



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0154690
(43) 공개일자 2022년11월22일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6806 (2018.01) C12Q 1/6827 (2018.01)
C12Q 1/686 (2018.01) C12Q 1/6869 (2018.01)
(52) CPC특허분류
C12Q 1/6806 (2018.05)
C12Q 1/6827 (2018.05)
(21) 출원번호 10-2022-7031690
(22) 출원일자(국제) 2021년02월12일
심사청구일자 없음
(85) 번역문제출일자 2022년09월14일
(86) 국제출원번호 PCT/US2021/017937
(87) 국제공개번호 WO 2021/163546
국제공개일자 2021년08월19일
(30) 우선권주장
62/977,066 2020년02월14일 미국(US)

(71) 출원인
더 존스 홉킨스 유니버시티
미합중국 매릴랜드 21218 볼티모어 노쓰 찰스 스트리트 3400
(72) 발명자
파파도폴로스 니콜라스
미국 매릴랜드 21204, 토우슨, 혼크레스트 606
킨즐러 케네스 더블유
미국, 매릴랜드 21230, 볼티모어, 폰트 빌라스 노스 616
(74) 대리인
강명구

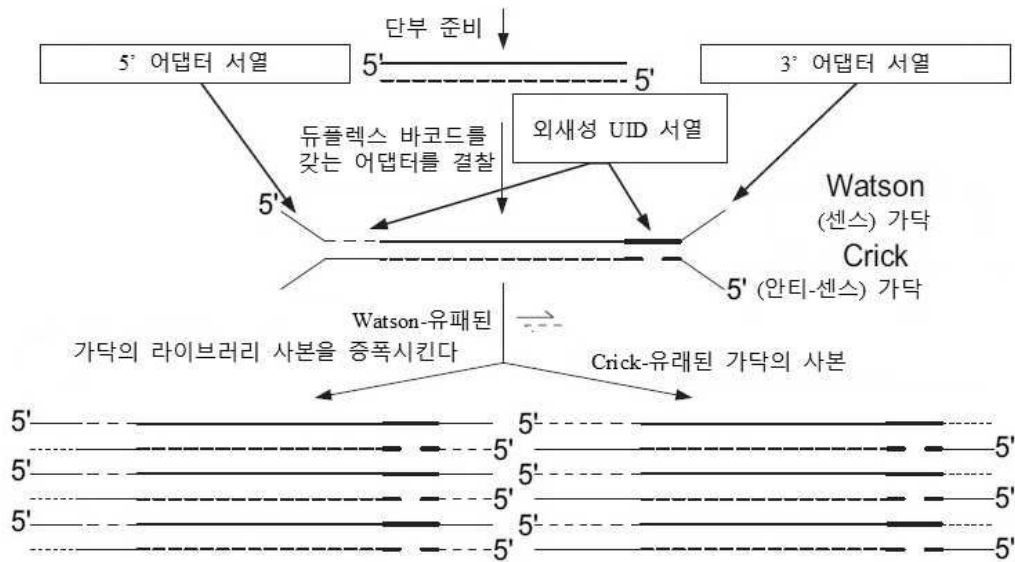
전체 청구항 수 : 총 212 항

(54) 발명의 명칭 핵산 평가 방법 및 재료

(57) 요약

퀵싱 라이브러리 준비 및 시퀀싱 작업흐름 (가령, 돌연변이 식별용)을 위한 시스템, 키트, 조성물 및 방법들이 본원에서 제공된다. 특정 구체예들에서, 템플릿의 두 가닥을 동일하게 마코딩하는 시스템 및 방법, 그리고 하이브리드화 캡처가 필요하지 않은 각 가닥의 PCR 기반 풍부화를 위한 시스템 및 방법들이 본원에서 제공된다.

대표도 - 도1



(52) CPC특허분류

C12Q 1/686 (2018.05)

C12Q 1/6869 (2018.05)

C12Q 2521/501 (2013.01)

C12Q 2521/531 (2013.01)

C12Q 2525/155 (2013.01)

C12Q 2525/186 (2019.08)

C12Q 2525/191 (2013.01)

C12Q 2531/113 (2013.01)

C12Q 2535/122 (2019.08)

(72) 발명자

보겔스테인 버트

미국, 매릴랜드 21208, 볼티모어, 브레튼 웨이
3700

코헨 조수아 데이비드

미국, 매릴랜드 21231, 볼티모터, 오르레안스 스트
리트 1650

명세서

청구범위

청구항 1

다음을 포함하는, 방법:

a) 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs)를 부착시키고,

이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) 다음을 포함하며: (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열, 그리고

이때 상기 3'PDSAs의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고,

b) 어닐링 5' 어댑터를 전술한 어닐링 부위에 어닐링시키고, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고;

c) 상기 외인성 UID 서열 및 전술한 제 1 세그먼트에 걸쳐 상기 5' 어댑터를 연장시키고, 이로써 전술한 외인성 UID 서열의 보체와 전술한 제 1 세그먼트의 보체가 생성되며, 그리고

d) 전술한 제 1 세그먼트의 전술한 보체의 3' 단부를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고, 이로써 어댑터-결찰된 이중-가닥으로 된 DNA 단편들 다수가 생성된다.

청구항 2

청구항 1에 있어서, 다음을 더 포함하는 방법: 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들 다수를 전술한 범용 3' 어댑터 서열에 상보적인 제 1 프라이머와 전술한 범용 5' 어댑터 서열의 보체에 상보적인 제 2 프라이머로 증폭시키고, 이로써 애플리콘이 생성되며, 이때 전술한 애플리콘은 이중-가닥으로 된 Watson 템플릿 다수와 이중-가닥으로 된 Crick 템플릿 다수를 포함한다.

청구항 3

청구항 2에 있어서, 다음을 더 포함하는 방법: 전술한 이중-가닥으로 된 Watson 템플릿을 Watson 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함한다.

청구항 4

청구항 3에 있어서, 다음을 더 포함하는 방법: 전술한 이중-가닥으로 된 Crick 템플릿을 Crick 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분의 보체에 상보적인 제 1 Crick 표적-선택적 프라이머, 그리고 (ii) 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함한다.

청구항 5

청구항 1에 있어서, 전술한 단일-가닥으로 된 3' 어댑터 (3'SSAs)가 생성되도록 하기 위해, 전술한 3'PDSAs의 제 2 가닥을 제거하는 단계를 더 포함하는, 방법.

청구항 6

청구항 5에 있어서, 이때 전술한 제 2 가닥의 제거 단계는 단계 b) 이후, 또는 단계 b) 이전, 또는 단계 b) 동안 일어나는, 방법.

청구항 7

청구항 5에 있어서, 이때 전술한 제 2 가닥은 하나 또는 그 이상의 데옥시우리딘을 포함하고, 이때 전술한 3'PDSAs의 제 2 가닥을 제거하는 단계는 3' 이중 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시켜, 전술한 제 2 가닥을 분해시키는 것을 포함하는, 방법.

청구항 8

청구항 5에 있어서, 이때 제 2 가닥을 제거하는 전술한 단계는 엑소뉴클레아제 활성을 갖는 중합효소에 의해 이루어지며, 이때 전술한 중합효소는 상기 외인성 UID 서열과 전술한 제 1 세그먼트에 걸쳐 5' 어댑터를 연장시키는, 방법.

청구항 9

청구항 2에 있어서, 하나 또는 그 이상의 전술한 앰플리콘의 서열 판독을 결정하는 단계를 더 포함하는, 방법.

청구항 10

청구항 9에 있어서, 이 서열 판독을 UID 패밀리에 할당하는 단계를 더 포함하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하는, 방법.

청구항 11

청구항 10에 있어서, R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하는 단계를 더 포함하는 방법.

청구항 12

청구항 11에 있어서, Watson 하위-패밀리의 적어도 50%가 해당 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 뉴클레오티드 서열을 식별해내는 단계를 더 포함하는, 방법.

청구항 13

청구항 12에 있어서, Crick 하위-패밀리의 적어도 50%가 해당 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 뉴클레오티드 서열을 식별해내는 단계를 더 포함하는, 방법.

청구항 14

청구항 12에 있어서, 상기 방법은 Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오티드 서열에서 돌연변이를 확인하는 단계를 더 포함하는, 방법.

청구항 15

청구항 14에 있어서, 상기 방법은 Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오티드 서열에서 돌연변이를 확인하는 단계를 더 포함하는, 방법.

청구항 16

청구항 15에 있어서, 뉴클레오티드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오티드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인하기 위한 명령을 더 포함하는, 방법.

청구항 17

청구항 10에 있어서, 이때 UID 패밀리 각 구성원은 동일한 내인성 UID 서열을 더 포함하며, 이때 상기 내인성 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함하는, 방법.

청구항 18

청구항 1에 있어서, 이때 전술한 이중-가닥으로 된 DNA 단편들의 집단은 블런트 단부를 갖는, 방법.

청구항 19

다음을 포함하는 시스템:

a) 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 결합되도록 의도된 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs) 집단,

이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고

이때 상기 3'PDSAs의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고; 그리고

b) 전술한 어닐링 부위에 어닐링되도록 의도된 5' 어댑터 집단, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함한다.

청구항 20

청구항 19에 있어서, 다음을 더 포함하는 시스템: c) 생물학적 샘플로부터 유래된 전술한 이중-가닥으로 된 DNA 단편들의 집단.

청구항 21

청구항 20에 있어서, 이때 전술한 이중-가닥으로 된 DNA 단편들의 집단은 블런트 단부를 갖는, 시스템.

청구항 22

청구항 19에 있어서, c) 단일-가닥으로 된 3' 어댑터 (3'SSAs)가 생성되도록 하기 위해, 전술한 3'PDSAs의 제 2 가닥을 분해시키는 시약을 더 포함하는 시스템.

청구항 23

청구항 19에 있어서, c) 전술한 범용 3' 어댑터 서열에 상보적인 제 1 프라이머, 그리고 전술한 범용 5' 어댑터 서열의 보체에 상보적인 제 2 프라이머를 더 포함하는 시스템.

청구항 24

청구항 19에 있어서, c) 전술한 범용 3' 어댑터 서열에 상보적인 Watson 고정 프라이머, 그리고 d) 전술한 범용 5' 어댑터 서열에 상보적인 Crick 고정 프라이머를 더 포함하는 시스템.

청구항 25

청구항 19에 있어서, 다음을 더 포함하는 시스템:

c) Watson 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 하나 또는 그 이상의 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 및 (ii) 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머를 포함하며, 이들 제 2 Watson 표적-선택적 프라이머의 각각은 하나 또는 그 이상의 표적-선택적 서열을 포함하고, 그리고

d) Crick 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분의 보체에 대해 상보적인 서열을 포함하는 하나 또는 그 이상의 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그

이상의 제 2 Crick 표적-선택적 프라이머를 포함하며, 상기 하나 또는 그 이상의 제 2 Crick 표적 선택적 프라이머의 각각은 상기 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함한다.

청구항 26

다음을 포함하는, 방법:

a) 다음을 포함하는 반응 혼합물을 형성시키고:

- i) 탈-인산화되고, 단부가 블런트화된(blunt-ended) 이중-가닥으로 된 DNA 단편 다수, 이때 상기 이중-가닥으로 된 DNA 단편들 각각은 Watson 가닥과 Crick 가닥을 포함하며;
- ii) 다수의 어댑터, 이때 전술한 어댑터 각각은 (5'에서 3'방향으로) 다음을 포함하고: A) 바코드, 및 B) 범용 3' 어댑터 서열; 그리고
- iii) 리가아제; 그리고

b) 전술한 반응 혼합물을 항온처리하며: i) 어댑터들은 Watson 가닥과 Crick 가닥의 3' 단부에 결합되며, 그리고 ii) 어댑터들은 Watson 가닥 또는 Crick 가닥의 5' 단부에 결합되지 않고, 이로써 이중-가닥으로 된 결합 산물이 생성된다.

청구항 27

청구항 26에 있어서, 이때 전술한 다수의 어댑터 각각은 독특한 바코드를 포함하는, 방법.

청구항 28

청구항 27에 있어서, 이때 전술한 이중-가닥으로 된 결합 산물 각각은 오직 한 개 바코드를 갖는 Watson 가닥과 오직 한 개 바코드 Crick 가닥을 포함하며, 이때 전술한 바코드는 전술한 Watson 가닥의 바코드와는 상이한, 방법.

청구항 29

포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한 방법에 있어서, 이때 상기 방법은 다음을 포함하는, 방법:

- A) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고;
- B) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 증폭시켜 증폭된 듀플렉스 시퀀싱 라이브러리를 생성시키고, 이때 상기 증폭 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 범용 프라이머 쌍을 접촉시키는 것을 포함하고;
- C) 임의선택적으로, 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 단일 가닥으로 된 DNA 라이브러리를 생성시키고;
- D) 임의선택적으로, 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 단일 가닥으로 된 DNA 라이브러리를 생성시키고;
- E) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머를 포함하는 프라이머 쌍을 이용하여 Watson 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키고;
- F) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머를 포함하는 프라이머 쌍을 이용하여 Crick 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키고;
- G) 시퀀싱 판독을 생성시키고, 표적 영역의 Watson 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Watson 가닥의 DNA 라이브러리로부터 증폭된 표적 영역을 시퀀싱하고;
- H) 시퀀싱 판독을 생성시키고, 표적 영역의 Crick 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Crick 가닥의 DNA 라이브러리로부터 증폭된 표적 영역을 시퀀싱하고;

I) 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 이 돌연변이가 존재하는 지를 결정하기 위해, 각 시퀀싱 판독에 존재하는 분자 바코드를 이용하여 상기 시퀀싱 판독을 그룹화시킨다.

청구항 30

청구항 29에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함하는, 방법:

i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 이중 어댑터를 결합시키는 단계, 이때 3' 이중 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며;

ii) 상기 분해가능한 3' 차단 그룹을 분해하는 단계;

iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계, 이때 상기 5' 이중 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 남겨두고; 그리고

iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다.

청구항 31

청구항 29에 있어서, 이때 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 DNA 라이브러리를 생성시키는 단계는 다음을 포함하는, 방법:

i) 태그된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 생성시키기 위해, 제 1 프라이머와 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 증폭된 듀플렉스 시퀀싱 라이브러리의 제 1 분액(aliquot)을 증폭시키는 단계, 이때 상기 제 1 프라이머는 Watson 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 태그(tag)를 포함하여;

ii) 단일 가닥으로 된 태그된 Watson 가닥과 단일 가닥으로 된 Crick 가닥을 생성시키기 위해, 태그된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변형시키는 단계; 그리고

iii) 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 DNA 라이브러리가 생성되도록 단일 가닥으로 된 태그된 Watson 가닥을 회수하는 단계.

청구항 32

청구항 29-31중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 포유류로부터 취한 샘플로부터 획득하고, 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 DNA 라이브러리를 생성시키는 단계는 다음을 포함하는, 방법:

i) 태그된 Crick 가닥을 갖는 이중 가닥으로 된 증폭 산물을 생성시키기 위해, 제 1 프라이머와 제 2 프라이머를 포함하는 프라이머 쌍을 이용하여 증폭된 듀플렉스 시퀀싱 라이브러리의 제 2 분액(aliquot)을 증폭시키는 단계, 이때 상기 제 1 프라이머는 Crick 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 태그(tag)를 포함하여;

ii) 단일 가닥으로 된 태그된 Crick 가닥과 단일 가닥으로 된 Watson 가닥을 생성시키기 위해, 태그된 Crick 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변형시키는 단계; 그리고

iii) 상기 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 DNA 라이브러리가 생성되도록 단일 가닥으로 된 태그된 Crick 가닥을 회수하는 단계.

청구항 33

청구항 29-32중 임의의 한 항에 있어서, 이때 포유류는 인간인, 방법.

청구항 34

청구항 29-33중 임의의 한 항에 있어서, 이때 상기 방법은 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 만들기 전, 다음의 단계를 더 포함하는 방법:

이중 가닥으로 된 DNA 단편들을 생성시키기 위해 상기 이중 가닥으로 된 DNA를 단편화시키는 단계;

상기 이중 가닥으로 된 DNA 단편들의 5' 단부를 탈인산화시키는 단계; 그리고

이중 가닥으로 된 DNA 단편들의 단부를 블런트화시키는 단계.

청구항 35

청구항 29-34중 임의의 한 항에 있어서, 이때상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 36

청구항 35에 있어서, 이때 상기 리가아제는 T4 DNA 리가아제인, 방법.

청구항 37

청구항 29-36중 임의의 한 항에 있어서, 이때 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함하는, 방법.

청구항 38

청구항 29-37중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 39

청구항 38에 있어서, 이때 상기 리가아제는 *대장균(Escherichia coli)* 리가아제인, 방법.

청구항 40

청구항 29-39중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에 단일 가닥으로 된 핵산의 갭을 채우는 단계는 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함하는, 방법.

청구항 41

청구항 40에 있어서, 이때 상기 중합효소는 Taq 중합효소인, 방법.

청구항 42

청구항 29-41중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 갭을 채우는 단계는 동시에 수행되는, 방법.

청구항 43

청구항 29-42중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 증폭시켜 증폭된 듀플렉스 시퀀싱 라이브러리를 생성시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이

중 가닥으로 된 DNA 단편에 범용 프라이머 쌍을 접촉시키는 것을 포함하는, 방법.

청구항 44

청구항 43에 있어서, 이때 이러한 증폭은 온전한-게놈 PCR을 포함하는, 방법.

청구항 45

청구항 29-44중 임의의 한 항에 있어서, 이때 상기 테그된 프라이머는 바이오티닐화된 프라이머이며, 그리고 이때 상기 바이오티닐화된 프라이머는 바이오티닐화된 단일 가닥으로 된 Watson 가닥과 바이오티닐화된 단일 가닥으로 된 Crick 가닥을 생성시킬 수 있는, 방법.

청구항 46

청구항 45에 있어서, 이때 상기 변성(denaturing) 단계는 NaOH 변성, 열 변성, 또는 이 둘의 조합을 포함하는, 방법.

청구항 47

청구항 45 또는 청구항 46에 있어서, 이때 상기 회수 단계는 상기 테그된 Watson 가닥에 스트렙타아비딘-기능화된 비드를 접촉시키고, 상기 테그된 Crick 가닥에 스트렙타아비딘-기능화된 비드를 접촉시키는 것을 포함하는, 방법.

청구항 48

청구항 47에 있어서, 이때 상기 회수 단계는 변성 비-테그된 Watson 가닥을 변성화시키고, 비-테그된 Watson 가닥을 변성화시키는 것을 더 포함하는, 방법.

청구항 49

청구항 47 또는 청구항 48에 있어서, 이때 상기 회수 단계는 상기 스트렙타아비딘-기능화된 비드로부터 바이오티닐화된 단일 가닥으로 된 Watson 가닥을 방출시키고, 상기 스트렙타아비딘-기능화된 비드로부터 바이오티닐화된 단일 가닥으로 된 Crick 가닥을 방출시키는 것을 더 포함하는, 방법.

청구항 50

청구항 29-44중 임의의 한 항에 있어서, 이때 상기 테그된 프라이머는 인산화된 프라이머이며, 그리고 이때 상기 인산화된 프라이머는 인산화된 단일 가닥으로 된 Watson 가닥과 인산화된 단일 가닥으로 된 Crick 가닥을 생성시킬 수 있는, 방법.

청구항 51

청구항 50에 있어서, 이때 상기 변성 단계들은 람다 엑소뉴클레아제 절단을 포함하는, 방법.

청구항 52

청구항 29-51중 임의의 한 항에 있어서, 이때 Watson 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머를 포함하는 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하는, 방법.

청구항 53

청구항 29-52중 임의의 한 항에 있어서, 이때 상기 시퀀싱 단계들은 페어드-엔드 시퀀싱을 포함하는, 방법.

청구항 54

포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한

방법에 있어서, 이때 상기 방법은 다음을 포함하는, 방법:

- A) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고;
- B) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편으로부터 Watson 가닥의 DNA 라이브러리와 Crick 가닥의 DNA 라이브러리를 생성시키고;
- C) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머 및 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 단일 가닥으로 된 Watson 가닥으로부터 표적 영역을 증폭시키고;
- D) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 단일 가닥으로 된 Crick 가닥으로부터 표적 영역을 증폭시키고;
- E) 시퀀싱 판독을 생성시키고, 표적 영역의 Watson 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Watson 가닥의 DNA 라이브러리로부터 증폭된 표적 영역을 시퀀싱하고;
- F) 시퀀싱 판독을 생성시키고, 표적 영역의 Crick 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Crick 가닥의 DNA 라이브러리로부터 증폭된 표적 영역을 시퀀싱하고;
- G) 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 이 돌연변이가 존재하는 지를 결정하기 위해, 각 시퀀싱 판독에 존재하는 분자 바코드를 이용하여 상기 시퀀싱 판독을 그룹화시킨다.

청구항 55

청구항 54에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플이며, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함하는, 방법:

- i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 이중 어댑터를 결합시키는 단계, 이때 3' 이중 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며;
- ii) 상기 분해가능한 3' 차단 그룹을 분해하는 단계;
- iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계, 이때 상기 5' 이중 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 남겨두고; 그리고
- iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다.

청구항 56

청구항 54에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 무-세포 DNA 샘플이며, 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 DNA 라이브러리와 Crick 가닥의 DNA 라이브러리를 생성시키는 다음을 포함하는 방법:

- i) 범용 제 1 프라이머와 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키는 단계, 이때 이러한 증폭 단계는 온전한-게놈 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 상기 프라이머 쌍을 접촉시키는 것을 포함하고, 이때 상기 제 1 프라이머 Watson 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 바이오티닐화되어, 바이오티닐화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물이 생성되며;
- ii) 스트렙타아비딘-기능화된 비드에 결합하는 조건 하에서, 스트렙타아비딘-기능화된 비드에 상기 바이오티닐

화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 접촉시키는 단계;

iii) 상기 스트렙타비딘-기능화된 비드에 결합된 단일 가닥으로 된 바이오티닐화된 Watson 가닥은 남겨두고, 단일 가닥으로 된 Crick 가닥을 방출시키기 위해, 상기 바이오티닐화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변형시키는 단계;

iv) 단일 가닥으로 된 Crick 가닥을 수거하는 단계;

v) 상기 스트렙타비딘-기능화된 비드로부터 단일 가닥으로 된 바이오티닐화된 Watson 가닥을 방출시키는 단계; 그리고

vi) 단일 가닥으로 된 바이오티닐화된 Watson 가닥을 수집하는 단계.

청구항 57

청구항 54-56중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 포유류로부터 취한 샘플로부터 획득하는, 단계.

청구항 58

청구항 54-57중 임의의 한 항에 있어서, 이때 상기 포유류는 인간인, 방법.

청구항 59

청구항 54-58중 임의의 한 항에 있어서, 이때 상기 방법은 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 만들기 전, 다음의 단계를 더 포함하는 방법:

이중 가닥으로 된 DNA 단편들을 생성시키기 위해 상기 이중 가닥으로 된 DNA를 단편화시키는 단계;

상기 이중 가닥으로 된 DNA 단편들의 5' 단부를 탈인산화시키는 단계; 그리고

이중 가닥으로 된 DNA 단편들의 단부를 블런트화시키는 단계.

청구항 60

청구항 54-59중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 61

청구항 60에 있어서, 이때 상기 리가아제는 T4 DNA 리가아제인, 방법.

청구항 62

청구항 54-61중 임의의 한 항에 있어서, 이때 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함하는, 방법.

청구항 63

청구항 54-62중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 64

청구항 63에 있어서, 이때 상기 리가아제는 대장균(*Escherichia coli*) 리가아제인, 방법.

청구항 65

청구항 54-64중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에

단일 가닥으로 된 핵산의 갭을 채우는 단계는 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함하는, 방법.

청구항 66

청구항 65에 있어서, 이때 상기 중합효소는 Taq-B 중합효소인, 방법.

청구항 67

청구항 54-66중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 갭을 채우는 단계는 동시에 수행되는, 방법.

청구항 68

청구항 54-67중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 프라이머 쌍을 접촉시키는 것을 포함하는, 방법.

청구항 69

청구항 68에 있어서, 이때 증폭 단계는 온전한-게놈 PCR을 포함하는, 방법.

청구항 70

청구항 54-69중 임의의 한 항에 있어서, 이때 Watson 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머를 포함하는 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하는, 방법.

청구항 71

청구항 54-70중 임의의 한 항에 있어서, 이때 상기 시퀀싱 단계들은 페어드-엔드 시퀀싱 또는 싱글-엔드 시퀀싱을 포함하는, 방법.

청구항 72

포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한 방법에 있어서, 이때 상기 방법은 다음을 포함하는, 방법:

- A) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고;
- B) 범용 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 증폭시키고, 이때 이러한 증폭 단계는 온전한-게놈 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 범용 프라이머 쌍을 접촉시키는 것을 포함하고;
- C) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 증폭된 이중 가닥으로 된 DNA 단편들의 Watson 가닥으로부터 표적 영역을 증폭시키고;
- D) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 증폭된 이중 가닥으로 된 DNA 단편들의 Crick 가닥으로부터 표적 영역을 증폭시키고;
- E) 시퀀싱 판독을 생성시키고, 표적 영역의 Watson 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Watson

가닥으로부터 증폭된 표적 영역을 시퀀싱하고;

F) 시퀀싱 판독을 생성시키고, 표적 영역의 Crick 가닥에 돌연변이의 존재 또는 부재를 알아내기 위해, Crick 가닥으로부터 증폭된 표적 영역을 시퀀싱하고;

G) 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 이 돌연변이가 존재하는 지를 결정하기 위해, 각 시퀀싱 판독에 존재하는 분자 바코드를 이용하여 상기 시퀀싱 판독을 그룹화시킨다.

청구항 73

청구항 72에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플이며, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함하는, 방법:

i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 이중 어댑터를 결합시키는 단계, 이때 3' 이중 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며;

ii) 상기 분해가능한 3' 차단 그룹을 분해하는 단계;

iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계, 이때 상기 5' 이중 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 남겨두고; 그리고

iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다.

청구항 74

청구항 73에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 무-세포 DNA 샘플인, 방법.

청구항 75

청구항 72-74중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플인, 방법.

청구항 76

청구항 72-75중 임의의 한 항에 있어서, 이때 상기 포유류는 인간인, 방법.

청구항 77

청구항 72-76중 임의의 한 항에 있어서, 이때 상기 방법은 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 만들기 전, 다음의 단계를 더 포함하는 방법:

이중 가닥으로 된 DNA 단편들을 생성시키기 위해 상기 이중 가닥으로 된 DNA를 단편화시키는 단계;

상기 이중 가닥으로 된 DNA 단편들의 5' 단부를 탈인산화시키는 단계; 그리고

이중 가닥으로 된 DNA 단편들의 단부를 블런트화시키는 단계.

청구항 78

청구항 72-77중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 79

청구항 50에 있어서, 이때 상기 리가아제는 T4 DNA 리가아제인, 방법.

청구항 80

청구항 72-79중 임의의 한 항에 있어서, 이때 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함하는, 방법.

청구항 81

청구항 72-80중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함하는, 방법.

청구항 82

청구항 81에 있어서, 이때 상기 리가아제는 *대장균*(*Escherichia coli*) 리가아제인, 방법.

청구항 83

청구항 72-82중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에 단일 가닥으로 된 핵산의 갭을 채우는 단계는 DNA 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함하는, 방법.

청구항 84

청구항 83에 있어서, 이때 DNA 중합효소는 Taq-B 중합효소인, 방법.

청구항 85

청구항 72-84중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 갭을 채우는 단계는 동시에 수행되는, 방법.

청구항 86

청구항 72-85중 임의의 한 항에 있어서, 이때 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 프라이머 쌍을 접촉시키는 것을 포함하는, 방법.

청구항 87

청구항 86에 있어서, 이때 증폭 단계는 온전한-게놈 PCR을 포함하는, 방법.

청구항 88

청구항 72-87중 임의의 한 항에 있어서, 이때 Watson 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의 DNA 라이브러리로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머를 포함하는 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하는, 방법.

청구항 89

청구항 72-88중 임의의 한 항에 있어서, 이때 상기 시퀀싱 단계들은 페어드-엔드 시퀀싱을 포함하는, 방법.

청구항 90

다음을 포함하는, 방법:

- a. 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터를 부착시키고, 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며;
- b. 상기 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키고, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고;
- c. 니크 해독(nick translation)-유사 반응을 실행하여, 3' 어댑터의 외인성 UID 서열에 걸쳐 5' 어댑터를 연장시키고, 이렇게 연장된 5' 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고;
- d. 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;
- e. 상기 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하고;
- f. 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- g. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하고;
- h. 뉴클레오타이드 서열은 Watson 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인되며;
- i. 뉴클레오타이드 서열은 Crick 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인되며;
- j. Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하고;
- k. Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하고; 그리고
- l. 뉴클레오타이드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오타이드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인한다.

청구항 91

청구항 90에 있어서, 이때 UID 패밀리의 각 구성원은 동일한 내인성 UID 서열을 더 포함하며, 이때 상기 내인성 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함하는, 방법.

청구항 92

청구항 91에 있어서, 이때 상기 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열은 적어도 8개, 10개, 또는 15개의 염기를 포함하는, 방법.

청구항 93

청구항 90-92중 임의의 한 항에 있어서, 이때 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특한, 방법.

청구항 94

청구항 90-92중 임의의 한 항에 있어서, 이때 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독 특하지 않는, 방법.

청구항 95

청구항 91-94중 임의의 한 항에 있어서, 이때 UID 패밀리 각 구성원은 동일한 내인성 UID 서열과 동일한 외인 성 UID 서열을 포함하는, 방법.

청구항 96

전술한 청구항들중 임의의 항에 있어서, 이때 단계 (d)는 단지 11회 주기의 PCR 증폭을 포함하는, 방법.

청구항 97

청구항 96에 있어서, 이때 단계 (d)는 단지 7회 주기의 PCR 증폭을 포함하는, 방법.

청구항 98

청구항 97에 있어서, 이때 단계 (d)는 5회를 넘지 않는 주기의 PCR 증폭을 포함하는, 방법.

청구항 99

전술한 청구항들중 임의의 항에 있어서, 이때 단계 (d)는 적어도 1회 주기의 PCR 증폭을 포함하는, 방법.

청구항 100

전술한 청구항들중 임의의 항에 있어서, 이때, 상기 서열 판독을 결정하기 전, 상기 앰플리콘은 하나 또는 그 이상의 표적 폴리뉴클레오티드에 대해 풍부화되는, 방법.

청구항 101

청구항 100에 있어서, 이때 상기 풍부화 단계는 다음을 포함하는, 방법:

- a. 표적 폴리뉴클레오티드 서열을 포함하는 Watson 가닥의 앰플리콘을 Watson 표적-선택적 프라이머 쌍의 제 1 세트에 임의선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택 적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열 의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이 머를 포함하며; 그리고
- b. 동일한 표적 폴리뉴클레오티드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍 의 제 1 세트에 임의선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적- 선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열 의 R1 시퀀싱 프라이머 부위이며], 그리고 (ii) 제 2 Watson 표적-선택적 프라이머 서열의 표적-선택적 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함한다.

청구항 102

청구항 101에 있어서, 비-표적 폴리뉴클레오티드로부터 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산 물을 정제해내는 단계를 더 포함하는, 방법.

청구항 103

청구항 102에 있어서, 이때 상기 정제 방법은 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 고정 서포트에 부착시키는 단계를 포함하는, 방법.

청구항 104

청구항 103에 있어서, 이때 상기 제 1 Watson 표적-선택적 프라이머와 제 1 Crick 표적-선택적 프라이머는 친화 력 결합 쌍의 제 1 구성원을 포함하고, 그리고 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을

포함하는, 방법.

청구항 105

청구항 104에 있어서, 이때 상기 제 1 구성원은 바이오틴과 상기 제 2 구성원은 스트렙타비딘인, 방법.

청구항 106

청구항 102-105중 임의의 한 항에 있어서, 이때 상기 고형 서포트는 비드, 웰, 막, 튜브, 컬럼, 플레이트, 세파로스, 자성 비드, 또는 칩을 포함하는, 방법.

청구항 107

청구항 102-106중 임의의 한 항에 있어서, 상기 고형 서포트에 부착되지 않은 폴리뉴클레오티드를 제거하는 단계를 포함하는, 방법.

청구항 108

청구항 101-107중 임의의 한 항에 있어서, 다음을 포함하는, 방법:

a. 상기 표적 Watson 증폭 산물을 Watson 표적-선택적 프라이머의 제 2 세트로 추가 증폭시키고, 이로써 표적 Watson 라이브러리 구성원을 창출시키고, 이때 상기 Watson 표적-선택적 프라이머의 제 2 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 3 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하는 제 4 Watson 표적-선택적 프라이머를 포함하고;

b. 상기 표적 Crick 증폭 산물을 Crick 표적-선택적 프라이머의 제 2 세트로 추가 증폭시키고, 이로써 표적 Crick 라이브러리 구성원을 창출시키고, 이때 상기 Crick 표적-선택적 프라이머의 제 2 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 5 Crick 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R1시퀀싱 프라이머 부위이며], 그리고 (ii) (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 상기 제 4 Watson 표적-선택적 프라이머의 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하는 제 4 Crick 표적-선택적 프라이머를 포함한다.

청구항 109

청구항 108에 있어서, 이때 상기 제 3 Watson 표적-선택적 프라이머와 Crick 표적-선택적 프라이머는 샘플 바코드 서열을 더 포함하는, 방법.

청구항 110

청구항 108 또는 109에 있어서, 이때 상기 제 3 Watson 표적-선택적 프라이머는 시퀀서(sequencer) 상에서 제 1 접목 프라이머에 혼성화될 수 있는 제 1 접목(grafting) 서열을 더 포함하고, 이때 상기 제 3 Crick 표적-선택적 프라이머는 이 시퀀서 상에서 제 2 접목 프라이머에 혼성화될 수 있는 제 2 접목 서열을 더 포함하는, 방법.

청구항 111

청구항 108-110중 임의의 한 항에 있어서, 이때 상기 제 4 Watson 표적-선택적 프라이머는 상기 제 2 접목 서열을 더 포함하고, 이때 상기 제 4 Crick 표적-선택적 프라이머는 상기 제 1 접목 서열을 더 포함하는, 방법.

청구항 112

청구항 110 또는 111에 있어서, 이때 상기 제 1 접목 서열은 P7 서열이며, 이때 상기 제 2 접목 서열은 P5 서열인, 방법.

청구항 113

청구항 101-112중 임의의 한 항에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 50%를 나타내는, 방법.

청구항 114

청구항 113에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 70%를 나타내는, 방법.

청구항 115

청구항 114에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 80%를 나타내는, 방법.

청구항 116

청구항 115에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 90%를 나타내는, 방법.

청구항 117

청구항 101-112중 임의의 한 항에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 50%를 나타내는, 방법.

청구항 118

청구항 117에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 70%를 나타내는, 방법.

청구항 119

청구항 118에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 80%를 나타내는, 방법.

청구항 120

청구항 119에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 90%를 나타내는, 방법.

청구항 121

다음을 포함하는, 방법:

- a. 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키는 단계, 이때 상기 어댑터는 외인성 UID를 포함하는 이중-가닥으로 된 일부분과 갈퀴형(forked) 일부분을 포함하고, 이때 갈퀴형 일부분은 (i) R2 시퀀싱 프라이머 부위를 포함하는 단일-가닥으로 된 3' 어댑터 서열 및 (ii) R1 시퀀싱 프라이머 부위를 포함하는 단일-가닥으로 된 5' 어댑터 서열을 포함하며;
- b. 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;
- c. 표적 폴리뉴클레오티드 서열을 포함하는 Watson 가닥의 앰플리콘을 Watson 표적-선택적 프라이머 쌍의 제 1 세트로 임의선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함하며; 그리고
- d. 동일한 표적 폴리뉴클레오티드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍의 제 1 세트로 임의선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위이며], 그리고 (ii) 제 2 Crick 표적-선택적 프라이머 서열의 표적-선택적 서열과

동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함하고;

- e. 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물의 서열 판독을 결정하고;
- f. 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- g. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하고;
- h. 뉴클레오타이드 서열은 Watson 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인되며;
- i. 뉴클레오타이드 서열은 Crick 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인되며; 그리고
- j. Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열과, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열 모두 동일한 돌연변이를 함유할 때, 피분석물 DNA 단편에서 돌연변이를 확인한다.

청구항 122

청구항 121에 있어서, 비-표적 폴리뉴클레오타이드로부터 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 정제해내는 단계를 포함하는, 방법.

청구항 123

청구항 122에 있어서, 이때 상기 정제 방법은 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 고품형 서포트에 부착시키는 단계를 포함하는, 방법.

청구항 124

청구항 123에 있어서, 이때 상기 제 1 Watson 표적-선택적 프라이머와 제 1 Crick 표적-선택적 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하고, 그리고 이때 상기 고품형 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함하는, 방법.

청구항 125

청구항 124에 있어서, 이때 상기 제 1 구성원은 바이오틴과 상기 제 2 구성원은 스트렙타비딘인, 방법.

청구항 126

청구항 122-125중 임의의 한 항에 있어서, 이때 상기 고품형 서포트는 비드, 웰, 막, 튜브, 컬럼, 플레이트, 세파로스, 자성 비드, 또는 칩을 포함하는, 방법.

청구항 127

청구항 122-126중 임의의 한 항에 있어서, 상기 고품형 서포트에 부착되지 않은 폴리뉴클레오타이드를 제거하는 단계를 포함하는, 방법.

청구항 128

청구항 121-127중 임의의 한 항에 있어서, 다음을 포함하는, 방법:

- a. 상기 표적 Watson 증폭 산물을 Watson 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Watson 라이브러리 구성원을 창출시키고, 이때 상기 Watson 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Watson 표적-선택적 프라이머를 포함하고;
- b. 상기 표적 Crick 증폭 산물을 Crick 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Crick 라이브러리 구성원을 창출시키고, 이때 상기 Crick 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어댑터 서열의 R1시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Crick 표적-선택적 프라이머, 그리고

(ii) (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 상기 제 4 Watson 표적-선택적 프라이머의 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Crick 표적-선택적 프라이머를 포함한다.

청구항 129

청구항 128에 있어서, 이때 상기 제 3 Watson 표적-선택적 프라이머와 Crick 표적-선택적 프라이머는 샘플 바코드 서열을 더 포함하는, 방법.

청구항 130

청구항 128 또는 129에 있어서, 이때 상기 제 3 Watson 표적-선택적 프라이머는 시퀀서(sequencer) 상에서 제 1 접목 프라이머에 혼성화될 수 있는 제 1 접목(grafting) 서열을 더 포함하고, 이때 상기 제 3 Crick 표적-선택적 프라이머는 이 시퀀서 상에서 제 2 접목 프라이머에 혼성화될 수 있는 제 2 접목 서열을 더 포함하는, 방법.

청구항 131

청구항 128-130중 임의의 한 항에 있어서, 이때 상기 제 4 Watson 표적-선택적 프라이머는 상기 제 2 접목 서열을 더 포함하고, 이때 상기 제 4 Crick 표적-선택적 프라이머는 상기 제 1 접목 서열을 더 포함하는, 방법.

청구항 132

청구항 130 또는 131에 있어서, 이때 상기 제 1 접목 서열은 P7 서열이며, 이때 상기 제 2 접목 서열은 P5 서열인, 방법.

청구항 133

청구항 121-131중 임의의 한 항에 있어서, 이때 상기 부착단계는 A-꼬리로 된(tailed) 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 집단에 부착시키는 단계를 포함하는, 방법.

청구항 134

청구항 133에 있어서, 이때 상기 부착단계는 A-꼬리로 된(tailed) 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 집단에 부착시키는 단계를 포함하는, 방법.

청구항 135

청구항 121-131중 임의의 한 항에 있어서, 이때 상기 부착단계는 다음을 포함하는, 방법:

- a. 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터를 부착시키고, 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 임의선택적으로, 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며; 그리고
- b. 상기 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키고, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고; 그리고
- c. 니크 해독(nick translation)-유사 반응을 실행하여, 3' 어댑터에 걸쳐 5' 어댑터를 연장시키고, 이렇게 연장된 5' 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연결시킨다.

청구항 136

청구항 121-135중 임의의 한 항에 있어서, 이때 상기 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열을 포함하는, 방법.

청구항 137

청구항 136에 있어서, 이때 상기 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열은 적어도 8개, 10개, 또는 15개의 염기를 포함하는, 방법.

청구항 138

청구항 121-136중 임의의 한 항에 있어서, 이때 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특한, 방법.

청구항 139

청구항 121-136중 임의의 한 항에 있어서, 이때 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특하지 않는, 방법.

청구항 140

청구항 136-139중 임의의 한 항에 있어서, 이때 UID 패밀리와 각 구성원은 동일한 내인성 UID 서열과 동일한 외인성 UID 서열을 포함하는, 방법.

청구항 141

청구항 121-140중 임의의 한 항에 있어서, 이때 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 11회를 넘지 않는 사이클의 PCR 증폭을 포함하는, 방법.

청구항 142

청구항 141에 있어서, 이때 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 7회를 넘지 않는 사이클의 PCR 증폭을 포함하는, 방법.

청구항 143

청구항 142에 있어서, 이때 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 5회를 넘지 않는 사이클의 PCR 증폭을 포함하는, 방법.

청구항 144

전술한 청구항들중 임의의 항에 있어서, 이때 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 적어도 1회 사이클의 PCR 증폭을 포함하는, 방법.

청구항 145

청구항 121-143중 임의의 한 항에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 50%를 나타내는, 방법.

청구항 146

청구항 145에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 70%를 나타내는, 방법.

청구항 147

청구항 146에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 80%를 나타내는, 방법.

청구항 148

청구항 147에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오티드의 적어도 90%를 나타내는, 방법.

청구항 149

청구항 121-143중 임의의 한 항에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 50%를 나타내는, 방법.

청구항 150

청구항 149에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 70%를 나타내는, 방법.

청구항 151

청구항 150에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 80%를 나타내는, 방법.

청구항 152

청구항 151에 있어서, 이때 상기 표적 Watson 라이브러리 구성원들과 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 90%를 나타내는, 방법.

청구항 153

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정으로 템플릿 분자의 양 단부 모두의 서열 결정이 가능하게 되는, 방법.

청구항 154

청구항 153에 있어서, 이때 템플릿 분자의 양 단부의 결정 단계는 페어드 엔드 시퀀싱을 포함하는, 방법.

청구항 155

전술한 청구항들중 임의의 항에 있어서, 이때 상기 서열 판독의 결정 단계는 이러한 서열 판독을 생성시키기 위한 템플릿의 길이에 걸쳐 단일 판독 시퀀싱을 포함하는, 방법.

청구항 156

전술한 청구항들중 임의의 항에 있어서, 이때 상기 서열 판독의 결정은 대규모 병렬 시퀀서를 이용한 시퀀싱을 포함하는, 방법.

청구항 157

청구항 156에 있어서, 이때 상기 대규모 병렬 시퀀서는 템플릿 폴리뉴클레오타이드의 양 단부로부터 서열 판독을 결정하도록 설정된, 방법.

청구항 158

전술한 청구항들중 임의의 항에 있어서, 이때 상기 이중-가닥으로 된 DNA 단편들의 집단은 길이가 약 50-600nt인 하나 또는 그 이상의 단편을 포함하는, 방법.

청구항 159

전술한 청구항들중 임의의 항에 있어서, 이때 상기 이중-가닥으로 된 DNA 단편들의 집단은 길이가 2000nt 미만, 1000nt 미만, 500nt 미만, 400nt 미만, 300nt 미만, 또는 250nt 미만인 하나 또는 그 이상의 단편을 포함하는, 방법.

청구항 160

청구항 101-159중 임의의 한 항에 있어서, 상기 방법은 초기 증폭-후 그리고 선택적 증폭-전, 상기 앰플리콘의 센스 가닥 및 안티센스 가닥에 대응하는 단일 가닥으로 된 (ss) DNA 라이브러리를 준비하는 단계를 더 포함하는, 방법.

청구항 161

청구항 160에 있어서, 이때 상기 ss DNA 라이브러리 준비는 다음을 포함하는, 방법:

- a. 두 개 프라이머를 이용하여 증폭 반응을 수행하고 [이때 두 개 프라이머중 오로지 한 개 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하며], 이로써 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥을 포함하는 증폭 산물이 창출되며;
- b. 상기 증폭 산물을 고정 서포트에 접착시키고, 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함하며;
- c. 변성 상기 증폭 산물을 변성시켜, 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥으로부터 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥을 분리해내고; 그리고
- d. 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 분리된 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 분리된 가닥을 정제한다.

청구항 162

청구항 161에 있어서, 이때 상기 친화력 결합 쌍의 제 1 구성원은 바이오틴이며, 상기 친화력 결합 쌍의 제 2 구성원은 스트렙타비딘인, 방법.

청구항 163

청구항 160에 있어서, 이때 상기 ss DNA 라이브러리 준비는 다음을 포함하는, 방법:

- a. 상기 앰플리콘을 두 가지 증폭 반응으로 분할시키고, 이로써 인산화된 가닥과 비-인산화된 가닥을 포함하는 증폭 산물이 창출되며, 이때 각 증폭 반응은 진행(forward) 프라이머와 역행(reverse) 프라이머를 이용하며, 이때 두 개 프라이머중 오로지 한 개만 인산화되며;
- b. 5' 포스페이트를 갖는 가닥만을 선택적으로 절단시키는 엑소뉴클레아제에 상기 증폭 산물을 접착시킨다.

청구항 164

청구항 163에 있어서, 이때:

- a. 상기 제 1 증폭 반응에서, 상기 진행 프라이머는 인산화되고, 상기 역행 프라이머는 비-인산화되고;
- b. 상기 제 2 증폭 반응에서, 상기 역행 프라이머는 인산화되고, 상기 진행 프라이머는 비-인산화되는, 방법.

청구항 165

청구항 163에 있어서, 이때 상기 엑소뉴클레아제는 람다 엑소뉴클레아제인, 방법.

청구항 166

청구항 163-165중 임의의 한 항에 있어서, 이때 상기 인산화는 상기 5' 부위에서 일어나는, 방법.

청구항 167

청구항 90-153중 임의의 한 항에 있어서, 이때 상기 초기 증폭은 다음을 포함하는, 방법:

- a. 프라이머 쌍으로 증폭시키고, 이로써 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥을 포함하는 증폭 산물이 창출되며, 이때 이 프라이머 쌍에서 두 개 프라이머중 오로지 한 개 프라이머만 친화력 결합 쌍의 제 1 구성원을 포함하고;
- b. 상기 증폭 산물을 고정 서포트에 접착시키고, 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함하며;
- c. 변성 상기 증폭 산물을 변성시켜, 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥으로부터 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥을 분리해내고; 그리고
- d. 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 분리된 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 분리된 가닥을 정제한다.

청구항 168

청구항 167에 있어서, 이때 상기 친화력 결합 쌍의 제 1 구성원은 바이오틴이며, 상기 친화력 결합 쌍의 제 2 구성원은 스트렙타비딘인, 방법.

청구항 169

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R2 서열의 하류이며, R1 서열의 상류일 경우, UID 패밀리 서열 관독은 Watson 하위-패밀리에 할당되는, 방법.

청구항 170

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R1 서열의 하류이며, R2 서열의 상류일 경우, UID 패밀리 서열 관독은 Crick 하위-패밀리에 할당되는, 방법.

청구항 171

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R2 서열에 더 근접하고, R1 서열에 덜 근접할 경우, UID 패밀리 서열 관독은 Watson 하위-패밀리에 할당되는, 방법.

청구항 172

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R1 서열에 더 근접하고, R2 서열에 덜 근접할 경우, UID 패밀리 서열 관독은 Crick 하위-패밀리에 할당되는, 방법.

청구항 173

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R2 서열의 바로 하류에 있거나, 또는 1-300개, 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오타이드 이내에 있을 경우, UID 패밀리 서열 관독은 Watson 하위-패밀리에 할당되는, 방법.

청구항 174

전술한 청구항들중 임의의 항에 있어서, 이때 상기 외인성 UID 서열이 R1 서열의 바로 하류에 있거나, 또는 1-300개, 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오타이드 이내에 있을 경우, UID 패밀리 서열 관독은 Crick 하위-패밀리에 할당되는, 방법.

청구항 175

전술한 청구항들중 임의의 항에 있어서, 이때 상기 이중-가닥으로 된 DNA 단편들의 집단은 생물학적 샘플로부터 유래된, 방법.

청구항 176

청구항 175에 있어서, 이때 상기 생물학적 샘플은 대상체로부터 획득되는, 방법.

청구항 177

청구항 176에 있어서, 이때 상기 대상체는 인간 대상체인, 방법.

청구항 178

청구항 175-177중 임의의 항에 있어서, 이때 상기 생물학적 샘플은 유체 샘플인, 방법.

청구항 179

청구항 178에 있어서, 이때 상기 유체 샘플은 전혈, 혈장, 혈청, 가래, 소변, 땀, 눈물, 복수, 정액 및 기관지 폐포 세척액으로부터 선택되는, 방법.

청구항 180

청구항 178에 있어서, 이때 상기 유체 샘플은 무-세포이거나, 필수적으로 무-세포 샘플인, 방법.

청구항 181

청구항 175-177중 임의의 항에 있어서, 이때 상기 생물학적 샘플은 고체 생물학적 샘플인, 방법.

청구항 182

청구항 181에 있어서, 이때 상기 고체 생물학적 샘플은 종양 샘플인, 방법.

청구항 183

전술한 청구항들중 임의의 항에 있어서, 이때 상기 확인된 돌연변이는 0.1% 또는 그 미만의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재하는, 방법.

청구항 184

청구항 183에 있어서, 이때 상기 확인된 돌연변이는 0.1% ~ 0.00001%의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재하는, 방법.

청구항 185

청구항 183에 있어서, 이때 상기 확인된 돌연변이는 0.1% ~ 0.01%의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재하는, 방법.

청구항 186

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 50%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 187

청구항 186에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 70%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 188

청구항 187에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 80%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 189

청구항 188에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 90%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 190

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 50%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 191

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 70%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 192

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 80%의 Watson 가닥과 Crick 가닥 모두로 부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 193

전술한 청구항들중 임의의 항에 있어서, 이때 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 90%의 Watson 가닥과 Crick 가닥 모두로 부터 서열 판독을 결정하는 것을 포함하는, 방법.

청구항 194

전술한 청구항들중 임의의 항에 있어서, 이때 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 피분석물 DNA 단편의 Watson 가닥과 Crick 가닥 모두에서 돌연변이가 확인이 필요없는, 돌연변이가 확인하는 대체 방법과 비교하였을 때, 적어도 2-배, 4-배, 5-배, 10-배, 20-배, 30-배, 40-배, 50-배, 60-배, 70-배, 80-배, 90-배, 또는 100-배 감소되는, 방법.

청구항 195

청구항 194에 있어서, 이때 상기 대체 방법은 표준 분자 바코딩 또는 표준 PCR-기반된 분자 바코딩을 포함하는, 방법.

청구항 196

청구항 195에 있어서, 이때 상기 대체 방법은 다음을 포함하는, 방법:

- 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키고, 이때 상기 어댑터는 독특한 외인성 UID를 포함하고;
- 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;
- 상기 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하고;
- 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- 뉴클레오티드 서열은 UID 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편을 정확하게 나타내는 것으로 확인되며; 그리고
- 피분석물 DNA 단편을 정확하게 나타내는 확인된 서열이 돌연변이가 결여된 참조 서열과 상이할 경우, 이 피분석물 DNA 단편에서 돌연변이가 확인된다.

청구항 197

전술한 청구항들중 임의의 항에 있어서, 이때 피분석물 DNA 단편들 선행 청구항들중 임의의 하나의 방법에 따라 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 단지 1×10^{-2} , 단지 1×10^{-3} , 단지 1×10^{-4} , 단지 1×10^{-5} , 단지 1×10^{-6} , 단지 5×10^{-6} , 또는 단지 1×10^{-7} 인, 방법.

청구항 198

핵산 샘플의 서열 판독 데이터를 분석하기 위한 컴퓨터 실행 가능 명령을 포함하는 컴퓨터 판독 가능 매체에 있어서, 이때 상기 데이터는 선행 청구항들중 임의의 한 항의 방법에 의해 생성된, 컴퓨터 판독가능한 매체.

청구항 199

청구항 198에 있어서, 다음을 위한 실행가능한 명령을 포함하는, 컴퓨터 판독 가능 매체:

- a. 서열 판독을 UID 패밀리에 할당하기 위한 명령, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- b. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하기 위한 명령;
- c. 뉴클레오타이드 서열은 Watson 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인하기 위한 명령;
- d. 뉴클레오타이드 서열은 Crick 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인하기 위한 명령;
- e. Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하기 위한 명령;
- f. Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하기 위한 명령;
- g. 뉴클레오타이드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오타이드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인하기 위한 명령.

청구항 200

청구항 199에 있어서, 외인성 UID 서열이 R2 시퀀싱 프라이머 결합 부위의 바로 하류에 있거나, 또는 1-300개 뉴클레오타이드 이내에 있을 경우, UID 패밀리의 서열 판독은 Watson 하위-패밀리로 할당하는, 컴퓨터 판독가능한 매체.

청구항 201

전술한 청구항들중 임의의 항에 있어서, 외인성 UID 서열이 R1 시퀀싱 프라이머 결합 부위의 바로 하류에 있거나, 또는 1-300개 뉴클레오타이드 이내에 있을 경우, UID 패밀리의 서열 판독은 Crick 하위-패밀리로 할당하는, 컴퓨터 판독가능한 매체.

청구항 202

전술한 청구항들중 임의의 항에 있어서, 상기 서열 판독을 참조 게놈에 맵핑하는 것을 포함하는, 컴퓨터 판독가능한 매체.

청구항 203

청구항 202에 있어서, 이때 상기 참조 게놈은 인간 참조 게놈인, 컴퓨터 판독가능한 매체.

청구항 204

전술한 청구항들중 임의의 항에 있어서, 해당 샘플에 돌연변이의 존재, 부재 또는 돌연변이의 양을 기반으로 하여 치료 옵션 보고서를 생성하기 위한 컴퓨터 실행가능한 명령을 더 포함하는, 컴퓨터 판독가능한 매체.

청구항 205

전술한 청구항들중 임의의 항에 있어서, 네트워크를 통해 데이터를 전송할 수 있는 컴퓨터 실행가능한 코드를 더 포함하는, 컴퓨터 판독가능한 매체.

청구항 206

다음을 포함하는, 컴퓨터 시스템:

- a. 핵산 샘플로부터 서열 데이터를 제공받도록 설정된 메모리 유닛(memory unit), 이때 상기 데이터는 전술한 청구항들중 임의의 한 항의 방법에 의해 생성되며;
- b. 상기 메모리 유닛에 통신 가능하도록 연결된 프로세서(processor), 이때 상기 프로세서는 전술한 청구항들중 임의의 항의 컴퓨터 판독가능한 매체를 포함한다.

청구항 207

청구항 206에 있어서, 데이터를 메모리 유닛으로 전달하도록 설정된 시퀀싱 시스템을 더 포함하는, 컴퓨터 시스템.

청구항 208

전술한 청구항들중 임의의 항에 있어서, 상기 보고서를 사용자에게 전달하거나 또는 사용자에게 전시하도록 설정된 사용자 인터페이스(user interface)를 더 포함하는, 컴퓨터 시스템.

청구항 209

전술한 청구항들중 임의의 항에 있어서, 네트워크를 통해 데이터 분석의 결과를 전송하도록 설정된 디지털 프로세서를 더 포함하는, 컴퓨터 시스템.

청구항 210

다음을 포함하는, 시스템:

- a. 생물학적 샘플로부터 유래된 이중-가닥으로 된 DNA 단편들의 집단;
- b. 전술한 청구항들중 임의의 항에서 기술된 바와 같은 3' 어댑터 집단;
- c. 전술한 청구항들중 임의의 항에서 기술된 바와 같은 5' 어댑터 집단;
- d. 니크 해독 유사 반응을 실행하기 위한 시약;
- e. 하나 또는 그 이상의 표적 폴리뉴클레오타이드에 대해 앰플리콘을 풍부하게 하기 위한 시약들; 그리고
- f. 시퀀싱 시스템.

청구항 211

청구항 210에 있어서, 전술한 청구항들중 임의의 항의 컴퓨터 시스템을 더 포함하는, 시스템.

청구항 212

다음을 포함하는 키트:

- a. Watson 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 하나 또는 그 이상의 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 (이때 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 각각은 표적-선택적 서열을 포함하고)를 포함하며, 임의선택적으로, 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며;
- b. Crick 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 하나 또는 그 이상의 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머를 포함하고, 임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위이며, 상기 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머 각각은 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하고;
- c. Watson 표적-선택적 프라이머 쌍의 제 2 세트, 이 세트는 (i) 상기 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머를 포함하며, 이때 상기 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하고; 그리고
- d. Crick 표적-선택적 프라이머의 제 2 세트, 이 세트는 (i) 상기 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머를 포함하며, 이때 상기 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오타이드에 선택

적인 표적-선택적 서열을 포함한다.

발명의 설명

기술 분야

[0001] 본 출원은 2020년 2월 4일자로 제출된 U.S. 가특허 출원 일련 번호 62/977,066에 대해 우선권을 주장하며, 이 출원은 이들 전문이 여기에 참고자료로 편입된다.

[0002] **연방 자금 지원에 관한 진술**

[0003] 본 발명은 National Institutes of Health에 의해 제공되는 CA062924, CA152753 및 CA230691에서 정부 지원으로 이루어졌다. 정부는 본 발명의 특정 권리를 보유한다.

[0004] **기술 분야**

[0005] 본 발명은 뉴클레오타이드 시퀀싱 분야에 관계한다. 구체적으로, 본 발명은 돌연변이 식별을 위한 시퀀싱 라이브러리 준비물 및 시퀀싱 작업흐름에 관계한다.

배경 기술

[0006] **배경 정보**

[0007] 회귀 돌연변이의 식별은 기본적인 생물학적 측면, 뿐만 아니라 환자 임상 관리 개선에 유용하다. 사용 분야에는 감염성 질환, 면역 레퍼토리 프로파일링, 고생물유전학(paleogenetics), 법의학, 노화, 비-침습적 산전 검사 및 암이 내포된다. 차세대 시퀀싱 (NGS) 기술은 이 응용 분야에 이론적으로 적합하며, 회귀 돌연변이 검출을 위한 다양한 NGS 접근 방식이 존재한다. 그러나, 기존 NGS 접근 방식의 경우, 시퀀싱 자체의 오류율이 너무 높기 때문에, 돌연변이, 특히 원래 샘플에서 낮은 빈도로 존재하는 돌연변이를 확실하게 검출할 수 없다.

[0008] 원래 템플릿 분자를 태그하기 위한 분자 바코드의 사용은 회귀 돌연변이 검출의 다양한 장애물을 극복하도록 설계되었다. 분자 바코딩으로, 각 태그된 분자의 PCR-생성된 자손의 중복(redundant) 시퀀싱이 수행되고, 시퀀싱 오류는 용이하게 인지된다 (Kinde et al., *Proc Natl Acad Sci U S A* 108:9530-9535 (2011)). 예를 들면, 바코딩화된 템플릿 분자의 자손에 대한 주어진 임계값이 동일한 돌연변이를 함유한다면, 그러면 이 돌연변이는 진정한 것으로 간주된다("슈퍼-돌연변이"). 상기 자손의 주어진 임계값 미만이 관심 돌연변이를 함유한다면, 이 돌연변이는 허상(artifact)으로 간주된다. 분자 바코드는 외인성과 내인성의 두 가지 유형으로 설명되어 왔다. 외인성 바코드 (또한 본원에서 외인성 UIDs로 지칭된)는 사전-명시된 뉴클레오타이드 또는 무작위 뉴클레오타이드를 포함하며, 라이브러리 준비 동안, 또는 PCR 동안 추가된다. 내인성 바코드 (또한 본원에서 내인성 UIDs로 지칭된)는 분석할 템플릿 DNA의 단편(가령, DNA의 무작위 전단(shearing)에 의해 생성된 단편, 또는 무-세포 유체 생물학적 샘플에 존재하는 단편)의 5' 단부 및 3' 단부에 있는 서열에 의해 형성된다. 이러한 바코드는 원래 출발 템플릿으로 앰플리콘을 역-추적하는 데 유용한 것으로 입증되었으며, 분자 계수를 허용하고, 임상-관련 샘플에서 진짜 돌연변이 식별을 개선한다.

[0009] 페어드 엔드 시퀀싱(paired end sequencing)을 위한 갈퀴형 어댑터(forked adapters)는 "듀플렉스 시퀀싱(duplex sequencing)"을 가능하게 하도록 개발되었고, 이때 원래 DNA 듀플렉스의 두 가닥(Watson 및 Crick) 각각은 시퀀싱 시 나타나는 5'에서 3' 방향으로 식별할 수 있다. 듀플렉스 시퀀싱은 라이브러리 준비 또는 시퀀싱 동안 돌연변이가 잘못 생성된 경우, DNA의 두 가닥 모두 동일한 돌연변이를 함유할 가능성이 극히 낮기 때문에, 시퀀싱 오류를 줄인다.

[0010] 그러나, 분자 바코딩의 적용 범위와 임상적 적용가능성을 제한하는 많은 문제가 있다. 예를 들면, 초기 템플릿 분자의 많은 분획을 각 가닥에서 동일한 바코드를 갖는 듀플렉스 분자로 변환시키는 것은 어렵다 (Schmitt et al., *Proc Natl Acad Sci U S A* 109:14508-14513 (2012); Schmitt et al., *Nat Methods* 12:423-425 (2015); 및 Newman et al., *Nat Biotechnol* 34:547-555 (2016)). 이 문제는 액체 생검에 일반적으로 사용되는 무-세포 혈장 DNA에서 발견되는 것과 같이, 초기 DNA의 양(가령, < 33ng)이 제한적일 때, 특히 문제가 된다.

[0011] 표적화된 시퀀싱 라이브러리의 준비에는 일반적으로 시퀀싱 템플릿에 대한 어댑터 부착, 라이브러리 증폭, 그리고 관심대상 표적에 대한 라이브러리를 풍부하게 하기 위한 하이브리드화 캡처가 관련된다. 하이브리드화 캡처

는 큰 관심 영역을 풍부하게 하는 데는 효과적이지만, 작은 표적 영역에 대해 잘 확장되지 않고 (Springer et al., *Elife* 7:doi:10.7554/eLife.32143 (2018)), 그리고 빈약한 듀플렉스 회수를 보인다(Wang et al., *Proc Natl Acad Sci U S A* 112:9704-9709 (2015); 및 Wang et al., *Elife* 5:doi:10.7554/eLife.15175 (2016)). 순차적인 캡처 라운드는 이러한 제한을 부분적으로 극복할 수 있지만, 그러나 이러한 개선에도 불구하고 듀플렉스 회수율은 일반적으로 ~1%이다. CRISPR-DS는 최대 15% 회수를 달성할 수 있지만, 무-세포 DNA에는 적용되지 않는다. 상기 표적화된 영역이 매우 작은 경우 (가령, 혈장에서 질환 모니터링에 필요한 것과 같이, 특정 관심 게놈의 한 개 위치 또는 몇 개의 위치), 또는 사용가능한 DNA의 양은 제한되어 있는 경우 (가령, < 33 ng, 혈장에서 흔히 발견됨), 캡처-기반 접근방식이 차선책이다.

[0012] 따라서, 액체 샘플과 같은 임상적으로 관련된 샘플에서 돌연변이, 가령, 희귀 돌연변이를 정확하게 식별할 수 있도록 하기 위해, 시퀀싱 라이브러리 준비 및 작업흐름에 개선이 필요하다.

발명의 내용

해결하려는 과제

[0013] 요약

[0014] 본 문서는 템플릿의 두 가닥을 동일하게 바코드하는 접근 방식을 제공하고, 하이브리드화 캡처가 필요하지 않은 각 가닥의 PCR-기반 풍부화(enrichment) 방법을 제공함으로써, 이러한 문제를 해결하는 방법 및 재료를 제공한다.

[0015] 본 문서는 이중 가닥으로 된 핵산(가령, DNA)의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이의 존재 감지에 사용할 수 있는 방법 및 재료에 관계한다. 일부 경우에서, 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이를 검출하는 방법에는 라이브러리에서 각 핵산의 각 단부 (가령, 5' 단부와 3' 단부) 상에 듀플렉스 분자 바코드를 갖는 듀플렉스 시퀀싱 라이브러리를 만들고, 듀플렉스 시퀀싱 라이브러리로부터 단일 가닥으로 된 Watson 가닥으로부터-꺼낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥-꺼낸 서열의 라이브러리를 만들고, 그리고 각 단일 가닥으로 된 라이브러리에서 이들 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이를 검출하는 단계가 내포될 수 있다.

[0016] 본원에서 입증된 바와 같이, 이중 가닥으로 된 핵산 템플릿의 Watson 가닥에 대응하는 단일 가닥으로 된 DNA 라이브러리와 이중 가닥으로 된 핵산 템플릿의 Crick 가닥에 대응하는 단일 가닥으로 된 DNA 라이브러리는 듀플렉스 분자 바코드가 통합된 시퀀싱 라이브러리로부터 생성될 수 있고, 각 단일 가닥으로 된 DNA 라이브러리는 가닥-특이적 고정된(anchored) PCR 방식을 이용하여 표적 영역에 대해 풍부화될 수 있고, 이들 표적 영역을 서열화시켜, 핵산의 해당 표적 영역 내 하나 또는 그 이상의 돌연변이의 존재 또는 부재를 검출할 수 있다. 예를 들면, 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이의 존재를 검출하는데 이용될 수 있는 본원에서 기술된 방법 및 재료는 오류없는 시퀀스 탐지용 시퀀싱 시스템 (SaferSeqS)이라고 지칭될 수 있으며, 여기에는 예를 들면, 이중 가닥으로 된 분자 바코드의 제자리(*in situ*) 생성 (가령, 도 22a 참고), 고정된 PCR을 통한 표적 풍부화 (가령, 도 22b 참고), 그리고 템플릿 분자의 제자리 재구축 (가령, 도 22c 참고)으로 라이브러리 구축이 내포될 수 있다. 원래 출발 템플릿에 존재하는 진짜(*Bona fide*)돌연변이는 동일한 초기 핵산 분자의 두 가닥 모두에서 발견되는 변형을 요구함으로써 식별될 수 있다.

[0017] 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이 (가령, 참(true) 체세포 돌연변이)를 검출하는 능력은 합리적인 비용으로, 동시에 다수 돌연변이를 정확하고, 효율적으로 평가할 수 있는 독특한, 미실현 기회를 제공한다. 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이의 존재를 검출하기 위한 본원에 기술된 방법 및 재료 (가령, SaferSeqS 방법)는 DNA 손상 양, 수행할 PCR의 양, 및/또는 DNA 손상 허상의 수를 최소화시키면서, 높은 신뢰도로 희귀 돌연변이 식별 가능하게 한다. "Watson 가닥" 및 "Crick 가닥"이라는 용어는 이중-가닥으로된 출발 핵산 서열에서 두 가닥을 분별하기 위해 단순히 사용된다는 점을 주지해야 한다. 이들 가닥 중 하나가 "Watson" 또는 "Crick"으로 표시될 수 있고, 그러면 또다른 가닥은 다른 이름으로 표시된다.

[0018] 일부 구체예들에서, 다음을 포함하는 방법들이 본원에서 제공된다: a) 다음을 포함하는 반응 혼합물을 만들고: i) 탈-인산화되고, 단부가 블런트화된(blunt-ended) 이중-가닥으로 된 DNA 단편 다수, 이때 상기 이중-가닥으로 된 DNA 단편들 각각은 Watson 가닥과 Crick 가닥을 포함하며; ii) 다수의 어댑터, 이때 이들 각 어댑터는 (5'에서 3'방향으로) A) 바코드, 및 B) 범용 3' 어댑터 서열을 포함하고; 그리고 iii) 리가아제(ligase); 그리고 b)

상기 반응 혼합물을 항온처리하며: i) 어댑터들은 Watson 가닥과 Crick 가닥의 3' 단부에 결합되며, 및 ii) 어댑터들은 Watson 가닥 또는 Crick 가닥의 5' 단부에 결합되지 않고, 이로써 이중-가닥으로 된 결합 산물이 생성된다.

[0019] 특정 구체예들에서, 상기 다수의 어댑터 각각은 독특한 바코드를 포함한다. 추가 구체예들에서, 상기 이중-가닥으로 된 결합 산물 각각은 오직 한 개 바코드를 갖는 Watson 가닥과 오직 한 개 바코드 Crick 가닥을 포함하며, 이때 이 바코드는 Watson 가닥의 바코드와는 상이하다. 추가 구체예들에서, 상기 방법은 c) 상기 이중-가닥으로 된 결합 산물의 적어도 일부분을 시퀀싱하는 단계를 더 포함한다.

[0020] 특정 구체예들에서, 다음을 포함하는 방법들이 본원에서 제공된다: a) 피분석물 DNA 샘플 안에 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs)를 부착시키고, 이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 3'PDSAs의 제 2 가닥 (5'에서 3'방향으로)은 (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 그리고 임의선택적으로, (ii) 3' 차단 그룹을 포함하고; b) 상기 어닐링 부위에 5' 어댑터를 어닐링시키고, 이때 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 및 (ii) 상기 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하며; c) 상기 외인성 UID 서열과 제 1 세그먼트에 걸쳐 5' 어댑터를 연장시키고, 이로써 상기 외인성 UID 서열의 보체와 상기 제 1 세그먼트의 보체가 생성되며, 그리고 d) 상기 제 1 세그먼트의 보체의 3' 단부를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고, 이로써 어댑터-결합된 이중-가닥으로 된 DNA 단편들 다수가 생성된다.

[0021] 일부 구체예들에서, 다음을 포함하는 방법들이 본원에서 제공된다: a) 피분석물 DNA 샘플 안에 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs)를 부착시키고, 이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 3'PDSAs의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 그리고 임의선택적으로, (ii) 3' 차단 그룹을 포함하고; b) 상기 어닐링 부위에 5' 어댑터를 어닐링시키고, 이때 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 및 (ii) 상기 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하며; c) 상기 외인성 UID 서열과 제 1 세그먼트에 걸쳐 5' 어댑터를 연장시키고, 이로써 상기 외인성 UID 서열의 보체와 상기 제 1 세그먼트의 보체가 생성되며, 그리고 d) 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 상기 제 1 세그먼트의 보체의 3' 단부를 공유적으로 연계시키고, 이로써 어댑터-결합된 이중-가닥으로 된 DNA 단편들 다수가 생성된다.

[0022] 일부 구체예들에서, 다음을 포함하는 방법들이 본원에서 제공된다: a) 피분석물 DNA 샘플 안에 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs)를 부착시키고, 이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 3'PDSAs의 제 2 가닥 (5'에서 3'방향으로)은 (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 그리고 임의선택적으로, (ii) 3' 차단 그룹을 포함하고; b) 상기 어닐링 부위에 5' 어댑터를 어닐링시키고, 이때 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 및 (ii) 상기 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하며; c) 상기 외인성 UID 서열과 3'PDSA의 제 1 세그먼트에 걸쳐 5' 어댑터를 연장시키고, 이로써 상기 외인성 UID 서열의 보체와 3'PDSA의 제 1 세그먼트의 보체가 생성되며, 그리고 d) 상기 3'PDSA의 제 1 세그먼트의 보체의 3' 단부를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고, 이로써 어댑터-결합된 이중-가닥으로 된 DNA 단편들 다수가 생성된다.

[0023] 일부 구체예들에서, 다음을 포함하는 시스템, 키트 및 조성물이 본원에서 제공된다: a) 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 결합되도록 설정된 부분적으로 이중-가닥으로 된 3' 어댑터 (3'PDSAs) 집단, 이때 상기 3'PDSAs의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 3'PDSAs의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하며; 그리고 b) 상기 어닐링 부위에 어닐링되도록 설정된 5' 어

댕터 집단, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 및 (ii) 3' 어댑터의 어닐링 부위에 상보적인 서열을 포함한다.

[0024] 추가 구체예들에서, 상기 시스템, 키트 및 조성물은 다음을 더 포함한다: c) 생물학적 샘플로부터 이중-가닥으로 된 DNA 단편들의 집단, 및/또는 c) 단일-가닥으로 된 3' 어댑터 (3'SSAs)가 생성되도록 하기 위해 상기 3'PDSAs의 제 2 가닥을 분해시키기 위한 시약; 및/또는 c) 상기 범용 3' 어댑터 서열에 상보적인 제 1 프라이머, 및 범용 5' 어댑터 서열의 보체에 상보적인 제 2 프라이머; 및/또는 c) 시퀀싱 시스템; 및/또는 c) 상기 범용 3' 어댑터 서열에 상보적인 Watson 고정 프라이머, 그리고 d) 상기 범용 5' 어댑터 서열의 보체에 상보적인 Crick 고정 프라이머 ; 및/또는 c) 제 1 세트의 Watson 표적-선택적 프라이머 쌍, 이 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 1 Watson 표적-선택적 프라이머, 및 (ii) 하나 또는 그 이상의 제 2 Watson 표적 선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 각각은 표적-선택적 서열을 포함하고; 및/또는 c) 제 1 세트의 Crick 표적-선택적 프라이머 쌍, 이 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 하나 또는 그 이상의 Crick 표적-선택적 프라이머, 및 (ii) 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 2 Crick 표적 선택적 프라이머 각각은 상기 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함한다.

[0025] 일부 구체예들에서, 상기 방법은 다음을 더 포함한다: 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들 다수를 상기 범용 3' 어댑터 서열에 상보적인 제 1 프라이머와 범용 5' 어댑터 서열의 보체에 상보적인 제 2 프라이머로 증폭시키고, 이로써 앰플리콘이 생성되며, 이때 상기 앰플리콘은 이중-가닥으로 된 Watson 템플릿 다수와 이중-가닥으로 된 Crick 템플릿 다수를 포함한다. 특정 구체예들에서, 상기 방법은 다음을 더 포함한다: 전술한 이중-가닥으로 된 Watson 템플릿을 Watson 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함한다. 추가 구체예들에서, 상기 방법은 다음을 더 포함한다: 전술한 이중-가닥으로 된 Crick 템플릿을 Crick 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분의 보체에 상보적인 제 1 Crick 표적-선택적 프라이머, 그리고 (ii) 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함한다. 특정 구체예들에서, 임의의 어댑터를 결찰하기-전, 상기 이중-가닥으로 된 DNA 단편들의 집단을 우라실-DNA 글리코실화효소와 DNA 글리코실화효소-리아제 엔도뉴클레아제 VIII의 혼합물과 함께 향온처리한다.

[0026] 일부 구체예들에서, 이용된 중합효소 (가령, 5' 어댑터 서열을 연장시키기 위한)는 5'에서 3'로의 엑소뉴클레아제 활성 (가령, 3'PDSAs의 제 2 가닥을 절단시킬 수 있는)을 갖는다. 다른 구체예들에서, 이용된 중합효소 (가령, 5' 어댑터 서열을 연장시키기 위한)는 5'에서 3'로의 엑소뉴클레아제 활성을 갖지 않는다.

[0027] 다른 구체예들에서, 상기 방법은 상기 단일-가닥으로 된 3' 어댑터 (3'SSAs)가 생성되도록 하기 위해, 3'PDSAs의 제 2 가닥을 제거하는 단계를 더 포함한다. 다른 구체예들에서, 이때 상기 제 2 가닥의 제거 단계는 단계 b) 이후, 또는 단계 b) 이전, 또는 단계 b) 동안 일어날 수 있다. 일부 구체예들에서, 상기 3'PDSAs의 제 2 가닥을 제거하는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시켜, 이 제 2 가닥을 분해시키는 것을 포함한다. 추가 구체예들에서, 제 2 가닥을 제거하는 단계는 엑소뉴클레아제 활성을 갖는 중합효소에 의해 이루어지며, 이때 상기 중합효소는 상기 외인성 UID 서열과 제 1 세그먼트에 걸쳐 5' 어댑터를 연장시킨다.

[0028] 추가 구체예들에서, 상기 방법은 하나 또는 그 이상의 상기 앰플리콘의 서열 판독을 결정하는 단계를 더 포함한다. 다른 구체예들에서, 상기 방법은 이 서열 판독을 UID 패밀리에 할당하는 것을 더 포함하며, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함한다. 특정 구체예들에서, 상기 방법은 R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하는 것을 더 포함한다. 다른 구체예들에서, 상기 방법은 Watson 하위-패밀리의 적어도 50% (가령, 50 ... 75 ... 95%)가 해당 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 뉴클레오티드 서열을 식별해내는 단계를 더 포함한다. 다른 구체예들에서, 상기 방법은 Crick 하위-패밀리의 적어도 50% (가령, 50 ... 75 ... 95%)가 해당 서열을 함유할 때, 피분석물

DNA 단편의 Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열을 식별해내는 단계를 더 포함한다.

[0029]

일부 구체예들에서, 상기 방법은 Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계를 더 포함한다. 추가적인 구체예들에서, 상기 방법은 Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계를 더 포함한다. 다른 구체예들에서, 상기 방법은 뉴클레오타이드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오타이드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인하는 단계를 더 포함한다. 일부 구체예들에서, UID 패밀리와 각 구성원은 동일한 내인성 UID 서열을 더 포함하며, 이때 상기 내인성 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함한다. 다른 구체예들에서, 상기 이중-가닥으로 된 DNA 단편들의 집단은 블런트(blunt) 단부를 갖는다.

[0030]

포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한 방법이 본원에서 제공되며, 이때 상기 방법은 다음을 포함한다: A) 상기 이중 가닥으로 된 DNA 단편의 각 단부에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고; B) 상기 이중 가닥으로 된 DNA 단편의 각 단부에 상기 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 증폭시켜, 증폭된 듀플렉스 시퀀싱 라이브러리를 생성시키고, 이때 이러한 증폭 단계는 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 온전한-계승 PCR 조건 하에서 상기 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 범용 프라이머 쌍을 접촉시키는 것을 포함하고; C) 임의선택적으로, 이러한 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 단일 가닥으로 된 DNA 라이브러리를 생성시키고; D) 임의선택적으로, 이러한 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 단일 가닥으로 된 DNA 라이브러리를 생성시키고; E) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여, Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 해당 표적 영역을 증폭시키고; F) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여, Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 해당 표적 영역을 증폭시키고; G) Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리) (가령, DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리))로부터 증폭된 표적 영역을 시퀀싱하여 시퀀싱 판독을 만들고, 이 표적 영역의 Watson 가닥 상에 돌연변이의 존재 또는 부재를 알아내고; H) Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리) (가령, DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리))로부터 증폭된 표적 영역을 시퀀싱하여 시퀀싱 판독을 만들고, 이 표적 영역의 Crick 가닥 상에 돌연변이의 존재 또는 부재를 알아내고; 그리고 I) 각 시퀀싱 판독에 존재하는 분자 바코드에 따라 시퀀싱 판독을 그룹별로 분류하여, 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 돌연변이가 존재하는 지를 결정한다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함한다: i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 듀플렉스 어댑터를 결합시키고, 이때 3' 듀플렉스 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며; ii) 상기 분해가능한 3' 차단 그룹을 분해시키고; iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키고, 이때 상기 5' 듀플렉스 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 남겨두고; 그리고 iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 갭을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다. 일부 구체예들에서, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리) (가령, DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리))를 생성시키는 단계는 다음을 포함한다: i) 제 1 프라이머와 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 증폭된 듀플렉스 시퀀싱 라이브러리의 제 1 분액(aliquot)을 증폭시키고, 이때 상기 제 1 프라이머는 Watson 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 태그(tag)를 포함하여, 태그된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물이 생성되며; ii) 태그된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변형시켜, 단일 가닥

으로 된 태그된 Watson 가닥과 단일 가닥으로 된 Crick 가닥을 생성시키고; 그리고 iii) 단일 가닥으로 된 태그된 Watson 가닥을 회수시켜, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)를 생성시킨다.

[0031] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 포유류의 샘플로부터 획득되며, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)를 생성시키는 단계는 다음을 포함한다: i) 제 1 프라이머와 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 증폭된 듀플렉스 시퀀싱 라이브러리의 제 1 분액을 증폭시키고, 이때 상기 제 1 프라이머는 Crick 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 태그(tag)를 포함하여, 태그된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물이 생성되며; ii) 태그된 Crick 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변성시켜, 단일 가닥으로 된 태그된 Crick 가닥과 단일 가닥으로 된 Watson 가닥을 생성시키고; 그리고 iii) 단일 가닥으로 된 태그된 Crick 가닥을 회수시켜, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)를 생성시킨다. 일부 구체예들에서, 포유류는 인간이다.

[0032] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 생성시키기 전, 이 방법은 다음을 더 포함한다: 상기 이중 가닥으로 된 DNA를 단편화시켜, 이중 가닥으로 된 DNA 단편들을 만들고; 상기 이중 가닥으로 된 DNA 단편들의 5' 단부들을 탈-인산화시키고; 그리고 이중 가닥으로 된 DNA 단편들의 단부를 블런팅시킨다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 리가아제는 T4 DNA 리가아제다.

[0033] 일부 구체예들에서, 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 리가아제는 *대장균(Escherichia coli)* 리가아제다.

[0034] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에 단일 가닥으로 된 핵산의 갭을 채우는 단계는 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 중합효소는 Taq 중합효소이다.

[0035] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 갭을 채우는 단계는 동시에 수행된다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 증폭시켜 증폭된 듀플렉스 시퀀싱 라이브러리를 생성시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 범용 프라이머 쌍을 접촉시키는 것을 포함한다. 일부 구체예들에서, 이러한 증폭 단계는 온전한-계능 PCR을 포함한다. 일부 구체예들에서, 상기 태그된 프라이머는 바이오티닐화된 프라이머이며, 그리고 이때 상기 바이오티닐화된 프라이머는 바이오티닐화된 단일 가닥으로 된 Watson 가닥과 바이오티닐화된 단일 가닥으로 된 Crick 가닥을 생성시킬 수 있다. 일부 구체예들에서, 상기 변성(denaturing) 단계는 NaOH 변성, 열 변성, 또는 이 둘의 조합을 포함한다.

[0036] 일부 구체예들에서, 회수 단계는 상기 태그된 Watson 가닥에 스트렙타비딘-기능화된 비드를 접촉시키고, 상기 태그된 Crick 가닥에 스트렙타비딘-기능화된 비드를 접촉시키는 것을 포함한다. 일부 구체예들에서, 회수 단계는 변성 비-태그된 Watson 가닥을 변성화시키고, 비-태그된 Watson 가닥을 변성화시키는 것을 더 포함한다. 일부 구체예들에서, 회수 단계는 상기 스트렙타비딘-기능화된 비드로부터 바이오티닐화된 단일 가닥으로 된 Watson 가닥을 방출시키고, 상기 스트렙타비딘-기능화된 비드로부터 바이오티닐화된 단일 가닥으로 된 Crick 가닥을 방출시키는 것들 더 포함한다. 일부 구체예들에서, 상기 태그된 프라이머는 인산화된 프라이머이며, 이 인산화된 프라이머는 인산화된 단일 가닥으로 된 Watson 가닥과 인산화된 단일 가닥으로 된 Crick 가닥을 생성시킬 수 있다. 일부 구체예들에서, 상기 변성 단계는 람다 엑소뉴클레아제 절단(digestion)을 포함한다.

[0037] 일부 구체예들에서, Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의

DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함한다. 일부 구체예들에서, 상기 시퀀싱 단계는 페어드-엔드(paired-end) 시퀀싱을 포함한다.

[0038] 포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한 방법이 또한 본원에서 제공되며, 이때 상기 방법은 다음을 포함한다: A) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고; B) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편으로부터 증폭된 듀플렉스 시퀀싱 라이브러리에서 Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)와 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)를 생성시키고; C) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 단일 가닥으로 된 Watson 가닥으로 부터 표적 영역을 증폭시키고; D) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 단일 가닥으로 된 Crick 가닥으로 부터 표적 영역을 증폭시키고; E) Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 증폭된 이 표적 영역을 시퀀싱하여, 시퀀싱 판독을 만들고, 해당 표적 영역의 Watson 가닥에 돌연변이의 존재 또는 부재를 알아내고; F) Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 증폭된 이 표적 영역을 시퀀싱하여, 시퀀싱 판독을 만들고, 해당 표적 영역의 Crick 가닥에 돌연변이의 존재 또는 부재를 알아내고; 그리고 G) 각 시퀀싱 판독에 존재하는 분자 바코드에 따라 시퀀싱 판독을 그룹별로 분류하여, 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 돌연변이가 존재하는 지를 결정한다.

[0039] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플이며, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함한다: i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 이중 어댑터를 결합시키고, 이때 3' 이중 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며; ii) 상기 분해가능한 3' 차단 그룹을 분해시키고; iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키고, 이때 상기 5' 이중 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 남겨두고; 그리고 iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다.

[0040] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 무-세포 DNA 샘플이며, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편으로부터 증폭된 듀플렉스 시퀀싱 라이브러리에서 Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)와 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)를 생성시키는 단계는 다음을 포함한다: i) 범용 제 1 프라이머와 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키고, 이때 이러한 증폭 단계는 온전한-게놈 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 상기 프라이머 쌍을 접촉시키는 것을 포함하고, 이때 상기 제 1 프라이머 Watson 가닥에 혼성화될 수 있으며, 그리고 이때 상기 제 1 프라이머는 바이오티닐화되어, 바이오티닐화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물이 생성되며; ii) 바이오티닐화된 Watson 가닥이 스트렙타아비딘-기능화된 비드에 결합하는 조건 하에서 바이오티닐화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물에 스트렙타아비딘-기능화된 비드를 접촉시키고; iii) 바이오티닐화된 Watson 가닥을 갖는 이중 가닥으로 된 증폭 산물을 변성시켜, 스트렙타아비딘-기능화된 비드에 결합된 단일 가닥으로 된 바이오티닐화된 Watson 가닥을 남겨두고, 단일 가닥으로 된 Crick 가닥을 방출시키고; iv) 단일 가닥으로 된 Crick 가닥을 수집하고; v) 상기 스트렙타아비딘-기능화된 비드로부터 단일 가닥으로 된 바이오티닐화된 Watson 가닥을 방출시키고; 그리고 vi) 단일 가닥으로 된 바이

오티닐화된 Watson 가닥을 수집한다.

- [0041] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 포유류로부터 취한 샘플에서 획득된다. 일부 구체예들에서, 포유류는 인간이다.
- [0042] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 생성시키기 전, 이 방법은 다음을 더 포함한다: 상기 이중 가닥으로 된 DNA를 단편화시켜, 이중 가닥으로 된 DNA 단편들을 만들고; 상기 이중 가닥으로 된 DNA 단편들의 5' 단부들을 탈-인산화시키고; 그리고 이중 가닥으로 된 DNA 단편들의 단부를 블런팅시킨다.
- [0043] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 리가아제는 T4 DNA 리가아제이다. 일부 구체예들에서, 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 리가아제는 *대장균* (*Escherichia coli*) 리가아제이다.
- [0044] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에 단일 가닥으로 된 핵산의 갭을 채우는 단계는 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 중합효소는 Taq-B 중합효소이다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 갭을 채우는 단계는 동시에 수행된다.
- [0045] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 프라이머 쌍을 접촉시키는 것을 포함한다. 일부 구체예들에서, 증폭 단계는 온전한-게놈 PCR을 포함한다. 일부 구체예들에서, Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함한다. 일부 구체예들에서, 상기 시퀀싱 단계는 페어드-엔드 시퀀싱 또는 싱글-엔드(single-end) 시퀀싱을 포함한다.
- [0046] 포유류의 샘플로부터 획득된 이중 가닥으로 된 DNA 템플릿의 표적 영역에서 돌연변이의 존재 또는 부재를 탐지하고, 그리고 이 돌연변이가 상기 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 존재하는 지를 결정하기 위한 방법이 또한 본원에서 제공되며, 이때 상기 방법은 다음을 포함한다: A) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키고; B) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 범용 프라이머 쌍을 이용하여 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 증폭시키고, 이때 이러한 증폭 단계는 온전한-게놈 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 프라이머 쌍을 접촉시키는 것을 포함하고; C) 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 각각 갖는 증폭된 이중 가닥으로 된 DNA 단편들의 Watson 가닥으로부터 표적 영역을 증폭시키고; D) 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 프라이머 쌍을 이용하여 듀플렉스 분자 바코드를 각각 갖는 증폭된 이중 가닥으로 된 DNA 단편들의 Crick 가닥으로부터 표적 영역을 증폭시키고; E) Watson 가닥으로부터 증폭된 표적 영역을 시퀀싱하여 시퀀싱 판독을 만들고, 표적 영역의 Watson 가닥에 돌연변이의 존재 또는 부재를 알아내고; F) Crick 가닥으로부터 증폭된 표적 영역을 시퀀싱하여 시퀀싱 판독을 만들고, 표적 영역의 Crick 가닥에 돌연변이의 존재 또는 부재를 알아내고; G) 각 시퀀싱 판독에 존재하는 분자 바코드에 따라 시퀀싱 판독을 그룹별로 분류하여, 이중 가닥으로 된 DNA 템플릿의 두 가닥 모두에 돌연변이가

존재하는 지를 결정한다.

- [0047] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플이며, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 각각 갖는 이중 가닥으로 된 DNA 단편들을 생성시키는 단계는 다음을 포함한다: i) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득한 이중 가닥으로 된 DNA 단편의 3' 단부에 3' 이중 어댑터를 결합시키고, 이때 3' 이중 어댑터는 a) 5' 포스페이트, 제 1 분자 바코드, 및 3' 올리고뉴클레오타이드를 포함하는 제 1 올리고뉴클레오타이드, 이에 어닐링된 b) 분해가능한 3' 차단 그룹을 포함하는 제 2 올리고뉴클레오타이드를 포함하고, 이때 3' 올리고뉴클레오타이드와 제 2 올리고뉴클레오타이드 서열은 서로 상보적이며; ii) 상기 분해가능한 3' 차단 그룹을 분해시키고; iii) 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키고, 이때 상기 5' 이중 어댑터는 제 2 분자 바코드를 포함하는 올리고뉴클레오타이드를 포함하며, 이때 상기 제 2 분자 바코드는 상기 제 1 분자 바코드와는 상이하며, 이때 상기 5' 어댑터는 상기 제 1 분자 바코드의 상류인 이중 가닥으로 된 DNA 단편에 결합되고, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 남겨두고; 그리고 iv) 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이에 단일 가닥으로 된 핵산 겹을 채워, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편을 생성시킨다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 무-세포 DNA 샘플이다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿은 게놈 DNA 샘플이다. 일부 구체예들에서, 포유류는 인간이다.
- [0048] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 생성시키기 전, 이 방법은 다음을 더 포함한다: 상기 이중 가닥으로 된 DNA를 단편화시켜, 이중 가닥으로 된 DNA 단편들을 만들고; 상기 이중 가닥으로 된 DNA 단편들의 5' 단부들을 탈-인산화시키고; 그리고 이중 가닥으로 된 DNA 단편들의 단부를 블런팅시킨다.
- [0049] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 3' 단부에 3' 듀플렉스 어댑터를 결합시키는 단계는 리가아제 존재 하에서 3' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 리가아제는 T4 DNA 리가아제이다. 일부 구체예들에서, 상기 분해가능한 3' 차단 그룹을 분해시키는 단계는 3' 듀플렉스 어댑터에 우라실-DNA 글리코실화효소 (UDG)를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편의 각 탈-인산화된 5' 단부에 5' 어댑터를 결합시키는 단계는 리가아제 존재 하에서 5' 듀플렉스 어댑터와 상기 이중 가닥으로 된 DNA 템플릿으로부터 획득된 이중 가닥으로 된 DNA 단편을 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 리가아제는 *대장균* (*Escherichia coli*) 리가아제이다.
- [0050] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 5' 어댑터 사이에 단일 가닥으로 된 핵산의 겹을 채우는 단계는 DNA 중합효소 및 dNTPs 존재 하에서 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터를 접촉시키는 것을 포함한다. 일부 구체예들에서, 상기 DNA 중합효소는 Taq-B 중합효소이다.
- [0051] 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편들의 각 5' 단부에 상기 5' 어댑터를 결합시키는 단계, 그리고 상기 이중 가닥으로 된 DNA 단편의 5' 단부와 상기 5' 어댑터 사이의 겹을 채우는 단계는 동시에 수행된다. 일부 구체예들에서, 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 갖는 이중 가닥으로 된 DNA 단편을 증폭시키는 단계는 PCR 조건 하에서 상기 이중 가닥으로 된 DNA 단편의 각 단부 상에서 듀플렉스 분자 바코드를 포함하는 이중 가닥으로 된 DNA 단편에 프라이머 쌍을 접촉시키는 것을 포함한다. 일부 구체예들에서, 증폭 단계는 온전한-게놈 PCR을 포함한다. 일부 구체예들에서, Watson 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 3' 듀플렉스 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함하며; 그리고 이때 Crick 가닥의 DNA 라이브러리 (가령, 단일 가닥으로 된 DNA 라이브러리)로부터 표적 영역을 증폭시키는 단계는 해당 표적 영역에 혼성화될 수 있는 제 1 프라이머와 5' 어댑터에 혼성화될 수 있는 제 2 프라이머로 구성된 제 2 프라이머 쌍을 이용한 제 2 증폭을 더 포함한다. 일부 구체예들에서, 상기 시퀀싱 단계는 페어드-엔드(paired-end) 시퀀싱을 포함한다.
- [0052] 다음을 포함하는 방법들이 본원에 또한 제공된다:
- [0053] a. 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터를 부착시키고, 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위,

및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며;

[0054] b. 상기 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키고, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고;

[0055] c. 니크 해독(nick translation)-유사 반응을 실행하여, 3' 어댑터의 외인성 UID 서열에 걸쳐 5' 어댑터를 연장시키고 (가령, DNA 중합효소를 이용하여), 이렇게 연장된 5' 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고 (가령, 리가아제를 이용하여);

[0056] d. 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;

[0057] e. 상기 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하고;

[0058] f. 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;

[0059] g. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하고;

[0060] h. 뉴클레오타이드 서열은 Watson 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인되며;

[0061] i. 뉴클레오타이드 서열은 Crick 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인되며;

[0062] j. Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하고;

[0063] k. Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하고; 그리고

[0064] l. 뉴클레오타이드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오타이드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인한다.

[0065] 일부 구체예들에서, UID 패밀리의 각 구성원은 동일한 내인성 UID 서열을 더 포함하며, 이때 상기 내인성 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함한다. 일부 구체예들에서, 상기 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열은 적어도 8개, 10개, 또는 15개 염기를 포함한다. 일부 구체예들에서, 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특하다. 일부 구체예들에서, 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특하지 않다. 일부 구체예들에서, UID 패밀리의 각 구성원은 동일한 내인성 UID 서열과 동일한 외인성 UID 서열을 포함한다. 일부 구체예들에서, 단계 (d)는 11회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 단계 (d)는 7회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 단계 (d)는 5회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 단계 (d)는 적어도 1회 사이클의 PCR 증폭을 포함한다.

[0066] 일부 구체예들에서, 상기 서열 판독을 결정하기 전, 상기 앰플리콘은 하나 또는 그 이상의 표적 폴리뉴클레오타이드에 대해 풍부화된. 일부 구체예들에서, 상기 풍부화 단계는 다음을 포함한다:

[0067] a. 표적 폴리뉴클레오타이드 서열을 포함하는 Watson 가닥의 앰플리콘을 Watson 표적-선택적 프라이머 쌍의 제 1 세트에 임의선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함하며; 그리고

- [0068] b. 동일한 표적 폴리뉴클레오타이드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍의 제 1 세트와 임의선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위이며], 그리고 (ii) 제 2 Watson 표적-선택적 프라이머 서열의 표적-선택적 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함한다.
- [0069] 일부 구체예들에서, 상기 방법은 비-표적 폴리뉴클레오타이드로부터 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 정제해내는 단계를 더 포함한다. 일부 구체예들에서, 상기 방법은 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 고형 서포트에 부착시키는 단계를 더 포함한다. 일부 구체예들에서, 상기 제 1 Watson 표적-선택적 프라이머와 제 1 Crick 표적-선택적 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하고, 그리고 이때 상기 고형 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함한다. 일부 구체예들에서, 상기 제 1 구성원은 바이오틴이며, 상기 제 2 구성원은 스트렙타비딘이다. 일부 구체예들에서, 상기 고형 지지대는 비드, 웰(well), 막, 튜브, 컬럼, 플레이트, 세파로즈, 자석 비드, 또는 칩을 포함한다. 일부 구체예들에서, 상기 방법은 상기 고형 서포트에 부착되지 않은 폴리뉴클레오타이드를 제거하는 단계를 더 포함한다.
- [0070] 일부 구체예들에서, 상기 방법은 다음을 더 포함한다:
- [0071] a. 상기 표적 Watson 증폭 산물을 Watson 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Watson 라이브러리 구성원을 창출시키고, 이때 상기 Watson 표적-선택적 프라이머의 제 2 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 3 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Watson 표적-선택적 프라이머를 포함하고;
- [0072] b. 상기 표적 Crick 증폭 산물을 Crick 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Crick 라이브러리 구성원을 창출시키고, 이때 상기 Crick 표적-선택적 프라이머의 제 2 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 5 Crick 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R1시퀀싱 프라이머 부위이며], 그리고 (ii) (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 상기 제 4 Watson 표적-선택적 프라이머의 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Crick 표적-선택적 프라이머를 포함하고;
- [0073] 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머와 Crick 표적-선택적 프라이머는 샘플 바코드 서열을 더 포함한다. 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머는 시퀀서(sequencer) 상에서 제 1 접목 프라이머에 혼성화될 수 있는 제 1 접목(grafting) 서열을 더 포함하고, 이때 상기 제 3 Crick 표적-선택적 프라이머는 이 시퀀서 상에서 제 2 접목 프라이머에 혼성화될 수 있는 제 2 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 4 Watson 표적-선택적 프라이머는 상기 제 2 접목 서열을 더 포함하고, 이때 상기 제 4 Crick 표적-선택적 프라이머는 상기 제 1 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 1 접목 서열은 P7 서열이며, 이때 상기 제 2 접목 서열은 P5 서열이다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 50%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 70%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 80%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 50%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 70%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 80%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 90%를 나타낸다.
- [0074] 다음을 포함하는 방법들이 또한 본원에서 제공된다:

- [0075] a. 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키는 단계, 이때 상기 어댑터는 외인성 UID를 포함하는 이중-가닥으로 된 일부분과 갈퀴형(forked) 일부분을 포함하고, 이때 갈퀴형 일부분은 (i) R2 시퀀싱 프라이머 부위를 포함하는 단일-가닥으로 된 3' 어댑터 서열 및 (ii) R1 시퀀싱 프라이머 부위를 포함하는 단일-가닥으로 된 5' 어댑터 서열을 포함하며;
- [0076] b. 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;
- [0077] c. 표적 폴리뉴클레오타이드 서열을 포함하는 Watson 가닥의 앰플리콘을 Watson 표적-선택적 프라이머 쌍의 제 1 세트에 임의선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 상기 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며], 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함하며; 그리고
- [0078] d. 동일한 표적 폴리뉴클레오타이드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍의 제 1 세트에 임의선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 상기 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머 [임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위이며], 그리고 (ii) 제 2 Crick 표적-선택적 프라이머 서열의 표적-선택적 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함하고;
- [0079] e. 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물의 서열 판독을 결정하고;
- [0080] f. 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- [0081] g. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하고;
- [0082] h. 뉴클레오타이드 서열은 Watson 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인되며;
- [0083] i. 뉴클레오타이드 서열은 Crick 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인되며; 그리고
- [0084] j. Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열과, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열 모두 동일한 돌연변이를 함유할 때, 피분석물 DNA 단편에서 돌연변이를 확인한다.
- [0085] 일부 구체예들에서, 상기 방법은 비-표적 폴리뉴클레오타이드로부터 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 정제해내는 단계를 더 포함한다. 일부 구체예들에서, 상기 방법은 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 고정 서포트에 부착시키는 단계를 더 포함한다. 일부 구체예들에서, 상기 제 1 Watson 표적-선택적 프라이머와 제 1 Crick 표적-선택적 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하고, 그리고 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함한다. 일부 구체예들에서, 상기 제 1 구성원은 바이오틴이며, 상기 제 2 구성원은 스트렙타비딘이다. 일부 구체예들에서, 상기 고정 지지대는 비드, 웰(well), 막, 튜브, 컬럼, 플레이트, 세파로즈, 자석 비드, 또는 칩을 포함한다. 일부 구체예들에서, 상기 방법은 상기 고정 서포트에 부착되지 않은 폴리뉴클레오타이드를 제거하는 단계를 더 포함한다.
- [0086] 일부 구체예들에서, 상기 방법은 다음을 더 포함한다:
- [0087] a. 상기 표적 Watson 증폭 산물을 Watson 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Watson 라이브러리 구성원을 창출시키고, 이때 상기 Watson 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Watson 표적-선택적 프라이머를 포함하고;
- [0088] b. 상기 표적 Crick 증폭 산물을 Crick 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Crick 라이브러리 구성원을 창출시키고, 이때 상기 Crick 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어

덱터 서열의 R1시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Crick 표적-선택적 프라이머, 그리고 (ii) (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 상기 제 4 Watson 표적-선택적 프라이머의 동일한 표적 폴리뉴클레오타이드에 선택적인 표적-선택적 서열을 포함하는 제 4 Crick 표적-선택적 프라이머를 포함하고;

[0089] 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머와 Crick 표적-선택적 프라이머는 샘플 바코드 서열을 더 포함한다. 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머는 시퀀서(sequencer) 상에서 제 1 접목 프라이머에 혼성화될 수 있는 제 1 접목(grafting) 서열을 더 포함하고, 이때 상기 제 3 Crick 표적-선택적 프라이머는 이 시퀀서 상에서 제 2 접목 프라이머에 혼성화될 수 있는 제 2 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 4 Watson 표적-선택적 프라이머는 상기 제 2 접목 서열을 더 포함하고, 이때 상기 제 4 Crick 표적-선택적 프라이머는 상기 제 1 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 1 접목 서열은 P7 서열이며, 이때 상기 제 2 접목 서열은 P5 서열이다. 일부 구체예들에서, 상기 부착단계는 A-꼬리로 된(tailed) 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 집단에 부착시키는 단계를 포함한다. 일부 구체예들에서, 상기 부착단계는 이 집단에 있는 DNA 단편들의 양 단부에 A-꼬리로 된 어댑터를 부착시키는 단계를 포함한다.

[0090] 일부 구체예들에서, 상기 부착단계는 다음을 포함한다:

[0091] a. 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터를 부착시키고, 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 임의선택적으로, 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며; 그리고

[0092] b. 상기 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키고, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고; 그리고

[0093] c. 니크 해독(nick translation)-유사 반응을 실행하여, 3' 어댑터에 걸쳐 5' 어댑터를 연장시키고 (가령, DNA 중합효소를 이용하여), 이렇게 연장된 5' 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부에 공유적으로 연계시키고 (가령, 리가아제를 이용하여);

[0094] 일부 구체예들에서, 상기 UID 서열은 이 집단으로부터 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열을 포함한다. 일부 구체예들에서, 상기 이중-가닥으로 된 DNA 단편의 단부를 포함하는 내인성 UID 서열은 적어도 8개, 10개, 또는 15개 염기를 포함한다. 일부 구체예들에서, 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 독특하다. 일부 구체예들에서, 상기 외인성 UID 서열은 각 이중-가닥으로 된 DNA 단편에 대해 독특하지 않다. 일부 구체예들에서, UID 패밀리들의 각 구성원은 동일한 내인성 UID 서열과 동일한 외인성 UID 서열을 포함한다.

[0095] 일부 구체예들에서, 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 11회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 7회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 5회를 넘지 않는 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 앰플리콘을 만들기 위해 상기 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시키는 단계는 적어도 1회 사이클의 PCR 증폭을 포함한다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 50%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 70%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 80%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 상기 이중-가닥으로 된 DNA 단편들의 집단에서 표적 폴리뉴클레오타이드의 적어도 90%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 50%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 70%를 나타낸다.

다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 80%를 나타낸다. 일부 구체예들에서, 상기 표적 Watson 라이브러리 구성원들과 상기 표적 Crick 라이브러리 구성원들은 전체 DNA 단편들의 집단의 적어도 90%를 나타낸다.

[0096] 일부 구체예들에서, 상기 서열 판독의 결정으로 템플릿 분자의 양 단부 모두의 서열 결정이 가능하다. 일부 구체예들에서, 템플릿 분자의 양 단부의 결정 단계는 페어드-엔드 시퀀싱을 포함한다. 일부 구체예들에서, 상기 서열 판독의 결정 단계는 이러한 서열 판독을 생성시키기 위한 템플릿의 길이에 걸쳐 단일 판독 시퀀싱을 포함한다. 일부 구체예들에서, 상기 서열 판독의 결정 단계는 대규모 병렬(parallel) 시퀀서를 사용한 시퀀싱을 포함한다. 일부 구체예들에서, 상기 대규모 병렬 시퀀서는 템플릿 폴리뉴클레오타이드의 양 단부로부터 서열 판독을 결정하도록 설정된다. 일부 구체예들에서, 상기 이중-가닥으로 된 DNA 단편들의 집단은 길이가 약 50-600nt 인 하나 또는 그 이상의 단편을 포함한다. 일부 구체예들에서, 상기 이중-가닥으로 된 DNA 단편들의 집단은 길이가 2000nt 미만, 1000nt 미만, 500nt 미만, 400nt 미만, 300nt 미만, 또는 250nt 미만인 하나 또는 그 이상의 단편을 포함한다.

[0097] 일부 구체예들에서, 본원에서 제공된 방법은 초기 증폭-후 그리고 선택적 증폭-전, 상기 앰플리콘의 센스 가닥 및 안티센스 가닥에 대응하는 단일 가닥으로 된 (ss) DNA 라이브러리를 준비하는 단계를 더 포함한다. 일부 구체예들에서, 상기 ss DNA 라이브러리 준비 단계는 다음을 포함한다:

[0098] a. 두 개 프라이머를 이용하여 증폭 반응을 수행하고 [이때 두 개 프라이머중 오로지 한 개 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하며], 이로써 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥을 포함하는 증폭 산물이 창출되며;

[0099] b. 상기 증폭 산물을 고정 서포트에 접착시키고, 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함하며;

[0100] c. 변성 상기 증폭 산물을 변성시켜, 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥으로부터 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥을 분리해내고; 그리고

[0101] d. 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 분리된 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 분리된 가닥을 정제한다.

[0102] 일부 구체예들에서, 상기 친화력 결합 쌍의 제 1 구성원은 바이오틴이며, 상기 친화력 결합 쌍의 제 2 구성원은 스트렙타비딘이다. 일부 구체예들에서, 상기 ss DNA 라이브러리 준비 단계는 다음을 포함한다:

[0103] a. 상기 앰플리콘을 두 가지 증폭 반응으로 분할시키고, 이로써 인산화된 가닥과 비-인산화된 가닥을 포함하는 증폭 산물이 창출되며, 이때 각 증폭 반응은 진행(forward) 프라이머와 역행(reverse) 프라이머를 이용하며, 이때 두 개 프라이머중 오로지 한 개만 인산화되며;

[0104] b. 5' 포스페이트를 갖는 가닥만을 선택적으로 절단시키는 엑소뉴클레아제에 상기 증폭 산물을 접착시킨다.

[0105] 일부 구체예들에서:

[0106] a. 상기 제 1 증폭 반응에서, 상기 진행 프라이머는 인산화되고, 상기 역행 프라이머는 비-인산화되고;

[0107] b. 상기 제 2 증폭 반응에서, 상기 역행 프라이머는 인산화되고, 상기 진행 프라이머는 비-인산화된다.

[0108] 일부 구체예들에서, 상기 엑소뉴클레아제는 람다 엑소뉴클레아제이다. 일부 구체예들에서, 상기 인산화는 5' 부위에서 이루어진다.

[0109] 일부 구체예들에서, 상기 초기 증폭은 다음을 포함한다:

[0110] a. 프라이머 쌍으로 증폭시키고, 이로써 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥을 포함하는 증폭 산물이 창출되며, 이때 이 프라이머 쌍에서 두 개 프라이머중 오로지 한 개 프라이머만 친화력 결합 쌍의 제 1 구성원을 포함하고;

[0111] b. 상기 증폭 산물을 고정 서포트에 접착시키고, 이때 상기 고정 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함하며;

[0112] c. 변성 상기 증폭 산물을 변성시켜, 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 가닥으로부터 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 가닥을 분리해내고; 그리고

- [0113] d. 상기 친화력 결합 쌍의 제 1 구성원을 포함하는 분리된 가닥과 상기 친화력 결합 쌍의 제 1 구성원을 포함하지 않는 분리된 가닥을 정제한다.
- [0114] 일부 구체예들에서, 상기 친화력 결합 쌍의 제 1 구성원은 바이오틴이며, 상기 친화력 결합 쌍의 제 2 구성원은 스트렙타비딘이다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열의 하류이며, R1 서열의 상류일 경우, UID 패밀리의 서열 판독은 Watson 하위-패밀리에 할당된다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열의 하류이며, R2 서열의 상류일 경우, UID 패밀리의 서열 판독은 Crick 하위-패밀리에 할당된다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열에 더 근접하고, R1 서열에 덜 근접할 경우, UID 패밀리의 서열 판독은 Watson 하위-패밀리에 할당된다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열에 더 근접하고, R2 서열에 덜 근접할 경우, UID 패밀리의 서열 판독은 Crick 하위-패밀리에 할당된다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열의 바로 하류에 있거나, 또는 1-300개, 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오티드 이내에 있을 경우, UID 패밀리의 서열 판독은 Watson 하위-패밀리에 할당된다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열의 바로 하류에 있거나, 또는 1-300개, 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오티드 이내에 있을 경우, UID 패밀리의 서열 판독은 Crick 하위-패밀리에 할당된다.
- [0115] 일부 구체예들에서, 상기 이중-가닥으로 된 DNA 단편들의 집단은 생물학적 샘플로부터 유래한다. 일부 구체예들에서, 상기 생물학적 샘플은 대상체로부터 획득된다.
- [0116] 일부 구체예들에서, 상기 대상체는 인간 대상체이다. 일부 구체예들에서, 상기 생물학적 샘플은 유체 샘플이다. 일부 구체예들에서, 상기 유체 샘플은 전혈, 혈장, 혈청, 가래, 소변, 땀, 눈물, 복수, 정액 및 기관지폐포 세척액으로부터 선택된다. 일부 구체예들에서, 상기 유체 샘플은 무-세포 샘플 또는 필수적으로 무-세포 샘플이다. 일부 구체예들에서, 상기 생물학적 샘플은 고체의 생물학적 샘플이다. 일부 구체예들에서, 상기 고체 생물학적 샘플은 중앙 샘플이다.
- [0117] 일부 구체예들에서, 상기 확인된 돌연변이는 0.1% 또는 그 미만의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재한다. 일부 구체예들에서, 상기 확인된 돌연변이는 0.1% ~ 0.00001%의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재한다. 일부 구체예들에서, 상기 확인된 돌연변이는 0.1% ~ 0.01%의 빈도로 상기 이중-가닥으로 된 DNA 단편들의 집단에 존재한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 50%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 70%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 80%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 표적 폴리뉴클레오티드를 포함하는 이중-가닥으로 된 DNA 단편들의 적어도 90%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 적어도 50%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 적어도 70%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 적어도 80%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다. 일부 구체예들에서, 서열 판독의 결정 단계는 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 적어도 90%의 Watson 가닥과 Crick 가닥 모두로부터 서열 판독을 결정하는 것을 포함한다.
- [0118] 일부 구체예들에서, 선행 청구항들중 임의의 하나의 방법에 따라 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 피분석물 DNA 단편의 Watson 가닥과 Crick 가닥 모두에서 돌연변이가 확인이 필요없는, 돌연변이가 확인하는 대체 방법과 비교하였을 때, 적어도 2-배, 4-배, 5-배, 10-배, 20-배, 30-배, 40-배, 50-배, 60-배, 70-배, 80-배, 90-배, 또는 100-배 감소된다. 일부 구체예들에서, 상기 대체 방법은 표준 분자 바코딩 또는 표준 PCR-기반된 분자 바코딩을 포함한다. 일부 구체예들에서, 상기 대체 방법은 다음을 포함한다:
- [0119] a. 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키고, 이때 상기 어댑터는 독특한 외인성 UID를 포함하고;

- [0120] b. 초기(initial) 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만들고;
- [0121] c. 상기 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하고;
- [0122] d. 이 서열 판독을 UID 패밀리에 할당하고, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- [0123] e. 뉴클레오타이드 서열은 UID 패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편을 정확하게 나타내는 것으로 확인되며; 그리고
- [0124] f. 피분석물 DNA 단편을 정확하게 나타내는 확인된 서열이 돌연변이가 결여된 참조 서열과 상이할 경우, 이 피분석물 DNA 단편에서 돌연변이가 확인된다.
- [0125] 일부 구체예들에서, 피분석물 선행 청구항들중 임의의 하나의 방법에 따라 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 단지 1×10^{-2} , 단지 1×10^{-3} , 단지 1×10^{-4} , 단지 1×10^{-5} , 단지 1×10^{-6} , 단지 5×10^{-6} , 또는 단지 1×10^{-7} 이다.
- [0126] 핵산 샘플의 서열 판독 데이터를 분석하기 위한 컴퓨터 실행 가능 명령을 포함하는 컴퓨터 판독 가능 매체가 본원에서 또한 제공되며, 이때 상기 데이터는 선행 청구항들중 임의의 한 항의 방법에 의해 생성된다. 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 다음을 위한 실행가능한 명령을 포함한다:
- [0127] a. 서열 판독을 UID 패밀리에 할당하기 위한 명령, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고;
- [0128] b. R1 판독 서열과 R2 판독 서열에 대해 상기 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하기 위한 명령;
- [0129] c. 뉴클레오타이드 서열은 Watson 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 확인하기 위한 명령;
- [0130] d. 뉴클레오타이드 서열은 Crick 하위-패밀리 구성원들의 임계값 백분율이 서열을 함유할 때, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 확인하기 위한 명령;
- [0131] e. Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하기 위한 명령;
- [0132] f. Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 결여된 참조 서열과 상이할 때, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하기 위한 명령;
- [0133] g. 뉴클레오타이드 서열에서 돌연변이가 Watson 가닥을 정확하게 나타내고, 뉴클레오타이드 서열에서 Crick 가닥을 정확하게 나타내는 돌연변이가 동일한 돌연변이일 때, 피분석물 DNA 단편에서 돌연변이를 확인하기 위한 명령.
- [0134] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 외인성 UID 서열이 R2 시퀀싱 프라이머 결합 부위의 바로 하류에 있거나, 또는 1-300개 뉴클레오타이드 이내에 있을 경우, UID 패밀리의 서열 판독은 Watson 하위-패밀리로 할당하는 실행가능한 명령을 포함한다. 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 상기 외인성 UID 서열이 R1 시퀀싱 프라이머 결합 부위의 바로 하류에 있거나, 또는 1-300개 뉴클레오타이드 이내에 있을 경우, UID 패밀리의 서열 판독은 Crick 하위-패밀리로 할당하는 실행가능한 명령을 포함한다. 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 참조 게놈에 상기 서열 판독을 맵핑하기 위한 실행가능한 명령을 포함한다. 일부 구체예들에서, 상기 참조 게놈은 인간 참조 게놈이다.
- [0135] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 해당 샘플에 돌연변이의 존재, 부재 또는 돌연변이의 양을 기반으로 하여 치료 옵션 보고서를 생성하기 위한 컴퓨터 실행가능한 명령을 더 포함한다. 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 네트워크를 통해 데이터를 전송할 수 있는 컴퓨터 실행가능한 코드를 더 포함한다.
- [0136] 다음을 포함하는 컴퓨터 시스템이 또한 본원에서 제공된다:
- [0137] a. 핵산 샘플로부터 서열 데이터를 제공받도록 설정된 메모리 유닛(memory unit), 이때 상기 데이터는 선행

청구항들중 임의의 한 항의 방법에 의해 생성되며;

- [0138] b. 상기 메모리 유닛에 통신 가능하도록 연결된 프로세서(processor), 이때 상기 프로세서는 전술한 청구항들중 임의의 항의 컴퓨터 판독가능한 매체를 포함한다.
- [0139] 일부 구체예들에서, 상기 컴퓨터 시스템은 데이터를 메모리 유닛으로 전달하도록 설정된 시퀀싱 시스템을 더 포함한다. 일부 구체예들에서, 상기 컴퓨터 시스템은 상기 보고서를 사용자에게 전달하거나 또는 사용자에게 전 시하도록 설정된 사용자 인터페이스(user interface)를 더 포함한다. 일부 구체예들에서, 상기 컴퓨터 시스템은 네트워크를 통해 데이터 분석의 결과를 전송하도록 설정된 디지털 프로세서를 더 포함한다.
- [0140] 다음을 포함하는 시스템이 또한 본원에서 제공된다:
- [0141] a. 생물학적 샘플로부터 유래된 이중-가닥으로 된 DNA 단편들의 집단;
- [0142] b. 전술한 청구항들중 임의의 항에서 기술된 바와 같은 3' 어댑터 집단;
- [0143] c. 전술한 청구항들중 임의의 항에서 기술된 바와 같은 5' 어댑터 집단;
- [0144] d. 니크 해독-유사 반응 (가령, DNA 중합효소, 코헤시브 엔드(cohesive end)-특이적 리가아제, 및 우라실-DNA 글리코실화효소를 이용하여)을 실행하기 위한 시약들;
- [0145] e. 하나 또는 그 이상의 표적 폴리뉴클레오티드에 대해 앰플리콘을 풍부하게 하기 위한 시약들; 그리고
- [0146] f. 시퀀싱 시스템.
- [0147] 일부 구체예들에서, 상기 시스템은 전술한 청구항들중 임의의 항의 컴퓨터 시스템을 더 포함한다.
- [0148] 다음을 포함하는 키트가 또한 본원에서 제공된다:
- [0149] a. Watson 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 하나 또는 그 이상의 상기 범용 3' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 (이때 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 각각은 표적-선택적 서열을 포함하고)를 포함하며, 임의선택적으로, 이때 상기 범용 3' 어댑터 서열의 일부분은 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위이며;
- [0150] b. Crick 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 상기 범용 5' 어댑터 서열의 일부분에 상보적인 서열을 포함하는 하나 또는 그 이상의 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머를 포함하고, 임의선택적으로 이때 상기 범용 5' 어댑터 서열의 일부분은 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위이며, 상기 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머 각각은 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하고;
- [0151] c. Watson 표적-선택적 프라이머 쌍의 제 2 세트, 이 세트는 (i) 상기 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머를 포함하며, 이때 상기 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하고; 그리고
- [0152] d. Crick 표적-선택적 프라이머의 제 2 세트, 이 세트는 (i) 상기 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머를 포함하며, 이때 상기 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함한다.
- [0153] 다른 언급이 없는 한, 본 명세서에서 이용된 모든 기술적 그리고 과학적 용어들은 본 발명이 속하는 분야의 통상적 숙련자에 의해 흔히 이해되는 것과 동일한 의미를 갖는다. 본 명세서에 기술된 것과 유사한 또는 동등한 방법 및 물질이 발명의 실시예에 사용될 수 있지만, 적합한 방법 및 물질은 하기에 기술된다. 본 명세서에 언급된 모든 간행물, 특허 출원, 특허 및 다른 참고문헌들은 그 전문이 참고문헌으로 포함된다. 충돌이 있는 경우, 정의를 비롯하여 본 명세서의 내용이 우선한다. 또한, 재료, 방법, 그리고 실험은 단지 예시이고 제한하는 것으로 의도되지 않는다.
- [0154] 본 발명의 하나 또는 그 이상의 구체예들에 관한 세부사항은 첨부 도면과 하기 상세한 설명에서 설명된다. 본

발명의 그 외 다른 특징, 목적 및 이점들은 하기 상세한 설명과 도면 및 청구범위로부터 명확해 질 것이다.

도면의 간단한 설명

[0155]

도 1은 예시적인 듀플렉스 고정된 PCR 방법의 개략도를 담고 있다. 분자 바코드를 갖는 듀플렉스 어댑터는 블런트화된 단부들을 갖는 핵산 단편들의 단부들에 결합되어, 듀플렉스 시퀀싱 라이브러리가 생성되며, 상기 듀플렉스 시퀀싱 라이브러리는 PCR을 거쳐 증폭된 듀플렉스 시퀀싱 라이브러리가 생성된다. 상기 증폭된 듀플렉스 시퀀싱 라이브러리에서 해당 증폭 산물을 두 분액으로 나누고, 각 분액은 PCR을 거치며, 여기에서 Watson 가닥은 상기 제 1 분액으로부터 증폭되며, Crick 가닥은 상기 제 2 분액으로부터 증폭된다.

도 2는 라이브러리 증폭의 예시적인 제 2 라운드의 개략도를 담고 있는데, 여기에서 도 1에서 제 1 분액으로부터 증폭된 Watson 가닥은 제 1 프라이머는 바이오틀화되고, 제 2 프라이머는 바이오틀화안된, 프라이머 쌍을 이용하여 PCR를 거치면, 단일 가닥으로 된 DNA 라이브러리가 생성되는데, 이 라이브러리를 이용하여 Watson 가닥을 증폭시키고, 이 가닥을 평가할 수 있다.

도 3은 라이브러리 증폭의 예시적인 제 3 라운드의 개략도를 담고 있는데, 여기에서 도 1에서 제 1 분액으로부터 증폭된 Crick 가닥은 제 1 프라이머는 바이오틀화안되고, 제 2 프라이머는 바이오틀화된, 프라이머 쌍을 이용하여 PCR를 거치면, 단일 가닥으로 된 DNA 라이브러리가 생성되는데, 이 라이브러리를 이용하여 Crick 가닥을 증폭시키고, 이 가닥을 평가할 수 있다.

도 4는 예시적인 Watson 증폭의 개략도를 담고 있다.

도 5는 예시적인 Crick 증폭의 개략도를 담고 있다.

도 6은 예시적으로 증폭된 Watson 가닥과 예시적으로 증폭된 Crick 가닥의 개략도를 담고 있다.

도 7은 예시적인 네스티드(nested) Watson 증폭의 개략도를 담고 있다.

도 8은 예시적인 네스티드 Crick 증폭의 개략도를 담고 있다.

도 9는 5' 포스페이트의 예시적인 제거의 개략도를 담고 있다.

도 10은 블런트 단부 증폭 산물을 생성하기 위해, 5' 오버행(overhang)을 갖는 증폭 단편의 3' 단부를 채우는 예시적인 개략도를 담고 있다.

도 11은 3SpC3 스페이서, 분자 바코드를 함유하는 외인성 UID 서열, 그리고 3' 차단 그룹에 혼성화된 3' 올리고뉴클레오타이드 (dT) [이는 우라실-DNA 글리코실화효소 (UDG)에 의해 분해될 수 있음]가 내포된 예시적인 3' 듀플렉스 어댑터의 개략도를 담고 있다.

도 12는 3' 듀플렉스 어댑터를 이용한 예시적인 어댑터 결합의 개략도를 담고 있다. 3' 듀플렉스 어댑터의 5' 포스페이트는 핵산 템플릿의 3' 단부에 결합된다.

도 13은 예시적인 5' 어댑터 결합의 개략도를 담고 있다. 단일 반응에서, 3' 듀플렉스 어댑터의 차단 그룹이 분해되고, 5' 어댑터는 Nick-해독 반응을 통하여 핵산 템플릿의 5' 단부에 결합된다.

도 14는 예시적인 라이브러리 PCR 증폭의 개략도를 담고 있다.

도 15는 예시적인 Watson 증폭의 개략도를 담고 있다.

도 16은 예시적인 네스티드 Watson 증폭의 개략도를 담고 있다.

도 17은 예시적인 Crick 증폭의 개략도를 담고 있다.

도 18은 예시적인 네스티드 Crick 증폭의 개략도를 담고 있다.

도 19는 예시적인 듀플렉스 고정된 PCR에 의해 생성된 최종 증폭 산물의 개략도를 담고 있다.

도 20은 예시적인 듀플렉스 고정된 PCR에 의해 생성된 최종 증폭 산물을 사용하여, 유입(유입) 핵산의 Crick 가닥으로부터 Watson 가닥을 구별하기 위해, 페어드-엔드 시퀀싱을 사용할 수 있는 방법을 보여주는 개략도를 담고 있다.

도 21은 예시적인 듀플렉스 고정된 PCR에 의해 생성된 최종 증폭 산물을 사용하여, 유입(유입) 핵산의 Crick 가닥으로부터 Watson 가닥을 구별하기 위해, 페어드-엔드 시퀀싱을 사용할 수 있는 방법을 보여주는 개략도를 담

고 있다.

도 22는 예시적인 SaferSeqS 방법의 개요에 대한 개략도를 담고 있다. (a) 라이브러리 준비는 DNA 템플릿 분자가 탈인산화되고, 블런트화되는, 단부 복구 (1단계)로 시작된다. 그 다음, 고유 식별자(UID) 서열 (좁거나 또는 넓은 대각선 크로스-해칭)을 함유하는 3' 어댑터는 3' 단편 단부에 결합된다 (2단계). 상기 UID 시퀀스는 5' 어댑터의 확장 및 결합 시, 이중 가닥으로 된 바코드로 변환된다 (3단계). 마지막으로, 각 원래 템플릿 분자의 중복성(redundant) PCR 복사본은 라이브러리 증폭 중에 생성된다(4단계). (b) 표적 풍부화는 가닥 특이적 절반-네스티드(hemi-nested) PCR로 달성된다. 상기 증폭된 라이브러리는 DNA 가닥 중 하나에서 파생된 산물을 선택적으로 증폭하는 Watson- 및 Crick 특이적 반응으로 분할된다(단계 5). 두 번째 네스티드 PCR을 통해 추가적인 표적-적중 특이성 및 샘플 바코드 통합이 달성된다(단계 7). 상기 최종 PCR 산물 (단계 8)은 페어드-엔드 시퀀싱을 거친다 (단계 9). 상기 내인성 바코드는 라이브러리 구축 전, 템플릿 단편의 단부를 나타낸다. (c) 시퀀싱-후, 판독은 Watson 가닥 또는 Crick 가닥으로부터 파생된 것으로 결정된다. 원래 템플릿 분자들의 각 가닥은 동일한 외인성 바코드로 태그가 지정되고, 동일한 내인성 바코드를 갖기 때문에, 동일한 부모계 DNA 듀플렉스의 두 가닥 각각으로부터 파생된 판독은 듀플렉스 패밀리로 함께 그룹화될 수 있다. 이들 가닥의 오른쪽 단부에 있는 상이한 크로스-해치(cross-hatch) 및 스티플 패턴(스티플 패턴)은 서로 상이한 바코드를 나타낸다. 나타난 실시예들에서, 각 듀플렉스 패밀리는 8개의 구성원을 보유하는데, 4개는 Watson 가닥을 나타내고, 4개는 Crick 가닥을 나타낸다. 이 논문에서 설명하는 실제 실험에서, 각 패밀리에는 Watson 가닥에서 최소 2개의 구성원과 Crick 가닥에서 2개 구성원을 함유하며, 이때 실제 수는 시퀀싱 심도(depth)에 따라 달라진다. 진짜 돌연변이 패밀리 내에서 별표로 표시되는 진짜(*Bona fide*) 돌연변이는 DNA 듀플렉스의 양쪽 부모 가닥에 모두 존재하고, 따라서 Watson 패밀리 및 Crick 패밀리에서 모두 발견된다. 대조적으로, 시퀀싱 오류 패밀리의 PCR내에 별표로 표시되는 PCR 또는 시퀀싱 오류는 두 가닥 중 하나로부터의 판독 하위-집합으로 제한된다. Watson 가닥-특이적 (손상된 Watson 가닥 패밀리 안에 별표) 그리고 Crick 가닥-특이적 (손상된 Crick 가닥 패밀리 안에 별표) 허상은 Watson 패밀리 또는 Crick 패밀리의 모든 사본에서 발견되지만, 이 둘 모두에서는 발견되지 않는다.

도 23은 SaferSeqS의 분석 수행능을 보여주는 그래프를 담고 있다. 알려진 돌연변이를 함유하는 암의 DNA가 10%에서 0.001%까지 다양한 비율로 건강한 공여자의 백혈구 DNA와 10%에서부터 0.001%로 비율을 낮추면서 혼합될 때, 예상되는 빈도에 대해 SaferSeqS에 의해 결정된 돌연변이 대립유전자 빈도 (MAF). 관심대상 돌연변이에 대한 특이성을 결정하기 위해, 0% 대조군 샘플도 분석하였다. 실선은 y절편이 0으로 고정된 선형 회귀 모델의 핏을 나타낸다 (기울기 = 0.776, $R^2 > 0.999$, $P = 3.95 \times 10^{-15}$).

도 24. SaferSeqS를 통한 높은 듀플렉스 회수 및 효율적인 표적 풍부화. 33ng의 혼합 cfDNA 샘플을 *TP53*의 3가지 다른 돌연변이(p.L264fs, p.P190L, 또는 p.R342X) 중 하나에 대해 분석했다. cfDNA 샘플 당 3개 라이브러리가 준비되었고, 각 라이브러리는 ~11 ng의 cfDNA를 함유한다. (a) 듀플렉스 패밀리의 중앙값(즉, 동일한 내인성 바코드 및 외인성 바코드를 함유하는 Watson 가닥 및 Crick 가닥 모두)은 원래 템플릿 분자 수의 89% (범위: 65% ~ 102%)였다. (b) 정확한 표적 판독의 중앙값 비율은 80%(범위: 72% ~ 91%)였다. 하부 힌지와 및 상부 힌지는 25번째 및 75번째 백분위수에 해당하고, 휘스커(whiskers)는 사분위수 범위의 1.5배까지 확장된다. 개별 포인트는 쉽게 시각화할 수 있도록 무작위 분산으로 오버레이된다.

도 25는 액체 생검 샘플에서 예시적인 돌연변이의 검출을 보여주는 그래프를 담고 있다. 암 환자의 무-세포 혈장 DNA와 혼합된 건강한 개체들의 33ng의 혈장 무-세포 DNA 분석. 돌연변이의 높은 빈도(~0.5-1%), 낮은 빈도(~0.01-0.1%)의 돌연변이 또는 돌연변이 없음이 생성되도록 혼합물을 만들었다. 혼합된 *TP53 p.R342X* 샘플은 (a) SafeSeqs 또는 (b) SaferSeqS로 분석되었다. 유사하게, 혼합된 *TP53 p.L264fs* 샘플은 (c) SafeSeqs 및 (d) SaferSeqS로 분석되었고, 혼합된 *TP53 p.P190L* 샘플은 (e) SafeSeqs 및 (f) SaferSeqS로 분석되었다. 돌연변이 번호는 SafeSeqS로 관찰된 153개의 고유한 돌연변이 각각을 나타낸다 (표 8에서 특정된). SaferSeqS에 의해 검출된 단일 슈퍼칼리돌연변이체(supercalimutant)는 SafeSeqS에 의해 분석된 게놈 영역 밖에 있었고(표 9), 따라서 표시되지 않는다.

도 26. 가닥-애그노스틱(agnostic), 결합-기반된 분자 바코딩 방법과 비교하였을 때, SaferSeqS에서 오류. 암 환자의 무-세포 혈장 DNA와 혼합된 건강한 개체들의 33ng의 혈장 무-세포 DNA 분석. 돌연변이의 높은 빈도(~0.5-1%), 낮은 빈도(~0.01-0.1%)의 돌연변이 또는 돌연변이 없음이 생성되도록 혼합물을 만들었다. 혼합된 *TP53 p.R342X* 샘플은 SaferSeqS로 분석되었지만, 그러나 (a) 가닥 정보는 가닥-애그노스틱, 결합-기반된 분자 바코딩 방법을 모방하기 위해, 이 분석에서 무시되었고, 또는 (b) 가닥 정보는 돌연변이 소환(calling) 동안 고

려되었다. 유사하게, 혼합된 TP53 p.L264fs 샘플은 (c) 가닥 정보의 고려없이 분석되었고, (d) SaferSeqS로 분석되었다. 혼합된 TP53 p.P190L 샘플은 유사하게 (e) 가닥 정보의 고려없이 분석되었고, (f) SaferSeqS로 분석되었다. 돌연변이 번호는 보충 표 3에 정의되어 있다. 별표는 혼합된 돌연변이를 나타낸다. SaferSeqS에 의해 확인된 단일의 예상치 못한 슈퍼칼리돌연변이체는 (e)에 나타낸다.

도 27. 암 환자들의 혈장 샘플 평가. 0.01% ~ 0.1% 사이의 빈도로 8개의 알려진 돌연변이가 품고 있는 5명의 암 환자로부터 얻은 혈장 무-세포 DNA 샘플을 이전에 설명한 PCR 기반 분자 바코딩 방법("SaferSeqS"이 아닌 "SafeSeqS")과 SaferSeqS로 분석했다. 돌연변이 번호는 표 11에 정의되어 있다. 별표는 예상된 돌연변이를 나타낸다. SaferSeqS에 의해 검출된 단일 예상된 슈퍼칼리돌연변이체(supercalimutant)는 SafeSeqS에 의해 분석된 게놈 영역 밖에 있었고(표 11), 따라서 표시되지 않는다.

도 28. 듀플렉스 회수에 대한 PCR 효율성 및 주기(cycle) 수의 효과 원래 DNA 듀플렉스(y-축)의 두 가닥을 모두 회수할 확률은 라이브러리 증폭 주기 수(x-축)에 대해 플롯된다. 이 도면에서 각 창은 창 상단에 표시된 고정된 PCR 효율을 나타낸다. 가닥-특이적 PCRs에 사용된 라이브러리 증폭 산물의 비율을 나타낸다. 라이브러리 증폭 주기 수는 1에서 11까지 다양했다. PCR 효율은 100%에서부터 50%까지 10%씩 증분 변화되었다. 각 가닥-특이적 PCRs에서 사용하는 라이브러리 증폭 산물의 비율은 50%에서 1.4%까지 다양했다. 확률적 모델링을 실시 예 2에 기재된 바와 같이, 수행하였다.

도 29는 예시적인 암 유발 유전자 돌연변이의 검출을 위한 다중 패널을 보여주는 그래프를 담고 있다. 멀티플렉스 패널 내에서 성공적으로 증폭된 36개 앰플리콘의 회수 및 적용범위(coverage). 가로축은 두 번째 유전자-특이적 프라이머(GSP2)의 3' 단부의 하류 위치를 표시한다. 3' 프라이머 단부로부터 거리가 증가함에 따라 적용범위가 점차적으로 감소하는 것은 유입 DNA 단편화 패턴의 결과다. 특정 앰플리콘 길이를 갖는 판독의 이론적 회수에 관한 세부사항은 실시 예 2에서 논의된다.

도 30. 암에서 흔히 돌연변이되는 드라이버 유전자 영역을 분석하기 위해, 멀티플렉스 패널에 사용된 48개 프라이머 쌍의 수행능. 가닥-특이적 PCRs에 사용된 48개의 SaferSeqS 프라이머 쌍 각각에 대한 적중 판독의 비율(즉, 의도한 표적에 매핑되는 전체 판독중에서 비율). 프라이머는 각 유전자-특이적 PCR에서 등몰 농도로 사용되었다.

도 31. 62개 프라이머 쌍의 수행능. 현재까지 테스트된 62개의 SaferSeqS 프라이머 쌍 각각에 대한 적중 판독의 비율(즉, 의도한 표적에 매핑되는 전체 판독중에서 비율). 62개 쌍 중 50개 (81%)가 50% 이상의 적중율(on-target rate)을 나타낸다. 제시된 결과는 프라이머 디자인에 대한 단일 시도를 반영한다.

도 32는 본 명세서에 기재된 방법에 따라 사용자가 핵산 샘플을 분석할 수 있도록 채택된 예시적인 컴퓨터 시스템을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0156]

본 명세서 및 첨부된 청구범위에서 사용되는, 단수형 부정관사("a", "an") 및 상기("the")는 문맥에서 달리 명확히 언급이 없는 한, 복수형을 포함함을 유념하여야 한다.

[0157]

"뉴클레오타이드" 및 "nt"는 본원에서 호환되며, 핵산을 포함하는 생물학적 분자를 일반적으로 지칭한다. 뉴클레오타이드는 알려진 퓨린 및 피리미딘 염기를 함유하는 모이어티를 가질 수 있다. 뉴클레오타이드는 변형된 다른 헤테로사이클릭 염기를 가질 수 있다. 이러한 변형에는 예를 들어, 메틸화된 퓨린 또는 피리미딘, 아실화된 퓨린 또는 피리미딘, 알킬화된 리보스, 또는 기타 헤테로사이클이 내포된다. "폴리뉴클레오타이드", "핵산", 및 "올리고뉴클레오타이드"는 호환사용될 수 있다. 이들은 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드, 또는 이들의 유사체로써, 임의의 길이의 뉴클레오타이드의 중합체 형태를 지칭할 수 있다. 폴리뉴클레오타이드는 임의의 3-차원 구조를 가질 수 있으며, 알려지거나 또는 알려지지 않은 모든 기능을 수행할 수 있다. 다음은 폴리뉴클레오타이드의 비-제한적인 예시들이다: 유전자 또는 유전자 단편의 코딩 영역 또는 non-코딩 영역, 링키지 분석으로 특정된 좌(좌들), 엑손, 인트론, 메신저 RNA(mRNA), 전달 RNA, 리보솜 RNA, 리보자임, cDNA, 재조합 폴리뉴클레오타이드, 분기된(branched) 폴리뉴클레오타이드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브, 그리고 프라이머. 폴리뉴클레오타이드는 비-천연 발생 서열을 포함할 수 있다. 폴리뉴클레오타이드는 변형된 뉴클레오타이드, 이를 테면, 메틸화된 뉴클레오타이드와 뉴클레오타이드 유사체들을 포함할 수 있다. 존재하는 경우, 뉴클레오타이드 구조에 대한 변형은 중합체의 조립 전 또는 후에 부여될 수 있다. 뉴클레오타이드 서열은 비-뉴클레오타이드 구성요소들에 의해 중단될 수 있다. 올리고뉴클레오타이드는 중합화 후 추가 변형될 수 있는데, 이

를 태면, 라벨링(labeling) 구성요소의 콘주게이션에 의해 변형될 수 있다.

- [0158] "프라이머"는 일반적으로 뉴클레오타이드 서열(가령, 올리고뉴클레오타이드)을 포함하는 폴리뉴클레오타이드 분자이며, 일반적으로 유리 3'-OH 기가 있고, 템플릿 서열 (이를 태면, 표적 폴리뉴클레오타이드, 또는 프라이머 연장 산물)과 혼성화되고, 이 템플릿에 상보적인 폴리뉴클레오타이드의 중합을 촉진할 수 있다.
- [0159] 본원에 사용된 용어 "포유류"에는 인간 및 비-인간 둘 모두가 내포되며, 인간, 비-인간 영장류, 개과, 고양이과, 쥐과, 소, 말 및 돼지가 내포되지만, 그러나 이에 국한되지는 않는다.
- [0160] 개요
- [0161] 이 문서는 핵산 샘플에 존재하는 돌연변이를 정확하게 식별하는 데 유용한 방법 및 재료에 관한 것이다. 일부 측면들에서, 상기 방법은 이중 가닥으로 된 핵산 템플릿의 Watson 가닥과 Crick 가닥 모두에 돌연변이가 존재할 때, 이 돌연변이를 확인하는 단계를 포함한다. 이러한 방법은 예를 들어, DNA 손상, PCR 및 기타 시퀀싱 허상에서 유래한 인공물과 진정한 돌연변이를 구별하는 데 특히 유용하여, 높은 신뢰도로 돌연변이를 식별할 수 있다.
- [0162] 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 오류율은 낮으면서, 하나 또는 그 이상의 돌연변이를 검출할 수 있다. 예를 들면, 본 명세서에 기술된 방법 및 재료는 약 1% 미만의 오류율로, 핵산 템플릿에서 핵산 돌연변이의 존재 또는 부재 확인에 사용될 수 있다 (가령, 약 0.1% 미만, 약 0.05% 미만, 또는 약 0.01% 미만). 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 핵산 템플릿에서 핵산 돌연변이의 존재 또는 부재 확인에 사용될 수 있으며, 오류율은 약 0.001% 내지 약 0.01%이다. 일부 구체예들에서, 본원에 기술된 방법에 따라 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 단지 1×10^{-2} , 단지 1×10^{-3} , 단지 1×10^{-4} , 단지 1×10^{-5} , 단지 1×10^{-6} , 단지 5×10^{-6} , 또는 단지 1×10^{-7} 이다. 일부 구체예들에서, 본원에 기술된 방법에 따라 피분석물 DNA 단편들에서 하나 또는 그 이상의 돌연변이를 확인하는 것과 관련된 오류율은 피분석물 DNA 단편의 Watson 가닥과 Crick 가닥 모두에서 돌연변이 확인이 필요없는, 돌연변이 확인하는 대체 방법과 비교하였을 때, 적어도 2-배, 4-배, 5-배, 10-배, 20-배, 30-배, 40-배, 50-배, 60-배, 70-배, 80-배, 90-배, 또는 100-배 감소된다.
- [0163] 일부 구체예들에서, 상기 대체 방법은 표준 분자 바코딩 또는 표준 PCR-기반된 분자 바코딩하고, 이어서 시퀀싱하는 것을 포함한다. 특정 구체예들에서, 상기 대체 방법은 다음을 포함한다: (a) 피분석물 DNA 샘플에서 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키는 단계, 이때 이 어댑터는 독특한 외인성 UID를 포함하고; (b) 초기 증폭을 실행하여, 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만드는 단계; (c) 상기 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하는 단계; (d) 이 서열 판독을 UID 패밀리에 할당하는 단계, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고; (e) UID 패밀리 구성원의 임계 백분율이 서열을 함유할 때, 뉴클레오타이드는 분석물 DNA 단편을 정확하게 나타내는 것으로 확인하는 단계; 그리고 (f) 분석물 DNA 단편을 정확하게 나타내는 것으로 확인된 서열이 돌연변이가 결여된 참조 서열과 상이한 경우, 이 분석물 DNA 단편에서 돌연변이를 확인하는 단계.
- [0164] 일부 경우에서, 본 명세서에 기술된 방법 및 재료를 이용하여 효과적인 듀플렉스 회수를 할 수 있다. 예를 들면, 본원에 기술된 방법은 이중 가닥으로 된 핵산 템플릿의 Watson 가닥 및 Crick 가닥 모두에서 유래된 PCR 증폭 산물을 회수하는 데 사용될 수 있다. 일부 경우에서, 상기 본원에 기술된 방법을 이용하여 적어도 50% (가령, 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 82%, 약 85%, 약 88%, 약 90%, 약 93%, 약 95%, 약 97%, 약 99%, 또는 100%) 듀플렉스 회수를 이룰 수 있다.
- [0165] 일부 경우에서, 본 명세서에 기술된 방법 및 재료를 이용하여 대립유전자 빈도가 낮은 돌연변이를 확인할 수 있다. 예를 들면, 본 명세서에 기술된 방법을 이용하여 약 1% 미만의 낮은 대립유전자 빈도를 갖는 돌연변이를 확인할 수 있다. (가령, 약 0.1% 미만, 약 0.05% 미만, 또는 약 0.01% 미만). 일부 경우에서, 상기 본원에 기술된 방법은 약 0.001%의 낮은 대립유전자 빈도를 갖는 돌연변이를 검출하는 데 사용될 수 있다.
- [0166] 일부 경우에서, 본원에 기술된 방법은 분석물 핵산 샘플에 존재하는 0.1% 또는 그 미만의 빈도의 돌연변이를 검출하는 데 사용될 수 있다. 일부 구체예들에서, 본원에 기술된 방법은 분석물 핵산 샘플에 존재하는 0.1% ~ 0.00001%의 빈도의 돌연변이를 검출하는 데 사용될 수 있다. 일부 구체예들에서, 본원에 기술된 방법은 분석물 핵산 샘플에 존재하는 0.1% ~ 0.01%의 빈도의 돌연변이를 검출하는 데 사용될 수 있다.

[0167] 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 최소한의 배경 허상 돌연변이가 있는 (또는 전혀 없는) 돌연변이를 감지하는 데 사용할 수 있다. 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 0.01% 미만의 배경 허상 돌연변이가 있는 돌연변이를 감지하는 데 사용할 수 있다. 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 배경 허상 돌연변이가 없는 돌연변이를 감지하는 데 사용할 수 있다.

[0168] 일부 경우에서, 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이를 검출하는 방법에는 라이브러리에서 각 핵산의 각 단부 (가령, 5' 단부와 3' 단부) 상에 듀플렉스 분자 바코드를 갖는 듀플렉스 시퀀싱 라이브러리를 만들고, 이중 시퀀싱 라이브러리로부터 단일 가닥으로 된 Watson 가닥으로부터-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들고, 그리고 각 단일 가닥으로 된 라이브러리에서 이들 이중 가닥으로 된 핵산의 두 가닥 모두에 존재하는 하나 또는 그 이상의 돌연변이를 검출하는 단계가 내포될 수 있다. 3' 듀플렉스 어댑터에 존재하는 제 1 분자 바코드와 5' 어댑터에 존재하는 제 2 분자 바코드를 이용하여 Watson 가닥으로부터 파생된 증폭 산물과 Crick 가닥으로부터 파생된 증폭 산물을 구별할 수 있다.

[0169] 일부 경우에서, 돌연변이를 확인하는 방법은 다음을 포함한다: (a) 부분적으로 이중-가닥으로 된 3' 어댑터를 피분석물 DNA 샘플 안에 이중-가닥으로 된 DNA 단편들의 집단의 Watson 가닥과 Crick 가닥의 양 단부에 부착시키는 단계, 이때 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열포함하고, 그리고 이때 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하고, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며; (b) 상기 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키는 단계, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 상기 범용 3' 어댑터 서열에 상보적이지 않으면서, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 및 (ii) 상기 5' 어댑터에 대한 어닐링 부위에 상보적인 서열을 포함하며; (c) 3' 어댑터의 외인성 UID 서열에 걸쳐 5' 어댑터를 연장시키기 위해 (가령, DNA 중합효소를 이용하여) 니크 해독-유사 반응을 실행하고, 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부 (가령, 리가아제를 이용하여)에 연장된 5' 어댑터를 공유적으로 연계시키는 단계; (d) 초기 증폭을 실행하여 이 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만드는 단계; (e) 하나 또는 그 이상의 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들의 하나 또는 그 이상의 앰플리콘의 서열 판독을 결정하는 단계; (f) 이 서열 판독을 UID 패밀리에 할당하는 단계, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고; (g) R1 판독 서열과 R2 판독 서열에 대해 외인성 UID 서열의 공간적 상관관계에 근거하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하는 단계; (h) Watson 하위-패밀리 구성원의 임계값 백분율이 서열을 포함하는 경우, 분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; (i) Crick 하위-패밀리 구성원의 임계값 백분율이 서열을 포함할 때, 분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; (j) Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이한 경우, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계; (k) Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이한 경우, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계; 그리고 (l) Watson 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이와 Crick 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이가 동일한 돌연변이인 경우, 분석물 DNA 단편의 돌연변이를 확인하는 단계.

[0170] 일부 경우에서, 돌연변이를 확인하는 방법은 다음을 포함한다: (a) 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키는 단계, 이때 이 어댑터는 외인성 UID를 포함하는 이중-가닥으로 된 일부분과 갈퀴형 일부분을 포함하며, 이때 갈퀴형 일부분은 (i) R2 시퀀싱 프라이머 부위를 포함하는 단일-가닥으로 된 3' 어댑터 서열과 (ii) R1 시퀀싱 프라이머 부위를 포함하며; (b) 초기 증폭을 실행하여 어댑터-결찰된, 이중-가닥으로 된 DNA 단편들을 증폭시켜 앰플리콘을 만드는 단계; (c) 표적 폴리뉴클레오타이드 서열을 포함하는 Watson 가닥의 앰플리콘은 Watson 표적-선택적 프라이머 쌍의 제 1 세트를 이용하여 선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되는 단계, 이때 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 및 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함하며; (d) 동일한 표적 폴리뉴클레오타이드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되는 단계, 이때 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 범용 5' 어댑터 서열

의 R1 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 1 Crick 표적-선택적 프라이머, 및 (ii) 상기 제 2 Crick 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함하며; (e) 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물의 서열 판독을 결정하는 단계; (f) 이 서열 판독을 UID 패밀리에 할당하는 단계, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고; (g) R1 판독 서열과 R2 판독 서열에 대해 외인성 UID 서열의 공간적 상관관계에 근거하여, 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리 및 Crick 하위-패밀리로 할당하는 단계; (h) Watson 패밀리 구성원의 임계값 백분율이 서열을 포함하는 경우, 분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; (i) Crick 패밀리 구성원의 임계값 백분율이 서열을 포함할 때, 분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; 그리고 (j) Watson 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이와 Crick 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이 모두가 동일한 돌연변이인 경우, 분석물 DNA 단편의 돌연변이를 확인하는 단계.

[0171] 일부 경우에서, 본 명세서에 기술된 방법 및 재료는 이중 가닥으로 된 핵산의 각 가닥을 독립적으로 평가하는데 사용될 수 있다. 예를 들면, 본원에서 기술된 바와 같이, 핵산 돌연변이가 이중 가닥으로 된 핵산의 평가된 가닥에 독립적으로 확인될 때, 본원에 기술된 방법 및 재료를 이용하여 핵산 돌연변이가 상기 이중 가닥으로 된 핵산의 어느 가닥에서 시작된 것인지를 결정할 수 있다.

[0172] 임의의 적합한 방법을 이용하여 듀플렉스 시퀀싱 라이브러리를 생성시킬 수 있다. 본원에서 사용된 바와 같이, 듀플렉스 시퀀싱 라이브러리는 이 라이브러리에서 각 핵산 단편의 한쪽 단부 (가령, 상기 5' 단부 및/또는 3' 단부) 상에 있는 듀플렉스 분자 바코드가 내포된 다수의 핵산 단편들이며, 이중 가닥으로 된 핵산의 두 가닥 모두의 서열화를 허용할 수 있다. 일부 경우에서, 핵산 샘플은 단편화되어, 핵산 단편이 생성될 수 있고, 생성된 핵산 단편들을 이용하여 듀플렉스 시퀀싱 라이브러리를 만들 수 있다. 듀플렉스 시퀀싱 라이브러리를 생성하기 위해 사용되는 핵산 단편들은 본원에서 유입 핵산으로 또한 지칭될 수 있다. 예를 들면, 듀플렉스 시퀀싱 라이브러리를 생성하는 데 사용되는 핵산 단편이 DNA 단편들인 경우, 이들 DNA 단편들 또한 유입 DNA으로 지칭될 수 있다. 듀플렉스 시퀀싱 라이브러리에는 임의의 적합한 갯수의 핵산 단편들이 내포될 것이다. 일부 경우에서, 듀플렉스 시퀀싱 라이브러리의 생성 단계에는 라이브러리에서 핵산 템플릿을 단편화시키고, 각 핵산 단편의 각 단부에 결합시키는 단계가 내포될 수 있다.

[0173] *피분석물 핵산 샘플*

[0174] 피분석물 핵산 샘플 안에 핵산 템플릿은 임의의 유형의 핵산 (가령, DNA, RNA, 및 DNA/RNA 하이브리드)을 포함할 수 있다. 일부 경우에서, 핵산 템플릿은 이중-가닥으로 된 DNA 템플릿일 수 있다. 상기 본원에 기술된 방법을 위한 템플릿으로 이용될 수 있는 예시적인 핵산에는 게놈 DNA, 순환하는 유리 DNA (cfDNA; 가령, 순환하는 종양 DNA (ctDNA), 및 무-세포 태아 DNA (cffDNA))가 내포되나, 이에 국한되지 않는다.

[0175] 일부 구체예들에서, 상기 핵산 샘플 안에 핵산 템플릿은 핵산 단편들, 가령, DNA 단편들이다. 일부 구체예들에서, DNA 단편의 단부들은 해당 단편의 내인성 독특한 식별자로 이용될 수 있는 독특한 서열을 나타낸다. 일부 구체예들에서, 상기 단편들은 수작업으로 생산된다. 일부 구체예들에서, 상기 단편들은 전단, 가령, 효소적 전단, 화학적 수단에 의한 전단, 음향적 전단, 분무, 원심분리적 전단, 포인트-싱크(point-sink) 전단, 바늘 전단, 초음파처리, 제한 엔도뉴클레아제, 비-특이적 뉴클레아제 (가령, DNase I), 및 이와 유사한 것들에 의해 생성된다. 일부 구체예들에서, 상기 단편들은 수작업으로 만들어지지 않는다. 일부 구체예들에서, 상기 단편들은 cfDNA 샘플에서 온다.

[0176] 일부 구체예들에서, 상기 핵산 샘플에서 핵산 단편은 길이를 갖는다. 이 길이는 약 4-1000개의 뉴클레오타이드일 수 있다. 이 길이는 약 60-300개의 뉴클레오타이드일 수 있다. 이 길이는 약 60-200개의 뉴클레오타이드일 수 있다. 이 길이는 약 140-170개의 뉴클레오타이드일 수 있다. 이 길이는 500 미만의, 400 미만의, 300 미만의, 250 미만의 nt, 또는 200 미만의 nt일 수 있다.

[0177] 일부 구체예들에서, 핵산 템플릿의 단부들은 내인성 UIDs로 이용된다. 이를 테면, 가령, 전반적인 템플릿 길이, 분할(partition) 또는 출발 핵산 샘플에서 핵산 템플릿의 복잡성, 및 이와 유사한 것들과 같은 인자들을 이용하여 핵산 템플릿을 독특하게 식별해내기 위해 필요한 내인성 UID의 길이를 담당자는 결정할 수 있다. 일부 구체예들에서, 핵산 템플릿의 양단부의 10-500개 뉴클레오타이드를 내인성 UIDs로 이용한다. 일부 구체예들에서, 핵산 템플릿의 양단부의 15-100개 뉴클레오타이드를 내인성 UIDs로 이용한다. 일부 구체예들에서, 핵산 템플릿의 양단부의 15-40개 뉴클레오타이드를 내인성 UIDs로 이용한다. 일부 구체예들에서, 핵산 템플릿의 양단부의 적어도 10개 뉴클레오타이드를 내인성 UIDs로 이용한다. 일부 구체예들에서, 핵산 템플릿의 양단부의 적어도 15

개 뉴클레오타이드를 내인성 UIDs로 이용한다. 일부 구체예들에서, 핵산 템플릿의 오직 한 개 단부를 내인성 UID로 이용한다.

[0178] 일부 구체예들에서, 핵산 템플릿은 하나 또는 그 이상의 표적 폴리뉴클레오타이드를 포함한다. 용어 "표적 폴리뉴클레오타이드", "표적 영역", "관심대상의 핵산 템플릿", "원하는 좌(locus)", "원하는 템플릿", 또는 "표적"은 본원에서 호환 사용되며, 연구되는 관심대상의 폴리뉴클레오타이드를 지칭한다. 특정 구체예들에서, 표적 폴리뉴클레오타이드는 관심대상의, 그리고 연구 중인 하나 또는 그 이상의 서열을 함유한다. 표적 폴리뉴클레오타이드는 예를 들면, 게놈 서열을 포함할 수 있다. 상기 표적 폴리뉴클레오타이드는 이의 존재, 그 양, 및/또는 뉴클레오타이드 서열, 또는 이들 서열에서 변화를 결정하고자 하는 표적 서열을 포함할 수 있다.

[0179] 상기 표적 폴리뉴클레오타이드는 질환과 관련된 유전자 영역일 수 있다. 일부 구체예들에서, 이 유전자는 약물화 가능한 표적이다. 용어 "약물화가능한(druggable) 표적"이란, 본원에서 사용된 바와 같이, 질환 치료법에 의해 조절되는 유전자 경로 또는 세포 경로를 일반적으로 지칭한다. 상기 질환은 암일 수 있다. 따라서, 상기 유전자는 종양의 암-관련된 유전자일 수 있다.

[0180] 일부 구체예들에서, 본 명세서에서 핵산 샘플로도 지칭되는 유입 핵산은 생물학적 샘플로부터 얻었다. 상기 생물학적 샘플은 대상체로부터 획득될 수 있다. 일부 구체예들에서, 상기 대상체는 포유류다. 본원에 기재된 방법에서 핵산을 수득하고, 핵산 템플릿으로 사용할 수 있는 포유동물의 예시에는 인간, 비-인간 영장류(가령, 원숭이), 개, 고양이, 양, 토끼, 마우스, 햄스터 및 랫(rat)가 포함되나, 이에 국한되지는 않는다. 일부 구체예들에서, 상기 대상체는 인간 대상체다. 일부 구체예들에서, 상기 대상체는 식물이다.

[0181] 생물학적 샘플에는 혈장, 혈청, 혈액, 조직, 종양 샘플, 대변, 가래, 타액, 소변, 땀, 눈물, 복수, 기관지폐포 세척액, 정액, 고고학 표본 및 법의학 샘플이 내포되지만, 이에 국한되지 않는다. 특정 구체예들에서, 상기 생물학적 샘플은 고체 생물학적 샘플, 가령, 종양 샘플이다. 일부 구체예들에서, 상기 고체 생물학적 샘플이 처리된다. 상기 고체 생물학적 샘플은 포르말린 용액에 고정시킨 후, 파라핀에 매립(가령, FFPE 샘플임)하여 처리할 수 있다. 처리는 대안적으로, 프로브-기반 분석을 수행하기 전, 이 샘플을 동결시키는 것을 포함할 수 있다. 일부 구체예들에서, 상기 샘플은 고정되지도 않고, 또는 동결되지도 않는다. 고정안된, 동결안된 샘플은 핵산의 보존을 위해 구성된 저장 용액에 저장될 수 있다(단지 예로서).

[0182] 일부 구체예들에서, 상기 생물학적 샘플은 액체 생물학적 샘플이다. 액체 생물학적 샘플에는 혈장, 혈청, 혈액, 가래, 타액, 소변, 땀, 눈물, 복수, 기관지폐포 세척액 및 정액이 내포되지만, 이에 국한되지는 않는다. 일부 구체예들에서, 상기 액체 생물학적 샘플은 무-세포, 또는 실질적으로 무-세포 샘플이다. 특정 구체예들에서, 상기 생물학적 샘플은 혈장 샘플 또는 혈청 샘플이다. 일부 구체예들에서, 상기 액체 생물학적 샘플은 전혈(whole blood) 샘플이다. 일부 구체예들에서, 상기 액체 생물학적 샘플은 말초 단핵 혈액 세포를 포함한다.

[0183] 일부 구체예들에서, 핵산 샘플은 상기 생물학적 샘플로부터 분리되었고, 정제되었다. 핵산은 당업계에 공지된 임의의 수단을 사용하여, 생물학적 샘플로부터 분리되고, 정제될 수 있다. 예를 들면, 생물학적 샘플은 세포로부터 핵산을 방출하거나, 또는 해당 생물학적 샘플의 원치 않는 구성 요소들(가령, 단백질, 세포 벽, 기타 오염원)로부터 핵산을 분리하기 위해 처리될 수 있다. 예를 들면, 핵산은 액체 추출(가령, Trizol, DNAzol) 기술을 사용하여, 생물학적 샘플에서 추출할 수 있다. 핵산은 상업적으로 이용 가능한 키트(가령, Qiagen DNeasy 키트, QIAamp 키트, Qiagen Midi 키트, QIAprep 스핀 키트)를 사용하여 또한 추출할 수도 있다.

[0184] 일부 구체예들에서, 상기 생물학적 샘플은 소량의 핵산을 포함한다. 일부 구체예들에서, 상기 생물학적 샘플은 약 500 나노그램 (ng) 미만의 핵산을 포함한다. 예를 들면, 상기 생물학적 샘플은 약 30 ng ~ 약 40 ng의 핵산을 포함한다.

[0185] 핵산은 원심분리를 비롯한, 공지된 방법에 의해 농축될 수 있다(단지 예로서). 핵산은 정제를 목적으로, 선택적 막(가령, 실리카)에 결합될 수 있다. 핵산은 원하는 길이의 단편, 예를 들어 길이가 1000개 미만, 500개 미만, 400개 미만, 300개 미만, 200개 미만 또는 100개 미만의 염기쌍 단편으로 또한 풍부화시킬 수 있다. 크기를 기반으로 한 이러한 풍부화는 예를 들어, PEG-유도된 침전, 전기영동 겔 또는 크로마토그래피 물질(Huber et al. (1993) 핵산 Res. 21:1061-6), 겔 여과 크로마토그래피, TSK 겔(Kato et al. (1984) J. Biochem, 95:83-86)을 사용하여 수행할 수 있고, 이들 공개문헌은 본원의 참고자료에 편입된다.

[0186] 생물학적 샘플로부터 추출된 폴리뉴클레오타이드는 당업계에 공지된 임의의 방법을 사용하여, 선택적으로 침전시키거나 또는 농축될 수 있다.

[0187] 일부 구체예들에서, 상기 핵산 샘플은 약 35 ng 미만의 핵산을 포함한다. 예를 들면, 상기 핵산 샘플에는 약 1

ng ~ 약 35 ng의 핵산 (가령, 약 1 ng ~ 약 30 ng, 약 1 ng ~ 약 25 ng, 약 1 ng ~ 약 20 ng, 약 1 ng ~ 약 15 ng, 약 1 ng ~ 약 10 ng, 약 1 ng ~ 약 5 ng, 약 5 ng ~ 약 35 ng, 약 10 ng ~ 약 35 ng, 약 15 ng ~ 약 35 ng, 약 20 ng ~ 약 35 ng, 약 25 ng ~ 약 35 ng, 약 30 ng ~ 약 35 ng, 약 5 ng ~ 약 30 ng, 약 10 ng ~ 약 25 ng, 약 15 ng ~ 약 20 ng, 약 5 ng ~ 약 10 ng, 약 10 ng ~ 약 15 ng, 약 15 ng ~ 약 20 ng, 약 20 ng ~ 약 25 ng, 또는 약 25 ng ~ 약 30 ng의 핵산)이 내포될 수 있다. 일부 경우에서, 핵산 샘플에는 약 수백개 이상의 뉴클레오타이드 핵산이 내포된 게놈의 핵산이 내포될 수 있다.

[0188] 일부 경우에서, 핵산 샘플은 본질적으로 오염이 없을 수 있다. 예를 들면, 핵산 샘플이 cfDNA 템플릿인 경우, cfDNA는 본질적으로 게놈 DNA 오염이 없을 수 있다. 일부 경우에서, 기본적으로 게놈 DNA 오염이 없는 cfDNA 샘플에는 최소한의 고분자량(가령, > 1000bp) DNA가 내포될 수 있거나, 또는 전혀 없을 수 있다. 일부 경우에서, 본원에 기술된 방법에는 핵산 샘플에 본질적으로 오염이 없는지 여부를 결정하는 단계가 내포될 수 있다. 핵산 샘플에 본질적으로 오염이 없는지 여부를 결정하기 위한 임의의 적절한 방법을 사용할 수 있다. 핵산 샘플에 본질적으로 오염이 없는지 여부를 결정하는 데 사용할 수 있는 방법의 예시로는 예를 들면, TapeStation 시스템 및 Bioanalyzer가 내포된다. 예를 들면, TapeStation 시스템 및/또는 Bioanalyzer를 사용하여 cfDNA 샘플에 기본적으로 게놈 DNA 오염이 없는지 확인하는 경우, ~180bp (가령, 모노뉴클레오타이드 DNA에 상응하는)의 주요 피크는 핵산 샘플에 기본적으로 게놈 DNA 오염이 없음을 나타내는 데 사용할 수 있다.

[0189] 일부 경우에서, 듀플렉스 시퀀싱 라이브러리(가령, 3' 듀플렉스 어댑터를 이들 핵산 단편의 3' 단부에 부착시키기 전)를 생성하는데 이용될 수 있는 핵산 단편들은 단부-복구될 수 있다(end-repaired). 임의의 적합한 방법을 이용하여 핵산 템플릿을 단부-복구시킬 수 있다. 예를 들면, 블런트화 반응 (가령, 블런트 단부 결찰) 및/또는 탈인산화 반응을 이용하여 핵산 템플릿을 단부-복구시킬 수 있다. 일부 경우에서, 블런트화(blunting)에는 단일 가닥으로 된 영역을 채우는 단계가 내포될 수 있다. 일부 경우에서, 블런트화에는 단일 가닥으로 된 영역을 분해시키는 단계가 내포될 수 있다. 일부 경우에서, 블런트화 및 탈인산화 반응을 이용하여 핵산 템플릿을 단부-복구시킬 수 있다 (도 9 및/또는 도 10에 나타냄).

[0190] 어댑터(Adapters)

[0191] 일부 구체예들에서, 상기 방법은 어댑터-부착된, 이중-가닥으로 된 DNA 단편들의 집단을 만들기 위해, 이중-가닥으로 된 DNA 단편들의 집단에 어댑터를 부착시키는 단계를 포함한다.

[0192] 일부 구체예들에서, 상기 어댑터는 외인성 UID를 포함하는 이중-가닥으로 된 일부분과 갈퀴형 일부분을 포함하며, 이 갈퀴형 일부분은 (i) 단일-가닥으로 된 3' 어댑터 서열, 그리고 (ii) 단일-가닥으로 된 5' 어댑터 서열을 포함한다. 일부 구체예들에서, 단일-가닥으로 된 3' 어댑터 서열은 단일-가닥으로 된 5' 어댑터 서열에 상보적이지 않다. 일부 구체예들에서, 상기 3' 어댑터 서열은 제 2 (가령, R2) 시퀀싱 프라이머 부위를 포함하고, 상기 5' 어댑터 서열은 제 1 (가령, R1) 시퀀싱 프라이머 부위를 포함한다. "R1" 시퀀싱 프라이머 부위와 "R2" 시퀀싱 프라이머 부위는 페어드 엔드 관독을 만들어내는 시퀀싱 시스템에 의해 이용되는데, 예를 들어, 시퀀싱될 DNA 단편들의 반대 단부로부터의 관독을 생성하는 시퀀싱 시스템에 의해 사용되는 것으로 이해되어야 한다. 일부 구체예들에서, 상기 R1 시퀀싱 프라이머를 이용하여 DNA 단편들의 제 1 단부로부터의 제 1 관독 집단을 생산하고, 상기 R2 시퀀싱 프라이머를 이용하여 이들 DNA 단편의 반대 단부로부터의 제 2 관독 집단을 생산한다. 상기 제 1 집단을 본원에서 "R1" 또는 "Read 1" 관독이라고 지칭한다. 상기 제 2 집단을 본원에서 "R2" 또는 "Read 2" 관독이라고 지칭한다. 상기 R1 관독과 R2 관독은 이중-가닥으로 된 피분석물 DNA 단편의 각 가닥에 상응하는 "관독 쌍(read pairs)" 또는 "메이트 쌍(mate pairs)"으로 정렬될 수 있다.

[0193] 특정 시퀀싱 시스템, 가령, Illumina는 "R1" 프라이머 및 "R2" 프라이머, 그리고 "R1" 관독 및 "R2" 관독을 이용한다. 본 출원의 목적상 "R1" 및 "R2", 그리고 "Read 1" 및 "Read 2"라는 용어는 특정 시퀀싱 플랫폼과 관련하여 언급되는 방법에 제한되지 않는다는 점에 유의해야 한다. 예를 들면, Illumina 시퀀서를 사용하는 경우, 본원에 기술된 "R2" 프라이머 및 대응하는 R2 관독은 Illumina "R2" 프라이머 및 관독을 지칭할 수 있거나, 또는 본원에 기술된 공개된 "R1" 프라이머 및 해당 R1 리드는 다른 Illumina 프라이머 및 관독을 지칭하는 한, "R1" 프라이머 및 관독을 지칭할 수 있다. 명료하게 하기 위해, 본원에서 제공된 "R2" 프라이머가 "R1" 관독을 생성시키는 Illumina "R1" 프라이머인 일부 구체예들에서, 본원에서 제공된 대응하는 "R1" 프라이머는 "R2" 관독을 생성시키는 Illumina "R2" 프라이머이다. 명료하게 하기 위해, 본원에서 제공된 "R2" 프라이머가 "R2" 관독을 제공하는 Illumina "R2" 프라이머인 일부 구체예들에서, 본원에서 제공된 "R1" 프라이머는 "R1" 관독을 제공하는 Illumina "R1" 프라이머이다.

[0194] 일부 구체예들에서, 상기 외인성 UID는 핵산 샘플에서 각 이중-가닥으로 된 DNA 단편에 독특하다. 일부 구체

예들에서, 상기 외인성 UID는 각 이중-가닥으로 된 DNA 단편에 독특하지 않다.

- [0195] 일부 구체예들에서, 상기 외인성 UID는 길이를 갖는다. 이 길이는 약 2-4000개의 nt일 수 있다. 이 길이는 약 6-100개의 nt일 수 있다. 이 길이는 약 8-50개의 nt일 수 있다. 이 길이는 약 10-20개의 nt일 수 있다. 이 길이는 약 12-14개의 nt일 수 있다. 일부 구체예들에서, 상기 외인성 UID의 길이는 해당 분자들을 독특하게 바코드화시키는데 충분하며, 상기 외인성 UID의 길이/서열은 하류 증폭 단계들을 간섭하지 않는다.
- [0196] 일부 구체예들에서, 상기 외인성 UID 서열은 이 핵산 템플릿에 존재하지 않는다. 일부 구체예들에서, 상기 외인성 UID 서열은 원하는 좌를 품고 있는 원하는 테블릿 안에 존재하지 않는다. 이러한 독특한 서열은 예를 들어, 컴퓨터 판독가능한 매체에 의해 무작위로 생성될 수 있고, 예를 들어, EMBL, GenBank 또는 DDBJ와 같은 공지된 뉴클레오타이드 데이터베이스에 대응하여 BLASTing함으로써, 선별될 수 있다. 일부 구체예들에서, 외인성 UID 서열은 핵산 템플릿 안에 존재한다. 이러한 경우, 이 서열 판독에서 외인성 UID 서열의 위치를 이용하여 해당 핵산 템플릿 안에 서열로부터 외인성 UID 서열을 분별해낸다
- [0197] 일부 구체예들에서, 상기 외인성 UID 서열은 무작위 서열이다. 일부 구체예들에서, 상기 외인성 UID 서열은 무작위 N-mer 서열이다. 예를 들면, 상기 외인성 UID 서열의 길이가 6개의 nt라면, 그러면, 이 서열은 무작위 헥사머(hexamer)일 수 있다. 상기 외인성 UID 서열의 길이가 12개의 nt라면, 그러면 이 서열은 무작위 12-mer일 수 있다.
- [0198] 외인성 UIDs는 뉴클레오타이드를 무작위로 추가하여, 식별자로 사용될 수 있는 길이를 갖는 서열을 형성시킴으로써, 만들 수 있다. 각 추가 위치에서, 4개의 데옥시리보뉴클레오타이드 중 하나를 선별할 수 있다. 대안적으로, 3개, 2개, 또는 1개의 데옥시리보뉴클레오타이드 중 하나로부터의 선별이 사용될 수 있다. 따라서, UID는 특정 위치에서 전적으로 무작위적이거나, 어느 정도 무작위적이거나, 전혀-무작위적이지 않을 수 있다.
- [0199] 일부 구체예들에서, 상기 외인성 UIDs는 무작위 N-mer가 아니지만, 그러나 사전-결정된 외인성 UID 서열 세트로부터 선별된다.
- [0200] 본원에 개시된 방법에 사용하기에 적합한 예시적인 외인성 UIDs는 PCT/US2012/033207에 기재되어 있으며, 이는 그 전문이 본원의 참고자료에 편입된다.
- [0201] 본원에 기술된 갈퀴형 어댑터는 당업계에 공지된 임의의 수단에 의해 이중-가닥으로 된 DNA 단편들에 부착될 수 있다.
- [0202] 일부 구체예들에서, 상기 갈퀴형 어댑터는 다음의 단계들에 의해 이중-가닥으로 된 DNA 단편들에 부착된다: (a) 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 부분적으로 이중-가닥으로 된 3' 어댑터를 부착시키는 단계, 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 1 가닥은 (5'-3'방향으로) (i) 제 1 세그먼트, (ii) 외인성 UID 서열, (iii) 5' 어댑터용 어닐링 부위, 및 (iv) R2 시퀀싱 프라이머 부위를 포함하는 범용 3' 어댑터 서열을 포함하고, 그리고 이때 상기 부분적으로 이중-가닥으로 된 3' 어댑터의 제 2 가닥은 (5'에서 3'방향으로) (i) 상기 제 1 세그먼트에 상보적인 세그먼트, 및 (ii) 3' 차단 그룹을 포함하며, 임의선택적으로 이때 상기 제 2 가닥은 분해가능하며; (b) 어닐링 부위를 통하여 3' 어댑터에 5' 어댑터를 어닐링시키는 단계, 이때 상기 5' 어댑터는 (5'에서 3'방향으로) (i) 범용 3' 어댑터 서열에 상보적이지 않으며, R1 시퀀싱 프라이머 부위를 포함하는 범용 5' 어댑터 서열, 그리고 (ii) 상기 5' 어댑터용 어닐링 부위에 상보적인 서열을 포함하고; 그리고 (c) Nick 해독-유사 반응을 실행하여, 3' 어댑터의 외인성 UID 서열에 걸쳐 5' 어댑터를 연장시키고 (가령, DNA 중합효소를 이용하여), 그리고 이렇게 연장된 5' 어댑터를 상기 이중-가닥으로 된 DNA 단편들의 Watson 가닥과 Crick 가닥의 5' 단부 (가령, 리가아제를 이용하여)에 공유적으로 연계시키는 단계.
- [0203] 일부 구체예들에서, 상기 갈퀴형 어댑터는 다음의 단계에 의해 이중-가닥으로 된 DNA 단편들에 부착된다: (a) 이중-가닥으로 된 DNA 단편들의 집단에서 Watson 가닥과 Crick 가닥 모두의 3' 단부에 3' 듀플렉스 어댑터를 부착시키는 단계.본원에서 기술된 바와 같이, 3' 듀플렉스 어댑터 (부분적으로 이중 가닥으로 된 3' 어댑터로도 또한 지칭되는)는 제 2 올리고뉴클레오타이드 (또한 본원에서 "제 2 가닥"으로도 지칭됨)에 어닐링된(혼성화된) 제 1 올리고뉴클레오타이드 (또한 본원에서 "제 1 가닥"으로도 지칭됨)를 보유할 수 있는 분자 바코드가 내포된 올리고뉴클레오타이드 복합체이며, 따라서 상기 3' 듀플렉스 어댑터의 일부분 (가령, 제 1 일부분)은 이중 가닥이며, 상기 3' 듀플렉스 어댑터의 일부분 (가령, 제 2 일부분)은 단일 가닥이다. 일부 경우에서, 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드는 이 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드에 존재하는 뉴클레오타이드에 서로 상보적인 제 2 올리고뉴클레오타이드를 포함하는 제 1 세그먼트를 포함한다 (가령, 상

기 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드와 상기 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드는 상보성 영역에서 어닐링될 수 있다). 3' 듀플렉스 어댑터의 예시적인 구조는 도 11에 나타난 것과 같을 수 있다.

- [0204] 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드는 5' 포스페이트와 분자 바코드가 내포된 올리고뉴클레오타이드일 수 있다. 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에는 임의의 적합한 갯수의 뉴클레오타이드가 내포될 수 있다. 임의의 적합한 분자 바코드는 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에 내포될 수 있다. 일부 경우에서, 분자 바코드에는 무작위 서열이 내포될 수 있다. 일부 경우에서, 분자 바코드에는 고정된(fixed) 서열이 내포될 수 있다. 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에 내포될 수 있는 예시적인 분자 바코드에는 통합된 DNA 기술에서 이용가능한 IDT 8, IDT 10, ILMN 8, ILMN 10이 내포되지만, 이에 국한되지 않는다. 임의의 적합한 유형의 분자 바코드가 이용될 수 있다. 일부 경우에서, 분자 바코드는 외인성 UID 서열을 포함한다. 외인성 UIDs가 본원에서 기술된다. 5' 포스페이트 및 분자 바코드가 내포되며, 그리고 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에 내포될 수 있는 올리고뉴클레오타이드의 예시에는 ATAAACGACGGCNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAG*T*C (이때 별표는 포스포로티오에이트 결합을 나타냄; 서열 식별 번호: 1)가 내포되며, 여기에서 NNNNNNNNNNNNN (서열 식별 번호: 2)는 분자 바코드이며, 이 분자 바코드에서 뉴클레오타이드의 수는 0 ~ 약 25개 일 수 있다.
- [0205] 일부 구체예들에서, 상기 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드는 5' 어댑터용 어닐링 부위를 포함한다.
- [0206] 일부 구체예들에서, 상기 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드는 범용 3' 어댑터 서열을 포함한다. 일부 구체예들에서, 상기 범용 3' 어댑터 서열은 R2 시퀀싱 프라이머 부위를 포함한다.
- [0207] 일부 경우에서, 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에는 PCR 동안 연장을 방지하거나, 또는 감소시키기 위한, 하나 또는 그 이상의 특성이 또한 내포될 수 있다. PCR 동안 확장을 방지하거나 또는 감소시키는 특성은 임의의 유형의 특성 (가령, 화학적 변형)일 수 있다. PCR 동안 연장을 방지하거나, 또는 감소시킬 수 있고, 본원에 기재된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에 포함될 수 있는 특성의 예시에는 3SpC3 및 3Phos가 내포되나, 이에 국한되지는 않는다. PCR 동안 연장을 방지하거나, 또는 감소시킬 수 있는 특성은 본원에서 기술된 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드에서 임의의 적합한 위치의 올리고뉴클레오타이드로 통합될 수 있다. 일부 경우에서, PCR 동안 확장을 방지하거나, 또는 감소시킬 수 있는 분자가 올리고뉴클레오타이드 내에 내부적으로 통합될 수 있다. 일부 경우에서, PCR 동안 연장을 방지하거나, 또는 감소시키기 위한 분자는 이 올리고뉴클레오타이드에서 단부(가령, 상기 5' 단부)에 통합될 수 있다.
- [0208] 특정 구체예들에서, 상기 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드는 5' 포스페이트, 이 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드에 존재하는 뉴클레오타이드에 서로 상보적인 뉴클레오타이드를 포함하는 제 1 세그먼트, 외인성 UID 서열, 5' 어댑터용 어닐링 부위, 그리고 범용 3' 어댑터 서열을 포함한다.
- [0209] 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드는 차단된 3' 그룹 (가령, 두 어댑터의 이량체화를 감소시키거나, 또는 제거하기 위한)이 내포된 올리고뉴클레오타이드일 수 있다. 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드에는 임의의 적합한 갯수의 뉴클레오타이드가 내포될 수 있다. 일부 구체예들에서, 상기 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드는 상기 3' 듀플렉스 어댑터의 제 1 올리고뉴클레오타이드의 제 1 세그먼트에 상보적이다. 차단된 3' 그룹이 내포되며, 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드가 내포될 수 있는 예시적인 올리고뉴클레오타이드에는 GCCGUCGUUUAdT (서열 식별 번호: 3)이 내포되나, 이에 국한되지 않는다.
- [0210] 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드는 분해가능한 것일 수 있다. 임의의 적합한 방법을 이용하여 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드를 분해할 수 있다. 예를 들면, UDGr를 이용하여 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오타이드를 분해할 수 있다.
- [0211] 일부 경우에서, 본원에서 기술된 3' 듀플렉스 어댑터에는 서열 GCCGUCGUUUAdT (서열 식별 번호: 3)이 내포된 제 2 올리고뉴클레오타이드에 어닐링된 서열 ATAAACGACGGCNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAG*T*C/3SpC3 (서열 식별 번호: 1)이 내포될 수 있다.
- [0212] 일부 경우에서, 본원에서 기술된 3' 듀플렉스 어댑터에는 상업적으로 이용가능한 어댑터가 내포될 수 있다. 본

원에서 기술된 3' 듀플렉스 어댑터로 이용될 수 있는 (또는 이를 만드는데 이용될 수 있는) 예시적인 상업적으로 이용가능한 어댑터에는 Accel-NGS 2S DNA 라이브러리 키트 (Swift Biosciences, cat. # 21024)에 있는 어댑터가 내포되나, 이에 국한되지 않는다. 일부 경우에서, 본원에서 기술된 3' 듀플렉스 어댑터는 실시예 1에서 기술된 것과 같은 것일 수 있다.

[0213] 상기 3' 어댑터는 임의의 적합한 방법을 이용하여 상기 이중-가닥으로 된 DNA 단편들의 3' 단부에 부착(가령, 공유적으로 부착)될 수 있다. 일부 구체예들에서, 상기 3' 어댑터는 절찰에 의해 부착된다. 일부 구체예들에서, 상기 절찰 단계는 리가아제 사용을 포함한다. 3' 어댑터를 각 핵산 단편의 3' 단부에 부착시키기 위해 이용될 수 있는 리가아제의 예시에는 T4 DNA 리가아제, *E. coli* 리가아제 (가령, Enzyme Y3), CircLigase I, CircLigase II, Taq-리가아제, T3 리가아제, T7 리가아제, 및 9N 리가아제가 내포되나, 이에 국한되지 않는다.

[0214] 상기 3' 듀플렉스 어댑터가 각 핵산 단편의 3' 단부에 부착(가령, 공유적으로 부착)된 후, 본원에서 기술된 3' 듀플렉스 어댑터의 제 2 올리고뉴클레오티드는 분해될 수 있고, 5' 어댑터는 상기 각 핵산 단편의 5' 단부에 부착(가령, 공유적으로 부착)될 수 있다. 일부 구체예들에서, 상기 5' 어댑터 서열은 상기 3' 어댑터의 제 1 올리고뉴클레오티드에 상보적이지 않다. 일부 구체예들에서, 상기 5' 어댑터 서열은 5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 상기 3' 어댑터의 어닐링 부위를 포함한다.

[0215] 일부 구체예들에서, 상기 5' 어댑터의 부착은 단계는 상기 어닐링 부위를 통하여 상기 3' 어댑터에 상기 5' 어댑터를 어닐링시키는 단계를 포함한다.

[0216] 5' 어댑터는 3' 듀플렉스 어댑터의 분자 바코드의 상류 핵산 단편에 어닐링될 수 있고, 상기 3' 듀플렉스 어댑터의 일부분 (가령, 분자 바코드)을 함유하는 갭(가령, 단일 가닥으로 된 핵산 단편)이 해당 핵산 단편에 존재하게 된다. 상기 3' 듀플렉스 어댑터의 일부분을 함유하는 갭이 채워질 수 있다 (가령, 이중 가닥으로 된 핵산 단편을 만들기 위해). 임의의 적합한 방법을 이용하여 단일 가닥으로 된 갭을 채울 수 있다. 핵산 단편 상에 단일 가닥으로 된 갭을 채우는데 이용될 수 있는 방법의 예시에는 중합효소, 이를 테면, DNA 중합효소 (가령, Taq 중합효소, 이를 테면 Taq-B 중합효소) 및 Nick-해독 반응 (가령, 리가아제, 이를 테면, *E. coli* 리가아제와 중합효소, 이를 테면, DNA 중합효소 모두가 내포됨)가 내포되나, 이에 국한되지 않는다. 핵산 단편 상에 단일 가닥으로 된 갭을 채우는 방법에 중합효소를 제공하는 것이 내포된 경우에서, 상기 방법에는 데옥시리보뉴클레오티드 트리포스페이트 (dNTPs; 가령, dATP, dGTP, dCTP, 및 dTTP)를 제공하는 것이 또한 내포될 수 있다. 일부 경우에서, 5' 어댑터를 상기 각 핵산 단편의 5' 단부에 부착시키는 단계와 단일 가닥으로 된 갭을 채우는 단계는 동시에 실행될 수 있다 (가령, 단일 반응 튜브안에서).

[0217] 일부 경우에서, 상기 어댑터를 템플릿에 부착시키는데 대체 방법이 이용될 수 있다. 예를 들면, 핵산 단편들을 단일 가닥 뉴클레아제 (가령, 오버행을 절단하기 위해)로 처리되고, 이어서 절찰시킨 후, 이 단편들을 이용하여 듀플렉스 시퀀싱 라이브러리를 준비할 수 있다. 예를 들면, 단일 뉴클레오티드를 각 핵산 단편의 3' 단부에 추가할 수 있고, 5' 단부에 상보성 염기를 함유하는 어댑터 (가령, 분자 바코드를 함유)를 각 핵산 단편에 절찰시켜, 어댑터-부착된 템플릿의 듀플렉스 시퀀싱 라이브러리를 준비할 수 있다.

[0218] 어댑터-부착된 템플릿의 초기 증폭

[0219] 어댑터 부착-후, 이 어댑터-부착된 템플릿을 초기 증폭 반응에서 증폭시킬 수 있다(가령, PCR 증폭된). 임의의 적합한 방법을 이용하여 어댑터-부착된 템플릿을 증폭시킬 수 있다. 어댑터-부착된 템플릿을 증폭시키는데 이용될 수 있는 예시적인 방법에는 온전한-게놈 PCR이 내포되나, 이에 국한되지 않는다.

[0220] 임의의 적합한 프라이머 쌍을 이용하여 어댑터-부착된 템플릿을 증폭시킬 수 있다. 일부 경우에서, 범용 프라이머 쌍이 이용될 수 있다. 프라이머에는 약 12개 뉴클레오티드 ~ 약 30개의 뉴클레오티드가 내포되지만, 이에 국한되지 않는다. 어댑터-부착된 템플릿을 증폭시키는데 이용될 수 있는 프라이머 쌍의 예시로는 실시예 1 및/또는 실시예 2에서 기술된 것들이 내포되나, 이에 국한되지 않는다.

[0221] 임의의 적합한 PCR 조건이 상기 초기 증폭에 이용될 수 있다. PCR 증폭은 변성 단계, 어닐링 단계 및 확장 단계들이 내포될 수 있다. 증폭 주기의 각 단계에는 임의의 적합한 조건이 내포될 수 있다. 일부 경우에서, 변성 단계에는 약 90℃ ~ 약 105℃ (가령, 약 94℃ ~ 약 98℃)의 온도, 및 약 1 초 ~ 약 5 분 (가령, 약 10 초 ~ 약 1 분)의 시간이 내포될 수 있다. 예를 들면, 변성 단계에는 약 10 초 동안 약 98℃의 온도가 내포될 수 있다. 일부 경우에서, 어닐링 단계에는 약 50℃ ~ 약 72℃의 온도, 및 약 30 초 ~ 약 90 초의 시간이 내포될 수 있다. 일부 경우에서, 연장 단계에는 약 55℃ ~ 약 80℃의 온도, 및 생성되는 앰플리콘의 kb당 약 15 초 ~ 생성되는 앰플리콘의 kb당 약 30 초의 시간이 내포될 수 있다. 일부 경우에서, 어닐링 단계 및 연장 단계는 단

일 주기에서 실행될 수 있다. 예를 들면, 어닐링 단계 및 연장 단계에는 약 75 초 동안 약 65℃의 온도가 내포될 수 있다.

[0222] 상기 초기 증폭에 이용된 PCR 조건에는 임의의 적합한 횟수의 PCR 증폭 주기가 내포될 수 있다. 일부 경우에서, PCR 증폭에는 약 1 ~ 약 50회 주기가 내포될 수 있다. 일부 구체예들에서, 상기 PCR 증폭은 11회를 넘지 않는 주기를 포함한다. 일부 구체예들에서, 상기 PCR 증폭은 7회를 넘지 않는 주기를 포함한다. 일부 구체예들에서, 상기 PCR 증폭은 5회를 넘지 않는 주기를 포함한다.

[0223] 일부 경우에서, PCR 조건에 열에 의해-활성화된 중합효소가 내포될 경우, PCR 증폭에는 초기화(initialization) 단계가 또한 내포될 수 있다. 예를 들면, PCR 증폭에는 이들 PCR 증폭 주기 실행에 앞서, 초기화 단계가 또한 내포될 수 있다. 일부 경우에서, 초기화 단계에는 약 94℃ ~ 약 98℃의 온도, 및 약 15 초 ~ 약 1 분의 시간이 내포될 수 있다. 예를 들면, 초기화 단계에는 약 30초 동안 약 98℃의 온도가 내포될 수 있다.

[0224] 일부 경우에서, PCR 증폭에는 대기(hold) 단계가 또한 내포될 수 있다. 예를 들면, PCR 증폭에는 PCR 증폭 주기를 실행한 후, 임의선택적으로 임의의 최종 연장 단계를 실행한 후, 단계가 또한 내포될 수 있다. 일부 경우에서, 대기 단계에는 무기한의 시간 동안 약 4℃ ~ 약 15℃의 온도가 내포될 수 있다.

[0225] 일부 경우에서, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리 (가령, 증폭된 듀플렉스 시퀀싱 라이브러리)를 정제시킬 수 있다. 임의의 적합한 방법을 이용하여 듀플렉스 시퀀싱 라이브러리를 정제시킬 수 있다. 듀플렉스 시퀀싱 라이브러리를 정제시키는데 이용될 수 있는 예시적인 방법에는 자성 비드 (가령, 고체상 가역 고정화 (SPRI) 자성 비드)가 내포되나, 이에 국한되지 않는다.

[0226] *임의선택적 ssDNA 라이브러리 준비*

[0227] 일부 경우에서, 듀플렉스 시퀀싱 라이브러리를 이용하여 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리 및 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있다. 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 생성함으로써, (가령, 결합된 서열, 이를 떼면, 3' 듀플렉스 어댑터 또는 5' 어댑터에 상보적인 프라이머로부터) 비-특이적 증폭을 최소화시킬 수 있다. 임의의 적합한 방법을 이용하여, (가령, 본원에서 기술된 바와 같이 만들어진 듀플렉스 시퀀싱 라이브러리로부터) 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있다. 일부 경우에서, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 상기 증폭 산물을 적어도 두 개 분액으로 나누고, 각 분액을 PCR 증폭시켜서, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있는데 여기에서 Watson 가닥은 제 1 분액으로부터 증폭되며, Crick 가닥은 제 2 분액으로부터 증폭된다. 예를 들면, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 증폭 산물의 제 1 분액은 프라이머 쌍을 이용한 PCR 증폭을 거쳐 Watson 가닥의 단일 가닥으로 된 라이브러리를 생성시킬 수 있으며, 여기에서 제 1 프라이머는 바이오틀화되고, 제 2 프라이머는 바이오틀화안되고, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 증폭 산물의 제 2 분액은 프라이머 쌍을 이용한 PCR 증폭을 거쳐 Crick 가닥의 단일 가닥으로 된 라이브러리를 생성시킬 수 있으며, 여기에서 제 1 프라이머는 바이오틀화안되고, 제 2 프라이머는 바이오틀화된다. 일부 경우에서, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리는 도 2 및 도 3에서 나타난 것과 같이 만들어질 수 있다.

[0228] 임의의 적합한 방법을 이용하여, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있다. 예를 들면, 증폭된 듀플렉스 시퀀싱 라이브러리로부터 증폭 산물은 제 1 PCR 증폭과 제 2 PCR 증폭으로 분리될 수 있는데, 이때 PCR 프라이머 쌍에서 두 개 프라이머중 오로지 한 개만 태그된다. 예를 들면, 제 1 PCR 증폭은 태그된 프라이머 (가령, 제 1 프라이머) 및 태그안된 프라이머 (가령, 제 2 프라이머)가 내포된 프라이머 쌍을 이용할 수 있고, 제 2 PCR 증폭은 태그안된 프라이머 (가령, 제 1 프라이머) 및 태그된 프라이머 (가령, 제 2 프라이머)가 내포된 프라이머 쌍을 이용할 수 있다. 프라이머 태그는 상기 태그된 프라이머로부터 생성된 PCR 증폭 산물을 회수할 수 있는 임의의 태그일 수 있다. 일부 경우에서, 태그된 프라이머는 바이오틀화된 프라이머일 수 있고, 바이오틀화된 프라이머로부터 생성된 PCR 산물은 스트랩트아비딘을 이용하여 회수될 수 있다. 예를 들면, 바이오틀화된 프라이머와 바이오틀화안된 프라이머가 내포된 프라이머 쌍을 이용한 PCR 증폭에서 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리가 생성될 수 있다. 일부 경우에서, 태그된 프라이머는 인산화된 프라이머일 수 있고, 상기 인산화된 프라이머로부터 생성된 PCR 증폭 산물은 람다 뉴클레아제를 이용하여 회수될 수 있다. 예를 들면, 인산화된 프라이머

와 비-인산화된 프라이머가 내포된 프라이머 쌍을 이용한 PCR 증폭에서 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리가 생성될 수 있다.

[0229] 임의의 적합한 프라이머 쌍을 이용하여, (가령, 본원에서 기술된 바와 같이 만들어진 듀플렉스 시퀀싱 라이브러리로부터) 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있다. 프라이머에는 약 12개 뉴클레오티드 ~ 약 30개의 뉴클레오티드가 내포되지만, 만이에 국한되지 않는다. 일부 경우에서, 프라이머 쌍에는 본원에서 기술된 바와 같이 생성된 증폭 산물 (가령, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 듀플렉스 시퀀싱 라이브러리에 있는 핵산 단편에 대한 제 2 분자 바코드가 내포된 5' 어댑터에 증폭-전 결합시킴으로써) 상에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 하는 (가령, 표적으로 하여, 이에 결합하고) 적어도 하나의 프라이머가 내포될 수 있다. 본원에 기술된 바와 같이, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만드는데 이용될 수 있는 프라이머 쌍의 예시에는 P5 프라이머와 P7 프라이머가 내포되나, 이에 국한되지 않는다.

[0230] 임의의 적합한 PCR 조건을 이용하여, (가령, 본원에서 기술된 바와 같이 만들어진 듀플렉스 시퀀싱 라이브러리로부터) 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만들 수 있다. PCR 증폭은 변성 단계, 어닐링 단계 및 확장 단계들이 내포될 수 있다. 증폭 주기의 각 단계에는 임의의 적합한 조건이 내포될 수 있다. 일부 경우에서, 변성 단계에는 약 90℃ ~ 약 105℃의 온도, 및 약 1 초 ~ 약 5 분의 시간이 내포될 수 있다. 예를 들면, 변성 단계에는 약 10 초 동안 약 98℃의 온도가 내포될 수 있다. 일부 경우에서, 어닐링 단계에는 약 50℃ ~ 약 72℃의 온도, 및 약 30 초 ~ 약 90 초의 시간이 내포될 수 있다. 일부 경우에서, 연장 단계에는 약 55℃ ~ 약 80℃의 온도, 및 생성되는 앰플리콘의 kb당 약 15 초 ~ 생성되는 앰플리콘의 kb당 약 30 초의 시간이 내포될 수 있다. 일부 경우에서, 연장 단계는 사용되는 중합효소의 진행성을 반영한다. 일부 경우에서, 어닐링 단계 및 연장 단계는 단일 주기에서 실행될 수 있다. 예를 들면, 어닐링 단계 및 연장 단계에는 약 75 초 동안 약 65℃의 온도가 내포될 수 있다.

[0231] (가령, 본원에 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리로부터) 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리를 만드는데 이용되는 PCR 조건에는 임의의 적합한 횟수의 PCR 증폭 주기가 내포될 수 있다. 일부 경우에서, PCR 증폭에는 약 1 ~ 약 50회 주기가 내포될 수 있으나, 이에 국한되지 않는다. 예를 들면, PCR 증폭에는 약 4회 증폭 주기가 내포될 수 있다.

[0232] 일부 경우에서, PCR 조건에 열에 의해-활성화된 중합효소가 내포될 경우, PCR 증폭에는 초기화(initialization) 단계가 또한 내포될 수 있다. 예를 들면, PCR 증폭에는 이들 PCR 증폭 주기 실행에 앞서, 초기화 단계가 또한 내포될 수 있다. 일부 경우에서, 초기화 단계에는 약 94℃ ~ 약 98℃의 온도, 및 약 15 초 ~ 약 1 분의 시간이 내포될 수 있다. 예를 들면, 초기화 단계에는 약 30초 동안 약 98℃의 온도가 내포될 수 있다.

[0233] 일부 경우에서, PCR 증폭에는 대기(hold) 단계가 또한 내포될 수 있다. 예를 들면, PCR 증폭에는 PCR 증폭 주기를 실행한 후, 임의선택적으로 임의의 최종 연장 단계를 실행한 후, 단계가 또한 내포될 수 있다. 일부 경우에서, 대기 단계에는 무기한의 시간 동안 약 4℃ ~ 약 15℃의 온도가 내포될 수 있다.

[0234] 임의의 적합한 방법을 이용하여 중 가닥으로 된 증폭 산물을 단일 가닥으로 된 증폭 산물로 분리시킬 수 있다. 일부 경우에서, 이중 가닥으로 된 증폭 산물을 변성시켜, 이중 가닥으로 된 증폭 산물을 두 개의 단일 가닥으로 된 증폭 산물로 분리시킬 수 있다. 이중 가닥으로 된 증폭 산물을 단일 가닥으로 된 증폭 산물로 분리시키는데 이용될 수 있는 방법의 예시에는 열 변성, 화학적 (가령, NaOH) 변성, 및 염 변성이 내포되나, 이에 국한되지 않는다.

[0235] PCR 증폭 후, 테그된 PCR 증폭 산물을 회수할 수 있다. 테그된 프라이머를 이용하여 생성된 테그된 PCR 증폭 산물은 임의의 적합한 방법을 이용하여 회수할 수 있다. 여기에서 테그된 프라이머가 바이오티닐화된 프라이머인 경우에, 스트렙타아비딘 (가령, 스트렙타아비딘-기능화된 비드)를 이용하여 바이오티닐화된 증폭 산물 (가령, 바이오티닐화된 프라이머로부터 생성된)을 회수할 수 있다. 예를 들면, 증폭된 듀플렉스 시퀀싱 라이브러리가 바이오티닐화된 제 1 프라이머와 바이오티닐화안된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 1 PCR 증폭과 바이오티닐화안된 제 1 프라이머와 바이오티닐화된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 2 PCR 증폭에서 추가 증폭될 때, 상기 제 1 PCR 증폭으로부터 생성된 바이오티닐화된 증폭 산물은 스트렙타아비딘-기능화된 비드 (가령, 스트렙타아비딘-기능화된 비드의 제 1 세트)에 결합될 수 있고, 상기 제 2 PCR 증폭으로부터 생성된 바이오티닐화된 증폭 산물은 스트렙타아비딘-기능화된 비드 (가령, 스트렙타아비딘-기능화된

비드의 제 2 세트)에 결합될 수 있고, 상기 이중 가닥으로 된 증폭 산물은 단일 가닥의 증폭 산물로 분리될 수 있다(가령, 변성될 수 있다). 일부 경우에서, 바이오티닐화된 PCR 증폭 산물을 회수하는 단계에는 상기 스트랩트아비딘 (가령, 스트랩트아비딘-기능화된 비드)으로부터 바이오티닐화된 PCR 증폭 산물을 방출시키는 단계가 또한 내포될 수 있다. 바이오티닐화된 제 1 프라이머와 바이오티닐화안된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 1 PCR 증폭, 그리고 바이오티닐화안된 제 1 프라이머와 바이오티닐화된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 2 PCR 증폭에 의해 생성된 이중 가닥으로 된 증폭 산물의 분리로 바이오티닐화된 프라이머로부터 생성된 단일 가닥으로 된 증폭 산물은 스트랩트아비딘-기능화된 비드에 결합된 채로 유지되도록 할 수 있고, 한편 바이오티닐화안된 프라이머로부터 생성된 단일 가닥으로 된 증폭 산물은 스트랩트아비딘-기능화된 비드로 부터 변성되고(가령, 변성되고, 분해되어), 이로써 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리의 듀플렉스 시퀀싱 라이브러리가 생성되도록 할 수 있다.

[0236] 테그된 프라이머가 인산화된 프라이머인 경우들에서, 상기 인산화된 증폭 산물 (가령, 상기 인산화된 프라이머로부터 생성된)은 엑소뉴클레아제 (가령, 람다 엑소뉴클레아제)를 이용하여 회수될 수 있다. 예를 들면, 증폭된 듀플렉스 시퀀싱 라이브러리가 인산화된 제 1 프라이머와 인산화안된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 1 PCR 증폭, 그리고 인산화안된 제 1 프라이머와 인산화된 제 2 프라이머 내포된 프라이머 쌍을 이용한 제 2 PCR 증폭에서 추가 증폭될 때, 상기 이중 가닥으로 된 증폭 산물은 단일 가닥의 증폭 산물로 분리될 수 있다. 인산화된 제 1 프라이머와 인산화안된 제 2 프라이머가 내포된 프라이머 쌍을 이용한 제 1 PCR 증폭, 그리고 인산화안된 제 1 프라이머와 인산화된 제 2 프라이머 내포된 프라이머 쌍을 이용한 제 2 PCR 증폭에 의해 생성된 이중 가닥으로 된 증폭 산물의 분리로 상기 비-인산화된 프라이머로부터 생성된 단일 가닥으로 된 증폭 산물이 회수될 수 있고, 한편상기 인산화된 프라이머로부터 생성된 단일 가닥으로 된 증폭 산물은 람다 엑소뉴클레아제에 의해 분해될 수 있고, 이로써 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리의 듀플렉스 시퀀싱 라이브러리가 생성될 수 있다.

[0237] 표적 풍부화

[0238] 본원의 방법들중 임의의 하나의 방법의 일부 구체예들에서, 상기 초기 증폭에 의해 생성된 앰플리콘은 하나 또는 그 이상의 표적 폴리뉴클레오티드에 대해 풍부화된다. 일부 구체예들에서, 표적 풍부화-전, 단일-가닥으로 된 DNA 라이브러리는 초기 증폭에 의해 생성된 앰플리콘으로부터 준비된다. 단일-가닥으로 된 DNA 라이브러리를 생산하는 예시적인 방법들이 본원에 기술된다.

[0239] 임의의 적합한 방법을 이용하여 증폭 산물의 라이브러리(가령, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리)로부터 표적 영역을 증폭시킬 수 있다. 일부 경우에서, 증폭 산물의 라이브러리를 프라이머 쌍을 이용하여 PCR 증폭시킴으로써 증폭 산물의 라이브러리로부터 표적 영역이 증폭될 수 있고, 여기에서 프라이머 쌍은 본원에서 기술된 바와 같이 생성된 (가령, 증폭-전, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 듀플렉스 시퀀싱 라이브러리 안에 있는 핵산 단편에 결합시킴으로써) 증폭 산물에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 할 수 있는 (가령, 표적으로 하고, 이에 결합하는) 프라이머 (가령, 제 1 프라이머)와 표적 영역 (가령, 관심대상 영역)을 표적으로 할 수 있는 (가령, 표적으로 하고, 이에 결합하는) 프라이머 (가령, 제 2 프라이머)이다. 일부 경우에서, 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리는 도 4 및 도 5에서 나타난 것과 같이 만들어질 수 있다. 일부 경우에서, 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리와 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리는 실시예 2에서 기술된 바와 같이 만들어질 수 있다.

[0240] 일부 경우에서, 단일 PCR 증폭에서 증폭 산물의 라이브러리 (가령, 기술된 바와 같이 생성된, 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끝어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끝어낸 서열의 라이브러리)로부터 표적 영역이 증폭될 수 있다. 예를 들면, 단일 PCR 증폭에서 프라이머 쌍을 이용하여 증폭 산물의 라이브러리로부터 표적 영역이 증폭될 수 있고, 여기에서 프라이머 쌍에는 본원에서 기술된 바와 같이 생성된 (가령, 증폭-전, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 듀플렉스 시퀀싱 라이브러리 안에 있는 핵산 단편에 결합시킴으로써) 증폭 산물에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 할 수 있는 프라이머 (가령, 제 1 프라이머)와 표적 영역 (가령, 관심대상 영역)을 표적으로 할 수 있는 프라이머 (가령, 제 2 프라이머)가 내포된다. 예를 들면, 도 4, 도 5, 도 15, 및 도 17에 나타난 것과 같이, 단일 PCR 증폭에서 증폭 산물의 라이브러

리로부터 표적 영역이 증폭될 수 있다.

[0241] 일부 경우에서, 다중 PCR 증폭에서 증폭 산물의 라이브러리 (가령, 기술된 바와 같이 생성된, 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리)로부터 표적 영역이 증폭될 수 있다. 다중 PCR 증폭 (가령, 제 1 PCR 증폭에 이어, 후속, 네스티드 PCR 증폭)을 이용하여 표적 영역의 증폭 특이성을 증가시킬 수 있다. 예를 들면, 일련의 PCR 증폭에서 증폭 산물의 라이브러리로부터 표적 영역이 증폭될 수 있는데, 여기에서 제 1 PCR 증폭은 본원에서 기술된 바와 같이 생성된 증폭 산물에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 할 수 있는 제 1 프라이머 (가령, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 듀플렉스 시퀀싱 라이브러리에 있는 핵산 단편에 증폭 전, 결합시킴으로써), 그리고 표적 영역을 표적으로 할 수 있는 제 2 프라이머가 내포된 프라이머 쌍을 이용하고, 상기 제 1 PCR 증폭에서 생성된 증폭 산물을 후속, 네스티드 PCR 증폭을 거치게 하고, 후속, 네스티드 PCR 증폭은 프라이머 쌍은 본원에서 기술된 바와 같이 생성된 (가령, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 증폭 전, 듀플렉스 시퀀싱 라이브러리에 있는 핵산 단편에 결합시킴으로써) 증폭 산물에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 하는 제 1 프라이머와 상기 제 1 PCR 증폭에서 생성된 증폭 산물에 존재하는 표적 영역에 있는 핵산 서열을 표적으로 하는 제 2 프라이머를 이용한다. 예를 들면, 도 7, 도 8, 도 16, 및 도 18에 나타난 것과 같이, 일련의 PCR 증폭에서 증폭 산물의 라이브러리로부터 표적 영역이 증폭될 수 있다.

[0242] 임의의 적합한 프라이머 쌍을 이용하여 증폭 산물의 라이브러리(가령, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리)로부터 표적 영역을 증폭시킬 수 있다. 프라이머에는 약 12개 뉴클레오티드 ~ 약 30개의 뉴클레오티드가 내포되지만, 만이에 국한되지 않는다. 일부 경우에서, 프라이머 쌍에는 본원에서 기술된 바와 같이 생성된 (가령, 증폭-전, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 듀플렉스 시퀀싱 라이브러리 안에 있는 핵산 단편에 결합시킴으로써) 증폭 산물에 존재하는 어댑터 서열 (가령, 분자 바코드를 함유하는 어댑터 서열)을 표적으로 할 수 있는 (가령, 표적으로 하고, 이에 결합하는) 프라이머 (가령, 제 1 프라이머)와 표적 영역 (가령, 관심대상 영역)을 표적으로 할 수 있는 (가령, 표적으로 하고, 이에 결합하는) 프라이머 (가령, 제 2 프라이머)가 내포될 수 있다. 본원에서 기술된 바와 같이 생성된 (가령, 제 1 분자 바코드가 내포된 3' 듀플렉스 어댑터와 제 2 분자 바코드가 내포된 5' 어댑터를 듀플렉스 시퀀싱 라이브러리에 있는 핵산 단편에 증폭-전 결합시킴으로써) 증폭 산물에 존재하는 분자 바코드를 함유하는 어댑터 서열을 표적으로 할 수 있는 프라이머의 예시에는 5 인덱스 프라이머와 i7 인덱스 프라이머가 내포되나, 이에 국한되지 않는다. 표적 영역을 표적으로 할 수 있는 프라이머에는 상기 표적 영역에 상보적인 서열이 내포된다. 표적 영역이 TP53을 인코딩하는 핵산인 경우, TP53을 인코딩하는 핵산을 표적으로 할 수 있는 프라이머의 예시에는 TP53_342_GSP1 및 TP53_GSP2이 내포되나, 이에 국한되지 않는다. 일부 경우에서, 표적 영역이 TP53을 인코딩하는 핵산일 경우, TP53을 인코딩하는 핵산을 표적으로 하는 프라이머는 실시예 2에서 기술된 바와 같은 것일 수 있다.

[0243] 일부 경우에서, 증폭 산물의 라이브러리 (가령, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리)로부터 표적 영역을 증폭시키는데 이용된 프라이머 쌍중 하나의 프라이머 또는 두 프라이머 모두에는 하나 또는 그 이상의 분자 바코드가 내포될 수 있다.

[0244] 일부 경우에서, 증폭 산물의 라이브러리 (가령, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리)로부터 표적 영역을 증폭시키는데 이용된 프라이머 쌍중 하나의 프라이머 또는 두 프라이머 모두에는 하나 또는 그 이상의 접목(graft) 서열 (가령, 차세대 시퀀싱용 접목 서열)이 내포될 수 있다.

[0245] 한 측면에서, 상기 표적 풍부화 단계는 다음을 포함한다: (a)표적 폴리뉴클레오티드 서열을 포함하는 Watson 가닥의 앰플리콘을 Watson 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Watson 증폭 산물이 창출되며, 이때 Watson 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보성 서열을 포함하는 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 표적-선택적 서열을 포함하는 제 2 Watson 표적-선택적 프라이머를 포함하고; 그리고 (b) 동일한 표적 폴리뉴클레오티드 서열을 포함하는 Crick 가닥의 앰플리콘을 Crick 표적-선택적 프라이머 쌍의 제 1 세트로 선택적으로 증폭시키고, 이로써 표적 Crick 증폭 산물이 창출되며, 이때 Crick 표적-선택적 프라이머 쌍의 제 1 세트는 (i) 범용 5' 어

댕터 서열의 R1 시퀀싱 프라이머 부위에 상보적 서열을 포함하는 제 1 Crick 표적-선택적 프라이머, 그리고 (ii) 상기 제 2 Watson 표적-선택적 프라이머 서열과 동일한 동일한 표적-선택적 서열을 포함하는 제 2 Crick 표적-선택적 프라이머를 포함한다.

[0246] 일부 구체예들에서, 상기 방법은 비-표적 폴리뉴클레오티드로부터 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 정제해내는 단계를 더 포함한다. 일부 구체예들에서, 상기 정제 방법은 상기 표적 Watson 증폭 산물과 상기 표적 Crick 증폭 산물을 고형 서포트에 부착시키는 단계를 더 포함한다. 일부 구체예들에서, 상기 제 1 Watson 표적-선택적 프라이머와 제 1 Crick 표적-선택적 프라이머는 친화력 결합 쌍의 제 1 구성원을 포함하고, 그리고 이때 상기 고형 서포트는 상기 친화력 결합 쌍의 제 2 구성원을 포함한다. 일부 구체예들에서, 상기 제 1 구성원은 바이오틴이며, 상기 제 2 구성원은 스트렙타비딘이다. 일부 구체예들에서, 상기 고형 지지대는 비드, 웰(well), 막, 튜브, 컬럼, 플레이트, 세파로즈, 자석 비드, 또는 칩을 포함한다. 일부 구체예들에서, 상기 방법은 상기 고형 서포트에 부착되지 않은 폴리뉴클레오티드를 제거하는 단계를 포함한다.

[0247] 일부 구체예들에서, 상기 방법은 다음을 더 포함한다: (a) 상기 표적 Watson 증폭 산물을 Watson 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Watson 라이브러리 구성원을 창출시키고, 이때 상기 Watson 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하는 제 4 Watson 표적-선택적 프라이머를 포함하고; (b) 상기 표적 Crick 증폭 산물을 Crick 표적-선택적 프라이머의 제 2 세트에 추가 증폭시키고, 이로써 표적 Crick 라이브러리 구성원을 창출시키고, 이때 상기 Crick 표적-선택적 프라이머의 제 2 세트는 (i) 범용 3' 어댑터 서열의 R1 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 제 3 Crick 표적-선택적 프라이머, 그리고 (ii) (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위와 상기 제 4 Watson 표적-선택적 프라이머의 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하는 제 4 Crick 표적-선택적 프라이머를 포함한다.

[0248] 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머와 Crick 표적-선택적 프라이머는 샘플 바코드 서열을 더 포함한다. 일부 구체예들에서, 상기 제 3 Watson 표적-선택적 프라이머는 시퀀서(sequencer) 상에서 제 1 접목 프라이머에 혼성화될 수 있는 제 1 접목(grafting) 서열을 더 포함하고, 이때 상기 제 3 Crick 표적-선택적 프라이머는 이 시퀀서 상에서 제 2 접목 프라이머에 혼성화될 수 있는 제 2 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 4 Watson 표적-선택적 프라이머는 상기 제 2 접목 서열을 더 포함하고, 이때 상기 제 4 Crick 표적-선택적 프라이머는 상기 제 1 접목 서열을 더 포함한다. 일부 구체예들에서, 상기 제 1 접목 서열은 P7 서열이며, 이때 상기 제 2 접목 서열은 P5 서열이다.

[0249] 임의의 적합한 PCR 조건을 이용하여 본원에서 기술된 바와 같이, 증폭된 표적 영역을 만들 수 있다 (가령, 증폭 산물의 라이브러리, 이를 테면, 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 생성된 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리로부터). 예시적인 PCR 조건은 본원에서 기술된다. 본원에서 기술된 바와 같이 증폭된 표적 영역을 만드는데 이용된 PCR 조건 (가령, 증폭 산물의 라이브러리, 이를 테면, 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 생성된 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리로부터)에는 임의의 적합한 횟수의 PCR 증폭 주기가 내포될 수 있다. 일부 경우에서, PCR 증폭에는 약 1 ~ 약 50회 주기가 내포될 수 있으나, 이에 국한되지 않는다. 예를 들면, 증폭된 표적 영역의 PCR 증폭에 단일 PCR 증폭이 내포될 경우, 상기 PCR 증폭에는 약 18회 증폭 주기가 내포될 수 있다. 예를 들면, 증폭된 표적 영역의 PCR 증폭에 제 1 PCR 증폭 그리고 후속적으로, 네스티드 PCR 증폭이 내포될 경우, 상기 제 1 PCR 증폭에는 약 18회 증폭 주기가 내포될 수 있고, 상기 후속적 네스티드 PCR 증폭에는 약 10회 증폭 주기가 내포될 수 있다.

[0250] 예시적인 표적

[0251] 임의의 적합한 표적 영역 (가령, 관심대상 영역)은 증폭 산물의 라이브러리 (가령, 본원에서 기술된 바와 같이 생성된 듀플렉스 시퀀싱 라이브러리, 단일 가닥으로 된 Watson 가닥-끌어낸 서열의 라이브러리, 또는 단일 가닥으로 된 Crick-가닥 끌어낸 서열의 라이브러리)로부터 증폭되고, 하나 또는 그 이상의 돌연변이가 존재하는지 또는 없는 지에 대해 평가할 수 있다. 일부 경우에서, 표적 영역은 핵산의 핵산일 수 있고, 이때 하나 또는 그 이상의 돌연변이는 질환 또는 장애와 연관된다. 증폭될 수 있고, 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 평가될 수 있는 표적 영역의 예시에는 종양 단백질 p53 (TP53)을 인코딩하는 핵산, 유방암 1

(BRCA1)를 인코딩하는 핵산, BRCA2, 포스파타제 및 텐신 호모로그 (PTEN) 폴리펩티드를 인코딩하는 핵산, AKT1 폴리펩티드를 인코딩하는 핵산, APC 폴리펩티드를 인코딩하는 핵산, CDKN2A 폴리펩티드를 인코딩하는 핵산, EGFR 폴리펩티드를 인코딩하는 핵산, FBXW7 폴리펩티드를 인코딩하는 핵산, GNAS 폴리펩티드를 인코딩하는 핵산, KRAS 폴리펩티드를 인코딩하는 핵산, NRAS 폴리펩티드를 인코딩하는 핵산, PIK3CA 폴리펩티드를 인코딩하는 핵산, BRAF 폴리펩티드를 인코딩하는 핵산, CTNNB1 폴리펩티드를 인코딩하는 핵산, FGFR2 폴리펩티드를 인코딩하는 핵산, HRAS 폴리펩티드를 인코딩하는 핵산, 및 PPP2R1A 폴리펩티드를 인코딩하는 핵산이 내포된다. 일부 경우에서, 증폭될 수 있고, 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 평가될 수 있는 표적 영역은 TP53을 인코딩하는 핵산일 수 있다. 예를 들면, TP53을 인코딩하는 핵산은 실시예 2에서 기술된 바와 같이, 증폭될 수 있고, 평가될 수 있다.

[0252] 임의의 적합한 방법을 이용하여, 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 평가될 수 있는 표적 영역 (가령, 증폭된 표적 영역)을 평가할 수 있다. 일부 경우에서, 하나 또는 그 이상의 시퀀싱 방법을 이용하여 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 평가될 수 있는 표적 영역을 평가할 수 있다.

[0253] 서열 결정

[0254] 일부 경우에서, 하나 또는 그 이상의 시퀀싱 방법을 이용하여 증폭된 표적 영역을 평가하여, 이들 돌연변이 (들)이 Watson 가닥과 Crick 가닥 모두에 존재하는 지 여부를 결정할 수 있다. 일부 경우에서, 시퀀싱 판독을 이용하여 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 증폭된 표적 영역을 평가하고, 이들 돌연변이 (들)이 Watson 가닥과 Crick 가닥 모두에 존재하는 지 여부를 결정할 수 있다. 본원에서 기술된 바와 같이, 하나 또는 그 이상의 돌연변이의 존재 또는 부재에 대해 평가하는데 이용될 수 있는 시퀀싱 방법에는 단일 판독 시퀀싱, 페어드-엔드 시퀀싱, NGS, 및 딥(deep) 시퀀싱이 내포되나, 이에 국한되지 않는다. 일부 구체예들에서, 단일 판독 시퀀싱은 템플릿의 길이에 걸쳐 시퀀싱하는 것을 포함하고, 이 서열 판독이 생성된다. 일부 구체예들에서, 상기 시퀀싱은 페어드 엔드 시퀀싱을 포함한다. 일부 구체예들에서, 상기 시퀀싱은 대규모 병렬 시퀀서를 이용하여 실행된다. 일부 구체예들에서, 상기 대규모 병렬 시퀀서는 템플릿 폴리뉴클레오티드의 양 단부로부터 서열 판독을 결정하도록 설정된다.

[0255] 서열 판독의 분석

[0256] 일부 구체예들에서, 상기 서열 판독 참조 게놈에 맵핑된다.

[0257] 일부 구체예들에서, 상기 서열 판독은 UID 패밀리에 할당된다. UID 패밀리는 원래 템플릿, 가령, 핵산 샘플로부터 원래 이중-가닥으로 된 DNA 단편으로부터 기인된 앰플리콘의 서열 판독을 포함할 수 있다.

[0258] 일부 구체예들에서, UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함한다. 일부 구체예들에서, UID 패밀리의 각 구성원은 동일한 내인성 UID 서열을 더 포함한다. 내인성 UIDs가 본원에서 기술된다.

[0259] 일부 구체예들에서, UID 패밀리의 각 구성원은 동일한 외인성 UID 서열과 동일한 내인성 UID 서열을 더 포함한다. 일부 구체예들에서, 상기 외인성 UID 서열과 내인성 UID 서열의 조합은 이 UID 패밀리에 독특하다. 일부 구체예들에서, 상기 외인성 UID 서열과 내인성 UID 서열의 조합은 해당 핵산 샘플에서 제시되는 또다른 UID 패밀리에 존재하지 않는다.

[0260] UID 패밀리의 구성원의 수는 시퀀싱 심도에 따라 달라진다. 일부 구체예들에서, UID 패밀리는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 또는 1000개 구성원을 포함한다. 일부 구체예들에서, UID 패밀리는 약 2-1000개 구성원, 약 2-500개 구성원, 약 2-100개 구성원, 약 2-50개 구성원, 또는 약 2-20개 구성원을 포함한다.

[0261] 일부 구체예들에서, 개별 UID 패밀리의 서열 판독은 Watson 하위-패밀리와 Crick 하위-패밀리에 할당된다. 일부 구체예들에서, 개별 UID 패밀리의 서열 판독은 어댑터 서열에 대해 삽입물의 방향을 기반으로 하여, Watson 하위-패밀리와 Crick 하위-패밀리에 할당된다. 일부 구체예들에서, 어댑터 서열에 대해 삽입물의 방향은 해당 서열 판독이 "판독 쌍" 또는 "메이트 쌍"으로 정렬된 방식에 따라 결정된다.

[0262] 일부 구체예들에서, 상기 서열 판독이 Watson 하위-패밀리와 Crick 하위-패밀리로 할당은 R1 판독 서열과 R2 판독 서열에 대해 외인성 UID 서열의 공간적 상관관계를 기반으로 한다. 일부 구체예들에서, Watson 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R2 서열의 하류이며, R1 서열의 상류인 것에 의해 특징화된다. 일부 구

체예들에서, Crick 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R1 서열의 하류이며, R2 서열의 상류인 것에 의해 특징화된다. 일부 구체예들에서, Watson 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R2 서열에는 더 근접하며, R1 서열에 덜 근접한 것으로 특징화된다. 일부 구체예들에서, Crick 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R1 서열에는 더 근접하며, R2 서열에 덜 근접한 것으로 특징화된다. 일부 구체예들에서, Watson 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R2 서열의 바로 하류이거나, 또는 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오타이드 이내에 있는 것으로 특징화된다. 일부 구체예들에서, Crick 하위-패밀리의 구성원은 상기 외인성 UID 서열이 R1 서열의 바로 하류이거나, 또는 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오타이드 이내에 있는 것으로 특징화된다.

[0263] 일부 구체예들에서, UID 하위-패밀리 (가령, Watson 하위-패밀리 및/또는 Crick 하위-패밀리)는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 또는 500개 구성원을 포함한다. 일부 구체예들에서, UID 하위-패밀리 (가령, Watson 하위-패밀리 및/또는 Crick 하위-패밀리)는 약 2-500개 구성원, 약 2-100개 구성원, 약 2-50개 구성원, 약 2-20개 구성원, 또는 약 2-10개 구성원을 포함한다.

[0264] 일부 구체예들에서, 뉴클레오타이드 서열은 Watson 하위-패밀리의 구성원의 임계값 백분율(또는 임계값을 초과하는 백분율)이 이 서열을 함유할 때, 피분석물 DNA 단편, 가령, 해당 핵산 샘플로부터 이중 가닥으로 된 DNA 단편의 Watson 가닥을 정확하게 제시하는 것으로 결정된다. 일부 구체예들에서, 뉴클레오타이드 서열은 Crick 하위-패밀리의 구성원의 임계값 백분율(또는 임계값을 초과하는 백분율)이 이 서열을 함유할 때, 피분석물 DNA 단편, 가령, 해당 핵산 샘플로부터 이중 가닥으로 된 DNA 단편의 Crick 가닥을 정확하게 제시하는 것으로 결정된다.

[0265] 임계값은 예를 들어, 하위-패밀리의 구성원의 수, 시퀀싱 실험의 구체적 목적, 및 시퀀싱 실험의 특정 매개변수에 기초하여, 숙련된 기술자에 의해 결정될 수 있다. 일부 구체예들에서, 상기 임계값은 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%로 설정된다. 특정 구체예들에서, 상기 임계값은 50%로 설정된다. 오로지 예시로서, 임계값이 50%로 설정된 구체예에서, 뉴클레오타이드 서열은 하위-패밀리 구성원의 적어도 50%가 이 서열을 함유할 때, 피분석물 DNA 단편, 가령, 해당 핵산 샘플로부터 이중 가닥으로 된 DNA 단편의 Watson 가닥 또는 Crick 가닥을 정확하게 제시하는 것으로 결정된다. 오로지 예시로서, 임계값이 50%로 설정된 구체예에서, 뉴클레오타이드 서열은 하위-패밀리 구성원의 50% 이상이 이 서열을 함유할 때, 피분석물 DNA 단편, 가령, 해당 핵산 샘플로부터 이중 가닥으로 된 DNA 단편의 Watson 가닥 또는 Crick 가닥을 정확하게 제시하는 것으로 결정된다.

[0266] 일부 구체예들에서, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 서열은 돌연변이를 보유하는 것으로 판정된다. 일부 구체예들에서, 피분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이할 때, 해당 서열은 돌연변이를 보유하는 것으로 판정된다.

[0267] 일부 구체예들에서, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 서열은 돌연변이를 보유하는 것으로 판정된다. 일부 구체예들에서, 피분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이할 때, 해당 서열은 돌연변이를 보유하는 것으로 판정된다.

[0268] 일부 구체예들에서, Watson 가닥을 정확하게 나타내는 서열과 Crick 가닥을 정확하게 나타내는 서열이 동일한 돌연변이를 포함할 때, 피분석물 DNA 단편은 돌연변이를 보유하는 것으로 판정된다.

[0269] 일부 경우에서, 상기 증폭된 표적 영역의 페어드-엔드 시퀀싱 판독 내 분자 바코드의 위치를 이용하여 증폭된 표적 영역이 이중 가닥으로 된 핵산 템플릿의 어느 가닥으로부터 유래된 것인지를 구별할 수 있다. 예를 들면, 증폭된 표적 영역의 제 1 페어드-엔드 시퀀싱 판독이 마지막 판독인 경우, 상기 증폭된 표적 영역은 해당 핵산 템플릿의 센스 가닥으로부터 유래된 것으로 확인될 수 있고, 증폭된 표적 영역의 제 1 페어드-엔드 시퀀싱 판독이 첫 판독인 경우, 상기 증폭된 표적 영역은 해당 핵산 템플릿의 안티센스 가닥으로부터 유래된 것으로 확인될 수 있다. 예를 들면, 증폭된 표적 영역의 제 2 페어드-엔드 시퀀싱 판독이 첫 판독인 경우, 상기 증폭된 표적 영역은 해당 핵산 템플릿의 안티센스 가닥으로부터 유래된 것으로 확인될 수 있고, 증폭된 표적 영역의 제 2 페어드-엔드 시퀀싱 판독이 마지막 판독인 경우, 상기 증폭된 표적 영역은 해당 핵산 템플릿의 센스 가닥으로부터 유래된 것으로 확인될 수 있다. 일부 경우에서, 도 20 및 도 21에 나타난 것과 같이, 페어드-엔드 시퀀싱을 이용하여 Watson 가닥으로부터 유래된 증폭 산물과 Crick 가닥으로부터 유래된 증폭 산물을 구별할 수 있다.

- [0270] 표적 영역의 시퀀싱 (가령, 본원에서 기술된 바와 같이 증폭된 표적 영역) 후, 시퀀싱 판독을 참조 게놈에 정렬시키고, 각 시퀀싱 판독에 존재하는 분자 바코드에 따라 그룹화시킬 수 있다. 일부 경우에서, 동일한 분자 바코드가 내포되며, 이중 가닥으로 된 핵산 템플릿의 Watson 가닥과 Crick 가닥 모두 (가령, 표적 영역의 Watson 가닥과 Crick 가닥 모두)에 맵핑되는 시퀀싱 판독은 듀플렉스 서포트를 보유하는 것으로 확인될 수 있다. 예를 들면, 시퀀싱 판독이 표적 영역에 동일한 분자 바코드가 내포되며, 표적 영역의 Watson 가닥과 Crick 가닥 모두에 맵핑되는 표적 영역에 하나 또는 그 이상의 돌연변이가 존재하는 것으로 나타낼 경우, 이들 돌연변이(들)은 듀플렉스 서포트를 보유하는 것으로 확인될 수 있다.
- [0271] *키트*
- [0272] 키트가 또한 본원에서 제공된다. 상기 키트는 하나 또는 그 이상의 표적 폴리뉴클레오티드의 증폭을 위한 프라이머 쌍을 포함할 수 있다.
- [0273] 일부 구체예들에서, 상기 키트는 다음을 포함한다: (a) Watson 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 1 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 2 Watson 표적-선택적 프라이머 각각은 표적-선택적 서열을 포함하고; (b) Crick 표적-선택적 프라이머 쌍의 제 1 세트, 이 세트는 (i) 범용 5' 어댑터 서열의 R1 시퀀싱 프라이머 부위에 상보적 서열을 포함하는 하나 또는 그 이상의 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 2 Crick 표적-선택적 프라이머 각각은 상기 제 2 Watson 표적-선택적 프라이머 서열과 동일한 표적-선택적 서열을 포함하며; (c) Watson 표적-선택적 프라이머의 제 2 세트 쌍, 이때 이 쌍은 (i) 범용 3' 어댑터 서열의 R2 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Watson 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 4 Watson 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R1 시퀀싱 프라이머 부위와 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함하며; 그리고 (d) 제 2 세트의 Crick 표적-선택적 프라이머는 (i) 범용 3' 어댑터 서열의 R1 시퀀싱 프라이머 부위에 상보적인 서열을 포함하는 하나 또는 그 이상의 제 3 Crick 표적-선택적 프라이머, 그리고 (ii) 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머를 포함하며, 이때 하나 또는 그 이상의 제 4 Crick 표적-선택적 프라이머 각각은 (5'에서 3'방향으로) R2 시퀀싱 프라이머 부위 및 동일한 표적 폴리뉴클레오티드에 선택적인 표적-선택적 서열을 포함한다.
- [0274] 상기 키트는 다수의 표적 폴리뉴클레오티드의 다중 증폭을 위한 프라이머 쌍을 포함할 수 있다.
- [0275] *컴퓨터 판독가능한 매체*
- [0276] 본원에 기술된 방법들중 임의의 방법을 실행하도록 설정된, 컴퓨터 실행가능한 지침을 포함하는 컴퓨터 판독가능한 매체가 본원에서 또한 제공된다. 상기 컴퓨터 판독가능한 매체는 핵산 샘플로부터 서열 데이터를 분석하기 위한 컴퓨터 실행가능한 지침을 포함할 수 있고, 이때 데이터는 선행 청구항들중 임의의 한 항의 방법에 의해 생성된다.
- [0277] 상기 컴퓨터 판독가능한 매체는 반-자동 또는 자동화된 서열 데이터 분석 방법을 구현할 수 있다.
- [0278] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 다음 단계들을 위한 실행가능한 지침을 포함한다: (a) 서열 판독을 UID 패밀리로 할당하는 단계, 이때 UID 패밀리의 각 구성원은 동일한 외인성 UID 서열을 포함하고; (b) 각 UID 패밀리의 서열 판독을 Watson 하위-패밀리와 Crick 하위-패밀리로 할당하는 단계; (c) Watson 하위-패밀리 구성원의 임계값 백분율이 서열을 포함하는 경우, 분석물 DNA 단편의 Watson 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; (d) Crick 하위-패밀리 구성원의 임계값 백분율이 서열을 포함할 때, 분석물 DNA 단편의 Crick 가닥을 정확하게 나타내는 것으로 뉴클레오타이드 서열을 확인하는 단계; (e) Watson 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이한 경우, Watson 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계; (f) Crick 가닥을 정확하게 나타내는 서열이 돌연변이가 없는 참조 서열과 상이한 경우, Crick 가닥을 정확하게 나타내는 뉴클레오타이드 서열에서 돌연변이를 확인하는 단계; 그리고 (g) Watson 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이와 Crick 가닥을 정확히 나타내는 뉴클레오타이드 서열의 돌연변이가 동일한 돌연변이인 경우, 분석물 DNA 단편의 돌연변이를 확인하는 단계.
- [0279] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 R1 판독 서열과 R2 판독 서열에 대해 외인성 UID 서열의 공간적 상관관계를 기반으로 하여, UID 패밀리 구성원을 Watson 하위-패밀리 또는 Crick 하위-패밀리에 할당하기

위한 실행가능한 코드를 포함한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열의 하류와 R1 서열의 상류일 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Watson 하위-패밀리에 할당한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열의 하류와 R2 서열의 상류일 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Crick 하위-패밀리에 할당한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열에 더 근접하고, R1 서열에 덜 근접할 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Watson 하위-패밀리에 할당한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열에 더 근접하고, R2 서열에 덜 근접할 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Crick 하위-패밀리에 할당한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R2 서열의 바로 하류에 있거나, 또는 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오티드 이내에 있을 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Watson 하위-패밀리에 할당한다. 일부 구체예들에서, 상기 외인성 UID 서열이 R1 서열의 바로 하류에 있거나, 또는 1-70개, 1-60개, 1-50개, 1-40개, 1-30개, 1-20개, 1-10개, 또는 1-5개 뉴클레오티드 이내에 있을 경우, 상기 컴퓨터 실행가능한 코드는 UID 패밀리 구성원을 Crick 하위-패밀리에 할당한다.

[0280] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 참조 게놈에 상기 서열 판독을 맵핑하기 위한 실행가능한 코드를 포함한다. 일부 구체예들에서, 상기 참조 게놈은 인간 참조 게놈이다.

[0281] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 해당 샘플에 돌연변이의 존재, 부재 또는 돌연변이의 양을 기반으로 하여, 질환 상태, 예후 또는 치료 진단 보고서를 생성하기 위한 실행 코드를 포함한다. 일부 구체예들에서, 상기 질환은 암이다.

[0282] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 해당 샘플에 돌연변이의 존재, 부재 또는 돌연변이의 양을 기반으로 하여, 치료 옵션 보고서를 생성하기 위한 실행 코드를 포함한다.

[0283] 일부 구체예들에서, 상기 컴퓨터 판독가능한 매체는 네트워크를 통해 데이터를 전송할 수 있는 실행가능한 코드를 포함한다.

[0284] 컴퓨터 시스템

[0285] 컴퓨터 시스템이 또한 본원에서 제공된다. 일부 구체예들에서, 상기 컴퓨터 시스템은 핵산 샘플로부터 본원에 기재된 방법에 의해 생성된 서열 데이터를 수신 및 저장하도록 설정된 메모리 유닛; 그리고 저장 메모리에 통신 가능하게 결합된 프로세서를 포함하며, 여기서 이 프로세서는 본 명세서에 개시된 컴퓨터 판독가능 매체를 포함한다.

[0286] 도 32는 본 명세서에 기재된 임의의 방법에 따라 사용자가 핵산 샘플을 분석할 수 있도록 채택된 예시적인 컴퓨터 시스템(900)을 도시한다. 시스템(900)에는 본원에 설명된 예시적인 방법을 구현하도록 프로그래밍된 중앙 컴퓨터 서버(901)가 내포된다. 상기 서버 (901)는 단일 코어 또는 멀티 코어 프로세서일 수 있는 중앙 처리 장치(CPU, 또한 "프로세서") (905), 또는 병렬 처리를 위한 복수의 프로세서가 내포된다. 상기 서버(901)에는 또한 메모리(910)(예를 들어, 랜덤 액세스 메모리, 읽기 전용 메모리, 플래시 메모리); 전자 저장 유닛(915)(예를 들어, 하드 디스크); 하나 이상의 다른 시스템, 예를 들어, 시퀀싱 시스템과 통신하기 위한 통신 인터페이스(920)(예를 들어, 네트워크 어댑터); 및 캐시, 다른 메모리, 데이터 저장, 및/또는 전자 디스플레이 어댑터를 포함할 수 있는 주변 장치(925)가 내포된다. 메모리(910), 저장 유닛(915), 인터페이스(920), 및 주변 장치(925)는 마더보드와 같은 통신 버스(실선)를 통해 프로세서(905)와 통신한다. 저장 유닛(915)은 데이터를 저장하기 위한 데이터 저장 유닛일 수 있다. 서버(901)는 통신 인터페이스(920)의 도움으로 컴퓨터 네트워크("네트워크")(930)에 작동 가능하게 연결된다. 네트워크(930)는 인터넷, 인트라넷 및/또는 엑스트라넷, 인터넷과 통신하는 인트라넷 및/또는 엑스트라넷, 통신 또는 데이터 네트워크일 수 있다. 일부 경우에, 네트워크(930)는 서버(901)의 도움으로 피어-투-피어 네트워크를 구현할 수 있으며, 이는 서버(901)에 연결된 장치가 클라이언트 또는 서버로서 동작하도록 할 수 있다.

[0287] 저장 유닛(915)은 서열 데이터, 바코드 서열 데이터, 또는 본 발명과 관련된 데이터의 임의의 측면과 같은 파일을 저장할 수 있다. 데이터 저장 유닛(915)은 가상 그리드에서 셀의 위치에 관한 데이터와 결합될 수 있다.

[0288] 서버는 네트워크(930)를 통해 하나 또는 그 이상의 원격 컴퓨터 시스템과 통신할 수 있다. 상기 하나 또는 그 이상의 원격 컴퓨터 시스템은 예를 들어 개인용 컴퓨터, 랩탑, 태블릿, 전화기, 스마트폰, 또는 개인용 정보 단말기일 수 있다.

[0289] 일부 상황에서, 이 시스템(900)에는 단일 서버(901)가 내포된다. 다른 상황들에서, 이 시스템에는 인트라넷, 엑

스트라넷 및/또는 인터넷을 통해 서로 통신하는 여러 서버들이 내포된다.

- [0290] 이 서버(901)는 서열 데이터, 핵산 샘플에 관한 데이터, 생물학적 샘플에 관한 데이터, 대상체에 관한 데이터, 및/또는 잠재적 관련성이 있는 기타 정보를 저장하도록 구성될 수 있다. 이러한 정보는 저장 유닛(915) 또는 서버(901)에 저장될 수 있고, 이러한 데이터는 네트워크를 통해 전송될 수 있다.
- [0291] 본원에 설명된 방법은 예를 들어, 메모리(910) 또는 전자 저장 유닛(915)과 같은 서버(901)의 전자 저장 위치에 저장된 기계(가령, 컴퓨터 프로세서) 컴퓨터 판독 가능 매체(또는 소프트웨어)를 통해 구현될 수 있다. 사용 중, 상기 코드는 프로세서(905)에 의해 실행될 수 있다.
- [0292] 일부 경우에, 코드는 저장 유닛 (915)으로부터 검색되고 프로세서 (905)에 의한 액세스 준비를 위해 메모리 (910)에 저장될 수 있다. 일부 경우에서, 전자 저장 유닛 (915)은 배제 될 수 있고, 기계 실행가능 명령은 메모리 (910)에 저장된다. 대안적으로, 상기 코드는 제 2 컴퓨터 시스템(940)에서 실행될 수 있다.
- [0293] 서버(901)와 같이 여기에 제공된 시스템 및 방법의 측면들은 프로그래밍으로 구현될 수 있다. 기술의 다양한 양상들은 일반적으로 기계 (또는 프로세서) 실행가능 코드 및/또는 기계 판독 가능 매체 유형(가령, 컴퓨터 판독 가능한 매체)에 포함되거나 또는 내장되는 관련 데이터 형태의 "제품" 또는 "제조 품목"으로 간주될 수 있다. 기계 실행가능 코드는 메모리 (가령, 읽기 전용 메모리, 랜덤 액세스 메모리, 및 플래시 메모리) 또는 하드 디스크와 같은 전자 저장 장치에 저장 될 수 있다. "저장" 유형 매체는, 컴퓨터, 프로세서 등의 유형 메모리 중 일부 또는 전부, 또는 이와 관련된 모듈들, 가령, 소프트웨어 프로그래밍을 위해 언제든지 비-일시적 저장을 제공할 수 있는, 다양한 반도체 메모리, 테이프 드라이브, 디스크 드라이브 등을 포함 할 수 있다. 소프트웨어의 전부 또는 일부는 때때로 인터넷 또는 다양한 기타 통신 네트워크를 통해 통신 될 수 있다. 예를 들어, 이러한 통신은 하나의 컴퓨터 또는 프로세서에서 다른 컴퓨터로, 예를 들어, 관리 서버 또는 호스트 컴퓨터에서 애플리케이션 서버의 컴퓨터 플랫폼으로 소프트웨어를 로드하도록 할 수 있다. 따라서 소프트웨어 요소들이 탑재될 수 있는 또 다른 유형의 매체는 유선 및 광 유선 네트워크를 통해 그리고 다양한 무선 링크를 통해 로컬 장치 간의 물리적 인터페이스를 통해 사용되는 것과 같은 광, 전기 및 전자기파를 포함한다. 유선 또는 무선 링크, 광 링크 등과 같이 이러한 파동을 전달하는 물리적 요소도 소프트웨어를 탑재한 매체로 간주 될 수 있다. 본 명세서에서 사용되는, 컴퓨터 또는 기계 "판독가능 매체"와 같은 용어는, 비 일시적인 유형의 "저장" 매체에 제한되지 않는 한, 실행을 위해 프로세서에 명령을 제공하는데 참여하는 임의의 매체를 지칭한다.
- [0294] 따라서, 컴퓨터 실행 코드와 같은 기계 판독가능 매체는 유형 저장 매체, 반송파 매체 또는 물리적 전송 매체를 비롯한 (그러나 이에 제한되지 않음) 많은 형태를 취할 수 있다. 비-휘발성 저장 매체에는 예를 들어, 상기 시스템을 구현하는데 사용될 수 있는 임의의 컴퓨터(들) 등 내의 임의의 저장 장치와 같은 광학 또는 자기 디스크가 내포될 수 있다. 유형의 전송 매체에는 동축 케이블; 컴퓨터 시스템 내 버스를 구성하는 와이어들을 비롯한 구리 와이어 및 광섬유가 포함된다. 반송파 전송 매체는 전기 또는 전자기 신호, 또는 무선 주파수 (RF) 및 적외선 (IR) 데이터 통신 중에 생성되는 것과 같은 음향 또는 광파의 형태를 취할 수 있다. 따라서 컴퓨터 판독 가능 매체의 일반적인 형태에는 예를 들어 다음이 포함된다: 플로피 디스크, 플렉서블 디스크, 하드 디스크, 자기 테이프(tape), 기타 자기 매체, CD-ROM, DVD 또는 DVD-ROM, 기타 광학 매체, 펀치 카드 용지 테이프, 구멍 패턴이 있는 기타 물리적 저장 매체, RAM, ROM, PROM 및 EPROM, FLASH-EPROM, 기타 메모리 칩 또는 카트리지, 데이터 또는 명령어를 전송하는 반송파, 이러한 반송파를 전송하는 케이블 또는 링크, 또는 컴퓨터가 프로그래밍 코드 및/또는 데이터를 판독할 수 있는 기타 매체. 이러한 형태의 컴퓨터 판독가능 매체 중 다수는 실행을 위해 하나 이상의 명령의 하나 이상의 시퀀스를 프로세서에 전달하는 것과 관련 될 수 있다.
- [0295] 분석 결과는 그래픽 사용자 인터페이스와 같은 사용자 인터페이스의 도움으로 사용자에게 표시될 수 있다.
- [0296] 본 발명은 하기 실시예에서 추가로 기술될 것이지만, 청구 범위에 기재된 본 발명의 범위를 제한하지 않는다.
- [0297] **실시예**
- [0298] **실시예 1: 듀플렉스 고정된 PCR**
- [0299] **재료 및 방법**
- [0300] *듀플렉스 고정된 PCR 라이브러리 준비물*
- [0301] 이 프로토콜에 따라, Swift Accel-NGS 2S PCR-Free Library Kit(Cat. # 20024 및 20096)와 특정 절두된 어댑터 및 프라이머를 사용하여 이중 라이브러리를 준비할 수 있다. 일부 경우에서, Illumina 기기에서 시퀀싱하기 위

해, 별도의 PCR을 통해 전장 P5 및 P7 접목 서열을 라이브러리에 추가할 수 있다.

[0302] 이 프로토콜은 PCR 튜브용이지만, PCR 플레이트로 확장할 수 있다.

[0303] 재료:

[0304] 1. Swift Accel-NGS 2S PCR-Free Library Kit(Cat. # 20024 및 20096)

[0305] 2. 3' Swift N14 어댑터 1 v3A

[0306] a. /5Phos/ATAAACGACGGCNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAG*T*C/3SpC3/ (서열 식별 번호: 1)

[0307] b. Order PAGE, IDT에서 정제된, 1 μ mole 합성 규모, 동결건조

[0308] c. /3SpC3/는 /3Phos/로 치환될 수 있고, 포스포로티오에이트 링키지는 필요하지 않으며, 올리고는 HPLC 정제될 수 있음.

[0309] 3. 3' Swift 어댑터 2 v3'dT

[0310] a. GCCGUCGUUUUA/33dT/ (서열 식별 번호: 3)

[0311] b. /33dT/는 3'-deoxyT에 대한 IDT 오프-카탈로그 변형이다.

[0312] c. Order PAGE, IDT에서 정제된, 1 μ mole 합성 규모, 동결건조

[0313] 4. 5' Swift 어댑터

[0314] a. /5SpC3/A*C*ACTCTTTCCTACACGACGCTCTCCGATCT (서열 식별 번호: 4)

[0315] b. Order PAGE, IDT에서 정제된, 1 μ mole 합성 규모, 동결건조

[0316] c. /5SpC3/ 및 포스포로티오에이트 연결이 필요하지 않으며, 올리고는 HPLC 정제되어야 한다.

[0317] d. 2S 이중 인덱싱 키트(Cat. No. 28096)의 truB2 시약으로 대체 가능.

[0318] 5. NEB Ultra II Q5 마스터 믹스 (Cat. No. M0544L)

[0319] 6. 절두된 P5 프라이머

[0320] a. ACACCTTTTCCTACACGACGCTCTCCGATCT (서열 식별 번호: 4)

[0321] b. 변형 필요 없음, IDT에서 탈염, IDTE에서 100 μ M.

[0322] 7. 절두된 P7 프라이머

[0323] a. GACTGGAGTTCAGACGTGTGCTCTCCGATCT (서열 식별 번호: 5)

[0324] b. 변형 필요 없음, IDT에서 탈염, IDTE에서 100 μ M.

[0325] 8. SPRIselect 비드 (Beckman Coulter, Cat. No. B23317/B23318/B23319)

[0326] 9. 80% EtOH (샘플 당 ~2 mL)

[0327] 10. PCR 튜브 스트립 (가령, GeneMate VWR Cat. No. 490003-710)

[0328] 11. 자성 랙 (가령, Permagen MSRLV08)

[0329] 12. USER Enzyme (NEB Cat. No. M5505L) - 이것은 우라실-DNA 글리코실화효소와 DNA 글리코실화효소-리아제 엔도뉴클레아제 VIII의 혼합물이다.

[0330] 사용자 지정 어댑터 준비(대규모 배치의 경우, 한 번 수행 가능):

[0331] 1. Swift의 truB2 시약을 사용하지 않는 경우, Low EDTA TE(Swift 2S 키트에 내포됨)에서 5' Swift 어댑터를 42 μ M으로 재현탁시킴.

[0332] 2. Low EDTA TE(Swift 2S 키트에 포함)에서 3' Swift N14 어댑터 1 v3A를 100 μ M로 재현탁시킴. 후속 사용을 위해 -20°C에서 보관.

[0333] 3. Low EDTA TE(Swift 2S 키트에 포함)에서 3' Swift 어댑터 2 v3'dT를 100 μ M로 재현탁시킴. 후속 사용을

위해 -20℃에서 보관.

[0334] 4. 실온에서 100 μl의 각 올리고를 혼합하여, 3' Swift N14 어댑터 1 v3A에서 3' Swift 어댑터 2 v3'dT를 어닐링한다. 튜브를 3' Swift N14 v3'dT 듀플렉스 어댑터, 50 μM로 라벨링한다. 상기 3' 듀플렉스 어댑터의 최종 농도는 50 μM이다. 사용하기 전에 실온에서 적어도 5분 동안 항온처리한다. 후속 사용을 위해 -20℃에서 보관.

[0335] 기술적 명기:

[0336] 효소 튜브는 -20℃ 보관에서 꺼내어, 피펫팅 전, 효소가 4℃에 도달하도록 약 10분 동안 얼음 위에 둘 수 있다. -20℃에서 효소를 피펫팅하면 효소 시약이 부족할 수 있다.

[0337] 시약을 4℃로 해동한 후, 시약(효소 제외)을 잠시 볼텍싱하여 잘 혼합시킬 수 있다. 개방 전, 내용물을 수집하기 위해 마이크로원자의 모든 튜브를 회전시킨다.

[0338] 피펫팅 손실을 보상하기 위해, 5% 초과 볼륨을 사용하여 얼음과 스케일 볼륨에 모든 시약 마스터 믹스를 어셈블리시킨다.

[0339] 프로토콜 전체에 명시된 대로 시약을 지정된 순서로 마스터 믹스에 추가해야 한다.

[0340] 시약들은 미리 준비할 수 있다(가령, 크기 선택 단계에서 자기 비드가 마르지 않도록 하기 위해).

[0341] 단계 1: 템플릿 복구

[0342] 1. 11ng의 cfDNA 샘플을 0.2mL PCR 튜브로 옮기고, 필요한 경우 Low EDTA TE를 사용하여 샘플의 부피를 37 μl의 최종 부피로 조정한다.

[0343] 2. 각 샘플에 3 μl의 USER Enzyme을 추가한다.

[0344] 3. 볼텍싱하여 혼합하고, 부드럽게 회전시켜 모든 액체를 튜브 바닥으로 모은다.

[0345] 4. 37℃에서 15분 동안 프로그래밍된 열순환기에 샘플을 두고, 뚜껑을 열을 이용하여 연다.

[0346] 단계 2: 말단 복구 1

[0347] 1. 임의의 응결을 수집하기 위해, 샘플을 부드럽게 회전시킨다.

[0348] 2. 40 μl DNA 샘플을 함유하는 각 샘플에 사전-혼합된 복구 I 마스터 믹스 20 μl(표 1 참조)을 추가한다.

[0349] 표 1. 단부 복구 I 마스터 믹스

시약	샘플당 용적
Low EDTA TE	13 μl
완충액 W1	6 μl
효소 W2	1 μl
총 용적	20 μl

[0350]

[0351] 3. 볼텍싱하여 혼합하고, 부드럽게 회전시켜 열순환기에 넣고, 아래 설명된 순서대로 복구 I 열순환기 프로그램을 실행한다.

[0352] a. 37℃, 5 분, 뚜껑 가열하여 열기(뚜껑은 75℃로 설정)

[0353] b. 65℃, 2 분, 뚜껑 가열하여 열기(뚜껑은 75℃로 설정)

[0354] c. 37℃, 5 분, 뚜껑 가열하여 열기(뚜껑은 75℃로 설정)

[0355] 4. 열순환기 프로그램이 완료된 후, 튜브를 부드럽게 회전시켜 응결물을 수집한다.

[0356] 5. SPRIselect 비드 120 μl(2.0X)을 추가하여, 복구 I 반응을 정리한다. 볼텍싱하여 혼합한다. 부드럽게 회전시켜 비드를 수집하고, 실온에서 5분 동안 항온처리한다.

[0357] 6. 5분 동안 마그네틱 랙에 샘플을 놓아 비드를 수집한다.

[0358] 7. 펠릿을 흘트리지지 않고, 상층액을 제거하고 폐기한다.

[0359] 8. 마그네틱 랙에 있는 동안 샘플에 새로 준비된 80% 에탄올 용액 180 μ l를 추가한다. 펠릿을 흘트리지지 않도록 주의한다. 30초 동안 항온처리한 다음, P20 피펫으로 에탄올 용액을 조심스럽게 제거한다.

[0360] 9. 80% 에탄올 용액으로 두 번째 세척을 위해 위의 단계를 반복한다.

[0361] 10. P20 피펫으로 잔류 에탄올 용액을 제거하고, 비드를 ~30초 동안 건조시킨다. 비드가 마르지 않도록 주의하고, 즉시 단부 복구 2의 1단계로 진행한다.

[0362] 단계 3: 단부 복구 2

[0363] 1. 각 샘플에 대한 비드에 사전-혼합된 복구 II 마스터 믹스(표 2 참조) 50 μ l를 추가하고 균질해질 때까지 볼텍싱하여 혼합한다.

[0364] 표 2. 단부 복구 II 마스터 믹스

시약	샘플당 용적
Low EDTA TE	30 μ l
완충액 G1	5 μ l
시약 G2	13 μ l
효소 G3	1 μ l
효소 G4	1 μ l
총 용적	50 μ l

[0365]

[0366] 2. 20℃에서 20분 동안 프로그래밍된 열순환기에 샘플을 두고, 뚜껑을 열을 이용하여 연다.

[0367] 3. 열순환기 프로그램이 완료된 후, 튜브를 부드럽게 회전시켜 응결물을 수집한다.

[0368] 4. 90 μ l(1.8X)의 PEG/NaCl 용액을 추가하여, 복구 2 반응을 정리한다. 볼텍싱하여 혼합한다. 부드럽게 회전시켜 비드를 수집하고, 실온에서 5분 동안 항온처리한다.

[0369] 5. 5분 동안 마그네틱 랙에 샘플을 놓아 비드를 수집한다.

[0370] 6. 펠릿을 흘트리지지 않고, 상층액을 제거하고 폐기한다.

[0371] 7. 마그네틱 랙에 있는 동안 샘플에 새로 준비된 80% 에탄올 용액 180 μ l를 추가한다. 펠릿을 흘트리지지 않도록 주의한다. 30초 동안 항온처리한 다음, P20 피펫으로 에탄올 용액을 조심스럽게 제거한다.

[0372] 8. 80% 에탄올 용액으로 두 번째 세척을 위해 위의 단계를 반복한다.

[0373] 9. P20 피펫으로 잔류 에탄올 용액을 제거하고, 비드를 ~30초 동안 건조시킨다. 비드가 과도하게 건조되지 않도록 주의하고, 즉시 결찰 1의 1단계로 진행한다.

[0374] 단계 4: 결찰 1

[0375] 1. 각 샘플에 대한 비드에 사전-혼합된 복구 II 마스터 믹스(표 3 참조) 30 μ l를 추가하고 균질해질 때까지 볼텍싱하여 혼합한다.

표 3. 결찰 I 마스터 믹스

시약	샘플당 용적
Low EDTA TE	20 μ l
3' Swift N14 v3'dT 듀플렉스 어댑터, 50 μ M	5 μ l
완충액 Y1	3 μ l
효소 Y3	2 μ l
총 용적	30 μ l

2. 25℃에서 15분 동안 프로그래밍된 열순환기에 샘플을 두고, 뚜껑을 열을 이용하여 연다.
3. 열순환기 프로그램이 완료된 후, 튜브를 부드럽게 회전시켜 응결물을 수집한다.
4. 36 μ l (1.8X)의 PEG/NaCl 용액을 추가하여, 결찰 1 반응을 정리한다. 볼텍싱하여 혼합한다. 부드럽게 회전시켜 비드를 수집하고, 실온에서 5분 동안 항온처리한다.
5. 5분 동안 마그네틱 랙에 샘플을 놓아 비드를 수집한다.
6. 펠릿을 흘트리지지 않고, 상층액을 제거하고 폐기한다.
7. 마그네틱 랙에 있는 동안 샘플에 새로 준비된 80% 에탄올 용액 180 μ l를 추가한다. 펠릿을 흘트리지지 않도록 주의한다. 30초 동안 항온처리한 다음, P20 피펫으로 에탄올 용액을 조심스럽게 제거한다.
8. 80% 에탄올 용액으로 두 번째 세척을 위해 위의 단계를 반복한다.
9. P20 피펫으로 잔류 에탄올 용액을 제거하고, 비드를 ~30초 동안 건조시킨다. 비드가 과도하게 건조되지 않도록 주의하고, 즉시 결찰 2의 1단계로 진행한다.

단계 5: 결찰 2

1. 각 샘플에 대한 비드에 사전-혼합된 결찰 II 마스터 믹스(표 4 참조) 50 μ l를 추가하고 균질해질 때까지 볼텍싱하여 혼합한다.

표 4. 결찰 II 마스터 믹스

시약	샘플당 용적
Low EDTA TE	30 μ l
시약 B1	5 μ l
5' Swift 어댑터, 42 μ M 또는 시약 truB2	2 μ l
시약 B3	9 μ l
효소 B4	1 μ l
시약 B5	2 μ l
시약 B6	1 μ l
총 용적	50 μ l

2. 40℃에서 10분 동안 프로그래밍된 열순환기에 샘플을 두고, 뚜껑을 열을 이용하여 연다.
3. 열순환기 프로그램이 완료된 후, 튜브를 부드럽게 회전시켜 응결물을 수집한다.

4. 52.5 μ l(1.8X)의 PEG/NaCl 용액을 추가하여, 결찰 1 반응을 정리한다. 볼텍싱하여 혼합한다. 부드럽게 회전시켜 비드를 수집하고, 실온에서 5분 동안 항온처리한다.

5. 5분 동안 마그네틱 랙에 샘플을 놓아 비드를 수집한다.

6. 펠릿을 흘트리지지 않고, 상층액을 제거하고 폐기한다.

7. 마그네틱 랙에 있는 동안 샘플에 새로 준비된 80% 에탄올 용액 180 μ l를 추가한다. 펠릿을 흘트리지지 않도록 주의한다. 30초 동안 항온처리한 다음, P20 피펫으로 에탄올 용액을 조심스럽게 제거한다.

8. 80% 에탄올 용액으로 두 번째 세척을 위해 위의 단계를 반복한다.

9. P20 피펫으로 잔류 에탄올 용액을 제거하고, 비드를 ~30초 동안 건조시킨다. 비드를 과도하게 건조시키지 않도록 주의하고 즉시 Low EDTA TE 24 μ l에 비드를 재현탁시킨다. 볼텍싱하여 혼합하고, 2분 동안 항온처리한다.

10. 부드럽게 회전시켜 비드를 수집하고, 2분 동안 마그네틱 랙에 비드를 수집한다.

단계 6: PCR-라이브러리 증폭

1. 각 샘플에 대해 깨끗한 튜브에 미리 혼합된 PCR-라이브러리 증폭 마스터 믹스(표 5 참조) 26 μ l를 추가한다.

표 5. PCR-라이브러리 증폭 마스터 믹스

시약	샘플당 용적
NEB Ultra II Q5 마스터 믹스 (2x)	25 μ l
절두된 P5 프라이머, 100 μ M	0.5 μ l
절두된 P7 프라이머, 100 μ M	0.5 μ l
총 용적	26 μ l

2. 최종 결찰-후 라이브러리를 함유하는 상층액을 PCR-라이브러리 증폭 마스터 믹스로 조심스럽게 옮긴다.

3. P20 피펫을 사용하여, 임의의 잔여분 최종 결찰-후 라이브러리를 옮긴다. 가능한 한 많은 상층액을 옮기도록 한다.

4. 볼텍싱하여 혼합하고, 부드럽게 회전시켜 열순환기에 넣고, 아래 설명된 순서대로 PCR-라이브러리 증폭 프로그램을 실행한다.

표 6. 예시적인 PCR-라이브러리 증폭 열순환기 프로그램

단계	온도	시간
초기 변성	98 °C	30 초
5 PCR 주기 (4 GOTO 단계들)	98 °C	10 초
	65 °C	75 초
최종 연장	65 °C	5 분
대기	4 °C	대기

5. SPRIselect 비드 90 μ l(1.8X)을 추가하여, PCR-라이브러리 증폭 반응을 정리한다. 볼텍싱하여 혼합한다. 부드럽게 회전시켜 비드를 수집하고, 실온에서 5분 동안 항온처리한다.

6. 5분 동안 마그네틱 랙에 샘플을 놓아 비드를 수집한다.

7. 펠릿을 흘트리지지 않고, 상층액을 제거하고 폐기한다.

- [0411] 8. 마그네틱 랙에 있는 동안 샘플에 새로 준비된 80% 에탄올 용액 180 μ l를 추가한다. 펠릿을 홀트리지지 않도록 주의한다. 30초 동안 항온처리한 다음, P20 피펫으로 에탄올 용액을 조심스럽게 제거한다.
- [0412] 9. 80% 에탄올 용액으로 두 번째 세척을 위해 위의 단계를 반복한다.
- [0413] 10. P20 피펫으로 잔류 에탄올 용액을 제거하고, 비드를 ~30초 동안 건조시킨다. 비드를 과도하게 건조시키지 않도록 주의하고 즉시 Low EDTA TE의 47 μ l에 비드를 재현탁시킨다. 볼텍싱하여 혼합하고, 2분 동안 항온 처리한다.
- [0414] 11. 부드럽게 회전시켜 비드를 수집하고, 2분 동안 마그네틱 랙에 비드를 수집한다.
- [0415] 12. 최종 PCR-증폭된 라이브러리를 함유하는 상층액을 임의의 비드를 나르지 않도록 조심하며 깨끗한 튜브로 옮긴다.
- [0416] 13. 1 μ l의 상기 증폭된 라이브러리를 TapeStation으로 분석한다. 눈에 띄는 피크는 어댑터-결찰된 단핵구 DNA(180bp + 60bp + 59bp)에 상응하는 ~300bp에 있어야 한다.
- [0417] 14. -20 °C에서 라이브러리를 보관한다.
- [0418] **듀플렉스 고정된 PCR을 이용하여 회귀 돌연변이의 정확하고, 효과적인 검출.**
- [0419] 듀플렉스 분자 바코드가 통합된 시퀀싱 라이브러리는 이중 가닥으로 된 유입 DNA에 두 개 어댑터 분자를 연속적으로 결합시킴으로써 생성되었다. 제 1, 유입 DNA는 블런트화 및 탈인산화 반응을 통하여 단부-복구되었다 (도 9 및 도 10). 단부-복구 후, 차단된 3' 그룹 (3' oligo #2)을 갖는 짧은 올리고뉴클레오타이드에 어닐링된 5' 포스페이트 (3' oligo #1)를 함유하는 3' 어댑터를 상기 유입 DNA의 각 3' 단부에 결합시켰다(도 12). 상기 올리고뉴클레오타이드중 하나는 3' 차단 그룹을 함유하고 있기 때문에, 오로지 5' 포스페이트 (3' oligo #1)를 함유하는 올리고뉴클레오타이드만 상기 유입 DNA의 3' 단부에 공유적으로 부착된다. 상기 부착된 3' 올리고뉴클레오타이드는 독특하게 각 가닥을 라벨링시키는 분자 바코드를 또한 함유한다 (도 11). 그 다음, 3' 차단 그룹을 함유하는 3' oligo는 분해되고, 5' 어댑터 oligo는 니크 해독-유사(nick-translation-like) 반응을 통하여 각 5' 단부에 결합시켰다. 특이적으로, 상기 5' 어댑터는 3' 어댑터 oligo #1 상의 분자 바코드 바로 상류에 갭을 남기면서 어닐링된다. 이 갭은 니크 해독-유사(nick-translation-like) 반응 동안 채워지고, 밀봉되어, DNA 단편의 각 단부에 제자리(in-situ)에서 이중 분자 바코드를 생성한다 (도 13). 생성된 결합된 산물을 정제하고, 초기 온전환-게놈 PCR을 통하여 증폭시켰다 (도 14).
- [0420] 상기 초기 온전환-게놈 PCR 후, 해당 산물을 임의선택적으로 정제시키고, 센스 가닥과 안티센스 가닥에 상응하는 단일 가닥으로 된 (ss) NA 라이브러리가 생성되었다 (도 2 및 도 3).
- [0421] 가닥-특이적 고정된 PCR 방식을 이용하여 원하는 표적에 대해 상기 증폭된 DNA 라이브러리를 풍부화시켰다. 이러한 PCR 풍부화는 원하는 관심대상 영역을 표적으로 하는 단일 프라이머와 상기 결합된 어댑터 서열을 표적으로 하는 제 2 프라이머를 이용하였다 (도 4, 5, 15, 17). 표적 풍부화의 특이성을 증가시키기 위해, 제 2 네스티드 PCR은 원하는 관심대상 영역을 표적으로 하는 단일 프라이머와 상기 결합된 어댑터 서열을 표적으로 하는 제 2 프라이머를 이용하여 수행될 수 있다 (도 7, 8, 16, 18). 표적 풍부화의 특이성을 증가시키기 위해, 제 2 네스티드 PCR을 이용하여 샘플 바코드와 차세대 시퀀싱에 필요한 필수 접목 서열을 통합할 수 있다. 그런 다음, 생성된 라이브러리를 정량화하고, 정규화하고, 시퀀싱한다.
- [0422] 시퀀싱 후, 판독값을 게놈에 정렬하고, 분자 바코드로 그룹화했다. 표적의 센스 및 안티센스 가닥 모두에 매핑되는 동일한 분자 바코드를 갖는 판독을 함유하는 단편은 "듀플렉스 서포트(duplex support)"를 갖도록 설계되었다. 돌연변이는 두 가닥 모두에 존재하는 경우에만 점수를 매겼다. (도 20 및 21).
- [0423] **실시예 2: DNA의 Watson 가닥과 Crick 가닥의 표적화된 DNA 시퀀싱**
- [0424] 회귀 핵산 서열의 식별 및 정량화는 생물학 및 임상 의학의 많은 영역에서 중요하다. 본 실시예에서는 (i) 템플릿 분자의 Watson 및 Crick 가닥에 동일한 분자 바코드를 효율적으로 도입하고, (ii) 새로운 가닥 특이적 PCR 분석으로 관심 게놈 영역을 강화하여 이 문제를 해결하는 방법(SaferSeq라고 함)을 설명한다. 단일 앰플리콘 내에서 또는 동시에 여러 앰플리콘 내에서 돌연변이를 평가하는 데 적용할 수 있으며, 혈장에서 발견되는 것과 같은 제한된 양의 DNA를 평가할 수 있으며, 기존 PCR-기반 분자 바코드 접근 방식의 오류율을 적어도 2배 감소시킬 수 있다.

[0425] **결과**

[0426] 일반적으로 라이브러리 구축과 관련된 비효율성과 도입된 오류를 해결하기 위해, 어댑터 서열을 3' 및 5' DNA 단편 단부에 순차적으로 연결하고, 제자리에서 이중 가닥으로 된 분자 바코드 생성을 포함하는 전략이 설계되었다 (도 22a). 분자 바코드의 제자리 생성은 새로운 라이브러리 준비 방법의 핵심 혁신이다. 이중 가닥 분자 바코드의 제자리 생성에 사용되는 효소는 각 DNA 단편을 고유하게 바코드하고, 이중 어댑터를 효소적으로 준비할 필요가 없다 (도 22a, 단계 2 및 단계 3). 이 어댑터는 상기 외인성 분자 바코드로 14개 무작위 뉴클레오타이드 스트레치를 함유하였다 (독특한 식별자 서열 [UID]). 이 어댑터-결찰된 단편들은 두 개의 원래 DNA 가닥의 중복 사본을 생성하기 위해 제한된 수의 PCR 주기를 겪었다 (도 22a, 단계 4). 명확성을 위해, 이 예시적인 구체예에서 UCSC 참조 서열(genome.ucsc.edu/에서 이용가능)은 "Watson" 가닥으로 임의로 정의되었고, 이의 역 상보체는 "Crick" 가닥으로 정의되었다.

[0427] 이 프로토콜의 또 다른 혁신은 풍부화를 위한 절반-네스티드 PCR 기반 접근 방식을 사용하는 것이다. 절반-네스티드 PCR이 이전에 표적 풍부화에 사용되었지만(가령, Zheng et al., 2014, Nat Med 20:1479-1484 참조), 이를 이중 시퀀싱에 적용하려면 주요 변경이 필요했다. 구체적으로, Watson 가닥에 대한 것과 Crick 가닥에 대한 두 가지 개별 PCR이 수행되었다. 두 가지 PCR은 모두 동일한 유전자-특이적 프라이머를 사용했지만, 각각 다른 고정 프라이머를 사용했다. 각 가닥에서 파생된 PCR 복제물은 외인성 UID에 대한 삽입물의방향으로 구별할 수 있다(도 22b).

[0428] 시퀀싱 후, 원래 DNA 듀플렉스의 각 가닥에 해당하는 판독값을 Watson 패밀리와 Crick 패밀리로 그룹화시켰다. 각 패밀리 구성원은 초기 템플릿 단편의 한쪽 단부에 있는 서열을 나타내는 동일한 내인성 바코드와 라이브러리 구축 중에 제자리에 도입된 동일한 외인성 UID를 보유하였다. Watson 가닥 패밀리의 >80%에 존재하는 돌연변이를 "Watson 슈퍼뮤턴트(supermutants)"라고 한다. Crick 가닥 패밀리의 >80%에 존재하는 돌연변이를 "Crick 슈퍼뮤턴트(supermutants)"라고 한다. 동일한 UID를 가진 Watson 패밀리 및 Crick 패밀리 ("듀플렉스 패밀리")의 >80%에 존재하는 것들은 "슈퍼칼리파질주의적(supercalifragilisticexpialidocious) 뮤턴트"라고 하며, 여기에서는 "슈퍼칼리뮤턴트"라고 한다(도 22c).

[0429] SaferSeqS의 초기 시연으로써, 공지의 돌연변이를 갖는 DNA를 10%에서 0%까지 다양한 비율로 정상 개인의 백혈구 DNA에 스파이킹시키는 혼합 실험이 수행되었다. 이러한 혼합물은 검정 당 15,400, 150, 15, 15, 8 또는 0개의 슈퍼칼리뮤턴트를 생성할 것으로 예측되었다. 표적 판독 (즉, 의도된 앰플리콘으로 구성된 판독)의 비율은 88%이었고, 하이브리드 캡처- 기반 접근 방식으로 달성할 수 있는 것보다 훨씬 높다 (가령, Samorodnitsky et al., 2015 *Hum Mutat* 36:903-914 참고). 더욱이, 예상되는 대립유전자 빈도와 관찰된 대립유전자 빈도 사이의 강한 상관관계는 5 자릿수(orders of magnitude)에 걸쳐 입증되었다. (도 23, Pearson's $r > 0.999$, $p = 2.02 \times 10^{-12}$). 사전-명시된 혼합 변이체에 상응하는 단일 돌연변이체는 정상 개체의 DNA에서 관찰되지 않았으며, 이는 관심 돌연변이에 대한 매우 높은 특이성을 나타낸다. 쿼리된(queried) 염기 뿐만이 아니라, 앰플리콘 내의 임의의 염기에 대해서도 특이성이 결정되었다. 모든 DNA 샘플 중에서 쿼리된 총 37,747,670개의 염기에 걸쳐 6개의 슈퍼칼리뮤턴트만 관찰되었으며, 이는 1.59×10^{-7} 슈퍼칼리뮤턴트/bp의 돌연변이 빈도를 나타낸다 (표 7).

표 7. 분석적 민감도 및 특이성 검증 실험에서 확인된 돌연변이.

*좌표는 인간 참조 게놈 hg19 릴리스를 나타낸다 (Genome Reference Consortium GRCh37, Feb 2009).

샘플	염색체	위치*	참조 염기	변이체 염기	유전자	cDNA 변화	단백질 변화	# 슈퍼칼리뮤턴트	듀플렉스 깊이	MAF
10% Spike-in	chr17	7579507	T	G	TP53	c.180A>C	p.P60P	3,067	37,686	8.14E+00
1% Spike-in	chr17	7579507	T	G	TP53	c.180A>C	p.P60P	280	34,997	8.00E-01
1% Spike-in	chr17	7579560	A	G	TP53	c.127T>C	p.L43L	1	1,481	6.75E-02
0.1% Spike-in	chr17	7579507	T	G	TP53	c.180A>C	p.P60P	28	36,764	7.62E-02
0.01% Spike-in	chr17	7579507	T	G	TP53	c.180A>C	p.P60P	3	37,872	7.92E-03
0.001% Spike-in	chr17	7579507	T	G	TP53	c.180A>C	p.P60P	2	206,682	9.68E-04
0.001% Spike-in	chr17	7579518	C	G	TP53	c.169G>C	p.D57H	1	193,317	5.17E-04
0.001% Spike-in	chr17	7579547	G	A	TP53	c.140C>T	p.P47L	1	153,733	6.50E-04
0.001% Spike-in	chr17	7579575	G	-	TP53	c.112del	p.Q38del	1	37,951	2.63E-03
0.001% Spike-in	chr17	7579716	G	A	TP53	c.80C>T	p.P27L	1	14,207	7.04E-03
0% Spike-in	chr17	7579508	G	C	TP53	c.179C>G	p.P60R	1	200,754	4.98E-04

그런 다음, SaferSeqS가 DNA 양이 제한적인 임상 샘플에 적용될 수 있는지 여부를 결정하기 위해 구하였다. 예를 들면, 액체 생검에 사용되는 10mL의 무-세포 혈장 DNA 샘플에는 33ng의 적은 양의 DNA가 종종 존재한다. 이들 샘플에 있는 DNA 템플릿 분자의 대다수는 야생형이며, 낮은 중앙 부하를 가진 환자의 샘플에 존재하는 10,000개의 야생형 템플릿 중 하나 또는 두 개의 돌연변이 템플릿이 있다. 이 극도로 적은 수의 돌연변이 템플릿을 민감하게 검출하려면, 해당 분석이 시작 분자를 효율적으로 회수해야 한다.

이러한 어려운 상황에서 SaferSeqS를 평가하기 위해, 임상 샘플에서 일반적으로 관찰되는 돌연변이 빈도를 모방하기 위해, 암 환자의 무-세포 혈장 DNA를 정상 개인의 무-세포 혈장 DNA와 혼합했다. 이들 실험에서, 33ng의 각 샘플을 TP53의 3가지 상이한 돌연변이 중 하나에 대해 분석했다. 27개의 실험 조건(3개의 TP53 앰플리콘 × 3개의 샘플 × 3개의 분취량/샘플)에 걸친 표적 판독의 중앙값 비율은 80%였다 (범위: 72% ~ 91%) (도 24a). 듀플렉스 패밀리의 중앙값(즉, 동일한 내인성 바코드 및 외인성 바코드를 함유하는 Watson 가닥 및 Crick 가닥 모두)은 원래 템플릿 분자 수의 89% (범위: 65% ~ 102%)였다 (도 24b). 더욱이, 6개의 혼합된 샘플 모두에서, 관심 있는 슈퍼칼리뮤턴트는 예상 빈도에서 식별되었다 (도 25b,d,e, 표 9). 이러한 예상 빈도의 돌연변이는 이전에 설명한 분자 바코딩 방법을 사용하여 동일한 샘플에서도 또한 확인되었다 ("SaferSeqS"보다는 "SafeSeqS") (도 25a,b,c, 표 8). SaferSeqS의 장점은 이의 특이성이다. 이전에 설명한 방법으로 관찰된 153개의 별개 돌연변이를 나타내는 총 1,406개의 슈퍼뮤턴트가 있었으며, 이는 9.39×10^{-6} 슈퍼뮤턴트/bp의 평균

오류율을 반영한다 (도 25a,b,c, 표 8). 이러한 돌연변이의 대부분은 아마도 두 가닥 중 오직 하나의 가닥에서 초기 바코딩 주기 동안 발생한 중합효소 오류였을 것이다. 유사하게, 만약 슈퍼칼리뮤턴트 이외의 Watson 슈퍼뮤턴트 또는 Crick 슈퍼뮤턴트 (즉, 두 가닥중 오직 하나의 가닥에서만 관찰된 것들, 도 22c)만을 고려하면, 6.56×10^{-6} 슈퍼뮤턴트/bp의 오류율이 관찰되었다 (도 26, 표 9). 대조적으로, SaferSeqS로 쿼리된 4,947,725개의 염기중에서 오직 한 개의 슈퍼칼리뮤턴트가 검출되면, 전반적인 돌연변이율은 2.02×10^{-7} 을 나타낸다(표 9). SaferSeqS와 이전에 설명한 분자 바코딩 방법 간의 이러한 특이성 차이 (즉, 시퀀싱 전에 분자 바코드를 통합하기 위해 직접 PCR 또는 어댑터 결찰을 사용하는 것)는 매우 유의미하다 ($P < 3.5 \times 10^{-10}$, SaferSeqS를 다른 각각의 방법과 비교하는 비율에 대한 양측-Z 검정)

표 8. SafeSeqS 및 SaferSeqS에 의해 확인된 돌연변이 비교

(별첨 A 참고)

표 9. 가닥-에그노스틱 분자 바코드 및 SaferSeqS에 의해 확인된 돌연변이 비교

(별첨 B 참고)

SaferSeqS의 임상 적용 가능성을 추가로 입증하기 위해 중앙 부하이 최소화된 5명의 암 환자를 평가했다. 각각의 경우에, 도처에서 설명한 바와 같이, 원발성 종양(혈장보다는)의 돌연변이가 확인되었다. (Tie et al., *Sci Transl Med* 8:346ra392 (2016)). 이들 환자의 혈장을 두 개의 동일한 분취량으로 나누고, 도처에서 기술된 바와 같은 분자 바코딩 방법으로 한 개 분취량을 평가했고(Kinde et al., *Proc Natl Acad Sci U S A* 108: 9530-9535 (2011)), 다른 분취량은 SaferSeqS로 평가했다. 두 경우에서, 모두 관심 있는 돌연변이를 표적으로 하는 작은 애플리콘을 생성하는 프라이머가 설계되었다. 이전에 설명된 바코딩 방법으로 평가한 결과, 혈장 샘플에는 원래 원발성 종양에서 확인된 8개의 돌연변이가 응집되어 있음이 밝혀졌다. 혈장에서 이러한 돌연변이들의 빈도는 0.01%에서 0.1%까지 다양했다. (도 27, 표 10). 8개의 공지된 돌연변이 외에도, 이전에 설명된 방법으로 최대 0.013%의 빈도로 존재하는 334개의 별개의 돌연변이를 확인했고, 이들 중 어느 것도 이들 환자의 원발성 종양에서 발견되지 않았다. 이들 334개의 돌연변이는 10,347개의 슈퍼뮤턴트로 구성되었으며, 이는 1.23×10^{-5} 슈퍼뮤턴트/bp의 평균 오류율을 반영한다 (도 27, 표 10). SaferSeqS를 사용하여, 원발성 종양에서 발견된 8개의 돌연변이가 이전에 설명된 방법에서 발견된 것과 유사한 빈도로 5명의 환자 모두에서 검출되었다 (도 27, 표 10). 그러나, 8,707,755개의 쿼리된 염기 중에서, SaferSeqS로 1개의 추가 슈퍼칼리,뮤턴트(334개의 돌연변이가 아닌)만이 확인되었으며 평균 오류율은 1.15×10^{-7} 이었다 (표 10). 이전에 설명된 분자 바코딩 방법에 비해, 특이성이 >100-배 이상 개선된 것은 상당히 유의미적이었다 ($P < 2.2 \times 10^{-16}$, 비율에 대한 양측 Z 검정).

표 10. 암 환자로부터 얻은 혈장 샘플에서 SafeSeqS 및 SaferSeqS로 확인된 돌연변이.

(별첨 C 참고)

다음으로, SaferSeqS가 다양한 시퀀싱 애플리케이션에 유용할 수 있는 여러 표적을 동시에 분석할 수 있는지 여부를 조사했다. SaferSeqS는 두 가지 유형의 다중화를 허용하는데, 하나는 별도의 PCR 반응에서 다중 표적을 분석하고, 다른 하나는 동일한 PCR 반응에서 다중 표적을 분석한다. 종종 Watson 가닥과 Crick 가닥-끌어낸 복사본이 라이브러리 증폭 중에 생성되기 때문에, 이 라이브러리는 샘플 복구에 부정적인 영향을 미치지 않고, 여러 PCR 반응으로 분할될 수 있다. 예를 들면, PCR 효율이 70%라고 가정하면, DNA 라이브러리가 11회 PCR 사이클로 증폭되는 경우, 10% 미만의 복구 손실로 최대 22개의 표적을 개별적으로 분석할 수 있다 (도 28). 실제로 라이브러리의 100% 또는 4.4%가 분석되었다. 표적 적중률은 라이브러리의 4.4%를 100% 사용하든지 간에 유사했으며, 판독의 82%와 92%는 의도한 영역에 적절하게 매핑되었다. 복구된 듀플렉스 패밀리의 수도 유사하여, 100% 및 4.4% 라이브러리 분할에서 7,825 및 6,769가 복구되었다.

위에서 설명한 다중화 접근 방식은 제한된 수의 표적을 동시에 분석하는 데 유용하지만, 많은 게놈 영역을 평가하는 응용 프로그램에는 소수의 PCR 반응으로 다중화하는 것이 내포될 수 있다. 이러한 맥락에서 SaferSeqS의 다중화 기능을 평가하기 위해, 48개의 프라이머가 암에서 일반적으로 돌연변이되는 드라이버 유전자 영역을 쿼리하도록 설계되었다 (표 11). 이들 프라이머는 2개의 반응으로 결합되었다: 하나는 25개를 표적으로 하고, 다른 하나는 23개 영역을 표적으로 한다. 48개의 프라이머 쌍 각각은 의도한 표적을 특이적으로 증폭시켰고 (도 30), 이중 36개는 듀플렉스 패밀리의 수가 단일 반응으로 확인된 것의 50% 이상인 점에서 성공적인 것으로 판단

되었다. 이들 36개 중, Watson에서 파생된 판독의 중앙값 표적 적중율은 95%(범위: 39% ~ 97%)이고, Crick에서 파생된 판독의 표적 적중율 중앙값은 95%(범위: 39% ~ 98 %)이었다. 가장 중요한 것은, 표적이 입력 분자의 비교적 균일한 회수를 보여주었고, 변동 계수는 겨우 17%였다 (도 29). 시퀀싱된 앰플리콘의 길이(중앙값 77 bp, 사분위수 범위: 71 - 83 bp)도 모든 앰플리콘에서 유사했으며, 무-세포 혈장 DNA의 초기 크기가 ~167 bp ± 10.4 bp인 것과 일치했다 (도 29).

표 11. 멀티플렉스 패널의 구성 및 GSP 프라이머 서열.

*좌표는 인간 참조 게놈 hg19 릴리스를 나타낸다 (Genome Reference Consortium GRCh37, Feb 2009).
프라이머 서열 내의 별표는 포스포로티오에이트 결합을 나타낸다.

앰플리콘	멀티플렉스 세트 #	GSP1 서열	서열 식별 번호:	GSP2 서열	서열 식별 번호:	GSP2 Chrom	GSP2 가닥	GSP2 시작 위치*	GSP2 종료 위치*
AKT1_17	1	TGCGTGGCTCTCACC*A*C	6	CACCCGCACGTCTGT *A*G	7	chr14	-	105,246,55 4	105,246,57 0
APC_1310	1	ACAGGATCTTCAGCTGAC* C*T	8	CTTCAGCTGACCTAG TTCC*A*A	9	chr5	-	112,175,22 3	112,175,24 3
CDKN2A_5 1	1	CCTGGCTCTGACCATCTG *T*T	10	CCATTCTGTTCTCT GG*C*A	11	chr9	-	21,971,209 21,971,228	21,971,228
EGFR_856	1	AACACCGCAGCATGTC*A* A	12	GCAGCATGTCAAGA TCACA*G*A	13	chr7	+	55,259,486	55,259,506
FBXW7_36 9	1	TTGTGCAGAGTTCAAGTTAC CTT*A*G	14	GAGTTCAGTTACCT AGGAGATTTG*A*G	15	chr4	+	153,251,87 2	153,251,89 8
FBXW7_48 1	1	GGCCTGTCTCAATATCCCA *A*A	16	GTCTCAATATCCCA ACCCTAA*G*A	17	chr4	+	153,247,33 3	153,247,35 6
GNAS_202	1	GGAACCTGGTCTCAAGAT TC*C*A	18	GGTCTCAAGATTCC AGAAATC*A*G	19	chr20	-	57,484,426	57,484,449
KRAS_14	1	TCTGAATTAGCTGTATCGT	20	TATCGTCAAGGCACT	21	chr12	+	25,398,256	25,398,276

KRAS_144	1	CAA*G*G CTGTATTATTTCAGTGTTA CTTACCTGT*C*T	22	CTTG*C*C AGTGTACTTACCTG TCTTGTCT*T	23	chr12	+	25,378,536	25,378,559
NRAS_12	1	GACTGAGTACAACTGGT G*G*T	24	AACTGGTGGTGGTT GG*A*G	25	chr1	-	115,258,75 1	115,258,76 9
PIK3CA_80	1	TTTCGTAAGTGTACTCAA GAA*G*C	26	AGTGTACTCAAGAA GCAGAAA*G*G	27	chr3	+	178,916,82 7	178,916,85 0
PIK3CA_34 8	1	GCATCAGCAATTTGACTTTA CCTTAT*C*A	28	CATTTGACTTTACCTT ATCAATGTCT*C*G	29	chr3	-	178,921,56 3	178,921,59 0
PIK3CA_10 53	1	GCATGCTGTTTAAATTGTGT G*G*A	30	TGTGTGGAAGATCC AATCCAT*T*T	31	chr3	-	178,952,10 6	178,952,12 8
PTEN_145	1	CCATAGAAATCTAGGGCC TC*T*T	32	TAGGGCTCTTGTGC CTT*A	33	chr10	-	89,692,953	89,692,972
TP53_25	1	GGCCTGCCCTTCCAA*T*G	34	TTCCAATGGATCCAC TC*A*C	35	chr17	+	7,579,820	7,579,838
TP53_60	1	GCAGCCTCTGGCATT*C*T	36	GGCATTCTGGGAGC TTC*A*T	37	chr17	+	7,579,487	7,579,505
TP53_125	1	GCCAGGCATTGAAGTCTC* A*T	38	ATTGAAGTCTCATGG AAGC*C*A	39	chr17	+	7,579,272	7,579,292
TP53_161	1	CCGTCATGTGCTGTGA*C*T	40	GTGCTGTGACTGCTT GTA*G*A	41	chr17	+	7,578,426	7,578,445
TP53_196	1	CCAAATACTCCACACGCA	42	CACACGCAAAATTTC	43	chr17	+	7,578,240	7,578,259

[0443]

		A*A*T		TTC*A					
TP53_245	1	GGAGTCTTCCAGTGTGATG A*T*G	44	GTGATGATGGTGAG GATG*G*G	45	chr17	+	7,577,514	7,577,533
TP53_264	1	CACAAACACGCACCTC*A* A	46	ACGCACCTCAAAGC TGT*T*C	47	chr17	+	7,577,119	7,577,137
TP53_305	1	AAC TGACACCC TTGGTC*T*C	48	CCTCCACCGCTTCTT G*T*C	49	chr17	+	7,576,989	7,577,006
TP53_330	1	GAAACTTTCCACTTGATAA GAGG*T*C	50	TGATAAGAGGTCCC AAGACT*T*A	51	chr17	+	7,576,827	7,576,848
TP53_365	1	AGTAGGGCCAGGAAG*G* G	52	GGGCTGAGGTCACT C*A*C	53	chr17	+	7,573,910	7,573,926
TP53_390	1	TGGGAGGCTGTCAGT*G*G	54	GGGAACAAGAAAGTG GAGAA*T*G	55	chr17	+	7,572,906	7,572,926
APC_1458	2	TGCTTAGGTCCACTCTCT* *T	56	GTCCACTCTCTCTCTT TTCAG*C*A	57	chr5	-	112,175,66 8	112,175,69 0
BRAF_602	2	ACAACTGTTCAAACCTGATG G*G*A	58	ATGGGACCCACTCC AT*C*G	59	chr7	+	140,453,11 1	140,453,12 8
CTNNB1_3 2	2	TTAGTCACTGGCAGCAAC* A*G	60	GGCAGCAACAGTCT TAC*C*T	61	chr3	+	41,266,077	41,266,095
FBXW7_47 2	2	TAGAGGAAGAAGTCCCAA C*C*A	62	GAAGTCCCAACCAT GACAAG*A*T	63	chr4	+	153,249,32 8	153,249,34 9
FBXW7_50	2	CTAACCAACCCCTCTGCCA*	64	CCCTCTGCGCATCAT	65	chr4	+	153,247,26	153,247,28

[0444]

6		T [*] C		ATTGA [*] A [*] C			2	3
FGFR2_250	2	ACTCCTCCTTTCTTCCCT [*] C *T	66	TTTCTTCCCTCTCTCC ACC [*] A [*] G	chr10	-	123,279,68 4	123,279,70 4
HRAS_10	2	GGAGCGATGACGGAATAT AA [*] G [*] C	68	GGAATATAAGCTGG TGGTGG [*] T [*] G	chr11	-	534,296	534,317
KRAS_62	2	CTCATGTACTGGTCCCTCA T [*] T [*] G	70	GTCCCTCATTCGCACT GTAC [*] T [*] C	chr12	+	25,380,251	25,380,271
NRAS_59	2	GTGAAACCTGTTTGTGGGA C [*] A [*] T	72	CCTGTTTGTGGACA TACTGGAT [*] A [*] C	chr1	-	115,256,53 8	115,256,56 2
PIK3CA_11	2	CACGGTTGCCTACTGG [*] T [*] T	74	GTGGCTACTGGTTC AATTACT [*] T [*] T	chr3	-	178,916,91 1	178,916,93 4
PIK3CA_54	2	GAACAGCTCAAGCAATT CTAC [*] A [*] C	76	AATTCTACACGAGA TCCTCTC [*] T [*] C	chr3	+	178,936,05 7	178,936,08 0
PPP2R1A_1	2	CGACTCCCAGGTACTTC [*] C *G	78	TACTCCGGAACCTG TGC [*] T [*] C	chr19	+	52,715,940	52,715,959
PTEN_130	2	CATGTTGCAGCAATTCACT *G [*] T	80	CAATTCACTGTAAAG CTGGA [*] A [*] G [*] G	chr10	+	89,692,878	89,692,901
TP53_36	2	GGGACAGCATCAAAATCAT C [*] C [*] A	82	ATCAAATCATCCATT GCTTG [*] G [*] G	chr17	+	7,579,556	7,579,577
TP53_113	2	GTCACAGACTGGCTGT [*] C *C	84	TTGGCTGTCCAGAA T [*] G [*] C	chr17	+	7,579,328	7,579,345

[0445]

TP53_138	2	ATCAACCCACAGCTGC*A* C	86	TGCACAGGGCAGGT C*T*T	87	chr17	+	7,578,499	7,578,515
TP53_180	2	CCAGCTGCTCACCAT*C*G	88	TCACCATCGCTATCT GAG*C*A	89	chr17	+	7,578,367	7,578,386
TP53_220	2	TGACAACCAACCCCTTAAC*C *C	90	CAGTTGCAAAACCAG ACCTC*A*G	91	chr17	+	7,578,161	7,578,181
TP53_260	2	GTGCAGGGTGGCAAG*T*G	92	GGCAAGTGGCTCCT G*A*C	93	chr17	+	7,577,482	7,577,498
TP53_282	2	CCTTCTTGGGAGATTCT C*T*T	94	CGGAGATTCTCTCC TCTG*T*G	95	chr17	+	7,577,069	7,577,089
TP53_315	2	AAATATTCTCCATCCAGTG GTT*T*C	96	CCAGTGGTTTCTTCT TTGG*C*T	97	chr17	+	7,576,876	7,576,896
TP53_342	2	GGGCATCCTTGAGTTCC*A *A	98	CTTGAGTTCCAAGGC CTC*A*T	99	chr17	+	7,573,974	7,573,993
TP53_380	2	TCAGGCCCTTCTGTCTTG* A*A	100	CCTTCTGTCTTGAAC ATGAGT*T*T	101	chr17	+	7,572,943	7,572,965

[0446]

[0447]

두 가지 예시적인 방법을 사용하여 다중 앰플리콘을 평가할 수 있다. 첫 번째 방법은 상이한 웰에서 병렬 앰플리콘 특이적 PCR과 관련된다. 질병 재발을 모니터링하기 위한 액체 생검의 경우, 일반적으로 몇 가지 드라이버 유전자 돌연변이만 관찰되는 경우, 이 전략은 다중 PCR 반응에서 일반적으로 발생하는 프라이머 또는 기타 문제 간의 교차 하이브리드화에 대한 우려 없이 쉽게 적용할 수 있다. 관심 돌연변이를 알 수 없는 경우의 스크리닝과 같은 액체 생검의 다른 응용 분야에서는 더 많은 앰플리콘, 예를 들어, 각 PCR 웰에서 여러 프라이머 쌍의 조합을 평가하는 것이 유용하다. 이 실시예에서는 SaferSeqS를 사용하여 단일 웰에서 최소 18개의 앰플리콘을 효과적으로 분석할 수 있음을 보여주었고, 듀플렉스 시퀀싱을 포함하지 않는 절반-네스티드 PCR 전략은 최대 313개의 앰플리콘을 공동-증폭시킬 수 있음을 보여주었다.

[0448]

희귀 유전자 변형의 효율적인 검출 및 정량화를 허용함으로써, SaferSeqS는 매우 민감하고, 특정한 DNA-기반 분자 진단의 개발을 가능하게 하고, 뿐만 아니라 다양한 중요한 기본 과학적 질문에 답하는 데 도움이 될 수 있다.

[0449]

방법

- [0450] 혈장 및 말초 혈액 DNA 샘플
- [0451] 제조업체가 지정한 대로, cfPure MAX 무-세포 DNA 추출 키트(BioChain, 카탈로그 번호 K5011625MA)를 사용하여 10mL 혈장에서 DNA를 정제했다. 말초 WBCs로부터 DNA를 제조사가 명시한 바와 같이, QIAasympyony DSP DNA Midi Kit(Qiagen, cat. # 937255)로 정제했다. 모든 샘플로부터 정제된 DNA를 도처에서 설명한 대로 정량화했다 (가령, Douville et al., 2019 *bioRxiv*, 660258 참고).
- [0452] 라이브러리 준비
- [0453] 유입 DNA 단편들을 효율적으로 회수시키고, 동시에 이중 가닥으로 된 분자 바코드를 통합할 수 있는 맞춤형 라이브러리 준비 작업흐름이 개발되었다. 간단히 말해서, 다음과 같이 변형된 Accel-NGS 2S DNA 라이브러리 키트 (Swift Biosciences, cat. # 21024)를 사용하여, 무-세포 DNA 또는 말초 WBC DNA로 듀플렉스 시퀀싱 라이브러리를 준비했다: 1) DNA는 우라실 염기를 절단하기 위해, 37℃에서 15분 동안 3유닛의 USER 효소(New England BioLabs, cat. # M5505L)로 전-처리되었다; 2) 각 반응 후에 사용된 SPRI 비드/PEG NaCl 비율은 단부 복구 1, 단부 복구 2, 절찰 1 및 절찰 2에 대해 각각 2.0X, 1.8X, 1.2X 및 1.05X이었다; 3) 맞춤형 50 μ M 3' 어댑터 (표 12)가 시약 Y2를 대체했다; 4) 맞춤형 42 μ M 5' 어댑터 (표 12)가 시약 B2를 대체했다. 라이브러리는 후속적으로 절찰된 어댑터 (표 12)를 표적으로 하는 프라이머를 사용하여, 50 μ L 반응에서 PCR 증폭되었다. 반응 조건은 다음과 같다: 1X NEBNext Ultra II Q5 마스터 믹스 (New England BioLabs, cat. # M0544L), 2 μ M 범용 전행 프라이머, 및 2 μ M 범용 역행 프라이머 (표 12). 라이브러리는 다음 프로토콜에 따라 계획된 실험 수에 따라, 5, 7 또는 11 주기의 PCR로 증폭되었다: 98℃에서 30초, 98℃에서 10초, 65℃에서 75초, 그리고 4℃에서 대기. 5 또는 7 주기가 사용된 경우, 라이브러리는 단일 50 μ L 반응으로 증폭되었다. 11 주기가 사용된 경우, 라이브러리를 8개의 분취량으로 나누고, 각각 추가로 0.5단위의 Q5® Hot Start High-Fidelity DNA 중합 효소(New England BioLabs, cat. # M0493L), 1 μ L의 10 mM dNTPs (New England BioLabs, cat. # N0447L), 그리고 0.4 μ L의 25 mM MgCl₂ 용액 (New England BioLabs, cat. # B9021S).이 보충된 8개의 50 μ L 반응으로 증폭했다. 이 산물은 1.8X SPRI 비드(Beckman Coulter cat. # B23317)로 정제하였고, EB 완충액(Qiagen)으로 용리시켰다.

표 12. 라이브리리 구축, 가닥-특이적 PCR 검정, 및 시퀀싱용 올리고뉴클레오타이드

프라이머 서열 내의 별표는 포스포로티오에이트 결합을 나타내며, 다른 사용자 정의 변형은 "참고" 열에 표시된다.

Oligo 이름	Oligo 서열	서열 식별 번호:	목적/용도	주석
3' N14 어댑터 Oligo #1	/5Phos/ATAAACGACGCGC>NNNNNNNNNNNA GATCGGAAGAGCACACGTCTGAACTCCAG*T*C/3SpC 3/	102	UID 서열을 함유하는 3' 어댑터의 1/2	/5Phos/는 5' 포스페이트 변형이며; /3SpC3/는 3' 3-탄소 스페이서 변형이다
3' 어댑터 Oligo #2	GCCGUCGUUUUA/33dT/	103	3' 블록과 분해가능한 잔기를 함유하는 3' 어댑터의 1/2	/33dT/는 3'-데옥시티미딘 변형된 염기이고, U는 데옥시우리딘을 나타낸다.
5' 어댑터	ACACTCTTCCCTACACGACGCTCTCCGATCT	104	5' 어댑터 결합	
라이브리리 증폭 프라이머 #1	ACACTCTTCCCTACACGACGCTCTCCGATCT	105	라이브리리 증폭	
라이브리리 증폭 프라이머 #2	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT	106	라이브리리 증폭	
GSP1 프라이머	표 11 참고	107	제 1 네스티드 유전자-특이적 PCR 프라이머	Watson 및 Crick 제 1 네스티드 PCR 에 사용된 동일한 프라이머
Watson	GACTGGAGTTCAGACGT*G*T	108	Watson-특이적 고정 PCR	

[0454]

GSP1 고정 프라이머			프라이머	
Crick GSP1 고정 프라이머	109	ACACTCTTCCCTACACG*A*C	Crick-특이적 고정 PCR 프라이머	
Watson GSP2 고정 프라이머	110	AATGATACGGGACCAACCGAGATCTACACXXXXXXX CGACGTAAACGACGGCCAGT[GSP2]	Watson 제 2 네스티드 유전자-특이적 PCR 프라이머	[GSP2]는 유전자 특이적 부분의 서열을 나타내고(표 11 참조), X는 인덱스 서열을 나타낸다.
Crick GSP2 고정 프라이머	111	CAAGCAGAAGACGGCATACGAGATXXXXXXXCACAC AGGAAACAGCTATGACCATG[GSP2]	Crick 제 2 네스티드 유전자-특이적 PCR 프라이머	[GSP2]는 유전자 특이적 부분의 서열을 나타내고(표 11 참조), X는 인덱스 서열을 나타낸다.
Watson GSP2 고정/인덱싱 프라이머	112	CAAGCAGAAGACGGCATACGAGATXXXXXXXGTGAC TGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	Watson 제 2 네스티드 고정 및 인덱싱 프라이머	Xs는 인덱스 서열을 나타낸다.
Crick GSP2 고정/인덱싱 프라이머	113	AATGATACGGGACCAACCGAGATCTACACXXXXXXX ACACTCTTCCCTACACGACGCTCTTCCGAT*C*T	Crick 제 2 네스티드 고정 및 인덱싱 프라이머	Xs는 인덱스 서열을 나타낸다.
맞춤 판독 1 시퀀싱 프라이머	114	GAGATCTACACNNNNNNNNCGACGTAAACGACGG CCA*G*T	Illumina 페어드-엔드 시퀀싱	Illumina 표준 판독 1 시퀀싱 프라이머로 스파이크

[0455]

맞춤 인덱스 시퀀싱 프라이머	CATGGTCATAGCTGTTTCCTGTG* ^T G	115	Illumina 페어드-엔드 시퀀싱	Illumina 표준 인덱스 시퀀싱 프라이머로 스파이크
맞춤 판독 2 시퀀싱 프라이머	GGCATACGAGATNNNNNNNNNCACAGGAAACAGC TATGACCA* ^T G	116	Illumina 페어드-엔드 시퀀싱	Illumina 표준 판독 2 시퀀싱 프라이머로 스파이크

[0456]

[0457]

[0458]

라이브러리 구축

일반적으로 라이브러리 구축과 관련된 비효율성을 해결하기 위해, 어댑터 서열을 3' 및 5' DNA 단편 단부에 순차적으로 연결하고, 제자리에서 이중 가닥으로 된 분자 바코드 생성을 포함하는 전략이 설계되었다 (도 22a). DNA 단부들이 탈인산화되고, 복구된 후(도 22a, 1단계), 어댑터가 DNA 단편의 3' 단부에 부착되었다(도 22a, 단계 2). 이 어댑터는 3' DNA 단부에 선택적으로 결합되고, 어댑터-이량체 형성을 방지하는 단부 변형이 있는 부분적으로 이중-가닥으로 된 DNA 단편이었다. 특히, 이 어댑터는 5' 포스페이트 단부 변형을 함유하는 하나의 올리고뉴클레오타이드(표 12, 3' N14 어댑터 올리고 #1)로 구성되어 있으며, 이는 3' 차단기를 함유하는 또 다른 올리고뉴클레오타이드 및 데옥시티미딘을 대체하는 데옥시우리딘에 혼성화되었다. (표 12, 3' N14 어댑터 Oligo #2). 이 디자인은 이량체 또는 연쇄체(concatemer) 형성의 상당한 위험 없이, 3' 단부에 대한 효율적인 부착을 촉진하는 결합 반응에서 고농도의 어댑터 사용을 가능하게 했다. 더욱이, 이 어댑터는 두 개 올리고뉴클레오타이드 중 하나에서 상기 이중 UID의 한 가닥이 제 기능을 못하게 하였던 14개의 무작위 뉴클레오타이드 스트레치를 함유하였다. 상기 3' 어댑터의 결합에 이어서, 제 2 어댑터(표 12, 5' 어댑터)는 DNA 중합효소, 코헤시브 엔드-

특이적 리가아제, 및 우라실-DNA 글리코실화효소로 구성된 다크 해독-유사 반응을 통하여 상기 5' DNA 단편에 결합되었다 (도 22a, 단계 3). 이들 효소의 협력 작용으로 UID의 상보적 가닥이 합성되었고, 3' 어댑터의 차단 부분이 분해되었고, 확장된 어댑터가 5' DNA 단편 단부에 결합되었다. 이중 가닥 분자 바코드의 제자리(*in situ*) 생성에 사용되는 효소는 각 DNA 단편을 고유하게 바코드하고, 이중 어댑터를 효소적으로 준비할 필요가 없었는데, 이것은 유입 DNA 회수에 부정적인 영향을 미치는 것으로 알려져 있다. 최종적으로, 이 어댑터-결합된 단편들은 두 개의 원래 DNA 가닥의 중복 사본(UID "패밀리")을 생성하기 위해 제한된 수의 PCR 주기를 겪었다 (도 22a, 단계 4).

[0459] 라이브러리 증폭 주기 수 및 효율성의 영향

[0460] PCR 주기 수와 라이브러리 증폭 중 복제 효율성을 조정하여, SaferSeqS 매개변수를 최적화할 수 있다. SaferSeqS는 표적 풍부화를 위해 중복 Watson 가닥과 Crick 가닥으로부터 파생된 사본을 특정 가닥 특이적 PCR로 분할하는 것을 포함할 수 있기 때문에, 바람직한 구체예들에서, 듀플렉스 복구의 높은 가능성을 보장하려면, 필요한 수의 복사본을 생성해야 한다. 예를 들면, 100% 효율을 가정할 때, 1회의 PCR 주기 후에 각 템플릿 DNA 듀플렉스는 두 개의 이중 가닥으로 된 사본(각 사본은 가닥을 나타냄)으로 변환되고, 하나의 Watson 가닥-끌어낸 사본은 Watson-특이적 PCR로 분할되고, 하나의 Crick 가닥-끌어낸 사본은 Crick-특이적 PCR로 분할되도록 이 두 사본을 적절하게 배포될 확률은 25%에 불과하다. PCR 주기의 수를 늘리거나, 또는 증폭 효율을 높이면 더 많은 리던던트(redundant) 사본이 생성되어, 원래 DNA 듀플렉스 복구 확률이 높아진다.

[0461] 효율적인 듀플렉스 회수에 필요한 PCR 주기 수와 증폭 효율을 추정하기 위해, 확률 모델이 개발되었다. 이 모델은 세 단계로 구성되었다: 1) 라이브러리 증폭 동안 생성된 PCR 자손의 수 시뮬레이션; 2) 이러한 PCR 사본을 Watson 가닥과 Crick 가닥 특이적 반응으로 무작위 분할; 그리고 3) 듀플렉스 회수, 즉 Watson 가닥-특이적 반응으로 분할된 적어도 하나 이상의 Watson 가닥 유래 사본과 Crick 가닥-특이적 반응으로 분할된 적어도 하나의 Crick 가닥 유래 사본 가닥으로 분할된 적어도 하나의 Crick 가닥 유래 사본을 갖는 원래 DNA 듀플렉스의 비율을 결정한다.

[0462] 각 라이브러리 증폭 주기 동안 생성된 원래 템플릿 가닥의 PCR 사본 수는 이항 분포(binomial distribution)를 따른다. 첫 번째 PCR 주기의 경우, 가닥-특이적 사본의 수를 1로 초기화시켰다. 첫 번째 라이브러리 증폭 주기는 단순히 두 개의 원래 템플릿 가닥을 변형시키고 물리적으로 구별되는 이중 가닥 형태로 변환하는 역할을 하기 때문에 숫자는 1(2가 아니라)로 초기화되었다는 점에 유의해야 한다. PCR의 후속 i 번째 주기 동안, 각각의 n_i PCR 사본은 확률 p (즉, 증폭 효율)로 복제시켜, $n_i + \text{Binom}(n_i, p)$ 과 동일한 총 n_{i+1} PCR 사본을 생성할 수 있다. 이 과정은 i PCR 주기 후에 생성된 자손의 수를 시뮬레이션하기 위해 반복적으로 반복되었다. 공식적으로, 생성된 총 PCR 사본 수는 다음과 같이 표현할 수 있다:

$$n_i = \sum_{j=1}^{i-1} \text{Binom}(n_j, p); n_1 = 1$$

[0463]

[0464] 라이브러리 증폭 후, 각각의 원래 DNA 듀플렉스는 위에서 설명한 대로, Watson 가닥의 $n_{i,W}$ 사본과 Crick 가닥의 $n_{i,C}$ 사본을 생성하도록 증폭되었다. 각각의 $n_{i,W}$ 사본 및 $n_{i,C}$ 사본은 각 반응에 사용된 라이브러리의 비율과 동일한 확률 q 로 Watson 가닥- 및 Crick 가닥-특이적 PCR 반응으로 무작위로 분할된다. 라이브러리를 단일 Watson 가닥-특이적, 그리고 단일 Crick 가닥-특이적 PCR로 나눌 때 q 는 50%이다. 라이브러리가 두 개의 Watson 가닥-특이적, 그리고 Crick 가닥-특이적 PCR로 분할되는 경우, q 는 25%이다. 적절한 가닥 특이적 PCR (각각 k 번째 Watson-특이적- 또는 Crick-특이적 PCR의 경우 $N_{k,W}$ 또는 $N_{k,C}$)으로 분할된 PCR 사본의 수는 이항 분포에서 가져오며, 각각 Watson 사본 및 Crick 사본에 대한 $n_{i,W}$ 또는 $n_{i,C}$ "시도" 및 "성공" 확률 q 이다. 따라서, 적어도 하나의 Watson-끌어낸 PCR 사본을 k 번째 Watson 특정 PCR 반응으로 분할할 확률은 다음과 같다:

$$P(N_{k,W} > 0) = 1 - (1 - q)^{n_{i,W}}$$

[0465]

[0466] 유사하게, 적어도 하나의 Crick-끌어낸 PCR 사본을 k 번째 Crick 특정 PCR 반응으로 분할할 확률은 다음과 같다:

$$P(N_{k,C} > 0) = 1 - (1 - q)^{n_{i,C}}$$

[0467]

[0468] 원래 DNA 이중 가닥의 두 가닥은 $N_{k,W}$ 및 $N_{k,C}$ 가 0보다 큰 경우에만, 회수될 수 있다. PCR 자손의 분할이 독립적

이기 때문에, 듀플렉스 회수 확률은 다음과 같이 예측된다:

$$P(N_{k,W} > 0, N_{k,C} > 0) = [1 - (1 - q)^{n_{i,W}}][1 - (1 - q)^{n_{i,C}}]$$

[0469]

우리는 PCR 효율을 100%에서 50%로, 라이브러리 증폭 주기 수를 1에서 11로, 각 반응에 사용된 라이브러리의 비율을 50%에서 1.4%로 변경시켰다. 각 조건에 대해, 위에서 설명한 프로세스의 10,000번의 시뮬레이션을 수행하였고, 도 28에 평균 듀플렉스 회수에 대해 보고하였다.

[0471]

고정된 절반-네스티드 PCR을 통한 단편 크기 및 회수

[0472]

고정된 절반-네스티드 PCR은 이론적으로 기존의 앰플리콘 PCR보다 템플릿 분자의 더 높은 회수율을 보여준다. 기존의 앰플리콘 PCR에서 템플릿 분자는 전행 및 역행 프라이머 결합 부위와 앰플리콘을 특징하는 중간 서열을 모두 함유해야 한다. 대조적으로, 고정된 절반-네스티드 PCR에서 주형 분자는 회수를 위해, 두 개의 유전자-특이적 프라이머 결합 부위의 결합만 있으면 된다. SaferSeqS에 사용되는 네스티드 유전자 특이적 프라이머의 복합된 공간(footprint)은 대략적으로 30bp인데 비해, cfDNA 프로파일링을 위해 SafeSeqS에서 사용하는 앰플리콘 길이는 일반적으로 70-80bp이다. 공식적으로, 일정한 무작위 단편 시작/종료 좌표를 가정하면, 길이가 L 인

템플릿 분자를 회수할 확률은 $\frac{L-r}{L}$ 이며, 여기서 r 은 전통적인 PCR의 경우 앰플리콘 길이이거나, 또는 고정된 절반-네스티드 PCR의 경우, 유전자 특이적 프라이머의 복합된 범위의 길이이다. 따라서, ~167bp 크기의 무-세포 DNA 단편의 경우, 고정된 절반-네스티드 PCR은 이론적으로 기존 앰플리콘 PCR보다 ~25% 더 많은 원래 템플릿 단편을 회수할 수 있다. 더욱이, 전행 및 역행 프라이머의 위치에 따라 미리-특정된 산물 크기를 생성하는 기존의 앰플리콘 PCR과 달리, 고정된 절반-네스티드는 유전자 특이적 프라이머의 위치에 의해 지시되는 단편 말단 중 오직 하나만을 갖는 다양한 길이의 단편을 만든다. 일정하게 무작위 시작/끝 좌표를 가진 길이 L 의 템플릿

분자를 가정하면, 고정된 절반-네스티드 PCR 후 관찰된 단편 길이는 $\frac{L-r}{2}$ 가 되며, 여기서 r 은 유전자 특이적 프라이머의 복합된 범위의 길이이다.

[0473]

SaferSeqS 생물정보학 파이프라인의 예시적인 구체예

[0474]

SaferSeqS 생물정보학 파이프라인의 예시적인 구체예에서, 각 샘플에 대한 Watson 판독 및 Crick 판독은 단일 BAM 파일로 병합되었고, 메이트 쌍을 쉽게 추출할 수 있도록 SAMtools를 사용하여 판독 이름별로 정렬되었다. 사용자 정의 Python 스크립트는 듀플렉스 패밀리의 후속 재구성 및 Watson 슈퍼뮤턴트, Crick 슈퍼뮤턴트 및 슈퍼칼리뮤턴트의 식별에 사용되었다.

[0475]

먼저, 판독을 UID 패밀리로 그룹화하면서 비트와이즈(bitwise) 플래그(즉, FLAG 필드)의 값을 검사함으로써, Watson 가닥 및 Crick 가닥에서 파생된 판독을 기록했다. 비트와이즈 플래그별 값 99 및 147을 함유하는 판독은 Watson 가닥에서 파생되고, 비트와이즈 플래그 83 및 163을 함유하는 판독은 Crick 가닥에서 파생된다. 임의의 다른 비트와이즈 플래그 값이 있는 판독은 후속 분석에서 제외되었다. 비트와이즈 플래그는 매핑 동안 읽기 쌍에 할당되는 숫자 값이다. 이들 값은 판독 메이트가 서로에 대해 게놈에 정렬되는 방식을 나타낸다. 예를 들면, 판독이 참조 가닥에 매핑되고, 이의 메이트가 역-(상보적) 가닥에 매핑되는 경우, 이 판독 쌍은 Watson 가닥에서 파생되었다. 유사하게, 판독이 역-(상보) 가닥에 매핑되고, 이의 메이트가 참조 가닥에 매핑되면 이 판독 쌍은 Crick 가닥에서 파생되었다.

[0476]

둘째, 내인성 분자 바코드(즉, 단편의 단부 좌표)의 결정을 개선하기 위해, UID 패밀리 그룹화 동안 두 가지 추가 품질 관리 기준이 부과되었다: 1) 단편 단부들의 5' 또는 3'에서 소프트 클리핑(soft clipping)이 있는 판독은 배제되었으며, 2) 판독에는 하나 이상의 미스매치가 없는 외인성 UID 바로 다음에 예상되는 일정한 태그 서열(GCCGTCGTTTAT; 서열 식별 번호: 117)를 함유하도록 요구되었다.

[0477]

세째, 이 실시예에서 가능한 외인성 UID 서열의 수가 시작 템플릿 분자의 수를 크게 초과하기 때문에, 두 분자가 동일한 외인성 UID 서열을 공유하지만, 상이한 내인성 UIDs를 갖는, "바코드 충돌(barcode collisions)"은 극히 드물어야 한다. 특히, 예상되는 바코드 충돌 수는 전통적인 "출생 문제(birthday problem)"에서 계산할 수 있으며, 다음과 같다:

$$E[X] = n \left\{ 1 - \left(1 - \frac{1}{N} \right)^{n-1} \right\}$$

[0478]

[0479] 여기서 n 은 템플릿 분자의 수와 같고, N 은 가능한 바코드의 수와 같다. 14bp 외인성 UID 서열(총 268,435,456개의 가능한 서열로 구성) 및 10,000개의 게놈 등가물의 경우, 예상 충돌 수는 0.37 또는 유입 0.0037%이다. 따라서, 이 실시예의 경우, 각 외인성 UID 서열이 하나의 내인성 UID와만 연합되어야 했다. 외인성 UID가 하나 이상의 내인성 UID와 연결된 경우, 가장 큰 패밀리리는 보존시키고, 다른 모든 패밀리리는 삭제되었다.

[0480] 다른 실험 설계 매개변수에서 독특하지 않은 외생 UIDs가 사용될 수 있고, 독특하지 않은 외인성 UID는 UID 패밀리에 할당하기 위해 내인성 UID와 함께 사용될 수 있다.

[0481] 최종적으로, 외인성 바코드 자체가 PCR 및 시퀀싱 오류에 취약하기 때문에, UMI-도구 네트워크 인접 방법을 사용하여 UID 서열의 오류 수정하고, UID 패밀리들을 재-그룹화했다.

[0482] 판독이 UID 패밀리로 어셈블리된 후, Watson 슈퍼뮤턴트, Crick 슈퍼뮤턴트 및 슈퍼칼리뮤턴트는 도처에서 설명된 대로 호출되었다. 공통적 다형성을 배제하기 위해, Genome Aggregation Database (gnomeAD)에서 0.1%보다 큰 대립유전자 빈도로 존재하는 모든 돌연변이들이 배제되었다. 슈퍼칼리뮤턴트를 포함하는 판독은 가능한 정렬 허상을 배제하도록, 최종적인 수작업 검사를 거쳤다.

[0483] 비-클론성 체세포 돌연변이 비율의 예측

[0484] 이 연구에 사용된 DNA는 평균 연령 30세의 개인 집단에서 얻었다. 그 결과, 이 샘플에서 비-클론성 체세포 단일 염기 치환의 예상 빈도는 이배체(diploid) 게놈당 426개, 또는 bp당 대략적으로 7×10^{-8} 돌연변이다. 이 연구에서, 우리는 SaferSeqS를 사용하여 건강한 대조군으로부터 파생된 DNA에서 총 42,695,395개의 염기를 평가했다. 이 42,695,395개의 염기 중에서, 12×10^{-8} 의 돌연변이 빈도를 나타내는 5개의 단일 염기 치환 슈퍼칼리뮤턴트가 탐지되었다. 관찰된 슈퍼칼리뮤턴트의 빈도가 건강한 혈액 세포에서 비-클론성 체세포 돌연변이율의 기존 추정치와 일치하는지 확인하기 위해, 다음과 같은 정확한 단측 이항 p-값을 산출하였다:

$$P(X \geq 5) = 1 - \sum_{k=0}^4 \binom{42,695,395}{k} (7 \times 10^{-8})^k (1 - 7 \times 10^{-8})^{42,695,395-k} = 0.19$$

[0485]

[0486] 따라서, 관찰된 슈퍼칼리뮤턴트의 수와 건강한 조혈 줄기 세포에서 발생하는 연령-연합된 비-클론성 체세포 돌연변이의 예측된 수 사이에는 통계적으로 유의한 차이가 없다.

[0487] 고정 절반-네스티드 PCR

[0488] 관심대상 영역의 표적 풍부화는 듀플렉스 시퀀싱에 필요한 고정된 절반-네스티드 PCR의 결정적 변형을 사용하여 달성되었다. 이 맞춤형 스탠드 특이적 분석을 개발하는 동안 주기의 횟수, 프라이머 농도 및 증합효소 제형을 비롯한, 다양한 반응 조건이 최적화되었다. 최종 최적화된 프로토콜은 다음과 같다: PCR의 첫 번째 라운드는 다음 조건에서 50 μ L 반응에서 수행되었다: 1X NEBNext Ultra II Q5 마스터 믹스 (New England BioLabs, cat. # M0544L), 2 μ M GSP1 프라이머, 및 Watson 가닥 증폭용 2 μ M P7 짧은 고정 프라이머 GSP1 프라이머는 각 앰플리콘에 대해 특이적이었고, P7 짧은 앵커 프라이머는 모든 앰플리콘의 Watson 가닥에 대한 앵커 프라이머로 사용되었다 (표 11 및 표 12). Crick 가닥은 P5 짧은 프라이머 앵커 프라이머가 P7 짧은 프라이머로 대체되었다는 점을 제외하고, 별도의 웰에서 동일한 방식으로 증폭되었다. Watson 가닥의 증폭에 사용된 GSP1 프라이머는 Crick 가닥에 사용된 GSP1 프라이머와 동일하며; Watson 가닥과 Crick 가닥 PCR의 유일한 차이점은 앵커 프라이머이다. 두 반응(Watson 가닥 및 Crick 가닥)은 상기에서 설명된 열순환 프로토콜에 따라 19회로 증폭되었다.

[0489] Watson 가닥의 경우, 1차 PCR에 사용된 동일한 반응 조건을 사용하여 50 μ L 반응에서 2라운드 PCR이 형성되었다. 차이는 다음과 같다: (i) 제 1 고정 Watson 가닥 PCR의 1% 산물이 템플릿(상기 제 1 PCR의 템플릿으로 사용된 라이브러리를 대신하여)으로 사용되었고, 그리고 (ii) 프라이머: 유전자-특이적 프라이머 GSP2는 GSP1 유전자-특이적 프라이머로 대체되었고, 앵커 P5 인덱싱 프라이머는 P7 짧은 고정 프라이머로 대체되었다. 상기 제 2 round of PCR for Crick 가닥을 위한 제 2 라운드 PCR은 다음을 제외하고 동일하게 수행되었다: (i) 템플릿: 상기 제 1 Crick 가닥 PCR이 템플릿으로 사용되었고, 그리고 (ii) 프라이머: 고정 P7 인덱싱 프라이머는 고정 P5 인덱싱 프라이머로 대체되었다. 두 반응 모두(Watson 가닥 및 Crick 가닥)은 상기 설명된 열순환 프로토콜에 따라 17 사이클로 증폭되었다. 제 2 라운드 PCR에 사용된 프라이머의 서열은 표 12에 나열되어 있다. 제 2 PCR 라운드의 산물을 풀링하고, 시퀀싱 전에 1.8X SPRI 비드로 정제했다.

- [0490] 단일 반응 내에서 여러 표적이 동시에 증폭되는 실험의 경우, PCR 조건은 다음을 제외하고, 상기에서 기술된 것과 동일하였다: (i) 각 유전자-특이적 프라이머는 0.25 μ M의 최종 농도로 내포되었고, 그리고 (ii) 고정 프라이머의 농도는 표적 당 0.25 μ M의 최종 농도로 내포되었다 (가령, 25개 표적이 공동-증폭된다면, 최종 농도는 6.25 μ M임).
- [0491] 시퀀싱
- [0492] 제작자가 설명한 바와 같이, KAPA 라이브러리 정량화 키트 (KAPA Biosystems, cat. # KK4824)를 이용하여 라이브러리 농도를 결정하였다. Illumina MiSeq 기기에서 8개의 염기 듀얼-인덱싱과 함께, 2 x 75 페어드-엔드 판독을 사용하여 시퀀싱을 수행했다. 듀얼-인덱싱된 PhiX 대조군 라이브러리(SeqMatic cat. # TM-502-ND)는 모든 주기에 걸쳐 염기 다양성을 보장하기 위해, 전체 템플릿의 25%에 스파이키-인되었다. 맞춤형 판독 1, 인덱스, 및 판독 2 시퀀싱 프라이머 (표 12)에 표준 Illumina 시퀀싱 프라이머 1 μ M의 최종 농도로 조합하였다.
- [0493] 돌연변이 호출 및 SaferSeqS 분석 파이프라인
- [0494] SafeSeqS 데이터의 분석은 사용자 정의 Python 스크립트를 사용하여, 도처에서 설명된 대로 수행되었다 (가령, Kinde et al., 2011 *Proc Natl Acad Sci U S A* 108:9530-9535 참고). 시퀀싱 판독은 Picard의 IlluminaBasecallsToSam(broadinstitute.github.io/picard)을 사용하여, 처음 14개 뉴클레오타이드를 UID 시퀀스로 추출해내고, 어댑터 시퀀스를 마스킹하여 초기 프로세싱이 수행되었다. 그런 다음, 판독을 BWA-MEM(버전 0.7.17)을 사용하여 hg19 참조 게놈에 매핑시키고, SAMtools를 사용하여 UID 서열 별로 소팅했다. UID 패밀리는 이들이 2개 또는 그 이상의 판독으로 구성되며, 리드의 >90%가 예상 프라이머 서열이 있는 참조 게놈에 매핑된 경우, 점수가 매겨졌다. "슈퍼뮤턴트"는 매핑된 판독의 >95%에 존재하고, 평균 Phred 점수가 25보다 큰 돌연변이로 식별되었다.
- [0495] SaferSeqS 분석을 위해 맞춤형 분석 파이프라인이 개발되었다. 간단히 말해서, 판독을 탈-다중화시키고, 인덱스 서열을 사용하여 판독이 파생된 가닥을 식별해내었다. 명확성과 간결성을 위해 Watson 가닥으로부터 파생된 판독을 "Watson 판독"이라고 하고, Crick 가닥으로부터 파생된 판독을 "Crick 판독"이라고 한다. Watson 판독의 경우, 상기 판독 1의 처음 14개 염기가 UID 서열로 추출되었다. 삽입물의 방향이 Crick 가닥에 대해 반대이기 때문에, 판독 2의 처음 14개 염기가 Crick 판독에 대한 UID 서열로 추출되었다. 어댑터 서열은 Picard의 IlluminaBasecallsToSam(broadinstitute.github.io/picard)을 사용하여 마스킹되었고, 생성된 판독의 템플릿-특이적 일부는 BWA-MEM (version 0.7.17)을 이용하여 hg19 참조 게놈에 매핑되었다. 정렬 후, 매핑된 Watson 판독 및 Crick 판독은 SAMtools를 사용하여 병합 및 정렬시켰다.
- [0496] Python 스크립트는 듀플렉스 패밀리의 후속 재구성 및 Watson 슈퍼뮤턴트, Crick 슈퍼뮤턴트 및 슈퍼칼리뮤턴트의 식별에 사용되었다. 도처에서 설명한 대로, 분자 바코드 서열 내에서 PCR 및 시퀀싱 오류를 수정한 후 (Smith et al., 2017 *Genome Res* 27:491-499 참고), 동일한 듀플렉스 패밀리에 속하는 Watson 판독 및 Crick 판독을 그룹화하여, 원래 템플릿 분자의 서열을 재구성했다. 라이브러리 구축의 최종 복구 단계로부터 파생된 허상을 배제하기 위해, 3' 어댑터 서열에서 10개 미만의 염기는 돌연변이 분석에 고려되지 않았다. Watson 슈퍼뮤턴트와 Crick 슈퍼뮤턴트는 각각 듀플렉스 패밀리의 Watson 판독 또는 Crick 판독의 > 80%에 존재하는 돌연변이로 정의되었다. 슈퍼칼리뮤턴트는 동일한 UID를 가진 Watson 패밀리 및 Crick 패밀리의 >80%에 존재하는 돌연변이로 정의되었다.
- [0497] 통계적 분석
- [0498] 연속 변수(continuous variables)는 정중값과 그 범위로 보고되었고, 범주형 변수(categorical variables)는 정수와 백분율로 보고되었다. 모든 통계 테스트는 R의 stats 패키지를 사용하여 수행되었다 (version 3.5.1).
- [0499] 이러한 결과는 SaferSeqS가 극히 높은 특이성을 갖는 희귀 돌연변이를 탐지할 수 있음을 보여준다. 이 기술은 확장성이 뛰어나고, 비용 효율적이며, 처리량이 많은 자동화에 적합하다. SaferSeqS는 기존 듀플렉스 시퀀싱 기술에 비해 유입 회수율이 최대 5~75배 향상되었으며, 제한된 양의 출발 물질에 적용할 수 있고, 분자 바코드를 사용하는 표준 PCR-기반 접근 방식에 비해 오류 수정이 >50-배 향상되었다 (도 23, 표 8). 슈퍼칼리뮤턴트가 아닌 Watson 슈퍼뮤턴트 또는 Crick 슈퍼뮤턴트만 사용하는 최적의 결찰-기반 기술에 비해 오류 수정이 또한 >50-배 향상되었다 (도 26, 표 9). 두 감소 모두다 이를 테면, 암의 스크리닝에서, 그리고 최소 잔류 질환 설정에서, 단일 또는 매우 낮은 사본 수에 존재하는 돌연변이의 검출에 유용하다. 최종적으로, SaferSeqS는 듀플렉스 시퀀싱을 통합하고 있기 때문에, 이것은 단일 앰플리콘 분석을 위한 디지털 액적(droplet) PCR보다 훨씬 더 민감하고, 디지털 액적 PCR과 달리, 고도로 다중화될 수 있다.

[0500] 기타 구체예들

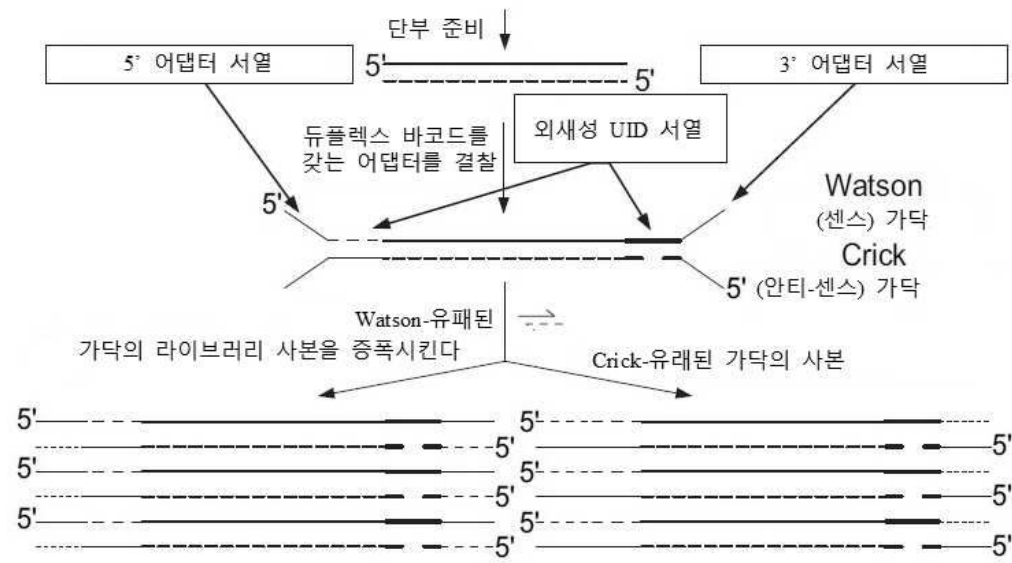
[0501] 본 발명이 그의 상세한 설명과 관련하여 설명되었지만, 전술한 설명은 예시하기 위한 것이며, 본 발명의 범위를 제한하지 않으며, 본 발명의 범위는 첨부된 청구항의 범위에 의해 특정된다. 다른 측면들, 장점들 및 수정들은 다음 청구 범위의 범주 내에 있다:

[0502] 참고자료로의 편입

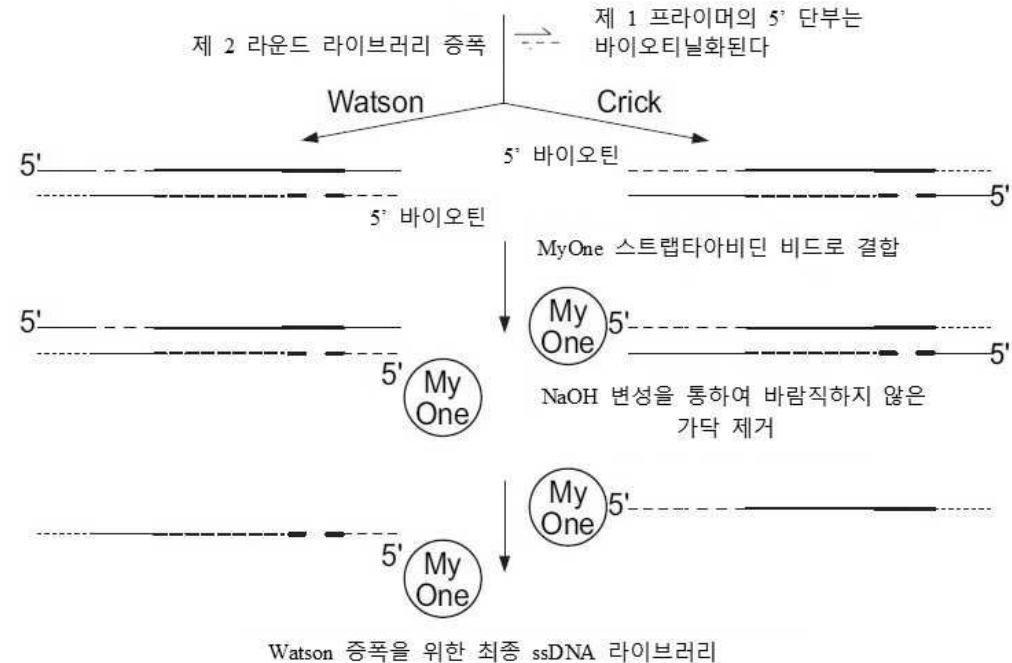
[0503] 본 명세서의 본문 내에 인용된 모든 참고 문헌, 등록된 특허 및 특허 출원은 모든 목적을 위해 전체적으로 참고로 포함된다.

도면

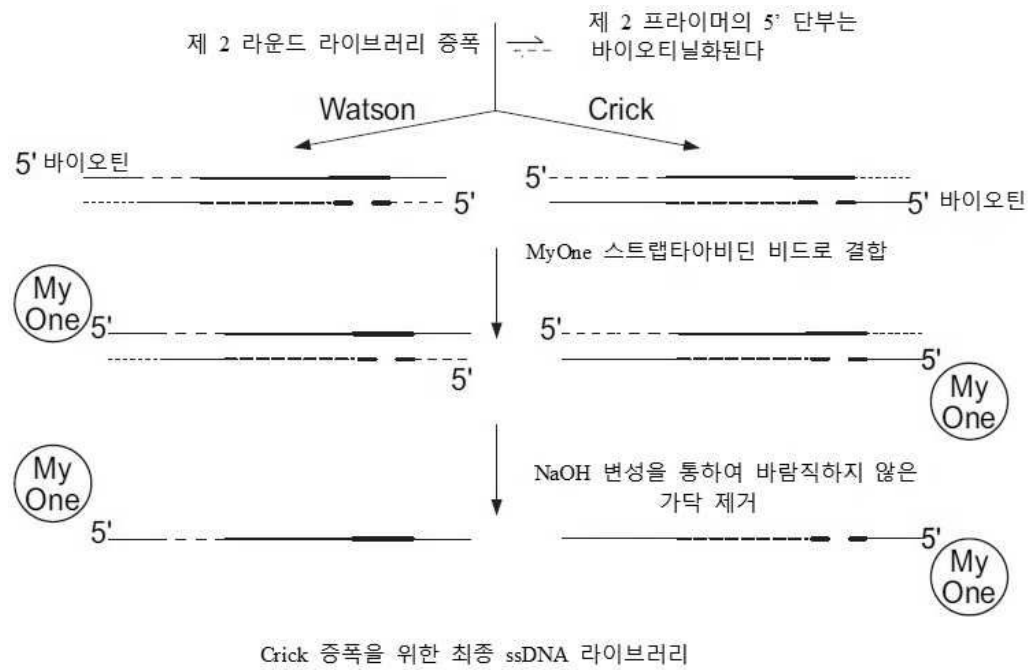
도면1



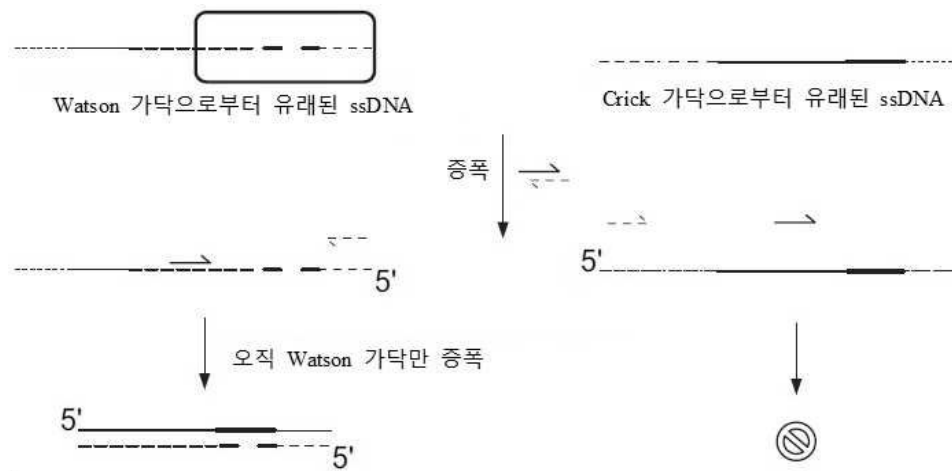
도면2



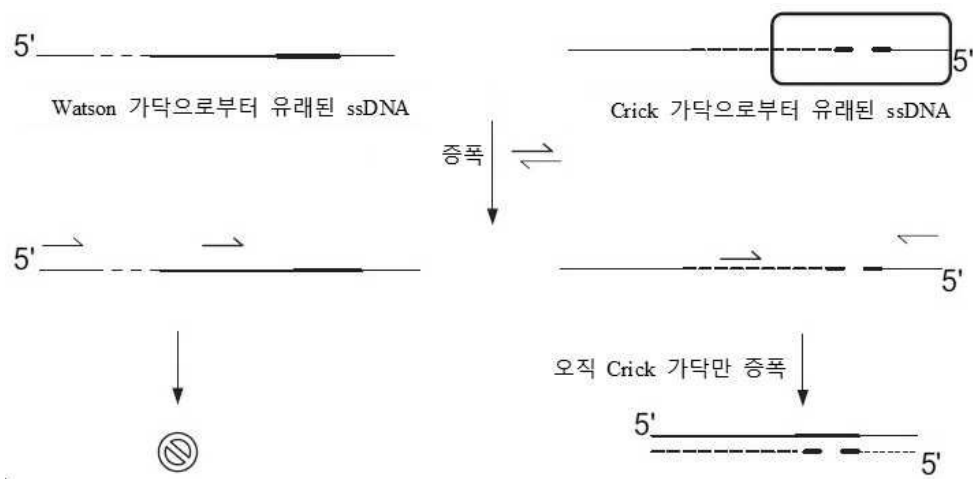
도면3



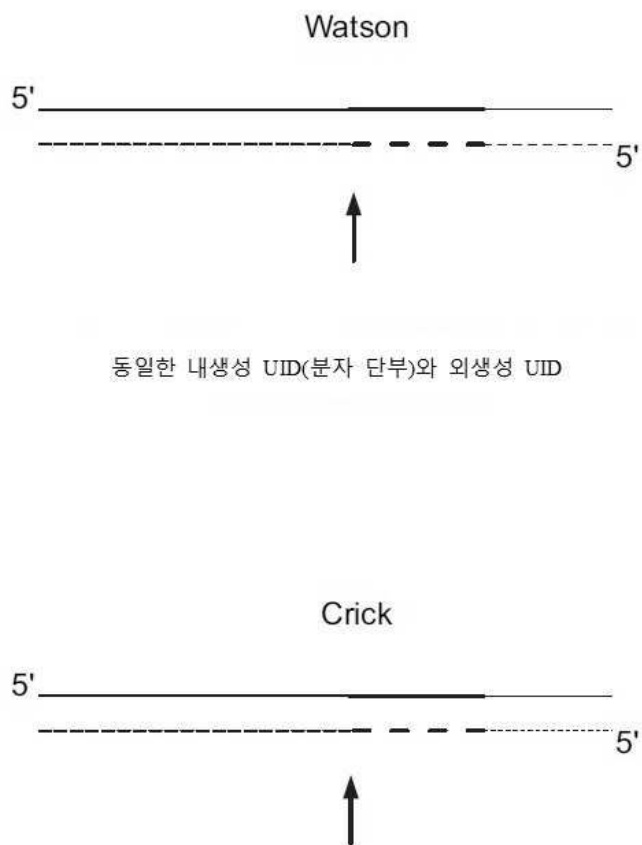
도면4



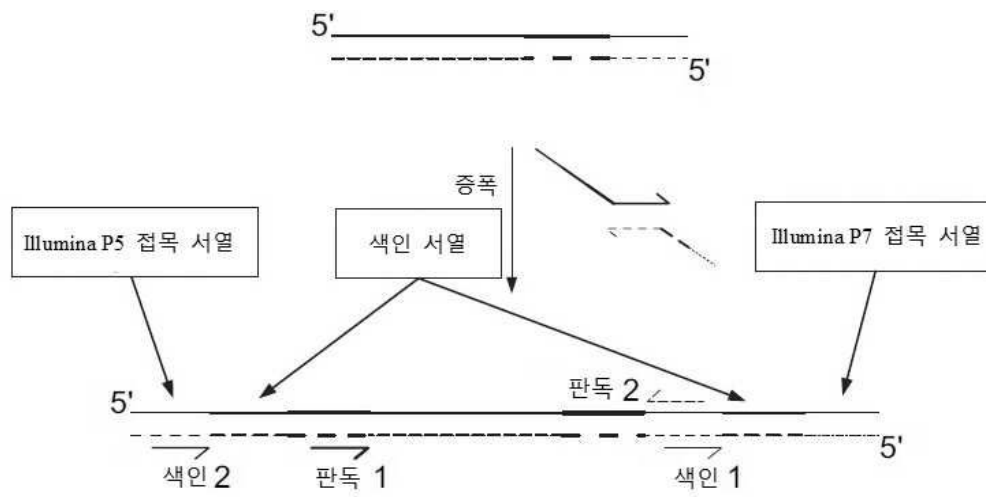
도면5



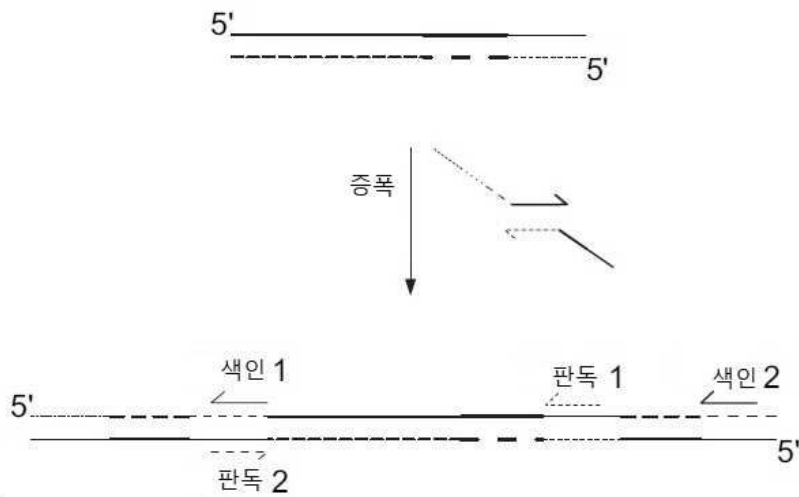
도면6



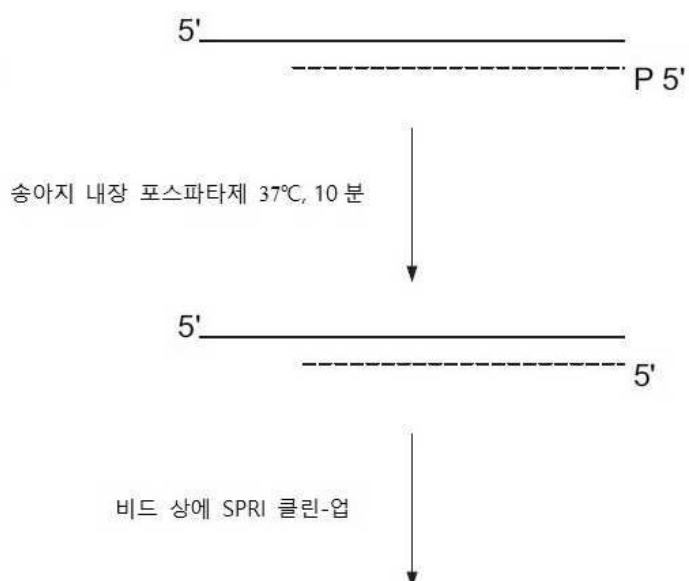
도면7



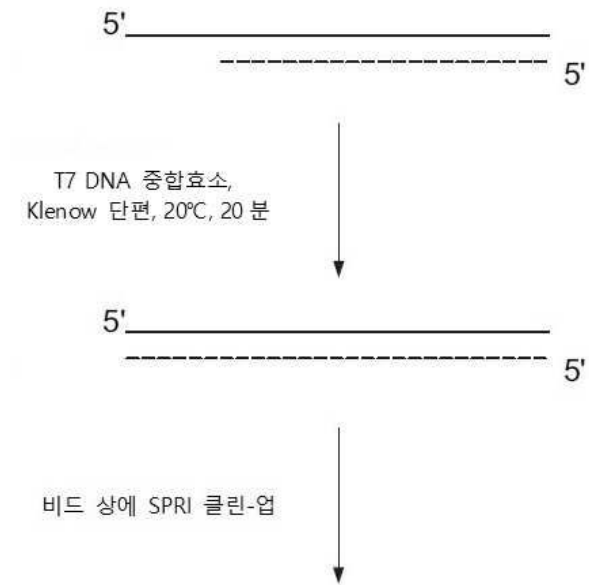
도면8



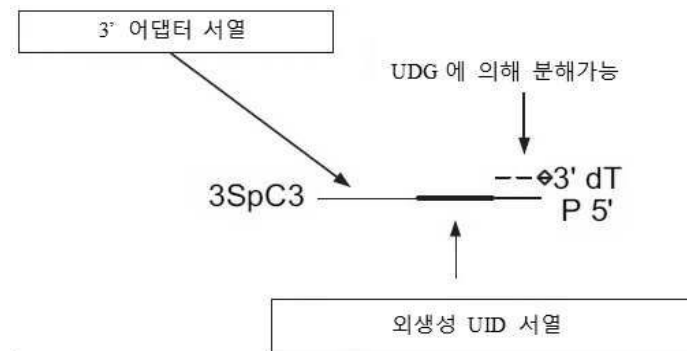
도면9



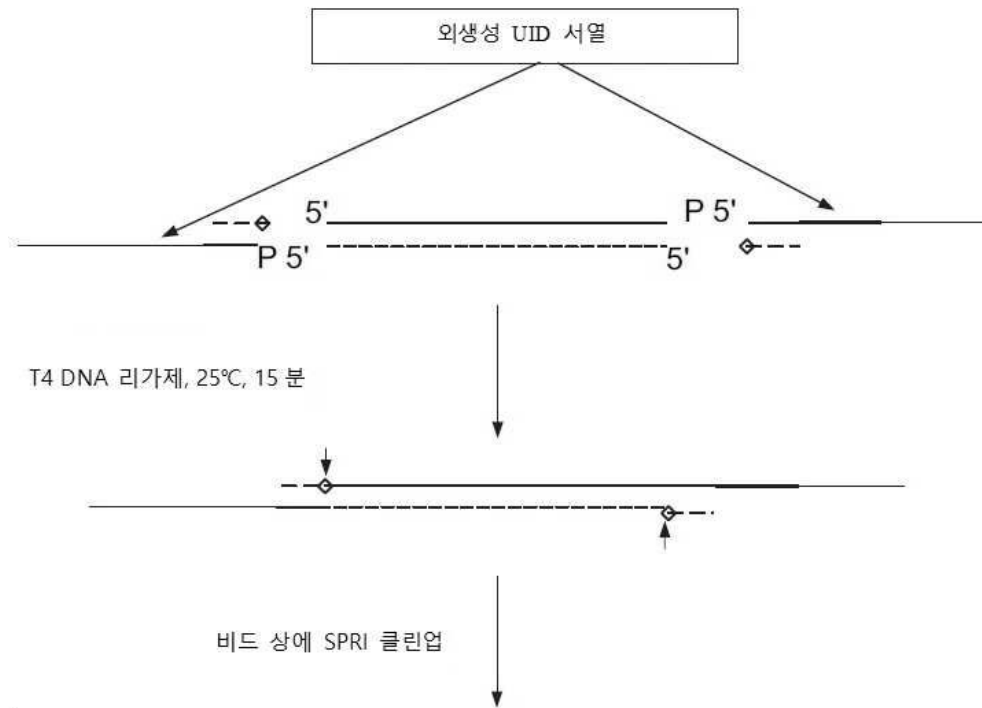
도면10



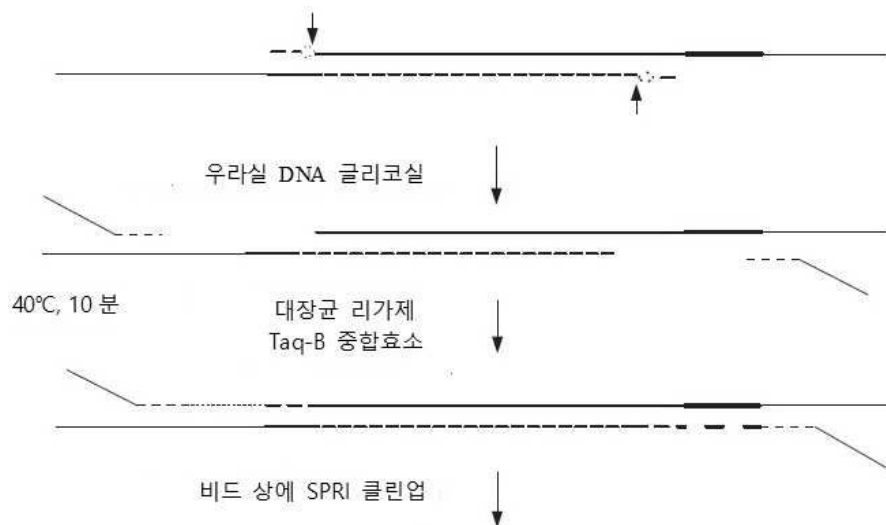
도면11



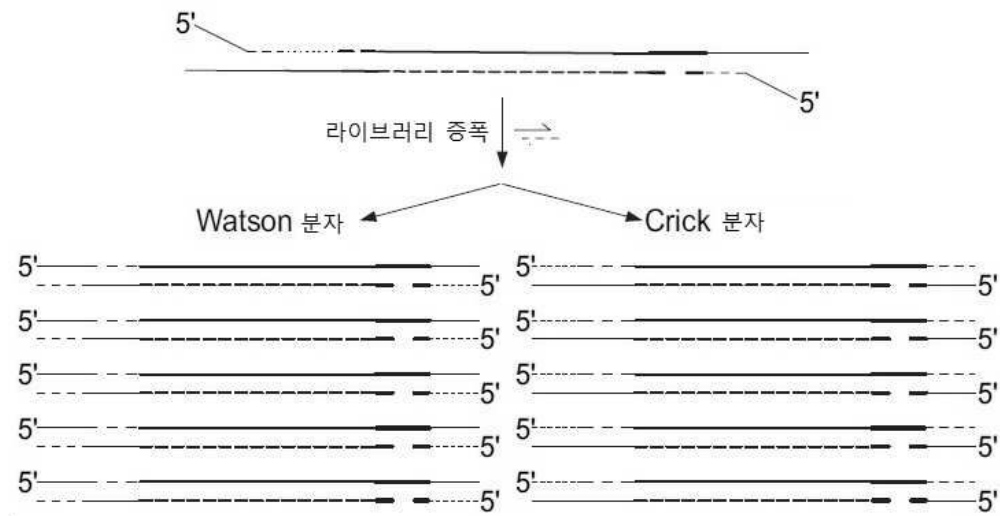
도면12



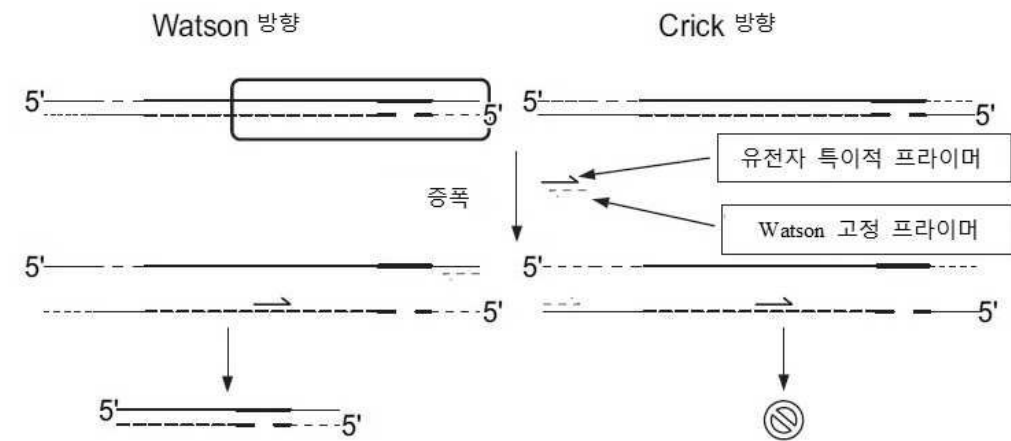
도면13



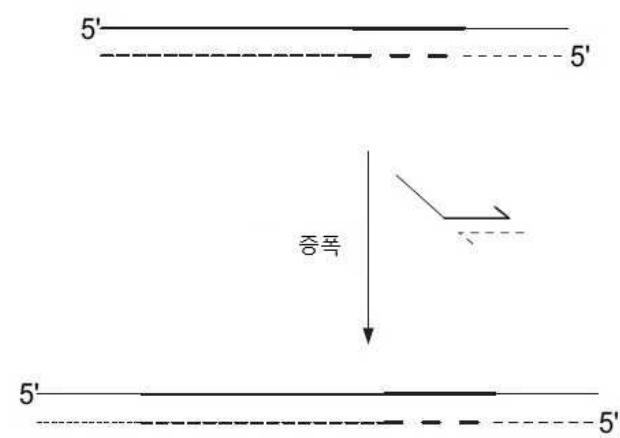
도면14



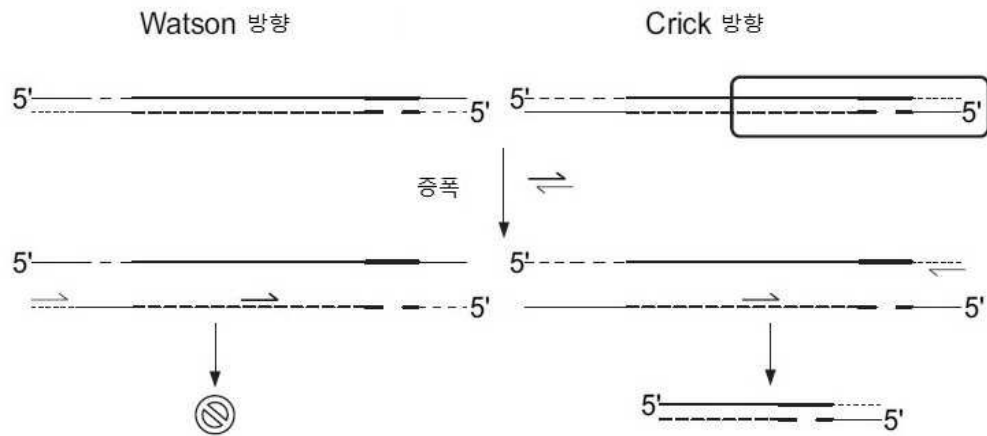
도면15



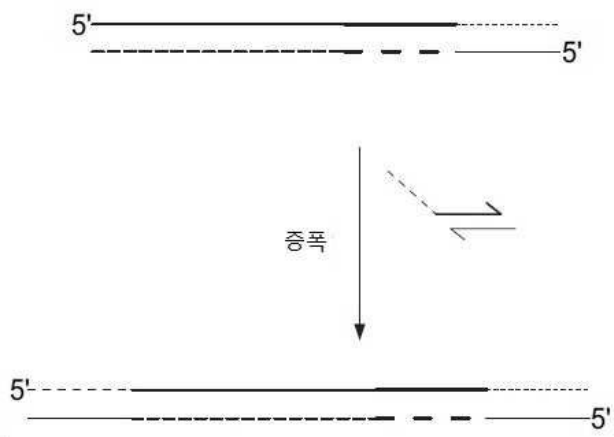
도면16



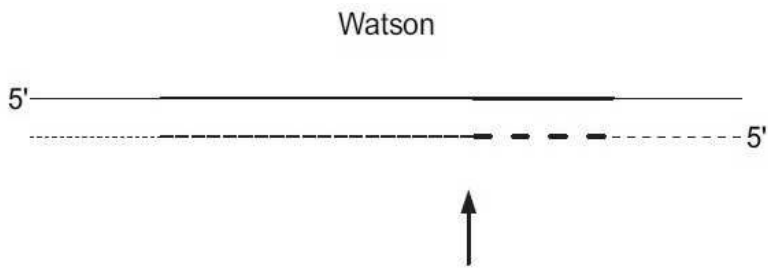
도면17



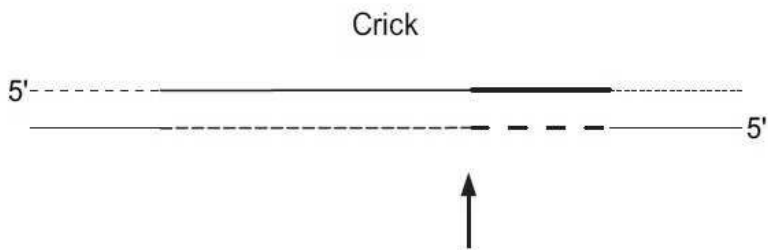
도면18



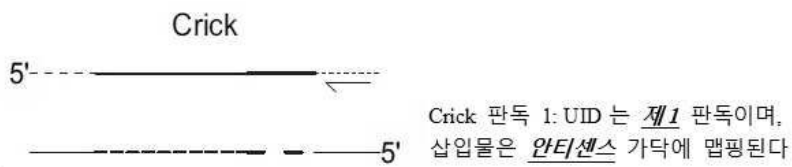
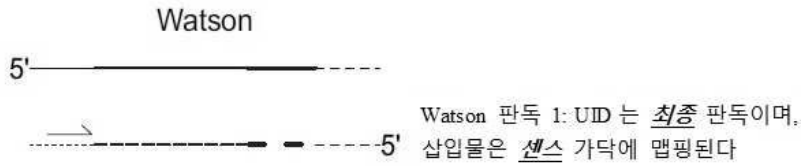
도면19



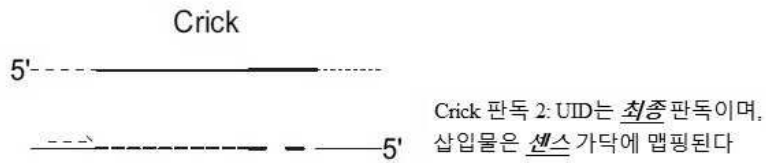
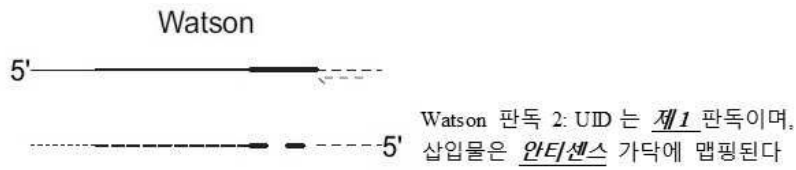
동일한 내생성 UID(분자 단부)와 외생성 UID



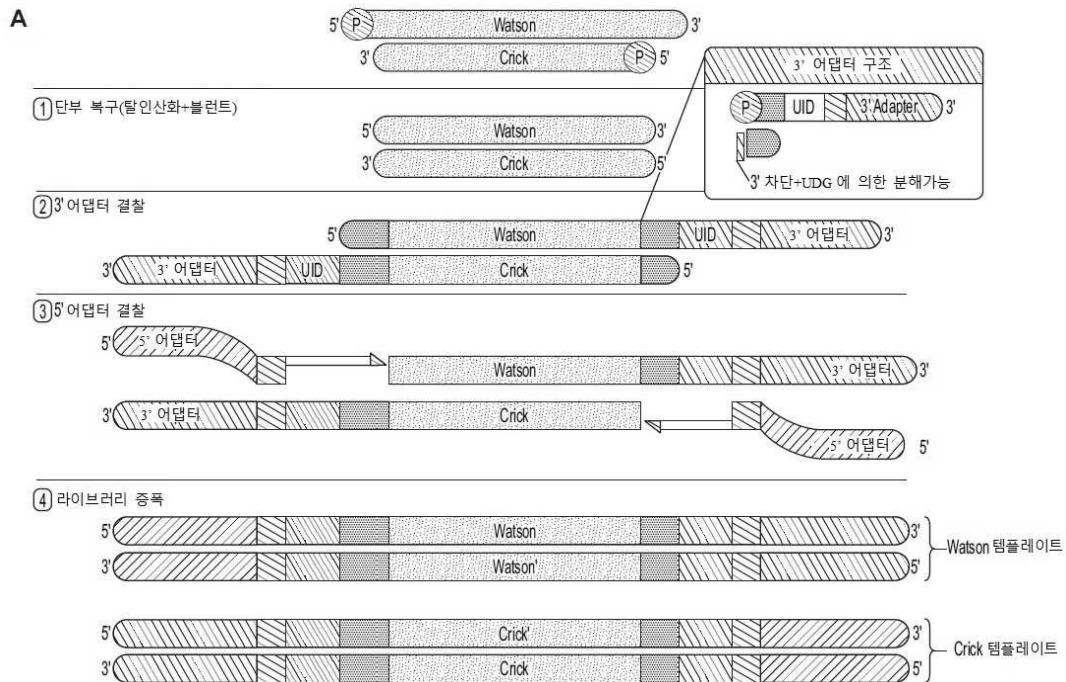
도면20



도면21

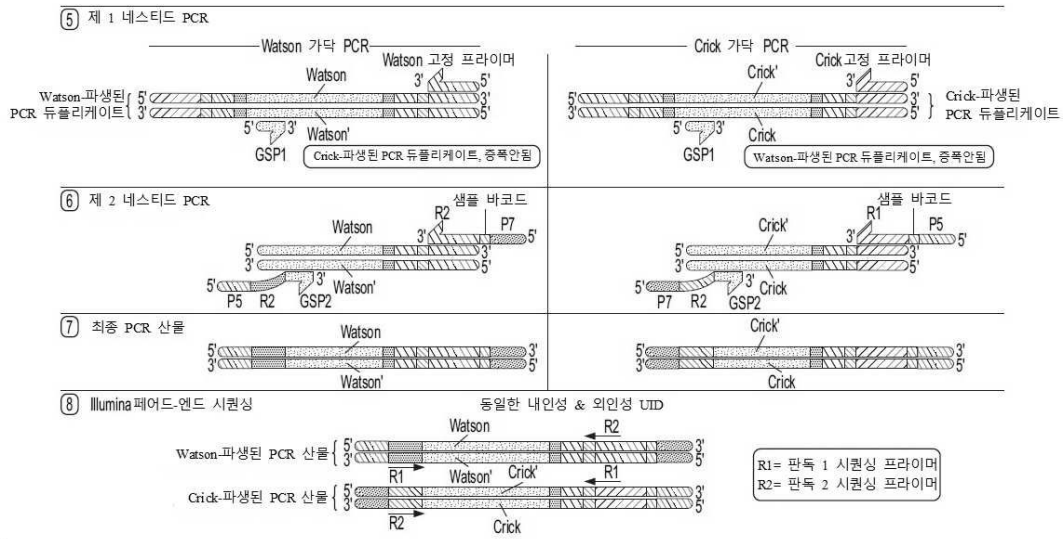


도면22a



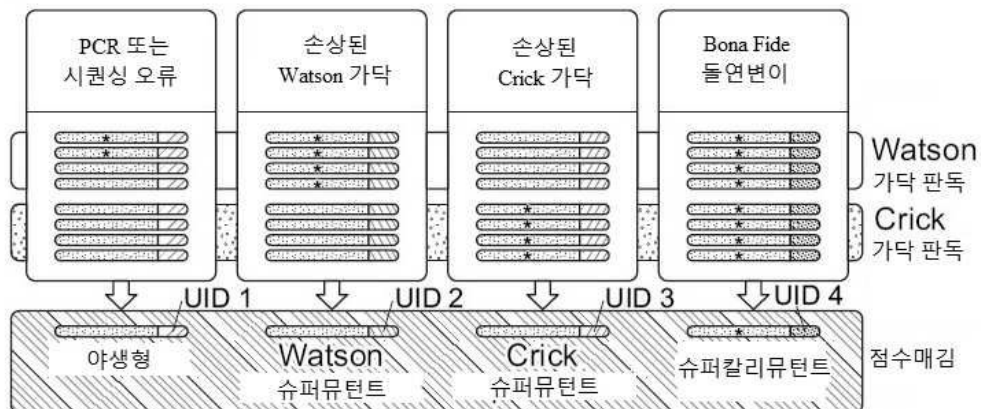
도면22b

B

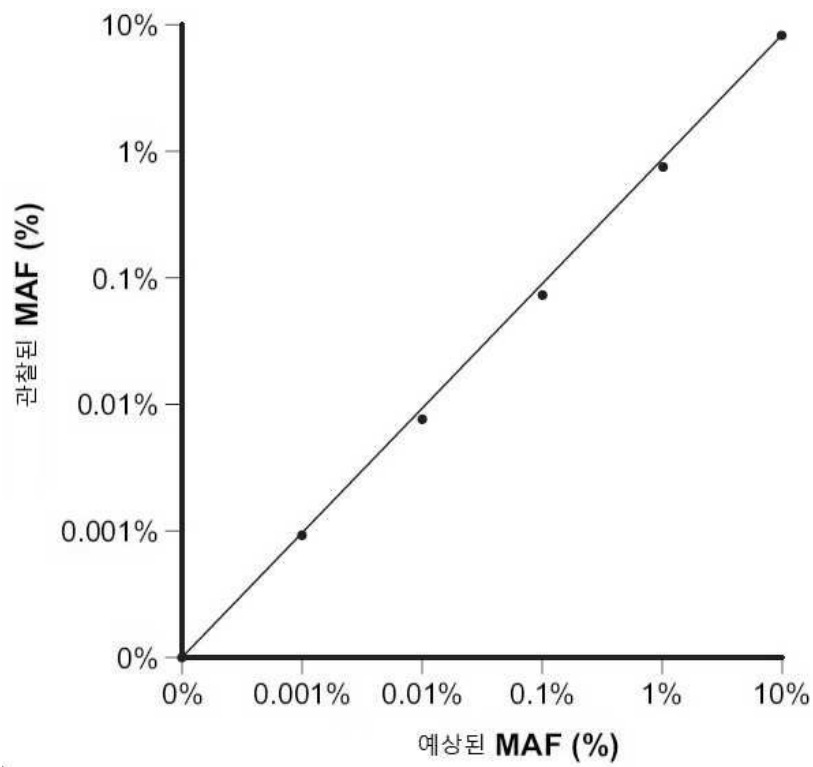


도면22c

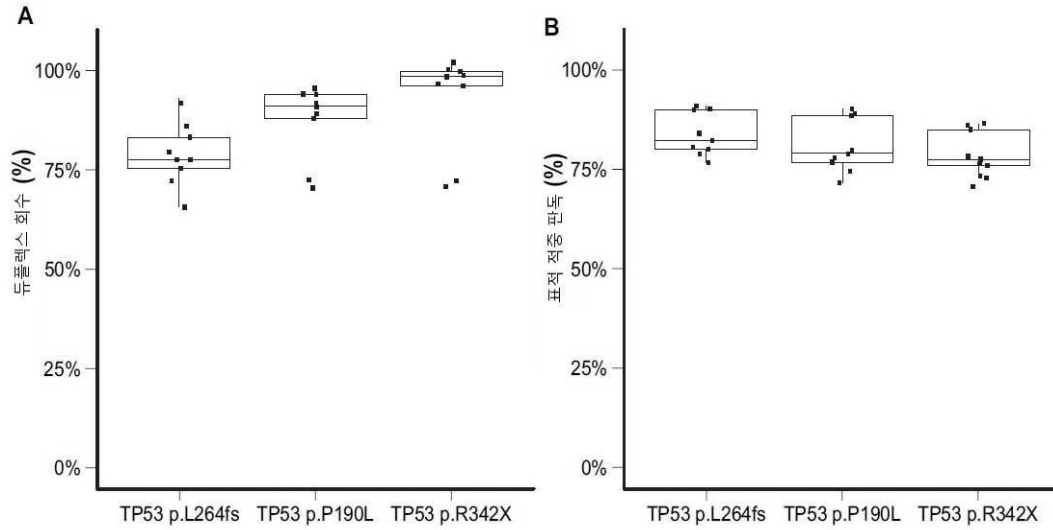
C



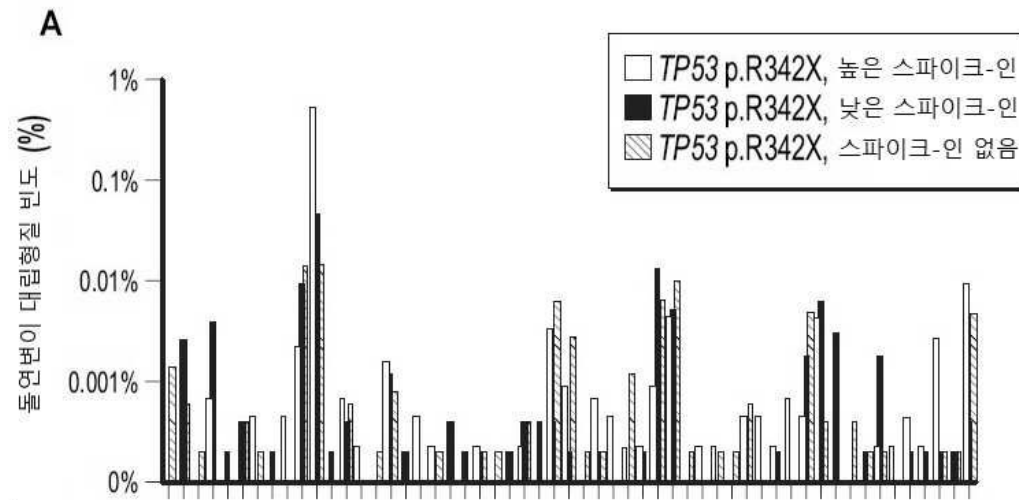
도면23



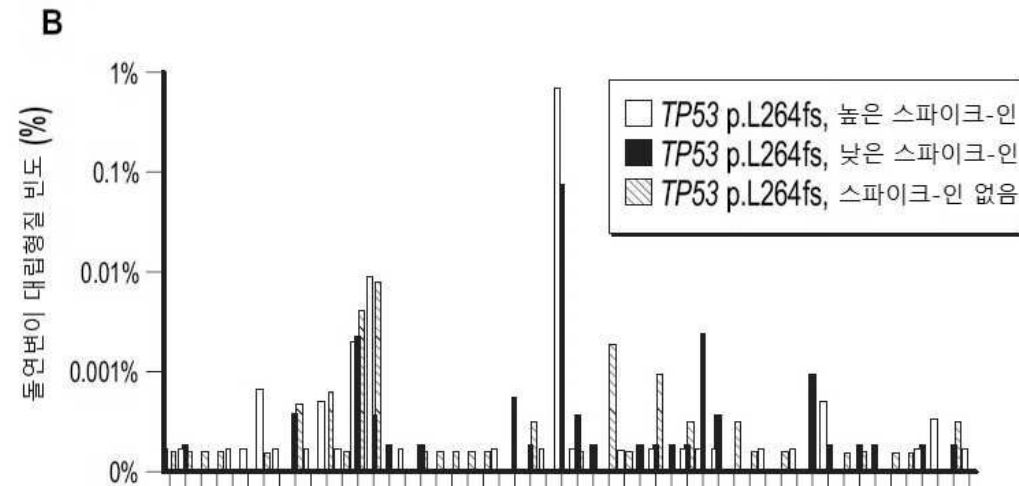
도면24



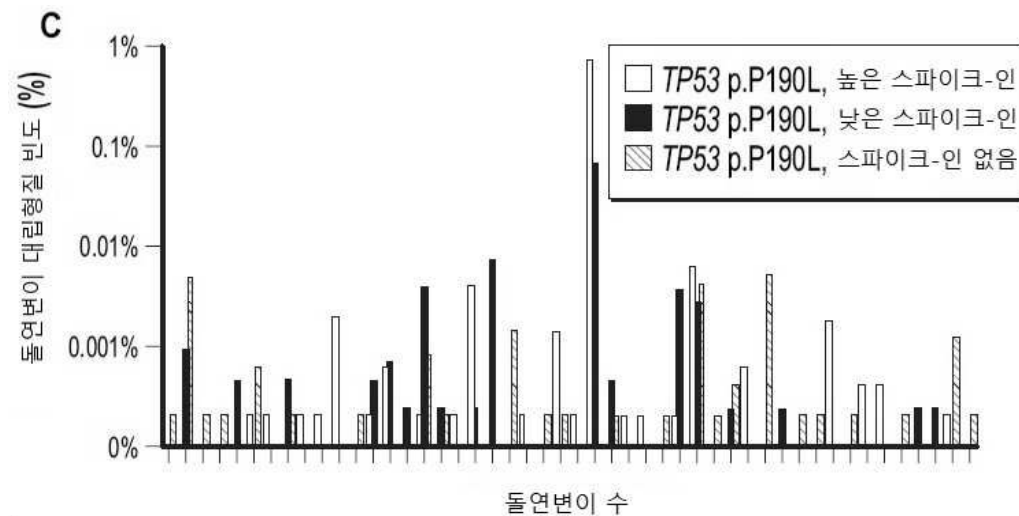
도면25a



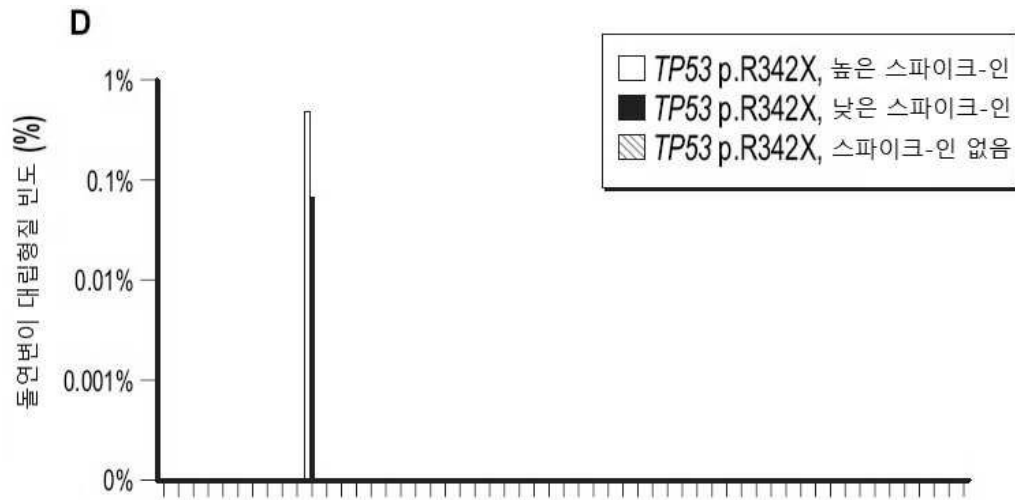
도면25b



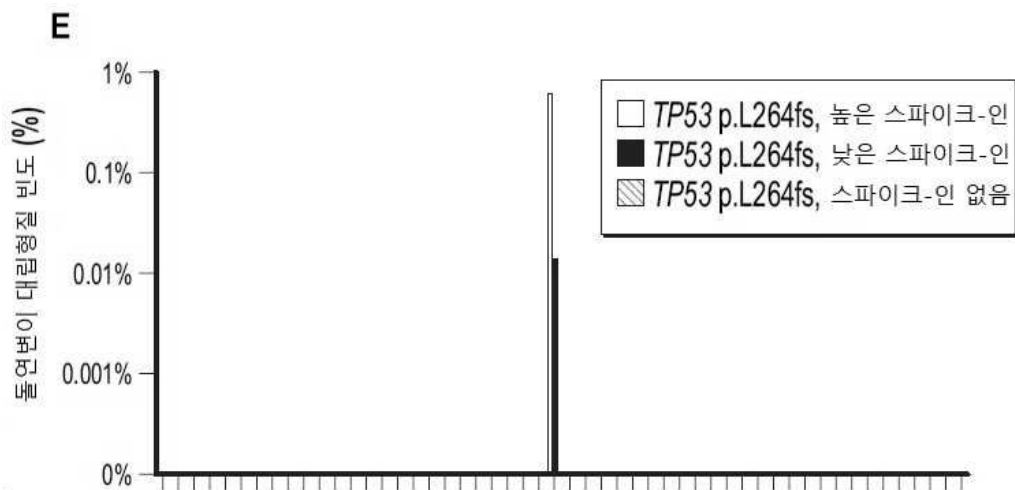
도면25c



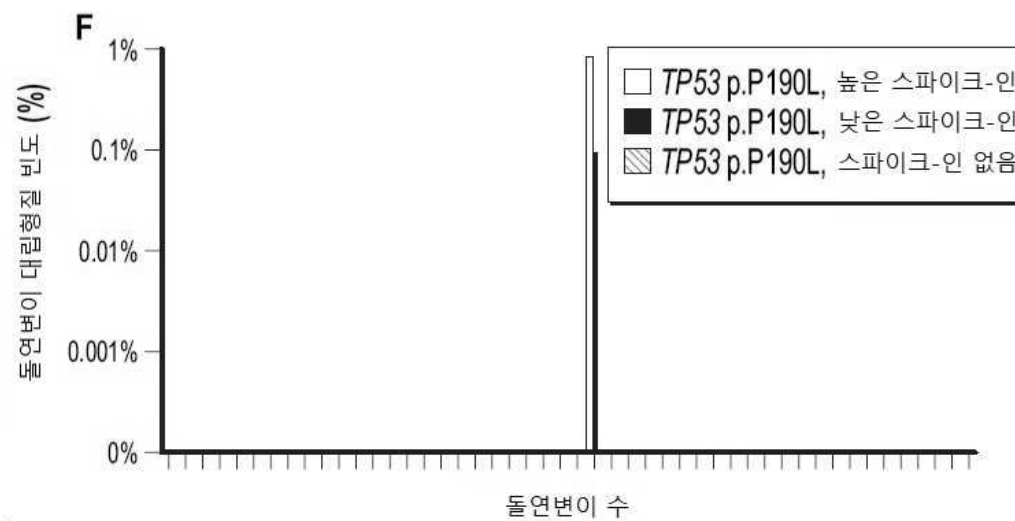
도면25d



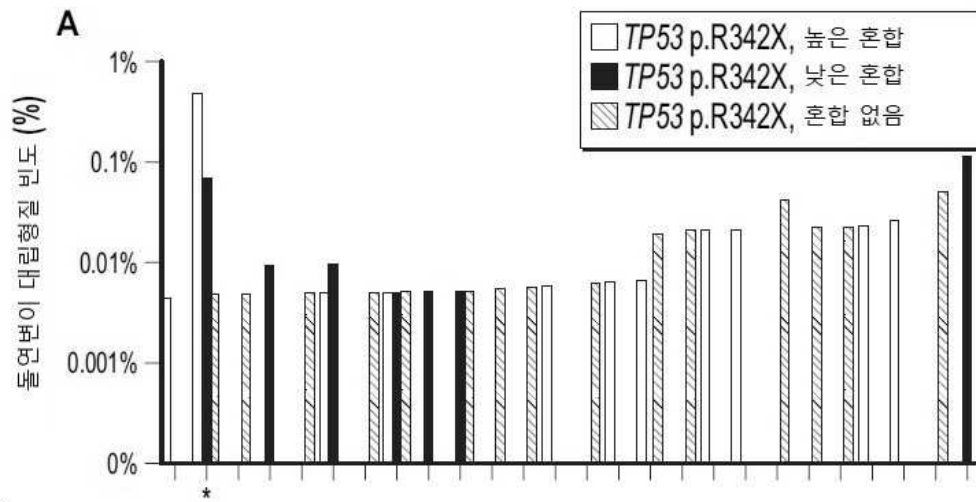
도면25e



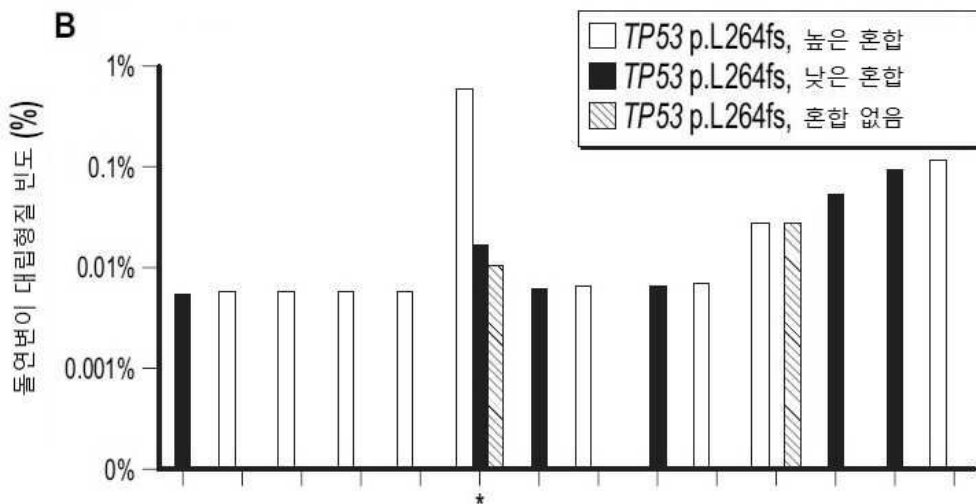
도면25f



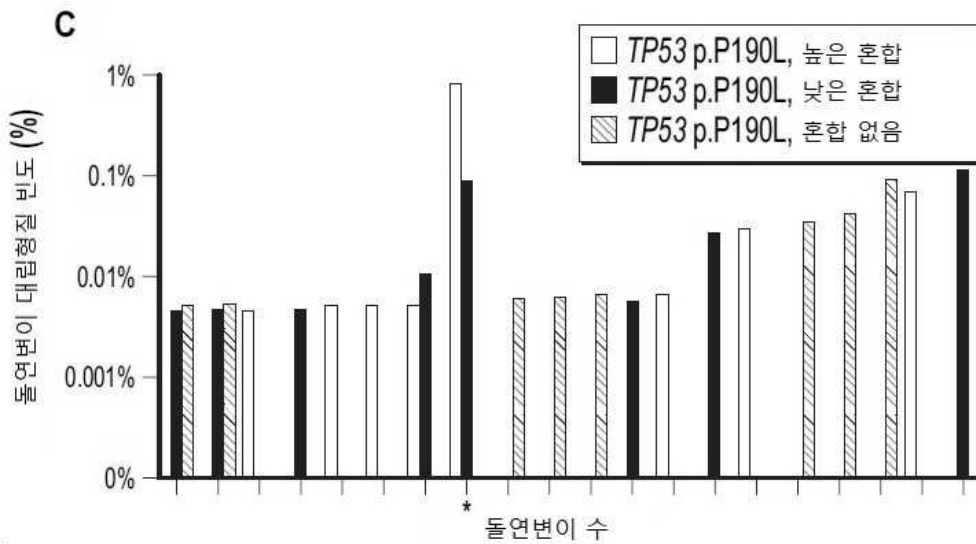
도면26a



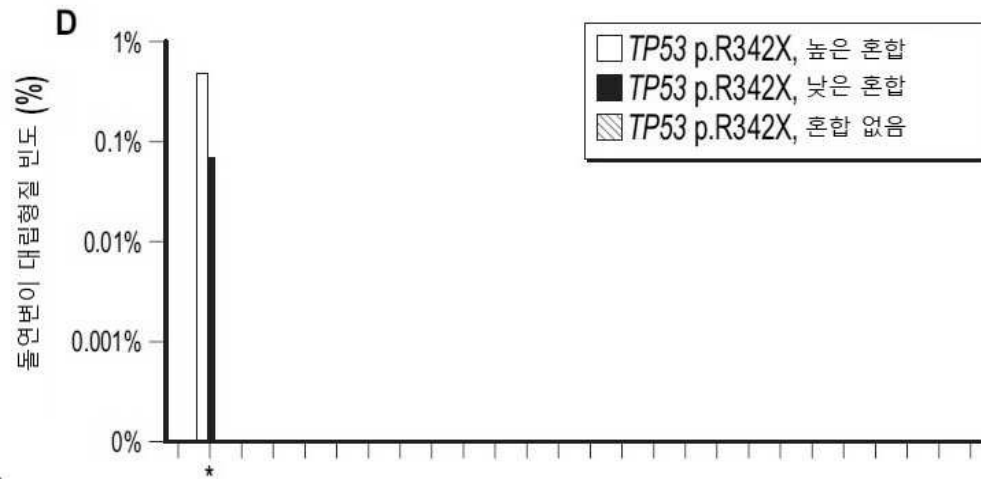
도면26b



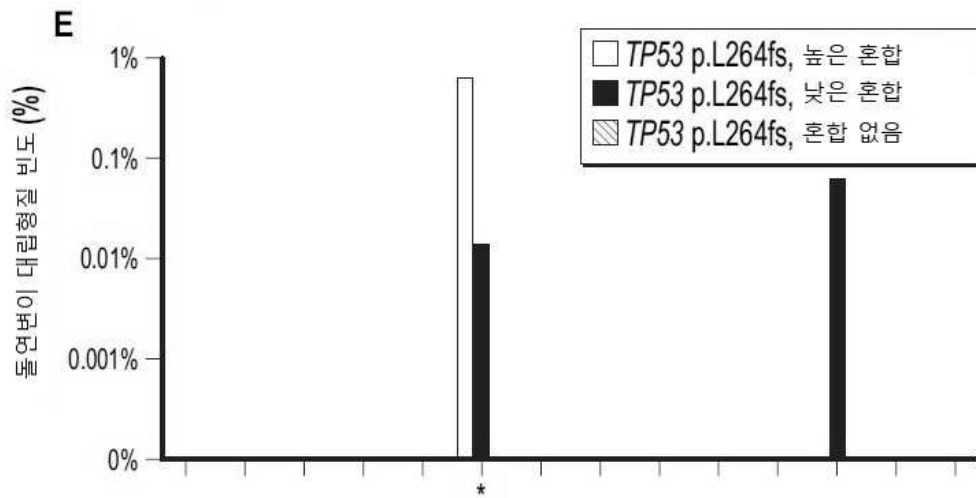
도면26c



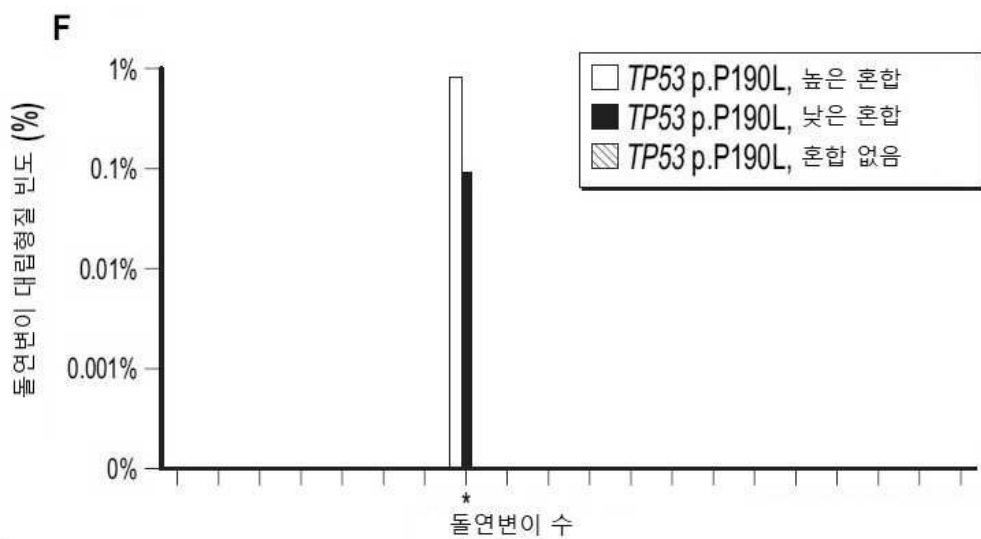
도면26d



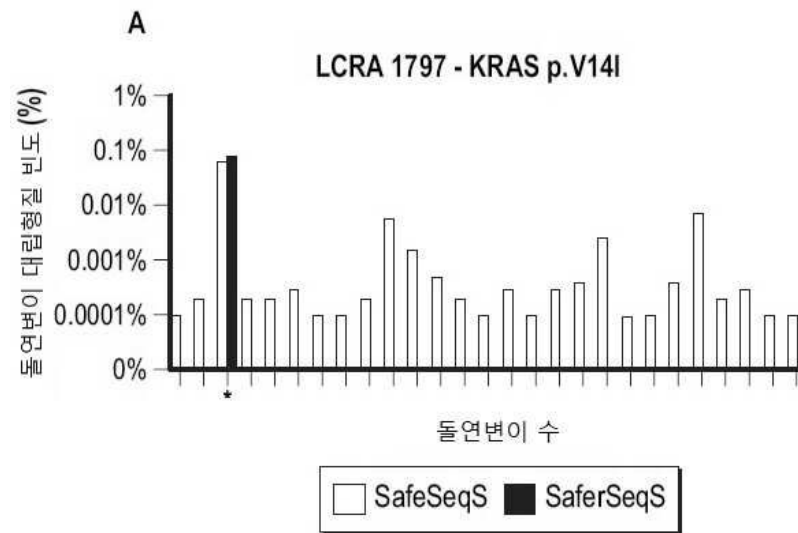
도면26e



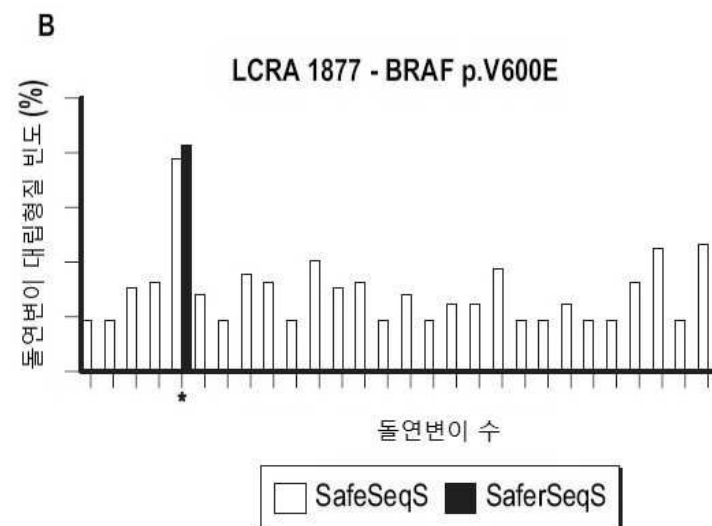
도면26f



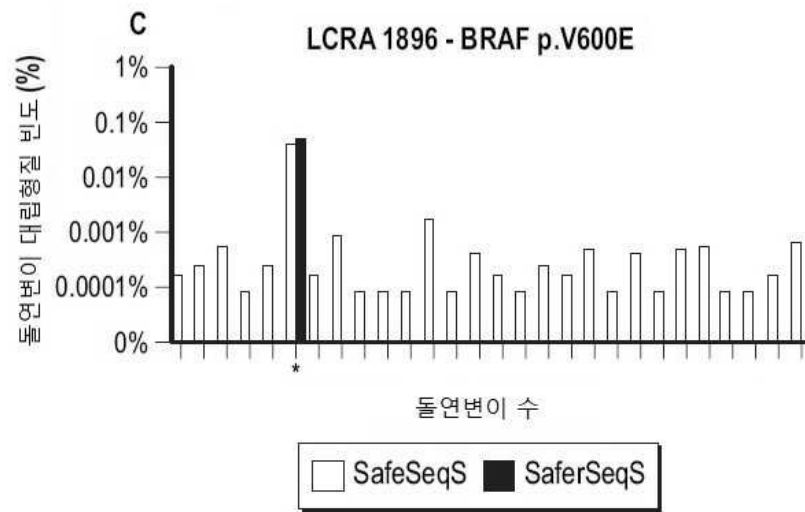
도면27a



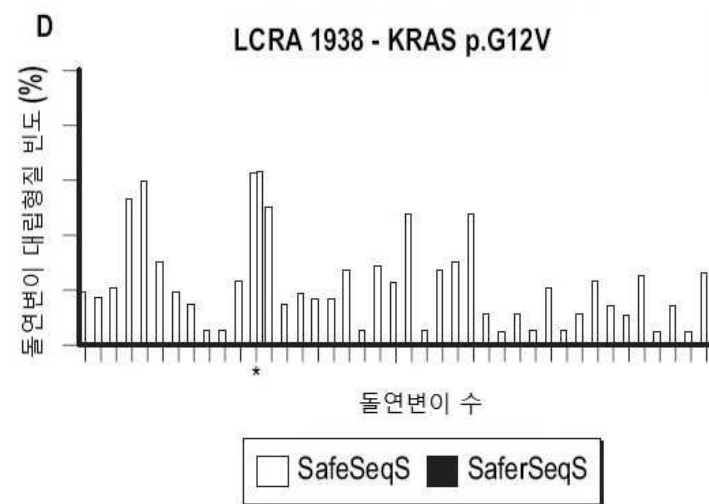
도면27b



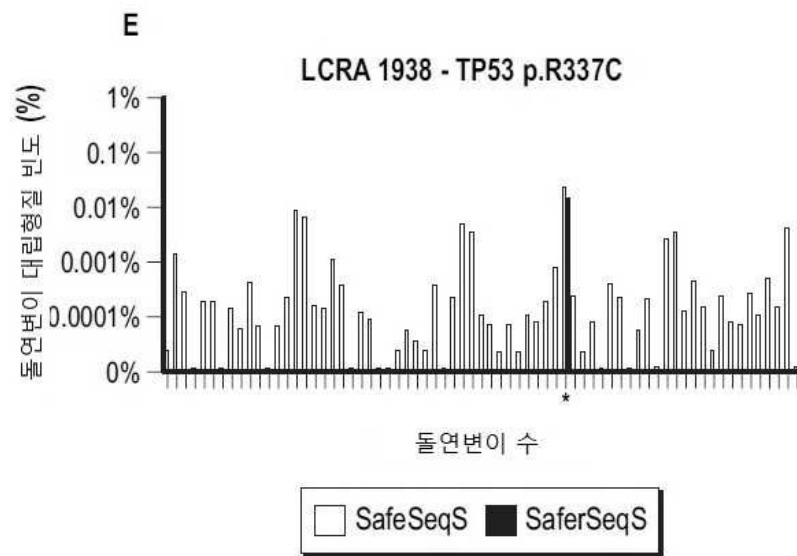
도면27c



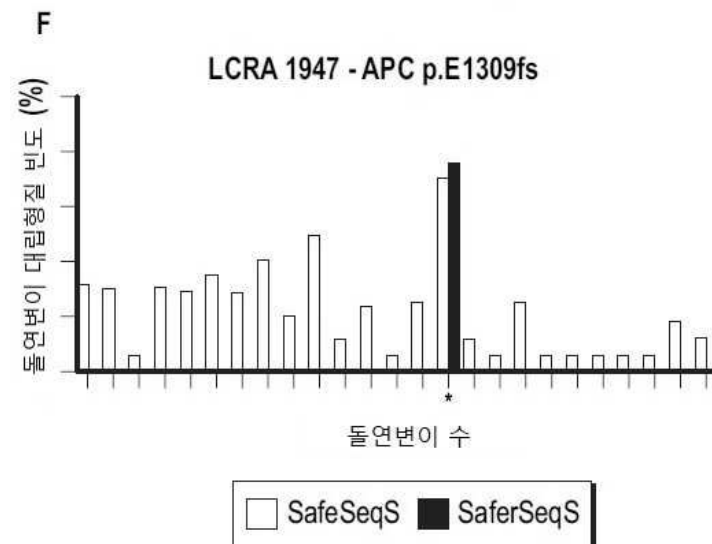
도면27d



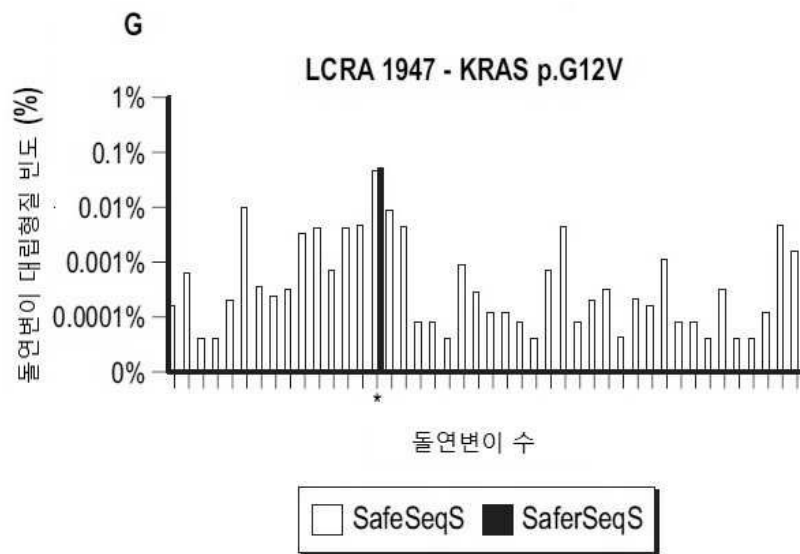
도면27e



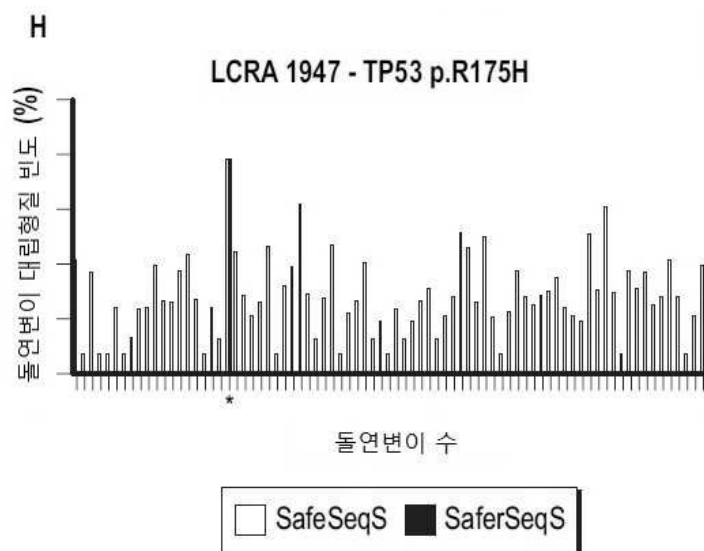
도면27f



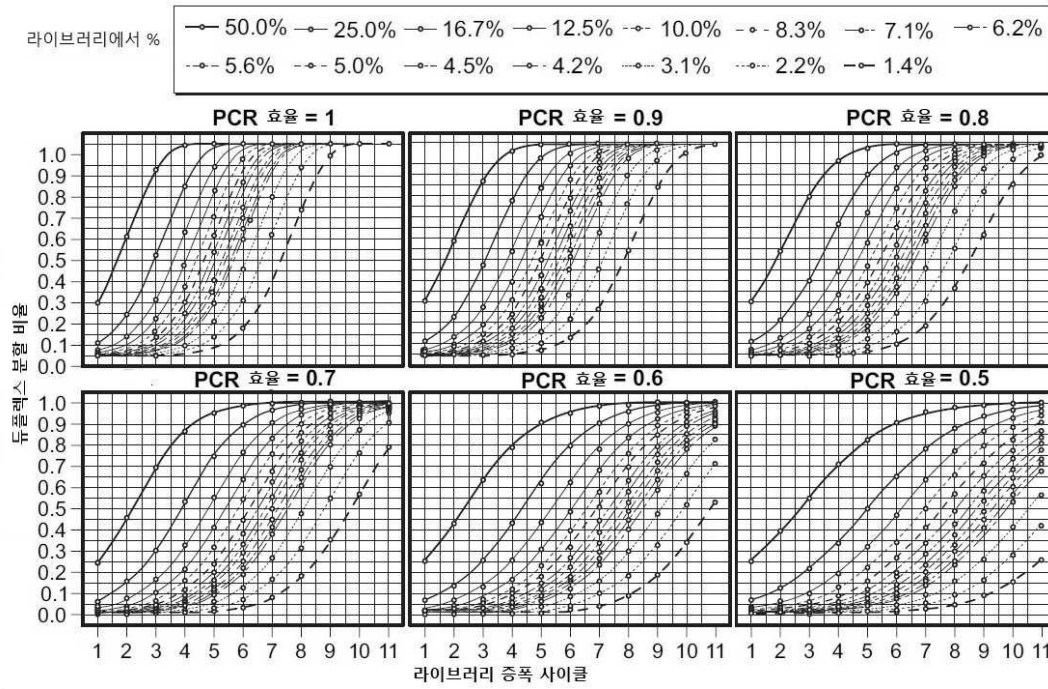
도면27g



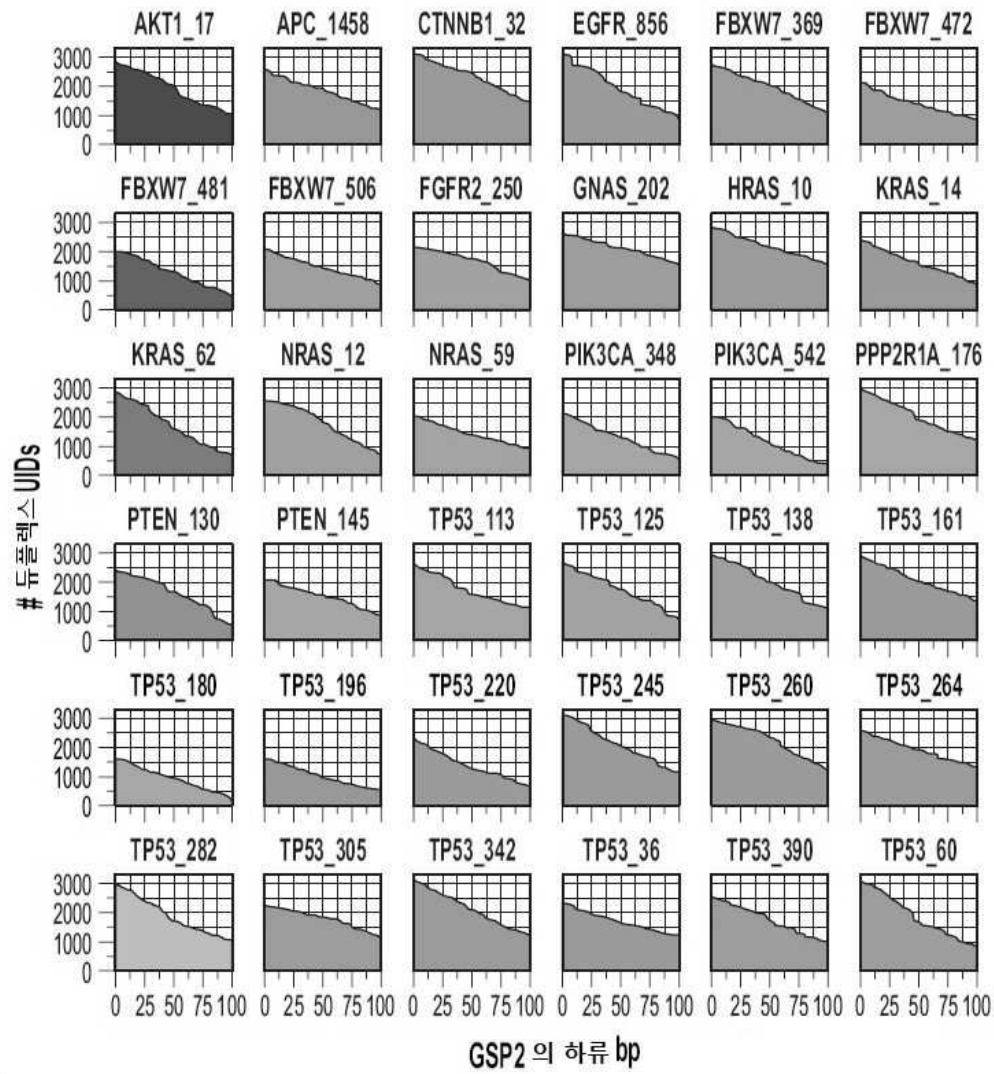
도면27h



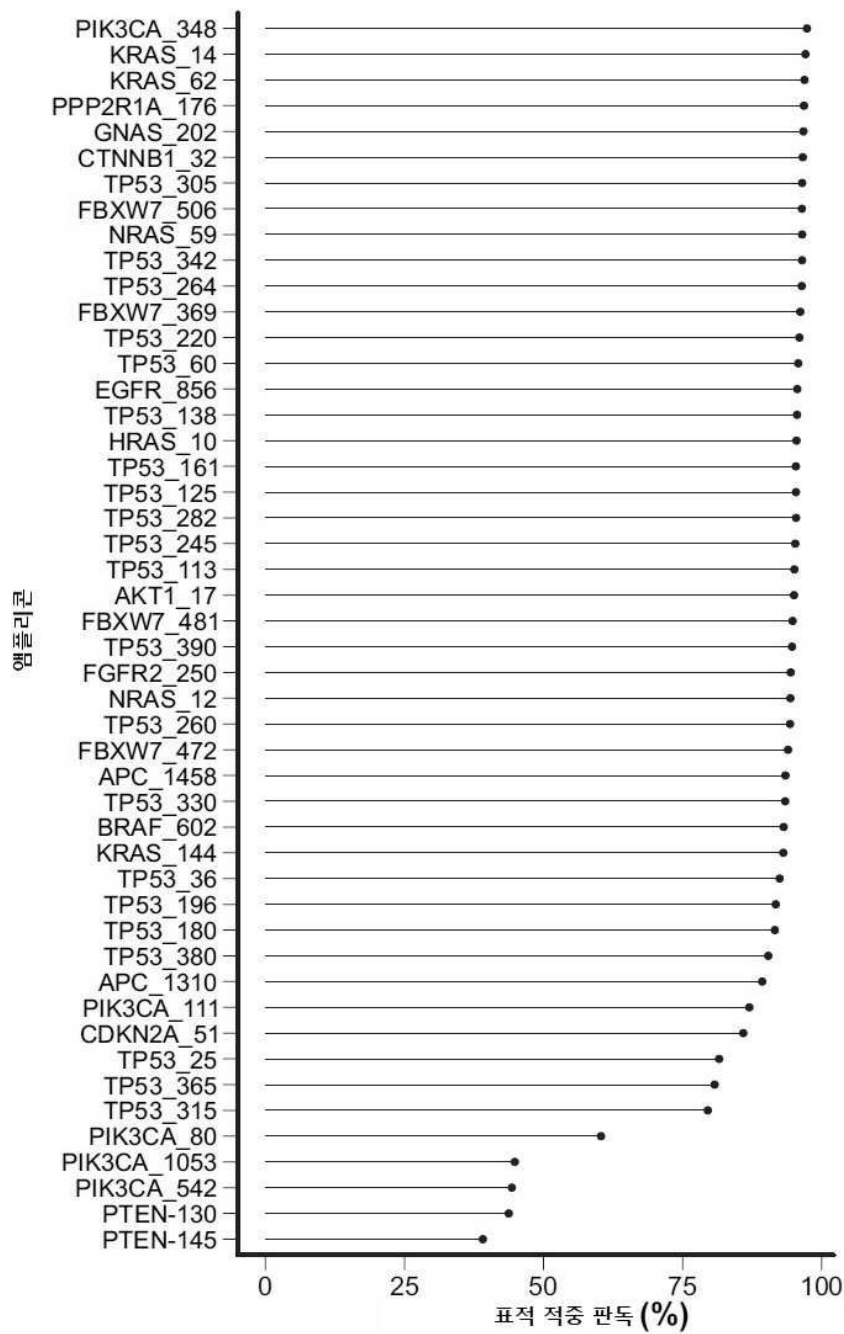
도면28



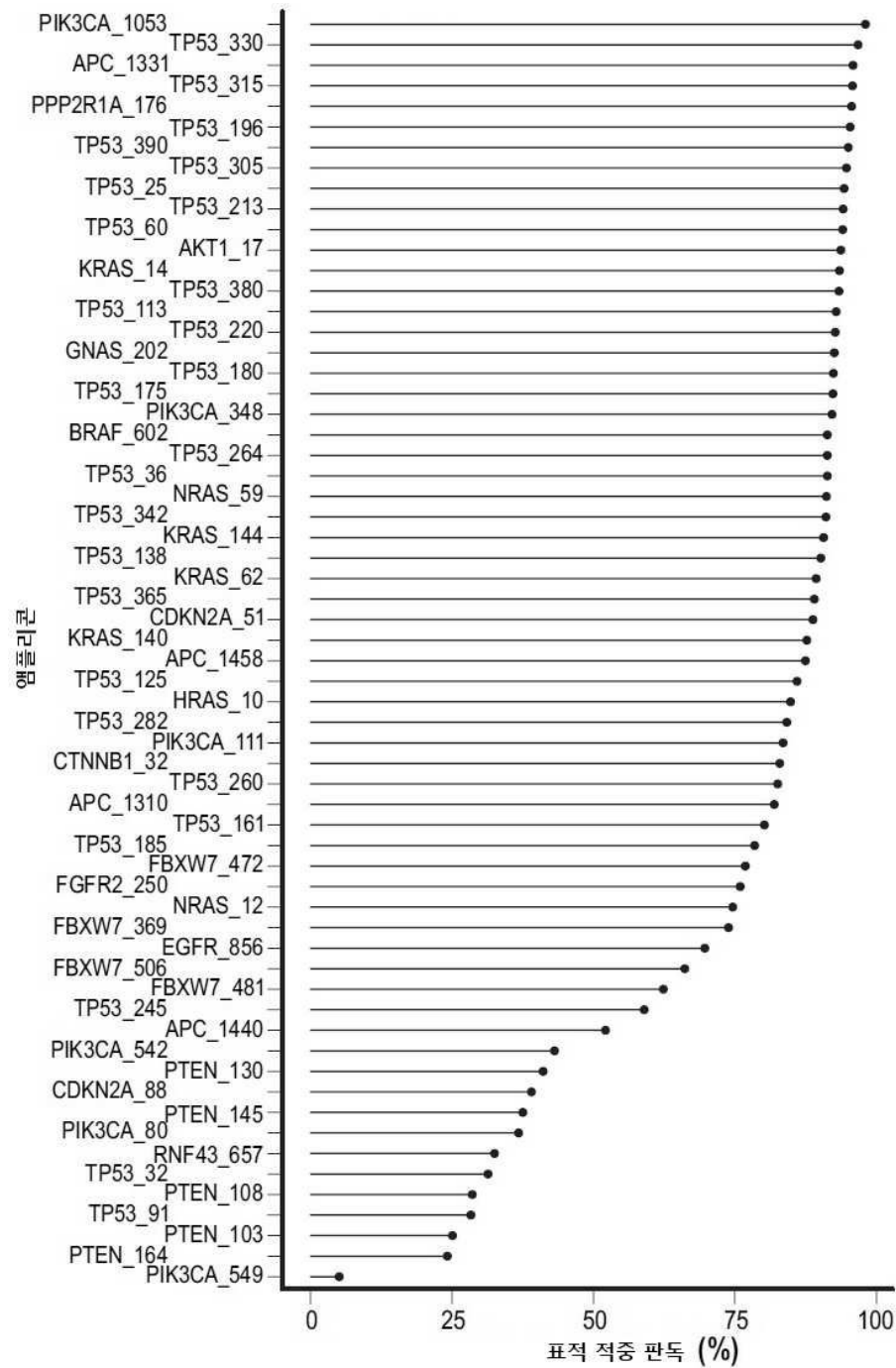
도면29



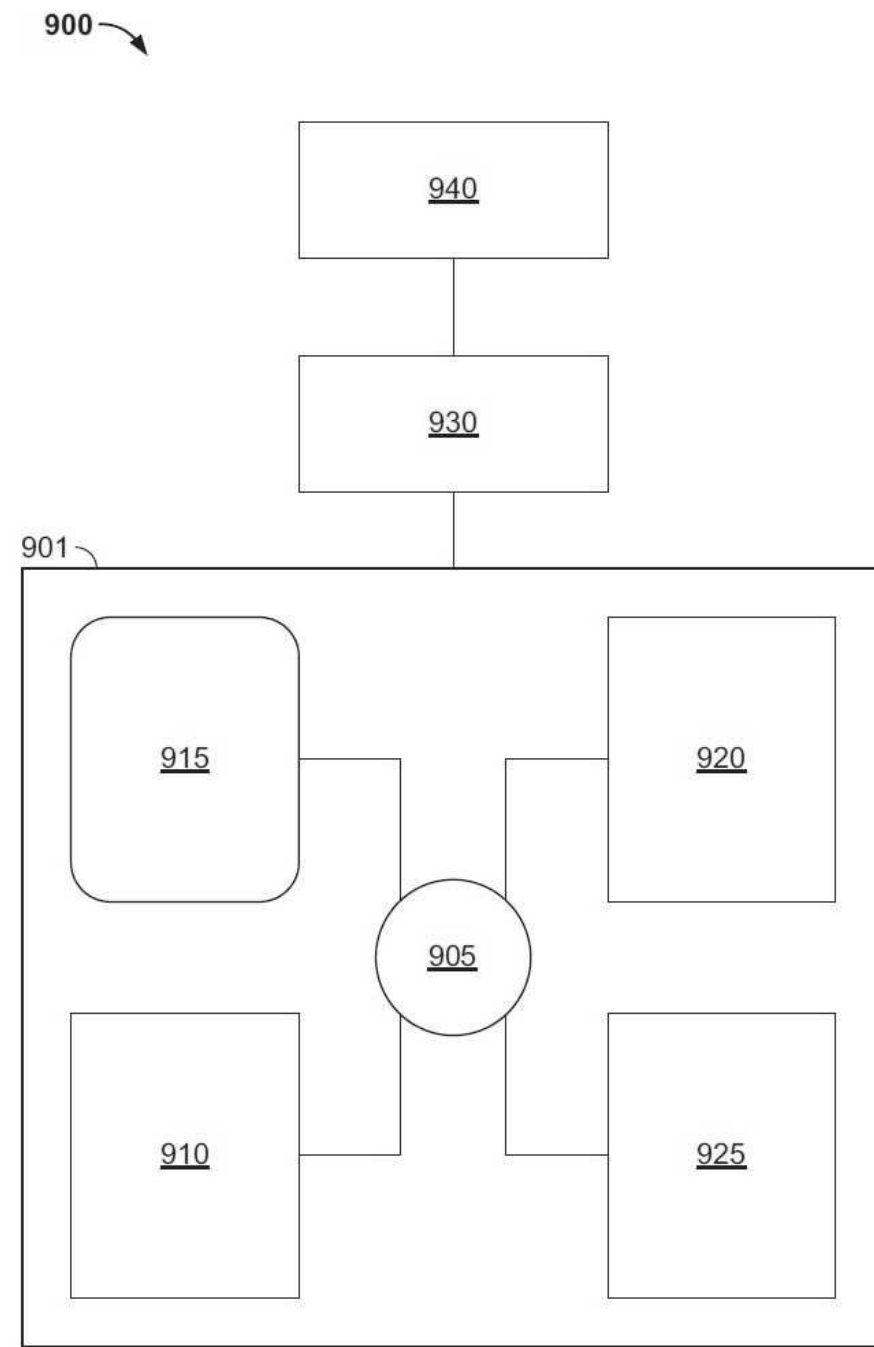
도면30



도면31



도면32



서열 목록

SEQUENCE LISTING

<110> The Johns Hopkins University

<120> METHODS AND MATERIALS FOR ASSESSING NUCLEIC ACIDS

<130> 44807-0341W01

<140> PCT/US2021/017937

<141> 2021-02-12

<150> 62/977,066

<151> 2020-02-14

<160> 117

<170> PatentIn version 3.5

<210> 1

<211> 59

<212> DNA

<213> Artificial

<220><223> 3' adapter oligonucleotide having a 5' phosphate and a molecular barcode

<220><221> misc_feature

<222> (57)..(58)

<223> Residues 57-58 are connected by a phosphorothioate bond

<220><221> misc_feature

<222> (58)..(59)

<223> Residues 58-59 are connected by a phosphorothioate bond

<220><221> misc_feature

<222> (14)..(27)

<223> Residues 14-27 are a molecular barcode

<400> 1

ataaaacgac ggcnnnnnnn nnnnnnnnaga tcggaagagc acacgtctga actccagtc 59

<210> 2

<211> 14

<212> DNA

<213> Artificial

<220><223> Molecular barcode

<220><221> misc_feature

<222> (1)..(14)

<223> n is a, c, g, or t

<400> 2

nnnnnnnnnn nnnn 14

<210> 3

<211> 13

<212> DNA

<213> Artificial

<220><223> 3' adapter oligonucleotide having a blocked 3' group

<220><221> misc_feature

<222> (13)..(13)

<223> residue 13 is a 3'-deoxythymidine

<220><221> misc_feature

<222> (5)..(5)

<223> U at residues 5, 8, 9, 10, 11 is a deoxyuridine

<220><221> misc_feature

<222> (8)..(8)

<223> U at residues 5, 8, 9, 10, 11 is a deoxyuridine

<220><221> misc_feature

<222> (9)..(9)

<223> U at residues 5, 8, 9, 10, 11 is a deoxyuridine

<220><221> misc_feature

<222> (10)..(10)

<223> U at residues 5, 8, 9, 10, 11 is a deoxyuridine

<220><221> misc_feature

<222> (11)..(11)

<223> U at residues 5, 8, 9, 10, 11 is a deoxyuridine

<400> 3

gccgucguuu uat

13

<210> 4

<211> 33

<212> DNA

<213> artificial

<220><223> oligonucleotide primer

<400> 4

acactctttc cctacacgac gctcttccga tct

33

<210> 5

<211> 32

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer
 <400> 5
 gactggagtt cagacgtgtg ctcttcgat ct 32
 <210> 6
 <211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 6
 tgcgtggctc tcaccac 17
 <210> 7
 <211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 7
 caccgcacg tctgtag 17
 <210> 8
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 8
 acaggatctt cagctgacct 20
 <210> 9
 <211> 21
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 9
 cttcagctga cctagttcca a 21
 <210> 10
 <211> 21
 <212> DNA

<213> Artificial	
<220><223> oligonucleotide primer	
<400> 10	
cctggctctg accattctgt t	21
<210> 11	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 11	
ccattctgtt ctctctggca	20
<210> 12	
<211> 18	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 12	
aacaccgcag catgtcaa	18
<210> 13	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 13	
gcagcatgtc aagatcacag a	21
<210> 14	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 14	
ttgtgcagag ttcagttacc ttag	24
<210> 15	

<211> 27
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 15
 gagttcagtt accttaggag atttgag 27
 <210> 16
 <211> 21
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 16
 ggcctgtctc aatatcccaa a 21

 <210> 17
 <211> 24
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 17
 gtctcaatat cccaaaccct aaga 24
 <210> 18
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 18
 ggaacttggt ctcaaagatt cca 23
 <210> 19
 <211> 24
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 19

ggtctcaaag attccagaag tcag	24
<210> 20	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 20	
tctgaattag ctgtatcgtc aagg	24
<210> 21	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 21	
tatcgtaag gcactcttgc c	21
<210> 22	
<211> 31	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 22	
ctgtatttat ttcagtgtta cttacctgtc t	31
<210> 23	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 23	
agtgttactt acctgtcttg tctt	24
<210> 24	
<211> 21	
<212> DNA	
<213> Artificial	

<220><223> oligonucleotide primer	
<400> 24	
gactgagtac aaactggtgg t	21
<210> 25	
<211> 19	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 25	
aaactggtgg tggttggag	19
<210> 26	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 26	
tttcgtaagt gttactcaag aagc	24
<210> 27	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 27	
agtgttactc aagaagcaga aagg	24
<210> 28	
<211> 27	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 28	
gcatcagcat ttgactttac cttatca	27
<210> 29	
<211> 28	

<212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 29
 catttgactt taccttatca atgtctcg 28
 <210> 30
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 30
 gcatgctgtt taattgtgtg ga 22
 <210> 31
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 31
 tgtgtggaag atccaatcca ttt 23

 <210> 32
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 32
 ccatagaaat ctagggcctc tt 22
 <210> 33
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 33
 tagggcctct tgtgccttta 20
 <210> 34

<211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 34
 ggccctgcct tccaatg 17

<210> 35
 <211> 19
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 35
 ttccaatgga tccactcac 19

<210> 36
 <211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 36
 gcagcctctg gcattct 17

<210> 37
 <211> 19
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 37
 ggcattctgg gagcttcat 19

<210> 38
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 38

gccaggcatt gaagtctcat	20
<210> 39	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 39	
attgaagtct catggaagcc a	21
<210> 40	
<211> 18	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 40	
ccgtcatgtg ctgtgact	18
<210> 41	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 41	
gtgctgtgac tgctttaga	20
<210> 42	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 42	
ccaaatactc cacacgcaa t	21
<210> 43	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	

<400> 43	
cacacgcaaa ttccttcca	20
<210> 44	
<211> 22	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 44	
ggagtcttcc agtgtgatga tg	22
<210> 45	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 45	
gtgatgatgg tgaggatggg	20
<210> 46	
<211> 18	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 46	
cacaaacacg cacctcaa	18
<210> 47	
<211> 19	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 47	
acgcacctca aagctgttc	19
<210> 48	
<211> 18	
<212> DNA	

<213> Artificial

<220><223> oligonucleotide primer

<400> 48

aactgcaccc ttggtctc 18

<210> 49

<211> 18

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 49

cctccaccgc ttcttgtc 18

<210> 50

<211> 25

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 50

gaaactttcc acttgataag aggtc 25

<210> 51

<211> 22

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 51

tgataagagg tccaagact ta 22

<210> 52

<211> 17

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 52

agtagggcca ggaaggg 17

<210> 53

<211> 17	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 53	
gggctgaggt cactcac	17
<210> 54	
<211> 17	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 54	
tgggaggctg tcagtgg	17
<210> 55	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 55	
gggaacaaga agtggagaat g	21
<210> 56	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 56	
tgcttaggtc cactctctct	20
<210> 57	
<211> 23	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 57	
gtccactctc tctcttttca gca	23

<210> 58
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 58
 acaactgttc aaactgatgg ga 22

<210> 59
 <211> 18
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 59
 atgggaccca ctccatcg 18

<210> 60
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 60
 ttagtcactg gcagcaacag 20

<210> 61
 <211> 19
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 61
 ggcagcaaca gtcttacct 19

<210> 62
 <211> 21
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer

<400> 62	
tagaggaaga agtcccaacc a	21
<210> 63	
<211> 22	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 63	
gaagtcccaa ccatgacaag at	22
<210> 64	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 64	
ctaacaaccc tctgccatc	20
<210> 65	
<211> 22	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 65	
ccctcctgcc atcatattga ac	22
<210> 66	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 66	
actcctcctt tcttcctct	20
<210> 67	
<211> 21	
<212> DNA	
<213> Artificial	

<220><223> oligonucleotide primer
 <400> 67
 tttcttcct ctctccacca g 21

<210> 68
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 68
 ggagcgatga cggaatataa gc 22
 <210> 69
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 69
 ggaatataag ctggtggtgg tg 22
 <210> 70
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 70
 ctcatgtact ggtccctcat tg 22

<210> 71
 <211> 21
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 71
 gtcctcatt gcactgtact c 21
 <210> 72
 <211> 22

<212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 72
 gtgaaacctg tttgttggac at 22
 <210> 73
 <211> 25
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 73
 cctgtttgtt ggacatactg gatac 25

 <210> 74
 <211> 18
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 74
 cacggttgcc tactggtt 18
 <210> 75
 <211> 24
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 75
 gttgcctact ggttcaatta cttt 24
 <210> 76
 <211> 25
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 76
 gaacagctca aagcaatttc tacac 25

<210> 77	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 77	
aatttctaca cgagatcctc tctc	24
<210> 78	
<211> 19	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 78	
cgactcccag gtacttccg	19
<210> 79	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 79	
tacttccgga acctgtgctc	20
<210> 80	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 80	
catgttgcag caattcactg t	21
<210> 81	
<211> 24	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 81	

caattcactg taaagctgga aagg	24
<210> 82	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 82	
gggacagcat caaatcatcc a	21
<210> 83	
<211> 22	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 83	
atcaaatcat ccattgcttg gg	22
<210> 84	
<211> 19	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 84	
gtcacagact ttggtgtcc	19
<210> 85	
<211> 18	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 85	
ttggctgtcc cagaatgc	18
<210> 86	
<211> 18	
<212> DNA	
<213> Artificial	

<220><223> oligonucleotide primer
 <400> 86
 atcaaccac agctgcac 18
 <210> 87
 <211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 87
 tgcacagggc aggtctt 17
 <210> 88
 <211> 17
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 88
 ccagctgctc accatcg 17
 <210> 89
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 89
 tcaccatcg tatctgagca 20
 <210> 90
 <211> 19
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 90
 tgacaaccac ccttaaccc 19
 <210> 91
 <211> 21
 <212> DNA

<213> Artificial	
<220><223> oligonucleotide primer	
<400> 91	
cagttgcaaa ccagacctca g	21
<210> 92	
<211> 17	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 92	
gtgcagggtg gcaagtg	17
<210> 93	
<211> 17	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 93	
ggcaagtggc tctgac	17
<210> 94	
<211> 22	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 94	
cctttcttgc ggagattctc tt	22
<210> 95	
<211> 21	
<212> DNA	
<213> Artificial	
<220><223> oligonucleotide primer	
<400> 95	
cggagattct cttcctctgt g	21
<210> 96	

<211> 24
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 96
 aaatattctc catccagtgg ttcc 24
 <210> 97
 <211> 21
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 97
 ccagtgggtt cttctttggc t 21

 <210> 98
 <211> 19
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 98
 gggcatcctt gatttccaa 19
 <210> 99
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 99
 cttgagttcc aaggcctcat 20
 <210> 100
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223> oligonucleotide primer
 <400> 100

tcaggccctt ctgtcttgaa 20

<210> 101

<211> 23

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 101

ccttctgtct tgaacatgag ttt 23

<210> 102

<400> 102

000

<210> 103

<400> 103

000

<210> 104

<211> 33

<212> DNA

<213> Artificial

<220><223> 5' adapter oligonucleotide

<400> 104

acactctttc cctacacgac gctcttccga tct 33

<210> 105

<211> 33

<212> DNA

<213> Artificial

<220><223> 5' adapter oligonucleotide

<400> 105

acactctttc cctacacgac gctcttccga tct 33

<210> 106

<211> 32

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 106
gactggagtt cagacgtgtg ctcttccgat ct 32
<210> 107
<400> 107
000
<210> 108
<211> 19
<212> DNA
<213> Artificial
<220><223> oligonucleotide primer
<400> 108
gactggagtt cagacgtgt 19
<210> 109
<211> 20

<212> DNA
<213> Artificial
<220><223> oligonucleotide primer
<400> 109
acactctttc cctacacgac 20
<210> 110
<211> 58
<212> DNA
<213> Artificial
<220><223> Watson second nested gene-specific PCR primer
<220><221> misc_feature
<222> (30)..(37)
<223> n at residues 30-37 are an index sequence
<220><221> misc_feature
<222> (58)..(58)
<223> residue 58 is conjugated to a sequence of gene-specific portion of the oligonucleotide

<400> 110
aatgatacgg cgaccaccga gatctacacn nnnnnnncga cgtaaacga cggccagt 58
<210> 111

<211> 57
 <212> DNA
 <213> Artificial
 <220><223> Crick second nested gene-specific PCR primer
 <220><221> misc_feature
 <222> (25)..(32)
 <223> n at residues 25-32 are an index sequence
 <220><221> misc_feature
 <222> (58)..(58)
 <223> residue 58 is conjugated to a sequence of gene-specific portion of the oligonucleotide
 <400> 111
 caagcagaag acggcatacg agatnnnnnn nncacacagg aaacagctat gaccatg 57

<210> 112
 <211> 66
 <212> DNA
 <213> Artificial
 <220><223> Watson second nested anchor and indexing primer
 <220><221> misc_feature
 <222> (25)..(32)
 <223> n at residues 25-32 are an index sequence
 <400> 112
 caagcagaag acggcatacg agatnnnnnn nngtgactgg agttcagacg tgtgctcttc 60
 cgatct 66

<210> 113
 <211> 70
 <212> DNA
 <213> Artificial
 <220><223> Crick second nested anchor and indexing primer
 <220><221>
 > misc_feature
 <222> (30)..(37)
 <223> n at residues 30-37 are an index sequence
 <400> 113
 aatgatacgg cgaccaccga gatctacacn nnnnnnnaca ctcttttcct acacgacgt 60

cttccgatct 70

<210> 114

<211> 40

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<220><221> misc_feature

<222> (12)..(19)

<223> n is a, c, g, or t

<400> 114

gagatctaca cnnnnnnnnc gacgtaaaac gacggccagt 40

<210> 115

<211> 25

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<400> 115

catggtcata gctgtttcct gtgtg 25

<210> 116

<211> 45

<212> DNA

<213> Artificial

<220><223> oligonucleotide primer

<220><221> misc_feature

<222> (13)..(20)

<223> n is a, c, g, or t

<400> 116

ggcatacgag atnnnnnnnnn cacacaggaa acagctatga ccatg 45

<210> 117

<211> 13

<212> DNA

<213> Artificial

<220><223>

oligonucleotide tag

<400> 117

gccgtcgttt tat

13