



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2022-0052995  
(43) 공개일자 2022년04월28일

- (51) 국제특허분류(Int. Cl.)  
G16B 40/10 (2019.01) C12Q 1/68 (2018.01)  
G16B 50/00 (2019.01)
- (52) CPC특허분류  
G16B 40/10 (2019.02)  
C12Q 1/68 (2018.05)
- (21) 출원번호 10-2022-7010110
- (22) 출원일자(국제) 2020년08월26일  
심사청구일자 없음
- (85) 번역문제출일자 2022년03월25일
- (86) 국제출원번호 PCT/US2020/047994
- (87) 국제공개번호 WO 2021/041540  
국제공개일자 2021년03월04일
- (30) 우선권주장  
62/892,176 2019년08월27일 미국(US)

- (71) 출원인  
애플 바이오시스템즈, 인코포레이티드  
미국, 캘리포니아 94588, 플레젠턴, 오웬스 드라이브 6670
- (72) 발명자  
스테이커, 브라이언  
미국, 캘리포니아 94588, 플레젠턴, 오웬스 드라이브 6670  
볼링거, 데니스  
미국, 캘리포니아 94588, 플레젠턴, 오웬스 드라이브 6670
- (74) 대리인  
김영철, 김 순 영

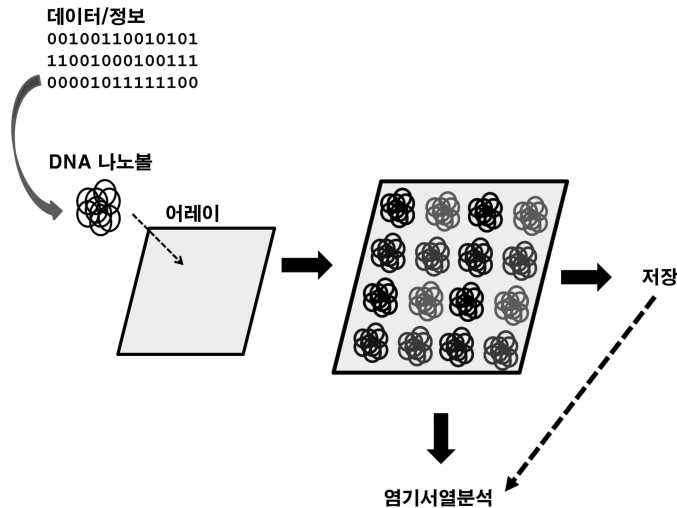
전체 청구항 수 : 총 73 항

(54) 발명의 명칭 **핵산 분자를 이용한 데이터 저장을 위한 시스템 및 방법**

**(57) 요약**

핵산 분자에 대한 데이터 및/또는 정보를 저장하고, 핵산 분자를 저장하고, 데이터 및/또는 정보를 검색하기 위한 방법 및 시스템은 본 명세서에 개시된다. 이러한 방법과 시스템은 데이터 검색의 효율성과 정확성을 향상시키는 것을 포함하여 데이터 저장을 위한 광범위한 응용 프로그램을 가지고 있다.

**대표도** - 도1



(52) CPC특허분류

**G16B 50/00** (2019.02)

C12Q 2563/179 (2013.01)

C12Q 2563/185 (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

데이터를 저장하는 방법으로서,

(a) 핵산 서열에서 상기 데이터를 인코딩하는 단계;

(b) 하나 이상의 핵산 분자를 생성하는 단계 -상기 하나 이상의 핵산 분자 중 하나의 핵산 분자는 상기 핵산 서열 및 헤더 서열의 적어도 일부를 포함하고, 상기 헤더 서열은 상기 핵산 서열의 상기 적어도 일부에 특이적인 서열을 포함하고, 상기 헤더 서열은 상기 핵산 서열의 상기 적어도 일부를 식별하기 위한 핵산 식별 반응의 개시를 허용하도록 구성됨-; 및

상기 하나 이상의 핵산 분자 또는 이의 유도체를 기관 상에 배치된 어레이에 저장하는 단계를 포함하는, 데이터를 저장하는 방법.

#### 청구항 2

제1항에 있어서, 상기 핵산 식별 반응은 염기서열분석(sequencing) 반응인, 데이터를 저장하는 방법.

#### 청구항 3

제1항에 있어서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 선형인, 데이터를 저장하는 방법.

#### 청구항 4

제1항에 있어서, 상기 하나 이상의 핵산 분자 또는 이의 유도체를 보존하는 단계를 더 포함하는, 데이터를 저장하는 방법.

#### 청구항 5

제4항에 있어서, 상기 보존하는 단계는 리오피라이제이션(lyophilization) 또는 동결건조하는 단계를 포함하는, 데이터를 저장하는 방법.

#### 청구항 6

제1항에 있어서, (b) 단계는 하나 이상의 증폭 산물을 형성하기 위해 상기 핵산 서열의 상기 적어도 일부를 증폭하는 단계를 더 포함하고, 상기 하나 이상의 핵산 분자는 상기 하나 이상의 증폭 산물을 포함하는, 데이터를 저장하는 방법.

#### 청구항 7

제6항에 있어서, 상기 증폭하는 단계는 롤링 서클 복제(rolling circle amplification)를 수행하는 단계를 포함하는, 데이터를 저장하는 방법.

#### 청구항 8

제6항에 있어서, 상기 증폭하는 단계는 브릿지 증폭(bridge amplification)을 수행하는 단계를 포함하는, 데이터를 저장하는 방법.

#### 청구항 9

제1항에 있어서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 연결된 핵산 분자들을 포함하는, 데이터를 저장하는 방법.

#### 청구항 10

제1항에 있어서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 상기 기관 상에 밀집하게 배치되고, 상기 하

나 이상의 핵산 분자 또는 상기 하나 이상의 핵산 분자 또는 이의 유도체의 유도체 및 인접한 핵산 분자 또는 이의 유도체 사이의 거리는 500 nm 미만인, 데이터를 저장하는 방법.

**청구항 11**

제10항에 있어서, 상기 거리는 중심 간 거리(center-to-center distance)를 포함하는, 데이터를 저장하는 방법.

**청구항 12**

제1항에 있어서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 제곱 마이크로미터 당 약 4 내지 약 25개의 핵산 분자 또는 이의 유도체의 밀도로 상기 기판 상에 배치되는, 데이터를 저장하는 방법.

**청구항 13**

제1항에 있어서, 상기 데이터를 검색하는 단계를 더 포함하는, 데이터를 저장하는 방법.

**청구항 14**

제13항에 있어서, 상기 검색하는 단계는 상기 하나 이상의 핵산 분자 또는 이의 유도체를 염기서열분석하는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 15**

제14항에 있어서, 상기 염기서열분석하는 단계는 검출 시스템을 이용하여 하나 이상의 통합된 핵산을 검출하는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 16**

제14항에 있어서, 상기 검출 시스템은 전기적 검출 시스템을 포함하는, 데이터를 저장하는 방법.

**청구항 17**

제16항에 있어서, 상기 전기적 검출 시스템은 트랜지스터를 포함하는, 데이터를 저장하는 방법.

**청구항 18**

제14항에 있어서, 상기 검출 시스템은 광학적 검출 시스템을 포함하는, 데이터를 저장하는 방법.

**청구항 19**

제18항에 있어서, 상기 광학적 검출 시스템은 광학 스캐닝 시스템을 포함하는, 데이터를 저장하는 방법.

**청구항 20**

제18항에 있어서, 상기 광학적 검출 시스템 상에서 검출된 상기 하나 이상의 통합된 핵산으로부터 생성된 신호의 파장은 상기 광학적 검출 시스템의 픽셀보다 2배 이상 큰, 데이터를 저장하는 방법.

**청구항 21**

제1항에 있어서, 상기 어레이는 정렬되는, 데이터를 저장하는 방법.

**청구항 22**

제1항에 있어서, 상기 어레이는 비정렬되는, 데이터를 저장하는 방법.

**청구항 23**

제1항에 있어서, 상기 시작 부위는 핵산 프라이머에 상보적인 핵산 서열을 포함하는, 데이터를 저장하는 방법.

**청구항 24**

제6항에 있어서, 상기 증폭하는 단계는 상기 저장하는 단계 이전에 발생하는, 데이터를 저장하는 방법.

**청구항 25**

데이터를 저장하는 방법으로서,

(a) 핵산 서열에서 상기 데이터를 인코딩하는 단계;

(b) 상기 핵산 서열을 포함하는 하나 이상의 핵산 분자를 생성하는 단계; 및

(c) 어레이를 제공하기 위해, 기관 상에 배치된 어레이에 상기 하나 이상의 핵산 분자를 저장하는 단계 -상기 어레이가 광학 스캐닝 시스템을 사용하여 이미지화되는 경우, 상기 하나 이상의 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀의 크기의 2배보다 큼-를 포함하는, 데이터를 저장하는 방법.

**청구항 26**

제25항에 있어서, 상기 하나 이상의 핵산 분자는 선형인, 데이터를 저장하는 방법.

**청구항 27**

제25항에 있어서, (b) 단계는 상기 핵산 서열의 적어도 일부를 포함하는 하나 이상의 선형 핵산 분자를 생성하는 단계; 상기 하나 이상의 선형 핵산 분자를 원형화하는 단계; 및 하나 이상의 연결된 핵산 분자를 생성하기 위해 롤링 서클 복제로 증폭하는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 28**

제25항에 있어서, (b) 단계는

상기 핵산 서열, 제1 어댑터 서열, 및 제2 어댑터 서열을 포함하는 하나 이상의 선형 핵산 분자를 생성하는 단계-상기 제1 및 상기 제2 어댑터 서열은 하나 이상의 원형 핵산 분자의 형성을 가능하게 함-; 및

상기 하나 이상의 원형 핵산 분자를 증폭시키는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 29**

제28항에 있어서, 상기 선형 핵산 분자는 하나 이상의 기능적 서열을 포함하는, 데이터를 저장하는 방법.

**청구항 30**

제28항에 있어서, 상기 하나 이상의 연쇄동일서열적(concatemeric) 핵산 분자는 롤링 서클 복제에 의해 생성되는, 데이터를 저장하는 방법.

**청구항 31**

제25항에 있어서, (c) 단계는 상기 연쇄동일서열적 핵산 분자를 상기 기관 상에 배치하는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 32**

제31항에 있어서, 상기 하나 이상의 연쇄동일서열적 핵산 분자는 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda / (2 * NA)$ 의 측정치 미만인 밀도로 배치되는, 데이터를 저장하는 방법.

**청구항 33**

제25항에 있어서, 상기 방법은 상기 기관을 보존하는 단계를 더 포함하는, 데이터를 저장하는 방법.

**청구항 34**

제33항에 있어서, 상기 보존하는 단계는 리오피라이제이션 또는 동결건조하는 단계를 포함하는, 데이터를 저장하는 방법.

**청구항 35**

제25항에 있어서, 상기 기판은 실리콘을 포함하는, 데이터를 저장하는 방법.

**청구항 36**

제25항에 있어서, 상기 기판은 유리를 포함하는, 데이터를 저장하는 방법.

**청구항 37**

제36항에 있어서, 상기 기판은 2개의 유리 조각을 포함하는, 데이터를 저장하는 방법.

**청구항 38**

제25항에 있어서, 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 데이터를 검색하는 단계를 더 포함하는, 데이터를 저장하는 방법.

**청구항 39**

제25항에 있어서, 상기 어레이는 정렬되는, 데이터를 저장하는 방법.

**청구항 40**

제25항에 있어서, 상기 어레이는 비정렬되는, 데이터를 저장하는 방법.

**청구항 41**

제39항에 있어서, 상기 정렬은 무작위적인, 데이터를 저장하는 방법.

**청구항 42**

핵산 분자를 기판에 배치하는 단계-상기 핵산 분자 또는 이의 유도체는 데이터를 인코딩 함-를 포함하는, 데이터를 저장하는 방법.

**청구항 43**

제42항에 있어서, 상기 핵산 분자 또는 이의 유도체는 핵산 연쇄체를 포함하는, 데이터를 저장하는 방법.

**청구항 44**

제42항에 있어서, 상기 핵산 분자 또는 이의 유도체는 밀도로 배치되고, 상기 기판이 광학 스캐닝 시스템을 사용하여 이미징되는 경우, 상기 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀 크기의 2배보다 큰, 데이터를 저장하는 방법.

**청구항 45**

제42항에 있어서, 상기 기판은 실리콘을 포함하는, 데이터를 저장하는 방법.

**청구항 46**

제42항에 있어서, 상기 기판은 유리를 포함하는, 데이터를 저장하는 방법.

**청구항 47**

제46항에 있어서, 상기 기판은 2개의 유리 조각을 포함하는, 데이터를 저장하는 방법.

**청구항 48**

제42항에 있어서, 상기 데이터는 염기서열분석 이전에 증폭 없이 상기 핵산 분자로부터 검색되는, 데이터를 저장하는 방법.

**청구항 49**

정보의 하나 이상의 비트를 저장하는 방법으로서, 상기 방법은:

- (a) 복수의 뉴클레오타이드에서 상기 하나 이상의 정보의 비트를 인코딩하는 단계;
- (b) 상기 복수의 뉴클레오타이드를 하나 이상의 프라이머에 결합하는 단계;
- (c) 상기 복수의 뉴클레오타이드를 약 300 내지 약 1,000개의 뉴클레오타이드의 길이로 합성하는 단계;
- (d) 상기 복수의 뉴클레오타이드를 원형화하는 단계;
- (e) 하나 이상의 핵산 분자를 생성하기 위해 롤링 서클 복제에 의해 상기 복수의 원형 분자를 증폭하는 단계; 및
- (f) 상기 하나 이상의 핵산 분자를 기관 상에 배치하는 단계를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 50**

정보의 하나 이상의 비트를 저장하는 방법으로서, 상기 방법은:

- (a) 상기 하나 이상의 정보의 비트를 인코딩하는 선형 핵산 분자를 합성하는 단계  
-상기 선형 핵산 분자는:
  - i. 상기 하나 이상의 정보의 비트를 인코딩하는 핵산 서열,
  - ii. 5' 어댑터 시퀀스,
  - iii. 3' 어댑터 시퀀스,
  - iv. 선택적인 하나 이상의 추가 기능적 서열을 포함하고-
- (b) 상기 선형 핵산 분자로부터 원형 핵산 분자를 생성하는 단계,
- (c) 상기 원형 핵산 분자를 증폭하여 상기 원형 핵산 분자의 하나 이상의 사본을 포함하는 증폭된 핵산 분자를 생성하기 위해 상기 원형 핵산 분자를 증폭하는 단계,
- (d) 상기 증폭된 핵산 분자를 기관 상에 배치하는 단계를 포함하는, 정보의 하나 이상의 비트를 저장하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 51**

제50항에 있어서, 상기 기관은 패터닝되는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 52**

제50항에 있어서, 상기 기관은 패터닝 되지 않는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 53**

제50항에 있어서, 상기 방법은 상기 하나 이상의 기관을 보존하는 단계를 더 포함하는, 하나 이상의 비트를 저장하는 방법.

**청구항 54**

제53항에 있어서, 상기 보존하는 단계는 리오필라이제이션 또는 동결건조하는 단계를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 55**

제50항에 있어서, 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 하나 이상의 정보의 비트를 검색하는 단계를 더 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 56**

제50항에 있어서, 상기 하나 이상의 정보의 비트를 검색하는 단계는 핵산 식별 반응을 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 57**

제51항에 있어서, 복구된 하나 이상의 정보의 비트에 오류 정정을 적용하는 단계를 더 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 58**

제52항에 있어서, 상기 오류 정정은 리드-솔로몬 부호(Reed-Solomon code)를 사용하는 단계를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 59**

제50항에 있어서, 상기 정보의 비트는 이진 비트를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 60**

제50항에 있어서, 상기 정보의 비트는 이진 비트를 포함하고

(a) 단계는 정보의 상기 이진 비트를 정보의 제4기 비트로 전사하는 단계를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 61**

제50항에 있어서, 상기 5' 어댑터 서열, 3' 어댑터 서열, 또는 둘 모두는 바코드 부호 서열을 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 62**

제50항에 있어서, 상기 하나 이상의 기능적 서열은 바코드 부호 서열, 태그 서열, 범용 프라이머 서열, 고유 식별자 서열, 또는 추가 어댑터 서열로 이루어진 그룹으로부터 선택되는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 63**

제50항에 있어서, 상기 원형 핵산 분자는 상기 5' 어댑터 및 상기 3' 어댑터를 결합함으로써 생성되는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 64**

제50항에 있어서, 상기 원형 핵산 분자는 롤링 서클 반응에 의해 증폭되는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 65**

제50항에 있어서, 상기 증폭된 핵산 분자는 핵산 연쇄체(concatemer)인, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 66**

제50항에 있어서, 상기 증폭된 핵산 분자는 밀집하게 배치되고, 상기 기관이 광학 스캐닝 시스템을 사용하여 이미징되는 경우, 상기 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀 크기의 2배보다 큰, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 67**

제50항에 있어서, 상기 기관은 실리콘을 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 68**

제50항에 있어서, 상기 기관은 유리를 포함하는, 정보의 하나 이상의 비트를 저장하는 방법.

**청구항 69**

제1항 내지 제68항 중 어느 한 항에 있어서, 상기 어레이는 제1 및 제2 유리 기판을 포함하는, 방법.

**청구항 70**

제1항 내지 제69항 중 어느 한 항에 있어서, 상기 방법은 제1항 내지 제69항 중 어느 한 항의 방법을 구현하도록 프로그래밍된 컴퓨터 시스템에 의해 자동화되는, 방법.

**청구항 71**

제1항 내지 제70항 중 어느 한 항의 방법을 구현하도록 프로그래밍 된, 컴퓨터 시스템.

**청구항 72**

복수의 핵산 서열을 포함하는 -상기 복수의 핵산 서열의 적어도 일부는 적어도 1GB의 데이터를 인코딩하고, 상기 핵산 분자는 상기 핵산 분자의 1% 이하가 1년의 기간에 걸쳐 분해되도록 하는 안정성을 갖음-, 핵산 분자.

**청구항 73**

제72항에 있어서, 복수의 헤더 서열을 더 포함하고, 상기 복수의 헤더 서열의 헤더 서열은 상기 핵산 서열의 적어도 일부의 염기서열분석을 허용하여 상기 1GB의 데이터를 검색하도록 구성된, 핵산분자.

**발명의 설명**

**기술 분야**

[0001] 본 출원은 2019년 8월 27일에 출원된 미국 가출원 제62/892,176호의 이익을 주장하며, 그 전체가 본 명세서에 참조로 포함된다.

**배경 기술**

[0002] 세계의 빅 데이터 과제 및 문제에 대한 규모와 복잡성은 빠르게 증가하고 있다. 이러한 과제를 해결하는 것은 굉장한 기술적이고 재정적인 문제를 제기한다. 예를 들어, 엑사바이트(Exabyte) 크기의 데이터 저장 센터는 막대한 양의 리소스를 소모하고 부담이 된다. 현재의 엑사바이트 크기의 데이터 저장 장치는 대형 창고를 필요로 하고 메가와트의 전력을 소모하고, 구축, 운영 및 유지하기 위해 수십억 달러의 비용이 든다. 이러한 자원 집약 모델은 향후 확장에 대한 실용적이거나 실현 가능한 다루기 쉬운 경로를 제공하지 못한다.

**발명의 내용**

**해결하려는 과제**

[0003] 본 발명은 종래의 저장 기술에 비해 물리적 공간, 전력 및 비용 요건 대비 감소된 리소스 풋프린트를 제공하고 확장 가능한 핵산 매개 데이터 저장 방법을 제공한다.

**과제의 해결 수단**

[0004] 본 명세서에 기술된 시스템 및 방법은 1) 어레이는 염기서열분석하는 단계/관독하는 단계 이전에 핵산 서열의 증폭이 없는 레디-투-리드 방식으로 생성될 수 있고, 2) 데이터 정보를 인코딩하는 핵산은 고밀도 어레이에 저장될 수 있고, 여기서 하나 이상의 핵산 분자 사이의 거리는 빛의 회절 한계 미만인 핵산 저장의 장점을 제공할 수 있다.

[0005] 본 명세서에 기술된 본 개시의 태양은 핵산 서열에서 상기 데이터를 인코딩하는 단계; 하나 이상의 핵산 분자를 생성하는 단계 -상기 하나 이상의 핵산 분자 중 하나의 핵산 분자는 상기 핵산 서열 및 헤더 서열의 적어도 일부를 포함하고, 상기 헤더 서열은 상기 핵산 서열의 상기 적어도 일부에 특이적인 서열을 포함하고, 상기 헤더 서열은 상기 핵산 서열의 상기 적어도 일부를 식별하기 위한 핵산 식별 반응의 개시를 허용하도록 구성됨-; 상기 하나 이상의 핵산 분자 또는 이의 유도체를 기판 상에 배치된 어레이에 저장하는 단계를 포함하는, 데이터를 저장하는 방법을 제공한다. 일부 실시예에서, 상기 핵산 식별 반응은 염기서열분석(sequencing) 반응이다. 일부 실시예에서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 선형이다. 일부 실시예에서, 상기 방법은 상기 하

나 이상의 핵산 분자 또는 이의 유도체를 보존하는 단계를 더 포함한다. 일부 실시예에서, 상기 보존하는 단계는 리오피라이제이션(lyophilization) 또는 동결건조하는 단계를 포함한다. 일부 실시예에서, (b) 단계는 하나 이상의 증폭 산물을 형성하기 위해 상기 핵산 서열의 상기 적어도 일부를 증폭하는 단계를 더 포함하고, 상기 하나 이상의 핵산 분자는 상기 하나 이상의 증폭 산물을 포함한다. 일부 실시예에서, 상기 증폭 단계는 롤링 서클 복제(rolling circle amplification)를 수행하는 단계를 포함한다. 일부 실시예에서, 상기 증폭하는 단계는 브릿지 증폭(bridge amplification)을 수행하는 단계를 포함한다. 일부 실시예에서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 연결된 핵산 분자들을 포함한다. 일부 실시예에서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 상기 기관 상에 밀집하게 배치되고, 상기 하나 이상의 핵산 분자 또는 상기 하나 이상의 핵산 분자 또는 이의 유도체의 유도체 및 인접한 핵산 분자 또는 이의 유도체 사이의 거리는 500 nm 미만이다. 일부 실시예에서, 상기 거리는 중심 간 거리(center-to-center distance)를 포함한다. 일부 실시예에서, 상기 하나 이상의 핵산 분자 또는 이의 유도체는 제곱 마이크로미터 당 약 4 내지 약 25개의 핵산 분자 또는 이의 유도체의 밀도로 상기 기관 상에 배치된다. 일부 실시예에서, 상기 방법은 상기 데이터를 검색하는 단계를 더 포함한다. 일부 실시예에서, 상기 검색하는 단계는 상기 하나 이상의 핵산 분자 또는 이의 유도체를 염기서열분석하는 단계를 포함한다. 일부 실시예에서, 상기 염기서열분석하는 단계는 검출 시스템을 이용하여 하나 이상의 통합된 핵산을 검출하는 단계를 포함한다. 일부 실시예에서, 상기 검출 시스템은 전기적 검출 시스템을 포함한다. 일부 실시예에서, 상기 전기적 검출 시스템은 트랜지스터를 포함한다. 일부 실시예에서, 상기 검출 시스템은 광학적 검출 시스템을 포함한다. 일부 실시예에서, 상기 광학적 검출 시스템은 광학 스캐닝 시스템을 포함한다. 일부 실시예에서, 상기 광학적 검출 시스템 상에서 검출된 상기 하나 이상의 통합된 핵산으로부터 생성된 신호의 파장은 상기 광학적 검출 시스템의 픽셀보다 2배 이상 크다. 일부 실시예에서, 상기 어레이는 정렬된다. 일부 실시예에서, 상기 어레이는 비정렬된다. 일부 실시예에서, 상기 시작 부위는 핵산 프라이머에 상보적인 핵산 서열을 포함한다. 일부 실시예에서, 상기 증폭하는 단계는 상기 저장하는 단계 이전에 발생한다.

[0006] 본 명세서에 기재된 본 발명의 다른 태양은 핵산 서열에서 상기 데이터를 인코딩하는 단계; 상기 핵산 서열을 포함하는 하나 이상의 핵산 분자를 생성하는 단계; 및 상기 어레이를 제공하기 위해, 기관 상에 배치된 어레이에 상기 하나 이상의 핵산 분자를 저장하는 단계 -상기 어레이가 광학 스캐닝 시스템을 사용하여 이미지화되는 경우, 상기 하나 이상의 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀의 크기의 2배보다 큼-를 포함하는, 데이터를 저장하는 방법을 제공한다. 일부 실시예에서, 상기 하나 이상의 핵산 분자는 선형이다. 일부 실시예에서, (b) 단계는 상기 핵산 서열의 적어도 일부를 포함하는 하나 이상의 선형 핵산 분자를 생성하는 단계; 상기 하나 이상의 선형 핵산 분자를 원형화하는 단계; 및 하나 이상의 연결된 핵산 분자를 생성하기 위해 롤링 서클 복제로 증폭하는 단계를 포함한다. 일부 실시예에서, 상기 (b) 단계는 상기 핵산 서열, 제1 어댑터 서열, 및 제2 어댑터 서열을 포함하는 하나 이상의 선형 핵산 분자를 생성하는 단계-상기 제1 및 상기 제2 어댑터 서열은 하나 이상의 원형 핵산 분자의 형성을 가능하게 함-; 및 상기 하나 이상의 원형 핵산 분자를 증폭시키는 단계를 포함한다. 일부 실시예에서, 상기 선형 핵산 분자는 하나 이상의 기능적 서열을 포함한다. 일부 실시예에서, 상기 하나 이상의 연쇄동일서열적(concatemeric) 핵산 분자는 롤링 서클 복제에 의해 생성된다. 일부 실시예에서, (c) 단계는 상기 연쇄동일서열적 핵산 분자를 상기 기관 상에 배치하는 단계를 포함한다. 일부 실시예에서, 상기 하나 이상의 연쇄동일서열적 핵산 분자는 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다. 일부 실시예에서, 상기 방법은 상기 기관을 보존하는 단계를 더 포함한다. 일부 실시예에서, 상기 보존하는 단계는 리오피라이제이션 또는 동결건조하는 단계를 포함한다. 일부 실시예에서, 상기 기관은 실리콘을 포함한다. 일부 실시예에서, 상기 기관은 유리를 포함한다. 일부 실시예에서, 상기 기관은 2개의 유리 조각을 포함한다. 일부 실시예에서, 상기 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 데이터를 검색하는 단계를 더 포함한다. 일부 실시예에서, 상기 어레이는 정렬된다. 일부 실시예에서, 상기 어레이는 비정렬된다. 일부 실시예에서, 상기 정렬은 무작위적이다.

[0007] 본 명세서에서 기재된 본 발명의 다른 태양은 핵산 분자를 기관에 배치하는 단계-상기 핵산 분자 또는 이의 유도체는 상기 데이터를 인코딩 함-를 포함하는 데이터를 저장하는 방법을 제공한다. 일부 실시예에서, 상기 핵산 분자 또는 이의 유도체는 핵산 연쇄체를 포함한다. 일부 실시예에서, 상기 핵산 분자 또는 이의 유도체는 밀도로 배치되고, 상기 기관이 광학 스캐닝 시스템을 사용하여 이미지되는 경우, 상기 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀 크기의 2배보다 크다. 일부 실시예에서, 상기 기관은 실리콘을 포함한다. 일부 실시예에서, 상기 기관은 유리를 포함한다. 일부 실시예에서, 상기 기관은 2개의 유리 조각을 포함한다. 일부 실시예에서, 상기 데이터는 염기서열분석 이전에 증폭 없이 상기 핵산 분자로부터 검색된다.

[0008] 본 명세서에서 기재된 본 발명의 다른 태양은 정보의 하나 이상의 비트를 저장하는 방법을 제공하고, 상기 방법

은: 복수의 뉴클레오티드에서 상기 하나 이상의 정보의 비트를 인코딩하는 단계; 상기 복수의 뉴클레오티드를 하나 이상의 프라이머에 결합하는 단계; 상기 복수의 뉴클레오티드를 약 300 내지 약 1,000개의 뉴클레오티드의 길이로 합성하는 단계; 상기 복수의 뉴클레오티드를 원형화하는 단계; 하나 이상의 핵산 분자를 생성하기 위해 롤링 서클 복제에 의해 상기 복수의 원형 분자를 증폭하는 단계; 및 상기 하나 이상의 핵산 분자를 기판 상에 배치하는 단계를 포함한다.

[0009] 본 명세서에서 기재된 본 발명의 다른 태양은 정보의 하나 이상의 비트를 저장하는 방법을 제공하고, 상기 방법은: 상기 하나 이상의 정보의 비트를 인코딩하는 선형 핵산 분자를 합성하는 단계, 상기 선형 핵산 분자는: 상기 하나 이상의 정보의 비트를 인코딩하는 핵산 서열, 5' 어댑터 시퀀스, 3' 어댑터 시퀀스, 선택적인 하나 이상의 추가 기능적 서열을 포함하고, 상기 선형 핵산 분자로부터 원형 핵산 분자를 생성하는 단계, 상기 원형 핵산 분자를 증폭하여 상기 원형 핵산 분자의 하나 이상의 사본을 포함하는 증폭된 핵산 분자를 생성하기 위해 상기 원형 핵산 분자를 증폭하는 단계; 상기 증폭된 핵산 분자를 기판 상에 배치하는 단계를 포함한다. 일부 실시예에서, 상기 기판은 패터닝된다. 일부 실시예에서, 상기 기판은 패터닝 되지 않는다. 일부 실시예에서, 상기 보존하는 단계는 리오필라이제이션 또는 동결건조하는 단계를 포함한다. 일부 실시예에서, 상기 방법은 상기 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 하나 이상의 정보의 비트를 검색하는 단계를 더 포함한다. 일부 실시예에서, 상기 하나 이상의 정보의 비트를 검색하는 단계는 핵산 식별 반응을 포함한다. 일부 실시예에서, 상기 방법은 복구된 하나 이상의 정보의 비트에 오류 정정을 적용하는 단계를 더 포함한다. 일부 실시예에서, 상기 오류 정정은 리드-솔로몬 부호(Reed-Solomon code)를 사용하는 단계를 포함한다. 일부 실시예에서, 상기 정보의 비트는 이진 비트를 포함한다. 일부 실시예에서, 상기 정보의 비트는 이진 비트를 포함하고, (a) 단계는 정보의 상기 이진 비트를 정보의 제4기 비트로 전사하는 단계를 포함한다. 일부 실시예에서, 상기 5' 어댑터 서열, 3' 어댑터 서열, 또는 둘 모두는 바코드 부호 서열을 포함한다. 일부 실시예에서, 상기 하나 이상의 기능적 서열은 바코드 부호 서열, 태그 서열, 범용 프라이머 서열, 고유 식별자 서열, 또는 추가 어댑터 서열로 이루어진 그룹으로부터 선택된다. 일부 실시예에서, 상기 원형 핵산 분자는 상기 5' 어댑터 및 상기 3' 어댑터를 결합함으로써 생성된다. 일부 실시예에서, 상기 원형 핵산 분자는 롤링 서클 반응에 의해 증폭된다. 일부 실시예에서, 상기 증폭된 핵산 분자는 핵산 연쇄체(concatemer)다. 일부 실시예에서, 상기 증폭된 핵산 분자는 밀집하게 배치되고, 상기 기판이 광학 스캐닝 시스템을 사용하여 이미징되는 경우, 상기 핵산 분자 또는 이의 유도체로부터 생성된 신호의 파장은 상기 광학 스캐닝 시스템의 픽셀 크기의 2 배보다 크다. 일부 실시예에서, 상기 기판은 실리콘을 포함한다. 일부 실시예에서, 상기 기판은 유리를 포함한다. 선행하는 실시예들 중 어느 하나의 방법에 있어서, 상기 어레이는 제1 및 제2 유리 기판을 포함한다. 선행하는 실시예 중 어느 하나에 있어서, 방법은 선행하는 실시예들 중 어느 하나의 방법을 구현하도록 프로그래밍 된 컴퓨터 시스템에 의해 자동화된다.

[0010] 본 명세서에서 기술된 본 발명의 다른 태양은 컴퓨터 시스템을 제공하고, 상기 컴퓨터 시스템은 선행하는 실시예 중 어느 하나의 방법을 구현하도록 프로그래밍 된 컴퓨터 시스템이다.

[0011] 본 명세서에 기재된 본 발명의 다른 태양은 복수의 핵산 서열을 포함하는 -상기 복수의 핵산 서열의 적어도 일부는 적어도 1GB의 데이터를 인코딩하고, 상기 핵산 분자는 상기 핵산 분자의 1% 이하가 1년의 기간에 걸쳐 분해되도록 하는 안정성을 갖음-, 핵산 분자를 제공한다. 선행하는 실시예의 핵산분자는 복수의 헤더 서열을 더 포함하고, 상기 복수의 헤더 서열의 헤더 서열은 상기 핵산 서열의 적어도 일부의 염기서열분석을 허용하여 상기 1GB의 데이터를 검색하도록 구성된다.

[0012] 본 명세서에 기재된 본 발명의 다른 태양은 (a) 핵산 서열에서 상기 데이터를 인코딩하는 단계; (b) 상기 핵산 서열을 포함하는 하나 이상의 핵산 분자를 생성하는 단계; 및 (c) 상기 하나 이상의 핵산 분자를 기판 상에 배치된 어레이에 저장하는 단계를 포함한다. 일부 실시예에서, 상기 하나 이상의 핵산 분자는 원형이다. 일부 실시예에서, (b) 단계는 상기 핵산 서열의 적어도 일부를 포함하는 하나 이상의 원형 핵산 분자를 생성하는 단계를 및 롤링 서클 복제에 의해 상기 하나 이상의 원형 핵산 분자를 증폭하여 개별 핵산 분자의 하나 이상의 연결된 사본을 생성하는 단계를 포함한다. 일부 실시예에서, (b) 단계는 상기 핵산 서열, 제1 어댑터 서열, 및 제2 어댑터 서열을 포함하는 하나 이상의 선형 핵산 분자를 생성하는 단계 -상기 제1 및 상기 제2 어댑터 서열은 하나 이상의 원형 핵산 분자의 형성을 가능함-; 및 상기 하나 이상의 원형 핵산 분자를 증폭하는 단계를 포함한다. 일부 실시예에서, 상기 선형 핵산 분자는 하나 이상의 기능적 서열을 포함한다. 일부 실시예에서, 하나 이상의 연결된 핵산 분자는 롤링 서클 복제에 의해 증폭된다. 일부 실시예에서, (c) 단계는 상기 기판 상에 핵산 분자의 상기 연결된 카피를 배치하는 단계를 포함한다. 일부 실시예에서, 상기 하나 이상의 연결된 핵산 분자는 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다. 일부 실시예에

서, 상기 방법은 상기 기판을 보존하는 단계를 더 포함한다. 일부 실시예에서, 상기 보존하는 단계는 리오필라이제이션 또는 동결건조하는 단계를 포함한다. 일부 실시예에서, 상기 기판은 실리콘을 포함한다. 일부 실시예에서, 상기 기판은 유리를 포함한다. 일부 실시예에서, 상기 기판은 2개의 유리 조각을 포함한다. 일부 실시예에서, 상기 방법은 상기 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 데이터를 검색하는 단계를 더 포함한다.

[0013] 본 명세서에 기재된 다른 태양은 핵산 분자를 기판에 배치하는 단계를 포함하고, 상기 핵산 분자는 상기 데이터를 인코딩하는, 데이터를 저장하는 방법을 제공한다. 일부 실시예에서, 상기 핵산 분자는 핵산 연쇄체를 포함한다. 일부 실시예에서, 상기 연쇄체 분자는 제1 및 제2 원형 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다. 일부 실시예에서, 상기 기판은 실리콘을 포함한다. 일부 실시예에서, 상기 기판은 유리를 포함한다. 일부 실시예에서, 상기 기판은 2개의 유리 조각을 포함한다. 일부 실시예에서, 상기 데이터는 염기서열분석하는 단계 이전에 원형화 또는 증폭 없이 핵산 분자로부터 검색된다.

[0014] 본 명세서에 기재된 다른 태양은 하나 이상의 정보의 비트를 저장하는 방법을 제공하고, 상기 방법은: 복수의 뉴클레오티드에서 상기 하나 이상의 정보의 비트를 인코딩하는 단계; 상기 다수의 뉴클레오티드를 하나 이상의 프라이머에 결합하는 단계; 상기 복수의 뉴클레오티드를 약 300 내지 약 1,000개의 뉴클레오티드의 범위로 합성하는 단계; 상기 복수의 뉴클레오티드를 원형화하는 단계; 및, 상기 복수의 뉴클레오티드를 기판 상에 배치하는 단계를 포함한다.

[0015] 본 명세서에 기재된 다른 태양은 하나 이상의 정보의 비트를 저장하는 방법을 제공하고, 상기 방법은: 상기 하나 이상의 정보의 비트를 인코딩하는 선형 핵산 분자를 합성하는 단계 -상기 선형 핵산 분자는: 핵산 서열 상기 하나 이상의 정보의 비트, 5' 어댑터 서열, 3' 어댑터 서열, 및 선택적인 하나 이상의 추가 기능적 서열을 포함함-, 상기 선형 핵산 분자로부터 원형 핵산 분자를 생성하는 단계, 상기 원형 핵산 분자를 증폭하여 원형 핵산 분자의 하나 이상의 사본을 포함하는 제2 핵산 분자를 생성하는 단계, 어레이 상에 상기 제2 핵산 분자를 배치하는 단계를 포함한다. 일부 실시예에서, 상기 방법은 하나 이상의 기판 상에 상기 어레이를 배치하는 단계를 더 포함한다. 일부 실시예에서, 상기 방법은 상기 하나 이상의 기판을 보존하는 단계를 더 포함한다. 일부 실시예에서, 상기 보존하는 단계는 리오필라이제이션 또는 동결건조하는 단계를 포함한다. 일부 실시예에서, 상기 방법은 상기 검색하는 단계 이전에 증폭 없이 상기 하나 이상의 핵산 분자로부터 상기 하나 이상의 정보의 비트를 검색하는 단계를 더 포함한다. 일부 실시예에서, 상기 정보의 하나 이상의 비트는 염기서열분석 반응에 의해 상기 어레이로부터 복구된다. 일부 실시예에서, 상기 방법은 복구된 하나 이상의 정보의 비트에 오류 정정을 적용하는 단계를 더 포함한다. 일부 실시예에서, 상기 오류 정정은 리드-솔로몬 부호(Reed-Solomon code)를 이용하는 단계를 포함한다. 일부 실시예에서, 상기 정보의 하나 이상의 비트는 염기서열분석하는 단계 이전에 증폭 복제 반응 없이 상기 어레이로부터 검색된다. 일부 실시예에서, 상기 정보의 비트는 이진 비트를 포함한다. 일부 실시예에서, 상기 정보의 비트는 이진 비트를 포함하고 (a) 단계는 정보의 상기 이진 비트를 정보의 제4기 비트로 전사하는 단계를 포함한다. 일부 실시예에서, 상기 어댑터 서열은 바코드 부호 서열을 포함한다. 일부 실시예에서, 상기 하나 이상의 기능적 서열은 바코드 부호 서열, 태그 서열, 범용 프라이머 서열, 고유 식별자 서열, 또는 추가 어댑터 서열로 이루어진 그룹으로부터 선택된다. 일부 실시예에서, 상기 원형 핵산 분자는 상기 5' 어댑터 및 상기 3' 어댑터를 결합함으로써 생성된다. 일부 실시예에서, 상기 원형 핵산 분자는 롤링 서클 PCR 반응에 의해 증폭된다. 일부 실시예에서, 상기 제2 핵산 분자는 핵산 연쇄체다. 일부 실시예에서, 상기 제2 핵산 분자는 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다. 일부 실시예에서, 상기 어레이는 실리콘화된 기판을 포함한다. 일부 실시예에서, 상기 어레이는 유리 기판을 포함한다. 일부 실시예에서, 상기 어레이는 제1 및 제2 유리 기판을 포함한다. 일부 실시예에서, 상기 방법은 선행 청구항 중 어느 한 항의 방법을 구현하도록 프로그래밍된 컴퓨터 시스템에 의해 자동화된다.

[0016] 본 명세서에 설명된 다른 태양은 컴퓨터 시스템을 제공하고, 상기 컴퓨터 시스템은 본 명세서에 설명된 방법을 구현하도록 프로그래밍된다.

[0017] 본 명세서에 기재된 다른 태양은 복수의 핵산 분자는 적어도 일부가 1기가바이트(gigabytes; GB) 이상의 데이터를 인코딩하는 핵산 서열을 포함하고, 상기 핵산 분자는 상기 핵산 서열의 1% 이하가 1년의 기간에 걸쳐 분해되도록 하는 안정성을 갖는다. 일부 실시예에서, 핵산 분자는 원형이다. 일부 실시예에서, 핵산 분자는 복수의 헤더 서열을 더 포함하고, 상기 복수의 헤더 서열의 헤더 서열은 상기 핵산 서열의 상기 적어도 일부의 염기서열 분석을 허용하여 상기 1GB의 데이터를 검색하도록 구성된다.

[0018] 본 명세서에 기재된 다른 태양은 (a) 핵산 서열의 데이터를 인코딩하는 단계; (b) 핵산 서열을 포함하는 핵산

분자를 생성하는 단계; 및 (c) 어레이 상에 핵산 분자를 저장하는 단계를 포함한다. 일부 실시예에서, 핵산 분자는 원형이다. 일부 실시예에서, 핵산 분자는 핵산 연쇄체다. 일부 실시예에서, (b) 단계는 핵산 서열의 적어도 일부를 포함하는 선형 핵산 분자를 생성하는 단계, 및 선형 핵산 분자의 말단을 서로 연결하여 원형 핵산 분자를 생성하는 단계를 포함한다. 다른 실시예에 (b) 단계는 (i) 선형 핵산 분자, 제1 어댑터 서열, 및 제2 어댑터 서열을 포함하는 선형 핵산 분자를 생성하는 단계 -상기 제1 및 제2 어댑터 서열은 원형 핵산 분자의 형성을 가능하게 함-; 및 (ii) 원형 핵산 분자를 증폭하여 핵산 연쇄체를 생성하는 단계를 포함한다. 일부 실시예에서, 선형 핵산 분자는 기능적 서열을 포함한다. 일부 실시예에서, 선형 핵산 분자는 복수의 기능적 서열을 포함한다.

[0019] 일부 실시예에서, 핵산 연쇄체는 롤링 서클 복제에 의해 생성된다. 일부 실시예에서, (c) 단계는 핵산 분자를 기관 상에 배치하는 단계를 포함한다. 일부 실시예에서, 핵산 분자는 밀집하게 배치되고 2개 이상의 핵산 분자 사이의 평균 거리는  $\lambda/(2*NA)$ 의 측정치 미만이다. 일부 실시예에서 어레이는 실리콘 기관을 포함한다. 일부 실시예에서 어레이는 유리 기관을 포함한다.

[0020] 일부 실시예에서, 데이터는 염기서열분석하는 단계 이전에 증합효소 연쇄 반응 증폭 없이 핵산 분자로부터 검색된다.

[0021] 다른 태양에서, 핵산 분자를 기관에 고정화하거나 배치하는 단계를 포함하는, 데이터를 저장하는 방법이 개시되고, 상기 핵산 분자는 데이터를 인코딩한다. 일부 실시예에서, 핵산 분자는 핵산 연쇄체를 포함한다. 일부 실시예에서 핵산 분자는 고정화되거나 밀집하게 배치되고 상기 제1 및 제2 핵산 분자 사이의 평균 거리는  $\lambda/(2*NA)$ 의 측정치 미만이다. 일부 실시예에서 기관은 실리콘을 포함한다. 일부 실시예에서 기관은 유리를 포함한다. 일부 실시예에서 데이터는 염기서열분석하는 단계 이전에 증폭 없이 핵산 분자로부터 검색된다.

[0022] 다른 태양에서, 하나 이상의 정보의 비트를 저장하는 방법이 개시되고, 상기 방법은 (a) 복수의 뉴클레오티드에서 하나 이상의 정보의 비트를 인코딩하는 단계, (b) 상기 복수의 뉴클레오티드를 하나 이상의 프라이머에 결합하는 단계, (c) 약 300 내지 약 1,000개의 뉴클레오티드 범위로 상기 복수의 뉴클레오티드를 합성하는 단계, (d) 상기 복수의 뉴클레오티드를 원형화하는 단계, (e) 상기 복수의 뉴클레오티드를 기관 상에 배치하는 단계를 포함한다.

[0023] 다른 태양에서, 하나 이상의 정보의 비트를 저장하는 방법이 개시되고, 상기 방법은 (a) 상기 하나 이상의 정보의 비트를 인코딩하는 선형 핵산 분자를 합성하는 단계 -상기 선형 핵산 분자는 (i) 데이터를 인코딩하는 핵산 서열, (ii) 5' 어댑터 서열, (iii) 3' 어댑터 서열, 및 (iv) 선택적인 하나 이상의 추가 기능적 서열을 포함함-, (b) 상기 선형 핵산 분자로부터 원형 핵산 분자를 생성하는 단계, (c) 상기 원형 핵산 분자의 하나를 초과하는 사본을 포함하는 제2 핵산 분자를 생성하기 위해 상기 원형 핵산 분자의 증폭하는 단계, (d) 패터닝되거나 패터닝되지 않은 어레이 상에 제2 핵산 분자를 고정화 또는 배치하는 단계를 포함한다.

[0024] 일부 실시예에서 정보는 염기서열분석 반응에 의해 어레이로부터 복구된다. 일부 실시예에서, 정보를 복구하는 단계는 복구된 하나 이상의 정보의 비트에 오류 정정을 적용하는 단계를 더 포함한다. 일부 실시예에서, 오류 정정은 리드 솔로몬 부호를 사용하는 단계를 포함한다. 일부 실시예에서 정보는 염기서열분석하는 단계 이전에 증폭 복제 반응 없이 어레이로부터 검색된다. 일부 실시예에서, 정보의 비트는 이진 비트를 포함한다.

[0025] 일부 실시예에서 정보의 비트는 이진 비트를 포함하고 (a) 단계는 정보의 이진 비트를 정보의 제4기 비트로 전사하는 단계를 포함한다. 일부 실시예에서 어댑터 서열은 바코드 부호 서열을 포함한다. 하나 이상의 기능적 서열은 바코드 부호 서열, 태그 서열, 범용 프라이머 서열, 고유 식별자 서열, 또는 추가 어댑터 서열로 이루어진 그룹으로부터 선택된다. 일부 실시예에서, 원형 핵산 분자는 5' 어댑터와 3' 어댑터를 연결함으로써 생성된다. 일부 실시예에서, 원형 핵산 분자는 롤링 서클 반응에 의해 증폭된다. 일부 실시예에서, 제2 핵산 분자는 핵산 연쇄체다. 일부 실시예에서, 제2 핵산 분자는 고정화되거나 밀도로 기관 상에 배치되고, 상기 2개 이상의 핵산 분자 사이의 평균 거리는  $\lambda/(2*NA)$ 의 측정치 미만이다.

[0026] 일부 실시예에서 어레이는 실리콘화된 기관을 포함한다. 일부 실시예에서 어레이는 유리 기관을 포함한다. 일부 실시예에서 어레이는 제1 및 제2 유리 기관을 포함한다.

[0027] 본 발명의 다른 태양은 하나 이상의 컴퓨터 프로세서에 의한 실행 시, 상술한 또는 본 명세서 다른 곳의 임의의 방법을 구현하는, 기계 실행 가능 부호를 포함하는 비일시적 컴퓨터 판독 가능 매체를 제공한다.

[0028] 본 발명의 다른 태양은 하나 이상의 컴퓨터 프로세서 및 이에 연결된 컴퓨터 메모리를 포함하는 시스템을 제공한다. 컴퓨터 메모리는 하나 이상의 컴퓨터 프로세서에 의해 실행되는 경우 상술한 또는 본 명세서 다른 곳의

임의의 방법을 구현하는 기계 실행 가능 부호를 포함한다.

[0029] 본 발명의 추가적인 태양 및 이점은 다음의 상세한 설명으로부터 본 발명의 기술분야에 속한 통상의 기술자에게 용이하게 명백해질 것이고, 본 발명의 예시적인 실시예만이 도시되고 설명된다. 실현될 바와 같이, 본 발명은 기타 상이한 실시예가 가능하고, 이의 여러 세부사항은 본 발명으로부터 벗어남이 없이 다양한 자명한 태양에서 수정될 수 있다. 따라서, 도면 및 설명은 본질적으로 예시적인 것으로 간주되어야 하며 제한적인 것으로 간주되어서는 안 된다.

**도면의 간단한 설명**

[0030] 본 발명의 신규한 특징은 첨부된 청구범위에 구체적으로 설명되어 있다. 본 발명의 특징 및 이점에 대한 더 나은 이해는 본 발명의 원리가 활용되는 예시적인 실시예를 설명하는 다음의 상세한 설명과 첨부 도면(또한 본 명세서의 "도면" 및 "도")을 참조하여 얻을 수 있고, 그 중:

도 1은 핵산 분자의 정보 또는 데이터 비트를 인코딩하고 핵산 분자를 어레이에 배치하기 위한 개략도를 묘사한다. 이후, 어레이는 기관 위에 배치되고 장기 저장을 위해 저장되거나, 염기서열분석 되거나, 저장된 후 염기서열분석 된다.

도 2는 본 명세서에 설명된 시스템 및 방법을 자동화하기 위해 컴퓨터 시스템을 활용하기 위한 개략도를 묘사한다.

**발명을 실시하기 위한 구체적인 내용**

[0031] 본 명세서에 언급된 모든 간행물, 특허 및 특허 출원은 각각의 개별 간행물, 특허 또는 특허 출원은 마치 각각의 개별 간행물, 특허 또는 특허출원이 각각 참조에 의해 포함되는 것으로 개별적으로 명시되는 경우, 본 명세서에 참조로 포함되는 것이다. 참조로 포함된 간행물 및 특허 또는 특허 출원이 본 명세서에 포함된 개시 내용에 반하는 경우, 명세서는 그러한 모순되는 자료를 대체 및/또는 우선하도록 의도된다.

[0032] 본 발명의 다양한 실시예들이 본 명세서에 도시되고 설명되었으나, 그러한 실시예들은 본 발명의 기술 분야에 속한 통상의 기술자에게 예시로서 제공된다는 것이 명백할 것이다. 많은 변형, 변경 및 대체는 본 발명을 벗어남이 없이 본 발명의 기술 분야에 속한 통상의 기술자에게 일어날 수 있다. 본 명세서에 설명된 본 발명의 실시예에 대한 다양한 대안들이 사용될 수 있음을 이해해야 한다.

[0033] 본 명세서에서 사용된, 용어 "연쇄체(concatemer)"는 원형 핵산 분자의 사본을 의미한다. 연쇄체는 선형 핵산 분자의 말단이 원형 핵산 분자를 달성하기 위해 연결된 후 롤링 서클 복제(rolling circle amplification)에 의해 증폭되는 원형 핵산 분자로부터 생성될 수 있다. 연쇄체는 전체 분자에 걸쳐 반복되는 핵산의 단일 서열을 포함할 수 있거나, 연쇄체는 각각의 별개의 서열 또는 반복되는 서열의 세트가 어댑터 서열 또는 영역에 의해 분리되는 핵산 서열의 상이한 서열을 포함할 수 있다.

[0034] 본 명세서에서 사용되는, "염기서열분석을 위한 기기"란 핵산 분자 염기서열분석 분야의 통상의 기술자에게 익숙한 하드웨어, 소프트웨어, 시약, 이미징 모듈 및/또는 이들의 임의의 조합을 포함하는 기기를 의미한다.

[0035] 본 명세서에 사용된 "분석물"은 분석에 적합한 임의의 하나 이상의 분자를 의미한다. 핵산 분자, 단백질, 펩티드 등을 포함하나, 이에 제한되는 것은 아니다. 본 명세서에 기재된 개시 내용 전반에 걸쳐, 용어 "분석물(들)"은 본 발명의 범위를 변경함 없이 "핵산(들)" 및/또는 "핵산 분자(들)" 및/또는 "원형 핵산 분자(들)" 및/또는 연쇄체와 상호교환 가능하게 사용될 수 있다.

[0036] 본 명세서에서 사용된, "헤더 서열(들)"은 별개의 염기서열분석 프라이머로 주소 지정이 가능한 공지된 서열을 의미한다.

[0037] 용어 "적어도", "초과" 또는 "이상"은 일련의 2개 이상의 숫자 값에서 첫 번째 숫자 값 앞에 올 때마다 용어 "적어도", "초과" 또는 "이상"은 일련의 숫자 값 있는 각 숫자 값에 적용된다. 예를 들어, 1, 2 또는 3이상은 1 이상, 2이상, 3이상인 것이다.

[0038] 용어 "보다 크지 않음", "미만" 또는 "이하"는 일련의 두 개 이상의 숫자 값에서 첫 번째 숫자 값 앞에 올 때마다, 용어 "보다 크지 않음", "미만" 또는 "이하"는 일련의 숫자 값에 있는 각 숫자 값에 적용된다. 예를 들어, 3, 2 또는 1이하는 3이하이거나, 2이하 이거나, 1이하이다.

[0039] 일 경우, 방법은 (a) 핵산 서열의 데이터를 인코딩하는 단계; (b) 핵산 서열을 포함하는 핵산 분자를 생성하는

단계; 및 (c) 정렬되는 또는 정렬되지 않은 어레이에 핵산 분자 분석물을 저장하는 단계를 포함한다. 일 예시에서, 핵산 분자는 원형이다. 일 예시에서, 핵산 분자는 핵산 연쇄체다. 일 예시에서, (b) 단계는 핵산 서열의 적어도 일부를 포함하는 선형 핵산 분자를 생성하는 단계, 및 선형 핵산 분자의 말단을 서로 연결하여 원형 핵산 분자를 생성하는 단계를 포함한다. 다른 예시에서 (b) 단계는 (i) 선형 핵산 분자, 제1 어댑터 서열, 및 제2 어댑터 서열을 포함하는 선형 핵산 분자를 생성하는 단계를 포함하며, 상기 제1 및 제2 어댑터 서열은 원형 핵산 분자의 형성을 가능하게 하고, (ii) 원형 핵산 분자를 증폭하여 핵산 연쇄체를 생성하는 단계를 포함한다. 일부 예시에서, 선형 핵산 분자는 기능적 서열을 포함한다. 일부 예시에서, 선형 핵산 분자는 복수의 기능적 서열을 포함한다.

[0040] 일 예시에서, 핵산 연쇄체는 롤링 서클 복제에 의해 생성된다. 일 예시에서, (c) 단계는 기관에 분석물 핵산 분자를 배치하는 것을 포함한다. 일부 예시에서, 분석물은 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다. 일부 예시에서, 어레이는 실리콘 기관을 포함한다. 일부 예시에서, 어레이는 유리 기관을 포함한다.

[0041] 일 예시에서, 데이터는 염기서열분석 이전 증폭 없이 핵산 분자에서 검색된다.

[0042] 일 경우, 핵산 분자를 기관에 고정화하거나 배치하는 단계를 포함하는 데이터 저장 방법이 개시되고, 상기 핵산 분자는 데이터를 인코딩한다. 일 예시에서, 핵산 분자는 핵산 연쇄체를 포함한다. 일 예시에서, 원형 핵산 분자는 고정화되거나 밀도로 배치되며, 상기 제1 와 제2 원형 핵산 분자 사이의 평균 거리는  $\lambda/(2*NA)$ 의 측정치 미만이다. 일부 예시에서, 기관은 실리콘을 포함한다. 일부 예시에서, 기관은 유리를 포함한다. 일 예시에서 데이터는 염기서열분석 이전에 중합효소 연쇄 반응(polymerase chain reaction; PCR) 증폭 없이 핵산 분자에서 검색된다.

[0043] 일 경우, 방법은 정보의 하나 이상의 비트를 저장하는 단계를 포함하고, 방법은 (a) 복수의 뉴클레오티드에서 정보의 하나 이상의 비트를 인코딩하는 단계, (b) 복수의 뉴클레오티드를 하나 이상의 프라이머에 연결하는 단계, (c) 복수의 뉴클레오티드를 약 300 내지 약 1,000개의 뉴클레오티드 범위로 합성하는 단계, (d) 복수의 분석물을 원형화(또는 원형화하지 않음)하는 단계; 및 (e) 복수의 분석물을 기관 상에 배치하는 단계를 포함한다.

[0044] 네 번째 경우에서, 방법은 하나 이상의 정보의 비트를 저장하는 단계를 포함하고, 상기 방법은 (a) 하나 이상의 정보의 비트를 인코딩하는 선형 핵산 분자를 합성하는 단계, 상기, 선형 핵산 분자는 (i) 데이터를 암호화하는 핵산 서열, (ii) 5' 어댑터 서열, (iii) 3' 어댑터 서열, 및 (iv) 선택적인 하나 이상의 추가 기능 서열을 포함하고, (b) 선형 핵산 분자로부터 원형 핵산 분자를 생성하는 단계, (c) 원형 핵산 분자를 증폭하여 원형 핵산 분자의 하나 이상의 사본을 포함하는 분석물을 생성하는 단계, (d) 어레이 상에 분석물을 고정화시키거나 배치하는 단계를 포함한다.

[0045] 일 예시에서 정보는 염기서열분석 반응에 의해 어레이에서 복구된다. 일 예시에서, 정보를 복구하는 단계는 복구된 하나 이상의 정보의 비트에 오류 정정을 적용하는 단계를 더 포함한다. 일부 예시에서 오류 정정은 리드 솔로몬 부호(Reed-Solomon code)를 사용하는 단계를 포함한다. 일 예시에서 정보는 염기서열분석 이전에 증폭 복제 반응 없이 어레이에서 검색된다.

[0046] 일 예시에서, 정보의 비트는 이진 비트를 포함한다. 일 예시에서 정보의 비트는 이진 비트를 포함하고 (a) 단계는 정보의 이진 비트를 정보의 제4기(quaternary) 비트로 전사하는 단계를 포함한다. 일 예시에서 어댑터 서열은 바코드 부호 서열을 포함한다. 하나 이상의 기능적 서열은 바코드 부호 서열, 태그 서열, 범용 프라이머 서열, 고유 식별자 서열, 또는 추가 어댑터 서열로 이루어진 그룹으로부터 선택된다. 일 예시에서, 원형 핵산 분자는 5' 어댑터와 3' 어댑터를 결합함으로써(ligating) 생성된다. 일 예시에서, 원형 핵산 분자는 롤링 서클 PCR 반응에 의해 증폭된다. 일 예시에서, 제2 핵산 분자는 핵산 연쇄체다. 일 예시에서, 제2 핵산 분자는 2개 이상의 핵산 분자 사이의 평균 거리가  $\lambda/(2*NA)$ 의 측정치 미만인 밀도로 배치된다.

[0047] 일 예시에서, 어레이는 실리콘화 된 기관을 포함한다. 일 예시에서 어레이는 유리 기관을 포함한다. 일 예시에서 어레이는 제1 및 제2 유리 기관을 포함한다.

[0048] 염기서열분석 기술에는 일루미나(Illumina) 및 컴플리트 제노믹스(Complete Genomics)와 같은 회사에서 개발한 이미지 기반 시스템과 아이온 토렌트(Ion Torrent) 및 옥스포드 나노포어(Oxford Nanopore)와 같은 회사에서 개발한 전기 기반 시스템이 포함된다. 이미지 기반 염기서열분석 시스템은 현재 모든 기존 염기서열분석 기술 중 가장 낮은 염기서열분석 비용을 가지고 있다. 이미지 기반 시스템은 높은 처리량의 이미징 광학 장치와 저렴한 소모품의 조합을 통해 저비용을 달성한다. 그러나, 종래 기술의 광학적 검출 시스템은 부분적으로는 광학 시스

템의 회절 한계로 인해, 약 1 마이크로에서 인접하여 분해 가능한 분자 사이의 최소 중심 간 거리(center-to-center spacing)를 갖는다. 일부 실시예에서, 순환 검출, 분석물의 정확한 위치 결정, 이미지화된 신호의 고도로 정확한 디컨볼루션을 위한 위치 정보의 사용을 사용하여 기존 생화학을 사용하여 이미지 기반 염기서열분석 시스템에 대해 상당히 낮은 비용을 달성하는 방법이 본 명세서에 설명되어 있다. 회절 한계 아래에서 작동하는 증가된 패킹 밀도를 수용한다.

**[0049] 장기 저장을 위해 기판에 핵산 분자 배치하는 단계**

**[0050]** 암호화된 핵산 분자에 대한 정보를 저장하고 장기 저장을 위해 핵산 분자를 처리하기 위한 시스템 및 방법이 본 명세서에 제공된다. 본 명세서에 기술된 시스템 및 방법은 핵산 분자가 상업적으로 실행 가능한 속도로 분해되거나 분해되지 않도록 핵산 분자를 보존하는 처리 기술에 관한 것이다.

**[0051]** 일부 실시예에서, 핵산 분자는 빠르고 정확한 검색을 보장하기 위해 저장된 정보 세그먼트 및 필요한 정보(예: 리드 솔로몬 부호 또는 중복성)를 포함하는 일련의 세그먼트 또는 단일 세그먼트로 처리된다. 핵산 분자에 대한 분절 길이는 정확한 합성(합성에 의한 염기서열분석 기술 또는 기타 염기서열분석 접근법에 의해)과 염기서열분석 기술 및 기기(들)에 의한 정확한 검색을 모두 보장하도록 선택된다. 일부 실시예에서, 정보 세그먼트는 합성 및 검색 모두에 적합한 크기를 갖는 50-75개의 염기의 범위 내에 있다. 일부 실시예에서, 정보 세그먼트의 길이는 약 30개 염기에서 약 140개 염기의 길이이다.

**[0052]** 일부 실시예에서, 정보 세그먼트의 길이는 약 30개 염기 내지 약 40개 염기, 약 30개 염기 내지 약 50개 염기, 약 30개 염기 내지 약 60개 염기, 약 30개 염기 내지 약 70개 염기, 약 30개 염기 내지 약 80개 염기, 약 30개 염기 내지 약 90개 염기, 약 30개 염기 내지 약 100개 염기, 약 30개 염기 내지 약 110개 염기, 약 30개 염기 내지 약 120개 염기, 약 30개 염기 내지 약 130개 염기, 약 30개 염기 내지 약 140개 염기, 약 40개 염기 내지 약 50개 염기, 약 40개 염기 내지 약 60개 염기, 약 40개 염기 내지 약 70개 염기, 약 40개 염기 내지 약 80개 염기, 약 40개 염기 내지 약 90개 염기, 약 40개 염기 내지 약 100개 염기, 약 40개 염기 내지 약 110개 염기, 약 40개 염기 내지 약 120개 염기, 약 40개 염기 내지 약 130개 염기, 약 40개 염기 내지 약 140개 염기, 약 50개 염기 내지 약 60개 염기, 약 50개 염기 내지 약 70개 염기, 약 50개 염기 내지 약 80개 염기, 약 50개 염기 내지 약 90개 염기, 약 50개 염기 내지 약 100개 염기, 약 50개 염기 내지 약 100개 염기, 약 50개 염기 내지 약 120개 염기, 약 50개 염기 내지 약 130개 염기, 약 50개 염기 내지 약 140개 염기, 약 60개 염기 내지 약 70개 염기, 약 60개 염기 내지 약 80개 염기, 약 60개 염기 내지 약 90개 염기, 약 60개 염기 내지 약 100개 염기, 약 60개 염기 내지 약 110개 염기, 약 60개 염기 내지 약 120개 염기, 약 60개 염기 내지 약 130개 염기, 약 60개 염기 내지 약 140개 염기, 약 70개 염기 내지 약 80개 염기, 약 70개 염기 내지 약 90개 염기, 약 70개 염기 내지 약 100개 염기, 약 70개 염기 내지 약 110개 염기, 약 70개 염기 내지 약 120개 염기, 약 70개 염기 내지 약 130개 염기, 약 70개 염기 내지 약 140개 염기, 약 80개 염기 내지 약 90개 염기, 약 80개 염기 내지 약 100개 염기, 약 80개 염기 내지 약 110개 염기, 약 80개 염기 내지 약 120개 염기, 약 80개 염기 내지 약 130개 염기, 약 80개 염기 내지 약 140개 염기, 약 90개 염기 내지 약 100개 염기, 약 90개 염기 내지 약 110개 염기, 약 90개 염기 내지 약 120개 염기, 약 90개 염기 내지 약 130개 염기, 약 90개 염기 내지 약 140개 염기, 약 100개 염기 내지 약 110개 염기, 약 100개 염기 내지 약 120개 염기, 약 100개 염기 내지 약 130개 염기, 약 100개 염기 내지 약 140개 염기, 약 110개 염기 내지 약 120개 염기, 약 110개 염기 내지 약 130개 염기, 약 110개 염기 내지 약 140개 염기, 약 120개 염기 내지 약 130개 염기, 약 120개 염기 내지 약 140개 염기, 또는 약 130개 염기 내지 약 140개 염기의 길이이다. 일부 실시예에서, 정보 세그먼트는 길이는 약 30개 염기, 약 40개 염기, 약 50개 염기, 약 60개 염기, 약 70개 염기, 약 80개 염기, 약 90개 염기, 약 100개 염기, 약 110개 염기, 약 120개 염기, 약 130개 염기, 또는 약 140개 염기의 길이이다. 일부 실시예에서, 정보 세그먼트는 적어도 약 30개 염기, 약 40개 염기, 약 50개 염기, 약 60개 염기, 약 70개 염기, 약 80개 염기, 약 90개 염기, 약 100개 염기, 110개 염기, 약 120개 염기, 또는 약 130개 염기의 길이이다. 일부 실시예에서, 정보 세그먼트의 길이는 최대 약 40개 염기, 약 50개 염기, 약 60개 염기, 약 70개 염기, 약 80개 염기, 약 90개 염기, 약 100개 염기, 약 110개 염기, 약 120개 염기, 약 130개 염기, 또는 약 140개 염기의 길이이다.

**[0053]** 일부 실시예에서, 핵산 분자는 예를 들어, 염기서열분석 및 검출을 위해 적절한 기판에 대한 부착 및 롤링 서클 복제에 의해, 원형 핵산 분자(예: CAT 또는 연쇄체)로의 후속 전환을 위한 적절한 어댑터에 부착된다(US20150330974 또는 US20160201119 및/또는 US10378053에 따라). 공통 서열은 핵산 분자의 염기서열분석 및 원형화 프라이밍에 적합한 서열을 최소한으로 포함한다. 일부 실시예에서, 원형화된 핵산 분자의 전체 길이는 300 - 1,000 개의 염기의 범위 내에 있다. 일부 실시예에서, 원형화된 핵산 분자의 길이는 상이한 염기서열분석 프라이머(본 명세서에서 "헤더 서열"로 지칭됨)로 주소 지정이 가능한 서열에 의해 분리된, 동일한 서클 내 다

수의 정보 세그먼트를 추가함으로써 달성될 수 있다. 일부 실시예에서, 원형화된 핵산 분자의 길이는 적절한 크기를 갖도록 염기서열분석되지 않을 스테퍼(stuffer) 단편을 도입함으로써 달성될 수 있다.

[0054] 일부 실시예에서, 원형화된 핵산 분자의 길이는 약 200개 염기 내지 약 1,200개 염기이다. 일부 실시예에서, 원형화된 핵산 분자의 길이는 약 200개 염기 내지 약 300개 염기, 약 200개 염기 내지 약 400개 염기, 약 200개 염기 내지 약 500개 염기, 약 200개 염기 내지 약 600개 염기, 약 200개 염기 내지 약 700개 염기, 약 200개 염기 내지 약 800개 염기, 약 200개 염기 내지 약 900개 염기, 약 200개 염기 내지 약 1,000개 염기, 약 200개 염기 내지 약 1,100개 염기, 약 200개 염기 내지 약 1,200개 염기, 약 300개 염기 내지 약 400개 염기, 약 300개 염기 내지 약 500개 염기, 약 300개 염기 내지 약 600개 염기, 약 300개 염기 내지 약 700개 염기, 약 300개 염기 내지 약 800개 염기, 약 300개 염기 내지 약 900개 염기, 약 300개 염기 내지 약 1000개 염기, 약 300개 염기 내지 약 1,100개 염기, 약 300개 염기 내지 약 1,200개 염기, 약 400개 염기 내지 약 500개 염기, 약 400개 염기 내지 약 600개 염기, 약 400개 염기 내지 약 700개 염기, 약 400개 염기 내지 약 800개 염기, 약 400개 염기 내지 약 900개 염기, 약 400개 염기 내지 약 1,000개 염기, 약 400개 염기 내지 약 1,100개 염기, 약 400개 염기 내지 약 1,200개 염기, 약 500개 염기 내지 약 600개 염기, 약 500개 염기 내지 약 700개 염기, 약 500개 염기 내지 약 800개 염기, 약 500개 염기 내지 약 900개 염기, 약 500개 염기 내지 약 1,000개 염기, 약 500개 염기 내지 약 1,100개 염기, 약 500개 염기 내지 약 1,200개 염기, 약 600개 염기 내지 약 700개 염기, 약 600개 염기 내지 약 800개 염기, 약 600개 염기 내지 약 900개 염기, 약 600개 염기 내지 약 1,000개 염기, 약 600개 염기 내지 약 1,100개 염기, 약 600개 염기 내지 약 1,200개 염기, 약 700개 염기 내지 약 800개 염기, 약 700개 염기 내지 약 900개 염기, 약 700개의 염기 내지 약 1,000개 염기, 약 700개 염기 내지 약 1,100개 염기, 약 700개 염기 내지 약 1,200개 염기, 약 800개 염기 내지 약 900개 염기, 약 800개 염기 내지 약 1,000개 염기, 약 800개 염기 내지 약 1,100개 염기, 약 800개 염기 내지 약 1,200개 염기, 약 900개 염기 내지 약 1,000개 염기, 약 900개 염기 내지 약 1,100개 염기, 약 900개 염기 내지 약 1,200개 염기, 또는 약 1,100개 염기 내지 약 1,200개 염기이다.

[0055] 일부 실시예에서, 원형화된 핵산 분자의 길이는 약 200개 염기, 약 300개 염기, 약 400개 염기, 약 500개 염기, 약 600개 염기, 약 700개 염기, 약 800개 염기, 약 900개 염기, 약 1,000개 염기, 약 1,100개 염기, 또는 약 1,200개 염기이다. 일부 실시예에서, 원형화된 핵산 분자의 길이는 약 200개 염기, 약 300개 염기, 약 400개 염기, 약 500개 염기, 약 600개 염기, 약 700개 염기, 약 800개 염기, 약 900개 염기, 약 1,000개의 염기, 또는 약 1,100개의 염기이다. 일부 실시예에서, 원형화된 핵산 분자의 길이는 최대 약 300개 염기, 약 400개 염기, 약 500개 염기, 약 600개 염기, 약 700개 염기, 약 800개 염기, 약 900개 염기, 약 1,000개 염기, 약 1,100개 염기, 또는 약 1,200개 염기이다.

[0056] 일부 실시예에서, 원형 핵산 분자는 (염기서열분석용 칩과 같은) 기관 위에 배치된다. 일부 실시예에서, 하나 이상의 핵산 분자가 기관 상에 배치된 후, 기관은 장기간 저장을 위해 가공되어야 할 것이다. 일부 실시예에서, 상기 가공은 기관을 건조시키는 것을 포함한다. 일부 실시예에서, 상기 가공은 리오피라이제이션(lyophilization; 동결건조) 또는 크라이오데시케이션(cryodesiccation; 동결 건조)과 같은, 동결 건조를 포함한다. 리오피라이제이션은 제품을 동결하고, 압력을 낮추고, 승화에 의해 얼음을 제거하는 방법을 포함할 수 있는 저온 탈수 공정을 포함하는 동결건조 공정을 이용할 수 있다. 일부 실시예에서, 건조 공정 이전에, 원형 핵산 분자가 배치된 기관은 건조 공정 및 건조 공정으로부터의 회복되는 동안 안정성을 보장하기 위해 (후부처리로서) 처리된다. 일부 실시예에서, 상기 처리는 원형 핵산 분자를 안정화시키기 위해 기관의 표면을 예: BSA 또는 덱스트란 설페이트(Dextran Sulfate)로 코팅하는 것 뿐만 아니라, 동결 건조 동안 얼음 결정 형성 및 재수화동안 충격으로부터 기관을 안정화하고 보호하기 위해 당(예: 만니톨(mannitol), 수크로스(sucrose), 트레할로스(trehalose), 젓당(lactose), 맥아당(maltose), 포도당(glucose), 글리신(glycine), 글리세롤(glycerol) 등)과 같은, 적절한 부형제 및 적절한 완충제를 도입하는 것을 포함한다.

[0057] 일부 실시예에서, 핵산 분자의 증폭(예: 롤링 서클 복제)은 핵산 분자를 포함하는 기관(들)에 대한 장기 저장 이전에 발생한다. 일부 실시예에서, 핵산 분자의 증폭은 핵산 분자가 배치된 기관에서 발생한다. 일부 실시예에서, 증폭은 브릿지 증폭(bridge amplification)이다. 일부 실시예에서, 핵산 분자의 증폭(예: 롤링 서클 복제)은 핵산 분자를 기관에 배치하기 전에 발생한다. 일부 실시예에서, 증폭은 롤링 서클 복제다.

[0058] 일부 실시예에서, 원형 핵산 분자는 저장을 위해 복수의 슬라이드에 배치된다. 일부 실시예에서 슬라이드는 구별되는 복수의 레인 및/또는 트랙이 있다. 일부 실시예에서, 고유 헤더 시퀀스는 정보를 포함하는 특정 시퀀스에 대한 위치 정보를 식별하는 데 사용된다. 일부 실시예에서, 위치 정보는 주어진 정보의 세트를 저장하기 위

해 사용되는 모든 헤더 시퀀스에 대한 정보를 포함하는 카탈로그에서 발견된다. 일부 실시예에서, 최종 검색을 위해 설정된 정보는 저장을 위해 기관/슬라이드에 배치된 핵산 분자에 포함되는 반면, 핵산 분자의 복수의 사본들은 백업 정보로서 별도로 저장된다. 일부 실시예에서, 정보 저장 과정의 미래 준비에 더하여, 각 레인에 대응하는 핵산 분자는 별도로 건조되고 백업으로서 저장된다. 일부 실시예에서, 백업 핵산 분자는 가공되고 저장된 슬라이드에 있는 정보를 복구할 수 없는 경우에 적절하게 후속적으로 처리될 수 있다.

[0059] 일부 실시예에서, 보존된 핵산의 분해율은 연간 약 0.05 % 내지 연간 약 2 %이다. 일부 실시예에서, 보존된 핵산의 분해율은 연간 약 2 % 내지 연간 약 1 %, 연간 약 2 % 내지 연간 약 0.9 %, 연간 약 2 % 내지 연간 약 0.8 %, 연간 약 2 % 내지 연간 약 0.7 %, 연간 약 2 % 내지 연간 약 0.6 %, 연간 약 2 % 내지 연간 약 0.5 %, 연간 약 2 %에서 연간 약 0.4 %, 연간 약 2 % 내지 연간 약 0.3 %, 연간 약 2 % 내지 연간 약 0.2 %, 연간 약 2 % 내지 연간 약 0.1 %, 연간 약 2 % 내지 연간 약 0.05 %, 연간 약 1 % 내지 연간 약 0.9 %, 연간 약 1 % 내지 연간 약 0.8 %, 연간 약 1 % 내지 연간 약 0.7 %, 연간 약 1 % 내지 연간 약 0.6 %, 연간 약 1 % 내지 연간 약 0.5 %, 연간 약 1 % 내지 연간 약 0.4 %, 연간 약 1 % 내지 연간 약 0.3 %, 연간 약 1 % 내지 연간 약 0.2 %, 연간 약 1 % 내지 연간 약 0.1 %, 연간 약 1 % 내지 연간 약 0.05 %, 연간 약 0.9 % 내지 연간 약 0.8 %, 연간 약 0.9 % 내지 연간 약 0.7 %, 연간 약 0.9 % 내지 연간 약 0.6 %, 연간 약 0.9 % 내지 연간 약 0.5 %, 연간 약 0.9 % 내지 연간 약 0.4 %, 연간 약 0.9 % 내지 연간 약 0.3 %, 연간 약 0.9 % 내지 연간 약 0.2 %, 연간 약 0.9 % 내지 연간 약 0.1 %, 연간 약 0.9 % 내지 연간 약 0.05 %, 연간 약 0.8 % 내지 연간 약 0.7 %, 연간 약 0.8 % 내지 연간 약 0.6 %, 연간 약 0.8 % 내지 연간 약 0.5 %, 연간 약 0.8 % 내지 연간 약 0.4 %, 연간 약 0.8 % 내지 연간 약 0.3 %, 연간 약 0.8 % 내지 연간 약 0.2 %, 연간 약 0.8 % 내지 연간 약 0.1 %, 연간 약 0.8 % 내지 연간 약 0.05 %, 연간 약 0.7 % 내지 연간 약 0.6 %, 연간 약 0.7 % 내지 연간 약 0.5 %, 연간 약 0.7 % 내지 연간 약 0.4 %, 연간 약 0.7 % 내지 연간 약 0.3 %, 연간 약 0.7 % 내지 연간 약 0.2 %, 연간 약 0.7 % 내지 연간 약 0.1 %, 연간 약 0.7 % 내지 연간 약 0.05 %, 연간 약 0.6 % 내지 연간 약 0.5 %, 연간 약 0.6 % 내지 연간 약 0.4 %, 연간 약 0.6 % 내지 연간 약 0.3 %, 연간 약 0.6 % 내지 연간 약 0.2 %, 연간 약 0.6 % 내지 연간 약 0.1 %, 연간 약 0.6 % 내지 연간 약 0.05 %, 연간 약 0.5 % 내지 연간 약 0.4 %, 연간 약 0.5 % 내지 연간 약 0.3 %, 연간 약 0.5 % 내지 연간 약 0.2 %, 연간 약 0.5 % 내지 연간 약 0.1 %, 연간 약 0.5 % 내지 연간 약 0.05 %, 연간 약 0.4 % 내지 연간 약 0.3 %, 연간 약 0.4 % 내지 연간 약 0.2 %, 연간 약 0.4 % 내지 연간 약 0.1 %, 연간 약 0.4 % 내지 연간 약 0.05 %, 연간 약 0.3 % 내지 연간 약 0.2 %, 연간 약 0.3 % 내지 연간 약 0.1 %, 연간 약 0.3 % 내지 연간 약 0.05 %, 또는 연간 약 0.1 %에서 연간 약 0.05 %이다. 일부 실시예에서, 보존된 핵산의 분해율은 연간 약 2%, 연간 약 1%, 연간 약 0.9%, 연간 약 0.8%, 연간 약 0.7%, 연간 약 0.6%이다. , 연간 약 0.5%, 연간 약 0.4%, 연간 약 0.3%, 연간 약 0.2%, 연간 약 0.1%, 또는 연간 약 0.05%이다. 일부 실시예에서, 보존된 핵산의 분해율은 적어도 연간 약 2%, 연간 약 1%, 연간 약 0.9%, 연간 약 0.8%, 연간 약 0.7%, 연간 약 0.6%, 연간 약 0.5%, 연간 약 0.4%, 연간 약 0.3%, 연간 약 0.2%, 또는 연간 약 0.1%이다. 일부 실시예에서, 보존된 핵산의 분해율은 최대 연간 약 1%, 연간 약 0.9%, 연간 약 0.8%, 연간 약 0.7%, 연간 약 0.6%, 연간 약 0.5%, 연간 약 0.4%, 연간 약 0.3%, 연간 약 0.2%, 연간 약 0.1%, 또는 연간 약 0.05%이다.

[0060] 일부 실시예에서, 핵산 분자를 포함하는 기관은 하나 이상의 데이터 센터에 저장된다. 일부 실시예에서, 하나 이상의 데이터 센터는 기관을 포함하고 유지하도록 구성된 복수의 장착 가능한 랙을 포함한다. 일부 실시예에서, 하나 이상의 데이터 센터는 핵산 분자를 염기서열분석하기 위한 하나 이상의 기기를 포함한다(합성에 의한 염기서열분석 또는 다른 차세대 염기서열분석 기술 또는 기타 핵산 분자 염기서열분석 기술). 일부 실시예에서, 핵산 분자를 염기서열분석하기 위한 기기는 랙에 장착 가능하도록 구성된다. 일부 실시예에서, 하나 이상의 데이터 센터는 핵산 분자를 염기서열분석 하기 위한 기기로서의 전달 및 완전 자동화된 기관 저장을 지원하도록 구성된다.

[0061] 일부 실시예에서, 본 명세서에 설명된 시스템 및 방법은 저장된 정보를 검색의 지연 시간을 줄인다(전송에 대한 데이터 요청). 일부 실시예에서, 데이터 검색을 위한 시간은 약 1시간 내지 약 12시간으로 단축된다. 일부 실시예에서, 데이터 검색 시간은 약 1시간 내지 약 2시간, 약 1시간 내지 3시간, 약 1시간 내지 약 4시간, 약 1시간 내지 약 5시간, 약 1시간 내지 약 6시간, 약 1시간 내지 약 7시간, 약 1시간 내지 8시간, 약 1시간 내지 약 9시간, 약 1시간 내지 약 10시간, 약 1시간 내지 약 11시간, 약 1시간 내지 약 12시간, 약 2시간 내지 약 3시간, 약 2시간 내지 약 4시간, 약 2시간 내지 약 5시간, 약 2시간 내지 약 6시간, 약 2시간 내지 약 7시간, 약 2시간 내지 약 8시간, 약 2시간 내지 약 9시간, 약 2시간 내지 약 10시간, 약 2시간 내지 약 11시간, 약 2시간 내지 약 12시간, 약 3시간 내지 약 4시간, 약 3시간 내지 약 5시간, 약 3시간 내지 약 6시간, 약 3시간 내지 약 7시간, 약 3시간 내지 약 8시간, 약 3시간 내지 약 9시간, 약 3시간 내지 약 10시간, 약 3시간 내지 약 11시간, 약

3시간 내지 약 12시간, 약 4시간 내지 약 5시간, 약4시간 내지 약6시간, 약4시간 내지 약7시간, 약4시간 내지 약8시간, 약 4시간 내지 약9시간, 약4시간 내지 약10시간, 약4시간 내지 약11시간, 약4시간 내지 약12시간, 약5시간 내지 약6시간, 약5시간 내지 약7시간, 약 5시간 내지 약8시간, 약 5시간 내지 약 9시간, 약 5시간 내지 약 10시간, 약 5시간 내지 약 11시간, 약 5시간 내지 약 12시간, 약 6시간 내지 약 7시간, 약 6시간 내지 약 8시간, 약 6시간 내지 약 9시간, 약 6시간 내지 약 10시간, 약 6시간 내지 약 11시간, 약 6시간 내지 약 12시간, 약 7시간 내지 약 8시간, 약 7시간 내지 9시간, 약 7시간 내지 약 10시간, 약 7시간 내지 약 11시간, 약 7시간 내지 약 12시간, 약 8시간 내지 약 9시간, 약 8시간 내지 약 10시간, 약 8시간 내지 약 11시간, 약 8시간 내지 약 12시간, 약 9시간 내지 약 19시간, 약 9시간 내지 약 11시간, 약 9시간 내지 약 12시간, 약 10시간 내지 약 11시간, 약 10시간 내지 약 12시간 또는 약 11시간 내지 약 12시간이다. 일부 실시예에서, 데이터 검색을 위한 시간은 약 1시간, 약 2시간, 약 3시간, 약 4시간, 약 5시간, 약 6시간, 약 7시간, 약 8시간, 약 9시간, 약 10시간, 약 11시간, 또는 약 12시간으로 감소된다. 일부 실시예에서, 데이터 검색을 위한 시간은 적어도 약 1시간, 약 2시간, 약 3시간, 약 4시간, 약 5시간, 약 6시간, 약 7시간, 약 8시간, 약 9시간, 약 10시간, 또는 약 11시간으로 감소된다. 일부 실시예에서, 데이터 검색을 위한 시간은 최대 약 2시간, 약 3시간, 약 4시간, 약 5시간, 약 6시간, 약 7시간, 약 8시간, 약 9시간, 약 10시간, 약 11시간, 또는 약 12시간으로 감소된다.

[0062] **정보 검색**

[0063] 본 명세서에 설명된 데이터 저장 시스템 및 방법의 한 가지 이점은 핵산 분자 및 기관은 본 명세서에서 설명된 시스템 및 방법에 의해 처리(배치 및 보존)되는 경우, 저장된 데이터에 대한 검색은 샘플 준비(예: 증폭)를 거의 필요로 하지 않는다. 일부 실시예에서, 샘플 준비는 핵산을 기관에 배치하는 것을 포함한다. 일부 실시예에서, 샘플 준비는 핵산 분자의 증폭을 포함한다. 일부 실시예에서, 샘플 준비는 중합효소 연쇄 반응 증폭을 포함한다. 일부 실시예에서, 샘플 준비는 핵산 분자를 염기서열분석(합성에 의한 염기서열분석 또는 기타 차세대 염기서열분석 기술 또는 기타 핵산 분자 염기서열분석 기술)에 적합한 시약에 노출시키는 것을 포함한다. 본 명세서에서 기술된 바와 같이, 특정 관심의 정보를 인코딩하는 핵산 분자는 장기 저장 이전에 증폭된다. 따라서, 정보 검색이 필요한 경우, 저장되고, 증폭된 핵산 분자는 단지 재수화(장기 저장 기술이 동결건조를 포함하는 경우) 및 검색되기를 바라는 정보를 인코딩하는 시퀀스에 대응하는 헤더 서열(들)에 특정한 핵산 신장 반응(extension reaction) 프라이머와 접촉만이 필요하다.

[0064] 일부 실시예에서, 본 명세서에 기재된 시스템 및 방법을 이용하는 경우, 현재 핵산 분자 염기서열분석 시스템 및 방법(예: 일루미나(Illumina)®, 컴플리트 제노믹스(Complete Genomics)®, 비지아이(BGI)® 또는 다른 핵산 염기서열분석 회사에서 사용하는 최신 염기서열분석 시스템 및 방법)의 시약 요구 사항과 비교하는 경우, 염기서열분석에 적합한 시약 요구량이 약 1배 내지 약 12배로 감소된다. 일부 실시예에서, 본 명세서에 기재된 시스템 및 방법을 사용하는 경우, 염기서열분석에 적합한 시약의 요구사항은 약 1배 내지 약 2배, 약 1배 내지 약 3배, 약 1배 내지 약 4배, 약 1배 내지 약 5배, 약 1배 내지 약 6배, 약 1배 내지 약 7배, 약 1배 내지 약 8배, 약 1배 내지 약 9배, 약 1배 내지 약 10배, 약 1배 내지 약 11배, 약 1배 내지 약 12배, 약 2배 내지 약 3배, 약 2배 내지 약 4배, 약 2배 내지 약 5배, 약 2배 내지 약 6배, 약 2배 내지 약 7배, 약 2배 내지 약 8배, 약 2배 내지 약 9배, 약 2배 내지 약 10배, 약 2배 내지 약 11배, 약 2배 내지 약 12배, 약 3배 내지 약 4배, 약 3배 내지 약 5배, 약 3배 내지 약 6배, 약 3배 내지 약 7배, 약 3배 내지 약 8배, 약 3배 내지 약 9배, 약 3배 내지 약 10배, 약 3배 내지 약 11배, 약 3배 내지 약 12배, 약 4배 내지 약 5배, 약 4배 내지 약 6배, 약 4배 내지 약 7배, 약 4배 내지 약 8배, 약 4배 내지 약 9배, 약 4배 내지 약 10배, 약 4배 내지 약 11배, 약 4배 내지 약 12배, 약 5배 내지 약 6배, 약 5배 내지 약 7배, 약 5배 내지 약 8배, 약 5배 내지 약 9배, 약 5배 내지 약 10배, 약 5배 내지 약 11배, 약 5배 내지 약 12배, 약 6배 내지 약 7배, 약 6배 내지 약 8배, 약 6배 내지 약 9배, 약 6배 내지 약 10배, 약 6배 내지 약 11배, 약 6배 내지 약 12배, 약 7배 내지 약 8배, 약 7배 내지 약 9배, 약 7배 내지 약 10배, 약 7배 내지 약 11배, 약 7배 내지 약 12배, 약 8배 내지 약 9배, 약 8배 내지 약 10배, 약 8배 내지 약 11배, 약 8배 내지 약 12배, 약 9배 내지 약 10배, 약 9배 내지 약 11배, 약 9배 내지 약 12배, 약 10배 내지 약 11배, 약 10배 내지 약 12배, 또는 약 11배 내지 약 12배 까지 감소된다. 일부 실시예에서, 본 명세서에 기재된 시스템 및 방법을 사용하는 경우, 염기서열분석에 적합한 시약의 요구사항은 약 1배, 약 2배, 약 3배, 약 4배, 약 5배, 약 6배, 약 7배, 약 8배, 약 9배, 약 10배, 약 11배, 또는 약 12배 까지 감소된다. 일부 실시예에서, 본 명세서에 기재된 시스템 및 방법을 사용하는 경우, 염기서열분석에 적절한 시약의 요구사항은 적어도 약 1배, 약 2배, 약 3배, 약 4배, 약 5배, 약 6배, 약 7배, 약 8배, 약 9배, 약 10배, 또는 약 11배까지 감소된다. 일부 실시예에서, 본 명세서에 기재된 시스템 및 방법을 사용하는 경우, 염기서열분석에 적절한 시약의 요구사항은 최대 약 2배, 약 3배, 약 4배, 약 5배, 약 6배, 약 7배, 약 8배, 약 9배, 약 10

배, 약 11배, 또는 약 12배 까지 감소된다.

[0065] 일부 실시예에서, 핵산 분자 및/또는 기관의 재수화 이후 저장된 정보에 대한 검색 또는 판독이 가능하다. 일부 실시예에서, 저장된 정보에 대한 검색 또는 판독은 핵산 분자를 염기서열분석하고 검출하는 것을 포함한다 (US20150330974 또는 US20160201119 및/또는 US10378053에 따라).

[0066] 회절 한계 미만(예: 미만 =  $\lambda / 2 \cdot NA$ )으로 중심 간 거리로 표면에 고정 또는 배치된 분석물의 신호 이미징을 용이하게 하는 시스템 및 방법이 제공된다. 이러한 시스템 및 방법은 고급 이미징 시스템을 사용하여 고해상도 이미지를 생성하고, 순환 검출을 사용하여 높은 정확도로 기관 상의 분자의 위치 결정을 용이하게 하고 이미지의 디컨볼루션을 사용하여 높은 정확도로 밀집하게 패킹된 표면 상의 각 분자에 대한 신호 동일성을 획득한다. 이러한 방법 및 시스템은 고밀도로 패킹된 기관에서 합성에 의한 단일 분자 염기서열분석을 허용하여 매우 효율적이고 매우 높은 처리량의 폴리뉴클레오티드 서열 결정을 높은 정확도로 제공한다.

[0067] 데이터 저장 비용의 감소를 달성하기 위해, 본 명세서에서는 회절 한계 미만의 밀도로 기관 표면에 고정되거나 배치된 폴리뉴클레오티드의 신뢰할 수 있는 염기서열분석을 용이하게 하는 방법 및 시스템을 제공한다. 이러한 고밀도 어레이는 시약의 보다 효율적인 사용을 허용하고 단위 면적당 데이터의 양을 증가시킨다. 또한, 검출 신뢰도의 증가는 염기서열분석 및 검출에서 오류를 식별하고 정정하기 위해 합성되어야 하는 클론 카피의 수를 감소시켜, 시약 비용과 데이터 처리 비용을 더 절감한다.

[0068] **기관의 표면 상의 분석물의 고밀도 분포**

[0069] 1,000달러 계층에 사용된 샘플 유효 피치와 제안된 피치를 비교하면, 새로운 어레이의 밀도는 170배 더 높아, 100배 더 높은 밀도를 달성하는 기준을 충족한다. 단위 면적당 이미징 지점 별 사본의 수는 또한 이전에 기존의 플랫폼보다 적어도 100배 더 낮은 기준을 충족한다. 이는 시약 비용이 기준선보다 100배 더 비용에 있어 효율적인 임을 보장하는 데 도움이 된다.

[0070] **밀집하게 패킹된 단일 생체 분자 이미지 생성 및 회절 한계**

[0071] 이미징 플랫폼을 위한 증가된 분자 밀도에 대한 주요한 제약은 회절 한계이다. 광학 시스템의 회절 한계에 대한 방정식은 다음과 같다 :  $D = \lambda / 2 \cdot NA$  여기서 D는 회절 한계,  $\lambda$ 는 빛의 파장, NA는 광학 시스템의 개구수 (numerical aperture)이다. 일반적인 공기 이미징 시스템의 NA는 0.6-0.8이다.  $\lambda=600$  nm를 사용하면, 회절 한계는 375 nm에서 500 nm 사이이다. 수침 시스템(water immersion)의 경우, NA는 ~1.0이며 회절 한계는 300nm이다.

[0072] 어레이 또는 생체분자를 포함하는 기타 기관 표면의 특징이 너무 가까운 경우, 두 개의 광학 신호가 상당히 중첩되어 이미지만으로는 안정적으로 해결할 수 없는 단일 얼룩만을 보게 된다. 이는 움직이는 기관의 부정확한 추적으로 인한 불선명함 또는 센서와 기관 표면 사이의 광로에서의 광학적 변화와 같은, 광학 이미징 시스템에 의해 발생된 오류로 인해 악화될 수 있다.

[0073] 현미경의 시료면의 한 지점에서 나오는 투과광 또는 형광 발광 파동은 대상 개구의 가장자리에서 회절되어, 파동을 효과적으로 확산시켜 유한한 중심 원반을 가지지만, 원점보다 큰 회절 패턴으로 확대된 점원(point source) 이미지를 생성한다. 따라서, 빛의 회절로 인해, 현미경 광학 시스템이 구조적인 세부 사항을 분해할 수 없는 하한이 있기 때문에 표본의 이미지는 표본에 존재하는 실제적인 세부 사항을 완벽하게 나타내지 못한다.

[0074] 현미경으로 서브 파장 구조를 관찰하는 것은 회절 한계로 인해 어렵다. 형광 단백질 또는 뉴클레오티드 단일 분자와 같은, 현미경의 점 개체(point object)는 간섭 작용에 의해 생성된 회절 패턴으로 이루어진 중계면(intermediate plane)에서 이미지를 생성한다. 고배율에서, 점 개체의 회절 패턴은 일련의 회절 고리들에 의해 둘러싸인 중심 지점(회절 원반)으로 이루어진 것으로 관찰된다. 결합된, 이 점원 회절 패턴을 에어리 원반(Airy disk)이라 지칭된다.

[0075] 에어리 패턴에서 중심 지점의 크기는 빛의 파장 및 대물렌즈의 개구각과 관련이 있다. 현미경 대물렌즈의 경우, 개구각은 대물렌즈가 표본으로부터 빛을 모을 수 있는 반각,  $\sin \theta$ 라는 용어를 포함하는, 개구수(NA)로 설명된다. 분해능 측면에서, 측면(x,y) 이미지 평면에서 회절의 반경 에어리 원반은 다음 공식으로 정의된다: 아베 분해능(Abbe Resolution)  $x,y = \lambda / 2 \cdot NA$ , 여기서  $\lambda$ 는 투과광에서 조명의 평균 파장 또는 형광에서 여기 파장 대역이다. 대물렌즈 개구수( $NA = n \cdot \sin(\theta)$ )는 이미징 매체(n; 일반적으로 공기, 물, 글리세린 또는 오일)의 굴절률에 개구각의 사인( $\sin(\theta)$ )을 곱하여 정의된다. 이 관계의 결과로, 점원에 의해 생성된 지점의 크기는 감소하는 파장 및 증가하는 개구수에 따라 감소하나, 항상 유한한 직경의 원반으로 남아 있다. 아베 분해능(즉, 아베 한

계)은 또한 본 명세서에서 회절 한계로 지칭되고 광학 시스템의 분해능 한계를 정의한다.

[0076] 2개의 에어리 원반 또는 포인트 스프레드(point-spread) 기능 사이의 거리가 이 값보다 큰 경우, 2개의 점원은 분리(resolve)된 것으로 간주된다(또한 쉽게 구별될 수 있음). 그렇지 않으면 에어리 원반은 병합되고 분리(resolve)되지 않은 것으로 간주된다.

[0077] 따라서, 파장  $\lambda$ 에서 단일 분자 검출이 가능한 점원에서 방출된 빛이, 굴절률이  $n$ 인 매질에서 이동하고 반각  $\theta$ 으로 지점으로 수렴하면 직경:  $d = \lambda / 2 * NA$ 인 회절 한계 지점이 만들어진다. 부근의 녹색광과 1개의 NA(Numerical Aperture; 개구수)를 고려하면, 이는 기존 이미징 기술로 이미지화할 수 있는 표면의 뉴클레오티드 및 단일 분자 단백질과 같은 분석물의 밀도를 제한하는, 회절 한계는 대략  $d = \lambda / 2 = 250 \text{ nm} (0.25 \mu\text{m})$ 이다. 광학 현미경이 사용 가능한 최고 품질의 렌즈 요소가 장착되고, 완벽하게 정렬되어 있고, 가장 높은 개구수를 가질 수 있는 경우에도, 분해능은 최상의 시나리오에서 빛의 파장의 약 절반으로 제한된다.

[0078] **디컨볼루션**

[0079] 디컨볼루션은 기록된 데이터에 대한 컨볼루션의 효과를 반전시키는 데 사용되는 알고리즘 기반 프로세스이다. 디컨볼루션의 개념은 신호 처리 및 이미지 처리 기술에서 널리 사용된다. 이러한 기술은 결국 많은 과학 및 공학 분야에서 널리 사용되기 때문에 디컨볼루션은 많은 분야를 찾는다.

[0080] 광학 및 이미징에서, 용어 "디컨볼루션"은 광학 현미경, 전자 현미경, 망원경, 또는 기타 이미징 기기에서 발생하는 광학 왜곡을 반전시켜 보다 선명한 이미지를 생성하는 프로세스를 지칭하기 위해 구체적으로 사용된다. 이는 일반적으로 현미경 이미지 처리 기술의 일부로서, 소프트웨어 알고리즘에 의해 디지털 영역에서 수행된다.

[0081] 일반적인 방법은 기기를 통과하는 광학 경로가 광학적으로 완벽하고, 포인트 스프레드 기능(Point Spread Function; PFS)로 구성되어 있다고 가정하는 것이다. 즉, 이론적인 점광원(또는 다른 점과원)이 기기를 통과하는 경로의 관점에서 왜곡을 설명하는 수학적 함수이다. 일반적으로, 이러한 점원은 최종 이미지에 작은 영역의 흐릿함을 제공한다. 이 함수가 결정될 수 있다면, 이후 이의 역함수 또는 보완 함수를 계산하고, 획득된 이미지를 그것으로 컨볼루션하는 문제이다. 디컨볼루션은 푸리에 공동 도메인(Fourier co-domain) 내 분할에 매핑된다. 이것은 푸리에 변환의 대상이 되는 실험 데이터에 디컨볼루션을 쉽게 적용할 수 있게 한다. 예시는 시간 영역에서 데이터가 기록되나, 주파수 영역에서 데이터가 분석되는, NMR 분광법이다. 지수 함수에 의한 시간 영역 데이터의 분할은 주파수 영역에서 로렌지안(Lorentzian lines) 라인의 폭을 줄이는 효과가 있다. 결과는 왜곡되지 않은 원본 이미지이다.

[0082] 그러나, 회절 한계 이미징의 경우, 포인트 스프레드 기능이 공지되었더라도, 디컨볼루션은 신호를 추가로 정제하여 회절 한계를 넘어 분해력을 개선하는 데도 필요하다. 나이퀴스트 거리(Nyquist distance)보다 짧은 거리에서 두 물체를 확실하게 분리하는 것은 매우 어렵다. 그러나, 나이퀴스트 거리보다 훨씬 짧은 거리로 분리된 물체를 안정적으로 검출하기 위해 순환 검출, 분석물 위치 결정, 정렬 및 디컨볼루션을 사용하는 방법 및 시스템이 본 명세서에 설명되어 있다.

[0083] **염기서열분석**

[0084] 광학 검출 이미징 시스템은 회절 한계적이므로, 염기서열분석에 일반적으로 사용되는 발형광단(fluorophores)을 사용하여 이론상 최대 분해력이  $\sim 300 \text{ nm}$ 이다. 현재까지, 최고의 염기서열분석 시스템은 어레이에서 인접한 폴리 뉴클레오티드 사이의 중심 간 거리가  $\sim 600 \text{ nm}$ , 또는 회절 한계의  $\sim 2$ 배였었다. 이러한 2배의 인자는 위치 오류를 초래할 수 있는 강도, 어레이 및 생물학적 변화를 설명하는 데 필요하다. 염기서열분석의 경우, 본 명세서에 설명된 시스템 및 방법의 목적은 광학 시스템의 회절 한계 아래의 중심 간 거리로 기관에서 염기서열분석 된 폴리 뉴클레오티드를 분석하는 것이다.

[0085] 본 명세서에 설명된 바와 같이, 우리는 부분적으로 높은 정확도(예:  $10 \text{ nm RMS}$  또는 그 이하)로 각 분석물의 위치를 식별하여 서브 회절 한계 이미징을 달성하는 방법 및 시스템을 제공한다. 이에 비해, 최첨단 초분해능 시스템(Harvard/STORM)은 이러한 시스템보다 2배 이상 악화된  $20 \text{ nm RMS}$ 까지의 정확도로만 위치를 식별할 수 있다. 따라서, 본 명세서에 공개된 방법 및 시스템은 서브-회절 한계 이미징을 통해 기관 상의 밀집하게 패키징된 분자를 식별하여 효소 단위당 높은 데이터 속도, 단위 시간당 데이터 속도 및 높은 데이터 정확도를 달성할 수 있다. 이러한 서브 회절 한계 이미징 기술은 본 명세서에 설명된 순환 검출을 사용하는 기술에 광범위하게 적용할 수 있다.

[0086] **이미징 및 순환 검출**

- [0087] 본 명세서에서 설명된 바와 같이, 각각의 검출 방법 및 시스템은 서브 회절 한계 이미징을 달성하기 위해 순환 검출을 필요로 했다. 순환 검출에는 가시광 광학 신호를 방출할 수 있는 검출 가능한 라벨에 결합된, 항체 또는 뉴클레오티드같은, 결합 및 이미징 또는 프로브를 포함한다. 서로 다른 사이클에서 필드의 일련의 이미지로부터 위치 정보를 사용함으로써, 밀집하게 패킹된 기관으로부터 신호를 분해하기 위한 디컨볼루션은 광학 이미징의 회절 한계로 인해 가려진 신호로부터 개별 광학 신호를 식별하는 데 효과적으로 사용될 수 있다. 여러 사이클 후에 분자의 정확한 위치는 점점 더 정확해질 것이다. 이러한 정보를 사용하여, 픽셀 이산화 효과로 인해 발생하는 누화 행렬(crosstalk matrix)의 공지된 비대칭에 관하여 누화 정정을 돕기 위해 추가 계산이 수행될 수 있다.
- [0088] 순환 프로브 결합 및 광학 검출을 사용하는 방법 및 시스템은 2015년 11월 19일에 발행된 미국 공개 번호 2015/0330974, 단일 분자 검출을 이용한 분자 분석물의 디지털 분석(Digital Analysis of Molecular Analytes Using Single Molecule Detection)에 설명되어 있으며, 이는 그 전체는 본 명세서에 참조로 포함된다.
- [0089] 일부 실시예에서, 원시 이미지는 오버샘플링 된 이미지의 보다 정확한 판단을 용이하게 하기 위해 적어도 나이퀴스트 한계에 있는 샘플링을 사용하여 획득된다. 나이퀴스트 한계(오버샘플링)를 초과하여 샘플링하여 이미지를 나타내는 데 사용되는 픽셀 수를 증가시키는 것은 이미지 처리 및 표시에 사용할 수 있는 픽셀 데이터가 증가시킨다.
- [0090] 이론적으로, 대역폭이 제한된 신호는 나이퀴스트 속도 또는 그 이상으로 샘플링 되는 경우 완벽하게 재구성될 수 있다. 나이퀴스트율(Nyquist rate)은 신호에서 가장 높은 주파수 성분의 2배로 정의된다. 오버샘플링은 해상도를 향상시키고 노이즈를 줄이며 앨리어싱(aliasing) 제거 필터 성능 요구 사항을 완화함으로써 앨리어싱 및 위상 왜곡을 방지하는 데 도움이 된다. 신호가 나이퀴스트율의 N배로 샘플링 되는 경우 신호는 N개의 인자만큼 오버샘플링될 수 있다.
- [0091] 따라서, 일부 실시예에서, 각각의 이미지는 관찰되는 빛의 파장의 절반 이하인 픽셀 크기로 촬영된다. 다시 말해서, 광학적 검출 시스템에서 검출된 하나 이상의 검출 가능한 라벨로부터 생성된 신호의 파장은 광학적 검출 시스템의 픽셀의 2배보다 크다. 예를 들어, 일부 실시예에서, 162.5 nm x 162.5 nm의 픽셀 크기는 나이퀴스트 한계 이상에서 샘플링을 달성하기 위해 검출에 사용된다. 기관의 원시 이미징 동안 적어도 나이퀴스트 한계의 주파수에서 샘플링하는 것은 본 명세서에 설명된 시스템 또는 방법의 해상도를 최적화하는 데 선호된다. 이것은 높은 정확도로 회절 한계 아래에서 기관의 특징을 분해하기 위해 본 명세서에 설명된 디컨볼루션 방법 및 광학 시스템과 함께 수행될 수 있다.
- [0092] **오류-정정 방법들**
- [0093] 상술한 광학적 및 전기적 검출 방법에서, 신호 결합 및/또는 검출에서 오류가 발생할 수 있다. 일부 경우에, 오류율은 5분의 1만큼 높을 수 있다(예: 형광성 신호 5개 중 1개가 올바르게 읽음). 이것은 5회 사이클 시퀀스마다 한 번의 오류에 대응하는 것이다. 실제 오류율은 20%만큼 높지 않을 수 있으나, 몇 퍼센트의 오류율은 가능하다. 일반적으로, 오류율은 샘플에서 분석 물질의 유형 및 사용된 프로브의 유형을 포함한 많은 요인에 달려 있다. 예를 들어, 전기적 검출 방법에서, 꼬리 영역은 사이클 동안 aptamer의 대응하는 프로브 영역에 적절하게 결합하지 않을 수 있다. 광학적 검출 방법에서, 항체 프로브는 표적에 결합하지 않거나 잘못된 표적에 결합할 수 있다.
- [0094] 검출된 신호의 오류를 설명하고 패리티 비트와 같은, 정보의 추가 비트를 얻기 위해 추가 사이클이 생성된다. 정보의 추가 비트는 오류 정정 부호를 사용하여 오류를 정정하기 위해 사용된다. 일 실시예에서, 오류 정정 부호는 시스템에서 오류를 검출하고 정정하기 위해 사용되는 비-이진 순환 부호인 리드-솔로몬 부호다. 다른 실시예에서, 다양한 다른 오류 정정 부호가 사용될 수 있다. 기타 오류 정정 부호는, 예를 들어, 블록 부호(block codes), 컨볼루션 부호(convolution codes), 골레이 부호(Golay codes), 해밍 부호(Hamming codes), BCH 부호, AN 부호, 리드-몰러 부호(Reed-Muller codes), 고평 부호(Goppa codes), 아마라드 부호(Hadamard codes), 왈쉬 부호(Walsh codes), 하젤버거 부호(Hagelbarger codes), 극 부호(polar codes), 반복 부호(repetition code), 반복 누적 부호(repeat-accumulate codes), 소거 부호(eraser codes), 온라인 부호(online codes), 그룹 부호(group codes), 익스팬더 부호(expander codes), 고정 가중치 부호(constant-weight code), 토네이도 부호(tornado codes), 저밀도 패리티 검사 부호(low-density parity check codes), 최대 거리 부호(maximum distance codes), 버스트 오류 부호(burst error codes), 루비 변환 부호(luby transform codes), 분수 부호(fountain codes) 및 랩터 부호(raptor codes)를 포함한다. 오류 제어 코딩(Error Control Coding, 2nd Ed., S.Lin and DJ Costello, Prentice Hall, New York, 2004.)을 참조한다. 사이클을 추가하고 정보의 추가 비트를

획득함으로써 오류 정정을 위한 방법을 보여주는 예시는 아래에 제공된다.

- [0095] 광학적 검출 방법
- [0096] 일부 실시예에서, 기관은 N개의 표적 분석물을 포함하는 분석물과 결합된다. N 개의 표적 분석물을 검출하기 위해, 프로브 결합 및 신호 검출의 M회의 사이클이 선택된다. M 회의 사이클 각각은 하나 이상의 패스를 포함하고, 각 패스는 N개의 세트의 프로브를 포함하므로, 각 프로브 세트는 N개의 표적 분석물 중 하나에 특이적으로 결합한다. 특정 실시예에서, N개의 표적 분석물에 대한 N개의 세트의 프로브가 있다.
- [0097] 각 사이클에는, 각 패스에 대한 프로브 세트를 도입하기 위한 기 결정된 순서가 있다. 일부 실시예에서, 프로브 세트에 대한 기 결정된 순서는 무작위적 순서이다. 다른 실시예에서, 프로브 세트에 대한 기 결정된 순서는 비 무작위적 순서이다. 일 실시예에서, 비무작위적 순서는 컴퓨터 프로세서에 의해 선택될 수 있다. 기 결정된 순서는 각 표적 분석물에 대한 키(key)에 표시된다. 키는 프로브 세트의 순서를 포함하도록 생성되고, 프로브의 순서는 각각의 표적 분석물을 식별하기 위한 부호로 디지털화된다.
- [0098] 일부 실시예에서, 정렬되는 프로브의 각각의 세트는 표적 분석물을 검출하기 위한 별도의 태그와 연관되고, 별도의 태그의 수는 N개의 표적 분석물의 수보다 적다. 그러한 경우, 각각의 N개의 표적 분석물은 M회 사이클에 대한 M 태그의 시퀀스와 일치된다. 태그의 정렬되는 시퀀스는 식별 부호로서 표적 분석물과 연관된다.
- [0099] 일 실시예에서, 상기 방법은 X개의 상이한 색상의 형광으로 태깅된 프로브를 이용하여 기관 상의 N개의 상이한 종류의 표적 분석물을 세기 위해 프로브 풀을 라벨링하기 위한 다음 단계를 포함한다:
- [0100] 1. 염기-X개의 수를 사용하여 N개의 표적(또는 그들의 프로브) 목록에 번호를 지정한다.
- [0101] 2. 형광성 태그를 0에서 X-1까지 염기-X 번호와 연결한다. (예를 들어, 0, 1, 2, 3은 적색, 청색, 녹색, 황색에 대응한다.)
- [0102] 3.  $XC > N$ 이 되는 C를 구한다.
- [0103] 4. N개의 표적을 식별하려면 최소한 C개의 프로브 풀이 필요하다. 인덱스  $k=1$ 로 C 프로브 풀에 라벨을 지정한다.
- [0104] 5. k번째 프로브 풀에서, 1단계에서 생성된 목록에서 프로브의 표적을 식별하는 염기-X 수의 k번째 염기-X 번호에 대응하는 색상의 형광성 태그로 각 프로브에 라벨을 지정한다.
- [0105] 예를 들어,  $N=10,000$ 개의 표적 분석물과 4개의 형광성 태그가 있는 경우 염기 4개를 선택할 수 있다. 4개의 형광성 태그 색상은 각각 숫자 0, 1, 2 및 3으로 지정된다. 예를 들어, 숫자 0, 1, 2, 3은 적색, 청색, 녹색 및 황색에 대응한다.
- [0106] 염기 4개가 선택되는 경우, 각 형광색은 2 비트(0 및 1, 여기서 0=신호 없음 및 1=신호 존재)로 표시되며, 표적 분석물을 식별하기 위한 부호로 사용되는 7가지 색상이 있다. 예를 들어, 단백질 A는 "청색, 녹색, 녹색, 청색, 청색, 황색, 황색"의 색상 조합 및 순서를 나타내는 "1221133"의 부호로 식별될 수 있다. 7가지 가능한 색상에 대해, 표적 분석물에 대한 총 14개 비트의 정보가 있다( $7 \times 2=14$ 비트).
- [0107] 다음으로, C는  $4C > 10,000$ 이 되도록 선택된다. 이러한 경우, C는 10,000개의 표적을 식별하기 위한 7개의 프로브 풀이 있도록 7개가 될 수 있다(10,000보다 큰,  $47=16,384$ ). 길이 C의 색상 시퀀스는 C개의 서로 다른 프로브 풀이 구성되어야 함을 의미한다. 7개의 프로브 풀은  $k=1$ 에서 7로 라벨이 지정된다. 이후 각 프로브는 k번째 염기 및 X-번호에 대응하는 형광성 태그로 라벨이 지정된다. 예를 들어, 부호 "1221133"에서 세 번째 프로브는 3 번째 염기-4번째 번호가 될 것이고, 이는 녹색에 해당한다.
- [0108] 광학적으로 검출된 프로브의 정량화
- [0109] 검출 과정 이후, 신호는 각 프로브 풀로부터 계산되고, 신호의 존재 또는 부재 및 신호의 색상은 기관의 각 위치에 대해 기록될 수 있다.
- [0110] 검출 가능한 신호로부터, N개의 별개의 표적 분석물에 대해 M 사이클 각각에서 K개의 비트의 정보가 얻어진다.  $K \times M=L$  정보의 비트 및  $L \log_2(N)$ 이 되도록, K개의 비트 정보는 L개의 총 정보의 비트를 결정하는 데 사용된다. L개의 비트의 정보는 N개의 별개의 표적 분석물의 정체성 (및 존재)을 결정하는 데 사용된다. 단 하나의 사이클 ( $M=1$ )이 수행되는 경우,  $K \times 1=L$ 이다. 그러나, 분석물 당 L개의 총 정보의 비트를 생성하기 위해 다중 사이클 ( $M>1$ )이 수행될 수 있다. 각 후속 주기는 표적 분석물을 식별하는 데 사용되는 추가 광학 신호 정보를

제공한다.

- [0111] 실제로, 신호 내 오류는 발생하고, 이는 표적 분석물의 식별에 대한 정확도를 혼란스럽게 한다. 예를 들어, 프로브는 잘못된 표적(예: 거짓 양성)에 결합하거나 올바른 표적(예: 거짓 음성)에 결합하지 못할 수 있다. 광학적 및 전기적 신호 검출의 오류를 설명하기 위해 후술하는 방법이 제공된다.
- [0112] 전기적 검출 방법
- [0113] 다른 실시예에서, 전기적 검출 방법을 이용하여 기관 상의 표적 분석물의 존재를 검출한다. 표적 분석물은 올리고뉴클레오티드 꼬리 영역으로 태그되고 올리고뉴클레오티드 태그는 용액 내 수소 이온 농도를 측정하는 이온 민감 전계 효과 트랜지스터(ion-sensitive field-effect transistors; ISFET 또는 pH 센서)를 이용하여 검출된다.
- [0114] ISFET는 분석물의 식별 및 특성화를 위해 민감하고 특정한 전기적 검출 시스템을 제공한다. 일 실시예에서, 본 명세서에 개시된 전기적 검출 방법은 컴퓨터(예: 프로세서)에 의해 수행된다. 용액의 이온 농도는 ISFET의 전극에 의해 대수의(logarithmic) 전위로 변환될 수 있으며, 전기 출력 신호는 검출 및 측정될 수 있다.
- [0115] 이전에 ISFET는 DNA 염기서열분석을 용이하게 하는 데 사용되었다. 단일 가닥 DNA가 이중 가닥 DNA로 효소적 전환되는 동안, 수소 이온은 각 뉴클레오티드가 DNA 분자에 부가되는 경우 방출된다. ISFET는 이렇게 방출된 수소 이온을 검출하고 뉴클레오티드가 DNA 분자에 추가된 때를 판단할 수 있다. 뉴클레오시드 삼인산(nucleoside triphosphates; dATP, dCTP, dGTP 및 dTTP)의 통합을 동기화함으로써, DNA 시퀀스 또한 결정될 수 있다. 예를 들어, 단일 가닥 DNA 템플릿이 dATP에 노출되었을 때 전기적 출력 신호가 검출되지 않지만, 전기 출력 신호가 dGTP의 존재 하에 검출되는 경우, DNA 서열은 상기 위치에서 상보적인 시토신(cytosine) 염기로 구성된다.
- [0116] 한 실시예에서, ISFET는 프로브의 꼬리 영역을 검출한 후 대응하는 표적 분석물을 식별하기 위해 사용된다. 예를 들어, 표적 분석물은 하나 이상의 ISFET를 포함하는 집적 회로 칩과 같은, 기관에 고정될 수 있다. 대응하는 프로브(예: 압타머 및 꼬리 영역)가 첨가되고 표적 분석물에 특이적으로 결합하는 경우, 꼬리 영역의 전사를 위해 뉴클레오티드 및 효소(중합 효소)가 첨가된다. ISFET는 방출 수소 이온을 전기 출력 신호로서 검출하고 dNTP가 꼬리 영역에 포함되는 경우 이온 농도의 변화를 측정한다. 방출된 수소 이온의 양은 꼬리 영역의 길이와 정지점에 대응하고, 꼬리 영역에 대한 이러한 정보는 다양한 태그를 구별하는 데 사용할 수 있다.
- [0117] 꼬리 영역의 가장 단순한 유형은 완전히 하나의 단일중합체 염기 영역으로 구성된 것이다. 이 경우, 4개의 가능한 꼬리 영역이 있다: 폴리-A 꼬리, 폴리-C 꼬리, 폴리-G 꼬리 및 폴리-T 꼬리. 그러나, 꼬리 영역에서 넓은 다양성을 갖는 것이 종종 바람직하다.
- [0118] 꼬리 영역에서 다양성을 생성하는 한 가지 방법은 꼬리 영역의 단일중합체 염기 영역 내에 정지 염기를 제공하는 것이다. 정지 염기는 하나 이상의 뉴클레오티드가 단일중합체 염기 영역 내의 염기와 구별되는 염기로 구성되도록, 단일중합체 염기 영역에 인접한 적어도 하나의 뉴클레오티드를 포함하는 꼬리 영역의 일부이다. 일 실시예에서, 정지 염기는 1개의 뉴클레오티드다. 다른 실시예에서, 정지 염기는 다수의 뉴클레오티드를 포함한다. 일반적으로, 정지 염기는 2개의 단일중합체 염기 영역에 의해 측면에 배치된다. 일 실시예에서, 정지 염기에 인접한 2개의 단일중합체 염기 영역은 동일한 염기로 구성된다. 다른 실시예에서, 2개의 단일중합체 염기 영역은 2개의 상이한 염기로 구성된다. 다른 실시예에서, 꼬리 영역은 하나 이상의 정지 염기를 포함한다.
- [0119] 일 예시에서, ISFET는 100개의 수소 이온의 최소 임계값 수를 검출할 수 있다. 표적 분석물 1은 꼬리 영역 길이 총 201개 뉴클레오티드에 대해, 100-뉴클레오티드 폴리-A 꼬리, 그 다음에는 하나의 시토신 염기, 그 다음에는 또 다른 100개 뉴클레오티드 폴리-A 꼬리로 구성된 꼬리 영역을 갖는 조성물에 결합된다. 표적 분석물 2는 200-뉴클레오티드 폴리-A 꼬리로 구성된 꼬리 영역을 갖는 조성물에 결합된다. dTTP의 첨가 시 및 폴리뉴클레오티드 합성에 도움이 되는 조건 하에, 표적 분석물1과 연결된 꼬리 영역의 합성은 100개의 수소 이온을 방출할 것이며, 이는 200개의 수소 이온을 방출할, 표적 분석물 2와 연결된 꼬리 영역의 폴리뉴클레오티드 합성과 구별될 수 있다. ISFET는 각 꼬리 영역에 대해 서로 다른 전기 출력 신호를 검출한다. 게다가, dGTP가 첨가된 후, 더 많은 dTTP가 첨가되는 경우, 표적 분석물 1과 관련된 꼬리 영역은 추가 폴리뉴클레오티드 합성으로 인해 하나, 이후 100개 더 많은 수소 이온을 방출할 것이다. 꼬리 영역 구성물에 기반한 특정 뉴클레오시드 삼인산염의 첨가로 생성된 고유한 전기 출력 신호는 ISFET가 각 꼬리 영역에서 수소 이온을 검출할 수 있도록 하고, 대응하는 정보를 이용하여 꼬리 영역과 대응하는 표적 분석물을 식별할 수 있다.
- [0120] 다양한 길이의 단일중합체 염기 영역, 정지 염기 및 이들의 조합은 샘플의 각 분석물에 고유하게 태그하는 데 이용될 수 있다. 기관에서 표적 분석물을 식별하기 위한 압타머 및 꼬리 영역의 전기적 검출에 대한 추가 설명

은 그 전체가 참고로 포함되는 미국 특허출원 출원번호 2016/020119에 기재되어 있다.

- [0121] 일부 실시예에서, 기관(들) 상에 저장된 데이터 카탈로그 내 대량의 정보는 몇 가지 레벨의 내장 중복성(built-in redundancy)을 생성한다. 일부 실시예에서, 정보 세분화의 제1 레벨은 데이터의 각 정보 세그먼트에 대한 슬라이드, 레인 및 특정 염기서열분석 프라이밍 부위에 포함된다. 일부 실시예에서, 본 명세서에서 설명된 검색을 위해 최적화 되기 위해 생성되는 개별 레인은 다양한 조합으로 저장된다.
- [0122] **본 명세서에 설명된 시스템 및 방법의 컴퓨터 자동화**
- [0123] 본 발명은 본 발명의 방법을 구현하도록 프로그래밍된 컴퓨터 시스템을 제공한다. 도 2는 데이터 센터 내의 장착 가능한 랙에 기관을 배치하고 염기서열분석을 위해 데이터 센터 내에 포함된 기기로 기관을 검색 및 전달하도록 프로그래밍되거나 그 외의 방법으로 구성된 컴퓨터 시스템(201)을 도시한다. 컴퓨터 시스템(201)은, 예를 들어, 데이터 센터의 온도 및 데이터 센터 내에 저장된 기관의 구성과 같은, 본 개시의 다양한 태양을 조절할 수 있다. 컴퓨터 시스템(201)은 사용자의 전자 장치이거나 전자 장치에 대해 원격으로 위치한 컴퓨터 시스템일 수 있다. 전자 장치는 모바일 전자 장치일 수 있다.
- [0124] 컴퓨터 시스템(201)은 단일 코어 또는 멀티 코어 프로세서, 또는 병렬 처리를 위한 복수의 프로세서일 수 있는, 중앙 처리 장치(Central Processing Unit; CPU, 또한 본 명세서에서 "프로세서" 및 "컴퓨터 프로세서")(205)를 포함한다. 컴퓨터 시스템(201)은 또한 메모리 또는 메모리 위치(210)(예: 랜덤 액세스 메모리, 읽기 전용 메모리, 플래시 메모리), 전자 저장부(215)(예: 하드 디스크), 하나 이상의 다른 시스템과 통신하기 위한 통신 인터페이스(220)(예: 네트워크 어댑터) 캐시, 다른 메모리, 데이터 저장 및/또는 전자 디스플레이 어댑터와 같은 주변 장치(225)를 포함할 수 있다. 메모리(210), 저장부(215), 인터페이스(220) 및 주변 장치(225)는 메인보드와 같은, 통신 버스(실선)를 통해 CPU(205)와 통신한다. 저장부(215)는 데이터를 저장하기 위한 데이터 저장부(또는 데이터 저장소)일 수 있다. 컴퓨터 시스템(201)은 통신 인터페이스(220)의 도움으로 컴퓨터 네트워크("네트워크")(230)에 작동 가능하게 연결될 수 있다. 네트워크(230)는 인터넷(Internet), 인터넷(internet) 및/또는 엑스트라넷(extranet), 또는 인터넷(Internet)과 통신하는 인트라넷(intranet) 및/또는 엑스트라넷일 수 있다. 일부 경우에 네트워크(230)는 통신 및/또는 데이터 네트워크이다. 네트워크(230)는 클라우드 컴퓨팅과 같은, 분산 컴퓨팅을 가능하게 할 수 있는 하나 이상의 컴퓨터 서버를 포함할 수 있다. 네트워크(230)는 일부 경우에 컴퓨터 시스템(201)의 도움으로 컴퓨터 시스템(201)에 연결된 장치가 클라이언트 또는 서버로 동작하도록 할 수 있게 하는, 피어-투-피어(peer-to-peer) 네트워크를 구현할 수 있다. 일부 실시예에서, 네트워크(230)는 장착 가능한 저장 랙 및 염기서열분석을 위한 기기로 기관을 기계적으로 운반하기 위한 기기를 포함한다. 일부 실시예에서, 네트워크(230)는 염기서열분석을 위한 기기를 포함한다.
- [0125] CPU(205)는 프로그램 또는 소프트웨어로 구현될 수 있는 일련의 기계 판독 가능 명령어를 실행할 수 있다. 명령어는 메모리(210)와 같은, 메모리 위치에 저장될 수 있다. 명령어는 본 발명의 방법을 구현하도록 CPU(205)를 후속적으로 프로그래밍하거나 다른 방법으로 구성할 수 있는, CPU(205)에 지시할 수 있다. CPU(205)에 의해 수행되는 작업의 예시들은 페치(fetch), 디코딩(decode), 실행 및 쓰기 저장(writeback)을 포함할 수 있다.
- [0126] CPU(205)는 집적 회로와 같은, 회로의 일부일 수 있다. 시스템(201)의 하나 이상의 다른 구성요소는 상기 회로에 포함될 수 있다. 일부 경우에, 상기 회로는 주문형 집적 회로(Application Specific Integrated Circuit ASIC)이다.
- [0127] 저장부(215)는 드라이버, 라이브러리 및 저장된 프로그램과 같은, 파일을 저장할 수 있다. 저장부(215)는 예를 들어, 사용자 선호도 및 사용자 프로그램 및 핵산 염기서열분석 판독과 같은, 사용자 데이터를 저장할 수 있다. 일부 경우에 컴퓨터 시스템(201)은 인트라넷 또는 인터넷을 통해 컴퓨터 시스템(201)과 통신하는 원격 서버에 위치하는 것과 같이, 컴퓨터 시스템(201) 외부에 있는 하나 이상의 추가 데이터 저장 장치를 포함할 수 있다. .
- [0128] 컴퓨터 시스템(201)은 네트워크(230)를 통해 하나 이상의 원격 컴퓨터 시스템과 통신할 수 있다. 예를 들어, 컴퓨터 시스템(201)은 사용자의 원격 컴퓨터 시스템(예를 들어, 염기서열분석을 위한 기기)과 통신할 수 있다. 원격 컴퓨터 시스템의 예시에는 개인용 컴퓨터(예: 휴대용 PC), 슬라이드 또는 태블릿 PC(예: 애플® 아이패드, 삼성® 갤럭시 탭), 전화, 스마트폰(예: 애플® 아이폰, 안드로이드 지원 장치, 블랙베리®), 또는 개인 디지털 비서를 포함한다. 사용자는 네트워크(230)를 통해 컴퓨터 시스템(201)에 액세스할 수 있다.
- [0129] 본 명세서에 설명된 방법은 예를 들어, 메모리(210) 또는 전자 저장부(215)와 같은, 컴퓨터 시스템(201)의 전자 저장 위치에 저장된 기계(예: 컴퓨터 프로세서) 실행 부호를 통해 구현될 수 있다. 기계 실행 가능 또는 기계 판독 가능 부호는 소프트웨어의 형태로 제공될 수 있다. 사용 중에 부호는 프로세서(205)에 의해 실행될 수 있

다.

- [0130] 일부 경우에, 부호는 저장부(215)로부터 검색될 수 있고 프로세서(205)에 의한 준비된 액세스를 위해 메모리(210)에 저장될 수 있다. 일부 상황에서, 전자 저장부(215)는 배제될 수 있고, 기계 실행 가능 명령어는 메모리(210)에 저장된다. 부호는 미리 컴파일 되어 부호를 실행하도록 구성된 프로세서가 있는 기계와 함께 사용하도록 구성되거나 런타임 중에 컴파일 될 수 있다. 부호는 미리 컴파일 되거나 컴파일 된 대로 부호를 실행할 수 있게 하기 위해 선택될 수 있는 프로그래밍 언어로 제공될 수 있다.
- [0131] 컴퓨터 시스템(201)과 같은, 본 명세서에 제공된 시스템 및 방법의 태양은 프로그래밍으로 구현될 수 있다. 기술의 다양한 태양은 일반적으로 기계(또는 프로세서) 실행 가능 부호 및/또는 기계 판독 가능 매체의 유형에 내장되거나 수행되는 관련된 데이터 형태의 "제품" 또는 "제조품"으로 간주될 수 있다. 기계 실행 가능 부호는 메모리(예: 읽기 전용 메모리, 랜덤 액세스 메모리, 플래시 메모리) 또는 하드 디스크와 같은 전자 저장 장치에 저장될 수 있다. "저장" 유형의 매체는 소프트웨어 프로그래밍을 위해 언제든지 비밀시적 저장을 제공할 수 있는 다양한 반도체 메모리, 테이프 드라이브, 디스크 드라이브 등과 같은, 컴퓨터, 프로세서 등의 유형 메모리 또는 관련 모듈의 일부 또는 전부를 포함할 수 있다. 소프트웨어의 전체 또는 일부는 때때로 인터넷 또는 다양한 기타 통신 네트워크를 통해 통신될 수 있다. 이러한 통신은, 예를 들어, 어느 컴퓨터 또는 프로세서에서 다른 컴퓨터 또는 프로세서로, 예를 들어, 관리 서버 또는 호스트 컴퓨터에서 응용 프로그램 서버의 컴퓨터 플랫폼으로 소프트웨어를 로드하는 것을 가능하게 할 수 있다. 따라서, 소프트웨어 요소를 포함할 수 있는 다른 유형의 매체는 유선 및 광학 유선 네트워크 및 다양한 에어 링크들(air-links)을 통해 로컬 장치 간의 물리적 인터페이스 전반에 걸쳐 사용되는, 광학, 전기 및 전자파를 포함한다. 유선 또는 무선 링크, 광 링크 등과 같은, 이러한 전파를 전달하는 물리적 요소도 소프트웨어를 포함하는 매체로 간주될 수 있다. 본 명세서에 사용된 바와 같이, 비밀시적 유형의 "저장" 매체로 제한되지 않는 한, 컴퓨터 또는 기계 "판독 가능 매체"와 같은, 용어는 실행을 위해 프로세서에 명령을 제공하는 데 참여하는 임의의 매체를 의미한다.
- [0132] 따라서, 컴퓨터 실행 가능 부호와 같은, 기계 판독 가능 매체는 유형의 저장 매체, 반송파 매체 또는 물리적 전송 매체를 포함하지만 이에 제한되지 않는, 많은 형태를 취할 수 있다. 비휘발성 저장 매체는, 예를 들어, 도면에 도시된, 데이터베이스 등을 구현하는 데 사용될 수 있는, 임의의 컴퓨터(들) 내 임의의 저장 장치 등과 같은, 광학 또는 자기 디스크를 포함한다. 휘발성 저장 매체는 이러한 컴퓨터 플랫폼의 메인 메모리와 같은, 동적 메모리를 포함한다. 유형의(tangible) 전송 매체는 동축 케이블, 컴퓨터 시스템 내에서 버스를 구성하는 전선을 포함하는 구리선 및 광섬유를 포함한다. 반송파 전송 매체는 전기 또는 전자기 신호, 또는 무선 주파수(Radio Frequency; RF) 및 적외선(Infrared; IR) 데이터 통신 중에 생성되는 것과 같은 음향 또는 광파의 형태를 취할 수 있다. 따라서 일반적인 형태의 컴퓨터 판독 가능 매체는 예를 들어 다음과 같은 사항이 포함된다: 플로피 디스크, 플렉서블 디스크, 하드 디스크, 자기 테이프, 기타 자기 매체, CD-ROM, DVD 또는 DVD-ROM, 기타 광학 매체, 펀치 카드 종이 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, 이러한 반송파를 전송하는 케이블 또는 링크, 또는 컴퓨터에서 읽을 수 있는 기타 매체 프로그래밍 부호 및/또는 데이터. 이러한 형태의 컴퓨터 판독 가능 매체 중 다수는 실행을 위한 프로세서에 하나 이상의 명령어의 하나 이상의 시퀀스를 전달하는 것과 관련될 수 있다.
- [0133] 컴퓨터 시스템(201)은 예를 들어, 핵산 분자 염기서열분석의 결과를, 제공하기 위한 사용자 인터페이스(UI)(240)를 포함하는 전자 디스플레이(235)를 포함하거나 이와 통신할 수 있다. UI의 예에는 그래픽 사용자 인터페이스(Graphical User Interface; GUI) 및 웹 기반 사용자 인터페이스가 포함되나, 이에 제한되는 것은 아니다.
- [0134] 본 발명의 방법 및 시스템은 하나 이상의 알고리즘을 통해 구현될 수 있다. 알고리즘은 중앙 처리부(205)에 의해 실행되는 경우 소프트웨어를 통해 구현될 수 있다. 알고리즘은, 예를 들어, 저장을 위한 장착 가능한 랙과 염기서열분석을 위한 기기로 기관이 이동되는 속도를 생성할 수 있다.
- [0135] 본 발명의 바람직한 실시예들은 본 명세서에 도시되고 설명되었으나, 그러한 실시예는 단지 예시로서 제공되는 것이 본 발명의 기술분야에 속한 통상의 기술자에게 명백할 것이다. 본 발명은 명세서 내에 제공된 특정 예시에 의해 제한되는 것으로 의도되는 것은 아니다. 본 발명은 전술한 명세서를 참조하여 설명되었으나, 본 명세서의 실시예의 설명 및 예시는 제한적인 의미로 해석되어서는 안 된다. 수많은 변형, 변경 및 대체는 본 발명을 벗어나지 않고 본 발명의 기술분야에 속한 통상의 기술자에게 발생할 것이다. 또한, 본 발명의 모든 태양들은 다양한 조건 및 변수에 의존하는 본 명세서에 기재된 특정 묘사, 구성 또는 상대적 비율에 제한되지 않는다는

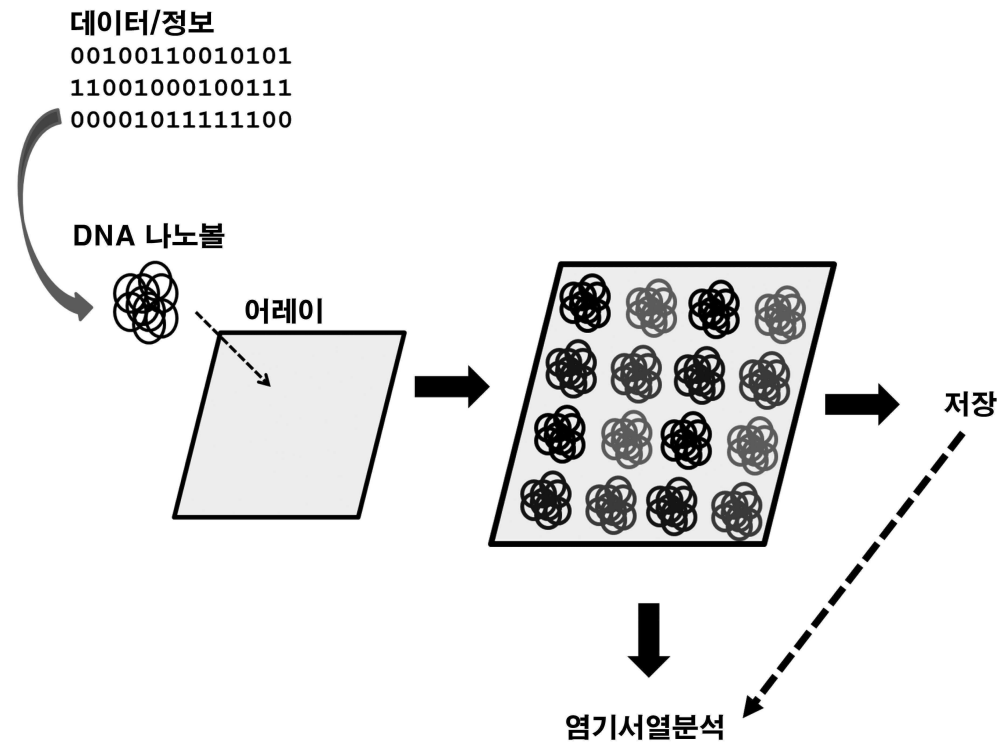
것을 이해해야 한다. 본 명세서에 기술된 본 발명의 실시예에 대한 다양한 대안이 본 발명을 실시하는데 이용될 수 있음을 이해해야 한다. 따라서, 본 발명은 그러한 대안, 수정, 변형 또는 균등물도 포함하는 것으로 고려된다. 청구범위는 본 발명의 범위를 정의하고 이러한 청구범위 및 그 균등물의 범위 내의 방법 및 구조는 이에 의해 포함되는 것으로 의도된다.

[0136] 본 명세서에 제공된 방법 및 시스템은 예를 들어, 미국 특허 공개 번호 20150330974 및 20180274028 설명된 것과, 같은, 기타 방법 및 시스템과 결합되거나 이에 의해 수정될 수 있고, 이들 각각은 참조로 본 명세서에 통합된다.

[0137] 본 발명의 바람직한 실시예는 본 명세서에 도시되고 설명되었으나, 그러한 실시예는 단지 예시로서 제공된다는 것이 본 발명의 기술분야에 속한 통상의 기술자에게 명백할 것이다. 본 발명은 명세서 내에 제공된 특정 예에 의해 제한되는 것으로 의도되는 것은 아니다. 본 발명은 전술한 명세서를 참조하여 설명되었으나, 본 명세서의 실시예에 대한 설명 및 예시는 제한적인 의미로 해석되어서는 안 된다. 수많은 변형, 변경 및 대체 사항들은 이제 본 발명을 벗어나지 않고 본 발명의 기술분야에 속한 통상의 기술자에게 발생할 것이다. 또한, 본 발명의 모든 태양들은 다양한 조건 및 변수에 의존하는 본 명세서에 기재된 특정 묘사, 구성 또는 상대적 비율에 제한되지 않는다는 것을 이해해야 한다. 본 명세서에 기술된 본 발명의 실시예에 대한 다양한 대안들은 본 발명을 실시하는 데 사용될 수 있음을 이해해야 한다. 따라서, 본 발명은 그러한 대안, 수정, 변형 또는 등가물 또한 포함할 것으로 예상된다. 다음 청구범위는 본 발명의 범위를 정의하고 이러한 청구범위 및 그 균등물의 범위 내의 방법 및 구조는 이에 의해 포함되는 것으로 의도된다.

**도면**

**도면1**



도면2

