



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0065295
(43) 공개일자 2023년05월11일

(51) 국제특허분류(Int. Cl.)
C12N 15/113 (2010.01) C12N 15/10 (2017.01)
(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 15/1065 (2013.01)
(21) 출원번호 10-2023-7011639
(22) 출원일자(국제) 2021년09월08일
심사청구일자 없음
(85) 번역문제출일자 2023년04월05일
(86) 국제출원번호 PCT/US2021/049448
(87) 국제공개번호 WO 2022/055984
국제공개일자 2022년03월17일
(30) 우선권주장
63/075,543 2020년09월08일 미국(US)

(71) 출원인
레졸루션 바이오사이언스, 인크.
미국 98033 워싱턴주 커클랜드 스위트 200 커클랜드
웨이 550
(72) 발명자
구오, 지안난
미국 98033 워싱턴 커클랜드 커클랜드 웨이 550
스위트 200 레졸루션 바이오사이언스 인크. 내
헤르난데즈, 제니퍼
미국 98033 워싱턴 커클랜드 커클랜드 웨이 550
스위트 200 레졸루션 바이오사이언스 인크. 내
(74) 대리인
양영준, 이상남

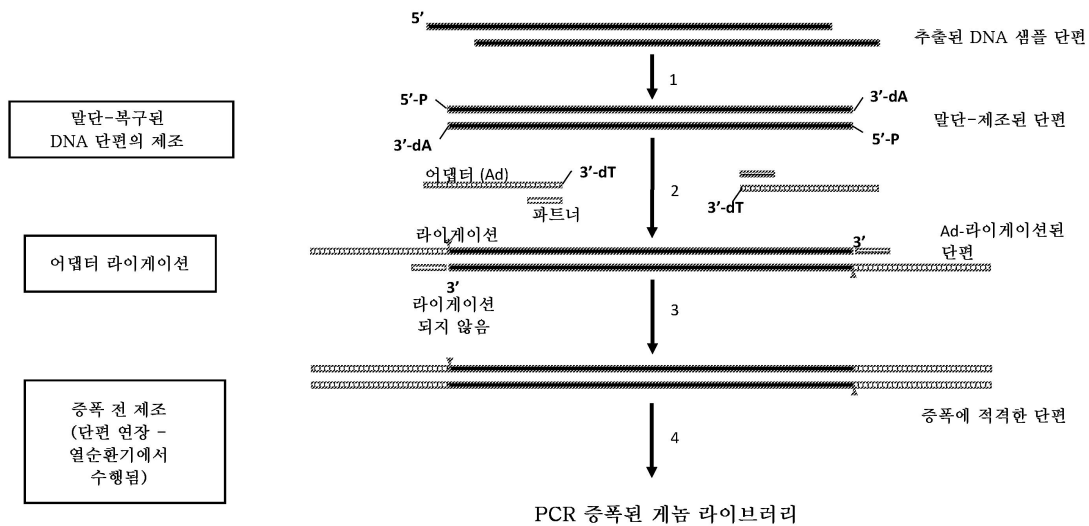
전체 청구항 수 : 총 21 항

(54) 발명의 명칭 유전자 라이브러리 및 유전자 분석의 고효율성 구축을 위한 어댑터 및 방법

(57) 요약

본 개시내용은 세포성 핵산의 멀티플렉스화 검출 및 분석을 위한 조성물 및 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 본 개시내용의 방법에서 사용하기 위한 다기능성 어댑터를 제공한다. 일부 실시양태에서, 본 개시내용의 조성물 및 방법은 자동화가능하다.

대표도



명세서

청구범위

청구항 1

하기를 포함하는 다기능성 어댑터이며:

- a) 라이게이션 가닥 올리고뉴클레오타이드, 및
- b) 라이게이션 가닥 올리고뉴클레오타이드의 3' 말단에 있는 영역에 혼성화하고 그와 듀플렉스를 형성할 수 있는 비-라이게이션 가닥 올리고뉴클레오타이드;

샘플로부터의 dsDNA 단편과 접촉시, 라이게이션 가닥 올리고뉴클레오타이드는 dsDNA 단편의 각각의 가닥의 5' 말단에 라이게이션하고;

라이게이션 가닥 올리고뉴클레오타이드는 하기를 포함하고:

- i) 3' 말단 오버행;
- ii) 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 증폭 영역;
- iii) 고유의 다기능성 ID 영역;
- iv) 고유의 분자 식별자 (UMI) 멀티플라이어; 및
- v) 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보적인 폴리뉴클레오타이드 서열을 포함하는 앵커 영역;

dsDNA 단편은 각각의 가닥의 5' 말단에서 포스페이트 기 및 각각의 가닥의 3' 말단에서 오버행을 포함하고;

다기능성 ID 영역 및 UMI 멀티플라이어의 조합은 dsDNA 단편을 식별하고;

다기능성 ID 영역은 샘플을 식별하는 것인 다기능성 어댑터.

청구항 2

제1항에 있어서,

라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dT 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dA 오버행을 포함하거나;

라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dA 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dT 오버행을 포함하거나;

라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dC 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dG 오버행을 포함하거나; 또는

라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dG 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dC 오버행을 포함하는 것인 다기능성 어댑터.

청구항 3

제1항 또는 제2항에 있어서, 비-라이게이션 가닥 올리고뉴클레오타이드가 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 포함하고, 비-라이게이션 가닥이 듀플렉스로부터 대체될 수 있는 것인 다기능성 어댑터.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

증폭 영역이 25개 뉴클레오타이드 길이이고;

다기능성 ID 영역이 8개 뉴클레오타이드 길이이고;
 UMI 멀티플라이어가 3개 뉴클레오타이드 길이이고;
 앵커 영역이 10개 뉴클레오타이드 길이이고;
 UMI 멀티플라이어다 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있고;
 앵커 영역이 4가지 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.

청구항 5

다기능성 어댑터 및 dsDNA 단편을 포함하는 복합체이며, 다기능성 어댑터는 제1항 내지 제4항 중 어느 한 항의 다기능성 어댑터 중 어느 하나로부터 선택되는 것인 복합체.

청구항 6

a) 복수개의 다기능성 어댑터를 복수개의 dsDNA 단편과 라이게이션하여, 복수개의 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 단계이며, 여기서 각각의 복수개의 다기능성 어댑터는 제1항 내지 제4항 중 어느 한 항의 다기능성 어댑터 중 어느 하나로부터 선택되는 것인 단계; 임의적으로,

b) 단계 (a)로부터의 복수개의 복합체를 하나 이상의 효소와 접촉시켜, 복수개의 연속된 어댑터-태그 부착된 DNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 형성하는 단계

를 포함하는, 어댑터-태그 부착된 DNA 라이브러리를 제조하는 방법.

청구항 7

제6항에 있어서, 복수개의 복합체의 각각의 다기능성 어댑터/dsDNA 단편 복합체가 dsDNA 단편의 각각의 말단에 라이게이션된 다기능성 어댑터를 포함하는 것인 방법.

청구항 8

제6항 또는 제7항에 있어서, 복수개의 dsDNA 단편이 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 또는 탈메틸화 DNA를 포함하는 것인 방법.

청구항 9

제6항 내지 제8항 중 어느 한 항에 있어서, 복수개의 dsDNA 단편이 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구되는 것인 방법.

청구항 10

제6항 내지 제9항 중 어느 한 항에 있어서, 비-라이게이션 가닥 올리고뉴클레오타이드가 단계 (b)에서 다기능성 어댑터/dsDNA 단편 복합체로부터 대체되는 것인 방법.

청구항 11

제6항 내지 제10항 중 어느 한 항에 있어서, 복수개의 연속된 어댑터-태그 부착된 DNA 단편을 증폭시켜, 복수개의 증폭된 연속된 어댑터-태그 부착된 dsDNA 단편을 포함하는 증폭된 어댑터-태그 부착된 DNA 라이브러리를 생성하는 것을 포함하는 방법.

청구항 12

제11항에 있어서, 1개 이상의 프라이머가 증폭을 위해 사용되고, 1개 이상의 프라이머가 어댑터의 프라이머-결합 영역에 혼성화하는 범용 프라이머 결합 서열을 포함하는 것인 방법.

청구항 13

제6항 내지 제12항 중 어느 한 항의 방법에 따라 생성된 어댑터-태그 부착된 DNA 라이브러리.

청구항 14

a) 제13항의 어댑터-태그 부착된 DNA 라이브러리를 1개 이상의 다기능성 포획 프로브와 혼성화하여, 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 형성하는 단계이며, 여기서 각각의 다기능성 포획 프로브는 하기를 포함하는 것인 단계:

i) 파트너 올리고뉴클레오타이드에 혼성화할 수 있는 제1 영역, 여기서 임의적으로 제1 영역은 PCR 프라이머 결합 부위를 포함하는 꼬리 서열을 포함함;

ii) 어댑터-태그 부착된 DNA 라이브러리에서 표적 영역에 혼성화할 수 있는 제2 영역;

b) 단계 (a)로부터의 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 단리하는 단계이며, 여기서 각각의 단리된 포획 프로브/어댑터-태그 부착된 DNA 복합체는 포획 프로브 및 어댑터-태그 부착된 DNA 단편을 포함하는 것인 단계;

c) 단계 (b)로부터의 단리된 포획 프로브/DNA 단편 복합체를 효소적으로 처리하여, 혼성체 분자를 포함하는 프로브-포획된 DNA 라이브러리를 생성하는 단계이며, 여기서 각각의 혼성체 분자는 하기를 포함하는 것인 단계를 포함하는:

i) 포획 프로브 또는 그의 상보체의 적어도 일부분;

ii) DNA 단편 또는 그의 상보체의 적어도 일부분; 및

iii) 어댑터

프로브-포획된 라이브러리를 제조하는 방법.

청구항 15

제14항에 있어서, (c)의 효소적 처리 단계가 주형으로서 복합체의 어댑터-태그 부착된 DNA 단편을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함하는 것인 방법.

청구항 16

제14항 또는 제15항에 있어서, 적어도 1개의 포획 프로브가 표적 영역에서 특이적인 영역의 하류에서 혼성화하고, 적어도 1개의 포획 프로브가 표적 영역에서 특이적인 영역의 상류에서 혼성화하는 것인 방법.

청구항 17

제14항 내지 제16항 중 어느 한 항에 있어서,

d) 단계 (c)로부터의 혼성체 분자에 대해 PCR을 수행하여, 증폭된 혼성체 분자를 생성하는 것을 추가로 포함하는 방법.

청구항 18

제14항 내지 제17항 중 어느 한 항에 따라 생성된 혼성체 분자 또는 증폭된 혼성체 분자를 포함하는 프로브-포획된 라이브러리.

청구항 19

제18항의 프로브-포획된 라이브러리에 대해 표적화된 유전자 분석을 수행하는 것을 포함하는 방법.

청구항 20

제19항에 있어서, 표적화된 유전자 분석이 서열 분석 또는 카피 수 분석을 포함하는 것인 방법.

청구항 21

제19항 또는 제20항에 있어서, 각각의 혼성체 분자에서 포획 프로브 영역 중 전부 또는 일부를 시퀀싱하는 것인 방법.

발명의 설명

기술 분야

관련 출원의 상호 참조

본 출원은 2020년 9월 8일에 출원한 미국 가출원 일련 번호 63/075,543을 우선권 주장하며, 그의 전문은 모든 목적을 위해 본원에 참조로 포함된다.

기술분야

본 개시내용은 유전자 라이브러리의 고효율성 구축을 위한 조성물 및 방법, 및 그의 사용 방법에 관한 것이다. 본원에 기재된 조성물 및 방법을 이용하여 생성되는 유전자 라이브러리는 유전자 분석을 위해 사용될 수 있다.

서열 목록

본 출원은 ASCII 포맷으로 전자적으로 제출된 서열 목록을 포함하며, 그의 전문이 본원에 참조로 포함된다. 2021년 9월 7일에 생성된 상기 ASCII 사본은 CLFK_006_01W0_SeqList_ST25의 명칭을 갖고, 약 20 KB 크기이다.

배경 기술

차세대 시퀀싱 (NGS)은 유전자 변화를 확인하기 위해 다양한 임상 환경에서 사용될 수 있다. NGS는 크게 샘플 전처리, 라이브러리 제조, 시퀀싱, 및 생물정보학의 과정 요소로 나뉘어진다. 현재, 샘플 전처리 및 라이브러리 제조는 대부분 자동화없이 수행되는 노동 집약적인 과정이다. 라이브러리 제조 프로토콜은 일반적으로 다단계 과정으로 구성되며, 고가의 시약 및 상당한 실습 시간을 필요로 한다. NGS에서 이러한 병목현상을 다루기 위해, 샘플 멀티플렉스화, 고처리량, 및 증가된 민감도를 가능하게 하는 자동화된 NGS 방법이 절실히 요구된다.

발명의 내용

본 개시내용은 a) 라이게이션 가닥 올리고뉴클레오타이드, 및 b) 라이게이션 가닥 올리고뉴클레오타이드의 3' 말단에 있는 영역에 혼성화하고 그와 듀플렉스를 형성할 수 있는 비-라이게이션 가닥 올리고뉴클레오타이드를 포함하는 다기능성 어댑터를 제공하며; 라이게이션 가닥 올리고뉴클레오타이드는 샘플로부터의 이중-가닥 DNA (dsDNA) 단편과 접촉시 dsDNA 단편의 각각의 가닥의 5' 말단에 라이게이션하고; 라이게이션 가닥 올리고뉴클레오타이드는 (i) 3' 말단 오버행; (ii) 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 증폭 영역; (iii) 고유의 다기능성 ID 영역; (iv) 고유의 분자 식별자 (UMI) 멀티플라이어; 및 (v) 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보적인 폴리뉴클레오타이드 서열을 포함하는 앵커 영역을 포함하고; dsDNA 단편은 각각의 가닥의 5' 말단에서 포스페이트 기 및 각각의 가닥의 3' 말단에서 오버행을 포함하고; 각각의 dsDNA 단편은 다기능성 ID 영역 및 UMI 멀티플라이어의 조합에 의해 식별될 수 있고; 샘플은 다기능성 ID 영역에 의해 식별될 수 있다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 3' 말단에서 dT 오버행을 포함하고, dsDNA 단편은 각각의 가닥의 3' 말단에서 dA 오버행을 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 3' 말단에서 dA 오버행을 포함하고, dsDNA 단편은 각각의 가닥의 3' 말단에서 dT 오버행을 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 3' 말단에서 dC 오버행을 포함하고, dsDNA 단편은 각각의 가닥의 3' 말단에서 dG 오버행을 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 3' 말단에서 dG 오버행을 포함하고, dsDNA 단편은 각각의 가닥의 3' 말단에서 dC 오버행을 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드에서 증폭 영역은 PCR, LAMP, NASBA, SDA, RCA, 또는 LCR을 위한 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 비-라이게이션 가닥 올리고뉴클레오타이드는 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 포함한다.

본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 샘플은 포유동물로부터 분리되거나 또는 유래된다. 일부 실시양태에서, 포유동물은 인간 질환을 위한 동물 모델이다. 일부 실시양태에서, 포유동물은 마우스, 래트, 기

니 피그, 토끼, 돼지, 고양이, 개, 양 또는 말이다. 일부 실시양태에서, 포유동물은 비-인간 영장류 (NHP)이다. 일부 실시양태에서, 포유동물은 인간이다.

[0016] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 샘플은 하나 이상의 세포 유형으로부터 분리되거나 또는 유래된다. 일부 실시양태에서, 샘플은 하나 이상의 조직 유형으로부터 분리되거나 또는 유래된다. 일부 실시양태에서, 샘플은 하나 이상의 공급원으로부터 분리되거나 또는 유래된다. 일부 실시양태에서, 하나 이상의 공급원은 공여자를 포함한다. 일부 실시양태에서, 공여자는 인간이다. 일부 실시양태에서, 공여자는 건강한 또는 대조군 공여자이다. 일부 실시양태에서, 하나 이상의 공급원은 (예를 들어 임상 시험의) 환자 또는 대상체이다. 일부 실시양태에서, 환자 또는 대상체는 인간이다. 일부 실시양태에서, 환자 또는 대상체는 건강한 또는 대조군 환자 또는 대상체이다. 일부 실시양태에서, 환자 또는 대상체는 시험 환자 또는 대상체이다. 일부 실시양태에서, 환자 또는 대상체는 질환 또는 장애의 징후 또는 증상을 나타낸다. 일부 실시양태에서, 환자 또는 대상체는 임신 중이다. 일부 실시양태에서, 환자 또는 대상체는 질환 또는 장애의 가족력 또는 유전자 마커를 나타낸다.

[0017] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 샘플은 조직 생검이다. 일부 실시양태에서, 조직 생검은 종양 또는 종양으로 의심되는 조직으로부터 취해진다.

[0018] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 또는 탈메틸화 DNA이다. 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 및 탈메틸화 DNA 중 하나 이상을 포함한다. 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 게놈 DNA (gDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 상보성 DNA (cDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 미토콘드리아 DNA를 포함한다. 일부 실시양태에서, dsDNA 단편은 메틸화 DNA를 포함한다. 일부 실시양태에서, dsDNA 단편은 탈메틸화 DNA를 포함한다.

[0019] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, dsDNA는 샘플 또는 시험 샘플로부터 분리되거나 또는 유래된다. 일부 실시양태에서, 샘플 또는 시험 샘플은 생물학적 샘플을 포함한다. 일부 실시양태에서, 생물학적 샘플은 하기로 이루어진 군으로부터 선택된 생물학적 유체를 포함한다: 양수, 혈액, 혈장, 혈청, 정액, 림프액, 뇌척수액, 안구액, 소변, 타액, 대변, 점액, 눈물 및 땀. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 양수를 포함한다. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 전혈, 혈장, 버피 코트, 및 혈청 중 하나 이상을 포함한다. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 림프액을 포함한다. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 뇌척수액을 포함한다. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 소변을 포함한다. 일부 실시양태에서, 생물학적 샘플 또는 생물학적 유체는 타액, 대변, 점액, 눈물 및 땀 중 하나 이상을 포함한다.

[0020] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, dsDNA 단편은 게놈 DNA를 단편화하여 적어도 하나의 DNA 단편을 생성하는 것을 포함하는 방법에 의해 수득된다. 일부 실시양태에서, 상기 방법은 단편화 단계 전에 적어도 하나의 세포를 포함하는 샘플로부터 게놈 DNA를 분리하는 것을 추가로 포함한다. 일부 실시양태에서, 단편화는 게놈 DNA를 적어도 하나의 효소와 접촉시키는 것을 포함하며, 효소는 게놈 DNA를 소화시켜 적어도 하나의 DNA 단편을 생성한다. 일부 실시양태에서, 단편화는 기계적 스트레스를 게놈 DNA에 적용하여 적어도 하나의 DNA 단편을 생성하는 것을 포함한다. 일부 실시양태에서, 단편화는 게놈 DNA를 하나 이상의 화합물과 접촉시켜 게놈 DNA의 하나 이상의 결합을 화학적으로 파괴하는 것을 포함한다. 일부 실시양태에서, 기계적 스트레스는 게놈 DNA를 초음파 처리하여 적어도 하나의 DNA 단편을 생성하는 것을 포함한다. 일부 실시양태에서, 단편화 단계 후에, 상기 방법은 적어도 하나의 DNA 단편 및 효소를 접촉시키는 것을 추가로 포함하고, 효소는 DNA 단편의 하나의 또는 둘 다의 말단을 소화시켜 하나 이상의 평활 말단(들)을 포함하는 DNA 단편을 생성한다. 일부 실시양태에서, 단편화 단계 후에, 상기 방법은 적어도 하나의 DNA 단편의 하나의 또는 둘 다의 평활 말단에 테옥시리보핵산 아데닌 (dA) 꼬리를 부착시키는 것을 추가로 포함한다. 일부 실시양태에서, 단편화 단계 후에, 상기 방법은 적어도 하나의 DNA 단편의 하나의 또는 둘 다의 말단을 인산화시키는 것을 추가로 포함한다. 일부 실시양태에서, 단편화 단계 후에, 상기 방법은 추가로 꼬리 부착 및 인산화 단계를 동시에 또는 순차적으로 포함한다. 일부 실시양태에서, 단편화 단계 후에, 상기 방법은 추가로 꼬리 부착 및 인산화 단계를 순차적으로 포함한다. 일부 실시양태에서, 인산화 단계 후에 꼬리 부착 단계가 있다. 일부 실시양태에서, 꼬리 부착 단계 후에 인산화 단계가 있다.

[0021] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, dsDNA 단편은 (a) 시험 샘플로부터 게놈 DNA를 분리하는

단계; 및 (b) 게놈 DNA를 단편화하여 게놈 DNA 단편을 수득하는 단계를 포함하는 단계에 의해 수득된다. 일부 실시양태에서, 단계 (b)는 게놈 DNA를 적어도 하나의 소화 효소와 접촉시킴으로써 수행된다. 일부 실시양태에서, 단계 (b)는 게놈 DNA에 기계적 스트레스를 적용함으로써 수행된다. 일부 실시양태에서, 기계적 스트레스는 게놈 DNA를 초음파 처리함으로써 적용된다.

[0022] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, dsDNA 단편은 (a) 시험 샘플로부터 세포성 DNA를 분리하는 단계; 및 (b) 세포성 DNA를 단편화하여 게놈 DNA 단편을 수득하는 단계를 포함하는 단계에 의해 수득된다. 일부 실시양태에서, 단계 (b)는 세포성 DNA를 적어도 하나의 소화 효소와 접촉시킴으로써 수행된다. 일부 실시양태에서, 단계 (b)는 세포성 DNA에 기계적 스트레스를 적용함으로써 수행된다. 일부 실시양태에서, 기계적 스트레스는 세포성 DNA를 초음파 처리함으로써 수행된다.

[0023] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 증폭 영역은 10 내지 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 증폭 영역은 20 내지 30개 뉴클레오타이드 길이이다. 일부 실시양태에서, 증폭 영역은 25개 뉴클레오타이드 길이이다.

[0024] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 다기능성 ID 영역은 3 내지 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 3 내지 15개 뉴클레오타이드 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 8개 뉴클레오타이드 길이이다.

[0025] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, UMI 멀티플라이어는 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있다. 일부 실시양태에서, UMI 멀티플라이어는 1 내지 5개 뉴클레오타이드 길이이다. 일부 실시양태에서, UMI 멀티플라이어는 3개 뉴클레오타이드 길이이고, 64가지 가능한 뉴클레오타이드 서열 중 1가지를 포함한다.

[0026] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 앵커 영역은 1 내지 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 5 내지 25개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 10개 뉴클레오타이드 길이이다.

[0027] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터는 복수개의 dsDNA 단편에 라이게이션된다.

[0028] 일부 실시양태에서, dsDNA 단편은 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구된다.

[0029] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 증폭 영역은 동일한 뉴클레오타이드 서열을 포함한다. 일부 실시양태에서, 동일한 뉴클레오타이드 서열은 PCR 프라이머 결합 부위이다.

[0030] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 2 내지 10,000가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 50 내지 500가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 100 내지 400가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 60가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다.

[0031] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 8개 뉴클레오타이드 길이이다.

[0032] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터는 64 내지 2,560,000가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다.

[0033] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터는 3840가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하고, 각각의 뉴클레오타이드 서열은 적어도 2의 해밍(Hamming) 거리만큼 3840가지의 고유의 뉴클레오타이드 서열의 임의의 다른 서열과 별개이다.

[0034] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 각각의 복수개의 다기능성 어댑터는 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있는 UMI 멀티플라이어를 포함한다.

[0035] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의

UMI 멀티플라이어는 1 내지 5개 뉴클레오타이드 길이이다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 UMI 멀티플라이어는 3개 뉴클레오타이드 길이이다.

[0036] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 앵커 영역은 4가지 뉴클레오타이드 서열 중 1가지를 포함하고, 주어진 서열의 각각의 다기능성 ID 영역은 4가지 앵커 영역 중 각각의 1가지와 쌍을 형성할 수 있다.

[0037] 본 개시내용의 다기능성 어댑터의 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 증폭 영역은 동일한 뉴클레오타이드 서열을 포함하고; 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 8개 뉴클레오타이드 길이이고; 각각의 다기능성 ID 영역의 뉴클레오타이드 서열은 적어도 2의 해밍 거리만큼 복수개의 다기능성 어댑터의 임의의 다른 다기능성 ID 영역의 뉴클레오타이드 서열과 별개이고; 각각의 복수개의 다기능성 어댑터는 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있는 UMI 멀티플라이어를 포함하고, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 UMI 멀티플라이어는 3개 뉴클레오타이드 길이이고, 각각의 가능한 뉴클레오타이드 서열의 UMI 멀티플라이어는 복수개의 다기능성 어댑터의 각각의 다기능성 ID 영역과 쌍을 형성하고, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 앵커 영역은 4가지 뉴클레오타이드 서열 중 1가지를 포함하고, 주어진 서열의 각각의 다기능성 ID 영역은 4가지 앵커 영역 중 각각의 1가지와 쌍을 형성할 수 있다.

[0038] 본 개시내용은 다기능성 어댑터 및 dsDNA 단편을 포함하는 복합체를 제공하며, 다기능성 어댑터는 개시된 다기능성 어댑터 중 어느 하나로부터 선택된다.

[0039] 본 개시내용은 본 개시내용의 복수개의 다기능성 어댑터를 제공한다. 일부 실시양태에서, 복수개는 폴로도 지칭될 수 있다. 일부 실시양태에서, 복수개의 다기능성 어댑터는 샘플에 적용되는 어댑터 세트를 포함한다. 일부 실시양태에서, 샘플에 적용되는 어댑터 세트 내에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터는 고유의 ID 영역 또는 고유의 UMI를 함유한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 다기능성 어댑터의 수는 샘플 또는 샘플의 세포성 DNA 표적을 수용하기 위해 증가되거나 또는 감소될 수 있다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 다기능성 어댑터의 수는 샘플의 세포성 DNA 표적을 검출하고/거나 분석하기 위해 필요한 멀티플렉스화 수준에 상응하도록 증가되거나 또는 감소될 수 있다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 다기능성 어댑터의 수는 샘플에 적용되는 복수개의 다기능성 어댑터 내에서 고유의 ID 영역 또는 고유의 UMI의 수를 증가시키거나 또는 감소시킴으로써 증가되거나 또는 감소될 수 있다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 다기능성 어댑터의 수는 ID 영역 또는 앵커 영역의 뉴클레오타이드를 증가시키거나 또는 감소시킴으로써 증가되거나 또는 감소될 수 있다.

[0040] 본 개시내용은 (a) 복수개의 다기능성 어댑터를 복수개의 dsDNA 단편과 라이게이션하여, 복수개의 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 단계이며, 여기서 다기능성 어댑터는 개시된 다기능성 어댑터 중 어느 하나로부터 선택되는 것인 단계; (b) 단계 (a)로부터의 다기능성 어댑터/dsDNA 단편 복합체를 하나 이상의 효소와 접촉시켜, 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 형성하는 단계를 포함하는, 어댑터-태그 부착된 DNA 라이브러리를 제조하는 방법을 제공한다. 일부 실시양태에서, 복수개의 복합체의 각각의 다기능성 어댑터/dsDNA 단편 복합체는 dsDNA 단편의 각각의 말단에 라이게이션된 다기능성 어댑터를 포함한다.

[0041] 본 개시내용의 방법의 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 또는 메틸화 DNA, 또는 탈메틸화 DNA이다. 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 및 탈메틸화 DNA 중 하나 이상을 포함한다. 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 게놈 DNA (gDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 상보성 DNA (cDNA)를 포함한다. 일부 실시양태에서, dsDNA 단편은 미토콘드리아 DNA를 포함한다. 일부 실시양태에서, dsDNA 단편은 메틸화 DNA를 포함한다. 일부 실시양태에서, dsDNA 단편은 탈메틸화 DNA를 포함한다.

[0042] 본 개시내용의 방법의 일부 실시양태에서, 복수개의 dsDNA 단편은 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구된다.

[0043] 본 개시내용의 방법의 일부 실시양태에서, 복수개의 dsDNA 단편은 저통과(low pass) 전체 게놈 라이브러리, 앵플리콘 라이브러리, 전체 엑솜 라이브러리, cDNA 라이브러리, 또는 메틸화 DNA 라이브러리로 이루어진 목록으로부터 선택된 라이브러리로부터 취득된다.

- [0044] 본 개시내용의 방법의 일부 실시양태에서, 비-라이게이션 가닥 올리고뉴클레오티드는 다기능성 어댑터/dsDNA 단편 복합체로부터 대체된다.
- [0045] 본 개시내용의 방법의 일부 실시양태에서, 하나 이상의 효소는 DNA 리가제 또는 RNA 리가제를 포함한다. 일부 실시양태에서, DNA 리가제는 T4 DNA 리가제 또는 Taq DNA 리가제를 포함한다.
- [0046] 본 개시내용은 (a) 복수개의 다기능성 어댑터를 복수개의 dsDNA 단편과 라이게이션하여, 복수개의 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 단계이며, 여기서 다기능성 어댑터는 개시된 다기능성 어댑터 중 어느 하나로부터 선택되는 것인 단계; (b) 단계 (a)로부터의 다기능성 어댑터/dsDNA 단편 복합체를 하나 이상의 효소와 접촉시켜 연속된 어댑터-태그 부착된 dsDNA 단편을 형성하고; 연속된 어댑터-태그 부착된 dsDNA 단편을 증폭시켜 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 생성하는 단계를 포함하는, 어댑터-태그 부착된 DNA 라이브러리를 제조하는 방법을 제공한다.
- [0047] 본 개시내용의 방법의 일부 실시양태에서, 1개 이상의 프라이머가 증폭을 위해 사용된다. 일부 실시양태에서, 1개 이상의 프라이머는 어댑터의 프라이머-결합 영역에 혼성화하는 범용 프라이머 결합 서열을 포함한다.
- [0048] 본 개시내용은 개시된 방법 중 어느 하나에 따라 생성된 어댑터-태그 부착된 DNA 라이브러리를 제공한다.
- [0049] 본 개시내용은 (a) 개시된 방법 중 어느 하나에 따라 생성된 어댑터-태그 부착된 DNA 라이브러리를 1개 이상의 다기능성 포획 프로브와 혼성화하여, 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 형성하는 단계이며, 여기서 각각의 다기능성 포획 프로브는 (i) 파트너 올리고뉴클레오티드에 혼성화할 수 있는 제1 영역 (여기서, 임의적으로 제1 영역은 PCR 프라이머 결합 부위를 포함하는 꼬리 서열을 포함함); 및 (ii) 태그 부착된 유전자 DNA 라이브러리에서 특이적인 표적 영역에 혼성화할 수 있는 제2 영역을 포함하는 것인 단계; (b) 단계 (a)로부터의 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 단리하는 단계이며, 여기서 각각의 단리된 포획 프로브/어댑터-태그 부착된 DNA 복합체는 포획 프로브 및 어댑터-태그 부착된 DNA 단편을 포함하는 것인 단계; (c) 단계 (b)로부터의 하나 이상의 단리된 포획 프로브/어댑터-태그 부착된 DNA 복합체를 효소적으로 처리하여, 하나 이상의 어댑터-태그 부착된 혼성체 핵산 분자 (혼성체 분자)를 생성하는 단계이며, 여기서 각각의 혼성체 분자는 포획 프로브, 및 표적화된 유전자 서열에 혼성화된 포획 프로브로부터 3'에 있는 어댑터-태그 부착된 DNA 단편의 상보체를 포함하는 것인 단계를 포함하는, 혼성체 분자의 라이브러리를 제조하는 방법을 제공한다. 일부 실시양태에서, 상기 방법은 (d) 단계 (c)로부터의 혼성체 분자에 대해 PCR을 수행하여, 증폭된 혼성체 분자를 포함하는 표적화된 유전자 라이브러리를 생성하는 단계를 추가로 포함한다. 일부 실시양태에서, (c)의 효소적 처리 단계는 주형으로서 복합체의 어댑터-태그 부착된 DNA 단편을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함한다.
- [0050] 개시된 방법 중 어느 하나의 일부 실시양태에서, 적어도 1개의 포획 프로브는 표적화된 유전자 서열의 하류에서 혼성화하고, 적어도 1개의 포획 프로브는 표적화된 유전자 서열의 상류에서 혼성화한다.
- [0051] 개시된 방법 중 어느 하나의 일부 실시양태에서, 포획 프로브는 시퀀싱 프라이머 인식 서열을 포함한다.
- [0052] 일부 실시양태에서, 본 개시내용은 개시된 방법 중 어느 하나에 따라 생성된 포획 프로브/어댑터-태그 부착된 DNA 복합체를 제공한다.
- [0053] 일부 실시양태에서, 본 개시내용은 개시된 방법 중 어느 하나에 따라 생성된 혼성체 분자의 라이브러리를 제공한다.
- [0054] 일부 실시양태에서, 본 개시내용은 개시된 방법 중 어느 하나에 따라 생성된 표적화된 유전자 라이브러리를 제공한다.
- [0055] 본 개시내용의 일부 실시양태는 개시된 방법 중 어느 하나에 따라 생성된 혼성체 분자의 라이브러리에 대해 표적화된 유전자 분석을 수행하는 것을 포함하는 방법에 관한 것이다. 일부 실시양태에서, 표적화된 유전자 분석은 서열 분석이다. 일부 실시양태에서, 표적화된 유전자 분석은 카피 수 분석이다.
- [0056] 본 개시내용의 일부 실시양태는 개시된 방법 중 어느 하나에 따라 생성된 표적화된 유전자 라이브러리에 대해 표적화된 유전자 분석을 수행하는 것을 포함하는 방법에 관한 것이다. 일부 실시양태에서, 표적화된 유전자 분석은 서열 분석이다. 일부 실시양태에서, 표적화된 유전자 분석은 카피 수 분석이다.
- [0057] 개시된 방법 중 어느 하나의 일부 실시양태에서, 혼성체 분자에서 포획 프로브 영역은 시퀀싱된다. 개시된 방법 중 어느 하나의 일부 실시양태에서, 혼성체 분자에서 전체 포획 프로브 영역은 시퀀싱된다. 개시된 방법 중

어느 하나의 일부 실시양태에서, 혼성체 분자에서 포획 프로브 영역의 일부분은 시퀀싱된다.

[0058] 이들 및 다른 측면은 하기 제시된 상세한 설명에서 더욱 상세하게 다루어진다.

도면의 간단한 설명

[0059] **도 1**은 본 개시내용의 예시적인 다기능성 어댑터를 도시하는 개략도이다. 예시적인 다기능성 어댑터는 5'에서 3'으로 25개 뉴클레오타이드 증폭 영역, 8개 뉴클레오타이드 ID 영역, 3개 뉴클레오타이드 UMI 멀티플라이어 영역, 및 10개 뉴클레오타이드 앵커 영역을 포함한다. 다기능성 어댑터는 또한 3' 말단에서 dT 오버행을 포함한다. 다기능성 어댑터의 각각의 영역의 길이는 하기 기재된 바와 같이 달라질 수 있다.

도 2는 본 개시내용의 방법의 전반적인 워크플로우의 실시양태를 도시하는 개략도이다. 이 워크플로우의 단계는 하기에 더욱 상세하게 기재될 것이다.

도 3은 본 개시내용의 방법의 일부 실시양태에 따라 어댑터-태그 부착된 DNA 라이브러리를 생성하는 예시적인 과정을 도시하는 개략도이다. 단계 1에서, cfDNA는 NEB 넥스트 울트라 II 엔드 리페어(NEB Next Ultra II End Repair)®/dA-꼬리 부착 모듈을 이용하여 말단 복구된다. 이어서, 염기를 3' 말단에 부가하고, 5' 말단을 인산화시킨다(단계 2). 단계 2는 열순환기를 사용하여 수행될 수 있다. 단계 3에서, 3' dT 말단 오버행을 갖는 다기능성 어댑터는 5' 말단 및 3' 말단 각각에 커플링된다. 임의적인 단계 4에서(도시되지 않음), 친화도 비드를 이용하여 어댑터-라이게이션된 DNA 가닥으로부터 라이게이션되지 않은 단편 및 어댑터를 분리할 수 있다. 일부 실시양태에서, 상기 방법은 단계 3에서 단계 5로 직접적으로 증폭을 진행한다. 이 단계에서, 울트라 II Q5® 효소는 단편 상에서 오버행을 연장시켜 이중 가닥 라이브러리를 제조하고, 표준 증폭 주기를 이용하여 단편을 증폭시킨다. 임의적으로, 추가의 3분 연장 단계를 초기에 수행하여, 단편의 5' 및 3' 말단에서 오버행을 채울 수 있다.

도 4a-도 4c는 어댑터-태그 부착된 DNA 라이브러리를 생성하는 비교기 과정(자동화되도록 설계되지 않음)을 본 개시내용의 방법에 따른 어댑터-태그 부착된 DNA 라이브러리를 생성하는 자동화가능한 과정과 비교하는 일련의 개략도 및 표를 제공한다. **도 4a**는 비교기 과정에서 수행된 단계, 뿐만 아니라 사용된 시약 및 열순환기 프로그램을 도시한다. **도 4b**는 자동화가능한 과정에서 수행된 단계, 뿐만 아니라 사용된 시약 및 열순환기 프로그램을 도시한다. **도 4c**는 비교기 및 자동화가능한 과정 각각을 이용하여 진행된, 예시적인 야생형 cfDNA 샘플 뿐만 아니라 2개의 대조군 샘플로부터의 깊이 데이터를 제공한다. 자동화가능한 과정을 이용할 때 깊이에서의 % 증가 또한 나타낸다.

도 5는 본 개시내용의 방법의 일부 실시양태에 따라 어댑터-태그 부착된 DNA 라이브러리의 제조 및 증폭을 위한 예시적인 방법을 도시하는 개략도이다.

도 6은 본 개시내용의 방법의 일부 실시양태에 따라 포획 프로브 혼성화 및 연장을 도시하는 개략도이다.

도 7은 본 개시내용의 방법의 일부 실시양태에 따라 표적화된(포획된)라이브러리의 증폭을 도시하는 개략도이다.

도 8은 어댑터 분자, "혼성체 분자" 및 시퀀싱 앰플리콘(NGS의 경우)을 비교하는 개략도이다. 어댑터 분자는 5'에서 3'으로 25개 뉴클레오타이드 증폭 영역, 8개 뉴클레오타이드 ID 영역, 3개 뉴클레오타이드 UMI 멀티플라이어 영역, 10개 뉴클레오타이드 앵커 영역, 및 3' 말단에서 dT 오버행을 포함한다. 혼성체 분자는 5'에서 3'으로 정방향 프라이머(FP), 어댑터 영역(증폭 영역, ID 영역, UMI 멀티플라이어, 및 앵커 영역을 포함함), 라이브러리 단편, 다기능성 포획 프로브(MCP), 및 역방향 프라이머(RP)를 포함한다. FP 및 RP는 혼성체 분자의 증폭을 위해 사용된다. 도면에 도시된 바와 같이, 시퀀싱 리드 1의 경우, 시퀀싱은 증폭 영역의 시작부에서 개시되고, 시퀀싱 앰플리콘을 따라 5'에서 3'으로 진행된다. 리드 2의 경우, 시퀀싱은 다기능성 포획 프로브의 말단에서 개시되고, 시퀀싱 앰플리콘을 따라 3'에서 5'으로 진행된다.

도 9는 비교기 및 자동화가능한 과정에 대한 어댑터 앵커 분포를 도시하는 그래프이다.

도 10a-10b는 cfDNA 라이브러리 구축의 예시적인 과정에서 DNA 단편에 대한 어댑터의 고효율성 부착을 도시하는 그래프 쌍을 제공한다. **도 10a**는 투입 cfDNA의 바이오애널리라이저(Bioanalyzer) 추적에 대한 데이터를 제공하고; **도 10b**는 라이브러리 구축 후에 cfDNA의 바이오애널리라이저 추적을 제공하며, 대부분의(> 50%) cfDNA가 2개의 바코드(어댑터)를 포함함을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0060] 본 개시내용의 조성물 및 방법은 전체 과정의 효율성을 증가시켜 고처리량 분석을 가능하게 하도록 멀티플렉스화된 핵산 검출 및 시퀀싱, 뿐만 아니라 자동화가능한 과정에 대해 이전에 오랫동안 느꼈지만 충족되지 않는 요구를 해결한다. 본 개시내용의 조성물 및 방법은 다양한 차세대 시퀀싱 (NGS) 과정과 함께 사용될 수 있다.
- [0061] 자동화가능한 DNA 라이브러리 제조 및 시퀀싱 과정의 속도 및 정확도는 암을 비롯한 말기 질환의 신속한 검출 및 진단, 및 전과 전에 고감염성 질환의 조기 검출에 특히 중요하다. 유전 질환은 신속하고 정확한 검출시에는 치료될 수 있거나 또는 심지어 그의 발병을 예방할 수 있다. 더욱이, 치료 효능의 모니터링은 질환 진행 또는 그의 관해의 생체마커를 추적하기 위해 종종 신속하고 정확한 결과를 필요로 한다.
- [0062] 본 개시내용은 자동화를 가능하게 하는 유전자 라이브러리 및 유전자 분석의 고효율성 구축을 위한 어댑터 및 방법을 제공한다. 진단 목적을 위한 분석 외에도, 본 개시내용의 방법 및 조성물은 임의의 핵산 샘플의 분석에서 사용될 수 있다. 한 예로서, 본 개시내용의 방법 및 조성물은 하나 이상의 종의 집단-규모 시퀀싱에 사용되어, 집단 수준에서 유전자 변동을 확인할 수 있고, 예를 들어 진화, 농업 및 생물학적 연구 분야의 질문을 다룰 수 있다.
- [0063] 특히 수많은 샘플에 대한 고도로 멀티플렉스화된 반응이 병렬로 최적으로 수행되는 이들 상황에서, 본 개시내용의 조성물 및 방법은 큰 샘플 집단에 대해 분석을 수행하기 위해, 예를 들어 질환 또는 감염의 기원을 추적하기 위해 효율성을 제공한다.
- [0064] 본 개시내용의 조성물 및 방법은 핵산 단편을 검출하고 분석하기 위해 필요한 단계를 최소화하도록 설계된다. 더욱이, 본 개시내용의 조성물 및 방법은 한 단계에서 또 다른 단계로 샘플의 조작을 단순화하도록 설계되어, 일부 상황에서는 다중 단계가 동일한 반응 용기에서 순차적으로 발생할 수 있게 한다. 추가로, 본 개시내용의 조성물 및 방법은 다른 상업적인 과정과 비교하여 더 적은 반응 부피에서 사용될 수 있으며, 이로써 유전 물질의 회수를 감소시킬 수 있다. 이는 출발 유전 물질이 부족하거나 또는 제한될 때, 예를 들어 무세포 DNA (cfDNA) 또는 고대 DNA를 사용할 때 특히 중요하다.
- [0065] 일부 실시양태에서, 본 개시내용은 카피 수 변화, 단일 뉴클레오타이드 변이체, (SNV), 짧은 (40 bp 미만) 삽입 및 결실 (indel), 및 게놈 재배열, 예를 들어 유전자 융합, 예컨대 종양원성 유전자 융합, 역위, 전좌를 비롯한 (이로 제한되지 않는) 다중 유형의 DNA 변화의 검출을 가능하게 하는 어댑터 설계 및 그의 사용 방법을 제공한다.
- [0066] 일부 실시양태에서, 본 개시내용은 간소화된 워크플로우에 따라 태그 부착된 DNA 라이브러리를 제조하는 방법을 제공한다. 이들 방법은 예를 들어 NGS 라이브러리 제조, 표적 포획 과정에 의한 관심 유전자좌의 농축, 유전 물질의 시퀀싱, 및 유전자 분석을 위한 샘플 취급 로봇을 사용하여 고처리량 처리 및 자동화에 특히 유용하다.
- [0067] 개시된 방법에서 개시된 조성물의 사용은 놀랍게도 클로닝 효율성을 증가시키고, 균일한 어댑터 분포를 개선시키고, 서열 리드 뿐만 아니라 게놈 등가물의 더 큰 깊이/커버리지의 관점에서 성능을 개선시키는데 효과적이다.
- [0068] 그 결과, 본 개시내용에 제공된 방법은 표준 워크플로우 (예컨대 자동화 불가능한 워크플로우)와 비교하여 적어도 감소된 단계의 수, 더 짧은 처리 시간, 작업자 오류에 대한 더 낮은 위험, 감소된 시약 수, 더 적은 반응 부피, 및 더 낮은 비용의 우수한 성질을 가지며, 이로써 이러한 방법 및 워크플로우의 상업화 및 자동화를 실현가능하게 한다.
- [0069] 본 개시내용의 방법의 일부 실시양태에서, 상기 방법은 자동화가능한 과정으로 지칭된다. 본 개시내용의 조성물 및 방법이 자동화된 기기에서의 사용을 위해 설계되었지만, 본 개시내용의 조성물 및 방법은 자동화될 필요가 없으며, 명확성을 위해 자동화되지 않은 수단에 의해 또는 자동화되지 않은 기기에서 또한 수행될 수 있다. 비교를 위한 기준을 제공하기 위해, 본 개시내용은 "비교기" 과정 (구체적으로 자동화를 위해 또는 자동화된 기기에서의 사용을 위해 설계되지 않은 과정)을 제공한다. 예를 들어 본 개시내용의 비교기 과정과 함께 정렬될 때, 본 개시내용의 "자동화가능한" 과정은 멀티플렉스화된 핵산 검출 및 분석의 원하는 결과를 보존하면서 몇몇 단계를 제거한다.
- [0070] 본 개시내용의 일부 실시양태의 조성물 및 방법에서, 말단-복구는 단일 단계로 수행되고; 어댑터 라이게이션은 단일 단계로 수행되고; 어댑터-태그 부착된 DNA 단편의 연장 및 증폭은 단일 단계로 수행된다. 일부 실시양태에서, 자동화가능한 과정은 또한 라이브러리 제조 시간을 감소시키고, 사용되는 시약의 수를 감소시키고, 반응을 위한 부피를 감소시킨다. 더 적은 반응 부피가 샘플링 로봇에 의해 취급될 수 있는 미세역가 플레이트 또는

튜브 스트립에서 수행될 수 있기 때문에, 특히 반응 부피의 감소는 자동화를 용이하게 한다. 도 4a-4c는 어댑터-태그 부착된 DNA 라이브러리를 생성하기 위한 예시적인 비교기 과정 및 본 개시내용의 어댑터-태그 부착된 DNA 라이브러리를 생성하기 위한 본 개시내용의 자동화가능한 과정을 설명하는 개략도를 제공한다.

[0071] 정의

[0072] 본 개시내용에서 달리 정의되지 않는다면, 본 출원에서 사용된 과학적 및 기술적 용어는 관련 기술분야의 통상의 기술자에 의해 일반적으로 이해되는 의미를 가져야 한다. 일반적으로, 본 개시내용에 기재된 화학, 분자 생물학, 세포 및 암 생물학, 면역학, 미생물학, 약리학, 단백질 및 핵산 화학과 관련하여 사용되는 명명법 및 그의 기술은 관련 기술분야에 널리 공지되고 일반적으로 사용되는 것들이다. 본원에 언급된 모든 공보, 특허 출원, 특허 및 다른 참고문헌은 그들의 전문이 본원에 참조로 포함된다.

[0073] 본 개시내용에서 사용된 바와 같이, 하기 용어들은 달리 명시되지 않는다면 그들에게 부여된 의미를 갖는다.

[0074] 단수 형태는 그의 문법적 등가물 중 1개 또는 1개 초과 (즉, 적어도 1개)를 지칭하기 위해 본 개시내용에서 사용된다. 예를 들어, "요소"는 1개의 요소 또는 1개 초과 요소의 요소를 의미한다.

[0075] 대안 (예를 들어, "또는")의 사용은 대안 중 하나, 둘 다, 또는 그들의 임의의 조합을 의미하는 것으로 이해되어야 한다.

[0076] 용어 "및/또는"은 대안 중 하나 또는 둘 다를 의미하는 것으로 이해되어야 한다.

[0077] 본 개시내용에서 사용된 바와 같이, 용어 "약" 또는 "대략"은 기준 분량, 수준, 값, 개수, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이의 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% 또는 1%만큼 달라지는 분량, 수준, 값, 개수, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이를 지칭한다. 일부 실시양태에서, 용어 "약" 또는 "대략"은 기준 분량, 수준, 값, 개수, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이에 대해 $\pm 15\%$, $\pm 10\%$, $\pm 9\%$, $\pm 8\%$, $\pm 7\%$, $\pm 6\%$, $\pm 5\%$, $\pm 4\%$, $\pm 3\%$, $\pm 2\%$, 또는 $\pm 1\%$ 인 분량, 수준, 값, 개수, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이의 범위를 지칭한다.

[0078] 본 개시내용에서 사용된 바와 같이, 용어 "단리된"은 그의 본래의 상태에서 일반적으로 그와 동반된 성분이 실질적으로 또는 본질적으로 없는 물질을 의미한다. 일부 실시양태에서, 용어 "수득된" 또는 "유래된"은 단리된과 동의어로 사용된다.

[0079] 본원에 사용된 바와 같이, "대상체", "개체" 또는 "환자"에는 본원에서 고려되는 조성물에 의해 검출되거나 또는 식별될 수 있는 병태의 증상을 나타내는 임의의 동물이 포함된다. 적합한 대상체에는 실험실 동물 (예컨대 마우스, 래트, 토끼, 또는 기니 피그), 농장 동물 (예컨대 말, 소, 양, 돼지), 및 가축 또는 애완동물 (예컨대 고양이 또는 개)이 포함된다. 일부 실시양태에서, 대상체는 포유동물이다. 일부 실시양태에서, 대상체는 비-인간 영장류이고, 바람직한 실시양태에서, 대상체는 인간이다.

[0080] "포함하는", "본질적으로 이루어지는" 및 "이루어지는"과 같은 전환 문구는 미국 특허청의 특허 심사 절차 매뉴얼에 설명된 관례적인 정의를 따른다 (MPEP 2111.03 참조).

[0081] 어댑터 설계

[0082] 자동화 (예를 들어, 샘플-취급 로봇의 사용)에 적합한 고처리량 능력을 달성하기 위해, 본 개시내용의 어댑터 및 관련 방법은 일부 실시양태에서 하기 특징을 포함한다: (i) 1 단계 부착; (ii) 고효율성 부착; (iii) 균일한 어댑터 분포; (iv) 샘플 멀티플렉스화 및 샘플 식별의 수용; (v) 많은 수의 고유의 분자 식별자 (UMI). 예를 들어, 본 개시내용의 어댑터 및 방법의 일부 실시양태는 하기에 제공된다:

[0083] 1 단계 부착: 일부 실시양태에서, 전장 다기능성 어댑터는 1 단계에서 DNA 단편에 부착될 수 있다. "전장" 다기능성 어댑터는 적어도 4개의 영역을 포함할 수 있다: 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 제1 증폭 영역, 고유의 분자 식별자 (UMI)를 포함하는 제2 다기능성 ID 영역, UMI 멀티플라이어를 포함하는 제3 영역, 및 앵커 영역을 포함하는 제4 영역. 전장 다기능성 어댑터의 부착은 앵커가 먼저 부착된 다음, 어댑터의 나머지 영역이 부착되는 단계적인 방식으로 어댑터 라이게이션의 필요성을 제거할 수 있다 (예를 들어, 도 4a의 비교기 과정에서 어댑터 라이게이션의 단계적인 방식 참조).

[0084] 고효율성 부착: 일부 실시양태에서, 다기능성 어댑터는 고효율성으로 DNA 단편에 부착될 수 있다. 본 개시내용의 목적을 위해, 어댑터 부착의 효율성은 투입 DNA 단편에서 어댑터-태그 부착된 DNA 라이브러리 분자로의 전환율을 지칭한다. 예를 들어, DNA 단편은 부착된 어댑터의 ID 영역에 의해 식별될 수 있고, DNA 단편이 어댑터에

부착되지 않는다면 ID 영역을 사용하여 식별할 수 없다. 따라서, 어댑터 부착의 보다 높은 효율성은 라이브러리 전환 과정에서 손실되는 투입 DNA 단편의 수를 감소시킬 수 있다. 이는 예를 들어 여러 종양학 적용 및 다른 유전 질환 (예를 들어 다발성 경화증, 류마티스성 관절염, 알츠하이머(Alzheimer) 질환)과 관련하여 분석되는 샘플에서 이용가능한 DNA의 양이 제한되는 상황에서 특히 유용하다. 이러한 상황에서, DNA 변경 (예를 들어, 단일 뉴클레오타이드 변이체 (SNV), indel, 카피 수 변화, DNA 재배열, 임의적으로 종양/암과 관련됨)의 발생은 전형적으로 드물고, 따라서 검출하기 어려울 수 있다. 이들 DNA 단편에 대한 본 개시내용의 어댑터의 고도로 효율적인 부착은 이러한 드문 변동의 포획을 용이하게 할 수 있다. 일부 실시양태에서, 투입 DNA 단편의 적어도 50%는 다기능성 어댑터의 부착에 의해 어댑터-태그 부착된 DNA 라이브러리 분자로 전환된다. 도 10은 cfDNA 라이브러리 구축의 예시적인 과정에서 DNA 단편에 대한 어댑터의 고효율성 부착에 대한 데이터를 제공한다.

[0085] 균일한 어댑터 분포: 생물정보학적 분석은 샘플내 프로브 성능 및 샘플간 프로브 성능을 분석할 수 있다. 샘플에서 어댑터 풀 사이의 성능 변동은 분석 민감도에 부정적인 영향을 미칠 수 있다. 서열 리드에 의해 측정되는 바와 같이 태그 부착된 DNA 라이브러리 및 포획 프로브 라이브러리에서 균일한 어댑터 분포가 바람직하다. 일부 실시양태에서, 어댑터-태그 부착된 DNA 라이브러리에서 어댑터의 분포에 편향 가능성이 있으며, 일부 어댑터는 DNA 단편에 라이게이션하는데 덜 효율적일 수 있거나 또는 어댑터 풀에서 다른 것과 비교하여 덜 효율적으로 증폭될 수 있다. 이는 시퀀싱 동안에 이들 덜 효율적인 어댑터의 더 적은 앰플리콘 및 더 적은 리드를 일으킬 수 있다. 이러한 편향된 분포가 어댑터 풀에서 덜 효율적인 어댑터의 양을 증가시킴으로써 용인되거나 또는 보상되어, 태그 부착된 DNA 라이브러리 및 시퀀싱 리드에서 어댑터의 더욱 균형잡힌 표현을 제공할 수 있지만, 본 개시내용의 조성물 및 방법은 이러한 보상을 제거하는 옵션을 제공한다. 본원에 개시된 어댑터 및 방법은 균일한 어댑터 분포를 달성하는 예상치 못한 이익을 제공할 수 있고, 각각의 어댑터는 시퀀싱 결과에서 대략 동일한 비를 나타낸다. 이 균일한 어댑터 분포는 증가된 민감도를 제공한다.

[0086] 일부 실시양태에서, 균일한 어댑터 분포는 어댑터의 각각의 풀에서 모두 표현되는 다중 유형의 앵커 영역을 가짐으로써 달성될 수 있다.

[0087] 일부 실시양태에서, 균일한 어댑터 분포는 어댑터의 각각의 풀에 대해 무작위로 선택된 고유의 ID 영역 (각각의 ID 영역은 샘플 및 그에 부착된 DNA 단편 둘 다를 식별함)을 가짐으로써 달성될 수 있다.

[0088] 샘플 멀티플렉스화 및 샘플 식별의 수용: 샘플 멀티플렉스화 (즉, 상이한 샘플을 동시에 실행시키는 능력)를 달성하기 위해, 일부 실시양태에서, 동일한 풀 내의 각각의 고유의 어댑터가 동일한 샘플에 부착된 것인 고유의 어댑터의 풀을 구축한다. 서열 카운트 관점에서, 어댑터의 풀의 각각의 고유의 어댑터가 풀에서 다른 모든 어댑터와 본질적으로 동일한 거동을 갖는 것이 유익하다. 이를 달성하기 위해, 일부 실시양태에서, 각각의 ID 영역은 ID 영역과 임의의 다른 ID 영역 사이에 2의 해밍 거리를 가지며, 따라서 리드가 잘못된 샘플에 가짜로 배정될 기회를 감소시킨다. 일부 실시양태에서, 어댑터의 각각의 풀은 특이적인 앵커 영역과 쌍을 형성하는 추가의 풀로 분할되어, 샘플 역멀티플렉스화에서 오류 가능성을 추가로 감소시킬 수 있다. 예를 들어, 2의 해밍 거리를 갖는 8쌍체 태그에서, 가능한 서열의 총 개수는 16,384개이다. 용어 "쌍 형성"이 2가지 상이한 폴리뉴클레오타이드 서열 또는 상이한 폴리뉴클레오타이드 서열을 포함하는 DNA 영역과 관련하여 사용될 때, 2가지 상이한 폴리뉴클레오타이드 서열 또는 상이한 폴리뉴클레오타이드 서열을 포함하는 DNA 영역은 동일한 폴리뉴클레오타이드 상에 존재함을 의미한다. 예를 들어, DNA의 특정한 ID 영역이 DNA의 특정한 증폭 영역과 쌍을 형성한다고 할 때, 이는 ID 영역 및 증폭 태그가 동일한 DNA 폴리뉴클레오타이드 분자 상에 존재함을 의미한다.

[0089] 높은 수의 고유의 분자 식별자 (UMI): 어댑터가 분자 생물학 관점에서 기능적으로 등가인 것이 일반적으로 유익하지만, 어댑터가 고유의 게놈 단편의 식별을 증가시키는 매우 많은 수의 고유의 분자 식별자 (UMI) ($\geq 10,000$)를 갖는 것 또한 바람직하다. 이러한 맥락에서, "증가"는 고유하게 유래된 단편을 식별하는 능력이 증가되는 것을 의미한다. 각각의 게놈 클론 단편은 이중-가닥 DNA가 절단된 게놈 서열의 위치에 상응하는 단편화 부위의 특정한 쌍을 갖는다. 각각의 클론이 상이한 절단 부위를 가질 가능성이 있기 때문에, 이 절단 부위를 이용하여 고유의 게놈 클론을 구별할 수 있다. 그러나, 수천 개의 독립적인 클론을 갖는 라이브러리에서, 고유하게 유래된 단편은 종종 정확하게 동일한 절단 부위를 가질 수 있다. 동일한 절단 부위를 공유하는 게놈 클론 (즉, 단편)은 동일한 샘플로부터 유래된 다른 클론 서열과 관련하여 고유한 것으로 또는 중복되는 것으로 분류될 수 있다. 서열 태그의 높은 다양성을 도입하는 어댑터를 부착시킴으로써, 동일한 절단 부위를 공유하는 상이한 게놈 클론은 고유한 것으로 식별될 가능성이 더 크다. 일부 실시양태에서, UMI는 다기능성 ID 영역과 UMI 멀티플라이어의 조합에 의해 생성된다. 즉, 각각의 고유의 DNA 단편은 다기능성 ID 영역 및 UMI 멀티플라이어의 조합에 의해 식별될 수 있다 (즉, UMI에 의해 식별됨). 추가로, UMI 및 절단 부위의 조합은 고유의 분자 식

별자 요소 (UMIE)를 생성하고, 이는 서열 리드를 중복 리드 또는 고유의 리드로 분류하는 것을 용이하게 한다. 일부 실시양태는 UMI 멀티플라이어가 전체 UMI 복잡성을 증가시키거나 또는 감소시키기 위해 더 길거나 또는 짧은 서열을 포함할 수 있음을 고려한다. 일부 실시양태에서, 각각의 고유의 DNA 단편은 다기능성 ID 영역 단독에 의해 식별될 수 있다.

- [0090] 용어 "어댑터", "다기능성 어댑터" 및 "어댑터 모듈"은 상호교환적으로 사용될 수 있고, DNA 또는 RNA 분자의 말단에 라이게이션될 수 있는 짧은 단일- 또는 이중-가닥 올리고뉴클레오타이드를 지칭한다. 전형적으로, 본원에 기재된 어댑터는 적어도 5가지 요소를 포함한다: (i) 3' 말단 오버행; (ii) 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 증폭 영역; (iii) 고유의 다기능성 ID 영역; (iv) 고유의 분자 식별자 (UMI) 멀티플라이어; 및 (v) 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보성인 폴리뉴클레오타이드 서열을 포함하는 앵커 영역. 도 1은 본원에 기재된 일부 실시양태에 따른 다기능성 어댑터의 예시적인 구성을 제공한다 (라이게이션 가닥 올리고뉴클레오타이드만이 도시됨).
- [0091] 일부 실시양태에서, 어댑터는 하나 이상의 증폭 영역, 하나 이상의 다기능성 ID 영역, 하나 이상의 UMI 멀티플라이어, 및 하나 이상의 앵커 영역을 포함한다. 일부 실시양태에서, 어댑터는 5'에서 3'의 순서로 증폭 영역, 다기능성 ID 영역, UMI 멀티플라이어, 앵커 영역, 및 3' 말단 오버행을 포함한다.
- [0092] 일부 실시양태에서, UMI 멀티플라이어는 다기능성 ID 영역 내에 함유되어 있고, 어댑터는 5'에서 3'의 순서로 증폭 영역, 통합된 다기능성 ID 영역 / UMI 멀티플라이어 영역, 앵커 영역, 및 3' 말단 오버행을 포함한다.
- [0093] 일부 실시양태에서, 다기능성 어댑터는 하나 이상의 증폭 영역, 하나 이상의 ID 영역, 하나 이상의 UMI 멀티플라이어, 하나 이상의 앵커 영역, 및 효율적인 라이게이션 기질인 3' 오버행에서 하나 이상의 뉴클레오타이드를 포함한다. 추가의 실시양태에서, 어댑터 모듈은 하나 이상의 시퀀싱 프라이머 결합 부위를 추가로 포함한다.
- [0094] 본 개시내용의 조성물 및 방법에서 사용될 수 있는 예시적인 어댑터의 구조는 표 2 및 표 3에 제공된다. 예를 들어, 일부 실시양태에서, 어댑터의 라이게이션 가닥은 하기 구조를 포함할 수 있다: AMP-ID 영역/UMI 멀티플라이어-ACGTATGCCA (서열식별번호 (SEQ ID NO): 2)-3'dT. 일부 실시양태에서, 어댑터의 라이게이션 가닥은 하기 구조를 포함할 수 있다: AMP-ID 영역/UMI 멀티플라이어-CTAGCGTTAC (서열식별번호: 3)-3'dT. 일부 실시양태에서, 어댑터의 라이게이션 가닥은 하기 구조를 포함할 수 있다: AMP-ID 영역/UMI 멀티플라이어-GATCGACATG (서열식별번호: 4)-3'dT. 일부 실시양태에서, 어댑터의 라이게이션 가닥은 하기 구조를 포함할 수 있다: AMP-ID 영역/UMI 멀티플라이어-TGCATCAGGT (서열식별번호: 5)-3'dT. 일부 실시양태에서, 어댑터의 비-라이게이션 가닥 앵커 영역은 서열 TGGCATACGT (서열식별번호: 6)을 포함할 수 있다. 일부 실시양태에서, 어댑터의 비-라이게이션 가닥 앵커 영역은 서열 GTAACGCTAG (서열식별번호: 7)을 포함할 수 있다. 일부 실시양태에서, 어댑터의 비-라이게이션 가닥 앵커 영역은 서열 CATGTCGATC (서열식별번호: 8)을 포함할 수 있다. 일부 실시양태에서, 어댑터의 비-라이게이션 가닥 앵커 영역은 서열 ACCTGATGCA (서열식별번호: 9)를 포함할 수 있다.
- [0095] 일부 실시양태에서, 어댑터는 3' dT 오버행을 갖는 라이게이션 가닥을 포함할 수 있다. 일부 실시양태에서, 3' dT 오버행을 갖는 라이게이션 가닥은 표 3에 나타난 서열 중 어느 하나를 포함할 수 있다. 예를 들어, 3' dT 오버행을 갖는 라이게이션 가닥은 서열식별번호: 10 내지 69 중 어느 하나의 서열을 포함할 수 있다. 이들 라이게이션 가닥 서열 내에서 "NNN"은 3개 뉴클레오타이드 UMI 멀티플라이어를 나타내고, 여기서 각각의 N은 A, G, C, 또는 T 중 어느 하나로부터 선택될 수 있다. 일부 실시양태에서, 3' dT 오버행을 갖는 라이게이션 가닥은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과 뉴클레오타이드 치환을 갖는 서열식별번호: 10 내지 69 중 어느 하나의 서열을 포함할 수 있다.
- [0096] **라이게이션 가닥 올리고뉴클레오타이드**
- [0097] 용어 "라이게이션 가닥 올리고뉴클레오타이드" 및 "라이게이션 가닥"은 상호교환적으로 사용된다.
- [0098] 일부 실시양태에서, 본 개시내용은 (i) 3' 말단 오버행; (ii) 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 증폭 영역; (iii) 고유의 다기능성 ID 영역; (iv) 고유의 분자 식별자 (UMI) 멀티플라이어; 및 (v) 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보성인 폴리뉴클레오타이드 서열을 포함하는 앵커 영역을 포함하는 라이게이션 가닥 올리고뉴클레오타이드를 제공한다.
- [0099] 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 5' 말단에서 인산화되지 않는다.
- [0100] 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 약 30개 뉴클레오타이드 내지 약 70개 뉴클레오타이드 길이이다. 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 약 35 내지 약 65개 뉴클레오타이드, 약

40 내지 약 60개 뉴클레오타이드, 또는 약 40 내지 약 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 약 47개 뉴클레오타이드 길이이다.

[0101] 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 30개 뉴클레오타이드 내지 70개 뉴클레오타이드 길이이다. 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 35 내지 65개 뉴클레오타이드, 40 내지 60개 뉴클레오타이드, 또는 40 내지 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 47개 뉴클레오타이드 길이이다.

[0102] **비-라이게이션 가닥**

[0103] 용어 "비-라이게이션 가닥 올리고뉴클레오타이드" 및 "비-라이게이션 가닥"은 상호교환적으로 사용된다.

[0104] 비-라이게이션 가닥 올리고뉴클레오타이드는 라이게이션 가닥 올리고뉴클레오타이드의 3' 말단에 있는 영역에 혼성화하고 그와 듀플렉스를 형성할 수 있다. 비-라이게이션 가닥은 듀플렉스를 형성하기 위해 라이게이션 가닥의 적어도 일부분에 대해 상보성이다. 이 듀플렉스 구조는 dsDNA의 5' 말단을 라이게이션 가닥에 라이게이션하는 것을 용이하게 할 수 있다.

[0105] 일부 실시양태에서, 비-라이게이션 가닥은 인산화되지 않는다. 비-라이게이션 가닥의 인산화 결여는 비-라이게이션 가닥이 DNA 단편의 3' 말단에 부착하는 것을 방지할 수 있고, 어댑터 이량체의 형성을 감소시킬 수 있다.

[0106] 일부 실시양태에서, 비-라이게이션 가닥은 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 임의적으로 포함할 수 있다. 일부 실시양태에서, 변형은 화학적 변형이다.

[0107] **3' 말단 오버행**

[0108] 용어 "3' 말단 오버행"은 폴리뉴클레오타이드의 3' 말단에 있는 하나 이상의 뉴클레오타이드 오버행 또는 꼬리를 지칭한다.

[0109] 일부 실시양태에서, 라이게이션 가닥 올리고뉴클레오타이드는 3' 말단에서 dT 오버행을 포함한다.

[0110] 일부 실시양태에서, 상보성 오버행 (예를 들어 dA-오버행/꼬리)을 갖는 DNA 단편에 대한 다기능성 어댑터의 효율적인 라이게이션을 유도하기 위해, 3' 말단 오버행 (예를 들어, dT 꼬리)은 라이게이션 가닥의 라이게이션을 보조하여 DNA 단편의 5' 말단에 라이게이션된다.

[0111] **증폭 영역**

[0112] 용어 "증폭 영역"은 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 어댑터 분자의 요소를 지칭한다. 프라이머 인식 부위는 관련 기술분야에 공지된 임의의 증폭, 예컨대 [Fakruddin et al. "Nucleic acid amplification: Alternative methods of polymerase chain reaction" *J Pharm Bioallied Sci.* 2013 Oct-Dec; 5(4): 245-252]에 개시된 방법에 적합한 임의의 프라이머일 수 있다. 예를 들어, 이러한 증폭 방법에는 PCR (폴리머라제 연쇄 반응), LAMP (루프-매개된 등온 증폭), NASBA (핵산 서열-기반 증폭), SDA (가닥 대체 증폭), RCA (회전 환 증폭), LCR (리가제 연쇄 반응)이 포함될 수 있다.

[0113] 일부 실시양태에서, 어댑터는 DNA 라이브러리의 단일-프라이머 증폭을 위한 1개 이상의 프라이머 인식 서열을 포함하는 증폭 영역을 포함한다. 일부 실시양태에서, 증폭 영역은 DNA 라이브러리의 단일-프라이머 증폭을 위한 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과인 프라이머 인식 서열을 포함한다. 일부 실시양태에서, 증폭 영역은 ACA2 프라이머 (서열식별번호: 70)에 대한 PCR 프라이머 결합 부위를 포함한다.

[0114] 일부 실시양태에서, 증폭 영역은 약 5 내지 약 50개 뉴클레오타이드, 약 10 내지 약 45개 뉴클레오타이드, 약 15 내지 약 40개 뉴클레오타이드, 또는 약 20 내지 약 30개 뉴클레오타이드 길이이다. 일부 실시양태에서, 증폭 영역은 약 10개 뉴클레오타이드, 약 11개 뉴클레오타이드, 약 12개 뉴클레오타이드, 약 13개 뉴클레오타이드, 약 14개 뉴클레오타이드, 약 15개 뉴클레오타이드, 약 16개 뉴클레오타이드, 약 17개 뉴클레오타이드, 약 18개 뉴클레오타이드, 약 19개 뉴클레오타이드, 약 20개 뉴클레오타이드, 약 21개 뉴클레오타이드, 약 22개 뉴클레오타이드, 약 23개 뉴클레오타이드, 약 24개 뉴클레오타이드, 약 25개 뉴클레오타이드, 약 26개 뉴클레오타이드, 약 27개 뉴클레오타이드, 약 28개 뉴클레오타이드, 약 29개 뉴클레오타이드, 약 30개 뉴클레오타이드, 약 31개 뉴클레오타이드, 약 32개 뉴클레오타이드, 약 33개 뉴클레오타이드, 약 34개 뉴클레오타이드, 약 35개 뉴클레오타이드, 약 36개 뉴클레오타이드, 약 37개 뉴클레오타이드, 약 38개 뉴클레오타이드, 약 39개 뉴클레오타이드, 또는 약 40개 뉴클레오타이드 또는 그 초과인 길이이다. 일부 실시양태에서, 증폭 영역은 약 25개 뉴클레오타이드 길이이다.

- [0115] 일부 실시양태에서, 증폭 영역은 5 내지 50개 뉴클레오타이드, 10 내지 45개 뉴클레오타이드, 15 내지 40개 뉴클레오타이드, 또는 20 내지 30개 뉴클레오타이드 길이이다. 일부 실시양태에서, 증폭 영역은 10개 뉴클레오타이드, 11개 뉴클레오타이드, 12개 뉴클레오타이드, 13개 뉴클레오타이드, 14개 뉴클레오타이드, 15개 뉴클레오타이드, 16개 뉴클레오타이드, 17개 뉴클레오타이드, 18개 뉴클레오타이드, 19개 뉴클레오타이드, 20개 뉴클레오타이드, 21개 뉴클레오타이드, 22개 뉴클레오타이드, 23개 뉴클레오타이드, 24개 뉴클레오타이드, 25개 뉴클레오타이드, 26개 뉴클레오타이드, 27개 뉴클레오타이드, 28개 뉴클레오타이드, 29개 뉴클레오타이드, 30개 뉴클레오타이드, 31개 뉴클레오타이드, 32개 뉴클레오타이드, 33개 뉴클레오타이드, 34개 뉴클레오타이드, 35개 뉴클레오타이드, 36개 뉴클레오타이드, 37개 뉴클레오타이드, 38개 뉴클레오타이드, 39개 뉴클레오타이드, 또는 40개 뉴클레오타이드 또는 그 초과 길이이다. 일부 실시양태에서, 증폭 영역은 25개 뉴클레오타이드 길이이다.
- [0116] **다기능성 ID 영역**
- [0117] 용어 "다기능성 ID 영역" 및 "ID 영역"은 상호교환적으로 사용되며, 특정한 DNA 단편 뿐만 아니라 그가 유래된 샘플을 고유하게 식별하는 폴리뉴클레오타이드 서열을 포함하는 어댑터의 요소를 지칭한다.
- [0118] 일부 실시양태에서, 다기능성 ID 영역은 약 3 내지 약 50개 뉴클레오타이드, 약 3 내지 약 25개 뉴클레오타이드, 또는 약 5 내지 약 15개 뉴클레오타이드 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 약 3개 뉴클레오타이드, 4개 뉴클레오타이드, 약 5개 뉴클레오타이드, 약 6개 뉴클레오타이드, 약 7개 뉴클레오타이드, 약 8개 뉴클레오타이드, 약 9개 뉴클레오타이드, 약 10개 뉴클레오타이드, 약 11개 뉴클레오타이드, 약 12개 뉴클레오타이드, 약 13개 뉴클레오타이드, 약 14개 뉴클레오타이드, 약 15개 뉴클레오타이드, 약 16개 뉴클레오타이드, 약 17개 뉴클레오타이드, 약 18개 뉴클레오타이드, 약 19개 뉴클레오타이드, 또는 약 20개 뉴클레오타이드 또는 그 초과 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 약 8개 뉴클레오타이드 길이이다.
- [0119] 일부 실시양태에서, 다기능성 ID 영역은 3 내지 50개 뉴클레오타이드, 3 내지 25개 뉴클레오타이드, 또는 5 내지 15개 뉴클레오타이드 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 3개 뉴클레오타이드, 4개 뉴클레오타이드, 5개 뉴클레오타이드, 6개 뉴클레오타이드, 7개 뉴클레오타이드, 8개 뉴클레오타이드, 9개 뉴클레오타이드, 10개 뉴클레오타이드, 11개 뉴클레오타이드, 12개 뉴클레오타이드, 13개 뉴클레오타이드, 14개 뉴클레오타이드, 15개 뉴클레오타이드, 16개 뉴클레오타이드, 17개 뉴클레오타이드, 18개 뉴클레오타이드, 19개 뉴클레오타이드, 또는 20개 뉴클레오타이드 또는 그 초과 길이이다. 일부 실시양태에서, 다기능성 ID 영역은 8개 뉴클레오타이드 길이이다.
- [0120] 일부 실시양태에서, 다기능성 ID 영역은 약 2 내지 약 10,000가지의 고유의 뉴클레오타이드 서열, 약 50 내지 약 500가지의 고유의 뉴클레오타이드 서열, 또는 약 100 내지 약 400가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 약 60가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다.
- [0121] 일부 실시양태에서, 다기능성 ID 영역은 2 내지 10,000가지의 고유의 뉴클레오타이드 서열, 50 내지 500가지의 고유의 뉴클레오타이드 서열, 또는 100 내지 400가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다. 일부 실시양태에서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역은 60가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다.
- [0122] 일부 실시양태에서, 다기능성 어댑터는 64 내지 2,560,000가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함한다.
- [0123] 일부 실시양태에서, 어댑터의 미리 지정된 풀 (복수개)이 제공된다. 이러한 미리 지정된 풀을 사용하여 단일 샘플을 나타낸다. 즉, 어댑터 올리고뉴클레오타이드의 각각의 풀에서의 각각의 어댑터 서열은 다른 샘플을 확인하기 위해 사용되는 다른 모든 풀에서의 각각의 어댑터 서열과 구별된다. 관련 기술분야의 기술자는 어댑터 올리고뉴클레오타이드에 대해 가능한 미리 지정된 풀에서 별개의 올리고뉴클레오타이드의 수가 다기능성 ID 영역 및/또는 UMI 멀티플라이어의 길이에 따라 좌우될 것임을 인식할 것이다. "복수개"는 복수개의 동일한 어댑터 모듈, 또는 상이한 어댑터 모듈의 풀을 지칭할 수 있다.
- [0124] 일부 실시양태에서, ID 영역은 개별 샘플, 예를 들어 계층 라이브러리 공급원을 식별한다. 일부 실시양태에서, 각각의 샘플은 약 64가지 내지 약 250만 가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일부 실시양태에서, 각각의 샘플은 64가지 내지 250만 가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일부 실시양태에서, 각각의 샘플은 약 3,840가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일부 실시양태에서, 각각의 샘플은 3,840가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일부 실시양태에서, 각각의 샘플은 약 1 내지 약 60가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일

부 실시양태에서, 각각의 샘플은 1 내지 60가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정된다. 일부 실시양태에서, 각각의 샘플은 60가지의 고유의 어댑터의 복수개 (미리 지정된 풀)로 배정되고, 60가지의 고유의 어댑터의 각각의 미리 지정된 풀은 4개의 세트 (각각의 세트는 15가지의 고유의 어댑터를 포함함)로 추가로 나뉘어지고, 한 세트의 각각의 다기능성 ID 영역은 4가지의 앵커 서열 중 1가지와 쌍을 형성한다. 따라서, 샘플은 다기능성 ID 영역 및 앵커 영역의 조합에 의해 식별될 수 있다.

[0125] 일부 실시양태에서, 각각의 다기능성 ID 영역의 뉴클레오타이드 서열은 적어도 2의 해밍 거리만큼 복수개의 다기능성 어댑터의 임의의 다른 다기능성 ID 영역의 뉴클레오타이드 서열과 별개이다 (한 ID 영역을 또 다른 것으로 변화시키기 위해 적어도 2개의 염기 변화가 필요함을 의미함).

[0126] 일부 실시양태에서, ID 영역은 그가 부착된 개별 DNA 단편을 식별하며, 따라서 ID 영역은 또한 한 예에서 카피 수 분석을 위한 클론 다양성을 열거할 수 있는 단편 태그의 역할을 한다.

[0127] 일부 실시양태에서, 다기능성 ID 영역은 8개 뉴클레오타이드 길이이고, 240가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하며, UMI 멀티플라이어는 3개 뉴클레오타이드 서열 길이이고, 따라서 고유의 어댑터 서열의 총 개수는 $240 \times 4^3 = 3840 = 15,360$ 일 것이다. 따라서, 일부 실시양태에서, 각각의 샘플은 DNA 단편 식별을 위한 1~15,360가지의 고유의 어댑터의 범위를 갖는 어댑터 세트로 배정될 수 있다.

[0128] 일부 실시양태에서, 다기능성 ID 영역은 8개 뉴클레오타이드 길이이고, 60가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하며, UMI 멀티플라이어는 3개 뉴클레오타이드 길이이고, 각각의 뉴클레오타이드 서열은 적어도 2의 해밍 거리만큼 3840가지의 고유의 뉴클레오타이드 서열의 임의의 다른 서열과 별개이다.

[0129] 따라서, 다기능성 ID 영역은 샘플과 DNA 단편 둘 다의 식별에 기여한다. 이는 샘플 멀티플렉스화를 가능하게 하도록 서열을 식별하기 위해 무작위로 생성된 태그, 및 별도의 바코드 또는 시퀀서 인덱싱을 이용하는 관련 기술분야에서 사용되는 현재의 시스템과는 매우 대조적이다.

[0130] UMI 멀티플라이어

[0131] 가능한 서열 태그 (UMI)의 다양성을 추가로 증가시키기 위해, UMI 멀티플라이어는 어댑터에 포함된다. UMI 멀티플라이어는 무작위 염기의 짧은 서열이고 (예를 들어 NNN이며, 여기서 각각의 N은 A, C, G, 및 T 중 어느 하나로부터 선택될 수 있음), UMI와 조합될 때, 어댑터 풀에서 어댑터 서열의 다양성 및 총 개수를 증가시킨다. 일부 실시양태에서, 어댑터는 UMI 멀티플라이어이며, UMI 멀티플라이어는 ID 영역에 인접하거나 또는 그 안에 함유되어 있다. 일부 실시양태에서, 어댑터는 8개 뉴클레오타이드 길이인 ID 영역 및 3개 뉴클레오타이드 길이인 UMI 멀티플라이어를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 3개 뉴클레오타이드 길이이고, 64가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 ID 영역에 인접하여 위치하거나 또는 그 안에 함유되어 있다.

[0132] 일부 실시양태에서, UMI 멀티플라이어의 각각의 뉴클레오타이드 위치는 아데닌, 구아닌, 시토신, 또는 티민 중 어느 하나를 포함할 수 있다. 따라서, 일부 실시양태에서, n개의 뉴클레오타이드를 포함하는 UMI 멀티플라이어는 임의의 4^n 가지의 가능한 뉴클레오타이드 서열을 포함할 수 있다. 일부 실시양태에서, UMI 멀티플라이어는 1개 뉴클레오타이드 길이이고, 4가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 2개 뉴클레오타이드 길이이고, 16가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 3개 뉴클레오타이드 길이이고, 64가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 4개 뉴클레오타이드 길이이고, 256가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 5개 뉴클레오타이드 길이이고, 1,024가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 6개 뉴클레오타이드 길이이고, 4,096가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 7개 뉴클레오타이드 길이이고, 16,384가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 8개 뉴클레오타이드 길이이고, 65,536가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 9개 뉴클레오타이드 길이이고, 262,144가지의 가능한 서열 중 1가지를 포함한다. 일부 실시양태에서, UMI 멀티플라이어는 10개 이상의 뉴클레오타이드 길이이고, 1,048,576가지 이상의 가능한 서열 중 1가지를 포함한다.

[0133] 일부 실시양태에서, UMI 멀티플라이어는 적어도 1, 적어도 2, 적어도 3, 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 또는 적어도 10개 뉴클레오타이드 길이이다. 일부 실시양태에서, UMI 멀티플라이어는 1 내지 5개 뉴클레오타이드 길이이다.

- [0134] **앵커 영역**
- [0135] 용어 "앵커 영역" 및 "앵커 서열"은 상호교환적으로 사용되며, 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보적인 폴리뉴클레오타이드 서열을 지칭한다. 일부 실시양태에서, 앵커 영역은 링커로도 지칭된다. 앵커 영역은 일부 실시양태에서 하기 성질 중 하나 이상을 포함한다:
- [0136] (1) 각각의 앵커 서열은 연장부 내의 각각의 부위에서 4가지 가능한 DNA 염기 각각을 집합적으로 나타내는 2가지 이상의 고유의 앵커 유형의 풀의 일부일 수 있고; 균형잡힌 염기 표현인 이 특징은 일부 실시양태에서 시퀀싱 리드에서 적절한 염기 호출을 보정하는데 유용하다. 앵커 서열의 총 유형의 수는 검출 모드의 총 수와 일치해야 한다. 예를 들어, 4가지 색상이 일루미나(Illumina)® 시퀀싱에서 검출되고, 따라서 4가지 유형의 앵커 서열이 사용될 수 있다. 최대 민감도를 달성하기 위해, 각각의 검출 모드를 이용할 수 있다. 본 개시내용의 조성물 및 방법은 광-기반 검출, 효소-기반 검출, 및 자성 검출을 비롯하여 이로 제한되지 않는 관련 기술분야에 공지된 임의의 검출 모드에서 사용될 수 있다.
- [0137] (2) 각각의 앵커 서열은 4가지 가능한 염기 중 2가지만으로 구성될 수 있으며, 이들은 구체적으로 동일한 수의 A + C 또는 동일한 수의 G + T로부터 선택되고; 2가지 염기만으로 형성된 앵커 서열은 앵커 서열이 적절한 어댑터 기능을 방해하는 이차 구조 형성에 참여할 가능성을 감소시킨다.
- [0138] (3) 각각의 앵커 서열이 동일한 수의 A + C 또는 G + T로 구성되기 때문에, 각각의 앵커 서열은 풀에서 다른 모든 앵커 서열과 대략 동일한 용융 온도 및 듀플렉스 안정성을 공유할 수 있다.
- [0139] (4) 각각의 유형의 앵커 서열 (A/T/G/C로 종결됨)은 시퀀싱 리드에서 대략 동등하게, 예를 들어 대략 등몰량으로 분포될 수 있다 (즉, 풀의 약 25%는 A로 종결되는, 약 25%는 T로 종결되는, 약 25%는 G로 종결되는, 약 25%는 C로 종결되는 어댑터 서열을 가짐).
- [0140] 일부 실시양태에서, 어댑터 모듈은 상이한 앵커 유형을 함유하는 등몰량의 어댑터에서 (예를 들어 등몰량의 앵커 1, 앵커 2, 앵커 3, 앵커 4) 더욱 균일한 어댑터 분포를 제공하기 위해 DNA 단편과 혼합된다. 예시적인 앵커 서열에는 하기가 포함되나 이로 제한되지 않는다: 앵커 1 ACGTATGCCA (서열식별번호: 2); 앵커 2 CTAGCGTTAC (서열식별번호: 3); 앵커 3 GATCGACATG (서열식별번호: 4); 및 앵커 4 TGCATCAGGT (서열식별번호: 5).
- [0141] 일부 실시양태에서, 어댑터 서열은 3' 말단 (3' T 오버행)에서 T 뉴클레오타이드로 종결된다. 일부 실시양태에서, 어댑터는 3' 말단의 마지막 2개 뉴클레오타이드로서 TT를 갖는다. 일부 실시양태에서, 어댑터는 3' 말단의 마지막 2개 뉴클레오타이드로서 AT, CT, 또는 GT를 갖는다.
- [0142] 일반적으로, 앵커 유형의 이상적인 분포는 동일한 분포 백분율 (즉, 100%를 앵커 유형의 수로 나눔)을 갖는 각각의 앵커 유형을 생성하고, 이는 모든 DNA 단편에서 상이한 앵커를 포함하는 상이한 어댑터의 "균일한" 분포를 생성한다. 예를 들어, 4가지 앵커 유형의 이상적인 분포는 각각의 앵커 유형의 약 25% 분포를 생성한다. 일부 실시양태에서, 주어진 풀의 앵커 서열은 약 5% 내지 약 75%의 분포 백분율을 갖는다 (즉, 가장 드문 앵커 유형의 분포 %는 약 5%이고, 가장 빈번한 앵커 유형의 분포 %는 약 75%임). 일부 실시양태에서, 주어진 풀의 각각의 앵커 서열은 약 50%, 약 34%, 약 28%, 약 27%, 약 23%, 약 14%, 또는 약 9%의 분포 %를 갖는다. 일부 실시양태에서, 본원에 기재된 자동화가능한 과정의 각각의 앵커 서열의 분포 백분율은 비교기 과정 (즉, 자동화를 위해 설계되지 않은 과정, 표 4 참조)과 비교하여 상응하는 이상적인 분포 백분율에 적어도 5%, 적어도 10%, 적어도 25%, 또는 적어도 20% 더 가깝다.
- [0143] 일부 실시양태에서, 복수개의 어댑터는 3' dT 오버행을 포함하고, 3' 말단의 마지막 2개의 2개 뉴클레오타이드로서 TT를 갖는 앵커를 갖는 어댑터를 다량으로 포함할 수 있다. 이러한 어댑터는 풀에서 다른 앵커 유형의 1X, 2X, 4X, 5X, 6X, 7X, 8X, 9X, 또는 10X 초과일 수 있고, 이는 시퀀싱 리드에서 어댑터의 더욱 균일한 분포를 생성한다.
- [0144] 일부 실시양태에서, 복수개의 어댑터는 1개 초과 앵커 서열을 포함할 수 있다. 예를 들어, 복수개의 어댑터는 동시에 사용되는 4가지 상이한 앵커 서열을 포함할 수 있다. 이들 앵커 서열은 또한 샘플 역멀티플렉싱 동안에 오류를 저하시키기 위해 사용될 수 있다. 또한, 리드 내에서 서열의 위치가 고정되고, 따라서 ID 영역 및 앵커는 하류 고려를 위해 포함 여파기를 통과하도록 시퀀싱 리드 내에서 고정된 위치를 가져야 한다.
- [0145] 일부 실시양태에서, 앵커 영역은 1 내지 50개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 4 내지 40개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 5 내지 25개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 적어도 4개 뉴클레오타이드, 적어도 6개 뉴클레오타이드, 적어도 8개 뉴클레오타이드, 적어

도 10개 뉴클레오타이드, 적어도 12개 뉴클레오타이드, 적어도 14개 뉴클레오타이드, 또는 적어도 16개 뉴클레오타이드 길이이다. 일부 실시양태에서, 앵커 영역은 10개 뉴클레오타이드 길이이다.

[0146] 예시적인 워크플로우

[0147] 본 개시내용의 방법의 예시적인 워크플로우는 하기에 제공되며, 도 2에 도시되어 있다.

[0148] 1. 말단-복구

[0149] 일부 실시양태에서, 투입 DNA 단편을 "말단 복구 DNA 단편"으로 전환시켜, 말단 복구 DNA 단편이 단일 반응 혼합물에서 5' 포스페이트 기 및 3' dA 뉴클레오타이드 오버행을 갖도록 한다 (단일 단계). 상업적으로 입수가능한 키트 (예를 들어 NEB 울트라 II 엔드 리페어®/dA 꼬리 부착 모듈 E7546L)를 사용하여 DNA 단편을 말단-복구할 수 있거나, 또는 개시된 개별 효소 및 완충제 중 하나 이상을 5' 포스페이트 기 및 3' dA 뉴클레오타이드 오버행을 갖는 말단 복구 DNA 단편의 제조를 위해 조합할 수 있다.

[0150] 일부 실시양태에서, 말단-복구 반응 부피는 50 μ L 미만이다.

[0151] 2. 어댑터 라이게이션

[0152] 일부 실시양태에서, 다기능성 어댑터의 풀을 하나 이상의 샘플 (멀티플렉스화)로부터의 말단 복구 dsDNA 단편에 라이게이션하여, dsDNA 단편의 5' 말단에 어댑터를 부착시킨다.

[0153] 일부 실시양태에서, 라이게이션 반응 부피는 100 μ L 미만이다.

[0154] 일부 실시양태에서, 어댑터-태그 부착된 DNA 단편은 100 μ L 미만의 반응 부피로 분리되고 세척된다.

[0155] 3. 연장

[0156] 일부 실시양태에서, 3' dA-꼬리 부착 DNA 단편은 DNA 단편의 3' 말단으로부터 연장되고, 주형으로서 DNA 단편의 5' 말단에 부착된 라이게이션 가닥을 사용하여 비-라이게이션 가닥을 대체시켜, 증폭에 적합한 연속된 어댑터-태그 부착된 dsDNA 단편을 제조한다. "연속된 어댑터-태그 부착된 dsDNA 단편"의 집합은 증폭되지 않은 어댑터-태그 부착된 DNA 라이브러리이다. 도 3은 본 개시내용의 일부 방법에 따른 어댑터-태그 부착된 DNA 라이브러리를 생성하는 예시적인 과정을 도시하는 개략도를 제공한다.

[0157] 일부 실시양태에서, 연장 반응 부피는 100 μ L 미만이다.

[0158] 4. 증폭

[0159] 일부 실시양태에서, 증폭되지 않은 어댑터-태그 부착된 DNA 라이브러리는 어댑터의 증폭 영역에서 증폭 프라이머 결합 부위를 인식하는 단일 프라이머에 의해 PCR 증폭되어, 증폭된 어댑터-태그 부착된 DNA 라이브러리를 생성한다. 일부 실시양태에서, 단일 프라이머는 서열식별번호: 70의 서열을 포함한다. 본 개시내용의 목적을 위해, "증폭된 어댑터-태그 부착된-DNA 라이브러리", "증폭된 태그 부착된 DNA 라이브러리" 및 "라이브러리 생성물 증폭된 (LPA)"은 상호교환적으로 사용된다. 도 5는 본 개시내용의 방법에 따른 어댑터-태그 부착된 DNA 라이브러리의 제조 및 증폭을 도시하는 개략도를 제공한다.

[0160] 일부 실시양태에서, 증폭 반응 부피는 100 μ L 미만이다.

[0161] 일부 실시양태에서, 증폭된 태그 부착된 DNA 라이브러리는 100 μ L 미만의 부피로 추가로 분리되고 세척된다.

[0162] 일부 실시양태에서, 증폭은 표 7에 제공된 조건에 따라 수행된다. 예를 들어, 일부 실시양태에서, 본원에 기재된 방법은 1) 69°C의 어닐링 온도하에 2개의 별도의 튜브로 나뉘어진 라이브러리의 증폭을 수행하거나; 2) 65°C의 어닐링 온도하에 2개의 별도의 튜브로 나뉘어진 라이브러리의 증폭을 수행하거나; 또는 3) 65°C의 어닐링 온도하에 라이브러리를 나누지 않고 (1개의 튜브에서) 증폭을 수행하는 것을 포함한다.

[0163] 5. 유전자좌/유전자좌들의 포획 및 분리

[0164] 일부 실시양태에서, 비오틴화 파트너 올리고뉴클레오타이드에 적어도 부분적으로 듀플렉스화된 그의 꼬리 영역 (제1 영역)을 갖는 다기능성 포획 프로브는 증폭된 태그 부착된 DNA 라이브러리에 혼성화되어, 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 형성한다.

[0165] 6. 증폭된 태그 부착된 DNA 분자-포획 프로브 모듈 복합체의 분리

[0166] 일부 실시양태에서, 포획 프로브/어댑터-태그 부착된 DNA 복합체 (즉, 포획된 단편)는 자성 스트렙타비딘-비드

를 사용하여 혼성화되지 않은 단편 (즉, 포획되지 않은 단편)으로부터 분리된다.

7. 포획 프로브 연장

일부 실시양태에서, 복합체에서 비드-지시된 포획 프로브는 주형으로서 태그 부착된 DNA 단편을 사용하여 3' 말단으로부터 연장되어, 어댑터-태그 부착된 혼성체 핵산 분자 (혼성체 분자)를 생성하고, 각각의 혼성체 분자는 포획 프로브, 및 표적화된 유전자 서열에 혼성화된 포획 프로브로부터 3'에 있는 어댑터-태그 부착된 DNA 단편의 상보체를 포함한다.

일부 실시양태에서, 변성은 혼성체 분자를 자성 비드에서 용액으로 방출시킨다. 도 6은 포획 프로브 혼성화 및 연장을 도시하는 개략도를 제공한다.

8. 혼성체 분자의 증폭

일부 실시양태에서, 정방향 프라이머 (FP) (서열식별번호: 71)는 혼성체 분자 내의 어댑터 태그의 증폭 영역에서 프라이머 결합 부위에 혼성화하고, 주형으로서 혼성체 분자를 사용하여 포획 프로브를 5' → 3'으로 연장시켜, 연속된 이중 가닥 혼성체 분자를 제조한다.

일부 실시양태에서, 연속된 이중 가닥 혼성체 분자에서 FP-연장된 가닥은 변성되고, 역방향 프라이머 (RP) (서열식별번호: 72)는 혼성체 분자에서 통합된 포획 프로브 모듈 꼬리 영역에서 변성된 FP-연장된 분자/가닥에 혼성화한다.

일부 실시양태에서, RP는 주형으로서 혼성체 분자를 사용하여 5' → 3'으로 연장되어, 일루미나® 시퀀싱 또는 관련 기술분야에 공지된 임의의 다른 방법에 의한 시퀀싱을 위해 준비된 연속된 이중 가닥 혼성체 분자를 제조한다. 도 7은 표적화된 (포획된) 라이브러리의 증폭을 도시하는 개략도를 제공한다.

일부 실시양태에서, 시퀀싱 프라이머는 서로 상이하다. 일부 실시양태에서, 혼성체 분자의 각각의 말단은 시퀀싱 프라이머, 예컨대 P5 및 P7 시퀀싱 프라이머, 또는 다른 일루미나® 시퀀싱 프라이머에 의해 인식되는 시퀀싱 프라이머 결합 부위를 우세하게 포함한다. 증폭된 시퀀싱 준비된 혼성체 분자의 집합은 "표적화된 유전자 라이브러리", "표적화된 라이브러리", 또는 "프로브 포획된 라이브러리 (PCL)"로 지칭된다.

일부 실시양태에서, 증폭 반응 부피는 100 μL 미만이다.

일부 실시양태에서, 증폭된 혼성체 분자는 단리되고 세척된다.

9. 시퀀싱

일부 실시양태에서, 증폭된 혼성체 분자의 차세대 시퀀싱 (NGS)은 일루미나® 넥스트세크(NextSeq)® 550 시퀀서를 이용하여 수행된다.

일부 실시양태에서, NGS는 증폭되지 않은 어댑터-태그 부착된 DNA 라이브러리, 증폭된 태그 부착된 DNA 라이브러리, 혼성체 분자의 라이브러리 (증폭되지 않은 표적화된 라이브러리), 및/또는 증폭된 표적화된 라이브러리에 대해 수행될 수 있다.

일부 실시양태에서, 시퀀싱 리드 1 (151 nt 길이) 및 시퀀싱 리드 2 (17 nt 길이)는 맞춤형 정방향 및 역방향 시퀀싱 프라이머를 사용하여 수행된다.

시퀀싱은 제조자의 지침에 따라 맞춤형 프라이머, 정방향 Seq 프라이머 및 역방향 Seq 프라이머 62를 사용하여 일루미나 넥스트세크550에서 수행되었다.

정방향 Seq 프라이머:

CAAGCAGAAGACGGCATACGAGATGTGACTGGCACGGGACCAGAGAATTCGAATACA (서열식별번호: 73)

역방향 Seq 프라이머 62:

GTGACTGGCACGGGACCAGAGAATTCGAATACA (서열식별번호: 74)

10. 유전자 분석

일부 실시양태에서, 증폭 프라이머 또는 시퀀싱 프라이머를 사용하는 시퀀싱에 적용될 수 있는 본 개시내용의 방법에 따라 생성된 혼성체 분자 또는 임의의 분자는 유전자 분석에 적용된다.

- [0188] 일부 실시양태에서, 서열 리드 1 및 2는 유전자 분석을 위해 사용된다.
- [0189] 일부 실시양태에서, 생물정보학적 분석은 유전자 변이체, 예컨대 카피 수, SNV, Indel, 유전자 및 염색체 재배열을 확인하기 위해 수행된다.
- [0190] **상세한 방법**
- [0191] **1. 어댑터-태그 부착된 DNA 라이브러리 제조**
- [0192] 일부 실시양태에서, 본 개시내용에서 고려되는 방법은 dsDNA 단편을 하나 이상의 말단-복구 효소로 처리하여 말단 복구 DNA를 생성하고, 하나 이상의 어댑터를 말단 복구 DNA의 각각의 말단에 부착하여 어댑터-태그 부착된 DNA 라이브러리를 생성하는 것을 포함하는, 어댑터-태그 부착된 DNA 라이브러리를 생성하는 것을 포함한다.
- [0193] **DNA 샘플 제조**
- [0194] 본 개시내용에서 사용된 바와 같이, 용어 "DNA"는 데옥시리보핵산을 지칭한다. 일부 실시양태에서, 용어 DNA는 게놈 DNA, 재조합 DNA, 합성 DNA, 또는 cDNA를 지칭한다. 일부 실시양태에서, DNA는 게놈 DNA 또는 cDNA를 지칭한다. 일부 실시양태에서, DNA는 "표적 영역"을 포함한다. 본원에서 고려되는 DNA 라이브러리에는 RNA로부터 구축된 게놈 DNA 라이브러리 및 cDNA 라이브러리 구축물, 예를 들어 RNA 발현 라이브러리가 포함된다. 일부 실시양태에서, DNA 라이브러리는 하나 이상의 추가의 DNA 서열 및/또는 태그를 포함한다.
- [0195] 본 개시내용에서 사용된 바와 같이, 용어 "순환 DNA", "순환 무세포 DNA" 및 "무세포 DNA"는 종종 상호교환적으로 사용되고, 세포의 DNA, 세포로부터 압출된 DNA, 또는 괴사 또는 아포토시스 세포로부터 방출된 DNA인 DNA를 지칭한다. 이 용어는 종종 "세포성 게놈 DNA" 또는 "세포성 DNA"와 대조적으로 사용되며, 이는 본원에서 상호교환적으로 사용되고, 세포 내에 함유되고 (즉, 뉴클레아제), 세포를 용해시키거나 또는 그의 온전성을 달리 파괴함으로써 본원에 기재된 분자 생물학적 기술로만 접근할 수 있는 게놈 DNA를 지칭한다.
- [0196] 본 개시내용에서 제공된 조성물 및 방법은 전형적으로 소량으로 수득되는 귀중한 생물학적 샘플, 예컨대 암 조직 생검 샘플 또는 전형적으로 유체 (예를 들어 소변, CSF, 전혈, 혈장, 타액)인 "액체 생검" 샘플의 제조에 특히 적합하다.
- [0197] 일부 실시양태에서, 라이브러리를 제조하기 위해 사용되는 DNA의 양은 임의의 적합한 양일 수 있다. 일부 실시양태에서, 양은 약 1 pg 내지 약 500 ng, 약 1 ng 내지 약 400 ng, 약 5 ng 내지 약 300 ng, 약 10 ng 내지 약 250 ng, 또는 약 20 ng 내지 약 200 ng이다. 일부 실시양태에서, DNA 양은 약 5 ng 내지 약 50 ng이다.
- [0198] 일부 실시양태에서, 라이브러리를 제조하기 위해 사용되는 DNA의 양은 임의의 적합한 양일 수 있다. 일부 실시양태에서, 양은 1 pg 내지 500 ng, 1 ng 내지 400 ng, 5 ng 내지 300 ng, 10 ng 내지 250 ng, 또는 20 ng 내지 200 ng이다. 일부 실시양태에서, DNA 양은 5 ng 내지 50 ng이다.
- [0199] 일부 실시양태에서, 본 개시내용에서 고려되는 방법 및 조성물은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 또는 탈메틸화 DNA로부터 선택된 dsDNA를 사용한다.
- [0200] 일부 실시양태에서, 본원에서 고려되는 유전자 분석 방법은 cfDNA 또는 단편화된 세포성 게놈 DNA를 하나 이상의 말단-복구 효소로 처리하여 말단 복구 DNA를 생성하고, 하나 이상의 어댑터를 말단 복구 DNA의 각각의 말단에 부착하여 DNA 라이브러리를 생성하는 것을 포함하는, DNA 라이브러리를 생성하는 것을 포함한다.
- [0201] 일부 실시양태에서, 본원에서 고려되는 방법 및 조성물은 피분석물로서 게놈 DNA를 사용하여 카피 수의 변화를 효율적으로 분석하고, 검출하고, 진단하고/거나, 모니터링하도록 설계된다. 일부 실시양태에서, 카피 수 분석은 시험 샘플, 예를 들어 생물학적 샘플, 예컨대 조직 생검으로부터 수득된 게놈 DNA로부터 게놈 DNA 라이브러리를 생성함으로써 수행된다. 일부 실시양태에서, 게놈 DNA는 순환 또는 무세포 DNA이다. 일부 실시양태에서, 게놈 DNA는 세포성 게놈 DNA이다.
- [0202] 일부 실시양태에서, 게놈 DNA는 골수, 식도, 위, 십이지장, 직장, 결장, 회장, 췌장, 폐, 간, 전립선, 뇌, 신경, 수막 조직, 신장 조직, 자궁내막 조직, 자궁경부 조직, 유방, 림프절, 근육 및 피부를 비롯하여 이로 제한되지 않는 조직으로부터 취한 조직 샘플 또는 생검으로부터 수득된다. 일부 실시양태에서, 조직 샘플은 종양 또는 의심되는 종양의 생검이다. 일부 실시양태에서, 종양은 암성이거나 또는 암성으로 의심된다. 일부 실시양태에서, 조직 샘플은 암 세포 또는 암성인 것으로 의심되는 세포이다.
- [0203] 세포로부터 또는 세포로 구성된 생물학적 조직으로부터 게놈 DNA를 정제하는 방법은 관련 기술분야에 널리 공지

되어 있고, 통상의 기술자는 조직 또는 조직이 수득되는 조건에 따라 최적의 절차 또는 시판 키트를 인식할 것이다. 일부 실시양태는 조직으로부터 세포성 DNA의 정제가 예를 들어 화학적 및 물리적 방법, 예컨대 블렌딩, 분쇄 또는 초음파 처리에 의해 조직 샘플 내의 세포성 DNA를 노출시키기 위해 세포 파괴 또는 세포 용해하고; 세포 용해의 역할도 하는 세제 또는 계면활성제를 첨가함으로써 막 지질을 제거하고, 임의적으로 예를 들어 프로테아제를 첨가함으로써 단백질을 제거하고; 예를 들어 RNase를 첨가함으로써 RNA를 제거하고; 예를 들어 세포 용해 단계 동안에 사용된 세제, 단백질, 염 및 시약으로부터 DNA를 정제하는 것을 필요로 할 것임을 고려한다. DNA 정제는 예를 들어 에탄올 또는 이소프로판올에 의한 침전에 의해; 페놀-클로로포름 추출에 의해 수행될 수 있다.

[0204] 일부 실시양태에서, 조직 및/또는 세포로부터 수득된 세포성 DNA는 본 개시내용에 기재된 게놈 DNA 라이브러리를 수득하고, 생성하고, 형성하고/거나, 생산하기 전에 및 또는 동안에 단편화된다. 관련 기술분야의 기술자는 DNA 단편화를 위한 몇몇 적합한 기술이 있음을 이해할 것이고, 차세대 시퀀싱을 비롯하여 이로 제한되지 않는 DNA 시퀀싱을 위한 게놈 DNA 라이브러리 생성의 목적을 위한 세포성 DNA의 단편화에 적합한 기술을 인식하고 확인할 수 있다. 일부 실시양태는 세포성 DNA가 물리적 단편화, 효소적 단편화, 및 화학적 전단을 비롯하여 이로 제한되지 않는 방법에 의해 라이브러리를 생성하는데 적절한 및/또는 충분한 길이의 단편으로 단편화될 수 있음을 고려한다.

[0205] 물리적 단편화에는 음향 전단, 초음파 처리, 및 유체역학적 전단이 포함될 수 있으나 이로 제한되지 않는다. 일부 실시양태에서, 세포성 DNA는 물리적 단편화에 의해 단편화된다. 일부 실시양태에서, 세포성 DNA는 음향 전단 또는 초음파 처리에 의해 단편화된다. 일부 실시양태는 음향 전단 및 초음파 처리가 세포성 DNA를 전단시키기 위해 사용되는 일반적인 물리적 방법임을 고려한다. 코바리스(Covaris)® 장비 (메사추세츠주 워번)는 DNA를 100-5kb bp로 파괴하는 음향 기기이다. 코바리스®는 또한 메이트-페어(Mate-Pair) 라이브러리를 위해 샘플을 6-20 kb로 처리하는 튜브 (지튜브(gTubes))를 제조한다. 바이오럽터(Bioruptor)® (뉴저지주 덴빌)는 염색질, DNA 전단 및 조직 파괴를 위해 사용되는 초음파 처리 기기이다. 적은 부피의 DNA는 150-1 kb 길이로 전단될 수 있다. 디지랩(Digilab)® (메사추세츠주 말보로)으로부터의 히드로쉬어(Hydroshear)®는 DNA를 전단시키기 위해 유체역학적 힘을 이용한다. 네블라이저 (라이프 테크놀로지스(Life Technologies)®, 뉴욕주 그랜드 아일랜드)는 또한 압축 공기를 사용하여 액체를 분무하여, 수초 내에 DNA를 100-3kb 단편으로 전단시키기 위해 사용될 수 있다. 네블라이징은 저렴하지만, 과정이 원래의 샘플로부터 세포성 DNA의 약 30%의 손실을 유발할 수 있다. 일부 실시양태에서, 세포성 DNA는 초음파 처리에 의해 단편화된다.

[0206] 효소적 단편화에는 제한 엔도뉴클레아제, 예를 들어 DNase I에 의한 처리, 또는 비특이적인 뉴클레아제에 의한 처리가 포함될 수 있으나 이로 제한되지 않는다. 일부 실시양태에서, 세포성 DNA는 효소적 단편화에 의해 단편화된다. 일부 실시양태에서, 세포성 DNA는 제한 엔도뉴클레아제에 의한 처리에 의해 단편화된다. 일부 실시양태에서, 세포성 DNA는 비특이적인 뉴클레아제에 의한 처리에 의해 단편화된다. 일부 실시양태에서, 세포성 DNA는 트랜스포사제에 의한 처리에 의해 단편화된다. 일부 실시양태는 세포성 DNA를 작은 조각으로 전단시키기 위한 효소적 방법에는 말토스 결합 단백질 (MBP)-T7 Endo I 및 비특이적인 뉴클레아제 비브리오 불니피쿠스 (*Vibrio vulnificus*) (Vvn) 뉴클레아제의 조합인 DNase I, 뉴 잉글랜드 바이오랩스(New England Biolabs)® (메사추세츠주 임스위치) 프래그멘타제(Fragmentase)® 및 넥스테라(Nextera)[™] 태그 형성 기술 (일루미나®, 캘리포니아주 샌디에고)이 포함된다. 비특이적인 뉴클레아제 및 T7 Endo의 조합은 비특이적인 Nick 및 카운터 Nick을 생성하기 위해 상승작용적으로 작동하여, Nick 부위로부터 8개 미만의 뉴클레오티드를 분리시키는 단편을 생성한다. 태그 형성은 트랜스포사제를 사용하여 동시에 어댑터를 단편화하고 이중 가닥 DNA에 삽입시킨다.

[0207] 화학적 단편화는 열 및 2가 금속 양이온에 의한 처리를 포함할 수 있다. 일부 실시양태에서, 게놈 DNA는 화학적 단편화에 의해 단편화된다. 일부 실시양태는 화학적 전단이 게놈 DNA와는 반대로 긴 RNA 단편의 파괴를 위해 더욱 일반적으로 사용됨을 고려한다. 화학적 단편화는 전형적으로 2가 금속 양이온 (마그네슘 또는 아연)에 의한 DNA의 열 소화를 통해 수행된다. DNA 단편의 길이는 인큐베이션 시간을 증가시키거나 또는 감소시킴으로써 조정될 수 있다.

[0208] 일부 실시양태에서, 게놈 DNA는 200 bp 단편을 생성하는데 적합한 울트라-소니케이터 (코바리스®)를 사용하여 초음파 처리에 의해 단편화될 수 있다.

[0209] 일부 실시양태에서, 생성된 단편은 상자성 앰퓨어 XP(AMPure XP)® 비드 (베크만(Beckman)®)에 의한 "양면" 비드 정제를 이용하여 추가로 정제되고 크기 선택될 수 있다.

[0210] 일부 실시양태에서, 전단된 세포주 DNA의 혼합물은 연구 목적에 적합하게 다양한 비를 가질 수 있고, 이들은 (X

및/또는 Y 염색체 상의 유전자를 고려하기 위해) 여성 및/또는 남성 대상체로부터의 WT cfDNA와 블렌딩하여, 단일 뉴클레오티드 변이체 (SNV), 예컨대 단일 유전자 다형성 (SNP), 삽입 및/또는 결실 (Indel), 유전자 배열, 예컨대 전좌, 융합, 역위, 복제 (카피 수 변화), 및 정의된 대립유전자 빈도 (AF)의 다른 변이체를 갖는 실험실-생성 샘플을 생성할 수 있다.

[0211] 일부 실시양태에서, 본 개시내용에서 고려되는 방법 및 조성물은 저통과 전체 게놈 라이브러리, 애플리콘 라이브러리, 전체 엑솜 라이브러리, cDNA 라이브러리, 또는 메틸화 DNA 라이브러리로부터 수득되는 dsDNA를 사용한다.

[0212] 일부 실시양태에서, 본 개시내용의 방법 및 조성물은 피분석물로서 표 1에 기재된 DNA 샘플 중 어느 하나를 사용한다. 예를 들어, 일부 실시양태에서, 본 개시내용에서 고려되는 방법 및 조성물은 피분석물로서 무세포 DNA (cfDNA)를 사용한다. 일부 실시양태에서, 피분석물로서 사용되는 DNA 샘플은 합성 DNA, 게놈 DNA, 또는 그의 혼합물을 포함한다. 일부 실시양태에서, 피분석물로서 사용되는 DNA 샘플은 HRD (상동성 복구 결핍) 유전자 변이체, 예컨대 하기 유전자 중 어느 하나에서의 변이체를 포함한다: ATM, BRCA1, BRCA2, FANCA, HDAC2, PALB2, ERBB2, TP53, EML4-A1k, EGFR. 일부 실시양태에서, 피분석물로서 사용되는 DNA 샘플은 폐암 유전자 변이체를 포함한다. 일부 실시양태에서, 피분석물로서 사용되는 DNA 샘플은 세포주, 예컨대 NA12878, PC-3 또는 H2228로부터의 DNA를 포함한다.

[0213] 일부 실시양태에서, 약 10 내지 약 250 ng의 샘플 DNA가 분석을 위해 사용된다. 예를 들어, 일부 실시양태에서, 약 1 내지 약 100 ng, 약 1 내지 약 50, 또는 약 1 내지 약 25 ng의 DNA가 사용된다. 일부 실시양태에서, 약 20, 약 25, 또는 약 50 ng의 DNA가 사용된다.

[0214] 일부 실시양태에서, 피분석물로서 사용되는 cfDNA의 크기 분포는 약 150 bp 내지 약 180 bp 단편의 범위이다. 일부 실시양태에서, cfDNA의 크기 분포는 150 bp 내지 180 bp 단편의 범위이다. cfDNA의 단편화는 엔도뉴클레올리틱 및/또는 엑소뉴클레올리틱 활성의 결과일 수 있고, cfDNA의 정확하고 신뢰할 수 있고 확고한 분석을 위한 엄청난 도전을 제시한다. cfDNA 분석을 위한 또 다른 도전은 혈류에서 약 15분 정도의 그의 짧은 반감기이다. 임의의 특정한 이론에 구애되기를 바라진 않지만, 본 개시내용은 부분적으로 cfDNA의 분석이 "액체 생검"과 유사하고 현재의 생물학적 과정의 실시간 스냅샷임을 고려한다.

[0215] 더욱이, cfDNA가 세포 내에서 발견되지 않고, 생물학적 유체 및 대변 샘플을 비롯하여 이로 제한되지 않는 수많은 적합한 공급원으로부터 수득될 수 있기 때문에, 이는 차세대 시퀀싱 분석, 예컨대 분석할 조직으로의 직접적인 접근을 방해하는 기존 제한에 적용되지 않는다.

[0216] 일부 실시양태에서, 본원에서 고려되는 유전자 분석 방법은 cfDNA를 하나 이상의 말단-복구 효소로 처리하여 말단 복구 cfDNA를 생성하고, 하나 이상의 어댑터를 말단 복구 cfDNA의 각각의 말단에 라이게이션하여 cfDNA 라이브러리를 생성하는 것을 포함하는, cfDNA 라이브러리를 생성하는 것을 포함한다.

[0217] 일부 실시양태에서 cfDNA를 단리하는데 적합한 공급원인 생물학적 유체의 예시적인 예에는 양수, 혈액, 혈장, 혈청, 정액, 림프액, 뇌척수액, 안구액, 소변, 타액, 점액, 및 땀이 포함되나 이로 제한되지 않는다.

[0218] 일부 실시양태에서, 생물학적 유체는 혈액 또는 혈장이다.

[0219] 일부 실시양태에서, 상업적으로 입수가능한 키트 및 통상의 기술자에게 공지된 다른 방법을 이용하여, 예를 들어 냉동 및/또는 EDTA, EGTA, 또는 2가 양이온에 대해 특이적인 다른 킬레이팅제를 비롯하여 이로 제한되지 않는 효소 킬레이팅제의 첨가에 의해 대상체의 생물학적 유체로부터 또는 이전에 수득되고 임의적으로 안정화된 생물학적 샘플로부터 cfDNA를 직접적으로 단리할 수 있다.

[0220] 일부 실시양태에서, 유전자 변이체 (코리엘 인스티튜트 포 메디컬 리서치(Coriell Institute for Medical Research) 또는 세라케어 라이프 사이언시즈, 인크.(SeraCare Life Sciences, Inc.))를 보유하는 불멸화된 세포로부터 단리된 무세포 DNA 또는 게놈 DNA (예를 들어 cfDNA 또는 gDNA)는 NGS 라이브러리 구축을 위해 사용될 수 있다.

[0221] 일부 실시양태에서, 무세포 DNA는 QIAmp DSP 순환 NA 키트 (키아젠(Qiagen))를 사용하여 혈장 샘플로부터 추출될 수 있다.

[0222] **단일-단계 DNA 말단-복구**

[0223] 본 개시내용의 DNA 단편이 처리된 형태로 수득될 수 있지만, 본 개시내용의 방법은 본 개시내용의 어댑터에 라

이게이션할 수 있는 DNA 단편을 획득하기 위해 생물학적 샘플의 처리를 가능하게 한다. 예를 들어, 본 개시내용의 DNA 단편의 처리된 형태에는 평활 말단, 평활 3' 말단, 평활 5' 말단, 데옥시리보핵산 아데닌 (dA)-꼬리, 3' 말단의 dA-꼬리, 5' 말단의 dA-꼬리, 인산화된 핵산, 3' 말단의 인산화된 핵산, 및 5' 말단의 인산화된 핵산 중 하나 이상을 포함하는 DNA 단편이 포함되나 이로 제한되지 않는다.

- [0224] 일부 실시양태에서, "말단 복구"는 탈인산화된, 내부 손상 복구된, 평활 말단인, 5' 인산화된 DNA 단편을 생성하기 위해, 또는 3' 오버행을 갖는 DNA 단편을 생성하기 위해 수행될 수 있다.
- [0225] DNA 단편의 처리를 포함하는 본 개시내용의 방법의 일부 실시양태에서, 말단을 평활화시키는 DNA 복구 반응, A-꼬리 부착 반응 및 인산화 반응 중 하나 이상을 단일 단계로 수행할 수 있다.
- [0226] 일부 실시양태에서, 게놈 DNA 라이브러리의 생성은 단리된 cfDNA 또는 단편화된 세포성 DNA의 말단-복구를 포함한다. 단편화된 cfDNA 또는 세포성 DNA는 말단-복구 효소에 의해 처리되어, 평활 말단, 5'-오버행, 또는 3'-오버행을 갖는 말단 복구 cfDNA를 생성한다. 일부 실시양태에서, 예를 들어 말단-복구 효소가 획득될 수 있다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA는 평활 말단을 함유하도록 처리된다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA 또는 cfDNA는 평활 말단을 함유하도록 추가로 변형된다. 일부 실시양태에서, 평활 말단을 함유하는 말단 복구 cfDNA 또는 세포성 DNA는 아데닌 (A)/티민 (T) 오버행을 함유하도록 추가로 처리될 수 있다. 일부 실시양태에서, 평활 말단을 함유하는 말단 복구 cfDNA 또는 세포성 DNA는 단일 염기 쌍 오버행으로서 아데닌 (A)/티민 (T) 오버행을 함유하도록 추가로 처리될 수 있다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA는 비-주형 3' 오버행을 갖는다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA는 3' 오버행을 함유하도록 처리된다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA는 3' 오버행을 함유하도록 말단 트랜스퍼라제 (TdT)에 의해 처리된다. 일부 실시양태에서, G-꼬리는 TdT에 의해 부가될 수 있다. 일부 실시양태에서, 말단 복구 cfDNA 또는 세포성 DNA는 임의의 공지된 제한 효소에 의한 (예를 들어, 효소 Sau3A 등에 의한) 부분적인 소화를 이용하여 오버행 말단을 함유하도록 처리된다.
- [0227] 일부 실시양태에서, DNA 단편의 탈인산화는 열불안정성 포스파타제, 예컨대 알칼리성 포스파타제에 의해 수행될 수 있다. 상업적인 예에는 APexTM 열불안정성 알칼리성 포스파타제, NTPhosTM 열불안정성 포스파타제, KTTM 열불안정성 포스파타제 및 슈림프 알칼리성 포스파타제 (SAP)가 포함된다.
- [0228] 일부 실시양태에서, 내부 DNA 손상은 DNA 단편에서 내부 손상을 복구할 수 있는 하나 이상의 복구 효소에 의해 복구될 수 있다. 예에는 Taq DNA 리가제, 엔도뉴클레아제 IV, Bst DNA 폴리머라제 Fpg, 우라실-DNA 글리콜라제 (UDG), T4 PDG, 및 엔도뉴클레아제 VIII가 포함된다. 일부 실시양태에서, 전술한 모든 효소가 사용될 수 있다. 전술한 효소의 상업적으로 입수가능한 각테일 (예를 들어 PreCR 효소 키트)을 사용할 수 있거나, 또는 하나 이상의 개별 효소를 임의의 조합으로 첨가하여 각테일을 제조할 수 있다. 일부 실시양태에서, DNA 내부 손상 복구는 수행되지 않을 수 있다.
- [0229] 일부 실시양태에서, dA-꼬리 부착을 위한 내부 DNA 손상 복구, 말단-복구, 및 말단 트랜스퍼라제 (TdT)는 단일 단계 및 단일 반응 혼합물에서 수행될 수 있다. 일부 실시양태에서, 상업적으로 입수가능한 키트, 예컨대 PreCR 효소 키트 또는 NEB로부터의 퀵 블런트 키트를 단일 단계 반응을 위해 사용할 수 있다.
- [0230] 일부 실시양태에서, DNA 말단 복구는 하나 이상의 말단-복구 효소를 사용하여 수행되어 평활 말단 DNA 단편을 생성할 수 있다. 효소에는 3'-5' 엑소뉴클레아제, 5' - 3' DNA 폴리머라제 (예를 들어 클레노브(Klenow) 단편), 및 5' FLAP 엔도뉴클레아제가 포함될 수 있다.
- [0231] 일부 실시양태에서, dA-꼬리 부착을 위한 DNA 말단-복구, 5' 인산화, 및 말단 트랜스퍼라제 (TdT)는 단일 단계 및 단일 반응 혼합물에서 수행되어, 3'-오버행 말단으로 5' 인산화되고, 예를 들어 5' 인산화되고 3'dA-꼬리 부착된 dsDNA 단편을 생성할 수 있다. 일부 실시양태에서, 상업적으로 입수가능한 키트, 예컨대 NEB로부터의 넥스트 울트라 II 엔드 리페어/dA-꼬리 부착 키트는 단일 단계 반응을 위해 사용될 수 있다.
- [0232] 일부 실시양태에서, 본 개시내용은 적당한 양의 단편화된 DNA 샘플이 하기 각각의 반응에 대한 효소 및 시약을 단일 혼합물로 조합함으로써 "단일-단계 말단 복구"될 수 있음을 고려한다: 탈인산화, 내부 DNA 손상 복구, 평활 말단 생성, 5' 말단 인산화, 및 3' 오버행 생성. 이 단일-단계 단일-반응 과정은 5' 인산화된 말단 및 3' 오버행을 갖는 말단 복구 이중 가닥 DNA 단편을 생성한다. 일부 실시양태에서, 3' 오버행은 dA 꼬리를 포함한다.

- [0233] 일부 실시양태에서, 말단 복구될 수 있는 DNA의 양은 임의의 적합한 양일 수 있다. 일부 실시양태에서, 말단 복구되는 DNA의 양은 1 ng 내지 500 ng, 5 ng 내지 400 ng, 10 ng 내지 300 ng, 15 ng 내지 250 ng, 또는 20 ng 내지 200 ng이다. 일부 실시양태에서, 말단 복구되는 DNA의 양은 20 ng 내지 50 ng이다.
- [0234] **말단 복구 DNA에 대한 어댑터 라이게이션**
- [0235] 일부 실시양태에서, 라이게이션 단계는 어댑터 모듈을 말단 복구 cfDNA에 라이게이션하여 "태그 부착된" cfDNA 라이브러리를 생성하는 것을 포함한다. 일부 실시양태에서, 단일 어댑터 모듈이 사용된다. 일부 실시양태에서, 2, 3, 4 또는 5개의 어댑터 모듈이 사용된다. 일부 실시양태에서, 동일한 서열의 어댑터 모듈은 단편화된 말단 복구 DNA의 각각의 말단에 라이게이션된다. 일부 실시양태에서, 동일하지 않은 서열의 어댑터 모듈은 각각의 단편화된 말단 복구 DNA의 두 말단에 라이게이션된다.
- [0236] 본원에서 고려되는 하나 이상의 어댑터의 라이게이션은 관련 기술분야의 통상의 기술자에게 공지된 방법에 의해 수행될 수 있다. 일부 실시양태에서, 본원에서 고려되는 하나 이상의 어댑터는 평활 말단을 포함하는 말단 복구 cfDNA에 라이게이션된다. 일부 실시양태에서, 하나 이상의 어댑터는 이용된 라이게이션 방법에 적당한 상보성 말단을 포함하는 말단 복구 cfDNA에 라이게이션된다. 일부 실시양태에서, 하나 이상의 어댑터는 3' 오버행을 포함하는 말단 복구 cfDNA에 라이게이션된다.
- [0237] 일부 실시양태에서, 복수개의 어댑터에 대한 게놈 DNA 단편의 부착은 앵커 영역의 적어도 일부분을 함유하는 올리고뉴클레오타이드에 말단 복구 cfDNA 또는 세포성 DNA 단편을 부착시키는 단계를 포함한다. 일부 실시양태에서, 올리고뉴클레오타이드는 전체 앵커 영역을 함유한다. 일부 실시양태에서, 올리고뉴클레오타이드는 파트너 가닥과 듀플렉스화된 5' 인산화 부착 가닥을 포함하는 DNA 듀플렉스이며, 파트너 가닥은 그의 3' 말단에서 화학적 변형에 의해 부착으로부터 차단되고, 부착 가닥은 게놈 DNA 단편에 부착된다. 이어서, 일부 실시양태에서, 앵커 영역의 적어도 일부분에 부착된 DNA 단편은 전장 어댑터 서열을 코딩하는 DNA 올리고뉴클레오타이드와 어닐링된다. 일부 실시양태에서, 하나 이상의 폴리뉴클레오타이드 키나제, 하나 이상의 DNA 리가제, 및/또는 하나 이상의 DNA 폴리머라제는 전장 어댑터 서열을 코딩하는 게놈 DNA 단편 및 DNA 올리고뉴클레오타이드에 부가된다. 일부 실시양태에서, 폴리뉴클레오타이드 키나제는 T4 폴리뉴클레오타이드 키나제이다. 일부 실시양태에서, DNA 리가제는 Taq DNA 리가제이다. 일부 실시양태에서, DNA 폴리머라제는 Taq 폴리머라제이다. 일부 실시양태에서, DNA 폴리머라제는 전장 Bst 폴리머라제이다.
- [0238] 일부 실시양태에서, 어댑터 및 DNA 단편은 라이게이션 완충제, 시약 및 라이게이션 효소, 예컨대 DNA 리가제 (예를 들어 T4 리가제 또는 Taq 리가제) 및/또는 RNA 리가제와 혼합될 수 있다. 이러한 리가제는 상기 기재된 바와 같이 3' 오버행을 갖는 단일 가닥 라이게이션 가닥을 단일 가닥 DNA 단편에 라이게이션하기 위해 사용될 수 있다.
- [0239] 일부 실시양태에서, 다기능성 어댑터의 라이게이션 가닥은 라이게이션 가닥의 3' 말단 오버행 및 DNA 단편의 3' 오버행의 보환을 통해 단일 단계에서 dsDNA 단편의 5' 말단에 라이게이션되는 반면에, 비-라이게이션 가닥은 DNA 단편의 3' 말단에 부착되지 않고 남아있다.
- [0240] 일부 실시양태에서, 라이게이션 단계는 다기능성 어댑터를 dsDNA 단편과 라이게이션하여 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 것을 포함한다. 일부 실시양태에서, 단일 어댑터가 사용된다. 일부 실시양태에서, 2, 3, 4 또는 5개의 어댑터가 사용된다. 일부 실시양태에서, 동일한 서열의 어댑터 모듈이 단편화된 말단 복구 DNA의 각각의 말단에 부착된다.
- [0241] 일부 실시양태에서, 동일한 어댑터는 DNA 단편의 양 말단에 부착된다. 일부 실시양태에서, 상이한 어댑터는 dsDNA 단편의 상이한 말단에 부착된다.
- [0242] 일부 실시양태에서, 라이게이션 단계는
- [0243] (a) 복수개의 다기능성 어댑터를 복수개의 dsDNA 단편과 라이게이션하여, 복수개의 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 단계이며, 여기서 다기능성 어댑터는 본 개시내용의 다기능성 어댑터 중 어느 하나인 단계;
- [0244] (b) 단계 (a)로부터의 어댑터/DNA 단편 복합체를 하나 이상의 효소와 접촉시켜, 복수개의 연속된 이중-가닥 어댑터-태그 부착된 DNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 생성하는 단계를 포함한다.
- [0245] 일부 실시양태에서, 단계 (b)에서 어댑터/DNA 단편 복합체는 주형으로서 라이게이션 가닥을 사용하는 DNA 폴리머라제 연장에 의해 연속된 이중 가닥 어댑터-태그 부착된 DNA 단편으로 제조된다.

- [0246] 일부 실시양태에서, 부착되지 않은 비-라이게이션 가닥은 주형으로서 라이게이션 가닥을 사용하는 DNA 단편의 5'-3' 폴리머라제 연장에 의해 대체된다. 일부 실시양태에서, 비-라이게이션 가닥은 임의적으로 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 포함할 수 있다.
- [0247] 일부 실시양태에서, 비-라이게이션 가닥은 주형으로서 라이게이션 가닥을 사용하는 DNA 폴리머라제 Nick-복구 (Nick 번역)에 의해 DNA 단편의 3' 말단에 라이게이션된다.
- [0248] 일부 실시양태에서, dsDNA 단편은 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 또는 메틸화 DNA, 또는 탈메틸화 DNA이다.
- [0249] 일부 실시양태에서, 복수개의 dsDNA 단편은 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구된다.
- [0250] 일부 실시양태에서, 복수개의 dsDNA 단편은 저통과 전체 게놈 라이브러리, 앰플리콘 라이브러리, 전체 엑솜 라이브러리, cDNA 라이브러리, 또는 메틸화 DNA 라이브러리로 이루어진 목록으로부터 선택된 라이브러리로부터 취득된다.
- [0251] 일부 실시양태에서, 어댑터 라이게이션 기간은 라이게이션에 적합한 임의의 기간일 수 있다. 일부 실시양태에서, 기간은 적어도 약 5분이다. 일부 실시양태에서, 기간은 약 5분 내지 약 72시간이다. 일부 실시양태에서, 기간은 약 5분 내지 약 2시간이다. 일부 실시양태에서, 라이게이션 기간은 약 1시간 미만, 약 30분 미만, 약 15분 미만, 또는 약 10분 미만이다.
- [0252] 일부 실시양태에서, 어댑터 라이게이션 기간은 라이게이션에 적합한 임의의 기간일 수 있다. 일부 실시양태에서, 기간은 적어도 5분이다. 일부 실시양태에서, 기간은 5분 내지 72시간이다. 일부 실시양태에서, 기간은 5분 내지 2시간이다. 일부 실시양태에서, 라이게이션 기간은 1시간 미만, 30분 미만, 15분 미만, 또는 10분 미만이다.
- [0253] 일부 실시양태에서, 어댑터 라이게이션 부피는 자동화 및 샘플 취급 로봇에 적합한 것이다. 일부 실시양태에서, 반응 부피는 약 1 μL 내지 약 1000 μL , 약 1 μL 내지 약 350 μL , 약 1 μL 내지 약 200 μL , 약 1 μL 내지 약 100 μL , 약 1 μL 내지 약 50 μL , 약 5 μL 내지 약 25 μL , 약 10 μL 내지 약 40 μL , 약 20 μL 내지 약 40 μL 이다. 일부 실시양태에서, 반응 부피는 약 100 μL 이다. 일부 실시양태에서, 부피는 약 30 μL 이다.
- [0254] 일부 실시양태에서, 어댑터 라이게이션 부피는 자동화 및 샘플 취급 로봇에 적합한 것이다. 일부 실시양태에서, 반응 부피는 1 μL 내지 1000 μL , 1 μL 내지 350 μL , 1 μL 내지 200 μL , 1 μL 내지 100 μL , 1 μL 내지 50 μL , 5 μL 및 25 μL , 10 μL 내지 40 μL , 20 μL 내지 40 μL 이다. 일부 실시양태에서, 반응 부피는 100 μL 이다. 일부 실시양태에서, 부피는 30 μL 이다.
- [0255] 일부 실시양태에서, 어댑터 라이게이션은 튜브의 스트립에서 또는 미세역가 플레이트 웰에서 또는 자동화된 및/또는 고처리량 처리를 가능하게 하는데 적합한 임의의 다른 포맷에서 수행될 수 있다.
- [0256] 일부 실시양태에서, 어댑터 양은 임의의 적합한 농도일 수 있다. 일부 실시양태에서, 어댑터의 농도는 적어도 0.01 μM 이다. 일부 실시양태에서, 어댑터의 농도는 약 0.01 μM 내지 약 200 μM , 약 0.01 μM 내지 약 50 μM , 약 0.1 μM 내지 약 50 μM , 약 0.2 μM 내지 약 20 μM , 약 0.2 μM 내지 약 10 μM , 약 1 μM 내지 약 10 μM , 또는 약 2 μM 내지 약 8 μM 이다. 일부 실시양태에서, 어댑터 농도는 적어도 약 2 μM 이다. 일부 실시양태에서, 어댑터 농도는 약 5 μM 이다.
- [0257] 일부 실시양태에서, 어댑터의 농도는 0.01 μM 내지 200 μM , 0.01 μM 내지 50 μM , 0.1 μM 내지 50 μM , 0.2 μM 내지 20 μM , 0.2 μM 내지 10 μM , 1 μM 내지 10 μM , 또는 2 μM 내지 8 μM 이다. 일부 실시양태에서, 어댑터 농도는 적어도 2 μM 이다. 일부 실시양태에서, 어댑터 농도는 5 μM 이다.
- [0258] 일부 실시양태에서, 라이게이션 반응 혼합물은 약 10°C 내지 약 30°C의 온도에서 인큐베이션될 수 있다. 일부 실시양태에서, 라이게이션 반응 혼합물은 약 20°C의 온도에서 인큐베이션될 수 있다.
- [0259] 일부 실시양태에서, 라이게이션 반응 혼합물은 10°C 내지 30°C의 온도에서 인큐베이션될 수 있다. 일부 실시양태에서, 라이게이션 반응 혼합물은 20°C의 온도에서 인큐베이션될 수 있다.
- [0260] 일부 실시양태에서, 라이게이션 후에, 어댑터-태그 부착된 DNA 분자는 분리되고 세척될 수 있다. 이는 어댑터-DNA 분자가 비드에 부착되어 유지되고 오염 물질이 세척되도록 DNA 정제 비드, 예컨대 앰플이어 XP® (베크만®) 또는 스펙트라(Spectra) 믹스를 사용하여 수행될 수 있다. 용리되고 정화된 상청액은 복수개의 어댑터-태그 부

착된 DNA 단편을 포함하는 단리된 라이브러리를 함유한다. 라이브러리를 함유하는 상청액은 증폭을 위해 신선한 PCR 튜브 또는 마이크로리터 플레이트 웰로 옮겨질 수 있다.

[0261] **DNA 라이브러리 증폭**

[0262] 일부 실시양태에서, dsDNA 단편의 5'-3' 연장을 먼저 수행하여 연속된 어댑터-태그 부착된 dsDNA 단편을 생성한 다음, 연속된 어댑터-태그 부착된 dsDNA 단편을 증폭시켜 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 생성한다.

[0263] 일부 실시양태에서, 연속된 어댑터-태그 부착된 dsDNA 단편을 형성하는 dsDNA 단편의 5'-3'연장의 제1 단계 및 연속된 어댑터-태그 부착된 dsDNA 단편을 증폭시키는 제2 단계는 조합되어, 단일 단계에서 어댑터-태그 부착된 DNA 라이브러리를 생성한다. DNA 라이브러리에서 어댑터-태그 부착된 dsDNA 단편은 동일한 증폭 영역을 포함하는 양 말단에서 어댑터에 의해 플랭킹되고, 증폭 영역의 서열은 단일 증폭 프라이머, 예컨대 PCR 증폭 프라이머에 의해 인식가능한 증폭 프라이머 결합 부위로서 기능할 수 있다.

[0264] 일부 실시양태에서, 부착되지 않은 비-라이게이션 가닥은 주형으로서 라이게이션 가닥을 사용하는 DNA 단편의 5'-3'폴리머라제 연장에 의해 대체된다. 일부 실시양태에서, 비-라이게이션 가닥은 임의적으로 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 포함한다.

[0265] 일부 실시양태에서, 비-라이게이션 가닥은 주형으로서 라이게이션 가닥을 사용하여 DNA 폴리머라제 Nick-복구에 의해 DNA 단편의 3' 말단에 라이게이션된다.

[0266] 일부 실시양태에서, DNA 폴리머라제가 사용된다. DNA 폴리머라제는 PCR 또는 항온/등온 증폭에 대해 호열성일 수 있다. 일부 실시양태에서, 5'-3' 연장을 위한 시약 및 효소 및 후속적인 증폭을 위한 효소를 함유하는 마스터 믹스 (Master Mix, MM)가 조합된다. 이러한 연장 및 증폭을 위한 상업적으로 입수가능한 효소 및 시약 키트에는 예를 들어 NEB 울트라 II 2X PCR 앰플리피케이션(NEB Ultra II 2X PCR Amplification)® 키트 (뉴 잉글랜드 바이오랩스®), 하이-피델리티 Q5(Hi-Fidelity Q5)® 효소 PCR (NEB), KAPA (로슈(Roche)®) KAPA 2X (로슈®), 트루세크 나노(TruSeq Nano)® (써모피셔(Thermofisher)®) 앰플리타크(AmpliTaq)® (써모피셔®)이 포함된다.

[0267] 일부 실시양태에서, 어댑터-태그 부착된 DNA 라이브러리의 일부는 증폭을 유도하는 단일 프라이머 서열과 함께 표준 PCR 기술을 이용하여 증폭될 것이다. 일부 실시양태에서, 단일 프라이머 서열은 약 25개 뉴클레오티드이고, 임의적으로 표준 이온 강도 조건하에 $\geq 55^{\circ}\text{C}$ 의 투영 T_m 을 갖는다. 일부 실시양태에서, 단일 프라이머 서열은 25개 뉴클레오티드이고, 임의적으로 표준 이온 강도 조건하에 $\geq 55^{\circ}\text{C}$ 의 투영 T_m 을 갖는다. 일부 실시양태에서, 단일 증폭 프라이머는 어댑터 모듈의 증폭 영역 내의 서열에 대해 상보성이다. 일부 실시양태에서, 단일 증폭 프라이머는 TGCAGGACCAGAGAATTCGAATACA (서열식별번호: 70)의 서열을 포함한다.

[0268] 일부 실시양태에서, 증폭은 관련 기술분야에 공지된 임의의 증폭, 예컨대 PCR (폴리머라제 연쇄 반응), LAMP (루프-매개된 등온 증폭), NASBA (핵산 서열-기반 증폭), SDA (가닥 대체 증폭), RCA (회전 환 증폭), LCR (리가제 연쇄 반응)에 의해 수행될 수 있다.

[0269] 일부 실시양태에서, 증폭 동안에 일부 앰플리콘은 단편의 양 말단에 라이게이션되는 어댑터로 인해 스태م-루프 구조를 형성할 것이다. 이 전략은 매우 짧은 생성물 (예를 들어, 프라이머 이량체)이 증폭되고 생성된 라이브러리를 편향시키는 것을 방지하는데 효율적이다.

[0270] 일부 실시양태에서, 초기 3분 인큐베이션 주기를 수행하여, 라이게이션 가닥 주형 연장에 의해 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편을 형성한다.

[0271] 일부 실시양태에서, PCR 증폭을 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편에 대해 수행하여, 증폭된 태그 부착된 DNA 라이브러리를 형성한다.

[0272] 일부 실시양태에서, 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편의 피코그램은 마이크로그램의 DNA 클론 (어댑터-태그 부착된 DNA 라이브러리)으로 증폭시키고, 이는 10,000배 증폭을 의미한다. 증폭된 생성물의 양은 관련 기술분야에 공지된 방법, 예를 들어 큐빗(Qubit) 2.0 또는 나노드롭(Nanodrop) 장비 상에서의 정량화를 이용하여 측정될 수 있다.

[0273] 일부 실시양태에서, 증폭된 어댑터-태그 부착된 DNA 라이브러리는 DNA 정제 비드를 사용하여 단리되고 세척 완충제, 예를 들어 트리스-EDTA 완충제 (TEZ) pH 8.0에 의해 세척될 수 있다. 정화된 상청액은 자동화 및/또는

고처리량에 적합한 신선한 PCR 튜브 또는 미세역가 플레이트 웰 또는 임의의 다른 포맷으로 옮겨질 수 있다.

[0274] 일반적으로, 라이브러리를 증폭시키기 위해 가능한 한 적은 PCR 주기를 이용하는 것이 바람직하다. 워크플로우 시간을 감소시키는 것 외에도, 이는 또한 PCR 동안에 편향이 도입될 위험을 제한한다. 본 개시내용의 방법에서 말단 복구, dA-꼬리 부착 및 어댑터 라이게이션의 증가된 효율성의 결과는 더 적은 PCR 주기가 시퀀싱 또는 다른 중간 하류 워크플로우에 필요한 라이브러리 수율을 달성하기 위해 필요하다는 것이다. 개시된 워크플로우 및 과정의 간소화는 감소된 처리 시간, 감소된 시약 수, 더 적은 장비/기계 사용, 및 감소된 비용과 같은 이점을 제공한다.

[0275] 이들 상기 과정, 어댑터, 및 태그 부착된 DNA 라이브러리는 임의의 시험 샘플에 존재하는 관심 유전자좌가 풍부한 포획 프로브 라이브러리를 제조하기 위해 사용될 수 있다.

[0276] 2. 표적 포획 및 단리

[0277] 일부 실시양태에서, 게놈 DNA, 예를 들어, 게놈 세포성 또는 cfDNA의 유전자 분석을 위한 방법은 DNA 라이브러리 클론의 하나 이상의 표적 유전자좌의 정량적인 유전자 분석을 포함한다. 정량적인 유전자 분석은 하기 단계 중 하나 이상 또는 모두를 포함한다: 표적 유전자좌를 포함하는 DNA 클론의 포획; 포획되고 표적화된 유전자좌의 증폭; 증폭되고 포획되고 표적화된 유전자좌의 시퀀싱; 및 생성된 서열 리드의 생물정보학적 분석. 본원에 사용된 바와 같이, 용어 "DNA 라이브러리 클론"은 DNA 라이브러리 단편을 지칭하고, 어댑터 및 게놈 DNA 단편의 조합은 고유의 DNA 서열 (예를 들어, 또 다른 DNA 라이브러리 클론과 구별될 수 있는 DNA 서열)을 생성한다.

[0278] 본 개시내용은 부분적으로 더 큰 프로브의 효율성 및 신뢰성을 유지하지만, 더 작은 DNA 단편을 포함하는 게놈 DNA 라이브러리, 예를 들어 cfDNA 클론 라이브러리에서 정보가 없는 서열 생성을 최소화하도록 설계된 포획 프로브 모듈을 고려한다.

[0279] 용어 "다기능성 포획 프로브" 및 "포획 프로브 모듈"은 상호교환적으로 사용된다. 일부 실시양태에서, "포획 프로브 모듈" 또는 "다기능성 포획 프로브"는 포획 프로브 서열 및 꼬리 서열을 포함하고, 포획 프로브 서열은 태그 부착된 유전자 DNA 라이브러리에서 표적 영역에 혼성화할 수 있다. 일부 실시양태에서, 다기능성 포획 프로브는 파트너 올리고뉴클레오타이드에 혼성화할 수 있는 제1 영역; 및 태그 부착된 유전자 DNA 라이브러리에서 특이적인 표적 영역에 혼성화할 수 있는 제2 영역을 포함하며, 여기서 임의적으로 제1 영역은 PCR 프라이머 결합 부위를 포함하는 꼬리 서열을 포함한다. 제1 영역은 또한 꼬리 영역으로도 명명될 수 있고, 제2 영역은 또한 포획 프로브 또는 포획 프로브 서열로도 명명될 수 있다.

[0280] 일부 실시양태에서, 포획 프로브 모듈은 꼬리 서열을 포함한다. 본원에 사용된 바와 같이, 용어 "꼬리 서열"은 포획 프로브 모듈의 5' 말단에 있는 폴리뉴클레오타이드를 지칭하고, 일부 실시양태에서 프라이머 결합 부위의 역할을 할 수 있다. 일부 실시양태에서, 포획 프로브는 시퀀싱 프라이머 결합 부위를 포함한다.

[0281] 일부 실시양태에서, 꼬리 서열은 약 5 내지 약 100개 뉴클레오타이드, 약 10 내지 약 100개 뉴클레오타이드, 약 5 내지 약 75개 뉴클레오타이드, 약 5 내지 약 50개 뉴클레오타이드, 약 5 내지 약 25개 뉴클레오타이드, 또는 약 5 내지 약 20개 뉴클레오타이드이다. 일부 실시양태에서, 제3 영역은 약 10 내지 약 50개 뉴클레오타이드, 약 15 내지 약 40개 뉴클레오타이드, 약 20 내지 약 30개 뉴클레오타이드 또는 약 20개 뉴클레오타이드, 또는 임의의 중간 개수의 뉴클레오타이드이다.

[0282] 일부 실시양태에서, 꼬리 서열은 약 30개 뉴클레오타이드, 약 31개 뉴클레오타이드, 약 32개 뉴클레오타이드, 약 33개 뉴클레오타이드, 약 34개 뉴클레오타이드, 약 35개 뉴클레오타이드, 약 36개 뉴클레오타이드, 약 37개 뉴클레오타이드, 약 38개 뉴클레오타이드, 약 39개 뉴클레오타이드, 또는 약 40개 뉴클레오타이드이다.

[0283] 일부 실시양태에서, 꼬리 서열은 5 내지 100개 뉴클레오타이드, 10 내지 100개 뉴클레오타이드, 5 내지 75개 뉴클레오타이드, 5 내지 50개 뉴클레오타이드, 5 내지 25개 뉴클레오타이드, 또는 5 내지 20개 뉴클레오타이드이다. 일부 실시양태에서, 제3 영역은 10 내지 50개 뉴클레오타이드, 15 내지 40개 뉴클레오타이드, 20 내지 30개 뉴클레오타이드 또는 20개 뉴클레오타이드, 또는 임의의 중간 개수의 뉴클레오타이드이다.

[0284] 일부 실시양태에서, 꼬리 서열은 30개 뉴클레오타이드, 31개 뉴클레오타이드, 32개 뉴클레오타이드, 33개 뉴클레오타이드, 34개 뉴클레오타이드, 35개 뉴클레오타이드, 36개 뉴클레오타이드, 37개 뉴클레오타이드, 38개 뉴클레오타이드, 39개 뉴클레오타이드, 또는 40개 뉴클레오타이드이다.

[0285] 일부 실시양태에서, 예시적인 파트너 올리고뉴클레오타이드는 GTGAAAACCAGGATCAACTCCCGTGCCAGTCACATCTCAGATGAGCT

(서열식별번호: 1)일 수 있으며, 3' 말단에 비오틴-TEG 변형을 갖는다.

- [0286] 연속된 어댑터-태그 부착된 DNA 단편 (증폭되지 않은) 및 태그 부착된 DNA 라이브러리 (증폭된) 각각은 하나 이상의 관심 유전자좌가 풍부한 혼성체 분자를 함유하는 라이브러리의 제조를 포함하는 다양한 시퀀싱-기반 유전자 분석에 유용하며, 증폭되지 않은 또는 증폭된 라이브러리 ("표적화된 라이브러리")일 수 있다.
- [0287] 상기 기재된 바와 같이 제조된 증폭되지 않은 어댑터-태그 부착된 DNA 단편 및/또는 증폭된 태그 부착된 DNA 라이브러리는 다기능성 포획 프로브 모듈에 혼성화되어 특이적인 유전자좌에 대해 표적화된 라이브러리, 즉, 표적화된 라이브러리를 생성할 수 있다. 어댑터-태그 부착된 DNA 단편은 1개 이상의 포획 프로브와 혼성화될 수 있다. 각각의 포획 프로브는 어댑터-태그 부착된 DNA 단편에서 동일한 유전자좌를 표적화할 수 있거나 또는 이들은 어댑터-태그 부착된 DNA 단편에서 상이한 유전자좌를 표적화할 수 있다. 일부 실시양태에서, 증폭된 태그 부착된 DNA 라이브러리 단편에서 복수개의 유전자좌가 표적화된다.
- [0288] 일부 실시양태에서, 포획 프로브는 세포성 DNA로부터 구축된 게놈 DNA 라이브러리와 함께 사용된다. 일부 실시양태에서, 포획 프로브는 cfDNA로부터 구축된 게놈 DNA 라이브러리와 함께 사용된다. cfDNA의 평균 크기가 약 150 내지 약 170 bp이고, 고도로 단편화되기 때문에, 일부 실시양태는 관심 DNA 표적 영역을 조사하기 위해 높은 밀도 및 비교적 짧은 포획 프로브의 사용을 포함하는 본원에서 고려되는 조성물 및 방법에 관한 것이다. 일부 실시양태에서, 포획 프로브는 모든 염색체 세그먼트에 걸쳐 균일한 밀도로 분포되는 DNA 표적 영역에 혼성화할 수 있다. 이러한 포획 프로브의 세트는 본원에서 "염색체 안정성 프로브"로 지칭된다. 염색체 안정성 프로브는 염색체 카피 수 (예를 들어, 염색체 배수성)의 게놈-수준 측정을 제공하기 위해 게놈-수준 규모에서 카피 수 변동을 조사하는데 사용된다.
- [0289] 고밀도 포획 프로브의 사용과 관련한 한 가지 특정한 우려는 일반적으로 포획 프로브가 특이적인 "서열 규칙"을 이용하여 설계된다는 것이다. 예를 들어, 중복 서열의 영역 또는 극심한 염기 조성 편향을 나타내는 영역은 일반적으로 포획 프로브의 설계에서 제외된다. 그러나, 포획 프로브 설계 규칙의 융통성의 부족이 프로브 성능에 실질적인 영향을 미치지 않는 것으로 밝혀졌다. 대조적으로, 위치 제한에 의해 엄격하게 선택된 포획 프로브는 온-타겟 서열 정보를 제공하였고; 매우 적은 오프-타겟 및 맵핑할 수 없는 리드 포획을 나타내며; 단지 몇개의 예외를 갖는 균일하고 유용한 온-타겟 리드를 생성한다. 더욱이, 가까운 프로브 간격에서 높은 중복성은 간헐적인 불량한 성능의 포획 프로브를 보상하는 것 이상이다.
- [0290] 일부 실시양태에서, 표적 영역은 복수개의 포획 프로브에 의해 표적화되고, 임의의 2개 이상의 포획 프로브는 서로 10개 뉴클레오타이드 이내, 서로 15개 뉴클레오타이드 이내, 서로 20개 뉴클레오타이드 이내, 서로 25개 뉴클레오타이드 이내, 서로 30개 뉴클레오타이드 이내, 서로 35개 뉴클레오타이드 이내, 서로 40개 뉴클레오타이드 이내, 서로 45개 뉴클레오타이드 이내, 또는 서로 50개 뉴클레오타이드 또는 그 초과 이내, 뿐만 아니라 모든 중간 뉴클레오타이드 길이에서 표적 영역에 결합하도록 설계된다.
- [0291] 일부 실시양태에서, 포획 프로브는 약 25개 뉴클레오타이드, 약 26개 뉴클레오타이드, 약 27개 뉴클레오타이드, 약 28개 뉴클레오타이드, 약 29개 뉴클레오타이드, 약 30개 뉴클레오타이드, 약 31개 뉴클레오타이드, 약 32개 뉴클레오타이드, 약 33개 뉴클레오타이드, 약 34개 뉴클레오타이드, 약 35개 뉴클레오타이드, 약 36개 뉴클레오타이드, 약 37개 뉴클레오타이드, 약 38개 뉴클레오타이드, 약 39개 뉴클레오타이드, 약 40개 뉴클레오타이드, 약 41개 뉴클레오타이드, 약 42개 뉴클레오타이드, 약 43개 뉴클레오타이드, 약 44개 뉴클레오타이드, 또는 약 45개 뉴클레오타이드이다.
- [0292] 일부 실시양태에서, 포획 프로브는 25개 뉴클레오타이드, 26개 뉴클레오타이드, 27개 뉴클레오타이드, 28개 뉴클레오타이드, 29개 뉴클레오타이드, 30개 뉴클레오타이드, 31개 뉴클레오타이드, 32개 뉴클레오타이드, 33개 뉴클레오타이드, 34개 뉴클레오타이드, 35개 뉴클레오타이드, 36개 뉴클레오타이드, 37개 뉴클레오타이드, 38개 뉴클레오타이드, 39개 뉴클레오타이드, 40개 뉴클레오타이드, 41개 뉴클레오타이드, 42개 뉴클레오타이드, 43개 뉴클레오타이드, 44개 뉴클레오타이드, 또는 45개 뉴클레오타이드이다.
- [0293] 일부 실시양태에서, 포획 프로브는 약 100개 뉴클레오타이드, 약 200개 뉴클레오타이드, 약 300개 뉴클레오타이드, 약 400개 뉴클레오타이드, 또는 약 100개 뉴클레오타이드이다. 또 다른 실시양태에서, 포획 프로브는 약 100개 뉴클레오타이드 내지 약 500개 뉴클레오타이드, 약 200개 뉴클레오타이드 내지 약 500개 뉴클레오타이드, 약 300개 뉴클레오타이드 내지 약 500개 뉴클레오타이드, 또는 약 400개 뉴클레오타이드 내지 약 500개 뉴클레오타이드, 또는 그의 임의의 중간 범위이다.
- [0294] 일부 실시양태에서, 포획 프로브는 100개 뉴클레오타이드, 200개 뉴클레오타이드, 300개 뉴클레오타이드, 400개 뉴클레오타이드, 또는 100개 뉴클레오타이드이다. 또 다른 실시양태에서, 포획 프로브는 100개 뉴클레오타이드 내지 500

개 뉴클레오타이드, 200개 뉴클레오타이드 내지 500개 뉴클레오타이드, 300개 뉴클레오타이드 내지 500개 뉴클레오타이드, 또는 400개 뉴클레오타이드 내지 500개 뉴클레오타이드, 또는 그의 임의의 중간 범위이다.

[0295] 특정한 실시양태에서, 포획 프로브는 60개 뉴클레오타이드이다. 또 다른 실시양태에서, 포획 프로브는 실질적으로 60개 뉴클레오타이드보다 작지만, 동일한 DNA 표적 영역을 표적화하는 60개 뉴클레오타이드 포획 프로브와 비슷하게 뿐만 아니라 더 양호하게 혼성화한다. 일부 실시양태에서, 포획 프로브는 40개 뉴클레오타이드이다.

[0296] 일부 실시양태에서, 포획 프로브 모듈은 포획 프로브에 혼성화하는 태그 부착된 및 또는 증폭된 게놈 DNA 라이브러리 (예를 들어, 세포성 또는 cfDNA 라이브러리)의 하나 이상의 포획된 단편의 단리 및/또는 정제를 가능하게 하도록 결합 쌍의 특이적인 구성원을 포함한다. 일부 실시양태에서, 포획 프로브 모듈은 비오틴 또는 또 다른 적합한 합텐, 예를 들어, 디니트로페놀, 디곡시게닌에 접합된다.

[0297] 일부 실시양태에서, 포획 프로브 모듈은 태그 부착된 및 임의적으로 증폭된 DNA 라이브러리에 혼성화되어 복합체를 형성한다. 일부 실시양태에서, 다기능성 포획 프로브 모듈은 실질적으로 DNA 라이브러리에서 특이적인 게놈 표적 영역에 혼성화한다.

[0298] 혼성화 또는 혼성화 조건에는 2개의 뉴클레오타이드 서열이 안정한 복합체를 형성하는 임의의 반응 조건; 예를 들어 안정한 태그 부착된 DNA 라이브러리-포획 프로브 모듈 복합체를 형성하는 태그 부착된 DNA 라이브러리 및 포획 프로브 모듈이 포함될 수 있다. 이러한 반응 조건은 관련 기술분야에 널리 공지되어 있고, 관련 기술분야의 기술자는 이러한 조건이 적당히 변형될 수 있음을 이해할 것이며, 예를 들어 더 짧은 길이의 포획 프로브에 대한 어닐링 온도는 감소된다. 실질적인 혼성화는 포획 프로브 복합체의 제2 영역이 태그 부착된 DNA 라이브러리의 영역에 대해 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 85%, 80%, 75%, 또는 70% 서열 동일성, 상동성 또는 상보성을 나타낼 때 발생할 수 있다.

[0299] 일부 실시양태에서, 포획 프로브 (즉, 표적 서열에 혼성화하는 영역)는 약 40개 뉴클레오타이드이고, 약 44°C 내지 약 47°C의 최적의 어닐링 온도를 갖는다.

[0300] 일부 실시양태에서, 포획 프로브 (즉, 표적 서열에 혼성화하는 영역)는 40개 뉴클레오타이드이고, 44°C 내지 47°C의 최적의 어닐링 온도를 갖는다.

[0301] 일부 실시양태에서, 본원에서 고려되는 방법은 태그 부착된 cfDNA 라이브러리-포획 프로브 모듈 복합체를 단리하는 것을 포함한다. 일부 실시양태에서, DNA 복합체를 단리하는 방법은 관련 기술분야의 기술자에게 널리 공지되어 있고 (예를 들어, [Ausubel *et al.*, *Current Protocols in Molecular Biology*, 2007-2012] 참조), 관련 기술분야의 기술자에 의해 적당하게 고려되는 임의의 방법은 본 개시내용의 방법과 관련하여 사용될 수 있다. 일부 실시양태에서, 복합체는 비오틴-스트렙타비딘 단리 기술을 이용하여 단리된다.

[0302] 3. 표적화된 라이브러리의 증폭

[0303] 일부 실시양태에서, 단리된 포획 프로브/어댑터-태그 부착된 DNA 복합체로부터 단일 가닥 3'-말단의 제거가 고려된다. 일부 실시양태에서, 상기 방법은 단일 가닥 3' 말단을 제거하기 위해 단리된 태그 부착된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 3'-5' 엑소뉴클레아제 효소적 처리를 포함한다.

[0304] 일부 다른 실시양태에서, 상기 방법은 주형으로서 단리된 태그 부착된 DNA 라이브러리 단편을 사용하여 다기능성 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함한다. 이 연장 과정에 적합한 효소는 임의의 호열성 열안정성 DNA 폴리머라제일 수 있다. 상업적으로 입수가능한 DNA 폴리머라제의 예에는 하이 피델리티 Q5 DNA 폴리머라제 (NEB®), NEB넥스트 울트라 PCR®, NEB넥스트 울트라 II PCR® (NEB), 및 KAPA 2X® (로슈®)가 포함된다.

[0305] 일부 다른 실시양태에서, 상기 방법은 5' FLAP 엔도뉴클레아제, DNA 중합 및 DNA 리가제에 의한 Nick 패쇄의 공동 작용을 통해 혼성체 포획 프로브-단리된 태그 부착된 DNA 표적 분자, 예를 들어 태그 부착된 cfDNA 표적 분자 또는 태그 부착된 세포성 DNA 표적 분자를 생성하는 것을 포함한다.

[0306] 다양한 효소가 단리된 태그 부착된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 3'-5' 엑소뉴클레아제 효소적 처리를 위해 사용될 수 있다. 일부 실시양태에서 사용될 수 있는 3'-5' 엑소뉴클레아제 효소적 활성을 나타내는 적합한 효소의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: T4 또는 엑소뉴클레아제 I, III, V ([Shevelev IV, Huebscher U., *Nat Rev Mol Cell Biol.* 3(5):364-76 (2002)] 또한 참조). 일부 실시양태에서, 3'-5' 엑소뉴클레아제 활성을 포함하는 효소는 T4 폴리머라제이다. 일부 실시양태에서, 3'-5' 엑소뉴클레아제 효소적 활성을 나타내고 프라이머 주형 연장이 가능한 효소, 예를 들어 T4 또는 엑소뉴클레아제 I,

III, V가 사용될 수 있다. *Id.*

- [0307] 일부 실시양태에서, 본원에서 고려되는 방법은 상기에서 및 본원의 다른 곳에서 논의된 3'-5' 엑소뉴클레아제 효소적으로 처리된 복합체에 대해 시퀀싱 및/또는 PCR을 수행하는 것을 포함한다. 일부 실시양태에서, 포획 프로브 모듈의 꼬리 부분은 혼성체 핵산 분자를 생성하기 위해 카피된다. 일부 실시양태에서, 생성된 혼성체 핵산 분자는 포획 프로브 모듈에 혼성화할 수 있는 표적 영역 및 포획 프로브 모듈 꼬리 서열의 상보체를 포함한다.
- [0308] 일부 실시양태에서, 유전자 분석은 a) 1개 이상의 포획 프로브 모듈을 복수개의 게놈 DNA 라이브러리 클론의 하나 이상의 표적 유전자좌에 혼성화하여, 1개 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 형성하는 단계; b) a)로부터 1개 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 단리하는 단계; c) 단계 b)로부터의 하나 이상의 단리된 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 효소적으로 처리하는 단계; d) c)로부터의 효소적으로 처리된 복합체에 대해 PCR을 수행하는 단계이며, 여기서 포획 프로브 모듈의 꼬리 부분은 증폭된 혼성체 핵산 분자를 생성하기 위해 카피되고, 증폭된 혼성체 핵산 분자는 포획 프로브에 혼성화할 수 있는 표적 게놈 유전자좌에서 표적 서열 및 포획 프로브 모듈 꼬리 서열의 상보체를 포함하는 것인 단계; 및 e) d)로부터의 증폭된 혼성체 핵산 분자에 대해 정량적인 유전자 분석을 수행하는 단계를 포함한다.
- [0309] 일부 실시양태에서, a) 1개 이상의 포획 프로브 모듈을 복수개의 DNA 라이브러리 클론의 하나 이상의 표적 유전자좌에 혼성화시켜 1개 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 형성하는 단계; b) a)로부터 1개 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 단리하는 단계; c) 단계 b)로부터의 하나 이상의 단리된 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 효소적으로 처리하는 단계; d) c)로부터의 효소적으로 처리된 복합체에 대해 PCR을 수행하는 단계이며, 여기서 포획 프로브 모듈의 꼬리 부분은 증폭된 혼성체 핵산 분자를 생성하기 위해 카피되고, 증폭된 혼성체 핵산 분자는 포획 프로브에 혼성화할 수 있는 표적 유전자좌에서 표적 서열 및 포획 프로브 모듈 꼬리 서열의 상보체를 포함하는 것인 단계; e) d)의 증폭된 혼성체 핵산 분자의 PCR 증폭을 수행하는 단계; f) e)의 PCR 반응을 정량화하는 단계이며, 여기서 정량화는 특이적인 표적 영역의 카피 수의 결정을 가능하게 하는 것인 단계를 포함하는, 특이적인 표적 유전자좌의 카피 수를 결정하는 방법이 고려된다.
- [0310] 일부 실시양태에서, 단계 c)의 효소적 처리는 3'-5' 엑소뉴클레아제 활성을 갖는 효소를 사용하여 b)로부터의 1개 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체에 대해 3'-5' 엑소뉴클레아제 효소적 처리를 수행하여 단일 가닥 3' 말단을 제거하거나; 5' FLAP 엔도뉴클레아제, DNA 중합, 및 DNA 리가제에 의한 Nick 폐쇄의 공동 작용을 통해 하나 이상의 혼성체 포획 프로브 모듈-cfDNA 라이브러리 클론 분자를 생성하거나; 또는 주형으로서 복합체에서 단리된 DNA 클론을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함한다.
- [0311] 일부 실시양태에서, 단계 c)의 효소적 처리는 주형으로서 복합체에서 단리된 DNA 클론을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함한다.
- [0312] 일부 실시양태에서, PCR은 관련 기술분야의 기술자에게 널리 공지된 임의의 표준 PCR 반응 조건을 이용하여 수행될 수 있다. 일부 실시양태에서, e)의 PCR 반응은 2가지 PCR 프라이머를 사용한다. 일부 실시양태에서, e)의 PCR 반응은 표적 유전자좌 내의 반복부에 혼성화하는 제1 PCR 프라이머를 사용한다. 특정한 실시양태에서, e)의 PCR 반응은 표적 유전자좌/꼬리 접합부에서 혼성체 핵산 분자에 혼성화하는 제2 PCR 프라이머를 사용한다. 일부 실시양태에서, e)의 PCR 반응은 표적 유전자좌/꼬리 접합부에서 표적 유전자좌에 혼성화하는 제1 PCR 프라이머 및 증폭된 혼성체 핵산 분자에 혼성화하는 제2 PCR 프라이머를 사용한다. 일부 실시양태에서, 제2 프라이머는 표적 유전자좌/꼬리 접합부에 혼성화하여, 프라이머의 적어도 하나 이상의 뉴클레오티드가 표적 유전자좌에 혼성화하고, 프라이머의 적어도 하나 이상의 뉴클레오티드가 꼬리 서열에 혼성화하도록 한다.
- [0313] 일부 실시양태에서, 증폭은 예컨대 루프 매개된 등온 증폭 (LAMP), 전체 게놈 증폭 (WGA), 가닥 대체 증폭 (SDA), 헬리카제-의존성 증폭 (HDA), 레콤비나제 폴리머라제 증폭 (RPA), 핵산 시퀀싱 기반 증폭 (NASBA), Nick 형성 효소 증폭 반응 (NEAR), 및 리가제 연쇄 반응 (LCR)에 의해 등온일 수 있다.
- [0314] 일부 실시양태에서, 등온 증폭을 위한 DNA 폴리머라제에는 DNA 폴리머라제, 예컨대 클레노브 단편, Bsu 대형 단편, 및 중간 온도 반응 (25-40°C)을 위한 phi29 및 더 높은 온도 (50-65°C) 반응을 위한 Bst DNA 폴리머라제의 대형 단편이 포함된다. LCR에 적합한 효소에는 열안정성 Taq 리가제 및 열안정성 DNA 폴리머라제, 예컨대 Taq 폴리머라제가 포함된다.
- [0315] 일부 실시양태에서, 단계 e)로부터 수득된 증폭된 혼성체 핵산 분자는 시퀀싱되고, 서열은 수평으로 정렬되고,

즉, 서로에 대해 정렬되지만 기준 서열에 대해서는 정렬되지 않는다. 일부 실시양태에서, 단계 a) 내지 e)는 1개 이상의 포획 프로브 모듈로 1회 이상 반복된다. 포획 프로브 모듈은 동일하거나 또는 상이할 수 있고, 표적 유전자좌의 cfDNA 가닥을 표적화하도록 설계된다. 일부 실시양태에서, 포획 프로브가 상이할 때, 이들은 태그 부착된 cfDNA 클론 라이브러리에서 표적 유전자좌 내의 중첩되는 또는 인접한 표적 서열에서 혼성화한다. 일부 실시양태에서, 고밀도 포획 프로브 전략이 이용되며, 여기서 복수개의 포획 프로브는 표적 유전자좌에 혼성화하고, 각각의 복수개의 포획 프로브는 모든 중간 거리를 비롯하여 태그 부착된 DNA 클론 라이브러리에서 표적 유전자좌에 혼성화하는 임의의 다른 포획 프로브의 약 5, 약 10, 약 15, 약 20, 약 25, 약 30, 약 35, 약 40, 약 45, 약 50, 약 100, 약 200 bp 또는 그 초과 내에서 표적 유전자좌에 혼성화한다. 일부 실시양태에서, 고밀도 포획 프로브 전략이 이용되며, 여기서 복수개의 포획 프로브는 표적 유전자좌에 혼성화하고, 각각의 복수개의 포획 프로브는 모든 중간 거리를 비롯하여 태그 부착된 DNA 클론 라이브러리에서 표적 유전자좌에 혼성화하는 임의의 다른 포획 프로브의 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 bp 또는 그 초과 내에서 표적 유전자좌에 혼성화한다.

[0316] 일부 실시양태에서, 상기 방법은 표적 유전자좌당 2개의 포획 프로브 모듈을 사용하여 수행될 수 있으며, 하나는 표적 영역의 상류에서 "왓슨(Watson)" 가닥 (비코딩 또는 주형 가닥)에 혼성화하고, 하나는 표적 영역의 하류에서 "크릭(Crick)" 가닥 (코딩 또는 비주형 가닥)에 혼성화한다.

[0317] 일부 실시양태에서, 본원에서 고려되는 방법은 표적 유전자좌당 임의의 수의 포획 프로브 모듈, 예를 들어 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 또는 그 초과 포획 프로브 모듈로 여러 번 추가로 수행될 수 있으며, 임의의 수는 임의의 조합으로 왓슨 또는 크릭 가닥에 혼성화한다. 일부 실시양태에서, 수득된 서열은 임의의 수많은 차이를 식별하기 위해 서로 정렬될 수 있다.

[0318] 일부 실시양태에서, 1개 이상의 포획 프로브 모듈을 사용하여 단일 반응에서 복수개의 표적 유전자좌, 예를 들어 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 10000, 50000, 100000, 500000개 또는 그 초과를 조사한다.

[0319] 일부 실시양태에서, 효소적 처리 단계 (상기 개시내용에서 단계 c)는 수행되지 않는다. 일부 실시양태에서, 단리된 포획 프로브/태그 부착된 DNA 단편 복합체는 직접적으로 증폭되며, DNA 폴리머라제는 5' → 3' 연장을 수행하여 혼성체 분자의 라이브러리를 형성한다. 일부 실시양태에서, 혼성체 분자를 함유하는 라이브러리는 시퀀싱 어댑터 (예를 들어 어댑터는 시퀀싱 프라이머, 예컨대 일루미나® 넥스트세크 NGS 기술의 P5 및 P7 시퀀싱 프라이머에 결합함)를 함유하는 정방향 및 역방향 프라이머를 사용하여 추가로 증폭되어, 시퀀싱 준비된 증폭된 혼성체 분자의 표적화된 라이브러리를 생성한다.

[0320] 따라서, 효소적 처리 단계를 제거함으로써, 표적화된 라이브러리는 개시된 방법을 이용하여 더 빠르게 생성된다. 이러한 더 빠른 방법은 태그 부착된 DNA 라이브러리 및 표적화된 라이브러리에서 DNA 단편에 존재하는 관심 유전자좌의 유전자 분석의 성능을 개선되고 더 빠르게 만든다. 관련 기술분야의 기술자는 유전자좌가 DNA 변경, 예를 들어 SNV, Indel, 유전자 재조직, 및 카피 수 변화를 위해 분석될 수 있음을 인식할 것이다.

[0321] 4. 계놈 등가물 수의 결정

[0322] 일부 실시양태에서, DNA의 유전자 분석 방법은 DNA 클론 라이브러리에서 계놈 등가물의 수를 결정하는 것을 포함한다. 본원에 사용된 바와 같이, 용어 "계놈 등가물"은 각각의 라이브러리에서 계놈 카피의 수를 지칭한다. 본원에서 고려되는 조성물 및 방법에 의해 충족하는 중요한 과제는 유전자 서열에서 드문 유전자 돌연변이 또는 차이를 검출하고 분석하기 위해 충분한 검정 민감도를 달성하는 것이다. 샘플별 기반으로 검정 민감도 값을 결정하기 위해, 시퀀싱 라이브러리에 존재하는 계놈 등가물의 수를 측정함으로써 각각의 샘플에 존재하는 상이한 별개의 서열의 수를 측정한다. 민감도를 확립하기 위해, 계놈 등가물의 수는 각각의 샘플 라이브러리에 대해 측정되어야 한다.

[0323] 계놈 등가물의 수는 qPCR 검정에 의해 또는 시퀀싱을 수행한 후 생물정보학-기반 카운팅을 이용하여 결정될 수 있다. 임상 샘플의 과정 흐름에서, 계놈 등가물의 qPCR 측정은 DNA 라이브러리를 위한 QC 단계로 이용된다. 이는 서열 분석 전에 검정 민감도에 대한 기대치를 확립하고, 샘플이 그의 상응하는 DNA 클론 라이브러리에서 계놈 등가물의 필요한 깊이가 부족한 경우 분석으로부터 제외되도록 한다. 궁극적으로, 계놈 등가물의 생물정보학-기반 카운팅을 또한 이용하여, 각각의 주어진 DNA 클론 라이브러리에 대해 계놈 등가물 (및 따라서 검정 민감도 및 위음성 추정치)을 식별한다.

[0324] 경험적 qPCR 검정 및 통계적 카운팅 검정은 상관관계가 잘 있어야 한다. 시퀀싱이 DNA 클론 라이브러리에서 서

열 깊이를 밝히지 못하는 경우, DNA 클론 라이브러리의 재처리 및/또는 추가의 시퀀싱이 필요할 수 있다.

[0325] 일부 실시양태에서, 세포성 DNA 또는 cfDNA 클론 라이브러리에서 게놈 등가물은 정량적인 PCR (qPCR) 검정을 이용하여 결정된다. 일부 실시양태에서, 공지된 농도의 표준 라이브러리를 사용하여 표준 곡선을 구축하고, qPCR 검정으로부터의 측정치를 생성된 표준 곡선에 대입하고, 게놈 등가물에 대한 값은 상기 대입으로부터 유래된다. 본 발명자들은 게놈에서 공통 서열, 예를 들어 반복부 서열에 특이적으로 혼성화하는 1개의 프라이머 및 어댑터에서 프라이머 결합 부위에 결합하는 또 다른 프라이머를 포함하는 qPCR "반복부-기반" 검정이 어댑터 특이적인 프라이머 (DNA 클론의 양 말단에 존재함)만을 사용하는 방법과 비교하여 게놈 등가물에서 8배 증가를 측정한 것을 발견하였다. 반복부-기반 검정에 의해 측정된 게놈 등가물의 수는 더 일관된 라이브러리별 성능, 및 시퀀싱 실행에서 게놈 등가물과 생물정보학적으로 카운팅된 태그 등가물의 qPCR 추정치 사이에 더 양호한 정렬을 제공한다.

[0326] 본원에서 고려되는 반복부-기반 게놈 등가물 검정에서 사용하기 위한 적합한 반복부의 예시적인 예에는 짧은 산재된 핵 요소 (SINE), 예를 들어 Alu 반복부; 긴 산재된 핵 요소 (LINEs), 예를 들어 LINE1, LINE2, LINE3; 미세부수체 반복부 요소, 예를 들어 짧은 직렬 반복부 (STR), 단순 서열 반복부 (SSR); 및 포유동물-수준 산재된 반복부 (MIR)가 포함되나 이로 제한되지 않는다.

[0327] 일부 실시양태에서, 반복부는 Alu 반복부이다.

[0328] 5. 시퀀싱

[0329] 일부 실시양태에서, 정량적인 유전자 분석은 상기에서 본원의 다른 곳에서 논의된 시퀀싱 복수개의 혼성체 핵산 분자를 포함하여, 충분한 시퀀싱 깊이를 생성하여 복수개의 고유의 시퀀싱 리드를 수득한다. 용어 "고유의 리드" 또는 "고유의 게놈 서열" (UGS)은 본원에서 상호교환적으로 사용되며, 개별 중복 리드를 함께 "패밀리"로 그룹화하여 식별된다. 중복 리드는 동일한 UMIE를 공유하고 (예를 들어, 게놈 서열 내에서 동일한 리드 코드 및 동일한 DNA 서열 출발 위치를 공유하고), 단일 부착 사건으로부터 유래되어, 서로의 증폭-유래된 "동기간"인 서열 리드이다. 중복 리드의 패밀리를 대표하는 단일 컨센서스는 고유의 리드 또는 UGS로서 정방향으로 운반된다. 각각의 고유의 리드 또는 UGS는 고유의 부착 사건으로 간주된다. 특정한 포획 프로브에 상응하는 고유의 리드의 합계는 해당 특정한 포획 프로브에 대한 "원시 게놈 깊이" (RGD)로 지칭된다. 각각의 포획 프로브는 패밀리로 그룹화함으로써 총 리드로부터 계산적으로 증류된 고유의 리드의 세트를 생성한다. 일부 실시양태에서, 혼성체 분자에서 전체 포획 프로브 영역을 시퀀싱한다. 일부 실시양태에서, 혼성체 분자에서 포획 프로브 영역의 일부를 시퀀싱한다.

[0330] 이어서, 주어진 샘플에 대한 고유의 리드 (예를 들어, 샘플에 대한 원시 게놈 깊이)는 프로브별 기반으로 관찰된 모든 고유의 리드의 평균으로서 계산된다. 각각의 고유의 리드가 고유의 게놈 DNA 클론으로부터 유래되어야 하기 때문에, 고유의 리드는 중요하다. 각각의 고유의 리드는 게놈 DNA의 반수체 등가물의 투입 및 분석을 나타낸다. 고유의 리드의 합계는 분석된 반수체 게놈의 합계이다. 이어서, 분석된 게놈의 수는 시퀀싱 검정의 민감도를 정의한다. 비제한적인 예로서, 평균 고유의 리드 카운트가 100개의 게놈 등가물인 경우, 해당 특정한 검정은 100개 중 1개의 돌연변이체 리드, 또는 1%를 검출할 수 있는 민감도를 갖는다. 이보다 작은 임의의 관찰은 옹호될 수 없다.

[0331] 명백한 카피 수 변화가 있는 경우 (예를 들어, 노이즈 프로브의 예)는 샘플 평균을 계산하기 위해 사용되는 데이터 세트로부터 제외된다. 여기서, "노이즈 프로브"는 큰 세트의 동일한 샘플 중에서 매우 가변적인 수의 고유의 리드 (예를 들어, 12 - 16개 샘플 복제물 중에서 매우 가변적인 수의 고유의 리드)를 포획하는 프로브를 지칭한다. 일부 실시양태에서, 노이즈 프로브와 연관된 고유의 리드의 수는 샘플에 대한 고유의 리드의 평균 수와 비교하여 50% 이상 증가된다. 일부 실시양태에서, 노이즈 프로브와 연관된 고유의 리드의 수는 샘플에 대한 고유의 리드의 평균 수와 비교하여 50% 이상 감소된다. 일부 실시양태에서, 특정한 분석에서 사용되는 프로브의 약 2% 내지 약 4%는 노이즈 프로브로서 식별되고, 주어진 샘플에 대한 고유의 리드의 평균 수를 결정하기 위해 계산으로부터 제외된다. 일부 실시양태에서, 특정한 분석에서 사용되는 프로브의 2% 내지 4%는 노이즈 프로브로서 식별되고, 주어진 샘플에 대한 고유의 리드의 평균 수를 결정하기 위해 계산으로부터 제외된다.

[0332] 일부 실시양태에서, 시퀀싱 리드는 "온-타겟 리드" 또는 "오프-타겟 리드"로서 식별된다. 온-타겟 리드는 게놈 라이브러리를 생성하기 위해 사용되는 포획 프로브 부근에서 맵핑되는 게놈 DNA 서열을 갖는다. 일부 실시양태에서, 각각의 게놈 서열이 특이적인 포획 프로브와 물리적으로 연결되고, 게놈 세그먼트 및 포획 프로브의 서열 둘 다 통합된 정보 조각으로서 결정되는 경우, 온-타겟 리드는 출발 좌표가 상응하는 포획 프로브의 3' 말단의

400 bp 내에, 더욱 일반적으로 200 bp 내에 맵핑되는 임의의 게놈 서열로 정의된다. 오프-타겟 리드는 포획 프로브에 비해 ≥ 500 개 염기 쌍 위치에서 기준 게놈에 대해 정렬되는 게놈 서열을 갖는 (그리고 더욱 자주 전체적으로 상이한 염색체에 맵핑되는) 것으로 정의된다.

[0333] 일부 실시양태에서, 정량적인 유전자 분석은 복수개의 샘플로부터 유래된 혼성체 핵산 분자의 멀티플렉스 시퀀싱을 포함한다.

[0334] 일부 실시양태에서, 정량적인 유전자 분석은 하나 이상의 또는 복수개의 태그 부착된 DNA 라이브러리 클론을 수득하고, 여기서 각각의 클론은 제1 DNA 서열 및 제2 DNA 서열을 포함하고, 제1 DNA 서열은 표적화된 유전자좌의 서열을 포함하고, 제2 DNA 서열은 포획 프로브 서열을 포함하고; 하나 이상의 클론에 대해 쌍 형성 말단 시퀀싱 반응을 수행하고 하나 이상의 시퀀싱 리드를 수득하거나, 또는 약 100, 약 200, 약 300, 약 400, 약 500개 초과 또는 그 초과 뉴클레오타이드의 단일의 긴 시퀀싱 리드가 수득되는 하나 이상의 클론에 대해 시퀀싱 반응을 수행하고, 여기서 리드는 제1 DNA 서열 및 제2 DNA 서열 둘 다를 식별하는데 충분하고; 시퀀싱 리드의 프로브 서열에 따라 하나 이상의 클론의 시퀀싱 리드를 배치하거나 또는 클러스터화하는 것을 포함한다.

[0335] 6. 생물정보학적 분석

[0336] 일부 실시양태에서, 정량적인 유전자 분석은 시퀀싱 리드의 생물정보학적 분석을 추가로 포함한다. 생물정보학적 분석은 시퀀싱을 위한 조성물 또는 방법의 부재하에 수행되는 임의의 순전히 정신적인 분석을 제외한다. 일부 실시양태에서, 생물정보학적 분석에는 서열 정렬; 게놈 등가물 분석; 단일 뉴클레오타이드 변이체 (SNV) 분석; 유전자 카피 수 변동 (CNV) 분석; 염색체 카피 수의 측정; 및 유전적 병변의 검출이 포함되나 이로 제한되지 않는다. 일부 실시양태에서, 생물정보학적 분석은 cfDNA 클론 라이브러리에서 분석되는 게놈 등가물의 수를 정량화하고; 표적 유전자좌의 유전적 상태를 검출하고; 표적 유전자좌에서 유전적 병변을 검출하고; 표적 유전자좌 내에서 카피 수 변동을 측정하는데 유용하다.

[0337] 서열 정렬은 서열 리드와 하나 이상의 인간 기준 DNA 서열 사이에서 수행될 수 있다. 일부 실시양태에서, 시퀀싱 정렬을 이용하여, 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재배열, 카피 수의 변화, 또는 유전자 융합의 검출을 비롯하여 이로 제한되지 않는 표적 유전자좌에서 유전적 병변을 검출할 수 있다. 원인 또는 예후 지표인 유전적 병변의 검출은 특정한 유전적 병태 또는 질환의 진단, 예후, 치료 및/또는 모니터링에 유용할 수 있다.

[0338] 용어 "표적 유전자좌" 및 "DNA 표적 영역"은 본원에서 상호교환적으로 사용되고, DNA 서열 내의 관심 영역을 지칭한다. 일부 실시양태에서, 표적화된 유전자 분석은 표적 유전자좌에 대해 수행된다. 일부 실시양태에서, DNA 표적 영역은 특정한 유전적 상태, 유전적 병태, 유전 질환; 태아 검사; 유전적 모자이크 현상, 친자 확인 검사; 약물 치료에 대한 반응의 예측; 의학적 병태의 진단 또는 모니터링; 미생물군 프로파일링; 병원균 스크리닝; 또는 장기 이식 모니터링과 연관된 유전자의 영역이다. 추가의 실시양태에서, DNA 표적 영역은 특정한 인간 염색체, 예컨대 특정한 상염색체 또는 X-연관 염색체, 또는 그의 영역 (예를 들어, 고유의 염색체 영역)과 연관된 DNA 서열이다.

[0339] 수평적 서열 분석으로 지칭되는, 기준 서열에 대해 정렬할 필요없이 수행될 수 있는 서열 정렬 분석 방법 또한 본원에서 고려된다. 이러한 분석은 본원에서 고려되는 방법 또는 임의의 다른 방법에 의해 생성된 임의의 서열에 대해 수행될 수 있다. 일부 실시양태에서, 서열 분석은 본원에서 고려되는 방법에 의해 수득되는 리드에 대해 서열 정렬을 수행하는 것을 포함한다.

[0340] 일부 실시양태에서, cfDNA 클론 라이브러리에서 게놈 등가물은 시퀀싱이 수행된 후 생물정보학-기반 카운팅을 이용하여 결정된다. 각각의 시퀀싱 리드는 특정한 포획 프로브와 회합되고, 각각의 포획 프로브에 배정된 리드의 집합은 그룹으로 분석된다. 그룹 내에서, 개별 리드의 세트는 게놈 서열 내에서 동일한 리드 코드 및 동일한 DNA 서열 출발 위치를 공유한다. 이들 개별 리드는 "패밀리"로 그룹화되고, 이 패밀리를 대표하는 단일 컨센서스는 "고유의 리드"로서 정방향으로 운반된다. 패밀리를 구성하는 모든 개별 리드는 단일 부착 사건으로부터 유래되며, 따라서 이들은 서로의 증폭-유래된 "동기간"이다. 각각의 고유의 리드는 고유의 부착 사건으로 고려되고, 고유의 리드의 합계는 분석된 게놈 등가물의 수와 등가물로 고려된다.

[0341] 고유의 클론의 수가 가능한 서열 조합의 총 수에 접근함에 따라, 확률은 동일한 코드 및 출발 부위 조합이 독립적인 사건에 의해 생성되고, 이들 독립적인 사건이 단일 패밀리 내에서 부적당하게 그룹화될 것임을 나타낸다. 최종 결과는 분석된 게놈 등가물의 과소평가일 것이며, 희귀 돌연변이체 리드는 동일한 식별자를 공유하는 야생형 리드와 중첩되기 때문에 시퀀싱 오류로서 폐기될 수 있다.

- [0342] 일부 실시양태에서, cfDNA 클론 라이브러리에 대한 정확한 분석을 제공하기 위해, 분석된 게놈 등가물의 수는 가능한 고유의 클론의 수의 약 1/10, 약 1/12, 약 1/14, 약 1/16, 약 1/18, 약 1/20, 약 1/25 또는 그 미만이다. 일부 실시양태에서, cfDNA 클론 라이브러리에 대한 정확한 분석을 제공하기 위해, 분석된 게놈 등가물의 수는 가능한 고유의 클론의 수의 1/10, 1/12, 1/14, 1/16, 1/18, 1/20, 1/25 또는 그 미만이다. 상기 나타난 절차가 단지 예시적이고 비제한적임을 이해해야 한다.
- [0343] 일부 실시양태에서, 분석될 게놈 등가물의 수는 증가될 필요가 있을 수 있다. 게놈 등가물의 깊이를 확장시키기 위해, 적어도 2가지 해결책이 고려된다. 첫번째 해결책은 샘플당 1개 초과와 어댑터 세트를 사용하는 것이다. 어댑터를 조합함으로써, 가능한 클론의 총 수를 배수적으로 확장시킬 수 있고, 따라서 게놈 투입의 편안한 한계를 확장시킬 수 있다. 두번째 해결책은 리드 코드를 1, 2, 3, 4, 또는 5개 또는 그 초과의 염기만큼 확장시키는 것이다. 다른 모든 리드 코드와 적어도 2개 염기만큼 상이한 가능한 리드 코드의 수는 $4^{(n-1)}$ 로 규모 조정되고, 여기서 n은 리드 코드 내의 염기의 수이다. 따라서, 비제한적인 예에서, 리드 코드가 5개 뉴클레오타이드이고 $4^{(5-1)} = 256$ 인 경우; 따라서 추가의 염기의 포함은 각각의 추가의 염기에 대해 이용가능한 레퍼토리를 4배 확장시킨다.
- [0344] 일부 실시양태에서, 정량적인 유전자 분석은 회귀 단일 뉴클레오타이드 변이체 (SNV)를 식별하기 위해 시퀀싱 리드의 생물정보학적 분석을 포함한다.
- [0345] 차세대 시퀀싱은 대략 0.02-0.02%의 고유 오류율을 가지며, 이는 어디든 1/200 내지 1/500 염기 호출이 부정확함을 의미한다. 이보다 낮은 빈도로, 예를 들어 1000개 서열당 1개의 빈도로 발생하는 변이체 및 다른 돌연변이를 검출하기 위해, 분자 주해 전략을 적용할 필요가 있다. 비제한적인 예로서, 표적화된 서열 포획 기술을 이용하는 5000개의 고유의 분자의 분석은 (>50,000 리드의 충분한 시퀀싱 깊이에서) 5000개의 고유의 리드의 집합을 생성하며, 각각의 고유의 리드는 모두 동일한 리드 코드를 갖는 리드의 "패밀리"에 속한다. 패밀리 내에서 발생하는 SNV는 회귀 변이체에 대한 후보이다. 이 동일한 변이체가 1개 초과와 패밀리에서 관찰되는 경우, 이는 출발 샘플 내에 존재하는 회귀 변이체에 대한 매우 강력한 후보가 된다. 대조적으로, 패밀리 내에서 산발적으로 발생하는 변이체는 시퀀싱 오류일 가능성이 높고, 1개 및 단지 1개의 패밀리 내에서 발생하는 변이체는 회귀하거나 또는 생체 외에서 발생하는 염기 변경의 결과 (예를 들어, DNA 염기의 산화 또는 PCR-도입된 오류)이다.
- [0346] 일부 실시양태에서, SNV를 검출하는 방법은 검정의 원하는 표적 민감도로서 10배 더 많은 게놈 투입 (게놈 또는 게놈 등가물)을 도입하는 것을 포함한다. 한 비제한적인 예에서, 원하는 민감도가 2% (2/100)인 경우, 실험 표적은 2000개 게놈의 투입이다.
- [0347] 일부 실시양태에서, 시퀀싱 데이터의 생물정보학적 분석을 이용하여, 유전적 상태, 병태 또는 질환, 유전적 모자이크 현상, 태아 검사, 친자 확인 검사, 약물 치료에 대한 반응의 예측, 의학적 병태의 진단 또는 모니터링, 미생물군 프로파일링, 병원균 스크리닝, 및 장기 이식 모니터링과 연관된 SNV를 검출하거나 또는 식별한다.
- [0348] **7. 카피 수 분석**
- [0349] 세포성 게놈 DNA (예를 들어 조직 생검 샘플로부터) 또는 cfDNA (예를 들어 혈액 샘플로부터)의 샘플 내에서 돌연변이 변화, SNP, 전좌, 역위, 결실, 카피 수의 변화 또는 다른 유전자 변동의 검출에 유용한 조성물 및 방법이 본원에 제공된다. 본원에 개시된 조성물 및 방법은 정교한 해상도로 생물학적 샘플 (예를 들어 혈액)로부터의 cfDNA에서 카피 수 변동을 검출하기 매우 어려운 검출에 특히 유용하다. 특히, 본 개시내용의 일부 실시양태는 어댑터에 부착된 게놈 DNA 단편으로 구성된 게놈 DNA 라이브러리를 생성하고, 복수개의 포획 프로브에 의해 DNA 표적 영역을 포획하고, DNA 표적 영역을 포함하는 DNA 라이브러리 단편을 단리하고, DNA 표적 영역의 정량적인 유전자 분석을 수행하여 DNA 표적 영역의 카피 수를 결정함으로써, 시험 샘플로부터 DNA 표적 영역의 카피 수를 검출하는 방법에 관한 것이다. 본원에 기재된 어댑터는 시퀀싱될 개별 DNA 단편의 식별, 뿐만 아니라 게놈 DNA의 샘플 또는 공급원의 식별을 가능하게 한다.
- [0350] 일부 실시양태에서, 본원에 개시된 표적-특이적인 카피 수 변화의 검출을 위한 조성물 및 방법은 직접적인 조직 생검 및 말초 혈액을 비롯하여 이로 제한되지 않는 몇몇 샘플 유형에 적용가능하다. 암 게놈학, 및 고형 종양의 분석을 위한 특정한 무세포 DNA (cfDNA) 검정의 맥락에서, 종양 DNA의 양은 종종 전체 DNA의 매우 작은 분획이다. 추가로, 카피 수 손실은 게놈 DNA 검정, 특히 카피 수 변화가 샘플로부터의 총 게놈 DNA의 일부에만 존재할 수 있는 게놈 DNA 검정, 예를 들어 cfDNA 검정에서는 검출하기가 어렵다. 예를 들어, 암 환자로부터 추출된 대부분의 무세포 DNA는 정상 공급원으로부터 유래될 것이고, 이때 카피 수 (남성 대상체에서 X-연관 유전

자의 경우는 제외)를 갖는다. 암 환자에서는, 예를 들어 혈장으로부터 추출된 순환 DNA의 2%가 종양으로부터 유래된 환자에서와 같이, 종양으로부터 유래된 DNA 분획이 종종 낮은 마이너 대립유전자 빈도를 갖는다. 종양 억제제 유전자 (예를 들어, 유방암에서는 BRCA1)의 1개 카피의 손실은 검출가능한 게놈 단편의 부재하에 마이너 대립유전자 빈도가 1%임을 의미한다. 이 시나리오에서, 조작된 카피 수 손실 검정은 100개 카피 (정상)와 99개 카피 (이형접합성 유전자 손실) 사이를 구별할 수 있어야 한다. 따라서, 일부 실시양태는 본원에 기재된 방법 및 조성물이 심지어 cfDNA의 맥락에서도 마이너 대립유전자 빈도에서 카피 수의 변화를 검출하기에 충분한 해상도로 카피 수 변화의 검출을 가능하게 함을 고려한다.

[0351] 일부 실시양태에서, DNA 표적 영역 DNA의 카피 수 분석을 위한 방법이 제공된다. 일부 실시양태에서, 카피 수 분석은 각각 게놈 DNA 단편 및 어댑터를 함유하는 DNA 라이브러리 단편의 게놈 DNA 라이브러리를 생성하고, DNA 표적 영역을 함유하는 DNA 라이브러리 단편을 단리하고, DNA 표적 영역의 정량적인 유전자 분석을 수행함으로써 수행된다. "정량적인 유전자 분석"이란 DNA 돌연변이, SNP, 전좌, 결실, 및 카피 수 변동 (CNV)을 비롯하여 이로 제한되지 않는 DNA (예를 들어, 유전자, 유전자좌, 표적 관심 영역 등)에서 변화를 정량화할 수 있는 임의의 분자 생물학적 기술에 의해 수행되는 분석을 의미한다. 일부 실시양태에서, 정량적인 유전자 분석은 시퀀싱, 예를 들어 차세대 시퀀싱에 의해 수행된다.

[0352] 일부 실시양태에서, 하나 이상의 또는 복수개의 클론을 수득하는 것을 포함하는, 카피 수 결정 분석을 위한 방법이 제공되며, 각각의 클론은 제1 DNA 서열 및 제2 DNA 서열을 포함하고, 제1 DNA 서열은 표적화된 유전자좌의 서열을 포함하고, 제2 DNA 서열은 포획 프로브 서열을 포함한다. 관련 실시양태에서, 하나 이상의 클론에 대한 쌍 형성 말단 시퀀싱 반응을 수행하고, 하나 이상의 시퀀싱 리드가 수득된다. 또 다른 실시양태에서, 하나 이상의 클론에 대한 시퀀싱 반응을 수행하고, 여기서 약 100개 초과 뉴클레오타이드의 단일의 긴 시퀀싱 리드가 수득되고, 리드는 제1 DNA 서열 및 제2 DNA 서열 둘 다를 식별하는데 충분하다. 하나 이상의 클론의 시퀀싱 리드는 시퀀싱 리드의 프로브 서열에 따라 배치되거나 또는 클러스터화될 수 있다.

[0353] 카피 수 분석에는 주어진 게놈 DNA 샘플에서 발생하는 특정한 유전자 또는 돌연변이의 카피 수를 검사하는 분석이 포함되나 이로 제한되지 않으며, 주어진 샘플에서 주어진 유전자 또는 서열 차이의 카피 수의 정량적인 결정이 추가로 포함될 수 있다. 일부 실시양태에서, 카피 수 분석을 이용하여, 유전적 상태, 병태 또는 질환, 태아 검사, 유전적 모자이크 현상, 친자 확인 검사, 약물 치료에 대한 반응의 예측, 의학적 병태의 진단 또는 모니터링, 미생물군 프로파일링, 병원균 스크리닝, 및 장기 이식 모니터링과 연관된 유전자 증폭을 검출하거나 또는 확인한다.

[0354] 일부 실시양태에서, 카피 수 분석을 이용하여 염색체 불안정성을 측정한다. 이러한 실시양태에서, 염색체 안정성 프로브를 포함하는 포획 프로브 세트를 사용하여, 모든 염색체 세트에 걸쳐 균일한 밀도로 카피 수 변동을 결정한다. 카피 수 분석이 각각의 염색체 안정성 프로브에 대해 수행된 다음, 염색체 안정성 프로브가 그들의 염색체 표적에 따라 배치된다. 이는 게놈에 걸쳐 카피 수 손실 또는 획득의 시각화를 가능하게 하고, 염색체 안정성의 척도의 역할을 할 수 있다.

[0355] 일부 실시양태에서, 시퀀싱 데이터의 생물정보학적 분석을 이용하여, 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재배열, 카피 수의 변화, 또는 유전자 융합의 검출을 비롯하여 이로 제한되지 않는 표적 유전자좌에서 하나 이상의 서열 또는 유전적 병변을 검출하거나 또는 확인한다. 원인 또는 예후 지표인 유전적 병변의 검출은 특정한 유전적 병태 또는 질환의 진단, 예후, 치료 및/또는 모니터링에 유용할 수 있다. 일부 실시양태에서, 유전적 병변은 유전적 상태, 병태 또는 질환, 태아 검사, 유전적 모자이크 현상, 친자 확인 검사, 약물 치료에 대한 반응의 예측, 의학적 병태의 진단 또는 모니터링, 미생물군 프로파일링, 병원균 스크리닝, 및 장기 이식 모니터링과 연관이 있다.

[0356] 일부 실시양태에서, 샘플에 존재하는 DNA 표적 영역의 카피 수는 정량적인 유전자 분석에 의해 결정된다. 일부 실시양태에서, DNA 표적 영역의 카피 수는 샘플에 존재하는 DNA 표적 영역의 카피의 양을 비교하고, 이를 공지된 카피 수를 갖는 하나 이상의 샘플에 존재하는 DNA 표적 영역의 양과 비교함으로써 결정된다.

[0357] 일부 실시양태는 본원에 기재된 조성물 및 방법이 게놈 DNA의 샘플에서 카피 수의 변화를 검출하는데 특히 유용하고, 샘플에서 총 게놈 DNA의 일부만이 카피 수의 변화를 갖는다는 것을 고려한다. 예를 들어, 유의한 종양 돌연변이는 대립유전자 빈도가 일반적으로 ~100%, 50% 또는 0%인 통상적인 SNP 유전자형 분석과는 대조적으로, 유의하게 50% 미만 (예를 들어, 0.1% 내지 >20%의 범위)인 마이너 대립유전자 빈도로 존재하는 샘플, 예를 들어 무세포 DNA의 샘플에 존재할 수 있다. 관련 기술분야의 기술자는 본원에 기재된 조성물 및 방법이 단일 뉴클레오타이드 변이체 (SNV), 짧은 (예를 들어, 40개 미만의 염기 쌍 (bp)) 삽입, 및 결실 (indel), 및 게놈 재배열,

예컨대 종양원성 유전자 융합을 비롯하여 다른 유형의 돌연변이를 검출하는데 또한 유용하다는 것을 인식할 것이다.

[0358] 일부 실시양태에서, 본원에 기재된 조성물 및/또는 방법은 샘플로부터의 총 게놈 DNA의 약 20% 미만, 약 19% 미만, 약 18% 미만, 약 17% 미만, 약 16% 미만, 약 15% 미만, 약 14% 미만, 약 13% 미만, 약 12% 미만, 약 11% 미만, 약 10% 미만, 약 9% 미만, 약 8% 미만, 약 7% 미만, 약 6% 미만, 약 5% 미만, 약 4% 미만, 약 3% 미만, 약 2% 미만, 약 1% 미만, 약 0.5% 미만, 약 0.2% 미만, 또는 약 0.1% 미만으로 존재하는 하나 이상의 DNA 표적 영역의 카피 수의 변화를 검출하고, 확인하고, 관찰하고/거나 밝히는데 유용하고, 그것이 가능하고, 그에 적합하고/거나 그를 할 수 있다. 일부 실시양태에서, 본원에 기재된 방법은 샘플로부터의 총 게놈 DNA의 약 0.01% 내지 약 100%, 약 0.01% 내지 약 50%, 또는 약 0.1% 내지 약 20%로 존재하는 하나 이상의 DNA 표적 영역의 카피 수의 변화를 검출하고, 확인하고, 관찰하고/거나 밝히는데 유용하고, 그것이 가능하고, 그에 적합하고/거나 그를 할 수 있다.

[0359] 일부 실시양태에서, 본원에 기재된 조성물 및/또는 방법은 샘플로부터의 총 게놈 DNA의 20% 미만, 19% 미만, 18% 미만, 17% 미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 10% 미만, 9% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만, 1% 미만, 0.5% 미만, 0.2% 미만, 또는 0.1% 미만으로 존재하는 하나 이상의 DNA 표적 영역의 카피 수의 변화를 검출하고, 확인하고, 관찰하고/거나 밝히는데 유용하고, 그것이 가능하고, 그에 적합하고/거나 그를 할 수 있다. 일부 실시양태에서, 샘플로부터의 총 게놈 DNA의 0.01% 내지 100%, 0.01% 내지 50%, 또는 0.1% 내지 20%로 존재하는 하나 이상의 DNA 표적 영역의 카피 수의 변화를 본원에 기재된 방법은 검출하고, 확인하고, 관찰하고/거나 밝히는데 유용하고, 그것이 가능하고, 그에 적합하고/거나 그를 할 수 있다.

[0360] 일부 실시양태에서, cfDNA의 유전자 분석을 위한 방법은 cfDNA 라이브러리를 생성하고 증폭시키고, cfDNA 라이브러리에서 게놈 등가물의 수를 결정하고; 하나 이상의 게놈 표적 유전자좌의 정량적인 유전자 분석을 수행하는 것을 포함한다.

[0361] 일부 실시양태는 본원에 기재된 임의의 방법 및 조성물이 게놈 DNA, 예를 들어 세포성 또는 cfDNA를 사용하여 유전적 상태, 유전적 병태, 유전 질환, 유전적 모자이크 현상, 태아 진단, 친자 확인 검사, 미생물군 프로파일링, 병원균 스크리닝, 및 장기 이식 모니터링을 효율적으로 분석하고, 검출하고, 진단하고/거나 모니터링하는데 사용하기에 효과적인 것으로 고려되며, 여기서 샘플에서 총 게놈 DNA 전체 또는 일부만이 관심 특징, 예를 들어 유전적 병변, 돌연변이, 단일 뉴클레오타이드 변이체 (SNV)를 갖는다. 일부 실시양태에서, 관심 특징은 질환 또는 병태와 연관된 유전적 특징이다. 예를 들어, 유의한 중앙 돌연변이는 대립유전자 빈도가 일반적으로 ~100%, 50% 또는 0%인 통상적인 SNP 유전자형 분석과는 대조적으로, 유의하게 50% 미만 (예를 들어 0.1% 내지 >20%의 범위)인 마이너 대립유전자 빈도로 존재하는 샘플, 예를 들어 cfDNA의 샘플에 존재할 수 있다.

[0362] 8. 임상적 적용

[0363] 일부 실시양태에서, 관심 영역에서 돌연변이 변화, SNP, 전좌, 역위, 결실, 카피 수의 변화 또는 다른 유전자 변동을 검출함으로써 대상체에서 병태 또는 질환을 검출하거나, 확인하거나, 예측하거나, 진단하거나 또는 모니터링하는 방법이 본원에 제공된다.

[0364] 일부 실시양태에서, 대상체에서 병태 또는 질환을 검출하거나, 확인하거나, 예측하거나, 진단하거나 또는 모니터링하는 방법이 본원에 제공된다.

[0365] 일부 실시양태에서, 대상체에서 유전적 상태, 병태 또는 질환을 검출하거나, 확인하거나, 예측하거나, 진단하거나 또는 모니터링하는 방법은 하나 이상의 표적 유전자좌에서 서열의 변화를 검출하거나 또는 확인하기 위해 DNA 클론 라이브러리에서 하나 이상의 표적 유전자좌의 정량적인 유전자 분석을 수행하는 것을 포함한다. 일부 실시양태에서, 변화는 카피 수의 변화이다.

[0366] 일부 실시양태에서, 유전적 상태, 병태 또는 질환을 검출하거나, 확인하거나, 예측하거나, 진단하거나 또는 모니터링하는 방법은 대상체의 생물학적 샘플로부터 세포성 DNA 또는 cfDNA를 분리하거나 또는 수득하고; 세포성 DNA 또는 cfDNA를 하나 이상의 말단-복구 효소로 처리하여 말단 복구 DNA를 생성하고; 말단 복구 DNA의 각각의 말단에 하나 이상의 어댑터를 부착하여 게놈 DNA 라이브러리를 생성하고; DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하고; DNA 클론 라이브러리에서 게놈 등가물의 수를 결정하고; DNA 클론 라이브러리에서 하나 이상의 표적 유전자좌의 정량적인 유전자 분석을 수행하여, 하나 이상의 표적 유전자좌에서 서열의 변화, 예를 들어 SNP, 전좌, 역위, 결실, 또는 카피 수의 변화를 검출하거나 또는 확인하는 것을 포함한다.

[0367] 일부 실시양태에서, 유전 질환; 유전적 모자이크 현상; 태아 검사; 친자 확인 검사; 친자 확인 검사; 약물 치료에 대한 반응의 예측; 의학적 병태의 진단 또는 모니터링; 미생물군 프로파일링; 병원균 스크리닝; 및 장기 이식 모니터링으로 이루어진 군으로부터 선택되는 유전적 상태, 또는 유전적 병태 또는 질환을 검출하거나, 확인하거나, 예측하거나, 진단하거나 또는 모니터링하는 방법은 대상체의 생물학적 샘플로부터 게놈 DNA를 획득하거나 또는 분리하고; DNA를 하나 이상의 말단-복구 효소로 처리하여 말단 복구 DNA를 생성하고; 말단 복구 DNA의 각각의 말단에 하나 이상의 어댑터를 부착하여 게놈 DNA 라이브러리를 생성하고; 게놈 DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하고; DNA 클론 라이브러리에서 게놈 등가물의 수를 결정하고; DNA 클론 라이브러리에서 하나 이상의 표적 유전자좌의 정량적인 유전자 분석을 수행하여, 하나 이상의 표적 유전자좌에서 서열에서 뉴클레오티드 전이 또는 전환, 뉴클레오티드 삽입 또는 결실, 게놈 재배열, 카피 수의 변화, 또는 유전자 융합을 검출하거나 또는 확인하는 것을 포함한다.

[0368] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 유전 질환의 예시적인 예에는 암, 알츠하이머 질환 (APOE1), 샤르코-마리-투스(Charcot-Marie-Tooth) 질환, 레버(Leber) 유전성 시신경병증 (LHON), 앤젤만(Angelman) 증후군 (UBE3A, 유비퀴틴-단백질 리가제 E3A), 프라더-윌리(Prader-Willi) 증후군 (염색체 15의 영역), β -지중해빈혈 (HBB, β -글로빈), 고셔(Gaucher) 질환 (유형 I) (GBA, 글루코세레브로시다제), 낭성 섬유증 (CFTR 상피 클로라이드 채널), 겸상 적혈구 질환 (HBB, β -글로빈), 테이-삭스(Tay-Sachs) 질환 (HEXA, 헥소사미니다제 A), 페닐케톤뇨증 (PAH, 페닐알라닌 히드롤라제), 가족성 고콜레스테롤혈증 (LDLR, 저밀도 지단백질 수용체), 성인 다낭성 신장 질환 (PKD1, 폴리스틴), 헌팅턴(Huntington) 질환 (HDD, 헌팅틴), 신경 섬유증증 유형 I (NF1, NF1 종양 억제 유전자), 근긴장성 이영양증 (DM, 미오토닌), 결절성 경화증 (TSC1, 투베린), 연골 무형성증 (FGFR3, 섬유모세포 성장 인자 수용체), 취약 X 증후군 (FMR1, RNA-결합 단백질), 뒤센(Duchenne)근이영양증 (DMD, 디스트로핀), 혈우병 A (F8C, 혈액 응고 인자 VIII), 레쉬-니한(Lesch-Nyhan) 증후군 (HPRT1, 하이포크산틴 구아닌 리보실트랜스퍼라제 1), 및 부신백질 이영양증 (ABCD1)이 포함되나 이로 제한되지 않는다.

[0369] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 암의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: B 세포암, 예를 들어 다발성 골수종, 흑색종, 유방암, 폐암 (예컨대 비-소세포 폐암종 또는 NSCLC), 기관지암, 결장직장암, 전립선암, 췌장암, 위암, 난소암, 비뇨기 방광암, 뇌암 또는 중추신경계암, 말초신경계암, 식도암, 자궁경부암, 자궁암 또는 자궁내막암, 구강암 또는 인두암, 간암, 신장암, 고환암, 담도암, 소장암 또는 맹장암, 침샘암, 갑상선암, 부신암, 골육종, 연골육종, 혈액 조직암, 선암종, 염증성 근섬유모세포 종양, 위장 간질성 종양 (GIST), 결장암, 다발성 골수종 (MM), 골수이형성 증후군 (MDS), 골수증식성 장애 (MPD), 급성 림프구성 백혈병 (ALL), 급성 골수구성 백혈병 (AML), 만성 골수구성 백혈병 (CML), 만성 림프구성 백혈병 (CLL), 진성 적혈구증가증, 호지킨(Hodgkin) 림프종, 비-호지킨 림프종 (NHL), 연조직 육종, 섬유육종, 점액육종, 지방육종, 골원성 육종, 척삭종, 혈관육종, 내피육종, 림프관육종, 림프관내피육종, 활막종, 중피종, 유잉(Ewing) 종양, 평활근육종, 횡문근육종, 편평상피 세포 암종, 기저 세포 암종, 선암종, 한선 암종, 피지선 암종, 유두 암종, 유두 선암종, 수질 암종, 기관지 암종, 신세포 암종, 간암, 담관 암종, 용모막 암종, 정상피종, 배아성 암종, 윌름스(Wilms) 종양, 방광 암종, 상피 암종, 신경교종, 성상세포종, 수모세포종, 두개인두종, 상의세포종, 송과체종, 혈관모세포종, 청신경종, 희소돌기아교세포종, 수막종, 신경모세포종, 망막모세포종, 여포성 림프종, 미만성 거대 B-세포 림프종, 맨틀 세포 림프종, 간세포성 암종, 갑상선암, 위암, 두경부암, 소세포암, 본태성 혈소판혈증, 원인 불명 골수 화생, 과호산구성 증후군, 전신 비만세포증, 가족성 과호산구성증, 만성 호산구성 백혈병, 신경내분비암, 유암종 등.

[0370] 일부 실시양태에서, 유전적 병변은 코스믹(Cosmic) 데이터베이스에서 주해된 병변 (병변 및 서열 데이터는 온라인으로 입수가능하고, 코스믹 웹사이트의 암 유전자 인구조사(Cancer Gene Census) 섹션으로부터 다운로드할 수 있음) 또는 캔서 게놈 아틀라스(Cancer Genome Atlas)에서 주해된 병변 (병변 및 서열 데이터는 온라인으로 입수가능하고, 캔서 게놈 아틀라스 웹사이트로부터 다운로드할 수 있음)이다.

[0371] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 암과 연관된 하나 이상의 유전적 병변을 보유하는 유전자의 예시적인 예에는 ABCB1, ABCC2, ABCC4, ABCG2, ABL1, ABL2, AKT1, AKT2, AKT3, ALDH4A1, ALK, APC, AR, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRAF, BRCA1, BRCA2, C1orf144, CARD11, CBL, CCND1, CCND2, CCND3, CCNE1, CDH1, CDH2, CDH20, CDH5, CDK4, CDK6, CDK8, CDKN2A, CDKN2B, CDKN2C, CEBPA, CHEK1, CHEK2, CRKL, CRLF2, CTNNB1, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DNMT3A, DOT1L, DPYD, EGFR,

EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, EPHX1, ERBB2, ERBB3, ERBB4, ERCC2, ERG, ESR1, ESR2, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FCGR3A, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, FOXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GSTP1, GUCY1A2, HOXA3, HRAS, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, ITPA, JAK1, JAK2, JAK3, JUN, KDR, KIT, KRAS, LRP1B, LRP2, LTK, MAN1B1, MAP2K1, MAP2K2, MAP2K4, MCL1, MDM2, MDM4, MEN1, MET, MITF, MLH1, MLL, MPL, MRE11A, MSH2, MSH6, MTHFR, MTOR, MUTYH, MYC, MYCL1, MYCN, NF1, NF2, NKX2-1, NOTCH1, NPM1, NQO1, NRAS, NRP2, NTRK1, NTRK3, PAK3, PAX5, PDGFRA, PDGFRB, PIK3CA, PIK3R1, PKHD1, PLCG1, PRKDC, PTCH1, PTEN, PTPN11, PTPRD, RAF1, RARA, RB1, RET, RICTOR, RPTOR, RUNX1, SLC19A1, SLC22A2, SLC01B3, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMO, SOD2, SOX10, SOX2, SRC, STK11, SULT1A1, TBX22, TET2, TGFBR2, TMPSR2, TNFRSF14, TOP1, TP53, TPMT, TSC1, TSC2, TYMS, UGT1A1, UMPS, USP9X, VHL, 및 WT1이 포함되나 이로 제한되지 않는다.

- [0372] 일부 실시양태에서, 유전적 병변은 뉴클레오티드 전이 또는 전환, 뉴클레오티드 삽입 또는 결실 (즉, indel), 게놈 재배열, 카피 수의 변화, 또는 유전자 융합을 포함한다. 일부 실시양태에서, 유전적 병변은 프레임시프트를 포함한다. 일부 실시양태에서, 유전적 병변은 스플라이싱의 변화를 포함한다. 일부 실시양태에서, 유전적 병변은 단일 뉴클레오티드 변동 (SNV)를 포함한다.
- [0373] 일부 실시양태에서, 유전적 병변은 ALK 유전자의 3' 코딩 영역을 또 다른 유전자에 융합시키는 유전자 융합이다.
- [0374] 일부 실시양태에서, 유전적 병변은 ALK 유전자의 3' 코딩 영역을 EML4 유전자에 융합시키는 유전자 융합이다.
- [0375] 일부 실시양태에서, 유전적 병변은 표 5 또는 표 6에 나타난 병변 중 어느 하나이다. 예를 들어, 유전적 병변은 TM 돌연변이, BRCA1 프레임시프트, BRCA2 프레임시프트, BRCA2 G4 돌연변이 (즉, G-사중화 구조의 형성을 유발하는 돌연변이), FANCA 스플라이스 돌연변이, HDAC2 프레임시프트, PALB2 Q479 돌연변이, 또는 ATM 프레임시프트일 수 있다. 일부 실시양태에서, 유전적 병변은 ERBB2 1655V 돌연변이, TP53 Q331 돌연변이, TP53 프레임시프트, EML4-ALK 융합, 또는 EGFR 증폭일 수 있다.
- [0376] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 태아 검사에 적합한 상태의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 다운 증후군 (삼염색체 21), 에드워드 증후군 (삼염색체 18), 파타우 증후군 (삼염색체 13), 클라인펠터 증후군 (XXY), 트리플 X 증후군, XYY 증후군, 삼염색체 8, 삼염색체 16, 터너 증후군 (XO), 로버트슨 전좌, 디조지 증후군 및 올프-허쉬호른 증후군.
- [0377] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 친자 확인 검사에 적합한 대립유전자의 예시적인 예에는 하기 16가지 이상이 포함되나 이로 제한되지 않는다: D20S1082, D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D4S2364, D9S1122, D2S1776, D10S1425, D3S3053, D5S2500, D1S1627, D3S4529, D2S441, D17S974, D6S1017, D4S2408, D9S2157, 아멜로게닌, D17S1301, D1GATA113, D18S853, D20S482, 및 D14S1434.
- [0378] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 약물 치료에 대한 반응을 예측하는데 적합한 유전자의 예시적인 예에는 하기 유전자 중 하나 이상이 포함되나 이로 제한되지 않는다: ABCB1 (ATP-결합 카세트, 서브-패밀리 B (MDR/TAP), 구성원 1), ACE (안지오텐신 I 전환 효소), ADH1A (알콜 데히드로게나제 1A (부류 I), 알파 폴리펩티드), ADH1B (알콜 데히드로게나제 IB (부류 I), 베타 폴리펩티드), ADH1C (알콜 데히드로게나제 1C (부류 I), 감마 폴리펩티드), ADRB1 (아드레날린성, 베타-1-, 수용체), ADRB2 (아드레날린성, 베타-2-, 수용체, 표면), AHR (아릴 탄화수소 수용체), ALDH1A1 (알데히드 데히드로게나제 1 패밀리, 구성원 A1), ALOX5 (아라키도네이트 5-리폭시게나제), BRCA1 (유방암 1, 조기 발병), COMT (카테콜-O-메틸트랜스퍼라제), CYP2A6 (시토크롬 P450, 패밀리 2, 서브패밀리 A, 폴리펩티드 6), CYP2B6 (시토크롬 P450, 패밀리 2, 서브패밀리 B, 폴리펩티드 6), CYP2C9 (시토크롬 P450, 패밀리 2, 서브패밀리 C, 폴리펩티드 9), CYP2C19 (시토크롬 P450, 패밀리 2, 서브패밀리 C, 폴리펩티드 19), CYP2D6 (시토크롬 P450, 패밀리 2, 서브패밀리 D, 폴리펩티드 6), CYP2J2 (시토크롬 P450, 패밀리 2, 서브패밀리 J, 폴리펩티드 2), CYP3A4 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 4), CYP3A5 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 5), DPYD (디히드로피리미딘 데히드로게나제), DRD2 (도파민 수용체 D2), F5 (응고 인자 V), GSTP1 (글루타티온 S-트랜스퍼라제 pi), HMGCR (3-히드록시-3-메틸글루타릴-조효소 A 리덕타제), KCNH2 (칼륨 전압-개폐 채널, 서브패밀리 H (eag-관련), 구성원 2), KCNJ11 (칼륨 내부-정류 채널, 서브패밀리 J, 구성원 11), MTHFR (5,10-메틸렌테트라하이드로폴레이트 리덕타제 (NADPH)), NQO1 (NAD(P)H 데히드로게나제,

퀴논 1), P2RY1 (푸린성 수용체 P2Y, G-단백질 커플링된, 1), P2RY12 (푸린성 수용체 P2Y, G-단백질 커플링된, 12), PTGIS (프로스타글란딘 I2 (프로스타시클린) 신타제), SCN5A (나트륨 채널, 전압-개폐, 유형 V, 알파 (긴 QT 증후군 3)), SLC19A1 (용질 운반체 패밀리 19 (폴레이트 수송체), 구성원 1), SLC01B1 (용질 운반체 유기 음이온 수송체 패밀리, 구성원 1B1), SULT1A1 (sul포트랜스퍼라제 패밀리, 세포질성, 1A, 페놀-선택, 구성원 1), TPMT (티오푸린 S-메틸트랜스퍼라제), TYMS (티미딜레이트 신테타제), UGT1A1 (UDP 글루쿠로노실트랜스퍼라제 1 패밀리, 폴리펩티드 A1), VDR (비타민 D (1,25- 디하이드록시비타민 D3) 수용체), VKORC1 (비타민 K 에폭시드 리덕타제 복합체, 서브유닛 1).

[0379] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 의학적 병태의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 졸중, 일과성 허혈성 발작, 외상성 뇌 손상, 심장 질환, 심장 마비, 협심증, 아테롬성 동맥경화증, 및 고혈압.

[0380] 본원에서 고려되는 조성물 및 방법에 의해 스크리닝될 수 있는 병원균의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 박테리아 진균, 및 바이러스.

[0381] 본원에서 고려되는 조성물 및 방법에 의해 스크리닝될 수 있는 박테리아 종의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 마이코박테리움(Mycobacterium) 종, 뉴모코커스(Pneumococcus) 종, 에스케리키아(Escherichia) 종, 캄필로박터(Campylobacter) 종, 코리네박테리움(Corynebacterium) 종, 클로스트리디움(Clostridium) 종, 스트렙토코커스(Streptococcus) 종, 스태필로코커스(Staphylococcus) 종, 슈도모나스(Pseudomonas) 종, 시겔라(Shigella) 종, 트레포네마(Treponema) 종, 또는 살모넬라(Salmonella) 종.

[0382] 종 본원에서 고려되는 조성물 및 방법에 의해 스크리닝될 수 있는 진균의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 아스페르길리스(Aspergillus) 종, 블라스토마이세스(Blastomyces) 종, 칸디다(Candida) 종, 코씨시오이데스(Coccidioides) 종, 크립토코커스(Cryptococcus) 종, 더마토파이테스(dermatophytes), 티네아(Tinea) 종, 트리코파이톤(Trichophyton) 종, 미크로스포룸(Microsporum) 종, 푸사리움(Fusarium) 종, 히스토플라스마(Histoplasma) 종, 뮤코로마이코티나(Mucoromycotina) 종, 뉴모시스티스(Pneumocystis) 종, 스포로트릭스(Sporothrix) 종, 엑세로필룸(Exserophilum) 종, 또는 클라도스포리움(Cladosporium) 종.

[0383] 본원에서 고려되는 조성물 및 방법에 의해 스크리닝될 수 있는 바이러스의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 인플루엔자 A, 예컨대 H1N1, H1N2, H3N2 및 H5N1 (조류 독감), 인플루엔자 B, 인플루엔자 C 바이러스, A형 간염 바이러스, B형 간염 바이러스, C형 간염 바이러스, D형 간염 바이러스, E형 간염 바이러스, 로타바이러스, 노워크(Norwalk) 바이러스 그룹의 임의의 바이러스, 장내 아데노바이러스, 파르보바이러스, 탕기열 바이러스, 원숭이 두창, 모노네가비라레스, 리사바이러스, 예컨대 광견병 바이러스, 라고스 박쥐 바이러스, 모콜라 바이러스, 두벤하게 바이러스, 유럽 박쥐 바이러스 1 & 2 및 호주 박쥐 바이러스, 에페메로바이러스, 베시쿨로바이러스, 수포성 구내염 바이러스 (VSV), 헤르페스 바이러스, 예컨대 단순 헤르페스 바이러스 유형 1 및 2, 대상포진, 시토메갈로바이러스, 엡스타인-바르 바이러스 (EBV), 인간 헤르페스 바이러스 (HHV), 인간 헤르페스 바이러스 유형 6 및 8, 몰로니 뮤린 백혈병 바이러스 (M-MuLV), 몰로니 뮤린 육종 바이러스 (MoMSV), 하르베이 뮤린 육종 바이러스 (HaMuSV), 뮤린 유방 종양 바이러스 (MuMTV), 긴팔 원숭이 백혈병 바이러스 (GaLV), 고양이 백혈병 바이러스 (FLV), 스푸마바이러스, 프렌드 뮤린 백혈병 바이러스, 뮤린 줄기 세포 바이러스 (MSCV) 및 라우스 육종 바이러스 (RSV), HIV (인간 면역 결핍 바이러스; 예컨대 HIV 유형 1, 및 HIV 유형 2), 비스나-마에디 바이러스 (VSV) 바이러스, 염소 관절염-뇌염 바이러스 (CAEV), 말 감염성 빈혈 바이러스 (EIAV), 고양이 면역 결핍 바이러스 (FIV), 소 면역 결핍 바이러스 (BIV), 및 유인원 면역 결핍 바이러스 (SIV), 유두종 바이러스, 뮤린 감마헤르페스바이러스, 아레나바이러스, 예컨대 아르헨티나 출혈 열 바이러스, 볼리비아 출혈 열 바이러스, 사비아-연관된 출혈 열 바이러스, 베네수엘라 출혈 열 바이러스, 라사 열 바이러스, 마추포 바이러스, 림프구성 맥락수막염 바이러스 (LCMV), 부냐비리다에, 예컨대 크림-콩고 출혈 열 바이러스, 한타바이러스, 출혈 열 동반 신 증후군 유행 바이러스, 리프트 밸리 열 바이러스, 필로비리다에 (필로바이러스), 예컨대 에볼라 출혈 열 및 마르부르크 출혈 열, 플라비비리다에, 예컨대 카이사누르 포레스트 질환 바이러스, 움스크 출혈 열 바이러스, 진드기-매개 뇌염 유행 바이러스 및 파라믹소비리다에, 예컨대 헨트라 바이러스 및 니파 바이러스, 대두창 및 소두창 (천연두), 알파바이러스, 예컨대 베네수엘라 말 뇌염 바이러스, 동부 말 뇌염 바이러스, 서부 말 뇌염 바이러스, SARS-연관된 코로나 바이러스 (SARS-CoV), 웨스트 나일 바이러스, 및 임의의 뇌염 유행 바이러스.

[0384] 본원에서 고려되는 조성물 및 방법에 의해 검출되거나, 확인되거나, 예측되거나, 진단되거나 또는 모니터링될 수 있는 이식 수용자에서 장기 이식을 모니터링하는데 적합한 유전자의 예시적인 예에는 하기 유전자 중 하나

이상이 포함되나 이로 제한되지 않는다: HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DP, 및 HLA-DQ.

- [0385] 일부 실시양태에서, 생물정보학적 분석을 이용하여, cfDNA 클론 라이브러리에서 분석되는 게놈 등가물의 수를 정량화하거나; 표적 유전자와 내에서 유전자 변이체를 검출하거나; 표적 유전자와 내에서 돌연변이를 검출하거나; 표적 유전자와 내에서 유전자 융합을 검출하거나; 또는 표적 유전자와 내에서 카피 수 변동을 측정한다.
- [0386] 일부 실시양태에서, 대상체의 생물학적 샘플로부터 게놈 DNA를 분리하거나 또는 수득하고; DNA를 하나 이상의 말단-복구 효소로 처리하여 말단 복구 DNA를 생성하고; 말단 복구 DNA의 각각의 말단에 하나 이상의 어댑터를 부착하여 DNA 라이브러리를 생성하고; DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하고; DNA 클론 라이브러리에서 게놈 등가물의 수를 결정하고; DNA 클론 라이브러리에서 유전 질환과 연관된 하나 이상의 생체 마커의 정량적인 유전자 분석을 수행하는 것을 포함하는, 유전 질환에 대한 동반 진단이 제공되며, 하나 이상의 생체마커 중 적어도 하나의 검출 또는 검출 실패는 대상체가 유전 질환에 대해 치료되어야 하는지 여부를 나타낸다. 일부 실시양태에서, DNA는 cfDNA이다. 일부 실시양태에서, DNA는 세포성 DNA이다.
- [0387] 본원에 사용된 바와 같이, 용어 "동반 진단"은 특정한 항암 요법과 관련된 진단 시험을 지칭한다. 특정한 실시양태에서, 진단 방법은 생물학적 샘플과 연관된 생체마커에서 유전적 병변을 검출하는 것을 포함하며, 이로써 환자가 항암 요법으로 치료되어야 하는지 또는 그렇지 않은지에 대한 신속한 확인을 가능하게 한다.
- [0388] 항암 요법에는 수술, 방사선 요법, 화학 요법, 항암 약물, 및 면역조절제가 포함되나 이로 제한되지 않는다.
- [0389] 항암 약물의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 알킬화제, 예컨대 티오테파 및 시클로포스파미드 (시톡산(CYTOXAN)TM); 알킬 술포네이트, 예컨대 부술폰, 임프로술폰 및 피포술폰; 아지리딘, 예컨대 벤조도파, 카르보쿠온, 메투레도파, 및 우레도파; 에틸렌이민 및 메틸아멜라민, 예컨대 알트레타민, 트리에틸렌멜라민, 트리에틸렌포스포라미드, 트리에틸렌티오포스포라미드 및 트리메틸올로멜라민 레슘; 질소 머스타드, 예컨대 클로람부실, 클로르나파진, 클로로포스파미드, 에스트라무스틴, 이포스파미드, 메클로레타민, 메클로레타민 옥시드 히드로클로라이드, 멜팔란, 노벤비킨, 페네스테린, 프레드니무스틴, 트로포스파미드, 우라실 머스타드; 니트로소우레아, 예컨대 카르무스틴, 클로로조토신, 포테무스틴, 로무스틴, 니무스틴, 라니무스틴; 항생제, 예컨대 아클라시노마이신, 악티노마이신, 오프라마이신, 아자세린, 블레오마이신, 캅티노마이신, 칼리케아미신, 카라비신, 카르미노마이신, 카르지노필린, 크로모마이신, 닥티노마이신, 다우노루비신, 데토루비신, 6-디아조-5-옥소-L-노르류신, 독소루비신 및 그의 PEG화 제형, 에피루비신, 에소루비신, 이다루비신, 마르셀로마이신, 미토마이신, 미코페놀산, 노갈라마이신, 올리보마이신, 페플로마이신, 포트피로마이신, 푸로마이신, 쿠엘라마이신, 로도루비신, 스트렙토니그린, 스트렙토조신, 투베르시딘, 우베니멕스, 지노스타틴, 조루비신; 항대사물, 예컨대 메토트렉세이트 및 5-플루오로우라실 (5-FU); 엽산 유사체, 예컨대 데노프테린, 메토트렉세이트, 프레로프테린, 트리메트렉세이트; 퓨린 유사체, 예컨대 플루다라빈, 6-머캅토프린, 티아미프린, 티오구아닌; 피리미딘 유사체, 예컨대 안시타빈, 아자시티딘, 6-아자우리딘, 카르모푸르, 시타라빈, 디데옥시우리딘, 독시플루리딘, 에노시타빈, 플록수리딘, 5-FU; 안드로젠, 예컨대 칼루스테론, 드로모스타놀론 프로피오네이트, 에피티오스타놀, 메피티오스탄, 테스톨락톤; 항부신제, 예컨대 아미노글루테티미드, 미토탄, 트릴로스탄; 엽산 보충제, 예컨대 프롤린산; 아세글라톤; 알도포스파미드 글리코시드; 아미노레볼린산; 암사크린; 베스트라부실; 비산트렌; 에다트락세이트; 데포파민; 데메콜신; 디아지쿠온; 엘포르미틴; 엘리프티늄 아세테이트; 에토글루시드; 질산갈륨; 히드록시우레아; 렌티난; 로니다민; 미토구아존; 미톡산트론; 모피다몰; 니트라크린; 펜토스타틴; 페나메트; 피라루비신; 포도필린산; 2-에틸히드라지드; 프로카르바진; PSK®; 라족산; 시조피란; 스피로게르마늄; 테누아존산; 트리아지쿠온; 2, 2', 2''-트리클로로트리메틸아민; 우레탄; 빈데신; 다카르바진; 만노무스틴; 미토브로니톨; 미톨락톨; 피포브로만; 가시토신; 아라비노시드 ("Ara-C"); 시클로포스파미드; 티오테파; 톡소이드, 예를 들어 파클리탁셀 (탁솔(TAXOL)®, 브리스톨-마이어스 스킵 온콜로지(Bristol-Myers Squibb Oncology), 뉴저지주 프린스턴) 및 독세탁셀 (탁소테레(TAXOTERE)®, 르네-폴랭크 로어(Rhne-Poulenc Rorer), 프랑스 앙토니); 클로람부실; 겐시타빈; 6-티오구아닌; 머캅토프린; 메토트렉세이트; 백금 유사체, 예컨대 시스플라틴 및 카르보플라틴; 빈블라스틴; 백금; 에토포시드 (VP-16); 이포스파미드; 미토마이신 C; 미톡산트론; 빈크리스틴; 비노렐빈; 나벨빈; 노반트론; 테니포시드; 아미노프테린; 크셀로다; 이반드로네이트; CPT-11; 토포아이스머라제 억제제 RFS 2000; 디플루오로메틸로미딘 (DMFO); 레티노산 유도체, 예컨대 타르그레틴(Targretin)TM (백사로텐), 판레틴(Panretin)TM (알리트레티노인); 온탁(ONTAK)TM (테닐류킨 디프티톡스); 에스페라미신; 카페시타빈; 및 상기 중 어느 것의 제약상 허용가능한 염, 산 또는 유도체. 암에 대한 호르몬 작용을 조절하거나 또는 억제하는 작용을 하는 항호르몬 작용제, 예컨대 항에스트로젠, 예를 들어 타목시펜, 팔록시펜, 아로마타제, 예컨대 억제제 4(5)-이미다졸, 4-히드록시타목시펜, 트리옥시펜, 케옥시펜, LY117018, 오나프리스톤, 및 토레미펜 (파레스톤); 및 항안드로젠, 예컨대 플루타미드, 닐루타미드, 비칼루

타미드, 류프롤리드, 및 고세렐린; 및 상기 중 어느 것의 제약상 허용가능한 염, 산 또는 유도체 또한 이 정의에 포함된다.

[0390] 면역조절제의 예시적인 예에는 하기가 포함되나 이로 제한되지 않는다: 시클로스포린, 타크롤리무스, 트레스페리무스, 피메크롤리무스, 시롤리무스, 베롤리무스, 라플루니무스, 라키니모드 및 이미퀴모드, 뿐만 아니라 그의 유사체, 유도체, 염, 이온 및 복합체.

[0391] 일부 실시양태에서, 항암 약물에는 폴리-ADP 리보스 폴리머라제 (PARP) 억제제가 포함될 수 있다. PARP 억제제의 예시적인 예에는 올라파립 (AZD-2281), 루카파립 (AG014699 또는 PF-01367338, 니라파립 (MK-4827), 탈라조파립 (BMN-673) 벨리파립 (ABT-888), CEP 9722, E7016, BGB-290, 3-아미노벤즈아미드가 포함되나 이로 제한되지 않는다.

[0392] 본 명세서에 인용된 모든 공보, 특허 출원 및 허여된 특허는 각각의 개별 공보, 특허 출원 또는 허여된 특허가 구체적으로 및 개별적으로 참조로 포함되는 것과 같이 본원에 참조로 포함된다. 특히, 국제 PCT 공개 번호 WO 2016/028316의 전체 내용은 구체적으로 참조로 포함된다.

[0393] 진술한 발명이 명확한 이해를 위해 설명 및 예시의 방식으로 다소 상세하게 기재되었지만, 관련 기술분야의 통상의 기술자에게는 본 발명의 교시내용에 비추어 첨부된 청구항의 개념 또는 범위를 벗어나지 않고 특정 변화 및 변형이 이루어질 수 있음을 용이하게 명백할 것이다. 하기 실시예는 제한하는 방식이 아니라 단지 설명하는 방식으로 제공된다. 관련 기술분야의 기술자는 본질적으로 유사한 결과를 산출하도록 변화되거나 또는 변형될 수 있는 다양한 중요하지 않은 파라미터를 용이하게 인식할 것이다.

[0394] 본 발명의 일부 실시양태의 실시는 구체적으로 반대로 나타내지 않는다면 관련 기술분야의 기술 내에 있는 화학, 생화학, 유기 화학, 분자 생물학, 미생물학, 재조합 DNA 기술, 유전학, 면역학 및 세포 생물학의 통상적인 방법을 이용할 것이며, 여러 가지가 설명의 목적으로 하기에 기재된다. 이러한 기술은 참고문헌에 충분히 설명되어 있다. 예를 들어, [Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (3rd Edition, 2001); Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual* (1982); Ausubel *et al.*, *Current Protocols in Molecular Biology* (John Wiley and Sons, updated July 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); *Transcription and Translation* (B. Hames & S. Higgins, Eds., 1984); Perbal, *A Practical Guide to Molecular Cloning* (1984); 및 Harlow and Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998)]을 참조한다.

[0395] 넘버링된 실시양태

[0396] 첨부된 청구항에도 불구하고, 하기 넘버링된 실시양태 또한 본 개시내용의 일부를 형성한다.

[0397] 1. 하기를 포함하는 다기능성 어댑터이며:

[0398] a. 라이게이션 가닥 올리고뉴클레오타이드, 및

[0399] b. 라이게이션 가닥 올리고뉴클레오타이드의 3' 말단에 있는 영역에 혼성화하고 그와 듀플렉스를 형성할 수 있는 비-라이게이션 가닥 올리고뉴클레오타이드;

[0400] 샘플로부터의 dsDNA 단편과 접촉시, 라이게이션 가닥 올리고뉴클레오타이드는 dsDNA 단편의 각각의 가닥의 5' 말단에 라이게이션하고;

[0401] 라이게이션 가닥 올리고뉴클레오타이드는 하기를 포함하고:

[0402] (i) 3' 말단 오버행;

[0403] (ii) 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 증폭 영역;

[0404] (iii) 고유의 다기능성 ID 영역;

[0405] (iv) 고유의 분자 식별자 (UMI) 멀티플라이어; 및

[0406] (v) 비-라이게이션 가닥 올리고뉴클레오타이드에 대해 적어도 부분적으로 상보적인 폴리뉴클레오타이드 서열을 포함

하는 앵커 영역;

- [0407] dsDNA 단편은 각각의 가닥의 5' 말단에서 포스페이트 기 및 각각의 가닥의 3' 말단에서 오버행을 포함하고;
- [0408] 각각의 dsDNA 단편은 다기능성 ID 영역 및 UMI 멀티플라이어의 조합에 의해 식별될 수 있고;
- [0409] 샘플은 다기능성 ID 영역에 의해 식별될 수 있는 것인 다기능성 어댑터.
- [0410] 2. 실시양태 1에 있어서,
- [0411] 라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dT 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dA 오버행을 포함하거나, 또는
- [0412] 라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dA 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dT 오버행을 포함하는 것인 다기능성 어댑터.
- [0413] 3. 실시양태 1에 있어서,
- [0414] 라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dC 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dG 오버행을 포함하거나, 또는
- [0415] 라이게이션 가닥 올리고뉴클레오타이드가 3' 말단에서 dG 오버행을 포함하고, dsDNA 단편이 각각의 가닥의 3' 말단에서 dC 오버행을 포함하는 것인 다기능성 어댑터.
- [0416] 4. 실시양태 1-3 중 어느 하나에 있어서, 라이게이션 가닥 올리고뉴클레오타이드에서 증폭 영역이 PCR, LAMP, NASBA, SDA, RCA, 또는 LCR을 위한 프라이머 인식 부위의 역할을 할 수 있는 폴리뉴클레오타이드 서열을 포함하는 것인 다기능성 어댑터.
- [0417] 5. 실시양태 1-4 중 어느 하나에 있어서, 비-라이게이션 가닥 올리고뉴클레오타이드가 dsDNA 단편의 5' 말단에 대한 라이게이션 및/또는 어댑터 이량체 형성을 방지하는 변형을 그의 3' 말단에 포함하는 것인 다기능성 어댑터.
- [0418] 6. 실시양태 1-5 중 어느 하나에 있어서, 샘플이 조직 생검인 다기능성 어댑터.
- [0419] 7. 실시양태 6에 있어서, 조직 생검이 종양 또는 종양으로 의심되는 조직으로부터 취해지는 것인 다기능성 어댑터.
- [0420] 8. 실시양태 1-7 중 어느 하나에 있어서, dsDNA 단편이 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 또는 탈메틸화 DNA를 포함하는 것인 다기능성 어댑터.
- [0421] 9. 실시양태 8에 있어서, dsDNA가 시험 샘플로부터 단리되거나 또는 생성되고; 시험 샘플이 양수, 혈액, 혈장, 혈청, 정액, 림프액, 뇌척수액, 안구액, 소변, 타액, 대변, 점액, 및 땀으로 이루어진 군으로부터 선택된 생물학적 샘플을 포함하는 것인 다기능성 어댑터.
- [0422] 10. 실시양태 1-9 중 어느 하나에 있어서, dsDNA 단편이 하기를 포함하는 단계에 의해 획득되는 것인 다기능성 어댑터:
- [0423] a.) 시험 샘플로부터 세포성 DNA를 단리하는 단계; 및
- [0424] b.) 세포성 DNA를 단편화하여 게놈 DNA 단편을 획득하는 단계.
- [0425] 11. 실시양태 10에 있어서, 단계 (b)가 세포성 DNA를 적어도 하나의 소화 효소와 접촉시킴으로써 수행되는 것인 다기능성 어댑터.
- [0426] 12. 실시양태 10에 있어서, 단계 (b)가 세포성 DNA에 기계적 스트레스를 적용함으로써 수행되는 것인 다기능성 어댑터.
- [0427] 13. 실시양태 12에 있어서, 기계적 스트레스가 세포성 DNA를 초음파 처리함으로써 적용되는 것인 다기능성 어댑터.
- [0428] 14. 실시양태 10에 있어서, 단계 (b)가 세포성 DNA를 하나 이상의 화합물과 접촉시켜 세포성 DNA의 하나 이상의 결합을 화학적으로 파괴시킴으로써 수행되는 것인 다기능성 어댑터.
- [0429] 15. 실시양태 1-14 중 어느 하나에 있어서, 증폭 영역이 10 내지 50개 뉴클레오타이드 길이인 다기능성 어댑터.

- [0430] 16. 실시양태 15에 있어서, 증폭 영역이 20 내지 30개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0431] 17. 실시양태 15에 있어서, 증폭 영역이 25개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0432] 18. 실시양태 1-17 중 어느 하나에 있어서, 다기능성 ID 영역이 3 내지 50개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0433] 19. 실시양태 18에 있어서, 다기능성 ID 영역이 3 내지 15개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0434] 20. 실시양태 18에 있어서, 다기능성 ID 영역이 8개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0435] 21. 실시양태 1-20 중 어느 하나에 있어서, UMI 멀티플라이어가 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있는 것인 다기능성 어댑터.
- [0436] 22. 실시양태 21에 있어서, UMI 멀티플라이어가 1 내지 5개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0437] 23. 실시양태 21에 있어서, UMI 멀티플라이어가 3개 뉴클레오타이드 길이이고, 64가지 가능한 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0438] 24. 실시양태 1-23 중 어느 하나에 있어서, 앵커 영역이 1 내지 50개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0439] 25. 실시양태 24에 있어서, 앵커 영역이 5 내지 25개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0440] 26. 실시양태 24에 있어서, 앵커 영역이 10개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0441] 27. 실시양태 1-26 중 어느 하나에 있어서, 복수개의 다기능성 어댑터가 복수개의 dsDNA 단편에 라이게이션되는 것인 다기능성 어댑터.
- [0442] 28. 실시양태 27에 있어서, dsDNA 단편이 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구되는 것인 다기능성 어댑터.
- [0443] 29. 실시양태 27 또는 28에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 증폭 영역이 동일한 뉴클레오타이드 서열을 포함하는 것인 다기능성 어댑터.
- [0444] 30. 실시양태 29에 있어서, 동일한 뉴클레오타이드 서열이 PCR 프라이머 결합 부위를 포함하는 것인 다기능성 어댑터.
- [0445] 31. 실시양태 27-30 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 2 내지 10,000가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0446] 32. 실시양태 31에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 50 내지 500가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0447] 33. 실시양태 31에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 100 내지 400가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0448] 34. 실시양태 31에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 60가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0449] 35. 실시양태 31-34 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 8개 뉴클레오타이드 길이인 다기능성 어댑터.
- [0450] 36. 실시양태 31-35 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터가 64 내지 2,560,000가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하는 것인 다기능성 어댑터.
- [0451] 37. 실시양태 31-36 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터가 3840가지의 고유의 뉴클레오타이드 서열 중 1가지를 포함하고, 각각의 뉴클레오타이드 서열이 적어도 2의 해밍 거리만큼 3840가지의 고유의 뉴클레오타이드 서열의 임의의 다른 서열과 별개인 것인 다기능성 어댑터.
- [0452] 38. 실시양태 31-37 중 어느 하나에 있어서, 각각의 복수개의 다기능성 어댑터가 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있는 UMI 멀티플라이어를 포함하는 것인 다기능성 어댑터.
- [0453] 39. 실시양태 31-38 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 UMI 멀티플라이어가 1 내지 5개 뉴클레오타이드 길이인 다기능성 어댑터.

- [0454] 40. 실시양태 39에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 UMI 멀티플라이어가 3개 뉴클레오티드 길이인 다기능성 어댑터.
- [0455] 41. 실시양태 31-40 중 어느 하나에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 앵커 영역이 4가지 뉴클레오티드 서열 중 1가지를 포함하고, 주어진 서열의 각각의 다기능성 ID 영역이 4가지 앵커 영역 중 각각의 1가지와 쌍을 형성할 수 있는 것인 다기능성 어댑터.
- [0456] 42. 실시양태 31에 있어서, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 증폭 영역이 동일한 뉴클레오티드 서열을 포함하고;
- [0457] 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 다기능성 ID 영역이 8개 뉴클레오티드 길이이고;
- [0458] 각각의 다기능성 ID 영역의 뉴클레오티드 서열이 적어도 2의 해밍 거리만큼 복수개의 다기능성 어댑터의 임의의 다른 다기능성 ID 영역의 뉴클레오티드 서열과 별개이고;
- [0459] 각각의 복수개의 다기능성 어댑터가 다기능성 ID 영역에 인접하거나 또는 그 안에 함유되어 있는 UMI 멀티플라이어를 포함하고, 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 UMI 멀티플라이어가 3개 뉴클레오티드 길이이고, 각각의 가능한 뉴클레오티드 서열의 UMI 멀티플라이어가 복수개의 다기능성 어댑터의 각각의 다기능성 ID 영역과 쌍을 형성하고,
- [0460] 복수개의 다기능성 어댑터의 각각의 다기능성 어댑터의 앵커 영역이 4가지 뉴클레오티드 서열 중 1가지를 포함하고, 주어진 서열의 각각의 다기능성 ID 영역이 4가지 앵커 영역 중 각각의 1가지와 쌍을 형성할 수 있는 것인 다기능성 어댑터.
- [0461] 43. 다기능성 어댑터 및 dsDNA 단편을 포함하는 복합체이며, 다기능성 어댑터는 실시양태 1-42의 다기능성 어댑터 중 어느 하나로부터 선택되는 것인 복합체.
- [0462] 44. a.) 복수개의 다기능성 어댑터를 복수개의 dsDNA 단편과 라이게이션하여, 복수개의 다기능성 어댑터/dsDNA 단편 복합체를 생성하는 단계이며, 여기서 각각의 복수개의 다기능성 어댑터는 실시양태 1-42의 다기능성 어댑터 중 어느 하나로부터 선택되고; 각각의 복수개의 복합체는 실시양태 43의 복합체를 포함하는 것인 단계; 임의적으로,
- [0463] b.) 단계 (a)로부터의 복수개의 복합체를 하나 이상의 효소와 접촉시켜, 복수개의 연속된 어댑터-태그 부착된 DNA 단편을 포함하는 어댑터-태그 부착된 DNA 라이브러리를 형성하는 단계를 포함하는, 어댑터-태그 부착된 DNA 라이브러리를 제조하는 방법.
- [0464] 45. 실시양태 44에 있어서, 복수개의 dsDNA 단편이 무세포 DNA (cfDNA), 게놈 DNA (gDNA), 상보성 DNA (cDNA), 미토콘드리아 DNA, 메틸화 DNA, 또는 탈메틸화 DNA를 포함하는 것인 방법.
- [0465] 46. 실시양태 44 또는 45에 있어서, 복수개의 dsDNA 단편이 복수개의 다기능성 어댑터와 라이게이션되기 전에 말단 복구되는 것인 방법.
- [0466] 47. 실시양태 44-46 중 어느 하나에 있어서, 복수개의 dsDNA 단편이 전체 게놈 라이브러리, 앰플리콘 라이브러리, 전체 엑솜 라이브러리, cDNA 라이브러리, 또는 메틸화 DNA 라이브러리로 이루어진 목록으로부터 선택된 라이브러리로부터 획득되는 것인 방법.
- [0467] 48. 실시양태 44-47 중 어느 하나에 있어서, 비-라이게이션 가닥 올리고뉴클레오티드가 단계 (b)에서 다기능성 어댑터/dsDNA 단편 복합체로부터 대체되는 것인 방법.
- [0468] 49. 실시양태 44-48 중 어느 하나에 있어서, 하나 이상의 효소가 DNA 리가제 또는 RNA 리가제를 포함하는 것인 방법.
- [0469] 50. 실시양태 49에 있어서, DNA 리가제가 T4 DNA 리가제 또는 Taq DNA 리가제를 포함하는 것인 방법.
- [0470] 51. 실시양태 44-50 중 어느 하나에 있어서, 복수개의 연속된 어댑터-태그 부착된 DNA 단편을 증폭시켜, 복수개의 증폭된 연속된 어댑터-태그 부착된 dsDNA 단편을 포함하는 증폭된 어댑터-태그 부착된 DNA 라이브러리를 생성하는 것을 추가로 포함하는 방법.
- [0471] 52. 실시양태 51에 있어서, 1개 이상의 프라이머가 증폭을 위해 사용되는 것인 방법.
- [0472] 53. 실시양태 52에 있어서, 1개 이상의 프라이머가 어댑터의 프라이머-결합 영역에 혼성화하는 범용 프라이머

결합 서열을 포함하는 것인 방법.

- [0473] 54. 실시양태 44-53 및 실시양태 67 중 어느 하나의 방법에 따라 생성된 어댑터-태그 부착된 DNA 라이브러리.
- [0474] 55. a.) 실시양태 54의 어댑터-태그 부착된 DNA 라이브러리를 1개 이상의 다기능성 포획 프로브와 혼성화하여, 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 형성하는 단계이며, 여기서 각각의 다기능성 포획 프로브는 하기를 포함하는 것인 단계:
- [0475] i.) 파트너 올리고뉴클레오타이드에 혼성화할 수 있는 제1 영역, 여기서 임의적으로 제1 영역은 PCR 프라이머 결합 부위를 포함하는 꼬리 서열을 포함함;
- [0476] ii.) 어댑터-태그 부착된 DNA 라이브러리에서 표적 영역에 혼성화할 수 있는 제2 영역;
- [0477] b.) 단계 (a)로부터의 1개 이상의 포획 프로브/어댑터-태그 부착된 DNA 복합체를 분리하는 단계이며, 여기서 각각의 분리된 포획 프로브/어댑터-태그 부착된 DNA 복합체는 포획 프로브 및 어댑터-태그 부착된 DNA 단편을 포함하는 것인 단계;
- [0478] c.) 단계 (b)로부터의 분리된 포획 프로브/DNA 단편 복합체를 효소적으로 처리하여, 혼성체 분자를 포함하는 프로브-포획된 DNA 라이브러리를 생성하는 단계이며, 여기서 각각의 혼성체 분자는 하기를 포함하는 것인 단계를 포함하는, 프로브-포획된 라이브러리를 제조하는 방법:
- [0479] i) 포획 프로브 또는 그의 상보체의 적어도 일부분;
- [0480] ii) DNA 단편 또는 그의 상보체의 적어도 일부분; 및
- [0481] iii) 어댑터.
- [0482] 56. 실시양태 55에 있어서, (c)의 효소적 처리 단계가 주형으로서 복합체의 어댑터-태그 부착된 DNA 단편을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행하는 것을 포함하는 것인 방법.
- [0483] 57. 실시양태 55 또는 실시양태 56에 있어서, 적어도 1개의 포획 프로브가 표적 영역에서 특이적인 영역의 하류에서 혼성화하고, 적어도 1개의 포획 프로브가 표적 영역에서 특이적인 영역의 상류에서 혼성화하는 것인 방법.
- [0484] 58. 실시양태 55-57 중 어느 하나에 있어서, 포획 프로브가 시퀀싱 프라이머 인식 서열을 포함하는 것인 방법.
- [0485] 59. 실시양태 55-58 중 어느 하나에 있어서,
- [0486] d.) 단계 (c)로부터의 혼성체 분자에 대해 PCR을 수행하여, 증폭된 혼성체 분자를 포함하는 증폭된 라이브러리를 생성하는 단계를 추가로 포함하는 방법.
- [0487] 60. 실시양태 55-58 중 어느 하나에 따라 생성된 혼성체 분자를 포함하는 프로브-포획된 라이브러리.
- [0488] 61. 실시양태 59에 따라 생성된 증폭된 프로브-포획된 라이브러리.
- [0489] 62. 실시양태 60의 혼성체 분자의 프로브-포획된 라이브러리에 대해 표적화된 유전자 분석을 수행하는 것을 포함하는 방법.
- [0490] 63. 실시양태 61의 증폭된 프로브-포획된 라이브러리에 대해 표적화된 유전자 분석을 수행하는 것을 포함하는 방법.
- [0491] 64. 실시양태 62 또는 63에 있어서, 표적화된 유전자 분석이 서열 분석을 포함하는 것인 방법.
- [0492] 65. 실시양태 62 또는 63에 있어서, 표적화된 유전자 분석이 카피 수 분석을 포함하는 것인 방법.
- [0493] 66. 실시양태 62-65 중 어느 하나에 있어서, 각각의 혼성체 분자에서 포획 프로브 영역 중 전부 또는 일부가 시퀀싱되는 것인 방법.
- [0494] 67. 실시양태 44-53 중 어느 하나에 있어서, 복수개의 복합체의 각각의 다기능성 어댑터/dsDNA 단편 복합체가 dsDNA 단편의 각각의 말단에 라이게이션된 다기능성 어댑터를 포함하는 것인 방법.
- [0495] **등가물**
- [0496] 본 발명이 상기 제시된 구체적인 실시양태와 관련하여 기재되었지만, 그의 여러 대안, 변형 및 다른 변동이 관련 기술분야의 통상의 기술자에게 명백할 것이다. 이러한 모든 대안, 변형 및 변동은 본 발명의 개념 및 범위를

내에 속하는 것으로 의도된다

[0497] 추가로, 본원에 기재된 임의의 방법이 본원에 기재된 임의의 작용제의 용도, 의약의 제조, 본원에 기재된 임의의 장애의 치료에 대해 스위스-유형 포맷으로 제작성될 수 있는 것으로 의도된다. 마찬가지로, 본원에 기재된 임의의 방법이 용도를 위한 화합물 청구항으로서 또는 화합물의 용도 청구항으로서 제작성될 수 있는 것으로 의도된다.

[0498] 본원에 기재된 모든 공보, 특허 및 특허 출원은 그들의 전문이 본원에 참조로 포함된다.

[0499] **실시예**

[0500] 본 개시내용은 하기 실시예에 의해 추가로 설명되며, 이는 본원에 기재된 구체적인 절차에 대한 범위 또는 개념으로 본 개시내용을 제한하는 것으로 해석되지 않아야 한다. 실시예가 특정 실시양태를 설명하기 위해 제공되며, 그에 의해 본 개시내용의 범위가 제한되지 않음을 이해해야 한다. 본 개시내용의 개념을 벗어나지 않고 관련 기술분야의 기술자에게 그 자체로 제안될 수 있는 그의 다양한 다른 실시양태, 변형 및 등가물에 의존할 수 있음을 추가로 이해해야 한다.

[0501] **실시예 1: DNA 라이브러리의 제조**

[0502] 유전자 변이체 (코리엘 인스티튜트 포 메디컬 리서치 또는 세라케어 라이프 사이언스즈, 인크.)를 보유하는 불멸화 세포로부터 분리된 무세포 DNA 및 게놈 DNA를 이 실시예에서 NGS 라이브러리 (어댑터-태그 부착된 DNA 라이브러리) 구축을 위해 사용하였다.

[0503] **표 1: 실험에서 사용된 샘플**

샘플 유형	샘플 투입 (ng)	샘플 설명
야생형 (cfDNA)	50	건강한 공여자의 혈액 샘플로부터 분리된 cfDNA
DNA 혼합물 1	20	HRD (상동성 복구 결핍) 유전자 변이체 (ATM, BRCA1 및 BRCA2, BRCA2, FANCA, HDAC2, 및 PALB2를 보유하는 게놈 및 합성 DNA의 혼합물. 게놈 DNA는 cfDNA와 병행하여 단편화되고 처리된다. (세라케어로부터 구입한 맞춤형 DNA 혼합물)
DNA 혼합물 2	25	폐암 유전자 변이체를 보유하는 게놈 DNA의 혼합물. 게놈 DNA는 cfDNA (ERBB2, TP53, EML4-Alk 융합 (융합), EGFR과 병행하여 단편화되고 처리된다. 사용되는 세포주 DNA에는 하기 세포주로부터의 DNA가 포함된다: NA12878, PC-3 및 H2228.
질환 샘플 1		야생형 cfDNA + DNA 혼합물 1
질환 샘플 2		야생형 cfDNA + DNA 혼합물 2

[0504]

[0505] 건강한 공여자로부터의 무세포 DNA는 QIAmp DSP 순환 NA 키트 (키아젠)를 사용하여 혈장 샘플 (표 1 참조)로부터 추출되었다.

[0506] 실험실-생성된 질환 샘플 1 및 질환 샘플 2을 사용하는 이점은 조성물이 하기에 상세하게 기재된 바와 같이 주의깊게 제어될 수 있고, 샘플 이용가능성이 본질적으로 제한되지 않는다는 것이다.

[0507] 게놈 DNA는 200 bp 단편을 생성하기 위한 설정에서 울트라 초음파 처리기 (코바리스®)를 사용하여 초음파 처리에 의해 전단된 다음, 상자성 앰플이어 XP® 비드 (베크만®)에 의해 "양면" 비드 정제를 이용하여 추가로 정제되

고 크기 선택되었다.

[0508] 단편화된 세포주 게놈 DNA 및 합성 DNA의 혼합물을 WT cfDNA와 조합하여, 공지된 단일 뉴클레오타이드 변이체 (SNV), 삽입 및 결실 (Indel 변이체), 카피 수 변이체 (CNV), 및 정의된 대립유전자 빈도 (AF)의 융합을 갖는 질환 샘플 1 및 질환 샘플 2를 생성하였다. 상기 표에 나열된 투입 샘플량의 적절한 조합을 정의된 백분율로 블렌딩하여, 낮은 대립유전자 빈도 (AF)의 검출을 가능하게 하고, 말단 복구하고, 하기 기재된 바와 같이 태그 부착된 DNA 라이브러리로 전환시켰다.

[0509] 실시예 2: 임의적인 단일-단계 DNA 말단-복구

[0510] 말단 복구 DNA 단편이 단일 반응 혼합물에서 5' 포스페이트 기 및 3' dA 뉴클레오타이드 오버행을 갖도록 투입 DNA 단편을 "말단 복구 DNA 단편"으로 전환시켰다 (단일-단계 말단 복구).

[0511] 상업적으로 입수가 가능한 키트 (NEB 울트라 II 엔드 리페어®/dA 꼬리 부착 모듈 (E7546L)를 사용하여, DNA 단편을 말단 복구하였다. 말단 복구 마스터 믹스® ("말단 복구 MM")를 단일 튜브 반응 혼합물에서 연장된 DNA 단편에 첨가하였다. 말단 복구 MM은 NEB넥스트 울트라 II 엔드 프랩 엔자임 믹스(NEBNext Ultra II End Prep Enzyme Mix)®와 NEB넥스트 울트라 II 엔드 프랩 리액션 버퍼(NEBNext Ultra II End Prep Reaction Buffer)®를 조합함으로써 제조하였고, 각각의 혼합물 또는 완충제는 NEB넥스트 울트라 II 엔드 프랩/dA-꼬리 부착 모듈 (뉴 잉글랜드 바이오랩스®)의 성분이다. 연장된 DNA 단편을 함유하는 반응 혼합물은 열순환기에서 하기 반응 조건하에 인큐베이션하였다: 20℃에서 15분 동안에 이어서 70℃에서 10분 동안 ("단일 단계 반응").

[0512] 일부 실시양태에서, 말단-복구/dA-꼬리 부착 단계는 단일 단계 반응이 이러한 반응을 수행하기 위한 제조자의 권장량에 비해 유의하게 적은 양의 말단 복구 마스터 믹스 (MM)를 사용하도록 최적화되었다. 일부 실시양태에서, 말단 복구 MM의 양의 감소는 또한 놀랍게도 본 개시내용의 어댑터 및 관찰된 생성된 NGS 라이브러리의 게놈 등가물을 사용하여 말단 복구 DNA 단편의 클로닝 효율성에 의해 입증되는 바와 같이 말단 복구 DNA 단편 (평균 > 3500 GEs)의 형성에 대해 불리한 영향을 미치지 않았다. 실제로, 놀랍게도, 클로닝 효율은 본 개시내용에 기재된 바와 같이 이 단일 단계 말단 복구 과정을 이용하여 증가되었다.

[0513] 실시예 3: 어댑터 라이게이션

[0514] 다기능성 어댑터 모듈의 3' dT-꼬리 부착 라이게이션 가닥의 풀을 상기 샘플로부터의 말단 복구 DNA 단편에 라이게이션하여, 단편의 5' 말단에 어댑터를 부착하였다. 상보성 비-라이게이션 가닥은 DNA 단편의 3' dA 꼬리 부착 말단에 라이게이션되지 않았다. 이 실험에서 사용된 어댑터에 대한 설명은 표 2 및 표 3에 제공된다.

[0515] 45uL의 말단 복구 반응 혼합물 (5' 인산화 말단 및 3'dA 뉴클레오타이드 오버행을 갖는 말단 복구 DNA 단편을 함유함)을 5.0 μL의 고유의 다기능성 어댑터 모듈의 풀 (5 μM) 및 30 μL의 NEB 울트라 II 라이게이션 믹스(NEB Ultra II Ligation Mix)® (뉴 잉글랜드 바이오랩스®, 미국 메사추세츠주)에 첨가하였다. 어댑터 모듈의 각각의 라이게이션 가닥은 47 nt 길이였고, (5' → 3') 증폭 영역 (AMP, 25 nt), 샘플 및 고유의 단편 둘 다를 식별할 수 있는 다기능성 ID 영역 (8 nt), UMI 멀티플라이어 (3 nt), 앵커 (10 nt), 및 3' dT 오버행을 포함한다. 이 실시예에서 사용된 라이게이션 가닥은 표 2에 제공된다. 어댑터 모듈의 풀은 각각의 어댑터 풀이 4가지 유형의 앵커 영역을 포함하는 등몰량의 어댑터 모듈을 함유하도록 제조되었고, 각각의 앵커 유형은 A, T, C, 및 G로부터 선택된 3' 말단 뉴클레오타이드를 갖는다.

[0516] 표 2: 어댑터 구조

어댑터 명칭	설명/서열
라이제이션 가닥/ 앵커 영역 1 (16-1)	AMP-ID 영역/UMI 멀티플라이어- ACGTATGCCA (SEQ ID NO: 2)-3'dT
라이제이션 가닥/ 앵커 영역 2 (16-2)	AMP-ID 영역/UMI 멀티플라이어- CTAGCGTTAC (SEQ ID NO: 3)-3'dT
라이제이션 가닥/ 앵커 영역 3 (16-3)	AMP-ID 영역/UMI 멀티플라이어- GATCGACATG (SEQ ID NO: 4)-3'dT
라이제이션 가닥/ 앵커 영역 4 (16-4)	AMP-ID 영역/UMI 멀티플라이어- TGCATCAGGT (SEQ ID NO: 5)-3'dT
비-라이제이션 가닥/앵커 영역 1 (16_1)	<u>TGGCATACGT</u> (SEQ ID NO: 6)
비-라이제이션 가닥/앵커 영역 2 (16_2)	<u>GTAACGCTAG</u> (SEQ ID NO: 7)
비-라이제이션 가닥/앵커 영역 3 (16_3)	<u>CATGTCGATC</u> (SEQ ID NO: 8)
비-라이제이션 가닥/앵커 영역 4 (16_4)	<u>ACCTGATGCA</u> (SEQ ID NO: 9)

[0517]

[0518] 반응 혼합물을 20℃에서 30분 동안 인큐베이션하여, 어댑터-태그 부착된 DNA 단편을 생성하였다.

[0519] 표 3. 증폭되지 않은 및 증폭된 태그 부착된 DNA 라이브러리를 제조하기 위해 사용된 예시적인 어댑터 서열

3' dT 오버행을 갖는 어댑터 모듈 라이게이션 가닥	SEQ ID NO
TGCAGGACCAGAGAATTCGAATACAAAAATCCTNNNACGTATGCCAT	10
TGCAGGACCAGAGAATTCGAATACAAATGATCTNNNACGTATGCCAT	11
TGCAGGACCAGAGAATTCGAATACAAGTAATAGNNNACGTATGCCAT	12
TGCAGGACCAGAGAATTCGAATACACACCTCCGNNNACGTATGCCAT	13
TGCAGGACCAGAGAATTCGAATACACGCCCCATNNNACGTATGCCAT	14
TGCAGGACCAGAGAATTCGAATACACTACCAAGNNNACGTATGCCAT	15
TGCAGGACCAGAGAATTCGAATACACTGTCGTTNNNACGTATGCCAT	16
TGCAGGACCAGAGAATTCGAATACAGCAAATGGNNNACGTATGCCAT	17
TGCAGGACCAGAGAATTCGAATACAGCTCGAGCNNNACGTATGCCAT	18
TGCAGGACCAGAGAATTCGAATACAGTCCACAANNNACGTATGCCAT	19
TGCAGGACCAGAGAATTCGAATACAGTTACCCTNNNACGTATGCCAT	20
TGCAGGACCAGAGAATTCGAATACATAGTTTTCNNNACGTATGCCAT	21
TGCAGGACCAGAGAATTCGAATACATCTCAGAGNNNACGTATGCCAT	22
TGCAGGACCAGAGAATTCGAATACATGACCTTCNNNACGTATGCCAT	23
TGCAGGACCAGAGAATTCGAATACATTACGGCANNNACGTATGCCAT	24
TGCAGGACCAGAGAATTCGAATACAAACAAAACNNNTGCATCAGGTT	25
TGCAGGACCAGAGAATTCGAATACAACACTGCANNNTGCATCAGGTT	26
TGCAGGACCAGAGAATTCGAATACAATCGCGATNNNTGCATCAGGTT	27
TGCAGGACCAGAGAATTCGAATACAATGGTGGANNNTGCATCAGGTT	28
TGCAGGACCAGAGAATTCGAATACACAACCTCTCNNNTGCATCAGGTT	29
TGCAGGACCAGAGAATTCGAATACACGCCCCGAANNNTGCATCAGGTT	30
TGCAGGACCAGAGAATTCGAATACACGTATGACNNNTGCATCAGGTT	31
TGCAGGACCAGAGAATTCGAATACAGAAACGACNNNTGCATCAGGTT	32
TGCAGGACCAGAGAATTCGAATACAGACTCTGANNNTGCATCAGGTT	33
TGCAGGACCAGAGAATTCGAATACAGTCACTCTNNNTGCATCAGGTT	34
TGCAGGACCAGAGAATTCGAATACATACTGGACNNNTGCATCAGGTT	35
TGCAGGACCAGAGAATTCGAATACATGCGATACNNNTGCATCAGGTT	36
TGCAGGACCAGAGAATTCGAATACATGTTAATGNNNTGCATCAGGTT	37

[0520]

TGCAGGACCAGAGAATTCTGAATACATTGTACTTNNNTGCATCAGGTT	38
TGCAGGACCAGAGAATTCTGAATACATTTGGCTCNNNTGCATCAGGTT	39
TGCAGGACCAGAGAATTCTGAATACAAACGCCTANNNGATCGACATGT	40
TGCAGGACCAGAGAATTCTGAATACAAAGTTTCANNNGATCGACATGT	41
TGCAGGACCAGAGAATTCTGAATACAACAGCGAANNNGATCGACATGT	42
TGCAGGACCAGAGAATTCTGAATACAAGCGCCTGNNNGATCGACATGT	43
TGCAGGACCAGAGAATTCTGAATACACAACCCTTNNNGATCGACATGT	44
TGCAGGACCAGAGAATTCTGAATACACAGAATAANNNGATCGACATGT	45
TGCAGGACCAGAGAATTCTGAATACACGGACACCNNNGATCGACATGT	46
TGCAGGACCAGAGAATTCTGAATACAGCCTATTCNNNGATCGACATGT	47
TGCAGGACCAGAGAATTCTGAATACAGCGTCCAGNNNGATCGACATGT	48
TGCAGGACCAGAGAATTCTGAATACAGGTACAAGNNNGATCGACATGT	49
TGCAGGACCAGAGAATTCTGAATACATAACCCTCNNNGATCGACATGT	50
TGCAGGACCAGAGAATTCTGAATACATAGGAGTGNNNGATCGACATGT	51
TGCAGGACCAGAGAATTCTGAATACATCCGCATTNNNGATCGACATGT	52
TGCAGGACCAGAGAATTCTGAATACATGCGTCAANNNGATCGACATGT	53
TGCAGGACCAGAGAATTCTGAATACATTGGTAATNNNGATCGACATGT	54
TGCAGGACCAGAGAATTCTGAATACAAATAGCTTNNNCTAGCGTTACT	55
TGCAGGACCAGAGAATTCTGAATACAAGAGAGAGNNNCTAGCGTTACT	56
TGCAGGACCAGAGAATTCTGAATACACAACCTGANNNCTAGCGTTACT	57
TGCAGGACCAGAGAATTCTGAATACACATATGGCNNNCTAGCGTTACT	58
TGCAGGACCAGAGAATTCTGAATACACCATATCCNNNCTAGCGTTACT	59
TGCAGGACCAGAGAATTCTGAATACACGAGGTCCNNNCTAGCGTTACT	60
TGCAGGACCAGAGAATTCTGAATACACGTCAATGNNNCTAGCGTTACT	61
TGCAGGACCAGAGAATTCTGAATACACTTATCATNNNCTAGCGTTACT	62
TGCAGGACCAGAGAATTCTGAATACAGCATTGACNNNCTAGCGTTACT	63
TGCAGGACCAGAGAATTCTGAATACAGGAGGTATNNNCTAGCGTTACT	64
TGCAGGACCAGAGAATTCTGAATACATAACAGTTNNNCTAGCGTTACT	65
TGCAGGACCAGAGAATTCTGAATACATCGAACACNNNCTAGCGTTACT	66
TGCAGGACCAGAGAATTCTGAATACATGCATAATNNNCTAGCGTTACT	67
TGCAGGACCAGAGAATTCTGAATACATGTCATAANNNCTAGCGTTACT	68
TGCAGGACCAGAGAATTCTGAATACATTGCGCGNNNCTAGCGTTACT	69

[0521]

[0522] * 표 3의 서열에서 NNN은 3개 뉴클레오티드 UMI 멀티플라이어를 나타내고, 각각의 N은 A, G, C, T 중 어느 하나로부터 선택될 수 있다.

[0523] 라이게이션 후, 100 μ L의 DNA 정제 비드 (엠펙어 XP®; 베크만®)를 라이게이션 믹스에 첨가하였다. 반응 혼합물을 실온에서 2분 동안 인큐베이션하였다. 비드를 자석 위에서 200 μ L의 80% 에탄올/물 (v/v)로 2회 세척하고, 공기 건조한 다음, 25 μ L의 TRIS-EDTA (TEZ)로 용리시켰다. 어댑터-태그 부착된 DNA 단편을 함유하는 용리되고 정화된 상청액 약 25 μ L를 증폭을 위해 신선한 PCR 튜브 또는 미세역가 플레이트 웰로 옮겨서, 어댑터-태그 부착된 DNA 라이브러리를 생성하였다.

[0524] 실시예 4: 태그 부착된 DNA 라이브러리 연장 및 증폭

[0525] 실시예 3에서 어댑터 라이게이션 후, 시약 및 호열성 DNA 폴리머라제 효소 (NEB 울트라 II 2X PCR 앰플리피케이션®; 뉴 잉글랜드 바이오랩스®)를 함유하는 75 μ L의 마스터 믹스 (MM)를 첨가하고, 하기 실행 파라미터를 이용하여 반응 혼합물을 증폭시켰다:

[0526] 60°C에서 30초 동안, 72°C에서 2분 동안, 98°C에서 30초 동안;

[0527] 8 주기의 98°C에서 30초 동안, 65°C에서 30초 동안 및 72°C에서 30초 동안.

[0528] 단일 증폭 프라이머를 사용하였다: TGCAGGACCAGAGAATTCTGAATACA (서열식별번호: 70).

- [0529] 초기 3분 인큐베이션 주기를 수행하여, 라이게이션-가닥-주형 연장에 의해 복수개의 연속된 어댑터-태그 부착된 dsDNA 단편을 형성한 후, 연속된 어댑터-태그 부착된 dsDNA 단편의 8 주기 PCR 증폭에 의해 어댑터 태그 부착된 DNA 단편 분자를 함유하는 증폭된 태그 부착된 DNA 라이브러리를 형성하였다.
- [0530] 증폭 후, 120 μ L의 DNA 정제 비드를 라이게이션 믹스에 첨가하였다. 반응 혼합물을 실온에서 2분 동안 인큐베이션하였다. 비드를 자석 위에서 200 μ L의 80% 에탄올/물 (v/v)로 2회 세척하고, 공기 건조하고, 14 μ L의 TEZ로 용리시켰다. 증폭된 태그 부착된 DNA 라이브러리를 함유하는 정화된 상청액을 신선한 PCR 튜브로 옮겼다.
- [0531] 이들 방법을 이용하여, 본 발명자들은 고효율성 라이브러리 제조 시간을 약 8 시간에서 약 3 내지 4 시간까지 감소시켰다. 실시예에서 사용된 효소를 비롯하여 효소/키트 시약의 수가 16개에서 4개로 감소함에 따라 복잡성이 감소되었다.
- [0532] 실시예 5: 포획 프로브 라이브러리 증폭 (예언적)
- [0533] 관심 유전자좌를 포획하고 농축시키기 위해, 실시예 2에 기재된 바와 같이 제조된 각각의 태그 부착된 DNA 라이브러리 (예를 들어 질환 샘플 1 및 질환 샘플 2)를 조합하고, 멀티플렉스화하고, 상동성 복구 결핍 유전자 (예를 들어 ATM, BRCA1, BRCA2, FANCA, HDAC2, 및 PALB2)에 대해 특이적인 다기능성 포획 프로브 모듈의 풀 또는 폐암 유전자 (예를 들어 ERBB2, TP53, EML4-ALK 융합, EGFR, MET)에 대해 특이적인 다기능성 포획 프로브 모듈의 풀에 혼성화한다.
- [0534] 다음으로, 100 μ L의 스트렙타비딘-코팅된 비드 (디나비즈 마이원 C1(Dynabeads MyOne C1))를 혼성화 반응과 조합하고, 실온에서 20분 동안 정치시킨다. 비드를 자석 위에서 수집하고, 200 μ L의 TEZ 완충제로 1회 세척한다. 세척된 비드를 40 μ L의 TEZ 완충제에서 재현탁시킨다. 160 μ L의 세척 완충제를 재현탁된 비드에 첨가하고, 혼합물을 45 $^{\circ}$ C에서 5분 동안 인큐베이션한다. 이어서, 자석을 사용하여 비드를 분리시키고, 200 μ L의 TEZ 완충제로 세척한다.
- [0535] 혼성화 후, 포획 프로브의 프라이머 연장을 이용하여, 포획된 게놈 서열, DNA 단편의 접합부에서 A/T 오버행, 및 부착된 어댑터 모듈을 카피하여, 혼성체 분자의 라이브러리를 형성한다. 이렇게 형성된 혼성체 분자는 한 말단에서는 포획 프로브 모듈 및 다른 말단에서는 어댑터 모듈에 의해 플랭킹된 DNA 단편을 포함한다.
- [0536] 온-비드 프로브 연장 후, PCR을 수행하여 일루미나® 시퀀싱 어댑터를 통합시킨다. 비드를 20 μ L의 TEZ에 재현탁시키고, PCR 마스터 믹스 (55 μ L의 울트라 II PCR 믹스, 5.5 μ L의 프라이머 F, 5.5 μ L의 프라이머 R)와 조합하고, 열순환기 상에 놓고, 하기 프로그램에 따라 실행한다: 60 $^{\circ}$ C에서 30초 동안; 72 $^{\circ}$ C에서 30초 동안; 98 $^{\circ}$ C에서 30초 동안; 5 주기의 98 $^{\circ}$ C에서 30초 동안; 65 $^{\circ}$ C에서 30초 동안; 72 $^{\circ}$ C에서 30초 동안.
- [0537] 다음으로, 비드를 자석 위에서 반응 혼합물로부터 분리한다. 상청액을 신선한 PCR 튜브로 옮기고, PCR MM과 조합하고, 하기 증폭 주기를 이용하여 열순환기 상에서 증폭시킨다: 10 주기의 98 $^{\circ}$ C에서 30초 동안; 65 $^{\circ}$ C에서 30초 동안; 72 $^{\circ}$ C에서 30초 동안.
- [0538] 정방향 프라이머: AATGATACGGCGACACCAGATCTACACGTCATGCAGGACCAGAGAATTTCGAATACA (서열식별번호: 71).
- [0539] 역방향 프라이머: CAAGCAGAAGACGGCATACGAGATGTGACTGGCACGGGAGTTGATCCTGGTTTTCAC (서열식별번호: 72).
- [0540] 혼성체 분자의 이 라이브러리를 PCR 증폭시켜, 각각의 샘플에 대해 증폭된 혼성체 분자를 함유하는 증폭되고 표적화된 DNA 라이브러리를 제공한다. 이들 증폭된 혼성체 분자는 도 8에 도시된 바와 같이 분자의 양 말단에 시퀀싱 프라이머 결합 부위를 함유한다는 점에서 "시퀀싱 준비된" 것이다.
- [0541] 실시예 6: 서열 분석
- [0542] 유전자 분석을 혼성체 분자에 대해 수행하였다. 시퀀싱 리드 1 (151 nt) 및 리드 2 (17 nt)를 유전자 분석을 위해 사용하였다. 적절한 클러스터 및 정렬 분석을 위해, 각각의 개별 서열 리드를 처리하여 A/T 뉴클레오타이드 삽입 (어댑터의 3' 말단 오버행 및 DNA 단편의 5' 말단 오버행으로부터 생성됨)을 생물정보학적으로 제외시켰다. A/T 삽입의 이러한 제외는 서열 리드를 생물정보학 방법을 이용하는 유전자 분석에 적용함으로써 수행되었다. 변이체 호출기는 중복 리드를 식별하고, 중복 리드를 단일 컨센서스 리드로 처리하였고, 이어서 이는 각각의 프로브 위치에서 정량화되었다. 변이체 호출기는 어댑터의 접합부 및 각각의 DNA 단편의 5' 말단을 추가로 식별하여, 삽입된 A/T 오버행을 생물정보학으로 제외시켜, 적절한 샘플-특이적인 DNA 단편 서열을 수득하였다. 유전자 분석 동안에 A/T 삽입의 제외는 품질을 증가시키고, 서열 리드의 정렬 오류 및/또는 부정확

한 클러스터화를 감소시켰다. 최종적으로, 통계적 유의성은 각각의 변수 측정에서 검출된 편차에 배정되었다.

실시예 7: 시퀀싱 깊이에 대한 개선

시퀀싱은 상기 생성된 태그 부착된 DNA 라이브러리에 대해 수행되었다. 태그 부착된 DNA 라이브러리를 인간 기준 게놈에 대해 정렬시키고, 의도된 표적에 대해 맵핑하였다.

비교기 과정 및 자동화가능한 과정을 이용하여 3개의 태그 부착된 DNA 라이브러리 (WT cfDNA, 질환 샘플 1, 및 질환 샘플 2)의 평균 깊이를 측정하였다 (도 4c 참조).

실시예 8: 균일한 어댑터 분포

본 실시예에서는, 서열 리드에 의해 측정되는 바와 같이 태그 부착된 DNA 라이브러리에서 특정 어댑터 서열의 포함에 대한 편향이 감소되었다. 자동화가능한 과정을 이용하는 라이브러리 제조는 비교기 과정을 이용하는 라이브러리 제조와 비교하여 개선된 어댑터 분포를 나타내었고, 이는 덜 효율적인 어댑터에 대한 보상의 필요성을 제거하였다. 생성된 앵커 분포는 도 9 및 표 4에 제시된다.

표 4: 어댑터 분포에 대한 라이브러리 제조 비교

	앵커 서열 분포	
	비교기 과정 % 분포	자동화가능한 과정 % 분포
16-1	50%	28%
16-2	14%	23%
16-3	9%	34%
16-4	27%	14%

실시예 9: 변이체 검출

표 6 및 7은 실시예 1에서 제조된 샘플 (질환 샘플 1 및 질환 샘플 2)에 대해 생성된 왓슨 (+) 또는 크릭 가닥 (-) 상의 각각의 변이체에 대한 시퀀싱 리드의 수를 나타낸다. 표 5에서 알 수 있는 바와 같이, 질환 샘플 1에 대해, 자동화가능한 과정을 이용하는 경우 변이체 + 가닥 (+ 가닥)에 대한 평균 리드는 94개인 반면에, 비교기 과정을 이용하는 경우에는 66개였다. 유사하게, 질환 샘플 2에 대해 자동화가능한 과정을 이용하는 경우 변이체 + 가닥 (+ 가닥)에 대한 평균 리드는 238개인 반면에, 질환 샘플 1에 대해 비교기 과정을 이용하는 경우에는 199개였다. 이들 결과는 각각의 변이체에 대해 증가된 수의 리드에 의해 측정될 수 있는 바와 같이 시험되고 검출된 각각의 변이체에 대해 태그 부착된 DNA 라이브러리 및 프로브 포획 라이브러리를 제조하는 과정이 훨씬 더 효율적이었음을 시사하고, 검정 민감도의 증가를 나타낸다.

표 5: 질환 샘플 1로부터 제조된 라이브러리를 사용하여 자동화가능한 과정과 비교기 과정 사이의 변이체 검출 비교

질환 샘플 1: 자동화가능한 과정						
	유형	돌연변이	WT(+)	WT(-)	Var(+)	Var(-)
Rep 1	Indel	ATM 프레임시프트	1784	2328	78	86
	Indel	BRCA1 프레임시프트	2762	2446	167	80
	Indel	BRCA2 프레임시프트	1084	3630	58	94
	SNV	BRCA2 G4*	1178	3016	85	117
	SNV	FANCA 스플라이스	2313	2944	125	123
	Indel	HDAC2 프레임시프트	2170	2380	76	101
	SNV	PALB2 Q479	2464	3079	110	186
Rep 2	Indel	ATM 프레임시프트	1672	2156	89	87
	Indel	BRCA1 프레임시프트	2608	2321	126	83
	Indel	BRCA2 프레임시프트	1130	3346	53	151
	SNV	BRCA2 G4*	1143	2763	73	105
	SNV	FANCA 스플라이스	2292	2980	127	123
	Indel	HDAC2 프레임시프트	1967	2145	79	86
	SNV	PALB2 Q479	2358	2925	70	129
	Avg		1923	2747	94	111
질환 샘플 1: 비교기 과정						
	유형	돌연변이	WT(+)	WT(-)	Var(+)	Var(-)
Rep 1	Indel	ATM 프레임시프트	1090	1839	49	86
	Indel	BRCA1 프레임시프트	1816	1652	71	60

[0553]

	Indel	BRCA2 프레임시프트	531	2768	40	117
	SNV	BRCA2 G4*	354	1884	32	98
	SNV	FANCA 스플라이스	2327	2386	115	91
	Indel	HDAC2 프레임시프트	1174	1318	82	54
	SNV	PALB2 Q479	1614	2198	65	64
Rep 2	Indel	ATM 프레임시프트	1139	1874	65	109
	Indel	BRCA1 프레임시프트	1974	1650	66	63
	Indel	BRCA2 프레임시프트	616	2913	32	102
	SNV	BRCA2 G4*	427	1980	45	93
	SNV	FANCA 스플라이스	2484	2628	86	116
	Indel	HDAC2 프레임시프트	1723	1381	104	70
	SNV	PALB2 Q479	1756	2378	65	102
	Avg		1402	2061	66	88

표 6: 질환 샘플 2로부터 제조된 라이브러리를 사용하여 자동화가능한 과정과 비교기 과정 사이의 변이체 검출 비교

질환 샘플 2: 자동화가능한 과정							
	유형	돌연변이	WT(+)	WT(-)	Var(+)	Var(-)	점수
Rep 1	SNV	ERBB2 1655V	1453	1483	419	472	
	SNV	TP53 Q331	2040	1679	270	216	
	Indel	TP53 프레임시프트	1197	2124	73	111	
	융합	EML4-ALK 융합	3457		171		
	CNV	EGFR 증폭					5.54E-17
Rep 2	SNV	ERBB2 1655V	1414	1422	491	534	
	SNV	TP53 Q331	2126	1811	243	240	
	Indel	TP53 프레임시프트	1226	2198	76	137	
	융합	EML4-ALK 융합	3157		161		
	CNV	EGFR 증폭					8.08E-17
	Avg		2009	1786	238	285	
질환 샘플 2: 비교기 과정							
	유형	돌연변이	WT(+)	WT(-)	Var(+)	Var(-)	점수
Rep 1	SNV	ERBB2 1655V	1093	1197	347	385	
	SNV	TP53 Q331	1650	1252	213	157	
	Indel	TP53 프레임시프트	909	1921	66	107	
	융합	EML4-ALK 융합	2879		151		
	CNV	EGFR 증폭					2.89E-11
Rep 2	SNV	ERBB2 1655V	1134	1171	395	371	
	SNV	TP53 Q331	1667	1297	190	100	
	Indel	TP53 프레임시프트	933	1875	53	115	
	융합	EML4-ALK 융합	2894		174		
	CNV	EGFR 증폭					1.71E-11
	Avg		1645	1452	199	218	

[0556]

[0557]

실시예 10: 게놈 라이브러리의 증폭에 대한 개선

[0558]

게놈 라이브러리를 증폭시키기 위한 조건은 3가지 상이한 조건하에 시험되었다:

[0559]

1) 69℃의 어닐링 온도하에 2개의 별도의 튜브로 나뉘어진 라이브러리의 증폭을 수행하고;

[0560]

2) 65℃의 어닐링 온도하에 2개의 별도의 튜브로 나뉘어진 라이브러리의 증폭을 수행하고;

[0561]

3) 65℃의 어닐링 온도하에 라이브러리를 나누지 않고 (1개의 튜브에서) 증폭을 수행한다 (표 7).

[0562]

65℃의 어닐링 온도하에 라이브러리를 나누지 않고 수행된 증폭 조건 (조건 3)은 잘 수행되었고, 라이브러리를 2개의 샘플 튜브로 나눌 필요성을 제거하고, 라이브러리 제조 과정을 간소화시켰다.

[0563]

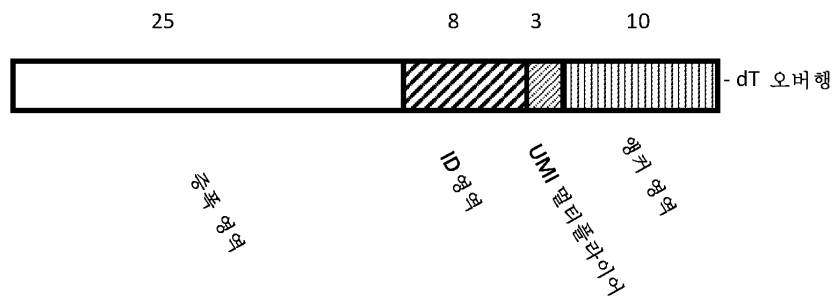
표 7: 증폭 조건의 최적화

	투입 GE	온 타겟 고유	온 타겟 Dup	오프 타겟	정렬되지 않은	총 리드	%오프 타겟	% 정렬되지 않은
2개의	1000	1655079	8376345	275174	72971	10379569	2.7	0.7
튜브,	4000	6192870	27466258	992782	282263	34934173	2.8	0.8
69 °C	10000	12209458	31194609	1248334	369438	45021839	2.8	0.8
어닐링	20000	18387168	30512373	1388364	415609	50703514	2.7	0.8
2개의	1000	1548719	5810708	191047	42991	7593465	2.5	0.6
튜브,	4000	5754578	19615062	712530	170191	26252361	2.7	0.6
65°	10000	13604693	31733716	1134323	298200	46770932	2.4	0.6
어닐링	20000	23349605	32085921	1277457	354548	57067531	2.2	0.6
1개의	1000	1727578	10688885	326101	68866	12811430	2.5	0.5
튜브,	4000	6517578	28182111	882117	202677	35784483	2.5	0.6
65°	10000	12933835	29309214	976804	248028	43467881	2.2	0.6
어닐링	20000	20496827	29537398	1093181	290146	51417552	2.1	0.6

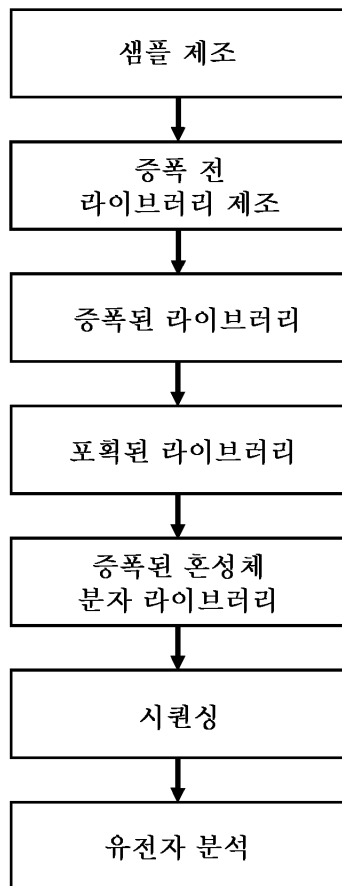
[0564]

도면

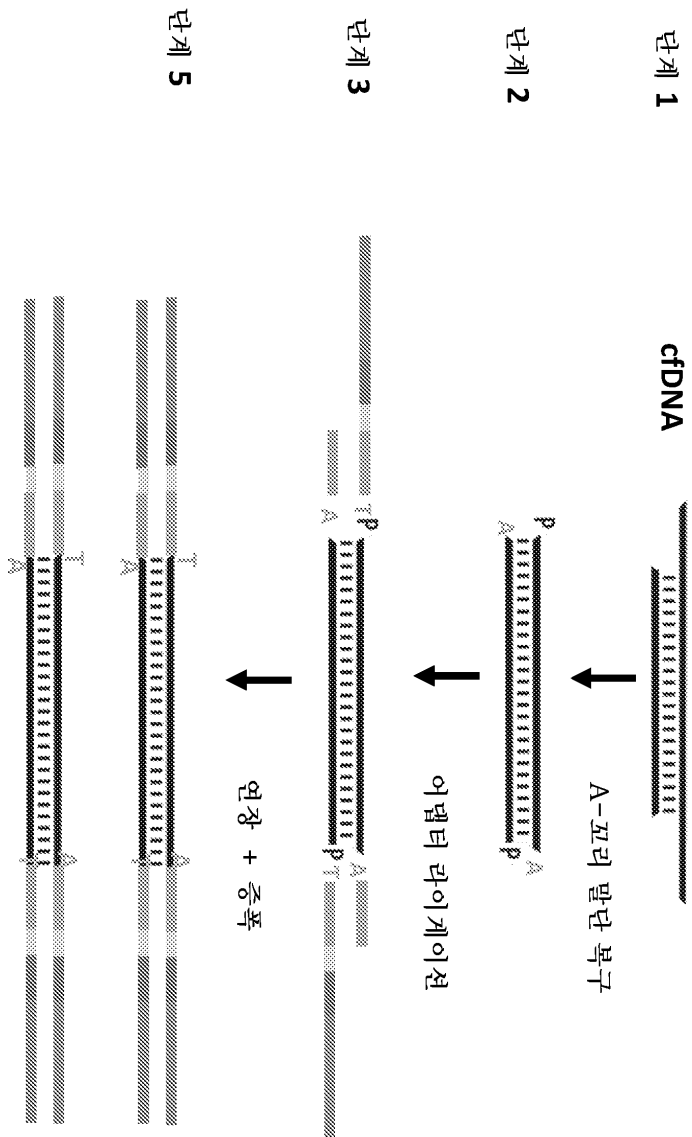
도면1



도면2



도면3



도면4a

비교기 과정

라이브러리를 생성하기 위한 시간: ~~~~~

96 웰 플레이트당 48가지 반응
PCR 전에 16가지 시약

1단계

① 무세포 DNA를 웰로부터 분리한다



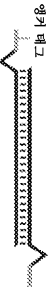
② 단편을 말인산화시킨다



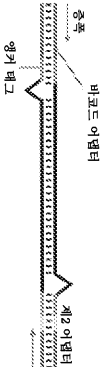
③ 말단을 평활하게 만든다



④ 앵커 태그 올리고를 각각의 단편의 양말단에 라이게이션시킨다



⑤ 혼성체 바코드 어댑터의 풀을 사용하여, 분자를 PCR 증폭될 수 있는 게놈 라이브러리 형태로 전환시킨다



단계

첨가

프로그램

말단 복구 1

완충제 5 uL
포스포타게 2.5 uL

37℃에서 15분
65℃에서 5분

말단 복구 2

Reg D: 0.5 uL
Rep Enz 1: 1 uL
Rep Enz 2: 0.5 uL

20℃에서 15분
70℃에서 10분

앵커 라이게이션

앵커: 20 uL
Lig Mix: 29 uL
Lig Reg: 1 uL

20℃에서 3시간
65℃에서 10분

정화

엘류아: 110 uL

바코드 라이게이션

바코드: 4 uL
풀: 10.1 uL
완충제 (Buff B로 B1): 4.4 uL
Reg N (Reg N으로 N1) 1.3 uL
Enz B1 E: (BBB로 B1 및 E2): 1.8 uL

37℃에서 15분
60℃에서 15분

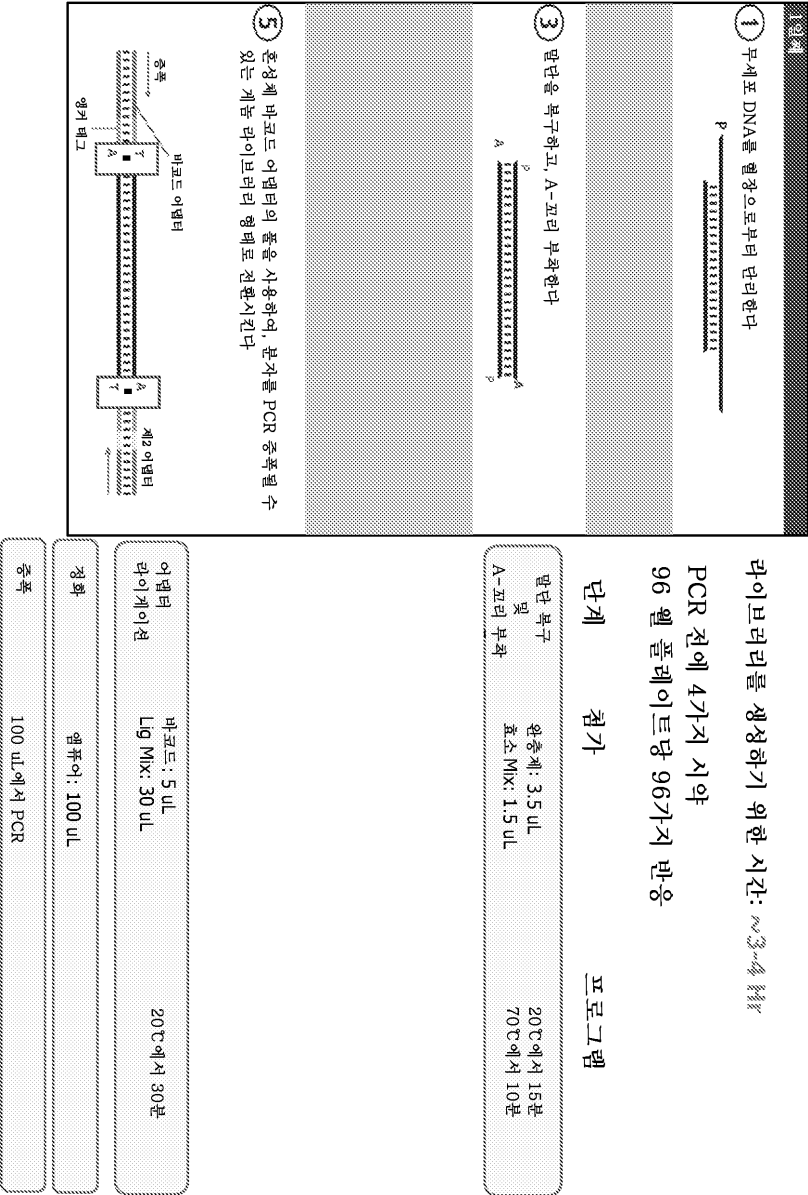
정화

엘류아: 60 uL

중복

220 uL에서 PCR

자동화가능한 과정

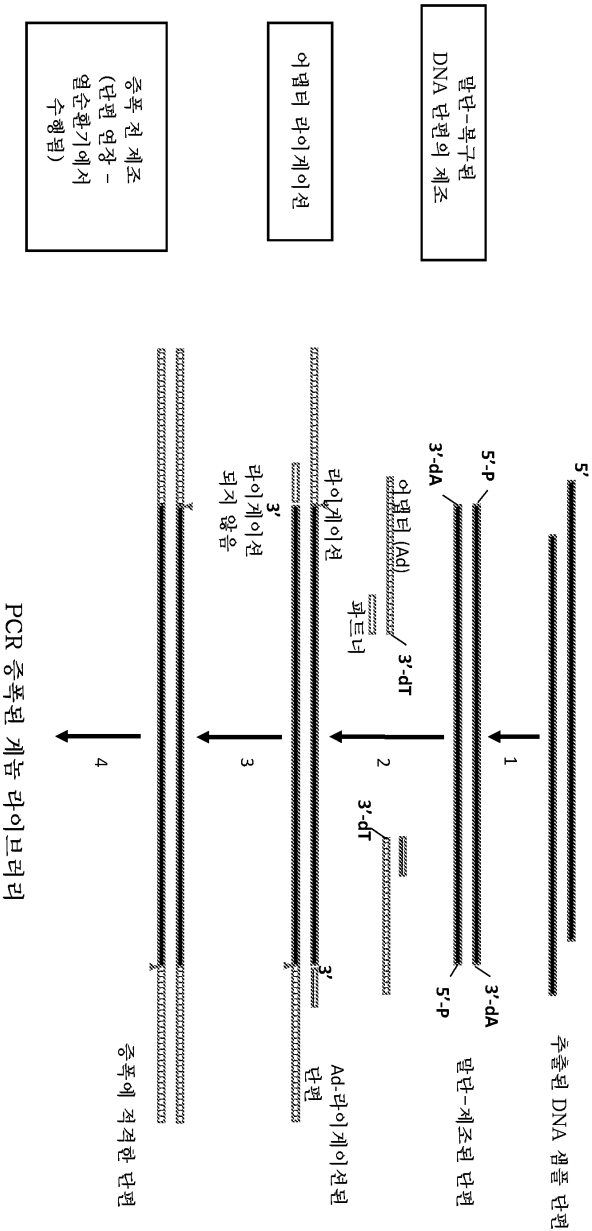


도면4b

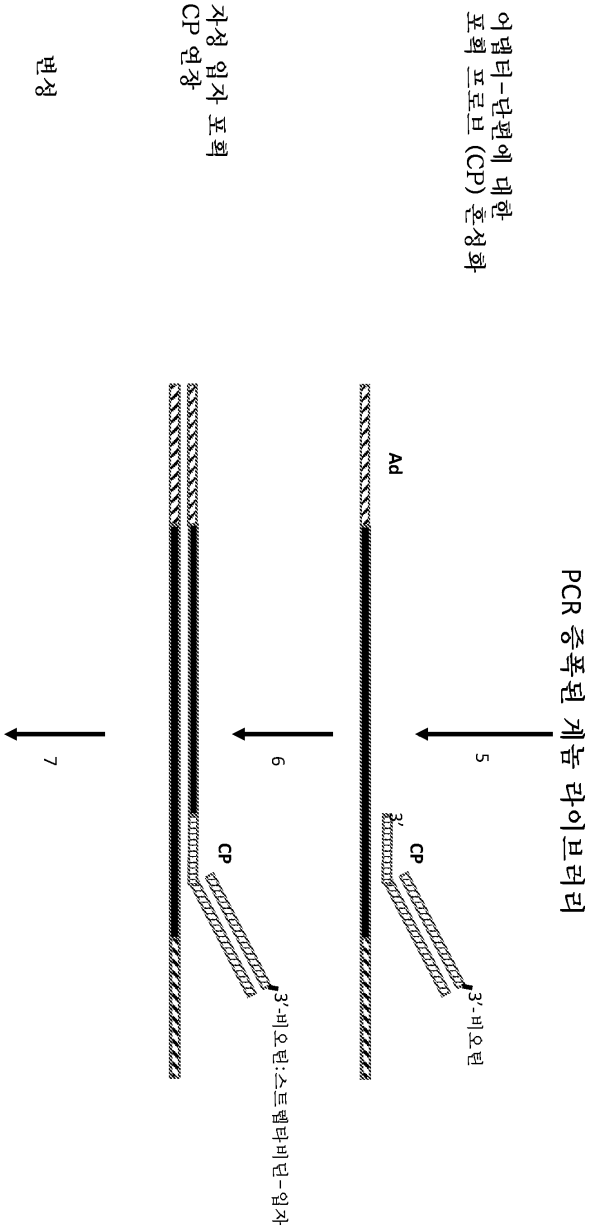
도면4c

WT cfDNA	비표기 길이	자동화가능한 길이	% 증가
양성 대조군 1	763	1,543	102%
양성 대조군 2	3,446	4,365	27%
양성 대조군 2	2,714	3,707	37%

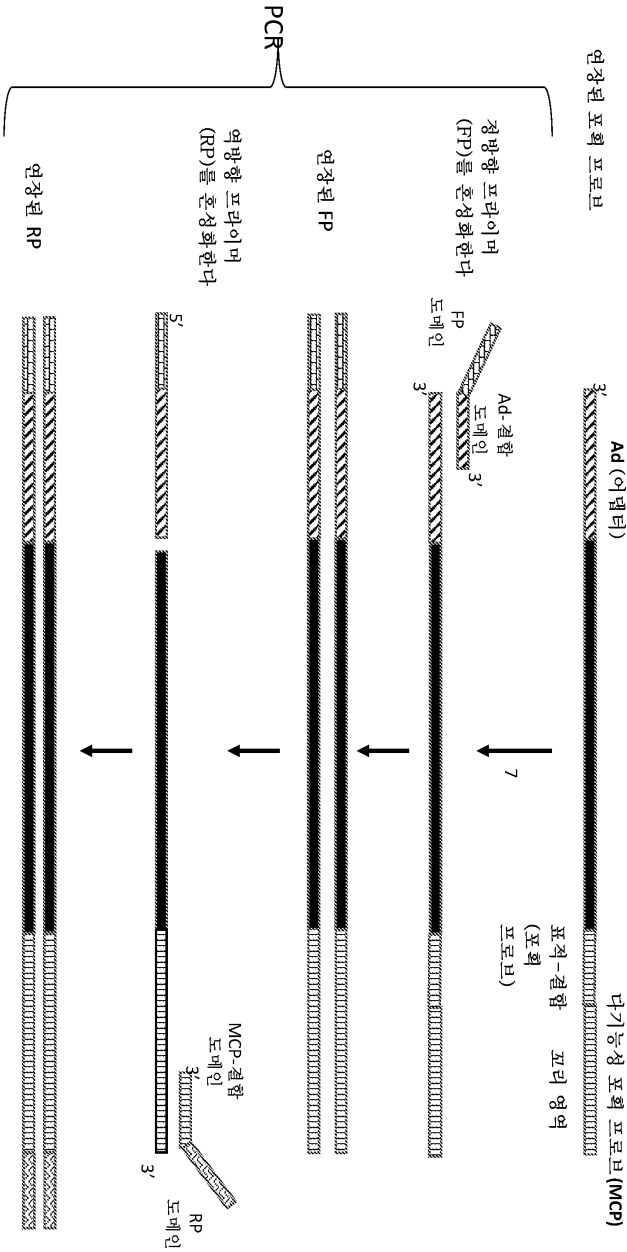
도면5



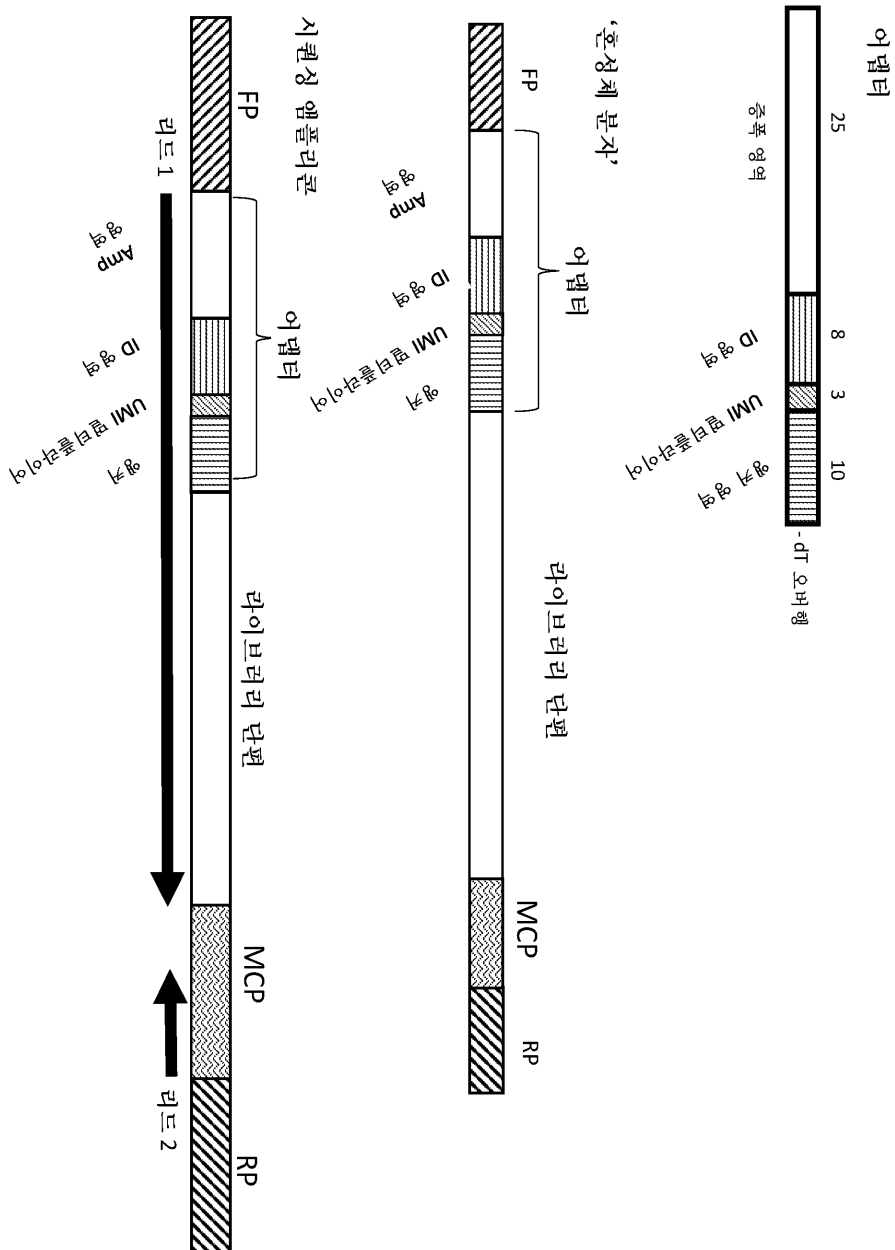
도면6



도면7



도면8



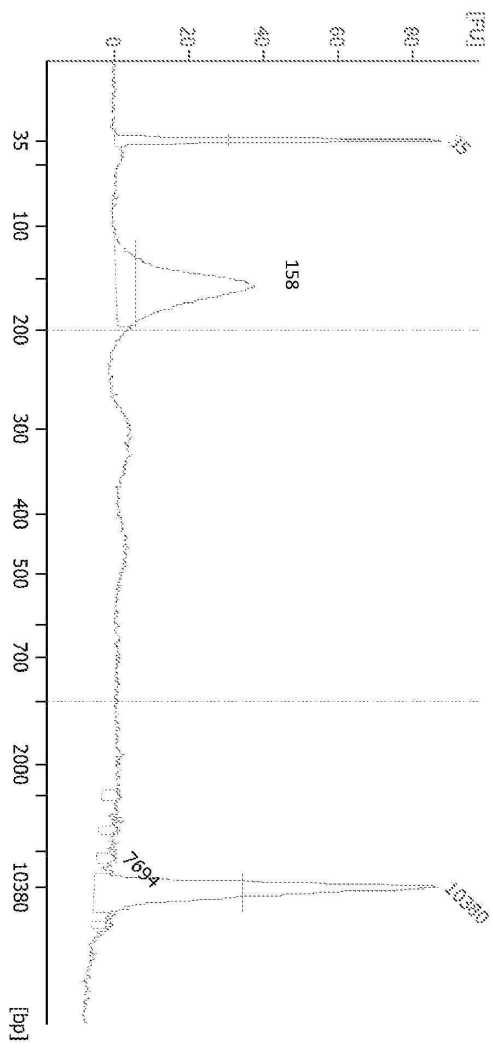
도면9

앵커 서열 분포		
	비교기	자동화가능한
3'A	50%	28%
3'C	14%	23%
3'G	9%	34%
3'T	27%	14%

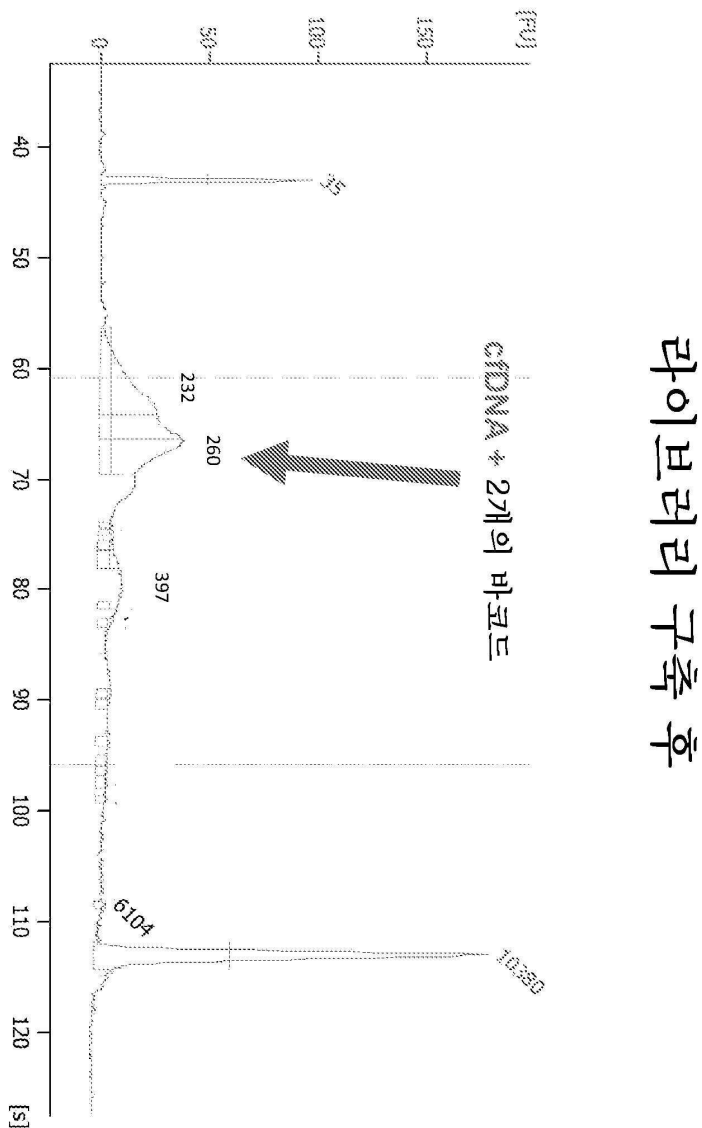
지정	앵커 서열
3'A	ACGTATGCCA (SEQ.ID NO: 2)
3'C	CTAGCGTTAC (SEQ.ID NO: 3)
3'G	GATCGACATG (SEQ.ID NO: 4)
3'T	TGCATCAGGT (SEQ.ID NO: 5)

투입 cDNA

도면10a



도면10b



서열 목록

SEQUENCE LISTING

<110> RESOLUTION BIOSCIENCE, INC.

<120> ADAPTORS AND METHODS FOR HIGH EFFICIENCY CONSTRUCTION OF GENETIC
LIBRARIES AND GENETIC ANALYSIS

<130> CLFK-006/01WO 318106-2088

<150> US 63/075,543

<151> 2020-09-08

<160> 74

<170> PatentIn version 3.5

<210> 1

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary partner oligonucleotide

<400> 1

gtgaaaacca ggatcaactc ccgtgccagt cacatctcag atgagct 47

<210> 2

<211>

> 10

<212> DNA

<213> Artificial Sequence

<220><223> ligation-anchor region 1

<400> 2

acgtatgccca 10

<210> 3

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> ligation-anchor region 2

<400> 3

ctagcgttac 10

<210> 4

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> ligation-anchor region 3

<400> 4

gatcgacatg 10

<210> 5

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> ligation-anchor region 4

<400> 5

tgcattcaggt	10
<210> 6	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> non-ligation-anchor region 1	
<400> 6	
tggcatacgt	10
<210> 7	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> non-ligation-anchor region 2	
<400> 7	
gtaacgctag	10
<210> 8	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> non-ligation-anchor region 3	
<400> 8	
catgtcgatc	10
<210> 9	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> non-ligation-anchor region 4	
<400> 9	
acctgatgca	10
<210> 10	
<211> 47	
<212> DNA	
<213> Artificial Sequence	
<220><223> exemplary ligation strand 1	

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 10

tgcaggacca gagaattcga atacaaaaat cctnnnacgt atgccat 47

<210> 11

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 2

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 11

tgcaggacca gagaattcga atacaaatga tctnnnacgt atgccat 47

<210> 12

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 3

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 12

tgcaggacca gagaattcga atacaagtaa tagnnnacgt atgccat 47

<210> 13

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 4

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 13
 tgcaggacca gagaattcga atacacacct ccgnnnacgt atgccat 47
 <210> 14
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223>
 > exemplary ligation strand 5
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 14
 tgcaggacca gagaattcga atacacgccc catnnnacgt atgccat 47
 <210> 15
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 6
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 15
 tgcaggacca gagaattcga atacactacc aagnnnacgt atgccat 47
 <210> 16
 <211> 47
 <212> DNA
 <213>
 Artificial Sequence
 <220><223> exemplary ligation strand 7
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 16
 tgcaggacca gagaattcga atacactgtc gttnnnacgt atgccat 47
 <210> 17

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 8
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 17
 tgcaggacca gagaattcga atacagcaaa tggnnnacgt atgccat 47
 <210> 18

 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 9
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 18
 tgcaggacca gagaattcga atacagctcg agcnnnacgt atgccat 47
 <210> 19
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 10
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 19
 tgcaggacca gagaattcga atacagtcca caannnacgt atgccat 47

 <210> 20
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220><223> exemplary ligation strand 11

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 20

tgcaggacca gagaattcga atacagttac cctnnnacgt atgccat 47

<210> 21

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 12

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 21

tgcaggacca gagaattcga atacatagtt ttcnnnacgt atgccat 47

<210> 22

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 13

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 22

tgcaggacca gagaattcga atacatctca gagnnnacgt atgccat 47

<210> 23

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 14

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 23

tgcaggacca gagaattcga atacatgacc ttcnnnacgt atgccat

47

<210> 24

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 15

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 24

tgcaggacca gagaattcga atacattacg gcannnacgt atgccat

47

<210> 25

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 16

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 25

tgcaggacca gagaattcga atacaacaa aacnnntgca tcaggtt

47

<210> 26

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 17

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 26

tgcaggacca gagaattcga atacaacact gcannntgca tcaggtt

47

<210> 27

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 18
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 27
 tgcaggacca gagaattcga atacaatcgc gatnnntgca tcaggtt 47

<210> 28
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 19
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 28
 tgcaggacca gagaattcga atacaatggt ggannntgca tcaggtt 47

<210> 29
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 20
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 29
 tgcaggacca gagaattcga atacacaact ctcnnntgca tcaggtt 47

<210> 30
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220><223> exemplary ligation strand 21

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 30

tgcaggacca gagaattcga atacacgccc gaannntgca tcaggtt 47

<210> 31

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 22

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 31

tgcaggacca gagaattcga atacacgtat gacnnntgca tcaggtt 47

<210> 32

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 23

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 32

tgcaggacca gagaattcga atacagaaac gacnnntgca tcaggtt 47

<210> 33

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 24

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 33

tgcaggacca gagaattcga atacagactc tgannntgca tcaggtt 47

<210> 34

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 25

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 34

tgcaggacca gagaattcga atacagtcac tctnnntgca tcaggtt 47

<210> 35

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 26

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 35

tgcaggacca gagaattcga atacatactg gacnnntgca tcaggtt 47

<210> 36

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 27

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 36

tgcaggacca gagaattcga atacatgcga tacnnntgca tcaggtt 47

<210> 37

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 28
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 37
 tgcaggacca gagaattcga atacatgtta atgnnntgca tcaggtt 47

<210> 38
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 29
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 38
 tgcaggacca gagaattcga atacattgta cttnnntgca tcaggtt 47

<210> 39
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 30
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 39
 tgcaggacca gagaattcga atacatttgg ctcnnntgca tcaggtt 47

<210> 40
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220><223> exemplary ligation strand 31

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 40

tgcaggacca gagaattcga atacaaacgc ctannngatc gacatgt 47

<210> 41

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 32

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 41

tgcaggacca gagaattcga atacaaagtt tcanngatc gacatgt 47

<210> 42

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 33

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 42

tgcaggacca gagaattcga atacaacagc gaannngatc gacatgt 47

<210> 43

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 34

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 43

tgcaggacca gagaattcga atacaagcgc ctgnnngatc gacatgt 47

<210> 44

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 35

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 44

tgcaggacca gagaattcga atacacaacc cttnnngatc gacatgt 47

<210> 45

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 36

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 45

tgcaggacca gagaattcga atacacagaa taannngatc gacatgt 47

<210> 46

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 37

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 46

tgcaggacca gagaattcga atacacggac accnnngatc gacatgt 47

<210> 47

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 38
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 47
 tgcaggacca gagaattcga atacagccta ttcnnngatc gacatgt 47

<210> 48
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 39
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 48
 tgcaggacca gagaattcga atacagcgtc cagnnngatc gacatgt 47

<210> 49
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 40
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 49
 tgcaggacca gagaattcga atacaggtac aagnnngatc gacatgt 47

<210> 50
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220><223> exemplary ligation strand 41

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 50

tgcaggacca gagaattcga atacataacc ctcnnngatc gacatgt 47

<210> 51

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 42

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 51

tgcaggacca gagaattcga atacatagga gtgnnngatc gacatgt 47

<210> 52

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 43

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 52

tgcaggacca gagaattcga atacatccgc attnnngatc gacatgt 47

<210> 53

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 44

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 53

tgcaggacca gagaattcga atacatgcgt caannngatc gacatgt

47

<210> 54

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 45

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 54

tgcaggacca gagaattcga atacattggt aatnnngatc gacatgt

47

<210> 55

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 46

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 55

tgcaggacca gagaattcga atacaatag cttnnnctag cgttact

47

<210> 56

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 47

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 56

tgcaggacca gagaattcga atacaagaga gagnnnctag cgttact

47

<210> 57

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 48
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 57
 tgcaggacca gagaattcga atacacaacc tgannnctag cgttact 47

<210> 58
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 49
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 58
 tgcaggacca gagaattcga atacacatat ggcnnnctag cgttact 47

<210> 59
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 50
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 59
 tgcaggacca gagaattcga atacaccata tccnnnctag cgttact 47

<210> 60
 <211> 47
 <212> DNA
 <213> Artificial Sequence

<220><223> exemplary ligation strand 51

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 60

tgcaggacca gagaattcga atacacgagg tccnnnctag cgttact 47

<210> 61

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 52

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 61

tgcaggacca gagaattcga atacacgtca atgnnnctag cgttact 47

<210> 62

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 53

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 62

tgcaggacca gagaattcga atacacttat catnnnctag cgttact 47

<210> 63

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 54

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 63

tgcaggacca gagaattcga atacagcatt gacnnnctag cgttact

47

<210> 64

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 55

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 64

tgcaggacca gagaattcga atacaggagg tatnnnctag cgttact

47

<210> 65

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 56

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 65

tgcaggacca gagaattcga atacataaca gttnnnctag cgttact

47

<210> 66

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> exemplary ligation strand 57

<220><221> misc_feature

<222> (34)..(36)

<223> n is any nucleotide

<400> 66

tgcaggacca gagaattcga atacatcgaa cacnnnctag cgttact

47

<210> 67

<211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 58
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 67
 tgcaggacca gagaattcga atacatgcat aatnnnctag cgttact 47

<210> 68
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 59
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 68
 tgcaggacca gagaattcga atacatgtca taannnctag cgttact 47

<210> 69
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> exemplary ligation strand 60
 <220><221> misc_feature
 <222> (34)..(36)
 <223> n is any nucleotide
 <400> 69
 tgcaggacca gagaattcga atacattgcg cggnnnctag cgttact 47

<210> 70
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220><223> single amplification primer

<400> 70

tgccaggacca gagaattcga ataca 25

<210> 71

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Forward Primer

<400> 71

aatgatacgg cgaccaccga gatctacacg tcatgcagga ccagagaatt cgaataca 58

<210> 72

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Reverse Primer

<400> 72

caagcagaag acggcatacg agatgtgact ggcacgggag ttgatcctgg ttttcac 57

<210> 73

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Forward Seq Primer

<400> 73

caagcagaag acggcatacg agatgtgact ggcacgggac cagagaattc gaatac 56

<210> 74

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Reverse Seq Primer 62

<400> 74

gtgactggca cgggaccaga gaattcgaat aca 33