(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년05월13일
(11) 등록번호 10-2806552
(24) 등록일자 2025년05월08일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6869 (2018.01)
(52) CPC특허분류
C12Q 1/6869 (2018.05)
C12Q 2525/179 (2013.01)
(21) 출원번호 10-2023-7014184(분할)
(22) 출원일자(국제) 2017년05월17일
심사청구일자 2023년05월23일
(85) 번역문제출일자 2023년04월26일
(65) 공개번호 10-2023-0065357
(43) 공개일자 2023년05월11일
(62) 원출원 특허 10-2018-7036548
원출원일자(국제) 2017년05월17일
심사청구일자 2020년05월14일
(86) 국제출원번호 PCT/EP2017/061902
(87) 국제공개번호 WO 2017/198742
국제공개일자 2017년11월23일
(30) 우선권주장
16169997.0 2016년05월17일
유럽특허청(EPO)(EP)
1701908.4 2017년02월06일 영국(GB)
(56) 선행기술조사문헌
WO2003052101 A1
US20090298049 A1
US20150322508 A1

(73) 특허권자
디네임-아이티 엔브이
벨기에, 비-3001 헤벌리, 가스톤 긴슬란 1
(72) 발명자
쿠펜스, 해리
벨기에, 비-1000 브뤼셀, 비이올렛스트라트 35 버스 11
(74) 대리인
특허법인 무한

전체 청구항 수 : 총 19 항 심사관 : 이준혁

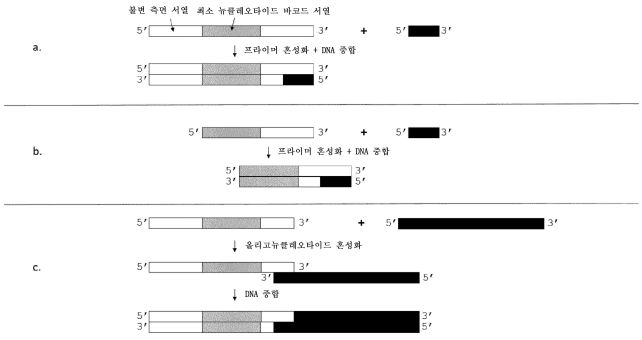
(54) 발명의 명칭 **시료의 식별 방법**

(57) 요 약

본 발명은 아이템을 표지하기 위한 핵산과 관련된 방법, 담체 및 벡터를 기재하고 있으며, 여기서, 각각의 담체는 시료 핵산 이외에, 표지를 위한 적어도 2개의 뉴클레오타이드 바코드 핵산을 포함하며, 각각의 뉴클레오타이드 바코드 핵산은 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 적

(뒷면에 계속)

대 표 도 - 도2



어도 2개의 상기 상이한 뉴클레오타이드 바코드 핵산들은 동일한 길이의 최소(minimal) 뉴클레오타이드 바코드 서열을 가지며, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 이전성(transferable) 분자 식별 바코드를 생성하고, 이로써 이러한 각각의 이전성 분자 식별 바코드가 수합물 내의 각각의 담체에 대해 상이하게 된다.

(52) CPC특허분류
C12Q 2535/122 (2019.08)
C12Q 2545/101 (2013.01)
C12Q 2563/179 (2013.01)
C12Q 2563/185 (2013.01)

명 세 서

청구범위

청구항 1

복수의 핵산을 함유하는 생물학적 시료 식별 방법으로,

상기 방법은:

a) 복수의 담체들을 제공하는 단계로서, 상기 담체는 각각 핵산을 함유하는 생물학적 시료를 포함하는 기판(substrate) 또는 용기이며,

여기서, 각각의 담체는 핵산을 포함하는 시료 외에도, 상기 담체를 표지하기 위해 적어도 2개의 뉴클레오타이드 바코드 핵산(nucleotide barcode nucleic acids, NBNA)을 포함하고, 이들 적어도 2개의 뉴클레오타이드 바코드 핵산은 각각 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열(minimal nucleotide barcode sequence, MNB)을 포함하며, 여기에서 적어도 2개의 상기 상이한 뉴클레오티드 바코드 핵산은 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖고,

이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 핵산을 포함하는 생물학적 시료를 함유하는 상기 각각 담체의 특이적 식별을 위한 이전성 분자 식별 바코드(transferable molecular identification barcode, TMIB)를 생성하고, 이로써 각각의 이전성 분자 식별 바코드는 각각의 담체에 대해 상이하며, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열(identifier sequences, IS)이 존재하고, 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열(extracting sequences, ES)이 존재하여, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하고,

각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지(macroscopic barcode label, MBL)를 함유하는, 단계,

b) 생물학적 시료의 핵산 내의 하나 이상의 표적 서열을 시퀀싱하고, 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산 내의 표적 서열을 시퀀싱하는 단계로서, 여기서, 생물학적 시료의 핵산 내에서 표적 서열의 시퀀싱 및 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 시퀀싱은 병행 시퀀싱 방법에 의해 수행되고, 병행 시퀀싱 방법 전에, 생물학적 시료의 핵산 내의 표적 서열 및 상이한 시료 유래의 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열을 풀링(pooling)하는 단계가 수행되는, 단계,

c) 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 확인하고 선별하는 단계로서, 상기 단계는 단계 b)에서 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 선별하는 단계를 포함하고, 여기서, 선별 단계는 최소 뉴클레오타이드 바코드 서열에 인접한 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 서열의 식별을 포함하는, 단계;

d) 단계 c)의 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 단리된 서열 내에서 최소 뉴클레오타이드 바코드 서열을 확인하고 선별하는 단계로서, 상기 단계는 한정된(defined) 길이에서 2개의 추출 서열들 사이에 존재하는 서열을 선별하거나 한정된 길이에서 1개의 추출 서열에 인접하여 존재하는 서열을 선별하고, 이들 선별된 서열을 최소 뉴클레오타이드 바코드 서열로서 확인하는 단계를 포함하는, 단계,

e) 단계 d)에서 확인된 최소 뉴클레오타이드 바코드 서열을, 담체에 제공된 거시적 바코드 표지를 기반으로 예상된 최소 뉴클레오타이드 바코드 서열과 비교함으로써, 시료 또는 오염을 식별하는 단계

를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 b) 단계는 어댑터를 시료 내의 표적 핵산, 및 뉴클레오타이드 바코드 핵산에 연결하는 단계, 및 연결된 생성물을 시퀀싱 주형으로서 사용하여 표적 서열 및 뉴클레오타이드 바코드 핵산을 시퀀싱하는 단계를 포함하는, 방법.

청구항 3

제1항 또는 제2항에 있어서,

상기 b) 단계는 상기 핵산 시료 내 표적 서열의 농화(enrichment) 단계를 수행하는 단계 및 뉴클레오타이드 바코드 핵산의 농화 단계를 수행하는 단계를 포함하는, 방법.

청구항 4

제3항에 있어서,

상기 농화 단계는 포착 또는 증폭을 사용하여 수행되며, 상기 증폭은 1-단계 PCR, 2-단계 PCR, 프라이머 연장 및 후속해서 연결(ligation) 및 PCR, 또는 서클러라이제이션(circularisation)-기반 증폭에서 선택된 것인, 방법.

청구항 5

제1항에 있어서,

상기 b) 단계는 시료의 표적 서열에 시료-특이적 풀링 바코드를 부착하고, 동일한 시료-특이적 풀링 바코드를 상기 시료 내의 뉴클레오타이드 바코드 핵산에 부착하는 단계를 포함하는, 방법.

청구항 6

아이템을 표지하기 위한 핵산을 포함하는 담체의 수합물(collection)로서, 상기 담체는 기판 또는 용기이고, 여기서, 각각의 담체가 시료 핵산 이외에, 표지를 위한 적어도 2개의 뉴클레오타이드 바코드 핵산들을 포함하며, 각각의 뉴클레오타이드 바코드 핵산이 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 핵산이 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 가지고, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 핵산을 포함하는 시료를 함유하는 상기 담체 각각의 특이적 식별을 위한 이전성 분자 식별 바코드를 생성하며, 이로써 각각의 이러한 이전성 분자 식별 바코드가 수합물 내의 각각의 담체에 대해 상이한 것이고,

상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하거나, 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하여, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하고, 각각의 담체가 상기 담체에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는 것을 특징으로 하는, 수합물.

청구항 7

제6항에 있어서,

상기 수합물이 적어도 100개의 담체, 또는 적어도 10000개의 담체를 포함하는, 수합물.

청구항 8

제6항 또는 제7항에 있어서,

상기 표지를 위한 뉴클레오타이드 바코드 핵산이, 핵산을 RNA로 전사시키기 위한 서열을 포함하지 않는, 수합물.

청구항 9

제6항에 있어서,

상기 시료가 DNA를 함유하는 시료 또는 RNA를 함유하는 시료인, 수합물.

청구항 10

제6항에 있어서,

최소 뉴클레오타이드 바코드 서열의 측면에, 증폭에 의해 최소 뉴클레오타이드 바코드 서열을 농화시키기 위한 올리고뉴클레오타이드 서열이 존재하거나,

최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 뉴클레오타이드 바코드 서열 내의 1개 또는 2개의 올리고뉴클레오타이드 결합 서열에서 혼성화-기반 서열 포착을 허용하는 하나 이상의 올리고뉴클레오타이드 결합 서열이 존재하는, 수합물.

청구항 11

제6항에 있어서,

뉴클레오타이드 바코드 핵산이 클로닝 벡터의 단편에 포함되는, 수합물.

청구항 12

제11항에 있어서,

상기 벡터 또는 벡터 단편이 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하거나, 상기 벡터 또는 벡터 단편이 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 수합물.

청구항 13

제6항에 있어서,

상기 뉴클레오타이드 바코드 핵산을 포함하는 핵산을 포착하거나 증폭시키기 위한 하나 이상의 프라이머를 추가로 포함하는, 수합물.

청구항 14

이전성 분자 식별 바코드를 포함하는, 기관 또는 용기인 담체의 수합물의 제조 방법으로서, 상기 방법은:

a) 적어도 4개 뉴클레오타이드의 최소 뉴클레오타이드 바코드 서열을 포함하는 상이한 뉴클레오타이드 바코드 핵산들의 제1 수합물을 제공하는 단계로서, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 핵산 서열들은 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 가지며, 이 중에서 최소 뉴클레오타이드 바코드 서열은 상기 수합물 내의 뉴클레오타이드 바코드 핵산들 사이에서 상이하고, 뉴클레오타이드 바코드 핵산 내에서 최소 뉴클레오타이드 바코드 서열의 식별을 허용하는 하나 이상의 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열 또는 하나 이상의 비-바이러스성 및 비-박테리아성 추출 서열을 보유하는, 단계,

b) 단계 a)의 적어도 2개의 뉴클레오타이드 바코드 핵산들의 조합을 각각의 담체에 첨가하여, 각각의 담체가 최소 뉴클레오타이드 바코드 서열들의 혼합물에서의 차이에 의해 한정되는 특이적 이전성 분자 식별 바코드를 갖는 담체들의 수합물을 수득하는 단계,

c) 각각의 담체를, 최소 뉴클레오타이드 바코드 서열들의 상이한 혼합물에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지로 표지하는 단계,

d) 거시적 바코드 표지와 최소 뉴클레오타이드 바코드 서열들의 혼합물을 포함하는 상기 이전성 분자 식별 바코드 사이의 관계를 저장하는 단계

를 포함하는, 방법.

청구항 15

제14항에 있어서,

상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하거나 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들

내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하는, 방법.

청구항 16

제14항 또는 제15항에 있어서,

상기 뉴클레오타이드 바코드 핵산이 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하거나, 상기 뉴클레오타이드 바코드 핵산이 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 방법.

청구항 17

이전성 분자 식별 바코드를 포함하는, 기판 또는 용기인 담체의 수합물의 제조 방법으로서, 상기 방법은:

a) 숙주에서 핵산 벡터들의 제1수합물을 제공하는 단계로서, 여기서 상기 벡터는 수합물 내에서 뉴클레오타이드 바코드 핵산 벡터들 사이에서 상이한 길이가 적어도 4개 뉴클레오타이드인 최소 뉴클레오타이드 바코드 서열을 갖는 뉴클레오타이드 바코드 서열을 포함하고, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 서열들이 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖고, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하는 하나 이상의 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열 또는 하나 이상의 비-바이러스성 및 비-박테리아성 추출 서열을 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면에 포함하는, 단계에 의해 제공되는 뉴클레오타이드 바코드 서열을 포함하는 벡터를 갖는 숙주들의 수합물을 제공하는 단계, 및

b) 숙주의 개별 콜로니를 제공하고, 복수의 콜로니들에 대한 핵산 벡터 내에서 뉴클레오타이드 바코드 서열을 시퀀싱하여, 단리된 콜로니들의 제2 수합물을 수득하는 단계로서, 여기서, 각각의 콜로니는 상이한 뉴클레오타이드 바코드 서열을 갖는 핵산 벡터를 포함하는 단계,

c) 콜로니의 선별을 위해, 상기 콜로니로부터 벡터를 단리하는 단계,

d) 담체의 수합물 내의 각각의 담체에, 상이한 뉴클레오타이드 바코드 서열을 갖는 적어도 2개의 핵산 벡터들의 조합을 첨가하여, 각각의 담체가 담체들 사이에서 최소 뉴클레오타이드 바코드 서열들의 혼합물에서의 차이에 의해 한정되는 특이적 이전성 분자 식별 바코드를 갖는 담체의 수합물을 수득하고, 벡터를 제한 효소에 의해 단편화하는 단계로서, 여기서, 적어도 2개의 상기 핵산 벡터들은 단계 b)의 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖는, 단계,

e) 각각의 담체를, 상이한 최소 뉴클레오타이드 바코드 서열들의 혼합물에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지로 표지하는 단계,

f) 거시적 바코드 표지와 최소 뉴클레오타이드 바코드 서열들의 서열을 포함하는 상기 이전성 분자 식별 바코드 사이의 관계를 저장하는 단계

를 포함하는, 방법.

청구항 18

제17항에 있어서,

상기 벡터가 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하거나, 상기 벡터가 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 방법.

청구항 19

기질 또는 용기인 담체 세트에서 뉴클레오타이드 바코드 핵산의 추적 방법으로서, 상기 방법은:

- 게놈 DNA 또는 RNA가 없으며, 상기 담체의 표지를 위한 적어도 2개의 뉴클레오타이드 바코드 핵산을 포함하는 복수의 담체들을 제공하는 단계로서, 여기서, 이들 적어도 2개의 뉴클레오타이드 바코드 핵산들은 각각 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 여기에서 적어도 2개의 상기 상이한 뉴클레오티드 바코드 핵산은 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖고,

이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 핵산을 포함하는 생물학적 시료를 함유하는 상기 담체 각각의 특이적 식별을 위한 이전성 분자 식별 바코드를 생성하며, 이로써 각각의 이전성 분자 식별 바코드는 각각의 담체에 대해 상이하고, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산 내의 최소 뉴클레오타이드 바코드 서

열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고, 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하여, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하고,

각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는 단계,

- 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산 내에서 표적 서열을 병행 시퀀싱 방법에 의해 시퀀싱하는 단계로서, 여기서, 병행 시퀀싱 방법 전에 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 풀링이 수행되는, 단계,

- 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 확인 및 선별하는 단계로서, 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 선별하는 단계를 포함하며, 여기서 선별 단계는 최소 뉴클레오타이드 바코드 서열에 인접한 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 서열의 식별을 포함하는, 단계;

- 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 단리된 서열 내에서 최소 뉴클레오타이드 바코드 서열을 확인 및 선별하는 단계로서, 한정된 길이에서 2개의 추출 서열들 사이에 존재하는 서열을 선별하거나 한정된 길이에서 1개의 추출 서열에 인접하여 존재하는 서열을 선별하고, 이들 선별된 서열을 최소 뉴클레오타이드 바코드 서열로서 확인하는 단계를 포함하는, 단계,

- 담체에 제공된 거시적 바코드 표지를 기반으로, 확인된 최소 뉴클레오타이드 바코드 서열을 예상된 최소 뉴클레오타이드 바코드 서열과 비교하는 단계

를 포함하는, 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 다수의 독특한 이전성(transferable) 분자 식별 바코드들의 생성 방법, 및 예컨대 생물학적 시료 내 아이템의 식별에 있어서 이들 바코드들의 용도, 및/또는 이들의 프로세싱에 관한 것이다. 본 발명은 프로세스에 대한 내부 조절(internal control)을 제공하는 문제점들, 예컨대 시료 바뀜 및 시료의 교차-오염, 시료의 추적성(traceability)을 모니터링하는 방식에 관한 것이다.

배 경 기 술

[0002] DNA 분자의 첨가를 통해 시료를 내부 식별하는 아이디어는 이미 오래되었다. 예를 들어, 특허 US5776737은 시료의 내부 식별을 위한 방법 및 조성물을 기재하고 있다.

[0003] US5776737은 독특한 DNA 코드를 경제적인 방식으로 수득하기 위해 DNA 분자의 혼합물들을 사용하는 힘을 인지하고 있다. DNA 코드의 특징화를 위해, Sanger 시퀀싱을 기반으로 하는 Pharmacia ALF 자동 시퀀서가 사용된다.

[0004] US6030657은 생성물 전환 및 생성물 위조에 대항하기 위한 생성물을 표지하기 위해 추가로 적외선(IR) 마커로 표지된 캡슐화된 DNA를 바이오마커로서 이용하는 표지/마킹 기술을 기재하고 있다. 실제 DNA 바이오마커 서열은 보안(security)을 위해 이차적인 고려 사항이었다.

[0005] EP1488039는 복수의 상이한 단일 가닥 DNA 서열들의 용도를 인지하고 있다. 이 문헌에서, DNA 바코드는 캐쉬 트랜스포트 박스(cash transport box)에 대한 보안 마커로서 사용된다. 그러나, 상이한 단일 가닥 유형의 DNA 서열들은 생성 동안 시작 시로부터 혼합되지 않는다. 오로지 1가지 유형의 DNA 올리고뉴클레오타이드만 입수 가능한 서로 다른 유형의 올리고뉴클레오타이드들로부터 선택되고, 오로지 하나의 카세트의 잉크 저장소(reservoir) 내로 삽입된다.

[0006] US20120115154는 게놈에 존재하는 것으로 공지되지 않은 하나 이상의 올리고뉴클레오타이드를 포함하는 참조 마커가 생물학적 시료에 첨가되는 방법을 개시하고 있다. 첨가되는 참조 마커는 단일 서열 또는 상이한 서열들의 혼합물이다. 그러나, 상이한 서열들의 혼합물을 사용하는 목적은, 훨씬 더 높은 수준의 특이성 및/또는 보안을 제공하기 위한 것이며, 따라서 많은 수의 독특한 참조 마커들을 경제적인 방식으로 제공하기 위한 것이 아니다.

[0007] US20120135413은 보안 마킹에서 올리고뉴클레오타이드 또는 바코드의 혼합물을 사용한다. 이 문헌에서, 혼합물을 사용하는 목적은 단지, 하나의 더 큰 합성 올리고뉴클레오타이드가 사용되는 경우에 수득할 수 있는 것보다 독특한 코드의 수와 유사할 충분히 많은 수의 독특한 코드를 생성하기 위해 더 짧은 올리고뉴클레오타이드들을 사용할 수 있기 위해서이다.

[0008] US6312911은, 모든 3개의 DNA 염기들이 문자 또는 기호를 나타내는 비밀 메세지를 암호화하기 위해 DNA 단편을 사용한다. 그런 다음, 비밀 DNA는 은폐(concealing) DNA들의 혼합물에 은폐된다. 비밀 DNA 코드가 프라이머 서열의 측면에 존재하는 경우, 이러한 코드는 증폭 및 시퀀싱에 의해 복잡한 은폐 혼합물로부터 특이적으로 지칭될 수 있다. 그런 다음, 인코딩된 메세지가, 보안 마커 내에서 DNA 단편의 시퀀싱 및 디코딩(decoding)에 대한 암호화 참조 표의 사용에 의해 디코딩된다.

[0009] US8785130은 유전 물질의 검출 및 식별 방법을 모니터링하기 위한 뉴클레오타이드 서열-기반 코드의 용도를 기재하고 있다. 이들은 상이한 DNA 서열들을 이용하지만, 이들은 모두 더 큰 단일 DNA 분자 상에 위치한다. 따라서, 해당 방법은 DNA 코드의 경제적인 생성을 위해 DNA 서열들의 혼합물을 사용하는 이득을 갖지 않는다.

[0010] WO2014005184는 상이한 핵산들의 혼합물을 사용하는 식별 또는 마킹 방법을 개시하고 있다. 그러나, 특징화를 위해, 이들은 상이한 크기를 갖는 복수의 증폭 생성물들을 생성하며, 이는 시퀀싱보다는 상이한 핵산 태그 서열들 사이의 구별에 대한 근거이다.

[0011] US20100285985는 복수의 보안 마커들의 생성 및 이들의 검출을 위한 방법 및 시스템을 기재하고 있다. 각각의 보안 마커는 DNA 주형 상에서 프라이머로서 사용되는 올리고뉴클레오타이드들의 혼합물이다. 따라서, 올리고뉴클레오타이드는 rtDNA(역(reverse) 주형 DNA) 올리고뉴클레오타이드로 지칭된다.

[0012] EP2201143은 이전성 분자 식별 바코드의 근거로서, 단일 DNA 분자보다는 상이한 DNA 분자들의 혼합물을 사용하는 힘을 인지하고 있다.

발명의 내용

해결하려는 과제

[0013] 전체 산업들은 다양한 제품 및 서비스를 포함하는 표지 주위로 구축되어 있다. 소매 산업에서, 대부분의 제품들은 1D(라인) 또는 2D 바코드 표지로 표지되어 있어서, 소매업체에서 제품 재고 관리 및 계산대 관리를 용이하게 한다. 택배 배달 서비스에서, 패키지에 바코드 표지가 표지되어 있어서, 전세계에 걸쳐 패키지의 운송은 자동화될 수 있고, 패키지 위치는 심지어 소비자에 의해 실시간으로 추적될 수 있다. 또한, 검사 실험실에서, 연구되는 시료 튜브가 바코드 표지로 표지된다.

[0014] 그러나, 이들 모든 아이템들(제품, 패키지, 시료 튜브 등)은 아이템의 외부에서 표지된다. 일단, 해당 아이템이 개봉되고 아이템의 내용물이 제거되면, 바코드와 아이템의 내용물과의 연관성은 상실된다. 식품 및 선박 패키지의 경우, 아이템의 언패킹(unpacking)은 통상적으로(대체로) 아이템의 생활사(lifecycle) 중 마지막 단계이고, 따라서 외부 바코드와의 연관성의 상실이 문제가 되지 않는다. 글나, 소정의 아이템, 예컨대 생물학적 시료를 함유하는 시료 튜브의 경우, 실제 프로세싱은 단지 그때부터 출발한다.

[0015] 본원에서, 본 발명자들은 아이템의 외부 및 내용물 둘 다에 바코드가 표지되는 해결방안을 제공한다. 아이템의 외부 파트는 물리적인 거시적 바코드 표지(예를 들어 광학 바코드 페이퍼 또는 RFID)로 표지되는 한편, 아이템의 내용물은 이전성 분자 식별 바코드 표지로 표지된다. 물리적 바코드 표지 및 이전성 분자 식별 바코드 표지는 둘 다 독특하고, 일대일 관계를 가진다. 바코드 표지들 중 하나가 공지되어 있는 경우, 다른 바코드가 또한 이러한 일대일 관계를 기반으로 공지된다. 바코드 표지들 중 하나(예를 들어 물리적 바코드 표지)가 심지어 제3의 바코드 표지와 연관되어 있는 경우, 다른 바코드(이러한 예에서는 이전성 분자 식별 바코드 표지) 또한, 이러한 제3의 바코드 표지와 연관되어 있다. 물리적 바코드 표지와는 대조적으로, 이전성 분자 식별 바코드 표지는 아이템의 내용물의 완전한 다운스트림 프로세싱 사슬에 이전되고, 프로세싱의 종료 시 판독된 다음 다시 다른 모든 연관된 바코드 표지들과 연관될 수 있다.

[0016] 대부분의 프로세스들은 특히 이전 단계가 존재하는 순간에 오차에 취약하다. 이는 또한, 진단 검사에도 적용되는데, 이러한 검사는 수여자(recipient)에서 환자로부터의 생물학적 시료, 예컨대 Vacutainer® 튜브 내의 혈액 시료를 단리하는 것에서 출발한다. 인쇄된 GS1 바코드 표지 및 바코드 스캐너의 사용은 오차를 최소화할 수 있지만, 항상 오차를 방지하는 것은 아니다. 사실상, 일부 아이템들은 인쇄된 바코드 표지로 표지될 수 없다. 예

를 들어, 생물학적 시료로부터 단리된 DNA가 중합효소 연쇄 반응(PCR)에 의해 증폭되어야 하는 경우, DNA는 작은 PCR 튜브로 이전된다. 인쇄된 바코드 표지는 이러한 작은 튜브에 부착되기에는 너무 큰 크기를 가지고/거나 PCR 프로세스에 악영향을 미칠 수 있다. 사실상, 페이퍼 표지를 PCR 튜브의 외벽에 고정하는 것은 PCR 튜브의 벽을 통한 효율적인 열 전달을 방해할 수 것이며, 이는 PCR에 악영향을 미치고 심지어 PCR 증폭을 방해할 수 있을 것이다. 그런 다음, PCR 튜브는 펜슬로 간단하게 표지된다. 그러나 그런 경우에도, PCR 튜브 상에 기재할 만한 장소는 제한되어 있어서, 독특한 코드의 경우 튜브에 기재되기 힘들 수 있다. 이따금, 펜슬 코드는 프로세싱 동안 펜슬로 표지된 튜브로부터 사라진다. 많은 실험실에서 표준 작업 절차(SOP; standard operating procedure)를 사용하고 LIMS(실험실 정보 관리 시스템)를 이용한다. 이들은 검사의 상이한 단계들 동안 수행되어야 하는 작업을 상세하게 기재하고 추적하며, 이는 사실상 오차의 기회를 감소시키지만, 이들 기재된 작업이 올바르게 실시되며 시료가 올바르게 취급되고 바뀌지 않음을 항상 보장하는 것은 아니다. 시료가 바뀐 경우, 결국 잘못된 검사 결과가 주어진 환자에 대해 기록된다. 특히 진단 시, 이들 오차는 관용될 수 없다. 이러한 시료 혼동은 이따금 병원 및 실험실에서 발생한다. 18,955명의 임산부들의 혈액 시료에서의 비-침습적 산전 검사의 발견을 보고하는 연구에서, 384명의 혈액 시료(2%)에서 혈액-수합 또는 표지 오차가 존재하였다(문헌[Norton et al., (2015) N. Engl. J. Med 372:1589-1597]). 이들 시료를 더 검사하지 않았기 때문에, 이들은 잘못된 검사 결과를 초래하지는 못했을 것이다. 그러나 이러한 경우, 검사 결과가 여전히 요망되는 경우에는 새로운 혈액 시료를 요청해야한 하고, 이로 인해 추가 시간이 소요되고 검사 결과가 지연되며, 결국에는 심지어 검사 결과를 기반으로 한 작업이 취해질 수 있는 시기를 넘길 수도 있다. 새로운 혈액 시료가 요청될 수 없는 경우, 검사 결과가 전혀 제공될 수 없다. 그러나, 검사 프로세스 중 심지어 다운스트림에서, 실험실에서 검사가 실제로 수행되는 경우, 오차가 발생한다. 소비자-직접(direct-to-consumer) 검사 회사는 유전자 데이터를 검수했을 때 96명 이하의 소비자들에서 이들 자신의 것이 아닌 것으로 나타난 실험실 혼동을 보고하였다. 이러한 혼동은, 단일 96-웰 플레이트를 시료의 프로세싱 동안 잘못 놓은 인간의 오차에 의해 유발되었다 (http://blog.23andme.com/23andme-and-you/update-from-23andme/).

[0017] 시료 바뀜 외에도, 시료 또는 이의 프로세싱된 시료 유도체가 또 다른 생물학적 시료 또는 이의 프로세싱된 시료 유도체로 인해 오염될 수 있으며, 이 또한 잘못된 검사 결과를 초래할 수 있다. 예를 들어, 유전자 검사에서 주어진 유전자좌에서 주어진 돌연변이에 대해 동형접합성인 주어진 시료가 동형접합성인 시료 또는 심지어 이형접합성인 시료에 의해 오염되게 된 경우, 주어진 해당 유전자좌에서 야생형 대립유전자에 대해, 주어진 해당 유전자좌에서 돌연변이에 대한 이형접합성 상태가 잘못 결론내려지고 기록될 수 있을 것이다. 모체 혈액에서 순환 태아 DNA를 분석하는 검사에서, 태아 분획은 적어도 4% 내지 10%이고, 이러한 분획 내에서 DNA 변칙(anomaly)이 총 DNA 백그라운드에 대해 검출되어야 한다. 따라서, 심지어 총 시료의 작은 오염도 검사를 방해하거나 심지어 잘못된 검사 결과를 초래할 수 있다. 혈액 시료 또는 다른 생물학적 시료 내에서 순환 종양 DNA를 분석하는 검사에도 마찬가지로 적용된다. 217개의 완전한 게놈을 시퀀싱한 연구에서, 7개의 시료(3.2%)가 오염성 DNA를 함유하는 것으로 확인되었다(문헌[Taylor et al., 2015]). 따라서, 시료의 오염 또는 혼동의 추적이 명백하게 요망되고 있다.

[0018] 상업 및 보안 분야에서 마킹 및/또는 추적 아이템이 또한 많은 관심을 끌고 있다. 예를 들어, 많은 제품들은 제품의 아이덴터티 또는 출처를 확인할 수 있는 태그로 마킹된다. 다른 상황에서, 제품은 이러한 제품을 추적할 수 있는 수단으로서 태그로 마킹된다. 이러한 마킹 및 추적 시스템은 또한, 물체가 하나의 위치에서 또 다른 위치로 옮겨감에 따라 이러한 물체의 경로 및/또는 시기를 추적하는 데 사용될 수 있다.

[0019] 마킹 시스템은 또한, 진품을 식별하고 이러한 진품을 위조품으로부터 구분하거나, 병행 무역 사례를 식별하는 데 사용될 수 있다.

[0020] 또한, 제품의 출처를 식별하는 것이 필요할 수 있는 경우, 예컨대 식품 산업에서 예컨대 또 다른 제품 또는 환경을 오염시키는 성분이 존재하는 상황에서 발생할 수 있는 경우가 존재한다.

[0021] 본 발명에서, 분자들의 혼합물은 하나 이상의 하기 기준을 충족하는 아이템의 독특한 내부 가용성 표지에 사용되며, 이는 이들 아이템 및/또는 이들의 프로세싱의 명백한 식별을 가능하게 한다.

[0022] 이러한 경우, DNA 분자들의 혼합물은 내부 표지에 사용된다.

[0023] - 단일 DNA 분자보다는 DNA 분자들의 혼합물이 사용되어, 독특한 DNA 코드의 경제적인 생성을 가능하게 하며, 이들 코드는 여기서 이전성 분자 식별 바코드로 지칭된다. 이러한 방식으로, 오로지 제한된 수의 DNA 분자만, 바람직하게는 오로지 1회 사용되는 다수의 독특한 DNA 코드의 경제적인 생성을 허용한다.

[0024] - 이러한 이전성 분자 식별 바코드가 아이템 또는 시료의 독특한 식별에 사용되기 때문에, 이들 바코드 자체가 가장 엄격한 품질 조건 하에 생성되는 것이 중요하다. 생성된 이전성 분자 식별 바코드, 또는 이러한 이전성 분자 식별 바코드를 함유하는 수합기(수합기)의 품질 관리는 해당 이전성 분자 식별 바코드/수합기의 파괴를 초래하며, 따라서 이러한 바코드/수합기는 더 이상 사용될 수 없다. 바코드/수합기가 여전히 사용되는 경우, 사실상 이것이 2회 또는 심지어 그보다 많이 사용될 경우, 이전성 분자 식별 바코드는 독특하게 사용되지 않는다. 이전성 분자 식별 바코드로서 단일의 독특한 DNA 분자는 생성 후 품질 관리를 허용하지 않는다. 이전성 분자 식별 바코드 혼합물이 제한된 수의 DNA 분자로부터 출발하여 생성되는 경우, 오로지 소수의 이전성 분자 식별 바코드 튜브가 품질 관리를 위해 생성 후 희생되어야 한다. 예상되는 서열이 이들 희생된 이전성 분자 식별 바코드에서 확인되는 경우, 다른 희생되지 않은 이전성 분자 식별 바코드 내의 서열 또한, 올바른 것으로 결론내려질 수 있다.

[0025] - DNA 분자 쌍이 사용되어, 이전성 분자 식별 바코드를 생성한다. DNA 분자의 혼합물이 소수의 DNA 분자로부터 출발하여 독특한 이전성 분자 식별 바코드의 생성에 사용되는 경우, 일부 이전성 분자 식별 바코드는 소정의 DNA 분자를 공통적으로 운반할 것이다. 사실상, 이전성 분자 식별 바코드는 여전히 독특한 혼합물이 되기 위해 1개의 DNA 분자를 제외하고 모든 DNA 분자들을 공유할 수 있다. 이전성 분자 식별 바코드의 프로세싱 동안, 혼합물 중 하나 이상의 DNA 분자가 프로세싱되는 데 실패하고 따라서 검출되지 않는 경우, 이전성 분자 식별 바코드는, 프로세싱된 DNA 분자를 공유하는 다른 모든 이전성 분자 식별 바코드들로부터 구별되지 못한다. 이러한 문제점은 DNA 분자 쌍, 또는 심지어 DNA 분자의 트리프렛(triplet) 또는 그 이상을 사용함으로써 모면하게 된다.

[0026] - 이전성 분자 식별 바코드는 가능하게는 다른 표적 핵산과 함께, 병행(병행) 방법에 의해 프로세싱 및 특징화되고, 따라서 DNA 그룹은 프로세싱 동안 결코 완전히 분리되거나 분할되지 않는다. 그렇지 않다면, 독립적인 이전(transfer) 단계가 프로세싱 동안 나란히 개시되며, 이는 바뀜에 취약하고, 따라서 나란히 개시된 각각의 단계들의 결과가 다시 하나의 결과로 조합되는 경우 이전성 분자 식별 바코드 및/또는 시료의 잘못된 특징화를 초래한다. 시퀀싱이 사용되는 경우, 병행 시퀀싱 방법이 사용되어야 한다.

[0027] - 각각의 이전성 분자 식별 바코드 내의 모든 DNA 분자들은 서열이 충분히 상이해야 한다. 사실상, 고도로 병행한 시퀀싱 방법이 사용되는 경우, 시퀀싱 오차는 단일 판독 수준에서 1% 내지 15% 다양하다. 혼합물이 오로지 1개의 뉴클레오타이드에 의해서만 상이한 2개 이상의 DNA 분자들을 함유하는 경우, 증폭 및/또는 시퀀싱 오차는 또 다른 이전성 분자 식별 바코드에 대해 이전성 분자 식별 바코드를 잘못 유형화할 것이다.

[0028] 본 발명의 일 양태는 생물학적 시료를 포함하는 복수의 핵산의 아이덴터티(identity)의 식별 방법에 관한 것으로서, 상기 방법은:

[0029] - 복수의 담체들을 제공하는 단계로서, 상기 담체는 각각 생물학적 시료를 포함하는 핵산을 함유하며, 여기서, 각각의 담체는 상기 담체를 표지하기 위해 적어도 2개의 핵산을 함유하고, 적어도 각각의 핵산은 길이가 적어도 4개 뉴클레오타이드인 상이한 뉴클레오타이드 바코드 서열을 포함하며, 이들 상이한 뉴클레오타이드 바코드 서열들의 조합은 이전성 분자 식별 바코드를 생성하고, 이로써 각각의 이전성 분자 식별 바코드는 수합물에서 각각의 담체에 대해 상이하며, 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에 하나 이상의 뉴클레오타이드 바코드 서열 식별자(identifier) 서열이 존재하며, 이는 상기 뉴클레오타이드 바코드 서열의 식별을 허용하고,

[0030] 각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 바코드 표지를 함유하는, 단계,

[0031] - 시료의 핵산 서열 내의 하나 이상의 표적 서열, 및 뉴클레오타이드 바코드 서열을 포함하는 핵산의 시퀀싱 파트를 시퀀싱하는 단계,

[0032] - 핵산 내의 시퀀싱된 뉴클레오타이드 바코드 서열로부터 각각의 담체의 이전성 분자 식별 바코드를 확인하는 단계로서, 상기 단계는 뉴클레오타이드 바코드 서열을 함유하는 서열을 서열 데이터 내에서 선별하고, 선별된 서열 데이터 내에서 뉴클레오타이드 바코드 서열을 확인하는 단계를 포함하며, 여기서, 선별 단계는 한정된(defined) 거리에서 뉴클레오타이드 바코드 서열의 측면에 존재하는 불변 서열을 기반으로, 뉴클레오타이드 바코드 서열을 포함하는 예정된 길이의 서열의 존재의 식별을 포함하는, 단계,

[0033] - 확인된 이전성 분자 식별 바코드를 담체에 제공된 바코드 표지와 비교함으로써, 시료의 아이덴터티를 식별하는 단계

[0034] 를 포함한다.

[0035] 이의 실시형태에서, 핵산을 포함하는 시료는 순환 DNA, 예컨대 태아 또는 종양 DNA를 포함한다.

[0036] 본 방법의 실시형태는 어댑터(adaptor)를 시료 내의 표적 서열 또는 이의 단편, 및 뉴클레오타이드 바코드 서열을 포함하는 핵산에 연결하는 단계, 및 연결된 어댑터를 시퀀싱 주형으로서 사용하여 표적 서열 및 뉴클레오타이드 바코드 서열을 시퀀싱하는 단계를 포함한다.

[0037] 본 방법의 실시형태는 상기 핵산 시료 내 표적 서열의 농화(enrichment) 단계를 수행하는 단계 및 바코드 서열의 농화 단계를 수행하는 단계를 포함한다.

[0038] 본 발명의 방법은 멀티플렉스 검정법을 수반한다.

[0039] 본 방법의 실시형태는 시료의 표적 서열에 시료-특이적 태그를 부착하고, 선택적으로 또한 동일한 시료-특이적 태그를 첨가된 상기 시료 내의 뉴클레오타이드 바코드 서열을 포함하는 핵산에 부착하는 단계를 포함한다.

[0040] 본 방법의 실시형태에서, 농화된 뉴클레오타이드 바코드 서열 및 농화된 표적 서열의 시퀀싱은 병행 시퀀싱 방법에 의해 수행된다.

[0041] 본 방법의 실시형태에서, 병행 시퀀싱 방법 전에, 농화된 뉴클레오타이드 바코드 서열 및 농화된 표적 서열을 상이한 시료들로부터 풀링(pooling)하는 단계가 수행된다.

[0042] 본 방법의 실시형태에서, 추가의 상이한 이전성 분자 식별 바코드를 한정하는 추가의 상이한 세트의 뉴클레오타이드 바코드 서열이 이러한 방법의 이후의 단계에서 폴리뉴클레오타이드 시료에 첨가된다.

[0043] 본 방법의 실시형태에서, 표적 서열의 농화를 위한 올리고뉴클레오타이드는 바코드의 증폭을 위한 농화를 위한 이전성 분자 식별 바코드 올리고뉴클레오타이드보다 과량으로 존재한다.

[0044] 또는 본 방법의 실시형태에서, 바코드의 증폭을 위한 농화를 위한 이전성 분자 식별 바코드 올리고뉴클레오타이드는 표적 서열의 농화를 위한 올리고뉴클레오타이드보다 과량으로 존재한다.

[0045] 본 방법의 실시형태에서, 뉴클레오타이드 바코드 서열을 포함하는 핵산은 시퀀싱되는 표적 DNA와 유사한 길이를 가진다.

[0046] 본 방법의 실시형태에서, 담체 내에서 뉴클레오타이드 바코드 서열은 이들 바코드 서열의 시퀀싱 전에는 담체의 사용자에게 미공지되어 있으며, 여기서, 확인된 이전성 분자 식별 바코드를 제공된 바코드 표지와 비교하는 단계는, 이전성 분자 식별 바코드와 바코드 표지 사이의 관계를 함유하는 데이터베이스를 참고함으로써 수행된다.

[0047] 본 발명의 또 다른 양태는 아이템에서 표지용 핵산을 포함하는 담체의 수합물에 관한 것으로서, 여기서, 각각의 담체는 적어도 2개의 표지용 핵산을 함유하고, 각각의 핵산은 길이가 적어도 4개 뉴클레오타이드인 상이한 뉴클레오타이드 바코드 서열을 포함하며, 이들 상이한 뉴클레오타이드 바코드 서열들의 조합은 이전성 분자 식별 바코드를 생성하고, 이로써 각각의 이전성 분자 식별 바코드는 수합물에서 각각의 담체에 대해 상이하며, 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열이 존재하며, 이는 상기 뉴클레오타이드 바코드 서열의 식별을 허용하는 것을 특징으로 하고,

[0048] 각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 바코드 표지를 함유한다.

[0049] 소정의 실시형태에서, 담체의 수합물은 상기 담체 상 또는 담체 내에서 생물학적 시료, 예컨대 DNA를 함유하는 시료에 적용하기에 적합하다.

[0050] 본 발명의 실시형태에서, 담체는 안정화제, 보존제, 세제, 중화제, 뉴클레아제 저해제, 환원제 또는 ?칭제(quenching agent) 중 하나 이상을 함유한다.

[0051] 소정의 실시형태에서, 각각의 담체는 상기 핵산을 적어도 3개 함유한다.

[0052] 소정의 실시형태에서, 뉴클레오타이드 바코드 서열은 이러한 뉴클레오타이드 바코드 서열을 포착 또는 증폭에 의해 농화시키기 위해 올리고뉴클레오타이드 서열의 측면에 존재한다.

[0053] 소정의 실시형태에서, 뉴클레오타이드 바코드 서열을 증폭시키기 위한 올리고뉴클레오타이드 서열은 1-단계 PCR, 예컨대 프라이머 연장 및 후속해서 연결, 또는 2-단계 PCR, 예컨대 프라이머 연장 및 후속해서 PCR, 서큘러라이제이션(circularisation)-기반 증폭 및 나노포어 시퀀싱으로 이루어진 군으로부터 선택되는 방법을 위한

것이다.

[0054] 소정의 실시형태에서, 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면에 하나 이상의 올리고뉴클레오타이드 결합 서열이 존재하여, 뉴클레오타이드 바코드 서열에서 하나 또는 둘 다의 올리고뉴클레오타이드 결합 서열의 혼성화-기반 서열 포착을 가능하게 한다.

[0055] 소정의 실시형태에서, 뉴클레오타이드 바코드 서열의 측면에는, 상기 바코드의 증폭 및 시퀀싱을 허용하는 PCR 프라이머에 대한 프라이머 결합 서열이 존재한다.

[0056] 소정의 실시형태에서, 뉴클레오타이드 바코드 서열의 측면에는, 상기 바코드를 시퀀싱하기 위한 프라이머 결합 서열이 존재한다.

[0057] 소정의 실시형태에서, 바코드 서열을 포함하는 핵산은 예를 들어, 상기 벡터의 단편화 또는 분해에 의해 수득된 클로닝 벡터의 단편에 포함된다.

[0058] 전형적인 실시형태에서, 담체는 생물학적 시료를 수용하기 위한 용기이다.

[0059] 전형적인 실시형태에서, 담체는 생물학적 시료를 적용하고/거나 고정하기 위한 기판(substrate)이다.

[0060] 상기 기재된 바와 같은 담체의 수합물은 100 내지 1 x 10^6개 담체, 10 x 10^6개 담체, 100 x 10^6개 담체, 100 x 10^6개 초과의 담체를 포함한다.

[0061] 또 다른 양태는 이전성 분자 식별 바코드를 포함하는 담체의 수합물의 제조 방법에 관한 것이며, 상기 방법은:

[0062] a) 길이가 적어도 4개 뉴클레오타이드인 뉴클레오타이드 바코드 서열, 및 핵산 내에서 뉴클레오타이드 바코드 서열의 존재의 식별을 허용하는 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열을 포함하는 상이한 핵산들의 제1 수합물을 제공하는 단계로서, 상기 바코드 서열은 상기 수합물 내의 핵산들 사이에서 상이한, 단계,

[0063] b) 단계 a)의 적어도 2개의 핵산들의 조합을 각각의 담체에 첨가하여, 각각의 담체가 뉴클레오타이드 바코드 서열에서의 차이에 의해 한정되는 독특한 이전성 분자 식별 바코드를 갖는 담체의 수합물을 수득하는 단계,

[0064] c) 각각의 담체를, 상이한 뉴클레오타이드 바코드 서열에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 표지로 표지하는 단계,

[0065] d) 표지와, 이전성 분자 식별 바코드에서 뉴클레오타이드 바코드 서열의 서열 사이의 관계를 저장하는 단계

[0066] 를 포함한다.

[0067] 이들 방법의 실시형태에서, 단계 c)에서의 표지 및 단계 d)에서의 저장은, 상이한 뉴클레오타이드 바코드 서열들이 확인될 때까지 담체의 후속적인 사용자가 표지와 이전성 분자 식별 바코드 사이의 관계를 추론할 수 없도록 수행된다.

[0068] 이들 방법의 실시형태에서, 핵산들의 제2 수합물은 단계 b) 전에, 상이한 뉴클레오타이드 바코드 서열들을 갖는 핵산들의 멀티플(multiple)을 한정함으로써 제조되며, 여기서, 단계 b)에서 적어도 3개 멀티플의 핵산들의 독특한 조합이 각각의 담체에 첨가된다.

[0069] 이들 방법의 실시형태에서, 제2 수합물은 멀티플의 핵산을 함께 첨가함으로써 제조된다.

[0070] 이들 방법의 실시형태에서, 멀티플은 2개 핵산들의 쌍이다.

[0071] 본 발명의 또 다른 양태는 이전성 분자 식별 바코드를 포함하는 벡터를 갖는 숙주들의 수합물의 제조 방법에 관한 것이며, 상기 방법은:

[0072] a) 숙주에서 핵산 벡터들의 제1 수합물을 제공하는 단계로서, 여기서, 이러한 벡터는, 수합물 내에서 핵산 벡터들 사이에서 상이한 길이가 적어도 4개 뉴클레오타이드인 뉴클레오타이드 바코드 서열들을 포함하고, 상기 바코드의 식별을 허용하는 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열을 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면에 포함하는, 단계,

[0073] b) 숙주의 개별 콜로니를 제공하고, 복수의 콜로니들에 대한 핵산 벡터 내에서 바코드를 시퀀싱하여, 단리된 콜로니들의 제2 수합물을 수득하는 단계로서, 여기서, 각각의 콜로니는 상이한 뉴클레오타이드 바코드 서열을 갖는 핵산 벡터를 포함하는, 단계

[0074] 를 포함한다.

[0075] 본 발명의 또 다른 양태는 이전성 분자 식별 바코드를 포함하는 담체의 수합물의 제조 방법에 관한 것이며, 상기 방법은:

[0076] a) 이전의 청구항들에서 한정된 바와 같은 숙주들의 수합물을 제공하는 단계,

[0077] b) 콜로니의 선별을 위해, 이러한 콜로니로부터 벡터를 단리하는 단계,

[0078] c) 담체의 수합물 내의 각각의 담체에, 단계 b)의 상이한 뉴클레오타이드 바코드 서열을 갖는 적어도 2개의 핵산 벡터들의 조합을 첨가하여, 각각의 담체가 담체들 사이에서 뉴클레오타이드 바코드 서열의 차이에 의해 한정되는 독특한 이전성 분자 식별 바코드를 갖는 담체의 수합물을 수득하는 단계,

[0079] d) 각각의 담체를, 상이한 뉴클레오타이드 바코드 서열에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 표지로 표지하는 단계,

[0080] e) 표지와, 이전성 분자 식별 바코드에서 뉴클레오타이드 바코드 서열의 서열 사이의 관계를 저장하는 단계

[0081] 를 포함한다.

[0082] 이들 방법의 실시형태에서, 단계 c) 후, 벡터를 제한 효소에 의해 단편화한다.

[0083] 이들 방법의 실시형태에서, 단계 d)에서의 표지 및 단계 e)에서의 저장은, 상이한 뉴클레오타이드 바코드 서열들이 확인될 때까지 담체의 후속적인 사용자가 표지와 이전성 분자 식별 바코드 사이의 관계를 추론할 수 없도록 수행된다.

[0084] 이들 방법의 실시형태에서, 단계 c) 전에, 벡터들의 추가 수합물은 단리된 핵산 벡터들의 멀티플을 한정함으로써 단계 b)의 수합으로부터 제조되며, 여기서, 멀티플에서 각각의 벡터는 상이한 뉴클레오타이드 바코드 서열을 가진다.

[0085] 나아가, 본 발명은 하기 항목을 포함한다:

[0086] 1. 생물학적 시료를 포함하는 복수의 핵산의 아이덴터티(identity)의 식별 방법으로서, 상기 방법은:

[0087] - 복수의 담체들을 제공하는 단계로서, 상기 담체는 각각 생물학적 시료를 포함하는 핵산을 함유하는 기판 또는 용기이며,

[0088] 여기서, 각각의 담체는 시료를 포함하는 핵산 외에도, 상기 담체를 표지하기 위해 적어도 2개의 뉴클레오타이드 바코드 핵산을 포함하고, 이들 적어도 2개의 뉴클레오타이드 바코드 핵산은 각각 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 이전성 분자 식별 바코드를 생성하고, 이로써 각각의 이전성 분자 식별 바코드는 각각의 담체에 대해 상이하며, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고, 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하여, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하고,

[0089] 각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는, 단계,

[0090] - 생물학적 시료의 핵산 내의 하나 이상의 표적 서열을 시퀀싱하고, 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산 내의 표적 서열을 시퀀싱하는 단계로서, 여기서, 생물학적 시료의 핵산 내에서 표적 서열의 시퀀싱 및 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 시퀀싱은 병행 시퀀싱 방법에 의해 수행되고, 병행 시퀀싱 방법 전에 선택적으로, 생물학적 시료의 핵산 내의 표적 서열 및 상이한 시료 유래의 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열을 풀링하는 단계가 수행되는, 단계,

[0091] - 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 확인하고 선별하는 단계로서, 상기 단계는 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 선별하는 단계를 포함하고, 여기서, 선별 단계는 최소 뉴클레오타이드 바코드 서열에 인접한 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열

을 갖는 서열의 식별을 포함하는, 단계;

[0092] - 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 단리된 서열 내에서 최소 뉴클레오타이드 바코드 서열을 확인하고 선별하는 단계로서, 상기 단계는 한정된 길이에서 2개의 추출 서열들 사이에 존재하는 서열을 선별하거나 한정된 길이에서 1개의 추출 서열에 인접하여 존재하는 서열을 선별하고, 이들 선별된 서열을 최소 뉴클레오타이드 바코드 서열로서 확인하는 단계를 포함하는, 단계,

[0093] - 확인된 최소 뉴클레오타이드 바코드 서열을, 담체에 제공된 거시적 바코드 표지를 기반으로 예상된 최소 뉴클레오타이드 바코드 서열과 비교함으로써, 시료의 아이덴터티 및/또는 오염을 식별하는 단계

[0094] 를 포함하는, 방법.

[0095] 2. 항목 1에 있어서, 어댑터를 시료 내의 표적 핵산, 및 뉴클레오타이드 바코드 서열을 포함하는 핵산에 연결하는 단계, 및 연결된 생성물을 시퀀싱 주형으로서 사용하여 표적 서열 및 뉴클레오타이드 바코드 서열을 시퀀싱하는 단계를 포함하는, 방법.

[0096] 3. 항목 1 또는 2에 있어서, 상기 핵산 시료 내 표적 서열의 농화 단계를 수행하는 단계 및 뉴클레오타이드 바코드 서열의 농화 단계를 수행하는 단계를 포함하는, 방법.

[0097] 4. 항목 1 내지 3 중 어느 한 항목에 있어서, 시료의 표적 서열에 시료-특이적 풀링 바코드를 부착하고, 선택적으로 또한 동일한 시료-특이적 풀링 바코드를 상기 시료 내의 뉴클레오타이드 바코드 서열을 포함하는 핵산에 부착하는 단계를 포함하는, 방법.

[0098] 5. 항목 1 내지 4 중 어느 한 항목에 있어서, 뉴클레오타이드 바코드 서열이, 시퀀싱되는 표적 핵산 또는 농화된 표적 뉴클레오타이드 서열과 유사한 길이를 가진, 방법.

[0099] 6. 항목 1 내지 5 중 어느 한 항목에 있어서, 상기 방법이 시료를 수합하는 단계 및 상기 시료로부터 핵산을 단리하는 단계를 포함하고, 표지용 적어도 2개 뉴클레오타이드 바코드 핵산들의 제1 세트가 수합된 시료에 첨가되고, 표지용 적어도 2개 뉴클레오타이드 바코드 핵산들의 제2 세트가 단리된 핵산에 첨가되는, 방법.

[0100] 7. 아이템을 표지하기 위한 핵산을 포함하는 담체의 수합물로서, 상기 담체는 기판 또는 용기이고, 여기서, 각각의 담체가 시료 핵산 이외에, 표지를 위한 적어도 2개의 뉴클레오타이드 바코드 핵산들을 포함하며, 각각의 뉴클레오타이드 바코드 핵산이 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 핵산이 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 가지고, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합이 이전성 분자 식별 바코드를 생성하며, 이로써 각각의 이러한 이전성 분자 식별 바코드가 수합물 내의 각각의 담체에 대해 상이한 것이고,

[0101] 상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고/거나 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하여, 상기 뉴클레오타이드 바코드 서열의 식별을 허용하고, 각각의 담체가 상기 담체에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는 것을 특징으로 하는, 수합물.

[0102] 8. 항목 7에 있어서, 상기 표지용 뉴클레오타이드 바코드 핵산이, 핵산을 RNA로 전사시키기 위한 서열을 포함하지 않는, 수합물.

[0103] 9. 항목 7 또는 8에 있어서, 담체가 상기 담체 상의 또는 담체 내의 생물학적 시료의 적용에 적합한, 수합물.

[0104] 10. 항목 7 내지 9 중 어느 한 항목에 있어서, 상기 시료가 DNA를 함유하는 시료인, 수합물.

[0105] 11. 항목 7 내지 9 중 어느 한 항목에 있어서, 상기 시료가 RNA를 함유하는 시료인, 수합물.

[0106] 12. 항목 7 내지 11 중 어느 한 항목에 있어서, 최소 뉴클레오타이드 바코드 서열의 측면에, 증폭에 의해 최소 뉴클레오타이드 바코드 서열을 농화시키기 위한 올리고뉴클레오타이드 서열이 존재하거나,

[0107] 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 뉴클레오타이드 바코드 서열 내의 1개 또는 2개의 올리고뉴클레오타이드 결합 서열에서 혼성화-기반 서열 포착을 허용하는 하나 이상의 올리고뉴클레오타이드 결합 서열이 존재하는, 수합물.

[0108] 13. 항목 12에 있어서, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산 내의 최소 뉴클레오타이드 바코드 서열

의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고/거나 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하는, 수합물.

[0109] 14. 항목 7 내지 13 중 어느 한 항목에 있어서, 증폭이 1-단계 PCR, 2-단계 PCR, 프라이머 연장 및 후속해서 연결 및 PCR, 서큘러라이제이션-기반 증폭 및 나노포어 시퀀싱으로 이루어진 군으로부터 선택되는, 수합물.

[0110] 15. 항목 7 내지 14 중 어느 한 항목에 있어서, 최소 뉴클레오타이드 바코드 서열의 측면에, 상기 최소 뉴클레오타이드 바코드 서열의 증폭 및 시퀀싱을 허용하고 선택적으로 뉴클레오타이드 바코드 서열 식별자 및 추출 서열의 증폭 및 시퀀싱을 허용하는 PCR 프라이머에 대한 프라이머 결합 서열이 존재하는, 수합물.

[0111] 16. 항목 7 내지 15 중 어느 한 항목에 있어서, 뉴클레오타이드 바코드 핵산이 클로닝 벡터의 단편에 포함되는, 수합물.

[0112] 17. 항목 7 내지 16 중 어느 한 항목에 있어서, 뉴클레오타이드 바코드 핵산이 상기 벡터의 단편화 또는 분해에 의해 수득되는 단편인, 수합물.

[0113] 18. 항목 16 또는 17에 있어서, 상기 벡터 또는 벡터 단편이 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하는, 수합물.

[0114] 19. 항목 16 또는 17에 있어서, 상기 벡터 또는 벡터 단편이 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 수합물.

[0115] 20. 항목 7 내지 19 중 어느 한 항목에 있어서, 상기 뉴클레오타이드 바코드 서열을 포함하는 핵산을 포착하기 위한 하나 이상의 프라이머를 추가로 포함하는, 수합물.

[0116] 21. 항목 7 내지 20 중 어느 한 항목에 있어서, 상기 뉴클레오타이드 바코드 서열을 포함하는 핵산을 증폭시키기 위한 하나 이상의 프라이머를 추가로 포함하는, 수합물.

[0117] 22. 이전성 분자 식별 바코드를 포함하는, 기판 또는 용기인 담체의 수합물의 제조 방법으로서, 상기 방법은:

[0118] a) 적어도 4개 뉴클레오타이드인 최소 뉴클레오타이드 바코드 서열, 및 뉴클레오타이드 바코드 핵산 내에서 최소 뉴클레오타이드 바코드 서열의 식별을 허용하는 하나 이상의 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열 및/또는 하나 이상의 비-바이러스성 및 비-박테리아성 추출 서열을 포함하는 상이한 뉴클레오타이드 바코드 핵산들의 제1 수합물을 제공하는 단계로서, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 핵산 서열들은 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 가지며, 이 중에서 최소 뉴클레오타이드 바코드 서열은 상기 수합물 내의 뉴클레오타이드 바코드 핵산들 사이에서 상이한, 단계,

[0119] b) 단계 a)의 적어도 2개의 뉴클레오타이드 바코드 핵산들의 조합을 각각의 담체에 첨가하여, 각각의 담체가 최소 뉴클레오타이드 바코드 서열들의 혼합물에서의 차이에 의해 한정되는 독특한 이전성 분자 식별 바코드를 갖는 담체들의 수합물을 수득하는 단계,

[0120] c) 각각의 담체를, 최소 뉴클레오타이드 바코드 서열들의 상이한 혼합물에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지로 표지하는 단계,

[0121] d) 거시적 표지와, 이전성 분자 식별 바코드에서 최소 뉴클레오타이드 바코드 서열들의 혼합물 사이의 관계를 저장하는 단계

[0122] 를 포함하는, 방법.

[0123] 23. 항목 22에 있어서, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고/거나 한쪽 면 또는 양쪽 면의 측면에 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하는, 방법.

[0124] 24. 항목 22 또는 23에 있어서, 상이한 핵산들의 상기 제1 수합물이 핵산을 RNA로 전사시키기 위한 서열을 포함하지 않는, 방법.

[0125] 25. 항목 22 내지 24 중 어느 한 항목에 있어서, 뉴클레오타이드 바코드 핵산이 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하는, 방법.

[0126] 26. 항목 22 내지 25 중 어느 한 항목에 있어서, 상기 뉴클레오타이드 바코드 핵산이 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 방법.

[0127] 27. 항목 22 내지 26 중 어느 한 항목에 있어서, 단계 c)에서의 표지 및 단계 d)에서의 저장이, 상이한 뉴클레오타이드 바코드 서열들이 확인될 때까지 담체의 후속적인 사용자가 거시적 바코드 표지와 이전성 분자 식별 바코드 사이의 관계를 추론할 수 없도록 수행되는, 방법.

[0128] 28. 뉴클레오타이드 바코드 서열을 포함하는 벡터를 갖는 숙주들의 수합물의 제조 방법으로서, 상기 방법은:

[0129] a) 숙주에서 핵산 벡터들의 제1 수합물을 제공하는 단계로서, 여기서, 벡터는, 수합물 내에서 뉴클레오타이드 바코드 핵산 벡터들 사이에서 상이한 길이가 적어도 4개 뉴클레오타이드인 최소 뉴클레오타이드 바코드 서열을 갖는 뉴클레오타이드 바코드 서열을 포함하고, 적어도 2개의 상기 상이한 뉴클레오타이드 바코드 서열들이 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖고, 상기 최소 뉴클레오타이드 바코드 서열의 식별을 허용하는 하나 이상의 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열 및/또는 하나 이상의 비-바이러스성 및 비-박테리아성 추출 서열을 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면에 포함하는, 단계,

[0130] b) 숙주의 개별 콜로니를 제공하고, 복수의 콜로니들에 대한 핵산 벡터 내에서 뉴클레오타이드 바코드 서열을 시퀀싱하여, 단리된 콜로니들의 제2 수합물을 수득하는 단계로서, 여기서, 각각의 콜로니는 상이한 뉴클레오타이드 바코드 서열을 갖는 핵산 벡터를 포함하는, 단계

[0131] 를 포함하는, 방법.

[0132] 29 항목 28에 있어서, 상기 핵산 벡터가 핵산을 RNA로 전사시키기 위한 서열을 포함하지 않는, 방법.

[0133] 30. 항목 28 또는 29에 있어서, 상기 벡터가 서열 번호: 1 내지 서열 번호: 20으로 이루어진 군으로부터 선택되는 서열을 포함하는, 방법.

[0134] 31. 항목 28 또는 29에 있어서, 상기 벡터가 서열 번호: 1 및 서열 번호: 11의 서열을 포함하는, 방법.

[0135] 32. 이전성 분자 식별 바코드를 포함하는, 기판 또는 용기인 담체의 수합물의 제조 방법으로서, 상기 방법은:

[0136] a) 항목 28에서 한정된 바와 같은 숙주들의 수합물을 제공하는 단계,

[0137] b) 콜로니의 선별을 위해, 상기 콜로니로부터 벡터를 단리하는 단계,

[0138] c) 담체의 수합물 내의 각각의 담체에, 상이한 뉴클레오타이드 바코드 서열을 갖는 적어도 2개의 핵산 벡터들의 조합을 첨가하여, 각각의 담체가 담체들 사이에서 최소 뉴클레오타이드 바코드 서열들의 혼합물에서의 차이에 의해 한정되는 독특한 이전성 분자 식별 바코드를 갖는 담체의 수합물을 수득하고, 선택적으로 벡터를 제한 효소에 의해 단편화하는 단계로서, 여기서, 적어도 2개의 상기 핵산 벡터들은 단계 b)의 동일한 길이의 최소 뉴클레오타이드 바코드 서열을 갖는, 단계,

[0139] d) 각각의 담체를, 상이한 최소 뉴클레오타이드 바코드 서열들의 혼합물에 의해 한정되는 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지로 표지하는 단계,

[0140] e) 거시적 바코드 표지와, 이전성 분자 식별 바코드에서 최소 뉴클레오타이드 바코드 서열의 서열 사이의 관계를 저장하는 단계

[0141] 를 포함하는, 방법.

[0142] 33. 담체 세트에서 뉴클레오타이드 바코드 핵산의 추적 방법으로서, 상기 방법은:

[0143] - 게놈 DNA 또는 RNA가 없으며, 상기 담체의 표지를 위한 적어도 2개의 뉴클레오타이드 바코드 핵산을 포함하는 복수의 담체들을 제공하는 단계로서, 여기서, 이들 적어도 2개의 뉴클레오타이드 바코드 핵산들은 각각 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 이전성 분자 식별 바코드를 생성하며, 이로써 각각의 이전성 분자 식별 바코드는 각각의 담체에 대해 상이하고, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 뉴클레오타이드 바코드 서열 식별자 서열이 존재하고, 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하여, 상기 최소

뉴클레오타이드 바코드 서열의 식별을 허용하고,

[0144] 각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는 단계,

[0145] - 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산 내에서 표적 서열을 병행 시퀀싱 방법에 의해 시퀀싱하는 단계로서, 여기서, 병행 시퀀싱 방법 전에 선택적으로, 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 풀링이 수행되는, 단계,

[0146] - 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 확인 및 선별하는 단계로서, 수득된 서열 데이터로부터 뉴클레오타이드 바코드 핵산 유래의 서열을 선별하는 단계를 포함하며, 여기서 선별 단계는 최소 뉴클레오타이드 바코드 서열에 인접한 하나 이상의 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 서열의 식별을 포함하는, 단계;

[0147] - 뉴클레오타이드 바코드 서열 식별자 서열을 갖는 단리된 서열 내에서 최소 뉴클레오타이드 바코드 서열을 확인 및 선별하는 단계로서, 한정된 길이에서 2개의 추출 서열들 사이에 존재하는 서열을 선별하거나 한정된 길이에서 1개의 추출 서열에 인접하여 존재하는 서열을 선별하고, 이들 선별된 서열을 최소 뉴클레오타이드 바코드 서열로서 확인하는 단계를 포함하는, 단계,

[0148] - 담체에 제공된 거시적 바코드 표지를 기반으로, 확인된 최소 뉴클레오타이드 바코드 서열을 예상된 최소 뉴클레오타이드 바코드 서열과 비교하는 단계

[0149] 를 포함하는, 방법.

[0150] 34. 생물학적 시료를 포함하는 복수의 핵산들의 아이덴터티의 식별 방법으로서, 상기 방법은:

[0151] - 기판 또는 용기인 복수의 담체들을 제공하는 단계로서, 상기 담체는 각각 생물학적 시료를 포함하는 핵산을 함유하며,

[0152] 여기서, 각각의 담체는 시료를 포함하는 핵산 외에도, 상기 담체를 표지하기 위해 적어도 2개의 뉴클레오타이드 바코드 핵산을 함유하고, 이들 적어도 2개의 뉴클레오타이드 바코드 핵산들은 각각 길이가 적어도 4개 뉴클레오타이드인 상이한 최소 뉴클레오타이드 바코드 서열을 포함하며, 이들 상이한 뉴클레오타이드 바코드 핵산들의 조합은 이전성 분자 식별 바코드를 생성하고, 이로써 각각의 이전성 분자 식별 바코드는 각각의 담체에 대해 상이하며, 상기 적어도 2개의 뉴클레오타이드 바코드 핵산들 내의 최소 뉴클레오타이드 바코드 서열의 한쪽 면 또는 양쪽 면의 측면에, 모든 뉴클레오타이드 바코드 핵산들 내에서 동일한 비-바이러스성 및 비-박테리아성 추출 서열이 존재하는 것을 특징으로 하고,

[0153] 각각의 담체는 상기 담체 상에 적용된 이전성 분자 식별 바코드에 상응하는 거시적 바코드 표지를 함유하는, 단계,

[0154] - 생물학적 시료의 핵산 내의 하나 이상의 표적 서열을 시퀀싱하고, 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열을 시퀀싱하는 단계로서, 여기서, 생물학적 시료의 핵산 내의 표적 서열의 시퀀싱 및 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 시퀀싱은 병행 시퀀싱 방법에 의해 수행되며, 여기서, 병행 시퀀싱 방법 전에 선택적으로, 생물학적 시료의 핵산 내의 농화된 표적 서열, 및 상이한 시료 유래의 최소 뉴클레오타이드 바코드 서열을 포함하는 뉴클레오타이드 바코드 핵산의 표적 서열의 풀링이 수행되는, 단계,

[0155] - 수득된 서열 데이터로부터 최소 뉴클레오타이드 바코드 서열을 확인 및 선별하는 단계로서, 상기 단계는 한정된 길이에서 2개의 추출 서열들 사이에 존재하는 서열을 선별하거나, 한정된 길이에서 1개의 추출 서열에 인접하여 존재하는 서열을 선별하고, 이들 선별된 서열을 최소 뉴클레오타이드 바코드 서열로서 확인하는 단계,

[0156] - 담체에 제공된 거시적 바코드 표지를 기반으로, 확인된 최소 뉴클레오타이드 바코드 서열을 예상된 최소 뉴클레오타이드 바코드 서열과 비교함으로써, 시료의 아이덴터티 및/또는 오염이 없는 시료를 식별하는 단계

[0157] 를 포함하는, 방법.

도면의 간단한 설명

[0158] **도 1.** NGS 시퀀싱에서 특징화되는 단일 가닥 뉴클레오타이드 바코드 올리고뉴클레오타이드의 제조이다.

도 2. 단일 가닥 올리고뉴클레오타이드로부터 이중 가닥 뉴클레오타이드 바코드의 제조이다.

도 3. 단일 가닥 또는 이중 가닥 뉴클레오타이드 바코드 서열의 예들이다.

도 4. 뉴클레오타이드 바코드 플라스미드 및 이의 선형화이다.

도 5. a. 2-단계 PCR 프로토콜에서; b. 프라이머-연장-연결/PCR 프로토콜에서; c. 단편화된 DNA 연결 방법, 및 후속해서 바코드 뉴클레오타이드 서열의 혼성화에 의한 포착에서, 고도 병행 시퀀싱을 위한 선형화된 뉴클레오타이드 바코드 플라스미드로부터 시퀀싱 주형의 제조이다.

도 6. 분할 멀티플렉스 반응을 사용하는 검정법에서 이전성 뉴클레오타이드 바코드의 용도이다. 이러한 예에서, 2개의 분할 멀티플렉스 반응이 수행된다. 각각의 멀티플렉스 PCR 반응에 첨가된 뉴클레오타이드 바코드 서열에 관한 2개의 프라이머들 중 적어도 하나는 최소 바코드 서열에 대한 측면 서열(들)에 상이한 프라이머 결합 부위를 가진다. 각각의 멀티플렉스 반응에서 수득된 뉴클레오타이드 바코드 앰플리콘은 최소 바코드 영역에 대한 1개 또는 2개의 인접 서열에 상이한 길이 및/또는 측면 서열 조성물을 가지며, 따라서 이들의 기원이 확인될 수 있다. 예상되는 바코드가 검정법의 종료 시 관찰되고 올바른 길이 및/또는 측면 서열 조성물을 가진 2개의 앰플리콘들이 모두 관찰되는 경우, 시료들 사이에서 및 분할 반응들 사이에서 시료 바뀜이 발생하지 않았다. 예상되는 바코드가 전혀 관찰되지 않는 경우, 2개의 멀티플렉스들은 모두 또 다른 시료의 2개의 멀티플렉스들과 바뀐다. 예상되는 바코드 외에도, 또 다른 바코드가 관찰되지만, 예상되는 바코드는 오로지 1개의 멀티플렉스의 예상되는 길이 및/또는 측면 서열 조성물을 갖는 앰플리콘에서 관찰되지만, 예상되지 않은 바코드가 오로지 제2 멀티플렉스의 올바른 길이 및/또는 측면 서열 조성물을 가진 앰플리콘에서 관찰되는 경우, 오로지 제2 멀티플렉스만 또 다른 시료의 제2 멀티플렉스와 바뀌었다.

도 7은 조사 하에 표적 영역의 농화를 위해 2-단계 PCR 프로토콜을 사용하는 NGS 검정법에서 이전성 분자 식별 바코드의 용도이다; 7a. 시료 바뀜이 없으며, 오염이 없고, 7b. 시료 바뀜이 있으며, 7c. 시료 오염이 있다.

도 8은 최종 유효 유전자 검사 보고 시까지, 이전성 분자 식별 바코드를 함유하는 혈액 수합기 튜브로부터 출발하는, 유전자 검사에서 가능한 작업 흐름을 보여준다. 조사 하에서 뉴클레오타이드 바코드 서열 유래의 시퀀싱된 서열 및 표적 핵산 유래의 시퀀싱된 서열 둘 다의 생물정보학 프로세싱이 병행하여 수행된다.

도 9는 최종 유효 유전자 검사 보고 시까지, 생물학적 시료에 이전되는 이전성 분자 식별 바코드를 함유하는 마이크로튜브로부터 출발하는, 유전자 검사에서 가능한 작업 흐름을 보여준다.

도 10은 최종 유효 유전자 검사 보고 시까지, 이전성 분자 식별 바코드를 함유하는 혈액 수합기 튜브로부터 출발하는, 유전자 검사에서 가능한 작업 흐름을 보여지만, 여기서 뉴클레오타이드 바코드 서열 유래의 시퀀싱된 서열의 생물정보학 프로세싱이 먼저 수행되고, 조사 하에 표적 핵산 유래의 시퀀싱된 서열의 생물정보학 프로세싱은 오로지 뉴클레오타이드 바코드의 유도된 서열 판독의 결과 성과 분석, 즉 시료 바뀜 및/또는 오염이 존재하는지의 여부에 따라 출발된다.

도 11은 총 프로세스의 서브프로세스들을 품질 관리 하기 위해 생물학적 시료 또는 이의 유도체가 총 검사 프로세스 내의 2개의 상이한 단계에서 상이한 이전성 분자 식별 바코드를 이용하여 스파이킹된(spiked) 가능한 작업 흐름을 보여준다.

발명을 실시하기 위한 구체적인 내용

[0159] 명세서 및 청구항을 포함하여 본 출원에 사용되는 용어는 다르게 명시되지 않는 한 하기에 나타낸 바와 같이 한정된다. 이들 용어는 명확성을 위해 구체적으로 한정되지만, 모든 정의들은 당업자가 이들 용어를 어떻게 이해하는지와 일치한다.

[0160] 단수형("a", "an" 및 "the")은 문맥상 명확하게 다르게 가리키지 않는 한, 복수형을 포함함을 주지해야 한다.

[0161] 변수를 지칭할 때, 용어 "본원에 사용된 바와 같이", "본원에 한정된 것들" 및 "상기 한정된 것들"은 참조에 의해, 변수의 광범위한 정의, 뿐만 아니라 있다면 바람직한, 보다 바람직한 및 가장 바람직한 정의를 포함한다.

[0162] 본원에 사용된 바와 같이 용어 "**핵산**"은 뉴클레오타이드, 예를 들어 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드로 구성된 중합체를 지칭한다. 핵산은 또한, 합성적으로 생성된 화합물을 포함하지만, 이러한 핵산은 변이체 당-포스페이트 백본(폴리아미드(예를 들어 펩타이드 핵산(PNA), 연결된 핵산(LNA), 폴리모르폴리노 중합체), 및/또는 하나 이상의 염기의 변이체를 포함하며, 그러나 이러한 핵산은 2개의 천연 발생 핵산들과 유사하

게 서열 특이적 방식으로 천연 발생 핵산과 여전히 혼성화할 수 있으며, 즉 혼성화 반응, Pi 전자 스태킹(stacking) 및 수소 결합을 통한 협동 작용, 예컨대 왓슨-크릭 염기쌍 형성 상호작용, 동요 상호작용(Wobble interaction) 등에 참여할 수 있다. 이들 핵산은 단일 가닥 또는 이중 가닥, 또는 심지어 트리플렛 DNA 또는 그 이상의 복잡한 구조일 수 있다. 용어 "핵산"은 특정 핵산 또는 이러한 핵산의 상보체(complement)인 뉴클레오타이드 서열을 포함하는 핵산, 특정 핵산과 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 또는 95% 초과의 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 핵산, 또는 특정 핵산의 상보체와 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 또는 95% 초과의 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 핵산일 수 있다.

[0163] 핵산은 천연 발생 핵산, 게놈 기원의 핵산, 미토콘드리아 핵산, cDNA 기원의(mRNA로부터 유래된) 핵산, 박테리아, 바이러스, 진균류 유래의 핵산, 합성 기원의 핵산, 비-천연 발생 핵산, DNA 및/또는 RNA의 유사체, 및/또는 유도체, 및/또는 이들 중 임의의 것의 조합일 수 있다. 핵산 변형은, 추가의 전하, 분극성(polarizability), 수소 결합, 정전기 상호작용 및 작용기를 혼입하는 화학 기를 개별 핵산 염기 또는 핵산 전체에 첨가하는 단계를 포함할 수 있다. 이러한 변형은 예를 들어 안정성을 개선하기 위해 5' 및/또는 3' 말단에 차단기를 포함하는 것, 염기 변형 예컨대 2'-위치 당 변형, 5-위치 피리미딘 변형, 8-위치 퓨린 변형, 시토신 외향고리(exocyclic) 아민에서의 변형, 5-브로모-우라실의 치환, 백본 변형, 비통상적인 염기쌍 형성 조합, 예컨대 이소염기, 이소시티딘 및 이소구아니딘, 메틸화 등을 포함할 수 있다. 다른 유형의 핵산들이 고려된다.

[0164] 핵산(들)은 완전히 화학적 합성 프로세스, 예컨대 고체상-매개 화학적 합성, 생물학적 출처, 예컨대 핵산을 생성하는 임의의 종으로부터의 단리를 통해, 또는 분자생물학 툴(예컨대 DNA 복제, 연결, PCR 증폭, 역전사 또는 이들 프로세스들의 조합)에 의한 핵산의 조작을 포함하는 프로세스, 또는 이들의 조합으로부터 유래될 수 있다.

[0165] 핵산 변형은 또 다른 분자에 의한 단리 및/또는 검출을 직접적으로 또는 간접적으로 용이하게 할 수 있으며, 이러한 또 다른 분자에 다시 다른 분자가 결합될 수 있다. 이러한 변형은 스트렙타비딘과 상호작용할 수 있는 하나 이상의 비오틴 기일 수 있을 것이다. 이러한 목적을 위한 다른 상호작용 분자는 비오틴/아비딘, 비오틴/비오틴-결합-분자(예를 들어 NEUTRAVIDINT 변형된 아비딘(Pierce Chemicals Rockford, IL), 글루타티온 S-트랜스퍼라제(GST)/글루타티온, 항체/항원, 항체/항체-결합-분자, 디옥시게닌/항-디옥시게닌, DNP(2,4-디니트로페닐)/항-DNP 항체, 말토스-결합-단백질/말토스, 킬레이트화(예를 들어 (Co2+, Ni2+)/폴리히스티딘, 플루로닉 커플링 기술일 것이다.

[0166] 본원에 사용된 바와 같이 용어 "**데옥시리보핵산**" 및 "**DNA**"는 데옥시리보뉴클레오타이드로 구성된 중합체를 의미한다.

[0167] 본원에 사용된 바와 같이 용어 "**리보핵산**" 및 "**RNA**"는 리보뉴클레오타이드로 구성된 중합체를 의미한다.

[0168] 본원에 사용된 바와 같이 용어 "**올리고뉴클레오타이드**"는 길이가 상대적으로 짧은, 일반적으로 200개 뉴클레오타이드보다 짧은, 보다 특히 100개 뉴클레오타이드보다 짧은, 가장 특히 50개 뉴클레오타이드보다 짧지만, 일반적으로 5개 뉴클레오타이드보다 긴 핵산을 지칭한다. 전형적으로, 올리고뉴클레오타이드는 단일 가닥 DNA 분자이다. 이러한 올리고뉴클레오타이드는 예컨대 비오틴화되는 변형, 5' 인산화되는 변형을 운반할 수 있다. '올리고뉴클레오타이드'의 동의어는 '올리고'이다.

[0169] 본원에 사용된 바와 같이 용어 "**표적 핵산**"은 특징화될 아이템, 예컨대 생물학적 시료에 존재하는 핵산 또는 (하나의) 이의 일부(들) 또는 핵산(들) 또는 이의 일부들이다. 대부분의 경우, 핵산(들)의 이러한 일부 또는 일부들은 특징화 프로세스의 표적이다.

[0170] 본원에 사용된 바와 같이 용어 "**표적 뉴클레오타이드 서열**"은 표적 핵산의 뉴클레오타이드 서열을 포함하는 분자, 예컨대 표적 핵산을 증폭시킴으로써 수득되는 증폭 생성물, 표적 핵산의 시퀀싱에 의해 수득되는 시퀀싱 생성물, RNA 표적 핵산의 역전사 시 생성되는 cDNA를 지칭한다.

[0171] 본원에 사용된 바와 같이 용어 "**뉴클레오타이드 바코드**"는 바코드 또는 식별 수단으로서 사용되는 특정 서열을 갖는 표적 핵산 또는 이의 일부를 지칭한다. 상이한 뉴클레오타이드 바코드들은 상이한 바코드 서열들을 가지며, 이들 바코드 서열은 상이한 유형의 뉴클레오타이드 바코드로 지칭된다. 뉴클레오타이드 바코드가 DNA로 구성되는 경우, 이러한 바코드는 "**DNA-유형 뉴클레오타이드 바코드**"로 지칭되고, 뉴클레오타이드 바코드가 RNA로 구성되는 경우, 이러한 바코드는 "**RNA-유형 뉴클레오타이드 바코드**"로 지칭된다. 실제 바코드 서열인 본원에 사용된 바와 같이 용어 "**최소 뉴클레오타이드 바코드**"는 주어진 유형의 뉴클레오타이드 바코드에서 동일한 불변 서열의 측면에 존재할 것이다. 이들 측면 불변 서열들은 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않거나(따라서, 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의

의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다.

[0172] 본원에 사용된 바와 같이 용어 **"이전성 분자 식별 바코드"**는 단일 유형의 뉴클레오타이드 바코드, 또는 상이한 유형의 뉴클레오타이드 바코드들의 혼합물을 지칭한다. 상이한 유형의 뉴클레오타이드 바코드들의 혼합물의 이전성 분자 식별 바코드 구축물에서, 이전성 분자 식별 바코드는, 여전히 독특한 이전성 분자 식별 바코드이기 위해, 다른 이전성 분자 식별 바코드와 공통으로 1개를 제외한 모든 뉴클레오타이드 바코드들을 운반할 수 있다. 바코드의 동의어는 지표(index), 태그, MID(분자 식별자)일 수 있다. 용어 '이전성 분자 식별 바코드'는 또한, 주어진 순간에 불용성 상 내에서 존재하지만 프로세싱 동안 (다시) 가용성으로 되는 하나 이상의 뉴클레오타이드 바코드를 지칭한다. 이전성 분자 식별 바코드가 DNA로 구성되는 경우, 이러한 바코드는 **"DNA-유형 이전성 분자 식별 바코드"**로 지칭되고, 이전성 분자 식별 바코드가 RNA로 구성되는 경우, 이러한 바코드는 **"RNA-유형 이전성 분자 식별 바코드"**로 지칭된다.

[0173] 본원에 사용된 바와 같이 용어 **"거시적 바코드 표지"**는 아이템(예를 들어 수령체 홀더)의 표지에 사용되는 인쇄된 바코드 페이퍼 표지(예를 들어 광학 1D(라인) 또는 2D 바코드 페이퍼 표지) 또는 RFID(무선 주파수 식별 바코드) 바코드 표지를 지칭한다. 이러한 표지는 상이한 수단들, 예컨대 수령체 홀더의 벽 외부에서 독특한 페이저 바코드의 부착에 의해 수행될 수 있다. 수령체 홀더가 독특한 이전성 분자 식별 바코드를 함유하는 경우, 독특한 거시적 바코드 표지는 상응하는 이전성 분자 식별 바코드와 명백하게 연관이 있고 연결될 수 있다.

[0174] 본원에 사용된 바와 같이 용어 **"상보적인"**은 2개의 뉴클레오타이드들 사이에서 정확한 쌍을 형성하는 능력을 지칭한다. 핵산의 주어진 위치에서 뉴클레오타이드가 또 다른 핵산의 뉴클레오타이드와 수소 결합을 할 수 있는 경우, 2개의 핵산들은 해당 위치에서 서로 상보적인 것으로 간주된다. 2개의 단일 가닥 핵산 분자들 사이에서 상보성은, 오로지 일부의 뉴클레오타이드만 결합할 수 있는 "부분적"일 수 있거나, 단일 가닥 분자들 사이에 전반적인 상보성이 존재하는 경우 이러한 상보성은 완전할 수 있다. 핵산 가닥들 사이에서 상보성의 정도는 핵산 가닥들 사이에서 혼성화의 효율 및 강도에 유의한 영향을 미친다.

[0175] 본원에 사용된 바와 같이 용어 **"특이적인 혼성화"**는 한정된 엄격한 조건 하에 혼성화 혼합물에 존재하는 다른 핵산 또는 뉴클레오타이드 서열에 실질적으로 결합하는 일 없이 표적 핵산 또는 표적 뉴클레오타이드 서열에 핵산이 결합하는 것을 지칭한다. 당업자는, 혼성화 조건의 엄격성의 완화가, 서열 미스매치가 관용되는 것을 허용함을 인지하고 있다. 혼성화는 엄격한 혼성화 조건 하에 수행된다. 구어 "엄격한 혼성화 조건"은 일반적으로, 한정된 이온 강도 및 pH에서 특이적인 서열에 대해 용융점(T_m)보다 약 5℃ 내지 약 20℃ 또는 25℃ 낮은 온도의 온도 범위를 지칭한다. 본원에 사용된 바와 같이, T_m은 이중 가닥 핵산 분자 집단이 단일 가닥으로 절반으로 해리되는 온도이다. 핵산의 T_m의 계산 방법은 당업계에 잘 공지되어 있다(문헌[A Laboratory Manual, by Sambrook and Russel, 3rd Edition, Cold Spring Harbor Laboratory Press, 2001]). 하이브리드의 용융점(및 따라서 엄격한 혼성화 조건)은 용액 내에 존재하거나 고정된 다양한 인자들, 예컨대 프라이머 또는 프로브의 길이 및 성질(DNA, RNA, 염기 조성물) 및 표적 핵산의 성질(DNA, RNA, 염기 조성물) 등, 뿐만 아니라 염 및 다른 구성성분들의 농도(예를 들어 포름아미드, 덱스트란 설페이트, 폴리에틸렌 글리콜의 존재 또는 부재)에 의해 영향을 받는다.

[0176] 본원에 사용된 바와 같이 용어 **"프라이머"**는 적절한 완충제 내에서 적합한 온도에서 적절한 조건 하에(즉, 4개의 상이한 뉴클레오사이드 트리포스페이트 및 중합 작용제, 예컨대 DNA 또는 RNA 중합효소 또는 역전사효소의 존재 시) 핵산과 혼성화(또한 "어닐링"이라고도 함)하고 뉴클레오타이드(DNA 또는 RNA) 중합을 위한 개시 부위로서 역할을 할 수 있는 올리고뉴클레오타이드를 지칭한다.

[0177] 본원에 사용된 바와 같이 용어 **"프라이머 결합 부위"** 또는 **"프라이머 부위"**는, 프라이머가 혼성화하여 뉴클레오타이드 합성을 준비시키는(prime) 표적 핵산 또는 표적 뉴클레오타이드 서열의 분절을 지칭한다. 이전성 분자 식별 바코드 내의 프라이머 결합 부위는 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않거나(따라서 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다. 프라이머가 결합하는 표적 핵산 또는 표적 뉴클레오타이드 서열의 분절은 본원에서, 올리고뉴클레오타이드 결합 서열로도 지칭된다. 프라이머 결합 부위는 전형적으로 적어도 5개 뉴클레오타이드 길이, 보다 전형적으로 10 내지 30개 뉴클레오타이드 범위, 또는 심지어 그보다 길다. 더 짧은 프라이머 결합 부위는 일반적으로, 프라이머와 주형 사이에서 충

분히 안정한 하이브리드 복합체를 형성하기 위해 더 낮은 온도를 필요로 한다. 프라이머가 주형의 정확한 서열을 반영할 필요는 없으나, 주형과 혼성화하기 위해서는 충분히 상보적이어야 한다. 프라이머 또는 이의 일부가 핵산 내의 뉴클레오타이드 서열에 혼성화하는 경우, 이러한 프라이머는 또 다른 핵산에 어닐링하는 것으로 일컬어진다. 프라이머가 특정 뉴클레오타이드 서열에 혼성화한다는 말이, 상기 프라이머가 해당 뉴클레오타이드 서열에 완전히 또는 독점적으로 혼성화한다는 것을 내포하는 것은 아니다.

[0178] 본원에 사용된 바와 같이 용어 "**프라이머 쌍**"은, 증폭되는 DNA 서열의 5' 말단의 상보체와 혼성화하는 5' "업스트림 프라이머", 및 증폭되는 서열의 3' 말단과 혼성화하는 3' "업스트림 프라이머"를 포함하는 프라이머 세트를 지칭한다. 뉴클레오타이드 바코드 서열에 혼성화하는 프라이머는 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않는 서열을 가지거나(따라서 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다.

[0179] 당업자에 의해 인지되는 바와 같이, 용어 "업스트림" 및 "다운스트림"은 제한하려는 것이 아니라, 그보다는 예시적인 배향을 제공한다. 동의어는 정방향 및 역방향 프라이머, 좌측 및 우측 프라이머, +(플러스) 및 -(마이너스) 프라이머, 5' 및 3' 프라이머이다. 하나의 프라이머가 플러스 DNA 가닥 내에 프라이머 결합 부위를 가지고 제2 프라이머가 표적 핵산의 마이너스 DNA 가닥 내에 프라이머 결합 부위를 갖는 "프라이머 쌍"은 PCR 반응을 준비시킬 수 있다. "프라이머 쌍"은 또한, 2개의 프라이머들이 표적 핵산의 동일한 DNA 가닥(플러스 또는 마이너스 DNA 가닥) 내에 프라이머 결합 부위를 갖는 프라이머 쌍, 예컨대 연결 사슬 반응 검정법 또는 프라이머-연장-연결 검정법에서 사용되는 프라이머 쌍을 지칭할 수 있다.

[0180] 프라이머는, 증폭 후 검출되는 대부분의 앰플리콘들이, 예상된 길이보다 상이한 길이를 갖는 앰플리콘을 생성하는 표적 핵산 내에서의 프라이밍으로 인한 앰플리콘과는 대조적으로, 이들 앰플리콘이 표적 핵산의 각각의 말단에서 예상된 부위에서 프라이밍으로 인한 것이라는 측면에서 "예상된 길이"를 갖도록 선택된다. 다양한 실시형태에서, 프라이머는, 수득된 앰플리콘의 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99%가 예상된 길이를 갖도록 선택된다.

[0181] 본원에 사용된 바와 같이 용어 "**어댑터**"는, 표적 뉴클레오타이드 서열에 첨가되는 하나 이상의 예정된 기능을 갖는 예정된 뉴클레오타이드 서열을 지칭하고, 따라서 심지어 표적 뉴클레오타이드 서열의 일부가 될 수 있다. 어댑터는 표적 핵산 또는 표적 뉴클레오타이드 서열의 한쪽 말단 또는 양쪽 말단에 첨가될 수 있다. 어댑터가 양쪽 말단에 첨가되는 경우, 이들 어댑터는 서열에 대해 동일하거나 상이할 수 있다. 양쪽 부위에서 어댑터들이 측면에 존재하는 표적 뉴클레오타이드 서열은 선형 또는 원형 형태를 가질 수 있다. 첨가되는 어댑터는 하나 이상의 특이적인 유형의 예정된 기능을 가질 수 있다. 따라서, 본원에 사용된 바와 같이 용어 "어댑터"는 하나의 어댑터 또는 다수의 어댑터들을 지칭할 수 있다. 1개 초과의 기능이 포함되는 경우, 기능들은 동일한 유형 또는 상이한 유형(들)일 수 있다. 상이한 유형의 어댑터들은 전체 어댑터 내의 임의의 위치에 혼입될 수 있다. 어댑터의 동의어는 예를 들어, 뉴클레오타이드 어댑터, 유니버셜 어댑터, 태그, 뉴클레오타이드 태그, 유니버셜 태그이다.

[0182] 예정된 어댑터 또는 어댑터 기능의 유형의 예들은 비제한적으로, 프라이머 서열, DNA 합성을 위한 프라이밍 결합 또는 어닐링 부위, 시퀀싱을 위한 프라이밍 결합 또는 어닐링 부위, 올리고뉴클레오타이드에 대한 혼성화 부위, 하나 이상의 제한 효소에 대한 인지 부위, 바코드 서열, 고정 서열, 나노포어 시퀀싱을 위한 모터 단백질에 대한 리더-어댑터(Oxford Nanopore Technologies), 또는 후속적인 프로세싱에 유용한 다른 인지 또는 결합 서열, 상기 기재된 하나 이상의 어댑터를 연결하는 링커 또는 스페이서일 것이다. 나아가 본원에 사용된 바와 같이, 특이적인 어댑터 서열에 대한 참조는 또한, 임의의 이러한 서열에 대한 상보체를 지칭하며, 따라서 상보적인 복제 시 특이적인 기재된 서열이 수득될 것이다.

[0183] 상이한 어댑터들이 전체 어댑터에 존재하는 경우, 이러한 상이한 어댑터들을 구축하는 뉴클레오타이드 서열 단위는 비-중첩 상이한 이웃 서열 단위들 및/또는 중첩 서열 단위로서 위치할 수 있다. 예를 들어, DNA 합성 반응을 위한 프라이머 결합 부위로서 사용될 20개 뉴클레오타이드 길이의 어댑터 기능기, 및 주어진 표적 뉴클레오타이드 서열의 포착/단리에 사용되는 20개 뉴클레오타이드 길이의 어댑터 기능기는 중첩될 수 있고, 예를 들어 10개의 뉴클레오타이드 서열을 공통으로 가져서 2개 어댑터 모두의 조합된 서열이 40개 뉴클레오타이드 대신에 30개 뉴클레오타이드 길이가 된다.

[0184] 본원에 사용된 바와 같이 용어 '**연결 어댑터**'는 완전히 또는 부분적으로 이중 가닥인 DNA 분자를 지칭한다. 일반적으로, 이들 연결 어댑터는 다른 DNA 분자에 연결하는 데 사용된다. 이들은 대체로, 2개의 가능하게는 부분적인, 상보적인 혼성화 올리고뉴클레오타이드들의 혼합물로부터 제조된다. 부분적으로 이중 가닥 연결 어댑터는 헤어핀 어댑터를 가질 수 있다. 헤어핀 어댑터의 기능은, 연결 후 DNA 분자들이 이들의 5' 말단에서 혼성화하지 않는 것을 방지하는 것, 예를 들어 연속(concatenation)을 방지하는 것이다. 또 다른 헤어핀 기능은, 헤어핀-단백질이 결합하여 나노포어 시퀀싱(Oxford Nanopore Technologies)을 용이하게 하는 헤어핀-어댑터이다. 연결 어댑터는 또한, 특허 US7803550에 기재된 바와 같이 역 반복부(inverted repeat) 및 루프를 포함하는 단일 줄기-루프 올리고뉴클레오타이드로부터 생성될 수 있다. 이들 모든 연결 어댑터들은 실제 연결 부위에서 1개 이상의 비-상보적인 뉴클레오타이드를 가질 것이며, 이는 방향성(directional) 연결을 용이하게 하고/거나 허용할 것이다. 예를 들어, 연결 어댑터는, 연결을 용이하게 하기 위해 이중 가닥 표적 뉴클레오타이드 서열의 3' A-오버행과 혼성화할 수 있는 3' T-오버행을 가질 것이다.

[0185] 본원에 사용된 바와 같이 용어 '**풀링 바코드**'는, 상이한 프로세싱된 DNA 시료들의 보다 효율적인(시간이 덜 소요되는) 및/또는 경제학적 풀링을 위해 표적 뉴클레오타이드 서열을 마크하는 바코드를 지칭한다. 대체로 이들 풀링 바코드는 DNA 합성 또는 증폭에 사용되는 프라이머 또는 연결 어댑터에 존재한다. 따라서, 풀링 바코드는 어댑터 기능기를 표적 뉴클레오타이드 서열에 첨가하며, 그런 다음 이는 생성된 표적 뉴클레오타이드 서열에 대한 정보를 인코딩한다. 예를 들어, 상이한 풀링 바코드들(상이한 바코드 서열들을 가짐)은, 상이한 개체들 유래의 다수의 상이한 시료들 각각으로부터 하나 이상의 표적 핵산을 증폭시키는 데 사용될 수 있다. 예를 들어, 상이한 풀링 바코드들(상이한 바코드 서열들을 가짐)은 생물학적 시료 유래의 다수의 상이한 개별 세포들 각각으로부터 하나 이상의 표적 핵산을 증폭시키는 데 사용될 수 있다. 따라서, 풀링 바코드 뉴클레오타이드 서열은 각각 결과적인 표적 뉴클레오타이드 서열의 시료 또는 세포 기원을 가리킨다. 이는, 다운스트림 프로세스에서 상이한 시료들 유래의 상이한 유형의 풀링-바코드 표적 뉴클레오타이드 서열들의 조합을 허용한다. 이는, 일단 풀링이 수행되면, 각각의 시료에 대한 작업 흐름들의 총 수를 모든 풀링된 시료에 대해 하나의 단일 작업 흐름으로 단순화시킨다. 하나의 적용은 고도 병행 시퀀싱에 의한 상이한 표적 뉴클레오타이드 서열들의 시퀀싱일 것이다. 고도 병행 시퀀싱 방법 및 장비의 시퀀싱 결과물은 방대하고, 많은 경우 단일 시료의 경우 너무 많다. 고도 병행 시퀀싱 장비의 전용량(full capacity)은, 상이한 시료들/개체들 유래의 상이한 풀링-바코드 표적 뉴클레오타이드 서열들을 조합함으로써 가장 경제적인 방식으로 사용될 수 있다. 시퀀싱 후, 표적 뉴클레오타이드 서열로부터 수득된 상이한 서열들은 존재하는 풀링-바코드 서열에 따라 그룹으로 나뉘고, 따라서 원래 시료에 배정된 다음, 다운스트림 작업 흐름에서 다시 추가로 개별적으로 프로세싱되고 분석될 수 있다. 이들 풀링 바코드는, 본 발명의 기반이 되는 "이전성 분자 식별 바코드"와 상이하다. 바코드의 동의어는 지표, 태그, 뉴클레오타이드 태그, MID(분자 식별자)일 수 있다.

[0186] 풀링 바코드는 이전성 분자 식별 바코드를 그와 같이 또는 예컨대 개체의 게놈 DNA (유래의) 다른 표적 핵산 또는 표적 뉴클레오타이드 서열과 조합하여 포함하는 임의의 표적 핵산 또는 표적 뉴클레오타이드 서열에 첨가될 수 있다. 그런 다음, 이전성 분자 식별 바코드 유래의 표적 뉴클레오타이드 서열은 2개의 바코드 서열들인 이전성 분자 식별 바코드 서열 유래의 최소 바코드 서열 및 풀링 바코드 서열 유래의 바코드 서열을 가질 것이다. 풀링 바코드는 하나의 말단에서 표적 뉴클레오타이드 서열의 측면에 존재하는 하나의 바코드일 수 있거나, 각각의 말단에서 표적 뉴클레오타이드 서열의 측면에 존재하는 2개의 분할된 바코드일 수 있다. 분할된 바코드인 경우, 2개의 분할된 바코드들은 동일하거나 상이할 수 있다(이중 지표화). 상이한 경우, 2개의 분할된 바코드들은 조합된 하나의 풀링 바코드를 확인하고, 조합은 독특하다.

[0187] 본원에 사용된 바와 같이 용어 "**증폭**"은 적어도 하나의 표적 핵산의 적어도 일부가 비제한적으로 핵산 서열의 증폭을 위한 광범위한 기술들을 포함하여 전형적으로 주형-의존적 방식으로 복제되는 (예를 들어 선형적으로, 기하급수적으로, 등온적으로, 서모사이클링(thermocycling)) 임의의 수단을 지칭한다. 증폭 단계를 수행하기 위한 예시적인 수단은 DNA 중합효소 반응, 프라이머 연장, 역방향 전사, PCR, 연결효소 사슬 반응(LCR), 올리고뉴클레오타이드 연결 검정법(OLA), 연결효소 검출 반응(LDR), 연결 및 후속해서 Q-레플리카제 증폭, 서큘러라이제이션(원형ization)-기반 DNA 합성 또는 증폭(HaloPlex™), 분자 인버젼 프로브(MIP) DNA 합성, 가닥 대체 증폭(SDA), 과분지화된 가닥 대체 증폭, 다중 대체 증폭(MDA), 롤링 서클 증폭(RCA), 루프 매개 등온 증폭(LAMP), 스마트 증폭 프로세스(SMAP), 핵산의 등온 및 키메라 프라이머-개시 증폭(ICAN®), 핵산 가닥-기반 증폭(NASBA), 전사-매개 증폭(TMA) 등, 예컨대 멀티플렉스 버전 및 이의 조합, 예를 들어 비제한적으로 PCR/PCR(2-단계 PCR), 프라이머 연장/OLA, 프라이머 연장/OLA/PCR, OLA/PCR, MIP/PCR, LDR/PCR, PCR/PCR/PCR(예를 들어 PCR/(네스티드(nested)-)PCR/(풀링-)PCR), PCR/PCR/LDR, PCR/LDR, LCR/PCR, PCR/LCR(또한 조합된 사슬 반응--CCR로도 공지

되어 있음) 등을 포함한다. 이러한 기술의 설명은 다른 출처들 중에서도 문헌[Ausbel et al.; PCR Primer: A Laboratory Manual, Diffenbach, Ed., Cold Spring Harbor Press (1995); The Nucleic Acid Protocols Handbook, R. Rapley, ed., Humana Press, Totowa, N.J. (2002); Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press (1990)]에서 확인될 수 있다.

[0188] DNA 또는 RNA의 증폭 및 프로세싱에 사용될 수 있는 DNA 중합효소, DNA 연결효소, 역전사효소 및 이들의 돌연변이체 및 변이체의 예로는 비제한적으로: DNA 중합효소 I, DNA 중합효소 I 큰(large) Klenow 단편, T4 DNA 중합효소, T7 DNA 중합효소, 종결 데옥시뉴클레오티딜 트랜스퍼라제, T4 DNA 연결효소, Taq DNA 중합효소, AmpliTaq Gold®, Taq DNA 중합효소 하이 피델리티, Tfl DNA 중합효소, Tli DNA 중합효소, Tth DNA 중합효소, Vent® DNA 중합효소, phi29 DNA 중합효소, Bst DNA 중합효소, Taq DNA 연결효소, Pfu DNA 연결효소, AMV 역전사효소, MMLV 역전사효소 등이 있다.

[0189] 본원에 사용된 바와 같이 용어 "**앰플리콘**"은 특정 핵산 서열을 핵산 증폭 기술, 예컨대 PCR에 의해 증폭시킴으로써 수득되는 표적 뉴클레오타이드 서열 또는 이의 수합물(집단)을 지칭한다. 용어 "앰플리콘"은 광범위하게는, 임의의 증폭 방법에 의해 생성되는 분자들의 임의의 수합물을 포함한다.

[0190] 본원에 사용된 바와 같이 용어 "**시퀀싱 주형 제조**"는, (하나의) 표적 핵산(들), 또는 표적 뉴클레오타이드 서열(들)이 시퀀싱을 위해 제조되는 방법 및 반응을 지칭한다. 일반적으로, 이러한 제조의 종료 시, 하나 이상의 어댑터의 측면에 존재하는 선형 또는 원형 표적 뉴클레오타이드 서열이 수득된다. 일부 어댑터는 시퀀싱 주형의 제조에만 필요한 한편, 다른 어댑터는 실제 시퀀싱, 또는 이의 조합에만 필요하다. 이때, 대부분의 고도 병행 시퀀싱 기술들은, 이들의 각각의 시퀀싱 플랫폼 상에서 시퀀싱을 수행하기 위해 특이적인 각각의 어댑터 서열을 필요로 한다.

[0191] 이러한 어댑터는 주어진 단계에서 증폭 방법에 의해 첨가될 수 있다. 예를 들어, 적어도 하나의 프라이머가 표적-특이적인 결합 부위 및 이러한 표적-특이적인 부위의 5' 말단에 위치한 어댑터를 포함하고, 제2 프라이머가 표적-특이적인 부위만 포함하거나 또는 표적-특이적인 부위 및 이러한 표적-특이적인 부위의 5' 말단에 위치한 어댑터를 포함하는 PCR에 의해서이다. 프라이머 부위에 관하여 본원에 사용된 바와 같이, 용어 "표적-특이적인" 뉴클레오타이드 서열은 적합한 어닐링 조건 하에 표적 핵산 또는 표적 뉴클레오타이드 서열 내의 프라이머 결합 부위에 특이적으로 어닐링할 수 있는 서열을 지칭한다.

[0192] 대안적으로, 하나 이상의 어댑터는 또한 예를 들어, 표적 핵산 또는 표적 뉴클레오타이드 서열의 한쪽 말단 또는 양쪽 말단에서 연결 어댑터의 연결 반응에 의해 첨가될 수 있다. 대부분의 적용에서, 표적 핵산이 우선, 예를 들어 물리적 수단(예를 들어 초음파처리, 온도)에 의해, 효소적 수단에 의해, 태그화(tagmentation)에 의해 단편화된다.

[0193] 시퀀싱 주형의 제조에 사용되는 어댑터는 심지어, 원형 표적 뉴클레오타이드 서열 주형의 제조를 위해 큰 5' 및 3' 오버행들을 가질 것이다. HaloPlex™ 검정법에서, 더 큰 5' 및 3' 오버행들은 표적화된 DNA 제한 단편의 양쪽 말단에 혼성화하며, 이로써 표적화된 단편을 가이드하여 원형 DNA 분자를 형성한다. 어댑터는 하나 이상의 다른 어댑터 기능기를 함유할 수 있으며, 이러한 기능기들은 시퀀싱 주형의 추가의 프로세싱 및 제조 및/또는 실제 시퀀싱의 수행에 필요하며, 이들은 서큘러라이제이션 후 표적 뉴클레오타이드 서열의 측면에 존재한다.

[0194] 상이한 시퀀싱 주형 제조 방법(검정법, 패널, 키트)은 예를 들어 제한 없이: TruSeq® DNA 시료 제조(Illumina), Nextera® DNA 시료 제조(Illumina), 프라이머-연장-연결에 의한 TruSeq 앰플리콘 제조(Illumina), TruSeq 가닥 mRNA 라이브러리 제조(Illumina), TruSeq RNA Access 라이브러리 제조(Illumina), TruSeq 표적화된 RNA 발현(Illumina), Ion Xpress™ 플러스 단편 라이브러리 제조(Ion Torrent, Thermo Fisher Scientific), Ion AmpliSeq™ DNA 및 RNA 라이브러리 제조(Ion Torrent, Thermo Fisher Scientific), SOLiD™ 단편 라이브러리 제조(Thermo Fisher Scientific), Titanimum 라이브러리 제조(454 Life Sciences, Roche), DNA 나노볼(DNB) 라이브러리 제조(Complete Genomics, BGI), SMRTbell 주형 제조(Pacific Biosciences), MinION 라이브러리 제조(Oxford Nanopore Technologies), GeneRead 라이브러리 제조(Qiagen), GeneRead DNAseq 유전자 라이브러리 제조(Qiagen), SureSelect^XT 라이브러리 제조(Agilent Technologies), 올리고뉴클레오타이드-Selective 시퀀싱(OS-Seq™; Blueprint Genetics), NEBNext® 라이브러리 제조(New England BioLabs), Access Array™ 표적화된 라이브러리 농화(Fluidigm), SmartChip™ 라이브러리 제조(Wafergen Biosystems), 리시퀀싱을 위한 특이적인 표

적의 멀티플렉스 증폭(MASTR)(Multiplicom), Devyser 멀티플렉스 PCR NGS 검정법(Devyser), HEAT-Seq 표적 농화(Roche), KAPA 라이브러리 제조(및 Hyper Prep 및 Hyper 플러스)(Kapabiosystems), ThruPLEX®, PicoPLEX™, TransPLEX® 라이브러리 제조(Rubicon Genomics), Accel-NGS DNA 라이브러리 제조(Swift Biosciences), Accel-앰플리콘™ 패널 제조(Swift Biosciences), Archer FusionPlex™ 및 변이체Plex™ 라이브러리 제조(Archerdx), Immunoseq® 및 Clonoseq 라이브러리 제조(Adaptive biotechnologies), 단일 프라이머 농화 기술(SPET)에 의한 라이브러리 제조(NuGEN), QuantSeq-Flex 표적화된 RNA 제조(Lexogen) 등이 있다.

[0195] 시퀀싱 주형 제조의 경우, 이전성 분자 식별 바코드는 표적 핵산 그 자체일 수 있거나, 개체 또는 환자, 동물, 식물, 박테리아, 바이러스 또는 진균류의 다른 표적 핵산, 예컨대 DNA 또는 RNA와 혼합될 수 있다. 당업자는, 이전성 분자 식별 바코드가, 시퀀싱 주형을 표적 핵산으로부터 제조할 수 있는 임의의 방법, 검정법 또는 키트에 적용된다는 것을 이해할 것이다.

[0196] 본원에 사용된 바와 같이 용어 "**프로브**"는 하나 이상의 유형의 화학 결합을 통해, 일반적으로 상보적인 염기쌍 형성을 통해, 통상적으로 수소 결합 형성을 통해 상보적인 서열의 표적 핵산에 결합하여, 듀플렉스 구조를 형성할 수 있는 핵산을 지칭한다. 이러한 프로브는 프로브 결합 부위에 결합하거나 혼성화한다. 프로브는, 특히 일단 프로브가 이의 상보적인 표적에 혼성화하면, 프로브의 손쉬운 검출을 허용하기 위해 검출 가능한 표지로 표지될 수 있다. 대안적으로, 프로브는 비표지될 수 있으나, 표지된 리간드와의 특이적인 결합에 의해 직접적으로 또는 간접적으로 검출될 수 있다. 프로브는 크기가 유의하게 다양할 수 있다. 일반적으로, 프로브는 적어도 7 내지 15개 뉴클레오타이드 길이이다. 다른 프로브는 적어도 20, 30, 또는 40개 뉴클레오타이드 길이이다. 더욱 다른 프로브는 다소 더 길며, 적어도 50, 60, 70, 80 또는 90개 뉴클레오타이드 길이이다. 보다 다른 프로브는 더욱 더 길고, 적어도 100, 150, 200 또는 그 이상의 개수의개 뉴클레오타이드 길이이다. 프로브는 또한, 상기 값들 중 임의의 값에 의해 경계가 생기는 임의의 범위 내의 임의의 길이일 수 있다(예를 들어 15 내지 20개 뉴클레오타이드 길이).

[0197] 프로브는 표적 핵산 서열에 완벽하게 상보적일 수 있거나, 완벽한 것보다 덜 상보적일 수 있다. 소정의 실시형태에서, 프라이머는 적어도 7개 뉴클레오타이드, 보다 전형적으로 10 내지 30개 뉴클레오타이드 범위의 서열, 종종 적어도 14 내지 25개 뉴클레오타이드 범위의 서열에 걸쳐 표적 핵산 서열의 상보체와 적어도 50% 동일성을 가지고, 보다 종종 적어도 65% 동일성, 적어도 75% 동일성, 적어도 85% 동일성, 적어도 90% 동일성, 또는 적어도 95%, 96%, 97%. 98%, 또는 99% 동일성을 가진다. 소정의 염기(예를 들어 프라이머의 3' 염기)는 일반적으로 표적 핵산 서열의 상응하는 염기에 바람직하게는 완벽하게 상보적인 것으로 이해될 것이다. 프라이머 및 프로브는 전형적으로, 엄격한 혼성화 조건 하에 표적 서열에 어닐링한다.

[0198] 본원에 사용된 바와 같이 용어 "**포착 올리고뉴클레오타이드**"는, 핵산들의 보다 복잡한 혼합물로부터 프로세싱될 뿐일 표적 핵산 또는 표적 뉴클레오타이드 서열의 관심 특정 표적에 혼성화하는 하나 이상의 올리고뉴클레오타이드 또는 프로브를 지칭한다. 포착 올리고뉴클레오타이드가 결합하는 표적의 특이적인 서열은 본원에서 올리고뉴클레오타이드 결합 서열로도 지칭될 것이다.

[0199] 이러한 방식으로, 오로지 관심 게놈 영역만 예컨대 시퀀싱, 포착 올리고뉴클레오타이드와의 혼성화-기반 서열 포착을 통한 관심 게놈의 DNA 영역의 단리에 의해 프로세싱될 것이다. 표적 핵산 내의 올리고뉴클레오타이드 결합 부위는 '**포착 서열**'로 지칭된다. 뉴클레오타이드 바코드 서열 내의 포착 서열은 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않는 서열이거나(따라서 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만으로 상동성인 서열을 가진다.

[0200] 포착 프로브는 단리를 용이하게 하기 위해 변형을 함유할 수 있다. 대부분의 경우, 이들 프로브는 비오틴화된다. 예를 들어, 관심 핵산의 표적은 하나 이상의 유전자의 엑손 서열, 또는 심지어 엑솜(exome)으로 공지된 게놈의 모든 유전자들 중 대부분의 엑손 서열일 수 있다. 표적 뉴클레오타이드 서열은 시퀀싱 주형 라이브러리 제조 방법에 의해 제조되었을 수 있으며, 이러한 방법에서 표적 뉴클레오타이드 서열은 예를 들어 총 뉴클레오타이드 서열의 1% 내지 3% 미만을 나타낸다. 총 시퀀싱 주형 라이브러리가 시퀀싱될 경우, 수득된 서열 중 단지 1% 내지 3%만 관심 서열이고 사용될 것이다. 총 라이브러리 유래의 관심 표적 뉴클레오타이드 서열은, 시퀀싱 전에 관심 영역의 이들 뉴클레오타이드 서열에 대한 포착 올리고뉴클레오타이드를 사용하여 특이적인 혼성화에 의해 선택적으로 농화될 수 있다. 포착은 용액 내에서 또는 물리적 지지체(어레이) 상에서 수행될 수 있

다. 포착 올리고뉴클레오타이드가 비오틴화되는 경우, 혼성화된 관심 단편은, 스트렙타비딘-코팅된 비드의 사용을 통해 관심이 아닌 비-혼성화된 단편으로부터 쉽게 단리될 수 있다. 특이적인 핵산 표적은 또한, 이전성 분자 식별 바코드 유래의 뉴클레오타이드 바코드일 수 있을 것이며, 이러한 바코드에 대해 이전성 분자 식별 바코드 내의 불변 서열 영역에 대한 단일 포착 올리고뉴클레오타이드가 디자인 및 제조될 수 있으며, 따라서 모든 유형의 뉴클레오타이드 바코드들은 존재하는 상이한 최소 바코드 서열과는 상관없이 단리 및 특징화될 수 있다. 이전성 분자 식별 바코드는 그와 같이 포착될 수 있으며, 이러한 분자 식별 바코드가 다른 핵산 표적, 예컨대 개체 또는 환자의 DNA와 혼합되는 경우 다른 표적 핵산과 조합된다.

[0201] 본원에 사용된 바와 같이 용어 DNA 혼합물(예를 들어 게놈, 이전성 분자 식별 바코드와 혼합된 게놈) 내의 (하나의) 특정 영역(들)의 '**농화**'는 핵산 내의 표적 핵산으로부터 증폭 또는 (혼성화-기반 서열) 포착을 통해 표적 뉴클레오타이드 서열을 생성 및/또는 단리하는 것을 지칭한다.

[0202] 상이한 포착 방법, 검정법, 키트는 예를 들어 비제한적으로: TruSight 시퀀싱 패널(Illumina), Nextera Rapid 포착 키트(Illumina), TargetSeq™ Exome 농화(Thermo Fisher Scientific), HaloPlex™ 농화(Agilent Technologies), SureSelect 표적 농화(Agilent Technologies), SeqCap EZ 농화(Roche NimbleGen), xGEN® 표적 포착(Integrated DNA Technologies) 등이다.

[0203] 본원에 사용된 바와 같이 용어 "**균등화(equalizing) 올리고뉴클레오타이드**"는 소정의 또는 모든 표적 뉴클레오타이드 서열들에서 확인되는 불변 서열에 특이적으로 혼성화하는 올리고뉴클레오타이드 또는 프로브를 지칭한다. 균등화 올리고뉴클레오타이드가 표적 핵산에서 결합할 수 있는 서열은 '균등화 서열'로 지칭된다. 균등화 서열은 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않는 서열이거나(따라서 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 또는 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다.

[0204] 균등화 올리고뉴클레오타이드는, 상이한 프로세싱된 시료 내의 표적 뉴클레오타이드 서열을 보다 균등한 수준까지 정규화(normalize)하는 데 사용된다. 이전성 분자 식별 바코드 내의 뉴클레오타이드 바코드 서열은 그와 같이 균등화될 수 있거나, 또는 이전성 분자 식별 바코드가 다른 핵산 표적, 예컨대 개체 또는 환자의 DNA와 혼합되는 경우 다른 표적 뉴클레오타이드 서열과 함께 균등화될 수 있을 것이다. 뉴클레오타이드 바코드 서열은 심지어, 다른 표적 뉴클레오타이드 서열 유래의 뉴클레오타이드 바코드 서열의 상이한 균등화를 허용하고, 다른 뉴클레오타이드 서열 유래의 뉴클레오타이드 바코드 서열의 균등화를 더 미세 조정(fine tune)하기 위해, 예를 들어 뉴클레오타이드 바코드 서열 내의 불변 서열 영역에 존재하는 균등화 서열을 통해, 다른 첨가된 핵산 유래의 뉴클레오타이드 서열에 존재하지 않는 제2 균등화 서열을 가질 수 있다. 균등화 올리고뉴클레오타이드는 용이한 추가의 프로세싱, 예컨대 스트렙타비딘-코팅된 비드를 통한 단리를 용이하게 하기 위해 비오틴화될 수 있다.

[0205] 본원에 사용된 바와 같이 용어 "**고도 병행 시퀀싱**"은 대량 병행 프로세싱의 개념을 사용하여 DNA 시퀀싱의 고-처리량 접근법을 지칭한다. 많은 고도 병행 시퀀싱 플랫폼들은 엔지니어링 배치 및 시퀀싱 화학의 측면에서 상이하다. 이들은 대체로, 유동 세포 내의 공간적으로 분리되며 클론적으로 증폭된 DNA 주형 또는 단일 DNA 분자를 통해 대량 병행 시퀀싱의 기술적 패러다임을 공유한다. 사용되는 동의어는 예를 들어, 차세대 시퀀싱(NGS), 2세대 시퀀싱, 3세대 시퀀싱, 대량적 병행 시퀀싱, 대량 병행 시퀀싱이다.

[0206] 상이한 고도 병행 시퀀싱 화학 및 플랫폼들은 예를 들어 비제한적으로: 피로시퀀싱(pyrosequencing), GS FLX(454 Life Sciences, Roche); 합성에 의한 시퀀싱, 가역적인 염료 종결자(Reversible Dye Terminator), HiSeq, MiSeq(Illumina); 올리고뉴클레오타이드 사슬 연결, SOLiD((Thermo Fisher Scientific), 양성자 검출을 기반으로 하는 이온 반도체 시퀀싱, Ion PGM™, Ion 양성자™, Ion S5™(Ion Torrent, Thermo Fisher Scientific) 및 GenapSys, 포토다이오드, 파이퍼플라이(Firefly)에 의한 형광 검출을 기반으로 하는 이온 반도체 시퀀싱(Illumina), 올리고뉴클레오타이드 언체인드(Unchained) 연결(Complete Genomics, BGI), 가역적인 염료 종결자, Heliscope(Helicos Biosciences), 포스포연결된(phospholinked) 형광 뉴클레오타이드, 실시간 SMRT® DNA 시퀀싱, Pacbio RS(Pacific Biosciences), 나노포어 시퀀싱, MinION™, PromethION™, GridION™(Oxford Nanopore Technologies), NanoTag 나노포어-기반 시퀀싱(Genia Technologies/Roche), Xpansion에 의한 시퀀싱(SBX, Stratos Genomics) 등이다. 당업자는, 이전성 분자 식별 바코드가, 핵산 또는 이의 부산물의 검출이 임의의 물리적, 화학적 및/또는 효소적 프로세싱 또는 이의 특성을 기반으로 하는 핵산의 임의의 병행 시퀀싱 방법

에 적용될 수 있음을 이해할 것이다.

[0207] 본원에 사용된 바와 같이 용어 '**뉴클레오타이드 바코드 서열 식별자 서열**', 축약해서 또한 '**식별자 서열**'은 DNA 분자 또는 이의 시퀀싱된 서열을 식별하기 위해 뉴클레오타이드 바코드 서열로서, 뉴클레오타이드 바코드 서열의 한쪽 또는 양쪽의 측면 불변 서열 내의 하나 이상의 어댑터 서열을 지칭한다. 뉴클레오타이드 바코드 서열 식별자 서열은 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않는 서열이거나(따라서, 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다. 이전성 분자 식별 바코드의 상이한 배치(batch)들은 상이한 뉴클레오타이드 바코드 서열 식별자 서열을 가질 것이고, 각각 상이한 적용들 및/또는 동일한 적용 내의 상이한 단계들에서 사용되어, 뉴클레오타이드 바코드 서열 식별자 서열은 상이한 배치를 식별하고, 따라서 적용 및/또는 적용 내의 단계를 식별한다.

[0208] 본원에 사용된 바와 같이 용어 '**추출 서열**'은 실제 최소 바코드 서열을 추출하기 위한, 뉴클레오타이드 바코드 서열 내의 최소 바코드 서열에 대해 한쪽 또는 양쪽 측면 서열에서 하나 이상의 어댑터 서열을 지칭한다. 추출 서열은 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되지 않는 서열이거나(따라서, 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음), 임의의 천연 발생 게놈, 바이러스, 박테리아 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만 상동성인 서열을 가진다.

[0209] 추출 서열은 뉴클레오타이드 바코드 서열 식별자 서열과 동일하거나 중첩일 수 있다. 전형적으로, 시퀀싱된 뉴클레오타이드 바코드 서열의 생물정보학적 분석은 2개의 단계, 또는 심지어 2개의 정보 파이프라인 또는 프로그램을 필요로 한다. 제1 프로그램은 모든 시퀀싱된 서열들로부터 뉴클레오타이드 바코드 서열 식별자 서열을 통해 시퀀싱된 뉴클레오타이드 바코드 서열을 단리하고, 제2 프로그램은 이들 시퀀싱된 뉴클레오타이드 바코드 서열 내의 최소 뉴클레오타이드 바코드 서열을 추출 서열을 통해 추출한다. 추출 서열이 뉴클레오타이드 바코드 서열 식별자 서열이 아니라 이전성 분자 식별 바코드의 상이한 배치들 사이에서 동일한 경우, 상이한 제1 생물정보학적 프로그램(프로그램 내의 상이한 설정임)이 식별자 서열에 따라 이전성 분자 식별 바코드의 각각의 배치에 사용되어야 하지만, 동일한 제2 정보 프로그램(프로그램 내의 동일한 설정임)은 이전성 분자 식별 바코드의 상이한 배치들(상이한 뉴클레오타이드 바코드 서열 식별자 서열을 가짐)로부터 모든 최소 뉴클레오타이드 바코드 서열들을 추출하는 데 사용될 수 있을 것이다.

[0210] 본원에 사용된 바와 같이 용어 "**담체(들)**"는 이전성 분자 식별 바코드를 함유하는 기판 및 용기를 지칭한다. 담체는 생물학적 시료의 수합에 사용될 수 있을 것이다. 담체는 생물학적 시료를 수합하는 또 다른 담체에 이의 내용물을 이전하는 데 사용될 수 있을 것이다.

[0211] 본 발명의 일부는 독특한 코드를 생성하기 위해 유전자 코드(A, C, T, G, U의 서열은 핵산에 존재하는 염기, 즉, 각각 아데닌, 시토신, 티로신, 구아닌 및 우라실을 나타냄)를 이용하며, 본원에서 이는 뉴클레오타이드 바코드로 지칭된다. 이들 코드는 특정 종류, 기원, 프로세싱 또는 치료의 아이템, 예컨대 인간, 동물, 식물, 박테리아, 바이러스 또는 진균류 기원의 생물학적 시료의 식별자에 사용될 수 있고, 이전성 분자 식별 바코드로서 사용된다. 이러한 생물학적 시료는 유기체, 전체 세포, 단일 세포, 세포 조제물, 및 임의의 유기체, 조직, 세포 또는 환경 유래의 세포-무함유 조성물을 포함하여 임의의 적합한 위치로부터 수득될 수 있다. 생물학적 시료는 또한, 세포-무함유, 예컨대 순환 핵산(예를 들어 DNA, RNA), 예컨대 혈액 내 순환 종양 DNA, 또는 임산부의 혈액 내 순환 태아 DNA일 수 있다. 생물학적 시료는 환경적 생검, 흡인물, 포르말린-고정 포매된 조직, 공기, 농업 시료, 토양 시료, 석유 시료, 물 시료 또는 먼지 시료로부터 수득될 수 있다. 일부 경우, 시료는 체액으로부터 수득될 수 있으며, 이러한 체액은 혈액, 소변, 대변, 혈청, 림프, 침, 점막 분비물, 땀, 중추신경계액, 질액 또는 정액을 포함할 수 있다. 시료는 또한, 제작된 제품, 예컨대 화장품, 식품(예컨대 육류, 우유, 와인, 올리브유), 퍼스널 케어 제품 등으로부터 수득될 수 있다. 시료는 재조합 클로닝, 폴리뉴클레오타이드 증폭, 중합효소 사슬 반응(PCR) 증폭, 정제 방법(예컨대 게놈 DNA 또는 RNA의 정제) 및 합성 반응을 포함하여 실험 조작의 생성물일 수 있다.

[0212] 짧은 DNA 분자 또는 올리고뉴클레오타이드는 유전자 코드의 "문자"를 임의의 요망되는 순서로 갖도록 제조될 수 있고, DNA 분자의 이들 문자의 특정한 조합은 특정 의미를 갖는 것으로 지정될 수 있다.

[0213] 작은 올리고뉴클레오타이드의 생성을 위한 하나의 바람직한 방식은, 천연 또는 화학적으로 변형된 뉴클레오사이드의 보호된 포스포아미다이트, 또는 적은 정도로는 비-뉴클레오사이드 화합물의 보호된 포스포아미다이트인 빌

딩 블록을 사용하는 화학적 합성에 의해서이다. 올리고뉴클레오타이드 사슬 어셈블리는 합성 동안 합성 사이클로 지칭되는 일상적인 절차에 따라 3'-말단으로부터 5'-말단 방향으로 진행된다. 단일 합성 사이클의 완료의 결과, 1개의 뉴클레오타이드 잔기가 성장하는 사슬에 첨가된다. 각각의 합성 단계의 100% 미만의 수율 및 부반응의 발생은 프로세스의 효율의 실제적인 한계를 설정하여, 합성 올리고뉴클레오타이드의 최대 길이가 200개 뉴클레오타이드 잔기를 초과하기 힘들다. 이러한 절차에서, 올리고뉴클레오타이드는 1개씩 생성된다.

[0214] 올리고뉴클레오타이드는 또한, 여러 가지 기술들, 예컨대 미리-제조된 마스크를 사용하는 포토리소그래피, 동적 마이크로미러 장치를 사용하는 포토리소그래피, 잉크젯 프린팅, 마이크로전극 어레이 상에서의 전기화학을 사용하여 마이크로어레이 상에서 병행 생성될 수 있다.

[0215] 합성될 수 있는 상이한 올리고뉴클레오타이드의 수는 방대하다. 예를 들어, 5개 뉴클레오타이드 길이의 올리고뉴클레오타이드가 생성되는 경우, 총 1,024(4^5)개의 상이한 올리고뉴클레오타이드가 생성될 수 있다. 10개 뉴클레오타이드 길이의 올리고뉴클레오타이드가 생성되는 경우, 1 x 10^6개(4^{10}) 초과의 상이한 올리고뉴클레오타이드가 생성될 수 있다(표 1).

표 1

[0216]

길이	상이한 뉴클레오타이드 서열의 수
1	4
2	16
3	64
4	256
5	1024
6	4096
7	16384
8	65536
9	262144
10	1048576
15	1073741824
20	1099511627776
25	1125899906842620

[0217] ***표 1.*** *DNA 서열의 길이에 따른 상이한 뉴클레오타이드 서열의 수.*이러한 각각의 뉴클레오타이드 바코드 서열은 하나의 단일 아이템, 예컨대 생물학적 시료를 표지하기 위한 이전성 분자 식별 바코드로서 사용될 수 있을 것이다. 그러나, 이러한 높은 수의 뉴클레오타이드 바코드의 합성은 꽤 시간이 소요되고 비용이 든다.

[0219] *보다 경제적인 바람직한 방식은, 시료 또는 아이템을 마크하는 데 사용될 수 있는 이전성 분자 식별 바코드를 구성하기 위해 1개 초과의 DNA 분자르르 사용하는 것일 것이다. 1개의 이전성 분자 식별 바코드 당 3개의 뉴클레오타이드 바코드가 사용되는 경우, 30개의 상이한 뉴클레오타이드 바코드가 1,000개의 상이한 독특한 이전성 분자 식별 바코드의 생성을 허용한다. 1개의 이전성 분자 식별 바코드 당 6개의 뉴클레오타이드 바코드가 사용되는 경우, 60개의 상이한 뉴클레오타이드 바코드가 1 x 10^6개의 상이한 독특한 이전성 분자 식별 바코드의 생성을 허용하는 등이다(표 2).

표 2

[0220]

이전성 분자 식별 바코드 당 상이한 뉴클레오타이드 바코드 (a)	필요한 뉴클레오타이드 바코드의 수	상이한 이전성 분자 식별 바코드의 수
2	20	100
3	30	1000
4	40	10000
5	50	100000
6	60	1000000
7	70	10000000
8	80	100000000
9	90	1000000000
10	100	10000000000

15	150	1000000000000000
20	200	10000000000000000000
25	250	100000000000000000000000000

[0221] (a) 10개 뉴클레오타이드 바코드 세트가 사용되며; 각각의 뉴클레오타이드 바코드에 대해, 당업자는 10개 뉴클레오타이드 바코드들 사이에서 선택할 수 있다.***표 2.*** *이전성 분자 식별 바코드 당 사용되는 뉴클레오타이드 바코드의 수의 함수로서, 제조될 수 있는 상이한 이전성 분자 식별 바코드의 수.*

[0222] 따라서, 이전성 분자 식별 바코드의 생성을 위해 뉴클레오타이드 바코드들의 혼합물을 사용하는 것이 가장 경제적이다. 1개 이전성 분자 식별 바코드 당 더 많은 뉴클레오타이드 바코드가 사용될수록, 프로세스는 더욱 경제적이게 된다. 각각이 6개 뉴클레오타이드 바코드로 구성된 많은 이전성 분자 식별 바코드를 생성하기 위해, 단지 60개의 상이한 뉴클레오타이드 바코드가, 6개 뉴클레오타이드 바코드의 모든 조합들로 구축된 1 x 10^6개의 상이한 이전성 분자 식별 바코드의 생성에 필요할 것이다. 이는, 3개의 DNA 분자로만 구축된 1 x 10^6개의 상이한 이전성 분자 식별 바코드의 구축에 30,000개의 상이한 뉴클레오타이드 바코드가 필요할 것이라는 점과는 대조적이다(표 3).

표 3

[0223]

이전성 분자 식별 바코드 당 상이한 뉴클레오타이드 바코드 (a)	필요한 뉴클레오타이드 바코드의 수	상이한 이전성 분자 식별 바코드의 수	1 x 10^6개 이전성 분자 식별 바코드	10 x 10^6개 이전성 분자 식별 바코드	100 x 10^6개 이전성 분자 식별 바코드
2	20	100	200000	2000000	20000000
3	30	1000	30000	300000	3000000
4	40	10000	4000	40000	400000
5	50	100000	500	5000	50000
6	60	1000000	60	600	6000
7	70	10000000		70	700
8	80	100000000			80

[0224] (a) 10개 뉴클레오타이드 바코드의 세트가 사용된다; 각각의 뉴클레오타이드 바코드에 대해, 당업자는 10개 뉴클레오타이드 바코드들 사이에서 선택할 수 있다.***표 3.*** *이전성 분자 식별 바코드에서 뉴클레오타이드 바코드의 수에 따라, 주어진 수 x 10^6개 이전성 분자 식별 바코드를 생성하는 데 필요한 상이한 이전성 분자 식별 바코드의 수.*

[0225] 따라서, 독특한 이전성 분자 식별 바코드를 제조하고 사용하는 기술적 및 경제적 실현 가능성은 상이한 뉴클레오타이드 바코드들의 조합 효과에 있다. 그런 다음, 하나의 이전성 분자 식별 바코드는 상이한 뉴클레오타이드 바코드들의 혼합물이다. 뉴클레오타이드 바코드들의 조합된 혼합물이 독특한 한, 이전성 분자 식별 바코드가 독특하다. 이전성 분자 식별 바코드가 x개의 뉴클레오타이드 바코드로 구축되는 경우, 2개의 이전성 분자 식별 바코드들은 이들이 x-1개의 뉴클레오타이드 바코드를 공통으로 갖지만 x번째의 뉴클레오타이드 바코드가 상이한 경우 여전히 독특할 수 있다.

[0226] 이전성 분자 식별 바코드를 이용하기 위해, 당업자는 결국 이전성 분자 식별 바코드에서 뉴클레오타이드 바코드를 검출하며, 식별하고/거나 특징화해야 한다. 이전성 분자 식별 바코드가 그와 같이 사용되는 경우, 오로지 뉴클레오타이드 바코드만 특징화될 필요가 있다. 이전성 분자 식별 바코드가 보다 복잡한 적용, 예컨대 유전자 검사에 사용되는 경우, 조사 하의 뉴클레오타이드 바코드 및 다른 핵산 표적 둘 다 특징화될 필요가 있다. 이는 당업자에게 공지되고 문헌[Molecular Cloning: A Laboratory Manual, by Sambrook and Russel, 3rd Edition, Cold Spring Harbor Laboratory Press, 2001(이의 개시내용은 원용에 의해 본 명세서에 포함됨)]에 기재된 바와 같은 분자 기술, 예컨대 DNA 합성, 중합, 연결, PCR, RT-PCR, 시퀀싱에 의해 수행될 수 있다.

[0227] 특징화 목적을 위한 단일 가닥 올리고뉴클레오타이드 바코드들의 프로세싱은 일부 단계에서 이중 가닥 DNA 분자로의 이들의 전환을 필요로 한다. 도 1에 기재된 바와 같이 하나의 바람직한 방식은 2개의, 적어도 부분적으로, 이중 가닥 연결 어댑터들의 연결이다. 이중 가닥 어댑터는 특이적인 다운스트림 프로세싱을 위해 다른 어댑터 서열을 운반할 수 있다. 부분적으로 이중 가닥은 또한, 다른 특징, 예컨대 헤어핀 구조 및 루프 구조를 운반할

수 있다. 뉴클레오타이드 바코드가 우선 인산화된다. 대안적으로, 이미 인산화된 뉴클레오타이드 바코드가 이전성 분자 식별 바코드의 제조에 사용된다. 하나의 이중 가닥 어댑터는, 단일 가닥 뉴클레오타이드 바코드의 한쪽 말단에 결합할 수 있는 5' 뉴클레오타이드 오버행을 가지고, 제2 이중 가닥 어댑터는 뉴클레오타이드 바코드의 다른쪽 말단에 결합할 수 있는 3' 뉴클레오타이드 오버행을 가진다. 전자의 이중 가닥 어댑터에서, 5' 뉴클레오타이드 오버행을 갖는 DNA 가닥은 이의 5' 말단에서 인산화된다. 3' 오버행을 갖는 후자의 이중 가닥 어댑터에서, 3' 오버행이 없는 반대쪽 가닥이 이의 5' 말단에서 인산화된다. 이들 이중 가닥 어댑터가 단일 가닥 뉴클레오타이드 바코드 서열에 혼성화된 후, 새로운 상보적인 DNA 가닥이 DNA 중합효소를 이용하여 합성될 수 있으며, 여기서, 3' 오버행 어댑터에서 3' 뉴클레오타이드 오버행을 갖는 DNA 가닥이, 다른쪽 말단에서 어댑터의 5" 뉴클레오타이드 오버행에 도달할 때까지 결합된 뉴클레오타이드 바코드를 주형으로서 사용하여 연장된다. 그런 다음, DNA 연결효소가 3개의 오픈 닉(nick)을 연결하는 데 사용되어, 완전한 이중 가닥 뉴클레오타이드 바코드가 수득된다. 2개의 이중 가닥 어댑터들에서 오버행은 1개 또는 그 이상의개 뉴클레오타이드 길이일 수 있다.

[0228] 2개 어댑터들의 오버행의 길이는, 길이가 동일하거나 상이할 수 있다. 오로지 1개의 1 뉴클레오타이드 오버행을 갖는 이중 가닥 연결 어댑터가 사용되는 경우, 당업자는 4개의 상이한 어댑터, 즉, A 오버행을 갖는 어댑터, C 오버행을 갖는 어댑터, G 오버행을 갖는 어댑터 및 T 오버행을 갖는 어댑터를 필요로 한다. 이들 어댑터는 물론 혼합될 수 있다. 5' 오버행 이중 가닥 연결 어댑터 및 3' 오버행 이중 가닥 연결 어댑터가 둘 다 1-뉴클레오타이드 오버행을 갖는 경우, 따라서 당업자는 임의의 서열 이중 가닥을 갖는 임의의 뉴클레오타이드 바코드를 제조하기 위해 8개의 상이한 어댑터들의 혼합물을 필요로 한다.

[0229] 이전성 분자 식별 바코드의 생성을 위해 단일 가닥 뉴클레오타이드 바코드를 사용하기 보다는, 이중 가닥 뉴클레오타이드 바코드가 이전성 분자 식별 바코드를 생성하기 위해 시작부터 사용된다. 이들 바코드는 특이적인 다운스트림 프로세싱 기능 및 적용을 위해 어댑터 서열을 운반할 수 있다. 이점은, 프로세싱, 특징화 및/또는 식별에 대한 일부 또는 모든 필요한 특징들은, 특징화를 위한 이들의 프로세싱 동안 연결에 의해 이들 특징을 부착시키기보다는, 이미 뉴클레오타이드 바코드 분자에 존재한다는 점이다.

[0230] 도 2는 이중 가닥 뉴클레오타이드 바코드를 구축하는 3개의 바람직한 방식들을 기재하고 있다. 제1 전략(도 2a)에서, 상이한 독특한 최소 뉴클레오타이드 바코드 서열 N_x(N은 임의의 뉴클레오타이드이며, x는 뉴클레오타이드의 수임)를 운반하는 더 긴 단일 가닥 올리고뉴클레오타이드가 사용되며, 이는 각각의 올리고뉴클레오타이드에 대해 공통적으로 불변 서열의 측면에 존재한다. 각각의 올리고뉴클레오타이드에서, 독특한 서열은 결국 최소 뉴클레오타이드 바코드 서열일 것이다. 불변 서열 또는 이의 일부는 특이적인 다운스트림 프로세싱 기능을 위해 다른 어댑터 기능을 가질 수 있다. 그런 다음, 이들 올리고뉴클레오타이드는 N_x 서열에 대해 측면 불변 서열 영역 3'에서 프라이머 결합 부위를 갖는 프라이머를 사용하여 DNA 합성 반응에 의해 이중 가닥이 될 수 있으며, 따라서 새로운 상보적인 DNA 가닥이 다른 제2 불변 서열 영역의 말단까지 N_x 서열에 걸쳐 합성된다. 프라이머가 3' 불변 서열 영역의 완전 말단에서 우측에 결합할 때, 완전 이중 가닥 DNA 분자가 수득되며, 그렇지 않다면 수득된 이중 가닥 뉴클레오타이드 바코드 분자는 하나의 말단에서 점착성일 것이다. 제2 전략(도 2b)에서, 각각의 유형의 올리고뉴클레오타이드에 특징적인 상이한 독특한 최소 뉴클레오타이드 바코드 서열 N_x, 및 후속해서 불변 서열을 운반하는 단일 가닥 올리고뉴클레오타이드가 사용되며, 따라서 최소 바코드 서열 영역은 단지 불변 서열 영역과 함께 하나의(3') 부위에서 측면에 존재한다. 본원에서 또한, 불변 서열 또는 이의 일부는 특이적인 다운스트림 프로세싱 기능을 위해 다른 어댑터 기능을 가질 수 있다. 그런 다음, 이들 올리고뉴클레오타이드는 최소 바코드 서열에 대해 측면 불변 서열 영역 3'에서 프라이머 결합 부위를 갖는 프라이머를 사용하여 DNA 합성 반응에 의해 이중 가닥이 될 수 있으며, 따라서 새로운 상보적인 DNA 가닥이 최소 바코드 서열 단독 및 그곳의 말단에 걸쳐 합성된다. 프라이머가 3' 불변 서열 영역의 완전 말단에서 우측에 결합할 때, 완전 이중 가닥 DNA 분자가 수득되며, 그렇지 않다면 수득된 이중 가닥 뉴클레오타이드 바코드 분자는 하나의 말단에서 점착성일 것이다. 매우 유사하게, 이러한 이중 가닥 DNA 분자는 다운스트림 프로세싱, 예컨대 불변 서열 영역이 원래 존재하지 않은 부위에서 말단에 불변 서열을 또한 첨가하기 위해 이중 가닥 연결 어댑터를 이용한 연결 단계를 필요로 할 것이다. 제3 전략(도 2c)은 제1 전략의 대안이며, 여기서, 상이한 독특한 최소 뉴클레오타이드 바코드 서열 N_x를 운반하는 더 긴 단일 가닥 올리고뉴클레오타이드가 사용되며, 이는 각각의 올리고뉴클레오타이드에 대해 공통적으로 불변 서열의 측면에 존재한다. 각각의 올리고뉴클레오타이드에서, 독특한 서열 N_x는 최소 뉴클레오타이드 바코드 서열일 것이다. 3' 불변 서열은 또 다른 올리고뉴클레오타이드와의 혼성화에 충분한 5' 불변 서열과 비교하여 더 작거나 더 큰 크기를 가질 수 있다. 제1 단일 가닥 올리고뉴클레오타이드의 최소 바코드 서열 영역에 대해 3'에서 불변 서열 영역에 상보적인 서열 또는 이의 일부를 이의 3' 말단에 운반하지만 추가의 5' 불변

서열을 운반하는 제2 단일 가닥 올리고뉴클레오타이드가 조합된다. 여기서 다시, 단일 가닥 올리고뉴클레오타이드 또는 이의 일부 내의 불변 서열은 특이적인 다운스트림 프로세싱을 위해 다른 어댑터 기능을 가질 수 있다. 이들의 상보적인 공유된 불변 3' 서열 영역들에서 2개의 유형의 단일 가닥 올리고뉴클레오타이드들의 혼성화 후, 새로운 DNA 가닥이, 다른 혼성화 올리고뉴클레오타이드가 주형으로서 사용되는 각각의 올리고뉴클레오타이드로부터 연장하여 합성되어, 완전한 이중 가닥 뉴클레오타이드 바코드가 수득된다. 선택적으로 하나 이상의 추가의 라운드(round)의 DNA 합성 반응이 불변 서열과 함께 (하나의) 추가의 올리고뉴클레오타이드(들)에 의해 수행될 수 있으며, 이는 특이적인 다운스트림 프로세싱을 위한 어댑터 기능을 운반할 수 있고, 이는 이전 라운드에서 생성된 DNA 분자 내에서 3'에 위치한 불변 서열 영역에 상보적인 서열을 이의 3' 말단에서 운반한다. 이러한 방식으로, 단일 올리고뉴클레오타이드의 전형적인 길이를 초과할 수 있는 긴 불변 어댑터 영역을 갖는 뉴클레오타이드 바코드가 생성될 수 있다. 일련의 새로운 반응에서 각각의 새로운 올리고뉴클레오타이드를 첨가하기보다는, 이러한 모든 올리고뉴클레오타이드들이 또한 하나의 단일 반응에서 혼합될 수 있고, DNA 합성이 열안정성(thermostable) DNA 중합효소를 이용하여 다수의 사이클 동안 수행될 수 있다. 최소 바코드 서열 및 완전 측면 서열을 함유하는 이중 가닥 바코드가 또한, 합성 생물학에서 사용되는 방법, 예컨대 gBlocks® Gene 단편(Integrated DNA Technologies)을 사용하여 하나의 진행에서 합성될 수 있다.

[0231] 이러한 방식으로 수득되는 상이한 어댑터 기능을 갖는 단일 가닥 또는 이중 가닥 뉴클레오타이드 바코드의 예들은 비제한적으로 도 3에 나타나 있다.

[0232] 하나 또는 둘 다의 측면 불변 서열은 분자생물학 프로토콜 및 툴에서 사용되는 자연상에서 확인되지 않는(인간, 동물, 식물, 박테리아, 바이러스, 진균류에서 확인되지 않으며, 심지어 클로닝 벡터(벡터 백본)에서 확인되지 않는) 인공 서열이다. 시퀀싱 적용에서, 불변 측면 서열 또는 (하나의) 이의 일부(들)와 동일한 뉴클레오타이드 서열이 확인되는 경우, 이러한 서열은 뉴클레오타이드 바코드 서열로부터 기원하고 이러한 방식으로 뉴클레오타이드 바코드 서열이 식별될 수 있는 것으로 결론내릴 수 있다. 특히 이전성 분자 식별 바코드가 다른 표적 핵산과 혼합되는 경우, 뉴클레오타이드 바코드 서열로부터 기원하는 서열 판독이 식별될 수 있다. 사실상, 뉴클레오타이드 바코드로부터 수득되는 서열 판독은 예를 들어 상이한 생물정보학적 파이프라인(예를 들어 뉴클레오타이드 바코드의 식별 및 정량화를 위한 파이프라인, 다른 표적 핵산에 대해 기원하는 서열 판독의 맵핑 및 변이체 콜링(calling)을 위한 파이프라인)에 의해 상이하게 프로세싱될 수 있다. 불변 측면 서열은 바람직하게는, 작은 GC% 함량 변이를 가지며, 바람직하게는 35% 내지 65%의 범위로 가진다. 예를 들어, 이러한 측면 서열의 X 뉴클레오타이드(X 뉴클레오타이드는 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45 및/또는 50임)의 임의의 연속 서열은 35-60%, 40-60%, 45-60%, 50-60%, 55-65%의 바람직한 GC-함량을 가진다. 예를 들어, 이러한 측면 서열의 X 뉴클레오타이드(X 뉴클레오타이드는 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45 및/또는 50임)의 임의의 연속 서열은 50℃ 내지 75℃, 55℃ 내지 75℃, 60℃ 내지 75℃, 60℃ 내지 70℃의 바람직한 Tm을 가진다. 이러한 방식으로, 임의의 요망되는 올리고뉴클레오타이드 결합 부위(예를 들어 증폭 또는 포착을 위한 프라이머에 대한)는 가장 유연하고 효율적인 방식으로 디자인될 수 있다. 사실상, 많은 유형의 유전자 NGS 검사들(상이한 회사들 유래)은 상이한 기술들을 사용하여 이용 가능하며, 이러한 기술들에서, 관심 게놈의 표적 영역은 시퀀싱을 위해, 예를 들어 멀티플렉스 PCR, 프라이머-연장-연결 등을 통해 농화된다. 각각의 유형의 검사는 이들의 특정한 검사 조건을 사용한다. 예를 들어, 하나의 유형의 검사는, 모든 프라이머들이 주어진 대략 동일한 Tm을 갖는 멀티플렉스 증폭을 사용할 수 있는 한편, 또 다른 벤더(vendor)는 또 다른 Tm을 갖는 프라이머를 사용한다. 또한, 수득된 검사에서 수득되는 앰플리콘은 주어진 특정한 크기 범위를 가질 수 있어서, 모든 앰플리콘들은 대체로 균등하게 증폭되고, 선택된 크기는 상이한 검사들 사이에서 다양할 것이다. 사실상, 더 작은 앰플리콘은 일반적으로, 더 큰 앰플리콘보다 효율적으로 증폭되어, 수득된 앰플리콘의 길이 또한, 주어진 검사의 특정한 특징일 소정의 범위에서 유지된다. 이들 검사가 품질 관리를 위해 이전성 분자 식별 바코드를 사용하기 원한다면, 뉴클레오타이드 바코드 서열을 농화시키는 이들의 검정법에 프라이머가 첨가되거나 포함되어야 한다. 그런 다음, 이들의 주어진 검정법에서 뉴클레오타이드 바코드의 농화에 사용되는 프라이머(들)가 바람직하게는, 이들의 검정법에서 다른 프라이머와 마찬가지로 동일한 특징, 예컨대 Tm을 가져야 한다. 더 큰 연속 GC 함량 범위를 불변 측면 영역에 가짐으로써, 이전성 분자 식별 바코드의 농화를 허용하는 (하나의) 프라이머(들)의 선택 및 첨가는 임의의 검사에서 가장 쉽게 통합된다. 매우 낮거나(<20%) 또는 매우 높은 (>70%) GC 기준을 갖는 서열 블록이 존재할 경우, 주어진 Tm(대체로 30-55℃의 범위임)을 갖는 (하나의) 프라이머(들) 결합 부위, 및 주어진 특정한 크기 범위의 앰플리콘이 수득되는 위치를 확인하는 기준은 수득하기 어렵거나 심지어 일부 검사의 경우 불가능할 것이다.

[0233] 예를 들어 도 3 및 도 4에 기재된 바와 같이 제한 효소 인지 부위 RE1로부터 출발하고 최소 뉴클레오타이드 바

코드 서열 전에 종결되는 단리된 업스트림 불변 측면 서열의 예는 서열 [서열 번호:1], [서열 번호:2], [서열 번호:3], [서열 번호:4], [서열 번호:5], [서열 번호:6], [서열 번호:7], [서열 번호:8], [서열 번호:9], [서열 번호:10]이다.

[0234] 예를 들어 도 3 및 도 4에 기재된 바와 같이 최소 뉴클레오타이드 바코드 서열 이후부터 출발하고 제한 효소 인지 부위 RE2에서 종결되는 단리된 다운스트림 불변 측면 서열의 예는 서열 [서열 번호:11], [서열 번호:12], [서열 번호:13], [서열 번호:14], [서열 번호:15], [서열 번호:16], [서열 번호:17], [서열 번호:18], [서열 번호:19], [서열 번호:20]이다.

[0235] 서열 [서열 번호:1] 내지 [서열 번호:20]의 변이체는 벡터 내의 클로닝 부위에 따라 대안적인 제한 부위 인지 서열을 가진다.

[0236] 대안적으로, [서열 번호:1] 내지 [서열 번호:10]은 다운스트림 불변 측면 서열이고, [서열 번호:11] 내지 [서열 번호:20]은 업스트림 불변 측면 서열이다.

[0237] 대안적으로, 하나 이상의 [서열 번호:1] 내지 [서열 번호:20] 서열은 각각의 [서열 번호:1] 내지 [서열 번호:20] 서열의 역방향 상보 서열일 수 있다.

[0238] 대안적으로, 업스트림 및/또는 다운스트림 불변 측면 서열은 서열 [서열 번호:1] 내지 [서열 번호:20]이며, 각각의 서열 [서열 번호:1] 내지 [서열 번호:20]의 서열 동일성과 70% 초과, 80% 초과, 90% 초과, 95% 초과, 97% 초과 또는 99% 초과의 서열 동일성을 보여주는 서열이다. 서열 동일성의 차이는 예를 들어, 제한 효소에 대한 인지 부위를 첨가하거나 결실시킨 결과일 수 있다.

[0239] 보다 대안적으로, 서열 [서열 번호:1] 내지 [서열 번호:20]을 포함하는 업스트림 및/또는 다운스트림 불변 측면 서열은, 지시된 제한 효소 인지 서열과 불변 서열 사이 및/또는 불변 서열과 최소 뉴클레오타이드 바코드 서열 사이에 추가의 뉴클레오타이드 서열이 존재해서이다.

[0240] 보다 대안적으로, 업스트림 및/또는 다운스트림 불변 측면 서열은 서열 [서열 번호:1] 내지 [서열 번호:20]의 단편, 즉, 적어도 200개 뉴클레오타이드, 적어도 300개 뉴클레오타이드, 적어도 350개 뉴클레오타이드, 적어도 375개 뉴클레오타이드, 또는 적어도 390개 뉴클레오타이드의 단편을 포함하거나 구성된다.

[0241] 이들 이중 가닥 DNA 분자는 또한, 플라스미드 또는 다른 복제 구축물에서 더 클로닝될 수 있다. 이들 이중 가닥 DNA 분자의 불변 측면 서열 내에 제한 효소에 대한 인지 부위를 갖는 어댑터는 플라스미드에서 이들 이중 가닥 DNA 분자의 클로닝을 용이하게 할 수 있다. 예를 들어, 불변 측면 서열들은 둘 다, 동일한 제한 효소, 또는 2개의 상이한 제한 효소들에 대한 인지 부위를 가질 것이다. 분해 후 점착성 DNA 말단을 생성하는, 각각의 측면 부위 내의 2개의 상이한 제한 효소들에 대한 인지 부위들은, 이들 인지 부위가 플라스미드 내에서 이중 가닥 DNA 분자의 효율적인 방향성 클로닝을 허용하기 때문에 바람직하다.

[0242] 이전성 분자 식별 바코드를 생성하기 위한 플라스미드의 사용은, 많은 이전성 분자 식별 바코드의 완전한 로트(lot)가 오로지 1개의 단일 올리고뉴클레오타이드 또는 1개의 합성 DNA 단편으로부터 출발할 수 있다는 이점을 가진다. 사실상, 주어진 독특한 뉴클레오타이드 바코드 서열을 갖는 단일 올리고뉴클레오타이드, 또는 주어진 뉴클레오타이드 바코드 서열을 갖는 이중 가닥 뉴클레오타이드 바코드 분자의 구축에 사용되는 올리고뉴클레오타이드가 동일한 불변 서열의 측면에 존재하는 경우, 당업자는 이들 상이한 올리고뉴클레오타이드를 하나씩 합성할 필요가 있다. 복제 분자, 예컨대 플라스미드 내에서 이전성 분자 식별 바코드가 사용되는 경우, 이러한 합성은 피해질 수 있고 보다 경제적으로 될 것이다. 그런 다음, 오로지 1개의 단일 올리고뉴클레오타이드 합성 반응이 요망된다. 올리고뉴클레오타이드 합성 반응 동안, 1개 초과의 뉴클레오타이드 또는 심지어 모든 4개의 가능한 뉴클레오타이드들이 합성 단계의 주어진 사이클 동안 첨가되고, 올리고뉴클레오타이드에 혼입될 수 있다. 모든 4개의 빌딩 블록들(N)이 연속된 수의 사이클에 걸쳐 첨가되는 경우, 이들 사이클의 수와 동일한 길이 (x)의 모든 가능한 서열들이 합성될 것이고, 따라서 해당 길이 N_x의 모든 가능한 무작위 서열을 갖는 올리고뉴클레오타이드가 단일 합성 반응에서 수득될 것이다. 1개 초과의 뉴클레오타이드가 첨가된 경우, 선행 사이클 및 후속 사이클 동안 단일 뉴클레오타이드가 첨가되고 혼입되는 경우, 올리고뉴클레오타이드들의 혼합물이 무작위 뉴클레오타이드 바코드 서열(N_x)을 갖는 단일 튜브 내에서 수득될 것이며, 이들 모든 바코드 서열들의 측면에 동일한 불변 서열이 존재한다. 유사하게는, 합성 DNA 단편들의 완전한 혼합물, 예컨대 gBlocks® Gene 단편(Integrated DNA Technologies)이 생성될 수 있으며, 이러한 생성물은 상이한 최소 바코드 서열 및 동일한 완전 측면 서열을 함유한다.

[0243] 실제 최소 뉴클레오타이드 바코드의 길이는 적용에 따라 임의의 적합한 길이일 수 있다. 모든 가능한 뉴클레오타이드들이 10개 사이클의 과정에 걸쳐 허용되는 경우, 4^{10}(1,048,576)개의 상이한 뉴클레오타이드 바코드 서열들이 단일 올리고뉴클레오타이드 합성 반응에서 합성될 수 있다. 모든 가능한 뉴클레오타이드들이 25개 사이클의 과정에 걸쳐 허용되는 경우, 4^{25}(1,125,899,906,842,620)개의 상이한 뉴클레오타이드 바코드 서열들이 단일 올리고뉴클레오타이드 합성 반응에서 합성될 수 있다(표 1). 일부 경우, 실제 최소 바코드 서열은 약 2 내지 약 500개 뉴클레오타이드 길이, 약 2 내지 약 100개 뉴클레오타이드 길이, 약 2 내지 약 50개 뉴클레오타이드 길이, 약 2 내지 약 25개 뉴클레오타이드 길이, 약 6 내지 약 25개 뉴클레오타이드 길이, 또는 약 4 내지 25개 뉴클레오타이드 길이일 수 있다. 일부 경우, 최소 바코드 서열은 약 10, 20, 100, 500, 750, 1000, 5000 또는 10000개 초과의 뉴클레오타이드 길이이다.

[0244] 이러한 단일 가닥 올리고뉴클레오타이드들의 혼합물의 최소 바코드 N_x 서열에 대해 불변 측면 서열 영역에 위치한 동일한 어댑터는 상기 기재된 바와 같이 제한된 수의 반응 또는 심지어 단일 반응에서 이중 가닥 DNA 분자로의 이들의 전환을 허용한다. 이론적으로 모든 가능한 최소 바코드 N_x 서열들을 갖는 이중 가닥 DNA 분자의 이러한 혼합물 또는 라이브러리로부터, 플라스미드 라이브러리는 이들 모든 가능한 최소 바코드 N_x 서열들을 함유하도록 상기 기재된 바와 같이 제조될 수 있다. 그런 다음, 이들 모든 가능한 최소 바코드 N_x 서열을 함유하는 이러한 플라스미드 라이브러리는 박테리아 세포의 형질변환에 사용될 수 있으며, 따라서 이들 모든 가능한 최소 바코드 N_x 서열들을 갖는 이들 상이한 플라스미드를 운반하는 박테리아 라이브러리가 수득된다. 이러한 박테리아 라이브러리로부터, 글리세롤 스탁(stock)이 향후 사용을 위해 생성될 수 있다. 각각의 수준에서, 즉, 임의의 가능한 최소 바코드 N_x 서열을 갖는 올리고뉴클레오타이드, 이의 이중 가닥 DNA 분자, 이의 플라스미드, 이의 박테리아 배양물은 사용 수명을 위해 단일 튜브 내에서, 당업자에게 공지되고 문헌[Molecular Cloning: A Laboratory Manual, by Sambrook and Russel, 3rd Edition, Cold Spring Harbor Laboratory Press, 2001]에 기재된 바와 같은 간단한 방법을 이용하여 수득된다. 각각의 뉴클레오타이드 바코드 조성물 포맷이 이전성 분자 식별 바코드의 생성에 사용될 수 있다. 심지어 뉴클레오타이드 바코드를 운반하는 박테리아 세포가 이전성 분자 식별 바코드의 생성에 사용되고, 그와 같이 적용에 사용될 수 있다. 플라스미드에 의해 형질변환된 박테리아가 본원에 기재되어 있으나, 임의의 복제 구축물, 및 복제 구축물을 위한 숙주가 사용될 수 있다.

[0245] 모든 이용 가능한 N_x 무작위 서열의 하나의 주어진 최소 뉴클레오타이드 바코드 서열을 운반하는 플라스미드는 필요할 때마다 한천 플레이트 배양 후 이러한 박테리아 배양물로부터 쉽게 수득될 수 있으며, 그런 다음 이들 배양물로부터 개별 콜로니를 골라내, 당업자에게 공지된 바와 같은 방법을 사용하여 주어진 최소 뉴클레오타이드 바코드 서열을 운반하는 플라스미드를 충분한 양으로 수합하기 위해 더 배양할 수 있다. 각각의 이전성 분자 식별 바코드가 6개의 상이한 뉴클레오타이드 바코드 플라스미드로 구축되는 경우, 오로지 60개의 박테리아 콜로니를 골라내고, 단리한 다음 성장시키고, 1 x 10^6개 상이한 이전성 분자 식별 바코드를 생성하기 위해 플라스미드 프렙(prep)해야 한다(표 2 참조).

[0246] 이전성 분자 식별 바코드의 적용에서, 뉴클레오타이드 바(bar) 서열이 결국 특징화되고 식별되어야 한다. 고도 병행 시퀀싱이 이들 서열의 특징화 방법일 수 있을 것이다. 최소 바코드 서열은 불변 측면 서열들 사이에서 이러한 바코드 서열의 위치에 의해 식별 가능할 것이다. 현재 대체로 사용되는 상업적인 고도 병행 시퀀싱 기술에 의해 수득되는 전형적인 서열 판독 길이는 100-150개 뉴클레오타이드 또는 그 이상이다. 또한, 최소 바코드 서열 영역의 측면 서열의 적어도 일부는, 특히 이들이 시퀀싱을 위한 다른 표적 핵산을 함유하는 혼합물에 존재하는 경우 뉴클레오타이드 바코드 서열을 식별하고 프로세스하기 위해 시퀀싱되어야 한다. 가능하게는 또한, 풀링 바코드 서열이 시퀀싱되어야 한다. 전형적으로 100-150개 뉴클레오타이드 길이인 서열 판독에서 이들 서열을 뺀 후, 이는 여전히 50-100개 뉴클레오타이드 길이의 최소 뉴클레오타이드 바코드 서열을 허용하며, 즉, 4^{50}-4^{100}개의 독특한 뉴클레오타이드 바코드 서열을 허용한다. 실질적인 이유에서, 상대적으로 더 짧은 길이 범위를 갖는 최소 뉴클레오타이드 바코드 서열을 사용하는 것이 관심있을 것이다. 첫째, 25개 뉴클레오타이드 길이의 최소 뉴클레오타이드 바코드 서열로 이미 수득될 수 있는 상이한 독특한 뉴클레오타이드 바코드 서열의 수는, 필요로 할 뉴클레오타이드 바코드 서열의 수를 이미 초과한다. 둘째, 소정의 고도 병행 시퀀싱 적용은 심지어 100-150개 뉴클레오타이드 길이 서열 판독을 필요로 하지 않을 것이다. 예를 들어, 고도 병행 시퀀싱 전사체 연구에서, 오로지 25-35개 뉴클레오타이드 길이의 짧은 서열 판독만 전사체 식별을 허용하는 데 필요할 수 있다. 단순히 뉴클레오타이드 위치 25-35를 능가하여 수득되는 뉴클레오타이드 바코드 서열을 무시함으로써 더 긴 최소 뉴

클레오타이드 바코드 서열도 이러한 적용에 사용될 수 있긴 하지만, 최소 뉴클레오타이드 바코드 서열의 처음 25개 뉴클레오타이드가 이미 매우 독특한 성향이 있을 것이기 때문에, 이러한 데이터의 분석은 더 복잡할 수 있고, 추가의 (생물)정보학적 프로세싱 툴을 필요로 할 것이다. 25개 뉴클레오타이드인 최소 뉴클레오타이드 바코드 서열로 이미 수득될 수 있는 상이한 뉴클레오타이드 바코드의 수가 이미 충분하다는 점을 고려하면, 추가의 (생물)정보학적 프로세싱은 불필요한 추가 노력이다.

[0247] 예를 들어 한쪽 또는 양쪽 불변 측면 서열에 위치하는 뉴클레오타이드 바코드 서열 식별자 서열을 통해 어떤 뉴클레오타이드 서열이 뉴클레오타이드 바코드로부터 유래되는지를 단순히 식별하기 보다는, 실제 최소 바코드 서열은 예컨대 정성화(아이덴터티) 및/또는 정량화(수) 목적을 위해 식별될 필요가 있을 수 있다. 예컨대 유전자 검사에서, 이전성 DNA 바코드는 품질 관리를 위해 돌연변이에 대해 분석되는 다른 핵산과 혼합된다.

[0248] 최소 뉴클레오타이드 바코드 서열은, 양쪽 불변 측면 서열에 위치하는 추출 서열 내/사이에서 상기 바코드 서열의 위치, 또는 하나의 불변 측면 서열에 위치하는 추출 서열로부터 출발하거나 추출 서열에서 종결되는 상기 바코드 서열의 위치에 의해 식별될 수 있다.

[0249] 불변 측면 서열 내의 임의의 서열이 사용될 수 있을 것이지만, 가장 효율적인 서열은 최소 뉴클레오타이드 바코드 서열의 바로 측면에 존재하는 뉴클레오타이드이다. 예를 들어, 최소 바코드 서열이 25개 뉴클레오타이드 서열 길이인 경우, 앞의(preceding) 7개 뉴클레오타이드는 제1 추출 서열로서 사용될 수 있고, 이후의 7개 뉴클레오타이드는 제2 추출 서열로서 사용될 수 있다. 보다 구체적으로, 모든 서열 판독들이 25개 뉴클레오타이드에 의해 분리된 2개의 추출 서열들에 대해 입증되며, 이후, 25개 뉴클레오타이드의 서열이 추출되고, 최소 바코드 서열로서 확인된다. 당연하게도, 혼합된 표적 핵산 유래의 다른 표적 뉴클레오타이드 서열들 중 일부 또한, 이들 기준을 충족할 것이며, 따라서 잘못된 최소 바코드 서열은 추출된다. 7개 뉴클레오타이드 길이의 2개의 주어진 추출 서열들이 25개 뉴클레오타이드에 의해 분리되는 가능성은 1/16807이다(표 4).

표 4

[0250]

길이 추출 서열	하나의 부위에서 추출 서열의 확률	양쪽 부위들에서 추출 서열의 확률
1	1	1
2	16	32
3	81	243
4	256	1024
5	625	3125
6	1296	7776
7	2401	16807
8	4096	32768
9	6561	59049
10	10000	100000
15	50625	759375
20	160000	3200000
25	390625	9765625
30	810000	24300000
35	1500625	52521875
40	2560000	102400000
45	4100625	184528125
50	6250000	312500000

[0251] ***표 4.*** *최소 바코드 서열의 한쪽 또는 양쪽에서 측면에 존재할 때, 주어진 추출 서열이 이의 길이를 기반으로 확인되는 확률.*이들 위양성을 갖는 확률은 추출 서열의 길이를 증가시킴으로써 감소된다. 이러한 위양성을 방지하기 위해, 더 긴 추출 서열이 바람직하며(표 4); 최상의 서열은 완전 불변 측면 서열이다.

[0252] 위(false) 최소 바코드 서열의 확인은 다른 혼합된 표적 핵산으로 인한 복잡성 및 추출 서열의 감소된 크기에 따라 증가한다. 혼합된 표적 핵산이 더 복잡할수록(예를 들어 이전성 분자 식별 바코드가 하나 또는 소수의 유전자의 표적 농화와 비교하여(versus) 단편화된 완전 게놈과 혼합되는 경우), 소정의 단편이 추출 서열의 기준을 충족할 수 있는 기회가 높아진다. 당연하게도, 뉴클레오타이드 바코드 서열과 다른 표적 뉴클레오타이드 서열의 혼합물에서, 뉴클레오타이드 바코드 서열 유래의 표적 뉴클레오타이드 서열가 뉴클레오타이드 바코드 서열

식별자 서열(추출 서열과 동일할 것임)을 통해 먼저 단리될 수 있고, 제2 단계 특징화/프로세스에서 단리된 뉴클레오타이드 바코드 서열 내의 실제 최소 바코드 서열만 단리될 수 있다. 이러한 방식으로, 당업자는 더 작은 추출 서열이 사용되는 경우, 위(false) 최소 바코드 서열의 추출을 심지어 방지할 것이다.

[0253] 최소 바코드 서열의 추출이 추출 서열(들)의 정확한 서열을 기반으로 할 수 있기 때문에, 불변 측면 시퀀싱, 보다 구체적으로 추출 서열 영역에 도입된 DNA 합성, 증폭 및 시퀀싱 오차는 정확하게 매치하지 않을 것이며, 따라서 최소 바코드 서열이 추출되지 않을 수 있다. 본원에서, 추출 서열의 길이가 짧을수록, 증폭 및/또는 시퀀싱 오차가 추출 서열의 영역에서 발생할 기회가 낮아진다. 서열 내의 시퀀싱된 염기의 품질은 단편의 말단으로 갈수록 감소하며, 따라서 낮은 신호를 갖는 시퀀싱된 단편 뉴클레오타이드의 말단에서 다듬어져서 제거될 것이다. 조사 하의 게놈 DNA, 및 따라서 또한 시료에 첨가되는 뉴클레오타이드 바코드가 시퀀싱 라이브러리의 제조를 위해 단편화(예를 들어 어댑터의 연결을 통해)될 경우, 뉴클레오타이드 바코드 서열 유래의 시퀀싱 단편의 상대적인 출발 및 말단은 달라진다. 말단이 최소 바코드 서열에 위치하는 경우, 이러한 최소 바코드 서열은 완전히 시퀀싱되지 않으며, 따라서 심지어 확인되지 않을 수 있다. 서열의 말단이 최소 바코드 서열 이후의 단지 소수의 뉴클레오타이드에 위치하는 경우, 단지 일부의 불변 측면 서열만 확인될 것이다. 추출 서열의 길이가 말단에서 수득되는 서열보다 긴 경우, 최소 뉴클레오타이드 바코드 서열이 완전히 시퀀싱되더라도, 당업자는 추출 서열이 인지되지 않기 때문에 최소 바코드 서열을 또한 확인할 수 없다. 이는 또한, 추출 서열이 최소 바코드 서열 바로 옆에 가장 잘 위치함, 즉, 최소 바코드 서열 바로 앞 및 바로 뒤에 위치함을 의미한다.

[0254] 따라서, 너무 긴 추출 서열은 증폭 및 시퀀싱 오차, 및/또는 시퀀싱된 뉴클레오타이드의 불량한 품질 때문에 최소 바코드 서열의 식별을 놓칠 것이다. 너무 짧은 추출 서열은 위양성 최소 바코드 서열의 확인을 초래할 것이다. 결국, 둘 다의 곤란(pitfall) 사이, 따라서 사용되는 추출 뉴클레오타이드 서열의 길이에서 균형이 확인되어야 한다.

[0255] 좌측 불변 추출 서열과 우측 불변 추출 서열에서/사이에서 최소 뉴클레오타이드 바코드 서열의 위치에 의해 이러한 바코드 서열을 확인하기 보다는, 당업자는 또한, 오로지 1개의 불변 추출 서열을 사용할 수 있을 것이다. 업스트림 추출 서열이 사용되는 경우, 이러한 추출 서열의 다운스트림에서 주어진 길이가 최소 뉴클레오타이드 바코드 서열을 확인하는 데 사용되고, 다운스트림 추출 서열이 사용되는 경우, 이러한 추출 서열의 업스트림에서 주어진 길이가 최소 뉴클레오타이드 바코드 서열을 확인하는 데 사용된다. 사실상, 최소 뉴클레오타이드 바코드 서열에 대한 7개 뉴클레오타이드 길이 업스트림 추출 서열 및 7개 뉴클레오타이드 길이 다운스트림 추출 서열의 상기 예에서, 최소 뉴클레오타이드 바코드 서열의 확인을 위해, 조사 하의 해당 시료 내의 다른 핵산 유래의 시퀀싱된 서열의 비특이적인 분석과 비교하여 뉴클레오타이드 바코드 서열 유래의 시퀀싱된 서열의 특이적인 분석의 동일한 엄격성/정확성은 14개 뉴클레오타이드 길이($1/4^{14}$ = 1/268435456)의 오로지 1개(업스트림 또는 다운스트림)의 추출 서열을 사용한 경우 수득된다.

[0256] 심지어 동일한 양/농도의 이전성 분자 식별 바코드가 사용되는 경우에도, 선형화가 수행되는 장소가 또한, 주어진 시퀀싱 실험에서 더 많은 최소 뉴클레오타이드 바코드 서열을 회수하기 위한 수행되는 수단이 될 것이다. 예를 들어, DNA 단편의 말단은 내부 단편보다 초음파처리에 의한 단편화에 덜 취약하다. 표적 시퀀싱 주형이 포착에 의해 제조되고 포착 올리고가 단편의 내부에 위치하는 경우, 포착 올리고가 향하는 뉴클레오타이드 및 이의 측면 서열에서의 판독 깊이는 가우스 분포(Gaussian distribution)를 가질 것이다. 표적 시퀀싱 주형이 포착에 의해 제조되고 포착 올리고가 이러한 단편의 말단에 위치하는 경우(예를 들어 제한 효소에 의한 선형화 때문에), 포착 올리고가 향하는 뉴클레오타이드 및 이의 측면 서열에서의 판독 깊이는 가우스 분포를 갖지 않을 것이다. 따라서, 훨씬 더 높은 판독 깊이가 선형화 포인트로부터 출발하여 수득될 것이고, 최소 뉴클레오타이드 바코드 서열이 이러한 출발 포인트에 근접하게 위치하는 경우, 동일한 양의 뉴클레오타이드 바코드 및 따라서 동일한 양의 이전성 분자 식별 바코드에 대해 더 높은 수의 최소 뉴클레오타이드 서열이 수득될 것이다.

[0257] 바람직하게는, 이전성 분자 식별 바코드 내의 상이한 최소 뉴클레오타이드 바코드 서열들은 동일한 길이(예를 들어 25개 뉴클레오타이드 길이)를 가진다. 모든 최소 뉴클레오타이드 바코드 서열들이 동일한 길이를 갖는 경우, 이들의 다운스트림 (생물)정보학적 프로세싱은 더 단순하다. 그러나, 이전성 분자 식별 바코드 내의 최소 뉴클레오타이드 바코드 서열은 상이한 길이를 가질 수 있으며, 이는, 보다 복잡한 적용뿐만 아니라 보다 복잡한 다운스트림 프로세싱을 허용할 수 있다. 이전성 분자 식별 바코드가 상이한 길이의 최소 뉴클레오타이드 바코드 서열을 함유하지만 이들 중 가장 작은 것이 바코드의 상이한 최소 분자 식별을 무제한 수로 생성하기에 여전히 충분히 긴 경우, 덜 복잡한 (생물)정보학적 프로세싱이 여전히, 더 긴 최소 분자 바코드 내의 뉴클레오타이드를 무시함으로써 사용될 것이다. 예를 들어, 이전성 분자 식별 바코드가 25nt 내지 30nt 길이의 최소 뉴클레오타이

드 바코드를 함유하는 경우, 당업자는 25개 뉴클레오타이드만(따라서, 25nt 최소 뉴클레오타이드 바코드 서열 내의 모든 뉴클레오타이드들 및 30nt 최소 뉴클레오타이드 바코드 서열 내의 처음 25개 뉴클레오타이드만) 분석할 것이다. 동일한 길이 및 상이한 길이의 최소 뉴클레오타이드 바코드 서열들을 갖는 뉴클레오타이드 바코드들의 혼합물이 사용되는 경우, 더 긴 길이의 최소 뉴클레오타이드 바코드 서열이 (생물)정보학적 수단에 의해 다듬어져서 제거되며; 이때부터 적어도 2개의 최소 뉴클레오타이드 바코드 서열들이 동일한 길이만 가지며, (병행) 시퀀싱이 사용되어, 동일한 길이를 갖는 최소 뉴클레오타이드 바코드 서열을 구별하고 식별할 수 있다.

[0258] 이전성 분자 식별 바코드의 생성을 위한 2개의 상이한 최소 뉴클레오타이드 바코드 서열들은, 이들 서열이 1개의 뉴클레오타이드 위치에서 뉴클레오타이드에 대해 상이할 때, 이론적으로 상이한 서열이다. 그러나, DNA 서열의 특징화에 사용되는 대부분의 DNA 합성, 증폭 및 시퀀싱 기술은 오차에 취약하다. 고도 병행 시퀀싱 기술은 0.1%-15% 이하의 시퀀싱 오차율을 가진다. 이전성 분자 식별 바코드의 주어진 로트가 뉴클레오타이드 바코드로부터 구축되며 이러한 바코드에서 예를 들어 2개의 최소 뉴클레오타이드 바코드 서열들이 1개의 뉴클레오타이드 위치에서 뉴클레오타이드에 대해 상이한 경우, 이들 뉴클레오타이드 바코드 서열들 중 하나의 서열이 뉴클레오타이드 위치에서의 시퀀싱 오차 때문에 다른 뉴클레오타이드 바코드 서열처럼 잘못 유형화될 수 있으며, 이에 대해 2개의 최소 뉴클레오타이드 바코드 서열들은 상이하여 다른 최소 뉴클레오타이드 바코드 서열의 서열이 잘못 결론내려진다.

[0259] 따라서, 이전성 분자 식별 바코드의 주어진 로트의 생성에 사용되는 상이한 뉴클레오타이드 바코드들은 이들의 최소 뉴클레오타이드 바코드 서열에서 충분히 상이할 필요가 있으며, 따라서 1개의 뉴클레오타이드 바코드 서열이 증폭 및/또는 시퀀싱 오차 때문에 또 다른 뉴클레오타이드 바코드 서열로 전환되는 일은 거의 일어나지 않는다. 단일 가닥 올리고뉴클레오타이드 또는 이의 이중 가닥 구축물이 그와 같이 상기 기재된 바와 같이 사용되는 경우, 이러한 원거리의 무관한(unrelated) 서열은 각각의 올리고뉴클레오타이드의 실제 합성 전에 개시로부터 우측에 있도록 디자인되어야 한다. 예를 들어 뉴클레오타이드 바코드 플라스미드가 사용되는 경우, 충분히 원거리의 무관한 서열을 갖는 뉴클레오타이드 바코드 플라스미드는 모든 가능한 뉴클레오타이드 바코드 서열들을 운반하는 플라스미드의 라이브러리로부터 선택되어야 한다.

[0260] 따라서, 예를 들어 1×10^6개의 이전성 분자 식별 바코드의 로트를 생성하는 데 60개의 뉴클레오타이드 바코드 플라스미드가 필요한 경우, 충분히 상이한 서열을 갖는 최상의 60개 플라스미드가 선택될 수 있도록 선택할 수 있기 위해 60개 초과의 플라스미드가 실제로 필요할 것이다.

[0261] 가장 상이한 서열의 선택은 예를 들어, 계통발생(phylogenetic) 분석에 의해 수득될 수 있다(문헌[De Bruyn et al., 2014]). 계통발생 트리(tree)는, 서열 세트 사이에서 이들의 물리적 또는 유전적 특징에서의 유사성 및 차이를 기반으로 추론된 관계를 보여주는 분지도(branching diagram) 또는 트리이다. 이러한 계통발생 트리로부터, 가장 큰 유전자 변화를 갖거나 가장 높은 유전자 거리를 갖는 서열이 선택될 수 있다.

[0262] 충분히 상이한 뉴클레오타이드 바코드 서열을 선택하는 것 외에도, 증폭으로부터의 오차 및/또는 시퀀싱 오차가 보정될 수 있는 서열이 선택될 수 있을 것이다. 이는, 오차 보정 알고리즘 및 코드의 사용을 통해 달성될 수 있다. 2개의 인기있는 오차-보정 코드 세트는 해밍(Hamming) 코드(문헌[Hamady et al., 2008]) 및 레벤시테인(Levenshtein) 코드(문헌[Buschmann and Bystrykh, 2013])이다.

[0263] 따라서, 뉴클레오타이드 바코드 서열은 고도 병행 시퀀싱에 의해 식별될 수 있다. 이전성 분자 식별 바코드가 6개 이하의 뉴클레오타이드 바코드 서열로 구축되는 경우, 시퀀싱은 이들 6개 뉴클레오타이드 바코드 서열을 보여줄 뿐만 아니라, 증폭 및/또는 시퀀싱 오차 때문에 이들 6개 뉴클레오타이드 바코드 서열로부터 벗어나는 다른 서열을 보여줘야 한다. 각각의 서열이 확인되는 횟수가 (예를 들어 히스토그램에서) 확인되는 경우, 참(true) 6개 뉴클레오타이드 바코드 서열이, 시퀀싱 오차를 갖는 서열보다 훨씬 더 높은 빈도로 확인될 것이다. 빈도에 대한 역치 수준은 뉴클레오타이드 바코드 서열을 보유하거나 폐기하도록 설정될 수 있다. 시퀀싱 오차를 갖는 서열은 역치 수준보다 낮은 빈도로 확인될 것이며, 따라서 이들 서열은 뉴클레오타이드 바코드의 확인을 위한 추가의 분석을 위해 무시되고 폐기되며, 따라서 실제로 존재하는 이전성 분자 식별 바코드만 결론내려질 것이다.

[0264] 이용 가능한 서열들 중에서, 작은 분획의 서열들은 이들 서열이 증폭 및/또는 시퀀싱 오차를 운반하기 때문에 분석에서 상실된다. 그러나, 고도 병행 시퀀싱으로부터의 서열 판독 수의 결과는 방대하여, 증폭 및 시퀀싱 오차 없이 여전히 충분한 서열 판독이 수득될 것이며, 따라서 실제 뉴클레오타이드 및 이전성 분자 식별 바코드가 여전히 확인될 수 있다.

[0265] 이전성 분자 식별 바코드의 로트를 생성하기 위한 뉴클레오타이드 바코드 서열이 오차-보정될 수 있게 선택된 경우, 시퀀싱 오차를 갖는 서열은 오차-보정 알고리즘에 의해 가능하게는 보정될 수 있으며, 따라서 증폭 및/또는 시퀀싱 오차 때문에 작은 서열 상실이 존재할 뿐이다.

[0266] 그러나, 주어진 로트의 이전성 분자 식별 바코드의 생성을 위한 뉴클레오타이드 바코드 서열이 오차 보정을 허용할 방법에 의해 선택되지 않더라도, 증폭 및/또는 시퀀싱 오차를 갖는 서열은 여전히 분석에서 회수될 수 있다. 예를 들어, 서열이 낮은 수로 관찰되지만 이웃 서열을 높은 수로 갖는 경우, 낮은 수의 서열이 추정된 돌연변이율을 기반으로 높은 수의 서열의 증폭 및/또는 시퀀싱 오차를 통해 발생하는 경향이 대체로이며, 덜 풍부한 서열의 계수는 더 높은 풍부한 이웃 서열 덕분이며, 전환되고 계수될 수 있다(문헌[Akmaev 및 Wang, 2004]).

[0267] 주어진 로트의 이전성 분자 식별 바코드의 생성에 사용되는 뉴클레오타이드 바코드 플라스미드는 생성 후 선택적으로 선형화되고, 실제 적용에서 선형화된 형태로 사용된다. 이러한 방식으로, 플라스미드의 우연한 복제가 방지된다. 더욱이, 선형 뉴클레오타이드 바코드 서열의 이전성 분자 식별 바코드 구축물의 사용은 원형화된, 가능하게는 초나선(supercoiled) 뉴클레오타이드 바코드 서열에 걸쳐 시료 바코드의 적용 시 보다 효율적일 수 있다. 이후에 생성되는 뉴클레오타이드 바코드 혼합물로 구성되는 대량의 이전성 분자 식별 바코드들을 선형화하기 보다는, 제한된 수의 뉴클레오타이드 바코드 플라스미드들을 이전성 분자 식별 바코드에 혼합하기 전에 이들 플라스미드를 선형화하는 것이 보다 경제적이다. 뉴클레오타이드 바코드 플라스미드의 선형화는, 플라스미드 내에서, 그러나 최소 뉴클레오타이드 바코드 서열 및 이의 (부분적) 측면 뉴클레오타이드의 외부에서 하나 이상의 부위에서 절단하는 하나 이상의 제한 효소를 이용한 분해에 의해 수득될 수 있다(도 4). 플라스미드가 오로지, 플라스미드 내의 오로지 1개의 부위에서 절단하는 1개의 제한 효소에 의해 절단될 경우, 절단에 실패한 미량의 플라스미드는 원형으로 남아 있고 존재한다. 이는 하나 이상의 제한 효소를 사용함으로써 방지되며, 따라서, 1개 초과의 부위가 플라스미드 내에서 절단된다. 그런 다음, 모든 부위들이 단일 플라스미드 내에서 절단에 실패하는 기회가 고도로 발생하지 않는다. 바람직하게는, 제한 효소(들)는 분해 후, 열-불활성화된다. 사실상, 이전성 분자 식별 바코드가 유전자 검사에 사용되고 게놈 DNA와 혼합되는 경우, 여전히 활성인 제한 효소가 게놈 DNA를 여전히 절단할 수 있을 것이다. 뉴클레오타이드 바코드 플라스미드가 1개 초과의 부위에서 절단되는 경우, 상이한 선형 단편들이 수득될 것이며, 이들 단편들 중에서 오로지 1개의 단편만 실제 최소 뉴클레오타이드 바코드 서열을 함유한다. 뉴클레오타이드 바코드 플라스미드의 모든 분해 단편들은 적용에서 그와 같이 사용될 수 있다. 대안적으로, 최소 뉴클레오타이드 바코드 서열을 함유하는 오로지 분해된 단편만 단리되고 적용에 사용될 것이다. 바람직하게는, 뉴클레오타이드 바코드 분자의 DNA 농도가 분해 전 또는 후에 확인되며, 따라서 뉴클레오타이드 바코드가 동일한 농도까지 정규화될 수 있으며, 이로써 상이한 뉴클레오타이드 바코드들이 보다 등몰 수준으로 확인되는 이전성 분자 식별 바코드가 생성된다. 제한 효소는, 실제 최소 뉴클레오타이드 바코드가 뉴클레오타이드 바코드 플라스미드의 분해 후에 큰 크기, 중간 크기 또는 작은 크기의 단편에 위치하도록 선택될 수 있다. 예를 들어, 이전성 분자 식별 바코드가, 유전자 검사에 사용되기 위해 게놈 DNA가 단리될 표지 혈액 시료에 사용될 경우, 실제 최소 뉴클레오타이드 바코드 서열은 바람직하게는, 더 큰 DNA 단편에 위치할 수 있다. 사실상, 이러한 생물학적 시료의 게놈은 더 큰 게놈 DNA 단편에서 제시될 것이며, 이는 게놈 DNA 추출 키트를 이용하여 보다 효율적으로 추출될 것이다. 더 작은 단편은, 더 짧은 단편이 DNA 추출 프로토콜에서 제거되는 단편화된 DNA로서 간주되기 때문에 이러한 게놈 DNA 추출을 이용하여 효율적으로 보유되지 않을 것이다. 예컨대 실시간 SMRT® DNA 시퀀싱 및 나노포어 시퀀싱에 의해 더 긴 서열 판독을 생성하는 고도 병행 시퀀싱 기술이 사용되는 경우, 최소 바코드 서열은 바람직하게는 또한, 더 긴 선형화된 단편에 위치하여, 이러한 서열은 이들 각각의 DNA 서열 주형 조제물과 보다 융화성이게 된다. 이전성 분자 식별 바코드가 예를 들어 3염색체성 21에 대한 비-침습적 산전 검사(NIPT)를 위해, 임산부로부터 수득된 혈액 시료의 표지에 사용되는 경우, 태아 순환 DNA가 단리될 것이다. 자유(free) 순환 DNA가 태아 유래의 작은 크기의 단편화된 DNA이기 때문에, 최소 뉴클레오타이드 바코드 서열은 바람직하게는 작은 분해 생성물 상에 위치할 수 있으며, 이들 서열은 태아 DNA와 함께 추출될 수 있다. 순환 종양 DNA의 검사에도 동일하게 고려될 필요가 있다. 또한, 포르말린-고정 파라핀 포매된 조직(FFPE)으로부터 수득된 DNA는 보다 작은 증폭 가능한 크기의 DNA 단편을 함유하며, 따라서, 이러한 적용에서, 최소 뉴클레오타이드 바코드 서열은 또한 바람직하게는 작은 분해 생성물 상에 위치할 수 있으며, 이로써 이들 서열은 FFPE DNA와 함께 추출될 수 있다. 따라서, 이전성 분자 식별 바코드의 생성에 사용되는 뉴클레오타이드 바코드 플라스미드는 적용을 기반으로 상이하게 분해될 수 있으며, 따라서 주어진 로트의 이전성 분자 식별 바코드가 주어진 적용 세트에 대해 생성되며, 심지어 1개의 주어진 적용에 대해서만 생성된다. 도 5는, 분자 뉴클레오타이드 서열이 어떻게 NGS 시퀀싱을 위한 주형으로서 제조되는지 그 예를 보여준다.

[0268] 따라서, 이전성 분자 식별 바코드의 중요한 적용은, 유전자 검사가 수행될 생물학적 시료의 표지이다. 시료는

하나 이상의 생물학적 물질, 예를 들어 혈액, 혈장, 혈청, 소변, 침, 가래, 대변, 점막 분비물, 눈물, 윤활액(synovial fluid), 뇌척수액, 복막액 또는 다른 유체로부터 유래될 수 있거나 이들의 초기 형태로 존재할 수 있다. 시료는 세포를 포함할 수 있거나, 세포가 없을 수 있다. 이전성 분자 식별 바코드는 시료 바뀜 및 오염을 방지하기 위해 유전자 검사의 절대 품질 보증을 허용한다. 보다 구체적으로, 시료 바뀜 및 오염은 방지되지만, 이전성 분자 식별 바코드는 이들이 발생할 경우 이들 바뀜 및 오염을 검출할 수 있게 한다. 바람직하게는, 이전성 분자 식별 바코드는 가능한 한 일찍 유전자 검사 프로세스에 첨가된다. 이전성 분자 식별 바코드 서열은 완료 프로세스의 종료 시 확인되어야 하고, 이는 품질 보증을 보장한다. 이러한 방식으로, 전체 프로세스는 이전성 분자 식별 바코드가 첨가된 순간부터 품질이 보증된다. 가능한 가장 빠른 순간은, 분석 표본이 수합되는 때일 것이며, 예를 들어 이전성 분자 식별 바코드는 이미 수합기 튜브(예를 들어 혈액 시료용 수합 튜브) 내에 존재한다. 품질 보증 외에도, 이러한 이전성 분자 식별 바코드는 실험의 자동화 및 보고 프로토콜을 허용한다. 사실상, 이전성 분자 식별 바코드를 함유하는 아이템 상의 물리적 바코드가 시료의 모든 정보가 저장된 LIMS 시스템에 연결되어 있는 경우, 시퀀싱 장비는 조사 하에 표적 뉴클레오타이드 산(acid)에서의 돌연변이를 판독할 뿐만 아니라, 이러한 돌연변이를 갖고 있는 환자의 성명 및 연관된 이전성 분자 식별 바코드를 시료에 첨가한다.

[0269] 이러한 방법은, 시료를, 생물학적 시료가 채취된 장소에서부터 시료 분석이 발생할 임상 실험실(예를 들어 생물학적 시료를 채취한 장소로부터 적어도 100 m, 1000 m, 10,000 m, 100,000 m 위치한 곳)까지 이전성 분자 식별 바코드와 접촉된 채로 운송하는 단계를 포함할 수 있다. 이러한 방식으로, 소비자가 여전히 자신의 시료의 수득된 데이터를 완전히 관리하기 때문에, 실제 검사는 임의의 문제점 없이 핵심 실험실까지 외부 위탁될 수 있다. 소비자가 이러한 핵심 실험실의 시퀀싱 데이터를 수령하는 경우, 소비자는 어떤 뉴클레오타이드 바코드 서열이 해당 시료의 시퀀싱 데이터에 존재해야 하는지 알고 있으며, 따라서 소비자는 자신이, 외부 위탁한 또는 핵심 실험실에 보낸 자신의 시료로부터 데이터를 수령함을 확신한다. 임의의 다른 뉴클레오타이드 바코드 서열의 발견은, 시퀀싱 데이터가 자신의 시료로부터 나온 것이 아니고/거나 시퀀싱 데이터가 또 다른 시료의 서열로 오염됨을 결론내릴 것이다.

[0270] 심지어 DNA 시료가 시퀀싱으로 옮겨지기 전에, 이러한 시료는 이미 또 다른 시료의 DNA에 의해 오염될 수 있다. 217개의 완전 게놈을 시퀀싱한 연구에서, 7개의 시료(3.2%)가 오염 DNA를 함유하는 것으로 확인되었다(문헌[Taylor et al., 2015]). 총 DNA 중 주어진 분획의 DNA, 예컨대 순환 태아 또는 종양 DNA에서 돌연변이를 찾고자 하는 검사에서, 이러한 오염은 검사 결과를 부정적으로 방해할 것이며, 심지어 잘못된 검사 결과를 초래할 것이다. DNA 추출은 종종 자동화된 시스템, 예컨대 Chemagic™ 장비(PerkinElmer), QIAcube 장비(Qiagen)에서 수행된다. DNA 추출 동안, 시료 및/또는 프로세싱 튜브는 특정 순간에 개방되어 있거나, 심지어 완전 프로세스 전 기간에 걸쳐 이러한 DNA 추출 시스템에서 개방되어 있다. 용액은 이들 튜브 사이에서 이전되고 튜브 내에서 교반되며, 시료들 사이에서 오염시킬 수 있는 에어로졸을 발생시킬 것이다. 이들 모든 프로세스들은 환경으로부터 단리된, 심지어 사용중이지 않은 시스템의 작은 챔버에 국한된다. 이러한 시스템이 원심분리를 이용하는 경우, 에어로졸은 국한된 챔버 내에서 균일하게 확산된다. 따라서, 챔버의 국한된 부피를 고려하면, 오염성 에어로졸의 분획은 개방된 실험실에서보다 훨씬 더 높아질 수 있다. 따라서, 수년에 걸쳐 이들 장비의 연속 사용은, 새로 프로세싱된 시료를 오염시킬 국한된 챔버 내에서 오염성 DNA의 점진적인 구축을 초래할 수 있다. 따라서, 이전성 분자 식별 바코드의 사용은 당업자가, 추출된 DNA 시료를 정량화하고 이의 품질에 점수를 매길 수 있게 할 것이다. 따라서 특히, 작은 분획의 표적 DNA, 예컨대 순환 태아 또는 종양 DNA를 분석하는 검사에서, 이전성 분자 식별 바코드는 결국, 검사된 DNA의 품질 점수, 및 따라서 전체 검사에 대한 품질 점수를 제공할 것이다.

[0271] 품질 관리 또는 보증 적용에 사용되는 이전성 분자 식별 바코드는, 이전성 분자 식별 바코드 그 자체가 최상의 품질 보증 조건 하에 생성될 것을 필요로 한다. 1×10^6개 또는 그보다 많은 독특한 이전성 분자 식별 바코드를 생성하는 데 있어서 단지 소수의 뉴클레오타이드 바코드 서열이 필요하다는 사실을 고려하면, 이들 뉴클레오타이드 바코드 분자를 함유하는 단지 소수의 튜브 및/또는 플레이트가 취급되어야 한다. 더욱이, 이들은, 생성 동안 거시적 페이퍼 1D 또는 2D 바코드 표지로 쉽게 표지될 수 있으며, 이러한 표지는 생성 동안 스캐닝에 의해 식별될 수 있다. 뉴클레오타이드 바코드를 함유하는 이러한 소수의 튜브 및/또는 플레이트로부터 제조된 이전성 분자 식별 바코드와 함께 튜브 또는 임의의 수령 용기의 제조는 보다 기본적인 로봇 시스템을 이용하여 수행될 수 있다. 이러한 완전한 생성 프로세스가 또한, LIMS 시스템에 연결되는 경우, 이전성 분자 식별 바코드의 상이한 수령 튜브 또는 용기들이 매우 낮은 오차 위험성으로 생성될 수 있다.

[0272] 예를 들어 상이한 이전성 분자 식별 바코드를 갖는 1×10^6개의 상이한 수령체들이 제한된 수의 60개 뉴클레오타이드 바코드로부터 생성된다는 사실은 심지어, 생성 후 이전성 분자 식별 바코드의 품질 검사를 허용한다. 사

실상, 1 x 10^6개 이전성 분자 식별 바코드의 각각의 로트에서 1 x 10^6개 이전성 분자 식별 바코드들의 모든 조합들의 생성이 고정된 순서의 예정된 프로세스에서 수행되는 경우, 당업자는 예상된 뉴클레오타이드 바코드가 선택된 튜브 내에 존재하는지 입증하기 위해 완전 로트 또는 배치의 특정 튜브를 선택할 수 있으며, 이후, 모든 60개의 뉴클레오타이드 바코드들이 예상된 1 x 10^6개 상이한 튜브 또는 수령 용기 각각에서 종료되었음을 결론내릴 수 있다.

[0273] 생성 후 이러한 품질 관리는, 단일 뉴클레오타이드 바코드 서열이 이전성 분자 식별 바코드로서 사용되는 경우, 올바른 뉴클레오타이드 바코드 서열이 존재하였음을 확인하기 위해 생성 후의 품질 관리가, 단일 뉴클레오타이드 바코드 서열을 함유하는 모든 튜브가 검사되어야 함을 의미할 것이기 때문에, 즉, 각각의 이전성 분자 식별 바코드 용기가 희생되어야 하고 따라서 더 이상 사용될 수 없음을 의미할 것이기 때문에 가능하지 않을 것이며; 사용되는 경우, 이전성 분자 식별 바코드는 1회 사용되지 않을 것이다. 제한된 수의 뉴클레오타이드 바코드로부터 이전성 분자 식별 바코드를 생성하기 위해 뉴클레오타이드 바코드들의 혼합물을 사용함으로써, 단지 최소 수의 이전성 분자 식별 바코드 튜브 또는 수령 용기가 품질 관리를 위해 희생될 필요가 있다. 1 x 10^6개 상이한 가용성 바코드를 생성하기 위해 60개 뉴클레오타이드 바코드가 사용되는 경우, 각각의 용기가 6개의 뉴클레오타이드 바코드 서열을 함유하기 때문에, 심지어 60개 미만의 이전성 분자 식별 바코드 튜브 또는 수령 용기가 희생될 필요가 있으며, 따라서, 희생되는 1개의 이전성 분자 식별 바코드 용기는 이미 6개 뉴클레오타이드 바코드 서열에 대한 정보를 제공한다. 실제로, 1 x 10^6개 이전성 분자 식별 바코드 튜브 또는 수령 용기의 로트가 한번에 생성되지는 않을 것이지만, 더 작은 수, 예를 들어 1개 생성 라운드 당 10.000 튜브가 생성될 것이다. 따라서, 주어진 수의 이전성 분자 식별 바코드 용기는 각각의 생성 라운드로부터 검사되는 경향이 있으며, 따라서 조합된 모든 생성 라운드에 대한 검사되는 이전성 분자 식별 바코드의 총 수는 실제로 60의 수를 초과할 것이며, 이로써, 생성된 이전성 분자 식별 바코드의 중복성(redundancy)을 증가시키고 따라서 더 양호한 품질 관리 검사를 초래할 것이다. 1 x 10^6개 이전성 분자 식별 바코드 튜브 또는 수령체의 생성을 위해, 100x100개 이전성 분자 식별 바코드의 총 100개 더 작은 로트가 생성되고, 각각의 더 작은 로트에 대해 40개 이전성 분자 식별 바코드가 생성 후 입증되는 경우, 4000개 이전성 분자 식별 바코드가 1 x 10^6개 이전성 분자 식별 바코드의 완전한 로트의 생성 후 입증된다. 이는 여전히, 1 x 10^6개 이전성 분자 식별 바코드의 해당 로트의 생성 프로세스의 매우 완전하고 중복된 품질 관리 프로세스를 수득하기 위해, 생성 후에만 입증되고 희생되는 총 로트의 모든 이전성 분자 식별 바코드들의 작은 분획이다.

[0274] 따라서, 이러한 예에서, 60개 뉴클레오타이드 바코드 세트는, 각각이 6개 뉴클레오타이드 바코드로 구축된 1 x 10^6개 이전성 분자 식별 바코드의 생성을 허용한다. 이는 또한, 60개 뉴클레오타이드 바코드들이 각각 100개의 이전성 분자 식별 바코드에서 확인될 것임을 내포한다. 이는 또한, 5개의 주어진 뉴클레오타이드 바코드들의 임의의 조합이 100개의 이전성 분자 식별 바코드에서 확인될 것이지만 후자의 100개의 이전성 분자 식별 바코드들은 각각 6번째 뉴클레오타이드 바코드에 대해 상이할 것임을 내포한다. 이전성 분자 식별 바코드가 특징화될 필요가 있고 이를 위해 PCR에 의한 증폭이 필요한 경우, 이전성 분자 식별 바코드 내의 주어진 뉴클레오타이드 바코드는 증폭되지 않을 것이며, 따라서 해당 뉴클레오타이드 바코드 유래의 상응하는 표적 뉴클레오타이드 서열(앰플리콘)은 확인되지 않을 것이며; 앰플리콘 드롭아웃(dropout)으로 공지된 현상이다. 상기 예에서, 6번째 뉴클레오타이드 바코드가 증폭 및/또는 시퀀싱에 실패하여 오로지 5개의 뉴클레오타이드 바코드 서열만 특징화될 수 있는 경우, 당업자는, 가능한 100개의 이전성 분자 식별 바코드들 중 하나가 존재하는지 확인할 수 없을 것이다. 이러한 잠재적인 문제점은 모든 뉴클레오타이드 바코드에 적용되며, 즉, 특정 뉴클레오타이드 바코드가 앰플리콘 드롭아웃때문에 검출되는 데 실패한 경우, 100개의 잠재적인 이전성 분자 식별 바코드들이 존재하며, 이에 대해, 어떤 이전성 분자 식별 바코드가 사실상 존재하였는지 결론내릴 수 없다. 증폭 및/또는 시퀀싱 오차의 상기 기재된 관용, 예컨대 오차 보정 툴의 사용은 앰플리콘-드롭아웃에 대해 보정할 수 없다.

[0275] 앰플리콘-드롭아웃 때문에 정확한 이전성 분자 식별 바코드를 특징화하고 식별하는 것의 실패는, 이전성 분자 식별 바코드 혼합물을 생성하기 위해 단일 뉴클레오타이드 바코드 대신에 뉴클레오타이드 바코드 쌍을 이용하여 작업함으로써 회피될 수 있으며, 이는 시스템을 보다 중복적으로 만들 수 있다. 뉴클레오타이드 바코드 쌍 중 하나의 뉴클레오타이드 바코드가 증폭 및/또는 시퀀싱에 실패한 경우, 해당 주어진 쌍의 다른 뉴클레오타이드 바코드가 여전히 특징화되고 식별될 것이다. 주어진 뉴클레오타이드 바코드 쌍의 뉴클레오타이드 바코드 중 하나, 또는 주어진 뉴클레오타이드 바코드 쌍의 뉴클레오타이드 바코드 둘 다는, 당업자가 어떤 뉴클레오타이드 바코드 쌍이 존재하였는지 명백하게 확인할 수 있게 한다. 각각이 6개의 뉴클레오타이드 바코드 쌍(총 12개 뉴

클레오타이드 바코드)으로 구축된 1 x 10^6개 독특한 시료 바 코드의 생성을 위해, 당업자는 1 x 10^6개 상이한 독특한 이전성 분자 식별 바코드의 완전한 로트를 구축하기 위해 120개의 뉴클레오타이드 바코드를 필요로 한다(표 5 참조). 1 x 10^6개 이전성 분자 식별 바코드의 생성을 위해, 120개의 뉴클레오타이드 바코드를 사용하고 제조하는 것(즉, 플라스미드 프렙)은, 60개 뉴클레오타이드 바코드가 제조될 필요가 있는 경우에 필요한 노력과 비교하여 유의한 추가의 노력이 되기 힘들다.

표 5

[0276]

이전성 분자 식별 바코드 당 상이한 뉴클레오타이드 바코드 쌍 (a)	필요한 뉴클레오타이드 바코드 쌍의 수	상이한 이전성 분자 식별 바코드의 수
2	40	100
3	60	1000
4	80	10000
5	100	100000
6	120	1000000
7	140	10000000
8	160	100000000
9	180	1000000000
10	200	10000000000
15	300	1000000000000000
20	400	100000000000000000000
25	500	10000000000000000000000000

[0277] (a) 10개의 뉴클레오타이드 바코드 쌍의 세트가 사용된다; 각각의 뉴클레오타이드 바코드 쌍에 대해, 당업자는 10개의 뉴클레오타이드 바코드 쌍들 사이에서 선택할 수 있다.***표 5.*** *이전성 분자 식별 바코드 당 사용되는 뉴클레오타이드 바코드 쌍의 수의 함수에서, 제조될 수 있는 상이한 이전성 분자 식별 바코드의 수.*

[0278] 뉴클레오타이드 바코드 서열 쌍을 사용할 때의 추가의 이점은, 오염이 훨씬 더 높은, 심지어 완전한 확실성으로 결론내려질 수 있다는 점이다. 사실상, 매우 낮은 오염이 존재하고 확인되는 경우, 뉴클레오타이드 바코드 서열 쌍의 최소 뉴클레오타이드 바코드 서열 둘 다의 확인은 오염의 검출 시 서로 확인시켜 준다.

[0279] 시료 바뀜이 예상치 못한 뉴클레오타이드 바코드를 높은 비율로 초래할 것이고(및 더 높은 비율에서 존재해야 하는 예상된 뉴클레오타이드 바코드의 사라짐) 오염이 전형적으로 예상된 뉴클레오타이드 바코드보다 예상치 못한 뉴클레오타이드 바코드를 훨씬 더 작은 비율로 초래하기 때문에, 정보학적 파이프라인은, 높은 비율로 확인되는 뉴클레오타이드 바코드가 '비-쌍(non-pair) 모드'로 분석되어 앰플리콘 드롭아웃이 검출될 수 있으며 한편 낮은 비율로 확인되는 뉴클레오타이드 바코드가 쌍-모드로 분석되어 뉴클레오타이드 바코드 쌍의 최소 뉴클레오타이드 바코드 서열이 둘 다 오염을 결론내리기 위해 검출될 필요가 있도록 프로그래밍될 수 있을 것이다.

[0280] 뉴클레오타이드 바코드 쌍을 이용하여 작업하기 보다는, 당업자는 시스템을 보다 더 중복성이고 강력하게 만들기 위해, 뉴클레오타이드 바코드 서열들의 3중, 4중 또는 그 이상의 조합을 사용할 것이다.

[0281] 뉴클레오타이드 바코드 플라스미드 프렙이 N_x 무작위 라이브러리로부터 사용되는 경우, 최소 뉴클레오타이드 바코드 서열을 식별하기 위해 뉴클레오타이드 바코드가 먼저 시퀀싱될 필요가 있다. 개별 뉴클레오타이드 바코드의 증폭 및 시퀀싱 단계는 사실상 이미, 증폭 및/또는 시퀀싱되기에 어려울 것이고 이전성 분자 식별 바코드가 적용에서 특징화되는 경우 앰플리콘 드롭아웃을 초래할 수 있을 뉴클레오타이드 바코드를 능가하여 이전성 분자 식별 바코드의 생성에 적합한 뉴클레오타이드 바코드를 수득하기 위해 제1 선택 단계일 것이다.

[0282] 이러한 선택 기준, 심지어 추가의 선택 기준을 통과하는 뉴클레오타이드 바코드가 사용될 수 있을 것이다. 예를 들어, 높은 GC 및/또는 AT 함량을 갖는 것으로 고르고 특징화하는 뉴클레오타이드 바코드 서열은 이전성 분자 식별 바코드의 추가 생성에 포함되지 않을 것이다. 사실상, 이러한 서열은 증폭되기 어렵거나 실패하고 앰플리콘 드롭아웃을 초래하고/거나 시퀀싱되는 데 실패할 수 있는 것으로 공지되어 있다. 주어진 뉴클레오타이드 바코드가 이미 상기 기재된 제1 기준 동안에 증폭 및/또는 시퀀싱 조절 단계를 통과한 한편, 이는, 오로지 1개의 뉴클레오타이드 바코드를 고농도로 함유하는 균질한 DNA 프렙에 존재하였기 때문에 수행될 것이며, 이는 다른 뉴클레오타이드 바코드와 조합되고/거나 더 낮은 농도로 사용되는 적용들에서 실제로 사용될 때 더 이상 그러한

경우가 아니다. 이의 후자의 참 적용 환경에서, 주어진 뉴클레오타이드 바코드는 여전히 증폭 및/시퀀싱되는 데 실패할 것이다.

[0283] 이전성 분자 식별 바코드의 생성을 위한 소정의 뉴클레오타이드 바코드 서열에 대한 다른 선택 기준은 동일한 뉴클레오타이드들의 스트레치(stretch), 예를 들어 동일한 뉴클레오타이드(6, 7 또는 그 이상; 또는 심지어 5, 4, 3 또는 2)들의 열(row)을 운반하는 서열일 것이다. 사실상, 이전성 분자 식별 바코드에서 뉴클레오타이드 바코드의 특징화에 사용될 소정의 고도 병행 시퀀싱 기술, 예컨대 피로시퀀싱 및 이온 반도체 시퀀싱에서, 동일한 뉴클레오타이드들의 스트레치에서 뉴클레오타이드의 정확한 수가 항상 올바르게 확인될 수는 없다. 이러한 서열이 1개 초과의 유형의 서열, 즉 시퀀싱 오차를 갖는 서열을 생성하고 이것이 상기 기재된 바와 같이 보정될 수 있다고 하더라도, 이전성 분자 식별 바코드의 생성을 위한 로트에서 이러한 뉴클레오타이드 바코드를 포함시키지 않는 것이 바람직할 것이며, 이것이, 가장 뉴클레오타이드 바코드를 선택하기 위해 뉴클레오타이드 바코드 풀의 생성을 위해 일부 추가의 뉴클레오타이드 바코드를 골라냄으로써 최소의 노력으로 수행될 수 있다면 그러할 것이다.

[0284] 임의의 이전성 분자 식별 바코드는 아이템 또는 시료의 표지에 오로지 1회 사용될 것이며, 따라서 각각의 아이템 또는 시료는, 수밀(watertight) 품질 보증된 표지 시스템을 수득하기 위해 독특하게 표지된다. 일단 1,000,000개의 이전성 분자 식별 바코드 튜브 또는 수령 용기가 생성되면, 120개의 새로운 뉴클레오타이드 바코드 플라스미드 프렙을 선택하고 제조하는 것은 이전성 분자 식별 바코드의 다음 로트의 생성을 위한 단지 최소의 노력이다.

[0285] 모든 이전성 분자 식별 바코드는 또한, 다운스트림 프로세싱에 사용될 식별 명칭 또는 코드를 얻을 필요가 있으며, 이는 데이터베이스에 저장될 것이다. 이러한 데이터베이스는 선택적으로, 중심 위치에 위치하고, 클라우드-기반으로 접근될 수 있다. 주어진 로트가 1,000,000개의 이전성 분자 식별 바코드를 함유하는 경우, 이러한 데이터베이스는 1 x 10^6개 상이한 열들을 함유할 것이며, 따라서 보다 클 것이다. 매우 가능성 있게, 소프트웨어 프로그램(예컨대 알고리즘)이 다운스트림 프로세싱에 사용될 것이다. 그런 다음, 실제 뉴클레오타이드 바코드 서열 또는 이전성 분자 식별 바코드가 이들 소프트웨어 프로그램을 이용하여 프로세싱되고/거나 이들 데이터베이스에 연결될 필요가 있다. 이러한 목적을 위해 예를 들어, 각각의 로트 넘버에 대해, 120개의 명칭 및 뉴클레오타이드 바코드의 최소 뉴클레오타이드 바코드 서열을 열거하는 파일(예컨대 txt.file, csv.file)이 생성될 수 있을 것이다. 주어진 로트의 모든 1,000,000개의 이전성 분자 식별 바코드들이 동일한 고정된 예정된 순서로 생성될 것이기 때문에, 임의의 이전성 분자 식별 바코드 내의 12개의 뉴클레오타이드 바코드 서열들은, 동일한 예정된 순서로 이전성 분자 식별 바코드의 로트 넘버의 생성에 사용된 뉴클레오타이드 바코드의 120개 명칭 및 최소 뉴클레오타이드 바코드 서열을 열거하고 이의 파일명에서 주어진 로트 넘버를 운반하는 이러한 작은 파일을 사용하여 소프트웨어 알고리즘에 의해 추론될 수 있다. 1,000,000개의 이전성 분자 식별 바코드의 각각에 존재하는 실제 최소 뉴클레오타이드 바코드 서열을 추론할 수 있는 알고리즘과 조합하여 120개의 열들을 포함하는 이러한 더 작은 데이터베이스는, 각각의 열이 1 x 10^6개 이전성 분자 식별 바코드의 각각에 12개 분자 뉴클레오타이드 바코드를 기재하는 1 x 10^6개 열들의 표를 각각의 로트에 대해 생성하고 사용하는 것보다 실용적일 것이다. 1개 로트 당 1개의 표를 사용하는 대신, 모든 로트 넘버들에서 사용되는 모든 최소 뉴클레오타이드 바코드 서열들을 열거하는 1개의 단일 표가 사용될 것이다. 그런 다음, 이러한 단일 표는, 명칭 및 각각의 최소 뉴클레오타이드 바코드 서열을 열거하는 컬럼들 외에 제3의 컬럼을 필요로 하며, 이는, 최소 뉴클레오타이드 바코드 서열들이 각각 사용되는 각각의 로트 넘버를 기재하고 있다. 심지어 추가의 컬럼, 예컨대 불변 측면 서열 내의 어댑터 정보(예를 들어 뉴클레오타이드 바코드 서열 식별자 서열, 추출 서열)가 추가될 것이다. 심지어 로트 넘버의 정보(표에서 여유분의 열 및 여유분의 컬럼을 생성함), 및 가능하게는 추가 정보(표에서 여유분의 컬럼을 생성함)가 단일 표에서 조합되는 경우, 이러한 단일 표는, 모든 이전성 분자 식별 바코드들 및 이들의 연관된 최소 뉴클레오타이드 바코드 서열들을 열거하는 표보다 훨씬 더 작다.

[0286] 당업자가, 예를 들어 가장 상이한 최소 뉴클레오타이드 바코드 서열 및 따라서 가장 무관한 서열을 갖는 120개의 뉴클레오타이드 바코드들을 선택해야 하는 경우, 당업자는 이들로부터 더 많은 수의 뉴클레오타이드 바코드들을 선택할 필요가 있다. 당업자가 선택할 수 있는 최소 뉴클레오타이드 바코드 서열의 수가 클수록, 보다 완벽하게 최적인 무관한 서열이 선택될 수 있다. 그러나, 선택될 최소 뉴클레오타이드 바코드 서열의 풀에 더 많은 서열들이 첨가되는 경우, 가장 무관한 서열의 선택에 있어서 이득(gain)은 감소된다. 각각이 12개 뉴클레오타이드 바코드로 구축된 1 x 10^6개 이전성 분자 식별 바코드의 주어진 로트의 생성을 위해 서열 내에서 가능한 한 상이한 25개 뉴클레오타이드 길이인 120개 최소 뉴클레오타이드 바코드 서열의 선택을 위한, 맨하탄 거리

(Manhattan distance)를 기반으로 하는 시뮬레이션 분석, m-메도이드(m-medoid) 클러스터링은, 약 400개의 뉴클레오타이드 바코드가 선택에 실제적으로 필요할 것이라고 결론내렸다. 선택하기 위해 추가의 최소 뉴클레오타이드 바코드 서열을 풀에 첨가하는 것은, 가장 최적의 무관한 서열을 수득하는 데 있어서 임의의 추가의 개선을 초래하기 어려웠다.

[0287] 120개 뉴클레오타이드 바코드보다는 400개 뉴클레오타이드 바코드의 제조는 여전히 약 1 x 10^6개 독특한 이전성 분자 식별 바코드의 생성을 위한 제한된 노력인데, 왜냐하면 특히 이러한 여유분의 노력이 단지, 1 x 10^6개 이전성 분자 식별 바코드의 바로 처음 로트의 생성에 필요하기 때문이다. 사실상, 1 x 10^6개 이전성 분자 식별 바코드가 생성되면, 1 x 10^6개 이전성 분자 식별 바코드의 다음 로트가 생성될 필요가 있다. 제1 생성 로트의 생성에 사용된 동일한 120개 뉴클레오타이드 바코드들은 1 x 10^6개 시료 바코드의 제2 로트에 대해 생성될 수 있다. 그러나, 이는, 동일한 이전성 분자 식별 바코드가 상이한 로트들에 걸쳐 생성되고 1회 초과로 사용될 것이기 때문에 바람직하지 못하다. 동일한 뉴클레오타이드 바코드가 제2 로트의 생성을 위해 생성되더라도, 이전의 로트에서 생성에 사용된 이들의 프렙은 대체로 고갈되는 경향이 있으며, 따라서 새로운 플라스미드 프렙이 제조될 필요가 있다. 이전성 분자 식별 바코드의 새로운 로트의 생성에는, 120개의 새로운 뉴클레오타이드 바코드들이 바람직하다. 따라서, 박테리아 글리세롤 스탁의 플레이트 콜로니를 수행하고 콜로니 고르기(picking)를 수행하는 추가의 노력은 최소이다. 더욱이, 새로운 뉴클레오타이드 바코드를 이용하여 이전성 분자 식별 바코드의 새로운 로트의 생성을 위해, 400개의 새로운 뉴클레오타이드 바코드가 필요하지는 않을 것이다. 단지 120개의 새로운 뉴클레오타이드 바코드가 필요할 것이며, 이들 바코드는 뉴클레오타이드 바코드의 이전의 로트의 생성을 위해 선택되지 않은 280개의 뉴클레오타이드 바코드 플라스미드에 첨가될 것이다. 이들 280+120개의 뉴클레오타이드 바코드로부터, 다시 120개의 대체로 무관한 뉴클레오타이드 바코드들이 이전성 분자 식별 바코드의 새로운 로트의 생성을 위해 선택될 것이며, 그렇게 되는 등이다.

[0288] 가능하게는 다른 표적 핵산과 조합된 이전성 분자 식별 바코드의 병행 시퀀싱 프로세싱 방법은 절대 품질 관리를 수득하기 위해 조건부이다. 이는, 오로지 1개 유형의 서열을 시퀀싱하는 Sanger 시퀀싱과는 대조적이다. 각각의 뉴클레오타이드 바코드, 및 가능하게는 하나 이상의 유전자의 다른 표적 핵산(들), 예컨대 하나 이상의 엑손(의 각각)의 Sanger 시퀀싱은 별도의 시퀀싱 반응을 필요로 할 것이다. 상이한 단일 Sanger 시퀀싱의 분할은, 특히 많은 이전성 분자 식별 바코드 및/또는 시료들이 동시에 프로세싱되는 경우 올바르게 수행되는 것이 보장되지 않는다. 사실상 결국, 상이한 분할 Sanger 시퀀싱 프로세스 또는 이의 수득된 특징화된 데이터는 다시 조합될 필요가 있다. 각각의 Sanger 시퀀싱 반응이 또한, 시퀀싱 주형의 농화, 예컨대 PCR을 필요로 하기 때문에, 분할 프로세스의 개시는 이미, 실제 시퀀싱보다 완전 프로세스에서 훨씬 더 이른 순간에 출발하여, 심지어 더 많은 단계들이 각각의 분할 프로세스에 관여하고, 이때 바뀜 및 오염 오차가 발생할 수 있다. 프로세스에서 분할이 올바르게 수행되지 않은 경우, 잘못 재조합된 프로세스 및/또는 결과가 수득될 것이다.

[0289] 1 x 10^6개의 상이한 이전성 분자 식별 바코드들이 60개의 상이한 뉴클레오타이드 바코드들로부터 구축되는 본 발명자들의 예에서, 모든 뉴클레오타이드 바코드 서열은 해당 로트의 100개의 이전성 분자 식별 바코드에서 확인될 것이다. 비-병행 시퀀싱 방법이 이들 이전성 분자 식별 바코드의 특징화에 사용되는 경우, 분할 프로세스들 사이에서 바뀜이 결국, 상이한 분할 시퀀싱 반응들의 결과들을 조합한 후, 뉴클레오타이드 바코드 서열 및 따라서 유효 이전성 분자 식별 바코드의 유효 조합을 초래할 수 있지만, 올바르지 않은 이전성 분자 식별 바코드는 초래하지 않을 수 있다.

[0290] 한편, 본원에 기재되고 병행 시퀀싱 프로세싱 방법에 사용되는 이전성 분자 식별 바코드는 여전히, 전체 프로세스에서 소정의 프로세스의 분할을 허용하고, 여전히 절대 품질 관리를 수득한다. 그러나, 분할 프로세스는 다시 (동일한) 병행 프로세싱 특징(예를 들어 멀티플렉스 포맷)을 가지고 공통적으로(최소 바코드 서열) 하나의 특징을 가질 필요가 있다. 예를 들어, Ion AmpliSeq™ Exome 검정법(ThermoFisher Scientific)은 12개의 프라이머 풀에 걸쳐 약 300,000개의 프라이머 쌍들을 사용한다. 몇몇 GeneRead DNAseq 표적화된 검정법(Qiagen)들은 4개의 프라이머 풀에 걸쳐 약 2,000개 초과의 프라이머 쌍들을 사용한다. 각각의 이들 분할 풀들은 여전히 복잡한 멀티플렉스 증폭이고, 따라서 여전히 병행 성질을 가진다. 뉴클레오타이드 바코드의 농화를 위해 이들 각각의 풀에 2개의 프라이머가 첨가되는 경우, 및 6개 뉴클레오타이드 바코드로 구축된 이전성 분자 식별 바코드가 본래 생물학적 시료에 존재한 경우, 단지 모든 예상된 6개의 뉴클레오타이드 바코드 서열들이 결국 확인된다면 검사 결과가 유효하다. 검정법에 대한 주어진 세트의 멀티플렉스 반응에서 하나 이상의 멀티플렉스 반응이 프로세싱되는 상이한 생물학적 시료들 사이에서 바뀌는 경우, 6개 초과의 예상되는 뉴클레오타이드 바코드 서열들이

확인될 것이며, 따라서 검사는 유효하지 않은 것으로 마킹될 수 있다. 4개의 멀티플렉스 증폭의 풀을 사용하는 검정법에서 하나의 멀티플렉스가, 이들의 이전성 분자 식별 바코드에서 임의의 뉴클레오타이드 바코드를 공유하지 않는 상이한 생물학적 시료들 사이에서 바뀐 경우, 예상되는 6개 뉴클레오타이드 바코드들은 모든 수득된 뉴클레오타이드 바코드 서열들 중 약 75%를 나타낼 것이며, 한편, 6개 이하의 추가의 뉴클레오타이드 바코드들은 모든 수득된 뉴클레오타이드 바코드 서열들 중 약 25%를 나타낼 것이다. 증폭 농화 단계 후(시퀀싱 전), 6개 이하의 예상된 뉴클레오타이드 바코드 서열들은 해당 바뀐 풀에서 확인되지 않을 것이지만, 6개 이하의 예상치 못한 뉴클레오타이드 바코드 서열들은 해당 바뀐 풀에서 100%를 나타낼 것이다. 시퀀싱 전에 모든 4개의 풀들이 동일하게 혼합되며, 여기서 3개의 바뀌지 않은 풀들에서 6개의 예상된 뉴클레오타이드 바코드들이 각각 100% 분획으로 존재하기 때문에, 총 6개 이하의 예상치 못한 뉴클레오타이드 바코드들은 결국 약 25%를 나타낼 것이다. 이러한 단일 풀의 바뀜은 총 시료의 보다 전반적인 오염, 예컨대 하나의 시료 중 약 25%가 또 다른 시료에 의해 DNA 오염되는 것으로부터 구별될 수 없다. 해당 경우, 모든 4개의 증폭 농화 풀들에서, 6개 이하의 예상치 못한 뉴클레오타이드 바코드들은 약 25%를 나타낼 것이며, 따라서 또한, 모든 4개의 증폭 농화 풀들이 시퀀싱 전에 조합되는 경우 약 25%를 나타낼 것이다. NGS 검정법이 단지 2개의 풀들을 사용하는 경우, 시료의 이전성 분자 식별 바코드들인 예상된 뉴클레오타이드 바코드 및 예상치 못한 뉴클레오타이드 바코드 둘 다 내에서 임의의 뉴클레오타이드 바코드를 공유하지 않는 2개의 생물학적 시료들 사이에서 단일 풀의 바뀜은 모든 뉴클레오타이드 바코드 서열 판독의 약 50%를 나타낼 것이다. 주어진 전체 검정법에 사용되는 풀이 많을수록, 오염 검출에 대해 덜 민감해지고, 가능하게는 분자 뉴클레오타이드 바코드가 오염 검출에 대해 절대 품질 관리를 더이상 보장하지 않는 정도까지 민감해진다. 이는, 비례적인 더 높은 농화, 및 따라서 다른 핵산 표적 서열에 걸쳐 뉴클레오타이드 바코드의 상대적인 더 깊은 시퀀싱에 의해 극복될 수 있다.

[0291] 이는 또한, 표적 농화를 위해 상이한 풀들을 사용하는 검정법에 대해, 각각의 풀에 2개의 동일하지 않은 프라이머들이 뉴클레오타이드 바코드의 농화를 위해 첨가되는 경우 극복될 수 있다. 실제 최소 뉴클레오타이드 바코드 서열들은 불변 서열의 측면에 존재한다. 멀티플렉스 증폭 검정법에서, 주어진 풀에 첨가되는 1개 또는 2개의 프라이머를 측면 불변 서열 내의 상이한 결합 부위까지 표적화함으로써, 각각의 풀로부터 유래된 각각의 뉴클레오타이드 바코드 서열은 이의 주어진 특징, 주어진 상이한 길이의 측면 서열 및/또는 서열 내용을 가질 것이다(도 6). 4개의 프라이머 농화 풀들을 사용하는 NGS 검사의 종료 시 주어진 6개의 뉴클레오타이드 바코드 서열들이 확인되는 경우, 및 모든 예상되는 4개 유형의 측면 서열들이 이들 모든 뉴클레오타이드 바코드 서열들에서 확인되는 경우, 예상되는 6개 뉴클레오타이드 바코드들이 사실상 4개의 분할 풀들 각각에 존재하였으며, 따라서 상이한 시료들 사이에서 단일 풀들의 바뀜이 또한 배제될 수 있다고 결론내릴수 있다.

[0292] 주어진 풀에서 사용되는 2개의 프라이머들 중 하나는 상이한 풀들 사이에서 공유될 것이다. 다른 풀들로부터 구별될 수 있는 주어진 풀 내에서 농화될 조합된 측면 서열의 정도를 결정할 것은 주어진 프라이머들의 조합이다.

[0293] 이러한 포맷은 표적 농화를 위해 프라이머 풀을 사용하는 NGS 검정법에서 보다 민감한 오염 검출을 허용할 뿐만 아니라, 총 DNA 시료에서 전반적인 오염으로부터 상이한 시료들 사이에서 분할 검사 프로세스에서 하나 이상의 풀들의 바뀜을 구별할 수 있게 할 것이다.

[0294] 이전성 분자 식별 바코드는 고체 기판 또는 용기, 예컨대 의료 및 법의학적 적용에서 시료 수합에 사용되는 키트의 수합 기판에 첨가될 수 있다. 또는 보다 양호하게는, 이전성 분자 식별 바코드는 이미 이러한 수합 용기 내에서 생성된다. 이전성 분자 식별 바코드는 핵산 서열의 수령에 적합한 키트의 요소에 직접적으로 첨가될 수 있다. 이러한 요소는 일반적으로, 분석되는 미공지 DNA 시료를 수령하기도 할 요소와 동일하거나 유사하다. 이전성 분자 식별 바코드는 수용액, 분말, 젤, 수지, 라미네이트, 스프레이로서 적용되거나, 캡슐, 제올라이트 내에 포획된 형태, 또는 임의의 다른 적합한 형태로 적용될 수 있다. 당연하게도, 뉴클레오타이드 바코드가 수용액으로서 적용되지 않는 경우, 이러한 바코드는 주어진 순간에 가용성이 아니다. 그러나, 이들 바코드는 프로세싱 동안에 다시 가용성 및 이전성으로 된다. 이전성 분자 식별 바코드는 또한, 수합 용기의 벽 상에 코팅되거나 스팟(spot)될 수 있거나, 면봉 또는 키트의 다른 요소 내로 함침될 수 있다. 예를 들어, 이전성 분자 식별 바코드는 용기, 예컨대 혈액 수합에 사용되는 Vacutainer®(Becton Dickinson), 마이크로튜브, 마이크로타이터플레이트 내의 웰에 첨가될 수 있다. 이전성 분자 식별 바코드는 또한, 카드, 예컨대 Whatman PLC사에 의해 제작된 FTA™ 클래식 카드 상에 스팟될 수 있다. 이러한 유형의 키트는 FTA™ 페이퍼를 포함하며, 이러한 페이퍼에 이전성 분자 식별 바코드가 제작 동안에 첨가되거나, 시료 수합에 사용되는 경우 후속적으로 첨가될 수 있다. 유사하게는, 이러한 바코드는 Guthrie 카드 상에 스팟될 수 있다. 이 단락에서 언급된 기판 및 용기는 "담체(들)"로 지칭된다.

[0295] 이전성 분자 식별 바코드는 시료 제조, 저장 또는 프로세싱에 사용되는 작용제 또는 프로세스, 예컨대 (하나의) 안정화제(들), 보존제(들), 세제(들), 중화제(들), 저해제(들), 환원제(들), ?칭제(들) 또는 이들의 조합을 함유하는 (하나의) 용액(들)과 조합될 수 있다. 이들 요소는 생물학적 시료, 예컨대 이의 핵산과의 직접적인(1차) 상호작용, 또는 이러한 1차 상호작용의 결과인 이차 생성물, 또는 심지어 더욱 다운스트림 상호작용, 예컨대 2차 또는 3차 상호작용 때문에 생성되는 (하나의) 요소(들)와의 상호작용을 통해 이들의 효과를 가질 것이다.

[0296] 예를 들어, 이러한 화합물은 전혈 세포의 응고를 방지하는 헤파린, 에틸렌디아민 테트라아세트산(EDTA), 시트레이트, 옥살레이트, 헤파린 및 이들의 임의의 조합으로 이루어진 군으로부터 선택되는 항응고제일 수 있을 것이며, 이는 백혈구 세포 집단으로부터의 DNA 방출을 감소시키는 것으로 생각된다.

[0297] 예를 들어, 이러한 화합물은 디에틸 피로카르보네이트, 에탄올, 아우린트리카르복실산(ATA), 포름아미드, 바나딜-리보뉴클레오사이드 복합체, 마칼로이드(macaloid), 2-아미노-2-하이드록시메틸-프로판-1,3-디올(TRIS), 에틸렌디아민 테트라아세트산(EDTA), 프로테이나제 K, 헤파린, 하이드록실아민-산소-큐프릭 이온(cupric ion), 벤토나이트, 암모늄 설페이트, 디티오트레이톨(DTT), 베타-머캅토에탄올, 시스테인, 디티오에리트리톨, 트리스(2-카르복시에틸) 포스펜 하이드로클로라이드, 2가 양이온, 예컨대 Mg^{+2}, Mn^{+2}, Zn^{+2}, Fe^{+2}, Ca^{+2}, Cu^{+2}, 및 이들의 임의의 조합으로 이루어진 군으로부터 선택되는 뉴클레아제 저해제일 수 있을 것이다.

[0298] 예를 들어, 이러한 화합물은 포름알데하이드 방출제 보존제이며, 예컨대 디아졸리디닐 우레아, 이미다졸리디닐 우레아, 디메토일롤-5,5-디메틸히단토인, 디메틸올 우레아, 2-브로모-2.-니트로프로판-1,3-디올, 옥사졸리딘, 소듐 하이드록시메틸 글리시네이트, 5-하이드록시메톡시메틸-1-1아자-3,7-디옥사비사이클로[3.3.0]옥탄, 5-하이드록시메틸-1-1 아자-3,7디옥사비사이클로[3.3.0]옥탄, 5-하이드록시폴리[메틸렌옥시]메틸-1-1아자-3,7디옥사비사이클로[3.3.0]옥탄, 4차 아다만틴 및 이들의 임의의 조합으로 이루어진 군으로부터 선택되는 것이 사용될 수 있다(US5459073). 포름알데하이드는 종종 세포막을 안정화시키는 데 사용되고, 따라서 이의 사용은 세포 용해를 감소시킬 수 있을 것이다. 포름알데하이드는 또한, DNase 및 RNase를 저해함으로써, 세포-무함유 핵산의 보존 및 안정성을 증가시키는 것으로 생각되어 왔다.

[0299] 예를 들어, ?칭 화합물은, 포름알데하이드의 전자 부족 작용기와 반응할 수 있는 적어도 하나의 작용기를 포함하는 화합물(예를 들어, 포름알데하이드와 반응하여 메틸올 및/또는 이민 시프(Schiff) 염기를 형성하는 아민 화합물, 또는 포름알데하이드와 반응하여 환형 아세탈을 형성하는 cis-디올 화합물)일 수 있을 것이다. 이러한 ?칭 화합물은 아미노산, 알킬 아민, 폴리아민, 1차 아민, 2차 아민, 암모늄염 또는 이들의 조합으로부터 선택되는 성분일 수 있을 것이다. 보다 구체적으로 글리신, 라이신, 에틸렌 디아민, 아르기닌, 우레아, 아디닌, 구아닌, 시토신, 티민, 스페르미딘(spermidine) 또는 이들의 임의의 조합으로부터 선택된다. 이러한 ?칭 화합물은 임의의 자유(free) 포름알데하이드를 제거하는 데 유용하다.

[0300] 이전성 분자 식별 바코드는 보존제, 항응고제 및 ?칭 화합물의 조합에 첨가될 수 있다.

[0301] 본원에서 또한, 이전성 분자 식별 바코드는 이미, 안정화 용액을 함유하는 수합 용기 내에서 생성될 수 있다. 예를 들어, 이전성 분자 식별 바코드는 혈액 시료의 수합에 사용되며 시료의 유핵 혈액 세포가 안정화되고 순환 태아 또는 종양 DNA를 분석하는 검사에 사용되는 세포-무함유 DNA BCT® 혈액 수합 튜브(Streck 또는 CFGenome), PAXgene 혈액 ccfDNA 튜브(PreAnalytiX), RNA가 안정화되는 PAXgene 혈액 RNA 수합 튜브(PreAnalytiX), 침의 DNA가 안정화되는 Oragene-DNA 침 수합 시스템(DNA Genotek Inc.) 등에서 에서 생성될 수 있다.

[0302] 이전성 분자 식별 바코드와 함께, 시료 바뀜 및/또는 오염을 검출하는 것 이외의 품질 목적을 위한 다른 분자가 존재할 것이다. 이전성 분자 식별 바코드를 함유하는 담체는 또한, 예를 들어 NIPT 검사에서 3염색체성을 검사하기 위한 3염색체성 대조군(예를 들어 CFGenome사의 Trizo21), 예를 들어 RNA 검사에 사용되는 스파이크-인(Spike-in) RNA 변이체(SIRV) 대조군(Lexogen)를 함유할 수 있다.

[0303] 이전성 분자 식별 바코드를 함유하는 용기 또는 수령체는 바람직하게는, 실제 이전성 분자 식별 바코드 및 따라서 또한 이의 뉴클레오타이드 바코드 서열에 일대일로 연결된 코드가 부착된 독특한 광학적 거시적 1D 또는 2D 바코드 페이퍼 표지를 가진다. 실질적인 이유에서, 예를 들어 12개의 25-뉴클레오타이드 길이 서열의 실제 서열은 더 작은 페이퍼 표지 상에서 인쇄될 수 없다. 이러한 실질적인 이유와는 별도로, 적어도 프로세스의 소정의 기(phase) 동안 시료를 프로세싱하는 당업자는 이러한 정보로부터 맹검이며, 따라서 이전성 분자 식별 바코드가 편견 없는 프로세스에 사용되는 것이 흥미로울 것이다. 이러한 포맷은 시료의 조작 및 이의 프로세스에 어느 정도의 여지를 남겨 둔다. 단지 프로세스의 바로 끝부분에서, 뉴클레오타이드 바코드의 실제 서열은 프로세스의

해석, 유효화에 필요하다.

[0304] 부착된 광학적 거시적 바코드 표지는 상이한 포맷을 가질 수 있을 것이다. 이러한 표지는 예를 들어 관통될 수 있으며, 제거될 수 있으며, 2개의 (부분적으로) 동일한 파트들에서 인쇄될 수 있으며, 이들 파트 중 하나의 파트는 제거 가능하다. 그런 다음, 제거될 수 있는 표지는 시료가 수합될 때 환자의 시료 용기 또는 수령체 상에 놓일 수 있으며, 이러한 경우, 시료 용기는 여전히 이전성 분자 식별 바코드를 함유하지 않은 채 사용된다.

[0305] 당업자는, 이전성 분자 식별 바코드의 실제 서열의 정보가 생성 장소로부터 적용(소비자) 장소까지 이전되는 상이한 프로토콜을 생각할 수 있을 것이다. 가능한 방식은, 소비자가 어느 순간, 이전성 분자 식별 바코드를 또한 생성하는 엔터티(entity)에 위치하는 가능성이 큰 클라우드를 통해 서버에 접속해야 하는 방식이다. 바람직하게는, 이전성 분자 식별 바코드, 및 따라서 연관된 뉴클레오타이드 바코드 서열은 예를 들어 OWASP(오픈 소스 웹 어플리케이션 보안 프로젝트; The Open Web Application Security Project) 권고사항에 따라 데이터 전송 동안 클라우드에서 보호된다. 가능하게는, 이들은 또한, 데이터 프라이버시 및 의료 정보를 산업보호하기 위한 보안 규정을 제공하는 HIPAA(의료보험의 이동성 및 신뢰성에 관한 법률; Health Insurance Portability and Accountability Act) 법령을 준수해야 한다. 그러나, 상기 기재된 이전성 분자 식별 바코드 시스템의 특성을 고려하면, 이러한 이전성 분자 식별 바코드 서버가 정보 또는 국소 소비자(예를 들어 병원) 코드 또는 정보에서 임의의 환자를 필요로 하지 않기 때문에 환자 프라이버시 정보는 이미 보장됨을 주지해야 한다. 예를 들어, 소비자 장소에서 수득된 오로지 수득된 뉴클레오타이드 바코드 서열, 뿐만 아니라 소비자가 수령한 페이퍼 식별 코드는 중심 뉴클레오타이드 바코드 데이터베이스에서의 입증을 위해 소비자 장소로부터 이전성 분자 식별 바코드 생성 및/또는 관리 장소 또는 회사까지 이전되며, 이후 이전성 분자 식별 바코드 생성 및/또는 관리 장소는, 페이퍼 코드와 수득된 뉴클레오타이드 서열 판독 사이에 매치가 존재하거나 존재하지 않는 경우 정보를 소비자에게 다시 보낸다.

[0306] 또는 수득된 페이퍼 식별 코드는 소비자에 의해 이전성 분자 식별 바코드 생성 및/또는 관리 장소 또는 회사로 보내지고, 이후 생성 장소는 예를 들어 해당 페이퍼 식별 코드와 연관된 실제 최소 뉴클레오타이드 바코드 서열을 소비자에게 보내고, 여기서, 매치의 존재 또는 부재의 확인이 소비자 장소에서 수행된다. 따라서, 환자 정보, 소비자 장소에서의 환자 코드, 및 이전성 분자 식별 바코드와의 연관성은 소비자 장소, 예컨대 병원에서 독점적으로 남아 있다.

[0307] 클라우드 서버로부터 이러한 정보를 수득하는 것은, 자유롭거나 제약될 수 있을 것이다. 튜브 상의 광학적 거시적 바코드 표지는 서버에게 올바른 이전성 분자 식별 바코드 서열을 제공하기에 충분한 정보를 제공한다. 이러한 코드는 이미, 이러한 클라우드 서버에의 접근을 수득하는 데 충분할 수 있을 것이다. 그러나, 추가의 코드, 즉, 모든 이전성 분자 식별 바코드에 독특한 코드는 접근을 얻기 위해 소비자에게 주어지고 필요할 것이며, 따라서 이전성 분자 식별 바코드의 실제 서열을 수득하기 위한 추가의 보안 수준이 제자리에 존재한다. 사실상, 이전성 분자 식별 바코드의 모든 로트가 제한된 수의 뉴클레오타이드 바코드 서열로 구축되며, 이러한 바코드 서열은 각각의 로트에 대해 동일한 고정된 예정된 프로토콜에서 생성되고 코딩되기 때문에, 로트로부터 주어진 수의 이전성 분자 식별 바코드 및 연관된 코드명이 사용되고 따라서 주어진 소비자에게 공지된 경우, 해당 주어진 로트의 임의의 이전성 분자 식별 바코드의 서열을 확인하는 것이 가능할 것이다. 이러한 추가의 코드를 이용하여, 소비자 요청은 사실상, 주어진 거시적 바코드 표지 및 이전성 분자 식별 바코드(조합)의 홀더 및 소유자에 의해 요청된 바와 같이 식별될 수 있다.

[0308] 요청 시 목록 상에서 이전성 분자 식별 바코드의 뉴클레오타이드 바코드의 서열의 정보를 하나씩 수득하기 보다는, 또 다른 옵션은 이전성 분자 식별 바코드가 기원하는 해당 주어진 로트의 생성에 사용된 상이한 서열의 뉴클레오타이드 바코드들 및 이들의 연관된 명칭을 모두 클라우드 서버로부터 수득하는 것일 것이다. 사실상, 각각의 이전성 분자 식별 바코드를 함유하는 용기의 명명을 포함하여 이전성 분자 식별 바코드의 각각의 로트가 고정된 예정된 프로토콜 내에서 고정된 순서로 소수의 개별 뉴클레오타이드 바코드 분자로 구축된다는 사실을 고려하면, 가능하게는 소프트웨어 패키지 내에서 또는 이의 일부에서 이러한 예정된 프로토콜에 맞춤화된 알고리즘이 또한, 사용자에게 제공될 수 있을 것이다. 그런 다음, 이러한 알고리즘은, 주어진 로트에 대한 뉴클레오타이드 바코드의 서열들 및 이들의 연결된 명칭들을 함유하는 파일(예를 들어 csv 파일 또는 txt 파일)과 함께, 소비자에 의해 주어진 이전성 분자 식별 바코드의 뉴클레오타이드 바코드의 실제 서열을 수득하는 데에 사용될 수 있다. 그러나, 소비자는 또한, 해당하는 동일한 로트의 모든 이전성 분자 식별 바코드들의 뉴클레오타이드 바코드의 서열을 추론할 수 있을 것이다. 당연하게도, 해당 시료에서 사용되지 않은 추가의 뉴클레오타이드 바코드 또는 이전성 분자 식별 바코드의 서열을 제공하는 것은, 당업자라면 누구나, 이전성 분자 식별 바코드를 함유하는 주어진 용기에 존재하는 서열을 해당 이전성 분자 식별 바코드 용기에 부착된 광학적 거시적 바코드

표지를 기반으로 쉽게 추론할 수 있기 때문에, 보안이 낮은 포맷 플랫폼을 제공한다.

[0309] 따라서, 상이한 소프트웨어 툴들이 이전성 분자 식별 바코드의 사용을 위해, 소비자에 의해 사용되는 독자적인 툴로서 개발될 수 있거나, 클라우드 서버 툴의 일부이며, 여기서, 단일 이전성 분자 식별 바코드의 정보는 단독으로 수득되며, 주어진 이전성 분자 식별 바코드 로트의 일부 또는 모든 이전성 분자 식별 바코드들의 정보가 수득되거나, 또는 심지어 이전성 분자 식별 바코드의 모든 로트들의 모든 이전성 분자 식별 바코드의 정보가 수득된다. 주어진 시료에서 실제로 사용되지 않는 이전성 분자 식별 바코드의 정보를 갖는 것은 사실상, 소정의 적용에서 흥미로울 것이다. 이러한 적용의 일례는 오염의 검출이다. 12개의 바코드 서열로 구축된 이전성 분자 식별 바코드가 사용되는 경우, 추가의 바코드 서열의 식별은 오염을 검출하고 보호하기 위해 소프트웨어 툴을 허용할 것이다. 시료가, 예를 들어 12개의 뉴클레오타이드 바코드 서열 중 임의의 서열을 공유하지 않는 또 다른 시료에 의해 오염되는 경우, 및 오염이 50%보다 작거나 큰 경우, 각각의 이전성 분자 식별 바코드에 속하는 추가의 바코드가 확인될 수 있다. 예를 들어, 추가의 바코드가 다른 뉴클레오타이드 바코드와 비교하여 약 1%의 빈도로 확인되는 경우, 이들은 오염성 이전성 분자 식별 바코드에 속할 가능성이 매우 높을 것이다. 따라서, 뉴클레오타이드 바코드가 관찰되는 비율은 오염도에 대한 추가 정보를 제공한다. 실제로, 2개의 상이한 이전성 분자 식별 바코드들은 일부 뉴클레오타이드 바코드를 공유할 것이다. 사실상, 2개의 상이한 이전성 분자 식별 바코드들이 1개의 최소 바코드 서열을 제외하고는 모두 공유하는 경우, 이들 바코드들은 독특하고 상이하다. 그런 다음, 오염은 공유되지 않은 최소 바코드 서열을 통해서만 검출된다. 따라서, 소비자는, 이들의 시료의 프로세싱 및 검정법을 개선하기 위해 오염의 출처에 관심을 가질 것이며, 따라서 이러한 오염은 향후 더 이상 발생하지 않는다. 또 다른 소프트웨어 툴은, 서열이 상이한 로트의 일부인 것으로 관찰되더라도, 오염된 서열에 대한 정보를 제공할 수 있을 것이다. 이러한 목적을 위해, 소비자는 해당하는 각각의 시료에 사용되지 않거나 아마도 해당 소비자의 임의의 시료에 사용되지 않는 이전성 분자 식별 바코드의 서열을 수득할 수 있어야 하고, 따라서 오염의 외부 출처가 관여한다.

[0310] 아이템 또는 생물학적 시료 용기는 단지, 시료의 외부에서 표지, 예를 들어 법의학적 분석 목적을 위한 혈액 수합기 튜브 또는 도핑 분석 목적을 위한 소변 수합기 용기 상, 수합기 용기의 외벽 상의 페이퍼 표지에 의해 표지되는 경우, 표지는 쉽게 조작될 수 있는데, 예컨대 표지가 제거되거나 심지어 또 다른 표지로 대체될 수 있다. 이는, 시료를 제공하는 사람의 일부에게, 오차 또는 악의가 이들의 시료를 잘못 취급할 수 있을 것이며 따라서 사기 또는 범죄 행위에 연루시킬 수 있을 것이라는 염려를 초래하였다. 법의학적 분석을 위한 생물학적 시료 내에서 DNA가 공여자의 개별적인 식별로서 역할을 하는 한편, 시료가 수득된 방식 및 시기에 대해서는 언급되지 않는다. 이러한 '내부'-표지는 법의학, 친자 확인, 도핑 검사 등에서 매우 중요할 것이며, 이후 표지의 조작은 훨씬 더 어렵거나 심지어 더이상 가능하지 않ㄴ다. 심지어 '내부'-표지는 검사 시점 그 자체에서 확인될 필요는 없으며, 예컨대 도핑 검사에서 실제 도핑 검사는 '내부'-표지를 검출할 검사와 상이한 검사이고, '내부'-표지는 오로지 의심이 있는 경우에만 확인된다. 시료 검사 결과에 의문이 생기는 경우, 실제 검사보다 훨씬 더 이후에, 예컨대 법정에 가기 전에 수행되었으며, 시료가 취해지고 운송된 방식 및/또는 시기에 관해, 당업자는 '내부 표지'의 아이덴터티를 다시 검사할 수 있을 것이다.

[0311] 선택적인 2개의 상이한 분자 식별 바코드들은 심지어 동일한 순간에 첨가되어, 심지어 추가의 보안 조치를 제공할 수 있다. 예를 들어, 수령체에 존재하고 시료를 단리하는 파티(party)를 나타내는 분자 식별 바코드, 및 수령체 내에서 생물학적 시료를 제공한 사람을 나타내는 여유분의 분자 식별 바코드가 사용된다. 이는, 분자 식별 바코드가 도핑 검사에 사용되는 시료에 첨가되는 경우 크게 흥미로울 수 있을 것이며, 여기서, 남/여 운동선수들이 또한, 자신이 시료를 제공하는 순간에 자신이 자신의 단리된 생물학적 시료에 첨가한 분자 식별 바코드(예를 들어 튜브 내의 용액으로서)를 소유한다. 이러한 방식으로, 이전성 분자 식별 바코드는, 남/여 운동선수가 자신의 취해진 생물학적 시료에 자신의 서명을 사인하고 따라서 승인하는 (분자) 서명 툴로서 사용되고, 이는 분자 식별 바코드의 형태의 서명이다.

[0312] 상이한 유형의 정보, 또는 상이한 유형의 정보들의 조합이 이전성 분자 식별 바코드에 배정될 수 있다: 예를 들어, 개체, 환자, 의사, 소비자, 병원, 제공자의 정보; 전화번호, 이메일 주소, 공간적 정보(이전성 분자 식별 바코드가 아이템에 첨가된 위치), 시간(이전성 분자 식별 바코드가 아이쳄에 첨가된 때), 시료 유형(혈액, 침, 소변, 대변), 검사 순서, 검사 식별.

[0313] 이전성 분자 식별 바코드는 또한, 표적 핵산의 분석에 사용되는 프로세싱 파이프라인, 예컨대 DNA 또는 RNA 기반 검사를 위한 생물정보학적 프로세싱 파이프라인에 배정될 수 있고, 심지어 자동 (생물정보학적) 프로토콜을 유도할 수 있다. 전체 게놈의 생물정보학적 분석은 심지어 고용량 하드웨어를 이용하더라도, 1개 시료 당 CPU 시간(시(hour))이 소요되고, 따라서 매우 시간 소모적이다. 더 작고 더 신속한 시퀀싱 시스템의 개발에 의해,

전체 게놈이 아니라 단지 표적 영역, 예컨대 단일 유전자 또는 유전자 세트가 분석되는 많은 적용들이 존재한다. 더 작은 표적 영역의 생물정보학적 분석은 요구사항이 더 적다. 그러나, 전형적인 진핵 게놈에는 적어도 20.000개의 상이한 유전자들이 존재한다. 따라서, 당업자는 수득된 시퀀싱 판독이 분석되어야 하는 유전자가 어떤 유전자인지 명시할 필요가 있다. 이전성 분자 식별 바코드가 이들 각각의 유전자에 배정되는 경우, 시스템이 또한, 분석될 필요가 있는 표적 DNA, 예컨대 유전자 또는 유전자 세트의 정보를 판독할 것이기 때문에 생물정보학 파이프라인은 더욱 더 자동화된다. 필수적으로, DNA 이전성 분자 식별 바코드는 당업자가 DNAware라고 지칭할 수 있을 만한 것으로 되며, 이는 시퀀싱 하드웨어 및 소프트웨어 외에도 필요한 것이다. 특히, NGS 시퀀서가 더욱 소형화되고 결국 시퀀싱을 수행하기 위해 노트북 또는 스마트폰 내로 저장될 필요가 있는 작은 하드웨어로 되는 경우, 임의의 첨가된 층, 예컨대 DNAware는 시퀀싱 검정법을 더욱 자동화시키고 접근 가능하게 만든다.

[0314] 상이한 이전성 분자 식별 바코드는 상이한 기능을 가질 것이며, 단일 검정법에 사용될 것이고 따라서 조합될 것이다. 예를 들어, 시료 바뀜 및 오염을 검출하기 위해 시료를 표지하는 이전성 분자 식별 바코드, 및 어떤 생물정보학적 파이프라인이 검정법에 사용되는지에 대한 정보를 제공하는 또 다른 유형의 이전성 분자 식별 바코드가 조합될 수 있다.

[0315] 이전성 분자 식별 바코드가 구축되는 서열을 수득하기 위한 보다 보안된 형태는, 뉴클레오타이드 바코드 분자가 1개의 최소 바코드 서열 대신에 2개 또는 그 이상의 최소 바코드 서열을 함유하는 이전성 분자 식별 바코드의 구축에 사용된 경우, 수득될 수 있을 것이다. 이들 2개의 최소 바코드 서열은 동일한 길이 또는 상이한 길이를 가질 수 있을 것이다. 2개의 독특한 서열을 갖는 이러한 뉴클레오타이드 바코드는 서로 인접해 있을 수 있거나, 링커에 의해 분리되거나 또는 중첩될 수 있을 것이다. 1개의 최소 뉴클레오타이드 바코드 서열을 갖는 뉴클레오타이드 바코드는 대체로, 양쪽 부위에서 불변 어댑터 및 서열의 측면에 존재한다. 이들 어댑터는 결국, 뉴클레오타이드 바코드의 증폭 및/또는 시퀀싱에 필요하다. 따라서, 모든 상이한 뉴클레오타이드 바코드들은 동일한 어댑터를 운반한다. 그런 다음, 이들 어댑터를 갖는 임의의 뉴클레오타이드 바코드의 실제 최소 바코드 서열은 당업자에 의해 쉽게 확인될 수 있다. 이는, 뉴클레오타이드 바코드에서 2개의 독특한 바코드 서열을 사용함으로써 회피될 수 있다. 1개의 바코드 서열은 제2 바코드 서열을 특징화하고 식별하기 위해 증폭 및/또는 시퀀싱 프라이머에 대한 결합 부위로서 사용될 수 있을 것이다. 주어진 이전성 분자 식별 바코드의 제조를 위해, 상이한 뉴클레오타이드 바코드는 동일한 제1 바코드 서열을 운반하며, 한편 제2 바코드는 불변이 아니고, 오로지 1개의 바코드만 운반하는 뉴클레오타이드 바코드에 대해 상기에서 기재된 최소 바코드 서열과 동등하다. 그런 다음, 사용자는, (제2) 최소 바코드 서열의 특징화에 필요한 프라이머를 알기 위해 증폭 및/또는 시퀀싱에 대한 프라이머 결합 부위로서 사용되는 제1 바코드 서열의 정확한 서열을 알 필요가 있다. 이는, 제1 바코드 서열의 서열이 공지되어 있는 경우에만 확인될 수 있다. 그런 다음, 제1 바코드 서열(이러한 경우 프라이머 서열)의 정보는 소정의 사람 및/또는 적용으로 제약되고 확보될 수 있을 것이다. 2개의 최소 바코드 서열을 갖는 이러한 구축물이 플라스미드에서 구축되는 경우, 당업자는 여전히, 예컨대 플라스미드의 벡터 백본 내에서 제1 바코드 서열 영역의 측면에 존재하는 증폭 및/또는 시퀀싱을 위한 프라이머를 구축함으로써 실제 제1 바코드 서열을 여전히 풀어낼(unravel) 것이다. 주어진 안심 제1 바코드 서열 및 상이한 가능한 제2 바코드 서열을 갖는 이러한 뉴클레오타이드 바코드가 제1 바코드 영역 및 제2 바코드 영역 둘 다에서 무작위 서열을 갖는 플라스미드의 혼합물 내에서 숨겨지는 경우, 제1 바코드는 거의 풀어지지 않을 수 있으며, 특히 이들 구축물이 무작위 제1 서열을 갖는 상이한 플라스미드 각각의 농도에서 존재하는 경우에, 심지어 당업자에 의해서가 아니라도, 그러할 수 있다.

[0316] 2개의 독특한 최소 바코드 서열을 갖는 뉴클레오타이드 바코드 서열을 사용하는 또 다른 적용은, 제1의 주어진 독특한 서열이 주어진 적용에 사용되는 한편 또 다른 주어진 제1 서열이 또 다른 적용에 사용되는 적용일 것이다.

[0317] 이전성 분자 식별 바코드는 이들 상이한 수준의 적용, 접근성 및/또는 보안 중 하나를 갖는 적용에서 생성, 제공 및 사용될 수 있을 것이다. 사실상, 상이한 로트들이 생성되는 경우, 상이한 적용, 접근성 및/또는 보안 수준이 주어진 로트에 배정될 수 있을 것이다.

[0319] *아이템, 시료 및/또는 프로세스는, 이전성 분자 식별 바코드가 첨가된 순간부터 완전히 품질-보증된다. 이전성 분자 식별 바코드가 프로세스에서 더 일찍 첨가될수록, 프로세스가 더 일찍 따라서 더 많이 완전히 품질-보증된다. 또 다른 이전성 분자 식별 바코드, 또는 이전성 분자 식별 바코드들의 혼합물이 프로세스의 종료 시 확인되는 경우, 시료 바뀜 및/또는 오염이 전체 프로세스 중 어딘가에 있는 것으로 결론내려진다. 그러나, 시료 바뀜 및/또는 오염이 발생한 장소가 프로세스 중 정확히 어디인지는 알려지지 않을 것이다. 당업자는, 시료 바뀜 및/또는 오염이 발생한 장소가 프로세스 중 정확히 어디인지에 대한 검색을, 상이한 시점들에서 이전성 분자 식별

바코드를 이용하여 프로레싱된 아이템 또는 시료를 다시 스파이킹함으로써 좁힐 수 있을 것이다. 따라서, 당업자는 이전성 분자 식별 바코드를 전체 프로세스 중 2개 또는 심지어 그 이상의 시점들에서 아이템/시료 및 이의 프로세싱에 첨가한다. 예를 들어, 이전성 분자 식별 바코드가 출발 시에 아이템 또는 시료에 첨가되었고 또 다른 이전성 분자 식별 바코드가 해당 프로세스의 중간에 해당 프로세싱된 아이템 또는 시료에 첨가되었으며 올바른 제2 이전성 분자 식별 바코드가 제1 이전성 분자 식별 바코드가 아니라 프로세스의 종료 시에 확인되는 경우, 프로세스의 제2 기에서 바뀜이 발생하지 않은 것으로 결론내릴 수 있다. 전형적인 유전자 검사는 몇몇 서브프로세스들로 구축되며, 각각은 다시 일반적으로 상이한 단계들로 구축된다. 제1 프로세스에서 DNA가 생물학적 시료, 예컨대 혈액으로부터 추출되고, 제2 기에서 DNA 돌연변이의 검색이 표적 영역의 농화 및 시퀀싱에 의해 수행된다. DNA 추출 및 시퀀싱은 종종 실험실에서 상이한 위치들, 또는 심지어 동일한 연구소/회사에서 상이한 실험실들, 또는 심지어 상이한 연구소/회사들에서 상이한 실험실들에서 수행된다. 각각의 상이한 실험실들에서, 상이한 사람들이 관여하고 이들의 책임감을 가진다. 프로세스 중 상이한 단계에서 이전성 분자 식별 바코드를 첨가하는 것은 예컨대 (프로세싱된) 시료가 상이한 실험실에 도달하는 경우, 시료 바뀜 또는 오염을 주어진 서브프로세스 및 실험실까지 추적할 수 있게 한다. 이러한 방식으로, 오차가 발생한 장소, 및 어떤 실험실이 책임이 있으며 예를 들어 새로운 검사를 수행하기 위한 비용을 지불해야 하는지에 대한 논의는 없을 것이다. 이는, 많은 실험실들이 그들의 시퀀싱을 외부위탁하는 NGS 서열 코어 설비/회사들에게 있어 매우 중요할 것이다.

[0320] 이전성 분자 식별 바코드가 6개의 뉴클레오타이드 바코드 쌍으로 구축되는 경우, 12개 뉴클레오타이드 바코드들은 각각 약 8.3%의 분획을 가진다(오염이 존재하지 않는 경우). 2개의 이전성 분자 식별 바코드가 임의의 뉴클레오타이드 바코드 서열을 공유하지 않는 주어진 프로세스에 첨가되는 경우, 24개의 뉴클레오타이드 바코드들은 각각 단지 약 4.15%의 분획에서 확인될 것이다. 더 많은 이전성 분자 식별 바코드들이 주어진 프로세스에 첨가되는 경우, 각각의 뉴클레오타이드 바코드의 분획은 심지어 더 작아진다. 프로세스의 종료 시 확인되는 각각의 뉴클레오타이드 바코드의 분획이 작을수록, 오염 검출에 대해 덜 민감해진다. 불변 측면 서열 중 1개 또는 둘 다가 배치들 사이에서 부분적으로 또는 완전히 상이한, 상이한 배치 생성들로부터의 이전성 분자 식별 바코드의 사용, 예컨대 완전 프로세스 중 각각의 상이한 단계에 첨가되는 1개 배치 당 상이한 뉴클레오타이드 바코드 서열 식별자 서열의 존재가 이러한 문제점을 극복할 것이다. 이러한 경우, 프로세스 중 주어진 시점에 첨가되는 각각의 이전성 분자 식별 바코드에 대해, 각각의 뉴클레오타이드 바코드의 분획은 프로세스의 종료 시 8.3%인 채로 남을 것이다.

[0321] 상이한 불변 측면 서열을 갖는 상이한 이전성 분자 식별 바코드가 주어진 프로세스에서 상이한 단계에서 사용되기보다는, 상이한 불변 측면 서열을 갖는 이러한 상이한 이전성 분자 식별 바코드는 시료에서 동일한 포인트에 첨가될 것이다. 예를 들어, 취해지는 혈액 시료는 상이한 불변 측면 서열을 갖는 상이한 이전성 분자 식별 바코드를 사용하여 상이한 검사에서 사용될 것이다. 이들 각각의 검사는 주어진 불변 측면 서열을 갖는 주어진 이전성 분자 식별 바코드를 사용할 것이다. 이러한 경우, 상이한 불변 측면 서열을 갖는 이전성 분자 식별 바코드들의 혼합물이 사용될 것이다. 예를 들어, 차세대 시퀀싱 실험실들은 이들이 검출하고자 하는 오염 수준에 대해 상이한 버전, 작업 흐름 및 프로토콜을 가질 것이다. 예를 들어, 이러한 정보는, 혈액 시료가 의사에 의해 채혈되는 순간에는 공지되지 않을 수 있다. 예를 들어, 이러한 정보는, 시료의 DNA가 추출되고 품질이 공지될 때까지 공지되지 않을 수 있다. 이러한 경우, 다량으로 존재할 때 뉴클레오타이드 바코드가 매우 민감한 오염 검출을 허용하는 주어진 불변 측면 시퀀싱을 갖는 이전성 분자 식별 바코드, 및 소량으로 사용될 때 뉴클레오타이드 바코드가 덜 민감한 오염 검출을 허용하는 다른 주어진 불변 측면 서열을 갖는 이전성 분자 식별 바코드가 혈액 시료에 첨가된다. 또 다른 예는, 혈액 시료가 DNA 및 RNA 분석에 사용되는 것일 수 있을 것이다. 이러한 경우, DNA-유형 이전성 분자 식별 바코드와 RNA-유형 이전성 분자 식별 바코드의 혼합물이 생물학적 시료, 예컨대 혈액에 첨가될 수 있다.

[0322] 이전성 분자 식별 바코드가 시료에 첨가되는 경우, 뉴클레오타이드 바코드는 해당 (프로세싱된) 시료 내에서 조사 하에 다른 표적 핵산을 압도해서는 안된다. 사실상, 당업자는, 돌연변이 검출에 대한 최고 민감성을 최저 비용으로 수득하기 위해, 표적 핵산의 서열 판독을 가장 높은 수로 수득하고자 한다. 한편, 이전성 분자 식별 바코드 내의 너무 낮은 수의 각각의 뉴클레오타이드 바코드는 당업자가 올바른 이전성 분자 식별 바코드를 검출하거나 추론할 수 없게 할 것이며, 그 이유는 각각의 뉴클레오타이드 바코드가 바람직하게는, 조사 하에 표적 핵산(들)에 등몰량으로 존재해야 한다. 진핵생물 기원의 생물학적 시료는 이배체(2N)이다. 6개 뉴클레오타이드 바코드 쌍으로 구축된 이전성 분자 식별 바코드가 사용되는 경우, 이들 바코드는 표적 핵산에 대해 등몰량이기 위해 실제로 6N에서 존재해야 한다. 혈액 시료가 10 ml Vacutainer® 튜브에서 취해지는 경우, 취해지는 혈액의 부

피가 항상 10 ml인 것은 아니다. 보다 중요하게는, 혈액 시료 내에서 DNA의 출처인 백혈구 세포의 수는 세포에서 1 ml 당 측정되는 경우 개체들 사이에서 다양하다. 또한, 심지어 정확하게 동일한 부피의 혈액이 개체들 사이에서 취해지는 경우, 이용 가능한 DNA 및 추출된 DNA의 양은 결국 시료들 사이에서 크게 다양할 수 있다. 따라서, 이전성 분자 식별 바코드의 하나의 고정된 양이 모든 시료들에 대해 가장 최적의 양은 아닐 것이다. 예를 들어, 혈액 시료의 혈장 내의 순환 DNA의 양은 해당 혈액 시료 유래의 백혈구 세포 내의 DNA의 양보다 훨씬 낮다. 따라서 바람직하게는, 더 낮은 양의 이전성 분자 식별 바코드가, 게놈 DNA의 수합에 사용되는 담체보다 cfDNA의 수합에 사용되는 담체에 첨가된다.

[0323] 시료의 시퀀싱을 위한 대부분의 주형 제조는 포착 올리고를 통한 표적 핵산의 포착에 의한 농화, 또는 올리고뉴클레오타이드를 사용하는 증폭 기술에 의한 농화를 필요로 한다. 그런 다음, 뉴클레오타이드 바코드가 농화될 필요가 있으며, 각각 뉴클레오타이드 바코드에 대한 포착 올리고 또는 뉴클레오타이드 바코드의 증폭에 대한 프라이머에 의해 농화될 필요가 있다. 농화되는 뉴클레오타이드 바코드의 수는 이들 포착 올리고뉴클레오타이드 또는 프라이머를 통해, 또는 보다 구체적으로 이들 포착 올리고뉴클레오타이드 또는 프라이머의 농축을 통해 최적의 양까지 정규화될 수 있다. 이러한 경우, 예를 들어 포착 올리고뉴클레오타이드는 이들의 뉴클레오타이드 바코드 표적보다 낮은 농도로 존재하며, 일단 이들 모든 올리고뉴클레오타이드들이 뉴클레오타이드 바코드를 확인하면, 미반응된 뉴클레오타이드 바코드가 포착되지 않고 세척된다. 이러한 방식으로, 다른 표적 핵산과 비교하여 과량의 뉴클레오타이드 바코드 서열이 제거될 수 있다. 유사하게는, 다른 표적 핵산의 증폭에 대한 프라이머와 비교하여 증폭 농화 프로토콜에서 뉴클레오타이드 바코드 서열을 증폭시키기 위한 프라이머의 양을 제한하는 것은, 전부는 아닌지만 즉 과량의 표적 뉴클레오타이드 바코드가 이들의 프라이머를 확인할 것이고 증폭에 참여하지 않을 것이라는 사실을 초래할 것이다. 이러한 방식으로, 뉴클레오타이드 바코드에 대한 고정된 농도의 올리고뉴클레오타이드는 과량의 뉴클레오타이드 바코드를 걷어낼 것이다.

[0324] 실제로, 표준 포착 또는 증폭 프로토콜에서, 올리고뉴클레오타이드의 농도는 이미, 이들이 포착하거나 증폭해야 하는 표적의 농도보다 훨씬 더 높다. 사실상, 올리고뉴클레오타이드들 및 이들의 각각의 DNA 표적들이 일대일의 분자 관계로 존재하는 것은 아니다. 과량의 뉴클레오타이드 바코드를 걷어내기 위해, 포착 또는 증폭을 위한 이들 뉴클레오타이드 바코드 서열에 대한 올리고뉴클레오타이드의 농도가 다른 표적 뉴클레오타이드 산에 대한 올리고뉴클레오타이드의 농도와 비교하여 감소된다. 그런 다음, 뉴클레오타이드 바코드 서열 대 다른 표적 뉴클레오타이드 서열의 상대적인 균등화는, 농화에 참여하는 표적 뉴클레오타이드 바코드의 수를 제한함으로써 달성되는 것이 아니라, 역학적으로(kinetically) 덜 바람직한 뉴클레오타이드 바코드 서열을 확인함으로써 달성된다.

[0325] 뉴클레오타이드 바코드 서열이 다른 표적 뉴클레오타이드 서열보다 낮은 농도로 존재하는 경우, 존재하는 이전성 분자 식별 바코드를 추론하기 위해 시퀀싱 후 충분하지 않은 뉴클레오타이드 바코드 서열이 수득될 위험이 있어서, 뉴클레오타이드 바코드 서열 대 표적 핵산의 최적 비율은 뉴클레오타이드 바코드 서열의 농화에 사용되는 올리고뉴클레오타이드(포착 올리고뉴클레오타이드 또는 증폭 프라이머)의 농도를 상대적으로 증가시킴으로써 수득될 수 있다.

[0326] 당업자가 오염을 검출하는 데 매우 민감한 시퀀싱 프로토콜을 개발하고자 하는 경우, 이는, 이전성 분자 식별 바코드 농도의 양을 증가시키지 않으면서, 예컨대 더 높은 양(농도)의 이전성 분자 식별 바코드를 함유하는 생물학적 시료에 대한 다른 시료 용기를 사용하여, 뉴클레오타이드 바코드 서열을 포착 또는 증폭에 의해 농화시키기 위해 올리고뉴클레오타이드의 농도를 증가시킴으로써 달성될 수 있다.

[0327] 전체 게놈이 예컨대 전체 게놈 시퀀싱 및 비-침습적인 산전 검사, 원형 종양 DNA에 대한 검사에서 시퀀싱되는 경우, 포착 또는 증폭에 의한 농화를 위해 어떠한 올리고뉴클레오타이드도 첨가되지 않는다. 그렇지만, 뉴클레오타이드 바코드 서열에 대한 포착 올리고뉴클레오타이드가, 전체 게놈 서열에 대한 가능하게는 과량의 뉴클레오타이드 바코드 서열을 제거하는 데 사용될 수 있다.

[0328] 이전성 분자 식별 바코드는 심지어, NGS 주형 제조 검정법에서 이들을 제조하지 않으면서 고도 병행 시퀀싱 적용에 사용될 것이다. 이는, 고도 병행 시퀀싱에 의한 증폭 및/또는 시퀀싱을 위한 모든 프라이머 결합 부위들이 이미 이전성 분자 식별 바코드의 불변 측면 서열에 존재하고 따라서 NGS 주형 제조 동안 부착될 필요가 없는 경우, 달성될 수 있다. 그런 다음, 오로지 최소 주형 제조 검정법, 예컨대 단일 실험에서 상이한 시료들을 분석하기 위한 풀링 지표의 혼입이 수행될 것이다.

[0329] 많은 적용들에서, 특히 표적화된 시퀀싱에서, 상이한 시료/라이브러리가 단일 시퀀싱 실험에서 풀링된다. 이러한 풀링된 라이브러리에서, 라이브러리 대 라이브러리 가변성은 가능한 한 낮아야 한다. 하나의 라이브러리가 작은 농도로 존재하는 경우, 더 적은 양의 시퀀싱 데이터(최악의 경우 불충분한 판독 깊이를 초래함)가 이러한

시료/라이브러리에 대해 수득될 것이며, 따라서 돌연변이가 해당 시료에서 놓치게 될 것이며, 이로써 시료는 가능하게는, 추가의 시퀀싱 판독을 수득하여 요망되는 판독 깊이를 수득하기 위해 다시 시퀀싱될 필요가 있다. 이는 또한, 해당 시료의 시퀀싱 비용을 증가시킬 것이다. 하나의 시료/라이브러리에 대해 너무 많은 시퀀싱 데이터(및 따라서 너무 높은 판독 깊이)가 필요한 것보다 더 수득되는 경우, 더 많은 시퀀싱된 서열이 수득될 것이고 필요하지 않아도 비용이 지불될 것이기 때문에 검사에 더 많은 비용이 든다. 따라서, 풀링된 시료/라이브러리는 균등화되어야 한다. 이러한 균등화는, 시퀀싱되는 DNA 단편의 분자량이 각각의 라이브러리에 대해 고려되거나 추정되는 경우, 농도의 확인, 및 따라서 DNA 분자의 수의 확인을 통해 수행될 수 있다. 상이한 라이브러리가 상이한 표적 핵산(및 따라서 시퀀싱될 매우 가능성 있게 상이한 크기의 DNA 단편)을 갖는 경우, 분자의 수는 표적 크기에 대해 보정되어야 한다. 각각의 라이브러리의 이들 농도 측정, 계산 및 후속적인 희석 실험, 및 후속적으로 이에 맞는 라이브러리의 혼합은 매우 시간 소모적이다.

[0330] 풀링되는 균등화 상이한 시퀀싱 라이브러리는 또한, (하나의) 균등화 올리고뉴클레오타이드(들)에 의해 달성될 수 있다. 이는 단순하며 심리스 비드(seamless bead)-기반의 해결방안으로서, 임의의 차세대 시퀀싱 작업 흐름에 필요한 바와 같은 라이브러리 정규화를 위한 라이브러리 정량화 및 라이브러리 희석에 대한 필요성을 대체하고, 라이브러리-대-라이브러리 가변성을 최소화한다. 균등화 올리고뉴클레오타이드는 모든 표적 뉴클레오타이드 서열들에 결합할 수 있다. 대부분의 적용에서, 균등화 프로브에 대한 이러한 결합 부위는, 프라이머가 균등화 올리고에 대한 이러한 결합 부위와 함께 어댑터를 운반하는 표적 핵산의 농화 동안 증폭에 의해 혼입되거나, 어댑터가 균등화 올리고에 대한 결합 부위를 운반하는 어댑터의 단편화 및 연결에 의해 시퀀싱 라이브러리의 제조 동안 증폭에 의해 혼입된다. 균등화 결합 부위는 증폭 및/또는 시퀀싱 프라이머에 대한 결합 부위와 동일할 것이다. 균등화 서열이 프라이머 또는 연결 어댑터에 존재한 경우, 미반응된 프라이머 또는 연결 어댑터가 먼저 제거될 필요가 있으며, 따라서 오로지 참 시퀀싱 주형만 균등화된다. 시료가 이전성 분자 식별 바코드 유래의 표적 뉴클레오타이드 서열 및 다른 표적 핵산을 함유하는 경우, 이전성 분자 식별 바코드 유래의 표적 뉴클레오타이드 서열의 균등화가 단독으로, 뉴클레오타이드 바코드의 불변 측면 서열에 대한 균등화 올리고뉴클레오타이드에 의해 달성될 수 있다. 균등화 올리고뉴클레오타이드들인, 뉴클레오타이드 바코드 서열 내의 불변 서열에 대한 제1 균등화 올리고뉴클레오타이드와, 프라이머에 의한 증폭 동안 또는 연결 어댑터에 의한 연결 동안 혼입되는 서열에 대한 제2 균등화 올리고뉴클레오타이드의 조합은 조사 하에 뉴클레오타이드 바코드 서열 유래의 표적 뉴클레오타이드 서열 대 다른 표적 핵산 유래의 다른 표적 뉴클레오타이드 서열의 상대적인 비율을 더 미세 조정할 수 있게 한다. 균등화 올리고뉴클레오타이드가 비오틴화되는 경우, 당업자는 이러한 동일한 수의 균등화 올리고뉴클레오타이드(들)를 각각의 시퀀싱 라이브러리에 첨가함으로써 라이브러리-대-라이브러리 가변성이 최소화된 시퀀싱 라이브러리를 수득하기 위한 용이한 비드-기반 해결방안을 수행할 수 있으며, 균등화 올리고뉴클레오타이드의 이러한 수는 수득될 수 있는 표적 뉴클레오타이드 서열의 수보다 낮은 경우, 각각의 시퀀싱 라이브러리 내의 과량의 수의 시퀀싱 주형 분자가 제거될 것이고, 이러한 방식으로, 시퀀싱 실험에서 풀링된 상이한 시퀀싱 라이브러리들이 보다 최적의 방식으로 균등화될 것이다.

[0331] 또한, 옥스포드 나노포어 시퀀싱에서 적용된 바와 같이 실시간에서 리드 언틸 시퀀싱(Read Until Sequencing)은 뉴클레오타이드 바코드 서열 대 다른 표적 뉴클레오타이드 서열의 서열 판독의 가장 최적의 비율을 수득하기 위한 또 다른 방식일 것이다. 리드 언틸 시퀀싱에서, 각각의 포어에서 무엇이 시퀀싱되는지 모니터링되며, 이는 사용자에게 시퀀싱을 계속할지 마칠지에 대한 옵션을 제공한다. 각각의 나노포어 채널이 개별적으로 다뤄질 수 있기 때문에, 각각의 포어에서의 시퀀싱 반응은 단순히 중단될 수 있으며, 이로써 새로운 DNA 가닥이 시퀀싱을 위해 해당 포어에 접근할 수 있다. 각각의 최소 바코드 서열의 측면에 불변 측면 어댑터 서열이 존재하기 때문에, 리드 언틸 시퀀싱은 뉴클레오타이드 바코드 서열의 불변 어댑터 서열의 존재 또는 부재를 모니터링하여, 시퀀싱된 뉴클레오타이드 바코드 서열의 수의 추적을 유지할 수 있다.

[0332] 라이브러리 제조 프로토콜에서 뉴클레오타이드 바코드 서열의 포착 또는 증폭을 위해 올리고뉴클레오타이드의 농도를 조정함으로써, 또는 선택적인 시퀀싱에 의해, 각각이 주어진 적용 또는 기술에 대해 상이한 양의 최적화된 이전성 분자 식별 바코드를 갖는 생물학적 시료의 수합을 위해 너무 많은 용기를 제조할 필요가 없을 것이다. 특이적인 다운스트림 프로세싱에 대해 상이한 농도의 이전성 분자 식별 바코드를 갖는 용기가 여전히 제조될 필요가 있는 경우, 이러한 수는 2개(높음 또는 낮음) 또는 3개(높음, 중간, 낮음) 농도 유형으로 제한되는 것으로 예상된다.

[0333] 따라서, 이전성 분자 식별 바코드는 검출될 필요가 있는 오염 수준에 따라 상이한 양으로 제조되고 사용될 것이다. 사실상, 당업자가 검출하고자 하는 오염 수준이 낮을수록, 필요한 NGS 시퀀싱은 더 깊고, 따라서 또한 아이템 또는 시료에 첨가되어야 했던 뉴클레오타이드 바코드의 양이 많아진다.

[0334] 이전성 분자 식별 바코드가 유전자 검사에서 오염을 검출하는 경우, 이러한 오염은 검사에 사용된 임의의 성분으로부터 기원할 수 있다. 사실상, 생물학적 시료들(검사 프로세스에서 혈액 시료, DNA 시료, 또는 이의 프로세싱된 생성물) 사이에서의 오염 외에도, 오염은 또한, 검사(DNA 추출, NGS 주형 제조 및/또는 시퀀싱 반응)에 사용된 임의의 생성물로부터 기원할 수 있을 것이다. 예를 들어, 상이한 시료들이 풀링 바코드의 사용을 통해 단일 시퀀싱 실험에서 조합되는 경우, 풀링 바코드가 또한, 서로들 사이에서 오염될 수 있을 것이다. 풀링 바코드는 올리고뉴클레오타이드 프라이머의 사용을 통해 시퀀싱 주형에 혼입될 수 있다. 상이한 풀링 바코드 프라이머들의 시리즈가 생성되는 경우, 합성, 가용화, 이후의 풀링 바코드 올리고뉴클레오타이드 동안 선행 풀링 바코드 올리고뉴클레오타이드 유래의 임의의 약간의 잔여물은 오염된 풀링 바코드 프라이머를 초래한다(예를 들어 풀링 바코드 프라이머 2가 풀링 바코드 프라이머 1에 의해 오염되었음). 2개의 상이한 시료들이 하나의 실험에서 시료 1에 대해 풀링 바코드 프라이머 1 및 시료 2에 대해 풀링 바코드 프라이머 2를 사용하여 프로세싱되고 시퀀싱되는 경우, 풀링 바코드 프라이머 2가 풀링 바코드 프라이머 1에 의해 오염되었다. 시퀀싱 후, 2개의 시료들 각각에 대한 상이한 시퀀싱된 서열들을 생물정보학 프로세스에서 각각의 풀링 바코드에 따라 그룹으로 나누고 프로세싱하는 경우, 시료 2의 일부 판독은 시료 1의 판독을 함유하는 파일 내에서 나타날 것이며, 이들 판독이 오염된 풀링 바코드 1과 연관이 있었기 때문이다. 따라서, 분석 후, 시료 1은 프로세싱 동안 시료 2에 의해 오염된 것으로 나타날 것이지만, 두 시료 모두 제조 동안 기술자에 의해 올바르게 프로세싱되고 오염이 시료들 사이에서 발생하지 않은 것이 가능할 것이며, 한편 사실상 풀링 바코드는 오염되었다. 따라서, 이전성 분자 식별 바코드들은 또한, 이들의 생성 프로세스 및 최종 생성물의 품질 관리를 위해 DNA 추출 키트, NGS 주형 제조 키트, 올리고뉴클레오타이드의 제조 회사들에 의해 사용될 것이다.

[0335] 뉴클레오타이드 바코드 서열에 대한 포착 올리고뉴클레오타이드는 또한, DNA 추출이 전체 생물학적 시료 상에서 수행되지 않는 프로토콜에서도 매우 필요할 수 있을 것이다. 본원에서, 생물학적 시료의 세포에 존재하는 게놈 DNA뿐만 아니라 첨가되는 DNA 뉴클레오타이드 바코드인 임의의 DNA가 분석될 것이다. 검정법을 예컨대 (미소유체) 시스템 상에서 소형화하는 경우, DNA 단리는 전체 생물학적 시료 상에서 수행되지 않을 것이다. 예를 들어, DNA 추출은 단리된 세포, 병원체(박테리아, 바이러스) 상에서만 수행될 수 잇다. 이전성 분자 식별 바코드가 이러한 생물학적 시료에 존재하는 경우, 이들 바코드는 세포 단리 단계에서 보유되지 않을 것이다. 뉴클레오타이드 바코드의 포착에 대한, 따라서 포착을 위한 포착 올리고뉴클레오타이드는 병원체-크기 또는 세포-크기의 비드(-구조)에 결합될 수 있으며, 따라서 뉴클레오타이드 바코드가 또한 보유될 것이다. 예를 들어 이러한 세포, 박테리아 또는 바이러스가 자기 비드에 결합된 폴리클로날 항체를 통해 포착되는 경우, 이들 비드는, 뉴클레오타이드 바코드에 대한 포착 올리고뉴클레오타이드가 결합된 자기 비드와 쉽게 혼합될 수 있으며, 따라서 프로토콜 또는 검정법은 이러한 검정법에서 이전성 분자 식별 바코드의 사용을 허용하기 위해 변형될 필요가 거의 없다.

[0336] DNA-유형 이전성 분자 식별 바코드는 또한, RNA가 cDNA로 전환된 순간부터 RNA 검사에 사용될 수 있다. 사실상, 생물학적 시료로부터의 많은 RNA 추출 프로토콜들은 (잔여) DNA를 예를 들어 DNase에 의해 파괴시켜, DNA-유형 이전성 분자 식별 바코드가 생물학적 시료에 존재하는 경우 파괴될 것이며, 따라서 이들은 RNA가 cDNA로 전환된 경우에만 첨가될 수 잇다. RNA 검사에서 RNA를 cDNA로 전환시키기 전에 업스트림 프로세싱 단계의 품질 보증을 위해, RNA-유형 뉴클레오타이드 바코드의 혼합물로 구축된 RNA-유형 이전성 분자 식별 바코드가 사용되어야 한다.

[0337] RNA-유형 뉴클레오타이드 바코드는 심지어, DNA-유형 뉴클레오타이드 바코드에서 최소 뉴클레오타이드 바코드 영역에 대한 업스트림 불변 측면 서열이 프라이밍 RNA 합성을 위한 프로모터를 함유하는 경우, DNA-유형 뉴클레오타이드 바코드로부터 제조될 수 있다. 이러한 어댑터는 예를 들어 T7 프로모터일 수 있을 것이다. RNA-유형 뉴클레오타이드 바코드에서 최소 뉴클레오타이드 바코드 영역에 대한 다운스트림 불변 측면 서열이 A-잔기들(예를 들어 20, 25, 50, 75, 100, 150, 200개 또는 그 이상의 A 잔기들)의 스트레치를 함유하는 경우, RNA-유형 이전성 분자 식별 바코드의 RNA-유형 뉴클레오타이드 바코드가 또한, mRNA 단리를 필요로 하여 RNA-유형 뉴클레오타이드 바코드가 mRNA 단리 단계에서 mRNA와 함께 단리되는 RNA 검사에서 프로세싱될 수 있다.

[0338] 진단 검사에 적용되는 것과는 별개로, 이전성 분자 식별 바코드는 또한, 연구 검사에 사용될 수 있을 것이다. 예를 들어 발현 벡터에 의해 형질감염되고 고도 병행 시퀀싱 검정법에 의해 연구되는 생물학적 시료(예를 들어 세포주)가 사용되는 연구 프로젝트가 사용된다. 이는, 측면 불변 서열이 박테리아 또는 바이러스 DNA(따라서 클로닝 벡터, 또는 보다 구체적으로 클로닝 벡터 백본에서 확인되지 않음) 내에서 인코딩되지 않거나, 또는 임의의 천연 발생 게놈, 박테리아 또는 바이러스 DNA에서 인코딩되는 서열과 1% 미만, 2% 미만, 3% 미만, 4% 미만, 5% 미만, 10% 미만, 15% 미만, 20% 미만, 25% 미만, 30% 미만, 40% 미만, 50% 미만으로 상동성인 서열을 갖는

또 다른 이유이다. 그렇지 않다면,이들은 고도 병행 시퀀싱을 방해하고, 또한 시퀀싱될 것이며, 이로써 당업자가 분석하고자 하는 시퀀싱된 서열을 희생시키고 비용을 들여 필요하지 않은 발현 벡터로부터 시퀀싱된 서열을 생성할 것이며, 그 이유는 당업자가 원하는 수의 시퀀싱된 서열을 수득하기 위해서는 더 많은 시퀀싱이 필요하기 때문이다.

[0339] 게놈 및/또는 전사체 수준에서 검사에 적용하는 것과는 별도로, 이전성 분자 식별 바코드는 또한, 임의의 종의 검사에 적용될 수 있다. 일례는, 식품을 이의 고기, 식물, 박테리아, 진균류 함량에 관하여, 심지어 완전 생성 사슬이라도, 식품의 품질 및 조성을 확인하는 검사이다. 이는 예를 들어, 또 다른 동물의 고기가 존재하며 식품 상에 표시된 주어진 동물의 고기와 혼합된 가짜의 검출을 허용할 것이다. 이는, 종들 사이에서 고도로 다양하고 따라서 이들 사이에서 종의 구별 및 따라서 검출을 허용하는 게놈 및/또는 미토콘드리아 DNA 영역의 서열 분석을 통해 분자 수준에서 수행될 수 있다. 본원에서 또한, 식품 시료가 이러한 분자 시퀀싱 검사를 위해 취해진 경우, 식품 시료들 사이에서 식품 시료 바뀜 및/또는 오염은 이전성 분자 식별 바코드의 사용(첨가)에 의해 유사하게 방지/모니터링될 수 있다.

[0340] 이전성 분자 식별 바코드는 또한, 산업 제품, 당해 기술분야의 연구, 유물, 보안 및 환경적 오염물, 산업 생산 프로세스의 품질 관리 등을 표지하는 데 사용될 수 있다. 더 높은 농도로 사용되는 이전성 분자 식별 바코드는 심지어, 증폭 필요성을 배제할 것이며, 따라서 이전성 분자 식별 바코드는 단일 분자 시퀀싱 장치 상에서 직접적으로 시퀀싱될 수 있을 것이며, 이로써 시퀀싱 주형 제조가 대체로 필요하지 않다. 이러한 포맷은 휴대용의 핸드헬드 단일 분자 시퀀싱 장치에서 매우 매력적일 수 있다. 이러한 이전성 분자 식별 바코드는 용이한 단리 및/또는 정제, 예컨대 비오틴기의 부착을 용이하게 하기 위해 추가의 특성을 가질 것이며, 따라서 이전성 분자 식별 바코드는 (복잡한) 혼합물로부터 쉽게 단리될 수 있다.

[0341] 분자 식별 DNA 바코드는 또한, 조사 하의 단백질(들)을 검출하는 앱타머(aptamer)(들)가 이용 가능한 경우, 단백질 검정법에 사용될 수 있을 것이다.

[0342] **실시예 1. NGS에서 이전성 분자 식별 바코드의 사용**

[0343] 예시를 위해, 도 7은, 조사 하의 핵산의 표적 영역의 농화를 위해 2-단계 PCR 프로토콜을 사용하는 NGS 검정법에서 이전성 분자 식별 바코드가 어떻게 시료 바뀜 또는 오염을 검출할 것인지 보여준다.

[0344] 이전성 분자 식별 바코드를 환자 시료에 첨가하였으며, 이 중 2개의 표적 핵산 영역 A 및 B가 관심 영역이다. 2-단계 멀티플렉스 PCR 프로토콜에 의해, 2개의 표적 핵산 영역 및 최소 뉴클레오타이드 바코드, 및 이전성 분자 식별 바코드의 측면 서열을 증폭시킨다. 이를 위해, 모든 3개의 표적 핵산 영역들의 측면에 존재하는 프라이머 결합 부위에 관한 앰플리콘-특이적인 프라이머를 증폭시킨다. 모든 앰플리콘-특이적인 프라이머들은 5' 유니버셜 어댑터 서열을 운반한다(정방향 프라이머에 대한 하나의 유형의 유니버셜 어댑터 서열 및 역방향 프라이머에 대한 또 다른 유형의 유니버셜 어댑터 서열). 제2 PCR 단계에서, 풀링 바코드를 상기 제1 PCR 단계에서 수득한 모든 앰플리콘들에 혼입시킨다. 1개 시료 당 오로지 1쌍의 상이한 풀링 바코드 프라이머를 사용한다. 풀링 바코드 프라이머는, 제1 PCR 단계 후 3개의 유형의 앰플리콘 내에 혼입된 유니버셜 어댑터 서열 내의 프라이머 결합 부위, 지표 서열, 및 NGS 시퀀싱을 위한 시료의 추가 프로세싱에 필요한 5' 어댑터 서열을 가진다. 각각의 시료에 대해, 상이한 풀링 바코드를 사용한다. 모든 3개의 앰플리콘들(조사 하의 2개의 표적 핵산 및 이전성 분자 식별 바코드로부터 유래됨)에서, 동일한 지표를 주어진 시료에 혼입시킨다. 그런 다음, 모든 상이한 시료들을 혼합하고 시퀀싱하여, 시퀀싱 칩 및 시약의 사용을 완전히 경제적으로 만든다. 시퀀싱 후, 모든 시퀀싱된 서열들을 1개 또는 소수의 큰 파일에서 수득한다. (생물)정보학적 수단에 의해, 모든 시퀀싱된 서열들을 풀링 바코드 서열에 따라 상이한 파일 내에서 분리한다. 따라서, 각각의 환자 시료에 특이적인 동일한 풀링 바코드 서열을 갖는 모든 서열들을 그룹으로 나누고, 별도의 파일에 저장한 다음, 개별적으로 추가로 프로세싱한다. 각각의 지표/환자 파일에 대해, 시료 바뀜 또는 오염이 없는지 알아보기 위해 최소 뉴클레오타이드 바코드 서열을 특징화할 것이다. 시료 바뀜 및/또는 오염이 확인되지 않는 경우, 유전자 검사, 즉 해당 시료 내의 2개의 표적 뉴클레오타이드 서열들에서 확인된 돌연변이는 유효하다.

[0345] 4명의 환자 시료가 나타나 있다. 제1 시료에 이전성 분자 식별 바코드 I을 첨가하였고, 풀링 바코드 1을 사용하였으며; 제2 시료에 이전성 분자 식별 바코드 II를 첨가하였고, 풀링 바코드 2를 사용하였으며; 제3 시료에 이전성 분자 식별 바코드 III을 첨가하였고, 풀링 바코드 3을 사용하였고; 제4 시료에 이전성 분자 식별 바코드 IV를 첨가하였고, 풀링 바코드 4를 사용하였다. 도 7a에서, 시료 바뀜 또는 오염이 관찰되지 않으며, 도 7b에서, 시료 바뀜이 프로세싱 동안 제1 시료와 제4 시료 사이에서의 발생하였다. 도 7c에서, 제4 시료는 프로세싱 동안 제1 시료에 의해 오염된다.

[0346] **실시예 2. 유전자 검사 작업 흐름**

[0347] 도 8은 이전성 분자 식별 바코드를 함유하는 혈액 수합기 튜브로부터 출발하여 최종 유효 유전자 검사 보고서까지, 유전자 검사에서의 작업 흐름을 보여준다,

[0348] 혈액 수합기 튜브는 독특한 이전성 분자 식별 바코드 및 일대일로 연결된 광학적 거시적 바코드 표지를 함유한다. 독특한 분자 식별 바코드 및 광학적 바코드, 및 이들의 일대일 연관성은 데이터베이스에 저장되고, 이러한 데이터베이스는 클라우드를 통해 접근 가능하다. 환자로부터, 혈액을 이러한 혈액 수합기 튜브 내에서 수합한다. 이러한 혈액 수합기 튜브에, 환자(예를 들어 병원에 의해 참여된)로부터 또 다른(제2) 광학적 바코드 페이퍼 표지를 부착하여, 혈액 수합기 튜브에는 이제 2개의 광학적 바코드 페이퍼 표지가 부착되어 있다. 환자의 광학적 바코드 표지를 LIMS 시스템에 연결하는 경우, 및 광학적 바코드 페이퍼 표지를 둘 다 스캐닝하는 경우, 제1 광학적 바코드 표지의 정보 또한 LIMS에 연결되며, 이전성 분자 식별 바코드와의 이의 일대일 연관성 및 또한 이전성 분자 식별 바코드가 고려된다. 따라서, 이전성 분자 식별 바코드 내의 최소 바코드 서열은 환자의 성명의 이전성 가명이다. 그런 다음, 이러한 시료를 시퀀싱을 위해 프로세싱하고, 보다 구체적으로 DNA 추출을 수행하며, 관심 표적 핵산 영역(조사 하의 게놈의 DNA 영역 및 뉴클레오타이드 바코드 서열)을 농화시키고, 농화된 서열을 시퀀싱 주형으로서 제조한다. 전체 게놈 시퀀싱 또는 원형 DNA 시퀀싱을 수행하는 경우, 농화를 수행하지 않는다. 각각의 시료의 서열 주형의 제조 동안, 풀링 바코드를 혼입시킨다. 그런 다음, 모든 시퀀싱 주형들을 풀링하고 시퀀싱한다. 시퀀싱 후, 동일한 풀링 바코드를 갖는 모든 서열들을 단일 파일로 그룹으로 만들고 더 분석한다. 이전성 분자 식별 바코드로부터 유래된 서열 판독을 뉴클레오타이드 바코드 서열 식별자 서열의 존재를 통해 식별할 수 있다. 이들 서열 판독들을 서브그룹으로 나눌 수 있고, 최소 뉴클레오타이드 바코드 서열을 예를 들어 이들의 '추출' 서열을 통해 특징화할 수 있다. 조사 하의 표적 핵산 유래의 다른 서열을 맵핑 및 변이체 콜링을 통해 돌연변이에 대해 특징화할 것이며, 이로써 유전자 검사 결과를 수득한다. 그런 다음, 본래의 제1 광학적 바코드 표지 페이퍼 코드를 제시함으로써 클라우드 데이터베이스에 접속하고, 이후, 클라우드 서버는 연관된 최소 뉴클레오타이드 바코드 서열을 소비자의 실험실에 전송한다. 이들 전송된 최소 뉴클레오타이드 바코드 서열은 서열 판독으로부터 수득된 최소 바코드 서열을 이용하여 입증될 것이다. 예상된 최소 바코드 서열이 서열 판독에서 확인되고 추가의 최소 바코드 서열이 확인되지 않는 경우, 시료 바뀜 또는 오염이 시료의 프로세싱 동안 발생하지 않았으며, 따라서 유전자 검사 결과는 유효하다. 심지어 병원에서 생성된 환자 광학적 페이퍼 표지 코드가 아니라도 환자 정보는 실험실/병원 외부로 이전되지 않고, 오로지, 환자의 혈액이 수합된 본래 혈액 수합기 튜브에 부착된 광학적 바코드 페이퍼 표지 코드만 이전됨을 주지해야 한다.

[0349] 도 9는 이전성 분자 식별 바코드르 함유하는 마이크로튜브로부터 출발하여 최종 유효 유전자 검사 보고서까지 유전자 검사에서의 작업 흐름을 보여준다.

[0350] 마이크로튜브 튜브는 독특한 이전성 분자 식별 바코드 및 일대일로 연관된 광학적 거시적 바코드 표지를 함유한다. 혈액 시료를 표준 진공채혈관 내에서 채혈한다. 채혈 직후, 또는 진공채혈관이 유전자 실험실에 도달한 때에, 마이크로튜브의 이전성 분자 식별 바코드 용액, 및 연관된 거시적 바코드 표지를 혈액 시료를 포함하는 진공채혈관에 이전시킨다. 대안은, 연관된 거시적 바코드 표지가 이전되지 않고 즉시 스캐닝되는 것이다. 그런 다음, 검사를 도 8에 기재된 바와 유사하게 진행한다.

[0351] 도 10은 도 8에 기재된 바와 유사한 작업 흐름을 보여주지만, 여기서 시퀀싱된 서열의 프로세싱은 상이하다. 뉴클레오타이드 바코드 유래의 시퀀싱된 서열 및 표적 시료 핵산의 시퀀싱된 서열을 도 8 및 도 9에 기재된 바와 같이 병행하여 분석하지 않는다. 오로지 뉴클레오타이드 바코드의 시퀀싱된 서열만 더욱이 클라우드 상에서 완전히 분석하고, 시료 바뀜이 검출되지 않는 경우에만, 표적 시료 핵산의 시퀀싱된 서열의 생물정보학적 분석을 개시한다. 또한, 오염이 검출되는 경우, 표적 시료 핵산의 시퀀싱된 서열의 생물정보학적 분석을 개시하고 그 결과를 검출 및 분석하지만, 검출된 오염 수준의 맥락에서 해석한다. 생물정보학적 파이프라인은 상이한 일련의 및/또는 병행 단계들(알고리즘)에서 주문될 수 있고, 최소 뉴클레오타이드 바코드 서열을 제공하는 기관의 클라우드 상에서 또는 클라이언트 장소(예를 들어 병원)에서 부분적으로 또는 완전히 수행될 수 있음이 분명하다.

[0352] **실시예 3. 이전성 분자 식별 바코드를 이용한, 상이한 서브프로세스에서 시료 검사 프로세스의 스파이킹**

[0353] 도 11은, 혈액을 채혈할 때, 이전성 분자 식별 바코드로 표지한 2개의 시료를 보여준다. 혈액 시료 1을 이전성 분자 식별 바코드 1로 표지하고, 혈액 시료 2를 이전성 분자 식별 바코드 2로 표지한다. 두 시료 모두의 DNA를 DNA 추출 설비 내에서 추출한다. 그런 다음, 추출된 DNA 시료를 다시 이전성 분자 식별 바코드로 표지한다. DNA 시료 1을 이전성 분자 식별 바코드 A로 표지하며, DNA 시료 2를 이전성 분자 식별 바코드 B로 표지한다. 그런 다음, DNA 시료를 차세대 시퀀싱 설비로 보낸다. 혈액 또는 DNA의 표지에 각가 사용된 이전성 분자 식별 바코드

들은, 이들이 구별될 수 있는 상이한 뉴클레오타이드 바코드 서열 식별자 서열을 각각 갖는 뉴클레오타이드 바코드로 구축된다. 시료 바뀜 및/또는 오염이 발생한 경우, 시료 바뀜 및/또는 오염이 발생한 서브프로세스(및 실험실)를 추적할 수 있다.

[0354] 시료 바뀜 또는 오염이 시료 1과 시료 2에서/사이에서 발생하지 않은 경우, 최소 뉴클레오타이드 바코드 서열 1 및 A가 시료 1에서 확인될 것이고, 최소 뉴클레오타이드 바코드 서열 2 및 B가 시료 2에서 확인될 것이다.

[0355] 시료 1이 DNA 추출 동안 시료 2에 의해 오염된 경우, 최소 뉴클레오타이드 바코드 서열 1, 2 및 A가 시료 1에서 확인될 것이다.

[0356] 시료 1이 NGS 주형 제조 및 시퀀싱 동안 시료 2에 의해 오염된 경우, 최소 뉴클레오타이드 바코드 서열 1, 2, A 및 B가 시료 1에서 확인될 것이다.

[0357] **실시예 4. 업스트림 불변 서열 및 다운스트림 불변 서열**

[0358] 이 실시예는 본 출원에서 기재되고 예를 들어 도 3 및 도 4에 도시된 바와 같은 단리된 업스트림 불변 서열 및 다운스트림 불변 서열의 실시형태를 보여준다.

[0359] 도 4를 참조로 하여, 하기 도시된 업스트림 불변 "서열 1"[서열 번호: 1]은 제한 부위 RE1(밑줄)으로부터 검정색으로 표시된 최소 바코드 서열까지 위치한 서열을 보여준다.

[0360] 도 4를 참조로 하여, 하기 도시된 다운스트림 "서열 2"[서열 번호: 11]는 검정색으로 표시된 최소 바코드 서열로부터 제한 부위 RE2(밑줄)까지 위치한 서열을 보여준다.

[0361] 서열 번호: 2 및 서열 번호: 12는 폴리 A 꼬리가 없는 변이체이다.

[0362] 이들의 변이체는 벡터 내의 클로닝 부위에 따라 대안적인 제한 부위 인지 서열을 가진다.

[0363] 대안적으로, "서열 1"은 다운스트림 서열이고, "서열 2"는 업스트림 서열이다.

[0364] 대안적으로, 1개 또는 2개의 다운스트림 서열은 하기 도시된 서열의 역방향 상보 서열일 수 있다.

[0365] 대안적인 실시형태에서, 도시된 다운스트림 및/또는 업스트림 서열은 서열 동일성을, 70% 초과, 80% 초과, 90% 초과, 95% 초과, 97% 초과 또는 99% 초과의 서열 동일성을 보여주는 서열이다. 서열 동일성의 차이는 예를 들어, 제한 효소에 대한 인지 부위를 첨가하거나 결실시킨 결과일 수 있다.

[0366] 보다 대안적인 실시형태는 지시된 제한 부위와 불변 서열 사이에서 및/또는 불변 서열과 최소 바코드 서열 사이에서 추가의 뉴클레오타이드 서열의 존재에 의해 하기 도시된 서열을 포함하는 불변 서열이다.

[0367] 보다 다른 실시형태는, 도시된 "서열 1" 및 "서열 2"의 단편, 즉, 적어도 200개 뉴클레오타이드, 적어도 300개 뉴클레오타이드, 적어도 350개 뉴클레오타이드, 적어도 375개 뉴클레오타이드, 또는 적어도 390개 뉴클레오타이드의 단편을 포함하거나 구성된 불변 서열이다.

[0368] 이들 서열 및 대안적인 서열의 비제한적인 세트는 서열 번호: 1 내지 20으로 도시되어 있다. "서열 1" 및 "서열 2"[서열 번호 1 및 1]에 대한 상기 설명은 다른 도시된 서열에 대해서도 동등하게 적용 가능하다.

[서열 번호 :1]

[0369] aagctttgtggatgtacaagtccacaccatgtacactagacgcagcctgtacagatatccatccagtg
tactcactgtcgacacggatccaatgcccgggttctgatagacgaacgacgagatgtgcagtgacttc
gaggatcccagatgtgcacgtagtgcaggtagcttgaatgactactacgcctgtagcatcatcacgta
gactcgtacagctacatgacggtagctagattgacgactcaagcatgctagtgtcgttactgacctga
tgacacagtcgatgcgaccttaatacgactcactatagggtcaacaagaccctgcagatcccgggatc
cgcctcttaagctgcgcaggccaggaattgcacgtcagagcactaaggccgccaccatggc

[서열 번호 : 2]

aagctttgtggatgtacaagtccacaccatgtacactagacgcagcctgtacagatatccatccagtg
tactcactgtcgacacggatccaatgcccgggttctgatagacgaacgacgagatgtgcagtgacttc
gaggatcccagatgtgcacgtagtgcaggtagcttgaatgactactacgcctgtagcatcatcacgta
gactcgtacagctacatgacggtagctagattgacgactcaagcatgctagtgtcgttactgacctga
tgacacagtcgatgcgaccttaaatgctgagagattagggtcaacaagaccctgcagatcccgggatc
cgcctcttaagctgcgcaggccaggaattgcacgtcagagcactaaggccgccaccatggc

[서열 번호 : 3]

aagcttctctcgccagctatttaaagtacgagtcgggaggccttagcacgaactgatttttccagcct
gagtgctgttcttgcatgtaccttctatctaacgacgtccgtaataggaagtataccaggtcgaacta
acgactcctttgccgtagcgagtgtttcgccaaaagtgtctgggtctactggccaccgtccagcattt
ctatgcccgtaccaggacccttcgtgtaatcccccatggattttcaagaattgaggaaaagtcacgtc
tccaaggccctacagggccagcggatactttgaaagcgacgataatatggtcgcttatttcatccaag
ccccgcgctaaacatggattttgggatgctatcccgaaagtacgacttggctccaaaggcc

[서열 번호 : 4]

aagcttaacttcagctgaagacccgttttcgatccgcggcgagcccggagtgtaaaacgatagacgtg
atgcttcggtcttctcacccccttcgaggtcataacatttttgtcatgattgccgtagtgctgatagtc
ctgagtctaaggcattcaatacaacgtacctcaggtcaattagactgtccatgactcatcttccgaag
cgcagaatgatacgcagttctcactagttgggacctgctcgacgtccggttaaggcggatttaactaa
gcatagggtaccgtcacctgggcaactgaaaatggcctctgtgacgcaagatgcatgttcggtcagct
cgttcaaagacggtatgaaatagagtagacatcagtacatcactcggacaggagcacctat

[서열 번호 : 5]

aagcttgcgcaactttgacgaaatgttggccaatagcatacccgaacaccgcagggttaatgcctaca
gctagtgttagtcgttccggtagacatctgttaaagccggaagctcgcccgactgtacgaaatcacat
ctaactatacaactgcgccactttgcaaatcgagtcacgacgacctgtcccttacggtgcccatttgc
gctgtaatgccgatcacttcacacaaacaaggcgcttgagagctcgaacttaggcgatgagggacaag
tggtacccaagctccaatagtagaatgtgtaccatagggccgcggcgagccgcctttgtatcctgaaa
aaattctcatcggcagcgcagtttattatttagttggaagcattagtgaacataacagcgc

[서열 번호 : 6]

aagcttccgtggtgggcagaagagcctagcttactctttatttaaaaacgccagtagaatttggtcgg
gaggatacgatccactgtccaacataaataacccgctgtagcctttacacattcacgggttaagtgta
gtgcgtgttctgtgtttctggtttgaataactgttcccactgtcttgaggatcgattctggccaaaat
gtatgaccctctacataggatgtacccctggggtaggacggaatcgattacgacccctgatgataatg
accaatcgtgacggtcggtgtctactgacttcgcctacatccgacgatcctggctaggcgggttgaga
[0370] acatcacggtattggggatcgggatgcgcgatcgcgataatgtggacttcgcaggtagtag

[서열 번호 :7]

aagcttggtggagcgcaaattctatttctgagttgcggcgtcagttgccattgaagtgcccgagctgc
atagtctcacggtgagtcctcttgtacgaccactagatgcaatgaagcgtgcatggagcgccactctg
caataaaagccgaaacgctctgtaaacaagattaatgtctcgtgatgctctgaaaccgtttacctaac
acgaacgataagacgcaacatcttccagagatgattacccgacacgctaatgaccgttatcactcccc
gcacatctgagcgtacttttgaagtcccgaggattgtcacggactaaatacctcgaatatcctgaac
taccttgccaatggagggaaggacagggacacgctgtcggtactttgtaggcatttgggt

[서열 번호 :8]

aagcttagctcgacgcacaatccaacaagtagcactgctgtctactaagcaacgtaatgatccattca
gacgagtttggaatgatctgcctcacccaaagcattaggcagccccctagctttctataggagaccga
aagagcatgagagagaactccctgatgacttactgactgcgtgatggttggctccgggacgcgcaacg
caacactttgtgtggcacgtaacttgtcgcacatatgtaatagcttcaaacccgcctcgtcttctggt
gtgcgctcgttcatttaatcgaatagattcctctctctactgctggtcaagggcgtattggaaataac
aagcaagctcctccgagctgagctacgagtcgatccgcccatgttccctcattatcgtctg

[서열 번호 :9]

aagcttgacctgtagcaccgcaaataatcattgctaatacgattcaagaatcgccctcgttatttgta
ttcacaggtgacccttggcttctactctaacacctaaggctgatccaactcagacttaagcggcgcag
ccgcaaatgtaatatgttcactgagagagagacgacggctccgtaggtcgaacattcaggtagctgga
gagatcattgcttagcatggcgctcgcggatctgttactgcaaatggcaacagactagaaaacaggcc
taatatgatctcggaattttcgcctaacacgctcctttgactggctgtgaggcctaagcgattctggc
agcgctgtgacttatcaagacacgcatgtcactacttgaccggcatcgtgccactctacgc

[서열 번호 :10]

aagcttttctctgcaacaggcgactatcggggccgggtgccaatctttcaaaagtgtgtaaacgtgcg
accgccagatgtcatgattcaatgtcttacctcgggctatcgtcataataagtttctaccgtaaggca
cgccctaaggacgttccgaataaacacgcacccccccgtcgtttcagaaatctcattaccggctgaca
tgcctttagatacctcagagaaatctaaccacgtgtgttacgactgacgtctcaaagagacgagctgc
tcctagctttcctattggagtatctgtgcctcttgtgtcgggatttagtggatcaatatgctccccta
cgataggtaagatttacccgttcgtcaattagagagccgggttttattattcggtcggcag

[서열 번호 :11]

catcatcaccatcaccattgatctcccagctgtgacacaaataagctagcccggggcagcatggaggt
taaaattgtgcatccgaccggccaggatacgtaatattaatgcgcaccgcgcactgaagaatatgatc
gaggctcgctgtagcagcactcagaaaaaaaaaaaaaaaaaaaaaaaaaagagtgaataacactcagat
ctcgggggcgtgaatgctaaacatacacagagcacgcggtgatgtataccgctatgtcggtcatgtgc

[0371]

tacctacagaagagctaggagtggatgagcactacacggtttcgggctaagaccatactctcacacgt
gtggatgactcgagacagcagtgtcagagcatgtagctctagagatgacacgatgaattc

[서열 번호 :12]

catcatcaccatcaccattgatctcccagctgtgacacaaataagctagcccggggcagcatggaggt
taaaattgtgcatccgaccggccaggatacgtaatattaatgcgcaccgcgcactgaagaatatgatc
gaggctcgctgtagcagcactcagaaccgtaataggaagtataccaaaagagtgaataacactcagat
ctcgggggcgtgaatgctaaacatacacagagcacgcggtgatgtataccgctatgtcggtcatgtgc
tacctacagaagagctaggagtggatgagcactacacggtttcgggctaagaccatactctcacacgt
gtggatgactcgagacagcagtgtcagagcatgtagctctagagatgacacgatgaattc

[서열 번호 :13]

cctctacggctccgtatcttaagacaaatgcgttctcgtaggtttgcttctacgtgatcatccggggt
ggtaatccgccctcgatctcctaaggatgaaaagggttagttgggccgaatttagttgatcgataagc
tgacggaaatctttactagcggataagctcatcccttcctgggtcaagatgcgagctagtacggccgc
gtcgctaatctcaatgaccattaactttgcgtagccatgtgtgctgctgcggagcgatactattaatt
gccctttcagttctggttccattgcactctgaaggatctccagtttgtcggaatatcacgtaagaacg
cttggcagaaaaagtctctatgctgtaacgcctcgacgtgaaactcgacaatgtgaattc

[서열 번호 :14]

cgtgaggggcacggcgagggagatcacaatatactgtcgtcgtttgatttcggaacagagccaacggg
ttcgggtgtcttgtgtgcttcactacatgacctcggtaaccagcagatttggtccaccgggtttgtgc
tggatttaggacaaggcgaaatatcatgatatacacagcatcgctttgccgttacatttttggcagcc
aaatggatcagaggctggtggggattacaccaccttgcccttacattggctaacgttttcaacacgtg
ttcctaaaatgtcagtcatgtccccccacacactatagcgctgagtcgatggagatcaaatgaggaat
cgaccggaaaccttggtgtcactgcctatgcgccggcaatgaacaaaccgaagtgaattc

[서열 번호 :15]

aagctggtagcatatggatagctggcatgttcagataattgctatctggtatccccacggatgctgat
ggctgatctttaaggtaaatgacattcgttgtctttacgcgccacagtgttgggccaagcagtctagt
catccagggtcatgctgagtctgcctcgtagcttaaactgttctaccattacgcggtcacgagccgtg
acatctcctatttacctggcacggttgcggtggcttgtaccgctccagatattataggagtcaagtct
aatgtcttatttatgcgagcgtcataggaccttgtccaataaattgaaaggatacgcccgagctgtgg
tagctgttagtgacggcatattgccgagggagccatcgaatgcaatgttgattcgaattc

[서열 번호 :16]

gtgctgttttgttcctcagttcgatacgacctaggaactgatggcgggctacccggatgatctcgatt
tgttctctcatgatagcaacggcgtcaagcgtcagtcttgtctcgatggagggtcgagtagatttggc
[0372] ttggatctttctcgtgtaaagtaaatccctgccagaggaccgagctggacggcgaagaagtttttttta

tctctgcacttcgaacgataagcgtcgtctccctggtcgcaaacatgggcccaaattggcttgcgatt
gttaaactaccggagtttttaatcgcctaaaccgcggagttaatccatgcaaccaagccagtaggatg
aagaagtgcgtccagtcgatcgttagtgcctggaatttctcttatcggcatcaagaattc

[서열 번호 :17]
tgtccgctctctagcagaagttgtaagttttaactcagtaggctgctactgaggggattgaacgcatg
ttatttgggttagtggtaataaatgactgtctcaggcgccatgctagagaacaattttgctggtttgc
ttacatggagacactagtctggtaccgcaccactcatggaatcaagcgtggtaggcccattgtttacg
tacgagccggctgcatgagggcacatagcatctggataaggcccgagagacagaggtctgccgagttt
tacgataccatagctgttgcgccttgcattgctatcggttttacctgtcgtctccggcagacggttta
ttcctcactcaattaattggctagtgcggctggttatccaacaagcgcattagtgaattc

[서열 번호 :18]
ccggtaacttgctcctgggacgcttaaatggcaattttaaaggaggcgaccgacccccctaacctaag
gatggtacttggtgaatactatcaaccacctccgtgacggcggccaattcaatcctgtaacgcgtgtc
gtaaaagttcagtttgtcgcagggtcgagttacccgtaatcctgggaacgcccccccaatccgcttca
gggctatatgccacacttgaaatcggaagtatcttggcttgagtatagtctggcgtggtaccacacat
ctacagtgaggtgaaaggcgcttctggcaaggtacgttctgcctgacagaattattcgcattagtgga
tgcgtccctggagtgcgtaaagcacactcggcagatgagtgctcggagcggactgaattc

[서열 번호 :19]
tagatgttgtacctgacaaccttctccctgcaaagcgggtgcctaaagatgttgttacatactccagg
cctcgatatggtccaatcaaaatcccatcggaccagcgttggaaagtagcacataagcgtgagacctc
aggagatccgtgtataagtgaatactggcattgggggtagttactagtgccgttcaatcggggaatga
ctcgggacataacgtctctaatctatatgagggtaccatattcaccgtaaaagactagagtccaattt
ggcctttcctcttagggaagagagtacaaaccgaaaacctggcgatcacgcctgcacagcagaatctt
gcctcgtttgtgtatcattgtggcagaggagcctttaagacatgcgaatagatcgaattc

[서열 번호 :20]
gattttgtcgtaaaacgatcatcatgagatcaagttcgtagaagccctgtcatatttaggagtttgat
gatcggcgcgagtgtaagtagcacaccgtattccaccgtgtttacctaacgcgactgcacagtactgg
caggtaacgtacaaactcatacaagggtttccacctctggcatgcttcttcggtatctcgttcgatgt
cgcattaatgcgttgaggaatggggttcatctggtcagggtctgaccgtttgtaaactaggtgacgag
cctgcggacctgatgtttaatctagcgccctttatggaaatctgttacgcgcagccagatgtgttgta
tcgagggatgtctaggtcctacacgcgacgatgaaacgggttcgtgtcggataggaattc

[0373]

[0374] **참조문헌**

[0375] 상기에서 인용된 모든 특허 및 공개문헌들은 이들 전체가 원용에 의해 본 명세서에 포함된다.

[0376] 23andMe. http://blog.23andme.com/23andme-and-you/update-from-23andme/

[0377] Akmaev VR, Wang CJ. (2004) Bioinformatics 20:1254-1263.

[0378] De Bruyn A, Martin DP, Lefeuvre P. (2014) Methods Mol. Biol. 1115:257-277.

[0379] Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. (2008) Nat Methods 5:235-237.

[0380] Buschmann T, Bystrykh LV. (2013) BMC Bioinformatics 14:272.

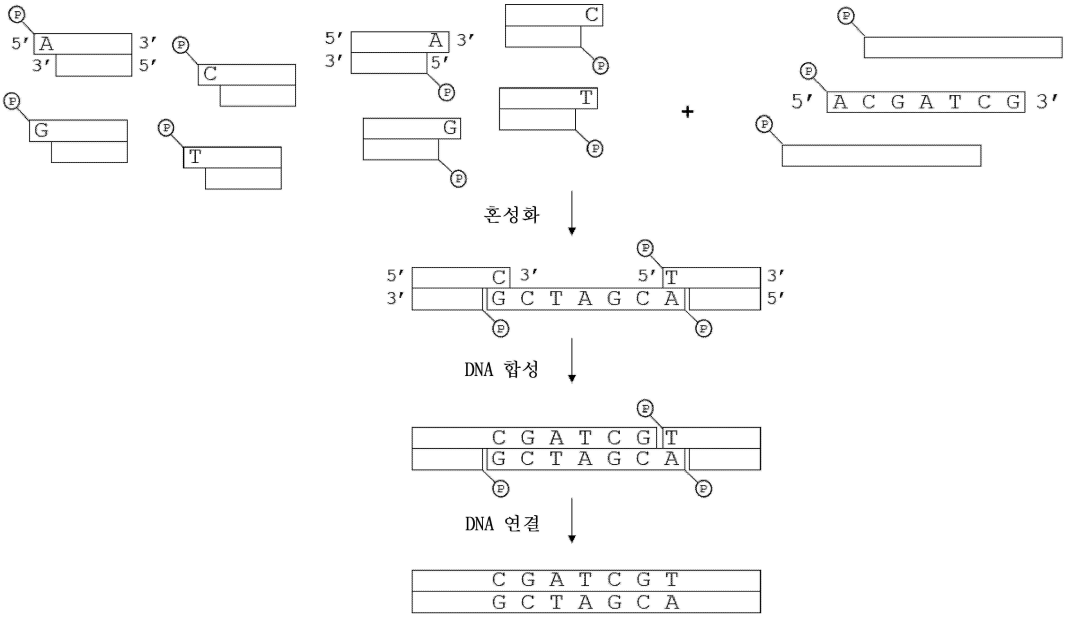
[0381] Sambrook and Russel: 'Molecular Cloning: A Laboratory Manual', 3rd Edition, 2001, Cold Spring Harbor Laboratory Press.

[0382] Norton et al., (2015) N. Engl. J. Med 372:1589-1597.

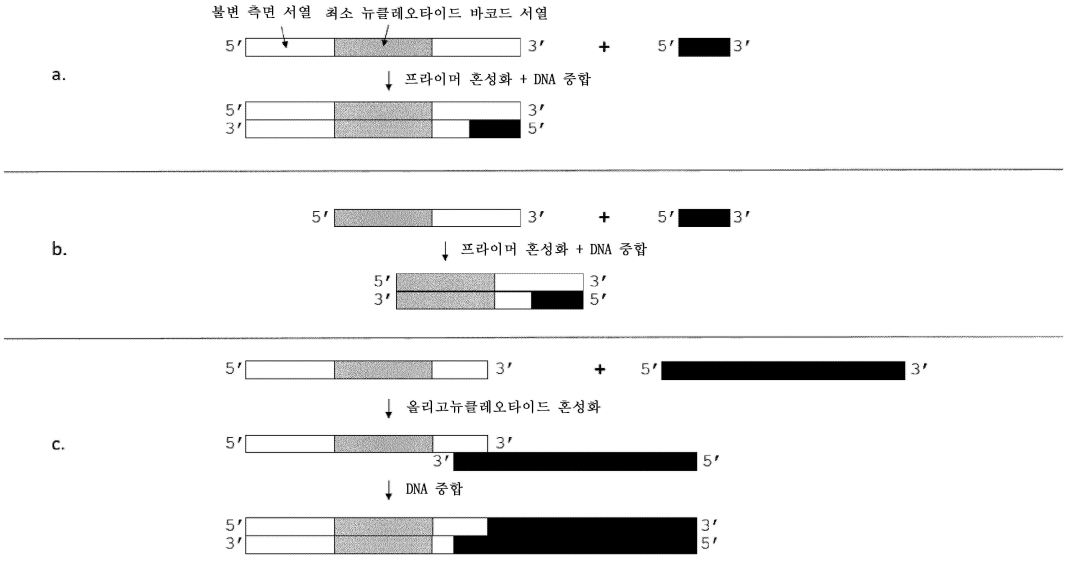
[0383] Taylor et al. (2015) Nat. Genetics 47:717-726.

도면

도면1



도면2

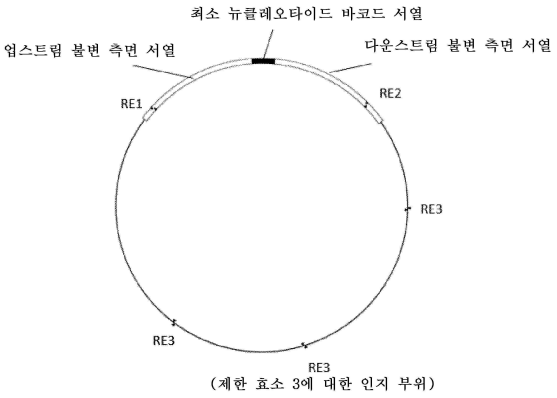


도면3

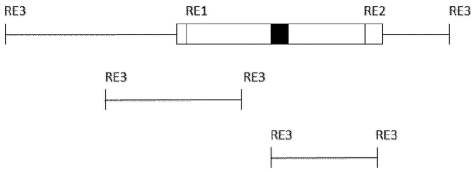
CFS1				MNB		CFS2			
CFS				MNB	L1	CS	L2		
L1	AP1	L2		MNB	L3	CS	AP2	L4	
L1	AP1	L2	ES1	MNB	ES2	L3	CS	AP2	L4
L1	AP1	L2	ES1	L3	MNB	L4	ES2	L5	CS	AP2	L6
L1	ASP1	L2	ES1	MNB	ES2	L3	CS	ASP2	L4
L1	ASP1	L2	NBSIS1	ES1	MNB	ES2	NBSIS2	L3	CS	ASP2

CFS: 불변 측면 서열
CS: 포착 서열
ES: 추출 서열
L: 링커 서열
MNB: 최소 뉴클레오타이드 바코드 서열
AP: 증폭 프라이머 결합 부위
ASP: 증폭 및 시퀀싱 프라이머 결합 부위
NBSIS: 뉴클레오타이드 바코드 서열 식별자 서열

도면4

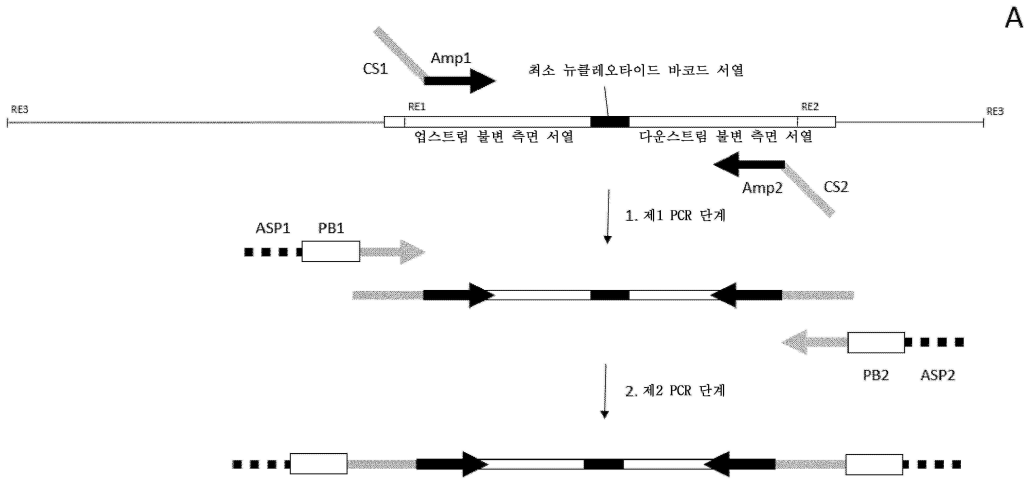


원형 뉴클레오타이드 바코드 플라스미드



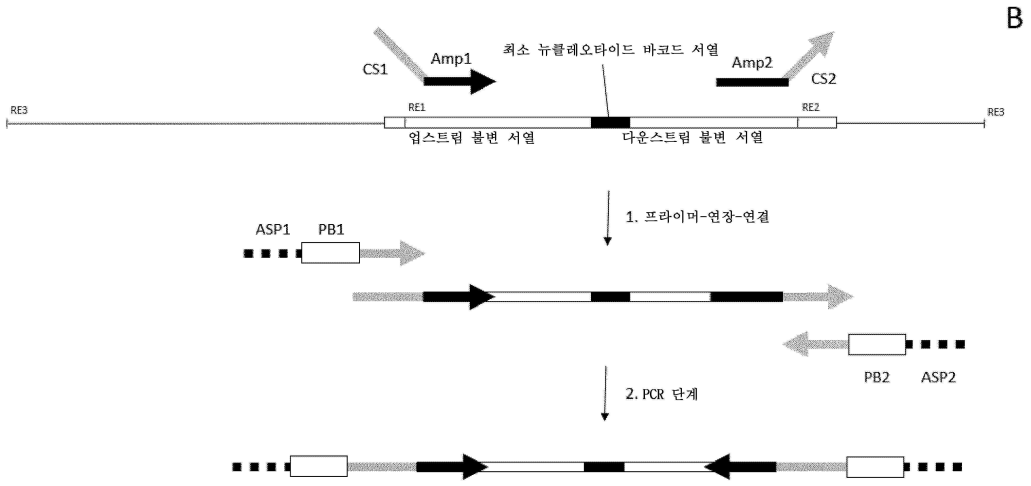
RE3 선형화된 뉴클레오타이드 바코드 서열

도면5a



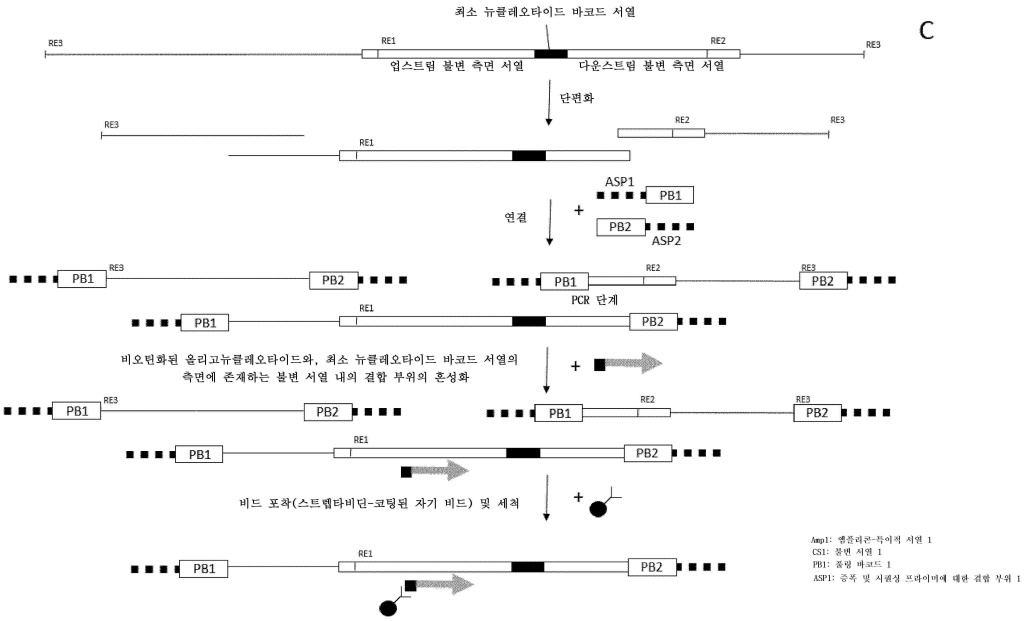
Amp1: 앰플리콘-특이적 서열 1
CS1: 불변 서열 1
PB1: 폴링 바코드 1
ASP1: 증폭 및 시퀀싱 프라이머에 대한 결합 부위 1

도면5b

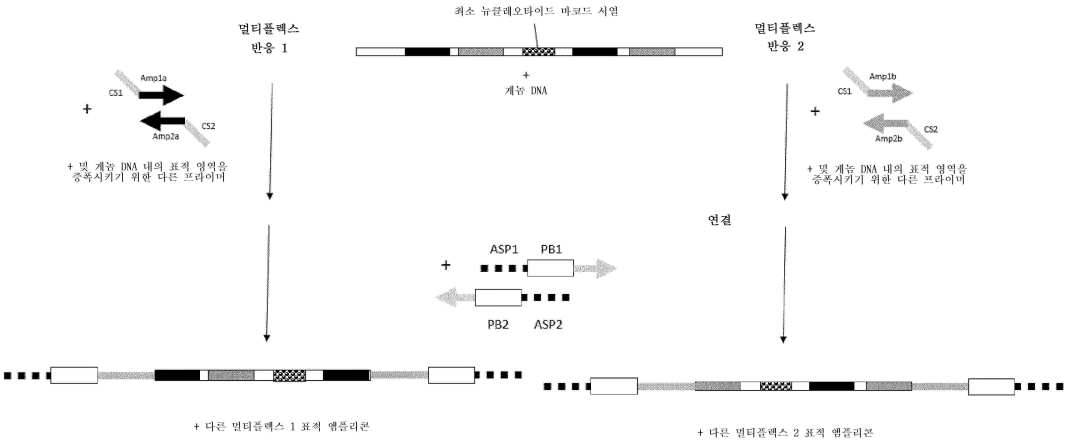


Amp1: 앰플리콘-특이적 서열 1
CS1: 불변 서열 1
PB1: 폴링 바코드 1
ASP1: 시퀀싱 프라이머에 대한 결합 부위 1

도면5c

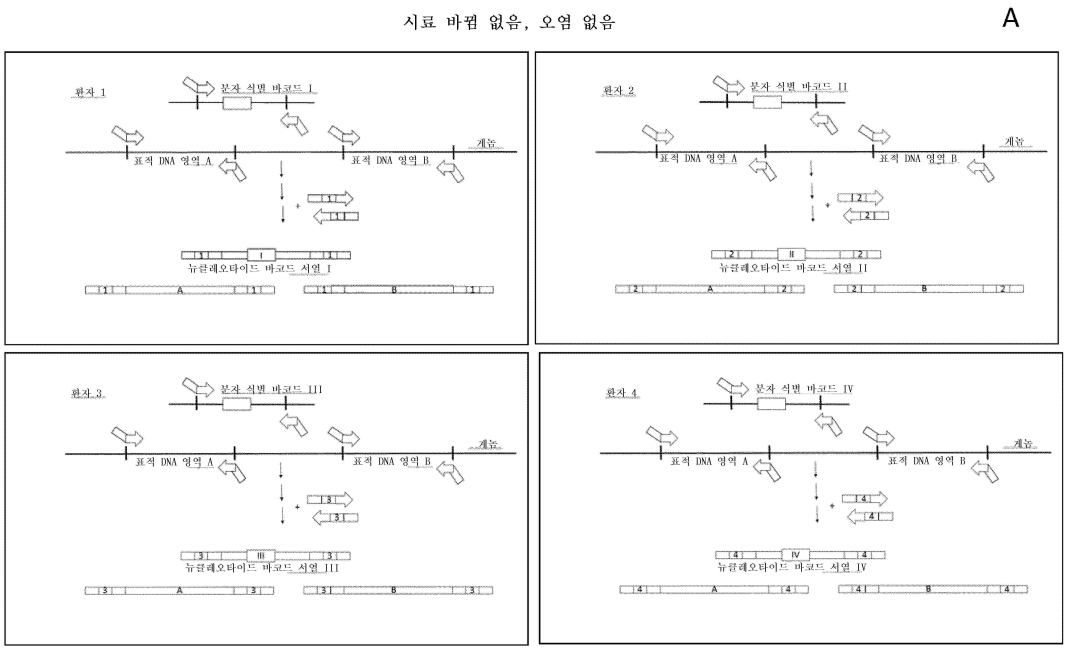


도면6

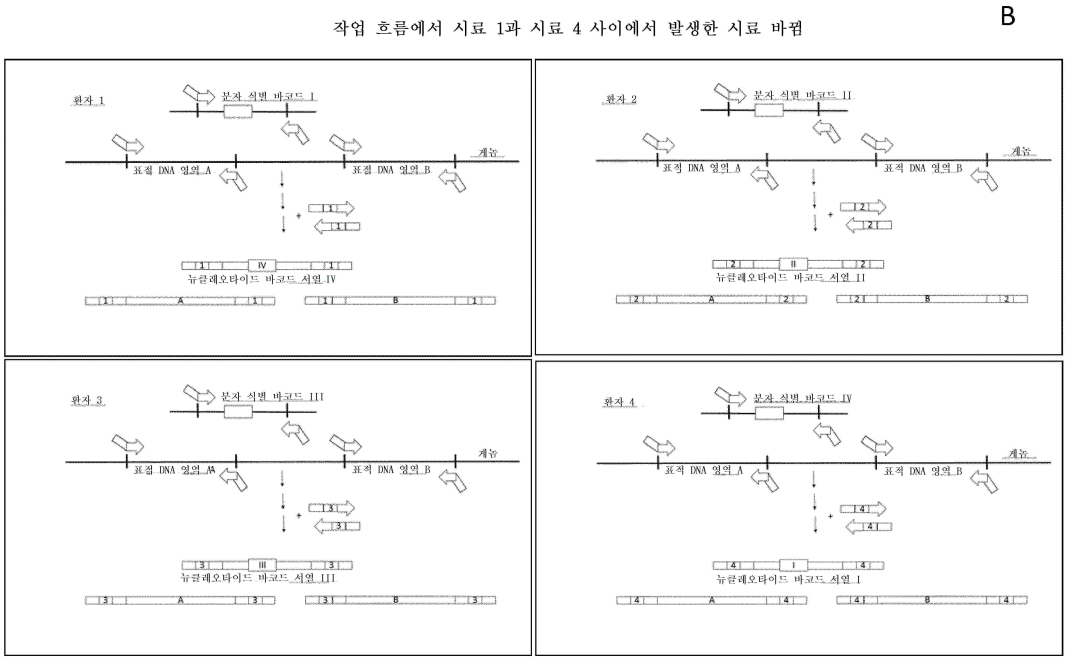


Amp1a: 앰플리콘-특이적 서열 1a
CS1: 불변 서열 1
PB1: 품질 바코드 1
ASP1: 증폭 및 시퀀싱 프라이머에 대한 결합 부위 1

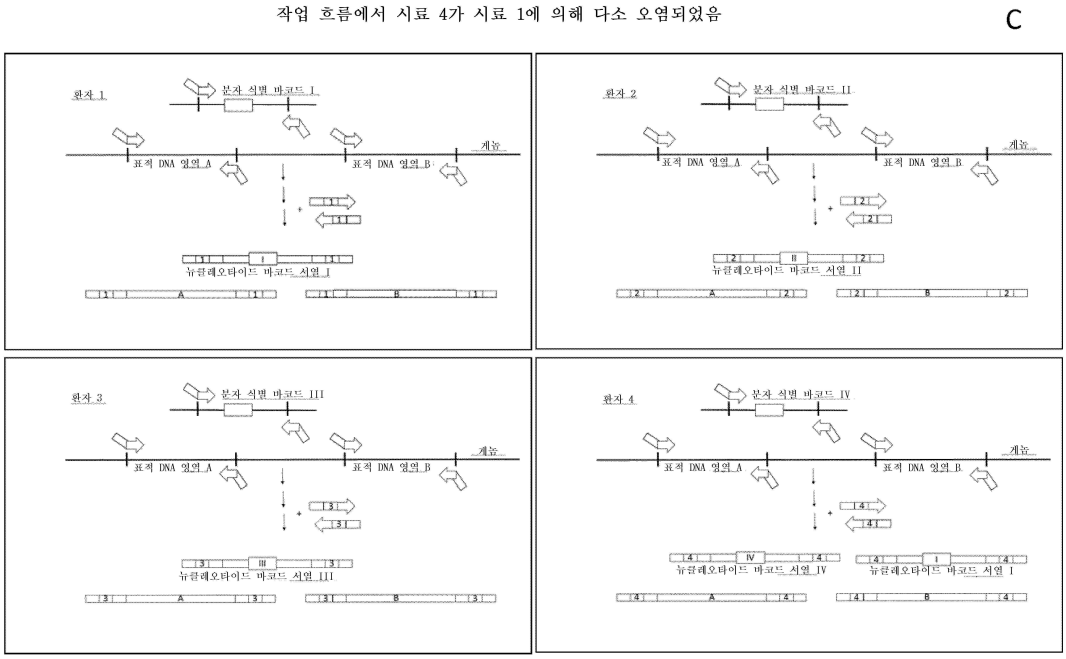
도면7a


시료 바꿈 없음, 오염 없음
A
환자 1
분자 식별 바코드 I
표적 DNA 영역 A
표적 DNA 영역 B
게놈
뉴클레오타이드 바코드 서열 I
환자 2
분자 식별 바코드 II
뉴클레오타이드 바코드 서열 II
환자 3
분자 식별 바코드 III
뉴클레오타이드 바코드 서열 III
환자 4
분자 식별 바코드 IV
뉴클레오타이드 바코드 서열 IV

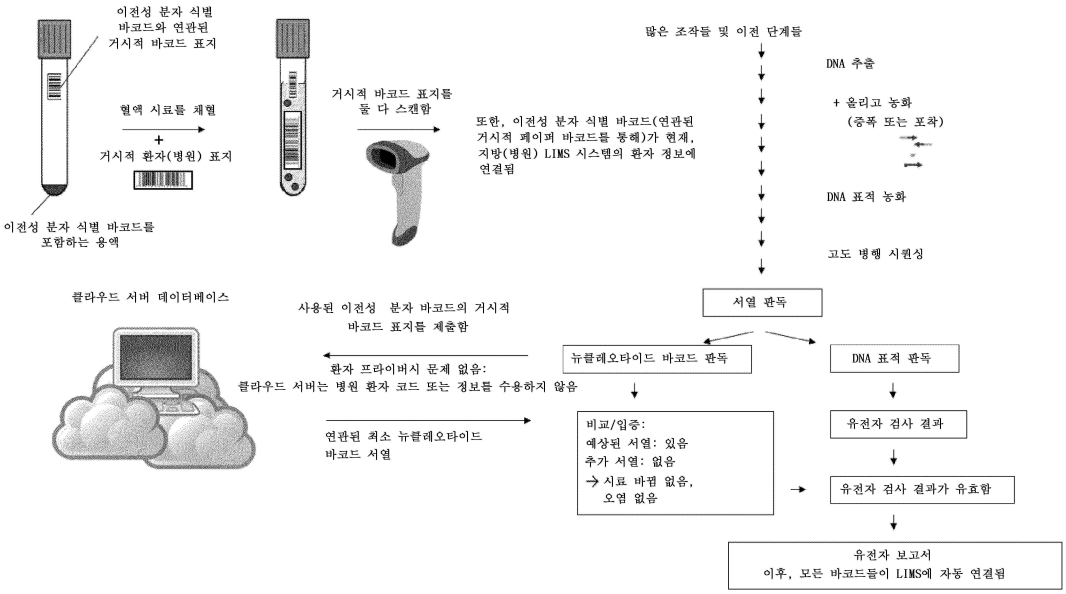
도면7b


작업 흐름에서 시료 1과 시료 4 사이에서 발생한 시료 바꿈
B
환자 1
분자 식별 바코드 I
표적 DNA 영역 A
표적 DNA 영역 B
게놈
뉴클레오타이드 바코드 서열 IV
환자 2
분자 식별 바코드 II
뉴클레오타이드 바코드 서열 II
환자 3
분자 식별 바코드 III
뉴클레오타이드 바코드 서열 III
환자 4
분자 식별 바코드 IV
뉴클레오타이드 바코드 서열 I

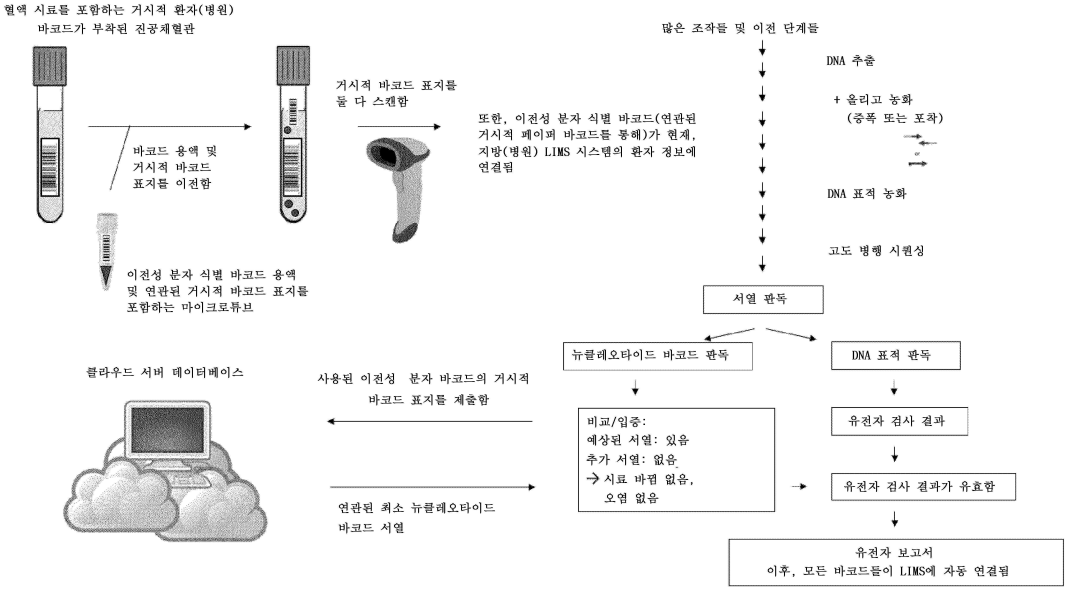
도면7c


작업 흐름에서 시료 4가 시료 1에 의해 다소 오염되었음
C
환자 1
분자 식별 바코드 I
표적 DNA 영역 A
표적 DNA 영역 B
게놈
뉴클레오타이드 바코드 서열 I
환자 2
분자 식별 바코드 II
뉴클레오타이드 바코드 서열 II
환자 3
분자 식별 바코드 III
뉴클레오타이드 바코드 서열 III
환자 4
분자 식별 바코드 IV
뉴클레오타이드 바코드 서열 IV
뉴클레오타이드 바코드 서열 I

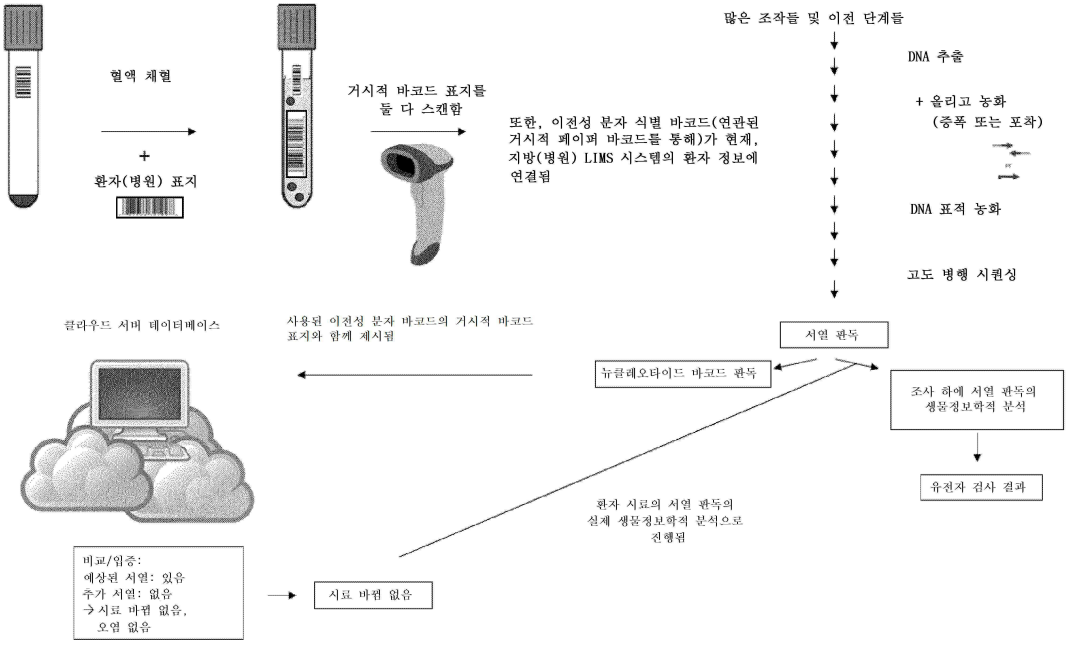
도면8


이전성 분자 식별 바코드와 연관된 거시적 바코드 표지
혈액 시료를 채혈
+
거시적 환자(병원) 표지
이전성 분자 식별 바코드를 포함하는 용액
거시적 바코드 표지를 둘 다 스캔함
또한, 이전성 분자 식별 바코드(연관된 거시적 페이퍼 바코드를 통해)가 현재, 지방(병원) LIMS 시스템의 환자 정보에 연결됨
많은 조작들 및 이전 단계들
DNA 추출
+ 올리고 농화
(중폭 또는 포착)
DNA 표적 농화
고도 병행 시퀀싱
서열 판독
뉴클레오타이드 바코드 판독
DNA 표적 판독
유전자 검사 결과
유전자 검사 결과가 유효함
유전자 보고서
이후, 모든 바코드들이 LIMS에 자동 연결됨
비교/입증:
예상된 서열: 있음
추가 서열: 없음
→ 시료 바뀜 없음,
오염 없음
클라우드 서버 데이터베이스
사용된 이전성 분자 바코드의 거시적 바코드 표지를 제출함
환자 프라이버시 문제 없음:
클라우드 서버는 병원 환자 코드 또는 정보를 수용하지 않음
연관된 최소 뉴클레오타이드 바코드 서열

도면9


혈액 시료를 포함하는 거시적 환자(병원)
바코드가 부착된 진공채혈관
바코드 용액 및
거시적 바코드
표지를 이전함
이전성 분자 식별 바코드 용액
및 연관된 거시적 바코드 표지를
포함하는 마이크로튜브
거시적 바코드 표지를
둘 다 스캔함
또한, 이전성 분자 식별 바코드(연관된
거시적 페이퍼 바코드를 통해)가 현재,
지방(병원) LIMS 시스템의 환자 정보에
연결됨
많은 조작들 및 이전 단계들
DNA 추출
+ 올리고 농화
(증폭 또는 포착)
DNA 표적 농화
고도 병행 시퀀싱
서열 판독
뉴클레오타이드 바코드 판독
DNA 표적 판독
유전자 검사 결과
유전자 검사 결과가 유효함
유전자 보고서
이후, 모든 바코드들이 LIMS에 자동 연결됨
비교/입증:
예상된 서열: 있음
추가 서열: 없음
→시료 바뀜 없음,
오염 없음
클라우드 서버 데이터베이스
사용된 이전성 분자 바코드의 거시적
바코드 표지를 제출함
연관된 최소 뉴클레오타이드
바코드 서열

도면10


혈액 채혈
+
환자(병원) 표지
거시적 바코드 표지를
둘 다 스캔함
또한, 이전성 분자 식별 바코드(연관된
거시적 페이퍼 바코드를 통해)가 현재,
지방(병원) LIMS 시스템의 환자 정보에
연결됨
많은 조작들 및 이전 단계들
DNA 추출
+ 올리고 농화
(증폭 또는 포착)
DNA 표적 농화
고도 병행 시퀀싱
서열 판독
뉴클레오타이드 바코드 판독
조사 하에 서열 판독의
생물정보학적 분석
유전자 검사 결과
환자 시료의 서열 판독의
실제 생물정보학적 분석으로
진행됨
클라우드 서버 데이터베이스
사용된 이전성 분자 바코드의 거시적 바코드
표지와 함께 제시됨
비교/입증:
예상된 서열: 있음
추가 서열: 없음
→시료 바뀜 없음,
오염 없음
시료 바뀜 없음

도면11



DNA 추출 설비 NGS 설비

오염

없음

	시료 1	시료 2
혈액	1	2
DNA	A	B

DNA 추출 동안

	시료 1		시료 2
혈액	1	2	2
DNA	A		B

NGS 제조/시퀀싱 동안

	시료 1		시료 2
혈액	1	2	2
DNA	A	B	B

서 열 목 록

SEQUENCE LISTING

<110> DName-iT NV

Cuppens, Harry

<120> METHODS FOR IDENTIFICATION OF SAMPLES

<130> DNA3713PCT

<150> GB1701908.4

<151> 2017-02-06

<150> EP16169997.0

<151> 2016-05-17

<160> 20

<170> PatentIn version 3.5

<210> 1

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 1

aagctttgtg gatgtacaag tccacaccat gtacactaga cgcagcctgt acagatatcc 60

atccagtgta ctcactgtcg acacggatcc aatgcccggg ttctgataga cgaacgacga 120

gatgtgcagt gacttcgagg atcccagatg tgcacgtagt gcaggtagct tgaatgacta 180

ctacgcctgt agcatcatca cgtagactcg tacagctaca tgacggtagc tagattgacg 240

actcaagcat gctagtgtcg ttactgacct gatgacacag tcgatgcgac cttaatacga 300

ctcactatag ggtcaacaag accctgcaga tcccgggatc cgcctcttaa gctgcgcagg 360

ccaggaattg cacgtcagag cactaaggcc gccaccatgg c 401

<210> 2

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 2

aagctttgtg gatgtacaag tccacaccat gtacactaga cgcagcctgt acagatatcc 60

atccagtgta ctcactgtcg acacggatcc aatgcccggg ttctgataga cgaacgacga 120

gatgtgcagt gacttcgagg atcccagatg tgcacgtagt gcaggtagct tgaatgacta 180

ctacgcctgt agcatcatca cgtagactcg tacagctaca tgacggtagc tagattgacg 240

actcaagcat gctagtgtcg ttactgacct gatgacacag tcgatgcgac cttaaatgct 300

gagagattag ggtcaacaag accctgcaga tcccgggatc cgcctcttaa gctgcgcagg 360

ccaggaattg cacgtcagag cactaaggcc gccaccatgg c 401

<210> 3

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> constant sequence

<400> 3

aagcttctct cgccagctat ttaaagtacg agtcgggagg ccttagcacg aactgatttt 60

tccagcctga gtgctgttct tgcatgtacc ttctatctaa cgacgtccgt aataggaagt 120

ataccaggtc gaactaacga ctcctttgcc gtagcgagtg tttcgccaaa agtgtctggg 180

tctactggcc accgtccagc atttctatgc ccgtaccagg acccttcgtg taatccccca 240

tggattttca agaattgagg aaaagtcacg tctccaaggc cctacagggc cagcggatac 300

tttgaaagcg acgataatat ggtcgcttat ttcatccaag ccccgcgcta aacatggatt 360

ttgggatgct atcccgaaag tacgacttgg ctccaaaggc c 401

<210> 4

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 4

aagcttaact tcagctgaag acccgttttc gatccgcggc gagcccggag tgtaaaacga 60

tagacgtgat gcttcggtct tctcacccct tcgaggtcat aacatttttg tcatgattgc 120

cgtagtgctg atagtcctga gtctaaggca ttcaatacaa cgtacctcag gtcaattaga 180

ctgtccatga ctcatcttcc gaagcgcaga atgatacgca gttctcacta gttgggacct 240

gctcgacgtc cggttaaggc ggatttaact aagcataggg taccgtcacc tgggcaactg 300

aaaatggcct ctgtgacgca agatgcatgt tcggtcagct cgttcaaaga cggtatgaaa 360

tagagtagac atcagtacat cactcggaca ggagcaccta t 401

<210> 5

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 5

aagcttgcgc aactttgacg aaatgttggc caatagcata cccgaacacc gcagggttaa 60

tgcctacagc tagtgttagt cgttccggta gacatctgtt aaagccggaa gctcgcccga 120

ctgtacgaaa tcacatctaa ctatacaact gcgccacttt gcaaatcgag tcacgacgac 180

ctgtcccttа cggtgcccat ttgcgctgta atgccgatca cttcacacaa acaaggcgct 240

tgagagctcg aacttaggcg atgagggaca agtggtaccc aagctccaat agtagaatgt 300

gtaccatagg gccgcggcga gccgcctttg tatcctgaaa aaattctcat cggcagcgca 360

gtttattatt tagttggaag cattagtgaa cataacagcg c 401

<210> 6

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 6

aagcttccgt ggtgggcaga agagcctagc ttactcttta tttaaaaacg ccagtagaat 60

ttggtcggga ggatacgatc cactgtccaa cataaataac ccgctgtagc ctttacacat 120

tcacgggtta agtgtagtgc gtgttctgtg tttctggttt gaataactgt tcccactgtc 180

ttgaggatcg attctggcca aaatgtatga ccctctacat aggatgtacc cctggggtag 240

gacggaatcg attacgaccc ctgatgataa tgaccaatcg tgacggtcgg tgtctactga 300

cttcgcctac atccgacgat cctggctagg cgggttgaga acatcacggt attggggatc 360

gggatgcgcg atcgcgataa tgtggacttc gcaggtagta g 401

<210> 7

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 7

aagcttggtg gagcgcaaat tctatttctg agttgcggcg tcagttgcca ttgaagtgcc 60

cgagctgcat agtctcacgg tgagtcctct tgtacgacca ctagatgcaa tgaagcgtgc 120

atggagcgcc actctgcaat aaaagccgaa acgctctgta aacaagatta atgtctcgtg 180

atgctctgaa accgtttacc taacacgaac gataagacgc aacatcttcc agagatgatt 240

acccgacacg ctaatgaccg ttatcactcc ccgcacatct gagcgtactt tttgaagtcc 300

cgaggattgt cacggactaa atacctcgaa tatcctgaac tacctttgcc aatggaggga 360

aggacaggga cacgctgtcg gtactttgta ggcatttggg t 401

<210> 8

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 8

aagcttagct cgacgcacaa tccaacaagt agcactgctg tctactaagc aacgtaatga 60

tccattcaga cgagtttgga atgatctgcc tcacccaaag cattaggcag ccccctagct 120

ttctatagga gaccgaaaga gcatgagaga gaactccctg atgacttact gactgcgtga 180

tggttggctc cgggacgcgc aacgcaacac tttgtgtggc acgtaacttg tcgcacatat 240

gtaatagctt caaacccgcc tcgtcttctg gtgtgcgctc gttcatttaa tcgaatagat 300

tcctctctct actgctggtc aagggcgtat tggaaataac aagcaagctc ctccgagctg 360

agctacgagt cgatccgccc atgttccctc attatcgtct g 401

<210> 9

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 9

aagcttgacc tgtagcaccg caaataatca ttgctaatac gattcaagaa tcgccctcgt 60

tatttgtatt cacaggtgac ccttggcttc tactctaaca cctaaggctg atccaactca 120

gacttaagcg gcgcagccgc aaatgtaata tgttcactga gagagagacg acggctccgt 180

aggtcgaaca ttcaggtagc tggagagatc attgcttagc atggcgctcg cggatctgtt 240

actgcaaatg gcaacagact agaaaacagg cctaatatga tctcggaatt ttcgcctaac 300

acgctccttt gactggctgt gaggcctaag cgattctggc agcgctgtga cttatcaaga 360

cacgcatgtc actacttgac cggcatcgtg ccactctacg c 401

<210> 10

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 10

aagcttttct ctgcaacagg cgactatcgg ggccgggtgc caatctttca aaagtgtgta 60

aacgtgcgac cgccagatgt catgattcaa tgtcttacct cgggctatcg tcataataag 120

tttctaccgt aaggcacgcc ctaaggacgt tccgaataaa cacgcacccc cccgtcgttt 180

cagaaatctc attaccggct gacatgcctt tagatacctc agagaaatct aaccacgtgt 240

gttacgactg acgtctcaaa gagacgagct gctcctagct ttcctattgg agtatctgtg 300

cctcttgtgt cgggatttag tggatcaata tgctccccta cgataggtaa gatttacccg 360

ttcgtcaatt agagagccgg gttttattat tcggtcggca g 401

<210> 11

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 11

catcatcacc atcaccattg atctcccagc tgtgacacaa ataagctagc ccggggcagc 60

atggaggtta aaattgtgca tccgaccggc caggatacgt aatattaatg cgcaccgcgc 120

actgaagaat atgatcgagg ctcgctgtag cagcactcag aaaaaaaaaa aaaaaaaaaa 180

aaaaagagtg aataacactc agatctcggg ggcgtgaatg ctaaacatac acagagcacg 240

cggtgatgta taccgctatg tcggtcatgt gctacctaca gaagagctag gagtggatga 300

gcactacacg gtttcgggct aagaccatac tctcacacgt gtggatgact cgagacagca 360

gtgtcagagc atgtagctct agagatgaca cgatgaattc 400

<210> 12

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 12

catcatcacc atcaccattg atctcccagc tgtgacacaa ataagctagc ccggggcagc 60

atggaggtta aaattgtgca tccgaccggc caggatacgt aatattaatg cgcaccgcgc 120

actgaagaat atgatcgagg ctcgctgtag cagcactcag aaccgtaata ggaagtatac 180

caaaagagtg aataacactc agatctcggg ggcgtgaatg ctaaacatac acagagcacg 240

cggtgatgta taccgctatg tcggtcatgt gctacctaca gaagagctag gagtggatga 300

gcactacacg gtttcgggct aagaccatac tctcacacgt gtggatgact cgagacagca 360

gtgtcagagc atgtagctct agagatgaca cgatgaattc 400

<210> 13

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 13

cctctacggc tccgtatctt aagacaaatg cgttctcgta ggtttgcttc tacgtgatca 60

tccggggtgg taatccgccc tcgatctcct aaggatgaaa agggttagtt gggccgaatt 120

tagttgatcg ataagctgac ggaaatcttt actagcggat aagctcatcc cttcctgggt 180

caagatgcga gctagtacgg ccgcgtcgct aatctcaatg accattaact ttgcgtagcc 240

atgtgtgctg ctgcggagcg atactattaa ttgcccttc agttctggtt ccattgcact 300

ctgaaggatc tccagtttgt cggaatatca cgtaagaacg cttggcagaa aaagtctcta 360

tgctgtaacg cctcgacgtg aaactcgaca atgtgaattc 400

<210> 14

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 14

cgtgaggggc acggcgaggg agatcacaat atactgtcgt cgtttgattt cggaacagag 60

ccaacgggtt cgggtgtctt gtgtgcttca ctacatgacc tcggtaacca gcagatttgg 120

tccaccgggt ttgtgctgga tttaggacaa ggcgaaatat catgatatac acagcatcgc 180

tttgccgtta catttttggc agccaaatgg atcagaggct ggtggggatt acaccacctt 240

gcccttacat tggctaacgt tttcaacacg tgttcctaaa atgtcagtca tgtcccccca 300

cacactatag cgctgagtcg atggagatca aatgaggaat cgaccggaaa ccttggtgtc 360

actgcctatg cgccggcaat gaacaaaccg aagtgaattc 400

<210> 15

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 15

aagctggtag catatggata gctggcatgt tcagataatt gctatctggt atccccacgg 60

atgctgatgg ctgatcttta aggtaaatga cattcgttgt ctttacgcgc cacagtgttg 120

ggccaagcag tctagtcatc cagggtcatg ctgagtctgc ctcgtagctt aaactgttct 180

accattacgc ggtcacgagc cgtgacatct cctatttacc tggcacggtt gcggtggctt 240

gtaccgctcc agatattata ggagtcaagt ctaatgtctt atttatgcga gcgtcatagg 300

accttgtcca ataaattgaa aggatacgcc cgagctgtgg tagctgttag tgacggcata 360

ttgccgaggg agccatcgaa tgcaatgttg attcgaattc 400

<210> 16

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 16

gtgctgtttt gttcctcagt tcgatacgac ctaggaactg atggcgggct acccggatga 60

tctcgatttg ttctctcatg atagcaacgg cgtcaagcgt cagtcttgtc tcgatggagg 120

gtcgagtaga tttggcttgg atctttctcg tgtaaagtaa atccctgcca gaggaccgag 180

ctggacggcg aagaagtttt tttatctctg cacttcgaac gataagcgtc gtctccctgg 240

tcgcaaacat gggcccaaat tggcttgcga ttgttaaact accggagttt ttaatcgcct 300

aaaccgcgga gttaatccat gcaaccaagc cagtaggatg aagaagtgcg tccagtcgat 360

cgttagtgcc tggaatttct cttatcggca tcaagaattc 400

<210> 17

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 17

tgtccgctct ctagcagaag ttgtaagttt taactcagta ggctgctact gaggggattg 60

aacgcatgtt atttgggtta gtggtaataa atgactgtct caggcgccat gctagagaac 120

aattttgctg gtttgcttac atggagacac tagtctggta ccgcaccact catggaatca 180

agcgtggtag gcccattgtt tacgtacgag ccggctgcat gagggcacat agcatctgga 240

taaggcccga gagacagagg tctgccgagt tttacgatac catagctgtt gcgccttgca 300

ttgctatcgg ttttacctgt cgtctccggc agacggttta ttcctcactc aattaattgg 360

ctagtgcggc tggttatcca acaagcgcat tagtgaattc 400

<210> 18

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 18

ccggtaactt gctcctggga cgcttaaatg gcaattttaa aggaggcgac cgacccccct 60

aacctaagga tggtacttgg tgaatactat caaccacctc cgtgacggcg gccaattcaa 120

tcctgtaacg cgtgtcgtaa aagttcagtt tgtcgcaggg tcgagttacc cgtaatcctg 180

ggaacgcccc cccaatccgc ttcagggcta tatgccacac ttgaaatcgg aagtatcttg 240

gcttgagtat agtctggcgt ggtaccacac atctacagtg aggtgaaagg cgcttctggc 300

aaggtacgtt ctgcctgaca gaattattcg cattagtgga tgcgtccctg gagtgcgtaa 360

agcacactcg gcagatgagt gctcggagcg gactgaattc 400

<210> 19

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> barcode flanking region

<400> 19

tagatgttgt acctgacaac cttctccctg caaagcgggt gcctaaagat gttgttacat 60

actccaggcc tcgatatggt ccaatcaaaa tcccatcgga ccagcgttgg aaagtagcac 120

ataagcgtga gacctcagga gatccgtgta taagtgaata ctggcattgg gggtagttac 180

tagtgccgtt caatcgggga atgactcggg acataacgtc tctaatctat atgagggtac 240

catattcacc gtaaaagact agagtccaat ttggcctttc ctcttaggga agagagtaca 300

aaccgaaaac ctggcgatca cgcctgcaca gcagaatctt gcctcgtttg tgtatcattg 360

tggcagagga gcctttaaga catgcgaata gatcgaattc 400

<210> 20

<211> 400

<212> DNA

<213> Artificial Sequence

<220><223> constant sequence

<400> 20

gattttgtcg taaaacgatc atcatgagat caagttcgta gaagccctgt catatttagg 60

agtttgatga tcggcgcgag tgtaagtagc acaccgtatt ccaccgtgtt tacctaacgc 120

gactgcacag tactggcagg taacgtacaa actcatacaa gggtttccac ctctggcatg 180

cttcttcggt atctcgttcg atgtcgcatt aatgcgttga ggaatggggt tcatctggtc 240

agggtctgac cgtttgtaaa ctaggtgacg agcctgcgga cctgatgttt aatctagcgc 300

cctttatgga aatctgttac gcgcagccag atgtgttgta tcgagggatg tctaggtcct 360

acacgcgacg atgaaacggg ttcgtgtcgg ataggaattc 400