



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년03월27일
(11) 등록번호 10-2788327
(24) 등록일자 2025년03월25일

- (51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) C12Q 1/6827 (2018.01)
C12Q 1/6855 (2018.01) C12Q 1/6883 (2018.01)
C12Q 1/6886 (2018.01)
- (52) CPC특허분류
C12N 15/1093 (2013.01)
C12Q 1/6827 (2018.05)
- (21) 출원번호 10-2018-7016475
- (22) 출원일자(국제) 2016년11월10일
심사청구일자 2021년11월09일
- (85) 번역문제출일자 2018년06월11일
- (65) 공개번호 10-2018-0113973
- (43) 공개일자 2018년10월17일
- (86) 국제출원번호 PCT/US2016/061395
- (87) 국제공개번호 WO 2017/083562
국제공개일자 2017년05월18일
- (30) 우선권주장
62/254,110 2015년11월11일 미국(US)
- (56) 선행기술조사문헌
WO2015117040 A1*
US20140274731 A1
WO2015134552 A1
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
레졸루션 바이오사이언스, 인크.
미국 98033 워싱턴주 커클랜드 스위트 200 커클랜드 웨이 550
- (72) 발명자
레이먼드, 크리스토퍼 케이.
미국 98115 워싱턴주 시애틀 노쓰이스트 에이티엑스 스트리트 2626
림, 리 피.
미국 98034 워싱턴주 커클랜드 아파트먼트 에이304 노쓰이스트 원헌드레드트웬티스 스트리트 9520
- (74) 대리인
양영준, 이상남

전체 청구항 수 : 총 52 항

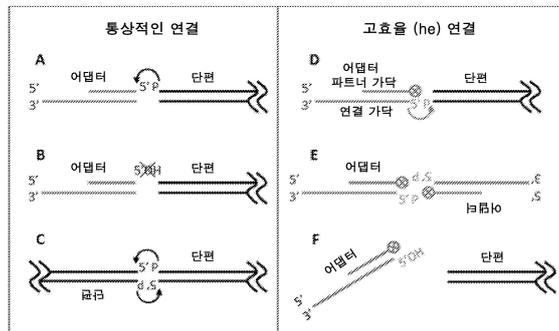
심사관 : 최지다

(54) 발명의 명칭 DNA 라이브러리의 고효율 작제

(57) 요약

본 발명은 효율적인 DNA 라이브러리 작제 및 상기 라이브러리의 표적화된 유전학적 분석을 위한 방법을 제공한다. 라이브러리는 대상체에서 유전학적 질환을 예측하거나, 진단하거나, 모니터링하는데 유용하다.

대표도



(52) CPC특허분류

C12Q 1/6855 (2018.05)

C12Q 1/6883 (2022.01)

C12Q 1/6886 (2022.01)

C12Q 2525/191 (2013.01)

명세서

청구범위

청구항 1

DNA 라이브러리를 작제하기 위한 방법으로서:

- (a) 하나 이상의 DNA 단편의 말단 포스페이트 잔기를 제거하여 탈인산화된 DNA 단편을 생성하는 단계;
- (b) 상기 탈인산화된 DNA 단편을 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA 단편을 생성하는 단계;
- (c) 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 상기 말단-복구된 DNA 단편의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는

(i) 제1 DNA 리가제를 사용하여 상기 말단-복구된 DNA 단편의 각각의 가닥의 3' 말단으로 연결된 앵커 서열을 포함하는 연결 가닥 올리고뉴클레오타이드; 및

(ii) 상기 연결 가닥 올리고뉴클레오타이드에 적어도 부분적으로 하이브리드화하는 비-연결 파트너 가닥 올리고뉴클레오타이드

를 포함하는 것인, 단계;

- (d) 상기 프리-어댑터/말단-복구된 DNA 복합체로부터의 상기 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 치환하여 어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 어댑터는 상기 연결 가닥 올리고뉴클레오타이드 및 상기 복구 올리고뉴클레오타이드를 포함하는, 단계; 및

- (e) 상기 어댑터/말단-복구된 DNA 복합체를 폴리뉴클레오타이드 키나제 및 제2 DNA 리가제로 처리하여 연속 dsDNA 단편의 라이브러리를 형성하는 단계로서, 상기 폴리뉴클레오타이드 키나제는 상기 말단-복구된 DNA 단편의 5' 말단에 포스페이트를 부가하고, 상기 제2 DNA 리가제는 상기 말단-복구된 DNA 단편의 각각의 가닥의 5' 말단에 복구 올리고뉴클레오타이드를 연결하는 것인, 단계

를 포함하는, DNA 라이브러리를 작제하기 위한 방법.

청구항 2

청구항 1에 있어서, 상기 어댑터/말단-복구된 DNA 복합체를 DNA 폴리머라제로 처리하는 단계를 추가로 포함하며, 상기 DNA 폴리머라제는 상기 복구 올리고뉴클레오타이드를 주형으로서 사용하여 프라이머 연장에 의해 상기 연결 가닥 올리고뉴클레오타이드를 연장시키는 것인, 방법.

청구항 3

청구항 1 또는 2에 있어서, 상기 비-연결 파트너 가닥 올리고뉴클레오타이드가 상기 말단-복구된 DNA 단편의 5' 말단으로의 연결 및 어댑터 이량체 형성 중 적어도 하나를 차단하는 3' 말단에서의 변형을 포함하는 것인, 방법.

청구항 4

청구항 1 또는 2에 있어서, 상기 하나 이상의 DNA 단편의 공급원이 게놈 DNA (gDNA), 상보적 DNA (cDNA), 및 세포-부재 DNA (cfDNA)로 이루어진 군으로부터 선택되는 DNA인, 방법.

청구항 5

청구항 4에 있어서, 상기 DNA의 공급원이 양수 샘플, 혈액 샘플, 피부 샘플, 모발 샘플, 모낭 샘플, 침 샘플, 점액 샘플, 경구 점막 샘플, 질 점막 샘플, 땀 샘플, 눈물 샘플, 상피 조직 샘플, 소변 샘플, 정액 (semen) 샘플, 정액 (seminal fluid) 샘플, 정액장 (seminal plasma) 샘플, 전립선액 샘플, 사정전액 (쿠퍼액 (Cowper's fluid)) 샘플, 배설물 샘플, 생검 샘플, 복수 샘플, 뇌척수액 샘플, 림프액 샘플, 조직 추출물 샘플, 혈장 샘플, 혈청 샘플, 림프액 샘플, 안구액 샘플, 대변 샘플 및 포르말린-고정된, 파라핀 매립된 (FFPE) 샘플로 이

루어진 군으로부터 선택되는 생물학적 샘플인, 방법.

청구항 6

청구항 5에 있어서, 상기 생물학적 샘플로부터 상기 DNA를 단리시키고, 상기 DNA를 단편화시키고, 단계 (c) 전에 하나 이상의 DNA 단편의 손상을 복구시키는 단계를 포함하는, 방법.

청구항 7

청구항 6에 있어서, 상기 손상이 탈아민화된 사이토신 (우라실), 무염기 부위, 구아닌의 O⁶MeG로의 메틸화, DNA 닉, 갭, 또는 티민 이량체인, 방법.

청구항 8

청구항 1 또는 2에 있어서, 상기 연결 가닥 올리고뉴클레오타이드가

- (i) 최소한 하나의 특유한 판독 코드;
 - (ii) 최소한 하나의 연속 dsDNA 라이브러리 분자의 PCR 증폭을 위한 PCR 프라이머 결합 부위;
 - (iii) 샘플 멀티플렉싱을 위한 최소한 하나의 샘플 코드; 및
 - (iv) DNA 서열분석을 위한 최소한 하나의 프라이머 결합 부위
- 중 적어도 하나를 포함하는 것인, 방법.

청구항 9

청구항 1 또는 2에 있어서, 상기 복구 올리고뉴클레오타이드가

- (i) 최소한 하나의 특유한 판독 코드;
 - (ii) 최소한 하나의 연속 dsDNA 라이브러리 분자의 PCR 증폭을 위한 최소한 하나의 프라이머 결합 부위;
 - (iii) 샘플 멀티플렉싱을 위한 최소한 하나의 샘플 코드; 및
 - (iv) DNA 서열분석을 위한 최소한 하나의 프라이머 결합 부위
- 중 적어도 하나 및 앵커 서열을 포함하는 것인, 방법.

청구항 10

청구항 1 또는 2에 있어서, 상기 연결 가닥 올리고뉴클레오타이드 및 상기 복구 올리고뉴클레오타이드 각각이 앵커 서열을 포함하고, 상기 연결 가닥 올리고뉴클레오타이드의 앵커 서열이 상기 복구 올리고뉴클레오타이드의 앵커 서열에 적어도 부분적으로 상보적인 것인, 방법.

청구항 11

청구항 1 또는 2에 있어서, 상기 연결 가닥 올리고뉴클레오타이드 및 상기 복구 올리고뉴클레오타이드 각각이 최소한 하나의 태그된 dsDNA 단편의 증폭을 위한 PCR 프라이머 결합 부위를 포함하고, 상기 연결 가닥 올리고뉴클레오타이드의 PCR 프라이머 결합 부위가 상기 복구 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 상보적인 것인, 방법.

청구항 12

청구항 1 또는 2에 있어서, 상기 연결 가닥 올리고뉴클레오타이드 및 상기 복구 올리고뉴클레오타이드 각각이 최소한 하나의 태그된 dsDNA 단편의 증폭을 위한 PCR 프라이머 결합 부위를 포함하고, 상기 연결 가닥 올리고뉴클레오타이드의 PCR 프라이머 결합 부위가 상기 복구 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 상보적이지 않은 것인, 방법.

청구항 13

청구항 12에 있어서, 상기 연결 가닥 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 결합하는 프라이머가

상기 복구 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 결합하지 않는 것인, 방법.

청구항 14

청구항 1 또는 2에 있어서, 상기 연결 가닥 올리고뉴클레오타이드가 상기 복구 올리고뉴클레오타이드보다 짧은 것인, 방법.

청구항 15

청구항 1 또는 2에 있어서, 상기 제1 DNA 리가제가 상기 제2 DNA 리가제와 상이한 것인, 방법.

청구항 16

청구항 15에 있어서, 상기 제2 DNA 리가제가 열안정한 닉-특이적 리가제인, 방법.

청구항 17

청구항 1 또는 2에 있어서, 단계 (b)가 평활 말단, 5'-오버행, 또는 3'-오버행을 포함하는 말단-복구된 DNA 단편을 생성하는 것인, 방법.

청구항 18

청구항 1 또는 2에 있어서, 단계 (c)가 제1 온도에서 수행되고, 단계 (d)가 제2 온도에서 수행되고, 상기 제2 온도가 상기 제1 온도보다 더 높아서, 상기 연결 가닥 올리고뉴클레오타이드로부터 상기 비-연결 파트너 가닥 올리고뉴클레오타이드의 치환을 유도하는 것인, 방법.

청구항 19

청구항 18에 있어서, 상기 제1 온도가 22 °C 이하이고, 상기 제2 온도가 37 °C 이상인, 방법.

청구항 20

청구항 1 또는 2에 있어서, 상기 어댑터/말단-복구된 DNA 복합체가 다수의 연결 가닥 올리고뉴클레오타이드 중 및 다수의 복구 올리고뉴클레오타이드 중 중 적어도 하나를 포함하는 것인, 방법.

청구항 21

청구항 1 또는 2에 있어서, 어댑터 연결 효율이 탈인산화된 어댑터 분자가 인산화된 DNA 단편에 연결되는 방법과 비교하여 증가된 것인, 방법.

청구항 22

청구항 1 또는 2에 있어서, 상기 연속 dsDNA 단편의 라이브러리가 DNA 클론 라이브러리를 생성하도록 증폭되는 것인, 방법.

청구항 23

청구항 22에 있어서, 상기 DNA 클론 라이브러리에서 다수의 유전학적 유전자좌 상에서 정량적 유전학적 분석을 수행하는 것을 포함하고, 여기서 상기 정량적 유전학적 분석이

(a) 최소한 하나의 포획 프로브 모듈을 상기 DNA 클론 라이브러리 내 다수의 유전학적 유전자좌로 하이브리드화하여 포획 프로브 모듈-DNA 클론 복합체를 형성하는 단계로서, 상기 포획 프로브 모듈이 포획 프로브 서열 및 꼬리 서열을 포함하는 것인, 단계; 및

(b) 상기 포획 프로브 모듈-DNA 클론 복합체를 단리하는 단계

를 포함하는 것인, 방법.

청구항 24

청구항 23에 있어서, 상기 포획 프로브 모듈-DNA 클론 복합체에서 단리된 DNA 클론을 주형으로서 사용하여 상기 포획 프로브 모듈의 5'-3' DNA 폴리머라제 연장을 수행하는 단계를 포함하는 방법.

청구항 25

청구항 23에 있어서, 상기 정량적 유전학적 분석이 다수의 서열분석 판독을 생성하기 위해 DNA 서열분석을 포함하는 것인 방법.

청구항 26

청구항 25에 있어서, 상기 정량적 유전학적 분석이 다수의 서열분석 판독의 생물정보학적 분석을 포함하는 것인 방법.

청구항 27

청구항 26에 있어서, 상기 생물정보학적 분석이

- (a) 상기 DNA 클론 라이브러리에서 분석된 게놈 등가물의 수를 정량하기 위해;
- (b) 표적 유전학적 유전자좌에서 유전학적 변이체를 검출하기 위해;
- (c) 표적 유전학적 유전자좌 내에서 돌연변이체를 검출하기 위해;
- (d) 표적 유전학적 유전자좌 내에서 유전학적 융합을 검출하기 위해; 또는
- (e) 표적 유전학적 유전자좌 내에서 카피수 변동을 측정하기 위해

사용되는 것인, 방법.

청구항 28

청구항 23에 있어서, 상기 정량적 유전학적 분석을 사용하여 유전학적 질환을 유발하거나 이와 연관된 하나 이상의 유전학적 병변을 동정하거나 검출하고, 상기 유전학적 병변이 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재정렬, 카피수 변화 또는 유전자 융합을 포함하는 것인, 방법.

청구항 29

- (a) 폴리뉴클레오타이드 키나제 및 DNA 리가제를 포함하는 제1 조성물;
- (b) 이중가닥 DNA (dsDNA) 프리-어댑터를 포함하는 제2 조성물로서, 상기 프리-어댑터가

- (i) (1) 최소한 하나의 특유한 판독 코드;
- (2) PCR 프라이머 결합 부위; 및
- (3) 샘플 멀티플렉싱을 위한 최소한 하나의 샘플 코드

중 적어도 하나 및 앵커 서열을 포함하는 연결 가닥 올리고뉴클레오타이드; 및

- (ii) 상기 연결 가닥 올리고뉴클레오타이드에 적어도 부분적으로 하이브리드화하는 비-연결 파트너 가닥 올리고뉴클레오타이드

를 포함하는 것인, 제2 조성물;

- (c) 복구 올리고뉴클레오타이드를 포함하는 제3 조성물로서, 상기 복구 올리고뉴클레오타이드가

- (1) 최소한 하나의 특유한 판독 코드;
- (2) PCR 프라이머 결합 부위;
- (3) 샘플 멀티플렉싱을 위한 최소한 하나의 샘플 코드; 및
- (4) 상기 연결 가닥 올리고뉴클레오타이드의 앵커 서열에 적어도 부분적으로 상보적인 앵커 서열

을 포함하는 것인, 제3 조성물

을 포함하며,

상기 폴리뉴클레오타이드 키나제는 말단-복구된 DNA 단편의 5' 말단에 포스페이트를 부가하여 인산화된 말단-복구된 DNA 단편을 생성할 수 있고, 상기 DNA 리가제는 상기 인산화된 말단-복구된 DNA 단편의 각각의 가닥의 5'

말단에 상기 복구 올리고뉴클레오타이드를 연결하여 연속 dsDNA 단편을 형성할 수 있는 것인, 인풋 DNA로부터 DNA 라이브러리를 작제하기 위한 다수의 조성물.

청구항 30

청구항 29에 있어서, 상기 제1 조성물이 DNA 폴리머라제를 포함하고, 상기 DNA 폴리머라제가 상기 복구 올리고뉴클레오타이드를 주형으로서 사용하여 프라이머 연장에 의해 상기 연결 가닥 올리고뉴클레오타이드를 연장시킬 수 있는 것인, 다수의 조성물.

청구항 31

청구항 29 또는 30에 있어서, 상기 비-연결 파트너 가닥 올리고뉴클레오타이드가 말단-복구된 DNA 단편의 5' 말단으로의 연결 및 어댑터 이량체 형성 중 적어도 하나를 차단하는 3' 말단에서의 변형을 포함하는 것인, 다수의 조성물.

청구항 32

청구항 29 또는 30에 있어서, 대상체로부터 DNA를 단리시켜 인풋 DNA를 생성시키기 위한 성분을 포함하는 제4 조성물을 추가로 포함하는, 다수의 조성물.

청구항 33

청구항 29 또는 30에 있어서, 포스파타제를 포함하는 제5 조성물을 추가로 포함하는, 다수의 조성물.

청구항 34

청구항 33에 있어서, 상기 포스파타제가 열-불안정한 것인, 다수의 조성물.

청구항 35

청구항 33에 있어서, 상기 포스파타제가 새우 알칼린 포스파타제인, 다수의 조성물.

청구항 36

청구항 29 또는 30에 있어서, 하나 이상의 말단-복구 효소를 포함하는 제6 조성물을 추가로 포함하는, 다수의 조성물.

청구항 37

청구항 36에 있어서, 상기 하나 이상의 말단-복구 효소가 평활 말단, 5'-오버행, 또는 3'-오버행을 포함하는 말단-복구된 DNA 단편을 생성할 수 있는 것인, 다수의 조성물.

청구항 38

청구항 36에 있어서, 상기 하나 이상의 말단-복구 효소가 T4 DNA 리가제를 포함하는 것인, 다수의 조성물.

청구항 39

청구항 29 또는 30에 있어서, 상기 제1 조성물의 DNA 리가제가 열안정한 너-특이적 리가제인, 다수의 조성물.

청구항 40

청구항 29 또는 30에 있어서, 상기 프리-어댑터의 연결 가닥을 말단-복구된 DNA 단편의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성할 수 있는 DNA 리가제를 포함하는 제7 조성물을 추가로 포함하는, 다수의 조성물.

청구항 41

청구항 40에 있어서, 상기 제1 조성물의 DNA 리가제가 상기 제7 조성물의 DNA 리가제와 상이한 것인, 다수의 조성물.

청구항 42

청구항 29 또는 30에 있어서, 상기 제2 조성물이 다수의 연결 가닥 올리고뉴클레오타이드 종을 포함하는 것인, 다수의 조성물.

청구항 43

청구항 29 또는 30에 있어서, 상기 제3 조성물이 다수의 복구 올리고뉴클레오타이드 종을 포함하는 것인, 다수의 조성물.

청구항 44

청구항 29 또는 30에 있어서, 상기 DNA 라이브러리를 PCR 증폭시켜 DNA 클론 라이브러리를 생성하기 위한 성분을 포함하는 제8 조성물을 추가로 포함하는, 다수의 조성물.

청구항 45

청구항 44에 있어서, 상기 복구 올리고뉴클레오타이드가 PCR 프라이머 결합 부위를 포함하고, 상기 제8 조성물이 상기 복구 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 결합하는 프라이머를 포함하는 것인, 다수의 조성물.

청구항 46

청구항 44에 있어서, 상기 DNA 클론 라이브러리에서 다수의 유전학적 유전자좌 상에서 정량적 유전학적 분석을 수행하기 위한 성분을 포함하는 제9 조성물을 추가로 포함하는, 다수의 조성물.

청구항 47

청구항 46에 있어서, 상기 제9 조성물이 상기 DNA 클론 라이브러리 내 다수의 유전학적 유전자좌로 하이브리드화하여 포획 프로브 모듈-DNA 클론 복합체를 형성할 수 있는 최소한 하나의 포획 프로브 모듈을 포함하고, 상기 포획 프로브 모듈이 포획 프로브 서열 및 꼬리 서열을 포함하는 것인, 다수의 조성물.

청구항 48

청구항 47에 있어서, 상기 포획 프로브 모듈이 상기 포획 프로브 모듈-DNA 클론 복합체의 단리 또는 정제를 용이하게 하기 위한 결합 쌍의 특정 구성원을 포함하는 것인, 다수의 조성물.

청구항 49

청구항 47에 있어서, 상기 제9 조성물이 포획 프로브 모듈의 꼬리 서열에 하이브리드화하는 파트너 올리고뉴클레오타이드를 추가로 포함하는 것인, 다수의 조성물.

청구항 50

청구항 49에 있어서, 상기 파트너 올리고뉴클레오타이드가 비오틴 분자를 포함하는 것인, 다수의 조성물.

청구항 51

청구항 47에 있어서, 상기 포획 프로브 모듈-DNA 클론 복합체를 단리하기 위한 성분을 포함하는 제10 조성물을 추가로 포함하는, 다수의 조성물.

청구항 52

청구항 50에 있어서, 상기 포획 프로브 모듈-DNA 클론 복합체를 단리하기 위한 성분을 포함하는 제10 조성물을 추가로 포함하고, 상기 제10 조성물이 스트렙타비딘을 포함하는 것인, 다수의 조성물.

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

발명의 설명

배경 기술

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2015년 11월 11일자로 출원된 미국 가출원 일련 번호 제62/254,110호에 대한 우선권을 주장하고 이의 이의 전문이 본원에 참조로 인용된다.

[0003] 전자적으로 제출된 텍스트 파일의 기재

[0004] 이와 함께 전자적으로 제출된 텍스트 파일의 내용은 이의 전문이 본원에 참조로 인용된다: 서열 목록 (파일명: CLFK_003_01WO_ST25, 기록일: 2016년 11월 10일, 파일 크기: 23kB)의 컴퓨터 판독 가능한 포맷 카피.

[0005] 기술분야

[0006] 본 발명은 일반적으로 DNA 라이브러리를 작제하기 위한 개선된 조성물 및 방법에 관한 것이다. 특히, 본 발명은 정량적 유전학적 분석을 위해 DNA 클론 라이브러리를 효율적으로 작제하는 것에 관한 것이다.

[0007] 관련 기술의 기재

[0008] 다운스트림 분석을 위해 목적하는 다양한 DNA 표본은 소량으로 수거한다. 예를 들어, 전혈의 혈장 분획으로부터 수거된 세포-부재 DNA (cfDNA)는 일반적으로 혈장 mL 당 나노그램 양으로 존재한다. 하나의 이배체 인간 게놈이 6 피코그램의 무게를 갖는 경우, 이것은 총 수백개 내지 수천개의 게놈 정보가 단일 채혈로부터 단리될 수 있음을 의미한다.

[0009] 암 환자에서, 종양 DNA는 총 순환하는 DNA의 $\leq 0.1\%$ 내지 $\geq 10\%$ 범위의 고도의 가변 양으로 혈류에 유입된다. 혈류는 단지 몇 나노그램의 DNA를 함유하고 종양 게놈이 총 순환하는 DNA의 0.1%로 존재하는 경우, 단지 총 1개 내지 10개 카피수의 종양 게놈이 존재한다. 서열 분석에 의해 종양 DNA를 명확하게 동정하기 위해서는, 2개 이상의 카피수의 종양-특이적 유전학적 병변을 관찰할 필요가 있다. 그러나, 0.1% 범위에서 종양 DNA의 정확한 검출을 의미하는 DNA의 검출 민감성을 최대화할 필요성은 아직 성취되지 않았다.

[0010] 이들 고려사항은 혈액을 사용하여 고체 종양의 신뢰할 수 있는 유전학적 분석이 희귀 게놈 단편들을 단리하고 분석하는 능력에 의해 부분적으로 지배된다는 근본적인 문제점을 조명한다. 더욱이, 많은 치료학적으로 실행 가능한 종양 병변은 유전자 융합, DNA 서열의 유의적 삽입 또는 결실 및/또는 유전자 카피수의 변화를 포함한다. 상기 변화는 PCR에 의한 분석에 불응하고, 여기서, 2개의 인접한 프라이머 결합 부위가 공지되어야만 하고 카피

수의 변화는 많은 횃수의 표적 증폭에 의해 모호해진다.

[0011] 현재, 표적 회수 방법은 순환하는 종양 DNA에서 잠재적 병변의 포괄적 분석을 위해 사용된다. 상기 회수 방법은 DNA 클론 라이브러리의 생성에 의존한다. 불행하게도, 이들 DNA 라이브러리를 생성하기 위한 현재 방법은 비효율적이고 단지 작은 %의 DNA 단편이 유용한 라이브러리 클론으로 성공적으로 전환된다.

발명의 내용

[0012] 본 발명은 일반적으로 DNA 어댑터를 DNA 단편으로 고효율적으로 부착시켜 정량적 유전학적 분석을 위해 DNA 라이브러리를 생성하는 조성물 및 방법에 관한 것이다.

[0013] 다양한 구현예에서, 하나 이상의 DNA 단편으로의 어댑터 연결의 효율을 증가시키기 위한 방법이 제공되고: 상기 방법은 하나 이상의 DNA 단편의 말단 포스페이트 잔기들을 제거하는 단계; 탈인산화된 DNA 단편을 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA를 생성하는 단계; 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 DNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 DNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 DNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 및 어댑터/말단-복구된 DNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 DNA 라이브러리를 형성하는 단계를 포함하고; 여기서, 어댑터 연결의 효율은 탈인산화된 어댑터 분자가 인산화된 DNA 단편에 연결되어 있는 방법과 비교하여 증가된다.

[0014] 다양한 구현예에서, DNA 라이브러리를 작제하기 위한 방법이 제공되고: 상기 방법은 하나 이상의 DNA 단편의 말단 포스페이트 잔기들을 제거하는 단계; 탈인산화된 DNA 단편을 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA를 생성하는 단계; 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 DNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 DNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 DNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 및 어댑터/말단-복구된 DNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 DNA 라이브러리를 형성하는 단계를 포함한다.

[0015] 특정 구현예에서, 비-연결 파트너 가닥 올리고뉴클레오타이드는 말단-복구된 DNA 및/또는 어댑터 이량체 형성의 5' 말단으로 이의 연결을 차단하는 3' 말단에서의 변형을 포함한다.

[0016] 특정 구현예에서, 하나 이상의 DNA 단편의 공급원은 하기로 이루어진 균으로부터 선택되는 DNA이다: 게놈 DNA (gDNA), 상보적 DNA (cDNA), 및 세포-부재 DNA (cfDNA).

[0017] 추가 구현예에서, DNA의 공급원은 하기로 이루어진 균으로부터 선택되는 생물학적 샘플이다: 혈액, 피부, 모발, 모낭, 침, 경구 점막, 질 점막, 땀, 눈물, 상피 조직, 소변, 정액 (semen), 정액 (seminal fluid), 정액장 (seminal plasma), 전립선액, 사정전액 (쿠퍼액 (Cowper's fluid)), 배설물, 생검, 복수, 뇌척수액, 림프액, 및 조직 추출물 샘플 또는 생검 샘플.

[0018] 특정 구현예에서, DNA의 공급원은 하기로 이루어진 균으로부터 선택되는 생물학적 샘플이다: 양수, 혈액, 혈장, 혈청, 정액, 림프액, 뇌척수액, 안구액, 소변, 침, 대변, 점액, 및 땀.

[0019] 추가의 구현예에서, 상기 방법은 대상체의 생물학적 샘플로부터 DNA를 단리시킴을 추가로 포함한다.

[0020] 일부 구현예에서, 상기 방법은 대상체의 생물학적 샘플로부터 DNA를 단편화시킴을 추가로 포함한다.

[0021] 특정 구현예에서, 상기 방법은 연결 전 하나 이상의 DNA 단편의 손상을 복구함을 추가로 포함한다.

[0022] 특정 구현예에서, 손상은 탈아민화된 사이토신 (우라실), 무염기 부위, 구아닌의 O6MeG로의 메틸화, DNA Nick, 갭, 또는 티민 이량체이다.

[0023] 다양한 구현예에서, cfDNA 라이브러리를 작제하는 방법이 제공되고: 상기 방법은 cfDNA를 대상체의 생물학적 샘플로부터 단리시키거나 수득하는 단계; cfDNA의 말단 포스페이트 잔기들을 제거하는 단계; 탈인산화된 cfDNA를

하나 이상의 말단-복구 효소로 처리하여 말단-복구된 cfDNA를 생성하고 임의로 DNA 손상을 복구시키는 단계; 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 cfDNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 cfDNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 cfDNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 cfDNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 cfDNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 및 어댑터/말단-복구된 cfDNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 cfDNA 라이브러리를 형성하는 단계; 및 cfDNA 라이브러리를 증폭시켜 세포-부재 DNA 클론 라이브러리를 생성하는 단계를 포함한다.

- [0024] 특정 구현예에서, 상기 연결 가닥 올리고뉴클레오타이드는 어댑터 이량체 형성을 차단시키기 위한 하나 이상의 변형을 포함하고, 임의로, 여기서, 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단의 변형은 어댑터 이량체 형성을 막는다.
- [0025] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 앵커 서열, 판독 코드 또는 PCR 프라이머 결합 부위를 포함한다.
- [0026] 추가의 구현예에서, 연결 가닥 올리고뉴클레오타이드는 앵커 서열, 판독 코드 및 PCR 프라이머 결합 부위를 포함한다.
- [0027] 일부 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 연속, 이중가닥, DNA 라이브러리 분자의 PCR 증폭을 위한 하나 이상의 PCR 프라이머 결합 부위를 포함한다.
- [0028] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 특유의 판독 코드를 포함한다.
- [0029] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 샘플 멀티플렉싱을 위한 하나 이상의 샘플 코드를 포함한다.
- [0030] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 DNA 서열분석을 위한 하나 이상의 서열을 포함한다.
- [0031] 추가의 구현예에서, 연결 가닥 올리고뉴클레오타이드는 앵커 서열을 포함한다.
- [0032] 추가의 구현예에서, 복구 올리고뉴클레오타이드는 앵커 서열, 판독 코드 또는 PCR 프라이머 결합 부위를 포함한다.
- [0033] 특정 구현예에서, 복구 올리고뉴클레오타이드는 앵커 서열, 판독 코드 및 PCR 프라이머 결합 부위를 포함한다.
- [0034] 특정 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 연속, 이중가닥, DNA 라이브러리 분자의 PCR 증폭을 위한 하나 이상의 프라이머 결합 부위를 포함한다.
- [0035] 일부 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 특유의 판독 코드를 포함한다.
- [0036] 특정 구현예에서, 복구 올리고뉴클레오타이드는 샘플 멀티플렉싱을 위한 하나 이상의 샘플 코드를 포함한다.
- [0037] 특정 구현예에서, 복구 올리고뉴클레오타이드는 DNA 서열분석을 위한 하나 이상의 서열을 포함한다.
- [0038] 추가의 구현예에서, 연결 가닥 올리고뉴클레오타이드는 복구 올리고뉴클레오타이드에 상보적이다.
- [0039] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드의 앵커 서열은 복구 올리고뉴클레오타이드의 앵커 서열에 상보적이다.
- [0040] 추가의 구현예에서, 연결 가닥 올리고뉴클레오타이드의 PCR 프라이머 결합 부위는 복구 올리고뉴클레오타이드의 PCR 프라이머 결합 부위에 상보적이다.
- [0041] 특정 구현예에서, 하나 이상의 어댑터는 다수의 연결 가닥 올리고뉴클레오타이드 종을 포함한다.
- [0042] 일부 구현예에서, 하나 이상의 어댑터는 다수의 복구 올리고뉴클레오타이드 종을 포함한다.
- [0043] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드의 프라이머 결합 부위는 복구 올리고뉴클레오타이드의 프라이머 결합 부위에 상보적이지 않다.
- [0044] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드의 프라이머 결합 부위는 복구 올리고뉴클레오타이드의 프라이머 결합 부위와 실질적으로 상이하다.

- [0045] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드의 프라이머 결합 부위에 결합하는 프라이머는 복구 올리고뉴클레오타이드의 프라이머 결합 부위에 실질적으로 결합하지 않는다.
- [0046] 특정 구현예에서, DNA 라이브러리는 증폭되어 DNA 클론 라이브러리를 생성한다.
- [0047] 추가의 구현예에서, qPCR은 DNA 클론 라이브러리에 대해 수행하고 qPCR 측정은 공지된 계놈 증가물의 표준물과 비교하여 DNA 클론 라이브러리의 계놈 증가물을 결정한다.
- [0048] 특정 구현예에서, qPCR은 Alu 서열에 결합하는 프라이머 및 어댑터 중 서열에 결합하는 프라이머를 사용하여 수행한다.
- [0049] 일부 구현예에서, 정량적 유전학적 분석은 DNA 클론 라이브러리에서 다수의 유전학적 유전자좌 상에서 수행된다.
- [0050] 특정 구현예에서, 정량적 유전학적 분석은 다수의 DNA 클론 라이브러리에서 다수의 유전학적 유전자좌 상에서 수행된다.
- [0051] 특정 구현예에서, 정량적 유전학적 분석은 표적 유전학적 유전자좌로 하나 이상의 포획 프로브를 하이브리드화하여 포획 프로브 모듈-DNA 클론 복합체를 형성함을 포함한다.
- [0052] 특정 구현예에서, 정량적 유전학적 분석은 포획 프로브-DNA 클론 복합체를 단리함을 포함한다.
- [0053] 추가의 구현예에서, 정량적 유전학적 분석은 단리된 포획 프로브-DNA 클론 복합체에서 DNA 클론 서열의 증폭을 포함한다.
- [0054] 특정 구현예에서, 정량적 유전학적 분석은 DNA 서열분석으로 다수의 서열분석 판독을 생성함을 포함한다.
- [0055] 추가의 구현예에서, 상기 방법은 추가로 다수의 서열분석 판독의 생물정보학적 분석을 포함한다.
- [0056] 특정 구현예에서, 정량적 유전학적 분석은 DNA 클론 라이브러리에서 다수의 유전학적 유전자좌에 대해 수행하고, 생물정보학적 분석을 사용하여 DNA 클론 라이브러리에서 분석된 계놈 증가물의 수를 정량하고; 표적 유전학적 유전자좌에서 유전학적 변이체를 검출하고; 표적 유전학적 유전자좌 내 돌연변이를 검출하고; 표적 유전학적 유전자좌내 유전학적 융합을 검출하고/하거나 표적 유전학적 유전자좌 내 카피수 변동을 측정한다.
- [0057] 특정 구현예에서, 정량적 유전학적 분석을 사용하여 유전학적 질환을 유발하거나 이와 연관된 하나 이상의 유전학적 병변을 동정하거나 검출한다.
- [0058] 특정 구현예에서, 유전학적 병변은 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 계놈 재정렬, 카피수 변화 또는 유전자 융합을 포함한다.
- [0059] 특정 구현예에서, 유전학적 질환은 암이다.
- [0060] 추가의 구현예에서, 정량적 유전학적 분석을 사용하여 태아 cfDNA에서 하나 이상의 유전학적 변이체 또는 하나 이상의 표적 유전학적 유전자좌의 유전학적 병변을 동정하거나 검출한다.
- [0061] 일부 구현예에서, 포획 프로브는 포획 프로브 모듈에서 꼬리 서열에 하이브리드화하는 함텐-표지된 파트너 올리고뉴클레오타이드와 임의로 이중가닥을 형성하는 포획 프로브 모듈의 성분이다.
- [0062] 다양한 구현예에서, 다음 단계를 포함하여 대상체에서 유전학적 질환을 예측하거나, 진단하거나, 모니터링하는 방법으로서, 상기 방법은: 대상체의 생물학적 샘플로부터 DNA를 단리시키거나 수득하는 단계; DNA의 말단 포스페이트 잔기들을 제거하는 단계; 탈인산화된 DNA를 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA를 생성하는 단계; 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 DNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 DNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 DNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 어댑터/말단-복구된 DNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 DNA 라이브러리를 형성하는 단계; DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하는 단계; DNA 클론 라이브러리에서 계놈 증가물의 수를 결정하는 단계; 및 DNA 클론 라이브러리에서 유전학적 질환과 연관된 하나 이상의 표적 유전자좌의 정량적 유전학적 분석

을 수행하는 단계를 포함하고, 여기서, 하나 이상의 표적 유전학적 유전자좌에서 하나 이상의 유전학적 병변의 동정 또는 검출은 유전학적 질환을 예후하거나, 진단하거나, 이의 진행을 모니터링하는 것이다.

- [0063] 특정 구현예에서, DNA는 게놈 DNA, 포르말린-고정된, 파라핀 매립된 (FFPE) 샘플로부터의 DNA, cDNA 또는 cfDNA이다.
- [0064] 특정 구현예에서, cfDNA는 하기로 이루어진 군으로부터 선택되는 생물학적 샘플로부터 단리된다: 양수, 혈액, 혈장, 혈청, 정액, 림프액, 뇌척수액, 안구액, 소변, 침, 대변, 점액, 및 땀.
- [0065] 추가의 구현예에서, 유전학적 병변은 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재정렬, 카피수 변화 또는 유전자 융합을 포함한다.
- [0066] 특정 구현예에서, 유전학적 질환은 암이다.
- [0067] 다양한 구현예에서, 다음 단계를 포함하여 유전학적 질환을 동반 진단하는 것이 제공된다: 대상체의 생물학적 샘플로부터 DNA를 단리시키거나 수득하는 단계; DNA의 말단 포스페이트 잔기들을 제거하는 단계; 탈인산화된 DNA를 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA를 생성하는 단계; 하나 이상의 이중가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 DNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 DNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 DNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 cfDNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 어댑터/말단-복구된 DNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 DNA 라이브러리를 형성하는 단계; DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하는 단계; DNA 클론 라이브러리에서 게놈 등가물의 수를 결정하는 단계; 및 DNA 클론 라이브러리에서 유전학적 질환과 연관된 하나 이상의 바이오마커의 정량적 유전학적 분석을 수행하는 단계를 포함하고, 여기서, 하나 이상의 바이오마커의 검출 또는 검출 실패는 대상체가 유전학적 질환에 대해 치료되어야만 하는지를 지적한다.
- [0068] 추가의 구현예에서, DNA는 게놈 DNA, 포르말린-고정된, 파라핀 매립된 (FFPE) 샘플로부터의 DNA, cDNA 또는 cfDNA이다.
- [0069] 특정 구현예에서, cfDNA는 하기로 이루어진 군으로부터 선택되는 생물학적 샘플로부터 단리된다: 양수, 혈액, 혈장, 혈청, 정액 (semen), 림프액, 뇌척수액, 안구액, 소변, 침, 대변, 점액, 및 땀.
- [0070] 특정 구현예에서, 바이오마커는 유전학적 병변이다.
- [0071] 추가의 구현예에서, 유전학적 병변은 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재정렬, 카피수 변화 또는 유전자 융합을 포함한다.
- [0072] 특정 구현예에서, 유전학적 질환은 암이다.

도면의 간단한 설명

[0073] **도 1**은 통상적인 것과 고효율 (HE) 연결 기술을 보여준다. (A) 표적 DNA 단편은 비인산화된 이중나선 어댑터를 사용한 DNA 연결 전에 5' 인산화된다. (B) 통상의 연결 방법을 사용한 공통된 비효율은 포스페이트 그룹이 없는 5' 표적 DNA 단편 말단은 비인산화된 이중나선 어댑터와 연결되는데 실패한다. (C) 표적 DNA 단편 서로 간의 연결. (D) 요구되는 5' 포스페이트 그룹을 포함하는 이중나선 어댑터는 표적 DNA 단편의 3' 말단에 연결한다. 어댑터의 파트너 가닥 올리고뉴클레오타이드의 해지된 씨클은 3' 차단 그룹을 나타낸다. (E) 3' 차단 그룹은 또한 어댑터가 서로 연결되지 못하도록 한다. (F) 5' 포스페이트가 없는 드문 어댑터 이중나선은 표적 단편에 연결되는데 실패한다.

도 2는 RsaI 분해된, 탈인산화된 pUC19 플라스미드가 일련의 9개의 상이한 HE 어댑터에 완전히 연결된 아가로스 겔의 대표적 이미지를 보여준다. 연결되지 않은 단편은 740, 500, 300, 및 150 bp (상부에서 하부로)의 분자량 (MW) 마커 옆에 있는 대조군 레인에 지적된다. 3개의 벡터 단편의 이동에서 완전한 쉬프트 (화살표)는 모든 9개의 어댑터에 걸친 양쪽 말단으로의 어댑터의 완전한 연결을 지적한다. 이들 결과는 HE 연결 기술이 일반화될 수 있음을 보여준다.

도 3은 HE 연결된 작제물을 완료하는 예시적 방법을 보여준다. (A) HE 연결 생성물은 단편의 3' 말단에 부착된

3' 연장된 어댑터 올리고뉴클레오타이드이다. (B) 초기 연결 생성물을 "복구하기" 위한 하나의 전략은 복구 올리고뉴클레오타이드 (상부 가닥: 녹색)를 추가하는 것이고; T4 폴리뉴클레오타이드 키나제는 포스페이트 (P)를 표적 단편의 5' 말단으로 추가할 수 있고, *Taq* DNA 리가제와 같은 닉 밀봉 리가제를 사용하여 복구 올리고뉴클레오타이드를 표적 단편으로 연결할 수 있다. (C) 대안적 전략은 5'에서 3' 엑소뉴클레아제 활성을 갖는 DNA 폴리머라제 (예를 들어, *BstI* DNA 폴리머라제), 및 *Taq* DNA 리가제에 의해 상보적 어댑터 올리고뉴클레오타이드를 조합하는 것이다. 상기 *BstI* DNA 폴리머라제는 표적 단편으로부터 5' 염기를 제거하여 5' 포스페이트를 노출시켜 *Taq* DNA 리가제에 의한 닉-밀봉 연결시킴에 의해 복구 올리고뉴클레오타이드를 연장시킨다. (D) 상보적 어댑터 올리고뉴클레오타이드는 또한 *BstI* DNA 폴리머라제를 사용한 본래의 HE 연결 가닥으로의 추가의 서열 특징을 도입하도록 디자인될 수 있다.

도 4는 HE 연결 기술을 사용한 DNA 라이브러리의 제조를 보여준다. (A) DNA 단편들은 포스페이트 그룹 (P)를 가질 수 있거나 가질 수 없는 이중성 "래그된" 말단을 가질 수 있다. 포스파타제를 사용한 DNA 단편의 처리는 노출된 5' 및 3' 포스페이트를 제거한다. 이어서 DNA는 탈아민화된 시토신 (U), 무염기 부위 (↑), 및 티민 이량체와 같은 DNA 손상을 복구시키고 5' 또는 3' 오버행을 평활 말단으로 "폴리싱"하는 효소로 처리될 수 있다. (B) 어댑터는 2단계로 DNA 단편으로 첨가된다. 첫번째로, 5' 인산화된 연결 가닥 및 3' 차단된 파트너 가닥을 포함하는 이중나선 어댑터는 표적 단편에 연결한다. ~ 30 °C의 용융 온도를 갖는 파트너 가닥은 ≥ 37 °C의 온도에서 수행하는 후속적 단계에서 제거한다. 두번째로, 복구 올리고뉴클레오타이드는 어댑터 연결된 단편에 어닐링하고; 복구 올리고뉴클레오타이드는 키나제/리가제 전략 또는 폴리머라제/엑소뉴클레아제/리가제 전략을 사용하여 표적 단편의 5' 말단으로 공유 부착하게 된다 (도 3). 초기 연결 가닥의 프라이머 연장은 다운스트림 분석을 위해 적합한 전장 어댑터 이중나선으로의 복구 올리고 정보를 카피한다.

도 5는 이원 PCR 프라이머 적용된 DNA 단편을 생성하기 위한 HE 연결 기술 전략을 보여준다. 상기 도식에서, 연결 가닥은 독립적 프라이머 결합 부위로서 작용하는 추가의 서열 (프라이머 2)을 갖는다. 상기 복구 올리고뉴클레오타이드는 연결 가닥 부분에 상보적이고 제2 PCR 프라이머 결합 부위로서 작용하는 그 자신의 분기 서열 (프라이머 1)을 갖는다. 완전히 완성된 어댑터는 보다 통상적인 범용 이원-프라이머 PCR 방법을 사용하여 DNA 표본 단편의 증폭을 가능하게 한다.

발명을 실시하기 위한 구체적인 내용

- [0074] **A. 개요**
- [0075] 본 발명은 부분적으로 다운스트림 분석을 위해 상기 DNA 단편을 클로닝하기 위해 개선된 고도의 효율적 방법을 위한 정량적 유전학적 분석 분야에서의 긴급한 필요성을 해결하기 위한 조성물 및 방법을 고려한다.
- [0076] DNA 분석을 위한 현재 방법은 특수화된 어댑터의 DNA 단편으로의 연결을 포함한다 (도 1). 통상적인 기술에서, 표적 DNA 단편은 비인산화된 이중나선 어댑터를 사용한 공유 연결시킬 수 있는 DNA 연결 전에 5' 인산화된다. 표적 DNA 단편 및 어댑터는 평활 말단화될 수 있거나 이들은 상보적 오버행 (예를 들어, T/A)을 공유할 수 있다. (도 1a). 이것은 모든 표적 DNA 단편의 양 말단이 인산화됨을 보장할 수 없고, 인산화되지 않은 말단이 연결될 수 없고 이들 표적 단편이 후속적 라이브러리로부터 상실되기 때문에 심각한 결점이다. (도 1b). 비-제한적인 예로서, 표적 DNA 단편 말단의 70 %가 5' 포스페이트를 갖는 경우, 이어서 최대치에서 단편의 49 % (0.7 x 0.7 x 100 %)만이 단편의 양 말단 상에 연결될 수 있고 양 말단으로의 연결은 클로닝을 위해 요구된다. 추가로, 표적 DNA 단편 상에 5' 포스페이트의 존재는 DNA 단편이 서로 연결될 수 있는 별도의 목적하지 않는 인공물을 촉진시킨다 (도 1c). 이것은 질환-특이적 염색체 재배열의 검출을 혼동시킬 수 있는 인위적 염색체 서열 융합 반응을 생성한다.
- [0077] 다양한 구현예에서, 본 발명은 부분적으로 어댑터 서열을 표적 DNA 단편에 효율적으로 부착시키기 위한 조성물 및 방법을 고려한다. 특정 구현예에서, 포스페이트는 표적 DNA 단편의 5' 및 3' 말단 둘 다로부터 제거한다. 이들 탈인산화된 단편은 이어서 평활 DNA 말단을 생성하는 효소 및 임의로 DNA 상에 가해질 수 있는 많은 유형의 DNA 손상, 예를 들어, 탈아민화된 시토신 (우라실), 무염기 부위, 구아닌의 ⁶O⁶MeG로의 메틸화, 닉, 이중가닥 브레이크, 또는 티민 이량체를 복구시키는 효소로 처리된다. 상기 어댑터는 비-연결 파트너 가닥 올리고뉴클레오타이드와 이중나선을 형성하는 연결 가닥 올리고뉴클레오타이드를 포함한다. 어댑터의 연결 가닥은 표적 DNA 단편으로의 연결을 위해 요구되는 5' 포스페이트 그룹을 갖고 파트너 가닥은 3' 차단 그룹을 포함한다 (도 1d). 상기 3' 차단 그룹은 어댑터:어댑터 이량체의 형성을 방지한다 (도 1e). DNA 단편과 함께, 모든 어댑터 서열이 5' 포스페이트를 갖지는 않는다 (용매 노출된 말단 포스페이트 결합은 고유적으로 화학적으로 불안정하다). 상

기 비인산화된 어댑터가 존재하지만 이들은 단지 일시적으로 연결 기작에 관여하고 (도 1f); 단편과 상기 어댑터의 비생산적 쌍 형성은 신속하게 분해되고 생산적 공유 부착을 제공할 수 있는 어댑터:표적 DNA 단편 쌍 형성에 의해 대체된다. 궁극적으로, 표적 DNA 단편 ~100 %는 어댑터 분자와 양 말단 상에서 부착하게 되고, 이것은 본원에서 DNA 라이브러리를 작제하기 위해 고려되는 고효율의 조성물 및 방법을 설명한다.

[0078] 다양한 구현예에서, 본원에서 DNA 라이브러리의 고효율 작제를 위해 고려되는 조성물 및 방법은 다양한 생물학적 공급원으로부터 가용한 DNA를 사용한, 신규한 포괄적 골격 어드레스 분자 유전학적 분석을 제공한다. 정제된 DNA의 클로닝은 다운스트림 분석을 고지하고 수득한 클론 라이브러리의 증폭을 가능하게 하는 태그된 DNA 서열을 도입한다. 표적 특이적 올리고뉴클레오타이드를 사용한 하이브리드 포획을 사용하여 후속적 분석을 위한 특이적 서열을 검색한다. 라이브러리에 존재하는 게놈 수의 독립적 측정은 각각의 샘플에 적용되고, 이들 검정은 검정의 민감성을 평가하기 위한 수단을 제공한다. 본원에서 고려되는 상기 검정은 유전학적 상태, 병태 또는 질환의 분석, 검출, 진단 또는 모니터링을 위해 신뢰할 수 있으며 재현성이 있으며 강력한 방법을 제공한다.

[0079] 본 발명의 특정 구현예의 수행은, 달리 반대로 구체적으로 나타내지 않는 한, 당업계의 기술 분야에 있고 이의 많은 부분이 설명을 목적으로 하기 되어 있는 화학, 생화학, 유기 화학, 분자 생물학, 미생물학, 재조합 DNA 기술, 유전학, 면역학, 및 세포 생물학을 사용할 것이다. 상기 기술은 문헌에서 완전하게 설명된다. 예를 들어, 문헌 (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); and Harlow and Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998))을 참조한다.

[0080] **B. 정의**

[0081] 달리 정의되지 않는 경우, 본원에서 사용된 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술 분야의 통상의 기술자에 의해 통상적으로 이해되는 바와 동일한 의미를 갖는다. 본원에 기재된 것들과 유사하거나 동등한 임의의 방법 및 물질이 본 발명의 수행 또는 시험에 사용될 수 있지만, 조성물, 방법 및 물질의 바람직한 구현예는 본원에 기재된다. 본 발명의 목적을 위해, 다음의 용어는 하기에 정의된다.

[0082] 관사 "a", "an" 및 "the"는 본원에서 관사의 문법적 대상의 하나 또는 하나 초과 (즉, 적어도 하나)를 언급하기 위해 사용된다. 예를 들어, "요소"는 하나의 요소 또는 하나 초과의 요소를 의미한다.

[0083] 대안물 (예를 들어, "또는")의 사용은 대안물 중 하나, 둘 다, 또는 이들의 임의의 조합을 의미하는 것으로 이해되어야 한다.

[0084] 용어 "및/또는"은 대안물 중 하나, 또는 둘 다를 의미하는 것으로 이해되어야만 한다.

[0085] 본원에 사용된 바와 같은 용어 "약" 또는 "대략적으로"는 참조 양, 수준, 값, 수, 빈도, 백분율, 차원, 크기, 양, 중량 또는 길이에 비해 15 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 % 또는 1 % 정도로 많이 다양한 양, 수준, 값, 수, 빈도, 백분율, 차원, 크기, 양, 중량 또는 길이를 언급한다. 하나의 구현예에서, 용어 "약" 또는 "대략적으로"는 참조 양, 수준, 값, 수, 빈도, 백분율, 차원, 크기, 양, 중량 또는 길이의 ± 15 %, ± 10 %, ± 9 %, ± 8 %, ± 7 %, ± 6 %, ± 5 %, ± 4 %, ± 3 %, ± 2 %, 또는 ± 1 % 양, 수준, 값, 수, 빈도, 백분율, 차원, 크기, 양, 중량 또는 길이를 언급한다.

[0086] 본원 명세서 전반에 걸쳐, 달리 요구되지 않는 한, 용어 "포함한다 (comprise)", "포함한다 (comprises)" 및 "포함하는"은 진술된 단계 또는 요소 또는 단계 또는 요소 그룹의 내포를 의미하고 임의의 다른 단계 또는 요소 또는 단계 또는 요소들의 그룹을 배제하는 것이 아닌 것으로 이해된다. 특정 구현예에서, 용어 "포함한다 (include)", "갖는다 (has)", "함유한다 (contains)", 및 "포함한다 (comprise)"는 동의어로 사용된다.

[0087] "로 이루어진"이란, "로 이루어진"란 문구에 뒤따르는 무엇이든지 간에 이를 포함하고 이에 제한되는 것임을 의미한다. 따라서, "로 이루어진"이란 문구는 열거된 요소들이 요구되거나 의무적이고 어떠한 다른 요소들이 존재할 수 없음을 지적한다.

- [0088] "필수적으로 이루어진"이란, 상기 문구 뒤에 열거된 임의의 요소들을 포함함을 의미하고 열거된 요소들에 대한 개시에 특정된 활성 또는 작용을 간섭하지 않거나 기여하는 다른 요소들로 제한됨을 의미한다. 따라서, "필수적으로 이루어진"이란 문구는 열거된 요소들이 요구되거나 의무적이지만 어떠한 다른 요소들도 임의적이지 않으며 이들이 열거된 요소들의 활성 또는 작용에 영향을 미치는지 여부에 따라 존재할 수 있거나 존재하지 않을 수도 있음을 지적한다.
- [0089] 본 명세서 전체에서의 "하나의 구현예", "한 구현예", "특정 구현예", "관련 구현예", "특정 구현예", "추가적 구현예" 또는 "추가적 구현예" 또는 이들의 조합은 구현예와 관련하여 기재된 특정 특징, 구조 또는 특성이 본 발명의 적어도 하나의 구현예에 포함됨을 의미한다. 따라서, 본 명세서 전체에서의 다양한 곳에서의 이전 문구의 출현은 반드시 전부 동일한 구현예를 언급하는 것은 아니다. 추가로, 특정 특징, 구조 또는 특성은 하나 이상의 구현예에서 임의의 적합한 방식으로 조합될 수 있다.
- [0090] 본원에 사용된 바와 같은 용어 "단리된"은 이의 본래 형태에서 통상적으로 동반되는 성분들이 실질적으로 또는 본질적으로 없는 물질을 의미한다. 특정 구현예에서, 용어 "수득된" 또는 "유래된"은 단리된과 동의어로 사용된다.
- [0091] 본원에 사용된 바와 같은 용어 "DNA"는 데옥시리보핵산을 언급한다. 다양한 구현예에서, 용어 DNA는 게놈 DNA, 재조합 DNA, 합성 DNA, 상보성 DNA (cDNA), 또는 세포-부재 DNA (cfDNA)를 언급한다. 하나의 구현예에서, DNA는 게놈 DNA 또는 cDNA를 언급한다. 하나의 구현예에서, DNA는 cfDNA를 언급한다. 특정 구현예에서, DNA는 특정 구현예에서 표적 DNA 단편으로서 또한 언급되는 "표적 영역"을 포함하는 DNA 단편이다. 본원에서 고려되는 DNA 라이브러리는 게놈 DNA 라이브러리, cfDNA 라이브러리, 및 RNA, 예를 들어, RNA 발현 라이브러리로부터 작제된 cDNA 라이브러리를 포함한다. 다양한 구현예에서, DNA 라이브러리는 하나 이상의 추가적 DNA 서열 및/또는 태그를 포함한다.
- [0092] "표적 유전학적 유전자좌" 또는 "DNA 표적 영역"은 DNA 서열 내에서 목적하는 영역을 언급한다. 다양한 구현예에서, 표적화된 유전학적 분석은 표적 유전학적 유전자좌 상에서 수행된다. 특정 구현예에서, DNA 표적 영역은 특정 유전학적 상태, 유전학적 병태, 유전학적 질환; 태아 시험; 유전학적 기작, 친부 확인 시험; 약물 치료에 대한 반응 예측; 의학적 병태의 진단 또는 모니터링; 마이크로바이옴 프로파일링; 병원체 스크리닝; 또는 기관 이식 모니터링과 관련된 유전자 영역이다.
- [0093] 본원에 사용된 바와 같은 용어 "순환하는 DNA", "순환하는 세포-부재 DNA" 및 "세포 부재 DNA"는 흔히 상호교환적으로 사용되고 세포로부터 추출된 DNA, 또는 괴사 또는 아포토시스 세포로부터 방출된 DNA인 세포의 DNA인 DNA를 언급한다.
- [0094] 본원에 사용된 바와 같은 "대상체", "개체", 또는 "환자"는 본원에 고려된 조성물로 검출되거나 동정될 수 있는 병태의 증상을 나타내는 임의의 동물을 포함한다. 적합한 대상체는 실험실 동물 (예를 들어, 마우스, 래트, 토끼, 또는 기니아 피그), 농장 동물 (예를 들어, 말, 소, 양, 돼지), 및 가축 또는 애완동물 (예를 들어, 고양이 또는 개)을 포함한다. 특정 구현예에서, 대상체는 포유동물이다. 특정 구현예에서, 대상체는 비-인간 영장류이고, 바람직한 구현예에서, 대상체는 인간이다.
- [0095] "반응 용기"는 본원에서 고려된 반응들 중 하나를 수행하기에 적합한 컨테이너를 의미한다. 특정 구현예에 사용하기에 적합한 반응 용기의 예시적 예는 시험 튜브, 미세원심 튜브 (예를 들어, PCR 튜브), 미세역가 플레이트 (예를 들어, 96웰 플레이트, 384 웰 플레이트, 1536 웰 플레이트), 슬라이드, 플레이트, 어레이 및 마이크로어레이를 포함하지만 이에 제한되지 않는다.
- [0096] **C. DNA 라이브러리의 고효율 작제**
- [0097] 특정 구현예에서, 본원에 고려된 DNA 라이브러리를 작제하는 방법은 표적 DNA 단편으로의 어댑터의 고효율 연결을 포함한다.
- [0098] **(a) DNA 공급원**
- [0099] 본원에서 고려되는 방법 및 조성물은 유전학적 상태, 유전학적 병태, 유전학적 질환, 유전학적 모자이크 현상, 태아 진단제, 친부 확인 시험, 마이크로바이옴 프로파일링, 병원체 스크리닝 및 분석물로서 DNA를 사용한 기관 이식 모니터링을 효율적으로 분석하고, 검출하고, 진단하고/하거나 모니터링하기 위해 디자인된다. 본원에 고려된 조성물 및 방법에 사용하기에 적합한 DNA는 당업자에게 공지된 임의의 공급원으로부터 기원할 수 있다. 특정 구현예에서, DNA는 임의의 공급원으로부터 단리된 게놈 DNA, RNA로부터 합성된 카피 DNA (cDNA) 또는 세포

-부재 DNA (cfDNA)이다.

- [0100] 일부 구현예에서, DNA는 고분자량 DNA (>1000 bp)이다. 본원에 고려된 조성물 및 방법에서 고분자량 DNA의 용도는 흔히 단편화 단계를 포함한다. 고분자량 DNA는 약 25 내지 약 750 염기쌍, 약 25 내지 약 500 염기쌍, 약 25 내지 약 250 염기쌍, 약 25 내지 약 200 염기쌍, 약 25 내지 약 150 염기쌍, 약 25 내지 약 100 염기쌍, 약 25 내지 약 50 염기쌍, 약 100 내지 약 200 염기쌍, 약 150 내지 약 180 염기쌍, 약 150 염기쌍, 약 155 염기쌍, 약 160 염기쌍, 약 165 염기쌍, 약 170 염기쌍, 약 175 염기쌍, 또는 약 180 염기쌍으로 단편화될 수 있다.
- [0101] 본원에 고려된 조성물 및 방법의 특정 구현예에 사용하기에 적합한 DNA를 단편화시키기 위한 예시적 방법은 다음을 포함하지만 이에 제한되지 않는다: 진단 가공, 초음파 처리, 제한 효소 분해를 포함하는 효소적 분해, 및 다른 방법. 특정 구현예에서, DNA를 단편화시키기 위해 당업계에 공지된 임의의 방법은 본 발명과 함께 사용될 수 있다.
- [0102] 본원에 고려되는 조성물 및 방법의 특정 구현예에서 사용하기에 적합한 게놈 DNA 및 RNA (cDNA를 생성하기 위한)의 예시적 공급원은 하기로 이루어진 군으로부터 선택된 생물학적 샘플을 포함하지만 이에 제한되지 않는다: 뇌 조직, 골 조직, 안구 조직, 후각 조직, 근육 조직, 심장 조직, 폐 조직, 간 조직, 췌장 조직, 신장 조직, 위 조직, 장 조직, 결장 조직, 혈액, 피부, 모발, 모낭, 침, 경구 점막, 질 점막, 땀, 눈물, 상피 조직, 소변, 정액 (semen), 정액 (seminal fluid), 정액장, 전립선액, 사정전액 (쿠퍼액), 배설물, 생검, 복수, 뇌척수액, 림프액, 및 조직 추출물 샘플 또는 생검 샘플 등.
- [0103] 특정 구현예에서, DNA는 cfDNA이다. cfDNA의 크기 분포는 약 150 bp 내지 약 180 bp 단편의 범위이다. 단편화는 엔도뉴클레오타이드분해 및/또는 엑소뉴클레오타이드분해 활성의 결과일 수 있고 cfDNA의 정확하고 믿을 수 있고 강력한 분석에 대해 엄청난 과제를 부여한다. cfDNA를 분석하기 위한 또 다른 과제는 약 15분 정도로 혈류내 이의 짧은 반감기이다. 임의의 특정 이론에 국한되는 것 없이, 본 발명은 부분적으로 cfDNA의 분석이 "액체 생검"과 같고 현재 생물학적 과정의 실시간 스냅샷임을 고려한다.
- [0104] 일부 구현예에서, 혈장 분획으로부터 단리된 cfDNA는 수거 프로토콜 동안에 용해하는 핵화된 혈액 세포로부터 유리되는 긴 (>10 킬로베이스 쌍) 고분자량 게놈 DNA로 실질적으로 오염될 수 있다. 이러한 긴 오염 DNA는 단편화되지 않은 상태로 남아있는 경우 잘 클로닝하지 않고 증폭되지 않으며 따라서 다운스트림 라이브러리 제조 동안에 상실된다. 그러나, 특정 구현예에서, DNA 단편화의 부재에서, 본원에서 고려되는 고효율 DNA 라이브러리 작제 방법은 DNA 표본 중에 존재하는 단편 크기의 수거물로부터 보다 짧은 (<1000 bp) 단편을 선택적으로 클로닝한다. 임의의 특정 이론에 국한되는 것 없이, 길고 짧은 단편의 블렌드인 DNA 표본으로부터의 짧은 cfDNA 단편의 선택적 클로닝은 액체 생검의 작제에서 유리하다.
- [0105] 특정 구현예에서 cfDNA를 이로부터 단리하기에 적합한 공급원인 생물학적 샘플의 예시적 예는 양수, 혈액, 혈장, 혈청, 정액 (semen), 림프액, 뇌척수액, 안구액, 소변, 침, 점액 및 땀을 포함하지만 이에 제한되지 않는다.
- [0106] 특정 구현예에서, 생물학적 샘플은 혈액 또는 혈장이다.
- [0107] 특정 구현예에서, DNA 샘플은 매립된 조직, 예를 들어, FFPE 또는 미세 바늘 흡인물로부터, 존재하는 마이크로바이옴 서열에 관한 정보를 얻기 위해 의도된 스왑으로부터, 모발, 의류, 지문 등과 같은 법의학 샘플로부터, 또는 낮은-인풋 DNA 샘플로부터 라이브러리를 작제하기 위해 특히 효율적인 본원에 고려된 라이브러리 작제 방법을 필요로 하는 DNA의 임의의 다른 공급원으로부터 유래될 수 있다.
- [0108] 특정 구현예에서, 당업자에게 공지된 시판되는 키트 및 다른 방법을 사용하여 환자의 생물학적 샘플로부터 또는 이전에 수득되고 임의로 안정화된 생물학적 샘플로부터 예를 들어, EDTA, EGTA, 또는 2가 양이온에 특이적인 다른 킬레이팅 제제를 포함하지만 이에 제한되지 않는 효소 킬레이팅 제제의 동결 및/또는 첨가에 의해 직접적으로 cfDNA를 단리할 수 있다.
- [0109] **(b) 인풋 DNA의 탈인산화**
- [0110] 특정 구현예에서, 인풋 DNA, 예를 들어, 표적 DNA 단편은 먼저 말단 포스페이트 잔기를 제거하는 열-불안정 포스파타제로 처리한다. 예를 들어, 도 4a를 참조한다.
- [0111] 본원에 고려된 조성물 및 방법의 특정 구현예에 사용하기에 적합한 열-불안정 포스파타제의 예시적 예는 APex™ 열-불안정 알칼린 포스파타제 (Epicentre Biotechnologies), NTPhos™ 열불안정 포스파타제 (Epicentre Biotechnologies), HK™ 열불안정 포스파타제 (Epicentre Biotechnologies), 및 새우 알칼린 포스파타제 (SAP;

NEB)를 포함하지만 이에 제한되지 않는다.

- [0112] 하나의 구현예에서, 열-불안정 포스파타제는 SAP이다.
- [0113] **(c) 표적 DNA 단편에서 DNA 손상을 역전시키는 것**
- [0114] 특정 구현예에서, 인풋 DNA 또는 탈인산화된 DNA는 또한 사이토신의 우라실로의 탈아민화, 구아닌으로의 산화적 첨가, 티미딘 이량체, 무염기성 부위를 유도하는 염기의 상실, 이중나선 DNA 등의 하나의 가닥 상에 Nick 또는 겹과 같은 DNA 손상의 통상의 공급원을 역전시키는 하나 이상의 효소로 처리한다. 예를 들어, 도 4a를 참조한다.
- [0115] 하나의 구현예에서, DNA에 대한 내부 손상은 하기의 효소 중 하나 이상을 포함하는 조성물을 사용하여 역전시킨다: Taq DNA 리가제, 엔도뉴클레아제 IV, Bst DNA 폴리머라제, Fpg (8-옥소구아닌 DNA 글리코실라제), 우라실-DNA 글리코실라제 (UDG), T4 PDG (T4 엔도뉴클레아제 V), 엔도뉴클레아제 VIII, 및 T4 DNA 폴리머라제.
- [0116] 하나의 구현예에서, DNA에 대한 내부 손상은 Taq DNA 리가제, 엔도뉴클레아제 IV, Bst DNA 폴리머라제, Fpg, 우라실-DNA 글리코실라제 (UDG), T4 PDG (T4 엔도뉴클레아제 V), 엔도뉴클레아제 VIII, 및 T4 DNA 폴리머라제를 포함하는 조성물을 사용하여 역전시킨다.
- [0117] **(d) 말단-복구된 DNA를 생성시키는 것**
- [0118] 특정 구현예에서, 본원에 고려된 조성물 및 방법은 말단-복구된 DNA 단편을 생성시킴을 포함한다. 특정 구현예에서, DNA 단편은 말단-복구되어 평활 말단, 5'-오버행, 또는 3'-오버행을 갖는 말단-복구된 DNA 단편을 생성한다. 예를 들어, 도 4a를 참조한다. 특정 구현예에서, DNA는 cfDNA이다.
- [0119] 일부 구현예에서, 말단-복구된 DNA는 평활 말단을 함유한다. 일부 구현예에서, 말단-복구된 DNA는 평활 말단을 함유하도록 프로세싱된다. 바람직한 구현예에서, DNA 단편은 하나 이상의 말단-복구 효소에 의해 말단-복구되어 평활 말단을 갖는 말단-복구된 DNA 단편을 생성시킨다.
- [0120] 본원에 고려된 조성물 및 방법의 특정 구현예에서 평활-말단화된 DNA 단편을 생성시키기에 적합한 말단-복구 효소의 예시적 예는 중합 활성 및 3' → 5' 엑소뉴클레아제 활성을 보유하지만 5' → 3' 엑소뉴클레아제 활성이 없는 DNA 폴리머라제 (예를 들어, T4 DNA 폴리머라제, DNA 폴리머라제 I의 Klenow 단편, 등)를 포함한다. DNA 폴리머라제를 사용하여 5' 오버행을 채우거나 3' 오버행을 "초우 백"하고 평활 말단을 갖는 DNA 단편을 잔류시킨다.
- [0121] 일부 구현예에서, 말단-복구된 DNA의 평활 말단은 단일 염기쌍 오버행을 함유하도록 추가로 변형시킨다. 일부 구현예에서, 평활 말단을 함유하는 말단-복구된 DNA는 추가로 프로세싱하여 아데닌 (A)/티민 (T) 오버행을 함유할 수 있다. 일부 구현예에서, 평활 말단을 함유하는 말단-복구된 DNA는 추가로 프로세싱하여 단일 염기쌍 오버행으로서 아데닌 (A)/티민 (T) 오버행을 함유할 수 있다. 일부 구현예에서, 말단-복구된 DNA는 비-주형 3' 오버행을 갖는다. 일부 구현예에서, 말단-복구된 DNA는 3' 오버행을 함유하도록 프로세싱된다. 일부 구현예에서, 말단-복구된 DNA는 3' 오버행을 함유하도록 말단 트랜스퍼라제 (TdT)를 사용하여 프로세싱된다. 일부 구현예에서, G-꼬리는 TdT에 의해 부가될 수 있다. 일부 구현예에서, 말단-복구된 DNA는 임의의 공지된 제한 효소를 사용한 (예를 들어, 효소 Sau3A 등을 사용한) 부분 분해를 사용하여 오버행 말단을 함유하도록 프로세싱된다.
- [0122] **(e) 프리-어댑터의 말단-복구된 DNA로의 연결**
- [0123] 특정 구현예에서, 본원에 고려된 조성물 및 방법은 말단-복구된 DNA의 각각의 말단으로 이중나선 DNA 프리-어댑터를 연결시킴을 포함한다.
- [0124] 본원에 사용된 바와 같은 용어 "프리-어댑터"는 연결 가닥 올리고뉴클레오타이드 및 파트너 가닥 올리고뉴클레오타이드를 포함하는 이중가닥 DNA 분자 또는 DNA 이중나선을 언급한다. 프리-어댑터는 임의의 적합한 리가제를 사용하여 말단-복구된 DNA 단편으로 연결될 수 있다. 하나의 구현예에서, 리가제는 T4 DNA 리가제이다. 예를 들어, 도 4b 및 5를 참조한다.
- [0125] 상기 "연결 가닥 올리고뉴클레오타이드"는 5' 포스페이트를 포함하는 폴리뉴클레오타이드이고 이는 말단-복구된 DNA 단편의 각각의 3' 말단에 연결될 수 있다.
- [0126] "파트너 가닥 올리고뉴클레오타이드"는 연결 가닥 올리고뉴클레오타이드의 뉴클레오타이드 일부 또는 전체에 상보적이고 어닐링한다. 파트너 가닥 올리고뉴클레오타이드는 파트너 가닥 올리고뉴클레오타이드가 또 다른 어댑

터로 또는 표적 DNA 단편의 인산화된 5' 말단으로 연결되지 못하도록 하거나 이를 실질적으로 억제하는 이의 3' 말단에서의 변형을 포함한다. 연결을 차단할 수 있는 파트너 가닥의 3' 말단의 화학적 변형은 디데옥시 리보스 뉴클레오타이드 유사체, 2-하이드록실 테옥시리보스 리보스 유사체, 및 리보스 당에 대한 광범위한 화학적 변형을 포함하지만 이에 제한되지 않는다.

[0127] 여러 고려사항은 프리-어댑터에 사용되는 서열 디자인 및 연결 가닥 올리고뉴클레오타이드의 내용물에 도입한다. 연결 가닥 올리고뉴클레오타이드는 DNA 리가제가 현재 합성 능력의 한계점을 밀어내는 올리고뉴클레오타이드 (>200 nt)에 대해 활성 (~5 nt)인 온도에서 안정한 DNA 이중나선을 형성하기 위해 요구되는 최소 길이로부터 길이가 다양할 수 있다. 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 약 8 내지 약 60 뉴클레오타이드 또는 약 8 내지 약 15 뉴클레오타이드이다.

[0128] DNA 단편의 NGS 분석과 관련된 추가의 고려 사항으로서, 연결 가닥에 의해 혼입된 DNA 염기는 DNA 서열분석 수행 전반에 걸쳐 DNA 염기 콜을 측정하기 위해 서열분석 장비에 의해 사용된다. 장비 및 이들 장비의 소프트웨어는 모든 4개의 DNA 염기가 서열분석되는 초기 8 - 15 뉴클레오타이드의 길이 전반에 걸친 모든 염기 위치에 존재하고 이것은 흔히 연결 어댑터 가닥에 매립된 염기를 포함할 것을 요구한다. 이러한 이유 때문에, 연결 가닥 서열의 길이에 걸친 모든 4개의 염기를 상호적으로 갖는 4개의 연결 가닥 세트가 흔히 사용된다. 상기 연결 가닥 올리고뉴클레오타이드의 비제한적인 예는 표 1 및 2에 나타난다.

[0129] 다양한 다른 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하기의 요소들을 포함한다: (i) 단일-프라이머 라이브러리 증폭을 위한 PCR 프라이머 결합 부위; (ii) 특유하게 동정된 각각의 서열분석 판독에 작용하는 5 뉴클레오타이드 판독 코드; (iii) 샘플 동정 서열로서 작용하여 서열분석 수행 내에 샘플 멀티플렉싱을 가능하게 하여 서열분석 판독에서 적당한 염기 콜의 측정을 가능하게 하고 파트너 가닥 올리고뉴클레오타이드로의 하이브리드화를 위한 앵커로서 작용하는 8 내지 15 뉴클레오타이드 앵커 서열.

[0130] 다양한 다른 구현예에서, 연결 가닥 올리고뉴클레오타이드는 샘플 동정 서열로서 작용하여 서열분석 수행 내에 샘플 멀티플렉싱을 가능하게 하여 서열분석 판독에서 적당한 염기 콜의 측정을 가능하게 하고 파트너 가닥 올리고뉴클레오타이드로의 하이브리드화를 위한 앵커로서 작용하는 8 내지 15 뉴클레오타이드 앵커 서열을 포함한다.

[0131] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 PCR 프라이머 서열, 하나 이상의 판독 코드, 하나 이상의 샘플 코드, 하나 이상의 앵커 서열, 또는 효율적인 연결 기질인 2개 이상의 3' 뉴클레오타이드를 포함한다. 추가의 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 서열분석 프라이머 결합 부위를 추가로 포함한다.

[0132] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 DNA 라이브러리의 증폭을 위한 하나 이상의 PCR 프라이머 결합 서열을 포함한다. 하나의 구현예에서, PCR 프라이머 결합 서열은 약 12 내지 약 40 뉴클레오타이드, 약 18 내지 약 40 뉴클레오타이드, 약 20 내지 약 35 뉴클레오타이드, 또는 약 20 내지 약 30 뉴클레오타이드이다. 또 다른 구현예에서, PCR 프라이머 결합 서열은 약 12 이상 뉴클레오타이드, 약 13 이상 뉴클레오타이드, 약 14 이상 뉴클레오타이드, 약 15 이상 뉴클레오타이드, 약 16 이상 뉴클레오타이드, 약 17 이상 뉴클레오타이드, 약 18 이상 뉴클레오타이드, 약 19 이상 뉴클레오타이드, 약 20 이상 뉴클레오타이드, 약 21 이상 뉴클레오타이드, 약 22 이상 뉴클레오타이드, 약 23 이상 뉴클레오타이드, 약 24 이상 뉴클레오타이드, 약 25 이상 뉴클레오타이드, 약 26 이상 뉴클레오타이드, 약 27 이상 뉴클레오타이드, 약 28 이상 뉴클레오타이드, 약 29 이상 뉴클레오타이드, 약 30 이상 뉴클레오타이드, 약 31 이상 뉴클레오타이드, 약 32 이상 뉴클레오타이드, 약 33 이상 뉴클레오타이드, 약 34 이상 뉴클레오타이드, 약 35 이상 뉴클레오타이드, 약 36 이상 뉴클레오타이드, 약 37 이상 뉴클레오타이드, 약 38 이상 뉴클레오타이드, 약 39 이상 뉴클레오타이드, 또는 약 40 이상 뉴클레오타이드이다.

[0133] 하나의 구현예에서, PCR 프라이머 결합 서열은 약 25 뉴클레오타이드이다.

[0134] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 판독 코드 서열을 포함한다. 본원에 사용된 바와 같은 용어 "판독 코드"는 특유의 서열분석 판독을 동정하기 위해 사용되는 폴리뉴클레오타이드를 언급한다. 하나의 구현예에서, 판독 코드는 뉴클레오타이드의 무작위 서열이다. 하나의 구현예에서, 판독 코드는 약 1 이상 뉴클레오타이드, 약 2 이상 뉴클레오타이드, 약 3 이상 뉴클레오타이드, 약 4 이상 뉴클레오타이드, 약 5 이상 뉴클레오타이드, 약 6 이상 뉴클레오타이드, 약 7 이상 뉴클레오타이드, 약 8 이상 뉴클레오타이드, 약 9 이상 뉴클레오타이드, 약 10 이상 뉴클레오타이드이다.

- [0135] 비제한적인 예로서, 5 뉴클레오타이드 판독 코드는 256의 가능한 특유의 서열로 이루어지고 여기서 각각의 선택된 코드는 세트내 모든 다른 코드와 2 뉴클레오타이드 상이하다. 이러한 특징은 특유하고 독특한 판독이 코드 영역에서 서열분석 오차로 인해 고유한 것으로 나타난 판독으로부터 구분되도록 할 수 있다. 특정 구현예에서, 특정 서열 조합으로 인해 경험적으로 어댑터 기능을 방해하는 것으로 결정된 코드는 사용으로부터 배제될 수 있는데, 예를 들어, 256의 7개 코드는 G 뉴클레오타이드의 과잉제공을 가졌고 배제되었다.
- [0136] 다른 구현예에서, 5, 6, 7, 8, 9, 10 이상의 뉴클레오타이드의 각각의 판독 코드는 모든 다른 판독 코드와 2, 3, 4, 또는 5 뉴클레오타이드 만큼 상이할 수 있다.
- [0137] 하나의 구현예에서, 판독 코드는 약 5 뉴클레오타이드이고 임의로 모든 다른 판독 코드와 2 뉴클레오타이드 만큼 상이하다.
- [0138] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 샘플 코드 서열을 포함한다. 본원에 사용된 바와 같은 용어 "샘플 코드"는 샘플을 동정하기 위해 사용되는 폴리뉴클레오타이드를 언급한다. 샘플 코드는 또한 각각의 샘플 코드가 샘플에 고유하기 때문에 멀티플렉스 서열분석 반응을 확립하는데 유용하고 따라서 멀티플렉스된 서열분석 반응 내에 특정 샘플로부터의 판독을 동정하기 위해 사용될 수 있다.
- [0139] 하나의 구현예에서, 샘플 코드는 약 1 이상, 약 2 이상 뉴클레오타이드, 약 3 이상 뉴클레오타이드, 약 4 이상 뉴클레오타이드, 또는 약 5 이상 뉴클레오타이드인 서열을 포함한다. 또 다른 구현예에서, 2, 3, 4, 5 이상의 뉴클레오타이드의 각각의 샘플 코드는 모든 다른 샘플 코드와 2, 3, 4, 또는 5 뉴클레오타이드 만큼 상이할 수 있다.
- [0140] 하나의 구현예에서, 샘플 코드는 약 3 뉴클레오타이드이고 다른 샘플에 사용되는 모든 다른 샘플 코드와 2 뉴클레오타이드 만큼 상이하다.
- [0141] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하나 이상의 앵커 서열을 포함한다. 본원에 사용된 바와 같은 "앵커 서열"은 파트너 가닥 올리고뉴클레오타이드에 하이브리드화하고 하기의 성질을 포함하는 적어도 8 뉴클레오타이드, 적어도 10 뉴클레오타이드, 적어도 12 뉴클레오타이드, 적어도 14 뉴클레오타이드, 또는 적어도 16 뉴클레오타이드의 뉴클레오타이드 서열을 언급한다: (1) 각각의 앵커 서열은 연장 내 각각의 부위에서 4개의 가능한 DNA 염기 각각을 총체적으로 나타내는 4개의 앵커 서열 계열의 일부이고; 이러한 특징의 균형화된 염기 제공은 특정 구현예에서 서열분석 판독에서 적당한 염기 콜링을 측정하기 위해 유용하고; (2) 각각의 앵커 서열은 동등한 수의 A+C 및 G+T로 구성되고, 따라서, 각각의 앵커 서열은 대략적으로 4개 세트에서 모든 다른 앵커 서열과 동일한 용융 온도 및 이중나선 안정성을 공유한다. 하나의 구현예에서, 앵커 서열 또는 이의 일부는 또한 샘플을 동정하고, 샘플이 서열분석 수행 내 멀티플렉싱할 수 있고, 서열분석 판독에서 적당한 염기 콜링의 측정을 가능하게 하는 작용을 하고 파트너 가닥 올리고뉴클레오타이드로의 하이브리드화를 위한 앵커로서 작용한다.
- [0142] 추가로, 여러 고려사항은 비-연결 파트너 가닥 올리고뉴클레오타이드의 디자인에 포함된다. 파트너 가닥 올리고뉴클레오타이드는 인산화된 평활 말단을 형성하는 영역에서 연결 가닥 올리고뉴클레오타이드에 적어도 부분적으로 상보적 (>5 nt) 이다. 두번째로, 파트너 가닥 올리고뉴클레오타이드의 3' 말단은 특히 자가-연결된 어댑터 이량체의 형성에서 올리고뉴클레오타이드가 연결 기질이 되도록 하는 것을 차단하거나 실질적으로 억제하도록 변형된다. 파트너 가닥 올리고뉴클레오타이드는 연결이 수행되는 온도 ($\leq 22\text{ }^{\circ}\text{C}$)에서 연결 가닥과 안정한 이중나선을 형성하도록 디자인되지만 또한 복구 올리고뉴클레오타이드가 어댑터에 혼입되는 온도 ($\geq 37\text{ }^{\circ}\text{C}$)에서 연결 가닥 올리고뉴클레오타이드로부터 분리되도록 디자인된다. 이러한 디자인 고려사항은 반응이 복구 올리고뉴클레오타이드에 의해 매개되는, 연결로부터 어댑터 완성 단계로 전환되기 때문에 어댑터/말단-복구된 DNA 복합체의 생성에서 도 4b 및 5에 나타난 분리된 파트너 가닥 올리고뉴클레오타이드로서 도시된다.
- [0143] 특정 구현예에서, 본원에 고려되는 조성물 및 방법은 연결 단계를 포함하고, 여기서, 프리-어댑터는 말단-복구된 DNA에 연결되어 "태그된" DNA 라이브러리를 생성한다. 일부 구현예에서, 프리-어댑터의 단일 종이 사용된다. 일부 구현예에서, 프리-어댑터의 2개, 3개, 4개 또는 5개 종이 사용된다. 일부 구현예에서, 동일한 서열의 프리-어댑터는 단편화된 말단-복구된 DNA의 각각의 말단으로 연결된다.
- [0144] 하나의 구현예에서, 다수의 프리-어댑터 종은 말단-복구된 DNA 라이브러리에 연결된다. 다수의 프리-어댑터 각각은 DNA 라이브러리의 증폭을 위해 하나 이상의 프라이머 결합 증폭, 하나 이상의 판독 코드 서열, 샘플 멀티플렉싱을 위한 하나 이상의 서열, 하나 이상의 앵커 서열, 또는 DNA 서열분석을 위한 하나 이상의 서열을 포함할 수 있다.

[0145] (f) 어댑터/말단-복구된 DNA 복합체의 형성

- [0146] 특정 구현예에서, 본원에 고려된 조성물 및 방법은 프리-어댑터/말단-복구된 DNA 복합체로부터 파트너 가닥 올리고뉴클레오타이드를 치환하고 상기 치환된 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 DNA 복합체를 생성함을 포함한다. 예를 들어, 도 3을 참조한다. 특정 구현예에서, 어댑터의 디자인은 단일 프라이머 또는 이원 프라이머 증폭 전략이 가능하도록 조작될 수 있다. 예를 들어, 도 4a 및 5를 참조한다.
- [0147] 특정 구현예에서, 본원에 고려되는 조성물 및 방법은 연결 단계를 포함하고, 여기서, 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는 어댑터는 말단-복구된 DNA에 연결되어 "태그된" DNA 라이브러리를 생성한다. 일부 구현예에서, 어댑터의 단일 종이 사용된다. 일부 구현예에서, 어댑터의 2개, 3개, 4개 또는 5개 종이 사용된다. 일부 구현예에서, 동일한 서열의 어댑터는 단편화된 말단-복구된 DNA의 각각의 말단으로 연결된다.
- [0148] 파트너 가닥 올리고뉴클레오타이드의 디자인 고려사항은 이것이 복구 올리고뉴클레오타이드가 연결 가닥 올리고뉴클레오타이드로 어닐링되는 온도 (예를 들어, > 37 °C)에서 및 효소 단계가 복구 올리고뉴클레오타이드를 어댑터/말단-복구된 DNA 복합체로 혼입되어 연속성 이중가닥 DNA 라이브러리 분자를 생성하기 위해 수행되는 온도 (예를 들어, > 37 °C)에서 연결 가닥 올리고뉴클레오타이드로 분리되기 때문에 프리-어댑터/말단-복구된 DNA 복합체로부터 치환되도록 한다.
- [0149] 본원에 사용된 바와 같은 용어 "복구 올리고뉴클레오타이드"는 연결 가닥 올리고뉴클레오타이드의 뉴클레오타이드 일부 또는 전부에 상보적이고 어닐링하는 폴리뉴클레오타이드 서열을 언급한다. 복구 올리고뉴클레오타이드는 DNA 리가제가 현재 합성 능력의 한계점을 밀어내는 올리고뉴클레오타이드 (> 200 nt)에 대해 활성 (~8 nt)인 온도에서 안정한 DNA 이중나선을 형성하기 위해 요구되는 최소 길이로부터 길이가 다양할 수 있다. 특정 구현예에서, "복구 올리고뉴클레오타이드"는 연결 가닥 올리고뉴클레오타이드에 필수적으로 존재하지 않는 추가의 기능성 DNA 서열을 포함한다.
- [0150] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 약 8 내지 약 15 뉴클레오타이드이고 복구 올리고뉴클레오타이드는 35 내지 60 뉴클레오타이드이다. 상기 디자인에서, 리간드 가닥 올리고뉴클레오타이드의 서열은 프라이머 연장에 의해 연장되고 복구 올리고뉴클레오타이드에 상보적인 뉴클레오타이드 서열을 생성한다. 이러한 디자인은 동일한 PCR 프라이머 결합 부위를 생성한다. 동일한 PCR 프라이머 결합 부위는 단일 프라이머 라이브러리 증폭 전략을 가능하게 한다. 예를 들어, 도 3d 및 4a를 참조한다.
- [0151] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 약 35 내지 약 60 뉴클레오타이드이고 복구 올리고뉴클레오타이드는 연결 가닥 뉴클레오타이드에 완전히 상보적이다. 동일한 PCR 프라이머 결합 부위는 단일 프라이머 라이브러리 증폭 전략을 가능하게 한다. 예를 들어, 도 4a를 참조한다.
- [0152] 특정 구현예에서, 연결 가닥 올리고뉴클레오타이드는 약 35 내지 약 60 뉴클레오타이드이고 복구 올리고뉴클레오타이드는 약 35 내지 약 60 뉴클레오타이드이고 2개의 올리고뉴클레오타이드는 PCR 프라이머 결합 부위를 제외하고는 상보적이다. 상이한 PCR 프라이머 결합 부위는 이원 프라이머 라이브러리 증폭 전략을 가능하게 한다. 예를 들어, 도 5를 참조한다.
- [0153] 바람직한 구현예에서, 연결 가닥 올리고뉴클레오타이드는 하기의 요소들을 포함한다: (i) 단일-프라이머 라이브러리 증폭을 위한 PCR 프라이머 결합 부위; (ii) 특유하게 동정된 각각의 서열분석 판독에 작용하는 5 뉴클레오타이드 판독 코드; (iii) 연결 가닥 올리고뉴클레오타이드의 앵커 서열에 부분적으로 또는 완전히 상보적인 8 내지 15 뉴클레오타이드 앵커 서열.
- [0154] 다른 구현예에서, 연결 가닥 올리고뉴클레오타이드는 연결 가닥 올리고뉴클레오타이드의 앵커 서열에 부분적으로 또는 완전히 상보적인 8 내지 15 뉴클레오타이드 앵커 서열을 포함한다.
- [0155] 특정 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 PCR 프라이머 서열, 하나 이상의 판독 코드, 하나 이상의 샘플 코드, 하나 이상의 앵커 서열 또는 효율적인 연결 기질인 2개 이상의 3' 뉴클레오타이드를 포함한다. 추가의 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 서열분석 프라이머 결합 부위를 추가로 포함한다.
- [0156] 특정 구현예에서, 복구 올리고뉴클레오타이드는 (i) 연결 가닥 올리고뉴클레오타이드에서 PCR 프라이머 결합 부위에 상보적인 하나 이상의 PCR 프라이머 결합 서열 (단일-프라이머 DNA 라이브러리 증폭을 가능하게 하는) 또

는 (ii) 연결 가닥 올리고뉴클레오타이드에서 PCR 프라이머 결합 부위에 상보적이지 않은 하나 이상의 PCR 프라이머 결합 서열 (이원-프라이머 DNA 라이브러리 증폭을 가능하게 하는)을 포함한다. 하나의 구현예에서, PCR 프라이머 결합 서열은 약 12 내지 약 40 뉴클레오타이드, 약 18 내지 약 40 뉴클레오타이드, 약 20 내지 약 35 뉴클레오타이드, 또는 약 20 내지 약 30 뉴클레오타이드이다. 또 다른 구현예에서, PCR 프라이머 결합 서열은 약 12 이상 뉴클레오타이드, 약 13 이상 뉴클레오타이드, 약 14 이상 뉴클레오타이드, 약 15 이상 뉴클레오타이드, 약 16 이상 뉴클레오타이드, 약 17 이상 뉴클레오타이드, 약 18 이상 뉴클레오타이드, 약 19 이상 뉴클레오타이드, 약 20 이상 뉴클레오타이드, 약 21 이상 뉴클레오타이드, 약 22 이상 뉴클레오타이드, 약 23 이상 뉴클레오타이드, 약 24 이상 뉴클레오타이드, 약 25 이상 뉴클레오타이드, 약 26 이상 뉴클레오타이드, 약 27 이상 뉴클레오타이드, 약 28 이상 뉴클레오타이드, 약 29 이상 뉴클레오타이드, 약 30 이상 뉴클레오타이드, 약 31 이상 뉴클레오타이드, 약 32 이상 뉴클레오타이드, 약 33 이상 뉴클레오타이드, 약 34 이상 뉴클레오타이드, 약 35 이상 뉴클레오타이드, 약 36 이상 뉴클레오타이드, 약 37 이상 뉴클레오타이드, 약 38 이상 뉴클레오타이드, 약 39 이상 뉴클레오타이드, 또는 약 40 이상 뉴클레오타이드이다.

- [0157] 하나의 구현예에서, PCR 프라이머 결합 서열은 약 25 뉴클레오타이드이다.
- [0158] 특정 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 판독 코드 서열을 포함한다. 하나의 구현예에서, 판독 코드는 뉴클레오타이드의 무작위 서열이다. 하나의 구현예에서, 판독 코드는 약 1 이상 뉴클레오타이드, 약 2 이상 뉴클레오타이드, 약 3 이상 뉴클레오타이드, 약 4 이상 뉴클레오타이드, 약 5 이상 뉴클레오타이드, 약 6 이상 뉴클레오타이드, 약 7 이상 뉴클레오타이드, 약 8 이상 뉴클레오타이드, 약 9 이상 뉴클레오타이드, 약 10 이상 뉴클레오타이드이다.
- [0159] 비제한적인 예로서, 5 뉴클레오타이드 판독 코드는 256의 가능한 특유의 서열로 이루어지고 여기서 각각의 선택된 코드는 세트내 모든 다른 코드와 2 뉴클레오타이드 상이하다. 이러한 특징은 특유하고 독특한 판독이 코드 영역에서 서열분석 오차로 인해 고유한 것으로 나타난 판독으로부터 구분되도록 할 수 있다. 특정 구현예에서, 특정 서열 조합으로 인해 경험적으로 어댑터 기능을 방해하는 것으로 결정된 코드는 사용으로부터 배제될 수 있는데, 예를 들어, 256의 7개 코드는 G 뉴클레오타이드의 과잉제공을 가졌고 배제되었다.
- [0160] 다른 구현예에서, 5, 6, 7, 8, 9, 10 이상의 뉴클레오타이드의 각각의 판독 코드는 모든 다른 판독 코드와 2, 3, 4, 또는 5 뉴클레오타이드 만큼 상이할 수 있다.
- [0161] 하나의 구현예에서, 판독 코드는 약 5 뉴클레오타이드이고 임의로 모든 다른 판독 코드와 2 뉴클레오타이드 만큼 상이하다.
- [0162] 특정 구현예에서, 복구 올리고뉴클레오타이드는 하나 이상의 샘플 코드 서열을 포함한다. 하나의 구현예에서, 샘플 코드는 약 1 이상, 약 2 이상 뉴클레오타이드, 약 3 이상 뉴클레오타이드, 약 4 이상 뉴클레오타이드, 또는 약 5 이상 뉴클레오타이드인 서열을 포함한다. 또 다른 구현예에서, 2, 3, 4, 5 이상의 뉴클레오타이드의 각각의 샘플 코드는 모든 다른 샘플 코드와 2, 3, 4, 또는 5 뉴클레오타이드 만큼 상이할 수 있다.
- [0163] 하나의 구현예에서, 샘플 코드는 약 3 뉴클레오타이드이고 다른 샘플에 사용되는 모든 다른 샘플 코드와 2 뉴클레오타이드 만큼 상이하다.
- [0164] 특정 구현예에서, 복구 올리고뉴클레오타이드는 연결 가닥 올리고뉴클레오타이드의 하나 이상의 앵커 서열에 상보적인 하나 이상의 앵커 서열을 포함한다.
- [0165] 임의의 특정 이론에 국한되는 것 없이, 적어도 2개의 예시적 전략은 복구 올리고뉴클레오타이드를 어댑터/말단-복구 DNA 복합체에 혼입시키기 위해 고려된다.
- [0166] 하나의 구현예에서, 파트너 가닥 올리고뉴클레오타이드는 프리-어댑터/말단-복구된 DNA 복합체로부터 치환되고; 복구 올리고뉴클레오타이드는 부가되고 연결 가닥에 어닐링되도록 하고; 폴리뉴클레오타이드 키나제, 예를 들어, T4 폴리뉴클레오타이드 키나제는 포스페이트 그룹을 말단-복구된 DNA 단편의 5' 말단에 부가하기 위해 사용되고; DNA 리가제는 복구 올리고뉴클레오타이드의 5' 말단과 말단-복구된 DNA 단편의 3' 말단 사이에 존재하는 틈을 복구하기 위해 사용된다. 특정 구현예에서, DNA 리가제는 Taq DNA 리가제, 이. 콜라이 DNA 리가제, 9도 노쓰 리가제 (NEB), 및 인산화된 틈을 밀봉할 수 있는 임의의 다른 리가제를 포함하지만 이에 제한되지 않는, 넓은 범위의 온도에 걸쳐 활성을 갖는 열-안정한 너-특이적 리가제이다. 예를 들어, 도 3a 및 3b를 참조한다.
- [0167] 또 다른 구현예에서, 파트너 가닥 올리고뉴클레오타이드는 프리-어댑터/말단-복구된 DNA 복합체로부터

치환되고; 복구 올리고뉴클레오타이드는 부가되고 연결 가닥에 어닐링되도록 하고; 5'→3' 엑소뉴클레아제 활성을 갖는 (및 어떠한 고유 3'→5' 엑소뉴클레아제 활성도 갖지 않는) 낮은 프로세싱의 DNA 폴리머라제는 연결 가닥 올리고뉴클레오타이드의 3' 말단을 연장하고 추가로 5'→3' 엑소뉴클레아제 활성으로 탈인산화된 5' 말단 뉴클레오타이드 및 인접한 뉴클레오타이드를 제거하여 연결 가능한 5' 포스페이트 그룹을 노출시키고 이들을 효소가 분리되는 경우 너 뒤에 남게되는 혼입된 염기로 대체하고; DNA 리가제, 예를 들어, *Taq* DNA 리가제는 너을 복구하기 위해 사용된다.

[0168] 본원에 고려된 조성물 및 방법의 특정 구현예에 사용하기에 적합한 낮은 프로세싱의 DNA 폴리머라제의 예시적 예는 *Taq* DNA 폴리머라제, 및 *BstI* DNA 폴리머라제를 포함하지만 이에 제한되지 않는다.

[0169] **D. DNA 라이브러리 증폭**

[0170] 특정 구현예에서, 본원에 고려된 방법은 DNA 클론 라이브러리 또는 DNA 클론의 라이브러리를 생성하기 위한 DNA 라이브러리의 증폭을 포함한다. 특정 구현예에서, DNA는 cfDNA이다. DNA 라이브러리의 각각의 분자는 말단-복구된 DNA의 각각의 말단에 연결된 어댑터를 포함하고, 각각의 어댑터는 하나 이상의 PCR 프라이머 결합 부위를 포함한다. 하나의 구현예에서, 상이한 어댑터는 말단-복구된 DNA의 상이한 말단에 연결된다.

[0171] 하나의 구현예에서, 동일한 어댑터는 DNA의 양 말단에 연결된다. 동일한 어댑터의 말단-복구된 DNA의 양 말단으로의 연결은 단일 프라이머 서열을 사용한 PCR 증폭을 가능하게 한다. 특정 구현예에서, 어댑터 연결된-DNA 라이브러리의 일부는 단일 프라이머 서열 구동 증폭과 함께 표준 PCR 기술을 사용하여 증폭될 것이다. 하나의 구현예에서, 단일 프라이머 서열은 임의로 표준 이온 강도 조건하에서 ≥ 55 °C의 프로젝트된 T_m 과 함께 약 25 뉴클레오타이드이다.

[0172] 하나의 구현예에서, 말단-복구된 DNA 단편의 3' 말단에 연결된 어댑터는 말단-복구된 DNA 단편의 5' 말단에 연결된 어댑터로부터 상이한 PCR 프라이머 결합 부위를 포함한다. 특정 구현예에서, 어댑터 연결된-DNA 라이브러리의 일부는 2개의 프라이머 구동 증폭과 함께 표준 PCR 기술을 사용하여 증폭될 것이다.

[0173] 특정 구현예에서, 피코그램의 초기 DNA 라이브러리는 마이크로그램의 DNA 클론으로 증폭되고 이는 10,000-배 증폭을 의미한다. 증폭된 생성물의 양은 당업계에 공지된 방법, 예를 들어, Qubit 2.0 또는 Nanodrop 장비 상에서의 정량을 사용하여 측정될 수 있다.

[0174] **E. DNA의 유전학적 분석 방법**

[0175] 다양한 구현예에서, DNA의 유전학적 분석을 위한 방법이 제공된다. 특정 구현예에서, DNA는 cfDNA이다. cfDNA는 혈장 또는 다른 체액에서 발견되는 세포-부재 DNA이다.

[0176] 특정 구현예에서, DNA의 유전학적 분석 방법은 DNA 라이브러리를 생성하고 증폭시키고, DNA 라이브러리에서 게놈 등가물의 수를 결정하고, 하나 이상의 게놈 표적 유전자좌의 정량적 유전학적 분석을 수행함을 포함한다.

[0177] **1. 게놈 등가물의 수를 결정하는 것**

[0178] 다양한 구현예에서, DNA의 유전학적 분석을 위한 방법은 DNA 클론 라이브러리에서 게놈 등가물의 수를 결정함을 포함한다. 본원에 사용된 바와 같은 용어 "게놈 등가물"은 각각의 라이브러리에서 게놈 카피수를 언급한다. 본원에서 고려되는 조성물 및 방법에 의해 충족되는 중요한 과제는 유전학적 서열에서 희귀 유전학적 돌연변이 또는 차이를 검출하고 분석하기에 충분한 검정 민감성을 성취하는 것이다. 샘플별 기준에 대한 검정 민감성 값을 결정하기 위해, 각각의 샘플에 존재하는 상이하고 구분된 서열의 수는 서열분석 라이브러리에 존재하는 게놈 등가물의 수를 측정함에 의해 측정된다. 민감성을 확립하기 위해, 게놈 등가물의 수는 각각의 샘플 라이브러리를 위해 측정되어야만 한다.

[0179] 게놈 등가물의 수는 서열분석이 수행된 후 qPCR 검정에 의해 또는 생물정보 기반 계수에 의해 결정될 수 있다. 임상 샘플의 프로세스 흐름에서, 게놈 등가물의 qPCR 측정은 DNA 라이브러리를 위한 QC 단계로서 사용된다. 이것은 서열분석 전 검정 민감성에 대한 예측을 확립하고 이의 상응하는 DNA 클론 라이브러리가 요구되는 정도의 게놈 등가물이 없는 경우 샘플이 분석으로부터 배제되도록 한다. 궁극적으로, 게놈 등가물의 생물정보학 기반 계수는 또한 각각의 주어진 DNA 클론 라이브러리에 대한 게놈 등가물을 동정하고- 따라서 검정 민감성 및 거짓 음성 평가를 위해 사용된다.

[0180] 경험적 qPCR 검정 및 통계학적 계수 검정은 매우 상호 관련이 있어야 한다. 서열분석이 DNA 클론 라이브러리에 서열 정도를 밝히는데 실패하는 경우에, DNA 클론 라이브러리 및/또는 추가의 서열분석의 재프로세싱이 요구

될 수 있다.

[0181] 하나의 구현예에서, DNA 클론 라이브러리에서 게놈 등가물은 정량적 PCR (qPCR) 검정을 사용하여 결정된다. 특정 구현예에서, 공지된 농도의 표준 라이브러리는 표준 곡선을 작성하기 위해 사용되고 qPCR 검정으로부터의 측정은 수득한 표준 곡선에 피팅하고 게놈 등가물에 대한 값은 상기 피트로부터 유래된다. 반복체 기반 검정에 의해 측정된 게놈 등가물의 수는 서열분석 수행에서 게놈 등가물 및 생물정보학적 계수된 태그 등가물의 qPCR 평가 간에 보다 일관된 라이브러리 대 라이브러리 수행능과 보다 우수한 정렬을 제공한다.

[0182] 반복체 기반 게놈 등가물 검정에서 사용하기에 적합한 반복체의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 짧은 산재된 핵 요소들 (SINE), 예를 들어, Alu 반복체; 긴 산재된 핵 요소들 (LINE), 예를 들어, LINE1, LINE2, LINE3; 마이크로위성 (microsatellite) 반복체 요소들, 예를 들어, 짧은 탠덤 반복체 (STR), 단순한 서열 반복체 (SSR); 및 포유동물-광범위 산재된 반복체 (MIR).

[0183] 하나의 구현예에서, 반복체는 Alu 반복체이다.

[0184] **2. 정량적 유전학적 분석**

[0185] 다양한 구현예에서, DNA의 유전학적 분석을 위한 방법은 DNA 라이브러리 클론의 하나 이상의 표적 유전학적 유전자좌의 정량적 유전학적 분석을 포함한다. 정량적 유전학적 분석은 하기의 단계 중 하나 이상 또는 모두를 포함한다: 표적 유전학적 유전자좌를 포함하는 DNA 클론을 포획하는 단계; 상기 포획된 표적화된 유전학적 유전자좌를 증폭시키는 단계; 상기 증폭된 포획된 유전학적 유전자좌를 서열분석하는 단계; 및 수득한 서열 판독을 생물정보학적으로 분석하는 단계.

[0186] **(a) 표적 유전학적 유전자좌의 포획**

[0187] 본 발명은 부분적으로 다기능성이고 보다 큰 프로브의 효율 및 신뢰성을 보유하도록 디자인되지만 DNA 클론 라이브러리에서 비정보학적 서열 생성을 최소화하는 포획 프로브 모듈을 고려한다. "포획 프로브 모듈"은 포획 프로브 서열 및 꼬리 서열을 포함하는 폴리뉴클레오타이드를 언급한다. 특정 구현예에서, 포획 프로브 모듈 서열 또는 이의 일부는 하나 이상의 서열분석 프라이머를 위한 프라이머 결합 부위로서 작용한다.

[0188] 특정 구현예에서, 포획 프로브 모듈은 포획 프로브를 포함한다. 본원에 사용된 바와 같은 "포획 프로브"는 특이적 DNA 표적 영역에 하이브리드화할 수 있는 영역을 포함하는 폴리뉴클레오타이드를 언급한다. DNA의 평균 크기가 상대적으로 작고 고도로 단편화되어 있기 때문에, 본원에 고려된 조성물 및 방법은 목적하는 DNA 표적 영역을 검색하기 위한 고밀도 및 상대적으로 짧은 포획 프로브의 용도를 포함한다.

[0189] 특정 구현예에서, 포획 프로브 모듈은 임의로 합텐을 포함하고 꼬리 서열과 하이브리드화하여 포획 프로브 모듈 이중나선을 생성시키는 파트너 올리고뉴클레오타이드와 조합된다.

[0190] 고밀도 포획 프로브를 사용하는 것의 하나의 특정 관심사는 일반적으로 포획 프로브가 특이적 "서열 법칙"을 사용하여 디자인된다는 것이다. 예를 들어, 중첩 서열 영역 또는 극단의 염기 조성 편중을 나타내는 영역은 일반적으로 포획 프로브를 디자인하는데 배제된다. 그러나, 본 발명자들은 포획 프로브 디자인 법칙에서 유연성의 기여가 프로브 수행능에 실질적으로 영향을 주지 않음을 발견하였다. 대조적으로, 위치적 속박에 의해 엄중하게 선택된 포획 프로브는 온-표적 서열 정보를 제공하였고; 매우 적은 오프-표적 및 맵핑될 수 없는 판독 포획을 나타내고; 단지 극소수를 제외한 균일하고 유용한 온-표적 판독을 생성한다. 더욱이, 보다 큰 공간으로 존재하는 밀접한 프로브에서 높은 중복성은 드문 불량한-수행 포획 프로브를 보상한다.

[0191] 특정 구현예에서, 표적 영역은 다수의 포획 프로브에 의해 표적화되고, 여기서, 2개 이상의 포획 프로브는 서로 10 이상 뉴클레오타이드 이내, 서로 15 이상 뉴클레오타이드 이내, 서로 20 이상 뉴클레오타이드 이내, 서로 25 이상 뉴클레오타이드 이내, 서로 30 이상 뉴클레오타이드 이내, 서로 35 이상 뉴클레오타이드 이내, 서로 40 이상 뉴클레오타이드 이내, 서로 45 이상 뉴클레오타이드 이내, 또는 서로 50 이상 뉴클레오타이드 이내 및 모든 간섭 뉴클레오타이드 길이에서 표적 영역에 결합하도록 디자인된다.

[0192] 하나의 구현예에서, 포획 프로브는 약 25 뉴클레오타이드, 약 26 뉴클레오타이드, 약 27 뉴클레오타이드, 약 28 뉴클레오타이드, 약 29 뉴클레오타이드, 약 30 뉴클레오타이드, 약 31 뉴클레오타이드, 약 32 뉴클레오타이드, 약 33 뉴클레오타이드, 약 34 뉴클레오타이드, 약 35 뉴클레오타이드, 약 36 뉴클레오타이드, 약 37 뉴클레오타이드, 약 38 뉴클레오타이드, 약 39 뉴클레오타이드, 약 40 뉴클레오타이드, 약 41 뉴클레오타이드, 약 42 뉴클레오타이드, 약 43 뉴클레오타이드, 약 44 뉴클레오타이드, 또는 약 45 뉴클레오타이드이다.

- [0193] 하나의 구현예에서, 포획 프로브는 약 100 뉴클레오타이드, 약 200 뉴클레오타이드, 약 300 뉴클레오타이드, 약 400 뉴클레오타이드, 또는 약 100 뉴클레오타이드이다. 또 다른 구현예에서, 포획 프로브는 약 100 뉴클레오타이드 내지 약 500 뉴클레오타이드, 약 200 뉴클레오타이드 내지 약 500 뉴클레오타이드, 약 300 뉴클레오타이드 내지 약 500 뉴클레오타이드, 또는 약 400 뉴클레오타이드 내지 약 500 뉴클레오타이드, 또는 이의 임의의 중재 범위이다.
- [0194] 특정 구현예에서, 포획 프로브는 60 뉴클레오타이드이다.
- [0195] 특정 구현예에서, 포획 프로브는 60 뉴클레오타이드가 아니다.
- [0196] 또 다른 구현예에서, 포획 프로브는 실질적으로 60 뉴클레오타이드 보다 작지만 동일한 DNA 표적 영역을 표적화하는 60 뉴클레오타이드 포획 프로브와 상응하게, 및 보다 우수하게 하이브리드화한다.
- [0197] 특정 구현예에서, 포획 프로브는 40 뉴클레오타이드이다.
- [0198] 특정 구현예에서, 포획 프로브 모듈은 꼬리 서열을 포함한다. 본원에 사용된 바와 같은 용어 "꼬리 서열"은 특정 구현예에서 프라이머 결합 부위로서 작용할 수 있는, 포획 프로브 모듈의 5' 말단에서 폴리뉴클레오타이드를 언급한다. 특정 구현예에서, 서열분석 프라이머는 꼬리 영역에서 프라이머 결합 부위에 결합한다.
- [0199] 특정 구현예에서, 꼬리 서열은 약 5 내지 약 100 뉴클레오타이드, 약 10 내지 약 100 뉴클레오타이드, 약 5 내지 약 75 뉴클레오타이드, 약 5 내지 약 50 뉴클레오타이드, 약 5 내지 약 25 뉴클레오타이드, 또는 약 5 내지 약 20 뉴클레오타이드이다. 특정 구현예에서, 제3 영역은 약 10 내지 약 50 뉴클레오타이드, 약 15 내지 약 40 뉴클레오타이드, 약 20 내지 약 30 뉴클레오타이드 또는 약 20 뉴클레오타이드, 또는 임의의 중재 수의 뉴클레오타이드이다.
- [0200] 특정 구현예에서, 꼬리 서열은 약 30 뉴클레오타이드, 약 31 뉴클레오타이드, 약 32 뉴클레오타이드, 약 33 뉴클레오타이드, 약 34 뉴클레오타이드, 약 35 뉴클레오타이드, 약 36 뉴클레오타이드, 약 37 뉴클레오타이드, 약 38 뉴클레오타이드, 약 39 뉴클레오타이드, 또는 약 40 뉴클레오타이드이다.
- [0201] 다양한 구현예에서, 포획 프로브 모듈은 포획 프로브에 하이브리드화하는 태그되고/되거나 증폭된 DNA 라이브러리의 하나 이상의 포획된 단편의 단리 및/또는 정제를 가능하게 하는 결합 쌍의 특정 구성원을 포함한다. 특정 구현예에서, 포획 프로브 모듈은 비오틴 또는 또 다른 적합한 합텐, 예를 들어, 디니트로페놀, 디콕시게닌으로의 접합체이다.
- [0202] 다양한 구현예에서, 포획 프로브 모듈은 태그되고 임의로 증폭된 DNA 라이브러리에 하이브리드화하여 복합체를 형성한다. 일부 구현예에서, 다기능성 포획 프로브 모듈은 DNA 라이브러리에서 특이적 게놈 표적 영역에 실질적으로 하이브리드화한다.
- [0203] 하이브리드화 또는 하이브리드화 조건은 임의의 반응 조건을 포함할 수 있고, 여기서, 2개의 뉴클레오타이드 서열은 안정한 복합체; 예를 들어, 태그된 DNA 라이브러리 및 안정한 태그된 DNA 라이브러리-포획 프로브 모듈 복합체를 형성하는 포획 프로브 모듈을 형성한다. 상기 반응 조건은 당업계에 널리 공지되어 있고 당업자는 상기 조건이 적당히, 예를 들어, 보다 짧은 길이 포획 프로브를 갖는 감소된 어닐링 온도로 변형될 수 있고 본 발명의 범위내에 있음을 인식할 것이다. 실질적 하이브리드화는 포획 프로브 복합체의 제2 영역이 태그된 DNA 라이브러리 영역과 100 %, 99 %, 98 %, 97 %, 96 %, 95 %, 94 %, 93 %, 92 %, 91 %, 90 %, 89 %, 88 %, 85 %, 80 %, 75 %, 또는 70 % 서열 동일성, 상동성 또는 상보성을 나타내는 경우 일어날 수 있다.
- [0204] 특정 구현예에서, 포획 프로브는 약 40 뉴클레오타이드이고 약 44°C 내지 약 47 °C의 최적의 어닐링 온도를 갖는다.
- [0205] 특정 구현예에서, 본원에 고려된 방법은 태그된 DNA 라이브러리-포획 프로브 모듈 복합체를 단리시킴을 포함한다. 특정 구현예에서, DNA 복합체를 단리시키기 위한 방법은 당업자에게 널리 공지되어 있고 당업자에 의해 적당할 것으로 간주된 임의의 방법이 본 발명의 방법과 함께 사용될 수 있다 (Ausubel *et al.*, *Current Protocols in Molecular Biology*, 2007-2012). 특정 구현예에서, 복합체는 비오틴-스트렙타비딘 단리 기술을 사용하여 단리시킨다. 일부 구현예에서, 다기능성 포획 프로브 모듈의 꼬리 서열에 하이브리드화할 수 있는 포획 파트너 올리고뉴클레오타이드는 DNA 복합체 단리 방법에 사용하기 위한 칼럼, 비드 또는 기판에 연결된 스트렙타비딘과 상호작용할 수 있는 5'-말단 또는 3'-말단에서 비오틴을 함유하도록 변형된다.
- [0206] 하나의 구현예에서, 다기능성 포획 프로브 모듈의 꼬리 서열에 하이브리드화할 수 있는 포획 파트너 올리고뉴클

레오타이드는 DNA 복합체 단리 방법에 사용하기 위한 칼럼, 비드 또는 다른 기판에 연결된 스트랩타비딘과 상호 작용할 수 있는 3'-말단에서 비오틴을 함유하도록 변형된다.

- [0207] 특정 구현예에서, 다기능성 포획 프로브 모듈의 꼬리 서열은 포획 파트너 올리고뉴클레오타이드에 결합한다. 일부 구현예에서, 다기능성 포획 프로브 모듈은 태그된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 형성 전 포획 파트너 올리고뉴클레오타이드에 결합한다. 일부 구현예에서, 다기능성 포획 프로브 모듈은 태그된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 형성 후 포획 파트너 올리고뉴클레오타이드에 결합한다. 일부 구현예에서, 다기능성 포획 프로브 모듈은 태그된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 형성과 동시에 포획 파트너 올리고뉴클레오타이드에 결합한다. 일부 구현예에서, 포획 파트너 올리고뉴클레오타이드는 화학적으로 변형된다. 하나의 구현예에서, 상기 포획 파트너 올리고뉴클레오타이드는 합텐을 5' 또는 3' 말단에 부가함에 의해 변형된다. 하나의 구현예에서 합텐은 비오틴이다.
- [0208] 특정 구현예에서, 단리된 태그된 DNA 라이브러리-포획 프로브 모듈 복합체로부터 단일 가닥의 3'-말단의 제거가 고려된다. 특정 구현예에서, 상기 방법은 단일 가닥의 3' 말단을 제거하기 위해 단리된 태그된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 3'-5' 엑소뉴클레아제 효소적 프로세싱을 포함한다.
- [0209] 특정 다른 구현예에서, 상기 방법은 주형으로서 단리된 태그된 DNA 라이브러리 단편을 사용하는 다기능성 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행함을 포함한다.
- [0210] 특정 다른 구현예에서, 상기 방법은 5' FLAP 엔도뉴클레아제, DNA 중합 및 DNA 리가제에 의한 Nick 폐쇄의 협력 작용을 통한 하이브리드 포획 프로브-단리된 태그된 DNA 표적 분자를 생성함을 포함한다.
- [0211] 다양한 효소는 단리된 태그된 DNA 라이브러리-다기능성 포획 프로브 모듈 복합체의 3'-5' 엑소뉴클레아제 효소적 프로세싱을 위해 사용될 수 있다. 특정 구현예에서 사용될 수 있는 3'-5' 엑소뉴클레아제 효소적 활성을 나타내는 적합한 효소의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: T4 또는 엑소뉴클레아제 I, III, V (또한 문헌 (Shevelev IV, Huebscher U., "The 3' 5' exonucleases," *Nat Rev Mol Cell Biol.* 3(5):364-76 (2002)을 참조한다). 특정 구현예에서, 3'-5' 엑소뉴클레아제 활성을 포함하는 효소는 T4이다. 특정 구현예에서, 3'-5' 엑소뉴클레아제 효소적 활성을 나타내고 프라이머 주형 연장을 수행할 수 있는 효소가 사용될 수 있고 예를 들어 T4 또는 엑소뉴클레아제 I, III, V. *Id*를 포함한다.
- [0212] 일부 구현예에서, 본원에 고려된 방법은 상기 문헌 및 본원에서 기타 위치에서 논의된 3'-5' 엑소뉴클레아제 효소적으로 프로세싱된 복합체에 대해 서열분석 및/또는 PCR을 수행함을 포함한다. 특정 구현예에서, 포획 프로브 분자의 꼬리 부분은 하이브리드 핵산 분자를 생성하기 위해 카피된다. 하나의 구현예에서, 생성된 하이브리드 핵산 분자는 포획 프로브 모듈 및 포획 프로브 모듈 꼬리 서열의 상보체에 하이브리드화할 수 있는 표적 영역을 포함한다.
- [0213] 특정 구현예에서, 유전학적 분석은 a) 하나 이상의 포획 프로브 모듈을 다수의 DNA 라이브러리 클론에서 하나 이상의 표적 유전학적 유전자좌에 하이브리드화시켜 하나 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 형성하는 단계; b) a)로부터 하나 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 단리하는 단계; c) 단계 b)로부터 하나 이상의 단리된 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 효소적으로 프로세싱하는 단계; d) c)로부터 효소적으로 프로세싱된 복합체에 대해 PCR을 수행하는 단계로서, 포획 프로브 분자의 꼬리 부분은 증폭된 하이브리드 핵산 분자를 생성시키기 위해 카피되고 상기 증폭된 하이브리드 핵산 분자는 포획 프로브 및 포획 프로브 모듈 꼬리 서열의 상보체에 하이브리드화할 수 있는 표적 게놈 유전자좌에서 표적 서열을 포함하는, 단계; 및 e) d)로부터 증폭된 하이브리드 핵산 분자에 대해 정량적 유전학적 분석을 수행하는 단계를 포함한다.
- [0214] 특정 구현예에서, 특이적 표적 유전학적 유전자좌의 카피수를 결정하기 위한 방법이 고려되고 다음 단계를 포함한다: a) 하나 이상의 포획 프로브 모듈을 다수의 DNA 라이브러리 클론에서 하나 이상의 표적 유전학적 유전자좌에 하이브리드화시켜 하나 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 형성하는 단계; b) a)로부터 하나 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 단리하는 단계; c) 단계 b)로부터 하나 이상의 단리된 포획 프로브 모듈-DNA 라이브러리 클론 복합체를 효소적으로 프로세싱하는 단계; d) c)로부터 효소적으로 프로세싱된 복합체에 대해 PCR을 수행하는 단계로서, 포획 프로브 분자의 꼬리 부분은 증폭된 하이브리드 핵산 분자를 생성시키기 위해 카피되고 상기 증폭된 하이브리드 핵산 분자는 포획 프로브 및 포획 프로브 모듈 꼬리 서열의 상보체에 하이브리드화할 수 있는 표적 유전학적 유전자좌에서 표적 서열을 포함하는, 단계; e) d)에서 증폭된 하이브리드 핵산 분자의 PCR 증폭을 수행하는 단계; 및 f) e)에서 PCR 반응을 정량하는 단계로서, 상

기 정량은 특이적 표적 영역의 카피수의 결정을 가능하게 한다.

- [0215] 하나의 구현예에서, 단계 c)의 효소적 프로세싱은 단일 가닥의 3' 말단을 제거하는 3'-5' 엑소뉴클레아제 활성을 갖는 효소를 사용하여 b)로부터의 하나 이상의 포획 프로브 모듈-DNA 라이브러리 클론 복합체에 대해 3'-5' 엑소뉴클레아제 효소적 프로세싱을 수행하는 단계; 5' FLAP 엔도뉴클레아제, DNA 중합 및 DNA 리가제에 의한 니페쇄의 협력 작용을 통해 하나 이상의 하이브리드 포획 프로브 모듈-DNA 라이브러리 클론 분자를 생성하는 단계; 또는 주형으로서 복합체에서 단리된 DNA 클론을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행함을 포함한다.
- [0216] 하나의 구현예에서, 단계 c)의 효소적 프로세싱은 주형으로서 복합체에서 단리된 DNA 클론을 사용하여 포획 프로브의 5'-3' DNA 폴리머라제 연장을 수행함을 포함한다.
- [0217] 특정 구현예에서, PCR은 당업자에게 널리 공지된 임의의 표준 PCR 반응 조건을 사용하여 수행될 수 있다. 특정 구현예에서, e)에서 PCR 반응은 2개의 PCR 프라이머를 사용한다. 하나의 구현예에서, e)에서 PCR 반응은 표적 유전학적 유전자좌 내 반복체에 하이브리드화하는 제1 PCR 프라이머를 사용한다. 특정 구현예에서, e)에서 PCR 반응은 표적 유전학적 유전자좌/꼬리 접합부에서 하이브리드 핵산 분자에 하이브리드화하는 제2 PCR 프라이머를 사용한다. 특정 구현예에서, e)에서 PCR 반응은 표적 유전학적 유전자좌에 하이브리드화하는 제1 PCR 프라이머를 사용하고 제2 PCR 프라이머는 표적 유전학적 유전자좌/꼬리 접합부에서 증폭된 하이브리드 핵산 분자에 하이브리드화한다. 특정 구현예에서, 제2 프라이머는 표적 유전학적 유전자좌/꼬리 접합부에 하이브리드화하여 프라이머의 적어도 하나 이상의 뉴클레오타이드는 표적 유전학적 유전자좌에 하이브리드화하고 프라이머의 적어도 하나 이상의 뉴클레오타이드는 꼬리 서열에 하이브리드화한다.
- [0218] 특정 구현예에서, 단계 e)로부터 수득된 증폭된 하이브리드 핵산 분자는 서열분석되고 서열은 수평적으로 정렬되고, 즉 서로 정렬되지만 참조 서열에 정렬되지 않는다. 특정 구현예에서, 단계 a) 내지 e)는 하나 이상의 포획 프로브 모듈과 함께 1회 이상 반복한다. 포획 프로브 모듈은 동일하거나 상이할 수 있고 표적 유전학적 유전자좌의 DNA 가닥에 표적화하도록 디자인될 수 있다. 일부 구현예에서, 포획 프로브가 상이한 경우, 이들은 태그된 DNA 클론 라이브러리 내에서 중첩하거나 인접한 표적 서열에서 하이브리드화한다. 하나의 구현예에서, 고밀도 포획 프로브 전략이 사용되고, 여기서, 다수의 포획 프로브는 표적 유전학적 유전자좌에 하이브리드화하고 다수의 포획 프로브 각각은 모든 중재 거리를 포함하는 태그된 DNA 클론 라이브러리에서 표적 유전학적 유전자좌에 하이브리드화하는 임의의 다른 포획 프로브의 약 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 bp 내 표적 유전학적 유전자좌에 하이브리드화한다.
- [0219] 일부 구현예에서, 상기 방법은 표적 유전학적 유전자좌 당 2개의 포획 프로브 모듈을 사용하여 수행될 수 있고, 여기서, 하나는 표적 영역의 업스트림에 "왓슨" 가닥 (비-암호화 또는 주형 가닥)에 하이브리드화하고 하나는 표적 영역의 다운스트림에 "클릭" 가닥 (암호화 또는 비-주형 가닥)에 하이브리드화한다.
- [0220] 특정 구현예에서, 본원에 고려된 방법은 이의 임의의 수가 임의의 조합으로 왓슨 또는 클릭 가닥에 하이브리드화하는 표적 유전학적 유전자좌 당 임의의 수의 포획 프로브 모듈, 예를 들어 2, 3, 4, 5, 6, 7, 8, 9, 또는 10 개 이상의 포획 프로브 모듈로 수회 추가로 수행될 수 있다. 일부 구현예에서, 수득된 서열은 임의의 다수의 차이를 동정하기 위해 서로 정렬될 수 있다.
- [0221] 특정 구현예에서, 다수의 표적 유전학적 유전자좌는 하나 이상의 포획 프로브 모듈을 사용하여 단일 반응에서 예를 들어, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 10000, 50000, 100000, 500000회 이상 검색된다.
- [0222] **(b) 서열분석**
- [0223] 특정 구현예에서, 정량적 유전학적 분석은 상기 본원에서 기타 다른 위치에서 논의된 바와 같이 다수의 하이브리드 핵산 분자를 서열 분석하여 다수의 특유의 서열분석 판독을 수득하기 위해 충분한 서열분석 정도를 생성함을 포함한다. 특유의 판독은 모두가 DNA 내 동일한 판독 코드 및 서열 개시점을 공유하는 판독 "계열"로부터의 단일 컨센서스 판독으로서 정의된다. 각각의 포획 프로브는 계열들로 분류함에 의해 총 판독으로부터 계산적으로 생성된 특유의 판독 세트를 생성시킨다. 소정의 샘플에 대한 특유의 판독은 이어서 프로브별 기준으로 관찰된 특유의 판독의 평균으로서 계산된다. 명백한 카피수 변화가 있는 경우는 평균을 계산하기 위해 사용된 데이터 세트로부터 제외시킨다. 특유의 판독은 각각의 특유의 판독이 특유의 DNA 클론으로부터 유래되어야만 하기 때문에 중요하다. 각각의 특유의 판독은 게놈 DNA의 반수체 증가물의 인풋 및 분석을 나타낸다. 특유의 판독은 분석된 반수체 게놈의 합이다. 이어서 분석된 게놈의 수는 서열분석 검정의 민감성을 정의한다. 비제한

적인 예로서, 평균 특유의 판독 계수가 100 게놈 등가물인 경우, 상기 특정 검정은 100, 또는 1%에서 하나의 돌연변이체 판독을 검출할 수 있는 민감성을 갖는다. 상기 미만의 임의의 관찰은 논리적으로 옹호할 수 없다.

[0224] 특정 구현예에서, 정량적 유전학적 분석은 다수의 샘플로부터 유래된 하이브리드 핵산 분자의 멀티플렉스 서열 분석을 포함한다.

[0225] 다양한 구현예에서, 정량적 유전학적 분석은 하나 이상 또는 다수의 태그된 DNA 라이브러리 클론을 수득하는 단계로서, 각각의 클론은 제1 DNA 서열 및 제2 DNA 서열을 포함하고, 제1 DNA 서열은 표적화된 유전학적 유전자좌에서 서열을 포함하고 제2 DNA 서열은 포획 프로브 서열을 포함하는, 단계; 하나 이상의 클론에 대해 쌍을 이룬 말단 서열분석 반응을 수행하고 하나 이상의 서열분석 판독을 수득하거나 하나 이상의 클론에 대해 서열분석 반응을 수행하는 단계로서, 약 100, 200, 300, 400, 500 이상의 뉴클레오타이드 보다 큰 단일의 긴 서열분석 판독이 수득되고, 상기 판독은 제1 DNA 서열과 제2 DNA 서열 둘 다를 동정하기에 충분한, 단계; 및 서열분석 판독의 프로브 서열에 따라 하나 이상의 클론의 서열분석 판독을 배열하거나 클러스터링하는 단계를 포함한다.

[0226] **(c) 생물정보학적 분석**

[0227] 다양한 구현예에서, 정량적 유전학적 분석은 서열분석 판독의 생물정보학적 분석을 추가로 포함한다. 생물정보학적 분석은 서열분석을 위한 조성물 또는 방법의 부재하에 수행되는 임의의 순수한 정신 분석을 배제한다. 특정 구현예에서, 생물정보학적 분석은 다음을 포함하지만 이에 제한되지 않는다: 서열 정렬; 게놈 등가물 분석; 단일 뉴클레오타이드 변이체 (SNV) 분석; 유전자 카피수 변화 (CNV) 분석; 및 유전학적 병변의 검출. 특정 구현예에서, 생물정보학적 분석은 DNA 클론 라이브러리에서 분석된 게놈 등가물의 수를 정량하고; 표적 유전학적 유전자좌의 유전학적 상태를 검출하고; 표적 유전학적 유전자좌에서 유전학적 병변을 검출하고; 표적 유전학적 유전자좌 내 카피수 변동을 측정하는데 유용하다.

[0228] 서열 정렬은 서열 판독과 하나 이상의 인간 참조 DNA 서열 간에 수행될 수 있다. 특정 구현예에서, 서열분석 정렬을 사용하여 표적 유전학적 유전자좌에서 유전학적 병변을 검출할 수 있고 이는 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재배열, 카피수 변화 또는 유전자 융합의 검출을 포함하지만 이에 제한되지 않는다. 원인이 되거나 예후적 지시기인 유전학적 병변의 검출은 특정 유전학적 병태 또는 질환의 진단, 예후, 치료 및/또는 모니터링에 유용할 수 있다.

[0229] 또한 본원에 고려되는 것은 서열 정렬 분석을 위한 방법이고 이는 표준 서열로 정렬시킬 필요 없이 수행될 수 있고 이는 본원에서 수평 서열 분석으로서 언급된다. 상기 분석은 본원에 고려되는 방법 또는 임의의 다른 방법에 의해 생성된 임의의 서열에 대해 수행될 수 있다. 특정 구현예에서, 서열 분석은 본원에 고려된 방법에 의해 수득된 판독에 대한 서열 정렬을 수행함을 포함한다.

[0230] 하나의 구현예에서, DNA 클론 라이브러리에서 게놈 등가물은 서열분석이 수행된 후 생물정보학적 기반 계수를 사용하여 결정된다. 각각의 서열분석 판독은 특정 포획 프로브와 관련되고, 각각의 포획 프로브에 할당된 판독의 수집물은 그룹 별로 분석된다. 그룹 내, 개별 판독 세트는 게놈 서열 내 동일한 판독 코드 및 동일한 DNA 서열 개시 위치를 공유한다. 이들 개별 판독은 "계열"로 분류되고 상기 계열을 대표하는 단일 컨센서스는 "특유의 판독"으로서 앞으로 이동된다. 계열을 구성하는 개별 판독 모두는 단일 연결 반응으로부터 유래되고 따라서 이들은 서로의 증폭-유래된 "시블링"이다. 각각의 특유의 판독은 특유의 연결 반응으로서 고려되고 특유의 판독 합은 분석되는 게놈 등가물의 수와 동등한 것으로 고려된다.

[0231] 특유의 클론의 수가 가능한 서열 조합의 총 수에 접근함에 따라서, 가능성은 동일한 코드 및 개시 부위 조합이 독립적인 반응에 의해 생성되고 이들 독립적 반응이 부적절하게 단일 계열 내에서 분류됨을 지적한다. 최종적 결과는 분석되는 게놈 등가물의 과소평가이고 희귀 돌연변이체 판독은 이들이 동일한 확인번호를 갖는 야생형 판독과 중첩하기 때문에 서열분석 오류로서 버려질 수 있다.

[0232] 특정 구현예에서, DNA 클론 라이브러리에 대해 정확한 분석을 제공하기 위해, 분석된 게놈 등가물의 수는 가능한 특유의 클론의 수의 약 1/10, 약 1/12, 약 1/14, 약 1/16, 약 1/18, 약 1/20, 약 1/25 이하이다. 전술한 과정이 주로 설명을 위한 것이고 제한하는 것이 아닌 것으로 이해되어야만 한다.

[0233] 일부 구현예에서, 분석될 게놈 등가물의 수는 증가될 필요가 있을 수 있다. 게놈 등가물의 정도를 확장하기 위해, 적어도 2개의 해결책이 고려된다. 제1 해결책은 샘플 당 1개 초과 어댑터 세트를 사용하는 것이다. 어댑터를 조합함에 의해, 가능한 클론의 총 수 및 따라서 게놈 인풋의 간편성 한계를 증가적으로 확장시킬 수 있다. 제2 해결책은 1, 2, 3, 4, 또는 5 이상의 염기 정도로 판독 코드를 확장시키는 것이다. 모든 다른 판독

코드로부터 적어도 2개의 염기 정도로 상이한 가능한 판독 코드의 수는 $4^{(n-1)}$ 의 스케일로 나타내고 여기서 n은 판독 코드 내 염기의 수이다. 따라서, 비제한적인 예에서, 판독 코드가 5 뉴클레오타이드이고 $4^{(5-1)} = 256$ 인 경우; 따라서 추가 염기의 내포는 각각의 추가의 염기에 대해 4개 인자 정도로 가용한 레퍼토리를 확장한다.

[0234] 하나의 구현예에서, 정량적 유전학적 분석은 희귀 단일 뉴클레오타이드 변이체 (SNV)를 동정하기 위한 서열분석 판독의 생물정보학적 분석을 포함한다.

[0235] 차세대 서열분석은 대략 0.02-0.02 %의 고유 오류율을 갖는데, 이는 1/200 내지 1/500 염기 풀의 어느 곳에서도 부정확함을 의미한다. 이 보다 낮은 빈도로 나타나는 변이체 및 다른 돌연변이를 검출하기 위해, 예를 들어 1000 서열 당 1의 빈도에서, 분자 주석 전략을 적용할 필요가 있다. 비제한적인 예로서, 표적화된 서열 포획 기술을 사용한 5000 특유한 분자의 분석은 >50,000 판독의 충분한 서열 정도에서 5000 특유의 판독 수집물을 생성하고, 각각의 특유의 판독은 모두 동일한 판독 코드를 갖는 판독의 "계열"에 속한다. 계열 내에 존재하는 SNV는 희귀 변이체에 대한 후보물이다. 상기 동일한 변이체가 1개 초과 계열에서 관찰되는 경우, 이것은 출발 샘플 내에 존재하는 희귀 변이체가 되는 것에 대해 매우 강한 후보물이 된다. 대조적으로, 계열 내 산발적으로 존재하는 변이체는 서열분석 오류일 가능성이 있고 하나 및 단지 하나의 계열 내에 존재하는 변이체는 희귀하거나 생체외에서 일어나는 염기 변형의 결과 (예를 들어, DNA 염기의 산화 또는 PCR-도입된 오류)이다.

[0236] 하나의 구현예에서, SNV를 검출하는 방법은 검정의 목적하는 표적 민감성으로서 10배 초과 계열 인풋 (계놈 또는 계놈 등가물)을 도입함을 포함한다. 하나의 비제한적인 예에서, 목적하는 민감성이 2 % (100 중 2)인 경우, 실험적 표적은 2000 계놈의 인풋이다.

[0237] 특정 구현예에서, 서열분석 데이터의 생물정보학적 분석을 사용하여 유전학적 상태, 병태 또는 질환, 유전학적 모자이크 현상, 태아 시험, 친부 시험, 약물 치료에 대한 반응 예측, 의학적 병태의 진단 또는 모니터링, 마이크로바이옴 프로파일링, 병원체 스크리닝 및 기관 이식체의 모니터링과 관련된 SNV를 검출하거나 동정한다.

[0238] 다양한 구현예에서, 카피수 결정 분석을 위한 방법이 제공되고 이는 하나 이상 또는 다수의 클론을 수득함을 포함하고, 각각의 클론은 제1 DNA 서열 및 제2 DNA 서열을 포함하고, 여기서, 상기 제1 DNA 서열은 표적화된 유전학적 유전자좌에서 서열을 포함하고 제2 DNA 서열은 포획 프로브 서열을 포함한다. 관련 구현예에서, 하나 이상의 클론에 대한 쌍을 이룬 말단 서열분석 반응을 수행하고 하나 이상의 서열분석 판독을 수득한다. 또 다른 구현예에서, 하나 이상의 클론에 대한 서열분석 반응이 수행되고 여기서, 약 100 초과 뉴클레오타이드의 단일의 긴 서열분석 판독이 수득되고 상기 판독은 제1 DNA 서열과 제2 DNA 서열 둘 다를 동정하는데 충분하다. 하나 이상의 클론의 서열분석 판독은 서열분석 판독의 프로브 서열에 따라 정렬되거나 클러스터링될 수 있다.

[0239] 카피수 분석은 특정 유전자의 카피 수를 조사하는 분석, 또는 소정의 계놈 DNA 샘플에 존재하고 소정의 샘플에서 소정의 유전자 또는 서열 차이의 카피수의 정량적 결정을 추가로 포함할 수 있는 돌연변이체를 포함하지만 이에 제한되지 않는다. 특정 구현예에서, 카피수 분석은 유전학적 상태, 병태 또는 질환, 태아 시험, 유전학적 모자이크 현상, 친부 확인 시험, 약물 치료에 대한 반응 예측, 의학적 병태의 진단 또는 모니터링, 마이크로바이옴 프로파일링, 병원체 스크리닝, 및 기관 이식 모니터링과 관련된 유전자 증폭을 검출하거나 동정하기 위해 사용된다.

[0240] 특정 구현예에서, 서열분석 데이터의 생물정보학적 분석을 사용하여 표적 유전자좌에서 하나 이상의 서열 또는 유전학적 병변을 검출하거나 동정하고 이는 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 계놈 재배열, 카피수 변화 또는 유전자 융합의 검출을 포함하지만 이에 제한되지 않는다. 원인이 되거나 예측적 지시기인 유전학적 병변의 검출은 특정 유전학적 병태 또는 질환의 진단, 예후, 치료 및/또는 모니터링에 유용할 수 있다. 하나의 구현예에서, 유전학적 영역은 유전학적 상태, 병태, 질환, 태아 시험, 유전학적 모자이크 현상, 친부 확인 시험, 약물 치료에 대한 반응 예측, 의학적 병태의 진단 또는 모니터링, 마이크로바이옴 프로파일링, 병원체 스크리닝, 및 기관 이식 모니터링과 관련된다.

[0241] **F. 정량적 유전학적 분석의 임상 적용**

[0242] 다양한 구현예에서, 본 발명은 대상체에서 병태 또는 질환을 검출하거나, 동정하거나, 예측하거나, 진단하거나, 모니터링하는 방법을 고려한다.

[0243] 특정 구현예에서, 대상체에서 유전학적 상태, 병태 또는 질환을 검출하거나, 동정하거나, 예측하거나, 진단하거나, 모니터링하는 방법은 DNA 클론 라이브러리에서 하나 이상의 표적 유전학적 유전자좌의 정량적 유전학적 분석을 수행하여 하나 이상의 표적 유전학적 유전자좌에서 서열 내 변화를 검출하거나 변화시킴을 포함한다. 하

나의 구현예에서, DNA는 cfDNA이다.

[0244] 특정 구현예에서, 유전학적 질환; 유전학적 모자이크 현상; 태아 시험; 친부 확인 시험; 약물 치료에 대한 반응 예측; 의학적 병태의 진단 또는 모니터링; 마이크로바이옴 프로파일링; 병원체 스크리닝; 및 기관 이식 모니터링으로 이루어진 군으로부터 선택된 유전학적 상태 또는 유전학적 병태 또는 질환을 검출하거나, 동정하거나, 예측하거나, 진단하거나, 모니터링하는 방법은 DNA 클론 라이브러리에서 하나 이상의 표적 유전학적 유전자좌의 정량적 유전학적 분석을 수행하여 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재배열, 카피수 변화 또는 하나 이상의 표적 유전학적 유전자좌에서 서열내 유전자 융합을 검출하거나 동정함을 포함한다.

[0245] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 유전학적 질환의 예시적 예는 암, 알츠하이머 질환 (APOE1), 찰코트-마리-투스 (Charcot-Marie-Tooth) 질환, 레버 유전성 시각 신경병증 (LHON), 양겔만 증후군 (UBE3A, 유비퀴틴-단백질 리가제 E3A), 프레이더-윌 증후군 (Prader-Willi syndrome) (염색체 15 내 영역), β -지중해빈혈 (HBB, β -글로빈), 고체병 (I형) (GBA, 글루코세레브로시다제), 낭성 섬유증 (CFTR 상피 클로라이드 채널), 겸상적혈구세포 질환 (HBB, β -글로빈), 타이 사흐 질환 (Tay-Sachs disease) (HEXA, 헥소스아미다제 A), 페닐케톤뇨증 (PAH, 페닐알라닌 하이드롤리아제), 가족성 과콜레스테롤혈증 (LDLR, 저밀도 지질단백질 수용체), 성인 당뇨병 신장 질환 (PKD1, 폴리시스틴), 헌팅턴 질환 (HDD, 헌팅틴), 신경섬유종증 I형 (NF1, NF1 중앙 서프래서 유전자), 근긴장성이영양증 (DM, 미오톤), 결절성 경화증 (TSC1, 튜버린), 연골무형성증 (FGFR3, 섬유아세포 성장 인자 수용체), 취약한 X 증후군 (FMR1, RNA-결합 단백질), 두첸 (Duchenne) 근육성이형성증 (DMD, 디스트로핀), 혈우병 A (F8C, 혈액 응고 인자 VIII), 레슈-니한 증후군 (Lesch-Nyhan syndrome) (HPRT1, 하이포산틴 구아닌 리보실트랜스퍼라제 1), 및 부신백질이영양증 (ABCD1)을 포함하지만 이에 제한되지 않는다.

[0246] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 암의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: B 세포 암, 예를 들어, 다발성 골수종, 흑색종, 유방암, 폐암 (예를 들어, 비-소 세포 폐 암종 또는 NSCLC), 기관지 암, 결장직장암, 전립선암, 췌장암, 위암, 난소암, 방광암, 뇌암 또는 중추 신경계 암, 말초 신경계 암, 식도암, 경부암, 자궁 또는 자궁내막암, 구강암 또는 인두암, 간암, 신장암, 고환암, 담관암, 소장 또는 맹장암, 침샘암, 갑상선암, 부신암, 골육종, 연골육종, 혈액학적 조직암, 선암종, 염증성 근섬유아세포 종양, 위장 기질 종양 (GIST), 결장암, 다발성 골수종 (MM), 골수이형성 증후군 (MDS), 골수증식성 장애 (MPD), 급성 림프구 백혈병 (ALL), 급성 골수구 백혈병 (AML), 만성 골수구 백혈병 (CML), 만성 림프구 백혈병 (CLL), 적혈구증가증 베라 (polycythemia Vera), 호지킨 림프종, 비-호지킨 림프종 (NHL), 연질-조직 육종, 섬유육종, 점액육종, 지방육종, 골육종, 척색종, 혈관육종, 내피육종, 림프구혈관육종, 림프구혈관내피육종, 활막종, 중피종, 어빙 종양 (Ewing's tumor), 평활근육종, 횡문근육종, 편평 세포암종, 기저 세포 암종, 선암종, 땀샘 암종, 피지선 암종, 유두암종, 유두 선암종, 수질암종, 기관지 암종, 신장 세포 암종, 간암종, 담관암종, 응모막암종, 정상피종, 배아 암종, 빌름 종양 (Wilms' tumor), 방광 암종, 상피암종, 신경교종, 성상세포종, 수모세포종, 두개인두종, 척수내 종양, 송과체종, 혈관아세포종, 청신경종, 뿔피 교종, 뇌수막종, 신경모세포종, 망막모세포종, 여포성 림프종, 확산 대형 B-세포 림프종, 맨틀 세포 림프종, 간 세포 암종, 갑상선암, 위암, 두경부암, 소세포 암, 본태성 혈소판증가증, 병인불명 골수 화생, 호산구증가증후군, 전신비만세포종, 가족성 과호산구증가증, 만성 호산성 백혈병, 신경내분비암, 유암종 등.

[0247] 하나의 구현예에서, 유전학적 병변은 코스믹 데이터베이스에 주석달린 병변 (병변 및 서열 데이터는 사이트 (cancer.sanger.ac.uk/cosmic/census)로부터 다운로드될 수 있다) 또는 암 게놈 아틀라스에 주석달린 병변 (병변 및 서열 데이터는 사이트 (tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp)로부터 다운로드될 수 있다)이다.

[0248] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 암과 관련된 하나 이상의 유전학적 병변을 함유하는 유전자좌의 예시적 예는 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, CTNBN1, 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, TGFB2, TMPRSS2, TNFRSF14, TOP1, TP53, TPMT, TSC1, TSC2, TYMS, UGT1A1, UMPS, USP9X, VHL, 및 WT1을 포함하지만 이에 제한되지 않는다.

[0249] 특정 구현예에서, 유전학적 병변은 뉴클레오타이드 전이 또는 전환, 뉴클레오타이드 삽입 또는 결실, 게놈 재정렬, 카피수 변화 또는 유전자 융합을 포함한다.

[0250] 하나의 구현예에서, 유전학적 병변은 ALK 유전자의 3' 암호화 영역을 또 다른 유전자로 융합하는 유전자 융합이다.

[0251] 하나의 구현예에서, 유전학적 병변은 ALK 유전자의 3' 암호화 영역을 EML4 유전자로 융합하는 유전자 융합이다.

[0252] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 태아 시험에 적합한 병태의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 다운 증후군 (Trisomy 21), 에드워드 증후군 (Trisomy 18), 파타우 증후군 (Trisomy 13), 클리네펠터 증후군 (Klinefelter's Syndrome) (XXY), 트리플 X 증후군, XYY 증후군, Trisomy 8, Trisomy 16, 터너 증후군 (XO), 로버트소니안 전좌, 디조지 증후군 및 울프-히르쇼른 증후군.

[0253] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 친부 확인 시험을 위해 적합한 대립유전자의 예시적 예는 다음 중 16개 이상을 포함하지만 이에 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.

[0254] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 약물 치료에 대한 반응을 예측하기에 적합한 유전자의 예시적 예는 다음 유전자 중 하나 이상을 포함하지만 이에 제한되지 않는다: ABCB1 (ATP-결합 카세트, 서브-패밀리 B (MDR/TAP), 구성원 1), ACE (안지오텐신 I 전환 효소), ADH1A (알콜 데하이드로게나제 1A (부류 I), 알파 폴리펩타이드), ADH1B (알콜 데하이드로게나제 IB (부류 I), 베타 폴리펩타이드), ADH1C (알콜 데하이드로게나제 1C (부류 I), 감마 폴리펩타이드), ADRB1 (아드레날린 작용성, 베타-1-, 수용체), ADRB2 (아드레날린 작용성, 베타-2-, 수용체, 표면), AHR (아릴 탄화수소 수용체), ALDH1A1 (알데하이드 데하이드로게나제 1 계열, 구성원 A1), ALOX5 (아라키도네이트 5-리폭시게나제), BRCA1 (유방암 1, 어얼리 온셋), COMT (카테콜-O-메틸트랜스퍼라제), CYP2A6 (시토크롬 P450, 계열 2, 서브패밀리 A, 폴리펩타이드 6), CYP2B6 (시토크롬 P450, 계열 2, 서브패밀리 B, 폴리펩타이드 6), CYP2C9 (시토크롬 P450, 계열 2, 서브패밀리 C, 폴리펩타이드 9), CYP2C19 (시토크롬 P450, 계열 2, 서브패밀리 C, 폴리펩타이드 19), CYP2D6 (시토크롬 P450, 계열 2, 서브패밀리 D, 폴리펩타이드 6), CYP2J2 (시토크롬 P450, 계열 2, 서브패밀리 J, 폴리펩타이드 2), CYP3A4 (시토크롬 P450, 계열 3, 서브패밀리 A, 폴리펩타이드 4), CYP3A5 (시토크롬 P450, 계열 3, 서브패밀리 A, 폴리펩타이드 5), DPYD (디하이드로피리미딘 데하이드로게나제), DRD2 (도파민 수용체 D2), F5 (응고 인자 V), GSTP1 (글루타티온 S-트랜스퍼라제 pi), HMGCR (3-하이드록시-3-메틸글루타릴-코엔자임 A 리덕타제), KCNH2 (칼륨 전압-게이팅된 채널, 서브패밀리 H (eag-관련된), 구성원 2), KCNJ11 (칼륨 내향-정류 채널, 서브패밀리 J, 구성원 11), MTHFR (5,10-메틸렌테트라하이드로폴레이트 리덕타제 (NADPH)), NQO1 (NAD(P)H 데하이드로게나제, 키논 1), P2RY1 (푸린 수용체 P2Y, G-단백질 커플링된, 1), P2RY12 (푸린 수용체 P2Y, G-단백질 커플링된, 12), PTGIS (프로스타글란딘 12 (프로스타사이클린) 신타제), SCN5A (나트륨 채널, 전압-게이팅된, V형, 알파 (긴 QT 증후군 3)), SLC19A1 (용질 캐리어 계열 19 (플레이트 수용체), 구성원 1), SLC01B1 (용질 캐리어 유기 음이온 수송체 계열, 구성원 1B1), SULT1A1 (설포트랜스퍼라제 계열, 시토졸릭, 1A, 페놀-선택, 구성원 1), TPMT (티오푸린 S-메틸트랜스퍼라제), TYMS (티미딜레이트 신데타제), UGT1A1 (UDP 글루쿠로노실트랜스퍼라제 1 계열, 폴리펩타이드 A1), VDR (비타민 D (1,25- 디하이드록시비타민 D3) 수용체), VKORC1 (비타민 K 에폭사이드 리덕타제 복합체, 서브유닛 1).

[0255] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 의학 적 병태의 예시적 예는 뇌졸중, 일시적 허혈 허혈 발작, 외상 뇌 손상, 심장 질환, 심장 발작, 협심증, 죽상동맥경화증, 및 고혈압을 포함하지만 이에 제한되지 않는다.

- [0256] 본원에 고려된 조성물 및 방법으로 스크리닝될 수 있는 병원체의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 세균 진균류, 및 바이러스.
- [0257] 본원에 고려된 조성물 및 방법으로 스크리닝될 수 있는 세균 종의 예시적 예는 마이코박테리움 종 (Mycobacterium spp.), 뉴모코커스 종 (Pneumococcus spp.), 에스케리치아 종 (Escherichia spp.), 캄필로박터 종 (Campylobacter spp.), 코리네박테리움 종 (Corynebacterium spp.), 클로스트리디움 종 (Clostridium spp.), 스트렙토코커스 종 (Streptococcus spp.), 스태필로코커스 종 (Staphylococcus spp.), 슈도모나스 종 (Pseudomonas spp.), 쉬겔라 종 (Shigella spp.), 트레포네마 종 (Treponema spp.), 또는 살모넬라 종 (Salmonella spp.)을 포함하지만 이에 제한되지 않는다.
- [0258] 본원에 고려된 조성물 및 방법으로 스크리닝될 수 있는 진균류 종의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 아스페르길러스 종 (Aspergillus spp.), 블라스토마이세스 종 (Blastomyces spp.), 캔디다 종 (Candida spp.), 코시시오이데스 종 (Coccidioides spp.), 크립토코커스 종 (Cryptococcus spp.), 더마토피트 (dermatophytes), 티네아 종 (Tinea spp.), 트리코피톤 종 (Trichophyton spp.), 마이크로스포룸 종 (Microsporium spp.), 푸사리움 종 (Fusarium spp.), 히스토플라스마 종 (Histoplasma spp.), 무코로마이코티나 종 (Mucoromycotina spp.), 뉴모시스티스 종 (Pneumocystis spp.), 스포로트릭스 종 (Sporothrix spp.), 엑세로필룸 종 (Exserophilum spp.), 또는 클라도스포리움 종 (Cladosporium spp.).
- [0259] 본원에 고려된 조성물 및 방법으로 스크리닝될 수 있는 바이러스의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 인플루엔자 A, 예를 들어, H1N1, H1N2, H3N2 및 H5N1 (조류 독감), 인플루엔자 B, 인플루엔자 C 바이러스, 간염 A형 바이러스, 간염 B형 바이러스, 간염 C형 바이러스, 간염 D형 바이러스, 간염 E형 바이러스, 로타바이러스, 노르윙크 바이러스 군의 임의의 바이러스, 장 아데노바이러스, 파르보바이러스, 땀귀열 바이러스, 몽키 폭스, 모노네가비탈레스, 리사바이러스, 예를 들어, 람비 바이러스, 라고스 배트 바이러스, 모콜라 바이러스, 두벤하게 바이러스, 유로핀 배트 바이러스 1 & 2 및 오스트레일리안 배트 바이러스, 에페메로바이러스, 비시콜로바이러스, 비시콜라 스토마티티스 바이러스 (VSV), 헤르페스 바이러스, 예를 들어, 헤르페스 심플렉스 바이러스 1형 및 2형, 바리셀라 조스터, 사이토메갈로바이러스, 엡슈타인-바르 바이러스 (EBV), 인간 헤르페스 바이러스 (HHV), 인간 헤르페스바이러스 6형 및 8형, 몰로니 뮤린 백혈병 바이러스 (M-MuLV), 몰로니 뮤린 육종 바이러스 (MoMSV), 하르베이 뮤린 육종 바이러스 (HaMuSV), 뮤린 포유동물 종양 바이러스 (MuMTV), 기본 에이프 백혈병 바이러스 (GaLV), 고양이 백혈병 바이러스 (FLV), 스푸마바이러스, 프론트 뮤린 백혈병 바이러스, 뮤린 줄기 세포 바이러스 (MSCV) 및 라우스 육종 바이러스(RSV), HIV (인간 면역결핍 바이러스; HIV 1형, 및 2형 HIV를 포함하는), 비스나-매디 바이러스 (VMV) 바이러스, 카프린 관절염-뇌염 바이러스 (CAEV), 말 감염성 빈혈 바이러스 (EIAV), 고양이 면역결핍 바이러스 (FIV), 소 면역 결핍 바이러스 (BIV), 및 시미안 면역 결핍 바이러스 (SIV), 파필로마 바이러스, 뮤린 감마헤르페스바이러스, 아레나바이러스, 예를 들어, 아르젠틴 출혈 열 바이러스, 볼리비아 출혈 열 바이러스, 사비아-관련 출혈열 바이러스, 베네주엘란 출혈열 바이러스, 라사열 바이러스, 마추포 바이러스, 림프구 맥락수막염 바이러스 (LCMV), 부냐비리대, 예를 들어, 크리메안-콩고 출혈열 바이러스, 한타바이러스, 신장 증후군 유발 바이러스와 함께 출혈열, 리프트 발리 열 바이러스, 에볼라 출혈열 및 마르부르크 출혈열을 포함하는 필로비리대 (필로바이러스), 카이사누르 포레스트 질환 바이러스를 포함하는 플라비비리대, 움스크 출혈열 바이러스, 진드기 매개 뇌염 유발 바이러스 및 과라믹소비리대, 예를 들어, 헨드라 바이러스 및 니파 바이러스, 바리올라 메이저 및 바리올라 마이너 (천연두), 알파바이러스, 예를 들어, 베네주엘란 말 뇌염 바이러스, 이스턴 말 뇌염 바이러스, 웨스턴 말 뇌염 바이러스, SARS-관련 코로나바이러스 (SARS-CoV), 웨스트 나일 바이러스, 및 임의의 뇌염 유발 바이러스.
- [0260] 본원에 고려된 조성물 및 방법으로 검출되거나, 동정되거나, 예측되거나, 진단되거나, 모니터링될 수 있는 이식 수용자에서 기관 이식체를 모니터링하기에 적합한 유전자의 예시적 예는 다음 유전자 중 하나 이상을 포함하지만 이에 제한되지 않는다: HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DP, 및 HLA-DQ.
- [0261] 특정 구현예에서, 생물정보학적 분석을 사용하여 DNA 클론 라이브러리에서 분석된 게놈 증가물의 수를 정량하거나; 생표적 유전학적 유전자좌에서 유전학적 변이체를 검출하거나; 표적 유전학적 유전자좌 내 돌연변이를 검출하거나; 표적 유전학적 유전자좌내 유전학적 융합을 검출하거나; 표적 유전학적 유전자좌 내 카피수 변동을 측정한다.
- [0262] **A. 동반 진단제**
- [0263] 다양한 구현예에서, 다음을 포함하는, 유전학적 질환을 위한 동반 진단제가 제공된다: 대상체의 생물학적 샘플로부터 DNA, 예를 들어, cfDNA를 단리시키거나 수득하는 단계; DNA의 말단 포스페이트 잔기들을 제거하는 단계;

탈인산화된 DNA를 하나 이상의 말단-복구 효소로 처리하여 말단-복구된 DNA를 생성하는 단계; 하나 이상의 이중 가닥 DNA (dsDNA) 프리-어댑터를 말단-복구된 DNA의 각각의 가닥의 3' 말단으로 연결하여 프리-어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 dsDNA 프리-어댑터는 말단-복구된 DNA의 각각의 가닥, 및 비-연결 파트너 가닥 올리고뉴클레오타이드의 3' 말단에 연결된 연결 가닥 올리고뉴클레오타이드를 포함하는, 단계; 프리-어댑터/말단-복구된 DNA 복합체로부터의 비-연결 파트너 가닥 올리고뉴클레오타이드를 복구 올리고뉴클레오타이드로 대체하여 어댑터/말단-복구된 DNA 복합체를 형성하는 단계로서, 각각의 어댑터는 연결 가닥 올리고뉴클레오타이드 및 복구 올리고뉴클레오타이드를 포함하는, 단계; 어댑터/말단-복구된 DNA 복합체를 하나 이상의 효소로 처리하여 연속 이중가닥 DNA 라이브러리를 형성하는 단계; DNA 라이브러리를 증폭시켜 DNA 클론 라이브러리를 생성하는 단계; DNA 클론 라이브러리에서 게놈 증가물의 수를 결정하는 단계; 및 DNA 클론 라이브러리에서 유전학적 질환과 연관된 하나 이상의 바이오마커의 정량적 유전학적 분석을 수행하는 단계를 포함하고, 여기서, 하나 이상의 바이오마커의 검출 또는 검출 실패는 대상체가 유전학적 질환에 대해 치료되어야만 하는지를 지적한다.

[0264] 본원에 사용된 바와 같은 용어 "동반 진단제"는 특정 항암 치료요법에 연관된 진단제 시험을 언급한다. 특정 구현예에서, 진단 방법은 생물학적 샘플과 관련된 바이오마커에서 유전학적 병변의 검출을 포함함으로써 환자가 항암 치료요법으로 치료되어야만 하거나 치료되지 말아야 하는지의 즉각적인 동정을 가능하게 한다.

[0265] 항암 치료요법은 수술, 방사선, 화학치료제, 항암 약물 및 면역조절제를 포함하지만 이에 제한되지 않는다.

[0266] 항암 약물의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 알킬화제, 예를 들어 티오테파 및 사이클로포스파미드 (CYTOXAN™); 알킬 설포네이트, 예를 들어 부설판, 임프로셀판 및 피포셀판; 아지리딘, 예를 들어 벤조도파, 카보쿤, 메투레도파, 및 우레도파; 알트레타민, 트리에틸렌멜라민, 트리에틸렌포스포르아미드, 트리에틸렌티오포스포르아미드 및 트리메틸올로멜라민 레수메를 포함하는 에틸렌이민 및 메틸아멜라민; 질소 머스타드, 예를 들어, 클로람부실, 클로나파진, 클로로포스파미드, 에스트라무스틴, 이포스파미드, 메클로레타민, 메클로레타민 옥사이드 하이드로클로라이드, 멜팔란, 노벤비친, 페네스테린, 프레드니무스틴, 트로포스파미드, 우라실 머스타드; 니트로스우레아, 예를 들어 카무스틴, 클로로조도신, 포테무스틴, 로무스틴, 니무스틴, 라니무스틴; 항생제, 예를 들어, 아클라시노마이신, 악티노마이신, 아우트라마이신, 아자세린, 블레오마이신, 캅티노마이신, 칼케아미신, 카라비신, 카르미노마이신, 카지노필린, 크로모마이신, 닥티노마이신, 다우노루비신, 테토루비신, 6-디아조-5-옥소-L-노르류신, 독소루비신 및 이의 폐기화된 제형, 에피루비신, 에소루비신, 이다루비신, 마르셀로마이신, 미토마이신, 마이코페놀산, 노갈라마이신, 올리보마이신, 페플로마이신, 포트피로마이신, 푸로마이신, 쿠엘라마이신, 로도루비신, 스트렙토니그린, 스트렙토조신, 투버시딘, 우베니맥스, 지노스타틴, 조루비신; 항-대사물, 예를 들어 메토크세이트 및 5-플루오로우라실 (5-FU); 폴산 유사체, 예를 들어 데노프테린, 메토크세이트, 프테로프테린, 트리메토크세이트; 푸린 유사체, 예를 들어 플루다라빈, 6-머캅토프린, 티아미프린, 티오구아닌; 피리미딘 유사체, 예를 들어 안시타빈, 아자시티딘, 6-아자우리딘, 카모푸르, 시타라빈, 디데옥시우리딘, 독시플루리딘, 에노시타빈, 플록스우리딘, 5-FU; 안드로겐, 예를 들어, 칼루스테론, 드로모스타놀론 프로피오네이트, 에피티오스타놀, 메피티오스탄, 테스톨락톤; 항-아드레날, 예를 들어, 아미노글루테티미드, 미토탄, 트릴로스탄; 폴산 리프레니셔, 예를 들어, 프롤린산; 아세글라톤; 알도포스파미드 글리코시드; 아미노레볼린산; 암사크린; 베스트라부실; 비산트렌; 에다트락세이트; 데포파민; 데메코이신; 디아지쿠온; 엘포르미틴; 엘리프티늄 아세테이트; 에토글루시드; 질산갈륨; 하이드록시우레아; 렌티난; 요오니다민; 미토구아존; 미톡산트론; 모피다몰; 니트라크린; 펜토스타틴; 페나메트; 피라루비신; 포도필린산; 2-에틸하이드라지드; 프로카바진; PSK®; 라족산; 시조피란; 스피로게르마늄; 테누아존산; 트리아지쿠온; 2,2',2"-트리클로로트리에틸아민; 우레탄; 빈데신; 다카바진; 만노무스틴; 미토브로니톨; 미톨락톨; 피포브로만; 가시토신; 아라비노시드 ("Ara-C"); 사이클로포스파미드; 티오테파; 탁소이드, 예를 들어, 파클리탁셀 (TAXOL®, Bristol-Myers Squibb Oncology, Princeton, N.J.) 및 독세탁셀 (TAXOTERE®, Rhne-Poulenc Rorer, Antony, France); 클로람부실; 겐시타빈; 6-티오구아닌; 머캅토프린; 메토크세이트; 백금 유사체, 예를 들어, 시스플라틴 및 카보플라틴; 빈블라스틴; 백금; 에토포시드 (VP-16); 이포스파미드; 미토미오신 C; 미톡산트론; 빈크리스틴; 비노렐빈; 나벨빈; 노반트론; 테니포시드; 아미노프테린; 젤로다; 이반드로네이트; CPT-11; 토포이소머라제 억제제 RFS 2000; 디플루오로메틸오미틴 (DMFO); 레티노산 유도체, 예를 들어, Targretin™ (백사로텐), Panretin™ (알리트레티노인); ONTAK™ (테니류킨 디프티톡스); 에스페라미신; 카페시타빈; 및 약제학적으로 허용되는 염, 상기 임의의 산 또는 유도체. 상기 정의에서 또한 예를 들어, 타목시펜, 랄록시펜, 아로마타제 억제제 4(5)-이미다졸, 4-하이드록시타목시펜, 트리옥시펜, 케옥시펜, LY117018, 오나프리스톤, 및 토레미펜 (Fareston)을 포함하는 항-에스트로겐과 같은 암에 대한 호르몬 작용을 조절하거나 억제하는 작용을 하는 항-호르몬 제제; 및 플루타미드, 닐루타미드, 바이칼루타미드, 류프롤리드, 및 고세렐린과 같은 항-안드로겐; 및 상기 임의의 약제학적으로 허용되

는 염, 산 또는 유도체가 포함된다.

[0267] 면역조절제의 예시적 예는 다음을 포함하지만 이에 제한되지 않는다: 사이클로스포린, 타크롤리무스, 트레스페리무스, 피메크롤리무스, 시롤리무스, 베롤리무스, 라플루니무스, 라쿠이니모드 및 이미쿠이모드, 및 이들의 유사체, 유도체, 염, 이온 및 착물.

[0268] 본원에 인용된 모든 공보, 특허 출원 및 허여된 특허는 각각의 개별 공보, 특허 출원 또는 허여된 특허가 구체적으로 및 개별적으로 참조로 인용된 것 처럼 참조로 본원에 인용된다. 2013년 12월 10일자로 출원된 미국 특허 출원 제14/102,285호, 및 2014년 8월 22일자로 출원된 미국 특허 출원 번호 제14/466,741호는 각각 이의 전문이 본원에 참조로 인용된다.

[0269] 이전의 발명은 이해를 명백히 할 목적으로 설명 및 예시에 의해 일부 상세하게 기재되었지만, 특정 변화 및 변형이 첨부된 청구항의 취지 또는 범위로 부터 벗어나는 것 없이 이루어 질 수 있는 본 발명의 교시의 관점에서 당업자에게 용이하게 자명할 것이다. 하기의 실시예는 단지 설명을 목적으로 제공되고 이에 제한하려는 것은 아니다. 당업자는 필수적으로 유사한 결과를 획득하도록 변화되거나 변형될 수 있는 다양한 비임계 파라미터를 용이하게 인지할 것이다.

[0270] 실시예

[0271] 실시예 1:

[0272] 고효율 어댑터 연결을 위한 원리의 증거

[0273] 본 실시예는 본원에 고려된 고효율 연결 전략이 어댑터 이량체 형성의 부재하에 DNA 단편의 양 말단으로 연결시키는 직접적인 정량적 증거를 제공한다 (도 2).

[0274] 클로닝 벡터 pUC19로부터의 플라스미드 DNA는 제한 효소 RsaI로 분해시켜 평활 말단을 생성하고 안타르틱 알칼린 포스파타제로 탈인산화시켰다. 이들 평활 말단의 탈인산화된 DNA 단편은 9개의 상이한 고효율 어댑터의 수 집물에 연결시켰다(표 1). 모든 경우에, 단편 이동성에서 정량적 전환은 연결 반응 (화살표) 후 관찰되었고 이동성에서 전환은 각각의 DNA 단편의 각각의 말단으로 어댑터의 부착과 동등하였다. 본 실시예는 본원에 고려된 조성물 및 방법이 어댑터의 DNA 단편으로의 고효율 연결을 유도하여 DNA 라이브러리를 작제하는 전체 효율을 증가시킨다는 원리의 증거를 제공한다.

[0275] 표 1 시험 he 어댑터를 생성하기 위해 사용되는 연결 가닥 및 파트너 가닥 올리고뉴클레오타이드 서열

서열 번호:	어댑터 ID	연결 가닥 (5' 포스페이트, 5'에서 3')	서열 번호:	파트너 가닥 (5'에서 3', 3' ddC)
1	1	GAGGGTCTACCTTCTTNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	1	AAGGTAGACCCT
2	2	GCTCTAGACGTCATCGNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	2	TGACGCTAGAG
3	3	GGATACTCGTAGCGGCNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	3	GCTACGAGTATC
4	4	GTCACGAGTAGAGAAANNNTGTATTTCGAATTCTCTGGTCCTGCA	4	CTCTACTCGTGA
5	5	GAGGGTCTACCTTAGTNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	5	AAGGTAGACCCT
6	6	GCTCTAGACGTCAGAGNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	6	TGACGCTAGAG
7	7	GGATACTCGTAGCTTNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	7	GCTACGAGTATC
8	8	GTCACGAGTAGAGCCANNNTGTATTTCGAATTCTCTGGTCCTGCA	8	CTCTACTCGTGA
9	9	GAGGGTCTACCTTGCTNNNNNTGTATTTCGAATTCTCTGGTCCTGCA	9	AAGGTAGACCCT

[0276] 실시예 2:

[0278] 고효율 DNA 라이브러리 작제

[0279] I. 단편 말단 복구

[0280] 세포-부재 DNA (cfDNA) 단편 말단은 탈인산화되었고, DNA 이중나선에 대한 내부 손상은 복구되었고 DNA 말단은 평활 말단으로 "폴리싱"되었다. 수득한 단편은 "말단-복구된 단편"으로서 언급된다.

[0281] 상기 cfDNA 단편 말단은 81 μl의 정제된 cfDNA, 10 μl의 뉴 잉글랜드 바이오랩 (NEB) CutSmart 완충액 (B7204S), 및 5 μl의 NEB 새우 알칼린 포스파타제 (M0371)를 조합함에 의해 탈인산화시켰다. 반응 혼합물은 15분 동안 37 °C에서 항온처리하고 이어서 5분 동안 65 °C에서 항온처리하였다.

[0282] 탈인산화된 cfDNA 단편에 대한 내부 손상은 빙상에서 제조된, 4 μl의 하기의 혼합물을 첨가함에 의해 복구하였다: 1.1 μl의 10 mM dNTP 혼합물 (NEB N0447), 2.2 μl의 PreCR 효소 혼합물 (M0309) 및 1.1 μl의 T4 DNA

폴리머라제 (M0203). 상기 반응 혼합물은 15분 동안 20 °C에 항온처리하고 이어서 10분 동안 70 °C에서 항온처리하였다. 상기 방식으로 복구되고 폴리싱된 cfDNA 단편은 DNA 연결 반응에 직접적인 사용을 위해 준비한다.

[0283] II. 프리-어댑터 디자인

[0284] 여러 디자인 고려사항을 사용하여 연결 가닥 올리고뉴클레오타이드 및 차단된 3' 말단을 갖는 상보적 파트너 가닥 올리고뉴클레오타이드로 구성된 프리-어댑터를 생성하였다.

[0285] 본 실시예에 사용된 프리-어댑터는 하기의 특징을 갖도록 디자인하였다: 길이가 10 nt인 연결 가닥 올리고뉴클레오타이드; A/T 또는 G/C 잔기의 균등한 발란스; 프리-어댑터의 각각의 세트 내 각각의 염기 위치에서 나타낸 4개의 DNA 염기 각각; 50 mM Na⁺ (또는 K⁺), 10 mM MgCl₂ 중 ~37 °C인 10 염기 서열의 예측된 용융 온도; 각각의 프리-어댑터 서열의 제1 2개 염기로서 A/T 및 G/C nt 둘 다; 길이가 8 nt이고 2-하이드록실 리보스-변형된 DNA 염기 (MWG Eurofins)를 사용함에 의해 화학적으로 차단되고 50 mM Na⁺ 또는 K⁺, 10 mM MgCl₂ 중 ~25 °C의 용융 온도를 갖는 상보적 파트너 올리고뉴클레오타이드 서열.

[0286] 제위치에 디자인 속박에도 불구하고, 어댑터 세트의 경험적 수행능 스크리닝을 수행하였다. 현재 실험에서, 허용되는 수행능을 갖는 4개의 어댑터의 5개 세트를 동정하였다 (예를 들어, 표 2를 참조한다). "스코어"로 표지된 칼럼은 최상의 수행 어댑터에 상대적인 각각의 어댑터의 % 클로닝 효율을 보여준다 (세트 6-2).

[0287] 표 2 경험적으로 확인된 어댑터 세트

연결 가닥 올리고뉴클레오타이드		파트너 가닥 올리고뉴클레오타이드		스코어
명칭	서열	명칭	서열	
he_lig_5-1	/5Phos/CTGAGCTAGT	he_part_5-1	TAGCTCA[3-dG-Q]	60
he_lig_5-2	/5Phos/GACTCGATAG	he_part_5-2	ATCGAGT[3-dC-Q]	97
he_lig_5-3	/5Phos/TCAGATCGTC	he_part_5-3	CGATCTG[3-dA-Q]	27
he_lig_5-4	/5Phos/AGTCTAGCCA	he_part_5-4	GCTAGAC[3-dT-Q]	84
he_lig_6-1	/5Phos/GGATTACCT	he_part_6-1	GGTAATC[3-dC-Q]	70
he_lig_6-2	/5Phos/CTTACGGATG	he_part_6-2	TCCGTAA[3-dG-Q]	100
he_lig_6-3	/5Phos/ACCGATTGAC	he_part_6-3	CAATCGG[3-dT-Q]	74
he_lig_6-4	/5Phos/TAGCGCATGA	he_part_6-4	ATGCGCT[3-dA-Q]	48
he_lig_8-1	/5Phos/ATGTCCAGCT	he_part_8-1	CTGGACA[3-dT-Q]	24
he_lig_8-2	/5Phos/CACAGGTTAG	he_part_8-2	AACCTGT[3-dG-Q]	46
he_lig_8-3	/5Phos/TGACATGCTC	he_part_8-3	GCATGTC[3-dA-Q]	53
he_lig_8-4	/5Phos/GCTGTACAGA	he_part_8-4	TGTACAG[3-dC-Q]	48
he_lig_11-1	/5Phos/TCAAGTCGGT	he_part_11-1	CGACTTG[3-dA-Q]	83
he_lig_11-2	/5Phos/GTTCAGACTG	he_part_11-2	GTCTGAA[3-dC-Q]	95
he_lig_11-3	/5Phos/CAGGTCTAAC	he_part_11-3	TAGACCT[3-dG-Q]	33
he_lig_11-4	/5Phos/AGCTCAGTCA	he_part_11-4	ACTGAGC[3-dT-Q]	98
he_lig_12-1	/5Phos/GATCCGTACT	he_part_12-1	TACGGAT[3-dC-Q]	91
he_lig_12-2	/5Phos/ACAGTCGTAG	he_part_12-2	ACGACTG[3-dT-Q]	98
he_lig_12-3	/5Phos/TGGTAACCTC	he_part_12-3	GGTTACC[3-dA-Q]	85
he_lig_12-4	/5Phos/CTCAGTAGGA	he_part_12-4	CTACTGA[3-dG-Q]	93

[0288] III. 프리-어댑터의 연결

[0289] 프리-어댑터는 본 실시예의 단계 I에서 생성된 말단-복구된 단편에 연결하였다. 25 µl의 말단-복구된 단편은 10 µl의 10 µM 어댑터와 조합하였다. 전형적으로, 1 내지 4 연결 반응은 반응에 첨가된 별도의 어댑터의 수에 의존하여 수행하였다. 15 µl의 연결 각테일 (5 µl의 10X T4-DNA 연결 완충액, 7.5 µl의 50 % PEG8000, 및 2.5 µl의 HC T4 DNA 리가제 (NEB; M0202))은 최종 용적 50 µl로 각각의 연결 반응에 첨가하였다. 반응물은 혼합하고 60분 동안 20 °C에서 항온처리하고 이어서 10분 동안 65 °C에 항온처리하고 이어서 실온으로 냉각시켰다.

[0291] 연결 반응 후, 50 µl의 TEzero (10 mM Tris pH 8.0, 0.1 mM EDTA, 0.05% Tween 20) 및 120 µl의 DNA 정제 비드는 각각의 반응물에 첨가하고 잘 혼합하였다. 상기 반응물은 10분 동안 실온에서 항온처리하고 이어서 비드는 200 µl의 70 % 에탄올/물 (v/v)로 2회 세척하고, 간략히 공기 건조시키고 (~5 분), 20 µl의 TEzero로 용출시

켰다.

[0292] IV. 복구 올리고뉴클레오타이드

[0293] 본 실시예에 사용된 복구 올리고뉴클레오타이드의 완전한 목록은 표 3에 나타난다. 각각의 복구 올리고뉴클레오타이드는 249개 개별 올리고뉴클레오타이드 풀이다. 각각의 복구 올리고뉴클레오타이드에서 불변 서열은 PCR 프라이머 결합 부위를 나타내고 표 3의 좌측 부분에 나타난다.

[0294] 249개 올리고뉴클레오타이드 각각은 복구 올리고뉴클레오타이드 서열에서 "XXXXX"로서 나타낸, 5개 뉴클레오타이드 샘플 코드를 포함한다. 5개 뉴클레오타이드 샘플은 표 3의 우측 부분에 나타난다. 5개 뉴클레오타이드 코드는 세트 내 모든 다른 코드와는 상이한 2개 염기 변화인 것으로 선택된 256 가능한 특유의 서열로 이루어진다. 이러한 특징은 특유하고 독특한 판독이 코드 영역에서 서열분석 오차로 인해 고유한 것으로 나타난 판독으로부터 구분되도록 할 수 있다. G 잔기가 과제공되고 경험적으로 어댑터 기능을 방해하는 것으로 나타난 7개 코드를 제거하고 249개 랜덤 코드를 남겼다.

[0295]

표 3 복구 올리고 및 이들의 관련된 판독 코드

연결 올리고와 상용성인 전장 복구 올리고 (RO)						
명칭	서열					
RO_5-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXACTAGCTCAG					
RO_5-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCTATCGAGTC					
RO_5-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGACGATCTGA					
RO_5-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTGCTAGACT					
RO_6-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXAGGGTAATCC					
RO_6-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCATCCGTAAG					
RO_6-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGTCAATCGGT					
RO_6-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCATGCGCTA					
RO_7-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXACTGCTAGCA					
RO_7-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCAGCGATCAT					
RO_7-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGTATCGATG					
RO_7-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTGATAGCTGC					
RO_8-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXAGCTGGACAT					
RO_8-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCTAACCTGTG					
RO_8-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGAGCATGTCA					
RO_8-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCTGTACAGC					
RO_11-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXACCGACTTGA					
RO_11-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCAGTCTGAAC					
RO_11-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGTTAGACCTG					
RO_11-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTGACTGAGCT					
RO_12-1	TGCAGGACCAGAGAATTTCGAATACAXXXXXAGTACGGATC					
RO_12-2	TGCAGGACCAGAGAATTTCGAATACAXXXXXCTACGACTGT					
RO_12-3	TGCAGGACCAGAGAATTTCGAATACAXXXXXGAGGTTACCA					
RO_12-4	TGCAGGACCAGAGAATTTCGAATACAXXXXXTCTACTGAG					
XXXXX - 차이 2 서열 코드						
Seq	Seq	Seq	Seq	Seq	Seq	Seq
CGGGT	CACGG	GTGAT	ATCAG	GATAC	TGTCT	CCAAA
CGGTG	CAGCG	GTGTA	ATCGA	GATCA	TGTTC	CCACC
CGTGG	CAGGC	GTTAG	ATGAC	GCAAT	TTCGT	CCATT
GCGGT	CCAGG	GTTGA	ATGCA	GCATA	TTCTG	CCCAC
GCGTG	CCGAG	TAGGT	CAAGT	GCCCT	TTGCT	CCCCA
GCTGG	CCGGA	TAGTG	CAATG	GCCTC	TTGTC	CCTAT
GGCGT	CGACG	TATGG	CAGAT	GCTAA	TTTCG	CCTTA
GGCTG	CGAGC	TGAGT	CAGTA	GCTCC	TTTGC	CTACT
GGGCT	CGCAG	TGATG	CATAG	GCTTT	AAAAA	CTATC
GGGTC	CGCGA	TGGAT	CATGA	GTAAC	AAACC	CTCAT
GGTCG	CGGAC	TGGTA	CCCGT	GTACA	AAATT	CTCTA
GGTGC	CGGCA	TGTAG	CCCTG	GTCAA	AACAC	CTTAC

[0296]

GTCGG	GAAAAG	TGTGA	CCGCT	GTCCC	AACCA	CTTCA
GTGCG	GAAGA	TTAGG	CCGTC	GTCTT	AATAT	TAAAT
GTGGC	GACCG	TTGAG	CCTCG	GTTCT	AATTA	TAATA
TGCGG	GACGC	TTGGA	CCTGC	GTTTC	ACAAC	TACCT
TGGCG	GAGAA	AACGT	CGAAT	TAACG	ACACA	TACTC
TGGGC	GAGCC	AACTG	CGATA	TAAGC	ACCAA	TATAA
AAAGG	GAGTT	AAGCT	CGCCT	TACAG	ACCCC	TATCC
AAGAG	GATGT	AAGTC	CGCTC	TACGA	ACCTT	TATTT
AAGGA	GATTG	AATCG	CGTAA	TAGAC	ACTCT	TCACT
ACCGG	GCACG	AATGC	CGTCC	TAGCA	ACTTC	TCATC
ACGCG	GCAGC	ACAGT	CGTTT	TCAAG	ATAAT	TCCAT
ACGGC	GCCAG	ACATG	CTAAG	TCAGA	ATATA	TCCTA
AGAAG	GCCGA	ACGAT	CTAGA	TCCCG	ATCCT	TCTAC
AGAGA	GCGAC	ACGTA	CTCCG	TCCGC	ATCTC	TCTCA
AGCCG	GCGCA	ACTAG	CTCGC	TCGAA	ATTAA	TTAAA
AGCGC	GGAAA	ACTGA	CTGAA	TCGCC	ATTCC	TTACC
AGGAA	GGACC	AGACT	CTGCC	TCGTT	ATTTT	TTATT
AGGCC	GGATT	AGATC	CTGTT	TCTGT	CAAAC	TTCAC
AGGTT	GGCAC	AGCAT	CTTGT	TCTTG	CAACA	TTCCA
AGTGT	GGCCA	AGCTA	CTTTG	TGAAC	CACAA	TTTAT
AGTTG	GGTAT	AGTAC	GAACT	TGACA	CACCC	TTTTA
ATGGT	GGTTA	AGTCA	GAATC	TGCAA	CACTT	
ATGTG	GTAGT	ATACG	GACAT	TGCCC	CATCT	
ATTGG	GTATG	ATAGC	GACTA	TGCTT	CATTC	

[0297]

[0298]

V. 복구 올리고뉴클레오타이드의 프리-어댑터 라이브러리의 첨가

[0299]

라이브러리 작제는 복구 올리고뉴클레오타이드를 어댑터에 첨가함에 의해 완료하였다. 본 실시예에서 설명된 복구 올리고뉴클레오타이드는 PCR 프라이머 결합 부위; 샘플 코드; 및 서열분석 판독에서 적당한 염기 콜의 보정을 가능하게 하고 연결 가닥 올리고뉴클레오타이드로의 하이브리드화를 위해 앵커로서 작용하는 서열을 동정하기 위한 수단으로서 작용하는 무작위 서열 표지인 앵커 서열을 함유한다.

[0300]

4 μl의 1 μM 풀의 복구 올리고뉴클레오타이드 (예를 들어, 표 3을 참조한다)는 본 실시예의 단계 III으로부터의 20 μl의 정제된 연결 혼합물에 첨가하였다.

[0301]

이어서, 40 μl의 복구 올리고뉴클레오타이드 반응은 병상에서 24 μl의 복구 올리고뉴클레오타이드/리가제 혼합물과 16 μl의 하기 혼합물을 배합함에 의해 제조하였다: 11 μl의 물, 4.4 μl의 완충액 "B" (190 μl의 CutSmart 완충액 (NEB; B7204) 및 10 μl의 1 M 디티오프라이트톨 (DTT; Sigma-Aldrich 646563), 1.32 μl의 뉴클레오타이드 혼합물 "N" (50 μl의 10 mM dNTP 혼합물 (NEB; N0447)과 25 μl의 100X NAD+(NEB; B9007)를 배합한다) 및 0.88 μl의 효소 혼합물 "E" (20 μl의 T4 폴리펩타이드 키나제 (NEB; M0201), 10 μl의 전장 *Bst*I 폴리머라제 (NEB; M0328) 및 10 μl의 *Taq* DNA 리가제 (NEB; M0208)를 배합한다). 반응 혼합물은 15분 동안 37 °C에서 항온처리하고 이어서 15분 동안 65 °C에서 항온처리하였다.

[0302]

반응물은 열 순환기로부터 제거하고 48 ul의 비드 재현탁 용액 (19 % PEG8000, 2 M NaCl, 10 mM Tris pH 8.0, 10 mM EDTA, 0.1 % Tween 20)을 반응물에 첨가하고 10분 동안 실온에서 항온처리하였다. 상기 비드는 200 μl의 70 % 에탄올로 2회 세척하고, 간략하게 공기 건조시키고 (~5분) 25 μl의 TEzero 중에 재현탁시켰다. 자기를 사용하여 비드를 위치시키고 세정된 DNA 라이브러리는 새로운 반응 용기로 이동시켰다.

[0303]

VI. 개요

[0304]

본원의 전반에 걸쳐 고려되고 실시예 1에 기재된 방법을 사용하여 작제된 수득한 dNA 라이브러리는 차세대 서열 분석, qPCR 분석 및 하나 이상의 표적 유전학적 유전자좌의 다른 정량적 유전학적 분석을 위해 준비되고 적합한 증폭이다.

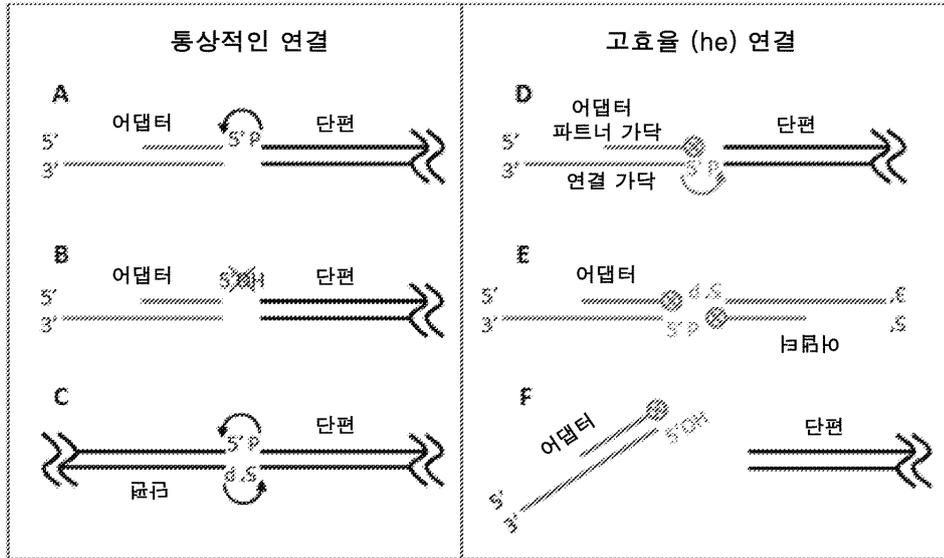
[0305]

일반적으로, 하기의 청구항에서, 사용된 용어는 청구 사항을 본원 명세서 및 청구항에 기재된 특정 구현예로 제한하는 것으로 해석되지 말아야 하고 그러한 청구항이 부여되는 전체 범위의 등가물과 함께 모든 가능한 구현예

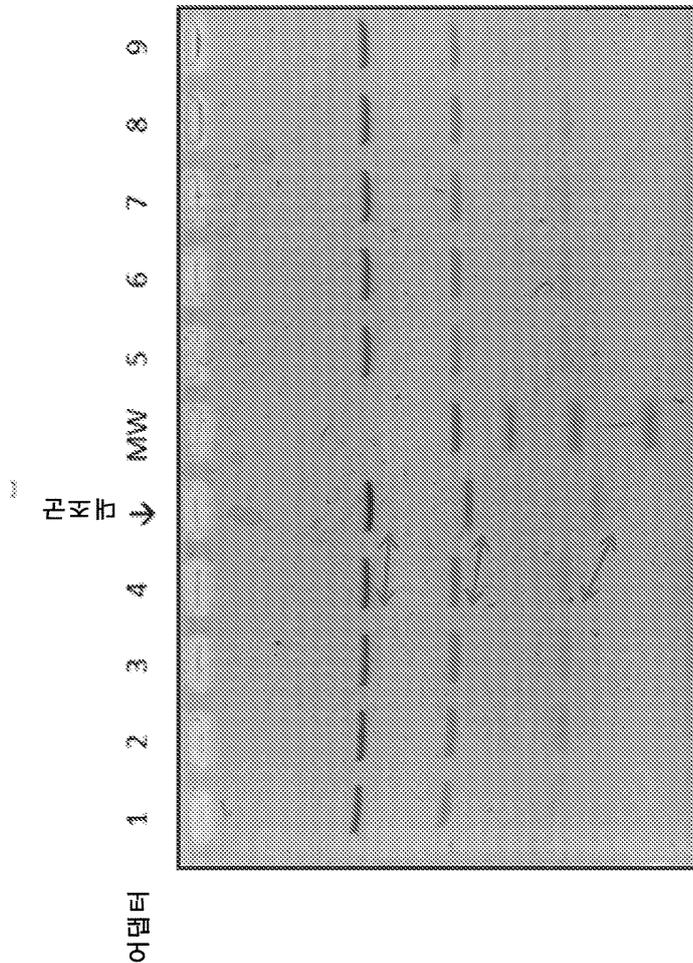
를 포함하는 것으로 해석되어야 한다. 따라서, 청구항은 개시내용에 의해 제한되지 않는다.

도면

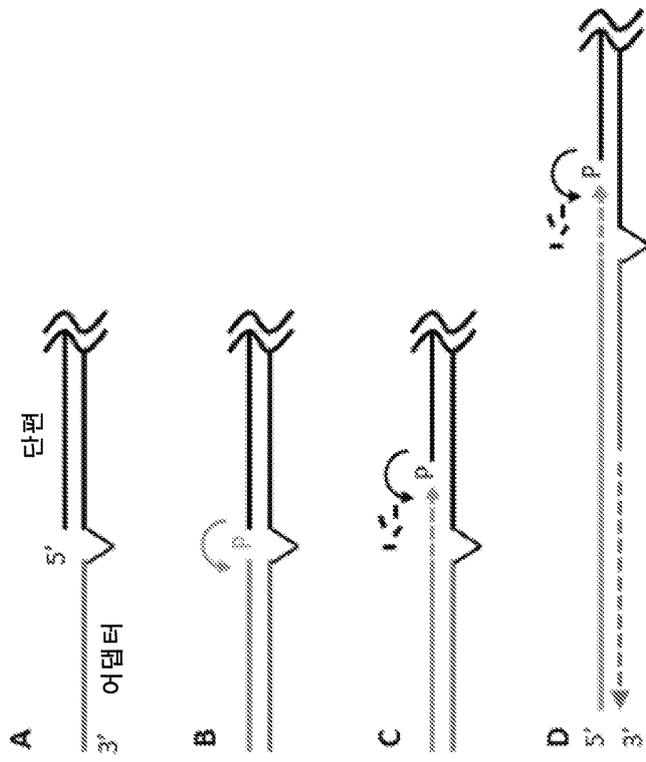
도면1



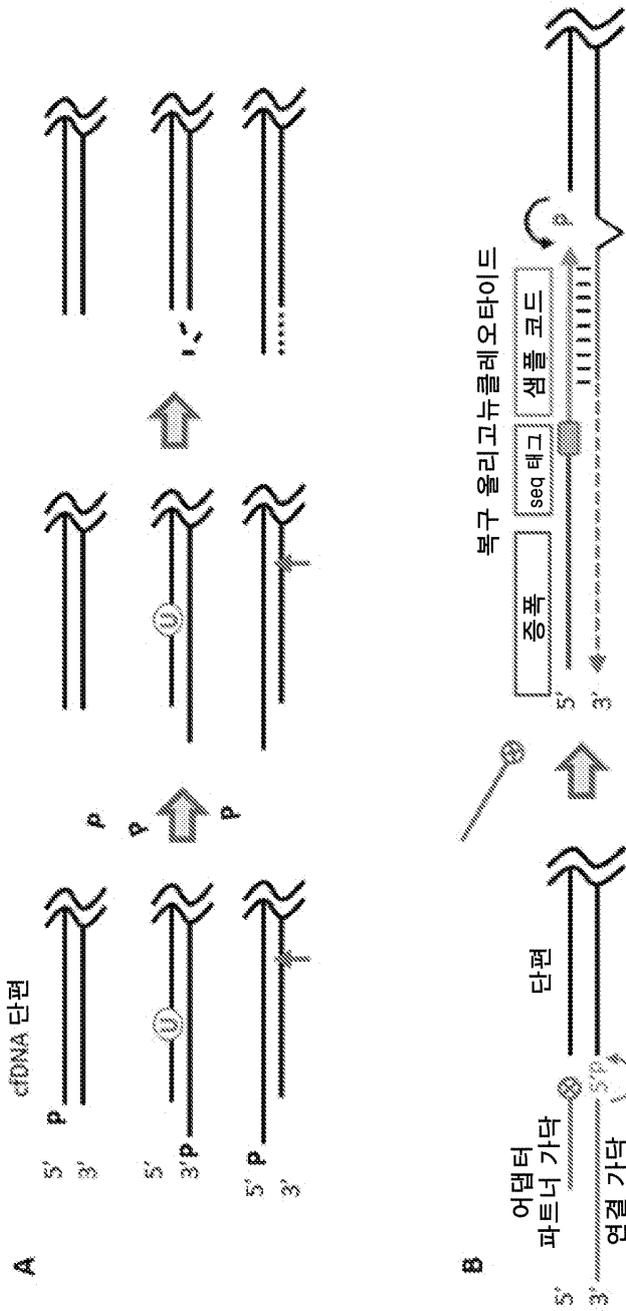
도면2



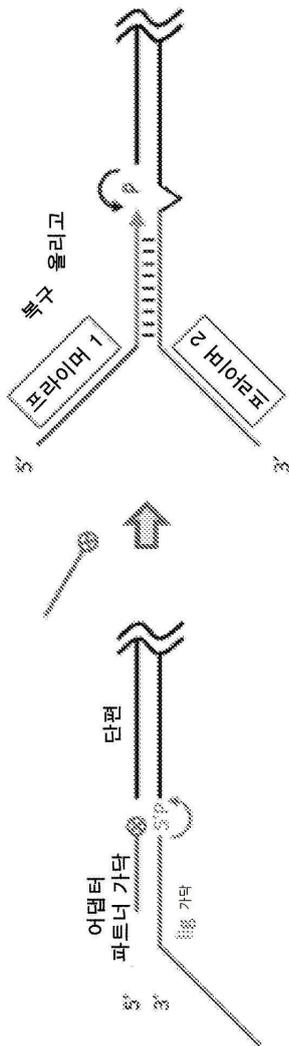
도면3



도면4



도면5



서열 목록

- <110> Resolution Bioscience, Inc.
Raymond, Christopher K
Lim, Lee P
Armour, Christopher D
- <120> HIGH EFFICIENCY CONSTRUCTION OF DNA LIBRARIES
- <130> CLFK-003/01W0
- <150> US 62/254,110
- <151> 2015-11-11
- <160> 62
- <170> KoPatent In 3.0
- <210> 1
- <211> 46
- <212> DNA

<213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400>
 > 1
 gagggctctac cttctttnnn ntgtattcga attctctggt cctgca 46
 <210> 2
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400> 2
 gctctagacg tcatcgnnnn ntgtattcga attctctggt cctgca 46
 <210> 3
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400> 3
 ggatactcgt agcggcnnnn ntgtattcga attctctggt cctgca 46
 <210> 4
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature

<222> (17)..(21)
 <223> n is A, C, T or G
 <400> 4
 gtcacgagta gagaaanmn ntgtattcga attctctggt cctgca 46
 <210> 5
 <211> 46
 <212> DNA

 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400> 5
 gagggctctac cttagtnnnn ntgtattcga attctctggt cctgca 46
 <210> 6
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400> 6
 gctctagacg tcagagnnnn ntgtattcga attctctggt cctgca 46

 <210> 7
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <220><221> misc_feature
 <222> (17)..(21)
 <223> n is A, C, T or G
 <400> 7

ggatactcgt agcttcnmmn ntgtattcga attctctggg cctgca 46

<210> 8

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized ligation strand oligonucleotide

<220><221> misc_feature

<222> (17)..(21)

<223> n is A, C, T or G

<400> 8

gtcacgagta gagccannnn ntgtattcga attctctggg cctgca 46

<210> 9

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized ligation strand oligonucleotide

<220><221> misc_feature

<222> (17)..(21)

<223> n is A, C, T or G

<400> 9

gagggtctac cttgctnnnn ntgtattcga attctctggg cctgca 46

<210> 10

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized partner strand oligonucleotide

<400> 10

aaggtagacc ct 12

<210> 11

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized partner strand oligonucleotide

<400> 11
 tgacgtctag ag 12
 <210> 12
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400> 12
 gctacgagta tc 12

 <210> 13
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400> 13
 ctctactcgt ga 12
 <210> 14
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400> 14
 aaggtagacc ct 12
 <210> 15
 <211> 12
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthesized partner strand oligonucleotide
 <400> 15
 tgacgtctag ag 12
 <210> 16
 <211> 12
 <212> DNA

<213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400> 16
 gctacgagta tc 12
 <210> 17
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400>
 > 17
 ctctactcgt ga 12
 <210> 18
 <211> 12
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized partner strand oligonucleotide
 <400> 18
 aaggtagacc ct 12
 <210> 19
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 19
 ctgagctagt 10

 <210> 20
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 20
 gactcgatag 10
 <210> 21

<211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 21
 tcagatcgtc 10
 <210> 22
 <211> 10
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthesized ligation strand oligonucleotide
 <400> 22
 agtctagcca 10
 <210> 23
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 23
 ggattaccct 10
 <210> 24
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide

 <400> 24
 cttacggatg 10
 <210> 25
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 25

accgattgac	10
<210> 26	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthesized ligation strand oligonucleotide	
<400> 26	
tagcgcata	10
<210> 27	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthesized ligation strand oligonucleotide	
<400> 27	
atgtccagct	10
<210> 28	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthesized ligation strand oligonucleotide	
<400> 28	
cacaggttag	10
<210> 29	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthesized ligation strand oligonucleotide	
<400> 29	
tgacatgctc	10
<210> 30	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223> Synthesized ligation strand oligonucleotide
 <400> 30
 gctgtacaga 10
 <210> 31
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 31
 tcaagtcggt 10
 <210> 32
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 32
 gttcagactg 10
 <210> 33
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 33
 caggtctaac 10
 <210> 34
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 34
 agctcagtca 10
 <210> 35
 <211> 10

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 35
 gatccgtact 10
 <210> 36
 <211> 10
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthesized ligation strand oligonucleotide
 <400> 36
 acagtcgtag 10
 <210> 37
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide
 <400> 37
 tggtaacctc 10
 <210> 38
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized ligation strand oligonucleotide

 <400> 38
 ctcagtagga 10
 <210> 39
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 39

tgcaggacca gagaattcga atacannnnn actagctcag 40

<210> 40

<211> 40

<

212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 40

tgcaggacca gagaattcga atacannnnn ctatcgagtc 40

<210> 41

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 41

tgcaggacca gagaattcga atacannnnn gacgatctga 40

<210> 42

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 42
 tgcaggacca gagaattcga atacannnnn tggctagact 40

<210> 43
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 43
 tgcaggacca gagaattcga atacannnnn agggtaatcc 40

<210> 44
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 44
 tgcaggacca gagaattcga atacannnnn catccgtaag 40

<210> 45
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 45
 tgcaggacca gagaattcga atacannnnn gtcaatcgg 40
 <210> 46
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 46
 tgcaggacca gagaattcga atacannnnn tcatgcgcta 40
 <210> 47
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 47
 tgcaggacca gagaattcga atacannnnn actgctagca 40
 <210> 48
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 48
 tgcaggacca gagaattcga atacannnnn cagcgatcat 40
 <210> 49
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 49
 tgcaggacca gagaattcga atacannnnn gtcacgatg 40
 <210> 50
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 50
 tgcaggacca gagaattcga atacannnnn tgatagctgc 40
 <210> 51
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 51

tgcaggacca gagaattcga atacannnnn agctggacat 40

<210> 52

<211> 40

<

212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 52

tgcaggacca gagaattcga atacannnnn ctaacctgtg 40

<210> 53

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 53

tgcaggacca gagaattcga atacannnnn gagcatgtca 40

<210> 54

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 54
 tgcaggacca gagaattcga atacannnnn tctgtacagc 40

<210> 55
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 55
 tgcaggacca gagaattcga atacannnnn accgacttga 40

<210> 56
 <211> 40
 <
 212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

<222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 56
 tgcaggacca gagaattcga atacannnnn cagtctgaac 40

<210> 57
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 57

tgcaggacca gagaattcga atacannnnn gttagacctg 40

<210> 58

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 58

tgcaggacca gagaattcga atacannnnn tgactgagct 40

<210> 59

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide

<220><221> misc_feature

<222> (26)..(30)

<223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

<400> 59

tgcaggacca gagaattcga atacannnnn agtacggatc 40

<210> 60

<211> 40

<

212> DNA

<213> Artificial Sequence

<220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 60
 tgcaggacca gagaattcga atacannnnn ctacgactgt 40
 <210> 61
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature

 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3
 <400> 61
 tgcaggacca gagaattcga atacannnnn gaggttacca 40
 <210> 62
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthesized repair oligonucleotide
 <220><221> misc_feature
 <222> (26)..(30)
 <223> n is A, C, T or G, where the specific groups of bases from positions 26 to 30 are listed in Table 3

 <400> 62
 tgcaggacca gagaattcga atacannnnn tcctactgag 40