



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0153382  
(43) 공개일자 2024년10월22일

(51) 국제특허분류(Int. Cl.)  
G11C 13/00 (2006.01) G11C 11/4063 (2006.01)  
G11C 11/409 (2015.01)  
(52) CPC특허분류  
G11C 13/0019 (2013.01)  
G11C 11/4063 (2013.01)  
(21) 출원번호 10-2024-7032021  
(22) 출원일자(국제) 2023년03월03일  
심사청구일자 없음  
(85) 번역문제출일자 2024년09월25일  
(86) 국제출원번호 PCT/US2023/014508  
(87) 국제공개번호 WO 2023/168085  
국제공개일자 2023년09월07일  
(30) 우선권주장  
63/316,812 2022년03월04일 미국(US)  
(뒷면에 계속)

(71) 출원인  
카탈로그 테크놀로지스, 인크.  
미국, 02129 매사추세츠, 찰스타운, 스위트 127,  
메인 스트리트 529  
(72) 발명자  
바라다라잘루, 가네슈쿠마르  
미국, 매사추세츠 02129, 찰스타운 스위트 127,  
메인 스트리트 529, 씨/오 카탈로그  
테크놀로지스, 아이엔씨.  
캄바라, 트레이시  
미국, 매사추세츠 02129, 찰스타운 스위트 127,  
메인 스트리트 529, 씨/오 카탈로그  
테크놀로지스, 아이엔씨.  
(뒷면에 계속)  
(74) 대리인  
강명구

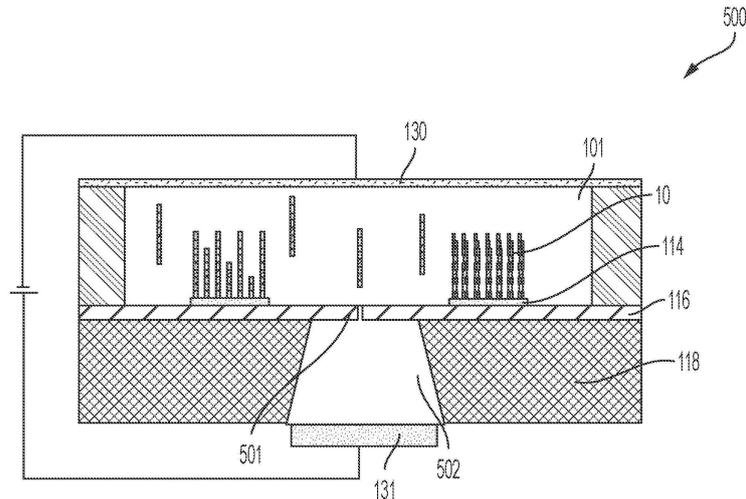
전체 청구항 수 : 총 153 항

(54) 발명의 명칭 **핵산-기반 데이터 저장 및 처리를 위한 DNA 마이크로어레이 및 구성요소 수준 시퀀싱**

(57) 요약

기술에는 핵산 분자(가령, DNA)를 사용하여 디지털 정보를 기록하고, 저장하고, 판독하며, 계산을 수행하는 시스템, 장치 및 방법이 포함된다. 예를 들어, 이러한 기술에는 디지털 정보를 기록하고, 저장하며, 불러오고, 판독하며, 계산/조작하기 위한 하나 이상의 개별 또는 블록 주소 지정 가능 전극 마이크로어레이 또는 나노어레이를 포함하는 장치가 포함된다.

대표도 - 도9a



(52) CPC특허분류

*G11C 11/409* (2018.05)  
*G11C 13/0023* (2013.01)  
*G11C 13/003* (2013.01)  
*G11C 13/004* (2013.01)  
*G11C 13/0069* (2013.01)

(30) 우선권주장

63/326,598 2022년04월01일 미국(US)  
63/329,111 2022년04월08일 미국(US)  
63/333,698 2022년04월22일 미국(US)

(72) 발명자

**바티아, 스왑닐 피.**

미국, 매사추세츠 02129, 찰스타운 스위트 127, 메인 스트리트 529, 씨/오 카탈로그 테크놀로지스, 아이엔씨.

**밈, 셴**

미국, 매사추세츠 02129, 찰스타운 스위트 127, 메인 스트리트 529, 씨/오 카탈로그 테크놀로지스, 아이엔씨.

**라미레즈-타피아, 루이스**

미국, 매사추세츠 02129, 찰스타운 스위트 127, 메인 스트리트 529, 씨/오 카탈로그 테크놀로지스, 아이엔씨.

---

## 명세서

### 청구범위

#### 청구항 1

디지털 정보를 핵산 서열로 변환하기 위한 시스템으로서, 상기 시스템은,

핵산 분자의 풀을 포함하는 유체를 보관하도록 구성된 출발지 저장소,

복수의 전기 전도성 판 및 상대 전극을 포함하는 메인 채널 - 상기 복수의 전기 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치됨 - ,

도착지 저장소,

출발지 저장소와 메인 채널과 유체 연통하는 입력 채널 - 상기 입력 채널은 출발지 저장소로부터의 제1 복수의 핵산 분자를 포함하는 제1 유체 볼륨을 메인 채널로 분배하도록 구성됨 - ,

상기 메인 채널 및 도착지 저장소와 유체 연통하는 출력 채널 - 상기 출력 채널은 메인 채널로부터의 제2 유체 볼륨을 도착지 저장소로 분배하도록 구성됨 - 을 포함하는, 시스템.

#### 청구항 2

제1항에 있어서, 상기 메인 채널은 반응 챔버를 포함하는, 시스템.

#### 청구항 3

제1항에 있어서, 상기 메인 채널은 반응 챔버인, 시스템.

#### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 복수의 셀 - 각 셀은 복수의 전기 전도성 판 중 하나씩 및 상대 전극의 일부분을 포함함 - 을 포함하는, 시스템.

#### 청구항 5

제1항 또는 제2항에 있어서, 각 셀은 기저 층 및 제1 유전체 층을 포함하며, 상기 제1 유전체 층은 상기 기저 층과 복수의 전기 전도성 판 사이에 배치되는, 시스템.

#### 청구항 6

제1항 또는 제2항에 있어서, 각 셀은 기저 층 및 제1 유전체 층 - 복수의 전기 전도성 판은 상기 기저 층과 제1 유전체 층 사이에 배치됨 - 을 포함하는, 시스템.

#### 청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 기저 층은 반도체 층을 포함하는, 시스템.

#### 청구항 8

제7항에 있어서, 상기 기저 층은 반도체 층에 부착된 비-반도체 층(non-semiconducting layer)을 포함하는, 시스템.

#### 청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 기저 층은 전도성 층을 포함하는, 시스템.

#### 청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 상기 기저 층은 절연 층을 포함하는, 시스템.

**청구항 11**

제1항 내지 제10항 중 어느 한 항에 있어서, 상기 기저 층은 투명한, 시스템.

**청구항 12**

제11항에 있어서, 상기 기저 층은 유리를 포함하는, 시스템.

**청구항 13**

제1항 내지 제12항 중 어느 한 항에 있어서, 상기 기저 층은 가열기 요소를 포함하는, 시스템.

**청구항 14**

제13항에 있어서, 상기 가열기 요소는 저항성 가열 요소를 포함하는, 시스템.

**청구항 15**

제4항 내지 제14항 중 어느 한 항에 있어서, 각 셀은 제2 유전체 층을 포함하고, 상기 제2 유전체 층은 상대 전극 상에 배치되는, 시스템.

**청구항 16**

제15항에 있어서, 복수의 전기 전도성 판과 제2 유전체 층은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되는, 시스템.

**청구항 17**

제4항 내지 제16항 중 어느 한 항에 있어서, 각 셀은 전기 전도성 판에 부착된 복수의 핵산 어댑터를 포함하는, 시스템.

**청구항 18**

제1항 내지 제17항 중 어느 한 항에 있어서, 제3 유전체 층을 포함하고, 상기 제3 유전체 층은 복수의 전기 전도성 판 상에 배치되는, 시스템.

**청구항 19**

제18항에 있어서, 각 셀은 제3 유전체 층에 부착된 복수의 핵산 어댑터를 포함하는, 시스템.

**청구항 20**

제4항 내지 제19항 중 어느 한 항에 있어서, 각 셀은 복수의 전기 전도성 판 중 하나 및 복수의 카운터 전극 중 하나를 포함하며, 복수의 상대 전극 각각은 복수의 전기 전도성 판 중 하나의 반대편에 배치되는, 시스템.

**청구항 21**

제16항 내지 제18항 중 어느 한 항에 있어서, 셀은 메인 채널의 제2 차원 및 제3 차원을 따라 2차원 어레이로 배열되고, 상기 어레이는 셀의 행 및 열을 갖는, 시스템.

**청구항 22**

제1항 내지 제21항 중 어느 한 항에 있어서, 각 전기 전도성 판은 전압원에 전기적으로 연결되는, 시스템.

**청구항 23**

제1항 내지 제22항 중 어느 한 항에 있어서, 복수의 스위치를 포함하는 제어 시스템 - 각 스위치는 복수의 전기 전도성 판 중 하나씩 및 전압원에 전기적으로 연결됨 - 을 포함하는, 시스템.

**청구항 24**

제23항에 있어서, 어레이의 각 셀 양단의 전압은 복수의 스위치 중 하나를 작동시킴으로써 개별적으로 제어 가

능한, 시스템.

**청구항 25**

제1항 내지 제24항 중 어느 한 항에 있어서, 유체는 유전체 유체인, 시스템.

**청구항 26**

제1항 내지 제24항 중 어느 한 항에 있어서, 유체는 전도성 유체인, 시스템.

**청구항 27**

제1항 내지 제26항 중 어느 한 항에 있어서, 메인 채널과 유체 연통하는 버퍼 저장소를 포함하는, 시스템.

**청구항 28**

제1항 내지 제27항 중 어느 한 항에 있어서, 메인 채널과 유체 연통하는 리가제 저장소를 포함하는, 시스템.

**청구항 29**

제1항 내지 제28항 중 어느 한 항에 있어서, 입력 채널, 메인 채널, 및 출력 채널 중 하나 이상을 통해 유체 볼륨을 펌프하도록 구성된 펌프를 포함하는, 시스템.

**청구항 30**

제1항 내지 제29항 중 어느 한 항에 있어서, 메인 채널을 통해 유체 볼륨의 흐름을 제어하기 위한 밸브를 포함하는, 시스템.

**청구항 31**

제1항 내지 제30항 중 어느 한 항에 있어서, 복수의 출발지 저장소 - 복수의 출발지 저장소의 각각은 실질적으로 동일한 핵산 분자의 모집단을 갖는 유체 볼륨을 가짐 - 를 포함하는, 시스템.

**청구항 32**

제31항에 있어서, 복수의 출발지 저장소의 각각은 실질적으로 동일한 핵산 분자의 상이한 모집단을 포함하는, 시스템.

**청구항 33**

제1항 내지 제32항 중 어느 한 항에 있어서, 메인 채널과 유체 연통하는 복수의 도착지 저장소를 포함하는, 시스템.

**청구항 34**

제1항 내지 제33항 중 어느 한 항에 있어서, 핵산 분자는 디지털 정보를 인코딩하는, 시스템.

**청구항 35**

제1항 내지 제34항 중 어느 한 항에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 인코딩하는 식별자 핵산 분자를 포함하는, 시스템.

**청구항 36**

제1항 내지 제35항 중 어느 한 항에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 인코딩하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 시스템.

**청구항 37**

제35항 또는 제36항에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 시스템.

**청구항 38**

제1항 내지 제37항 중 어느 한 항에 있어서, 메인 채널은 복수의 유체 연결된 반응 챔버를 포함하는, 시스템.

**청구항 39**

제21항 내지 제38항 중 어느 한 항에 있어서, 2차원 어레이는 둘 이상의 블록으로 분할되는, 시스템.

**청구항 40**

제39항에 있어서, 복수의 유체 연결된 반응 챔버의 각각은 하나씩의 블록을 하우징하는, 시스템.

**청구항 41**

제1항 내지 제40항 중 어느 한 항에 있어서, 제2 유체 볼륨은 제2 복수의 핵산 분자를 포함하는, 시스템.

**청구항 42**

제4항 내지 제41항 중 어느 한 항에 있어서, 제2 복수의 핵산 분자는 셀로부터 방출된 핵산 분자를 포함하는, 시스템.

**청구항 43**

디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,

청구항 제5항 내지 제42항 중 어느 한 항에 따른 시스템,

제1 유전체 층에 배치된 시퀀싱 장치 - 상기 시퀀싱 장치는 메인 채널과 유체 연통하는 유입구 및 캐비티(cavity)와 유체 연통하는 배출구를 가짐 -, 및

배출구의 다운스트림에 배치된 기저 전극을 포함하는, 시스템.

**청구항 44**

제43항에 있어서, 시퀀싱 장치는 나노포어(nanopore)를 포함하는, 시스템.

**청구항 45**

제43항에 있어서, 시퀀싱 장치는 나노채널을 포함하는, 시스템.

**청구항 46**

제43항에 있어서, 시퀀싱 장치는 고체 상태 멤브레인 내에 형성된 나노포어 또는 나노채널을 포함하는, 시스템.

**청구항 47**

제43항 내지 제46항 중 어느 한 항에 있어서, 캐비티는 메인 채널의 기저 층 내에 배치되는, 시스템.

**청구항 48**

제43항 내지 제47항 중 어느 한 항에 있어서, 메인 채널 및 캐비티는 전해질 용액을 포함하는, 시스템.

**청구항 49**

제44항 내지 제48항 중 어느 한 항에 있어서, 나노포어 또는 나노채널은 알파-헤몰리신( $\alpha$ HL) 또는 마이코박테리움 스메그마티스 포린 A(MspA)를 포함하는, 시스템.

**청구항 50**

제43항 내지 제49항 중 어느 한 항에 있어서, 기저 전극 및 상대 전극에 의해 생성된 전기장은 나노포어 또는 나노채널 양단에 100 mV 보다 큰 차동 전위를 갖는, 시스템.

**청구항 51**

제43항 내지 제50항 중 어느 한 항에 있어서, 복수의 나노포어 또는 나노채널을 포함하는, 시스템.

**청구항 52**

제51항에 있어서, 각 블록은 복수의 시퀀싱 장치 중 하나씩을 포함하는, 시스템.

**청구항 53**

디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,

청구항 제5항 내지 제42항 중 어느 한 항에 따른 시스템,

메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 나노채널을 포함하는 시퀀싱 장치,

배출구의 다운스트림에 배치된 중앙 전극,

셀이 블록 전극과 중앙 전극 사이에 배치되도록 배치되는 블록 전극, 및

핵산이 나노채널을 통해 전좌하는 동안 전류 변화를 검출하도록 구성된 나노채널 센서를 포함하는, 시스템.

**청구항 54**

제53항에 있어서, 블록 전극 및 중앙 전극에 의해 생성되는 전기장은 나노채널 양단에 100 mV 보다 큰 차동 전위를 갖는, 시스템.

**청구항 55**

디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,

청구항 제5항 내지 제42항 중 어느 한 항에 따른 시스템, 및

제로 모드 도파관 관독기를 포함하는 시퀀싱 장치를 포함하는, 시스템.

**청구항 56**

제55항에 있어서, 유전체 층은 투명한, 시스템.

**청구항 57**

제55항 또는 제56항에 있어서, 상기 유전체 층은 도파관 채널을 포함하는, 시스템.

**청구항 58**

제57항에 있어서, 도파관 채널은 그 내부에 고정화된 중합효소를 포함하는, 시스템.

**청구항 59**

제57항 또는 제58항에 있어서, 상기 도파관 채널은 프라이머(primer) 및 형광 라벨링된 뉴클레오티드의 세트를 포함하는, 시스템.

**청구항 60**

제55항 내지 제59항 중 어느 한 항에 있어서, 상기 시스템은 단일 가닥 DNA 분자의 상보적 가닥의 합성 동안 형광 라벨링된 뉴클레오티드의 혼입에 의해 생성된 형광 신호를 검출하도록 구성된 검출기를 포함하는, 시스템.

**청구항 61**

디지털 정보를 핵산 서열로 코딩하기 위한 방법으로서, 상기 방법은,

출발지 저장소 내 유체에 현탁된 핵산 분자의 풀을 획득하는 단계,

출발지 저장소로부터, 입력 채널을 통해, 복수의 핵산 분자를 포함하는 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 메인 채널은 복수의 셀을 포함하고, 각 셀은 복수의 전기 전도성 판 중 하나씩과 상대 전극의 일부를 포함하며, 복수의 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되고,

각 전기 전도성 판은 그 상에 부착된 복수의 핵산 어댑터를 포함함 - , 및

복수의 셀 중 제1 셀에 결합 전압(binding voltage)을 인가함으로써, 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제1 전기 전도성 판의 핵산 어댑터에 결합시키는 단계를 포함하는, 방법.

**청구항 62**

제61항에 있어서, 핵산 분자의 복수의 풀을 획득하는 단계 - 핵산 분자의 복수의 풀의 각각은 복수의 출발지 저장소 중 하나 내 유체에 현탁됨 - 를 포함하는, 방법.

**청구항 63**

제61항 또는 제62항에 있어서, 결합 전압의 인가 후, 버퍼 저장소로부터 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소(waste reservoir) 내로 유동시키는 단계를 포함하는, 방법.

**청구항 64**

제62항 또는 제63항에 있어서,

제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 핵산 어댑터에 결합된 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.

**청구항 65**

제64항에 있어서,

제3 출발지 저장소로부터, 입력 채널을 통해, 제3 복수의 핵산 분자를 포함하는 제3 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제3 복수의 핵산 분자의 일부를 제2 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.

**청구항 66**

제62항 또는 제63항에 있어서,

제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

복수의 셀 중 제2 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제2 전기 전도성 판의 핵산 어댑터에 결합하는 단계를 포함하는, 방법.

**청구항 67**

제66항에 있어서,

제3 출발지 저장소로부터, 입력 채널을 통해, 제3 복수의 핵산 분자를 포함하는 제3 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

복수의 셀 중 제2 셀에 결합 전압을 인가하여, 제3 복수의 핵산 분자의 일부를 제2 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.

**청구항 68**

제61항 내지 제67항 중 어느 한 항에 있어서,

품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계.

**청구항 69**

제61항 내지 제67항 중 어느 한 항에 있어서,

품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 둘 이상에 인가하는 단계를 포함하는, 방법.

**청구항 70**

제68항 또는 제69항에 있어서,

제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계를 포함하는, 방법.

**청구항 71**

제68항 또는 제69항에 있어서,

제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 둘 이상에 인가하는 단계를 포함하는, 방법.

**청구항 72**

제68항 또는 제69항에 있어서,

제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 각 셀로 인가하는 단계를 포함하는, 방법.

**청구항 73**

제68항 내지 제72항 중 어느 한 항에 있어서, 형광 검출기를 사용해, 복수의 셀 중 하나 이상의 형광량을 측정하는 단계를 포함하는, 방법.

**청구항 74**

제61항 내지 제73항 중 어느 한 항에 있어서, 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계를 포함하는, 방법.

**청구항 75**

제61항 내지 제74항 중 어느 한 항에 있어서, 핵산 분자 분자의 결합은 점착 말단 결합을 포함하는, 방법.

**청구항 76**

제61항 내지 제75항 중 어느 한 항에 있어서, 핵산 분자의 결합은 평활 말단 결합을 포함하는, 방법.

**청구항 77**

제61항 내지 제76항 중 어느 한 항에 있어서, 결합 전압의 각 인가 후, 버퍼 저장소로부터 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소(waste reservoir) 내로 유동시키는 단계를 포함하는, 방법.

**청구항 78**

제61항 내지 제77항 중 어느 한 항에 있어서, 복수의 셀 중 하나 이상에 방출 전압(release voltage)을 인가하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.

**청구항 79**

제61항 내지 제77항 중 어느 한 항에 있어서, 복수의 셀 중 하나 이상에서 전기장을 형성하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.

**청구항 80**

제61항 내지 제77항 중 어느 한 항에 있어서, 복수의 셀 중 하나 이상을 가열하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.

**청구항 81**

제61항 내지 제77항 중 어느 한 항에 있어서, 효소 저장소로부터, 효소를 포함하는 유체의 볼륨을 메인 채널을 통해 유동시키는 단계 및 상기 효소가 핵산 어댑터와 반응하게 하여, 복수의 핵산 분자의 일부를 방출하는 단계를 포함하는, 방법.

**청구항 82**

제78항 내지 제81항 중 어느 한 항에 있어서, 상기 메인 채널로부터, 출력 채널을 통해, 출력 유체 볼륨을 도착지 저장소 내로 유동시키는 단계를 포함하는, 방법.

**청구항 83**

제61항 내지 제82항 중 어느 한 항에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자를 포함하는, 방법.

**청구항 84**

제61항 내지 제82항 중 어느 한 항에 있어서, 상기 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 방법.

**청구항 85**

제83항 또는 제84항에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 방법.

**청구항 86**

핵산 서열 내 디지털 정보를 처리하기 위한 방법으로서, 상기 방법은,

- (i) 출발지 저장소 내 유체에 현탁된 핵산 분자의 풀을 획득하는 단계,
- (ii) 출발지 저장소로부터, 입력 채널을 통해, 복수의 핵산 분자를 포함하는 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 메인 채널은 복수의 셀을 포함하고, 각 셀은 복수의 전기 전도성 판 중 하나씩과 상대 전극의 일부를 포함하며, 복수의 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되고, 각 전기 전도성 판은 그 상에 부착된 복수의 핵산 어댑터를 포함함 - ,
- (iii) 복수의 셀 중 제1 셀에 결합 전압(binding voltage)을 인가함으로써, 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제1 전기 전도성 판의 핵산 어댑터에 결합시키는 단계,
- (iv) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
- (v) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계,
- (vi) 제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계,
- (vii) 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 핵산 어댑터에 결합된 복수의 핵산 분자의 일부에 결합하는 단계, 및

- (viii) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
- (ix) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계를 포함하는, 방법.

**청구항 87**

제86항에 있어서,

- (x) 제 $n$  출발지 저장소로부터, 입력 채널을 통해, 제 $n$  복수의 핵산 분자를 포함하는 제 $n$  입력 유체 볼륨을 메인 채널 내로 유동시키는 단계,
  - (xi) 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제 $n$  복수의 핵산 분자의 일부를 제 $(n-1)$  복수의 핵산 분자의 결합된 일부에 결합하는 단계,
  - (xii) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
  - (xiii) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계, 및
- 단계 x - xiii를  $n$ 회 수행하는 단계 -  $n$ 은 3이상임 - 를 포함하는, 방법.

**청구항 88**

제86항 또는 제87항에 있어서,

- (xiv) 메인 채널을 저장 장치에 저장하는 단계를 포함하는, 방법.

**청구항 89**

제86항 내지 제88항 중 어느 한 항에 있어서, 셀에 방출 전압(release voltage)을 인가하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.

**청구항 90**

제89항에 있어서, 상기 메인 채널로부터, 출력 채널을 통해, 출력 유체 볼륨을 도착지 저장소 내로 유동시키는 단계를 포함하는, 방법.

**청구항 91**

제90항에 있어서, 도착지 저장소를 저장 장치에 저장하는 단계를 포함하는, 방법.

**청구항 92**

제86항 내지 제91항 중 어느 한 항에 있어서,

단계 (ix) 또는 (xiii) 중 하나 이상 후에, 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계를 포함하는, 방법.

**청구항 93**

제86항 내지 제92항 중 어느 한 항에 있어서,

방출 전압을 셀에 인가하여, 결합된 핵산 분자를 방출하는 단계,

메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 시퀀싱 장치로 방출된 핵산 분자를 지향시키는 단계,

방출된 핵산 분자를 시퀀싱 장치를 통해 캐비티 내로 지향시키는 단계,

시퀀싱 장치에서 복수의 전압을 측정하는 단계, 및

측정된 복수의 전압에 기초하여 염기 콜링(base calling)을 수행하는 단계를 포함하는, 방법.

**청구항 94**

제93항에 있어서, 방출된 핵산 분자는 단일 가닥 DNA 분자이거나 이를 포함하는, 방법.

**청구항 95**

제93항 또는 제94항에 있어서, 상대 전극과 시퀀싱 장치의 배출구의 다운스트림에 배치된 캐비티 전극 사이에 전압을 인가하는 단계를 포함하는, 방법.

**청구항 96**

제86항 내지 제92항 중 어느 한 항에 있어서,

방출 전압을 복수의 셀에 인가하여, 복수의 셀의 각각으로부터 결합된 핵산 분자를 방출하는 단계,

연산자 출발지 저장소로부터, 입력 채널을 통해, 복수의 연산자 핵산 분자를 포함하는 연산자 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 연산자 핵산 분자는 논리 연산자에 대응함 - , 및

연산자 핵산 분자와 방출된 핵산 분자의 화학 반응을 수행하여, 복수의 결과 핵산 분자를 생성하는 단계를 포함하는, 방법.

**청구항 97**

제96항에 있어서,

메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 시퀀싱 장치로 결과 핵산 분자를 지향시키는 단계,

결과 핵산 분자를 시퀀싱 장치를 통해 캐비티 내로 지향시키는 단계,

시퀀싱 장치에서 복수의 전압을 측정하는 단계, 및

측정된 복수의 전압에 기초하여 염기 콜링(base calling)을 수행하는 단계를 포함하는, 방법.

**청구항 98**

제86항 내지 제97항 중 어느 한 항에 있어서, 핵산 분자는 길이 L의 심볼의 스트링에 대응하는 디지털 정보를 인코딩하는 식별자 핵산 분자를 포함하는, 방법.

**청구항 99**

제86항 내지 제97항 중 어느 한 항에 있어서, 상기 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 방법.

**청구항 100**

제98항 또는 제99항에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 방법.

**청구항 101**

핵산 서열을 판독하기 위한 장치로서, 상기 장치는,

기관 내에 배치되고 입력 가닥을 포함하는 입력 핵산 분자를 수신하도록 구성된 나노-채널, 및

상기 나노-채널 상에 또는 내에 배치된 센서 장치 - 상기 센서 장치는 전자 감지 장치를 포함하고, 상기 전자 감지 장치는 게이트 전압을 갖는 전자 게이트를 갖고, 게이트 전압은 게이트의 소스-드레인 전류를 변경하기 위해 입력 핵산 분자의 이동 판독 구성요소의 전하에 의해 변조될 수 있음 - 를 포함하는, 장치.

**청구항 102**

제101항에 있어서, 상기 장치는 제5항 내지 제42항 중 어느 한 항의 시스템의 일부인, 장치.

**청구항 103**

제101항에 있어서, 상기 장치는 메인 채널 내에 배치되는, 장치.

**청구항 104**

제101항 내지 제103항 중 어느 한 항에 있어서, 상기 센서 장치는 금속-옥사이드-반도체 전계 효과 트랜지스터 (MOSFET)거나 이를 포함하는, 장치.

**청구항 105**

제101항 내지 제103항 중 어느 한 항에 있어서, 상기 센서 장치는 전해질 옥사이드 전계 효과 트랜지스터 (EOSFET)거나 이를 포함하는, 장치.

**청구항 106**

제101항 내지 제105항 중 어느 한 항에 있어서, 판독 구성요소는 단일 가닥 핵산 분자의 섹션에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는, 장치.

**청구항 107**

제101항 내지 제106항 중 어느 한 항에 있어서, 복수의 판독 구성요소 - 각 판독 구성요소는 입력 가닥의 상보적 섹션에 혼성화되어 입력 핵산 분자를 형성하도록 구성됨 - 를 포함하는, 장치.

**청구항 108**

제101항 내지 제107항 중 어느 한 항에 있어서, 제1 판독 구성요소는 제1 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되며, 제2 판독 구성요소는 제2 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되는, 장치.

**청구항 109**

제108항에 있어서, 제1 판독 구성요소는, 게이트를 통해 전좌할 때, 게이트에서의 소스-드레인 전류의 제1 변화를 야기하고, 제2 판독 구성요소는, 게이트를 통해 전좌할 때, 게이트에서의 소스-드레인 전류의 제2 변화를 야기하는, 장치.

**청구항 110**

제109항에 있어서, 상기 제1 변화는 상기 제2 변화와 상이한, 장치.

**청구항 111**

제101항 내지 제108항 중 어느 한 항에 있어서, 시작 판독 구성요소 및 종료 판독 구성요소 - 상기 시작 판독 구성요소는 입력 가닥의 제1 말단에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하고 종료 판독 구성요소는 입력 가닥의 제2 말단에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함함 - 를 포함하는, 장치.

**청구항 112**

제101항 내지 제111항 중 어느 한 항에 있어서, 입력 가닥은 디지털 정보를 인코딩하는, 장치.

**청구항 113**

제101항 내지 제112항 중 어느 한 항에 있어서, 입력 가닥은 하나 이상의 식별자 구성요소를 포함하고, 각 식별자 구성요소는 디지털 정보를 인코딩하는 핵산 식별자의 구성요소인, 장치.

**청구항 114**

제101항 내지 제113항 중 어느 한 항에 있어서, 입력 가닥은 제1 식별자 구성요소에 대응하는 제1 입력 서열 및 제2 식별자 구성요소에 대응하는 제2 입력 서열을 포함하는, 장치.

**청구항 115**

제101항 내지 제114항 중 어느 한 항에 있어서, 제1 식별자 구성요소 및 제2 식별자 구성요소의 적어도 일부분에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는 오버랩 판독 구성요소를 포함하는, 장치.

**청구항 116**

제115항에 있어서, 상기 오버랩 판독 구성요소는 플랩(flap)을 형성하는 비-상보적 핵산 섹션 및 상기 플랩의 말단에 2차 분자 구조를 갖는 핵산 섹션을 포함하는, 장치.

**청구항 117**

제115항에 있어서, 상기 오버랩 판독 구성요소는 플랩을 형성하는 비-상보적 구성요소 및 상기 플랩에 혼성화되는 2차 분자 구조를 갖는 구성요소를 포함하는, 장치.

**청구항 118**

제101항 내지 제117항 중 어느 한 항에 있어서, 상기 센서 장치는 하나 이상의 전자 신호 처리 장치이거나 이를 포함하는, 장치.

**청구항 119**

제101항 내지 제118항 중 어느 한 항에 있어서, 판독 구성요소는 2차 분자 구조를 갖는 섹션을 포함하는 단일 가닥 핵산 분자인, 장치.

**청구항 120**

제101항 내지 제118항 중 어느 한 항에 있어서, 판독 구성요소는 펩티드 압타머를 포함하는, 장치.

**청구항 121**

제101항 내지 제118항 중 어느 한 항에 있어서, 판독 구성요소는 덴드리머를 포함하는, 장치.

**청구항 122**

제101항 내지 제118항 중 어느 한 항에 있어서, 판독 구성요소는 단백질을 포함하는, 장치.

**청구항 123**

제101항 내지 제122항 중 어느 한 항에 있어서, 복수의 센서 장치를 포함하는, 시스템.

**청구항 124**

제101항 내지 제122항 중 어느 한 항에 있어서, 센서 장치는 복수의 전자 감지 장치를 포함하는, 장치.

**청구항 125**

제123항 또는 제124항에 있어서, 복수의 센서 장치 또는 복수의 전자 감지 장치는 나노-채널의 경로를 따라 직렬로 배열되는, 장치.

**청구항 126**

제124항에 있어서, 복수의 센서 장치 또는 복수의 전자 감지 장치는 둘 이상의 판독 구성요소를 동시에 판독하도록 구성되는, 장치.

**청구항 127**

핵산 서열을 판독하기 위한 방법으로서, 상기 방법은  
청구항 제101항 내지 제126항 중 어느 한 항에 따른 장치를 제공하는 단계, 및  
나노-채널을 통해 판독 구성요소를 전좌시키는 단계를 포함하는, 방법.

**청구항 128**

핵산 서열을 판독하기 위한 장치로서, 상기 장치는,

기질 내에 배치되고 입력 가닥을 포함하는 입력 핵산 분자를 수신하도록 구성된 나노-채널, 및

상기 나노-채널 상에 또는 내에 배치되는 센서 장치 - 상기 센서 장치는 광학 센싱 장치를 포함하고, 상기 광학 센싱 장치는 입력 핵산 분자의 전좌 판독 구성요소로부터 광 신호를 검출하도록 구성됨 - 를 포함하는, 장치.

**청구항 129**

제128항에 있어서, 상기 장치는 제5항 내지 제42항 중 어느 한 항의 시스템의 일부인, 장치.

**청구항 130**

제129항에 있어서, 상기 장치는 메인 채널 내에 배치되는, 장치.

**청구항 131**

제128항 내지 제130항 중 어느 한 항에 있어서, 광학 감지 장치는 형광 측정 장치인, 장치.

**청구항 132**

제128항 내지 제131항 중 어느 한 항에 있어서, 광학 감지 장치는 광소자, 카메라, 또는 광자 카운터 중 하나 이상을 포함하는, 장치.

**청구항 133**

제128항 내지 제132항 중 어느 한 항에 있어서, 판독 구성요소는 단일 가닥 핵산 분자의 섹션에 혼성화되도록 구성된 단일 가닥 핵산 분자 및 발광 요소를 포함하는, 장치.

**청구항 134**

제128항 내지 제133항 중 어느 한 항에 있어서, 복수의 판독 구성요소 - 각 판독 구성요소는 입력 가닥의 상보적 섹션에 혼성화되어 입력 핵산 분자를 형성하도록 구성됨 - 를 포함하는, 장치.

**청구항 135**

제128항 내지 제134항 중 어느 한 항에 있어서, 제1 판독 구성요소는 제1 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되며, 제2 판독 구성요소는 제2 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되는, 장치.

**청구항 136**

제135항에 있어서, 제1 판독 구성요소는 제1 광 신호를 발산하도록 구성된 발광 요소를 포함하고 제2 판독 구성요소는 제2 광 신호를 발산하도록 구성된 발광 요소를 포함하는, 장치.

**청구항 137**

제136항에 있어서, 제1 광 신호는 제2 광 신호와 상이한, 장치.

**청구항 138**

제137항에 있어서, 제1 광 신호는 제2 광 신호보다 큰 강도를 갖는, 장치.

**청구항 139**

제137항에 있어서, 제1 광 신호는 제2 광 신호와 상이한 색상을 갖는, 장치.

**청구항 140**

제128항 내지 제139항 중 어느 한 항에 있어서, 입력 가닥은 디지털 정보를 인코딩하는, 장치.

**청구항 141**

제128항 내지 제140항 중 어느 한 항에 있어서, 입력 가닥은 하나 이상의 식별자 구성요소를 포함하고, 각 식별자 구성요소는 디지털 정보를 인코딩하는 핵산 식별자의 구성요소인, 장치.

**청구항 142**

제128항 내지 제141항 중 어느 한 항에 있어서, 입력 가닥은 제1 식별자 구성요소에 대응하는 제1 입력 서열 및 제2 식별자 구성요소에 대응하는 제2 입력 서열을 포함하는, 장치.

**청구항 143**

제128항 내지 제142항 중 어느 한 항에 있어서, 제1 식별자 구성요소 및 제2 식별자 구성요소의 적어도 일부분에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는 오버랩 관독 구성요소를 포함하는, 장치.

**청구항 144**

제143항에 있어서, 상기 오버랩 관독 구성요소는 플랩(flap)을 형성하는 비-상보적 핵산 섹션 및 상기 플랩의 말단에 부착된 발광 요소를 포함하는, 장치.

**청구항 145**

제144항에 있어서, 상기 오버랩 관독 구성요소는 플랩을 형성하는 비-상보적 구성요소 및 상기 플랩에 혼성화되는 2차 분자 구조 및 발광 요소를 갖는 구성요소를 포함하는, 장치.

**청구항 146**

제128항 내지 제145항 중 어느 한 항에 있어서, 센서 장치는 복수의 광학 감지 장치를 포함하는, 장치.

**청구항 147**

제128항 내지 제146항 중 어느 한 항에 있어서, 발광 요소는 형광단인, 장치.

**청구항 148**

제128항 내지 제147항 중 어느 한 항에 있어서, 상기 센서 장치는 하나 이상의 전자 신호 처리 장치이거나 이를 포함하는, 장치.

**청구항 149**

제128항 내지 제148항 중 어느 한 항에 있어서, 복수의 센서 장치를 포함하는, 장치.

**청구항 150**

제128항 내지 제148항 중 어느 한 항에 있어서, 센서 장치는 복수의 광학 감지 장치를 포함하는, 장치.

**청구항 151**

제149항 또는 제150항에 있어서, 복수의 센서 장치 또는 복수의 광학 감지 장치는 나노-채널의 경로를 따라 직렬로 배열되는, 장치.

**청구항 152**

제150항에 있어서, 복수의 센서 장치 또는 복수의 광학 감지 장치는 둘 이상의 관독 구성요소를 동시에 관독하도록 구성되는, 장치.

**청구항 153**

핵산 서열을 관독하기 위한 방법으로서, 상기 방법은

청구항 제128항 내지 제152항 중 어느 한 항에 따른 장치를 제공하는 단계, 및

나노-채널을 통해 관독 구성요소를 전좌시키는 단계를 포함하는, 방법.

**발명의 설명**

**기술 분야**

[0001] 관련 출원에 대한 교차 참조

[0002] 본 출원은 2022년03월04일에 출원된 미국 가특허출원 번호 63/316,812, 발명의 명칭 "DNA MICROARRAYS", 2022년04월01일에 출원된 미국 가특허출원 번호 63/326,598, 발명의 명칭 "COMPONENT LEVEL SEQUENCING", 2022년04월08일에 출원된 미국 가특허출원 번호 63/329,111, 발명의 명칭 "MULTISENSOR COMPONENT LEVEL SEQUENCING", 및 2022년04월22일에 출원된 미국 가특허출원 번호 63/333,698, 발명의 명칭 "MULTISENSOR COMPONENT LEVEL SEQUENCING"의 우선권 및 이익을 주장한다. 앞서 언급된 각 출원의 전체 내용은 본 명세서에 참조로 포함된다.

**배경 기술**

[0003] 핵산 디지털 데이터 스토리지는 자기 테이프 또는 하드 드라이브 저장 시스템보다 더 높은 밀도로 저장된 데이터를 갖는 장기간 동안 정보를 인코딩 및 저장하기 위한 안정한 방식이다. 또한, 춥고 건조한 조건에서 저장된 핵산 분자에 저장된 디지털 데이터는 60,000년 이상 동안 검색될 수 있다. 핵산 분자에 저장된 디지털 데이터를 액세스하기 위해, 핵산 분자가 시퀀싱될 수 있다. 이와 같이, 핵산 디지털 데이터 스토리지는 빈번하게 액세스되지는 않지만 장기간 저장하거나 보관해야 하는 정보의 양이 많을 수 있는 데이터를 저장하는 이상적인 방법일 수 있다.

**발명의 내용**

**과제의 해결 수단**

[0004] 현재의 DNA 데이터 저장 기술은 핵산 분자(가령, DNA)에 정보를 기록하거나 핵산 분자에 인코딩된 정보를 읽는 것과 같은 문제의 한 측면에 초점을 맞춘다. 본 명세서에는 핵산 분자에 디지털 정보를 기록하고, 핵산 분자에 인코딩된 디지털 정보를 읽고, 디지털 정보를 인코딩한 핵산 분자를 저장하고, 핵산 분자에서 계산 연산을 수행하는 통합 플랫폼이 기재되어 있다. 본 명세서에 기재된 기술은 다음과 같다: 이상적인 버퍼 조건에서 핵산 분자(가령, DNA) 조립을 위한 완전히 통합된 핵산 분자(가령, DNA) 기록기/판독기/저장/계산 장치, 시스템 및 방법, 결찰 효율의 신속한 정량화를 위한 시스템 및 방법, 완전히 형성된 핵산 분자(가령, DNA) 분자의 신속한 정화를 위한 시스템 및 방법, 및 핵산 분자(가령, DNA)로의 대규모 데이터 세트에 대한 높은 처리율 기록을 위한 시스템 및 방법. 또한, 본 명세서에 기재된 기술은 다음과 같다: 핵산 서열을 판독하기 위한 장치로서, 상기 장치는, 기질 내에 배치되고 입력 가닥을 포함하는 입력 핵산 분자를 수신하도록 구성된 나노-채널, 및 상기 나노-채널 상에 또는 내에 배치된 센서 장치 - 상기 센서 장치는 전자 감지 장치를 포함하고, 상기 전자 감지 장치는 게이트 전압을 갖는 전자 게이트를 갖고, 게이트 전압은 게이트의 소스-드레인 전류를 변경하기 위해 입력 핵산 분자의 전하 판독 구성요소의 전하에 의해 변조될 수 있음 - 를 포함하는, 장치.

[0005] 참조에 의한 통합

[0006] 본 명세서에 언급된 모든 간행물, 특허 및 특허 출원은 각각의 개별 간행물, 특허 또는 특허 출원이 참조로 포함되도록 구체적이고 개별적으로 표시된 것과 동일한 정도로 참조로 여기에 포함된다. 참조로 포함된 간행물, 특허 또는 특허 출원이 명세서에 포함된 공개 내용과 모순되는 경우, 명세서가 그러한 모순되는 자료를 대체하거나 우선시되도록 의도된다.

**도면의 간단한 설명**

[0007] 본 발명의 신규한 특징은 첨부된 청구범위에서 구체적으로 설명된다. 본 발명의 특징과 장점에 대한 더 나은 이해는 본 발명의 원리가 활용되는 예시적인 실시예를 설명하는 다음의 상세한 설명과 첨부 도면(또한 "도면" 및 "도")을 참조하여 얻어질 것이다.

도 1은 전기 전도성 판 및 상대 전극을 포함하는 복수의 셀을 포함하는 예시적 채널의 개략적 사시도를 나타낸다.

도 2는 본 명세서에 기재된 기술과 함께 사용될 수 있는 소형화된 금속 판의 예시적 (마이크로)어레이의 개략적인 평면도를 보여준다.

도 3은 본 명세서에 기재된 기술과 함께 사용될 수 있는 전극 마이크로어레이의 개략적인 평면도를 보여주며, 여기서 어레이는 복수의 블록으로 세분화된다.

도 4a는 본 명세서에 기재된 기술과 함께 사용될 수 있는 5x5 전극 마이크로어레이의 개략적인 평면도를 보여준다. 도 4b는 4x4 전극 어레이를 갖는 예시적 동적 랜덤 액세스 메모리(DRAM: dynamic random access memory) 셀 어레이의 개략도이다.

도 5는 마이크로 어레이의 예시적 셀의 단면도를 보여준다.

도 6은 예시적인 셀 구성의 개략적인 단면도를 보여준다.

도 7은 본 명세서에 기재된 바와 같이 디지털 정보를 핵산 서열로 변환하기 위한 예시 시스템의 개략도를 보여준다.

도 8은 본 명세서에 기재된 바와 같이 디지털 정보를 핵산 서열로 변환하기 위한 예시 시스템의 개략도를 보여준다.

도 9a는 본 명세서에 기재된 기술과 함께 사용될 수 있는 나노포어 판독기 모듈의 개략적인 단면도를 보여준다.

도 9b는 본 명세서에 기재된 기술과 함께 사용될 수 있는 예시적인 나노포어 판독기 모듈의 개략적인 평면도를 보여준다.

도 10a는 본 명세서에 기재된 기술과 함께 사용될 수 있는 나노채널 판독기 모듈의 개략적인 평면도를 보여준다. 도 10b는 본 명세서에 기재된 기술과 함께 사용될 수 있는 나노채널 판독기 모듈의 개략적인 단면도를 보여준다.

도 11a는 본 명세서에 기재된 기술과 함께 사용될 수 있는 제로-모드 도파관 모듈의 개략적인 단면도를 보여준다. 도 11b는 본 명세서에 기재된 기술과 함께 사용될 수 있는 예시적 제로-모드 도파관 모듈의 개략적인 평면도를 보여준다.

도 12는 본 명세서에 기재된 기술과 함께 사용될 수 있는 예시적인 나노-포어 시퀀싱 모듈의 개략적인 단면도를 보여준다.

도 13은 본 명세서에 기재된 기술과 함께 사용될 수 있는 예시적인 나노-포어 시퀀싱 장치의 개략적인 사시도이다.

도 14는 본 명세서에 기재된 기술과 함께 사용될 수 있는 예시적인 나노-포어 시퀀싱 장치의 개략적인 평면도이다.

도 15는 디지털 정보를 핵산 서열로 변환하기 위한 예시 시스템의 개략도를 보여준다.

도 16은 본 명세서에 기재된 바와 같은 예시적인 DNA 쓰기 프로세스의 흐름도이다.

도 17은 본 명세서에 기재된 바와 같이 나노포어 판독기를 사용한 예시적인 DNA 판독 프로세스의 흐름도이다.

도 18은 본 명세서에 기재된 바와 같이 제로 모드 도파관 판독기를 사용하는 예시적인 DNA 판독 프로세스의 흐름도이다.

도 19는 본 명세서에 기재된 기술에 사용하기 위한 복수의 구성요소 층의 예시 핵산 분자의 개략적 그림을 보여준다.

도 20은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 2개의 셀이 있는 채널을 통해 0번째( $A_0^0$ ) 구성요소 층의 핵산 분자를 유동시키는 공정의 개략적 그림을 보여준다.

도 21은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 버퍼 행금 공정의 개략적 그림을 보여준다.

도 22는 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 2개의 셀이 있는 채널을 통해 0번째 구성요소( $A_0^1$ ) 층의 핵산 분자를 유동시키는 공정의 개략적 그림을 보여준다.

도 23은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 버퍼 행금 공정의 개략적 그림을 보여준다.

도 24는 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 두 개의 셀이 있는 채널을 통해 첫 번째 구성요소( $B_1^0$ ) 층의 핵산 분자를 유동시키게 하는 공정의 개략적 그림을 보여준다.

도 25는 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 두 개의 셀이 있는 채널을 통해 두 번째 구성요소( $C_2^0$ ) 층의 핵산 분자를 유동시키게 하는 공정의 개략적 그림을 보여준다.

도 26는 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 두 개의 셀이 있는 채널을 통해 두 번째 구성요소( $C_2^1$ ) 층의 핵산 분자를 유동시키게 하는 공정의 개략적 그림을 보여준다.

도 27은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 두 개의 셀이 있는 채널을 통한 최종 버퍼 행균 공정의 개략적 그림을 보여준다.

도 28은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 두 개의 셀이 있는 채널에서 품질 관리 단계에 대한 개략적 설명을 보여준다.

도 29는 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 결찰 효율을 결정하기 위한 두 개의 셀 스텝이 있는 채널에서의 품질 관리의 개략적 설명을 보여준다.

도 30은 본 명세서에 기재된 바와 같은 예시적인 핵산 조립 공정에서 불완전한 생성물의 분포를 결정하기 위한 두 개의 셀이 있는 채널에서의 품질 관리 단계의 개략적 설명을 보여준다.

도 31은 본 명세서에 기재된 예시 핵산 조립 공정에서 두 개의 셀이 있는 채널에서의 이진 결과(데이터 있음 대 (vs) 데이터 없음) 판독의 개략적 그림을 보여준다.

도 32는 예시 QC 공정에서 획득된 전극 어레이의 개략적인 형광 지도를 보여준다.

도 33은 두 개의 셀이 있는 채널에서 불완전한 핵산 생성물을 제거하는 개략적 그림을 보여준다.

도 34는 본 명세서에 기재된 바와 같이 두 개의 셀이 있는 채널에서의 예시적 데이터 불러오기 단계의 개략적 그림을 보여준다.

도 35는 본 명세서에 기재된 바와 같이 두 개의 셀이 있는 채널에서의 예시적 계산 단계의 개략적 그림을 보여준다.

도 36은 판독 구성요소를 포함하는 예시적 핵산 분자를 갖는 예시 나노-채널 및 MOSFET 감지 장치의 개략적인 단면도를 나타낸다.

도 37은 소스-드레인 전류의 변화를 생성하는 MOSFET의 예시적 게이트를 통해 전좌하는 제1 판독 구성요소를 갖는 나노-채널의 개략적 단면도를 도시한다.

도 38은 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 나타낸다.

도 39는 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 설명하는데, 여기서 분자는 핵산 시퀀스의 경계를 나타내는 큰 전기 신호를 유도하는 첫 번째 및 마지막 판독 구성요소를 가진다.

도 40는 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 설명하는데, 여기서 분자는 핵산 시퀀스의 경계를 나타내는 작은 전기 신호의 패턴을 유도하는 첫 번째 및 마지막 판독 구성요소를 가진다.

도 41은 네 개의 영역을 갖는 예시적 판독 구성요소를 설명한다.

도 42는 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 기록 프로세스에 사용되는 것과 같은 네 개 영역을 갖는 판독 구성요소를 가진다.

도 43은 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 나타내고, 여기서 분자는 혼성화된 2차 판독 구성요소와 함께 판독 구성요소를 가진

다.

도 44는 형광단을 포함하는 판독 구성요소를 포함하는 예시 핵산 분자를 갖는 광학 형광 측정 장치가 있는 예시적 나노-채널의 개략적 단면도를 보여준다.

도 45는 본 명세서에 기재된 바와 같이 광학 형광 측정 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 형광 강도 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 기록 프로세스에 사용되는 것과 같은 형광단을 갖는 판독 구성요소를 가진다.

도 46는 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 쓰기 프로세스에 사용되는 것과 같은, 층 경계에서 판독 구성요소를 가진다.

도 47은 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 쓰기 프로세스에 사용되는 것과 같은, 층 경계에서 감소된 개수의 판독 구성요소를 가진다.

도 48는 본 명세서에 기재된 바와 같이 광학 형광 측정 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 형광 강도 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 기록 프로세스에 사용되는 것과 같은, 형광단을 갖는 층 경계부에서 판독 구성요소를 가진다.

도 49는 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 기록 프로세스에 사용되는 것과 같은, 압타머 및 펩티드를 갖는 판독 구성요소를 가진다.

도 50은 본 명세서에 기재된 바와 같이 나노-채널 센서 장치를 통해 예시적인 핵산 분자(가령, DNA)가 전좌할 때 측정된 전류 변화의 시퀀스를 도시하는데, 여기서 분자는 본 명세서에 기재된 바와 같이 DNA 기록 프로세스에 사용되는 것과 같은, 덴드리머를 갖는 판독 구성요소를 가진다.

도 51은 본 명세서에 기재된 바와 같이 예시적 나노-채널 및 센서 장치를 통해 전좌하는 다중 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여준다.

도 52는 예시적 나노-채널과 "느린" 센서 장치를 통해 전좌하는 다중 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여준다.

도 53은 예시적 나노-채널과 다중 센서 장치를 통해 전좌하는 다중 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여주며, 각 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))는 적어도 하나의 정보를 판독한다.

도 54는 예시적 나노-채널과 여러 센서 장치를 통해 전좌하는 다수의 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여주며, 여기서 각 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))는 핵산에서 판독된 정보 중 일부를 놓칠 수 있다.

도 55는 예시적 나노-채널과 n개의 다중 센서 장치를 통해 전좌하는 다중 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여준다.

도 56은 다중 센서 장치 클러스터를 갖춘 예시적 나노-채널을 통해 전좌하는 다중 판독 구성요소를 갖는 예시적 핵산 분자의 개략도를 보여준다.

도 57은 핵산 서열에 저장된 디지털 정보를 인코딩, 기록, 액세스, 질의, 판독 및 디코딩하는 프로세스의 개요를 개략적으로 도시한다.

도 58a 및 58b는 객체 또는 식별자(가령, 핵산 분자)를 사용하여 "데이터 앳 어드레스(data at address)"라고 하는, 디지털 데이터를 인코딩하는 예시적인 방법을 개략적으로 예시한다. 도 58a는 식별자를 생성하기 위해 순위 객체(또는 어드레스 객체)를 바이트-값 객체(또는 데이터 객체)와 결합하는 것을 도시한다. 도 58b는 데이터 앳 어드레스 방법의 실시예를 도시하며, 여기서 순위 객체와 바이트-값 객체는 그 자체가 타 객체의 조합적 연결이다.

도 59a 및 59b는 객체 또는 식별자(가령, 핵산 서열)를 사용하여 디지털 정보를 인코딩하는 예시적인 방법을 개략적으로 예시한다. 도 59a는 식별자로서 순위 객체를 사용하여 디지털 정보를 인코딩하는 것을 도시한다. 도

59b는 인코딩 방법의 실시예를 도시하며, 여기서, 주소 객체는 그 자체가 타 객체의 조합적 연결이다.

도 60은 가능한 식별자의 조합 공간(C, x-축)과 주어진 크기의 정보를 저장하기 위해 구성될 수 있는 식별자의 평균 개수(k, y-축) 간의 관계의 로그 공간에서 등고선 플롯을 보여준다.

도 61는 핵산 서열(가령, 테옥시리보핵산)에 정보를 기록하기 위한 방법의 개요를 개략적으로 예시한다.

도 62a 및 62b는 개별 구성요소(가령, 핵산 서열)를 조합적으로 조립함으로써 식별자(가령, 핵산 분자)를 구축하기 위한 "곱 방식"라고 하는 예시적인 방법을 예시한다. 도 62a는 곱 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 도 62b는 곱 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다.

도 63은 구성요소(예를 들어, 핵산 서열)로부터 식별자(예를 들어, 핵산 분자)를 구성하기 위한 중첩 확장 중합 효소 연쇄 반응의 사용을 개략적으로 예시한다.

도 64은 구성요소(예를 들어, 핵산 서열)로부터 식별자(예를 들어, 핵산 분자)를 구성하기 위한 점착 말단 결찰의 사용을 개략적으로 예시한다.

도 65는 구성요소(가령, 핵산 서열)로부터 식별자(가령, 핵산 분자)를 구성하기 위한 재조합효소 조립의 사용을 개략적으로 예시한다.

도 66a 및 66b는 주형 지시 결찰을 보여준다. 도 66a는 구성요소(예를 들어, 핵산 서열)로부터 식별자(예를 들어, 핵산 분자)를 구성하기 위한 주형 지시 결찰의 사용을 개략적으로 예시한다. 도 66b는 하나의 풀링된 주형 지시 결찰 반응에서 6개의 핵산 서열(예를 들어, 구성요소)로부터 각각 조합적으로 조립된 256개의 개별 핵산 서열의 복제수(존재비)에 대한 히스토그램을 보여준다.

도 67a - 67g는 순열된 구성요소(가령, 핵산 서열)로 식별자(가령, 핵산 분자)를 구성하기 위한 "순열 방식"으로 불리는 예시적인 방법을 개략적으로 예시한다. 도 67a는 순열 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 도 67b는 순열 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 도 67c는 주형 지시 결찰을 이용한 순열 방식의 예시적인 구현을 보여준다. 도 67d는 도 67c의 구현 방법이 순열 및 반복된 구성요소로 식별자를 구성하도록 수정될 수 있는 방법의 예를 도시한다. 도 67e는 도 67d의 예시적인 구현예가 핵산 크기 선택으로 제거될 수 있는 원치 않는 부산물을 어떻게 초래할 수 있는지를 보여준다. 도 67f는 순열 및 반복된 구성요소로 식별자를 구성하기 위해 주형 지시 결찰 및 크기 선택을 사용하는 방법의 또 다른 예를 보여준다. 도 67g는 크기 선택이 원치 않는 부산물로부터 특정 식별자를 분리하는 데 실패할 수 있는 경우의 예를 보여준다.

도 68a - 68d는 더 많은 개수  $M$ 의 가능한 구성요소 중 임의의 개수  $k$ 의 조립된 구성요소(가령, 핵산 서열)를 갖는 식별자(가령, 핵산 분자)를 구성하기 위한 "MchooseK" 방식이라 지칭되는, 예시적 방법을 개략적으로 도시한다. 도 68a는 MchooseK 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 도 68b는 MchooseK 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 도 68c는 주형 지시 결찰을 사용하는 MchooseK 방식의 예시적인 구현을 보여준다. 도 68d는 도 68c의 예시적인 구현예가 핵산 크기 선택으로 제거될 수 있는 원치 않는 부산물을 어떻게 초래할 수 있는지를 보여준다.

도 69a 및 도 69b는 분할된 구성요소로 식별자를 구성하기 위한 "분할 방식"으로 지칭되는 예시적인 방법을 개략적으로 도시한다. 도 69a는 분할 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 보여준다. 도 69b는 주형 지시 결찰을 사용한 분할 방식의 구현 예를 보여준다.

도 70a 및 도 70b는 다수의 가능한 구성요소로부터의 구성요소의 임의의 스트링으로 구성된 식별자를 구성하기 위한 "비제한 스트링(unconstrained string)"(또는 USS) 방식으로 지칭되는 예시적인 방법을 개략적으로 나타낸다. 도 70a는 USS 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 도 70b는 주형 지시 결찰을 사용하는 USS 방식의 예시적인 구현을 보여준다.

도 71a 및 도 72b는 부모 식별자로부터 구성요소를 제거함으로써 식별자를 구성하기 위한 "구성요소 결실"이라고 불리는 예시적인 방법을 개략적으로 예시한다. 도 71a는 구성요소 결실 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 도 71b는 이중 가닥 표적화된 절단 및 복구를 사용하는 구성요소 결실 방식의 예시적인 구현을 보여준다.

도 72는 재조합효소를 부모 식별자에 적용함으로써 추가의 식별자가 구성될 수 있는 재조합효소 인식 부위를 갖는 부모 식별자를 개략적으로 도시한다.

도 73a - 73c는 더 많은 수의 식별자 중 다수의 특정 식별자를 액세스함으로써 핵산 서열에 저장된 정보의 일부를 액세스하기 위한 예시적 방법의 개요를 개략적으로 도시한다. 도 73a는 특정된 구성요소를 포함하는 식별자를 액세스하기 위해 중합효소 연쇄 반응, 친화성 태깅된 프로브, 및 분해 표적화 프로브를 사용하는 예시적인 방법을 보여준다. 도 73b는 다수의 특정된 구성요소를 포함하는 식별자를 액세스하기 위해 'OR' 또는 'AND' 연산을 수행하기 위해 중합효소 연쇄 반응을 사용하는 방법의 예를 보여준다. 도 73c는 다수의 특정된 구성요소를 포함하는 식별자를 액세스하기 위해 'OR' 또는 'AND' 연산을 수행하기 위해 친화성 태그를 사용하는 예시적인 방법을 도시한다.

도 74a 및 74b는 핵산 분자에 인코딩된 데이터의 인코딩, 기록 및 판독의 예를 보여준다. 도 74a는 5,856 비트의 데이터를 인코딩하고, 기록하고, 판독하는 예를 보여준다. 도 74b는 62,824 비트의 데이터를 인코딩하고, 기록하고, 판독하는 예를 보여준다.

도 75은 본 명세서에 제공된 방법을 구현하도록 프로그래밍되거나 그 밖의 다른 방식으로 구성된 컴퓨터 시스템을 보여준다.

도 76은 이중 가닥 구성요소의 단일 부모 세트로부터의 임의의 두 개의 선택된 이중-가닥 구성요소를 조립하는 예시적인 방식을 도시한다.

도 77은 두 개의 올리고 X와 Y로 만들어진 가능한 접착 말단 구성요소 구조를 보여준다.

도 78는 다수의 기능적 부분을 갖는 구성요소로부터 식별자를 구축하는 예를 보여준다.

도 79a - 79b는 PCR 기반 랜덤 액세스에 대한 식별자 순위의 효과 예시를 보여준다.

도 80a - 80b는 PCR 기반 랜덤 액세스에 대한 불균일한 구성요소 분포를 갖는 식별자 아키텍처의 예시적인 효과를 보여준다.

도 81는 PCR 기반 랜덤 액세스에 대한 식별자 아키텍처에서의 층 증가의 예시적인 효과를 도시한다.

도 82은 9개 심볼의 알파벳에 대한 다중-빈 위치 인코딩 방식의 예를 보여준다.

도 83은 4-비트 스트링의 9개의 가능한 메시지 중 임의의 것의 인코딩을 가능하게 하는 2개의 식별자의 식별자 라이브러리 및 3개의 빈의 빈 세트를 갖춘 다중-빈 식별자 분포 인코딩 방식의 예를 보여준다.

도 84은 6-비트 스트링의 64개 가능한 메시지 중 임의의 것의 인코딩을 가능하게 하는 2개의 식별자의 라이브러리 및 3개의 빈의 빈 세트를 갖춘 식별자의 재사용을 이용하는 다중-빈 식별자 분포 인코딩 방식의 예를 도시한다.

도 85는 정수 분할을 사용하여 DNA의 정보를 인코딩하는 예를 보여준다.

도 86은 소스 비트스트림을 작성자에 의해 해석될 빌드 프로그램 사양으로 준비하고 변환하기 위한 알고리즘 모듈을 포함하는 인코딩 파이프라인의 예를 보여준다.

도 87은 식별자 라이브러리를 직렬화된 포맷으로 표현하기 위한 데이터 구조의 하나의 실시예를 도시한다.

도 88는 식별자 풀에 정의된 연산을 사용하여 계산하기 위해 준비된 두 개의 소스 비트스트림과 범용 식별자 라이브러리의 예를 보여준다.

도 89은 식별자 라이브러리가 시험관 내(*in vitro*) 계산을 위한 플랫폼으로서 사용될 수 있는 방법을 설명하는 식별자의 풀에 대해 수행되는 논리 연산의 세 가지 예에 대한 입력 및 결과를 보여준다.

도 90a - 90g는 이미지 파일을 저장하고 이를 다양한 해상도로 읽는 예를 보여준다.

도 91는 랜덤 비트 스트링을 생성하는 데 사용될 수 있는 엔트로피를 생성하기 위한 예시적인 방법을 도시한다.

도 92a - 92c는 엔트로피(랜덤 비트 스트링)를 생성하고 저장하기 위한 예시적인 방법을 보여준다.

도 93a - 93b는 입력을 사용하여 랜덤 비트 스트링을 구성하고 액세스하는 방법의 예를 보여준다.

도 94은 물리적 DNA 키를 사용하여 아티팩트에 대한 액세스를 보호하고 인증하는 예시적인 방법을 보여준다.

**발명을 실시하기 위한 구체적인 내용**

본 발명의 다양한 실시예가 본 명세서에 도시되고 설명되었지만, 이러한 실시예는 단지 예로서 제공된다는 것이

[0008]

통상의 기술자에게 명백할 것이다. 본 발명을 벗어나지 않으면서 통상의 기술자가 다양한 변형, 변화 및 대체를 할 수 있다. 본 명세서에 기재된 본 발명의 실시예에 대한 다양한 대안이 채용될 수 있다는 것이 이해되어야 한다.

- [0009] 본 명세서에서 사용되는 "심볼(symbol)"이라는 용어는 일반적으로 디지털 정보의 단위를 나타내는 것을 의미한다. 디지털 정보는 심볼의 스트링으로 분할되거나 번역될 수 있다. 예를 들어, 심볼은 비트일 수 있고 비트는 '0' 또는 '1'의 값을 가질 수 있다.
- [0010] 본 명세서에서 사용된 용어 "개별(distinct)" 또는 "고유한(unique)"은 일반적으로 그룹 내의 다른 객체와 구별될 수 있는 객체를 의미한다. 예를 들어, 개별 또는 고유한 핵산 서열은 임의의 타 핵산 서열과 동일한 서열을 갖지 않는 핵산 서열일 수 있다. 개별 또는 고유한 핵산 분자는 임의의 타 핵산 분자와 동일한 서열을 갖지 않을 수도 있다. 개별, 또는 고유한 핵산 서열 또는 분자는 타 핵산 서열 또는 분자와 유사성의 영역을 공유할 수 있다.
- [0011] 본 명세서에서 사용될 때 용어 "구성요소"는 일반적으로 핵산 서열을 지칭한다. 구성요소는 개별 핵산 서열일 수 있다. 구성요소는 다른 핵산 서열 또는 분자를 생성하기 위해 하나 이상의 다른 구성요소와 연결되거나 조립될 수 있다.
- [0012] 본 명세서에 사용될 때, 용어 "층(layer)"은 일반적으로 구성요소의 그룹 또는 풀을 지칭한다. 각 층은 한 층의 구성요소가 다른 층의 구성요소와 상이하도록 구별되는 구성요소 세트를 포함할 수 있다. 하나 이상의 층의 구성요소가 조립되어 하나 이상의 식별자를 생성할 수 있다.
- [0013] 본 명세서에서 사용될 때 용어 "식별자"는 일반적으로 더 큰 비트-스트링 내에서 비트-스트링의 위치 및 값을 나타내는 핵산 분자 또는 핵산 서열을 지칭한다. 더 일반적으로, 식별자는 심볼의 스트링 내 한 심볼을 나타내거나 대응하는 임의의 객체를 지칭할 수 있다. 일부 실시예에서, 식별자는 하나 또는 다수의 연결된 구성요소를 포함할 수 있다.
- [0014] 본 명세서에 사용될 때 "조합 공간"이라는 용어는 일반적으로 객체, 가령, 구성요소의 시작 세트로부터 생성될 수 있는 가능한 모든 개별 식별자의 세트와 식별자를 형성하기 위해 해당 객체를 수정하는 방법에 대한 허용 가능한 규칙 세트를 지칭한다. 구성요소를 조립하거나 연결함으로써 만들어지는 식별자의 조합 공간의 크기는 구성요소의 층의 수, 각 층에서의 구성요소의 수, 및 식별자를 생성하는 데 사용되는 특정 조립 방법에 따라 달라질 수 있다.
- [0015] 본 명세서에서 사용되는 "식별자 순위"라는 용어는 일반적으로 세트 내 식별자의 순서를 정의하는 관계를 지칭한다.
- [0016] 본 명세서에서 사용될 때 용어 "식별자 라이브러리"는 일반적으로 디지털 정보를 나타내는 심볼 스트링에서의 심볼에 대응하는 식별자의 모음을 지칭한다. 일부 실시예에서, 식별자 라이브러리에 주어진 식별자가 없다는 것이 특정 위치에서의 심볼 값을 나타낼 수 있다. 하나 이상의 식별자 라이브러리는 식별자의 풀, 그룹 또는 세트로 조합될 수 있다. 각각의 식별자 라이브러리는 식별자 라이브러리를 식별하는 고유의 바코드를 포함할 수 있다.
- [0017] 본 명세서에서 사용될 때 용어 "핵산"은 일반적으로 데옥시리보핵산(DNA), 리보핵산(RNA), 또는 이들의 변이체를 지칭한다. 핵산은 아데노신(A), 시토신(C), 구아닌(G), 티민(T) 및 우라실(U), 또는 이들의 변이체로부터 선택되는 하나 이상의 서브유닛을 포함할 수 있다. 뉴클레오티드는 A, C, G, T, 또는 U, 또는 이의 변이체를 포함할 수 있다. 뉴클레오티드는 성장하는 핵산 가닥에 포함될 수 있는 임의의 서브유닛을 포함할 수 있다. 이러한 서브유닛은 A, C, G, T, 또는 U, 또는 하나 이상의 상보적 A, C, G, T 또는 U에 특정적일 수 있는 그 밖의 다른 임의의 서브유닛, 또는 퓨린(즉, A 또는 G, 또는 이의 변이체) 또는 피리미딘(즉, C, T 또는 U, 또는 이의 변이체)에 상보적인 서브유닛일 수 있다. 일부 예를 들어, 핵산은 단일 가닥 또는 이중 가닥일 수 있으며, 일부 경우에 핵산은 원형이다.
- [0018] 본 명세서에서 사용될 때 용어 "핵산 분자" 또는 "핵산 서열"은 일반적으로 데옥시리보뉴클레오티드(DNA) 또는 리보뉴클레오티드(RNA)와 같은 다양한 길이를 가질 수 있는 중합체 형태의 뉴클레오티드 또는 폴리뉴클레오티드, 또는 이의 유사체를 지칭한다. "핵산 서열"이라는 용어는 폴리뉴클레오티드의 알파벳순 표현을 지칭할 수 있으며, 대안으로, 상기 용어는 물리적 폴리뉴클레오티드 자체에 적용될 수 있다. 이 알파벳 표현은 중앙 처리 장치가 있는 컴퓨터의 데이터베이스에 입력할 수 있으며 핵산 서열 또는 핵산 분자를 디지털 정보를 인코딩하는 심볼 또는 비트에 매핑하는 데 사용할 수 있다. 핵산 서열 또는 올리고뉴클레오티드는 하나 이상

의 비표준 뉴클레오티드(들), 뉴클레오티드 유사체(들) 및/또는 변형된 뉴클레오티드를 포함할 수 있다.

- [0019] 본 명세서에 사용될 때 "올리고뉴클레오티드"는 일반적으로 단일 가닥 핵산 서열을 의미하며, 일반적으로 다음의 4개의 뉴클레오티드 염기의 특정 서열로 구성된다: 아데닌(A), 시토신(C), 구아닌(G), 및 티민(T) 또는 폴리뉴클레오티드가 RNA인 경우 우라실(U).
- [0020] 변형된 뉴클레오티드의 비제한적 예로는 디아미노퓨린, 5-플루오로우라실, 5-브로모우라실, 5-클로로우라실, 5-요오도우라실, 히포크산틴, 잔틴, 4-아세틸시토신, 5-(카르복시히드록실메틸)우라실, 5-카르복시메틸아미노메틸-2-티오우리딘, 5-카르복시메틸아미노메틸우라실, 디하이드로우라실, 베타-D-갈락토실케오신, 이노신, N6-이소펜테닐아데닌, 1-메틸구아닌, 1-메틸이노신, 2,2-디메틸구아닌, 2-메틸아데닌, 2-메틸구아닌, 3-메틸시토신, 5-메틸시토신, N6-아데닌, 7-메틸구아닌, 5-메틸아미노메틸우라실, 5-메톡시아미노메틸-2-티오우리딘, 베타-D-만노실케오신, 5'-메톡시카르복시메틸우라실, 5-메톡시우라실, 2-메틸티오-D46-이소펜테닐아데닌, 우라실-5-옥시아세트산(v), 와이부톡신, 슈도우라실, 쿠오신, 2-티오시토신, 5-메틸-2-티오우리딘, 2-티오우리딘, 4-티오우리딘, 5-메틸우리딘, 우라실-5-옥시아세트산 메틸에스테르, 우라실-5-옥시아세트산(v), 5-메틸-2-티오우리딘, 3-(3-아미노-3-N-2-카르복시프로필)우라실, (acp3)w, 2,6-디아미노퓨린 등이 있다. 핵산 분자는 또한 염기 잔기에서(가령, 상보적 뉴클레오티드와 수소 결합을 형성하는 데 일반적으로 이용 가능한 하나 이상의 원자 및/또는 일반적으로 상보적 뉴클레오티드와 수소 결합을 형성할 수 없는 하나 이상의 원자에서), 당 잔기 또는 포스페이트 골격에서도 변경될 수 있다. 핵산 분자는 또한 아민 변형된 기, 가령, 아민 반응성 잔기의 공유 부착을 허용하기 위해 아미노알릴-dUTP(aa-dUTP) 및 아미노핵실아크릴아미드-dCTP(aha-dCTP), 가령, N-히드록시 숙신이미드 에스테르(NHS)를 함유할 수 있다.
- [0021] 본 명세서에서 사용될 때 용어 "프라이머(primer)"는 일반적으로 핵산 합성, 가령, 중합효소 연쇄 반응(PCR)을 위한 시작점으로서 역할 하는 핵산의 가닥을 지칭한다. 예를 들어, DNA 샘플을 복제하는 동안, 복제를 촉매하는 효소는 DNA 샘플에 부착된 프라이머의 3'-말단에서 복제를 시작하고 반대 가닥을 복제한다. 프라이머 설계에 대한 세부사항을 포함하여 PCR에 대한 자세한 내용은 화학적 방법 섹션 D를 참조할 수 있다.
- [0022] 본 명세서에 사용된 용어 "중합효소" 또는 "중합효소 효소"는 일반적으로 중합효소 반응을 촉매할 수 있는 임의의 효소를 의미한다. 중합효소의 비제한적 예를 들면, 핵산 중합효소가 있다. 중합효소는 자연적으로 발생하거나 합성될 수 있다. 중합효소의 예는  $\Phi$ 29 중합효소 또는 이의 유도체이다. 일부 경우에, 전사효소 또는 리가제(즉, 결합 형성을 촉매하는 효소)가 중합효소와 함께 또는 중합효소의 대안으로서 사용되어 새로운 핵산 서열을 구성할 수 있다. 중합효소의 예에는 DNA 중합효소, RNA 중합효소, 열안정성 중합효소, 야생형 중합효소, 변형된 중합효소, 대장균 DNA 중합효소 I, T7 DNA 중합효소, 박테리오파지 T4 DNA 중합효소  $\Phi$ 29(phi29) DNA 중합효소, Taq 중합효소, Tth 중합효소, Tli 중합효소, Pfu 중합효소 Pwo 중합효소, VENT 중합효소, DEEPVENT 중합효소, Ex-Taq 중합효소, LA-Taw 중합효소, Sso 중합효소 Poc 중합효소, Pab 중합효소, Mth 중합효소 ES4 중합효소, Tru 중합효소, Tac 중합효소, Tne 중합효소, Tma 중합효소, Tca 중합효소, Tih 중합효소, Tfi 중합효소, 백금 Taq 중합효소, Tbr 중합효소, Tfl 중합효소, Pfutubo 중합효소, Pyrobest 중합효소, KOD 중합효소, Bst 중합효소, Sac 중합효소, 3'에서 5' 엑소뉴클레아제 활성을 갖는 Klenow 단편 중합효소, 및 이의 변형, 수정된 산물 및 파생물이 있다. PCR에 사용할 수 있는 추가 중합효소와 중합효소 특성이 PCR에 어떤 영향을 미칠 수 있는지에 대한 자세한 내용은 화학적 방법 섹션 D를 참조할 수 있다.
- [0023] 본 명세서에 사용될 때 용어 "중"은 일반적으로 동일한 서열의 하나 이상의 DNA 분자(들)을 지칭한다. "중"이 복수 의미로 사용되는 경우, 복수의 중에 포함된 모든 중은 개별 서열을 가지고 있다고 가정할 수 있지만 이는 때때로 "중" 대신 "개별 중"을 써서 명시적으로 나타낼 수 있다.
- [0024] 디지털 정보, 가령, 이진 코드 형태의 컴퓨터 데이터가 서열 또는 심볼 스트링을 포함할 수 있다. 이진 코드는 예를 들어 비트라고 하는 일반적으로 0과 1인 2개의 이진 심볼을 갖는 이진수 시스템을 사용하여 텍스트 또는 컴퓨터 프로세서 명령을 인코딩하거나 나타낼 수 있다. 디지털 정보는 비-이진 심볼(non-binary symbol)의 시퀀스를 포함할 수 있는 비-이진 코드의 형태로 표현될 수 있다. 각 인코딩된 심볼은 고유한 비트 스트링(또는 "바이트")에 다시 할당될 수 있으며 고유한 비트 스트링 또는 바이트는 바이트의 스트링 또는 바이트 스트림(byte stream)으로 배열될 수 있다. 주어진 비트에 대한 비트 값은 두 개의 심볼(가령, 0 또는 1) 중 하나일 수 있다. N 비트의 스트링으로 구성될 수 있는 바이트는 총  $2^N$ 개의 고유 바이트-값을 가질 수 있다. 예를 들어, 8비트로 구성된 바이트는 총  $2^8$ 개 또는 256개의 가능한 고유 바이트-값을 생성할 수 있으며, 256개의 바이트 각각은 바이트로 인코딩될 수 있는 256개의 가능한 개별 심볼, 문자 또는 명령 중 하나에 대응할 수 있다. 미가공 데이터(가령, 텍스트 파일 및 컴퓨터 명령)는 바이트의 스트링 또는 바이트 스트림으로 표현될 수 있다. 미가공 데이

터로 구성된 zip) 파일 또는 압축 데이터 파일은 바이트 스트림으로도 저장될 수 있으며, 이들 파일은 압축 형식의 바이트 스트림으로 저장된 다음 컴퓨터에서 판독되기 전에 미가공 데이터로 압축해제될 수 있다.

- [0025] 본 개시 내용의 방법 및 시스템은 복수의 식별자로 컴퓨터 데이터 또는 정보를 인코딩하는 데 사용될 수 있으며, 이들 각각은 원본 정보의 하나 이상의 비트를 나타낼 수 있다. 일부 예에서, 본 개시의 방법 및 시스템은 각각 원본 정보의 2비트를 나타내는 식별자를 사용하여 데이터 또는 정보를 인코딩한다.
- [0026] 디지털 정보를 핵산으로 인코딩하는 이전 방법은 비용이 많이 들고 시간이 많이 소요될 수 있는 핵산의 염기별 합성에 의존해 왔다. 대체 방법은 디지털 정보를 인코딩하기 위한 염기별 핵산 합성에 대한 의존도를 줄임으로써 효율성을 향상시키고, 디지털 정보 저장의 상업적 생존 가능성을 향상시키며, 모든 새로운 정보 저장 요청에 대해 개별 핵산 서열의 신규(de novo) 합성을 제거할 수 있다.
- [0027] 새로운 방법은 복수의 식별자, 또는 핵산 서열에서 염기별 또는 신규(de-novo) 핵산 합성(가령, 포스포르아미다이트 합성)을 의존하는 대신 구성요소의 조합 배열을 포함하는 디지털 정보(가령, 이진 코드)를 인코딩할 수 있다. 따라서, 새로운 전략은 정보 저장의 첫 번째 요청에 대해 개별 핵산 서열(또는 구성요소)의 제1 세트를 생성할 수 있으며, 이후 후속 정보 저장 요청에 대해 동일한 핵산 서열(또는 구성요소)을 재사용할 수 있다. 이들 접근 방식은 정보를 DNA로 인코딩하고 기록하는 과정에서 핵산 서열의 신규 합성 역할을 줄임으로써 DNA-기반 정보 저장 비용을 크게 줄일 수 있다. 더욱이, 각 염기를 각 신장 핵산에 주기적으로 전달할 수 있는 염기별 합성, 가령, 포스포라미다이트 화학 기반 또는 주형이 없는 중합효소 기반 핵산 신장과 달리, 정보를 DNA로 변환하는 새로운 방법은 구성요소의 식별자 구성을 사용하여 작성하는 것은 주기적 핵산 신장을 반드시 사용하지 않는 고도로 병렬화 가능한 프로세스이다. 따라서 새로운 방법은 기존 방법에 비해 디지털 정보를 DNA에 기록하는 속도를 높일 수 있다.
- [0028] 핵산 분자(가령, DNA)를 사용하여 디지털 정보를 기록하고, 저장하며, 판독하고, 계산을 수행하는 시스템, 장치 및 방법을 포함하는 기술이 본 명세서에 기재된다. 예를 들어, 이러한 기술에는 디지털 정보를 기록하고, 저장하며, 불러오고, 판독하며, 계산/조작하기 위한 하나 이상의 개별 또는 블록 주소 지정 가능 전극 마이크로어레이 또는 나노어레이를 포함하는 장치가 포함된다.
- [0029] 이 기술에는 데이터를 인코딩하거나 다른 인코딩 방식을 사용하여 데이터를 직접 표현하기 위해 국소 전기장을 사용하여 올리고뉴클레오타이드 단편을 조합적으로 조립하는 방법이 포함된다. 이 기술에는 (가령, 결할 효율성을 계산함으로써) 조립 프로세스의 품질 관리를 수행하기 위한 방법이 포함된다. 이 기술에는 전기장이나 열을 사용하여 합성된 핵산 분자, 가령, DNA의 후처리(가령, 짧은 불완전 생성물 필터링)를 수행하는 방법이 포함된다. 이 기술에는 주소 지정 가능한 위치로부터의 핵산 분자에 인코딩된 데이터를 불러오기 위한 방법이 포함된다. 이 기술에는 핵산 분자, 가령, DNA로부터 저장되어 있거나 불러와진 디지털 정보를 판독하기 위한 방법이 포함된다. 이 기술에는 하나 이상의 저장된 위치로부터 검색된 핵산 분자, 가령, DNA를 사용하여 계산을 수행하는 방법이 포함된다.
- [0030] 이 기술에는 DNA 저장 및/또는 계산과 관련된 응용분야를 위해 핵산 분자, 가령, DNA를 조작하기 위해 국소 전기장을 사용하는 기술이 포함된다. 본 명세서에 기재된 기술에는 마이크로어레이 또는 그 밖의 다른 소형화된 어레이와 같은 어레이가 포함되며, 이는 마이크로 또는 나노 규모이다. 마이크로어레이의 각 요소(가령, 아래에 설명될 판)는 0.1 마이크로미터<sup>2</sup> 내지 1 mm<sup>2</sup>, 1 마이크로미터<sup>2</sup> 내지 10,000 마이크로미터<sup>2</sup>, 10 마이크로미터<sup>2</sup> 내지 1,000 마이크로미터<sup>2</sup>, 또는 100 마이크로미터<sup>2</sup> 내지 500 마이크로미터<sup>2</sup>의 면적을 가질 수 있다. 나노-어레이의 각 요소(가령, 아래에 설명될 판)는 0.1 나노미터<sup>2</sup> 내지 1 마이크로미터<sup>2</sup>, 1 나노미터<sup>2</sup> 내지 10,000 나노미터<sup>2</sup>, 10 나노미터<sup>2</sup> 내지 1,000 나노미터<sup>2</sup>, 또는 100 나노미터<sup>2</sup> 내지 500 나노미터<sup>2</sup>의 면적을 가질 수 있다. 본 명세서에 기재된 기술에는 서로 가까이 배치된 한 쌍의 대전된 구조물에 전압이 인가됐을 때 유도되는 전기장을 포함한다. 핵산, 가령 DNA는 음전하를 띤 분자다. 핵산이 전기장에 놓일 때 분자에 힘이 유도되는데, 이 힘은 분자의 전하와 전기장의 세기에 따라 달라진다.
- [0031] 일부 구현예에서, 본 명세서에 기재된 기술은 전기장-중속적 힘을 이용하여 특정 마이크로어레이 위치에서 핵산, 가령, DNA를 선택적으로 끌어당기거나 밀어낸다. 이러한 국소화된 전기장은 (소형화된) 전도성 판, 가령, 금속판을 충전하고 이를 상대 전극과 평행하게 배치함으로써, 생성될 수 있다. 도 1은 복수의 전기 전도성 판(110)과 상대 전극(130)을 포함하는 예시 채널(100)의 개략적인 사시도를 보여준다. 복수의 핵산 분자를 함유한 유체는 판(110) 및 상대 전극(130)과 평행하게 유동할 수 있으며, 예를 들어 핵산을 저장소(reservoir)에서 어

레이로 운반할 수 있다. 복수의 전도성 판(110)과 상대 전극(130)은 채널의 (제1) 차원을 따라 서로 반대편에 배치된다. 각 판(110)과 상대 전극(130)의 일부는 셀(150)을 형성한다. 셀의 전기 전도성 판은 마이크로 스팟(micro spot)이라고도 불린다. 판(110)과 셀(130)은 표면에 여러 개의 어레이 및 패턴, 예를 들어 정사각형 어레이, 직사각형 어레이 또는 원형 어레이로 배열될 수 있다. 도 2는 소형화된 금속 판(110)의 예시 5x5 (마이크로)어레이(200)의 개략적인 평면도를 보여준다. 도 3은 확대되어 복수의 블록(본 설명 사례에서는 25개 셀로 구성된 9개 블록)으로 세분화된 예시 전극 마이크로어레이의 개략적인 평면도를 보여준다. 마이크로어레이 셀 각각은 개별적으로 또는 하나 이상의 다른 세포, 가령, 하나의 블록 내 모든 셀과 함께 어드레싱될 수 있다.

[0032] 어레이 내의 각 셀은 예를 들어, 본 명세서에 기재된 바와 같이 디지털 정보를 복수의 식별자 분자로 인코딩하는 방식을 사용해, 인코딩된 디지털 정보를 저장하는 데 사용될 수 있는 하나 이상의 핵산 분자를 선택적으로 끌어당기거나 밀어내도록 구성되어 있다. 핵산을 조작하기 위해(가령, 본 명세서에 아래에 기재되어 있고 도면 57-94에 도시된 정보 읽기, 쓰기 및/또는 계산 작업), 마이크로어레이 셀 각각은 도 4a 및 4b에 도시된 것과 같은 어드레싱 방식을 사용하여 독립적으로 어드레싱되고 충전될 수 있다. 도 4a에는 25개의 판(110)을 포함하는 예시적 5x5 어레이(202)가 도시되어 있다. 도 4b는 16개의 판(110a)을 갖는 예시적 동적 랜덤-액세스 메모리(DRAM) 셀 어레이(203)의 개략도이다. 행 디코더(222)와 열 디코더(224)를 사용하여 행 라인과 열 라인이 어드레싱될 때, 대응하는 셀(150)/판(111)이 켜지고, 예를 들어 셀에 전기장이 유도된다. 셀은 커패시터 역할을 할 수 있다. 셀의 전압 레벨에 따라, 셀 내의 커패시터는 충전되거나 방전된다. 각 셀 내 평행 판 커패시터의 상대 전극(130)(상부 판)은 용액 내의 핵산 분자(가령, DNA)를 끌어들이거나 밀어내는 데 사용될 수 있다.

[0033] 도 5는 본 명세서에 기재된 바와 같은 셀(150)의 예시 단일 판(110)의 단면도를 보여준다. 예시 판(110)은 전기 전도성 판, 가령, 가령, a) 어댑터 분자(112), 가령, 본 명세서에 기재된 층 0 구성요소에 상보적인 어댑터에 대한 고체 지지부, 및/또는 b) (가령, 상대 전극(130)과 함께) 전기장을 생성하기 위한 전극 쌍의 전극들 중 하나를 제공하는 금속 부분(114)을 포함한다. 셀(150)의 판(110)은 또한 기저 층(118) 및 상기 기저 층(118)과 전기 전도성 금속 부분(114) 사이에 배치된 유전체 층(116)을 포함한다. 기저 층(118)은 도체, 반도체 또는 둘 다이거나 이를 포함할 수 있다. 도 6은 판(110)의 전기 전도성 층(가령, 금속 층(114))과 상대 전극(130)이 어떻게 연결되어 전기장을 생성하고 이를 통해 힘을 생성할 수 있는지를 설명하는 예시 셀(150)을 보여준다. 상대 전극은 예를 들어 전기 전도성 판을 마주보도록 배치된 유전체 층(132)을 포함할 수 있다.

[0034] 본 명세서에 설명된 시스템은 1개, 2개 또는 3개(또는 그 이상)의 유전체 층, 가령, 1) 기저 층(118) 상에 배치된 유전체 층, 가령, 유전체 층(116), 2) 전도성 층(전도성 판)(도시되지 않음) 상에 배치된 유전체, 및/또는 3) 상대 전극(130) 상에 배치된 유전체, 가령, 유전체 층(132)을 포함할 수 있다. 일부 구현에서, 채널(100)은 서로 평행한 2개의 전극(가령, 전기 전도성 판(110)과 상대 전극(130)) 사이에 형성된 전해질로 채워진 반응 챔버이거나 이를 포함한다. 이들 두 전극 사이에 직류(DC) 전압이 인가될 때, 전극들 사이에 전기장이 생성되고 애노드와 캐소드 사이에 전류가 흐른다. 전기 전도성 판(110)은 캐소드일 수 있고 상대 전극(130)이 애노드일 수 있으며, 그 반대의 경우도 마찬가지다. 전자의 흐름으로 인해 발생하는 전류와는 달리, 이 경우 전류는 이온의 흐름으로 인해 이온 전류가 발생한다. 이온 전류는 전도성 물질이나 유체, 가령, 전해질, 전선, 또는 플라즈마에서 관찰될 수 있는 전하의 흐름을 말한다. 일부 구현에서, 유전체 층이 한쪽 전극 또는 양쪽 전극 상에 배치되는 경우(가령, 유전체 층(116 및/또는 132)), 반응 챔버는 평행판 커패시터처럼 작동할 수 있으며 전극들 사이에 DC 전류가 흐르지 않는다. 이러한 구성은 평행인 판들 사이에 일정한 전기장을 생성한다. 이 전기장은 전기장 내의 모든 하전 입자에 유도력을 발생시킨다. 예를 들어 음전하를 띤 분자인 핵산, 즉 DNA가 전해질에 현탁되어 있으면 전기장에 놓이면 힘을 받을 수 있다. 이 메커니즘은 전기장이 존재하는 상황에서 핵산 분자, 가령, DNA 분자의 인력이나 반발력을 높이는 데 사용될 수 있다. 일부 구현예에서, 셀(150)과 그 유전체 층은 셀에 부착된 이중 가닥 뉴클레오티드를 변형시키기에 충분히 강한 전기장을 생성하도록 구성된다.

[0035] 유전체의 품질과 두께는 평행인 판들, 가령, 전기 전도성 판(110)과 상대 전극(130) 사이의 전류 흐름을 조절하거나 방지하는 데 역할을 할 수 있다. 예를 들어, 제조 중, 예를 들어 유전체(산화물) 증착 중에, 금속 이온의 존재로 인해 전하 트랩이 형성될 수 있으며, 이는 산화물의 무결성을 감소시키고 풀-프렌켈 전류(Poole-Frenkel current)라고 하는 전류 흐름을 촉진하는 데 영향을 줄 수 있다. 이러한 트랩 기반 전류는 더 두꺼운(100 nm) 유전체 층에서도 발생할 수 있다. 건조 염소 열 산화물의 증착을 포함하는 방법이 이러한 금속 이온을 제거하여 전류 흐름을 방지할 수 있다. 이러한 높은 무결성 산화물이 있더라도, 얇은 산화물은 터널링을 통해 약간의 전류를 누출할 수 있다. 두께가 2 nm 미만인 산화물 층은 직접 터널링에 기여할 수 있는 반면, 두께가 5-20 nm인 산화물 층은 파울러-노르드하임(Fowler-Nordheim) 터널링(정확한 삼각형 또는 둥근 삼각형 장벽을 통과하는 전자의 파동 역학적 터널링)에 기여하는 것으로 알려져 있다. 일부 구현에서, 본 명세서에 기재된 장치는 두께가

적어도 100 nm(가령, 100 내지 200 nm, 200 내지 400 nm, 또는 400 내지 1000 nm)의 두께인 하나 이상의 유전체 층(가령, 층(116 및/또는 132))을 포함할 수 있다. 일부 구현에서, 본 명세서에 기재된 장치는 산화물(가령, 높은 무결성 산화물)을 포함하는 하나 이상의 유전체 층(가령, 층(116 및/또는 132))을 포함할 수 있으며, 예를 들면: 1) 기저 층(118) 상에 배치되는 유전체 층, 가령, 유전체 층(116), 2) 전도성 층(전도성 판)(도시되지 않음) 상에 배치되는 유전체, 및/또는 3) 평행인 전도성 판 사이에 임의의 전류 흐름을 방지하기 위해, 적어도 100 nm의 두께를 갖는, 상대 전극(130) 상에 배치되는 유전체, 가령, 유전체 층(132).

[0036] 디지털 정보를 핵산 서열로 변환하거나 핵산 서열에서 디지털 정보를 변환하기 위한 예시적 시스템은, 앞서 기재된 바와 같이, 핵산, 가령, DNA를 처리하기 위한 유체 시스템에 통합된 어레이, 가령, 마이크로 어레이(200, 201, 202 또는 203)를 포함한다. 도 7에 도시된 바와 같은 예시적인 시스템은 핵산 분자의 풀을 포함하는 유체를 보관하도록 구성된 출발지 저장소(source reservoir)(300)를 포함한다. 시스템은 앞서 기재된 바와 같이, 전기 전도성 판과 상대 전극을 포함하는 복수의 셀(150)을 포함하는 메인 채널(101)을 포함하며, 복수의 전도성 판과 상대 전극은 메인 채널(101)의 첫 번째 차원을 따라 서로 반대편에 배치된다. 이 시스템은 표적 핵산 또는 폐기된 핵산을 포함하는 유체를 보관하기 위한 도착지 저장소(destination reservoir)(400)를 포함한다. 본 시스템은 출발지 저장소(300)와 유체 연통하는 입력 채널(102) 및 메인 채널(101)을 포함하며, 상기 입력 채널(102)은 출발지 저장소(300)로부터의 (제1) 복수의 핵산 분자를 포함하는 제1 유체 볼륨을 메인 채널(101)로 분배하도록 구성된다. 시스템은 메인 채널(101) 및 도착지 저장소(400)와 유체 연통하는 출력 채널(103)을 포함하며, 상기 출력 채널(103)은 메인 채널(101)로부터의 제2 유체 볼륨을 도착지 저장소(400)로 분배하도록 구성된다. 시스템은 본 명세서에 아래에 기재되어 있고 도 57 내지 94에 도시된 바와 같이 정보 판독, 기록 및/또는 계산 작업에 사용될 수 있다.

[0037] 본 명세서에 기재된 시스템의 예시적 구현은 도 8에 도시되어 있다. 예시적 시스템은 다음을 포함하는 구획을 포함할 수 있는 출발지 저장소(300)를 포함한다: i) 디지털 정보를 인코딩하기 위한 하나 이상의 식별자 분자의 핵산(가령, DNA) 구성요소(가령, 구성요소 A-D)를 함유하는 유체를 위한 하나 이상의 구획(이러한 유체를 DNA 잉크라고도 함) ii) 품질 관리(QC) 층의 핵산(가령, DNA) 분자를 함유하는 유체를 위한 하나 이상의 구획; 및 iii) 버퍼 용액(가령, 세척 버퍼액)을 위한 하나 이상의 구획. 예시적 시스템은 버퍼의 선택적 흐름을 제어하기 위한 하나 이상의 밸브(105)와 압력 구동 흐름을 유도할 수 있는 하나 이상의 유체 펌프(106)를 포함한다. 메인 채널(101)은 셀(150)의 마이크로어레이를 포함하는 유입구(입력 채널(102))와 배출구(출력 채널(103))를 갖춘 하나 이상의 반응 챔버 중 하나로서 (가령, "칩"의 형태로) 구성된다. 예시적 시스템은 하나 이상의 도착지(가령, 폐기물) 저장소(400)를 포함한다. 예시적 구현에서, 유체는 펌프(106)에 의해 유도되고 밸브(105)에 의해 제어되는 압력 구동 흐름에 기초하여 DNA 잉크/버퍼 저장소(300)에서 폐기물 저장소(400)로 이동한다. 예를 들어, 구성요소 뉴클레오티드 0의 예시 층 0(A)이 반응 챔버로 이동되려면 펌프(106)가 압력 구동 흐름을 유도하고 층 0(A)의 DNA 잉크를 담고 있는 저장소를 제외한 모든 DNA 잉크 저장소에 대해 밸브(105)가 닫혀야 한다. 시스템은 본 명세서에 아래와 같이 기재되어 있고 도 57 내지 94에 도시된 바와 같이 정보 판독, 기록 및/또는 계산 작업에 사용될 수 있다.

[0038] 일부 구현예에서, 본 명세서에 기재된 시스템은 하나 이상의 핵산 판독 장치, 예를 들어 하나 이상의 판독기 모듈을 포함한다. 판독기 모듈은 나노채널, 나노포어, 또는 제로 모드 도파관 또는 이들의 조합일 수 있거나 이를 포함할 수 있다.

[0039] 일부 구현예에서, 본 명세서에 기재된 바와 같이 핵산에 인코딩된 디지털 정보를 기록 및/또는 판독하기 위한 시스템은 하나 이상의 나노포어 판독기를 포함할 수 있다. 나노포어 판독기는 핵산이 단백질 나노포어를 통과할 때 전류의 변화를 모니터링한다. 그 결과로 나오는 전기 신호는 특정 핵산(가령, DNA 또는 RNA) 서열을 제공하기 위해 디코딩된다. 본 명세서에 기재된 기술과 함께 사용될 수 있는 나노포어 판독기 모듈 500의 예가 도 9a 및 9b에 도시되어 있다. 나노포어 판독기(500)(가령, 나노포어 판독기 모듈)는 판(110) 및 하나 이상의 나노포어(501)를 갖는 복수의 셀(150)을 포함한다. 나노포어(501)는 예를 들어, 도 9b에 도시된 바와 같이 셀(150)의 블록, 예를 들어 블록의 중앙에 위치할 수 있다. 도 9a의 개략적 단면도는 기저 층(118)과 유전체(산화물)(116)를 포함하는 예시 기관에 대한 나노포어(501)의 위치를 보여준다. 판독기는 메인 채널(101)과 캐비티(502)를 분리하는 층(가령, 유전체(116))의 반대편에 배치된 캐비티(502)를 포함한다. 전해질은 메인 채널(101)과 캐비티(502)를 채운다. 메인 채널(101)은 본 명세서에 기재된 대로 화학적 판독, 기록 및/또는 컴퓨팅 작업을 위한 반응 챔버 역할을 한다.

[0040] 예를 들어 판독 작업을 수행하기 위해, 전기 전도성 판(110)(금속 층(114))과 상대(상부) 전극(130) 사이에 전기장을 인가하여, 핵산(10)(가령, ssDNA)을 방출함으로써, 핵산(10)에 저장된 데이터가, 가령, dsDNA가 하나 이

상의 셀(150)로부터 추출된다. (가령, 그 직후) 전압이 상대 전극(130)과 기본 전극(131) 사이의 캐비티에 인가된다. 이 전압은 셀(150)로부터의 ssDNA 분자가 나노포어(501)를 통해 전좌하도록 강제한다. 용액 내에 DNA가 전혀 없을 경우, 관찰되는 전류는 주로 나노채널을 통과하는 이온의 흐름으로 인해 발생한다. ssDNA가 용액 내에 존재하고 나노포어(501)을 통해 전좌할 때, 포어 내의 저항이 일시적으로 증가하여 염기가 나노포어(501)을 통과할 때 전류가 감소한다. 전류의 크기는 나노포어(501)를 통과하는 염기의 종류에 따라 달라진다. 측정된 전류는 전기 감지 회로를 통해 컴퓨터로 보고되어 추가 분석 및 시퀀스 호출을 위해 사용된다.

[0041] 일부 구현예에서, 본 명세서에 기재된 바와 같이 핵산에 인코딩된 디지털 정보를 기록 및/또는 판독하기 위한 시스템은 하나 이상의 나노채널 판독기를 포함할 수 있다. 나노채널은 나노포어와 비슷한 방식으로 작동한다. 본 명세서에 기재된 기술과 함께 사용될 수 있는 나노채널 판독기 모듈(1000)의 예가 도 10a 및 10b에 도시되어 있다. 나노채널 판독기(1000)는 판(110)을 포함하는 셀(150)의 블록 내에 배치되며 중앙 전극(1030)을 포함한다. 기질(1005)(가령, 기저 층 및 유전체(산화물))과 나노채널 영역을 제외한 아래의 벽에 의해 지지되는 얇은 물질 층으로 형성된 천장(1040) 사이에 하나 이상의 나노채널(1010)이 형성된다. 각 나노채널에는 또한 나노채널(1010)을 통한 DNA 분자의 전좌를 전기적으로 검출하는 전용 센서(1020)가 포함되어 있다. 블록의 경계를 정의하는 블록 전극(1050)은 이와 중앙 전극(1030) 사이에 전기장을 인가하여 DNA 분자가 나노채널(1010)을 통해, 예를 들어 배출구 및 저장소(400) 방향으로 전좌하도록 강제하는 데 사용된다. 일부 구현예에서, 중앙 전극(1030)은 DNA 분자를 배출구 및 저장소(400) 쪽으로 전좌시키도록 구성된 유체 시스템의 일부인 하나 이상의 유체 채널을 포함할 수 있다.

[0042] 일부 구현예에서, 셀(150)은 판(110)의 금속 층(114)의 표면 상에 고정화된 dsDNA 분자를 포함하며, 한 가닥은 금속 층(114)의 표면에 고정되어 있다. 예를 들어 읽기 작업을 수행하기 위해, 각 셀(150)의 상대 전극(130)과 전기 전도성 금속 층(114) 사이에 전압이 인가된다. (가령, 그 직후)블록 전극(1050)과 중앙 전극(1030) 사이에 전압이 인가된다. 이 전압은 ssDNA 분자가 나노채널(1010)을 통해 전좌하도록 강제한다. 각 나노채널(1010) 내의 센서(1020)는 DNA가 나노채널을 통해 전좌하는 동안 전류의 변화를 검출한다. 이들 전류의 변화는 전기 검출 회로에 의해 판독된다. 그런 다음 이 데이터는 추가 분석 및 서열 콜링을 위해 컴퓨터로 전송된다.

[0043] 일부 구현예에서, 본 명세서에 기재된 바와 같이 핵산에 인코딩된 디지털 정보를 기록 및/또는 판독하기 위한 시스템은 하나 이상의 제로 모드 도파관(zero mode waveguide)(ZMW) 판독기를 포함할 수 있다. 제로 모드 도파관이란, 광의 파장에 비해 모든 차원에서 작은 공간으로 광 에너지를 유도하는 광학 도파관을 말한다. 제로 모드 도파관은 얇은 금속 필름으로 제조된 광학 나노구조를 포함할 수 있으며, 이를 통해 여기 공간(excitation volume)을 아토리터(attoliter) 범위로 제한할 수 있다. 이러한 작은 밀폐 공간 덕분에 생리학적으로 적절한 농도의 형광 표지된 생체 분자에서 단일 분자 형광 실험을 수행할 수 있다.

[0044] 본 명세서에 기재된 기술과 함께 사용될 수 있는 제로 모드 도파관 판독기 모듈(600)의 예가 도 11a 및 11b에 도시되어 있다. 제로 모드 도파관 판독기(600)(가령, 제로 모드 도파관 판독기 모듈)은 판(110)과 하나 이상의 나노포어 제로 모드 도파관(601)을 갖는 복수의 셀(150)을 포함한다. 제로 모드 도파관은 셀(150)의 블록 내에, 가령, 도 11b에 도시된 바와 같이 블록의 중앙에 위치할 수 있다. 도 11a의 개략적 단면도는 기저 층(118)과 유전체(산화물)(116)를 포함할 수 있는 예시 기판에 대한 도파관(601)의 위치를 보여준다. 도파관(601)은 투명하고 기저 층(118) 위에 위치하는 유전체 층(116) 상에 또는 그 내부에 생성될 수 있다. ssDNA 분자가 도파관(601) 내부에 포획될 때, 새로운 가닥이 합성될 때 혼입된 각 염기에 대해 형광 신호가 방출된다. 이 신호는 기저 층(118) 내 캐비티(602)를 통해 광학적 여기 및 검출 시스템(603)에 의해 검출될 수 있다.

[0045] 셀(150)에 대한 데이터를 판독하기 위해, 전기 전도성 판(110)(가령, 금속 층(114))과 상대 전극(130) 사이에 전기장을 인가함으로써 해당 셀로부터 ssDNA가 방출된다. 방출된 ssDNA 분자(10)는 제로 모드 도파관(601)으로 확산되거나 셀(150)과 제로 모드 도파관 전극 사이에 전기장을 인가함으로써 제로 모드 도파관(601)으로 강제로 이동될 수 있다. 중합효소는 제로 모드 도파관 내부에서 고정화되어 있다. ssDNA가 제로 모드 도파관에 도달할 때, 중합효소는 프라이머와 형광 라벨링된 뉴클레오티드의 존재 하에 상보적 가닥을 합성할 수 있다. 각 뉴클레오티드의 혼입이 광학 검출 시스템(603)에 의해 검출될 수 있는 형광 신호를 생성한다. 디지털화된 정보는 추가 분석 및 서열 콜링을 위해 컴퓨터(CPU/GPU)로 전송될 수 있다.

[0046] 본 명세서에 기재된 기술은 상기 설명한 바와 같은 센서 장치를 포함한다. 일부 구현예에서, 센서 장치는 나노-채널의 단부, 예를 들어 원위(하류) 단부에 배치된다. 이러한 센서 장치의 예로는 전기/전자 감지 장치, 예를 들어, 전해질 산화물 전계 효과 트랜지스터(EOSFET)가 있다.

[0047] 앞서 설명한 바와 같이, 전하 기반 DNA 시퀀싱은 고체 상태 또는 유기 나노포어를 사용하여 DNA 분자가 나노포

어를 통해 전좌할 때 이온 전류를 측정하여 수행될 수 있다. 이온 전류 측정 기술은 민감도와 확장성이 제한적일 수 있다. 본 명세서에는 나노-포어 기반 핵산 시퀀싱의 감도와 확장성을 개선하기 위한 나노-포어와 결합된 평면 전해질 산화물 전계 효과 트랜지스터(EOSFET)를 포함하는 장치를 포함하는 기술이 설명되어 있다.

[0048] 일반적으로 나노-포어 FET 장치는 염기 단위 시퀀싱의 요건을 충족하도록 설계된다. 단일 DNA 염기는 펼쳐질 때 길이가 약 0.7 nm이고, 나선형으로 되었을 때 길이는 약 0.35 nm이다. 염기 단위 시퀀싱을 수행하려면, FET 채널(가령, n-채널)의 높이가, 핵산(가령, DNA) 분자가 공극(pore)을 통해 전좌할 때 전하 변화를 검출할 수 있을 만큼 충분히 민감하도록 0.7 nm 미만이어야 한다. 이런 FET 구조를 제작하는 것은 번거로울 수 있다. 예를 들어, FET의 소스 및/또는 드레인을 도핑(화학적으로 다른 원소와 결합시켜 반도체 물질의 전기 전도도를 조절)하기 위한 어닐링은 고온에서 수행되는데, 이로 인해 채널로 이온이 확산되어 채널의 감도가 떨어질 수 있다.

[0049] 본 명세서에는 구성요소 수준에서 핵산(가령, DNA) 분자를 시퀀싱하는 방법을 포함하는 기술이 설명되어 있다. 식별자(식별자 구성요소) 내의 핵산(가령, DNA) 구성요소의 한 예는 약 30개의 염기 길이일 수 있다. 이는 컴팩트한 2차 구조를 가진 판독 구성요소가 펼쳐진 상태에서 약 30개의 염기 또는 21 nm 길이가 될 것임을 의미한다. 이는 약 10-20 nm의 채널 높이(또는 길이)를 가진 나노-포어 FET가 판독 구성요소의 2차 구조 내의 전하를 민감하게 측정하는 데 충분하다는 것을 의미한다. 채널 치수가 증가하면 이온 주입 및 어닐링과 같은 기존 방법을 사용하여 더 작은 채널을 만드는 것보다 제작 공정이 간단해지므로 나노-포어 기반 시퀀싱의 견고성 및/또는 민감성이 향상된다.

[0050] 본 명세서에 기재된 기술과 함께 사용될 수 있는 결합된 나노포어-전계 효과 트랜지스터(FET) 판독기 모듈(700)의 예가 도 12에 도시되어 있다. 나노포어-FET 판독기(700)(가령, 나노포어-FET 판독기 모듈)는 판(110)과 하나 이상의 나노포어(701)를 갖는 복수의 셀(150)을 포함한다. 나노포어(701)는 예를 들어, 셀(150)의 블록, 예를 들어, 도 9b의 나노포어 판독기(500)에 대해 설명된 바와 같이 블록의 중앙에 위치할 수 있다. 도 12의 개략적 단면도는 기저 층(118)과 유전체(산화물)(116)를 포함하는 예시 기판에 대한 나노포어(701)의 위치를 보여준다. 판독기는 메인 채널(101)과 캐비티(702)를 분리하는 층(가령, 유전체(116))의 반대편 상에 배치된 캐비티(702)를 포함한다. 핵산 분자(10)를 함유한 전해질은 메인 채널(101)과 캐비티(702)를 채운다. 메인 채널(101)은 본 명세서에 기재된 대로 화학적 판독, 기록 및/또는 컴퓨팅 작업을 위한 반응 챔버 역할을 한다. 나노포어 판독기(500) 또는 나노포어-FET 판독기(700)와 같은 장치는 예를 들어 습식 화학 에칭을 통해 캐비티(502, 702)가 구성된 실리콘 웨이퍼로부터 제작될 수 있다. 메인 채널(101)은 예를 들어, 다른 실리콘 웨이퍼를 캐비티 또는 폴리머 구조와 웰(well)-타입 구조를 결합하여 생성될 수 있다. 챔버는 유전체 층 또는 멤브레인(116), 예를 들어, 실리콘 디옥사이드 또는 실리콘 니트라이드 멤브레인에 의해 분리된다. 예시 유전체(116)는 직경이 10nm 미만인 나노-포어를 포함한다. 일부 구현예에서, 나노-포어의 직경은 5 내지 10 nm이다. 일부 구현예에서, 나노-포어의 직경이 20 nm 미만이다. 예를 들어, 도 12-14에 도시된 바와 같이, 전계 효과 트랜지스터(FET)(703)가 멤브레인에 부착된다.

[0051] 나노포어-FET(703)의 예시적 구성의 구조가 도 13 및 14에 도시되어 있다. 예시적 FET로는 n-채널 고갈형 전해질 산화물 전계효과 반도체(EOSFET)가 있다. 기존 MOSFET과 달리, 게이트(금속 층)가 전해질 용액으로 대체되었다. FET는 고농도로 도핑된 n형 실리콘(실리콘과 다른 원소가 결합되어 전자가 음전하를 띠는 상태)의 소스 영역(711), 고농도로 도핑된 n형 실리콘의 드레인 영역(712), 반대 극성(p형)의 약하게 도핑되거나 도핑되지 않은 실리콘의 기판 영역(713), 및 약하게 도핑된 n형 실리콘으로 형성된 좁은 n-채널(714)을 포함한다. 이러한 구성은 게이트 전압이 0일 때 소스(711)와 드레인(712) 사이에 n-채널 714를 형성하여, 드레인-소스 전압(drain-to-source voltage)이 인가될 때 소스와 드레인 사이에 전류가 흐르게 된다. 게이트에 음 전압이 인가될 때, n-채널 폭이 줄어들고 다수 캐리어(전자)가 고갈되어 소스-드레인 전류가 감소한다. n-채널은 산화물 층(유전체(116)) 근처에 음전하를 띠므로써 고갈될 수 있으며, 이를 통해 게이트(전해질)와 채널 사이에 전기장이 유도된다. 음 전하를 도입하는 것은 외부에서 음 전압을 인가하는 것과 같다.

[0052] 앞서 기재된 EOSFET 기술은 본 명세서에 기재된 시퀀싱 기술, 가령, 구성요소 수준 시퀀싱과 함께 사용될 수 있다. 일부 구현예에서, 음 전하는 본 명세서에서 이하에서 기재될 바와 같이, 핵산(가령, DNA) 식별자 구성요소의 포스페이트 백본 및/또는 단일 가닥 식별자 모듈에 혼성화된 판독 구성요소 내 2차 구조로부터 기원한다. 식별자 또는 식별자-판독 구성요소 복합체가 나노-포어-FET를 통해 전좌할 때, 소스-드레인 전류는 분자에 존재하는 전하에 따라 감소한다. 2차 구조가 없는 예시적 영역이 적은 전류 감소를 생성할 수 있지만, 2차 구조를 포함한 전하가 높은 영역에서는 큰 전류 감소를 생성할 수 있다. 채널 도핑 농도에 따라, 소스-드레인 전류의 감소는 식별자 및/또는 2차 구조의 전하 크기에 비례할 수 있다. 따라서 입력 시퀀스의 각 구성요소가 다른 양의

음전하로 태깅되는 경우, 나노포어-FET가 사용되어 구성요소 수준에서 핵산(가령, DNA) 분자의 시퀀스를 검출할 수 있다.

[0053] 도 14는 본 명세서에 기재된 바와 같은 예시 나노포어-FET의 구성요소의 치수를 나타낸다. 소스(711)와 드레인(712)은 저농도로 도핑되거나 도핑되지 않은 p형 실리콘으로 만들어진, 기관(713)의 비교적 넓은 영역에 의해 분리된 고농도로 도핑된 n형 실리콘이다. 이러한 분리는 소스-드레인 전압이 인가될 때 소스와 드레인 사이의 전자 터널링 효과를 최소화한다. 채널(714) 근처의 소스 및 드레인 영역은 폭이 5~10nm가 되도록 설계되었다. 일부 구현에서, 채널 근처의 드레인 영역의 폭이 1 내지 50 nm가 될 수 있다. 예시 나노포어-FET의 채널 근처의 기관 영역은 나노포어(701)와 길이가 같을 수 있고, 소스 및 드레인 영역과 너비가 같을 수 있다. 소스 영역과 드레인 영역 사이의 나노포어 근처 기관을 저농도로 도핑함으로써 3 내지 5nm 크기의 작은 n형 채널을 형성할 수 있다. 소스(711), 드레인(712) 및 기관(713)에 대한 실리콘 층(두께 10 내지 20 nm)은 비정질 실리콘 FET를 형성하기 위해 증착되거나 SOI(Silicon-On-Insulator) 웨이퍼의 실리콘 층을 얇게 만들어 얻을 수 있다. 채널의 두께(높이)는 채널 길이(소스와 드레인 간의 거리)보다 센서 감도에 더 큰 영향을 미칠 수 있다. 일부 구현에서는, 채널의 길이는 나노-포어(701)의 직경보다 작거나 같습니다. 채널이 얇을수록 센서에 대한 민감도가 더 높아진다. SOI 웨이퍼의 경우, 이온 주입과 어닐링을 이용한 도핑 방법이 수행될 수 있다. 비정질 실리콘 FET의 경우, 별도의 도핑 단계를 거치지 않고도 다양한 농도의 도핑된 실리콘을 증착할 수 있다. 두께가 5 내지 10 nm 인 유전체 층(SiO<sub>2</sub> 또는 Si<sub>3</sub>N<sub>4</sub>)이 소스, 드레인, 기관 및 채널 영역을 덮을 수 있다. 유전체 층은 전해질에서 분자의 비특이적 흡착을 방지하는 분자로 부동태화될 수 있다.

[0054] 본 명세서에 기재된 바와 같은 EOSFET 감지 장치, 예를 들어, 나노포어-FET(703)은 10, 20, 30, 40, 50, 100, 1,000개 또는 그 이상의 EOSFET 감지 장치를 포함하는 감지 장치의 어레이 중 하나가 될 수 있다.

[0055] 본 명세서에 기재된 예시 시스템의 아키텍처 다이어그램이 도 15에 도시되어 있다. 예시 시스템은 본 명세서에 설명된 대로 마이크로어레이 칩, 가령, 하나 이상의 반응 챔버 중 하나로 구성된 메인 채널(101)을 포함하는 칩을 포함한다. 칩은 마이크로어레이 전극 또는 셀(가령, 셀(150))을 블록으로 그룹화하여, 가령, 도 3에 도시된 바와 같이 단위 또는 블록으로 배열하여 포함한다. 핵산(가령, DNA) 분자는 전기 전도성 판, 가령, 판(110)(마이크로 스팟이라고도 함)에 조립 및/또는 배치될 수 있다. 예시 시스템은 판독기 모듈을 포함한다. 판독기 모듈은 본 명세서에 기재된 바와 같이, 나노포어(가령, 나노포어 판독기(500)), 나노채널(가령, 나노채널 어레이(1000)), 제로 모드 도파관(가령, 제로 모드 도파관 판독기(600)) 또는 나노포어 FET 판독기(700)를 포함하거나 이를 포함할 수 있으며, 예를 들어 각 블록 내부 또는 블록 사이에 위치할 수 있다. 일부 구현에서, 하나 이상의 판독기 모듈이 각 블록에 포함되거나 유동적으로 연결될 수 있다. 일부 구현에서는, 하나 이상의 판독기 모듈이 전체 칩의 하나 이상 또는 모든 셀에 연결될 수 있다. 일부 구현에서는, 판독기 모듈은 칩의 별도 부분에 위치하거나, 마이크로어레이 칩에 유체 및 전기적으로 연결된 별도 판독기 칩 상에 위치할 수도 있다. 마이크로어레이 칩은 하나 이상의 가열기 요소, 예를 들어 저항 가열기 또는 펠티에 가열기(Peltier heater)(가령, 장치의 한 쪽에서 다른 쪽으로 열을 전달하는 고체 능동 히트 펌프) 및 특정 온도를 유지하기 위한 가열기 제어기를 포함할 수 있다. 가열기는 하나 이상의 전기 전도성 판, 가령, 판(110)을 가열하도록 구성될 수 있다. 일부 구현에서, 각 셀이 별도의 히터 요소를 포함하거나 연결되어 각 셀이 독립적으로 가열될 수 있다. 일부 구현에서는, 두 개 이상의 셀(150)이 가열기 요소를 공유한다.

[0056] 본 명세서에 기재된 마이크로어레이 칩은 메인 채널(가령, 메인 채널(101))의 일부가 될 수 있으며 유체 라인 또는 채널의 세트(가령, 입력 채널 102 및/또는 출력 채널(103))에 연결되어 칩을 하나 이상의 유체 저장소(가령, 출발지 저장소(300) 및/또는 도착지 저장소(400))에 연결할 수 있다. 본 명세서에 설명된 시스템은 하나 이상의 펌프(가령, 펌프(106))와 밸브(가령, 밸브(105))를 포함하는 유체 펌핑 시스템을 포함할 수 있으며, 이를 통해 시스템을 통한 유체의 흐름을 제어할 수 있다. 저장소 및 대응하는 유체 라인은 흐름 방향에 따라 입력 또는 출력으로 지정될 수 있다. 유체 펌핑 시스템의 밸브와 펌프는 저장소와 칩 사이의 액체 흐름을 제어하는 데 사용할 수 있다. 유체 펌핑 시스템의 작동은 메모리와 CPU를 포함한 컴퓨팅 시스템에 의해 제어될 수 있다. 제어기나 드라이버 회로는 전선을 통해 컴퓨터와 칩 사이에 연결되어 신호(가령, 아날로그 신호)를 전달할 수 있다. 제어기는 통신 버스(가령, I2C, SPI)를 통해 컴퓨터로부터 명령을 수신하고 이를 마이크로유체 칩에 전달하는 전기 신호로 변환하여, 예를 들어, 한 세트 이상의 셀에 전기장을 유도하거나 가열기를 활성화한다. 제어기는 또한 주어진 셀의 전압을 읽고 컴퓨터에 보고할 수 있다. 제어기 회로는 아날로그와 디지털 구성요소의 조합일 수 있다.

[0057] 본 명세서에 기재된 시스템은 앞서 기재된 판독기 모듈 및/또는 셀로부터 신호를 판독하고 해당 신호를 추가 분

석 또는 저장을 위해 컴퓨터의 CPU 또는 GPU(그래픽 처리 장치)로 전송하기 위한 전기 감지 회로를 포함할 수 있다. 이 검출 회로는 전선을 통해 칩에 연결되고, 통신 버스를 통해 컴퓨터에 연결될 수 있다. 나노포어 또는 나노채널이 판독 기술로 사용되면 회로망은 핵산(가령, DNA) 분자가 나노포어 또는 나노채널을 통해 전좌할 때 아날로그 전류를 측정하고 이를 디지털 값으로 변환하여 CPU나 GPU에 보고할 수 있다. 이 시스템은 광학 구성요소(가령, 렌즈, 광섬유, 편광판 등)와 광학 검출기(가령, 카메라, 광자 카운터 등)를 포함하는 광학 검출 시스템을 포함할 수 있으며, 이를 통해 셀이나 판독 모듈에서 나오는 광학 신호를 측정하고 디지털화된 값을 컴퓨터에 보고한다. 예를 들어, 제로 모드 도파관이 판독 기술로 사용되는 경우, 광학 시스템 내의 카메라가 도파관에서 나오는 형광 강도를 검출하고 형광 신호를 디지털 신호로 변환할 수 있다.

[0058] 본 명세서에 기재된 시스템은 예를 들어 도 15에 도시된 바와 같은 컴퓨터 시스템을 포함할 수 있다. 컴퓨터 시스템은 CPU, GPU, 메모리, 저장장치, 주변장치 및/또는 연관된 소프트웨어를 포함할 수 있다. 컴퓨터가 사용되어 시스템과 시스템의 모든 구성요소를 제어할 수 있다. 이 시스템의 GPU는 칩이 생성한 데이터를 기반으로 머신 러닝을 포함한 분석을 수행하는 데 사용될 수 있다.

[0059] 본 명세서에 기재된 마이크로어레이 칩을 포함하는 시스템은 다양한 구성으로 사용되어 DNA에 인코딩된 데이터에 대한 기록, 판독, 저장, 계산 및/또는 QC를 수행할 수 있다. 본 명세서에는 시스템의 하나 이상의 구성요소를 활용하는 예시 워크플로가 설명되어 있다. 모든 작업은 컴퓨터(프로세서와 명령이 저장된 메모리 포함)에서 하위 시스템이나 구성요소에 전송된 명령어를 통해 제어할 수 있다. 예를 들어, 워크플로우 명령문에 "제어기가 ...를 발행합니다"라고 명시되어 있는 경우, 컴퓨터가 제어기에 명령/커맨드를 전송하고, 제어기는 해당 명령을 처리하여 칩이나 다른 구성요소와 신호를 주고받는다 하는 것을 의미합니다.

[0060] 본 명세서에 설명된 바와 같이 시스템을 운영하기 위한 예시적인 워크플로우는 다음과 같은 프로세스를 포함할 수 있다(핵산 분자의 예로 DNA가 사용됨).

[0061] 일부 구현예에서, 본 명세서에 기재된 기술은 DNA를 사용하여 데이터를 기록하기 위한 시스템 및 방법을 포함한다. 도 16에 예시 작업 흐름이 나와 있다. 하나 이상의 출발지 저장소(가령, 출발지 저장소(300))는 사전-합성된 올리고뉴클레오티드(구성요소라고 함, 가령, DNA(10))가 들어 있으며, 그 길이는 약 30개 염기이다. 복수의 셀(150)을 갖는 칩의 온도는 전기 전도성 판(110) 상에 배치된 구성요소 및 어댑터 분자(112)의 점착 말단의 녹는점( $T_m$ )보다  $5^{\circ}\text{C}$  낮은 온도까지 가열기를 사용하여 상승된다. 제어기는 특정 셀에 전압을 인가한다. 펌핑 시스템(가령, 펌프(106))은 하나의 저장소의 내용물(핵산 구성요소가 포함된 유체)을 반응 챔버로 구성된 메인 채널(101)로 유동시킨다. 이 용액은 칩 상의 어댑터에 대한 구성요소의 혼성화를 가능하게 하기 위해 유한한 시간 동안 반응 챔버에 보관된다. 반응 챔버를 세척하기 위해, 펌핑 시스템은 반응 챔버의 내용물(혼성화되지 않은 임의의 구성요소를 포함)을 폐기물 저장소(가령, 도착지 저장소(400))로 유동시킨다. 그런 다음 펌핑 시스템은 출발지 저장소에서 세척 버퍼액을 반응 챔버로 유동시키고 폐기물 저장소로 유동시킨다. 특정 셀에 전압을 인가하고, 구성요소가 함유된 용액을 유동시키며, 구성요소의 혼성화를 허용하기 위해 용액을 유지하고, 챔버를 씻어내는 단계는 데이터 세트에 사용된 구성요소가 함유된 모든 입력 저장소에 대해 반복된다. 펌핑 시스템(가령, 펌프(106))은 리가제를 함유한 용액을 출발지 저장소로부터 반응 챔버로 유동시킨다. 용액은 결합 반응이 완료될 때까지 한정된 시간 동안 반응 챔버에 보관된다. 펌핑 시스템은 세척 버퍼액을 반응 챔버로 유동시키고 폐기물 저장소로 유동시킨다. 이제 셀은 DNA에 기록될 데이터를 보유하게 되었다. 이 데이터는 선택적으로 추출되어 다른 곳에, 가령, 본 명세서에 설명된 대로 저장될 수 있다.

[0062] 일부 구현예에서, 본 명세서에 기재된 기술에는 칩 외부 저장 또는 처리를 위한 데이터 추출을 위한 시스템 및 방법이 포함된다. 이 사용 사례에서는, 앞서 설명한 대로, 데이터가 이미 칩 상의 뉴클레오티드에 기록되어 칩에 저장되어 있다고 가정한다. 펌핑 시스템(가령, 펌프(106))은 출발지 저장소(300)로부터 버퍼액을 유동시킨다. 다음 기술 (a)-(c) 중 하나 이상을 사용하여 데이터가 추출될 수 있다: (a) 칩의 온도가 전체 길이 분자의  $T_m$ 보다 높게 증가된다. 이를 통해 데이터가 포함된 단일 가닥 DNA를 녹일 수 있다. 칩 상에 남아 있는 DNA에는 여전히 인코딩된 데이터가 담겨 있다. (b) 제어기는 전압이 하나 이상의 셀(150)에 인가되게 하여 한 가닥의 DNA가 표면으로부터 녹아내리고 데이터가 포함된 다른 가닥이 칩의 판(110)에 남게 된다. (c) 데이터가 포함된 dsDNA를 방출하기 위한 어댑터(112)의 바닥의 제한 소화 효소가 칩으로 유동할 수 있다. 펌핑 시스템은 반응 챔버로부터 수집 저장소(가령, 도착지 저장소(400))로 버퍼액을 유동시킨다.

[0063] 일부 구현예에서, 본 명세서에 기재된 기술에는 나노포어 판독기 모듈, 가령, 나노포어 판독기 모듈(500)이 있는 칩에 저장된 데이터를 판독하기 위한 시스템 및 방법이 포함된다. 도 17에 예시 작업 흐름이 나와 있다. 펌핑 시스템(가령, 펌프(106))은 출발지 저장소(300)에서 칩의 반응 챔버로서 구성된 메인 채널(101)로 판독기 버

퍼액을 유동시킨다. 제어기는 ssDNA를 추출하기 위해 데이터가 판독되어 올 셀(150)에 전압이 인가되도록 한다. 제어기는 칩의 상대 전극(130)과 기저 전극(131) 사이에 전압이 인가되도록 한다. 분자는 나노포어(501)를 통해 전좌하여 나노포어 내 전류 변화를 발생시킨다. 전기 검출 회로는 전류를 측정하고 그 전류 값을 컴퓨터에 보고한다. CPU/GPU는 전류 값에 대한 분석을 수행하고 전류 판독치를 DNA 서열(가령, 염기 콜링)로 변환한다. CPU/GPU는 모든 서열을 분석하고 데이터를 디코딩한다.

[0064] 일부 구현에서, 본 명세서에 기재된 기술에는 제로 모드 도파관 판독기 모듈, 가령, 도파관 판독기 모듈(600)을 사용하여 칩 상에 저장된 데이터를 판독하기 위한 시스템 및 방법이 포함된다. 도 18에 예시 작업 흐름이 나와 있다. 펌핑 시스템(가령, 펌프(106))은 출발지 저장소(300)로부터 칩의 반응 챔버로서 구성된 메인 채널(101)로 판독기 버퍼액을 유동시킨다. 제어기는 ssDNA를 추출하기 위해 데이터가 판독되어 올 셀(150)에 전압을 인가한다. 분자는 도파관(601)으로 확산되거나, 도파관과 상대 전극(130) 사이에 전압이 인가된다. 분자는 앞서 설명한 바와 같이 광 펄스를 생성함으로써 도파관(601)에 의해 분석된다. 광학 검출 시스템(603)은 광 강도를 측정하고, 광학 신호를 디지털 신호로 변환하고, 컴퓨터에 데이터를 보고한다. CPU/GPU는 광 강도의 디지털 값을 분석하고 이 값을 DNA 서열로 변환한다(가령, 염기 콜링 기술을 적용). CPU/GPU는 서열을 분석하고 데이터를 디코딩한다.

[0065] 일부 구현에서, 본 명세서에 기재된 기술에는 칩에 저장된 정보를 계산하기 위한 시스템 및 방법이 포함된다. 펌핑 시스템(가령, 펌프(106))은 출발지 저장소(300)로부터 칩의 반응 챔버로 구성된 메인 채널(101)로 버퍼액을 유동시킨다. 제어기/CPU는 데이터가 판독되어 올 하나 이상의 셀(150)(출발지)에 전압을 인가하여, 출발지 셀로부터 DNA 분자(10)를 방출한다. 제어기/CPU는 하나 이상의 상이한 셀(150)(도착지)에 상이한 전압을 인가해 분자를 도착지 셀로 강제로 이동시킨다. 펌핑 시스템은 출발지 저장소(300)에서 효소와 형광단을 유동시켜 도착지 셀 상의 분자에 작용한다. 광학 검출 시스템은 출발지 셀과 도착지 셀에서 광학 신호를 검출하고 해당 신호를 컴퓨터에 보고한다. 출발지 분자는 피연산자이고, 효소와 형광단은 연산자이며, 도착지 셀 내 분자는 연산의 결과이다.

[0066] 일부 구현에서, 본 명세서에 기재된 기술에는 본 명세서에 기재된 대로 기록 단계 후 불완전한 생성물 분자를 정리하기 위한 시스템 및 방법이 포함된다. 펌핑 시스템(가령, 펌프(106))은 입력 저장소에서 칩의 반응 챔버로 구성된 메인 채널(101)로 버퍼액을 유동시킨다. 제어기는 모든 셀(150)에 특정 전압을 인가하여 불완전한 생성물만 녹인다(가령, 불완전한 생성물을 변성시키기에 충분히 강하지만 완전한 생성물을 변성시키기에는 너무 약한 전기장을 인가함). 불완전한 생성물은 완전히 형성된 생성물의 녹는점보다 낮은 온도에서 열 변성될 수도 있다. 펌핑 시스템은 반응 챔버로부터 도착지 저장소(400), 가령, 폐기물 배출 저장소로 버퍼액을 유동시킨다. 칩 상에 남아 있는 dsDNA는 본 명세서에 기재된 사용 사례에 따라 판독하거나 저장하기 위해 추출될 수 있는 전체 길이의 분자이다.

[0067] 일부 구현에서, 본 명세서에 기재된 기술에는 기록(write)의 효율성을 측정하기 위한 시스템 및 방법이 포함된다. 펌핑 시스템(가령, 펌프(106))은 태깅된 DNA 분자를 함유하는 품질 관리(QC) 버퍼액을 출발지 저장소(300)에서 칩의 반응 챔버로서 구성된 메인 채널(101)로 유동시킨다. 광학 시스템은 각 셀(150)의 형광을 측정하고, 형광 값을 정량화하고, 디지털화된 정보를 컴퓨터로 전송한다. CPU/GPU는 셀(150)에 대한 분석을 수행하고, 기록 효율성(셀(150)에서 가능한 전체 분자에 대한 전체 길이의 분자의 백분율)을 계산한다. 제어기는 QC 분자를 제거하기 위해 모든 셀(150)에 전압을 인가한다. 펌핑 시스템은 반응 챔버를 세척하기 위해 반응 챔버를 통해 버퍼액을 도착지 저장소(400)로 유동시킨다.

[0068] 일부 구현에서, 본 명세서에 기재된 기술에는 기록된 분자의 길이 분포를 측정하기 위한 시스템 및 방법이 포함된다. 펌핑 시스템(가령, 펌프(106))은 상이한 층에 대응하는 상이한 형광단을 갖는 DNA 분자의 혼합물을 포함하는 QC 버퍼액을 칩의 반응 챔버로서 구성된 메인 채널(101)로 유동시킨다. 분자들은 하나 이상의 셀(150)에 의해 포획된다(선택 사항). 광학 시스템은 각 색상에 대해 각 셀로부터의 형광 값을 측정하고, 형광을 정량화하고, 디지털화된 정보를 컴퓨터로 전송한다. CPU/GPU는 셀(150)에 대한 분석을 수행하고 분자 길이의 분포를 계산한다.

[0069] 예시

[0070] 본 명세서에는 디지털 정보를 인코딩하기 위해 결찰을 사용하여 핵산 분자(가령, DNA)의 더 작은 단편을 조립하는 접근 방식을 기반으로 핵산(가령, DNA)을 사용하여 디지털 정보를 기록하고, 저장하며, 판독하고, 계산하는 기술이 설명되어 있다. 본 명세서에는 핵산에 디지털 정보를 인코딩하는 시스템 및 방법이 설명되어 있다. 이른바 "식별자"에 디지털 정보를 인코딩하기 위해 결찰될 수 있는 예시 구성요소는 도 19에 나와 있다. 구성요소는

중양의 이중 가닥 영역에 고유한 서열이 포함되어 있고 한 층의 오버행(overhang)이 인접한 층의 구성요소에 상보적인 "층"으로 그룹화될 수 있다. 일부 구현에서는, 가장자리 층(층 0 및 층 n(가령, 3))은 평활 말단(blunt end)을 가질 수 있다. 일부 구현에서는 가장자리 층이 점착 말단(sticky end)을 갖도록 조정될 수 있다. 이 기술에는 층 n에 상보적이며 형광단을 함유하는 세그먼트를 갖는 단일 가닥 핵산(가령, DNA)을 갖는 선택적 QC 층이 포함된다.

[0071] 본 명세서에 기재된 시스템 및 방법의 한 가지 예시적 응용 분야는 결찰을 통해 짧은 DNA 단편을 조합적으로 조립하여 더 긴 (식별자) 분자를 구축하는 것이다. 분자를 구축하기 위해, 하나 이상의 셀(150)에 국소 전기장을 인가하여 상이한 셀 또는 위치에 있는 DNA 분자를 끌어당기거나 밀어내는 동안, 핵산(가령, 이른바 "DNA 잉크")을 함유한 용액이 본 명세서에 기재된 바와 같이 칩의 반응 챔버로서 구성된 메인 채널(101)로 순차적으로 유동된다. 도 20 내지 27에 제시된 예에서, DNA 분자의 두 가지 독특한 조합이 조립되었다. 먼저, 첫 번째 위치(가령, 셀)(도 20의 왼쪽 셀)에 인력을 인가하고 두 번째 위치(가령, 셀)(도 20의 오른쪽 셀)에 척력을 인가하는 동안  $A_0^0$  (층 0, 구성요소 0)으로 지정된 핵산 분자(10)를 함유하는 DNA 잉크가 시스템을 통해 유동된다. 이는 위치 1의 표면에 있는 어댑터에 혼성화되는  $A_0^0$  를 야기한다. 그런 다음, 버퍼액 행굼을 통해 통합되지 않은 DNA 구성요소를 제거한다(도 21). 다음으로, 위치 2에서는 인력을, 위치 1에서는 척력을 인가하는 동안  $A_0^1$  로 지정된 핵산 분자를 함유한 DNA 잉크가 시스템을 통해 유동된다. 이는 위치 2의 표면에 있는 어댑터에 혼성화되는  $A_0^1$  를 야기한다(도 22). 버퍼액 행굼은 통합되지 않은 DNA 구성요소를 씻어 낸다(도 23).

[0072] 이 예에서 다음 단계는  $B_1^0$  로 지정된 핵산 분자를 함유하는 DNA 잉크를 유동시키고 위치 1과 위치 2 모두에 인력을 인가한 후 버퍼액 행굼을 수행하는 것이다(도 24). 이로 인해 두 위치(왼쪽 셀과 오른쪽 셀) 모두에서 층 0 구성요소에 혼성화되는  $B_1^0$  가 도출된다. DNA 잉크를 유동시키고 인력이나 척력을 인가하여 과잉 핵산(가령, DNA)을 씻어내는 이러한 접근 방식을 계속하면 핵산(가령, DNA) 단편의 고유한 조합이 생성된다(도 25-27). 이들 단편은 결찰 반응을 통해 고유한 핵산(가령, DNA) 서열을 형성하도록 결찰될 수 있다. 지금까지 설명한 단계는 정보에 대한 기록 과정을 구성한다. 이 스테이지에서는, 조립된 DNA 가닥이 칩 상의 각자의 위치에 저장될 수 있다. 이 경우, 칩은 전통적인 전자 메모리 칩과 비슷하지만 정보가 핵산(가령, DNA)에 인코딩되어 있다. 핵산(가령, DNA) 가닥은 열, 전기장 또는 효소 제한을 적용하는 방법 등을 통해 추출할 수 있으며, 이를 수집하여 용액이나 동결 건조하여 보관할 수 있다.

[0073] 일부 구현에서, 기술에는 조립된 핵산(가령, DNA) 가닥에 대해 취할 수 있는 하나 이상의 품질 관리(QC) 조치가 포함된다. 예를 들어, 예를 들어, QC 층을 형광단과 혼성화하고 방출된 광자를 정량화함으로써(도 27), 핵산(가령, DNA)의 정량화는 임의의 위치(셀(150))에서 수행될 수 있다. 결찰 기반 조립의 문제들 중 하나는 반응이 효율적이지 않을 수 있다는 것이다. 완전히 형성된 핵산(가령, DNA) 생성물의 수는 구성요소 핵산 및/또는 그 생성물의 총량의 일부에 불과할 수 있다. 반응을 완료하지 못한 더 작은 단편은 도 29에 도시된 바와 같이 셀에 부착될 수 있다. 나머지 생성물에 비교되는 완전히 형성된 생성물의 백분율이 결찰 효율성을 나타낼 수 있다. 도 29의 왼쪽 셀은 형광단을 포함하는 전장 식별자가 적기 때문에 형광량이 적은 반면, 오른쪽 셀은 모든 식별자에 형광단이 있다. 결찰 효율성을 정량화하기 위한 기존 접근 방식에는 몇 가지 시간-소모적 작업, 가령, 모나치 클린업(monarch cleanup), 겔 추출 및/또는 모든 단계에서 손실이 발생하는 qPCR이 필요하다. 본 명세서에 기재된 예시적 방법에서 이들 단계가 후처리 단계 없이 단일 형광 기반 판독으로 대체된다. 이 과정을 통해 며칠이 아니라 몇 시간에 달하는 후속처리 시간을 절약할 수 있다.

[0074] 일부 구현에서 불완전하게 형성된 생성물의 분포가 평가될 수 있다. 이 메트릭은 상이한 길이의 핵산(가령, DNA) 가닥에 대한 결찰 효율성을 나타낼 수 있다. 기존 방법은 DNA 정제/클린업 및 (자동화된) 전기영동 실행이 필요할 수 있는데, 이는 모든 단계에서 손실이 발생하고 전기영동의 감도가 낮은 단점이 있다. 본 명세서에 기재된 기술을 구현하는 일부 경우, 서로 다른 형광단을 서로 다른 길이의 핵산(가령, DNA)에 부착하고 방출된 광자를 정량화함으로써 자동화된 방식으로 분산형 QC 판독 결과를 얻을 수 있다(도 30). 도 30의 왼쪽 셀은 셀의 다른 층에 혼성화된 다른 형광단의 존재로 인해 다른 파장에서 형광을 나타내는 반면, 오른쪽 셀은 모든 식별자가 동일한 형광단을 가지고 있다. 이 장치가 메모리 칩으로 사용되는 경우, 가령, 도 31에서 각각 첫 번째 위치

및 두 번째 위치에 나타난 바와 같이, 한 가지 바람직한 빠른 판독 결과는 데이터가 있는 위치나 데이터가 없는 위치를 식별하는 것일 수 있다. QC 층의 형광은, 도 32에 나타난 바와 같이, 예시 어레이의 형광 맵을 얻는 데 사용될 수 있다.

[0075] 일부 구현에서, 본 명세서에 기재된 방법을 통해 핵산(가령, DNA)에 기록된 정보를 읽기 위한 후처리 단계는 불완전한 결찰 생성물을 제거하는 것을 포함할 수 있다. 불완전한 제품으로 인해 시퀀싱 및 판독 중에 노이즈가 증가할 수 있다. 기존 방법은 겔 추출을 통해 노이즈를 줄이는데, 이는 수동적이고 신뢰할 수 없는 방법이며 상당한 차이가 있다. 본 명세서에 기재된 기술을 구현하는 일부 방법에서는, 완전히 형성된 생성물의 녹는점 이하에서 불완전한 생성물을 열 변성시켜 불완전한 생성물을 분리하고 제거하는 단계를 포함한다. 도 33은 왼쪽 셀에서 불완전한 생성물을 제거하는 방법을 보여준다. 일부 구현에서는, 완전한 생성물을 변성시키는 데 필요한 강도보다 낮은 강도의 전기장을 사용하여 힘을 가함으로써 불완전한 생성물을 변성시킬 수 있다.

[0076] 일부 구현에서는, 완전히 형성된 생성물이 정제되면 데이터를 각각의 위치(셀)에 있는 핵산(가령, DNA)에 저장하거나 별도로 검색하여 저장할 수 있다. 이후의 처리, 저장 또는 계산을 위해 데이터를 추출하기 위해, 가령, 국소 전기장을 인가함으로써, 임의의 특정 위치로부터의 데이터가 개별적으로 추출될 수 있다. 도 34는 왼쪽 셀에서 전체 길이 식별자를 검색하는 방법을 보여준다. 데이터가 불러와지면, 예를 들어, 본 명세서에 기재된 바와 같이, 동일한 칩에서 (통합된) 나노포어, 나노채널 및/또는 제로 모드 도파관 기반 기술을 사용하여 시퀀싱될 수 있다. 또한, 주어진 위치로부터 불러와진 데이터는, 도 35에 도시된 바와 같이, 칩 내부의 계산에 사용될 수 있다. 이 예에서, "덧셈(addition)" 연산자는 상이한 위치로부터 불러와진 두 분자의 연결(concatenation)에 대한 형광 기반 판독과 함께 표시된다. 이 연산은, 현재 계산 단계는 핵산(가령, DNA)을 사용하여 수행되는 점을 제외하고, 메모리 칩의 상이한 위치에서 전자 데이터를 검색하고 메모리에서 덧셈을 수행하는 것과 유사하다. 연결 등의 계산 단계는 용액에서 수행될 수도 있거나 하나 이상의 셀에서 수행될 수도 있다.

[0077] 일부 구현에서는, 데이터는 뉴클레오티드 서열로 인코딩되거나 핵산의 길이로 인코딩될 수 있다. 핵산은 표면(가령, 셀)에 혼성화될 수 있으며 핵산의 전기적 특성(따라서 핵산에 인코딩된 정보)이 측정될 수 있다. 계산은 또한 서로 다른 셀로부터 유래한 핵산 가닥을 "결과 셀"에 공동 배치하여, 계산을 수행함으로써, 계산이 수행될 수 있다(가령, 덧셈 단계 - 두 개의 서로 다른 입력 셀로부터 유래한 식별자가 있으면 추가가 성공했음을 나타냄). 본 명세서에 기재된 계산 과정은 (가령, 선택적 변성을 통해) 가역적일 수 있어서 정보를 지우고 다시 쓸 수 있다. 일부 구현에서는 계산이나 연산은 두 개의 결합된 핵산 분자를 연결하는 작업을 포함할 수 있다. 전기 신호나 형광 신호의 변화가 검출되고 이를 통해 정보를 추출할 수 있다. 앞서 기재된 어레이 기술과 함께 사용될 수 있는 예시적인 데이터 저장, 불러오기 및/또는 계산 기술이 본 명세서에서 아래에 기술되어 있다.

[0078] 핵산 분자(가령, DNA)를 사용하여 디지털 정보를 기록하고, 저장하며, 판독하고, 계산을 수행하는 시스템, 장치 및 방법을 포함하는 기술이 본 명세서에 기재된다. 예를 들어, 이러한 기술에는 나노포어, 나노-채널 및/또는 센서를 사용하여 전하 핵산 가닥의 하나 이상의 구성요소를 검출하는 핵산 서열을 판독하기 위한 장치 및 방법이 포함된다. 본 명세서에 기재된 판독 장치 및 방법은 독립형 장치일 수 있거나 이를 포함할 수도 있고, 하나 이상의 개별 또는 블록-주소 지정 전극 마이크로어레이 또는 나노어레이(가령, 어레이(200, 201, 202, 203))를 포함하는 장치에 통합되어 디지털 정보를 기록하고, 저장하며, 불러오고, 판독하며, 계산/조작할 수도 있다.

[0079] 현재의 DNA 시퀀싱 기술은 한 번에 하나의 염기를 읽어서 판독 속도가 느리다. 이러한 기술의 예로는 합성 시퀀싱 기술, Illumina® 유형 시퀀싱, Oxford Nanopore® 유형 시퀀싱 기술 등이 있다. 예를 들어, 서열에 반복이 있거나 잘 알려진 패턴이 있는 특정 응용 분야의 경우, 한 번에 한 염기씩 핵산(가령, DNA)을 시퀀싱할 필요가 없을 수 있다. 대신, 본 명세서에 기재된 바와 같이 더 높은 수준, 예를 들어 염기 세트(가령, 식별자의 "구성요소")에서 핵산 서열(가령, DNA)을 판독하는 접근 방식은 매우 효율적이고 빠를 수 있다. 기존 접근 방식은 DNA가 시퀀싱될 수 있기 전에 절단 효소 및 RecA 단백질을 사용하여 핵산 서열(가령, DNA)을 수정해야 하기 때문에 번거로울 수 있다. 본 명세서에 기재된 기술은 단일 혼성화 단계만 포함할 수 있기 때문에 더 단순화된 접근 방식을 제공할 수 있다.

[0080] 본 명세서에 기재된 기술은 핵산 분자(가령, DNA)로부터 정보를 빠르게 시퀀싱하고/하거나 추출하는 단순화된 접근 방식을 제공할 수 있다. 본 명세서에 기재된 기술은 전류 기반 감지 기술을 사용할 수 있으며, 이를 통해 광학적 방법을 사용하는 기술에 비해 더 저렴한 시퀀싱이 가능하다. 본 명세서에 기재된 기술은 생물학적 분자(가령, 단백질)를 사용하는 다른 기술, 예를 들어 단백질 나노-포어에 비해 제조의 확장성을 향상시킬 수 있는 나노-채널을 사용할 수 있다. 고체 물질로부터 만들어진 나노-채널의 수명은 생물학적 나노포어 기술의 수명보다 더 길 수 있다.

- [0081] 본 명세서에는 본 명세서에 기재된 바와 같이 디지털 정보를 인코딩하는 염기 세트(가령, 식별자의 "구성요소")에서 핵산(가령, DNA) 서열을 판독하기 위한 장치 및 방법이 기술되어 있다. 기존의 시퀀싱 방법은 한 번에 하나의 염기를 판독하고 광학적 또는 전기적 방법을 사용하여 염기를 검출한 후 각 염기에 대한 염기 콜링을 수행한다. 본 명세서에는 한 번에 하나의 염기를 판독하는 대신 한 번에 하나 이상의 염기 세트(가령, "구성요소")를 읽어서 판독 프로세스를 가속화하는 기술이 설명되어 있다.
- [0082] 본 명세서에 기재된 기술은 (A) 시퀀싱될 구성요소(가령, 디지털 정보를 인코딩하는 핵산 식별자의 구성요소인 식별자 구성요소)에 상보적인 가령, 3' 및 5' 말단 상의 정의된 수의 염기를 갖는 단일 가닥 핵산(가령, DNA)을 포함하고 (B) 전기 전하를 제공하기 위해 컴팩트한 구조로 조직되는 (가령, 3' 말단과 5' 말단) 사이에 핵산(가령, DNA)의 서열을 포함하는 판독 구성요소 핵산(가령, DNA)을 포함한다. 판독 구성요소의 전하량은 다양할 수 있으며 DNA의 상보적인 가닥의 전하량보다 높을 수 있다.
- [0083] 본 명세서에 기재된 기술에는 시퀀싱될 DNA의 개별 구성요소(가령, 식별자 구성요소)에 상보적인 일련의 판독 구성요소가 포함되며, 각 판독 구성요소는 고유한 전하량을 가진다.
- [0084] 본 명세서에 기재된 기술에는 전자 장치, 가령, 금속 산화물 반도체 전계 효과 트랜지스터(MOSFET))를 포함하는 센서 장치가 포함되며, 이 트랜지스터의 게이트 전압은 전좌 판독 구성요소의 전하를 통해 변조되어 트랜지스터의 소스-드레인 전류를 변경할 수 있다. 센서 장치는 나노 채널이나 나노포어의 먼 쪽(다운스트림) 끝에 배치될 수 있다. 일부 구현에서, 센서 장치는 나노 채널 또는 나노포어의 가까운 쪽(업스트림) 끝에 배치될 수 있다.
- [0085] 본 명세서에 기재된 기술에는 시퀀싱될 핵산 분자에 판독 구성요소를 혼성화시키고 나노-채널 또는 나노포어를 통해 DNA를 전좌시킴으로써 센서를 사용하여 서열을 판독하는 방식으로 핵산 분자(가령, DNA)를 시퀀싱하는 방법이 포함된다.
- [0086] 본 명세서에 기재된 기술에는 개별적으로가 아닌 염기 세트(가령, 구성요소, 가령, 식별자 구성요소) 단위로 핵산 분자(가령, DNA)를 시퀀싱하는 장치 및 방법이 포함된다. 시퀀싱될 핵산 분자(가령, DNA)는 이중 가닥이거나 단일 가닥 형태일 수 있다. 일부 구현예에서, 시퀀싱될 핵산 분자(가령, DNA)는 단일 가닥 형태이다. 단일 가닥 핵산 분자는 이중 가닥 핵산 분자(가령, DNA)를 열을 가해 녹인 다음, 예를 들어 셸(150)의 판(110)을 가열하여 단일 가닥을 분리함으로써 쉽게 얻을 수 있다. 본 명세서에 기재된 기술에는 "판독 구성요소"가 포함되며, 이는 "2차 구조", 가령, 컴팩트한 구조를 형성하고 3' 말단과 5' 말단에 단일 가닥으로 유지되는 핵산 서열을 갖는 단일 가닥 핵산 분자(가령, DNA) 서열이거나 이를 포함한다. 식별자 구성요소는 약 500개 염기 길이의 고유한 서열일 수 있다. 일부 구현에서 식별자 구성요소는 10 내지 5000개 염기, 100 내지 1000개 염기, 또는 200 내지 800개 염기 길이의 고유한 서열일 수 있다. 본 명세서에 기재된 바와 같은 판독 구성요소는 시퀀싱될 핵산 분자(가령, DNA)의 특정 자리에 상보적인데, 예를 들어, 3' 및 5' 말단의 식별자 구성요소가 12-염기로 상보적이다. 일부 구현예에서, 3' 및/또는 5' 말단에서의 보체는 1 염기(b 또는 bp) 내지 50 b, 2 b 내지 40 b, 3 b 내지 30 b, 또는 4 b 내지 20 b의 길이를 가질 수 있다. 일부 구현예에서, 2차 구조는 10 b 내지 50 b, 50 b 내지 100 b, 100 b 내지 200 b, 200 b 내지 300 b, 또는 300 b 초과 길이를 가질 수 있다. 일부 구현예에서, 이러한 2차 구조는 예를 들어 루프, 파, 끈, 꼬임, 권선, 접힘, 매듭, 실물치 또는 이들의 임의의 조합의 형태를 가질 수 있다. 일부 구현에서, 판독 구성요소가 단일 2차 구조를 가진다. 일부 구현에서, 판독 구성요소는 둘 이상의 2차 구조, 가령, 2, 3, 4, 5개 이상을 가진다. 일부 구현에서, 이러한 2차 구조는 판독 구성요소의 3'과 5' 말단 사이, 판독 구성요소의 3' 말단, 판독 구성요소의 5' 말단 또는 이들의 조합에 위치할 수 있다. 판독 구성요소의 3' 및 5' 말단은 시퀀싱될 핵산 분자(가령, DNA)의 특정 자리에 결합된다. 이들 자리는 알려진 서열이나 패턴을 나타내는데, 예를 들어, 본 명세서에 설명된 대로, 디지털 정보(가령, 식별자 구성요소)를 인코딩하는 자리이다. 2차 구조는 그 치밀한 구조로 인해 가령, 나노입자처럼 작은 볼륨 안에 큰 전하를 제공한다. 대량의 디지털 정보를 작은 볼륨에 저장할 수 있는데, 예를 들어, 64킬로바이트에서 1메가바이트 이상까지 다양한 규모의 도서관이 저장될 수 있다. 시퀀싱될 단일 가닥 핵산 분자(가령, DNA)는 이후 판독 구성요소와 혼성화되어 도 36에 도시된 바와 같은 구조가 생성된다. 이 분자가 인가된 전기장에 의해 나노-채널(또는 나노포어)을 통해 전좌될 때, 핵산 분자(가령, DNA)는 도 36에서 보이는 것과 같이 펼쳐질 수 있는데, 이를 통해 판독 구성요소가 한 번에 하나의 판독 구성요소씩 감지 장치의 게이트를 통과하거나 지나갈 수 있다.
- [0087] 여기에 설명된 기술에는 센서 장치가 포함된다. 일부 구현예에서, 센서 장치는 나노 채널(또는 나노포어)의 끝, 예를 들어 원위(다운스트림) 끝에 배치된다. 이러한 센서 장치의 예로는 전기/전자 감지 장치, 가령, MOSFET가 있다. MOSFET은 소스, 드레인, 및 게이트로 구성되며, 게이트 전극을 포함할 수 있다. 게이트 전압이 MOSFET에 인가되면, MOSFET이 켜지면서 소스와 드레인 사이에 전류가 흐를 수 있다. 게이트 전압이 변하면, 소스-드레인

전류도 그에 따라 변한다. 게이트 전압은 게이트 전극 위 및/또는 근처에 전하가 도입되면 교란될 수 있다. 시퀀싱될 핵산 분자(가령, DNA)(가령, "입력 서열"을 갖는 "입력 가닥")가 혼성화된 판독 구성요소를 포함하고 MOSFET의 게이트 위 및/또는 근처로 전하하는 경우, 소스에서 드레인으로의 전류 변화를 감지할 수 있다. 전류의 변화는 판독 구성요소의 2차 구조에 있는 전하량에 따라 달라질 수 있다. 2차 구조의 부피를 제한하고 작은 부피에 큰 전하를 채우면 도 37에서 보는 것과 같이 전류가 급격히 변할 수 있다. 서로 다른 2차 구조, 예를 들어 서로 다른 길이의 핵산은 서로 다른 전하를 가질 수 있다. 따라서 전류를 측정함으로써 해당 전하가 식별될 수 있고, 이를 통해 판독 구성요소가 식별될 수 있다.

[0088] 일부 구현에서, 센서 장치는 하나 이상의 전자 신호 처리 장치이거나 이를 포함한다. 일부 구현에서, 센서 장치는 본 명세서에 설명된 대로 컴퓨팅 시스템의 하나 이상의 프로세서를 포함하거나 이에 (전기적으로) 연결된다. 컴퓨팅 시스템은 전자 감지 장치 또는 광학 감지 장치로부터 수신된 신호를 처리하도록 구성될 수 있는데, 예를 들어, 염기 콜링을 수행하거나 전기 또는 광학 신호를 서열 정보, 가령, 입력 서열의 서열 정보로 변환하는 하나 이상의 단계를 수행할 수 있다.

[0089] 일부 구현에서, 전체 핵산 분자(가령, DNA, 예를 들어, 식별자 분자)가 나노-채널 또는 나노포어를 통해 전하할 때, 측정된 전류는 시퀀싱될 DNA에 혼성화되는 다양한 판독 구성요소의 전하량에 따라 변한다(도 38 참조). 예를 들어, 판독 구성요소의 서열이 길수록 전하량이 커지고 신호가 더 두드러진다. 전류 변화의 순서를 분석하면 입력 서열의 핵산 분자(가령, DNA) 서열의 변화를 직접적으로 나타낼 수 있다.

[0090] 핵산 분자는 액체 용액, 예를 들어, 상기 설명한 바와 같은 메인 채널(101)에 현탁될 수 있다. 본 명세서에 기재된 기술과 함께 사용되는 액체 용액은 통합되지 않은 판독 구성요소 및 혼성화된 판독 구성요소가 있는 핵산 분자(가령, DNA) 가닥을 포함할 수 있다. 본 명세서에는 (가령, 현탁액 내) 개별 판독 구성요소에서 발생하는 신호와 시퀀싱될 핵산 분자(가령, DNA, 가령, 식별자 분자)에서 발생하는 신호를 구별하는 기술이 설명되어 있다. 예를 들어, 시퀀싱될 핵산 분자(가령, DNA)는 한쪽 또는 양쪽 말단에서 알려진 시작 및/또는 종료 서열, 예를 들어 판독 구성요소에 큰 전하를 갖는 서열과 연결 및/또는 혼성화될 수 있다. 시작/종료 서열에서의 큰 전하로 인해 전류가 크게 변한다(도 39). 전류 변화의 시퀀스를 분석하고 시작/종료 전류 수준을 식별함으로써, 핵산 분자(가령, DNA, 가령, 식별자 분자) 서열에 대한 전류 변화가 개별 통합되지 않은 판독 구성요소나 불완전한 생성물과 구별될 수 있다. 일부 구현에서, 통합되지 않은 판독 구성요소로 인한 노이즈를 최소화하기 위해 하나 이상의 분자 정화 방법을 사용하여 통합되지 않은 판독 구성요소를 필터링할 수 있다.

[0091] 일부 구현에서, 도 40에 도시된 바와 같이, 서열(가령, 식별자 분자의 입력 서열)의 경계부(시작/종료)는 전류의 특징적인 패턴을 생성하는 복수의 작은 2차 구조를 갖는 시작 및/또는 종료 서열을 사용하여 식별될 수 있다.

[0092] 일부 구현에서, 2차 구조를 갖는 판독 구성요소는 핵산 "기록" 프로세스의 일부로 적용될 수 있는데, 예를 들어, 입력 가닥이 조립될 때, 예를 들어, 본 명세서에 설명된 바와 같이, 복수의 식별자 구성요소로부터 조립될 때이다. 예를 들어, 본 명세서에 설명된 것처럼 이른바 "식별자"에 디지털 정보를 인코딩하기 위해 연결할 수 있는 예시 식별자 구성요소이다. 구성요소는 앞서 설명한 대로 "층"으로 그룹화될 수 있다.

[0093] 하나의 구현에서, 도 41(디자인 A)에 도시된 바와 같이, 식별자의 하나의 가닥은 복수의 식별자 구성요소(가령, 구성요소 A 및 B)를 포함한다. 예시적 상보적 가닥(판독 구성요소)에는 다음의 4개의 고유한 영역이 포함된다: (1) 이전(층 n) 구성요소에 상보적 영역, (2) 현재(층 n+1) 구성요소에 상보적 영역, (3) 비-상보적 구성요소(가령, 플랩), 및 (4) 2차 구조, 예를 들어, 높은 전하 밀도(가령, 2차 구조가 없는 DNA 분자의 전하 밀도보다 높은 전하 밀도)를 갖는 컴팩트한 2차 구조. 영역 1과 2는 상이한 층으로부터의 식별자 구성요소 A와 B가 혼성화되어 리가제가 식별자 구성요소 사이에 틈을 연결할 수 있도록 한다. 영역 3은 A나 B의 어떤 서열과도 상보적이지 않아 플랩을 생성할 수 있다. 영역 4는 본 명세서에 기재된 바와 같이 2차 구조이거나 이를 포함할 수 있으며, 여기에는 자가-혼성화에 의해 컴팩트한 2차 구조를 형성할 수 있는 (긴) 서열이 포함될 수 있다. 컴팩트한 구조는 단위 부피당 큰 전하 밀도(가령, 2차 구조가 없는 DNA 분자의 전하 밀도보다 높은 전하 밀도)를 보일 수 있다. 일부 구현에서, 2차 구조는 10 내지 50개의 염기 또는 염기 쌍(bp), 50 내지 100 bp, 100 내지 200 bp, 200 내지 300 bp, 또는 300 bp 초과 길이를 가질 수 있다.

[0094] 도 41(디자인 A)에 도시된 판독 구성요소는 본 명세서에 기재된 바와 같이 기록 프로세스 동안 조립된 식별자 분자의 한 가닥을 구성할 수 있다(가령, 도 42에 도시된 바와 같음). 서로 다른 2차 구조는 판독기, 예를 들어, 본 명세서에 설명된 나노-채널(또는 나노포어) 및 MOSFET 센서 장치를 포함하는 장치에서 서로 다른 전기 신호를 유도한다. 예를 들어, 2차 구조가 길수록 MOSFET에 흐르는 전류는 커진다. 이 기술의 장점은 읽기 프로세스

중에 판독 구성요소와 식별자 분자가 생성되면 이후의 판독 단계에서 혼성화 단계가 필요 없다는 것이다. 라이브러리에 있는 분자는 나노 채널 기반(또는 나노포어 기반) 판독기 장치에 직접 공급되어 후처리 없이 판독을 수행할 수 있다. 이 기술은 핵산(가령, DNA)의 저장 및 검색 파이프라인에서 기록과 판독 단계 사이에 있는 일부 또는 모든 후처리 단계를 제거하기 때문에, 예를 들어 핵산(가령, DNA)에서 디지털 정보를 판독하고, 기록하며, 계산하고, 저장하는 기술에 상당한 가치를 더할 수 있다.

[0095] 일부 구현에서, 입력 가닥은 본 명세서에 설명된 대로 식별자 구성요소로부터 조립된다. 상보적 가닥은, 도 43(디자인 B)에 예시된 바와 같이, 앞서 기재된 디자인 A의 구성요소의 1-3 영역만을 포함하는 구성요소(가령, "1차 판독 구성요소")를 포함한다. 일부 구현에서, 이 입력 가닥은 기록 프로세스 동안 식별자 구성요소의 혼성화(가령, 1차 판독 구성요소) 및 결찰을 허용하도록 구성된다. 서열을 판독하기 위해, 앞서 설명한 바와 같은 2차 구조를 갖는 2차 판독 구성요소가 1차 판독 구성요소의 플랩(3)에 혼성화될 수 있다. 이제 핵산 분자(가령, DNA)가 앞서 기재된 바와 같이 판독될 수 있다.

[0096] 일부 구현예에서, 본 명세서에 기재된 기술은 전기 신호 대신 광 신호를 사용하여 뉴클레오티드 분자(가령, 식별자)를 판독할 수 있다. 예를 들어, 발광 태그(가령, 형광단)는 앞서 설명한 바와 같이 2차 핵산 구조를 대신하여 또는 2차 핵산 구조와 함께 사용될 수 있다(가령, 디자인 A 및 B). 도 44(디자인 C)에 도시된 예시적인 구현에서, 판독 구성요소는 상기의 디자인 A에 대해 설명한 바와 같이 4개의 영역을 포함하도록 구조화될 수 있으나, 영역 4는 2차 구조를 갖는 핵산(가령, DNA) 대신 형광단이다. 일부 구현예에서, 본 명세서에 기재된 바와 같이 판독 장치의 나노-채널(또는 나노포어)은 광 신호를 검출하도록 구성될 수 있다. 일부 구현에서는 나노-채널(또는 나노포어)이 광학 검사를 위한 창을 포함하도록 구성될 수 있다. 예를 들어, 렌즈, 광소자, 카메라, 광자 계수기 또는 그 밖의 다른 광학 검출 또는 이미지 처리 도구 중 하나 이상을 포함하는 형광 측정 시스템을 사용하여 검출이 수행될 수 있다. 핵산 분자가 인가된 전기장에 의해 나노채널(또는 나노포어)을 통해 전좌되면, 핵산 분자(가령, DNA)는 도 44에 도시된 바와 같이 펼쳐질 수 있으며, 이를 통해 판독된 구성요소나 나노채널(나노포어)의 창을 한 번에 하나씩 통과할 수 있다. 일부 구현에서는 각 판독 구성요소의 형광이 창을 통과할 때 측정될 수 있다. 다양한 색상이나 다양한 광 강도(또는 둘 다)가 사용되어 각 판독 구성요소에 개별적 광학 특성을 부여할 수 있다. 가령, 도 45에 도시된 바와 같이, 관찰된 일련의 광학 신호는 식별자 서열로 변환될 수 있다.

[0097] 하나의 구현에서, 판독 구성요소는 도 41(디자인 A)에 나타난 것과 같이 판독 구성요소와 유사하게 구성된 경계 판독 구성요소가 될 수 있는데, 다만 영역 2는 식별자 구성요소 B의 주요 부분("페이로드 부분")을 포함하지 않도록 잘려 있다. 따라서, (경계) 판독 구성요소는 각 층의 단일 식별자 구성요소에 대응하지 않고, 도 46(디자인 D)에 도시된 바와 같이 인접 층의 식별자 구성요소의 조합에 대응한다. 이러한 접근 방식의 장점은 판독 구성요소에 필요한 재료를 줄이고 판독 프로세스를 예를 들어 2배 정도 빠르게 할 수 있다는 것이다.

[0098] 일부 구현에서는, 가령, 도 47에 도시된 바와 같이, 주어진 식별자 구성요소에 필요한 판독 구성요소의 수가 앞서 설명한 이전 설계에 필요한 것의 절반일 수 있다. 예를 들어, 각 판독 구성요소는 (두 층 사이의 경계에서) 혼성화된 두 층 모두를 식별한다. 이는 이전 설계의 두 배 속도로 읽기 작업이 수행될 수 있음을 의미한다. 예를 들어, 식별자 구성요소에 홀수 개의 층이 있는 경우, 필요한 추가 판독 구성요소가 하나 있다. 일부 구현에서는 이러한 접근 방식에는 인접 층으로부터의 구성요소의 조합에 대응하는 더욱 고유한 판독 구성요소 서열이 필요할 수 있다.

[0099] 일부 구현에서, 예를 들어, 도 48(디자인 E)에 도시된 바와 같이, 상기 설명한 바와 같은 경계 판독 구성요소(가령, 디자인 D)는 전기 신호 대신 광 신호를 사용할 수 있다. 예를 들어, 발광 태그, 가령, 형광단을 하전된 핵산(가령, DNA) 2차 구조 대신 사용할 수 있다. 나노-채널(또는 나노포어) 설계는 설계 C에 대해 설명된 대로 될 수 있다. 마찬가지로, 도 46 및 47에 도시된 바와 같이 일부 구현에서는 필요한 읽기 구성요소의 수도 줄어들 수 있다(가령, 설계 C에서 구현된 것의 절반).

[0100] 일부 구현에서, 예를 들어 도 49(디자인 F)에 도시된 바와 같이, 판독 구성요소는 다른 분자, 예를 들어 펩티드 및/또는 단백질에 결합할 수 있는 능력을 가진 작은 핵산(가령, DNA) 분자(가령, 압타머)를 포함할 수 있다. 일부 구현에서, 본 명세서에 기재된 기술에는 앞서 기재된 바와 같이, 나노-채널(또는 나노포어)에 의해 검출될 수 있는 높은 순(이동) 전하(변형되지 않은 핵산 분자와 비교 시)를 갖는 분자를 결합할 수 있는 압타머 라이브러리가 포함된다. 이 설계의 장점은 일부 구현에서는 핵산(가령, DNA) 프로브의 크기가 120개 염기보다 크지 않을 수 있다는 것이다. 그러나 순 전하는 사용된 특정 압타머를 변경함으로써 제어될 수 있다. 예를 들어, 여러 개의 음성 아미노산을 갖는 서열을 갖는 펩티드가 구현될 수 있으며, 이 펩티드를 이용하여 2차 구조가 생성될

수 있다. 각 판독 구성요소를 감지하기 위한 다양한 순 전하가 사용되어 서열을 판독할 수 있다. 일부 구현에서는, 가령, 디자인 A에 대해 앞서 설명된 바와 같이 상이한 펩티드 서열이 사용되어 각 판독 구성요소에 대해 상이한 순 전하를 적용할 수 있다.

[0101] 일부 구현에서, 예를 들어, 도 50(디자인 G)에 도시된 바와 같이, 상기 설명한 바와 같이 판독 구성요소의 신호 순 전하를 증가시키기 위해(가령, 디자인 A) 분지 구조, 예를 들어 덴드리머가 사용될 수 있다. 일부 구현예에서, 이러한 분지 구조의 분지의 끝은 예를 들어 스트렙타비딘, 비오틴 또는 프로브와 같은 유사한 결합 파트너와 같은 하나 이상의 다른 분자를 부착함으로써 변형될 수 있다. 이들 분자는 프로브의 크기를 증가시키고, 앞서 설명한 바와 같이 핵산 분자(가령, DNA) 서열을 감지하고 판독하기 위한 판독 구성요소의 순 전하를 증가 또는 감소시킬 수 있는 다른 분자의 부착을 허용할 수 있다.

[0102] 일부 구현예에서, 본 명세서에 기재된 기술은 다중센서 핵산(가령, DNA) 판독기로서 배열 또는 구성될 수 있다.

[0103] 본 명세서에는 현재 기술(가령, 현재 나노포어 시퀀싱 기술)보다 수십 배 더 빠른 속도로 핵산(가령, DNA) 서열을 판독하기 위한 장치 및 방법을 포함하는 기술이 설명되어 있다. 앞서 논의한 바와 같이, 현재의 시퀀싱 기술은 한 번에 하나의 염기씩 전기적 또는 광학적 신호를 감지하여 핵산(가령, DNA) 서열을 추론한다. 전기적 접근 방식은 핵산 분자(가령, DNA)가 나노포어를 통해 전좌할 때 이온 전류를 측정하거나 나노 전극이 있는 나노 채널을 통해 전좌할 때 분자의 전하를 측정하는 것을 포함한다. 전자 회로는 전류의 변화를 처리하여 DNA 서열을 도출할 수 있다. 현재 이들 회로는 DNA 전좌의 속도(가령, 초당 100만 개의 염기)에 미치지 못한다. 따라서 핵산 분자(가령, DNA)는 나노포어를 통해 전좌함에 따라 검출 회로의 속도와 맞춰 이동 속도가 느려지는 것이 일반적이다. 본 명세서에는 탠덤 또는 병렬로 작동하고 머신 러닝 알고리즘을 활용하는 여러 개의 "느린" 센서의 정보에 기초하여 핵산(가령, DNA) 서열을 해독하는 장치 및 방법을 포함한 기술이 설명되어 있다.

[0104] 일부 구현에서는 여러 개의 센서 장치(가령, "느린" 센서 장치)가 전좌 경로(가령, 나노포어 또는 나노 채널)를 따라 직렬로 배열되어 시퀀싱 프로세스를 몇 배나 빠르게 진행할 수 있다. 일부 구현에서, 복수의 전자 감지 장치 또는 복수의 광학 감지 장치, 또는 이들의 조합이 전좌 경로(가령, 나노포어 또는 나노 채널)를 따라 직렬로 배열된다. 핵산 분자의 전좌 속도보다 낮은 속도로 작동하는 단일 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치)(도면에서 "센서"로 표시)가 분자 내의 각 염기를 검출하지 못할 수 있는 반면, 가령, 도 51-53에 도시된 바와 같이, 적절하게 배열된 센서 장치 어레이(또는 전자 감지 장치 또는 광학 감지 장치)는 집단적으로 분자 내의 모든 염기(또는 본 명세서에서 위에 기술된 바와 같은 판독 구성요소)를 재구성하는 데 충분한 샘플을 수집할 수 있다. 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치)가 신호를 샘플링할 때, 도 51에 도시된 바와 같이, 단일 염기(또는 판독 구성요소)를 전혀 검출하지 못하거나, 잠재적으로 단일 염기(또는 판독 구성요소)만 검출할 수 있다. "느린" 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))가 핵산 분자로부터의 정보를 분해할 수 없는 또 다른 시나리오가 도 52에 나와 있다. 이 경우, 센서의 샘플링 시간이 단일 염기(또는 판독 구성요소)의 전좌 시간보다 훨씬 긴 경우, 센서는 개별 염기(또는 판독 구성요소)를 구별할 수 있는 분해능이 없을 수 있다. 전좌 경로를 따라 센서(가령, 센서 장치(또는 전자 감지 장치나 광학 감지 장치))의 수를 늘리면 다양한 센서의 일부 정보를 머신 러닝 등을 사용하여 조합하고 더욱 정확한 정보를 추론할 수 있다.

[0105] 일부 구현예에서, 본 명세서에 기재된 기술을 사용하여 판독될 수 있는 핵산(가령, DNA) 분자는, 가령, 앞서 기재되고 도 36-37에 도시된 바와 같이, 단일 가닥 핵산 분자(가령, DNA), 이중 가닥 핵산 분자(가령, DNA) 또는 2차 구조 또는 판독 구성요소를 포함하는 상보적 가닥을 갖는 하이브리드 핵산 분자일 수 있다. 일부 구현에서, 센서 또는 센서 장치는 전기/전자 감지 장치, 예를 들어, 나노 채널을 따라(또는 나노 채널 내부 또는 그 위에) 배치된 하나 이상의 MOSFET 센서 및/또는 하나 이상의 저항성 센서, 또는 나노포어 내부 또는 그 위에 배치된 하나 이상의 이온 전류 센서일 수 있거나 이를 포함할 수 있다. 일련의 센서(가령, 센서 장치(또는 전자 감지 장치나 광학 감지 장치))를 핵산(가령, DNA) 전좌 경로를 따라 정의된 간격으로 배열할 수 있다. 예시 간격은 1 내지 10 nm, 10 내지 100 nm, 100 내지 300 nm, 200 내지 500 nm, 500 내지 1000 nm일 수 있다. 간격은 일정할 수도 있고, 가변적일 수도 있다. 예를 들어, 일련의 MOSFET 센서를 500nm의 일정한 간격으로 배열할 수 있다. (가령, 나노채널의) 전좌 경로의 폭은 임의의 주어진 시간에 채널을 따라 단 하나의 분자만 전좌될 수 있도록 선택될 수 있다. 도 53은 전체 감지 용량(센서 수 곱하기 각 센서의 속도)이 전좌 속도(가령, 센서 속도: 초당 하나의 판독 구성요소 판독, 전좌 속도: 초당 5개의 판독 구성요소)와 동일한 예시 장치를 보여준다. 이 예시 사례에서, 각 센서(가령, 센서 장치(또는 전자 감지 장치나 광학 감지 장치))로부터의 신호는 센서 융합 및 머신 러닝 알고리즘을 사용하여 시퀀싱된 핵산(가령, DNA)의 서열을 도출하는 데 수집될 수 있다. 도 53에 도시된 예는 각 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))가 전좌 DNA로부터 적어도 하나의 정

보(가령, 염기 또는 판독 구성요소)를 판독하는 이상적인 경우의 예를 보여준다.

- [0106] 도 54는 센서 어레이의 전체 감지 용량이 전좌 속도와 동일하지만 각 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))가 핵산으로부터 판독될 정보 중 일부를 놓칠 수 있는 사례를 보여준다. 이 경우, 각 센서로부터 수집된 정보는 원하는 서열을 재현하지 못할 수 있다. 일부 구현에서, 다수의 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))가 전좌 경로를 따라 사용될 수 있다(가령, 도 55에 도시된 바와 같음). 이렇게 하면, 가령, 중복 스캐닝 및/또는 누락된 정보의 빈틈 채우기에 의해, 다수의 센서로부터 얻은 부분적인 정보가 조립된 정보의 정확도를 증가시킬 수 있다. 일부 구현에서, 판독 구성요소는 핵산 분자의 길이를 따라 상대적인 위치 정보를 포함할 수 있다. 일부 구현에서, 판독 구성요소의 세트가 동시에, 단독으로 또는 해당 구성요소를 서로 읽는 센서의 상대적 위치 정보와 함께, 판독 구성요소의 스트링의 값 및 위치에 대한 정보를 제공할 수 있다(가령, 센서 2와 센서 4가 각각 판독한 판독 구성요소 B와 판독 구성요소 D는 구성요소 B와 D가 두 개의 센서 간격만큼 떨어져 있음을 나타낼 수 있음). 각 구성요소를 여러 번 스캔하면 판독 구성요소 및/또는 염기 서열을 구성하는 데 사용될 수 있는 정보의 양이 증가될 수 있다. 일부 구현에서는 센서의 수가 많을수록 정확도가 높아진다.
- [0107] 일부 구현에서는, 센서(가령, 센서 장치(또는 전자 감지 장치 또는 광학 감지 장치))의 밀도를 높여서 정확도를 향상시킬 수 있다. 예를 들어, 센서가 서로 매우 가깝게 배열된 경우(가령, 3-5nm 간격), 센서 세트("클러스터")가 사용되어 가령, 도 56에서 설명한 대로 클러스터당 분자로부터의 두 개 이상의 정보(가령, 염기 또는 판독 구성요소)를 검출할 수 있다. 일부 구현에서, 판독 구성요소의 세트가 동시에, 단독으로 또는 해당 구성요소를 서로 읽는 개별 센서 및/또는 클러스터의 상대적 위치 정보와 함께, 판독 구성요소의 스트링의 값 및 위치에 대한 정보를 제공할 수 있다(가령, 센서 2와 센서 4가 각각 판독한 판독 구성요소 B와 판독 구성요소 D는 구성요소 B와 D가 두 개의 센서 간격만큼 떨어져 있음을 나타낼 수 있음). 각 구성요소 또는 구성요소의 세트를 여러 번 스캔하면 판독 구성요소 및/또는 염기 서열을 구성하는 데 사용될 수 있는 정보의 양이 증가될 수 있다. 다시 말해, 이러한 클러스터의 수가 증가하면 판독된 정보의 정확도가 높아진다.
- [0108] 일부 구현에서, 머신 러닝 알고리즘을 염기 콜링이나 하나 이상의 센서(가령, 센서 장치(또는 전자 감지 장치나 광학 감지 장치))에서 수신한 신호를 핵산 서열 정보로 변환하기 위한 그 밖의 다른 작업에서 사용할 수 있다.
- [0109] 일부 구현예에서, 본 명세서에 제공된 기술은 자연스러운 전좌 속도(초당 100만 개의 염기)와 동일한 속도로 머신 러닝을 사용하여 핵산(가령, DNA) 서열을 판독하는 것을 제공할 수 있다. 현재 상업적으로 이용 가능한 DNA 시퀀싱 속도는 초당 420개 염기이다. 본 명세서에 기재된 기술은 최소한 3배 이상 빠른 속도로 핵산 서열을 판독할 수 있다.
- [0110] 본 명세서는 디지털 정보의 검색 및 액세스의 효율성을 향상시키기 위해 다양한 방식으로 디지털 정보를 핵산 분자에 저장하기 위한 시스템 및 방법을 제공한다. 예를 들어, 구성요소 핵산 분자(가령, 구성요소)가 선택되고 서로 연결(concatenate)되어 식별자 핵산 분자(가령, 식별자)를 형성하며, 각각은 특정 심볼(가령, 비트 또는 비트 시리즈), 또는 심볼 스트링(가령, 비트스트림)에서의 그 심볼의 위치(가령, 순위 또는 주소)에 대응한다. 이러한 구성요소는 디지털 데이터를 표현하기 위한 효율적인 방식을 제공하기 위해 구조적인 방식으로 구성될 수 있다. 예를 들어, 구성요소의 구조는 구성요소 분자가 자체적으로 모아지도록 하거나, 그렇지 않으면 다수의 구성요소 분자가 동일한 구획에 보관되거나 분배된 후에 사전 결정된 순서로 스스로 정렬될 수 있다.
- [0111] 디지털 정보를 핵산 분자에 저장하기 위한 방법이 본 명세서에 제공되며, 상기 방법은 (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 다음에 의해 제1 식별자 핵산 분자를 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 분자의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 분자를 선택하는 것, (2) M개의 선택된 구성요소 핵산 분자를 하나의 구획으로 보관하는 것 -, (3) (2)의 M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하여, 제1 및 제2 층으로부터의 구성요소 핵산 분자가 식별자 핵산 분자의 제1 및 제2 말단 분자에 대응하며, 제3 층 내 구성요소 핵산 분자가 식별자 핵산 분자의 제3 분자에 대응하여, 제1 식별자 핵산 분자의 M개의 층의 물리적 순서를 정의하도록, 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치하는 제3 분자를 갖는 제1 식별자 핵산 분자를 형성함 -, (c) 복수의 추가 식별자 핵산 분자를 형성하는 단계 - 추가 식별자 핵산 분자 각각은 (1) 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, (2) 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 연속 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산

분자를 선택할 수 있게 함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계를 포함한다.

- [0112] 일부 구현에서, 식별자 핵산 분자 집단은 동일한 표적 분자를 공유하는 반면, 동일한 풀에 있는 다른 식별자 핵산 분자는 다른 표적 분자를 가질 수 있다. 적어도 하나의 추가 식별자 핵산 분자의 제1 및 제2 말단 분자 중 적어도 하나는 (b)의 제1 식별자 핵산 분자의 표적 분자와 동일할 수 있다. 일부 구현예에서, M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하는 것은 구성요소 핵산 분자의 결합을 포함한다.
- [0113] 일부 구현예에서, 각 층의 구성요소 핵산 분자는 다른 층의 구성요소 핵산 분자의 적어도 하나의 점착 말단에 상보적인 적어도 하나의 점착 말단을 포함하여, (b)와 (c)의 식별자 핵산 분자의 형성을 위한 점착 말단 결합을 가능하게 한다. 예를 들어, 각 층(예를 들어, A, B, C) 내의 모든 구성요소는 서로 동일한 점착 말단을 가질 수 있으며, 층 A 내 모든 구성요소의 하나의 점착 말단은 층 B 내 모든 구성요소의 하나의 점착 말단과 상보적이다. 더욱이, 층 B 내 모든 구성요소의 다른 점착 말단은 층 C 내 모든 구성요소의 하나의 점착 말단에 상보적일 수 있는 등이다. 일부 구현예에서, (c)의 적어도 하나의 추가 식별자 핵산 분자의 제1 분자는 (b)의 식별자 핵산 분자의 제1 말단 분자와 동일하고, (c)의 적어도 하나의 추가 식별자 핵산 분자의 제2 말단 분자는 (b)의 식별자 핵산 분자의 제2 말단 분자와 동일하다.
- [0114] 일부 구현예에서, 방법은 프로브를 사용하여 제1 식별자 핵산 분자 내의 적어도 일부 식별자 핵산 분자와 복수의 추가 식별자 핵산 분자를 표적 분자에 혼성화하여 연속적인 심볼 위치를 갖는 각각의 심볼에 대응하는 식별자 핵산 분자를 선택하는 단계를 더 포함한다. 인접한 심볼 위치를 갖는 심볼은 서로 인접해 있으며 유사한 이웃에 있기 때문에 유사한 특성을 공유할 수 있다. 따라서, 동일한 프로브를 사용하여 서로 가까이 위치하는 식별자 핵산 분자를 선택하는 것이 바람직할 수 있다. 일부 구현예에서, 방법은 연속 심볼 위치를 갖는 각각의 심볼에 대응하는 적어도 2개의 식별자 핵산 분자를 증폭시키기 위해 단일 PCR 반응을 적용하는 단계를 더 포함한다. 일부 구현에서, 인접한 심볼 위치를 갖는 각각의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자는 식별자 핵산 분자의 세 번째 분자에 있는 특정 구성요소 핵산 분자를 표적으로 하는 또 다른 PCR 반응에 의해 추가로 증폭될 수 있다.
- [0115] 일부 구현예에서, 각 층의 구성요소 핵산 분자는 제1 및 제2 말단 영역으로 구성되고, M개 층 중 하나로부터의 각 구성요소 핵산 분자의 제1 말단 영역은 M개 층 중 또 다른 층으로부터의 임의의 구성요소 핵산 분자의 제2 말단 영역에 결합하도록 구성된다. 일부 구현예에서 M은 3보다 크거나 같다. 일부 구현예에서, 심볼 스트링 내의 각 심볼 위치는 대응하는 상이한 식별자 핵산 분자를 가진다. 일부 구현예에서, (b) 및 (c)의 식별자 핵산 분자는 각각의 M개 층으로부터 하나의 구성요소 핵산 분자를 포함하는 가능한 식별자 핵산 분자의 조합 공간의 서브세트를 나타낸다.
- [0116] 일부 구현예에서, (d)의 풀에서 식별자 핵산 분자의 존재 또는 부재는 심볼의 스트링 내 대응하는 각각의 심볼 위치의 심볼 값을 나타낸다. 예를 들어, 식별자가 있으면 해당 심볼 위치의 심볼 값이 1임을 나타내고, 없으면 심볼 값이 0임을 나타낼 수 있으며, 그 반대일 수도 있다. 일부 구현예에서, 인접한 심볼 위치를 갖는 심볼은 유사한 디지털 정보를 인코딩한다. 일부 구현예에서, M개 층 각각의 구성요소 핵산 분자 수의 분포는 불균일하다. 예를 들어, 한 층은 식별자 핵산 분자를 생성하기 위한 가능한 순열의 수 및/또는 다양성을 조정하기 위해 다른 층보다 더 많은 구성요소 핵산 분자를 가질 수 있다.
- [0117] 일부 구현예에서, 제3 층이 제1 층 또는 제2 층 중 어느 하나보다 더 많은 구성요소 핵산 분자를 포함할 때, (d)의 풀을 액세스하기 위해 사용된 PCR 쿼리는 제3 층이 제1 층 또는 제2 층보다 더 적은 구성요소 핵산 분자를 포함한 경우보다 액세스된 식별자 핵산 분자의 더 큰 풀을 도출한다.
- [0118] 일부 구현예에서, 제3 층이 제1 층 또는 제2 층 중 어느 하나보다 더 적은 구성요소 핵산 분자를 포함할 때, (d)의 풀을 액세스하기 위해 사용된 PCR 쿼리는 제3 층이 제1 층 또는 제2 층보다 더 많은 구성요소 핵산 분자를 포함한 경우보다 액세스된 식별자 핵산 분자의 더 작은 풀을 도출하고, 액세스된 식별자 핵산 분자의 더 작은 풀은 심볼의 스트링의 심볼로의 더 높은 액세스 분해능에 대응한다.
- [0119] 일부 구현예에서, 제1 층은 가장 높은 우선순위를 갖고, 제2 층은 두 번째로 높은 우선순위를 가지며, 나머지 M-2 층은 제1 말단 분자와 제2 말단 분자 사이에 대응하는 구성요소 핵산 분자를 가진다. 일부 구현예에서, (d)의 풀은 하나의 PCR 반응에서 제1 및 제2 말단 분자에 특정 구성요소 핵산 분자를 갖는 풀의 모든 식별자 핵산 분자에 접근하는 데 사용될 수 있다.
- [0120] 하나의 양태에서, 본 개시는 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 (a) 디지

털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 갖고, 디지털 정보는 벡터의 모음에 의해 나타나는 이미지 데이터를 포함함 - , (b) 다음에 의해 제1 식별자 핵산 분자를 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 분자의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 분자를 선택하는 것, (2) M개의 선택된 구성요소 핵산 분자를 하나의 구획으로 보관하는 것 - , 및 (3) (2)의 M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하여, 제1 및 제2 층으로부터의 구성요소 핵산 분자가 식별자 핵산 분자의 제1 및 제2 말단 분자에 대응하며, 제3 층 내 구성요소 핵산 분자가 식별자 핵산 분자의 제3 분자에 대응하여, 제1 식별자 핵산 분자의 M개의 층의 물리적 순서를 정의하도록, 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치하는 제3 분자를 갖는 제1 식별자 핵산 분자를 형성함 - .

[0121] 일부 구현예에서, 상기 방법은 상기 단계(a), (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계를 포함한다.

[0122] 일부 구현예에서, M개 층 중 적어도 일부는 이미지 데이터의 상이한 특징에 대응한다. 일부 구현예에서, 상이한 특징은 x 좌표, y 좌표, 강도 값 또는 강도 값 범위를 포함한다. 이미지 데이터를 핵산 분자에 저장하면 여기에 설명된 임의의 액세스 방식과 같은 랜덤 액세스 방식을 사용하여 색상 값에 대해 임의의 이웃 픽셀에 대해 쿼리를 허용할 수 있다. 일부 구현예에서, 이미지 데이터를 핵산 분자에 저장하는 것은 이미지 데이터가 이미지 데이터의 원래 해상도의 일부로 디코딩되는 것을 허용한다.

[0123] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가지며, 디지털 정보는 벡터의 모음에 의해 나타내어지는 이미지 데이터를 포함함 - , (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관함으로써 제1 식별자 핵산 분자를 형성하는 단계 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계. 이미지 데이터를 핵산 분자에 저장하면 임의의 접근 방식을 사용하여 임의의 이웃 픽셀에서 색상 값을 쿼리할 수 있다.

[0124] 일부 구현예에서, 이미지 데이터를 핵산 분자에 저장하는 것은 이미지 데이터가 이미지 데이터의 원래 해상도의 비율로 디코딩되는 것을 허용하고, 그 비율로 이미지 데이터를 디코딩하는 것은 관심 프레임을 식별하기 위한 감시 이미지 아카이브 또는 비디오 아카이브에서 특정 시각적 특징을 검색하는 데 사용된다.

[0125] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 - , (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 하고, M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하여 (b)의 식별자 핵산 분자를 형성하는 것은 클릭 화학을 사용하는 것을 포함함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계. 디지털 정보를 저장하는 방법의 단계 (c)는 일반적으로 앞서 언급한 바와 같이 제1 및 제2 말단 분자와 제3 분자를 갖는 분자의 형성을 수행하지 않고 각각 심볼 위치에 대응하는 복수의 식별자 핵산 분자를 형성하는 것을 포함할 수 있다.

- [0126] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 -, 클릭 화학을 사용해 M개의 선택된 구성요소 핵산 분자를 물리적으로 조립함으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 -, (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계, 및 (e) 풀에 수집된 데이터를 삭제하는 단계. 일부 구현예에서, 단계 (c)는 복수의 식별자 핵산 분자를 물리적으로 조립하는 단계 - 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 하고, M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하여 (b)의 식별자 핵산 분자를 형성하는 것은 클릭 화학을 사용하는 것을 포함함 - 를 포함한다.
- [0127] 일부 구현예에서, 방법은 데이터를 선택적으로 삭제하기 위해 서열 특이적 프로브를 사용하여 (d)의 풀로부터 식별자 핵산 분자를 풀다운 선택하는 단계를 더 포함한다. 일부 구현예에서, 선택된 식별자 핵산 분자는 CRISPR 기반 방법을 사용하여 선택적으로 삭제된다. 일부 구현예에서, 방법은 (d)의 풀에서 식별자 핵산 분자를 단독화하여 데이터에 접근할 수 없게 하거나 판독하기 어렵거나 불가능하게 만들어 데이터를 비선택적으로 삭제하는 단계를 더 포함한다. 일부 구현예에서, 방법은 데이터를 비선택적으로 삭제하기 위해 (d)의 풀로부터 식별자 핵산 분자를 분해하기 위해 초음파 처리, 오토클레이빙, 표백제, 염기, 산, 에티뮴 브로마이드 또는 기타 DNA 변형제를 사용한 처리, 방사선 조사, 연소 및 비특이적 뉴클레아제 소화를 사용하는 것을 더 포함한다.
- [0128] 하나의 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 심볼의 스트링을 고정된 길이보다 크지 않은 크기의 하나 이상의 블록으로 나누는 단계, (c) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 -, M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (d) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 -, 및 (e) (d) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.
- [0129] 일부 구현예에서, 상기 (d)의 복수의 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 한다.
- [0130] 일부 구현예에서, 방법은 심볼 스트링, 처리 요구사항, 또는 디지털 정보의 의도된 적용예에 기초하여 각 블록의 크기를 결정하는 단계를 더 포함한다. 일부 구현예에서, 방법은 각 블록의 해시를 계산하는 단계를 더 포함한다. 일부 구현예에서, 방법은 하나 이상의 에러 검출 및 정정을 각각의 블록에 적용하고 하나 이상의 에러 보호 바이트를 계산하는 단계를 더 포함한다. 일부 구현예에서, 방법은 인코딩 또는 디코딩 동안 화학적 조건을 최적화하는 코드워드 세트에 하나 이상의 블록을 매핑하는 단계를 더 포함한다. 일부 구현예에서, 코드워드 세트는 고정된 수의 식별자 핵산 분자가 기록기 시스템의 각 반응 구획에서 조립되고 각 반응 구획 내에서 그리고 반응 구획 전체에 걸쳐 대략 동일한 농도로 조립되도록 고정된 가중치를 가진다.
- [0131] 하나의 양태에서, 본 개시내용은 핵산 분자에 저장되어 있는 디지털 정보에 대한 계산을 수행하는 방법을 제공한다. 중요한 점은 분자 풀에서 실제 디지털 정보를 읽거나 디코딩하지 않고도 해당 계산을 수행할 수 있다는 것이다. 계산에는 AND, OR, NOT 또는 NAND 연산과 같은 부울 논리 게이트의 조합이 포함될 수 있다. 구체적으로, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은, (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 -, M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산

분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계, 및 (e) (d)의 식별자 핵산 분자를 사용하여 심볼의 스트링에 대한 부울 논리 연산, 가령, AND, OR, NOT 또는 NAND을 포함하는 계산을 수행하여 핵산 분자의 새로운 풀을 생성하는 단계. 새로운 핵산 분자 풀은 계산 결과 또는 출력을 나타낼 수 있다.

[0132] 일부 구현예에서, 상기 (c)의 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 분자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 분자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 한다.

[0133] 일부 구현예에서, 심볼 스트링 내의 심볼 중 임의의 것을 얻기 위해 식별자 핵산 분자 중 임의의 것을 디코딩하지 않고 (d)의 식별자 핵산 분자 풀에 대해 계산이 수행된다. 일부 구현예에서, 계산 수행에는 혼성화 및 절단을 포함한 일련의 화학적 작업이 포함된다.

[0134] 일부 구현예에서, (a)의 심볼 스트링은 a로 표시되고 서브-비트스트림 s를 포함하며, (d)의 풀에 있는 복수의 식별자 핵산 분자는 이중 가닥이고 dsA로 표시되며, 방법은 dsB로 표시되고 서브 비트스트림 t를 포함하는 b로 표시되는 또 다른 심볼 스트링을 나타내는 또 다른 복수의 식별자 핵산 분자의 풀을 획득하는 단계를 더 포함하며, 여기서 계산은 dsA 및 dsB에 대한 일련의 단계를 수행함으로써 서브-비트스트림 s 및 t에 대해 수행된다. 일부 구현예에서, dsA 및 dsB에 대한 일련의 단계는 초기화 단계를 수행하는 것을 포함하며, 상기 초기화 단계는 dsA의 이중 가닥 식별자 핵산 분자를 A로 표시된 양성 단일 가닥 형태로 변환하는 단계, dsA의 이중 가닥 식별자 핵산 분자를 A\*로 표시된 음성 단일 가닥 형태로 변환하는 단계 - A\*는A의 역 보체임 - , dsB의 이중 가닥 식별자 핵산 분자를 B의 양성 단일 가닥 형태로 변환하는 단계, dsB의 이중 가닥 식별자 핵산 분자를 B\*으로 표시된 음성 단일 가닥 형태로 변환하는 단계 - B\*은 B의 역 보체임 - , s에 대응하는 dsA의 식별자 핵산 분자로서 dsP를 선택하는 단계, s에 대응하는 A의 식별자 핵산 분자로서 P를 선택하는 단계, t에 대응하는 dsB의 식별자 핵산 분자로서 dsQ를 선택하는 단계, 및 t에 대응하는 B\*의 식별자 핵산 분자로서 Q\*를 선택하는 단계를 포함한다.

[0135] 일부 구현예에서, 계산은 AND 연산이고, dsA 및 dsB에 대한 일련의 단계는 A와 B\*를 결합하여 a와 b 사이에 AND 연산을 수행하고, 상보적인 핵산 분자를 혼성화하고, 핵산 분자의 새로운 풀로서의 핵산 분자로서 완전히 상보적인 이중 가닥을 선택하는 것을 더 포함한다. 일부 구현예에서, 계산은 OR 연산이고, dsA 및 dsB에 대한 일련의 단계는 P와 Q\*를 결합하여 a와 t 사이에 AND 연산을 수행하고, 상보적인 핵산 분자를 혼성화하고, 핵산 분자의 새로운 풀로서의 핵산 분자로서 완전히 상보적인 이중 가닥을 선택하는 것을 더 포함한다.

[0136] 일부 구현예에서, 완전히 상보적인 핵산 분자를 선택하는 것은 크로마토그래피, 겔 전기영동, 단일 가닥 특이적 엔도뉴클레아제, 단일 가닥 특이적 엑소뉴클레아제, 또는 이들의 조합을 사용하는 것을 포함한다.

[0137] 일부 구현예에서, 계산은 OR 연산이고, dsA와 dsB에 대한 일련의 단계는 dsA와 dsB를 결합하여 a와 b 사이의 OR 연산을 수행하여 새로운 핵산 분자 풀을 생성하는 것을 포함한다. 일부 구현예에서, 계산은 OR 연산이고, dsA와 dsB에 대한 일련의 단계는 dsP와 dsQ를 결합하여 s와 t 사이의 OR 연산을 수행하여 새로운 핵산 분자 풀을 생성하는 것을 포함한다.

[0138] 일부 구현예에서, 방법은 A 또는 dsA를 업데이트하여 핵산 분자의 새로운 풀을 포함함으로써 A 또는 dsA가 작업의 출력을 나타낼 수 있게 하는 단계를 더 포함한다.

[0139] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 - , (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계, 및 (c) (b)와 (c)의 식별자 핵산 분자를 개별 빈(bin)으로 파티셔닝하는 단계 - 각 빈은 상이한 심볼 값에 대응함 - .

[0140] 일부 구현예에서, (b)에서 제1 식별자 핵산 분자를 형성하는 것은 다음을 포함한다: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 분자의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 분자를 선택하는 것, (2) M개의 선택된 구성요소 핵산 분자를 하나의 구획으로 보관하는 것 - , (3) (2)의 M개의 선택된 구

성요소 핵산 분자를 물리적으로 조립하여, 제1 및 제2 층으로부터의 구성요소 핵산 분자가 식별자 핵산 분자의 제1 및 제2 말단 분자에 대응하며, 제3 층 내 구성요소 핵산 분자가 식별자 핵산 분자의 제3 분자에 대응하여, 제1 식별자 핵산 분자의 M개의 층의 물리적 순서를 정의하도록, 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치하는 제3 문자를 갖는 제1 식별자 핵산 문자를 형성함 - . 일부 구현예에서, 특정 심볼 값을 갖는 각 심볼의 심볼 위치는 해당 값을 위해 예약된 빈에 기록되며, 빈은 (2)의 구획이다.

[0141] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 - , (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계, 및 (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0142] 일부 구현예에서, 상기 단계 (c)는 복수의 식별자 핵산 분자를 형성하는 단계를 포함하며, 식별자 핵산 분자 각각은 제1 및 제2 말단 분자 및 상기 제1 말단 분자와 상기 제2 말단 분자 사이에 위치한 제3 문자를 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 분자의 제1 말단 분자, 제2 말단 분자, 및 제3 문자가 (b)에서의 제1 식별자 핵산 분자의 타깃 분자와 동일하여, 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 한다.

[0143] 일부 구현예에서, M개의 선택된 구성요소 중 개별 구성요소는 다수의 부분을 포함하며, 각 부분은 핵산 분자를 포함하고, 각 부분은 하나 이상의 화학적 방법에 의해 동일한 식별자에 연결된다. 일부 구현예에서, 상기 다수의 부분 각각은 서로 다른 데이터 저장 작업을 위해 별도의 기능적 목적을 제공한다. 일부 구현예에서, 상기 기능적 목적은 시퀀싱의 용이성 및 핵산 혼성화에 의한 접근의 용이성을 포함한다. 일부 구현예에서, 제1 식별자 핵산 분자를 형성하는 것은 dCas9-데아미나제와 같은 염기 편집기를 적용하여 모 식별자에서 하나 이상의 염기를 프로그램적으로 돌연변이화하는 것을 포함한다.

[0144] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 - , (b) 염기 편집기를 적용하여 모 식별자의 하나 이상의 염기를 프로그램적으로 돌연변이시켜 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각의 식별자 핵산 분자는 각자의 심볼 위치에 대응함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계. 예를 들어, (b)에 적용된 기본 편집기 중 하나는 dCas9-데아미나제이다.

[0145] 하나의 양태에서, 본 개시내용은 하나 이상의 랜덤 프로세스로부터 생성된 디지털 정보를 핵산 분자에 저장하는 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 - , (b) M개의 선택된 구성요소 핵산 분자를 하나의 구획에 보관하고 - M개의 선택된 구성요소 핵산 분자는 M개의 상이한 층으로 분리된 개별 구성요소 핵산 분자의 세트로부터 선택됨 - , M개의 선택된 구성요소 핵산 분자를 물리적으로 모음으로써, 제1 식별자 핵산 분자를 형성하는 단계, (c) 복수의 식별자 핵산 분자를 형성하는 단계 - 각각은 각자의 심볼 위치에 대응함 - , 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0146] 일부 구현에서, 본 개시 내용은 위의 방법 또는 위의 방법 중 임의의 적용 분야를 제공하며, 여기서 적용 분야는 정보의 암호화, 개체의 인증, 또는 무작위화를 포함하는 애플리케이션에서 엔트로피 소스로서의 이의 사용을 포함한다. 일부 구현예에서, 하나 이상의 분리된 식별자 라이브러리의 식별자는 개체 또는 물리적 위치를 고유하게 식별하는 데 사용된다.

[0147] 하나의 양태에서, 본 개시내용은 다수의 랜덤 DNA 중의 파티션에서 디지털 정보를 인코딩하는 방법을 제공한다.

[0148] 하나의 양태에서, 본 개시내용은 가능한 DNA 중의 대규모 조합 풀로부터 DNA 중을 무작위로 샘플링하고 시퀀싱함으로써 랜덤 데이터를 생성하는 방법을 제공한다.

[0149] 하나의 양태에서, 본 개시내용은 가능한 DNA 중의 대규모 조합 풀로부터 DNA 중의 서브세트를 무작위로 샘플링하고 시퀀싱함으로써 랜덤 데이터를 생성하고 저장하는 방법을 제공한다.

- [0150] 일부 구현예에서, DNA 중의 상기 서브세트는 증폭되어 각 종의 다중 복사본을 생성한다. 일부 구현예에서, 에러 체크 및 정정을 위한 핵산 분자가 DNA 중의 상기 서브세트에 추가되어 강력한 향후 관독이 가능해진다. 일부 구현예에서, 상기 DNA 중의 서브세트는 고유한 분자로 바코드화되고 바코드화된 DNA 중 서브세트의 풀에 결합된다. 일부 구현예에서, 상기 바코드화된 DNA 중 서브세트 풀 내의 DNA 중의 특정 서브세트는 PCR 또는 핵산 포획을 위한 입력 핵산 프로브로 접근 가능하다.
- [0151] 하나의 양태에서, 본 개시는 다음을 포함하는 시스템으로 아티팩트를 보안하고 인증하는 방법을 제공한다: (1) 정의된 세트의 DNA 중 서브세트로 구성된 DNA 키, 및 (2) 키를 받아들이고 일치하는 키를 검색하여 해당 아티팩트를 로컬로 잠금 해제하거나 해시된 토큰을 반환하여 다른 곳에서 아티팩트에 액세스하는 DNA 관독기. 일부 구현예에서, 방법은 생물학적 적용을 위해 DNA 단편을 조합적으로 조립하는 단계를 더 포함한다.
- [0152] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 분자에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 다음에 의해 제1 식별자 핵산 분자를 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 분자의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 분자를 선택하는 것, (2) M개의 선택된 구성요소 핵산 분자를 하나의 구획으로 보관하는 것 -, (3) (2)의 M개의 선택된 구성요소 핵산 분자를 물리적으로 조립하여 특정된 구성요소를 포함하는 제1 식별자 핵산 분자를 형성하는 것 - 특정된 구성요소는 적어도 하나의 표적 분자를 포함하여 특정된 구성요소를 함유하는 식별자의 액세스를 가능하게 함 -, (c) 각각 특정된 구성요소를 갖는 복수의 추가 식별자 핵산 분자를 물리적으로 조립하는 단계 - 특정된 구성요소는 (b)의 제1 식별자 핵산 분자의 적어도 하나의 표적 분자를 포함함으로써, 프로브가 심볼의 스트링 내 연속 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 분자를 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 분자를 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.
- [0153] 하나의 양태에서, 본 개시내용은 정보를 핵산 서열로 인코딩하는 방법을 제공한다. 정보를 핵산 서열로 인코딩하기 위한 방법은 (a) 정보를 심볼의 스트링으로 번역하는 단계, (b) 심볼의 스트링을 복수의 식별자로 매핑하는 단계, 및 (c) 복수의 식별자의 서브세트를 적어도 포함하는 식별자 라이브러리를 구축하는 단계를 포함할 수 있다. 복수의 식별자 중 개별 식별자는 하나 이상의 구성요소를 포함할 수 있다. 하나 이상의 구성요소의 개별 구성요소는 핵산 서열을 포함할 수 있다. 심볼의 스트링 내 각 위치에 있는 각 심볼은 고유한 식별자에 대응할 수 있다. 개별 식별자는 심볼의 스트링 내 개별 위치에 있는 개별 심볼에 대응할 수 있다. 또한, 심볼의 스트링 내 각 위치에서의 하나의 심볼이 식별자의 부재에 대응할 수 있다. 예를 들어, '0'과 '1'의 이진 심볼의 스트링 (가령 비트)에서 '0'의 각 발생은 식별자가 없음에 대응할 수 있다.
- [0154] 또 다른 양태에서, 본 개시내용은 핵산 기반 컴퓨터 데이터 저장을 위한 방법을 제공한다. 핵산 기반 컴퓨터 데이터 저장을 위한 방법은 (a) 컴퓨터 데이터를 수신하는 단계, (b) 컴퓨터 데이터를 인코딩하는 핵산 서열을 포함하는 핵산 분자를 합성하는 단계, 및 (c) 핵산 서열을 갖는 핵산 분자를 저장하는 단계를 포함할 수 있다. 컴퓨터 데이터는 각각의 핵산 분자의 서열이 아니라 적어도 합성된 핵산 분자의 서브세트에 인코딩될 수 있다.
- [0155] 또 다른 양태에서, 본 개시내용은 핵산 서열에 정보를 기록하고 저장하기 위한 방법을 제공한다. 방법은 (a) 정보를 나타내는 가상 식별자 라이브러리를 수신 또는 인코딩하는 단계, (b) 식별자 라이브러리를 물리적으로 구성하는 단계, 및 (c) 하나 이상의 별도 위치에 식별자 라이브러리의 하나 이상의 물리적 사본을 저장하는 단계를 포함할 수 있다. 식별자 라이브러리의 개별 식별자는 하나 이상의 구성요소를 포함할 수 있다. 하나 이상의 구성요소의 개별 구성요소는 핵산 서열을 포함할 수 있다.
- [0156] 또 다른 양태에서, 본 개시내용은 핵산 기반 컴퓨터 데이터 저장을 위한 방법을 제공한다. 핵산 기반 컴퓨터 데이터 저장을 위한 방법은 (a) 컴퓨터 데이터를 수신하는 단계, (b) 컴퓨터 데이터를 인코딩하는 적어도 하나의 핵산 서열을 포함하는 핵산 분자를 합성하는 단계, 및 (c) 적어도 하나의 핵산 서열을 포함하는 핵산 분자를 저장하는 단계를 포함할 수 있다. 핵산 분자를 합성하는 것은 염기별 핵산 합성이 없을 수 있다.
- [0157] 또 다른 양태에서, 본 개시내용은 핵산 서열에 정보를 기록하고 저장하기 위한 방법을 제공한다. 정보를 핵산 서열에 기록하고 저장하는 방법은 (a) 정보를 나타내는 가상 식별자 라이브러리를 수신하거나 인코딩하는 단계, (b) 식별자 라이브러리를 물리적으로 구성하는 단계, 및 (c) 식별자 라이브러리의 하나 이상의 물리적 복사본을 하나 이상의 개별 위치에 저장하는 단계를 포함할 수 있다. 식별자 라이브러리의 개별 식별자는 하나 이상의 구성요소를 포함할 수 있다. 하나 이상의 구성요소의 개별 구성요소는 핵산 서열을 포함할 수 있다.

[0158] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 다음에 의해 제1 식별자 핵산 서열을 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 서열의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 서열을 선택하는 것, (2) M개의 선택된 구성요소 핵산 서열을 하나의 구획으로 보관하는 것 -, (3) (2)의 M개의 선택된 구성요소 핵산 서열을 물리적으로 조립하여, 제1 및 제2 층으로부터의 구성요소 핵산 서열이 식별자 핵산 서열의 제1 및 제2 말단 서열에 대응하며, 제3 층 내 구성요소 핵산 서열이 식별자 핵산 서열의 제3 서열에 대응하여, 제1 식별자 핵산 서열의 M개의 층의 물리적 순서를 정의하도록, 제1 및 제2 말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치하는 제3 서열을 갖는 제1 식별자 핵산 서열을 형성함 -, (c) 복수의 추가 식별자 핵산 서열을 형성하는 단계 - 추가 식별자 핵산 서열 각각은 (1) 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, (2) 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 프로브가 심볼의 스트링 내 연속 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0159] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은, (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가지며, 디지털 정보는 벡터의 모음에 의해 나타내어지는 이미지 데이터를 포함함 -, (b) M개의 선택된 구성요소 핵산 서열을 하나의 구획에 예치함으로써 제1 식별자 핵산 서열을 형성하는 단계 - M개의 선택된 구성요소 핵산 서열은 M개의 상이한 층으로 분리된 개별 구성요소 핵산 서열의 세트로부터 선택됨 -, (c) 복수의 식별자 핵산 서열을 형성하는 단계 - 추가 식별자 핵산 서열 각각은 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 단일 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계 - 이미지 데이터를 핵산 서열로 저장함으로써 랜덤 액세스 스킴을 사용해 픽셀의 임의의 이웃이 색상 값을 질의 받을 수 있음 - 를 포함하는, 방법.

[0160] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은, (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) M개의 선택된 구성요소 핵산 서열을 하나의 구획에 보관함으로써 제1 식별자 핵산 서열을 형성하는 단계 - M개의 선택된 구성요소 핵산 서열은 M개의 상이한 층으로 분리된 개별 구성요소 핵산 서열의 세트로부터 선택됨 -, (c) 복수의 식별자 핵산 서열을 물리적으로 조립하는 단계 - 추가 식별자 핵산 서열 각각은 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 단일 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0161] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 심볼의 스트링을 고정된 길이보다 크지 않은 크기의 하나 이상의 블록으로 나누는 단계, (c) M개의 선택된 구성요소 핵산 서열을 하나의 구획에 보관함으로써 제1 식별자 핵산 서열을 형성하는 단계 - M개의 선택된 구성요소 핵산 서열은 M개의 상이한 층으로 분리된 개별 구성요소 핵산 서열의 세트로부터 선택됨 -, (d) 복수의 식별자 핵산 서열을 물리적으로 조립하는 단계 - 추가 식별자 핵산 서열 각각은 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 단일 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (e) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0162] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은 다

음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) M개의 선택된 구성요소 핵산 서열을 하나의 구획에 보관함으로써 제1 식별자 핵산 서열을 형성하는 단계 - M개의 선택된 구성요소 핵산 서열은 M개의 상이한 층으로 분리된 개별 구성요소 핵산 서열의 세트로부터 선택됨 -, (c) 복수의 식별자 핵산 서열을 물리적으로 조립하는 단계 - 추가 식별자 핵산 서열 각각은 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 단일 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계, 및 (e) (d)의 식별자 핵산 서열을 사용하여 심볼의 스트링에 대한 부울 논리 연산, 가령, AND, OR, NOT 또는 NAND을 포함하는 계산을 수행하여 핵산 분자의 새로운 풀을 생성하는 단계.

[0163] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 다음에 의해 제1 식별자 핵산 서열을 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 서열의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 서열을 선택하는 것, (2) M개의 선택된 구성요소 핵산 서열을 하나의 구획으로 보관하는 것 -, (c) 복수의 식별자 핵산 서열을 물리적으로 조립하는 단계 - 추가 식별자 핵산 서열 각각은 제1 및 제2말단 서열 및 상기 제1 말단 서열과 상기 제2 말단 서열 사이에 위치한 제3 서열을 가지며, 각자의 심볼 위치에 대응하며, 적어도 하나의 추가 식별자 핵산 서열의 제1 말단 서열, 제2 말단 서열, 및 제3 서열이 (b)에서의 제1 식별자 핵산 서열의 타깃 서열과 동일하여, 단일 프로브가 심볼의 스트링 내 관련 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0164] 또 다른 양태에서, 본 개시내용은 디지털 정보를 핵산 서열에 저장하기 위한 방법을 제공하며, 상기 방법은 다음을 포함한다: (a) 디지털 정보를 심볼의 스트링으로서 수신하는 단계 - 심볼의 스트링 내 각 심볼은 심볼 값 및 심볼의 스트링 내의 심볼 위치를 가짐 -, (b) 다음에 의해 제1 식별자 핵산 서열을 형성하는 단계: (1) M개의 상이한 층으로 분리되는 개별 구성요소 핵산 서열의 세트로부터 M개의 층 각각으로부터의 하나씩의 구성요소 핵산 서열을 선택하는 것, (2) M개의 선택된 구성요소 핵산 서열을 하나의 구획으로 보관하는 것 -, (3) (2)의 M개의 선택된 구성요소 핵산 서열을 물리적으로 조립하여 특정된 구성요소를 포함하는 제1 식별자 핵산 서열을 형성하는 것 - 특정된 구성요소는 적어도 하나의 표적 서열을 포함하여 특정된 구성요소를 함유하는 식별자의 액세스를 가능하게 함 -, (c) 각각 특정된 구성요소를 갖는 복수의 추가 식별자 핵산 서열을 물리적으로 조립하는 단계 - 특정된 구성요소는 (b)의 제1 식별자 핵산 서열의 적어도 하나의 표적 서열을 포함함으로써, 프로브가 심볼의 스트링 내 연속 심볼 위치를 갖는 각자의 심볼에 대응하는 적어도 두 개의 식별자 핵산 서열을 선택할 수 있게 함 -, 및 (d) (b) 및 (c)의 식별자 핵산 서열을 분말, 액체, 또는 고체 형태를 갖는 풀에 수집하는 단계.

[0165] 도 57은 정보를 핵산 서열로 인코딩하고, 정보를 핵산 서열에 기록하고, 핵산 서열에 기록된 정보를 판독하고, 판독된 정보를 디코딩하는 개요 프로세스를 도시한다. 디지털 정보 또는 데이터는 심볼의 하나 이상의 스트링으로 변환될 수 있다. 예시에서, 심볼은 비트이고 각 비트는 '0' 또는 '1'의 값을 가질 수 있다. 각 심볼은 해당 심볼을 나타내는 객체(가령, 식별자)에 매핑되거나 인코딩될 수 있다. 각 심볼은 개별 식별자로 나타내어질 수 있다. 개별 식별자는 구성요소로 구성된 핵산 분자일 수 있다. 구성요소는 핵산 서열일 수 있다. 디지털 정보는 정보에 대응하는 식별자 라이브러리를 생성함으로써 핵산 서열에 기록될 수 있다. 식별자 라이브러리는 디지털 정보의 각 심볼에 대응하는 식별자를 물리적으로 구성함으로써 물리적으로 생성될 수 있다. 디지털 정보의 전부 또는 일부가 한 번에 액세스될 수 있다. 예를 들어, 식별자의 서브세트가 식별자 라이브러리로부터 액세스된다. 식별자의 서브세트는 식별자를 시퀀싱하고 식별함으로써 판독될 수 있다. 식별된 식별자는 해당 심볼과 연관되어 디지털 데이터를 디코딩할 수 있다.

[0166] 도 57의 접근 방식을 사용하여 정보를 인코딩하고 판독하기 위한 방법은 예를 들어, 비트 스트림을 수신하고 식별자 순위 또는 핵산 인덱스를 사용하여 비트 스트림의 각 1비트(비트 값이 '1'인 비트)를 개별 핵산 식별자에 매핑하는 것을 포함할 수 있다. 비트 값 1에 대응하는 식별자의 복사본을 포함하는(비트 값 0에 대한 식별자는 제외) 핵산 샘플 풀 또는 식별자 라이브러리를 구축한다. 샘플을 판독하는 것은 분자 생물학적 방법(가령, 시퀀싱, 혼성화, PCR 등)을 사용하고, 어떤 식별자가 식별자 라이브러리에 표현되는지 결정하고, 해당 식별자에 대

응하는 비트에 '1'의 비트 값을 할당하고, 그 밖의 다른 곳에 '0'의 비트 값을 할당함으로써(각 식별자가 대응하는 원본 비트스트림의 비트를 식별하기 위해 식별자 순위를 다시 참조함) 정보를 본래의 인코딩된 비트 스트림으로 디코딩하는 것을 포함할 수 있다.

[0167] N개의 개별 비트의 스트링을 인코딩하면 동일한 수의 고유한 핵산 서열을 가능한 식별자로 사용할 수 있다. 정보 인코딩에 대한 이러한 접근 방식은 저장할 각각의 새로운 정보 항목(N 비트의 스트링)에 대한 식별자(가령, 핵산 분자)의 신규 합성을 사용할 수 있다. 다른 경우에서, 저장할 각각의 새로운 정보에 대한 식별자(N개 이하)를 새로 합성하는 비용은 일회성 신규 합성 및 가능한 모든 식별자의 후속 유지 관리를 통해 감소되어, 새로운 정보를 인코딩하는 것이 사전-합성된(또는 사전-제조된) 식별자를 기계적으로 선택 및 혼합하여 식별자 라이브러리를 형성하는 것을 포함할 수 있다. 다른 경우, (1) 저장할 각각의 새로운 정보에 대한 최대 N개의 식별자의 신규 합성 또는 (2) 저장할 각각의 새로운 정보에 대한 N개의 가능한 식별자로부터의 유지 및 선택, 또는 임의의 조합의 비용 모두가, 다수(N개 미만, 및 일부 경우 N개 훨씬 미만)의 핵산 서열을 합성하고 유지한 다음 효소 작용을 통해 이들 서열을 수정하여 저장할 각각의 새로운 정보에 대한 최대의 N개의 식별자를 생성함으로써, 감소될 수 있다.

[0168] 식별자는 판독, 기록, 액세스, 복사 및 삭제 작업의 용이성을 위해 합리적으로 설계되고 선택될 수 있다. 식별자는 기록 오류, 돌연변이, 성능 저하 및 읽기 오류를 최소화하도록 설계되고 선택될 수 있다. 합성 핵산 라이브러리(가령, 식별자 라이브러리)를 포함하는 DNA 서열의 합리적인 설계에 대해서는 화학적 방법 섹션 H를 참조할 수 있다.

[0169] 도 58a 및 58b는 객체 또는 식별자(가령, 핵산 분자)에 디지털 데이터를 인코딩하는 "데이터 옛 어드레스(data at address)"로 지칭되는 예시적 방법을 개략적으로 설명한다. 도 58a는 바이트-값을 특정하는 단일 구성요소와 식별자 순위를 특정하는 단일 구성요소를 연결하거나 조립함으로써 개별 식별자가 구성되는 식별자 라이브러리로 비트 스트림을 인코딩하는 것을 도시한다. 일반적으로, 데이터 옛 어드레스 방법은 다음의 두 개의 객체를 포함함으로써 정보를 모듈식으로 인코딩하는 식별자를 사용한다: 하나의 객체, 즉, 바이트-값을 식별하는 "바이트-값 객체"(또는 "데이터 객체") 및 하나의 객체, 즉, 식별자 순위(또는 원본 비트-스트림 내 바이트의 상대 위치)를 식별하는 "순위 객체"(또는 "주소 객체"). 도 58b는 데이터 옛 어드레스 방법의 예시를 도시하며, 여기서, 각각의 순위 객체가 구성요소의 세트로부터 조합적으로 구성될 수 있으며 각각의 바이트-값 객체가 구성요소의 세트로부터 조합적으로 구성될 수 있다. 순위 및 바이트-값 객체의 이러한 조합 구성에 의해, 객체가 단일 구성요소만으로부터 만들어진 경우(도 58a)보다 더 많은 정보가 식별자에 기록될 수 있다.

[0170] 도 59a 및 59b는 객체 또는 식별자(예를 들어, 핵산 서열)의 디지털 정보를 인코딩하는 또 다른 예시적인 방법을 개략적으로 도시한다. 도 59a는 비트 스트림을 식별자 라이브러리로 인코딩하는 것을 도시하며, 여기서 식별자는 식별자 순위를 특정하는 단일 구성요소로부터 구성된다. 특정 순위(또는 주소)에 식별자가 있으면 비트 값 '1'을 지정하고 특정 순위(또는 주소)에 식별자가 없으면 비트 값 '0'을 지정한다. 이러한 유형의 인코딩은 순위(원본 비트 스트림 내 비트의 상대 위치)만 인코딩하는 식별자를 사용할 수 있으며 식별자 라이브러리에서 해당 식별자의 존재 여부를 사용하여 '1' 또는 '0'의 비트 값을 각각 인코딩할 수 있다. 정보를 판독하고 디코딩하는 것은 식별자 라이브러리에 존재하는 식별자를 식별하는 것, 비트 값 '1'을 대응하는 순위에 할당하는 것, 비트 값 '0'을 그 외 다른 곳에 할당하는 것 등을 포함할 수 있다. 도 59b는 각각의 가능한 조합 구성이 순위를 특정하도록 각 식별자가 구성요소의 세트로부터 조합적으로 구성될 수 있는 예시적인 인코딩 방법을 도시한다. 이러한 조합 구성은 식별자가 단일 구성요소만으로 만들어진 경우(가령, 도 59a)보다 더 많은 정보가 식별자에 기록될 수 있도록 한다. 예를 들어, 구성요소 세트는 5개의 개별 구성요소를 포함할 수 있다. 5개의 개별 구성요소는 조립되어 10개의 개별 식별자를 생성할 수 있으며, 각각은 5개의 구성요소 중 2개를 포함한다. 10개의 개별 식별자는 각각 비트스트림 내 비트의 위치에 대응하는 순위(또는 주소)를 가질 수 있다. 식별자 라이브러리는 비트-값 '1'의 위치에 대응하는 10개의 가능한 식별자의 서브세트를 포함하고, 길이가 10인 비트 스트림 내 비트-값 '0'의 위치에 대응하는 10개의 가능한 식별자의 서브세트를 배제할 수 있다.

[0171] 도 60은 가능한 식별자의 조합 공간(C, x축)과 도 59a 및 도 59b에 도시된 인코딩 방법을 사용하여 비트 단위의 주어진 원래 크기의 정보(D, 등고선)를 저장하도록 물리적으로 구성될 식별자의 평균 개수(k, y축) 사이의 관계를 로그 공간으로 나타낸 등고선 플롯을 보여준다. 이 플롯은 크기 D의 원본 정보가 C 비트의 스트링(C는 D보다 클 수 있음)으로 재코딩되며, 여기서 비트 수 k는 '1'의 비트 값을 가짐을 가정한다. 또한, 플롯은 정보-핵산 인코딩이 재코딩된 비트 스트림에 대해 수행되며 비트-값이 '1'인 위치에 대한 식별자가 구성되고 비트-값이 '0'인 위치에 대한 식별자가 구성되지 않음을 가정한다. 가정에 따르면, 가능한 식별자의 조합 공간은 재코딩된 비트 스트림의 모든 위치를 식별하기 위한 크기 C를 가지며, 크기 D의 비트 스트림을 인코딩하는 데 사용되는

식별자의 개수는  $D = \log_2(C\text{choose}k)$ 이도록 정해지며, 여기서,  $C\text{choose}k$  은 C개의 가능성 중에서 k개의 정렬되지 않은 결과를 선택하는 방법의 수에 대한 수학 공식일 수 있다. 따라서 가능한 식별자의 조합 공간이 주어진 정보의 크기(비트 단위) 이상으로 증가함에 따라, 감소하는 수의 물리적으로 구성된 식별자가 주어진 정보를 저장하는 데 사용될 수 있다.

[0172] 도 61는 정보를 핵산 서열에 기록하는 개략적 방법을 보여준다. 정보를 기록하기 전에 정보는 심볼의 스트링으로 변환되고 복수의 식별자로 인코딩될 수 있다. 정보를 기록하는 것은 가능한 식별자를 생성하기 위한 반응을 설정하는 것을 포함할 수 있다. 입력을 한 구획에 보관함으로써 반응이 설정될 수 있다. 입력은 핵산, 구성요소, 주형, 효소 또는 화학적 시약을 포함할 수 있다. 구획은 웰, 튜브, 표면 상의 위치, 미세유체 장치 내 챔버, 또는 에멀전 내의 액적일 수 있다. 다중 구획에서 다수의 반응이 설정될 수 있다. 프로그래밍된 온도 배양 또는 순환을 통해 반응이 진행되어 식별자를 생성할 수 있다. 반응은 선택적으로 또는 편재적으로 제거(가령, 삭제)될 수 있다. 반응은 하나의 풀에서 식별자를 수집하기 위해 선택적으로 또는 편재적으로 중단되고, 통합되고, 정제될 수도 있다. 다수의 식별자 라이브러리로부터의 식별자가 동일한 풀에 수집될 수 있다. 개별 식별자는 자신이 속한 식별자 라이브러리를 식별하는 바코드나 태그를 포함할 수 있다. 대안으로 또는 추가로, 바코드는 인코딩된 정보에 대한 메타데이터를 포함할 수 있다. 보충 핵산 또는 식별자가 식별자 라이브러리와 함께 식별자 풀에 포함될 수도 있다. 보충 핵산 또는 식별자는 인코딩된 정보에 대한 메타데이터를 포함하거나 인코딩된 정보를 난독화하거나 숨기는 역할을 할 수 있다.

[0173] 식별자 순위(가령, 핵산 인덱스)는 식별자의 순서를 결정하기 위한 방법 또는 키를 포함할 수 있다. 상기 방법은 모든 식별자 및 이들의 대응하는 순위가 있는 록업 테이블을 포함할 수 있다. 방법은 또한 식별자를 구성하는 모든 구성요소의 순위를 갖는 검색 테이블 및 이러한 구성요소의 조합을 포함하는 임의의 식별자의 순서를 결정하기 위한 기능을 포함할 수 있다. 이러한 방법은 사전순 정렬이라고 할 수 있으며 사전의 단어를 알파벳순으로 정렬하는 방식과 유사할 수 있다. 데이터 옛 어드레스 인코딩 방법에서 식별자 순위(식별자의 순위 객체에 의해 인코딩됨)는 비트 스트림 내에서의 바이트(식별자의 바이트 값 객체에 의해 인코딩됨)의 위치를 결정하는데 사용될 수 있다. 다른 방법으로, 현재 식별자에 대한 식별자 순위(전체 식별자 자체에 의해 인코딩됨)를 사용하여 비트스트림 내에서 비트값 '1'의 위치를 결정할 수 있다.

[0174] 키는 샘플 내 식별자(가령, 핵산 분자)의 고유한 서브세트에 개별 바이트를 할당할 수 있다. 예를 들어, 간단한 형태에서, 키는 비트의 위치를 특정하는 고유한 핵산 서열에 바이트의 각 비트를 할당할 수 있으며, 그런 다음 샘플 내 해당 핵산 서열의 존재 여부에 따라 각각 1 또는 0의 비트-값을 특정할 수 있다. 핵산 샘플로부터의 인코딩된 정보를 판독하는 것은 시퀀싱, 혼성화 또는 PCR을 포함하는 다양한 분자 생물학 기술을 포함할 수 있다. 일부 실시예에서, 인코딩된 데이터세트를 판독하는 것은 데이터세트의 일부를 재구성하거나 각 핵산 샘플로부터 전체 인코딩된 데이터세트를 재구성하는 것을 포함할 수 있다. 서열이 판독될 수 있는 경우, 고유한 핵산 서열의 존재 또는 부재와 함께 핵산 인덱스가 사용될 수 있으며 핵산 샘플은 비트 스트림(가령, 각 비트 스트링, 바이트, 바이트 또는 바이트 스트링)으로 디코딩될 수 있다.

[0175] 식별자는 구성요소 핵산 서열을 조합적으로 조합함으로써 구성될 수 있다. 예를 들어, 정보는 정의된 분자 그룹(가령, 조합 공간)으로부터 핵산 분자(가령, 식별자)의 세트를 취함으로써 인코딩될 수 있다. 정의된 분자 그룹의 각각의 가능한 식별자는 층으로 분할될 수 있는 구성요소의 사전 제작된 세트로부터의 핵산 서열(가령, 구성요소)의 조합체일 수 있다. 각 개별 식별자는 모든 층으로부터의 하나의 구성요소를 고정된 순서로 연결함으로써 구성될 수 있다. 예를 들어, M개의 층이 있고 각 층이 n개의 구성요소를 가질 수 있는 경우, 최대  $C = n^M$ 개의 고유 식별자가 구성될 수 있으며 최대  $2^C$ 개의 상이한 정보 또는 C 비트가 인코딩되고 저장될 수 있다. 예를 들어 메가비트 정보를 저장하려면  $1 \times 10^6$ 개의 개별 식별자 또는  $C = 1 \times 10^6$  크기의 조합 공간을 사용할 수 있다. 이 예의 식별자는 다양한 방식으로 구성된 다양한 구성요소로부터 조립될 수 있다. 조립체는 각각  $n = 1 \times 10^3$ 개의 구성요소를 포함하는  $M = 2$ 개의 사전 제작된 층으로부터 만들어질 수 있다. 대안으로, 조립체는 각각  $n = 1 \times 10^2$ 개의 구성요소를 포함하는  $M = 3$ 개의 층으로부터 만들어질 수 있다. 일부 구현예에서, 조립체는  $M=2$ ,  $M=3$ ,  $M=4$ ,  $M=5$  또는 그 이상의 층으로 만들어질 수 있다. 이 예에서 알 수 있듯이, 더 많은 수의 층을 사용하여 동일한 양의 정보를 인코딩하면 전체 구성요소의 수가 더 작아질 수 있다. 전체 구성요소의 수를 적게 사용하는 것이 기록 비용 측면에서 유리할 수 있다.

[0176] 하나의 예에서, 각각 x 및 y 구성요소(가령, 핵산 서열)를 각각 갖는 고유한 핵산 서열 또는 층, X 및 Y의 두 세트에 시작할 수 있다. X로부터의 각 핵산 서열은 Y로부터의 각 핵산 서열로 조립될 수 있다. 두 개의 세트에

유지되는 핵산 서열의 총 수는  $x$ 와  $y$ 의 합(sum)일 수 있지만, 생성될 수 있는 핵산 분자, 따라서 가능한 식별자의 총 수가  $x$ 와  $y$ 의 곱(product)일 수 있다.  $X$ 로부터의 서열이 임의의 순서로  $Y$ 의 서열에 조립될 수 있는 경우 훨씬 더 많은 핵산 서열(가령, 식별자)이 생성될 수 있다. 예를 들어, 생성된 핵산 서열(가령, 식별자)의 수는 조립 순서가 프로그래밍 가능한 경우  $x$ 와  $y$ 의 곱의 두 배가 될 수 있다. 생성될 수 있는 모든 가능한 핵산 서열 세트는  $XY$ 로 지칭될 수 있다.  $XY$ 의 고유한 핵산 서열의 조립된 단위 순서는 개별 5' 및 3' 말단을 가진 핵산을 사용하여 제어될 수 있으며, 제한 분해, 결찰, 중합효소 연쇄 반응(PCR) 및 시퀀싱은 서열의 개별 5' 및 3' 말단에 대해 발생할 수 있다. 이러한 접근 방식은 조립 산물의 조합 및 순서로 정보를 인코딩함으로써  $N$ 개의 개별 비트를 인코딩하는 데 사용되는 핵산 서열(가령, 구성요소)의 총 수를 줄일 수 있다. 예를 들어, 100 비트의 정보를 인코딩하기 위해, 10개의 개별 핵산 분자(가령, 구성요소)의 두 개의 층을 고정된 순서로 조립하여  $10 \times 10$  또는 100개의 개별 핵산 분자(가령, 식별자)를 생성할 수 있거나, 5개의 개별 핵산 분자(가령, 구성요소)의 하나의 층과 10개의 개별 핵산 분자(가령, 구성요소)의 또 다른 층이 임의의 순서로 조립되어 100개의 개별 핵산 분자(가령, 식별자)를 생성할 수 있다.

[0177] 각 층 내의 핵산 서열(예를 들어, 구성요소)은 층안에 고유한(또는 개별) 서열 또는 바코드, 한쪽 말단에 공통 혼성화 영역, 또 다른 다른 말단에 또 다른 공통 혼성화 영역을 포함할 수 있다. 바코드는 층 내의 모든 서열을 고유하게 식별하기에 충분한 수의 뉴클레오티드를 포함할 수 있다. 예를 들어, 바코드 내의 각 염기 위치에 대해 일반적으로 4개의 가능한 뉴클레오티드가 있다. 따라서 3개 염기 바코드는  $4^3 = 64$ 개의 핵산 서열을 고유하게 식별할 수 있다. 바코드는 랜덤하게 생성되도록 설계될 수 있다. 대안으로, 바코드는 식별자 구성 화학 또는 시퀀싱에 복잡함을 야기할 수 있는 서열을 방지하도록 설계될 수 있다. 추가적으로, 바코드는 각각이 다른 바코드로부터 최소 해밍 거리를 가질 수 있도록 설계될 수 있으며, 이로써 염기 분해능 돌연변이 또는 판독 오류가 바코드의 적절한 식별을 방해할 가능성을 줄일 수 있다. DNA 서열의 합리적인 설계에 대해서는 화학적 방법 섹션 H를 참조할 수 있다.

[0178] 핵산 서열(예를 들어, 구성요소)의 하나의 말단에 있는 혼성화 영역은 각 층마다 상이할 수 있지만, 혼성화 영역은 층 내의 각 구성원에 대해 동일할 수 있다. 인접한 층은 서로 상호 작용할 수 있도록 구성요소에 상보적인 혼성화 영역이 있는 층이다. 예를 들어, 층  $X$ 로부터의 모든 구성요소는 상보적인 혼성화 영역을 가질 수 있으므로 층  $Y$ 로부터의 임의의 구성요소에 부착될 수 있다. 반대편 말단의 혼성화 영역은 제1 말단의 혼성화 영역과 동일한 목적을 수행할 수 있다. 예를 들어, 층  $Y$ 로부터의 임의의 구성요소는 한쪽 말단 상의 층  $X$ 의 임의의 구성요소에 부착되고 반대쪽 말단 상의 층  $Z$ 의 임의의 구성요소에 부착될 수 있다.

[0179] 도 62a 및 62b는 각각의 층으로부터 개별 구성요소(가령, 핵산 서열)를 고정된 순서로 조합적으로 조립함으로써 식별자(가령, 핵산 분자)를 구축하기 위한 "곱 방식(product scheme)"이라고 하는 예시적인 방법을 예시한다. 도 62a는 곱 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 각 층으로부터의 단일 구성요소를 고정된 순서로 조합함으로써 식별자가 구성될 수 있다. 각각  $N$ 개의 구성요소를 포함하는  $M$ 개의 층에 대해  $N^M$ 개의 가능한 식별자가 있다. 도 62b는 곱 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 예를 들어, 조합 공간은 각각 3개의 개별 구성요소를 포함하는 3개의 층으로부터 생성될 수 있다. 구성요소는 각 층으로부터의 하나씩의 구성요소가 고정된 순서로 결합될 수 있도록 결합될 수 있다. 이 조립 방법의 전체 조합 공간은 27개의 가능한 식별자로 구성될 수 있다.

[0180] 도 63-66은 곱 방식(도 62 참조)을 구현하기 위한 화학적 방법을 예시한다. 도 63-66에 도시된 방법은, 둘 이상의 개별 구성요소를 고정된 방식으로 조립하기 위한 임의의 다른 방법과 함께, 사용되어 임의의 하나 이상의 식별자를 식별자 라이브러리를 생성할 수 있다. 식별자는 본 명세서에 개시된 방법 또는 시스템 동안 임의의 시점에서, 도 63-66에서 기재된 구현 방법 중 임의의 것을 사용해 구성될 수 있다. 어떤 경우에는, 디지털 정보가 인코딩되거나 기록되기 전에 가능한 식별자의 조합 공간의 전체 또는 일부가 구성될 수 있으며, 그런 다음 기록 프로세스는 기존 세트로부터의 (정보를 인코딩하는) 식별자를 기계적으로 선택하고 풀링하는 것을 포함할 수 있다. 다른 경우에, 식별자는 데이터 인코딩 또는 기록 프로세스의 하나 이상의 단계가 발생한 후에(즉, 정보가 기록되는 동안) 구성될 수 있다.

[0181] 효소 반응은 상이한 층 또는 세트로부터의 구성요소를 조립하는 데 사용될 수 있다. 각 층의 구성요소(가령, 핵산 서열)가 인접한 층의 구성요소에 대한 특정 혼성화 또는 부착 영역을 갖기 때문에 조립은 원 팻 반응(one pot reaction)으로 발생할 수 있다. 예를 들어, 층  $X$ 로부터의 핵산 서열(가령, 구성요소)  $X1$ , 층  $Y$ 로부터의 핵산 서열  $Y1$ , 및 층  $Z$ 로부터의 핵산 서열  $Z1$ 은 조립된 핵산 분자(예를 들어, 식별자)  $X1Y1Z1$ 을 형성할 수 있다. 추가로, 다수의 핵산 분자(예를 들어, 식별자)는 각 층으로부터의 다수의 핵산 서열을 포함함으로써 하나의 반

응으로 조립될 수 있다. 예를 들어, 이전 예시의 원 포트 반응에 Y1과 Y2를 모두 포함하면 두 개의 조립된 생성물(가령, 식별자) X1Y1Z1 및 X1Y2Z1이 생성될 수 있다. 이 반응 다중화는 물리적으로 구성된 복수의 식별자에 대한 기록 시간을 단축하는데 사용될 수 있다. 조립 효율성과 관련된 DNA 서열의 합리적인 설계에 대한 자세한 내용은 화학적 방법 섹션 H를 참조할 수 있다. 핵산 서열의 조립은 약 1일, 12시간, 10시간, 9시간, 8시간, 7시간, 6시간, 5시간, 4시간, 3시간, 2시간 또는 1시간 이하의 기간에 수행될 수 있다. 인코딩된 데이터의 정확도는 적어도 약 90%, 95%, 96%, 97%, 98%, 99% 이상일 수 있다.

[0182] 식별자는 도 63에 예시된 바와 같이 OEPCR(overlap Extension Polymerase Chain Reaction)을 사용하는 곱 방식에 따라 구성될 수 있다. 각 층의 각 구성요소는 인접 층으로부터의 구성요소의 서열 말단 상에 공통 혼성화 영역과 상동성 및/또는 상보적일 수 있는 서열 말단 상의 공통 혼성화 영역을 갖는 이중 가닥 또는 단일 가닥(도면에 도시됨) 핵산 서열을 포함할 수 있다. 개별 식별자는 구성요소  $X_1 - X_A$ 를 포함하는 층 X(또는 층 1)로부터의 하나의 구성요소(가령, 고유 서열),  $Y_1 - Y_A$ 을 포함하는 층 Y(또는 층 2)로부터의 두 번째 구성요소(가령, 고유 서열), 및  $Z_1 - Z_B$ 를 포함하는 층 Z(또는 층 3)으로부터의 세 번째 구성요소(가령, 고유 서열)를 연결함으로써 구성될 수 있다. 층 X로부터의 구성요소는 층 Y로부터의 구성요소 상의 3' 말단과 상보성을 공유하는 3' 말단을 가질 수 있다. 따라서 층 X와 Y의 단일 가닥 구성요소는 3' 말단에서 함께 어닐링될 수 있으며 PCR을 사용하여 이중 가닥 핵산 분자를 생성하도록 확장될 수 있다. 생성된 이중 가닥 핵산 분자는 용융되어 층 Z로부터의 구성요소의 3' 말단과 상보성을 공유하는 3' 말단을 생성할 수 있다. 층 Z로부터의 구성요소는 생성된 핵산 분자와 어닐링될 수 있으며 고정된 순서로 층 X, Y, 및 Z로부터의 단일 구성요소를 포함하는 고유 식별자를 생성하도록 확장될 수 있다. OEPCR에 대한 화학적 방법 섹션 A를 참조할 수 있다. DNA 크기 선택(가령, 겔 추출, 화학적 방법 섹션 E 참조) 또는 최외각 층 측면에 있는 프라이머를 사용한 중합효소 연쇄 반응(PCR)(화학적 방법 섹션 D 참조)이 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 완전히 조립된 식별자 산물을 분리할 수 있다. 두 개의 최외각 층 각각에 대해 하나씩, 두 개의 프로브를 사용한 순차적 핵산 포획이 또한 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 완전히 조립된 식별자 산물을 분리할 수 있다(화학적 방법 섹션 F 참조).

[0183] 식별자는 도 64에 도시된 바와 같이 점착 말단 결찰을 사용하여 곱 방식에 따라 조립될 수 있다. 단일 가닥 3' 오버행은 갖는 이중 가닥 구성요소(가령, 이중 가닥 DNA(dsDNA))를 각각 포함하는 3개의 층이 사용되어 개별 식별자를 조립할 수 있다. 예를 들어, 식별자는 구성요소  $X_1 - X_A$ 를 포함하는 층 X(또는 층 1)로부터의 하나의 구성요소,  $Y_1 - Y_B$ 를 포함하는 층 Y(또는 층 2)의 두 번째 구성요소, 및  $Z_1 - Z_C$ 를 포함하는 층 Z(또는 층 3)으로부터의 세 번째 구성요소를 포함한다. 층 X로부터의 구성요소를 층 Y로부터의 구성요소와 결합하기 위해, 층 X의 구성요소는 도 64의 **a**로 라벨링되는 공통 3' 오버행을 포함할 수 있고, 층 Y의 구성요소가 공통적인, 상보적 3' 오버행인 **a\***를 포함할 수 있다. 층 Y로부터의 구성요소를 층 Z로부터의 구성요소와 결합하기 위해, 층 Y의 요소는 도 64의 **b**로 라벨링된 공통 3' 오버행을 포함할 수 있고, 층 Z의 요소는 공통의 상보적인 3' 오버행인 **b\***를 포함할 수 있다. 층 X의 구성요소의 3' 오버행은 층 Y 구성요소의 3' 말단에 상보적일 수 있고 층 Y 구성요소의 다른 3' 오버행은 층 Z 구성요소의 3' 말단에 상보적일 수 있어 구성요소가 혼성화되고 결찰될 수 있다. 따라서 층 X로부터의 구성요소는 층 X 또는 층 Z의 다른 구성요소와 혼성화될 수 없으며 마찬가지로 층 Y의 구성요소는 층 Y의 다른 요소와 혼성화될 수 없다. 또한 층 Y로부터의 단일 구성요소는 완전한 식별자의 형성을 보장하면서 층 X의 단일 구성요소 및 층 Z의 단일 구성요소로 결찰될 수 있다. 점착 말단 결찰에 대해서는 화학적 방법 섹션 B를 참조할 수 있다. DNA 크기 선택(가령, 겔 추출, 화학적 방법 섹션 E 참조) 또는 최외각 층 측면에 있는 프라이머를 사용한 중합효소 연쇄 반응(PCR)(화학적 방법 섹션 D 참조)이 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다. 두 개의 최외각 층 각각에 대해 하나씩, 두 개의 프로브를 사용한 순차적 핵산 포획이 또한 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다(화학적 방법 섹션 F 참조).

[0184] 점착 말단 결찰을 위한 점착 말단은 각 층의 구성요소를 제한 엔도뉴클레아제로 처리하여 생성될 수 있다(제한 효소 반응에 대한 자세한 내용은 화학적 방법 섹션 C 참조). 일부 실시예에서, 다수의 층의 구성요소는 구성요소의 하나의 "부모" 세트로부터 생성될 수 있다. 예를 들어, 이중 가닥 구성요소의 단일 부모 세트가 각 말단 상의 상보적인 제한 부위(가령, BamHI 및 BglIII에 대한 제한 부위)를 가질 수 있는 실시예가 있다. 조립을 위해 임의의 2개 구성요소가 선택될 수 있고, 하나 또는 다른 상보적 제한 효소(가령, BglIII 또는 BamHI)로 개별적으로 소화되어 함께 결찰될 수 있는 상보적인 점착 말단을 생성하여 불활성 흉터를 도출할 수 있다. 생성물 핵산 서열은 각 말단에 상보적 제한 부위(예를 들어, 5' 말단 상의 BamHI 및 3' 말단 상의 BglIII)를 포함할 수 있고,

동일한 프로세스에 따라 부모 세트로부터의 또 다른 구성요소에 추가로 결합될 수 있다. 이 프로세스는 무한정 순환될 수 있다(도 76). 부모가 N개의 구성요소를 포함하는 경우, 각 주기는 곱 방식에 N개의 구성요소의 추가 층을 추가하는 것과 동일할 수 있다.

[0185] 세트 X(가령, dsDNA의 세트 1)의 요소와 세트 Y(가령, dsDNA의 세트 2)의 요소를 포함하는 핵산의 서열을 구성하기 위해 결합을 사용하는 방법은 이중 가닥 서열의 2개 이상의 폴(가령, dsDNA의 세트 1 및 dsDNA의 세트 2)을 얻거나 구성하는 단계를 포함할 수 있으며, 제1 세트(가령, dsDNA의 세트 1)는 점착 말단(가령, **a**)을 포함하고 제2 세트(가령, dsDNA의 세트 2)는 제1 세트의 점착 말단에 상보적인 점착 말단(가령, **a\***)을 포함한다. 제1 세트(가령, dsDNA의 세트 1)로부터의 임의의 DNA와 제2 세트(가령, dsDNA의 세트 2)로부터의 DNA의 임의의 서브 세트가 조합되고 조립된 다음, 함께 결합되어 제1 세트로부터의 요소와 제2 세트로부터의 요소를 갖는 단일 이중 가닥 DNA를 형성할 수 있다.

[0186] 식별자는 도 65에 도시된 바와 같이 부위 특정적 재조합을 사용하여 곱 방식에 따라 조립될 수 있다. 식별자는 세 가지 상이한 층으로부터의 구성요소를 조립함으로써 구성될 수 있다. 층 X(또는 층 1)의 구성요소는 분자의 하나의 측 상에 attB<sub>x</sub> 재조합효소 부위가 있는 이중 가닥 분자를 포함할 수 있고, 층 Y(또는 층 2)로부터의 구성요소는 하나의 측 상에 attP<sub>x</sub> 재조합효소 부위가 있는 이중 가닥 분자를 포함할 수 있으며, 층 Z(또는 층 3)의 구성요소는 분자의 하나의 측 상의 attP<sub>y</sub> 재조합효소 부위를 포함할 수 있다. 한 쌍 내의 attB 및 attP 부위는 아래 첨자로 표시된 바와 같이 해당하는 재조합 효소의 존재 하에서 재조합될 수 있다. 층 X로부터의 하나의 구성요소가 층 Y로부터의 하나의 구성요소와 연관되고, 층 Y로부터의 하나의 구성요소가 층 Z로부터의 하나의 구성요소와 연관되도록 각각의 층으로부터의 하나씩의 구성요소가 조합될 수 있다. 하나 이상의 재조합효소의 적용이 구성요소를 재조합하여 정렬된 구성요소를 포함하는 이중 가닥 식별자를 생성할 수 있다. DNA 크기 선택(가령, 겔 추출) 또는 최외곽 층 측면에 있는 프라이머를 사용한 PCR이 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다. 일반적으로 다중 직교 attB 및 attP 쌍이 사용될 수 있으며, 각 쌍은 추가 층으로부터의 구성요소를 조립하는 데 사용될 수 있다. 큰 세린 계열의 재조합효소의 경우, 재조합효소당 최대 6개의 직교 attB 및 attP 쌍이 생성될 수 있으며, 다수의 직교 재조합효소도 구현될 수 있다. 예를 들어, 12개의 직교 attB 및 attP 쌍, 즉 BxbI 및 PhiC31과 같은 두 개의 큰 세린 재조합효소 각각으로부터의 6개의 직교 쌍을 사용하여 13개의 층이 조립될 수 있다. attB와 attP 쌍의 직교성은 한 쌍의 attB 사이트가 다른 쌍의 attP 사이트와 반응하지 않도록 보장한다. 이를 통해 서로 다른 층의 구성요소를 고정된 순서로 조립할 수 있다. 재조합효소 매개 재조합 반응은 구현된 재조합효소 시스템에 따라 가역적이거나 비가역적일 수 있다. 예를 들어, 큰 세린 재조합효소 계열은 고에너지 보조인자를 필요로 하지 않고 비가역적 재조합 반응을 촉매하는 반면, 티로신 재조합효소 계열은 가역적 반응을 촉매한다.

[0187] 식별자는 도 66a에 도시된 바와 같이 주형 지정 결합(TDL)을 사용하는 곱 방식에 따라 구성될 수 있다. 주형 지정 결합은 "주형" 또는 "스테인플"이라고 불리는 단일 가닥 핵산 서열을 활용하여 구성요소의 정렬된 결합을 촉진하여 식별자를 형성할 수 있다. 주형은 인접 층으로부터의 구성요소에 동시에 혼성화되어 리가제가 이를 결합하는 동안 서로 인접하게 유지한다(3' 말단 대 5' 말단). 도 66a의 예에서 단일 가닥 구성요소의 3개 층 또는 세트가 결합된다. 서열 a\*에 상보적인, 3' 말단에서 공통 서열 a를 공유하는 구성요소의 제1 층(예를 들어, 층 X 또는 층 1), 서열 b\* 및 c\*에 상보적인 공통 시퀀스 b 및 c를 각각 5' 및 3' 말단에서 공유하는 구성요소의 두 번째 층(가령, 층 Y 또는 층 2), 서열 d\*에 상보적일 수 있는 5' 말단에서 공통 서열 d를 공유하는 구성요소의 세 번째 층(예를 들어, 층 Z 또는 층 3), 및 서열 a\*b\*(5'에서 3')를 포함하는 첫 번째 스테이플과 서열 c\*d\*(5'에서 3')를 포함하는 두 번째 스테이플을 갖는 두 개의 주형 또는 "스테인플"의 세트. 이 예에서, 각 층의 하나 이상의 구성요소가 선택되어 스테이플과의 반응으로 혼합될 수 있으며, 이는 상보적 어닐링에 의해 정의된 순서로 각 층으로부터의 하나씩의 구성요소를 결합하여 식별자를 형성하는 것을 촉진할 수 있다. TDL에 대해서는 화학적 방법 섹션 B를 참조할 수 있다. DNA 크기 선택(가령, 겔 추출, 화학적 방법 섹션 E 참조) 또는 최외곽 층 측면에 있는 프라이머를 사용한 중합효소 연쇄 반응(PCR)(화학적 방법 섹션 D 참조)이 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다. 두 개의 최외곽 층 각각에 대해 하나씩, 두 개의 프로브를 사용한 순차적 핵산 포획이 또한 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다(화학적 방법 섹션 F 참조).

[0188] 도 66b는 6층 TDL로 각각 조립된 256개의 개별 핵산 서열의 복제수(풍부함)에 대한 히스토그램을 보여준다. 외곽 층(첫 번째 및 마지막 층)에는 각각 하나의 구성요소가 있고 각 내부 층(나머지 4개의 4개 층)에는 4개의 구성요소가 있다. 각 외곽 층 구성요소는 10개의 염기 혼성화 영역을 포함하여 28개의 염기였다. 각각의 내부 층

구성요소는 5' 말단 상의 10 염기 공통 혼성화 영역, 10 염기 가변(바코드) 영역, 및 3' 말단 상의 10 염기 공통 혼성화 영역을 포함하는 30개 염기였다. 3개의 주형 가닥 각각의 길이는 20개 염기였다. 모든 256개의 개별 서열은 모든 구성요소와 주형, T4 폴리뉴클레오티드 키나제(구성요소 인산화용), 및 T4 리가아제, ATP 및 기타 적절한 반응 시약을 포함하는 하나의 반응으로 다중 방식으로 조립되었다. 반응물이 37도에서 30분 동안 배양된 후 실온에서 1시간 동안 배양됐다. PCR을 통해 반응 생성물에 시퀀싱 어댑터(sequencing adapter)가 추가되었고 Illumina MiSeq 장비를 사용하여 생성물이 시퀀싱됐다. 192910개의 총 조립된 서열 리드 중 각각의 개별 조립된 서열의 상대적 복제수가 나타난다. 이 방법의 다른 실시예에는 이중 가닥 구성요소를 사용할 수 있으며, 여기서 구성요소는 초기에 용융되어 스테이플에 어닐링될 수 있는 단일 가닥 버전을 형성할 수 있다. 이 방법의 또 다른 실시예 또는 파생예(즉, TDL)가 곱 방식에서 달성될 수 있는 것보다 더 복잡한 식별자의 조합 공간을 구성하는 데 사용될 수 있다.

[0189] 식별자는 골든 게이트 조립체, 깃스 조립체 및 리가아제 순환 반응 조립체를 포함한 다양한 기타 화학적 구현을 사용하여 제품 체계에 따라 구성될 수 있다.

[0190] 도 67a 및 67b는 순열된 구성요소(가령, 핵산 서열)로 식별자(가령, 핵산 분자)를 구성하기 위한 "순열 방식"으로 불리는 예시적인 방법을 개략적으로 예시한다. 도 67a는 순열 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 각 층으로부터의 단일 구성요소를 프로그램 가능한 순서로 조합함으로써 식별자가 구성될 수 있다. 도 67b는 순열 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 예를 들어, 크기 6의 조합 공간은 각각 하나의 개별 구성요소를 포함하는 3개의 층으로부터 생성될 수 있다. 구성요소는 임의의 순서로 연결될 수 있다. 일반적으로, 각각  $M$ 개의 구성요소를 갖는  $M$ 개의 층을 사용하면 순열 방식을 통해 총  $N^M$ 개의 조합 공간이 가능해진다.

[0191] 도 67c는 주형 지정 결찰(TDL, 화학적 방법 섹션 B 참조)을 사용한 순열 방식의 예시적인 구현을 도시한다. 여러 층으로부터의 구성요소는 가장자리 스캐폴드(scaffold)라고도 하는 고정된 왼쪽 말단과 오른쪽 말단 구성요소 사이에 조립된다. 이들 가장자리 스캐폴드는 조합 공간의 모든 식별자에 대해 동일하므로 구현을 위한 반응 마스터 믹스의 일부로 추가될 수 있다. 상이한 층으로부터의 구성요소가 반응의 식별자에 통합되는 순서가 반응을 위해 선택된 주형에 따라 달라지도록 임의의 두 층 또는 스캐폴드 사이의 임의의 가능한 접합에 대한 주형 또는 스테이플이 존재한다.  $M$ 개의 층에 대한 임의의 가능한 층 순열을 가능하게 하기 위해, 모든 가능한 접합(스캐폴드와의 접합 포함)에 대해  $M+2M$ 개의 개별 선택 가능한 스테이플이 있을 수 있다. 이들 주형 중  $M$ 개(회색으로 음영 처리됨)는 층과 그 자체 사이의 접합을 형성하며 본 명세서에 설명된 순열 조립의 목적을 위해 제외될 수 있다. 그러나, 이들을 포함시키면 도 67d-g에 예시된 바와 같이 반복 구성요소를 포함하는 식별자로 더 큰 조합 공간을 가능하게 할 수 있다. DNA 크기 선택(가령, 겔 추출, 화학적 방법 섹션 E 참조) 또는 최외각 층 측면에 있는 프라이머를 사용한 증합효소 연쇄 반응(PCR)(화학적 방법 섹션 D 참조)이 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다. 두 개의 최외각 층 각각에 대해 하나씩, 두 개의 프로브를 사용한 순차적 핵산 포획이 또한 구현되어 반응에서 형성될 수 있는 다른 부산물로부터 식별자 산물을 분리할 수 있다(화학적 방법 섹션 F 참조).

[0192] 도 67d-g는 반복되는 구성요소를 갖는 식별자의 특정 인스턴스를 포함하도록 순열 방식이 어떻게 확장될 수 있는지에 대한 예시적인 방법을 도시한다. 도 67d는 도 67c의 구현 형태가 순열 및 반복 구성요소와 함께 어떻게 사용될 수 있는지에 대한 예를 도시한다. 예를 들어, 식별자는 두 개의 개별 구성요소로부터 조립된 총 세 개의 구성요소를 포함할 수 있다. 이 예에서, 층의 구성요소가 식별자에 여러 번 나타날 수 있다. 동일한 구성요소의 인접 연결은 동일한 구성요소, 가령, 도면에서  $a*b*(5'에서 3')$  스테이플의 3' 말단과 5' 말단 모두에 대해 인접한 상보적 혼성화 영역이 있는 스테이플을 사용하여 달성할 수 있다. 일반적으로  $M$ 개의 층에 대해,  $M$ 개의 이러한 스테이플이 있다. 이러한 구현에 반복된 구성요소를 통합하면 도 67e에 도시된 바와 같이, 가장자리 스캐폴드 사이에 조립되는 길이가 2개를 초과하는(즉, 1개, 2개, 3개, 4개 이상의 구성요소를 포함하는) 핵산 서열을 생성할 수 있다. 도 67e는 도 67d의 예시적인 구현 방법이 식별자 외에 가장자리 스캐폴드들 사이에 조립되는 비표적 핵산 서열을 도출할 수 있다. 적절한 식별자가 가장자리 상의 동일한 프라이머 결합 부위를 공유하기 때문에 PCR에 의해 비표적 핵산 서열로부터 분리될 수 없다. 그러나 이 예에서는, (가령, 모든 구성요소가 동일한 길이를 갖는 경우) 각각의 조립된 핵산 서열이 고유한 길이를 갖도록 설계될 수 있기 때문에, DNA 크기 선택(예를 들어, 겔 추출을 사용하여)이 구현되어 비표적 서열로부터 표적 식별자(가령, 위에서 두 번째 서열)를 분리할 수 있다. 크기 선택에 대해서는 화학적 방법 섹션 E를 참조할 수 있다. 도 67f는 반복된 구성요소로 식별자를 구성하는 것이 동일한 반응에서 가장자리 서열은 동일하지만 길이가 다른 다중 핵산 서열을 생성할 수 있

는 또 다른 예를 보여준다. 이 방법에서는 교대 패턴으로 한 층의 구성요소를 다른 층의 구성요소와 조립하는 주형이 사용될 수 있다. 도 67e에 도시된 방법을 이용할 때, 크기 선택은 설계된 길이의 식별자를 선택하는 데 사용될 수 있다. 도 67g는 반복된 구성요소로 식별자를 구성하는 것이 동일한 가장자리 서열을 갖고 일부 핵산 서열(예를 들어 위에서 세 번째와 네 번째, 위에서 여섯 번째와 일곱 번째)에 대해 동일한 길이를 갖는 다중 핵산 서열을 생성할 수 있는 예를 보여준다. 이 예에서, PCR 및 DNA 크기 선택이 구현되더라도, 다른 하나를 구성하지 않고 하나를 구성하는 것이 불가능할 수 있으므로 동일한 길이를 공유하는 핵산 서열은 둘 다 개별 식별자에서 제외될 수 있다.

[0193] 도 68a - 68d는 더 많은 개수  $M$ 의 가능한 구성요소 중 임의의 개수  $k$ 의 조립된 구성요소(가령, 핵산 서열)를 갖는 식별자(가령, 핵산 분자)를 구성하기 위한 "MchooseK" 방식이라 지칭되는, 예시적 방법을 개략적으로 도시한다. 도 68a는 MchooseK 방식을 사용하여 구성된 식별자의 아키텍처를 도시한다. 이 방법을 사용하면 모든 층의 임의의 서브세트에 있는 각 층에서 하나의 구성요소를 조립함으로써 식별자가 구성된다(가령,  $M$ 개의 가능한 층 중  $k$  층에서 구성요소 선택). 도 68b는 MchooseK 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 도시한다. 이 조립 방식에서 조합 공간은  $M$ 개의 층, 층당  $N$ 개의 구성요소, 및  $k$ 개의 구성요소의 식별자 길이에 대한  $N^k M \text{choose} K$  가능한 식별자를 포함할 수 있다. 예를 들어, 각각 하나의 구성요소를 포함하는 5개의 층이 있는 경우, 각각 2개의 구성요소를 포함하는 최대 10개의 개별 식별자가 조립될 수 있다.

[0194] MchooseK 방식은 도 68c에 도시된 바와 같이 주형 지정 결찰(화학적 방법 섹션 B 참조)을 사용하여 구현될 수 있다. 순열 방식(도 67c)에 대한 TDL 구현과 마찬가지로, 이 예의 구성요소는 반응 마스터 믹스에 포함될 수도 있고 포함되지 않을 수도 있는 가장자리 스캐폴드 사이에 조립된다. 구성요소는  $M$ 개의 층, 예를 들어 2에서  $M$ 까지 미리 정의된 순위를 갖는  $M = 4$ 개의 층으로 분할될 수 있으며, 여기서 왼쪽 가장자리 스캐폴드는 순위 1일 수 있고 오른쪽 가장자리 스캐폴드는 순위  $M+1$ 일 수 있다. 주형은 각각 낮은 순위에서 높은 순위로 임의의 두 구성요소의 3'에서 5' 연결을 위한 핵산 서열을 포함한다. 이러한 주형이  $((M+1)^2 + M+1)/2$ 개 있다. 개별 층으로부터 임의의  $K$  구성요소의 개별 식별자가 결찰 반응에서 선택된 구성요소를 순위 순서로 가장자리 스캐폴드와 함께  $K$  구성요소를 가져오는 데 사용되는 상응하는  $K+1$  스테이플과 결합함으로써 구축될 수 있다. 이러한 반응 설정은 가장자리 스캐폴드 사이의 표적 식별자에 해당하는 핵산 서열을 생성할 수 있다. 대안으로, 모든 주형을 포함하는 반응 혼합물을 선택된 구성요소와 결합하여 표적 식별자를 조립할 수 있다. 이 대안적인 방법은 도 68d에 예시된 바와 같이 동일한 가장자리 서열을 갖지만 길이가 개별적인(모든 구성요소 길이가 동일한 경우) 다양한 핵산 서열을 생성할 수 있다. 표적 식별자(하단)는 크기별로 부산물 핵산 서열로부터 분리될 수 있다. 핵산 크기 선택에 대해서는 화학적 방법 섹션 E를 참조할 수 있다.

[0195] 도 69a 및 도 69b는 분할된 구성요소로 식별자를 구성하기 위한 "분할 방식(partition scheme)"으로 지칭되는 예시적인 방법을 개략적으로 예시한다. 도 69a는 분할 방식을 사용하여 구성될 수 있는 식별자의 조합 공간의 예를 보여준다. 개별 식별자는 서로 다른 층의 두 구성요소 사이에 파티션(특별히 분류된 구성요소)을 선택적으로 배치하여 고정된 순서로 각 층의 하나의 구성요소를 조립하여 구성될 수 있다. 예를 들어, 구성요소의 세트는 하나의 파티션 구성요소와 각각 하나의 구성요소를 포함하는 4개의 층으로 구성될 수 있다. 각 층으로부터의 구성요소는 고정된 순서로 조합될 수 있으며 단일 파티션 구성요소는 층들 사이의 다양한 위치에서 조립될 수 있다. 이 조합 공간의 식별자는 파티션 구성요소를 포함하지 않고, 첫 번째와 두 번째 층의 구성요소들 사이의 파티션 구성요소, 두 번째와 세 번째 층의 구성요소들 사이의 파티션 등을 포함하여 8개의 가능한 식별자의 조합 공간을 만들 수 있다. 일반적으로, 각각  $N$ 개의 구성요소를 갖는  $M$ 개의 층과  $p$ 개의 파티션 구성요소를 사용하면  $N^k (p+1)^{M-1}$ 개의 가능한 식별자가 구성될 수 있다. 이 방법은 다양한 길이의 식별자를 생성할 수 있다.

[0196] 도 69b는 주형 지정 결찰을 사용하는 파티션 방식의 구현 예를 보여준다(화학적 방법 섹션 B 참조). 주형은 고정된 순서로  $M$ 개의 층 각각으로부터의 하나씩 구성요소를 함께 결찰하기 위한 핵산 서열을 포함한다. 각 파티션 구성요소에 대해, 파티션 구성요소가 임의의 인접한 두 층으로부터의 구성요소들 사이에 결찰할 수 있도록 하는 추가 주형 쌍이 존재한다. 예를 들어 하나의 쌍에서의 하나의 주형(가령, 서열  $g*b*(5'에서 3')$  포함)이 층 1의 3' 말단(서열  $b$  포함)을 분할 구성요소의 5' 말단(서열  $q$  포함)으로 결찰할 수 있도록 그리고 상기 쌍에서의 제 2 주형(가령, 서열  $c*h*(5' to 3')$  포함)이 분할 구성요소의 3' 말단(서열  $h$  포함)을 층 2의 5' 말단(서열  $c$  포함)으로 결찰할 수 있도록 주형 쌍이 이뤄진다. 인접한 층의 임의의 두 구성요소들 사이에 파티션을 삽입하기 위해, 해당 층을 함께 결찰하기 위한 표준 주형이 반응에서 제외될 수 있으며 해당 위치에서 파티션을 결찰하기 위한 주형 쌍을 반응에서 선택할 수 있다. 현재 예에서, 층 1과 층 2 사이의 파티션 구성요소를 표적으로 하는 것은 주형  $c*b*(5'에서 3')$ 보다 주형 쌍  $c*h*(5'에서 3')$  및  $g*b*(5'에서 3')$ 를 사용하여 반응을 선택할 수 있

다. 구성요소는 반응 혼합물에 포함될 수 있는 가장자리 스캐폴드들 사이에 조립될 수 있다(각각 첫 번째 및 M 번째 층에 결합하기 위한 해당 주형과 함께). 일반적으로 총 약  $M-1+2*p*(M-1)$ 개의 선택 가능한 주형이 M개 층과 p개 파티션 구성요소에 대해 이 방법에 사용될 수 있다. 이러한 파티셔닝 방식의 구현은 동일한 가장자리 서열을 갖지만 길이가 다른 반응에서 다양한 핵산 서열을 생성할 수 있다. 표적 식별자는 DNA 크기 선택을 통해 부산물 핵산 서열로부터 분리될 수 있다. 구체적으로, 정확히 M개의 층 구성요소를 갖는 정확히 하나의 핵산 서열 생성물이 있을 수 있다. 층 구성요소가 파티션 구성요소에 비해 충분히 크게 설계되면, 전역 크기 선택 영역을 정의함으로써, 식별자(그리고 비표적 부산물 중 아무것도 없음)가 식별자 내 구성요소의 특정 파티셔닝에 무관하게 선택될 수 있음으로써, 다수의 반응으로부터의 다수의 파티셔닝된 식별자가 동일한 크기 선택 단계에서 분리될 수 있다. 핵산 크기 선택에 대해서는 화학적 방법 섹션 E를 참조할 수 있다.

[0197] 도 70a 및 도 70b는 다수의 가능한 구성요소로부터의 구성요소의 임의의 스트링으로 구성된 식별자를 구성하기 위한 "비제한 스트링(unconstrained string)"(또는 USS) 방식으로 지칭되는 예시적인 방법을 개략적으로 나타낸다. 도 70a는 비제한 스트링 방식을 사용하여 구성될 수 있는 3-구성요소(또는 4-스캐폴드) 길이 식별자의 조합 공간의 예를 보여준다. 비제한 스트링 방식은 하나 이상의 층에서 각각 가져온 하나 이상의 개별 구성요소를 사용하여 길이가 K 구성요소인 개별 식별자를 구성하며, 여기서 각 개별 구성요소는 식별자의 K 구성요소 위치 중 하나에 나타날 수 있다(반복 허용). 예를 들어, 각각 하나의 구성요소를 포함하는 두 개의 층에 대해, 8개의 가능한 3-구성요소 길이 식별자가 있다. 일반적으로, 각각 하나씩의 구성요소를 가진 M개의 층에는 길이 K 구성요소의  $M^K$ 개의 가능한 식별자가 있다. 도 70b는 주형 지정 결합을 사용하여 비제한 스트링 방식의 구현 예를 보여준다(화학적 방법 섹션 B 참조). 이 방법에서는 K+1 단일 가닥 및 정렬된 스캐폴드 DNA 구성요소(2개의 가장자리 스캐폴드 및 K-1개의 내부 스캐폴드 포함)가 반응 혼합물에 존재한다. 개별 식별자는 인접한 스캐폴드의 모든 쌍 사이에 연결된 단일 구성요소를 포함한다. 예를 들어, 스캐폴드 A와 B 사이에 결합된 구성요소, 스캐폴드 C와 D 사이에 결합된 구성요소 등 모든 K개의 인접한 스캐폴드 접합부가 구성요소에 의해 점유될 때까지 계속된다. 반응에서는, 상이한 층으로부터의 선택된 구성요소가 선택된 스테이플 쌍과 함께 스캐폴드에 도입되어 적절한 스캐폴드에 조립되도록 지시한다. 예를 들어, 스테이플  $a^*L^*$  (5'에서 3') 및  $A^*b^*$  (5'에서 3') 쌍은 5' 말단 영역 'a' 및 3' 말단 영역 'b'이 있는 층 1 구성요소에게 L과 A 스캐폴드 사이에 결합할 것을 지정한다. 일반적으로 M개의 층과 K+1개의 스캐폴드의 경우,  $2^M * M^K$ 개의 선택 가능 스테이플이 사용되어 길이 K의 임의의 USS 식별자를 구성할 수 있다. 구성요소를 5' 말단 상의 스캐폴드에 연결하는 스테이플이 동일한 구성요소를 3' 말단 상의 스캐폴드에 연결하는 스테이플로부터 분리되어 있기 때문에, 핵산 부산물이 동일한 가장자리 스캐폴드와의 반응에서 표적 식별자로서 형성될 수 있지만, K개 미만의 구성요소(K+1개 미만의 스캐폴드) 또는 K개 초과 구성요소(K+1개 초과 스캐폴드)가 포함되어 있다. 표적 식별자는 정확히 K개의 구성요소(K+1개의 스캐폴드)로 형성될 수 있으므로 모든 구성요소의 길이가 동일하도록 설계되고 모든 스캐폴드의 길이가 동일하도록 설계된 경우 DNA 크기 선택과 같은 기술을 통해 선택할 수 있다. 핵산 크기 선택에 대해서는 화학적 방법 섹션 E를 참조할 수 있다. 층당 하나의 구성요소가 있을 수 있는 제한되지 않는 스트링 방식의 특정 구현예에서, 해당 구성요소는 (1) 식별 바코드, (2) 스캐폴드로의 5' 말단의 스테이플-매개 결합을 위한 혼성화 영역, 및 (3) 스캐폴드로의 3' 말단의 스테이플 매개 결합에 대한 혼성화 영역에 대한 3가지 모든 역할을 수행하는 단일 개별 핵산 서열만을 포함한다.

[0198] 도 70b에 도시된 내부 스캐폴드는 구성요소로서의 스캐폴드의 스테이플 매개 5' 결합 및 또 다른(반드시 개별적인 것은 아닌) 구성요소로서의 스캐폴드의 스테이플 매개 3' 결합 모두에 대해 동일한 혼성화 서열을 사용하도록 설계될 수 있다. 따라서 도 70b에 도시된 1-스캐폴드, 2-스테이플 적층 혼성화 이벤트는 스캐폴드와 각 스테이플 사이에서 발생하여 5' 구성요소 결합 및 3' 구성요소 결합을 모두 가능하게 하는 통계적 앞뒤 혼성화 이벤트를 나타낸다. 제한되지 않는 스트링 방식의 다른 구현예에서, 스캐폴드는 2개의 연결된 혼성화 영역, 즉 스테이플 매개 3' 결합을 위한 개별 3' 혼성화 영역과 스테이플 매개 5' 결합을 위한 개별 5' 혼성화 영역으로 설계될 수 있다.

[0199] 도 71a 및 71b는 모 식별자로부터 핵산 서열(또는 구성요소)을 삭제함으로써 식별자를 구성하기 위한 "구성요소 삭제 방식"으로 지칭되는 예시적인 방법을 개략적으로 예시한다. 도 71a는 구성요소 삭제 방식을 사용하여 구성될 수 있는 가능한 식별자의 조합 공간의 예를 보여준다. 이 예에서 부모 식별자는 여러 구성요소로 구성될 수 있다. 부모 식별자는 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50개 이상의 구성요소를 포함할 수 있다. 개별 식별자는 N개의 가능한 구성요소에서 임의의 수의 구성요소를 선택적으로 삭제하여, 크기  $2^N$ 의 "전체" 조합 공간을 생성하거나, N개의 가능한 구성요소에서 고정된 개수인 K개의 구성요소를 삭제하여 크기  $N\text{choose}K$ 의 " $N\text{choose}K$ "를 생성함으로써 구성될 수 있다. 3개의 구성요소가 있는 부모 식별자가 있는 예에서, 전체 조합 공

간은 8이 될 수 있고 3choose2 조합 공간은 3이 될 수 있다.

[0200] 도 71b는 이중 가닥 표적화 절단 및 복구(DSTCR)를 사용하는 구성요소 삭제 방식의 예시적인 구현을 보여준다. 부모 서열은 뉴클레아제 특이적 표적 부위(길이가 4개 이하의 염기일 수 있음) 옆에 있는 구성요소를 포함하는 단일 가닥 DNA 기질일 수 있으며, 여기서 모 서열은 표적 부위에 대응하는 하나 이상의 이중 가닥 특이적 뉴클레아제와 함께 배양될 수 있다. 개별 구성요소는 모체의 구성요소 DNA(및 인접 뉴클레아제 부위)에 결합하는 상보적인 단일 가닥 DNA(또는 절단 주형)를 사용하여 삭제의 표적이 될 수 있으며, 따라서 뉴클레아제에 의해 양쪽 말단 모두에서 절단될 수 있는 모체에 안정한 이중 가닥 서열을 형성한다. 또 다른 단일 가닥 DNA(또는 복구 주형)는 부모의 분리된 말단(구성요소 서열이 그 사이에 있었던 것)에 혼성화하고 결찰을 위해 직접적으로 또는 대체 서열에 의해 연결되도록 함으로써 부모에는 더 이상 활성 뉴클레아제 표적 사이트가 포함되어 있지 않는다. 우리는 이 방법을 "이중 가닥 표적 절단"(DSTC)이라고 한다. 크기 선택은 특정 개수의 삭제된 구성요소가 있는 식별자를 선택하는 데 사용될 수 있다. 핵산 크기 선택에 대해서는 화학적 방법 섹션 E를 참조할 수 있다.

[0201] 대안으로 또는 추가로, 부모 식별자는 두 개의 구성요소가 동일한 서열의 측면에 위치하지 않도록 스페이서 서열에 의해 분리된 구성요소를 포함하는 이중 또는 단일 가닥 핵산 기질일 수 있다. 부모 식별자는 Cas9 뉴클레아제와 함께 배양될 수 있다. 개별 구성요소는 구성요소의 가장자리에 결합하고 측면 부위에서 Cas9 매개 절단을 가능하게 하는 가이드 리보핵산(절단 주형)을 사용하여 삭제 대상이 될 수 있다. 단일 가닥 핵산(복구 주형)은 부모 식별자의 결과적인 분리된 말단(가령, 구성요소 서열이 있었던 말단 사이)에 혼성화하여 결찰을 위해 이들을 하나로 모을 수 있다. 결찰은 직접적으로 수행되거나 대체 서열로 말단을 연결하여 부모의 결찰된 서열이 더 이상 Cas9의 표적이 될 수 있는 스페이서 서열을 포함하지 않도록 할 수 있다. 우리는 이 방법을 "서열 특이적 표적 절단 및 복구" 또는 "SSTCR"이라고 부른다.

[0202] 식별자는 DSTCR의 파생물을 사용하여 상위 식별자에 구성요소를 삽입하여 구성할 수 있다. 부모 식별자는 뉴클레아제 특이적 표적 부위(길이가 4개 이하의 염기일 수 있음)를 포함하는 단일 가닥 핵산 기질일 수 있으며, 각각은 별개의 핵산 서열 내에 내장되어 있다. 부모 식별자는 표적 부위에 대응하는 하나 이상의 이중 가닥 특이적 뉴클레아제와 함께 배양될 수 있다. 부모 식별자의 개별 표적 부위는 표적 부위와 부모 식별자의 별개의 주변 핵산 서열에 결합하여 이중 가닥 부위를 형성하는 상보적인 단일 가닥 핵산(절단 주형)을 사용하여 구성요소 삽입을 위해 표적화될 수 있다. 이중 가닥 부위는 뉴클레아제에 의해 절단될 수 있다. 또 다른 단일 가닥 핵산(복구 주형)은 부모 식별자의 분리된 말단에 혼성화하여 결찰을 위해 이들을 하나로 모을 수 있으며, 구성요소 서열에 의해 연결되어 부모의 결찰된 서열은 더 이상 활성 뉴클레아제 표적 부위를 포함하지 않는다. 대안으로 SSTCR의 파생물을 사용하여 구성요소를 상위 식별자에 삽입할 수 있다. 부모 식별자는 이중 가닥 또는 단일 가닥 핵산일 수 있으며 부모 식별자는 Cas9 뉴클레아제와 함께 배양될 수 있다. 부모 식별자 상의 개별 부위는 가이드 RNA(절단 주형)를 사용하여 절단의 표적이 될 수 있다. 단일 가닥 핵산(복구 주형)은 모 식별자의 분리된 말단에 혼성화하여 결찰을 위해 함께 모을 수 있으며 구성요소 서열에 의해 연결되어 모 식별자의 결찰된 서열은 더 이상 활성 뉴클레아제 표적 부위를 포함하지 않는다. 크기 선택을 사용하여 특정 수의 구성요소 삽입이 있는 식별자를 선택할 수 있다.

[0203] 도 72는 제조합효소 인식 부위를 갖는 부모 식별자를 개략적으로 예시한다. 다양한 패턴의 인식 부위는 다양한 제조합효소에 의해 인식될 수 있다. 특정 세트의 제조합효소에 대한 모든 인식 부위는 제조합효소가 적용되면 그 사이의 핵산이 제거될 수 있도록 배열된다. 도 72에 도시된 핵산 가닥은 적용되는 제조합효소의 서브세트에 따라  $2^5=32$ 개의 상이한 서열을 채택할 수 있다. 일부 실시예에서, 도 72에 도시된 바와 같이, DNA의 세그먼트를 잘라내고, 이동하고, 반전시키고, 전치시키는 제조합효소를 사용하여 독특한 분자가 생성되어 다른 핵산 분자를 생성할 수 있다. 일반적으로,  $N$ 개의 제조합효소를 사용하면 부모로부터  $2^N$ 개의 가능한 식별자가 만들어질 수 있다. 일부 실시예에서, 하나의 제조합효소의 적용이 하류 제조합효소가 적용될 때 발생하는 제조합 사건의 유형에 영향을 미치도록, 상이한 제조합효소로부터의 인식 부위의 다수의 직교 쌍이 중첩 방식으로 모 식별자 상에 배열될 수 있다(본 명세서에 참조로서 포함되는 Roquet et al., Synthetic recombinase-based state machines in living cells, Science 353 (6297): aad8559 (2016)를 참조할 수 있다). 이러한 시스템은  $N$ 개의 제조합효소의 모든 정렬,  $N!$ 에 대해 서로 다른 식별자를 구성할 수 있다. 제조합효소는 Flp 및 Cre와 같은 티로신 계열이거나 PhiC31, BxbI, TP901 또는 A118과 같은 대규모 세린 제조합효소 계열일 수 있다. 큰 세린 제조합효소 계열의 제조합효소를 사용하는 것은 비가역적 제조합을 촉진하고 따라서 다른 제조합효소보다 더 효율적으로 식별자를 생성할 수 있기 때문에 유리할 수 있다.

- [0204] 일부 경우에, 다수의 재조합효소를 별개의 순서로 적용함으로써 단일 핵산 서열이 다수의 별개의 핵산 서열이 되도록 프로그래밍될 수 있다. 대략  $\sim e^M$ 개의 개별 핵산 서열은 재조합효소의 수  $M$ 이 큰 세린 재조합효소 계열에 대해 7 이하일 수 있는 경우,  $M$  재조합효소를 다른 서브세트 및 이의 순서로 적용함으로써 생성될 수 있다. 재조합효소의 수  $M$ 이 7보다 클 수 있는 경우, 생성될 수 있는 서열의 수는 대략  $3.9^M$ 에 가까우며, 예를 들어, 본 명세서에 참조로서 그 전체가 포함되는 Roquet et al., Synthetic recombinase-based state machines in living cells, Science 353 (6297): aad8559 (2016)을 참조할 수 있다. 하나의 공통 서열에서 다른 DNA 서열을 생산하기 위한 추가 방법에는 CRISPR-Cas, TALENs 및 징크 핑거 뉴클레아제(Zinc Finger Nucleases)와 같은 표적 핵산 편집 효소가 포함될 수 있다. 재조합효소, 표적화된 편집 효소 등에 의해 생성된 서열은 임의의 이전 방법, 예를 들어 본 출원의 임의의 도면 및 개시내용에 개시된 방법과 함께 사용될 수 있다.
- [0205] 인코딩될 정보의 비트스트림이 임의의 단일 핵산 분자에 의해 인코딩될 수 있는 것보다 큰 경우, 정보는 분할되어 핵산 서열 바코드로 인덱싱될 수 있다. 더욱이,  $N$ 개의 핵산 분자의 세트로부터 크기  $k$ 개의 핵산 분자의 임의의 서브세트가 선택되어  $\log_2(N\text{choose}k)$  비트의 정보를 생성할 수 있다. 바코드는 더 긴 비트 스트림을 인코딩하기 위해 크기  $k$ 의 서브세트 내의 핵산 분자에 조립될 수 있다. 예를 들어,  $M$ 개의 바코드가  $M * \log_2(N\text{choose}k)$  비트의 정보를 생성하는 데 사용될 수 있다. 세트에 있는 이용 가능한 핵산 분자의 수  $N$ 과 이용 가능한 바코드의 수  $M$ 이 주어지면, 정보를 인코딩하기 위해 풀에서 분자의 총 수를 최소화하도록 크기  $k = k_0$ 의 서브세트가 선택될 수 있다. 디지털 정보를 인코딩하기 위한 방법은 비트 스트림을 분할하고 개별 요소를 인코딩하기 위한 단계를 포함할 수 있다. 예를 들어, 6비트를 포함하는 비트 스트림은 각 구성요소가 2비트로 구성되는 3개의 구성요소로 분할될 수 있다. 각각의 2비트 구성요소는 바코드로 정보 카세트를 형성할 수 있으며 함께 그룹화되거나 풀링되어 정보 카세트의 하이퍼 풀을 형성할 수 있다.
- [0206] 바코드는 인코딩할 디지털 정보의 양이 하나의 풀에만 들어갈 수 있는 양을 초과하는 경우 정보 색인화를 용이하게 할 수 있다. 더 긴 비트 스트링 및/또는 다중 바이트를 포함하는 정보는 도 59에 개시된 접근 방식을 총함함으로써, 가령, 핵산 인덱스를 사용해 인코딩된 고유 핵산 서열을 갖는 태그를 포함시킴으로써, 인코딩될 수 있다. 정보 카세트 또는 식별자 라이브러리는 주어진 서열이 해당하는 비트 스트림의 구성요소 또는 구성요소들을 나타내는 바코드 또는 태그 외에 위치 및 비트 값 정보를 제공하는 고유한 핵산 서열을 포함하는 질소 함유 염기 또는 핵산 서열을 포함할 수 있다. 정보 카세트는 하나 이상의 고유한 핵산 서열뿐만 아니라 바코드 또는 태그를 포함할 수 있다. 정보 카세트 상의 바코드 또는 태그는 정보 카세트 및 정보 카세트에 포함된 모든 시퀀스에 대한 참조를 제공할 수 있다. 예를 들어, 정보 카세트 상의 태그 또는 바코드는 고유 시퀀스가 비트 스트림의 어느 부분 또는 비트 스트림의 비트 구성요소에 대한 정보(예를 들어, 비트 값 및 비트 위치 정보)를 인코딩하는지 나타낼 수 있다.
- [0207] 바코드를 사용하면, 가능한 식별자의 조합 공간 크기보다 더 많은 비트 단위의 정보를 풀에 인코딩할 수 있다. 예를 들어, 10 비트 시퀀스는 두 개의 바이트 세트로 분리될 수 있으며, 각 바이트는 5 비트로 구성된다. 각 바이트는 5개의 가능한 개별 식별자의 세트에 매핑될 수 있다. 초기에, 각 바이트에 대해 생성된 식별자가 동일할 수 있지만 별도의 풀에 보관되거나 정보를 읽는 사람이 특정 핵산 서열이 어느 바이트에 속하는지 알 수 없을 수도 있다. 그러나 각 식별자는 인코딩된 정보가 적용되는 바이트에 대응하는 라벨로 바코드가 지정되거나 태그가 지정될 수 있고(가령, 바코드 1은 처음 5 비트를 제공하기 위해 핵산 풀의 서열에 부착될 수 있고 바코드 2는 두 번째 5 비트를 제공하기 위해 핵산 풀 내 서열에 부착될 수 있음), 그런 다음 2 바이트에 대응하는 식별자가 하나의 풀(가령, "하이퍼-풀" 또는 하나 이상의 식별자 라이브러리)로 조합될 수 있다. 하나 이상의 조합 식별자 라이브러리의 각 식별자 라이브러리는 주어진 식별자를 주어진 식별자 라이브러리에 속하는 것으로 식별하는 개별 바코드를 포함할 수 있다. 식별자 라이브러리 내 각 식별자에 바코드를 추가하기 위한 방법은 PCR, Gibson, 절찰 또는 주어진 바코드(가령, 바코드 1)가 주어진 핵산 샘플 풀에 부착될 수 있게 하는(가령, 바코드 1을 핵산 샘플 풀 1에 부착하고 바코드 2를 핵산 샘플 풀 2에 부착함) 그 밖의 다른 임의의 접근 방식을 사용하는 것을 포함할 수 있다. 하이퍼-풀로부터의 샘플은 시퀀싱 방법으로 판독될 수 있으며, 바코드나 태그를 사용하여 시퀀싱 정보를 파싱할 수 있다.  $M$ 개의 바코드 세트와  $N$ 개의 가능한 식별자(조합 공간)가 있는 식별자 라이브러리와 바코드를 사용하는 방법은  $M$ 과  $N$ 의 곱과 동일한 길이의 비트 스트림을 인코딩할 수 있다.
- [0208] 일부 실시예에서, 식별자 라이브러리는 웰(well)의 어레이에 저장될 수 있다. 웰의 어레이는  $n$ 개의 열과  $q$ 개의 행을 갖는 것으로 정의될 수 있으며, 각 웰은 하이퍼-풀에 2개 이상의 식별자 라이브러리를 포함할 수 있다. 각각의 웰에 인코딩된 정보는 각각의 웰에 포함된 정보보다  $n \times q$  더 큰 크기의 하나의 큰 연속 정보를 구성할 수

있다. 웰의 어레이의 웰 중 하나 이상으로부터 분취량을 채취할 수 있으며, 시퀀싱, 혼성화 또는 PCR을 사용하여 인코딩이 판독될 수 있다.

[0209] 핵산 샘플 풀, 하이퍼-풀, 식별자 라이브러리, 식별자 라이브러리의 그룹, 또는 핵산 샘플 풀이나 하이퍼-풀을 포함하는 웰은 정보 비트에 대응하는 고유한 핵산 분자(가령, 식별자) 및 복수의 보충 핵산 서열을 포함할 수 있다. 보충 핵산 서열은 인코딩된 데이터에 대응하지 않을 수 있다(예를 들어, 비트 값에 대응하지 않음). 보충 핵산 샘플은 샘플 풀에 저장된 정보를 마스킹하거나 인코딩할 수 있다. 보충 핵산 서열은 생물학적 공급원로부터 유래되거나 합성적으로 생산될 수 있다. 생물학적 공급원으로부터 유래된 보충 핵산 서열은 무작위로 단편화된 핵산 서열 또는 합리적으로 단편화된 서열을 포함할 수 있다. 특히 합성으로 인코딩된 정보(예를 들어, 식별자의 조합 공간)가 자연 유전 정보(예를 들어, 단편화된 게놈)와 닮도록 만들어진 경우, 생물학적으로 유래된 보충 핵산은 합성으로 인코딩된 정보와 함께 천연 유전 정보를 제공함으로써 시료 풀 내의 데이터 포함 핵산을 숨기거나 모호하게 할 수 있다. 하나의 예에서, 식별자는 생물학적 공급원에서 유래되고, 보충 핵산은 생물학적 공급원에서 유래된다. 샘플 풀은 여러 세트의 식별자와 보충 핵산 서열을 포함할 수 있다. 각 식별자 세트와 보충 핵산 서열은 서로 다른 유기체에서 유래될 수 있다. 하나의 예에서, 식별자는 하나 이상의 유기체로부터 유래되고, 보충 핵산 서열은 단일의 상이한 유기체로부터 유래된다. 보충 핵산 서열은 또한 하나 이상의 유기체로부터 유래될 수 있고, 식별자는 보충 핵산이 유래되는 유기체와는 다른 단일 유기체로부터 유래될 수 있다. 식별자와 보충 핵산 서열 둘 다는 다수의 서로 다른 유기체로부터 유래될 수 있다. 식별자를 보충 핵산 서열과 구별하기 위해 키가 사용될 수 있다.

[0210] 보충 핵산 서열은 기록된 정보에 대한 메타데이터를 저장할 수 있다. 메타데이터는 원본 정보의 출처 및/또는 원본 정보의 의도된 수신자를 결정 및/또는 승인하기 위한 추가 정보를 포함할 수 있다. 메타데이터는 원본 정보의 형식, 원본 정보를 인코딩하고 기록하는 데 사용된 도구 및 방법, 원본 정보를 식별자에 기록한 날짜 및 시간에 대한 추가 정보를 포함할 수 있다. 메타데이터는 원본 정보의 형식, 원본 정보를 인코딩하고 기록하는 데 사용된 도구 및 방법, 원본 정보를 핵산 서열에 기록한 날짜 및 시간에 대한 추가 정보를 포함할 수 있다. 메타데이터는 정보를 핵산 서열에 기록한 후 원래 정보에 적용된 수정에 대한 추가 정보를 포함할 수 있다. 메타데이터는 원본 정보에 대한 주석 또는 외부 정보에 대한 하나 이상의 참조를 포함할 수 있다. 대안으로 또는 추가로, 메타데이터는 식별자에 부착된 하나 이상의 바코드 또는 태그에 저장될 수 있다.

[0211] 식별자 풀의 식별자는 길이가 서로 동일하거나 유사하거나 다를 수 있다. 보충 핵산 서열은 식별자의 길이보다 작거나, 실질적으로 동일하거나, 더 큰 길이를 가질 수 있다. 보충 핵산 서열은 식별자의 평균 길이의 1개 염기 이내, 2개 염기 이내, 3개 염기 이내, 4개 염기 이내, 5개 염기 이내, 6개 염기 이내, 7개 염기 이내, 8개 염기 이내, 9개 염기 이내, 10개 염기 이내, 또는 그 이상의 염기 이내인 평균 길이를 가질 수 있다. 하나의 예에서, 보충 핵산 서열은 식별자와 길이가 동일하거나 실질적으로 동일합니다. 보충 핵산 서열의 농도는 식별자 라이브러리에 있는 식별자의 농도보다 낮거나, 실질적으로 동일하거나, 높을 수 있다. 보충 핵산의 농도는 식별자의 농도보다 약 1%, 10%, 20%, 40%, 60%, 80%, 100%, 125%, 150%, 175%, 200%, 1000%,  $1 \times 10^4$ %,  $1 \times 10^5$ %,  $1 \times 10^6$ %,  $1 \times 10^7$ %,  $1 \times 10^8$ % 이하보다 낮거나 동일할 수 있다. 보충 핵산의 농도는 식별자의 농도보다 약 1%, 10%, 20%, 40%, 60%, 80%, 100%, 125%, 150%, 175%, 200%, 1000%,  $1 \times 10^4$ %,  $1 \times 10^5$ %,  $1 \times 10^6$ %,  $1 \times 10^7$ %,  $1 \times 10^8$ % 이상보다 크거나 동일할 수 있다. 농도가 높을수록 데이터를 난독화하거나 숨기는 데 도움이 될 수 있다. 하나의 예에서, 보충 핵산 서열의 농도는 식별자 풀에 있는 식별자의 농도보다 실질적으로 더 높다(예를 들어,  $1 \times 10^8$ % 더 높음).

[0212] 또 다른 양태에서, 본 개시내용은 핵산 서열(들)에 인코딩된 정보를 복사하기 위한 방법을 제공한다. 핵산 서열(들)에 인코딩된 정보를 복사하기 위한 방법은 (a) 식별자 라이브러리를 제공하는 단계 및 (b) 식별자 라이브러리의 하나 이상의 복사본을 구성하는 단계를 포함할 수 있다. 식별자 라이브러리는 더 큰 조합 공간으로부터의 복수의 식별자의 서브세트를 포함할 수 있다. 복수의 식별자의 각각의 개별 식별자는 심볼의 스트링의 개별 심볼에 대응할 수 있다. 식별자는 하나 이상의 구성요소를 포함할 수 있다. 구성요소는 핵산 서열을 포함할 수 있다.

[0213] 또 다른 양태에서, 본 개시내용은 핵산 서열에 인코딩된 정보를 액세스하기 위한 방법을 제공한다. 핵산 서열에 인코딩된 정보를 액세스하기 위한 방법은 (a) 식별자 라이브러리를 제공하는 단계, 및 (b) 식별자 라이브러리로부터 식별자 라이브러리에 존재하는 식별자의 일부 또는 서브세트를 추출하는 단계를 포함할 수 있다. 식별자 라이브러리는 더 큰 조합 공간으로부터의 복수의 식별자의 서브세트를 포함할 수 있다. 복수의 식별자의 각각의

개별 식별자는 심볼의 스트링의 개별 심볼에 대응할 수 있다. 식별자는 하나 이상의 구성요소를 포함할 수 있다. 구성요소는 핵산 서열을 포함할 수 있다.

[0214] 정보는 본 문서의 다른 곳에 설명된 대로 하나 이상의 식별자 라이브러리에 기록될 수 있다. 식별자는 본 명세서의 다른 곳에 설명된 방법을 사용하여 구성될 수 있다. 저장된 데이터는 식별자 라이브러리 또는 하나 이상의 식별자 라이브러리에 개별 식별자의 복제본을 생성하여 복제할 수 있다. 식별자의 일부가 복제될 수도 있고 전체 라이브러리가 복제될 수도 있다. 복제는 식별자 라이브러리의 식별자를 증폭하여 수행할 수 있다. 하나 이상의 식별자 라이브러리가 결합될 때, 단일 식별자 라이브러리 또는 다수의 식별자 라이브러리가 복제될 수 있다. 식별자 라이브러리가 보충 핵산 서열을 포함하는 경우, 보충 핵산 서열은 복제될 수도 있고 복제되지 않을 수도 있다.

[0215] 식별자 라이브러리의 식별자는 하나 이상의 공통 프라이머 결합 부위를 포함하도록 구성될 수 있다. 하나 이상의 결합 부위는 각 식별자의 가장자리에 위치하거나 각 식별자 전체에 걸쳐 엮일 수 있다. 프라이머 결합 부위는 식별자 라이브러리 특이적 프라이머 쌍 또는 범용 프라이머 쌍이 식별자에 결합하여 증폭되도록 할 수 있다. 식별자 라이브러리 내의 모든 식별자 또는 하나 이상의 식별자 라이브러리에 있는 모든 식별자는 여러 PCR 주기에 의해 여러 번 복제될 수 있다. 전통적인 PCR이 사용되어 식별자를 복제할 수 있으며 식별자는 각 PCR 주기마다 기하급수적으로 복제될 수 있다. 식별자의 복제 수는 PCR 주기마다 기하급수적으로 증가할 수 있다. 선형 PCR은 식별자를 복제하는 데 사용될 수 있으며 식별자는 각 PCR 주기마다 선형적으로 복제될 수 있다. 식별자 복제의 수는 각 PCR 주기에 따라 선형적으로 증가할 수 있다. 식별자는 PCR 증폭 전에 원형 벡터에 결합될 수 있다. 원 벡터는 식별자 삽입 부위의 각 말단에 바코드를 포함할 수 있다. 식별자 증폭을 위한 PCR 프라이머는 바코드 가장자리가 증폭 산물의 식별자와 함께 포함되도록 벡터에 프라이밍되도록 설계될 수 있다. 증폭 중에, 식별자 간의 재조합으로 인해 각 가장자리에 상관되지 않은 바코드를 포함하는 식별자가 복제될 수 있다. 비상관 바코드는 식별자 판독 시 검출될 수 있다. 비상관 바코드를 포함하는 식별자는 위양성으로 간주될 수 있으며 정보 디코딩 프로세스 중에 무시될 수 있다. 화학적 방법 섹션 D를 참조할 수 있다.

[0216] 정보는 각 정보 비트를 고유한 핵산 분자에 할당함으로써 인코딩될 수 있다. 예를 들어, 각각 2개의 핵산 서열을 포함하는 3개의 샘플 세트(X, Y 및 Z)는 8개의 고유한 핵산 분자로 조립되어 8비트의 데이터를 인코딩할 수 있다.

[0217]  $N1 = X1Y1Z1$

[0218]  $N2 = X1Y1Z2$

[0219]  $N3 = X1Y2Z1$

[0220]  $N4 = X1Y2Z2$

[0221]  $N5 = X2Y1Z1$

[0222]  $N6 = X2Y1Z2$

[0223]  $N7 = X2Y2Z1$

[0224]  $N8 = X2Y2Z2$

[0225] 그런 다음 스트링의 각 비트가 대응하는 핵산 분자에 할당될 수 있다(예를 들어, N1은 첫 번째 비트를 특정할 수 있고, N2는 두 번째 비트를 특정할 수 있으며, N3은 세 번째 비트를 특정할 수 있는 등). 전체 비트 스트링은 '1'의 비트 값에 해당하는 핵산 분자가 조합 또는 풀에 포함되는 핵산 분자의 조합에 할당될 수 있다. 예를 들어, UTF-8 코딩에서 문자 'K'는 4개의 핵산 분자(가령, 앞선 예시에서, X1Y1Z2, X2Y1Z1, X2Y2Z1, 및 X2Y2Z2)의 존재로 인코딩될 수 있는 8비트 스트링 코드 01001011로 표시될 수 있다.

[0226] 정보는 시퀀싱이나 혼성화 분석을 통해 액세스될 수 있다. 예를 들어, 프라이머 또는 프로브는 핵산 서열의 공통 영역 또는 바코드 영역에 결합하도록 설계될 수 있다. 이는 핵산 분자의 임의 영역의 증폭을 가능하게 할 수 있다. 증폭 산물은 증폭 산물의 서열을 분석하거나 혼성화 분석을 통해 판독할 수 있다. 문자 'K'를 인코딩하는 상기의 예에서, 데이터의 전반부가 관심 대상인 경우 X1 핵산 서열의 바코드 영역에 특이적인 프라이머와 Z 세트의 공통 영역에 결합하는 프라이머가 사용되어 핵산 분자를 증폭시킬 수 있다. 이는 0100을 인코딩할 수 있는 시퀀스 Y1Z2를 반환할 수 있다. Y1 핵산 서열의 바코드 영역에 결합하는 프라이머와 Z 세트의 공통 서열에 결합하는 프라이머를 사용하여 핵산 분자를 추가로 증폭함으로써 해당 데이터의 서브스트링이 액세스될 수 있다. 이

는 서브스트링 01을 인코딩하는 Z2 핵산 서열을 반환할 수 있다. 대안으로, 시퀀싱 없이 특정 핵산 서열의 존재 여부를 체크함으로써 데이터가 액세스될 수 있다. 예를 들어, Y2 바코드에 특이적인 프라이머를 사용한 증폭은 Y2 바코드에 대한 증폭 산물을 생성할 수 있지만 Y1 바코드에 대한 증폭 산물은 생성하지 않을 수 있다. Y2 증폭 산물의 존재는 비트 값 '1'을 시그널링할 수 있다. 대안으로, Y2 증폭 산물이 없다는 것은 비트 값 '0'을 시그널링할 수 있다.

[0227] PCR 기반 방법이 사용되어 식별자 또는 핵산 샘플 풀의 데이터를 액세스하고 복제할 수 있다. 풀 또는 하이퍼-풀의 식별자 옆에 있는 공통 프라이머 결합 사이트를 사용하면, 정보를 포함하는 핵산이 쉽게 복제될 수 있다. 대안으로, 등은 증폭과 같은 다른 핵산 증폭 접근 방식을 사용하여 샘플 풀 또는 하이퍼-풀(가령, 식별자 라이브러리)에서 데이터를 쉽게 복제할 수도 있다. 핵산 증폭에 대해서는 화학적 방법 섹션 D를 참조할 수 있다. 샘플이 하이퍼-풀을 포함하는 경우 정보의 특정 서브세트(가령, 특정 바코드와 관련된 모든 핵산)은 정방향에서 식별자의 한쪽 가장자리에 특정 바코드와 결합하는 프라이머를, 역방향에서 식별자의 반대쪽 가장자리에 있는 공통 서열과 결합하는 또 다른 프라이머와 함께, 사용함으로써 액세스되고 검색될 수 있다. 다양한 판독 방법이 사용되어 인코딩된 핵산에서 정보를 가져올 수 있다, 예를 들어 마이크로어레이(또는 임의의 유형의 형광 혼성화), 디지털 PCR, 정량적 PCR(qPCR) 및 다양한 시퀀싱 플랫폼이 추가로 사용되어 인코딩된 서열을 판독하고 확장에 의해 디지털로 인코딩된 데이터를 읽을 수 있다.

[0228] 핵산 분자(가령, 식별자)에 저장된 정보를 액세스하는 것이 식별자 라이브러리 또는 식별자 풀에서 비표적 식별자의 일부를 선택적으로 제거하거나, 예를 들어 다수의 식별자 라이브러리의 풀에서 식별자 라이브러리의 모든 식별자를 선택적으로 제거함으로써 수행될 수 있다. 본 명세서에서 사용될 때, "액세스" 및 "쿼리"는 상호교환적으로 사용될 수 있다. 데이터 액세스는 식별자 라이브러리나 식별자 풀에서 대상 식별자를 선택적으로 캡처하여 수행할 수도 있다. 표적화된 식별자는 더 큰 정보 내의 관심 데이터에 대응할 수 있다. 식별자의 풀은 보충 핵산 분자를 포함할 수 있다. 보충 핵산 분자는 인코딩된 정보에 대한 메타데이터를 포함할 수 있거나 정보에 대응하는 식별자를 인코딩하거나 마스킹하는 데 사용될 수 있다. 보충 핵산 분자는 표적 식별자를 액세스하는 동안 추출될 수도 있고 추출되지 않을 수도 있다. 도 73a - 73c는 더 많은 수의 식별자 중 다수의 특정 식별자를 액세스함으로써 핵산 서열에 저장된 정보의 일부를 액세스하기 위한 예시적 방법의 개요를 개략적으로 도시한다. 도 73a는 특정된 구성요소를 포함하는 식별자를 액세스하기 위해 중합효소 연쇄 반응, 친화성 태깅된 프로브, 및 분해 표적화 프로브를 사용하는 예시적인 방법을 보여준다. PCR 기반 액세스의 경우, 식별자 풀(가령, 식별자 라이브러리)은 각 말단에 공통 서열, 각 말단에 가변 서열, 또는 각 말단에 공통 서열이나 가변 서열 중 하나를 갖는 식별자를 포함할 수 있다. 공통 서열 또는 가변 서열은 프라이머 결합 부위일 수 있다. 하나 이상의 프라이머가 식별자 가장자리의 공통 또는 가변 영역에 결합할 수 있다. 프라이머가 결합된 식별자는 PCR에 의해 증폭될 수 있다. 증폭된 식별자의 수는 증폭되지 않은 식별자보다 훨씬 더 많을 수 있다. 판독하는 동안 증폭된 식별자가 식별될 수 있다. 식별자 라이브러리로부터의 식별자는 해당 라이브러리와 구별되는 한쪽 또는 양쪽 말단 상의 서열을 포함할 수 있으므로, 단일 라이브러리가 둘 이상의 식별자 라이브러리 그룹이나 풀에서 선택적으로 액세스될 수 있다.

[0229] 친화성-태그 기반 액세스를 위해, 핵산 포착으로 지칭될 수 있는 프로세스의 경우, 풀의 식별자를 구성하는 구성요소는 하나 이상의 프로브와 상보성을 공유할 수 있다. 하나 이상의 프로브는 액세스될 식별자에 결합하거나 혼성화할 수 있다. 프로브는 친화성 태그를 포함할 수 있다. 친화성 태그는 고체-상 기관, 가령, 막, 웰, 컬럼 또는 비드 상에 포획될 수 있다. 고체상 기질로서 비드를 사용하는 경우, 친화성 태그는 비드에 결합하여 비드, 적어도 하나의 프로브 및 적어도 하나의 식별자를 포함하는 복합체를 생성할 수 있다. 비드는 자석일 수 있으며 자석과 함께 액세스할 식별자를 수집하고 격리할 수 있다. 판독하기 전에 변성 조건 하에서 식별자가 비드에서 제거될 수 있다. 대안으로 또는 추가로, 비드는 비표적 식별자를 수집하고 이를 별도의 용기로 세척하여 판독할 수 있는 풀의 나머지 부분으로부터 분리할 수 있다. 컬럼을 사용할 때 친화성 태그가 컬럼에 결합될 수 있다. 액세스될 식별자는 포착을 위해 컬럼에 결합될 수 있다. 컬럼 경계 식별자는 판독 전에 컬럼으로부터 용출되거나 변성될 수 있다. 대안으로, 비표적 식별자는 선택적으로 컬럼에 표적화될 수 있는 반면 표적 식별자는 컬럼을 통과해 유동할 수 있다. 고체상 기질에 결합된 식별자는 예를 들어 산, 염기, 산화, 환원, 열, 빛, 금속 이온 촉매 작용, 치환 또는 제거 화학과 같은 조건에 노출시킴으로써 또는 효소 절단에 의해 고체상 기질에서 제거될 수 있다. 특정 구현예에서, 액세스될 식별자는 절단 가능한 연결 모이어티를 통해 고체 지지체에 부착될 수 있다. 예를 들어, 고체상 기질은 표적 식별자에 대한 공유 부착을 위한 절단 가능한 링커를 제공하도록 기능화될 수 있다. 링커 모이어티는 길이가 6개 이상의 원자일 수 있다. 일부 구현예에서, 절단 가능한 링커는 TOPS(합성당 2개의 올리고뉴클레오타이드) 링커, 아미노 링커, 화학적으로 절단 가능한 링커, 또는 광절단 가능한 링커일 수 있다. 표적화된 식별자를 액세스하는 것은 하나 이상의 프로브를 식별자 풀에 동시에 적용하거나 하

나 이상의 프로브를 식별자 풀에 순차적으로 적용하는 것을 포함할 수 있다. 핵산 포획에 대해서는 화학적 방법 섹션 F를 참조할 수 있다.

[0230] 분해 기반 액세스의 경우, 풀의 식별자를 구성하는 구성요소는 하나 이상의 분해 표적화 프로브와 상보성을 공유할 수 있다. 프로브는 식별자의 개별 구성요소에 결합하거나 혼성화할 수 있다. 프로브는 엔도뉴클레아제와 같은 분해 효소에 대한 표적이 될 수 있다. 예를 들어, 하나 이상의 식별자 라이브러리가 조합될 수 있다. 프로브의 세트는 식별자 라이브러리 중 하나와 혼성화될 수 있다. 프로브의 세트는 RNA를 포함할 수 있고, RNA는 Cas9 효소를 안내할 수 있다. Cas9 효소는 하나 이상의 식별자 라이브러리에 도입될 수 있다. 프로브와 혼성화된 식별자는 Cas9 효소에 의해 분해될 수 있다. 액세스된 식별자는 분해 효소에 의해 분해되지 않을 수 있다. 또 다른 예에서, 식별자는 단일 가닥일 수 있고 식별자 라이브러리는 액세스되지 않는 식별자를 선택적으로 분해하는 S1 뉴클레아제와 같은 단일 가닥 특이적 엔도뉴클레아제(들)와 결합될 수 있다. 액세스된 식별자는 단일 가닥 특이적 엔도뉴클레아제(들)에 의한 분해로부터 보호하기 위해 상보적인 식별자 세트와 혼성화될 수 있다. 액세스할 식별자는 크기 선택 크로마토그래피(가령, 아가로스 겔 전기영동)와 같은 크기 선택을 통해 분해 산물로부터 분리될 수 있다. 대안으로 또는 추가로, 분해되지 않은 식별자는 분해 산물이 증폭되지 않도록 선택적으로 증폭(가령, PCR을 사용하여)될 수 있다. 분해되지 않은 식별자는 분해되지 않은 식별자의 각 말단에 혼성화되므로 분해되거나 절단된 식별자의 각 말단에는 혼성화되지 않는 프라이머를 사용하여 증폭될 수 있다.

[0231] 도 73b는 다중 구성요소를 포함하는 식별자를 액세스하기 위해 'OR' 또는 'AND' 연산을 수행하기 위해 중합효소 연쇄 반응을 사용하는 예시적인 방법을 보여준다. 예를 들어, 두 개의 정방향 프라이머가 왼쪽 말단 상에 식별자의 개별 세트를 결합하는 경우, 이러한 식별자 세트의 결합에 대한 'OR' 증폭은 오른쪽 말단 상의 모든 식별자를 결합하는 역방향 프라이머를 갖는 다중 PCR 반응에서 두 개의 정방향 프라이머를 함께 사용함으로써 달성될 수 있다. 다른 예에서, 하나의 정방향 프라이머가 왼쪽 말단에 있는 식별자의 세트와 결합하고 하나의 역방향 프라이머가 오른쪽 말단에 있는 식별자 세트와 결합하는 경우, 두 식별자 세트의 교차점에 대한 'AND' 증폭은, PCR 반응에서 정방향 프라이머와 역방향 프라이머를 함께 프라이머 쌍으로 사용함으로써 이뤄질 수 있다.

[0232] 도 73c는 다중 구성요소를 포함하는 식별자를 액세스하기 위해 'OR' 또는 'AND' 연산을 수행하기 위해 친화성 태그를 사용하는 예시적인 방법을 도시한다. 예를 들어, 친화성 프로브 'P1'이 구성요소 'C1'를 갖는 모든 식별자를 포착하고 다른 친화성 프로브 'P2'가 구성요소 'C2'를 갖는 모든 식별자를 포착하는 경우, C1 또는 C2를 갖는 모든 식별자의 세트는 ('OR' 연산에 대응하는) P1 및 P2를 동시에 사용함으로써 포착될 수 있다. 동일한 구성요소와 프로브를 사용하는 또 다른 예에서 C1 및 C2를 갖는 모든 식별자의 세트는 ('AND' 연산에 대응하는) P1와 P2를 순차적으로 사용함으로써 캡처될 수 있다.

[0233] 또 다른 양태에서, 본 개시내용은 핵산 서열에 코딩된 정보를 판독하기 위한 방법을 제공한다. 핵산 서열에 인코딩된 정보를 판독하기 위한 방법은 (a) 식별자 라이브러리를 제공하는 단계, (b) 식별자 라이브러리에 존재하는 식별자를 식별하는 단계, (c) 식별자 라이브러리에 존재하는 식별자로부터 심볼의 스트링을 생성하는 단계 및 (d) 심볼의 스트링으로부터 정보를 컴파일하는 단계를 포함할 수 있다. 식별자 라이브러리는 조합 공간으로부터의 복수의 식별자의 서브세트를 포함할 수 있다. 식별자의 서브세트의 각각의 개별 식별자는 심볼의 스트링 내 개별 심볼에 대응할 수 있다. 식별자는 하나 이상의 구성요소를 포함할 수 있다. 구성요소는 핵산 서열을 포함할 수 있다.

[0234] 정보는 본 문서의 다른 곳에 설명된 대로 하나 이상의 식별자 라이브러리에 기록될 수 있다. 식별자는 본 명세서의 다른 곳에 설명된 방법을 사용하여 구성될 수 있다. 저장된 데이터는 본 문서의 다른 곳에서 설명한 방법을 사용하여 복사되고 액세스될 수 있다.

[0235] 식별자는 인코딩된 심볼의 위치, 인코딩된 심볼의 값, 또는 인코딩된 심볼의 위치와 값 모두에 관한 정보를 포함할 수 있다. 식별자는 인코딩된 심볼의 위치와 관련된 정보를 포함할 수 있으며 식별자 라이브러리 내 식별자가 존재 또는 부재는 심볼의 값을 나타낼 수 있다. 식별자 라이브러리 내 식별자의 존재는 이진 스트링 내 첫 번째 심볼 값(가령, 제1 비트 값)을 나타낼 수 있고 식별자 라이브러리 내 식별자의 부재는 이진 스트링 내 두 번째 심볼 값(가령, 두 번째 비트 값)을 나타낼 수 있다. 이진 시스템에서, 비트 값을 식별자 라이브러리 내 식별자의 존재 또는 부재에 기초하는 것은 조립된 식별자의 수를 감소시킬 수 있고, 따라서 기록 시간을 감소시킬 수 있다. 예를 들어, 식별자의 존재는 매핑된 위치에서의 비트 값 '1'을 나타낼 수 있고, 식별자의 부재는 매핑된 위치에서의 비트 값 '0'을 나타낼 수 있다.

[0236] 정보에 대한 심볼(가령, 비트 값)을 생성하는 것은 심볼(가령, 비트)이 매핑되거나 인코딩될 수 있는 식별자의 존재 또는 부재를 식별하는 것을 포함할 수 있다. 식별자의 존재 또는 부재를 결정하는 것은 존재하는 식별자를

시퀀싱하거나 혼성화 어레이를 사용하여 식별자의 존재를 검출하는 것을 포함할 수 있다. 예에서, 인코딩된 서열을 디코딩하고 판독하는 것은 시퀀싱 플랫폼을 사용하여 수행될 수 있다. 시퀀싱 플랫폼의 예시가 그 전체가 본 명세서에 참조로서 포함되는 2014년08월21일에 출원된 미국 특허 출원 번호 14/465,685이자 2014년12월18일로 공개된 미국 특허 공개 번호 2014-0371100 A1인 발명의 명칭 "METHOD OF NUCLEIC ACID AMPLIFICATION", 2013년05월02일에 출원된 미국 특허 출원 번호 13/886,234이자 2013년09월05일에 공개된 미국 특허 공개 번호 2013-0231254 A1인 발명의 명칭 "METHOD OF NUCLEIC ACID AMPLIFICATION", 및 2009년03월09일에 출원된 미국 특허 출원 번호 12/400,593이자 2009년10월08일에 공개된 미국 특허 번호 US 2009-0253141 A1인 발명의 명칭 "METHODS AND APPARATUSES FOR ANALYZING POLYNUCLEOTIDE SEQUENCES"에 기재되어 있다.

[0237] 하나의 예에서, 핵산 인코딩 데이터를 디코딩하는 것은 핵산 가닥의 염기별 시퀀싱, 가령, Illumina® 시퀀싱에 의해, 또는 특정 핵산 서열의 존재 또는 부재를 나타내는 시퀀싱 기법, 모세관 전기영동에 의한 단편화 분석을 사용함으로써, 달성될 수 있다. 시퀀싱은 가역적 종결자(reversible terminator)의 사용을 채용할 수 있다. 시퀀싱은 자연 또는 비자연(예를 들어, 조작된) 뉴클레오티드 또는 뉴클레오티드 유사체의 사용을 채용할 수 있다. 대안으로 또는 추가로, 핵산 서열을 디코딩하는 것은 다양한 분석 기법, 비제한적 예를 들면, 광학적, 전기화학적, 또는 화학적 신호를 생성하는 임의의 방법을 사용하여 수행될 수 있다. 다양한 시퀀싱 방식, 비제한적 예를 들면, 중합효소 연쇄반응(PCR), 디지털 PCR, Sanger 시퀀싱, 고처리량 시퀀싱, 합성별 시퀀싱, 단일 분자 시퀀싱, 결합별 시퀀싱, RNA-Seq(Illumina), 차세대 시퀀싱, 디지털 유전자 발현(Helicos), Clonal Single MicroArray(Solexa), 샷건 시퀀싱, Maxim-Gilbert 시퀀싱 또는 대규모 병렬 시퀀싱이 사용될 수 있다.

[0238] 다양한 판독 방법이 사용되어 인코딩된 핵산에서 정보를 가져올 수 있다. 예를 들어, 마이크로어레이(또는 모든 종류의 형광 혼성화), 디지털 PCR, 정량적 PCR(qPCR) 및 다양한 시퀀싱 플랫폼이 추가로 사용되어 인코딩된 서열을 판독하고 더 나아가 디지털로 인코딩된 데이터를 판독할 수 있다.

[0239] 식별자 라이브러리는 정보에 관한 메타데이터를 제공하거나, 정보를 암호화하거나 마스킹하거나, 메타데이터를 제공하고 정보를 마스킹하는 보충 핵산 서열을 더 포함할 수 있다. 보충 핵산은 식별자의 식별과 동시에 식별될 수 있다. 대안으로, 보충 핵산은 식별자를 식별하기 전이나 후에 식별될 수 있다. 예를 들어, 인코딩된 정보를 판독하는 동안 보충 핵산이 식별되지 않는다. 보충 핵산 서열은 식별자와 구별되지 않을 수 있다. 식별자 인덱스 또는 키가 사용되어 식별자와 보충 핵산 분자를 구별할 수 있다.

[0240] 입력 비트 스트링을 재코딩하여 더 적은 수의 핵산 분자를 사용함으로써 데이터 인코딩 및 디코딩의 효율성을 높일 수 있다. 예를 들어, 인코딩 방법에 의해 3개의 핵산 분자(가령, 식별자)에 매핑될 수 있는 '111' 서브스트링의 발생률이 높은 입력 스트링이 수신되는 경우, 핵산 분자의 널(null) 세트에 매핑될 수 있는 '000' 서브스트링으로 재코딩될 수 있다. '000'의 대체 입력 서브스트링이 또한 '111'로 재코딩될 수 있다. 이 재코딩 방법은 데이터 세트에서 '1'의 수가 감소할 수 있으므로 데이터를 인코딩하는 데 사용되는 핵산 분자의 총량을 줄일 수 있다. 이 예에서, 새로운 매핑 지침을 지정하는 코드북을 수용하기 위해 데이터세트의 전체 크기가 증가될 수 있다. 인코딩 및 디코딩 효율성을 높이는 또 다른 방법은 입력 스트링을 재코딩하여 가변 길이를 줄이는 것일 수 있다. 예를 들어, '111'은 '00'으로 재코딩될 수 있으며, 이는 데이터세트의 크기를 축소하고 데이터세트에서 '1'의 수를 줄일 수 있다.

[0241] 핵산 인코딩된 데이터를 디코딩하는 속도 및 효율성은 검출 용이성을 위해 식별자를 구체적으로 설계함으로써 제어(가령, 증가)될 수 있다. 예를 들어, 검출 용이성을 위해 설계된 핵산 서열(가령, 식별자)은 광학적, 전기화학적, 화학적 또는 물리적 특성을 기반으로 쿨 및 검출이 더 쉬운 뉴클레오티드의 대부분을 포함하는 핵산 서열을 포함할 수 있다. 조작된 핵산 서열은 단일 가닥 또는 이중 가닥일 수 있다. 조작된 핵산 서열은 핵산 서열의 검출 가능한 특성을 개선하는 합성 또는 비천연 뉴클레오티드를 포함할 수 있다. 조작된 핵산 서열은 모든 천연 뉴클레오티드, 모든 합성 또는 비천연 뉴클레오티드, 또는 천연, 합성 및 비천연 뉴클레오티드의 조합을 포함할 수 있다. 합성 뉴클레오티드는 뉴클레오티드 유사체, 가령, 펩티드 핵산, 잠금 핵산, 글리콜 핵산 및 트레오스 핵산을 포함할 수 있다. 비천연 뉴클레오티드는 3-메톡시-2-나프탈기를 함유한 인공 뉴클레오시드인 dNm 및 6-메틸이소퀴놀린-1-티온-2-일기를 함유한 인공 뉴클레오시드인 d5SICS를 포함할 수 있다. 조작된 핵산 서열은 강화된 광학 특성과 같은 단일 강화 특성을 위해 설계될 수 있거나, 설계된 핵산 서열은 강화된 광학적 및 전기화학적 특성 또는 강화된 광학적 및 화학적 특성과 같은 다중 강화된 특성으로 설계될 수 있다. DNA 설계에 대한 화학적 방법 섹션 H를 참조할 수 있다.

[0242] 조작된 핵산 서열은 핵산 서열의 광학적, 전기화학적, 화학적 또는 물리적 특성을 개선하지 않는 반응성 천연, 합성 및 비천연 뉴클레오티드를 포함할 수 있다. 핵산 서열의 반응성 구성요소는 핵산 서열에 개선된 특성을 부

여하는 화학적 잔기의 첨가를 가능하게 할 수 있다. 각각의 핵산 서열은 단일 화학적 부분을 포함할 수 있거나 다수의 화학적 부분을 포함할 수 있다. 예시적인 화학적 부분은 형광성 잔기, 화학발광성 잔기, 산성 또는 염기성 잔기, 소수성 또는 친수성 잔기, 및 핵산 서열의 산화 상태 또는 반응성을 변경하는 잔기가 포함될 수 있으나 이에 제한되지는 않는다.

[0243] 시퀀싱 플랫폼은 핵산 서열로 인코딩된 정보를 디코딩하고 판독하기 위해 특별히 설계될 수 있다. 시퀀싱 플랫폼은 단일 또는 이중 가닥 핵산 분자의 시퀀싱 전용일 수 있다. 시퀀싱 플랫폼은 개별 염기를 판독함으로써(가령, 염기별 시퀀싱) 또는 핵산 분자(가령, 식별자) 내에 통합된 전체 핵산 서열(가령, 구성요소)의 존재 또는 부재를 검출함으로써 핵산 인코딩된 데이터를 디코딩할 수 있다. 시퀀싱 플랫폼은 난잡한 시약의 사용, 리드(read) 길이의 증가, 검출 가능한 화학적 잔기의 추가에 의한 특정 핵산 서열의 검출을 포함할 수 있다. 시퀀싱 중에 더 난잡한 시약을 사용하면 더 빠른 염기 호출을 활성화하여 판독 효율성을 높일 수 있으며 결과적으로 시퀀싱 시간이 줄어들 수 있다. 증가된 리드 길이의 사용은 리드당 디코딩될 인코딩된 핵산의 더 긴 서열을 가능하게 할 수 있다. 검출 가능한 화학적 잔기 태그의 첨가는 화학적 잔기의 존재 또는 부재에 의해 핵산 서열의 존재 또는 부재의 검출을 가능하게 할 수 있다. 예를 들어, 정보 비트를 인코딩하는 각 핵산 서열에는 고유한 광학적, 전기화학적 또는 화학적 신호를 생성하는 화학적 부분이 태그로 지정될 수 있다. 해당 고유한 광학적, 전기화학적 또는 화학적 신호의 존재 여부는 '0' 또는 '1' 비트 값을 나타낼 수 있다. 핵산 서열은 단일 화학적 잔기 또는 다중 화학적 잔기를 포함할 수 있다. 화학적 잔기는 데이터를 인코딩하기 위해 핵산 서열을 사용하기 전에 핵산 서열에 첨가될 수 있다. 대안으로 또는 추가로, 화학적 잔기는 데이터를 인코딩한 후, 그러나 데이터를 디코딩하기 전에 핵산 서열에 추가될 수 있다. 화학적 잔기 태그는 핵산 서열에 직접 추가될 수 있거나, 핵산 서열은 합성 또는 비천연 뉴클레오타이드 앵커를 포함할 수 있고 화학적 부분 태그는 해당 앵커에 추가될 수 있다.

[0244] 인코딩 및 디코딩 오류를 최소화하거나 검출하기 위해 고유 코드가 적용될 수 있다. 인코딩 및 디코딩 오류는 위음성(가령, 무작위 샘플링에 포함되지 않은 핵산 분자 또는 식별자)으로 인해 발생할 수 있다. 오류 검출 코드의 예는 식별자 라이브러리에 포함된 연속 가능한 식별자 세트의 식별자 수를 계산하는 체크섬 서열일 수 있다. 식별자 라이브러리를 읽는 동안 체크섬은 연속된 식별자 집합에서 검색할 것으로 예상되는 식별자 수를 나타낼 수 있으며, 예상 개수가 충족될 때까지 읽기를 위해 식별자를 계속 샘플링할 수 있다. 일부 실시예에서, 체크섬 시퀀스는 R개의 식별자의 모든 연속 세트에 대해 포함될 수 있으며, 여기서 R은 크기가 동일하거나 1, 2, 5, 10, 50, 100, 200, 500 또는 1000보다 크거나 1000, 500, 200, 100, 50, 10, 5 또는 2보다 작을 수 있다. R의 값이 작을수록 오류 검출 성능이 향상된다. 일부 실시예에서, 체크섬은 보충 핵산 서열일 수 있다. 예를 들어, 7개의 핵산 서열(가령, 구성요소)을 포함하는 세트는 두 그룹, 즉, 곱 방식에 의한 식별자를 구성하기 위한 핵산 서열(총 X의 구성요소 X1-X3 및 총 Y의 Y1-Y3) 및 보충 체크섬에 대한 핵산 서열(X4-X7 및 Y4-Y7)로 나뉠 수 있다. 체크섬 서열 X4-X7은 총 X의 0개, 1개, 2개 또는 3개의 서열이 총 Y의 각 멤버와 조립되는지 여부를 나타낼 수 있다. 대안으로, 체크섬 서열 Y4-Y7은 총 Y의 0개, 1개, 2개 또는 3개의 서열이 총 X의 각 멤버와 조립되는지 여부를 나타낼 수 있다. 이 예에서, 식별자 {X1Y1, X1Y3, X2Y1, X2Y2, X2Y3}를 갖는 원본 식별자 라이브러리가 체크섬을 포함하도록 보완되어 다음의 풀이 될 수 있다: {X1Y1, X1Y3, X2Y1, X2Y2, X2Y3, X1Y6, X2Y7, X3Y4, X6Y1, X5Y2, X6Y3}. 체크섬 서열은 오류 정정에도 사용될 수 있다. 예를 들어, 위의 데이터 세트에서 X1Y1이 없고 X1Y6 및 X6Y1이 있으면 X1Y1 핵산 분자가 데이터 세트에 없다는 추론이 가능해진다. 체크섬 서열은 식별자 라이브러리의 샘플링 또는 식별자 라이브러리의 액세스된 부분에서 식별자가 누락되었는지 여부를 나타낼 수 있다. 체크섬 서열이 누락된 경우 PCR 또는 친화성 태깅된 프로브 혼성화와 같은 액세스 방법을 통해 이를 증폭 및/또는 분리할 수 있다. 일부 실시예에서, 체크섬은 보충 핵산 서열이 아닐 수도 있다. 체크섬은 식별자로 표현되도록 정보에 직접 코딩될 수 있다.

[0245] 예를 들어 곱 방식에서 단일 구성요소가 아닌 구성요소의 회문 쌍을 사용하여 식별자를 회문식으로 구성하면 데이터 인코딩 및 디코딩의 노이즈가 줄어들 수 있다. 그런 다음, 상이한 층으로부터의 구성요소의 쌍은 회문 방식(가령, 구성요소 X 및 Y에 대해 XY 대신 YXY)으로 서로 조립될 수 있다. 이 회문 방법은 더 많은 수의 층(가령 XYZ 대신 ZYXYZ)로 확장될 수 있으며 식별자들 간의 잘못된 교차 반응을 감지할 수 있다.

[0246] 식별자에 과잉(예를 들어, 엄청난 과잉)의 보충 핵산 서열을 추가하면 시퀀싱이 인코딩된 식별자를 복구하는 것을 방지할 수 있다. 정보를 디코딩하기 전에, 식별자는 보충 핵산 서열로부터 강화될 수 있다. 예를 들어, 식별자 말단에 특이적인 프라이머를 사용하는 핵산 증폭 반응에 의해 식별자가 강화될 수 있다. 대안으로, 또는 추가로, 특정 프라이머를 사용하는 시퀀싱(가령, 합성에 의한 시퀀싱)을 통해 샘플 풀을 강화하지 않고도 정보를 디코딩할 수 있다. 두 가지 디코딩 방법 모두, 디코딩 키가 없거나 식별자 구성에 대해 알지 못하면 정보를 강

화하거나 디코딩하는 것이 어려울 수 있다. 친화성 태그 기반 프로브를 사용하는 것과 같은 대체 접근 방법도 사용될 수 있다.

- [0247] 디지털 정보를 핵산(가령, DNA)으로 인코딩하기 위한 시스템은 파일 및 데이터(가령, 미가공 데이터, 압축된 zip 파일, 정수 데이터 및 그 밖의 다른 형태의 데이터)를 바이트로 변환하고 바이트를 핵산, 통상 DNA, 또는 이들의 조합의 세그먼트 또는 서열로 인코딩하기 위한 시스템, 방법 및 장치를 포함할 수 있다.
- [0248] 하나의 양태에서, 본 개시는 핵산을 사용하여 바이너리 서열 데이터를 인코딩하기 위한 시스템을 제공한다. 핵산을 사용하여 바이너리 서열 데이터를 인코딩하기 위한 시스템은 장치 및 하나 이상의 컴퓨터 프로세서를 포함할 수 있다. 장치는 식별자 라이브러리를 구성하도록 구성될 수 있다. 하나 이상의 컴퓨터 프로세서는 (i) 정보를 심볼의 스트링으로 변환하고, (ii) 심볼의 스트링을 복수의 식별자로 매핑하며, (iii) 적어도 복수의 식별자의 서브세트를 포함하는 식별자 라이브러리를 구성하도록 개별적 또는 집합적으로 프로그램될 수 있다. 복수의 식별자 중 개별 식별자는 심볼의 스트링의 개별 심볼에 대응될 수 있다. 복수의 식별자 중 개별 식별자는 하나 이상의 구성요소를 포함할 수 있다. 하나 이상의 구성요소의 개별 구성요소는 핵산 서열을 포함할 수 있다.
- [0249] 다른 양태에서, 본 개시는 핵산을 사용하여 이진 서열 데이터를 판독하기 위한 시스템을 제공한다. 핵산을 사용하여 이진 서열 데이터를 판독하기 위한 시스템은 데이터베이스 및 하나 이상의 컴퓨터 프로세서를 포함할 수 있다. 데이터베이스는 정보를 인코딩하는 식별자 라이브러리를 저장할 수 있다. 하나 이상의 컴퓨터 프로세서는 (i) 식별자 라이브러리 내 식별자를 식별하고, (ii) (i)에서 식별된 식별자로부터 복수의 심볼을 생성하며, (iii) 복수의 심볼로부터 정보를 컴파일하도록 개별적으로 또는 집합적으로 프로그램될 수 있다. 식별자 라이브러리는 복수의 식별자의 서브세트를 포함할 수 있다. 복수의 식별자의 각각의 개별 식별자는 심볼의 스트링의 개별 심볼에 대응할 수 있다. 식별자는 하나 이상의 구성요소를 포함할 수 있다. 구성요소는 핵산 서열을 포함할 수 있다.
- [0250] 디지털 데이터를 인코딩하기 위해 시스템을 사용하는 방법의 비제한적인 실시예는 바이트 스트림의 형태로 디지털 정보를 수신하기 위한 단계를 포함할 수 있다. 바이트 스트림을 개별 바이트로 파싱(parsing)하고, 핵산 인덱스(또는 식별자 순위)를 사용하여 바이트 내의 비트 위치를 매핑하고, 비트 값 1 또는 비트 값 0에 대응하는 서열을 식별자로 인코딩하는 단계. 디지털 데이터를 검색하기 위한 단계는 하나 이상의 비트에 매핑되는 핵산 샘플 또는 핵산 서열(가령, 식별자)을 포함하는 핵산 풀을 시퀀싱하고, 식별자 순위를 참조하여 식별자가 핵산 풀에 존재하는지 여부를 확인하고 각각의 서열에 대한 위치 및 비트-값 정보를 디지털 정보의 서열을 포함하는 바이트로 디코딩하는 것을 포함할 수 있다.
- [0251] 핵산 분자에 인코딩 및 기록된 정보를 인코딩, 기록, 복사, 액세스, 판독 및 디코딩하기 위한 시스템은 단일 통합 장치일 수 있거나 앞서 언급한 작업 중 하나 이상을 실행하도록 구성된 다중 장치일 수 있다. 정보를 핵산 분자(가령 식별자)로 인코딩하고 기록하기 위한 시스템은 장치와 하나 이상의 컴퓨터 프로세서를 포함할 수 있다. 하나 이상의 컴퓨터 프로세서는 정보를 심볼의 스트링(가령, 비트의 스트링)으로 파싱하도록 프로그램될 수 있다. 컴퓨터 프로세서는 식별자 순위를 생성할 수 있다. 컴퓨터 프로세서는 심볼을 두 개 이상의 카테고리로 분류할 수 있다. 하나의 카테고리는 식별자 라이브러리에 해당 식별자가 있음을 나타내는 심볼을 포함하고, 다른 카테고리는 식별자 라이브러리에 해당 식별자가 없음을 나타내는 심볼을 포함할 수 있다. 컴퓨터 프로세서는 식별자 라이브러리에 식별자가 존재하면 표시될 심볼에 대응하는 식별자를 조립하도록 장치에 지시할 수 있다.
- [0252] 장치는 복수의 영역, 섹션 또는 파티션을 포함할 수 있다. 식별자를 조립하기 위한 시약 및 구성요소가 장치의 하나 이상의 영역, 섹션 또는 파티션에 저장될 수 있다. 층은 장치 섹션의 별도 영역에 저장될 수 있다. 층은 하나 이상의 고유 구성요소를 포함할 수 있다. 한 층 내 구성요소는 다른 층 내 구성요소에 비해 고유할 수 있다. 영역 또는 섹션은 베셀(vessel)을 포함할 수 있고 파티션은 웰(well)을 포함할 수 있다. 각 층은 별도의 베셀 또는 파티션에 저장될 수 있다. 각 시약 또는 핵산 서열은 별도의 베셀 또는 파티션에 저장될 수 있다. 대안으로 또는 추가로 시약을 결합하여 식별자 구성을 위한 마스터 믹스를 형성할 수도 있다. 장치는 장치의 한 섹션에서 시약, 구성요소 및 주형을 전달하여 다른 섹션에 결합할 수 있다. 장치는 조립 반응을 완료하기 위한 조건을 제공할 수 있다. 예를 들어, 장치는 가열, 교반 및 반응 진행 감지 기능을 제공할 수 있다. 구성된 식별자는 식별자의 하나 이상의 말단에 바코드, 공통 서열, 가변 서열 또는 태그를 추가하기 위해 하나 이상의 후속 반응을 거치도록 지시될 수 있다. 그런 다음 식별자는 영역이나 파티션으로 전달되어 식별자 라이브러리를 생성할 수 있다. 하나 이상의 식별자 라이브러리가 장치의 각 영역, 섹션 또는 개별 파티션에 저장될 수 있다. 장치는 압력, 진공 또는 흡입을 사용하여 유체(가령, 시약, 구성요소, 주형)를 전달할 수 있다.
- [0253] 식별자 라이브러리는 장치에 저장되거나 별도의 데이터베이스로 이동될 수 있다. 데이터베이스는 하나 이상의

식별자 라이브러리를 포함할 수 있다. 데이터베이스는 식별자 라이브러리의 장기 저장을 위한 조건(가령, 식별자의 열화를 줄이기 위한 조건)을 제공할 수 있다. 식별자 라이브러리는 분말, 액체 또는 고체 형태로 저장될 수 있다. 보다 안정적인 보관을 위해 식별자의 수용액을 동결건조할 수 있다(동결건조에 대한 자세한 내용은 화학적 방법 섹션 G 참조). 대안으로, 식별자는 산소가 없는 상태(가령, 혐기성 보관 조건)에 보관될 수 있다. 데이터베이스는 자외선 차단, 온도 감소(가령, 냉장 또는 냉동), 분해되는 화학물질 및 효소로부터의 보호 기능을 제공할 수 있다. 데이터베이스로 전송되기 전에 식별자 라이브러리를 동결건조하거나 냉동할 수 있다. 식별자 라이브러리는 뉴클레아제를 불활성화하기 위한 EDTA(에틸렌디아민테트라아세트산) 및/또는 핵산 분자의 안정성을 유지하기 위한 버퍼액을 포함할 수 있다.

[0254] 데이터베이스는 정보를 식별자에 기록하거나, 정보를 복사하거나, 정보에 액세스하거나, 정보를 읽는 장치에 연결되거나, 포함되거나, 분리될 수 있다. 식별자 라이브러리의 일부는 복사, 액세스 또는 판독 전에 데이터베이스로부터 제거될 수 있다. 데이터베이스로부터 정보를 복사하는 장치는 정보를 기록하는 장치와 동일하거나 다를 수 있다. 정보를 복사하는 장치는 장치에서 식별자 라이브러리의 부분표본을 추출하고 해당 부분표본을 시약 및 구성요소와 결합하여 식별자 라이브러리의 일부 또는 전체를 증폭할 수 있다. 장치는 증폭 반응의 온도, 압력 및 교반을 제어할 수 있다. 장치는 구획을 포함할 수 있으며, 식별자 라이브러리를 포함하는 구획에서 하나 이상의 증폭 반응이 일어날 수 있다. 장치는 한 번에 둘 이상의 식별자 풀을 복사할 수 있다.

[0255] 복사된 식별자는 복사 장치에서 액세스 장치로 전송될 수 있다. 액세스 장치는 복사 장치와 동일한 장치일 수 있다. 액세스 장치는 별도의 영역, 섹션 또는 파티션을 포함할 수 있다. 액세스 장치는 친화성 태그에 결합된 식별자를 분리하기 위한 하나 이상의 컬럼, 비드 저장소 또는 자기 영역을 가질 수 있다(핵산 포획에 관한 화학적 방법 섹션 F 참조). 대안으로 또는 추가로, 액세스 장치는 하나 이상의 크기 선택 유닛을 가질 수 있다. 크기 선택 유닛은 아가로스 겔 전기영동 또는 핵산 분자의 크기 선택을 위한 임의의 다른 방법을 포함할 수 있다(핵산 크기 선택에 대한 자세한 내용은 화학적 방법 섹션 E 참조). 복사 및 추출은 장치의 동일한 영역 또는 장치의 상이한 영역에서 수행될 수 있다(핵산 증폭에 대해서는 화학적 방법 섹션 D 참조).

[0256] 액세스된 데이터는 동일한 장치에서 읽힐 수도 있고, 액세스된 데이터가 다른 장치로 전송될 수도 있다. 판독 장치는 식별자를 검출하고 식별하기 위한 검출 유닛을 포함할 수 있다. 검출 유닛은 시퀀서, 혼성화 어레이, 또는 식별자의 존재 또는 부재를 식별하기 위한 그 밖의 다른 유닛의 일부일 수 있다. 시퀀싱 플랫폼은 핵산 서열로 인코딩된 정보를 디코딩하고 판독하기 위해 특별히 설계될 수 있다. 시퀀싱 플랫폼은 단일 또는 이중 가닥 핵산 분자의 시퀀싱 전용일 수 있다. 시퀀싱 플랫폼은 개별 염기를 판독함으로써(가령, 염기별 시퀀싱) 또는 핵산 분자(가령, 식별자) 내에 통합된 전체 핵산 서열(가령, 구성요소)의 존재 또는 부재를 검출함으로써 핵산 인코딩된 데이터를 디코딩할 수 있다. 대안으로, 시퀀싱 플랫폼은 Illumina® 시퀀싱 또는 모세관 전기영동에 의한 단편화 분석과 같은 시스템일 수 있다. 대안으로 또는 추가로, 핵산 서열의 디코딩은 장치에 의해 구현되는 다양한 분석 기술을 사용하여 수행될 수 있으며, 여기에는 광학적, 전기화학적 또는 화학적 신호를 생성하는 모든 방법이 포함되지만 이에 국한되지는 않는다.

[0257] 핵산 분자의 정보 저장은 장기 정보 저장, 민감한 정보 저장 및 의료 정보 저장을 포함하되 이에 국한되지 않는 다양한 응용 분야를 가질 수 있다. 예를 들어, 개인의 의료 정보(가령, 병력 및 기록)가 핵산 분자에 저장되어 개인에게 전달될 수 있다. 정보는 신체 외부(가령, 웨어러블 장치)에 저장되거나 신체 내부(가령, 피하 캡슐)에 저장될 수 있다. 환자가 진료실이나 병원에 입원하면 장치나 캡슐에서 샘플을 채취하고 핵산 서열 분석기를 사용하여 정보를 해독할 수 있다. 의료 기록을 핵산 분자로 개인별로 저장하는 것은 컴퓨터 및 클라우드 기반 저장 시스템에 대한 대안을 제공할 수 있다. 개인의 의료 기록을 핵산 분자로 저장하면 의료 기록이 해킹당하는 사례나 빈도가 줄어들 수 있다. 의료 기록의 캡슐 기반 저장에 사용되는 핵산 분자는 인간 게놈 서열에서 유래될 수 있다. 인간 게놈 서열의 사용은 캡슐 고장 및 누출의 경우 핵산 서열의 면역원성을 감소시킬 수 있다.

[0258] 본 개시는 본 개시의 방법을 구현하도록 프로그래밍된 컴퓨터 시스템을 제공한다. 도 75는 디지털 정보를 핵산 서열로 인코딩하고/하거나 핵산 서열로부터 유래된 정보를 판독(예를 들어, 디코딩)하도록 프로그래밍되거나 달리 구성된 컴퓨터 시스템(1901)을 도시한다. 컴퓨터 시스템(1901)은 예를 들어 인코딩된 비트스트림 또는 바이트 스트림으로부터 주어진 비트 또는 바이트에 대한 비트 값 및 비트 위치 정보와 같은 본 개시의 인코딩 및 디코딩 절차의 다양한 측면을 조절할 수 있다.

[0259] 컴퓨터 시스템(1901)은 단일 코어 또는 멀티 코어 프로세서, 또는 병렬 처리를 위한 복수의 프로세서일 수 있는 중앙 처리 장치(CPU, 또한 "프로세서" 및 "컴퓨터 프로세서")(1905)를 포함한다. 컴퓨터 시스템(1901)은 또한 통신을 위한 메모리 또는 메모리 위치(1910)(가령, 랜덤 액세스 메모리, 리드 온리 메모리, 플래시 메모리), 전

자 저장 장치(1915)(가령, 하드 디스크), 하나 이상의 다른 시스템과 통신하기 위한 통신 인터페이스(1920)(가령, 네트워크 어댑터), 및 주변 장치(1925), 가령, 캐시, 그 밖의 다른 메모리, 데이터 저장소 및/또는 전자 디스플레이 어댑터를 포함한다. 메모리(1910), 저장 유닛(1915), 인터페이스(1920) 및 주변 장치(1925)는 마더보드와 같은 통신 버스(실선)를 통해 CPU(1905)와 통신한다. 저장 유닛(1915)은 데이터를 저장하기 위한 데이터 저장 유닛(또는 데이터 레포지토리)일 수 있다. 컴퓨터 시스템(1901)은 통신 인터페이스(1920)의 도움으로 컴퓨터 네트워크("네트워크")(1930)에 작동 가능하게 연결될 수 있다. 네트워크(1930)는 인터넷, 인터넷 및/또는 엑스트라넷, 또는 인터넷과 통신하는 인트라넷 및/또는 엑스트라넷일 수 있다. 어떤 경우에는 네트워크(1930)는 통신 및/또는 데이터 네트워크이다. 네트워크(1930)는 분산 컴퓨팅을 가능하게 할 수 있는 하나 이상의 컴퓨터 서버, 가령, 클라우드 컴퓨팅을 포함할 수 있다. 네트워크(1930)는 어떤 경우에는 컴퓨터 시스템(1901)의 도움으로 피어-투-피어 네트워크를 구현할 수 있으며, 이는 컴퓨터 시스템(1901)에 연결된 장치가 클라이언트 또는 서버로 동작할 수 있도록 할 수 있다.

[0260] CPU(1905)는 프로그램이나 소프트웨어로 구현될 수 있는 일련의 기계 판독 가능 명령을 실행할 수 있다. 명령은 메모리(1910)와 같은 메모리 위치에 저장될 수 있다. 명령은 CPU(1905)로 전달될 수 있으며, 상기 명령은 본 개시의 방법을 구현하기 위해 CPU(1905)를 후속적으로 프로그래밍하거나 구성할 수 있다. CPU(1905)에 의해 수행되는 작업의 예로는 인출(fetch), 디코딩(decode), 실행(execute) 및 라이트백(writeback)이 포함될 수 있다.

[0261] CPU(1905)는 회로, 가령, 집적 회로의 일부일 수 있다. 시스템(1901)의 하나 이상의 다른 구성요소가 회로에 포함될 수 있다. 어떤 경우에는, 회로가 주문형 집적 회로(ASIC)이다.

[0262] 저장 유닛(1915)은 파일, 가령, 드라이버, 라이브러리, 저장된 프로그램을 저장할 수 있다. 저장 유닛(1915)은 사용자 데이터, 예를 들어, 사용자 선호도, 사용자 프로그램 등을 저장할 수 있다. 일부 경우에 컴퓨터 시스템(1901)은 인트라넷 또는 인터넷을 통해 컴퓨터 시스템(1901)과 통신하는 원격 서버에 위치하는 것과 같이 컴퓨터 시스템(1901) 외부에 있는 하나 이상의 추가 데이터 저장 장치를 포함할 수 있다.

[0263] 컴퓨터 시스템(1901)은 네트워크(1930)를 통해 하나 이상의 원격 컴퓨터 시스템과 통신할 수 있다. 예를 들어, 컴퓨터 시스템(1901)은 사용자의 원격 컴퓨터 시스템 또는 핵산 서열로 인코딩되거나 디코딩된 데이터를 분석하는 과정에서 사용자가 사용할 수 있는 다른 장치 및/또는 기계(가령, 시퀀서 또는 핵산 서열에서 질소 염기의 순서를 화학적으로 결정하기 위한 다른 시스템)와 통신할 수 있다. 원격 컴퓨터 시스템의 예로는 개인용 컴퓨터(가령, 휴대용 PC), 슬레이트 또는 태블릿 PC(가령, Apple® iPad, Samsung® Galaxy Tab), 전화기, 스마트폰(가령, Apple® iPhone, Android 지원 장치, BlackBerry®), 또는 개인 디지털 보조 장치가 있다. 사용자는 네트워크(1930)를 통해 컴퓨터 시스템(1901)을 액세스할 수 있다.

[0264] 본 명세서에 기재된 방법은 예를 들어 메모리(1910) 또는 전자 저장 장치(1915)와 같은 컴퓨터 시스템(1901)의 전자 저장 위치에 저장된 기계(가령, 컴퓨터 프로세서) 실행 코드를 통해 구현될 수 있다. 기계 실행 가능 코드 또는 기계 판독 가능 코드는 소프트웨어 형태로 제공될 수 있다. 사용 중에 코드는 프로세서(1905)에 의해 실행될 수 있다. 일부 경우에, 코드는 저장 유닛(1915)으로부터 검색되어 프로세서(1905)에 의한 즉시 액세스를 위해 메모리(1910)에 저장될 수 있다. 일부 상황에서는 전자 저장 유닛(1915)이 배제될 수 있으며 기계 실행 가능 명령이 메모리(1910)에 저장된다.

[0265] 코드는 코드를 실행하도록 조정된 프로세서가 있는 기계와 함께 사용하기 위해 사전 컴파일 및 구성될 수 있거나 런타임 중에 컴파일될 수 있다. 코드는 사전 컴파일된 방식이나 컴파일된 대로 실행되도록 선택할 수 있는 프로그래밍 언어로 제공될 수 있다.

[0266] 컴퓨터 시스템(1901)과 같이 여기에 제공된 시스템 및 방법의 양태가 프로그래밍으로 구현될 수 있다. 기술의 다양한 측면은 일반적으로 기계(또는 프로세서) 실행 코드 및/또는 기계 판독 가능 매체 유형에 전달되거나 구현되는 관련 데이터 형태의 "제품" 또는 "물품"으로 간주될 수 있다. 기계 실행 가능 코드는 메모리(가령, 리드온리 메모리, 랜덤 액세스 메모리, 플래시 메모리) 또는 하드 디스크와 같은 전자 저장 장치에 저장될 수 있다. "스토리지" 유형의 미디어는 컴퓨터, 프로세서 등의 유형 메모리 또는 다양한 반도체 메모리, 테이프 드라이브, 디스크 드라이브 등과 같은 관련 모듈의 일부 또는 전부를 포함할 수 있으며, 이는 소프트웨어 프로그래밍을 위한 임의의 때에 비일시적 저장을 제공할 수 있다. 소프트웨어의 전체 또는 일부는 때때로 인터넷이나 기타 다양한 통신 네트워크를 통해 전달될 수 있다. 예를 들어, 이러한 통신을 통해 한 컴퓨터 또는 프로세서에서 다른 컴퓨터 또는 프로세서로, 예를 들어 관리 서버 또는 호스트 컴퓨터에서 애플리케이션 서버의 컴퓨터 플랫폼으로 소프트웨어를 로드할 수 있다. 따라서 소프트웨어 요소를 포함할 수 있는 또 다른 유형의 미디어에는 로컬 장치 간의 물리적 인터페이스, 유선 및 광학 유선 네트워크 및 다양한 무선 링크를 통해 사용되는 것과 같은

광학, 전기 및 전자기파가 포함된다. 유무선 링크, 광 링크 등과 같이 이러한 파동을 전달하는 물리적 요소도 소프트웨어를 담고 있는 미디어로 간주될 수 있다. 본 명세서에 사용될 때, 비일시적, 유형의 "저장" 매체로 제한되지 않는 한, 컴퓨터 또는 기계의 "판독 가능한 매체"와 같은 용어는 실행을 위해 프로세서에 명령을 제공하는 데 참여하는 모든 매체를 의미한다.

[0267] 따라서, 컴퓨터 실행 가능 코드와 같은 기계 판독 가능 매체는 유형의 저장 매체, 반송파 매체 또는 물리적 전송 매체를 포함하지만 이에 제한되지 않는 다양한 형태를 취할 수 있다. 비휘발성 저장 매체는 예를 들어, 도면에 도시된 데이터베이스 등을 구현하는 데 사용될 수 있는 임의의 컴퓨터(들) 등의 임의의 저장 장치와 같은 광학 또는 자기 디스크를 포함한다. 휘발성 저장 매체는 컴퓨터 플랫폼의 메인 메모리와 같은 동적 메모리를 포함한다. 유형의 전송 매체는 동축 케이블, 컴퓨터 시스템 내의 버스를 구성하는 전선을 포함한 구리선 및 광섬유를 포함한다. 반송파 전송 매체는 전기 또는 전자기 신호, 무선 주파수(RF) 및 적외선(IR) 데이터 통신 중에 생성되는 것과 같은 음향 또는 광파의 형태를 취할 수 있다. 따라서 컴퓨터 판독 가능 매체의 일반적인 형태는 플로피 디스크, 유연한 디스크, 하드 디스크, 자기 테이프, 기타 자기 매체, CD-ROM, DVD 또는 DVD-ROM, 기타 광학 매체, 펀치 카드 용지 등이 포함된다. 테이프, 구멍 패턴이 있는 기타 물리적 저장 매체, RAM, ROM, PROM 및 EPROM, FLASH-EPROM, 기타 메모리 칩 또는 카트리지, 데이터 또는 명령을 전송하는 반송파, 그러한 캐리어를 전송하는 케이블 또는 링크 웨이브 또는 컴퓨터가 프로그래밍 코드 및/또는 데이터를 읽을 수 있는 기타 매체를 포함한다. 이러한 형태의 컴퓨터 판독 가능 매체 중 다수는 실행을 위해 하나 이상의 명령의 하나 이상의 시퀀스를 프로세서에 전달하는 것과 관련될 수 있다.

[0268] 컴퓨터 시스템(1901)은 예를 들어 서열 출력 데이터, 가령, 크로마토그래프, 서열, 및 핵산, 미가공 데이터, 과일 및 압축 또는 압축해제된 집 파일을 DNA 저장된 데이터로 인코딩 또는 디코딩하는 기계 또는 컴퓨터 시스템에 의해 인코딩되거나 판독되는 비트, 바이트, 비트 스트림을 제공하기 위한 사용자 인터페이스(UI)(1940)를 포함하는 전자 디스플레이(1935)를 포함하거나 이와 통신할 수 있다. UI의 예로는 그래픽 사용자 인터페이스(GUI) 및 웹 기반 사용자 인터페이스가 포함되나 이에 국한되지 않는다.

[0269] 본 개시의 방법 및 시스템은 하나 이상의 알고리즘을 통해 구현될 수 있다. 알고리즘은 중앙 처리 장치(1905)에 의해 실행될 때 소프트웨어를 통해 구현될 수 있다. 예를 들어, 알고리즘은 디지털 정보를 인코딩하기 전에 원시 데이터 또는 집(zip) 파일 압축 데이터로부터 디지털 정보를 코딩하기 위한 맞춤형 방법을 결정하기 위해 DNA 인덱스 및 원시 데이터 또는 집 파일 압축 또는 압축 해제 데이터와 함께 사용될 수 있다.

[0270] **화학적 방법**

[0271] A - 중첩 확장 PCR(OEPCR) 어셈블리 OEPCR에서 중합효소와 dNTP(dATP, dTTP, dCTP, dGTP 또는 이들의 변이체 또는 유사체를 포함하는 데옥시뉴클레오티드 삼인산염)를 포함하는 반응에서 구성요소가 조립된다. 구성요소는 단일 가닥 또는 이중 가닥 핵산일 수 있다. 서로 인접하게 조립될 구성요소는 상보적인 3' 말단, 상보적인 5' 말단, 또는 하나의 구성요소의 5' 말단과 인접한 구성요소의 3' 말단 사이에 상동성을 가질 수 있다. "혼성화 영역"으로 불리는 이들 말단 영역은 OEPCR 동안 구성요소 사이의 혼성화된 접합의 형성을 촉진하기 위한 것이며, 여기서 하나의 입력 구성요소(또는 그 보체)의 3' 말단은 의도된 인접 구성요소(또는 이의 보체)의 3' 말단에 혼성화된다. 이어서, 중합효소 연장에 의해 조립된 이중 가닥 산물이 형성될 수 있다. 이 산물은 후속 혼성화 및 확장을 통해 더 많은 구성요소로 조립될 수 있다. 도 63은 3개의 핵산을 조립하기 위한 OEPCR의 예시적인 개략도를 예시한다.

[0272] 일부 실시예에서, OEPCR은 3가지 온도, 즉 용융 온도, 어닐링 온도 및 연장 온도 사이의 순환을 포함할 수 있다. 용융 온도는 이중 가닥 핵산을 단일 가닥 핵산으로 전환할 뿐만 아니라 구성요소 내에서 또는 구성요소들 간에 2차 구조 또는 혼성화의 형성을 제거하기 위한 것이다. 일반적으로 용융 온도는 섭씨 95도 이상으로 높다. 일부 실시예에서 용융 온도는 적어도 섭씨 96, 97, 98, 99, 100, 101, 102, 103, 104 또는 105도 이상일 수 있다. 다른 실시예에서 용융 온도는 최대 섭씨 95, 94, 93, 92, 91 또는 90도일 수 있다. 용융 온도가 높을수록 핵산과 그 2차 구조의 해리가 향상될 수 있지만, 핵산이나 중합효소의 분해와 같은 부작용이 발생할 수도 있다. 용융 온도는 적어도 1, 2, 3, 4, 5초 또는 그 이상, 예를 들어 30초, 1분, 2분 또는 3분 동안 반응에 적용될 수 있다.

[0273] 어닐링 온도는 의도된 인접 구성요소(또는 그 보체)의 상보적인 3' 말단 사이의 혼성화 형성을 촉진하기 위한 것이다. 일부 실시예에서, 어닐링 온도는 의도된 혼성화된 핵산 형성의 계산된 용융 온도와 일치할 수 있다. 다른 실시예에서, 어닐링 온도는 상기 용융 온도의 섭씨 10도 이내일 수 있다. 일부 실시예에서, 어닐링 온도는 섭씨 25, 30, 50, 55, 60, 65, 또는 70도 이상일 수 있다. 용융 온도는 구성요소들 사이의 의도된 혼성화 영역

의 순서에 따라 달라질 수 있다. 더 긴 혼성화 영역일수록 더 높은 용융 온도를 가지며, 더 높은 구아닌 또는 시토신 뉴클레오티드 함량을 갖는 혼성화 영역일수록 더 높은 용융점을 가질 수 있다. 따라서 특정 어닐링 온도에서 최적으로 조립되도록 의도된 OEPCR 반응용 구성요소를 설계하는 것이 가능할 수 있다. 어닐링 온도는 적어도 1초, 5초, 10초, 15초, 20초, 25초 또는 30초 이상 동안 반응에 적용될 수 있다.

[0274] 연장 온도는 하나 이상의 중합효소에 의해 촉매되는 혼성화된 3' 말단의 핵산 사슬 연장을 시작하고 촉진하기 위한 것이다. 일부 구현예에서, 연장 온도는 중합효소가 핵산 결합 강도, 연장 속도, 연장 안정성 또는 충실도 측면에서 최적으로 기능하는 온도로 설정될 수 있다. 일부 실시예에서 연장 온도는 적어도 섭씨 30도, 40도, 50도, 60도 또는 70도 이상일 수 있다. 어닐링 온도는 적어도 1초, 5초, 10초, 15초, 20초, 25초, 30초, 40초, 50초 또는 60초 이상 동안 반응에 적용될 수 있다. 권장 연장 시간은 예상 연장의 킬로베이스당 약 15 내지 45 초일 수 있다.

[0275] OEPCR의 일부 실시예에서, 어닐링 온도와 연장 온도는 동일할 수 있다. 따라서 3단계 온도 사이클 대신 2단계 온도 사이클이 사용될 수 있다. 결합된 어닐링 및 연장 온도의 예로는 섭씨 60, 65 또는 72도가 있다.

[0276] 일부 실시예에서, OEPCR은 하나의 온도 사이클로 수행될 수 있다. 그러한 실시예는 단 두 개의 구성요소의 의도된 조립을 포함할 수 있다. 다른 실시예에서, OEPCR은 다중 온도 사이클로 수행될 수 있다. OEPCR의 모든 특정 핵산은 하나의 주기에서 최대 하나의 다른 핵산에만 조립될 수 있다. 이는 조립(또는 연장 또는 연장)이 핵산의 3' 말단에서만 발생하고 각 핵산에는 3' 말단이 하나만 있기 때문이다. 따라서 여러 구성요소를 조립하려면 여러 온도 주기가 필요할 수 있다. 예를 들어, 4개의 구성요소를 조립하려면 3회의 온도 사이클이 필요할 수 있다. 6개의 구성요소를 조립하려면 5회의 온도 사이클이 필요할 수 있다. 10개의 구성요소를 조립하려면 9회의 온도 사이클이 필요할 수 있다. 일부 실시예에서, 필요한 최소치보다 더 많은 온도 사이클을 사용하면 조립 효율성이 증가할 수 있다. 예를 들어, 두 개의 구성요소를 조립하기 위해 4개의 온도 사이클을 사용하면 하나의 온도 사이클만 사용하는 것보다 더 많은 산물을 생산할 수 있다. 이는 구성요소의 혼성화 및 신장이 각 주기의 전체 구성요소 수의 일부에서 발생하는 통계적 이벤트이기 때문이다. 따라서 조립된 구성요소의 전체 비율은 사이클이 증가함에 따라 증가할 수 있다.

[0277] 온도 사이클링 고려사항 외에도 OEPCR의 핵산 서열 설계는 서로의 조립 효율성에 영향을 미칠 수 있다. 긴 혼성화 영역을 갖는 핵산은 짧은 혼성화 영역을 갖는 핵산에 비해 주어진 어닐링 온도에서 더 효율적으로 혼성화할 수 있다. 이는 더 긴 혼성화 제품이 더 많은 수의 안정적인 염기쌍을 포함하고 따라서 더 짧은 혼성화 제품보다 전체적으로 더 안정적인 혼성화 제품일 수 있기 때문이다. 혼성화 영역은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 이상의 염기 길이를 가질 수 있다.

[0278] 구아닌 또는 시토신 함량이 높은 혼성화 영역은 구아닌 또는 시토신 함량이 낮은 혼성화 영역보다 주어진 온도에서 더 효율적으로 혼성화할 수 있다. 이는 구아닌이 아데닌이 티민과 형성하는 것보다 시토신과 더 안정적인 염기쌍을 형성하기 때문이다. 혼성화 영역은 0%에서 100% 사이의 구아닌 또는 시토신 함량(GC 함량이라고도 함)을 가질 수 있다.

[0279] 혼성화 영역 길이 및 GC 함량 외에도 OEPCR의 효율성에 영향을 미칠 수 있는 핵산 서열 설계의 더 많은 측면이 있다. 예를 들어, 구성요소 내의 바람직하지 않은 2차 구조의 형성은 의도된 인접 구성요소와의 혼성화 생성물을 형성하는 능력을 방해할 수 있다. 이들 2차 구조는 헤어핀 루프(hairpin loop)를 포함할 수 있다. 핵산에 대한 가능한 2차 구조의 유형과 그 안정성(가령, 용융 온도)은 서열을 기반으로 예측할 수 있다. 설계 공간 검색 알고리즘이 사용되어 잠재적으로 억제적인 2차 구조가 있는 서열을 피하면서 효율적인 OEPCR을 위한 적절한 길이와 GC 함량 기준을 충족하는 핵산 서열을 결정할 수 있다. 설계 공간 검색 알고리즘에는 유전자 알고리즘, 휴리스틱 검색 알고리즘, 금기 검색과 같은 메타 휴리스틱 검색 전략, 분기 및 경계 검색 알고리즘, 동적 프로그래밍 기반 알고리즘, 제한된 조합 최적화 알고리즘, 경사 하강 기반 알고리즘, 무작위 검색 알고리즘 또는 이들의 조합이 포함될 수 있다.

[0280] 마찬가지로, 동종이량체(동일한 서열의 핵산 분자와 혼성화하는 핵산 분자) 및 원치 않는 이종이량체(의도된 조립 파트너를 제외하고 다른 핵산 서열과 혼성화하는 핵산 서열)의 형성은 OEPCR을 방해할 수 있다. 핵산 내의 2차 구조와 유사하게, 동종이량체 및 이종이량체의 형성은 계산 방법 및 설계 공간 검색 알고리즘을 사용하여 핵산 설계 중에 예측되고 설명될 수 있다.

[0281] 더 긴 핵산 서열 또는 더 높은 GC 함량은 OEPCR을 통해 원치 않는 2차 구조, 동종이량체 및 이종이량체의 형성을 증가시킬 수 있다. 따라서, 일부 실시예에서, 더 짧은 핵산 서열 또는 더 낮은 GC 함량의 사용은 더 높은 조

립 효율을 초래할 수 있다. 이들 설계 원칙은 보다 효율적인 조립을 위해 긴 혼성화 영역이나 높은 GC 함량을 사용하는 설계 전략에 반대될 수 있다. 따라서, 일부 실시예에서, OEPCR은 높은 GC 함량을 갖는 긴 혼성화 영역을 사용하고 낮은 GC 함량을 갖는 짧은 비혼성화 영역을 사용함으로써 최적화될 수 있다. 핵산의 전체 길이는 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90 또는 100개 염기 또는 그 이상일 수 있다. 일부 실시예에서, 조립 효율이 최적화되는 핵산의 혼성화 영역에 대한 최적의 길이 및 최적의 GC 함량이 있을 수 있다.

[0282] OEPCR 반응에서 더 많은 수의 개별 핵산이 예상 조립 효율을 방해할 수 있다. 이는 더 많은 수의 별개의 핵산 서열이 특히 이중이량체 형태로 바람직하지 않은 분자 상호작용에 대한 더 높은 확률을 생성할 수 있기 때문이다. 따라서 다수의 구성요소를 조립하는 OEPCR의 일부 구현예에서, 핵산 서열 제약은 효율적인 조립을 위해 더욱 엄격해질 수 있다.

[0283] 예상되는 최종 조립 산물을 증폭하기 위한 프라이머가 OEPCR 반응에 포함될 수 있다. 그런 다음 OEPCR 반응은 구성요소 사이에 더 많은 조립체를 생성할 뿐만 아니라 기존 PCR 방식으로 전체 조립된 산물을 기하급수적으로 증폭하여 조립된 제품의 수율을 향상시키기 위해 더 많은 온도 주기로 수행될 수 있다(화학적 방법 섹션 D를 참조할 수 있다).

[0284] 조립 효율성을 향상시키기 위해 OEPCR 반응에 첨가제가 포함될 수 있다. 예를 들어, 베타인, 디메틸 설펝사이드(DMSO), 비이온성 세제, 포름아미드, 마그네슘, 소 혈청 알부민(BSA) 또는 이들의 조합의 첨가가 있다. 첨가물 함량(체적당 중량)은 적어도 0%, 1%, 5%, 10%, 20% 이상일 수 있다.

[0285] OEPCR을 위해 다양한 중합효소가 사용될 수 있다. 중합효소는 자연적으로 발생하거나 합성될 수 있다. 중합효소의 예는  $\Phi$ 29 중합효소 또는 이의 유도체이다. 일부 경우에, 전사효소 또는 리가제(즉, 결합 형성을 촉매하는 효소)가 중합효소와 함께 또는 중합효소의 대안으로서 사용되어 새로운 핵산 서열을 구성할 수 있다. 중합효소의 예로는 DNA 중합효소, RNA 중합효소, 열안정성 중합효소, 야생형 중합효소, 변형된 중합효소, E.coli DNA 중합효소 I, T7 DNA 중합효소, 박테리오파지 T4 DNA 중합효소  $\Phi$ 29 ( $\phi$ 29) DNA 중합효소, Taq 중합효소, Tth 중합효소, Tli 중합효소, Pfu 중합효소 Pwo 중합효소, VENT 중합효소, DEEPVENT 중합효소, Ex-Taq 중합효소, LA-Taw 중합효소, Sso 중합효소 Poc 중합효소, Pab 중합효소, Mth 중합효소 ES4 중합효소, Tru 중합효소, Tac 중합효소, Tne 중합효소, Tma 중합효소, Tca 중합효소, Tih 중합효소, Tfi 중합효소, 백금 Taq 중합효소, Tbr 중합효소, Phusion 중합효소, KAPA 중합효소, Q5 중합효소, Tfi 중합효소, Pfutubo 중합효소, Pyrobest 중합효소, KOD 중합효소, Bst 중합효소, Sac 중합효소, 3' 내지 5' 엑소뉴클레아제 활성을 갖는 Klenow 단편 중합효소 및 이들의 변이, 변형 산물 및 유도체를 포함하나, 이에 한정되지는 않는다. 상이한 중합효소는 상이한 온도에서 안정적이고 최적으로 기능할 수 있다. 또한, 상이한 중합효소는 상이한 특성을 가진다. 예를 들어, Phusion 중합효소와 같은 일부 중합효소는 3'에서 5' 엑소뉴클레아제 활성을 나타낼 수 있으며, 이는 핵산 신장 동안 더 높은 충실도에 기여할 수 있다. 일부 중합효소는 신장(elongation) 동안 주요 서열을 대체할 수 있는 반면, 다른 중합효소는 이를 분해하거나 신장을 중단시킬 수 있다. Taq과 같은 일부 중합효소는 핵산 서열의 3' 말단에 아데닌 염기를 통합한다. 이 과정을 A-테일링(A-tailing)이라고 하며, 아데닌 염기를 추가하면 의도된 인접 구성요소 간의 설계된 3' 상보성을 방해할 수 있으므로 OEPCR을 억제할 수 있다.

[0286] OEPCR은 중합효소 순환 조립(또는 PCA)이라고도 한다.

[0287] B - 결찰 조립 결찰 조립에서, 하나 이상의 리가제 효소와 추가 보조인자를 포함하는 반응에서 별도의 핵산이 조립된다. 보조인자에는 아데노신 삼인산염(ATP), 디티오프레이톨(DTT) 또는 마그네슘 이온( $Mg^{2+}$ )이 포함될 수 있다. 결찰(ligation) 동안, 하나의 핵산 가닥의 3'-말단은 다른 핵산 가닥의 5'-말단에 공유적으로 연결되어 조립된 핵산을 형성한다. 결찰 반응의 구성요소는 무딘 말단 이중 가닥 DNA(dsDNA), 단일 가닥 DNA(ssDNA) 또는 부분적으로 혼성화된 단일 가닥 DNA일 수 있다. 핵산의 말단을 하나로 모으는 전략은 리가제 효소에 대한 생존 기질의 빈도를 증가시켜 리가제 반응의 효율성을 향상시키는 데 사용될 수 있다. 무딘 말단의 dsDNA 분자는 리가제 효소가 작용할 수 있는 소수성 스택을 형성하는 경향이 있지만, 핵산을 하나로 모으는 보다 성공적인 전략은 조립되려 의도되는 구성요소의 오버행에 대한 상보성을 갖는 5' 또는 3' 단일 가닥 오버행을 갖는 핵산 구성요소를 사용하는 것일 수 있다. 후자의 경우, 염기-염기 혼성화로 인해 보다 안정적인 핵산 이중가닥이 형성될 수 있다.

[0288] 이중 가닥 핵산의 한쪽 끝에 오버행 가닥이 있는 경우, 동일한 끝의 다른 가닥은 "캐비티(cavity)"로 지칭될 수 있다. 캐비티와 오버행이 함께 "접착성 말단(cohesive-end)"라고도 알려진 "점착 말단(sticky end)"을 형성한다. 점착 말단은 3' 오버행 및 5' 캐비티일 수도 있고, 5' 오버행 및 3' 캐비티일 수도 있다. 2개의 의도된 인접한 구성요소들 사이의 점착 말단은 두 점착 말단의 오버행이 혼성화되어 각 오버행이 다른 구성요소 상

의 캐비티의 시작부분에 직접 인접하게 끝나도록 상보성을 갖도록 설계될 수 있다. 이는 리가제의 작용에 의해 "실링"(포스포디에스테르 결합을 통해 공유 결합)될 수 있는 "닉(nick)"(이중 가닥 DNA 파손)을 형성한다. 3개의 핵산을 조립하기 위한 점착 말단 결합의 예시 도식은 도 17에 나와 있다. 한쪽 가닥이나 다른 쪽 가닥 또는 둘 모두 상의 닉이 실링될 수 있다. 열역학적으로, 점착 말단을 형성하는 분자의 상단 및 하단 가닥은 연계된 상태와 해리된 상태 사이를 이동할 수 있으므로 점착 말단은 일시적인 형성일 수 있다. 그러나 두 구성요소 사이의 점착 말단 이중 가닥의 한 가닥을 따라 있는 닉이 실링되면, 반대 가닥의 구성원이 분리되더라도 해당 공유 결합은 그대로 유지된다. 그런 다음 연결된 가닥은 반대쪽 가닥의 의도된 인접 구성원이 결합할 수 있고 다시 한번 실링될 수 있는 닉을 형성할 수 있는 주형(template)이 될 수 있다.

[0289] 점착 말단은 하나 이상의 엔도뉴클레아제로 dsDNA를 분해함으로써 생성될 수 있다. 엔도뉴클레아제(제한 효소라고도 함)는 dsDNA 분자의 한쪽 또는 양쪽 말단에서 특정 부위(제한 부위라고도 함)를 표적으로 삼아 엇갈린 절단(때때로 소화라고도 함)을 생성하여 점착 말단을 남겨둘 수 있다. 제한 소화에 대해서는 화학적 방법 섹션 C를 참조할 수 있다. 소화는 회문형 오버행(자체 역보체인 서열이 있는 오버행)을 남길 수 있다. 그렇다면, 동일한 엔도뉴클레아제로 소화된 두 구성요소는 리가제와 조립될 수 있는 상보적인 점착 말단을 형성할 수 있다. 엔도뉴클레아제와 리가제가 호환되는 경우 동일한 반응에서 소화와 결합이 함께 발생할 수 있다. 반응은 섭씨 4, 10, 16, 25 또는 37도와 같은 균일한 온도에서 일어날 수 있다. 또는 반응은 섭씨 16도에서 37도 사이와 같이 여러 온도 사이에서 순환될 수 있다. 여러 온도 사이를 순환하면 주기의 여러 부분 동안 소화와 결합이 각각 최적의 온도에서 진행될 수 있다.

[0290] 소화와 결합을 별도의 반응으로 수행하는 것이 유익할 수 있다. 예를 들어, 원하는 리가제와 원하는 엔도뉴클레아제가 서로 다른 조건에서 최적으로 기능하는 경우이다. 또는 예를 들어, 결합된 생성물이 엔도뉴클레아제에 대한 새로운 제한 부위를 형성하는 경우이다. 이러한 경우, 제한 소화를 수행한 후 결합(ligation)을 별도로 수행하는 것이 더 나을 수 있으며, 아마도 결합 전에 제한 효소를 제거하는 것이 더 유리할 수 있다. 핵산은 페놀-클로로포름 추출, 에탄올 침전, 자성 비드 포획 및/또는 실리카막 흡착, 세척 및 용리를 통해 효소로부터 분리될 수 있다. 여러 엔도뉴클레아제가 동일한 반응에 사용될 수 있지만, 엔도뉴클레아제가 서로 간섭하지 않고 유사한 반응 조건에서 기능하도록 주의를 기울여야 한다. 두 개의 엔도뉴클레아제를 사용하면 dsDNA 구성요소의 양쪽 말단에 직교(비상보적) 점착 말단을 만들 수 있다.

[0291] 엔도뉴클레아제 소화는 인산화된 5' 말단과 함께 점착 말단을 남길 것이다. 리가제는 인산화된 5' 말단에서만 기능할 수 있으며, 인산화되지 않은 5' 말단에서는 기능할 수 없다. 따라서 소화와 결합 사이에 중간 5' 인산화 단계가 필요하지 않을 수 있다. 점착 말단 상에 회문 오버행이 있는 소화된 dsDNA 구성요소는 자체적으로 결합될 수 있다. 자가 결합을 방지하기 위해, 결합 전에 상기 dsDNA 구성요소를 탈인산화하는 것이 유익할 수 있다.

[0292] 다수의 엔도뉴클레아제는 서로 다른 제한 부위를 표적으로 삼을 수 있지만 호환 가능한 오버행(서로의 역보완인 오버행)을 남길 수 있다. 두 개의 이러한 엔도뉴클레아제로 생성된 점착 말단의 결합의 생성물은 결합 부위에 어느 엔도뉴클레아제에 대한 제한 부위도 포함하지 않는 조립된 생성물을 생성할 수 있다. 이러한 엔도뉴클레아제는 반복적인 소화-결합 주기를 수행함으로써 단 두 개의 엔도뉴클레아제를 사용하여 여러 구성요소를 프로그래밍 방식으로 조립할 수 있는 바이오브릭 조립과 같은 조립 방법의 기초를 형성한다. 도 76은 호환 가능한 오버행을 갖는 엔도뉴클레아제 BamHI 및 BglIII를 사용하는 소화-결합 주기의 예를 예시한다.

[0293] 일부 구현예에서, 점착 말단을 생성하는 데 사용되는 엔도뉴클레아제는 IIS 유형 제한 효소일 수 있다. 이들 효소는 제한 부위에서 특정 방향으로 고정된 수의 염기를 절단하므로 이들이 생성하는 오버행의 서열을 맞춤화할 수 있다. 오버행 서열은 회문식일 필요는 없다. 동일한 유형의 IIS 제한 효소가 사용되어 동일한 반응 또는 여러 반응에서 여러 개의 상이한 점착 말단을 생성할 수 있다. 더욱이, 하나 또는 다중 유형의 IIS 제한 효소가 사용되어 동일한 반응 또는 다중 반응에서 호환 가능한 오버행을 갖는 구성요소를 생성할 수 있다. 유형 IIS 제한 효소에 의해 생성된 두 개의 점착 말단 사이의 결합 부위는 새로운 제한 부위를 형성하지 않도록 설계될 수 있다. 또한, 유형 IIS 제한 효소 부위는 dsDNA에 위치하여 제한 효소가 점착 말단을 갖는 구성요소를 생성할 때 자신의 제한 부위를 절단할 수 있다. 따라서 IIS 제한 효소 유형에서 생성된 여러 구성요소 간의 결합 생성물은 어떠한 제한 부위도 포함하지 않을 수 있다.

[0294] 유형 IIS 제한 효소는 리가제와 함께 반응에서 혼합되어 구성요소 소화 및 결합을 함께 수행할 수 있다. 반응의 온도는 최적의 소화 및 결합을 촉진하기 위해 두 개 이상의 값 사이에서 순환될 수 있다. 예를 들어, 소화는 섭씨 37도에서 최적으로 수행될 수 있고, 결합은 섭씨 16도에서 최적으로 수행될 수 있다. 보다 일반적으로, 반응은 적어도 섭씨 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 또는 65도 이상의 온도 값 사이에서 순환

될 수 있다. 조합된 소화 및 결찰 반응이 사용되어 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 이상의 구성요소를 조합할 수 있다. 유형 IIS 제한 효소를 활용하여 점착 말단을 생성하는 조립 반응의 예로는 Golden Gate Assembly(Golden Gate Cloning이라고도 함) 또는 Modular Cloning(MoClo라고도 함)이 있다.

[0295] 결찰의 일부 실시예에서, 엑소뉴클레아제를 사용하여 점착 말단을 갖는 구성요소가 생성될 수 있다. 3' 엑소뉴클레아제는 dsDNA의 3' 말단을 추잉백하여 5' 오버행을 생성하는 데 사용될 수 있다. 마찬가지로, 5' 엑소뉴클레아제가 dsDNA의 5' 말단을 추잉백하여 3' 오버행을 생성하는 데 사용될 수 있다. 상이한 엑소뉴클레아제는 상이한 특성을 가질 수 있다. 예를 들어, ssDNA에 작용하는지 여부, 인산화된 또는 비인산화된 5' 말단에 작용하는지 여부, 닉(nick)에서 시작할 수 있는지 여부, 또는 5' 캐비티, 3' 캐비티, 5' 오버행 또는 3' 오버행에서 활동을 시작할 수 있는지 여부에 따라, 엑소뉴클레아제는 뉴클레아제 활성 방향(5'에서 3' 또는 3'에서 5')이 상이할 수 있다. 다양한 유형의 엑소뉴클레아제에는 람다 엑소뉴클레아제, RecJ<sub>I</sub>, 엑소뉴클레아제 III, 엑소뉴클레아제 I, 엑소뉴클레아제 T, 엑소뉴클레아제 V, 엑소뉴클레아제 VIII, 엑소뉴클레아제 VII, 뉴클레아제 BAL\_31, T5 엑소뉴클레아제, 및 T7 엑소뉴클레아제가 포함된다.

[0296] 엑소뉴클레아제는 리가제와 함께 반응에 사용되어 여러 구성요소를 조립할 수 있다. 반응은 고정된 온도 또는 여러 온도 사이의 주기에서 발생할 수 있으며, 각각은 리가제 또는 엑소뉴클레아제에 이상적이다. 중합효소는 리가제 및 5'-to-3' 엑소뉴클레아제와의 조립 반응에 포함될 수 있다. 이러한 반응에서의 구성요소는 서로 인접하여 조립되도록 의도된 구성요소가 가장자리에서 상동성 서열을 공유하도록 설계될 수 있다. 예를 들어, 구성요소 Y와 조립될 구성요소 X는 5'-z-3' 형태의 3' 가장자리 서열을 가질 수 있고, 구성요소 Y는 5'-z-3' 형태의 5' 가장자리 서열을 가질 수 있고, 여기서 z는 임의의 핵산 서열이다. 우리는 '깁슨 오버랩(gibson overlap)'과 같은 형태의 상동 가장자리 서열을 참조한다. 5' 엑소뉴클레아제는 깁슨 오버랩이 있는 dsDNA 구성요소의 5' 말단을 추잉백할 때 서로 혼성화되는 호환 가능한 3' 오버행을 생성한다. 그런 다음 혼성화된 3' 말단은 중합효소의 작용에 의해 주형 구성요소의 말단까지 또는 한 구성요소의 확장된 3' 오버행이 인접한 구성요소의 5' 캐비티와 만나는 지점까지 확장되어, 리가제에 의해 실링될 수 있는 닉을 형성할 수 있다. 중합효소, 리가제, 및 엑소뉴클레아제가 함께 사용되는 이러한 조립 반응이 종종 "깁슨 조립(Gibson assembly)"이라고 한다. 깁슨 조립은 T5 엑소뉴클레아제, Phusion 중합효소 및 Taq 리가제를 사용하고 반응물을 섭씨 50도에서 배양하여 수행할 수 있다. 상기 경우, 호열성 리가제인 Taq를 사용하면 반응에서 세 가지 유형의 효소 모두에 적합한 온도인 섭씨 50도에서 반응이 진행될 수 있다.

[0297] "깁슨 조립"이라는 용어는 일반적으로 중합효소, 리가제 및 엑소뉴클레아제를 포함하는 모든 조립 반응을 의미할 수 있다. 깁슨 조립은 적어도 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상의 구성요소를 조립하는 데 사용될 수 있다. 깁슨 조립은 1단계, 등온 반응 또는 하나 이상의 온도 배양을 통한 다단계 반응으로 발생할 수 있다. 예를 들어, 깁슨 조립은 최소 30, 40, 50, 60 또는 70도 이하의 온도에서 발생할 수 있다. 깁슨 조립을 위한 배양 시간은 적어도 1, 5, 10, 20, 40 또는 80분일 수 있다.

[0298] 깁슨 조립 반응은 의도된 인접 구성요소들 사이의 깁슨 중첩이 특정 길이이고 헤어핀, 동종이량체 또는 원치 않는 이종이량체와 같은 바람직하지 않은 혼성화 사건을 피하는 서열과 같은 서열 특징을 가질 때 최적으로 발생할 수 있다. 일반적으로, 적어도 20개 베이스의 깁슨 오버랩이 권장된다. 그러나 깁슨 오버랩은 길이가 적어도 1, 2, 3, 5, 10, 20, 30, 40, 50, 60, 100 또는 그 이상의 염기일 수 있다. 깁슨 오버랩의 GC 함량은 0%에서 100% 사이일 수 있다.

[0299] 깁슨 조립은 일반적으로 5' 엑소뉴클레아제로 설명되지만 반응은 3' 엑소뉴클레아제에서도 발생할 수 있다. 3' 엑소뉴클레아제가 dsDNA 구성요소의 3' 말단을 추잉백하면서 중합효소는 3' 말단을 확장함으로써 해당 작용을 방해한다. 이러한 동적 과정은 두 구성요소(깁슨 중첩을 공유함)의 5' 오버행(엑소뉴클레아제에 의해 생성됨)이 혼성화되고 중합효소가 한 구성요소의 3' 말단을 인접 구성요소의 5' 말단과 만날 만큼 충분히 멀리 확장할 때까지 계속될 수 있으며, 따라서 리가제에 의해 봉인될 수 있는 닉이 남겨질 수 있다.

[0300] 결찰의 일부 실시예에서, 점착 말단을 갖는 구성요소는 완전한 상보성을 공유하지 않는 2개의 단일 가닥 핵산 또는 올리고를 함께 혼합함으로써 효소적으로가 아니라 합성적으로 생성될 수 있다. 예를 들어, 두 개의 올리고, 올리고 X와 올리고 Y는 하나 또는 두 올리고 모두를 구성하는 더 큰 염기 스트링의 서브스트링을 형성하는 연속적인 상보 염기 스트링을 따라 완전히 혼성화하도록 설계될 수 있다. 이 상보적인 염기 스트링을 "인덱스 영역"이라고 한다. 인덱스 영역이 올리고 X 전체와 올리고 Y의 5' 말단만을 차지하는 경우, 올리고는 함께 한 쪽에는 평활 말단이 있고 다른 한 쪽에는 올리고 Y의 3' 오버행이 있는 점착 말단이 있는 구성요소를 형성한

다(도 77a). 인텍스 영역이 올리고 X 전체와 올리고 Y의 3' 말단만을 차지하는 경우, 올리고는 함께 한 쪽에는 평할 말단이 있고 다른 한 쪽에는 올리고 Y의 5' 오버행이 있는 점착 말단이 있는 구성요소를 형성한다(도 77b). 인텍스 영역이 올리고 X 전체를 차지하고 올리고 Y의 어느 쪽 말단도 차지하지 않는 경우(인텍스 영역이 올리고 Y의 중간에 내장되어 있음을 의미), 올리고는 함께 한 쪽에 올리고 Y로부터의 3' 오버행이 있는 점착 말단이 있고 다른 쪽에 올리고 Y로부터의 5' 오버행을 갖는 구성요소를 형성한다(도 77c). 인텍스 영역이 올리고 X의 5' 말단과 올리고 Y의 5' 말단만 차지하는 경우, 올리고는 함께 한 쪽에는 올리고 Y로부터의 3' 오버행이 있는 점착 말단이 있고 다른 쪽에는 올리고 X로부터의 3' 오버행이 있는 구성요소를 형성한다(도 77d). 인텍스 영역이 올리고 X의 3' 말단과 올리고 Y의 3' 말단만 차지하는 경우, 올리고는 함께 한 쪽에는 올리고 Y로부터의 5' 오버행이 있는 점착 말단이 있고 다른 쪽에는 올리고 X로부터의 5' 오버행이 있는 구성요소를 형성한다(도 77e). 전술한 예에서, 오버행의 서열은 인텍스 영역 외부의 올리고 서열에 의해 정의된다. 이들 오버행 서열은 결찰을 위해 구성요소가 혼성화되는 영역이기 때문에 혼성화 영역으로 지칭될 수 있다.

[0301] 점착 말단 결찰에서 올리고의 인텍스 영역과 혼성화 영역은 구성요소의 적절한 조립을 촉진하도록 설계될 수 있다. 오버행이 긴 구성요소는 오버행이 짧은 구성요소에 비해 주어진 어닐링 온도에서 서로 더 효율적으로 혼성화할 수 있다. 오버행은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30개 이상의 염기 길이를 가질 수 있다.

[0302] 높은 구아닌 또는 시토신 함량을 포함하는 오버행을 갖는 구성요소는 낮은 구아닌 또는 시토신 함량을 포함하는 오버행을 갖는 구성요소보다 주어진 온도에서 상보적 구성요소에 더 효율적으로 혼성화할 수 있다. 이는 구아닌이 아데닌이 티민과 형성하는 것보다 시토신과 더 안정적인 염기쌍을 형성하기 때문이다. 오버행의 구아닌 또는 시토신 함량(GC 함량이라고도 함)은 0%에서 100% 사이일 수 있다.

[0303] 오버행 서열과 마찬가지로, 올리고 인텍스 영역의 GC 함량과 길이도 결찰 효율성에 영향을 미칠 수 있다. 이는 각 구성요소의 상단과 하단 가닥을 안정적으로 묶어주면 점착 말단의 구성요소가 더욱 효율적으로 조립될 수 있기 때문이다. 따라서 인텍스 영역은 더 높은 GC 함량, 더 긴 시퀀스 및 더 높은 용융 온도를 촉진하는 기타 기능으로 설계될 수 있다. 그러나 인텍스 영역과 오버행 서열 모두에 대해 올리고 설계에는 결찰 조립의 효율성에 영향을 미칠 수 있는 더 많은 측면이 있다. 예를 들어, 구성요소 내에 원하지 않는 2차 구조가 형성되면 의도된 인접 구성요소와 조립된 생성물을 형성하는 능력이 방해받을 수 있다. 이는 인텍스 영역, 오버행 서열, 또는 둘 모두에서의 2차 구조로 인해 발생할 수 있다. 이들 2차 구조는 헤어핀 루프(hairpin loop)를 포함할 수 있다. 올리고에 대한 가능한 2차 구조 유형과 안정성(가령, 접합 온도)은 서열을 기반으로 예측될 수 있다. 설계 공간 검색 알고리즘이 사용되어 잠재적으로 억제할 수 있는 2차 구조가 있는 시퀀스를 피하면서 효과적인 구성요소 형성을 위한 적절한 길이와 GC 함량 기준을 충족하는 올리고 시퀀스를 결정할 수 있다. 설계 공간 검색 알고리즘에는 유전자 알고리즘, 휴리스틱 검색 알고리즘, 금기 검색과 같은 메타 휴리스틱 검색 전략, 분기 및 경계 검색 알고리즘, 동적 프로그래밍 기반 알고리즘, 제한된 조합 최적화 알고리즘, 경사 하강 기반 알고리즘, 무작위 검색 알고리즘 또는 이들의 조합이 포함될 수 있다.

[0304] 마찬가지로, 동종이량체(동일한 서열의 올리고와 혼성화하는 올리고) 및 원치 않는 이종이량체(의도된 조립 파트너를 제외한 다른 올리고와 혼성화하는 올리고)의 형성은 결찰을 방해할 수 있다. 구성요소 내의 2차 구조와 유사하게, 동종이량체 및 이종이량체의 형성은 계산 방법 및 설계 공간 검색 알고리즘을 사용하여 구성요소 설계 중에 예측되고 설명될 수 있다.

[0305] 올리고 서열이 길거나 GC 함량이 높을수록 결찰 반응 내에서 원치 않는 2차 구조, 동종이량체 및 이종이량체의 형성이 증가할 수 있다. 따라서 일부 실시예에서 더 짧은 올리고 또는 더 낮은 GC 함량을 사용하면 조립 효율성이 더 높아질 수 있다. 이러한 설계 원칙은 보다 효율적인 조립을 위해 긴 올리고 또는 높은 GC 함량을 사용하는 설계 전략에 반대될 수 있다. 따라서, 결찰 조립의 효율성이 최적화되도록 각 구성요소를 구성하는 올리고에 대한 최적의 길이와 최적의 GC 함량이 있을 수 있다. 결찰에 사용되는 올리고의 전체 길이는 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90 또는 100개 염기 또는 그 이상일 수 있다. 결찰에 사용되는 올리고의 전체 GC 함량은 0%에서 100% 사이일 수 있다.

[0306] 점착 말단 결찰 외에도, 스테이플(또는 주형 또는 브리지) 가닥을 사용하여 단일 가닥 핵산들 간에 결찰이 발생할 수도 있다. 이 방법은 스테이플 가닥 결찰(SSL), 주형 지정 결찰(TDL) 또는 브리지 가닥 결찰이라고 할 수 있다. 3개의 핵산을 조립하기 위한 TDL의 예시적 개략도는 도 66a에 도시되어 있다. TDL에서는 두 개의 단일 가닥 핵산이 주형에 인접하게 혼성화되어 결찰에 의해 밀봉될 수 있는 틈을 형성한다. 점착 말단 결찰에 대한 동일한 핵산 설계 고려 사항이 TDL에도 적용된다. 주형과 의도된 상보적 핵산 서열 사이의 더 강한 혼성화는 증가

된 결찰 효율로 이어질 수 있다. 따라서 주형 양쪽의 혼성화 안정성(또는 용융 온도)을 향상시키는 서열 특징은 결찰 효율을 향상시킬 수 있다. 이러한 특징에는 더 긴 서열 길이와 더 높은 GC 함량이 포함될 수 있다. 주형을 포함한 TDL의 핵산 길이는 적어도 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 또는 100개 염기 또는 그 이상일 수 있다. 주형을 포함한 핵산의 GC 함량은 0%에서 100% 사이일 수 있다.

[0307] TDL에서는, 점착 말단 연결과 마찬가지로 서열 공간 검색 알고리즘이 포함된 핵산 구조 예측 소프트웨어를 사용하여 원치 않는 2차 구조를 피하는 구성요소 및 주형 서열을 설계하는 데 주의를 기울일 수 있다. TDL의 구성요소는 이중 가닥이 아닌 단일 가닥일 수 있으므로 노출된 염기로 인해 원치 않는 2차 구조(점착 말단 결찰과 비교하여)가 발생할 가능성이 더 높을 수 있다.

[0308] TDL은 또한 평활 말단 dsDNA 구성요소를 사용하여 수행될 수도 있다. 이러한 반응에서, 스테이플 가닥이 두 개의 단일 가닥 핵산을 적절하게 연결하기 위해 스테이플은 먼저 전체 단일 가닥 상보체를 대체하거나 부분적으로 대체해야 할 수 있다. dsDNA 구성요소와의 TDL 반응을 촉진하기 위해 dsDNA는 초기에 고온에서 배양하여 용융될 수 있다. 그런 다음 반응물이 냉각되어 스테이플 가닥이 적절한 핵산 상보체에 어닐링될 수 있다. 이 과정은 dsDNA 구성요소에 비해 상대적으로 높은 농도의 주형을 사용하여 훨씬 더 효율적으로 이루어질 수 있으며, 따라서 주형이 결합을 위해 적절한 전체 길이의 ssDNA 보체와 경쟁할 수 있게 된다. 두 개의 ssDNA 가닥이 주형과 리가제에 의해 조립되면, 조립된 핵산은 반대편 전장 ssDNA 보체에 대한 주형이 될 수 있다. 따라서 평활 말단 dsDNA와 TDL의 연결은 여러 차례의 용융(더 높은 온도에서 배양) 및 어닐링(낮은 온도에서 배양)을 통해 개선될 수 있다. 이 과정을 리가제 순환 반응(LCR)이라고 한다. 적절한 용융 및 어닐링 온도는 핵산 서열에 따라 달라진다. 용융 및 어닐링 온도는 적어도 섭씨 4, 10, 20, 20, 30, 40, 50, 60, 70, 80, 90 또는 100도 이상일 수 있다. 온도 사이클의 수는 적어도 1, 5, 10, 15, 20, 15, 30 또는 그 이상일 수 있다.

[0309] 모든 결찰은 고정 온도 반응 또는 다중 온도 반응에서 수행될 수 있다. 결찰 온도는 적어도 섭씨 0, 4, 10, 20, 20, 30, 40, 50 또는 60도 이상일 수 있다. 리가제 활성을 위한 최적 온도는 리가제 유형에 따라 다를 수 있다. 또한, 반응에서 구성요소가 인접하거나 혼성화되는 속도는 해당 핵산 서열에 따라 다를 수 있다. 배양 온도가 높을수록 확산 속도가 빨라지고 구성요소가 일시적으로 인접하거나 혼성화되는 빈도가 높아진다. 그러나 온도가 증가하면 염기쌍 사이의 수소 결합이 파괴되어 인접하거나 혼성화된 구성요소 이중체의 안정성이 감소할 수도 있다. 결찰을 위한 최적의 온도는 조립할 핵산의 수, 해당 핵산의 서열, 리가아제 유형 및 반응 첨가제와 같은 기타 요인에 따라 달라질 수 있다. 예를 들어, 4개 염기의 상보적인 오버행이 있는 두 개의 점착 말단 구성요소는 T4 리가제를 사용하는 25°C보다 T4 리가제를 사용하는 4°C에서 더 빠르게 조립될 수 있다. 그러나 25개 염기의 상보적 오버행이 있는 두 개의 점착 말단 구성요소는 T4 리가제를 사용하는 4°C에서보다 T4 리가제를 사용하는 섭씨 2도에서 더 빠르게 조립될 수 있으며 어떤 온도에서든 4-염기 오버행을 사용하는 결찰보다 더 빠를 수 있다. 결찰의 일부 구현예에서, 리가아제 첨가 전에 어닐링을 위한 구성요소를 가열하고 서서히 냉각시키는 것이 유익할 수 있다.

[0310] 결찰은 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 핵산을 조립하는 데 사용될 수 있다. 결찰 배양 시간은 최대 30초, 1분, 2분, 5분, 10분, 20분, 30분, 1시간 또는 그 이상일 수 있다. 배양 시간이 길수록 결찰 효율성이 향상될 수 있다.

[0311] 결찰에는 5' 인산화된 말단을 가진 핵산이 필요할 수 있다. 5' 인산화 말단이 없는 핵산 구성요소는 T4 폴리뉴클레오티드 키나제(또는 T4 PNK)와 같은 폴리뉴클레오티드 키나제와의 반응으로 인산화될 수 있다. ATP, 마그네슘 이온 또는 DTT와 같은 다른 보조 인자가 반응에 존재할 수 있다. 폴리뉴클레오티드 키나제 반응은 섭씨 37도에서 30분 동안 발생할 수 있다. 폴리뉴클레오티드 키나제 반응 온도는 적어도 섭씨 4, 10, 20, 20, 30, 40, 50 또는 60도일 수 있다. 폴리뉴클레오티드 키나제 반응 배양 시간은 최대 1분, 5분, 10분, 20분, 30분, 60분 이상일 수 있다. 대안으로, 핵산 구성요소는 변형된 5' 인산화를 사용하여 합성적으로(효소적으로 반대되는) 설계되고 제조될 수 있다. 5' 말단에 조립되는 핵산만 인산화가 필요할 수 있다. 예를 들어, TDL의 템플릿은 조립할 의도가 아니기 때문에 인산화되지 않을 수 있다.

[0312] 결찰 효율을 향상시키기 위해 결찰 반응에 첨가제가 포함될 수 있다. 예를 들어, 디메틸 설폭사이드(DMSO), 폴리에틸렌 글리콜(PEG), 1,2-프로판디올(1,2-Prd), 글리세롤, Tween-20 또는 이들의 조합의 첨가. PEG6000은 특히 효과적인 결찰 강화제일 수 있다. PEG6000은 밀집화제(crowding agent) 역할을 하여 결찰 효율성을 높일 수 있다. 예를 들어, PEG6000은 리가제 반응 용액에서 공간을 차지하고 리가제와 구성요소를 더 가깝게 만드는 응집된 결절을 형성할 수 있다. 첨가물 함량(체적당 중량)은 적어도 0%, 1%, 5%, 10%, 20% 이상일 수 있다.

[0313] 결찰에는 다양한 리가제가 사용될 수 있다. 리가제는 자연적으로 발생하거나 합성될 수 있다. 리가제의 예에는

T4 DNA 리가제, T7 DNA 리가제, T3 DNA 리가제, Taq DNA 리가제, 9<sup>o</sup>N<sup>TM</sup> DNA 리가제, 대장균 DNA 리가제, 및 SplintR DNA 리가제가 포함된다. 상이한 리가제가 다양한 온도에서 안정적이고 최적으로 기능할 수 있다. 예를 들어, Taq DNA 리가제는 열안정성이 있지만 T4 DNA 리가제는 그렇지 않다. 또한, 상이한 리가제는 상이한 특성을 가지고 있다. 예를 들어, T4 DNA 리가제는 평활 말단 dsDNA를 결합할 수 있지만 T7 DNA 리가제는 그렇지 않을 수 있다.

[0314] 결합을 사용하여 서열 분석 어댑터를 핵산 라이브러리에 부착할 수 있다. 예를 들어, 결합은 핵산 라이브러리의 각 구성원의 말단에 있는 공통 점착 말단 또는 스테이플을 사용하여 수행될 수 있다. 핵산 한쪽 말단의 점착 말단이나 스테이플이 다른 쪽 말단의 것과 다른 경우 시퀀싱 어댑터가 비대칭으로 결합될 수 있다. 예를 들어, 순방향 시퀀싱 어댑터는 핵산 라이브러리 구성원의 한쪽 말단에 결합될 수 있고 역방향 시퀀싱 어댑터는 핵산 라이브러리 구성원의 다른 말단에 결합될 수 있다. 대안으로, 평활 말단 결합을 사용하여 평활 말단 이중 가닥 핵산 라이브러리에 어댑터를 부착할 수 있다. 포크 어댑터는 각 말단이 동일한 평활 말단이나 점착 말단(가령, A-꼬리)이 있는 핵산 라이브러리에 어댑터를 비대칭적으로 연결하는 데 사용할 수 있다.

[0315] 결합은 열 불활성화(가령, 65°C에서 적어도 20분 배양), 변성제 첨가 또는 EDTA와 같은 킬레이트제(chelator) 첨가에 의해 억제될 수 있다.

[0316] C - 제한 소화(Restriction Digest). 제한 소화는 제한 엔도뉴클레아제(또는 제한 효소)가 핵산의 동족 제한 부위를 인식하고 이어서 상기 제한 부위를 포함하는 핵산을 절단(또는 소화)하는 반응이다. 유형 I, 유형 II, 유형 III 또는 유형 IV 제한 효소가 제한 소화에 사용될 수 있다. 유형 II 제한 효소는 핵산 분해에 가장 효율적인 제한 효소일 수 있다. 유형 II 제한 효소는 회문형 제한 부위를 인식하고 인식 부위 내의 핵산을 절단할 수 있다. 상기 제한 효소(및 이들의 제한 부위)의 예에는 Aat II(GACGTC), AfeI(AGCGCT), ApaI(GGGCCC), DpnI(GATC), EcoRI(GAATTC), NgeI(GCTAGC) 등이 포함된다. DpnI 및 AfeI과 같은 일부 제한 효소는 중앙의 제한 부위를 절단하여 평활 말단 dsDNA 산물을 남길 수 있다. EcoRI 및 AatII와 같은 다른 제한 효소는 제한 부위를 중심에서 벗어나서 점착 말단(또는 엇갈린 말단)을 갖는 dsDNA 산물을 남긴다. 일부 제한 효소는 불연속적인 제한 부위를 표적으로 삼을 수 있다. 예를 들어, 제한 효소 AlwNI는 제한 부위 CAGNNCTG를 인식하며, 여기서 N은 A, T, C 또는 G일 수 있다. 제한 부위의 길이는 적어도 2, 4, 6, 8, 10 또는 그 이상의 염기 길이일 수 있다.

[0317] 일부 유형 II 제한 효소는 제한 부위 외부의 핵산을 절단한다. 효소는 유형 IIS 또는 유형 IIG 제한 효소로 하위 분류될 수 있다. 상기 효소는 비회문적 제한 부위를 인식할 수 있다. 상기 제한 효소의 예에는 GAAAC를 인식하고 더 하류에 엇갈린 절단 2(동일 가닥) 및 6(반대 가닥) 염기를 생성하는 BbsI이 포함된다. 또 다른 예에는 GGCTC를 인식하고 더 하류에 엇갈린 절단 1(동일 가닥) 및 5(반대 가닥) 염기를 생성하는 BsaI이 포함된다. 상기 제한 효소는 골든 게이트 어셈블리(Golden Gate Assembly) 또는 모듈러 클로닝(MoClo)에 대해 사용될 수 있다. BcgI(유형 IIG 제한 효소)와 같은 일부 제한 효소는 인식 부위의 양쪽 말단에서 엇갈린 절단을 생성할 수 있다. 제한 효소는 인식 부위에서 최소한 1, 5, 10, 15, 20개 또는 그 이상의 염기를 분리하여 핵산을 절단할 수 있다. 상기 제한 효소는 인식 부위 외부에 엇갈린 절단을 생성할 수 있기 때문에 생성되는 핵산 오버행의 서열은 임의로 설계될 수 있다. 이는 생성된 핵산 오버행의 서열이 제한 부위의 서열에 결합되는 인식 부위 내에서 엇갈린 절단을 생성하는 제한 효소와 반대이다. 제한 소화에 의해 생성된 핵산 오버행은 적어도 1, 2, 3, 4, 5, 6, 7, 8개 이상의 염기 길이일 수 있다. 제한효소가 핵산을 절단할 때 생성되는 5' 말단에는 인산염이 포함된다.

[0318] 하나 이상의 핵산 서열이 제한 분해 반응에 포함될 수 있다. 마찬가지로, 제한 소화 반응에서는 하나 이상의 제한 효소가 함께 사용될 수 있다. 제한 소화물에는 칼륨 이온, 마그네슘 이온, 나트륨 이온, BSA, S-아테노실-L-메티오닌(SAM) 또는 이들의 조합을 포함하는 첨가제 및 보조인자가 포함될 수 있다. 제한 소화 반응은 섭씨 37도에서 1시간 동안 배양될 수 있다. 제한 소화 반응은 섭씨 0, 10, 20, 30, 40, 50 또는 60도 이상의 온도에서 배양될 수 있다. 최적의 소화 온도는 효소에 따라 달라질 수 있다. 제한 분해 반응은 최대 1분, 10분, 30분, 60분, 90분, 120분 이상 동안 배양될 수 있다. 배양 시간이 길어지면 소화가 증가할 수 있다.

[0319] D - 핵산 증폭. 핵산 증폭은 중합효소 연쇄반응 즉 PCR을 통해 수행될 수 있다. PCR에서, 시작 핵산 풀(주형 풀 또는 주형이라고 함)은 중합효소, 프라이머(짧은 핵산 프로브), 뉴클레오티드 트리 포스페이트(가령, dATP, dTTP, dCTP, dGTP 및 이의 유사체 또는 변형체), 및 추가 보조인자 및 첨가제, 가령, 베타인, DMSO 및 마그네슘 이온과 조합될 수 있다. 주형은 단일 가닥 또는 이중 가닥 핵산일 수 있다. 프라이머는 주형 풀의 표적 서열을 보완하고 이에 혼성화하기 위해 합성적으로 구축된 짧은 핵산 서열일 수 있다. 프라이머는 주형 풀에서 표적 서열을 포함하는 각각의 식별자 핵산 서열에 결합하여 표적 서열을 포함하는 식별자 핵산 서열만을 선택할 수 있

다. 일반적으로, PCR 반응에는 두 개의 프라이머가 있는데, 하나는 표적 주형의 상단 가닥에 있는 프라이머 결합 부위를 보완하기 위한 것이고, 다른 하나는 첫 번째 결합 부위 하류의 표적 주형의 하단 가닥에 있는 프라이머 결합 부위를 보완하기 위한 것이다. 이들 프라이머가 표적과 결합하는 5'-to-3' 방향은 그들 사이의 핵산 서열을 성공적으로 복제하고 기하급수적으로 증폭시키기 위해 서로 마주해야 한다. "PCR"은 전형적으로 상기 형태의 반응을 구체적으로 지칭할 수 있지만, 이는 또한 임의의 핵산 증폭 반응을 지칭하기 위해 보다 일반적으로 사용될 수도 있다.

[0320] 일부 실시예에서, PCR은 3가지 온도, 즉 용융 온도, 어닐링 온도 및 연장 온도 사이의 순환을 포함할 수 있다. 용융 온도는 이중 가닥 핵산을 단일 가닥 핵산으로 바꾸고 혼성화 생성물 및 2차 구조의 형성을 제거하기 위한 것이다. 일반적으로 용융 온도는 섭씨 95도 이상으로 높다. 일부 실시예에서 용융 온도는 적어도 섭씨 96, 97, 98, 99, 100, 101, 102, 103, 104 또는 105도 이상일 수 있다. 다른 실시예에서 용융 온도는 최대 섭씨 95, 94, 93, 92, 91 또는 90도일 수 있다. 용융 온도가 높을수록 핵산과 그 2차 구조의 해리가 향상되지만, 핵산이나 중합효소의 분해와 같은 부작용이 발생할 수도 있다. 용융 온도는 적어도 1, 2, 3, 4, 5초 또는 그 이상, 예를 들어 30초, 1분, 2분 또는 3분 동안 반응에 적용될 수 있다. 복잡하거나 긴 템플릿을 사용하는 PCR에는 더 긴 초기 용융 온도 단계가 권장될 수 있다.

[0321] 어닐링 온도는 프라이머와 표적 주형 사이의 혼성화 형성을 촉진하기 위한 것입니다. 일부 실시예에서, 어닐링 온도는 계산된 프라이머의 용융 온도와 일치할 수 있다. 다른 실시예에서, 어닐링 온도는 상기 용융 온도의 섭씨 10도 이내일 수 있다. 일부 실시예에서, 어닐링 온도는 섭씨 25, 30, 50, 55, 60, 65, 또는 70도 이상일 수 있다. 용융 온도는 프라이머의 서열에 따라 달라질 수 있다. 프라이머가 길수록 용융 온도가 더 높을 수 있고, 구아닌 또는 시토신 뉴클레오티드 함량이 높은 프라이머는 용융 온도가 더 높을 수 있다. 따라서 특정 어닐링 온도에서 최적으로 조립되도록 의도된 프라이머를 설계하는 것이 가능할 수 있다. 어닐링 온도는 적어도 1초, 5초, 10초, 15초, 20초, 25초 또는 30초 이상 동안 반응에 적용될 수 있다. 어닐링을 보장하기 위해 프라이머 농도는 높거나 포화된 양일 수 있다. 프라이머 농도는 500나노몰(nM)일 수 있다. 프라이머 농도는 최대 1nM, 10nM, 100nM, 1000nM 또는 그 이상일 수 있다.

[0322] 연장 온도는 하나 이상의 중합효소에 의해 촉매되는 프라이머의 3' 말단 핵산 사슬 연장을 시작하고 촉진하기 위한 것이다. 일부 구현예에서, 연장 온도는 중합효소가 핵산 결합 강도, 연장 속도, 연장 안정성 또는 충실도 측면에서 최적으로 기능하는 온도로 설정될 수 있다. 일부 실시예에서 연장 온도는 적어도 섭씨 30도, 40도, 50도, 60도 또는 70도 이상일 수 있다. 어닐링 온도는 적어도 1초, 5초, 10초, 15초, 20초, 25초, 30초, 40초, 50초 또는 60초 이상 동안 반응에 적용될 수 있다. 권장 연장 시간은 예상 신장의 킬로베이스당 약 15 내지 45 초일 수 있다.

[0323] PCR의 일부 실시예에서, 어닐링 온도와 연장 온도는 동일할 수 있다. 따라서 3단계 온도 사이클 대신 2단계 온도 사이클이 사용될 수 있다. 결합된 어닐링 및 연장 온도의 예로는 섭씨 60, 65 또는 72도가 있다.

[0324] 일부 실시예에서, PCR은 하나의 온도 사이클로 수행될 수 있다. 이러한 실시예에는 표적화된 단일 가닥 주형 핵산을 이중 가닥 핵산으로 바꾸는 것이 포함될 수 있다. 다른 실시예에서, PCR은 다중 온도 사이클로 수행될 수 있다. PCR이 효율적이라면, 표적 핵산 분자의 수가 각 주기마다 두 배로 증가하여 원래 주형 풀에서 표적 핵산 주형의 수가 기하급수적으로 증가할 것으로 예상된다. PCR의 효율성이 다를 수 있다. 따라서 매 라운드마다 복제되는 표적 핵산의 실제 비율은 100%보다 많거나 적을 수 있다. 각 PCR 주기마다 돌연변이 및 재조합 핵산과 같은 바람직하지 않은 아티팩트가 도입될 수 있다. 이러한 잠재적인 손상을 줄이기 위해 충실도가 높고 가공성이 높은 중합효소가 사용될 수 있다. 또한 제한된 수의 PCR 주기가 사용될 수 있다. PCR은 최대 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 또는 그 이상의 주기를 포함할 수 있다.

[0325] 일부 구현예에서, 다수의 개별 표적 핵산 서열이 하나의 PCR에서 함께 증폭될 수 있다. 각 표적 서열이 공통 프라이머 결합 부위를 갖는 경우, 모든 핵산 서열은 동일한 프라이머 세트를 사용하여 증폭될 수 있다. 대안으로, PCR은 각각의 별개의 핵산을 표적으로 삼도록 의도된 다수의 프라이머를 포함할 수 있다. 상기 PCR은 멀티플렉스 PCR로 지칭될 수 있다. PCR은 최대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 개별 프라이머를 포함할 수 있다. 여러 개의 서로 다른 핵산 표적을 사용한 PCR에서 각 PCR 주기는 표적 핵산의 상대적 분포를 변경할 수 있다. 예를 들어 균일한 분포가 왜곡되거나 불균일하게 분포될 수 있다. 이러한 잠재적인 손상을 줄이기 위해 최적의 중합효소(가령, 높은 충실도와 서열 견고성을 갖춘)와 최적의 PCR 조건을 사용할 수 있다. 어닐링, 연장 온도 및 시간과 같은 요소가 최적화될 수 있다. 또한 제한된 수의 PCR 주기가 사용될 수 있다.

[0326] PCR의 일부 실시예에서, 주형 내 표적 프라이머 결합 부위에 대한 염기 불일치를 갖는 프라이머를 사용하여 표

적 서열을 돌연변이화할 수 있다. PCR의 일부 구현예에서, 5' 말단에 추가 서열(오버행으로 알려짐)이 있는 프라이머를 사용하여 표적 핵산에 서열을 부착할 수 있다. 예를 들어, 5' 말단에 서열 분석 어댑터를 포함하는 프라이머를 사용하여 서열 분석을 위한 핵산 라이브러리를 준비 및/또는 증폭할 수 있다. 서열 분석 어댑터를 표적으로 삼는 프라이머가 사용되어 특정 서열 분석 기술을 위한 충분한 농축으로 핵산 라이브러리를 증폭할 수 있다.

[0327] 일부 실시예에서, 선형-PCR(또는 비대칭-PCR)이 사용되는데, 여기서 프라이머는 주형의 한 가닥(두 가닥 모두가 아님)만을 표적으로 삼는다. 선형 PCR에서는 각 사이클에서 복제된 핵산이 프라이머에 상보적이지 않으므로 프라이머가 이에 결합하지 않는다. 따라서 프라이머는 각 주기마다 원래 표적 템플릿만 복제하므로 선형(지수적 반대) 증폭이 이루어진다. 선형 PCR의 증폭은 기존(지수) PCR만큼 빠르지는 않지만 최대 수율은 더 높을 수 있다. 이론적으로 선형 PCR의 프라이머 농도는 기존 PCR처럼 주기가 증가하고 수율이 증가하는 제한 요인이 되지 않을 수 있다. 선형 후 지수 PCR(또는 LATE-PCR)은 특히 높은 수율이 가능할 수 있는 선형 PCR의 수정된 버전이다.

[0328] 핵산 증폭의 일부 실시예에서, 용융, 어닐링 및 연장 과정은 단일 온도에서 발생할 수 있다. 이러한 PCR은 등온 PCR로 지칭될 수 있다. 등온 PCR은 프라이머 결합을 위해 완전히 보완된 핵산 가닥을 서로 분리하거나 대체하기 위한 온도 독립적인 방법을 활용할 수 있다. 전략에는 루프 매개 등온 증폭, 가닥 치환 증폭, 헬리카제 의존 증폭 및 니킹 효소 증폭 반응이 포함된다. 등온 핵산 증폭은 최대 섭씨 20, 30, 40, 50, 60 또는 70도 이상의 온도에서 발생할 수 있다.

[0329] 일부 구현예에서, PCR은 샘플 내 핵산의 양을 정량화하기 위해 형광 프로브 또는 염료를 추가로 포함할 수 있다. 예를 들어, 염료는 이중 가닥 핵산에 삽입될 수 있다. 상기 염료의 예는 SYBR Green이다. 형광 프로브는 또한 형광 단위에 부착된 핵산 서열일 수도 있다. 형광 단위는 표적 핵산에 대한 프로브의 혼성화 및 확장 폴리머라제 단위로부터의 후속 변형 시 방출될 수 있다. 상기 프로브의 예에는 Taqman 프로브가 포함된다. 이러한 프로브는 샘플 내 핵산 농도를 정량화하기 위해 PCR 및 광학 측정 도구(여기 및 검출용)와 함께 사용될 수 있다. 이 과정을 정량적 PCR(qPCR) 또는 실시간 PCR(rtPCR)이라고 할 수 있다.

[0330] 일부 실시예에서 PCR은 여러 주형 분자의 풀보다는 단일 분자 주형(단일 분자 PCR이라고 할 수 있는 과정)에서 수행될 수 있다. 예를 들어, 에멀전-PCR(ePCR)은 오일 에멀전 내의 물방울 내에 단일 핵산 분자를 캡슐화하는 데 사용될 수 있다. 물방울은 PCR 시약도 포함할 수 있으며, 물방울은 PCR에 필요한 온도 사이클링이 가능한 온도 제어 환경에 유지될 수 있다. 이러한 방식으로 여러 자체 포함 PCR 반응이 높은 처리량으로 동시에 발생할 수 있다. 오일 에멀전의 안정성은 계면활성제를 사용하면 향상될 수 있다. 액적의 이동은 미세유체 채널을 통한 압력으로 제어될 수 있다. 미세유체 장치는 액적 생성, 액적 분할, 액적 병합, 물질 도입 액적 주입 및 액적 배양에 사용될 수 있다. 오일 에멀전의 물방울 크기는 최소 1피코리터(pL), 10pL, 100pL, 1나노리터(nL), 10nL, 100nL 이상일 수 있다.

[0331] 일부 실시예에서, 단일 분자 PCR은 고체상 기관에서 수행될 수 있다. 예에는 Illumina 고체상 증폭 방법 또는 그 변형이 포함된다. 주형 풀은 고상 기관에 노출될 수 있으며, 여기서 고상 기관은 특정 공간 해상도에서 주형을 고정할 수 있다. 그러면 브리지 증폭이 각 주형의 공간적 인접 내에서 발생할 수 있으며 이에 따라 기관에서 높은 처리량 방식으로 단일 분자가 증폭된다.

[0332] 처리량이 높은 단일 분자 PCR은 서로 간섭할 수 있는 서로 다른 핵산 풀을 증폭시키는 데 유용할 수 있다. 예를 들어, 여러 개의 서로 다른 핵산이 공통 서열 영역을 공유하는 경우 PCR 반응 중에 이 공통 영역을 따라 핵산 간의 재조합이 발생하여 새로운 재조합 핵산이 생성될 수 있다. 단일 분자 PCR은 서로 다른 핵산 서열을 구획화하여 상호 작용할 수 없으므로 이러한 잠재적인 증폭 오류를 방지한다. 단일 분자 PCR은 서열 분석을 위한 핵산을 준비하는 데 특히 유용할 수 있다. 단일 분자 PCR 매트릭스는 템플릿 풀 내 여러 표적의 절대 정량화에도 유용하다. 예를 들어, 디지털 PCR(또는 dPCR)은 별개의 단일 분자 PCR 증폭 신호의 빈도를 사용하여 샘플의 시작 핵산 분자 수를 추정한다.

[0333] PCR의 일부 구현예에서, 핵산 그룹은 모든 핵산에 공통적인 프라이머 결합 부위에 대한 프라이머를 사용하여 비차별적으로 증폭될 수 있다. 예를 들어, 프라이머 결합 부위에 대한 프라이머는 풀의 모든 핵산 측면에 위치한다. 합성 핵산 라이브러리는 일반 증폭을 위해 이러한 공통 부위를 사용하여 생성되거나 조립될 수 있다. 그러나 일부 구현예에서 PCR은 예를 들어 상기 표적화된 핵산 하위세트에만 나타나는 프라이머 결합 부위가 있는 프라이머를 사용하여 풀에서 표적화된 핵산 하위세트를 선택적으로 증폭하는 데 사용될 수 있다. 합성 핵산 라이브러리는 잠재적인 관심 하위 라이브러리에 속하는 핵산이 더 포괄적인 라이브러리로부터의 서브-라이브러리의

선택적 증폭을 위해 모두 해당 가장자리에서 공통 프라이머 결합 부위(하위 라이브러리 내에서는 공통이지만 다른 하위 라이브러리와는 구별됨)를 공유하도록 생성되거나 조립될 수 있다. 일부 구현예에서, PCR은 부분적으로 조립되거나 잘못 조립된(또는 의도하지 않거나 바람직하지 않은) 부산물로부터 완전히 조립되거나 잠재적으로 완전히 조립된 핵산을 선택적으로 증폭시키기 위해 핵산 조립 반응(가령, 결찰 또는 OEPCR)과 조합될 수 있다. 예를 들어, 조립은 전체 조립된 핵산 제품만이 증폭을 위해 필요한 두 개의 프라이머 결합 부위를 포함하도록 각 가장자리 서열의 프라이머 결합 부위와 핵산을 조립하는 것을 포함할 수 있다. 상기 예를 들어, 부분적으로 조립된 생성물은 프라이머 결합 부위가 있는 에지 서열 중 어느 것도 포함하지 않거나 하나만 포함할 수 있으므로 증폭되어서는 안 된다. 마찬가지로 잘못 조립된(또는 의도하지 않았거나 바람직하지 않은) 제품에는 모서리 시퀀스 중 하나만 포함되거나 하나만 포함되거나 두 모서리 시퀀스가 모두 포함되어 있지만 방향이 잘못되었거나 베이스의 양이 잘못되어 분리되어 있을 수 있다. 따라서 잘못 조립된 생성물은 증폭되거나 잘못된 길이의 제품을 생성하도록 증폭되어서는 안 된다. 후자의 경우 잘못된 길이의 증폭된 잘못 조립된 생성물은 아가로스 겔에서 DNA 전기영동 후 겔 추출과 같은 핵산 크기 선택 방법(화학적 방법 섹션 E 참조)을 통해 정확한 길이의 증폭된 완전히 조립된 산물로부터 분리될 수 있다.

[0334] PCR에는 핵산 증폭 효율을 높이기 위해 첨가제가 포함될 수 있다. 예를 들어, 베타인, 디메틸 설펝사이드(DMSO), 비이온성 세제, 포름아미드, 마그네슘, 소 혈청 알부민(BSA) 또는 이들의 조합의 첨가가 있다. 첨가물 함량(체적당 중량)은 적어도 0%, 1%, 5%, 10%, 20% 이상일 수 있다.

[0335] PCR에는 다양한 중합효소가 사용될 수 있다. 중합효소는 자연적으로 발생하거나 합성될 수 있다. 중합효소의 예는  $\Phi$ 29 중합효소 또는 이의 유도체이다. 일부 경우에, 전사효소 또는 리가제(즉, 결합 형성을 촉매하는 효소)가 중합효소와 함께 또는 중합효소의 대안으로서 사용되어 새로운 핵산 서열을 구성할 수 있다. 중합효소의 예로는 DNA 중합효소, RNA 중합효소, 열안정성 중합효소, 야생형 중합효소, 변형된 중합효소, E.coli DNA 중합효소 I, T7 DNA 중합효소, 박테리오파지 T4 DNA 중합효소  $\Phi$ 29 (phi29) DNA 중합효소, Taq 중합효소, Tth 중합효소, Tli 중합효소, Pfu 중합효소 Pwo 중합효소, VENT 중합효소, DEEPVENT 중합효소, Ex-Taq 중합효소, LA-Taq 중합효소, Sso 중합효소 Poc 중합효소, Pab 중합효소, Mth 중합효소 ES4 중합효소, Tru 중합효소, Tac 중합효소, Tne 중합효소, Tma 중합효소, Tca 중합효소, Tih 중합효소, Tfi 중합효소, 백금 Taq 중합효소, Tbr 중합효소, Phusion 중합효소, KAPA 중합효소, Q5 중합효소, Tfl 중합효소, Pfutubo 중합효소, Pyrobest 중합효소, KOD 중합효소, Bst 중합효소, Sac 중합효소, 3' 내지 5' 엑소뉴클레아제 활성을 갖는 Klenow 단편 중합효소 및 이들의 변이, 변형 산물 및 유도체를 포함하나, 이에 한정되지는 않는다. 상이한 중합효소는 상이한 온도에서 안정적이고 최적으로 기능할 수 있다. 또한, 상이한 중합효소는 상이한 특성을 가진다. 예를 들어, Phusion 중합효소와 같은 일부 중합효소는 3'에서 5' 엑소뉴클레아제 활성을 나타낼 수 있으며, 이는 핵산 신장 동안 더 높은 충실도에 기여할 수 있다. 일부 중합효소는 신장(elongation) 동안 주요 서열을 대체할 수 있는 반면, 다른 중합효소는 이를 분해하거나 신장을 중단시킬 수 있다. Taq과 같은 일부 중합효소는 핵산 서열의 3' 말단에 아데닌 염기를 통합한다. 또한 일부 중합효소는 다른 중합효소보다 더 높은 충실도와 진행성을 가질 수 있으며 증폭된 핵산 수율이 최소한의 돌연변이를 갖는 것이 중요하고 개별 핵산의 분포가 증폭 전반에 걸쳐 균일한 분포를 유지하는 것이 중요한 서열 분석 준비와 같은 PCR 응용 분야에 더 적합할 수 있다.

[0336] E - 크기 선택. 특정 크기의 핵산은 크기 선택 기술을 사용하여 샘플에서 선택될 수 있다. 일부 실시예에서, 크기 선택은 겔 전기영동 또는 크로마토그래피를 사용하여 수행될 수 있다. 핵산의 액체 샘플은 고정상 또는 겔(또는 매트릭스)의 한쪽 말단에 로드될 수 있다. 겔의 음극 단자가 핵산 샘플이 로드되는 단자이고 겔의 양극 단자가 반대 단자가 되도록 전압 차이가 겔 전체에 배치될 수 있다. 핵산은 음전하를 띤 인산 골격을 갖고 있기 때문에 겔을 거쳐 양극 말단으로 이동할 수 있다. 핵산의 크기는 겔을 통한 상대적인 이동 속도를 결정할 수 있다. 따라서 다양한 크기의 핵산이 이동하면서 겔에서 분해될 것이다. 전압 차이는 100V 또는 120V일 수 있다. 전압 차이는 최대 50V, 100V, 150V, 200V, 250V 이상일 수 있다. 전압 차이가 클수록 핵산 이동 속도와 크기 분해능이 높아질 수 있다. 그러나 전압 차이가 커지면 핵산이나 겔이 손상될 수도 있다. 더 큰 크기의 핵산을 분리하려면 더 큰 전압 차이가 권장될 수 있다. 일반적인 이주 시간(migration time)은 15분 내지 60분일 수 있다. 이주 시간은 최대 10분, 30분, 60분, 90분, 120분 이상일 수 있다. 전압이 높아지는 것과 유사하게 이동 시간이 길어지면 핵산 분해능이 향상될 수 있지만 핵산 손상이 증가할 수 있다. 더 큰 크기의 핵산을 분리하려면 더 긴 이동 시간이 권장될 수 있다. 예를 들어, 250염기 핵산에서 200염기 핵산을 분리하는 데에는 120V의 전압 차이와 30분의 이동 시간이면 충분할 수 있다.

[0337] 겔 또는 매트릭스의 특성이 크기 선택 과정에 영향을 미칠 수 있다. 겔은 일반적으로 TAE(Tris-acetate-EDTA) 또는 TBE(Tris-borate-EDTA)와 같은 전도성 버퍼액에 분산된 아가로스 또는 폴리아크릴아미드와 같은 고분자 물

겔을 포함한다. 겔 내 물질(가령, 아가로스 또는 아크릴아미드)의 함량(체적당 중량)은 최대 0.5%, 1%, 2%, 3%, 5%, 10%, 15%, 20%, 25% 이상일 수 있다. 함량이 높을수록 이주 속도가 느려질 수 있다. 더 작은 핵산을 분리하려면 더 높은 함량이 바람직할 수 있다. 아가로스 겔은 이중 가닥 DNA(dsDNA)를 해결하는 데 더 좋을 수 있다. 폴리아크릴아미드 겔은 단일 가닥 DNA(ssDNA)를 분석하는 데 더 적합할 수 있다. 바람직한 겔 조성은 핵산 유형 및 크기, 첨가제(가령, 염료, 염색제, 변성 용액 또는 로딩 버퍼액)의 호환성뿐만 아니라 예상되는 다운스트림 적용(가령, 겔 추출 후 결찰, PCR 또는 시퀀싱)에 따라 달라질 수 있다. 아가로스 겔은 폴리아크릴아미드 겔보다 겔 추출이 더 간단할 수 있다. TAE는 TBE만큼 좋은 전도체는 아니지만 추출 과정에서 붕산염(효소 억제제) 잔여물이 하류 효소 반응을 억제할 수 있기 때문에 겔 추출에 더 나을 수도 있다.

[0338] 겔은 SDS(나트륨 도데실 황산염) 또는 요소와 같은 변성 용액을 추가로 포함할 수 있다. 예를 들어 SDS는 단백질을 변성시키거나 잠재적으로 결합된 단백질에서 핵산을 분리하는 데 사용될 수 있다. 요소(urea)가 DNA의 2차 구조를 변성시키는 데 사용될 수 있다. 예를 들어, 요소는 dsDNA를 ssDNA로 변환할 수 있거나 요소는 접힌 ssDNA(가령, 헤어핀)를 접히지 않은 ssDNA로 변환할 수 있다. ssDNA를 정확하게 분리하기 위해 요소-폴리아크릴아미드 겔(TBE를 추가로 포함)을 사용할 수 있다.

[0339] 샘플은 다양한 형식의 겔에 통합될 수 있다. 일부 실시예에서 겔은 샘플을 수동으로 로드할 수 있는 웰을 포함할 수 있다. 하나의 겔에는 여러 핵산 샘플을 실행하기 위한 여러 웰이 있을 수 있다. 다른 실시예에서, 겔은 핵산 샘플(들)을 자동으로 로딩하는 미세유체 채널에 부착될 수 있다. 각 겔은 여러 미세유체 채널의 하류에 있을 수도 있고, 겔 자체가 각각 별도의 미세유체 채널을 차지할 수도 있다. 겔의 크기는 핵산 검출(또는 시각화)의 민감도에 영향을 미칠 수 있다. 예를 들어, 미세유체 채널(가령, 바이오분석기 또는 테이프스테이션) 내부의 얇은 겔 또는 겔은 핵산 검출 감도를 향상시킬 수 있다. 핵산 검출 단계는 올바른 크기의 핵산 단편을 선택하고 추출하는 데 중요할 수 있다.

[0340] 핵산 크기 참조를 위해 래더(ladder)가 겔에 로드될 수 있다. 래더는 핵산 샘플과 비교할 수 있는 다양한 크기의 마커를 포함할 수 있다. 래더마다 크기 범위와 해상도가 다를 수 있다. 예를 들어 50 베이스 래더는 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 및 600 베이스에 마커를 가질 수 있다. 상기 래더는 50 내지 600 염기 크기 범위 내의 핵산을 검출하고 선택하는 데 유용할 수 있다. 래더는 시료 내 다양한 크기의 핵산 농도를 추정하기 위한 표준으로도 사용될 수 있다.

[0341] 겔 전기영동(또는 크로마토그래피) 과정을 촉진하기 위해 핵산 샘플과 래더를 로딩 버퍼액과 혼합할 수 있다. 로딩 버퍼액에는 핵산 이동을 추적하는 데 도움이 되는 염료와 마커가 포함될 수 있다. 로딩 버퍼액은 핵산 샘플이 샘플 로딩 웰(런닝 버퍼액에 잠길 수 있음)의 바닥에 가라앉는 것을 보장하기 위해 실험 버퍼액(가령, TAE 또는 TBE)보다 밀도가 높은 샘플(가령, 글리세롤)을 추가로 포함할 수 있다. 로딩 버퍼액은 SDS 또는 요소와 같은 변성제를 추가로 포함할 수 있다. 로딩 버퍼액은 핵산의 안정성을 향상시키기 위한 시약을 추가로 포함할 수 있다. 예를 들어, 로딩 버퍼액은 뉴클레아제로부터 핵산을 보호하기 위해 EDTA가 포함될 수 있다.

[0342] 일부 실시예에서 겔은 핵산에 결합하고 다양한 크기의 핵산을 광학적으로 검출하는 데 사용될 수 있는 염료를 포함할 수 있다. 염료는 dsDNA, ssDNA 또는 둘 다에 특이적일 수 있다. 상이한 염료는 다양한 겔 물질과 호환될 수 있다. 일부 염료는 시각화하기 위해 광원(또는 전자기파)의 자극이 필요할 수 있다. 광원은 UV(자외선) 또는 청색광일 수 있다. 일부 실시예에서는 전기영동 전에 겔에 염료를 첨가할 수 있다. 다른 실시예에서, 전기영동 후에 겔에 염료가 추가될 수 있다. 염료의 예로는 EtBr(Ethidium Bromide), SYBR Safe, SYBR Gold, 은 염료, 또는 메틸렌 블루가 있다. 예를 들어, 특정 크기의 dsDNA를 시각화하는 신뢰할 수 있는 방법은 SYBR Safe 또는 EtBr 염색과 함께 agarose TAE 겔을 사용하는 것일 수 있다. 예를 들어, 특정 크기의 ssDNA를 시각화하는 신뢰할 수 있는 방법은 메틸렌 블루 또는 실버 염색이 포함된 요소-폴리아크릴아미드 TBE 겔을 사용하는 것일 수 있다.

[0343] 일부 실시예에서, 겔을 통한 핵산의 이동은 전기영동 이외의 다른 방법에 의해 유도될 수 있다. 예를 들어, 중력, 원심분리, 진공 또는 압력을 사용하여 핵산을 겔을 통해 이동시켜 크기에 따라 분리할 수 있다.

[0344] 특정 크기의 핵산은 핵산이 포함된 겔 밴드를 잘라내기 위해 칼날이나 면도기를 사용하여 겔에서 추출할 수 있다. 적절한 광학적 검출 기술과 DNA 사다리를 사용하여 절단이 특정 밴드에서 정확하게 발생하고 절단을 통해서로 다른 바람직하지 않은 크기 밴드에 속할 수 있는 핵산을 성공적으로 제외할 수 있다. 겔 밴드는 버퍼액과 함께 배양되어 용해될 수 있으며, 이에 따라 핵산이 완충 용액으로 방출된다. 열이나 물리적인 교반으로 인해 용해 속도가 빨라질 수 있다. 대안으로, 겔 밴드는 겔 용해를 요구하지 않고 DNA가 버퍼액으로 확산될 수 있을 만큼 충분히 오랫동안 버퍼액에서 배양될 수 있다. 그런 다음, 예를 들어 흡인 또는 원심분리에 의해 버퍼액을

남은 고상 겔로부터 분리할 수 있다. 그런 다음 페놀-클로로포름 추출, 에탄올 침전, 자기 비드 포획 및/또는 실리카 막 흡착, 세척 및 용리와 같은 표준 정제 또는 버퍼액 교환 기술을 사용하여 용액으로부터 핵산을 정제할 수 있다. 이 단계에서는 핵산도 농축될 수 있다.

[0345] 겔 절제의 대안으로, 특정 크기의 핵산이 겔에서 흘러내려 겔에서 분리할 수 있다. 이동하는 핵산은 겔에 내장되어 있거나 겔 끝에 있는 분지(또는 웰)를 통과할 수 있다. 이동 과정은 특정 크기의 핵산 그룹이 유역에 들어갈 때 샘플이 유역에서 수집되도록 시간을 정하거나 광학적으로 모니터링할 수 있다. 수집은 예를 들어 흡인을 통해 이루어질 수 있다. 그런 다음 페놀-클로로포름 추출, 에탄올 침전, 자기 비드 포획 및/또는 실리카 막 흡착, 세척 및 용리와 같은 표준 정제 또는 버퍼액 교환 기술을 사용하여 수집된 용액으로부터 핵산을 정제할 수 있다. 이 단계에서는 핵산도 농축될 수 있다.

[0346] 핵산 크기 선택을 위한 다른 방법에는 질량 분석법 또는 막 기반 여과가 포함될 수 있다. 막 기반 여과의 일부 실시예에서, 핵산은 dsDNA, ssDNA 또는 둘 다에 우선적으로 결합할 수 있는 막(예를 들어 실리카 막)을 통과한다. 막은 적어도 특정 크기의 핵산을 우선적으로 포획하도록 설계될 수 있다. 예를 들어, 막은 20, 30, 40, 50, 70, 90개 또는 그 이상의 염기로 구성된 핵산을 걸러내도록 설계될 수 있다. 상기 막 기반의 크기 선택 기술은 겔 전기영동이나 크로마토그래피만큼 엄격하지 않을 수 있다.

[0347] F - 핵산 포획. 친화성 태깅된 핵산은 핵산 포획을 위한 서열 특이적 프로브로 사용될 수 있다. 프로브는 핵산 풀 내의 표적 서열을 보완하도록 설계될 수 있다. 이어서, 프로브는 핵산 풀과 함께 배양되고 그 표적에 혼성화될 수 있다. 배양 온도는 혼성화를 촉진하기 위해 프로브의 용융 온도보다 낮을 수 있다. 배양 온도는 프로브의 용융 온도보다 섭씨 5, 10, 15, 20, 25도 이상 낮을 수 있다. 혼성화된 표적은 친화성 태그에 특이적으로 결합하는 고체상 기질에 포획될 수 있다. 고체상 기질은 멤브레인, 웰, 컬럼 또는 비드일 수 있다. 여러 차례 세척하면 표적에서 혼성화되지 않은 모든 핵산이 제거될 수 있다. 세척은 세척 중에 표적 서열의 안정적인 고정을 촉진하기 위해 프로브의 용융 온도보다 낮은 온도에서 발생할 수 있다. 세척 온도는 프로브의 용융 온도보다 최대 섭씨 5, 10, 15, 20, 25도 낮을 수 있다. 최종 용리 단계에서는 고체상 기질뿐만 아니라 친화성 태그가 지정된 프로브로부터 핵산 표적을 회수할 수 있다. 용리 단계는 핵산 표적이 용리 버퍼액으로 방출되는 것을 촉진하기 위해 프로브의 용융 온도보다 높은 온도에서 발생할 수 있다. 용리 온도는 프로브의 용융 온도보다 섭씨 5, 10, 15, 20, 25도 이상 높을 수 있다.

[0348] 특정 실시예에서, 고체상 기질에 결합된 올리고뉴클레오티드는 예를 들어 산, 염기, 산화, 환원, 열, 빛, 금속 이온 촉매작용, 치환 또는 제거 화학과 같은 조건에 노출시킴으로써 또는 효소 절단에 의해, 고체상 기질로부터 제거될 수 있다. 특정 구현예에서, 올리고뉴클레오티드는 절단 가능한 연결 모이어티를 통해 고체 지지체에 부착될 수 있다. 예를 들어, 고체 지지체는 표적화된 올리고뉴클레오티드에 대한 공유 부착을 위한 절단 가능한 링커를 제공하도록 기능화될 수 있다. 일부 구현예를 들어, 링커 모이어티는 6개 이상의 원자 길이를 가질 수 있다. 일부 구현예에서, 절단 가능한 링커는 TOPS(two oligonucleotides per synthesis) 링커, 아미노 링커, 또는 광절단 가능한 링커일 수 있다.

[0349] 일부 구현예에서, 비오틴은 스트렙타비딘에 의해 고체상 기질에 고정되는 친화성 태그로 사용될 수 있다. 핵산 포획 프로브로 사용하기 위한 비오틴화된 올리고뉴클레오티드가 설계되고 제조될 수 있다. 올리고뉴클레오티드는 5' 또는 3' 말단에서 비오틴화될 수 있다. 또한 티민 잔기 내부에서 비오틴화될 수도 있다. 올리고의 비오틴 증가는 스트렙타비딘 기질에 대한 더 강력한 포획으로 이어질 수 있다. 올리고의 3' 말단에 있는 비오틴은 PCR 중에 올리고가 확장되는 것을 차단할 수 있다. 비오틴 태그는 표준 비오틴의 변형일 수 있다. 예를 들어, 비오틴 변이체는 비오틴-TEG(트리에틸렌글리콜), 이중비오틴, PC비오틴, DesthioBiotin-TEG, 비오틴아지드 등이 될 수 있다. 이중 비오틴은 비오틴-스트렙타비딘 친화성을 증가시킬 수 있다. 비오틴-TEG는 비오틴 그룹을 TEG 링커에 의해 분리된 핵산에 부착한다. 이는 비오틴이 핵산 프로브의 기능, 예를 들어 표적에 대한 혼성화를 방해하는 것을 방지할 수 있다. 핵산 비오틴 링커도 프로브에 부착될 수 있다. 핵산 링커는 표적에 혼성화되도록 의도되지 않은 핵산 서열을 포함할 수 있다.

[0350] 비오틴화된 핵산 프로브는 표적에 얼마나 잘 혼성화할 수 있는지를 고려하여 설계될 수 있다. 더 높게 설계된 용융 온도를 갖는 핵산 프로브는 표적에 더 강하게 혼성화될 수 있다. 더 긴 핵산 프로브뿐만 아니라 더 높은 GC 함량을 갖는 프로브는 증가된 용융 온도로 인해 더 강하게 혼성화될 수 있다. 핵산 프로브의 길이는 적어도 5, 10, 15, 20, 30, 40, 50 또는 100개 염기 또는 그 이상일 수 있다. 핵산 프로브는 0 내지 100% 사이의 GC 함량을 가질 수 있다. 프로브의 녹는 온도가 스트렙타비딘 기질의 온도 허용 오차를 초과하지 않도록 주의해야 한다. 핵산 프로브는 헤어핀, 동종이량체 및 표적을 벗어난 핵산이 있는 이종이량체와 같은 억제성 2차 구조를

방지하도록 설계될 수 있다. 프로브 용융 온도와 표적을 벗어난 결합 사이에는 상충 관계가 있을 수 있다. 용융 온도가 높고 표적의 결합이 낮은 최적의 프로브 길이와 GC 함량이 있을 수 있다. 합성 핵산 라이브러리는 그 핵산이 효율적인 프로브 결합 부위를 포함하도록 설계될 수 있다.

- [0351] 고체상 스트랩타비딘 기질은 자기 비드일 수 있다. 자기 비드는 자기 스트립이나 플레이트를 사용하여 고정될 수 있다. 자기 스트립 또는 플레이트는 용기와 접촉하여 자기 비드를 용기에 고정시킬 수 있다. 반대로, 자기 스트립 또는 플레이트는 용기 벽에서 용액으로 자기 비드를 방출하기 위해 용기에서 제거될 수 있다. 상이한 비드 특성이 그 적용에 영향을 미칠 수 있다. 비드의 크기는 다양할 수 있다. 예를 들어 비드는 직경이 1~3마이크로미터(um) 사이일 수 있다. 비드의 직경은 최대 1, 2, 3, 4, 5, 10, 15, 20 또는 그 이상의 마이크로미터일 수 있다. 비드 표면은 소수성이거나 친수성일 수 있다. 비드는 차단 단백질, 예를 들어 BSA로 코팅될 수 있다. 사용하기 전에 비드를 세척하거나 차단 용액과 같은 첨가제로 전처리하여 비특이적으로 결합하는 핵산을 방지할 수 있다.
- [0352] 비오틴화된 프로브는 핵산 샘플 풀과 함께 배양 전에 자성 스트랩타비딘 비드에 결합될 수 있다. 이 프로세스를 직접 포획이라고 할 수 있다. 대안으로, 비오틴화된 프로브는 자성 스트랩타비딘 비드를 첨가하기 전에 핵산 샘플 풀과 함께 배양될 수 있다. 이 프로세스를 간접 포획이라고 할 수 있다. 간접 포획 방법은 목표 수율을 향상시킬 수 있다. 짧은 핵산 프로브는 자기 비드에 결합하는 데 더 짧은 시간이 필요할 수 있다.
- [0353] 핵산 샘플과 핵산 프로브의 최적 배양은 프로브의 용융 온도보다 섭씨 1~10도 이상 낮은 온도에서 발생할 수 있다. 배양 온도는 최대 섭씨 5, 10, 20, 30, 40, 50, 60, 70, 80도 이상일 수 있다. 권장되는 배양 시간은 1시간일 수 있다. 배양 시간은 최대 1, 5, 10, 20, 30, 60, 90, 120분 또는 그 이상일 수 있다. 배양 시간이 길수록 포획 효율성이 향상될 수 있다. 비오틴-스트랩타비딘 결합을 허용하기 위해 스트랩타비딘 비드를 첨가한 후 추가로 10분 동안 배양할 수 있다. 이 추가 시간은 최대 1, 5, 10, 20, 30, 60, 90, 120분 또는 그 이상일 수 있다. 배양은 나트륨 이온과 같은 첨가제가 포함된 완충 용액에서 발생할 수 있다.
- [0354] 핵산 풀이 단일 가닥 핵산(이중 가닥과 반대)인 경우 표적에 대한 프로브의 혼성화가 향상될 수 있다. dsDNA 풀에서 ssDNA 풀을 준비하려면 풀에 있는 모든 핵산 서열의 가장자리에 일반적으로 결합하는 하나의 프라이머를 사용하여 선형 PCR을 수행해야 할 수 있다. 핵산 풀이 합성적으로 생성되거나 조립된 경우, 이 공통 프라이머 결합 부위가 합성 설계에 포함될 수 있다. 선형 PCR의 생성물은 ssDNA가 될 것이다. 더 많은 주기의 선형 PCR을 통해 핵산 포획을 위한 더 많은 시작 ssDNA 주형이 생성될 수 있다. PCR의 화학적 방법 섹션 D를 참조할 수 있다.
- [0355] 핵산 프로브가 표적에 혼성화되고 자기 스트랩타비딘 비드에 결합된 후, 비드는 자석에 의해 고정될 수 있으며 여러 차례의 세척이 발생할 수 있다. 비표적 핵산을 제거하는 데는 3 내지 5회 세척이면 충분할 수 있지만, 더 많거나 적은 횟수의 세척이 사용될 수도 있다. 각각의 증분 세척은 비표적 핵산을 추가로 감소시킬 수 있지만 표적 핵산의 수율도 감소시킬 수 있다. 세척 단계 동안 프로브에 대한 표적 핵산의 적절한 혼성화를 촉진하기 위해 낮은 배양 온도가 사용될 수 있다. 섭씨 60, 50, 40, 30, 20, 10 또는 5도 이하의 낮은 온도를 사용할 수 있다. 세척 버퍼액은 나트륨 이온이 포함된 Tris 완충 용액을 포함할 수 있다.
- [0356] 자기 비드 결합 프로브로부터 혼성화된 표적의 최적 용리는 프로브의 용융 온도와 동일하거나 그보다 높은 온도에서 발생할 수 있다. 온도가 높을수록 표적과 프로브의 분리가 촉진된다. 용리 온도는 최대 섭씨 30, 40, 50, 60, 70, 80 또는 90도 이상일 수 있다. 용리 배양 시간은 최대 1, 2, 5, 10, 30, 60분 이상일 수 있다. 일반적인 배양 시간은 약 5분이지만 배양 시간이 길면 수율이 향상될 수 있다. 용리 버퍼액은 물이거나 EDTA와 같은 첨가제가 포함된 트리스 완충 용액일 수 있다.
- [0357] 별개의 부위 세트 중 적어도 하나 이상을 함유하는 표적 서열의 핵산 포획은 이들 부위 각각에 대해 다수의 별개의 프로브를 사용하는 하나의 반응으로 수행될 수 있다. 일련의 개별 부위의 모든 구성원을 포함하는 표적 서열의 핵산 포획은 일련의 포획 반응, 즉 특정 부위에 대한 프로브를 사용하여 각 개별 부위에 대한 하나의 반응으로 수행될 수 있다. 일련의 포획 반응 후 표적 수율은 낮을 수 있지만, 포획된 표적은 이후 PCR을 통해 증폭될 수 있다. 핵산 라이브러리가 합성적으로 설계된 경우, 표적은 PCR용 공통 프라이머 결합 부위를 사용하여 설계될 수 있다.
- [0358] 합성 핵산 라이브러리는 일반 핵산 포획을 위한 공통 프로브 결합 부위를 사용하여 생성되거나 조립될 수 있다. 이러한 공통 사이트는 조립 반응에서 완전히 조립되었거나 잠재적으로 완전히 조립된 핵산을 선택적으로 캡처하여 부분적으로 조립되거나 잘못 조립된(또는 의도하지 않았거나 바람직하지 않은) 부산물을 필터링하는 데 사용

될 수 있다. 예를 들어, 조립은 완전히 조립된 핵산 제품만이 각 프로브를 사용하여 일련의 두 가지 포획 반응을 통과하는 데 필요한 필수 두 개의 프로브 결합 부위를 포함하도록 각 모서리 서열에 프로브 결합 부위가 있는 핵산을 조립하는 것을 포함할 수 있다. 상기 예를 들어, 부분적으로 조립된 제품은 프로브 부위 중 어느 것도 포함하지 않거나 하나만 포함할 수 있으므로 궁극적으로 포획되지 않을 것이다. 마찬가지로 잘못 조립된(또는 의도하지 않았거나 바람직하지 않은) 제품에는 가장자리 시퀀스가 하나도 없거나 하나만 포함되어 있을 수 있다. 따라서, 상기 잘못 조립된 제품은 최종적으로 포획되지 않을 수 있다. 엄격함을 높이기 위해 어셈블리의 각 구성요소에 공통 프로브 결합 부위를 포함할 수 있다. 각 구성요소에 대한 프로브를 사용하는 일련의 후속 핵산 포획 반응에서는 조립 반응의 부산물로부터 완전히 조립된 제품(각 구성요소 포함)만 분리할 수 있다. 후속 PCR은 표적 강화를 향상시킬 수 있으며 후속 크기 선택은 표적 엄격성을 향상시킬 수 있다.

[0359] 일부 실시예에서, 핵산 포획은 풀로부터 표적화된 핵산 서브세트를 선택적으로 포획하기 위해 사용될 수 있다. 예를 들어, 상기 표적화된 핵산 서브세트에만 나타나는 결합 부위가 있는 프로브를 사용함으로써 가능하다. 합성 핵산 라이브러리는 잠재적인 관심 하위 라이브러리에 속하는 핵산이 모두 더 일반적인 라이브러리로부터의 서브-라이브러리의 선택적 포획을 위해 공통 프로브 결합 부위(하위 라이브러리 내에서는 공통이지만 다른 하위 라이브러리와는 구별됨)를 공유하도록 생성되거나 조립될 수 있다.

[0360] G - 동결건조. 동결건조는 탈수 프로세스이다. 핵산과 효소 모두 동결건조될 수 있다. 동결건조된 물질은 수명이 더 길 수 있다. 화학적 안정제와 같은 첨가제는 동결건조 공정을 통해 기능성 제품(가령, 활성 효소)을 유지하는 데 사용될 수 있다. 수크로스, 트레할로스 등의 이당류는 화학적 안정제로 사용될 수 있다.

[0361] H - DNA 설계. 합성 라이브러리(가령, 식별자 라이브러리)를 구축하기 위한 핵산(가령, 구성요소)의 서열은 합성, 시퀀싱 및 조립 복잡성을 방지하도록 설계될 수 있다. 더욱이, 합성 라이브러리를 구축하는 비용을 줄이고 합성 라이브러리가 저장될 수 있는 수명을 향상시키도록 설계될 수 있다.

[0362] 핵산은 합성하기 어려울 수 있는 긴 문자열의 단일중합체(또는 반복되는 염기 서열)를 방지하도록 설계될 수 있다. 핵산은 길이가 2, 3, 4, 5, 6, 7 이상인 단독중합체를 피하도록 설계될 수 있다. 더욱이, 핵산은 합성 과정을 방해할 수 있는 헤어핀 루프와 같은 2차 구조의 형성을 방지하도록 설계될 수 있다. 예를 들어, 예측 소프트웨어를 사용하여 안정한 2차 구조를 형성하지 않는 핵산 서열을 생성할 수 있다. 합성 라이브러리를 구축하기 위한 핵산은 짧게 설계될 수 있다. 길이가 긴 핵산은 합성하기가 더 어렵고 비용이 많이 들 수 있다. 핵산이 길수록 합성 중에 돌연변이가 발생할 확률이 더 높아질 수도 있다. 핵산(예를 들어, 구성요소)은 최대 5, 10, 15, 20, 25, 30, 40, 50, 60개 이상의 염기일 수 있다.

[0363] 조립 반응에서 구성요소가 되는 핵산은 조립 반응을 촉진하도록 설계될 수 있다. OEPCR 및 결찰 기반 조립 반응에 대한 핵산 서열 고려 사항에 대한 자세한 내용은 화학적 방법 섹션 A 및 B를 참조할 수 있다. 효율적인 조립 반응에는 일반적으로 인접한 구성요소 간의 혼성화가 포함된다. 잠재적인 표적외 혼성화를 피하면서 이들 표적 내 혼성화 사건을 촉진하도록 서열을 설계할 수 있다. 잠금 핵산(LNA)과 같은 핵산 염기 변형을 사용하여 표적 혼성화를 강화할 수 있다. 이들 변형된 핵산은 예를 들어 스테이플 가닥 결찰에서 스테이플로 또는 점착성 가닥 결찰에서 점착 말단으로 사용될 수 있다. 합성 핵산 라이브러리(또는 식별자 라이브러리)를 구축하는 데 사용될 수 있는 다른 변형된 염기에는 2,6-디아미노퓨린, 5-브로모 dU, 테옥시우리딘, 역전된 dT, 역전된 디테옥시-T, 디테옥시-C, 5-메틸 dC, 테옥실노신, Super T, Super G 또는 5-니트로인돌을 포함한다. 핵산은 동일하거나 다른 변형된 염기 중 하나 또는 여러 개를 포함할 수 있다. 상기 변형된 염기 중 일부는 용융 온도가 더 높은 천연 염기 유사체(가령, 5-메틸 dC 및 2,6-디아미노퓨린)이므로 조립 반응에서 특정 혼성화 사건을 촉진하는 데 유용할 수 있다. 상기 변형된 염기 중 일부는 모든 천연 염기에 결합할 수 있는 범용 염기(가령, 5-니트로인돌)이므로 바람직한 결합 부위 내에 가변 서열을 가질 수 있는 핵산과의 혼성화를 촉진하는 데 유용할 수 있다. 조립 반응에서의 유의한 역할 외에도, 이들 변형된 염기는 핵산 풀 내에서 표적 핵산에 대한 프라이머 및 프로브의 특이적 결합을 촉진할 수 있으므로 프라이머(가령, PCR용) 및 프로브(가령, 핵산 포획용)에 유용할 수 있다. 핵산 증폭(또는 PCR) 및 핵산 포획에 관한 추가 핵산 설계 고려 사항은 화학적 방법 섹션 D 및 F를 참조할 수 있다.

[0364] 핵산은 시퀀싱을 용이하게 하도록 설계될 수 있다. 예를 들어, 핵산은 2차 구조, 단독중합체의 연장, 반복적 서열, GC 함량이 너무 높거나 낮은 서열과 같은 일반적인 서열 분석 문제를 방지하도록 설계될 수 있다. 특정 시퀀서 또는 시퀀싱 방법이 오류에 취약할 수 있다. 합성 라이브러리(예를 들어, 식별자 라이브러리)를 구성하는 핵산 서열(또는 구성요소)은 서로 특정 해밍 거리를 갖도록 설계될 수 있다. 이러한 방식으로, 염기 분해 오류가 시퀀싱에서 높은 비율로 발생하는 경우에도 오류가 포함된 서열의 범위는 여전히 가장 가능성이 높은 핵산

(또는 구성요소)에 다시 매핑될 수 있다. 핵산 서열은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개 이상의 염기 돌연변이의 해밍 거리로 설계될 수 있다. 해밍 거리로부터의 대체 거리 측정법을 사용하여 설계된 핵산 사이의 최소 필수 거리를 정의할 수도 있다.

[0365] 일부 시퀀싱 방법 및 장비에서는 어댑터 서열이나 프라이머 결합 부위와 같은 특정 서열을 포함하는 입력 핵산이 필요할 수 있다. 이들 시퀀싱은 "방법-특이적 서열"로 지칭될 수 있다. 상기 시퀀싱 기기 및 방법에 대한 일반적인 준비 작업 흐름에는 방법-특이적 서열을 핵산 라이브러리에 조립하는 작업이 포함될 수 있다. 그러나 합성 핵산 라이브러리(가령, 식별자 라이브러리)가 특정 기기 또는 방법을 사용하여 서열 분석될 것이라는 것이 미리 알려진 경우, 이러한 방법-특이적 서열은 라이브러리(가령, 식별자 라이브러리)를 포함하는 핵산(가령, 구성요소)로 설계될 수 있다. 예를 들어, 시퀀싱 어댑터는 합성 핵산 라이브러리의 구성원이 개별 핵산 구성요소로부터 조립될 때와 동일한 반응 단계에서 합성 핵산 라이브러리의 구성원에 조립될 수 있다.

[0366] 핵산은 DNA 손상을 촉진할 수 있는 서열을 방지하도록 설계될 수 있다. 예를 들어, 부위 특이적 뉴클레아제 부위를 포함하는 서열은 피할 수 있다. 또 다른 예로서, UVB(자외선-B) 광은 인접한 티민이 피리미딘 이량체를 형성하게 하여 시퀀싱 및 PCR을 억제할 수 있다. 따라서 합성 핵산 라이브러리를 UVB에 노출된 환경에 보관하려는 경우 인접한 티민(즉, TT)을 피하도록 핵산 서열을 설계하는 것이 유리할 수 있다.

[0367] 화학적 방법 섹션에 포함된 모든 정보는 여기에 설명된 기술, 방법, 프로토콜, 시스템 및 프로세스를 지원하고 활성화하기 위한 것이다.

[0368] **아지드-알킨 변형이 있는 구성요소로부터 식별자를 조립하는 예시적 방법**

[0369] 2개 이상의 핵산 구성요소를 함께 결합하여 화학적 및/또는 생물학적 결합 방법을 사용하여 식별자를 생성할 수 있다. 일부 구현예에서, 효소 결합과 같은 생물학적 방법에 비해 "클릭 화학"과 같은 화학적 결합 방법에는 이점이 있을 수 있다.

[0370] 클릭 화학 또는 CuAAC(Copper-Catalyzed Azide-Alkyne Cycloaddition)는 Huisgen 1,3-쌍극성 고리 첨가 반응의 변형이다. 이 반응에서 알킨과 아지드 그룹이 반응하여 트리아졸 포스포디에스테르 모방물을 형성할 수 있다. 현재 방법은 Cu(I) 이온을 사용하여 이 반응의 특이성, 속도 및 수율을 높인다. 반응은 약 1분의 반응 완료 시간을 보고하는 일부 알킨으로 인해 빠를 수 있다. 반응 시간은 30, 60, 90, 120, 150 또는 180초 이상이 될 수 있다. 반응은 또한 강력하여 넓은 pH 범위에 대한 내성을 나타낼 수도 있다.

[0371] 클릭 화학을 사용한 화학적 결합은 주형(또는 스테이플 또는 부목) 올리고뉴클레오티드의 도움으로 두 개의 단일 가닥 핵산 구성요소 사이에서 발생할 수 있다. 대안으로, 공통적으로 상보적 오버행(또는 점착 말단)이 있는 경우 이중 가닥 핵산 구성요소 사이에 화학적 결합이 발생할 수도 있다. 클릭 화학을 이용한 화학적 결합은 전술된 곱 방식(도 62), 순열 방식(도 67), MchooseK 방식(도 68), 분할 방식(도 69) 또는 비제한 스트링 방식(도 70)에 따라 식별자를 구성하는 데 사용될 수 있다.

[0372] 클릭 화학을 사용하여 구성요소들을 결합하려면 한 구성요소가 하나 이상의 알킨 기를 갖고 다른 구성요소가 하나 이상의 아지드 기를 가져야 한다. 한 구성요소의 3' 말단이 다른 구성요소의 5' 말단에 결합되도록 상보적 변형이 인접한 구성요소에 위치하는 한 어느 변형이든 하나의 핵산 구성요소의 5' 또는 3' 말단에 위치할 수 있다.

[0373] 여러 가지 다른 유형의 알킨-아지드 결합이 클릭 화학에 사용될 수 있다. PCR과 같은 분자 생물학 방법과 호환되는 알킨-아지드 연결은 식별자 생성에 특히 적합할 수 있다. 특정 식별자 풀이 하나 이상의 알킨-아지드 결합을 포함하는 경우, 식별자는 PCR을 사용하여 자연 형태(염기 사이의 포스포디에스테르 결합 포함)로 복사될 수 있다.

[0374] 식별자를 구성하는 구성요소는 서로 다른 기능을 가진 두 개 이상의 부분으로 나누어질 수 있다. 예를 들어, 각 구성요소는 두 부분으로 구성될 수 있는데, 하나는 데이터 액세스를 위해 핵산 프로브에 혼성화하기 위한 긴 부분이고, 다른 하나는 시퀀싱 판독을 위한 짧은 부분이다. 두 부분은 서로 분리되어 각 가장자리의 식별자에 조립되도록 의도될 수 있으므로 최종 식별자 생성물은 기능적으로 서로 다른 두 개의 영역을 가진다. 한 쪽의 한 영역은 화학적 액세스를 위한 것이고 다른 쪽의 한 영역은 시퀀싱을 위한 것이다.

[0375] 도 78은 식별자의 점착 말단 결합 조립에 대한 이 개념의 예시 개략도를 제공하며, 여기서 각 층의 구성요소는 곱 방식에 따라 함께 모인다. 첫 번째 층은 결합된 2-부분 구성요소로 식별자 조립 프로세스를 핵화하고, 후속 층은 양쪽 가장자리에서 식별자로 조립되는 연결되지 않은 2-부분 구성요소로 구성된다. 점착 말단 위의 심볼은

각자의 순서를 나타낸다. 상이한 심볼이 있는 점착 말단들이 직교한다. 심볼 옆의 별표는 역 보체(reverse complement)를 나타낸다. 예를 들어, 'a'와 'a\*'는 서로 역보체이므로 결합 중에 혼성화하여 생성물을 형성한다.

**[0376] 기본 편집기를 사용하여 식별자를 구축하는 예시적 방법**

**[0377]** 염기 편집기가 사용되어 모 식별자 내의 특정 유전자좌에 위치한 염기를 프로그래밍 방식으로 돌연변이시켜 새로운 식별자를 구성할 수 있다. 하나의 실시예에서, 염기 편집기는 시스토신(C)을 우라실(U)로 전환시키는 시티딘 데아미나제에 융합된 dCas9 단백질일 수 있다. 모 식별자는 가이드 RNA(gRNA)가 결합하기 위한 여러 직교 표적 유전자좌로 설계될 수 있다. 표적 유전자좌에는 해당 유전자좌에 결합된 dCas9-데아미나제의 활성 범위 내에 하나 이상의 시토신이 포함될 수 있다. 활성 범위는 유전자좌 내의 1, 2, 3, 4, 5, 6개 이상의 염기일 수 있다. dCas9-데아미나제 및 특정 유전자좌에 대한 gRNA의 서브세트와 함께 모 식별자를 후속 배양하면 각 표적 유전자좌에서 하나 이상의 시토신에서 우라실로의 돌연변이가 발생할 수 있다. 또한, DNA 중합효소는 우라실을 티민으로 인식하므로, 돌연변이된 식별자에 대해 PCR을 수행하면 상보적인 돌연변이(구아닌에서 아데닌으로)가 발생할 수도 있다. N개의 직교 표적 유전자좌를 갖는 부모 식별자는 dCas9-데아미나제 및 N개의 gRNA의 다양한 서브세트(각각 부모의 개별 유전자좌를 표적으로 함)를 적용하여 2N개의 개별 딸 식별자 서열로 프로그래밍 방식으로 변환될 수 있다. 따라서 이 체계에서 구성된 가능한 식별자의 조합 공간은 N개의 gRNA 입력에 대한 N 비트의 정보를 저장할 수 있다.

**[0378]** 일부 구현예에서, 부모 서열의 임의의 주어진 표적 유전자좌는 증가된 돌연변이 효율을 촉진하기 위해 상부 및 하부 가닥 모두에 표적화된 시토신을 함유할 수 있다. 또한 효율적인 gRNA 타겟팅이 발생하려면 각 유전자좌가 PAM 사이트에 인접해야 합니다. 그러나 PAM 서열은 다양한 조작된 Cas9 변이체의 사용에 따라 달라질 수 있다.

**[0379]** dCas9-데아미나제 융합체는 두 개의 융합된 단백질 사이에 링커 서열을 포함할 수 있다. 효율적인 표적화 돌연변이를 위한 최적의 링커 길이는 16 아미노산 길이일 수 있다. 링커 길이는 적어도 0, 1, 5, 10, 15, 20, 25 이상의 아미노산 길이일 수 있다. 여러 시티딘 탈아미노효소 중 하나를 사용할 수 있다. 시티딘 데아미나제의 예로는 APOBEC1, AID, CDA1 또는 APOBEC3G가 있다. dCas9 대신 활성 Cas9 니카제(nickase)를 사용할 수 있지만 식별자 구성 반응에도 DNA 복구 효소를 포함해야 할 수도 있다.

**[0380]** 염기 편집기를 사용하여 식별자를 구성하는 또 다른 구현예에서, dCas9에 융합된 아데닌 데아미나제(dCas9에 융합된 시티딘 데아미나제와 반대이거나 이에 추가하여)를 사용하여 gRNA에 의해 액세스될 수 있는 모 식별자의 정의된 유전자좌에서 아데닌을 이노신으로 돌연변이화할 수 있다. 이노신은 DNA 중합효소에 의해 구아닌으로 해석된다. 따라서 염기 편집 유전자좌의 PCR은 반대 가닥의 시토신에 대한 상보적인 티민 돌연변이를 초래할 수 있다.

**[0381] DNA에 저장된 정보를 삭제하는 예시적 방법**

**[0382]** 핵산을 사용하여 저장된 데이터를 안정적으로 제거(또는 삭제)하는 기능은 보안, 개인 정보 보호 및 규제상의 이유로 유익할 수 있다. 데이터 삭제에는 핵산 내의 공유 결합을 끊거나, 핵산을 비가역적으로 변형하여 서열 분석 능력을 방해하거나, 되돌릴 수 없는 방식으로 캡슐화 또는 흡착하거나, 더 많은 핵산 또는 기타 물질을 추가하여 원래의 핵산 모음을 읽을 수 없게 하거나 또는 읽기가 불가능하게 하는 것이 포함될 수 있다. 이러한 방법은 선택적 또는 비선택적 방식으로 수행될 수 있다. 선택 프로세스는 삭제 프로세스와 별개일 수 있다. 예를 들어, 식별자 라이브러리로 시작하여 서열 특정 프로브를 사용하여 삭제할 식별자의 서브세트를 풀다운할 수 있다. 또 다른 예로서, 크기 또는 질량 대 전하 비율에 의한 선별 식별자의 정제는 다른 선택적 또는 비선택적 삭제 방법과 함께 수행될 수 있다.

**[0383]** 라이브러리에서 핵산을 삭제하는 선택적 방법에는 삭제를 위한 핵산 하위 집합을 풀다운하기 위한 서열 특이적 프로브의 사용, 하나 이상의 표적 서열을 포함하는 선별된 핵산을 절단하기 위한 CRISPR 기반 방법의 사용, 크기 또는 질량 대 전하 비율에 따라 핵산을 선택하는 정제 기술이 포함된다.

**[0384]** 정보를 인코딩하는 핵산을 라이브러리에서 삭제하는 비선택적 방법에는 초음파 처리, 오토클레이빙, 표백제, 염기, 산, 에티뮴 브로마이드 또는 기타 DNA 변형제 처리, 방사선 조사(가령, 자외선 사용), 연소 및 비특이적 뉴클라아제 분해(시험관 내 또는 생체 내), 가령, DNase I를 이용한 것이 포함된다. 난독화, 은닉화, 핵산의 액세스나 시퀀싱을 물리적으로 보호하기 위해 다른 방법이 사용될 수도 있다. 방법에는 캡슐화, 회석, 원래 핵산을 난독화하기 위한 무작위 핵산 추가, 및 핵산의 다운스트림 시퀀싱을 방지하는 다른 체계의 추가가 포함될 수 있다. 하나의 실시예에서, 핵산에 저장된 데이터는 오류가 발생하기 쉬운 중합효소, 예를 들어 교정 기능이 부족

한 증합효소에 의한 증폭으로 인해 난독화될 수 있다.

[0385] 정의된 가치 기간을 가진 핵산에 저장된 데이터의 경우 특정 시점에 데이터를 자동으로 삭제하는 방법을 사용하는 것이 유리할 수 있다. 예를 들어, 필수 규제 기간 이후 데이터가 삭제되도록 예약될 수 있다. 또 다른 예로, 데이터가 전송 중이고 제 시간에 목적지에 도달하지 못한 경우 데이터가 삭제되도록 예약될 수 있다. 하나의 실시예에서, 핵산의 계획된 결실은 정의된 속도로 또는 특정 시점에 즉시 작용하는 분해제의 사용을 수반할 수 있다. 또 다른 실시예에서, 핵산의 예정된 삭제는 시간이 지남에 따라 분해되는 핵산 캡슐 또는 보호 케이스의 사용을 포함할 수 있다. 또 다른 실시예에서, 핵산은 다양한 속도의 분해를 촉진하기 위해 다양한 온도나 환경에서 보관될 수 있다. 예를 들어 분해 속도를 높이기 위해 고온이나 높은 습도를 사용한다. 또 다른 실시예에서, 핵산은 더 빠른 분해를 위해 덜 안정한 형태로 전환될 수 있다. 예를 들어, DNA는 덜 안정한 RNA로 전환될 수 있다.

[0386] 핵산 결실의 확인은 시퀀싱, PCR 또는 정량적 PCR을 통해 달성될 수 있다.

[0387] **효율적인 랜덤 액세스를 위한 식별자 설계 및 순위화를 위한 예시적 방법**

[0388] 본 명세서에 설명된 시스템 및 방법은 인코딩되고 저장된 정보로부터 임의의 비트 분포의 효율적인 랜덤 액세스 검색을 허용한다. 라이브러리에 있는 식별자의 표적화된 서브셋을 증폭하기 위해 가장자리 층(또는 끝 서열)에 사용되는 구성요소별 프라이머와 함께 데이터가 저장되면 인코딩된 정보의 일부를 효율적으로 검색할 수 있다. 효율적인 액세스는 저장된 데이터로부터 선택된 정보 부분을 검색하는 데 필요한 PCR 단계 수를 줄이는 것이 포함될 수 있다. 예를 들어, 여기에 설명된 방법을 사용하여 저장된 데이터 세트에서 식별자는 L/2 미만의 순차적 PCR 단계로 액세스될 수 있으며, 여기서 L은 식별자를 포함하는 층의 수이다. 식별자 아키텍처와 식별자 순위 시스템은 식별자 풀의 무작위 액세스 속성에 영향을 미친다. 식별자의 순위는 식별자가 나타내는 비트의 위치에 대응한다. 식별자 순위는 전략적으로 정의될 수 있는 각 층에 나타날 수 있는 가능한 각 구성요소의 순서로부터 사전식으로 결정될 수 있다. 예를 들어, 식별자의 가장자리에 있는 층에는 식별자 중간에 있는 층보다 더 높은 우선순위가 할당될 수 있으므로 랜덤 액세스(가령, 식별자의 가장자리 층을 바이인딩하는 PCR 프라이머 사용)가 인코딩된 비트의 연속 또는 관련 스트레치에 대응하는 연속 순위를 갖는 식별자를 반환할 것이다. "우선순위"가 높을수록 액세스 심도가 낮아진다, 가령, 우선순위가 높은 요소는 우선순위가 낮은 요소보다 액세스하기 쉽다.

[0389] 식별자 아키텍처 및 식별자 순위 시스템은 식별자 풀에서 식별자의 특정 서브셋에 대한 무작위 액세스를 허용한다. 일부 구현에서, 식별자 풀의 각 식별자 핵산 서열은 심볼 스트링 내의 심볼 값 및 심볼 위치에 대응한다. 또한, 풀 내의 식별자 핵산 서열의 존재 또는 부재는 심볼 스트링 내의 대응하는 각각의 심볼 위치의 심볼 값을 나타낼 수 있다.

[0390] 특정 구현에서, 인접한 심볼 위치를 갖는 심볼은 유사한 디지털 정보를 인코딩한다. 본 명세서에 사용된 유사한 디지털 정보에는 동일한 구조의 데이터(즉, 이미지 데이터 또는 이진 코드 스트링)가 포함될 수 있다. 유사한 디지털 정보는 해당 정보에 포함된 데이터를 의미할 수도 있다. 예를 들어, 빨간색의 특정 강도로 인코딩된 모든 이미지 데이터 위치는 인접한 심볼 위치에서 함께 그룹화될 수 있다. 대안으로, 연속적인 심볼 위치를 갖는 심볼은 유사한 디지털 정보를 인코딩하지 않을 수도 있다. 예를 들어, 연속적인 심볼 위치는 x 좌표, y 좌표, 강도 값 또는 강도 값 범위와 같은 데이터(즉, 이미지 데이터)의 다양한 특징에 대응할 수 있다. 도 79은 3개 층 A, B, C의 곱 방식에 의해 생성된 식별자의 예를 보여주는데, 여기서 각 층은 2개의 구성요소 1과 2를 가진다. 3개의 층 A, B, C 각각의 구성요소가 해당 순서로 조립된다. 각 식별자의 순위는 각 층에 특정 순서를 할당한 다음 각 층 내의 각 구성요소에 특정 순서를 할당하고 식별자를 사전순으로 정렬하여 결정될 수 있다. 도 79a는 물리적 식별자에서 층이 정렬되는 것과 동일한 방식으로 층의 사전순 정렬을 정의함으로써 얻은 결과 순위를 보여준다. 식별자 풀(가령, 구성요소 A1 및 구성요소 C1)을 결합하는 프라이머를 사용하여 PCR 반응으로 이러한 식별자 풀을 쿼리하는 경우 액세스된 식별자는 비연속적인 순위를 가지므로, 한 번의 PCR 반응으로 연속적인 비트 스트링을 랜덤 액세스하는 것을 불가능하게 만들 수 있다. 본 명세서에 설명된 특정 구현에서, 식별자(예를 들어, 구성요소 A1 및 구성요소 C1)의 가장자리는 "말단 서열" 또는 "말단 분자"로 지칭된다. 그러나 연속된 스트레치 내의 비트는 종종 관련 정보를 인코딩하므로 연속된 비트 스트레치(연속적으로 순위가 매겨진 식별자로 표시됨)에 무작위로 액세스하는 것이 이상적이다. 연속적인 비트 스트레치 내의 각 비트는 프로브를 사용하여 액세스되어 복수의 식별자 핵산 서열 중 각 식별자 핵산 서열의 표적 말단 서열에 혼성되어 연속적인 심볼 위치를 갖는 각각의 심볼에 대응하는 식별자 핵산 서열을 선택할 수 있다. 도 79b는 식별자의 가장자리(또는 말단 서열)를 결합하는 프라이머를 사용하는 한 번의 PCR 반응으로 인접한 비트 스트레치의 질의를 가능

하게 하기 위해 층 A, B 및 C의 사전편찬 순서가 어떻게 변경될 수 있는지를 보여준다. 전략은 층의 물리적 순서와 동일한 사전식 층 순서를 사용하지 않는 것이다. 대신, 식별자의 가장자리(또는 말단 서열)에 있는 층에 더 높은 우선 순위의 사전 편찬 순서를 할당하고 식별자 중간에 있는 층에 더 낮은 우선 순위를 할당하는 것이 전략이다.

[0391] 조합 공간의 기본이 되는 분할 방식의 구성요소 분포는 PCR 반응에서 액세스할 수 있는 심볼 수에 영향을 미칠 수 있다. 도 80는 3개 층 A, B, C의 곱 방식에 의해 생성된 식별자의 예를 보여주며, 여기서 층 전체에 걸쳐 구성요소가 균일하지 않게 분포된다. 구체적으로 두 층에는 두 개의 구성요소 1과 2가 있고, 한 층에는 세 개의 구성요소 1, 2, 3이 있다. 앞서 언급한 식별자 순위 원칙에 따르면, 물리적 순서는 A, B, C이지만 층의 사전 편찬 순서는 A, C, B 순이다. 이는 식별자의 가장자리 층(또는 말단 서퀀스)를 결합하는 PCR 프라이머를 사용한 무작위 액세스가 연속 순위(연속적인 비트 범위에 해당)로 식별자를 반환하도록 하기 위한 것이다. 구체적으로, 특정 식별자 핵산 서열의 첫 번째 및 두 번째 말단 서열은 인접한 비트 스트레치에 대응하는 다중 식별자 핵산 서열 사이에서 공유된다. 도 80a는 더 많은 구성요소가 식별자의 중간 층(들)에 배치될 때 PCR 쿼리(각각 가장자리 구성요소(또는 말단 서열)를 결합하는 프라이머를 사용)로 인해 액세스된 식별자의 더 큰 풀이 생성될 수 있음을 보여준다. 이에 따라 한 번에 더 많은 비트에 액세스할 수 있다. 도 80b는 더 많은 구성요소가 식별자의 가장자리 층(또는 말단 서열(들)에 배치될 때 등가 PCR 쿼리로 인해 액세스된 식별자의 풀이 더 작아질 수 있음을 보여준다. 이에 따라 비트는 더 높은 해상도로 액세스될 수 있다.

[0392] 식별자 구성을 위한 곱 방식의 층의 수는 PCR 쿼리당 액세스할 수 있는 심볼 수에도 영향을 미칠 수 있다. 도 81는 5개 층(A, B, C, D, E)의 곱 방식에 의해 생성된 식별자의 예를 보여주며, 여기서 각 층은 2개의 구성요소(1과 2)를 가진다. 앞서 언급한 식별자 순위 원칙에 더해 층의 사전 편찬 순서는 최외부 층(A 및 E)에 가장 높은 우선순위를 할당하고, 두 번째에서 최외부 층(B 및 D)에 다음으로 높은 우선순위를, 중간 층(층 C)에 가장 낮은 우선순위를 할당한다. 본 명세서에서 사용된 바와 같이, 우선순위는 데이터 액세스의 깊이(또는 레벨)를 나타내며, 높은 우선순위는 얕은 깊이에 대응되고 낮은 우선순위는 깊은 깊이에 대응됩니다. 예를 들어, 책 모음에서 책(즉, 층 A 및 E)에 대한 액세스는 가장 높은 우선순위로 간주되고, 책 내의 한 챕터(즉, 층 B 및 D)에 대한 액세스는 다음으로 가장 높은 우선순위로 간주되며, 책의 챕터 내 단락(즉, 층 C)에 대한 액세스는 가장 낮은 우선순위로 간주된다. 더 많은 층이 있는 경우 층의 사전순 정렬은 이러한 방식으로 계속되므로 연속적이거나 관련된 비트 스트레치를 검색하는 데 더 적은 PCR 쿼리를 사용할 수 있다. 최외부 층(A1 및 E1)의 구성요소와 관련된 모든 식별자는 한 번의 PCR 반응으로 쿼리될 수 있다. 그런 다음 두 번째에서 가장 바깥쪽 층(B1 및 D1)의 구성요소를 결합하는 프라이머를 사용하는 추가 PCR 반응을 통해 더 높은 해상도(즉, 더 낮은 우선 순위 또는 더 깊은) 쿼리를 수행할 수 있다. 식별자 아키텍처에 더 많은 층이 있는 경우 순차 PCR 반응은 이러한 방식으로 계속되어 더 높은 해상도의 쿼리를 얻을 수 있다. 그러나 두 가지 순차 PCR 반응을 사용하여 A1, B1, D1 및 E1의 4개 구성요소와 관련된 모든 식별자를 쿼리하는 대신 사용할 수 있다. (특히 구성요소가 충분히 짧은 서열을 갖도록 설계된 경우) PCR 프라이머가 A1-B1과 E1-D1을 함께 결합하도록 설계될 수 있지만 그 자체로는 어느 구성요소도 결합하지 않아 결과 PCR 쿼리가 A1과 E1에 이어 B1과 D1이 순차적으로 PCR 쿼리된 것과 동일한 식별자이다.

[0393] **DNA 및 다중 빈을 사용하여 정보를 인코딩하는 예시적 방법**

[0394] 정보는 "다중 빈 방식"을 사용하여 DNA 식별자로 인코딩될 수 있다. 이러한 방식의 한 구현에서, b개의 빈(bin)이 있으며, 각각은 식별자의 서로소 집합을 유지한다. 각각의 빈은 라벨 또는 빈 라벨로 지칭될 수 있는 고유

$[\log_2 b]$  비트 심볼로 라벨링된다.  $l$  비트의 비트스트림은  $l / \log_2 b$  "워드"로 분할되며, 각각의 워드는 길이가  $[\log_2 b]$  비트를 가진다. 임의의 워드 w는 빈 라벨일 수 있다.

[0395] 구체적으로, 다중 빈 방식은 "다중 빈 위치 인코딩 방식"일 수 있다. 이 다중 빈 방식에서, 비트스트림에서 각 워드 w의 위치를 나타내기 위해 고유 식별자가 구성되고 라벨 w가 있는 고유 빈에 배치된다. 이 방식의 다중 빈

구현에서는  $l$  비트의 정보를 인코딩하기 위해  $l / \log_2 b$  식별자가 생성되고, 각 비트는 정확히 하나의 빈에 존재하는 정확히 하나의 식별자로 인코딩된다. 우리는 이것을 "다중 빈 위치 인코딩 방식"이라고 지칭한다.

[0396] 앞서 설명한 다중 빈 위치 인코딩 방식이 다음의 예를 통해 설명될 수 있다. 구두점을 포함하여 영어 알파벳의

고유한 심볼로 레이블이 지정된 35개의 빈을 고려할 수 있다. 영어 텍스트 단락의 인코딩은 다음과 같은 방식으로 수행된다. 각 심볼  $x$ 에 대해  $x$ 의 모든 발생은 단락에서 식별된다. 정수 주소는 텍스트의 각 문자에 오름차순으로 번호를 매겨 획득된다. 특정 심볼  $x$ 의 주소에 해당하는 모든 식별자가 생성되어  $x$ 라는 레이블이 붙은 단일 저장소에 수집된다. 따라서  $x$ 가 발생하는 텍스트의 모든 위치는  $x$ 라는 레이블이 붙은 저장소의 식별자로 표시된다.

[0397] 도 82은 다중 빈 위치 인코딩 방식의 예를 도시하며, 여기서 심볼 스트림의 각 유형의 심볼 위치는 해당 유형의 심볼에 대해 예약된 빈에 기록된다. 도면은 1로 라벨링된 "A BEACH CAFÉ" 문구의 예를 보여준다. 이 예에서는 9가지 유형의 심볼 "A", "B", "C", "D", "E", "F", "G", "H", 및 " " (공백을 나타냄)로 구성된 9개의 문자 알파벳을 가정한다. 이 알파벳의 각 심볼에는 해당 심볼에 해당하고 해당 심볼로 이름이 지정된 고유한 빈이 할당된다. 예를 들어 비어 있는 빈 "D"는 라벨 7로 표시된다. 예를 들어, 빈 "F"의 라벨은 라벨 6으로 나타난다. 인코딩될 문구는 알파벳의 심볼로 구분되고 라벨 3에 표시된 대로 식별자 라이브러리와 일대일 대응으로 매핑된다. 심볼이 나타날 때마다 해당 심볼에 대해 예약된 저장소에 해당 식별자가 추가된다. 예를 들어, 빈 A에는 인코딩할 문구("A BEACH CAFÉ", 강조 추가)에 "A" 심볼이 3번 나타나기 때문에 3개의 식별자(라벨 4)가 포함되어 있다. 더욱이, 빈 "A"에 있는 세 개의 식별자는 해당 심볼이 나타나는 위치를 표시한다. 문자 "B"와 "G"가 매핑된 문구("A BEACH CAFÉ")에 나타나지 않기 때문에 저장소 "D"와 "G"는 비어 있다.

[0398] 다중 빈 방식의 또 다른 구현에서,  $l$  비트의 비트스트림은  $1, 2, \dots, b$ 로 라벨링된  $b$ 개의 빈에 대한 식별자 배포에서 암시적으로 인코딩된다. 이 방식에서, 길이가  $l$  비트인 모든 비트스트림 세트와  $b$ 개의 빈으로의 모든  $d$ 개의 식별자 분포 세트 사이에 매핑이 설계된다.  $b$ 개의 빈에 대한  $d$  식별자의 분포는 정수 라벨의 벡터  $(b_1, b_2, \dots, b_d)$ 에서  $0 \leq b_i < b$ 이며: 각각의 비음성 정수  $b_i$ 가  $i$ 번째 식별자에 할당된 고유 빈의 라벨이다. 할당된 각각의 빈 라벨은  $b$ 개의 가능한 라벨 중에서 자유롭게 선택될 수 있으므로  $b^d$ 개의 가능한 분포가 있다.

[0399] 도 83은 정보 인코딩을 위한 식별자 분포의 사용에 기초한 다중 빈 방식의 예를 도시한다. 도 83은 두 개의 식별자(1로 라벨링됨)로 구성된 식별자 라이브러리와 세 개의 명명된 빈(0, 1, 2)으로 구성된 빈 컬렉션의 예를 보여준다. 반의 각 행(각 행은 명명된 세 개의 빈 0, 1, 2로 구성됨)은 세 개의 빈으로 분할된 두 식별자의 분포 예를 보여준다. 표(6으로 라벨링됨)는 고정되어 있지만 각 분포에 매핑된 임의의 비트스트림을 보여준다. 예를 들어, 3개의 빈으로 구성된 네 번째 행(5로 라벨링됨)은 두 개의 식별자가 1로 명명된 빈에 배치되고 0 및 2 빈은 비어 있는 분포를 보여준다. 이 분포는 비트스트림 0011에 임의로 매핑된다. 마찬가지로, 3개 빈의 두 번째 행은 두 식별자가 0과 1로 명명된 빈에 배치되고 세 번째 빈은 비어 있는 분포를 보여준다. 이 분포는 비트스트림 0001(3으로 라벨링됨)에 매핑된다. 다음 행은 1이라는 이름의 빈이 비어 있는 분포를 보여준다. 이는 비트스트림 0010에 대응한다. 그러한 비트스트림이 주어지면 해당 분포가 구성되고 보존된다. 이러한 방식으로, 충분한 수의 빈과 식별자를 사용하여 이 다중 빈 식별자 배포 방식을 사용하여 임의의 비트스트림을 인코딩할 수 있다.

[0400] 다중 빈 방식의 다른 실시예에서, 식별자는 하나보다 많은 빈에 존재할 수 있다. 이 방식에서,  $l$  비트의 비트스트림이  $1, 2, \dots, b$ 로 라벨링된 빈에 대한 식별자 배포에서 암묵적으로 인코딩된다. 이 방식에서 각 빈에는 식별자의 서브세트가 포함되어 있다. 따라서, 이 방식에서, 길이가  $l$  비트인 모든 비트스트림의 세트와 모든 식별자 서브세트의 세트 중 모든  $b$ -서브세트의 세트 간 매핑이 설계된다.  $b$ -서브세트는  $b$ 개의 원소를 포함하는 세트를 의미한다. 예를 들어, 조합 공간에 총  $d$ 개의 식별자가 있는 경우, 모든 식별자 서브세트의 세트는  $2^d$ 개의 세트를 포함하며,  $D$ 로 표시된다. 이 방식은 길이  $l$ 의 모든 비트스트림과  $b$ 개의 세트를 포함하는  $D$ 의 임의의 서브 세트 간 매핑을 사용하며,  $\log_2 2^{db}$  보다 크지 않은 길이의 비트스트림을 인코딩할 수 있다. 또 다른 실시예에서,  $\log_2 \left(\frac{2^d}{b}\right)$  각 빈은 개별 서브세트를 포함하며, 이 경우, 방식은 길이가  $\log_2 \left(\frac{2^d}{b}\right)$  보다 크지 않은 비트스트림을 인코딩할 수 있다.

[0401] 도 84은 정보를 인코딩하기 위한 식별자 분포의 사용에 기초한 다중 빈 방식의 예를 도시하며, 여기서 식별자는 하나보다 많은 빈에 나타날 수 있다. 우리는 이 방식을 재사용이 가능한 식별자 분포(Identifier Distributions with Reuse)라고 부른다. 도 84은 두 개의 식별자(8과 9로 라벨링됨)의 식별자 라이브러리와 세 개의 빈(빈 0,

1, 2)과 관련된 예를 보여준다. 2개의 식별자와 3개의 빈은 6개의 비트( $b_0b_1b_2b_3b_4b_5$ , 여기서 각  $b_x$ 는 비트스트림 내 단일 비트에 대응하고  $x$ 는 비트스트림의 각 비트 위치를 나타냄)를 코딩하는 데 사용된다. 도면의 상단에는 각각 비트  $b_0b_1$ (4로 라벨링됨),  $b_2b_3$  및  $b_4b_5$ 에 해당하는 가능한 식별자의 서브세트가 나와 있다. 식별자의 서브세트는 임의의 빈에 포함될 수 있다. 따라서 3개의 빈의 각각의 빈은 4개의 옵션, 즉 식별자 없음, 단일 식별자(8로 라벨링됨), 다른 식별자(9로 라벨링됨), 또는 두 식별자 모두(8 및 9)를 포함할 수 있다. 이 예에는 세 개의 빈이 포함되어 있으므로 각 서브세트는 각 행(라벨 2)에 세 번씩 표시된다. 세 개의 빈 각각은 정확히 하나의 서브세트를 포함할 수 있지만 모든 서브세트 트리플이 허용된다. 이는 서브세트들을 연결하는 선(라벨 3)으로 표시되는데, 즉, 왼쪽에서 오른쪽으로의 각 경로는 세 개의 저장소에 포함될 서브세트의 모음에 대응한다. 각 식별자 분포는 표(7로 라벨링)에 표시된 것처럼 특정 비트스트림에 매핑된다. 하나의 실시예에서, 비트스트림은 각 빈에 대해 서브세트를 00, 01, 10 및 11로 명명함으로써 추론될 수 있다. 따라서 예를 들어 라벨 5로 표시된 분포는 세 개의 빈 각각에 빈 식별자 서브세트를 포함하도록 선택하고 이 서브세트의 이름은 00이므로 비트스트림 000000에 대응한다. 마찬가지로, 라벨 6에 표시된 분포는 비트스트림 010110에 대응할 것인데, 왜냐하면 이 분포는 빈 0에 서브세트 01, 빈 1에 서브세트 01, 빈 2에 서브세트 10을 포함하도록 선택했기 때문이다. 도면은 64개의 가능한 분포 중 몇 가지 예를 더 보여준다(도면의 점선 항목으로 표시됨).

[0402] 다중 빈 인코딩 방식은 데이터의 보안 보관에 응용할 수 있는데, 왜냐하면 이러한 방식으로 인코딩된 데이터를 디코딩하려면 모든 빈에 대한 액세스 및 디코딩이 필요할 수 있기 때문이다. 예를 들어, 다중 빈 인코딩된 식별자 라이브러리를 소스 비트스트림에 다시 매핑하기 위해, 다중 빈 방식이 비트스트림을 일반적으로 빈의 적절한 서브세트로부터 소스 비트스트림의 임의의 유의미한 서브스트링을 디코딩하는 것을 가능하게 하지 않는 다중 빈의 식별자의 개별 분포에 매핑하므로 각 빈에 존재하는 식별자 세트를 얻는 것이 필요할 수 있다.

[0403] 다른 실시예에서, 소스 비트스트림은 다수의 직교 식별자 라이브러리를 사용하는 다중 빈 방식을 사용하여 인코딩될 수 있다. 결과적인 다중 빈 라이브러리는 일부 최소 카디널리티의 빈의 임의의 서브세트로부터의 디코딩을 가능하게 하는 방식으로 조합될 수 있다. 예를 들어, 소스 비트스트림은 5개의 직교 라이브러리와 각각 3개의 빈을 사용하여 인코딩될 수 있다. 결과적인 15개의 빈은 3개의 빈의 임의의 서브세트로부터 비트스트림의 디코딩을 가능하게 하는 방식으로 결합될 수 있다. 실제로, 빈은 물리적 위치, 가령, 튜브, 웰 또는 기관 상의 스팟일 수 있다.

[0404] 일부 실시예에서, 빈은 물리적 위치, 가령, 튜브, 웰, 또는 기관 상의 스팟일 수 있다. 다른 실시예에서 빈은 특정 바코드 시퀀스와 같이 컬렉션의 모든 식별자에 의해 공유되는 보다 추상적인 연관일 수 있다.

[0405] **DNA 및 정수 파티셔닝을 통한 정보를 인코딩하는 예시적 방법**

[0406] 우리는 DNA의 무작위 서열을 파티셔닝할 때 정보를 저장하는 인코딩 전략을 지칭하기 위해 "정수 분할" 방법이라는 용어를 사용한다. 도 85는 5 단계로 요약된 정수 분할 방법의 실시예를 도시한다. DNA는 회색 또는 검정색 막대와 심볼로 구성된 스트링으로 표시된다. 묘사된 각각의 DNA는 별개의 종을 나타낸다. "중"은 동일한 서열의 하나 이상의 DNA 분자(들)로 정의된다. "중"이 복수 의미로 사용되는 경우, 복수의 종에 포함된 모든 종은 개별 서열을 가지고 있다고 가정할 수 있지만 이는 때때로 "중" 대신 "개별 중"을 써서 명시적으로 나타낼 수 있다.

[0407] 방법 실시예의 단계 1에서, 각각 "카운트"라고 하는 매우 많은 수의 종의 풀로 시작한다. 카운트는 가장자리(검은색과 밝은 회색 막대)에 공통 서열이 있고 중간( $N...N$ )에 개별 서열이 있도록 설계될 수 있다. 축퇴성 올리고뉴클레오타이드 합성 전략을 사용하여 이러한 시작 카운트 풀을 신속하고 저렴한 방식으로 제조할 수 있다. 단계 2에서 카운트가 빈(단계 2에 있는 직사각형)으로 분할된다. 어떤 카운트가 어떤 빈으로 분할되는지는 중요하지 않으며, 중요한 것은 각 빈으로 분할되는 카운트 수이다. 따라서 시작 풀에서 무작위로 단일 카운트를 샘플링한 다음 이를 특정 빈(가령, 단계 2에 있는 5개 빈 중 하나)에 할당하여 분할이 발생할 수 있다. 작은 액적에서 단일 카운트가 풀로부터 샘플링될 수 있다. 빈은 반응 용기이다. 예를 들어, 빈은 기관 상의 미세유체 채널 또는 위치 내 챔버일 수 있다. 카운트는 미세유체 장치를 통해 챔버에 할당되거나 인쇄를 통해 기관의 위치에 할당될 수 있다. 각 빈은 바코드라고 하는 개별 DNA 종을 포함한다. 바코드는 가장자리(밝은 회색 막대와 어두운 회색 막대) 상의 공통 서열 및 각각의 빈을 식별하는 중앙의 개별 서열( $B_0, B_1, B_2, B_3, B_4, \dots$ )을 갖도록 설계될 수 있다. 단계 3에서, 바코드의 공통 가장자리 서열이 카운트의 공통 가장자리 서열로 조립된다. 예를 들어, 바코드의 공통 가장자리 서열은 점착 말단 결찰 또는 겹손 조립을 통해 조립되도록 구성될 수 있다. 단계 4에서 각 빈으로부터의 조립된 DNA 분자가 단계 5로 지시되는 저장을 위해 최종 풀로 통합된다. 최종 풀에서의 종은 카운트가 각 빈에 어떻게 분할되었는지에 대한 모든 정보를 포함한다. 이 정보는 시퀀싱을 통해 복구될 수 있다. 주어진 예에서, 시퀀싱 데이터는 9개의 카운트가 5개의 빈으로 분할되어 첫 번째 빈( $B_0$ )이 2개의 카운트

를 갖고, 두 번째 빈(B1)이 3개의 카운트를 가지며, 세 번째 빈(B2)이 1개의 카운트를 가지며, 네 번째 빈(B3)은 1개의 카운트를 가지며, 다섯 번째 빈(B4)은 2개의 카운트를 가짐을 의미할 수 있다. 이는 정수 "9"를 "구성"으로 알려진 정렬된 합계 "2+3+1+1+2"로 수학적으로 다시 쓰는 것과 같다. 이 방법의 파라미터가 항상 총 9개의 카운트와 5개의 빈을 갖도록 고정된 경우,  $13\text{choose}4$ 개의 가능한 구성이 있으므로 이 예에 기록된 특정 구성에는  $\log_2(13\text{choose}4)$  비트의 정보가 포함된다. 이 프로세스의 어느 시점에서든 저장되는 정보를 방해하지 않고 각 종의 여러 복사본이 존재하거나 생성될 수 있다(가령, PCR을 사용하여). 이를 통해 분해를 방지하고 시퀀싱을 용이하게 하기 위해 최종 풀을 증폭할 수 있다.

[0408] 일반적으로, 정수 파티션 시스템이  $n$ 개의 분할된 카운트 및  $k$ 개의 빈의 고정 파라미터 값을 갖는 경우, 방법은  $\log_2[(n+k-1)\text{choose}(k-1)]$  비트의 정보를 저장하도록 구현될 수 있다. 수학적으로, 정보는 시스템의 "약한 구성"의 수를 측정한다고 말한다. 그러나 이는 각 저장소의 바코드 순서를 알고 있는 경우에만 해당된다. 각 빈

$$\log_2\left[\sum_{j=1}^{j=k} P_j(n)\right]$$

의 바코드 시퀀스를 알 수 없는 경우(가령, 바코드 자체가 무작위 시퀀스인 경우) 방법을 를 저장하도록 구현될 수 있으며, 이때  $P_j(n)$ 은  $n$ 을 정확히  $j$  부분으로 분할한 수이다.

[0409] **DNA에 정보를 인코딩하기 위한 데이터 파이프라인 설계를 위한 예시적 방법**

[0410] DNA에 기록될 입력 비트스트림은 "코텍"으로 약칭되는 컴퓨터 인코딩-디코딩 파이프라인에 의해 처리된다. 도 86은 코텍의 예시적인 인코딩 부분의 상위 레벨 블록도를 도시한다. 소스 비트스트림과 이를 DNA에 기록하라는 요청을 수신하면 코텍은 소스 비트스트림을 블록 크기라고 알려진 고정 길이보다 크지 않은 크기의 하나 이상의 블록으로 나눈다. 코텍은 소스 비트스트림(즉, 심볼 스트림), 처리 요건 및 비트스트림 콘텐츠(즉, 디지털 정보)의 의도된 적용을 기반으로 적절한 블록 크기를 결정한다. 예를 들어, 100 Gbit 비트스트림은 각각 길이가 1Gbit인 100개의 블록, 또는 각각 길이가 100Mbit인 1000개의 블록으로 분할되거나 다른 방식으로 분할될 수 있다.

[0411] 코텍은 하나 이상의 해싱 알고리즘을 사용하여 각 블록의 해시를 계산할 수 있다. 해시 및 기타 메타데이터(가령, 블록 길이, 블록 주소)를 블록에 추가할 수 있다.

[0412] 코텍은 하나 이상의 에러 검출 및 정정 알고리즘을 각 블록에 적용하고 하나 이상의 에러 보호 바이트를 계산할 수 있다. 그런 다음 코텍은 원본 블록을 에러 보호 정보와 결합하여 에러 보호 블록을 얻을 수 있다. 예를 들어, 코텍은 블록의 비트에 콘볼루션 코딩을 적용하고 블록의 바이트 청크에 리드 솔로몬 또는 삭제 코딩을 적용하며 블록의 각 청크에 리드 솔로몬 또는 삭제 오류 방지 바이트를 추가할 수 있다. 코텍은 각 블록에 에러 보호 메타데이터를 추가할 수 있다.

[0413] 에러 보호 정보를 계산할 때 코텍은 에러 보호 계산을 수행하기 위해 특정 대수 필드 크기를 선택할 수 있다. 필드 크기는 소스 워드 길이를 나타낼 수 있으며, 이는 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 64 또는 128비트와 같은 임의의 비트 수일 수 있다. 소스 단어는 소스 비트스트림을 구성하는 연속된 비트 스트림(고정 길이)이다. 코텍은 계산 복잡성과 에러 보호 고려 사항을 기반으로 특정 필드 크기와 단어 길이를 선택할 수 있다. 예를 들어, 8비트 단어 길이는 계산상 효율적일 수 있지만 16비트 단어 길이는 더 나은 에러 보호 기능을 제공할 수 있다. 코텍은 하나 이상의 목적 함수에 기초하여 최적의 파라미터 값 세트를 식별하기 위해 검색 알고리즘을 사용할 수 있다. 예를 들어, 코텍은 기록기 하드웨어 시스템 내의 독립적인 반응 구획의 수, 파라미터 값의 특정 구성, 일부 다른 기능 또는 기능의 일부 조합 하에서 비트스트림을 인코딩하는 데 필요한 고유 식별자의 수를 비용 함수로서 사용할 수 있다.

[0414] 코텍은 쓰기 또는 읽기 성능을 향상시키기 위해 에러 보호 블록에 또 다른 인코딩 단계를 추가로 적용할 수 있다. 코텍은 에러 보호 블록의 각 단어를 새로운 코드워드에 매핑할 수 있다. 코텍은 검색 알고리즘을 사용하여 특정 속성 집합을 가진 코드워드 집합을 생성할 수 있다. 예를 들어, 코텍은 가변 길이이거나 동일한 고정 개수의 "1" 비트 값을 갖는 코드워드, 서로 지정된 해밍 거리를 갖는 코드워드, 또는 이러한 특징의 일부 조합을 생성할 수 있다. 코텍은 최상의 코드워드 길이, 가중치, 해밍 거리 또는 코드워드의 기타 특징을 결정할 때 소스 워드 길이, 기록기 하드웨어 속도 및 사용 가능한 전체 구성요소 수를 포함하는 일련의 파라미터를 사용할 수 있다. 코텍은 이러한 코드워드와 함께 에러 검출 또는 정정 정보의 또 다른 층을 포함할 수 있다. 예를 들어, 코텍은 정확히  $k$ 개의 "1" 비트 값을 갖는 길이  $n$ 의 코드워드를 생성할 수 있으며, 여기서 높은 비트 또는 낮은 비트로 알려진 비트 중 2개가 패리티 비트 역할을 하는데, 패리티 비트가 1일 때 높은 비트가 설정되며, 그렇지 않으면 낮은 비트가 설정된다. 이러한 에러 보호 비트의 하나 이상의 쌍은 코드워드의 다양한 부분을 보호할 수

있다.

[0415] 코텍은 인코딩 또는 디코딩 중에 최적화된 화학적 조건을 보장하기 위해 특정 코드워드 세트를 선택할 수 있다. 예를 들어, 코텍은 기록기 시스템의 각 반응 구획에 고정되고 동일한 수의 식별자가 조립되고 각 구획 내에서 그리고 구획 전체에 걸쳐 거의 동일한 농도로 조립되도록 고정 가중치의 코드워드를 생성할 수 있다. 코텍은 각 반응 구획이 동일한 수의 식별자를 조립하고 정수 개의 코드워드를 인코딩하도록 코드워드 길이 및 분할 방식을 선택할 수 있다.

[0416] 코텍은 여러 세트의 식별자를 사용하여 소스 비트스트림의 일부 또는 모든 비트를 인코딩하도록 선택할 수 있다. 식별자는 직교 식별자 라이브러리에서 나오거나 동일한 식별자 라이브러리에 속할 수 있다. 식별자는 소스 비트스트림 또는 소스 비트스트림으로부터의 비트 조합을 인코딩할 수 있다. 코텍은 비트 조합을 인코딩하는 여러 식별자 세트를 사용하여 모든 비트를 안정적으로 디코딩하는 데 필요한 샘플 크기를 줄일 수 있다.

[0417] 코텍은 각 소스 블록에 대해 하나 이상의 출력 블록을 생성할 수 있다. 출력 블록은 목록 또는 트리를 포함하는 다른 유형의 데이터 구조로 조립될 식별자 세트를 설명할 수 있다. 코텍은 장치에 지정된 식별자를 조합하도록 명령하는 하나 이상의 커맨드 파일을 생성할 수 있다. 예를 들어 코텍은 구성요소가 포함된 인크를 사용하여 액체 처리 로봇이나 인크젯 프린터를 제어하는 명령 파일을 생성할 수 있다. 코텍은 장치와 통신하고 장치의 정보를 기반으로 블록 파일을 최적화할 수 있다. 예를 들어, 장치는 어셈블리 조립 어려움을 보고할 수 있으며 코텍은 에러 보호 성능이 더 높은 새 블록 파일을 생성할 수 있다. 코텍은 블록 파일이나 명령을 파일로 전송하거나 네트워크를 통해 전송할 수 있다. 코텍은 하나 이상의 컴퓨터에서 계산 프로세스를 실행할 수 있다.

[0418] **정보 기록기에게 지시를 특정하는 예시적 방법**

[0419] 우리는 식별자 라이브러리를 구축하는 모든 시스템을 "기록기"라고 지칭한다. 예를 들어, 작성기의 일부 실시예는 인쇄 기반 방법을 사용하여 식별자 구성을 위한 구성요소를 함께 배치할 수 있다. 인쇄 기반 방법은 각각 하나 이상의 핵산 분자를 기관에 인쇄할 수 있는 하나 이상의 인쇄헤드의 사용을 포함할 수 있다.

[0420] 조립할 식별자 라이브러리가 특정되고 사양 파일의 세트를 통해 작성기로 전송된다. 블록 데이터 파일은 작성기가 생성할 식별자 세트를 특정한다. 블록 데이터 파일은 데이터 압축 알고리즘을 이용하여 압축될 수 있다. 블록을 구성하는 식별자는 트리, 트리, 리스트, 비트맵 등 직렬화된 데이터 구조의 형태로 지정될 수 있지만 이에 국한되지는 않는다.

[0421] 예를 들어, 곱 방식을 사용하여 생성될 식별자 라이브러리는 구성요소 라이브러리 파티션 방식(구성요소가 식별자 아키텍처의 층들로 분할되는 방식), 및 각각의 층에서 사용될 가능한 구성요소의 명칭의 목록을 포함하는 블록 메타데이터 파일로 특정될 수 있다. 블록 데이터 파일에는 트리의 루트에서 리프까지의 각 경로가 식별자를 나타내고 경로를 따라 있는 각 노드가 그 식별자의 층에서 사용될 구성요소 명칭을 특정하는 직렬화된 트리 데이터 구조로 구성되어 생성될 식별자가 포함될 수 있다. 블록 데이터 파일은 루트부터 시작하여 각 노드의 왼쪽 자식 노드를 방문하고 노드 자체를 방문한 다음 오른쪽 자식 노드를 방문하는 순서로 트리를 순회함으로써 이 트리의 직렬화로 구성될 수 있다.

[0422] 도 87은 식별자 라이브러리를 표현하기 위한 데이터 구조 및 직렬화의 실시예를 예시한다. 일부 비트스트림을 인코딩하는 식별자 라이브러리가 나타난다(라벨 11). 트리 루트에서 리프까지의 각 경로는 단일 식별자를 나타내며, 식별자의 구성요소는 경로를 따라 만나는 노드의 명칭으로 특정된다. 라벨 6은 주로 구성요소 명칭과 구분 심볼로 구성된 데이터 구조의 직렬화된 표현을 보여준다. 직렬화된 형식은 생성자-특이적 분할 방식(라벨 5)의 사양으로 시작된다. 이 경우 산물 구성은 각 층에 3, 2, 3, 5개의 구성요소를 포함하는 4개의 층으로 사용된다. 직렬화의 나머지 항목은 1로 표시된 것과 같은 데이터 구조의 경로를 스케치한다. 직렬화에서 4로 표시된 세그먼트는 트리의 루트에서 시작하여 첫 번째 층의 노드 0, 두 번째 층의 노드 0, 세 번째 층의 노드 0, 마지막 층의 리프 0으로 내려가는 경로를 스케치한다. 분할 방식은 4개의 층을 갖기 때문에 알고리즘은 이 단계에서 완전한 식별자가 출력될 수 있다고 추론한다. 보다 일반적으로 이 직렬화 세그먼트(7로 라벨링)는 최종 층의 모든 대체 구성요소를 특정한다. 특정 층의 식별자 라이브러리에 포함될 모든 대안이 나열될 때 이 상태를 표시하기 위해 구분 심볼(이 예에서는 마침표)가 직렬화에 포함된다. 그러면 트리의 경로(3으로 표시)에 표시된 대로 알고리즘이 층 위로 올라가도록 트리거된다. 직렬화에서 구성요소 식별자의 다음 세그먼트(16으로 라벨링)는 다음 식별자 집합을 설명한다. 이러한 방식으로, 전체 식별자 라이브러리는 컴팩트한 방식으로 플랫폼 직렬 파일로 표현될 수 있다.

[0423]

[0424] **식별자를 사용한 예시적 계산 방법**

[0425] 화학적 연산을 사용하여 식별자 라이브러리에 인코딩된 데이터에 대한 계산을 수행하는 것이 가능할 수 있다. 이러한 작업은 전체 아카이브의 서브세트 또는 전체 아카이브에서 병렬 방식으로 수행될 수 있으므로 그렇게 하는 것이 유리할 수 있다. 추가로, 계산은 데이터를 디코딩하지 않고 시험관 내에서 수행될 수 있으므로 계산을 허용하면서 비밀성을 보장할 수 있다. 일부 구현에서, AND, OR, NOT, NAND 등과 같은 부울 논리 연산을 포함하는 계산은 각 비트 위치를 나타내는 식별자를 사용하여 인코딩된 비트스트림에서 수행되며, 여기서 식별자의 존재는 '1'의 비트 값을 인코딩하고 식별자가 없으면 비트 값 '0'을 인코딩한다.

[0426] 일부 구현예에서, 모든 식별자는 단일 가닥 핵산 분자로 구성된다(또는 처음에는 이중 가닥 핵산 분자로 구성되었다가 단일 가닥 형태로 분리됨). 임의의 단일 가닥 식별자  $x$ 의 경우 식별자는  $x^*$ 에 의한  $x$ 의 역보체로 표시된다. 임의의 단일 가닥 식별자  $S$  세트에 대해  $S$ 에 있는 각 식별자의 역보체 세트를  $S^*$ 로 표시한다. 라이브러리에 있는 모든 가능한 단일 가닥 식별자 집합을  $U$ 로 표시하고, 역보체 집합을  $U^*$ 로 표시한다. 우리는 이러한 집합을 유니버스와 유니버스\*라고 부른다.  $U_s$  및  $U_s^*$ 는 유니버스와 유니버스\* 세트의 두 번째 쌍을 나타내며, 이러한 세트의 각 식별자는 화학적 방법으로 표적화하거나 선택할 수 있는 검색 영역으로 알려진 추가 핵산 서열로 강화된다.

[0427] 주어진 식별자 라이브러리에 대한 계산은 혼성화 및 절단을 포함하는 일련의 화학적 작업에 의해 구현될 수 있다. 이러한 작업의 추상화는 아래에 설명되어 있다. 각 작업은 식별자 풀을 입력으로 사용하여 작업을 수행하고 식별자 풀을 출력으로 반환한다.

[0428] 도입 예로서, 아래 표와 같이 제1 라이브러리(L1)와 제2 라이브러리(L2)는 각각 8비트를 포함할 수 있다. 두 라이브러리 간의 비트별 "OR" 연산과 두 라이브러리 간의 비트별 "AND" 연산 결과도 표시된다. 화학적 단계에 의해 수행되는 이러한 작업(및 추가 작업)의 세부 사항은 아래에서 더 자세히 설명된다.

비트	L1	L2	OR	AND
b0	1	0	1	0
b1	0	0	0	0
b2	1	1	1	1
b3	1	0	1	0
b4	0	1	1	0
b5	1	1	1	1
b6	0	1	1	0
b7	0	0	0	0

[0429] **표 1**

[0430] 각 라이브러리의 각 비트는 심볼 위치를 포함하는 식별자로 인코딩된다. 심볼 위치에 대한 식별자가 없으면 0을 나타내고 심볼 위치에 대한 식별자가 있으면 1을 나타낸다. 이 예에서 라이브러리의 식별자는 이중 가닥이다.

[0431] 두 라이브러리 L1 및 L2에서 OR 연산을 수행하려면 두 라이브러리 풀이 결합된다. 두 라이브러리의 식별자는 OR 작업을 위해 이중 가닥 상태로 남아 있을 수 있다. OR 연산은 L1 또는 L2에 1이 있는지 여부를 나타내기 때문에 두 풀의 조합은 완전히 결정된 OR 연산 출력이다(위의 OR 열에 표시된 대로). 동일한 심볼 위치에 대해 최대 2배의 식별자 복사본(원래 라이브러리에 비해)이 있으며 이는 여전히 해당 심볼 위치(즉, 심볼 위치

b5)에 1이 있음을 나타낸다. 일부 구현에서, 이중 가닥 식별자는 변성되어 2개의 단일 가닥(즉, 각각의 이중 가닥 식별자에 대해 하나의 센스 또는 "양성" 가닥과 하나의 안티센스 또는 "음성" 가닥)을 생성할 수 있다. 우리는 결과적인 두 개의 상보적인 단일 가닥을 "양성" 및 "음성" 가닥이라고 부른다. 일부 구현에서, 라이브러리의 하위 섹션이 선택될 수 있고, OR 연산이 수행될 수 있으며, OR 연산의 결과는 기존 라이브러리 중 하나 또는 둘 다의 기존 비트 값을 대체할 수 있다.

[0433] 두 개의 라이브러리 L1 및 L2에 대해 AND 연산을 수행하기 위해 이중 가닥 식별자를 먼저 변성하여 두 개의 단일 가닥(즉, 각 이중 가닥 식별자에 대해 하나의 센스 가닥과 하나의 안티센스 가닥)을 생성한다. 다시 한번, 우리는 결과적인 두 개의 상보적인 단일 가닥을 "양성" 및 "음성" 가닥이라고 부른다. 양성 가닥과 음성 가닥은 별도의 풀로 분리된다. 실제로 이는 양성 또는 음성 가닥에 대해 친화성 태그가 지정된 프로브를 사용하여 달성할 수 있다(핵산 포획에 대한 화학적 방법 섹션 F 참조). 식별자는 이러한 목적을 위해 공통 프로브 대상을 포함하도록 설계될 수 있다. 그런 다음 첫 번째 라이브러리의 이중 가닥 식별자의 양성 가닥(가령, 센스 가닥)과 두 번째 라이브러리의 이중 가닥 식별자의 음성 가닥(가령, 안티센스 가닥)이 함께 풀링되어 상보적 단일 가닥이 혼성화된다. 두 라이브러리(가령, 위 표에 표시된 L1 및 L2)에 기존 식별자가 있다고 가정하면, 결과적인 조합 풀은 혼성화가 발생한 후 DNA의 단일 가닥과 DNA의 이중 가닥의 조합을 갖게 된다. 완전 이중 가닥 식별자는 해당 식별자가 첫 번째 라이브러리 L1과 두 번째 라이브러리 L2 모두에 존재했음을 나타낸다. 풀에서 완전히 이중 가닥 식별자를 선택하여 AND 연산 출력을 생성할 수 있다. 예를 들어, 단일 가닥 식별자는 단일 가닥 식별자(및 부분적으로 단일 가닥)를 작은 단위로 절단하기 위해 S1 뉴클레아제 또는 녹두 뉴클레아제와 같은 단일 가닥 특이적 뉴클레아제를 사용하여 선택적으로 제거될 수 있다. 절단으로부터 보호되는 완전한 이중 가닥 식별자는 화학적 방법 섹션 F에 설명된 핵산 포획 기술 또는 화학적 방법 섹션 E에 설명된 크기 선택 기술과 같은 기술을 사용하여 분리될 수 있다. 예를 들어, 완전히 보완된 이중 가닥 DNA만이 특정 길이로 실행되도록 핵산 풀을 크로마토그래피 젤에서 실행할 수 있다. 결합된 풀 출력은 위 표의 AND 열로 표시된다. 이러한 AND 및 OR 연산을 수행하는 데 필요한 단계에 대한 세부 정보 및 추가 예는 아래에 설명되어 있다.

[0434] 본 명세서에 설명된 랜덤 액세스 방법은 라이브러리의 일부를 추출하는 데 사용될 수 있다. 예를 들어, 라이브러리의 서브섹션은 무작위 액세스를 통해 추출될 수 있다. 서브섹션에는 논리 연산(가령, OR 또는 AND)이 적용될 수 있다. 일부 구현에서는 결과적인 식별자 집합이 라이브러리 내 서브섹션의 원래 값을 대체할 수 있다.

[0435] Single(X) 작업은 식별자 풀(이중 가닥 및/또는 단일 가닥)을 가져와 단일 가닥 핵산 식별자만 반환한다(모든 이중 가닥 식별자 제거). double(X) 작업은 식별자 풀(이중 가닥 및/또는 단일 가닥)을 가져와 이중 가닥 식별자만 반환한다(모든 단일 가닥 식별자 제거). make-single(X) 및 make-single\*(X) 작업은 모든 이중 가닥 핵산 식별자를 단일 가닥 형태로 변환한다. (별표가 있는 버전은 음극 가닥을 반환하고 별표가 없는 버전은 양극 가닥을 반환한다.) get(X, q) 작업은 쿼리 q와 일치하는 모든 식별자의 풀을 반환한다. q = "all"인 경우 쿼리는 모든 식별자와 일치하고 작동한다. delete(X, q) 작업은 쿼리 q를 만족하는 모든 식별자(이중 가닥 또는 단일 가닥)를 삭제한다. 쿼리는 앞서 설명한 대로 랜덤 액세스를 통해 구현될 수 있다. Combine(P, Q) 작업은 P 또는 Q의 모든 식별자를 포함하는 풀을 반환한다. Y의 결과를 변수 이름 X에 할당하는 할당(X, Y) 작업을 정의한다. 간결하게 하기 위해 이 작업을 X = Y 형식으로도 표시한다. 할당 작업은 "오염" 문제 없이 변수를 재사용할 수 있는 이상적인 조건에서 실행된다고 가정한다.

[0436] 후속에서, 우리는 길이 l의 비트스트림 a와 b가 각각 이중 가닥 식별자 라이브러리 dsA와 dsB에 기록되었으며, 일부 부분-비트스트림  $s = a_1 \dots a_j$  및  $t = b_1 \dots b_j$ 에 대한 계산에 관심이 있으며, 계산 결과는 부분-비트스트림 s에 저장된다. 즉, initialize(dsA, dsB, s, t) 작업으로 표시되는 다음 작업이 처음에 지정된 순서로 실행되었다고 가정한다:

1	$A = \text{make-single}(dsA)$
2	$A^* = \text{make-single}^*(dsA)$
3	$B = \text{make-single}(dsB)$
4	$B^* = \text{make-single}^*(dsB)$
5	$P = \text{get}(A, "s")$
6	$Q = \text{get}(B^*, "t")$
7	$A = \text{delete}(A, "s")$
8	$B^* = \text{delete}(B^*, "t")$

[0437]

[0438]

도 88는 식별자 라이브러리를 사용한 컴퓨팅을 위한 예시적인 설정을 도시한다. 도면은 추상 트리 데이터 구조(4로 라벨링)로 그려진 식별자의 조합 공간 예시를 보여준다. 이 예에서 트리의 각 수준은 두 구성요소(2로 라벨링) 중에서 선택한다. 트리 루트의 각 경로는 고유 식별자(라벨 3의 예 참조)에 해당하며 순서(또는 순위)를 결정한다. 라벨 4는 단일 가닥 범용 식별자 라이브러리를 보여준다. 라벨 5는 예를 들어 "a"라고 불리는 특정 비트스트림을 인코딩하는 단일 가닥 식별자 라이브러리를 보여준다. 라벨 7은 7비트로 구성된 "s"라고 불리는 "a"의 하위 비트스트림을 보여준다. 마찬가지로, 라벨 10은 동일한 길이의 비트스트림 "b"의 하위 비트스트림 "t"를 보여준다.  $\text{initialize}(dsA, dsB, s, t)$ 를 계산하기 위한 초기화 절차에 설명된 대로 계산할 하위 비트스트림은 폴 P 및 Q(각각 6과 9로 표시됨)에서 사용 가능하고 계산할 준비가 되어 있다.

[0439]

비트스트림 s와 t의 비트의 비트별 논리 결합으로 정의된 연산  $\text{and}(s, t)$ 는 아래 연산 시퀀스를 사용하여 구현될 수 있다.

1	$R = \text{combine}(P, Q^*)$
2	$S = \text{double}(R)$
3	$T = \text{make-single}(S)$
4	$T^* = \text{make-single}^*(S)$
5	$A = \text{combine}(A, T)$
6	$A^* = \text{combine}(A, T^*)$

[0440]

[0441]

비트스트림 s의 비트에 대한 비트별 논리적 부정으로 정의된 연산  $\text{not}(s)$ 는 아래 연산 시퀀스를 사용하여 구현될 수 있다.

1	$R = \text{get}(U^*, "s")$
2	$S = \text{combine}(P, R)$
3	$T = \text{single}(S)$
4	$V = \text{make-single}(T)$
5	$A = \text{combine}(A, V)$
6	$A^* = \text{combine}(A^*, T)$

[0442]

[0443]

비트스트림 s와 t에 있는 비트의 비트별 논리적 분리로 정의된 연산  $\text{or}(s, t)$ 는 아래 연산 시퀀스를 사용하여 구현될 수 있다:

1	$R = \text{get}(B, "t")$
2	$A = \text{combine}(A, R)$
3	$A^* = \text{combine}(A^*, Q^*)$

[0444]

[0445]

일부 구현에서,  $\text{or}(s, t)$  연산은 폴에서 dsA와 dsB를 결합하여  $\text{or}(s, t)$  연산의 출력으로 지칭될 수 있는 식별자의

조합을 생성하는 것을 포함할 수 있다.

[0446] 비트스트림  $s$ 와  $t$ 의 비트 결합에 대한 비트별 논리 부정으로 정의된 연산  $nand(s, t)$ 는 아래 연산 시퀀스를 사용하여 구현될 수 있다.

1	$R = combine(P, Q^*)$
2	$S = single(R)$
3	$T = make-single(S)$
4	$T^* = make-single^*(S)$
5	$A = combine(A, T)$
6	$A^* = combine(A, T^*)$

[0447]

[0448] 하나의 실시예에서,  $single(X)$  연산은  $X$ 로부터의 단일 가닥 식별자가 범용 식별자에 혼성화되도록 먼저  $X$ 를  $U_s$  또는  $U_s^*$  와 조합하는 것을 포함할 수 있다. 더욱이,  $U_s$  및  $U_s^*$ 의 범용 식별자는 특별한 검색 영역을 갖기 때문에, 범용 식별자에 혼성화되는 이러한 분자는 표적화된 방식으로 액세스될 수 있다.

[0449] 하나의 실시예에서,  $double(X)$  연산은  $X$ 의 식별자를 S1 뉴클레아제와 같은 단일 가닥 특정 뉴클레아제로 처리한 다음 생성된 DNA 풀을 겔에서 실행하여 절단되지 않은 식별자만 분리하는 것을 포함할 수 있다(따라서 완전히 이중 가닥이 된다).

[0450] 도 89은 식별자 라이브러리에 의해 인코딩된 비트스트림 " $s$ " 및 " $t$ "에 대해 논리 연산이 어떻게 수행될 수 있는지의 예를 도시한다. 이 도면에서, 계산 중인 풀을 보완하는 범용 라이브러리(14로 라벨링)를 사용한다. AND/NAND라고 라벨링된 컬럼은 비트스트림 " $s$ "와 " $t$ "(각각 5와 7로 라벨링됨)의 결합을 계산할 수 있는 방법을 보여준다. 올바른 범용 라이브러리( $U$  또는  $U^*$ )를 사용하여 풀이 리포매팅되었다고 가정한다. 두 개의 풀이 결합되면 상보적인 단일 가닥 식별자가 혼성화되어 표시된 대로 이중 식별자를 형성한다(가령, 라벨 9). 결과적인 풀(10으로 라벨링)의 이중 가닥 식별자 컬렉션은 AND 계산의 결과를 인코딩한다: 이중 가닥 제품을 분리하면  $and(s, t)$ 의 식별자 라이브러리 표현이 제공된다. 대안으로, 단일 가닥 생성물을 분리하면  $nand(s, t)$ 의 식별자 라이브러리 표현이 제공된다. OR라고 라벨링된 열은 비트스트림 " $s$ "와 " $t$ "의 분리를 계산하는 방법을 보여준다. " $s$ "와 " $t$ "를 나타내는 식별자를 포함하는 풀이 결합되면 결과 라이브러리에는  $or(s, t)$ 의 표현이 포함된다. NOT로 라벨링된 열은 비트스트림 " $s$ "의 부정을 계산하는 방법을 보여준다. 여기서, 비트스트림 " $s$ "를 나타내는 단일 가닥 식별자 라이브러리는 상보적인 범용 식별자 라이브러리(라벨 15)와 조합된다. 결과(19로 라벨링됨)로서, 형성된 모든 이중 가닥 제품(가령, 18로 라벨링됨)은 " $s$ "에서 "1" 비트를 나타내며 폐기될 수 있다. 나머지 단일 가닥 제품(가령, 17로 라벨링됨)은 " $s$ "의 "0" 비트를 나타내므로  $not(s)$ 의 "1" 비트에 대응한다. 이러한 단일 가닥 제품은  $not(s)$ 의 식별자 라이브러리 표현을 제공하며 추가 계산에 사용될 수 있다.

[0451] **이미지 데이터를 인코딩하고 관독하는 예시적 방법**

[0452] 식별자 라이브러리는 인코딩된 비트스트림의 내용에 독립적이지만, 큰 크기와 자연스러운 장기적 사회적 가치로 인해 이미지 데이터를 보관하는 데 특히 유용할 수 있다. 따라서, 그러한 데이터를 위해 특별히 설계된 인코딩 방식 및 포맷을 사용하여 이미지 데이터를 인코딩하는 것이 유용할 수 있다. "이미지 데이터"는 묵시적 또는 명시적으로 어떤 차원의 벡터 집합으로 제시되고 지역성 속성을 갖는 데이터를 지칭한다: 제시된 벡터는 그들 사이의 거리에 대한 개념을 갖고 서로 가까운 벡터를 쿼리하고 연산하며, 또는 함께 해석된다. 예를 들어, 사진 이미지에서 각 픽셀은 픽셀의 위치와 해당 색상 값을 설명하는 벡터이며, 인근 픽셀은 일반적으로 사진에서 하나 이상의 객체 영역을 형성하므로 유닛으로서 해석되고 작동될 가능성이 높다.

[0453] 한 구현예에서, 이미지는 원본 다차원 이미지의 벡터가 공간 채우기 곡선과 같은 수학적 함수에 의해 정의된 선형 순서로 정렬되는 이미지 인코딩 방식을 사용하여 식별자 라이브러리에 매핑된다. 제시된 벡터의 일부 또는 모든 차원에 따른 가능한 값은 구성요소 라이브러리의 특정 구성요소에 매핑될 수 있으며 벡터의 일부 또는 모든 차원은 식별자 구성을 위한 곱 방식 내의 층에 매핑될 수 있다. 우리는 이것을 네이티브 이미지 인코딩 (*native image encoding*)이라고 부른다. 예를 들어, 그레이스케일 이미지의 폭  $x$  픽셀과 높이  $y$  픽셀이 식별자를 구성하기 위한 곱 방식에 매핑될 수 있으며, 여기서, 제1 층의 구성요소가 픽셀의  $x$ -좌표를 나타내고, 제2 층의 구성요소가 픽셀의  $y$ -좌표를 나타내며, 제3 층의 구성요소가 픽셀의 그레이스케일 강도를 나타낸다. 예를

들어, RGB 색상 이미지는 빨간색, 파란색 및 녹색 색상 채널 각각에 대해 하나씩 세 개의 직교 식별자 라이브리리를 사용하여 유사하게 표현될 수 있다. 다른 실시예에서, 색상-채도-값과 같은 다른 대체 색상 모델이 유사하게 표현될 수 있다. 다른 실시예에서, 픽셀의 위치를 지정하는 좌표는 제3 층의 구성요소가 각각 강도 값을 지정하는 대신 각각 강도를 지정하는 비트열의 비트 위치를 나타내는 경우를 제외하고 위에서 설명된 대로 표현될 수 있으며, 여기서 각 구성요소에 대한 식별자의 존재 여부는 각각 '1' 또는 '0' 값을 특정한다. 예를 들어, 전자의 실시예에서 제3 층은 특정 픽셀의 각 구성요소가 256개의 가능한 강도 값 중 1개를 특정하는 256개의 구성요소를 포함할 수 있고, 후자의 실시예에서 제3 층은 8개의 구성요소를 포함할 수 있으며, 여기서 이러한 구성요소의 각 서브세트는 특정 픽셀에서 256개의 가능한 강도 값 중 1개를 특정한다.

[0454] 일부 구현에서는 일부 또는 모든 구성요소가 값 범위와 연관된다. 예를 들어, 색상 값 층(제3 층)의 구성요소는 해당 색상 채널의 색상 값 간격을 나타내도록 정의될 수 있다. 예를 들어, 레드 채널 식별자의 세 번째 층의 각 구성요소는 특정 레드 색상 값이 아닌  $\pm 10$  포인트의 레드 색상 값 범위에 매핑될 수 있다.

[0455] 일부 구현에서, 이미지가 위에서 정의된 대로 인코딩되면 이미지의 임의의 데카르트 섹션(픽셀의 이웃)은 PCR 또는 혼성화 캡처와 같이 이전에 설명된 랜덤 액세스 방식을 사용하여 색상 값에 대해 쿼리될 수 있다. 더욱이, 인코딩 방식이 제3 층의 각 구성요소가 강도 값을 지정하도록 하는 것이라면, 임의의 색상 값은 랜덤 액세스 방식을 사용하여 연관된 픽셀 좌표에 대해 쿼리될 수 있다.

[0456] 일부 구현예에서, 네이티브 이미지 인코딩으로 인코딩된 이미지는 복수의 해상도로 디코딩될 수 있다. 예를 들어, 대략  $3xy$  식별자를 사용하여 RGB 색상 모델로 인코딩된  $x$  픽셀 너비와  $y$  픽셀 높이의 이미지는 절반의 식별자 중 균일하게 무작위인 서브세트를 샘플링함으로써 원래 해상도의 절반으로 디코딩될 수 있다. 원본 이미지의 콘텐츠는 이미지 처리 및 보간 기술을 사용하여 샘플링된 식별자로부터 더 낮은 해상도로 재구성될 수 있다. 이미지를 디코딩하는 데 더 작은 샘플이 사용되므로 디코딩 비용과 시간이 줄어든다.

[0457] 일부 구현에서, 다수의 이미지의 저해상도 디코딩 및 이미지 처리는 아카이브에서 관심 있는 이미지 또는 이미지 섹션을 식별하는 데 사용될 수 있다. 이어서 이러한 이미지 또는 이미지 섹션의 고해상도 디코딩이 이어질 수 있다. 이 기능 세트는 예를 들어 특정 시각적 기능을 찾고 있는 대규모 감시 이미지 아카이브를 분석하는 데 유용할 수 있다. 다른 응용 분야에서는 비디오 아카이브가 정적 이미지 프레임의 대규모 아카이브로 처리될 수 있다. 이 응용분야에서는 랜덤 액세스 및 저해상도 디코딩을 통해 관심 있는 프레임을 식별할 수 있다. 그런 다음 주변 프레임을 더 높은 해상도로 디코딩하여 관심 있는 비디오 세그먼트를 재구성할 수 있다. 이러한 방식으로 대용량 이미지나 비디오 아카이브를 수세기 동안 고밀도로 저장하고 동시에 저렴한 비용으로 쿼리할 수 있다.

[0458] 다음은 이미지 데이터 저장 및 다중 해상도 관독의 예를 설명한다. 압축되지 않은 이미지 파일은 각 식별자 또는 각 식별자의 인접한 그룹이 이미지의 픽셀을 나타내도록 식별자로 인코딩될 수 있다. 예를 들어, 이미지가 각 비트가 두 가지 색상(가령, 흰색 또는 검정색) 중 하나를 가질 수 있는 픽셀인 비트맵으로 저장되면 비트맵의 각 비트는 식별자로 표시될 수 있으며 존재 여부는 해당 식별자는 각각 하나의 색상 또는 다른 색상을 나타낼 수 있다. 이미지를 다시 읽으려면 식별자 라이브리리가 무작위로 샘플링될 수 있다(표준 차세대 시퀀싱 기술에서 예상하는 것처럼). 이미지의 다시 읽기 해상도는 읽기의 샘플 크기를 정의하여 지정할 수 있다. 따라서 이미지의 저해상도 버전은 고해상도 버전보다 저렴한 비용으로 다시 읽을 수 있다. 이는 이미지를 다시 읽는 목적에 미세한 이미지 세부정보가 필요하지 않을 때 유용할 수 있다. 대안으로, 이미지의 저해상도 버전 또는 여러 이미지를 검사하여 더 높은 해상도에서 쿼리(액세스)할 위치를 결정할 수 있다.

[0459] 다중 해상도 제어 다시 읽기의 이러한 원리를 추가로 입증하기 위해 비트맵으로 저장된 개의 예시 이미지(도 90)를 고려한다. 도 90a의 원본 이미지는 1476800 픽셀(1300x1136 픽셀)이며, 각각은 비트(흰색 또는 검정색)로 저장된다. 우리는 각 비트가 식별자이고 이미지가 검은색 픽셀에 대해서만 식별자를 구축하여 인코딩된 경우 어떤 일이 발생하는지 시뮬레이션한다. 이를 위해서는 131820개의 식별자가 필요하다. 도 90b는 전체 식별자 수(샘플 크기 1318200)의 10배에 대한 시뮬레이션 샘플링의 결과 이미지를 보여준다. 원본 이미지와 디테일이 비슷하다. 도 90c는 총 식별자 수(샘플 크기 131820)에 해당하는 숫자를 시뮬레이션하여 샘플링한 결과 이미지를 보여준다. 도 90d는 전체 식별자 수(13182 샘플 크기)보다 10배 적은 식별자의 시뮬레이션된 샘플링으로부터 얻은 결과 이미지를 보여준다. 검은색 픽셀이 너무 희박하기 때문에 이미지를 시각화하기가 어렵다. 원본을 다시 만드는 데 도움이 되도록 각 어두운 픽셀의 크기를 증폭할 수 있다. 도 90e는 각 검정색 픽셀이 25 픽셀로 증폭된 것을 제외하고는 동일한 이미지를 보여준다. 이 해상도에서는 원본 이미지의 일부 세부 사항, 예, 털 가닥이 손실될 수 있다. 그러나 눈과 코와 같은 더 거친 세부 사항은 여전히 볼 수 있다. 도 90f는 전체 식별자 수

(1318 샘플 크기)보다 100배 적은 식별자의 시뮬레이션된 샘플링으로부터 얻은 결과 이미지를 보여준다. 검은색 픽셀이 너무 희박하기 때문에 이미지를 시각화하기가 어렵다. 이번에도 원본을 다시 만드는 데 도움이 되도록 각각의 어두운 픽셀의 크기를 증폭할 수 있다. 도 90g는 각 검정색 픽셀이 25 픽셀로 증폭된 것을 제외하고는 동일한 이미지를 보여준다. 원본 이미지의 많은 세부 사항이 손실되었을 수 있지만 이미지에는 강아지의 모양과 색상 패턴에 대한 일부 세부 정보가 여전히 표시된다.

[0460] 이미지의 각 픽셀에 세 가지 이상의 가능한 색상이 있는 경우에도 동등한 다중 해상도 다시 읽기가 수행될 수 있다. 예를 들어, 각 픽셀이 2개가 아닌 256개의 가능한 색상을 갖는 경우 각 픽셀은 8개의 식별자의 서브셋으로 표시될 수 있다. 각 픽셀이 3개의 색상 채널, 가령, RGB, 각각 256개의 가능한 강도를 갖는 경우 이미지는 각 채널에 해당하는 3개의 직교 식별자 라이브러리와 함께 저장될 수 있다.

[0461] **DNA를 이용한 예시적 데이터 랜덤화, 암호화 및 인증 방법**

[0462] DNA를 사용하여 랜덤 비트스트림을 생성하고 저장하는 능력은 암호화 및 조합 알고리즘의 계산에 응용될 수 있다. DES(Data Encryption Standard)와 같은 많은 암호화 알고리즘에서는 보안을 보장하기 위해 랜덤 비트를 사용해야 한다. AES(Advanced Encryption Standard)와 같은 다른 암호화 알고리즘에는 암호화 키를 사용해야 한다. 일반적으로, 이들 랜덤 비트 및 키는 안전한 랜덤화 소스를 사용하여 생성되는데, 이는 랜덤 비트 또는 키의 체계적 패턴이나 편향이 암호화된 메시지를 공격하고 해독하는 데 악용될 수 있기 때문이다. 또한 암호화에 사용되는 키는 일반적으로 해독을 위해 보관해야 한다. 암호화 방법의 보안 강도는 알고리즘에 사용되는 키의 길이에 따라 달라지는데, 일반적으로 키가 길수록 암호화가 더 강력해진다. 일회용 패드와 같은 방법은 가장 안전한 암호화 방법 중 하나이지만 키 요구 사항이 길기 때문에 적용이 제한된다.

[0463] 이 문서에 설명된 방법은 길이가 수십, 수백, 수천, 수만 또는 그 이상의 비트일 수 있는 매우 큰 임의의 키 컬렉션을 생성하고 보관하는 데 사용될 수 있다. 한 실시예에서, 각각의 핵산 분자가 다음 설계를 만족하는 핵산 라이브러리가 생성될 수 있다: 이는  $k < n$  염기의 가변 영역과 함께  $n$  염기의 길이를 가진다. 가변 영역의 염기는 라이브러리를 구축하는 동안 랜덤으로 선택할 수 있다. 예를 들어,  $n$ 은 100이고  $k$ 는 80일 수 있으며, 따라서 크기가  $10^{50}$ 인 다양한 분자의 라이브러리가 잠재적으로 생성될 수 있다. 예를 들어, 1000개 분자 크기의 라이브러리의 무작위 샘플은 암호화에 사용될 수 있는 최대 1000비트 무작위 키를 얻기 위해 시퀀싱될 수 있다.

[0464] 또 다른 실시예에서, 앞서 설명한 핵산 키(키를 나타내는 핵산 분자)는 키 세트의 정렬된 컬렉션을 생성하는 식별자에 첨부될 수 있다. 순서가 지정된 키 세트는 암호화 컨텍스트에서 다양한 당사자가 키를 사용하는 순서를 동기화하는 데 사용될 수 있다. 예를 들어, 식별자 라이브러리는  $10^{12}$ 개의 고유 식별자를 얻기 위해 제품 체계를 사용하여 조합적으로 구성될 수 있다. 미세유체 방법을 사용하여 각 식별자를 핵산 키와 함께 배치하고 조립하여 고유 식별자와 무작위 키를 포함하는 핵산 샘플을 형성할 수 있다. 식별자 라이브러리의 식별자는 순서가 지정되어 있으므로 이제 키를 지정된 순서대로 정렬하고 액세스하고 순서를 지정할 수 있다.

[0465] 일부 구현에서, 식별자에 첨부된 키는 입력 식별자를 랜덤 비트의 스트림에 매핑하는 랜덤 함수를 인스턴스화하는 데 사용될 수 있다. 이러한 랜덤 함수는 해싱과 같이 값을 계산하기는 쉽지만 주어진 값에서 반전시키기는 어려운 함수가 필요한 적용 분야에 유용할 수 있다. 이러한 적용 분야에서는 각각 고유 식별자로 조합된 키 라이브러리가 랜덤 함수로 사용된다. 값이 해싱되어야 할 때 이는 식별자에 매핑된다. 다음으로, 혼성화 캡처 또는 PCR과 같은 무작위 액세스 방법을 사용하여 키 라이브러리에서 식별자에 액세스한다. 식별자는 무작위 염기 시퀀스로 구성된 키에 첨부된다. 이 키는 순서가 지정되어 비트 스트림으로 변환되며 무작위 함수의 출력으로 사용된다.

[0466] 핵산 분자 라이브러리는 저렴하고 빠르게 복사될 수 있고, 소량으로 은밀하게 운반될 수 있기 때문에, 위에서 설명한 대로 생성된 핵산 키 세트는 대량의 암호화 키를 지리적으로 모여 있지 않는 다수의 당사자 간에 안전하고 은밀한 방식으로 주기적으로 배포해야 하는 상황에서 유용할 수 있다. 또한 키는 매우 오랜 기간 동안 안정적으로 보관되므로 암호화된 보관 데이터를 안전하게 저장할 수 있다.

[0467] 도 91-94는 DNA에 저장된 랜덤 또는 암호화된 데이터를 생성, 저장, 액세스 및 사용하는 방법의 실시예를 도시한다. DNA는 회색, 검정색 막대와 심볼로 구성된 스트림으로 표시된다. 묘사된 각각의 DNA는 별개의 종을 나타낸다. "종"은 동일한 서열의 하나 이상의 DNA 분자(들)로 정의된다. "종"이 복수 의미로 사용되는 경우, 복수의 종에 포함된 모든 종은 개별 순서를 가지고 있다고 가정할 수 있지만, 때로는 "종" 대신 "개별 종"이라고 표기하여 이를 명시적으로 나타낸다.

[0468] 도 91는 DNA의 큰 조합 공간과 시퀀서를 사용하는 엔트로피(또는 무작위 데이터) 생성기의 예를 묘사한다. 이 방법은 종자라고 불리는 DNA 종의 무작위 풀로 시작된다. 시드는 정의된 조합 DNA 세트의 모든 종, 가령, 50개의 염기( $4^{50}$ 개 구성원을 가짐)이 있는 모든 DNA 종의 균일한 분포가 이상적으로 포함되어야 한다. 그러나 전체 조합 공간은 모든 구성원이 시드에 표시되기에는 너무 클 수 있으므로 시드에 전체 조합 공간 대신 조합 공간의 무작위 서브세트가 포함되는 것이 허용된다. 시드 종은 가장자리(검은색 및 연한 회색 막대)에 공통 서열이 있고 중간(N...N)에 별개의 서열이 있도록 설계될 수 있다. 축퇴성 올리고뉴클레오티드 합성 전략을 사용하여 이러한 시작 시드를 신속하고 저렴한 방식으로 제조할 수 있다. 공통 가장자리 서열은 PCR을 통해 시드의 증폭을 가능하게 하거나 특정 관독(또는 시퀀싱) 방법과의 호환성을 가능하게 할 수 있다. 축퇴성 올리고뉴클레오티드 합성의 대안으로, 조합 DNA 조립(1회 반응으로 다중화)을 사용하여 신속하고 저렴하게 시드를 생성할 수도 있다. 시퀀서는 종자에서 종을 무작위로 샘플링하며 무작위 순서로 수행한다. 주어진 시간에 시퀀서가 읽는 종에는 불확실성이 있기 때문에 시스템은 엔트로피 생성기로 분류될 수 있으며, 예를 들어 암호화 키와 같은 난수 또는 데이터의 랜덤 스트림을 생성하는 데 사용될 수 있다.

[0469] 도 92a는 랜덤하게 생성된 데이터를 DNA에 저장하는 방법의 예시적인 개략도를 예시한다. 이는 (1) 시드라고 불리는 DNA 종의 대규모 무작위 풀로 시작된다. 시드는 정의된 조합 DNA 세트의 모든 종, 가령, 50개의 염기( $4^{50}$ 개 구성원을 가짐)이 있는 모든 DNA 종의 균일한 분포가 이상적으로 포함되어야 한다. 그러나 전체 조합 공간은 모든 구성원이 시드에 표시되기에는 너무 클 수 있으므로 시드에 조합 공간의 랜덤 서브세트가 포함되는 것이 허용된다. 시드는 축퇴성 올리고뉴클레오티드 합성 또는 조합 DNA 조립으로부터 그 자체로 생성될 수 있다. (2) 랜덤 데이터(또는 엔트로피)는 시드에 있는 종의 무작위 서브세트를 취하여 생성된다. 예를 들어, 이는 시드 용액의 비례적, 부분적 부피를 취함으로써 달성될 수 있다. 예를 들어, 시드 용액이 마이크로리터( $\mu\text{L}$ )당 약 100만 종으로 구성된 경우 시드 용액(잘 혼합됐다고 가정)에서 1나노리터( $\text{nL}$ ) 분취량을 취하여 약 1,000개의 종의 랜덤 서브세트가 선택될 수 있다. 대안으로, 시드 용액의 분취량을 나노기공 막을 통해 흐르게 하고 막을 통과하는 종만을 수집함으로써 서브세트가 선택될 수 있다. 막을 통과하는 종의 수를 계산하는 것은 나노기공 사이의 전압 차이를 측정하여 달성할 수 있다. 이 프로세스는 원하는 수의 시그니처(가령, 100, 1000, 10000개 이상의 종 시그니처)가 검출될 때까지 계속될 수 있다. 또 다른 대체 방법으로는 단일 종을 작은 방울로 분리할 수 있다(가령, 오일 에멀전 사용). 단일 종을 갖는 작은 액적은 형광 시그니처에 의해 검출되고 일련의 미세유체 채널에 의해 수집 챔버로 분류될 수 있다. (3) 선택된 각 종을 식별자로 참조할 수 있으며, 나아가 "랜덤 식별자 라이브러리" 또는 RIL로 선택된 종의 전체 서브세트를 참조할 수도 있다. RIL의 정보를 안정화하고 분해로부터 보호하기 위해 RIL은 종의 말단에 있는 공통 서열에 결합하는 PCR 프라이머를 사용하여 증폭될 수 있다. RIL의 식별자(따라서 내부에 저장된 데이터)를 결정하기 위해 RIL의 순서가 지정될 수 있다. 실제 식별자는 정의된 노이즈 임계값보다 강화된 샘플의 종에 의해 정의될 수 있다. (4) RIL에 포함된 데이터가 결정되면 추가 오류 검사 및 오류 수정 종류가 RIL에 추가될 수 있다. 예를 들어, 예상되는 식별자 수(가령, 체크섬 또는 패리티 검사)에 대한 정보가 포함된 "정수 DNA"가 RIL에 추가될 수 있다. 정수 DNA를 통해 모든 정보를 복구하기 위해 RIL의 서열을 얼마나 깊게 배열해야 하는지 알 수 있다.

[0470] RIL은 고유한 DNA 태그로 바코드가 표시될 수 있다. 그런 다음 여러 바코드 RIL을 함께 모아서 특정 RIL에 고유한 DNA 태그에 대한 혼성화 분석(또는 PCR)을 통해 개별적으로 접근할 수 있다. 독특한 DNA 태그는 조합적으로 조립되거나 합성된 후 해당 RIL에 조립될 수 있다. 도 92b는 각각 100개의 랜덤 염기를 함유하는 4개의 종을 포

$$\log_2(4^{100} \text{choose } 4) \approx 725$$

함하는 예시적인 RIL을 보여준다. 가능한 종의 조합 공간은  $4^{100}$ 이므로 RIL은  $\approx$ 의 정보

비트를 포함할 수 있다. 도 92c는 또한 각각 100개의 무작위 염기를 함유하는 4개의 종을 포함하는 예시적인 RIL을 보여준다.  $4^{100}$ 개의 조합 공간(도 92b에서와 같이)에서 선택된 4개 종의 특정 비순차적 조합에 정보를 저

$$\log_2(4^{90}) = 180$$

장하는 대신, 각 종의 최종 90개의 랜덤 염기가 정보 비트를 저장하기 위해 예약될 수 있으며, 처음 10개의 랜덤 염기는 4개 종 각각에 저장된 정보 간의 상대적 순서를 설정하기 위해 예약될 수 있다. 상대 순서는 정의된 4개 염기 순서에 기초하여 10개 염기 스트링의 사전식 순서로 정의될 수 있다(영어 단어가 알파벳 문자 순서에 따라 정렬되는 방식과 유사함). RIL에 정보를 할당하는 이 방법은 도 92b에 설명된 방법보다 이진 스트링에 매핑하는 데 계산적으로 더 빠를 수 있다.

[0471] 이전 도면(도 92)에서는 여러 RIL을 바코드화하고 함께 풀링하는 전략에 대해 논의한다. 이를 통해 입력이 (개별 RIL에 액세스하기 위한) 바코드 혼성화 프로브에 대응하고 출력이 랜덤 데이터 스트링(대상 RIL에 의해 인코

딩됨)에 대응하는 입력-출력 매핑이 생성된다. 반면, 이 방법에서는 미리 정의된 바코드가 조합된 풀에서 검색되기 위해 랜덤 데이터로 조립되는데, 도 93a는 (데이터를 액세스하기 위한) 바코드가 랜덤 데이터 자체와 함께 랜덤하게 생성되는 랜덤 데이터 스트링과 핵산 프로브 사이의 입출력 매핑을 생성하는 다양한 방법을 보여준다. 예를 들어, 바코드는 하나 또는 여러 종의 양쪽 가장자리에 나타날 수 있는 한 쌍의 짧은 DNA 서열일 수 있다. 이 실시예에서, 가능한 바코드의 조합 공간은 각 바코드가 우연히 하나 이상의 종과 연관되도록 풀의 모든 가능한 종의 총 수에 비해 작을 수 있다. 예를 들어, 바코드가 종의 랜덤 DNA 서열의 각 가장자리에 있는 3개의 염기(공통 서열 옆에 위치)인 경우 가능한 바코드는  $4^6 = 4096$ 개이므로 이를 액세스하기 위해 구축될 수 있는 프라이머 쌍은  $4^6 = 4096$ 이다(12비트 입력에 대응). DNA 풀이 약 400,000종을 갖도록 선택되면 각 바코드는 평균 약 100종과 연관될 수 있다. 이 실시예에서, RIL은 각 바코드와 관련된 종의 서브세트에 의해 정의된다. 이전 예에 따르면, 각 종이 바코드에 사용된 염기(또는 서열) 외에 25개의 랜덤 염기(또는 랜덤 서열)를 포함하는 경우, 100종의 RIL과 연관된 바코드는 최대  $\approx \log_2(4^{25} \text{choose } 100) \approx 4475$  비트의 정보를 포함할 수 있다.

[0472] 도 93b는 바코드 RIL 풀로부터 저장된 랜덤 데이터에 액세스하고 관독하기 위한 방식의 구현을 보여준다. 시퀀서(또는 관독기)는 출력을 반환하기 전에 시퀀스 데이터를 조작하는 기능을 추가로 포함할 수 있다. 예를 들어, 해시 함수는 출력 데이터 스트링을 사용하여 역화학 쿼리를 수행하고 해당 입력을 찾는 것을 어렵게 만들 수 있다. 예를 들어 입력이 인증에 사용되는 키 또는 자격 증명인 경우 이 기능이 유용할 수 있다.

[0473] 쿼리 가능한(또는 액세스 가능한) 데이터의 랜덤 스트링을 생성하고 저장하는 방법은 (랜덤 데이터 스트링에서 생성된) 암호화 키를 생성하고 보관하는 데 특히 유용할 수 있다. 각 입력은 다른 암호화 키에 액세스하는 데 사용될 수 있다. 예를 들어, 각 입력은 개인 보관 데이터베이스의 특정 사용자, 시간 범위 및/또는 프로젝트에 해당할 수 있다. 개인 보관 데이터베이스의 암호화된 데이터(잠재적으로 매우 많은 양의 데이터에 달함)는 보관 서비스 제공업체에 의해 기존 매체에 저장될 수 있으며, 암호화 키는 소유자에 의해 DNA에 저장될 수 있다. 또한 특정 입력에 대한 화학적 액세스 프로토콜을 수행하는 데 필요한 잠재적인 대기 시간과 정교함으로 인해 해킹에 대한 암호화 방법의 보안 장벽이 높아질 수 있다.

[0474] 도 94은 아티팩트에 대한 액세스를 보호하고 인증하기 위한 예시적인 시스템을 도시한다. 시스템에는 가능한 종의 큰 풀에서 가져온 DNA 종의 특정 조합으로 구성된 물리적 키가 필요하다. "식별자 키"라고도 하는 종의 표적 조합은 예를 들어 조합 미세유체 채널, 전기습윤 또는 인쇄 장치에 의해 자동으로 생성되거나 피켓팅에 의해 수동으로 생성될 수 있다. 잠금 기능이 내장된 리더 또는 시퀀서는 일치하는 식별자 키를 확인하고 아티팩트에 대한 액세스를 활성화한다. 또는 리더는 아티팩트에 대한 액세스를 직접 잠금 해제하는 대신 아티팩트에 액세스하는 데 사용할 수 있는 토큰을 반환하는 자격 증명 토큰 시스템으로 작동할 수 있다. 토큰은 예를 들어 관독기 내에 내장된 해싱 함수에 의해 생성될 수 있다.

[0475] **DNA로 개체를 추적하고 개체에 태그를 지정하는 예시적 방법**

[0476] 용매에 용해된 식별자 라이브러리는 정보로 태그를 지정하기 위해 물리적 개체에 뿌리거나, 퍼거나, 분배하거나, 주입할 수 있다. 예를 들어, 고유 식별자 라이브러리를 사용하여 개체 유형의 개별 인스턴스에 태그를 지정할 수 있다. 개체의 식별자 라이브러리 태그는 고유한 바코드 역할을 할 수도 있고 제품 번호, 제조 또는 배송 날짜, 원산지 위치 또는 개체 이력, 가령, 예를 들어 이전 소유자의 거래 목록과 관련된 기타 정보와 같은 보다 정교한 정보를 포함할 수도 있다. 개체에 태그를 지정하기 위해 식별자를 사용하는 주요 이점은 식별자가 검출될 수 없고 내구성 있으며 수많은 개체 인스턴스에 개별적으로 태그를 지정하는 데 적합하다는 것이다.

[0477] 또 다른 실시예에서, 하나 이상의 물리적 위치는 각각 식별자 라이브러리의 고유 식별자로 태그가 지정될 수 있다. 예를 들어, 물리적 사이트 A, B, 및 C에는 식별자 라이브러리가 곳곳에 태그로 지정될 수 있다. 사이트 A를 방문하거나 사이트 A와 접촉하는 개체, 가령, 차량, 사람 또는 기타 개체는 의도적이든 아니든 식별자 라이브러리의 샘플을 선택할 수 있다. 나중에 개체를 액세스하면, 개체에서 샘플을 수집하여 화학적으로 처리하고 디코딩하여 해당 개체가 방문한 위치를 식별할 수 있다. 개체는 두 개 이상의 위치를 방문할 수 있으며 두 개 이상의 샘플을 수집할 수 있다. 식별자 라이브러리가 분리되어 있는 경우 개체가 방문한 위치 중 일부 또는 전체를 식별하는 데 유사한 프로세스가 사용될 수 있다. 이러한 방식은 개체를 은밀하게 추적하는 데 응용될 수 있다. 이 방식을 사용하면 식별자를 특별히 찾지 않는 한 검출될 수 없고 생물학적으로 불활성으로 설계될 수 있으며 수많은 사이트나 개체에 고유하게 태그를 지정하는 데 사용될 수 있다는 이점이 있다.

- [0478] 다른 실시예에서, 식별자 라이브러리는 개체를 태깅할 수 있다. 개체는 자신이 방문하는 사이트에 삽입된 식별자의 샘플을 남길 수 있다. 이러한 샘플은 어떤 개체가 사이트를 방문했는지 식별하기 위해 수집, 처리 및 디코딩될 수 있다.
- [0479] **조합 DNA 조립 방법 및 시스템의 예시적 응용**
- [0480] 대규모로 정의된 식별자 세트의 구성요소를 조합적으로 조립하기 위해 여기에 설명된 방법 및 시스템은 정보 기술(가령, 데이터 저장, 컴퓨팅 및 암호화)과 관련되어 설명되었다. 그러나 이러한 시스템과 방법은 처리량이 높은 조합 DNA 조립의 모든 응용 분야에 더 일반적으로 사용될 수 있다.
- [0481] 한 실시예를 들어, 우리는 아미노산 사슬을 암호화하는 조합 DNA의 라이브러리를 생성할 수 있다. 이러한 아미노산 사슬은 펩타이드 또는 단백질을 나타낼 수 있다. 조립을 위한 DNA 단편은 코돈 서열을 포함할 수 있다. 단편이 조립되는 연결점은 조합 라이브러리의 모든 구성원에 공통되는 기능적으로 또는 구조적으로 비활성 코돈일 수 있다. 대안으로, 단편이 조립되는 접합부는 나중에 처리된 펩티드 사슬로 번역되는 메신저 RNA로부터 최종적으로 제거되는 인트론일 수 있다. 특정 단편은 코돈이 아닐 수 있지만 오히려 (다른 조립된 바코드와 결합하여) 각 조합 코돈 문자열에 고유하게 태그를 지정하는 바코드 서열일 수 있다. 조립된 제품(바코드 + 코돈 문자열)은 함께 모아서 시험관 내 발현 분석을 위해 액적에 캡슐화하거나 함께 모아서 생체 내 발현 분석을 위해 세포로 변환할 수 있다. 분석은 형광 강도에 따라 액적/세포가 빈으로 분류될 수 있도록 형광 출력을 가질 수 있으며, 이어서 각 코돈 문자열을 특정 출력과 연관시킬 목적으로 DNA 바코드의 서열이 결정될 수 있다.
- [0482] 다른 실시예를 들어, 우리는 RNA를 암호화하는 조합 DNA의 라이브러리를 생성할 수 있다. 예를 들어, 조립된 DNA는 마이크로RNA 또는 CRISPR gRNA의 조합을 나타낼 수 있다. 시험관 내 또는 생체 내 풀링된 RNA 발현 분석은 액적 또는 세포, 그리고 어떤 액적 또는 세포가 어떤 RNA 서열을 포함하는지 추적하기 위한 바코드를 사용하여 위에서 설명한 대로 수행될 수 있다. 그러나 출력 자체가 RNA 염기서열 분석 데이터인 경우 일부 풀링된 분석은 물방울이나 세포 외부에서 수행될 수 있다. 이러한 통합 분석의 예로는 RNA 압타머 선별 및 검사(가령, SELEX)가 있다.
- [0483] 다른 실시예를 들어, 우리는 대사 경로에서 유전자를 암호화하는 조합 DNA의 라이브러리를 생성할 수 있다. 각 DNA 단편에는 유전자 발현 구조가 포함될 수 있다. 단편이 조립되는 접합부는 유전자 사이의 불활성 DNA 서열을 나타낼 수 있다. 시험관 내 또는 생체 내 통합 유전자 경로 발현 분석은 액적 또는 세포, 그리고 어떤 액적 또는 세포가 어떤 유전자 경로를 포함하는지 추적하기 위한 바코드를 사용하여 위에서 설명한 대로 수행될 수 있다.
- [0484] 다른 실시예를 들어, 우리는 유전자 조절 요소들의 다양한 조합을 갖는 조합 DNA의 라이브러리를 생성할 수 있다. 유전자 조절 요소의 예에는 5' 비번역 영역(UTR), 리보솜 결합 부위(RBS), 인트론, 엑손, 프로모터, 터미네이터 및 전사 인자(TF) 결합 부위가 포함된다. 시험관 내 또는 생체 내 풀링된 유전자 발현 분석은 액적 또는 세포, 그리고 어떤 액적 또는 세포가 어떤 유전자 조절 구성물을 포함하는지 추적하기 위한 바코드를 사용하여 위에서 설명한 대로 수행될 수 있다.
- [0485] 또 다른 실시예에서, 조합 DNA 압타머의 라이브러리가 생성될 수 있다. 리간드에 결합하는 DNA 압타머의 능력을 테스트하기 위해 분석을 수행할 수 있다.
- [0486] **예시 1: DNA 분자에 하나의 시를 인코딩, 기록 및 관독하기** 인코딩된 데이터는 시가 포함된 텍스트 파일이다. 데이터는 피펫을 사용하여 수동으로 인코딩되어 96개 구성요소로 구성된 두 개의 층으로부터의 DNA 구성요소를 함께 혼합하여 중첩 확장 PCR로 구현된 곱 방식을 사용해 식별자를 구성한다. 제1 층인 X는 총 96개의 DNA 구성요소를 포함한다. 제2 층인 Y도 총 96개의 구성요소를 포함한다. DNA를 기록 전에, 데이터는 이진수로 매핑된 다음 원본 데이터의 61비트로 구성된 모든 연속(인접한 분리) 스트링이 정확히 17비트 값이 1인 96비트 스트링으로 변환되는 균일한 가중치 형식으로 다시 코딩된다. 이 균일한 가중치 형식은 자연스러운 에러 체크 특성을 가질 수 있다. 그런 다음 데이터는 참조 맵을 형성하기 위해 96 x 96 테이블로 해시된다.
- [0487] 도 74a의 중앙 패널은 시를 복수의 식별자로 인코딩하는 96 x 96 테이블의 2차원 참조 맵을 보여준다. 어두운 점은 '1' 비트 값에 해당하고 흰색 점은 '0' 비트 값에 대응한다. 데이터는 96개 구성요소로 구성된 두 개의 층을 사용하여 식별자로 인코딩된다. 테이블의 각 X 값과 Y 값에는 구성요소가 할당되고 X 및 Y 구성요소는 '1' 값을 갖는 각 (X,Y) 좌표에 대한 중첩 확장 PCR을 사용하여 식별자로 조립된다. 가능한 각 (X,Y) 조립체의 존재 여부를 결정하기 위해 식별자 라이브러리를 시퀀싱하여 데이터를 다시 읽어(가령 디코딩) 사용했다.

- [0488] 도 74a의 오른쪽 패널은 시퀀싱에 의해 결정된 식별자 라이브러리에 존재하는 서열의 존재비에 대한 2차원 히트맵을 보여준다. 각 픽셀은 해당 X 및 Y 구성요소로 구성된 분자를 나타내며 해당 픽셀의 그레이스케일 강도는 다른 분자와 비교하여 해당 분자의 상대적 풍부함을 나타낸다. 식별자는 각 행에서 가장 풍부한 상위 17개(X, Y) 조합체로 간주된다(균일한 가중치 인코딩은 96비트의 각 연속 스트링이 정확히 17개의 '1' 값을 가질 수 있으므로 17개의 해당 식별자를 가질 수 있음을 보장하기 때문이다).
- [0489] **예시 2: 62824 비트 텍스트파일을 인코딩하기.** 인코딩될 데이터는 총 62824 비트에 달하는 3개의 시로 구성된 텍스트 파일이다. 데이터는 Labcyte Echo® Liquid Handler를 사용하여 인코딩되어 384개 구성요소로 구성된 두 개의 층의 DNA 구성요소를 함께 혼합하여 중첩 확장 PCR로 구현된 제품 체계를 사용하여 식별자를 구성한다. 제 1 층인 X는 총 384개의 DNA 구성요소를 포함한다. 제 2 층인 Y도 총 384개의 구성요소를 포함한다. DNA를 쓰기 전에 데이터를 이진수로 매핑한 후 다시 코딩하여 가중치(비트 값 수 '1')를 줄이고 체크섬을 포함시킨다. 체크섬은 192 비트 데이터의 모든 연속 스트링에 대한 체크섬에 해당하는 식별자가 있도록 설정된다. 재코딩된 데이터의 가중치는 약 10,100으로, 구성될 식별자의 개수에 대응한다. 그런 다음 데이터는 참조 맵을 형성하기 위해 384 x 384 테이블로 해시될 수 있다.
- [0490] 도 74b의 중앙 패널은 텍스트 파일을 복수의 식별자로 인코딩하는 384 x 384 테이블의 2차원 참조 맵을 보여준다. 각 좌표 (X,Y)는 X + (Y-1)\*192 위치의 데이터 비트에 대응한다. 검은색 점은 비트값 '1'에 해당하고 흰색 점은 비트값 '0'에 대응한다. 도면의 오른쪽 상의 검은 점은 체크섬이고, 도면의 위쪽의 검은 점 패턴은 코드북(가령, 데이터 디코딩을 위한 사전)이다. 테이블의 각 X 값과 Y 값에는 구성요소가 할당될 수 있으며 X 및 Y 구성요소는 '1' 값을 갖는 각 (X, Y) 좌표에 대한 오버랩 확장 PCR을 사용하여 식별자로 조립된다. 가능한 각 (X, Y) 조합체의 존재 여부를 결정하기 위해 식별자 라이브러리를 시퀀싱하여 데이터가 다시 판독(가령, 디코딩)되었다.
- [0491] 도 74b의 오른쪽 패널은 시퀀싱에 의해 결정된 식별자 라이브러리에 존재하는 서열의 존재비에 대한 2차원 히트맵을 보여준다. 각 픽셀은 해당 X 및 Y 구성요소로 구성된 분자를 나타내며 해당 픽셀의 그레이스케일 강도는 다른 분자와 비교하여 해당 분자의 상대적 풍부함을 나타낸다. 식별자는 각 행에서 가장 풍부한 상위 S개(X, Y) 어셈블리로 간주되며 여기서 각 행의 S는 체크섬 값일 수 있다.
- [0492] 일반적으로, 본 명세서에 설명된 주제 및 기능적 동작의 측면은 본 명세서에 개시된 구조 및 그 구조적 등가물을 포함하는 디지털 전자 회로, 컴퓨터 소프트웨어, 펌웨어 또는 하드웨어에서 구현될 수 있다. 본 명세서에 설명된 주제의 측면은 하나 이상의 컴퓨터 프로그램 제품, 즉 데이터 처리 장치에 의해 실행되거나 데이터 처리 장치의 동작을 제어하기 위해 컴퓨터 판독 가능 매체에 인코딩된 컴퓨터 프로그램 명령의 하나 이상의 모듈로 구현될 수 있다. 컴퓨터 판독 가능 매체는 기계 판독 가능 저장 장치, 기계 판독 가능 저장 기판, 메모리 장치, 기계 판독 가능 전파 신호에 영향을 미치는 물질의 구성, 또는 이들 중 하나 이상의 조합일 수 있다. "데이터 처리 장치"라는 용어는 예를 들어 프로그래밍 가능한 프로세서, 컴퓨터, 또는 다중 프로세서 또는 컴퓨터를 포함하여 데이터를 처리하기 위한 모든 장치, 장치 및 기계를 포함한다. 장치는 하드웨어 이외에 문제의 컴퓨터 프로그램에 대한 실행 환경을 생성하는 코드, 예를 들어 프로세서 펌웨어, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 체제, 또는 이들 중 하나 또는 그 조합을 구성하는 코드를 포함할 수 있다. 전파된 신호는 인공적으로 생성된 신호, 예를 들어 적절한 수신기 장치로 전송하기 위해 정보를 인코딩하기 위해 생성된 기계 생성 전기, 광학 또는 전자기 신호이다.
- [0493] 컴퓨터 프로그램(프로그램, 소프트웨어, 소프트웨어 응용 프로그램, 스크립트 또는 코드라고도 함)은 컴파일된 언어나 해석된 언어를 포함하여 모든 형태의 프로그래밍 언어로 작성될 수 있으며 임의의 형태로, 가령, 독립형 프로그램 또는 모듈, 구성요소, 서브루틴, 또는 컴퓨팅 환경에서 사용되기에 적합한 그 밖의 다른 유닛으로 배포될 수 있다. 컴퓨터 프로그램은 파일 시스템의 파일에 해당할 수 있다. 프로그램은 다른 프로그램이나 데이터(가령 마크업 언어 문서에 저장된 하나 이상의 스크립트)를 보유하는 파일의 일부, 해당 프로그램 전용 단일 파일 또는 여러 개의 조정된 파일(가령, 하나 이상의 모듈, 하위 프로그램 또는 코드 일부를 저장하는 파일)에 저장될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터 또는 한 사이트에 위치하거나 여러 사이트에 걸쳐 분산되고 통신 네트워크로 연결된 여러 컴퓨터에서 실행되도록 배포될 수 있다.
- [0494] 본 명세서에 설명된 프로세스 및 논리 흐름은 입력 데이터에 대해 작동하고 출력을 생성함으로써 기능을 수행하는 하나 이상의 컴퓨터 프로그램을 실행하는 하나 이상의 프로그래밍 가능한 프로세서에 의해 수행될 수 있다. 프로세스 및 논리 흐름은 또한 특수 목적 논리 회로, 예를 들어 FPGA(필드 프로그래밍 가능 게이트 어레이) 또는 ASIC(응용프로그램 특정 집적 회로)에 의해 수행될 수 있고 장치도 구현될 수 있다.

- [0495] 컴퓨터 프로그램의 실행에 적합한 프로세서에는 예를 들어 범용 및 특수 목적 마이크로프로세서, 그리고 모든 종류의 디지털 컴퓨터의 하나 이상의 프로세서가 포함된다. 일반적으로 프로세서는 읽기 전용 메모리나 랜덤 액세스 메모리 또는 둘 다로부터 명령과 데이터를 수신한다. 컴퓨터의 필수 요소는 명령을 수행하는 프로세서와 명령 및 데이터를 저장하는 하나 이상의 메모리 장치이다. 일반적으로, 컴퓨터는 또한 데이터를 저장하기 위한 하나 이상의 대용량 저장 장치, 예를 들어 자기, 광자기 디스크 또는 광 디스크로부터 데이터를 수신하거나 전송하거나 둘 모두를 포함하거나 작동 가능하게 결합될 것이다. 그러나 컴퓨터에 그러한 장치가 있을 필요는 없다.
- [0496]
- [0497] **구현 예**
- [0498] 항목 1. 디지털 정보를 핵산 서열로 변환하기 위한 시스템으로서, 상기 시스템은,
- [0499] 핵산 분자의 풀을 포함하는 유체를 보관하도록 구성된 출발지 저장소,
- [0500] 복수의 전기 전도성 판 및 상대 전극을 포함하는 메인 채널 - 상기 복수의 전기 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치됨 - ,
- [0501] 도착지 저장소,
- [0502] 출발지 저장소와 메인 채널과 유체 연통하는 입력 채널 - 상기 입력 채널은 출발지 저장소로부터의 제1 복수의 핵산 분자를 포함하는 제1 유체 볼륨을 메인 채널로 분배하도록 구성됨 - ,
- [0503] 상기 메인 채널 및 도착지 저장소와 유체 연통하는 출력 채널 - 상기 출력 채널은 메인 채널로부터의 제2 유체 볼륨을 도착지 저장소로 분배하도록 구성됨 - 을 포함하는, 시스템.
- [0504] 항목 2. 항목 1에 있어서, 상기 메인 채널은 반응 챔버를 포함하는, 시스템.
- [0505] 항목 3. 항목 1에 있어서, 상기 메인 채널은 반응 챔버인, 시스템.
- [0506] 항목 4. 항목 1 내지 3 중 어느 하나에 있어서, 복수의 셀 - 각 셀은 복수의 전기 전도성 판 중 하나씩 및 상대 전극의 일부분을 포함함 - 을 포함하는, 시스템.
- [0507] 항목 5. 항목 1 또는 2에 있어서, 각 셀은 기저 층 및 제1 유전체 층을 포함하며, 상기 제1 유전체 층은 상기 기저 층과 복수의 전기 전도성 판 사이에 배치되는, 시스템.
- [0508] 항목 6. 항목 1 또는 2에 있어서, 각 셀은 기저 층 및 제1 유전체 층 - 복수의 전기 전도성 판은 상기 기저 층과 제1 유전체 층 사이에 배치됨 - 을 포함하는, 시스템.
- [0509] 항목 7. 항목 1 내지 6 중 어느 하나에 있어서, 상기 기저 층은 반도체 층을 포함하는, 시스템.
- [0510] 항목 8. 항목 7에 있어서, 상기 기저 층은 반도체 층에 부착된 비-반도체 층(non-semiconducting layer)을 포함하는, 시스템.
- [0511] 항목 9. 항목 1 내지 8 중 어느 하나에 있어서, 상기 기저 층은 반도체 층을 포함하는, 시스템.
- [0512] 항목 10. 항목 1 내지 9 중 어느 하나에 있어서, 상기 기저 층은 절연 층을 포함하는, 시스템.
- [0513] 항목 11. 항목 1 내지 10 중 어느 하나에 있어서, 상기 기저 층은 투명한, 시스템.
- [0514] 항목 12. 항목 11에 있어서, 상기 기저 층은 유리를 포함하는, 시스템.
- [0515] 항목 13. 항목 1 내지 12 중 어느 하나에 있어서, 상기 기저 층은 가열기 요소를 포함하는, 시스템.
- [0516] 항목 14. 항목 13에 있어서, 상기 가열기 요소는 저항성 가열 요소를 포함하는, 시스템.
- [0517] 항목 15. 항목 4 내지 14 중 어느 하나에 있어서, 각 셀은 제2 유전체 층을 포함하고, 상기 제2 유전체 층은 상대 전극 상에 배치되는, 시스템.
- [0518] 항목 16. 항목 15에 있어서, 복수의 전기 전도성 판과 제2 유전체 층은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되는, 시스템.
- [0519] 항목 17. 항목 4 내지 16 중 어느 하나에 있어서, 각 셀은 전기 전도성 판에 부착된 복수의 핵산 어댑터를 포함

하는, 시스템.

- [0520] 항목 18. 항목 1 내지 17 중 어느 하나에 있어서, 제3 유전체 층을 포함하고, 상기 제3 유전체 층은 복수의 전기 전도성 판 상에 배치되는, 시스템.
- [0521] 항목 19. 항목 18에 있어서, 각 셀은 제3 유전체 층에 부착된 복수의 핵산 어댑터를 포함하는, 시스템.
- [0522] 항목 20. 항목 4 내지 19 중 어느 하나에 있어서, 각 셀은 복수의 전기 전도성 판 중 하나 및 복수의 카운터 전극 중 하나를 포함하며, 복수의 상대 전극 각각은 복수의 전기 전도성 판 중 하나의 반대편에 배치되는, 시스템.
- [0523] 항목 21. 항목 16 내지 18 중 어느 하나에 있어서, 셀은 메인 채널의 제2 차원 및 제3 차원을 따라 2차원 어레이로 배열되고, 상기 어레이는 셀의 행 및 열을 갖는, 시스템.
- [0524] 항목 22. 항목 1 내지 21 중 어느 하나에 있어서, 각 전기 전도성 판은 전압원에 전기적으로 연결되는, 시스템.
- [0525] 항목 23. 항목 1 내지 22 중 어느 하나에 있어서, 복수의 스위치를 포함하는 제어 시스템 - 각 스위치는 복수의 전기 전도성 판 중 하나씩 및 전압원에 전기적으로 연결됨 - 을 포함하는, 시스템.
- [0526] 항목 24. 항목 23에 있어서, 어레이의 각 셀 양단의 전압은 복수의 스위치 중 하나를 작동시킴으로써 개별적으로 제어 가능한, 시스템.
- [0527] 항목 25. 항목 1 내지 24 중 어느 하나에 있어서, 유체는 유전체 유체인, 시스템.
- [0528] 항목 26. 항목 1 내지 24 중 어느 하나에 있어서, 유체는 전도성 유체인, 시스템.
- [0529] 항목 27. 항목 1 내지 26 중 어느 하나에 있어서, 메인 채널과 유체 연통하는 버퍼 저장소를 포함하는, 시스템.
- [0530] 항목 28. 항목 1 내지 27 중 어느 하나에 있어서, 메인 채널과 유체 연통하는 리가제 저장소를 포함하는, 시스템.
- [0531] 항목 29. 항목 1 내지 28 중 어느 하나에 있어서, 입력 채널, 메인 채널, 및 출력 채널 중 하나 이상을 통해 유체 볼륨을 펌프하도록 구성된 펌프를 포함하는, 시스템.
- [0532] 항목 30. 항목 1 내지 29 중 어느 하나에 있어서, 메인 채널을 통해 유체 볼륨의 흐름을 제어하기 위한 밸브를 포함하는, 시스템.
- [0533] 항목 31. 항목 1 내지 30 중 어느 하나에 있어서, 복수의 출발지 저장소 - 복수의 출발지 저장소의 각각은 실질적으로 동일한 핵산 분자의 모집단을 갖는 유체 볼륨을 가짐 - 를 포함하는, 시스템.
- [0534] 항목 32. 항목 31에 있어서, 복수의 출발지 저장소의 각각은 실질적으로 동일한 핵산 분자의 상이한 모집단을 포함하는, 시스템.
- [0535] 항목 33. 항목 1 내지 32 중 어느 하나에 있어서, 메인 채널과 유체 연통하는 복수의 도착지 저장소를 포함하는, 시스템.
- [0536] 항목 34. 항목 1 내지 33 중 어느 하나에 있어서, 핵산 분자는 디지털 정보를 인코딩하는, 시스템.
- [0537] 항목 35. 항목 1 내지 34 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 인코딩하는 식별자 핵산 분자를 포함하는, 시스템.
- [0538] 항목 36. 항목 1 내지 35 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 인코딩하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 시스템.
- [0539] 항목 37. 항목 35 또는 36에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 시스템.
- [0540] 항목 38. 항목 1 내지 37 중 어느 하나에 있어서, 메인 채널은 복수의 유체 연결된 반응 챔버를 포함하는, 시스템.
- [0541] 항목 39. 항목 21 내지 38 중 어느 하나에 있어서, 2차원 어레이는 둘 이상의 블록으로 분할되는, 시스템.
- [0542] 항목 40. 항목 39에 있어서, 복수의 유체 연결된 반응 챔버의 각각은 하나씩의 블록을 하우징하는, 시스템.

- [0543] 항목 41. 항목 1 내지 40 중 어느 하나에 있어서, 제2 유체 블록은 제2 복수의 핵산 분자를 포함하는, 시스템.
- [0544] 항목 42. 항목 4 내지 41 중 어느 하나에 있어서, 제2 복수의 핵산 분자는 셀로부터 방출된 핵산 분자를 포함하는, 시스템.
- [0545] 항목 43. 디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,
- [0546] 항목 5 내지 42 중 어느 하나에 따른 시스템,
- [0547] 제1 유전체 층에 배치된 시퀀싱 장치 - 상기 시퀀싱 장치는 메인 채널과 유체 연통하는 유입구 및 캐비티(cavity)와 유체 연통하는 배출구를 가짐 - , 및
- [0548] 배출구의 다운스트림에 배치된 기저 전극을 포함하는, 시스템.
- [0549] 항목 44. 항목 43에 있어서, 시퀀싱 장치는 나노포어(nanopore)를 포함하는, 시스템.
- [0550] 항목 45. 항목 43에 있어서, 시퀀싱 장치는 나노채널을 포함하는, 시스템.
- [0551] 항목 46. 항목 43에 있어서, 시퀀싱 장치는 고체 상태 멤브레인 내에 형성된 나노포어 또는 나노채널을 포함하는, 시스템.
- [0552] 항목 47. 항목 43 내지 46 중 어느 하나에 있어서, 캐비티는 메인 채널의 기저 층 내에 배치되는, 시스템.
- [0553] 항목 48. 항목 43 내지 47 중 어느 하나에 있어서, 메인 채널 및 캐비티는 전해질 용액을 포함하는, 시스템.
- [0554] 항목 49. 항목 44 내지 48 중 어느 하나에 있어서, 나노포어 또는 나노채널은 알파-헤몰리신( $\alpha$ HL) 또는 마이코박테리움 스페그마티스 포린 A(MspA)를 포함하는, 시스템.
- [0555] 항목 50. 항목 43 내지 49 중 어느 하나에 있어서, 기저 전극 및 상대 전극에 의해 생성된 전기장은 나노포어 또는 나노채널 양단에 100 mV 보다 큰 차동 전위를 갖는, 시스템.
- [0556] 항목 51. 항목 43 내지 50 중 어느 하나에 있어서, 복수의 나노포어 또는 나노채널을 포함하는, 시스템.
- [0557] 항목 52. 항목 51에 있어서, 각 블록은 복수의 시퀀싱 장치 중 하나씩을 포함하는, 시스템.
- [0558] 항목 53. 디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,
- [0559] 항목 5 내지 42 중 어느 하나에 따른 시스템,
- [0560] 메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 나노채널을 포함하는 시퀀싱 장치,
- [0561] 배출구의 다운스트림에 배치된 중앙 전극,
- [0562] 셸이 블록 전극과 중앙 전극 사이에 배치되도록 배치되는 블록 전극, 및
- [0563] 핵산이 나노채널을 통해 전좌하는 동안 전류 변화를 검출하도록 구성된 나노채널 센서를 포함하는, 시스템.
- [0564]
- [0565] 항목 54. 제53항에 있어서, 블록 전극 및 중앙 전극에 의해 생성되는 전기장은 나노채널 양단에 100 mV 보다 큰 차동 전위를 갖는, 시스템.
- [0566] 항목 55. 디지털 정보를 핵산 서열로 디코딩하기 위한 시스템으로서, 상기 시스템은,
- [0567] 항목 5 내지 42 중 어느 하나에 따른 시스템, 및
- [0568] 제로 모드 도파관 관독기를 포함하는 시퀀싱 장치를 포함하는, 시스템.
- [0569] 항목 56. 항목 55에 있어서, 유전체 층은 투명한, 시스템.
- [0570] 항목 57. 항목 55 또는 56에 있어서, 유전체 층은 도파관 채널을 포함하는, 시스템.
- [0571] 항목 58. 항목 57에 있어서, 도파관 채널은 그 내부에 고정화된 증합효소를 포함하는, 시스템.
- [0572] 항목 59. 항목 57 또는 58에 있어서, 상기 도파관 채널은 프라이머(primer) 및 형광 라벨링된 뉴클레오티드의 세트를 포함하는, 시스템.

- [0573] 항목 60. 항목 55 내지 59 중 어느 하나에 있어서, 상기 시스템은 단일 가닥 DNA 분자의 상보적 가닥의 합성 동안 형광 라벨링된 뉴클레오티드의 혼입에 의해 생성된 형광 신호를 검출하도록 구성된 검출기를 포함하는, 시스템.
- [0574] 항목 61. 디지털 정보를 핵산 서열로 코딩하기 위한 방법으로서, 상기 방법은,
- [0575] 출발지 저장소 내 유체에 현탁된 핵산 분자의 풀을 획득하는 단계,
- [0576] 출발지 저장소로부터, 입력 채널을 통해, 복수의 핵산 분자를 포함하는 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 메인 채널은 복수의 셀을 포함하고, 각 셀은 복수의 전기 전도성 판 중 하나씩과 상대 전극의 일부를 포함하며, 복수의 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되고, 각 전기 전도성 판은 그 상에 부착된 복수의 핵산 어댑터를 포함함 - , 및
- [0577] 복수의 셀 중 제1 셀에 결합 전압(binding voltage)을 인가함으로써, 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제1 전기 전도성 판의 핵산 어댑터에 결합시키는 단계를 포함하는, 방법.
- [0578] 항목 62. 항목 61에 있어서, 핵산 분자의 복수의 풀을 획득하는 단계 - 핵산 분자의 복수의 풀의 각각은 복수의 출발지 저장소 중 하나 내 유체에 현탁됨 - 를 포함하는, 방법.
- [0579] 항목 63. 항목 61 또는 62에 있어서, 결합 전압의 인가 후, 버퍼 저장소로부터 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소(waste reservoir) 내로 유동시키는 단계를 포함하는, 방법.
- [0580] 항목 64. 항목 62 또는 63에 있어서,
- [0581] 제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0582] 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 핵산 어댑터에 결합된 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.
- [0583] 항목 65. 항목 64에 있어서,
- [0584] 제3 출발지 저장소로부터, 입력 채널을 통해, 제3 복수의 핵산 분자를 포함하는 제3 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0585] 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제3 복수의 핵산 분자의 일부를 제2 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.
- [0586] 항목 66. 항목 62 또는 63에 있어서,
- [0587] 제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0588] 복수의 셀 중 제2 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제2 전기 전도성 판의 핵산 어댑터에 결합하는 단계를 포함하는, 방법.
- [0589] 항목 67. 항목 66에 있어서,
- [0590] 제3 출발지 저장소로부터, 입력 채널을 통해, 제3 복수의 핵산 분자를 포함하는 제3 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0591] 복수의 셀 중 제2 셀에 결합 전압을 인가하여, 제3 복수의 핵산 분자의 일부를 제2 복수의 핵산 분자의 일부에 결합하는 단계를 포함하는, 방법.
- [0592] 68. 항목 61 내지 67 중 어느 하나에 있어서,
- [0593] 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0594] 결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계를 포함하는, 방법.
- [0595] 항목 69. 항목 61 내지 67 중 어느 하나에 있어서,
- [0596] 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산

분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및

- [0597] 결합 전압을 복수의 셀 중 둘 이상에 인가하는 단계를 포함하는, 방법.
- [0598] 항목 70. 항목 68 또는 69에 있어서,
- [0599] 제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0600] 결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계를 포함하는, 방법.
- [0601] 항목 71. 항목 68 또는 69에 있어서,
- [0602] 제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0603] 결합 전압을 복수의 셀 중 둘 이상에 인가하는 단계를 포함하는, 방법.
- [0604] 항목 72. 항목 68 또는 69에 있어서,
- [0605] 제2 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 제2 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 제2 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0606] 결합 전압을 복수의 셀 중 각 셀로 인가하는 단계를 포함하는, 방법.
- [0607] 항목 73. 항목 68 내지 72 중 어느 하나에 있어서, 형광 검출기를 사용해, 복수의 셀 중 하나 이상의 형광량을 측정하는 단계를 포함하는, 방법.
- [0608] 항목 74. 항목 61 내지 73 중 어느 하나에 있어서, 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계를 포함하는, 방법.
- [0609] 항목 75. 항목 61 내지 74 중 어느 하나에 있어서, 핵산 분자 분자의 결합은 짐작 말단 결합을 포함하는, 방법.
- [0610] 항목 76. 항목 61 내지 75 중 어느 하나에 있어서, 핵산 분자 분자의 결합은 평활 말단 결합을 포함하는, 방법.
- [0611] 항목 77. 항목 61 내지 76 중 어느 하나에 있어서, 결합 전압의 각각의 인가 후, 버퍼 저장소로부터 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소(waste reservoir) 내로 유동시키는 단계를 포함하는, 방법.
- [0612] 항목 78. 항목 61 내지 77 중 어느 하나에 있어서, 복수의 셀 중 하나 이상에 방출 전압(release voltage)을 인가하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.
- [0613] 항목 79. 항목 61 내지 77 중 어느 하나에 있어서, 복수의 셀 중 하나 이상에서 전기장을 형성하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.
- [0614] 항목 80. 항목 61 내지 77 중 어느 하나에 있어서, 복수의 셀 중 하나 이상을 가열하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.
- [0615] 항목 81. 항목 61 내지 77 중 어느 하나에 있어서, 효소 저장소로부터, 효소를 포함하는 유체의 볼륨을 메인 채널을 통해 유동시키는 단계 및 상기 효소가 핵산 어댑터와 반응하게 하여, 복수의 핵산 분자의 일부를 방출하는 단계를 포함하는, 방법.
- [0616] 항목 82. 항목 78 내지 81 중 어느 하나에 있어서, 상기 메인 채널로부터, 출력 채널을 통해, 출력 유체 볼륨을 도착지 저장소 내로 유동시키는 단계를 포함하는, 방법.
- [0617] 항목 83. 항목 61 내지 82 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자를 포함하는, 방법.
- [0618] 항목 84. 항목 61 내지 82 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 방법.
- [0619] 항목 85. 항목 83 또는 84에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 방법.

- [0620] 항목 86. 핵산 서열 내 디지털 정보를 처리하기 위한 방법으로서, 상기 방법은,
- [0621] (i) 출발지 저장소 내 유체에 현탁된 핵산 분자의 풀을 획득하는 단계,
- [0622] (ii) 출발지 저장소로부터, 입력 채널을 통해, 복수의 핵산 분자를 포함하는 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 메인 채널은 복수의 셀을 포함하고, 각 셀은 복수의 전기 전도성 판 중 하나씩과 상대 전극의 일부를 포함하며, 복수의 전도성 판 및 상대 전극은 메인 채널의 제1 차원을 따라 서로 반대편에 배치되고, 각 전기 전도성 판은 그 상에 부착된 복수의 핵산 어댑터를 포함함 - ,
- [0623] (iii) 복수의 셀 중 제1 셀에 결합 전압(binding voltage)을 인가함으로써, 복수의 핵산 분자의 일부를 복수의 전기 전도성 판 중 제1 전기 전도성 판의 핵산 어댑터에 결합시키는 단계,
- [0624] (iv) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
- [0625] (v) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계,
- [0626] (vi) 제2 출발지 저장소로부터, 입력 채널을 통해, 제2 복수의 핵산 분자를 포함하는 제2 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계,
- [0627] (vii) 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제2 복수의 핵산 분자의 일부를 핵산 어댑터에 결합된 복수의 핵산 분자의 일부에 결합하는 단계, 및
- [0628] (viii) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
- [0629] (ix) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계를 포함하는, 방법.
- [0630] 항목 87. 항목 86에 있어서,
- [0631] (x) 제n 출발지 저장소로부터, 입력 채널을 통해, 제n 복수의 핵산 분자를 포함하는 제n 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계,
- [0632] (xi) 복수의 셀 중 제1 셀에 결합 전압을 인가하여, 제n 복수의 핵산 분자의 일부를 제(n-1) 복수의 핵산 분자의 결합된 일부에 결합하는 단계,
- [0633] (xii) 버퍼 저장소로부터, 메인 채널을 통해 버퍼 유체의 볼륨을 폐기물 저장소 내로 유동시키는 단계,
- [0634] (xiii) 리가제 저장소로부터, 메인 채널을 통해 리가제를 포함하는 유체를 유동시키는 단계, 및
- [0635] 단계 x - xiii를 n회 수행하는 단계 - n은 3이상임 - 를 포함하는, 방법.
- [0636] 항목 88. 항목 86 또는 87에 있어서,
- [0637] (xiv) 메인 채널을 저장 장치에 저장하는 단계를 포함하는, 방법.
- [0638] 항목 89. 항목 86 내지 88 중 어느 하나에 있어서, 셀에 방출 전압(release voltage)을 인가하여, 결합된 핵산 분자를 방출하는 단계를 포함하는, 방법.
- [0639] 항목 90. 항목 89에 있어서, 상기 메인 채널로부터, 출력 채널을 통해, 출력 유체 볼륨을 도착지 저장소 내로 유동시키는 단계를 포함하는, 방법.
- [0640] 항목 91. 항목 90에 있어서, 도착지 저장소를 저장 장치에 저장하는 단계를 포함하는, 방법.
- [0641] 항목 92. 항목 86 내지 91 중 어느 하나에 있어서,
- [0642] 단계 (ix) 또는 (xiii) 중 하나 이상 후에, 품질 관리(QC) 출발지 저장소로부터, 입력 채널을 통해, 형광단(fluorophore)을 포함하는 복수의 태깅된 핵산 분자를 포함하는 QC 입력 유체 볼륨을 메인 채널 내로 유동시키는 단계, 및
- [0643] 결합 전압을 복수의 셀 중 제1 셀로 인가하는 단계를 포함하는, 방법.
- [0644] 항목 93. 항목 86 내지 92 중 어느 하나에 있어서,
- [0645] 방출 전압을 셀에 인가하여, 결합된 핵산 분자를 방출하는 단계,
- [0646] 메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 시퀀싱 장치로 방출된 핵산 분자

를 지향시키는 단계,

- [0647] 방출된 핵산 분자를 시퀀싱 장치를 통해 캐비티 내로 지향시키는 단계,
- [0648] 시퀀싱 장치에서 복수의 전압을 측정하는 단계, 및
- [0649] 측정된 복수의 전압에 기초하여 염기 콜링(base calling)을 수행하는 단계를 포함하는, 방법.
- [0650] 항목 94. 제93항에 있어서, 방출된 핵산 분자는 단일 가닥 DNA 분자이거나 이를 포함하는, 방법.
- [0651] 항목 95. 항목 93 또는 94에 있어서, 상대 전극과 시퀀싱 장치의 배출구의 다운스트림에 배치된 캐비티 전극 사이에 전압을 인가하는 단계를 포함하는, 방법.
- [0652] 항목 96. 항목 86 내지 92 중 어느 하나에 있어서,
- [0653] 방출 전압을 복수의 셀에 인가하여, 복수의 셀의 각각으로부터 결합된 핵산 분자를 방출하는 단계,
- [0654] 연산자 출발지 저장소로부터, 입력 채널을 통해, 복수의 연산자 핵산 분자를 포함하는 연산자 유체 볼륨을 메인 채널 내로 유동시키는 단계 - 상기 연산자 핵산 분자는 논리 연산자에 대응함 - , 및
- [0655] 연산자 핵산 분자와 방출된 핵산 분자의 화학 반응을 수행하여, 복수의 결과 핵산 분자를 생성하는 단계를 포함하는, 방법.
- [0657] 항목 97. 항목 96에 있어서,
- [0658] 메인 채널과 유체 연통하는 유입구 및 캐비티와 유체 연통하는 배출구를 갖는 시퀀싱 장치로 결과 핵산 분자를 지향시키는 단계,
- [0659] 결과 핵산 분자를 시퀀싱 장치를 통해 캐비티 내로 지향시키는 단계,
- [0660] 시퀀싱 장치에서 복수의 전압을 측정하는 단계, 및
- [0661] 측정된 복수의 전압에 기초하여 염기 콜링(base calling)을 수행하는 단계를 포함하는, 방법.
- [0662] 항목 98. 항목 86 내지 97 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링에 대응하는 디지털 정보를 인코딩하는 식별자 핵산 분자를 포함하는, 방법.
- [0663] 항목 99. 항목 86 내지 97 중 어느 하나에 있어서, 핵산 분자는 길이 L의 심볼의 스트링으로부터의 디지털 정보를 저장하는 식별자 핵산 분자의 복수의 구성요소 핵산 분자를 포함하는, 방법.
- [0664] 항목 100. 항목 98 또는 99에 있어서, 각 개별 식별자 핵산 분자는 심볼 값 및 심볼의 스트링 내 심볼 위치에 대응하며, 식별자 핵산 분자의 풀은 길이 L을 갖는 심볼의 임의의 스트링을 인코딩할 수 있는 식별자 라이브러리 내 식별자 핵산 서열의 서브세트에 대응하는, 방법.
- [0665] 항목 101. 핵산 서열을 판독하기 위한 장치로서, 상기 장치는,
- [0666] 기관 내에 배치되고 입력 가닥을 포함하는 입력 핵산 분자를 수신하도록 구성된 나노-채널, 및
- [0667] 상기 나노-채널 상에 또는 내에 배치된 센서 장치 - 상기 센서 장치는 전자 감지 장치를 포함하고, 상기 전자 감지 장치는 게이트 전압을 갖는 전자 게이트를 갖고, 게이트 전압은 게이트의 소스-드레인 전류를 변경하기 위해 입력 핵산 분자의 이동 판독 구성요소의 전하에 의해 변조될 수 있음 - 를 포함하는, 장치.
- [0668] 항목 102. 항목 101에 있어서, 상기 장치는 항목 5 내지 42 중 어느 하나의 시스템의 일부인, 장치.
- [0669] 항목 103. 항목 101에 있어서, 상기 장치는 메인 채널 내에 배치되는, 장치.
- [0670] 항목 104. 항목 101 내지 103 중 어느 하나에 있어서, 상기 센서 장치는 금속-옥사이드-반도체 전계 효과 트랜지스터(MOSFET)거나 이를 포함하는, 장치.
- [0671] 항목 105. 항목 101 내지 103 중 어느 하나에 있어서, 상기 센서 장치는 전해질 옥사이드 전계 효과 트랜지스터(EOSFET)거나 이를 포함하는, 장치.
- [0672] 항목 106. 항목 101 내지 105 중 어느 하나에 있어서, 판독 구성요소는 단일 가닥 핵산 분자의 섹션에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는, 장치.

- [0673] 항목 107. 항목 101 내지 106 중 어느 하나에 있어서, 복수의 판독 구성요소 - 각 판독 구성요소는 입력 가닥의 상보적 섹션에 혼성화되어 입력 핵산 분자를 형성하도록 구성됨 - 를 포함하는, 장치.
- [0674] 항목 108. 항목 101 내지 107 중 어느 하나에 있어서, 제1 판독 구성요소는 제1 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되며, 제2 판독 구성요소는 제2 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되는, 장치.
- [0675] 항목 109. 항목 108에 있어서, 제1 판독 구성요소는, 게이트를 통해 전좌할 때, 게이트에서의 소스-드레인 전류의 제1 변화를 야기하고, 제2 판독 구성요소는, 게이트를 통해 전좌할 때, 게이트에서의 소스-드레인 전류의 제2 변화를 야기하는, 장치.
- [0676] 항목 110. 항목 109에 있어서, 상기 제1 변화는 상기 제2 변화와 상이한, 장치.
- [0677] 항목 111. 항목 101 내지 108 중 어느 하나에 있어서, 시작 판독 구성요소 및 종료 판독 구성요소 - 상기 시작 판독 구성요소는 입력 가닥의 제1 말단에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하고 종료 판독 구성요소는 입력 가닥의 제2 말단에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함함 - 를 포함하는, 장치.
- [0678] 항목 112. 항목 101 내지 111 중 어느 하나에 있어서, 입력 가닥은 디지털 정보를 인코딩하는, 장치.
- [0679] 항목 113. 항목 101 내지 112 중 어느 하나에 있어서, 입력 가닥은 하나 이상의 식별자 구성요소를 포함하고, 각 식별자 구성요소는 디지털 정보를 인코딩하는 핵산 식별자의 구성요소인, 장치.
- [0680] 항목 114. 항목 101 내지 113 중 어느 하나에 있어서, 입력 가닥은 제1 식별자 구성요소에 대응하는 제1 입력 서열 및 제2 식별자 구성요소에 대응하는 제2 입력 서열을 포함하는, 장치.
- [0681] 항목 115. 항목 101 내지 114 중 어느 하나에 있어서, 제1 식별자 구성요소 및 제2 식별자 구성요소의 적어도 일부분에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는 오버랩 판독 구성요소를 포함하는, 장치.
- [0682] 항목 116. 항목 115에 있어서, 상기 오버랩 판독 구성요소는 플랩(flap)을 형성하는 비-상보적 핵산 섹션 및 상기 플랩의 말단에 2차 분자 구조를 갖는 핵산 섹션을 포함하는, 장치.
- [0683] 항목 117. 항목 115에 있어서, 상기 오버랩 판독 구성요소는 플랩을 형성하는 비-상보적 구성요소 및 상기 플랩에 혼성화되는 2차 분자 구조를 갖는 구성요소를 포함하는, 장치.
- [0684] 항목 118. 항목 101 내지 117 중 어느 하나에 있어서, 상기 센서 장치는 하나 이상의 전자 신호 처리 장치이거나 나 이를 포함하는, 장치.
- [0685] 항목 119. 항목 101 내지 118 중 어느 하나에 있어서, 판독 구성요소는 2차 분자 구조를 갖는 섹션을 포함하는 단일 가닥 핵산 분자인, 장치.
- [0686] 항목 120. 항목 101 내지 118 중 어느 하나에 있어서, 판독 구성요소는 펩티드 압타머를 포함하는, 장치.
- [0687] 항목 121. 항목 101 내지 118 중 어느 하나에 있어서, 판독 구성요소는 덴드리머를 포함하는, 장치.
- [0688] 항목 122. 항목 101 내지 118 중 어느 하나에 있어서, 판독 구성요소는 단백질질을 포함하는, 장치.
- [0689] 항목 123. 항목 101 내지 122 중 어느 하나에 있어서, 복수의 센서 장치를 포함하는, 장치.
- [0690] 항목 124. 항목 101 내지 122 중 어느 하나에 있어서, 센서 장치는 복수의 전자 감지 장치를 포함하는, 장치.
- [0691] 항목 125. 항목 123 또는 124에 있어서, 복수의 센서 장치 또는 복수의 전자 감지 장치는 나노-채널의 경로를 따라 직렬로 배열되는, 장치.
- [0692] 항목 126. 항목 124에 있어서, 복수의 센서 장치 또는 복수의 전자 감지 장치는 둘 이상의 판독 구성요소를 동시에 판독하도록 구성되는, 장치.
- [0693] 항목 127. 핵산 서열을 판독하기 위한 방법으로서, 상기 방법은
- [0694] 항목 101 내지 126 중 어느 하나에 따른 장치를 제공하는 단계, 및
- [0695] 나노-채널을 통해 판독 구성요소를 전좌시키는 단계를 포함하는, 방법.
- [0696] 항목 128. 핵산 서열을 판독하기 위한 장치로서, 상기 장치는,

- [0697] 기질 내에 배치되고 입력 가닥을 포함하는 입력 핵산 분자를 수신하도록 구성된 나노-채널, 및
- [0698] 상기 나노-채널 상에 또는 내에 배치되는 센서 장치 - 상기 센서 장치는 광학 센싱 장치를 포함하고, 상기 광학 센싱 장치는 입력 핵산 분자의 전좌 판독 구성요소로부터 광 신호를 검출하도록 구성된 - 를 포함하는, 장치.
- [0699] 항목 129. 항목 128에 있어서, 상기 장치는 항목 5 내지 42 중 어느 하나의 시스템의 일부인, 장치.
- [0700] 항목 130. 항목 129에 있어서, 상기 장치는 메인 채널 내에 배치되는, 장치.
- [0701] 항목 131. 항목 128 내지 130 중 어느 하나에 있어서, 광학 감지 장치는 형광 측정 장치인, 장치.
- [0702] 항목 132. 항목 128 내지 131 중 어느 하나에 있어서, 광학 감지 장치는 광소자, 카메라, 또는 광자 카운터 중 하나 이상을 포함하는, 장치.
- [0703] 항목 133. 항목 128 내지 132 중 어느 하나에 있어서, 판독 구성요소는 단일 가닥 핵산 분자의 섹션에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는, 장치.
- [0704] 항목 134. 항목 128 내지 133 중 어느 하나에 있어서, 복수의 판독 구성요소 - 각 판독 구성요소는 입력 가닥의 상보적 섹션에 혼성화되어 입력 핵산 분자를 형성하도록 구성됨 - 를 포함하는, 장치.
- [0705] 항목 135. 항목 128 내지 134 중 어느 하나에 있어서, 제1 판독 구성요소는 제1 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되며, 제2 판독 구성요소는 제2 입력 서열을 갖는 입력 가닥의 하나 이상의 섹션에 혼성화되도록 구성되는, 장치.
- [0706] 항목 136. 항목 135에 있어서, 제1 판독 구성요소는 제1 광 신호를 발산하도록 구성된 발광 요소를 포함하고 제2 판독 구성요소는 제2 광 신호를 발산하도록 구성된 발광 요소를 포함하는, 장치.
- [0707] 항목 137. 항목 136에 있어서, 제1 광 신호는 제2 광 신호와 상이한, 장치.
- [0708] 항목 138. 항목 137에 있어서, 제1 광 신호는 제2 광 신호보다 큰 강도를 갖는, 장치.
- [0709] 항목 139. 항목 137에 있어서, 제1 광 신호는 제2 광 신호와 상이한 색상을 갖는, 장치.
- [0710] 항목 140. 항목 128 내지 139 중 어느 하나에 있어서, 입력 가닥은 디지털 정보를 인코딩하는, 장치.
- [0711] 항목 141. 항목 128 내지 140 중 어느 하나에 있어서, 입력 가닥은 하나 이상의 식별자 구성요소를 포함하고, 각 식별자 구성요소는 디지털 정보를 인코딩하는 핵산 식별자의 구성요소인, 장치.
- [0712] 항목 142. 항목 128 내지 141 중 어느 하나에 있어서, 입력 가닥은 제1 식별자 구성요소에 대응하는 제1 입력 서열 및 제2 식별자 구성요소에 대응하는 제2 입력 서열을 포함하는, 장치.
- [0713] 항목 143. 항목 128 내지 142 중 어느 하나에 있어서, 제1 식별자 구성요소 및 제2 식별자 구성요소의 적어도 일부분에 혼성화되도록 구성된 단일 가닥 핵산 분자를 포함하는 오버랩 판독 구성요소를 포함하는, 장치.
- [0714] 항목 144. 항목 143에 있어서, 상기 오버랩 판독 구성요소는 플랩(flap)을 형성하는 비-상보적 핵산 섹션 및 상기 플랩의 말단에 부착된 발광 요소를 포함하는, 장치.
- [0715] 항목 145. 항목 144에 있어서, 상기 오버랩 판독 구성요소는 플랩을 형성하는 비-상보적 구성요소 및 상기 플랩에 혼성화되는 2차 분자 구조 및 발광 요소를 갖는 구성요소를 포함하는, 장치.
- [0716] 항목 146. 항목 128 내지 145 중 어느 하나에 있어서, 센서 장치는 복수의 광학 감지 장치를 포함하는, 장치.
- [0717] 항목 147. 항목 128 내지 146 중 어느 하나에 있어서, 발광 요소는 형광단인, 장치.
- [0718] 항목 148. 항목 128 내지 147 중 어느 하나에 있어서, 상기 센서 장치는 하나 이상의 전자 신호 처리 장치이거나 나 이를 포함하는, 장치.
- [0719] 항목 149. 항목 128 내지 148 중 어느 하나에 있어서, 복수의 센서 장치를 포함하는, 장치.
- [0720] 항목 150. 항목 128 내지 148 중 어느 하나에 있어서, 센서 장치는 복수의 광학 감지 장치를 포함하는, 장치.
- [0721] 항목 151. 항목 149 또는 150에 있어서, 복수의 센서 장치 또는 복수의 전자 감지 장치는 나노-채널의 경로를 따라 직렬로 배열되는, 장치.
- [0722] 항목 152. 항목 150에 있어서, 복수의 센서 장치 또는 복수의 광학 감지 장치는 둘 이상의 판독 구성요소를 동

시에 관독하도록 구성되는, 장치.

[0723] 항목 153. 핵산 서열을 관독하기 위한 방법으로서, 상기 방법은

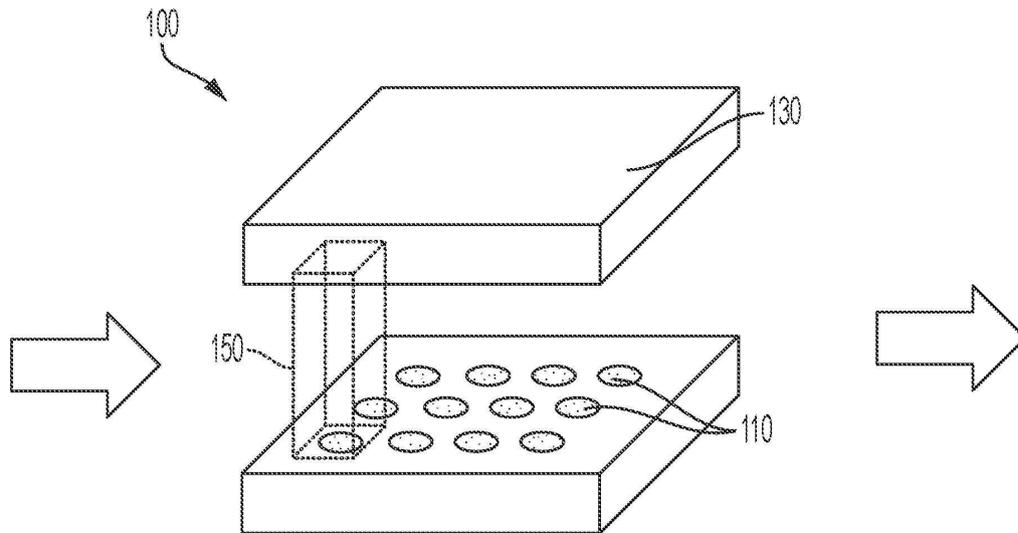
[0724] 항목 128 내지 152 중 어느 하나에 따른 장치를 제공하는 단계, 및

[0725] 나노-채널을 통해 관독 구성요소를 전좌시키는 단계를 포함하는, 방법.

[0726] 본 발명의 바람직한 실시예가 본 명세서에 도시되고 설명되었지만, 이러한 실시예는 단지 예로서 제공된다는 것이 통상의 기술자에게 명백할 것이다. 본 발명은 명세서 내에 제공된 특정 실시예에 의해 제한되도록 의도되지 않는다. 본 발명은 전술한 명세서를 참조하여 설명되었지만, 본 명세서의 실시예의 설명 및 예시는 제한적인 의미로 해석되는 것을 의미하지 않는다. 본 발명을 벗어나지 않으면서 당업자는 다양한 변형, 변화 및 대체를 할 수 있을 것이다. 또한, 본 발명의 모든 측면은 다양한 조건 및 변수에 따라 달라지는 본 명세서에 제시된 특정 묘사, 구성 또는 상대적 비율에 제한되지 않는다는 것이 이해되어야 한다. 본 명세서에 기술된 본 발명의 실시예에 대한 다양한 대안이 본 발명을 실시하는데 채용될 수 있다는 것을 이해해야 한다. 따라서 본 발명은 그러한 대안, 수정, 변형 또는 등가물도 포함해야 한다. 다음의 청구범위는 본 발명의 범위를 정의하고, 이들 청구범위 및 그 등가물 범위 내의 방법 및 구조가 이에 의해 포괄되도록 의도된다. 본 명세서에 인용된 모든 참고 문헌은 그 전체가 참고로 포함되어 있으며 본 출원의 일부를 구성한다.

**도면**

**도면1**

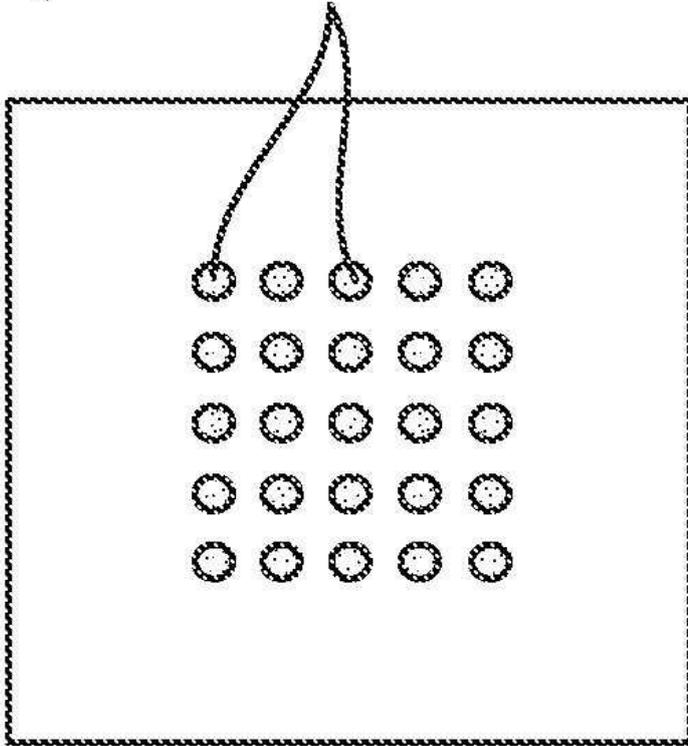


도면2

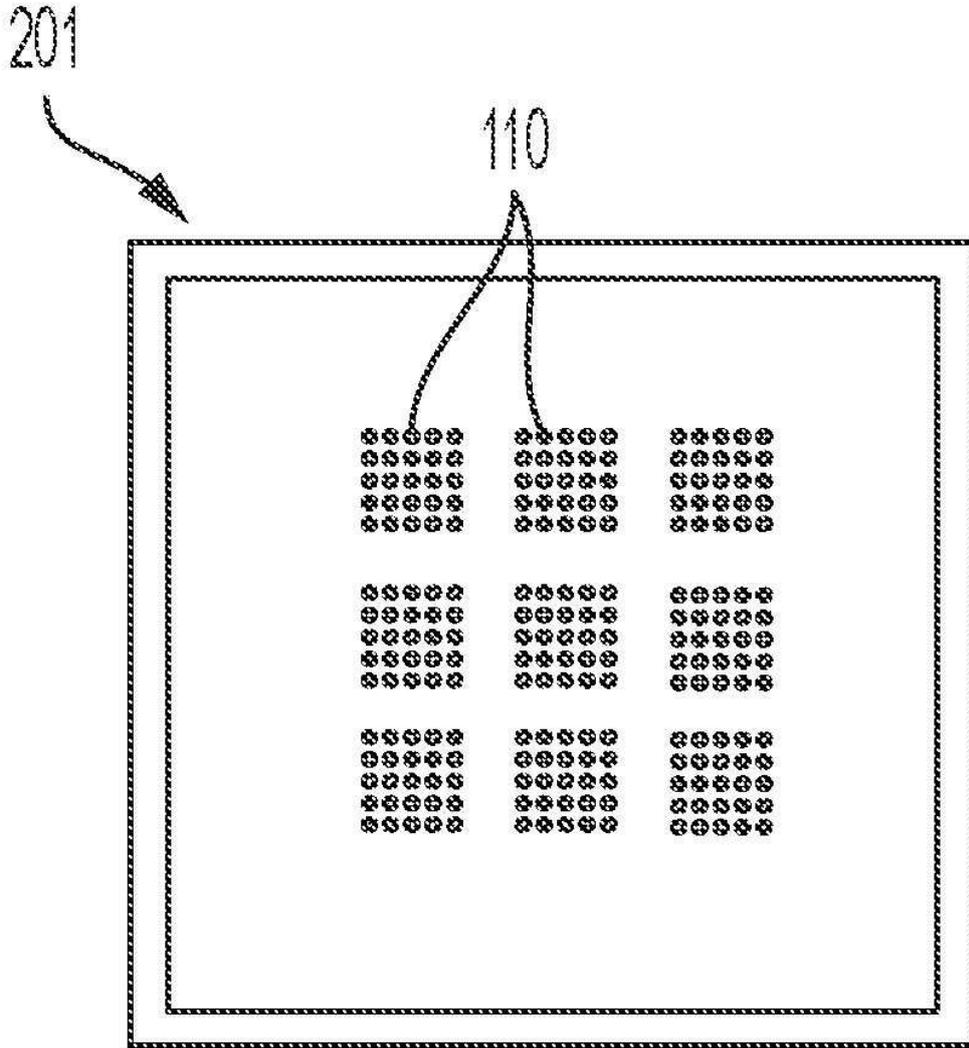
200



110



도면3

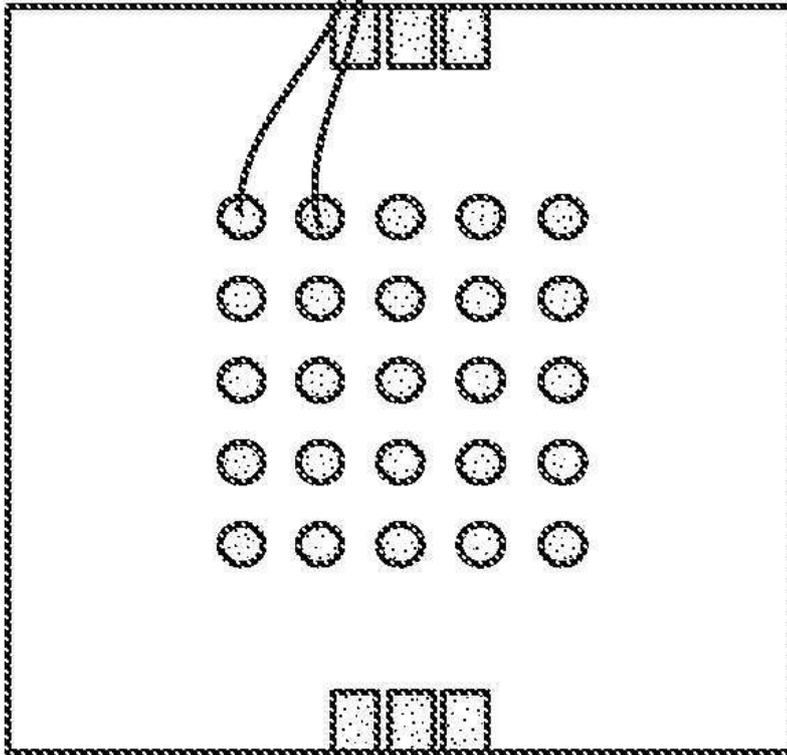


도면4a

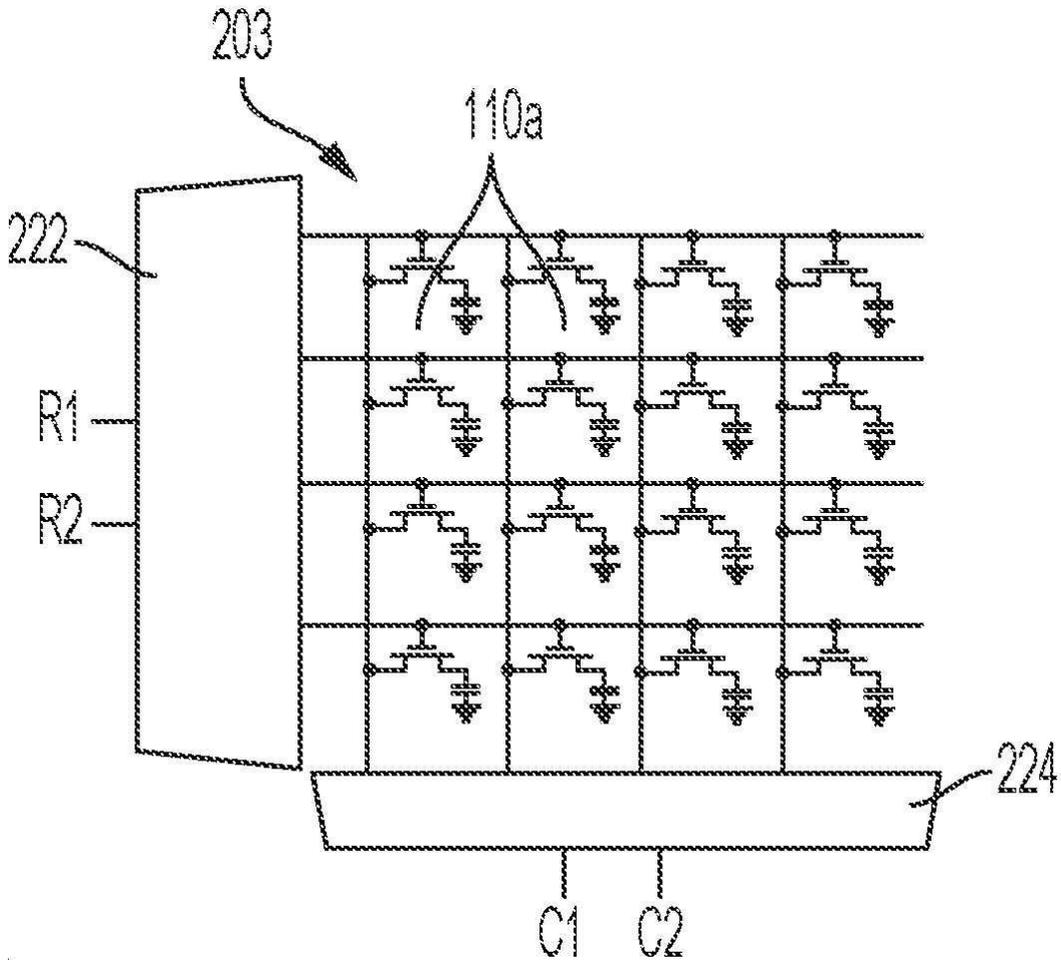
202



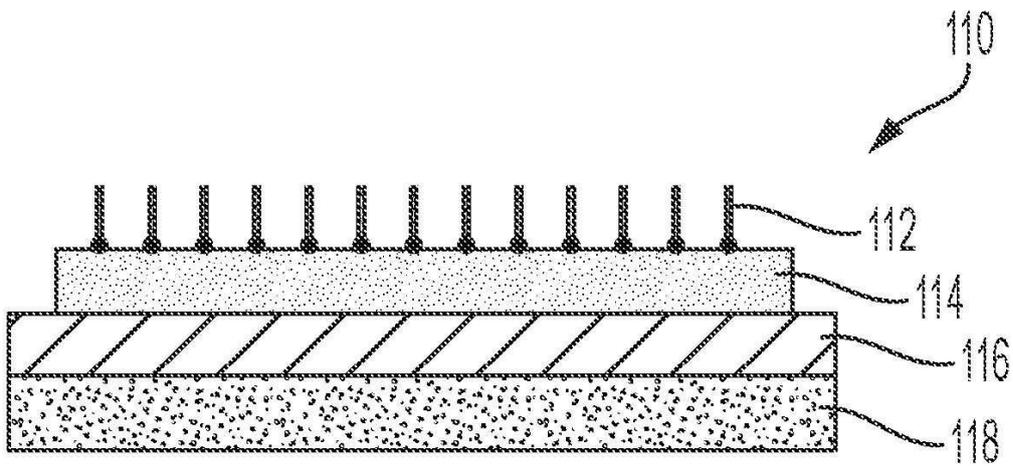
110



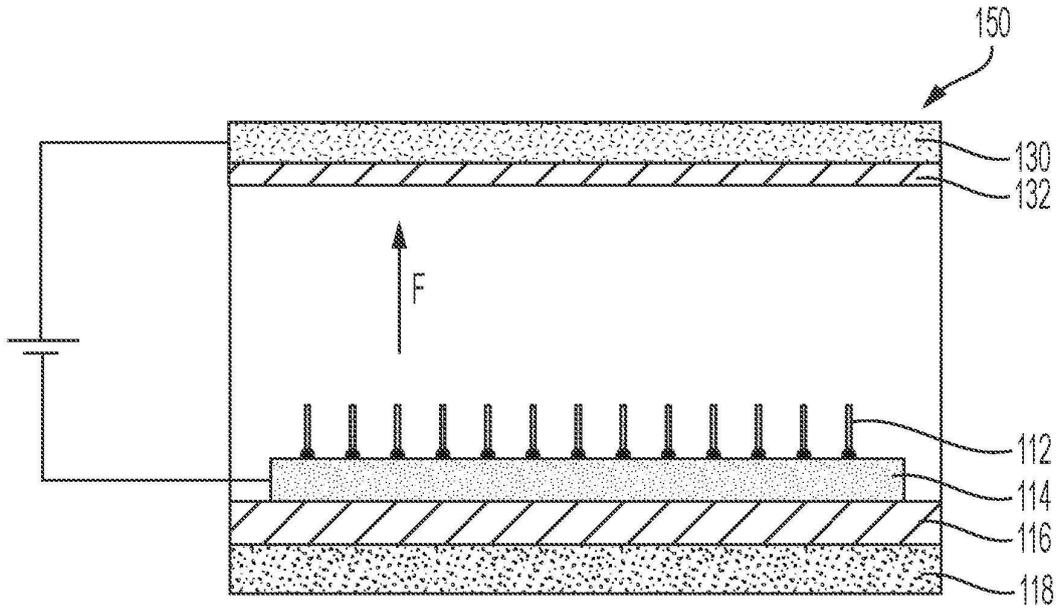
도면4b



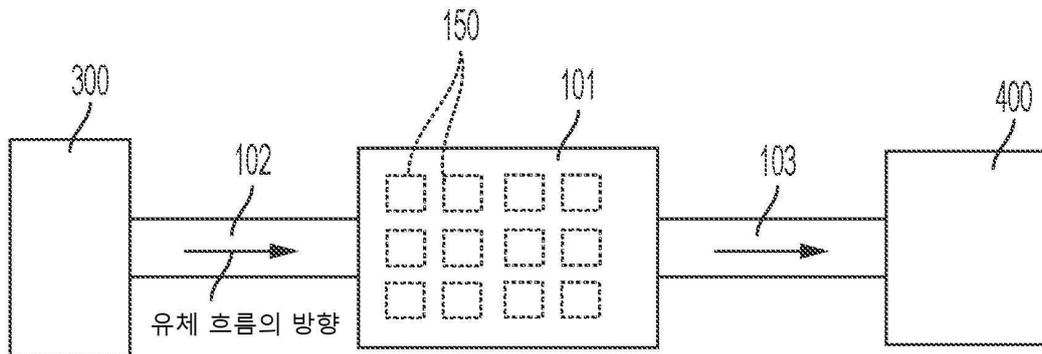
도면5



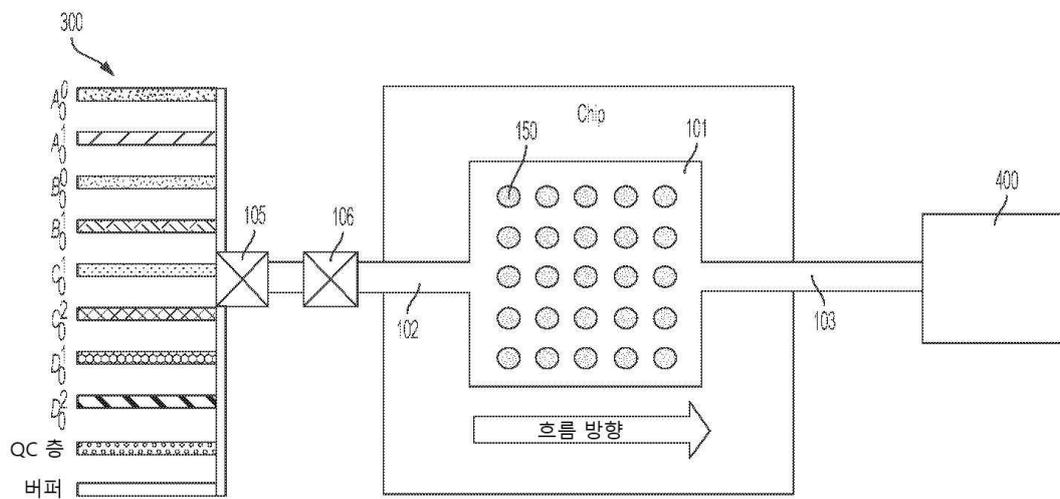
도면6



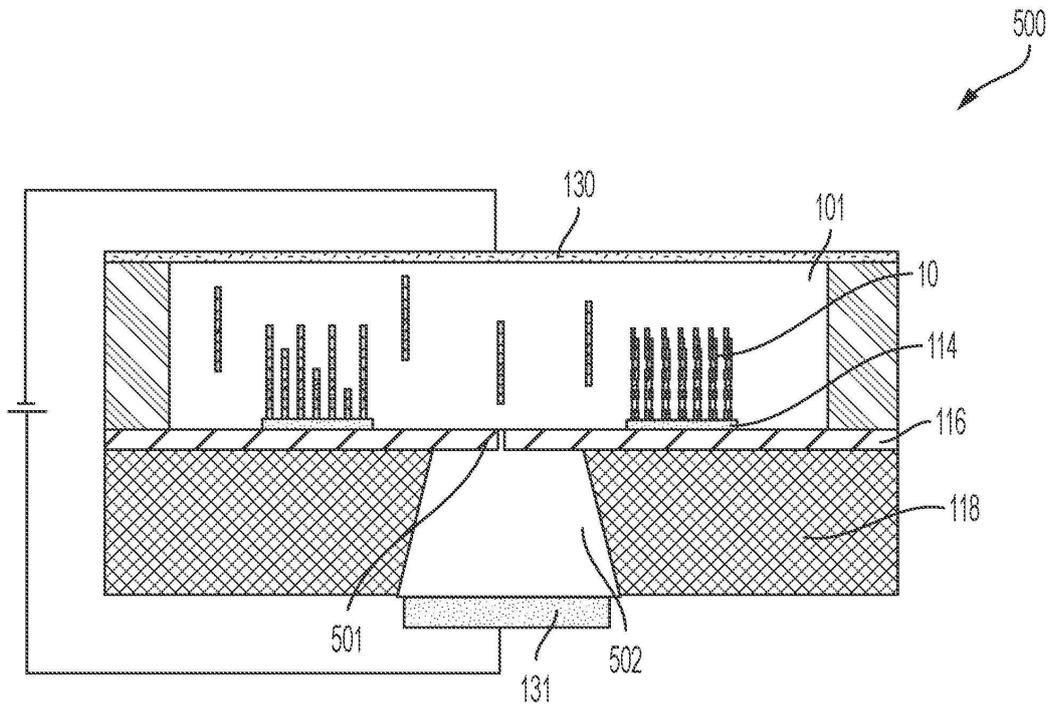
도면7



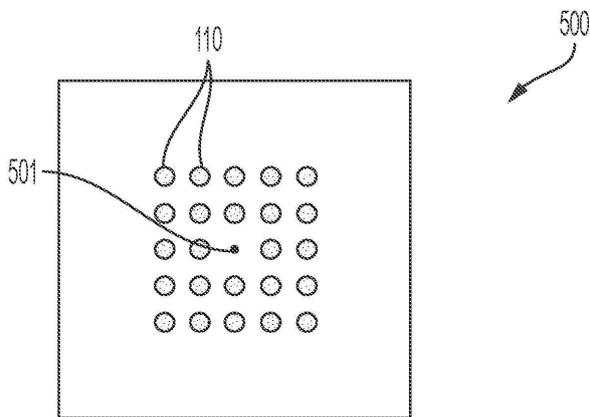
도면8



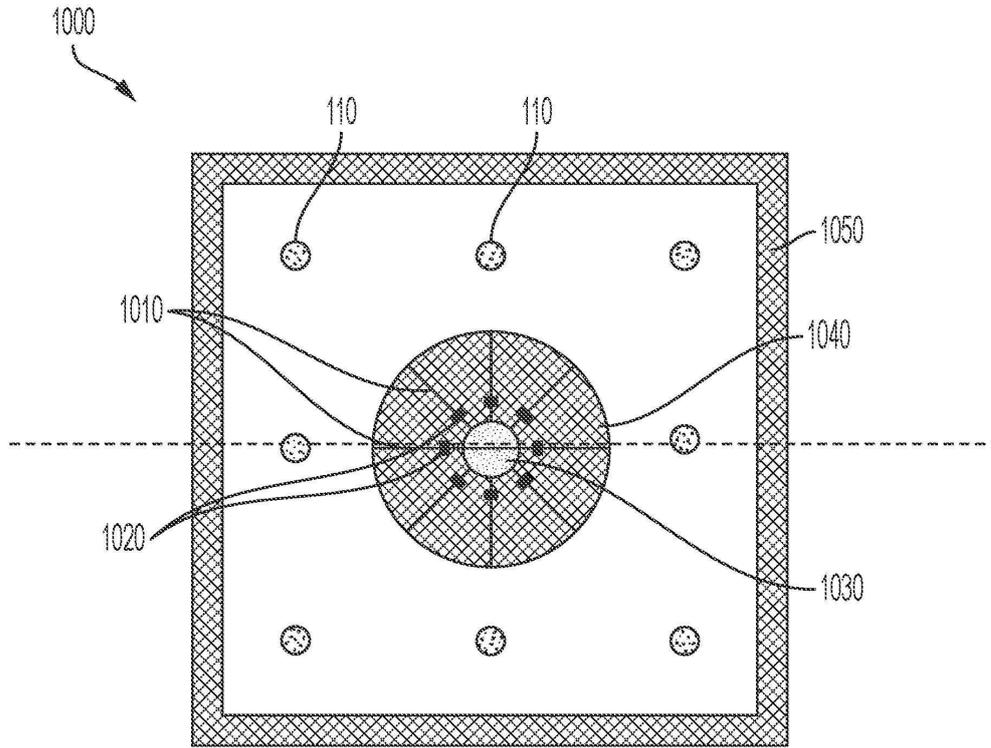
도면9a



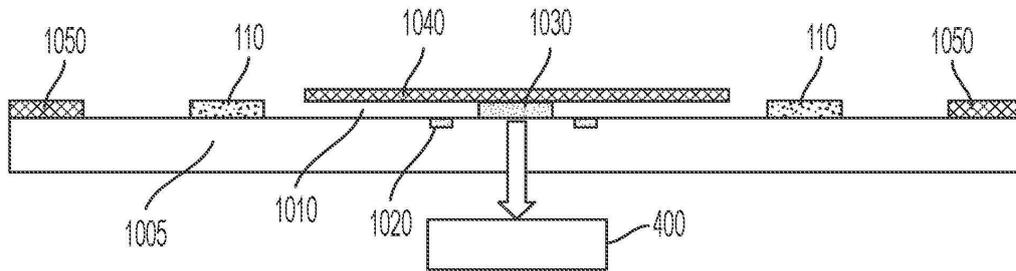
도면9b



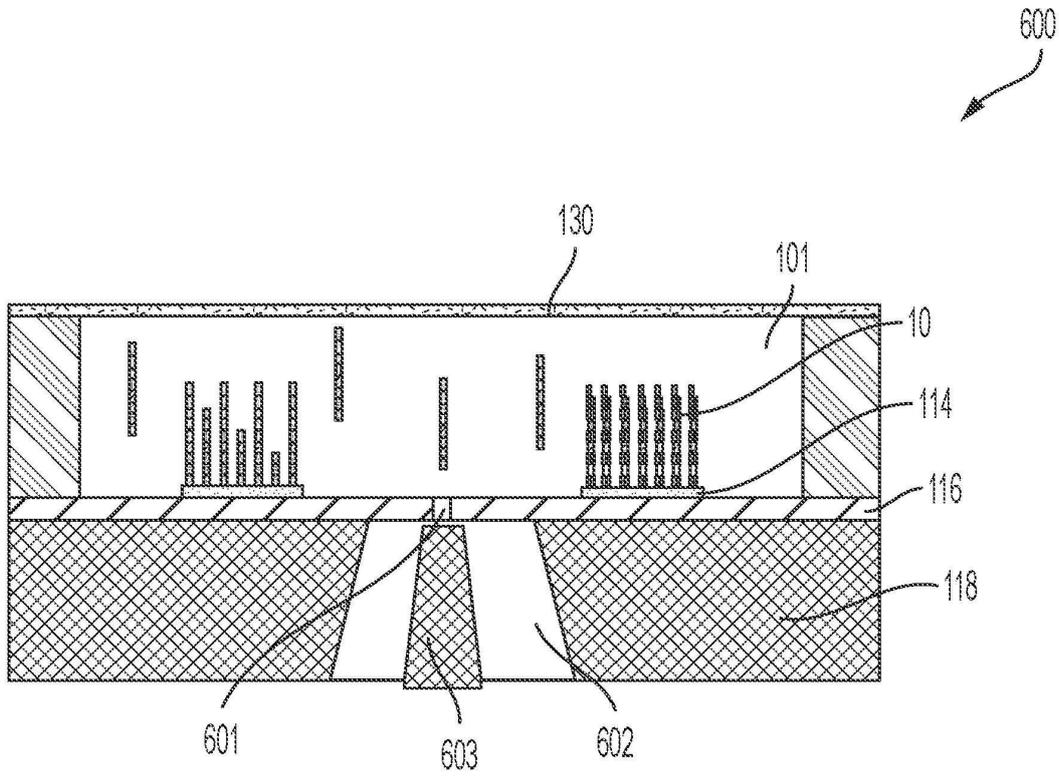
도면10a



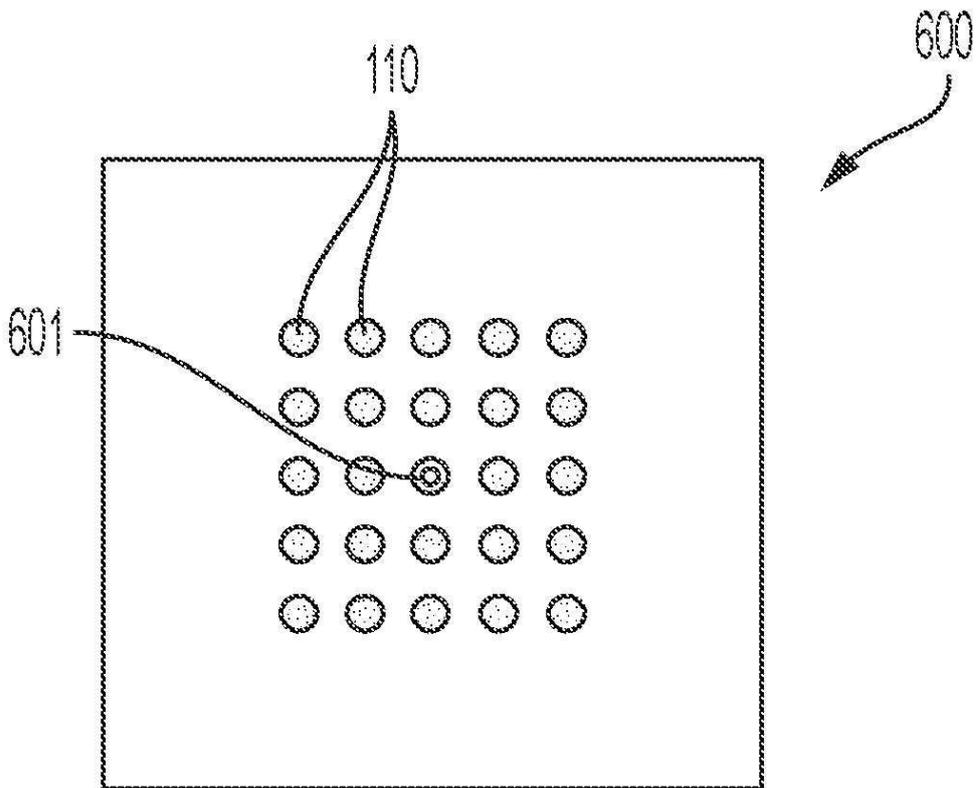
도면10b



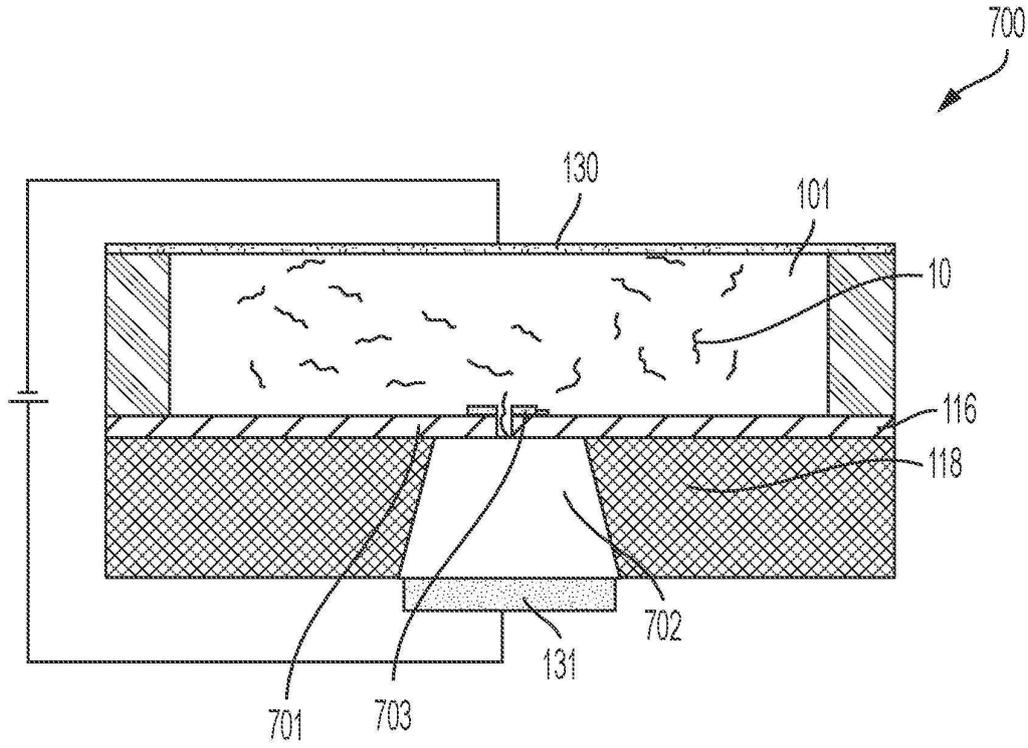
도면11a



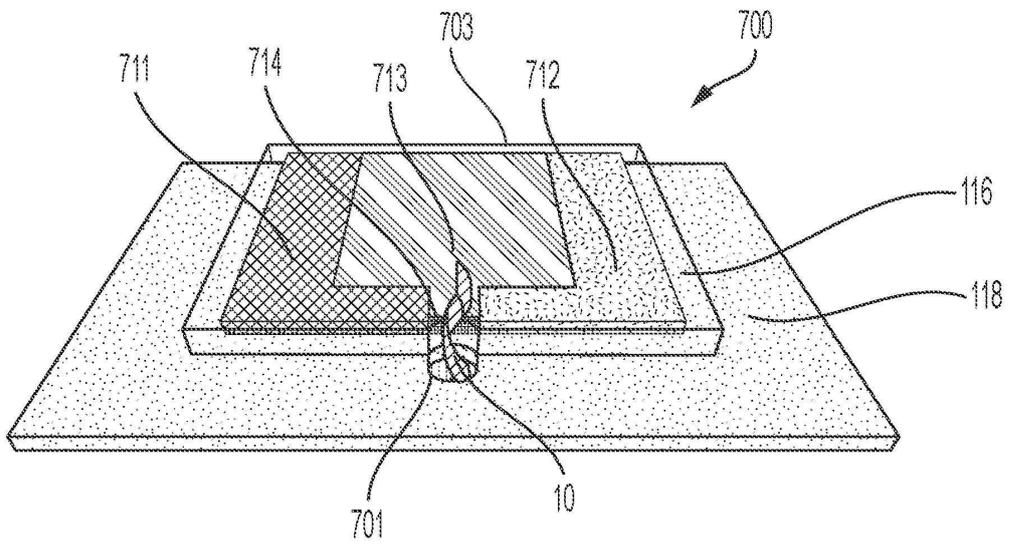
도면11b



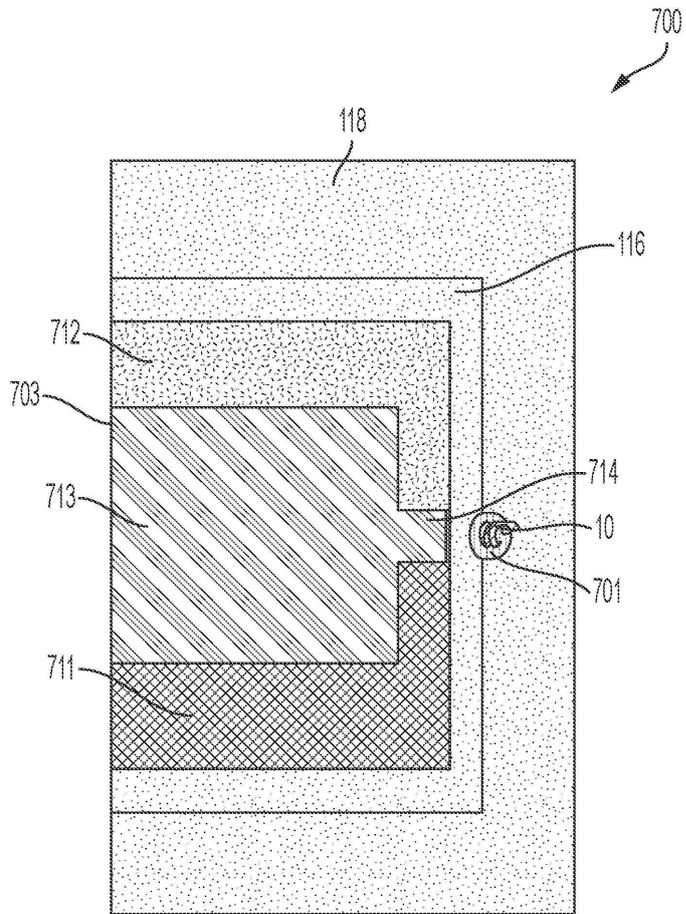
도면12



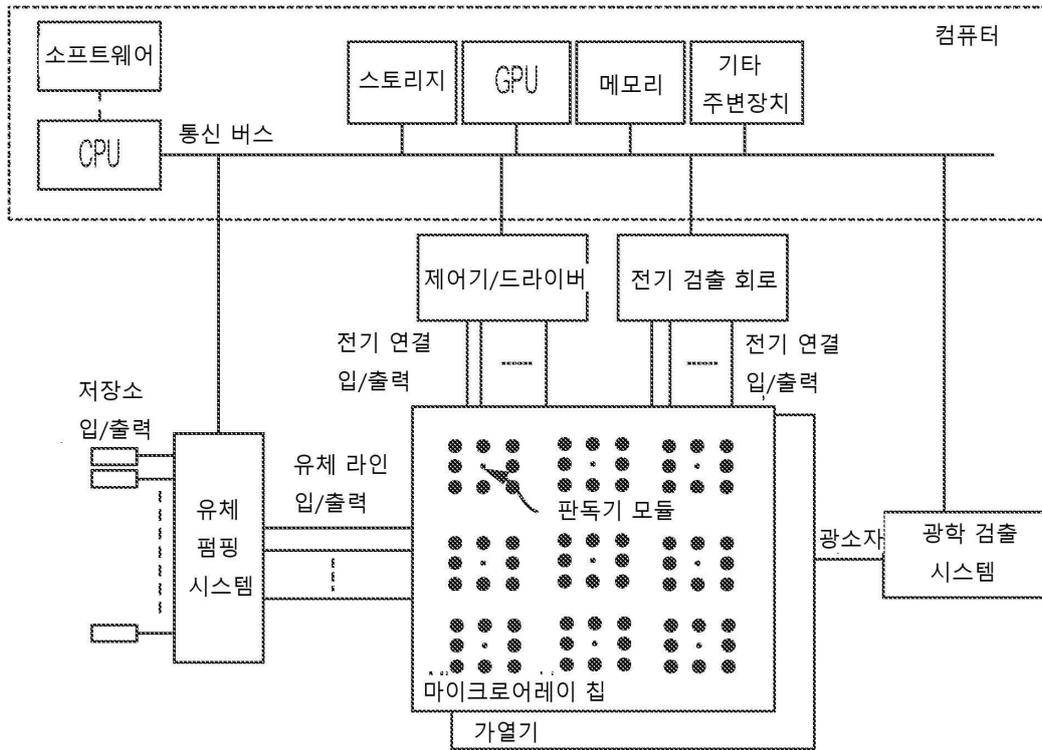
도면13



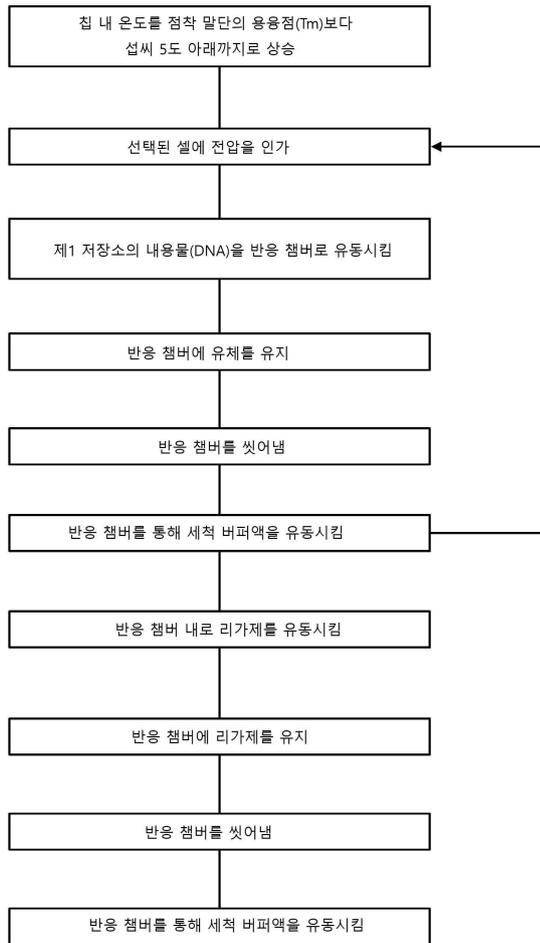
도면14



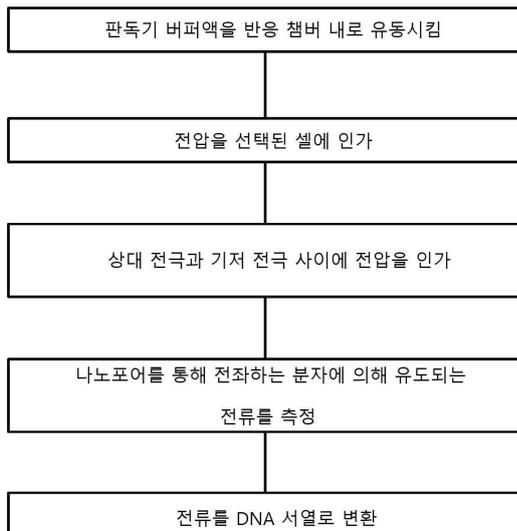
도면15



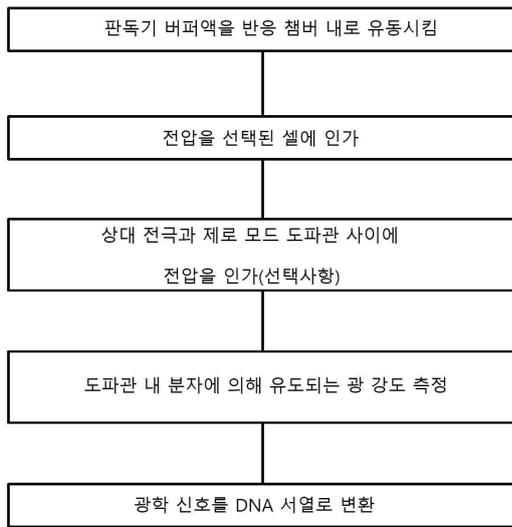
도면16



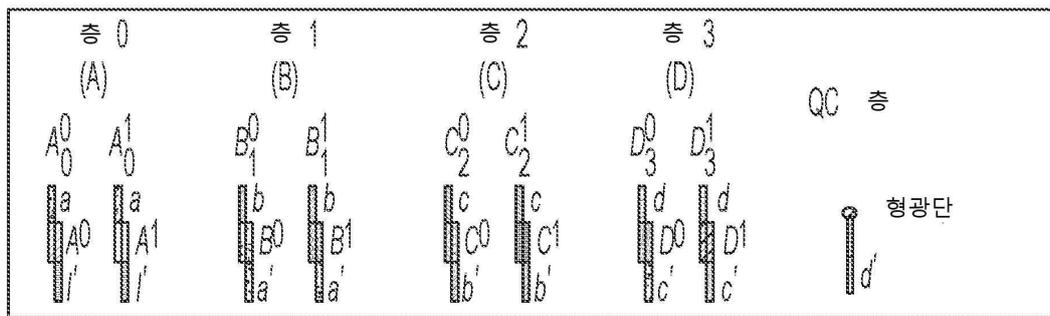
도면17



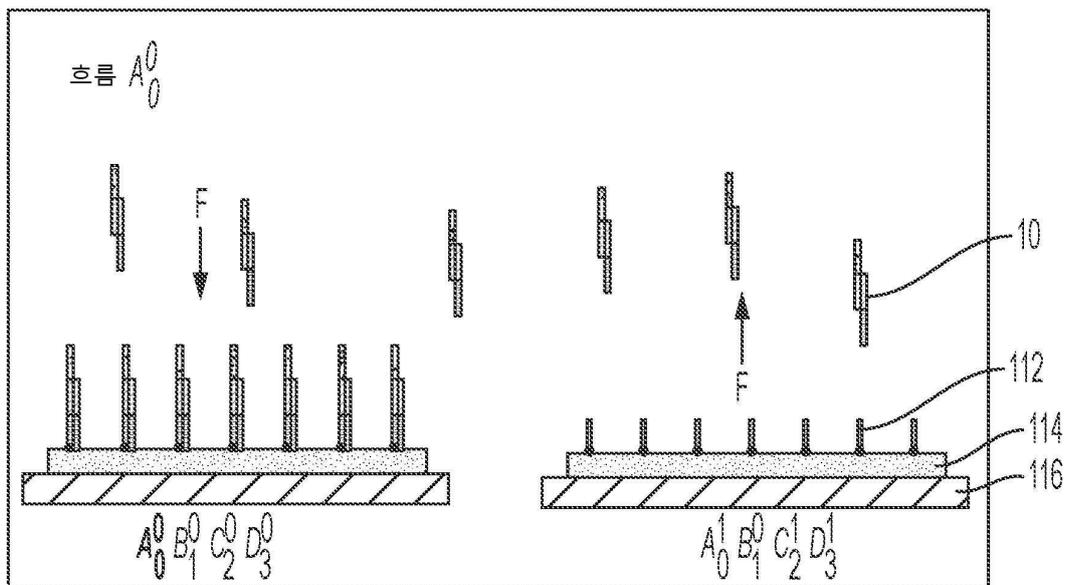
도면18



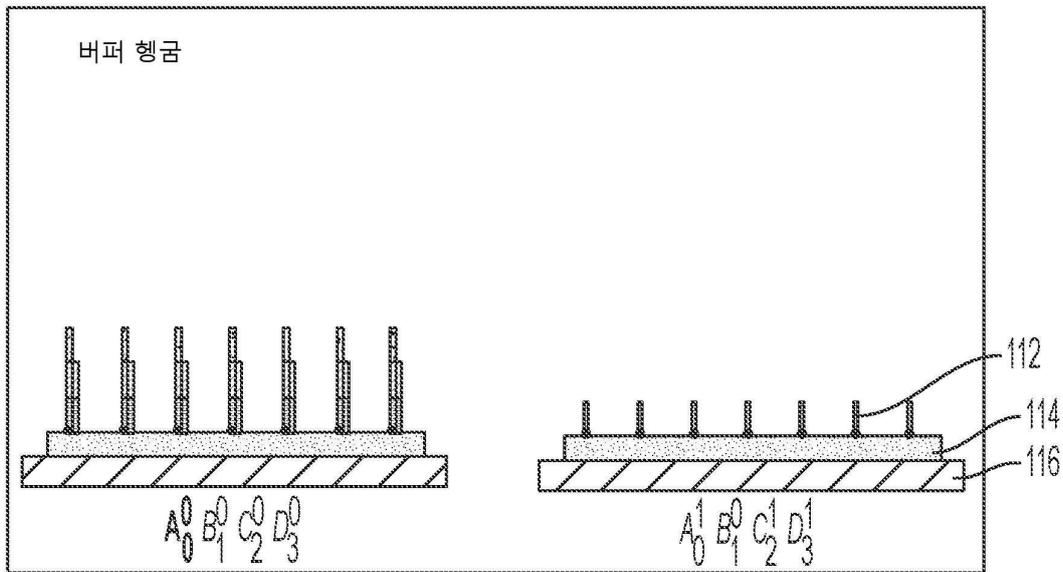
도면19



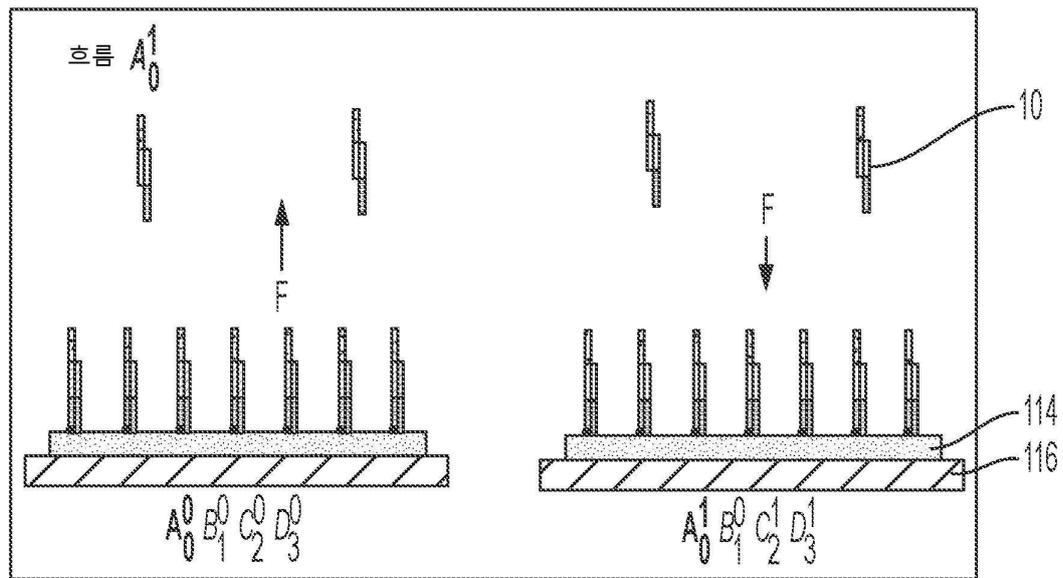
도면20



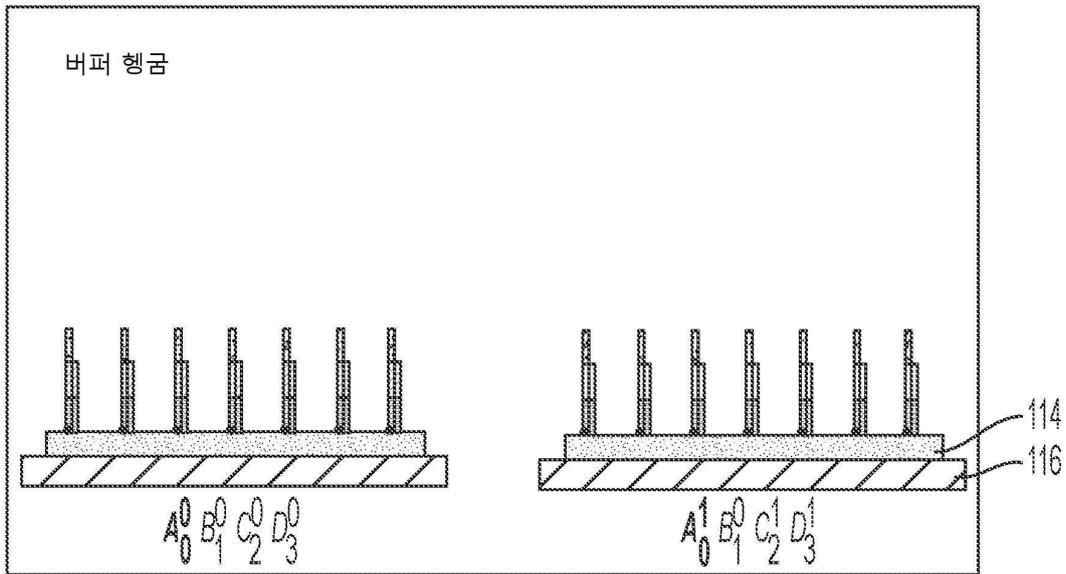
도면21



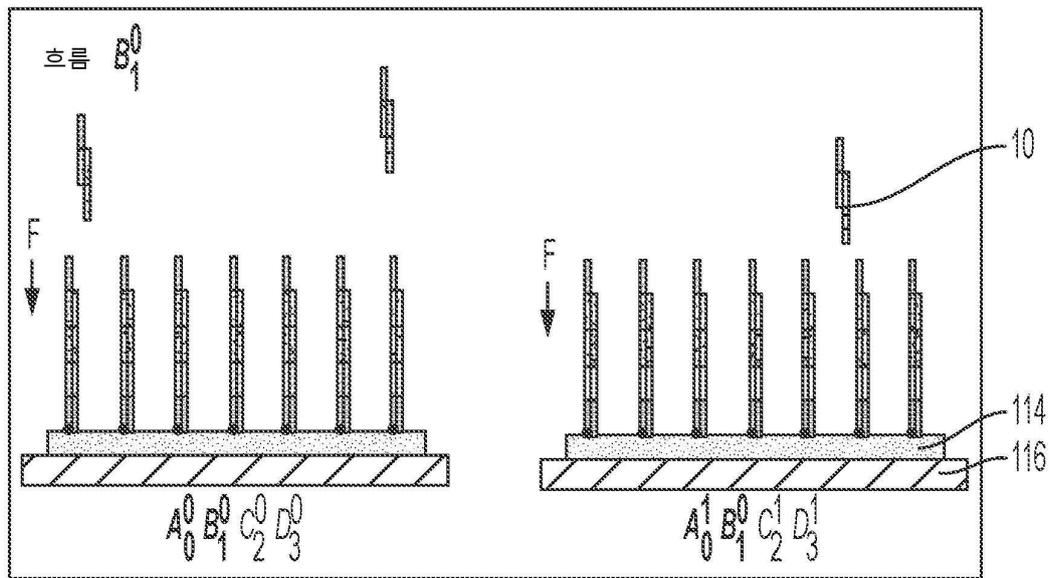
도면22



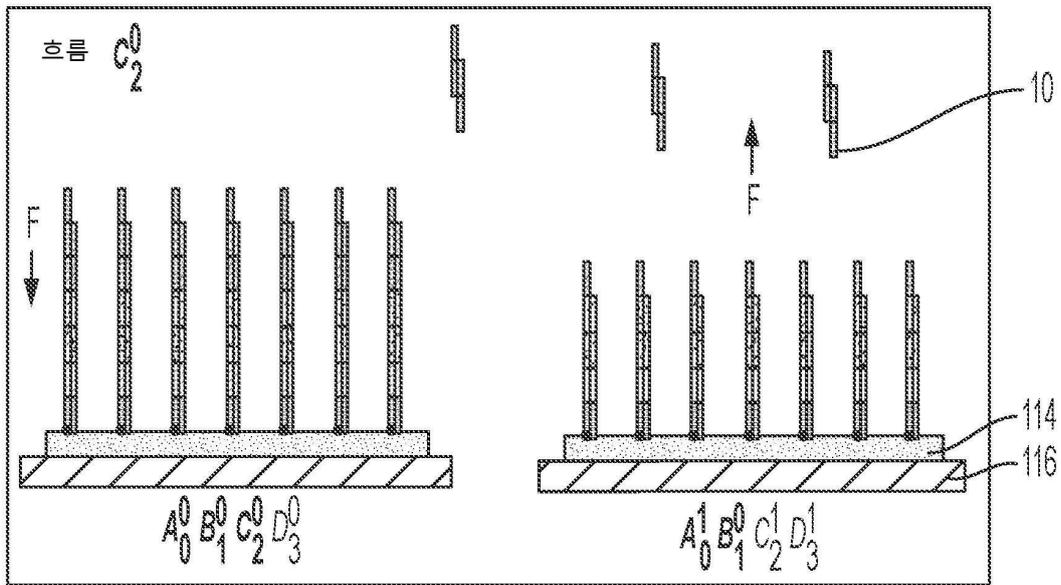
도면23



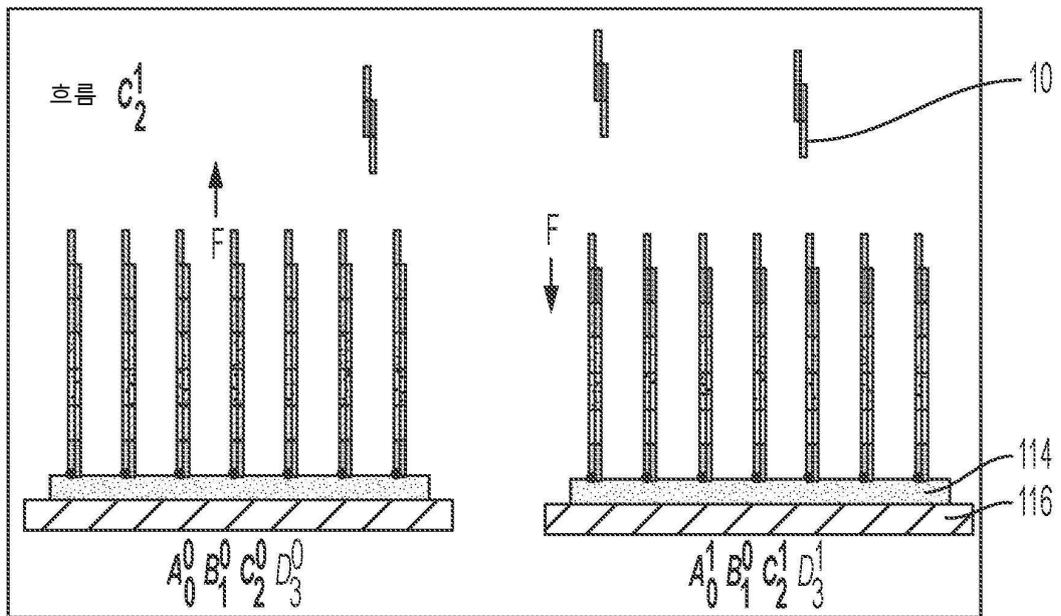
도면24



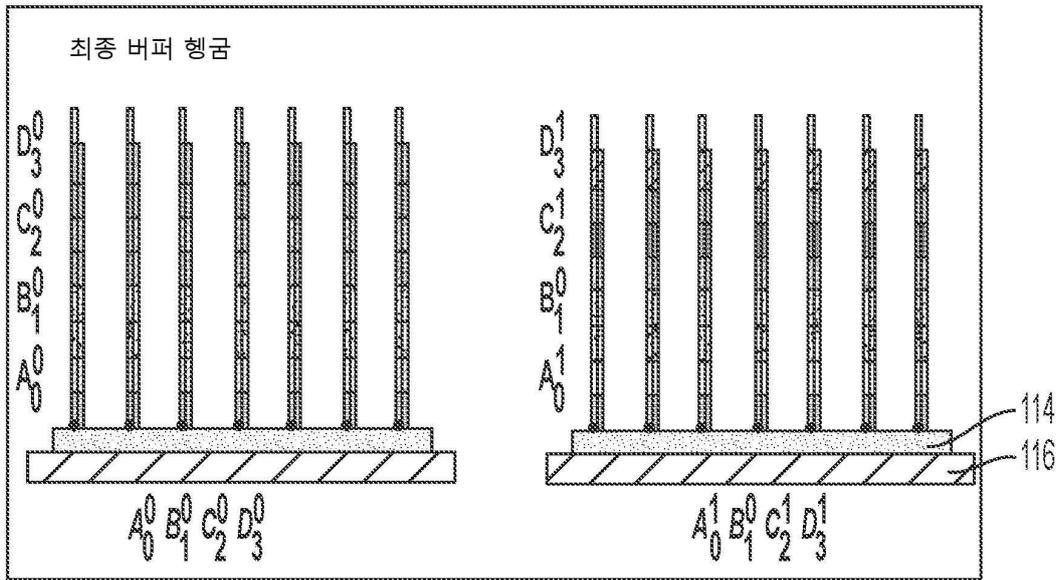
도면25



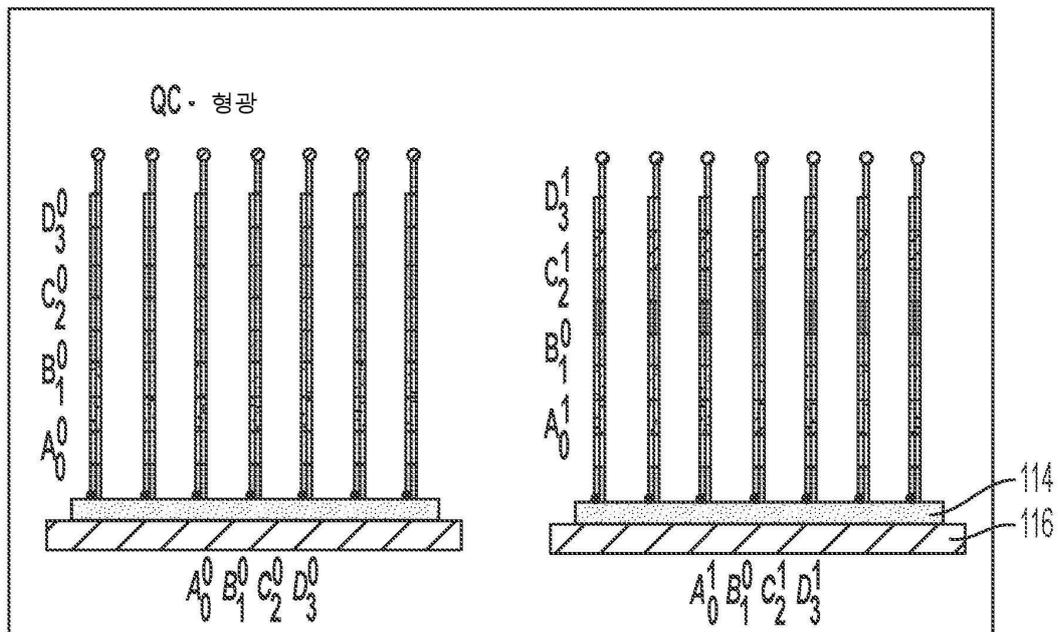
도면26



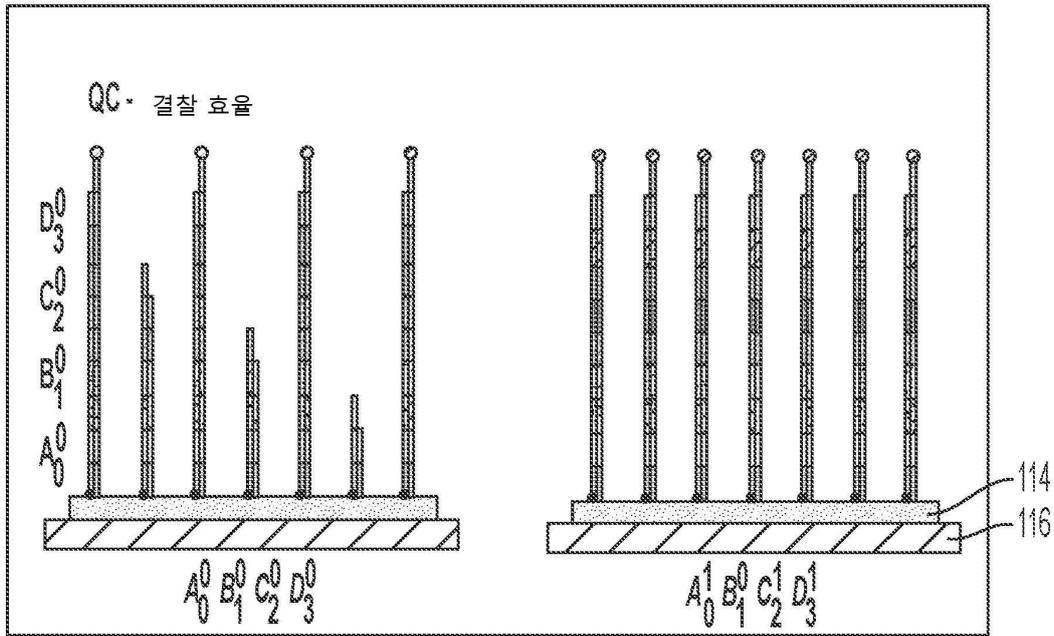
도면27



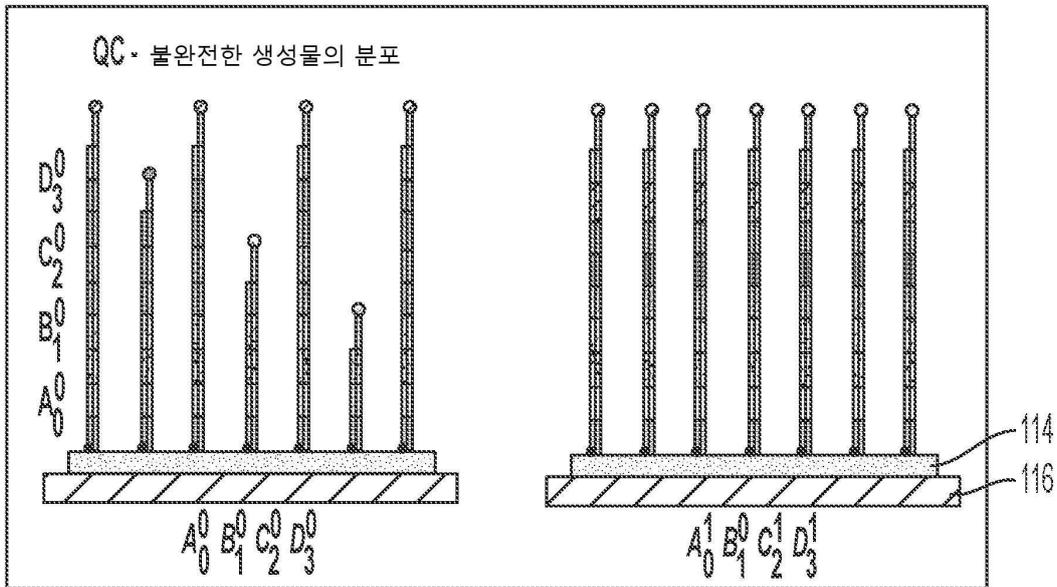
도면28



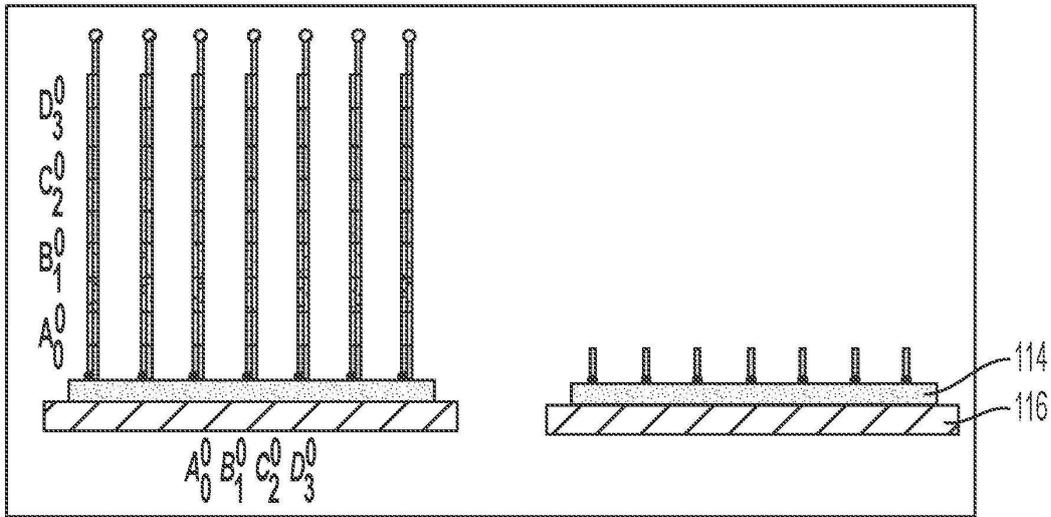
도면29



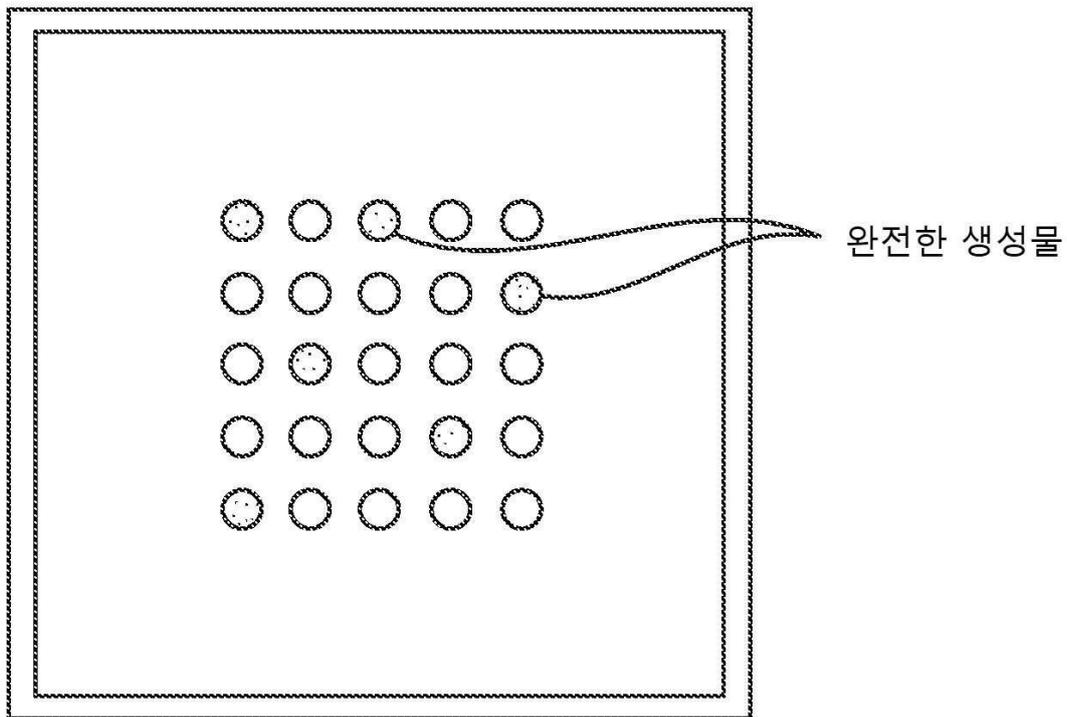
도면30



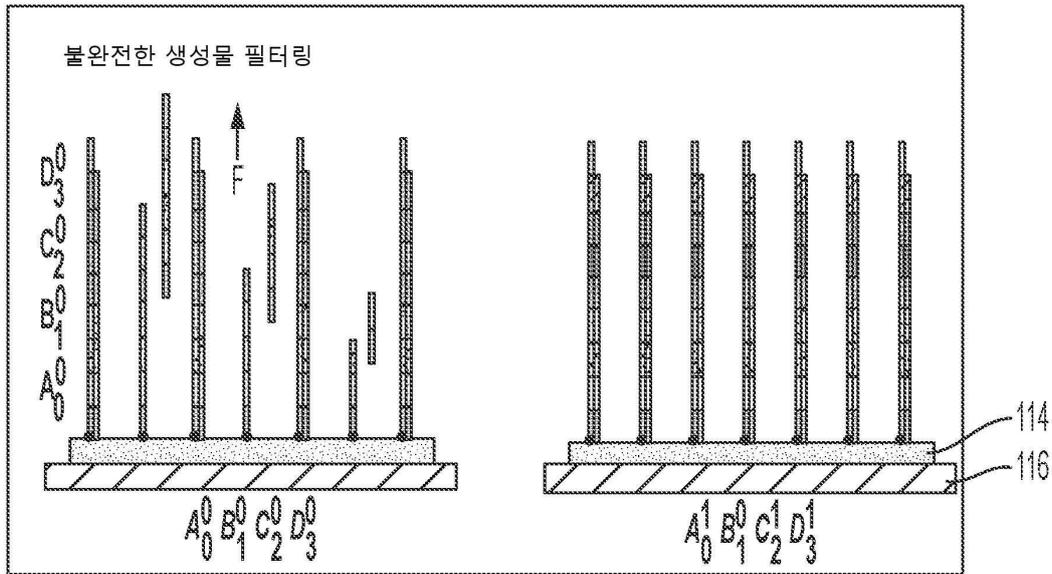
도면31



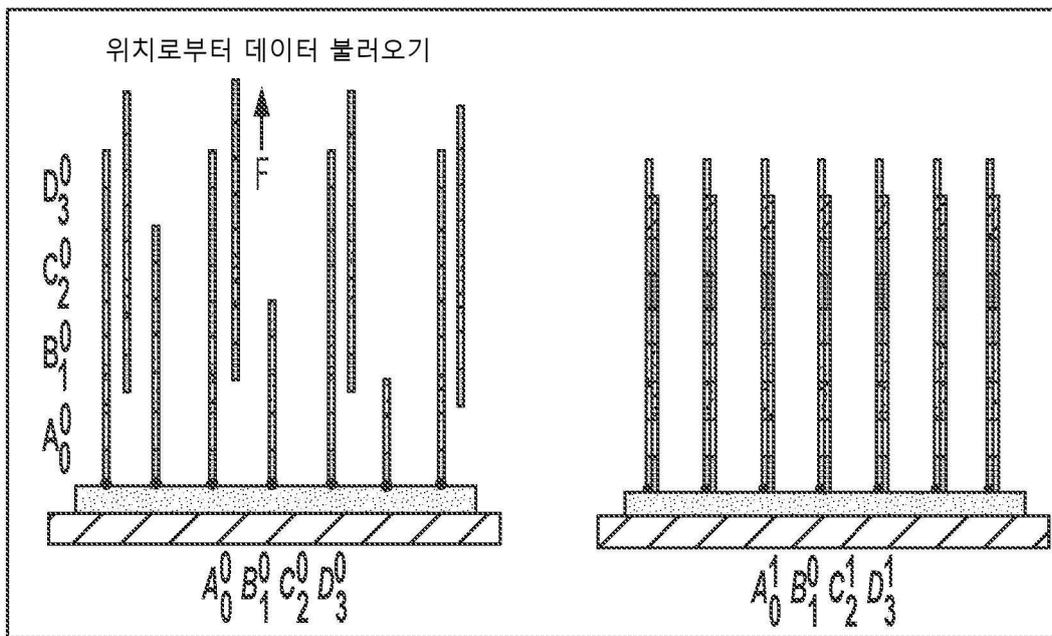
도면32



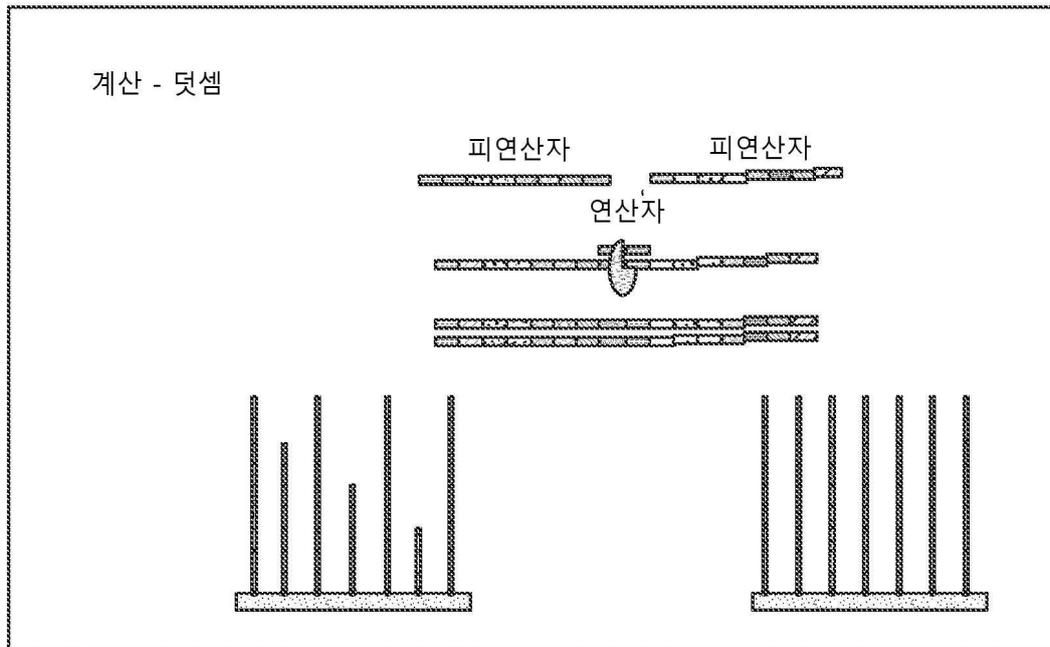
도면33



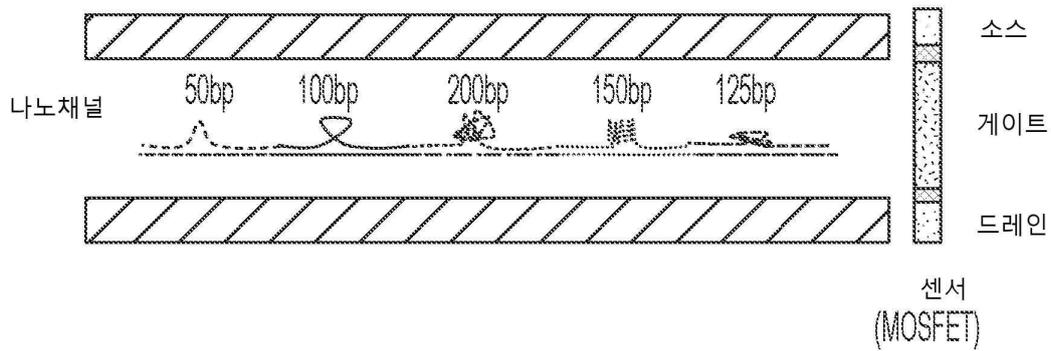
도면34



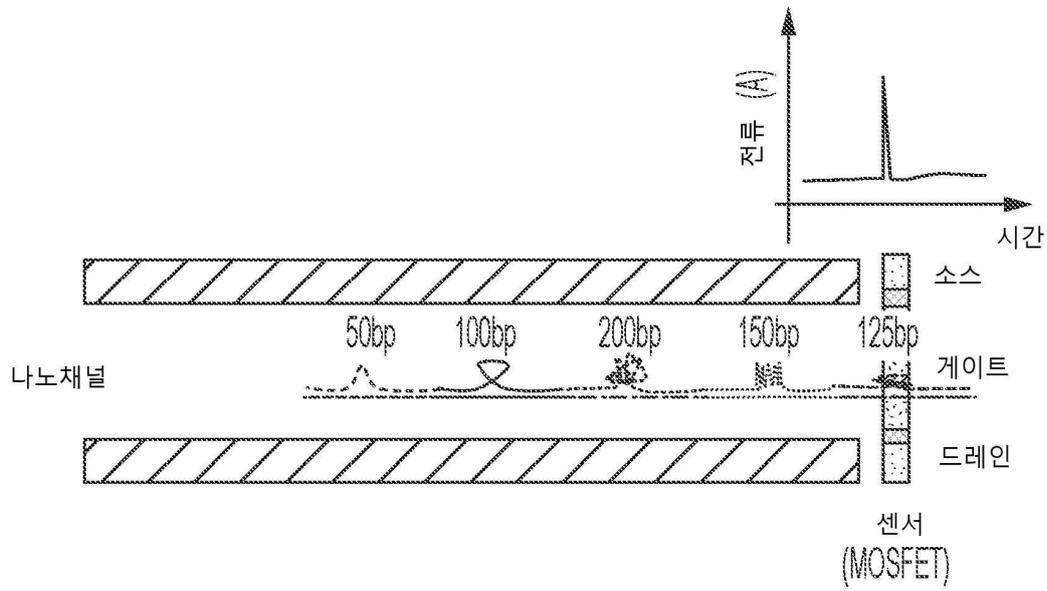
도면35



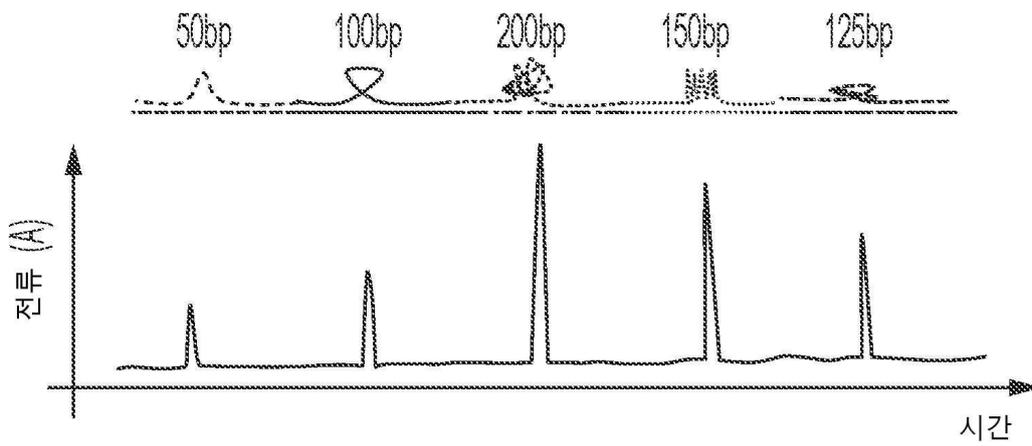
도면36



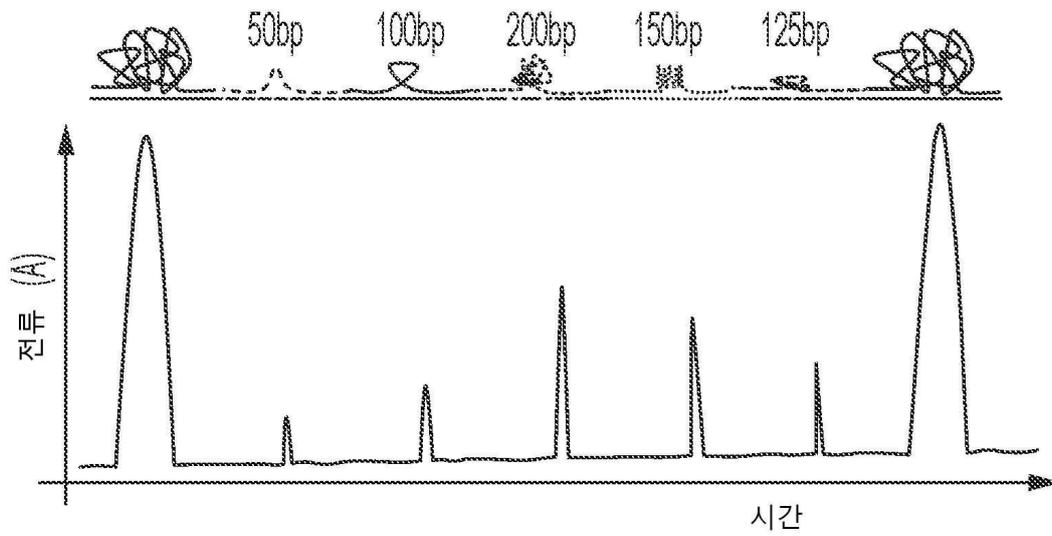
도면37



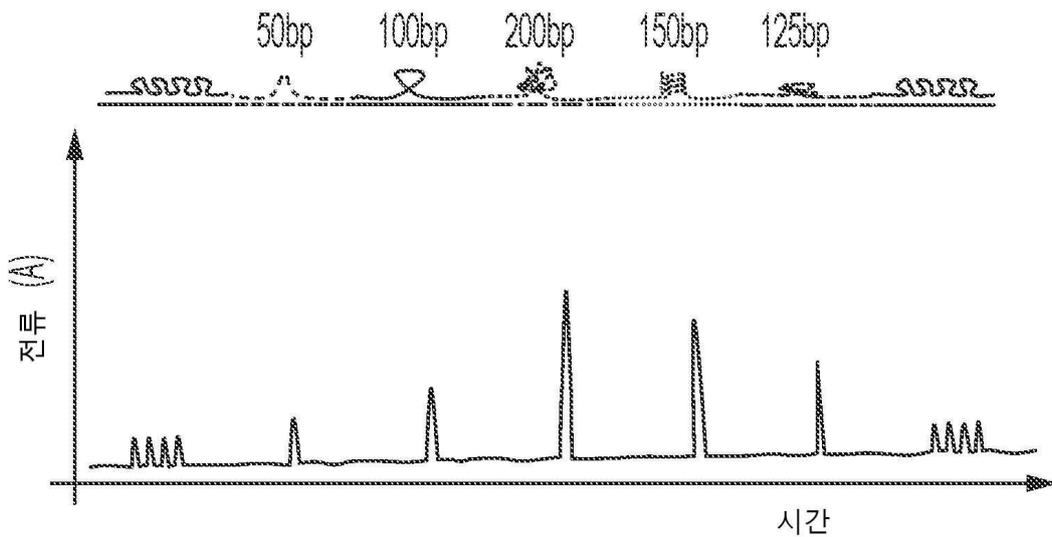
도면38



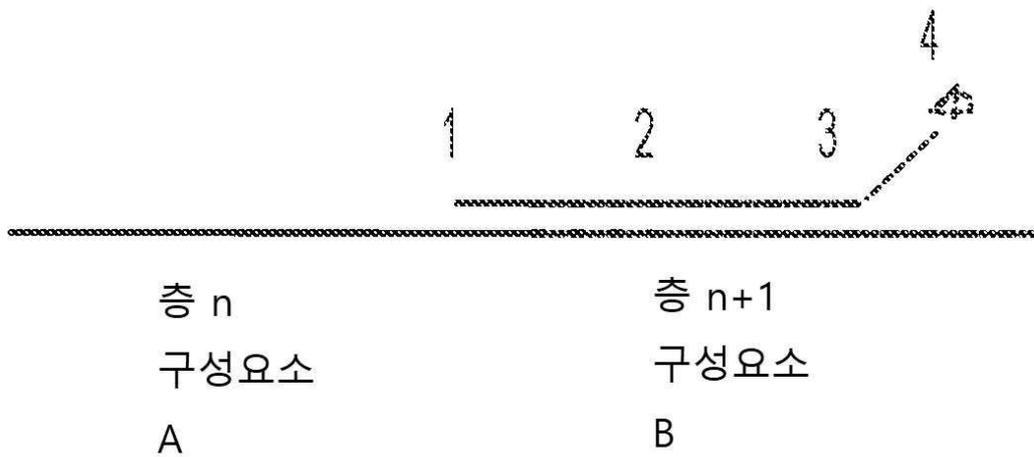
도면39



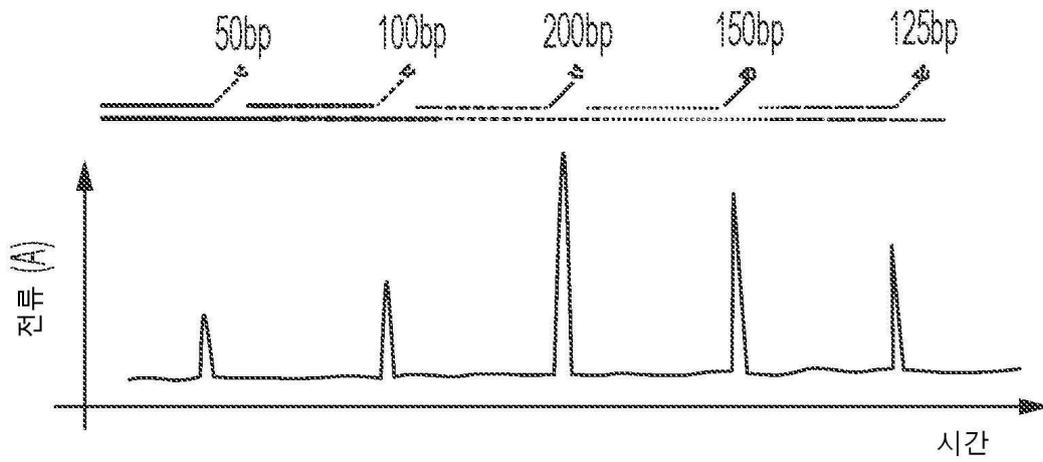
도면40



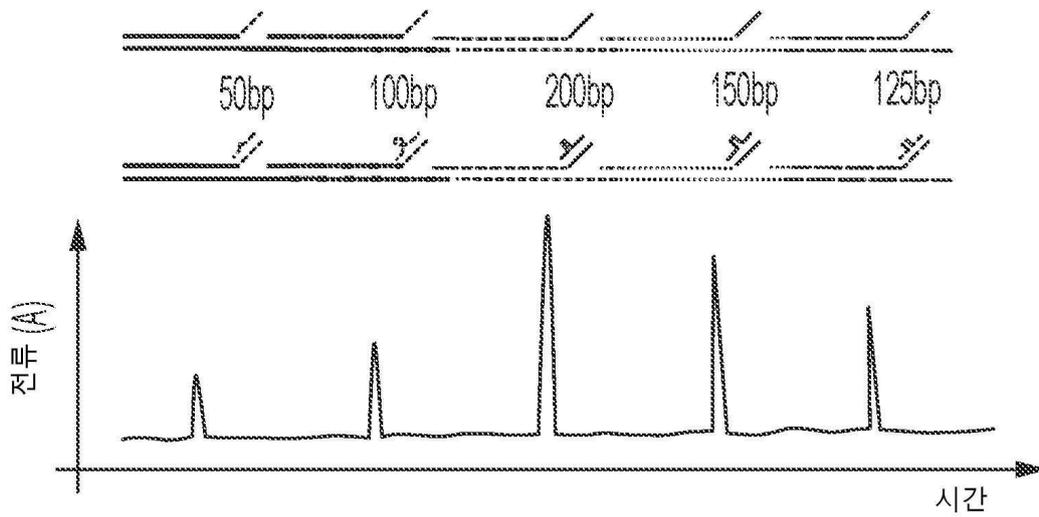
도면41



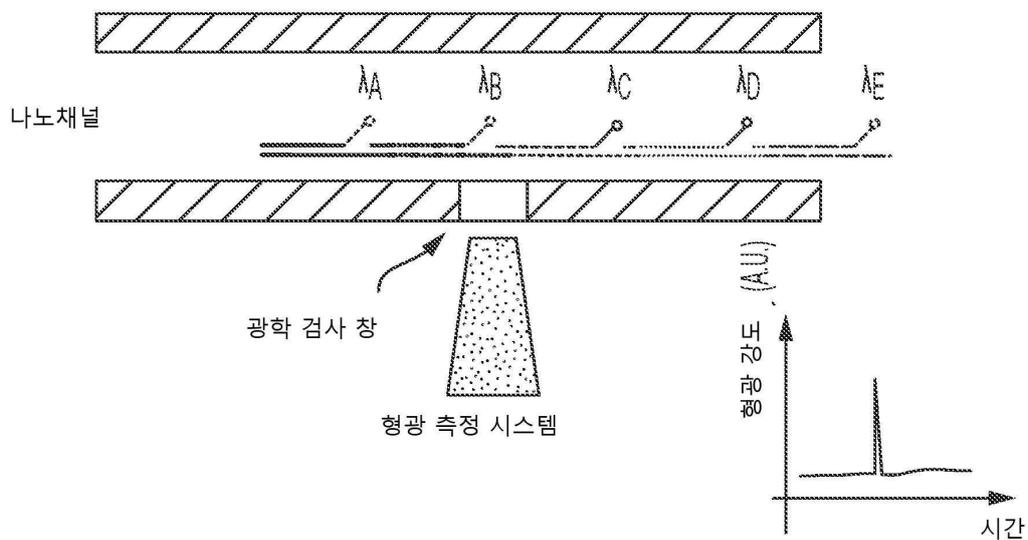
도면42



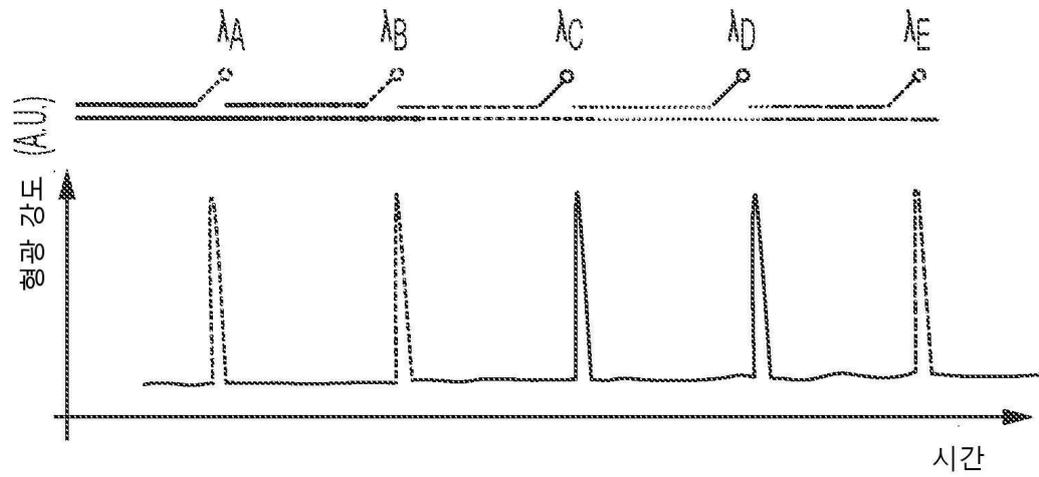
도면43



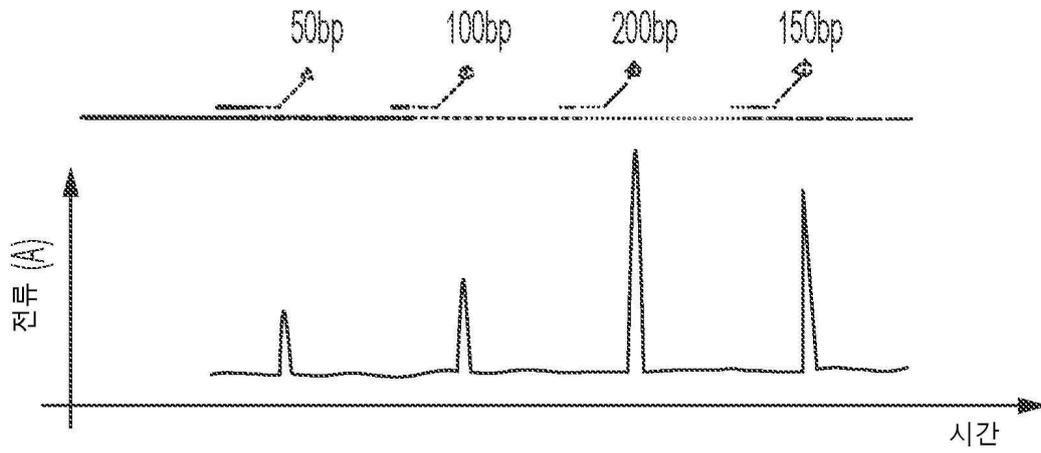
도면44



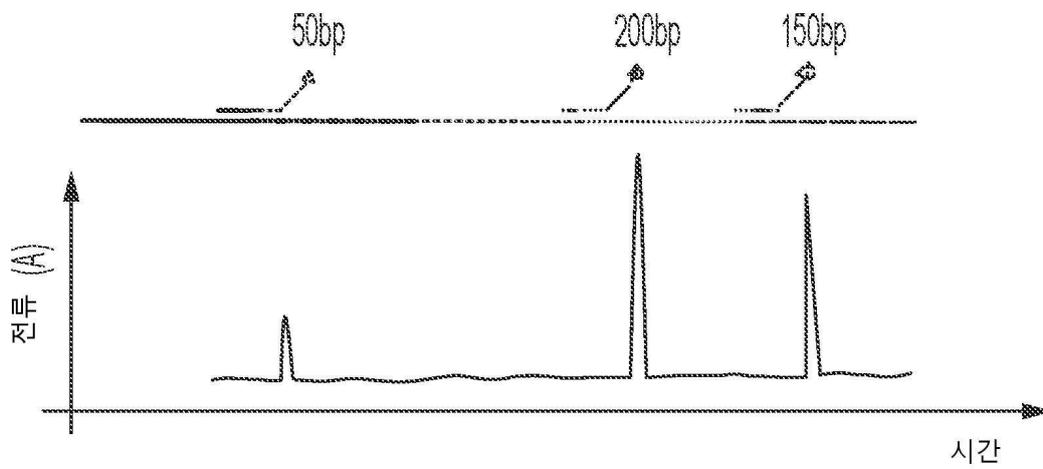
도면45



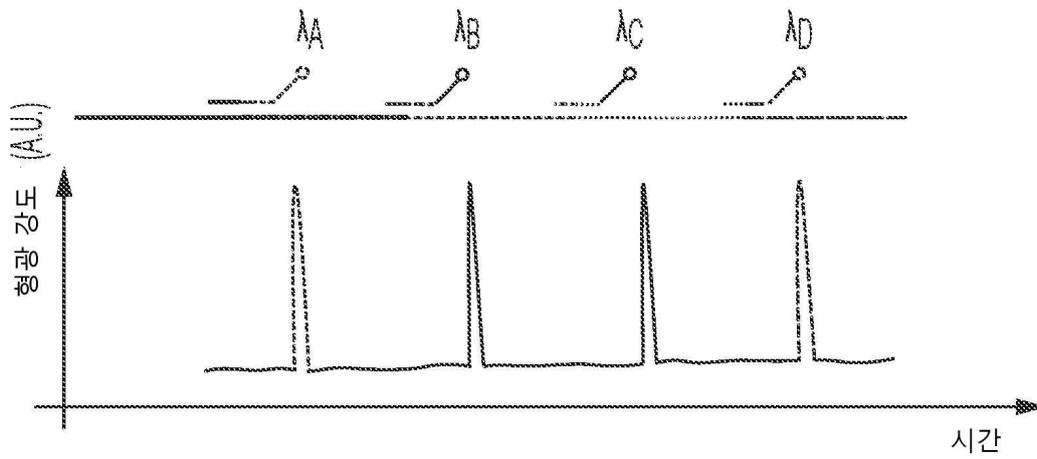
도면46



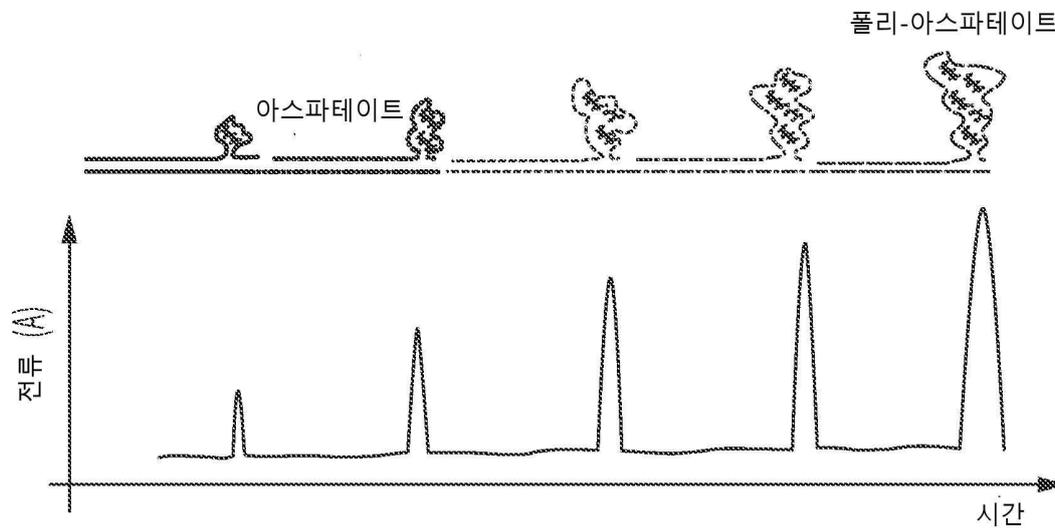
도면47



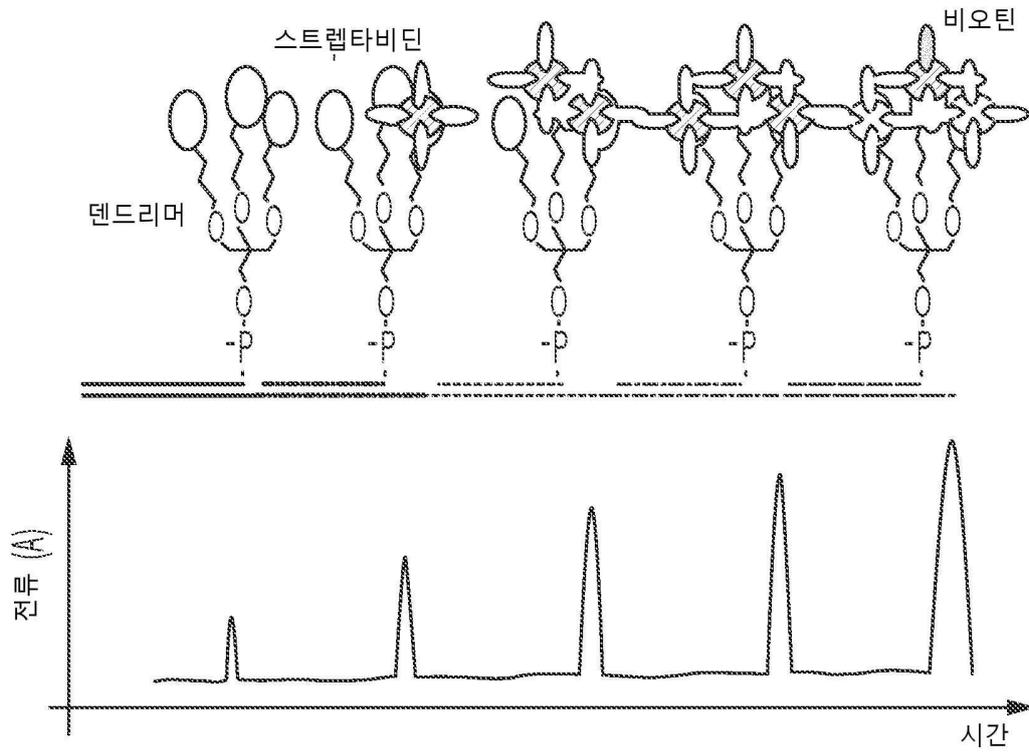
도면48



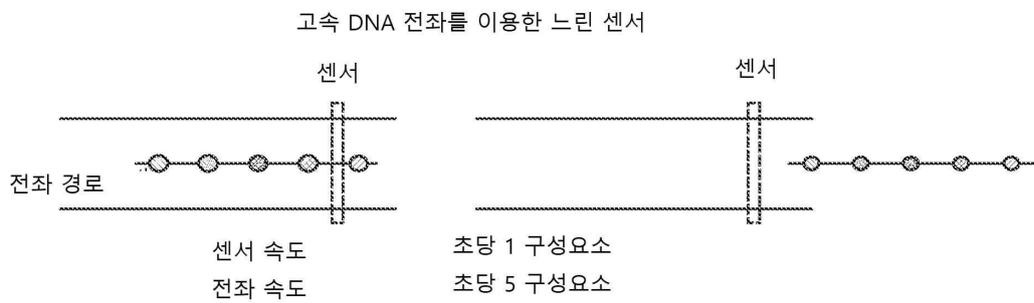
도면49



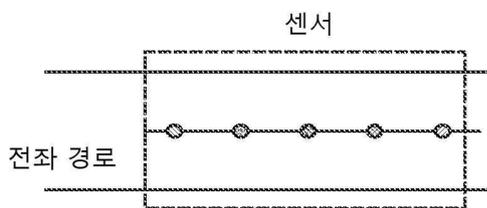
도면50



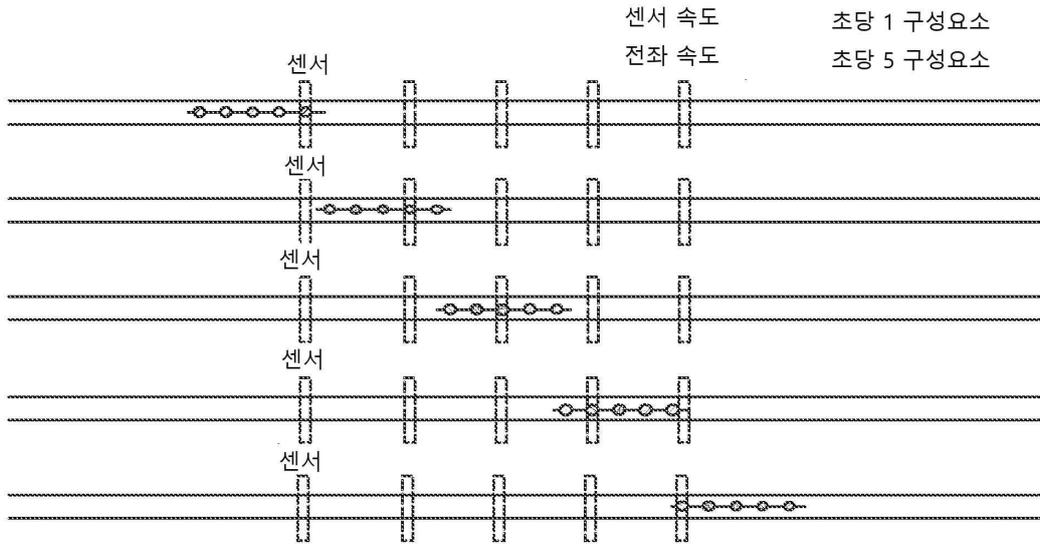
도면51



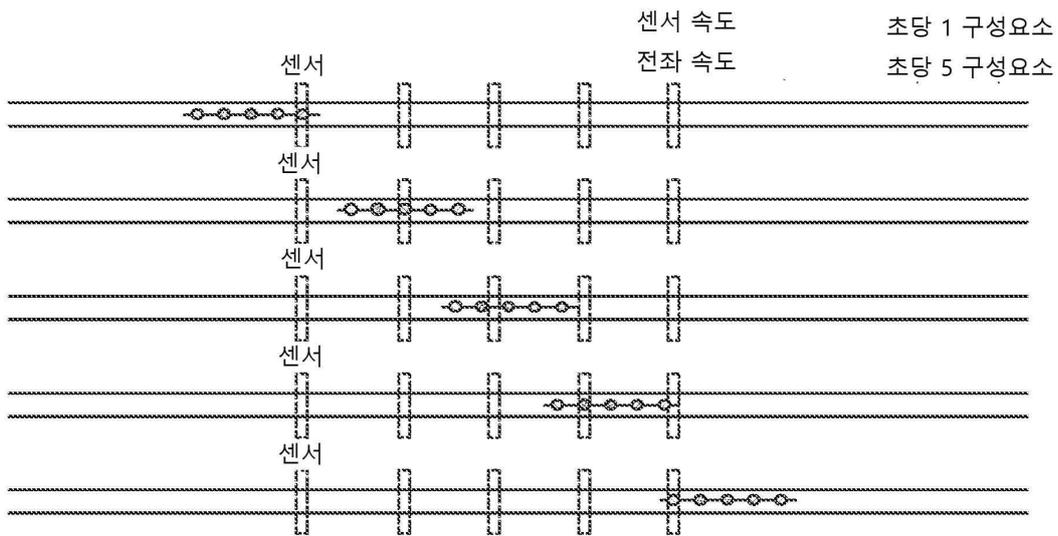
도면52



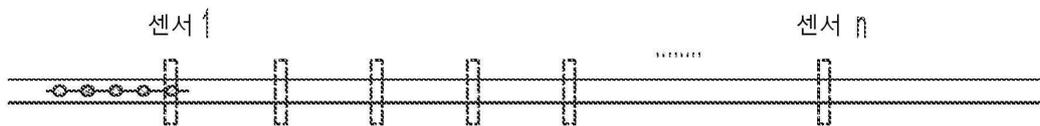
도면53



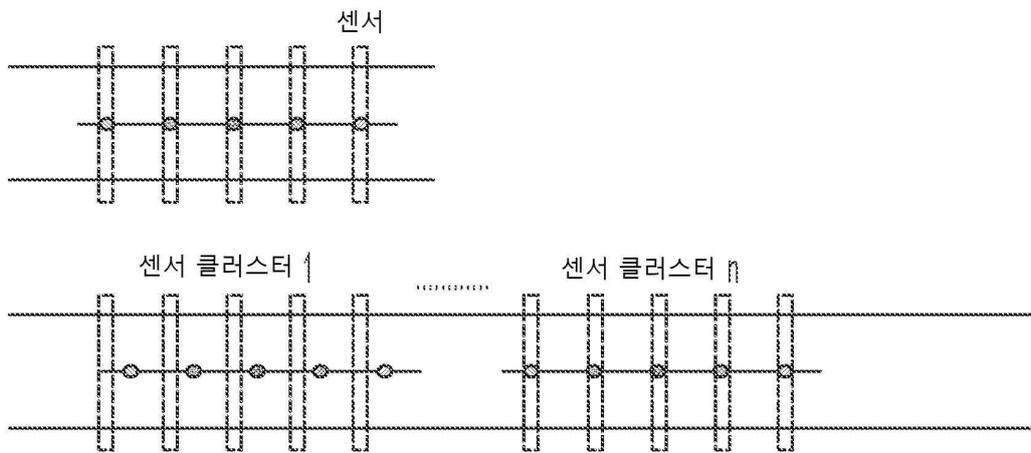
도면54



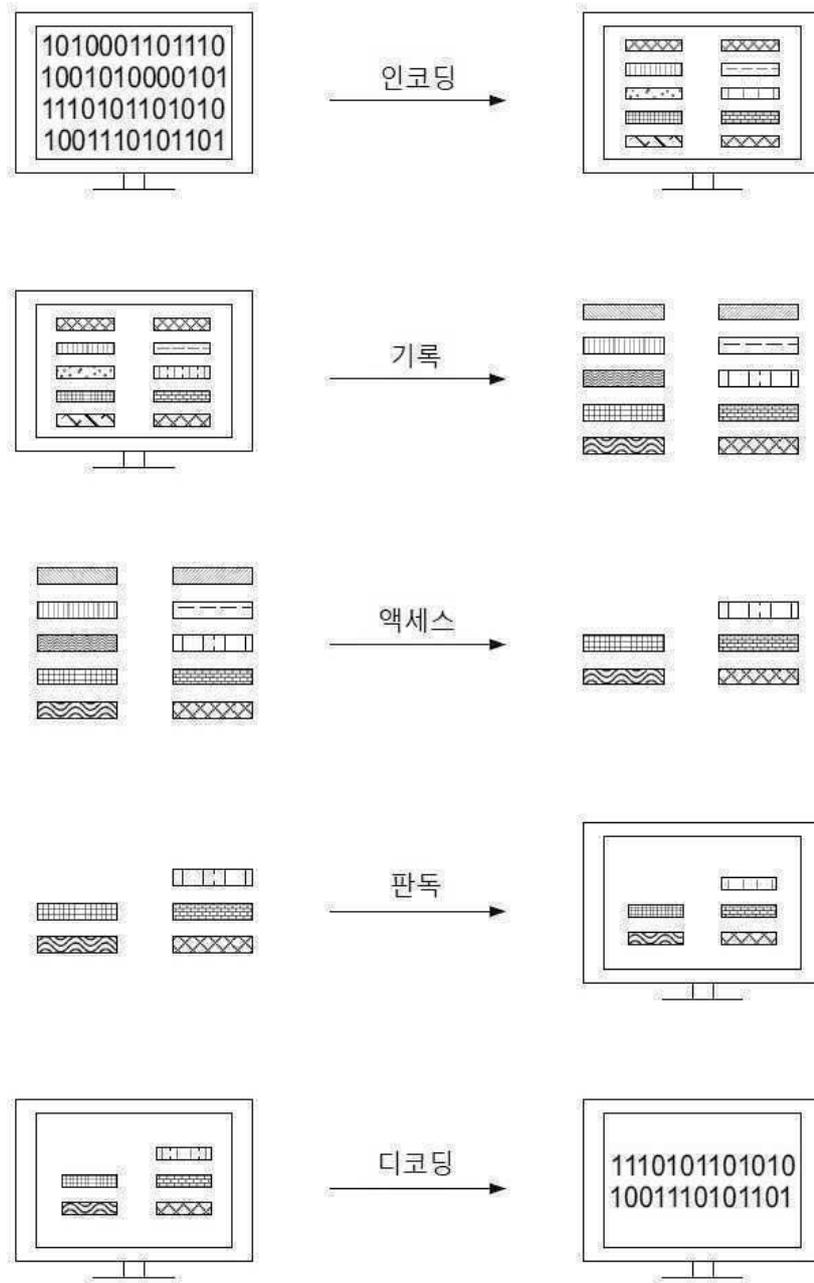
도면55



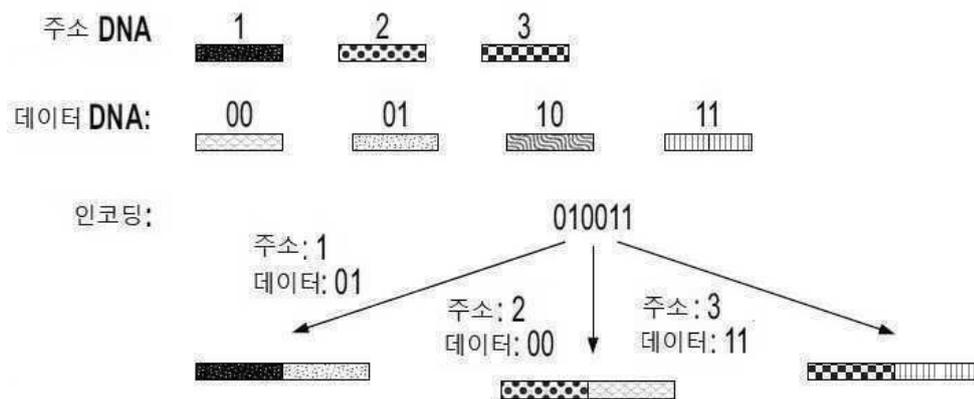
도면56



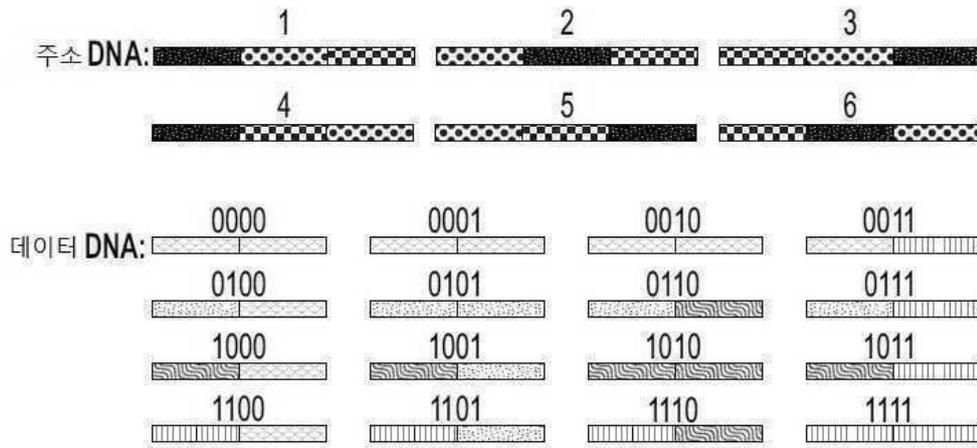
도면57



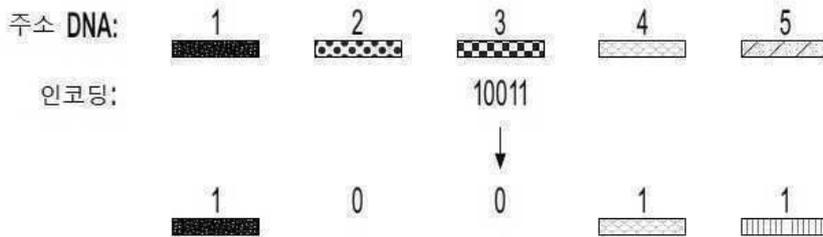
도면58a



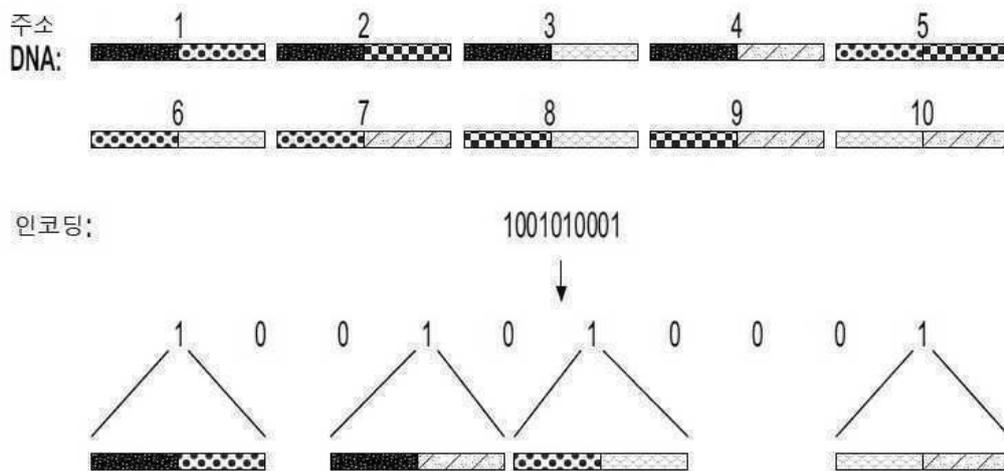
도면58b



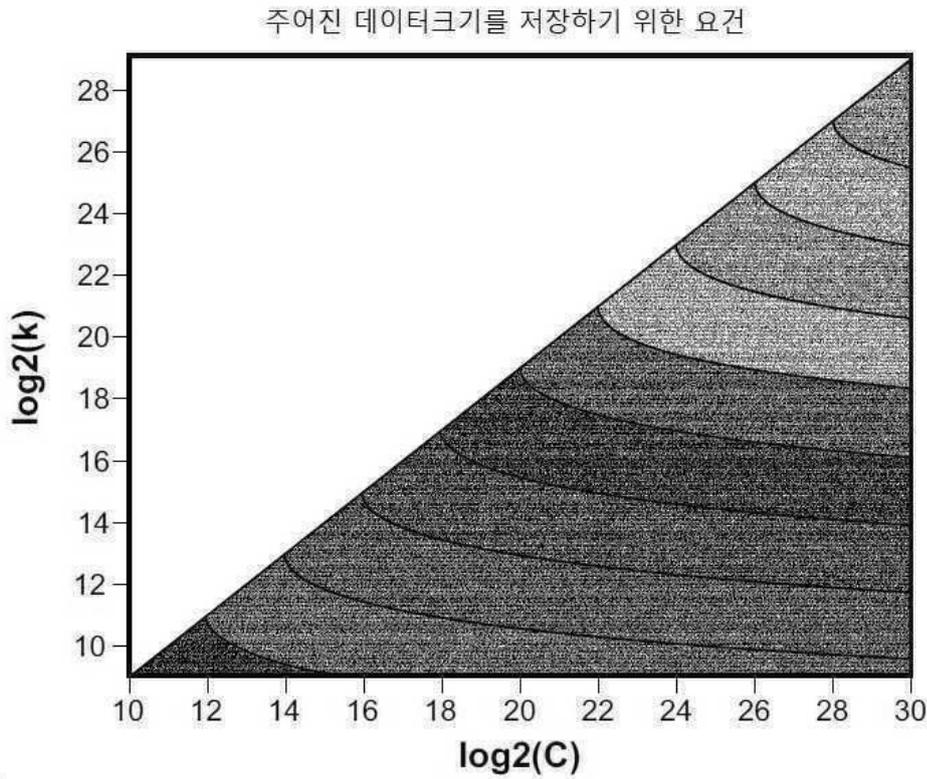
도면59a



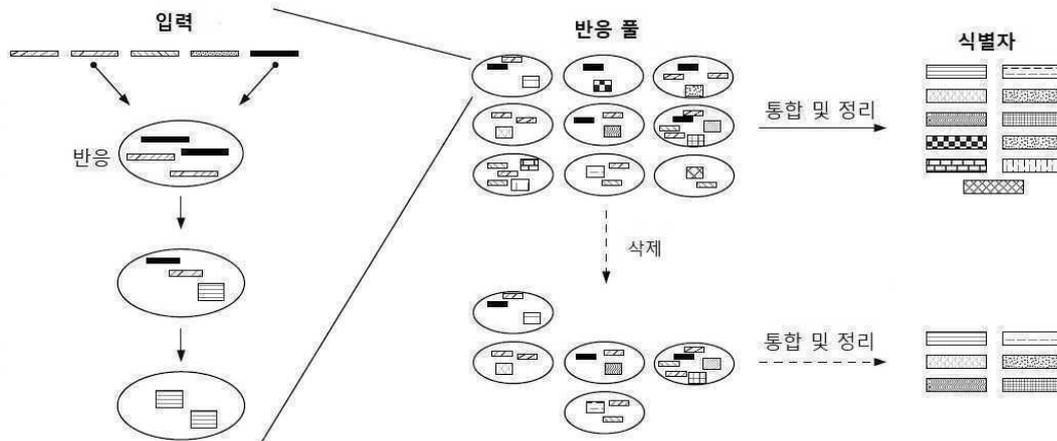
도면59b



도면60

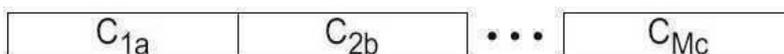


도면61



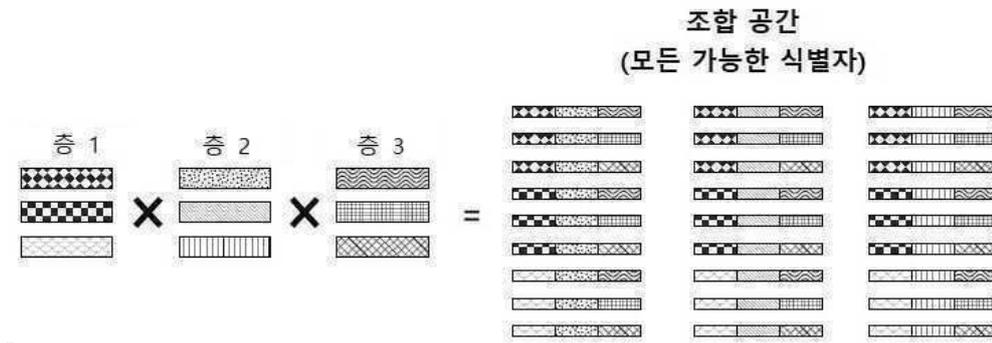
도면62a

$C_{xy}$ 는 층  $x$  내  $y$ 번째 구성요소이다.  $N$ 개의 구성요소를 각각 갖는  $M$ 개의 층을 갖는 시작 라이브러리에 대해, 식별자가 다음의 구조를 가질 것이다:

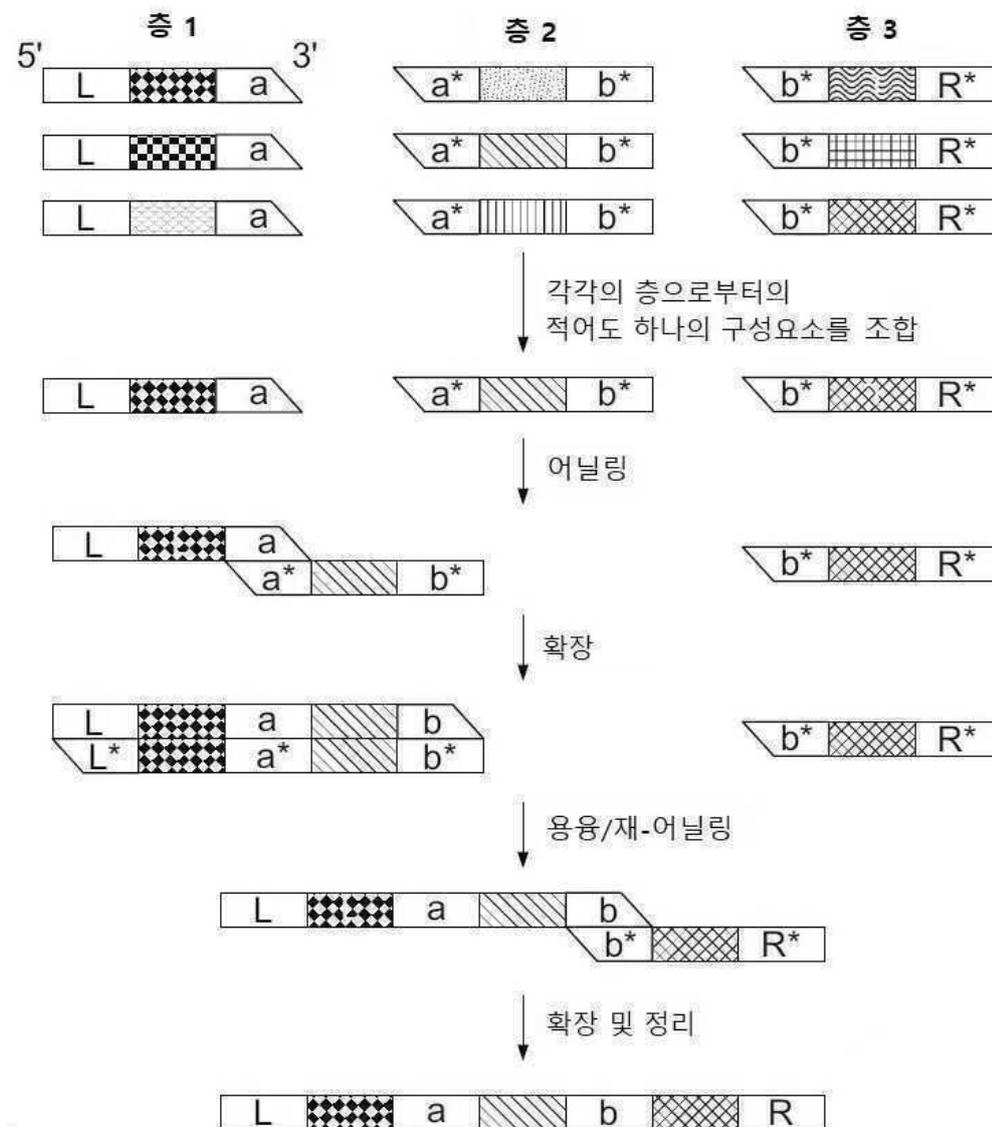


a, b, c는 세트  $\{1, 2, \dots, N\}$  내 원소를 나타냄.

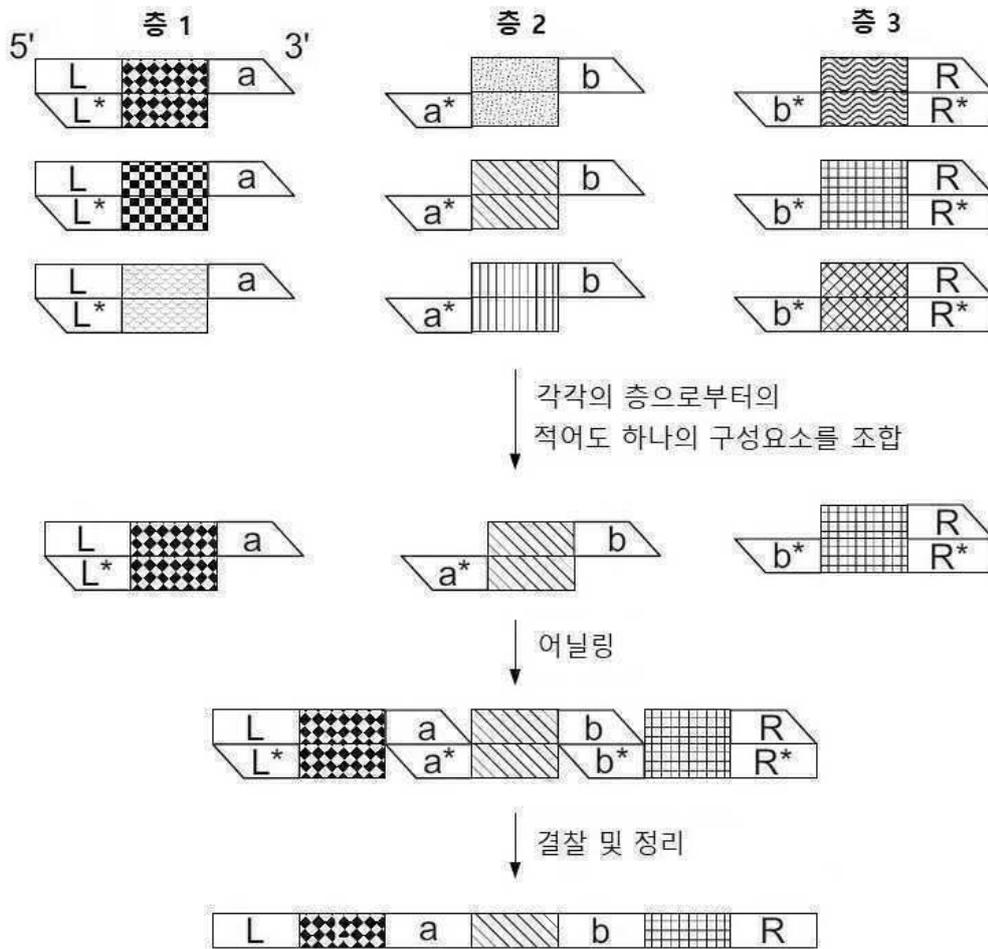
도면62b



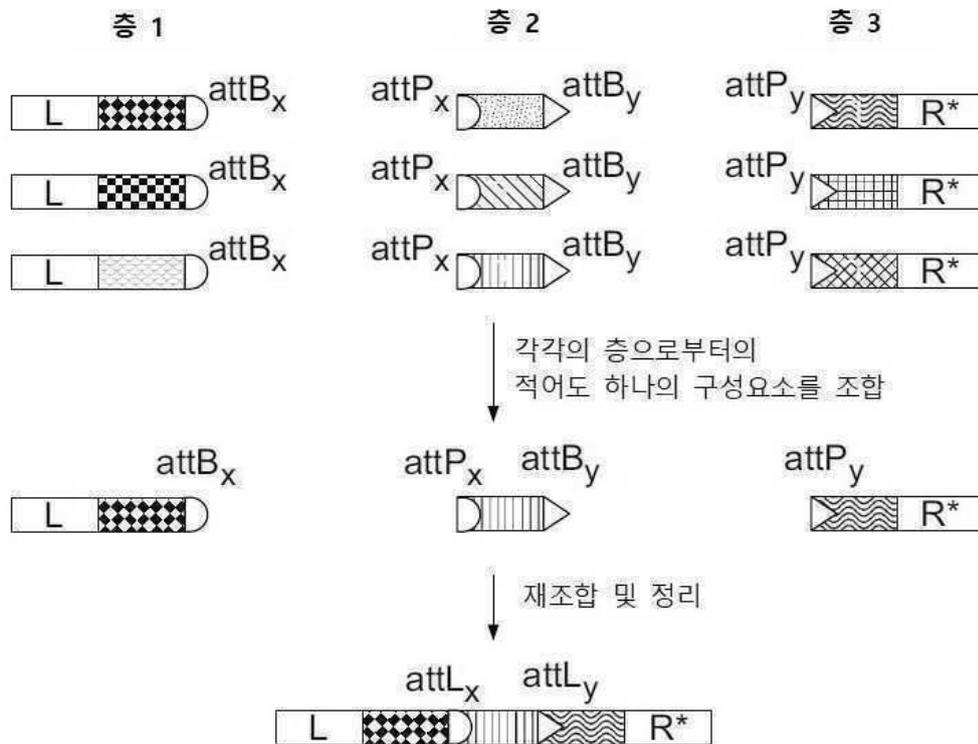
도면63



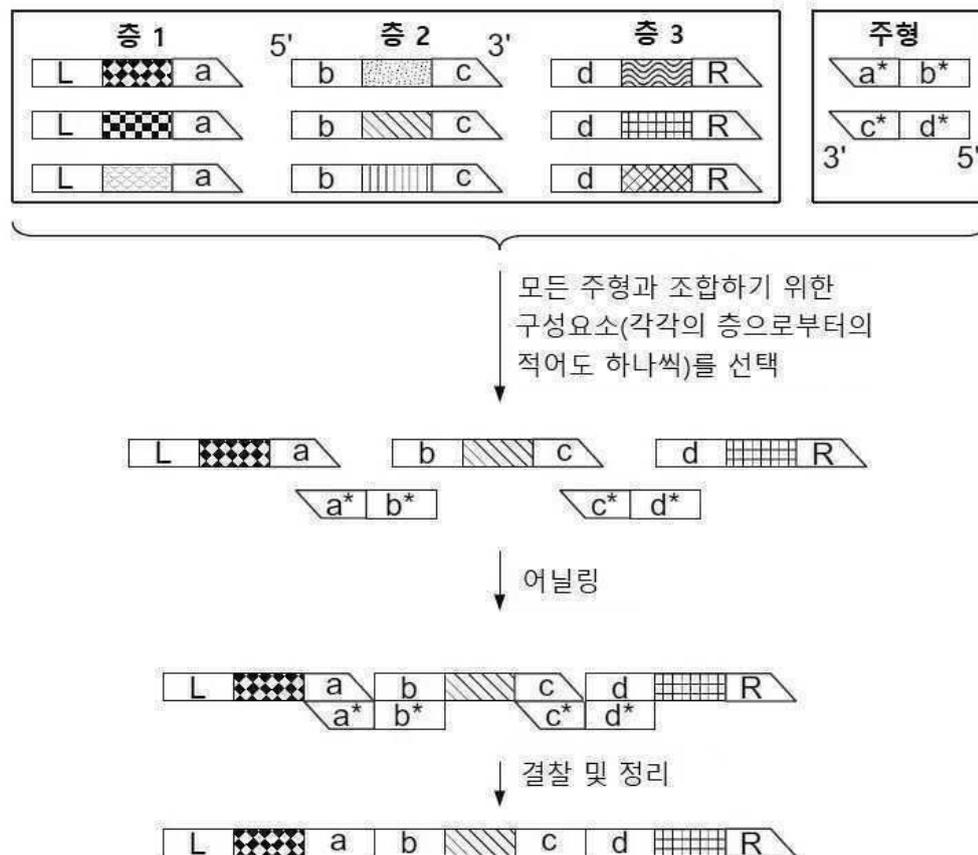
도면64



도면65

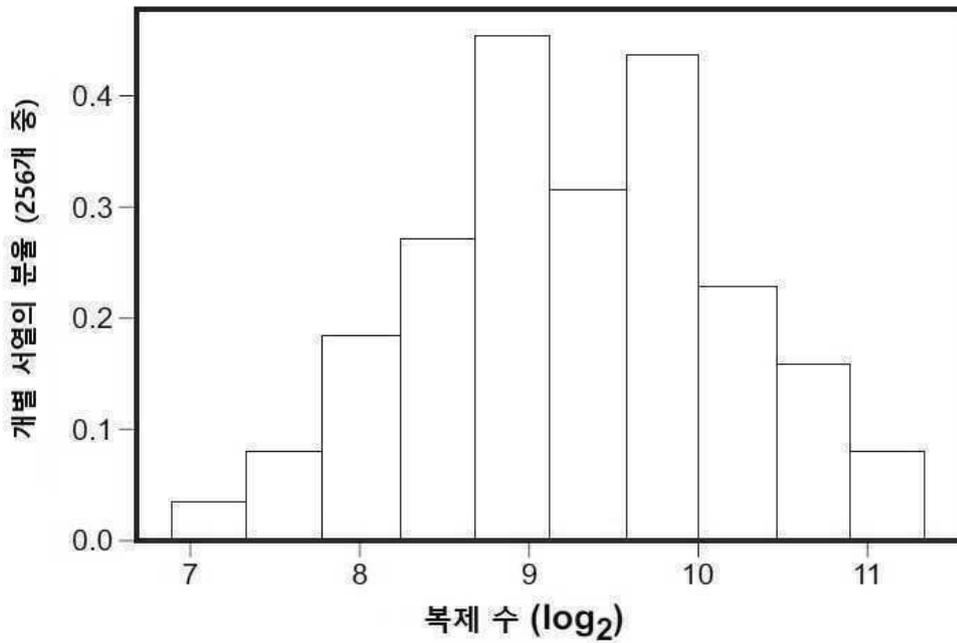


도면66a



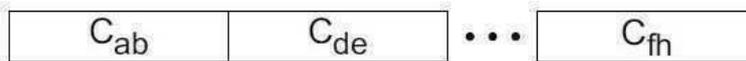
도면66b

6-층 TDL과 조립되는 256개의 개별 DNA 서열의 복제수



도면67a

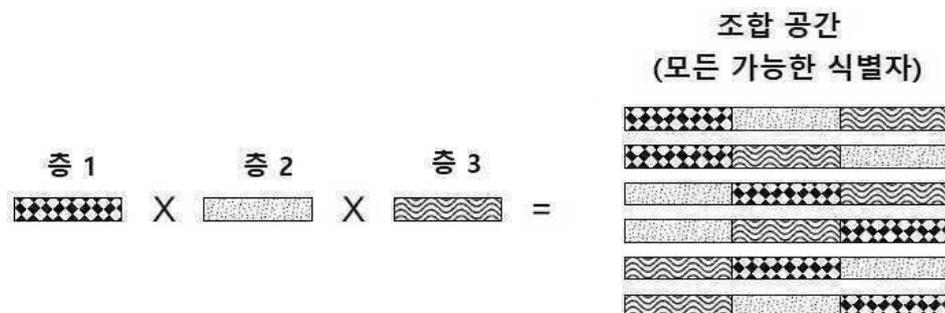
$C_{xy}$ 는 층  $x$  내  $y$ 번째 구성요소이다.  $N$ 개의 구성요소를 각각 갖는  $M$ 개의 층을 갖는 시작 라이브러리에 대해, 식별자가 다음의 구조를 가질 것이다:



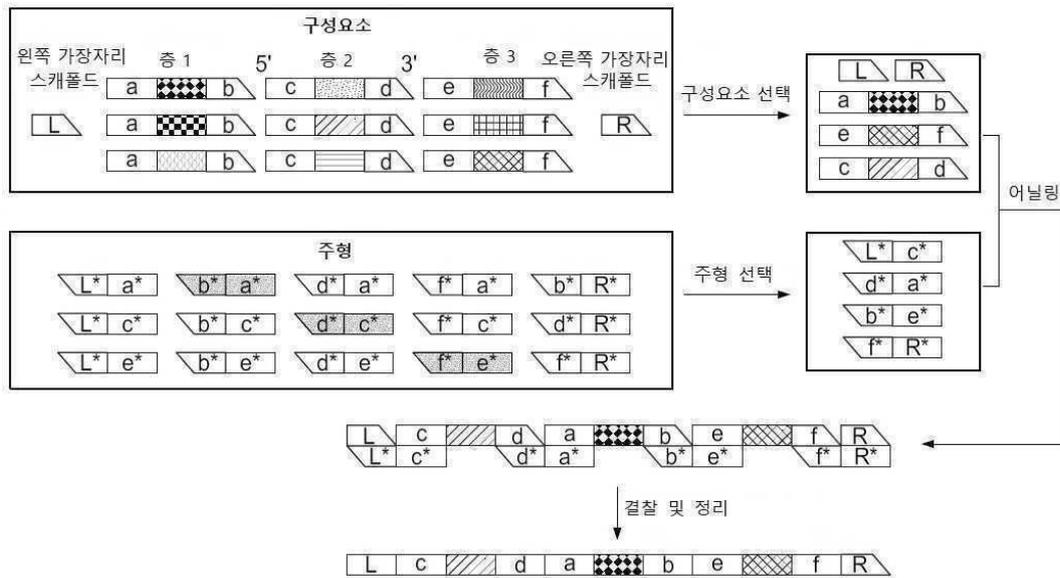
$b, e, h$ 는 세트  $\{1, 2, \dots, N\}$  내 원소를 나타내고,  $a, d, f$ 는 세트  $\{1, 2, \dots, M\}$  내 고유 원소임.

가능한 식별자의 조합 공간은  $N^M M!$ 임.

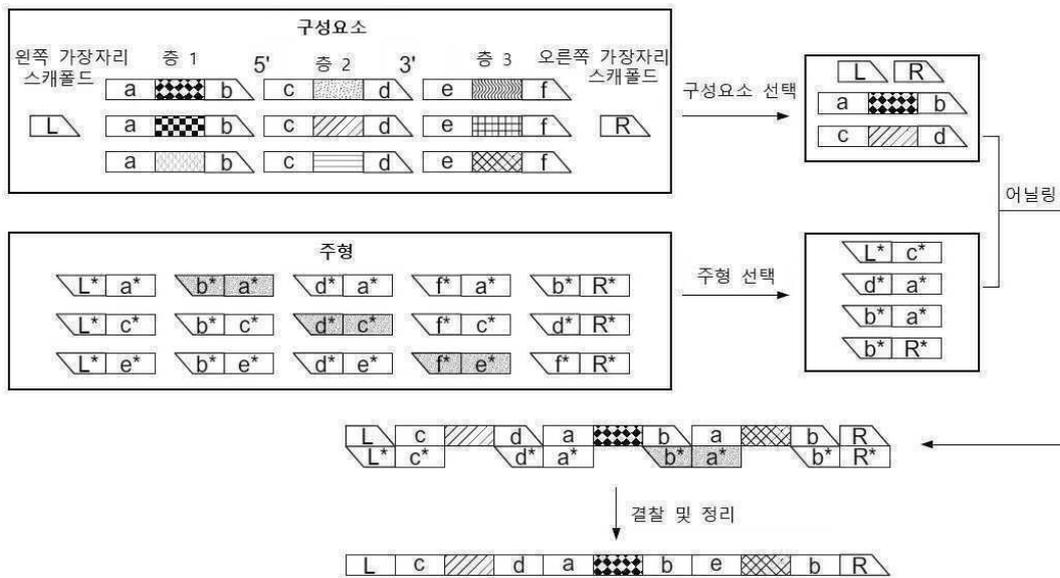
도면67b



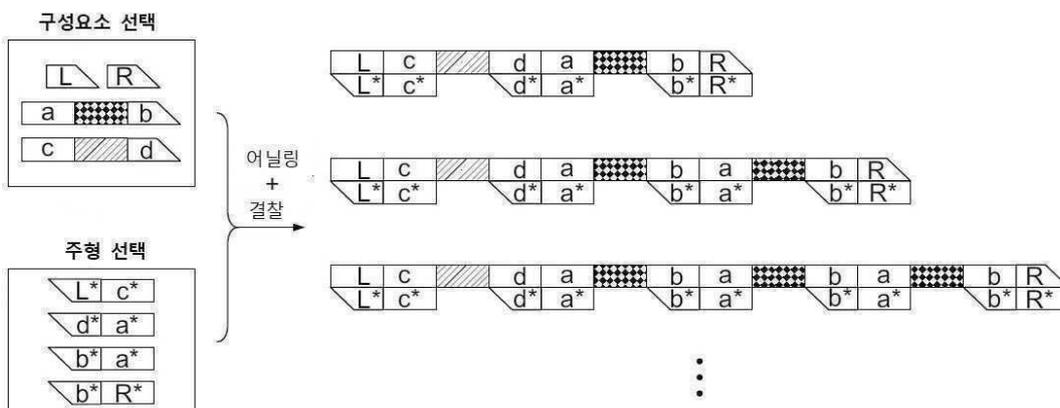
도면67c



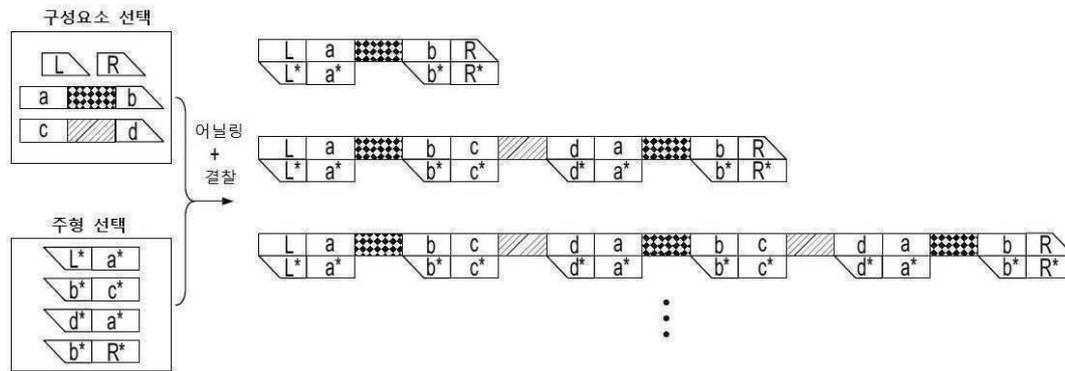
도면67d



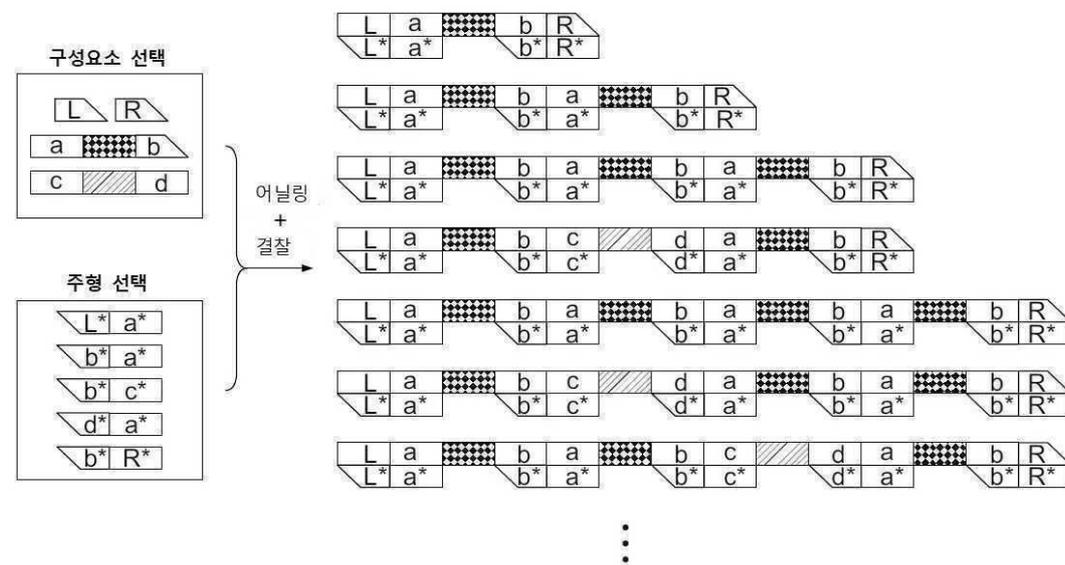
도면67e



도면67f



도면67g



도면68a

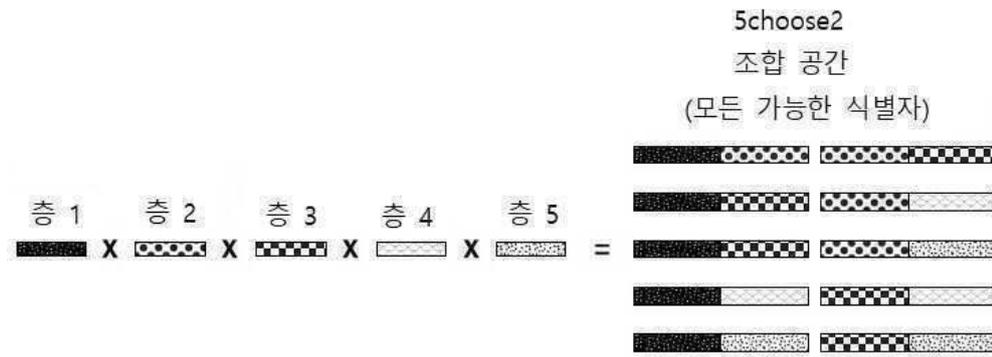
$C_{xy}$ 는 층  $x$  내  $y$ 번째 구성요소이다.  $N$ 개의 구성요소를 각각 갖는  $M$ 개의 층을 갖는 시작 라이브러리에 대해, 식별자가 고정된 수  $k$ 개의 구성요소를 조합하며 다음의 구조를 가질 것이다:



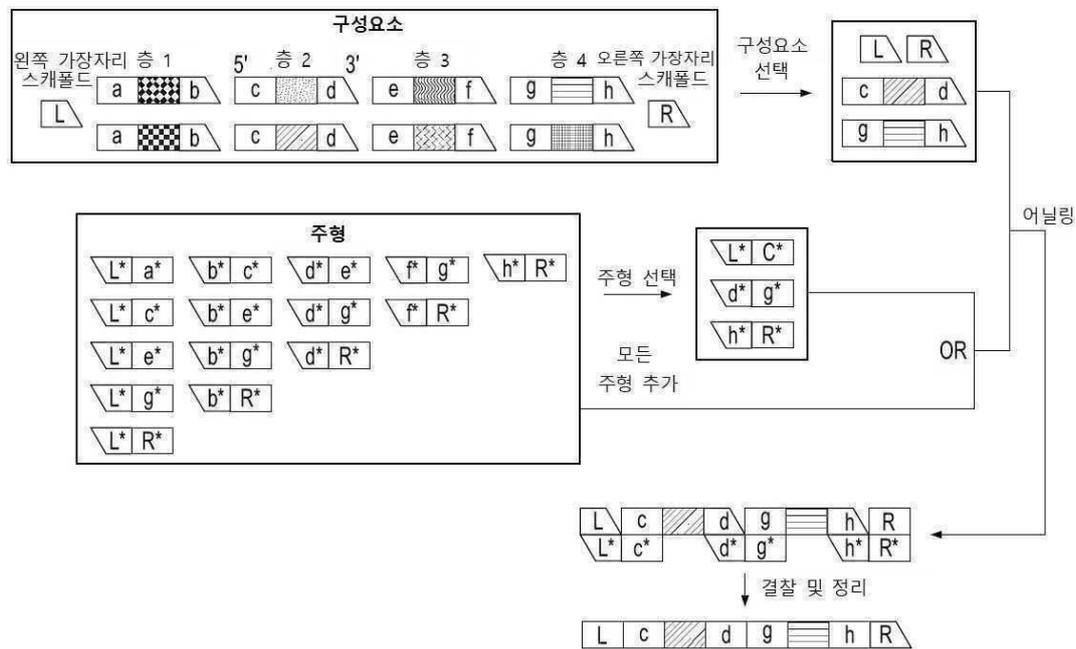
$b, e, h$ 는 세트  $\{1, 2, \dots, N\}$  내 원소를 나타내고, 요소  $a, d, f$ 는 세트  $\{1, 2, \dots, M\}$  내  $k$ 개의 고유한, 순위-정렬된 원소임.

가능한 식별자의 조합 공간은  $N^k M \text{choose } k$ 임.

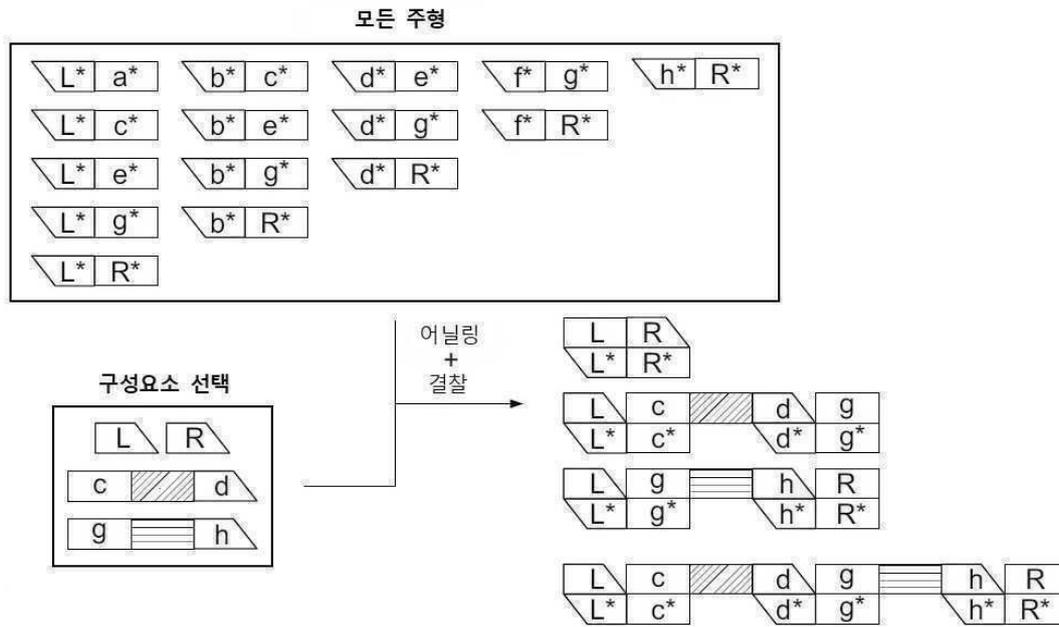
도면68b



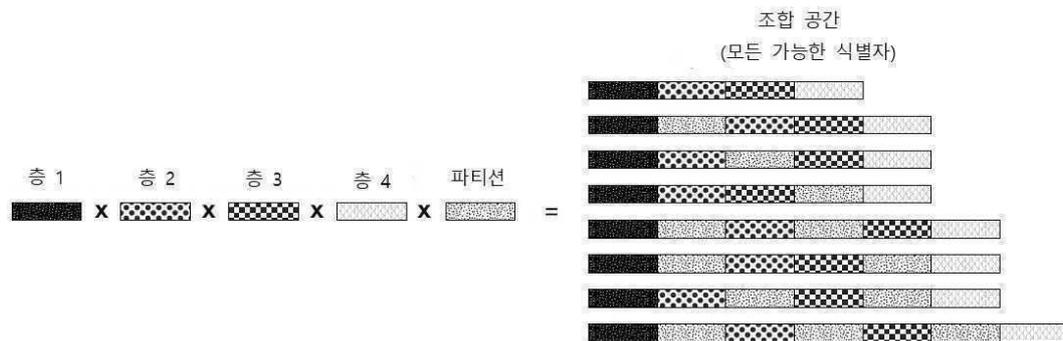
도면68c



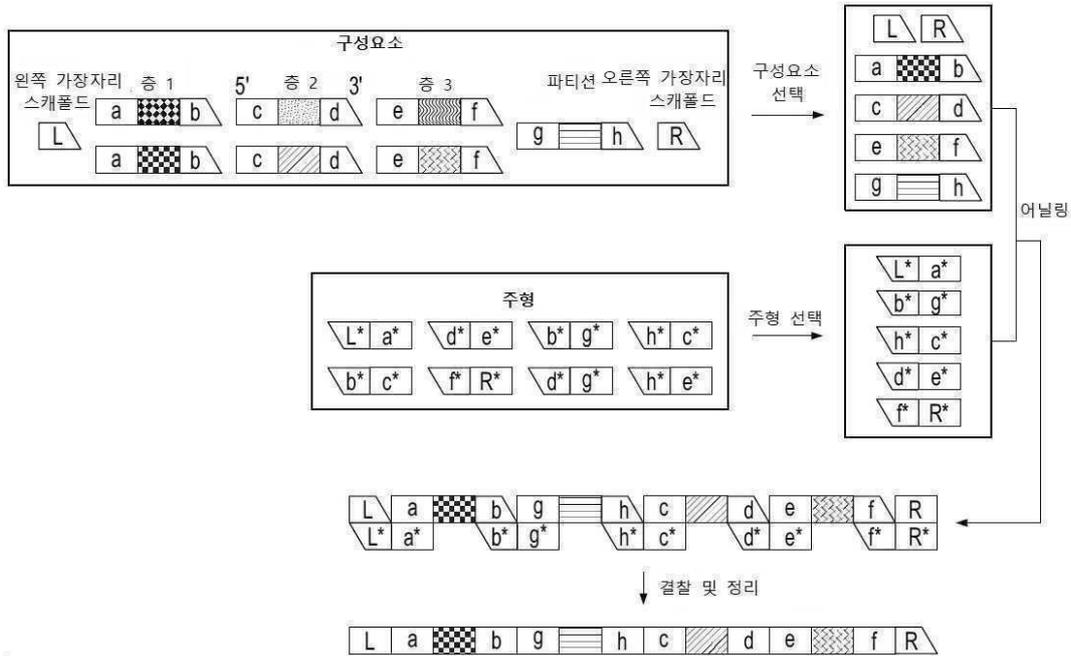
도면68d



도면69a

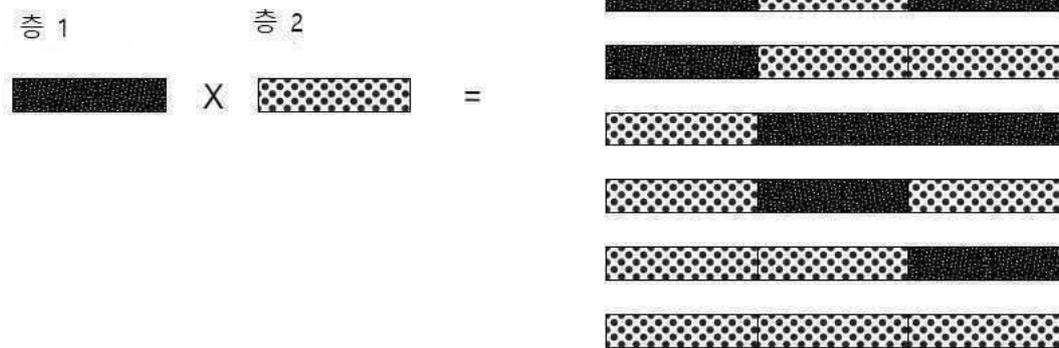


도면69b

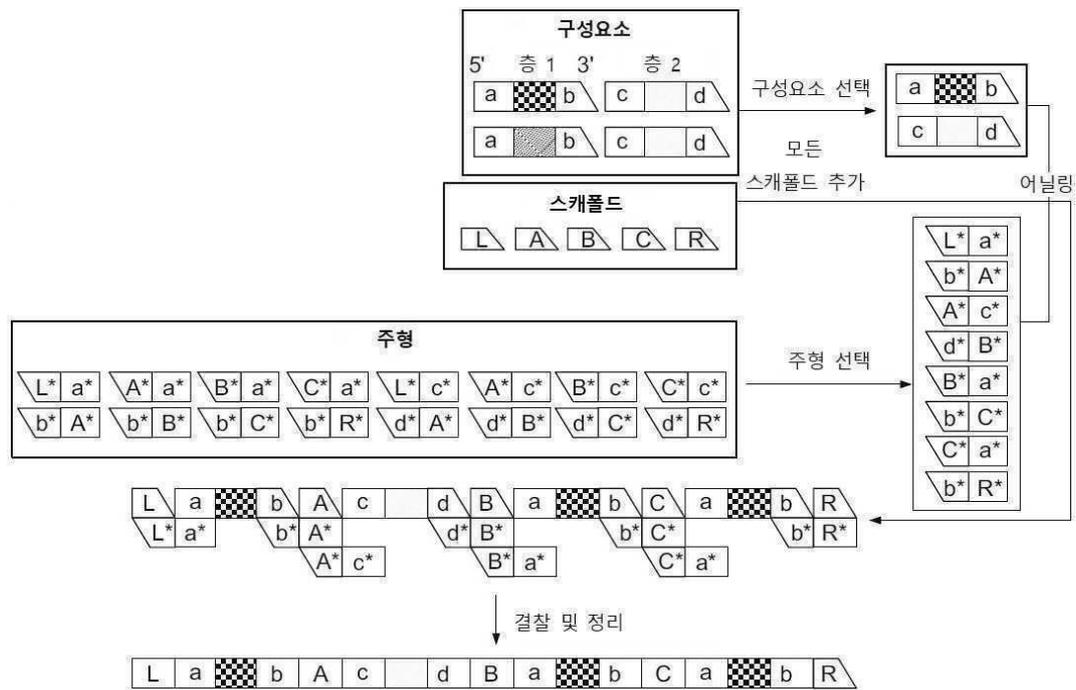


도면70a

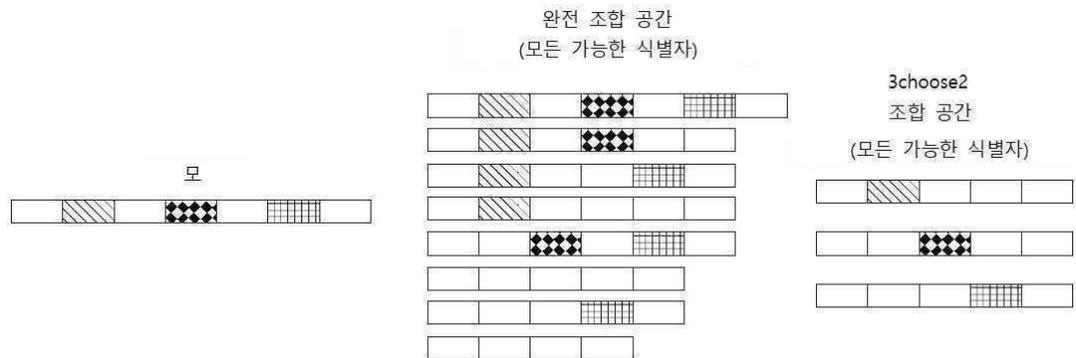
3-구성요소 길이  
조합 공간  
(모든 가능한 식별자)



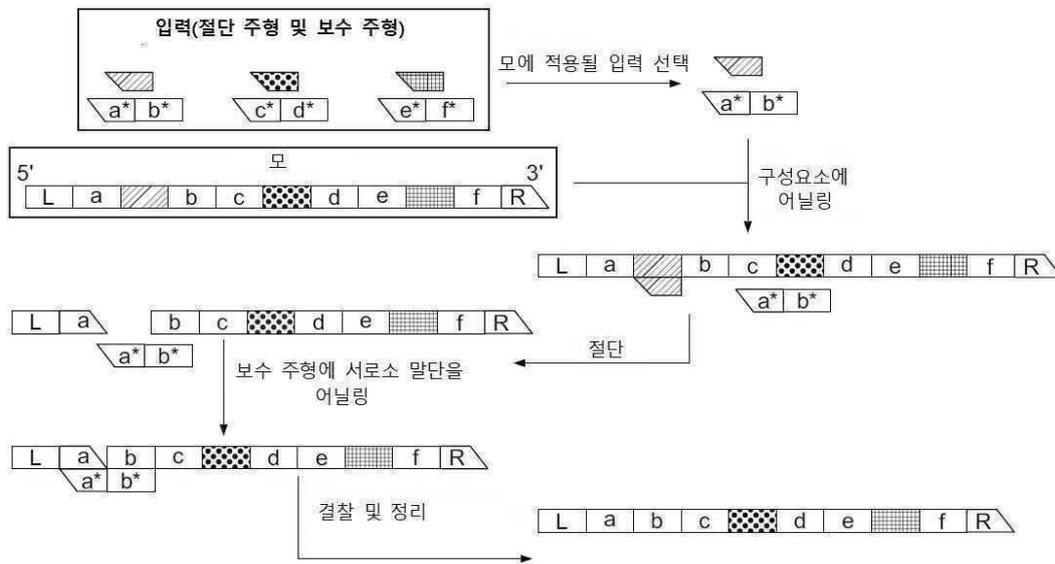
도면70b



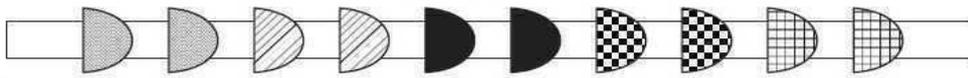
도면71a



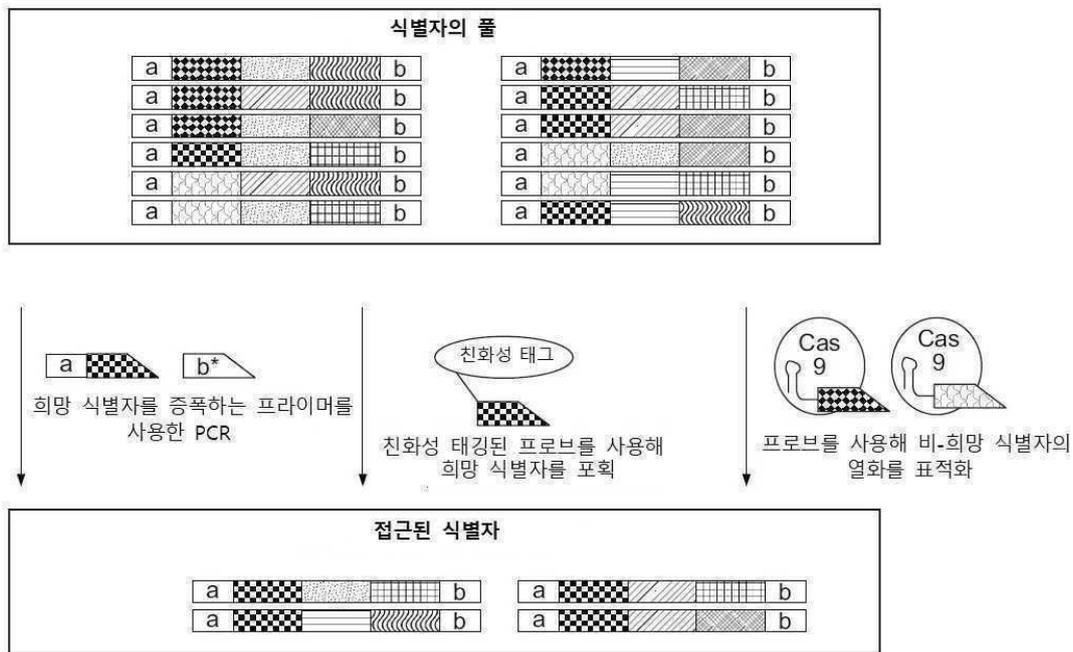
도면71b



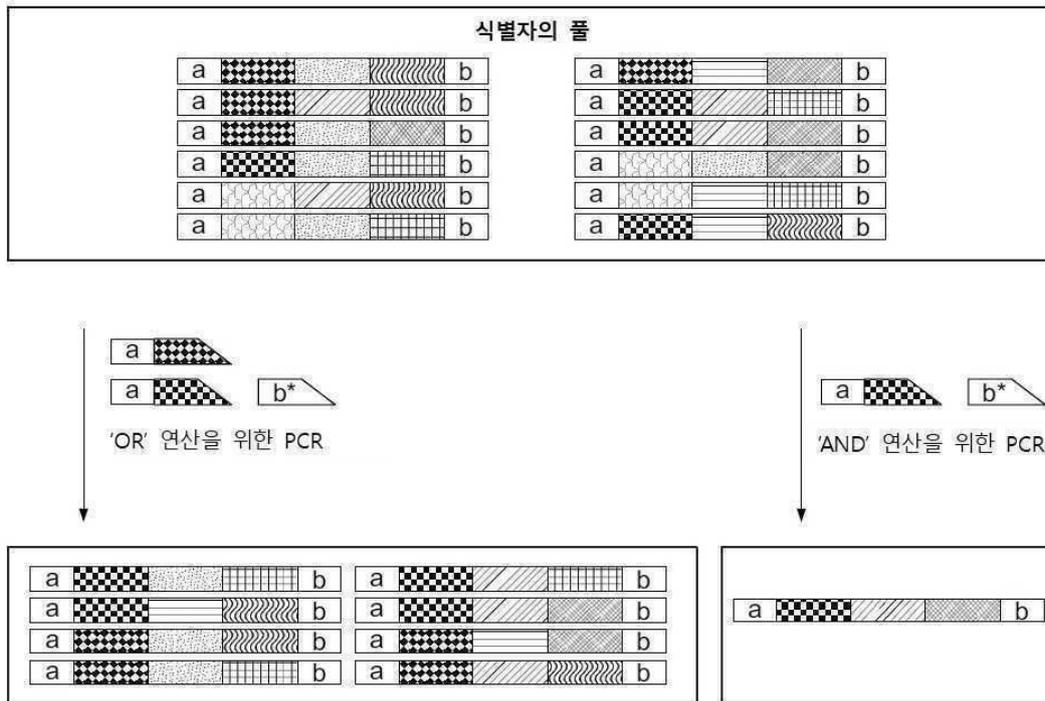
도면72



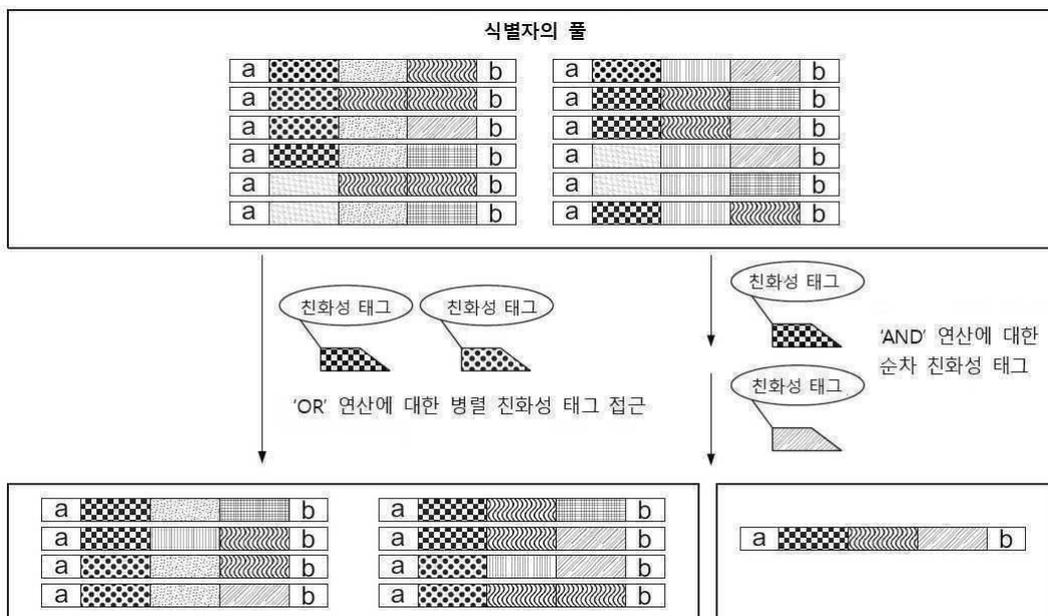
도면73a



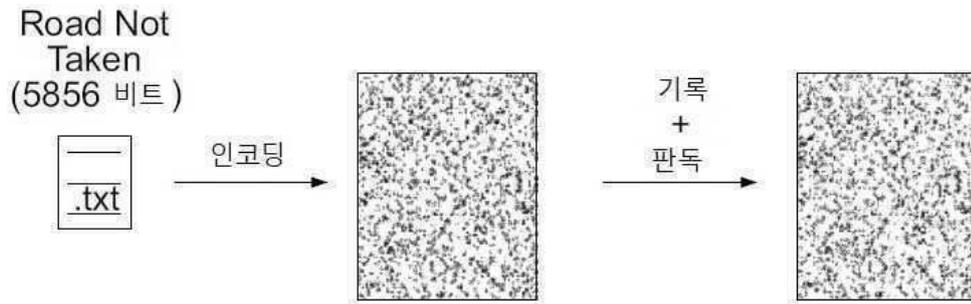
도면73b



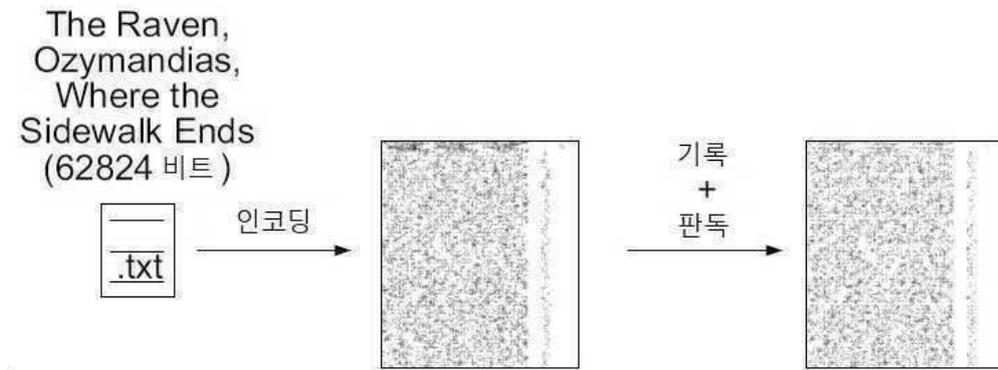
도면73c



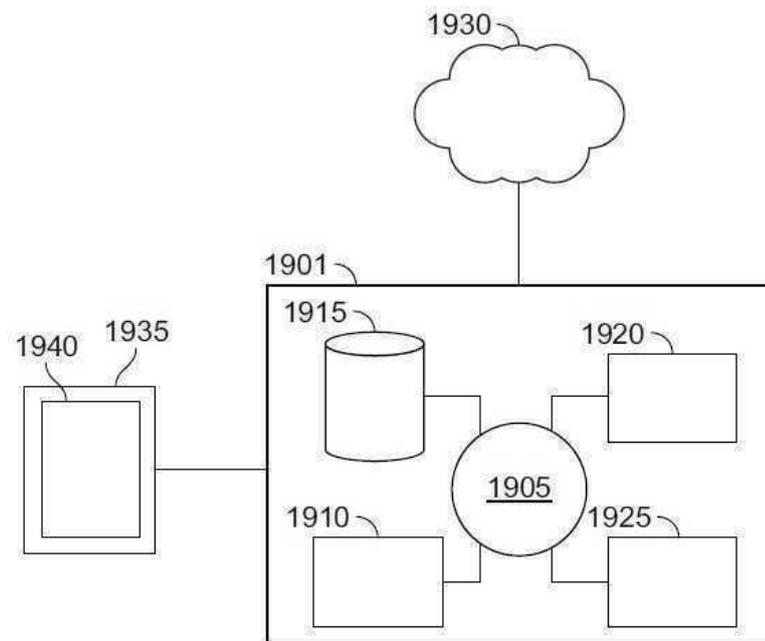
도면74a



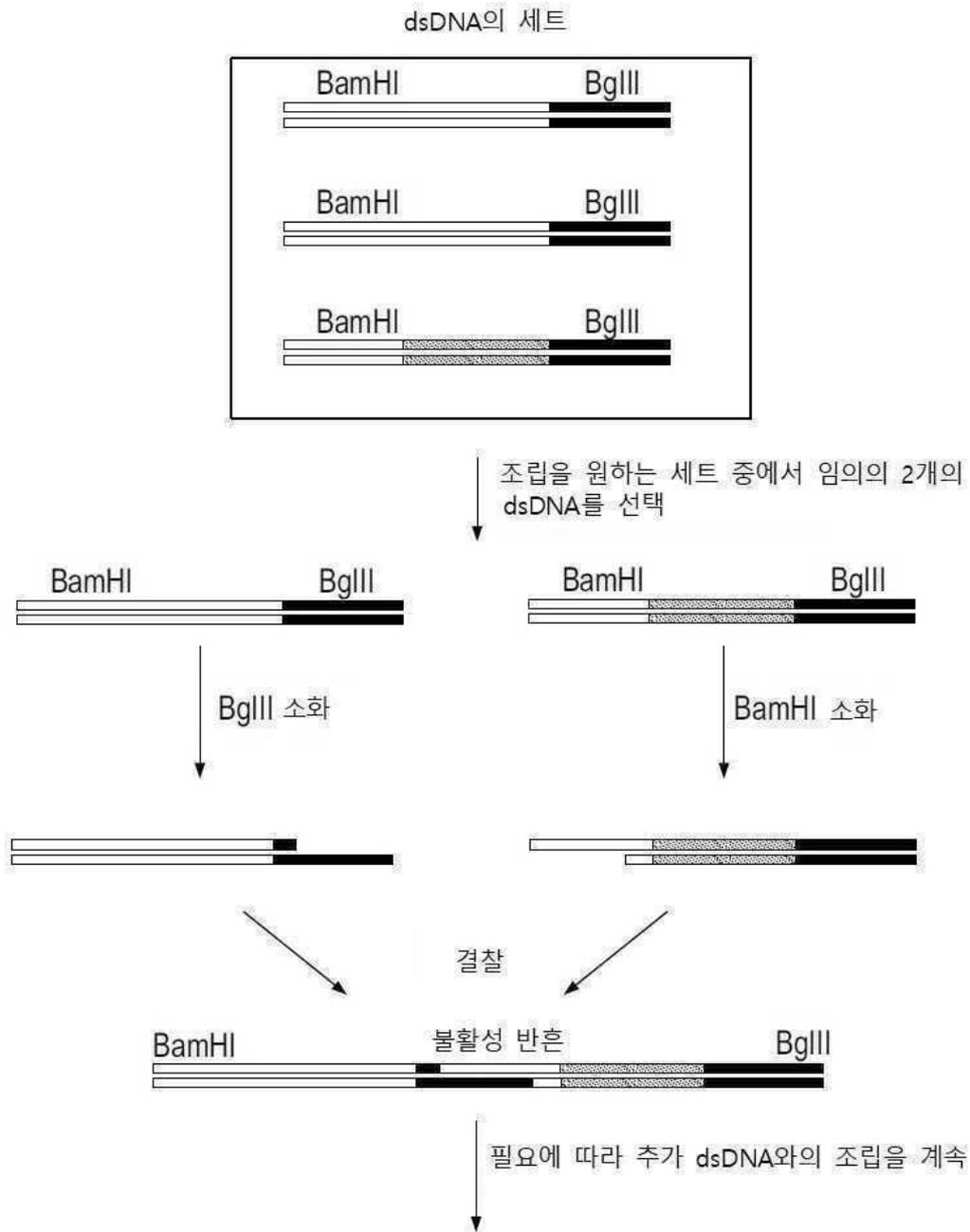
도면74b



도면75

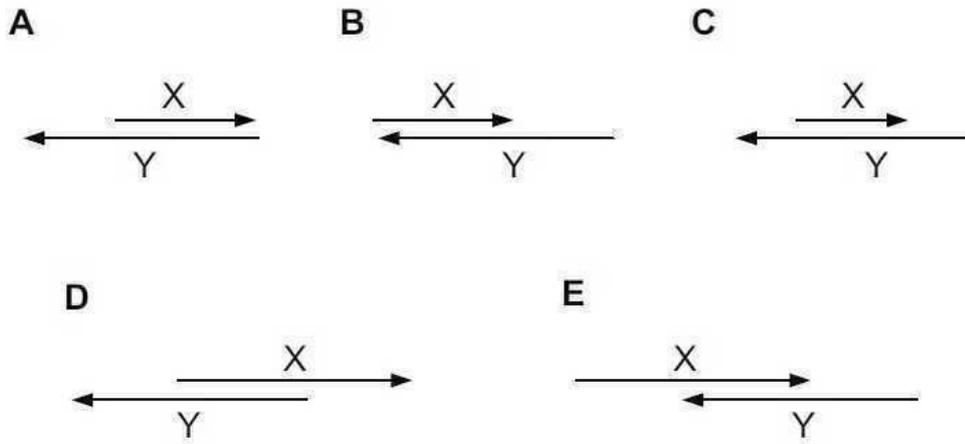


도면76

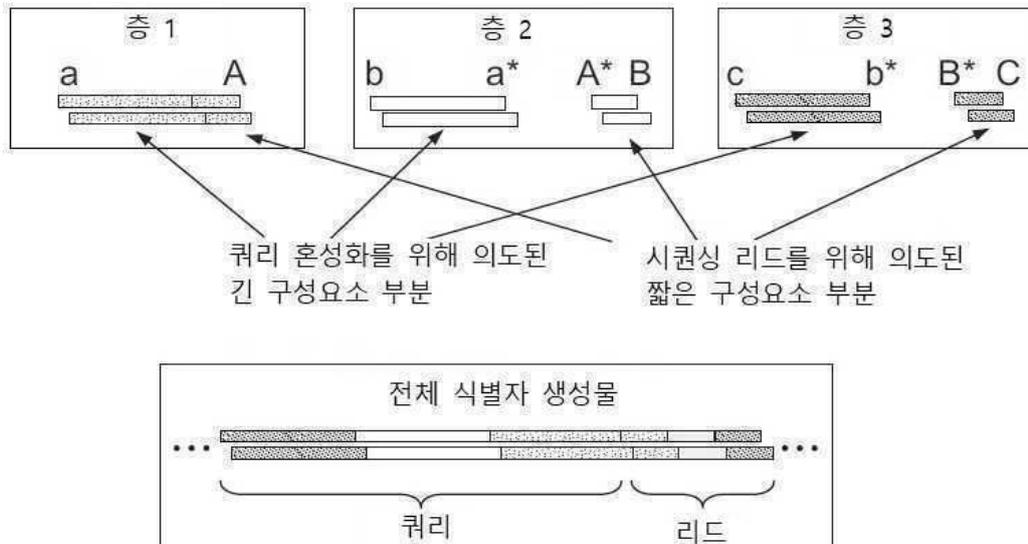


도면77

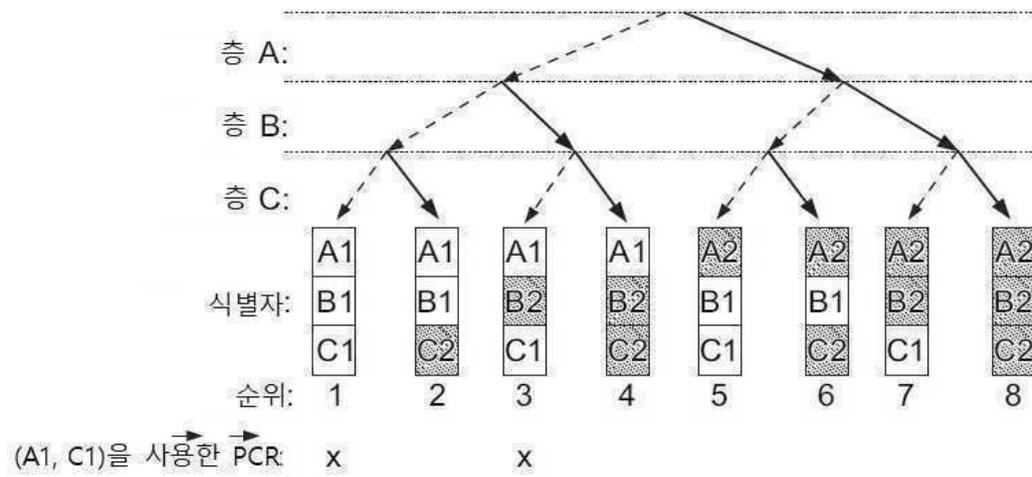
2개의 올리고 X 및 Y로부터 만들어진 가능한 점착성-말단 구성요소 구조.  
 화살표가 올리고를 나타냄. 화살표는 5'에서 3' 방향을 따라 가리킴.



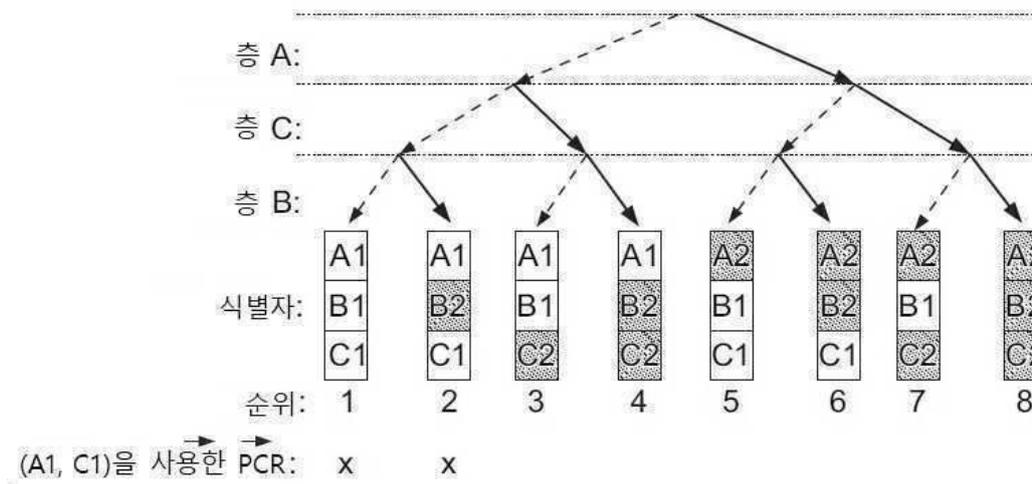
도면78



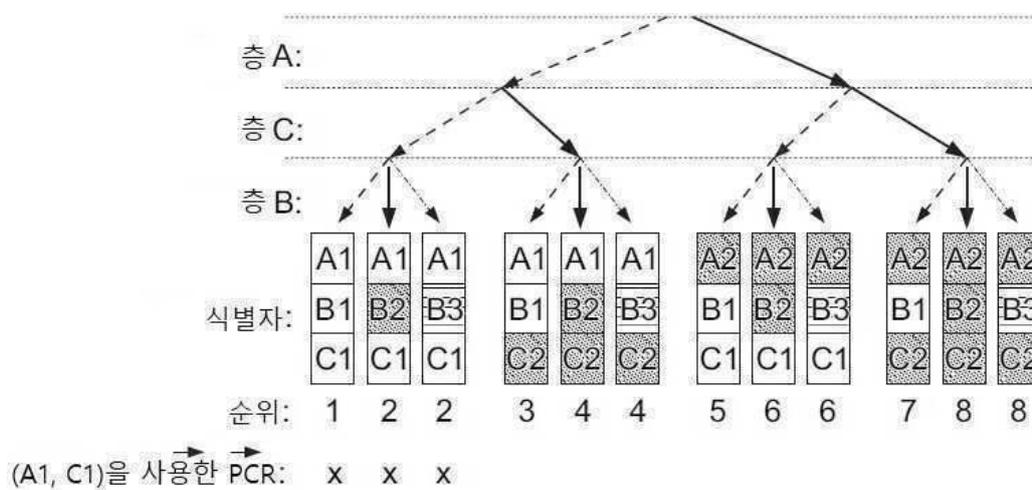
도면79a



도면79b

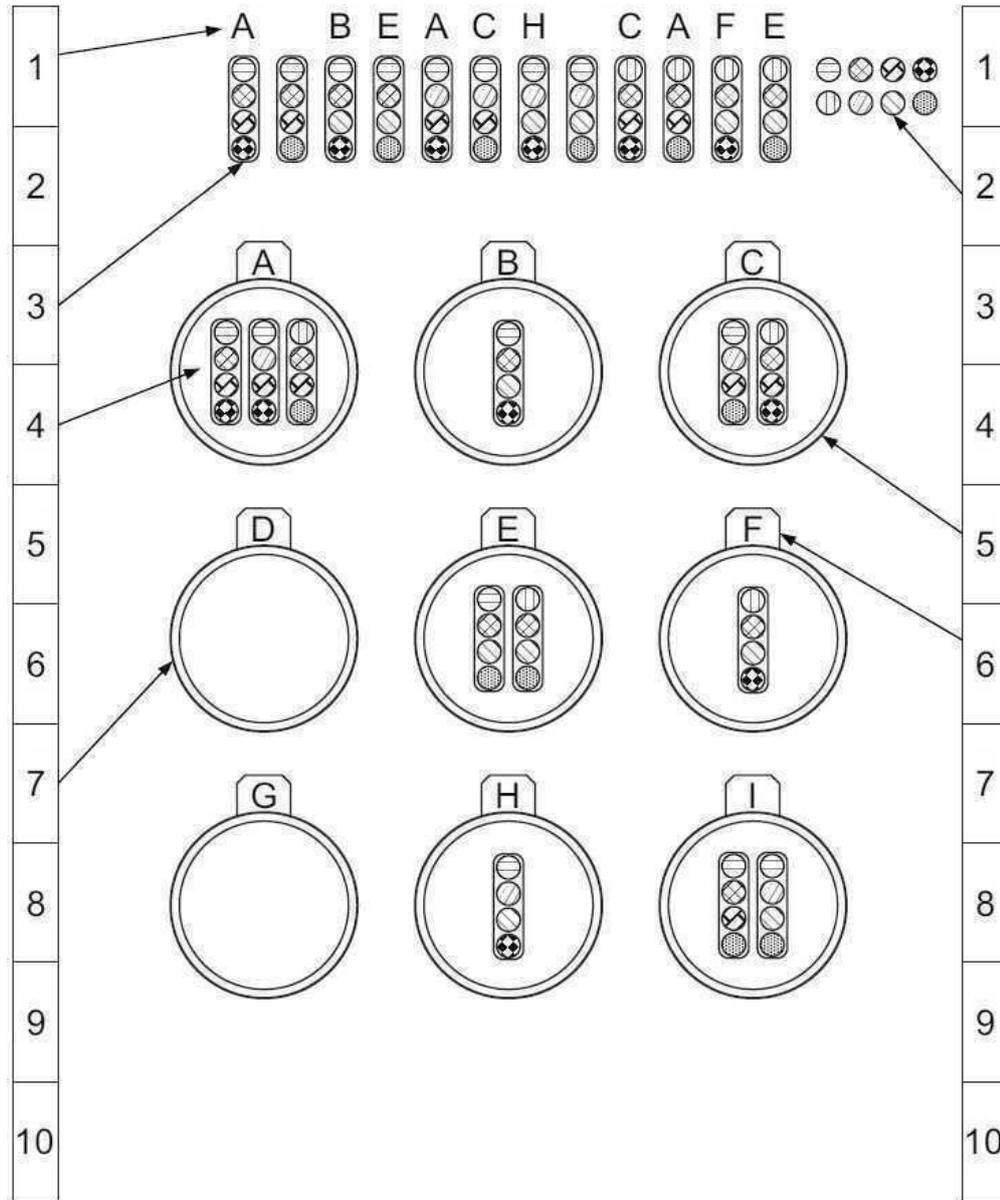


도면80a

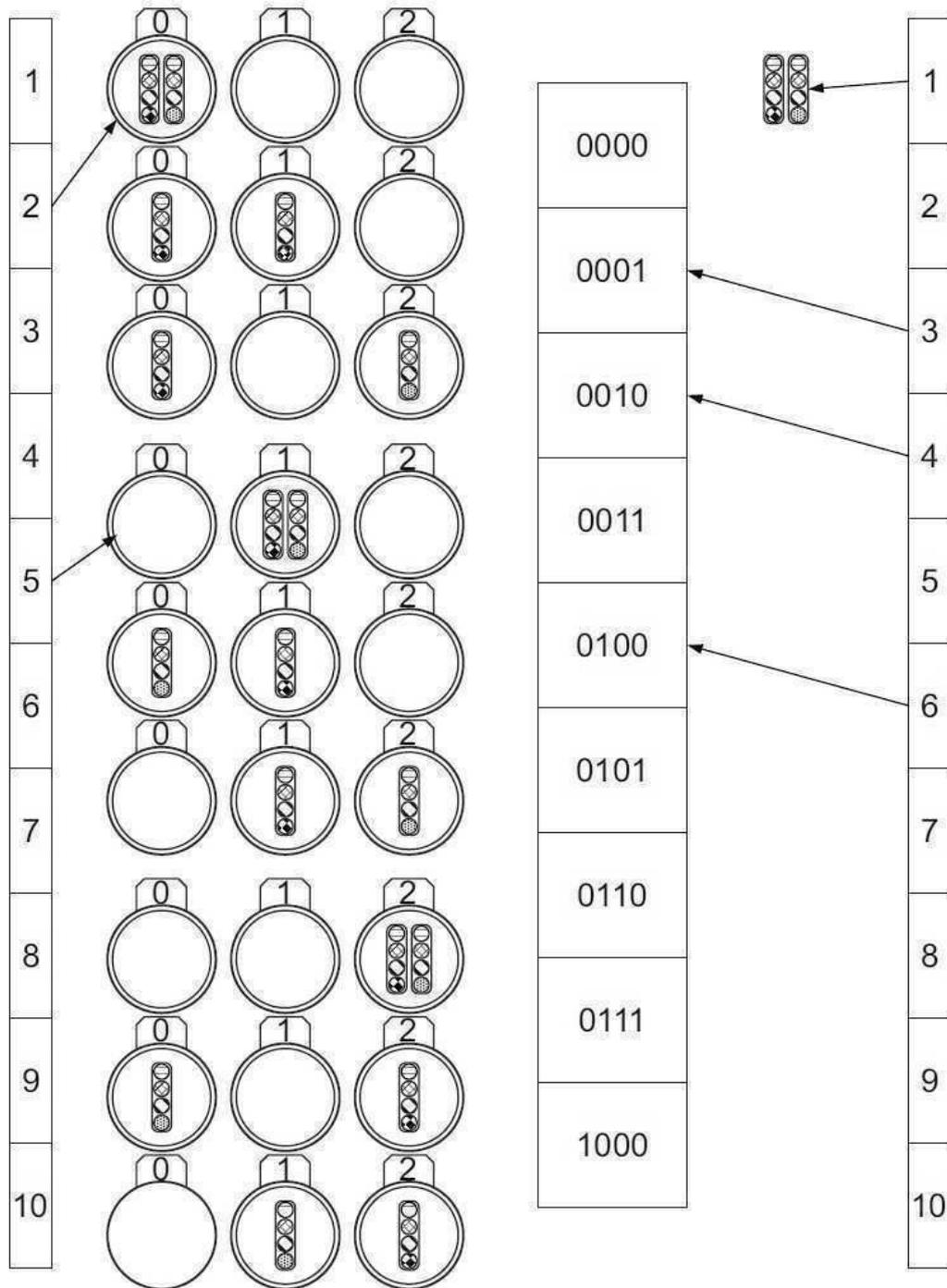




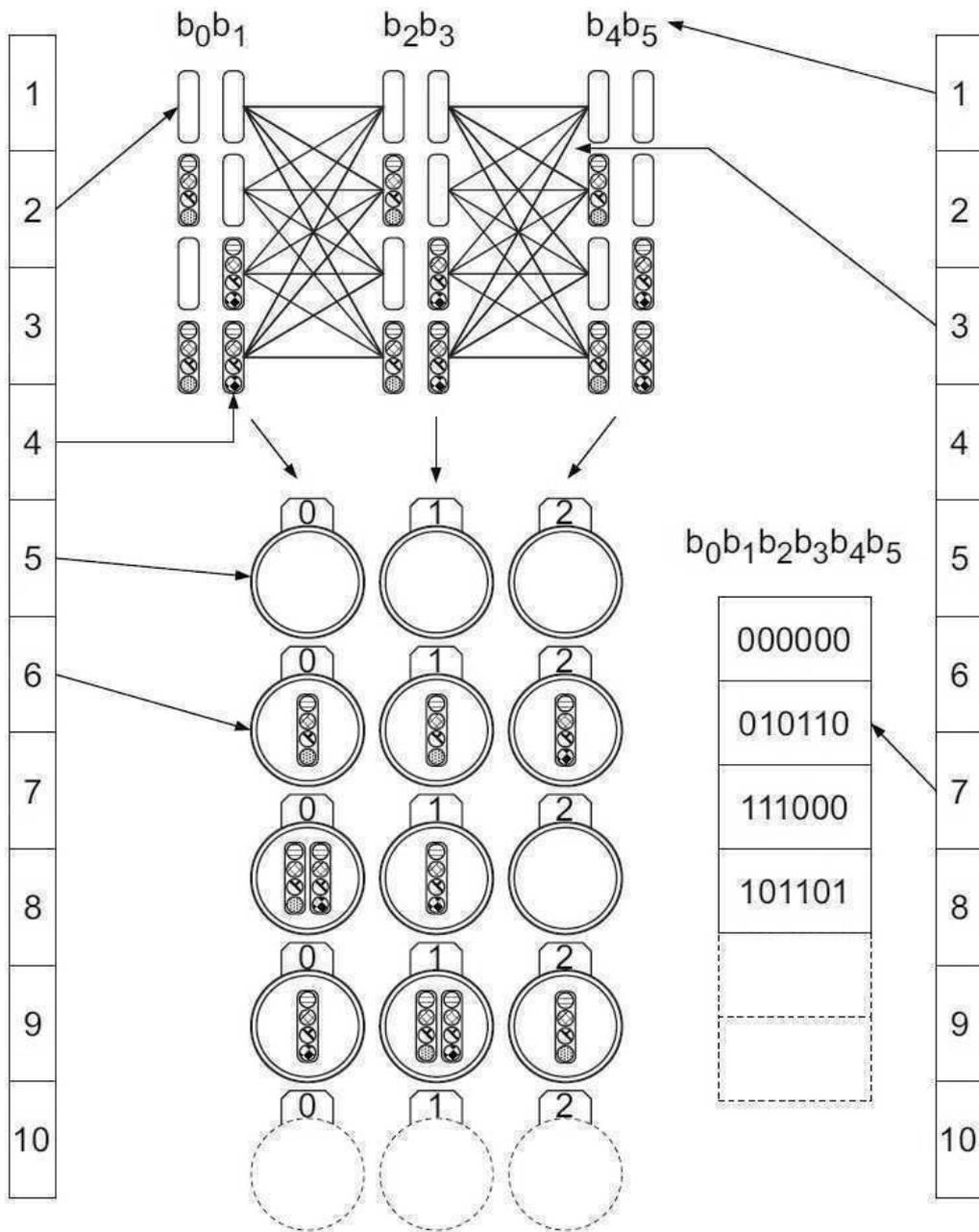
도면82



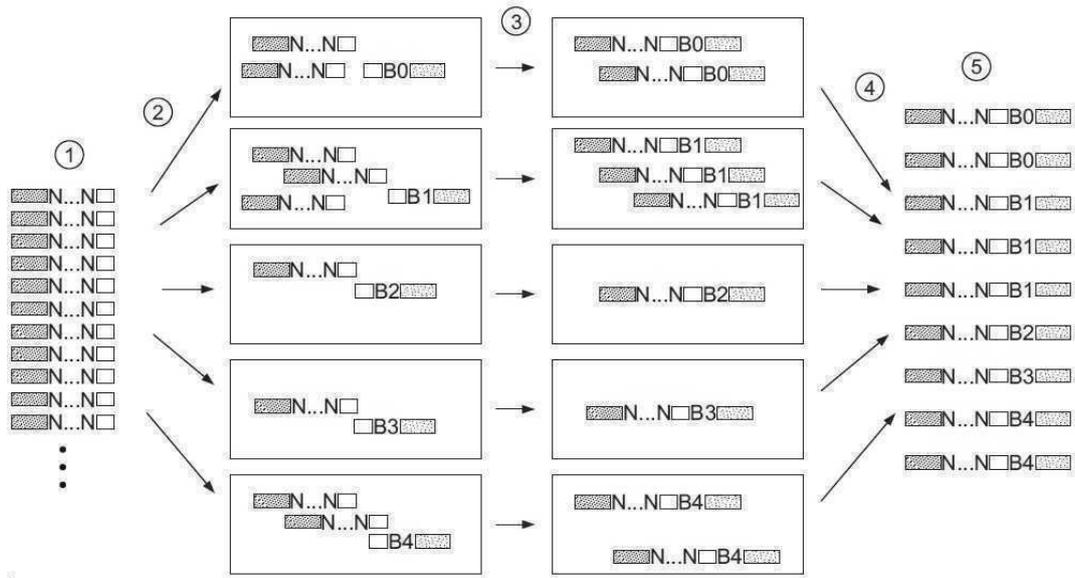
도면83



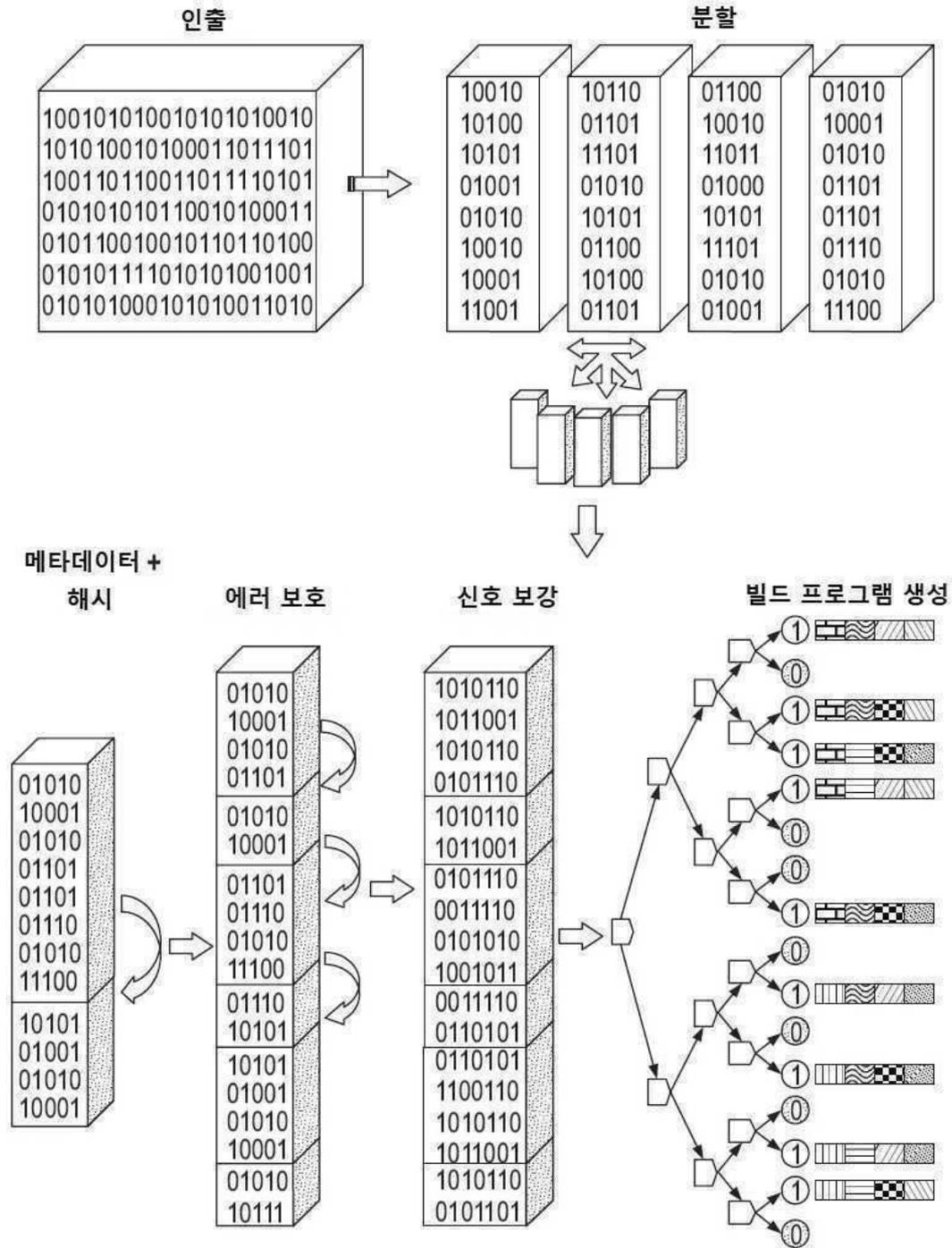
도면84



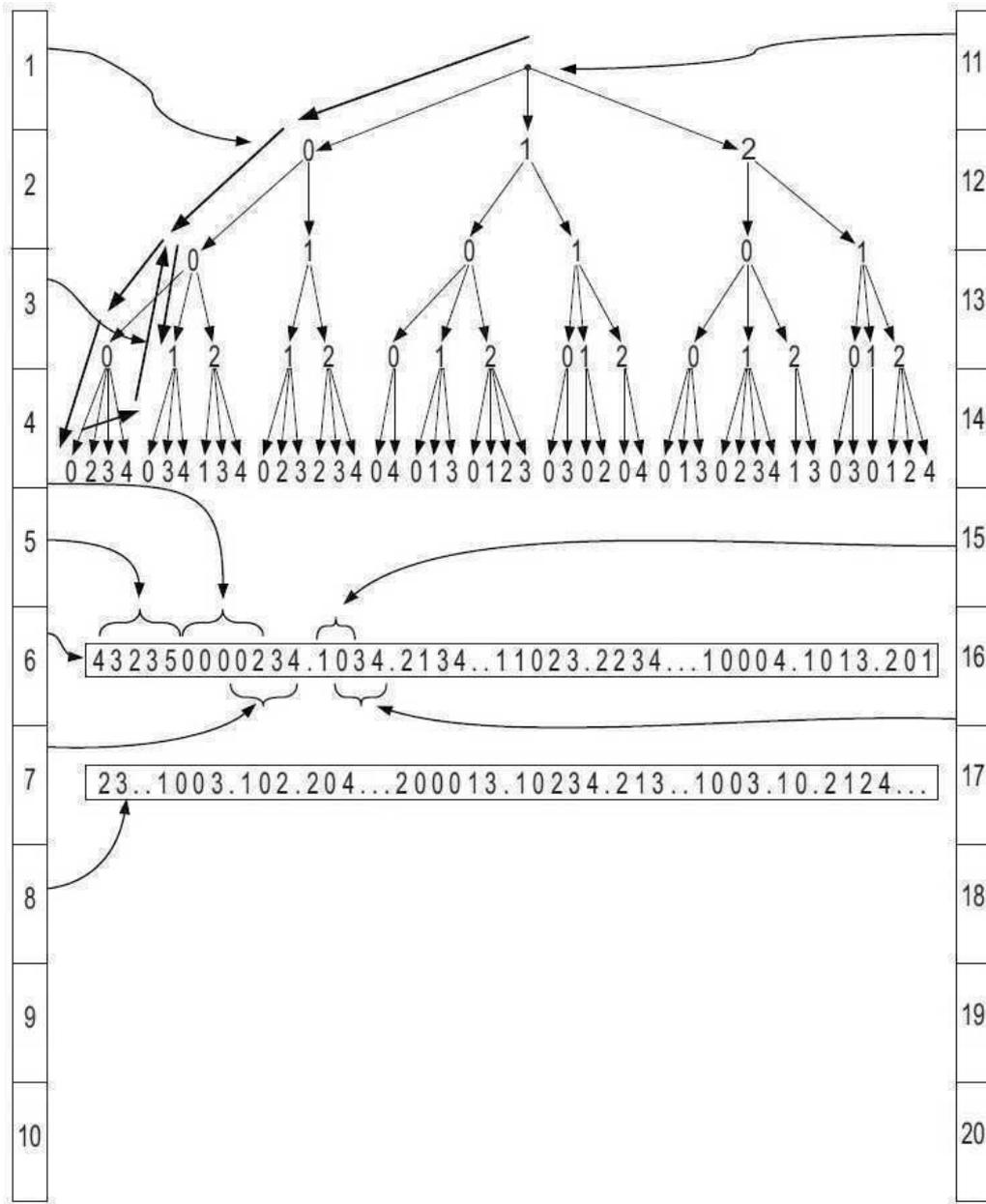
도면85



도면 86

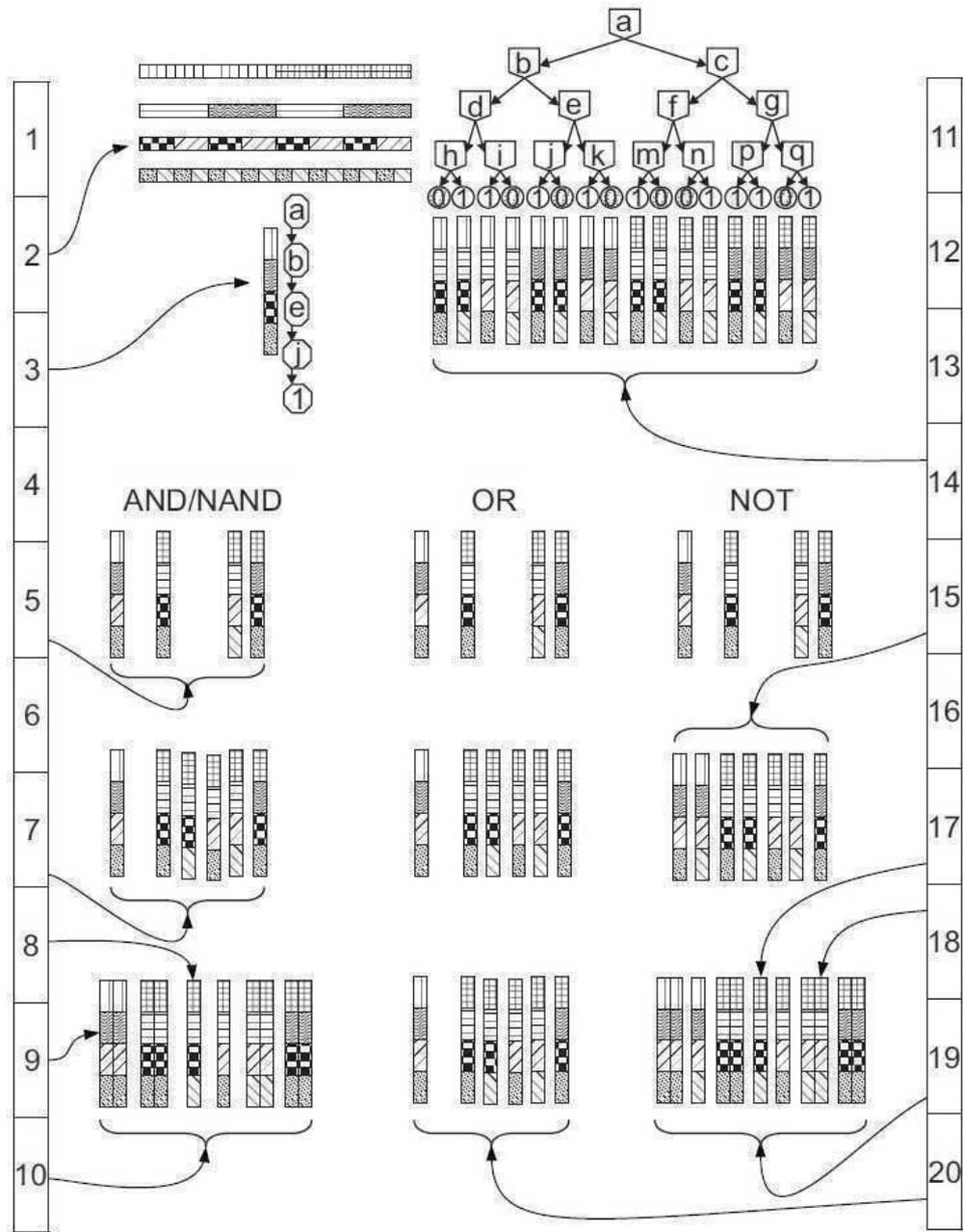


도면87

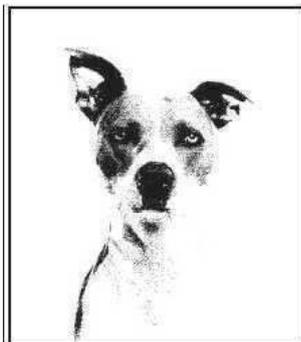




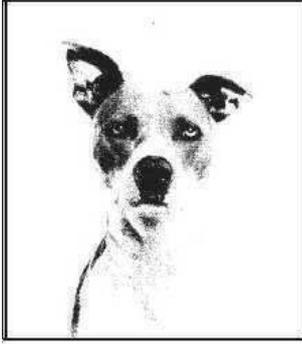
도면89



도면90a



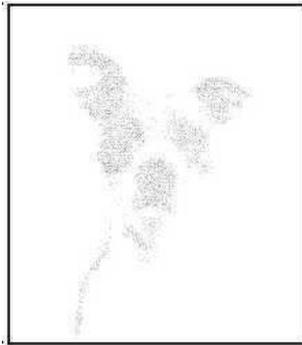
도면90b



도면90c



도면90d



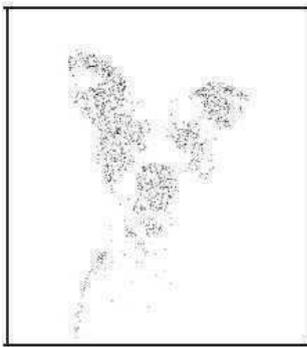
도면90e



도면90f

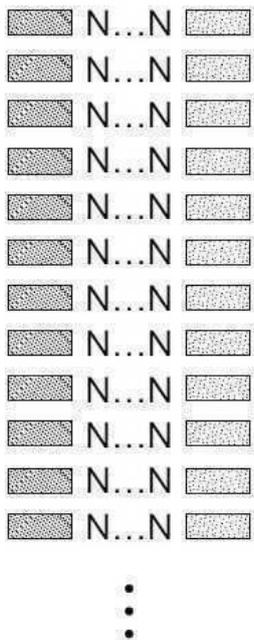


도면90g

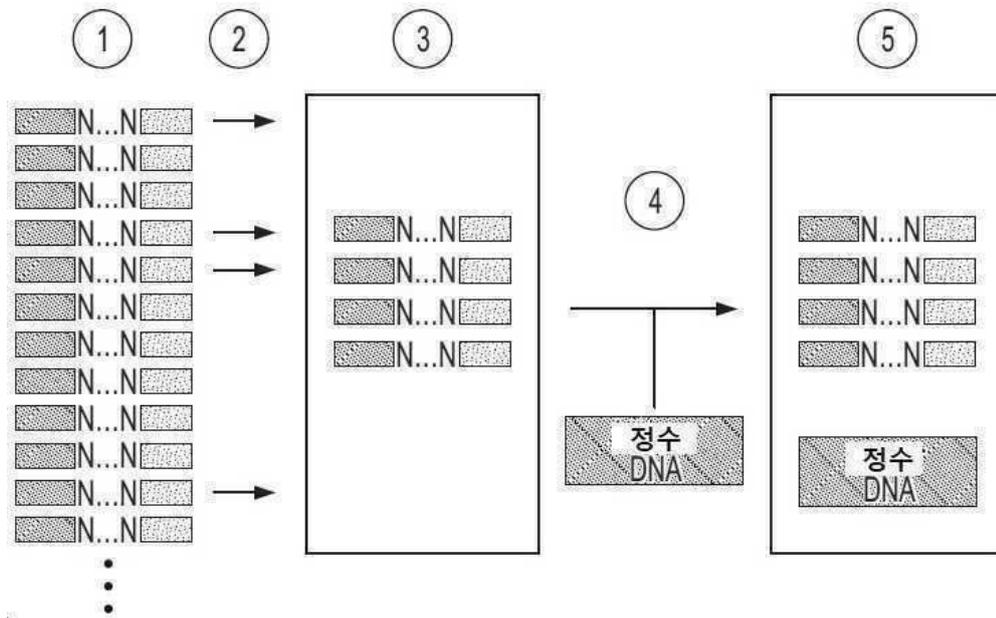


도면91

랜덤 DNA 풀  
(시드)

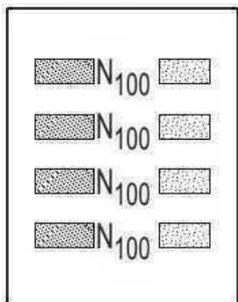


도면92a



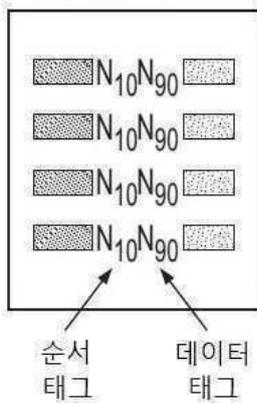
도면92b

$$\text{Log}_2(4^{100} \text{choose } 4) \sim 795 \text{ 비트}$$



도면92c

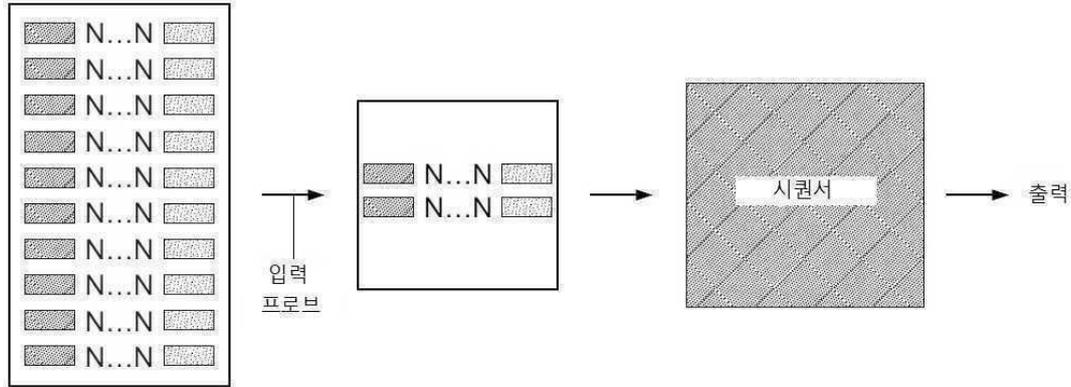
$$4 \text{Log}_2(4^{90}) \sim 720 \text{ 비트}$$



도면93a



도면93b



도면94

