법이다. 예를 들어, 일부 실시양태에서, 컴퓨터-구현 방법은 하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계; 상기 하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계; 상기 하나 이상의 프로세서를 사용하여 매치 스코어를 결정하는 단계; 및 상기 하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계를 포함한다.
- [0034] 하기를 포함하는 시스템이 또한 본원에 기재되어 있다: 하나 이상의 프로세서; 및 상기 기재된 임의의 방법을 구현하기 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체.
- [0035] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 별도의 흐름은 단일 염기 유형을 포함한다.
- [0036] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 별도의 흐름 중 적어도 하나는 2 또는 3개의 상이한 염기 유형을 포함한다.
- [0037] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 방법은 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 변이체 호출 파일을 생성 또는 업데이트하는 단계를 포함한다.
- [0038] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 방법은 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 보고서를 생성하는 단계를 포함한다. 일부 실시양태에서, 보고서는 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 텍스트, 확률, 수치 또는 그래픽 출력을 포함한다. 일부 실시양태에서, 방법은 환자 또는 환자의 의료 대리인에게 보고서를 제공하는 단계를 포함한다.
- [0039] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 짧은 유전자 변이체는 단일 뉴클레오티드 다형성을 포함한다.
- [0040] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 짧은 유전자 변이체는 indel을 포함한다.
- [0041] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 시험 샘플은 단편화된 DNA를 포함한다.
- [0042] 상기 기재된 임의의 방법 또는 시스템의 일부 실시양태에서, 시험 샘플은 무세포 DNA를 포함한다. 일부 실시양태에서, 무세포 DNA는 순환 종양 DNA (ctDNA)를 포함한다.
- [0043] 일부 실시양태에서, 핵산 분자를 시퀀싱하는 방법은 핵산 분자를 프라이머에 혼성화하여 혼성화된 주형을 형성

하는 단계; 5개 이상의 별도의 뉴클레오타이드 흐름을 포함하는 반복된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 표지된 비중결 뉴클레오타이드를 사용하여 프라이머를 연장하는 단계; 및 프라이머가 뉴클레오타이드 흐름에 의해 연장됨에 따라 혼입된 표지된 뉴클레오타이드로부터의 신호 또는 신호의 부재를 검출하는 단계를 포함한다. 일부 실시양태에서, 방법은 각 뉴클레오타이드 흐름 후에 신호 또는 신호의 부재를 검출하는 단계를 포함한다. 일부 실시양태에서, 방법은 복수의 핵산 분자를 시퀀싱하는 단계를 포함한다. 일부 실시양태에서, 복수의 핵산 분자는 유전자좌에 대해 상이한 시퀀싱 시작 위치를 갖는다. 일부 실시양태에서, 시험 샘플은 무세포 DNA이다. 일부 실시양태에서, 무세포 DNA는 순환 종양 DNA (ctDNA)를 포함한다. 일부 실시양태에서, 흐름-사이클 순서는 무작위 시퀀싱 시작 위치의 적어도 5%에서 가능한 SNP 순열의 50% 이상에 대해 2개 초과인 흐름 위치에서 신호 변화를 유도한다. 일부 실시양태에서, 유도된 신호 변화는 신호 강도의 변화이거나, 또는 새로운 실질적으로 제로 (또는 새로운 제로) 또는 새로운 실질적으로 넌-제로 (또는 새로운 넌-제로) 신호이다. 일부 실시양태에서, 유도된 신호 변화는 새로운 실질적으로 제로 (또는 새로운 제로) 또는 새로운 실질적으로 넌-제로 (또는 새로운 넌-제로) 신호이다. 일부 실시양태에서, 흐름-사이클 순서는 흐름 당 0.6 이상의 염기 혼입 효율을 갖는다.

도면의 간단한 설명

[0044]

도 1a는 T-A-C-G의 반복된 흐름-사이클 순서를 사용하여 TATGGTCGTCGA (서열식별번호(SEQ ID NO): 1)의 서열로 프라이머를 연장함으로써 획득된 시퀀싱 데이터를 보여준다. 시퀀싱 데이터는 연장된 프라이머 가닥을 대표하고, 상보적 주형 가닥에 대한 시퀀싱 정보는 쉽게 결정될 수 있고 효과적으로 동등하다.

도 1b는 각 흐름 위치에서 가장 높은 가능도에 기초하여 선택된 시퀀싱 데이터가 주어지면, 가장 가능성이 높은 서열과 함께 도 1a에 표시된 시퀀싱 데이터를 보여준다 (별표로 표시됨).

도 1c는 2개의 상이한 후보 서열: TATGGTCATCGA (서열식별번호: 2) (단힌 원) 및 TATGGTCGTCGA (서열식별번호: 1) (열린 원)을 표시하는 트레이스와 함께 도 1a에 표시된 시퀀싱 데이터를 보여준다. 시퀀싱 데이터가 주어진 서열과 매치할 가능성은 각 흐름 위치가 후보 서열과 매치할 가능성의 곱으로 결정될 수 있다.

도 2a는 2개의 후보 서열 H1 (서열식별번호: 5) 및 H2 (서열식별번호: 6) (각각 이들의 상보체로 표시됨)와 정렬된 시퀀싱 리드 R1 (서열식별번호: 1), R2 (서열식별번호: 3), 및 R3 (서열식별번호: 4) (각각 연장된 프라이머의 서열로 표시됨)의 정렬을 보여준다. 도 2b는 H1 (단힌 원) 및 H2 (열린 원)를 표시하는 트레이스와 함께 R1에 상응하는 시퀀싱 데이터를 보여준다. 도 2c는 H1 (단힌 원) 및 H2 (열린 원)를 표시하는 트레이스와 함께 R2에 상응하는 시퀀싱 데이터를 보여준다. 도 2d는 H1 (단힌 원) 및 H2 (열린 원)를 표시하는 트레이스와 함께 R3에 상응하는 시퀀싱 데이터를 보여준다.

도 3은 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 예시적인 방법의 흐름 차트를 보여준다.

도 4a는 제1 흐름-사이클 순서 (T-A-C-G)를 사용하여 핵산 분자를 시퀀싱함으로써 획득된 TATGGTCGTCGA (서열식별번호: 1)의 연장된 프라이머 서열을 갖는 핵산 분자로부터의 시퀀싱 데이터를 보여주고, 도 4b는 제2 흐름-사이클 순서 (A-G-C-T)를 사용하여 동일한 핵산 분자를 시퀀싱함으로써 획득된 시퀀싱 데이터를 보여준다. 또한, 도 4a 및 도 4b는 각각 제1 후보 서열 TATGGTCGTCGA (서열식별번호: 1) (단힌 원) 및 제2 후보 서열 TATGGTCATCGA (서열식별번호: 2) (열린 원)로부터의 트레이스를 보여준다. 도 4a 및 도 4b에 나타난 바와 같이, 흐름-사이클 순서의 차이는 주어진 흐름 위치에서 검출된 신호를 극적으로 변화시킬 수 있으며, 변이체의 컨텍스트에 대해 더 양호한 흐름 사이클을 사용할 때 더 유의한 신호 차이가 검출될 수 있다.

도 5는 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하기 위한 또 다른 예시적인 방법을 보여준다.

도 6은 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하기 위한 또 다른 예시적인 방법을 보여준다.

도 7은 본원에 기재된 방법을 구현하는데 사용될 수 있는, 한 실시양태에 따른 컴퓨팅 디바이스의 한 예를 예시한다.

도 8은 A-T-G-C 흐름 사이클 순서를 사용하여 시퀀싱된 가상의 핵산 분자로부터의 시퀀싱 데이터를 보여준다. 트레이스는 잠재적인 반수체형 서열 TATGGTCG-TCGA (서열식별번호: 7) (H1) 및 TATGGTCGATCG (서열식별번호: 8) (H2)을 사용하여 생성될 수 있으며, H1은 H2에 비해 1개의 염기 결실을 갖는다. 시퀀싱 데이터는 H2 후보 서열과 더 양호한 매치를 갖고, 이 서열에서 indel이 호출되지 않는다.

도 9는 4개의 예시적인 흐름 사이클 순서 (그 중 3개는 연장된 흐름 사이클 순서인 것을 포함함)에 대해, 무작위 시퀀싱 시작 위치가 주어지면 검출된 SNP 순열의 민감도를 보여준다. 도 9에서, x-축은 흐름 위상 (또는 단편화 시작 위치)의 분율을 나타내고, y-축은 2개 초과인 흐름 위치에서 신호 변화를 유도한 SNP 순열의 분율을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0045] 발명의 상세한 설명

[0046] 대상체로부터 유래된 시험 샘플에서 하나 이상의 짧은 유전자 변이체, 예컨대 단일 뉴클레오타이드 다형성 (SNP), 다중-뉴클레오타이드 다형성 (MNP), 또는 indel을 검출하는 방법이 본원에 기재되어 있다. 시험 샘플로부터의 시험 핵산 분자와 연관된 시험 시퀀싱 데이터는 분석되어 시험 시퀀싱 데이터 및 또 다른 서열 (예컨대, 시험 서열, 후보 서열 (또는 후보 반수체형 서열 및/또는 참조 서열) 사이의 매치를 결정하며, 이는 매치의 근접성을 나타내는 매치 스코어를 결정함으로써 반영될 수 있다 (예를 들어, 시험 시퀀싱 데이터가 주어지면, 시험 시퀀싱 데이터가 비교된 서열의 핵산 분자로부터 유래할 가능성). 그 후, 매치 스코어는 시험 샘플에서 짧은 유전자 변이체의 존재 또는 아이덴티티 또는 부재를 호출하는데 사용될 수 있다.

[0047] 시험 시퀀싱 데이터 세트는 계산적으로 효율적인 분석을 제공하도록 고유하게 구조화된다. 예를 들어, 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 생성될 수 있다. 그 후, 핵산 분자에 대한 시험 시퀀싱 데이터 세트는 특정 뉴클레오타이드의 흐름에 각각 상응하는 흐름 위치에서 흐름 신호를 포함한다. 이러한 고유하게 구조화된 데이터 세트를 사용하여, 핵산 분자 (또는 분자들)는 "베이스스페이스" ("뉴클레오타이드 공간" 또는 "서열 공간"으로도 지칭됨)가 아닌 "플로우스페이스"에서 분석될 수 있다. 플로우스페이스 데이터는 베이스스페이스 데이터에 의해 전달되지 않는 흐름-사이클 순서와 관련된 추가 정보에 의존한다. 플로우스페이스에서 수집된 데이터의 분석은 베이스스페이스로 전환되거나 수집된 데이터의 분석에 비해 적어도 2개의 장점을 제공한다. 첫째, 시험 핵산 분자에서 가장 일반적인 변이체 유형 (치환 SNP)은 플로우스페이스의 참조 서열과 비교할 때 2개 이상의 별개의 흐름 신호 (전체 흐름 사이클 또는 그 이상 동안 전파될 수 있음)를 초래할 것인 반면, 베이스스페이스의 서열을 분석할 때 하나의 데이터 신호만이 이용가능하다. 즉, 베이스스페이스에서, 각 염기 위치는 단일 신호와 연관되고, 변이체 염기는 변이체 염기의 신호에만 영향을 미치고, 인접한 신호에는 영향을 미치지 않는다. 플로우스페이스에서, 변이체는 다중 흐름 위치에 영향을 미칠 수 있고, 특정 변이체의 경우, 변이체는 참조 서열에 대한 후속 흐름도 신호의 이동을 유도하며, 이에 의해 사실상 변이체 검출의 지속적인 강화를 생성할 수 있다. 둘째, 플로우스페이스 데이터는 시험 핵산 분자의 서열 및 하나 이상의 후보 서열 사이의 직접적인 정렬 없이 하나 이상의 후보 플로우스페이스 서열과의 매치를 결정하기 위해 분석될 수 있다. 서열 정렬은 계산적으로 고가이고, 본원에 기재된 매치 분석을 사용하여 단순화될 수 있다.

[0048] 주어진 유전자 변이체에 대한 플로우스페이스의 다중-신호 표시기는 베이스스페이스 분석에서 식별될 수 있는 단일 신호 표시기보다 변이체 호출 정확도를 증가시킨다. 또한, 흐름 신호 차이의 수가 많을수록 변이체 호출이 검출될 가능성이 증가된다. 본원에서 추가로 논의된 바와 같이, 특정 상황에서, 높은 신뢰도로 미리 선택된 변이체를 호출하는 것이 바람직하고, 이러한 변이체 및/또는 흐름 순서는 원하는 수의 흐름 신호 차이가 생성되어 유전자 변이체를 확실하게 호출하는 것을 보장하도록 선택될 수 있다. 핵산 분자에 대한 시퀀싱 데이터 세트는 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하기 위해 후보 서열과 비교될 수 있다.

[0049] 베이스스페이스에서 후보 서열 (예컨대, 후보 반수체형 서열)에 대한 결정된 서열의 정렬은 계산적으로 고가이고, 현재 게놈 분석 툴 키트 (Genome Analysis Tool Kit; GATK) 하플로타입콜러(HaplotypeCaller)에서 가장 계산적으로 집약적인 단계이다. 하플로타입콜러 내에서, PairHMM은 각 시퀀싱 리드를 각 반수체형에 정렬하고, 염기 품질을 오류의 추정치로 사용하여 시퀀싱 리드가 주어진 반수체형의 가능성을 결정한다. 그러나, 본원에 기재된 방법과 함께 사용되는 데이터 세트의 구조는 오류 모드 가능성을 유지하며, 이는 변이체 호출을 계산적으로 더 효율적으로 만든다. 예를 들어, 주어진 유전자형 가능성은 유전자형을 갖는 서열과 정렬되는 각 흐름 위치에서의 가능성의 곱으로서 간단히 결정될 수 있다. 플로우스페이스 결정된 가능성은 계산적으로 더 효율적인 변이체 호출을 위해 하플로타입콜러의 PairHMM 모듈을 대체할 수 있다.

[0050] 시퀀싱 데이터 세트의 임의의 흐름 위치에 대한 흐름 신호는 임의의 염기 위치에서 핵산 분자를 시퀀싱하는데 사용되는 흐름 순서가 해당 위치에서 흐름 신호에 영향을 미칠 수 있다는 점에서 흐름-순서-의존적이다. 본원

에 추가로 기재된 바와 같이, 이 발견은 하나 이상의 방식으로 이용될 수 있다. 첫째, 동일한 유전자좌에서 중첩되는 핵산 분자의 무작위 단편화 (생체내 단편화, 예컨대 무세포 DNA, 또는 시험관내 단편화, 예컨대 초음파 처리 또는 효소적 소화에 의해)는 핵산 분자에 대한 (유전자좌에 비해) 다중 상이한 시퀀싱 시작 부위를 초래한다. 일부 경우에, 상이한 흐름 컨텍스트는 유전자좌에서 이용가능하다 (예를 들어, 상이한 흐름 순서로 재시퀀싱할 때, 또는 주기적인 흐름 순서를 사용할 때). 따라서, 유전자좌의 변이체는 다른 핵산 분자가 더 낮은-신뢰도 신호 (예를 들어, 단일 흐름 신호 변화)를 초래하더라도 변이체에 대한 높은 민감도 흐름 신호 (예를 들어, 참조 또는 비선택된 후보 서열과 비교하여 2개 이상의 흐름 신호 차이가 있음)를 갖는 단일 핵산 분자에 기초하여 올바르게 검출될 수 있다. 둘째, 주어진 핵산 분자는 제1 흐름 순서를 사용하여 시퀀싱될 수 있고, 제2 (상이한) 흐름 순서를 사용하여 재시퀀싱되어, 이에 따라 핵산 분자에 걸쳐 상이한 흐름 서열 컨텍스트를 제공할 수 있다. 하나의 흐름 순서를 사용하여 변이체를 갖는 핵산 분자와 변이체를 갖는 후보 서열의 가능성도 매치가 낮으면, 제2 흐름 순서를 사용하여 핵산 분자와 후보 서열의 가능성도 매치가 높을 수 있다. 셋째, 흐름 순서는 연장된 흐름 사이클일 수 있으며 (예를 들어, 한 사이클에 4개 초과 의 염기 유형 포함), 이는 단순히 4개의 염기 유형 A, C, T 및 G의 4개의 흐름 주기적인 반복이 아님을 의미한다. 일부 경우에, 반복 단위는 모든 가능한 2-염기 흐름 서열 (즉, 모든 X-Y 쌍은 반복 단위 내에 있음 (여기서, X는 모든 4개의 염기이고, Y는 각각 비-X 염기임)) 또는 3-염기 흐름 서열 (즉, 모든 가능한 X-Y-Z 순열은 반복 단위 내에 있음)을 포함하는 패턴과 같이 4개의 염기보다 더 길다. 넷째, 흐름 시퀀싱 순서는 특이적 유전자 변이체를 표적화하도록 선택될 수 있다.

[0051] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 시험 핵산 분자와 연관되고, 여기서 시험 시퀀싱 데이터 세트는 흐름 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 생성되었고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 시험 시퀀싱 데이터 세트가 하나 이상의 후보 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (c) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0052] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과 의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응하는 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0053] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나

이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

- [0054] 본원에 기재된 방법은 컴퓨터-구현 방법일 수 있고, 방법의 하나 이상의 단계는 예를 들어 하나 이상의 컴퓨터 프로세서를 사용하여 수행될 수 있다.
- [0055] 또한, 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 저장 매체가 본원에 제공되며, 하나 이상의 프로그램은 전자 디바이스의 하나 이상의 프로세서에 의해 실행될 때 전자 디바이스가 본원에 기재된 방법 중 임의의 하나 이상을 수행하게 하는 명령을 포함한다.
- [0056] 하나 이상의 프로세서, 메모리, 및 메모리에 저장된 하나 이상의 프로그램을 포함하는 전자 디바이스가 추가로 본원에 기재되어 있으며, 하나 이상의 프로그램은 하나 이상의 프로세서에 의해 실행되도록 구성된다. 하나 이상의 프로그램은 본원에 기재된 방법 중 임의의 하나 이상을 수행하기 위한 명령을 포함할 수 있다.
- [0057] 또한, 핵산 분자를 시퀀싱하는 방법이 본원에 기재되어 있다. 예를 들어, 핵산 분자를 시퀀싱하는 방법은 하기를 포함할 수 있다: 핵산 분자를 프라이머에 혼성화하여 혼성화된 주형을 형성하는 단계; 5개 이상의 별도의 뉴클레오타이드 흐름을 포함하는 반복된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 표지된 비종결 뉴클레오타이드를 사용하여 프라이머를 연장하는 단계; 및 프라이머가 뉴클레오타이드 흐름에 의해 연장됨에 따라 혼입된 표지된 뉴클레오타이드로부터의 신호 또는 신호의 부재를 검출하는 단계.
- [0058] 정의
- [0059] 본원에서 사용된 바와 같이, 단수 형태 "하나"는 문맥이 명백하게 달리 지시하지 않는 한 복수의 언급을 포함한다.
- [0060] 본원에서 "약" 값 또는 파라미터에 대한 언급은 해당 값 또는 파라미터 그 자체에 대한 변경을 포함 (및 기재) 한다. 예를 들어, "약 X"를 언급하는 기재는 "X"의 기재를 포함한다.
- [0061] 주어진 서열에 대한 "예상 시퀀싱 데이터" 또는 "예상 시퀀싱 데이터 세트"는 흐름 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 서열을 시퀀싱한 경우 생성될 계산된 시퀀싱 데이터를 지칭한다. 예상 시퀀싱 데이터 세트 또는 예상 시퀀싱 데이터 세트는 예를 들어 컴퓨터 모델링 (즉, 인실리코)에 의해 결정될 수 있다.
- [0062] "흐름 순서"는 비종결 뉴클레오타이드를 사용하여 핵산 분자를 시퀀싱하는데 사용되는 별도의 뉴클레오타이드 흐름의 순서를 지칭한다. 흐름 순서는 반복 단위의 사이클로 나눌 수 있으며, 반복 단위의 흐름 순서는 "흐름-사이클 순서"라고 한다. "흐름 위치"는 시퀀싱 공정 동안 주어진 별도의 뉴클레오타이드 흐름의 순차적 위치를 지칭한다.
- [0063] 용어 "개체," "환자" 및 "대상체"는 동의어로 사용되고, 인간을 포함하는 동물을 지칭한다.
- [0064] 본원에서 사용된 바와 같은 용어 "표지"는 또 다른 모이어티, 예를 들어 뉴클레오타이드 또는 뉴클레오타이드 유사체에 커플링되거나 커플링될 수 있는 검출가능한 모이어티를 지칭한다. 표지는 표지의 존재 또는 부재가 검출될 수 있도록 신호를 방출하거나 표지에 전달된 신호를 변경할 수 있다. 일부 경우에, 커플링은 절단가능, 예컨대 광-절단가능 (예를 들어, 자외선 하에 절단가능), 화학적으로-절단가능 (예를 들어, 환원제, 예컨대 디티오트레이톨 (DTT), 트리스(2-카르복시에틸)포스핀 (TCEP)을 통해) 또는 효소적으로 절단가능 (예를 들어, 에스테라제, 리파제, 펩티다제, 또는 프로테아제를 통해)할 수 있는 링커를 통해 이루어질 수 있다. 일부 실시양태에서, 표지는 형광단이다.
- [0065] "비종결 뉴클레오타이드"는 폴리머라제 또는 전사효소를 사용하여 폴리뉴클레오타이드의 3' 말단에 부착될 수 있고, 뉴클레오타이드로부터 보호기 또는 가역적 종결자를 제거할 필요 없이 폴리머라제 또는 전사효소를 사용하여 이에 부착된 또 다른 비종결 핵산을 가질 수 있는 핵산 모이어티이다. 자연 발생 핵산은 일종의 비종결 핵산이다. 비종결 핵산은 표지되거나 비표지될 수 있다.
- [0066] "뉴클레오타이드 흐름"은 하나 이상의 비종결 뉴클레오타이드 세트 (이는 표지될 수 있거나, 그 일부가 표지될 수 있음)를 지칭한다.
- [0067] "짧은 유전자 변이체"는 길이가 10개 이하의 연속 염기 (즉, 길이가 10, 9, 8, 7, 6, 5, 4, 3, 2 또는 1개의

염기(들)인 유전자 다형체 (즉, 돌연변이)를 설명하기 위해 본원에서 사용된다. 상기 용어는 단일 뉴클레오타이드 다형성 (SNP), 다중-뉴클레오타이드 다형성 (MNP), 및 길이가 10개 이하의 연속 염기인 indel을 포함한다.

- [0068] 본원에 기재된 본 발명의 측면 및 변경은 측면 및 변경으로 "이루어진" 및/또는 "본질적으로 이루어진"을 포함하는 것으로 이해된다.
- [0069] 값의 범위가 제공되는 경우, 해당 범위의 상한 및 하한 사이의 각 개재 값, 및 해당 언급 범위 내의 임의의 다른 언급된 또는 개재 값은 본 개시내용의 범위 내에 포함되는 것으로 이해되어야 한다. 언급된 범위가 상한 또는 하한을 포함하는 경우, 포함된 한계 중 어느 하나를 제외한 범위가 또한 본 개시내용에 포함된다.
- [0070] 본원에 기재된 분석 방법 중 일부는 서열을 참조 서열에 맵핑하고/거나, 서열 정보를 결정하고/거나, 서열 정보를 분석하는 것을 포함한다. 상보적 서열이 쉽게 결정 및/또는 분석될 수 있고, 본원에 제공된 설명이 상보적 서열을 참조하여 수행되는 분석 방법을 포함한다는 것이 관련 기술분야에서 잘 이해된다.
- [0071] 본원에서 사용된 섹션 제목은 조직화 목적만을 위한 것이며, 설명된 주제를 제한하는 것으로 해석되어서는 안된다. 설명은 관련 기술분야의 통상의 기술자가 본 발명을 만들고 사용할 수 있도록 제공되며 특허 출원 및 그 요건의 문맥에서 제공된다. 설명된 실시양태에 대한 다양한 변형은 관련 기술분야의 통상의 기술자에게 쉽게 명백할 것이고, 본원에서 일반적인 원리는 다른 실시양태에 적용될 수 있다. 그러므로, 본 발명은 표시된 실시양태에 제한되도록 의도되지 않지만, 본원에 기재된 원리 및 특징과 일치하는 가장 넓은 범위가 부여되어야 한다.
- [0072] 도면은 다양한 실시양태에 따른 공정을 예시한다. 예시적인 공정에서, 일부 블록은 임의로 조합되고, 일부 블록의 순서는 임의로 변경되고, 일부 블록은 임의로 생략된다. 일부 예에서, 추가 단계는 예시적인 공정과 조합하여 수행될 수 있다. 따라서, 예시된 (및 아래에서 더 자세히 설명된) 바와 같은 작동은 본래 예시적인 것이며, 이와 같이 제한적인 것으로 간주되어서는 안된다.
- [0073] 본원에 언급된 모든 간행물, 특허 및 특허 출원의 개시내용은 각각 그 전문이 본원에 참조로 포함된다. 참조로 포함된 임의의 참고문헌이 본 개시내용과 충돌하는 범위 내에서는 본 개시내용이 우선한다.
- [0074] 흐름 시퀀싱 방법
- [0075] 시퀀싱 데이터는 미리 결정된 흐름 사이클에 따라 주형 폴리뉴클레오타이드 분자에 결합된 프라이머를 연장하는 것을 포함하는 흐름 시퀀싱 방법을 사용하여 생성될 수 있으며, 여기서 임의의 주어진 흐름 위치에서, 단일 유형의 뉴클레오타이드가 연장 프라이머에 접근가능하다. 일부 실시양태에서, 특정 유형의 뉴클레오타이드 중 적어도 일부는 표지를 포함하고, 이는 표지된 뉴클레오타이드를 연장 프라이머에 혼입시 검출가능한 신호를 제공한다. 이러한 뉴클레오타이드가 연장된 프라이머에 혼입되는 생성된 서열은 주형 폴리뉴클레오타이드 분자의 서열의 역상보체여야 한다. 일부 실시양태에서, 예를 들어, 시퀀싱 데이터는 표지된 뉴클레오타이드를 사용하여 프라이머를 연장하고, 연장 프라이머에 혼입된 표지된 뉴클레오타이드의 존재 또는 부재를 검출하는 것을 포함하는 흐름 시퀀싱 방법을 사용하여 생성된다. 흐름 시퀀싱 방법은 또한 "합성에 의한 자연 시퀀싱" 또는 "합성에 의한 비종결 시퀀싱" 방법으로 지칭될 수 있다. 예시적인 방법은 미국 특허 번호 8,772,473에 기재되어 있으며, 이는 그 전문이 본원에 참조로 포함된다. 흐름 시퀀싱 방법을 참조하여 하기 설명이 제공되지만, 시퀀싱된 영역의 전부 또는 일부를 시퀀싱하기 위해 다른 시퀀싱 방법이 사용될 수 있음이 이해되어야 한다. 예를 들어, 본원에서 논의된 시퀀싱 데이터는 피로시퀀싱 방법을 사용하여 생성될 수 있다.
- [0076] 흐름 시퀀싱은 폴리뉴클레오타이드에 혼성화된 프라이머를 연장하기 위해 뉴클레오타이드를 사용하는 것을 포함한다. 주어진 염기 유형 (예를 들어, A, C, G, T, U 등)의 뉴클레오타이드는 상보적 염기가 주형 가닥에 존재하는 경우 프라이머를 연장하기 위해 혼성화된 주형과 혼합될 수 있다. 뉴클레오타이드는 예를 들어 비종결 뉴클레오타이드일 수 있다. 뉴클레오타이드가 비종결인 경우, 하나 초과인 연속 상보적 염기가 주형 가닥에 존재하는 경우 하나 초과인 연속 염기는 연장 프라이머 가닥에 혼입될 수 있다. 비종결 뉴클레오타이드는 3' 가역적 종결자를 갖는 뉴클레오타이드와 대조되며, 여기서 차단기는 일반적으로 연속 뉴클레오타이드가 부착되기 전에 제거된다. 상보적 염기가 주형 가닥에 존재하지 않는 경우, 주형 가닥에서 다음 염기에 상보적인 뉴클레오타이드가 도입될 때까지 프라이머 연장이 중단된다. 뉴클레오타이드의 적어도 일부는 혼입이 검출될 수 있도록 표지될 수 있다. 가장 일반적으로, 특정 실시양태에서 2 또는 3개의 상이한 유형의 뉴클레오타이드가 동시에 도입될 수 있지만, 한번에 단일 뉴클레오타이드 유형만이 도입된다 (즉, 별개로 첨가됨). 이 방법론은 가역적 종결자를 사용하는 시퀀싱 방법과 대조될 수 있으며, 여기서 프라이머 연장은 모든 단일 염기의 연장 후 종결자가 역전되어 다음 후속 염기의 혼입을 허용하기 전에 정지된다.

- [0077] 뉴클레오티드는 프라이머 연장 과정 동안 흐름 순서로 도입될 수 있으며, 이는 흐름 사이클로 추가로 나눌 수 있다. 흐름 사이클은 뉴클레오티드 흐름의 반복되는 순서이며, 임의의 길이일 수 있다. 뉴클레오티드는 단계별로 첨가되며, 이는 주형 가닥이 존재하는 상보적 염기의 시퀀싱 프라이머의 말단에 첨가된 뉴클레오티드의 혼입을 허용한다. 단지 예로서, 흐름 사이클의 흐름 순서는 A-T-G-C일 수 있거나, 흐름-사이클 순서는 A-T-C-G일 수 있다. 관련 기술분야의 통상의 기술자는 대안적인 순서를 쉽게 고려할 수 있다. 4개의 고유한 염기 유형(임의의 순서로 A, T, C 및 G)을 함유하는 흐름 사이클이 가장 일반적이지만, 흐름-사이클 순서는 임의의 길이일 수 있다. 일부 실시양태에서, 흐름 사이클은 흐름-사이클 순서로 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 별도의 뉴클레오티드 흐름을 포함한다. 단지 예로서, 흐름 사이클 순서는 T-C-A-C-G-A-T-G-C-A-T-G-C-T-A-G일 수 있으며, 이들 16개의 별도로 제공된 뉴클레오티드는 여러 사이클 동안 이 흐름-사이클 순서로 제공된다. 상이한 뉴클레오티드의 도입 사이에, 비혼입된 뉴클레오티드는 예를 들어 시퀀싱 플랫폼을 세척액으로 세척함으로써 제거될 수 있다.
- [0078] 폴리머라제는 주형-의존적 방식으로 프라이머의 말단에 하나 이상의 뉴클레오티드를 혼입함으로써 시퀀싱 프라이머를 연장하는데 사용될 수 있다. 일부 실시양태에서, 폴리머라제는 DNA 폴리머라제이다. 폴리머라제는 자연 발생 폴리머라제 또는 합성(예를 들어, 돌연변이체) 폴리머라제일 수 있다. 폴리머라제는 프라이머 연장의 초기 단계에서 첨가될 수 있지만, 보충 폴리머라제는 예를 들어 뉴클레오티드의 단계별 첨가와 함께 또는 다수의 흐름 사이클 후에 시퀀싱 동안 임의로 첨가될 수 있다. 예시적인 폴리머라제는 DNA 폴리머라제, RNA 폴리머라제, 열안정성 폴리머라제, 야생형 폴리머라제, 변형된 폴리머라제, Bst DNA 폴리머라제, Bst 2.0 DNA 폴리머라제, Bst 3.0 DNA 폴리머라제, Bsu DNA 폴리머라제, 이. 콜라이 DNA 폴리머라제 I, T7 DNA 폴리머라제, 박테리오파지 T4 DNA 폴리머라제 Φ 29 (phi29) DNA 폴리머라제, Taq 폴리머라제, Tth 폴리머라제, Tli 폴리머라제, Pfu 폴리머라제, 및 SeqAmp DNA 폴리머라제를 포함한다.
- [0079] 도입된 뉴클레오티드는 주형 가닥의 서열을 결정할 때 표지된 뉴클레오티드를 포함할 수 있고, 혼입된 표지된 핵산의 존재 또는 부재는 서열을 결정하기 위해 검출될 수 있다. 표지는 예를 들어 광학 활성 표지(예를 들어, 형광 표지) 또는 방사성 표지일 수 있고, 표지에 의해 방출되거나 변경된 신호는 검출기를 사용하여 검출될 수 있다. 주형 폴리뉴클레오티드에 혼성화된 프라이머에 혼입된 표지된 뉴클레오티드의 존재 또는 부재는 검출될 수 있으며, 이는 서열의 결정을 허용한다(예를 들어, 흐름도를 생성함으로써). 일부 실시양태에서, 표지된 뉴클레오티드는 형광성, 발광성 또는 다른 광-방출 모이어티로 표지된다. 일부 실시양태에서, 표지는 링커를 통해 뉴클레오티드에 부착된다. 일부 실시양태에서, 링커는 예를 들어 광화학적 또는 화학적 절단 반응을 통해 절단가능하다. 예를 들어, 표지는 연속 뉴클레오티드(들)의 검출 후에 및 혼입 전에 절단될 수 있다. 일부 실시양태에서, 표지(또는 링커)는 뉴클레오티드 염기에, 또는 DNA의 초기 가닥의 신장을 방해하지 않는 뉴클레오티드 상의 또 다른 부위에 부착된다. 일부 실시양태에서, 링커는 디설피드 또는 PEG-함유 모이어티를 포함한다.
- [0080] 일부 실시양태에서, 도입된 뉴클레오티드는 비표지된 뉴클레오티드만을 포함하고, 일부 실시양태에서 뉴클레오티드는 표지된 및 비표지된 뉴클레오티드의 혼합물을 포함한다. 예를 들어, 일부 실시양태에서, 총 뉴클레오티드와 비교하여 표지된 뉴클레오티드의 부분은 약 90% 이하, 약 80% 이하, 약 70% 이하, 약 60% 이하, 약 50% 이하, 약 40% 이하, 약 30% 이하, 약 20% 이하, 약 10% 이하, 약 5% 이하, 약 4% 이하, 약 3% 이하, 약 2.5% 이하, 약 2% 이하, 약 1.5% 이하, 약 1% 이하, 약 0.5% 이하, 약 0.25% 이하, 약 0.1% 이하, 약 0.05% 이하, 약 0.025% 이하, 또는 약 0.01% 이하이다. 일부 실시양태에서, 총 뉴클레오티드와 비교하여 표지된 뉴클레오티드의 부분은 약 100%, 약 95% 이상, 약 90% 이상, 약 80% 이상, 약 70% 이상, 약 60% 이상, 약 50% 이상, 약 40% 이상, 약 30% 이상, 약 20% 이상, 약 10% 이상, 약 5% 이상, 약 4% 이상, 약 3% 이상, 약 2.5% 이상, 약 2% 이상, 약 1.5% 이상, 약 1% 이상, 약 0.5% 이상, 약 0.25% 이상, 약 0.1% 이상, 약 0.05% 이상, 약 0.025% 이상, 또는 약 0.01% 이상이다. 일부 실시양태에서, 총 뉴클레오티드와 비교하여 표지된 뉴클레오티드의 부분은 약 0.01% 내지 약 100%, 예컨대 약 0.01% 내지 약 0.025%, 약 0.025% 내지 약 0.05%, 약 0.05% 내지 약 0.1%, 약 0.1% 내지 약 0.25%, 약 0.25% 내지 약 0.5%, 약 0.5% 내지 약 1%, 약 1% 내지 약 1.5%, 약 1.5% 내지 약 2%, 약 2% 내지 약 2.5%, 약 2.5% 내지 약 3%, 약 3% 내지 약 4%, 약 4% 내지 약 5%, 약 5% 내지 약 10%, 약 10% 내지 약 20%, 약 20% 내지 약 30%, 약 30% 내지 약 40%, 약 40% 내지 약 50%, 약 50% 내지 약 60%, 약 60% 내지 약 70%, 약 70% 내지 약 80%, 약 80% 내지 약 90%, 약 90% 내지 100% 미만, 또는 약 90% 내지 약 100%이다.
- [0081] 시퀀싱 데이터를 생성하기 전에, 폴리뉴클레오티드는 시퀀싱 프라이머에 혼성화되어 혼성화된 주형을 생성한다. 폴리뉴클레오티드는 시퀀싱 라이브러리 제조 동안 어댑터에 라이게이션될 수 있다. 어댑터는 시퀀싱 프라이머에 혼성화하는 혼성화 서열을 포함할 수 있다. 예를 들어, 어댑터의 혼성화 서열은 복수의 상이한 폴리뉴클레

오티드에 걸쳐 균일한 서열일 수 있고, 시퀀싱 프라이머는 균일한 시퀀싱 프라이머일 수 있다. 이는 시퀀싱 라이브러리에서 상이한 폴리뉴클레오티드의 다중화된 시퀀싱을 허용한다.

[0082] 폴리뉴클레오티드는 시퀀싱을 위해 표면 (예컨대 고체 지지체)에 부착될 수 있다. 폴리뉴클레오티드는 증폭되어 (예를 들어, 브릿지 증폭 또는 다른 증폭 기술에 의해) 폴리뉴클레오티드 시퀀싱 콜로니를 생성할 수 있다. 클러스터 내의 증폭된 폴리뉴클레오티드는 실질적으로 동일하거나 상보적이다 (폴리뉴클레오티드의 일부가 원래의 폴리뉴클레오티드와 반드시 동일하지 않을 수 있도록 증폭 공정 동안 일부 오류가 도입될 수 있음). 콜로니 형성은 검출기가 각 콜로니에 대한 표지된 뉴클레오티드의 혼입을 올바르게 검출할 수 있도록 신호 증폭을 허용한다. 일부 경우에, 콜로니는 에멀전 PCR을 사용하여 비드 상에 형성되고, 비드는 시퀀싱 표면에 걸쳐 분포된다. 시퀀싱을 위한 시스템 및 방법에 대한 예는 미국 특허 일련 번호 10,344,328에서 찾을 수 있으며, 이는 그 전문이 본원에 참조로 포함된다.

[0083] 폴리뉴클레오티드에 혼성화된 프라이머는 흐름 순서 (흐름-사이클 순서에 따라 순환적일 수 있음)에 따라 별도의 뉴클레오티드 흐름을 사용하여 핵산 분자를 통해 연장되고, 뉴클레오티드의 혼입은 상기 기재된 바와 같이 검출될 수 있으며, 이에 의해 핵산 분자에 대한 시퀀싱 데이터 세트를 생성한다.

[0084] 흐름 시퀀싱을 사용하는 프라이머 연장은 대략 수백 또는 심지어 수천 개의 염기 길이의 장거리 시퀀싱을 허용한다. 흐름 단계 또는 사이클의 수는 원하는 시퀀싱 길이를 획득하기 위해 증가 또는 감소될 수 있다. 프라이머의 연장은 하나 이상의 상이한 염기 유형을 갖는 뉴클레오티드를 사용하여 프라이머의 단계별 연장을 위한 하나 이상의 흐름 단계를 포함할 수 있다. 일부 실시양태에서, 프라이머의 연장은 1 내지 약 1000개의 흐름 단계, 예컨대 1 내지 약 10개의 흐름 단계, 약 10 내지 약 20개의 흐름 단계, 약 20 내지 약 50개의 흐름 단계, 약 50 내지 약 100개의 흐름 단계, 약 100 내지 약 250개의 흐름 단계, 약 250 내지 약 500개의 흐름 단계, 또는 약 500 내지 약 1000개의 흐름 단계를 포함한다. 흐름 단계는 동일한 또는 상이한 흐름 사이클로 세그먼트화될 수 있다. 프라이머에 혼입된 염기의 수는 시퀀싱된 영역의 서열, 및 프라이머를 연장하는데 사용되는 흐름 순서에 의존한다. 일부 실시양태에서, 시퀀싱된 영역은 길이가 약 1개 염기 내지 약 4000개 염기, 예컨대 길이가 약 1개 염기 내지 약 10개 염기, 길이가 약 10개 염기 내지 약 20개 염기, 길이가 약 20개 염기 내지 약 50개 염기, 길이가 약 50개 염기 내지 약 100개 염기, 길이가 약 100개 염기 내지 약 250개 염기, 길이가 약 250개 염기 내지 약 500개 염기, 길이가 약 500개 염기 내지 약 1000개 염기, 길이가 약 1000개 염기 내지 약 2000개 염기, 또는 길이가 약 2000개 염기 내지 약 4000개 염기이다.

[0085] 본원에 기재된 방법에서 사용된 폴리뉴클레오티드는 임의의 적합한 생물학적 공급원, 예를 들어 조직 샘플, 혈액 샘플, 혈장 샘플, 타액 샘플, 대변 샘플, 또는 소변 샘플로부터 획득될 수 있다. 폴리뉴클레오티드는 DNA 또는 RNA 폴리뉴클레오티드일 수 있다. 일부 실시양태에서, RNA 폴리뉴클레오티드는 폴리뉴클레오티드를 시퀀싱 프라이머에 혼성화하기 전에 DNA 폴리뉴클레오티드로 역전사된다. 일부 실시양태에서, 폴리뉴클레오티드는 무세포 DNA (cfDNA), 예컨대 순환 종양 DNA (ctDNA) 또는 태아 무세포 DNA이다. 핵산 분자는 예를 들어 생체 내에서 (예를 들어, cfDNA에서와 같이) 또는 시험관 내에서 (예를 들어, 초음파처리 또는 효소적 단편화에 의해) 무작위로 단편화될 수 있다.

[0086] 폴리뉴클레오티드의 라이브러리는 공지된 방법을 통해 제조될 수 있다. 일부 실시양태에서, 폴리뉴클레오티드는 어댑터 서열에 라이게이션될 수 있다. 어댑터 서열은 커플링된 시퀀싱 리드 쌍의 생성 동안 연장된 프라이머에 혼성화된 혼성화 서열을 포함할 수 있다.

[0087] 일부 실시양태에서, 시퀀싱 데이터는 시퀀싱 콜로니 (시퀀싱 클러스터로도 지칭됨)를 확립하기 전에 핵산 분자를 증폭하지 않고 획득된다. 시퀀싱 콜로니를 생성하는 방법은 브릿지 증폭 또는 에멀전 PCR을 포함한다. 샷건 시퀀싱 및 컨센서스 서열 호출에 의존하는 방법은 일반적으로 고유한 분자 식별자 (UMI)를 사용하여 핵산 분자를 표지하고, 핵산 분자를 증폭하여 독립적으로 시퀀싱된 동일한 핵산 분자의 수많은 카피를 생성한다. 그 후, 증폭된 핵산 분자는 표면에 부착되고 브릿지 증폭되어 독립적으로 시퀀싱된 시퀀싱 클러스터를 생성할 수 있다. 그 후, UMI는 독립적으로 시퀀싱된 핵산 분자를 연관시키는데 사용될 수 있다. 그러나, 증폭 공정은 예를 들어 DNA 폴리머라제의 제한된 충실도로 인해 핵산 분자에 오류를 도입할 수 있다. 일부 실시양태에서, 핵산 분자는 시퀀싱 데이터를 획득하기 위한 콜로니를 생성하기 위해 증폭 전에 증폭되지 않는다. 일부 실시양태에서, 핵산 시퀀싱 데이터는 고유한 분자 식별자 (UMI)를 사용하지 않고 획득된다.

[0088] 시퀀싱 데이터 세트 및 변이체 검출

[0089] 시퀀싱 데이터는 혼입된 뉴클레오티드의 검출 및 뉴클레오티드 도입의 순서에 기초하여 생성될 수 있다. 예를

들어, 흐르는 연장된 서열 (즉, 상응하는 주형 서열의 각 역상보체): CTG, CAG, CCG, CGT, 및 CAT (시퀀싱 방법에 적용되는 선행 서열 또는 후속 서열이 없다고 가정함), 및 T-A-C-G의 반복 흐름 사이클 (즉, 반복 사이클에서 T, A, C 및 G 뉴클레오타이드의 순차적 첨가)을 취한다. 주어진 흐름 위치에서 특정 유형의 뉴클레오타이드는 상보적 염기가 주형 폴리뉴클레오타이드에 존재하는 경우에만 프라이머에 혼입될 것이다. 예시적인 생성된 흐름도는 표 1에 표시되며, 여기서 1은 도입된 뉴클레오타이드의 혼입을 나타내고, 0은 도입된 뉴클레오타이드의 혼입이 없음을 나타낸다. 흐름도는 주형 가닥의 서열을 유도하는데 사용될 수 있다. 예를 들어, 본원에서 논의된 시퀀싱 데이터 (예를 들어, 흐름도)는 연장된 프라이머 가닥의 서열을 표시하고, 이의 역상보체는 주형 가닥의 서열을 표시하는 것으로 쉽게 결정될 수 있다. 표 1에서 별표 (*)는 추가 뉴클레오타이드가 연장된 시퀀싱 가닥 (예를 들어, 더 긴 주형 가닥)에 혼입되는 경우 시퀀싱 데이터에 신호가 존재할 수 있음을 나타낸다.

<표 1>

	사이클 1				사이클 2				사이클 3			
흐름 위치	1	2	3	4	5	6	7	8	9	10	11	12
흐름 내의 염기	T	A	C	G	T	A	C	G	T	A	C	G
연장된 서열: CTG	0	0	1	0	1	0	0	1	*	*	*	*
연장된 서열: CAG	0	0	1	0	0	1	0	1	*	*	*	*
연장된 서열: CCG	0	0	2	1	*	*	*	*	*	*	*	*
연장된 서열: CGT	0	0	1	1	1	*	*	*	*	*	*	*
연장된 서열: CAT	0	0	1	0	0	1	0	0	1	*	*	*

흐름도는 바이너리 또는 년-바이너리일 수 있다. 바이너리 흐름도는 혼입된 뉴클레오타이드의 존재 (1) 또는 부재 (0)를 검출한다. 년-바이너리 흐름도는 각 단계별 도입으로부터 혼입된 뉴클레오타이드의 수를 보다 정량적으로 결정할 수 있다. 예를 들어, CCG의 연장된 서열은 동일한 C 흐름 (예를 들어, 흐름 위치 3에서) 내에서 연장 프라이머에 2개의 C 염기의 혼입을 포함할 것이며, 표지된 염기에 의해 방출되는 신호는 단일 염기 혼입에 상응하는 강도 수준보다 더 큰 강도를 가질 것이다. 이는 표 1에 표시된다. 년-바이너리 흐름도는 또한 염기의 존재 또는 부재를 나타내고, 주어진 흐름 위치에서 각 연장 프라이머에 혼입될 가능성이 있는 염기의 수를 포함하는 추가 정보를 제공할 수 있다. 값이 정수일 필요는 없다. 일부 경우에, 값은 주어진 흐름 위치에서 혼입되는 염기의 수의 불확실성 및/또는 확률을 반영할 수 있다.

일부 실시양태에서, 시퀀싱 데이터 세트는 각 흐름 위치에 혼입된 시퀀싱된 핵산 분자에서 염기의 수를 나타내는 염기 카운트를 표시하는 흐름 신호를 포함한다. 예를 들어, 표 1에 나타난 바와 같이, T-A-C-G 흐름 사이클 순서를 사용하여 CTG 서열로 연장된 프라이머는 위치 3에서 1의 값을 가지며, 이는 해당 위치에서 1의 염기 카운트를 나타낸다 (1개의 염기는 C이며, 이는 시퀀싱된 주형 가닥에서 G에 상보적임). 또한 표 1에서, T-A-C-G 흐름 사이클 순서를 사용하여 CCG 서열로 연장된 프라이머는 위치 3에서 2의 값을 가지며, 이는 이 흐름 위치 동안 연장 프라이머에 대한 해당 위치에서 2의 염기 카운트를 나타낸다. 여기서, 2개의 염기는 연장 프라이머 서열에서 CCG 서열의 시작에 있는 C-C 서열을 지칭하고, 이는 주형 가닥에서 G-G 서열에 상보적이다.

시퀀싱 데이터 세트의 흐름 신호는 각 흐름 위치에서 하나 이상의 염기 카운트에 대한 가능도 또는 신뢰 구간을 나타내는 하나 이상의 통계적 파라미터를 포함할 수 있다. 일부 실시양태에서, 흐름 신호는 시퀀싱 공정 동안 검출된 아날로그 신호, 예컨대 시퀀싱 동안 시퀀싱 프라이머에 혼입된 하나 이상의 염기의 형광 신호로부터 결정된다. 일부 경우에, 아날로그 신호는 통계적 파라미터를 생성하기 위해 프로세싱될 수 있다. 예를 들어, 기계-학습 알고리즘은 공개된 국제 특허 출원 WO 2019084158 A1 (그 전문이 본원에 참조로 포함됨)에 기재된 바와 같이 아날로그 시퀀싱 신호의 컨텍스트 효과를 수정하는데 사용될 수 있다. 0 이상의 정수의 염기가 임의의 주어진 흐름 위치에 혼입되지만, 주어진 아날로그 신호는 아날로그 신호와 완벽하게 매치하지 않을 수 있다. 따라서, 검출된 신호가 주어지면, 흐름 위치에 혼입된 염기의 수의 가능도를 나타내는 통계적 파라미터가 결정될 수 있다. 단지 예로서, 표 1의 CCG 서열에 대해, 흐름 신호가 흐름 위치 3에 혼입된 2개의 염기를 나타낼 가능도는 0.999일 수 있고, 흐름 신호가 흐름 위치 3에 혼입된 1개의 염기를 나타낼 가능도는 0.001일 수 있다. 시퀀싱 데이터 세트는 각 흐름 위치에서 복수의 염기 카운트에 대한 가능도를 나타내는 통계적 파라미터를 포함하는 흐름 신호와 함께 희소 행렬로 포맷될 수 있다. 단지 예로서, T-A-C-G의 반복 흐름-사이클 순서를 사용하여 TATGGTCGTCGA (서열식별번호: 1)의 서열로 연장된 프라이머는 도 1a에 표시된 시퀀싱 데이터 세트를 초래할 수 있다. 통계적 파라미터 또는 가능도 값은 예를 들어 시퀀싱 동안 아날로그 신호의 검출 동안 존재하는 노이즈 또는 다른 아티팩트에 기초하여 달라질 수 있다. 일부 실시양태에서, 통계적 파라미터 또는 가능도가 미리 결

정된 역치 미만인 경우, 파라미터는 본원에서 추가로 논의된 통계적 분석을 돕기 위해 실질적으로 제로 (즉, 일부 매우 작은 값 또는 무시가능한 값)인 미리 결정된 년-제로 값으로 설정될 수 있으며, 여기서 진정한 제로 값은 계산적 오류를 발생시킬 수 있거나, 불가능도의 수준, 예를 들어 매우 가능성이 낮은 (0.0001) 및 상상할 수 없는 (0)을 불충분하게 구분할 수 있다.

[0095] 주어진 서열에 대한 시퀀싱 데이터 세트의 가능도를 나타내는 값은 서열 정렬 없이 시퀀싱 데이터 세트로부터 결정될 수 있다. 예를 들어, 데이터가 주어지면, 가장 가능성이 높은 서열은 도 1b에서 별로 표시된 바와 같이 각 흐름 위치에서 가장 높은 가능도를 갖는 염기 카운트를 선택함으로써 결정될 수 있다 (도 1a에 표시된 동일한 데이터 사용). 그러므로, 프라이머 연장 서열은 각 흐름 위치에서 가장 가능성이 높은 염기 카운트에 따라 결정될 수 있다: TATGGTCGTCGA (서열식별번호: 1). 이로부터, 역상보체 (즉, 주형 가닥)가 쉽게 결정될 수 있다. 또한, TATGGTCGTCGA (서열식별번호: 1) 서열 (또는 역상보체)이 주어지면, 이 시퀀싱 데이터 세트의 가능도는 각 흐름 위치에서 선택된 가능도의 곱으로 결정될 수 있다.

[0096] 핵산 분자와 연관된 시퀀싱 데이터 세트는 하나 이상 (예를 들어, 2, 3, 4, 5, 6개 또는 그 이상)의 가능한 후보 서열과 비교될 수 있다. 시퀀싱 데이터 세트 및 후보 서열 사이의 근접 매치 (하기 논의된 바와 같이 매치 스코어에 기초하여)는 시퀀싱 데이터 세트가 밀접하게 매치된 후보 서열과 동일한 서열을 갖는 핵산 분자로부터 발생하였을 가능성이 있음을 나타낸다. 일부 실시양태에서, 시퀀싱된 핵산 분자의 서열은 서열에 대한 유전자좌 (또는 하나 이상의 유전자좌)를 결정하기 위해 참조 서열에 맵핑될 수 있다 (예를 들어 버로우즈-휠러 정렬 (BWA) 알고리즘 또는 다른 적합한 정렬 알고리즘을 사용하여). 상기 논의된 바와 같이, 플로우스페이스의 시퀀싱 데이터 세트는 베이스스페이스로 쉽게 전환될 수 있고 (또는 흐름 순서가 공지된 경우 역도 또한 같음), 맵핑은 플로우스페이스 또는 베이스스페이스에서 수행될 수 있다. 맵핑된 서열에 상응하는 유전자좌 (또는 유전자좌들)는 하나 이상의 변이체 서열과 연관될 수 있으며, 이는 본원에 기재된 분석 방법에 대한 후보 서열 (또는 반수체형 서열)로서 작동할 수 있다. 본원에 기재된 방법의 한 가지 장점은 시퀀싱된 핵산 분자의 서열이 일반적으로 계산적으로 고가인 일부 경우에 정렬 알고리즘을 사용하여 각 후보 서열과 정렬될 필요가 없다는 것이다. 대신, 매치 스코어는 보다 계산적으로 효율적인 작동인 플로우스페이스의 시퀀싱 데이터를 사용하여 후보 서열 각각에 대해 결정될 수 있다.

[0097] 매치 스코어는 시퀀싱 데이터 세트가 후보 서열을 얼마나 잘 뒷받침하는지를 나타낸다. 예를 들어, 시퀀싱 데이터 세트가 후보 서열과 매치할 가능도를 나타내는 매치 스코어는 후보 서열에 대한 예상 시퀀싱 데이터가 주어지면 흐름 위치의 염기 카운트에 상응하는, 각 흐름 위치에서 통계적 파라미터 (예를 들어, 가능도)를 선택함으로써 결정될 수 있다. 선택된 통계적 파라미터의 곱은 매치 스코어를 제공할 수 있다. 예를 들어, 연장된 프라이머에 대한 도 1a에 표시된 시퀀싱 데이터 세트, 및 TATGGTCATCGA (서열식별번호: 2)의 후보 프라이머 연장 서열을 가정한다. 도 1c (도 1a에서 동일한 시퀀싱 데이터 세트를 나타냄)는 후보 서열에 대한 트레이스 (속이 찬 원)를 보여준다. 비교로서, TATGGTCGTCGA (서열식별번호: 1) 서열에 대한 트레이스 (도 1b 참조)는 열린 원을 사용하여 도 1c에 표시된다. 시퀀싱 데이터가 제1 후보 서열 TATGGTCATCGA (서열식별번호: 2)와 매치할 가능도를 나타내는 매치 스코어는 서열이 단일 염기 변이에 의해서만 달라진다 하더라도 시퀀싱 데이터가 제2 후보 서열 TATGGTCGTCGA (서열식별번호: 1)와 매치할 가능도를 나타내는 매치 스코어와 실질적으로 상이하다. 도 1c에 표시된 바와 같이, 트레이스 사이의 차이는 흐름 위치 12에서 관찰되고, 적어도 9개의 흐름 위치에 대해 전파된다 (그리고, 시퀀싱 데이터가 추가 흐름 위치에 걸쳐 연장된 경우 잠재적으로 더 길다). 하나 이상의 흐름 사이클에 걸친 이러한 계속된 전파는 "흐름 이동" 또는 "사이클 이동"으로 지칭될 수 있고, 시퀀싱 데이터 세트가 후보 서열과 매치하는 경우 일반적으로 매우 가능성이 낮은 사건이다.

[0098] 그 후, 각 시퀀싱 데이터 세트 및 후보 서열 (또는 각 후보 서열) 사이의 매치 스코어가 결정될 수 있다. 예를 들어, 시퀀싱 데이터 세트가 주어진 후보 서열과 매치할 가능도 $L(R_j | H_i)$ 는 주어진 후보 서열에 대한 각 흐름 위치에서 선택된 염기 카운트의 가능도 (예를 들어, 이의 곱)를 사용하여 결정될 수 있다.

[0099] 매치 스코어는 시험 시퀀싱 데이터 및/또는 시험 시퀀싱 데이터와 연관된 핵산 분자를 분류하는데 사용될 수 있다. 분류기는 핵산 분자가 변이체 (예를 들어, 후보 서열에 포함된 변이체)를 포함한다는 것, 핵산 분자가 변이체를 포함하지 않는다는 것을 나타낼 수 있거나, 널 호출을 나타낼 수 있다. 널 호출은 시험 시퀀싱 데이터와 연관된 핵산 분자에서 변이체의 존재 또는 부재를 나타내지 않지만, 대신 매치 스코어가 원하는 통계적 신뢰도로 호출을 하는데 사용될 수 없음을 나타낸다. 시험 시퀀싱 데이터 또는 핵산 분자는 예를 들어 매치 스코어가 원하는 신뢰도 역치를 초과하는 경우 변이체를 갖는 것으로 분류될 수 있다. 반대로, 시험 시퀀싱 데이터 또는 핵산 분자는 예를 들어 매치 스코어가 원하는 신뢰도 역치 미만인 경우 변이체를 갖지 않는 것으로 분류될

수 있다.

[0100] 상기 분석은 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하기 위해 적용될 수 있다. 시퀀싱 데이터 세트가 각 후보 서열과 매치할 가능성을 나타내는 매치 스코어가 결정될 수 있다. 예를 들어, 해당 흐름 위치에서 후보 서열의 염기 카운트에 상응하는, 시퀀싱 데이터 세트의 각 흐름 위치에서 통계적 파라미터는 각 후보 서열에 대해 선택될 수 있다. 일부 실시양태에서, 이 분석은 후보 서열이 시퀀싱된 시험 핵산 분자에 대한 시퀀싱 데이터 세트를 생성하는데 사용되는 동일한 흐름 순서를 사용하여 시퀀싱된다고 가정하여 후보 시퀀싱에 대한 예상 시퀀싱 데이터를 생성하는 것을 포함한다. 이는 후보 서열로 핵산 분자를 시퀀싱함으로써, 또는 후보 서열 및 흐름 순서에 기초하여 인실리코에서 후보 시퀀싱 데이터 세트를 생성함으로써 생성될 수 있다. 예시적인 후보 시퀀싱 데이터 세트는 하기 표시된 도 1c의 시험 데이터 시퀀싱 데이터 세트이며, 제1 후보 서열 (TATGGTCATCGA (서열식별번호: 2))은 속이 찬 원 트레이스에 상응하고, 제2 후보 서열 (TATGGTCGTCGA (서열식별번호: 1))은 열린 원 트레이스에 상응한다. 일부 실시양태에서, 예를 들어, 매치 스코어가 2개 이상의 상이한 후보 서열에 대해 결정되는 경우, 시험 시퀀싱 데이터 또는 핵산 분자는 2개 이상의 후보 서열 중 하나의 변이체를 갖는 것으로, 2개 이상의 후보 서열 중 하나의 변이체를 갖지 않는 것으로 분류될 수 있거나, 또는 2개 이상의 후보 서열 간에 널 호출이 수행될 수 있다 (예를 들어, 후보 서열 중 임의의 것에 대해 호출이 수행될 수 없는 경우, 또는 매치 스코어가 동일한 유전자좌에서 2개 이상의 상이한 변이체를 나타내는 경우).

[0101] 시퀀싱 데이터 세트에 대한 매치 스코어가 후보 서열에 대해 결정되면, 짧은 유전자 변이체를 갖는 후보 서열은 매치 스코어에 기초하여 선택될 수 있다 (예를 들어, 2개 이상의 후보 서열 중에서 가장 높은 가능성도 매치로 매치 스코어를 초래하는 후보 서열). 짧은 유전자 변이체는 예를 들어 개체의 하위집단 내에서 발견되는 변이체 또는 돌연변이, 또는 단일 또는 특정 개체에 고유한 변이체 또는 돌연변이일 수 있다. 짧은 유전자 변이체는 생식계열 변이체 또는 체세포 변이체일 수 있다. 짧은 유전자 변이체를 갖는 서열 핵산 분자로부터 발생하는 시퀀싱 데이터는 짧은 유전자 변이체를 갖는 후보 서열과 매치할 것이고, 해당 후보 서열이 선택될 수 있지만, 거부된 (또는 비선택된) 후보 서열(들)은 더 적은 가능성도 매치로 표시된 바와 같은 짧은 유전자 변이체를 포함하지 않는다 (해당 후보 서열에 대한 결정된 매치 스코어에 기초하여). 비선택된 후보 서열은 2개 이상의 연속 흐름 위치 또는 2개 이상의 비연속 흐름 위치일 수 있는 2개 이상의 흐름 위치에서 선택된 후보 서열 (이는 시퀀싱된 핵산 분자 시퀀싱 데이터 세트와 가장 잘 매치함)과 상이할 수 있다. 일부 실시양태에서, 비선택된 후보 서열은 3개 이상, 4개 이상, 5개 이상, 6개 이상, 7개 이상, 8개 이상, 9개 이상, 또는 10개 이상의 흐름 위치에서 선택된 후보 서열과 상이하다. 일부 실시양태에서, 비선택된 후보 서열은 1개 이상, 2개 이상, 3개 이상, 4개 이상, 또는 5개 이상의 흐름 사이클에 걸쳐 선택된 후보 서열과 상이하다. 일부 실시양태에서, 비선택된 후보 서열은 X개의 염기 위치에서 선택된 후보 서열과 상이하며, 여기서 서열 핵산 분자와 연관된 시퀀싱 데이터 세트는 (X+2)개 이상의 흐름 위치에서 비선택된 후보 서열과 상이하다. 선택된 및 비선택된 후보 서열 사이의 상이한 흐름 위치의 수의 증가 (여기서 시퀀싱된 핵산 분자 시퀀싱 데이터 세트는 선택된 후보 서열과 가장 잘 매치함)는 시퀀싱된 핵산 분자 시퀀싱 데이터 세트가 비선택된 후보 서열로 핵산 분자를 시퀀싱함으로써 야기될 가능성을 낮춘다.

[0102] 시퀀싱된 핵산 분자에 대한 시퀀싱 데이터 세트가 비선택된 후보 서열과 매치할 가능성도는 바람직하게는 0.05 미만, 0.04 미만, 0.03 미만, 0.02 미만, 0.01 미만, 0.005 미만, 0.001 미만, 0.0005 미만, 또는 0.0001 미만과 같이 낮다. 시퀀싱된 핵산 분자에 대한 시퀀싱 데이터 세트가 선택된 후보 서열과 매치할 가능성도는 바람직하게는 0.95 초과, 0.96 초과, 0.97 초과, 0.98 초과, 0.99 초과, 0.995 초과, 또는 0.999 초과와 같이 높다.

[0103] 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 일부 실시양태에서 복수의 시험 시퀀싱 데이터 세트를 분석하는 단계를 포함할 수 있으며, 각 시험 시퀀싱 데이터 세트는 시험 샘플에서 별도의 시험 핵산 분자와 연관된다. 핵산 분자는 예를 들어 핵산 분자의 서열이 참조 서열에 정렬된 경우 유전자좌에서 적어도 부분적으로 중첩된다. 핵산 분자의 적어도 일부는 상이한 시퀀싱 시작 위치 (유전자좌에 대해)를 가질 수 있으며, 이는 서열 내의 주어진 염기에 대한 상이한 흐름 위치 및/또는 상이한 흐름 순서 컨텍스트를 초래한다. 이러한 방식으로, 동일한 후보 서열은 복수의 시험 시퀀싱 데이터 세트를 분석하는데 사용될 수 있다. 각 후보 서열에 대해, 복수의 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타내는 매치 스코어가 결정될 수 있고, 가장 높은 가능성도 매치를 갖는 후보 서열 (및 이에 따라, 짧은 유전자 변이체 포함)이 선택될 수 있다. 복수의 시험 시퀀싱 데이터 세트를 사용하여 짧은 유전자 변이체를 검출하기 위한 예시적인 분석이 도 2a-2d에 표시된다. 도 2a에서, 3개의 시퀀싱된 시험 핵산 분자에 상응하는 서열 (R1, R2 및 R3, 각각은 연장된 프라이머의 서열로 표시됨)은 2개의 후보 서열 (H1 및 H2)과 연관된 중첩 유전자좌에서 참조 서열에 정렬된다. 도 2b, 도 2c 및 도 2d는 H1 (닫힌 원) 또는 H2 (열린 원)의 염기에 상응하는 시퀀싱 데이터 세트의 각 흐름 위치에서 선택된 통

계적 파라미터와 함께 각각 R1, R2 및 R3에 대한 예시적인 시퀀싱 데이터 세트를 보여준다.

[0104] 짧은 유전자 변이체의 존재 (또는 아이덴티티) 또는 부재는 하나 이상의 결정된 매치 스코어를 사용하여 시험 샘플에 대해 호출될 수 있다. 일부 실시양태에서, 예를 들어, 변이체를 갖는 것으로 분류된 단일 핵산 분자 (또는 연관된 시험 시퀀싱 데이터 세트)는 예를 들어 매치 스코어가 원하는 또는 미리 설정된 신뢰도로 후보 서열과의 매치를 나타내는 경우, 변이체의 존재, 아이덴티티 또는 부재를 호출하기에 충분할 수 있다. 일부 실시양태에서, 미리 결정된 수 (예를 들어, 1개 이상, 2개 이상, 3개 이상, 4개 이상, 5개 이상 등)의 핵산 분자 (또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트)는 변이체가 시험 샘플에 대해 호출되기 전에 변이체를 갖는 것으로 분류된다. 일부 실시양태에서, 핵산 분자 (또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트)의 수는 매치 스코어에 따라 동적으로 선택되고; 예를 들어, 높은 신뢰도 매치 스코어로 변이체를 갖는 것으로 분류된 단일 핵산 분자는 변이체를 호출하는데 사용될 수 있거나, 또는 더 낮은 신뢰도 매치 스코어로 변이체를 갖는 것으로 분류된 2개 이상의 핵산 분자는 변이체를 호출하는데 사용될 수 있다.

[0105] 임의로, 시퀀싱 데이터 세트에 대한 별도의 매치 스코어는 복수의 시험 시퀀싱 데이터 세트에 대한 매치 스코어를 결정하기 위해 집합적으로 분석된다. 예를 들어, 각 후보 서열에 대한 각 시험 시퀀싱 데이터 세트에 대한 매치 스코어가 본원에 기재된 방법을 사용하여 결정되면, 복수의 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타내는 매치 스코어는 공지된 베이지안 방법을 사용하여, 예를 들어 게놈 분석 툴킷 (Genome Analysis Toolkit; GATK)에 포함된 하플로타입콜러 (하플로타입콜러) 알고리즘을 사용하여 결정될 수 있고, 가장 높은 가능성도 매치를 갖는 후보 서열이 선택될 수 있다. 예를 들어, 문헌 [DePristo et al., A framework for variation discovery and genotyping using next-generation DNA sequencing data, Nature Genetics 43, 491-498 (2011)]; 및 [Poplin et al., Scaling accurate genetic variant discovery to tens of thousands of samples, bioRxiv, www.biorxiv.org/content/10.1101/201178v3 (July 24, 2018)]; [Hwang et al., Systematic comparison of variant calling pipelines using gold standard personal exome variants, Scientific Reports, vol. 5, no. 17875 (2015)]을 참조하며; 이들 각각의 내용은 본원에 포함된다.

[0106] 표적 변이체 및/또는 흐름-사이클 순서의 선택

[0107] 표적 짧은 유전자 변이체는 예를 들어 흐름 순서 및/또는 후보 서열을 선택하기 위한 기초로서 작용하도록 (즉, 표적 짧은 유전자 변이체를 미리 선택함으로써), 또는 다운스트림 분석을 위해 선택될 수 있다. 다운스트림 분석은 예를 들어 식별된 짧은 유전자 변이체를 포함하는 바이오마커 패널을 어셈블리하는 것을 포함할 수 있다. 바이오마커 패널은 시험 샘플과 연관된 개별 대상체에 대해 개인화될 수 있다. 예로서, 바이오마커 패널은 질병 (예를 들어 암)과 연관된 하나 이상의 짧은 유전자 변이체, 예를 들어 변이체 시그니처를 포함할 수 있다. 또 다른 예에서, 바이오마커 패널은 대상체에 대해 개인화되고, 대상체로부터의 샘플에서 이전에 검출된 하나 이상의 짧은 유전자 변이체를 포함하며, 이는 대상체에서 질병 (예컨대 암)에 기인할 수 있다.

[0108] 본원에 기재된 바와 같은 짧은 유전자 변이체를 식별하는 방법은 하나 이상의 표적 짧은 유전자 변이체가 미리 선택될 때 특히 유용할 수 있다. 주어진 짧은 유전자 변이체에 대한 검출 한계 (LOD)는 짧은 유전자 변이체의 서열 컨텍스트 (예를 들어, 표적 짧은 유전자 변이체 유전자좌를 플랭킹하는 핵산 분자의 서열) 및 핵산 분자를 시퀀싱하고 핵산 분자에 대한 시퀀싱 데이터 세트를 생성하는데 사용되는 흐름 순서 (또는 흐름 사이클 순서)에 의존할 수 있다. 즉, 주어진 흐름 순서, 짧은 유전자 변이체, 및 짧은 유전자 변이체 컨텍스트를 사용하여, 짧은 유전자 변이체를 갖는 핵산 분자 및 짧은 유전자 변이체를 갖지 않는 핵산 분자 (예를 들어, 참조 서열)에서 플로우스페이스에서 흐름 위치 분산의 수가 결정될 수 있다. 이는 특히 민감한 변이체의 선택 또는 높은 민감도로 특정 변이체를 검출할 수 있는 흐름 순서의 선택을 허용한다. 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 짧은 유전자 변이체를 갖지 않는 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 비교되어 표적 서열 및 참조 서열 사이에 존재하는 흐름 위치 차이의 수를 결정할 수 있다. 즉, 참조 서열은 표적 짧은 유전자 변이체를 제외하고는 표적 서열과 동일하다. 흐름 위치 차이의 수가 많을수록 해당 변이체에 대한 민감도 (즉, 검출 하한)가 더 높음을 나타낸다. 표적 및 참조 시퀀싱 데이터 세트는 표적 서열을 갖는 핵산 분자 및/또는 참조 서열을 갖는 핵산 분자를 실제로 시퀀싱함으로써 결정될 수 있거나, 또는 데이터 세트는 예상 시퀀싱 데이터 세트일 수 있다 (예를 들어, 인실리코에서 결정된 바와 같이).

[0109] 한 예에서, 특정 대상체 또는 암의 유전적 지문이 요구될 수 있지만, 대상체 또는 암의 게놈에서 짧은 유전자 변이체 각각 및 모두를 검출할 필요는 없다. 대신, 주어진 흐름 순서에 대해 특히 높은 민감도를 갖는 하나 이상의 짧은 유전자 변이체가 미리 선택될 수 있다. 민감한 변이체를 미리 선택함으로써, 시험 샘플에 대한 더 낮은 시퀀싱 깊이가 변이체를 확실하게 호출하는데 사용될 수 있다.

- [0110] 일부 실시양태에서, 시험 샘플에서 표적 짧은 유전자 변이체를 검출하는 방법은 표적 짧은 유전자 변이체를 선택하는 단계를 포함할 수 있으며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 이상의 비연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 연속적 또는 비연속적일 수 있는 3개 이상의 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 서열은 X개의 염기 위치에서 참조 서열과 상이하고, 여기서 표적 시퀀싱 데이터 세트는 (X+2)개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.
- [0111] 일부 실시양태에서, 시험 샘플에서 표적 짧은 유전자 변이체를 검출하는 방법은 표적 짧은 유전자 변이체를 선택하는 단계를 포함할 수 있으며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 표적 서열 및 참조 서열을 시퀀싱함으로써 수득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 이상의 비연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 연속적 또는 비연속적일 수 있는 3개 이상의 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 서열은 X개의 염기 위치에서 참조 서열과 상이하고, 여기서 표적 시퀀싱 데이터 세트는 (X+2)개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.
- [0112] 선택된 표적화된 짧은 유전자 변이체의 검출은 일반적으로 상기 논의된 바와 같이 진행할 수 있다. 예를 들어, 일부 실시양태에서, 표적 짧은 유전자 변이체의 유전자좌를 갖는 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 수득될 수 있다. 시퀀싱 데이터는 표적 및 참조 시퀀싱 데이터 세트를 생성하는데 사용되는 동일한 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 생성된다. 시험 시퀀싱 데이터 세트가 짧은 유전자 변이체를 갖는 표적 서열과 매치할 가능성을 나타내는 매치 스코어 (또는, 대안적으로 또는 추가적으로, 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어)가 결정되고, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재는 결정된 매치 스코어를 사용하여 호출될 수 있다.
- [0113] 일부 실시양태에서, 표적 짧은 유전자 변이체는 복수의 시험 시퀀싱 데이터 세트를 사용하여 시험 샘플에서 검출되며, 각 시험 시퀀싱 데이터 세트는 시험 샘플 내의 상이한 시험 핵산 분자와 연관된다. 분석된 시험 핵산 분자는 표적 짧은 유전자 변이체 유전자좌에서 중첩되고, 데이터 세트는 표적 짧은 유전자 변이체를 선택하는데 사용되는 동일한 흐름-사이클 순서를 사용하여 시험 핵산 분자를 시퀀싱함으로써 생성된다. 복수의 시험 시퀀싱 데이터 세트가 짧은 유전자 변이체를 갖는 표적 서열과 매치할 가능성을 나타내는 매치 스코어 (또는, 대안적으로 또는 추가적으로, 복수의 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어)가 결정되고, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재는 결정된 매치 스코어를 사용하여 호출될 수 있다.
- [0114] 일부 실시양태에서, 시퀀싱 데이터를 생성하는데 사용되는 흐름 순서 또는 흐름-사이클 순서가 미리 선택된다. 본원에서 논의된 바와 같이, 흐름 순서에서 변이체의 컨텍스트는 변이체 서열 및 비교된 (예를 들어, 참조) 서열 사이의 신호 차이에 영향을 미칠 수 있다. 선택된 표적 변이체를 검출할 가능성을 증가시키기 위해, 흐름 순서 또는 흐름-사이클 순서가 미리 선택될 수 있다.
- [0115] 도 3은 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 예시적인 방법의 흐름 차트를 보여준다. 단계 302에서, 표적 짧은 유전자 변이체가 선택된다. 표적 짧은 유전자 변이체는, 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터가 2개 초과인 흐름 위치에서 참조 서열과 연관된 시퀀싱 데이터 세트와 상이하도록 선택되고, 여기

서 흐름 위치는 뉴클레오타이드 흐름에 상응한다. 단계 304에서, 하나 이상의 시험 시퀀싱 데이터 세트는 예를 들어 하나 이상의 시험 시퀀싱 데이터 세트를 획득하기 위해 하나 이상의 시험 핵산 분자를 시퀀싱함으로써, 또는 하나 이상의 시험 시퀀싱 데이터 세트를 수신함으로써 획득된다. 각각의 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 시험 핵산 분자와 연관된다. 선택된 표적 짧은 유전자 변이체의 분석을 위해, 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관된 유전자좌와 적어도 부분적으로 중첩된다. 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정될 수 있으며 (또는 이전에 결정되었을 수 있음), 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함한다. 단계 306에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 매치 스코어가 결정된다. 매치 스코어는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타낸다. 대안적으로, 매치 스코어는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타낼 수 있다. 단계 308에서, 하나 이상의 결정된 매치 스코어는 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는데 사용된다.

[0116] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과인 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과인 흐름 위치 (예를 들어, 2개 초과인 연속 흐름 위치 또는 2개 초과인 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0117] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과인 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과인 흐름 위치 (예를 들어, 2개 초과인 연속 흐름 위치 또는 2개 초과인 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0118] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유

전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0119] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0120] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리

선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0121] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 미리 선택된 흐름-사이클 순서에 따라 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0122] 표적 변이체 및/또는 흐름-사이클 순서의 선택

[0123] 흐름-사이클 순서는 4개의 염기 흐름 사이클 (예를 들어, A, G, C 및 T 각각의 하나, 임의의 반복 순서로)로 제한될 필요는 없으며, 사이클에서 4개 초과와 염기 유형을 갖는 연장된 흐름 사이클일 수 있다. 연장된 사이클 순서는 시퀀싱 프라이머를 연장하기 위해 원하는 수의 사이클 동안 반복될 수 있다. 예로서, 일부 실시양태에서, 연장된 흐름 순서는 흐름 사이클 순서에서 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 별도의 뉴클레오타이드 흐름을 포함한다. 사이클은 A, G, C 및 T 각각을 적어도 하나 포함할 수 있지만, 사이클이 반복되기 전에 사이클 내에서 하나 이상의 염기 유형을 반복한다.

[0124] 연장된 흐름 사이클 순서는 4개의 반복된 염기를 갖는 흐름 사이클 순서보다 더 큰 비율의 작은 게놈 변이체 (예를 들어, SNP)를 검출하는데 유용할 수 있다. 예를 들어, XYZ → XQZ 형식의 치환 SNP의 192개의 유효한 구성 (여기서, Q≠Y (및 Q, X, Y 및 Z는 각각 A, C, G 및 T 중 어느 하나임))이 있다. 이들 중, 168개는 시퀀싱 데이터 세트 (예를 들어, 흐름도)에서 새로운 신호 (즉, 새로운 년-제로 신호 또는 새로운 제로 신호)를 생성할 수 있다. 민감한 흐름 순서와 조합된 새로운 제로 또는 년-제로 신호는 참조와 관련하여 변이체에서 동일한 후행 서열이 주어지면, 다중 흐름 위치 (예를 들어, 사이클의 길이 초과로 연장할 수 있는 흐름 이동 또는 사이클 이동)에 대해 전파되는 신호를 생성할 수 있다. 호모폴리머 길이 변화보다는 호모폴리머의 삽입 또는 결실이 신호 차이 전파를 초래할 수 있다는 점에 주목해야 한다. 나머지 24개의 변이체는 영향을 받는 흐름 위치에서 호모폴리머 길이 변화를 유발하지만, 이러한 변화는 전파된 신호 변화를 유발하지 않는다. 그러므로, 이론적 최대 87.5%의 SNP는 2개 초과와 흐름 위치에 대해 참조 (또는 후보) 서열과 상이한 새로운 신호를 초래할 수 있다. 상기 논의된 바와 같이, 전파된 신호 차이는 시험 시퀀싱 데이터 세트 및 틀리게 매치된 후보 서열 사이의 가능성도 차이를 증가시킨다. 또한, 전파된 신호 변화는 변이체에 걸친 흐름 순서에 의존한다.

[0125] 무작위로 단편화된 시험 샘플에서 핵산 분자를 시퀀싱하는 것은 시퀀싱 프라이머가 흐름 순서를 사용하여 연장될 때 변이체의 흐름 순서 컨텍스트에서 무작위 이동을 초래한다. 즉, 변이체의 흐름 위치는 시퀀싱된 핵산 분자의 시작 위치에 따라 달라질 수 있다. 핵산 분자 서열에서 모든 시퀀싱 시작 위치가 활용되더라도, 모든 흐름 사이클 조합이 SNP의 모든 87.5%에 대해 2개 초과와 흐름 위치에서 신호 변화를 검출할 수 있는 것은 아니다. 예를 들어, 4-염기 흐름 사이클 순서 T-A-C-G는 SNP의 41.7%에 대해 2개 초과와 흐름 위치에서 참조

시퀀싱 데이터 세트와 상이한 시험 시퀀싱 데이터 세트를 초래할 수 있다. 본원에서 추가로 논의된 바와 같이, 연장된 흐름 사이클 순서는 충분히 높은 시퀀싱 깊이 (즉, 충분히 많은 수의 시작 위치의 샘플링)가 주어지면, 모든 이론적 최대 SNP (즉, 가능한 SNP의 87.5%, 또는 호모폴리머 길이 변화를 초래하는 것 이외의 모든 SNP)가 시험 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트 사이의 2개 초과 흐름 위치에서 차이를 일으킬 수 있도록 설계되었다.

[0126] 연장된 시퀀싱 흐름 순서는 상이한 효율을 가질 수 있다 (즉, 인간 참조 게놈을 시퀀싱하는데 사용될 때 흐름 당 혼입의 평균 수). 일부 실시양태에서, 흐름 순서는 약 0.6 이상 (예컨대 약 0.62 이상, 약 0.64 이상, 약 0.65 이상, 약 0.66 이상, 또는 약 0.67 이상)의 효율을 갖는다. 일부 실시양태에서, 흐름 순서는 약 0.6 내지 약 0.7의 효율을 갖는다. 흐름 사이클 순서 및 해당 추정 효율의 예는 표 2에 나타난다.

[0127] 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 5%에 대한 SNP 순열의 약 50% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 5%에 대한 SNP 순열의 약 60% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자 (즉, "흐름 위상")와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 5%에 대한 SNP 순열의 약 70% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 5%에 대한 SNP 순열의 약 80% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다.

[0128] 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 10%에 대한 SNP 순열의 약 50% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 10%에 대한 SNP 순열의 약 60% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 10%에 대한 SNP 순열의 약 70% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 10%에 대한 SNP 순열의 약 80% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다.

[0129] 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 20%에 대한 SNP 순열의 약 50% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 20%에 대한 SNP 순열의 약 60% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 20%에 대한 SNP 순열의 약 70% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 20%에 대한 SNP 순열의 약 80% 내지 87.5%에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 흐름 위치에서 신호 차이를 생성하도록 선택된다.

- [0130] 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 30%에 대한 SNP 순열의 약 50% 내지 87.5% (또는 약 50% 내지 약 80%)에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 의 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 30%에 대한 SNP 순열의 약 60% 내지 87.5% (또는 약 60% 내지 약 80%)에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 의 흐름 위치에서 신호 차이를 생성하도록 선택된다. 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 무작위 시퀀싱 시작 위치의 적어도 30%에 대한 SNP 순열의 약 70% 내지 87.5% (또는 약 70% 내지 약 80%)에 대해 SNP에 의해 상이한 핵산 분자와 연관된 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 의 흐름 위치에서 신호 차이를 생성하도록 선택된다.
- [0131] 일부 실시양태에서, 연장된 시퀀싱 흐름 순서는 표 2의 연장된 시퀀싱 흐름 순서 중 임의의 하나이다. "이동 민감도"는 모든 가능한 SNP 순열에 대해 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 의 흐름 위치에서 신호 차이를 생성하기 위한 최대 민감도를 지칭한다. "최대 이동 민감도"는 민감도가 유지되는 흐름 위상의 가장 높은 비율에서 모든 가능한 SNP 순열에 대해 2개의 시퀀싱 데이터 세트 (예를 들어, 시험 또는 표적 시퀀싱 데이터 세트 및 후보 또는 참조 시퀀싱 데이터 세트) 사이의 2개 초과 의 흐름 위치에서 신호 차이를 생성하기 위한 최대 민감도를 지칭한다.

호들 사이를 순서	추정 효율	최대 이동 민감도	호들 위상의 5%에서 이동 민감도	호들 위상의 10%에서 이동 민감도	호들 위상의 20%에서 이동 민감도	호들 위상의 30%에서 이동 민감도
T-C-A-G-A-T-G-C-A-T-G-C-T-A-C-G	67.5%	82.3% @ 19%	82.3%	82.3%	75.0%	66.7%
T-C-A-C-G-A-T-G-C-A-T-G-C-T-A-G	67.5%	83.3% @ 12%	83.3%	83.3%	72.9%	62.5%
T-C-A-T-G-C-A-T-G-C-T-A-C-G-A-G	67.3%	82.3% @ 12%	82.3%	82.3%	72.9%	67.7%
T-C-A-G-T-A-C-G-A-T-G-C-A-T-G-C	67.3%	82.3% @ 12%	82.3%	82.3%	75.0%	63.5%
T-C-A-G-T-C-G-A-T-G-A-C-T-A-G-C	67.2%	81.3% @ 12%	81.3%	81.3%	74.0%	69.8%
T-C-A-T-C-G-A-C-T-G-A-G-C-T-A-G	67.2%	81.3% @ 12%	81.3%	81.3%	74.0%	69.8%
T-C-G-T-A-G-C-T-G-A-C-A-T-G-C-A	67.2%	83.3% @ 12%	83.3%	83.3%	75.0%	67.7%
T-C-G-T-A-G-C-A-T-G-C-T-A-C-G-A	67.0%	79.2% @ 25%	79.2%	79.2%	79.2%	75.0%
T-C-A-T-G-C-A-G-T-C-G-A-C-T-A-G	66.9%	83.3% @ 19%	83.3%	83.3%	75.0%	68.8%
T-C-A-T-G-C-A-T-C-G-T-A-C-G-A-G-C-T-A-G	66.7%	86.5% @ 7%	86.5%	85.4%	85.4%	69.8%
T-C-G-A-C-T-G-T-A-G-C-T-A-G-C-A	66.7%	82.3% @ 19%	82.3%	82.3%	75.0%	66.7%
T-C-A-C-G-A-T-G-C-T-A-G-C-T-A-G	66.5%	82.3% @ 12%	82.3%	82.3%	75.0%	67.7%

<표 2>

[0132]

[0133]

호름 사이클 순서	추정 효율	최대 이동 민감도	호름 위상의 5%에서 이동 민감도	호름 위상의 10%에서 이동 민감도	호름 위상의 20%에서 이동 민감도	호름 위상의 30%에서 이동 민감도
T-C-A-G-T-A-C-G-A-T-G-C-T-A-C-G	66.4%	83.3% @ 19%	83.3%	83.3%	75.0%	68.8%
T-C-G-A-C-T-A-G-C-A-T-G-C-A-T-G	66.0%	81.3% @ 12%	81.3%	81.3%	70.8%	62.5%
T-A-C-G	66.0%	41.7% @ 100%	41.7%	41.7%	41.7%	41.7%
T-C-A-G-C-T-G-A-C-T-A-G-T-C-A-T-G-A-C-T-A-G-C-G-A-T-C-G	65.7%	87.5% @ 11%	87.5%	87.5%	82.3%	75.0%
T-C-T-A-G-C-A-T-G-A-C-T-G-A-C-G	65.7%	83.3% @ 12%	83.3%	83.3%	71.9%	63.5%
T-C-G-A-C-T-A-T-G-C-A-T-G-C-A-G	65.5%	81.3% @ 19%	81.3%	81.3%	71.9%	63.5%
T-C-G-A-C-T-G-C-A-T-T-C-G-A-T-G-C-A-G-T-A-C-T-A-G	65.4%	87.5% @ 12%	87.5%	87.5%	82.3%	74.0%
T-C-A-C-T-G-A-C-G-T-A-G-C-T-A-T-G-C-A-T-C-G-A-G	65.3%	84.4% @ 17%	84.4%	84.4%	83.3%	76.0%
T-C-A-T-G-C-T-A-G-C-T-A-G-T-A-C-G-T-A-C-G-A-G-C-A-T-C-G	65.2%	86.5% @ 11%	86.5%	86.5%	82.3%	78.1%
T-C-G-A-T-G-C-A-T-C-G-T-A-C-T-A-G-C-A-G-T-G-A-C	65.2%	87.5% @ 8%	87.5%	86.5%	84.4%	71.9%
T-C-A-T-G-A-G-C-T-A-G-C-A-T-C-G-T-A-C-T-G-A-C-G	65.2%	87.5% @ 8%	87.5%	86.5%	81.3%	70.8%
T-C-A-G-C-A-T-G-T-A-C-T-G-A-T-G-C-A-T-C-G-A-G-C-T-A-C-G	65.0%	87.5% @ 11%	87.5%	87.5%	82.3%	77.1%
T-C-A-G-T-A-C-T-A-G-C-A-T-G-C-G-A-T-C-G-T-A-G-C-T-G-A-C	65.0%	86.5% @ 11%	86.5%	86.5%	78.1%	74.0%

호름 사이클 순서	추정 효율	최대 이동 민감도	호름 위상의 5%에서 이동 민감도	호름 위상의 10%에서 이동 민감도	호름 위상의 20%에서 이동 민감도	호름 위상의 30%에서 이동 민감도
T-C-A-C-G-T-A-G-C-T-A-T-G-C-T-G-A-C-T-G-A-C-A-T-G-A-C-T-A-G-C-G	64.6%	85.4% @ 9%	85.4%	84.4%	76.0%	61.5%
T-C-A-G-C-T-A-T-G-A-C-T-G-A-C-T-G-A-C-T-G-A-C-T-A-G-C-G	64.5%	85.4% @ 12%	85.4%	85.4%	77.1%	74.0%
T-C-A-G-C-T-A-C-T-G-C-A-T-G-A-C-T-G-A-C-T-A-G-C-G-A	64.5%	87.5% @ 14%	87.5%	87.5%	83.3%	70.8%
T-C-A-G-A-C-T-A-G-C-G-A-T-G-C-A-T-G-T-C-T-A-G-T-C-A-C-G	64.5%	86.5% @ 11%	86.5%	86.5%	83.3%	62.5%
T-C-A-T-C-G-A-C-T-G-C-G-A-T-G-C-T-A-G-T-A-C-A-G	64.4%	85.4% @ 17%	85.4%	85.4%	83.3%	72.9%
T-C-A-C-G-T-A-C-T-G-A-C-A-T-G-C-A-T-G-C-T-A-G-T-A-G-C-G-A-T-C-G	64.4%	85.4% @ 9%	85.4%	84.4%	83.3%	72.9%
T-C-A-G-T-G-C-T-A-C-G-T-C-A-C-G-A-T-C-A-G-A-T-G-C-T-A-G	64.4%	86.5% @ 11%	86.5%	86.5%	71.9%	67.7%
T-C-A-G-C-G-A-T-G-A-C-T-A-G-C-T-A-C-G-T-C-A-T-G	64.4%	85.4% @ 17%	85.4%	85.4%	84.4%	66.7%
T-C-A-T-G-C-T-A-C-G-A-G	64.4%	81.3% @ 17%	81.3%	81.3%	80.2%	66.7%
T-C-A-T-G-A-C-G-T-A-C-G-A-C-T-C-A-T-G-C-A-G-T-G-C-T-A-G	64.3%	85.4% @ 11%	85.4%	85.4%	82.3%	75.0%
T-C-A-G-T-C-G-A-T-G-C-T-A-C-T-G-C-A-T-A-C-G-T-C-G-A-T-G-A-C-A-G	64.3%	87.5% @ 9%	87.5%	86.5%	83.3%	74.0%
T-C-G-A-T-G-C-T-A-C-A-G	64.3%	81.3% @ 17%	81.3%	81.3%	80.2%	66.7%
T-C-A-G-T-C-G-A-C-A-T-G-C-A-T-C-G-A-T-A-C-G-T-G-C-T-A-G-C-T-A-G	64.2%	87.5% @ 9%	87.5%	86.5%	79.2%	70.8%

일부 실시양태에서, 핵산 분자를 시퀀싱하는 방법은 하기를 포함한다: (a) 핵산 분자를 프라이머에 혼성화하여 혼성화된 주형을 형성하는 단계; (b) 5개 이상의 별도의 뉴클레오티드 호름을 포함하는 반복된 호름-사이클 순서에 따라 별도의 뉴클레오티드 호름에 제공된 표지된 비종결 뉴클레오티드를 사용하여 프라이머를 연장하는 단계; 및 (c) 프라이머가 뉴클레오티드 호름에 의해 연장됨에 따라 혼입된 표지된 뉴클레오티드로부터의 신호 또는 신호의 부재를 검출하는 단계. 일부 실시양태에서, 호름-사이클 순서는 무작위 시퀀싱 시작 위치의 5%에서 가능한 SNP 순열의 50% 이상에 대해 2개 초과인 호름 위치에서 신호 변화를 유도한다. 일부 실시양태에서, 유도된 신호 변화는 신호 강도의 변화이거나, 또는 새로운 실질적으로 제로 (또는 새로운 제로) 또는 새로운 실질적으로 넌-제로 (또는 새로운 넌-제로) 신호이다. 일부 실시양태에서, 유도된 신호 변화는 새로운 실질적으로 제로 (또는 새로운 제로) 또는 새로운 실질적으로 넌-제로 (또는 새로운 넌-제로) 신호이다. 일부 실시양태에서, 호름-사이클 순서는 호름 당 0.6 이상의 염기 혼입 효율을 갖는다. 일부 실시양태에서, 호름-사이클은 표 2에 나열된 호름-사이클 순서 중 임의의 하나이다.

상이한 호름 순서를 사용한 재시퀀싱

검출된 짧은 유전자 변이체의 민감도는 핵산 분자를 시퀀싱하는데 사용되는 호름 사이클 순서에 의존하기 때문에, 본원에 기재된 방법은 2개 이상의 상이한 호름 사이클 순서를 사용하여 시퀀싱된 시험 핵산 분자 (또는 중첩 유전자좌를 갖는 복수의 핵산 분자)를 분석하도록 적응될 수 있다. 매치 스코어는 하나 이상의 후보 서열에 대한 2개 이상의 상이한 시퀀싱 데이터 세트 (상이한 호름 사이클 순서로부터 생성됨)의 매치에 기초하여 결정

될 수 있다. 변이체의 존재 또는 부재가 호출되고/거나, 후보 서열이 상기 논의된 바와 같이 매치 스코어에 기초하여 선택될 수 있다.

[0139] 상기 방법은 제1 흐름-사이클 순서를 사용하여 시퀀싱된 시험 샘플로부터 유래된 시험 핵산 분자와 연관된 제1 시험 시퀀싱 데이터 세트, 및 제2 흐름-사이클 순서를 사용하여 시퀀싱된 동일한 시험 핵산 분자와 연관된 제2 시험 시퀀싱 데이터 세트를 획득하는 단계를 포함할 수 있다. 예를 들어, 시험 핵산 분자는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 비중결 핵산 분자를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오타이드 흐름 후에 시퀀싱 프라이머로 뉴클레오타이드 혼입의 존재 또는 부재를 검출하여 제1 시험 시퀀싱 데이터 세트를 생성하는 단계; 연장된 시퀀싱 프라이머를 제거하는 단계; 및 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 비중결 뉴클레오타이드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오타이드 흐름 후에 시퀀싱 프라이머로 뉴클레오타이드 혼입의 존재 또는 부재를 검출하여 제2 시험 시퀀싱 데이터 세트를 생성함으로써 동일한 시험 핵산 분자를 시퀀싱하는 단계에 의해 시퀀싱될 수 있다.

[0140] 핵산 분자는 상이한 흐름-사이클 순서를 사용하여 시퀀싱되기 때문에, 시퀀싱 데이터 세트는 상이하다. 도 4a 및 도 4b는 제1 흐름-사이클 순서 (T-A-C-G) (도 4a) 및 제2 흐름-사이클 순서 (A-G-C-T) (도 4b)를 사용하여 결정된 TATGGTCGTCGA (서열식별번호: 1)의 연장된 프라이머 서열을 갖는 핵산 분자에 대한 예시적인 시퀀싱 데이터 세트를 보여준다. 알 수 있는 바와 같이, 도 4a 및 도 4b의 시퀀싱 데이터 세트는 핵산 분자 서열이 변하지 않더라도 흐름-사이클 순서의 차이로 인해 상이하다. 시퀀싱 데이터 세트 내에서, 제1 후보 연장된 프라이머 서열 TATGGTCGTCGA (서열식별번호: 1) (단편 원) 및 제2 후보 연장된 프라이머 서열 TATGGTCATCGA (서열식별번호: 2) (열린 원)의 염기 카운트에 상응하는, 각 흐름 위치에서 통계적 파라미터가 선택될 수 있다. 도 4a 및 도 4b는 흐름 사이클 순서가 변이체 검출 민감도에 대해 갖는 유의한 변화를 나타낸다. 예를 들어, 제1 흐름 사이클 순서를 사용하는 제1 후보 서열 및 제2 후보 서열 사이의 차이는 흐름 위치 12-20에서 명백한 반면 (도 4a), 제1 흐름 사이클 순서를 사용하는 제1 후보 서열 및 제2 후보 서열 사이의 차이는 위치 17 및 18에서만 명백하다 (도 4b).

[0141] 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트가 하나 이상의 후보 서열 (예를 들어, 미리 선택된 표적 짧은 유전자 변이체를 갖는 표적 서열, 미리 선택된 표적 짧은 유전자 변이체가 없는 서열을 갖는 참조 서열, 또는 다른 가능한 후보 서열 (예컨대 반수체형))과 매치할 가능성을 나타내는 매치 스코어가 결정될 수 있고, 표적 짧은 유전자 변이체의 존재 또는 부재가 호출될 수 있거나, 또는 후보 서열이 선택될 수 있다.

[0142] 본원에서 논의된 바와 같이, 이 공정은 공통 유전자좌에서 중첩되는 복수의 상이한 시험 핵산 분자를 시퀀싱할 때 사용될 수 있다. 예를 들어, 복수의 제1 시험 시퀀싱 데이터 세트 (각 시험 시퀀싱 데이터 세트는 제1 흐름 사이클 순서를 사용하여 시퀀싱된 시험 핵산 분자와 연관됨)가 획득될 수 있고, 복수의 제2 시험 시퀀싱 데이터 세트 (각 시험 시퀀싱 데이터 세트는 제2 흐름 사이클 순서를 사용하여 시퀀싱된 동일한 핵산 분자와 연관됨)가 획득될 수 있다. 제1 흐름 사이클 순서 및 제2 흐름 사이클 순서는 상이하다. 복수의 제1 시퀀싱 데이터 세트 및 복수의 제2 시퀀싱 데이터 세트가 하나 이상의 후보 서열 (예를 들어, 미리 선택된 표적 짧은 유전자 변이체를 갖는 표적 서열, 미리 선택된 표적 짧은 유전자 변이체가 없는 서열을 갖는 참조 서열, 또는 다른 가능한 후보 서열 (예컨대 반수체형))과 매치할 가능성을 나타내는 매치 스코어가 결정될 수 있고, 표적 짧은 유전자 변이체의 존재 또는 부재가 호출될 수 있거나, 또는 후보 서열이 선택될 수 있다.

[0143] 도 5는 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하기 위한 예시적인 방법을 보여준다. 단계 502에서, 하나 이상의 제1 시험 시퀀싱 데이터 세트가 획득된다. 하나 이상의 제1 시험 시퀀싱 데이터 세트는 예를 들어 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수신함으로써, 또는 하나 이상의 핵산 분자를 시퀀싱함으로써 획득될 수 있다. 제1 시험 시퀀싱 데이터 세트 각각은 시험 샘플로부터 유래된 상이한 핵산 분자와 연관된다. 제1 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정된다. 생성된 하나 이상의 제1 시험 시퀀싱 데이터 세트는 각각 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함한다. 단계 504에서, 하나 이상의 제2 시험 시퀀싱 데이터 세트가 획득된다. 하나 이상의 제2 시험 시퀀싱 데이터 세트는 예를 들어 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수신함으로써, 또는 하나 이상의 핵산 분자를 시퀀싱함으로써 획득될 수 있다. 제2 시험 시퀀싱 데이터 세트 각각은 제1 시험 시퀀싱 데이터 세트와 동일한 핵산 분자와 연관된다. 즉, 핵산 분자는 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트 둘 모두와 연관된다. 제2 시퀀싱 데이터 세트는 제1 흐름-사이클 순서와 상이한 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정된다. 생성된 하나 이상의 제2 시험 시퀀싱 데이터 세트는 각각 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포

함한다. 단계 506에서, 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 매치 스코어가 결정된다. 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 시퀀싱 데이터 세트 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치함을 나타낸다. 단계 508에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재는 결정된 매치 스코어를 사용하여 호출된다.

[0144] 도 6은 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하기 위한 또 다른 예시적인 방법을 보여준다. 단계 602에서, 표적 짧은 유전자 변이체가 선택된다. 표적 짧은 유전자 변이체는, 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서, 또는 둘 모두에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터가 2개 초과인 흐름 위치에서 참조 서열과 연관된 시퀀싱 데이터 세트와 상이하도록 선택되고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한다. 단계 604에서, 하나 이상의 제1 시험 시퀀싱 데이터 세트가 획득된다. 하나 이상의 제1 시험 시퀀싱 데이터 세트는 예를 들어 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수신함으로써, 또는 하나 이상의 핵산 분자를 시퀀싱함으로써 획득될 수 있다. 제1 시험 시퀀싱 데이터 세트 각각은 시험 샘플로부터 유래된 상이한 핵산 분자와 연관된다. 제1 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정된다. 생성된 하나 이상의 제1 시험 시퀀싱 데이터 세트는 각각 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함한다. 단계 606에서, 하나 이상의 제2 시험 시퀀싱 데이터 세트가 획득된다. 하나 이상의 제2 시험 시퀀싱 데이터 세트는 예를 들어 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수신함으로써, 또는 하나 이상의 핵산 분자를 시퀀싱함으로써 획득될 수 있다. 제2 시험 시퀀싱 데이터 세트 각각은 제1 시험 시퀀싱 데이터 세트와 동일한 핵산 분자와 연관된다. 즉, 핵산 분자는 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트 둘 모두와 연관된다. 제2 시퀀싱 데이터 세트는 제1 흐름-사이클 순서와 상이한 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정된다. 생성된 하나 이상의 제2 시험 시퀀싱 데이터 세트는 각각 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함한다. 단계 608에서, 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 매치 스코어가 결정된다. 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 시퀀싱 데이터 세트 또는 둘 모두가 하나 이상의 후보 서열 (이는 예를 들어 참조 서열을 포함할 수 있음)로부터의 후보 서열과 매치함을 나타낸다. 단계 610에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재는 결정된 매치 스코어를 사용하여 호출된다.

[0145] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0146] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (b) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여 (여기서, 제2 흐름-사이클 순서는 제1 흐름-사이클 순서와 상이함), 뉴클레오티드 흐름에 상응하는 흐름

를 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0147] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (d) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (e) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0148] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (b) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여 (여기서, 제2 흐름-사이클 순서는 제1 흐름-사이클 순서와 상이함), 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (d) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (e) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0149] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드

흐름에 상응한 것인 단계; (b) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (e) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0150] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (c) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (e) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0151] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열 (이는 참조 서열을 포함할 수 있음)에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과

매치할 가능성을 나타내는 것인 단계; (e) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (f) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0152] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 검출하는 방법은 하기를 포함한다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (c) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (e) 2개 이상의 상이한 후보 서열 (이는 참조 서열을 포함할 수 있음)로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (f) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0153] 시스템, 디바이스 및 보고서

[0154] 도면을 참조하여 기재된 것을 포함하여 상기 기재된 작동은 임의로 도 7에 도시된 하나 이상의 구성성분에 의해 구현된다. 다른 공정, 예를 들어 상기 기재된 작동의 전부 또는 일부의 조합 또는 하위조합이 도 7에 도시된 구성성분에 기초하여 구현될 수 있다는 것은 관련 기술분야의 통상의 기술자에게 명백할 것이다. 또한, 상기 방법, 기술, 시스템 및/또는 디바이스가 도 7에 도시된 구성성분에 의해 구현되고/거나 이에 의해 제공되는지 여부에 관계 없이 본원에 기재된 방법, 기술, 시스템 및 디바이스가 전체적으로 또는 부분적으로 서로 조합될 수 있는 방법이 관련 기술분야의 통상의 기술자에게 명백할 것이다.

[0155] 도 7은 한 실시양태에 따른 컴퓨팅 디바이스의 한 예를 예시한다. 디바이스 (700)는 네트워크에 연결된 호스트 컴퓨터일 수 있다. 디바이스 (700)는 클라이언트 컴퓨터 또는 서버일 수 있다. 도 7에 나타난 바와 같이, 디바이스 (700)는 임의의 적합한 유형의 마이크로프로세서-기반 디바이스, 예컨대 개인용 컴퓨터, 워크스테이션, 서버, 또는 핸드헬드 컴퓨팅 디바이스 (휴대용 전자 디바이스), 예컨대 전화 또는 태블릿일 수 있다. 디바이스는 예를 들어 프로세서 (710), 입력 디바이스 (720), 출력 디바이스 (730), 스토리지 (740), 및 통신 디바이스 (760) 중 하나 이상을 포함할 수 있다. 입력 디바이스 (720) 및 출력 디바이스 (730)는 일반적으로 상기 기재된 것들에 상응할 수 있고, 컴퓨터와 연결가능하거나 통합될 수 있다.

[0156] 입력 디바이스 (720)는 입력을 제공하는 임의의 적합한 디바이스, 예컨대 터치 스크린, 키보드 또는 키패드, 마우스, 또는 음성-인식 디바이스일 수 있다. 출력 디바이스 (730)는 출력을 제공하는 임의의 적합한 디바이스, 예컨대 터치 스크린, 햅틱 디바이스, 또는 스피커일 수 있다.

[0157] 스토리지 (740)는 저장을 제공하는 임의의 적합한 디바이스, 예컨대 RAM, 캐시, 하드 드라이브, 또는 제거가능

한 저장 디스크를 포함하는 전기, 자기 또는 광학 메모리일 수 있다. 통신 디바이스 (760)는 네트워크를 통해 신호를 전송 및 수신할 수 있는 임의의 적합한 디바이스, 예컨대 네트워크 인터페이스 칩 또는 디바이스를 포함할 수 있다. 컴퓨터의 구성성분은 임의의 적합한 방식으로, 예컨대 물리적 버스를 통해 또는 무선으로 연결될 수 있다.

[0158] 스토리지 (740)에 저장되고 프로세서 (710)에 의해 실행될 수 있는 소프트웨어 (750)는 예를 들어 본 개시내용의 기능성을 구현하는 프로그래밍을 포함할 수 있다 (예를 들어, 상기 기재된 바와 같은 디바이스에서 구현된 바와 같이).

[0159] 소프트웨어 (750)는 또한 명령 실행 시스템, 장치 또는 디바이스로부터 소프트웨어와 연관된 명령을 폐치하고 명령을 실행할 수 있는, 명령 실행 시스템, 장치 또는 디바이스, 예컨대 상기 기재된 것들에 의해 또는 이와 관련하여 사용하기 위한 임의의 비일시적 컴퓨터-판독가능 저장 매체 내에 저장 및/또는 전송될 수 있다. 본 개시내용의 문맥에서, 컴퓨터-판독가능 저장 매체는 명령 실행 시스템, 장치 또는 디바이스에 의해 또는 이와 관련하여 사용하기 위한 프로그래밍을 함유 또는 저장할 수 있는 임의의 매체, 예컨대 스토리지 (740)일 수 있다.

[0160] 소프트웨어 (750)는 또한 명령 실행 시스템, 장치 또는 디바이스로부터 소프트웨어와 연관된 명령을 폐치하고 명령을 실행할 수 있는, 명령 실행 시스템, 장치 또는 디바이스, 예컨대 상기 기재된 것들에 의해 또는 이와 관련하여 사용하기 위한 임의의 전송 매체 내에서 전파될 수 있다. 본 개시내용의 문맥에서, 전송 매체는 명령 실행 시스템, 장치 또는 디바이스에 의해 또는 이와 관련하여 사용하기 위한 프로그래밍을 통신, 전파 또는 전송할 수 있는 임의의 매체일 수 있다. 전송 판독가능 매체는 전자, 자기, 광학, 전자기 또는 적외선 유선 또는 무선 전파 매체를 포함할 수 있지만 이에 제한되지는 않는다.

[0161] 디바이스 (700)는 임의의 적합한 유형의 상호연결된 통신 시스템일 수 있는 네트워크에 연결될 수 있다. 네트워크는 임의의 적합한 통신 프로토콜을 구현할 수 있고, 임의의 적합한 보안 프로토콜에 의해 보장될 수 있다. 네트워크는 네트워크 신호의 전송 및 수신을 구현할 수 있는 임의의 적합한 배열의 네트워크 링크, 예컨대 무선 네트워크 연결, T1 또는 T3 라인, 케이블 네트워크, DSL, 또는 전화선을 포함할 수 있다.

[0162] 디바이스 (700)는 네트워크 상에서 작동하기에 적합한 임의의 운영 시스템을 구현할 수 있다. 소프트웨어 (750)는 임의의 적합한 프로그래밍 언어, 예컨대 C, C++, 자바 또는 파이썬으로 작성될 수 있다. 다양한 실시양태에서, 본 개시내용의 기능성을 구현하는 애플리케이션 소프트웨어는 예를 들어, 상이한 구성으로, 예컨대 클라이언트/서버 배열에서 또는 웹-기반 애플리케이션 또는 웹 서비스와 같은 웹 브라우저를 통해 전개될 수 있다.

[0163] 본원에 기재된 방법은 임의로 분석 방법을 사용하여 결정된 정보를 보고하는 단계 및/또는 분석 방법을 사용하여 결정된 정보를 함유하는 보고서를 생성하는 단계를 추가로 포함한다. 예를 들어, 일부 실시양태에서, 방법은 대상체로부터 유래된 폴리뉴클레오티드 (예를 들어, 대상체의 게놈 내의)의 변이체의 식별과 관련된 내용을 함유하는 보고서를 보고 또는 생성하는 단계를 추가로 포함한다. 보고된 정보 또는 보고서 내의 정보는 예를 들어 참조 서열에 맵핑된 커플링된 시퀀싱 리드 쌍의 유전자좌, 검출된 변이체 (예컨대 검출된 구조적 변이체 또는 검출된 SNP), 하나 이상의 어셈블리된 컨센서스 서열 및/또는 하나 이상의 어셈블리된 컨센서스 서열에 대한 검증 통계와 연관될 수 있다. 수신자, 예를 들어 임상의, 대상체, 또는 연구자에게 보고서가 배포될 수 있거나 이들에게 정보가 보고될 수 있다.

[0164] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성도

를 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0165]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0166]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0167]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에

제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0168] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0169] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱하여, 미리 선택된 흐름-사이클 순서에 따라 복수의 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되는 것인 단계; (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한

개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0170] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0171] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (b) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여 (여기서, 제2 흐름-사이클 순서는 제1 흐름-사이클 순서와 상이함), 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0172] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 수득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (d) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인

단계; 및 (e) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0173]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (b) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여 (여기서, 제2 흐름-사이클 순서는 제1 흐름-사이클 순서와 상이함), 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (d) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (e) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0174]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오타이드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (e) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0175]

일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 제1 흐름-사이클 순

서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (c) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및 (e) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

[0176] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한 것인 단계; (b) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열 (이는 참조 서열을 포함할 수 있음)에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (e) 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (f) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0177] 일부 실시양태에서, 하나 이상의 프로세서; 및 하기를 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체를 포함하는 시스템이 있다: (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오티드 흐름에 상응한 것인 단계; (b) 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 상이한 시험 핵산 분자와 연관된 것인 단계; (c) 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시험 샘플로부터 유래된 동일한 하나 이상의 시험 핵산 분자를 시퀀싱하여, 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는

단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트 중 하나와 동일한 시험 핵산 분자와 연관된 것인 단계; (d) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; (e) 2개 이상의 상이한 후보 서열 (이는 참조 서열을 포함할 수 있음)로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및 (f) 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열은 제1 흐름-사이클 순서 및/또는 제2 흐름-사이클 순서에 따라 2개 이상 (또는 3개 이상, 또는 하나 이상의 흐름-사이클에 걸쳐)의 흐름 위치 (이는 연속적 또는 비연속적일 수 있음)에서 선택된 후보 서열과 상이하다.

[0178]

일부 실시양태에서, 본원에 기재된 방법은 도 7에 예시된 구성성분 중 하나 이상을 사용하여 수행될 수 있는 컴퓨터-구현 방법이다. 예를 들어, 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서를 사용하여, 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서를 사용하여, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0179]

일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서를 사용하여, 표적 짧은 유전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서에서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와

연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0180]

일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서를 사용하여, 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서에서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0181]

일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서에서, 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서에서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0182]

일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서에서, 표적 짧은 유전자 변이체를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 수득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분

자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서에서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0183] 일부 실시양태에서, 시험 샘플에서 짧은 유전자 변이체를 검출하기 위한 컴퓨터-구현 방법은 하기를 포함한다: (a) 하나 이상의 프로세서에서, 표적 짧은 유전자 변이체 및 흐름-사이클 순서를 미리 선택하는 단계이며, 여기서 미리 선택된 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계; (b) 하나 이상의 프로세서에서, 하나 이상의 시험 시퀀싱 데이터 세트를 수신하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 미리 선택된 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 미리 선택된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계; (c) 하나 이상의 프로세서에서, 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및 (d) 하나 이상의 프로세서에서 및 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 미리 선택된 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계. 일부 실시양태에서, 방법은 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 추가로 포함하고, 바이오마커 패널은 표적 짧은 유전자 변이체를 포함한다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 2개 초과와 흐름 위치 (예를 들어, 2개 초과와 연속 흐름 위치 또는 2개 초과와 비연속 흐름 위치)에서 참조 시퀀싱 데이터 세트와 상이하다. 일부 실시양태에서, 표적 시퀀싱 데이터 세트는 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이하다.

[0184] 예시적인 실시양태

[0185] 하기 실시양태는 예시적이며, 청구된 발명의 범위를 제한하도록 의도되지 않는다.

[0186] 실시양태 1. 하기를 포함하는, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법:

[0187] (a) 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 초과와 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계;

[0188] (b) 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 시험 시퀀싱 데이터 세트는 시험 핵산 분자와 연관되고, 각 시험 핵산 분자는 표적 짧은 유전자 변이체와 연관되고 시험 샘플로부터 유래된 유전자좌와 적어도 부분적으로 중첩되고, 여기서 하나 이상의 시험 시퀀싱 데이터 세트는 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 시험 시퀀싱 데이터 세트는 복수의 흐름 위치에서 흐름 신호를 포함하는 것인 단계;

[0189] (c) 시험 시퀀싱 데이터 세트와 연관된 각 시험 핵산 분자에 대해, 핵산 분자와 연관된 시험 시퀀싱 데이터 세

트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어, 또는 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 및

- [0190] (d) 하나 이상의 결정된 매치 스코어를 사용하여, 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0191] 실시양태 2. 실시양태 1에 있어서, 수득하는 단계가 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 것을 포함하는 것인 방법.
- [0192] 실시양태 3. 실시양태 1 또는 실시양태 2에 있어서, 표적 짧은 유전자 변이체가 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택되는 것인 방법.
- [0193] 실시양태 4. 실시양태 1 또는 실시양태 2에 있어서, 표적 짧은 유전자 변이체가 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택되는 것인 방법.
- [0194] 실시양태 5. 실시양태 1-4 중 어느 하나에 있어서, 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 포함하고, 바이오마커 패널이 표적 짧은 유전자 변이체를 포함하는 것인 방법.
- [0195] 실시양태 6. 실시양태 1-5 중 어느 하나에 있어서, 흐름-사이클 순서를 선택하는 단계를 포함하는 방법.
- [0196] 실시양태 7. 실시양태 1-6 중 어느 하나에 있어서, 표적 시퀀싱 데이터 세트가 예상 표적 시퀀싱 데이터 세트이거나, 참조 시퀀싱 데이터 세트가 예상 참조 시퀀싱 데이터 세트인 방법.
- [0197] 실시양태 8. 실시양태 7에 있어서, 예상 표적 시퀀싱 데이터 세트 및 예상 참조 시퀀싱 데이터 세트가 인실리코에서 표적 서열 및 참조 서열을 시퀀싱함으로써 수득되는 것인 방법.
- [0198] 실시양태 9. 실시양태 1-8 중 어느 하나에 있어서, 표적 시퀀싱 데이터 세트가 2개 초과 of 비연속 흐름 위치에서 참조 시퀀싱 데이터와 상이한 것인 방법.
- [0199] 실시양태 10. 실시양태 1-9 중 어느 하나에 있어서, 표적 시퀀싱 데이터 세트가 2개 초과 of 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이한 것인 방법.
- [0200] 실시양태 11. 실시양태 1-10 중 어느 하나에 있어서, 표적 서열이 X개의 염기 위치에서 참조 서열과 상이하고, 여기서 표적 시퀀싱 데이터 세트가 (X+2)개 이상의 연속 흐름 위치에서 참조 시퀀싱 데이터와 상이한 것인 방법.
- [0201] 실시양태 12. 실시양태 11에 있어서, (X+2) 흐름 위치 차이가 0과 실질적으로 동일한 값 및 0보다 실질적으로 큰 값 사이의 차이를 포함하는 것인 방법.
- [0202] 실시양태 13. 실시양태 1-12 중 어느 하나에 있어서, 표적 시퀀싱 데이터 세트가 하나 이상의 흐름-사이클에 걸쳐 참조 시퀀싱 데이터 세트와 상이한 것인 방법.
- [0203] 실시양태 14. 실시양태 1-13 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함하는 것인 방법.
- [0204] 실시양태 15. 실시양태 1-14 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.
- [0205] 실시양태 16. 실시양태 1-15 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 복수의 염기 카운트에 대한 가능성을 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.
- [0206] 실시양태 17. 실시양태 16에 있어서, 단계 (c)가 하기를 포함하는 것인 방법:
- [0207] 해당 흐름 위치에서 표적 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계; 또는
- [0208] 해당 흐름 위치에서 참조 서열의 염기 카운트에 상응하는, 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하고, 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계.

- [0209] 실시양태 18. 실시양태 17에 있어서, 단계 (c)에서 결정된 매치 스코어가 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값인 방법.
- [0210] 실시양태 19. 실시양태 1-18 중 어느 하나에 있어서, 단계 (c)가 시험 시퀀싱 데이터 세트가 표적 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함하는 것인 방법.
- [0211] 실시양태 20. 실시양태 1-19 중 어느 하나에 있어서, 단계 (c)가 시험 시퀀싱 데이터 세트가 참조 서열과 매치할 가능성을 나타내는 매치 스코어를 결정하는 단계를 포함하는 것인 방법.
- [0212] 실시양태 21. 실시양태 1-20 중 어느 하나에 있어서, 하나 이상의 시험 시퀀싱 데이터 세트가 복수의 시험 시퀀싱 데이터 세트를 포함하는 것인 방법.
- [0213] 실시양태 22. 실시양태 21에 있어서, 표적 짧은 유전자 변이체의 존재 또는 부재가 하나 이상의 시험 시퀀싱 데이터 세트 각각에 대해 별도로 호출되는 것인 방법.
- [0214] 실시양태 23. 실시양태 21 또는 22에 있어서, 복수의 시험 시퀀싱 데이터 세트의 적어도 일부가 상이한 시퀀싱 시작 위치를 갖는 상이한 시험 핵산 분자와 연관된 것인 방법.
- [0215] 실시양태 24. 실시양태 1-23 중 어느 하나에 있어서, 흐름-사이클 순서가 동일한 순서로 반복되는 4개의 별도의 흐름을 포함하는 것인 방법.
- [0216] 실시양태 25. 실시양태 1-24 중 어느 하나에 있어서, 흐름-사이클 순서가 5개 이상의 별도의 흐름을 포함하는 것인 방법.
- [0217] 실시양태 26. 실시양태 1-25 중 어느 하나에 있어서, 하기를 포함하는, 컴퓨터-구현 방법인 방법:
- [0218] 하나 이상의 프로세서를 사용하여 표적 짧은 유전자 변이체를 선택하는 단계;
- [0219] 하나 이상의 프로세서에서 하나 이상의 시험 시퀀싱 데이터 세트를 수신함으로써 하나 이상의 시험 시퀀싱 데이터 세트를 획득하는 단계;
- [0220] 하나 이상의 프로세서를 사용하여 하나 이상의 매치 스코어를 결정하는 단계; 및
- [0221] 하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0222] 실시양태 27. 하기를 포함하는 시스템:
- [0223] 하나 이상의 프로세서; 및
- [0224] 실시양태 1-26 중 어느 하나의 방법을 구현하기 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체.
- [0225] 실시양태 28. 하기를 포함하는, 시험 샘플에서 짧은 유전자 변이체를 검출하는 방법:
- [0226] (a) 하나 이상의 제1 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제1 시험 시퀀싱 데이터 세트는 시험 샘플로부터 유래된 상이한 시험 핵산 분자와 연관되고, 여기서 제1 시험 시퀀싱 데이터 세트는 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 하나 이상의 제1 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계;
- [0227] (b) 하나 이상의 제2 시험 시퀀싱 데이터 세트를 획득하는 단계이며, 각 제2 시험 시퀀싱 데이터 세트는 제1 시험 시퀀싱 데이터 세트와 동일한 시험 핵산 분자와 연관되고, 여기서 제2 시험 시퀀싱 데이터 세트는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 제공된 비중결 뉴클레오티드를 사용하여 하나 이상의 시험 핵산 분자를 시퀀싱함으로써 결정되었고, 여기서 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서는 상이하고, 여기서 시험 시퀀싱 데이터 세트는 뉴클레오티드 흐름에 상응하는 흐름 위치에서 흐름 신호를 포함하는 것인 단계;
- [0228] (c) 각각의 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트에 대해, 하나 이상의 후보 서열에 대한 매치 스코어를 결정하는 단계이며, 여기서 매치 스코어는 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두가 하나 이상의 후보 서열로부터의 후보 서열과 매치할 가능성을 나타내는 것인 단계; 및
- [0229] (d) 결정된 매치 스코어를 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.

- [0230] 실시양태 29. 실시양태 28에 있어서, 제1 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계, 및 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비중결 뉴클레오타이드를 사용하여 시험 핵산 분자를 시퀀싱하는 단계를 포함하는 방법.
- [0231] 실시양태 30. 실시양태 28 또는 29에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성, 또는 제2 시험 시퀀싱 데이터 세트가 후보 서열과 매치할 가능성을 나타내는 것인 방법.
- [0232] 실시양태 31. 실시양태 28 또는 29에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트 둘 모두가 후보 서열과 매치할 가능성을 나타내는 것인 방법.
- [0233] 실시양태 32. 실시양태 28-31 중 어느 하나에 있어서, 하나 이상의 후보 서열이 2개 이상의 상이한 후보 서열을 포함하고, 방법이 제1 시퀀싱 데이터 세트 및 제2 시퀀싱 데이터 세트와 연관된 각 핵산 분자에 대해, 하기를 포함하는 것인 방법:
- [0234] 2개 이상의 상이한 후보 서열로부터 후보 서열을 선택하는 단계이며, 여기서 선택된 후보 서열은 제1 시험 시퀀싱 데이터 세트, 제2 시험 시퀀싱 데이터 세트, 또는 둘 모두와 가장 높은 가능성도 매치를 갖는 것인 단계; 및
- [0235] 선택된 후보 서열을 사용하여, 시험 샘플에서 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0236] 실시양태 33. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0237] 실시양태 34. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0238] 실시양태 35. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 비연속 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0239] 실시양태 36. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 비연속 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0240] 실시양태 37. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 연속 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0241] 실시양태 38. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 연속 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0242] 실시양태 39. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0243] 실시양태 40. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 3개 이상의 흐름 위치에서 선택된 후보 서열과 상이한 것인 방법.
- [0244] 실시양태 41. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 X개의 염기 위치에서 선택된 후보 서열과 상이하고, 여기서 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 (X+2)개 이상의 흐름 위치에서 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열과 상이한 것인 방법.
- [0245] 실시양태 42. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 X개의 염기 위치에서 선택된 후보 서열과 상이하고, 여기서 시험 핵산 분자와 연관된 시험 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 (X+2)개 이상의 흐름 위치에서 2개 이상의

상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열과 상이한 것인 방법.

- [0246] 실시양태 43. 실시양태 41 또는 42에 있어서, $(X+2)$ 흐름 위치 차이가 0과 실질적으로 동일한 값 및 0보다 실질적으로 큰 값 사이의 차이를 포함하는 것인 방법.
- [0247] 실시양태 44. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 선택된 후보 서열과 상이한 것인 방법.
- [0248] 실시양태 45. 실시양태 32에 있어서, 2개 이상의 상이한 후보 서열로부터의 적어도 하나의 비선택된 후보 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 하나 이상의 흐름-사이클에 걸쳐 선택된 후보 서열과 상이한 것인 방법.
- [0249] 실시양태 46. 실시양태 28-45 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 염기 카운트를 포함하는 것인 방법.
- [0250] 실시양태 47. 실시양태 28-46 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 적어도 하나의 염기 카운트에 대한 가능도를 나타내는 통계적 파라미터를 포함하고, 여기서 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.
- [0251] 실시양태 48. 실시양태 28-47 중 어느 하나에 있어서, 흐름 신호가 각 흐름 위치에서 복수의 염기 카운트에 대한 가능도를 나타내는 통계적 파라미터를 포함하고, 여기서 각 염기 카운트가 흐름 위치에서 시퀀싱된 시험 핵산 분자의 염기의 수를 나타내는 것인 방법.
- [0252] 실시양태 49. 실시양태 48에 있어서, 매치 스코어를 결정하는 단계가 하나 이상의 상이한 후보 서열 각각에 대해, 해당 흐름 위치에서 후보 서열의 염기 카운트에 상응하는, 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 각 흐름 위치에서 통계적 파라미터를 선택하는 것을 포함하는 것인 방법.
- [0253] 실시양태 50. 실시양태 49에 있어서, 하나 이상의 상이한 후보 서열에 대해, 각 흐름 위치에서 후보 서열의 염기 카운트를 포함하는 후보 시퀀싱 데이터 세트를 생성하는 단계를 포함하는 방법.
- [0254] 실시양태 51. 실시양태 50에 있어서, 후보 시퀀싱 데이터 세트가 인실리코에서 생성되는 것인 방법.
- [0255] 실시양태 52. 실시양태 49-51 중 어느 하나에 있어서, 매치 스코어가 제1 시험 시퀀싱 데이터 세트 및 제2 시험 시퀀싱 데이터 세트에서 흐름 위치에 걸쳐 선택된 통계적 파라미터의 조합된 값인 방법.
- [0256] 실시양태 53. 실시양태 28-52 중 어느 하나에 있어서, 시험 핵산 분자의 적어도 일부가 상이한 시퀀싱 시작 위치를 갖는 것인 방법.
- [0257] 실시양태 54. 실시양태 28-52 중 어느 하나에 있어서, 하기를 포함하는 방법:
- [0258] 표적 짧은 유전자 변이체를 선택하는 단계이며, 여기서 표적 짧은 유전자 변이체를 포함하는 표적 서열과 연관된 표적 시퀀싱 데이터 세트는 표적 시퀀싱 데이터 세트 및 참조 시퀀싱 데이터 세트가 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 표적 서열을 시퀀싱함으로써 획득될 때 2개 이상의 흐름 위치에서 참조 서열과 연관된 참조 시퀀싱 데이터 세트와 상이하고, 여기서 제1 흐름-사이클 순서는 제2 흐름-사이클 순서와 상이하고, 여기서 흐름 위치는 뉴클레오타이드 흐름에 상응한 것인 단계;
- [0259] 여기서 하나 이상의 후보 서열은 표적 서열 및 참조 서열을 포함한다.
- [0260] 실시양태 55. 실시양태 54에 있어서, 표적 짧은 유전자 변이체가 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하기 전에 미리 선택되는 것인 방법.
- [0261] 실시양태 56. 실시양태 54에 있어서, 표적 짧은 유전자 변이체가 호출의 신뢰도에 기초하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출한 후 선택되는 것인 방법.
- [0262] 실시양태 57. 실시양태 56에 있어서, 시험 샘플과 연관된 대상체에 대한 개인화된 바이오마커 패널을 생성하는 단계를 포함하고, 바이오마커 패널이 시험 샘플에 존재하는 표적 짧은 유전자 변이체를 포함하는 것인 방법.
- [0263] 실시양태 58. 실시양태 54-57 중 어느 하나에 있어서, 참조 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 별도의 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시퀀싱되었을 경우, 참조 시퀀싱 데이터 세

트가 예상 참조 시퀀싱 데이터 세트를 결정함으로써 획득되는 것인 방법.

- [0264] 실시양태 59. 실시양태 54-57 중 어느 하나에 있어서, 참조 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 별도의 흐름에 제공된 비종결 뉴클레오티드를 사용하여 시퀀싱되었을 경우, 참조 시퀀싱 데이터 세트가 예상 참조 시퀀싱 데이터 세트를 결정함으로써 획득되는 것인 방법.
- [0265] 실시양태 60. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0266] 실시양태 61. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 비연속 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0267] 실시양태 62. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 비연속 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0268] 실시양태 63. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 2개 이상의 연속 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0269] 실시양태 64. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 2개 이상의 연속 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0270] 실시양태 65. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 3개 이상의 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0271] 실시양태 66. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 3개 이상의 흐름 위치에서 참조 서열과 상이한 것인 방법.
- [0272] 실시양태 67. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서에 따라 하나 이상의 흐름-사이클에 걸쳐 참조 서열과 상이한 것인 방법.
- [0273] 실시양태 68. 실시양태 54-57 중 어느 하나에 있어서, 표적 서열이 제1 흐름-사이클 순서 및 제2 흐름-사이클 순서 둘 모두에 따라 하나 이상의 흐름-사이클에 걸쳐 참조 서열과 상이한 것인 방법.
- [0274] 실시양태 69. 실시양태 28-68 중 어느 하나에 있어서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서가 동일한 순서로 반복되는 4개의 별도의 흐름을 포함하는 것인 방법.
- [0275] 실시양태 70. 실시양태 28-68 중 어느 하나에 있어서, 제1 흐름-사이클 순서 또는 제2 흐름-사이클 순서가 동일한 순서로 반복되는 5개 이상의 별도의 흐름을 포함하는 것인 방법.
- [0276] 실시양태 71. 실시양태 28-70 중 어느 하나에 있어서, 하기를 포함하는 방법:
- [0277] 제1 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로의 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제 1 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 시험 핵산 분자를 시퀀싱하는 단계;
- [0278] 연장된 시퀀싱 프라이머를 제거하는 단계; 및
- [0279] 제2 흐름-사이클 순서에 따라 별도의 뉴클레오티드 흐름에 비종결 뉴클레오티드를 제공하고, 시퀀싱 프라이머를 연장하고, 각 뉴클레오티드 흐름 후에 시퀀싱 프라이머로의 뉴클레오티드 혼입의 존재 또는 부재를 검출하여 제 2 시험 시퀀싱 데이터 세트를 생성하는 것을 포함하는, 동일한 시험 핵산 분자를 시퀀싱하는 단계.
- [0280] 실시양태 72. 실시양태 28-71 중 어느 하나에 있어서, 하기를 포함하는, 컴퓨터-구현 방법인 방법:
- [0281] 하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계;
- [0282] 상기 하나 이상의 프로세서에서 하나 이상의 제1 시퀀싱 데이터 세트를 수신하는 단계;
- [0283] 상기 하나 이상의 프로세서를 사용하여 매치 스코어를 결정하는 단계; 및
- [0284] 상기 하나 이상의 프로세서를 사용하여 시험 샘플에서 표적 짧은 유전자 변이체의 존재 또는 부재를 호출하는 단계.
- [0285] 실시양태 73. 하기를 포함하는 시스템:

- [0286] 하나 이상의 프로세서; 및
- [0287] 실시양태 28-72 중 어느 하나의 방법을 구현하기 위한 명령을 포함하는 하나 이상의 프로그램을 저장하는 비일시적 컴퓨터-판독가능 매체.
- [0288] 실시양태 74. 실시양태 1-73 중 어느 하나에 있어서, 별도의 흐름이 단일 염기 유형을 포함하는 것인 방법 또는 시스템.
- [0289] 실시양태 75. 실시양태 1-74 중 어느 하나에 있어서, 별도의 흐름 중 적어도 하나가 2 또는 3개의 상이한 염기 유형을 포함하는 것인 방법 또는 시스템.
- [0290] 실시양태 76. 실시양태 1-75 중 어느 하나에 있어서, 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 변이체 호출 파일을 생성 또는 업데이트하는 단계를 포함하는 방법 또는 시스템.
- [0291] 실시양태 77. 실시양태 1-76 중 어느 하나에 있어서, 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 보고서를 생성하는 단계를 포함하는 방법 또는 시스템.
- [0292] 실시양태 78. 실시양태 77에 있어서, 보고서가 시험 샘플에서 짧은 유전자 변이체의 존재, 아이덴티티 또는 부재를 나타내는 텍스트, 확률, 수치 또는 그래픽 출력을 포함하는 것인 방법 또는 시스템.
- [0293] 실시양태 79. 실시양태 77 또는 78에 있어서, 환자 또는 환자의 의료 대리인에게 보고서를 제공하는 단계를 포함하는 방법 또는 시스템.
- [0294] 실시양태 78. 실시양태 1-77 중 어느 하나에 있어서, 짧은 유전자 변이체가 단일 뉴클레오타이드 다형성을 포함하는 것인 방법 또는 시스템.
- [0295] 실시양태 79. 실시양태 1-77 중 어느 하나에 있어서, 짧은 유전자 변이체가 indel을 포함하는 것인 방법 또는 시스템.
- [0296] 실시양태 80. 실시양태 1-79 중 어느 하나에 있어서, 시험 샘플이 단편화된 DNA를 포함하는 것인 방법 또는 시스템.
- [0297] 실시양태 81. 실시양태 1-80 중 어느 하나에 있어서, 시험 샘플이 무세포 DNA를 포함하는 것인 방법 또는 시스템.
- [0298] 실시양태 82. 실시양태 81에 있어서, 무세포 DNA가 순환 종양 DNA (ctDNA)를 포함하는 것인 방법 또는 시스템.
- [0299] 실시양태 83. 하기를 포함하는, 핵산 분자를 시퀀싱하는 방법:
- [0300] 핵산 분자를 프라이머에 혼성화하여 혼성화된 주형을 형성하는 단계;
- [0301] 5개 이상의 별도의 뉴클레오타이드 흐름을 포함하는 반복된 흐름-사이클 순서에 따라 별도의 뉴클레오타이드 흐름에 제공된 표지된 비중결 뉴클레오타이드를 사용하여 프라이머를 연장하는 단계; 및
- [0302] 프라이머가 뉴클레오타이드 흐름에 의해 연장됨에 따라 혼입된 표지된 뉴클레오타이드로부터의 신호 또는 신호의 부재를 검출하는 단계.
- [0303] 실시양태 84. 실시양태 83에 있어서, 각 뉴클레오타이드 흐름 후에 신호 또는 신호의 부재를 검출하는 단계를 포함하는 방법.
- [0304] 실시양태 85. 실시양태 83 또는 84에 있어서, 복수의 핵산 분자를 시퀀싱하는 단계를 포함하는 방법.
- [0305] 실시양태 86. 실시양태 85에 있어서, 복수의 핵산 분자가 유전자좌에 대해 상이한 시퀀싱 시작 위치를 갖는 것인 방법.
- [0306] 실시양태 87. 실시양태 83-86 중 어느 하나에 있어서, 시험 샘플이 무세포 DNA인 방법.
- [0307] 실시양태 88. 실시양태 83-86 중 어느 하나에 있어서, 무세포 DNA가 순환 종양 DNA (ctDNA)를 포함하는 것인 방법.
- [0308] 실시양태 89. 실시양태 83-86 중 어느 하나에 있어서, 흐름-사이클 순서가 무작위 시퀀싱 시작 위치의 5% 이상에서 가능한 SNP 순열의 50% 이상에 대해 2개 초과인 흐름 위치에서 신호 변화를 유도하는 것인 방법.
- [0309] 실시양태 90. 실시양태 83-86 중 어느 하나에 있어서, 흐름-사이클 순서가 흐름 당 0.6 이상의 염기 혼입 효율

을 갖는 것인 방법.

[0310]

실시에

[0311]

본 출원은 본 출원의 예시적인 실시양태로서 제공되는 하기 비제한적인 실시예를 참조하여 더 잘 이해될 수 있다. 하기 실시예는 실시양태를 보다 완전하게 예시하기 위해 제시되지만, 본 출원의 넓은 범위를 제한하는 것으로 결코 해석되어서는 안된다. 본 출원의 특정 실시양태가 본원에 도시되고 설명되었지만, 이러한 실시양태는 단지 예로서 제공된다는 것이 명백할 것이다. 본 발명의 사상 및 범위를 벗어나지 않으면서 관련 기술분야의 통상의 기술자에게 수많은 변경, 변화 및 치환이 발생할 수 있다. 본원에 기재된 실시양태에 대한 다양한 대안이 본원에 기재된 방법을 실시하는데 사용될 수 있음을 이해해야 한다.

[0312]

실시에 1 - SNP 검출

[0313]

가상의 핵산 분자는 흐름-사이클 순서 A-T-G-C에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시퀀싱되며, 결과적으로 도 1a에 표시된 시험 시퀀싱 데이터 세트가 생성된다. 시퀀싱 데이터 세트의 각 값은 각 흐름 위치에서 표시된 염기 카운트가 올바른 가능성을 나타낸다. 시퀀싱 데이터 세트에 기초하여, 예비 서열은 TATGGTCGTCGA (서열식별번호: 1)로 결정되며, 이는 참조 게놈의 유전자좌에 맵핑된다. 참조 게놈의 유전자좌는 잠재적인 반수체형 서열 TATGGTCGTCGA (서열식별번호: 1) (H1) 및 TATGGTCATCGA (서열식별번호: 2) (H2)와 연관된다. 각 흐름 위치에 대한 반수체형 서열의 염기 카운트와 연관된 가능성 값이 각 반수체형에 대해 선택된다. 각 반수체형이 주어진 시퀀싱 데이터 세트의 가능성은 각 흐름 위치에 대한 반수체형 서열의 염기 카운트와 연관된 가능성 값을 곱함으로써 결정된다. 시퀀싱 데이터 세트의 로그 가능성은 H1이 올바른 서열인 경우 -0.015이고, 시퀀싱 데이터 세트의 로그 가능성은 H2가 올바른 서열인 경우 -27.008이다. 그러므로, 이 핵산 분자에 대해 H1의 서열이 선택된다.

[0314]

실시에 2 - Indel 검출

[0315]

가상의 핵산 분자는 흐름-사이클 순서 A-T-G-C에 따라 별도의 뉴클레오타이드 흐름에 제공된 비종결 뉴클레오타이드를 사용하여 시퀀싱되며, 결과적으로 도 8에 표시된 시험 시퀀싱 데이터 세트가 생성된다. 시퀀싱 데이터 세트의 각 값은 각 흐름 위치에서 표시된 염기 카운트가 올바른 가능성을 나타낸다. 시퀀싱 데이터 세트에 기초하여 (즉, 각 흐름 위치에서 가장 가능성이 높은 염기 카운트를 선택함으로써), 예비 서열은 TATGGTCGATCG (서열식별번호: 8)로 결정되며, 이는 참조 게놈의 유전자좌에 맵핑된다. 참조 게놈의 유전자좌는 잠재적인 반수체형 서열 TATGGTCG-TCGA (서열식별번호: 7) (H1) 및 TATGGTCGATCG (서열식별번호: 8) (H2)와 연관된다. 각 흐름 위치에 대한 반수체형 서열의 염기 카운트와 연관된 가능성 값이 각 반수체형에 대해 선택된다. 각 반수체형이 주어진 시퀀싱 데이터 세트의 가능성은 각 흐름 위치에 대한 반수체형 서열의 염기 카운트와 연관된 가능성 값을 곱함으로써 결정된다. 시퀀싱 데이터 세트의 로그 가능성은 H1이 올바른 서열인 경우 -24.009이고, 시퀀싱 데이터 세트의 로그 가능성은 H2가 올바른 서열인 경우 -0.015이다. 그러므로, 이 핵산 분자에 대해 H2의 서열이 선택된다.

[0316]

실시에 3 - 연장된 시퀀싱 흐름 순서

[0317]

모든 가능한 SNP 세트 (XYZ → XQZ, 여기서 Q≠Y (및 Q, X, Y 및 Z는 각각 A, C, G 및 T 중 어느 하나임))에 대해 2개 초과 흐름 위치에서 신호 변화를 유도할 가능성에 대해 인실리코에서 백만개 초과 연장된 시퀀싱 흐름 순서를 시험하였다. 연장된 흐름 순서는 모든 유효한 2-염기 흐름 순열과 함께 최소 12개의 염기 서열을 갖도록 설계되었으며, 순차적 염기 반복부를 갖는 흐름 순서는 제거되었다. 2개 초과 흐름 위치에서 신호 변화를 유도하기 위해 연장된 흐름 순서의 민감도를 평가하기 위해 흐름 순서에 대한 모든 가능한 시작 위치를 시험하였다. 도 9 및 표 2는 이 분석의 예시적인 결과를 보여준다. 도 9에서, x-축은 흐름 위상 (또는 단편화 시작 위치)의 분율을 나타내고, y-축은 2개 초과 흐름 위치에서 신호 변화를 유도한 SNP 순열의 분율을 나타낸다. 여러 흐름 순서는 리드 (또는 흐름 시작 위치)의 대략 10%에 대해 모든 가능한 (87.5%) SNP 순열에서 2개 이상의 신호 차이를 유도한다. 4개의 염기 주기적인 흐름은 가능한 SNP의 42%에서만 사이클 이동을 유도하지만, 모든 리드 또는 흐름 위상에서 이를 수행한다. 생존성을 확립하기 위해 인간 참조 게놈의 백만 염기 서브세트에 대해 효율의 최종 평가를 수행하였다. 이는 실제 유기체의 패턴 및 편향이 주어지면 흐름 순서가 서열을 얼마나 효율적으로 연장하는지에 대한 실용적인 측정이다.

[0318]

실시에 4 - SNP 검출 정확도

[0319]

DNA 샘플 NA12878 (코리엘 의학연구소(Coriell Institute for Medical Research)로부터 입수가 가능한 샘플)의 게놈을 4개의 흐름 사이클 (T-A-C-G)에 따라 비종결 형광 표지된 뉴클레오타이드를 사용하여 시퀀싱하였다. 시퀀싱

실행은 176개 염기의 평균 길이를 갖는 415,900,002개의 리드를 생성하였다. 399,804,925개의 리드가 hg38 참조 게놈에 정렬되었다 (BWA, 버전 0.7.17-r1188 사용).

[0320] 정렬 후, 참조 게놈과 완벽하게 정렬된 리드 (178,634,625개의 리드) 또는 참조 게놈과의 단일 미스매치를 함유하고 20 이상의 맵핑 품질 스코어로 정렬된 리드 (27,265,661개의 리드)를 선택하였다. 즉, 193,904,639개는 예를 들어 참조 게놈에 대한 indel, 다중 미스매치, 또는 잠재적으로 잘못된 (아티팩트) 정렬을 갖기 때문에 추가 분석을 위해 배제되었다. 따라서, 27,265,661개의 리드는 참양성 NA12878 SNP, 뿐만 아니라 시퀀싱 오류로부터 발생한 임의의 위양성 SNP를 포함하는 것으로 추정되었다. 27,265,661개 리드의 이러한 풀로부터, 미스매치된 유전자좌에 걸쳐 있는 시퀀싱 리드를 1회 초과 제거하여 참양성 NA12878 SNP 변이체의 효과를 감소시켜, 깊이 1)의 미스매치를 함유하는 총 3,413,700개의 리드를 생성하였다.

[0321] 나머지 3,413,700개의 리드는 각각 하기 미스매치를 포함하였다: (1) 흐름도 흐름 신호가 흐름-사이클 순서에 기초하여 참조에 대해 하나의 전체 사이클 (예를 들어, 4개의 흐름 위치)만큼 이동하는 경우 사이클 이동을 유도할 것으로 예상되었음, (2) 상이한 흐름 사이클이 사용된 경우 잠재적으로 사이클 이동을 유도할 수 있었음 (예를 들어, 이는 흐름도에서 새로운 제로 또는 새로운 년-제로 신호를 생성함), 또는 (3) 흐름-사이클 순서에 관계없이 사이클 이동을 유도할 수 없었음. 3,413,700개의 미스매치 중 1,184,954개 (34%)가 사이클 이동을 유도한 반면, 1,546,588개 (43%)는 상이한 흐름 순서로 사이클 이동을 유도할 수 있었다 (즉, "잠재적인 사이클 이동"). 비교하여, 무작위 미스매치의 이론적 예상은 명목상 42% 사이클 이동 및 46% 잠재적인 사이클 이동 미스매치를 제한하였다. 전반적으로, 사이클 이동을 유도하는 미스매치 비율은 3.7×10^{-5} 사건/염기였고, 잠재적인 사이클 이동을 유도하는 미스매치 비율은 4.8×10^{-5} 사건/염기였다. 표 3은 사이클 이동을 유도하는 10개의 가장 빈번한 단일 미스매치 및 상대적 발생 백분율을 보여준다.

[0322] <표 3>

참조	리드	% 사례
TTT	TCT	7.18
AAA	AGA	7.18
GAG	GGG	4.63
CTC	CCC	4.62
CAG	CGG	4.12
CTG	CCG	4.09
AAC	AGC	3.86
GTT	GCT	3.83
CAT	CGT	3.63
GAT	GGT	3.62

[0323] 그 후, 3개의 상이한 클래스 (즉, 사이클 이동을 유도하거나, 사이클 이동을 잠재적으로 유도하거나, 또는 사이클 이동을 유도하지 않고 유도할 수 없음) 각각의 미스매치에 기초한 변이체 호출의 성능을 평가하였다. 리드는 BWA를 사용하여 참조 게놈에 정렬되었고, GATK (버전 4)의 하플로타입콜러 도구를 사용하여 변이체 호출을 수행하였다. 생성된 미스매치 호출은 10개 염기보다 긴 호모폴리머 내에서, 또는 10개 염기 이상의 길이를 갖는 호모폴리머에 인접한 10개 염기 내에서 변이체 호출을 폐기함으로써 필터링되었다.

[0325] 미스매치 호출은 각 클래스의 미스매치에 대한 결정된 정확도 $\#TP/(\#FP+\#FN+\#TP)$ 에 대한 병내 게놈 (GIAB) 프로젝트에 의해 동일한 NA12878에 대해 생성된 호출과 비교되었다. 시퀀싱 데이터는 표시된 평균 게놈 깊이로 무작위로 다운 샘플링되었다. 사이클 이동을 유도하는 미스매치 및 사이클 이동을 잠재적으로 유도하는 미스매치는 표 4에서 입증된 바와 같이 사이클 이동을 유도하지 않는 미스매치보다 더 높은 정확도를 가졌다.

[0326] <표 4>

미스매치 유형	30x	22x	15x	8x
사이클 이동	0.9834	0.981	0.981	0.9772
사이클 이동 없음	0.9799	0.9759	0.9775	0.9696
잠재적인 사이클 이동	0.9826	0.9808	0.9795	0.9767

[0327]

도면

도면1a

열기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	
0	0.0010	0.0010	0.9988	0.9988	0.0010	0.9988	0.9988	0.0001	0.0010	0.9988	0.0010	0.0010	0.0010	0.9988	0.0010	0.0010	0.9988	0.0010	0.9988	0.9988
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.9979	0.0010	0.9979	0.9979	0.9979	0.0010	0.9979	0.9979	0.0010	0.9979	0.0010	0.0010
2	0.0010	0.0010	0.0001	0.0001	0.0010	0.0001	0.0001	0.9979	0.0010	0.0001	0.0010	0.0010	0.0010	0.0001	0.0010	0.0010	0.0001	0.0010	0.0001	0.0001
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

도면1b

접기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.0010 T	0.0010 A	0.0001 C	0.0010 G	0.0010 T	0.0001 A	0.0001 C	0.0001 G	0.0010 T	0.0001 A	0.0010 C	0.0010 G	0.0010 T	0.0001 A	0.0010 C	0.0010 G	0.0001 T	0.0010 A	0.0001 C	0.0001 G
1	0.0001 ★	0.0001 ★	0.0010	0.0010 ★	0.0001 ★	0.0010	0.0001 ★	0.0010	0.0001 ★	0.0010	0.0001 ★	0.0001 ★	0.0001 ★	0.0010	0.0001 ★	0.0001 ★	0.0010	0.0001 ★	0.0010	0.0001 ★
2	0.0010	0.0010	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001 ★	0.0010	0.0001	0.0010	0.0001	0.0001	0.0001	0.0010	0.0010	0.0001	0.0010	0.0001	0.0001
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

도면1c

염기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
1	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
2	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
3	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010

도면2a

TATGGTCGTCGA	R1
TGGTCGTCGAGC	R2
TATATGGTCGTC	R3
TATATGGTCATCGAGCTAT	H1
TATATGGTCGTCGAGCTAT	H2
↑	

도면2b

R1 시퀀싱 데이터 세트

염기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.0010 T	0.0010 A	0.9988 C	0.9988 G	0.0010 T	0.9988 A	0.9988 C	0.0010 G	0.0010 T	0.9988 A	0.0010 C	0.0010 G	0.0010 T	0.9988 A	0.0010 C	0.0010 G	0.9988 T	0.0010 A	0.9988 C	0.9988 G
1	0.9979 ●	0.9979 ○	0.0010 ●	0.0010 ○	0.9979 ○	0.0010 ○	0.0010 ○	0.0010 ○	0.9979 ○	0.0010 ○	0.9979 ○	0.9979 ○	0.9979 ○	0.0010 ○	0.9979 ○	0.9979 ○	0.0010 ○	0.9979 ○	0.0010 ○	0.0010 ○
2	0.0010 ○	0.0010 ○	0.0001 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0001 ○	0.9979 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0010 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0001 ○
3	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○

도면2c

R2 시퀀싱 데이터 세트

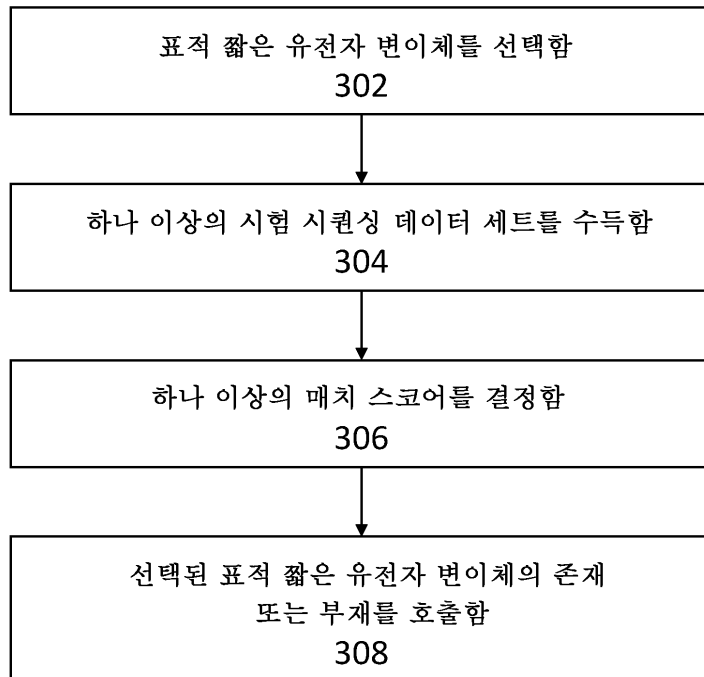
염기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.0010 T	0.9979 A	0.9988 C	0.0001 G	0.0010 T	0.9988 A	0.0010 C	0.0010 G	0.0010 T	0.9988 A	0.0010 C	0.0010 G	0.9988 T	0.0010 A	0.9979 C	0.0010 G	0.9988 T	0.9979 A	0.0010 C	0.9988 G
1	0.9979 ●	0.0010 ○	0.0010 ●	0.0010 ○	0.9979 ○	0.0010 ●	0.9988 ○	0.9988 ●	0.9979 ○	0.0010 ○	0.9979 ●	0.9979 ○	0.9988 ○	0.9988 ●	0.0010 ○	0.9979 ○	0.0010 ○	0.0010 ●	0.9988 ○	0.0010 ●
2	0.0010 ○	0.0001 ○	0.0001 ○	0.9988 ○	0.9979 ○	0.0001 ○	0.0010 ○	0.0010 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0001 ○	0.0010 ○	0.0001 ○
3	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0010 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○	0.0001 ○

도면2d

R3 시퀀싱 데이터 세트

염기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	T 0.0010	A 0.0010	C 0.0010	G 0.0010	T 0.0010	A 0.0010	C 0.0010	G 0.0010	T 0.0010	A 0.0010	C 0.0010	G 0.0010	T 0.0010	A 0.0010	C 0.0010	G 0.0010	T 0.0010	A 0.0010	C 0.0010	G 0.0010
1	● 0.0010	○ 0.0010	● 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010
2	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010
3	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.0010

도면3



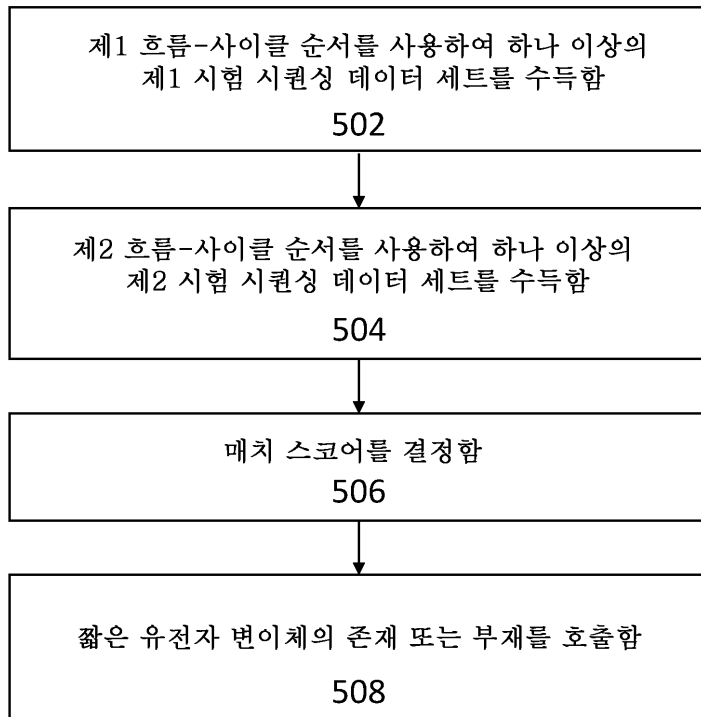
도면4a

열기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	T 0.0010	A 0.0010	C 0.9988 ○	G 0.9988 ○	T 0.0010	A 0.9988 ○	C 0.9988 ○	G 0.0001	T 0.0010	A 0.9988 ○	C 0.0010	G 0.0010 ○	T 0.0010 ○	A 0.9988 ●	C 0.0010 ○	G 0.0010 ○	T 0.9988 ●	A 0.0010 ○	C 0.9988 ●	G 0.9988 ●
1	● 0.9979 ○	● 0.9979 ○	○ 0.0010	○ 0.0010	○ 0.9979	○ 0.0010	○ 0.0010	○ 0.0010	○ 0.9979	○ 0.0010	○ 0.9979	○ 0.0010	○ 0.9979	○ 0.0010	○ 0.9979	○ 0.0010	○ 0.0001	○ 0.9979	○ 0.0010	○ 0.0010
2	○ 0.0010	○ 0.0010	○ 0.0001	○ 0.0001	○ 0.0010	○ 0.0001	○ 0.0001	○ 0.9979	○ 0.0010	○ 0.0001	○ 0.0010	○ 0.0010	○ 0.0001	○ 0.0001	○ 0.0010	○ 0.0010	○ 0.0001	○ 0.0010	○ 0.0001	○ 0.0001
3	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0010	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001	○ 0.0001

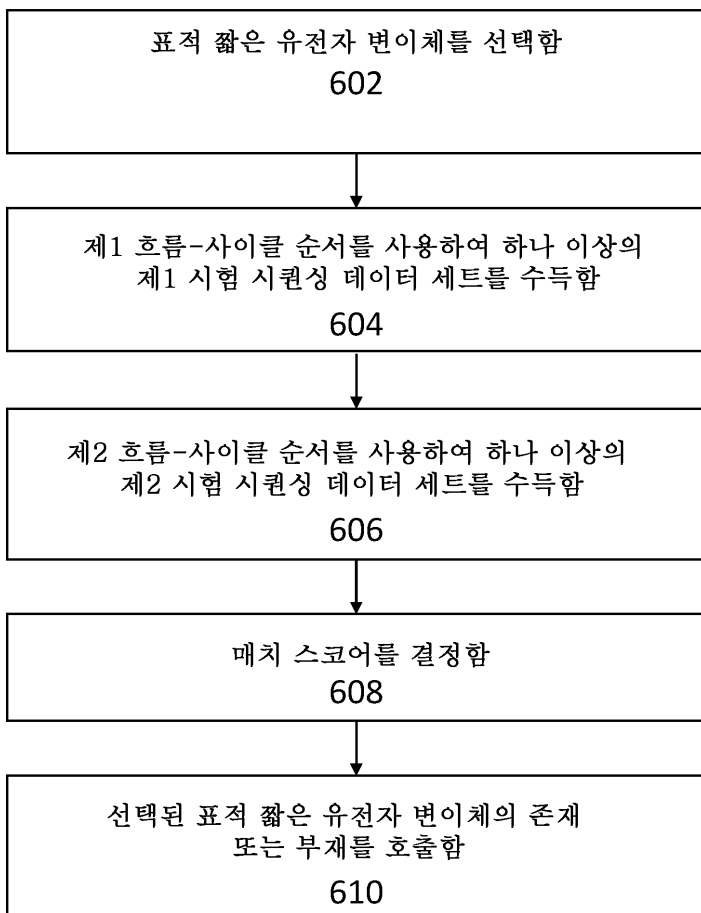
도면4b

검기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T
	0.9979 ●○	0.9979 ●○	0.9988 ●○	0.0010 ●○	0.0010 ●○	0.9988 ●○	0.9988 ●○	0.0010 ●○	0.9979 ●○	0.0001 ●○	0.9979 ●○	0.0010 ●○	0.9979 ●○	0.9988 ●○	0.0010 ●○	0.9979 ●○	0.9988 ●○	0.0010 ○●	0.9988 ○●	0.0010 ○●
1	0.0010	0.0010	0.0010	0.9988 ●○	0.9979 ●○	0.0010	0.0010	0.9979 ●○	0.0010	0.0010	0.0010	0.9979 ●○	0.0010	0.0010	0.9979 ●○	0.0010	0.0010	0.9979 ●○	0.0010	0.9988 ●○
2	0.0001	0.0010	0.0001	0.0010	0.0010	0.0001	0.0001	0.0010	0.0001	0.9979 ●○	0.0001	0.0010	0.0001	0.0001	0.0010	0.0001	0.0001	0.0010	0.0001	0.0010
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

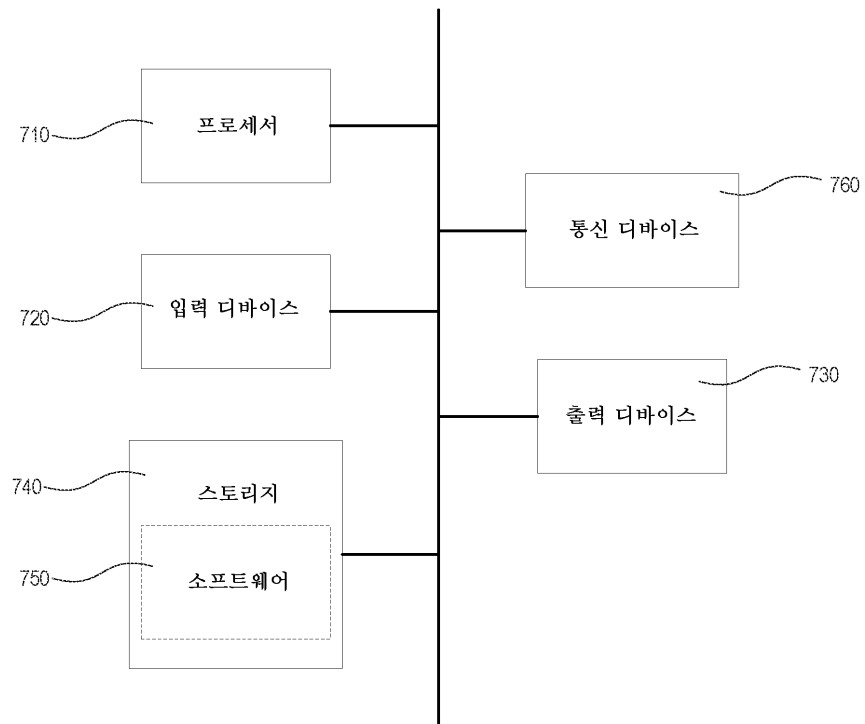
도면5



도면6



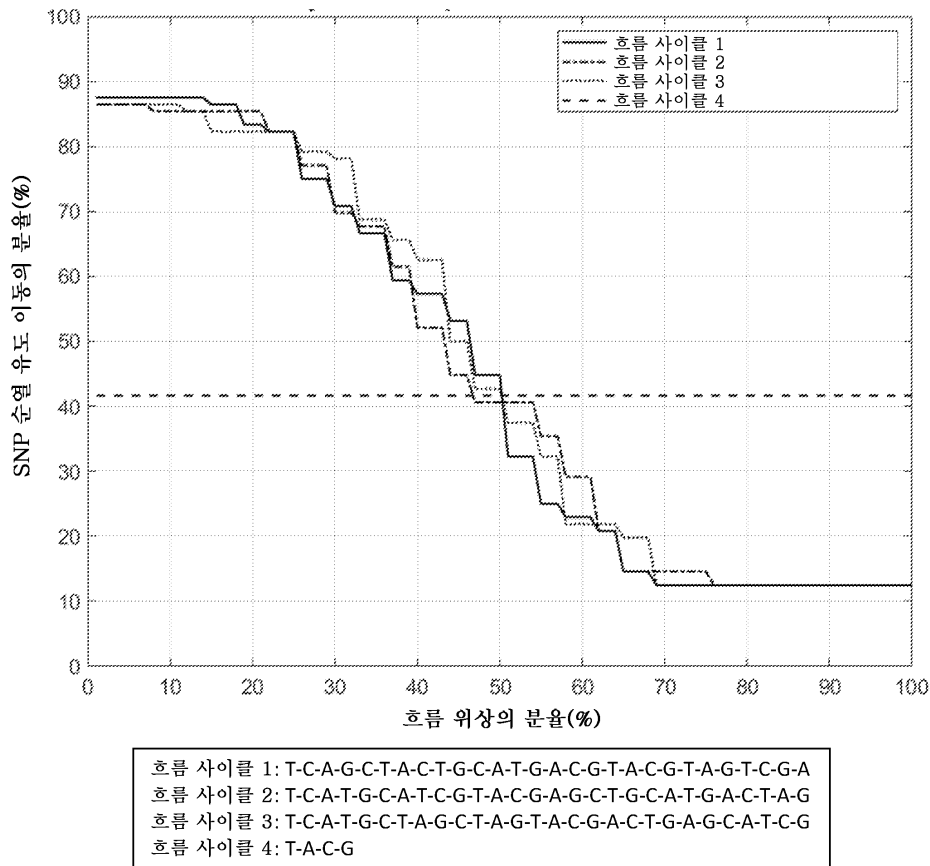
도면7



도면8

열기 카운트	사이클 1				사이클 2				사이클 3				사이클 4				사이클 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.0010	0.0010	0.9988	0.9988	0.0010	0.9988	0.9988	0.0001	0.0010	0.9988	0.0010	0.0010	0.9988	0.0010	0.9988	0.0010	0.9988	0.0010	0.9988	0.0010
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.9979	0.0010	0.9979	0.9979	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9979	0.9979
2	0.0010	0.0010	0.0001	0.0001	0.0010	0.0001	0.0001	0.9979	0.0010	0.0001	0.0010	0.0010	0.0001	0.0010	0.0001	0.0001	0.0010	0.0001	0.0010	0.0010
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

도면9



서열 목록

SEQUENCE LISTING

<110> Ultima Genomics, Inc.

<120> METHODS FOR DETECTING NUCLEIC ACID VARIANTS

<130> 16527-20005.40

<140> Not Yet Assigned

<141> Concurrently Herewith

<150> US 62/971,530

<151> 2020-02-07

<150> US 62/842,534

<151> 2019-05-03

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 1

tatggtcgtc ga 12

<210> 2

<211> 12

<212>

> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 2

tatggtcatc ga 12

<210> 3

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 3

tggtcgtcga gc 12

<210> 4

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 4

tatatggtcg tc 12

<

<210> 5

<211> 19

<212> DNA

<213> Artificial Sequence

<220>
 <223> Synthetic Construct
 <400> 5
 tatatggtca tcgagctat 19
 <210> 6
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 6
 tatatggtcg tcgagctat 19
 <210> 7
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 7
 tatggtcgtc ga 12
 <210> 8
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 8
 tatggtcgat cg 12