



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년09월25일
(11) 등록번호 10-2583062
(24) 등록일자 2023년09월21일

- (51) 국제특허분류(Int. Cl.)
G06N 3/12 (2023.01) B01J 19/00 (2018.01)
C12Q 1/6869 (2018.01) G11B 7/245 (2006.01)
G11C 7/00 (2006.01) G16B 30/10 (2019.01)
G16B 50/30 (2019.01)
- (52) CPC특허분류
G06N 3/123 (2023.01)
B01J 19/0046 (2013.01)
- (21) 출원번호 10-2019-7038747
- (22) 출원일자(국제) 2018년05월31일
심사청구일자 2021년05월31일
- (85) 번역문제출일자 2019년12월30일
- (65) 공개번호 10-2020-0027927
- (43) 공개일자 2020년03월13일
- (86) 국제출원번호 PCT/US2018/035365
- (87) 국제공개번호 WO 2018/222853
국제공개일자 2018년12월06일
- (30) 우선권주장
62/513,111 2017년05월31일 미국(US)
- (56) 선행기술조사문헌
US20170017436 A1*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
몰레큘러 어셈블리즈, 인크.
미국 92121 캘리포니아주 샌 디에고 사이언스 센터 드라이브 10355 스위트 130
- (72) 발명자
에프카비치, 윌리엄, 제이.
미국 94070 캘리포니아주 샌 카를로스 아로요 애비뉴 2089
홀덴, 매튜, 티.
미국 92121 캘리포니아주 샌 디에고 사이언스 센터 드라이브 10355 스위트 130
- (74) 대리인
양영준, 이상남

전체 청구항 수 : 총 22 항

심사관 : 양대경

(54) 발명의 명칭 동종중합체 인코딩된 핵산 메모리

(57) 요약

반복된 뉴클레오티드의 동종중합체 트랙의 시퀀스를 사용하여 디지털 데이터를 인코딩하는 핵산 메모리 가닥은 통상적인 디지털 DNA 저장 기술에 대해 보다 저렴하고 보다 신속한 대안을 제공한다. 동종중합체 트랙의 사용은 메모리 가닥에 인코딩된 데이터를 판독하기 위한 나노세공 시퀀싱과 같은 보다 낮은 충실도, 높은 처리량 시퀀싱 기술을 허용한다. 특수화된 합성 기술은 통상적인 단일 뉴클레오티드 시퀀스와 비교하여 동종중합체 트랙에 의해 제공되는 감소된 데이터 밀도에도 불구하고 대량의 데이터를 인코딩할 수 있는 긴 메모리 가닥의 합성을 허용한다.

대표도 - 도1



(52) CPC특허분류

C12Q 1/6869 (2018.05)

G11B 7/245 (2013.01)

G11C 7/00 (2013.01)

G16B 30/10 (2019.02)

G16B 50/30 (2019.02)

명세서

청구범위

청구항 1

데이터세트를 나타내는 인 실리코 비트의 시퀀스를 생성하는 단계;

규정된 조성을 갖는 복수개의 동종중합체 트랙 또는 이종중합체 트랙을 포함하는 핵산 메모리 가닥을 합성하는 단계이며, 각각의 동종중합체 또는 이종중합체 트랙은 데이터세트를 나타내는 비트의 시퀀스의 비트에 상응하고, 상기 데이터세트의 각각의 유닛은 베이스 2로 나타내지는 것인 단계

를 포함하는, 핵산 메모리 가닥을 사용하여 데이터를 기록하는 방법.

청구항 2

제1항에 있어서, 복수개의 동종중합체 트랙이 2 내지 10개의 반복된 뉴클레오티드를 포함하는 것인 방법.

청구항 3

삭제

청구항 4

데이터세트를 나타내는 인 실리코 비트의 시퀀스를 생성하는 단계;

규정된 조성을 갖는 복수개의 동종중합체 트랙 또는 이종중합체 트랙을 포함하는 핵산 메모리 가닥을 합성하는 단계이며, 각각의 동종중합체 또는 이종중합체 트랙은 데이터세트를 나타내는 비트의 시퀀스의 비트에 상응하고, 상기 데이터세트의 각각의 유닛은 베이스 3 또는 그 초과로 나타내지는 것인 단계

를 포함하는, 핵산 메모리 가닥을 사용하여 데이터를 기록하는 방법.

청구항 5

제1항에 있어서, 핵산 메모리 가닥이 길이가 적어도 약 200개의 뉴클레오티드 내지 길이가 약 5,000개의 뉴클레오티드인 방법.

청구항 6

제1항에 있어서, 상기 합성 단계가 dNTP 농도를 변화시킴으로써 트랙 길이를 제어하는 단계를 포함하는 것인 방법.

청구항 7

제1항에 있어서, 상기 합성 단계가 반응 시간을 변화시킴으로써 트랙 길이를 제어하는 단계를 포함하는 것인 방법.

청구항 8

제1항에 있어서, 복수개의 트랙이 둘 이상의 상이한 뉴클레오티드를 포함하는 것인 방법.

청구항 9

제1항에 있어서, 상기 합성 단계가 dNTP 비율을 변화시킴으로써 트랙 조성을 제어하는 단계를 포함하는 것인 방법.

청구항 10

제1항에 있어서,

핵산 메모리 가닥의 제1 단부를 변형시켜 나노세공 시퀀싱 시스템의 나노세공을 통한 제1 단부의 통과를 방지하

는 단계;

핵산 메모리 가닥의 제2 단부를 나노세공을 통해 통과시키는 단계; 및

핵산 메모리 가닥의 제2 단부를 변형시켜 나노세공을 통한 제2 단부의 통과를 방지하여, 나노세공에서 핵산 메모리 가닥을 수집하는 단계

를 추가로 포함하는 방법.

청구항 11

제1항에 있어서, 데이터세트가 텍스트 파일, 이미지 파일, 및 오디오 파일로 이루어진 군으로부터 선택되는 것인 방법.

청구항 12

제1항에 있어서, 상기 합성 단계가 템플릿-비의존적 합성을 포함하는 것인 방법.

청구항 13

제12항에 있어서, 뉴클레오티드 트랜스퍼라제 효소를 사용하여 상기 템플릿-비의존적 합성을 촉매하는 것인 방법.

청구항 14

제12항에 있어서, 폴리머라제 세타를 사용하여 상기 템플릿-비의존적 합성을 촉매하는 것인 방법.

청구항 15

제1항에 있어서, 핵산 메모리 가닥이 핵산 메모리 가닥의 데이터세트로의 변환을 방지하는 공유 결합된 화학 보호기를 포함하는 것인 방법.

청구항 16

규정된 조성을 갖는 복수개의 동중중합체 트랙 또는 이중중합체 트랙을 포함하는 핵산 메모리 가닥을 시퀀싱하는 단계;

핵산 메모리 가닥 시퀀스를 디지털화된 데이터로 변환하는 단계이며, 여기서 복수개의 동중중합체 또는 이중중합체 트랙 각각은 데이터 유닛에 상응하는 비트를 나타내는 것인 단계; 및

디지털화된 데이터 조각을 판독가능한 포맷으로 변환하는 단계

를 포함하는, 핵산 메모리 가닥으로부터 데이터를 판독하는 방법.

청구항 17

제16항에 있어서, 판독가능한 포맷을 디스플레이하는 단계를 추가로 포함하는 방법.

청구항 18

제16항에 있어서, 복수개의 동중중합체 트랙이 약 2 내지 약 10개의 뉴클레오티드 반복부를 포함하는 것인 방법.

청구항 19

제16항에 있어서, 데이터의 유닛이 베이스 2로 나타내지는 것인 방법.

청구항 20

제16항에 있어서, 데이터의 유닛이 베이스 3으로 나타내지는 것인 방법.

청구항 21

제16항에 있어서, 데이터의 유닛이 베이스 4로 나타내지는 것인 방법.

청구항 22

제16항에 있어서, 핵산 메모리 가닥이 길이가 적어도 약 200개의 뉴클레오티드 내지 길이가 약 5,000개의 뉴클레오티드인 방법.

청구항 23

제16항에 있어서, 핵산 메모리 가닥이 핵산 메모리 가닥의 데이터세트로의 변환을 방지하는 공유 결합된 화학 보호기를 포함하고, 방법이 핵산 메모리 가닥을 시퀀싱하기 전에 화학 보호기 중 하나 이상을 제거하는 단계를 추가로 포함하는 것인 방법.

발명의 설명

기술 분야

[0001] 관련 출원

[0002] 본 출원은 2017년 5월 31일에 출원된 미국 가특허출원 일련번호 62/513,111을 우선권 주장하며, 이 가출원의 내용은 그 전문이 본원에 참조로 포함된다.

[0003] 발명의 분야

[0004] 본 발명은 동종중합체 트랙을 포함하는 핵산 메모리 가닥에 데이터를 저장하기 위한 방법 및 장치에 관한 것이다.

배경 기술

[0005] DNA 디지털 저장은 DNA의 염기 서열을 사용하여 디지털 데이터를 나타내고 데이터를 인코딩하는 염기 서열에 상응하는 폴리뉴클레오티드의 DNA 합성을 통해 그 데이터를 저장하는 프로세스이다. DNA 디지털 저장은 통상적인 데이터 저장 방법에 비해 몇몇 이점을 제공하며 수천억 달러의 시장을 목표로 한다. 플래시 메모리 및 자성 테이프 상에 기록을 포함한 통상적인 데이터 저장 방법은 물리적 공간 요건, 희소 자원에 대한 의존성, 및 데이터 무결성과 관련된 문제를 일으킨다. DNA 디지털 저장은 상당히 더 낮은 에너지 요건으로 훨씬 더 큰 데이터 저장 밀도를 제공한다. 현재의 방법은 DNA에서 인코딩된 데이터를 정확하게 판독하기 위해 허용 오차가 거의 없는 고-충실도 시퀀싱 기술에 의존한다. 필요로 하는 시퀀싱 방법은 충실도 요건을 충족시키기 위해 비교적 느리고 비용이 많이 든다. 현재의 DNA 디지털 저장 기술의 예는 처치(Church) 등의 미국 특허 번호 9,384,320에 기재되어 있다 (본원에 참조로 포함됨). 시퀀싱 충실도를 증가시키기 위해, 처치에 의해 기재된 것들과 같은 현재의 방법은 시퀀스 반복부와 같이 판독하거나 기입하기 곤란한 특징부를 피하는 시퀀스를 사용하여 데이터를 인코딩한다.

[0006] 현재의 DNA 디지털 저장 기술의 합성 측면은 속도 부족, 독성 부산물의 생성, 및 고비용을 통해 기술의 채택을 추가로 제한한다. 대부분의 드 노보 핵산 서열은 천연 (또는 비천연) 핵산 염기에 상응하는 포스포르아미다이트 시약으로부터 구축된 서열의 순차적 탈보호 및 합성을 수반하는 고체상 포스포르아미다이트-기술을 사용하여 합성된다. 어레이-기반 포맷에서의 잉크젯 합성은 매우 저비용의 포스포르아미다이트 합성이 가능하나, 제조되는 가닥은 길이가 100-200개 염기로 제한되고, 인덱스 서열까지의 일부 길이를 희생하여야 하며, 펨토몰 이하(sub-femtomolar) 규모로 제조되어 후속 판독을 위한 충분한 물질을 제공하기 위해 합성 후 증폭을 필요로 한다. 통상적인 합성 기술을 사용하여, 길이가 200개 염기 쌍(bp) 초과인 핵산은 높은 속도의 파손 및 부반응을 경험한다. 게다가, 통상적인 합성 기술은 독성 부산물을 생성하고, 이 폐기물의 처분은 핵산 합성기의 이용 가능성을 제한하고 오히려 생산 비용을 증가시킨다. DNA 디지털 저장에서의 합성 및 판독과 관련된 이들 문제는 달리 유망한 기술에 대한 응용을 제한하였다.

발명의 내용

[0007] 본 발명은 디지털 데이터를 인코딩하는 동종중합체 트랙의 시퀀스를 사용하여 데이터를 저장하기 위한 시스템 및 방법을 제공한다. 반복된 염기 (예를 들어, 2-10개 뉴클레오티드)의 동종중합체 트랙을 사용한 데이터 시퀀스에서 각각의 비트를 나타내는 것은 보다 높은 처리량 및 보다 저렴한 시퀀싱 기술이 사용될 수 있게 한다. 시퀀스 판독은 동종중합체 트랙 사이의 전이를 식별하는 것에만 의존하고 각각의 개별 뉴클레오티드의 충실한 판독을 필요로 하지 않기 때문에, 시퀀싱 기술 예컨대 나노세공(nanopore) 시퀀싱, 제로-모드 도파관(ZMW) 단

일 분자 시퀀싱, 및 질량 분석법을 사용하여 속도를 증가시키고 비용을 감소시킬 수 있다.

- [0008] 본원에 기재된 바와 같은 동중중합체 트랙을 사용한 데이터의 기록은 핵산의 긴 가닥 (예를 들어, 5-10 kb)을 사용하여 가장 효율적으로 완수된다. 전통적인 합성 기술은 길이 제한적이지만, 예를 들어, 뉴클레오티드 트랜스퍼라제를 사용하는 템플릿-비의존적 폴리뉴클레오티드 합성은 감소된 비용으로 그리고 보다 낮은 폐기물 생산으로 긴 가닥을 합성할 수 있다. 효율적으로 합성된 ssDNA 메모리 가닥은 길이가 약 100 - 200개 뉴클레오티드보다 더 긴 ssDNA 가닥이 복잡하고 값비싼 라이게이션 또는 PCR 기술을 필요로 하고 dsDNA 중간체로부터 ssDNA를 단지 생산할 수 있기 때문에 통상적인 포스포르아미다이트 접근법과 비교하여 DNA 합성의 50%를 단지 필요로 한다. 본원에 참조로 포함된 에프캐비치(Efcavitch) 등의 미국 특허 번호 8,808,989를 참조한다. 데이터 인코딩은 표준 뉴클레오티드를 사용하여 기수(numerical base) 2, 3, 4에서 일 수 있거나 베이스 8, 10, 12, 또는 그 초과인 인코딩 체계를 생성하기 위해 임의의 수의 변형된 뉴클레오티드 유사체를 사용하여 데이터 밀도를 증가시킬 수 있다.
- [0009] 변형된 뉴클레오티드 유사체에 대한 제한은 이들이 선택된 합성 기술 (예를 들어, 말단 데옥시뉴클레오티드 트랜스퍼라제 (TdT))을 사용하여 혼입될 수 있고 선택된 시퀀싱 분석을 사용하여 서로 구별될 수 있다는 점이다. 일부 실시양태에서, 합성은 Mn^{2+} 의 존재하에 폴리머라제 세타를 사용하여 완수될 수 있다.
- [0010] 일관된 동중중합체 트랙 길이는 인식될 필요가 있는 개별 트랙 사이의 단지 전이이기 때문에 본 발명의 시스템 및 방법에 필수적이지 않다. 비록 트랙 길이가 변할 수 있을지라도, 본 발명의 합성 기술은 합성되는 올리고뉴클레오티드 메모리 가닥에 대한 데옥시뉴클레오티드 (dNTP)의 비를 조정하고 dNTP의 신생(nascent) 메모리 가닥으로의 노출 시간을 제어함으로써 평균 동중중합체 트랙 길이를 효과적으로 제어할 수 있다. 동중중합체 트랙의 길이는 관독 기술에 최적화될 수 있으며; 최고 데이터 저장 밀도는 단일 뉴클레오티드 관독 해상도로 달성되나, 최고 관독 속도 및 정확도는 주어진 시퀀싱 기술의 경우 뉴클레오티드 비트의 크기를 최소 검출가능한 길이 (예를 들어, 2-10 뉴클레오티드)로 확장함으로써 달성가능하다.
- [0011] 나노세공 시퀀싱을 사용하는 본 발명의 시스템 및 방법은 가닥의 한쪽 또는 양쪽 말단 상에 포함된 스토퍼와 같은 특수화된 메모리 가닥 구조물 (예를 들어, 헤어핀 또는 거대분자 부속물)을 사용할 수 있다. 다른 나노세공-기반 방법에서, 메모리 가닥은 원형화되고 인접한 나노세공들 사이에 스레드(thread)될 수 있다.
- [0012] 본 발명의 특정 측면은 핵산 메모리 가닥을 사용하여 데이터를 기록하는 방법을 포함한다. 방법의 단계는 올리고뉴클레오티드 시퀀스의 각각의 뉴클레오티드가 상기 데이터세트의 유닛에 상응하는 데이터세트를 나타내는 인실리코(in-silico) 올리고뉴클레오티드 시퀀스를 생성하는 단계를 포함할 수 있다. 이어서, 각각의 동중중합체 트랙이 올리고뉴클레오티드 시퀀스의 뉴클레오티드에 상응하는 복수개의 동중중합체 트랙을 포함하는 핵산 메모리 가닥을 합성할 수 있다. 복수개의 동중중합체 트랙은 3 내지 10개의 반복된 뉴클레오티드를 포함할 수 있다. 상기 데이터세트의 각각의 유닛은 특정 적용을 위해 원하는 바와 같이 베이스 2, 베이스 3, 베이스 4, 또는 그 초과로 나타내질 수 있다.
- [0013] 특정 실시양태에서, 핵산 메모리 가닥은 길이가 적어도 약 200개의 뉴클레오티드 내지 길이가 약 5,000개의 뉴클레오티드일 수 있다. 합성 단계는 dNTP 농도를 변화시킴으로써 동중중합체 길이를 제어하는 단계를 포함할 수 있다. 방법의 단계는 핵산 메모리 가닥의 제1 단부를 변형시켜 나노세공 시퀀싱 시스템의 나노세공을 통한 제1 단부의 통과를 방지하는 단계; 핵산 메모리 가닥의 제2 단부를 나노세공을 통해 통과시키는 단계; 및 핵산 메모리 가닥의 제2 단부를 변형시켜 나노세공을 통한 제2 단부의 통과를 방지하는 단계를 포함할 수 있다.
- [0014] 다른 실시양태는 정해진 수의 뉴클레오티드 유사체의 코딩 용량을 추가로 증가시키기 위해 규정된 화학량론 또는 조성의 이중중합체 트랙으로 구성된 메모리 가닥을 이용할 수 있다. 추가 실시양태는 구조상 유사하나 상이한 조건 예컨대 자외선 또는 가시광선, 산화제 또는 환원제, 알칼리성 또는 산성 pH, 또는 서열 특이적 뉴클레아제하에 제거될 수 있는 링커를 이용하는 뉴클레오티드 유사체를 사용하여, 적용가능한 프로세스에 대한 지식 없이 데이터를 그들에 위장함으로써 메모리 가닥에 인코딩된 데이터를 보호하고자 할 수 있다.
- [0015] 데이터세트는 텍스트 파일, 이미지 파일, 및 오디오 파일로 이루어진 군으로부터 선택될 수 있다. 합성 단계는 템플릿-비의존적 합성을 포함할 수 있다. 특정 실시양태에서, 뉴클레오티드 트랜스퍼라제 효소를 사용하여 상기 템플릿-비의존적 합성을 촉매할 수 있다. 일부 실시양태에서, 폴리머라제 세타를 사용하여 상기 템플릿-비의존적 합성을 촉매할 수 있다.
- [0016] 본 발명의 측면은 핵산 메모리 가닥으로부터 데이터를 관독하는 방법을 포함할 수 있다. 방법의 단계는 복수개의 동중중합체 트랙을 포함하는 핵산 메모리 가닥을 시퀀싱하는 단계; 핵산 메모리 가닥 시퀀스를 디지털화된

데이터로 변환하는 단계이며, 여기서 복수개의 동종중합체 트랙 각각은 데이터 유닛에 상응하는 뉴클레오티드를 나타내는 것인 단계; 및 디지털화된 데이터 조각을 판독가능한 포맷으로 변환하는 단계를 포함할 수 있다. 방법의 단계는 판독가능한 포맷을 디스플레이하는 단계를 포함할 수 있다. 복수개의 동종중합체 트랙은 약 2 내지 약 10개의 뉴클레오티드 반복부를 포함할 수 있다. 핵산 메모리 가닥은 길이가 적어도 약 200개의 뉴클레오티드 내지 약 5,000개의 뉴클레오티드일 수 있다.

[0017] 다양한 실시양태에서, 시퀀싱 단계는 나노세공 시퀀싱, 합성에 의한 시퀀싱, 또는 질량 분석법을 포함할 수 있다. 시퀀싱, 번역, 및 변환 단계는 핵산 메모리 가닥 상에서 1회 이상 반복될 수 있다.

[0018] 본 발명의 다른 측면은 하기 도면 및 상세한 설명의 고려시 통상의 기술자에게 명백하다.

도면의 간단한 설명

- [0019] 도 1은 동종중합체 트랙을 형성하는 데 사용되는 효소적 합성 사이클을 나타낸다.
- 도 2는 동종중합체 트랙을 가진 핵산 메모리 가닥을 합성하는 방법을 나타낸다.
- 도 3은 동종중합체 트랙을 가진 핵산 메모리 가닥으로부터 데이터를 판독하는 방법을 나타낸다.
- 도 4는 가닥/GB와 동종중합체 길이 및 가닥 길이의 관계를 나타낸다.
- 도 5는 메모리 가닥에서 구별가능한 뉴클레오티드 유사체의 수의 함수로서 500개의 중합체 트랙으로 구성된 DNA 가닥에 인코딩될 수 있는 데이터를 나타낸다.
- 도 6은 본 발명의 나노세공 포획된 핵산 메모리 가닥을 나타낸다.
- 도 7은 처리 조건에 대응하여 동종중합체 트랙의 가닥 내에 인코딩된 데이터를 변화시키기 위한 도식을 나타낸다.
- 도 8은 동종중합체 트랙을 가진 핵산 메모리 가닥을 합성하기 위한 시스템을 나타낸다.
- 도 9는 나노웰의 어레이 상에서 동종중합체 트랙을 가진 핵산 메모리 가닥의 병렬 합성을 위한 시스템을 나타낸다.
- 도 10은 시스템 내에 나타날 수 있는 구성요소의 보다 상세한 개략도를 제공한다.
- 도 11은 2개의 상이한 염기 조성 동종중합체 트랙의 효소-매개 합성의 분석을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0020] 본 발명은 디지털 데이터의 유닛에 상응하는 동종중합체 트랙을 갖는 핵산에 데이터를 기입하고 상기 핵산으로부터 데이터를 판독하기 위한 시스템 및 방법을 제공한다. 데이터-인코딩 시퀀스에서 각각의 뉴클레오티드를 반복 (예를 들어, 3-10 회)함으로써, 핵산 데이터 저장에서 보다 저렴한 실행을 결과할 수 있는 보다 낮은 충실도, 보다 높은 처리량 시퀀싱 기술을 허용하는 시퀀싱 판독에서 동종중합체 트랙 사이의 전이만이 관찰될 필요가 있다. 핵산 동종중합체 트랙 메모리 가닥을 합성하는 이점은 다음과 같다: 1) 판독을 위해 높은 처리량, 긴 판독 DNA 시퀀싱 기술의 사용을 가능하게 하는, 매우 긴 (5 - 10 kb) 가닥을 만드는 능력, 2) 시퀀싱 판독 기술에서 오류를 허용하는 능력 및 3) 통상적인 화학적 합성 방법의 것보다 훨씬 더 적은 비용으로 핵산 메모리 가닥을 제조하는 능력. 동종중합체 핵산 메모리 가닥의 사용은 템플릿-비의존적 TdT 효소 또는 폴리머라제 세타를 사용하여 효율적으로 생산될 수 있는 긴 (예를 들어, 5-10 kb) 가닥에서 가장 잘 실현되며, 여기서 동종중합체 트랙 길이는 노출 시간 및 dNTP 대 폴리뉴클레오티드 메모리 가닥 비를 변경함으로써 제어될 수 있다.

[0021] 효소적으로 매개된 접근법에 의해 데이터를 인코딩하기 위한 동종중합체의 합성은 종결자가 아닌 천연 또는 변형된 뉴클레오티드 트리포스페이트를 사용함으로써 쉽게 달성되어, 가능한 가장 단순하고 가장 신속한 DNA 합성 방법을 결과한다. 하나의 천연 또는 변형된 뉴클레오티드 트리포스페이트는 뉴클레오티딜 트랜스퍼라제를 사용하여 반응 구역으로 전달되고, 반응한 다음에, 완충제로 세척함으로써 제거되어 도 1에 도시된 바와 같이 데이터 저장의 하나의 "기입(write)" 사이클을 완료한다. 데이터 가닥 합성은 독성 또는 유해 화학 물질이 전혀 없는 전적으로 수성인 환경에서 발생하므로, 대규모 데이터 저장 센터에 적합한 실용적인 장치를 가능하게 한다.

[0022] 도 2는 특정 실시양태에 따라 동종중합체 트랙을 가진 핵산 메모리 가닥을 합성하는 방법 (101)을 나타낸다. 방법 (101)은 데이터세트를 나타내는 인 실리코 올리고뉴클레오티드 시퀀스 (103)를 생성하는 단계를 포함한다.

데이터세트는 텍스트, 이미지, 비디오, 오디오, 또는 디지털화될 수 있는 임의의 다른 단편의 정보를 나타낼 수 있는 디지털화된 데이터를 포함할 수 있다. 올리고뉴클레오타이드 시퀀스는 임의의 수의 천연 또는 변형된 뉴클레오타이드 또는 그의 유사체를 포함할 수 있고, 메모리 가닥에서 사용된 특유의 뉴클레오타이드 또는 유사체의 수에 따라 베이스 2, 베이스 3, 베이스 4, 또는 그 초과와 체계를 사용하여 데이터세트를 인코딩할 수 있다. 간단한 실시양태에서, 인코딩 체계는 통상적으로 일련의 0 및 1로 나타내지는 이진 데이터 체계에 상응할 수 있고 여기서 하나 이상의 뉴클레오타이드 또는 유사체는 0에 상응할 수 있고 하나 이상의 다른 뉴클레오타이드는 1에 상응할 수 있다. 이어서, 인 실리코 올리고뉴클레오타이드 시퀀스에서, 순번으로, 뉴클레오타이드에 각각 상응하는 일련의 동종중합체 트랙을 포함하는 핵산 메모리 가닥 (예를 들어, RNA, 단일 가닥, 또는 이중 가닥 DNA)이 합성될 수 있다 (105). 특정 실시양태에서, 추가 단계는 메모리 가닥의 한 단부를 변형하는 단계 (107), 나노세공을 통해 메모리 가닥을 스레드하는 단계 (109) 및 단부가 나노세공을 통과하는 것을 방지하기 위해 가닥의 다른 한 단부를 변형하는 단계 (111)를 포함한다.

[0023] 도 3은 동종중합체 트랙을 가진 핵산 메모리 가닥으로부터 데이터를 판독하는 방법 (203)을 나타낸다. 방법 (203)의 단계는 핵산 메모리 가닥에서 일련의 동종중합체 트랙을 시퀀싱하는 단계 (203), 시퀀스를 데이터세트로 변환하는 단계 (205), 데이터세트를 판독가능한 포맷 (예를 들어, 이미지, 비디오, 오디오 클립, 또는 한편의 텍스트)으로 변환하는 단계, 및 임의로 (예를 들어, 모니터 상에 또는 프린터 또는 다른 입력/출력 장치를 사용하여) 데이터의 판독가능한 포맷을 디스플레이하는 단계 (209)를 포함한다.

[0024] 바람직하게는, 본 발명의 시스템 및 방법은 단일 가닥 또는 이중 가닥이고 자연적으로 발생하거나 화학적 또는 효소적 합성에 의해 생성될 수 있는 긴 가닥의 DNA (5-10 kb)를 사용한다. 특정 실시양태에서, 핵산 메모리 가닥은 TdT를 사용하여 효소적으로 생성되어 2-10, 3-10, 4-10 개 뉴클레오타이드 또는 그 초과일 수 있는 일련의 동종중합체 트랙을 생성할 수 있다. 동종중합체 트랙은 각각 아데닌 (A), 구아닌 (G), 시토신 (C), 또는 티민 (T)으로 이루어질 수 있다. 교호 동종중합체 트랙의 시퀀스는 메모리 가닥에 저장될 데이터를 인코딩하는 데 사용될 수 있다.

[0025] 각각의 뉴클레오타이드 동종중합체 트랙은 사용된 염기의 수에 따라 다양한 양의 데이터를 나타낼 수 있다. 1 바이트를 만드는 데 필요한 비트의 수 (10진 256)는 하기 관계에 의해 정의된다: #비트/바이트 = $8/(\log_2(n))$, 여기서 n = 사용되는 기수. 각각의 트랙은 2 베이스 인코딩이 사용되는 경우 1 비트에 상응할 수 있거나 4 베이스 인코딩이 사용되는 경우 바이트의 1/4에 상응할 수 있다. 특정 실시양태에서, DNA 데이터 가닥은 2-10 개의 뉴클레오타이드 동종중합체 트랙 (베이스 2 데이터세트 표시를 사용하여)으로 구성될 수 있으며, 이는 333 비트 내지 100 비트가 999 내지 1000개의 염기 길이의 메모리 가닥으로 나타날 수 있게 할 것이다. 바람직한 실시양태에서, 데이터의 핵산 염기 인코딩은 하나의 뉴클레오타이드의 단일 동종중합체 트랙이 항상 상이한 뉴클레오타이드의 동종중합체 트랙에 인접하도록 하는 것일 수 있다. 예를 들어, 인코딩은 아데닌의 동종중합체 트랙이 또 다른 아데닌 동종중합체 트랙에 바로 앞에 또는 그에 뒤이어 있지 않도록 하는 것일 수 있다. 2개의 인접한 동종중합체 트랙이 동일한 뉴클레오타이드를 포함하는 경우에, 인코딩된 데이터 시퀀스에서 단일 뉴클레오타이드를 나타내는 평균 동종중합체 트랙보다 측정할 수 있게 더 긴 동종중합체 트랙이 합성될 수 있다. 그러한 보다 긴 트랙은, 예를 들어, 반응에서 dNTP의 농도를 증가시키거나 반응 시간을 증가시킴으로써, 하기 기재된 합성 반응의 조작을 통해 생성될 수 있다. 2개의 인접한 동일한 동종중합체 트랙의 정확한 길이는 판독 장치 (즉, 나노세공 시퀀서)를 사용하여 단일 동종중합체 트랙과 명확하게 구별되기에 충분할 정도로 단지 길 필요가 있다. 특정 실시양태에서, 비-뉴클레오타이드 동종중합체 스페이서는 A, G, C, 또는 T 동종중합체 트랙 사이에 첨가되어 인접한 동일한 뉴클레오타이드 동종중합체 트랙을 서로 분명히 구별할 수 있을 것이다. A, G, C, 및 T 동종중합체 트랙의 사용은 사 (4) 비트 인코딩 공간의 생성을 가능하게 하여, 단순히 2개의 뉴클레오타이드 (이진 코드로 0 및 1과 유사)를 사용하기 보다는 하나의 인접한 가닥에 저장될 수 있는 데이터 밀도를 증가시킨다. 예를 들어, A, G, C, 및 T가 베이스 4 체계로 사용되는 경우, 4개의 인접한 동종중합체 트랙은 256 숫자 (즉, 1 바이트)를 인코딩할 수 있다. 이러한 실시양태에서, 삼 (3) 뉴클레오타이드 길이의 동종중합체 트랙 또는 십 (10) 뉴클레오타이드 길이의 동종중합체 트랙이 각각 사용되는 경우 996 또는 1000 뉴클레오타이드 길이의 핵산 메모리 가닥에 83 바이트 또는 25 바이트가 나타내질 수 있다.

[0026] 다양한 실시양태에서, 베이스 8 또는 심지어 베이스 12 코딩 체계는 특유하게 변형된 뉴클레오타이드 또는 비-뉴클레오타이드 유사체의 동종중합체 트랙의 메모리 가닥으로의 혼입을 통해 이용될 수 있다. 그러한 변형된 뉴클레오타이드 또는 비-뉴클레오타이드 유사체는 판독 장치 예컨대 나노세공 시퀀서 또는 단일 분자 ZMW 시퀀서로 특유의 디지털 신호를 생성하여야 한다. 하기 논의된 바와 같이, TdT는 나노세공과 같은 판독 장치에 의해 제공되는 신호를 향상시킬 수 있는 매우 다양한 변형된 dNTP 유사체를 혼입할 수 있고, 따라서 그들 중에서 인코딩된

데이터를 가진 핵산 메모리 가닥을 생성하는 데 사용될 수 있다. 변형된 뉴클레오타이드의 동중중합체 (예를 들어, A*, G*, C* 및 T* 또는 A**, G**, C** 및 T**)는 TdT 및 또는 팔 (8) 비트 또는 심지어 십이 (12) 비트 인코딩 체계를 생성하기 위해 4개의 염기 각각의 변형된 dNTP 유사체 (예를 들어, dA*TP 또는 dA**TP)를 사용하여 합성될 수 있다. 보다 높은 베이스 (n) 인코딩은 데이터 압축을 허용하고 주어진 양의 정보를 인코딩하는 데 필요한 DNA 가닥의 수의 감소를 결과한다. 동중중합체 트랙의 길이, 기수 (n), 및 합성된 가닥 길이의 함수로서 데이터의 GB 당 DNA 가닥의 수를 결정하는 관계는 다음에 의해 정의된다: 도 4에 도시된 바와 같은 #가닥/GB = $(8/(\log_2(n)) * 10^9 * \text{동중중합체 길이} * 1/\text{가닥 길이})$.

[0027] 특유의 동중중합체 트랙의 수는 하나의 동중중합체 트랙을 또 다른 것으로부터 결정하는 판독 기술 (즉, 나노세공 또는 ZMW 단일 분자 시퀀싱)의 능력에 의해서만 제한될 수 있다. 나노세공을 통한 전위 동안에 이온 전류의 변화를 검출함으로써 변형되지 않은 뉴클레오타이드로 구성된 동중중합체의 검출에 관한 문헌에 몇몇 보고가 있다 (Venta et al, 2013; Feng et al, 2015). 나노세공에서 DNA의 체류 시간을 변경하는 변형은 구별가능하고 특징적인 이온-전류 신호를 생성할 것이다. 문헌 [Singer et al. 2010] 및 [Morin et al. 2016]은 5 또는 10kDa PEG로 기능화된 비공유 결합된 bisPNA 또는 γ PNA를 사용하여 나노세공에 의한 검출을 향상시킨다. 문헌 [Liu et al. 2015]은 체류 시간을 변형하고 특유의 신호를 생성하기 위해 확고히 8-옥소G 유사체를 선택적으로 생성하였다. dATP의 N6, dCTP의 N4, dGTP의 N2 또는 O6 및 dTTP의 O4 또는 N3에서 벌키한 변형을 혼입하는 것에 대한 TdT의 허용을 고려할 때, 그러한 위치에서의 아실 또는 알킬 변형이 스크리닝되고 선택되어 나노세공의 검출 양식 또는 ZMW 단일 분자 시퀀싱 기술을 향상시킬 수 있다. ZMW 단일 분자 접근법에서 DNA 폴리머라제의 활성 부위에서 변형된 뉴클레오타이드의 체류 시간을 향상시키거나 나노세공에서 차등 전류 차단을 향상시키는 변형된 뉴클레오타이드를 통해 검출이 개선될 수 있다. 다른 천연 및 비천연 퓨린 및 피리미딘 뉴클레오타이드 유사체는 이들이 판독 장치 예컨대 나노세공 시퀀서 또는 단일 분자 ZMW 시퀀서로 특유의 디지털 신호를 생성하는 경우 사용될 수 있다. 피리미딘 및 퓨린의 C5 또는 C7 각각에서의 변형은 이들이 판독 장치 예컨대 나노세공 시퀀서 또는 단일 분자 ZMW 시퀀서로 특유의 디지털 신호를 생성하는 경우 각각 사용될 수 있다. 적합한 변형된 뉴클레오타이드 트리포스페이트는 효소 연장 단계 동안에 신속하게 혼입되고 검출 단계 동안에 가능한 한 짧은 동중중합체로 치환-특이적 체류 시간을 제공하도록 선택된다. 비트 인코딩 공간을 확장하기에 적합한 A, G, C, 및 T 염기에 대한 변형의 예는 N6-벤조일 dA, N6-벤질 dA, N6-알킬 dA, N6-아실 dA, N6-치환 알킬 dA, N6-치환 아실 dA, N6-아릴 아실 dA, N6-치환 아릴 아실 dA, N2-알킬-dG, N2-아실 dG, N2-아릴 아실 dG, N2-치환 알킬 dG, N2-치환 아실 dG, N2-치환 아릴 아실 dG, O6 알킬 dG, O4 알킬 dT, N3 알킬 dT, N3 아실 dT, O6 치환 알킬 dG, O4 치환 알킬 dT, C5-프로파르길 아민 dT, C5-프로파르길 아민 dC, C7-프로파르길 아민 dA, C7-프로파르길 아민 dG, 치환 C5-프로파르길 아민 dT, 치환 C5-프로파르길 아민 dC, 치환 C7-프로파르길 아민 dA, 치환 C7-프로파르길 아민 dG를 포함할 수 있으나 이에 제한되지는 않는다. 치환의 바람직한 실시양태는 pH, 온도 및 반응성 종의 농도의 가장 극단적인 화학적 조건하에를 제외하고는 제거에 완전히 안정적인 공유 부착을 포함하나 이에 제한되지는 않는다. 특유의 전류 차단에 영향을 미치는 것이 가능한 치환은 알킬, 헤테로원자 치환 알킬, 방향족 탄화수소, 알킬 치환 방향족 탄화수소, 헤테로원자 치환 알킬 치환 방향족 탄화수소, 헤테로원자 치환 방향족 탄화수소, 벤질, 치환 벤질 또는 그의 조합을 포함할 수 있으나 이에 제한되지는 않는다. 일부 실시양태에서, 치환은 2 내지 450 개의 단량체 유닛으로 구성된 폴리에틸렌 글리콜일 수 있다. 일부 실시양태에서, 펩티드 또는 펩타이드로 구성된 치환은 특이적이고 식별가능한 방식으로 동중중합체의 체류 시간을 증가시키기에 적합할 수 있다. TdT와 같은 템플릿-비의존적 폴리머라제에 의한 변형된 뉴클레오타이드의 혼입 효율은 예컨대 Co⁺⁺, Zn⁺⁺, Mg⁺⁺, Mn⁺⁺, 또는 둘 이상의 상이한 금속 이온의 혼합물이나 이에 제한되지 않는 상이한 금속 이온 보조인자의 사용에 의해 조정될 수 있다. 각각의 변형된 뉴클레오타이드는 효소적 동중중합체 합성 동안에 최적 성능을 위해 상이한 금속 이온을 필요로 할 수 있다.

[0028] DNA 데이터 가닥의 장기간의 안정성이 필수적이기 때문에, 비-퓨린을 기재로 하는 동중중합체는 낮은 pH에서 탈퓨린화되기 때문에 이를 사용하는 것에 대한 뚜렷한 이점이 있다. 일부 실시양태에서, 동중중합체 비트는 단지 단일의 뉴클레오타이드 유형 (즉, 티민)으로만 구성될 수 있으며, 이는 2, 3, 4 또는 그 초과 상이한 화학적 기로 변형되어 각각이 특유의 전류 차단을 결과하는 동중중합체 트랙을 결과한다. 따라서, 4개의 특유의 변형제로 표지된 하나의 뉴클레오타이드는 A, G, C, T의 존재를 대신할 수 있다. 2, 3, 4 또는 그 초과 상이한 화학적 기를 가진 다른 3개의 뉴클레오타이드 중 단지 하나를 사용하는 다른 실시양태가 가능하다.

[0029] 변형된 뉴클레오타이드 유사체가 나노세공을 통한 단일 뉴클레오타이드의 통과를 위해 특유의 체류 시간을 유발하기에 충분한 경우, 또 다른 실시양태는 동중중합체 비트 대신에 단일 뉴클레오타이드 비트를 사용할 것이다. 단일의 변형된 뉴클레오타이드 비트는 DNA 가닥당 최대 정보 밀도를 허용함으로써 DNA 기반 데이터 저장의 비용을 감

소시키는 데 유리할 것이다.

[0030] 관독에 사용된 시퀀싱 기술이 하나의 동중중합체 트랙의 시작 및 정지를 다른 것과 분명히 구별할 수 있는 한, 동중중합체 트랙의 정확한 길이는 중대하지 않다. 비록 사용된 특유의 뉴클레오티드 또는 염기 (변형된 뉴클레오티드 또는 비-뉴클레오티드 유사체 포함)의 수를 증가시키고 따라서 정해진 양의 데이터를 수집하는 데 필요한 길이를 감소시키는 데 뚜렷한 합성 및 저장 밀도 이점이 있긴 하지만, DNA 데이터 저장 합성당 최저 비용은, 그러한 시약들이 분자 생물학 및 시퀀싱 분야에서 널리 사용되며 최저 제조 비용으로 매우 큰 배치로 생산되기 때문에, 효소적 합성 동안에 4개의 천연 뉴클레오티드 dNTP 단량체를 사용함으로써 달성될 수 있다. DNA 데이터 저장의 사용이 증가하고 유사체의 제조 규모가 또한 증가됨에 따라 특유의 동중중합체 트랙의 수를 증가시키기 위한 dNTP 유사체의 생산 비용이 감소될 수 있다.

[0031] 동중중합체 트랙 세그먼트를 합성하는 임의의 방법은 본 발명의 시스템 및 방법과 함께 사용될 수 있으나, 바람직한 실시양태는 템플릿-비의존적 효소 TdT를 사용한다. TdT는 동중중합체의 평균 크기가 신생 올리고뉴클레오티드 메모리 가닥에 대한 [dNTP]의 비에 의해 엄격하게 제어될 수 있는 포아송(Poisson) 분포를 가진 동중중합체를 신속하고 저렴하게 생성하는 한에 있어서는 특정 이점을 제공한다. 일부 실시양태에서, Mn^{2+} 의 존재하에 폴리머라제 세타는 동중중합체 트랙 핵산 메모리 가닥을 합성하기 위해 템플릿-비의존적 폴리머라제로서 사용될 수 있다. 또 다른 실시양태에서, 동중중합체 트랙 세그먼트의 길이는 과량의 dNTP를 반응 구역에 전달한 다음에 신중하게 제어된 시간 간격 후에 반응물을 제거함으로써 제어될 수 있다.

[0032] TdT는 변형하고자 하는 핵산 가닥의 3'-말단의 농도에 대한 dNTP 농도의 비를 제어함으로써 적정하게 규정된 길이의 동중중합체 트랙을 합성하는 능력을 입증하였다. 어레이-기반 포맷에서의 잉크젯 합성은 매우 저비용의 포스포라미다이트 합성이 가능하나, 제조되는 가닥은 길이가 100-200개 염기로 제한되고, 인덱스 서열까지의 일부 길이를 희생하여야 하며, 캡토몰 이하 규모로 제조되어 후속 관독을 위한 충분한 물질을 제공하기 위해 합성 후 증폭을 필요로 하고 비교적 비효율적인 짧은 관독 시퀀싱 관독 기술에 주로 적합하다.

[0033] 본 발명의 방법에 따라 합성된 단일 가닥 DNA의 가닥은 합성 동안에 또는 관독 동안에 헤어핀 또는 dsDNA의 방지에 유익할 수 있다. A:T 또는 G:C 염기 쌍의 한 구성원의 엑소시클릭 아민을 변형시켜 염기 쌍형성에 필요한 수소 결합을 방지함으로써 헤어핀 형성을 방지할 수 있다. 일부 실시양태에서, 엑소시클릭 아민은 아실화 또는 알킬화에 의해 변형될 수 있다. 염기 쌍형성을 방지하는 A, G 또는 C의 엑소시클릭 아민의 임의의 단순하고 안정적인 변형은 헤어핀 형성을 방지하는 데 사용될 수 있다. 특정 실시양태에서, 테옥시아데노신의 N6 및 테옥시구아노신의 N2는 아세틸 기로 아세틸화되어 염기 쌍형성을 방지할 수 있다. 일부 실시양태에서, 테옥시아데노신의 N6 및 테옥시시티딘의 N4는 염기 쌍형성 및 헤어핀 형성을 방지하도록 변형될 수 있다. 일부 실시양태에서, 테옥시구아노신의 O6 또는 테옥시티미딘의 O6은 염기 쌍형성 및 헤어핀 형성을 방지하도록 변형될 수 있다. 일부 실시양태에서, 테옥시티미딘의 O4 또는 테옥시티미딘의 N3은 염기 쌍형성 및 헤어핀 형성을 방지하도록 변형될 수 있다. 일부 실시양태에서, 고차 염기 인코딩 체계를 생성하기 위한 A, G, C 또는 T에 대한 변형은 염기 쌍형성 및 헤어핀 형성을 방지하는 목적을 또한 제공한다. 일부 실시양태에서, 동중중합체 비트는 특유의 전류 차단을 결과하고 가닥-내 또는 가닥-간 이중 가닥 영역의 형성을 방지하는 2, 3, 4 또는 그 초과 상이한 화학적 기로 변형된 단지 단일 뉴클레오티드 유형 (즉, 티민)으로 구성될 수 있다. 또 다른 실시양태에서, TdT 또는 다른 템플릿-비의존적 뉴클레오티드 트랜스퍼라제의 열안정성 버전을 사용하여 승온에서 가닥 합성을 수행할 수 있어서, 가닥-내 또는 가닥-간 이중 가닥 영역의 형성을 방지할 수 있다.

[0034] 동중중합체 트랙 길이의 제어는 길이가 2-10 뉴클레오티드의 동중중합체 트랙 길이의 재현가능한 범위를 생성하기 위해 dNTP 유사체의 혼입 속도의 결정 및 교정 후에 상기 기재된 바와 같이 임의의 유사체에 대해 최적화될 수 있다. A*, G*, C*, 및 T* 및 A**, G**, C** 및 T** 동중중합체 트랙의 사용은 팔 (8) 비트 또는 십이 (12) 비트 인코딩의 생성을 가능하게 하여, "0" 및 "1"을 인코딩하기 위해 단순히 2개의 뉴클레오티드를 사용하기 보다는 하나의 인접한 가닥에 저장될 수 있는 데이터 밀도를 증가시킨다. A, G, C, T, A*, G*, C*, 및 T*가 사용되는 경우 삼 (3) 인접한 동중중합체 트랙은 256 숫자 (즉, 1 바이트)를 인코딩할 수 있다. 이러한 실시양태에서, 삼 (3) 뉴클레오티드 길이의 동중중합체 트랙 또는 십 (10) 뉴클레오티드 길이의 동중중합체 트랙이 각각 사용되는 경우 999 또는 990 뉴클레오티드 길이의 핵산 메모리 가닥에 111 바이트 또는 33 바이트가 있을 것이다. A, G, C, T, A*, G*, C*, T*, A**, G**, C**, 및 T**가 사용되는 경우 이 (2) 인접한 동중중합체 트랙은 256 숫자 (즉, 1 바이트)를 인코딩할 수 있다. 이들 실시양태에서, 삼 (3) 뉴클레오티드 길이의 동중중합체 트랙 또는 십 (10) 뉴클레오티드 길이의 동중중합체 트랙이 각각 사용되는 경우 996 또는 1000 뉴클레오티드 길이의 핵산 메모리 가닥에 166 바이트 또는 50 바이트가 있을 것이다.

[0035] 특정 실시양태에서, 데이터는 보다 높은 수준의 데이터 압축을 달성하기 위해 무작위 시퀀스 및 규정된 조성의 메모리 가닥 이중중합체 트랙에 또한 인코딩될 수 있다. 이중중합체 스트레치(stretch)는 상이한 dNTP의 혼합물을 사용하는 효소적 반응으로 생성될 수 있으며, 여기서 dNTP 화학량론은 이중중합체 트랙의 조성을 제어하는데 사용된다. 이중중합체 트랙의 수 및 유형은 dNTP 유사체의 조합 및 상이한 트랙의 조성을 구별하기 위한 검출 양식의 능력에 의해서만 제한된다. m dNTP 유사체의 경우, 이중중합체 형성을 위한 $(m^2-m)/2$ 이원 조합이 있다. 각각의 이원 조합에 대해 2 가지 상이한 수준의 트랙 조성 (예를 들어, 유사체 A 및 B가 각각 대략 2:1 비로 존재하는 트랙 및 이들이 1:2 비로 존재하는 트랙)을 구별할 수 있는 검출 양식은 데이터가 m개의 유사체 세트로부터 베이스 m^2 의 속도로 인코딩될 수 있게 하여, 메모리 가닥의 코딩 용량을 효과적으로 배가시킨다. 도 5는 동중중합체 및 2 가지 수준의 트랙 조성을 가진 이원 이중중합체-기반 인코딩 체계를 사용하여 이용가능한 dNTP 유사체의 수의 함수로서 메모리 가닥에 저장될 수 있는 데이터를 도시한다.

[0036] 본 발명의 데이터 인코딩 가닥은 이들이 고 처리량 DNA 시퀀싱 기술에 의해 동중중합체 트랙 세그먼트 사이의 전이의 명확한 식별을 허용하기에 충분할 정도로 단지 길 필요가 있기 때문에 (약 2-10 뉴클레오티드) 정확히 규정된 동중중합체 길이를 반드시 필요로 할 필요는 없을 수 있다. 기존의 차세대 합성에 의한 서열분석 (Sequencing-by-Synthesis) (SBS) 시스템은 두 개의 인접한 동중중합체 트랙 사이의 전이를 쉽게 결정할 수 있다. 또한, 동중중합체 트랙의 정확한 길이는 동중중합체 비트의 정확한 검출에 중요하지 않다. 동일한 뉴클레오티드의 트랙의 사용은 현재 SBS 플랫폼에서 가장 일반적인 오류: 삽입 및 결실을 극복할 수 있는 이점을 제공한다. >2 nt의 동중중합체 트랙에서 하나의 뉴클레오티드의 결실은 진정한 동중중합체로 여전히 해석될 것이다. 마찬가지로, 동중중합체 트랙에 단일 뉴클레오티드의 삽입은 SBS 동안에 하나 초과의 뉴클레오티드의 삽입이 있음직하지 않은 사건이기 때문에 2개의 인접한 동중중합체로 잘못 해석되지 않을 것이다. 이 시퀀싱 허용 오차는 DNA 데이터 가닥에 의해 저장된 정보의 올바른 디코딩을 보장하는 데 필요한 시퀀싱 깊이를 감소시키는 이점을 제공한다. 기존의 나노세공 시스템은 그의 차등 전류 차단에 기초하여 A, G, C, 또는 T의 동중중합체 트랙을 서로 쉽게 구별할 수 있다. 특정 실시양태에서, 단일 분자 ZMW 시퀀싱은 선형 가닥 상의 동중중합체 트랙의 선형 순서를 결정하는 데 사용될 수 있다. 둘 중 어느 한쪽이든 시퀀싱 기술의 사용은 DNA 개시체가 단일 분자 ZMW 시퀀싱을 위한 프라이머를 제공하기 위해 합성된 단일 가닥 메모리 가닥의 5'-말단에서 자기-상보적 헤어핀과 같은 시퀀싱 판독 기술과 호환되는 특성을 가질 것을 필요로 할 수 있다. 나노세공 시퀀싱 기술은 "시작" 데이터 마크를 제공하기 위해 가닥의 5'-말단에 자기-상보적 헤어핀을 또한 필요로 할 수 있다. 다양한 실시양태에서, 판독 기술은 일루미나(Illumina) (캘리포니아주 샌디에이고)에 의해 제공되는 것과 같은 임의의 차세대 시퀀싱 방법일 수 있다. 일부 실시양태에서, 판독 또는 시퀀싱 기술은 질량 분석법에 기초할 수 있다. 판독 기술의 기술 특정 오류율은 두 개의 상이한 동중중합체 트랙 사이의 전이를 명확하게 검출할 수 있고/거나 두 개의 동일한 동중중합체 트랙이 서로 인접한 경우에 하나의 동중중합체 트랙 길이와 하나의 2x 길이의 차이를 명확하게 검출할 수 있는 한 중요하지 않다.

[0037] 특정 판독 기술은 본 발명의 구체적 적용에 기초하여 다른 것보다 바람직할 수 있다. 나노세공 시퀀싱과 같은 기술은 비파괴적일 수 있으며, 다중 판독 사이클에 적합한 핵산 메모리 가닥을 무손상으로 남겨둘 수 있다. ZMW 단일 분자 등과 같은 합성에 의한 서열분석 (SBS)에 의존하는 판독 기술은, 원래의 템플릿 가닥의 카피를 생성하며, 상보적 가닥을 제거하고 원래의 핵산 메모리 가닥을 후속의 판독 사이클을 위한 준비가 된 그의 원시적 상태로 되돌리기 위해 판독 후 조작 (즉, 용융에 의한 가닥 분리)을 필요로 할 수 있다. 질량 분석법과 같은 다른 판독 기술은 파괴적이며 샘플링 및 판독의 반복된 사이클 후에 핵산 메모리 가닥의 풀을 고갈시킬 것이다.

[0038] 다양한 실시양태에서, 핵산 메모리 가닥은 "스토퍼"를 포함할 수 있다. "스토퍼"는 나노세공을 통한 단일 가닥 또는 이중 가닥 핵산의 통과를 방지하는 거대분자 구축물일 수 있다 (Manrao, et al., 2012, Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase, Nature Biotechnology 30, 349-353 (본원에 참조로 포함됨)). phi29 DNA 폴리머라제와 같은 단백질은 단백질 나노세공의 시스 측 상에서 더 큰 세공 (약 6.3 nm)을 통해 끌어내어 지지 않을 정도로 충분히 크다. 더 작은 측의 세공 직경은 약 1.2 nm 폭인 것으로 추산된다. 스톱퍼는 본 발명의 핵산 메모리 가닥의 5'- 또는 3'- 말단 상에 어느 한쪽에 사용될 수 있다. 일부 적용에서, 핵산 분자의 양쪽 3'- 또는 5'- 말단 둘 다에 스톱퍼를 갖는 것이 바람직할 수 있다. 스톱퍼는 정보 인코딩 핵산이 공유 부착되는 돌출된 5'- 또는 3'- 돌출부를 가진 헤어핀 (스텝-루프) 구조로 이루어질 수 있다. 스톱퍼가 헤어핀으로 이루어지는 경우, ds 스텝의 길이는 나노세공을 통해 메모리 가닥을 전위시키는 데 사용되는 전기장에 의해 그에 가해지는 임의의 용융력에 저항하기에 충분히 길 수 있다. 특정 실시양태에서, 이중 가닥 줄기 영역의 하나의 염기는 그 염기쌍을 형성하는 그의 동족 염

기에 가교결합될 수 있어서, 헤어핀의 이중 가닥 스템 부분이 나노세공을 통해 나머지 분자를 전위시키는 전기장에 의해 그에 가해진 힘의 영향하에 용융되는 것이 불가능하다. 특정 실시양태에 따른 TdT 매개 핵산 메모리 합성에 헤어핀 스토퍼를 이용하기 위해, 스토퍼는 템플릿 비의존적 합성을 위한 TdT의 결합을 허용하기에 충분한 길이 (즉, >10 뉴클레오티드)의 3'-돌출부를 가질 수 있다.

[0039] 스토퍼는 핵산 분자의 3'- 또는 5' 말단에 부가될 수 있는 비-뉴클레오티드 거대분자 구축물로 이루어질 수 있다. 구축물은 폴리머라제 또는 트랜스퍼라제 예컨대 TdT에 의해 거대분자 종을 핵산의 3'-말단 상에 직접 접합에 의해 (Sorensen, et al., 2013, Enzymatic Ligation of Large Biomolecules to DNA, ACS Nano, 7(9):8098-8104 (본원에 참조로 포함됨)) 또는 그 기능성을 통해 핵산의 특이적 변형을 허용하는 기능화된 뉴클레오티드의 혼입에 의해 (Winz, et al., 2015, Nucleotidyl transferase assisted DNA labeling with different click chemistries, Nucleic Acids Res. 43(17):e110 (본원에 참조로 포함됨)) 합성될 수 있다. 5'-말단 스토퍼는 헤어핀의 직접 합성을 통해 또는 3' 내지 5' 올리고뉴클레오티드 합성의 마지막 단계로서 도입된 기능성 핸들의 2차 변형을 통해 올리고뉴클레오티드 어댑터의 화학적 합성시에 쉽게 도입될 수 있고 개시제로서 사용될 수 있다. 대안적으로, 5'-말단 스토퍼는 5'-말단을 통해 올리고뉴클레오티드 개시제를 자성 또는 비자성 비드 또는 입자 또는 나노입자에 부착하고, 메모리 가닥을 함유하는 동종중합체 트랙을 효소적으로 합성한 다음에 메모리 가닥을 자성 또는 비자성 비드 또는 입자 또는 나노입자에 부착된 채로 덩으로써 구축될 수 있다.

[0040] 스토퍼는 나머지 분자로부터 스토퍼의 절단을 허용하여 핵산 가닥이 나노세공 밖으로 수동으로 확산되거나 전압의 인가를 통해 나노세공 밖으로 전위되어 가닥이 회복될 수 있게 함으로써 추가로 변형될 수 있다.

[0041] 특정 실시양태에서, 템플릿 비의존적 폴리머라제 또는 트랜스퍼라제는 핵산의 사전-합성된 가닥을 변형시켜 "1 회 기입, 다수 판독(Write Once, Read Many)" 유형의 메모리 장치로서 나노세공 장치의 사용을 가능하게 하기 위해 사용될 수 있다. DNA 시퀀서로서 나노세공 장치의 사용과 연관된 고유의 문제 중 일부는 나노세공의 열악한 식별 때문에 이들이 초래하는 높은 오류율이다. 이는 세공을 통한 전위의 속도에 기인할 수 있거나 나노세공의 대략적인 깊이가 8 nm이어서, 다수의 염기가 동시에 세공에 존재할 수 있게 한다는 사실에 기인할 수 있다. 본 발명의 동종중합체 메모리 가닥은 동종중합체 반복부의 사용을 통해 이 문제를 해결하여, 엄격한 시퀀싱 정확도의 필요성을 감소시킨다. 특정 실시양태에서, 나노세공 시퀀싱의 단점은 이중 가닥 DNA 메모리 가닥의 한 말단에 헤어핀 어댑터를 시행하고 따라서 전위 및 염기 호출 과정 동안에, DNA 메모리 가닥의 각각의 센스가 판독될 수 있도록 하여 개별 염기와 그 상보적 가닥이 각각의 염기를 단지 한 번만 판독하는 오류율을 보상할 수 있도록 함으로써 해결될 수 있다. 특정 실시양태에서, 나노세공 시퀀싱 충실도는 세공 (예를 들어, 단백질 또는 고체 상태)을 가로질러 전위하지 않을 별키한 부속물을 가진 단일 가닥 또는 이중 가닥 핵산 분자의 각각의 말단 (5'- 및 3'-)의 적절한 변형에 의해 증가될 수 있다. 그 다음에, 분자는 세공 내에 포획되고 동일한 세공에서 동일한 분자의 다수의 판독을 허용하기 위해 앞뒤로 다수회 전위될 수 있으며, 따라서 판독 횟수의 제공에 의해 시퀀싱 오류율을 감소시킨다 (시퀀싱 판독 오류가 확률적 기원에 기인한 경우).

[0042] TdT와 같은 트랜스퍼라제는 DNA 분자의 3'-말단에 크고 벌키한 변형된 뉴클레오티드 유사체를 부가하는 데 사용될 수 있다. 특정 실시양태에서, WORM 나노세공 메모리 장치는 하기 단계를 사용하여 생성될 수 있다: (1) 상기 논의된 바와 같이 임의의 고밀도 인코딩 체계로 구체적 정보를 인코딩하는 단일 분자의 DNA를 생성하고, 나노세공을 통한 DNA 분자의 완전한 전위를 방지하는 벌키한 분자 구축물로 5'-말단을 공유 변형시키는 단계; (2) 5'-변형된 말단이 나노세공과 접촉하고 더 이상 전위될 수 없을 때까지 나노세공을 통해 DNA 분자를 스트레하는 단계; (3) 나노세공의 원환체 내에 분자를 효과적으로 포획하기 위해 하나 (또는 그 초과)의 벌키한 뉴클레오티드 유사체 ("스토퍼")를 DNA 분자의 3-말단에 공유적으로 첨가하기 위해 TdT 및 변형된 뉴클레오티드를 사용하는 단계; (4) 전류의 극성을 나노세공으로 역전시켜 3'- 변형되지 않은 임의의 DNA 분자를 일소하여 "포획된" (5'- 및 3'-변형) 핵산 가닥의 순수 집단을 생성하는 단계; (5) 세척 또는 다른 수단을 통해 나노세공 부근으로부터 임의의 포획되지 않은 핵산을 제거하는 단계; (6) 인가된 전압을 사용하여 "포획된" DNA 가닥을 한 방향 또는 두 방향으로 판독하는 단계 (잠재적으로 다수회 판독하여 오차율을 허용되는 수준으로 감소시키는 단계). 다양한 실시양태에서, 단계 6은 한 방향으로 전압 유도된 "판독", 및 나노세공을 통해 데이터 인코딩 핵산을 "되감기" 위해 반대 방향으로의 신속한 전위에 뒤이어 원래 방향으로 또 다른 전압 유도된 "판독"으로 이루어질 수 있다. 이 "판독" - "되감기" - "판독"의 사이클은 원하는 바와 같이 다수회 반복할 수 있다.

[0043] 일부 실시양태에서, 포획된 핵산 가닥은 전위 동안에 어느 방향으로든 판독될 수 있다. 일부 실시양태에서, 포획된 가닥은 분자의 한 말단 (5'- 또는 3'-)으로 전위될 수 있고, 그러한 판독의 극성이 보다 높은 정확도 판독을 제공할 수 있으므로 반대 방향으로 판독될 수 있다.

- [0044] 특정 실시양태에서, 원형화된 핵산 메모리 가닥은 상기 기재된 합성 방법을 사용하여 생성된 후 원형화될 수 있다. 원형화된 가닥은 별키한 거대분자 또는 합성된 가닥의 말단이 데이터 관독을 위한 시작 및 정지 지점을 지정하기 위해 연결된 특이적 동중중합체 시퀀스를 포함할 수 있다. 시작 및 정지 동중중합체 시퀀스는 또한 선형 핵산 가닥에서 사용될 수 있다. 원형화된 가닥 (305)이 도 6에 도시된 바와 같이 단일 막 (303) 상에 위치한 2개의 나노세공 (306 및 309) 사이에 물리적으로 포획되도록 원형화된 가닥 (305)은 2개의 인접한 나노세공 (306 및 309) 사이에 스레싱되었을 수 있다. 원형화된 메모리 가닥은 단일 뉴클레오티드의 시퀀스, 동중중합체 트랙 시퀀스, 변형된 뉴클레오티드 유사체의 시퀀스, 또는 그의 일부 조합물로서 디지털 정보를 인코딩할 수 있다. 하나의 나노세공 (309)은 정보-인코딩된 메모리 가닥이 세공을 통해 전위됨에 따라 전기 신호를 생성하는데 사용될 수 있으며, 한편 다른 하나의 나노세공 (307)은 DNA 분자가 막 (303) 및 제1 나노세공 (309)의 시스측으로 되돌아갈 수 있게 하는 포털로서 단순히 작용할 수 있다. 이 체계의 한 이점은 정보 인코딩 가닥이 반복적인 관독을 위해 재순환될 수 있고 다수회 관독될 수 있어서 임의의 가능한 관독 오류를 감소시킬 수 있다는 점이다.
- [0045] 다른 실시양태는 단지 구체적 조건 세트하에 액세스될 수 있도록 메모리 가닥에서 데이터를 인코딩할 수 있다. 이러한 경우에, 메모리 가닥은 절단가능한 링커로 부착된 변형을 함유하는 뉴클레오티드로 적어도 부분적으로 구성된다. 중합체 트랙이 올바른 처리 없이 나노세공을 통해 전위하는 경우, 전류 차단이 데이터를 인코딩하는 시퀀스와 상이하도록 변형 (예를 들어, 화학적 보호기) 및 링커를 선택할 수 있다. 도 7은 dG 뉴클레오티드에 대한 디설피드 및 아미드-연결된 변형을 사용하는 도식을 개괄하고, 메모리 가닥에 인코딩된 전류 차단 및 데이터가 처리 조건에 대응하여 어떻게 변할 수 있는지를 도시한다. G* 및 G**는 구조적으로 크기와 유연성이 유사하며 나노세공 플랫폼 상에서 유사한 전류 차단을 초래할 수 있지만, 다른 조건하에 제거된다. G* 및 G***는 구조적으로 상이하지만 변형은 동일한 제거 조건을 공유한다. 다른 실시양태는 상이한 처리 예컨대 구체적 파장의 빛, 산성 또는 알칼리성 pH, 산화 또는 환원 조건, 또는 서열-특이적 뉴클레아제로 절단가능한 다른 변형 또는 링커를 이용할 수 있다. 일부 실시양태는 암호화를 위해 또는 데이터에 대한 이전 액세스 또는 변경의 화학적 마커로서 메모리 가닥 변형의 존재 또는 부재를 사용할 수 있다. 대부분의 링커 절단 반응은 효과적으로 비가역적이므로, 이 접근법은 단일 분자가 중복 없이 데이터를 인코딩하기에 충분할 수 있는 1회 기입 다수 관독 시스템 가장 적합하다.
- [0046] 본 발명의 관독 체계에 유용한 많은 가능한 정보 인코딩 체계가 가능하며 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 분명할 수 있다.
- [0047] 플레이트의 웰 (예를 들어, 각각 1.5 μL 의 1536개 웰 플레이트) 내로의 드롭의 어쿠스틱 전달을 사용하여 합성이 완수될 수 있다. 다양한 실시양태에서, 핵산 메모리 가닥은 비드 또는 자성 비드 또는 표면 상에서 합성될 수 있고, 전장 합성이 완료된 후 비드 또는 자성 비드 또는 표면 상에 남겨지거나 적용에 따라 합성 지지체로부터 제거될 수 있다.
- [0048] 특정 실시양태에서, 긴 (5-10 kb) 데이터 가닥의 합성을 위한 시스템은 웰 (예를 들어, 다수의 나노리터 부피 웰)의 어레이에 잉크젯 전달을 사용할 수 있다. 다른 실시양태에서, 다수의 공압 제어 액츄에이터가 시약을 어레이의 각각의 위치로 동시에 전달하기 위해 각각의 웰 위에 위치될 수 있다. 각각의 액츄에이터는 DNA 데이터 가닥의 비드를 특징짓는 데 사용되는 템플릿-비의존적 폴리머라제로 제제화된 2개 이상의 뉴클레오티드 또는 변형된 뉴클레오티드 각각 중에서 선택하는 선택기 밸브에 의해 제공될 것이다. 하나 이상의 추가 선택기 밸브 포트는 필요한 경우 하나 이상의 세척 시약 전용일 것이다. 웰의 직경이 전달된 액체가 개방형 웰의 길이 내에 모세관 작용에 의해 포획되도록 하는 한, 나노리터 부피 웰의 어레이는 양단에서 개방될 수 있다. 각각의 라운드의 뉴클레오티드-효소 제제가 개방형 웰에 전달된 후, 세정 시약 또는 효소적 반응 정지 시약을 웰의 어레이의 하부 개구부를 가로질러 그리고 그를 통해 흐르게 할 수 있어 각각의 웰이 반응 혼합물로 세정되어 효소적 합성의 그 다음 사이클을 위한 어레이를 준비하도록 한다. 다른 실시양태에서, 그 다음 시약의 전달 전에 모세관 나노웰로부터 하나의 시약을 신속하게 제거하기 위해 진공 공급원이 사용된다.
- [0049] 특정 실시양태는 밸브-제어 시약 전달과 함께 고 병렬적 나노유체 챔버를 사용할 수 있다. 예시적인 미세유체 핵산 메모리 가닥 합성 장치가 예시적인 목적을 위한 것이며 일정한 비율로 확대축소된 것이 아닌 도 8에 도시되어 있다. 조절기 (257)를 포함한 미세유체 채널 (255)은, 저장소 (253)를 반응 챔버 (251)에 커플링하고 출구 채널 (259)은 반응 챔버 (251)로부터 폐기물을 배출하기 위한 조절기 (257)를 포함한다. 핵산 메모리 가닥 합성을 위한 미세유체 장치는, 예를 들어, 채널 (255), 저장소 (253), 및/또는 조절기 (257)를 포함할 수 있다. 핵산 메모리 가닥 합성은 미세유체 반응 챔버 (251)에서 일어날 수 있으며, 이는 반응 챔버의 내부 표면에 고정되거나 결합되고 폴리뉴클레오티드 개시제를 임의로 방출가능하게 결합시킬 수 있는 비드 또는 다른 기질을 포

함할 수 있는 다수의 고정된 합성된 뉴클레오티드 개시제를 포함할 수 있다. 반응 챔버 (251)는 시약이 반응 챔버 (254)에 첨가 및 제거될 수 있도록 적어도 하나의 흡입구 및 하나의 출구 채널 (259)을 포함할 수 있다. 반응 챔버 (251)는 최적의 그리고 재현가능한 효소적 합성 조건을 유지하도록 온도 제어되어야 한다. 미세유체 장치는 메모리 체인 코딩 체계에 사용될 각각의 dNTP 또는 유사체 각각에 대한 저장소 (253)를 포함할 수 있다. 이들 저장소 (253) 각각은 템플릿-비의존적 방식으로 DNA 또는 RNA 가닥을 연장시키는 적절한 양의 TdT 또는 임의의 다른 효소를 또한 포함할 수 있다. 추가 저장소 (253)는 세척 또는 다른 작업을 위한 시약을 함유할 수 있다.

[0050] 저장소 (253)는 별도의 채널 (255)을 통해 반응 챔버 (254)에 커플링될 수 있고, 각각의 채널 (255)을 통해 반응 챔버 (254)로의 시약 흐름은 게이트, 밸브, 압력 조절기, 또는 다른 수단의 사용을 통해 개별적으로 조절될 수 있다. 출구 채널 (259)을 통한 반응 챔버 (254)로부터의 흐름은 유사하게 조절될 수 있다. 저장소 (253)는 반응 챔버 (254) 내로 유입되도록 허용되는 시약의 부피에 기초하여 시약의 농도가 엄격하게 제어될 수 있도록 알려진 농도의 유체에 현탁된 상기에 기재된 dNTP, 변형된 dNTP 또는 그의 임의의 유사체를 보유할 수 있다. 따라서, 각각의 동종중합체 트랙의 길이는 시약 농도의 제어를 통해 관리될 수 있다.

[0051] 특정 경우에, 시약, 특히 dNTP 및 효소 시약이 재순환될 수 있다. 시약은 게이트, 밸브, 진공 펌프, 압력 조절기 또는 다른 조절기 (257)를 사용하여 역류를 유도함으로써 이들이 진입된 동일한 채널 (255)을 통해 반응 챔버 (254)로부터 그의 각각의 저장소 (253)로 다시 유입될 수 있다. 대안적으로, 시약은 독립적인 복귀 채널을 통해 반응 챔버 (254)로부터 그의 각각의 저장소 (253)로 복귀될 수 있다. 미세유체 장치는 게이트, 밸브, 압력, 또는 상기에 기재된 다른 조절기 (257)를 작동시킬 수 있는 제어기를 포함할 수 있다.

[0052] 예시적인 미세유체 핵산 메모리 가닥 합성 반응은 원하는 dNTP (본 발명의 핵산 메모리 체인의 데이터를 인코딩하는 데 사용된 임의의 구성요소 분자를 지칭하기 위해 전체에 걸쳐 사용됨) 시약을 출구 채널 (259) 또는 복귀 채널 (도시되지 않음)을 통해 반응 챔버 (254)로부터 NTP 시약을 제거하기 전에 미리 결정된 농도에서 및 미리 결정된 시간의 양 동안 (원하는 동종중합체 길이를 결과하도록 계산됨) 반응 챔버 (254) 내로 유동시키는 단계; 세척 시약을 반응 챔버 (254) 내로 유동시키는 단계; 출구 채널 (259)을 통해 반응 챔버 (254)로부터 세척 시약을 제거하는 단계; 원하는 동종중합체 트랙 비를 달성하기 위해 계산된 조건하에 원하는 메모리 가닥 시퀀스에서 그 다음 NTP 시약을 유동시키는 단계; 및 원하는 핵산 메모리 가닥이 합성될 때까지 반복하는 단계를 포함할 수 있다. 원하는 핵산 메모리 가닥이 합성된 후, 이는 반응 챔버 앵커 또는 기관으로부터 방출되어 출구 채널 (259) 또는 다른 수단을 통해 수집될 수 있다.

[0053] 사용가능한 양의 데이터를 인코딩하는 데 필요한 상당한 수의 동종중합체 인코딩된 DNA 가닥 때문에, 고 병렬적 DNA 합성 방법이 필요하다. 한 실시양태에서, 도 9에 도시된 바와 같이, 소수성 물질의 수평 및 수직 스트라이프 (9-2)를 패터닝하여 소수성 영역과 경계를 이루는 복수개의 친수성 웰 (?)을 형성함으로써, 웰 (9-1)의 어레이를 함유하는 플로우 셀이 적합한 기관 상에 형성된다. 친수성 웰의 전형적인 치수는 300 x 300 nm 내지 1000 x 1000 nm일 수 있다. 이 친수성 어레이는 플로어와 광학적으로 투명한 커버 사이에 적합한 치수의 갭을 가진 플로우 셀의 플로어를 형성한다. 천연 또는 변형된 뉴클레오티드 트리포스페이트 중 하나 및 임의의 필요한 보조인자인 저온 (즉, 최적의 효소-특이적 온도 미만의) 뉴클레오티딜 트랜스퍼라제 용액은, 정지시 유입구 (9-3)를 통해 플로우 셀 내로 유입되어, 유체 흐름의 중단 직후, 효소-뉴클레오티드 트리포스페이트 용액은 각각의 친수성 영역 위에 위치된 공간적으로 규정된 액적 (9-4) 내로 비딩되도록 한다. 성형 렌즈 (9-6)를 통한 IR 공급원 (9-5)은 DLP (디지털 광 프로젝션(digital light projection)) 장치 (9-8) 상으로 빔 (9-7)을 투사하며, 이는 구체적 뉴클레오티드가 첨가되도록 선택된 각각의 소수성계 제한된 액적 (9-4)에 IR 빔 (9-9)을 동시에 조종하기 위해 사용되어, 원하는 길이의 동종중합체를 합성하도록 규정된 기간 동안 폴리머라제 연장 반응 체제를 최대 효소 활성을 위한 온도로의 신속 가열을 결과한다. 일부 적합하게 규정된 반응 시간 후, IR 공급원은 차단되고 저온 세정 완충제는 플로우 셀에 신속하게 주입하고 출구 (9-10)를 통해 배출시켜, 반응물을 퀘칭하고 1회 "기입" 사이클을 마무리한다. 이 일련의 단계는 각각의 "기입" 사이클에 대해 다수회 반복되므로, 전장 동종중합체 데이터 가닥이 완료될 때까지, 각각의 신생 데이터 가닥이 그의 공간 위치 및 첨가될 선택된 뉴클레오티드에 따라 무작위로 액세스된다. 일부 실시양태에서, 1920 x 1080 조종가능한 미터를 가진 DLP 장치는 합성 플로우 셀에서 ~2M 합성 위치에 걸쳐 동시에 무작위로 액세스하기 위해 사용될 수 있다. 일부 실시양태에서, 사용된 템플릿-비의존적 폴리머라제는 실온보다 훨씬 높은 반응 온도로 호열성이므로, IR 공급원에 의한 액적의 신속한 가열 전에 소수성계 제한된 액적에서 효소 활성이 최소화된다. 일부 실시양태에서, 소수성계 규정된 친수성 웰을 갖는 플로우 셀의 바닥 표면은, IR 공급원에 의해 온도가 상승될 때까지 효소적 활성을 방지하기 위해 소수성 웰의 액적을 감소된 온도로 유지하는 냉각 장치에 인접해 있다.

- [0054] 또 다른 실시양태에서, 플로우 셀은 소수성 영역과 경계를 이루는 복수개의 친수성 스폿을 형성하는 소수성 물질의 수평 및 수직 스트라이프를 패터닝함으로써 적합한 기판 상에 형성된 웰의 어레이로 구성된다. 친수성 스폿의 전형적인 치수는 300 x 300 nm 내지 1000 x 1000 nm일 수 있다. 각각의 친수성 스폿은 개별적으로 주소 지정이 가능한 CMOS 히터 상에 위치된다. 이 친수성-CMOS 히터 어레이는 플로어와 광학적으로 투명한 커버 사이에 적합한 치수의 겹을 가진 플로우 셀의 플로어를 형성한다. 저온 (즉, 최적의 효소-특이적 온도 미만의) 뉴클레오티드 트랜스퍼라제 및 친연 또는 변형된 뉴클레오티드 트리포스페이트 중 하나의 용액은 플로우 셀 내로 유입되어, 유체 흐름의 중단 직후, 폴리머라제 연장 반응 용액이 연관된 CMOS 히터로 각각의 친수성 영역 위에 위치한 공간적으로 규정된 액적 내로 비딩되도록 한다. 구체적 뉴클레오티드를 첨가하도록 선택된 소수성계 제한된 액적 각각은 원하는 길이의 동중중합체를 합성하도록 규정된 기간 동안 최대 효소 활성을 위한 온도로 신속히 가열된다. 일부 적합하게 규정된 반응 시간 후, 히터는 차단되고, 저온 세정 완충제는 플로우 셀에 신속하게 주입하여, 반응물을 쉐킹하고 1회 "기입" 사이클을 마무리한다. 이 일련의 단계는 각각의 "기입" 사이클에 대해 다수회 반복되므로, 전장 동중중합체 데이터 가닥이 완료될 때까지, 각각의 신생 데이터 가닥이 그의 공간 위치 및 첨가될 선택된 뉴클레오티드에 따라 무작위로 액세스된다. 일부 실시양태에서, 사용된 효소는 실온보다 훨씬 높은 온도에서 최적의 온도를 갖는 호열성이므로, CMOS 히터에 의한 액적의 신속한 가열 전에 소수성계 제한된 드롭 내에 원치 않는 뉴클레오티드 첨가의 가능성이 낮도록 한다.
- [0055] 관련 기술분야의 통상의 기술자는 본 발명의 시스템 및 방법에 대해 필요하거나 가장 적합하다고 인식할 것인 바와 같이, 본 발명의 시스템 및 방법은 도 10에 도시된 바와 같은 컴퓨팅 장치를 포함할 수 있으며, 이는 하나 이상의 프로세서 (309) (예를 들어, 중앙 처리 장치 (CPU), 그래픽 처리 장치 (GPU) 등), 컴퓨터-판독가능 저장 장치 (307) (예를 들어, 메인 메모리, 정적 메모리 등), 또는 버스를 통해 서로 통신하는 그의 조합을 포함할 수 있다. 컴퓨팅 장치는 모바일 장치 (101) (예를 들어, 휴대폰), 개인용 컴퓨터 (901), 및 서버 컴퓨터 (511)를 포함할 수 있다. 다양한 실시양태에서, 컴퓨팅 장치는 네트워크 (517)를 통해 서로 통신하도록 구성될 수 있다.
- [0056] 컴퓨팅 장치는 메모리 가닥의 합성, 시퀀싱된 메모리 가닥의 판독, 및 인간 또는 기계-판독가능 포맷, 디지털화된 데이터, 및 핵산 시퀀스 사이의 데이터의 컴파일링 또는 번역을 본원에 기재된 다른 단계 중에서 제어하는 데 사용될 수 있다. 컴퓨팅 장치는 판독가능한 포맷의 데이터를 디스플레이하는 데 사용될 수 있다.
- [0057] 프로세서 (309)는 관련 기술분야에 공지된 임의의 적합한 프로세서, 예컨대 인텔(Intel)에 의해 상표 크세온 (XEON) E7 (캘리포니아주 산타클라라)로 판매된 프로세서 또는 AMD (캘리포니아주 서니테일)에 의해 상표 옵테론 (OPTERON) 6200으로 판매된 프로세서를 포함할 수 있다.
- [0058] 메모리 (307)는 바람직하게는 저장가능한 적어도 하나의 유형의(tangible), 비-일시적 매체: 시스템이 본원에 기재된 기능 (예를 들어, 본원에서 발견된 임의의 방법론 또는 기능을 구현하는 소프트웨어)을 수행하도록 실행 가능한 하나 이상의 명령 세트; 데이터 (예를 들어, 메모리 가닥에서 인코딩될 데이터); 또는 둘 다를 포함한다. 예시적인 실시양태에서 컴퓨터-판독가능 저장 장치는 단일 매체일 수 있지만, 용어 "컴퓨터-판독가능 저장 장치"는 명령 또는 데이터를 저장하는 단일 매체 또는 다중 매체 (예를 들어, 중앙 집중식 또는 분산 데이터베이스, 및/또는 연관 캐시 또는 서버)를 포함하는 것으로 간주되어야 한다. 따라서, 용어 "컴퓨터-판독가능 저장 장치"는 고체-상태 메모리(solid-state memory) (예를 들어, 가입자 식별 모듈 (SIM) 카드, 보안 디지털 카드 (SD 카드), 마이크로 SD 카드, 또는 고체-상태 드라이브 (SSD)), 광학 및 자성 매체, 하드 드라이브, 디스크 드라이브, 및 임의의 다른 유형의 저장 매체를, 제한 없이, 포함하는 것으로 간주되어야 한다.
- [0059] 임의의 적합한 서비스, 예컨대, 예를 들어, 아마존 웹 서비스(Amazon Web Services), 서버 (511)의 메모리 (307), 클라우드 저장, 또 다른 서버, 또는 다른 컴퓨터-판독가능 저장 장치 (527)에 사용될 수 있다. 클라우드 저장은 데이터가 논리적 풀에 저장되고 물리적 저장 장치 여러 서버 및 여러 위치에 걸쳐 있을 수 있는 데이터 저장 체계를 지칭할 수 있다. 저장 (527)은 호스팅 회사에 의해 소유되고 관리될 수 있다. 바람직하게는, 저장 (527)은 본원에 기재된 조작을 수행하고 지지하는 데 필요한 바와 같은 기록 (399)을 저장하는 데 사용된다.
- [0060] 본 발명에 따른 입력/출력 장치 (305)는 하나 이상의 비디오 디스플레이 유닛 (예를 들어, 액정 디스플레이 (LCD) 또는 음극선 관 (CRT) 모니터), 영숫자 입력 장치 (예를 들어, 키보드), 커서 제어 장치 (예를 들어, 마우스 또는 트랙패드) 디스크 드라이브 유닛, 신호 생성 장치 (예를 들어, 스피커), 터치스크린, 버튼, 가속도계, 마이크, 셀룰러 무선 주파수 안테나, 네트워크 인터페이스 장치 (이는 예를 들어 네트워크 인터페이스 카드 (NIC), 와이-파이(Wi-Fi) 카드, 또는 셀룰러 모듈일 수 있다), 또는 그의 임의의 조합을 포함할 수 있

다.

[0061] 관련 기술분야의 통상의 기술자는 임의의 적합한 개발 환경 또는 프로그래밍 언어가 본 발명의 다양한 시스템 및 방법에 대해 본원에 기재된 동작성을 허용하기 위해 이용될 수 있음을 인식할 것이다. 예를 들어, 본원에서의 시스템 및 방법은 펄(Perl), 피톤(Python), C++, C#, 자바(Java), 자바스크립트(JavaScript), 비주얼 베이식(Visual Basic), 루비 온 레일즈(Ruby on Rails), 그루비(Groovy) 및 그레일즈(Grails) 또는 임의의 다른 적합한 툴을 사용하여 시행될 수 있다. 컴퓨팅 장치 (101)의 경우, 고유 x코드(Code) 또는 안드로이드 자바(Android Java)를 사용하는 것이 바람직할 수 있다.

[0062] 도 11은 효소적 합성을 통해 제조된 2개의 상이한 단일 동중중합체 트랙의 폴리아크릴아미드 겔 전기영동 분석을 나타낸다. 레인 A는 모든 후속 레인에서 사용되는 출발 이십량체(20-mer) 올리고뉴클레오티드의 샘플이다. 레인 B는 이십량체 올리고뉴클레오티드 및 이십일량체(21-mer)의 형성을 나타내는 비가역적인 종결자 ddATP를 함유하는 TdT 반응으로부터의 샘플이다. 레인 C는 37°C에서 1분 후 이십량체 및 천연 뉴클레오티드 dATP를 함유하는 TdT 반응으로부터의 샘플이다. 레인 D는 37°C에서 5분 후 레인 D에서 동일한 반응 혼합물의 샘플이다. 레인 E는 37°C에서 15분 후 레인 C에서 동일한 반응 혼합물의 샘플이다. 그러한 3개의 레인은 TdT 연장 반응 동안에 ~5분 내에 입력 dATP의 소비 및 5 내지 15분 사이의 어떠한 동중중합체 성장도 관찰되지 않기 때문에 동중중합체 길이 제어를 예시한다. 레인 F는 37°C에서 1분 후 이십량체 올리고뉴클레오티드 및 뉴클레오티드 유사체 N6-벤조일-dATP를 함유하는 TdT 반응으로부터의 샘플이다. 레인 G는 37°C에서 5분 후 레인 F에서의 반응 혼합물의 샘플이다. 레인 H는 37°C에서 15분 후 레인 F에서 동일한 반응 혼합물의 샘플이다. 비록 레인 F-H에서 형성된 dA^{Bz} 동중중합체의 길이 사이에 질적 차이가 있긴 하지만, 심지어 N6-변형 dATP 유사체에서도 동일한 길이 제어가 입증된다.

[0063] 실시예

[0064] 건조 N₂ 하에 바이알을 N⁶-벤조일 -2'-데옥시아데노신 (0.055 g, 0.16 mmol)으로 충전시킴으로써 N⁶-벤조일-데옥시아데노신 트리포스페이트를 제조하고, 트리메틸 포스페이트 (0.435 mL)를 첨가하였다. 생성된 용액에 트리부틸아민 (0.077 mL, 0.32 mmol)을 첨가하고, 반응 혼합물을 -5°C에서 유지하면서 30분 동안 건조 N₂로 플라싱하였다. 이 바이알에 시린지를 통해 무수 옥시염화인 (0.018 mL, 0.19 mmol)을 첨가하고, 반응 혼합물을 3분 동안 -5°C에서 교반하였다. 무수 옥시염화인 (0.009 mL, 0.10 mmol)의 제2 분취액을 시린지를 통해 첨가하고 반응 혼합물을 8분 동안 -5°C에서 교반하였다. 제2 바이알을 트리부틸아민 피로포스페이트 (0.075 g, 0.14 mmol)로 충전하고, 건조 N₂로 플라싱하고, 무수 아세트니트릴 (0.609 mL)에 이어서 트리부틸아민 (0.231 mL, 0.97 mmol)을 첨가하였다. 제조된 트리부틸아민 피로포스페이트 혼합물을 -20°C로 냉각시키고 반응 혼합물에 첨가하고 10분 동안 반응시켰다. H₂O (4.35 mL)를 적가하여 반응물을 쉐킷하였다. 플라스크의 내용물을 0.87 mL의 H₂O와 합하고 디클로로메탄 (3 x 150 mL)으로 추출하였다. 수성 상을 농축 NH₄OH로 pH 6.5로 조정하고 4°C에서 12시간 동안 교반하였다. 혼합물을 50 mL의 물과 함께 250 mL 환저 플라스크로 옮기고, 감압하에 농축하였다. 잔류물을 40 mL 물에 용해시키고, 이온-교환 크로마토그래피 (AKTA FPLC, 프락토겔(Fractogel) DEAE 48 mL 칼럼 부피, 물 중 단계적 구배 0 → 70% TEAB, pH 7.5)를 통해 정제하였다. 원하는 생성물을 함유하는 분획을 풀링하고, 농도는 A260에 의한 것이고, 건조까지 물 (5 x 50 mL)로부터 반복 농축을 통해 잔류 트리에틸암모늄 비카르보네이트를 제거하여 감압하에 농축시켜 N⁶-벤조일-2'-데옥시아데노신 트리포스페이트를 제공하였다.

[0065] 뉴클레오티딜 트랜스퍼라제 (TdT)에 의한 변형된 뉴클레오티드로 구성된 동중중합체 트랙의 제어된 합성을 하기 방식으로 수행하였다. 각각 1 mM의 데옥시아데노신 트리포스페이트 (트리링크 바이오사이언시즈(TriLink Biosciences)) 및 N⁶-벤조일-데옥시아데노신 트리포스페이트의 스톱 용액을 H₂O에서 제조하였다.

[0066] 상이한 트리포스페이트 각각 0.5 μL (500 pmoles)를 1.5 μL (30U)의 시판되는 TdT (써모 사이언티픽(Thermo Scientific)), 0.5 μL (50 pmoles)의 5'-TAATAATAATAATTTTT-3' (IDT), 2 μL의 시판되는 TdT Rxn 완충제 (써모 사이언티픽-1M 칼륨 카르코딜레이트, 0.125M 트리스, 0.05% (v/v) 트리톤 X100, 5 mM C₆C₁₂ (25°C에서 pH 7.2)) 및 4.5 μL의 H₂O와 개별적으로 합하였다. 반응물을 37°C에서 인큐베이션하였다. 30 μL 분취액을 1 m, 5 m 및 15 m 후에 제거하고 각각 20 μL 5 mM EDTA로 쉐킷하였다. 각각의 샘플을 진공하에 건조시키고 100 μL H₂O에서 재구성하였다. 10 μL의 각각의 시점을 10 μL의 변성 로드 완충제 (100% 포름아미드 및 0.1% 오렌지 (Orange) G))와 혼합하고 1 mm x 20 cm x 14 cm 20% 폴리아크릴아미드 겔의 웰에 적용하였다. 400 V에서 3.5

시간 동안 전기영동 후, 밴드를 Sybr 골드(Sybr Gold) (써모 사이언티픽)로 가시화하고 UV 조명 (UV-차단 래튼 (Wratten) 2A 필터 405 nm 컷오프, 유브이피, 엘엘씨(UVP, LLC))하에 촬영하였다.

[0067] 다중 동중중합체 트랙의 합성을 위해, 개시제 올리고뉴클레오티드를 이전 반응물 및 세척액의 제거와 함께 다중 라운드의 효소적 합성이 산재될 수 있도록 비드에 부착시킬 수 있었다. 5'-비오틴-TAATAATAATAATTTT-3' (ID T)는 스트랩타비딘 코팅 자성 세파로스 마이크로비드 (지이 헬쓰케어 라이프 사이언시즈(GE Healthcare Life Sciences))와 함께 인큐베이션할 수 있었다. 올리고뉴클레오티드-충전된 비드는 비드 슬러리의 분취액을 제거하고 필터 컵으로 옮김으로써 제조할 수 있었다. 이어서, 비드를 1x PBS (2x 부피의 비드 슬러리를 사용) 볼텍스로 5회 세척하고 각각의 세정을 스피ندا운(spin down)할 수 있었다. 이어서, 1x PBS의 1/2 비드 슬러리 부피를 첨가할 수 있고, 비오틴닐화된 올리고뉴클레오티드를 1/10의 공개된 비드 결합능으로 스파이크할 수 있었다. 혼합물을 30분마다 볼텍싱하면서 2시간 동안 37°C에서 인큐베이션할 수 있었다. 2시간 후, 소량의 상청액을 제거할 수 있고 A260은 임의의 비결합 올리고뉴클레오티드에 대해 측정할 수 있었다. 일단 A260이 <10% 잔류 올리고뉴클레오티드를 나타내면, 비드를 MQ 수로 5x 세척할 수 있었다. 세척된 비드는 MQ 수에서 원하는 농도로 올릴 수 있었다.

[0068] 동중중합체 합성은 비드 결합된 올리고뉴클레오티드에 대해 2-10x 몰 당량의 원하는 dNTP를 사용하여 상기에 기재된 바와 같이 수행할 수 있었다. 비드, dNTP, TdT 및 완충제를 함유하는 반응 혼합물을 15분 동안 37°C에서 인큐베이션할 수 있었다. 반응을 10 μl EDTA로 정지시키고 물로 3x 세정하였다. 새로운 TdT 효소, dNTP, 완충제를 첨가함으로써 새로운 사이클의 동중중합체 합성을 개시하고 15분 동안 37°C에서 인큐베이션할 수 있었다. EDTA와의 반응을 정지시키고 물로 3x 세정한 후; 동중중합체 합성의 사이클은 원하는 만큼의 횟수로 반복할 수 있었다. 마지막 EDTA 쉼터 및 물로 3x 세정한 후, 지지체-결합 교호 동중중합체는 100 μl 농축 압모늄을 사용하여 고체 지지체로부터 절단될 수 있었고, 상청액은 Gen-vac에 의해 건조시킨 다음에, 상기 기재된 바와 같이 폴리아크릴아미드 겔을 사용하여 분석할 준비가 될 때까지 -20°C에서 보관할 수 있었다.

[0069] 참조에 의한 포함

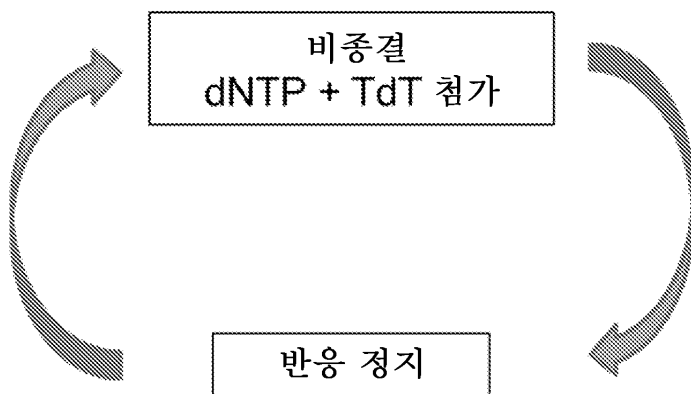
[0070] 특허, 특허 출원, 특허 간행물, 저널, 서적, 논문, 웹 콘텐츠와 같은 다른 문서에 대한 언급 및 인용이 본 개시 내용에 걸쳐 이루어졌다. 이러한 모든 문서는 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0071] 등가물

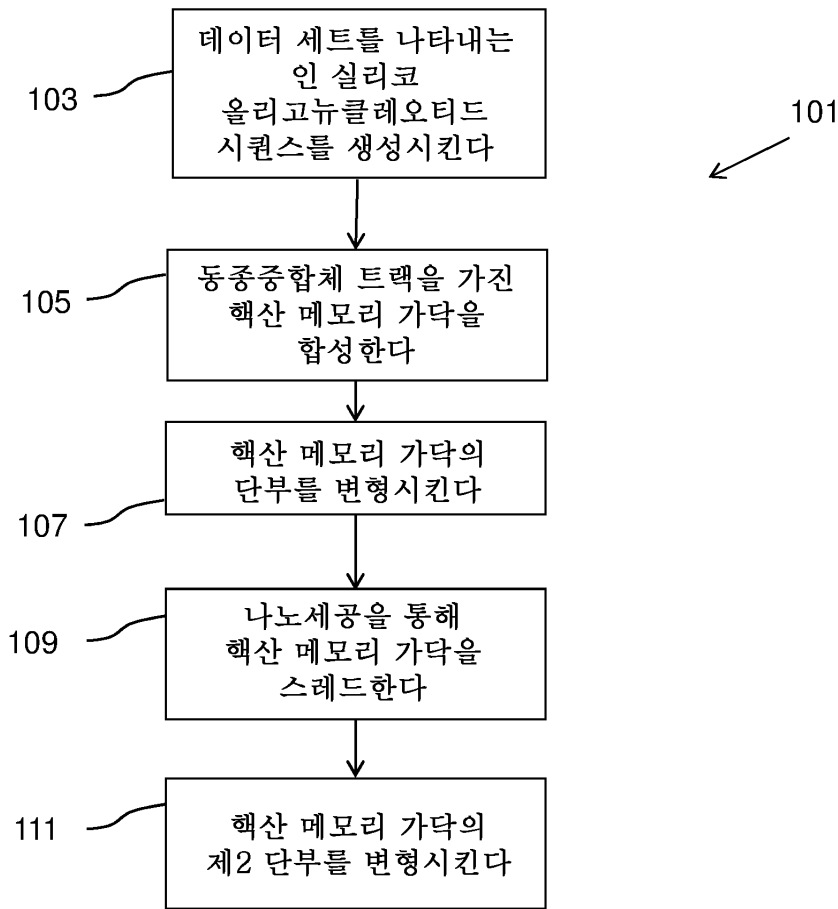
[0072] 본원에 도시되고 기재된 것들에 더하여, 본 발명의 다양한 변형 및 그의 많은 추가 실시양태는, 여기에 인용된 과학 및 특허 문헌에 대한 언급을 포함하여, 본 문서의 전체 내용으로부터 관련 기술분야의 통상의 기술자에게 명백해질 것이다. 본원의 주제는 그의 다양한 실시양태 및 그의 등가물에서 본 발명의 실시예에 적용될 수 있는 중요한 정보, 예시 및 안내를 함유한다.

도면

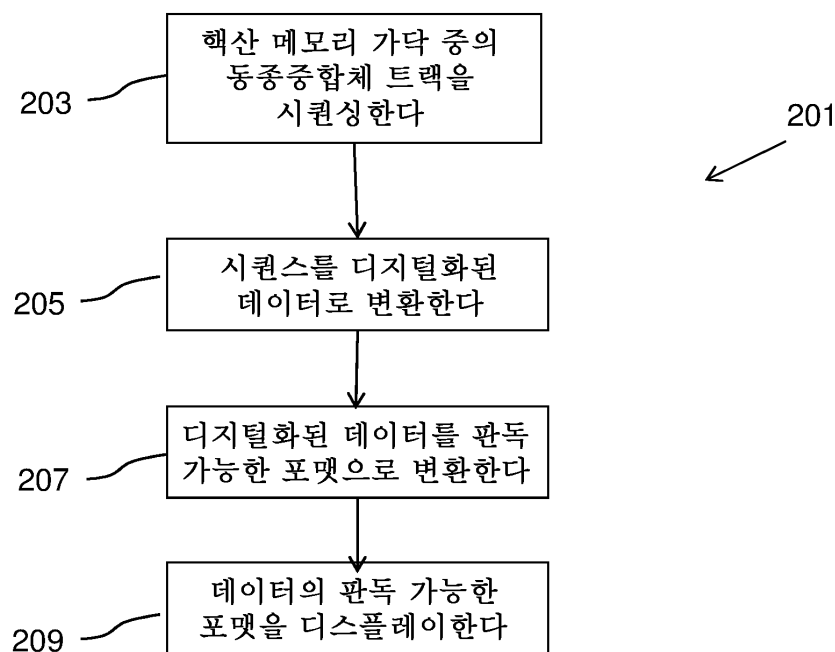
도면1



도면2



도면3



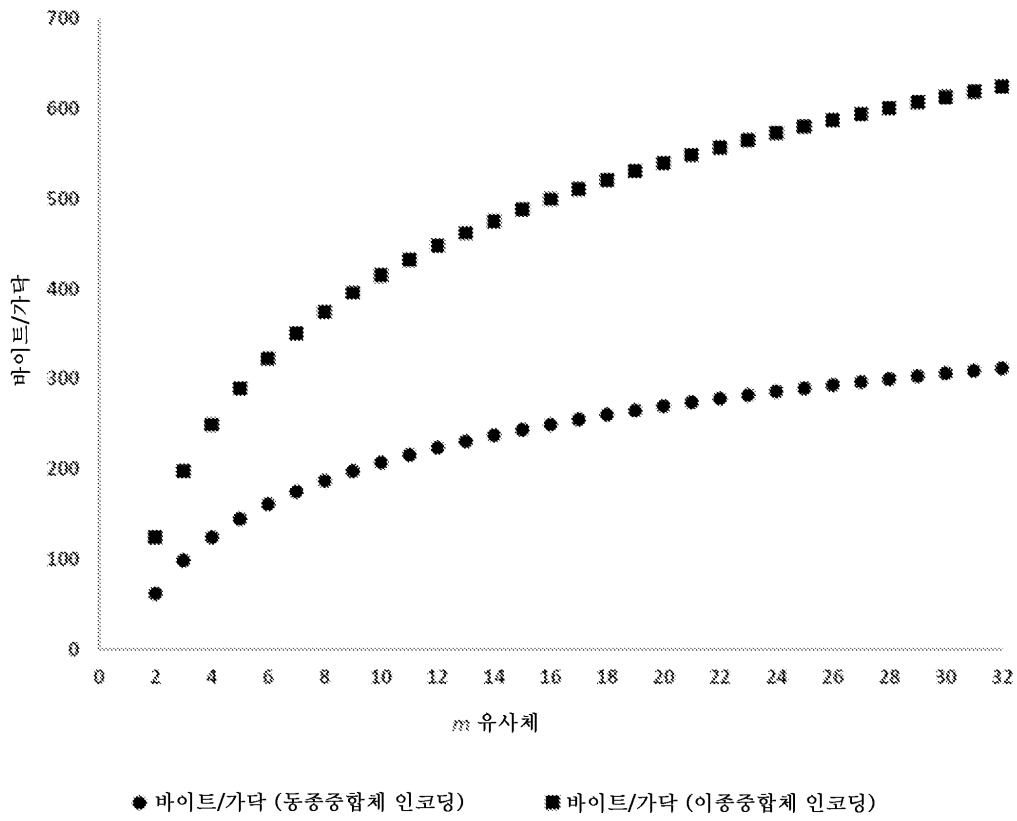
도면4

베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#
2	2	500	3.2E+07	2	5	500	8.8E+07	2	10	500	1.6E+08
4	2	500	1.8E+07	4	5	500	4.8E+07	4	10	500	8.8E+07
6	2	500	1.2E+07	6	5	500	3.1E+07	6	10	500	6.2E+07
8	2	500	1.1E+07	8	5	500	2.7E+07	8	10	500	5.3E+07
12	2	500	8.8E+06	12	5	500	2.2E+07	12	10	500	4.5E+07
16	2	500	8.2E+06	16	5	500	2.2E+07	16	10	500	4.2E+07

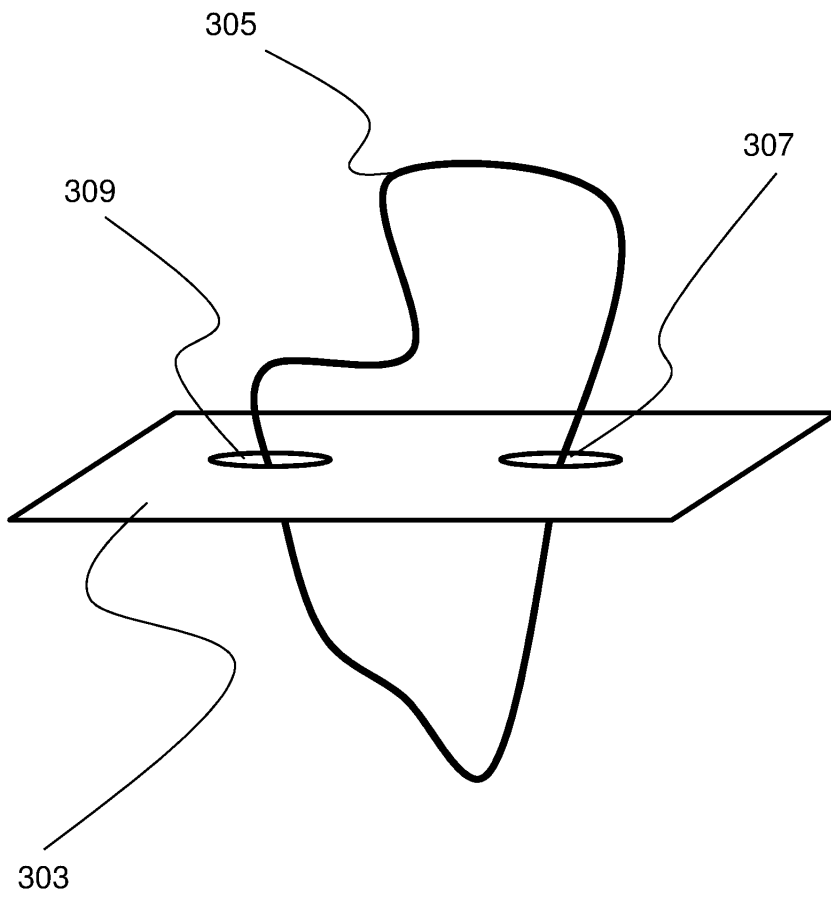
베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#
2	2	1000	1.8E+07	2	5	1000	4.8E+07	2	10	1000	8.8E+07
4	2	1000	8.8E+06	4	5	1000	2.8E+07	4	10	1000	4.8E+07
6	2	1000	6.2E+06	6	5	1000	1.5E+07	6	10	1000	3.1E+07
8	2	1000	5.3E+06	8	5	1000	1.2E+07	8	10	1000	2.7E+07
12	2	1000	4.5E+06	12	5	1000	1.1E+07	12	10	1000	2.1E+07
16	2	1000	4.2E+06	16	5	1000	1.2E+07	16	10	1000	2.2E+07

베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#	베이스#	동공층별 간이	가 나	*가 나 /#
2	2	10000	1.8E+06	2	5	10000	4.8E+06	2	10	10000	8.8E+06
4	2	10000	8.8E+05	4	5	10000	2.8E+06	4	10	10000	4.8E+06
6	2	10000	6.2E+05	6	5	10000	1.5E+06	6	10	10000	3.1E+06
8	2	10000	5.3E+05	8	5	10000	1.2E+06	8	10	10000	2.7E+06
12	2	10000	4.5E+05	12	5	10000	1.1E+06	12	10	10000	2.1E+06
16	2	10000	4.2E+05	16	5	10000	1.2E+06	16	10	10000	2.2E+06

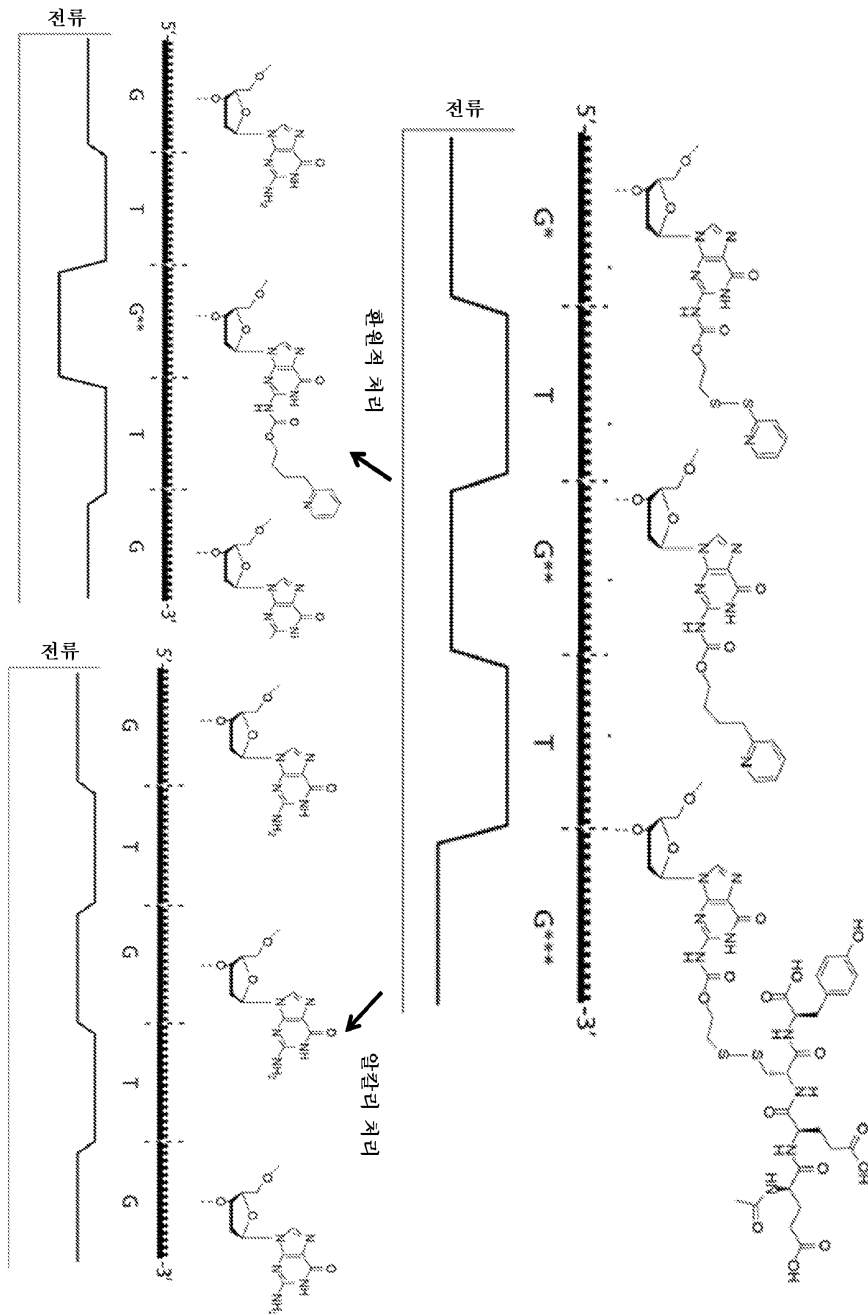
도면5



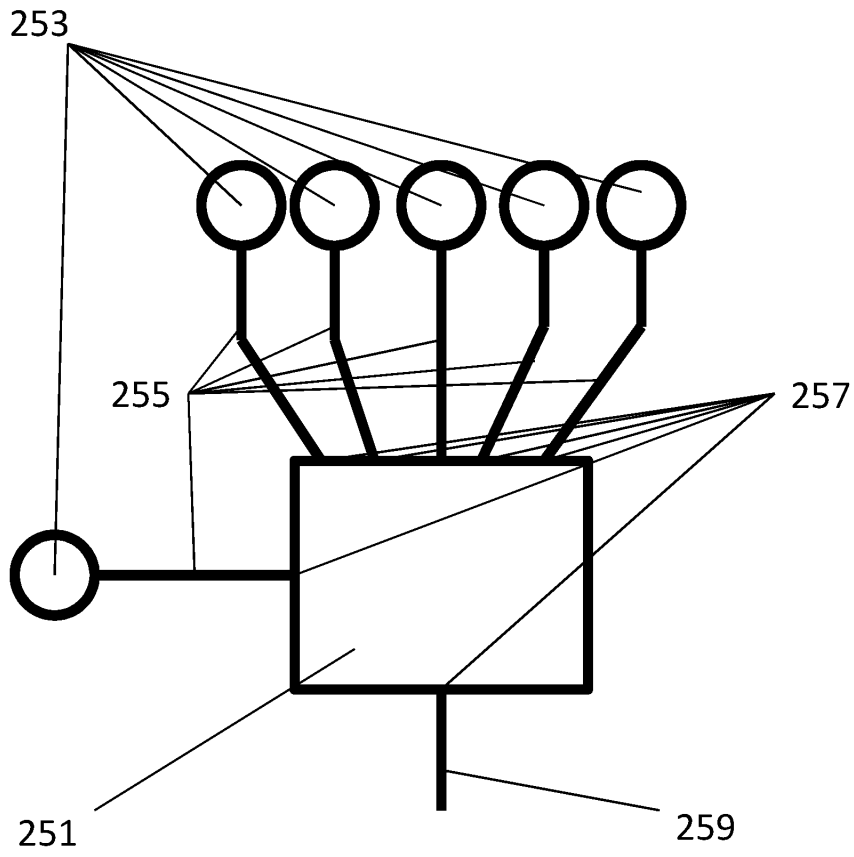
도면6



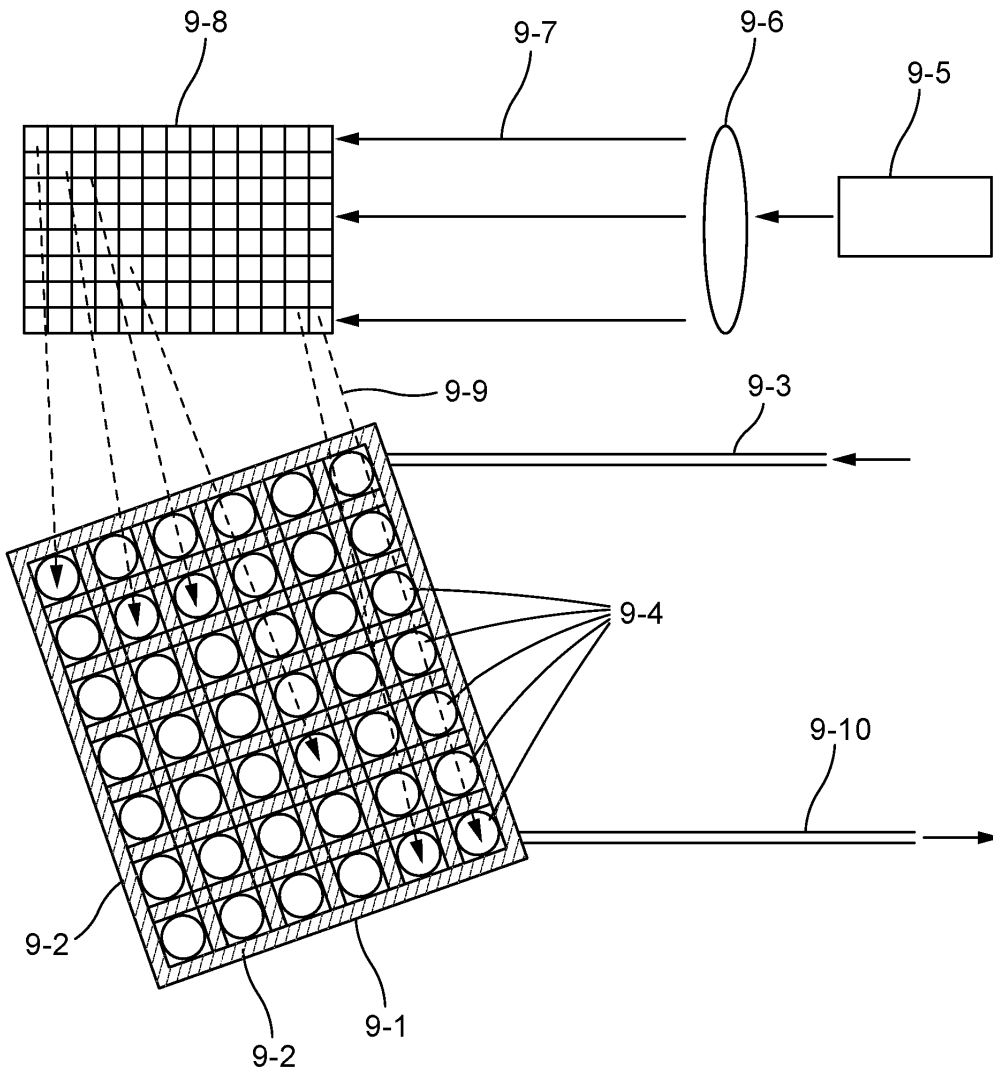
도면7



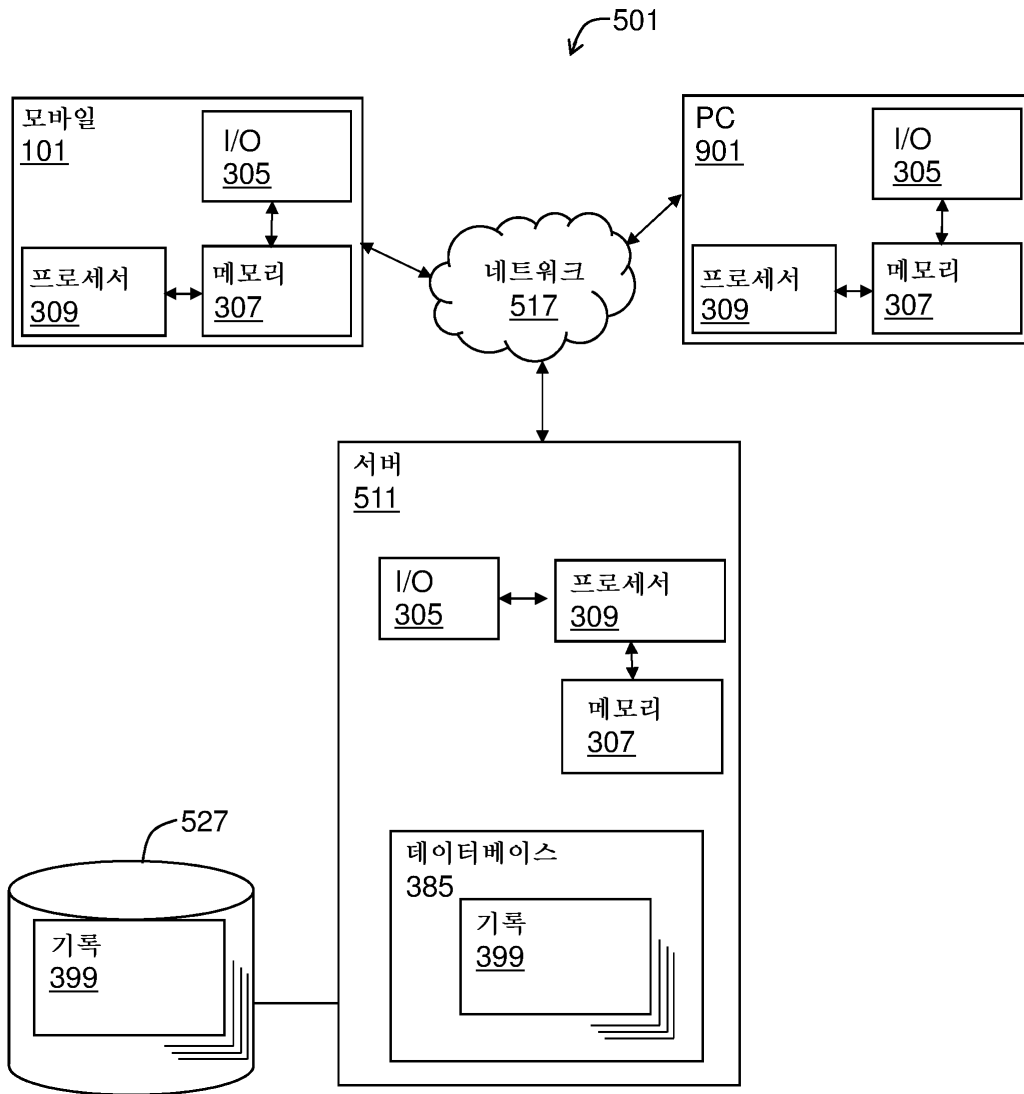
도면8



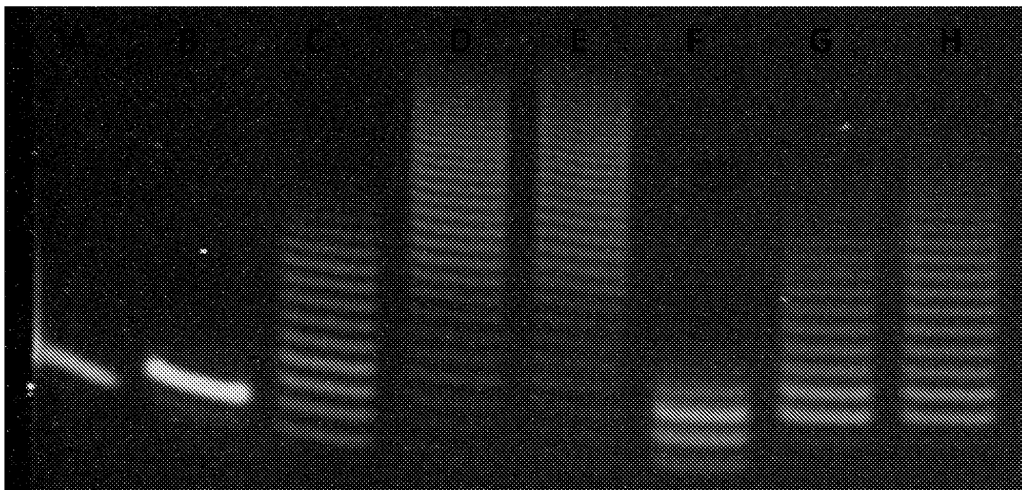
도면9



도면10



도면11



서열목록

SEQUENCE LISTING

<110> Molecular Assemblies, Inc.

Efcavitch, J. William

Holden, Matthew T.

<120> HOMOPOLYMER ENCODED NUCLEIC ACID MEMORY

<130> MOLA-004/01WO 31144/44

<140> PCT/US2018/035365

<141> 2018-05-31

<150> US 62/513,111

<151> 2017-05-31

<160> 1

<170> PatentIn version 3.5

<210> 1

<211> 17

<212> DNA

<213> Synthetic

<400> 1

taataataat aattttt

17