(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년11월27일
(11) 등록번호 10-2735452
(24) 등록일자 2024년11월25일

(51) 국제특허분류(Int. Cl.)
G16B 35/10 (2019.01) *C12N 15/10* (2017.01)
C12Q 1/6806 (2018.01) *C12Q 1/6869* (2018.01)
C12Q 1/6876 (2018.01) *G16B 25/10* (2019.01)
G16B 30/00 (2019.01)
(52) CPC특허분류
G16B 35/10 (2019.02)
C12N 15/1072 (2013.01)
(21) 출원번호 10-2023-7004940(분할)
(22) 출원일자(국제) 2018년11월05일
심사청구일자 2023년02월10일
(85) 번역문제출일자 2023년02월10일
(65) 공개번호 10-2023-0028569
(43) 공개일자 2023년02월28일
(62) 원출원 특허 10-2019-7036132
원출원일자(국제) 2018년11월05일
심사청구일자 2019년12월05일
(86) 국제출원번호 PCT/US2018/059255
(87) 국제공개번호 WO 2019/090251
국제공개일자 2019년05월09일
(30) 우선권주장
62/582,175 2017년11월06일 미국(US)
(56) 선행기술조사문헌
US20140141982 A1*
(뒷면에 계속)

(73) 특허권자
일루미나, 인코포레이티드
미국 캘리포니아 92122 샌디에고 일루미나 웨이 5200
(72) 발명자
비에셀리 존 에스.
미국 캘리포니아주 92122 샌디에이고 5200 일루미나 웨이
켈리 라이언 매튜
미국 캘리포니아주 92122 샌디에이고 5200 일루미나 웨이
(74) 대리인
특허법인아주김장리

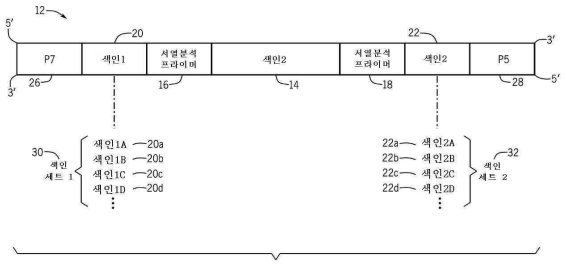
전체 청구항 수 : 총 21 항 심사관 : 김채연

(54) 발명의 명칭 **핵산 색인 기술**

(57) 요 약

예를 들면, 서열분석과 함께 사용하기 위한, 핵산의 색인을 위한 기술이 본 명세서에 나타난다. 기술은 색인된 핵산 단편을 개별적인 샘플로부터 생성하는 것을 포함하고, 이로써 핵산 단편의 각각의 색인 부위로 혼입된 색인 서열은 색인 서열과 구별 가능한 복수로부터 선택되고, 따라서 생성된 핵산 단편의 집단은 복수로부터의 각각의 색인 서열을 나타낸다. 이러한 방식으로, 단일 샘플로부터 생성된 색인된 핵산 단편은 낮은 서열 다양성과 연관된 색인 판독물 오류로 인한 잘못된 할당을 감소시키는 다양한 색인 서열의 혼합물에 의해 색인된다.

대 표 도 - 도1



(52) CPC특허분류
C12Q 1/6806 (2018.05)
C12Q 1/6869 (2018.05)
C12Q 1/6876 (2018.05)
G16B 25/10 (2019.02)
G16B 30/00 (2019.02)

(56) 선행기술조사문헌
US20160314242 A1*
M. Kircher 외, "Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform", Nucleic Acids Research, 40(1):e3. (2011.10.21.)
US10081807 B2
US20160319345 A1
*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

다중화 핵산 라이브러리 제조물(multiplexed nucleic acid library preparation)을 포함하는 기판;

이미지 데이터를 생성하기 위해 샘플 기판을 이미징하도록 구성된 이미저(imager); 및

상기 이미지 데이터를 처리하여 제1 샘플 및 제2 샘플의 시퀀싱 데이터를 생성하도록 구성된 컴퓨터를 포함하는, 시퀀싱 장치로서,

상기 제조물은,

제1 샘플로부터 제조된 제1 핵산 라이브러리로서, 상기 제1 핵산 라이브러리는 제1 복수의 핵산 단편들을 포함하고, 상기 제1 복수의 각각의 핵산 단편은 모두 서로 구별가능한 색인 서열들의 제1 세트로부터 선택된 적어도 2개의 상이한 색인 서열을 포함하고, 색인 서열들의 상기 제1 세트는 5' 색인 서열들 및 3' 색인 서열들을 포함하고, 색인 서열들의 상기 제1 세트는 상기 제1 복수의 개별 핵산 단편들에 배치되어 상기 5' 색인 서열들이 표적 서열의 3'가 아닌 표적 서열의 5'에 위치되고, 상기 3' 색인 서열들은 표적 서열의 5'가 아닌 표적 서열의 3'에 위치되고, 여기서, 상기 제1 복수의 핵산 단편들은 상기 제1 세트의 2개의 상이한 색인 서열의 둘 이상의 조합을 포함하고, 상기 제1 복수의 핵산 단편들의 둘 이상의 단편은 2개의 상이한 색인 서열의 동일한 조합을 가지고, 여기서 동일한 조합은 상기 제1 세트의 5' 색인 서열 및 3' 색인 서열을 포함하는 것인, 상기 제1 핵산 라이브러리; 및

제2 샘플로부터 제조된 제2 핵산 라이브러리로서, 상기 제2 핵산 라이브러리는 제2 복수의 핵산 단편들을 포함하고, 상기 제2 복수의 각각의 핵산 단편은 색인 서열들의 상기 제1 세트와, 그리고 모두 서로 구별 가능한 색인 서열들의 제2 세트로부터 선택된 적어도 2개의 상이한 색인 서열을 포함하고, 색인 서열들의 상기 제2 세트는 5' 색인 서열들 및 3' 색인 서열들을 포함하고, 색인 서열들의 상기 제2 세트는 상기 제2 복수의 개별 핵산 단편에 배치되어 상기 5' 색인 서열들이 표적 서열의 3'가 아닌 표적 서열의 5'에 위치되고, 상기 3' 색인 서열들은 표적 서열의 5'가 아닌 표적 서열의 3'에 위치되고,

여기서, 색인 서열들의 상기 제1 세트는 제2 샘플이 아닌 제1 샘플과 연관되고, 색인 서열들의 상기 제2 세트는 제1 샘플이 아닌 제2 샘플과 연관되고,

상기 제2 복수의 핵산 단편들은 2개의 상이한 색인 서열의 둘 이상의 조합을 포함하고, 상기 제2 복수의 핵산 단편들의 둘 이상의 단편은 2개의 상이한 색인 서열의 동일한 조합을 가지고, 여기서 동일한 조합은 상기 제2 세트의 5' 색인 서열 및 3' 색인 서열을 포함하는 것인, 상기 제2 핵산 라이브러리를 포함하는, 시퀀싱 장치.

청구항 2

제1항에 있어서, 제1 핵산 라이브러리의 제1 복수의 핵산 단편들 및 제2 핵산 라이브러리의 제2 복수의 핵산 단편들의 각각의 핵산 단편은 5' 어댑터 서열 및 3' 어댑터 서열을 포함하고, 2개의 상이한 색인 서열이 상기 5' 어댑터 서열 및 3' 어댑터 서열 사이에 위치하는, 시퀀싱 장치.

청구항 3

제2항에 있어서, 5' 어댑터 서열이 p5 어댑터 서열이고/이거나 3' 어댑터 서열이 p7 어댑터 서열인, 시퀀싱 장치.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 5' 색인 서열은 각 핵산 단편 상에 5' 어댑터 서열과 시퀀싱 프라이머 사이에 위치하는, 시퀀싱 장치.

청구항 5

제2항 또는 제3항에 있어서, 5' 어댑터 서열은 제1 핵산 라이브러리의 제1 복수의 핵산 단편들 및 제2 핵산 라

이브러리의 제2 복수의 핵산 단편들의 각각의 핵산 단편에서 동일한 서열인, 시퀀싱 장치.

청구항 6

제2항 또는 제3항에 있어서, 3' 어댑터 서열은 제1 핵산 라이브러리의 제1 복수의 핵산 단편들 및 제2 핵산 라이브러리의 제2 복수의 핵산 단편들의 각각의 핵산 단편에서 동일한 서열인, 시퀀싱 장치.

청구항 7

제1항 내지 제3항 중 어느 한 항에 있어서, 5' 색인 서열의 길이가 5-10 염기이고, 3' 색인 서열의 길이가 5-10 염기인, 시퀀싱 장치.

청구항 8

제1항 내지 제3항 중 어느 한 항에 있어서, 제1 세트의 5' 색인 서열들은 적어도 4개의 상이한 색인 서열을 포함하고, 제1 세트의 3' 색인 서열들은 적어도 4개의 상이한 색인 서열을 포함하는, 시퀀싱 장치.

청구항 9

제1항 내지 제3항 중 어느 한 항에 있어서, 제2 세트의 5' 색인 서열들은 적어도 4개의 상이한 색인 서열을 포함하고 제2 세트의 3' 색인 서열들은 적어도 4개의 상이한 색인 서열을 포함하는, 시퀀싱 장치.

청구항 10

제1항 내지 제3항 중 어느 한 항에 있어서, 기판은 제1 복수의 핵산 단편 및 제2 복수의 핵산 단편 중 적어도 일부의 영역에 상보적인 서열을 갖는 복수의 포획 프로브를 포함하는 것인, 시퀀싱 장치.

청구항 11

제10항에 있어서, 복수의 포획 프로브가 5' 어댑터에 상보적인 제1 포획 프로브 및 3' 어댑터에 상보적인 제2 포획 프로브를 포함하는 것인, 시퀀싱 장치.

청구항 12

제1항 내지 제3항 중 어느 한 항에 있어서, 제1 복수의 핵산 단편들 및 제2 복수의 핵산 단편들이 상기 복수의 포획 프로브의 각각의 포획 프로브에 대한 상보적 결합을 통해 기판 상에 고정화되는, 시퀀싱 장치.

청구항 13

복수의 샘플로부터의 복수의 핵산 단편을 포함하는 다중화된 핵산 라이브러리 제조물이 배치된 기판으로부터 이미지 데이터를 생성하도록 구성된 이미저; 및

프로세서를 포함하는 시퀀싱 장치로서,

상기 프로세서는,

이미지 데이터를 수신하고;

이미지 데이터에 기초하여 시퀀싱 데이터를 생성하고, 여기서 상기 시퀀싱 데이터는 복수의 핵산 단편들을 나타내는 복수의 시퀀싱 리드(sequencing reads)를 포함하며;

상기 복수의 시퀀싱 리드 중 한 시퀀싱 리드의 5' 색인 서열, 3' 색인 서열 및 표적 서열을 확인하고;

상기 5' 색인 서열이 복수의 샘플들 중 한 샘플과 연관된 복수의 5' 색인 서열들 중 하나임을 결정하고;

상기 3' 색인 서열이 복수의 샘플들 중 상기 샘플과 연관된 복수의 3' 색인 서열들 중 하나임을 결정하되, 여기서 상기 복수의 5' 색인 서열들 및 복수의 3' 색인 서열들은 서로, 또 상기 시퀀싱 리드와 상기 샘플을 연관짓기 위해 상기 복수의 샘플들 중 다른 샘플들의 5' 색인 서열들 및 3' 색인 서열들과 구별가능하며; 그리고

상기 시퀀싱 리드의 표적 서열이 복수의 샘플들 중 상기 샘플과 연관되어 있다는 표시를 생성하도록 구성되고, 여기서 상기 샘플과 연관된 복수의 시퀀싱 리드중 하나의 시퀀싱 리드 세트는 상기 5' 색인 서열과 상기 3' 색인 서열의 적어도 2개의 상이한 조합을 포함하고, 상기 샘플과 연관된 복수의 시퀀싱 리드의 다른 하나의 세트

는 상기 5' 색인 서열과 상기 3' 색인 서열의 동일한 조합을 포함하는, 시퀀싱 장치.

청구항 14

제13항에 있어서, 프로세서는 메모리로부터 상기 복수의 5' 색인 서열의 서열들에 액세스함으로써 상기 5' 색인 서열이 상기 샘플과 연관된 복수의 5' 색인 서열들 중 하나임을 결정하도록 구성된, 시퀀싱 장치.

청구항 15

제13항 또는 제14항에 있어서, 프로세서는 메모리로부터 복수의 3' 색인 서열의 서열들에 액세스함으로써 상기 3' 색인 서열이 상기 샘플과 연관된 복수의 3' 색인 서열들 중 하나임을 결정하도록 구성된, 시퀀싱 장치.

청구항 16

제13항 또는 제14항에 있어서, 상기 표시를 전시하도록 구성된 디스플레이를 포함하는, 시퀀싱 장치.

청구항 17

제13항 또는 제14항에 있어서, 프로세서는

상기 복수의 시퀀싱 리드 중 제2 시퀀싱 리드의 5' 색인 서열, 3' 색인 서열 및 표적 서열을 확인하고;

상기 5' 색인 서열이 복수의 샘플들 중 상기 샘플과 연관된 복수의 5' 색인 서열들 중 하나임을 결정하고;

상기 3' 색인 서열이 상기 복수의 샘플들 중 상기 샘플과 연관된 복수의 3' 색인 서열들 중 하나가 아님을 결정하고; 그리고

상기 표적 서열이 상기 복수의 샘플들 중 상기 샘플과 연관되어 있지 않다는 표시를 생성하도록 구성된 것인, 시퀀싱 장치.

청구항 18

제13항 또는 제14항에 있어서, 프로세서는

상기 복수의 시퀀싱 리드 중 제3 시퀀싱 리드의 5' 색인 서열, 3' 색인 서열 및 표적 서열을 확인하고;

상기 5' 색인 서열이 상기 복수의 샘플들 중 제2 샘플과 연관된 복수의 5' 색인 서열들 중 하나임을 결정하고;

상기 3' 색인 서열이 상기 복수의 샘플들 중 상기 제2 샘플과 연관된 복수의 3' 색인 서열들 중 하나임을 결정하고; 그리고

상기 표적 서열이 상기 복수의 샘플들 중 상기 제2 샘플과 연관되어 있다는 표시를 생성하도록 구성된 것인, 시퀀싱 장치.

청구항 19

제13항 또는 제14항에 있어서, 시퀀싱 데이터가 상기 복수의 시퀀싱 리드의 각 염기에 대한 염기 호출(base call)을 포함하는, 시퀀싱 장치.

청구항 20

제13항 또는 제14항에 있어서, 상기 샘플과 연관된 상기 복수의 5' 색인 서열들 및 상기 복수의 3' 색인 서열들을 확인하는 입력(input)을 수신하도록 구성된 사용자 인터페이스를 포함하는, 시퀀싱 장치.

청구항 21

제13항 또는 제14항에 있어서, 시퀀싱 장치는 1-채널 시퀀싱 장치인, 시퀀싱 장치.

발명의 설명

기 술 분 야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 미국 가특허 출원 제62/582,175호(출원일: 2017년 11월 6일, 발명의 명칭: "NUCLEIC ACID INDEXING TECHNIQUES")의 우선권 및 이의 유익을 주장하며, 이 기초출원의 개시내용은 본 명세서에 모든 목적을 위하여 그 전문이 참고로 원용된다.

배 경 기 술

[0003] 본 개시내용은 일반적으로 생물학적 샘플과 관련된 데이터 분야, 예를 들면, 서열 데이터에 관한 것이다. 더 특히, 개시내용은 핵산을 색인하고, 획득된 서열 데이터에서 색인된 서열을 분해하는 기술에 관한 것이다.

[0004] 유전자 서열분석(genetic sequencing)은 유전자 연구의 점점 더 중요한 영역이 되고 있고, 진단 및 다른 응용분야에서 장래의 용도를 약속한다. 일반적으로, 유전자 서열분석은 핵산을 위한 뉴클레오타이드, 예를 들면, RNA 또는 DNA의 단편의 순서를 결정하는 것을 포함한다. 차세대 서열분석 기술은 풀링된(pooled) 샘플 제조물, 예를 들면, 다중 샘플 제조물을 사용하여 더 높은 처리량의 서열분석을 촉진한다. 풀링된 샘플 제조물에 있어서, 각각의 개별적인 샘플은 풀링된 샘플로부터의 각각의 서열분석 판독물(sequencing read)이 풀(pool)에서 개별적인 샘플과 연관되거나 이의 결과로 볼 수 있도록 태그가 붙거나, 그렇지 않으면 표시될 수 있다. 그러나, 획득된 서열 데이터는 다양한 출처, 예를 들면, 샘플 결함, 샘플 제조물 오류, 및 서열분석 편향으로부터 도입된 오류 및 노이즈를 가질 수 있고, 이는 각각의 개별적인 샘플에 연관된 서열 데이터의 정확성을 감소시킬 수 있다. 따라서, 감소된 귀인 오류와 함께 다중 샘플 제조물에서 고처리량 서열분석을 허용하는 방법을 개발하는 것이 바람직하다.

발명의 내용

[0005] 본 개시내용은 생물학적 샘플의 샘플 색인을 위한 신규한 접근법을 제공한다. 본 명세서에 제공된 바와 같이, 샘플 색인은 획득된 서열분석이 특정한 샘플과 연관된다는 공정이다. 특히, 개시된 색인 기술은 샘플 간의 오염을 도입하지 않고, 임의로 낮은 수의 샘플(포지티브 샘플 추적을 위하여 하나로 줄임)로 강력하게 실행하고, 심지어 다중화 샘플을 교차하는 표상을 지지한다. 색인 상호 오염의 문제, 저-플렉스(low-plex) 샘플 색인 성능, 및 가변적인 색인 성능은 단편당 이중 색인을 사용하여 제1 및 제2 색인에 대한 다중 색인의 잠재적인 풀을 사용하여 샘플로부터 제조된 각각의 핵산 단편을 표지화하는 색인 전략으로 해결된다. 전통적인 조합 접근법과 반대로, 개시된 기술은 복수의 고유 색인을 제1 및 제2 색인 둘 다에서 각각의 샘플에 할당한다. 하나의 실시형태에 있어서, 색인 서열은 샘플 확인 및/또는 샘플 연관을 위하여 샘플로부터 핵산으로 혼입될 수 있는 외인성 핵산 서열이다. 예를 들면, 본 명세서에 제공된 색인 서열은 샘플 제조 동안 그리고/또는 서열분석 동안 관심대상 샘플로부터 유래된 핵산 단편으로 혼입될 수 있다. 각각의 색인 서열은 서열분석 반응에서 사용된 다른 색인 서열과 비교하여 구별 가능한 서열을 갖는다. 특정한 실시형태에 있어서, 색인 서열은 샘플 또는 샘플들의 내인성 핵산과 구별 가능하다. 색인 서열은 단일 또는 이중 가닥일 수 있고, 적어도 5개의 염기, 적어도 6개의 염기, 적어도 7개의 염기, 적어도 8개의 염기, 또는 그 이상일 수 있다. 특정한 실시형태에 있어서, 색인 서열은 5 내지 8개의 염기, 5 내지 10개의 염기, 5 내지 15개의 염기, 5 내지 25개의 염기, 8 내지 10개의 염기, 8 내지 12개의 염기, 8 내지 15개의 염기, 또는 8 내지 25개의 염기 등의 길이이다.

[0006] 개시된 기술은 색인된 서열분석을 제공하고, 이로써 특정한 샘플로부터 유래된 핵산 단편은 각각의 색인 부위에서 색인 서열의 미리 정해진 풀을 사용하여 색인된다. 특정한 실시형태는 핵산 단편당 2개의 분리된 색인 부위(즉, 이중-색인)의 맥락에서 개시되지만, 개시된 기술은 적어도 1개의 색인 부위, 적어도 2개의 색인 부위를 갖는 핵산 단편과 함께 사용될 수 있고, 각각의 핵산 단편에 대하여 3개, 4개, 또는 그 이상의 색인 부위가 혼입될 수 있다는 것이 이해되어야 한다. 다른 색인된 샘플이 각각의 색인 부위에서 색인 서열의 상이한 풀로 색인되는 한, 본 명세서에 제공된 바와 같은 색인된 샘플은 서열분석을 위하여 다른 색인된 샘플로 풀링될 수 있다. 따라서, 본 기술은 서열분석 반응에서 각각의 개별적인 샘플이 색인 서열의 할당된 고유 세트로 색인되는 풀링된 또는 다중-샘플 서열분석 반응을 허용하고, 획득된 서열분석 판독물이 할당된 고유 색인을 통해 본래의 적절한 샘플과 연관된다.

[0007] 본 기술은 샘플 단편당 및/또는 각각의 색인 부위에서 단일 색인 서열을 혼입하는 기술에 관하여 서열분석 정확성을 개선시킨다. 예를 들면, 서열분석 분석은, 예를 들면, 이중-색인된 단편에 대하여 제1 및 제2 색인 부위에서 또는 단일-색인된 단편에 대하여 단 하나의 색인 부위에서, 적절한 할당된 색인을 포함하지 않는 임의의 서열분석 판독물을 폐기하는 규칙 기반의 할당을 사용하여 수행될 수 있다. 각각의 샘플에 대하여 선택된 할당된 색인은 다양한 서열분석 기술을 사용하여 다양하고 구별 가능하게 설계된다. 예를 들면, 각각의 색인 서열은 내부 가변성, 이의 할당된 세트 내의 가변성, 및 풀에서 다른 샘플에 할당된 다른 색인 서열에 관한 가변성을 갖

는다. 이러한 방식으로, 색인의 다양성이 잘못된 판독물을 감소시킬 가능성이 있기 때문에 색인에서 단일 염기 판독물에서 서열분석 장치 오류는 더 적다. 하나의 예에서, 서열분석 실행이 낮은 복합성 또는 적은 샘플을 갖는 경우, 색인 서열 데이터는 그렇더라도, 각각의 색인 부위에서 상이한 색인 서열의 수가 샘플의 수와 1:1 상응을 갖는 경우, 예상되는 것보다 더 높은 다양성을 반영할 수 있다. 즉, 샘플 색인 부위당 단 하나의 단일 색인 서열을 사용하여 이로써 각각의 샘플이 각각의 색인 부위에서 오직 하나의 색인과 연관되어 1:1 상응을 수득하는 것보다는, 각각의 샘플은 각각의 색인 부위에서 다중 잠재적인 색인 서열과 연관된다. 따라서, 상이하거나 구별 가능한 색인 서열은 다수의 샘플과 관련하여 각각의 색인 부위에 대하여 적어도 2:1, 적어도 3:1, 또는 적어도 4:1의 비율로 나타난다. 하나의 예에서, 전용 색인 판독물을 혼입하는 서열분석 기술(즉, 오직 색인 서열의 업스트림을 표적화하는 색인 프라이머를 사용하고, 색인 서열의 길이에 상응하는 제한된 수의 염기를 서열분석함)에 있어서, 본 기술 및 개선된 색인 다양성은 더 정확한 염기 호출(base calling)을 야기할 수 있다.

[0008] 색인 서열 데이터의 정확성을 증가시킴으로써, 더 적은 서열분석 판독물은 규칙 기반의 할당에 따라 폐기되고, 이는 또한 심지어 낮은 농도 샘플 및 낮은 샘플 수의 경우에 서열분석의 정확성을 개선시킨다. 추가로, 색인 서열분석의 개선된 정확성을 통해 개선된 샘플 연관은 색인 호핑(hopping)(즉, 색인의 잘못된 확인을 통한 샘플에 대한 서열분석 판독물의 잘못된 할당)의 효과를 감소시킨다.

[0009] 하나의 실시형태에 있어서, 제1 핵산 라이브러리가 제1 복수의 핵산 단편을 포함하고, 제1 복수의 각각의 핵산 단편이 색인 서열의 제1 세트로부터 선택된 적어도 2개의 상이한 색인 서열을 포함하는, 제1 샘플로부터 제조된 제1 핵산 라이브러리; 및 제2 핵산 라이브러리가 복수의 핵산 단편을 포함하고, 제2 복수의 각각의 핵산 단편이 색인 서열의 제1 세트와 구별 가능한 색인 서열의 제2 세트로부터 선택된 적어도 2개의 상이한 색인 서열을 포함하는, 제2 샘플로부터 제조된 제2 핵산 라이브러리를 포함하는 샘플-색인된 핵산 라이브러리 제조물이 제공되고, 여기서 색인 서열은 색인 서열의 제1 색인 서열이 표적 서열의 5'에 위치하고 색인 서열의 제2 색인 서열이 표적 서열의 3'에 위치하도록 제1 복수 및 제2 복수의 개별적인 핵산 단편 상에 배열된다.

[0010] 또 다른 실시형태에 있어서, 샘플로부터 생성된 복수의 이중-색인된 핵산 단편을 제공하는 단계로서, 여기서 핵산 단편의 각각의 개별적인 핵산 단편은 5' 어댑터 서열, 5' 색인 서열, 3' 어댑터 서열, 및 3' 색인 서열을 포함하고, 여기서 샘플과 연관된 5' 색인 서열의 제1 세트로부터 선택된 복수의 상이한 5' 색인 서열 및 샘플과 연관된 3' 색인 서열의 제2 세트로부터 선택된 복수의 상이한 3' 색인 서열은 이중-색인된 핵산 단편에서 나타나고, 복수의 상이한 5' 색인 서열과 복수의 상이한 3' 색인 서열은 서로 구별 가능한, 상기 복수의 이중-색인된 핵산 단편을 제공하는 단계; 이중-색인된 핵산 단편의 서열을 나타내는 서열분석 데이터를 생성하는 단계; 및 제1 세트로부터 선택된 5' 색인 서열 및 제2 세트로부터 선택된 3' 색인 서열 둘 다를 포함하는 경우에만, 서열의 개별적인 서열을 샘플과 연관시키는 단계를 포함하는, 핵산 분자의 서열분석 방법이 제공된다.

[0011] 또 다른 실시형태에 있어서, 복수의 핵산 단편을 포함하고, 각각의 단편이 색인 서열 및 어댑터 서열을 포함하는 다중-샘플 라이브러리 제조물 키트가 제공된다. 복수의 핵산 단편은 제1 색인 세트로부터 선택된 색인 서열을 갖고 어댑터 서열이 제1 어댑터 서열인 핵산 단편을 포함하고, 제2 색인 세트로부터 선택된 색인 서열을 갖고 어댑터 서열이 제2 어댑터 서열인 핵산 단편을 포함하고, 제1 색인 세트 및 제2 색인 세트의 각각의 색인 서열이 제1 샘플-연관된 핵산 단편 세트에서 나타나는, 제1 샘플-연관된 핵산 단편 세트; 및 제3 색인 세트로부터 선택된 색인 서열을 갖고 어댑터 서열이 제1 어댑터 서열인 핵산 단편을 포함하고, 제4 색인 세트로부터 선택된 색인 서열을 갖고 어댑터 서열이 제2 어댑터 서열인 핵산 단편을 포함하고, 제3 색인 세트 및 제4 색인 세트의 각각의 색인 서열이 제2 샘플-연관된 핵산 단편 세트에서 나타나는, 제2 샘플-연관된 핵산 세트를 포함하고; 여기서 제1 색인 세트, 제2 색인 세트, 제3 색인 세트 및 제4 색인 세트는 각각 서로 구별 가능한 복수의 색인 서열을 포함한다.

[0012] 또 다른 실시형태에 있어서, 기판 상에 고정화된 복수의 핵산 포획 서열을 포함하는 서열분석 기판이 제공되되, 여기서 각각의 개별적인 핵산 포획 서열은 제1 어댑터 서열에 상보적인 제1 포획 서열 또는 제2 어댑터 서열에 상보적인 제2 포획 서열을 포함한다. 서열분석 기판은 또한 복수의 핵산 포획 서열의 각각의 핵산 포획 서열에 커플링된 복수의 핵산 단편을 포함하고, 여기서 복수의 핵산 단편의 각각의 개별적인 단편은 제1 어댑터 서열 및 제2 어댑터 서열을 포함하고, 여기서 복수의 핵산 단편의 각각의 개별적인 단편은 고유 색인 서열의 제1 세트의 하나의 색인 서열에 상보적인 제1 서열, 및 고유 색인 서열의 제2 세트의 하나의 색인 서열에 상보적인 제2 서열을 포함하고, 여기서 고유 색인 서열의 제1 세트 및 고유 색인 서열의 제2 세트는 복수의 핵산 단편이 이로부터 유래되는 오직 하나의 샘플과 연관되고, 제1 세트 및 제2 세트의 각각의 고유 색인 서열은 복수의 핵산 단편의 적어도 하나의 핵산 단편에 나타난다.

[0013] 또 다른 실시형태에 있어서, 핵산 분자의 서열분석 방법은 샘플로부터 생성된 복수의 이중-색인된 핵산 단편을 제공하는 단계로서, 여기서 이중-색인된 핵산 단편의 각각의 개별적인 핵산 단편은 이중-색인된 핵산 단편을 생성하기 위하여 샘플로부터 유래된 관심대상 서열, 5' 어댑터 서열, 5' 색인 서열, 3' 어댑터 서열, 및 3' 색인 서열을 포함하고, 여기서 샘플과 연관된 5' 색인 서열의 제1 세트로부터 선택된 복수의 상이한 5' 색인 서열, 및 샘플과 연관된 3' 색인 서열의 제2 세트로부터 선택된 복수의 상이한 3' 색인 서열은 이중-색인된 핵산 단편에서 나타나고, 복수의 상이한 5' 색인 서열과 복수의 상이한 3' 색인 서열은 서로 구별 가능한, 복수의 이중-색인된 핵산 단편을 제공하는 단계; 관심대상 서열을 나타내는 시퀀싱 데이터를 생성하는 단계; 5' 색인 서열 및 3' 색인 서열을 나타내는 서열분석 데이터를 생성하는 단계; 및 관심대상 개별적인 서열이 제1 세트로부터 선택된 5' 색인 서열 및 제2 세트로부터 선택된 3' 색인 서열 둘 다와 연관된 경우에만 관심대상 개별적인 서열을 샘플에 할당하는 단계를 포함한다.

도면의 간단한 설명

[0014] 도 1은 본 기술에 따른 색인 핵산 단편의 도식적 묘사도;

도 2는 본 기술에 따른 색인된 핵산 단편 라이브러리의 도식적 묘사도;

도 3은 본 기술에 따른 샘플의 색인된 핵산 라이브러리의 생성 방법의 흐름도;

도 4는 본 기술에 따른 풀링된 색인된 핵산 라이브러이의 서열분석 방법의 흐름도;

도 5는 본 기술에 따른 풀링된 색인된 핵산 단편 라이브러리의 도식적 묘사도;

도 6은 본 기술에 따른 색인 서열을 포함하는 서열분석 어댑터의 도식적 묘사도;

도 7은 본 기술에 따른 핵산 라이브러리의 서열분석 방법의 흐름도;

도 8은 본 기술에 따른 색인 세트의 예를 도시한 도면;

도 9는 본 기술에 따른 색인 세트의 예를 도시한 도면;

도 10은 본 기술에 따른 핵산 서열분석 키트를 도시한 도면;

도 11은 본 기술에 따른 서열분석 데이터를 획득하도록 구성된 서열분석 장치의 블록 다이어그램;

도 12는 본 기술에 따른 그래픽 사용자 인터페이스의 예를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0015] 본 기술은 색인된 핵산 및 이를 사용하는 서열분석 및 분석에 관한 것이다. 핵산의 서열분석은 후속적으로 분석되고 서열분석된 샘플에 관한 의미있는 정보를 제공하도록 분석되는 다량의 미가공 데이터를 생성한다. 서열분석 기술은 다중 샘플을 동시에 처리하도록 발전하였고, 이는 시간 및 비용 절약을 제공한다. 그러나, 이러한 기술은 데이터 처리 문제는 나타낸다. 다중-샘플 서열분석 실행에서 획득된 서열분석 데이터는 반드시 서열 어셈블리 전 본래의 이의 샘플과 연관되어야 하고, 분석은 각각의 개별적인 샘플에 대하여 수행될 수 있다. 그러나, 자주 이러한 샘플 연관은 내인성 핵산 서열 만을 사용하여 정확하여 완료될 수 없다. 따라서, 특정한 서열분석 기술은 서열분석 전에 고유 외인성 바코드 또는 색인 서열을 핵산으로 혼입하는데, 여기서 각각의 샘플은 고유 바코드 또는 색인과 연관된다. 서열분석 데이터가 획득된 후, 고유 바코드 또는 색인을 갖는 서열 판독물을 본래의 적절한 샘플에 할당된다.

[0016] 이러한 기술이 다중-샘플 서열분석 데이터의 할당을 촉진하지만, 서열분석 판독물의 샘플 할당에서 오류는 여전히 발생한다. 도입된 샘플 할당 오류의 출처와 상관없이, 서열의 잘못된 할당은 후속적인 게놈 어셈블리에서 부정확성 및/또는 잘못 할당된 데이터 상에서 수행된 데이터 분석을 야기할 수 있다. 색인 서열 데이터가 서열분석 장치에 의해 정확하게 획득되지 않는 경우, 핵산 단편(예를 들면, 50 내지 300bp 핵산 단편을 나타냄)의 서열분석 판독물은 부정확한 데이터를 기반으로 잘못된 샘플에 할당될 수 있다. 추가로, 특정한 서열분석 기술은 더 큰 정도의 색인 호핑 또는 샘플 사이의 색인의 분자 재조합과 연관될 수 있다. 색인 호핑은 주형 라이브러리에서 분자 사이에서 이동하는 이러한 분자(색인 영역을 포함)의 하나의 말단에 의해 유발될 수 있다. 색인 호핑은 낮은 비율(대략 0.05%)로 발생할 수 있지만, 심지어 낮은 수준의 색인 호핑은 임상적 맥락에서 관련이 있을 수 있다. 단일 색인 판독물의 경우, 이동은 서열분석 판독물의 또 다른 샘플에 대한 부정확한 할당을 야기할 것이고, 이는 결국 다운스트림 분석에서 오염을 야기할 것이다. 색인 호핑은 또한 조합 접근법(예를 들면, 동일한

서열을 색인 1 또는 색인 2에서 다중 샘플에 할당하지만, 색인 1 및 색인 2의 임의의 특정한 조합이 특정한 샘플에 고유하다는 것을 보장하는 것)을 사용하는 이중 색인의 경우에도 발생할 수 있다. 특정한 실시형태에 있어서, 본 기술은 제1 및 제2 색인 판독물 내의 고유 색인과 함께 사용된다. 그 경우, 심지어 교체가 발생하는 경우에도, 관찰된 색인 조합은 색인 쌍의 예상된 세트의 일원이 아닐 것이다.

[0017] 감소된 서열분석 판독물의 잘못된 할당을 야기하는 인덱싱 기술이 본 명세서에 제공된다. 각각의 샘플이 단일 바코드 또는 색인과 연관되는 기술과 대조적으로, 본 기술은 각각의 개별적인 샘플과 고유하게 연관되고 샘플 제조 동안 도입되는 복수의 색인 서열을 제공한다. 샘플로부터 제조된 개별적인 핵산 단편이 적어도 하나의 색인 서열을 갖고 모든 상이한 상이한 색인 서열이 각각의 샘플의 핵산 단편의 라이브러리에 나타나도록, 복수의 색인 서열은 색인 부위(또는 색인 부위들)에 도입된다. 이러한 방식으로, 해당 샘플을 기반으로 색인 서열 다양성이 도입된다. 이러한 다양성은 획득된 색인 서열 데이터의 정확성을 개선시킬 수 있다. 특히, 동시에 서열분석되는 2개 이상의 샘플로부터의 핵산 단편의 수백(수천)개를 나타내는 이미지 데이터를 획득하는 서열분석 장치는 샘플이 뉴클레오타이드 다양성이 부족할 때 염기 호출을 결정하는데 어려움을 가질 수 있고, 이는 결국 해결하기 어려운 획득된 이미지 데이터를 야기할 수 있다. 예를 들면, 특정한 서열분석 기술은 염기 호출을 만드는 이미지 신호 강도에서 차이를 평가한다. 색인 부위당 단일 색인 서열만을 갖는 샘플에 있어서, 색인 서열의 각각의 서열분석 주기에서 뉴클레오타이드는 총 샘플 수만큼 다양하고, 낮은 샘플 수 실행에 대하여 불충분하게 다양할 수 있다. 따라서, 하나의 실시형태에 있어서, 본 명세서에 제공된 인덱싱 기술은 색인 다양성을 통해 색인 서열 데이터의 개선된 획득을 야기하고, 심지어 낮은 복합성의 서열분석 실행에 대하여 정확한 서열분석을 허용한다.

[0018] 그 목적을 위하여, 도 1은 본 기술에 따른 샘플의 색인된 핵산 단편(12)의 도식적 묘사이다. 색인된 핵산 단편(12)은 서열분석 실행에 적합한 단편을 나타낸다. 색인된 핵산 단편(12)은 샘플로부터 유래된 핵산, 즉, 내인성 핵산의 삽입물(14)을 포함한다. 색인된 핵산 단편(12)은 또한 서열분석을 촉진하는 도입된 또는 외인성 서열을 포함한다. 이러한 서열은 삽입물(14)의 5' 및 3'이고, 보편적인 서열분석 프라이머의 하나 또는 두 가닥에 대한 결합을 허용하는 하나 이상의 서열분석 프라이머 서열(16, 18)을 포함할 수 있다. 색인된 핵산 단편(12)은 또한 제1 색인 부위(20) 및 제2 색인 부위(22)를 포함한다. 도시된 실시형태에 있어서, 색인된 핵산 단편(12)은 또한 제1 어댑터 서열(26) 및 제2 어댑터 서열(28)을 포함한다. 어댑터 서열 또는 서열들(26, 28)은 원하는 서열분석 형식을 기반으로 선택될 수 있거나, 예를 들면, 색인된 핵산 단편(12)의 플로우 셀 또는 서열분석 기판 부착을 촉진하는, 도시된 바와 같은 P7 및 P5 어댑터일 수 있다.

[0019] 도시된 색인된 핵산 단편(12)은 이중 가닥 단편일 수 있고, 색인 서열(20, 22)이 삽입물의 측면에 위치하도록 정방향 가닥에서 제1 색인 서열(20)은 삽입물(14)의 5'에 위치할 수 있고, 제2 색인 서열(22)은 삽입물(14)의 3'에 위치할 수 있다. 제1 색인 서열(20)은 제1 색인 세트(30)로부터 선택되고, 이는 색인 서열(20a, 20b, 20c, 20d)로 도시된 복수(예를 들면, 2, 3, 4, 또는 그 이상)의 색인 서열을 포함한다. 제2 색인 서열(22)은, 존재하는 경우, 제2 색인 서열로부터 선택되고, 이는 색인 서열(22a, 22b, 22c, 22d)로 도시된 복수의 색인 서열을 포함한다.

[0020] 색인 세트(예를 들면, 색인 세트(30) 또는 색인 세트(32)) 내에서, 개별적인 색인 서열(예를 들면, 색인 서열(20a, 20b, 20c, 20d))은 서로 상이하다. 즉, 이들은 본 명세서에 제공되는 바와 같이 서로 상이한 서열을 갖는다. 추가로, 제1 색인 세트(30)에서의 색인 서열은 또한 제2 색인 세트(32)에서의 색인 서열과 구별 가능하다. 특정한 실시형태에 있어서, 색인 서열(20, 22)은 샘플 또는 샘플들의 내인성 핵산과 구별 가능할 수 있다. 그러나, 다른 실시형태에 있어서, 색인 서열은 반드시 삽입물(14)과 구별 가능할 필요는 없을 수 있다.

[0021] 색인 서열은 단일 또는 이중 가닥일 수 있고, 적어도 5개의 염기, 적어도 6개의 염기, 적어도 7개의 염기, 적어도 8개의 염기, 또는 그 이상일 수 있다. 특정한 실시형태에 있어서, 색인 서열은 5 내지 8개의 염기, 5 내지 10개의 염기, 5 내지 15개의 염기, 5 내지 25개의 염기, 8 내지 10개의 염기, 8 내지 12개의 염기, 8 내지 15개의 염기, 또는 8 내지 25개의 염기 등의 길이이다. 추가로, 특정한 실시형태에 있어서, 색인 서열(예를 들면, 색인 서열(20, 22))은 30개 이하의 염기, 25개 이하의 염기, 20개 이하의 염기, 15개 이하의 염기의 길이이다. 본 명세서에 제공되는 바와 같은 색인 서열의 길이는 각각의 색인 세트 내에서(그리고 다른 색인 세트와 비교하여) 서열의 고유/구별 가능한 부분을 지칭할 수 있고, 서열분석 프라이머로서 제공될 수 있고 샘플로부터 모든 색인된 핵산 단편(12) 사이에 공통인 색인된 핵산 단편(12)의 인접한 공통 또는 보편적인 서열을 배제할 수 있다는 것이 이해되어야 한다.

[0022] 개시된 기술의 특정한 실시형태는 이중-색인된 서열분석 기술의 맥락에서 논의되지만, 본 명세서에 제공된 기술

은 또한 단일-색인된 서열분석의 맥락에서도 사용될 수 있다는 것이 이해되어야 한다. 예를 들면, 핵산 단편(12)은 색인 세트(예를 들면, 색인 세트(30))로부터 선택된 오직 하나의 색인 서열(예를 들면, 색인 서열(20 또는 22))을 혼입할 수 있다. 추가로, 원하는 서열분석 기술에 따라, 예를 들면, 쌍을 이룬 말단 또는 단일 판독물 서열분석에서 이중-가닥 핵산 단편(12)의 하나 또는 두 가닥으로부터 서열분석되는 색인 서열이 선택될 수 있다.

[0023] 도 2는 샘플로부터 유래되고 색인된 핵산 단편(12)을 포함하는 서열분석 라이브러리(40)의 도식적 묘사이다. 도시된 바와 같이, 라이브러리(40)는 색인 조합의 몇몇 상이한 구성을 갖는다. 예를 들면, 특정한 단편(12a)은 색인(20a) 및 색인(22a)의 조합으로 색인될 수 있고, 다른 단편(12b)은 색인(20b 및 22a)의 조합으로 색인될 수 있다. 샘플로부터 라이브러리(40)의 제조 동안, 각각의 색인 세트(예를 들면, 제1 색인 세트(30) 및 제2 색인 세트(32))에서 상이한 색인 서열(20, 22)을 포함하는 어댑터 혼합물은 임의의 정해진 삽입물(14)이 제1 색인 세트(30)의 개별적인 색인 서열(20a, 20b, 20c 또는 20d) 중 어느 하나 및 제2 색인 세트(32)의 개별적인 색인 서열(22a, 22b, 22c 또는 22d) 중 어느 하나에 의해 변형될 수 있도록 하는 일반적으로 무작위인 방식으로 삽입물(14)을 변형하는데 사용된다고 생각된다. 도 2는 라이브러리에 나타난 제1 색인 서열(20) 및 제2 색인 서열(22)의 가능한 조합의 부분을 보여준다. 추가로, 색인을 위해 사용되는 특정한 색인 세트에서 색인 서열의 총 수에 따라, 추가의 조합이 가능하다는 것이 이해되어야 한다. 특정한 샘플에 있어서, 각각의 상이한 삽입물(14)을 갖는 다중 색인된 핵산 단편(12)은 색인 서열(20, 22)의 동일한 구성 또는 조합을 가질 것이라는 것이 이해되어야 한다.

[0024] 도 3은 도 2의 색인된 라이브러리(40)의 제조 방법(50)의 실시형태의 흐름도이다. 관심대상 샘플의 획득(단계 52) 후, 샘플에서의 핵산을 단편화한다(단계 54). 단편화된 핵산은 색인 세트의 복수의 다양한 색인 서열과 접촉한다. 특히, 라이브러리 제조의 출발시, 하나 이상의 색인 세트는 특정한 샘플과 연관되거나 이에 할당될 수 있다. 그 다음, 샘플-할당된 색인 세트 또는 세트들만을 사용하여 라이브러리 제조물을 제조할 수 있다(단계 56). 특정한 실시형태에 있어서, 샘플-할당된 복수의 색인 서열, 예를 들면, 제1 색인 세트(30) 또는 제2 색인 세트(32)는 세트에서, 각각의 개별적인 색인 서열, 예를 들면, 색인 서열(20, 22)의 동등한 농도로 제공된다. 즉, 색인 세트(30)가 3개의 상이한 색인 서열(20a, 20b, 20c)을 갖는 경우, 이들은 서로에 대하여 1:1:1 비율로 제공되고, 하나의 실시형태에 있어서, 각각의 색인 서열(20)의 동등한 농도(또는 서로 관련하여 0.9 내지 1.1 범위의 대략 동등한 농도)를 갖는 혼합물로서 제공된다.

[0025] 단편화된 핵산은 샘플-할당된 색인 서열을 포함하는 어댑터에 의해 변형된다(단계 58)(예를 들면, 도 4에 도시된 바와 같음). 다른 실시형태에 있어서, 색인 서열은 어댑터를 추가하는 것과는 별개의 단계로 단편화된 핵산에 추가될 수 있다. 변형의 결과로서, 색인된 핵산 단편이 생성된다. 후속적으로 동등한 농도로 각각의 색인 세트의 개별적인 색인 서열의 제공은 색인된 핵산 단편(12) 내에서 각각의 개별적인 색인 서열의 비교적 동등한 혼입을 촉진할 수 있다.

[0026] 도 4는 샘플(60)로부터의 색인 핵산 단편(12)의 라이브러리의 라이브러리 제조물의 실시형태의 도식적 묘사이다. 그러나, 도시된 방법은 예시의 방식이고, 본 명세서에 제공된 바와 같은 색인된 핵산(12)이 다른 라이브러리 제조 기술, 예를 들면, 태그지정(tagmentation)을 사용하여 제조될 수 있다는 것이 이해되어야 한다. 특정한 실시형태에 있어서, 라이브러리는 서열분석 반응에서 사용될 수 있는 단일-가닥 주형 분자를 유도하는데 사용된다. 라이브러리는 이의 5' 및 3' 말단에서 공통 서열을 갖지만 하나 이상의 색인 부위 및 삽입물 부위에서 다양한 핵산 단편으로부터 형성될 수 있다. 하기 추가로 상세하게 설명되는 바와 같이, 라이브러리 내의 핵산 단편은 이의 5' 및 3' 말단에서(또는 이에 근접하게) 공통 서열의 영역을 함유할 수 있다. 특정한 실시형태에 있어서, 라이브러리의 핵산 단편은 라이브러리에서 각각의 개별적인 주형의 5' 말단에서 공통 서열이 상기 주형의 3' 말단에서의 공통 서열과 동일하지 않으며 이에 완전히 상보적이지 않도록 "갈라진다(forked)". 그러나, 다른 실시형태에 있어서, 어댑터는 갈라지지 않고, 예를 들면, 완전히 상보적이다.

[0027] 샘플(60)을 단편화하여 당업자에게 공지된 다수의 방법에 의해 블런트-말단화된(blunt-ended) 돌출형 말단을 갖는 단편화된 핵산(62)을 생성한다. 하나의 방법에서, 단편화된 DNA의 말단을 T4 DNA 중합효소 및 클레노브 중합효소에 의해 말단 복구한 다음, 폴리뉴클레오타이드 키나제 효소에 의해 인산화하여 인산화된 단편화된 핵산(64)을 생성한다. 그 다음, 단일 'A' 데옥시뉴클레오타이드를 Taq 중합효소 효소를 사용하여 DNA 분자의 양 3' 말단에 추가하여, 갈라진 어댑터의 이중-가닥 말단 상에 돌출된 1-염기 3' T'에 상보적인 1-염기 3' 돌출형 단편화된 핵산(68)을 생성한다.

[0028] 그 다음, 어댑터의 2개의 카피를 각각의 DNA 단편에 하나의 말단에 하나씩 결합시키는 적합한 리가제 효소(예를

들면, T4 DNA 리가제)를 사용하여 갈라진 어댑터(70)와 1-염기 3' 돌출형 단편화된 핵산(68) 사이의 결찰 반응을 수행하여 어댑터-표적 구조물(72)을 형성한다. 이러한 반응의 생성물은 바람직하게는 아가로스 겔 슬래브를 통한 전기영동 후, 어댑터의 크기보다 크기가 큰 DNA를 함유하는 아가로스의 부분의 제거에 의한 크기-포함 크로마토그래피를 포함하는 다수의 기술에 의해 결찰되지 않은 어댑터로부터 정제할 수 있다.

[0029] 특정한, 도시된 실시형태에 있어서, 라이브러리 제조물에서 구현된 갈라진 어댑터(70)는 제1 색인 서열(20) 및 제2 색인 서열(22)에서 다양한 색인 혼합물을 포함한다. 어댑터가 샘플에 커플링되거나 삽입물 핵산이 일반적으로 동일한 다른 기술과 대조적으로, 갈리진 어댑터(70)는, 예를 들면, 색인 세트(30) 및 색인 세트(32)으로부터의 색인 서열의 혼합물을 사용하여 제조된다. 따라서, 갈라진 어댑터(70)는 서로 모두 동일하지 않고, 제1 색인 서열(20) 및 제2 색인 서열(22)에서 색인 서열의 특정한 조합을 기반으로 다양하다. 그러나, 서열분석 프라이머(16, 18) 및 어댑터 서열은 갈라진 어댑터(70) 사이에서 공유되거나 공통적이다. 어댑터-표적 구조물(72)은 증폭되어 색인된 핵산 단편(12)을 생성할 수 있고, 이는 결국 변성되어 서열분석 전에 이중-가닥 구조를 단일-가닥 분자로 분리할 수 있다.

[0030] 색인된 핵산 단편(12)의 라이브러리(40)는 원하는 서열분석 형식에 따라, 단일 샘플로부터 제조되고, 다른 각각의 샘플로부터의 다른 라이브러리에 의해 풀링되고, 유사한 방식으로 제조될 수 있다. 도 5는 제1 색인 세트(30a, 30b, 30c)에 존재하는 구별 가능한 색인 서열, 및/또는 제2 색인 세트(32a, 32b, 32c)에 존재하는 구별 가능한 색인 서열을 통해 후속-서열분석을 풀링하고 구별할 수 있는 전용 또는 할당된 상이한 색인 세트(30, 32)를 갖는 라이브러리의 도식적 묘사이다. 라이브러리(40a)의 생성에서 사용되는 제1 색인 세트(30a)는 다른 라이브러리(40b, 40c)에 존재하는 다른 색인 세트(30b, 32b, 30c, 32c)에 존재하지 않고 겹치지 않는 색인 서열의 별개의 세트를 포함한다. 또한, 제1 색인 세트(30a)는 또한 동일한 라이브러리(40a)에 사용되는 제2 색인 세트(32a)와 별개이고 겹치지 않는다는 것이 이해되어야 한다.

[0031] 도 6은 상이한 샘플로부터의 풀링된 라이브러리(예를 들면, 도 5의 라이브러리)를 서열분석하고, 본 명세서에 제공된 바와 같은 색인 서열을 사용하여 서열분석 데이터를 정확한 샘플에 할당하는 방법(80)의 흐름도이다. 단계(82)에서, 색인된 개별적인 샘플 단편(예를 들면, 색인된 핵산 단편(12))을 복수의 색인 서열을 포함하는 적어도 하나의 색인 세트를 사용하여 개별적인 샘플로부터 제조한다. 색인된 개별적인 샘플 단편을 상이한(즉, 단계 82에서 언급된 색인과 구별 가능한) 색인 세트를 사용하여 제조된 상이한 샘플로부터의 다른 색인된 샘플 단편으로 풀링하여 단계 84에서 조합된 샘플 단편을 생성한다. 풀링 또는 조합은 서열분석 기판 상의 샘플 로딩 동안 발생할 수 있다. 하나의 실시형태에 있어서, 조합된 샘플은 플로우 셀의 상이한 레인으로 로딩된다. 플로우 셀의 레인이 서열분석 동안 서로 분리되기 때문에, 동일한 레인 내의 샘플이 동일한 색인 세트를 사용하여 색인되지 않는 한, 플로우 셀의 제1 레인에서의 샘플에 사용된 색인 세트는 상이한 레인의 다른 샘플을 색인하는데 사용될 수 있다.

[0032] 단계 86에서, 조합된 샘플 단편의 서열분석을 나타내는 서열분석 데이터가 획득되고, 단계 88에서, 서열분석 판독물은 서열분석 판독물이 할당된 색인 세트의 색인 서열을 포함하는 경우에만 개별적인 샘플과 연관된다. 특정한 실시형태에 있어서, 색인된 샘플 단편 각각이 제1 색인 서열 및 제2 색인 서열을 갖는 경우, 규칙 기반의 할당은 제1 색인 서열 및 제2 색인 서열 둘 다가 정해진 샘플에 대하여 할당된 색인 세트의 일원임을 요구한다. 오직 하나의 색인 부위에서의 실패는 색인 호핑에 의해 판독물을 제거하는 다운스트림 분석(예를 들면, 게놈 어셈블리)으로부터의 서열분석 판독물의 제거를 야기하는데 충분하다. 특정한 실시형태에 있어서, 실패한 서열분석 판독물은 품질 평가를 위하여 저장될 수 있다. 즉, 특정한 잠재적인 색인 서열은 더 큰 색인 실패와 연관될 수 있고, 이들은 재설계를 위하여 추적될 수 있다.

[0033] 도 7은 본 명세서에 제공된 바와 같은 서열분석 데이터를 획득하기 위하여 사용되는 서열분석 기술(90)의 실시형태의 도식적 묘사이다. 도시된 바와 같이, 서열분석된 핵산은 색인된 핵산 단편(12)의 변성으로부터 유도된 주형 가닥(96)과 함께 사용되는 것에 상보적인 포획 프로브(94)를 통해 기판(92) 상에 고정화된다. 제1 서열분석 판독물인 판독물 1은 서열분석 프라이머 서열(16, 18) 중 하나를 표적화하는(즉, 이에 상보적인) 판독물 1 프라이머(98)와 접촉을 통해 획득되는 삽입물(14)의 서열이다. 예를 들면, 주형 가닥(96)이 p7(또는 다른 5') 어댑터에 상보적인 포획 프로브(94)를 사용하여 포획되는 경우, 판독물 1 프라이머(98)는 서열분석 프라이머 서열(18)에 상보적이다.

[0034] 리드 1 생성물의 제거 후, 기술(90)은 또한 서열분석 프라이머(16)에 대하여 표적화될 수 있는 제1 색인 프라이머(100)를 사용하여 삽입물의 p7 측 상에서 색인 서열(20)의 제1 색인 판독물을 획득할 수 있다. 도시된 실시형태에 있어서, 제1 색인 판독물은 판독물 1과 동일한 가닥 상에 있다. 판독물 1 판독물은 100 내지 150개의 염기

를 가질 수 있고, 색인 판독물은 비교적 더 짧을 수 있고, 예를 들면, 8 내지 12개의 염기일 수 있거나, 색인 서열(20, 22)의 공지된 길이만큼 길 수 있다. 이러한 방식으로, 서열분석 리소스는 보존된다. 제1 색인 생성물의 제거 후 동일한 가닥으로부터 제2 색인 판독물을 획득하기 위하여, 포획 프로브(104)를 통해 3' 말단에 포획된 주형 가닥이 사용될 수 있다. 예를 들면, 제2 색인 판독물은 p5(또는 다른 3') 어댑터 또는 인접한 서열의 부분을 표적화하는 제2 색인 판독물 프라이머(106)에 의해 획득될 수 있다. 그러나, 다른 실시형태에 있어서, 제2 색인 판독물은 재합성된 상보적 가닥으로부터 획득될 수 있다. 제2 색인 판독물 생성물의 제거 후, 주형 가닥(96)에 상보적인 가닥(110)이 합성되고, 본래의 주형 가닥(96)은 제거된다. 후속적으로, 합성된 가닥은 판독물 2 프라이머(112)와 접촉하여 판독물 1 서열의 역 상보체인 판독물 2 서열을 수득한다. 제1 색인 프라이머(100), 제2 색인 프라이머(106), 및 판독물 1 및 판독물 2 프라이머(98, 112)는 주형 샘플의 유래 및 색인 서열과 상관없이 모든 주형 가닥(96)에 보편적인 것으로 생각된다. 획득된 제1 색인 및 제2 색인 판독물을 기반으로, 판독물 1 및 판독물 2 서열분석 데이터는 특정한 샘플과 연관될 수 있다.

[0035] 본 명세서에서 제공되는 바와 같이, 색인 세트는 색인 판독물에서 정해진 샘플에 대하여 추가의 다양성을 제공한다. 단일 색인 서열을 각각의 샘플에 할당하는 경우, 모든 뉴클레오타이드가 샘플의 수가 적을 때 색인 판독물의 임의의 정해진 주기로 나타나지 않을 위험성이 존재한다. 그것이 일어난 경우, 서열분석 실행은 강도 보정 및 염기 호출 알고리즘이 색인 주기에서 설계된 바에 따라 작동되지 않기 때문에 사용 가능한 데이터를 생성하는데 실패할 수 있고, 이는 판독물을 샘플에 할당하는 것의 실패를 야기한다. 이러한 시나리오의 가장 극단적인 예시는, 예를 들면, 포지티브 샘플 추적에 대하여 수행될 수 있고/거나 색인되지 않은 Phi X를 판독물로부터 제거하는 단일 샘플의 서열분석이다. 이러한 문제에 대한 해결책은 설계된 바와 같이 작동하는 일차 분석 파이프라인에 대한 충분한 뉴클레오타이드 표상이 존재하도록 각각의 샘플을 다중 색인 서열로 표지화하는 것이다. 도 8은 뉴클레오타이드(N_1, N_2 등)를 갖는 색인 세트(예를 들면, 색인 세트(30))의 표상이다. 색인 세트는 임의의 정해진 위치에 있어서 뉴클레오타이드가 색인 세트의 서열 내에서 다양하도록 선택될 수 있다. 예를 들면, 위치(120)에서, 뉴클레오타이드는 A, C, T, 및 G 중 3개 이상이 색인 세트를 교차하여 나타나도록 상이할 수 있다. 색인 세트(30)의 특정한 실시형태를 도시하는 도 9에서, 3개의 상이한 뉴클레오타이드가 위치(120)에서 나타나고, 4개의 상이한 뉴클레오타이드가 위치(122)에서 나타난다. 따라서, 색인 세트는 각각의 뉴클레오타이드 위치에서 적어도 3개의 상이한 뉴클레오타이드가 색인 세트를 교차하여 나타나도록 설계될 수 있다. 다른 실시형태에 있어서, 4개의 상이한 뉴클레오타이드가 색인 서열에서 뉴클레오타이드 위치의 적어도 절반에서 색인 세트를 교차하여 나타난다. 추가로, 특정한 실시형태에 있어서, 색인 서열은 내부적으로 다양하다. 즉, 개별적인 색인 서열의 인접한 뉴클레오타이드의 적어도 부분은 서로 상이하다.

[0036] 추가로, 특정한 색인은 다른 것도 수행할 수 없고, 이는 특정한 샘플이 심지어 샘플 입력의 농도가 동일한 경우에도 나타나지 않는 것을 야기한다. 다중 색인 서열을 각각의 샘플 상에 위치시키는 이득은 하나의 색인 서열의 불량한 성능의 전체 영향이 제한되는 것이다. 이러한 전략 내에서, 고성능 및 저성능 색인 서열을 함께 그룹화하여 상이한 샘플을 교차하는 표상의 균일성을 추가로 개선시키는 것이 가능하다.

[0037] 표 1 및 표 2는 색인 세트(예를 들면, 색인 세트(30, 32)) 및 본 기술에 따른 색인 세트를 만드는 개별적인 색인 서열(예를 들면, 색인 서열(20, 22))의 예시이다. 예를 들면, 각각의 개별적인 그룹 #하에 색인 서열(예를 들면, 색인 서열(20, 22))은 단일 색인 세트인 것으로 생각된다. 즉, 그룹 #0은 단일 색인 세트를 함께 형성하는 4개의 상이한 색인 서열 F7-001, F7-002, F7-003 및 F7-004를 포함한다. 그룹 #1은 단일 색인 세트를 함께 형성하는 추가의 4개의 상이한 색인 서열 F7-005, F7-006, F7-007 및 F7-008을 포함하는 등이다.

[0038] 하기 나타낸 색인 세트에서, 4개의 색인은 색인 세트당 표시된다. 그러나, 색인 세트의 크기는 변할 수 있고, 3개, 4개, 또는 그 이상의 개별적인 구별 가능한 색인 서열을 포함할 수 있다는 것이 이해되어야 한다. 나타낸 색인 세트는 임의의 그룹 내에서 뉴클레오타이드의 균형 잡힌 표상이 존재한다는 것을 보장하도록 선택되었다. 구체적으로, 임의의 그룹 내에서, 각각의 주기에서 두 번째로 빈도가 적은 뉴클레오타이드는 반드시 색인 서열의 적어도 25%로 나타나야 한다.

[0039] 표 1은 P7-측 색인 세트(30)일 수 있는 색인 세트를 나타낸다. 그러나, 특정한 실시형태에 있어서, 표 1의 서열은 색인 세트(32)로서 P5 측 상에 사용될 수 있다.

표 1

P7 색인 세트
색인 그룹 #0
F7-001,CCATACTA
F7-002,TGTGCGCT
F7-003,CACATTGC
F7-004,ATCCGGAG
색인 그룹 #1
F7-005,ACCTTAAC
F7-006,CAGCGCCT
F7-007,TGAACAGG
F7-008,GTGGCTCA
색인 그룹 #2
F7-009,GGTAACAC
F7-010,TTCTGATG
F7-011,ACGCTTGT
F7-012,GAATAGCA

[0040]

색인 그룹 #3
F7-013,GGATTCAA
F7-014,ATTGAACT
F7-015,TCCAAGGC
F7-016,TAACCTTG
색인 그룹 #4
F7-017,GAGACAAC
F7-018,TGTTAGGA
F7-019,ACAGTATG
F7-020,CTTCTTCT
색인 그룹 #5
F7-021,GTCGCCTT
F7-022,CGGTGAGA
F7-023,TCACTGAA
F7-024,AGCAATTG
색인 그룹 #6
F7-025,TAGGTTGA

[0041]

F7-026,CCTTCGCC
F7-027,GGATAATG
F7-028,TTCACAAT
색인 그룹 #7
F7-029,TGGCACGG
F7-030,ATATCTAC
F7-031,GCTCGGTT
F7-032,CTAATGTA
색인 그룹 #8
F7-033,AAGATGAA
F7-034,GCCGAATC
F7-035,AGATGCGG
F7-036,TTCCATCC
색인 그룹 #9
F7-037,CCTCACGT
F7-038,GTGTGGAC
F7-039,TCCGTTCG

[0042]

F7-040,TGAGGATA
색인 그룹 #10
F7-041,TGATTGCC
F7-042,AACAGCTT
F7-043,CATCATAA
F7-044,TTGTCAGC
색인 그룹 #11
F7-045,CGGAACTT
F7-046,GTACTTGG
F7-047,AACGCACC
F7-048,GCTTAGAG
색인 그룹 #12
F7-049,TAATCGAT
F7-050,CCGGTCCA
F7-051,ATTAGTTC
F7-052,CGCGAAGG
색인 그룹 #13

[0043]

F7-053,CTCTAGCT
F7-054,TCGATCTG
F7-055,AATGGAGA
F7-056,CGGCCTAT
색인 그룹 #14
F7-057,TGCCTCTT
F7-058,CCAGCTGC
F7-059,AAGTGTCA
F7-060,GTTATACG
색인 그룹 #15
F7-061,GTAGAGGC
F7-062,TAGTCCTA
F7-063,CCTAGAAT
F7-064,GACCTCCG

[0044]

[0045] 표 2는 P5-측 색인 세트(32)일 수 있는 색인 세트를 나타낸다. 특정한 실시형태에 있어서, 색인 세트는 샘플에 대하여 색인된 핵산 단편을 생성시키는 동일한 그룹 수를 갖는 P7 색인 세트(30)와 함께 사용되도록 설계된다. 예를 들면, 특정한 쌍 세트는 함께 사용되는 경우에 완료된 품질 평가를 가질 수 있고, 낮은 수준의 색인 호핑, 또는 일반적으로 고르게 분포된 증폭 수율과 연관될 수 있다. 다른 실시형태에 있어서, 임의의 정해진 P7-측(또는 5'-측) 색인 세트(30)는 임의의 또 다른 P5-측 색인 세트(32)와 함께 사용될 수 있다.

표 2

P5 색인 세트
색인 그룹 #0
F5-001,TCGCTCTA
F5-002,ATTGGAGG
F5-003,AACTAGAC
F5-004,CGGACTAT
색인 그룹 #1
F5-005,TCCTTAGG
F5-006,AGGAGGAA
F5-007,CAACACTC
F5-008,GTGGTTCT
색인 그룹 #2
F5-009,TGGTGGTT
F5-010,GCTACCGC
F5-011,AGCGTTCA
F5-012,CAATTATG

[0046]

색인 그룹 #3
F5-013,TTACAAGA
F5-014,GGTTCTAC
F5-015,ACGTGGCG
F5-016,CACAATTG
색인 그룹 #4
F5-017,CCATTCAA
F5-018,TAGCAGGC
F5-019,TTCAGATG
F5-020,AGTGCGGT
색인 그룹 #5
F5-021,ATTCAACT
F5-022,TCAAGGAG
F5-023,CACCTCGA
F5-024,AGGTATCC
색인 그룹 #6
F5-025,AAGCGCTT

[0047]

F5-026,TGTGTAGC
F5-027,CTATAGAG
F5-028,GCCACTTA
색인 그룹 #7
F5-029,TGTCCTTG
F5-030,AACATGGT
F5-031,GTGTGCAA
F5-032,CCTGAACA
색인 그룹 #8
F5-033,CTTCTTGC
F5-034,TAGGAATA
F5-035,AGTTGCCT
F5-036,TCCACAAT
색인 그룹 #9
F5-037,AGACCTCT
F5-038,CTCATGTA
F5-039,CCGTACGC

[0048]

F5-040,GAAGGTCG
색인 그룹 #10
F5-041,AGATTGAT
F5-042,TCTGATTC
F5-043,ATCCGCCA
F5-044,GAGACGAG
색인 그룹 #11
F5-045,CGCTAACT
F5-046,ATAGCTAG
F5-047,GCACTGGC
F5-048,TCGAGCGT
색인 그룹 #12
F5-049,GGCAAGTT
F5-050,CTGTTCCG
F5-051,ACCGGTGC
F5-052,GATCCATA
색인 그룹 #13

[0049]

F5-053,TTCACTCC
F5-054,ACTTACTA
F5-055,GAAGCAAT
F5-056,GGTCGTGA
색인 그룹 #14
F5-057,GCCGTCAT
F5-058,TGTCGGCC
F5-059,ATGAATGG
F5-060,CATGGATT
색인 그룹 #15
F5-061,AGCCTAAG
F5-062,CAAGCTGA
F5-063,GTGTAATT
F5-064,TGTAACAC

[0050]

[0051] 도 10은 본 기술과 함께 사용될 수 있는, 즉, 색인된 핵산 단편(12)을 개별적인 샘플로부터 제조하고, 특정한 실시형태에 있어서, 색인된 핵산 단편(12)을 서열분석하는, 서열분석을 위한 샘플 제조 키트(150)의 예이다. 샘플 제조 키트(150)는 제1 색인 세트(30) 및, 사용되는 경우, 제2 색인 세트(32)를 포함할 수 있다. 특정한 실시형태에 있어서, 제1 색인 세트(30) 및/또는 제2 색인 세트(32)는 프라이머 서열, 어댑터 서열 등과 같은 추가의 요소를 포함하는 어댑터 핵산의 형태로 제공될 수 있다. 제1 색인 세트(30) 및/또는 제2 색인 세트(32)는 각각의 개별적인 색인 서열이 대략 동등한 농도로 존재하며 정해진 샘플에 대한 색인 다양성이 사용자 오류에 취약하지 않도록 미리 혼합된 양으로 각각의 개별적인 컨테이너 내에 제공될 수 있다. 어댑터가 갈라진 이중-색인된 어댑터인 실시형태에 있어서, 색인 세트(30, 32)는 둘 다 단일 어댑터 핵산에 나타날 수 있고, 이는 단일 컨테이너로 제공될 수 있다. 샘플 제조 키트(150)는 또한 원하는 서열분석 형식과 함께 사용을 위한 적절한 프라이머(152)를 포함할 수 있다. 샘플 제조 키트(150)는 또한 하나 이상의 샘플 제조 효소, 완충제 및/또는 시약(154)을 포함할 수 있다. 샘플 제조 키트(150)는 단일 샘플로부터 라이브러리를 제조하기 위하여 미리 포장된 키트로서 제공될 수 있거나, 특정한 실시형태에 있어서, 복수의 상이한 색인 세트(30, 32)와 함께 다중-샘플 키트로서 제공될 수 있다.

[0052] 도 11은 본 명세서에 제공된 바와 같은 색인 기술을 사용하여 개별적인 샘플에 할당된 색인된 핵산으로부터의 서열분석 데이터(예를 들면, 서열분석 판독물, 판독물 1, 판독물 2, 색인 판독물, 색인 판독물 1, 색인 판독물 2, 다중-샘플 서열분석 데이터)를 획득하기 위하여 개시된 실시형태와 함께 사용될 수 있는 서열분석 장치(160)의 도식적 도면이다. 서열 장치(160)는 임의의 서열분석 기술, 예를 들면, 미국 특허 공개 공보 제2007/0166705호; 제2006/0188901호; 제2006/0240439호; 제2006/0281109호; 제2005/0100900호; 미국 특허 제7,057,026호; 제WO 05/065814호; 제WO 06/064199호; 제WO 07/010,251호(이들의 개시내용은 그들의 전문이 본 명세서에 참고로 포함됨)에 기재된 합성에 의한 서열분석(sequencing-by-synthesis) 방법을 포함하는 것들에 따라 구현될 수 있다. 대안적으로, 결찰에 의한 서열분석 기술이 서열분석 장치(160)에서 사용될 수 있다. 이러한 기술은 DNA 리가제를 사용하여 올리고뉴클레오타이드를 혼입하고, 이러한 올리고뉴클레오타이드의 혼입을 확인하고, 미국 특허 제6,969,488호; 미국 특허 제6,172,218호; 및 미국 특허 제6,306,597호(이들의 개시내용은 그들의 전문이 본 명세서에 참고로 포함됨)에 기재되어 있다. 몇몇 실시형태는 나노포어 서열분석을 사용할 수 있

고, 이로써 표적 핵산 가닥, 또는 뉴클레오타이드는 표적 핵산으로부터 나노포어를 통해 핵산 말단 분해적으로 제거된다. 표적 핵산 또는 뉴클레오타이드가 나노포어를 통과함에 따라, 각각의 유형의 염기는 기공의 전기 전도도에서 변동을 측정함으로써 확인될 수 있다(미국 특허 제7,001,792호; 문헌[Soni & Meller, *Clin. Chem.* 53, 1996-2001(2007); Healy, *Nanomed.* 2, 459-481(2007)]; 및 [Cockroft, et al. *J. Am. Chem. Soc.* 130, 818-820(2008], 이들의 개시내용은 그 전문이 본 명세서에 참고로 포함됨). 다른 실시형태는 뉴클레오타이드의 신장 생성물로의 혼입시 방출되는 양성자의 검출을 포함한다. 예를 들면, 방출된 양성자의 검출을 기반으로 한 서열분석은 이온 토렌트(Ion Torrent)(라이프 테크놀로지스(Life Technologies) 자회사, 미국 코네티컷주 길퍼드 소재)로부터 상업적으로 이용 가능한 전기 검출기 및 연관된 기술, 또는 제US 2009/0026082 A1호; 제US 2009/0127589 A1호; 제US 2010/0137143 A1호; 또는 제US 2010/0282617 A1호(이들 각각은 그들의 전문이 본 명세서에 참고로 포함됨)에 기재된 서열분석 방법 및 시스템을 사용할 수 있다. 특정한 실시형태는 DNA 중합효소 활성의 실시간 모니터링을 포함하는 방법을 이용할 수 있다. 뉴클레오타이드 혼입은 형광단-함유 중합효소와 γ-포스페이트-표지화된 뉴클레오타이드 사이의 형광성 공명 에너지 전지(FRET) 상호작용을 통해, 또는 예를 들면, 문헌[Levene et al. *Science* 299, 682-686(2003); Lundquist et al. *Opt. Lett.* 33, 1026-1028(2008); Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176-1181(2008)](이들의 개시내용은 그들의 전문이 본 명세서에 참고로 포함됨)에 기재된 바와 같은 제로모드 도파관에 의해 검출될 수 있다. 다른 적합하고 대안적인 기술은, 예를 들면, 형광 동소 서열분석(FISSEQ), 및 초 병렬 서명 서열분석(MPSS)을 포함한다. 특정한 실시형태에 있어서, 서열분석 장치(160)는 일루미나(Illumina, 미국 캘리포니아주 라호이아 소재)로부터의 HiSeq, MiSeq, 또는 HiScanSQ일 수 있다. 다른 실시형태에 있어서, 서열분석 장치(160)는 DNA 증착이 각각의 포토다이오드와 일대일로 정렬되도록 포토다이오드 상에 제작된 나노웰을 갖는 CMOS 센서를 사용하여 작동되도록 구성될 수 있다.

[0053] 서열분석 장치(160)는 4개의 뉴클레오타이드 중 오직 2개만이 표지화되고 임의의 정해진 이미지에 대하여 검출 가능한 "1-채널" 검출 장치일 수 있다. 예를 들면, 티민은 영구적인 형광 표지를 가질 수 있고, 아데닌은 검출 가능한 형태로 동일한 형광 표지를 사용한다. 구아닌은 영구적으로 어두울 수 있고, 시토신은 초기에는 어둡지만 주기 동안 추가된 표지를 가질 수 있다. 따라서, 오직 티민 및 아데닌만이 초기 이미지에서 검출 가능하고, 오직 티민 및 시토신만이 제2 이미지에서 검출 가능하도록, 각각의 주기는 초기 이미지 및 염료가 임의의 아데닌으로부터 절단되고 임의의 시토신에 추가된 제2 이미지를 포함할 수 있다. 두 이미지 모두에서 어두운 임의의 염기는 구아닌이고, 두 이미지 모두에서 검출 가능한 임의의 염기는 티민이다. 제1 이미지에서 검출 가능하지만 제2 이미지에서는 그렇지 않은 염기는 아데닌이고, 제1 이미지에서는 검출 가능하지 않지만 제2 이미지에서는 검출 가능한 염기는 시토신이다. 초기 이미지와 제2 이미로부터의 정보를 조합하여, 하나의 채널을 사용하여 모든 4종의 염기는 식별될 수 있다.

[0054] 도시된 실시형태에 있어서, 서열분석 장치(160)는 분리된 샘플 처리 장치(162) 및 관련 컴퓨터(164)를 포함한다. 그러나, 기재된 바와 같이, 이들은 단일 장치로서 구현될 수 있다. 추가로, 관련 컴퓨터(164)는 로컬이거나, 샘플 처리 장치(162)에 네트워크화될 수 있다. 도시된 실시형태에 있어서, 생물학적 샘플은 샘플 기판(170), 예를 들면, 플로우 셀 또는 슬라이드 상의 샘플 처리 장치(162)로 로딩될 수 있고, 이는 이미지화되어 서열 데이터를 생성한다. 예를 들면, 생물학적 샘플과 상호작용하는 시약은 이미징 모듈(172)에 의해 발생한 여기 빔과 반응하여 특정한 파장에서 형광을 내고, 따라서 이미징을 위하여 방사선을 돌려보낸다. 예를 들면, 형광 성분은 성분의 상보적 분자 또는 중합효소를 사용하여 올리고뉴클레오타이드로 혼입되는 형광 표지된 뉴클레오타이드에 혼성화되는 형광 표지된 핵산에 의해 생성될 수 있다. 당업자에 의해 인식될 것인 바와 같이, 샘플의 염료가 여기되는 파장 및 이들이 형광을 내는 파장은 특정한 염료의 흡수 및 방출 스펙트럼에 따라 좌우될 것이다. 이러한 되돌아간 방사선은 지향 광학을 통해 뒤로 전파될 수 있다. 이러한 레트로빔은 이미징 모듈(172)의 검출 광학을 향하여 지향될 수 있다.

[0055] 이미징 모듈 검출 광학은 임의의 적합한 기술을 기반으로 할 수 있고, 예를 들면, 장치 내의 광자 충격 위치를 기반으로 픽셀 단위의 이미지 데이터를 생성하는 고체 촬상 소자(charged coupled device: CCD) 센서일 수 있다. 그러나, 시간 지연 적분(time delay integration: TDI) 작업을 위하여 구성된 검출기 어레이, 상보적 금속 산화물 반도체(CMOS) 검출기, 애벌란시 포토다이오드(avalanche photodiode: APD) 검출기, 가이거(Geiger)식 광자 계수기, 또는 임의의 다른 적합한 검출기를 포함하지만 이에 한정되지 않는 임의의 다양한 다른 검출기가 또한 사용될 수 있다는 것이 이해될 것이다. TDI 방식 검출은 미국 특허 제7,329,860호(본 명세서에 참고로 포함됨)에 기재된 바와 같은 선 주사와 커플링될 수 있다. 다른 유용한 검출기는, 예를 들면, 다양한 핵산 서열분석 방법론의 맥락에서 본 명세서에 상기 제공된 참조 문헌에 기재되어 있다.

[0056] 이미징 모듈(172)은 프로세서 제어하에, 예를 들면, 프로세서(174)를 통할 수 있고, 샘플 수용 장치(162)는 또한 I/O 컨트롤(176), 내부 버스(78), 비휘발성 메모리(180), RAM(82) 및 메모리가 실행 가능한 명령어를 저장할 수 있도록 하는 임의의 다른 메모리 구조 및 도 11에 관하여 기재된 것들과 유사할 수 있는 다른 적합한 하드웨어 구성요소를 포함할 수 있다. 추가로, 관련 컴퓨터(164)는 또한 프로세서(184), I/O 컨트롤(186), 통신 모듈(84), 및 메모리 아키텍처가 실행 가능한 명령어(192)를 저장할 수 있도록 RAM(188) 및 비휘발성 메모리(190)를 포함하는 메모리 아키텍처를 포함할 수 있다. 하드웨어 구성요소는 내부 버스(194)에 의해 연결될 수 있고, 이는 또한 디스플레이(196)에 연결될 수 있다. 서열분석 장치(160)가 일체형 장치로서 구현되는 실시형태에 있어서, 특정한 중복된 하드웨어 요소는 제거될 수 있다.

[0057] 프로세서(184)는 개별적인 서열분석 판독물을 본 명세서에 제공된 기술에 따라 연관된 색인 서열 또는 서열들을 기반으로 샘플에 할당하도록 프로그래밍될 수 있다. 특정한 실시형태에 있어서, 이미징 모듈(172)에 의해 획득된 이미지 데이터를 기반으로, 서열분석 장치(160)는 서열분석 판독물의 각각의 염기에 대하여 염기 호출을 포함하는 서열분석 데이터를 생성하도록 구성될 수 있다. 추가로, 이미지 데이터를 기반으로, 심지어 연속적으로 수행되는 서열분석 판독물의 경우에도, 개별적인 판독물은 이미지 데이터를 통해 동일한 위치에, 따라서, 동일한 주형 가닥에 연결될 수 있다. 이러한 방식으로, 색인 서열분석 판독물은 본래의 샘플에 할당되기 전에 삽입물 서열의 서열분석 판독물과 연관될 수 있다. 프로세서(184)는 또한 서열분석 판독물의 샘플에의 할당에 후속적으로 특정한 샘플에 대한 삽입물에 상응하는 서열에 대한 다운스트림 분석을 수행하도록 프로그래밍될 수 있다.

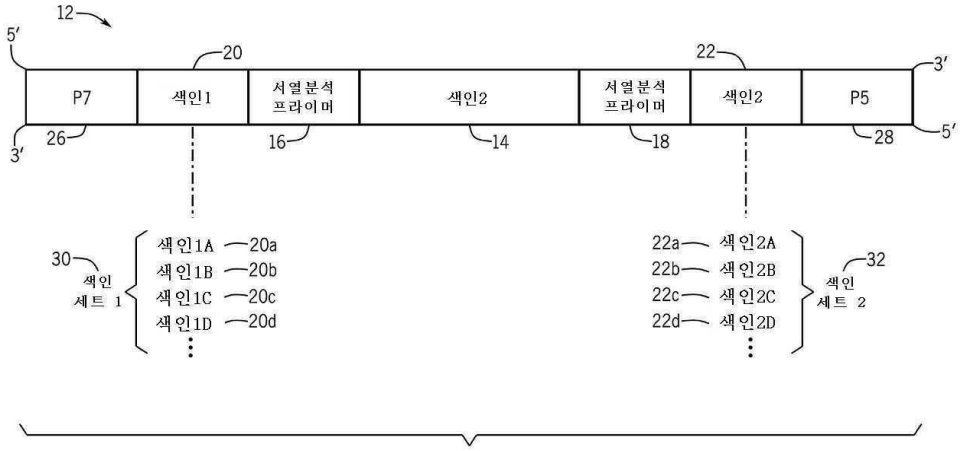
[0058] 도 12는 본 명세서에 제공된 바와 같은 색인된 핵산 단편을 사용하는 서열분석 반응에 관한 정보의 사용자 입력을 위하여 서열분석 장치(160)에 의해 생성될 수 있는 그래픽 사용자 인터페이스 스크린(200)의 예이다. 예를 들면, 사용자는 서열분석 실행에서 각각의 샘플의 명칭 또는 확인, 색인 부위의 수, 및 각각의 샘플에 사용되는 특정한 색인 세트 또는 세트들에 관한 입력을 제공할 수 있다. 하나의 실시형태에 있어서, 색인 세트는 상업적으로 이용 가능하고, 사용자 인터페이스 스크린(200)은 상업적으로 이용 가능한 색인 세트의 드롭-다운 메뉴를 제공한다. 그 다음, 각각의 개별적인 샘플은 선택된 하나 이상의 상업적으로 이용 가능한 색인 세트와 연관될 수 있다. 선택을 기반으로, 서열분석 장치(160)의 프로세서(예를 들면, 프로세서(184))는 메모리보터 선택된 색인 세트에 상응하는 저장된 색인 서열 정보에 접근하고, 접근된 색인 서열 정보를 사용하여 획득된 색인 서열 판독물을 서열분석 장치(160)에 의해 특정한 샘플에 할당한다. 일단 색인 서열 판독물 또는 판독물들을 기반으로 특정한 샘플에 할당되면, 삽입물을 나타내고 색인 서열 판독물의 이미지화된 위치와 연관된 서열분석 판독물은 특정한 샘플에 동시할당된다.

[0059] 개시된 실시형태의 기술적 효과는 개선되고 더 정확한 핵산 서열의 색인을 포함한다. 개선된 색인은 다중화(예를 들면, 다중-샘플)로부터 부정확하게 할당된 서열분석 판독물을 감소시켜 임상의에게 더 의미있는 정보를 줄 수 있다. 추가로, 본 기술과 연관된 서열분석 판독물 할당의 정확성에서의 개선은 상업적 및 시간 절약을 제공하는 고처리량 서열분석 전략을 촉진한다. 본 명세서에 제공된 바와 같은 색인 서열은 서열분석 범위 계수에 영향을 미치는 서열분석 데이터에 도입된 편향을 해결한다.

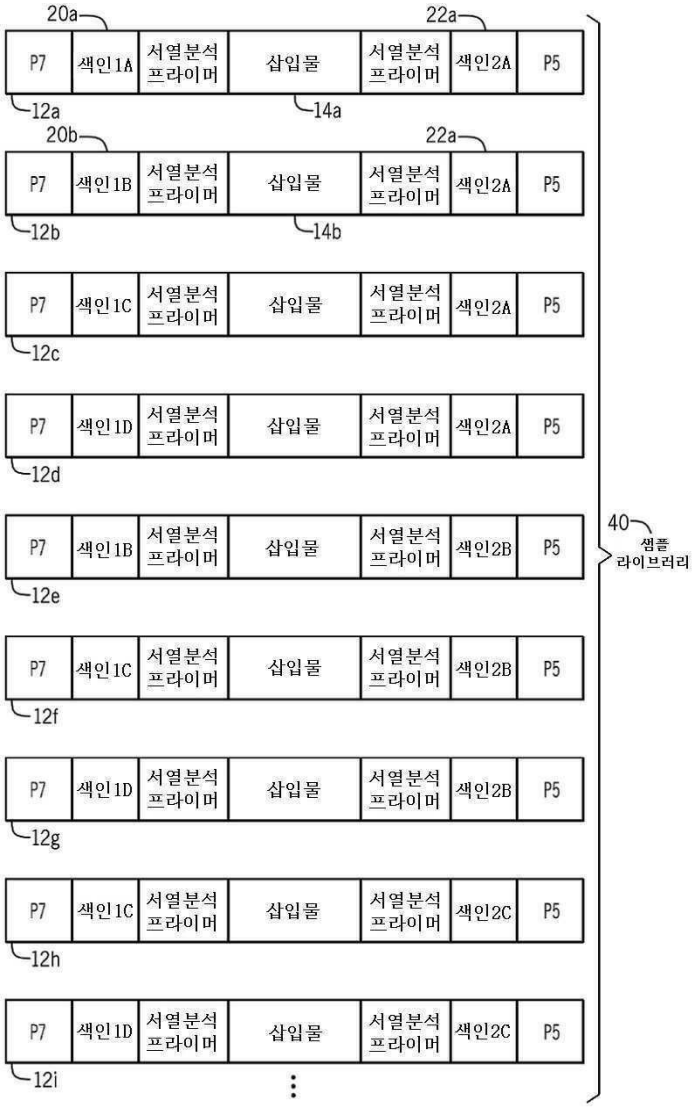
[0060] 본 개시내용의 오직 특정한 특징만이 본 명세서에서 예시되고 기재되었지만, 많은 변형 및 변화가 당업자에게 일어날 것이다. 따라서, 첨부된 청구범위는 본 개시내용의 진정한 취지에 속하는 모든 이러한 변형 및 변화를 포함한다고 의도되는 것으로 이해된다.

도면

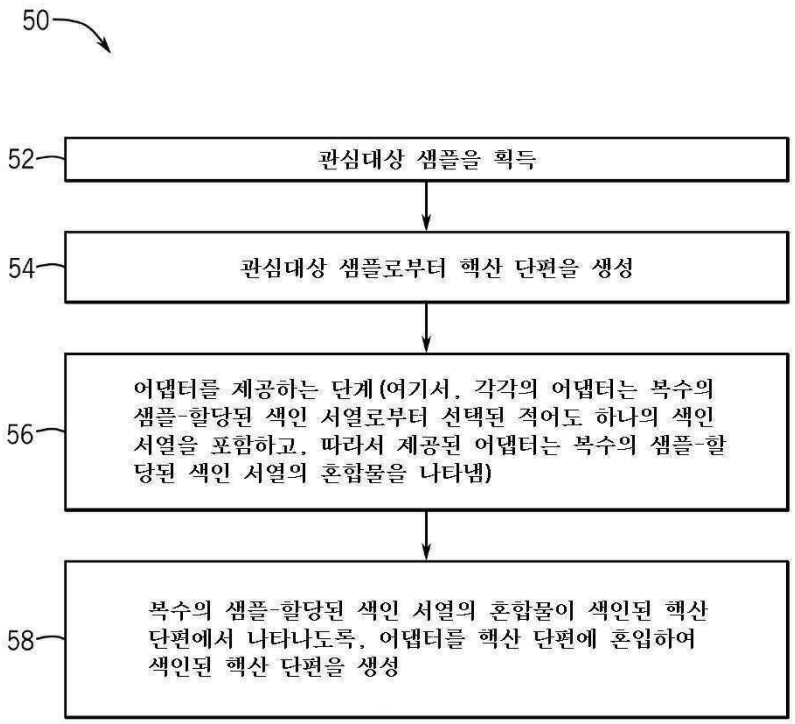
도면1


12
5′
20
22
3′
P7
색인1
서열분석 프라이머
색인2
서열분석 프라이머
색인2
P5
3′
26
16
14
18
28
5′
30
색인 세트 1
색인1A 20a
색인1B 20b
색인1C 20c
색인1D 20d
22a 색인2A
22b 색인2B
22c 색인2C
22d 색인2D
32
색인 세트 2

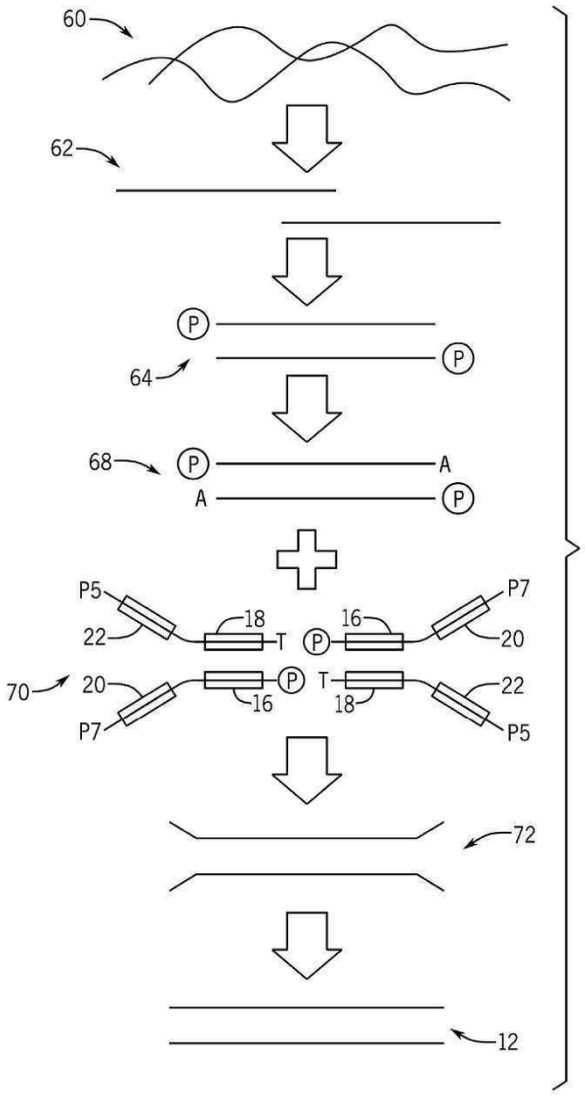
도면2


20a
22a
P7
색인1A
서열분석 프라이머
삽입물
서열분석 프라이머
색인2A
P5
12a
14a
20b
22a
P7
색인1B
서열분석 프라이머
삽입물
서열분석 프라이머
색인2A
P5
12b
14b
P7
색인1C
서열분석 프라이머
삽입물
서열분석 프라이머
색인2A
P5
12c
P7
색인1D
서열분석 프라이머
삽입물
서열분석 프라이머
색인2A
P5
12d
P7
색인1B
서열분석 프라이머
삽입물
서열분석 프라이머
색인2B
P5
12e
40
샘플 라이브러리
P7
색인1C
서열분석 프라이머
삽입물
서열분석 프라이머
색인2B
P5
12f
P7
색인1D
서열분석 프라이머
삽입물
서열분석 프라이머
색인2B
P5
12g
P7
색인1C
서열분석 프라이머
삽입물
서열분석 프라이머
색인2C
P5
12h
P7
색인1D
서열분석 프라이머
삽입물
서열분석 프라이머
색인2C
P5
12i

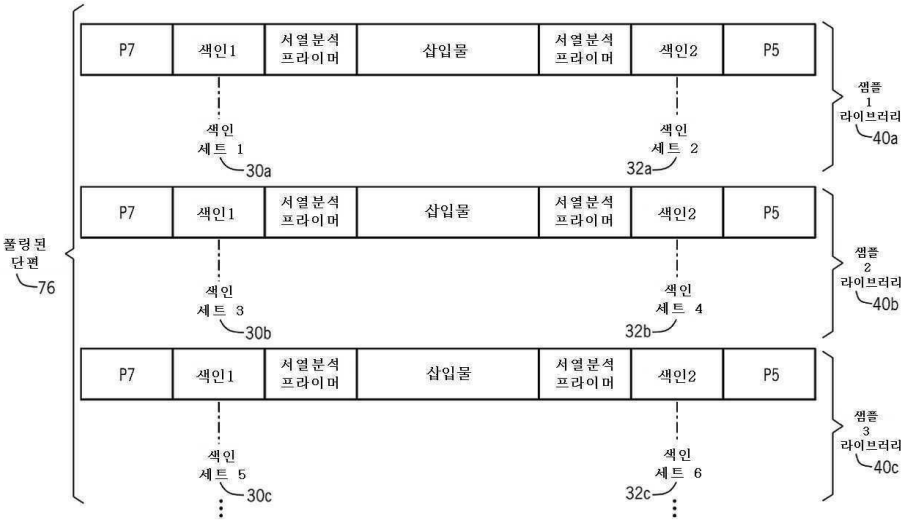
도면3


50
52
관심대상 샘플을 획득
54
관심대상 샘플로부터 핵산 단편을 생성
56
어댑터를 제공하는 단계(여기서, 각각의 어댑터는 복수의 샘플-할당된 색인 서열로부터 선택된 적어도 하나의 색인 서열을 포함하고, 따라서 제공된 어댑터는 복수의 샘플-할당된 색인 서열의 혼합물을 나타냄)
58
복수의 샘플-할당된 색인 서열의 혼합물이 색인된 핵산 단편에서 나타나도록, 어댑터를 핵산 단편에 혼입하여 색인된 핵산 단편을 생성

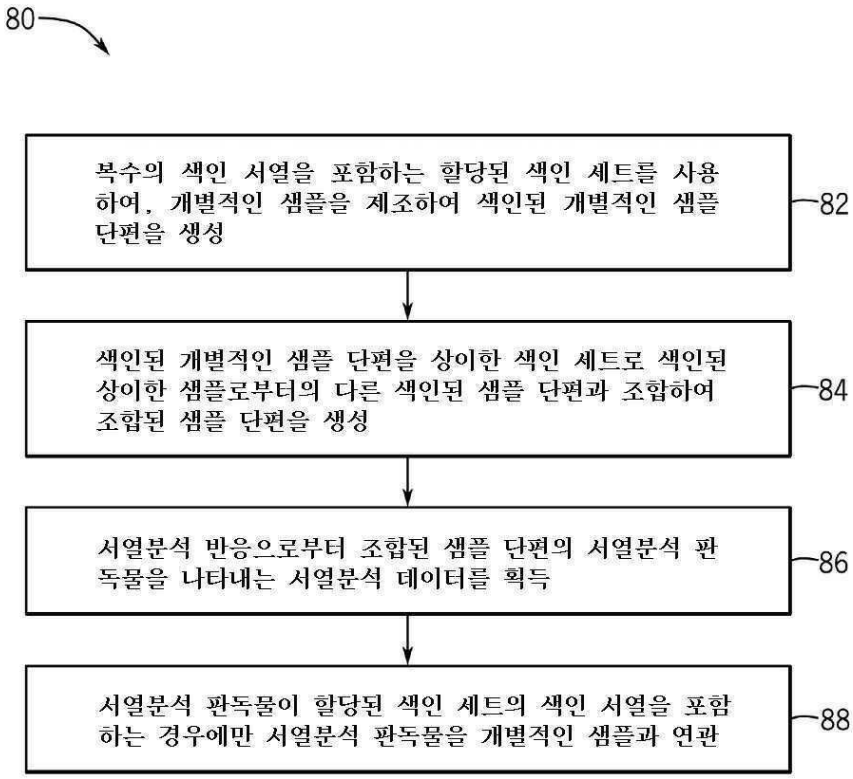
도면4


60
62
64
68
A
A
P5
P7
18
16
22
T
20
70
20
16
T
18
22
P7
P5
72
12

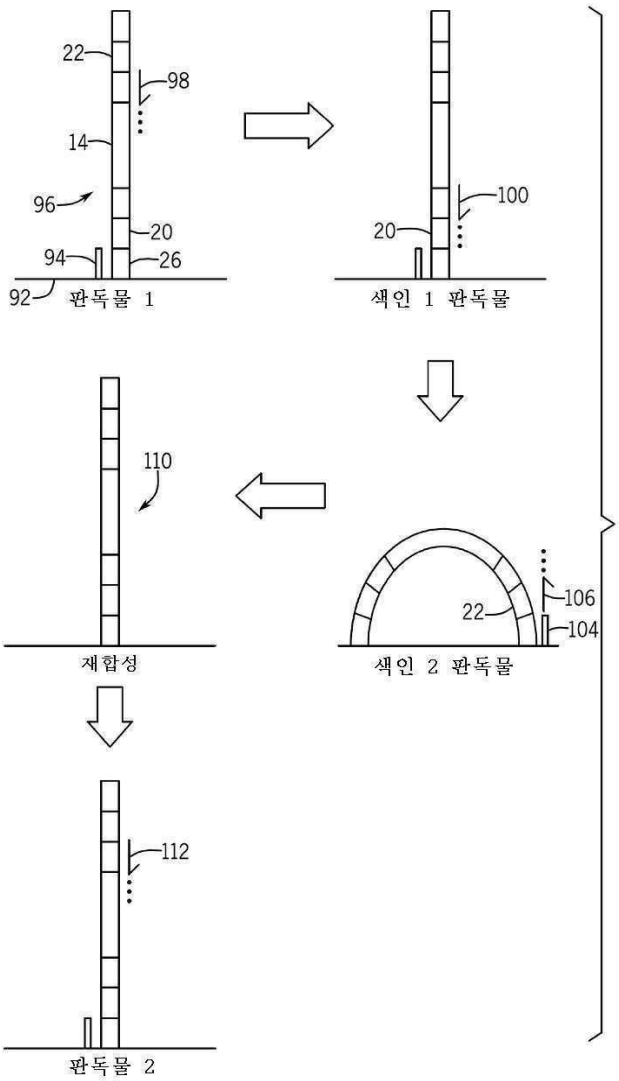
도면5


P7
색인1
서열분석 프라이머
삽입물
서열분석 프라이머
색인2
P5
색인 세트 1
30a
색인 세트 2
32a
샘플 1 라이브러리
40a
P7
색인1
서열분석 프라이머
삽입물
서열분석 프라이머
색인2
P5
색인 세트 3
30b
색인 세트 4
32b
샘플 2 라이브러리
40b
풀링된 단편
76
P7
색인1
서열분석 프라이머
삽입물
서열분석 프라이머
색인2
P5
색인 세트 5
30c
색인 세트 6
32c
샘플 3 라이브러리
40c

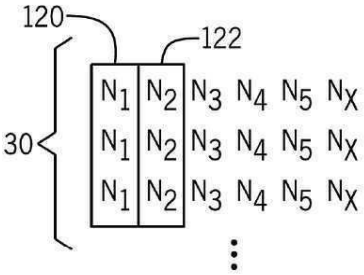
도면6


80
복수의 색인 서열을 포함하는 할당된 색인 세트를 사용하여, 개별적인 샘플을 제조하여 색인된 개별적인 샘플 단편을 생성
82
색인된 개별적인 샘플 단편을 상이한 색인 세트로 색인된 상이한 샘플로부터의 다른 색인된 샘플 단편과 조합하여 조합된 샘플 단편을 생성
84
서열분석 반응으로부터 조합된 샘플 단편의 서열분석 판독물을 나타내는 서열분석 데이터를 획득
86
서열분석 판독물이 할당된 색인 세트의 색인 서열을 포함하는 경우에만 서열분석 판독물을 개별적인 샘플과 연관
88

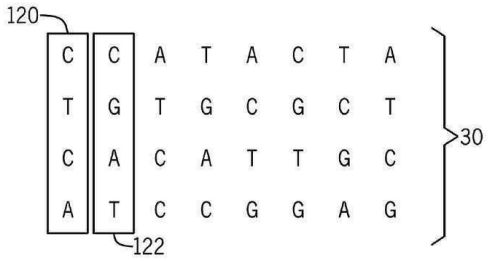
도면7


22
98
14
96
20
94
26
92
판독물 1
20
100
색인 1 판독물
110
재합성
22
106
104
색인 2 판독물
112
판독물 2

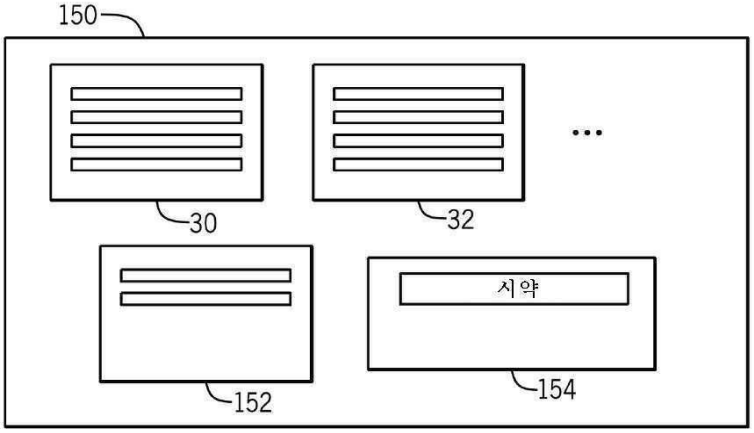
도면8


120
122
30
N1 N2 N3 N4 N5 NX
N1 N2 N3 N4 N5 NX
N1 N2 N3 N4 N5 NX

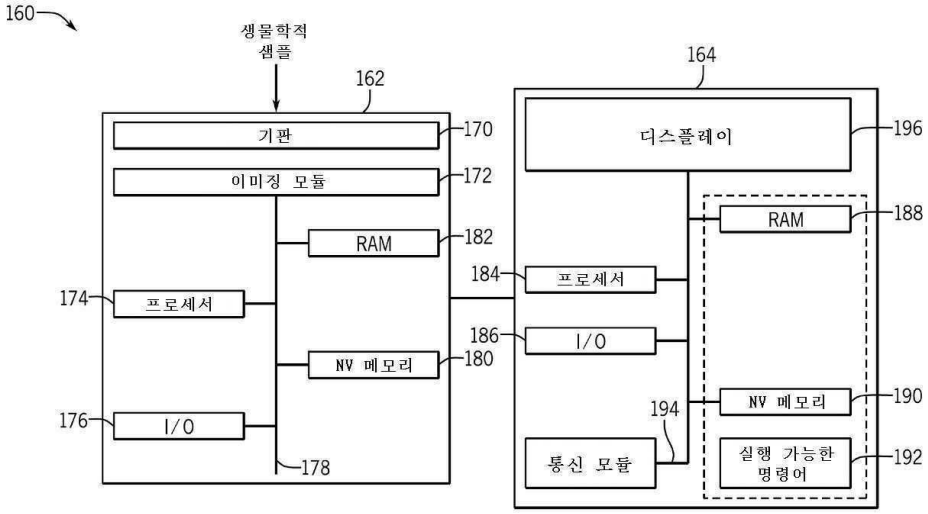
도면9



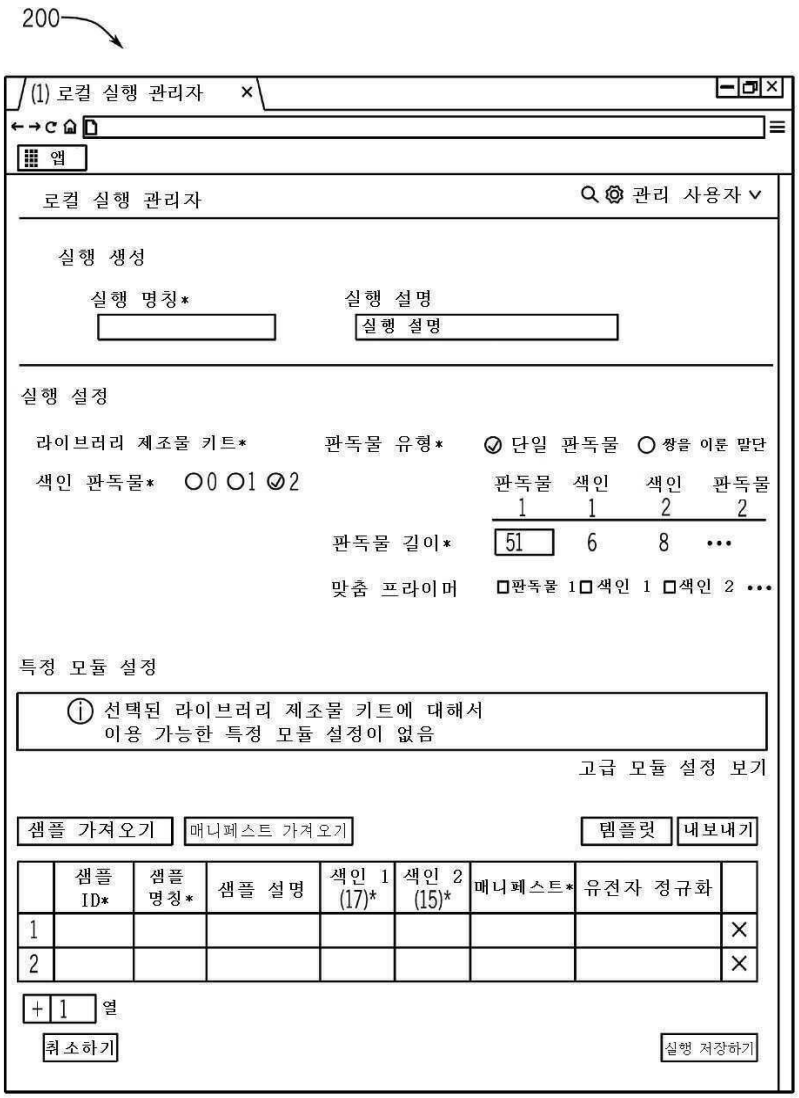
도면10



도면11



도면12



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 15

【변경전】

제13항 또는 제14항에 있어서, 프로세서는 상기 메모리로부터 복수의 3' 색인 서열의 서열들에 액세스함으로써 상기 3' 색인 서열이 상기 샘플과 연관된 복수의 3' 색인 서열들 중 하나임을 결정하도록 구성된, 시퀀싱 장치.

【변경후】

제13항 또는 제14항에 있어서, 프로세서는 메모리로부터 복수의 3' 색인 서열의 서열들에 액세스함으로써 상기 3' 색인 서열이 상기 샘플과 연관된 복수의 3' 색인 서열들 중 하나임을 결정하도록 구성된, 시퀀싱 장치.